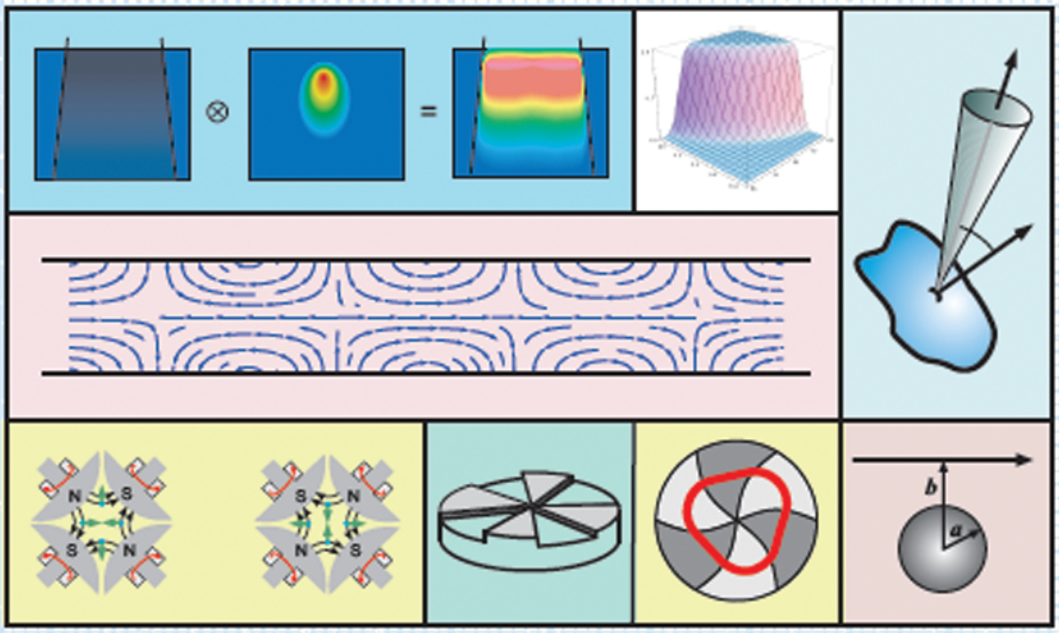
# TUTORIALS IN RADIOTHERAPY PHYSICS

## Advanced Topics with Problems and Solutions




Patrick N. McDermott

CRC Press
Taylor & Francis Group

# TUTORIALS IN RADIOTHERAPY PHYSICS

# TUTORIALS IN RADIOTHERAPY PHYSICS

## Advanced Topics with Problems and Solutions

**Patrick N. McDermott**

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742



Printed on acid-free paper
Version Date: 20151008

International Standard Book Number-13: 978-1-4822-5167-8 (Paperback)



**Library of Congress Cataloging-in-Publication Data**

Names: McDermott, Patrick N., author.
Title: Tutorials in radiotherapy physics : advanced topics with problems and solutions / Patrick N. McDermott.
Description: Boca Raton, FL : CRC Press, Taylor & Francis Group, 2016. | "2016 | Includes bibliographical references and index.
Identifiers: LCCN 2015034513| ISBN 9781482251678 | ISBN 1482251671
Subjects: LCSH: Medical physics--Problems, exercises, etc. | Radiotherapy--Problems, exercises, etc. | Linear accelerators in medicine.
Classification: LCC R895 .M336 2016 | DDC 616.07/57--dc23
LC record available at http://lccn.loc.gov/2015034513

**Visit the Taylor & Francis Web site at**
**http://www.taylorandfrancis.com**

**and the CRC Press Web site at**
**http://www.crcpress.com**

*Monica McDermott*

*My wife, who sees something in me, although I am still not quite sure what it is. This project would not have been possible without her support.*

# CONTENTS

# PREFACE

This book is intended to be a *textbook* covering selected advanced topics that are not covered in depth in any of the standard medical physics texts. The prerequisite is a bachelor's degree in physics and a graduate-level course in radiological physics at the level found in the textbooks by Attix (1986), Khan and Gibbons (2014), Podgorsak (2010), and Metcalfe et al. (2007). A course in radiobiology would be helpful for Chapter 5 but is not essential. This book is not meant to be a comprehensive review of the topics presented. Review articles already exist for some (but not all) of these topics, and they are summaries of current research. Fundamental, systematic, mathematical development is missing in such articles. This book is the place where one should begin when learning these topics. Reading review articles should then follow.

Many of these important topics are not well known or understood even by board-certified medical physicists with years of experience. Reading the scant existing literature is an exercise in frustration. While there are a few notable exceptions, problems are all too common: poor, nonexistent, or incomplete derivations and explanations; numerous errors; poor equation typography; failure to clearly and carefully define all mathematical symbols (if they are defined at all); internal contradictions or inconsistencies; nonsequiturs; and so forth. Physics is difficult enough to learn without these obstacles. I certainly do not claim perfection, but I have been very careful to try to minimize these problems. My approach is to start from first principles rather than writing down "immaculately conceived" equations whose provenance is mysterious. The material in this book cannot be found in any one place. It is a synthesis from a large variety of sources using consistent mathematical notation.

This book can be used for self-study, or it could be used by students in graduate medical physics programs or physics residency programs. It may also be of use in vendor training for linacs or treatment planning systems. The chapters can be read independently, although there are some links between them. It is my contention that a single course in radiation therapy physics (à la Khan) is not adequate for radiation therapy physicist preparation. Indeed, many graduate programs do have an advanced course. There are efforts underway by the American Association of Physicists in Medicine to improve teaching of graduate medical physics students. One of the biggest problems, however, is a lack of well-written texts covering topics that medical physicists need to know. This was frustrating for me as a student, and it continues to be a problem today.

Problems and questions appear at the end of each chapter. The solutions to the problems are given in the back of the book. I urge you to give the problems considerable thought before looking at the solutions. The struggle to solve problems is, after all, one of the ways that we learn. I believe that if you can solve the problems, you will have a good quantitative understanding of the material.

Chapter 1, "The Physics of Electron Acceleration in Medical Linacs," is for those physicists who are tired of comic book explanations about electrons surfing on waves and desire a real understanding of the mechanism of microwave acceleration of electrons. Commonly cited books in the medical physics literature do not provide a satisfactory explanation of this topic. The monograph by Karzmark, Nunan, and Tanabe (*Medical Electron Accelerators*) does provide some explanation; however, it is from an engineering perspective, and this book is now out of print. Basic physics textbooks on electromagnetic theory (e.g., Jackson) do not extend the discussion of waveguides to a consideration of their use in accelerating charged particles. The accelerator physics literature helps some, but the discussion there is concentrated on high-energy physics applications, not medical linacs. Medical linacs are different than those used for other applications. They have to be able to produce both x-ray beams and electron beams. They usually offer multiple beam energies, and they are mounted on a gantry that can rotate around a patient.

The study of proton therapy (Chapter 2) is perhaps the one exception to the critique of the existing literature described thus far. There are a number of monographs on this subject, and some of these are well written. There is no shortage of good references for this topic. Chapter 2 is a concise introduction for physicists that I believe is more thorough than presented in any of the introductory radiation therapy textbooks. The coverage here is intermediate between elementary textbooks and that found in extensive monographs.

Convolution and superposition dose calculation algorithms are now the predominant method of clinical photon dose computation. This topic is the subject of Chapter 3. There are a few older review articles on this subject, and it is briefly discussed in some of the elementary textbooks. These discussions are brief and sketchy. To gain a good understanding would require reading various reviews and original papers and synthesizing these sources. That is what I have done for you, the reader.

The common view of dose calculation algorithms held by many physicists is that there is Monte Carlo, the gold standard, and then there is everything else. If Monte Carlo is the gold standard, then perhaps the Boltzmann transport equation is the silver standard. Chapter 3 discusses this method of dose calculation. It is a little known but important topic because it is (1) a technique that rivals the accuracy of Monte Carlo but is much faster, (2) an option in a widely used commercial treatment planning system, (3) much faster than collapsed cone convolution for VMAT dose calculations, and (4) the basis for Fermi–Eyges theory, which is used in many commercial electron dose calculation algorithms. Electron pencil beam calculations are generally based on the solution of the Fermi–Eyges equation, which we will derive from the Boltzmann transport equation.

The title of Chapter 5 is "Tumor Control and Normal Tissue Complication Probability Models in Radiation Therapy." Although this topic is not classified as physics, I have included it because physicists have been heavily involved in the development of this subject and because it usually falls to the clinical physicist to understand and use these models. There are some excellent review articles on this

topic, but one is hard pressed to find literature that systematically develops this topic mathematically in a step-by-step fashion.

The American Association of Physicists in Medicine (AAPM) Report No. 249, "Essentials and Guidelines for Clinical Medical Physics Residency Training Programs," specifies that the physics resident "demonstrates an understanding of the theory of operation of megavoltage electron and proton accelerators currently used in radiation oncology treatment… e.g., linear accelerators (linacs), synchrotrons and cyclotrons." This could be fulfilled by reading and solving the end of chapter problems in Chapters 1 and 2. In addition, this report recommends that the physics resident "describes how the computer algorithm calculates dose for at least one major treatment planning system with regard to: i. Photon beams, ii. Electron beams." This is covered in Chapter 3 for photon beams and in Section 4.10 of Chapter 4 for electron beams. Report No. 249 calls for coverage of "biological evaluators (e.g., generalized equivalent uniform dose [gEUD], equivalent uniform dose [EUD], normal tissue complication probability [NTCP], and tumor control probability [TCP])." These topics are covered in Chapter 5.

What gives me the "nerve" to write a book on these topics when I have never published a single paper on any of them? I wrote this material so that I could learn about these topics myself. I turned it into a book because no one else is doing it. Often, the expert is not the best person to explain a subject. I am convinced that none of these topics are difficult to learn or understand, but they have not been explained well. You are the judge of whether I have been successful.

The superficial explanations given in elementary textbooks of the topics of this book are unsatisfying. One reason that individuals choose to study physics is because they enjoy understanding nature on a very fundamental level. This book is for people who like physics. In short, this is the book I wish I had when I was a student.

I welcome feedback and comments, and I will try to answer all e-mail communications, time permitting.

**Patrick McDermott, PhD**
*patrick.mcdermott@beaumont.edu*

## REFERENCES

Attix, F. 1986. *Introduction to Radiological Physics and Radiation Dosimetry*. New York: Wiley-Interscience.

Khan, F.M. and J.P. Gibbons. 2014. *The Physics of Radiation Therapy*, 5th ed. Philadelphia, Wolters-Kluwer.

Metcalfe, P., T. Kron, and P. Hoban, P. 2007. *The Physics of Radiotherapy X-Rays and Electrons*. Madison, WI: Medical Physics Publishing.

Podgorsak, E.B. 2010. *Radiation Physics for Medical Physicists*, 2nd ed. Berlin: Springer-Verlag.

# ACKNOWLEDGMENTS

I wish to thank Michael Derr for a careful review of Chapter 1 on linacs. Thanks to Leo Ding for his expert review of Chapter 2 on proton therapy. Thanks to Almut Troeller for a careful review of Chapter 5 and Colin Orton for encouraging comments about Chapter 5. Thanks to Di Yan for answering questions about NTCP and TCP modeling. Thanks to Luna Han, my editor, for believing in and supporting this project. Thanks to Michelle van Kampen, my editor, for tolerating my endless nit-picking. Finally, I would like to thank my wife Monica for her forbearance while I did the research and writing. Any mistakes in this book are mine alone.

# AUTHOR

**Patrick N. McDermott, PhD,** is the director of physics education at Beaumont Health in Royal Oak, Michigan, and an adjunct associate professor at Oakland University in Rochester, Michigan. Prior to accepting that position in 2005, he was an associate professor in the Department of Radiation Oncology at Wayne State University and a senior staff physicist at the Karmanos Cancer Institute in Detroit, Michigan. He taught in the graduate medical physics program at Wayne State for over ten years. He has won a number of teaching awards and was elected a fellow by the American Association of Physicists in Medicine in 2015. He earned a PhD in physics and astronomy from the University of Rochester in Rochester, New York, and an MS in radiological physics from Wayne State University. He is board certified by the American Board of Medical Physics in radiation oncology physics.

# 1

# THE PHYSICS OF ELECTRON ACCELERATION IN MEDICAL LINACS

## 1.1 INTRODUCTION

Let us estimate the sort of energy and beam current necessary to achieve a desirable penetration depth and dose rate. We first consider the beam energy. The deepest target in the human body is in the pelvis. Let us suppose that we need to reach a depth of 30 cm with no more than 50% beam attenuation. In this case, we need photons with energy such that the half-value layer thickness equals 30 cm in water. This corresponds to photons with an energy of approximately 5 MeV. The average energy in a bremsstrahlung spectrum is about one-third of the maximum energy. Therefore, the electron beam energy needs to be about 15 MeV. We would not want to use beam energies much larger than this because this would waste photons that would pass right through the patient and because we want the beam to be attenuated beyond the target to reduce the dose to normal tissues.

As the reader may already know, linac beams are pulsed (see Section 1.9). The beam currents quoted in this section are time averages over a timescale long compared to the individual pulses. Calculation of the necessary average beam current is most easily done for electron treatments. For photon treatments, we have the question of the efficiency of bremsstrahlung photon production in the target and the transmission of the flattening filter. Modern medical linear accelerators have a dose rate on the order of 600 cGy/min. The dose rate due to electrons is

$$\dot{D}\left[\mathrm{Gy\,s^{-1}}\right] = 1.6\times10^{-10}\dot{\Phi}\left[\mathrm{cm^{-2}\,s^{-1}}\right]\left(\frac{dT}{\rho dx}\right)_c\left[\mathrm{MeV/cm^{-2}\,g^{-1}}\right], \tag{1.1}$$

where $\dot{\Phi}$ is the fluence rate in $cm^{-2}\ s^{-1}$ and the mass collision stopping power is given in units of MeV $cm^2\ g^{-1}$ (Attix, 1986). For 10 MeV electrons, the mass collision stopping power is 2.0 MeV $cm^2\ g^{-1}$ in water. To obtain the desired dose rate, the fluence rate must be $3.2 \times 10^8\ cm^{-2}\ s^{-1}$. Let us suppose that the beam is 10 cm × 10 cm in cross-sectional area. In this case, we can calculate the necessary average beam current to be approximately 5 nA. In reality, the average beam current is actually considerably higher than this for the following reasons. To reduce the variation in the energy of the accelerated electrons, the beam passes through an energy slit in the bending magnet. According to Karzmark et al. (1993, p. 190), this energy slit may pass as little as 5% of the beam. This raises the necessary raw beam current for electron mode to approximately 0.1 μA. In addition, the electron gun current needs to be even higher than the beam current because only some fraction of the electrons are captured and accelerated.

For a photon treatment, the beam current needs to be between 100 and 1000 times greater than for an electron treatment. The required average beam current is up to 100 μA. Photon beam current requirements are complicated by the efficiency of x-ray production in the target and attenuation of x-rays in the flattening filter. The dose rate for a flattened x-ray beam (50 cm diameter field) is proportional to $I_{\text{avg}}V_e^{1.8}$, where $I_{\text{avg}}$ is the time-averaged beam current (in x-ray mode) and $V_e$ is the electron beam energy (Karzmark et al., 1993, p. 18). This implies that the average current required to produce a 6 MV beam having the same dose rate as an 18 MV beam is approximately a factor of 7 larger.[1]

If we think about how to accelerate electrons to high energy, we might naively first consider two charged plates with a potential difference of $V$ between them. For radiation therapy, we need electron energies on the order of 10 MeV. It is very difficult to produce a potential difference of 10 million volts between two electrodes. An attempt to generate such a potential difference is likely to lead to leakage current, corona discharge, and ultimately arcing. To avoid this problem, we might modify our initial thought and be a little more clever by accelerating the electrons repeatedly through a moderate potential difference, thereby accumulating a large energy. We could consider a series of plates with holes in them to allow the electrons to go through them as in Figure 1.1. Suppose that we place an oscillating potential difference across these plates. We ignore for now the problem of initial acceleration up to relativistic energies and assume that the electrons are traveling at the speed of light. Let us suppose that the distance between adjacent accelerating plates is $L$. The wavelength of the oscillating potential difference should be $\lambda = 2L$, and the corresponding frequency is $\nu = c/2L$. If the distance between the plates is 5 cm, then the required frequency is 3000 MHz.

This frequency is very high, in the microwave s-band of the electromagnetic spectrum. If we attempt to supply the accelerating voltage using wires (as shown in Figure 1.1), the wires will act like an antenna and radiate. Wires are therefore a very inefficient way to supply the necessary potential difference. Instead, a power transmission waveguide should be used.

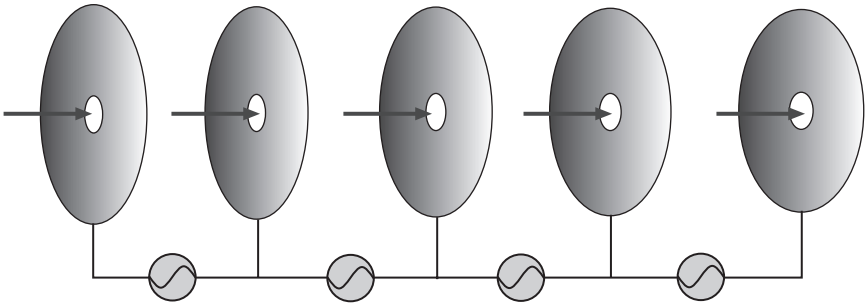


FIGURE 1.1 Initial naive idea for acceleration of electrons to high energies using a series of plates with a high potential difference between plates. The potential difference has to be oscillatory to maintain the correct phase between adjacent plates as the electrons move from left to right.

You will begin to see where we are going with this. It does not require a great leap in imagination to turn the arrangement of electrodes in Figure 1.1 into a waveguide by adding a tube to surround the metal plates. Such a tube will also enable us to provide the necessary vacuum.

It is important to distinguish between an accelerating waveguide and a power transmission waveguide. A power transmission waveguide carries radiofrequency (RF) power to the accelerating waveguide in which electrons are accelerated. Transmission waveguides are filled with a dielectric gas ($SF_6$) to avoid electrical arcing due to the intense electric fields. The accelerating waveguide is under high vacuum to prevent scattering of accelerated electrons. The vacuum pressure should be low enough that the mean free path of an electron is longer than the length of the waveguide. We will concentrate our attention on accelerating waveguides.

The abbreviation TW will be used for traveling wave linacs and SW for standing wave linacs. Elekta linacs are TW, and Varian and Siemens are SW. There is more to a linear accelerator than just a waveguide or resonant cavity—a linac is not just a waveguide. RF power needs to be fed into the waveguide, and for TW linacs, the RF must exit the waveguide. For an SW linac, RF power has to be fed into the guide and then reflected at the ends. In addition, the beam needs to be injected into the accelerator and then extracted. Furthermore, the beam will affect the fields in the waveguide. We first discuss waveguides and cavity oscillators; then in subsequent sections, we discuss how to turn these into accelerators.

Modern medical linacs have two or three x-ray energies and five or six different electron energies. The high-energy x-ray beam is sometimes called high X and the low-energy beam low X.

The various timescales associated with the operation of medical linacs can be confusing. The shortest timescale is associated with the period of the 3000 MHz s-band microwaves at 0.3 ns. Electrons are accelerated in bunches with a time interval between bunches that is on the order of the microwave period. This acceleration occurs only when the RF power is delivered to the waveguide—this occurs

in pulses of approximately 5 μs in duration that repeat at roughly 5 ms intervals. This will be discussed further in Section 1.9. Throughout the text, time averages will be taken over the various timescales.

In Section 1.2, we begin with a review of Maxwell's equations and the necessary boundary conditions. We then apply this to cylindrical waveguides in Section 1.3. Application of this material to traveling wave accelerators is covered in Section 1.4. Cavity oscillations are discussed in Section 1.5. The electromagnetic energy budget is considered in Section 1.6, including energy loss mechanisms. In Section 1.7, traveling wave accelerators are revisited, including a discussion of load lines and RF feedback. Section 1.8 is on standing wave accelerators. The pulsed operation of linacs and waveforms is covered in Section 1.9. The frequency stability and fabrication of waveguides are discussed in Section 1.10. The method used to change beam energy is considered in Section 1.11. Section 1.12 is a discussion of the differences between SW and TW linacs. In the final section, x-band linacs are discussed. Problems and questions appear at the end of the chapter, along with a list of symbols. In this chapter boldface symbols represent vectors. Useful references for this topic are Karzmark et al. (1993), Podgorsak et al. (1999), Podgorsak (2010), Ford (1986), Scharf (1994), Greene and Williams (1997), and Feynman et al. (1963, chapters 23 and 24; pure reading pleasure).

## 1.2 MAXWELL'S EQUATIONS

We begin by writing down Maxwell's equations in SI units:

$$\nabla \cdot \mathbf{D} = \rho \tag{1.2a}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.2b}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{1.2c}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}. \tag{1.2d}$$

We can also write these in integral form as follows:

$$\oint \mathbf{D} \cdot d\mathbf{A} = q \tag{1.3a}$$

$$\oint \mathbf{B} \cdot d\mathbf{A} = 0 \tag{1.3b}$$

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\frac{d}{dt}\int_A \mathbf{B} \cdot d\mathbf{A} \tag{1.3c}$$

$$\oint \mathbf{H} \cdot d\mathbf{l} = \int_A \mathbf{J} \cdot d\mathbf{A} + \frac{d}{dt}\int_A \mathbf{D} \cdot d\mathbf{A}. \tag{1.3d}$$

We will assume constitutive relations for the field vectors as follows:

$$\mathbf{D} = \varepsilon \mathbf{E}$$

$$\mathbf{B} = \mu \mathbf{H}$$

$$\mathbf{J} = \sigma \mathbf{E}. \tag{1.4}$$

### 1.2.1 BOUNDARY CONDITIONS ON FIELD VECTORS

In order to solve for the fields inside the waveguide, we need to know the boundary conditions that must be obeyed by the fields at the inner surface of the waveguide. We will use the integral form of Maxwell's equations (1.3) to derive these. We start with the equation for the displacement vector **D**. Figure 1.2 shows the boundary surface between the waveguide and the vacuum. We construct a "Gaussian pillbox," which spans this surface.

The end caps of the pillbox have a tiny area $a$. The **D** vector is presumed to be constant over the area of the tiny end cap. Applying the first of Maxwell's equations (in integral form [1.3a]) to this pillbox, we obtain

$$\mathbf{D} \cdot \mathbf{n}a - \mathbf{D}_c \cdot \mathbf{n}a + \int_{\text{sides}} \mathbf{D} \cdot d\mathbf{A} = \Sigma a, \tag{1.5}$$

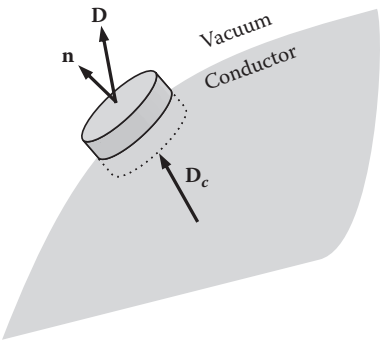



FIGURE 1.2 A Gaussian pillbox spanning the interface between a conductor and vacuum. **n** is a unit vector perpendicular to the surface.

where:

- $\mathbf{D}$ is the value of the displacement vector in the vacuum
- $\mathbf{D}_c$ is the value in the conductor
- $\mathbf{n}$ is a unit vector perpendicular and pointing outward (into the vacuum) from the surface

The surface charge density is Σ. We use this symbol so as to avoid confusion with the conductivity σ. The first two contributions on the left-hand side of Equation 1.5 are due to the end caps of the pillbox, and the third is due to the sides of the cylinder. If we take the limit in which the length of the cylinder shrinks to zero, the contribution to the surface integral from the sides will vanish and we will be left with

$$\mathbf{n}\cdot\left(\mathbf{D}-\mathbf{D}_c\right)=\Sigma. \tag{1.6}$$

In a completely analogous fashion, we can derive a boundary condition for **B** by using the second of the integral form of Maxwell's equations. The result is

$$\mathbf{n}\cdot\left(\mathbf{B}-\mathbf{B}_c\right)=0, \tag{1.7}$$

where:

- $\mathbf{B}$ is the magnetic field in the vacuum
- $\mathbf{B}_c$ is the magnetic field inside the conductor

We can derive a boundary condition for the electric field by using the third of the integral form of Maxwell's equations (1.3c). We construct a rectangular loop spanning the boundary as shown in Figure 1.3. The vector **l** is tangent to the surface and has length $l$, **E** is the electric field vector in vacuum, $\mathbf{E}_c$ is the electric field vector in the conductor, and $h_1 + h_2$ is the side length of the loop. The loop is very tiny, and

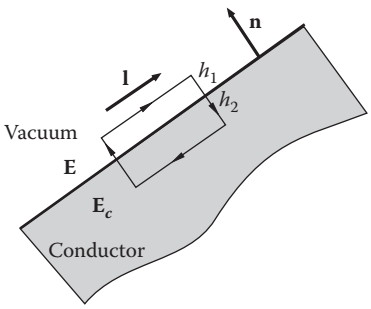



FIGURE 1.3 The loop over which the integral in Equation 1.8 is carried out. **l** is a unit vector tangent to the surface.

therefore **E** is presumed constant over the length of the loop. Applying the third Maxwell equation to the loop gives

$$\oint \mathbf{E}\cdot d\mathbf{l} = \mathbf{E}\cdot\mathbf{l} - \mathbf{E}_c\cdot\mathbf{l} + \mathbf{n}\cdot\left(h_1\mathbf{E} + h_2\mathbf{E}_c\right) - \mathbf{n}\cdot\left(h_1\mathbf{E} + h_2\mathbf{E}_c\right) = -\frac{d}{dt}\int_A \mathbf{B}\cdot\mathbf{dA}. \tag{1.8}$$

If we take the limit as $h_1$ and the $h_2$ values go to zero, the integral on the right-hand side of Equation 1.8 will go to zero as the area of the loop goes to zero. The result is $\mathbf{E}\cdot\mathbf{l} = \mathbf{E}_c\cdot\mathbf{l}$, or in other words, the tangential component of **E** is continuous across the interface. This condition can be written as

$$\mathbf{n}\times\left(\mathbf{E} - \mathbf{E}_c\right) = 0, \tag{1.9}$$

because $\mathbf{n}\times\mathbf{E}$ is perpendicular to **n** and therefore tangential to the surface.

The final boundary condition is obtained from the fourth Maxwell equation (1.3d) by again integrating around a loop as just demonstrated: $\mathbf{H}\cdot\mathbf{l} - \mathbf{H}_c\cdot\mathbf{l} = K$, where $K$ is the surface current; $K = \int \mathbf{J}\cdot d\mathbf{A}$. This condition on **H** can be expressed in vector form as follows:

$$\mathbf{n}\times\left(\mathbf{H} - \mathbf{H}_c\right) = \mathbf{K}. \tag{1.10}$$

For a perfect conductor, $\sigma\to\infty$ and $\mathbf{E}_c = 0$. This implies that $\partial\mathbf{B}_c/\partial t = 0$ from Faraday's law (Equation 1.2c). Therefore, only a static magnetic field can exist inside a perfect conductor. If we assume that $\mathbf{B}_c = 0$ at time $t = 0$, then $\mathbf{B}_c$ will remain zero. Even if there is a nonzero static magnetic field, it will not participate in wave motion. We may therefore conclude from Equation 1.7 that the normal component of **B** is zero just outside the surface of a perfect conductor. The boundary conditions for a perfect conductor are therefore

$$\mathbf{n}\times\mathbf{E} = 0$$

$$\mathbf{n}\cdot\mathbf{B} = 0. \tag{1.11}$$

These boundary conditions are a good approximation for real conductors with high conductivity, but they are inadequate if one wishes to evaluate power losses in a waveguide. We will return to this topic in Section 1.6.

## 1.3 CYLINDRICAL WAVEGUIDES

Maxwell's equations do not allow the existence of electromagnetic waves in free space with a longitudinal component of **E** (i.e., a nonzero component of **E** in the direction of propagation). Therefore, such waves cannot be used to accelerate charged particles in the direction of propagation. As we will see, the situation is different in a waveguide.

Certain wave modes can propagate down the waveguide with a longitudinal component of **E**. Such modes have the potential to accelerate charged particles. It remains to be shown how this can occur.

We assume a cylindrical waveguide in the form of a hollow right circular cylinder (see Figure 1.4). This pipe may be open at both ends or have end caps. We assume that the cylinder is long enough that we may ignore end effects. Initially, we will assume that the walls of the waveguide are perfectly conducting. We will later relax this assumption. The interior of the waveguide is evacuated.

We use cylindrical coordinates as shown in Figure 1.4 with the $z$ axis extending along the length of the cylinder. The inner surface of the cylinder is at $r = R$. We assume a sinusoidal time dependence $e^{-i\omega t}$ for all field vectors. Under these circumstances, Maxwell's equations (1.2) become (in vacuum)

$$\nabla \cdot \mathbf{E} = 0 \tag{1.12a}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.12b}$$

$$\nabla \times \mathbf{E} = i\omega \mathbf{B} \tag{1.12c}$$

$$\nabla \times \mathbf{B} = -\frac{i\omega}{c^2}\mathbf{E}. \tag{1.12d}$$

A wave equation can be derived for **E** and **B** by taking the curl of Equations 1.12c and 1.12d and using the mathematical relation

$$\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2 \mathbf{F}, \tag{1.13}$$

resulting in the wave equations

$$\left(\nabla^2 + \frac{\omega^2}{c^2}\right)\begin{Bmatrix}\mathbf{E}\\ \mathbf{B}\end{Bmatrix} = 0. \tag{1.14}$$

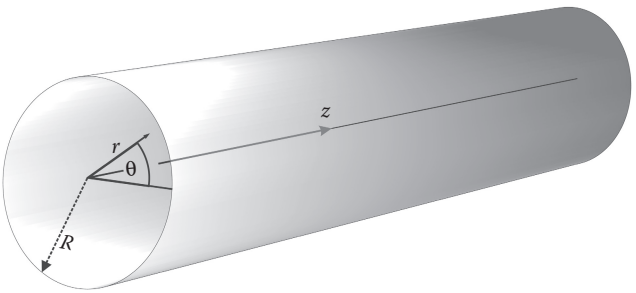



FIGURE 1.4 The geometry of a hollow cylindrical waveguide showing the coordinate system. The waveguide has an inner radius of $R$.

We first consider the case of traveling waves in a hollow, evacuated right circular cylinder with open ends. We can write **E** and **B** in the following form:

$$\mathbf{E}(r,\theta,z,t) = \mathbf{E}(r,\theta)e^{i(\pm kz-\omega t)}$$
$$\mathbf{B}(r,\theta,z,t) = \mathbf{B}(r,\theta)e^{i(\pm kz-\omega t)}, \tag{1.15}$$

where the plus sign is for wave propagation toward +$z$ and the minus sign is for propagation toward –$z$. Linear combinations of these expressions may be formed to give standing waves in the $z$ direction. We will consider this aspect later in the context of a discussion of microwave cavities.

It is useful to write

$$\nabla^2 = \nabla_t^2 + \frac{\partial^2}{\partial z^2},$$

where $\nabla_t^2$ is the piece of the Laplacian involving derivatives of transverse coordinates $r$ and $\theta$. In this case, we may write the wave Equations 1.14 as

$$\left[\nabla_t^2 + \gamma^2\right]\begin{Bmatrix}\mathbf{E}\\ \mathbf{B}\end{Bmatrix} = 0, \tag{1.16}$$

where

$$\gamma^2 = \frac{\omega^2}{c^2} - k^2; \tag{1.17}$$

$k$ is unknown and may be real or complex.

We now wish to show that once the longitudinal components of the fields ($E_z$ and $B_z$) are known, all other components can be easily calculated directly from them.

In cylindrical coordinates, for an arbitrary vector field **A**,

$$\nabla\times\mathbf{A} = \mathbf{e}_r\left(\frac{1}{r}\frac{\partial A_z}{\partial\theta} - \frac{\partial A_\theta}{\partial z}\right) + \mathbf{e}_\theta\left(\frac{\partial A_r}{\partial z} - \frac{\partial A_z}{\partial r}\right) + \mathbf{e}_z\left(\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial A_r}{\partial\theta}\right), \tag{1.18}$$

where $\mathbf{e}_r$, $\mathbf{e}_\theta$, and $\mathbf{e}_z$ are unit vectors in the $r$, $\theta$, and $z$ directions and $A_r$, $A_\theta$, and $A_z$ are the components of vector **A**. The components of Faraday's law (Equation 1.12c) can be written

$$\frac{1}{r}\frac{\partial E_z}{\partial\theta} \mp ikE_\theta = i\omega B_r \tag{1.19a}$$

$$\pm ikE_r - \frac{\partial E_z}{\partial r} = i\omega B_\theta \tag{1.19b}$$

$$\frac{\partial}{\partial r}(rE_\theta) - \frac{\partial E_r}{\partial \theta} = i\omega B_z. \tag{1.19c}$$

Likewise, the components of Ampère's law (Equation 1.12d) can be written as

$$\frac{1}{r}\frac{\partial B_z}{\partial \theta} \mp ikB_\theta = -\frac{i\omega}{c^2}E_r \tag{1.20a}$$

$$\pm ikB_r - \frac{\partial B_z}{\partial r} = -\frac{i\omega}{c^2}E_\theta \tag{1.20b}$$

$$\frac{\partial}{\partial r}(rB_\theta) - \frac{\partial B_r}{\partial \theta} = -\frac{i\omega}{c^2}E_z. \tag{1.20c}$$

If we substitute $E_r$ from Equation 1.19b into Equation 1.20a, we can solve for $B_\theta$:

$$B_\theta = \frac{ik}{\gamma^2}\left(\pm\frac{1}{r}\frac{\partial B_z}{\partial \theta} + \frac{\omega}{c^2 k}\frac{\partial E_z}{\partial r}\right). \tag{1.21}$$

If we substitute $E_\theta$ from Equation 1.20b into Equation 1.19a, we can solve for $B_r$:

$$B_r = \frac{i\omega}{c^2\gamma^2}\left(\pm\frac{kc^2}{\omega}\frac{\partial B_z}{\partial r} - \frac{1}{r}\frac{\partial E_z}{\partial \theta}\right). \tag{1.22}$$

We can find expressions for $E_\theta$ and $E_r$ in a similar fashion:

$$E_\theta = \frac{\pm ik}{\gamma^2}\left(\frac{1}{r}\frac{\partial E_z}{\partial \theta} \mp \frac{\omega}{k}\frac{\partial B_z}{\partial r}\right) \tag{1.23}$$

$$E_r = \frac{\pm ik}{\gamma^2}\left(\frac{\partial E_z}{\partial r} \pm \frac{\omega}{k}\frac{1}{r}\frac{\partial B_z}{\partial \theta}\right). \tag{1.24}$$

We need only solve Equation 1.16 for $E_z$ and $B_z$, and Equations 1.21 through 1.24 will then provide all other components of **E** and **B**.

For a perfect conductor the boundary conditions at the surface are given by Equations 1.11. The first of these equations implies that $E_z = 0$ and $E_\theta = 0$ at $r = R$,

and the second implies that $B_r = 0$ at the inner surface of the waveguide. Substituting these values into Equation 1.20b shows that $\partial B_z/\partial r = 0$. Thus, the boundary conditions for the $z$ components of the field vectors for perfect conductivity are

$$E_z(R,\theta) = 0$$
$$\left.\frac{\partial B_z}{\partial r}\right|_{r=R} = 0. \tag{1.25}$$

The differential equation (1.16) for **E** is the same as the differential equation for **B**, but the two boundary conditions above are different. There are two broad categories of solutions, one associated with each of the conditions in Equation 1.25. These are listed below as follows:

1. Transverse electric (TE)

$$E_z = 0 \text{ everywhere}$$
$$\left.\frac{\partial B_z}{\partial r}\right|_{r=R} = 0 \tag{1.26}$$

2. Transverse magnetic (TM)

$$E_z = 0 \text{ at } r = R$$
$$B_z = 0 \text{ everywhere.} \tag{1.27}$$

The transverse electric (TE) solution has $E_z = 0$ everywhere and is therefore not relevant to particle acceleration, although it is relevant to power transmission. We will not consider this class of solutions any further.

For the TM modes, $B_z = 0$, and therefore the magnetic field is transverse (hence the name). This class of solutions permits a longitudinal electric field that can be used to accelerate charged particles along the direction of wave propagation.

The equation that we wish to solve (Equation 1.16) for the TM modes is

$$\left(\nabla_t^2 + \gamma^2\right)E_z = 0, \tag{1.28}$$

with the boundary condition $E_z = 0$ at $r = R$. Inserting the form of the Laplacian, $\nabla_t^2$, into Equation 1.28 gives

$$\frac{\partial^2 E_z}{\partial r^2} + \frac{1}{r}\frac{\partial E_z}{\partial r} + \frac{1}{r^2}\frac{\partial^2 E_z}{\partial \theta^2} + \gamma^2 E_z = 0. \tag{1.29}$$

Equation 1.29 is a partial differential equation for $E_z$, and we will attempt to solve it using the method of separation of variables. We seek a solution of the form: $E_z(r, \theta, z, t) = \Re(r)\,\Theta\,(\theta)e^{i(\pm kz-\omega t)}$. Substitution of this into Equation 1.29 yields

$$\frac{r^2}{\Re}\frac{d^2\Re}{dr^2}+\frac{r}{\Re}\frac{d\Re}{dr}+\frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2}+\gamma^2r^2=0. \tag{1.30}$$

Following the technique of separation of variables, we reason that if this equation is valid for all values of $r$ and $\theta$, the term on the left-hand side, which depends only on $\theta$ (the second derivative of the $\Theta$ term), must actually be equal to a constant if Equation 1.30 is to be valid for *all* values of $\theta$. Let us call this constant $-m^2$. We may therefore write

$$\frac{d^2\Theta}{d\theta^2}+m^2\Theta=0, \tag{1.31}$$

and then Equation 1.30 becomes

$$\frac{d^2\Re}{dr^2}+\frac{1}{r}\frac{d\Re}{dr}+\left(\gamma^2-\frac{m^2}{r^2}\right)\Re=0. \tag{1.32}$$

Equation 1.31 is simply a harmonic oscillator equation with the solution

$$\Theta(\theta)=A\cos m\theta+B\sin m\theta, \tag{1.33}$$

where $A$ and $B$ are constants to be determined from the boundary conditions and $m$ describes the azimuthal variation in $E_z$.

Equation 1.32 is Bessel's equation of order $m$. This is a well known differential equation in mathematical physics, and it arises quite frequently in problems involving cylindrical symmetry. The solutions of this equation are a linear combination of Bessel and Neumann functions:

$$\Re(r)=CJ_m(\gamma r)+DN_m(\gamma r), \tag{1.34}$$

where:

$J_m$ is the Bessel function of order $m$
$N_m$ is the Neumann function of order $m$
$C$ and $D$ are constants

Bessel functions have well-known properties, and numerical values can be found in many standard mathematical tables (Arfken, 1985). The Neumann function $N_m$ is

singular at $r = 0$, and therefore the constant $D = 0$ in Equation 1.34. The complete solution for $E_z$ is thus

$$E_z = \Re(r)\Theta(\theta)e^{i(\pm kz-\omega t)} = CJ_m(\gamma r)\left[A\cos m\theta + B\sin m\theta\right]e^{i(\pm kz-\omega t)}. \tag{1.35}$$

The boundary condition in Equation 1.27 implies that $J_m(\gamma R) = 0$, and therefore $\gamma$ can only take on certain discrete values $\gamma_n$. Let $x_{mn} = \gamma_n R$ = the $n$th zero (or root) of the $m$th-order Bessel function. We may therefore write $\gamma_n = x_{mn}/R$. In Equation 1.35, the $m$ values describe the azimuthal variation in $E_z$ and $n$ represents the number of times that $E_z = 0$ between $r = 0$ and $r = R$. These transverse magnetic modes are designated as $TM_{mn}$ modes. The lowest-frequency $TM_{mn}$ mode is for $m = 0$, $n = 1$ ($TM_{01}$). This is the mode that is used in TW linacs, and therefore we will concentrate our attention on this mode. The zeros of Bessel functions are tabulated in standard mathematical compilations (Beyer, 1978). The first zero of the Bessel function with $m = 0$ has the value $x_{01} = 2.405$. Equation 1.35 for the $TM_{01}$ mode can be written

$$E_z = E_0 J_0\left(\frac{x_{01}r}{R}\right)e^{i(\pm kz-\omega t)}, \tag{1.36}$$

where $E_0$ is the amplitude of the electric field in the $z$ direction. Notice that there is no $\theta$ dependence because $m = 0$. A graph of $J_0$ is shown in Figure 1.5.

We can calculate the other components of the fields from Equations 1.21 through 1.24:

$$E_r = \frac{\mp ik}{\gamma_1}E_0 J_1\left(\frac{x_{01}r}{R}\right)e^{i(\pm kz-\omega t)} \tag{1.37}$$

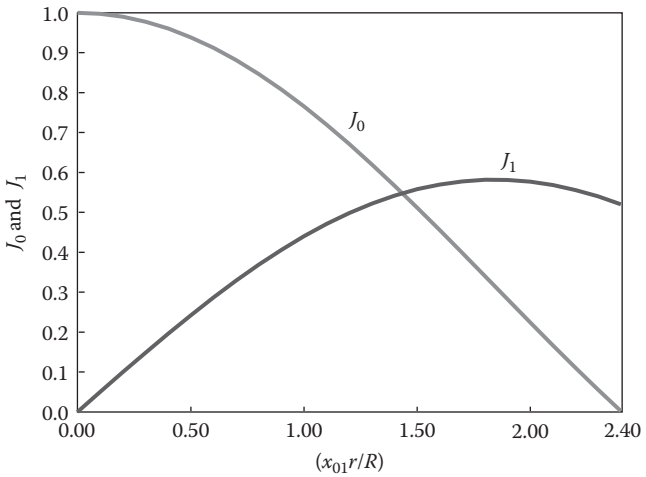



FIGURE 1.5 Graphs of Bessel functions of orders 0 and 1.

$$E_\theta = 0$$

$$B_r = 0 \tag{1.38}$$

$$B_\theta = \frac{-i\omega}{\gamma_1 c^2} E_0 J_1\left(\frac{x_{01} r}{R}\right) e^{i(\pm kz - \omega t)}$$

$$B_z = 0. \tag{1.39}$$

Note that $dJ_0(u)/du = -J_1(u)$, where $J_1$ is the Bessel function of order 1. These solutions (Equations 1.36 through 1.39) of Maxwell's equations assume no energy losses in the walls of the waveguide, that is, the walls are perfect conductors. At $r = 0$, $J_0$ has its maximum value and $J_1 = 0$; therefore, along the $z$ axis the electric field has its maximum value and it is purely longitudinal.

The condition for wave propagation is $k^2 > 0$; otherwise, the waves will not propagate but will be evanescent. The wave propagation condition can be written as (see Equation 1.17)

$$k^2 = \frac{\omega^2}{c^2} - \left(\frac{x_{mn}}{R}\right)^2 > 0. \tag{1.40}$$

Equation 1.40 is the *dispersion* relation for $TM_{mn}$ modes in a hollow cylindrical waveguide.

For frequencies less than the *cutoff frequency* $\omega_c = (x_{mn}/R)c$, $k$ is imaginary and there is no wave propagation. For wave propagation, $\omega > \omega_c$. At any given frequency, only a finite number of modes can propagate. For a TW accelerator, the dimensions of the guide are chosen so that only the lowest frequency mode can occur. For a cutoff frequency of $\omega_c = 2\pi \times 3000$ MHz, the inner diameter of the waveguide must be about 8 cm.

Let us compare the wavelength of the lowest-order mode to the diameter of the waveguide. For the $TM_{01}$ mode, $\omega_c = (2.405/R)c$. The free space wavelength of an electromagnetic wave having this frequency is $\lambda \sim 2.6\,R$, which is approximately the diameter of the waveguide. It is perhaps not surprising that something interesting happens when the wavelength is on the order of the diameter of the pipe. For very short wavelength, the electromagnetic waves are not affected by the presence of the waveguide. There is no upper limit to the frequency of propagation; after all, we can see light through the length of a pipe!

The phase velocity of the wave is given by the condition that $kz - \omega t = constant$. The phase velocity is defined as $dz/dt$, and therefore the phase velocity of propagating waves is

$$v_{ph} = \frac{\omega}{k} = \frac{c}{\sqrt{1 - \omega_c^2/\omega^2}}. \tag{1.41}$$

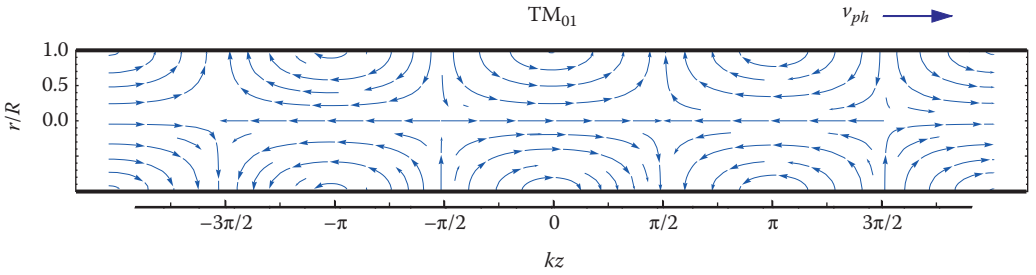



FIGURE 1.6 The electric field configuration in a plane containing the $z$ axis for a $TM_{01}$ traveling wave at some arbitrary instant in time and for a group velocity of 0.44 $c$. The horizontal scale is $kz$ and the vertical scale is $r/R$. Note the longitudinal nature of the field and that it has its maximum value on the axis of the waveguide.

Above the cutoff frequency, not only is the phase velocity greater than $c$, but it becomes infinite at cutoff. For TW particle acceleration, it is desirable for the particles to maintain a constant-phase relation with the waves so that they "see" a constant electric field. This is called synchronous particle acceleration. The condition for synchronism is $kz - \omega t = \varphi$, where $\varphi$ is a constant and $z$ is the position of the particle. On the waveguide axis (see Equation 1.36), $E_z = E_0 \cos(kz - \omega t) = E_0 \cos \varphi$, where for synchronous electrons $\varphi$ is constant. A phase velocity greater than $c$ presents an obvious problem for synchronous particle acceleration. When $v_{ph} > c$, the phase "rolls over" the particle and the net acceleration is zero. We will return to this difficulty shortly.

The group velocity is the speed at which an arbitrary waveform propagates or, equivalently, the speed at which energy is transferred. The group velocity is given by

$$v_g = \frac{d\omega}{dk} = c\sqrt{1 - \omega_c^2/\omega^2}. \tag{1.42}$$

For $\omega > \omega_c$, $v_g < c$ as expected. For $\omega \gg \omega_c$, $v_g = v_{ph} = c$ and $k = \omega/c$, the dispersion relation for a plane wave in free space.

Figure 1.6 shows a graphical representation of the electric field for a $TM_{01}$ mode with $v_g/c = 0.44$. The radial component of the electric field is zero on the axis ($r = 0$) of the waveguide.

## 1.4 TRAVELING WAVE ACCELERATORS I

A uniform open cylindrical waveguide cannot be used to accelerate charged particles because $v_{ph} > c$. It is necessary to decrease $v_{ph}$ below $c$ for certain modes and frequencies. This is accomplished by "loading" the waveguide by adding disks or iris plates, as shown in Figures 1.7 and 1.10. We will refer to a waveguide without disks as an open waveguide. Adding disks with central holes produces a sequence of cavities with coupling through the holes. The disks have a spacing $d$ and an opening

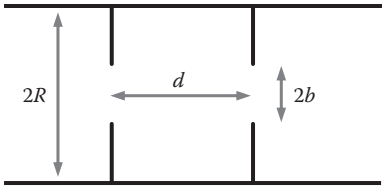



FIGURE 1.7 A cross section of a portion of a disk-loaded waveguide showing the disks and their dimensions.

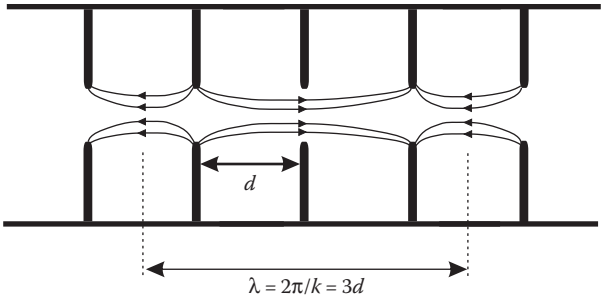



FIGURE 1.8 Shows how the presence of disks modifies the electric field lines in the waveguide. This shows the 2π/3 mode.

diameter of 2*b*. The values of *d* and *b* are chosen such that $v_{ph} < c$. The opening diameter ranges from 2.6 cm down to 1.9 cm for the constant-gradient (see Section 1.7.3) accelerator at the Stanford Linear Accelerator Center (SLAC) (see Karzmark et al., 1993, p. 71).[2] Figure 1.8 shows the effect on the fields of adding disks.

Julian Schwinger, one of the codevelopers of the theory of quantum electrodynamics, developed variational techniques for dealing with diaphragms and apertures in waveguides during World War II (Jackson, 1999).

Some fraction of the wave energy will be reflected at each disk. The distinction between a traveling wave and standing wave linac thus becomes somewhat blurred. In general, the reflections from successive disks will not be in phase and will therefore not combine to form a strong reflected wave. If, however, the extra distance traveled by the reflected waves from successive disks (namely 2*d*, forward and then back) is equal to an integer multiple of the wavelength $2d = n\lambda$ (or in terms of *k*, $k = \pm\pi/d$, $\pm 2\pi/d$, $\pm 3\pi/d$, etc.), there will be a strong reflected wave and standing waves can form. A TW linac must therefore run in a mode for which $k < \pi/d$.

In practice, ω is chosen so that $v_{ph}$ is approximately equal to *c* and the group velocity is quite low: $v_{ph}/v_g$ equals approximately 100. The reason for this may be that under these circumstances, only one resonant mode is available and the radial component of the electric field (which may be destabilizing) is small.

Equation 1.40 can be rewritten as

$$\omega^2 = \omega_c^2 + c^2k^2, \tag{1.43}$$

where $\omega_c$ is the cutoff frequency. Equation 1.43 is the dispersion relation for a uniform waveguide (no disks). Dispersion relations are common in physics. The name comes from the fact that a pulse consisting of a spread in frequencies will disperse because the phase velocities of the various components depend on the frequency (e.g., see Equation 1.41). The simplest dispersion relation is for the case of a plane electromagnetic wave traveling in free space: $\omega = ck$ or $\nu = c/\lambda$. The phase velocity of the wave is $v_{ph} = \omega/k$. In the case of a plane wave in free space, the phase velocity is independent of the frequency. Equation 1.43 is the equation of a hyperbola in a $\omega$ versus $k$ graph (see the solid black curve in Figure 1.9). In Equation 1.43, when $k \gg \omega_c/c$ (i.e., when the wavelength becomes very small compared to $R$), the curve approaches the asymptote $\omega = ck$, which is the free space dispersion relation. We expect that this should be the case. We can see from Figure 1.9 that the group velocity is zero at the cutoff frequency; that is, power does not propagate. For large values of $k$, the group velocity is equal to the phase velocity.

When disks are introduced into the waveguide, the dispersion relation changes (see the gray curve in Figure 1.9). For small values of $k$, it remains unchanged. The cutoff frequency therefore remains the same. The dispersion relation for the disk-loaded guide begins to depart significantly from the open-waveguide dispersion relation for $k \gtrsim 1/d$. As $k$ becomes larger, the wavelength becomes shorter until $\lambda = 2d$ ($k = \pi/d$). When this occurs, successive reflection at the disks interferes constructively, and we arrive at a standing wave situation. No power flows and $v_g = d\omega/dk = 0$. This implies an upper limit to the frequency, $\omega_{c2}$, of wave propagation in a disk-loaded guide.

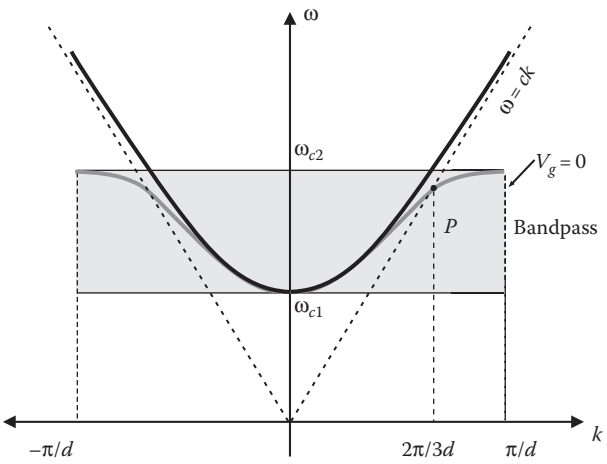



FIGURE 1.9 The dispersion relation for an open-bore waveguide without disks (solid black curve) and for a disk-loaded guide (gray). The two dashed diagonal lines show the dispersion relation for a plane electromagnetic wave in free space. At $k = \pm\pi/d$, where $d$ is the spacing between disks, the group velocity becomes zero. A disk-loaded waveguide operates at point $P$, where the phase velocity is about equal to $c$. It is common to run a TW accelerator in the $2\pi/3$ mode.

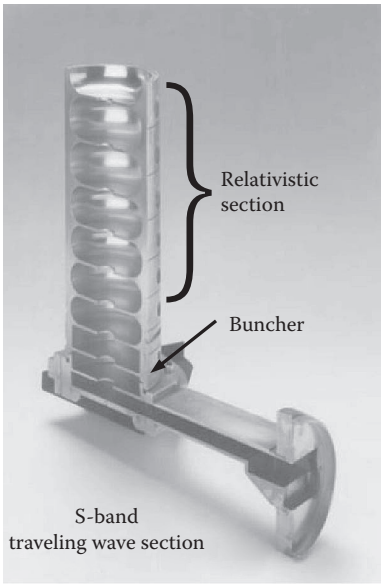



FIGURE 1.10 A cutaway view of a TW waveguide. The electrons enter the waveguide at the bottom. The structure protruding to the right at the bottom is the RF inlet. Notice that the spacing between disks is smaller at the bottom in the buncher section. This waveguide appears to be a constant-gradient structure (see Section 1.7.3) because the disk apertures become smaller toward the top. (Adapted from Thorson, T., Advanced acceleration and image guidance technologies, online PowerPoint presentation, 2007, http://195.135.200.83/allegatiifo/Congresso2007/19aprile/Thorson.pdf.)

Waves may propagate only within the bandpass frequency range shown in Figure 1.9. In an accelerator, the waveguide is operated at the point $P$ on the dispersion relation where the phase velocity $v_p \lesssim c$ and $v_g \ll$ c. Typically, $v_g/c = 0.01$. The optimum choice for the mode of operation is the $2\pi/3$ mode (Loew and Talman, 1983; Humphries, 2012).[3] This means that the disk spacing is given by $kd = 2\pi/3$, the wavelength is $3d$ (see Figure 1.8), and there are four disks per wavelength. Figure 1.10 shows a photo of a TW waveguide.

## 1.5 CAVITY OSCILLATIONS

We now consider a closed cylindrical cavity of radius $R$ and length $l$. The boundary conditions that were applied for the open cylinder at $r = R$ apply here as well. In addition, we have new boundary conditions that apply at the end caps ($z = 0$ and $z = l$). To provide an accelerating field, we only consider transverse magnetic solutions to Maxwell's equations in which $E_z$ is nonzero.

The boundary conditions at the end caps are that the transverse components of the electric field are zero and that the normal component of the magnetic field equals zero (see Equation 1.11):

$$z = 0 : E_\theta = 0, E_r = 0, B_z = 0$$
$$z = l : E_\theta = 0, E_r = 0, B_z = 0. \tag{1.44}$$

The boundary condition $B_z = 0$ is already obeyed once we restrict solutions to transverse magnetic modes.

The solutions that we derived previously for traveling waves in an open cylinder (Equations 1.36 through 1.39) remain valid, but they must be constrained by the additional boundary conditions at the end caps (Equation 1.44). We need to include traveling waves in the $+z$ direction and in the $-z$ direction. We expect these traveling waves to combine to form standing waves in the cavity.

For standing waves in the cylindrical cavity, we write $E_z$ as a linear combination of waves traveling to the right and to the left ($+z$ and $-z$). We drop the time dependence ($e^{-i\omega t}$) temporarily for brevity:

$$E_z = E_0 J_0 \left[ A e^{ikz} + B e^{-ikz} \right], \tag{1.45}$$

where $A$ and $B$ are constants to be determined. We will use Equations 1.21 through 1.24 to solve for the other field components. From Equation 1.23, $E_\theta = 0$ ($m = 0$, no $\theta$ dependence). The expression for $E_r$ is

$$E_r = \frac{\pm ik}{\gamma^2} \frac{\partial E_z}{\partial r}, \tag{1.46}$$

where the plus sign is for waves traveling to the right and the minus sign is for waves traveling to the left. Substituting Equation 1.45 into Equation 1.46,

$$E_r = \frac{ik}{\gamma^2} E_0 \left( \frac{x_{01}}{R} \right) J_1 \left[ B e^{-ikz} - A e^{ikz} \right]. \tag{1.47}$$

For the magnetic field, $B_r = 0$ (from Equation 1.22) and $B_\theta$ is

$$B_\theta = \frac{i\omega}{c^2\gamma^2} \frac{\partial E_z}{\partial r} = \frac{-i\omega}{c^2\gamma^2} E_0 \left( \frac{x_{01}}{R} \right) J_1 \left[ A e^{ikz} + B e^{-ikz} \right]. \tag{1.48}$$

These expressions for **E** and **B** already obey the boundary conditions at $r = R$. The boundary conditions at the end caps (Equations 1.44) must now be applied. The condition $E_r = 0$ at $z = 0$ implies that $B = A$, and therefore

$$E_r = \frac{2k}{\gamma^2} E_0 A \left( \frac{x_{01}}{R} \right) J_1 \sin kz. \tag{1.49}$$

The condition that $E_r = 0$ at $z = l$ implies that

$$k = \frac{p\pi}{l}, \text{ where } p = 0, 1, 2, \ldots. \tag{1.50}$$

We can absorb the constant $A$ into $E_0$ in Equation 1.49 without any loss of generality. The complete set of solutions for **E** and **B** can now be written (restoring the time dependence):

$$\begin{aligned} E_z &= E_0 J_0 \cos\frac{p\pi z}{l}\cos\omega t \\ E_\theta &= 0 \\ E_r &= \frac{2k}{\gamma^2}\left(\frac{x_{01}}{R}\right)E_0 J_1 \sin\frac{p\pi z}{l}\cos\omega t, \end{aligned} \tag{1.51}$$

and

$$\begin{aligned} B_z &= 0 \\ B_\theta &= \frac{2\omega}{c^2\gamma^2}E_0\left(\frac{x_{01}}{R}\right)J_1 \cos\frac{p\pi z}{l}\sin\omega t \\ B_r &= 0. \end{aligned} \tag{1.52}$$

In addition to the parameters $m$ and $n$, we now have the extra parameter $p$ that describes the number of nodes between the two end caps of the cylinder. As these modes are transverse magnetic, we can describe them as $TM_{mnp}$ modes. The solutions listed in Equations 1.51 and 1.52 are $TM_{01p}$ modes.

Now let us concentrate on the simplest mode, the $p = 0$ mode. It is this mode that is relevant for SW accelerators. When $p = 0$, this implies $k = 0$. This is a bit puzzling because it is not clear what the wavelength is in this case. As we will see, the wavelength is actually infinite, as implied by $k = 0$! The resonant frequency is

$$\omega_{010} = c\frac{x_{01}}{R}. \tag{1.53}$$

The resonant frequency is independent of $l$! This is really quite surprising. For standing waves on a vibrating string, the frequencies very definitely depend on the length of the string—likewise for standing sound waves in a pipe. This shows that analogies between electromagnetic oscillations in a resonant cavity and standing

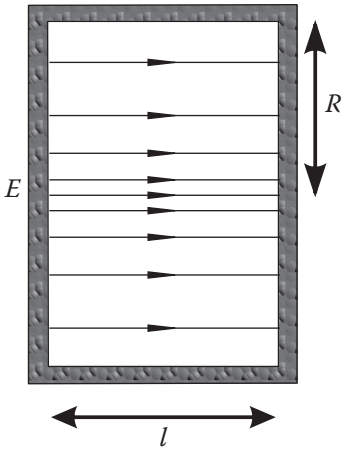



FIGURE 1.11 A cross section through a resonant cavity and the electric field distribution within for a $TM_{010}$ mode. The electric field is constant in the longitudinal direction. The aspect ratio of this cavity is $l/R = 1.3$.

waves on a string can be misleading. Equation 1.53 immediately fixes the inner radius for cavity oscillations at 3000 MHz. This radius is 3.83 cm.

When $p = 0$ ($TM_{010}$ mode), the only nonzero components of the fields are $E_z$ and $B_\theta$:

$$E_z = E_0 J_0\left(\frac{x_{01} r}{R}\right)\cos\omega t$$

$$B_\theta = \frac{E_0}{c} J_1\left(\frac{x_{01} r}{R}\right)\sin\omega t. \tag{1.54}$$

Note that at any given instant in time, $E_z$ and $B_\theta$ are spatially constant with $z$! Therefore, the wavelength is effectively infinite, consistent with $k = 0$. There is no radial component of **E**; the electric field is parallel to the axis of the waveguide and has its maximum value on the axis. Figure 1.11 shows a resonant cavity along with the electric field distribution.

## 1.6 ENERGY

Poynting's theorem is a statement of conservation of energy for electromagnetic fields and particles. Consider a region of volume $V$ bounded by a surface of area $A$. The rate at which energy is dissipated by electrical currents is $\int \mathbf{J}\cdot\mathbf{E}dV$. This will apply to the walls of the waveguide. The rate at which work is done on the electrons to be accelerated is $q\mathbf{v}\cdot\mathbf{E}$, where $\mathbf{v}$ is the velocity of these electrons. We assume that the accelerated electrons occur in small enough bunches that they themselves do not contribute

significantly to the electric field. A statement of Poynting's theorem (Jackson, 1999), including the rate of work done on the accelerated electrons, is

$$-q\mathbf{v}\cdot\mathbf{E}-\int \mathbf{J}\cdot\mathbf{E}\,dV=\frac{\partial}{\partial t}\int u dV+\oint_A \mathbf{S}\cdot d\mathbf{A}, \tag{1.55}$$

where

$$\begin{aligned} u &= \frac{1}{2}(\mathbf{E}\cdot\mathbf{D}+\mathbf{B}\cdot\mathbf{H}) \\ &= \frac{1}{2}\left(\varepsilon_0 E^2+\frac{1}{\mu_0}B^2\right) \end{aligned} \tag{1.56}$$

is the energy density stored in the electric and magnetic fields and

$$\mathbf{S}=\mathbf{E}\times\mathbf{H}, \tag{1.57}$$

is the Poynting vector. The Poynting vector represents the energy per unit time per unit area flowing across the surface bounding the volume. The direction of this vector is the direction of energy flow. Equation 1.55 states in words that the rate at which work is done on charged particles (the accelerated electrons and conducting electrons in the walls of the waveguide) is balanced by a decrease in the energy stored in the fields and the rate at which electromagnetic energy flows out of the volume.

It is important to understand that the vectors in Equations 1.55 through 1.57 must be real quantities. Given any two arbitrary complex vector functions **A** and **C**: Re (**A** × C) ≠ Re **A** × Re **C**, where Re indicates the real part. Let us consider the product of the real parts of any two complex vector quantities that vary with time dependence $e^{-i\omega t}$. The product can be either a cross product or a dot product. We can write

$$\begin{aligned} \mathbf{A}(\mathbf{r},t) &= \mathbf{A}(\mathbf{r})e^{-i\omega t} \\ \mathbf{C}(\mathbf{r},t) &= \mathbf{C}(\mathbf{r})e^{-i\omega t}. \end{aligned} \tag{1.58}$$

The product of the real parts of these quantities is

$$\begin{aligned} \mathrm{Re}\,\mathbf{A}\times\mathrm{Re}\,\mathbf{C} &= \frac{1}{2}\left(\mathbf{A}e^{-i\omega t}+\mathbf{A}^* e^{i\omega t}\right)\times\frac{1}{2}\left(\mathbf{C}e^{-i\omega t}+\mathbf{C}^* e^{i\omega t}\right) \\ &= \frac{1}{4}\left(\mathbf{A}\times\mathbf{C}e^{-2i\omega t}+\mathbf{A}\times\mathbf{C}^*+\mathbf{A}^*\times\mathbf{C}+\mathbf{A}^*\times\mathrm{C}^* e^{2i\omega t}\right) \\ &= \frac{1}{2}\mathrm{Re}\left(\mathbf{A}^*\times\mathbf{C}+\mathbf{A}\times\mathbf{C}e^{-2i\omega t}\right). \end{aligned} \tag{1.59}$$

We would now like to find the time average of the product of the harmonic functions in Equation 1.59 over the period of oscillation: $2\pi/\omega$. The numerical value of the period for s-band microwaves is 0.35 ns. The time average of a function $f(t)$ is

$$\langle f(t)\rangle = \frac{\omega}{2\pi}\int_0^{2\pi/\omega} f(t)\,dt. \tag{1.60}$$

Applying this definition to the last expression in Equation 1.59, we find that

$$\langle \mathrm{Re}\,\mathbf{A}\times \mathrm{Re}\,\mathbf{C}\rangle = \frac{1}{2}\mathrm{Re}\left(\mathbf{A}^*\times\mathbf{C}\right). \tag{1.61}$$

We can apply Equation 1.61 to the Poynting vector (Equation 1.57) to obtain the time-averaged power *loss* per unit area traversing a surface:

$$\frac{d\langle P_w\rangle}{da} = -\langle \mathbf{n}\cdot\mathbf{S}\rangle = -\frac{1}{2}\mathrm{Re}\left[\mathbf{n}\cdot\left(\mathbf{E}^*\times\mathbf{H}\right)\right], \tag{1.62}$$

where **n** is the unit vector normal to the surface.

The time-averaged energy density is

$$\langle u\rangle = \frac{1}{4}\left(\mathbf{E}^*\cdot\mathbf{E} + \frac{1}{\mu_0}\mathbf{B}^*\cdot\mathbf{B}\right). \tag{1.63}$$

We would like to be able to evaluate the power loss in the walls of an accelerating waveguide or cavity. Our solutions to Maxwell's equations for the traveling waves (Equations 1.36 through 1.39) and for the cavity (Equations 1.54) assume perfectly conducting walls. In this circumstance, there is no power dissipation in the walls. To evaluate the power loss in the walls of a real conductor, we need to find expressions for the fields in the thin boundary layer at the surface. To accomplish this, we follow the successive approximation scheme outlined by Jackson (1999). We start with the solutions for $\mathbf{H}_t$ and $\mathbf{E}_n$ for a perfect conductor just outside the conductor surface, where $t$ represents the tangential component and $n$ represents the normal component. We then find the fields in the surface boundary layer of the conductor and small corrections to the fields outside.

In a real conductor, the conductivity is finite. In such a conductor, there cannot truly be a surface current, for **J** would have to be infinite as the surface area of the loop in Figure 1.3 shrinks to zero. Instead, there is a thin boundary layer over which there is a finite current density; beyond this layer one expects **E** and **B** to take on values characteristic of a perfect conductor, that is, to go to zero. Equation 1.10 should be modified for a real conductor:

$$\mathbf{n}\times\left(\mathbf{H}-\mathbf{H}_c\right) = 0. \tag{1.64}$$

The boundary layer is expected to be thin, and therefore we assume that spatial variations normal to the conductor are much larger than parallel to the surface. We will also neglect the displacement current ($\partial \mathbf{D}/\partial t$ term in Equation 1.2). We can check this after the fact as an exercise. We will assume a harmonic time dependence of the form $e^{-i\omega t}$ and that $\mathbf{B} = \mu\mathbf{H}$. The curl equation for $\mathbf{E}$ in the conductor becomes

$$\nabla \times \mathbf{E}_c = -\frac{\partial \mathbf{B}_c}{\partial t} = i\omega\mu_c \mathbf{H}_c,$$

and therefore

$$\mathbf{H}_c = \frac{-i}{\mu_c \omega} \nabla \times \mathbf{E}_c, \tag{1.65}$$

where $\mu_c$ is the permeability in the conductor. From the other Maxwell curl equation, we have $\nabla \times \mathbf{H}_c = \sigma \mathbf{E}_c$, so that

$$\mathbf{E}_c = \frac{1}{\sigma} \nabla \times \mathbf{H}_c. \tag{1.66}$$

Let $\xi$ be the coordinate normal to and increasing inward into the wall of the conductor (see Figure 1.12). Then we can write $\nabla \cong -\mathbf{n}(\partial/\partial\xi)$. In this case, Equation 1.66 becomes

$$\mathbf{E}_c \cong -\frac{1}{\sigma} \mathbf{n} \times \frac{\partial \mathbf{H}_c}{\partial \xi}, \tag{1.67}$$

and Equation 1.65 for $\mathbf{H}_c$ is

$$\mathbf{H}_c \cong \frac{i}{\mu_c \omega} \mathbf{n} \times \frac{\partial \mathbf{E}_c}{\partial \xi}. \tag{1.68}$$

Equations 1.67 and 1.68 are two equations in the two unknowns, $\mathbf{E}_c$ and $\mathbf{H}_c$. By substituting Equation 1.67 for $\mathbf{E}_c$ into Equation 1.68, we can obtain (after a little bit of effort) the two equations

$$\frac{\partial^2}{\partial \xi^2}(\mathbf{n} \times \mathbf{H}_c) + \frac{2i}{\delta^2}(\mathbf{n} \times \mathbf{H}_c) \cong 0$$

$$\mathbf{n} \cdot \mathbf{H}_c = 0, \tag{1.69}$$

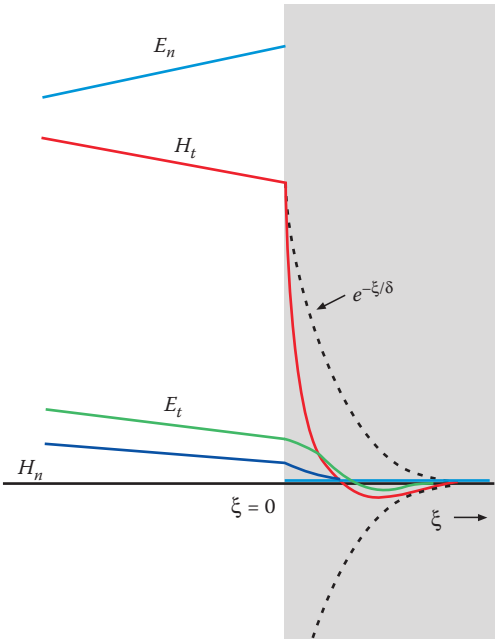



FIGURE 1.12 The fields inside a good but not perfect conductor. There are tangential components of **E** and **H** that rapidly decline toward zero with increasing depth ξ. The length scale over which the fields decline is the skin depth δ.

where

$$\delta = \left( \frac{2}{\mu_c \omega \sigma} \right)^{1/2} \tag{1.70}$$

is called the skin depth. Equations 1.69 are two equations for the two components of $\mathbf{H}_c$, the normal component and the tangential component. Let us interpret the second of these equations first. The normal component of $\mathbf{H}_c$ just inside the metal surface is 0, as is the case for a perfect conductor. The first Equation 1.69 is a differential equation for the tangential component of $\mathbf{H}_c$ with the initial conditions (1.64). The solution of this differential equation is

$$\mathbf{H}_c = \mathbf{H}_t e^{-\xi/\delta} e^{-i\xi/\delta}, \tag{1.71}$$

where $\mathbf{H}_t$ is the tangential magnetic field just outside the conductor. The electric field in the conductor may be obtained by substituting Equation 1.71 into Equation 1.67:

$$\mathbf{E}_c = \frac{(1-i)}{\sigma\delta} (\mathbf{n} \times \mathbf{H}_t) e^{-\xi/\delta} e^{i\xi/\delta}. \tag{1.72}$$

There is thus, in the boundary layer, a tangential component of the magnetic field and a tangential component of the electric field. These fields are spatially oscillatory and decay over a distance on the order of δ (see Figure 1.12). As the tangential components of **E** and **H** are continuous, the fields just outside the conductor are given by Equations 1.71 and 1.72 with $\xi = 0$. Note that in the limit of infinite conductivity, the fields are all zero inside the conductor.

The skin depth δ depends on the frequency and conductivity (see Equation 1.70). For a perfect conductor, $\delta = 0$, as expected. Let us estimate δ. For copper, $\sigma = 5.8 \times 10^7\ \Omega^{-1}\ \mathrm{m}^{-1}$ and $\mu_c = 1.25 \times 10^{-6}$ H/m. For $\omega = 2\pi \times 3000$ MHz (s-band microwaves used in medical linear accelerators), $\delta = 1.2 \times 10^{-6}$ m. The boundary layer is thin indeed!

We are now in a position to calculate the power lost by electromagnetic energy flow into the conducting walls of the waveguide or cavity. Substitute the expressions for the fields (Equations 1.71 and 1.72) into Equation 1.62):

$$\frac{d\langle P_w \rangle}{da} = \frac{1}{2\sigma\delta}|\mathbf{H}_t|^2. \tag{1.73}$$

The energy flowing into the surface should show up in the form of ohmic or resistive losses (sometimes called Joule heating) in the conductor (see the left side of Equation 1.55). The current density near the surface is $\mathbf{J} = \sigma\mathbf{E}_c$:

$$\mathbf{J} = \frac{1-i}{\delta}(\mathbf{n} \times \mathbf{H}_t)e^{-\frac{\xi}{\delta}(1-i)}. \tag{1.74}$$

The time-averaged rate of energy dissipation per unit volume in ohmic losses is

$$\frac{1}{2}\mathrm{Re}(\mathbf{J}^* \cdot \mathbf{E}) = \frac{1}{2\sigma}|\mathbf{J}|^2.$$

The total rate of energy dissipation in the conductor for the volume lying beneath an area $\Delta A$ is

$$\frac{1}{2\sigma}\Delta A\int_0^\infty |\mathbf{J}|^2 d\xi = \frac{\Delta A}{2\sigma\delta}|\mathbf{H}_t|^2, \tag{1.75}$$

the same as found previously (see Equation 1.73).

As Equation 1.74 shows, the current density is confined to a very thin layer just below the surface of the conductor. It is equivalent to an effective surface current $\mathbf{K}_{\mathbf{eff}}$:

$$\mathbf{K}_{\mathbf{eff}} = \int_0^\infty \mathbf{J} d\xi = \mathbf{n} \times \mathbf{H}_t. \tag{1.76}$$

Comparison with the boundary condition (Equation 1.10) for a perfect conductor shows that a real conductor behaves like a perfect conductor with a surface current $\mathbf{K}_{\mathbf{eff}}$. The power loss can be written in terms of the effective surface current:

$$\frac{dP}{da} = \frac{1}{2\sigma\delta}\left|\mathbf{K}_{\mathbf{eff}}\right|^2. \tag{1.77}$$

Equation 1.73 will allow us to estimate the energy losses in a real waveguide or resonant cavity provided that we have solved for the fields in the infinite conductivity case.

### 1.6.1 TRAVELING WAVE WAVEGUIDE

We can calculate the time-averaged energy per unit length stored in the fields for an open waveguide by using Equation 1.63 and integrating over the cross-sectional area of the waveguide:

$$\langle U_l \rangle = \int \langle u \rangle dA, \tag{1.78}$$

where $\langle u \rangle$ is the time average of the energy density and $dA = 2\pi r\, dr$. The following integrals of Bessel functions are needed to evaluate Equation 1.78 (and a number of integrals appearing later):

$$\int_0^R r J_0^2\left(\frac{x_{01}}{R} r\right) dr = \frac{R^2}{2} J_1(x_{01})$$

$$\int_0^R r J_1^2\left(\frac{x_{01}}{R} r\right) dr = \frac{R^2}{2} J_1(x_{01}). \tag{1.79}$$

It is curious but true that both of these integrals have the same value. The value of $J_1^2(x_{01})$ is 0.2695.[4] For the time-averaged energy per unit length we obtain

$$\langle U_l \rangle = \frac{\pi}{2}\left(\frac{\omega}{\omega_c}\right)^2 \varepsilon_0 E_0^2 R^2 J_1^2(x_{01}). \tag{1.80}$$

The time-averaged power flowing along the guide is computed from the Poynting vector:

$$\langle P \rangle = \int \langle S_z \rangle dA, \tag{1.81}$$

where $\langle S_z \rangle$ is the $z$ component of the time average of the Poynting vector and is equal to: ½Re ($E_r^* B_\theta / \mu_0$). The result of carrying out this integral is

$$\langle P \rangle = \frac{\pi}{2} \varepsilon_0 c \left( \frac{\omega}{\omega_c} \right)^2 E_0^2 R^2 J_1^2 (x_{01}) \sqrt{1 - \omega_c^2 / \omega^2}. \tag{1.82}$$

If we divide $\langle P \rangle$ by $\langle U_l \rangle$ we find that

$$\frac{\langle P \rangle}{\langle U_l \rangle} = v_g. \tag{1.83}$$

This result is generally true and provides a physical interpretation of the group velocity as the speed at which energy flows.

The time-averaged energy loss per unit length in the walls of the waveguide as a result of eddy current flow and Joule heating is given by Equation 1.73 ($da = 2\pi R\, dz$):

$$\left\langle \frac{dP_w}{dz} \right\rangle = -\frac{\pi}{\sigma \delta \mu_0^2 c^2} R \left( \frac{\omega}{\omega_c} \right)^2 E_0^2 J_1^2 (x_{01}). \tag{1.84}$$

## 1.6.2 CAVITY OSCILLATOR

In this section, we assume that the frequency always takes on the value of the resonant frequency ($\omega = \omega_{010}$). The total energy, $U$, stored in the fields in a resonant cavity is found by integrating Equation 1.56 over the volume of the cavity (fields given by Equation 1.54). The result is

$$U = \frac{\pi}{2} \varepsilon_0 E_0^2 R^2 l J_1^2 (x_{01}), \tag{1.85}$$

where $l$ is the length of a single resonant cavity. The energy is constant in time, as expected, since it was calculated based on the assumption of infinite conductivity. The stored energy is proportional to $E_0^2$ and the volume of the cavity. The total rate of power loss in the cavity walls, $P_w$, is given by Equation 1.73. This expression for the power loss per unit area must be integrated over the end caps and over the cylindrical side walls at $r = R$. The latter integration is actually a simple multiplication by the area, as the expression for the fields is independent of $\theta$ and $z$ for the $TM_{010}$ mode (see Equation 1.54). The result is

$$\langle P_w \rangle = \frac{\pi}{\sigma \delta c^2 \mu_0^2} R(R + l) E_0^2 J_1^2 (x_{01}). \tag{1.86}$$

The power dissipation in the walls is proportional to $E_0^2$ and the surface area of the walls. The rate of power dissipation in a resonant cavity can be described by a parameter $Q$. This quantity is frequently used to describe the damping of an oscillator such as a damped mechanical harmonic oscillator or an inductor, capacitor, and resistor (LCR) circuit. The definition of this quantity varies slightly from reference to reference. Here, we use the following definition:[5]

$$Q = \omega \frac{U}{\langle P_w \rangle}. \tag{1.87}$$

The power dissipated in the walls of a real cavity will lead to a decrease in $U$ with respect to time. If power dissipation in the cavity walls is the only energy loss mechanism, then the rate at which the field energy decays must equal $P_w$, and therefore

$$\frac{dU}{dt} = -\frac{\omega}{Q} U, \tag{1.88}$$

with the solution

$$U = U_0 e^{-\omega t/Q}, \tag{1.89}$$

where $U_0$ is the initial stored energy. The timescale for the decay of the stored energy is $t_d = Q/\omega = \text{period} \times Q/2\pi$. This provides a physical interpretation of $Q$: it is $2\pi$ times the decay timescale divided by the period.

We can calculate an explicit expression for $Q$ for a cylindrical resonant cavity by substituting the expressions in Equations 1.85 and 1.86 into Equation 1.87:

$$Q = \frac{l}{\delta(1 + l/R)}, \tag{1.90}$$

assuming that $\mu \approx \mu_0$. Thus, we see that $Q$ depends only on the geometry of the cavity, the conductivity of its walls, and the frequency. For a pillbox cavity of length 5.0 cm, frequency 3000 MHz ($R$ = 3.8 cm), $Q$ = 18,000. $Q$ values for SW accelerator resonant cavities are typically 15,000–20,000. The energy decay timescale $t_d \approx 1$ μs for a cavity with the parameters given above.

## 1.7 TRAVELING WAVE ACCELERATORS II

Electrons are injected into the waveguide with an energy of about 50 keV, corresponding to a speed of about 0.4$c$. In the first 30 cm of the guide, the electrons are accelerated from 0.4$c$ to 1.0$c$. The first section of the waveguide is called the buncher section.

This consists of six to eight cells. The remaining portion is called the relativistic section, and we will concentrate our attention on this section here (Figure 1.10).

The accelerating waveguide is embedded in a water jacket to carry away the heat dissipated in the copper walls. Focusing coils surround portions of this and are used to prevent the electron bunches from dispersing. These coils are designed to minimize the beam diameter. Steering coils are used to maintain the beam in the correct position. The steering coils are servoed to maintain proper beam position via feedback from the monitor ion chambers. Beam steering is affected by the earth's magnetic field, which must be compensated for as a function of gantry angle.

In a traveling wave linac, it is necessary to avoid the formation of standing waves. The RF exiting at the far end of the guide must not be reflected back into the waveguide. Reflections can be avoided by either (1) dissipating the residual RF in a resistive load or (2) RF feedback to the input end with a suitable phase adjustment to reinforce incoming power. RF feedback will be discussed in Section 1.7.4.

### 1.7.1 INPUT POWER, BEAM ENERGY, AND CURRENT

Our goal in this section is to learn the relationship between microwave input power, beam energy, and beam current. We begin with a qualitative discussion of the energy flow and losses in the accelerating waveguide. RF power is fed into the guide at the electron gun end. There are two sinks of RF power: the losses in the walls of the guide due to induced currents and the energy given to the electron beam. The RF power that is not either dissipated in the walls or absorbed by the beam exits at the other end of the waveguide. Wave reflections at the far end of the guide must be avoided; otherwise, a standing wave will be established.

It is important to emphasize that the solutions for **E** and **B** derived previously (Equations 1.36 through 1.39) are based on the assumption of perfectly conducting walls, and therefore these solutions do not include any energy losses.

Let us apply energy conservation to a disk-shaped volume of thickness $\Delta z$ in the waveguide, as shown in Figure 1.13.

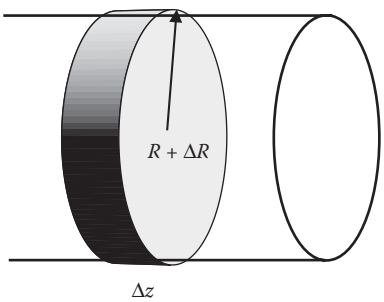



FIGURE 1.13 Conservation of energy applied to a thin disk inside the waveguide. The radius of the disk is slightly larger than the inner radius of the waveguide.

The radius of the disk is somewhat larger than $R$, so that it extends into the walls of the guide where the fields are zero over the circumference of the disk. We assume there is no electromagnetic power flow radially outward over the surface of the disk at: $R + \Delta R$ ($\Delta R \gg \delta$). The electromagnetic power flowing into the disk is $P(z) = \int S_z(z) dA$. The power flowing out of the disk is $P(z + \Delta z)$. Power losses occur in the walls of the waveguide due to induced current flow in the walls and the finite conductivity of copper. If the power dissipated per unit length in the walls is $dP_w/dz$, then the power lost in length $\Delta z$ is: $(dP_w/dz)\ \Delta z$. Energy is stored in the fields inside the disk. If the energy per unit length stored in the fields is $U_l$, then the time rate of change of the energy in the disk is $(\partial U_l/\partial t)\ \Delta z$.

The work done on the beam electrons inside the disk is $\Delta W = q\mathbf{E} \cdot \mathbf{e}_z\ \Delta z$. The rate at which work is done on the electrons in the beam is $\Delta W/\Delta t = (q/\Delta t)\mathbf{E} \cdot \mathbf{e}_z\ \Delta z = IE_z \Delta z$, where $I$ is the beam current and $E_z$ is the $z$ component of the electric field.[6] We can write an equation for energy conservation in the disk by putting all the pieces together:

$$P(z) = P(z + \Delta z) - \frac{dP_w}{dz}\Delta z - \frac{\partial U_l}{\partial t}\Delta z + IE_z \Delta z. \tag{1.91}$$

Dividing Equation 1.91 by $\Delta z$ and taking the limit in which $\Delta z \to 0$, we obtain

$$IE_z - \frac{dP_w}{dz} = \frac{\partial U_l}{\partial t} - \frac{dP}{dz}, \tag{1.92}$$

where $dP/dz$ is the rate of change of the RF power down the guide (and it is less than zero). Equation 1.92 can also be derived more rigorously from Poynting's theorem, although with more effort.

Equation 1.92 is a time-dependent instantaneous expression for the rate at which power is either lost or absorbed by different parts of the system. When microwave power is first fed into the waveguide, there are initial transients. After some period of time, the system reaches a steady state in which $\partial U_l/\partial t = 0$. It is presumed that energy is fed into the guide at a rate that matches the rate of Joule heating in the wall and the rate of power absorbed by the beam.

The solutions for **E** and **B** in Equations 1.36 through 1.39 assume perfectly conducting walls and ignore the energy absorbed by the electron beam. In this case, the left-hand side of Equation 1.92 is zero. In the steady-state case, therefore, $dP/dz$ is also zero. For perfectly conducting walls, the wavenumber $k$ is either real, in which case we have wave propagation, or purely imaginary, in which case wave propagation does not occur. If the walls have a large but finite conductivity, $k$ will be perturbed and can be written $k' = k + \beta + i\alpha$, where $k$ is the wavenumber for perfectly conducting walls and the terms $\beta$ and $\alpha$ represent perturbations. In particular, $i\alpha$ represents a small dissipation term due to power losses in the walls. When there are energy losses, $\langle P \rangle$ will decline with increasing $z$ down the waveguide. This implies that the

amplitude of the electric field will also decline with increasing $z$. We assume that $E_0$ in Equation 1.36 is no longer a constant but now a function of $z$, that is, $E_0 \rightarrow E(z)$. We will shortly use Equation 1.92 to find this function.

We wish to take the time average of Equation 1.92 over an interval on the order of 1 μs (the duration of an RF pulse; see Section 1.9). This timescale is very long compared to the period of the microwaves (about 0.3 ns). Let us consider the first term on the left-hand side of Equation 1.92, which is proportional to $\cos(kz - \omega t) = \cos \varphi$.[7] We assume that the electrons travel at virtually the speed of light and that therefore $\varphi$ is constant in time. This is the condition of synchronous acceleration. We further assume that $\varphi = 0$; that is, the electrons see the maximum value of the electric field. The time average of $P_w$ is written as $\langle P_w \rangle$, and the time average of $P$ is $\langle P \rangle$. The time average of Equation 1.92 can now be written

$$IE(z) - \left\langle \frac{dP_w}{dz} \right\rangle = -\left\langle \frac{dP}{dz} \right\rangle. \tag{1.93}$$

The time averages in Equation 1.93 have been calculated previously (for a waveguide without disks); see Equations 1.82 and 1.84. We now define the following parameters:

$$\alpha = \frac{-1}{2\langle P \rangle}\left\langle \frac{dP_w}{dz} \right\rangle \quad \text{and} \quad r_s = -\frac{E^2(z)}{(dP_w/dz)}, \tag{1.94}$$

where:

- $r_s$ is called the shunt impedance per unit length
- $\alpha$ is a parameter related to power loss in the walls of the guide

$\alpha$ and $r_s$ may depend on $z$. We also define

$$\tau = \int_0^L \alpha(z)\,dz, \tag{1.95}$$

where $L$ is the length of the waveguide. As we will see, the shunt impedance is a measure of the efficiency of the accelerator. The higher the shunt impedance, the lower the wall losses and the higher the electron energy for a given input power. A typical value of the shunt impedance is 50–60 MΩ/m. If we substitute Equations 1.94 into Equation 1.93, we get

$$\frac{dE}{dz} + \left[\alpha - \frac{1}{2}\left(\frac{r_s'}{r_s} + \frac{\alpha'}{\alpha}\right)\right]E = -\alpha I r_s, \tag{1.96}$$

where the primes denote $z$ derivatives. Equation 1.96 will allow us to determine $E(z)$ in various circumstances.

### 1.7.2 CONSTANT-IMPEDANCE LOAD LINE

In the constant-impedance case, $r_s$ and $\alpha$ are both constant with $z$. We may integrate Equation 1.96 to find $E(z)$, assuming $E(0) = E_0$ as an initial condition:

$$E(z) = E_0 e^{-\alpha z} - I r_s \left(1 - e^{-\alpha z}\right). \tag{1.97}$$

From Equation 1.95, $\tau = \alpha L$.

We can calculate the power carried by the RF as a function of position by inserting Equation 1.97 into Equations 1.94. In the absence of beam loading ($I = 0$),

$$\langle P \rangle = \langle P_0 \rangle e^{-2\alpha z}, \tag{1.98}$$

where

$$\langle P_0 \rangle = E_0^2 / (2\alpha r_s)$$

At the end of the waveguide, the RF power left over will be $\langle P \rangle = \langle P_0 \rangle e^{-2\tau}$.

We may now derive an expression for the effective accelerating potential of the waveguide:

$$V_e = \int_0^L E(z) \cos\varphi \, dz, \tag{1.99}$$

where:

- $L$ is the length of the waveguide
- $\varphi$ is the phase

Remember that the electrons are assumed to maintain a constant-phase relationship with the traveling wave. We will assume the optimum circumstance in which $\varphi = 0$. It is only the amplitude of the electric field that decreases as the electrons travel down the guide. If there were no losses in the walls and no energy absorbed by the beam (called beam loading), $V_e$ would simply be $E_0 L$.

The maximum field strength $E_0$ in Equation 1.97 is not easily measured. We may write $E_0$ in terms of $\langle P_0 \rangle$, the time-averaged input microwave power:

$$E_0 = \sqrt{2\tau \langle P_0 \rangle r_s / L}. \tag{1.100}$$

We may now substitute Equation 1.97 into Equation 1.99 and carry out the integration to obtain

$$V_e = \sqrt{2\tau} \frac{1 - e^{-\tau}}{\tau} \sqrt{\langle P_0 \rangle r_s L} - \left(1 - \frac{1 - e^{-\tau}}{\tau}\right) I r_s L. \tag{1.101}$$

Equation 1.101 is the "load line" for a constant-impedance TW accelerator. It tells us the accelerating potential in terms of the input microwave power and the beam current. This equation is of the form

$$V_e = V_0 - FI, \tag{1.102}$$

where:

$V_0$ is the maximum accelerating potential in the absence of beam loading
$F$ is a constant

If $r_s$ is in units of MΩ/m, $\langle P_0 \rangle$ in units of MW, $L$ in m, and $I$ in A, then $V_e$ is in units of MV. The second term in Equation 1.102 represents beam loading. It is possible to understand beam loading quite simply: when the current is increased (at constant input power), the energy available has to be shared by more electrons, and therefore each individual electron gets less energy.

The shunt impedance for a TW guide can be estimated from Equations 1.84 and 1.94. Assuming $E(z) = E_0$ and $v_g/c = 0.01$, we find that $r_s = 110$ MΩ/m for a smooth open waveguide without disks. A typical value for a disk-loaded guide is $r_s = 60$ MΩ/m. We can also calculate the value of α from Equation 1.94. For an open waveguide, this results in a value of approximately 0.1 $m^{-1}$. For a disk-loaded guide, a typical value is $\alpha = 0.2$ $m^{-1}$. Let us suppose that $L = 2$ m, and therefore $\tau = 0.4$; Equation 1.101 becomes

$$V_e = 8.1\sqrt{\langle P_0 \rangle} - 21I, \tag{1.103}$$

where:

$\langle P_0 \rangle$ is in MW
$I$ is in A
$V_e$ is in MV

In the zero-beam loading case, this linac will produce an 18 MV beam for a power input of $\langle P_0 \rangle = 5$ MW. The electric field associated with this power level (see Equation 1.100) is on the order of $10^7$ V/m! This corresponds to a potential difference of about 0.5 MV between disks if the disks are spaced about 5 cm apart.

The high microwave beam power of 5 MW is only available from the magnetron in pulses of a few microseconds in duration. Recall that the time average of $P$ is over one oscillation period of the fields (i.e., time = $1/\nu = 0.3$ ns). The time necessary to fill this waveguide with RF power is $t_f = L/v_g = 0.7$ μs.

From Equation 1.101 or 1.102, it can be seen that there is a beam current for which $V_e = 0$. If the current is increased beyond this point, power will flow out of the beam and back into the fields "in a fashion similar to a klystron!" (Ford, 1987). The beam power output is $I\ V_e$. This is a maximum for a beam current that corresponds to $V_e = V_0/2$. This can be shown by maximizing $I\ V_e$. This condition does not result in the maximum x-ray dose rate, however. The maximum x-ray dose rate depends on the details of bremsstrahlung x-ray production. The electron beam strikes a target

producing bremsstrahlung x-rays, which then pass through a flattening filter. The thick target bremsstrahlung yield is proportional to $V_e^3$ for relativistic electrons and directly proportional to the beam current $I$ (Ford, 1987). Thus, the total x-ray yield rate is $X = BI(V_e)V_e^3$, where $B$ is a constant determined by the accelerator target and flattening filter.[8] Using Equation 1.102, it is not difficult to show that the maximum x-ray output occurs when $V_e = (3/4)V_0$. For maximum dose rate, the linac should be operated at this point on the load line. Table 1.1 lists nominal parameters for a TW linac.

Figure 1.14 shows the load lines for an Elekta linac with three different values of the input power. The load lines in Figure 1.14 are not in good agreement with Equation 1.103, but these load lines are presumably for linacs with RF recirculation (see Section 1.7.4). Recirculation alters the load line. Notice that the maximum beam power (point P in Figure 1.14) occurs at output energy of about $V_0/2$. The maximum output occurs at approximately ¾ $V_0$.

### 1.7.3 CONSTANT-GRADIENT LOAD LINE

In the first half of this section, we will assume that there is no beam current. We assume a constant gradient, $\langle dP/dz \rangle$ = constant, and that the electric field amplitude is also constant, ($E(z) = E_0$), in the absence of a beam current. The power dissipation per unit length in the walls of the waveguide will be constant with $z$ under these circumstances. Assume that at the end of the guide: $\langle P \rangle = \langle P_0 \rangle e^{-2\tau}$, where $\langle P_0 \rangle$ is the input power at $z = 0$. We want the power exiting the guide to be the same as for the constant-impedance case so that we may compare them (presuming that $\tau$ is the same). The constant-gradient requirement implies

$$\left\langle \frac{dP}{dz} \right\rangle = \frac{\langle P(L) \rangle - \langle P_0 \rangle}{L} = -\frac{\langle P_0 \rangle}{L}\left(1 - e^{-2\tau}\right), \tag{1.104}$$

**TABLE 1.1 Nominal Parameters for a TW Linac**

| Parameter | Value |
|---|---|
| Max energy $V_0$ | 18 MV |
| Length $L$ | 2.0 m |
| Mode | $kd = 2\pi/3$ |
| Shunt impedance length ($r_s$) | 60 MΩ/m |
| Group velocity ($v_g/c$) | 0.01 |
| Maximum input power ($\langle P_0 \rangle$) | 5 MW |
| Attenuation $\tau$ | 0.4 |
| $I$ (at max x-ray output) | 200 mA |
| Max electric field $E_0$ | $10^7$ V/m |
| Accelerating gradient | 10 MeV/m |
| RF filling time | 0.7 μs |
| Vacuum | $10^{-10}$ bar |

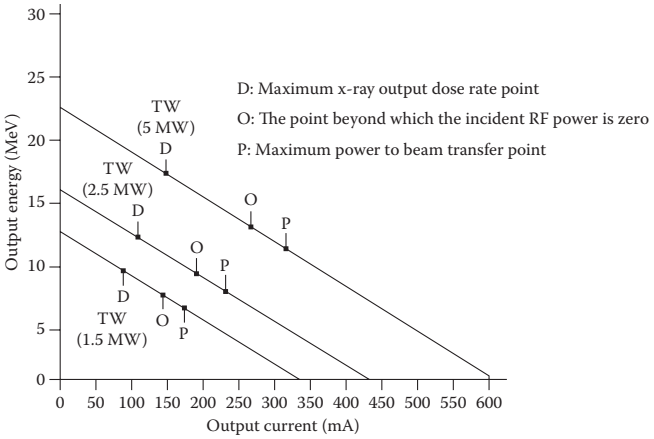



FIGURE 1.14 Load lines for an Elekta TW linac show the beam energy as a function of beam current (during a pulse) in x-ray mode. There are three load lines shown for input powers of 5, 2.5, and 1.5 MW for high X, medium X, and low X. The point at which the dose rate is maximum is indicated by a D. The point of maximum beam power is indicated by a P. (From *Elekta Corrective Maintenance Manual, HT and RF System*, 2011, Elekta Limited, Crawley, UK, section 4.45 p. 4–11. Courtesy of Elekta.)

and in turn this implies that

$$\langle P\rangle = \langle P_0\rangle\left[1-\left(1-e^{-2\tau}\right)\frac{z}{L}\right]. \tag{1.105}$$

Equation 1.94 implies that $r_s$ is constant. For the constant-impedance case, $r_s$ is also constant; however, $E(z)$ and $\langle dP/dz\rangle$ are not. Equation 1.94 may be used to find the $z$ dependence of $\alpha$:

$$\alpha = \frac{1-e^{-2\tau}}{2\left[L-\left(1-e^{-2\tau}\right)z\right]}. \tag{1.106}$$

$\alpha$ depends on the properties of the waveguide. Equation 1.106 and $r_s$ = constant are statements about the design of a constant-gradient waveguide, and therefore these conditions hold whether or not a beam current is present. A constant-gradient structure can be realized by decreasing the disk hole diameter with increasing $z$ to reduce the coupling between cells and make the energy stored per cell constant, thus resulting in a constant axial electric field strength (Ford, 1987). It appears as if the waveguide shown in Figure 1.10 is a constant-gradient design.

Let us now turn to the case in which there is a beam current. Equation 1.96 becomes

$$\frac{dE}{dz} = -\alpha I r_s, \tag{1.107}$$

where α is given by Equation 1.106. Equation 1.107 may be integrated to yield

$$E = \sqrt{r_s \langle P_0 \rangle (1 - e^{-2\tau})/L} + \frac{1}{2} r_s I \ln\left[1 - (1 - e^{-2\tau})\frac{z}{L}\right]. \tag{1.108}$$

We may now calculate the energy of the accelerated particles by using Equations 1.99 and 1.108:

$$V_e = \sqrt{(1 - e^{-2\tau})\langle P_0 \rangle r_s L} - \left(\frac{1}{2} - \frac{\tau e^{-2\tau}}{1 - e^{-2\tau}}\right) I r_s L. \tag{1.109}$$

This is the load line equation for a constant-gradient TW guide.

Inspection of an Elekta waveguide makes it appear as if it is a constant-impedance structure. The advantages given by Karzmark et al. (1993) for a constant-gradient structure are uniform power dissipation per unit length, higher beam power conversion efficiency, less sensitivity to frequency deviations, and less susceptibility to beam breakup. There is no explanation provided in this reference for the latter three advantages. A disadvantage is the extra difficulty in fabricating a structure that is not periodic.

### 1.7.4 RF RECIRCULATION

A TW accelerator does not use up all of the input RF power. The amount of RF power left at the end of the guide is on the order of $\langle P(L) \rangle / \langle P_0 \rangle \approx 30\%$ (no beam loading). This residual RF must not be allowed to be reflected; otherwise, standing waves will be set up in the guide. One solution is to simply dump the residual power in an "RF load." This may be a tapered section of waveguide where the RF is dissipated in the walls and in which a reflected wave is not formed. A water cooling line carries the heat away. The efficiency of a TW accelerator can be improved by recirculating some of the power that leaves the guide. Rather than dump all of the leftover power in a load, it can be recirculated. This makes it possible to produce a higher-energy beam. The power that leaves the waveguide is simply directed back to the input. The filling time for a TW linac is on the order of $L/v_g \sim 0.7$ μs. This is moderately short compared to the pulse length, and therefore it is feasible to recirculate the power. A schematic diagram showing the power flow path is displayed in Figure 1.15. The power flowing out of the end of the waveguide flows into a phase shifter. The phase shifter applies a phase shift to ensure that the recirculated power is in phase with the RF issuing from the magnetron so that they will add constructively. The phase shifter uses a mechanical plunger that must be in a different position for each beam energy. That portion of the returning power that is not in phase is dumped into a load.

High-energy Elekta linacs take advantage of power recirculation. This is shown in Figure 1.16, which is a diagram of the actual waveguide plumbing. An RF "isolater" prevents reflected RF from traveling back toward the magnetron. The RF

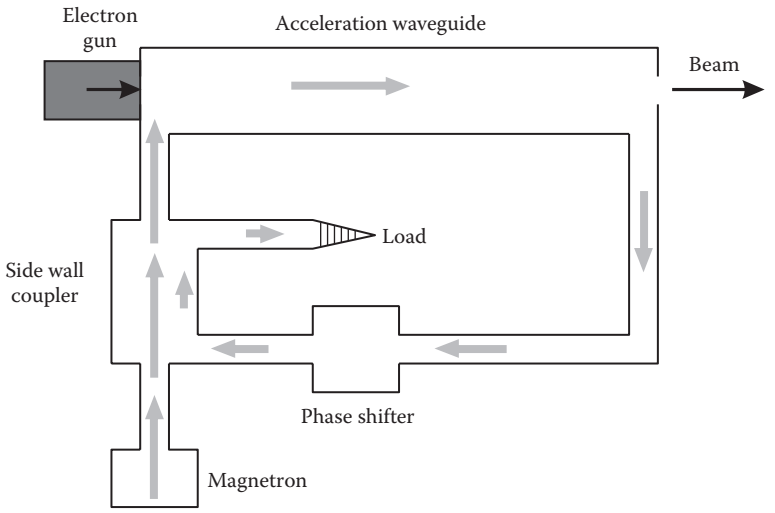



FIGURE 1.15 A schematic diagram of a TW accelerator in which the residual RF power is recirculated. The phase shifter ensures that power added to the source power from the magnetron is in phase with this power. The portion of the return power not in phase is dumped to an RF load.

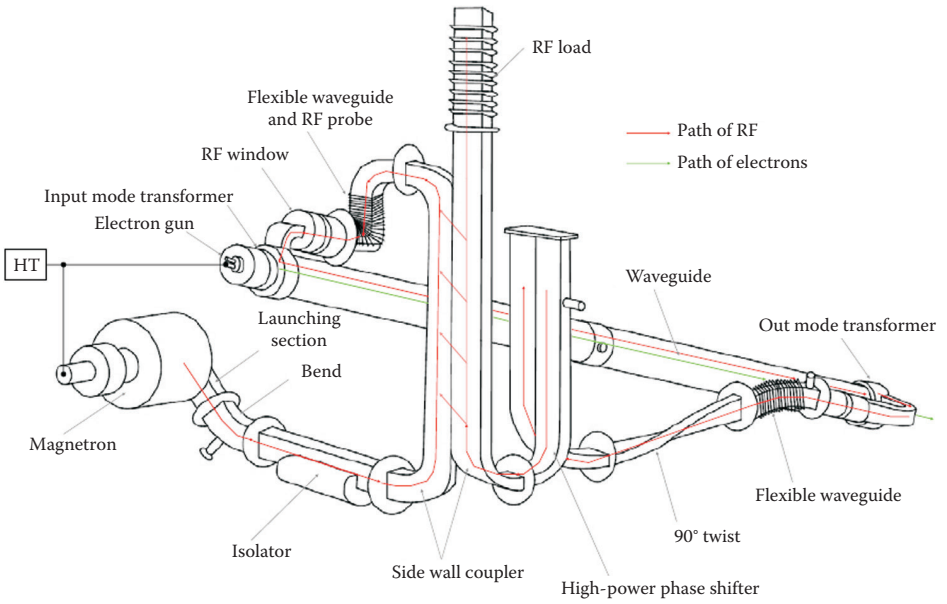



FIGURE 1.16 An Elekta high-energy linac with feedback. The path followed by the RF is in gray. The path followed by the electron beam is shown by the dashed line. The input/output mode transformers couple the waveguides with a circular cross section to the rectangular cross section transmission waveguides. The isolator prevents power from traveling back into the magnetron. (From *Elekta* Oncology Engineer 1, TTI Pre-Course Work, Digital Linear Accelerator - Beam Generation, EOE_1_PW5 Rev.A, 2009, Elekta Limited, Crawley, UK, p. 3. Courtesy of Elekta.)

load is where excess power (not in phase) is dumped. This is a tapered section of waveguide where the RF is converted to heat. The RF load is water cooled. The cooling line is shown wrapped around the waveguide. The input and output mode transformers couple the rectangular cross section power transmission waveguide to the circular cross section accelerating waveguide.

## 1.8 STANDING WAVE ACCELERATORS

The input end of the waveguide provides initial acceleration and bunching of the electrons. Recall from Section 1.5 that the resonant frequency of the cavities is independent of their length, and therefore bunching can be achieved by adjusting the axial length of the acceleration cavities at the input end. The buncher consists of just one or two cavities for the highest-energy x-ray beams or just one for lower-energy x-rays. In this section, we will concentrate our attention on the portion of the waveguide in which the electrons are highly relativistic. This portion is sometimes called the constant-velocity section.

A standing wave accelerator consists of a series of coupled cavity oscillators. The cavity oscillators are arranged end to end with a hole connecting them for both beam passage and field coupling. Before considering the accelerator as a whole, let us first examine acceleration of electrons in a single cavity. For a standing wave cavity, the electric field is changing temporally as the particle traverses the cavity. The effective accelerating potential is given by (see Equation 1.54 with $r = 0$)

$$V_0 = \int_{-l/2}^{+l/2} E_0 \cos \omega t \, dz, \tag{1.110}$$

where $l$ is the length of a single cavity, and we assume that the particle is in the middle of the cavity at time $t = 0$. In the absence of any energy losses, $E_0$ is a constant.[9] We can write $t = z/v$, where $v$ is the particle speed (presumed constant). We can now evaluate the integral in Equation 1.110:

$$V_0 = E_0 \frac{2v}{\omega} \sin \frac{\omega l}{2v} = E_0 l T, \tag{1.111}$$

where $T$ is called the *transit time factor* and

$$T = \frac{\sin w}{w} \text{ where } w = \frac{\omega l}{2v}. \tag{1.112}$$

The quantity $E_0T$ is like the average electric field experienced by electrons traversing the cavity. The accelerating potential is a maximum when $lT$ in Equation 1.111 is a maximum. The maximum possible value of $(\sin w)/w$ is 1.00. This will occur when the value of $w$ is small. For large values of $l$, $T$ becomes negative and the particle will

be decelerated. The largest value of $lT$ occurs for $w = \pi/2$.[10] In this case, $T = 2/\pi$, and therefore $l = 5.0$ cm. For these parameters, a 1.5 m accelerating waveguide requires approximately 30 cavities.

Figure 1.17 shows the electric field strength as a function of time as the particle traverses the cavity.

We define the shunt impedance per unit length of an SW resonant cavity as

$$r_s = \frac{E_0^2 l}{\langle P_w \rangle}. \tag{1.113}$$

Some references define the shunt impedance per unit length as $r_e = r_s T^2$. This is sometimes called the effective shunt impedance per unit length. For a cylindrical pillbox cavity of length $l$ and radius $R$ (see Equation 1.86),

$$r_s = \frac{\sigma \delta c^2 \mu_0^2}{\pi J_1^2(x_{01})} \frac{l/R}{R(1+l/R)}. \tag{1.114}$$

For a copper cylindrical cavity with values of $l = 5.0$ cm, $R = 3.83$ cm, $\nu = 3000$ MHz, the effective shunt impedance per unit length calculated from Equation 1.114) is about 110 MΩ/m. Typical values of the actual effective shunt impedance are 100–150 MΩ/m for a copper cavity at 3000 MHz and with $l = 5.0$ cm.

Equation 1.111 gives the energy of an electron that is accelerated across an isolated resonant pillbox cavity in the absence of beam loading. The energy of the accelerated electrons (no beam loading) can be expressed as

$$V_e = V_0 = T\sqrt{r_s l \langle P_w \rangle} = \sqrt{r_e l \langle P_w \rangle} \tag{1.115}$$

for a single cavity of length $l$.

The standing wave cavity has to be coupled to a transmission waveguide that supplies it with RF. A critically coupled cavity is one for which the rate of energy inflow is matched by the rate at which energy is absorbed by the walls.

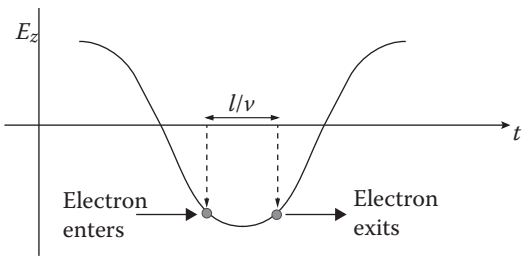



FIGURE 1.17 The acceleration of an electron traversing a resonant cavity of length $l$. The electric field strength in the cavity is spatially constant in the axial direction, but its amplitude oscillates in time. The particle crosses the cavity during a time interval in which the field strength is high.

In an SW linac, we have standing waves in each of the resonant cavities and the resonant cavities are coupled together. We have considered the resonant properties of individual cavities; we must now turn to a discussion of the coupling of these cavities.

### 1.8.1 COUPLED OSCILLATORS

So far we have considered only a single cavity. In an SW linac, the cavities are strung together in a line. Thus, the waveguide is a collection of coupled oscillators. The linac structure as a whole can sustain a large number of normal modes of oscillation. SW linacs can be analyzed by finding a suitable equivalent circuit model for each of the resonant cavities. Each of these circuit analog oscillators is then coupled to one another in a serial fashion (Ford, 1987).

There is a close analogy between coupled microwave cavities and coupled electrical circuits. One of the simplest circuits that illustrates the relevant behavior of a microwave cavity oscillator is an LC circuit. The resonant frequency of an isolated LC oscillator is $\omega = 1/\sqrt{LC}$, where $L$ is the inductance and $C$ is the capacitance.[11] To illustrate the effect of coupling, we consider the simplified case in which three cavities (circuits) are serially coupled to one another, as shown in Figure 1.18.

The circuits are magnetically coupled to one another through a mutual inductance $M$. As we will see later, a side-cavity SW waveguide does use magnetic coupling. We wish to analyze this coupled oscillator system and find the frequencies of oscillation of the system as a whole and the amplitude of the oscillation in each portion of the circuit.

We begin by applying Kirchhoff's loop law to each one of the oscillators:

$$\begin{aligned} L\dot{I}_1 + \frac{q_1}{C} + M\dot{I}_2 &= 0 \\ L\dot{I}_2 + \frac{q_2}{C} + M\dot{I}_1 + M\dot{I}_3 &= 0 \\ L\dot{I}_3 + \frac{q_3}{C} + M\dot{I}_2 &= 0, \end{aligned} \tag{1.116}$$

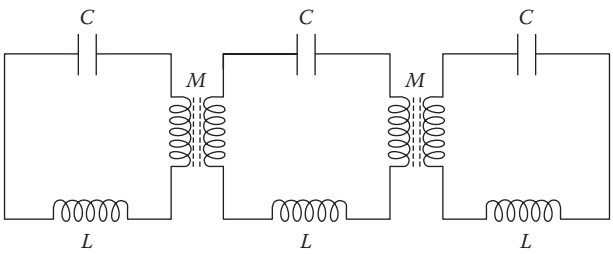



FIGURE 1.18 Three LC circuits coupled in a linear fashion are an analog to three coupled resonant cavities in an SW waveguide. Each of the LC circuits is identical. Coupling is through mutual inductance $M$. This system has three modes of oscillation, as described in the text.

where the $q$'s are the charges on the capacitors, $I$'s represent current in each loop, and the dots over the $I$'s represent time derivatives. Let us differentiate Equations 1.116 with respect to time to eliminate the $q$'s:

$$\begin{aligned} LC\ddot{I}_1 + I_1 + MC\ddot{I}_2 &= 0 \\ LC\ddot{I}_2 + I_2 + MC\ddot{I}_{1+} MC\ddot{I}_3 &= 0 \\ LC\ddot{I}_3 + I_3 + MC\ddot{I}_2 &= 0. \end{aligned} \tag{1.117}$$

We assume a solution of the form

$$\begin{aligned} I_1 &= A_1 e^{i\omega t} \\ I_2 &= A_2 e^{i\omega t} \\ I_3 &= A_3 e^{i\omega t}. \end{aligned} \tag{1.118}$$

Substituting Equations 1.118 into 1.117, we obtain

$$\begin{aligned} \left(\omega^2 LC - 1\right) A_1 + \left(\omega^2 MC\right) A_2 &= 0 \\ \left(\omega^2 MC\right) A_1 + \left(\omega^2 LC - 1\right) A_2 + \left(\omega^2 MC\right) A_3 &= 0 \\ \left(\omega^2 MC\right) A_2 + \left(\omega^2 LC - 1\right) A_3 &= 0. \end{aligned} \tag{1.119}$$

Equations 1.119 form a homogeneous system of linear algebraic equations in the three unknowns, $A_1$, $A_2$, and $A_3$. A necessary condition for a nontrivial solution is that the determinant of the matrix of the coefficients vanishes. Evaluation of the determinant of the coefficients in Equation 1.119 yields

$$\left(\omega^2 LC - 1\right)\left[\left(\omega^2 LC - 1\right)^2 - 2\left(\omega^2 MC\right)^2\right] = 0. \tag{1.120}$$

The roots of this equation are

$$\omega_0 = \frac{1}{\sqrt{LC}} \text{ and } \omega_{\pm} = \frac{\omega_0}{\left(1 \mp \sqrt{2} M/L\right)}. \tag{1.121}$$

The first of these, $\omega_0$, is the frequency of an isolated oscillator. The two frequencies $\omega_{\pm}$ are slightly above and slightly below the resonant frequency $\omega_0$ of an isolated oscillator. The system can oscillate with any one of these three frequencies. Notice

that as the coupling constant $M$ goes to zero, $\omega_{\pm} \rightarrow \omega_0$ as expected. Let us calculate the amplitudes of the currents ($A$ values) for the root $\omega_0$. Substitution of the expression for $\omega_0$ into Equation 1.119 shows that $A_2 = 0$ and $A_3 = -A_1$. There is no current in the middle oscillator, and the two oscillators on the ends are 180° or $\pi$ out of phase.

The mechanical analog of the coupled oscillators in Figure 1.18 are three coupled pendulums, as shown in Figure 1.19. For the mode shown in this figure, the central pendulum always remains at rest.

A similar phenomenon can occur when resonant cavities are coupled. The electric field in every other cavity is always (not just momentarily) zero. This feature of a standing wave linac seems surprising because we are accustomed to thinking of standing waves on a string. We have already seen in Section 1.5 that cavity oscillations do not act like standing waves on a string. In the case of a string, the node occurs at a single point, whereas for the coupled cavities, the node extends over the length of every other cavity. Standing wave modes are named for the phase difference between adjacent cavities: 0, $\pi/2$, $2\pi/3$, and so forth (see Figure 1.20). The length of the cavity can be matched to the particle velocity.

The node cavities do not contribute to acceleration; they serve only to couple microwave power from cavity to cavity, and therefore they may be moved off to the side and out of the beam line (see Figure 1.21). Such an arrangement is called a side-coupled standing wave linac. The cavities are coupled with a small hole that does not appreciably perturb the resonant properties. The two different types of cavities, coupling and accelerating, can then be optimized separately (see Figure 1.22). As the electric field in the coupling cavities is zero, the coupling is magnetic.

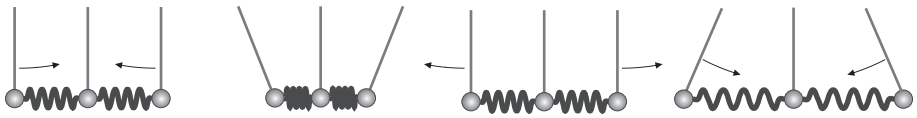


FIGURE 1.19 Three pendulums coupled by springs. This diagram shows the configuration of the pendulums at various instants for one of the normal modes of oscillation. For this normal mode, the central pendulum never moves.

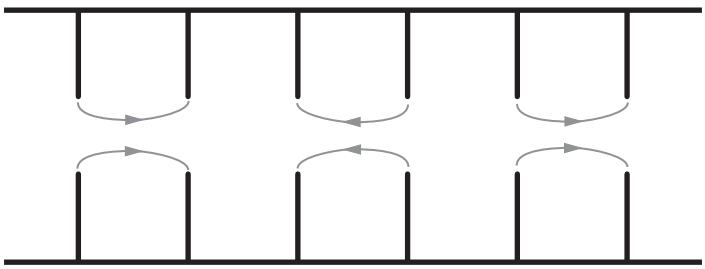


FIGURE 1.20 An SW waveguide operating in the $\pi/2$ mode. The phase shift between cavities is $\pi/2$. Every other cavity has a zero electric field at all times. This is the mode employed in standing wave medical linacs.

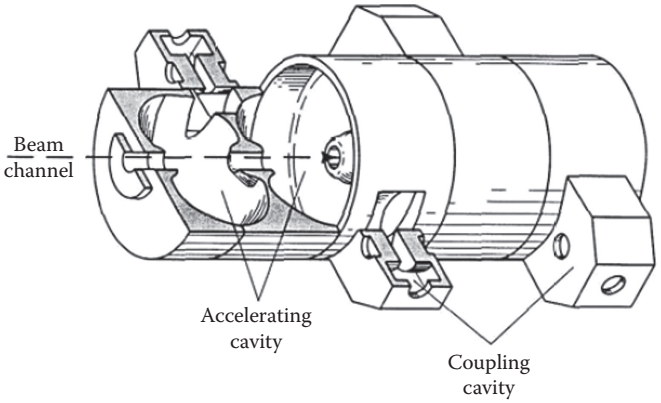



FIGURE 1.21 Perspective diagram showing a partial cutaway of a standing wave, side-cavity-coupled accelerator waveguide. Every other cavity has been moved off to the side. The purpose of the side cavities is to couple electromagnetic energy from one accelerating cavity to the next. (From Karzmark, C.J. and R.J. Morton, *A Primer on Theory and Operation of Linear Accelerators in Radiation Therapy*, 2nd ed, Madison, WI: Medical Physics Publishing, 1998, Fig. 33. Reprinted with permission from Medical Physics Publishing.)

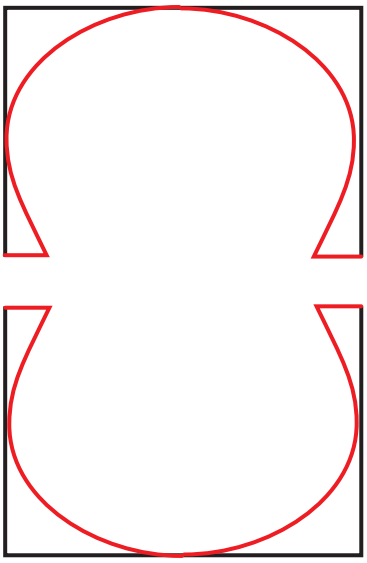


FIGURE 1.22 Optimization of cavity shape. The pillbox cavity is in black. The optimized cavity is in red. The surface area of the inside of the cavity is reduced to decrease power losses in the wall and increase $Q$. At the same time, the length of the cavity is (approximately) retained to maintain the energy gain of electrons that cross the cavity.

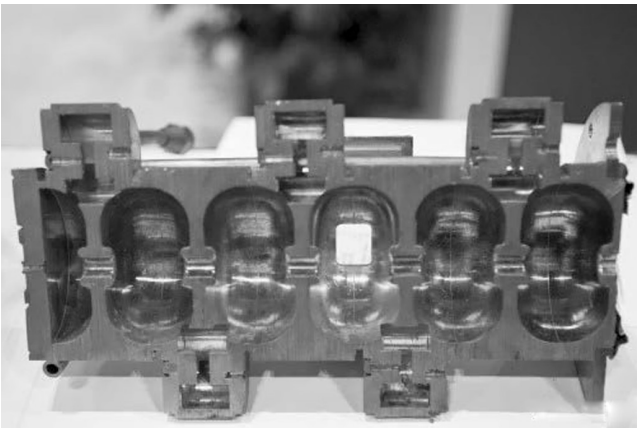


FIGURE 1.23 Photograph of a side-coupled standing wave waveguide that has been opened. There are 5(½) accelerating cavities. The RF coupling between accelerating cavities is through the side cavities (seen at the top and bottom). The electron gun is attached to the left side of the guide. RF is fed in through the aperture in the center cavity. (Courtesy of Dave Bullock, eecue.com.)

The accelerating cavity shape is optimized to minimize power losses in the walls. The idea is to reduce the surface area of the cavity without compromising the acceleration of the electrons. We want to keep the length of the cavity more or less the same. See Figure 1.22. Figure 1.23 shows a cutaway view of a side-cavity-coupled SW waveguide.

Figure 1.24 illustrates the operation of a side-cavity-coupled SW accelerator.

If the coupled cavities in an SW waveguide are relatively independent of one another, then the total energy gain (no beam loading) is given by $N$ times the expression in Equation 1.115, where $N$ is the number of *accelerating* cavities. If we further assume that the total length of the waveguide is $L = Nl$ and the total power dissipated in the walls is $\langle P_0 \rangle = N\langle P_w \rangle$, then the no-beam-load energy is

$$V_0 = \sqrt{r_e L \langle P_0 \rangle}, \tag{1.122}$$

where we have assumed side-cavity coupling and neglect any extra contribution to the length due to the side cavities.

## 1.8.2 STANDING WAVE LOAD LINE

The circuit diagram shown in Figure 1.25 is an analog for an SW accelerating waveguide. The RF power supply is in parallel with the shunt resistance and the beam load. This circuit does not apply to a TW waveguide because it does not account for RF power exiting the guide. We assume perfect coupling between the power source and the accelerating waveguide, meaning that all of the power $\langle P_0 \rangle$ supplied by the magnetron enters the waveguide. The beam energy is numerically equal to the

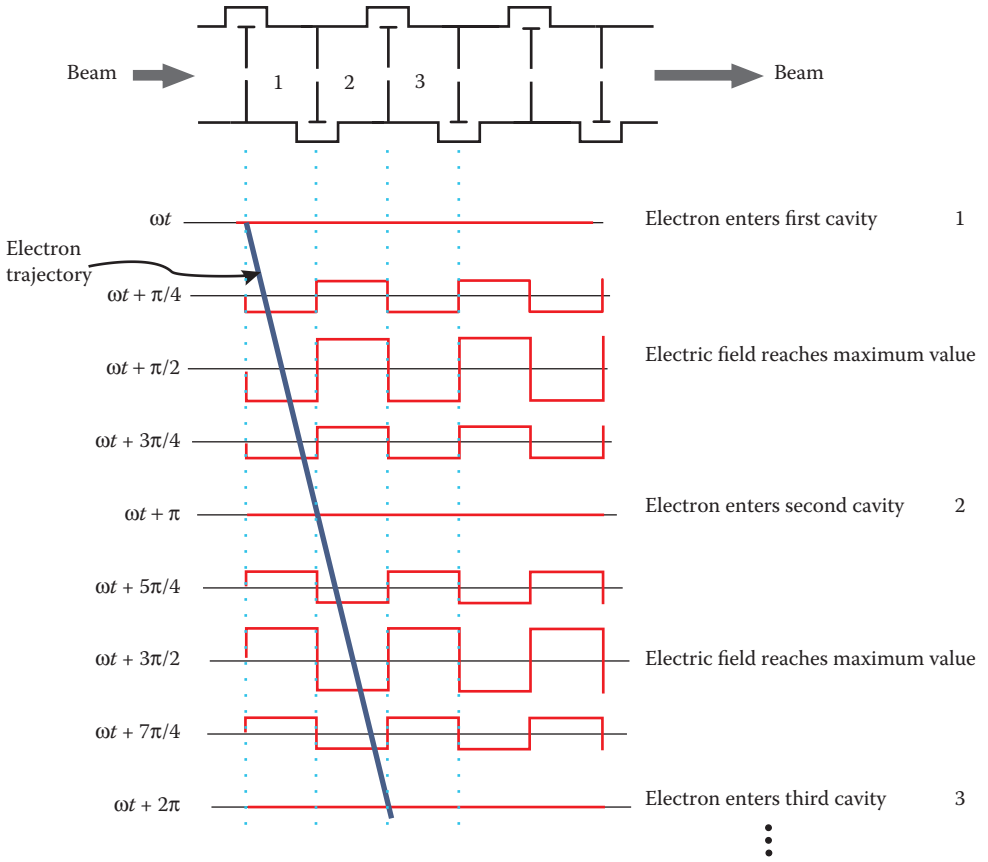



FIGURE 1.24 A side-cavity-coupled standing wave accelerator. The electric field configuration is shown at various instants in time. A negative electric field indicates that the electric field points toward the left. The trajectory of an electron bunch is illustrated by the oblique line. When an electron bunch enters cavity 1, the electric field is zero in that cavity. As the electrons traverse this cavity, the field strength rises and reaches a maximum when the electrons are near the middle of the cavity. As the electrons exit the cavity, the field strength has returned to zero. The electrons enter the second cavity just as the field in this cavity begins to go negative.

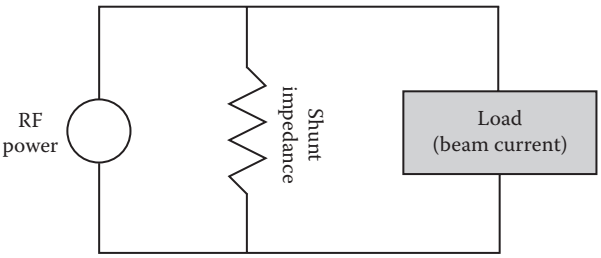



FIGURE 1.25 An equivalent circuit for an SW accelerator. The beam energy is numerically equal to the potential difference across the load. The shunt impedance drains power even in the absence of a beam current. When the beam is on, the input power must be shared between the shunt impedance and the beam.

potential difference across the load in the absence of beam loading. There is no current traveling down the beam load branch of the circuit when the electron beam current in the linac is zero. This is the reason for the name *shunt* resistance, because this "resistor" drains power even in the absence of beam current. It now becomes clearer why it is desirable to have a high shunt impedance—the higher the shunt impedance, the more power that is available to the beam.

In the no-load case, all of the power supplied by the power source (magnetron or klystron) is dissipated in the shunt resistance (walls of the waveguide) and $\langle P_0 \rangle = V_0^2/(r_e L)$; therefore,

$$V_0 = \sqrt{r_e L \langle P_0 \rangle}, \tag{1.123}$$

in agreement with Equation 1.122.

Beam loading is represented by a beam current $I$ traveling down the load branch of the circuit. Now the power must be shared; as a result, the potential difference will decease to a value $V_e$. We can write the equation for power sharing as

$$\langle P_0 \rangle = \frac{V_0^2}{r_e L} = \frac{V_e^2}{r_e L} + IV_e. \tag{1.124}$$

We can solve this equation for $V_e$ to obtain

$$V_e = V_0 \left\{ \sqrt{1 + \left( \frac{Ir_e L}{2V_0} \right)^2} - \left( \frac{Ir_e L}{2V_0} \right) \right\}. \tag{1.125}$$

To the extent that the equivalent circuit is a good analog of the accelerator, Equation 1.125 is the nonlinear load line for an SW accelerator. If $Ir_e L/(2V_0) \ll 1$, then to first order in this small quantity,

$$V_e \approx \sqrt{\langle P_0 \rangle r_e L} - \frac{1}{2} Ir_e L. \tag{1.126}$$

For typical linac parameters, $r_e = 100$ MΩ/m:

$$V_e = 10\sqrt{\langle P_0 \rangle L} - 50IL, \tag{1.127}$$

where $V_e$ is in units of MeV, $\langle P_0 \rangle$ is in units of MW, $I$ is in units of A, and $L$ is in units of m.

Karzmark et al. (1993) give data for a Varian Cl 1800 (no longer sold). The length of the waveguide is 1.5 m. There are 28½ accelerating cavities, the frequency of operation

is 2,856 MHz, the effective shunt impedance is 102 MΩ/m, the cavity $Q$ is 15,200, and the coupling factor between cavities is 0.04. The load line is illustrated in Figure 1.26. The load line for the 18 MV beam agrees well with Equation 1.127, with $L = 1.5$ m and $\langle P_0 \rangle = 3.8$ MW For the 6 MV load line, the effective length of the waveguide is shorter than 1.5 m (see Section 1.11). Table 1.2 gives nominal values of SW waveguide parameters.

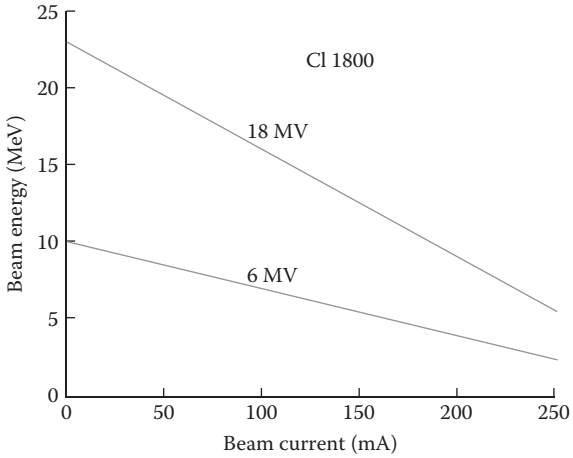



FIGURE 1.26 Load line for a Varian Cl 1800. The top load line is for an 18 MV beam with input power of 3.8 MW. The bottom load line is for a 6 MV beam with input power of 1.2 MW. (Data from Karzmark, C.J., et al., *Medical Electron Accelerators* [out of print], New York: McGraw Hill, 1993.)

**TABLE 1.2 Nominal Parameters for an SW Linac**

| Parameter | Value |
|---|---|
| Energy | 6/18 MV |
| Length | 1.5 m |
| Mode | π/2 |
| Maximum input power | 5 MW |
| $r_e$ (shunt impedance) | 100 MΩ/m |
| $T$ (transit time factor) | 0.91 |
| Accelerating gradient | 20 MeV/m |
| Number of cavities | 30 |
| Cavity length | 5.0 cm |
| Cavity diameter | 7.7 cm |
| Cavity $Q$ | 15,000 |
| Vacuum | $10^{-10}$ bar |

## 1.9 PULSED OPERATION AND WAVEFORMS

Linear accelerators operate in a pulsed mode. There are pulses on two timescales. Electrons are accelerated in bunches. The temporal spacing of these individual bunches is about equal to the microwave period ~0.3 ns. These subpulses are only delivered while the RF power is on. The RF itself is pulsed on a much longer timescale. The RF is pulsed because the power that is necessary to produce the strong electric fields required is quite large, on the order of 5 MW. Such very high instantaneous power levels are difficult to sustain for more than short intervals. A typical pulse duration is 5 μs, and the pulse repetition frequency (prf) is about 200 pulses/s (see Figure 1.27). The time averages discussed above (Section 1.7) are valid within an RF pulse because they are taken over the period of the microwaves, which is essentially instantaneous by comparison. The duty cycle for a linac is defined as $\zeta$ = RF pulse duration/RF pulse period (pulse duration × prf), and a typical value is $10^{-3}$. The peak current (averaged over the subpulses in an RF pulse) is typically about 100 mA in x-ray mode; therefore, the average current is $I_{avg} = I\,\zeta = 100$ μA. For electrons, the current is about 1 mA during a pulse. The gun current needs to be up to three times larger than the beam current because of inefficiency in capturing injected electrons. The dose rate depends on the beam current during an RF pulse and the duty cycle. At fixed energy, the dose rate is proportional to $I\,\zeta$ (the average beam current).

Karzmark et al. (1993) give a table (9-1) of average beam currents. For low X, the average beam current is about 100 μA with a prf of 300 $s^{-1}$ for a dose rate of 400 cGy/min. For high X, the average beam current is 30 μA with a prf of 150 $s^{-1}$ for a dose rate of 500 cGy/min. For electron beams, the average beam current is on the order of 100 nA for a dose rate of 500 cGy/min. Linac dose rates are changed by changing the prf. Changes in the beam pulse current would change the beam energy.

During RF injection in an SW linac, it takes about 1 μs for the fields to build up (filling time), during which there are transients. Therefore, the injection of electrons

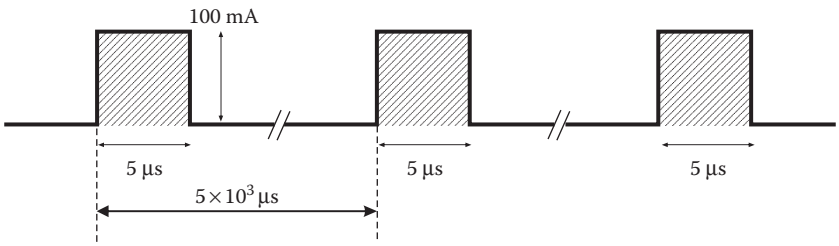



FIGURE 1.27 Pulsed operation of a medical linac in x-ray mode. The RF pulses are nominally 5 μs in duration. There are typically about 200 pulses/s. The peak beam current is approximately 100 mA in x-ray mode. The RF pulses consist of subpulses (not shown) that correspond to individual bunches of electrons. These subpulses are spaced about 0.3 ns apart.

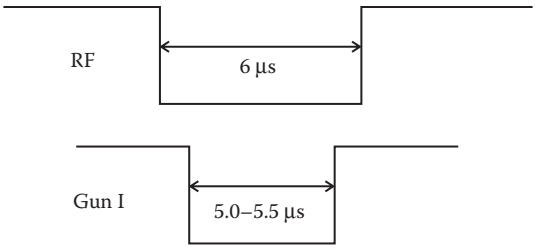



FIGURE 1.28 RF and gun current pulse applied to an SW waveguide. The RF turns on first and fills the guide before the gun current is applied.

is delayed. The RF pulse width is approximately 6 μs in duration. The gun current is on for about 5–5.5 μs (see Figure 1.28).

At the end of each pulse of RF power, the standing wave decays due to losses in the cavity. The timescale for this needs to be short compared to the time between RF pulses; otherwise, the incoming microwave pulse may not be in phase with the previous one. As we have seen in Section 1.6.2, the decay time for the fields in a cavity is on the order of 1 μs; this is much shorter than the time between RF pulses.

The average microwave power is on the order of $\zeta \times 5$ MW $= 5$ kW. It is for this reason that power consumption of a medical linac is not much higher than that of a household electric range (Podgorsak et al., 1999). The Elekta planning guide (2005) quotes a *total* machine power requirement of 30 kVA with beam on. The power factor is 0.6, which means that the true power consumption is 18 kW.[12] The Varian installation data package (for a Clinac *i*X, Trilogy, 2100C(D), 21EX and 23EX) (Varian, 2004) quotes a beam on load of 45 kVA (power factor 0.9). A household electric range can require up to about 12 kW.

## 1.10 FREQUENCY STABILITY AND FABRICATION OF WAVEGUIDE STRUCTURES

A common specification of electron beam energy is $d_{80}$, the depth at which the percent depth dose is 80%. The tolerance for the value of $d_{80}$ is typically ±2 mm. For therapeutic electron energies, the value of $d_{80}$ (in cm) is $V_e/3$ to a good approximation, where $V_e$ is in units of MeV. For a 20 MeV electron beam, it follows that the relative uncertainty in the beam energy $\Delta V_e/V_e = 3\%$. The necessary energy stability for electron beams is often quoted as ±1% (Ford, 1987; Scharf, 1994; Karzmark et al., 1993).

An energy slit is usually used in the bending magnet system to reduce the spread in the electron energy. According to Karzmark et al. (1993), the beam exiting the waveguide may have an energy spread of as much as 20%. The energy slit reduces the spread to about 1%. Beam current can be wasted in electron mode because the current requirements are already small compared to those of x-ray mode.

### 1.10.1 TRAVELING WAVE FREQUENCY STABILITY

A change in the phase will change the beam energy (see Equation 1.99). A change in phase can result from a change in frequency. Let us work out the sensitivity of the beam energy to the frequency of the microwaves. We assume a small change in phase $d\varphi$. Let us expand the cosine in Equation 1.99 around $\varphi = kz - \omega t = 0$: $\cos\varphi \approx 1 - \frac{1}{2}(d\varphi)^2$. We also have $d\varphi = z\,dk - t\,d\omega$, and therefore

$$d\varphi = \frac{z}{v_p}\left(\frac{v_p}{v_g} - 1\right)d\omega. \tag{1.128}$$

We can now relate the relative change in beam energy to relative changes in frequency using Equation 1.99, assuming that beam loading and wall losses are neglected (i.e., $\alpha = 0$ and $I = 0$):

$$\frac{\Delta V_e}{V_e} \approx -\frac{1}{6}\left(\frac{v_p}{v_g} - 1\right)^2 (kL)^2 \left(\frac{\Delta\omega}{\omega}\right)^2. \tag{1.129}$$

The value of $kd = 2\pi/3$, and therefore $kL = 2\pi/3\,(L/d)$, for $d = 3$ cm and an accelerator length of 2.2 m: $kL = 150$. If we demand an energy stability of 20%, this implies that $\Delta\nu/\nu = 7 \times 10^{-5}$. The dimensional tolerances that are often quoted (Ford, 1987; Scharf, 1994) associated with this are 0.01 mm. The coefficient of thermal expansion of copper is $17 \times 10^{-6}$ °C$^{-1}$. This corresponds to a temperature stability of approximately 2°C. This is difficult to accomplish given the power that is absorbed by the walls of the waveguide. In practice, stability is maintained by a combination of temperature control and input frequency tuning.

Waveguides are surrounded by a water jacket with flowing water. The purpose of the water jacket is not only to carry away excess heat, but also to maintain nearly constant temperature conditions. The internal cooling water for Elekta linacs is maintained at 30°C ± 1°C by a temperature servo probe.

### 1.10.2 STANDING WAVE FREQUENCY STABILITY

Our analysis of standing waves in a cavity in the absence of dissipation showed that there are discrete resonant frequencies at which the cavity will sustain electromagnetic oscillations. In the absence of dissipation, the cavity must be excited at precisely the resonant frequency in order to oscillate. With the presence of dissipation, there will instead be a narrow band of frequencies that will lead to excitation. Equation 1.89 shows how the energy decays with time. The energy is proportional to the square of the amplitude of the electric field. The time dependence of the electric field can therefore be written

$$E(t) = E_0 e^{-\omega_0 t/2Q} e^{-i(\omega_0 + \Delta_\omega)t}, \tag{1.130}$$

where $\omega_0$ is the resonant frequency, and we have allowed for the inclusion of a small shift $\Delta\omega$ in the resonant frequency due to damping (Jackson, 1999). This expression for the electric field incorporates a superposition of frequencies around $\omega = \omega_0 + \Delta\omega$ as follows:

$$E(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} E(\omega) e^{-i\omega t} d\omega. \tag{1.131}$$

The frequency components of the electric field in Equation 1.130 are given by

$$E(\omega) = \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} E_0 e^{-\omega_0 t/2Q} e^{-i(\omega_0 + \Delta\omega)t} dt. \tag{1.132}$$

This is an elementary integral. Evaluation of this integral (1.132) shows that the energy distribution in the cavity is given by

$$\left|E(\omega)\right|^2 \propto \frac{1}{\left(\omega - \omega_0 - \Delta\omega\right)^2 + \left(\omega_0/2Q\right)^2}. \tag{1.133}$$

Equation 1.133 represents a resonance line shape for the frequency distribution of the power in the cavity. A sketch of this is shown in Figure 1.29.

The full width at half maximum is $\Gamma = \omega_0/Q$. For a cavity with a $Q$ value of $10^4$, this is about $1.5 \times 10^6$ s$^{-1}$. For beam energy stability on the order of 20%, the energy in a cavity (Equation 1.133) cannot vary by more than 40% (see Equation 1.124). The 60% power level corresponds to approximately $\omega = (\omega_0 + \Delta\omega) \pm 0.41\ \Gamma$. This in turn corresponds to a change in frequency of $\Delta\nu \approx \pm 4$ MHz. The dimensions of the waveguide must be extremely accurate. From Equation 1.53, we see that $\Delta R = (\Delta\omega/\omega_0)R \approx 5 \times 10^{-3}$ cm. The value widely quoted in the literature is $10^{-3}$ cm (Karzmark, 1984). A single sheet of 20 wt paper has a thickness of $10^{-2}$ cm. The precision requirement is a tenth of this! Machining to this precision is not feasible using standard techniques (see Section 1.10.3).

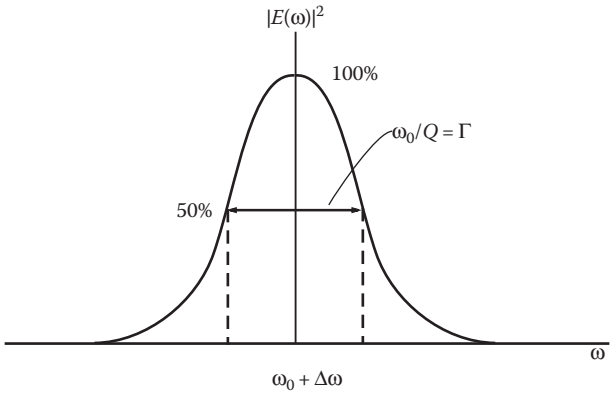



FIGURE 1.29 The line shape for a cavity resonator showing the distribution of energy with frequency. The full width at half maximum is given by $\omega_0/Q$. In the absence of dissipation, the line shape would be a $\delta$ function centered on the resonant frequency $\omega_0$.

The automatic frequency control (AFC) system is used to match the RF source frequency with the resonant frequency of the waveguide. Waveguide resonant frequency can change due to temperature changes, variations in input power, beam loading, and so forth. The AFC system utilizes feedback to adjust the RF source frequency. For a magnetron, a motorized plunger is used to adjust the frequency.

### 1.10.3 FABRICATION

Fabrication of accelerator waveguides is a difficult and exacting process. Waveguides are tested at low power and are fine-tuned by introducing mechanical deformations in order to adjust the resonant frequency of each cavity. Waveguides are constructed from 99.99% pure oxygen-free electronic (OFE) grade copper (Whittum, 1998). They are assembled in segments and then brazed together (Ford, 1987). Arcing in the guide due to metal burrs or imperfections could be very destructive. For this reason, guides must be carefully cleaned and polished. The cavities in a TW waveguide are fine-tuned after construction by mechanically squeezing them to produce slight ($10^{-3}$ cm) changes in the dimensions (Karzmark and Morton, 1998). In light of these considerations, it becomes easier to understand why the cost of an accelerating waveguide is in excess of $100,000.

## 1.11 CHANGING BEAM ENERGY

Modern medical linacs have two or three x-ray energies and five or six different electron energies. This discussion focuses on the mechanism by which the beam energy is changed. This may be different than the method used to *adjust* or fine-tune a particular beam energy The high-energy x-ray beam is sometimes called high X and the low-energy beam low X.

### 1.11.1 TRAVELING WAVE

The output energy depends on input power level, RF frequency, and beam current. The wave velocity is critically dependent on frequency (see point *P* in Figure 1.9). For stable operation, the frequency has to be accurately controlled, as we have seen. A small change in frequency can be used to produce large variation in electron energy. In a traveling wave waveguide with RF feedback, variable coupling is used to maintain the electric field in the buncher for both high and low X.

To change the energy of the x-ray beam, the accelerator is run at different RF power levels. See the Elekta load lines in Figure 1.14 showing the different powers used to produce beams of different energies.

There are several possibilities for changing electron beam energy. One method is to alter the position of the electrons with respect to the phase of the accelerating wave by changing the frequency slightly. This is known as detuning. In this way, the electrons will slip in phase, receiving less than the maximum acceleration. This makes it

possible to vary the energy of the electrons from a maximum value down to a few MeV. This appears to be the method used by Elekta service technicians to *adjust* electron beam energy (Elekta, 2009). Another possibility is beam loading. The beam current is increased, keeping the RF power constant, thereby reducing the beam energy.

### 1.11.2 STANDING WAVE[13]

There are numerous possible methods for varying the beam energy discussed in Karzmark et al. (1993). It is possible to change beam energy by varying the RF power into the guide or by increasing the beam current exploiting beam loading. The problem with these methods is that they lead to an energy spread in the accelerated beam. Any method used to change energy should avoid broadening the spectrum.

We describe the method used in Varian linacs to change the x-ray energy. The guide is divided into two sections. The first section bunches the electrons and positions them in phase. The second portion of the guide accelerates the electrons without causing undue spread in the energy. Whatever is done to the second portion of the guide to change the energy must not interfere with the job of the first portion.

Beam energy is changed by changing the ratio of power fed to the first and second portions of the waveguide. This can be accomplished with the use of an "energy switch." This is a moving mechanical plunger in a side cavity between the first and second portions of the guide (see Figure 1.30). The plunger does not fully block the RF, but rather reduces the coupling between the first and second portions of the guide, allowing some RF through to the second portion of the guide as follows:

1. High X: There is high coupling between the first and second portions of the guide, and the electric field is high in both portions of the guide (see Figure 1.31).
2. Low X: There is low coupling between the first and second portions of the guide so that the electric field is reduced in the second portion of the guide (see Figure 1.31).

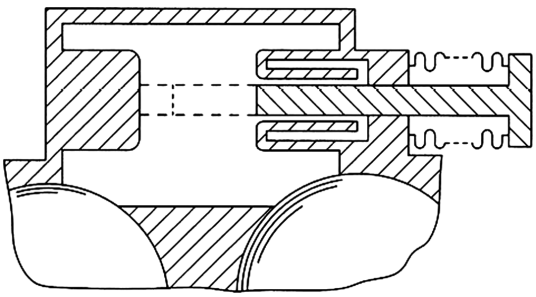


FIGURE 1.30 Energy switch for an SW accelerator. The dashed lines show the plunger inserted into one of the side coupling cavities part way down the waveguide. When the plunger is inserted, coupling between the first and second section of the guide is reduced. This in turn reduces the electric field strength in the second portion of the guide, lowering the final beam energy. See Figure 1.31. (From Karzmark, C.J., et al., *Medical Electron Accelerators* [out of print], New York: McGraw Hill, 1993, p. 194. Copyright and courtesy of McGraw-Hill Education.)

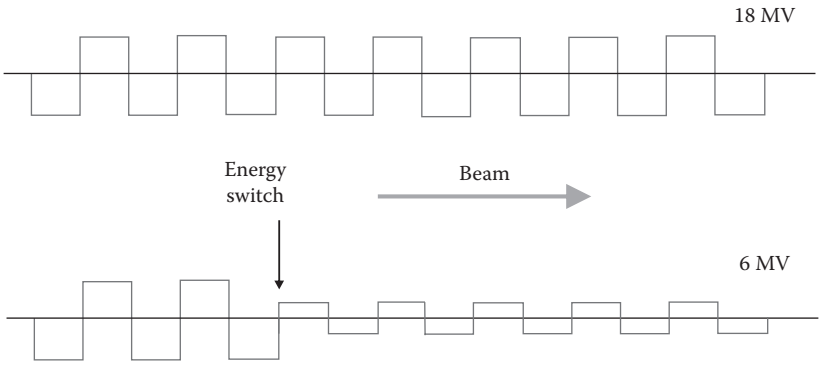



FIGURE 1.31 Shows the electric field along the length of the waveguide at some instant in time for an SW linac operating in the high-energy mode (18 MV, top portion of the figure) and low-energy mode (6 MV at bottom). In the bottom portion of the figure, the energy switch (see Figure 1.30) has been inserted. This causes a reduction in the electric field strength in the second portion of the guide and a consequent reduction in the final beam energy. (Based on figure 11-1 of Karzmark, C.J., et al., *Medical Electron Accelerators* [out of print], New York: McGraw Hill, 1993.)

The electric field in the first portion of the guide is the same in both cases to maintain optimum capture and bunching.

## 1.12 COMPARISON BETWEEN TW AND SW LINACS

In the book by Karzmark et al. (1993), a comparison is made between the virtues of SW and TW waveguides. These authors clearly prefer SW waveguides. Some of the reasons cited follow. We have already seen that SW structures are shorter than the equivalent energy traveling wave structure. The shape of the accelerating cavity can be optimized to maximize the electric field on the axis and minimize power losses in the walls. The coupling cavities can be optimized separately. Using half-wave or quarter-wave accelerating cavities results in a guide that is inherently mode stable, it is easier to avoid excitation of nearby resonant modes.

TW waveguides are longer than SW. The Elekta waveguide is 2.2 m in length and is oriented obliquely. TW waveguides do not require a klystron. High-energy SW linacs do require a klystron. Klystrons are more expensive and cannot be mounted on the gantry, although they do have a longer life. The Elekta linacs are capable of delivering three photon energies (e.g., 6, 10, and 18 MV).

There is little clinical distinction between TW and SW linacs in terms of performance characteristics, such as beam energies, dose rates, and ability to deliver intensity-modulated radiation therapy (IMRT) treatment.

According to Greene and Williams (1997), the mean power required by an SW machine is slightly higher because of the filling time. Given two 6 MV accelerators, the TW requires 2 MW and the SW requires 2.5 MW.

## 1.13 X-BAND LINACS

Most medical linacs operate in the s-band of the microwave spectrum at approximately 3 GHz. The x-band ranges from 8.0 to 12.0 GHz (Wikipedia). Increasing the microwave frequency allows construction of smaller and lighter linacs. Linear accelerators have been built that operate in the x-band of the microwave spectrum. The disadvantage is that it makes the already exacting dimensional tolerances even more difficult to achieve. The stability and fabrication difficulties discussed in Section 1.10 become worse.

The Mobetron interoperative radiation therapy unit uses a mobile x-band linac designed by Intraop Medical (Figure 1.32). The Mobetron can produce electron beams


FIGURE 1.32 The Mobetron employs a mobile x-band linac for intraoperative radiotherapy. This compact linac produces electron beams ranging in energy from 6 to 12 MeV. (Courtesy of Intraop Medical Corporation, Sunnyvale, CA.)

**TABLE 1.3 Comparison between s-Band and x-Band Linacs[a]**

| Band | Frequency | Wavelength | Shunt Impedance | Acceleration Gradient | 6 MeV Accelerator Length |
|---|---|---|---|---|---|
| s | 3 GHz | 10 cm | 85–130 MΩ/m | 20 MeV/m | 30 cm |
| x | 11.4 GHz | 3 cm | 110–160 MΩ/m | 100 MeV/m | 6–8 cm |

*Source:* Data from Thorson, T., Advanced acceleration and image guidance technologies, online PowerPoint presentation, 2007, http://195.135.200.83/allegatiifo/Congresso2007/19aprile/Thorson.pdf.

with energies of 6, 9, and 12 MeV. The CyberKnife is a stereotactic radiosurgery system that consists of a 6 MV x-band linac on a robotic arm. The CyberKnife linac operates at 9.3 GHz and weighs 285 lb. Both the Mobetron and the CyberKnife use a 2.0 MW magnetron.

Thorson (2007) has suggested a number of reasons that x-band linacs have not become more widespread: (1) familiarity with s-band components, (2) reliability of s-band RF sources, (3) availability of components, (4) manufacturing tolerances, and (5) cost. Table 1.3 shows a comparison between some s-band and x-band linac characteristics.

## QUESTIONS

1. Why does a low-energy beam require higher current than a high-energy beam for the same dose rate?
2. What is a typical electron gun voltage?
3. How is a standing wave in a waveguide different than a standing wave on a string?
4. Why do s-band cavities in an SW waveguide have a radius of 3.83 cm?
5. Describe the energy budget of a traveling wave linac.
6. Write a generic expression for a linac load line and explain the physical meaning of each term.
7. Why is a TW waveguide much longer than an SW waveguide?
8. Why is it desirable to have a high shunt impedance?
9. What is the relationship between shunt impedance and conductivity?
10. What are typical average beam currents for x-ray mode and electron mode?
11. What are typical instantaneous and average power requirements for an SW and a TW linac?
12. How is x-ray beam energy changed in an SW and a TW linac?
13. What are the practical differences between SW and TW waveguides?
14. Why is waveguide temperature stability so crucial?
15. Why are waveguides so expensive?

# PROBLEMS

1. Derive Equations 1.23 and 1.24.

2. Estimate the disk spacing in a TW waveguide operating in the "2π/3" (i.e., $k = 2\pi/3d$) mode.

3. Derive Equation 1.69 and show that the solution is given by Equation 1.71. The BAC-CAB rule for the vector cross product may be useful: $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A} \cdot \mathbf{C}) - \mathbf{C}(\mathbf{A} \cdot \mathbf{B})$.

4. Derive Equation 1.101 from Equations 1.97, 1.99, and 1.100.

5. Show that in the absence of beam loading and wall losses, Equation 1.101 reduces to $V_e = E_0 L$.

6. For a load line of the form $V_e = V_0 - F\,I$, where $F$ is a constant, (a) show that the beam power is maximum when $V_e = V_0/2$ and (b) show that the maximum x-ray dose rate during an RF pulse occurs for $V_e = (3/4)\,V_0$. Assume that the x-ray output is proportional to $IV_e^3$.

7. a. Calculate the value of α for a smooth-bore (no disks) TW waveguide. Use the condition $v_g = 0.01c$ and $R = 3.8$ cm (note that this implies $v_{ph} = 100c$!).

   b. Calculate the value of τ for $L = 2.0$ m.

   c. Calculate the residual power exiting the waveguide for no beam loading.

8. a. A constant-gradient TW guide has an electric field amplitude $E_0$ that remains constant down the guide (increasing $z$). In this case, one would expect that $V_0$ (beam energy in the absence of beam loading) should be $E_0 L$. Show this.

   b. For a constant-impedance TW guide, write $V_0$ in terms of $E_0 L$. Use typical values of waveguide parameters to compare the result to that found in part (a).

9. In going from Equation 1.125 to 1.127, it is assumed that $Ir_e L/(2V_0) \ll 1$. Check the validity of this for parameters $I = 100$ mA, $r_e = 100$ MΩ/m, $L = 1.5$ m, and $V_0 = 18$ MV.

10. Estimate the shunt impedance/length for an x-band SW guide operating at a frequency of 11.4 GHz with $w = \pi/2$. Compare your answer to the table entry in Section 1.13.

# SYMBOLS

| | |
|---|---|
| $a$ | Area |
| $A$ | Area |
| $d\mathbf{A}$ | Surface area vector, normal to surface |
| $\dot{D}$ | Dose rate |
| $b$ | Radius of disk iris |
| $\mathbf{B}$ | Magnetic field vector |
| $\mathbf{B}_c$ | Magnetic field vector in conductor |
| $B_r$, $B_\theta$, $B_z$ | Vector components of $\mathbf{B}$ |
| $c$ | Speed of light in vacuum |
| $C$ | Capacitance |
| $d$ | Distance between disks in TW guide |
| $d_{80}$ | Depth at which percent depth dose (PDD) is 80% for electrons |
| $\mathbf{D}$ | Electric displacement vector |
| $\mathbf{D}_c$ | Electric displacement vector in medium |
| $\mathbf{E}$ | Electric field vector |
| $\mathbf{E}_c$ | Electric field vector in conductor |
| $\mathbf{E}_n$ | Normal component of electric field |
| $\mathbf{e}_r$, $\mathbf{e}_\theta$, $\mathbf{e}_z$ | Unit vectors in $r$, $\theta$, and $z$ directions |
| $E_r$, $E_\theta$, $E_z$ | Vector components of $\mathbf{E}$ |
| $F$ | Slope of generic load line |
| $\mathbf{H}$ | Magnetic field strength |
| $\mathbf{H}_n$ | Normal component of $\mathbf{H}$ |
| $\mathbf{H}_c$ | Value of $\mathbf{H}$ in conductor |
| $\mathbf{H}_t$ | Component of $\mathbf{H}$ tangent to surface |
| $i$ | $\sqrt{-1}$ |
| $I$ | Average beam current during an RF pulse (or current in oscillator circuit, Section 1.9) |
| $\dot{I}$ | Time derivative of current in oscillator circuit |
| $\ddot{I}$ | Second time derivative of current in oscillator circuit |
| $I_{avg}$ | Average beam current (over many RF pulses) |
| $\mathbf{J}$ | Current density |
| $J_m$ | Bessel function of index $m$ |
| $k$ | Wavenumber |
| $\mathbf{K}$ | Surface current |
| $\mathbf{K}_{eff}$ | Effective value of $\mathbf{K}$ |

| | |
|---|---|
| $\mathbf{l}$ | Unit tangent vector to a surface |
| $l$ | Length of cavity oscillator in SW |
| $L$ | Waveguide length (or inductance; Section 1.9) |
| $m$ | Index of Bessel function $J_m$ |
| $M$ | Mutual inductance |
| $\mathbf{n}$ | Unit normal vector to a surface |
| $N_m$ | Neumann function of index $m$ |
| $P$ | Microwave power |
| $P_0$ | Initial microwave input power |
| $P_w$ | Power loss in walls |
| prf | Pulse repetition frequency |
| $q$ | Charge |
| $Q$ | Quality factor for cavity oscillator |
| $r$ | Radial coordinate in cylindrical coordinates |
| $r_s$ | Shunt impedance |
| $r_e$ | Effective shunt impedance |
| $R$ | Inner radius of waveguide or cavity |
| Re | Take the real part of |
| $\mathbf{S}$ | Poynting vector |
| $t$ | Time |
| $t_d$ | Decay time for energy in cavity |
| $T$ | Transit time factor |
| $u$ | Energy density stored in fields |
| $U$ | Energy stored in fields |
| $U_l$ | Energy/length |
| $\mathbf{v}$ | Electron velocity vector |
| $v_g$ | Group velocity |
| $v_{ph}$ | Phase velocity of wave |
| $V_e$ | Electron beam energy |
| $V_0$ | Beam energy without beam loading |
| $x_{mn}$ | $n$th root of Bessel function of order $m$ |
| $z$ | Cylindrical coordinate |
| $\alpha$ | Microwave electric field attenuation |
| $\beta$ | Perturbed value of real part of $k$ |
| $\Gamma$ | Full width at half maximum for power in cavity |
| $\gamma^2$ | $\omega^2/c^2 - k^2$ |
| $\Delta A$ | Area element |

| | |
|---|---|
| $\Delta z$ | Small increment in $z$ |
| $\delta$ | Skin depth |
| $\nabla_t^2$ | Transverse portion of Laplacian |
| $\varepsilon$ | Permittivity |
| $\varepsilon_0$ | Permittivity of free space |
| $\zeta$ | Duty cycle factor |
| * | Complex conjugate |
| ′ | Spatial derivative with respect to $z$ |
| $\Theta$ | Angular function in separation of variables |
| $\theta$ | Angular coordinate in cylindrical coordinates |
| $\lambda$ | Wavelength |
| $\mu_0$ | Permeability constant ($1.26 \times 10^{-6}$ H/m) |
| $\mu_c$ | Value of $\mu$ in conductor |
| $\nu$ | Frequency |
| $\xi$ | Coordinate perpendicular to surface |
| $\sigma$ | Conductivity |
| $\tau$ | Attenuation factor |
| $\varphi$ | $kz - \omega t$ |
| $\omega$ | $2\pi\nu$ angular frequency |
| $\omega_c$ | Cutoff angular frequency |
| $\omega_0$ | Resonant frequency of LC circuit |
| $\omega_\pm$ | Resonant frequencies of coupled circuits |
| $\mathfrak{R}$ | Radial function in separation of variables |
| $\Sigma$ | Surface charge density |
| $\dot{\Phi}$ | Fluence rate |
| $\langle\rangle$ | Time average over one period |
| $\Omega$ | Ohms |

# REFERENCES


Arfken, G. 1985. *Mathematical Methods for Physicists*, 3rd ed. Orlando, FL: Academic Press.

Attix, F. 1986. *Introduction to Radiological Physics and Radiation Dosimetry*. New York: Wiley.

Beyer, W.H. 1978. *Standard Mathematical Tables*, 25th ed. West Palm Beach, FL: CRC Press.

Elekta Limited. 2005. *Digital Accelerator: Planning Guide*. Crawley, UK: Elekta Limited.

Elekta Limited. 2009. Beam physics system. In *Digital Accelerator Corrective Maintenance Manual*, 6–58. Crawley, UK: Elekta Limited.

Feynman, R.P., R.B. Leighton, and M. Sands. 1964. *The Feynman Lectures on Physics*, vol. II, chaps. 23 and 24. Palo Alto, CA: Addison-Wesley.

Ford, J.C. 1987. Advances in accelerator design. In *Radiation Oncology Physics—1986*, ed. J.G. Kereiakes, H.R. Elso, and C.G. Born, 24. Medical Physics Monograph No. 15. New York: American Institute of Physics.
Greene, D. and P.C. Williams. 1997. *Linear Accelerators in Radiation Therapy*, 2nd ed. Bristol: Institute of Physics Publishing.
Humphries, S. 2012. *Principles of Charged Particle Acceleration.* Mineola, NY: Dover. http://eecue.com/a/1499/ASTR0-07.html (accessed November 5, 2014).
Jackson, J.D. 1999. *Classical Electrodynamics*, 3rd ed. New York: Wiley.
Karzmark, C.J. 1984. Advances in linear accelerator design for radiotherapy. *Med. Phys.* 11(2), 105–28.
Karzmark, C.J. and R.J. Morton. 1998. *A Primer on Theory and Operation of Linear Accelerators in Radiation Therapy*, 2nd ed. Madison, WI: Medical Physics Publishing.
Karzmark, C.J., C.S. Nunan, and E. Tanabe, 1993. *Medical Electron Accelerators* [out of print]. New York: McGraw Hill.
Khan, F.M. 2010. *The Physics of Radiation Therapy*, 4th ed. Philadelphia: Lippincott, Williams & Wilkins.
Loew, G.A. and Talman, R. 1983. *Elementary Principles of Linear Accelerators*. SLAC PUB 3221. www.slac.stanford.edu/cgi-wrap/getdoc.
Podgorsak, E.B., P. Metcalfe, and J. Van Dyk. 1999. Medical accelerators. In *Modern Technology of Radiation Oncology*, ed. J. Van Dyk, 349–435. Madison, WI: Medical Physics Publishing.
Scharf, W.H. 1994. *Biomedical Particle Accelerators.* New York: AIP Press.
Thorson, T. 2007. Advanced acceleration and image guidance technologies. Online PowerPoint presentation. http://195.135.200.83/allegatiifo/Congresso2007/19aprile/Thorson.pdf.
U.S. Particle Accelerator School (USPAS). Course materials. http://uspas.fnal.gov/lect_note.html.
Varian Medical Systems. 2004. Varian installation data package (Clinac *iX*, Trilogy, 2100 C(D), 21EX and 23EX), Section 1, 1.39.1. Palo Alto, CA: Varian Medical Systems.
Whittum, D.H. 1998. *Introduction to Microwave Linacs*. SLAC PUB 8026. http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-pub-8026.pdf.
Wikipedia. X band. http://en.wikipedia.org/wiki/X_band.

## ENDNOTES

1. "That is, it takes 7.2 times as much beam current and 2.4 times as much beam power to produce the same dose rate at 6 MV as at 18 MV flattened over a 50-cm diameter at 100-cm SAD." (Karzmark et al. 1993, p. 18). This result is somewhat at odds with the linac operating parameters given in table 9-1 of Karzmark et al. (1993), which shows a factor of 5 difference in average beam current between 6 and 18 MV beams.
2. This accelerator operates at a frequency of 2856 MHz, and the waveguide inner diameter is approximately 8.2 cm.
3. Under these circumstances, power flow is maximized and losses are minimized (Humphries, 2012).
4. This value was found by using the spreadsheet function for Bessel functions in Microsoft Excel.
5. In some references, $Q$ is $1/(2\pi)$ times the definition given above.
6. $I$ is the average beam current during an RF pulse (see Section 1.9). This is an average over many individual bunches of electrons.

7. We may assume that the peak beam current is constant on the timescale of an RF pulse (see Section 1.9).
8. The reader may notice a discrepancy here between this result quoted by Ford and the statement in Section 1.1 that the flattened x-ray intensity is proportional to the 1.8 power of $V_e$. The latter dependence comes from the book by Karzmark et al. (1993, p. 18), in which it is stated, "Thus, the flattened x-ray intensity is proportional to about the 1.8 power of x-ray energy." Ford (1987, p. 25) states, "Hence the x-ray yield is given by X = B I(V) $V^3$ where B is a constant determined by the accelerator target and flattening filter." The flattening filter transmission actually depends on the −0.8 power of the energy according to Karzmark et al. (1993); thus, the quantity B in the previous sentence is not a constant.
9. Ignoring the hole in the center of the cavity through which the beam must pass.
10. This is for fixed electric field strength. A more realistic constraint would be fixed *power*.
11. In this section, $L$ represents inductance and not waveguide length.
12. The real power (in kW) is the apparent power in kVA multiplied by the power factor.
13. This follows the discussion in Karzmark et al. (1993), chapter 11.

# 2

# PROTON THERAPY PHYSICS

## *Protons for Pedestrians*

## 2.1 INTRODUCTION

According to Smith (2009a), 35% of cancer patients die as a result of a lack of local disease control. This suggests that there may be room for significant improvement if radiation therapy treatments can become more conformal. The goal of proton therapy is to increase the rate of local control and reduce morbidity.

In a 2008 point/counterpoint article in the journal *Medical Physics* (Maughan and Van den Heuvel, 2008), the subject of contention was that "within the next 10–15 years protons will likely replace photons as the most common type of radiation for curative radiotherapy." In 2016, 8 years later, proton therapy has grown rapidly; however, it does not appear that it will replace conventional photon therapy any time soon. While the effectiveness of proton therapy is still hotly debated, there is no debate about the cost. The cost of a proton therapy facility is at least an order of magnitude greater than a linac-based facility. According to Nafziger (2011), the United States will not witness the construction of 100 or 200 proton therapy centers as some have speculated, but more likely only 20–30. If innovative accelerators can be built that are small and cheap, the cost of proton therapy may come down to the point that it rivals photon therapy and the outlook may change. At the time of this writing, there are about 15 operating proton centers in the United States, and 12 new centers under development (National Association for Proton Therapy, 2015) (Figure 2.1). There are now approximately 40 proton therapy facilities outside the United States, and more than 105,000 patients have been treated worldwide as of the end of 2013 (PTCOG, 2014).

In contrast to some of the other material in this text, there is an abundance of literature on proton therapy, much of which is very good quality. The company IBA maintains a list of proton therapy references on its website: http://www.iba-proton-therapy.com/more-resources-pt. There are books by Ma and Lomax (2012), Paganetti

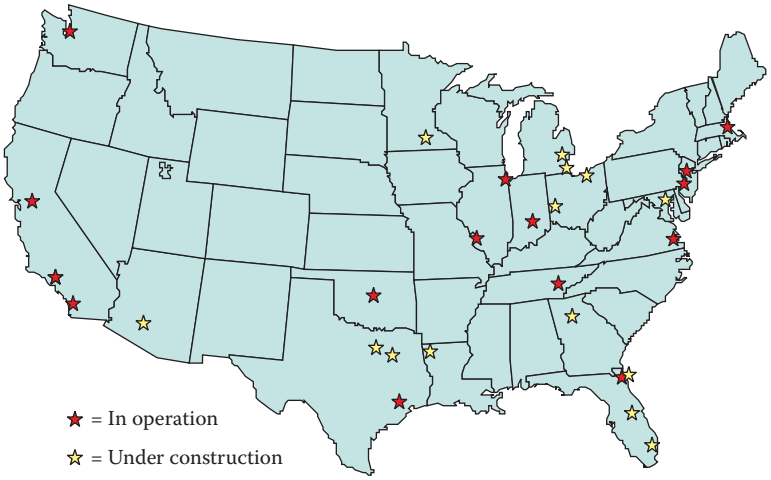



FIGURE 2.1 Existing proton therapy centers in the United States and centers under construction.

(2012b), Yajnik (2012), Metz and Thomas (2010), DeLaney and Kooy (2007), and Linz (2012). There are review chapters in *Modern Technology of Radiation Oncology*, Volumes I (Moyers, 1999) and III (Vatnitsky and Moyers, 2013), and a review article by Smith (2006). Chapters on proton therapy in the book by Goitein (2008) are especially a delight to read and are highly recommended. Two books on particle accelerators that are relevant to proton therapy are Humphries (2012) and Scharf (1994).

In Section 2.2, we provide a brief history of proton therapy. Section 2.3 contains a discussion of the interaction of protons with matter. Absorbed dose and the Bragg peak are covered in Section 2.4. A very brief overview of the radiobiological properties of protons is given in Section 2.5. The accelerators used for proton therapy are all circular machines, and it is therefore important to consider circular orbits of a charged particle in a magnetic field, including the stability of these orbits. This is the subject of Section 2.6. In Section 2.7, we discuss proton therapy accelerators and their characteristics. Once a proton beam is generated, it must be transported to the patient. This is described in Section 2.8. Section 2.9 involves a discussion of methods used to spread the raw narrow pencil beam. Dosimetric beam calibration is covered in Section 2.10. Dose calculation algorithms are discussed briefly in Section 2.11. Inhomogeneities play a crucial role in proton therapy, as discussed in Section 2.12. A sampling of dose distributions possible is explored in Section 2.13. The radiation shielding requirements for proton therapy are reviewed in Section 2.14. Finally, some new developments on the horizon are described in Section 2.15.

## 2.2 BRIEF HISTORY

The idea of using protons for radiation therapy is hardly new. This is, in fact, not only your father's radiation, but also your grandfather's. Physicists have been

producing beams of accelerated protons since at least the early 1930s, when Ernest Lawrence developed the cyclotron at the University of California–Berkeley for research in nuclear physics.[1] The physicist Robert Wilson made the first known suggestion that protons might have properties that are useful for radiation therapy in 1946 (Figure 2.2). He was one of the founders and the first director of Fermilab. When Wilson was called to appear before the Congressional Joint Committee on Atomic Energy, a senator asked him how a multi-million-dollar particle accelerator improved the security of the country. He answered, "It has nothing to do directly with defending our country, except to make it worth defending."

Patients were first treated with protons by C.A. Tobias and J.H. Lawrence (Ernest's physician brother) on the 184 in. cyclotron at Lawrence Berkeley Laboratory beginning in 1954 (Figure 2.3). They treated pituitary tumors in approximately 30 patients. In the 1950s, Larson and Leksell in Uppsala, Sweden, developed radiosurgical techniques for the treatment of brain tumors. They used a 180 MeV synchrocyclotron. They were the first to use range modulation and beam scanning to produce large treatment fields (in the lateral direction). They treated 73 patients. Lars Leksell later went on to develop the Gamma Knife. In 1961, Ray Kjellberg, a neurosurgeon at Massachusetts General Hospital (MGH), began treating intracranial targets at the Harvard Cyclotron Laboratory. Throughout the following decades, proton therapy proliferated internationally.

All of the early facilities used accelerators that were designed and built for physics research and were housed in laboratories. Many of these centers used the narrow native proton beam without any lateral beam spreading. They also used fixed beam (nonisocentric gantry) machines. These early facilities can be thought of as first-generation proton therapy centers. The first of the second-generation hospital-based

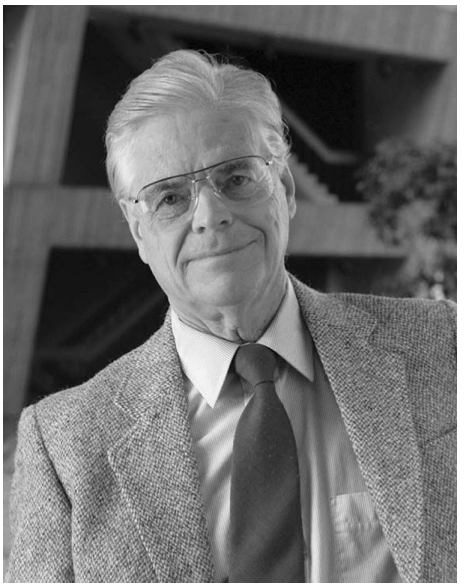


FIGURE 2.2 Robert R. Wilson (1914–2000) first suggested the use of protons for radiation therapy. (From Fermilab Neg. N. 89-0305-06. With permission.)

FIGURE 2.3 Ernest Lawrence at the controls of the 60 in. cyclotron. His brother John looks on. (Courtesy of Lawrence Berkeley National Laboratory.)

treatment centers is Loma Linda University Medical Center, which opened in 1990. The accelerator, a synchrotron, was built by Fermilab specifically for proton therapy. It has a large field size and there are isocentric gantries.

Clinical applications of proton therapy have included tumors at the base of the skull, head and neck tumors, localized prostate cancer, and inoperable early-stage lung cancer (Schultz-Ertner and Tsujii, 2007). Pediatric patients are favored for proton therapy because of the high dose conformity and low integral dose (see Section 2.8). Low-energy protons are used for treating eye tumors (uveal melanoma). Additional applications of proton therapy are for paraspinal tumors and other tumors in regions that are difficult to treat due to their proximity to critical normal organs. It has been estimated that 15% of all radiation therapy patients are candidates for proton therapy (McDonald and Fitzek, 2010).

Although the use of protons in radiation therapy has a long history, the technology for beam delivery and shaping is still evolving rapidly. One of the difficulties in describing proton therapy facilities is the large variety of approaches to beam acceleration and delivery. Each facility has been custom designed. This is changing now with the entry of commercial firms such as Mevion, IBA, Varian, ProTom, and Hitachi.

## 2.3 INTERACTION OF PROTONS WITH MATTER

In order to understand how protons can be used to treat patients and why proton therapy may be superior to other forms of radiation, it is first necessary to understand how they interact with matter. Charged particles interact very differently with matter than photons. High-energy photons used to treat patients interact perhaps a few

dozen times before completely expending their energy or exiting the patient. Photons lose their energy in large steps. They have no definite or specific range, and they are exponentially attenuated. Charged particles, on the other hand, lose energy in a very large number of small steps, to the point where the energy loss can almost be thought of as continuous. This is due to the long range of the Coulomb force; charged particles interact with large numbers of atoms simultaneously. Charged particles have a very definite range. Physicists classify charged particles for radiation therapy in two categories: electrons (and positrons) and everything else that we call heavy charged particles.[2] Protons fall into the latter category. The distinction between electrons and heavy particles is due to the much larger mass of heavy particles. The proton is the least massive of the heavy charged particles, and its mass is 2000 times larger than the electron mass. If we exclude nuclear reactions, which are relatively unimportant in the radiotherapy context, both electrons and protons interact with matter via Coulomb forces. They also have the same magnitude of charge. The difference, then, in the interactions of these particles *is due almost solely to the much larger mass of the proton*. Given the huge mass difference, it is not surprising that there are some significant differences in the nature of their behavior in matter.

On a microscopic scale, bombarding particles can interact with

1. Atomic electrons, either with the atom as a whole or with individual electrons
2. Atomic nucleus as a whole
3. Individual nucleons

The type of interaction depends on the classical impact parameter $b$ (Figure 2.4) in comparison to the radius of the atom $a$:

1. $b \gg a$: This is referred to as a soft collision. It is a collision with the atom as a whole, resulting in excitation or ionization and a small energy transfer of a few electron-volts. This is the most probable type of interaction, and because of the nature of the Coulomb interaction, these occur over a long range, and a passing charged particle can interact with many atoms simultaneously, leading to the continuous slowing-down approximation (CSDA) (see Chapter 4).

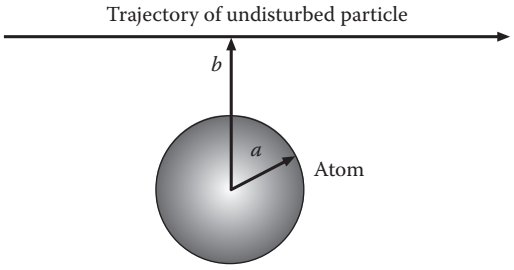



FIGURE 2.4 The size of the classical impact parameter $b$, in comparison to the radius of the atom, $a$, determines the nature of the interaction.

2. $b \approx a$: These are referred to as hard or "knock-on" collisions. In these collisions, the proton is more likely to interact with a single atomic electron transferring a considerable amount of energy, leading to ionization and the production of secondary electrons (delta rays). The range of secondary electrons is small, less than 1 mm.
3. $b \ll a$: This involves interactions with the atomic nucleus, possibly even with individual nucleons. Most of these interactions are just elastic scattering events. There is no bremsstrahlung emission for protons because they are too massive; it only happens for electrons. Also, protons do not scatter as easily as electrons. Multiple Coulomb scattering is less important than for electrons, but not negligible. Above about 100 MeV protons can interact with individual nucleons; charged particles may be ejected from the nucleus along with the emission of gamma rays. We now turn to a discussion of nuclear reactions.

Nuclear reactions must be considered for the following reasons. The attenuation of the beam by nuclear reactions, although small, is not insignificant. Nuclear reactions produce prompt neutron emission, which is of principal concern for radiation shielding. Nuclear reactions lead to activation of beam line components, beam modifiers, and the patient.

Nuclear reactions gradually reduce the fluence of protons in a proton beam. For 100 MeV protons in carbon, about 2.5% of the collisions are inelastic, resulting in energy loss. According to Smith (2009a), nuclear reactions in tissue, mainly on $^{16}O$, reduce the proton fluence by about 1% per centimeter. Therefore, at a depth of 20 cm, there is a loss of about 20% in the fluence. Nuclear reactions contribute to neutron production. The effects and outcome of proton-induced nuclear reactions depend on the energy of the bombarding protons, as described in the following paragraphs.

### 2.3.1 LOW ENERGY ($\lesssim$50 MeV)

In general, light elements emit gammas after proton bombardment via (p, γ) reactions. There are many sharp resonance peaks, as, for example, for $^{27}Al(p, \gamma)^{28}Si$. Also, (p, n) reactions are quite common. These may be thought of as substituting a proton for a neutron in the nucleus. One pair of neighboring isobars must be radioactive; therefore, a reaction involving a stable isotope must result in a radioactive target.[3] These reactions have a definite threshold and a high yield. The products of (p, n) reactions are generally positron emitters, as might be expected since positive charge has been added to the nucleus (see Section 2.15).

### 2.3.2 INTERMEDIATE ENERGY (50 < *T* < 100 MeV)

At intermediate energies, the proton may cause the emission of more than one particle, as 50 MeV exceeds the binding energy of a single nucleon (recall that the average binding energy is approximately 8 MeV per nucleon, except for very light elements).

The incoming particle may not be captured by the nucleus. The bombarding proton energy may be shared by many nucleons, causing them to "evaporate" from the nucleus.

### 2.3.3 HIGH ENERGY (>100 MeV)

High-energy protons have a de Broglie wavelength comparable to the range of the nuclear force. This energy is approaching the total binding energy of carbon, nitrogen, and oxygen nuclei that are common in tissue. At high energies, the products will all have mass numbers lower than that of the target. There are four types of reactions:

1. Spallation: A single nucleon or small group of nucleons are emitted. If the target mass is $A$, the residual nuclei have mass number 0.75–0.99 $A$.
2. Fission: A few individual nucleons are emitted and the original nucleus splits into roughly equal parts. Residual mass numbers are 0.30–0.65 $A$.
3. Fragmentation: Nucleon groups of $10 < A < 40$ are "blasted" from the target nucleus.
4. Secondary reactions: These are initiated by particles produced in the primary interaction, usually of type 1. Secondary reactions include all those possible with low-energy particles.

The rate of loss of kinetic energy by charged particles traversing matter is described by the stopping power, $(dT/dx)_{Y,T,Z}$, where $Y$ is the particle type and $Z$ is the atomic number of the medium. The collisional stopping power is the sum of the contributions from soft plus hard collisions. The radiative stopping power is negligible for protons.

The mass collision stopping power is defined as $S_C = dT/\rho dx$, and the units are MeV cm$^2$/g. The mass collision stopping power is the amount of energy lost in traversing a column of material 1 cm on a side with a mass of 1 g.

The total mass collision stopping power in units of MeV cm$^2$/g is given by (Attix, 1986):

$$S_C = 0.3071\frac{Zz^2}{A\beta^2}\left[13.8373 + \ln\left(\frac{\beta^2}{1-\beta^2}\right) - \beta^2 - \ln I\right] \text{MeV cm}^2/\text{g}, \tag{2.1}$$

where:

$\beta = v/c$, $v$ is the speed of the particle

$z$ is the charge of the bombarding particle

$Z$ is the atomic number of the medium

$A$ is the mass number of the medium

$I$ is the mean excitation potential of the medium (in eV, usually determined experimentally)

Note that $S_C$ is proportional to the electron density (electrons/g) of the medium. Shell corrections have not been included in Equation 2.1. Below some energy, K shell, and so on, electrons are no longer able to absorb energy, and therefore no longer participate in energy loss.

Figure 2.5 shows a plot of the proton mass stopping power as a function of kinetic energy for various media of radiological interest. High-energy protons have a relatively low stopping power. As these particles slow down, however, the stopping power rises dramatically. This is consistent with the velocity dependence given in Equation 2.1. In the range between 200 and 300 MeV, the stopping power for protons in water is about 4 MeV/cm. In comparison, the stopping power of therapeutic electrons is about 2 MeV/cm.

Important dependencies of the mass stopping power are

1. Medium: $S_c$ is directly proportional to the electron density (electrons/g), $n_e = N_A Z/A$, where $N_A$ is Avogadro's number.
2. Proportional to the square of the charge of the bombarding particle: heavy ions will have much higher stopping power; alpha particles have four times greater stopping power than protons for the same velocity.
3. Goes down slightly as $Z$ rises, because $Z/A$ declines slightly as $Z$ increases for stable nuclei. Also, $I$ goes up as $Z$ rises. The mass stopping power declines by about 20% in going from carbon to lead (Figure 2.5).

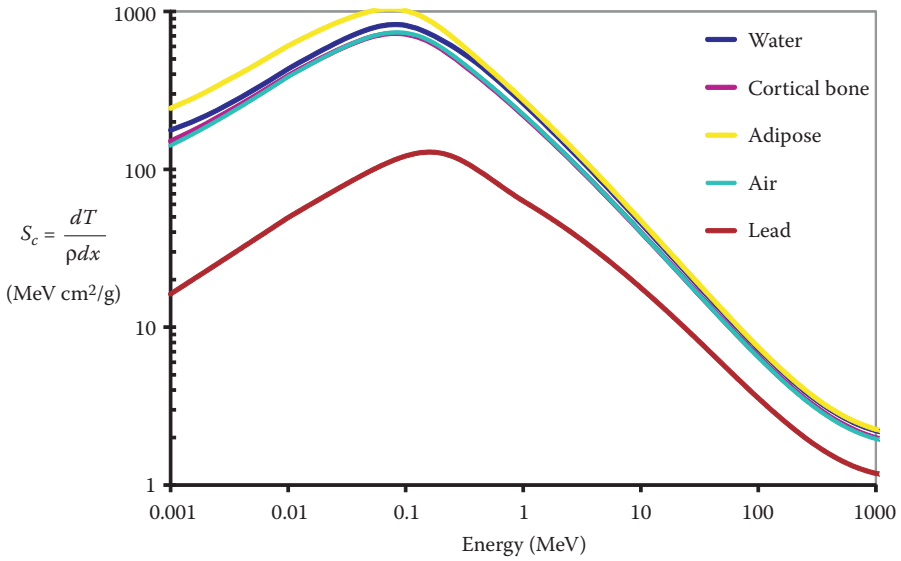



FIGURE 2.5 Mass stopping power for protons in water, cortical bone, adipose, air, and lead. Note that this is a log-log plot and that the stopping power is several orders of magnitude higher at low energy than at high energy. (Data from National Institute of Standards and Technology (NIST), Stopping-power and range tables for protons, Gaithersburg, MD: NIST, 2014, http://physics.nist.gov/PhysRefData/Star/Text/PSTAR.html.)

4. Independent of the mass of the bombarding particle: all heavy charged particles of the *same charge and velocity* have the same stopping power.
5. Rises dramatically as β goes to 0. This is due to the $1/v^2$ dependence.[4]

The *restricted stopping power* includes the effects of all soft collisions plus those hard collisions resulting in delta rays with energies less than a cutoff value Δ. The restricted stopping power is denoted as $(dT/\rho dx)_\Delta$. When the value of Δ approaches its maximum value, the restricted stopping power goes over to the mass collision stopping power because then all delta rays are included in the energy loss. The cutoff energy is usually chosen as 100 eV; beyond this energy, the delta rays can travel far enough to cause ionizations elsewhere.

The linear energy transfer (LET) is defined as

$$L_\Delta\left(\frac{\text{keV}}{\mu\text{m}}\right) = \frac{\rho}{10}\left[\left(\frac{dT}{\rho dx}\right)_\Delta \text{MeV cm}^2/\text{g}\right]. \tag{2.2}$$

The CSDA range is given by

$$\mathfrak{R}_{\text{CSDA}} = \int_0^{T_0}\left(\frac{dT}{\rho dx}\right)^{-1} dT, \tag{2.3}$$

where $T_0$ is the starting energy, and the units are g/cm$^2$. The CSDA range underestimates the actual range by 0.21% or less for protons.

The projected range is the mean of the maximum distance traveled in a direction perpendicular to the incident surface. The CSDA range is always greater than the projected range because of scattering.

Figure 2.6 shows the projected range of protons in a variety of materials. The deepest target in the human body is on the order of 30 cm. To treat a target at this depth (in water), the incident protons must have energies between about 200 and 250 MeV. Beams of less than 100 MeV have small penetration, but they have found application in treating ocular conditions such as uveal melanoma. A power law fit to the data in Figure 2.6 gives the projected proton range in water for energies between 10 and 300 MeV,

$$\mathcal{R}_p\,(\text{cm}) \approx 2.1\times10^{-3}\left[T\,(\text{MeV})\right]^{1.8}, \tag{2.4}$$

to an accuracy of better than 20%.

Figure 2.7 shows the fluence as a function of depth for various types of radiation.

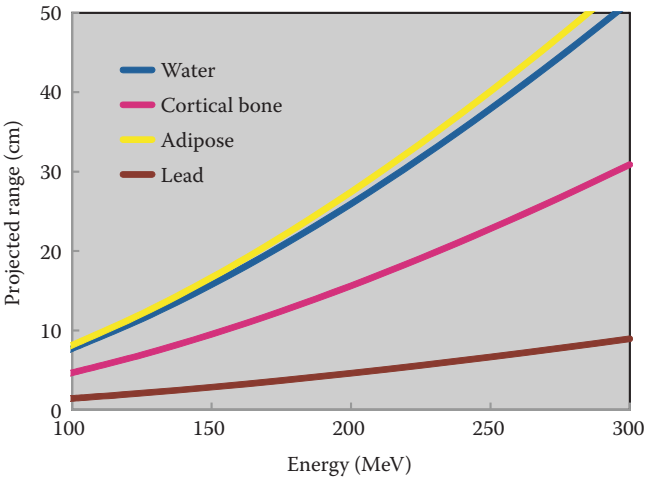



FIGURE 2.6 The projected proton range in a variety of materials. Protons with energy between 200 and 250 MeV have a projected range of approximately 30 cm in water. (Data from National Institute of Standards and Technology (NIST), Stopping-power and range tables for protons, Gaithersburg, MD: NIST, 2014, http://physics.nist.gov/PhysRefData/Star/Text/PSTAR.html.)

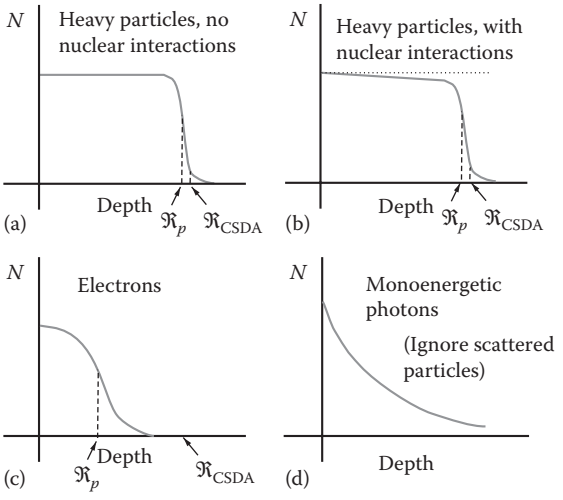



FIGURE 2.7 The fluence (or number) of particles as a function of depth for various types of radiation. The practical and CSDA ranges are also indicated. (a) Heavy charged particles with no nuclear reactions. (b) Heavy charged particles with nuclear reactions. The nuclear reactions remove some of the particles before they can reach the end of their normal range. (c) Electrons: these are easily scattered, and therefore they do not show the well-defined range seen for heavy charged particles. (d) Photons: these have no definite range and are exponentially attenuated.

## 2.4 ABSORBED DOSE AND THE BRAGG PEAK

For monoenergetic charged particles (with delta ray equilibrium), the dose is given by (Attix, 1986)

$$D(x) = \Phi(x)\left(\frac{dT}{\rho dx}\right), \tag{2.5}$$

where $\Phi(x)$ is the particle fluence at the depth $x$ and the mass collision stopping power is to be evaluated for the energy of the particles at depth. If the fluence remained constant out to the end of the track, then $D \propto (dT/\rho dx)$, and one would expect the dose to rise dramatically near the end of the particle track. If the mass collision stopping power is expressed in units of MeV cm$^2$/g and the fluence is in units of cm$^{-2}$, then (Attix, 1986)

$$D(x) = 1.6\times10^{-10}\Phi(x)\left(\frac{dT}{\rho dx}\right)\text{Gy}. \tag{2.6}$$

For a spectrum of particles (as one would realistically expect at depth from any real source), we introduce the differential fluence spectrum, defined as follows: $d\Phi = \Phi_T dT$ is the fluence of particles having kinetic energy between $T$ and $T + dT$. At any depth $x$, the dose is then given by

$$D(x) = \int_0^{T_{\max}} \Phi_T\left(\frac{dT}{\rho dx}\right)dT, \tag{2.7}$$

where $\Phi_T$ is the fluence spectrum at depth. The stopping power is to be evaluated at depth and $T_{\max}$ is the maximum energy at depth $x$.

Unfortunately, it is not an easy matter to calculate $\Phi_T$. To a first approximation, the fluence $\Phi$ is a constant with depth until it abruptly drops to zero at the end of the range of the particles (Figure 2.7).

Figure 2.8 shows a depth–dose curve for a monoenergetic 200 MeV proton beam in water. The dose is fairly constant with increasing depth until near the depth corresponding to the practical range. Here the dose rises sharply as a result of the increase in the stopping power. The dose reaches a peak and then turns over and declines sharply because of the rapidly decreasing fluence beyond the projected range (see Figure 2.7). The peak of the curve is called the Bragg peak. Beyond the Bragg peak, the dose drops to a negligible value. *There is no exit dose for proton beams.* This is almost like a free lunch—relatively little dose proximal to the Bragg peak and almost no dose beyond (called distal blocking). This presents obvious

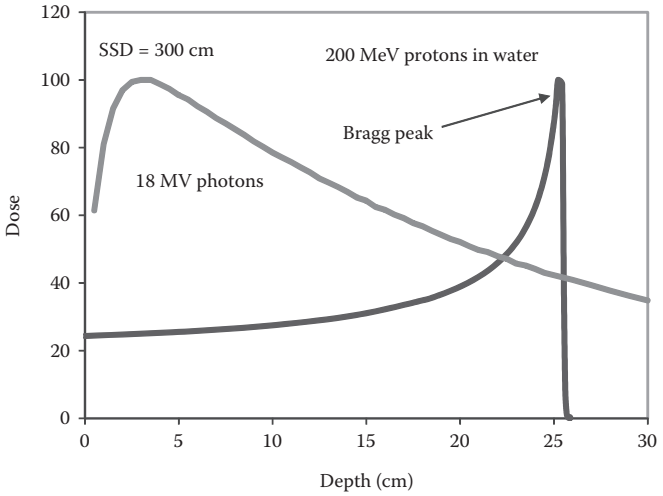



FIGURE 2.8 Depth–dose curve for 200 MeV protons in water (SSD = 300 cm). Also shown for comparison is a depth–dose curve for 18 MV photons (SSD = 100 cm). In the entrance region, the proton dose is relatively low and more or less constant with depth. Near the end of the proton range, the stopping power rises rapidly, and the dose reflects this dramatic rise. At the same time, the fluence is beginning to decline. The result is a peak in the dose and then a precipitous decline. This is called the Bragg peak. The dose is essentially zero for depths greater than about 26 cm.

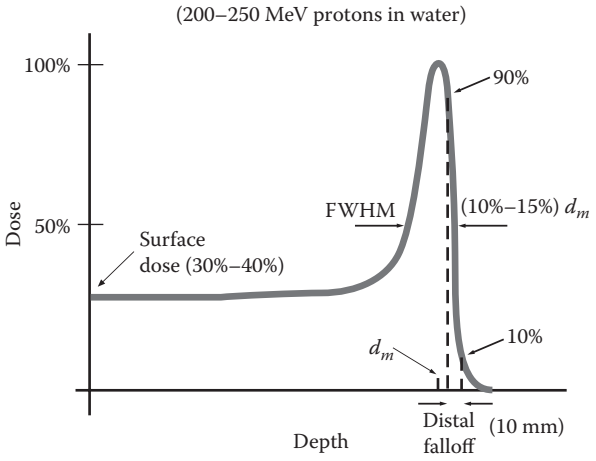



FIGURE 2.9 Properties of the Bragg peak for clinical proton beams ranging in energy between 200 and 250 MeV. (Data from Central axis depth dose data for use in radiotherapy, 1996, *Br. J. Radiol.*, Suppl. 25, 1996.)

advantages for radiation therapy. It also presents perils; a slight mispositioning of the patient or uncertainty in the range can result in a large error in the dose distal to the Bragg peak.

Figure 2.9 shows some properties of the Bragg peak for protons with energies between 200 and 250 MeV in water. The depth–dose curve is normalized to 100%

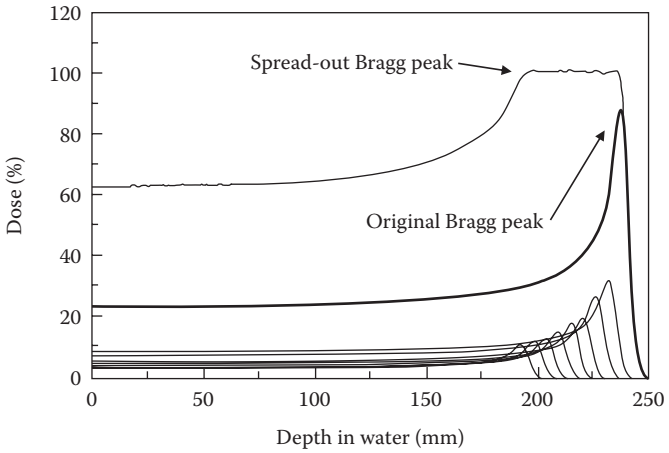



FIGURE 2.10 The Bragg peak is spread out by adding a number of beams with slightly different energies. The dose profiles of the contributing beams are shown at the bottom of the figure. The sum of all these gives the SOBP. The proximal dose plateau rises. (Courtesy of Dr. D. Jones, iThemba LABS, Faure, South Africa.)

at the location of maximum dose ($d_m$). The surface dose ranges from 30% to 40% of the peak value. The width of the Bragg peak is defined in terms of the full width at half maximum (FWHM). This ranges between 10% and 15% of the value of $d_m$ (a few centimeters). The falloff beyond the Bragg peak is measured in terms of the distance from 90% to 10%, and this is typically 10 mm.

The Bragg peak is too localized or narrow to treat most targets. Figure 2.9 shows that the width of the peak is on the order of a few centimeters. Most target volumes are larger than this. It is usually necessary to spread out the Bragg peak. This can be accomplished by adding a number of beams with varying energy, as shown in Figure 2.10. One way of doing this is to introduce an oscillating wedge into the beam or a range modulation wheel like a propeller (Figure 2.11). This produces a so-called spread-out Bragg peak (SOBP). There is a price to pay for this in that the proximal dose plateau rises.

The lateral distribution of dose is also affected by the large mass of the proton. The penumbra is fairly sharp because protons do not scatter easily. This contributes to the ability to treat targets in proximity to critical structures. The penumbra width depends on many factors. The lateral penumbra is dominated by multiple Coulomb scattering. This leads to a broadening of the penumbra with increasing depth. The total penumbra is about 6–7 mm (distance from 80% to 20%) at 15 cm depth (Smith, 2009a). The proton penumbra is less than the penumbra for a high-energy (15 MV) photon beam up to a depth of about 17 cm, and then it becomes larger (Smith, 2009a).

Small fields degrade the Bragg peak because of a lack of lateral charged particle equilibrium (Smith, 2009a). Large air gaps will also degrade the penumbra. For this reason, it is important that the beam-shaping aperture is positioned close to the patient.

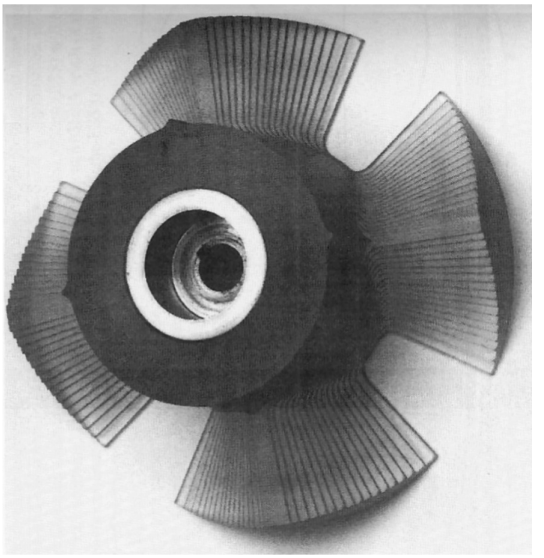


FIGURE 2.11 A propeller that can be inserted into the beam line to spread out the Bragg peak. The thickness of the propeller that is intercepted by the beam varies as the propeller rotates. Some portions of the propeller have zero thickness. This is for the Loma Linda eye beam line. It produces a uniform dose of more than 22 mm in depth. (Reprinted from Moyers, M.F., in *The Modern Technology of Radiation Oncology*, vol. I, ed. J. VanDyk, 823–69, Madison, WI: Medical Physics Publishing, 1999.)

## 2.5 A FEW WORDS ABOUT RADIOBIOLOGY

The attraction of protons for radiotherapy is due to their dose distribution, not radiobiology (as for, say, neutrons). The radiobiological properties of protons are "unremarkable" (Hall and Giaccia, 2012). The relative biological effectiveness (RBE) of protons is the same as that for 250 kV x-rays, 10%–15% more effective than Co-60 radiation. The oxygen enhancement ratio (OER) is the same as that for x-rays: 2.5–3.0. The biological properties of protons are consistent with their physical properties. Protons are sparsely ionizing, except at the end of their range. For 250 MeV protons, in the entrance plateau the LET is about 8 keV/μm for depths between 2.5 and 27 cm, although the LET does rise sharply at the end of the range. The LET in the SOBP is a mixture of the LET in many low LET plateau regions plus one (or a few) high LET Bragg peaks. There is a small increase in RBE in comparison to photons. Table 2.1 from Ternier (undated) shows RBE values for a variety of biological endpoints. It is remarkable how these RBE values cluster around a value of 1.1 regardless of biological endpoint or fractionation scheme. The International Commission on Radiation Units and Measurements (ICRU) Report No. 78 (2007) recommends a generic RBE of 1.1 when comparing photon and proton therapy doses. It is common to see reference to cobalt gray equivalent (CGE). This is the dose in gray multiplied by 1.1.

**TABLE 2.1 RBE Values of Modulated Proton Beams at the Bragg Peak Compared to $^{60}C$**

| Tissue | Endpoint | Proton Energy (MeV) | No. Fractions | RBE |
|---|---|---|---|---|
| Crypt cell[a] | Survival | 160 | 1 | 1.19 |
| Crypt cell[a] | Survival | 160 | 20 | 1.23 |
| Skin[a] | Acute reaction | 160 | 20 | 1.13 |
| Fibrosarcoma[b] | Survival | 160 | 1–10 | 1.16 |
| Mammary cancer [c] | $TCD_{50/120}$ | 160 | 1 | 1.11 |
| Lens[c] | Cataract | 160 | 1 | 1.09 |
| Lungs[c] | $LD_{50/100}$ | 160 | 1 | 1.02 |
| Testis[c] | Weight loss | 160 | 1 | 1.23 |
| Tail vertebrae[c] | Growth | 160 | 1 | 1.32 |
| Mouse[d] | $LD_{50/30}$ | 250 | 1 | 1.09 |
| Skin[d] | Contraction | 250 | 10 | 1.03 |

[a] Tepper, J., et al., *Int. J. Radiat. Oncol. Biol. Phys.* 2(11), 1115–1122, 1977.
[b] Urano, M., et al., *Int. J. Radiat. Oncol. Biol. Phys.* 6(9), s1187–1193, 1980.
[c] Urano, M., et al., *Int. J. Radiat. Oncol. Biol. Phys.* 10(4), 509–514, 1984.
[d] Tatsuzaki, H., et al., An RBE study of a proton beam at University of Tsukuba (vs. Co-60). *Proceedings of the XXI PTCOG Meeting*, Chiba, Japan 14–16 Nov. 1994, pp. 146–148, 1995.
TCD, tumor control dose; LD, lethal dose.

## 2.6 CIRCULAR CHARGED PARTICLE ORBITS AND STABILITY

We have already seen that the proton energy needed for therapeutic applications is on the order of 250 MeV. This kinetic energy is about 25% of the rest mass energy of 940 MeV. Although such protons are not highly relativistic, accurate calculations cannot neglect relativity either. Let us consider the motion of a charged ion of mass $m$ in a uniform magnetic field of magnitude $B_0$. Assume that the particle undergoes circular motion in a plane perpendicular to the direction of the magnetic field.

The relativistic equation for charged particle motion in a magnetic field is

$$\frac{d\vec{p}}{dt} = q\vec{v} \times \vec{B}, \tag{2.8}$$

where $q$ is the charge, $\vec{p} = m\vec{v}$, and $m = m_0 / \sqrt{1-\beta^2} = m_0\gamma$. We assume no change in the energy of the particle; therefore, $\beta$ is a constant and

$$m_0\gamma \frac{d\vec{v}}{dt} = q\vec{v} \times \vec{B}. \tag{2.9}$$

This equation is the same as the nonrelativistic equation of motion, except for the $\gamma$ term, which multiplies $m_0$. The angular frequency of revolution is

$$\omega_0 = \frac{qB_0}{\gamma m_0} = \frac{qc^2 B_0}{E}, \tag{2.10}$$

where $E = T + m_0c^2$ is the total relativistic energy.

This angular frequency is known as the cyclotron or gyro frequency. For a nonrelativistic particle ($\gamma \approx 1$), the frequency $\nu_0$ corresponding to this is

$$\nu_0\,[\mathrm{MHz}] = 15\left(\frac{Z}{A}\right)B_0\,[T], \tag{2.11}$$

where:

- $\nu_0$ is in units of MHz
- $Z$ is the charge on the ion
- $A$ is approximately the atomic mass number ($A = \mathrm{m}/m_p$, where $m_p$ is the mass of the proton)

and the magnetic field is in tesla.

For $B_0 = 1$ T, $\nu_0 = 15$ MHz. This frequency is in the radiofrequency (RF) range of the electromagnetic spectrum.

The kinetic energy of nonrelativistic ions is given by

$$T\,[\mathrm{MeV}] = 48\left(\frac{Z^2}{A}\right)R^2\,[\mathrm{m}]B_0^2\,[T], \tag{2.12}$$

where the kinetic energy $T$ is given in units of MeV, the radius $R$ is in units of meters, and the magnetic field strength is in units of tesla. As an example, to produce 30 MeV deuterons, $B_0 = 1$ T is needed over a 1.25 m radius (Humphries, 2012).

The radius of circular motion of a nonrelativistic particle of energy $T$ is

$$R = \frac{\sqrt{2m_0T}}{qB_0}. \tag{2.13}$$

For a kinetic energy of 250 MeV and a magnetic field strength of 1 T, the radius is about 5 m. The mass of a proton is three orders of magnitude greater than the mass of an electron. Furthermore, the therapeutic energy of protons is about an order of magnitude greater than for electrons. Therefore, the radius of curvature of the trajectory, in a fixed magnetic field, is about two orders of magnitude larger for therapeutic protons than for electrons. This is one of the reasons that proton facilities are so expensive. Gantries need to be huge (see Section 2.8).

A topic that is not normally addressed in elementary discussions of particle accelerators is the issue of orbital stability. Particle stability is crucial for the successful acceleration of charged particles. Some particles may have a small component of velocity in either the radial or vertical direction, or both. Any nondamped motion in the vertical direction may remove particles from the useful beam. This is described by the *emittance* of the beam. The emittance is a measure of the spread of the particles in both position and momentum (Wikipedia, 2014b). A beam with low emittance is bunched tightly, and all particles have nearly the same momentum. In addition to the emittance, there are inevitably small perturbations in the applied magnetic field. The "unrolled" trajectory of a proton in a cyclotron is about 4 km in length (Jongen, 2008). It is necessary for the protons (from a cyclotron) to reach a target of about 1 mm in size at the energy degrader.

For successful acceleration, particle orbits have to be stable in both the vertical and radial directions (see Figure 2.13). We will first state the conditions necessary for stabilty and then derive them. The derivation is not difficult. The conditions for stability are related to the radial gradient of the $z$ component of the magnetic field. The stability index $n$ is defined as

$$n(r_0) \equiv -\left(\frac{r}{B_z}\frac{\partial B_z}{\partial r}\right)\Bigg|_{r=r_0,\, z=0}, \tag{2.14}$$

where:

$r_0$ is the equilibrium radius of the particles
$z = 0$ is the median orbital plane of the motion

For vertical stability, $n > 0$. For radial stability, the condition is $n < 1$. The condition for both vertical and radial stability is therefore $0 < n < 1$. For stability, the vertical component of the magnetic field must decrease with increasing radial distance from the center of the circular orbit.

In cylindrical polar coordinates $r$, $\theta$, and $z$, the velocity is $\vec{v} = \dot{r}\hat{e}_r + r\dot{\theta}\hat{e}_\theta + \dot{z}\hat{e}_z$, and the acceleration is $\vec{a} = (\ddot{r} - r\dot{\theta}^2)\hat{e}_r + (2\dot{r}\dot{\theta} + r\ddot{\theta})\hat{e}_\theta + \ddot{z}\hat{e}_z$ (see Thornton and Marion, 2003, or any mechanics textbook), where $\hat{e}_r$, $\hat{e}_\theta$, and $\hat{e}_z$ are unit vectors in the $r$, $\theta$, and $z$ directions, respectively. The components of the equation of motion are (see Equation 2.8)

$$\frac{m}{q}\left(\ddot{r} - r\dot{\theta}^2\right) = r\dot{\theta}B_z - \dot{z}B_\theta \tag{2.15}$$

$$\frac{m}{q}\left(2\dot{r}\dot{\theta} + r\ddot{\theta}\right) = \dot{z}B_r - \dot{r}B_z \tag{2.16}$$

$$\frac{m}{q}\ddot{z} = \dot{r}B_\theta - r\dot{\theta}B_r. \tag{2.17}$$

In the special case in which $\dot{r} = \dot{z} = 0, B_r = B_\theta = 0,$ and $B_z = B_0$, $\left|\dot{\theta}\right| = \frac{qB_0}{m}$, as in Equation 2.10.

The magnetic field must obey Maxwell's equations:

$$\vec{\nabla} \cdot \vec{B} = 0 \tag{2.18}$$

and

$$\vec{\nabla} \times \vec{B} = 0. \tag{2.19}$$

We assume the field is azimuthally symmetric and therefore

$$B_\theta = 0 \text{ and } \frac{\partial B_z}{\partial \theta} = \frac{\partial B_r}{\partial \theta} = 0. \tag{2.20}$$

Under these circumstances,

$$\vec{\nabla} \cdot \vec{B} = \frac{1}{r}\frac{\partial}{\partial r}(rB_r) + \frac{\partial B_z}{\partial z} = 0, \tag{2.21}$$

$$\nabla \times \vec{B} = \left(\frac{\partial B_r}{\partial z} - \frac{\partial B_z}{\partial r}\right)\hat{e}_\theta = 0. \tag{2.22}$$

Expand $B_z$ $(r, z)$ and $B_r$ $(r, z)$ around the equilibrium circular orbit $r = r_0$ and $z = 0$,

$$B_z(r,z) = B_0 + \frac{\partial B_z}{\partial r}(r - r_0) + \frac{\partial B_z}{\partial z}z + \cdots \tag{2.23}$$

$$B_r(r,z) = \frac{\partial B_r}{\partial r}(r - r_0) + \frac{\partial B_r}{\partial z}z + \cdots \tag{2.24}$$

to leading order:

$$B_z = B_0 + b_1(r - r_0) - \frac{b_1}{2r_0}z^2 \tag{2.25}$$

and

$$B_r = b_1 z, \tag{2.26}$$

where

$$b_1 = \frac{\partial B_z}{\partial r} = \frac{\partial B_r}{\partial z} \text{ and } n = -\frac{r_0\, b_1}{B_0}. \tag{2.27}$$

Now perturb the equations of motion by introducing the terms

$$r = r' + r_0 : \dot{r} = \dot{r}' \tag{2.28}$$

$$\theta = \theta' + \theta_0 : \dot{\theta} = \dot{\theta}' - \omega_0 \tag{2.29}$$

$$z = z' + z_0 = z', \tag{2.30}$$

where $|r'/r_0| \ll 1$ and $|\theta'/\omega_0| \ll 1$. Substituting Equations 2.28 through 2.30 into Equations 2.15 through 2.17 and retaining only leading order terms yields (see Problem 6)

$$\ddot{r}' + \omega_0 r_0 \dot{\theta}' + n\omega_0^2 r' = 0 \tag{2.31}$$

$$r_0 \ddot{\theta}' - \dot{r}'\omega_0 = 0 \tag{2.32}$$

$$\ddot{z}' + n\omega_0^2 z' = 0. \tag{2.33}$$

For the $z'$ in Equation 2.33, if $n < 0$, the solution grows exponentially and the particles will be unstable to small perturbations. If, on the other hand, $n > 0$, Equation 2.33 describes harmonic motion in the $z$ direction (the vertical stability condition). The oscillation angular frequency about the midplane is $\Omega_z = \omega_0\sqrt{n}$. These vertical oscillations are called *betatron oscillations*.

Now, let us look at radial stability. Differentiate Equation 2.31 and use Equation 2.32 to eliminate $\ddot{\theta}'$:

$$\dddot{r}' + \omega_0^2 (1 - n)\dot{r}' = 0. \tag{2.34}$$

Let $\dot{r}' = u$:

$$\ddot{u} + \omega_0^2 (1 - n) u = 0. \tag{2.35}$$

The solution of this equation will be oscillatory provided that $n < 1$. The frequency of radial oscillations is $\Omega_r = \omega_0\sqrt{1 - n}$.

For stability in both the vertical and radial directions,

$$0 < n < 1. \tag{2.36}$$

## 2.7 PROTON THERAPY ACCELERATORS

Two of the major requirements for an ion therapy accelerator are sufficient energy to provide the necessary range in tissue and adequate beam current to provide a dose rate that will allow a therapeutic dose to be delivered in a reasonable amount of time.[5] We have already seen that we need protons with energies up to 250 MeV to reach depths of 30 cm in water. Regarding beam current, we need a sufficient number of particles per minute to have a reasonable dose rate. An electron linear accelerator can easily reach dose rates of 300 cGy/min. Let us estimate the beam current necessary to match this. In the proton Bragg peak, the energy is on the order of 25 MeV. The corresponding stopping power is on the order of 20 MeV $cm^2/g$ (see Figure 2.5). If we assume a field size of 20 cm × 20 cm, this requires an average beam current delivered to the patient of about 10 nA, based on Equation 2.5. This current is easily achieved, as we will see.

An additional design goal is to make the accelerator as small as reasonably possible so that it may be installed in a hospital setting. It should be easy to operate—avoiding the need for an engineer operator to provide constant beam tuning.

Medical linear accelerators can produce accelerating fields of 20 MeV/m (see Chapter 1) for electron acceleration. Assuming that the same gradients could be achieved for proton acceleration, production of 250 MeV protons would require a waveguide of 12–13 m in length. In an electron linac, the electrons are highly relativistic after acceleration through the first few cavities (see Chapter 1). Therapy protons are neither nonrelativistic nor highly relativistic, which would lead to a complicated variable cavity design for a linac. It is because of these difficulties that all proton therapy accelerators are circular machines.

There are two main types of accelerators that are used to accelerate protons for radiation therapy:

1. Cyclotrons
   a. Isochronous
   b. Synchrocyclotron
2. Synchrotrons

Features common to all accelerators include:

1. Ion source: Produce protons for acceleration (use hydrogen gas)
2. Injection: Get protons into the accelerator

3. Acceleration using electric fields
4. Extraction: Get protons out of accelerator
5. Beam transport or switchyard: Direct beam to various treatment rooms

Characteristics in common for all circular accelerators are that they have a vertical magnetic field to deflect the particles. The beam orbits are often not truly circular, in contrast to the assumption made in Section 2.6. Some synchrotrons have straight and circular sections. Another common feature is a gap between cavities where the particles are accelerated. Particle recirculation is the key to the operation of circular machines. The particles pass multiple times ($10^2$ to $>10^8$) through the accelerating gap (Humphries, 2012).

Accelerators can have superconducting or nonsuperconducting magnets. Superconducting magnets allow much higher magnet currents and therefore much higher magnetic fields, leading to more compact designs. This permits smaller and lighter magnets, allowing a cyclotron to be gantry mounted. The magnetic field strengths used in superconducting cyclotrons exceed the saturation field of the iron poles. As all magnetic dipoles are aligned under such circumstances, the field can be predicted more accurately (Humphrey, 2012). Another advantage of superconducting magnets is low power consumption. A disadvantage is the need for a cryostat and cooling with liquid helium.

Cyclotrons accelerate particles to a constant fixed native beam energy. These machines are therefore designed to produce the maximum necessary therapeutic energy potentially needed for the deepest anatomical site. A reduction in energy to treat more superficial targets requires an energy selection system. A block of material inserted into the beam serves this purpose. A synchrotron, in contrast, can produce an inherently variable energy beam.

## 2.7.1 CYCLOTRONS

### 2.7.1.1 UNIFORM FIELD (CLASSICAL) CYCLOTRON

The cyclotron was invented by Ernest Lawrence in the early 1930s for research in nuclear physics. Lawrence was awarded the Nobel Prize in Physics in 1939 for the development of the cyclotron and for the research carried out with it. The cyclotron exploits a clever idea: instead of accelerating charged particles once through a large potential difference, accelerate them many times through a small potential difference. Lawrence's original cyclotron used two hollow electrodes called *dees* because they resemble the shape of the letter D (Figure 2.12).

The particles are injected into the gap between the dees near the center. An oscillatory potential difference across the gap accelerates the particles. The acceleration results in a gain in energy. After crossing the gap, the ions pass into the interior of the dees, where they are shielded from the electric field. If the polarity of the dees oscillates at the cyclotron frequency (see Equation 2.10), and if the particles remain nonrelativistic, the particles will always be accelerated when they find themselves between the dees.

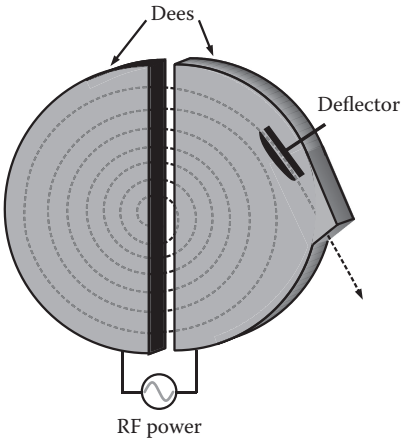



FIGURE 2.12 An overhead view of a classical cyclotron. The cyclotron has two electrodes called dees (because of their resemblance to the letter *D*). These electrodes are hollow. An oscillatory potential difference is applied between the dees. The particles are accelerated every time they pass through the gap between the dees. The applied magnetic field is perpendicular to the plane of the dees. The particles spiral outward in the magnetic field as they gain energy. The path followed by the particles has been superimposed on the diagram. The frequency of revolution of the particles remains constant with radius in the nonrelativistic limit. An electric field between the deflector plates deflects the beam so that it may be extracted from the cyclotron.

The magnetic field forces the particles into circular orbits that increase in radius as the particles are accelerated. The magnetic field between the magnet pole pieces is temporally constant and (almost) spatially uniform. The field is uniform in azimuth and almost constant in the radial direction. The protons are extracted from the cyclotron by using an electrostatic field applied by a deflector (Figure 2.12). A side view of the magnet is shown in Figure 2.13.

The crucial feature of the cyclotron is that the period (for nonrelativistic particles) is independent of the kinetic energy of the protons. In this circumstance, it is possible to have a constant frequency (RF) accelerating potential tuned to the cyclotron frequency of the protons for a given magnetic field strength. In this way, the protons can be repeatedly accelerated across the gap as they circulate. A typical potential difference between the dees is $10^5$ V. Each time a proton crosses the gap, it will acquire an additional $10^5$ eV. A proton crosses the gap twice each time it makes a round-trip. If the proton goes around 100 times, it will gain an amount of kinetic energy equal to $2 \times 100 \times 10^5$ eV = 20 MeV.

The ion source supplies ions continuously. Ions are captured from the source during roughly half the phase of the applied RF voltage. This results in the production of a continuous train of beam micropulses. An aperture located at the entrance to the accelerating gap (not shown in Figure 2.12) restricts the ions to a small range of phase space, and this limits the energy spread of the beam.

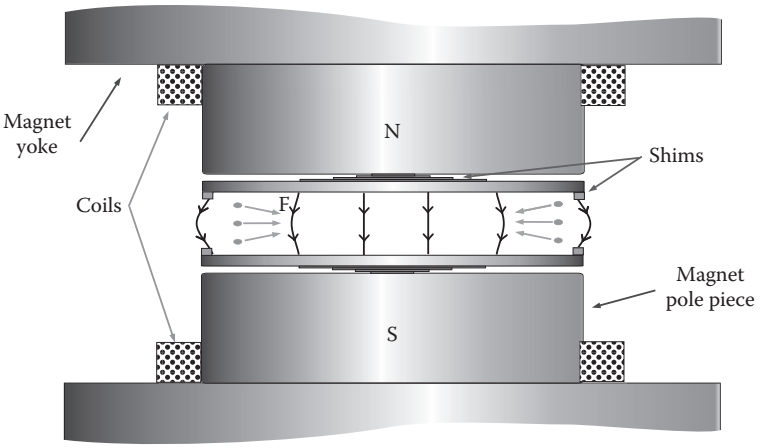



FIGURE 2.13 Side-view schematic diagram of a classical cyclotron showing the magnet yoke, the pole pieces, and the coils. The dees are not shown in this view. Shims may be used for fine adjustments of the magnetic field. The magnetic field lines bow out toward the edges. The vertical component of the magnetic field decreases slightly in the radial direction. This provides a stabilizing restoring force for particles that drift above or below the midplane, as shown. The particles move into the page on the right.

The maximum energy of the beam is limited by two factors: (1) relativity and (2) radial variations of the $\vec{B}$ field. The protons used for radiation therapy have kinetic energies up to 250 MeV. The rest mass energy of a proton is 940 MeV. The kinetic energy of therapy protons is a significant fraction of the rest mass. Under these circumstances, Newtonian mechanics begins to break down, and we must consider the effects of special relativity. In special relativity, an object's mass increases as its speed becomes a significant fraction of the speed of light. When this occurs, the cyclotron frequency is no longer constant with particle energy. This ruins the very basis for the classical cyclotron. This is one reason that cyclotrons are not useful for accelerating electrons. The requirement for constancy of the gyrofrequency limits the beam energy to about 15–20 MeV (Humphries, 2012).

The potential difference between the dees is given by

$$V = V_0 \sin \omega_{rf} t, \tag{2.37}$$

where:

$\omega_{rf}$ is the angular frequency of the applied voltage
$V_0$ is its amplitude

The phase of a particle at azimuthal position θ relative to the applied RF voltage is given by

$$\varphi = \omega_{rf} t - \theta(t). \tag{2.38}$$

The phase is defined in terms of the crossing time between the dees relative to the RF waveform. When φ = 90°, crossing occurs at maximum voltage and the particle gains the maximum possible kinetic energy. When φ = 0° or 180°, the accelerating potential is zero.

Suppose that the RF is set to the nonrelativistic gyrofrequency. Under these circumstances,

$$\dot{\varphi} = \frac{qB_0}{m_0}\left(1 - \frac{1}{\gamma}\right) > 0. \tag{2.39}$$

If the initial value of φ = 90, the value of φ will increase until φ = 180. When this occurs, the ions arrive at the gap between the dees when the accelerating voltage is zero. The ions will then be trapped at a particular energy and circulate at constant radius. A phase analysis of azimuthal particle motion (Humphries, 2012) shows that the maximum kinetic energy will be

$$T_{\max} = \left[\frac{16qV_0}{\pi m_0 c^2}\right]^{1/2} m_0 c^2, \tag{2.40}$$

where $V_0$ is the accelerating potential between the dees. The value of $V_0$ is typically 100 kV. For a proton, $T_{\max} \approx 20$ MeV.

There are two methods that are used to circumvent the problem presented by relativistic behavior. The first of these is to vary the frequency of the RF accelerating voltage in such a way as to compensate for the increase in proton mass. Cyclotrons that employ this strategy are called synchrocyclotrons. A disadvantage of this design is that the accelerator cannot generate a continuous beam of micropulses, but instead produces a cycled beam because the particles must be accelerated in groups.

A second method of overcoming relativistic behavior is to make the magnetic field increase at larger distances from the center of the cyclotron. If the gradient in the magnetic field is just right, the protons will continue to circulate with the same period as they gain energy and move out toward the perimeter of the cyclotron. Such a cyclotron design is referred to as isochronous (meaning equal time) or azimuthally varying field (AVF). An advantage of the isochronous cyclotron is that the RF can remain constant.

#### 2.7.1.2 AVF OR ISOCHRONOUS CYCLOTRON

Most proton therapy accelerators are isochronous cyclotrons. In an isochronous or AVF cyclotron, the orbital period is the same for all particles regardless of radius or energy, and the RF power operates at a single frequency. The magnetic field increases radially outward to maintain a constant period of orbital motion. As we have seen in Section 2.6, however, a positive field gradient leads to instability (associated with a negative index $n$; see Equations 2.14 and 2.36). The solution to the instabilty problem is to make the magnetic field nonaxisymmetric. This is accomplished by adding wedge-shaped

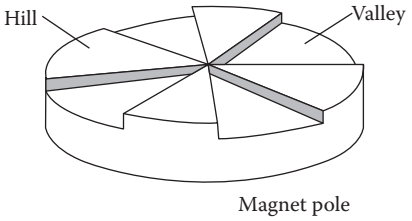



FIGURE 2.14 A magnet pole for an AVF cyclotron. Wedges have been added to the pole face to create hills and valleys. The magnetic field strength is higher in the hills than in the valleys. The azimuthal variations in the magnetic field strength have a focusing effect on the beam. This compensates for the instability introduced by the presence of a positive field gradient (negative index of stability).

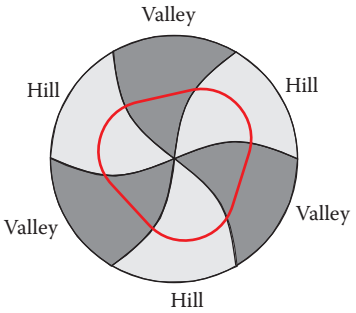



FIGURE 2.15 A cyclotron with segmented spiral-shaped pole pieces with hills and valleys. The magnetic field strength is larger in the hills than in the valleys. This provides the enhanced focusing necessary for a negative index isochronous cyclotron. The particle orbit is shown in red.

inserts on the magnetic pole pieces (Figure 2.14) with "hills" and "valleys" in the field strength. The extra horizontal component of $\vec{B}$ enhances vertical focusing. Under these circumstances, the particles can have stable orbits even in the presence of a negative field index. Modifying the wedge-shaped inserts to make them spiral increases the focusing. The orbits will no longer be circular (Figures 2.15 and 2.16).

Commercially available isochronous cyclotrons are provided by IBA (Ion Beam Applications) and Varian. More than half of proton therapy centers worldwide are IBA systems (http://www.iba-worldwide.com/?page=about).

We now describe some parameters of IBA cyclotrons (Figure 2.17). This discription is taken from Moyers (1999) and Jongen (2010). The Proteus 235 is a nonsuperconducting isochronous cyclotron with a maximum extraction energy of 235 MeV. The magnet is 4.3 m in diameter and weighs 220 tons. There are four magnetic field sectors (see Figure 2.16). The magnetic field in a hill is 2.9 T, and it is 0.9 T in a valley. There are two dees positioned in opposite magnetic field valleys. The RF is 106 MHz (Figure 2.18), which is the fourth harmonic of the proton rotation frequency. The voltage across the dees is variable: 60 kV in the center region and 130 kV near extraction. The maximum extracted beam current is 300 nA. The variable energy degrader is

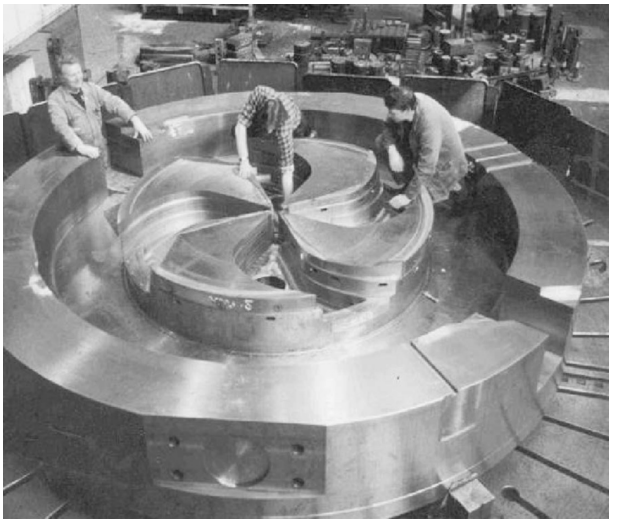


FIGURE 2.16 Magnet yoke for an isochronous cyclotron showing spiral pole pieces. The spiral pole pieces provide enhanced focusing. (Courtesy of IBA, Louvain-la-Neuve, Belgium, 2014, www.iba-worldwide.com. With permission.)

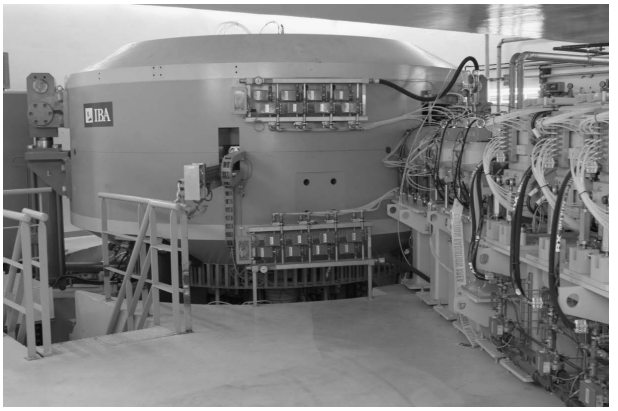



FIGURE 2.17 Photograph of an IBA isochronous cyclotron used for radiation therapy. The beam line may be seen on the right. The yoke lifting system, seen on the far left, can be used to open the cyclotron for service. (Courtesy of IBA, Louvain-la-Neuve, Belgium, 2014, www.iba-worldwide.com. With permission.)


FIGURE 2.18 Don't try to listen to kiss-fm near the cyclotron.

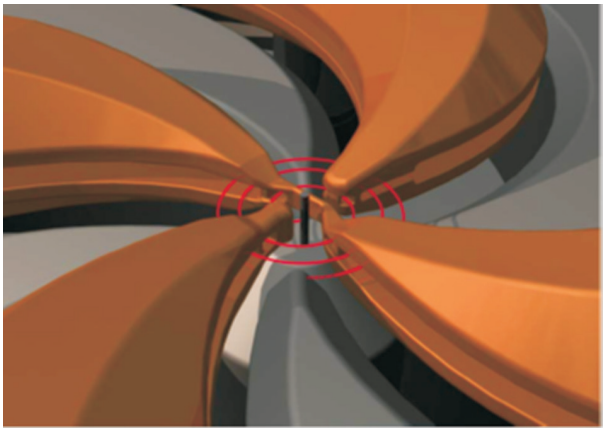


FIGURE 2.19 The Varian isochronous cyclotron with four spiral-shaped dees. The four dees provide four acceleration "kicks" per revolution. The "chimney" ion source is at the center. The path of the protons is shown in red. (Courtesy of Varian Medical, Palo Alto, CA, 2014.)

made of graphite, followed by a magnetic analyzer that selects the required energy width. The power consumption at full beam extraction is 450 kW.

Varian manufactures a 250 MeV superconducting isochronous cyclotron. The following description is based on the articles by Jongen (2010) and Röcken et al. (2010). Unlike IBA, the Varian unit is equipped with four RF cavities (dees) (Figure 2.19) to maximize energy gain per turn. A graphite variable energy degrader is used to adjust proton energy over the range from 70 to 250 MeV. The cyclotron is 3 m in diameter and weighs 90 tons. The magnetic field strength ranges from 2.4 to 4.0 T. The system is capable of beam scanning. The maximum extracted beam current is 800 nA. The RF is 73 MHz (this is the second harmonic). The potential difference between the dees varies between 80 and 130 kV (at the extraction radius).

#### 2.7.1.3 SYNCHROCYCLOTRONS

A synchrocyclotron has the same geometry as a cyclotron: a large magnet with circular pole pieces and an azimuthally symmetric magnetic field. The magnetic field is vertical with a positive field index. The RF is varied to maintain particle synchronism with the applied voltage. This requires a cycled rather than a continuous beam, and therefore the average current and dose rate are much lower. The ions make about 10,000–50,000 revolutions during acceleration (Humphries, 2012). The frequency of revolution of the protons decreases as the energy increases. The cycled beam is a disadvantage for pencil beam scanning because it makes scanning difficult.

Mevion Medical Systems manufactures a superconducting synchrocyclotron for radiation therapy. The MEVION S250 produces 250 MeV protons (Figure 2.20) and is gantry mounted. Gantry rotation is limited to 190° (Klein et al., 2012). The dose rate is 2–8 Gy/min. This compact accelerator is only 1.8 m in diameter, weighs 20 tons, and the entire system can be contained in a single room. The magnetic field at the center

FIGURE 2.20 The MEVION S250 is a superconducting synchrocyclotron. This compact accelerator is 1.8 m in diameter and is gantry mounted. (From Mevion Medical Systems, Littleton, MA, 2014.)

is 9 T (Jongen, 2010). The magnet current is on the order of 2000 A (Bloch et al., 2012). The magnet cryocooler does not use liquid helium (Jongen, 2010), certainly significant for a gantry-mounted structure.

The IBA ProteusONE® unit is a compact single-room system that employs a pencil beam scanning system and is capable of intensity-modulated proton therapy (IMPT). The maximum beam energy is 230 MeV. The accelerator is a superconducting synchrocyclotron referred to as S2C2 and has been described in detail by Kleeven et al. (2013). The S2C2 synchrocyclotron is expected to receive (FDA) 510(k) clearance in late 2016. The diameter of the yoke is 2.5 m, and the weight is 50 tons. The magnetic field is 5.7 T at the center and 5.0 T at extraction. The RF voltage is 10 kV, and the frequency ranges from 93 to 63 MHz.

### 2.7.2 SYNCHROTRONS

The other type of accelerator that is used for proton therapy is the synchrotron. Synchrotrons are the only accelerators in use for heavy ion therapy (Schippers, 2012). Synchrotron radiation from electrons limits the use of synchrotrons as electron

accelerators to an energy of about 12 GeV (Humphries, 2012). The Loma Linda accelerator is a synchrotron built by Fermilab specifically for therapeutic use.

A cyclotron is limited in the energy to which it can accelerate protons. The gyrofrequency of revolution is only constant in the nonrelativistic limit. Recall that the constancy of the gyrofrequency is the crucial feature of a classical cyclotron. When this no longer prevails, the protons will not reach the gap between the dees at the correct moment and thus will not be properly accelerated. We have seen that this limitation may be overcome to a degree by modifying the design of the cyclotron to make it a synchrocyclotron or an isochronous cyclotron.

Synchrotrons are not limited by relativistic behavior. Synchrotrons are shown in Figures 2.21 and 2.22. A synchrotron consists of a ring (not a perfect circle) in which

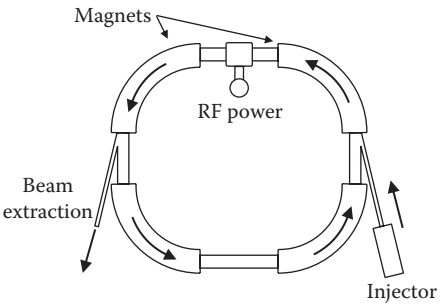



FIGURE 2.21 A simplified overhead schematic depiction of a proton synchrotron. The protons are accelerated in the RF cavity section of the ring. As the protons gain energy, both the RF and the magnetic field strength must increase in synchrony. Synchrotrons are the only method currently in widespread use to accelerate protons to the highest energies (>1 TeV).


FIGURE 2.22 Radiance 330® proton therapy system synchrotron. (Courtesy of ProTom International, Inc., Flower Mound, TX.)

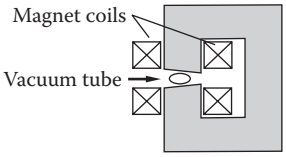



FIGURE 2.23 A simplified schematic cross section through a synchrotron showing the vacuum tube and bending magnets. The particles travel perpendicular to the page.

the particle orbit has a constant "radius." The magnitude of the magnetic field and the RF are varied to maintain the particles at constant orbital radius as they are accelerated. Synchrotrons are not limited in the energy of the particles that they can produce. The only limitations are practical: real estate and money. They can produce energies far higher than other circular accelerators. Humphries (2012) gives two reasons for this:

1. The betatron wavelength of the particles can be maintained constant as the particles are accelerated. It is therefore possible to avoid orbital resonances that limit the energy attainable with an isochronous cyclotron.
2. The *B* field need only extend over a small annulus (Figure 2.23) rather than the full circular volume. Small modular magnets can be used for this purpose rather than the huge monolithic behemoths necessary for cyclotrons.

The proton has the highest charge-to-mass ratio of any nucleon, and therefore protons can reach the highest energy/nucleon for a given magnetic field strength. The limit of the kinetic energy of the ions is set by the magnetic field strength and the radius of the ring. Synchrotrons are the only method currently in widespread use to accelerate protons to very high energies (>1 TeV). The Large Hadron Collider (LHC) at CERN is a synchrotron. The energy of the protons in the LHC is about 13 TeV, five orders of magnitude higher than the protons used for therapy.

The steps in beam production are

1. Inject protons from the ion source and the injector into the accelerator, fill the ring with ~$10^{11}$ protons (Schippers, 2012) at an energy of a few MeV.
2. Accelerate the particles in the ring to the desired energy.
3. Extract the beam into the beam line.
4. Ramp down, decelerate, and dump remaining protons; return to Step 1 and repeat.

The injector may consist of a small drift tube linac or an RF quadrupole (RFQ). The injection energy is 3–7 MeV. The protons travel around in an orbit of constant radius. As they travel around the ring, they gain energy. To maintain a constant

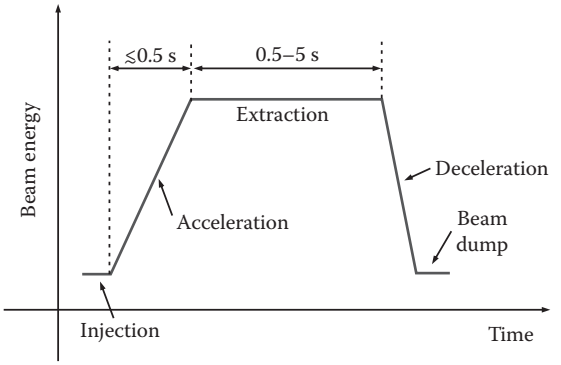



FIGURE 2.24 The acceleration cycle for a synchrotron. After injection at an energy of a few MeV, the particles are accelerated in 0.5 s or less. Over a period of 0.5–5 s, the particles are extracted and delivered to the patient. Any remaining particles are decelerated and dumped.

orbital radius, the magnetic field strength must increase. The diameter is on the order of 5–8 m. The acceleration occurs in RF acceleration cavities spaced around the synchrotron. These are in straight sections. The acceleration mechanism is basically the same as in the RF waveguide of a linac. The RF must increase in synchrony with the magnetic field to stay in phase with the circulating protons.

The acceleration cycle (for a slow extraction) synchrotron is shown in Figure 2.24. The extracted beam is not continuous. The protons are supplied in "spills" during extraction. The slow extraction shown in Figure 2.24 takes too long for energy modulation (Schippers, 2012).

The particles execute about $10^9$ revolutions, and therefore the magnets need to be very precisely aligned. Strong focusing with $n \gg 1$ produces a small beam diameter. Quadrupole magnets are used for focusing (see Section 2.8). Focusing can be accomplished separately from beam bending. Bending is accomplished with dipole magnets. Focusing is accomplished with quadrupole magnets grouped into a set. The arrangement of all the magnets around the circumference of the synchrotron is referred to as the focusing lattice.

Synchrotrons hold a number of advantages over cyclotrons. The energy can be easily varied with a synchrotron. In a cyclotron, the particles are accelerated until they reach the extraction radius, and thus they have a fixed energy for the applied magnetic field. No energy degrader is necessary for a synchrotron. The energy of the beam can be varied without the need for an energy selection system as used with cyclotrons. As a consequence, low-energy protons have the same intensity as high-energy protons. The absence of a degrader implies less neutron contamination and less radioactivity. A separate beam energy can be chosen for each gantry angle. A synchrotron has many small components that can be built in series rather than the massive monolithic magnet used for a cyclotron. Synchrotrons are therefore less massive.

The main drawbacks of synchrotrons are the complex operation cycle and low average beam current and hence dose rate. Synchrotrons require more space (diameter 5–8 m). In addition, the injection system requires preacceleration. Synchrotrons for radiation therapy are manufactured by Hitachi, Mitsubishi, ProTom International, and Optivus Proton Therapy.

The ProTom Radiance 330® synchrotron (Figure 2.22) accelerates protons to energies ranging from 70 to 250 MeV, with the possibility of extending this to 330 MeV. This system is based on technology developed at the Lebedev Physics Institute in Russia. The external ring diameter is about 5.0 m, and the weight of the accelerator is about 15 tons. This machine is capable of beam scanning with a variable pencil beam size. This accelerator received FDA 510(k) clearance in May 2014. According to Klein (2014), the time required for acceleration up to 330 MeV is 1 s, and the time for beam extraction is 0.1–10 s (Figure 2.24). Klein (2014) states that the maximum power consumption is about 100 kW, and that the average is about 50 kW. The dose rate is expected to be 2 Gy/min for large fields (40 cm × 30 cm) according to Wang et al. (2011).

The Hitachi synchrotron is called PROBEAT (Hitachi, 2014). This machine began clinical use at MD Anderson Cancer Center in 2008. It uses a 7 MeV linac as an injector. It has a patented RF extraction system that allows high-speed beam on and off for spot scanning. The accelerator progressively steps through a series of preset energies for spot scanning. The beam delivery can be synchronized with a patient's respiratory signal.

Optivus (San Bernardino, California) manufactures the Conforma 3000. This company grew out of the Loma Linda proton therapy center. This 250 MeV Fermilab-designed synchrotron has been upgraded with many enhancements over the years.

## 2.8 BEAM TRANSPORT AND GANTRIES

Once the beam is generated in the accelerator, it must be transported to the patient. There are a number of accelerator and gantry arrangements in use. An accelerator can feed one room or multiple rooms. The treatment time per fraction is on the order of 20 min. Most of the time a patient spends in the treatment room is devoted to setup. To avoid wasting expensive beam time, it is useful to deliver the beam to a number of treatment rooms sequentially so one patient can be treated while others are being set up. Some rooms may have a fixed nonrotating treatment beam and others a gantry that rotates through an arc that may be less than 360°. In some cases, the accelerator is gantry mounted. An eye beam uses a fixed gantry with a single scatterer. The treatment depth is less than about 35 mm, and therefore a short SSD is acceptable because inverse square attenuation is not important. If an accelerator feeds multiple rooms, a beam line with a switchyard is required. The beam must be deflected and focused

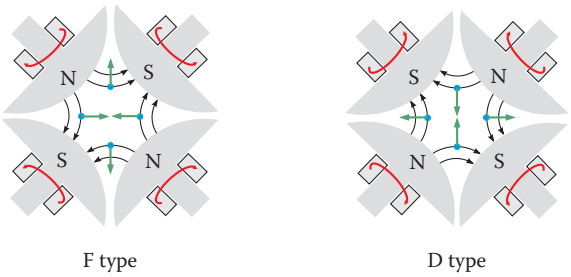



FIGURE 2.25 Cross section of quadrupole magnets used for beam focusing. The two types of arrangements used are F type and D type. In each case, the magnetic field at the center is zero, and thus particles there are unaffected. In an F type arrangement, particles that stray from the center in a horizontal direction experience a focusing force, whereas those that stray in the vertical direction are defocused. A D type quadrupole acts in just the opposite fashion. Sequential D and F type quadrupoles result in a net focusing effect.

periodically en route to the patient. The focusing compensates for any transverse particle motion and for the mutual repulsion of the positive ions. Beam bending or deflection is accomplished with a simple dipole magnetic field such as that produced by a "C" dipole, like the one depicted in Figure 2.23. Focusing requires the use of a more complex magnet arrangement consisting of successive combinations of quadrupole and sextapole fields.

The magnetic field arrangement for F and D type quadrupole magnets is shown in Figure 2.25. Focusing can be accomplished using successive combinations of F and D type fields (see Humphries, 2012). This is analogous to pairing convex and concave lenses.

One of the most difficult and complex aspects of proton therapy in a multiple room facility is beam delivery. The beam delivery system consists of the beam line and switchyard where the beam is diverted to various treatment rooms, as shown in Figures 2.26 and 2.27. A system is needed for selecting the appropriate room for beam delivery. It is necessary to develop rules for priority of beam switching. It must be verified that only one room can receive the beam. As the beam travels from the accelerator to the patient, it passes through many bending, focusing, and steering magnets; all must be monitored (interlocks) to ensure correct beam delivery to the patient. Beam control systems are needed to deal with the safety issues associated with beam transport from room to room. It is necessary to monitor bending magnet power supplies, beam stops, area emergency buttons, door interlocks, and so forth.

Rotating gantries are large. High-energy protons have a large radius of curvature in a magnetic field generated by room temperature magnets (see Equation 2.13). Large gantries are one of the reasons for the high cost of proton therapy. Some facilities have up to four gantries supplied by a single accelerator (PTCOG, 2014). An IBA gantry is shown in Figures 2.28 and 2.29.

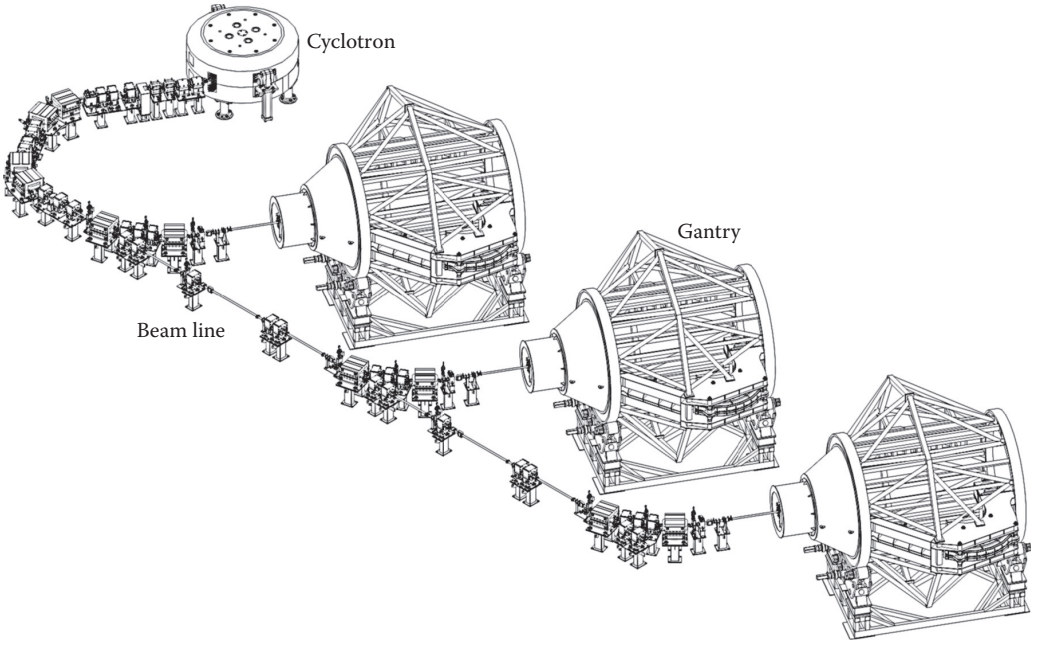



FIGURE 2.26 Beam line feeding three isocentric gantries. The magnets in the beam line are for steering and focusing. (Courtesy of IBA, Louvain-la-Neuve, Belgium, 2014, www.iba-worldwide.com.)

FIGURE 2.27 Photograph of a beam line. Quadrupole magnets in the beam line provide focusing. (Courtesy of IBA, Louvain-la-Neuve, Belgium, 2014, www.iba-worldwide.com.)

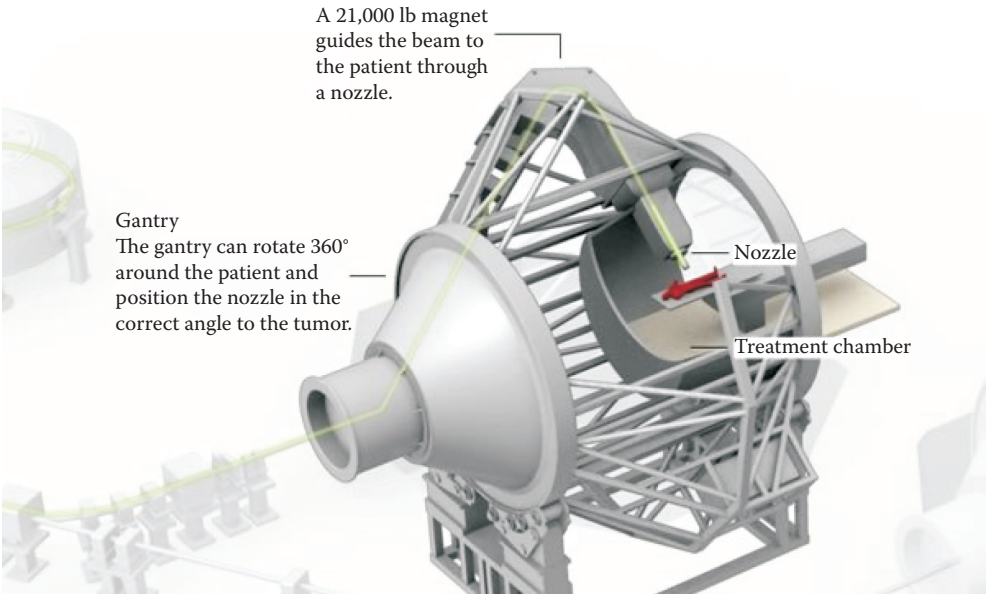



FIGURE 2.28 Diagram showing a drawing of the gantry in the photo in Figure 2.29. The proton beam follows the pale green path shown. (Courtesy of IBA, Louvain-la-Neuve, Belgium, 2014, www.iba-worldwide.com.)

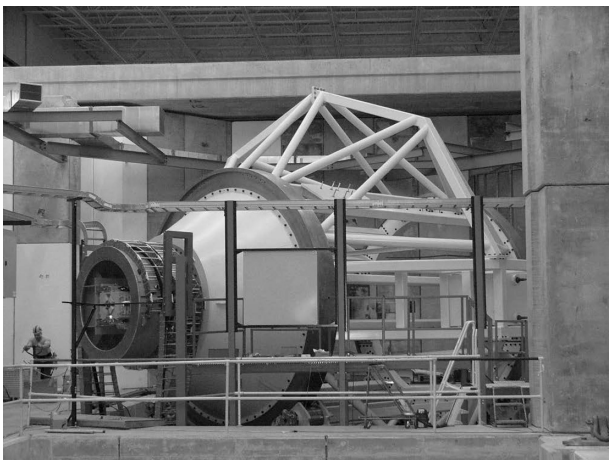


FIGURE 2.29 The gantry for an IBA cyclotron. The entire structure rotates on the two rings shown. The beam enters from the left at the center of the gantry structure. To appreciate the scale, notice the man standing in the lower left. (Courtesy of IBA, Louvain-la-Neuve, Belgium, 2014, www.iba-worldwide.com.)

## 2.9 LATERAL AND AXIAL BEAM SPREADING

The proton beam produced by the accelerator is very small in cross section. When the beam emerges from the accelerator, it may only be a few millimeters in diameter. There are a variety of methods that have been employed to spread the beam out in the lateral direction to produce a broad beam. This is conceptually no different than for electron beams produced by linacs. There are two main approaches to this problem: passive scattering and active scanning. The beam can be spread laterally by the use of a scattering element positioned upstream of the collimation system (see Figure 2.30). Electromagnetic scanning has also been used (active scanning, shown in Figure 2.33). These are the two basic techniques. Up until fairly recently, only the first technique had been used. It is technologically simpler but requires compromises in the dose distribution.

The traditional passive scattering method of beam spreading involves placement of a scattering element in the beam just as in conventional radiotherapy to spread electron beams laterally. A single lead foil placed in the beam will result in a Gaussian intensity profile with a field diameter of only about 30 mm, as shown in Figure 2.31. This is too small, except possibly for treatment of the eye. To provide

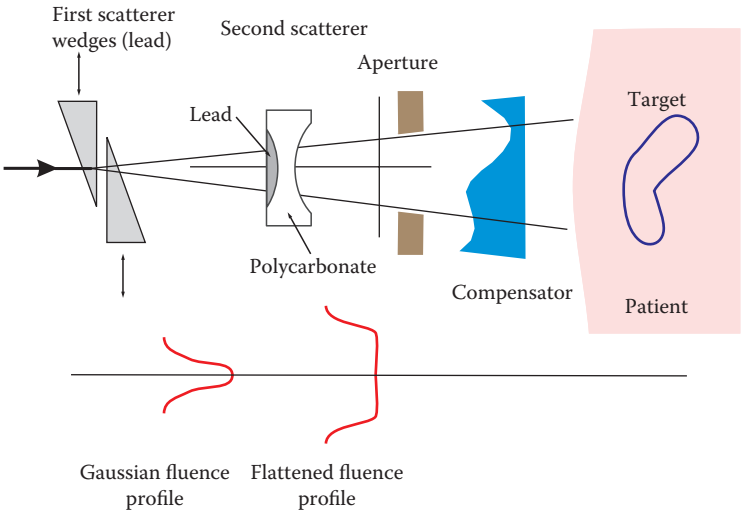



FIGURE 2.30 A passive scattering system. The first scatterer is a set of movable lead wedges. This creates a Gaussian beam profile. The second scatterer is thicker in the center than at the periphery and functions like a flattening filter. Lead provides the scattering, and the polycarbonate is shaped to ensure that the beam energy does not vary laterally. Passage through the second scatterer results in a flattened beam profile and a uniform energy across the beam. The beam then traverses a patient-specific collimator and compensator. The compensator is designed so that the range of the protons corresponds to the location of the distal edge of the target. (From McDermott, P. and C. Orton, *The Physics and Technology of Radiation Therapy*, Madison, WI: Medical Physics Publishing, 2010, fig. 20.41. With permission.)

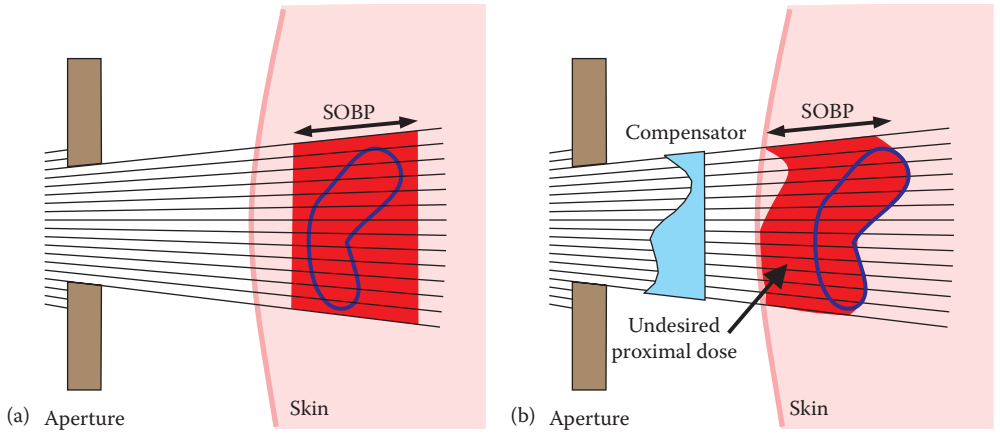



FIGURE 2.31 Customized patient treatment with a passive scattering system. (a) The SOBP prior to insertion of a compensator. (b) The SOBP after insertion of a customized compensator. The conformity with the distal edge of the tumor is excellent, but the shift in the SOBP produces an unwanted dose proximal to the target. (From McDermott, P. and C. Orton, *The Physics and Technology of Radiation Therapy*, Madison, WI: Medical Physics Publishing, 2010, fig. 20.41. With permission.)

additional spreading, a second lead scatterer can be introduced. The thick portion in the center reduces the Gaussian peak by scattering protons to the periphery of the beam. The second scatterer also includes a shaped piece of polycarbonate plastic to maintain uniform beam energy across the field. This maintains a uniform range laterally.

Range compensation is used to conform dose to the distal edge of the tumor. Compensators can be made of materials similar to tissue (acrylic or wax can be milled). This can also be accomplished with a scanning beam by raster scanning a narrow proton beam across the target while modulating energy and intensity. A custom compensator is placed in the passively scattered beam to conform the distal edge of the SOBP to the target. This is illustrated in Figure 2.31. Because of the fixed depth of modulation, conformity with the proximal edge of the tumor is poor. This is mitigated by the use of multiple beams. A photograph of a custom patient aperture and compensator is shown in Figure 2.32.

The second technique used for lateral beam spreading is active scanning. A pair of orthogonal magnets are used to deflect a proton pencil beam, as shown in Figure 2.33. The magnetic field is used to deflect the beam and "paint" the field. The dose can be delivered layer by layer by changing the energy between layers. The dose is delivered spot by spot or line by line and then layer by layer. This has the advantage that no physical patient-specific beam modifiers are necessary. In addition, none of the protons are "wasted" by filters, nor are any neutrons produced in those filters. The dose rate for active scanning is usually specified by the time necessary to deliver a uniform dose to a 10 cm cube (1 L). The treatment time

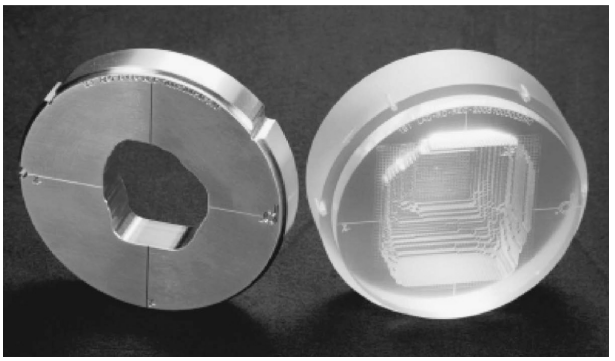


FIGURE 2.32 A patient-specific aperture and range compensator. The aperture shapes the beam laterally, and the range compensator shapes the distal edge of the Bragg peak. The aperture is made of brass. These must be stored for a time after use because they are activated. They are then recycled. The compensators are made out of acrylic. (From the *decimal point*, courtesy of www.dotdecimal.com. With permission.)

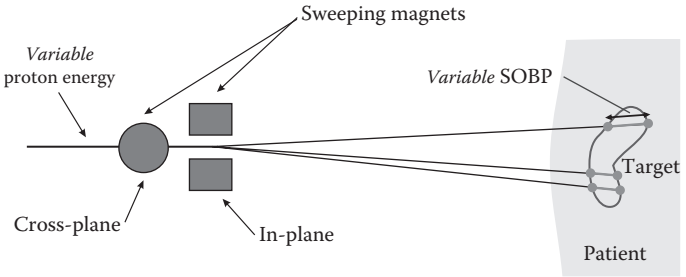



FIGURE 2.33 In active scanning, the target is painted with a pencil beam. The pencil beam is steered by a set of magnets. If the energy of the beam can also be changed dynamically, then the distal edge of the SOBP can be made to correspond to the distal boundary of the target. When the energy is varied simultaneously, the proximal edge of the SOBP can be made to conform to the proximal position of the target. (From McDermott, P. and C. Orton, *The Physics and Technology of Radiation Therapy*, Madison, WI: Medical Physics Publishing, 2010, fig. 20.44. With permission.)

depends not only on the beam current, but also on the time interval between layers when the beam energy is changing (Lu and Flanz, 2012).

Active scanning clearly requires a very sophisticated planning and delivery system. There are serious safety and quality assurance issues associated with this approach. These safety issues are related to the intensity of the pencil beam. Because the whole energy of the beam is concentrated in a small spot, a redundandant fast beam abort system is necessary. Tumor motion is also problematic with this technique. Remedies are to rescan the target (sometimes called repainting) or use gating. Rescanning should be performed on a timescale larger than the period of motion. A future possibility is to follow the tumor motion in real time and make scanning adjustments on the fly.

## 2.10 BEAM CALIBRATION

ICRU Report No. 78 (2007) recommends the International Atomic Energy Agency (IAEA, 2006) proton dosimetry calibration protocol referred to as Technical Report Series (TRS) 398: "Absorbed Dose Determination in External Beam Radiotherapy: An International Code of Practice for Dosimetry Based on Standards of Absorbed Dose to Water." The code of practice for proton beams is found in chapter 10 of this document. This protocol gives worksheets and provides the tables and graphs necessary for calibration of proton beams with energies between 50 and 250 MeV. The protocol utilizes an ionization chamber calibration for absorbed dose to water in a Co-60 beam. The ionization chamber may be either a cylindrical or a plane parallel model. For cylindrical chambers, the reference point is taken as the center of the cavity volume. For plane parallel chambers, the reference point is at the inner surface of the entrance window.

The reference depth is the depth at which the ion chamber reference point is placed. This is at the center of the SOBP, as shown in Figure 2.34. The width of the SOBP is normally defined by the distance between the 95% dose levels (Figure 2.34).

Proton dose calibration requires a beam quality specifier, as do photon beams. The quantity used in TRS 398 is called the residual range $R_{res}$. One of the advantages of this quantity is that it is easily measured. The residual range, measured at depth $z_{ref}$ (in units of g/cm$^2$), is defined as

$$R_{\text{res}} = R_p - z_{\text{ref}}, \tag{2.41}$$

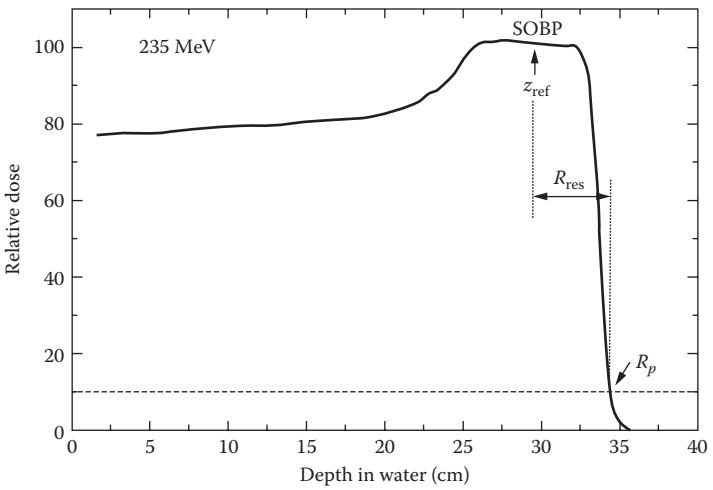



FIGURE 2.34 Shows $z_{\text{ref}}$, the reference depth of measurement for proton beam calibration. This point is at the center of the SOBP. The beam quality is defined by $R_{\text{res}}$, and $R_p$ is the practical range. (Reproduced from International Atomic Energy Agency (IAEA), Absorbed dose determination in external beam radiotherapy: An international code of practice for dosimetry based on standards of absorbed dose to water, Technical Report Series 398, in *Code of Practice for Proton Beams*, chap. 10, Vienna: IAEA, 2006. With permission.)

where $R_p$ is the practical range (in units of g cm$^{-2}$), which is defined as the depth at which the absorbed dose beyond the Bragg peak or SOBP falls to 10% of its maximum value (Figure 2.34).

The absorbed dose in water $D_{w,Q}$ at the reference depth $z_{\text{ref}}$ in the absence of the ion chamber from a proton beam of quality $Q$ is

$$D_{w,Q} = M_Q N_{D,w}^{Q_0} k_{Q,Q_0}, \tag{2.42}$$

where:

- $M_Q$ is the measured charge collected from the ion chamber at depth $z_{ref}$ corrected for pressure and temperature, electrometer calibration factor, polarity effect, and ion recombination
- $N_{D,w}^{Q_0}$ is the Co-60 calibration factor supplied for the user's ion chamber by a calibration lab
- $k_{Q,Q0}$ is the chamber-specific calibration factor for the chamber response for the user's beam quality $Q$ (as specified by Equation 2.41)

These numbers can be found in table 10.III of the protocol as a function of $R_{res}$ for a large number of commercially available ion chambers.

## 2.11 DOSE CALCULATION ALGORITHMS

Proton beam dose calculation algorithms are relatively simple. Unlike the case for photons, the range of secondary particles is relatively small. Commercial treatment planning systems that will handle proton dose calculations are Eclipse (Varian Medical Systems), Pinnacle (Philips Healthcare), XiO (Elekta), and RayStation (RaySearch Laboratories). Vendors vary in their capability for mixed modality planning and IMPT. Philips Pinnacle and Elekta XiO use a pencil beam model based on the work of Hong et al. (1996). RayStation uses a Fermi-Eyges-based pencil beam algorithm (see Chapter 4). A detailed discussion of pencil beam algorithms for photon beams can be found in Chapter 3. The water equivalent depth is used to account for inhomogeneities. Pencil beam algorithms can, in principle, take account of the lateral effects of inhomogeneities to some degree.

## 2.12 INHOMOGENEITIES

The effects of inhomogeneities on dose distributions are different from those for photons. For photons, the intensity is changed by the presence of an inhomogeneity. For protons, the intensity stays roughly the same (see Figure 2.7); rather, it is the range

that changes. Due to the sharp distal falloff of the beam, it is essential to include the effects of inhomogeneities; otherwise, large errors can result.

There are three phenomena caused by inhomogeneities:

1. Range modification
2. Edge scatter if the inhomogeneity does not fill the beam
3. Dose modification from thin slivers

The last effect is illustrated in Figure 2.35.

The presence of contrast agents and artifacts due to the presence of metals can be significant, as these will have an effect on range. The relative position of inhomogeneities can change due to tumor motion or shrinkage. One must look very carefully at this, and tumor shrinkage may require replanning.

Both the proton stopping power and, to a lesser extent, the fluence are affected by the tissue electron density. Computed tomography (CT) images provide electron density values just as they do for conventional photon therapy. This permits true dose compensation (as opposed to missing tissue compensation). It is necessary to have a relationship between the CT number and the stopping power.

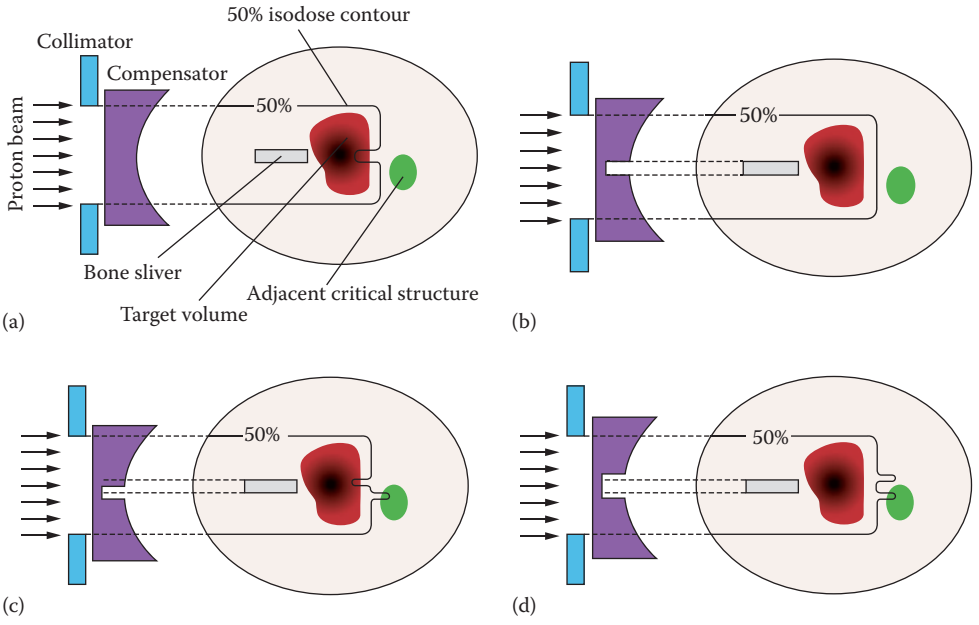



FIGURE 2.35 The effect of a thin sliver inhomogeneity. (a) There has been no modification of the compensator to account for the inhomogeneity. As a result, there is a cold spot in the target volume. (b) The compensator has been modified to adjust for the presence of the bone sliver. (c) What can happen if there is even a slight misalignment of the compensator. (d) "Opened" compensation. The target is fully covered in the event of a small misalignment, but the critical structure receives some dose. (Reproduced with permission from *Physics Today*, M. Goitein, et al., September 2002, p. 45, fig. 1. Copyright 2002, American Institute of Physics.)

For proton therapy, proton stopping powers are determined from CT scans, much like electron densities are determined for photon and electron therapy. This is not as straightforward as the determination of electron densities, however, since radiation used for CT scanning is x-rays, and x-ray attenuation is a simple function of electron density. Measured CT numbers in Hounsfield units (HUs) can be readily converted to electron densities. Ideally, one would want to use a CT unit that employs protons instead of x-rays, but no such machine is currently in use for proton treatment planning. Consequently, it is necessary to convert CT numbers to proton stopping powers, which depend on both the electron density and the chemical composition of tissues. This can be done using a calibration curve like the one published by Schaffner and Pedroni (1998) shown in Figure 2.36. The authors of this paper state that use of this calibration curve should lead to errors in prediction of the range of protons no greater than about 1.8% for bone and 1.1% for soft tissue. Calibration curves, however, may be a unique property of the scanner and the peak kilovoltage used, as is the case for the electron density–CT number calibration curves for photon and electron beam treatment planning.

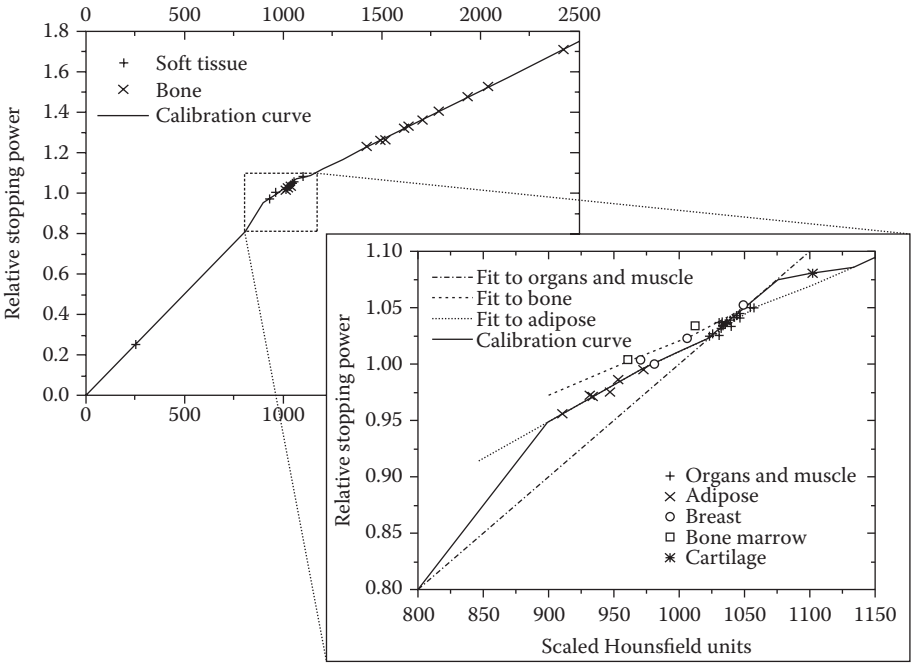



FIGURE 2.36 Relative stopping powers as a function of scaled HU for biological tissues grouped into soft tissue and bone (behind) and in the enlarged section of the graph (front) for specific soft tissues. The three broken lines show the three linear fits for soft tissue, adipose, and bone, and the solid line is the chosen calibration curve. (From B. Shaffner and E. Pedroni, The precision of proton range calculations in proton radiotherapy treatment planning: Experimental verification of the relation between CT-HU and proton stopping power, in *Physics in Medicine and Biology*, fig. 1, p. 1582, vol. 43, issue 6, pp. 1579–1592, 1998. © Institute of Physics and engineering in Medicine. Reproduced by permission of IOP Publishing. All rights reserved.)

## 2.13 DOSE DISTRIBUTIONS

Passive scattering provides poor conformity with the proximal edge of the target. To improve this situation, multiple beams are used. Proton plans generally use a smaller number of beams than photon plans. Distal blocking presents new options, such as the use of patch fields (Figure 2.37). Very careful patient alignment is called for, and care is obviously necessary in matching the range of the patch field with the lateral penumbra of the first field. An x-ray tube is often put in the beam line, and online corrections are made to adjust patient position. If the patch field passes through significant inhomogeneities, this is likely to increase the uncertainty in the range. The possibility of range errors and the resultant overlap make some clinics reluctant to use this technique.

One of the frequently cited advantages of proton therapy, particularly for pediatric patients, is low integral dose. The whole-body integral dose is the integral of the dose over all mass elements in the patient's body (including the target volume). The integral dose is therefore the total energy absorbed by the patient from the radiation. This is thought to be correlated with secondary radiation-induced cancers. The latency period for radiation-induced cancers is on the order of 5–7 years (Hall and Giaccia, 2012). The lower integral dose is due to the fact that there is almost no exit dose for protons and the lateral penumbra is generally smaller than that for photon beams. It is claimed that proton therapy reduces the integral dose by a factor of 2 or 3 in comparison to photon therapy (Paganetti, 2012a). The dose distribution is also clearly a factor in addition to the integral dose. This issue is not quite as straightforward as it might first appear. For passively scattered proton beams, the neutron contamination could conceivably make a significant contribution to secondary cancer induction because of the high RBE of the neutrons (Lomax, 2012, and references therein).

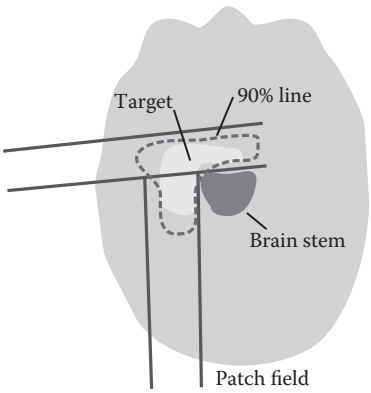



FIGURE 2.37 A patch field is used to fill in coverage of a tumor that wraps around the brain stem. The distal edge of the patch field is adjusted to coincide with the edge of the lateral beam. (From McDermott, P. and C. Orton, *The Physics and Technology of Radiation Therapy*, Madison, WI: Medical Physics Publishing, 2010, fig. 20.52. With permission.)

Neutron contamination is much lower for scanned beams because there is much less material in the beam line for this modality.

Some dose distributions are shown in Figures 2.38 and 2.39 that illustrate some of the principles associated with proton treatment planning.

Due to the uncertainties described in Section 2.12, the dose distributions calculated by a treatment planning system must be considered "ideal" dose distributions.

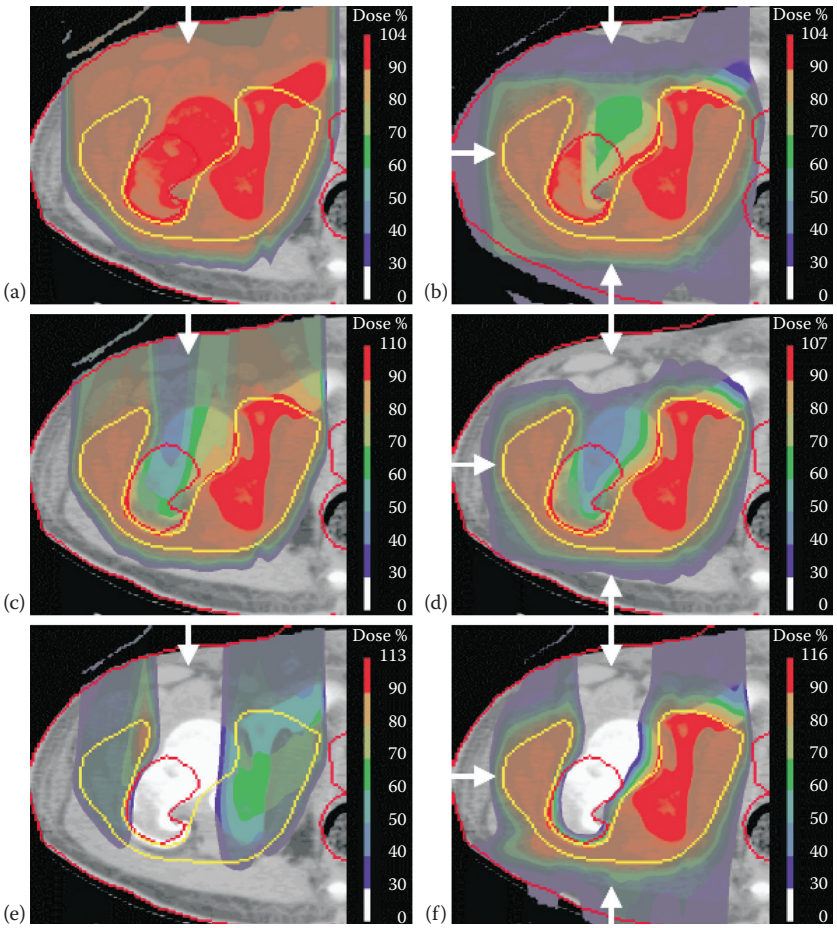



FIGURE 2.38 Treatment of Ewing's sarcoma with protons. The dose distribution is shown by the color wash (see legend on the right of each frame). The target is the yellow contour. Critical structures are outlined in red. The arrows show beam directions. (a) A single passively scattered beam is used. Note the good conformality laterally and at the distal edge of the target, but not at the proximal target border. (b) A three-field dose distribution from passively scattered beams. (c) A single field delivered by an actively scanned beam. (d) Three fields delivered by actively scanned beams. Each of these fields delivers a near-uniform dose. (e) One of the intensity-modulated fields shown in panel (f). (f) A three-field optimized intensity-modulated plan. The dose distribution is extraordinarily conformal. (Reproduced with permission from *Physics Today*, M. Goitein et al., September 2002, p. 50, fig 4. Copyright 2002, American Institute of Physics.)

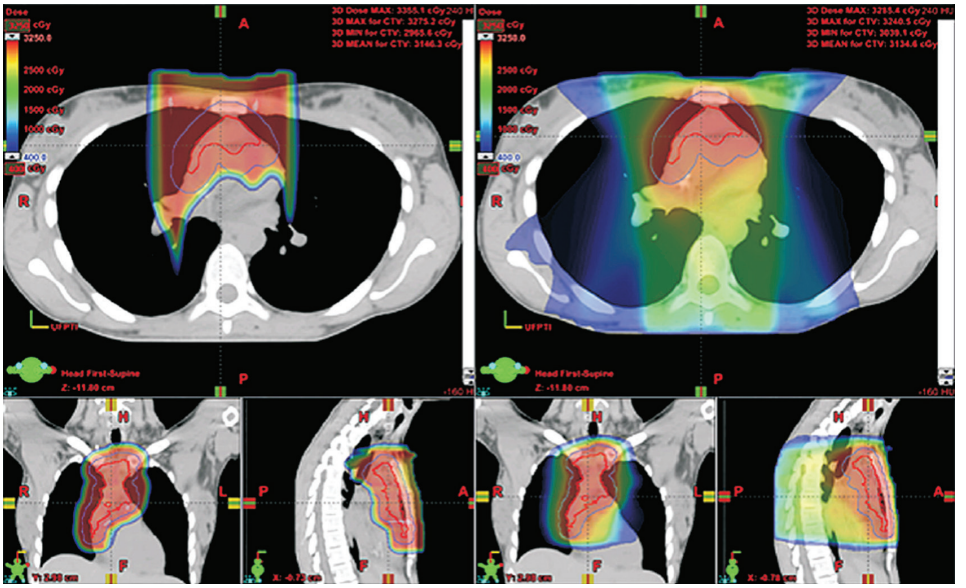


FIGURE 2.39 Color wash dose distribution for protons on the left and photon intensity-modulated radiation therapy (IMRT) on the right. The proton dose distribution is dramatically more conformal. The planning target volume (PTV) is shown as the light blue anterior midline contour. This is a Hodgkin's lymphoma patient. (From Hoppe, B.S., et al., Involved-node proton therapy in combined modality therapy for Hodgkin lymphoma: Results of a phase 2 study, *Int. J. Radiat. Oncol. Biol. Phys.* 89(5), 1053–59, 2014.)

Even proponents (Palta et al., 2009) admit that "what you see is not what you get." Proton therapy is more sensitive to CT number and stopping power accuracy, organ motion, and anatomical changes. For this reason, replanning may be needed more frequently for proton therapy than photon therapy, and image guidance appears to be almost mandatory.

## 2.14 RADIATION SHIELDING

This discussion of shielding is based on the excellent review article by Mukherjee (2012). See also NCRP No. 144 (2003). The example facility is the West German Proton Therapy Center in Essen.

The main concern is for neutrons and gamma rays produced when the proton beam strikes any material in the beam or along the beam line. Items placed in the beam include energy degraders, beam-shaping components made of brass or polystyrene, and of course patients. There is also the possibility of accidental beam loss, in which the proton beam strikes components of the accelerator or beam transport system. Proton irradiation results in prompt neutron and gamma emission (see Section 2.2). In turn, the neutrons lead to activation of beam line components and of the air in the treatment room.

The West German Proton Therapy Center is located about 8 m underground. Ordinary density concrete is used (2.35 g $cm^{-3}$ = 147 lb $ft^{-3}$) as shielding material, and walls and ceilings are 2.0–2.5 m thick (Figure 2.40).

Neutron production from proton irradiation has been estimated from Monte Carlo calculations. It is directly proportional to the proton beam current. The Monte Carlo calculation is based on irradiation of a 30 cm × 30 cm × 30 cm polystyrene phantom with a 3 mm diameter beam.[6] In the worst-case scenario, the neutron production rate at a distance of 1 m (from the point of beam incidence) in the forward direction is 13 mSv $h^{-1}$ $nA^{-1}$ for the 235 MeV proton beam energy. The mean energy of the neutrons is 68 MeV. Neutrons with this energy have a tenth value layer (TVL) thickness of 65 cm in concrete. It is assumed that the shielding thickness calculated for the neutrons is also adequate for the gamma rays produced in the treatment room and in the shielding material. The beam on time is assumed to be 500 h/year. This presumes that the beam is on one-fourth of the time for a 2000 h work year. The dose equivalent at the point of interest is given by

$$H = B\frac{WUT}{d^2}, \tag{2.43}$$

where:

- $B$ is the barrier transmission factor
- $W$ is the workload
- $U$ is the use factor
- $T$ is the occupancy factor
- $d$ is the distance (in m) from the *source of the neutrons* to the point of interest

The source of the neutrons is assumed to be at the isocenter.

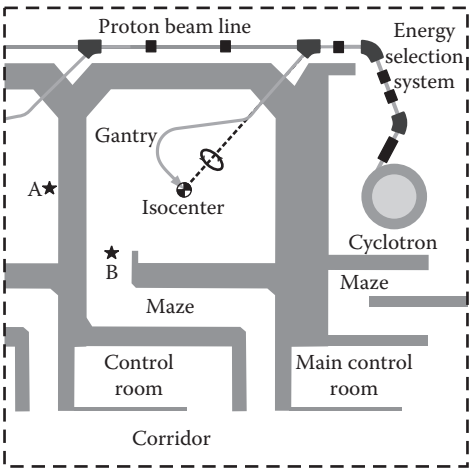



FIGURE 2.40 A plan view of a proton therapy treatment room, the control room, and the adjacent room housing the cyclotron. The barriers consist of ordinary density concrete.

A plan view of a treatment room is shown in Figure 2.40. The radiation level at the entrance door is the sum of the direct component that penetrates the maze barrier and the component that diffuses from the maze entrance (labeled B in the diagram). Shielding details are discussed in the paper by Mukherjee (2012) including consideration of "skyshine."

Mukherjee (2012) raises the issue of activation of the room air, in particular by thermalized neutron interactions with $^{40}$Ar (abundance of 0.46% by volume) via the reaction $^{40}$Ar(n, γ)$^{41}$Ar. Radioactive $^{41}$Ar is a gamma emitter (1.3 MeV) with a half-life of 1.8 h. This poses an internal hazard as it is inhaled. This has been evaluated indirectly by measuring the neutron levels in the room and by assuming an air exchange rate of eight times per hour. The calculated concentration is 26 Bq m$^{-3}$ ($7 \times 10^{-4}$ pCi/L), well below the permissible limit of 200 Bq m$^{-3}$.

Activated cyclotron parts potentially pose a risk to therapists who need to handle these items many times throughout a treatment day. The main concern is the brass patient beam apertures (see Figure 2.32). A dose reduction of 50% is possible by waiting 15 min before handling these items. After 1 week, the activity of these apertures drops to an insignificant level.

## 2.15 NEW DEVELOPMENTS

There are some accelerator advances on the horizon that may revolutionize proton therapy. These advances may significantly reduce the space requirements and cost of proton therapy.

### 2.15.1 DIELECTRIC WALL ACCELERATORS

Dielectric wall acceleration of protons is an outgrowth of nuclear weapons research at the Lawrence Livermore National Laboratory. This technology has been licensed to the Compact Particle Acceleration Corporation for development of a medical proton accelerator that will fit into an ordinary linac vault. The individual pulses from this accelerator can be varied in intensity, energy, and spot width, making IMPT possible. The accelerator consists of a hollow tube with alternating rings of electrical insulators and conductors. A transmission line sends brief pulses to the conductors as the proton bunch travels down the tube. The insulator can withstand very high electric fields for short periods of time without undergoing dielectric breakdown. It may be possible to accelerate protons to 200 MeV in a distance as short as 2 m, which would make proton tomotherapy a reality.

### 2.15.2 PROTON LASER ACCELERATORS

Very high-intensity short laser pulses can accelerate protons to energies sufficient for radiotherapy. Protons accelerated in this manner have a wide energy spectrum, however, so magnetic field energy separation is necessary in order to obtain the desired

SOBP for therapy. This means that most of the protons have to be removed from the beam before use, which reduces the intensity of the useful beam considerably. Because of this lack of efficiency, laser beams of extremely high intensity are required. Such high intensities have not yet been achieved, but several groups in several countries are working to make such beams a reality.

### 2.15.3 PET VALIDATED TREATMENT

Small amounts of positron emitters are produced in tissue irradiated with protons such as $^{11}C$, $^{15}O$, and $^{10}C$ (see Section 2.3). There is apparently enough of this to perform imaging studies. The measured positron activity is compared to the expected activity calculated by Monte Carlo techniques.

## 2.16 SUMMARY

Protons with energies of 200–250 MeV are needed (see Figure 2.41) to reach the deepest anatomical structures (about 30 cm). The stopping power of protons is proportional to $1/v^2$. This leads to a sharp rise in energy deposition at the end of the track where the fluence begins to drop. The net effect is that the dose rises sharply near the end of the track, reaches a peak (the Bragg peak), and then drops precipitously. The raw Bragg peak extends only over a few centimeters in depth (FWHM ~ (10% – 15%) $d_m$). For most targets, the dose distribution must be spread out in the depth direction. This can be accomplished by adding beams with different energies and amplitudes to obtain an SOBP.

The radiobiological properties of protons are unremarkable, and the RBE is 1.1 for a wide range of biological endpoints. The CGE is the dose in gray multiplied by 1.1.

Charged particles orbiting in a plane perpendicular to a constant magnetic field do so with the cyclotron or gyrofrequency $\omega_0 = qB_0/\gamma m_0$. This frequency is independent of the energy (or orbital radius) of the particles provided; they remain nonrelativistic. This is the basis for the classical cyclotron. The orbits are stable if the vertical magnetic field has a negative radial gradient. More specifically, the stability index $n$ must obey the condition $0 < n < 1$. Particles obeying this condition will be stable to perturbations in the vertical and radial directions. Perturbations in the vertical direction lead to oscillations called betatron oscillations.

The classical cyclotron consists of two D-shaped electrodes called dees placed in a uniform magnetic field. An RF voltage is imposed on the dees with a frequency equal to the cyclotron frequency. Classical cyclotrons are only able to accelerate protons to an energy of about 20 MeV before relativistic behavior interferes with further acceleration. This limitation can be overcome by either (1) introducing a positive radial gradient in the magnetic field or (2) varying the frequency of the applied voltage on the dees. The first solution leads to the AVF or isochronous cyclotron. The second solution is represented by the synchrocyclotron. The isochronous cyclotron retains the constant RF of the classical cyclotron. An azimuthal


FIGURE 2.41 Cartoon "Particles, particles, particles." (Copyright www.ScienceCartoonsPlus.com.)

variation of the magnetic field must be imposed to overcome the instability associated with the positive field gradient. A sectored magnet yoke provides hills and valleys in the field.

A synchrotron is only limited in beam energy by space and money. This accelerator requires preacceleration by the injector. After the protons are injected into the ring, they are accelerated in an RF cavity that occupies a portion of the ring. Small modular C-shaped dipole magnets provide the bending force. As the particles circulate, they gain energy and the magnetic field must increase to maintain the fixed ring radius. When the particles have been fully accelerated, they are extracted. Then the process must start over, and thus beam production is cyclic.

There are trade-offs in the choice of a synchrotron versus a cyclotron for proton therapy. Cyclotrons require a single large magnet and are therefore quite massive

(order of 100 tons), although superconducting cyclotrons are much less massive. Synchrotrons are larger in diameter (5–7 m) than cyclotrons, thus precluding gantry mounting. Cyclotrons accelerate protons to a fixed energy that is not inherently variable, unlike a synchrotron. Cyclotrons therefore require an energy degrader in the beam line for reduced energies. Energy degraders have the disadvantage that they contribute to unwanted neutron production, requiring greater shielding, and the neutrons lead to activation, raising the level of radioactivity and contributing to whole-body dose. Cyclotron beams are not cycled, unlike synchrocyclotrons and synchrotrons, and therefore the average beam currents (and thus dose rates) are higher.

Treatment rooms may have a fixed beam or a rotating gantry. The accelerator can feed multiple rooms or, when superconducting, be gantry mounted and self-contained. Due to the large orbital radius of protons in a magnetic field, accelerators feeding multiple rooms must have large gantries, contributing to the high cost of proton therapy. Multiple room arrangements with a single accelerator require a beam line and switching mechanism. Beam control and safety systems become more complex in this configuration.

Lateral beam spreading is accomplished either by a passive scattering system or by active beam scanning. In passive scattering systems, two scattering elements are frequently used to produce a flat beam. A custom aperture block defines the lateral dimensions of the beam. A range compensator is used to make the distal edge of the SOBP along various ray lines correspond to the distal depth of the target. This system leads to unwanted proximal dose, which is diluted by using multiple beams. In active scanning, a set of deflecting magnets is used to raster scan the target. The target can be painted one layer (in depth) at a time, changing the energy between layers. This is a complex scheme that requires various safety systems. It also raises issues of patient motion management. It can eliminate the need for patient beam apertures and compensators, reduce neutron contamination, and reduce dose proximal to the target.

The standard dose calibration protocol is the IAEA TRS 398. This protocol uses an ion chamber calibrated in water with a Co-60 beam. The ion chamber is positioned at a depth in the center of the SOBP. The beam quality is specified in terms of the residual range. This is the distance from the center of the Bragg peak to the point where the SOBP falls to 10% of its maximum value.

Dose calculation algorithms in commercial treatment planning systems are usually pencil beam algorithms.

Inhomogeneities play a crucial role in proton therapy. The effect of inhomogeneities for proton therapy is different than that for photons. For photon therapy, inhomogeneities modify the photon fluence. For protons, inhomogeneities modify the range rather than the fluence. Range uncertainties can result in errors in the intended position of the distal falloff of the Bragg peak. This can potentially lead to the overdose of an underlying critical structure.

## PROBLEMS

1. Calculate γ and $v/c$ for 250 MeV protons.

2. Calculate the de Broglie wavelength for 250 MeV protons (use nonrelativistic mechanics) and compare it to the range of the strong nuclear force (about $2 \times 10^{-15}$ m). The mass of a proton is $1.67 \times 10^{-27}$ kg, and Planck's constant is $6.62 \times 10^{-34}$ J·s.

3. A delivered or incident beam current of 20 nA forms a 10 cm × 10 cm proton beam. Estimate the dose rate in the Bragg peak in units of cGy/min, assuming that the fluence does not decline significantly with depth.

4. a. Estimate the ratio of the vertical velocity to the azimuthal velocity ($v_z/v$) that would result in an approximate vertical deflection of 1 mm for a 250 MeV proton in a cyclotron. Assume a perfectly uniform vertical magnetic field; the potential difference between the two dees of $10^5$ V and a cyclotron radius is 5 m.

   b. What prevents the particles in an isochronous cyclotron from crashing into the top or bottom of the dees?

5. a. Derive a relativistic formula for the kinetic energy of a charged particle of rest mass $m_0$ in uniform circular motion with radius $R$ in a magnetic field $B_0$.

   b. Calculate the kinetic energy of a proton in a circular accelerator of radius 2.5 m with an average magnetic field strength of 1 T.

6. Derive the perturbation equations, Equations 2.31 – 2.33, starting with Equations 2.15 through 2.17 and 2.25 through 2.27.

7. What is the ratio of the betatron oscillation frequency to the cyclotron frequency for an index value of $n = 0.2$?

8. For a synchrocyclotron, $B_0 = 5.7$ T at the center and 5.0 T at the extraction radius of 1.25 m. The energy of the proton bunch at extraction is 230 MeV.

   a. What is the approximate value of the stability index $n$?

   b. What is the frequency of the applied voltage at extraction?

9. Calculate the dose equivalent on an annual basis for point A in Figure 2.40. The thickness of the concrete is 2.2 m, and the distance from the source (passive scatterer) is 6.5 m. Assume that the proton beam energy is 235 MeV

and the average beam current is 10 nA. Assume an occupancy factor of 1.0 and a use factor of 1/4. How does the result compare with the U.S. Nuclear Regulatory Commission dose limit requirement for members of the public (1.0 mSv/year)?

## SYMBOLS

| | |
|---|---|
| $\vec{a}$ | Acceleration |
| $a$ | Radius of atom |
| $A$ | Atomic mass number |
| $b$ | Impact parameter |
| $b_1$ | Expansion parameter for B field |
| $\vec{B}$ | Magnetic field strength |
| $B_0$ | Vertical magnetic field strength |
| $B_z$ | $z$ component of magnetic field |
| $B_\theta$ | θ component of magnetic field |
| $c$ | Speed of light |
| $D$ | Dose |
| $d$ | Distance from source |
| $D_{w,Q}$ | Absorbed dose in water for beam quality $Q$ |
| $E$ | Total energy |
| $H$ | Dose equivalent |
| $I$ | Mean excitation potential of the medium |
| $k_{Q,Q'0}$ | Absorbed dose calibration conversion factor |
| $L_\Delta$ | Linear energy transfer (LET) |
| $m$ | Mass |
| $m_0$ | Rest mass |
| $M_Q$ | Corrected collected charge for beam calibration |
| $n$ | Stability index |
| $N_A$ | Avogadro's number |
| $N_{D,W}^{Q_0}$ | Absorbed dose calibration factor in water for beam quality $Q$ |
| $n_e$ | Electron density (electrons/g) |
| $\vec{p}$ | Momentum |
| $q$ | Charge |
| $R$ | Radius of orbit |
| $r$ | Radial coordinate |
| $r'$ | Perturbation in $r$ |
| $r_0$ | Equilibrium value of $r$ |
| $R_p$ | Practical range |
| $R_{\text{res}}$ | Residual range |
| $S_c$ | Mass collision stopping power |

| | |
|---|---|
| $t$ | Time |
| $T_0$ | Initial energy of charged particle |
| $T_{max}$ | Maximum kinetic energy |
| $U$ | Use factor |
| $v$ | Speed |
| $V$ | Potential difference |
| $V_0$ | Maximum potential difference |
| $W$ | Workload |
| $Y$ | Particle type |
| $Z$ | Atomic number of the medium |
| $z$ | Atomic number of bombarding particle |
| $z$ | Vertical coordinate |
| $z'$ | Perturbation in $z$ |
| $z_0$ | Equilibrium value of $z$ |
| $z_{ref}$ | Reference depth |
| $(dT/dx)_{Y,T,Z}$ | Stopping power |
| $(dT/dx)_{\Delta}$ | Restricted stopping power |
| $\mathfrak{R}_{CSDA}$ | CSDA range |
| $\beta$ | $v/c$ |
| $\gamma$ | Relativistic gamma |
| $\Delta$ | Cutoff energy |
| $\theta$ | Angular position |
| $\theta'$ | Angular position perturbation |
| $\nu_0$ | Cyclotron frequency |
| $\rho$ | Mass density |
| $\Phi$ | Particle fluence |
| $\varphi$ | Phase difference |
| $\Phi_T$ | Differential energy particle fluence spectrum |
| $\omega_0$ | Cyclotron angular frequency = $2\pi\nu_0$ |
| $\Omega_r$ | Radial oscillation frequency |
| $\omega_{rf}$ | Angular frequency of applied voltage |
| $\Omega_z$ | Betatron oscillation frequency |

# REFERENCES


Attix, F.H. 1986. *Introduction to Radiological Physics and Radiation Dosimetry.* New York: Wiley.

Bloch, C., P. Hill, K.L. Chen, A. Saito, and E.E. Klein. 2012. Startup of the Kling Center for Proton Therapy [abstract]. In *22nd International Conference on the Application of Accelerators in Research and Industry*, Fort Worth, TX, p. 150.

British Institute of Radiology. 1996. Central axis depth dose data for use in radiotherapy, 1996. *Br. J. Radiol.*, Suppl. 25.

DeLaney, T.F. and H.M. Kooy. 2007. *Proton and Charged Particle Radiotherapy*. Philadelphia: Lippincott, Williams and Wilkins.
Goitein, M. 2008. *Radiation Oncology: A Physicists-Eye View.* New York: Springer.
Goitein, M., A.J. Lomax, and E.S. Pedroni. 2002. Treating cancer with protons. *Physics Today*, September, p. 45.
Hall, E.J. and A.J. Giaccia. 2012. *Radiobiology for the Radiologist*. Philadelphia: Wolters Kluwer.
Hitachi. 2014. Proton beam therapy. Hitachi. http://www.hitachi-america.us/products/business/protonbeam/.
Hong, L., M. Goitein, M. Bucciolini, R. Comiskey, B. Gottschalk, S. Rosenthal, C. Serago, and M. Urie. 1996. A pencil beam algorithm for proton dose calculations. *Phys. Med. Biol.* 41, 1305–1330.
Hoppe, B.S., S. Flampouri, R. Zaiden, W. Slayton, E. Sandler, S. Ozdemir, N.H. Dang, et al. 2014. Involved-node proton therapy in combined modality therapy for Hodgkin lymphoma: Results of a phase 2 study. *Int. J. Radiat. Oncol. Biol. Phys.* 89(5), 1053–59.
Humphries, S. 2012. *Principles of Charged Particle Acceleration*. Mineola, NY: Dover.
International Atomic Energy Agency (IAEA). 2006. Absorbed dose determination in external beam radiotherapy: An international code of practice for dosimetry based on standards of absorbed dose to water. Technical Report Series 398. In *Code of Practice for Proton Beams*, chap. 10. Vienna: IAEA.
International Commission on Radiation Units and Measurements (ICRU). 2007. Prescribing, recording and reporting proton beam therapy. ICRU Report No. 78. *J. ICRU* 7(2).
Jongen, Y. 2008. Production of particle beams: Cyclotron. Presented at Particle Therapy Cooperative Group (PTCOG 47) Educational Workshop, Jacksonville, FL.
Jongen, Y. 2010. Review on cyclotrons for cancer therapy. In *Proceedings of CYCLOTRONS 2010*, Lanzhou, China, p. 398.
Kleeven, W., M. Abs, E. Forton, S. Henrotin, Y. Jongen, V. Nuttens, Y. Paradis, E. Pearson, S. Quets, J. Van de Walle, P. Verbruggen, S. Zaremba, M. Conjat, J. Wang, J. Flanz, and R.W. Hamm. 2011. Injection study of the ProTom 330 synchrotron with a 1.6 MeV RFQ linac. Presented at 19th Particles and Nuclei International Conference of MIT Bates, Boston, MA. http://web.mit.edu/panic11/talks/tuesday/PARALLEL-3H/4-1430/wang/317-0-panic11-talk317.pdf (accessed September 2014).
Kleeven, W., M. Abs, E. Forton, S. Henrotin, Y. Jongen, V. Nuttens, Y. Paradis, et al.; AIMA Development, Nice, France. 2013. *Proceedings of Cyclotrons 2013*, Vancouver, BC.
Klein, E.E. 2014. Slides, PROtom details. https://www.aapm.org/meetings/09PRS/documents/KleinV3.pdf.
Klein, E.E., C. Bloch, J. Pierburg, and J. Bardley. 2012. Single room proton radiotherapy. Poster at ASTRO 2012, Boston, MA. astro2012.abstractsnet.com/handouts/pdfs/011758.pdf.
Linz, U., ed. 2012. *Ion Beam Therapy: Fundamentals, Technology, Clinical Applications*. New York: Springer.
Lomax, A. 2012. Physics of treatment planning using scanned beams. In *Proton Therapy Physics*, ed. H. Paganetti, 335–80. Boca Raton, FL: CRC Press.
Lu, H. and J. Flanz. 2012. Characteristics of clinical proton beams. In *Proton Therapy Physics*, ed. H. Paganetti. Boca Raton, FL: CRC Press, pp. 103–124.
Ma, C.-M. and T. Lomax. 2012. *Proton and Carbon Ion Therapy*. Boca Raton, FL: CRC Press.
Maughan, R.L. and F. Van den Heuvel. 2008. Within the next 10–15 years protons will likely replace photons as the most common type of radiation for curative radiotherapy [point/counterpoint]. *Med. Phys.* 35(10), 4285.
McDermott, P. and C. Orton. 2010. *The Physics and Technology of Radiation Therapy*. Madison, WI: Medical Physics Publishing.
McDonald, M.W. and M.M. Fitzek. 2010. Proton therapy. *Curr. Prob. Cancer* 34, 257–296.

Metz, J.M. and C.R. Thomas, eds. 2010. Proton therapy. *Radiat. Med. Rounds* 1(3).
Moyers, M.F. 1999. Proton therapy. In *The Modern Technology of Radiation Oncology*, vol. I, ed. J. VanDyk, 823–69. Madison, WI: Medical Physics Publishing.
Mukherjee, B. 2012. Radiation safety issues relevant to proton therapy and radioisotope production medical cyclotrons. *Radiat. Prot. Environ.* 35(3–4), 126–34.
Nafziger, B. 2011. Proton therapy: Can you afford it? *DOTmedbusiness News*. www.dotmed.com.
National Association for Proton Therapy (NAPT). 2015. Number of proton centers. Washington, DC: NAPT. http://www.proton-therapy.org/map.htm.
National Council on Radiation Protection and Measurements (NCRP). 2003. Radiation protection for particle accelerator facilities. Report No. 144. Bethesda, MD: NCRP.
National Institute of Standards and Technology (NIST). Stopping-power and range tables for protons. Gaithersburg, MD: NIST. http://physics.nist.gov/PhysRefData/Star/Text/PSTAR.html (accessed September 2014).
Paganetti, H. 2012a. Late effects from scattered and secondary radiation. In *Proton Therapy Physics*, ed. H. Paganetti, 555–92. Boca Raton, FL: CRC Press.
Paganetti, H., ed. 2012b. *Proton Therapy Physics.* Boca Raton, FL: CRC Press.
Palta, J.R., D. Yeung, and R. Slopsema. 2009. Proton therapy treatment planning: Challenges and solutions. American Association of Physicists in Medicine Presentation, Anaheim, CA. http://www.aapm.org/meetings/amos2/pdf/42-11867-64697-834.pdf. Accessed February 2016.
Particle Therapy Cooperative Group (PTCOG). 2014. http://www.ptcog.ch/ (accessed July 2014).
Röcken, H., M. Abdel-Bary, E. Akcöltekin, P. Budz, T. Stephani, and J. Wittschen. 2010. The Varian 250 MeV superconducting compact proton cyclotron: Medical operation of the 2nd machine, production and commissioning status of machines no. 3 to 7. Presented at Proceedings of CYCLOTRONS 2010, Lanzhou, China.
Schaffner, B. and E. Pedroni. 1998. The precision of proton range calculations in proton radiotherapy treatment planning: Experimental verification of the relation between CT-HU and proton stopping power. *Phys. Med. Biol.* 43(6) 1579–92.
Scharf, W.H. 1994. *Biomedical Particle Accelerators*. New York: AIP Press.
Schippers, M. 2012. Proton accelerators. In *Proton Therapy Physics*, ed. H. Paganetti, 61–102. Boca Raton, FL: CRC Press.
Schultz-Ertner, D. and H. Tsujii. 2007. Particle radiation therapy using proton and heavier ion beams. *J. Clin. Oncol*. 25, 953–64.
Smith, A. 2009a. Proton physics and technology. AAPM 51st Annual Meeting, Anaheim, CA. http://www.aapm.org/meetings/2010AM/documents/SAMsSlideSetAAPM_2009_AlSmith_000.pdf. Accessed February 2016.
Smith, A.R. 2009b. Vision 20:20 proton therapy. *Med. Phys.* 36(2), 556.
Smith, A.R. 2006. Proton therapy. *Phys. Med. Biol*. 51, R494–504.
Thornton, S.T. and J.B. Marion. 2003. *Classical Dynamics of Particles and Systems*. Independence, KY: Cengage Learning.
Vatnitsky, S.M. and M.F. Moyers. 2013. Radiation therapy with light ions. In *The Modern Technology of Radiation Oncology*, vol. III, ed. J. VanDyk, 183–222. Madison, WI: Medical Physics Publishing.
Wang, F., J. Flanz, and R. Hamm. 2011. Injection study of the ProTom-Radiance 330 synchrotron with a 1.6 MeV RFQ linac. The 19th Particles and Nuclei International Conference, Cambridge, MA. Slide presentation.
Wikipedia. 2014a. Mattauch isobar rule. https://en.wikipedia.org/wiki/Mattauch_isobar_rule.
Wikipedia. 2014b. Beam emittance. https://en.wikipedia.org/wiki/Beam_emittance.
Wilson, R.R. 1946. Radiological use of fast protons. *Radiology* 47, 487–91.

Yajnik, S. 2012. *Proton Beam Therapy: How Protons Are Revolutionizing Cancer Treatment*. New York: Springer.

## ENDNOTES

1. It is likely that protons were being accelerated at the Cavendish laboratory in England prior to the development of the cyclotron.
2. Terminology can be a little confusing: although all ions are heavy charged particles in comparison to electrons, the ICRU defines "light ions" as nuclei with an atomic number less than or equal to 10 (neon).
3. The Mattauch isobar rule states that if two adjacent elements in the periodic table have isotopes that are isobars, at least one of the isobars must be radioactive (Wikipedia, 2014a).
4. This dependence clearly cannot continue down to $v \to 0$ (see Figure 2.5).
5. The time necessary to deliver a particular treatment depends on additional factors besides the beam current. See Section 2.9.
6. The neutron production rate listed in Mukherjee (2012, p. 128) is 13 μSv/h, but the value in figure 7 is listed as 13 mSv/h/nA. We assume that 13 mSv/h/nA is the correct value.

# 3

# CONVOLUTION/ SUPERPOSITION DOSE COMPUTATION ALGORITHMS

## 3.1 INTRODUCTION

There are three basic approaches to photon dose calculation: primary, correction based, and model based. Primary calculation methods rely on a direct computation of the dose from first principles. Examples are Monte Carlo and the Boltzmann transport equation (see Chapter 4). The correction approach is heavily based on measurements made in a water phantom, which are then "corrected" for patient-specific factors. Traditional dose calculation methods, involving percent depth dose, tissue air ratio, and tissue maximum ratio, are of this type. These methods are poorly suited for intensity-modulated radiation therapy (IMRT). IMRT dose calculations require the summation of a large number of small fields. The penumbral dose in these fields must be calculated very accurately in order that the sums be accurate. Correction-based algorithms are not capable of such accurate dose computation. We will not discuss these methods further here, as they have fallen out of use (at least in commercial-based treatment planning systems). The model-based approach involves a secondary computation of dose based on precomputed "kernels." This method explicitly takes into account beam characteristics such as energy, geometry, and presence of beam modifiers. Figure 3.1 shows the various contributions to the dose that must be considered in order to calculate an accurate dose distribution. This illustrates the complexity of the numerous physical processes that must be evaluated. Review articles pertaining to photon beam dose calculation methods can be found in Ahnesjö and Aspradakis (1999), Mackie et al. (1996, 2000, 2007), and Bloch and Altschuler (1995). The most comprehensive of these is the paper by Ahnesjö and Aspradakis (1999).

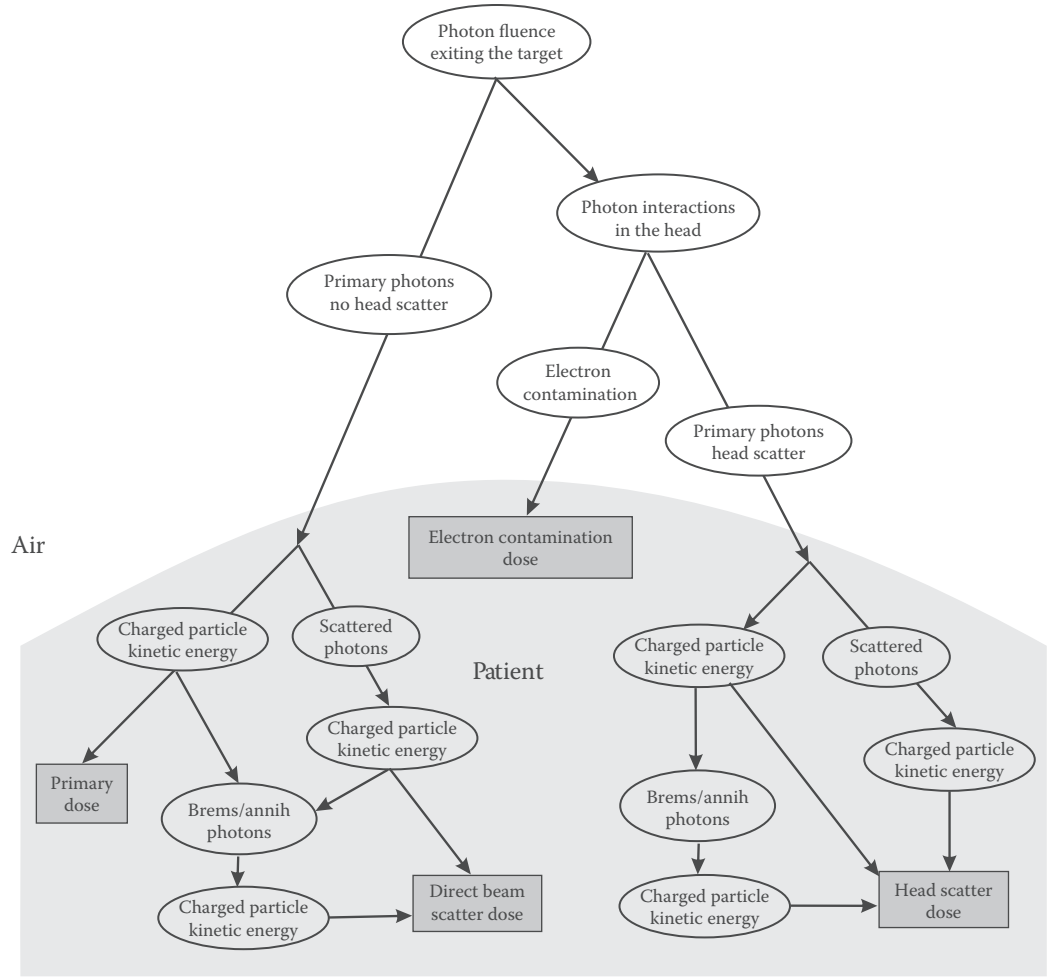



FIGURE 3.1 The contributions to the absorbed dose that must be calculated to obtain an accurate dose distribution. This shows the complexity of the various physical processes that must be considered. There are two sources of primary photons incident upon the patient: photons that come directly from the target without scattering and photons that are scattered in the linac head. Photon scattering in the head contributes to electron contamination of the beam. The "brems/annih" component refers to bremsstrahlung and annihilation photons. (Adapted from Ahnesjö, A. and M.M. Aspradakis, *Phys. Med. Biol.* 44, R99–155, 1999.)

Convolution/superposition (C/S) methods are model-based algorithms. Advantages of these techniques are that they lend themselves well to three-dimensional (3D) calculations, they can handle electronic disequilibrium (to some degree), and they are easily adaptable for intensity modulation techniques. Commercial treatment planning systems commonly offer more than one dose calculation algorithm, and a C/S algorithm is frequently one of the choices. As of this writing, commercial treatment planning systems offering superposition algorithms include Pinnacle[3] (Philips Healthcare), XiO (Elekta), Lexsell GammaPlan® 10 (Elekta), RayStation (RaySearch

Laboratories), DAO MRT (Prowess, Inc.), Eclipse (Varian Medical Systems), iPlan RT Dose 4.5 (Brainlab AG), and Mobius3D (Mobius Medical Systems).[1]

C/S dose calculations consist of four main steps:

1. Model the energy fluence originating in the head of the linac.
2. Represent the energy imparted to the medium by primary photons (terma; see Section 3.2).
3. Use the precomputed energy deposited about primary photon interaction sites to calculate the dose deposited by the incident photons (kernel; see Section 3.2).
4. The addition of electron contamination to more accurately model the dose in the buildup region.

We will consider each of these steps in some detail, including complications associated with realistic linac-generated radiation beams and with real finite heterogeneous patients. Among the complexities that must be dealt with are those in a long list given by Mackie et al. (1996): finite source size, angular distribution of photons, primary transmission, "extrafocal" radiation originating from the flattening filter and primary collimator, differential beam hardening with lateral distance in the flattening filter, curved multi-leaf collimator (MLC) leaf ends, leaf configuration, tongue-and-groove effect, leaf transmission, electron contamination, and tissue heterogeneities. This is a daunting list.

We will primarily use continuous variables throughout the description here. The reader understands that in practice, numerical calculations require discrete variables and integrals need to be turned into sums.

All photons that are incident on the surface of the phantom or patient are called primary, whether they have been scattered in the head or not.

In Section 3.2, we handle the simple case of monoenergetic beams and a homogeneous medium. In Section 3.3, we deal with a strategy for homogeneous media that allows a significant reduction in the CPU time to perform a dose calculation by taking advantage of the convolution theorem. In Section 3.4, we add the complication of a polyenergetic beam. In Section 3.5, we discuss beam modeling and primary photon transport in the medium. A description of the characteristics and calculation of point dose kernels is dealt with in Section 3.6. An analytical expression for the point dose kernel due to singly Compton scattered photons is derived in Section 3.7. The complication of a heterogeneous medium is considered in Section 3.8. In Section 3.9, we examine another strategy to speed computation, the use of pencil beam kernels instead of point dose kernels. In Section 3.10, we look at effects related to patient geometry. Section 3.11 is a discussion of the widely used collapsed cone convolution (CCC) approximation, which is another method for decreasing dose calculation time for point dose kernel calculations. A method is needed for the calculation of monitor units (MUs), and this is covered in Section 3.12. Dose calculation speed is considered in Section 3.13. There has been no attempt here to provide a detailed comparison between C/S algorithms and other algorithms or to assess their accuracy. There is ample literature on this topic, and it speaks for itself.

## 3.2 MONOENERGETIC BEAMS, HOMOGENEOUS MEDIUM

For simplicity, we begin by considering a monoenergetic beam of photons in a homogeneous medium. In charged particle equilibrium, the dose at a point in the medium can be written

$$D(\vec{r}) \overset{CPE}{=} K_c(\vec{r}) = \frac{\mu_{en}}{\rho} \Psi_T(\vec{r}), \tag{3.1}$$

where:

$K_c$ is the collision kerma
$\mu_{en}/\rho$ is the mass energy absorption coefficient
$\psi_T$ is the *total* energy fluence for all photons in the medium (Attix, 1986)

The energy fluence in Equation 3.1 includes both primary and scattered photons. Primary photons are those that have not scattered in the medium, although they may have been scattered in the linac head. Equation 3.1 assumes charged particle equilibrium; it therefore does not take account of the finite forward range of charged particles. As a first attempt to remedy this, assume the mean distance that charged particles travel downstream is $a$; then

$$D(\vec{r}) = \frac{\mu_{en}}{\rho} \Psi(\vec{r} - \vec{a}) = K_c(\vec{r}) e^{+\mu'|a|} = K_c(\vec{r})\beta, \tag{3.2}$$

where:

$\mu'$ is the slope of $K_c$ in the downstream direction at $\vec{r}$
$\vec{a}$ is a vector pointing downstream into the medium with a magnitude $a$

Equation 3.2 represents the condition known as transient charged particle equilibrium (TCPE) (see Attix, 1986). The quantity $\beta$ is slightly larger than 1.00. TCPE assumes that the dose at a point arises from photon interactions at some other discrete location. A simple extension of this idea is to allow multiple sites for interaction of photons that generate charged particles. This suggests an integral over the neighborhood of $\vec{r}$:

$$D(\vec{r}) = \int_V K_c(\vec{r}\,') A_c(\vec{r} - \vec{r}\,') d^3r', \tag{3.3}$$

where $A_c$ is a function that gives the contribution of charged particle energy that gets absorbed per unit volume at $\vec{r}$ from all photon interactions at $\vec{r}\,'$ (Figure 3.2). This function is called the kernel. Other commonly used names for the kernel are dose spread array, point spread function, or energy deposition function. Note that if $A_c = \delta(\vec{r} - \vec{r}\,')$ (an "on-the-spot" approximation), where $\delta$ is the Dirac delta function, then $D(\vec{r}) = K_c(\vec{r})$. The difficulty with the use of Equation 3.3 is that $K_c(\vec{r}\,')$ is due

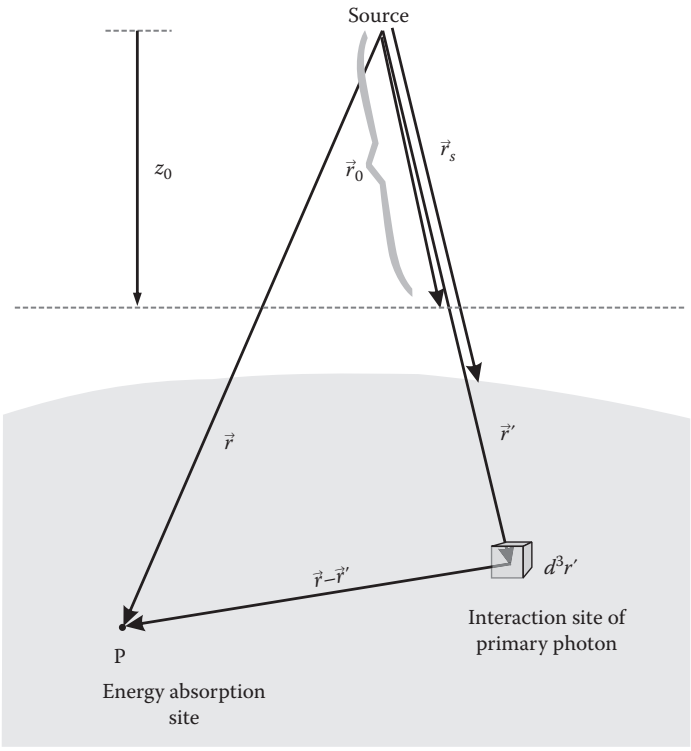



FIGURE 3.2 The geometry of the C/S point kernel method of dose calculation for point P located at $\vec{r}$. Some of the energy liberated by photons interacting at $\vec{r}'$ is absorbed at point P. $\vec{r}_s$ is a vector from the source of the radiation to the surface of the medium. To find the total dose at point *P*, one must integrate over all values of $\vec{r}'$. To find the dose at another point, the integration over all values of $\vec{r}'$ must be repeated.

to interactions of both primary and scattered photons at $\vec{r}'$ (see Equation 3.1), and it is therefore troublesome to compute (even if $A_c$ could be easily determined). The reason is that we have no easy way of knowing the total fluence $\psi_T$ in Equation 3.1. The primary fluence, consisting of those photons that have not been scattered in the medium, is, however, more easily determined, as we will see below. A useful strategy is therefore to replace $K_c$ in Equation 3.3 with a quantity directly related to the primary fluence and to include the effect of scattered photons in the kernel. The kernel will no longer be finite in extent because it now includes scattered photons.

A "superposition" equation that separates the primary photon transport and includes a kernel that accounts for scattered photons and electrons set in motion throughout the volume is

$$D(\vec{r}) = \int \frac{\mu}{\rho} \Psi(\vec{r}') A(\vec{r}, \vec{r}') d^3 r'$$

$$= \int T(\vec{r}') A(\vec{r}, \vec{r}') d^3 r', \tag{3.4}$$

where:

$\psi(\vec{r}')$ is the primary photon energy fluence *only*

$A(\vec{r},\vec{r}')$ is a kernel that now includes the contribution from scattered photons

Figure 3.2 illustrates the vector geometry used in the dose computation. Equation 3.4 represents a superposition of weighted responses to point irradiations.

The quantity

$$T(\vec{r}') = \left(\frac{\mu}{\rho}\right)\Psi(\vec{r}'), \tag{3.5}$$

is called the terma: total energy released in matter. It is the total energy released by *primary* photon interactions per unit mass; it includes the energy imparted to both charged particles (this part is the kerma) and scattered photons. In Equation 3.5, $\psi$ is the energy fluence of primary photons *only and does not include photons scattered in the medium*. Note that the kerma is equal to $(\mu_{en}/\rho)\Psi_T$. The terma is the energy lost or removed from the primary beam per unit mass.

It is important to emphasize that the kernel for a particular beam need only be computed once. Once the functional form or a table of numbers is in hand for the kernel, the dose distribution can be found in any phantom (or patient) provided that the primary energy fluence can be determined at all points within that phantom. The kernel as defined implicitly in Equation 3.4 is a point dose kernel. In Section 3.9, we consider a pencil beam kernel. The point dose kernel is the fraction of the energy of the primary photons interacting at $\vec{r}'$, which is absorbed at $\vec{r}$, per unit volume at $\vec{r}$. The units are therefore those of inverse volume. There are variations in the definition of the kernel in the literature (Table 3.1).

**TABLE 3.1 Selected Point Kernel References**

| Reference | Name of Kernel | Units | Comments |
|---|---|---|---|
| Mackie et al. (1988) | "Energy deposition kernel" | cGy MeV$^{-1}$ photon$^{-1}$ | May be the kernel used in Equation 3.4 divided by $\rho$ |
| Mackie et al. (1985) | "Primary dose spread arrays" | None | Dose deposited in the voxel normalized to the collision kerma in the interaction voxel |
| Ahnesjö et al. (1987) | Point spread function | cm$^{-3}$ | Appears to be the same kernel as used in Equation 3.4 |
| Ahnesjö (1989) | Point spread function | cm$^{-3}$ | Appears to be the same kernel as used in Equation 3.4 |
| Huang et al. (2013) | "Energy deposition kernels" | MeV cm$^2$ g$^{-1}$ photon$^{-1}$ | "Dose × $r^2$" |
| Boyer and Mok (1985) | "Interaction kernel" | "eV/g per photon interaction at the origin" | |

The kernel for a (infinite) homogeneous medium is a function only of the distance between the points at $\vec{r}$ and $\vec{r}'$ and may therefore be written as $A(\vec{r},\vec{r}') = A(\vec{r}-\vec{r}')$. Thus, in this case, Equation 3.4 may be written

$$D(\vec{r}) = \int_V T(\vec{r}')A(\vec{r}-\vec{r}')d^3r'. \tag{3.6}$$

Figure 3.3 illustrates the application of Equation 3.6 to the calculation of absorbed dose.

For the simple case in which the incident energy fluence originates from a point source, the primary energy fluence in the medium is given by

$$\Psi(\vec{r}') = \Psi(\vec{r}_0)\left(\frac{r_0}{r'}\right)^2 e^{-\mu(r'-r_s)}, \tag{3.7}$$

where $\Psi(\vec{r}_0)$ is the primary energy fluence in air on a plane at some standard vertical distance (e.g., $z_0$, the plane through the isocenter; Figure 3.2). The quantity $\vec{r}_s$ is a vector from the origin to the surface of the medium (Figure 3.2). The magnitude of this vector may vary from point to point on the surface depending on the off-axis distance and the contour of the surface of the medium. The dose may now be written as

$$D(\vec{r}) = \int_V \frac{\mu}{\rho}\Psi(\vec{r}_0)\left(\frac{r_0}{r'}\right)^2 e^{-\mu(r'-r_s)}A(\vec{r}-\vec{r}')d^3r'. \tag{3.8}$$

The quantity $\Psi(\vec{r}_0)$ must remain inside the integral because the energy fluence will vary with off-axis distance. We postpone a detailed discussion of the kernel until Section 3.6.

Let us consider the computation required to numerically evaluate the dose in a medium for a single beam using Equation 3.8. The integral must be turned into a sum. In order to compute this sum, we divide the medium into $N^3$ voxels. To calculate the dose in any one voxel, we must sum over the contributions from all $N^3$ voxels.

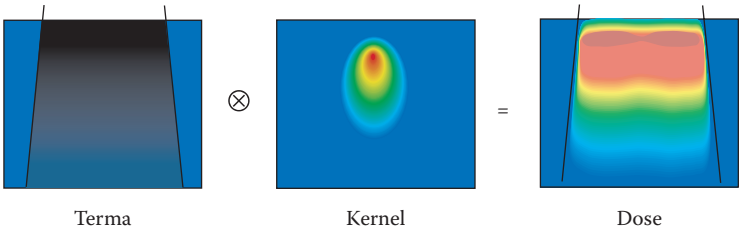



FIGURE 3.3 Illustration of the idea of the point kernel C/S method of dose calculation. The terma is first calculated throughout the phantom (or patient). This is then convolved or superposed with the kernel (see Equation 3.6) to calculate the dose distribution.

Therefore, to calculate the dose throughout the entire medium, we must perform a sum of $N^6$ terms. The number of voxels can be as large as $10^7$. In this case, the dose computation for a single beam requires a sum of $10^{14}$ numbers. Dose computation speed will be considered further in Section 3.8. To reduce the computational intensity of the problem, it is advisable to search for strategies to reduce the number of floating-point operations necessary. Some of these strategies involve approximations such as CCC or pencil beam convolution. These will be considered later. Another method involves application of the convolution theorem, to which we now turn.

## 3.3 CONVOLUTION INTEGRALS

The form of Equation 3.6 lends itself to an evaluation method that exploits Fourier transforms of convolution integrals. The general form of a convolution integral is

$$f \otimes g = \frac{1}{\sqrt{2\pi}} \int g(\vec{r}\,') f(\vec{r} - \vec{r}\,') d^3 r'. \tag{3.9}$$

This integral is a function of $\vec{r}$. The reader is likely to have encountered convolution integrals before, without perhaps realizing it. Consider Poisson's equation for the electrostatic potential, $\nabla^2 \psi = -\rho/\varepsilon_0$, well known to all physicists. The solution to this is

$$\psi(\vec{r}) = \frac{1}{4\pi\varepsilon_0} \int \frac{\rho(\vec{r}\,')}{|\vec{r} - \vec{r}\,'|} d^3 r'. \tag{3.10}$$

Equation 3.10 is simply a convolution of the charge distribution and a weighting function $\left(4\pi\varepsilon_0 |\vec{r} - \vec{r}\,'|\right)^{-1}$, which is analogous to the kernel in the dose computation problem.

A true convolution integral must have the form of Equation 3.9. The function $f$ is spatially invariant. Convolution requires spatial invariance of the kernel. In an infinite homogeneous medium, the pattern of energy absorption around a point of interaction is spatially invariant. If $A(\vec{r}, \vec{r}\,')$ is of the form $A(\vec{r} - \vec{r}\,')$ then the dose integral will have the form of a true convolution integral. Realistic dose integrals are not true convolution integrals, as we will see. The kernel is spatially invariant if it depends only on the relative geometric relationship between the interaction and dose deposition site.

The advantage of writing a dose integral in terms of a convolution is that computation time for the integral in Equation 3.6 can be significantly reduced by using Fourier transforms. If the kernels are spatially invariant, then the convolution theorem applies (Arfken et al., 2013). This theorem states that given two functions $f(\vec{r})$ and $g(\vec{r})$ with Fourier transforms $F(k)$ and $G(k)$,

$$f \otimes g = \Im^{-1}\{F \times G\}, \tag{3.11}$$

where $\mathfrak{F}^{-1}$ is the inverse Fourier transform. The Fourier transforms $F$ and $G$ will be a set of complex numbers. These numbers are to be multiplied together, and then the inverse Fourier transform is taken. One can exploit the speed offered by fast Fourier transform (FFT) numerical techniques.

With the use of Fourier transforms, the order of the number of operations required is considerably reduced from $N^6$ to $N^3 \log_2 N$ (Ahnesjö and Aspradakis, 1999).

All of this is conditional on the dose integral having the form of a true convolution integral. Kernels are not invariant in real patients because inhomogeneities are present (see Section 3.8). If a homogeneous dose calculation is acceptable, then a considerable speed advantage becomes available.

## 3.4 POLYENERGETIC BEAMS, HOMOGENEOUS MEDIUM

In this section, we add the complication of a polyenergetic beam, but the medium will remain homogeneous for now. Clinical radiation beams are polyenergetic with the exception of Co-60 radiation. The effects of a beam spectrum must be accounted for in the calculation of both the terma and the kernel. In addition, we must consider the complications of the change in energy with depth (depth hardening) of the primary radiation and off-axis energy softening due to the flattening filter.

We begin with a discussion of the terma by generalizing the treatment in Section 3.2. We introduce the differential energy fluence for the primary radiation. The energy fluence for photons having energies between $E$ and $E + dE$ is

$$d\Psi = \Psi_E(\vec{r}', E)dE, \tag{3.12}$$

where $\Psi_E$ is the differential energy fluence. This quantity describes the energy spectrum of the beam. The total energy fluence is

$$\Psi(\vec{r}') = \int_0^\infty \Psi_E(\vec{r}', E)dE. \tag{3.13}$$

The differential energy fluence is related to the differential particle fluence $\Phi_E$:

$$\Psi_E = E\Phi_E. \tag{3.14}$$

The average beam energy is given by

$$\bar{E}(\vec{r}') = \frac{1}{\Phi(\vec{r}')}\int_0^\infty \Psi_E(\vec{r}')dE, \tag{3.15}$$

where Φ is the total particle fluence:

$$\Phi = \int_0^\infty \Phi_E dE. \tag{3.16}$$

We may define an energy differential terma as follows:

$$T_E(\vec{r}', E) = \frac{\mu(E)}{\rho} \Psi_E(\vec{r}', E). \tag{3.17}$$

For a polyenergetic beam, each component of the differential energy fluence at the surface of the medium must be exponentially attenuated in the medium. We need to propagate or transport the differential terma throughout the irradiated medium as follows (assuming an ideal point source):

$$\Psi_E(\vec{r}', E) = \Psi_E(\vec{r}_0, E)\left(\frac{r_0}{r'}\right)^2 e^{-\mu(r'-r_s)}, \tag{3.18}$$

where μ depends on *E*.

There are a number of methods to account for the spectrum of energies in a clinical photon beam. Three of these are

1. Calculate a separate kernel for each energy bin (sometimes called the components method).
2. Calculate an energy-dependent kernel for a generic published spectrum (as in Equation 3.29). In this case, the kernel will only be valid for that specific spectrum.
3. Calculate an energy-averaged kernel over the user's beam entrance spectrum by using the results of number 1 and the user's beam spectrum.

In the first case (number 1), the expression for the dose may be written as

$$\begin{aligned} D(\vec{r}) &= \int_E \int_V T_E(\vec{r}') A(\vec{r}, \vec{r}'; E) d^3 r' dE \\ &= \int_E \int_V \frac{\mu}{\rho} \Psi_E(\vec{r}_0, E)\left(\frac{r_0}{r'}\right)^2 e^{-\mu(r'-r_s)} A(\vec{r}, \vec{r}'; E) d^3 r' dE, \end{aligned} \tag{3.19}$$

where *A* is the kernel for energy interval *E* to *E* + *dE*.

Evaluation of Equation 3.19 for the dose requires repeat spatial integration for each energy bin and is thus very computationally intensive. To determine the dose *at a single point*, not only do we have to integrate over every other point in the medium, but we also must repeat these integrations over the energy spectrum of

the incident radiation (which may change with depth). Using the average energy doesn't work well (Ahnesjö and Aspradakis, 1999). Fortunately, the number of energy bins for accurate numerical calculations may not be large. For example, Boyer et al. (1989) used five energy bins to model an 18 MV beam. These authors derived a kernel for each of five energies. The amplitude of the bin contribution is weighted by the spectrum of the beam.

The difficulty with approach 2 is that the spectrum depends on depth and lateral position. Furthermore, the beam spectrum varies from linac to linac.

The third approach has been followed by Papanikolaou et al. (1993). In this method, the dose is written as

$$D(\vec{r}) = \int_V \frac{\bar{\mu}(\vec{r}')}{\rho} \Psi(\vec{r}') \bar{A}(\vec{r}, \vec{r}') d^3 r', \tag{3.20}$$

where

$$\frac{\bar{\mu}(\vec{r}')}{\rho} = \frac{1}{\Psi(\vec{r}')} \int_E \frac{\mu(\vec{r}')}{\rho} \Psi_E(\vec{r}', E) dE, \tag{3.21}$$

and

$$\Psi(\vec{r}') = \left( \frac{r_0}{r'} \right)^2 \int_E \Psi_E(\vec{r}_0) e^{-\mu(r'-r_s)} dE, \tag{3.22}$$

and

$$\bar{A}(\vec{r}, \vec{r}') = \frac{\int_E \frac{\mu(\vec{r}')}{\rho} \Psi_E(\vec{r}') A(\vec{r}, \vec{r}'; E) dE}{\Psi(\vec{r}') \bar{\mu}(\vec{r}')/\rho}. \tag{3.23}$$

Equation 3.20 may be verified by substituting Equations 3.21 through 3.23 into Equation 3.20, thereby recovering Equation 3.19. As it stands, Equation 3.20 does not change the computational complexity of the problem because it still requires evaluation of the energy integral in Equation 3.23 over every point in the medium. If we integrate the energy-dependent kernel in Equation 3.23 only over the beam spectrum at $\vec{r}_0$, however, extra integrations are avoided. This requires the approximation that

$$\bar{A}(\vec{r}, \vec{r}') \approx \frac{\int_E \frac{\mu}{\rho} \Psi_E(\vec{r}_0) A(\vec{r}, \vec{r}'; E) dE}{\Psi(\vec{r}_0) \bar{\mu}/\rho}. \tag{3.24}$$

The kernel in Equation 3.24 can be computed in advance, stored, and then used for every beam calculation. This is the method used in the Pinnacle treatment planning system, wherein 15–20 discrete energies (Starkschall et al., 2000) are used for the integration.

## 3.5 INCIDENT ENERGY FLUENCE, BEAM MODELING, AND PRIMARY PHOTON TRANSPORT

The problem of characterizing the spectrum of radiation incident upon the patient can be separated from a calculation of the terma and the kernel. In fact, it is necessary to address this issue regardless of the method of dose computation. The energy fluence incident on the patient surface is used to calculate the energy fluence throughout the patient as in Equation 3.7. Most of the complexity of the convolution method involves computing this energy fluence. Mackie et al. (1996) have listed the features that need to be modeled:

1. Finite source size
2. Extrafocal radiation
3. Beam spectrum, including the change in the spectrum with off-axis position
4. Beam intensity variation across the field (e.g., horns)
5. Electron contamination
6. Transmission through collimator jaws
7. Scatter outside the field
8. MLC rounded leaf ends, tongue-and-groove effect, and MLC transmission
9. Wedges and compensators, including beam hardening

Points 1 and 2 are related. Extrafocal radiation is that radiation scattered from the primary collimator and flattening filter and not originating directly from the target (see Figures 3.1 and 3.5). The magnitude of the extrafocal radiation accounts for the variation in output with field size. This radiation originates downstream or distal to the radiation produced directly by the bremsstrahlung mechanism in the target. The energy fluence due to extrafocal radiation can amount to as much as 15% of the total, and therefore it must be modeled. Of this, 8%–12% (for large field sizes) originates as scattered photons in the flattening filter, and 3%–5% results from scattering off the primary collimator (Mackie et al., 2007). The presence of this radiation manifests itself as an extended source size. The native target source size of the region over which bremsstrahlung emission occurs in the target is only a few millimeters. It is a function of the size of the cross section of the electron beam striking the target and the scattering of these electrons in the target as they lose energy.

The collimator plus block field outline is modeled with a mask function. For the primary collimator, the mask function inside the field is 1.0, and outside, it is equal to the

primary collimator transmission. The mask function alone cannot model penumbral blurring of the field boundary. The finite source size and extrafocal radiation are modeled by convolving the mask function with a Gaussian aperture function that describes the width of the blurring (Mackie et al., 2007). Wedges cannot be accurately modeled with primary attenuation only because of scatter in the wedge (which leads to a field size dependence) and beam hardening. The incident energy fluence is altered by beam modifiers, wedges, blocks, MLC, and so forth. The MLC raises what Ahnesjö (2006) has aptly described as a multitude of "fine-print issues" that we will not address here.

In the Philips Pinnacle software, the beam energy spectrum is determined from measured depth dose data (McNutt, 2002). The software starts with a published spectrum and uses this to calculate depth–dose curves that are then compared to measurements. The initial spectrum is then adjusted until the calculated and measured depth–dose curves agree as closely as possible. Figure 3.4 shows spectra calculated by this method for an Elekta Agility linac with beam energies of 6, 10, and 18 MV.

Due to beam hardening on the central axis relative to points off axis, the beam energy spectrum varies across the beam because of the presence of the flattening filter (ignoring the anisotropy of bremsstrahlung emission).[2] Due to the conical shape of the flattening filter, the photon spectrum emerging along the central axis is hardened more than the spectrum emerging at an angle to the central axis. This results in a reduction in beam energy as the distance from the central axis increases.

The incident energy fluence can be calculated from measurements made in air with a dense buildup cap (see Equations 3.74 and 3.75) or by iterative deconvolution of the dose profile obtained in a phantom (Mackie et al., 1996). The incident energy fluence in air is characterized at some convenient source distance (perhaps isocenter) on a plane perpendicular to the central axis at $z = z_0$ (see Figure 3.2). For simplicity, we consider the

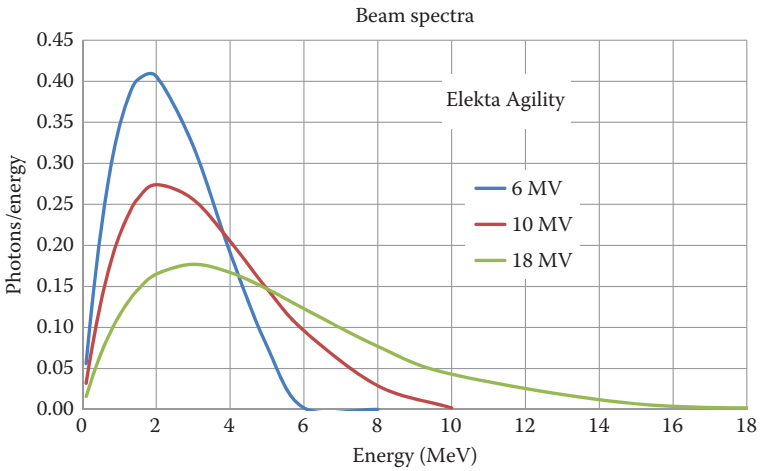



FIGURE 3.4 The beam spectra for 6, 10, and 18 MV beams from an Elekta Agility linac, as determined by the Pinnacle treatment planning system based on a fit to percent depth–dose curves. The vertical axis represents the relative number of photons per energy interval. The total area under each curve (proportional to the total fluence) is approximately the same.

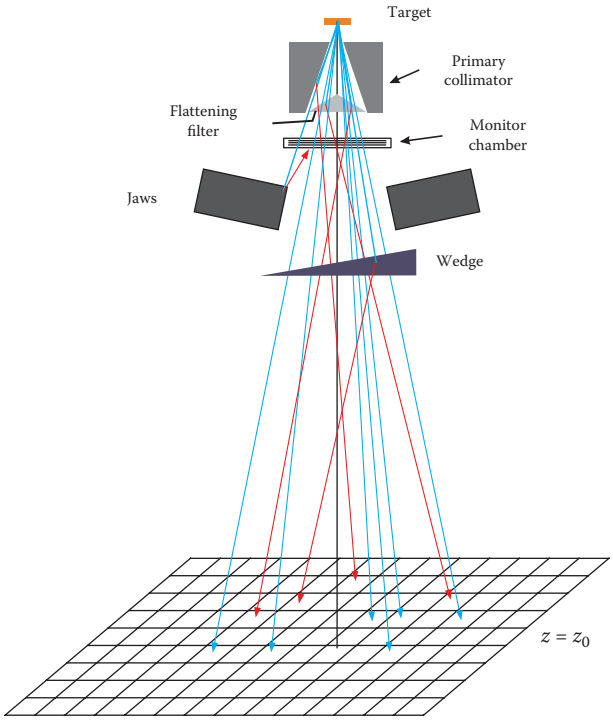



FIGURE 3.5 Energy fluence incident on the medium consists primarily of "direct" radiation (in blue) originating from bremsstrahlung emission in the target plus head scatter radiation (in red). Head scatter radiation is scattered predominantly from the primary collimator and the flattening filter (if present). It is necessary to determine the energy fluence in air on each grid element at a standard distance ($z = z_0$, isocenter) from the target. The energy distribution of these photons is needed, as well as (in principle) the direction of the photons.

monoenergetic case. The energy fluence at $\vec{r}_0$ may be described as $\Psi(\vec{r}_0) = \Psi(C; x, y, z_0)$, where $C$ represents all beam collimation and modulation variables (jaw setting, wedges, etc.). The energy fluence can be divided into two contributions: unscattered photons that come directly from the target and photons that are scattered in the head, $\Psi = \Psi_{usc} + \Psi_{hsc}$ (see Figures 3.1 and 3.5). The lateral distribution of unscattered energy fluence can be written in terms of a relative distribution function as follows: $\Psi_{usc}(C; x, y, z_0) = \Psi_0 f(C; x, y, z_0)$, where $\Psi_0$ is the total energy fluence on the central axis, in air, at a distance of $z_0$ from the target. The total incident energy fluence may be written as

$$\Psi(\vec{r}_0) = \Psi_0 \left[ f(C; x, y, z_0) + \frac{\Psi_{hsc}}{\Psi_0}(C; x, y, z_0) \right]. \tag{3.25}$$

The dose is then proportional to $\Psi_0$. We will make use of this fact when we discuss monitor unit (MU) calculations later.

The primary photons must be transported through the medium by ray tracing. The primary energy fluence in the medium may be written as

$$\Psi(\vec{r}') = \Psi(\vec{r}_0)\left(\frac{r_0}{r'}\right)^2 e^{-\mu(r'-r_s)}, \tag{3.26}$$

and therefore the terma for a monoenergetic beam in a homogeneous medium is

$$T(\vec{r}') = \Psi(\vec{r}_0)\left(\frac{\mu}{\rho}\right)\left(\frac{r_0}{r'}\right)^2 e^{-\mu(r'-r_s)}. \tag{3.27}$$

Surface dose is produced almost entirely by electron contamination. Electron contamination is not included in the kernel, and therefore the dose from this must be added separately to that computed by C/S. According to Mackie et al. (2007), the electron contamination component resembles an electron beam with a practical range approximately equal to the $d_{\max}$ of the photon beam. The electron contamination dose component is modeled in the Pinnacle (McNutt, 2002) treatment planning system as a function that declines linearly inward from the surface at first and then, beginning at some depth, declines exponentially.

## 3.6 POINT DOSE KERNELS

The absorbed dose in a medium can be divided into two general categories: primary dose and scatter dose. The primary component of the dose originates from the interaction of primary photons, which set charged particles in motion. The scatter component of the dose originates from photons that have been scattered at least once in the medium.

The kernel is sometimes divided up into a primary and a scatter component:

$$A(\vec{r},\vec{r}') = A_p(\vec{r},\vec{r}') + A_s(\vec{r},\vec{r}'). \tag{3.28}$$

The reason for this is the very different physics involved in the two contributions to the total kernel and practical considerations related to numerical calculation. For the primary kernel, we are dealing with electron transport, and for the scatter kernel, we are dealing with photon scattering (and then electron transport). Point dose kernals are almost universally calculated by the Monte Carlo technique. The Monte Carlo calculation forces interactions to occur at some fixed point (at $\vec{r}$) for each particle history and then records the energy absorbed in each voxel. The energy lost by primary photons at the interaction site is also recorded. If separate primary and scatter kernels are required, these are stored in separate arrays. Energy transport by all possible secondary particles must be described:

1. Single and multiply scattered photons
2. Photoelectrons
3. Compton recoil electrons
4. Auger electrons

5. $e^+/e^-$ pairs (and annihilation photons)
6. Bremsstrahlung photons
7. Photonuclear particles

Selected references for point kernel calculations and data are given in Table 3.1. The point kernel, as defined here, appears to correspond to the usage of Ahnesjö and coworkers (1987, 1989), who call this the "point spread function." Both the names and the units used in other references are different, and the precise definition of the kernel is not always clear.

Figure 3.6 shows the primary kernel calculated for 1.25 MeV primary photons (Mackie et al., 1988). This kernel was calculated by forcing monoenergetic photons to interact at the center of a sphere of water with a radius of 60 cm. The energy deposition is largely due to Compton recoil electrons, and since these electrons have a fairly short range, the kernel drops off rapidly around the interaction site.

Ahnesjö (1989) has fit Monte Carlo calculated data to an analytic expression for polyenergetic kernels of the form

$$A(\vec{r},\vec{r}') = \frac{A_\theta e^{-a_\theta|\vec{r}-\vec{r}'|} + B_\theta e^{-b_\theta|\vec{r}-\vec{r}'|}}{|\vec{r}-\vec{r}'|^2}, \tag{3.29}$$

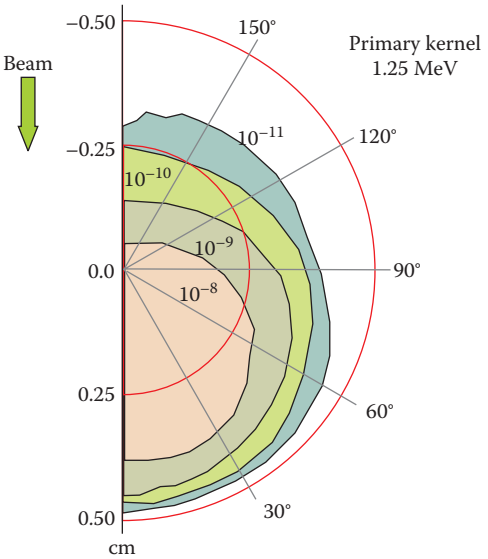



FIGURE 3.6 Primary energy deposition kernel for 1.25 MeV photons in water. The units are cGy $MeV^{-1}$ $photon^{-1}$. The beam is coming from above, and the photons interact at 0.0 on the vertical axis. Note the drop in value by more than three orders of magnitude within a distance of 0.5 cm. This corresponds roughly to the maximum range of electrons set in motion by 1.25 MeV photons. Compton scattering cannot set electrons in motion at angles larger than 90°. It is presumed that energy deposition at these angles is due primarily to backscattered electrons. (Adapted from Mackie, T.R., et al., *Phys. Med. Biol.* 33(1), 1–20, 1988.)

where $A_\theta$, $a_\theta$, $B_\theta$, and $b_\theta$ are functions of energy and scattering angle θ (see Figure 3.9) and where the first term is mainly primary dose and the second term is mainly scatter. This paper gives tables of these parameters for 4, 6, 10, 15, and 24 MV beams, along with the spectrum used for each beam. The form of Equation 3.29 is perhaps not surprising (at least for singly scattered photons), consisting of exponential attenuation over the distance from the interaction site, combined with inverse square attenuation along ray lines from the interaction site—the two main physical effects that are responsible for photon attenuation. The difficulty with a full-spectrum kernel like that given by Equation 3.29 is that it is only valid for the spectrum that it was computed for and may not be valid for a specific linac.

Figure 3.7 shows a contour plot of the total kernel from Ahnesjö et al. (1987) for a 10 MeV monoenergetic beam with respect to lateral and axial distance from

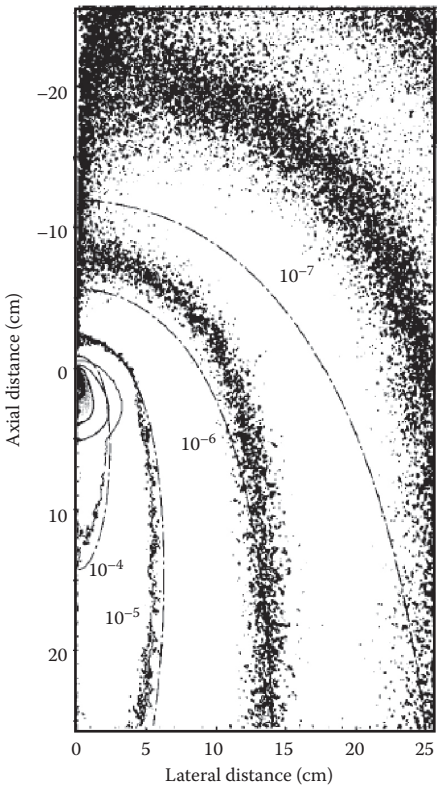



FIGURE 3.7 The total (includes both primary and scatter) point kernel for 10 MeV photons interacting in water (units $cm^{-3}$) The speckled data are from Monte Carlo calculations and the dash-dotted lines are the "first scatter terma," assuming only Compton interactions. Photons are incident from the top of the page traveling along the vertical axis and interacting at point (0, 0). Calculations were done in a cylindrical geometry with $5 \times 10^5$ photons. (From Ahnesjö et al., *Acta Oncol*. 26, 49–56, 1987, fig. 1c.)

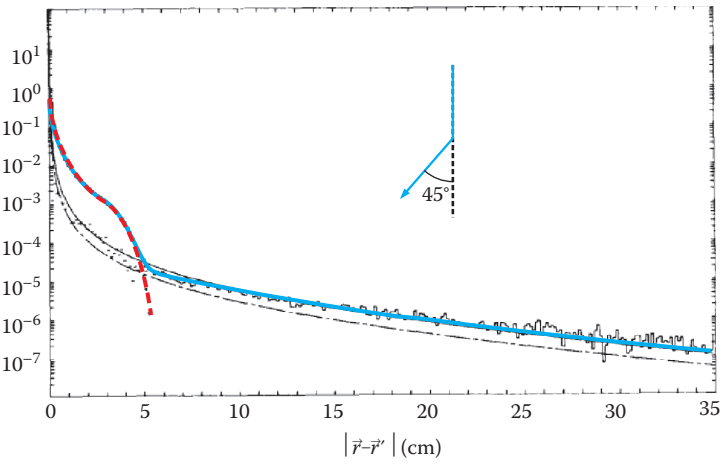



FIGURE 3.8 The point dose kernel for 10 MeV photons (see also Figure 3.7) for water in units of $cm^{-3}$ showing the radial dependence along a ray line at an angle of 45° from the path of the incident photon. The total kernel is shown by the solid line, and the primary component is shown by the dashed line. The primary component dominates within about 5 cm of the interaction site. This corresponds approximately to the maximum range of the secondary charged particles that deposit the primary dose component. The scatter component is dominant beyond a distance of 5 cm. (From Ahnesjö, A., et al., *Acta Oncol.* 26, 49–56, 1987, fig. 3.)

the interaction site. Each successive contour curve value drops by an order of magnitude.

Figure 3.8 shows a subset of the data from Figure 3.7. This is a log plot of the total scatter kernel for a 10 MeV beam as a function of distance from the scattering site along an axis at a 45° angle to the initial path of the photons. The dashed curve is the primary kernel due to charge particle transport. The solid curve is the total kernel, including both primary and scatter radiation. The primary contribution is dominant within a distance of 5 cm from the interaction site.

The maximum range of charged particles set in motion by primary photons is expected to be approximately $T_{max}/2$, where $T_{max}$ is the maximum energy of the electrons set in motion by Compton interactions. This is because the energy loss rate for MeV energy electrons in water is about 2 MeV/cm. This maximum range is approximately

$$\Re \approx \frac{T_{max}}{2} = \frac{E_0^2}{2E_0 + 0.511\,\text{MeV}}, \tag{3.30}$$

where $E_0$ is the energy of the primary photon in MeV. For 10 MeV photons, the maximum range is about 5 cm. We expect the kernel to be dominated by multiply scattered photons beyond a distance of approximately the mean free path, $1/\mu$. For 10 MeV photons in water, the value of $1/\mu$ is about 45 cm. This is off the scale in Figure 3.8.

## 3.7 ANALYTICAL DERIVATION OF A POINT KERNEL FOR SINGLY SCATTERED PHOTONS

It is possible to analytically derive an approximate expression for the scatter component of the dose deposition kernel. Although Equation 3.4 represents an implicit definition of the kernel, we need an explicit definition, consistent with this equation, that we can use to calculate the kernel. From Equation 3.4, we write

$$A = \frac{dD(\vec{r},\vec{r}')}{T(r')d^3r'}. \tag{3.31}$$

We can now express in words how $A$ is to be computed: calculate the energy per unit mass absorbed at $\vec{r}$ due to interactions at $\vec{r}'$ and divide by the energy lost by photons per unit mass in interactions at $\vec{r}'$. The units of the kernel are inverse length cubed (e.g., $cm^{-3}$).

We assume single photon scattering or, equivalently, that multiply scattered photons may be ignored. We assume that all interactions are Compton scattering. We further assume that the incident radiation is monodirectional and monoenergetic and that scattered (but not primary) photons produce charged particles that obey CPE.

Compton scattering is axisymmetric, and therefore we only need worry about the scattering angle θ shown in Figure 3.9. The differential cross section $d\sigma/d\Omega$ is the number of particles scattered through angle θ per unit solid angle, divided by the

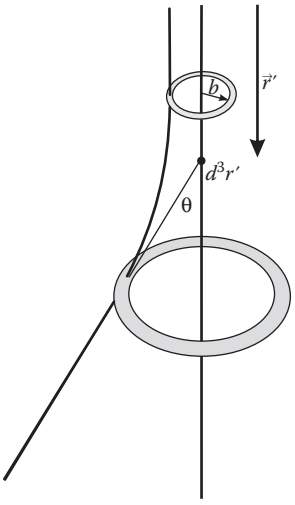



FIGURE 3.9 Illustration of the differential cross section. The photon is scattered through the angle θ. The quantity $b$ is the classical impact parameter.

incident fluence (number of particles incident per unit area). The number of photons scattered through angle θ by a single electron can therefore be written

$$dN_1 = \Phi_i \frac{d\sigma}{d\Omega} d\Omega, \tag{3.32}$$

where $\Phi_i$ is the incident photon fluence. The units of $d\sigma$ are cm$^2$. The differential scattering cross section is a function of θ and the incident photon energy and is given by the Klein–Nishina formula (Johns and Cunningham, 1983):

$$\frac{d\sigma}{d\Omega} = \frac{1}{2} r_e^2 \left(1 + \cos^2\theta\right) F_{KN}, \tag{3.33}$$

where

$$F_{KN} = \left\{\frac{1}{1+\alpha(1-\cos\theta)}\right\}^2 \left\{1 + \frac{\alpha^2(1-\cos\theta)^2}{\left[1+\alpha(1-\cos\theta)\right]\left(1+\cos^2\theta\right)}\right\}, \tag{3.34}$$

$r_e$ is the classical electron radius ($2.818 \times 10^{-13}$ cm) and $\alpha = E_0/(m_0c^2)$.

In Figure 3.9, photons are incident from above with uniform fluence $\Phi_i$ over a microscopic range of impact parameters *b*. According to classical physics, photons with different impact parameters will be scattered through different angles. At the scattering location, there is actually an infinitesimal volume containing $\rho_e\, d^3r'$ electrons, where $\rho_e$ is the number of electrons per unit volume. The total number of photons scattered by the electrons in the scattering element is therefore

$$dN = \Phi_i \frac{d\sigma}{d\Omega} d\Omega \rho_e\, d^3r'. \tag{3.35}$$

Now let us calculate the fluence at some macroscopic distance $h = |\vec{r} - \vec{r}'|$ over which the scattered photons may be attenuated. The element of solid angle $d\Omega = dA/h^2$, where $dA$ is the area that the photons are scattered into at a distance *h*. The fluence of particles scattered at angle θ reaching distance *h* is

$$d\Phi_s = \frac{dN}{dA} = \Phi_i e^{-\mu_s h} \frac{d\sigma}{d\Omega} \frac{dA}{h^2} \frac{1}{dA} \rho_e d^3r' = \Phi_i(\vec{r}') \rho_e \frac{d\sigma}{d\Omega} \frac{e^{-\mu_s h}}{h^2} d^3r', \tag{3.36}$$

where $\mu_s$ is the linear attenuation coefficient for the scattered photons.

If we assume that the scattered photons interact at the dose deposition point to produce a charged particle spectrum that exhibits CPE, then the scatter dose due to these photons will be given by

$$dD_s(\vec{r},\vec{r}') \overset{CPE}{=} dK_c = d\Phi_s E_s \frac{\mu_{en-s}}{\rho}, \tag{3.37}$$

where:

$E_s$ is the energy of the scattered photon

$(\mu_{en-s}/\rho)$ is the mass energy absorption coefficient for the scattered photons

The dose at the observation point due to the scattering element is therefore

$$dD_s(\vec{r},\vec{r}') = E_s \frac{\mu_{en-s}}{\rho} \Phi_i(\vec{r}') \rho_e \frac{d\sigma}{d\Omega} \frac{e^{-\mu_s h}}{h^2} d^3 r'. \tag{3.38}$$

For a monoenergetic primary beam of energy $E_0$, the terma is given by

$$T(\vec{r}') = \frac{dE_t}{dm'} = \frac{\mu}{\rho} E_0 \Phi(\vec{r}'). \tag{3.39}$$

Using Equations 3.31, 3.38, and 3.39, we can now write an expression for the scatter kernel (for singly scattered photons) in the homogeneous case:

$$A_{sh}(\vec{r}-\vec{r}') = \frac{E_s}{E_0} \frac{\mu_{en-s}}{\mu} \rho_e \frac{d\sigma}{d\Omega} \frac{e^{-\mu_s h}}{h^2}, \tag{3.40}$$

where subscript *sh* refers to scattered photons in a homogeneous medium. The quantities $E_s$, $\mu_{en-s}$, and $\mu_s$ are functions of θ.

The scattering kernel in Equation 3.40 is of the form $B_\theta e^{-b_\theta h}/h^2$ suggested by Ahnesjö (1989) (see Equation 3.29), where $B_\theta$ and $b_\theta$ are functions of θ. Contour lines of the scattering kernel given by Equation 3.40 are plotted in Figure 3.10 for 1.25 MeV monoenergetic photons. It is assumed that

$$\mu_{en-s} \approx \mu_{tr}. \tag{3.41}$$

We note that this scattering kernel has a close qualitative resemblance (similar teardrop shape) to the Monte Carlo calculated kernel shown in Figure 3.7.

## 3.8 HETEROGENEITIES

An inhomogeneous medium presents a difficult problem, and it is in general a challenging complication. The effect of tissue heterogeneities is very important for small field sizes. This is because of a lack of lateral equilibrium. The apertures employed

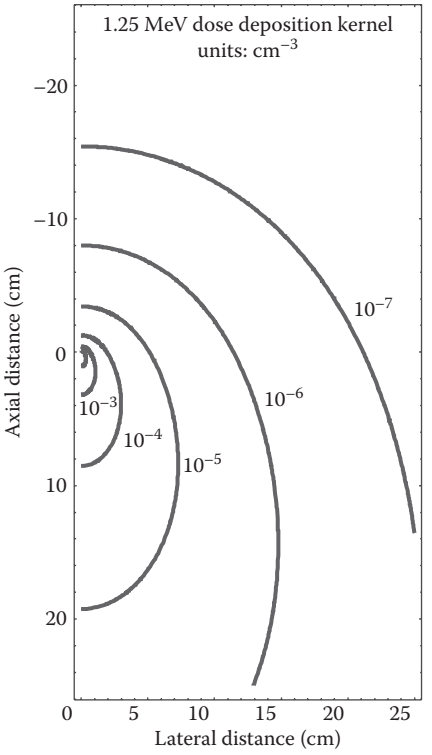



FIGURE 3.10 The scatter component of the point dose kernel in water for 1.25 MeV photons interacting at position (0, 0) and based on Equations 3.40 and 3.41. This assumes single Compton scattering only.

in IMRT delivery are as small as several square centimeters. IMRT requires dose calculation methods that can handle nonequilibrium effects. The effects of inhomogeneities must be accounted for in primary photon transport, scattered photon transport, and charged particle transport. The effects of heterogeneities must be taken into account for both the terma and the kernel. The terma represents the primary photon transport, and this is the easiest part of the problem to handle. For the kernel, we must account for effects on the charged particle component and the scattered photon component.

Knöös and McClean (2008) have described two types of algorithms with respect to heterogeneity corrections. Type A models are based on effective path length in a longitudinal direction only. This is characteristic of simple pencil beam models. Type B models account for both longitudinal and lateral scaling.

Let us deal first with primary photon transport. This is the terma part. Equation 3.7 describes the primary photon transport for a monoenergetic point source of radiation. The attenuation by the matter is represented by the exponential factor where the $r' - r_s$ term is the distance traveled through the medium. We may write this term as $e^{-\mu d}$, where $d$ is the distance traveled through the medium. In a heterogeneous

medium, $\mu = \mu(\vec{r}')$ and the expression for the attenuation must be modified. Suppose the photons travel along the $x$ axis from $x = 0$ to $x = d$; the expression for the attenuation becomes

$$\exp\left(-\int_0^d \mu(x)dx\right).$$

For Compton scattering $\mu/\rho = N_A Z\sigma_e/AW$, where $N_A$ is Avogadro's number, $Z$ is the atomic number, $AW$ is the atomic weight (g/mol), and $\sigma_e$ is the total Compton cross section per electron (the integral with respect to solid angle of $d\sigma/d\Omega$ in Equation 3.33).[3] The quantity $N_A Z/AW$ is the number of electrons per gram. We can now write the linear attenuation coefficient as

$$\mu(\vec{r}') = \rho N_A \frac{Z\sigma_e}{AW} = \rho_e(\vec{r}')\sigma_e,$$

where $\rho_e$ is the number of electrons per unit volume. The radiological distance is defined as

$$d_{rad} = \frac{1}{\rho_e^0\sigma_e}\int_0^d \rho_e(x)\sigma_e dx = \int_0^d n_e(x)dx = \bar{n}_e d, \tag{3.42}$$

where:

$d$ is the physical distance along the path
$\rho_e^0$ is the electron density of water
$n_e = \rho_e/\rho_e^0$ is the electron density relative to water (per unit volume)
$\bar{n}_e$ is the average electron density along the path

The attenuation factor may now be written as $e^{-\mu^0\bar{n}_e d}$, where $\mu^0$ is the linear attenuation coefficient for the given beam energy in water. Equation 3.7 for the primary fluence now becomes

$$\Psi(\vec{r}') = \Psi(\vec{r}_0)\left(\frac{r_0}{r'}\right)^2 e^{-\mu^0\bar{n}_e(r'-r_s)}. \tag{3.43}$$

We now turn to the more difficult task of evaluating the effect of inhomogeneities on the kernel. In principle, the presence of inhomogeneities requires a separate kernel for each point in the medium. This is clearly not practical, and it is customary instead to modify the homogeneous kernel for water. The simplest possible expedient is to spatially scale the kernel in accordance with radiological distances. In this case, all distances are replaced with radiological distances and Equation 3.6 becomes (Mackie et al., 2007)

$$D(\vec{r}) \stackrel{?}{=} \int_V T(\bar{n}_{e,\vec{r}'}\,\vec{r}')A\left[\bar{n}_{e,\vec{r}-\vec{r}'}(\vec{r}-\vec{r}')\right]d^3r', \tag{3.44}$$

where:

$\bar{n}_{e\vec{r}}r'$ is the radiological distance from the source to the interaction site (relative electron density weighted along the path of the vector)

$\bar{n}_{e,\vec{r}-\vec{r}'}|\vec{r}-\vec{r}'|$ is the radiological distance from the interaction site to the dose deposition site

There is a question mark over the equal sign because this equation has not been derived but rather *postulated*. In Equation 3.44, the influence of tissue heterogeneities on the dose calculations at point $\vec{r}$ is approximated by density weighted rays from $\vec{r}$ to all sites at which primary photons interact. The kernel in Equation 3.44 is no longer spatially invariant, as $\bar{n}_{e,\vec{r}-\vec{r}'}$ depends on the specific location of the path between $\vec{r}$ and $\vec{r}'$. The convolution theorem, as discussed in Section 3.3, no longer applies due to the lack of spatial invariance. Inhomogeneities can be included in the terma alone without violation of position invariance.

Equation 3.44 seems intuitively reasonable for first scatter photons, but the validity is not clear for the primary component of the kernel or for multiply scattered photons. Although this equation seems plausible for first scattered photons, let us see if we can derive it using the approach of Section 3.7. We will later return to a discussion of the primary component of the kernel.

To evaluate the effect of inhomogeneities on the scatter component of the kernel, we repeat the analysis of Section 3.7 for the single scatter case, but this time we do not assume a homogeneous medium. Equation 3.35 for the total number of photons scattered by the electrons in the scattering element becomes

$$dN = \Phi_i(\vec{r}')\frac{d\sigma}{d\Omega}d\Omega\rho_e(\vec{r}')d^3r'. \tag{3.45}$$

Equation 3.36 becomes

$$d\Phi_s = \Phi_i(\vec{r}')\rho_e(\vec{r}')\frac{d\sigma}{d\Omega}\frac{e^{-\mu_s^0\bar{n}_e h}}{h^2}d^3r', \tag{3.46}$$

where:

$\mu_s^0$ is the linear attenuation coefficient for the scattered photons in water

$\bar{n}_e h$ is the radiological distance for the scattered photons along the path between the scattering event at $\vec{r}'$ and the dose deposition location at $\vec{r}$ (see Figure 3.2)

The latter quantity depends on the relative electron density $n_e$ (per unit volume, relative to water) distribution along this line.

Equation 3.37 becomes

$$dD_s(\vec{r},\vec{r}') \overset{CPE}{=} dK_c = d\Phi_s E_s \frac{\mu_{en-s}(\vec{r})}{\rho(\vec{r})}, \tag{3.47}$$

where:

$E_s$ is the energy of the scattered photon

$(\mu_{en-s}/\rho)$ is the mass energy absorption coefficient for the scattered photons

The dose at the observation point due to the scattering element is therefore (see Equation 3.38)

$$dD_s(\vec{r},\vec{r}') = E_s \frac{\mu_{en-s}(\vec{r})}{\rho(\vec{r})} \Phi(\vec{r}') \frac{d\sigma}{d\Omega} \frac{\rho_e(\vec{r})}{h^2} e^{-\mu_s^0 \bar{n}_e h} d^3 r'. \tag{3.48}$$

For a monoenergetic primary beam of energy $E_0$, the terma is given by

$$T(\vec{r}') = \frac{dE_t}{dm'} = \frac{\mu(\vec{r}')}{\rho(\vec{r}')} E_0 \Phi(\vec{r}'). \tag{3.49}$$

Using Equations 3.31, 3.48, and 3.49 we can now write an expression for the scatter kernel in the inhomogeneous case:

$$A_s(\vec{r},\vec{r}') = \frac{E_s}{E_0} \frac{\rho(\vec{r}')}{\rho(\vec{r})} \frac{\mu_{en-s}(\vec{r})}{\mu(\vec{r}')} \rho_e(\vec{r}') \frac{d\sigma}{d\Omega} \frac{e^{-\bar{n}_e \mu_s^0 h}}{h^2}. \tag{3.50}$$

In the homogeneous case, the scatter kernel reduces to Equation 3.40:

$$A_{sh}(\vec{r}-\vec{r}') = \frac{E_s}{E_0} \frac{\mu_{en-s}^0}{\mu^0} \rho_e^0 \frac{d\sigma}{d\Omega} \frac{e^{-\mu_s^0 h}}{h^2}, \tag{3.51}$$

where superscript zero denotes water composition and a density of 1 g/cm$^3$.

In Equation 3.50, there are, in general, an infinite variety of possibilities depending on the distribution of mass and electron density in the patient. This implies the need for a different kernel for every situation. This is clearly not practical, and it is therefore desirable to be able to write the inhomogeneous kernel in terms of the homogenous kernel. The linear attenuation coefficients in Equation 3.50 are assumed proportional to the electron density $\rho_e$, and therefore

$$\frac{\mu_{en-s}(\vec{r})}{\mu(\vec{r}')} = \frac{n_e(\vec{r})}{n_e(\vec{r}')} \frac{\mu_{en-s}^0(\vec{r})}{\mu^0(\vec{r}')}. \tag{3.52}$$

With a little effort and consideration of Equations 3.50 through 3.52, it can be shown that

$$A_s(\vec{r},\vec{r}') = \frac{\rho(\vec{r}')}{\rho(\vec{r})} n_e(\vec{r}) \bar{n}_e^2 A_{sh}\left(\bar{n}_e(\vec{r}-\vec{r})\right). \tag{3.53}$$

Equation 3.53 is only strictly correct for singly scattered photons.

A further simplification to Equation 3.53 is possible if we make an assumption about the composition

$$n_e(\vec{r}) = \frac{\rho_e(\vec{r})}{\rho_e^0} = \frac{\rho(\vec{r})}{\rho^0} \frac{(Z/AW)}{(Z/AW)^0} \approx \frac{\rho(\vec{r})}{\rho^0}.$$

The quantity $Z/AW$ is typically about ½ for all elements except hydrogen, for which it is 1. Treatment planning systems usually compute dose to a material with the composition of water even within an inhomogeneity (Monte Carlo codes can compute the dose based on the specific composition). *It is assumed that within an inhomogeneity, the number of electrons per gram is the same as for water* even though the mass density may differ from 1.0 g/cm$^3$. Another way to state this is that it is almost always the case that treatment planning systems (that do not use Monte Carlo) calculate the dose within an inhomogeneity to a water cavity obeying Bragg–Grey conditions (Knöös et al., 2006). This is seldom stated and not widely understood. Table 3.2 shows electron density information for common materials and tissues. If we assume that the material at $\vec{r}$ has the composition of water, then

$$A_s(\vec{r}-\vec{r}') = \frac{\rho(\vec{r}')}{\rho^0} \bar{n}_e^2 A_{sh}\left[\bar{n}_e(\vec{r}-\vec{r}')\right]. \tag{3.54}$$

Equation 3.54 appears to be the same as Equation 3.13 of Ahnesjö et al. (1987); however, it appears to disagree with Ahnesjö and Aspradakis (1997, equation 37) because

**TABLE 3.2 Electron Densities**

| Material | ρ Density (g/cm$^3$) | Electron Density ($10^{23}$ g$^{-1}$) | Relative Electron Density (by Mass) | $\rho_e$, Electron Density ($10^{23}$ cm$^{-3}$) | $n_{e,r}$ Relative Electron Density (by Volume) |
|---|---|---|---|---|---|
| Water | 1.00 | 3.343 | 1.000 | 3.343 | 1.000 |
| Adipose | 0.92 | 3.363 | 1.006 | 3.094 | 0.926 |
| Cortical bone | 1.85 | 3.139 | 0.939 | 5.807 | 1.737 |
| Muscle | 1.04 | 3.308 | 0.990 | 3.440 | 1.029 |
| Air | $1.21 \times 10^{-3}$ | 3.006 | 0.899 | 0.004 | $1.09 \times 10^{-3}$ |

*Source:* Data from Attix, F.H., *Introduction to Radiological Physics and Radiation Dosimetry,* New York: Wiley, 1986.

we have $\rho(\vec{r}')$ where they have $\rho(\vec{r})$. We see that the kernel in Equation 3.54 is not simply scaled as in Equation 3.44.

There is no reason to expect Equation 3.54 to be valid for multiply scattered photons. The kernel contribution for multiply scattered photons is not expected to scale with the mean radiological path length between the interaction and deposition point because the photons do not traverse this path.

We see that the scatter kernel for singly scattered photons in an inhomogeneous medium can be calculated from a modified version of the scatter kernel for a homogeneous medium. The distance from the interaction point to the absorption point is scaled by the mean electron density along this line. The scatter kernel is stretched in regions of low density and compressed in regions of high density. This is illustrated in Figure 3.11.

We now consider the effects of inhomogeneities on the primary kernel. In this case, we are concerned with the transport of charged particles, chiefly electrons. According to Metcalfe et al. (2007), it is generally assumed that energy loss by secondary electrons in traveling from $\vec{r}'$ to $\vec{r}$ is dependent on the radiological path length. The dose from a differential energy fluence of charged particles is

$$D(\vec{r}) = \int_0^\infty (S_\rho)_c \, \Phi_E(\vec{r}, E) dE, \tag{3.55}$$

where:

$(S\rho)_c$ is the mass collision stopping power (see Chapter 4)

$\Phi_E$ is the differential energy fluence of charged particles (excluding delta rays) resulting from the interaction of primary photons with the medium (Attix, 1986)

Metcalfe et al. (2007) state that it is assumed that the elemental composition is the same throughout the medium (as noted earlier), even though the density may vary. The mass collision stopping power is independent of the mass density but does

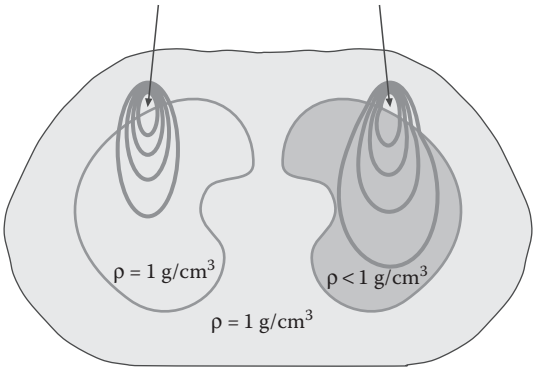



FIGURE 3.11 The qualitative effects of inhomogeneities on the kernel. The left side of the figure shows a kernel in a homogeneous medium. On the right-hand side, a low-density inhomogeneous region causes the same kernel to be stretched.

depend on $Z/AW$ (and thus the electron density per gram). If we assume that this electron density (per gram) does not vary (see Table 3.2), then the stopping power will be unaffected. It is not obvious that the differential charged particle energy fluence will depend solely on the radiological distance from $\vec{r}'$ to $\vec{r}$. This would be more plausible if electrons traveled in straight lines.

If we ignore the fact that multiply scattered photons and charged particles cannot truly be treated by scaling the homogeneous kernel as in Equation 3.54, then the most general case of the superposition integral for the dose becomes

$$D(\vec{r}) = \int_E \int_V T_E(\vec{r}') \frac{\rho(\vec{r}')}{\rho(\vec{r})} \bar{n}_e^2 A_h\left(E, \bar{n}_e(\vec{r} - \vec{r}')\right) d^3 r' dE. \tag{3.56}$$

## 3.9 PENCIL BEAMS

A considerable savings in computation time could be realized if the depth portion of the integral in Equation 3.6 could be carried out in advance. This is the idea of the pencil beam kernel. The pencil beam kernel gives the dose at a point from an infinitesimal pencil of radiation that enters the phantom at a point $\vec{r}_s$ (see Figure 3.12). A pencil beam kernel is the energy deposition in a semi-infinite medium from a point monodirectional beam. Calculation of the kernel essentially involves integrating the dose (Equation 3.6) first in depth and storing the results as a way of saving computation time.

We assume parallel monoenergetic pencil beams. We also assume that the medium is infinitely deep. The geometry is shown in Figure 3.12. The fluence at depth $z'$ is given by (see Equation 3.7)

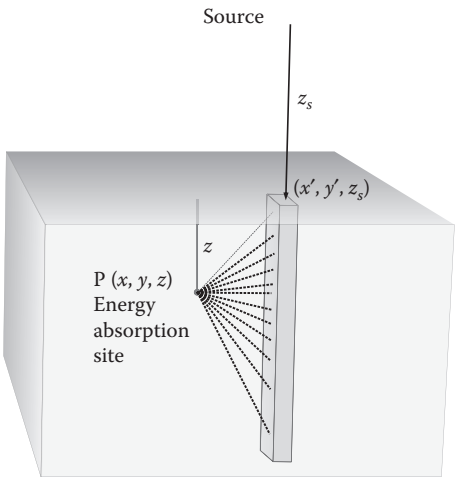



FIGURE 3.12 A pencil of radiation with small cross section is incident on a medium at point $(x', y', z_s)$ contributing to dose at point P. A pencil beam algorithm calculates the dose by summing the contributions over the beam aperture from all of the pencil beams weighted by the incident energy fluence distribution.

$$\Psi(\vec{r}') = \Psi(x', y', z_0)\left(\frac{z_0}{z'}\right)^2 e^{-\mu(z'-z_s)}. \tag{3.57}$$

Equation 3.8 for the dose may now be written as

$$D(x, y, z) = \int\int_{\text{area}} \Psi(x', y', z_0) P_z(x - x', y - y') dx' dy', \tag{3.58}$$

where

$$P_z(x - x', y - y') = \int_{z_0}^{\infty} \frac{\mu}{\rho}\left(\frac{z_0}{z'}\right)^2 e^{-\mu(z'-z_0)} A(x - x', y - y', z - z') dz'. \tag{3.59}$$

The dose is computed by integrating the pencil kernel multiplied by the incident fluence over the area of the beam aperture. This is illustrated in Figure 3.13. The pencil kernel is a function of the depth (see Figure 3.12). Note that Equation 3.58 is in the form of a convolution integral. Equation 3.59 assumes that the patient source to surface distance (SSD) is $z_s = z_0$ (at the central axis; see Figure 3.2) and that there are no variations in this quantity; that is, the patient skin surface is flat and perpendicular to the central axis. Equation 3.59 also assumes that the patient is infinite in depth and lateral extent. This will be discussed further in Section 3.10.

Pencil beam kernels can be derived from measurements (Chui and Mohan, 1988; Ceberg et al., 1996) or from Monte Carlo calculations (Mohan and Chui, 1987; Ahnesjö et al., 1992). Ahnesjö et al. (1992) have calculated pencil beam kernels from Monte Carlo simulations that they have fit to a function of the form

$$P_z = \frac{1}{u}\left(A_z e^{-a_z u} + B_z e^{-b_z u}\right), \tag{3.60}$$

where:

$u^2$ $= (x - x')^2 + (y - y')^2$

$u$ is the perpendicular distance from the axis of the pencil beam

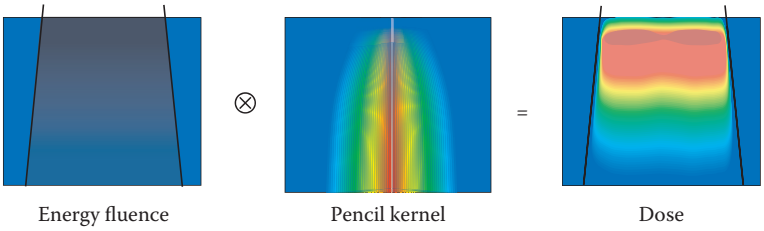



FIGURE 3.13 The pencil beam approach to dose calculation. The incident energy fluence is convolved with the energy deposition from a precalculated narrow pencil beam of radiation to get the total dose.

The quantities $A_z$, $a_z$, $B_z$, and $b_z$ are functions of $z$ (presumed SSD = 100 cm).[4] A density plot for a 5 MV kernel of this form is shown in Figure 3.14.

Heterogeneities can be handled by using the radiological depth when looking up the value of $P_z$. This, by itself, ignores the three-dimensional nature and effects of the inhomogeneities.

A finite-size pencil beam (FSPB) model divides the beam aperture into a series of small but finite (in cross section) pencil beams of identical cross-sectional area. The integral in Equation 3.58 is carried out over this small area assuming that the incident fluence is constant over that area. The FSPB are precomputed and then summed over the beam aperture weighted by the incident fluence at the location of each FSPB.

For an FSPB, the dose per unit fluence due to a single pencil of radiation centered at $(x_i, y_j)$ is given by

$$\frac{\Delta D_{Pb}(x,y,z)}{\Psi(x_i,y_j,z_0)} = \int_{\text{FSPB area}} \int P_z(x_i - x', y_j - y', z)\,dx'dy' = PB_z(x - x_i, y - y_j). \tag{3.61}$$

It is assumed that the incident energy fluence is constant over the cross section of the FSPB. The total dose is then given by

$$D(x,y,z) = \sum_{i,j} \Psi(x_i,y_j,z_0)\left(\frac{z_0}{z_{s:ij}}\right)^2 e^{-\mu_w t_{ij}} F_{ij} PB_z(x - x_i, y - y_j), \tag{3.62}$$

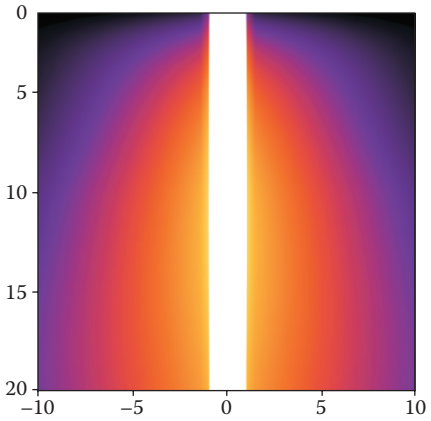



FIGURE 3.14 Density plot of the pencil beam kernel of Equation 3.60 from Ahnesjö et al. (1992) for a 5 MV beam. The vertical axis represents depth, and the horizontal axis is the perpendicular distance from the pencil ($u$ in Equation 3.60). The kernel is singular at $u = 0$, and this is the reason for the gap in that region. The scale is in cm.

where the sum is carried out over the beam aperture and $z_{s:ij}$ is the distance to the surface of the patient at location $(x_i, y_j)$, $\mu_w$ is the effective linear attenuation coefficient for a beam modifier (wedge), $t_{ij}$ is the thickness of the beam modifier along the ray line, and

$$F_{ij} = \left( \frac{z_0 + z}{z_{s:ij} + z} \right)^2 \left( \frac{z_{s:ij} + d_m}{z_0 + d_m} \right)^2, \tag{3.63}$$

where $d_m$ is the depth of maximum dose (see Bourland and Chaney, 1992). The quantity $F_{ij}$ is the Mayneord correction factor for the pencil beam at $(x_i, y_j)$.

The analytic anisotropic algorithm (AAA) used in the Varian Eclipse treatment planning software uses an FSPB algorithm. The software consists of two major components: a configuration module and the dose calculation engine. The purpose of the configuration module is to determine the user's beam energy spectrum and electron contamination. The dose calculation includes contributions from the primary fluence (that part not scattered in the head), scattered extrafocal photons, and electron contamination. The pencil kernel is represented by an analytic fit to a weighted sum of four Gaussians and thus results in great time savings. The equation for the kernel is (Sievenen et al.)

$$K_\beta = \sum_{k=0}^{3} \frac{c_k(z)}{\pi \sigma_k^2(z)} e^{-u^2/\sigma_k^2}(z), \tag{3.64}$$

where:

$c_k(z)$ are the weighting factors
$\sigma_k(z)$ are the standard deviations

These parameters are determined from Monte Carlo–calculated monoenergetic scatter kernels and the beam spectrum. The kernel is based on the energy spectrum and the radial dependence of the beam energy due to beam hardening in the flattening filter as derived in the configuration module. The FSPB are tilted to follow beam divergence. The beam profile is modeled in terms of the energy fluence as a function of radial distance from the central axis. The scatter kernels were computed using Monte Carlo calculations for monoenergetic pencil beams in various media. The polyenergetic kernel is computed as a weighted sum of the monoenergetic kernels. Heterogeneity corrections are based on effective longitudinal path length (along the pencil beam), with the addition of a lateral scaling based on density in the four main lateral directions. This allows for anisotropic corrections for inhomogeneities that are lateral to the pencil beam.

## 3.10 PATIENT GEOMETRY

For both point dose and pencil beam kernels, there is going to be a problem near any boundary of the medium if the kernel calculation is based on the assumption of an infinite medium. The boundary could be either a beam entrance surface, an exit surface, or an edge where there is "flash," such as in tangential breast irradiation.

Point dose kernels are generally calculated based on the assumption of a geometrically large or effectively infinite medium. The dose calculation integral (Equation 3.4) is carried out over the finite extent of the patient geometry.

There will be problems in the buildup region because the point kernel assumes that there is no upper boundary. Near the beam entry surface, the terma will be handled properly based on a ray trace to locate the position of the surface along all rays (see Figure 3.2). The kernel, however, will be inaccurate if it is calculated based on the assumption that there is no upper boundary. The difficulty is primarily due to the *assumed* presence of multiply scattered photons from the region above the surface of the medium. For primary and single scatter dose deposition, the entrance boundary should present no difficulty (except perhaps for the contribution of backscattered electrons).

Next, we consider exit surfaces. For multiply scattered photons, the kernel near an exit surface or a density interface will be incorrect. This will result in an overestimate of the dose near these surfaces (and near the surface for tangential irradiation). The kernel implicitly assumes that there is backscattering material where in reality there is none. For point dose kernels, this effect is on the order of several percent within 3 cm of the phantom boundary for 4 MV beams (Aspradakis, 1996). The effect is less pronounced for higher beam energy because, in this case, photons are more likely to be forward scattered (see Ahnesjö and Aspradakis, 1999, p. R129).

Pencil beam algorithms suffer from the same problems as described above for point dose kernels. For implementations assuming a flat patient entrance surface, the scatter can be incorrectly calculated as shown in Figure 3.15. Pencil beams are generally computed assuming a semi-infinite medium. The dose calculation using a pencil

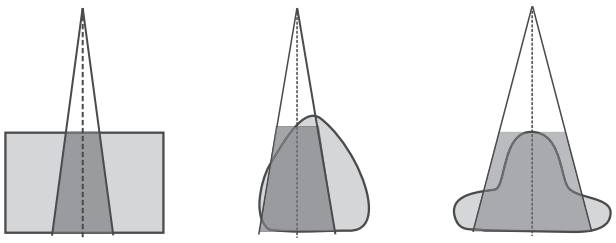



FIGURE 3.15 Pencil beam dose calculations that do not account for surface contour. (Adapted from Ahnesjö, A., Current concepts in dose calculations, slide presentation, 2006, http://www.sasro.ch/PastEvents/Lung_Seminar_2006/lung-ahn.pdf.)

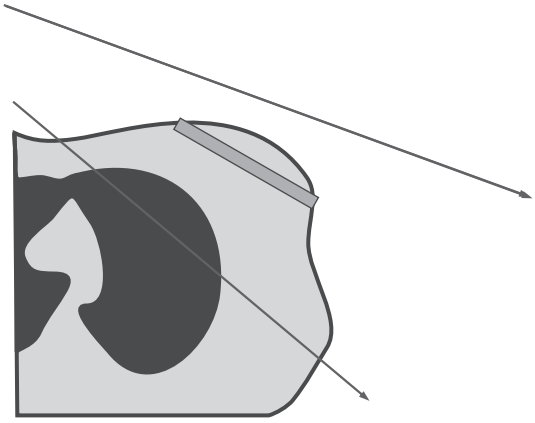


FIGURE 3.16 Pencil beam calculation of the dose for a tangential breast treatment. For the pencil beam shown, there is a deficit of tissue above and to the right of the pencil, contrary to the assumption on which the pencil kernel was calculated. When there is a lack of tissue to the side, a simple pencil beam algorithm will predict too much dose.

beam usually assumes that there are contributions from the dose from all sides of a dose pencil. Due to the irregular shape of a patient, there may not be. Pencil beams ignore skin surface contour. In addition, this does not account for build-down at the exit surface. Figure 3.16 illustrates the problem with respect to patient flash.

## 3.11 COLLAPSED CONE CONVOLUTION

The CCC method is a superposition approximation method for point kernel dose calculation that has found widespread use. Treatment planning systems using CCC include Pinnacle, GammaPlan 10, RayStation, Prowess DAO MRT, and Mobius3D.

Numerical computation of dose requires that continuous variable integrals be replaced by discrete sums. The collapsed cone approach assumes that all the energy scattered from a differential volume element into a small cone is absorbed along the line forming the axis of the cone. Thus, the cone is "collapsed" onto its axis. This is illustrated in Figure 3.17. The cone axis directions are specified by values of $\theta_m$ and $\varphi_n$, and the solid angle subtended by the cone is $\Omega_{mn}$. The sum of all the $\Omega_{mn}$ must be $4\pi$ steradians. Figure 3.17 shows a single direction.

The CCC approximation speeds up the computation time in comparison to brute-force integration. The goal is to calculate the dose at the $N^3$ lattice points. For a homogeneous medium and monoenergetic radiation source, this would require $N^6$ operations. In CCC, when one calculates the dose at a point, one replaces the sum over all the other points with a sum over the radiation contribution from all the cone directions numbering $M$. A value of $M$ of about 100 is reasonable (Mackie et al., 1996). The computational problem now becomes one of order $N^3M$. This is a considerable

improvement in speed and is comparable to that gained from convolution (see Section 3.3).

The consequences of the CCC approximation are illustrated in Figures 3.17 through 3.19. In Figure 3.18, voxels A and A′ are irradiated. A single-cone axis direction is shown in this figure for each of these voxels. For these discrete cones, some of the

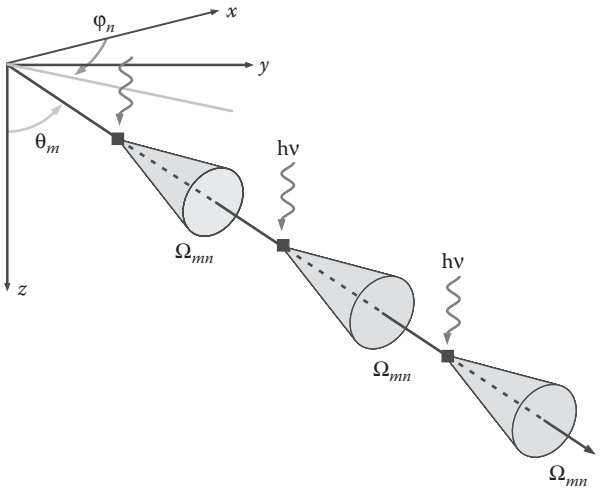



FIGURE 3.17 CCC approximation. Photons are incident from above on a few selected voxels (depicted as squares). A particular direction is shown, given by $\theta_m$ and $\varphi_n$. Photons actually scattered into the (small) solid angles $\Omega_{mn}$ are assumed to all travel along the direction of the common axis of the cones. Thus, the cones are collapsed. (Adapted from Ahnesjö, A., *Med. Phys.* 16(4), 577–92, 1989.)

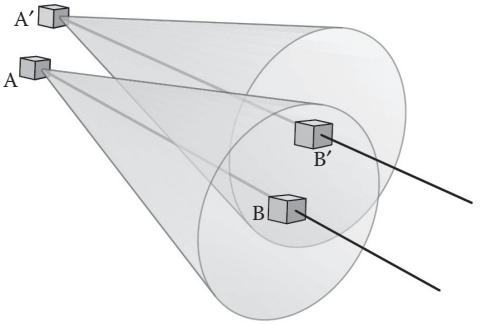



FIGURE 3.18 Consequences of the CCC approximation. Voxels A and A′ are irradiated and scatter photons into the discrete cones shown. The CCC approximation assumes that all the energy fluence is collapsed onto the two cone axes. As a result, scattered energy fluence from A that should be deposited into B′ is instead deposited into B. Likewise for A′ and B. This geometric effect becomes more pronounced at larger distances; however, the energy deposited in a voxel decreases very rapidly with increasing distance from the scatterer. (Adapted from Ahnesjö, A., *Med. Phys.* 16(4), 577–92, 1989.)

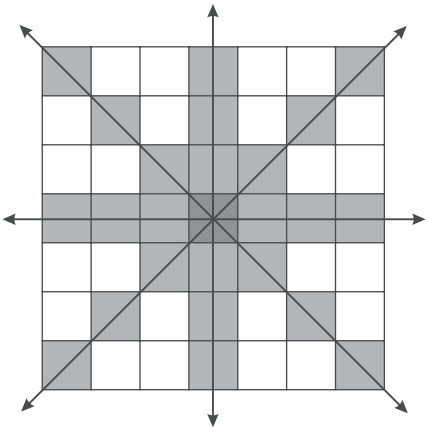


FIGURE 3.19 A two-dimensional illustration of the CCC approximation. Consider the contribution of scattered radiation from the central (dark gray) square to its surroundings. There are a total of eight "cone" angles represented by the arrows. Only the shaded boxes receive any energy deposition. In the CCC approximation with these eight cones, the unshaded boxes receive no energy deposition. If the number of cone angles is increased, then the boxes receiving no energy deposition will be farther away from the scattering element, where the missing contribution is much less. (Adapted from Mackie, T.R., et al., in *Teletherapy: Present and Future*, ed. T.R. Mackie and J.R. Palta, 103–35, Madison, WI: Advanced Medical Publishing, 1996.)

photon energy fluence scattered from A should be absorbed in voxel B′ but is instead absorbed in voxel B because the cone is collapsed onto the axis emanating from A. Likewise, some of the photon energy fluence scattered from A′ should be absorbed in voxel B but it is instead absorbed in voxel B′. The geometric accuracy of this method therefore decreases with increasing distance from the scattering voxel. On the other hand, the amount of energy deposition decreases rapidly with distance from the scattering voxel (as shown by Equation 3.29). In addition, this approximation is expected to be better for smaller, discrete solid angles. The solid angle in Figure 3.18 is quite large for the purpose of illustration. Figure 3.19 shows another representation of this effect in 2D. Figure 3.20 shows the collapsed cone axes used in the Elekta Leksell GammaPlan 10 dose calculation algorithm that is used to calculate Gamma Knife dose distributions.

The CCC can be considered from the point of view of a voxel that receives absorbed dose—each receiving voxel is at the intersection of collapsed cone lines, and these cone lines cover all $4\pi$ solid angles. Only energy emitted from volume elements on the cone axes is used to calculate the dose to the receiving voxel. Energy is received and transported from all elements located along the line by performing a sum of the contributions. The dose calculation volume has to be covered with a lattice of lines corresponding to the collapsed cones. This is illustrated in Figure 3.21 for a $3 \times 3 \times 3$ array of volume elements and 26 cone angles. In this way, rather than having to compute the contribution at a point from each of the other $N^3$ voxels, one only

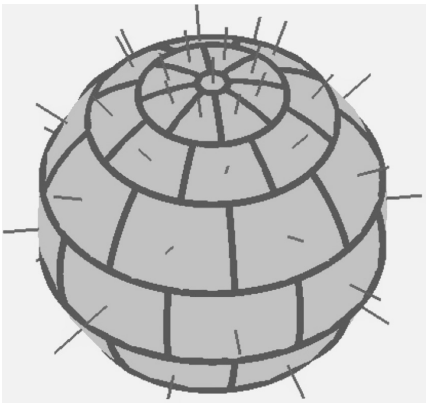


FIGURE 3.20 The collapsed cone axes (protruding from the sphere) used in the Elekta Leksell GammaPlan 10 dose calculation algorithm that is used for Gamma Knife dose calculations. The solid angle associated with each axis is bound by the longitudinal and latitudinal lines. There are a total of 42 directions. (From The Convolution Algorithm in Leksell GammaPlan® 10. Courtesy of Elekta, Stockholm.)

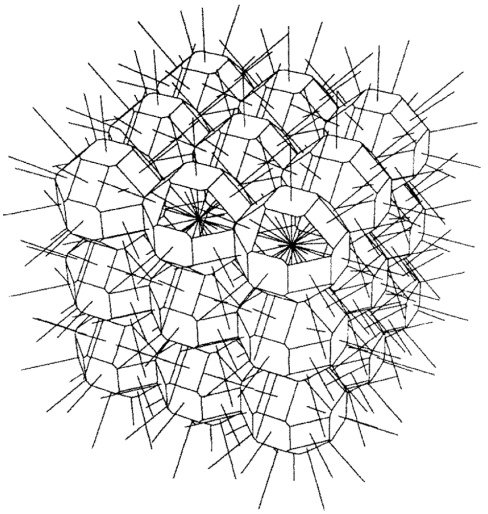


FIGURE 3.21 A CCC for an array of $3 \times 3 \times 3$ voxels. In this illustration, there are 26 different cone directions used for each voxel. (From Ahnesjö, A. and M.M. Aspradakis, *Phys. Med. Biol.* 44, R99–155, 1999, figure 12.)

need compute the contribution from each of the $M$ cone directions by carrying out a sum along these directions.

According to Mackie et al. (1996), about 100 collapsed cones are sufficient. Mobius3D uses 144. The Pinnacle treatment planning system uses a zenith bin count of 10 and an azimuthal bin count of 8. This is presumably the number of angular values $\theta_m$ and $\varphi_n$, respectively. These parameters cannot be modified by the user. The values of these parameters suggest that there are up to 80 collapsed cone beam directions.

## 3.12 CALCULATION OF MONITOR UNITS

In correction-based algorithms, monitor units (MUs) are calculated using percent depth dose, tissue maximum ratio (TMR), or tissue air ratio (TAR) values. An alternative method is necessary for model-based algorithms. In Section 3.5, it was shown that the dose is proportional to the incident energy fluence $\Psi_0 = \Psi(0, 0, z_0)$ for a monoenergetic beam. Although it is perhaps less obvious, we will presume that the same result holds for a polyenergetic beam. It is not possible to calculate the absolute dose without knowing the incident energy fluence. We can find the incident energy fluence by appealing to the absolute linac dose calibration. The calibration establishes the dose per MU in phantom under carefully specified standard conditions. If we ask the algorithm to calculate the dose per incident energy fluence under these same conditions, we can determine the incident energy fluence per MU. The details follow.

This discussion follows the description given by Ahnesjö and Aspradakis (1999). These authors include corrections for backscatter into the monitor ionization chambers. We ignore this complication here.

The number of MUs required to deliver a specified (prescribed) dose to a particular location is given by

$$M = \frac{D(x,y,z)}{\dot{D}(x,y,z)}, \tag{3.65}$$

where:

$D$ is the prescribed dose

$\dot{D}$ is the dose per MU at the prescription point

MU calculations always come down to a determination of the dose per MU.

The dose calculation algorithm is presumed to be capable of supplying the dose per energy fluence at any specified point. Note that the nature of the algorithm is irrelevant as long as the algorithm can calculate this quantity. Therefore, the method described here could be used for dose calculation algorithms other than superposition (e.g., Monte Carlo or Boltzmann transport equation). We define $w$ as follows:

$$w(x,y,z) = \frac{D(C;x,y,z)}{\Psi_0}, \tag{3.66}$$

where $\Psi_0 = \Psi(0, 0, z_0)$ is the (reference) incident energy fluence (see Equation 3.25) free in air and $C$ refers to the effects of all beam modifiers, including jaw setting and wedges.

The reference energy fluence $\Psi_0$ is tied to the dose calibration. Under dose calibration conditions,

$$\left(\frac{D}{M}\right)_{\text{cal}} = \frac{\Psi_0 w_{\text{cal}}}{M}. \tag{3.67}$$

The value of $(D/M)_{\text{cal}}$ is known from dose calibration measurements. Equation 3.67 may therefore be used to determine $\Psi_0/M$:

$$\frac{\Psi_0}{M} = \frac{(D/M)\text{cal}}{w_{\text{cal}}}. \tag{3.68}$$

The dose rate is then given by

$$\dot{D} = \left(\frac{D}{M}\right)_{\text{cal}} \left(\frac{w}{w_{\text{cal}}}\right). \tag{3.69}$$

Equation 3.69 is used along with Equation 3.65 to compute MUs.

The Pinnacle TPS defines the dose per MU as

$$\dot{D} = ND \cdot OF_c \cdot TTF \cdot (D/M)_{\text{cal}} \tag{3.70}$$

(Philips, 1993, p. 30), where $ND$ is the normalized dose, $OF_c$ is the computed correction factor determined during commissioning, and $TTF$ is the total transmission factor. The normalized dose $ND$ is defined as "the ratio of the dose per unit energy fluence at the prescription point to dose per unit energy fluence at the reference point for the calibration field (the point at which $(D/M)_{\text{cal}}$ was measured) as determined by the convolution superposition calculation for the treatment geometry." Based on this quotation, we interpret $ND$ to be given by

$$ND = \frac{D_d(C)}{\Psi(C)} \cdot \frac{\Psi(C_0)}{D_{d_m}(C_0)}, \tag{3 71}$$

where $C_0$ indicates calibration conditions.

We can rewrite this in the following fashion:

$$\begin{aligned} ND &= \frac{D_d(C)}{D_{d_m}(C)} \cdot \frac{D_{d_m}(C)\Psi(C_0)}{D_{d_m}(C_0)\Psi(C)} \\ &= TMR(d,C) \cdot S_{c,p}(C) \cdot \frac{1}{S_c(C)}, \end{aligned} \tag{3.72}$$

assuming that $S_c(C) = \Psi(C)/\Psi(C_0)$, where $S_{c,p}$ is the output factor (Khan, 2010), $S_c$ is the collimator scatter factor, and *TMR* is the tissue maximum ratio. The phantom scatter factor is $S_p = S_{c,p}/S_c$. Using this result along with Equations 3.70 and 3.72 and the assumption that $OF_c = S_c$,

$$\dot{D} = (D/M)_{\text{cal}} \cdot S_c(C) \cdot S_p(C) \cdot TMR(d,C) \cdot TTF. \tag{3.73}$$

This is the standard equation for the dose rate (per MU) using *TMR* (Khan, 2010). Unfortunately, it appears that it is not quite as simple as this. The Pinnacle[3] planning reference guide (Philips, 2013) states, *"$OF_c$ is an internal normalization factor that will not match measured $S_c$ values."*[5] The definition of this quantity is therefore unclear.

The problem may lie with the assumption that $S_c$ is the ratio of the fluences in Equation 3.72). The measured value of $S_c$ is based on ion chamber readings (with buildup cap) made in free space. For a *monoenergetic* beam, ignoring attenuation in the cap,

$$\frac{Q}{m_g} = \Psi(C).\left(\frac{\mu_{en}}{\rho}\right)_{\text{air}} \cdot \left(\frac{e}{\overline{W}}\right), \tag{3.74}$$

where:

$Q$ is the ionization produced in the gas of mass $m_g$ inside the ionization chamber

$\overline{W}$ is the mean energy necessary to produce an ion pair in air

Therefore, for the case of a monoenergetic beam, the reading of the ion chamber electrometer is directly proportional to the fluence. In the case of a polyenergetic beam,

$$\frac{Q}{m_g} = \left(\frac{e}{\overline{W}}\right)\int_E \Psi_E \left(\frac{\mu_{en}}{\rho}\right) dE. \tag{3.75}$$

In this case, $S_c$ will only be the ratio of the total energy fluence provided that the spectrum of the beam remains unchanged with changing field size. This may be approximately true, but it is unlikely to be strictly true.

## 3.13 DOSE CALCULATION SPEED

Suppose that we wish to calculate the dose for a matrix consisting of $N^3$ voxels. Straightforward brute-force application of the superposition algorithm for a homogeneous medium requires the calculation of $N^3$ dose contributions (one integral) for

each voxel. Calculating the dose at all the voxels in the matrix requires $N^6$ multiplications and additions for a monoenergetic beam. A $30 \times 30 \times 30$ cm$^3$ grid with a grid spacing of 3 mm has $N = 100$ and $N^3 = 10^6$ voxels.

Let us now allow for the presence of inhomogeneities. Assume that we do not calculate and store the radiological distance between all pairs of voxels. This would be prohibitive because it would require on the order of $N^6$ stored numbers (roughly 1 TB). For each of the $N^3$ terms in the sum needed to calculate the dose to a single point, it is also necessary to perform a sum to calculate the radiological distance (in Equation 3.54). On average, this will require a sum over about $N$ terms. Therefore, $N^4$ sums are required to calculate the dose to a single point when the medium is inhomogeneous. As there are $N^3$ points in the dose matrix, approximately $N^7$ operations are required to compute the dose to the entire matrix. Brute-force dose calculation of the dose throughout the entire matrix would therefore require approximately $100^7 = 10^{14}$ multiplication and addition operations.

For the CCC algorithm, the number of operations is proportional to $MN^3$, where $M$ is the number of collapsed cones (Ahnesjö, 1989). I have made an informal test of this using the Pinnacle treatment planning system. The test starts with a patient beam calculation for a 3 mm grid size and $154 \times 121 \times 137$ voxels $= 2.55 \times 10^6$. The spatial resolution was then increased to a 2 mm grid size. As expected, the number of voxels rises to $8.70 \times 10^6$. The time for CCC beam calculations for five beams was measured with a stopwatch. The expected ratio of the time for the two grid sizes is $(8.70/2.55)^3 = 3.4$. The average ratio measured with a stopwatch was 4.1. Some of this time may have been necessary for "overhead" to set up the dose calculation.

The primary kernel accounts for charged particle transport. The range of electrons set in motion by Compton interactions is relatively short compared to the mean free path of a scattered photon. To save computation time, the primary kernel can be computed over a fine grid extending only over a short distance compared to that of the scatter kernel. The scatter kernel may be calculated with a lower resolution extending over a large distance.

As for any algorithm, the speed of pencil beam calculations depends on the specific implemetation. It is reported by Hasenbalg et al. (2007) that the AAA algorithm is 7–11 times faster than CCC.

Full C/S calculations as used in IMRT and volumetric modulation arc therapy (VMAT) still require a significant amount of CPU time even on modern computing platforms. There is recent interest in the use of graphics processing units (GPUs) to increase the speed of dose computations. Extensive efforts in this direction have been reported by Jacques et al. (2010). The very brief description here is taken from that paper. Computer clock speeds are no longer doubling every 18 months, but the number of processing cores is increasing. GPUs are gaining the flexibility to run generalized algorithms. To exploit this, serial algorithms must be converted to parallel algorithms. Jacques et al. have adapted an improved C/S calculation that has been timed on a 3 GHz Pentium 4 with an NVIDIA GeForce GTX 280 GPU. Increases in dose calculation speeds of up to a factor of 100 have been measured. The Mobius3D CCC commercial

treatment planning software runs on a GPU. This company claims speed gains of up to a factor of 200 (Childress et al., 2014), along with a number of other improvements.

## 3.14 PINNACLE TREATMENT PLANNING SYSTEM[6]

The Philips Pinnacle treatment planning system is a widely used C/S-based software package. The user has a choice of three photon dose computation algorithms: fast convolve, adaptive convolve, or cc convolution (collapsed cone convolution). CCC is the most accurate of these. The adaptive convolution superposition algorithm is designed to speed up dose calculations by adaptively varying the spatial resolution of the dose grid. In regions where the terma gradient is high, the resolution is increased. In regions where it is low, a coarse grid is used. This strategy reduces the computation time by a factor of 2–3 (McNutt, 2002). For the most accurate MU calculations, full CCC is still recommended. The fast convolve option uses "fewer ray directions" for the scatter calculation than are used for adaptive convolve. It is presumed that this means fewer collapsed cones. This algorithm should not be used for MU calculations.

The CCC algorithm uses a zenith bin count of 10 and an azimuthal bin count of 8. These are presumably the numbers of angular values $\theta_m$ and $\varphi_n$, respectively. These parameters cannot be modified by the user. The values of these parameters suggest that there are up to 80 collapsed cone beam directions.

According to Huang et al. (2013), the kernels used by Pinnacle were computed by Mackie et al. (1988). Monoenergetic photons were forced to interact at the center of a sphere of water with a radius of 60 cm.

Pinnacle uses the method of Papanikolaou et al. (1993) to compute the energy-dependent kernel (see Section 3.4), wherein 15–20 discrete energies (Starkschall et al., 2000) are used for the integration. The beam energy spectrum is determined from measured depth dose data (McNutt, 2002). The software starts with a published spectrum and uses this to calculate depth–dose curves that are then compared to measurements. The initial spectrum is then adjusted until the calculated and measured depth–dose curves agree as closely as possible (Figure 3.4).

The electron contamination dose component is modeled in the Pinnacle (McNutt, 2002) treatment planning system as a function that declines linearly inward from the surface at first and then, beginning at some depth, declines exponentially.

Beam commissioning with the Pinnacle TPS is discussed by Starkschall et al. (2000). These authors describe the parameters that define a beam model in Pinnacle.

## 3.15 CONCLUSION

It is sometimes claimed that model-based dose calculations are primary calculations of dose. They are not. They are based on computations using kernels that must be precomputed. In addition, implementation is generally based on numerous

approximations and assumptions, beyond those normally associated with numerical calculations (i.e., discretization of continuous variables). The term *convolution* is misleading. No realistic patient dose calculations are based on mathematical convolution. They are rather based on superposition computations. In this author's opinion, direct calculations of dose are either Monte Carlo or Boltzmann transport algorithms (see Chapter 4). These are truly first principles calculations of absorbed dose. It is of course true that even these algorithms are based on some approximations.

It is difficult to make sweeping generalizations because of the many different commercial implementations of C/S algorithms; nevertheless, we can list some common assumptions and approximations that have been or are used (Huang et al., 2013) in *some* implementations. As Huang et al. (2013) point out, the implications of these assumptions and approximations are not always totally clear.

1. Kernel calculations assume an infinite medium and therefore do not take account of boundaries in the medium. This is discussed in Section 3.10.
2. Even when the medium is considered heterogeneous, it is still taken as water equivalent—meaning that the number of electrons per gram for all materials is assumed to be the same as for water. Table 3.2 shows that the electron density for cortical bone is 6% different than that for water.
3. Photons are monodirectional, meaning that they all diverge radially from the source in the target. This is implicitly assumed in Equation 3.7, which is based on a point source. This assumption is untrue for photons scattered in the head.
4. CCC approximation. This is discussed in Section 3.11.
5. The use of a spatially invariant kernel calculated for water. This kernel is scaled in the presence of inhomogeneities, as discussed in Section 3.8. It is not clear that the primary dose contribution or the contribution from multiply scattered photons can be computed in this manner. For the primary kernel, the use of radiological path length or, equivalently, the mean electron density between the interaction site and deposition site assumes that secondary electrons travel in straight lines and have a range that is given by the radiological distance. Electrons are easily scattered, and therefore this is not a good assumption. This can lead to discrepancies near interfaces.
6. The use of an energy-averaged kernel derived from energy averaging at a single location. An effective mean kernel is calculated using monoenergetic kernels along with the beam spectrum at some convenient location in air. This ignores the effect of spectrum changes throughout the patient, such as beam hardening. Some implementations add a correction factor for beam hardening.

The Imaging and Radiation Oncology Core (IROC) Houston Quality Assurance Center (formerly known as the Radiological Physics Center [RPC]) has presented a list (IROC, 2014) of treatment planning system algorithms tested by irradiation of their

lung phantom. This list is divided into "acceptable" and "unacceptable" categories for use in calculating dose within a heterogeneous medium. The list has been approved by the Radiation Therapy Oncology Group (RTOG) quality assurance group. Many of the algorithms in the unacceptable category are pencil beam algorithms. One exception is the Eclipse AAA algorithm, which is considered acceptable.

Comparisons of a number of treatment planning algorithms, including many C/S to Monte Carlo calculations, have been studied by Knöös et al. (2006).

## PROBLEMS AND QUESTIONS

1. What is the difference between collision kerma and terma?

2. What is the difference between a convolution and a superposition?

3. Consider a one-dimensional convolution calculation in which

$$D(x) = \int_0^\infty T(x')A(x - x')dx',$$

where $T(x') = T_0 e^{-\mu_0 x'}$, $A(x - x') = A_0 e^{-\mu|x-x'|}$, and $\mu > \mu_0$.

   a. Calculate $D(x)$.

   b. What is the value of $x$ for which $D(x)$ is a maximum?

   c. Plot $D(x)$ versus $x$, normalized to a maximum value of 100%, for $\mu_0 = 0.1$ cm$^{-1}$ and $\mu = 0.4$ cm$^{-1}$ for $x$ ranging from 0 to 30 cm.

4. Verify the validity of Equation 3.20 by substituting Equations 3.21 through 3.23 into it and thereby recovering Equation 3.19.

5. Derive Equation 3.53 for $A_s(\vec{r} - \vec{r}')$ for an inhomogeneous medium from Equations 3.50 through 3.52.

6. What is the difference between a pencil beam algorithm and a full point kernel C/S dose calculation? Why are pencil beam calculations faster than point C/S calculations? Why are they generally less accurate?

7. Why are simple pencil beam model implementations poor at accounting for heterogeneities?

8. The pencil beam kernel calculated by Ahnesjö et al. (1992) is of the form given in Equation 3.60.

   a. Derive an expression for the dose per unit energy fluence at depth $z$ for a circular field of radius $R$ irradiated by a uniform fluence (i.e., the beam profile is completely flat).

   b. For an 18 MV beam at a depth of 10 cm, the pencil beam fitting parameters for Equation 3.60 are $A_z = 5.48 \times 10^{-3}$ cm/g, $a_z = 2.49$ cm$^{-1}$, $B_z = 9.19 \times 10^{-5}$ cm/g, and $b_z = 0.219$ cm$^{-1}$. Calculate and plot the phantom scatter factor $S_p$ at a depth of $z = 10$ cm for circular fields with radii ranging from 1.0 to 22.0 cm. Include the radius 5.64 cm and normalize $S_p$ to 1.00 at this radius. A square field of $10 \times 10$ cm$^2$ is equivalent to a circular field of radius 5.64 cm.

   c. Table 3.3 contains measured values of $S_p$ at a depth of 10 cm for an 18 MV beam. Plot these points on the graph in (b).

9. The fluence spectrum of three linac photon beams is given in Figure 3.4 for nominal accelerating potentials (NAPs) of 6, 10, and 18 MV. The data used to make these graphs are given in Table 3.4. The data represent the differential energy fluence in some arbitrary units.

   a. Refer to this table and calculate the average energy of these beams. This may be done using a spreadsheet and performing the integration using the trapezoidal rule.

   b. How accurate is the rule of thumb that the average photon energy is the NAP divided by 3?

10. Do C/S calculations correctly account for build-down? Consider both point kernel and pencil kernel. Explain.

11. Consider a patient dose calculation in an $N \times N \times N$ grid. Roughly how much faster (by what factor) is an FSPB dose calculation (pencil beam cross section is $N^2$) than a CCC calculation?

12. In the CCC approximation, the grid size is halved (in all directions). Assuming that the number of collapsed cones remains the same, by what factor will the beam computation time rise?

**TABLE 3.3 Phantom Scatter Factors (18 MV)**

| $R$ (cm) | 1.7 | 2.8 | 4.5 | 5.6 | 7.9 | 11.3 | 14.1 | 16.9 | 22.6 |
|---|---|---|---|---|---|---|---|---|---|
| $S_p$ | 0.907 | 0.961 | 0.988 | 1.00 | 1.019 | 1.036 | 1.047 | 1.055 | 1.064 |

| TABLE 3.4 | Linac Energy Spectra, Relative Differential Fluence $\Phi_E$ | | |
|---|---|---|---|
| **Energy** | **6 MV** | **10 MV** | **18 MV** |
| 0.1 | 0.056 | 0.032 | 0.016 |
| 0.2 | 0.102 | 0.06 | 0.03 |
| 0.3 | 0.145 | 0.086 | 0.044 |
| 0.4 | 0.183 | 0.109 | 0.056 |
| 0.5 | 0.218 | 0.131 | 0.068 |
| 0.6 | 0.25 | 0.151 | 0.079 |
| 0.8 | 0.304 | 0.185 | 0.098 |
| 1 | 0.345 | 0.212 | 0.115 |
| 1.25 | 0.381 | 0.238 | 0.132 |
| 1.5 | 0.402 | 0.257 | 0.146 |
| 2 | 0.406 | 0.274 | 0.165 |
| 3 | 0.321 | 0.256 | 0.177 |
| 4 | 0.191 | 0.205 | 0.167 |
| 5 | 0.077 | 0.147 | 0.147 |
| 6 | 0.002 | 0.096 | 0.123 |
| 8 | | 0.029 | 0.077 |
| 10 | | 0.002 | 0.043 |
| 15 | | | 0.007 |
| 20 | | | 0.001 |

13. The dose grid in a CCC dose homogeneous calculation is reduced from 3 to 2 mm. By what factor will the dose calculation time rise?

14. How are MU calculated in C/S algorithms?

# SYMBOLS

| | |
|---|---|
| $A$ | General dose calculation kernel |
| $a$ | Shift in downstream direction |
| $\overline{A}$ | Energy-averaged kernel |
| $A_c$ | Kernel for collision kerma |
| $A_p$ | Primary component of point dose kernel |
| $A_s$ | Scatter component of point dose kernel |
| $A_{sh}$ | Scatter kernel, homogeneous |
| $AW$ | Atomic weight |
| $a_z$ | Pencil beam fitting parameter |
| $A_z$ | Pencil beam fitting parameter |
| $a_\theta$ | Fitting constant for kernel |
| $A_\theta$ | Fitting constant for kernel |

| | |
|---|---|
| $b_z$ | Pencil beam fitting parameter |
| $B_z$ | Pencil beam fitting parameter |
| $b_\theta$ | Fitting constant for kernel |
| $B_\theta$ | Fitting constant for kernel |
| $C$ | Represents collimation variables such as field size |
| $D$ | Absorbed dose |
| $d$ | Physical distance in the medium |
| $\dot{D}$ | Dose per monitor unit |
| $dA$ | Differential area element |
| $d_{max}$ | Depth of maximum dose |
| $d_{rad}$ | Radiological distance in the medium |
| $d\sigma/d\Omega$ | Differential cross section |
| $E$ | Energy |
| $E_0$ | Energy of primary photon |
| $E_s$ | Energy of scattered photon |
| $E_t$ | Total energy |
| $F$ | Fourier transform of $f$ |
| $f$ | Relative off-axis fluence distribution function |
| $F_{ij}$ | Mayneord factor at $(x_i, y_j)$ |
| $F_{KN}$ | Klein–Nishina factor |
| $G$ | Fourier transform of $g$ |
| $h$ | $= \lvert\vec{r} - \vec{r}'\rvert$ |
| $K_c$ | Collision kerma |
| $M$ | Monitor units |
| $m_g$ | Mass of gas in an ion chamber |
| $N$ | Number of grid points along a particular axis |
| $N_A$ | Avogadro's number |
| $n_e$ | Relative electron density (volume) relative to water |
| $\overline{n}_e$ | Average relative electron density along defined path |
| $PB_z$ | Finite pencil beam dose per unit fluence |
| $P_z$ | Pencil kernel |
| $Q$ | Charge |
| $R$ | Radius of circular radiation field |
| $r_e$ | Classical electron radius |
| $\vec{r}$ | Position vector |
| $\vec{r}_s$ | Position vector from radiation source to surface of medium |
| $\vec{r}'$ | Position vector of interaction site |
| $\mathfrak{R}$ | Max range of secondary electrons |
| $\vec{r}_0$ | Vector from source of radiation to a plane at $z = z_0$ |
| $(S_\rho)_c$ | Mass collision stopping power |
| $t_{ij}$ | Thickness of beam modifier along a ray-line |
| $T$ | Terma (total energy released in matter) |
| $T_E$ | Energy differential terma |

| | |
|---|---|
| $T_{max}$ | Maximum kinetic energy of charged particles |
| $u$ | Distance from pencil beam |
| $w$ | Dose per energy fluence |
| $\overline{W}$ | Mean energy to produce an ion pair |
| $x$ | Lateral position coordinate |
| $y$ | Lateral position coordinate |
| $Z$ | Atomic number |
| $z$ | Vertical coordinate |
| $z_0$ | $z$ coordinate of a plane at some standard distance (100 cm) from the source of radiation |
| $z_s$ | Distance to the surface of the medium (from source) |
| $\alpha$ | $E_0/(m_0c^2)$ |
| $\beta$ | TCPE coefficient |
| $\delta$ | Dirac delta function |
| $\Delta D_{pb}$ | Dose contribution from FSPB |
| $\theta$ | Scattering angle |
| $\mu$ | Linear attenuation coefficient |
| $\mu^0$ | Linear attenuation coefficient for water |
| $\mu_{en}$ | Energy absorption coefficient |
| $\mu_{en-s}$ | Energy absorption coefficient for singly scattered photons |
| $\mu_s$ | Linear attenuation coefficient for singly scattered photons |
| $\mu_{tr}$ | Energy transfer coefficient |
| $\mu'$ | Slope of $K_c$ |
| $\mu_s^0$ | Linear attenuation coefficient for scattered photons in water |
| $\overline{\mu}$ | Value of $\mu$ averaged over beam spectrum |
| $\nu$ | Frequency |
| $\rho$ | Charge density |
| $\rho$ | Mass density |
| $\rho^0$ | Mass density of water |
| $\rho_e$ | Volume electron density |
| $\rho_e^0$ | Volume electron density for water |
| $\sigma_e$ | Total Compton cross section |
| $\Phi$ | Particle fluence |
| $\Phi_E$ | Differential particle fluence |
| $\Phi_i$ | Incident fluence |
| $\Phi_s$ | Scattered fluence |
| $\psi$ | Electrostatic potential |
| $\Psi$ | Energy fluence of *primary* photons |
| $\Psi_0$ | Primary energy fluence on the central axis at distance $z_0$ from source |
| $\Psi_E$ | Differential energy fluence |
| $\Psi_{hsc}$ | Primary energy fluence scattered in the head |
| $\Psi_T$ | Total energy fluence of all photons in the medium: primary plus scatter |

$\Psi_{usc}$ Primary energy fluence unscattered in the head
$\Omega$ Solid angle
$\mathfrak{F}^{-1}$ Inverse Fourier transform
$\otimes$ Mathematical operator representing convolution
$\mathfrak{F}^{-1}$ Inverse Fourier transform

# REFERENCES


AAA photon dose calculation model in Eclipse™. Palo Alto, CA: Varian Medical Systems.

Ahnesjö, A. 1989. Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media. *Med. Phys.* 16(4), 577–92.

Ahnesjö, A. 2006. Current concepts in dose calculations, slide presentation. http://www.sasro.ch/PastEvents/Lung_Seminar_2006/lung-ahn.pdf (accessed June 5, 2014).

Ahnesjö, A., P. Andreo, and A. Brahme, 1987. Calculation and application of point spread functions for treatment planning with high energy photon beams. *Acta Oncol.* 26, 49–56.

Ahnesjö, A. and M.M. Aspradakis. 1999. Dose calculation for external photon beams in radiotherapy. *Phys. Med. Biol.* 44, R99–155.

Ahnesjö, A., M. Saxner, and A. Trepp. 1992. A pencil beam model for photon dose calculation. *Med. Phys.* 19(2), 263.

Arfken, G.B., H.J. Weber, and F.E. Harris. 2013. *Mathematical Methods for Physicists*, 7th ed. Oxford: Elsevier.

Aspradakis, M.M. 1996. A study to assess and improve dose computation in photon beam therapy. PhD thesis, University of Edinburgh, Edinburgh, UK.

Attix, F.H. 1986. *Introduction to Radiological Physics and Radiation Dosimetry.* New York: Wiley.

Bloch, P. and M.D. Altschuler. 1995. Three dimensional photon beam calculations. In *Radiation Therapy Physics*, ed. A.R. Smith, 33–42. Berlin: Springer-Verlag.

Bourland, J.D. and E.L. Chaney. 1992. A finite size pencil beam model for photon dose calculations in three dimensions. *Med. Phys.* 19, 1401–12.

Boyer, A. and E. Mok. 1985. A photon dose distribution model employing convolution calculations. *Med. Phys.* 12(2), 169–77.

Boyer, A.L., Y. Zhu, L. Wang, and P. Francois. 1989. Fast Fourier transform convolution calculations of x-ray isodose distributions in homogeneous media. *Med. Phys.* 16, 248–53.

Ceberg, C.P., B.E. Bjärngard, and T.C. Zhu. 1996. Experimental determination of the dose kernel in high energy x-ray beams. *Med. Phys.* 23, 505.

Childress, N., E. Stevens, D. Eklund, and M. Zhang. 2012. Mobius3D white paper: Dose calculation algorithm, rev. 0. Houston, TX: Mobius Medical Systems LP.

Chui, C. and R. Mohan. 1988. Extraction of pencil beam kernels by the deconvolution method. *Med. Phys.* 15, 138.

Hasenbalg, F., H. Neuenschwander, R. Mini, and E.J. Born. 2007. Collapsed cone convolution and analytical anisotropic algorithm dose calculations compared to VMC++ Monte Carlo simulations in clinical cases. *Phys. Med. Biol.* 52, 3679–91.

Huang, J.Y., D. Eklund, N.L. Childress, R.H. Howell, D. Mirkovic, D.S. Followill, and S.F. Kry. 2013. Investigation of various energy deposition kernel refinements for the convolution/superposition method. *Med. Phys.* 40(12), 121721.

IROC Houston Quality Assurance Center, Houston, TX. 2014. http://rpc.mdanderson.org/RPC/home.htm.
Jacques, R., R. Taylor, J. Wong, and T. McNutt. 2010. Towards real-time radiation therapy: GPU accelerated superposition/convolution. *Comput. Methods Programs Biomed.* 98, 285–92.
Johns, H.E. and J.R. Cunningham. 1983. *The Physics of Radiology*, 4th ed. Springfield: Charles C. Thomas.
Khan, F.M. 2010. *The Physics of Radiation Therapy*, 4th ed. Philadelphia: Wolters Kluwer.
Knöös, T. and B. McClean. 2008. Dose calculation algorithms in 3D CRT and IMRT. AAPM presentation. College Park, MD: American Association of Physicists in Medicine (AAPM). http://www.aapm.org/meetings/amos2/pdf/35-9832-6508-974.pdf (accessed June 11, 2014).
Knöös, T., E. Wieslander, L. Cozzi, C. Brink, A. Fogliata, D. Albers, H. Nystrom, and L. Lassen. 2006. Comparison of dose calculation algorithms for treatment planning in external photon beam therapy for clinical situations. *Phys. Med. Biol.* 51, 5785–807.
Mackie, T.R., A.F. Bielajew, D.W.O. Rogers, and J.J. Battista. 1988. Generation of photon energy deposition kernels using the EGS Monte Carlo code. *Phys. Med. Biol.* 33(1), 1–20.
Mackie, T.R., H.H. Liu, and E.C. McCullough. 2007. Treatment planning algorithms: Model based photon dose calculations. In *Treatment Planning in Radiation Oncology*, ed. F. Khan, 2nd ed., 63–77. Philadelphia: Lippincott, Williams and Wilkins.
Mackie, T.R., G.H. Olivera, P.J. Reckwerdt, and D.M. Shephard. 2000. Convolution/superposition photon dose algorithm. In *General Practice of Radiation Oncology Physics in the 21st Century*, ed. A.S. Shiu and D.E. Mellenberg, 39–56. Madison, WI: Medical Physics Publishing.
Mackie, T.R., P. Reckwerdt, T. McNutt, M. Gehring, and C. Sanders. 1996. Photon beam dose computations. In *Teletherapy: Present and Future*, ed. T.R. Mackie and J.R. Palta, 103–35. Madison, WI: Advanced Medical Publishing.
Mackie, T.R., J.W. Scrimger, and J.J. Battista. 1985. A convolution method of calculating dose for 15 MV x-rays. *Med. Phys.* 12(2), 188–96.
McNutt, T. 2002. Dose calculations. Pinnacle$^3$ White Paper. Cleveland, OH: Philips Medical Systems.
Metcalfe, P., T. Kron, and P. Hoban. 2007. *The Physics of Radiotherapy X-Rays and Electrons*. Madison, WI: Medical Physics Publishing.
Mohan, R. and C. Chui. 1987. Use of fast Fourier transforms in calculating dose distributions for irregularly shaped fields for three-dimensional treatment planning. *Med. Phys.* 14, 70–77.
Papanikolaou, P., T.R. Mackie, C. Meger-Wells, M. Gehring, and P. Reckwerdt. 1993. Investigation of the convolution method for polyenergetic spectra. *Med. Phys.* 20, 1327–36.
Philips. 2013. Philips Pinnacle$^3$. Reference guide, release 9.6, p. 30.
Sievenen, J., W. Ulmer, and W. Kaissl. (No date). AAA photon dose calculation model in Eclipse. RAD #7170A. Palo Alto, CA: Varian Medical Systems.
Starkschall, G., R.E. Steadham, R.A. Popple, S. Ahmad, and I.I. Rosen. 2000. Beam commissioning methodology for three dimensional convolution/superposition photon dose algorithm. *J. Appl. Clin. Med. Phys.* 1(1), 8–27.

# ENDNOTES

1. Remarks about commercial treatment planning systems are likely to become out of date quickly.

2. We will not consider flattening filter-free beams here.
3. $\sigma_e$ is independent of $Z$.
4. Note that exponents appear to be missing from some of the 18 MV data in table II of Ahnesjö et al. (1992).
5. Italics are mine.
6. Some of the material in this section is repeated from earlier sections.

# 4

# DETERMINISTIC RADIATION TRANSPORT

## *A Rival to Monte Carlo Methods*

## 4.1 INTRODUCTION

The radiation transport equation provides a deterministic solution for the differential fluence of particles in a radiation field. This is in contrast to the stochastic solution provided by the Monte Carlo method. The differential fluence can be used to compute the absorbed dose in the medium. The transport equation, sometimes called the Boltzmann transport equation (BTE), is an integro-differential equation. As we will see, the deterministic radiation transport approach involves "book keeping" in phase space.

The common view of dose calculations held by many physicists is that there is Monte Carlo, the gold standard, and then there is everything else. If Monte Carlo is the gold standard, then perhaps the BTE is the silver standard. If it were possible to solve the transport equation in its full generality, the solution would provide all possible macroscopic information about the radiation field without concern about statistics and microscopic fluctuations. The catch is that the transport equation is virtually impossible to solve in its full generality. It therefore becomes a question of the assumptions and approximations that are necessary to solve this equation.

The purpose of the development given here is to help the reader understand the deterministic transport approach to dose calculations. This includes both photons and electrons. In the last few years, the medical physics community has become interested in this approach to dose calculations. A widely used commercial treatment planning system now offers this method of photon dose calculations as an option.[1] The BTE can be used for both external beam and brachytherapy dose calculations.

Electron pencil beam calculations are generally based on the solutions of the Fermi–Eyges equation, which we will derive from the BTE. In addition to these practical applications, the discussion here also provides a deeper insight into the meaning of a number of radiation quantities.

We begin in Section 4.2 with a large number of definitions (for which we ask the reader's patience). The most difficult aspect of this subject is keeping track of the many different variables and their integrals. To aid in this, we have provided a table with symbol definitions at the end of the chapter. In Section 4.3, we introduce the differential fluence and demonstrate that it is a constant along a ray line in the absence of sources or scattering of radiation, regardless of the distance from the source. In Section 4.4, we show how the dose can be calculated from the differential fluence and other radiometric quantities. The linearized BTE is derived in Section 4.5. Some simple applications of the BTE are given to show that it leads to familiar results. In Section 4.6, the case in which the primary radiation consists of charged particles is considered, and the simplification provided by the continuous slowing-down approximation (CSDA) approximation is discussed in Section 4.7. In Section 4.8, the more difficult case of indirectly ionizing radiation is analyzed. Section 4.9 consists of a discussion of the efficacy of BTE-based dose calculations for radiation therapy treatment planning. The Fermi–Eyges theory of electron transport is derived from the BTE in Section 4.10. There are commercial treatment planning systems that use the Fermi–Eyges theory to calculate electron dose distributions. The Fermi–Eyges theory is also used by the RayStation treatment planning system to calculate dose distributions from proton beams.

The single best reference for this topic is NCRP Report No. 108: "Conceptual Basis for Calculations of Absorbed-Dose Distributions" (1991). Other useful references are "Review of Electron Beam Therapy Physics" (Hogstrom and Almond, 2006), Notes 8 (online notes), and "A Review on the Use of Grid-Based Boltzmann Equation Solvers for Dose Calculation in External Beam Treatment Planning" (Kan et al., 2013).

## 4.2 ABSORBED DOSE, KERMA, AND FLUENCE

In this section, some useful radiation measurement quantities are defined that the reader is likely familiar with (see Attix, 1986). These definitions are intended to be a review and to establish notation.

The absorbed dose is defined by

$$D = \frac{d\bar{\varepsilon}}{dm}, \tag{4.1}$$

where $d\bar{\varepsilon}$ is the mean energy absorbed in a mass element $dm$. The quantity $\bar{\varepsilon}$ is defined as

$$\bar{\varepsilon} = E_{in} - E_{out} + \sum Q, \tag{4.2}$$

where $E_{in}$ (and $E_{out}$) equals the sum of the energies of all charged and uncharged ionizing particles that enter (leave) the volume, *excluding rest mass energies.* For photons $E = h\nu$ and for all other particles, $E$ is the kinetic energy. The term $\Sigma Q$ is defined as the algebraic sum of all changes in the rest energy of nuclei and particles that occur in the volume. $Q$ is taken as positive ($Q > 0$) when there is a decrease in rest energy (mass converted to energy) and negative when there is an increase in the rest energy (energy converted to mass). As an example of this, in the pair production process, $\Sigma Q = -2m_0c^2$. For positron annihilation, a positron and an electron combine, leaving two annihilation photons, $\Sigma Q = +2m_0c^2$. The units of absorbed dose are J/kg and 1 J/kg = 1 Gy.

The term *kerma* is an acronym for kinetic energy released in matter. Kerma is defined for neutral particles (photons or neutrons) as follows:

$$K = \frac{d\varepsilon_{tr}}{dm}, \tag{4.3}$$

where $d\varepsilon_{tr}$ is the energy transferred to charged particles from uncharged particles in mass element $dm$. More specifically $\varepsilon_{tr}$ is defined as

$$\varepsilon_{tr} = (E_{in})_u - (E_{out})_u^{nonr} + \sum Q, \tag{4.4}$$

where:

$(E_{in})_u$ is the sum of all the energies of the uncharged particles entering the volume

$(E_{out})_u$ is the sum of all the energies of the uncharged particles leaving the volume except those that originated as a result of radiative losses of kinetic energy by charged particles in the volume

Radiative losses involve conversion of charged particle *kinetic* energy to photon energy. There are two processes by which this can occur: (1) bremsstrahlung and (2) in-flight annihilation of positrons. In the latter case, only the kinetic energy of the positron at the instant of annihilation is considered a radiative energy loss. The kerma is simply the kinetic energy transferred from uncharged particles to charged particles per unit mass. The units of kerma are the same as the units of absorbed dose.

The kerma can be split into two parts, depending on whether the energy imparted is expended in collisional interactions or radiative interactions (bremsstrahlung):

$$K = K_c + K_r, \tag{4.5}$$

where:

$K_c$ is the collision kerma

$K_r$ is the radiative kerma

We define the net energy transferred as

$$\varepsilon_{tr}^{n} = \varepsilon_{tr} - E_{u}^{tr}, \tag{4.6}$$

where $E_{u}^{tr}$ is the energy lost by radiative processes by the charged particles that originated in a specified volume, regardless of where the radiative losses occur. The collision kerma is then defined as

$$K_c = \frac{d\varepsilon_{tr}^{n}}{dm}. \tag{4.7}$$

The collision kerma is the net energy transferred to charged particles per unit mass by uncharged particles that originated in a specified volume, excluding both radiative losses and energy transferred from one charged particle to another. The radiative kerma can now be defined as

$$K_r = K + K_c. \tag{4.8}$$

We now define the total fluence, which we will need in the subsequent development. Let $dN$ be the number of rays striking an infinitesimally small sphere of great circle area $dA$ (Figure 4.1) centered on a point P in some arbitrary but specified time interval. The total fluence at point P is defined as

$$\Phi_T(\vec{r}) = \frac{dN}{dA}. \tag{4.9}$$

The units of fluence are $m^{-2}$.

The energy fluence, $\Psi$, is defined as

$$\Psi(\vec{r}) = \frac{dR}{dA}, \tag{4.10}$$

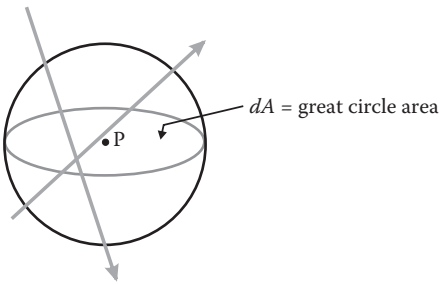



FIGURE 4.1 Definition of total particle fluence at point P. The diagram shows two crossing rays and the great circle area.

where $dR$ is the radiant energy incident on a sphere of great circle area $dA$. The units of energy fluence are J/m$^2$.

The energy distribution of the fluence, $\Phi_E$, and the energy fluence, $\Psi_E$, are given by

$$\Phi_E(\vec{r},E)=\frac{d\Phi_T}{dE} \tag{4.11}$$

and

$$\Psi_E(\vec{r},E)=\frac{d\Psi}{dE}, \tag{4.12}$$

where:

$d\Phi_T$ is the fluence of particles having energy between $E$ and $E + dE$

$d\Psi$ is the energy fluence of these particles

For monoenergetic radiation, the relationship between these two distributions is

$$\Psi_E = E\Phi_E. \tag{4.13}$$

The kerma can be written in terms of $\Phi_E$ as follows:

$$K=\int \Phi_E E\frac{\mu_{tr}}{\rho}dE, \tag{4.14}$$

where:

$\mu_{tr}/\rho$ is the mass energy transfer coefficient

$E$ is the kinetic energy of the uncharged particles

The concept of charged particle equilibrium (CPE) will be useful later in the text. A given specified volume is said to be in CPE if for every type and energy of charged particle leaving the volume, an identical charged particle with the same energy enters the volume.

The stopping power for charged particles of a particular specified type (e.g., electrons) and energy is the energy lost per unit path length (on average) in a specified medium. The stopping power is often written as

$$S=-\frac{dE}{dx}. \tag{4.15}$$

The mass stopping power, $S_\rho$, is the stopping power divided by the mass density. Loss of energy by electrons traversing a medium occurs as a result of collisions and

radiative interactions (bremsstrahlung), and therefore the mass stopping power may be divided into these two contributions:

$$S_\rho = -\frac{1}{\rho}\frac{dE}{dx} = \left(S_\rho\right)_{col} + \left(S_\rho\right)_{rad}, \tag{4.16}$$

where subscripts *col* and *rad* indicate collisional and radiative, respectively.

## 4.3 DIFFERENTIAL FLUENCE

We begin by defining a quantity called the differential fluence. Consider the infinitesimal surface element centered on point P, as shown in Figure 4.2. The vector $d\vec{S}$ describes this area element. The direction of $d\vec{S}$ is perpendicular to the surface, and the magnitude of the area is $dS$. Radiation may traverse the area element in Figure 4.2 in all different directions. Let us consider radiation passing through the area element that is only traveling into the solid angle $d\Omega$ centered on the unit vector $\hat{\Omega}$. Furthermore, let us restrict consideration to radiation having energy in the range from $E$ to $E + dE$.

The net number of particles transported by radiation in an energy interval from $E$ to $E + dE$ across an element of area $d\vec{S}$ into the solid angle $d\Omega$ about the direction specified by unit vector $\hat{\Omega}$ (in some arbitrary specified time interval, but *not* per unit time) at point P is

$$dN_{\Omega,E} = \Phi_{\Omega,E}\hat{\Omega}\cdot d\vec{S}\, d\Omega\, dE, \tag{4.17}$$

where $\hat{\Omega}\cdot d\vec{S} = \cos\theta' dS$. Equation 4.17 is an implicit definition for the differential fluence $\Phi_{\Omega,E}$. This quantity is not given a name in the International Commission on Radiation Units (ICRU) description of radiometric quantities (ICRU, 2011). The subscripts $\Omega$ and $E$ are to remind the reader that this quantity depends on direction $\hat{\Omega}$ and kinetic energy $E$. In Equation 4.17, $dN_{\Omega,E}$ is the net number of particles: $\hat{\Omega}\cdot d\vec{S}$ can be negative.

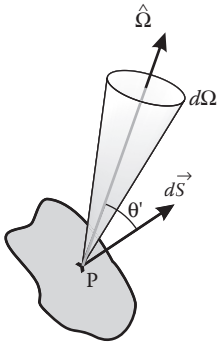



FIGURE 4.2 Diagram used to define the differential fluence $\Phi_{\Omega,E}$. The vector $d\vec{S}$ is perpendicular to the infinitesimal area element and $dS$ is the area. We consider radiation passing through this area element traveling into the solid angle $d\Omega$ centered on the unit vector $\hat{\Omega}$.

The units of differential fluence are $m^{-2}\ J^{-1}\ sr^{-1}$. Do not confuse the differential fluence (Equation 4.17) with the total fluence (Equation 4.9). The differential fluence is not the total fluence. It is only remotely related to the total fluence, and it has very different properties. The vector $\hat{\Omega}$ is the direction we are examining, not necessarily the direction of radiation flow; indeed, radiation may be moving in all directions.

The differential fluence can vary from point to point in space, and therefore, in general, $\Phi_{\Omega,E}$ depends on the location of point P in Figure 4.2. Let us suppose that the location of this point is specified by the position vector $\vec{r}$. The location of point P in spherical coordinates is given by $(r, \theta, \varphi)$ (Figure 4.3). The direction of the vector $\hat{\Omega}$ can be specified by two angles, $\theta'$ and $\phi'$, as shown in Figure 4.3. Notice that there are two coordinate systems involved. There are three coordinates associated with the location of the point of interest $\vec{r}$ and there are two coordinates ($\theta'$ and $\varphi'$) associated with the direction of particle movement $\hat{\Omega}$ at the location of point P (Figure 4.2). The vector $\hat{\Omega}$ is simply a direction at the location of point P in which some, all, or none of the radiation may flow. In general, radiation may be moving in all directions. The geometry is shown in spherical coordinates in Figure 4.3. It is important to understand the relationship between the two coordinate systems. This topic can be very confusing without a clear understanding of this.

In general, the differential fluence is a function of six variables, that is: $\Phi_{\Omega,E} = \Phi_{\Omega,E}(\vec{r}, \hat{\Omega}, E) = \Phi_{\Omega,E}(r, \theta, \varphi, \theta', \varphi', E)$. In the case of an isotropic radiation field, however, the differential fluence does not depend on direction and $\Phi_{\Omega,E} = \Phi_{\Omega,E}(\vec{r}, E)$. In a homogeneous radiation field, the differential fluence does not depend on position, and in this case: $\Phi_{\Omega,E} = \Phi_{\Omega,E}(\hat{\Omega}, E)$.

We now wish to show that the differential fluence is constant along every ray path (although it may be a different constant along different ray paths) in the absence of scattering, sources, or sinks of radiation along the ray path. Another way to state

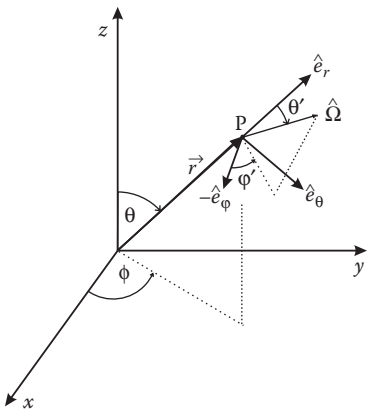



FIGURE 4.3 Two coordinate systems (spherical coordinates) used in the definition of the differential fluence. The first coordinate system ($r$, $\theta$, and $\varphi$) specifies the location of the point of observation P. The second coordinate system ($\theta'$ and $\varphi'$) defines the direction of interest. $\hat{\Omega}$ is a unit vector in this direction. Some, all, or none of the radiation may be traveling in this direction. This direction is defined in terms of $\theta'$, and $\varphi'$, $\hat{e}_r$, and $\hat{e}_\theta$ are unit vectors in the $r$ and $\theta$ directions, respectively.

this is that the differential fluence is independent of the distance between the source and observer in the absence of sources or sinks of radiation. Note how very different this is from the total fluence itself—we expect the total fluence to vary as the inverse square of the distance from a point source.

Refer to Figure 4.4. Let us write an expression for all of the particles of energy $E$ that flow out of the emitter from the area $dS_1$ centered on point $P_1$ and into the receiver (both assumed infinitesmal in size) centered on point $P_2$. The solid angle subtended by the receiver from the perspective of point $P_1$ is $d\Omega_1$. We assume that there is no scattering and no sources or sinks of radiation. We also assume that the radiation travels in straight lines. Under these circumstances, all of the particles that flow out of the emitter at point $P_1$ and into solid angle $d\Omega_1$ must flow through the receiver. This number is given by

$$dN_1 = \Phi_{\Omega,E}(1)\hat{\Omega}_1 \cdot d\vec{S}_1 d\Omega_1 dE, \tag{4.18}$$

where:

$d\Omega_1 = \cos\theta_2 dS_2/R^2$

$\Phi_{\Omega,E}(1)$ is the differential fluence at point $P_1$

Now let us write an expression for the number of particles flowing out of the receiver that have originated from $dS_1$. The number of particles flowing out of the receiver *that originate from* $dS_1$ is

$$dN_2 = \Phi_{\Omega,E}(2)\hat{\Omega}_2 \cdot d\vec{S}_2 d\Omega_2 dE. \tag{4.19}$$

These particles flow out into the solid angle shown to the right of the receiver in Figure 4.4. To ascertain the size of this solid angle, we trace rays backward to surface $dS_1$. Only rays inside the cone opening up to the left of the receiver can have originated from $dS_1$ and flowed out to the right through $dS_2$. Any larger cone will include radiation that did not originate from $dS_1$ Therefore, $d\Omega_2 = \cos\theta_1 dS_1/R^2$. As $dN_l = dN_2$, it follows that $\Phi_{\Omega,E}(1) = \Phi_{\Omega,E}(2)$. Therefore, in the absence of scattering, sources, or

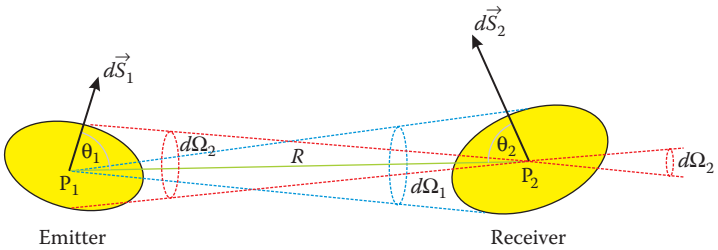



FIGURE 4.4 Diagram used to prove that the differential fluence is a constant along a ray line in the absence of scattering, sources, or sinks of radiation. The distance between point $P_1$ and $P_2$ is $R$. See text for an explanation.

sinks of radiation, the differential fluence is constant along a ray line independent of the distance from the source. Once again, this underscores the great distinction between differential fluence and total fluence itself. The total fluence clearly depends on the distance from the source.

The total fluence $\Phi_T$ can be calculated from the differential fluence $\Phi_{\Omega,E}$. In Equation 4.17, which defines $\Phi_{\Omega,E}$, $dN_{\Omega,E}$ is the *net* number of particles crossing the area $dS$ ($\hat{\Omega}\cdot d\vec{S}$ can be negative, depending on the direction of $\hat{\Omega}$). Imagine that the area $dS$ is a circle surrounding point P and that this circle is the great circle area of an infinitesmal sphere surrounding point P. The number of particles traversing this sphere is $\Phi_{\Omega,E}\, dS\, d\Omega\, dE$, and therefore the total fluence is

$$\Phi_T(\vec{r}) = \int_0^\infty \int_{4\pi} \Phi_{\Omega,E} d\Omega\, dE \tag{4.20}$$

(see Equation 4.9), where $d\Omega = \sin\theta' d\theta' d\varphi'$.

We define two other useful quantities:

$$\Phi_E(\vec{r},E) = \int_{4\pi} \Phi_{\Omega,E}\left(\vec{r},E,\hat{\Omega}\right) d\Omega$$

$$\Phi_\Omega\left(\vec{r},\hat{\Omega}\right) = \int_0^\infty \Phi_{\Omega,E}\left(\vec{r},E,\hat{\Omega}\right) dE. \tag{4.21}$$

The equation for $\Phi_E$ is equivalent to the definition given in Equation 4.11. For an isotropic radiation field (i.e., $\Phi_\Omega = \Phi_\Omega(r, \theta, \varphi)$), the total fluence is given by $\Phi_T = 4\pi\, \Phi_\Omega$.

Consider radiation that is traveling in a single direction $\hat{\Omega}(\theta_0', \varphi_0')$. In order to describe this mathematically, we define the delta function in spherical coordinates. Consider $f(\theta', \varphi')$; the delta function is defined so that

$$f(\theta_0', \varphi_0') = \int\int \delta(\theta' - \theta_0')\,\delta(\varphi' - \varphi_0')\, f(\theta', \varphi')\, d\theta' d\varphi'. \tag{4.22}$$

We wish to find the delta function in spherical coordinates such that

$$f(\theta_0', \varphi_0') = \int\int \delta\left(\hat{\Omega} - \hat{\Omega}_0\right) f(\theta', \varphi') \sin\theta' d\theta' d\varphi'. \tag{4.23}$$

Comparing Equations 4.22 and 4.23, we see that

$$\delta\left(\hat{\Omega} - \hat{\Omega}_0\right) = \frac{1}{\sin\theta'}\,\delta(\theta' - \theta_0')\,\delta(\varphi' - \varphi_0'). \tag{4.24}$$

We now give an example in which we recover the inverse square law from the definition of the differential fluence. Figure 4.5 shows a small spherical source of

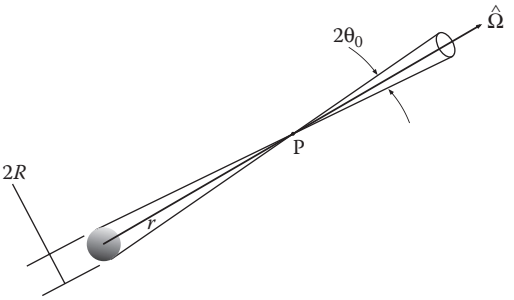



FIGURE 4.5 The inverse square law emerges by considering the total fluence of isotropic radiation originating from a small sphere of radius $R$. The distance between the center of the sphere and point P is $r$ and $r \gg R$.

monoenergetic radiation of radius $R$ located at the origin of coordinates. The sphere radiates particles isotropically. There are no sources or sinks of radiation outside the sphere. An observation point lies at location P at a distance of $r$ from the origin, where $r \gg R$. Let us compute $dN/dS$, the number of particles per unit area crossing point P. From the perspective of point P, the sphere subtends an angle $2\theta_0 = 2R/r$. The differential fluence must be zero for $\theta' > \theta_0$ and $\theta' < -\theta_0$ (see Figure 4.5), for there is no radiation approaching P from along these directions. Let us assume that $\Phi_\Omega(\vec{r}, \theta') = \Phi_{\Omega,0}$, a constant, for $-\theta_0 < \theta' < \theta_0$.

The total fluence at point P is given by

$$\Phi_T(\mathrm{P}) = \Phi_{\Omega,0}\int_0^{2\pi}\int_0^{\theta_0}\sin\theta' d\theta' d\varphi' = \pi\Phi_{\Omega,0}(1-\cos\theta_0) \approx \frac{\pi\Phi_{\Omega,0}R^2}{r^2}; \tag{4.25}$$

thus, we arrive at an inverse square law for the total fluence, as we expect.

## 4.4 CALCULATION OF DOSE FROM FUNDAMENTAL RADIOMETRIC QUANTITIES

We wish to derive an equation for the absorbed dose in terms of fundamental radiometric quantities. We will do this by considering conservation of energy in an arbitrary volume $V$. The *net* kinetic energy of particles flowing into the volume must equal the energy absorbed in the volume minus the kinetic energy of particles created in the volume and minus any decrease in rest mass energy of the matter in the volume (see Equation 4.2). It is possible that particles may be created inside $V$, for example, through radioactive decay. The quantity $S_0\left(\vec{r},E,\hat{\Omega}\right)d^3r\,dE\,d\Omega$ is the expected total number of particles produced in volume $d^3r$ at location $\vec{r}$, with energy in the range from $E$ to $E + dE$ traveling in the direction $\hat{\Omega}$.

An expression for conservation of energy follows:

$$-\oint d\vec{S}\cdot\int dE\int \hat{\Omega}E\Phi_{\Omega,E}d\Omega=\int\frac{d\bar{\varepsilon}}{dV}d^3r-\iiint ES_0\left(\vec{r},E,\hat{\Omega}\right)dEd\Omega d^3r-\int\frac{dQ}{dV}d^3r, \quad (4.26)$$

where the term on the left represents the net kinetic energy flowing *into* the volume, $d\bar{\varepsilon}/dV$ is the energy absorbed per unit volume, and $dQ/dV$ is the decrease in rest mass energy per unit volume. The surface integral on the left-hand side of Equation 4.26 can be transformed into a volume integral by using the divergence theorem; thus,

$$\int_V d^3r\left\{\left[\int dE\int d\Omega E\left(\hat{\Omega}\cdot\vec{\nabla}\Phi_{\Omega,E}-S_0\right)\right]-\frac{dQ}{dV}+\frac{d\bar{\varepsilon}}{dV}\right\}=0, \quad (4.27)$$

along with the result that $\vec{\nabla}\cdot\left[\hat{\Omega}\Phi_{\Omega,E}\right]=\hat{\Omega}\cdot\vec{\nabla}\Phi_{\Omega,E}$, provided that $\hat{\Omega}$ is a constant vector.

The volume $V$ is arbitrary, and therefore the integrand in Equation 4.27 must vanish. Let us also convert the volume element into a mass element by noting that the mass $dm = \rho\, d^3r$, where $\rho$ is the mass density. We can now write an expression for the dose:

$$D(\vec{r})=\frac{1}{\rho}\int dE\int d\Omega E\left[S_0\left(\vec{r},E,\hat{\Omega}\right)-\hat{\Omega}\cdot\vec{\nabla}\Phi_{\Omega,E}\left(\vec{r},E,\hat{\Omega}\right)\right]+\frac{dQ}{dm}. \quad (4.28)$$

Equation 4.28 provides a formal relationship between the absorbed dose $D(\vec{r})$ and the field quantities $S_0$ and $\Phi_{\Omega,E}$. The latter quantity must include *all types of particles* in the radiation field, both charged and uncharged. The fluence distribution, $\Phi_{\Omega,E}$, can be calculated from the transport equation. We therefore now turn our attention to a derivation of this equation.

## 4.5 TRANSPORT EQUATION

In radiation therapy, we are generally not concerned with time dependence. For this reason, we will skip the time-dependent form of the transport equation and proceed directly to the time-independent version. The transport equation is derived by considering the balance of the particle number in a phase space element $dE\, d\Omega$. Thus, we concentrate on particles in a volume $V$, having an energy in the range between $E$ and $E + dE$ and that are traveling into solid angle $d\Omega$ centered about direction $\hat{\Omega}$. Consider an arbitrary volume $V$; the net change over an arbitrary but specified time

interval in the number of particles *of a particular type*[2] in the phase space element $dE\,d\Omega$ is given by

$$\oint \Phi_{\Omega,E}\,\hat{\Omega}\cdot d\vec{S} = \int_V \left(\frac{\delta N_{\Omega,E}}{\delta V}\right)_{coll} d^3r + \int_V S_0\left(\vec{r},E,\hat{\Omega}\right)d^3r, \tag{4.29}$$

where the integrand in the first term on the right side of this equation represents a change per unit volume in the particles of the given type that occur as a result of collisions. The quantity $S_0$ in Equation 4.29, in this context, refers not only to "true" sources of particles, but also to the creation of particles governed by Equation 4.29 from other types of radiation that undergo a transformation to the first type. An example is provided by bremmstrahlung production. If the particles in Equation 4.29 are photons, those photons set electrons in motion. The electrons can in turn undergo bremsstrahlung photon production. Those bremsstrahlung photons should, at least in principle, be included in $S_0$ in Equation 4.29.

We can turn the surface integral in Equation 4.29 into a volume integral by invoking the divergence theorem. The result is that Equation 4.29 becomes

$$\int_V d^3r\left[\vec{\nabla}\cdot\left(\hat{\Omega}\Phi_{\Omega,E}\right) - \left(\frac{\delta N_{\Omega,E}}{\delta V}\right)_{coll} - S_0\right] = 0. \tag{4.30}$$

The volume $V$ is arbitrary, and therefore the integrand itself must be equal to zero:

$$\hat{\Omega}\cdot\vec{\nabla}\Phi_{\Omega,E} - \left(\frac{\delta N_{\Omega,E}}{\delta V}\right)_{coll} - S_0 = 0, \tag{4.31}$$

where we have assumed that the vector $\hat{\Omega}$ is independent of the position coordinates (i.e., can be considered a constant).

Let us consider the collision term in Equation 4.31 in more detail. Particles can be removed from the phase space element ($dE\,d\Omega$) by collisions through either absorption or out-scattering. Particles can be added to the phase space by in-scattering; that is, a particle is scattered in such a fashion that it ends up having energy in the range $E$ to $E + dE$ and is traveling in the direction $\hat{\Omega}$. We need to include all three of these contributions to the collision term in Equation 4.31.

The macroscopic cross section, $\mu(\vec{r}, E, \hat{\Omega})$, is defined as the probability per unit path length traveled that a particle at $\vec{r}$ (with energy in the range $E$ to $E + dE$, traveling in solid angle $d\Omega$ centered on $\hat{\Omega}$) will have *any* type of interaction. We assume an isotropic medium, and therefore $\mu(\vec{r}, E, \hat{\Omega}) = \mu(\vec{r}, E)$ (linear attenuation coefficient). The number of interactions per unit volume experienced at $\vec{r}$ by a particle of energy $E$ is $\mu\,\Phi_{\Omega,E}$. We now define $\mu_s(\vec{r}, E \to E', \hat{\Omega} \to \hat{\Omega}')dE\,d\Omega$ as the probability per unit length traveled that a particle with energy in the range $E$ to $E + dE$ and traveling in

direction $\hat{\Omega}$ will produce, as a result of an interaction at $\vec{r}$, a particle (including the primary itself, i.e., scattering) with energy $E'$ traveling in direction $\hat{\Omega}'$. The total macroscopic cross section can be written as a sum of an absorption term and a scattering term:

$$\mu(\vec{r},E) = \mu_a(\vec{r},E) + \int dE' \int_{4\pi} d\Omega' \mu_s\left(\vec{r},E \to E',\hat{\Omega} \to \hat{\Omega}'\right). \tag{4.32}$$

The $\mu_a$ term represents absorption. We will omit the explicit appearance of this term by considering absorption to be a special case of scattering with $E \to E' = 0$. Equation 4.32 represents removal of particles from phase space due to collisions. We can also have particles scatter in from other phase space elements, that is, $\mu_s(\vec{r}, E' \to E, \hat{\Omega}' \to \hat{\Omega})$. The collision term in Equation 4.31 can now be written as

$$\left(\frac{\delta N_{\Omega,E}}{\delta V}\right)_{coll} = -\mu(\vec{r},E)\Phi_{\Omega,E}\left(\vec{r},E,\hat{\Omega}\right) + \int dE' \int_{4\pi} d\Omega' \mu_s\left(\vec{r},E' \to E,\hat{\Omega}' \to \hat{\Omega}\right)\Phi_{\Omega',E'}\left(\vec{r},E',\hat{\Omega}'\right), \tag{4.33}$$

where the first term on the right-hand side in Equation 4.33 represents particles scattered out of the phase space element and the second term represents particles scattered in. We can substitute this collision term back into Equation 4.31:

$$\hat{\Omega}\cdot\vec{\nabla}\Phi_{\Omega,E} + \mu(\vec{r},E)\Phi_{\Omega,E} = \int dE' \int_{4\pi} d\Omega' \mu_s\left(\vec{r},E' \to E,\hat{\Omega}' \to \hat{\Omega}\right)\Phi_{\Omega',E'} + S_0\left(\vec{r},E,\hat{\Omega}\right). \tag{4.34}$$

The boundary condition at a "nonreentrant" surface is

$$\Phi_{\Omega,E} = 0 \quad \text{for} \quad \hat{\Omega}\cdot\hat{n} < 0, \tag{4.35}$$

where $\hat{n}$ is the outward directed normal vector to the surface. Equation 4.35 simply states that no particles are entering from outside the surface.

Equation 4.34 is the classical form of the transport equation. It is an integro-differential equation. This equation is linear in $\Phi_{\Omega,E}$. It is sometimes referred to as the linear BTE. This equation is for a single type of particle. If there are multiple species of particles, a BTE must be written for each type, including coupling terms (see Section 4.8).

The BTE has widespread application in a variety of fields. This equation is essentially the same as the radiative transfer equation commonly used in astrophysics to calculate the spectrum of radiation that emerges from astronomical objects, such as stars and accretions disks. The principal application is the study of the spectrum of radiation that emerges from stellar atmospheres (see Mihalas, 1978) (Figure 4.6). Astronomical observations, combined with the theory of radiative transport, have

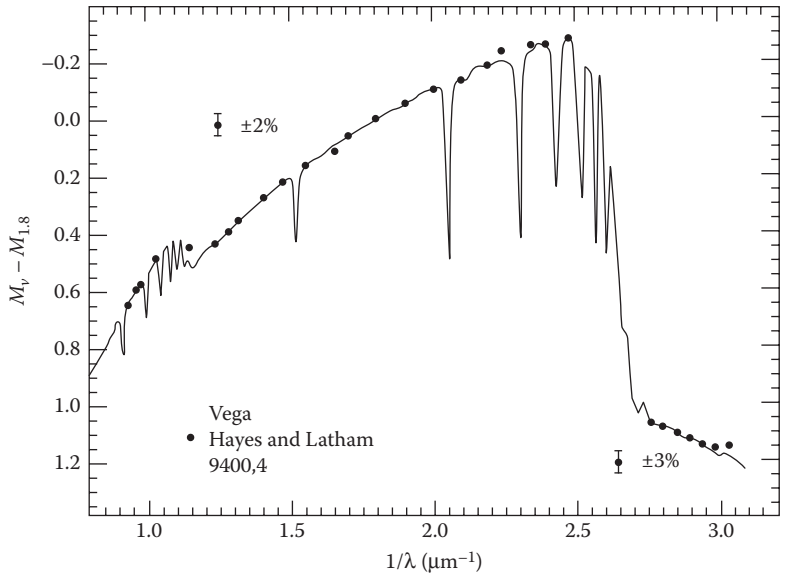



FIGURE 4.6 The spectrum of the bright star Vega. The dots are measurements and the solid curve is a "line-blanketed" model. (From Mihalas, D., *Stellar Atmospheres*, W.H. Freeman, New York, 1978.)

provided very detailed information about both the physical conditions and the composition of stellar atmospheres. The astrophysical literature pertaining to this topic is rich, detailed, and lengthy, extending back at least 100 years (Figure 4.7).

Equation 4.34 is the form of the BTE that we will use throughout. The divergence term on the left-hand side of Equation 4.30 has been written as $\hat{\Omega}\cdot\vec{\nabla}\Phi_{\Omega,E}$. This is valid *provided that* $\hat{\Omega}$ is a constant vector.[3] This is called the streaming term. It describes the behavior of the radiation when there is no scattering and no sources (streaming). The second term on the left-hand side of Equation 4.34 represents out-scattered particles (see also Equation 4.32). The first term on the right-hand side of Equation 4.34 represents in-scattered particles. The second term on the right-hand side of Equation 4.34 represents either true creation of the particles in question or creation from other particles.

Equation 4.34 is very difficult to solve because of the scattering and absorption terms. It is generally solved numerically using a discrete coordinate grid, discrete directions (discrete ordinates), and discrete energies (multigroups).

One of the simplest possible solutions of Equation 4.34 is for the case in which there is no scattering, absorption, or sources of radiation. In this instance, the streaming term in the BTE becomes

$$\hat{\Omega}\cdot\vec{\nabla}\Phi_{\Omega,E} = 0. \tag{4.36}$$

Equation 4.36 implies that the differential fluence is constant along a fixed direction $\hat{\Omega}$, as we have already deduced based on a physical argument. That is why the term

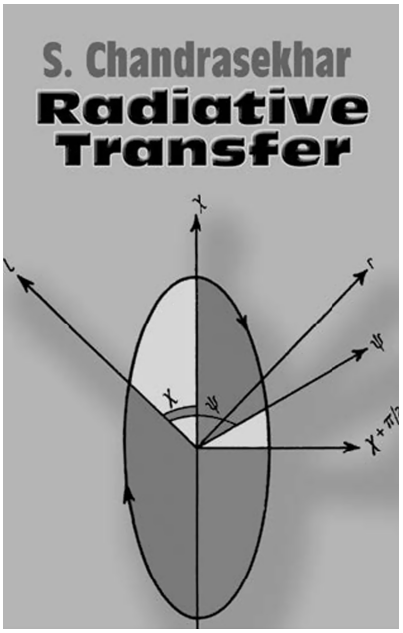



FIGURE 4.7 Monograph on radiation transport written by astrophysicist and Nobel laureate S. Chandrasekhar (2011) and first published in 1949.

on the left-hand side of Equation 4.36 is called the streaming term, because there is no change in $\Phi_{\Omega,E}$ in the direction of $\hat{\Omega}$.

We next examine another very simple solution of Equation 4.34. Let us suppose that a collimated, uniform, monoenergetic photon beam is perpendicularly incident on a flat semi-infinite homogeneous medium with attenuation coefficient μ (Figure 4.8).

We ignore in-scattering, and therefore the integral term on the right-hand side of Equation 4.34 is zero. This is the case of "narrow beam geometry." Let us also assume that there are no sources in the medium, and therefore $S_0$ is zero. The energy dependence is $\delta(E - E_0)$, and we can integrate over the energy in Equation 4.34 to get

$$\hat{\Omega} \cdot \vec{\nabla} \Phi_{\Omega} + \mu \Phi_{\Omega} = 0. \tag{4.37}$$

Assume that the radiation is traveling in the $+z$ direction at $z = 0$. This is an axisymmetric problem, and therefore $\Phi_{\Omega}(\vec{r}, \hat{\Omega}) = \Phi_{\Omega}(z, \theta')$, where $\theta'$ is the angle between the $z$-axis and $\hat{\Omega}$.

Equation 4.37 becomes

$$\cos\theta' \frac{d\Phi_{\Omega}}{dz} + \mu \Phi_{\Omega} = 0, \tag{4.38}$$

where μ is spatially constant. The boundary condition is

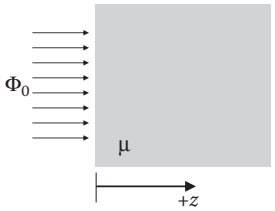



FIGURE 4.8 A monoenergetic, uniform photon beam is incident on a flat semi-infinite homogeneous medium. The surface of the medium lies at $z = 0$.

$$\Phi_\Omega(0,\theta') = A\delta\left(\hat{\Omega} - \hat{\Omega}_0\right) = \frac{A}{\sin\theta'}\delta(\theta'), \tag{4.39}$$

where $A$ is a constant. The solution of Equations 4.38 and 4.39 is

$$\Phi_\Omega(z,\theta') = A\frac{\delta(\theta')}{\sin\theta'}e^{-(\mu z/\cos\theta')}. \tag{4.40}$$

The total fluence is

$$\Phi_T(z) = \Phi_0 e^{-\mu z}, \tag{4.41}$$

where $\Phi_0 = 2\pi A$. Equation 4.41 is, of course, the well-known elementary result for simple exponential attenuation of a monoenergetic beam in narrow beam geometry (no in-scatter).

We now turn to the form of the streaming term when there is spherical symmetry—an example of this is a spherically symmetric source of radiation. In general, $\Phi_{\Omega,E}$ depends on $r$, θ, φ, θ′, and φ′. In the case of spherical symmetry, $\Phi_{\Omega,E}$ depends only on $r$ and θ′. One must use care when evaluating $\hat{\Omega}\cdot\nabla\Phi_{\Omega,E}$ in curvilinear coordinates (e.g., spherical). The difficulty arises because θ′ depends on θ (Figure 4.9). In the evaluation of the streaming term, $\hat{\Omega}$ is to be considered a constant vector (note the transition between Equations 4.30 and 4.31).

Figure 4.9 shows the geometry for the evaluation of the streaming term. The unit vector $\hat{\Omega}$ can be expressed in terms of the unit vectors shown in Figure 4.9:

$$\hat{\Omega} = \cos\theta'\hat{e}_r + \sin\theta'\hat{e}_\theta, \tag{4.42}$$

where $\theta' = \alpha - \theta$. The orientation of $\hat{\Omega}$ is fixed, and therefore α is constant. The gradient operator in spherical coordinates is

$$\vec{\nabla}\Phi_{\Omega,E} = \frac{\partial\Phi_{\Omega,E}}{\partial r}\hat{e}_r + \frac{1}{r}\frac{\partial\Phi_{\Omega,E}}{\partial\theta}\hat{e}_\theta, \tag{4.43}$$

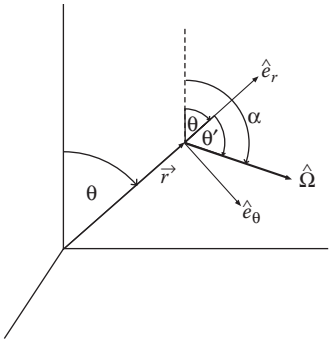



FIGURE 4.9 The geometry for evaluation of the streaming term in spherical coordinates in the spherically symmetric case. The angle θ′ will change as θ changes because $\hat{\Omega}$ is a constant vector (angle α is fixed).

and $\partial/\partial\theta' = -\partial/\partial\theta$; therefore,[4]

$$\hat{\Omega}\cdot\vec{\nabla}\Phi_{\Omega,E} = \cos\theta'\frac{\partial\Phi_{\Omega,E}}{\partial r} - \frac{\sin\theta'}{r}\frac{\partial\Phi_{\Omega,E}}{\partial\theta'}. \tag{4.44}$$

This result will be used later.

Let us consider a limiting case in which there is a source of radiation but no particle interactions (no absorption or scattering). In this case, the transfer equation becomes

$$\hat{\Omega}\cdot\nabla\Phi_{\Omega,E} = S_0\left(\vec{r},E,\hat{\Omega}\right). \tag{4.45}$$

We would like to find a solution of Equation 4.45 for $\Phi_{\Omega,E}$. Examine the geometry shown in Figure 4.10. The left-hand side of Equation 4.45 is the derivative of $\Phi_{\Omega,E}$ in the direction of $\hat{\Omega}$, that is, $d\Phi_{\Omega,E}/du$, where $u$ is the coordinate distance along the direction of $\hat{\Omega}$. To solve Equation 4.45, we need to write $S_0$ in terms of the coordinate $u$. Examining Figure 4.10, we see that $\vec{r} = \vec{r}' + u\hat{\Omega}$, and therefore Equation 4.45 becomes

$$\frac{d\Phi_{\Omega,E}}{du} = S_0\left(\vec{r}' + u\hat{\Omega},E\right). \tag{4.46}$$

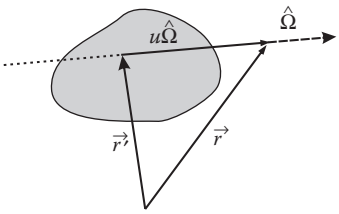



FIGURE 4.10 A source region (shaded) emits radiation. The $\Phi_{\Omega,E}$ is observed at location $\vec{r}$.

If there are no sources and no scattering or absorption, the derivative on the left-hand side of Equation 4.46 is zero and $\Phi_{\Omega,E}$ is a constant along the direction of $u$, as we have seen previously. In Equation 4.46, the primed coordinate is a source coordinate and the unprimed coordinate is the observation coordinate. The formal solution of this equation is

$$\Phi_{\Omega,E} = \int_{-\infty}^{u} S_0\left(\vec{r}' + w\hat{\Omega}, E\right) dw, \tag{4.47}$$

where $w$ is a dummy integration variable. The physical interpretation of Equation 4.47 is that radiation emitted from a point in the source region traveling along a line in the direction of $\hat{\Omega}$ continues to travel in that direction in the absence of scattering. If the line intersects the point of observation at $\vec{r}$, then it contributes to the differential fluence there. The differential fluence is calculated by "looking back" from the point of observation along the line in the opposite direction of $\hat{\Omega}$ and summing up all the contributions along that line.

As an example of an application of Equation 4.47, let us consider a source of radiation in the form of a sphere. The geometry is shown in Figure 4.11. The origin has been placed at the center of the sphere for convenience. Let us assume that our point of observation is on the $z$ axis (which can be in any direction because of symmetry). Assume that $S_0$ is spatially constant throughout the spherical volume and that particles are emitted isotropically. If $N(E)\,dE$ = the number of particles emitted by the sphere with energy between $E$ and $E + dE$, then $S_0(\vec{r}', E)$ is

$$S_0\left(\vec{r}', E\right) = \frac{3N(E)}{16\pi^2 R^3} \quad \left(r' \le R\right) \text{ and } S_0\left(\vec{r}', E\right) = 0 \quad \left(r' > R\right). \tag{4.48}$$

Referring to Equation 4.47, we have $\Phi_{\Omega,E} = S_0 u$, where $u$ is the length of the cord in Figure 4.11. The cord length is

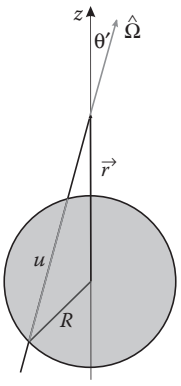



FIGURE 4.11 Isotropic radiation is uniformly emitted from a spherical source region of radius $R$. $\Phi_{\Omega,E}$ is observed at $\vec{r}$. The quantity $u$ represents the length of a chord.

$$2R\sqrt{1-\left(\frac{r}{R}\right)^2\sin^2\theta'}$$

and therefore

$$\Phi_{\Omega,E}=\frac{3N(E)}{8\pi^2R^2}\sqrt{1-\left(\frac{r}{R}\right)^2\sin^2\theta'}. \tag{4.49}$$

The value of $\Phi_{\Omega,E}$ is a maximum in the direction $\theta'=0$, as expected, and decreases to zero as $\theta'\rightarrow\sin^{-1}(R/r)$. The expression in Equation 4.49 can be integrated over all solid angle to give the energy fluence $\Phi_E$:

$$\Phi_E=\frac{3N(E)}{8\pi R^2}\int_0^{\sin^{-1}(R/r)}\int_0^{2\pi}\sin\theta'\sqrt{1-\left(\frac{r}{R}\right)^2\sin^2\theta'}\,d\varphi d\theta'=\frac{3N(E)}{8\pi R^2}\left\{\frac{x^2-1}{2x}\ln\left(\frac{x-1}{x-1}\right)+1\right\}, \tag{4.50}$$

where $x=(r/R)$. Far away from the sphere, where $x$ is large,

$$\Phi_E=\frac{N(E)}{4\pi r^2}, \tag{4.51}$$

and thus the energy fluence is simply the number of particles per unit energy emitted divided by the surface area of the sphere, which of course is an inverse square law dependence (see problem 2).

## 4.6 PRIMARY RADIATION CONSISTING OF CHARGED PARTICLES

Let us look at the form of the transfer equation in the case in which the primary radiation consists of charged particles.

First, we define

$$\mu_s\left(\vec{r},E\rightarrow E'\right)=\int_{4\pi}d\Omega'\mu_s\left(\vec{r},E\rightarrow E',\hat{\Omega}\rightarrow\hat{\Omega}'\right). \tag{4.52}$$

This quantity is the probability per unit distance traveled and per unit energy interval that a particle of energy $E$ scatters and ends up with an energy $E'$. Note that this is not the same as $\mu_s(\vec{r},E)$. Let us look at the case where there is only one

kind of charged particle (e.g., electrons or protons). This implies that there is no secondary charged particle transport (no delta rays), and therefore we set $S_0 = 0$ in Equation 4.34. Substitute the transport equation (4.34) into Equation 4.28 for the dose and set $Q = 0$:

$$D(\vec{r}) = \frac{1}{\rho}\int E\,dE\left[\Phi_E \int \mu_s\left(\vec{r}, E \to E'\right)dE' - \int dE' \int_{4\pi} d\Omega \int_{4\pi} d\Omega' \mu_s\left(\vec{r}, E' \to E, \hat{\Omega}' \to \hat{\Omega}\right)\Phi_{\Omega', E'}\right]. \tag{4.53}$$

Consider the second term in Equation 4.53. The integration variables $E$, $E'$, $\Omega$, and $\Omega'$ in this term are dummy variables; we may therefore switch primed and unprimed variables:

$$\begin{aligned} &-\frac{1}{\rho}\int dE \int E\,dE' \int_{4\pi} d\Omega \int_{4\pi} d\Omega' \mu_s\left(\vec{r}, E' \to E, \hat{\Omega}' \to \hat{\Omega}\right)\Phi_{\Omega', E'} \\ &= -\frac{1}{\rho}\int \Phi_E dE \int E' dE' \mu_s\left(\vec{r}, E \to E'\right). \end{aligned} \tag{4.54}$$

Substituting this back into Equation 4.53 and adding limits of integration gives

$$D(\vec{r}) = \frac{1}{\rho}\int_0^\infty \Phi_E dE\left[\int_0^\infty \left(E - E'\right)\mu_s\left(\vec{r}, E \to E'\right)dE'\right]. \tag{4.55}$$

The integrand of the term in square brackets in Equation 4.55 is the expected energy loss if the particle is scattered with final energy $E'$. The integral itself is the charged particle energy loss per unit length or the collisional stopping power of the medium.[5] The stopping power divided by the density is the mass stopping power. Therefore,

$$D(\vec{r}) = \int_0^\infty \left(S_\rho\right)_{\text{col}} \Phi_E\left(\vec{r}, E\right)dE, \tag{4.56}$$

where $\Phi_E$ does not include the fluence of secondary charged particles (i.e., delta rays).

## 4.7 CSDA APPROXIMATION

One of the features of the transport equation that makes it so difficult to solve is the in-scattering integral expression on the right-hand side of Equation 4.34. The form of

$\mu_s$ makes evaluation of this term troublesome. It is sometimes possible to write $\mu_s$ in a form that allows simplification of the problem. It is often the case that we deal with problems in which collisions are very frequent and that result in small energy transfers but no appreciable angular deflection. Another possibility is infrequent collisions that result in angular scatter but negligible energy transfer. This is the situation for high-energy electrons traversing matter. We can consider both of these possibilities by writing $\mu_s$ as

$$\mu_S\left(E'\to E,\hat{\Omega}'\to\hat{\Omega}\right)=\mu_S\left(E'\to E\right)\delta\left(\hat{\Omega}'-\hat{\Omega}\right)+\mu_S\left(\hat{\Omega}'\to\hat{\Omega}\right)\delta\left(E'-E\right). \tag{4.57}$$

We have suppressed the $r$ dependence of $\mu_s$ for brevity. In the case of frequent collisions with small energy transfer, we can express $\mu_s$ in terms of a continuous slowing down approximation (CSDA):

$$\mu_S\left(E'\to E\right)\approx\frac{1}{\Delta x}\delta\left(E'-E-\frac{dE}{dx}\Delta x\right), \tag{4.58}$$

where $dE/dx < 0$. The $1/\Delta x$ term in Equation 4.58 gives us units of probability per unit length. We now substitute Equations 4.57 and 4.58 into the integral on the right-hand side of the BTE (Equation 4.34). After a little manipulation (see problem 5), the result is

$$\int dE'\int_{4\pi}d\Omega'\mu_s\left(\vec{r},E'\to E,\hat{\Omega}'\to\hat{\Omega}\right)\Phi_{\Omega',E'}=\frac{1}{\Delta x}\Phi_{\Omega,E}+\frac{dE}{dx}\frac{\partial\Phi_{\Omega,E}}{\partial E}+\int d\Omega'\mu_s\left(\hat{\Omega}'\to\hat{\Omega}\right)\Phi_{\Omega',E}. \tag{4.59}$$

The second term on the left-hand side of Equation 4.34) becomes

$$\mu\left(\vec{r},E\right)\Phi_{\Omega,E}=\left[\frac{1}{\Delta x}+\int d\Omega'\mu_S\left(\hat{\Omega}'\to\hat{\Omega}\right)\right]\Phi_{\Omega,E}. \tag{4.60}$$

We have assumed that $\mu_s\left(\hat{\Omega}'\to\hat{\Omega}\right)=\mu_s\left(\hat{\Omega}'\to\hat{\Omega}\right)$ because $\mu_s$ depends only on the scattering angle $\hat{\Omega}'\cdot\hat{\Omega}$. Combining all the terms results in the BTE in the CSDA approximation:

$$\hat{\Omega}\cdot\vec{\nabla}\Phi_{\Omega,E}=\frac{dE}{dx}\frac{\partial\Phi_{\Omega,E}}{\partial E}+\int d\Omega'\mu_s\left(\hat{\Omega}'\to\hat{\Omega}\right)\left[\Phi_{\Omega',E}-\Phi_{\Omega,E}\right]+S_0\left(r,E,\hat{\Omega}\right). \tag{4.61}$$

This eliminates the integral over energy on the right-hand side of the BTE (Equation 4.34).

## 4.8 INDIRECTLY IONIZING RADIATION

Now let us examine a more complex case in which we have an indirectly ionizing radiation field consisting of neutral particles (e.g., photons or neutrons) and charged particles set in motion by the neutral particles (electrons, protons, etc.). Pair production is excluded.[6] We will write two coupled equations, one for the neutral particles and the other for the charged particles. Quantities associated with uncharged particles (photons) will have a subscript $u$, and quantities associated with the charged particles (electrons or positrons) will have subscript $c$. *In this section, we will drop the subscript* $\Omega,E$ in $\Phi_{\Omega,E}$ *for ease of writing.* The transport equations are

$$\hat{\Omega}\cdot\nabla\Phi_u+\mu_u\Phi_u=q_{u'u}+q_{cu}+S_{ou}, \tag{4.62a}$$

$$\hat{\Omega}\cdot\nabla\Phi_c+\mu_c\Phi_c=q_{c'c}+q_{uc}+S_{0c}, \tag{4.62b}$$

It must be remembered that $\Phi$ is *not* the total fluence, but rather the differential fluence $\Phi_{\Omega,E}$. We now define each one of the terms in Equations 4.62. First, we give the out-scattering terms for uncharged and charged particles, respectively:

$$\begin{aligned}\mu_u&=\int dE'_u\int d\Omega'\mu_s\left(E_u\to E'_u,\hat{\Omega}\to\hat{\Omega}'\right)\\ \mu_c&=\int dE'_c\int d\Omega'\mu_s\left(E_c\to E'_c,\hat{\Omega}\to\hat{\Omega}'\right).\end{aligned} \tag{4.63}$$

Next, we give the in-scattering terms:

$$\begin{aligned}q_{u'u}\left(E_u,\hat{\Omega}\right)&=\int dE'_u\int d\Omega'\mu_s\left(E'_u\to E_u,\hat{\Omega}'\to\hat{\Omega}\right)\Phi_{u'}\\ q_{c'c}\left(E_c,\hat{\Omega}\right)&=\int dE'_c\int d\Omega'\mu_s\left(E'_c\to E_c,\hat{\Omega}'\to\hat{\Omega}\right)\Phi_{c'}\end{aligned} \tag{4.64}$$

Finally, the terms that couple charged and uncharged particles are

$$\begin{aligned}q_{cu}\left(E_u,\Omega_u\right)&=\int dE_c\int d\Omega_c\mu_s\left(E_c\to E_u,\hat{\Omega}_c\to\hat{\Omega}_u\right)\Phi_c\\ q_{uc}\left(E_c,\Omega_c\right)&=\int dE_u\int d\Omega_u\mu_s\left(E_u\to E_c,\hat{\Omega}_u\to\hat{\Omega}_c\right)\Phi_u.\end{aligned} \tag{4.65}$$

The first of the two equations above (4.65) describes the production of uncharged particles from charged particles (i.e., bremsstrahlung radiation). The second

equation describes production of energetic charged particles from photons, namely, recoil electrons resulting from Compton scattering. In Equations 4.63 through 4.65, $\mu_s$ may depend on $\vec{r}$.

We now wish to compute the dose. We assume that there are no sources within the medium (that $S_{oc}$ and $S_{ou}$ are zero; there are no radionuclides in the medium undergoing radioactive decay) and that the source of radiation is external to the medium. We will also ignore the $Q_u$ and $Q_c$ terms in Equation 4.28 for the dose. Under these circumstances, the dose is given by

$$D(\vec{r}) = \frac{1}{\rho}\int E_u dE_u \int d\Omega_u \left[-\hat{\Omega}_u \cdot \vec{\nabla}\Phi_u\right] + \frac{1}{\rho}\int E_c dE_c \int d\Omega_c \left[-\hat{\Omega}_c \cdot \vec{\nabla}\Phi_c\right]. \tag{4.66}$$

Substitution of Equations 4.62 into the dose equation (4.66) yields

$$D(\vec{r}) = \frac{1}{\rho}\left\{\underbrace{\int E_u dE_u \int d\Omega_u \mu_u \Phi_u}_{1} - \underbrace{\int E_u dE_u \int d\Omega_u q_{u'u}}_{2} - \underbrace{\int E_u dE_u \int d\Omega_u q_{cu}}_{3}\right\} +$$

$$\frac{1}{\rho}\left\{\underbrace{\int E_c dE_c \int d\Omega_c \mu_c \Phi_c}_{4} - \underbrace{\int E_c dE_c \int d\Omega_c q_{c'c}}_{5} - \underbrace{\int E_c dE_c \int d\Omega_c q_{uc}}_{6}\right\}. \tag{4.67}$$

We have numbered the six integral terms in Equation 4.67 so that we may refer back to them. Let us consider terms 1 and 2 to begin with. We proceed as follows: Substitute Equation 4.63a into integral 1 and integrate over Ω and Ω′. Substitute $q_{u'u}$ (Equation 4.64a) into integral 2. This results in a quadruple integral over both primed and unprimed variables. The integration variables are dummy variables, and therefore they can be switched by making unprimed variables primed and vice versa. Then integrate over Ω and Ω′. Subtraction of integral 2 from integral 1 yields

$$\int dE_u \Phi_u(E_u) \int dE'_u (E_u - E'_u)\mu_s(E_u \to E'_u), \tag{4.68}$$

where:

$\Phi_u(E_u)$ is the differential energy fluence (Equation 4.21)

$\mu_s(E_u \to E'_u)$ is the probability per unit length and per unit energy interval of a scattering event in which the energy goes from $E_u$ to $E_u'$

Following NCRP 108 (1991), we define $q_u(E_u \to E'_u)$ as the change in rest mass energy after a transition in which the energy goes from $E_u$ to $E'_u$ and in which a secondary charged particle of kinetic energy $E_u - E'_u - q_u$ is produced. The energy

transfer coefficient $\mu_{tr}(E_u)$ is defined as the fraction of $E_u$ transferred to charged particles per unit distance traveled. This can be expressed as

$$\mu_{tr}(E_u) = \frac{1}{E_u}\int dE'_u\, \mu_s(E_u \to E'_u)\left[E_u - E'_u - q_u(E_u \to E'_u)\right]. \tag{4.69}$$

We will neglect the $q_u$ term and use Equation 4.69 to replace the second integral in Equation 4.68. We also carry along the 1/ρ term from Equation 4.67. The integral in Equation 4.68, with correction for the rest mass energies $q_u$, represents the sum of the initial kinetic energies of all the charged particles set in motion by the uncharged particles per unit mass. This is the kerma, which can be written as

$$K(\vec{r}) = \underbrace{\int E_u \Phi_u(\vec{r}, E_u)\frac{\mu_{tr}}{\rho}\, dE_u}_{1\text{–}2}. \tag{4.70}$$

We now turn our attention to integral 3 in Equation 4.67. If we substitute the expression for $q_{cu}$ (Equation 4.65) into this integral, integrate over $d\Omega_u$ and $d\Omega_c$, and multiply by 1/ρ, the result is

$$\frac{1}{\rho}\int \Phi_c(E_c)\int E_u \mu_s(E_c \to E_u)\, dE_u\, dE_c = K_r, \tag{4.71}$$

where $\mu_s(E_c \to E_u)$ is the probability per unit length per unit energy that a charged particle will produce a photon (i.e., bremsstrahlung production) of energy $E_u$ and $\Phi_c(E_c)$ is the differential energy fluence of charged particles. This integral represents the energy per unit mass radiated by charged particles. If we ignore secondary charged particles (i.e., delta rays), this is simply the radiative portion of the kerma.

Let us pause for a moment to consider the expression for the dose in the case in which CPE prevails. In CPE, the gradient of the differential fluence for the charged particles must be zero, that is $\vec{\nabla}\Phi_c = 0$ and therefore integrals 4–6 must algebraically sum to zero as they originate from the second term in Equation 4.66. We are then left with two contributions to the dose: the kerma (Equation 4.70) minus the radiative kerma (Equation 4.71), which is simply equal to the collision kerma:

$$D \overset{CPE}{\approx} K_c, \tag{4.72}$$

where $K_c$ is the collision kerma. Equation 4.72 is a result that the reader is likely to find familiar.

CPE is not strictly obeyed, particularly for high-energy photon beams. Therefore, let us continue our evaluation of the expression for the dose without invoking this assumption.

Let us consider integrals 4 and 5 in Equation 4.67. We proceed as we did with integrals 1 and 2. Substitute Equation 4.63 into integral 4 and integrate over Ω and Ω′. Substitute $q_{c'c}$ into integral 5. Switch unprimed and primed variables, and then integrate over Ω and Ω′. Subtraction of integral 5 from integral 4 yields

$$\frac{1}{\rho}\int \Phi_c(E_c)\int (E_c - E_c')\mu_s(E_c \rightarrow E_c')\,dE_c'\,dE_c, \tag{4.73}$$

where $\mu_c(E_c \rightarrow E_c')$ is the probability per unit length per unit energy of a charged particle with energy $E_c$ undergoing a transition to a charged particle of energy $E_c'$. The second integral in Equation 4.73 is the total energy loss per unit length for charged particles of energy $E_c$—in other words, it is the stopping power (as in Equation 4.55). Equation 4.73 can therefore be written

$$\int \Phi_c(E_c)\left(S_\rho\right)_{tot} dE_c. \tag{4.74}$$

This term represents the total energy per unit mass lost by charged particles.

The last term that we need to evaluate in Equation 4.67 is integral 6. Proceed as before by inserting the expression for $q_{uc}$ (Equation 4.65) into integral 6 and then integrating over $d\Omega_c$ and $d\Omega_u$, which results in

$$\frac{1}{\rho}\int \Phi_u(E_u)\int E_c\,\mu_s(E_u \rightarrow E_c)\,dE_c\,dE_u. \tag{4.75}$$

This is the total initial energy per unit mass of charged particles set in motion by uncharged particles (e.g., by Compton scattering) or the total kerma again. Adding all six integrals (algebraically) in Equation 4.67 for the dose leaves a net contribution of

$$D(\vec{r}) = \int \left(S_\rho\right)_{col} \Phi_c(E_c)\,dE_c, \tag{4.76}$$

which is the same as Equation 4.56. This shows that the charged particles deposit the dose. Uncharged particles set charged particles in motion, and it is ultimately these charged particles that deposit energy.

## 4.9 EFFICACY OF BTE-BASED DOSE CALCULATIONS

In this section, we will describe ACUROS®, a commercial treatment planning system dose computation algorithm based on the BTE. There are two versions of the ACUROS algorithm: ACUROS XB, for external beam, and ACUROS BV, for brachytherapy dose calculations. These BTE algorithms are offered as an option for the Varian Eclipse treatment planning system. The ACUROS XB algorithm received Food and Drug Administration (FDA) 510(k) clearance in late 2010. Commercial dose calculation algorithms have achieved a high level of accuracy over the last 10 years. In order to succeed, new algorithms need to be (1) more accurate, (2) faster, and/or (3) cheaper.

For Monte Carlo calculations, each individual particle is followed as it traverses the material, until the energy of the particle falls below some negligible value. To obtain reasonable average values, it may be necessary to follow hundreds of millions of particle histories. The BTE approach is free from statistical noise, and "small perturbations can be rapidly assessed." In addition, it requires much less CPU time. Even under ideal circumstances, Monte Carlo stochastic errors result from the simulation of a finite number of particles. In grid-based BTE algorithms, errors are due to (1) approximations made to the BTE and (2) errors introduced by the numerical solution.

Calculation times for BTE algorithms are only weakly dependent on the number of beams, whereas convolution/superposition algorithm computation times scale linearly with the number of beams. In BTE algorithms, the time for building the source model for all beams is proportional to the number of fields, but the rest of the calculation is not. BTE-based algorithms may therefore hold a significant speed advantage for arc therapy (e.g., volumetric modulated arc therapy [VMAT], see below) (Bush et al., 2011).

This description of the ACUROS (hereafter AXB) algorithm is from the paper by Vassiliev et al. (2010). We have converted their notation so that it is consistent with the notation used here. We have again dropped the $\Omega,E$ subscripts for economy. *Note:* $\Phi$ *is not the total fluence here, but* $\Phi_{\Omega,E}$.

We start with the transport equations in the form

$$\hat{\Omega}\cdot\nabla\Phi_u+\mu_u\Phi_u=q_{c'c}+S_{0u} \tag{4.77}$$

$$\hat{\Omega}\cdot\nabla\Phi_c+\mu_c\Phi_c-\frac{\partial}{\partial E}\left(S_r\Phi_c\right)=q_{c'c}+q_{uc}+S_{0c}, \tag{4.78}$$

where $S_r$ is the restricted collisional plus radiative stopping power. Equation 4.78 is the CSDA version of the charged particle Equation 4.62b.

Some of the assumptions built into these equations are

1. For pair production, the secondary particles are both assumed to be electrons instead of an electron and a positron. This avoids the need for a separate transport equation for the positrons.

2. Partial coupling is assumed between photons and electrons. Photons produce electrons, but electrons do not produce photons ($q_{cu}$ in Equation 4.62 is neglected). That is, there is no bremsstrahlung. The energy from photons produced by electrons is assumed to be deposited locally (not carried away by a bremsstrahlung photon).

The dose is given by

$$D = \int_0^\infty dE \int_{4\pi} d\hat{\Omega} \frac{\sigma_{ED}^e(\vec{r}, E)}{\rho} \Phi_c\left(\vec{r}, E, \hat{\Omega}\right), \tag{4.79}$$

where $\sigma_{ED}^e$ is the macroscopic energy deposition cross section in units of MeV/cm. This would appear to be related to the collisional stopping power (see Equation 4.55).

The photon sources incident on the patient are represented as point sources located at the treatment head target and possible other places to represent photon scatter, such as the flattening filter. In this case, the source term in Equation 4.77 can be written as $S_{ou} = q_u\left(E, \hat{\Omega}\right)\delta\left(\vec{r} - \vec{r}_p\right)$, where $\vec{r}_p$ is the location of the source. Further details of the implementation of Equations 4.77 and 4.78 can be found in Vassiliev et al. (2010).

A comprehensive review of the published assessments of the accuracy of BTE dose calculations is provided by Kan et al. (2013). We will describe three of these: Vassiliev et al. (2010), Bush et al. (2011), and Han et al. (2013).

Vassiliev et al. have tested the AXB algorithm against a Monte Carlo algorithm (based on EGSnrc) for both 6 and 18 MV beams. Two phantoms were used for this purpose: a slab phantom containing layers of "tissue," bone, and lung material and an anthropomorphic breast phantom. Details of the Monte Carlo statistical error are given in the paper. For the slab phantom, in locations where the dose is >10% of the maximum dose, the agreement between AXB and Monte Carlo (hereafter MC) is within 2% or 1 mm distance to agreement. For the breast phantom, the AXB dose agreed with MC to within 2% or 2 mm distance to agreement for 99.9% of voxels with a dose > 10% of the prescribed dose. All AXB beam calculations took less than 5 min on a workstation with two dual-core AMD Opteron processors.

Figure 4.12 shows the dose along the central axis from Vassiliev et al. (2010) for an 18 MV beam in the slab phantom. AXB appears to accurately handle interfaces between media.

Figure 4.13 shows a dose profile comparison for the breast phantom.

Bush et al. (2011) have studied the accuracy of AXB for single 6 and 18 MV beams in homogeneous and heterogeneous media. Calculated doses have been compared to measurements, the analytic anisotropic algorithm (AAA) (a pencil beam algorithm; see Chapter 3), and MC calculations using BEAMnrc/DOSXYZnrc. Open field measurements in a homogeneous phantom show agreement with AXB calculations "to within ±1.9% in the inner field region for all field sizes and energies" (Bush et al., 2011, p. 2208). AXB dose calculations in a heterogeneous interface phantom agree with MC within ±2% in lung ($\rho$ = 0.24 g/cm$^3$) and within 3% in low-density lung

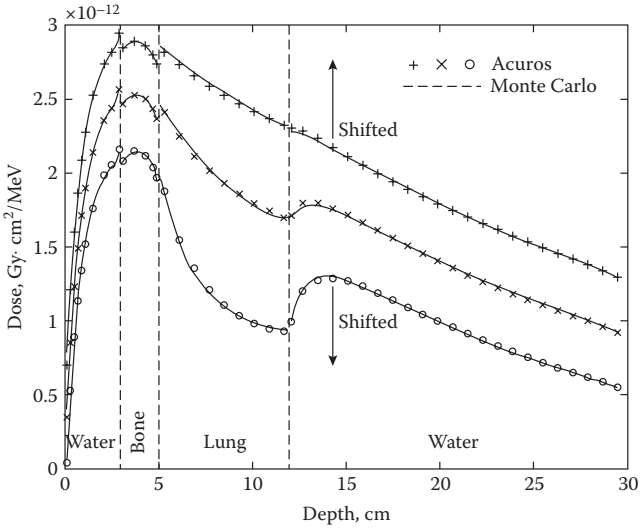



FIGURE 4.12 Validation of the ACUROS BTE algorithm against a Monte Carlo calculation as reported by Vassiliev et al. (2010). This is the dose per unit energy fluence along the central axis for an 18 MV beam incident on a slab phantom with inhomogeneous slabs as labeled. The phantom is 30 cm × 30 cm × 30 cm, and the SSD to the top of the phantom is 100 cm. The top curve is for a 10 cm × 10 cm beam, the middle curve is for a 5 cm × 5 cm beam, and the lower curve is for a 2.5 cm × 2.5 cm beam. The top and bottom curves have been shifted for ease of visualization. (From Vassiliev, O.N., et al., *Phys. Med. Biol.* 55, 581–98, 2010, figure 2.)

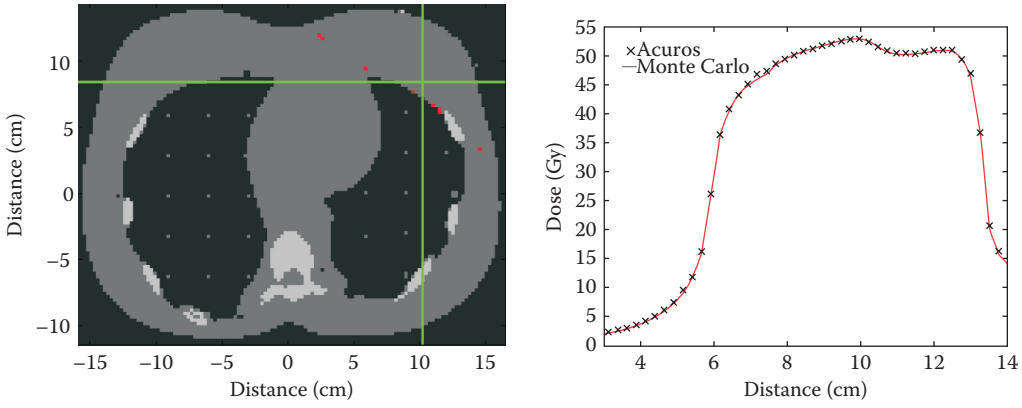



FIGURE 4.13 Validation of the ACUROS BTE algorithm for a breast phantom. The isocenter is at the intersection of the two green lines. The dose profile on the right shows the agreement between AXB and MC along the green line in the anterior and posterior direction. We presume this is for a 6 MV beam arrangement. (From Vassiliev, O.N., et al., *Phys. Med. Biol.* 55, 581–98, 2010, figures 4.8 and 4.10.)

($\rho = 0.10$ g/cm$^3$). In contrast to this, the AAA algorithm exhibited differences of up to 10% and 18%, respectively, in lung and low-density lung.

Figure 4.14 shows the Bush et al. (2011) comparison between an 18 MV (4 cm × 4 cm) beam dose calculation using MC, AXB, and AAA. The phantom used for these calculations is illustrated in Figure 4.15.

It is reported that AAA is a fast pencil beam algorithm. Hasenbalg et al. (2007) report AAA to be 7–11 times faster than collapsed cone convolution. The time

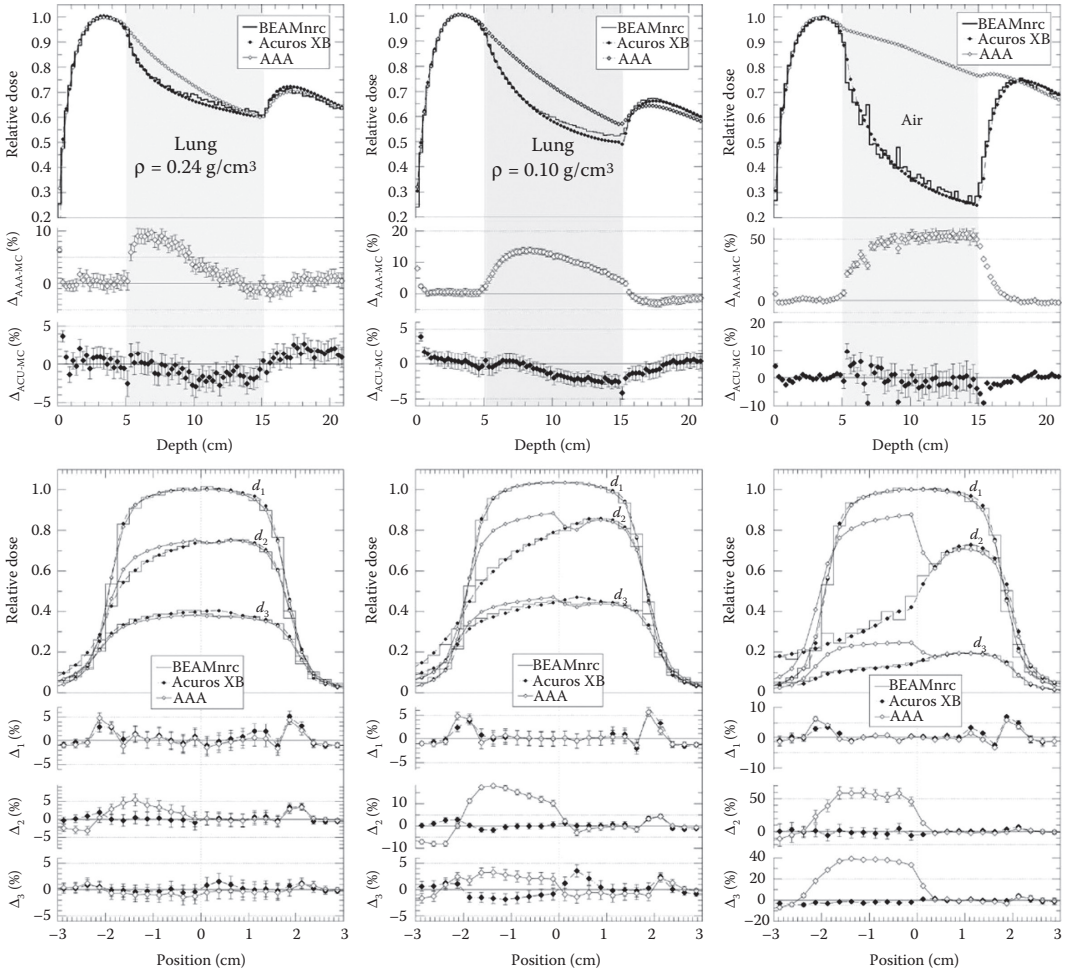



FIGURE 4.14 A 4 cm × 4 cm, 18 MV beam dose comparison between ACUROS XB, Monte Carlo (BEAMnrc), and a superposition/convolution algorithm (AAA) in a heterogeneous phantom (see Figure 4.15). The three columns show comparisons for three different heterogeneities: lung ($\rho = 0.24$ g/cm$^3$), low-density lung ($\rho = 0.10$ g/cm$^3$), and air. The top row shows depth–dose curves along with curves showing differences between the algorithms. The bottom row shows profiles above, in, and below the heterogeneous layer. Although perhaps not clinically relevant, the air cavity AAA dose shows deviations of up to 50% with respect to the MC calculations. The accuracy of ACUROS XB is much better. (From Bush, K., et al., *Med. Phys.* 38(4), 2208, 2011, figure 6.)

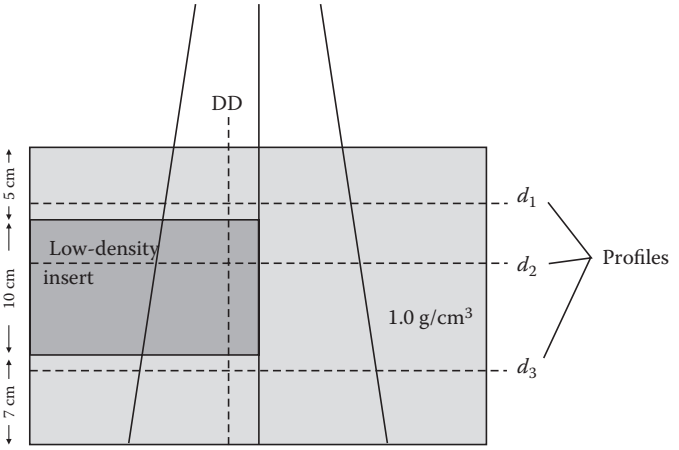



FIGURE 4.15 Phantom in which depth–dose curves and profiles have been computed for the purpose of comparing the ACUROS XB, AAA, and Monte Carlo dose calculation algorithms. The depth–dose curves and profiles are shown in Figure 4.14. Calculations were performed for low-density inserts of lung ($\rho = 0.24$ and 0.10 g/cm$^3$) and air. (Based on Bush, K., et al., *Med. Phys.* 38(4), 2208, 2011, figure 2.)

required for the AXB dose calculation of a static 10 cm × 10 cm 6 MV field with a 2 mm × 2 mm × 2 mm dose grid was 110 s on a Dell T5500 with dual Xeon quad-core CPUs. The equivalent calculation for the AAA algorithm took 8 s (Bush et al., 2011).

Convolution superposition algorithms are least accurate for high-energy narrow beams and for regions near water density and low-density interfaces. Bush et al. (2011, p. 2208) conclude that "the Acuros® XB algorithm is capable of modeling radiotherapy dose deposition with accuracy only previously achievable with Monte Carlo techniques."

Han et al. (2013) have compared IMRT and VMAT dose measurement in the Radiological Physics Center (RPC, known as IROC) thorax phantom with calculations using the AXB and AAA algorithms. Dose was measured using thermoluminescent dosimeters (TLDs) and Gafchromic film. The results are summarized in Table 4.1. Differences of up to 8% were found at lung–soft tissue interfaces.

Figures 4.16 and 4.17 are from Han et al. (2013) and show a gamma analysis comparison between AXB and AAA for 6 MV, RPC lung phantom plans for IMRT and VMAT, respectively.

**TABLE 4.1 Calculated (AXB and AAA) vs. Measured Doses**

| Comparison Method | AXB | AAA |
|---|---|---|
| TLD | 0.4%–4.4% | 2.5%–6.4% |
| γ index (±3%/3 mm) | 97%–98% | 94% |

*Source:* Han, T., et al., *Med. Phys.* 40(5), 2013, 051710-1–051710-11.

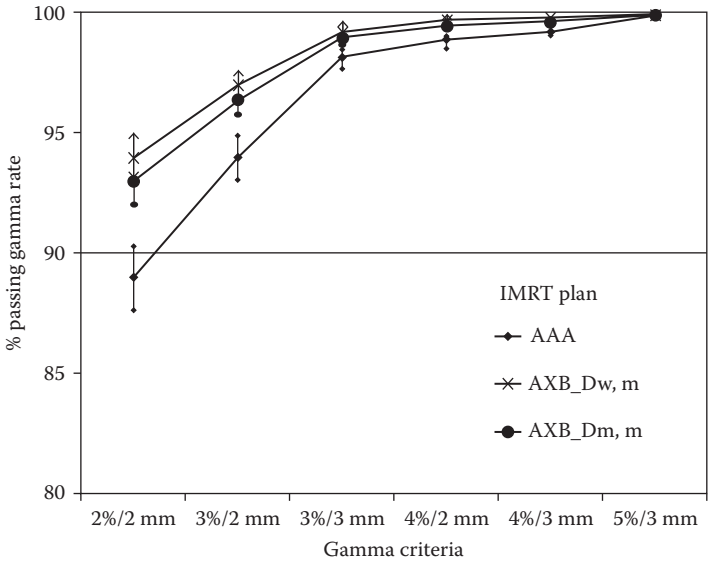



FIGURE 4.16 Gamma analysis comparison between dose measurements and ACUROS XB (AXB) and the AAA algorithm for a 6 MV IMRT plan in the RPC lung phantom. Dw,m and Dm,m represent dose to water in the medium and dose to the medium in the medium. (From Han, T., et al., *Med. Phys.* 40(5), 2013, fig. 6.)

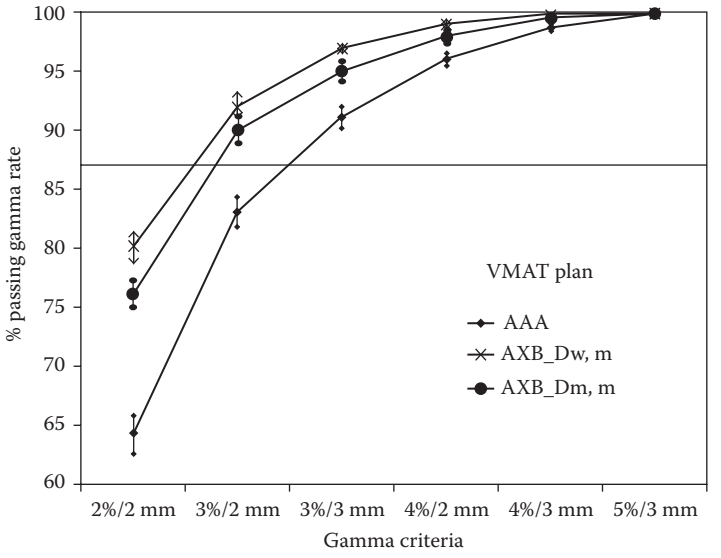



FIGURE 4.17 Gamma analysis comparison between dose measurements and ACUROS XB and the AAA algorithm for a 6 MV VMAT plan in the RPC lung phantom. The Dw,m and Dm,m represent dose to water in the medium and dose to the medium in the medium. (From Han, T., et al., *Med. Phys.* 40(5), 2013, fig. 6.)

A comparison was made between the calculation times for AAA and AXB running on a Dell T5500 with dual 2.27 GHz quad-core Intel processors (64 bit, Windows 7). Table IV of Han et al. (2013) shows that the AAA algorithm required about 0.8 min for an IMRT plan. This is to be compared to 2.5 min for the AXB calculation. This is dose calculation time and does not include optimization. For VMAT plans, the time ratio is more than reversed, with AAA requiring 16 min and AXB taking about 4 min.

In summary, AXB rivals the accuracy of MC calculations and may be significantly more accurate than pencil beam algorithms near interfaces between unit density and low-density tissue, particularly for high energies and small field sizes. AXB calculations are significantly faster than pencil beam computations for VMAT and may therefore hold a decided advantage in both speed and accuracy for VMAT lung treatment planning.

## 4.10 FERMI–EYGES THEORY AND ELECTRON PENCIL BEAM DOSE CALCULATIONS

In this section, we derive the Fermi–Eyges equation that is the basis for the electron dose calculations used by some commercial treatment planning systems.[7] We assume an electron pencil beam traveling in the +$z$ direction and that scattering is through small angles. This is reasonable, as the vast majority of interactions are soft collisions. The derivation here relies heavily on "Notes 8" found on the web (any errors are my responsibility).

Fermi lectured on this topic in 1940 in the context of the investigation of cosmic rays. His intent was to study charged particle transport in the atmosphere. He assumed that the particles did not undergo any energy loss. Apparently, Fermi did not feel that this research was significant enough to even bother publishing. Perhaps what others would have considered a major accomplishment, Fermi felt was obvious and trivial. The theory was extended by Eyges in 1948 to include energy losses.

We begin with the BTE in the form (Equation 4.34) with no absorption and no sources. We also neglect energy losses, in which case, $\mu_s(\vec{r},E' \to E,\hat{\Omega} \to \hat{\Omega}') = \mu_s(\vec{r},\hat{\Omega} \to \hat{\Omega}')\delta(E' - E)$ and so on:[8]

$$\hat{\Omega} \cdot \vec{\nabla}\Phi_\Omega = \int_{4\pi} d\Omega' \left[ \mu_s\left(\vec{r},\hat{\Omega}' \to \hat{\Omega}\right)\Phi_{\Omega'} - \mu_s\left(\vec{r},\hat{\Omega} \to \hat{\Omega}'\right)\Phi_\Omega \right] \tag{4.80}$$

*In this section, we will hereafter drop the* $\Omega$ *subscript on* $\Phi$.

We define projection angles $\theta_x$ and $\theta_y$ as the scattering angles projected onto the $x$-$z$ plane and the $y$-$z$ plane, respectively (Figure 4.18):

$$\tan\theta_x = \frac{\Omega_x'}{\Omega_z'} = \tan\theta' \cos\varphi'$$

$$\tan\theta_y = \frac{\Omega_y'}{\Omega_z'} = \tan\theta' \sin\varphi'. \tag{4.81}$$

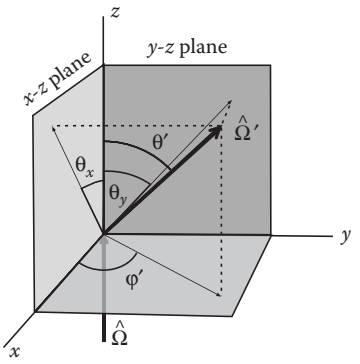



FIGURE 4.18 The geometry for the derivation of the Fermi–Eyges equation from the BTE. A pencil beam of electrons travels upward along the –$z$ axis and scatter through an angle $\theta'$ and $\varphi'$. The scattering angle $\theta'$ is small despite the fact that it is not shown that way in the diagram. The projected angles $\theta_x$ and $\theta_y$ are defined in Equation 4.81.

We assume that the scattering angle $\theta'$ is small, and therefore the streaming term becomes

$$\hat{\Omega}\cdot\vec{\nabla}\Phi \approx \Omega_z'\left[\theta_x\frac{\partial\Phi}{\partial x}+\theta_y\frac{\partial\Phi}{\partial y}+\frac{\partial\Phi}{\partial z}\right]. \tag{4.82}$$

Now let us look at the right-hand side of Equation 4.80. The quantity:
$d\Omega' = \sin\theta'\, d\theta'\, d\varphi' \approx \theta'\, d\theta'\, d\varphi' \approx d\theta_x\, d\theta_y - \varphi'\, d\theta_x\, d\theta_x + \cdots \approx d\theta_x\, d\theta_y$ ($\varphi' \ll 1$).

Let us write $\hat{\Omega}' = \hat{\Omega} + \vec{\xi}$, and it can be shown that $\theta_x \approx \xi_x$ and $\theta_y \approx \xi_y$, and therefore $d\theta_x d\theta_y \approx d\xi_x\, d\xi_y$. The quantity $\mu_s$ depends only on the angle of scattering and consequently $\mu_s(\vec{r},\hat{\Omega}\to\hat{\Omega}') = \mu_s(\vec{r},\hat{\Omega}'\to\hat{\Omega}) = \mu_s(\vec{r},\vec{\xi})$. We can now write the transport equation as

$$\Omega_z'\left[\theta_x\frac{\partial\Phi}{\partial x}+\theta_y\frac{\partial\Phi}{\partial y}+\frac{\partial\Phi}{\partial z}\right] \approx \int \mu_s\left(\vec{r},\vec{\xi}\right)\left[\Phi\left(\vec{r},\hat{\Omega}+\vec{\xi}\right)-\Phi\left(\vec{r},\hat{\Omega}\right)\right]d\xi_x d\xi_y. \tag{4.83}$$

We can expand $\Phi\left(\vec{r},\hat{\Omega}+\vec{\xi}\right)$ to second order in powers of the small quantities $\xi_x$ and $\xi_y$ as follows:

$$\Phi\left(\vec{r},\hat{\Omega}+\vec{\xi}\right) \approx \Phi\left(\vec{r},\hat{\Omega}\right)+\frac{\partial\Phi}{\partial\Omega_x}\xi_x+\frac{\partial\Phi}{\partial\Omega_y}\xi_y+\frac{1}{2}\frac{\partial^2\Phi}{\partial\Omega_x^2}\xi_x^2+\frac{1}{2}\frac{\partial^2\Phi}{\partial\Omega_y^2}\xi_y^2+\frac{\partial^2\Phi}{\partial\Omega_x\partial\Omega_y}\xi_x\xi_y+\cdots. \tag{4.84}$$

We note that $\partial\Phi/\partial\Omega_x \approx \partial\Phi/\partial\theta_x$, $\partial\Phi/\partial\Omega_y \approx \partial\Phi/\partial\theta_y$ and so forth. We now substitute Equation 4.84 into Equation 4.83. The terms in the integral in Equation 4.83 that are linear in $\xi_x$ or $\xi_y$ are odd functions (this also includes the $\xi_x\xi_y$ term), and therefore

the integrals of these terms are zero. The integrals involving $\xi_x^2$ and $\xi_y^2$ have the same value because of symmetry. Equation 4.83 can now be written as

$$\theta_x \frac{\partial \Phi}{\partial x} + \theta_y \frac{\partial \Phi}{\partial y} + \frac{\partial \Phi}{\partial z} = \frac{1}{2}\left[\frac{\partial^2 \Phi}{\partial \theta_x^2} + \frac{\partial^2 \Phi}{\partial \theta_y^2}\right] \int \mu_s\left(\vec{\xi}\right) \xi_{x \text{ or } y}^2 d\xi_x d\xi_y \tag{4.85}$$

$\left(\Omega_z' \approx 1\right)$. We now define the scattering power:

$$T(z) = \int \theta'^2 \mu_s\left(\theta'\right) d\Omega'; \tag{4.86}$$

it is related to the mean of the square of the scattering angle per unit length, and it is a measure of the ability of the medium to scatter electrons. The units of $T$ are radian$^2$/cm. From Equations 4.81, we can see that $\theta_x^2 + \theta_y^2 \approx \theta'^2$. The mean values of $\theta_x^2$ and $\theta_y^2$ are equal, and therefore Equation 4.85 becomes

$$\theta_x \frac{\partial \Phi}{\partial x} + \theta_y \frac{\partial \Phi}{\partial y} + \frac{\partial \Phi}{\partial z} = \frac{1}{4} T(z)\left[\frac{\partial^2 \Phi}{\partial \theta_x^2} + \frac{\partial^2 \Phi}{\partial \theta_y^2}\right]. \tag{4.87}$$

The scattering power $T$ depends on the electron energy and therefore depth $z$ in the medium. Equation 4.87 is the Fermi–Eyges equation.[9]

The Fermi–Eyges equation is symmetric with respect to $x$ and $y$, $\theta_x$ and $\theta_y$; therefore, let us concentrate on the solution in the $x$-$z$ plane:

$$\theta_x \frac{\partial \Phi}{\partial x} + \frac{\partial \Phi}{\partial z} = \frac{1}{4} T(z)\left[\frac{\partial^2 \Phi}{\partial \theta_x^2}\right]. \tag{4.88}$$

The differential fluence is now a function of $z$, $x$, and $\theta_x$. The quantity $\Phi(z, x, \theta_x)$ $dx\, d\theta_x$ is the number of particles at depth $z$ with lateral displacement between $x$ and $x + dx$ traveling at an angle between $\theta_x$ and $\theta_x + d\theta_x$.

Before we discuss the solution of Equation 4.88, we pause to consider the angular scattering power (Equation 4.86) in more detail. The mass angular scattering power is the scattering power divided by the density, that is, $T/\rho$. This quantity is often written as $d\theta^2/\rho ds$, and the units are radian$^2$ cm$^2$/g. The mass angular scattering power is proportional to atomic number, and it varies approximately as the inverse square of the kinetic energy for kinetic energy that is large compared to 0.5 MeV (Podgorsak, 2010). Li and Rogers (1995) have compiled values of $T/\rho$ as a function of energy for a variety of materials.

These values of $T$ are plotted in Figure 4.19 for water for energies ranging from 0.5 to 20 MeV. For 5 MeV electrons in water, $T = 0.28$ cm$^{-1}$. It is clear from the log-log

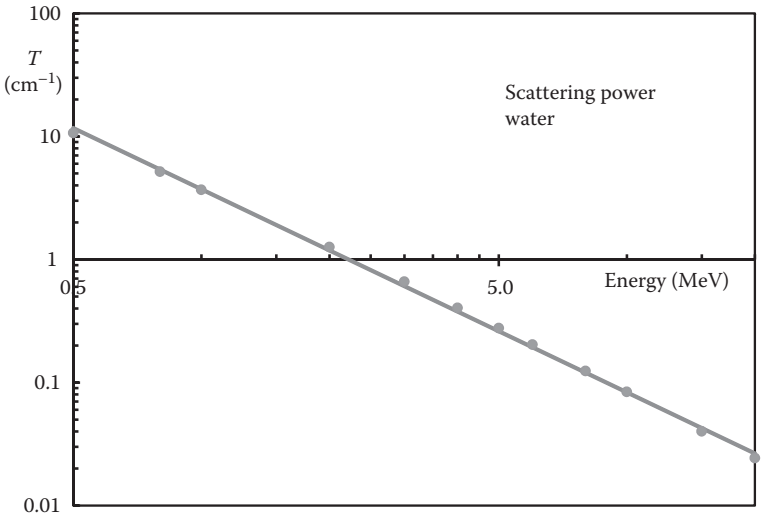



FIGURE 4.19 A plot of log $T$ vs. log $E$ for water for energies from 0.5 to 20 MeV. Values of $T$ (data points) are from Li and Rogers (1995). The line shows a power law fit to the data over this range in energy: $T = 3.72\ E^{-1.65}$. Units of $T$ are cm$^{-1}$.

plot of Figure 4.19 that $T$ follows a power law in $E$. The best fit to these data (the line in Figure 4.19) is $T = 3.72\ E^{-1.65}$, where $E$ is in units of MeV and $T$ is in units of cm$^{-1}$.

Eyges (1948) has solved Equation 4.88 using the method of Fourier transforms:

$$\Phi(z,x,\theta_x) = \frac{1}{4\pi\sqrt{B(z)}} \exp\left\{\frac{-A_2(z)\theta_x^2 + 2A_1(z)x\theta_x - A_0(z)x^2}{4B(z)}\right\}, \tag{4.89}$$

where

$$B(z) = A_0A_2 - A_1^2$$

$$A_n(z) = \frac{1}{4}\int_0^z (z-z')^n\, T(z')\,dz'. \tag{4.90}$$

This solution can be verified by substitution into Equation 4.88 (see problem 6). In the case for which $T(z) = T_0$ (a constant),

$$A_n(z) = \frac{1}{4}\frac{T_0}{n+1} z^{n+1}. \tag{4.91}$$

Equation 4.89 for $\Phi$ can be written as the product of two Gaussians:

$$\Phi(z,x,\theta_x) = \frac{1}{2\sqrt{\pi A_2}} \exp\left\{-\frac{x^2}{4A_2}\right\} \frac{1}{\sqrt{2\pi\sigma_{\theta_x}^2}} \exp\left\{-\frac{(\theta_x-\theta_p)^2}{2\sigma_{\theta_x}^2}\right\}, \tag{4.92}$$

where

$$\sigma_{\theta_x}^2 = 2\left(A_0 - \frac{A_1^2}{A_2}\right) = \frac{2B}{A_2};\ \text{and}\ \theta_p = \frac{xA_1}{A_2}. \tag{4.93}$$

This is not quite a clean separation of variables into a product of a function of $x$ and a function of $\theta_x$ because $\theta_p$ depends on $x$. The first Gaussian in the product of Gaussians in Equation 4.92 is independent of $\theta_x$, and it describes the distribution of the fluence in the $x$ direction with a width related to the standard deviation $\sqrt{2A_2(z)}$. For any particular point in the medium $(x, z)$, the angular dependence is described by the second Gaussian, with standard deviation of $\sigma_{\theta_x}$ centered around a most probable angle $\theta_p$. Figure 4.20 illustrates these dependencies.

The total fluence may be calculated as follows:

$$\Phi_T(x,y,z) = \int_{-\pi}^{+\pi}\int_{-\pi}^{+\pi} \Phi_\Omega(z,x,y,\theta_x,\theta_y)\,d\theta_x d\theta_y. \tag{4.94}$$

Because of symmetry in the $x$-$y$ plane, the integral over $\theta_x$ and the integral over $\theta_y$ are identical. The integral over $\theta_x$ is

$$\int_{-\pi}^{+\pi} \Phi_\Omega(z,x,\theta_x)\,d\theta_x = \frac{1}{4\sqrt{\pi A_2}} e^{-x^2/4A_2}\left[\text{erf}\left(\frac{\pi - xA_1/A_2}{2\sqrt{B/A_2}}\right) + \text{erf}\left(\frac{\pi + xA_1/A_2}{2\sqrt{B/A_2}}\right)\right], \tag{4.95}$$

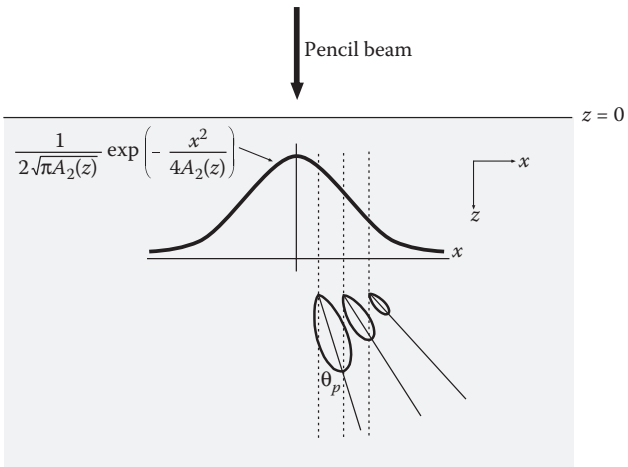



FIGURE 4.20 The differential fluence resulting from a pencil beam of electrons incident on a water phantom at $(x, y, z) = (0, 0, 0)$ traveling in the $+z$ direction. The differential fluence distribution is a (quasi) Gaussian (see Equation 4.92) in the lateral direction with standard deviation increasing with depth. The angular distribution is also Gaussian around a central value $\theta_p$, the most probable direction. The value of $\theta_p$ is proportional to the value of $x$.

where erf represents the error function. For a discussion of the properties of the error function, see Section 5.2. In the literature, it is assumed that the arguments of the error functions in Equation 4.95 are large (compared to 1) and positive (Jette, 1995). In this case, the term in the square brackets in Equation 4.95 will reduce to 2, and thus

$$\int_{-\pi}^{-\pi} \Phi_\Omega(z,x,\theta_x)\,d\theta_x \approx \frac{1}{2\sqrt{\pi A_2(z)}} e^{-x^2/4A_2(z)}. \tag{4.96}$$

It is not hard to see that in three dimensions (complete symmetry between $x$ and $y$), the total fluence may be expressed as

$$\Phi_T(x,y,z) \approx \frac{1}{4\pi A_2(z)} e^{-(x^2+y^2)/4A_2(z)}. \tag{4.97}$$

The integral of $\Phi_T$ in Equation 4.97 over the entire $x$-$y$ plane represents the total number of particles at any depth:

$$N_0 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi_T(x,y,z)\,dx\,dy = 1. \tag{4.98}$$

This presents a problem for the Fermi–Eyges theory in that the number of particles is independent of the depth regardless of the functional form of $A_2(z)$. There is no absorption built into the Fermi–Eyges theory. If this is to be applied to electron beam dose calculations, it will be necessary to force the electrons to have a finite range.

For $T(z)$ = constant in Equation 4.97, the limit in which $z \to 0$ is

$$\lim_{z\to 0} \Phi_T(x,y,z) = \lim_{A_2\to 0}\left(\frac{1}{2\sqrt{\pi A_2}} e^{-x^2/4A_2}\right)\left(\frac{1}{2\sqrt{\pi A_2}} e^{-y^2/4A_2}\right) = \delta(x)\delta(y). \tag{4.99}$$

The pencil beam is a delta function in $x$ and $y$ at the surface, as we expect.

Figure 4.21 shows the spatial distribution of the fluence for a pencil beam (Equation 4.97). It is assumed that $T_0 = 0.7$ cm$^{-1}$ (a constant). The fluence has a characteristic teardrop shape.

Let us assume that a broad electron beam consists of a sum of pencil beams of the type shown in Figure 4.21 incident upon a medium. At any particular depth, the energy of the electrons falls within a fairly narrow range (except perhaps at the end of the range), and the stopping power is relatively insensitive to the energy; therefore, the dose (per electron) from a single pencil beam is given by

$$d(x,y,z) = \int \Phi_E S_\rho(E)\,dE \approx \Phi(x,y,z) S_\rho\left[E(z)\right], \tag{4.100}$$

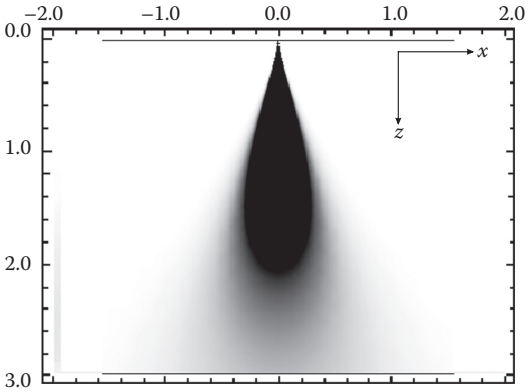



FIGURE 4.21 The spatial distribution of the pencil beam fluence (Equation 4.97) as a density plot (like radiographic film). We have assumed that $T(z)$ is constant and has a value of 0.7 radians$^2$/cm. The fluence has a characteristic teardrop shape. The width of the teardrop increases with increasing scattering power. Units are in cm.

where:

$S_\rho$ is the (restricted) mass collision stopping power (see Equations 4.55 and 4.56)
$E(z)$ is the mean electron energy at depth $z$ (see Equation 4.101)

The mean energy at depth $z$ is given by

$$E(z) = E_0\left(1 - \frac{z}{R_p}\right), \tag{4.101}$$

where:

$E_0$ is the incident energy
$R_p$ is the practical range (Khan and Gibbons, 2014)

The dose calculated using Equation 4.100 does not take account of secondary electrons (delta rays) or any x-ray contamination of the beam.

We can calculate the dose distribution for a broad beam normally incident on a phantom (with no angular divergence in the *incident* beam) by summing parallel pencil beams over the beam cross section:

$$D(x,y,z) = \iint\limits_{\substack{\text{collimator}\\ \text{at } z}} \Phi_T(x',y',0)\,d(x'-x,y'-y,z)\,dx'dy', \tag{4.102}$$

where $\Phi_T(x', y', 0)$ is the incident fluence. For numerical implementation in a treatment planning system, the double integral in Equation 4.102 would be replaced by a double

summation over $x'$ and $y'$. For a perfectly flat beam (incident fluence independent of off-axis position) with field size $L \times W$ (at the surface),

$$D(x,y,z) = \frac{S_\rho(z)\Phi_T(0)}{4\pi A_2(z)} \int_{-W/2}^{+W/2} \int_{-L/2}^{+L/2} \exp\left\{-\frac{1}{4A_2(z)}\left[(x-x')^2 + (y-y')^2\right]\right\} dx'dy', \quad (4.103)$$

where $\Phi_T(0)$ is the incident fluence. The integral in Equation 4.103 can be carried out analytically, and the result is

$$D(x,y,z) = \frac{1}{4} S_\rho(z)\Phi_T(0)\left\{\text{erf}\left[\frac{L(z)-2x}{2\sqrt{A_2(z)}}\right] + \text{erf}\left[\frac{L(z)+2x}{2\sqrt{A_2(z)}}\right]\right\} \times$$

$$\left\{\text{erf}\left[\frac{W(z)-2y}{2\sqrt{A_2(z)}}\right] + \text{erf}\left[\frac{W(z)+2y}{2\sqrt{A_2(z)}}\right]\right\}, \quad (4.104)$$

where

$$L(z) = L\left(1+\frac{z}{SSD}\right)$$

$$W(z) = W\left(1+\frac{z}{SSD}\right). \quad (4.105)$$

We note that the introduction of beam divergence here (Equation 4.105) is not totally consistent with the previous assumption of parallel, perpendicularly incident pencil beams.

The function $A_2(z)$ plays an important role in Equation 4.104. The value of $A_2(0) = 0$. This is not surprising, as the pencil beam is a delta function at the surface. We can calculate values of $A_2(z)$ by using Equation 4.90 along with the approximate expression for $T(z) = 3.72\ E^{-1.65}$ obtained from the fit in Figure 4.19. A plot of $2\sqrt{A_2}$ versus $z/R_p$ appears in Figure 4.22. This plot is not expected to be valid near $z/R_p = 1$, as the fit to $T(z)$ is only valid down to an energy of 0.5 MeV.

Let us examine the properties of Equation 4.104. For values of $|x| > 2$, erf($x$) has a constant value of 1.000 (or –1.000 if $x < 0$) to within 0.5% or less (see Figure 5.2). In Equation 4.104, the error function terms involving $x$ have a value of 1.000 ± 0.005 except in the regions $L/2–2\sqrt{A_2} < x < L/2 + 2\sqrt{A_2}$ and $–L/2 – 2\sqrt{A_2} < x < –L/2 + 2\sqrt{A_2}$. These are the penumbral regions. A similar statement holds for the $y$-dependent terms in Equation 4.104. As we have seen in Figure 4.22, the value of $2\sqrt{A_2}$ is less than a few centimeters, and therefore $2\sqrt{A_2}$ is generally small compared to $L/2$ or $W/2$.

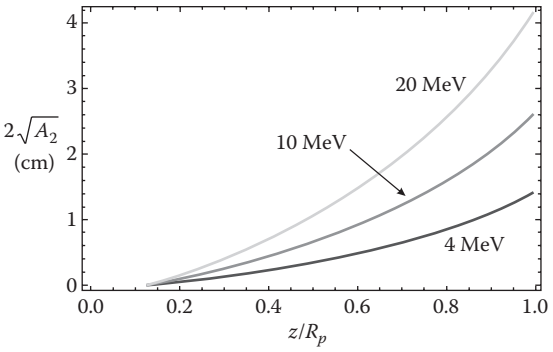



FIGURE 4.22 A plot of $2\sqrt{A_2}$ (in units of cm) as a function of $z/R_p$ for electrons in water with incident energies 4, 10, and 20 MeV. The value of $2\sqrt{A_2}$ is generally less than a few centimeters over the energy range of interest in radiation therapy.

For this reason, Equation 4.104 can be written approximately in terms of the central axis depth dose $D_0(0, 0, z)$:

$$D(x,y,z)=\frac{1}{4}D_0(0,0,z)\left\{\operatorname{erf}\left[\frac{L(z)-2x}{2\sqrt{A_2(z)}}\right]+\operatorname{erf}\left[\frac{L(z)+2x}{2\sqrt{A_2(z)}}\right]\right\}$$
$$\left\{\operatorname{erf}\left[\frac{W(z)-2y}{2\sqrt{A_2(z)}}\right]+\operatorname{erf}\left[\frac{W(z)+2y}{2\sqrt{A_2(z)}}\right]\right\}. \tag{4.106}$$

The region over which erf ≈ 1 in Equation 4.106 is the umbral region of the dose distribution. Equation 4.106 is expected to be a fairly good approximation except for small field sizes in which the dose distribution is completely penumbra.

In implementations of the Fermi–Eyges pencil beam in treatment planning systems, the values of $D_0(0, 0, z)$ are taken from measured central axis depth dose data. Equation 4.106 assumes that the incident fluence is uniform across the cross section of the beam (i.e., $\Phi(x', y', 0) = \Phi_0$) in Equation 4.102. This assumption can clearly be relaxed in a numerical implementation of this scheme.

An isodose plot of the dose given by Equation 4.106 is shown in Figure 4.23 for a 10 × 10 electron beam in which $T_0$ is assumed constant with a value of 0.7 (as in Figure 4.21), and $D_0(0, 0, z)$ is replaced by a measured depth–dose curve for a 10 MeV electron beam. The value $T_0 = 0.7$ gives a penumbra of about 10 mm, close to measured values. Figure 4.24 (to be compared to Figure 4.23) shows the isodose lines calculated using the Philips Pinnacle treatment planning software. The simple dose calculation model of Figure 4.23 is surprisingly realistic. It shows the characteristic behavior of low-energy electron beams in which the high isodose lines (e.g., 90%) are pinched inward and low isodose lines bulge out. For more accurate examples, see the

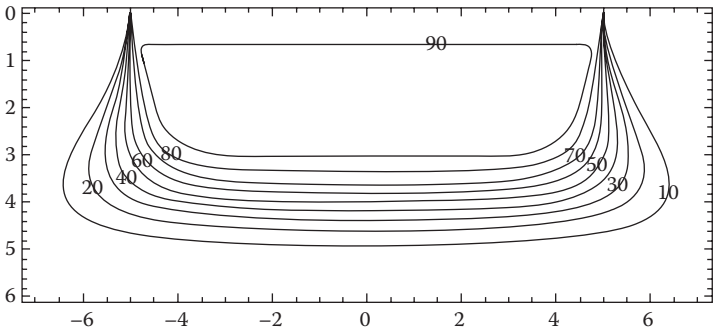



FIGURE 4.23 A calculated dose distribution based on Equation 4.106 showing isodose lines ranging from 90% down to 10% for a 10 cm × 10 cm, SSD = 100 cm, 10 MeV electron beam based on the assumptions described in the text. It is assumed that $T(z)$ has a constant value of 0.7 and that the incident fluence is uniform. The appearance is surprisingly realistic given these artificial assumptions. High-value isodose lines (such as 90%) are pinched inward and low-value isodose lines (10%) bulge out.

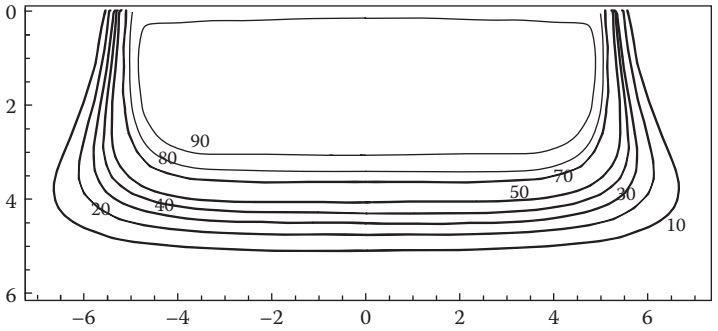



FIGURE 4.24 A Philips Pinnacle treatment planning system calculation of a 10 cm × 10 cm, SSD = 100 cm, 10 MeV electron beam dose distribution in a water phantom. This is to be compared to Figure 4.23, which shows a dose distribution calculated using Equation 4.106.

paper by Hogstrom et al. (1981). For a discussion of how tissue inhomogeneities affect this, see Hogstrom et al. (1981).

We can inquire about the length of the side of the equivalent square, $X$, for a rectangular field of dimensions $L \times W$. The condition for the equivalent square is $D_{\text{rect}}(0, 0, z) = D_{\text{square}}(0, 0, z)$. Let us assume that the equivalent square has a side length of $X$; from Equation 4.106, this condition requires that

$$\left\{\text{erf}\left[\frac{X(z)/2}{\sqrt{A_2(z)}}\right]\right\}^2 = \text{erf}\left[\frac{L(z)/2}{\sqrt{A_2(z)}}\right]\text{erf}\left[\frac{W(z)/2}{\sqrt{A_2(z)}}\right]. \tag{4.107}$$

As $A_2$ depends on $z$, there is, strictly speaking, no equivalent square; however, notice that

$$D_{L\times W}(0,0,z)=\left[D_{L\times L}(0,0,z)\times D_{W\times W}(0,0,z)\right]^{1/2}. \tag{4.108}$$

This is the well-known square root rule (see Khan and Gibbons, 2014). Equation 4.108 is quite useful for evaluating electron cutouts. For an irregularly shaped field, one makes a rough estimate of the dimensions of the equivalent rectangular field ($L \times W$). From this and a table of cutout factors for square fields ($L \times L$ and $W \times W$), the cutout factor for the irregularly shaped field can be estimated.

The Fermi–Eyges theory has been applied to the calculation of electron dose distributions by Brahme et al. (1981), Hogstrom et al. (1981), Jette et al. (1983), and others. This has been extended to layered homogeneous media and arbitrary field shapes. It is not surprising that there are limitations to the accuracy of the electron dose calculations described here. Near the end of the electron range, the scattering angles are expected to become large and electron transport is expected to become a diffusive process. This theory also has difficulty dealing with localized inhomogeneities. The Philips Pinnacle treatment planning system uses the Hogstrom pencil beam model, although the details of the implementation are not public (Philips Medical System, 2013).

## 4.11 CONCLUSION

The BTE is a deterministic integro-differential equation for the differential fluence of a radiation field. The differential fluence can be used to derive the dose distribution. The BTE is no less fundamental than Monte Carlo calculations, and it is not subject to stochastic variations. The BTE rivals Monte Carlo algorithms for accuracy, and it is faster. The BTE dose calculation option in the Varian Eclipse treatment planning system is more accurate than simple implementations of superposition algorithms. The BTE is faster than superposition algorithms for VMAT. The BTE can be used to derive the Fermi–Eyges electron transport equation. This electron transport model is used for electron dose calculations in some commercial treatment planning systems.

## PROBLEMS

1. Derive Equation 4.28 for the dose from Equation 4.26.

2. Derive the energy fluence equation for a point source from the integral in Equation 4.50. Far away from the source or when the source is small use the approximation that $\sin\theta' \approx \theta'$ in the integral in Equation 4.50.

3. Calculate $\Phi_{\Omega,E}$ on the axis of symmetry at a distance $d$ from a thin disk source of radius $R$ and thickness $h$ that radiates isotropically with $N(E)\,dE$ = total number of particles having energy between $E$ and $E + dE$ emitted. Find the total fluence $\Phi_T$ and show that when the distance from the disk is large, it acts like a point source.

4. a. Verify that Equation 4.49 for $\Phi_{\Omega,E}$ for a spherically symmetric source obeys the equation $\hat{\Omega}\cdot\vec{\nabla}\Phi_{\Omega,E} = 0$.

   b. Make a plot of $\Phi_E\,(4\pi R^2/N)$ versus $(r/R)$ for the spherical source of Figure 4.11. On the same graph, also plot this quantity for a point source. Use a domain of $1 < (r/R) < 2$.

5. Derive Equation 4.59 using Equations 4.57 and 4.58.

6. Show that the solution to the Fermi–Eyges equation (4.88) is given by Equations 4.89 and 4.90. Hint: You will need to use the Liebnitz rule for differentiation:

$$\frac{d}{d\alpha}\int_{a(\alpha)}^{b(\alpha)} f(x,\alpha)\,dx = \frac{db}{d\alpha} f\big(b(\alpha),\alpha\big) - \frac{da}{d\alpha} f\big(a(\alpha),\alpha\big) + \int_{a(\alpha)}^{b(\alpha)} \frac{\partial}{\partial\alpha} f(x,\alpha)\,dx$$

7. Find an expression for $\sigma^2_{\theta_x}$ and $\theta_p$ for the case in which $T = T_0$ (a constant).

8. Show that when $T = T_0$, a constant (no energy loss), the solution to the Fermi–Eyges equation (4.88) is given by

$$\Phi(z,x,\theta_x) = \frac{2\sqrt{3}}{\pi T z^2}\exp\left\{-\frac{4}{Tz}\left(3\left(\frac{x}{z}\right)^3 - 3\left(\frac{x}{z}\right)\theta_x + \theta_x^2\right)\right\}$$

9. Set up an Excel spreadsheet to plot an electron beam profile (Equation 4.106) using $T_0 = 0.5$ at a depth of $z = 3$ cm for a 10 cm × 10 cm field with SSD = 100 cm. Estimate the penumbra width (80%–20%).

10. a. Calculate the dose distribution on the central axis for a circular electron beam of radius $R$ that is perpendicularly incident on a large water phantom. Use Equation 4.102 and assume that the incident beam profile is completely "flat" (incident fluence is independent of lateral position). Assume that $S_\rho(z)$ is constant with depth (a fairly reasonable approximation).

    b. Discuss the form of the resulting expression for the depth dose. Is it realistic?

# SYMBOLS

| | |
|---|---|
| $A_n(z)$ | Parameter that appears in solution to Fermi–Eyges equation |
| $B(z)$ | Parameter that appears in solution to Fermi–Eyges equation |
| $D$ | Absorbed dose |
| $d(x, y, z)$ | Dose from a single electron pencil beam per electron |
| $\hat{e}_r, \hat{e}_\theta, \hat{e}_$ | Unit vectors in spherical coordinates |
| $E$ | Kinetic energy |
| $E_0$ | Initial incident electron energy |
| $K$ | Kerma |
| $K_c$ | Collision kerma |
| $L$ | Length of an electron field at the surface |
| $L(z)$ | Length of rectangular cross section electron beam |
| $m$ | Mass |
| $N(E)$ | Number of particles per unit energy interval |
| $\hat{n}$ | Unit normal vector to a surface |
| $N_0$ | Total number of particles |
| $Q$ | Change in rest energy |
| $q_u$ | Change in rest mass energy after scattering event in which a charged particle is produced with kinetic energy $E_u - E_{u'} - q_u$ |
| $q_{uu'}$ | Scattering of uncharged particle |
| $q_{c'c}$ | Scattering of charged particle |
| $q_{cu}$ | Production of uncharged particle from charged particle |
| $q_{uc}$ | Production of charged particle from uncharged particle |
| $\vec{r}$ | Position vector |
| $r$ | Radial coordinate |
| $R$ | Radius of sphere |
| $R_p$ | Electron practical range |
| $S$ | Stopping power (total) |
| $S_\rho$ | Mass stopping power |
| $S_o$ | Source of particles |
| SSD | Source-to-surface distance |
| $t$ | Time |
| $T(z)$ | Scattering power |
| $u$ | Coordinate distance along direction $\hat{\Omega}$ |
| $\vec{v}$ | Velocity |
| $V$ | Volume |
| $W$ | Width of an electron field at the surface |
| $X$ | Length of the side of the equivalent square of an electron field |
| $z$ | Depth in the medium |
| $\varepsilon_{tr}^n$ | Net energy transferred to charged particles from uncharged particles |

| | |
|---|---|
| $\hat{\Omega}(\theta', \phi')$ | Unit direction vector |
| $\Phi_{\Omega,E}(\vec{r}, E, \hat{\Omega})$ | Differential fluence |
| $\Phi_T(\vec{r})$ | Total fluence |
| $\Phi_E(\vec{r}, E)$ | Differential energy fluence |
| $\Phi_\Omega(\vec{r}, \hat{\Omega})$ | Differential angular fluence |
| $\Psi(\vec{r})$ | Energy fluence |
| $\Psi_E(\vec{r}, E)$ | Differential energy fluence |
| $d\vec{S}$ | Area element |
| $\mu(\vec{r}, E)$ | Attenuation coefficient |
| $\vec{\xi}$ | Difference vector between initial and scattered directions of particles |
| $\sigma_{\theta_x}$ | Standard deviation of angular dependence of differential fluence in solution of Fermi–Eyges equation |
| $(S_\rho)_{col}$ | Collisional mass stopping power |
| $(S_\rho)_{rad}$ | Radiative mass stopping power |
| $\varepsilon$ | Mean absorbed energy |
| $\varepsilon_{tr}$ | Energy transferred to charged particles |
| $\theta_p$ | Most probable angle in differential fluence solution to Fermi–Eyges equation |
| $\theta_x$, $\theta_y$ | Projection of scattering angle onto *x*-*z* and *y*-*z* planes, respectively |
| $\theta'$ | Direction coordinate |
| $\mu_a$ | Absorption portion of attenuation coefficient |
| $\mu_s$ | Scattering portion of attenuation coefficient |
| $\mu_{tr}/\rho$ | Mass energy transfer coefficient |
| $\varphi'$ | Direction coordinate |
| $d\Omega$ | Differential element of solid angle |

# REFERENCES


Attix, F.H. 1986. *Introduction to Radiological Physics and Radiation Dosimetry*. New York: Wiley.

Brahme, A., J. Lax, and P. Andreo. 1981. Electron beam dose planning using discrete Gaussian beams: Mathematical background. *Acta Radiol. Oncol.* 20, 147–58.

Bush, K., I.M. Gagne, S. Zavgorodni, W. Ansbacher, and W. Beckham, 2011. Dosimetric validation of Acuros® XB with Monte Carlo methods for photon dose calculations. *Med. Phys.* 38(4), 2208.

Chandrasekhar, S. 2011. *Radiative Transfer*. New York: Dover.

Eyges, L. 1948. Multiple scattering with energy loss. *Phys. Rev.* 74, 1534–35.

Han, T., et al. 2013. Dosimetric impact of Acuros XB deterministic radiation transport algorithm for heterogeneous dose calculation in lung cancer. *Med. Phys.* 40(5).

Hasenbalg, F., H. Neuenschwander, R. Mini, and E.J. Born. 2007. Collapsed cone convolution and analytical anisotropic algorithm dose calculations compared to VMC++ Monte Carlo simulations in clinical cases. *Phys. Med. Biol.* 52, 3679–91.

Hogstrom, K.R. and P.R. Almond. 2006. Review of electron beam therapy physics. Electron beam dose calculations. *Phys. Med. Biol.* 51, R455–89.

Hogstrom, K.R., M.D. Mills, and M.D. Almond. 1981. *Phys. Med. Biol.* 26(3), 445–59.
International Commission on Radiation Units (ICRU). 2011. Fundamental units and quantities for ionizing radiation. *J. ICRU* 11(1), Report 85.
Jette, D. 1995. *Electron Beam Dose Calculations in Radiation Therapy Physics*, ed. A.R. Smith. Berlin: Springer-Verlag.
Jette, D., A. Pagnamenta, L.H. Lanzl, and M. Rozenfeld. The application of multiple scattering theory to therapeutic electron dosimetry. 1983. *Med. Phys.* 10(2), 141–46.
Johns, H.E. and J.R. Cunningham. 1983. *The Physics of Radiology*, 4th ed. Springfield, IL: Charles C. Thomas.
Kan, M.W.K., P.K.N. Yu, and L.H.T. Leung. 2013. A review on the use of grid-based Boltzmann equation solvers for dose calculation in external beam treatment planning. *BioMed Res. Int.* 1–10.
Khan, F.M. and J.P. Gibbons. 2014. *The Physics of Radiation Therapy*, 5th ed. Baltimore: Lippincott, Williams and Wilkins.
Li, X.A. and D.W.O. Rogers. 1995. Electron mass scattering powers: Monte Carlo and analytical calculations. *Med. Phys.* 22(5), 531–41.
McMaster University. Radiative transport. Ch. 8 in Advanced Radiation Physics (MED PHYS 775) lecture notes. http://www.science.mcmaster.ca/medphys/images/files/courses/775/ch8.pdf.
Mihalas, D. 1978. *Stellar Atmospheres*. New York: W.H. Freeman.
National Council on Radiation Protection and Measurements (NCRP). 1991. Conceptual basis for calculations of absorbed-dose distributions. Report No. 108. Bethesda, MD: NCRP.
Philips Medical Systems. 2013. Instructions for use: Physics Pinnacle 3, release 9.6. Cleveland, OH: Philips Medical Systems.
Podgorsak, E.B. 2010. *Radiation Physics for Medical Physicists*, 2nd ed. Berlin: Springer.
Vassiliev, O.N., et al. 2010. Validation of a new grid-based Boltzmann equation solver for dose calculation in radiotherapy with photon beams. *Phys. Med. Biol.* 55, 581–98.

## ENDNOTES

1. Varian Eclipse ACUROS XB Advanced Dose Calculation.
2. All the quantities in this section refer to a particular type of particle rather than all types of particles, as in the previous section.
3. Note that the vector $\hat{\Omega} = \Omega_r \hat{e}_r + \Omega_\theta \hat{e}_\theta + \Omega_\varphi \hat{e}_\varphi$, where $\Omega_r$ and so on are constants (independent of θ and φ), is *not* a constant vector.
4. Note that there appears to be a sign error for this expression in Equation 6.28 of NCRP (1991).
5. The probability of scattering becomes zero when $E' > E$.
6. We invite the ambitious reader to generalize the discussion here to include this.
7. There does not appear to be a consensus on the pronunciation of Eyges. One possibility is that Eyges rhymes with egregious.
8. The reader may notice a rather conspicuous contradiction. We stated that Eyges extended Fermi's analysis by including energy losses. Energy losses will be put back in later in (what appears to this author) a non-self-consistent fashion.
9. The energy dependence has now crept back in, in a somewhat arbitrary fashion.

# 5

# TUMOR CONTROL AND NORMAL TISSUE COMPLICATION PROBABILITY MODELS IN RADIATION THERAPY

A model is a lie that helps you see the truth.

**—Howard Skipper**

## 5.1 INTRODUCTION

Absorbed dose is a physical quantity that is at best a surrogate for treatment outcome. What really matters to patients is the likelihood that their tumor will be eradicated and the chance of complications. There are a variety of biological indices that address the issue of outcome, such as tumor control probability (TCP), normal tissue complication probability (NTCP), probability of uncomplicated control, and equivalent uniform dose (EUD). It is to be noted that local tumor control is a necessary but insufficient condition for successful treatment, and therefore TCP is perhaps overemphasized. Local tumor control is only of palliative benefit if the patient succumbs to distant metastasis.

Biological treatment indices are based on two types of models: mechanistic models and empirical models. Empirical models are based on mathematical

functions that are simply fit to observed data. One of the difficulties associated with this is that observed data at high complication probability are rare, as, of course, complications are assiduously avoided. Empirical models offer no guarantee that the probabilities are accurate outside the range of measured values. The mechanistic models are based on cell survival (the linear quadratic model), tissue architecture, and probability theory. These models have a number of free, unknown parameters.

Physicists like to play with mathematics, and therefore there is sometimes a tendency to take mechanistic models too seriously. The reason for this is that mathematics is an extremely effective tool for physicists. In fact, it is the language of physics. Eugene Wigner (1960), one of the great theoretical physicists of the twentieth century, has written on the "unreasonable effectiveness of mathematics." Biological systems are extremely complex. Mathematical descriptions of physical systems are often based on derivations from first principles. Physical systems are generally simpler, and therefore mathematical descriptions of them can be extraordinarily accurate. The medical physicist Goitein (2008) has commented on this. He feels that medical physicists take mechanistic models too seriously and radiation oncologists do not take them seriously enough.

Biological indices are based on numerous assumptions and simplifications. Therefore, they should all be taken with a giant grain of salt. They are generally not very accurate in making absolute outcome predictions. One may therefore wonder why we bother with them. It is because they may be of use in comparing rival treatment plans *when there is no clear dosimetric preference.* Biological models are not yet widely used clinically, although a task group report from the American Association of Physicists in Medicine (AAPM) on clinical application has appeared within the past few years (Li et al., 2012). The models force us to think carefully about some of the factors governing tumor control and complications. In addition, they may provide guidance for gathering future data on complications and tumor control.

All of the models for both NTCP and TCP are based on dose–response curves in which the probability is plotted as a function of dose. Response curves are sigmoidal in shape (Figure 5.1), and there are two quantities that are widely quoted to characterize these curves: $D_{50}$ and $\gamma_{50}$. $D_{50}$ is the dose for which there is a 50% response (either NTCP = 50% or TCP = 50%). The quantity $\gamma_{50}$ is a dimensionless parameter that describes the slope of the dose–response curve at a dose of $D_{50}$:

$$\gamma_{50} = D\left.\frac{\partial(\mathrm{N})\mathrm{TCP}}{\partial D}\right|_{D_{50}} \tag{5.1}$$

There are many mathematical functions that correspond to a sigmoid shape. For specific values of $D_{50}$ and $\gamma_{50}$, *there is no unique mathematical function describing a sigmoid shape.*

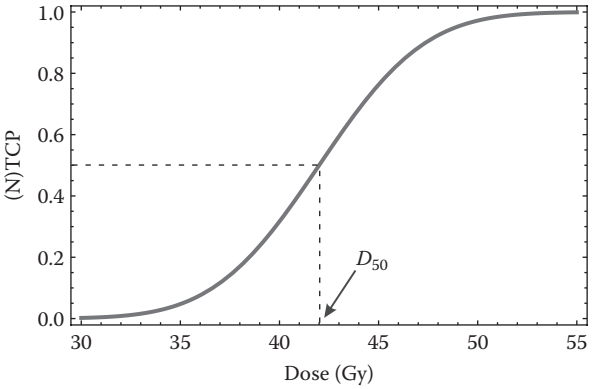



FIGURE 5.1 A dose–response curve is a plot of TCP or NTCP vs. dose. These curves are sigmoidal in shape and are often characterized by $D_{50}$ and $\gamma_{50}$. The quantity $D_{50}$ is the dose at which the probability is 50% (42 Gy in this example). The quantity $\gamma_{50}$ is related to the slope of the curve evaluated at $D_{50}$ (see Equation 5.1; $\gamma_{50} = 4.0$ for this curve).

There are numerous questions that we would like biological models to address, among them are:

1. If two dose volume histograms (DVHs) for organs at risk from rival treatment plans cross, which one is preferred?
2. Is it better to give a small amount of dose to a large part of an organ at risk (OAR) or a large dose to a small amount of volume?
3. How do mechanistic models relate to empirical models? Do mechanistic models "predict" empirical models?
4. Do mechanistic models predict a threshold volume for normal tissue complication, and if so, what are the values?
5. How does TCP depend on tumor volume?
6. For a target, when, if ever, does the TCP depend only on the minimum dose?
7. What are the effects of a small cold spot on TCP?
8. How harmful is it to delay a patient's treatment by one cell doubling time?

There are some definitions to be introduced here at the outset that are enumerated in the following bulleted list:

- Partially uniform irradiation: (For an OAR) a portion of the volume receives a constant dose and the remainder receives no dose (zero). In this situation, there are only two dose values, $D$ and 0. This nomenclature is somewhat of an oxymoron. This is an idealization because it implies an infinite gradient in the dose. In an inhomogeneous irradiation, there are more than two values of the dose in the OAR.
- Clonogenic cell: A cell that is capable of sustained proliferation or reproduction.

The descriptions in the literature of various biological indices are usually expressed in terms of discrete volumes. This is understandable because computerized treatment planning systems calculate doses in a volumetrically discrete dose matrix. The actual dose distribution in a patient is a continuous variable, however (at least down to the microdosimetry level). There are two reasons why it may make sense to describe biological indices in terms of a continuous dose distribution: (1) real dose distributions are continuous, and (2) the mathematics of continuous functions is often easier to work with than the mathematics of discrete functions. It is usually easier to derive analytic results using continuous variables. We will switch back and forth between continuous and discrete descriptions as convenient.

One of the most up-to-date, comprehensive reviews of normal tissue complication data is the quantitative analysis of normal tissue effects in the clinic (QUANTEC) report (see Marks et al., 2010b). This represents a multidisciplinary effort sponsored jointly by the AAPM and the American Society for Radiation Oncology (ASTRO). The report is available on the AAPM website (http:/www.aapm.org). Other references are the AAPM TG 166 report: "The Use and QA of Biologically Related Models for Treatment Planning" (Li et al., 2012), *The Physics of Radiotherapy X-Rays and Electrons* (Metcalfe et al., 2007, chapter 14), "Biological Indices for Evaluation and Optimization of IMRT" (Yorke, 2003), "Radiobiological Modeling for Treatment Planning" (Moiseenko et al., 2005), *Radiation Oncology: A Physicist's-Eye View* (Goitein, 2008, chapter 5), "Dose-Volume Considerations: An Update for Use in Treatment Planning" (Yorke et al., 2013), and *The Physics of Conformal Radiotherapy* (Webb, 1997, chapter 5).

Before we discuss models, we very briefly review some elements of probability theory (Section 5.2). Section 5.3 contains a discussion of analytic forms of DVHs. Normal tissue complication probabilities are discussed in Section 3.4, including both empirical and mechanistic models. In Section 5.5, we discuss TCP models, both empirical and mechanistic. In the case of mechanistic models, we include an analysis of both intra- and intertumor variations in radiosensitivity. The probability of uncomplicated control is considered in Section 5.6. A list of symbols used in this chapter can be found at the end.

## 5.2 SOME ELEMENTS OF PROBABILITY THEORY

Mechanistic descriptions of NTCP and TCP employ probability calculations. We therefore review *very briefly* some of the necessary probability theory.

For two successive *independent* events, the probability that both $A$ and $B$ will happen is $p(A)\,p(B)$, where $p(A)$ is the probability of $A$ and $p(B)$ is the probability of $B$. The probability that either $A$ or $B$ occurs is $p(A) + p(B)$, provided that $A$ and $B$ are mutually exclusive.

The binomial probability distribution applies to the situation in which there are independent repeated trials with constant probabilities (Bernoulli trials) and there are only two outcomes: success or failure. An example is a coin toss. The probability

of success is $p$, and the probability of failure is $q = 1 - p$. The probability $f(x)$ of exactly $x$ successes in $m$ trials is

$$f(x) = {}_mC_x p^x q^{m-x}, \tag{5.2}$$

where

$${}_mC_x = \frac{m!}{x!(m-x)!}. \tag{5.3}$$

The average of the binomial distribution is

$$\mu = \sum_{x=0}^{x=m} x f(x) = mp, \tag{5.4}$$

and the standard deviation is $\sigma = \sqrt{mpq}$.

The cumulative binomial distribution is given by

$$P(\le y) = \sum_{j=0}^{\lfloor y \rfloor} {}_mC_j p^j (1-p)^{m-j}, \tag{5.5}$$

where $\lfloor y \rfloor$ is the greatest integer less than or equal to $y$. This is the probability of getting $y$ successes or fewer in $m$ trials.

For a very large number of trials in which the probability $p$ is quite small, the binomial distribution can be approximated by the Poisson distribution ($p \to 0$, $m$ very large, but $mp$ remains finite):

$$f(x) = \frac{\mu^x e^{-\mu}}{x!}. \tag{5.6}$$

The binomial distribution can be approximated by the Gaussian or normal distribution when $|mp - x| \ll mp$, that is when $x$ is not very far from the average. In this case $x$ is treated as a continuous variable:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \tag{5.7}$$

where $f(x)\,dx$ is the probability that $x$ lies between $x$ and $x + dx$, and $\mu$ and $\sigma$ are as given previously for the binomial distribution.

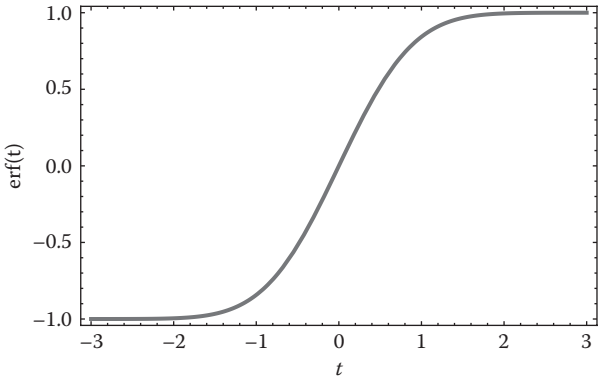



FIGURE 5.2 The error function erf($t$) has the characteristic sigmoid shape that is typical of NTCP and TCP curves. See Equation 5.8.

The error function is related to the cumulative normal probability distribution. The error function is defined as

$$\mathrm{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-y^2} dy. \tag{5.8}$$

Numerical values of the error function (Figure 5.2) can be found in standard statistical tables or in spreadsheet software such as Excel. Some properties of the error function that are useful include erf(–$t$) = –erf($t$), an odd function; erf(+∞) = 1; and erf(–∞) = –1. From the definition of the error function,

$$\frac{d}{dz}\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} e^{-z^2}. \tag{5.9}$$

## 5.3 DVHs

We assume that the reader has some familiarity with DVHs and their clinical interpretation. DVHs play an important role in treatment plan evaluation and are related to some of the biological indices discussed here. DVHs provide information about how the dose is statistically distributed over the volume.

Let $h(D)dD$ be the amount of volume with a dose between $D$ and $D + dD$. The total volume, $V$, is therefore

$$V = \int_0^{D_{\max}} h(D) dD, \tag{5.10}$$

where $D_{max}$ is the maximum dose received by a point in the volume. If the dose is spatially constant throughout the volume, then $h(D)$ is proportional to a delta function. A graph of $h(D)$ versus $D$ is an ordinary histogram or a differential DVH.

The average dose is

$$\bar{D} = \frac{1}{V}\int_0^{D_{max}} Dh(D)dD. \tag{5.11}$$

The cumulative DVH (cDVH), $V(D)$, is the volume receiving a dose of at *least* $D$. This can be calculated from a differential DVH as follows:

$$V(D) = \int_D^{D_{max}} h(D)dD. \tag{5.12}$$

It is the area under the differential DVH from $D$ to $D_{max}$.

Expressed in dimensionless form (relative volume), the cDVH will be written as $H(D) = V(D)/V$, where

$$H(D) = 1 - \frac{1}{V}\int_0^{D} h(D)dD. \tag{5.13}$$

The differential DVH $h(D)$ is proportional to the (negative) of the slope of the cDVH $H(D)$:

$$h(D) = -V\frac{dH}{dD}. \tag{5.14}$$

The average value of the dose is simply the area under the cDVH curve:

$$\bar{D} = \int_0^{D_{max}} H(D)dD. \tag{5.15}$$

This can be proven by substituting Equation 5.14 into Equation 5.11.

In the clinic, we do not have an analytic expression for $H(D)$, but rather a dose matrix calculated by treatment planning software. In this case, the integrals in Equations 5.10 through 5.15 must be converted to sums over the voxels in the structure of interest. Nevertheless, we will find the integrals in this section useful for deriving general results.

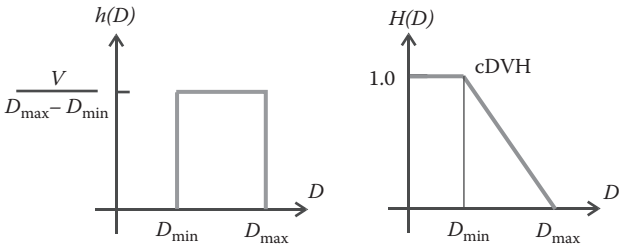



FIGURE 5.3 Archetypal differential DVH (left) and corresponding cDVH (right) for either a target or an OAR. For a target, $D_{min}$ should not be much less than $D_{max}$. For an OAR, $D_{min}$ may be as small as zero. Although real patient cDVHs may show more structure (bumps, wiggles, etc.), they are frequently of this general form.

For a constant dose $D = D_t$, such as in a target volume, $h(D)$ is a Dirac delta function and $H(D)$ is a step function with $H(D) = 1$ for $D < D_t$ and $H(D) = 0$ for $D > D_t$. Figure 5.3 shows a more general dose distribution in which $h(D) = V/(D_{max} - D_{min})$ for $D_{min} < D < D_{max}$ and zero everywhere else. In this case, $H(D) = 1$ for $D < D_{min}$, $H(D) = 0$ for $D > D_{max}$, and between $D_{min}$ and $D_{max}$ we have

$$H(D) = 1 - \frac{1}{\Delta D}(D - D_{min}), \tag{5.16}$$

where $\Delta D = D_{max} - D_{min}$. For $D_{min} \approx D_{max}$, this will be like a target cDVH. On the other hand, if $D_{min} \ll D_{max}$ (or $D_{min} = 0$), this is like a cDVH for an OAR. Actual patient target cDVHs may have curvature and "wiggles," but they are often qualitatively similar to Figure 5.3. Therefore, Figure 5.3 may be considered an archetypal cDVH for either a target or a normal structure.

## 5.4 NORMAL TISSUE COMPLICATION PROBABILITY

The probability of complication in normal tissues due to exposure to radiation depends on a large number of factors, such as the amount and distribution of the dose. The dose received by an OAR is not likely to be the full prescribed dose per fraction. The dose distribution in an OAR is frequently very nonuniform by design as part of the effort to decrease complications. The effect on the complication probability of the volume of the OAR irradiated and the dose it receives is called a volume effect. One of the major facets of NTCP models is an attempt to deal with the volume effect.

Other confounding and contributing factors include

- Dose per fraction, which may vary with fraction due to boosts.
- Chemotherapy: Concurrent chemo exacerbates the severity of normal tissue complications.

- Host factors, for example, chronic liver disease, lifestyle, and genetic factors.
- Follow-up duration.
- The endpoint: There can be various types of complications for a given OAR. The endpoint must be carefully specified.

Some useful dose descriptors for evaluation of complication probability are the mean dose $\bar{D}$, the maximum dose $D_{max}$, the equivalent uniform dose (EUD), and the cutoff volume. An example of the cutoff volume is $V_{20}$, the volume (or percent volume) receiving at least 20 Gy.

As discussed in the introduction, there are two types of models for normal tissue complication:

1. Empirical models
2. Mechanistic models

Empirical models are based on phenomenological mathematical fits to clinical data. Mechanistic models are based on cell survival curves and statistical and probability analysis.

One criticism of the models discussed here (both empirical and mechanistic) is that they are binary—a complication either occurs or it does not. The models do not take into account the severity of the OAR response to irradiation.

## 5.4.1 EMPIRICAL MODELS

Empirical models are based on the use of a fitting function that reproduces the NTCP data. These models should therefore be reasonably accurate over the range of data to which they are fit; there is no guarantee, however, that they will be in the least bit accurate outside the range of the data used to obtain the fits.

### 5.4.1.1 DOSE VOLUME EFFECTS

Both mechanistic and empirical models require a method for dealing with nonuniform dose distributions. In mechanistic models, the volume dependence is built into the construction of the model. For empirical models, however, a separate analysis of dose volume effects must be developed. We now consider this problem. There are two types of spatially variable dose distributions that we will consider: partially uniform irradiation (see introduction) and inhomogeneous irradiation, in which the dose varies arbitrarily throughout the entire volume. Let us consider the simpler case first.

An early formulation of the dose volume effect applies to an OAR that receives a uniform irradiation to only a part of the OAR. Let $\nu$ equal the volume fraction of the OAR that receives any dose, $D(\nu = 1) = D(1)$ the uniform tolerance dose received by the full organ (this could be $D_{50}$), and $D(\nu)$ = the tolerance dose for the partial irradiation of fractional volume $\nu$. It is assumed, but not often stated, that the dose per fraction is the same in both cases. $D(1)$ can be used to read off or look up the NTCP from

a graph of whole-organ NTCP as a function of dose, such as shown in Figure 5.1. It has been found empirically that roughly

$$D(\mathrm{v}) = D(1)\mathrm{v}^{-n}, \tag{5.17}$$

where $n$ is a tissue-specific parameter describing the magnitude of the volume effect ($n > 0$ always; otherwise, the tolerance dose for partial irradiation would be less than that for whole-organ irradiation). Burman et al. (1991) compiled data for 28 normal tissues with defined endpoints. The values of $n$ were found to range from 0.01 (ear) to 0.87 (lung). When $n$ is small, the relative tolerance dose depends weakly on the irradiated volume. As $n$ becomes large, the tolerance dose depends sensitively on the fraction of the volume irradiated.

#### 5.4.1.2 EQUIVALENT UNIFORM DOSE

Tissues and organs invariably receive nonuniform doses. It would be very useful if one could calculate a value for a uniform dose that would lead to the same NTCP or TCP as the nonuniform dose distribution. Such a dose is called the EUD. There are many formulas that have been developed to calculate the EUD. The specific formula depends on the NTCP or TCP model. One of the most widely used expressions for the EUD is sometimes called the generalized equivalent uniform dose. It is a generalization of the formulation expressed by Equation 5.17, where the EUD corresponds to $D(1)$. It is the EUD that is believed to yield the same complication probability as the nonuniform dose (corresponding to $D(\mathrm{v})$ in Equation 5.17). It is computed by carrying out a sum over all of the voxels in the OAR. The definition is

$$\mathrm{EUD} = \left[\frac{1}{N}\sum_{i=1}^{i=N} D_i^{1/n}\right]^n, \tag{5.18}$$

where:

- $D_i$ is the dose received by voxel $i$ and there are a total of $N$ voxels
- $n$ is the volume effect parameter, as in Equation 5.17.[1]

It is assumed that the voxels are sufficiently small that the dose is constant within a voxel and that all voxels have the same volume. In some references, the parameter $n$ is replaced by a parameter $1/a$.

If the OAR consists of two portions, one of which receives a uniform dose $D(\mathrm{v})$ and the other a zero dose (partially uniform irradiation), then $\mathrm{EUD} = D(\mathrm{v})\mathrm{v}^n = D(1)$ as in Equation 5.17. Therefore, EUD is a generalization of Equation 5.17. The EUD is a type of average dose. If $n = 1$, the EUD is simply the arithmetic mean dose for the OAR. When $n = ½$, the EUD is the root mean square value of the dose. In the limit in which $n$ is small and positive (i.e., $n \to 0^+$), $\mathrm{EUD} \to D_{\max}$, the maximum dose in the OAR.

In the case of a target, and in the limit in which $n$ is small and negative (i.e., $n \to 0^-$), EUD $\to D_{min}$, the minimum dose.

The EUD can be calculated from either the differential or cDVH, as follows (see problem 5):

$$\mathrm{EUD}^{1/n} = \frac{1}{V}\int_0^{D_{max}} h(D)D^{1/n}dD = \frac{1}{n}\int_0^{D_{max}} H(D)D^{\frac{1}{n}-1}dD. \tag{5.19}$$

Note that if the dose is constant throughout ($D = D_{max}$), $H(D) = 1.0$ between $D = 0$ and $D = D_{max}$, and therefore EUD = $D_{max}$, as expected.

In the situation where we have a DVH of the sort shown in Figure 5.3, $H(D) = 1$ for $D < D_{min}$ and $H(D) = 0$ for $D > D_{max}$; between $D_{min}$ and $D_{max}$, $H(D)$ is given by Equation 5.16. Under these circumstances, the integral in Equation 5.19 can be easily evaluated:

$$\mathrm{EUD} = \left[\frac{D_{max}^{1/n+1} - D_{min}^{1/n+1}}{(1+1/n)\Delta D}\right]^n. \tag{5.20}$$

In the case where $n = 1$, this is simply the average dose or the area under the cDVH curve. When $n$ is positive and becomes small, EUD approaches $D_{max}$, as expected (see Problem 6). In the case where $n$ is negative and becomes small, EUD approaches $D_{min}$. The latter limit only applies to a target.

Figure 5.4 shows a graph of EUD/$D_{max}$ versus $D_{min}/D_{max}$ based on Equation 5.20 and with values of $n$ taken from Table 5.1 for liver ($n = 0.86$), kidney ($n = 0.7$), rectum

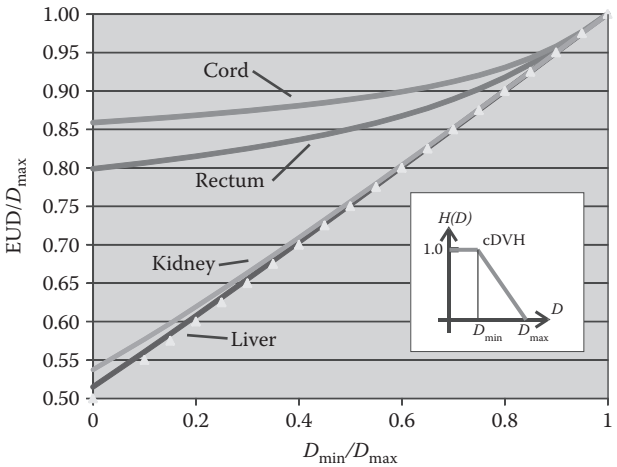



FIGURE 5.4 EUD/$D_{max}$ vs. $D_{min}/D_{max}$ for various values of $n$ (see text) corresponding to cervical cord, rectum, liver, and kidneys. The triangles show the average dose. These graphs are based on Equation 5.20 for the cDVH shown in the inset. The EUDs for liver and kidney are close to the average dose, but the cord and rectum have EUDs much closer to $D_{max}$. The $n$ values are from Table 5.1.

**TABLE 5.1 Selected OAR Model Parameters**

| Organ/Tissue | $D_{50}$ (Gy) (95% CI) | $m$[a] | $\gamma_{50}$ | Model | $n$ (95% CI) | $\alpha/\beta$ (Gy) (95% CI) | Endpoint | Reference |
|---|---|---|---|---|---|---|---|---|
| Lungs | 31.4 (29.0–34.7) | 0.45 (0.39–0.51) | 0.89 | LKB | 1.03 (±0.36) | 4.0 (±0.9)[b] | Radiation pneumonitis | Marks et al. (2010a) |
| Liver | 43.0 (39.8–46.1) | 0.12–0.31 | 1.9 | LKB | 0.86–1.1 | 2.0–3.0 | Radiation-induced liver disease | Pan et al. (2010) |
| Kidney (bilateral) | 28[c] | 0.10[c] | 4.0 | LKB | 0.7[c] | ? | Radiation-induced kidney injury | Dawson et al. (2010) |
| Rectum | 76.9 (±3.2) | 0.13 | 3.1 | LKB | 0.09 (±0.05) | 3.0 | ≥Grade 2 toxicity, rectal bleeding | Michalski et al. (2010) |
| Cervical spinal cord | 69.4 (±3.1) | 0.085 | 4.7 | Logistic | 0.05 | 0.87 (±0.33) | Myelopathy grade ≥ 2 | Kirkpatrick et al. (2010), Schultheiss (2008) |

[a] See Equation 5.24.
[b] Uncertainty is ±1 standard error.
[c] Based on data from Emami et al. (1991).

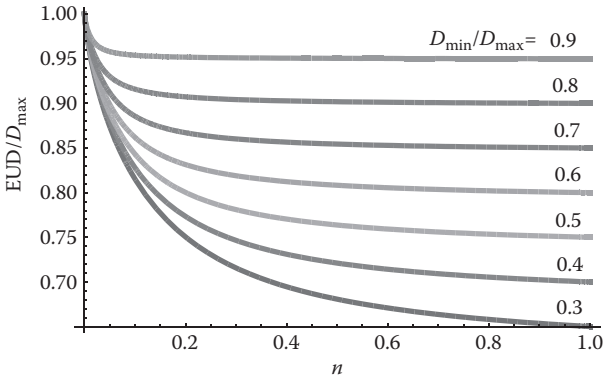



FIGURE 5.5 A graph of EUD/$D_{max}$ vs. $n$ for various values of $D_{min}/D_{max}$ based on Equation 5.20. When $n \gtrsim 0.5$, the value of $\mathrm{EUD} \approx \bar{D}$ unless $D_{min}/D_{max} \leq 0.3$. When $D_{min}/D_{max}$ is greater than about 0.8, the EUD is equal to the average dose unless $n$ is very small.

($n = 0.09$), cervical cord ($n = 0.05$), and the average dose ($n = 1$, triangles). The liver and kidney have an EUD almost equal to the average dose, whereas the rectum and spinal cord have an EUD much higher than average, even for $D_{min} = 0$.

Figure 5.5 shows EUD/$D_{max}$ based on Equation 5.20 as a function of $n$ for various values of $D_{min}/D_{max}$ ranging from 0.3 to 0.9. For values of $n \gtrsim 0.5$, the value of $\mathrm{EUD}/D_{max} \approx \bar{D}/D_{max}$ unless $D_{min}/D_{max} \lesssim 0.3$.

We can now answer question (1) posed in the introduction: If two DVHs from rival treatment plans cross, which one is preferred? The one with the smallest EUD is favored. To the extent that $n \approx 1$ (and average dose determines complications), the cDVH with the least area is preferred. For small values of $n$, the cDVH having the smallest $D_{max}$ is best.

#### 5.4.1.3 EMPIRICAL EXPRESSIONS FOR NTCP

Historically, empirical models were generally derived from data for complication probability under uniform irradiation. There are two parameters that are frequently cited to summarize the dose–response relationship:

1. $D_{50}$ = tolerance dose resulting in 50% complication rate
2. Slope of the response curve at $D_{50}$

One must specify the type of tissue and the particular endpoint. $D_{50}$ is not always well known since physicians generally avoid giving doses anywhere near that required to cause a 50% chance of a complication.

The Lyman model (sometimes called the integrated normal or probit model) describes the NTCP for an OAR irradiated to a uniform dose over a partial volume v

(Lyman, 1985; Webb, 1993). The partial volume is the fraction of the volume receiving a dose $D$, the remainder receiving zero dose. In the Lyman model, NTCP is given by

$$\text{NTCP} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx, \tag{5.21}$$

where $x$ is a dummy integration variable and

$$t = \frac{D - D_{50}(\nu)}{m D_{50}(\nu)}, \tag{5.22}$$

where:

$D_{50}(\nu)$ is the dose for which NTCP = 50% for partial volume $\nu$
$m$ is a quantity related to the slope of the NTCP curve at $D_{50}$ and thus characterizes the shape of the NTCP curve

The slope of the NTCP curve at $D_{50}$ can be described by the dimensionless quantity

$$\gamma_{50} = \left. \frac{\partial \text{NTCP}}{\partial \ln D} \right|_{D_{50}}. \tag{5.23}$$

The relation between $\gamma_{50}$ and $m$ is

$$\gamma_{50} = \frac{1}{m\sqrt{2\pi}}. \tag{5.24}$$

Although Equation 5.21 may seem rather nonintuitive, it is related to the error function (see Figure 5.2 and Equation 5.8), and it does produce a sigmoidal NTCP versus dose curve. In terms of the error function,

$$\text{NTCP} = \frac{1}{2}\left[1 + \text{erf}\left(\frac{t}{\sqrt{2}}\right)\right]. \tag{5.25}$$

When $\nu = 1.0$ and $D = D_{50}$, NTCP = 0.5, as expected. Note that when $D = 0$, NTCP is not zero, as $t = -1/m$; therefore, this model may not be accurate for small values of NTCP.

Three parameters are necessary for a calculation of NTCP for partially uniform irradiation (see definition in Section 5.1): $D_{50}(1)$, $n$, and $m$. Burman et al. (1991) have tabulated these values for 28 tissues based on $D_5$ and $D_{50}$ data of Emami et al. (1991). These factors have been updated for many OAR in the QUANTEC report. Table 5.1 provides data for a few OAR. We had hoped to include the bladder in this table, but according to Viswanathan et al. (2010, p. S120), "no quantifiable models are available that satisfactorily describe the observed serious late bladder toxicity after EBRT."

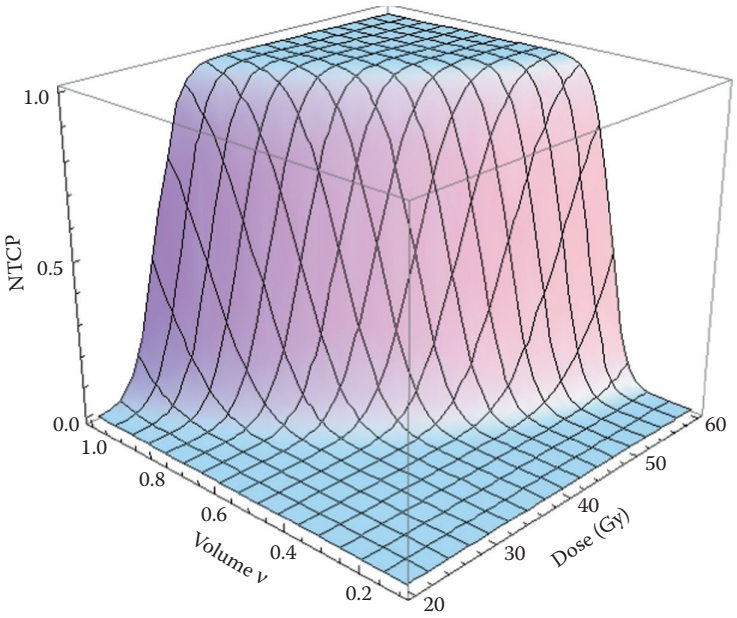



FIGURE 5.6 The NTCP for kidneys as a function of dose and partial volume ν for a Lyman model with parameters $D_{50}(1) = 28$ Gy, $m = 0.10$ ($\gamma_{50} = 4.0$), and $n = 0.7$. The partial volume is the fraction of the volume receiving a uniform dose; the remainder of the volume receives no dose.

Figure 5.6 shows a three-dimensional plot of NTCP as a function of dose and partial volume for radiation-induced kidney injury (bilateral irradiation). The parameters are $D_{50}(1) = 28$ Gy, $\gamma_{50} = 4.0$, and $n = 0.7$. Figure 5.7 shows a contour plot corresponding to Figure 5.6 for constant values of NTCP in a partial volume–dose plane. Figure 5.8 shows Lyman-based NTCP dose–response curves for the data in Table 5.1.

Another empirical expression for NTCP is the logistic function:

$$\text{NTCP} = \frac{1}{1+\left(D_{50}/D\right)^{4\gamma_{50}}}. \tag{5.26}$$

Although the graph of this formula for NTCP has the same general shape as the Lyman formula, the two curves are not the same. For the logistic formula, NTCP = 0 when $D = 0$.

It is important to understand that values of $D_{50}$ and $\gamma_{50}$ are model dependent. Values derived by fitting a particular model (e.g., Lyman–Kutcher–Burman [LKB] or logistic) should be used with caution, if at all, in another model because predictions made with another model may differ significantly, particularly for low doses (i.e., $\ll D_{50}$).

To handle inhomogeneous irradiation within the context of the Lyman model, some method is needed to find a value of $D$ to substitute into Equation 5.22. The cDVH describes the distribution of dose, and therefore some procedure is needed

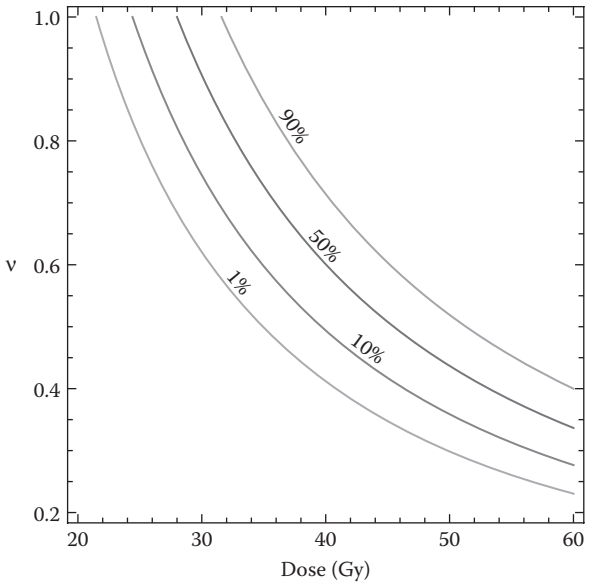



FIGURE 5.7 A contour plot corresponding to Figure 5.6 showing lines of constant NTCP as a function of dose and partial volume irradiated for kidneys. The upper right is the dose volume region to be avoided.

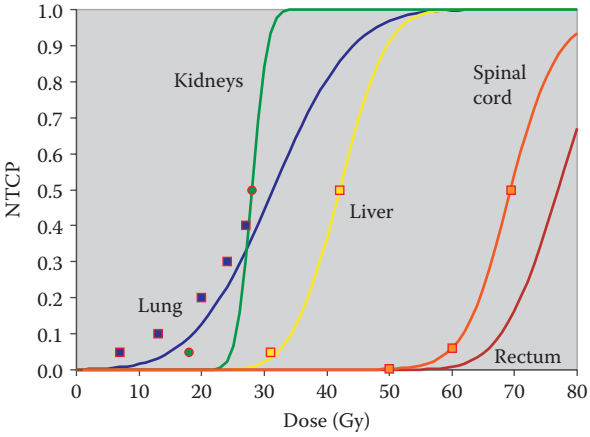



FIGURE 5.8 NTCP data based on the Lyman model with parameters from Table 5.1. Individual recommended dose volume constraint points have been included from the QUANTEC tables. The individual points tend to be above (higher NTCP) the model curves, presumably to err on the side of caution. Data points for lung, liver, and kidney are mean doses. The dose points for the spinal cord are maximum doses.

to condense the entire cDVH into a value of $D$, an equivalent dose for whole-organ irradiation. This is called DVH reduction, and there are numerous algorithms that have been proposed in the literature to accomplish this. One possibility is to set $D$ = EUD and substitute $D_{50}(1)$ for $D_{50}(v)$. Therefore, for inhomogeneous irradiation,

$$t = \frac{\mathrm{EUD} - D_{50}(1)}{mD_{50}(1)}. \tag{5.27}$$

Equation 5.18 or 5.19 can be used along with the cDVH to compute the EUD. The Lyman model, along with Equation 5.27, is frequently referred to as the LKB model.

## 5.4.2 MECHANISTIC MODELS

### 5.4.2.1 LINEAR QUADRATIC CELL SURVIVAL

This is very a brief review of cell survival probability. For more detail, the reader is referred to the excellent texts by Hall and Giaccia (2012) and Joiner and van der Kogel (2009).

The most commonly used expression for the probability (or fraction) of cells of a particular type that survive a single absorbed dose of radiation $d$ is given by

$$S = e^{-\alpha d - \beta d^2}, \tag{5.28}$$

where α and β are parameters that depend on the type of cell irradiated. It is remarkable that there is so much discussion of Equation 5.28 in the literature, yet so few references actually quote *any* values for α or β. Common values for tumor cells are $\alpha = 0.30\ \mathrm{Gy}^{-1}$ and $\beta = 0.03\ \mathrm{Gy}^{-2}$. For an OAR, nominal values are $\alpha = 0.15\ \mathrm{Gy}^{-1}$ and $\beta = 0.05\ \mathrm{Gy}^{-2}$. Another parameter that is frequently cited to describe cell survival is the fraction of cells that survive a dose of 2 Gy: $SF_2$. It is known that values of α and β vary from one individual patient to another. Values of α and β from *in vitro* cell culture studies may not be relevant for *in vivo* cell survival (Yorke, 2003).

Determination of α *in vivo* is difficult because tumor or normal tissue is a mixture of cells with varying sensitivity. From *in vitro* measurements and for low linear energy transfer (LET) radiation, α ranges from 0.1 to 2 $\mathrm{Gy}^{-1}$ for a large variety of cell types (Yorke, 2003). *In vivo* determination of α/β is easier because it is based on dose fractionation studies. Deschavanne and Fertil (1996) list *in vitro* values of $SF_2$ for almost 700 tumors and normal cells. α/β is often assumed to be 10 for human tumors; prostate is an exception, with α/β approximately equal to 1.5–3.0 Gy. Late-responding normal tissues have α/β approximately equal to 1–4 Gy, such as the spinal cord, kidney, and liver.

Therapeutic radiation is almost always given in installments or "fractions." The proportion of cells that will survive $f$ installments of radiation, each of dose $d$, is

$$S_f = e^{-D(\alpha+\beta d)}, \tag{5.29}$$

where the total dose is $D$ and $d = D/f$ is the dose per fraction, which is usually 2 Gy. Equation 5.29 is presumed to give the average fraction of surviving cells in a colony of

a large number of cells. Equation 5.29 assumes that each fraction is independent and that cell sensitivity remains constant throughout the treatment.

Normal tissue complication data are usually based on (or corrected to) $d = 2$ Gy per fraction. Physicians generally make an attempt to spare OAR, and therefore these structures rarely receive a full dose of 2 Gy per fraction. The dose absorbed by OAR is also frequently nonuniform; most of the volume receives far less than the standard 1.8–2.0 Gy. Fractionation effects should therefore be important for highly inhomogeneous dose distributions. The original LKB model does not appear to take proper account of fractionation effects for inhomogeneous dose distributions. Mechanistic models can account for this effect provided that the proper value of $\alpha/\beta$ is built in (e.g., see Equation 5.50).

#### 5.4.2.2 TISSUE ARCHITECTURE

Tissues and organs can be thought of as consisting of functional subunits (FSUs). These are elementary components of the organ. Examples of FSUs that have been *suggested* are nephrons for the kidney, alveoli for the lungs, and groups of oligodendrocytes for the spinal cord. Oligodendrocytes are neuroglial cells that send out processes that spiral around axons, thus forming a myelin covering (Spence and Mason, 1992). Some organs and tissues do not have such a structure. An example is skin. In this case, an FSU is a single cell. We will assume that an OAR is composed of FSUs that carry out the function of the organ.

Let us assume that each FSU consists of $k$ cells. We would like to calculate the probability of inactivating *all* the cells (or perhaps all the clonogenic cells) in an FSU.[2] When this occurs, the FSU is assumed to be destroyed. If one clonogenic cell survives, then the FSU can repopulate. The average number of cells in an FSU surviving a dose $D$ is $kS_f$, where $S_f$ is given by Equation 5.29. The probability of no cells surviving is

$$P = \left(1 - S_f\right)^k. \tag{5.30}$$

In the limit in which $k \to \infty$ and $S_f \to 0$ (but $kS_f$ remains finite), the probability is given approximately by the Poisson distribution (see Equation 5.6):

$$P_i \approx f(0) = e^{-\mu} = e^{-kS_f}. \tag{5.31}$$

For an OAR, this may not always be valid, particularly when complication rates are low.

The response of an OAR to irradiation depends on the organization of the FSUs. In a *serial* structure, a complication occurs if one or more FSUs is inactivated. In such a structure, each FSU is critical to the function of the organ, and therefore such structures are sometimes called *critical element* structures. A serial structure is like a chain: if any link fails, then the entire chain will break. Another analogy is to a string of

light bulbs wired in series. If any one of the light bulbs fails, the entire string will fail. Tissues with a small volume index $n$ are thought to be serial in nature. The prime example of a serial structure is the spinal cord. One difficulty with this conception is that the nature of an FSU is not always clear. For example, it is often assumed that the spinal cord is a serial tissue, but what is the nature of an FSU for spinal cord?

Another type of tissue architecture is a parallel structure. In this case, each FSU operates independently. An example for the kidneys is a nephron. Inactivation of some FSUs may make the organ less efficient, but it will still be capable of performing its function. A critical number of FSUs need to be destroyed before the organ fails. A parallel tissue is more like a rope (Goitein, 2008). A number of strands can be destroyed without the rope failing. If enough strands are destroyed, however, the rope will fail. Another analogy is to a string of light bulbs wired in parallel. If one light bulb goes out, the rest remain on. Parallel tissue architecture is sometimes referred to as a *critical volume* structure. Tissues with large values of $n$ (~1) are thought to be parallel in nature.

Some organs and tissues probably do not fall into either a parallel or serial category. The Kallman et al. (1992) relative seriality model describes an OAR as a mixture of serial and parallel FSUs.

#### 5.4.2.3 SERIAL ORGANS: MECHANISTIC MODELS FOR UNIFORM IRRADIATION

The NTCP for a serial organ is the probability that one or more FSUs are inactivated. Let us suppose that $N_0$ is the total number of FSU in the OAR. We assume that the FSUs are small enough that the dose is constant throughout an FSU (i.e., either $k$ is relatively small or the dose is uniform throughout the OAR). Let $P_i$ = the probability of destroying an $\text{FSU}_i$ (see Equation 5.30), which is a function of the dose received by that FSU.

The probability that an FSU is *not* destroyed is $1 - P_i$. The probability that *none* of the FSUs is destroyed is

$$\prod_{i=1}^{i=N_0}(1-P_i),$$

assuming independent probabilities. The NTCP is the probability that one or more are destroyed:

$$\text{NTCP} = 1 - \prod_{i=1}^{N_0}(1-P_i). \tag{5.32}$$

Now suppose that a subvolume $N$ of these receive a spatially constant dose and the remainder receives zero dose. The partial volume $v = N/N_0$. If the portion of the OAR that is irradiated receives the same dose throughout (assuming all cells are equally sensitive), then $P_i = P$ and

$$\text{NTCP} = 1-(1-P)^{N} = 1-\left(1-\left(1-S_f\right)^{k}\right)^{\nu N_0}. \tag{5.33}$$

In this case, the NTCP is given by

$$\text{NTCP} = 1-\left[1-\left(1-e^{-D(\alpha+\beta d)}\right)^{k}\right]^{\nu N_0}. \tag{5.34}$$

Issues of fractionation reduce to the simple biologically equivalent dose (BED) calculations of the linear quadratic model. If NTCP is the same for different fractionations ($k$ and $N$ remain fixed), then the cell survival term $S_f$ in Equation 5.33 must be equal in the two different fractionation schemes.

The two extreme limiting situations are $k = 1$ and $N_0 = 1$. In the case where $k = 1$, each cell is an FSU, and therefore only one cell must be destroyed for a complication to occur. When $N_0 = 1$, the whole organ is the FSU, and all the cells have to be killed for a complication to occur.

One of the problems with the model represented by Equation 5.34 is that we have no idea of the values of $N_0$ and $k$. The nature of an FSU for a serial organ is uncertain (if indeed it has any meaning). The product $kN_0$, however, represents the total number of (clonogenic?) cells in the organ. An upper limit to this number is the total number of cells in the organ. A constraint can therefore be placed on the model:

$$N_0 k = N_c, \tag{5.35}$$

where $N_c$ is the total number of cells of the relevant type in the structure. If this model is to be taken seriously, then the number of cells as given by Equation 5.35 should be of a realistic order of magnitude.

The dose that leads to a given complication probability is found by solving Equation 5.34 for the dose. The result is

$$D = -\frac{1}{\alpha+\beta d}\ln\left[1-\left(1-(1-\text{NTCP})^{1/(N_0\nu)}\right)^{1/k}\right], \tag{5.36}$$

and in particular for NTCP = 50%,

$$D_{50}(\nu) = -\frac{1}{\alpha+\beta d}\ln\left[1-\left(1-2^{-1/(N_0\nu)}\right)^{1/k}\right], \tag{5.37}$$

assuming the dose per fraction is fixed. In the limit as $N_0 \to \infty$,

$$\mathrm{NTCP} = 1 - 2^{-(D/D_{50})^k}. \tag{5.38}$$

The value of $\gamma_{50}$ is (see Equation 5.1)

$$\gamma_{50} = \frac{1}{2} kN\left(2^{1/N} - 1\right)\left[1 - \left(1 - 2^{-1/N}\right)^{-1/k}\right]\ln\left[1 - \left(1 - 2^{-1/N}\right)^{1/k}\right]. \tag{5.39}$$

The dose per fraction $d$ has been held constant in evaluating the derivative for the calculation of $\gamma_{50}$.[3] Note that $\gamma_{50}$ is independent of $\alpha$ and $\beta$ under these circumstances. When $N \gg 1$, $\gamma_{50} \approx \frac{1}{2} k \ln 2$, and thus $\gamma_{50}$ is independent of $N$ for large values of $N$. Equations 5.34, 5.35, 5.37, and 5.39 are fairly general, and it appears as if $N_0$, $k$, $\alpha$, and $\beta$ are uniquely determined once $\gamma_{50}$, $D_{50}$, $N_c$, and $\alpha/\beta$ are specified.

There are two ways to approach the problem of modeling serial OARs. The first is to make reasonable estimates (guesses) of the values of the parameters and substitute these into the equations for the model and see what comes out. The second, and more difficult, approach is to use observed complication rates to find model parameters that reproduce these values. We begin with the first approach, and then later in this section, we will follow the second approach. We will assume that $\nu = 1.0$ or, equivalently, $N = N_0$, throughout the remainder of this section.

Figure 5.9 shows graphs of NTCP for a serial organ in which $N_0 k = 10^8$ cells for $N_0 = 1$, 10, 100, and 1000. We have assumed $d = 2$ Gy, $\alpha = 0.15$ Gy$^{-1}$, and $\beta = 0.05$ Gy$^{-2}$. These curves have the characteristic sigmoid shape. The more FSUs in the OAR, the lower the tolerance dose. When $N_0 = 1$, there is only one FSU, and all $10^8$ cells must be inactivated for a complication to occur. As the number of FSU grows, the NTCP curve becomes steeper. For large $N_0$, and for these values of $\alpha$ and $\beta$, the NTCP is almost

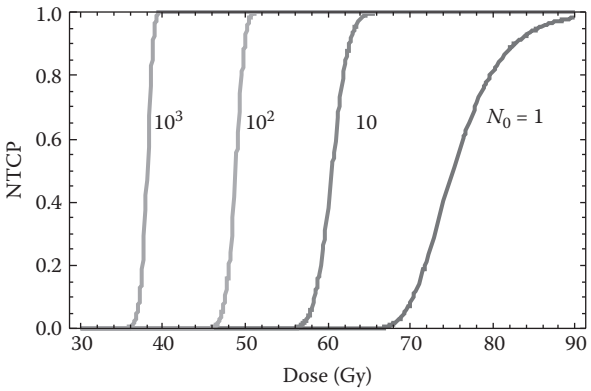



FIGURE 5.9 NTCP for $kN_0 = 10^8$ cells, $d = 2$ Gy, $\alpha = 0.15$ Gy$^{-1}$, and $\beta = 0.05$ Gy$^{-2}$, where $N_0$ is the number of FSU in the OAR and $k$ is the number of cells per FSU. From the left, we have $N_0 = 1000$, 100, 10, and 1. For large $N_0$ and for these values of $\alpha$ and $\beta$, the NTCP curve is almost a step function.

a step function. NTCP becomes almost binary—it is either 0% or above a threshold 100%. As we will discuss shortly, this is not what is actually observed.

Figure 5.10 shows NTCP for various values of α (α/β held fixed at 3.0). As expected, $D_{50}$ declines as α goes up, but $\gamma_{50}$ appears to be insensitive to variations in α as expected from Equation 5.39.

Figure 5.11 provides information about allowable values of $N_0$ and $k$. This figure is a contour plot of $\gamma_{50}$ in a log $N_0$–log $k$ plane. Known values of $\gamma_{50}$ are $2 < \gamma_{50} < 5$. Values

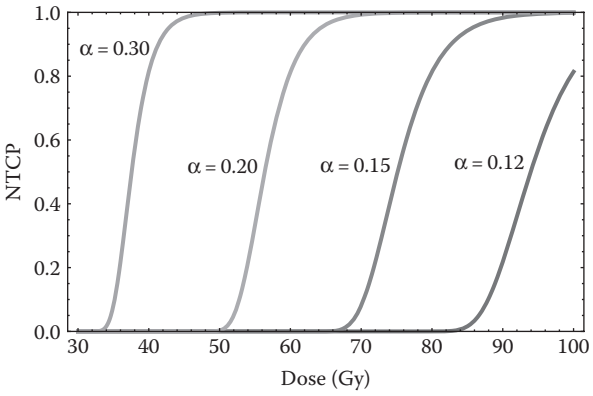



FIGURE 5.10 NTCP for $k = 10^8$, $N_0 = 1$, and various values of α (keeping $\alpha/\beta = 3$). From the left, the values of α are 0.30, 0.20, 0.15, and 0.12. This shows that NTCP and $D_{50}$ are quite sensitive to the value of α, but $\gamma_{50}$ is not.

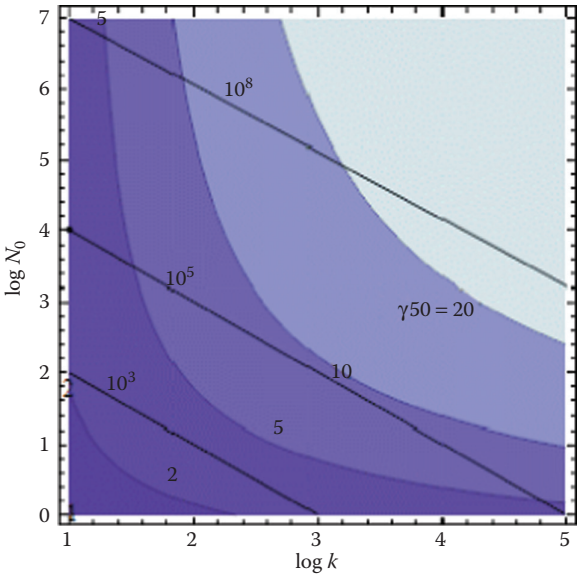



FIGURE 5.11 A contour plot of $\gamma_{50}$ as a function of log $k$ and log $N_0$. Observed values of $\gamma_{50}$ are less than 10. Values of $N_0$ and $k$ are therefore restricted to either small values of $k$ and large values of $N_0$ or vice versa. The diagonal lines are of constants $N_0k = 10^8$, $10^5$, and $10^3$.

of $N_0$ and $k$ are therefore restricted to either small values of $k$ and large values of $N_0$ or vice versa.

**5.4.2.3.1 Cervical Spinal Cord Models** The quintessential serial organ is the spinal cord. According to Hall and Giaccia (2012), a 5% complication rate occurs for $D_5 = 50$ Gy for 10 cm length irradiated. At 70 Gy, the complication incidence is about 50%, depending critically on the dose per fraction. It also depends sensitively on the length irradiated for small lengths, but if the length exceeds a few centimeters, the irradiated volume has little effect. NTCP data from Kirkpatrick et al. (2010) and Schultheiss (2008) for cervical cord myelopathy are contained in Table 5.2. These data are corrected for overall survival as a function of time. These authors suggest that $\alpha/\beta = 0.87$, considerably less than the standard value of 3.0 generally used for normal tissues. Schultheiss (2008) fit the data in Table 5.2 to a logistic model (see Equation 5.26), obtaining values of $D_{50} = 69.4$ Gy and $\gamma_{50} = 4.7$. We note that the "pathogenesis of injury is thought to be primarily from vascular or glial cell injury" (Kirkpatrick et al., 2010, p. S43). It has been suggested that the rectum may also have serial architecture based on the small value of the dose volume index $n = 0.09$.

The mechanistic model of Equations 5.34 and 5.35 requires a value for the number of relevant cells in the OAR. We can make a crude order of magnitude estimate of the number of cells in the cervical spinal cord. A typical mammalian cell is about 10 microns in diameter. Let us assume overhead for packing fraction and other nonrelevant cell types of a factor of 10; that is, the relevant clonogenic cells occupy 1/10 of the spinal cord volume. If we also assume that the cord is 2 cm in diameter and 10 cm long, then the number of cells expected is on the order of $kN_0 \sim 10^8$.

We have made approximate fits to the data in Table 5.2 for the mechanistic model of Equations 5.34 and 5.35. We have fit log (1 – NTCP) using the function NonlinearModelFit available with the software Mathematica. We have used $\alpha/\beta = 0.87$ as suggested by Schultheiss (2008). In our fits to Equation 5.34, we used the constraint that the total number of cells is $N_c$, where $N_c$ takes on values of $10^7$, $10^8$, and $10^9$. We caution that there are essentially only three or four meaningful data points in Table 5.2, as the NTCP at 45, 49.8, and 56.6 Gy are listed as zero. It is presumed that the true values are small but not actually zero (unless one postulates a threshold). The NTCP values for 56.6 and 49.8 Gy were based on only 19 and 22 patients, respectively. For this reason, *any model derived strictly from the data in Table 5.2 may not be accurate at low complication probabilities*, which is just where we would like to know these values.

**TABLE 5.2 Cervical Cord Myelopathy NTCP Data**

| Dose (Gy)[a] | 45.0 | 49.8 | 56.6 | 60.0 | 68.6 | 72.8 |
|---|---|---|---|---|---|---|
| NTCP | 0.00 | 0.00 | 0.00 | 0.10 | 0.44 | 0.62 |

*Source:* Data from Kirkpatrick, J., et al. *Int. J. Radiat. Oncol. Biol. Phys.* 76(3, Suppl.), S42–49, 2010.

[a] Equivalent dose assuming 2.0 Gy per fraction and $\alpha/\beta = 0.87$.

**TABLE 5.3 Mechanistic Cervical Cord Models**[a]

| $N_c = N_0 k$ | $\beta$ (Gy$^{-2}$) | $N_0$ | $k$ | $D_{50}$ (Gy) | $\gamma_{50}$ |
|---|---|---|---|---|---|
| $10^7$ | $2.96 \times 10^{-3}$ | $5.83 \times 10^5$ | 17.1 | 70.4 | 4.3 |
| $10^8$ | $1.68 \times 10^{-3}$ | $7.67 \times 10^6$ | 13.0 | 70.2 | 3.8 |
| $10^9$ | $1.32 \times 10^{-3}$ | $7.82 \times 10^7$ | 12.8 | 70.6 | 3.9 |

[a] All models assume $\alpha/\beta = 0.87$.

We have not used the points at 45 and 49.8 Gy in our fits, as this may artificially suppress NTCP values at low doses. For the data point at dose 56.6 Gy, we assume that the actual NTCP $\lesssim 5\%$ (not zero, as listed in the table) as this point is based on only 19 patients.

The fitting data are listed in Table 5.3. The models of Table 5.3 along with the logistic fit from Schultheiss (2008) and the data from Table 5.2 are plotted in Figure 5.12.

We note that the value of $k$ in Table 5.3 remains roughly constant at about 15 (see also Figure 5.11), even though the value of $N_0$ spans two orders of magnitude. This is because $k$ largely determines $\gamma_{50}$, independent of $N_0$ for large $N_0$ (see comments following Equation 5.39). As we will see, $k$ is closely related to the volume index $n$. The biological meaning (if any) of these $k$ values of approximately 15 is not clear to this author.

The values of $\beta$ in Table 5.3 are smaller than the nominal *in vitro* value of 0.05 Gy$^{-2}$ by more than one order of magnitude. We argue that these values are not well known

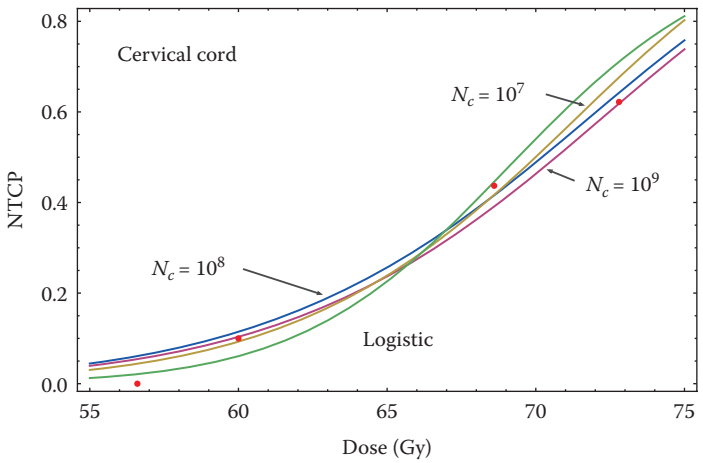



FIGURE 5.12 NTCP curves for the cervical spinal cord, based on fits to observed complication rates (dots) from Kirkpatrick et al. (2010). Fits have been made to a logistic model and to various mechanistic models having a total number of cells of $10^7$, $10^8$, and $10^9$. The model parameters for the mechanistic models appear in Table 5.3. The Schultheiss (2008) logistic model fit has parameters $D_{50} = 69.4$ Gy and $\gamma_{50} = 4.7$. The mechanistic models predict a higher NTCP at low doses than the logistic model because the mechanistic fits have not assumed that the probability is zero at a dose of 56.6 Gy.

and that *in vitro* values may not be applicable here. It seems plausible that $N_c$ may vary between patients, although not by two orders of magnitude, as in Table 5.3.

In clinical application, we are primarily concerned with small values of NTCP as clinicians make a determined effort to minimize complications, particularly for the spinal cord. Clinical applications therefore require NTCP values at small values of NTCP. Radiation oncologists are careful to keep OAR doses well below $D_{50}$. Notice, however, that there is a substantial difference between the predictions of the Schultheiss (2008) logistic and our mechanistic models at small NTCP. The mechanistic models have higher NTCP values at small doses. This is simply a reflection of the fact that we have not assumed that the complication probability is zero at low doses. At 45 Gy, the Schultheiss logistic model predicts NTCP = 0.03%, whereas the mechanistic models of Table 5.3 predict between 0.2% and 0.5%, about an order of magnitude larger. At 50 Gy, the traditional cervical cord dose limit, the logistic model predicts 0.2% and the mechanistic models predict 0.9%–1.5%.

#### 5.4.2.4 SERIAL ORGANS: PARTIAL VOLUME EFFECTS

If $D_{50}(v) = D_{50}(1)\, v^{-n}$, then

$$n = -\frac{\partial \ln D_{50}}{\partial \ln v}. \tag{5.40}$$

Evaluation of the derivative of $D_{50}(v)$, given by Equation 5.37, shows that $n$ is not completely independent of v, and therefore $D_{50}(v)$ cannot be an exact power law. As we will show below, however, under some circumstances, $D_{50}(v)$ follows a power law very closely.

The ratio $D_{50}(v)/D_{50}(1)$ is independent of α, β, and $d$, as can be seen from Equation 5.37, and depends only on $N_0$ and $k$. For $N_0 \gg 1$, $1 - 2^{-1/N_0} \approx (\ln 2)/N_0$, and if $(\ln 2/N_0)^{1/k} \ll 1$, then from Equation 5.37, $D_{50}(v)/D_{50}(1) \approx v^{-1/k}$, which in turn implies that $n \approx 1/k$. The last approximation is of marginal validity; for $N_0 = 5.8 \times 10^5$ and $k = 17.1$, $(\ln 2/N_0)^{1/k} \approx 0.45$. In Section 5.4.2.3, we saw that when $N_0 \gg 1$, $\gamma_{50} \approx \frac{1}{2} k \ln 2$. The serial model therefore predicts a simple relationship between $n$ and $\gamma_{50}$, namely, that $n \approx \ln 2/(2\ \gamma_{50})$, when $N_0 \gg 1$.

Figure 5.13 shows a log-log plot of $D_{50}(v)$ versus v for the models of Table 5.3. For comparison, the relation $D_{50}(v) = 69.4\ v^{-0.05}$ is also plotted, which uses the accepted value of $n$ for the cervical cord. A power law relationship has a linear graph in a log-log plot. The mechanistic models of Table 5.3 are seen to exhibit power law behavior for $D_{50}(v)$ to a good approximation. The slope of the mechanistic graphs is approximately $n = 0.09$, whereas the "accepted" value for spinal cord is $n = 0.05$. We note that the value of $1/k = 0.08$ ($k = 13$) is close to the value for the slope of the mechanistic graphs.

If the probability of complication is known to be $NTCP_1$ for the whole organ irradiated to a dose $D$, then the NTCP when a fractional volume v is irradiated to the same dose (and the remainder receives zero dose) is (see Equation 5.34)

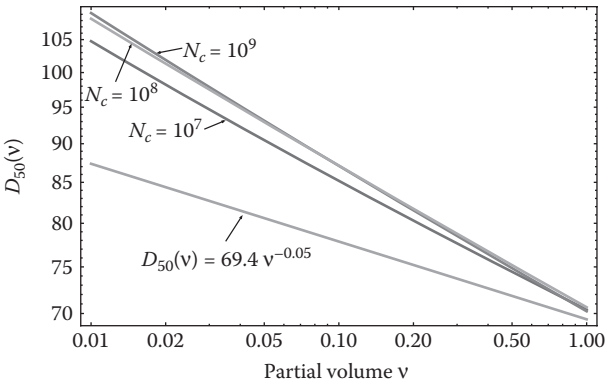



FIGURE 5.13 A log-log plot of $D_{50}(\nu)$ vs. the partial volume ν for the mechanistic models of Table 5.3. The relation $D_{50}(\nu) = \nu^{-0\,05}$ is also plotted, where the volume index $n = 0.05$ is the accepted value for cervical spinal cord. A power law relation will appear as a straight line on a log-log graph. The mechanistic model graphs are seen to be nearly straight lines, showing that $D_{50}$ does follow a power law for these models. The power law index for these models is approximately $n = 0.09$.

$$\text{NTCP}_\nu = 1 - (1 - \text{NTCP}_1)^\nu \tag{5.41}$$

When $\text{NTCP}_1$ is small, $\text{NTCP}_\nu$ becomes (to first order in ν) = $\nu\text{NTCP}_1$; that is, $\text{NTCP}_\nu$ is proportional to the volume irradiated.

Figure 5.14 shows the dose that elicits various values of NTCP as a function of the partial volume irradiated for the first cervical spinal cord model listed in Table 5.3. The graphs in this figure show that values of NTCP do not vary much with volume

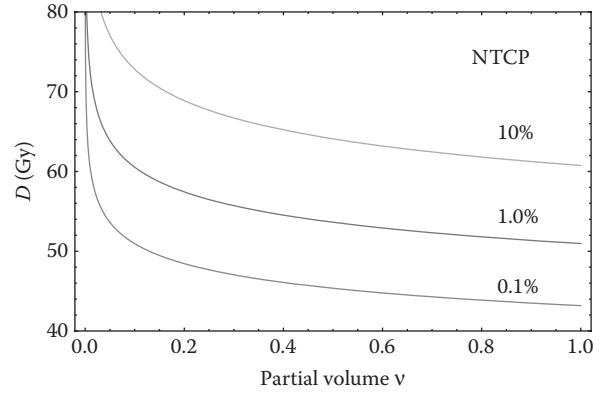



FIGURE 5.14 The dose required for a given value of NTCP as a function of partial volume ν irradiated. The model parameters are $N_0 = 5.83 \times 10^5$, $k = 17.1$, $\alpha = 2.31 \times 10^{-3}$, and $\beta = 2.96 \times 10^{-3}$. The NTCP values of the three curves are 0.1%, 1%, and 10%. This clearly shows that there is little volume dependence for this model except for very small partial volumes of perhaps 10% or less. When ν < 10%, volume effects become significant. This is consistent with clinical data.

except when the partial volume is less than about 10%. This is consistent with the statement by Hall and Giaccia (2012) for spinal cord that NTCP is very sensitive to the length for lengths less than a few centimeters, but that NTCP is independent of length above this.

The mechanistic models of this section fit clinical data for cervical spinal cord reasonably accurately and predict power law behavior for $D_{50}(\nu)$ with a roughly correct value of $n$.[4] These models are consistent with an order of magnitude limiting value of the number of cells in the cervical cord. In addition, the partial volume graphs of Figure 5.14 show that there is little volume effect for cervical spinal cord until $\nu \sim 0.10$, in agreement with experimental data. One criticism is that the values of $\alpha$ and $\beta$ are much smaller than expected. In Section 5.4.2.5, we will generalize this model to inhomogeneous irradiation.

#### 5.4.2.5 SERIAL ORGANS: INHOMOGENEOUS IRRADIATION

We will assume that the dose received by all of the $k$ cells in an arbitrary FSU is the same. This should be reasonable provided that $k/N_0 \ll 1$. Let the number of FSU receiving dose $D_i$ be $\nu_i N_0$, then the generalization of Equation 5.32 is

$$\mathrm{NTCP} = 1 - \prod_{i=1}^{M} \left(1 - P_i\right)^{\nu_i N_0}, \tag{5.42}$$

where $M$ is the number of different discrete dose levels. The probability $P_i = (1 - S)^k$ and $S$ is now given by

$$S = e^{-D_i(\alpha + \beta D_i/f)}, \tag{5.43}$$

where $f$ is the number of fractions. For an inhomogeneous irradiation, the dose per fraction will vary throughout the OAR, and therefore we express $S$ in terms of the number of fractions. Equation 5.42 can be written as

$$\ln\left(1 - \mathrm{NTCP}\right) = N_0 \sum_{i=1}^{M} \nu_i \ln\left(1 - P_i\right). \tag{5.44}$$

We can turn the discrete description in Equation 5.44 into a continuous description by noting that the fractional volume having dose between $D$ and $D + dD$ is $\{h(D)/V\}dD$; this corresponds to $\nu_i$ in Equation 5.44. We may therefore rewrite this equation as

$$\ln\left(1 - \mathrm{NTCP}\right) = \frac{N_0}{V} \int_0^{D_{\max}} h(D) \ln\left[1 - \left(1 - S\right)^k\right] dD. \tag{5.45}$$

In terms of the cDVH, this can be written (using integration by parts)

$$\ln(1-\mathrm{NTCP}) = -kN_0 \int_0^{D_{\max}} (\alpha + 2\beta D/f) H(D) \frac{S(1-S)^{k-1}}{1-(1-S)^k} dD. \tag{5.46}$$

The uniform dose that is equivalent is independent of $N_0$. The uniform dose equivalent for inhomogeneous irradiation can be found by setting the NTCP for uniform irradiation (Equation 5.34) equal to the NTCP for inhomohogeneous irradiation. Alternatively, we may set ln(1 – NTCP) equal in the two cases. First, we will make some approximations. For an OAR, $D_i(\alpha + \beta D_i/f) \ll 1$ provided that

$$D_i \ll \frac{1}{2} f \left(\frac{\alpha}{\beta}\right) \left[\sqrt{1 + \frac{4\beta}{f\alpha^2}} - 1\right]. \tag{5.47}$$

For the model parameters of Table 5.3 ($N_c = 10^7$), this will be true for doses much less than 82 Gy, and therefore the condition of Equation 5.47 seems reasonable, at least in this case. In addition to this approximation, let us examine the limit in which $N_0 \gg 1$ and $(\ln 2/N_0)^{1/k} \ll 1$. In this case, we have seen that $k \approx 1/n$. Under these circumstances for uniform irradiation (from Equation 5.34),

$$\ln(1-\mathrm{NTCP}) \approx -N_0 \left[D(\alpha + \beta D/f)\right]^{1/n}. \tag{5.48}$$

For inhomogeneous irradiation (from Equation 5.44),

$$\ln(1-\mathrm{NTCP_v}) \approx -N_0 \sum_{i=1}^{M} \mathrm{v}_i D_i (\alpha + \beta D_i/f)^{1/n}. \tag{5.49}$$

The sum in Equation 5.49 may be replaced by $1/N$ times the sum over the total number of voxels $N$, assuming the dose is constant in each voxel. If we now set the expressions for ln(1 – NTCP) in Equations 5.48 and 5.49 equal, we find

$$D_e \left(\frac{\alpha}{\beta} + \frac{D_e}{f}\right) = \left\{\frac{1}{N} \sum_{i=1}^{N} \left[D_i \left(\frac{\alpha}{\beta} + \frac{D_i}{f}\right)\right]^{1/n}\right\}^n, \tag{5.50}$$

where $D_e$ is the uniform equivalent dose. Examination of Equation 5.18 for the EUD shows that Equation 5.50 is a generalization of the generalized EUD. Equation 5.50 takes account of variable fractionation experienced by different voxels in the OAR,

which the EUD of Equation 5.18 does not. For small doses in which $D \ll f\,(\alpha/\beta)$, the uniform equivalent dose of Equation 5.50 reduces to the EUD.

An expression can be easily derived for the NTCP of an OAR irradiated to a uniform dose $D_0$ over a fraction $1 - \nu_h$, with a uniform hot spot of fractional volume $\nu_h$, which receives a dose $D_h$ (refer to Equation 5.44):

$$\mathrm{NTCP}_h = 1 - \left(1 - \mathrm{NTCP}_0\right)^{1-\nu_h} \left(1 - \left(1 - S_h\right)^k\right)^{\nu_h N_0}, \tag{5.51}$$

where $\mathrm{NTCP}_0$ is the probability for uniform irradiation of the entire volume with dose $D_0$ and

$$S_h = e^{-D_h(\alpha+\beta D_h/f)}.$$

We have examined the effect of a small hot spot for a cervical spinal cord model ($N_c = 10^7$) (see Table 5.3) using Equation 5.51. The model parameters are $N_0 = 5.83 \times 10^5$, $k = 17.1$, $\beta = 2.96 \times 10^{-3}$, $\alpha/\beta = 0.87$, and $f = 25$ fractions. For this model, $\mathrm{NTCP}_0 = 1\%$ corresponds to a dose of 50.6 Gy. Assume the cord receives a uniform dose of 50.6 Gy except for a small hot spot (1%, 2%, and 5% in fractional volume) that receives a variable dose. The results of this analysis appear in Figure 5.15. The graph shows that a 1% (in volume) hot spot must receive 62 Gy to double the NTCP to 2%.

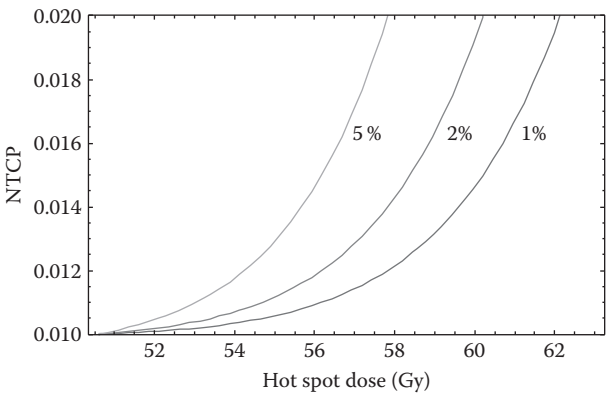



FIGURE 5.15 The effect of a hot spot for a serial organ. This shows the NTCP vs. the maximum dose for a cervical spinal cord model in which $N_0 = 5.83 \times 10^5$, $k = 17.1$, $\beta = 2.96 \times 10^{-3}$, and $\alpha/\beta = 0.87$. It is assumed that the dose is delivered in 25 fractions. The cord receives a uniform dose of 50.6 Gy, except for a small percentage that receives a maximum uniform dose as plotted along the horizontal axis. The three curves show 1%, 2%, and 5% (volume) hot spots, respectively. A 1% (in volume) hot spot would require a dose of more than 62 Gy to double the NTCP from 1% to 2%.

### 5.4.2.6 PARALLEL ARCHITECTURE: UNIFORM IRRADIATION

Among those organs that are considered to have parallel architecture are kidneys, lungs, and liver. For kidneys, an FSU is thought to be a nephron, and for lungs, it could be an alveolus (although Hall and Giaccia [2012] state that it is an entire pulmonary lobule). For liver, the nature of the FSU is not clear. According to Hall and Giaccia (2012), for the kidneys there are about 1000 stem cells in an FSU. An average human kidney contains between $9 \times 10^5$ and $1.0 \times 10^6$ nephrons (Bertram et al., 2011), although specific individuals can have as low as $2 \times 10^5$ nephrons per kidney. In six adult human lungs, the mean number of alveoli is $4.8 \times 10^8$. The mean volume of a single alveolus has been found to be rather constant at $4.2 \times 10^6$ μm$^3$ (Ochs et al., 2004). An alveolus is approximately a hollow sphere with a thin wall to permit gaseous diffusion. From these data, we can estimate the minimum number of cells in a alveolus. If a typical human cell has a diameter of 10 μm and we assume that cells tile the wall of the alveolus, then each alveolus should contain at least 1600 cells. This assumes that the wall thickness is one cell thick.

Let us suppose that a parallel organ contains $N_0$ FSU. Let $P$ be the probability of eradicating an FSU by delivery of dose $D$ (see Equation 5.30). Let us suppose that there is a threshold number of FSU, $L$, that must be eradicated before a complication occurs. This is like a rope that fails when $L$ strands out of $N_0$ break. Note that this implies a *volume threshold*; that is, a complication *cannot* occur unless a partial volume $v > L/N_0$ is irradiated regardless of the dose delivered. Some organs that are thought to have parallel architecture come in pairs (e.g., kidneys and lungs). It is commonly known that an individual can survive the loss of one kidney. Therefore, it seems possible that $L/N_0$ might be greater than ½ for such an organ pair.[5]

The probability of destroying exactly $j$ FSU out of $N_0$ irradiated is given by the binomial distribution (Equation 5.2):

$$ {}_{N_0}C_j P^j (1-P)^{N_0-j}. $$

A complication occurs only if more than $L$ FSU are eradicated, and thus

$$ \mathrm{NTCP} = 1 - \sum_{j=0}^{j=L-1} {}_{N_0}C_j P^j (1-P)^{N_0-j}. \tag{5.52} $$

When $L = 1$, Equation 5.52 reduces to Equation 5.33 for the serial model. *Thus, the serial model is a special case of the parallel model.*

As $N_0$ and $L$ are presumed to be large numbers, let us assume that the binomial distribution can be replaced with the normal distribution (Equation 5.7). We will consider the validity of this assumption later. We replace the sum in Equation 5.52 with an integral:

$$\text{NTCP} = 1 - \frac{1}{\sigma\sqrt{2\pi}} \int_0^L e^{-(z-\mu)^2/2\sigma^2} dz, \tag{5.53}$$

where $z$ (like $j$ in Equation 5.52) is a dummy variable, $\mu = N_0 P$, and $\sigma = \sqrt{N_0 P(1-P)}$. Equation 5.53 may be written as

$$\text{NTCP} = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\mu - L}{\sigma\sqrt{2}}\right)\right], \tag{5.54}$$

where $L < N_0$. This assumes that $P \gg 2/N_0$, and therefore erf $(\mu/(\sigma\sqrt{2})) \approx 1$. When $\mu = L$, NTCP = 50% as expected. At $D_{50}$, $L = \mu = N_0 P$, and therefore $P = L/N_0$ at $D_{50}$.

We can relate the mechanistic model here to empirical models by deriving equations for $D_{50}$ and $\gamma_{50}$. We can derive an expression for $D_{50}$ by setting $\mu = L$ in Equation 5.54. The result is

$$D_{50}(1) = -\frac{1}{\alpha + \beta d} \ln\left[1 - \left(\frac{L}{N_0}\right)^{1/k}\right], \tag{5.55}$$

assuming that the dose per fraction $d$ is constant. Notice that $D_{50}$ depends only on the ratio of $L/N_0$ and not on $N_0$ and $L$ separately. $\gamma_{50}$ can be derived by differentiating Equation 5.54:

$$\gamma_{50} = -\frac{k\sqrt{N_0}\left[1 - x^{1/k}\right] x^{1-1/k} \ln\left[1 - x^{1/k}\right]}{\sqrt{2\pi x(1-x)}}, \tag{5.56}$$

where $x = L/N_0$ and we have assumed that $d$ is a constant. Notice that $\gamma_{50}$ does not depend on $\alpha$ and $\beta$.

Let us look at some of the implications of the model represented by Equations 5.54 through 5.56. For kidney, $D_{50} = 28$ Gy (from LKB fit). If we assume $\alpha = 0.15$ Gy$^{-1}$, $\beta = 0.05$ Gy$^{-2}$, $d = 2$ Gy, and $k = 1000$, then $L/N_0 = 0.40$ from Equation 5.55. If we substitute $N_0 = 10^6$, $x = 0.40$, and $k = 1000$ into Equation 5.56, we get $\gamma_{50} = 2090$. The actual observed value of $\gamma_{50}$ for the kidneys is approximately 4. We can repeat this exercise for lung using $D_{50} = 31.4$ Gy, $\alpha/\beta = 4.0$ Gy, $\beta = 0.05$ Gy$^{-2}$, and $k = 5 \times 1600$. Solving for $L/N_0$, we get $L/N_0 = 0.52$ from Equation 5.55.[6] Smaller values of $k$ give unrealistically high values of $L/N_0$. If we also assume that $N_0 = 5 \times 10^8$, Equation 5.56 predicts that $\gamma_{50} = 57{,}150$, whereas the observed value is 0.89. This is a stunningly inaccurate prediction. Adjusting the values of $\alpha$ and $\beta$ does not help because $\gamma_{50}$ does not depend on them (providing that the dose per fraction is constant).

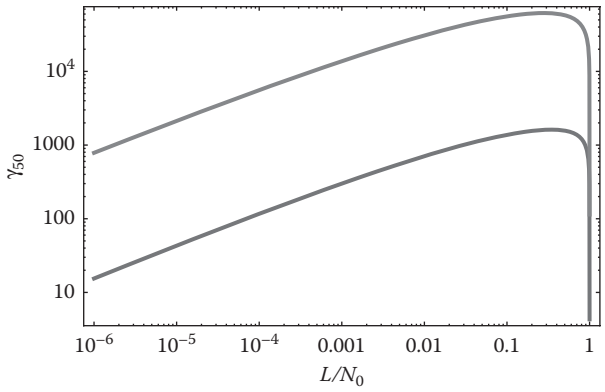



FIGURE 5.16 $\gamma_{50}$ as a function of $x = L/N_0$ for $N_0 = 10^6$, $k = 200$ (bottom curve), and $N_0 = 5 \times 10^8$, $k = 8000$ (top curve). The calculated values of $\gamma_{50}$ are orders of magnitude larger than measured values except for very tiny values of $x$, values for which $L$ approaches 1 or less.

Figure 5.16 shows $\gamma_{50}$ as a function of $L/N_0$ for two sets of ($N_0$, $k$) values. This graph shows that there is no reasonable combination of parameters that can reproduce measured values of $\gamma_{50}$.

Could the problem of large values of $\gamma_{50}$ be due to faulty approximations used in deriving the model? There are two approximations. The first is that $|N_0P - j| \ll N_0P$ is the number of FSUs killed by the radiation. This approximation is used to replace the binomial distribution with the normal distribution in Equation 5.52. This is certainly not true when $j$ is small; however, when $j$ is small, the terms in the sum are small, and it is *plausible* that they do not contribute very much. When $j$ is on the order of $L$, $P$ is roughly $L/N_0$, and in this case, the approximation is valid. The second approximation is implicit in going from Equation 5.53 to Equation 5.54, and that is the assumption that $\text{erf}(\mu/(\sigma\sqrt{2})) \approx 1$. This will be correct when the argument of erf is somewhat greater than 1.0 (see Figure 5.2). This occurs when $P > 2/N_0$, which should be valid for most doses of interest.

The very large values of $\gamma_{50}$ predicted by Equation 5.56 require a modification (or abandonment) of the simple model of Equations 5.54 through 5.56. Intrapatient variability in $\alpha$ and $\beta$ would seem logical in that the cells making up an FSU are likely to be heterogeneous in nature; however, $\gamma_{50}$ does not depend on $\alpha$ or $\beta$. Variations in $N$, $L$, or $k$ within any reasonable bounds do not seem to help either (see Figure 5.16). An alternative scheme that reduces the predicted value of $\gamma_{50}$ is to invoke interpatient variability. The value of $L/N_0$ may vary from patient to patient, depending on health status, age, adjuvant chemotherapy, and so forth. In addition, values of $\alpha$ and $\beta$ may vary between individuals. If Equation 5.56 is correct for individuals, it implies that there is a sharp dose threshold for complications for parallel structures. In this case, any effort to reduce the dose to the OAR for a specific patient may make the difference between having a complication and not having one.

NTCP data are for populations, not individuals. An individual either has a complication or does not, although we acknowledge that the severity of the complication may vary. Equations 5.54 through 5.56 apply to individuals.

It seems very plausible that $D_{50}$ may vary between individual patients, depending on individual health status or morbidity. Examination of Equation 5.55 shows that any variation in radiosensitivity or $L/N_0$ will lead to variability of $D_{50}$. For the kidneys, it is known that the number of nephrons $N_0$ varies dramatically among individuals (Bertram et al., 2011). If one assumes that a certain number of nephrons $L_0$ are necessary for proper function, then $L/N_0$ may vary significantly from patient to patient. Let us suppose that $D_{50}$ varies between individuals following a normal distribution with mean value $\langle D_{50}\rangle$ and standard deviation $\sigma_{50}$. The very large values of $\gamma_{50}$ show that Equation 5.54 for NTCP is essentially a step function. Therefore, Equation 5.54, can be replaced by NTCP = $u(D - D_{50})$, where $u$ is the unit step function: $u = 0$ for $D < D_{50}$ and $u = 1.0$ for $D > D_{50}$. If we average these step functions over the normal distribution,

$$\begin{aligned}\langle \text{NTCP}\rangle &= \frac{1}{\sigma_{50}\sqrt{2\pi}}\int_0^\infty u(D-D_{50})\,e^{-\frac{1}{2}\left(\frac{D_{50}-\langle D_{50}\rangle}{\sigma_{50}}\right)^2} dD_{50} \\ &= \frac{1}{2}\left[\text{erf}\left(\frac{D-\langle D_{50}\rangle}{\sqrt{2}\sigma_{50}}\right)+\text{erf}\left(\frac{\langle D_{50}\rangle}{\sqrt{2}\sigma_{50}}\right)\right],\end{aligned} \tag{5.57}$$

where $\gamma_{50}$ for the population can be found by differentiating Equation 5.57:

$$\langle \gamma_{50}\rangle = \frac{\langle D_{50}\rangle}{\sigma_{50}\sqrt{2\pi}}. \tag{5.58}$$

The value of $\langle D_{50}\rangle/(\sqrt{2}\sigma_{50})$ in Equation 5.58 is expected to be greater than 1.0. Under these circumstances, the second term (on the second line) in Equation 5.57 is expected to have a value of approximately 1.0, and therefore

$$\langle \text{NTCP}\rangle = \frac{1}{2}\left[1+\text{erf}\left(\frac{D-\langle D_{50}\rangle}{\sqrt{2}\sigma_{50}}\right)\right]. \tag{5.59}$$

Examination of Equations 5.58 and 5.24 shows that $\sigma_{50} = m\langle D_{50}\rangle$. Comparing Equation 5.59 with Equations 5.22 and 5.25 shows that our model for population averaged parallel architecture is the same as the Lyman empirical model; that is, *the Lyman model for a population is equivalent to a dose–response curve that is a step function for individuals with values of $D_{50}$ averaged over a normal distribution*. This is illustrated in Figure 5.17.

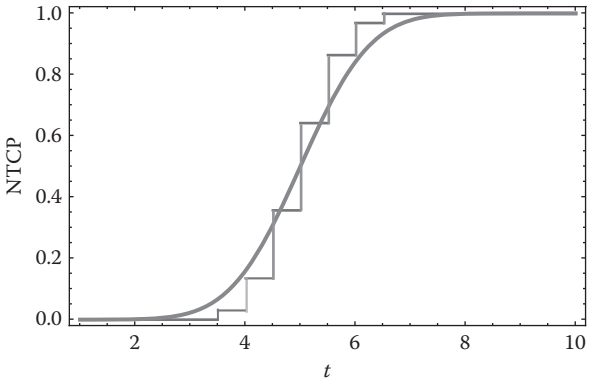



FIGURE 5.17 The construction of a continuous sigmoidal dose–response curve from a series of step functions. The amplitude of each step is modulated by a normal distribution. In this case, seven steps have been summed (stepped graph). The average is at $t = 5$ and the standard deviation is 1.0. For comparison, the smooth curve is for the addition of an infinite number of steps given by Equation 5.59.

If we perform an interpatient average on the parallel model, then Equation 5.55 for $D_{50}(1)$, although valid for an individual, is not valid for the mean value of $D_{50}$. For any function $F(z)$, $\langle F(z)\rangle \neq F(\mu)$, where $\mu$ is the mean value of $z$. If the random variable $z$ is distributed according to the normal distribution, then to the second order,

$$\langle F(z)\rangle \approx F(\mu) + \frac{1}{2}\sigma^2 F''(\mu), \tag{5.60}$$

where $\sigma$ is the standard deviation of $x$ and $F''$ is the second derivative of $F$ with respect to $z$. The ratio of the second term to the first term in Equation 5.60 is on the order of $(\sigma/\mu)^2$. To the extent that this is much less than 1, the second term in Equation 5.60 can be neglected. The quantity $\alpha + \beta d$ and $(L/N_0)$ cannot strictly obey a normal distribution because $\alpha + \beta d > 0$ and $0 < L/N_0 < 1$. For a lognormal distribution, if $F(z) = 1/z$ (see Equation 5.55), where $z = \alpha+\beta d$, then $\langle F(z)\rangle = 1/\mu$. Based on this discussion, we will assume for expedience that

$$\left\langle D_{50}(1)\right\rangle \approx -\frac{1}{\langle \alpha + \beta d\rangle} \ln\left[1 - \left\langle \frac{L}{N_0}\right\rangle^{1/k}\right]. \tag{5.61}$$

We are now in a position to calculate model parameters for some selected OAR (see Table 5.4). The input are the following data: $D_{50}(1)$, $\gamma_{50}$, and various values of: $\alpha + 2\beta$ ($= -1/2 \ln SF_2$). The dose per fraction $d$ is assumed to be 2 Gy throughout. Values of $\alpha + 2\beta$ are chosen from a plausible range of possibilities or from *in vitro* data. We also know that $L/N_0$ must be significantly less than 1.0. For example, $L/N_0$ cannot be 0.98; otherwise, this would be clinically apparent. The quantity $L/N_0$ only

**TABLE 5.4 Parallel Architecture Models for Selected OAR[a]**

| Parallel Organ Data | | | | | | Parallel Model Parameters | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Organ (Endpoint)** | **$D_{50}$ (Gy) (95% CI)** | **$\gamma_{50}$** | **$n$ (95% CI)** | **$\alpha/\beta$ (Gy)** | **Reference** | **$\alpha_{50}/D_{50}$[e]** | **$\langle\alpha + 2\beta\rangle$[d] $(Gy)^{-1}$** | **$k$** | **$L/N_0$[f]** | **$\tilde{n}$[c]** |
| Kidney (RT-induced kidney injury) | 28 | 4.0 | 0.7 | ? | Dawson et al. (2010) | 0.10 | 0.25 | 1000 | 0.40 | 0.23 |
| | | | | | | | 0.25 | 379 | 0.71 | 0.70 |
| | | | | | | | 0.20[b] | 113 | 0.66 | 0.70 |
| Lung (pneumonitis) | 31.4 (29.0–34.7) | 0.89 (0.78–1.02) (95% CI) | 1.03 (0.67–1.39) | 4.0 | Marks et al. (2010a) | 0.45 | 0.30 | 8000 | 0.52 | 0.27 |
| | | | | | | | 0.25 | 1600 | 0.54 | 0.33 |
| | | | | | | | 0.25 | 3200 | 0.29 | 0.15 |
| Liver (radiation-induced liver disease) | 43 (39.8–46.1) | 2.3 (1.3–3.3) Range reported | 0.86–1.1 | 2.0–3.0 | Pan et al. (2010) | 0.17 | 0.25 | 8003 | 0.84 | 0.98 |
| | | | | | | | 0.30 | 58,150 | 0.29 | 0.98 |
| | | | | | | | 0.20 | 1140 | 0.81 | 0.98 |

[a] All parallel organ data on the left-hand side of the table are based on the LKB model.
[b] Based on in *vitro* value from Deschavanne and Fertil (1996).
[c] Average value of volume exponent n; see Equation 5.63.
[d] It is assumed that the dose per fraction is 2 Gy throughout.
[e] Computed from Equation 5.58.
[f] Computed from Equation 5.61.

has reasonable values over a fairly narrow range in values of $\alpha + 2\beta$. The quantity $\sigma_{50}$ is computed from Equation 5.58, and $(L/N_0)^{1/k}$ is computed from Equation 5.61. The value of $L/N_0$ is then computed for various values of $k$.

Table 5.4 shows a variety of models for the kidney, lung, and liver with a range of parameters. Some examples follow. For the lungs, $\langle D_{50}(1)\rangle = 31.4$ Gy (based on LKB), and according to Marks et al. (2010a), $\alpha/\beta = 4.0$. If we assume that $k = 8000$ and $\beta = 0.05$, then $L/N_0 = 0.52$ and $\sigma_{50} = 14$ Gy. For the kidneys, if $\langle D_{50}(1)\rangle = 28$ Gy, $k = 1000$, $\alpha/\beta = 3.0$ Gy, and $\beta = 0.05$ Gy$^{-2}$, then $L/N_0 = 0.40$ and $\sigma_{50} = 2.8$ Gy.

#### 5.4.2.7 PARALLEL ARCHITECTURE: PARTIALLY UNIFORM IRRADIATION

This is the case where a fraction of the organ $v$ receives a uniform dose $D$ (at 2 Gy per fraction) and the remainder receives no dose. To examine this case, we replace $N_0$ in Equation 5.61 by $vN_0$. Due to the threshold, if $v < L/N_0$, there will be no complication regardless of how large the dose. As we have commented above, for some paired organs, such as the kidneys, it is likely that $v > ½$. Figure 5.18 illustrates the "forbidden" region for kidneys in a $v$-$D$ plane. The shaded area has an NTCP above 5%.

We can evaluate the putative power law dependence of $D_{50}$ by plotting $D_{50}(1)/D_{50}(v)$, assuming that the dose per fraction, $d$, is the same in both cases. This quantity is only defined above the threshold: $v > L/N_0$. A log-log plot is shown in Figure 5.19 for various combinations of parameters $(L/N_0)$ and $k$. The graphs are clearly not straight lines, as would be expected of a power law. A charitable view is that these plots are

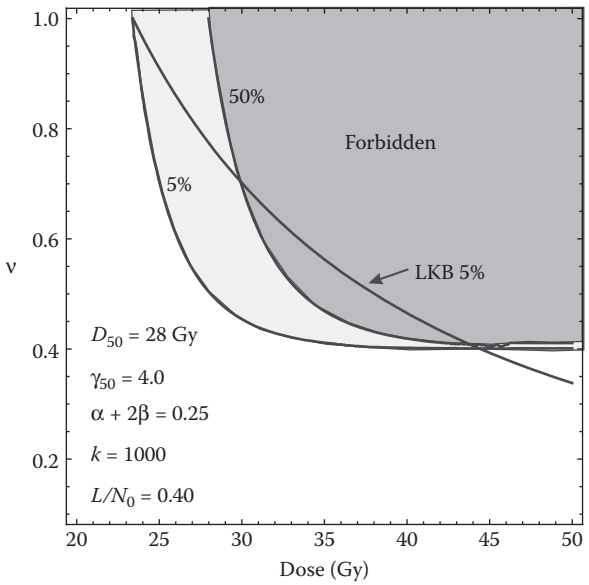



FIGURE 5.18 Shows NTCP contours for a kidney model with NTCP = 5% and 50% in a partial volume–dose plane. The shaded area is to be avoided if one wishes to hold complication probability to less than 5%. For this model, $L/N_0 = 0.40$, and therefore when the partial volume is below this value, no amount of dose will cause a complication. For comparison, the corresponding LKB model with NTCP = 5% is also shown.

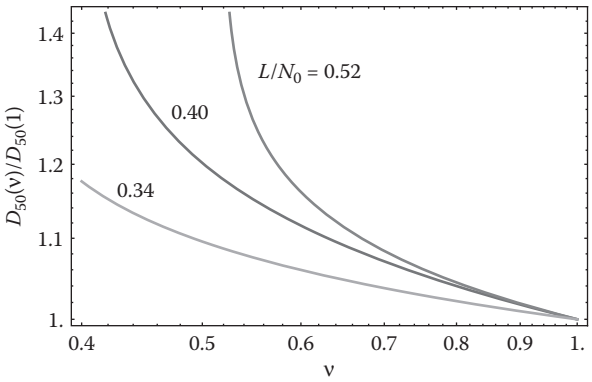



FIGURE 5.19 A log-log plot of $D_{50}(\nu)/D_{50}(1)$ vs. ν for $(L/N_0, k)$ of (0.34, 50,000), (0.40, 1000), and (0.52, 8000). The graphs are clearly not straight lines, as would be expected for a power law. They are roughly linear for $\nu \approx 1.0$.

very roughly linear for values of ν well above the threshold where complications are most likely to be seen.

Equation 5.40 may be used to calculate the value of the power law exponent:

$$n(\nu) = \frac{-(x/\nu)^{1/k}}{k\left(1-(x/\nu)^{1/k}\right)\ln\left(1-(x/\nu)^{1/k}\right)}, \tag{5.62}$$

which of course would be a constant if $D_{50}(\nu)$ were a true power law. This exponent actually becomes singular when $\nu = x = L/N_0$. We can define an average of sorts by evaluating $n$ at a value of ν, which is halfway between $x$ and 1.0 (Figure 5.19):

$$\tilde{n} = n\left(\frac{1+x}{2}\right). \tag{5.63}$$

Some of the values of $k$ in Table 5.4 have been chosen to give the observed values of $\tilde{n}$ by using Equations 5.62 and 5.63 as constraints.

There appears to be no solution to Equations 5.58 and 5.61 for the *in vitro* value of $\alpha + 2\beta = 0.46$ for liver (Deschavanne and Fertil, 1996). If the value of $\alpha + 2\beta$ is much less than the values in Table 5.4, then $k \to 1$, and if $\alpha + 2\beta$ is somewhat larger than the values in Table 5.4, then $L/N_0 \to 1$. These are limiting values.

#### 5.4.2.8 PARALLEL ARCHITECTURE: INHOMOGENEOUS IRRADIATION

This is a difficult problem to analyze because of the challenge of calculating the probability of killing > $L$ FSU. In principle, each FSU could receive a unique dose $D_i$. For each FSU, there will be a different probability of inactivation $p_i = (1 - S_i)^k$. We then face the mathematical problem of determining the probability that $L$ or more FSU will be eradicated. This is analogous to the following problem. There are $N$ coins, each of which has a probability $p_i$ of landing heads up. What is the probability that $L$ or more

of these coins will land heads up? This difficult mathematical problem has been discussed by Jackson et al. (1993).

## 5.5 TUMOR CONTROL PROBABILITY

Yorke (2003) describes some of the difficulties with TCP models. These include a frequent lack of detailed knowledge of the dose and the distribution of dose actually delivered to the patient. Confusion between local failure and nearby recurrence can be a confounding issue. Soft tissue tumors have a cell number density of about $10^9$ cm$^{-3}$, but not all of these cells are clonogenic. The fraction of the cells associated with tumor growth is not known. Reasonable models can be constructed that agree with observed data for uniform dose distributions but differ greatly for nonuniform irradiation.

### 5.5.1 TCP EMPIRICAL MODELS

The EUD (see Equation 5.18) is used by several commercial treatment planning systems for optimization or plan evaluation (Li et al., 2012). It is a simple, single-parameter quantity that is easily computed from the DVH (see Equation 5.19). For a tumor, the value of the volume index $n$ must be negative, and it is small. The value is typically about $n \approx -0.10$ (Yorke et al., 2013).[7] The smaller the value of $|n|$, the closer the value of the EUD is to the minimum dose.

Although the parameters $D_{50}$ and $\gamma_{50}$ are most frequently mentioned in the literature, the parameter $D_{95}$ may be more relevant. Various models can have the same $D_{50}$ and $\gamma_{50}$ but very different values of $D_{95}$. Clinical data indicate that the values of $D_{50}$ lie between 40 and 70 Gy and that the values of $\gamma_{50}$ range between 1 and 3 (for the model of Equation 5.64) (Okunieff et al., 1995).

Table 5.5 contains selected $D_{50}$ and $\gamma_{50}$ data from the paper by Okunieff et al. (1995), along with some modeling data that will be discussed in the next section. This reference reports a mean value of $D_{50}$ of 50 Gy ± 18.4 (1 SD) and a mean $\gamma_{50}$ value of 3.2 ± 7.6 (1 SD). For the mechanistic model considered here, $D_{50}$ and $\gamma_{50}$ depend on the volume of the tumor, but volumes are either not reported or only vaguely given. Variations in volume may reduce individual values of $\gamma_{50}$. The $\gamma_{50}$ value quoted in Table 5.5 for nasopharynx seems unusually high; however, there is an additional entry in the Okunieff tables for nasopharynx (T1 + T2, single institution) with $D_{50} = 59$ and $\gamma_{50} = 47.3$, an even larger value. Neither mechanistic model described below can accommodate the high $\gamma_{50}$ value for nasopharynx.

Okunieff et al. obtained $D_{50}$ and $\gamma_{50}$ values based on data fits to the following TCP formula:

$$\mathrm{TCP} = \frac{e^{(D-D_{50})/k}}{1+e^{(D-D_{50})/k}}, \tag{5.64}$$

where $\gamma_{50} = D_{50}/(4k)$.[8] Note that TCP is not zero when $D = 0$.

**TABLE 5.5 TCP Data for Selected Tumors**

| TCP Data from Okunieff et al. (1995) | | | | | | Intertumor Variation Models[a] $\rho V = 10^8$ | | Homogeneous Model[b] ($\sigma_e = 0$) | |
|---|---|---|---|---|---|---|---|---|---|
| **Tumor Site** | **Comments** | $D_{50}$ **(Gy)** | $\gamma_{50}$ | $D_{95}$[c] **(Gy)** | $\alpha/\beta$ **(Gy)** | $\langle\bar{\alpha}\rangle$ **(Gy$^{-1}$)** | $\sigma_e$ **(Gy$^{-1}$)** | $N$ | $\alpha$ **(Gy$^{-1}$)** |
| Prostate | T0 and T1, multi[d] | 28.3 | 1.00 | 49 | 1.4 | 0.671 | 0.264 | 12.4 | 0.102 |
| Anal | ≤4 to 5 cm, multi[d] | 32.9 | 0.58 | 74 | 10 | 0.577 | 0.395 | 3.7 | 0.051 |
| Breast | T1 + T2, adjuvant[e] | 30.9 | 1.30 | 48 | 2.8 | 0.614 | 0.184 | 30 | 0.121 |
| Lung | All NSCLC, single[f] | 52.0 | 1.81 | 73 | 10 | 0.365 | 0.0766 | 129 | 0.100 |
| Nasopharynx | Node < 3 cm | 50.0 | 17.5 | 52 | 10 | No solution | | $6 \times 10^{21}$ | 1.01 |

[a] Based on solution of Equations 5.77 for $D_{50}$ and 5.78 for $\gamma_{50}$.
[b] Based on solution of Equations 5.69 and 5.70.
[c] $D_{95}$ from Equation 5.64.
[d] Multi = based on data from multiple institutions.
[e] Breast adjuvant, volume assumed small.
[f] Single = based on data from a single institution.

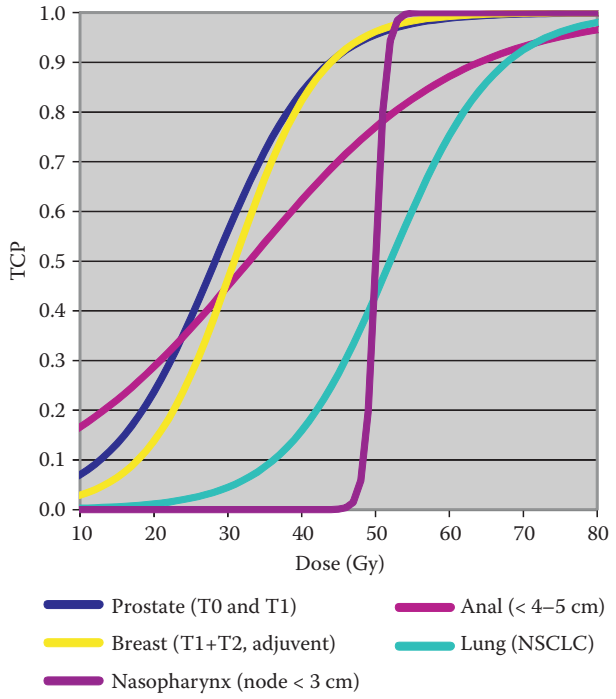



FIGURE 5.20 TCP data from Okunieff et al. (1995) for selected tumors. These curves are based on fits to Equation 5.64 for $D_{50}$ and $\gamma_{50}$. All fits except for lung (and possibly breast) are based on data from multiple institutions. This averaging may have the effect of reducing $\gamma_{50}$.

Figure 5.20 shows a plot of Equation 5.64 for the tumors in Table 5.5. The anal tumors have $D_{50} = 32.9$ Gy, but a small value of $\gamma_{50}$ (0.58), and therefore $D_{95}$ is relatively high (~75 Gy). The nasopharynx tumors have a very steep dose response, and therefore the dose must exceed a threshold of about 48 Gy for any likelihood of tumor control.

## 5.5.2 TCP MECHANISTIC MODELS

### 5.5.2.1 HOMOGENEOUS CASE

For tumors, the generic value of $\alpha/\beta$ is approximately 10. The ratio of the first term to the second term in the exponent of the cell survival, Equation 5.29 is $\alpha/(\beta d) \sim 5$ for typical fractionation when $d = 2$ Gy. It is therefore common to neglect the $\beta$ term in Equation 5.29. This is perhaps invalid for prostate cancer, as there is evidence that $\alpha/\beta$ may be as low as 1.4 for prostate (Miralbell et al., 2012). For breast cancer, $\alpha/\beta = 2.8$ has recently been quoted (Qi et al., 2011). Based on these data, we will not neglect the $\beta$ term until a later section, when it becomes difficult to proceed without doing do. The neglect of the $\beta$ term will not be valid for large doses per fraction (stereotactic radiosurgery or hypofractionation).

For tumor control, an FSU is assumed to consist of a single clonogenic cell. The TCP is assumed to be the probability that no cell survives. If a tumor consists of

$N$ clonogenic cells, the average number surviving an absorbed dose $D$ is $\mu = N S_f$, where $S_f$ is given by Equation 5.29. It is probably reasonable to use the Poisson probability here because we are interested in values of TCP such that the probability of individual cell survival is small and the number of cells is large. The probability that no cells survive ($x = 0$) in Equation 5.6 is $f(0) = e^{-\mu}$. Under these circumstances, the TCP is given by

$$\mathrm{TCP} = \exp\left[-Ne^{-D(\alpha+\beta d)}\right]. \tag{5.65}$$

Note that the effect of the neglect of the β term in Equation 5.65 would imply that TCP does not depend on the number of fractions *f*.

Equation 5.65 ignores proliferation. According to Hall and Giaccia (2012), the doubling time of human tumors spans a large range, and the median value is about 60 days, although values as short as 3 days (Trott et al., 1985) have been reported for Hodgkin's disease. If tumor growth is exponential, then the number of cells at time $t$ is proportional to $e^{\eta t}$, where $\eta = (\ln 2)/T_2$ and $T_2$ is the doubling time. The number of cells $N_c(t)$ surviving after time $t$ during which $f$ fractions are delivered is

$$N_c(t) = N_c(0)e^{\eta t}e^{-D(\alpha+\beta d)}, \tag{5.66}$$

where $N_c(0)$ is the number at the beginning of treatment. Proliferation is unimportant when $\eta t \ll D(\alpha + \beta d)$ or when $t \ll D(\alpha + \beta d)T_2/\ln 2$. If we assume that $\alpha = 0.35\ \mathrm{Gy}^{-1}$, $\alpha/\beta = 10$ Gy, $d = 2$ Gy, $D = 70$ Gy, and $T_2 = 60$ days, we find that when $t \ll 2500$ days, proliferation is unimportant. For these parameters, tumor proliferation during treatment is unimportant, although this may not be true under all circumstances.[9] We will neglect proliferation from this point forward.

Let us suppose that $N = \rho V$, where ρ is the number density of clonogenic cells and $V$ is the volume of the tumor.[10] We can now write

$$\mathrm{TCP} = \exp\left[-\rho V e^{-D(\alpha+\beta d)}\right]. \tag{5.67}$$

A typical mammalian cell is 10 μm in diameter. If there was perfect cell packing with no interstices between cells, this would lead to a cell density of about $2 \times 10^9$ $\mathrm{cm}^{-3}$. Not all cells are clonogenic, and there is not perfect packing; therefore, this is an upper limit to the clonogenic cell density.

Cell lines established from biopsies can be used to measure α and β *in vitro*, but clearly, the *in vitro* environment does not reproduce the *in vivo* environment. It is not difficult to infer values of α/β from fractionation studies; however, separate values of α and β are difficult to come by. For low LET radiation, *in vitro* measurements yield values of α between 0.1 $\mathrm{Gy}^{-1}$ (highly radioresistant) and 2 $\mathrm{Gy}^{-1}$ (very radiosensitive). Commonly used values in the literature are $\alpha = 0.35\ \mathrm{Gy}^{-1}$ and $\rho = 10^7\ \mathrm{cm}^{-3}$ (Nahum and Tait, 1992).

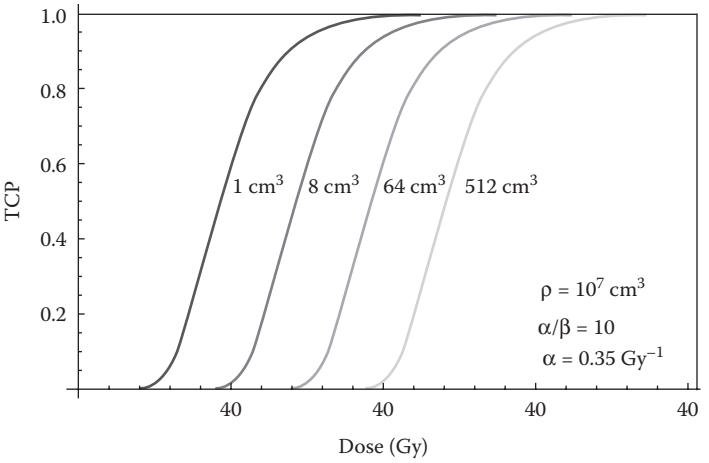



FIGURE 5.21 Tumor control probability–response curves for $\alpha = 0.35\ \text{Gy}^{-1}$, $\alpha/\beta = 10$, and $\rho = 10^7\ \text{cm}^{-3}$ and tumor volumes of 1, 8, 64, and 512 $\text{cm}^3$.

We can solve Equation 5.67 for the dose necessary to achieve a given TCP as follows:

$$D = \frac{1}{\alpha + \beta d}\left[\ln(\rho V) - \ln\left(\ln\left(\frac{1}{\text{TCP}}\right)\right)\right], \tag{5.68}$$

where the dose per fraction is assumed constant. The dose for a given TCP depends sensitively on α, but rather insensitively on the tumor volume. Figure 5.21 shows the TCP as a function of dose for $\rho = 10^7$ and $\alpha = 0.35\ \text{Gy}^{-1}$ for tumor volumes of 1, 8, 64, and 512 $\text{cm}^3$.

An expression for $D_{50}$ is easily derived from Equation 5.68:

$$D_{50} = \frac{1}{\alpha + \beta d}\ln\left(\frac{\rho V}{\ln 2}\right), \tag{5.69}$$

assuming that $d$ is constant. For the model described by Equation 5.67, $\gamma_{50}$ is given by

$$\gamma_{50} = \frac{\ln 2}{2}\ln\left(\frac{\rho V}{\ln 2}\right). \tag{5.70}$$

Notice that $\gamma_{50}$ is solely determined by $N = \rho V$.

Although $D_{50}$ and $\gamma_{50}$ both depend on the volume of the tumor, the ratio does not. The ratio of these two quantities predicts the value of $\alpha + \beta d$.

For $N = 10^9$, Equation 5.70 predicts $\gamma_{50} \approx 7$. Observed values of $\gamma_{50} \approx 2$ (or even less) appear to present a problem, although we note that there are some large observed values (see Table 5.5, nasopharynx). Equation 5.70 implies that $N$ is on the order of

200 if typical values of $\gamma_{50} \approx 2$. This is eight orders of magnitude less than the number of cells in a small 10 $cm^3$ tumor. In addition, Equation 5.69 then predicts an unusually small value of $\alpha$. If $D_{50} \approx 40 - 70$ Gy, then $\alpha \leq 0.1$ $Gy^{-1}$. For $D_{50} = 50$ Gy and $\gamma_{50} \approx 2$, Equations 5.69 and 5.70 predict $N = 222$ and $\alpha = 0.115$ $Gy^{-1}$. In turn, Equation 5.68 predicts that $D_{95}$ (dose required for TCP = 95%) is 73 Gy. For some tumors, the situation is even more extreme. For anal tumors (Table 5.5), $N = 3.7$ and $\alpha = 0.051$ $Gy^{-1}$. Large observed values of $\gamma_{50}$ present the opposite problem. For nasopharynx, $\gamma_{50} = 17.5$ (see Table 5.5) implies a value of $N$ on the order of $10^{22}$, which is highly unrealistic. These models are summarized in Table 5.5 (right-hand two columns labeled homogeneous model). Figure 5.22 shows $\gamma_{50}$ as a function of tumor volume for a variety of clonogen densities ranging from $10^2$ to $10^9$ $cm^{-3}$. The value of $\gamma_{50}$ is relatively independent of volume, and observed values are only consistent with small values of $\rho$. Only values of $\rho$ on the order of $10^2$ $cm^{-3}$ can reproduce (average) observed values of $\gamma_{50}$, which are approximately 2.

Two possible solutions to this puzzle of small $N$ (or $\rho$) values have been discussed in the literature. The first of these is to accept the results. In fact, there is evidence that tumors may consist of a small number of radioresistant clonogens. A number of authors are cited by Yorke (2003) as believing that tumors are dominated by a small number of radioresistant clonogens, while the vast majority of cells are either not clonogens or are killed by low doses. The second solution recognizes that tumor control data are based on populations, not individuals. Individuals are either controlled locally or not. TCP data represent an average over many individuals who may have tumors of differing radiosensitivity. It seems as if other possibilities exist also, for example, averaging over tumor volume $V$ between patients, dose variations throughout the tumor, or variation of clonogen density within a tumor.

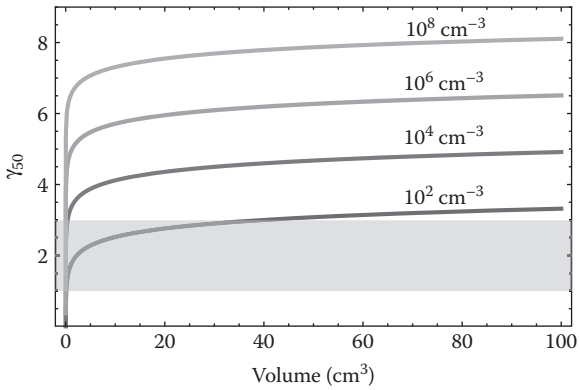



FIGURE 5.22 $\gamma_{50}$ as a function of the tumor volume (in $cm^3$) for various values of the clonogen density $\rho$ ranging from $10^8$ down to $10^2$ $cm^{-3}$. It is seen that the value of $\gamma_{50}$ is relatively constant with volume except for very small volumes (less than a few cc). Only very small values of $\rho$ (on the order of $10^2$ $cm^{-3}$) can reproduce (average) observed values of $\gamma_{50}$, which fall in the shaded range between about 1.0 and 3.0.

If the number of radioresistant clonogens in a tumor is really on the order of several hundred, this is quite astonishing. We are using large doses to kill just several hundred cells. That makes radiation therapy a very blunt instrument indeed.

### 5.5.2.2 TCP MECHANISTIC MODELS: TUMOR HETEROGENEITY

**5.5.2.2.1 Intratumor Variation** It is known that individual tumors are highly heterogeneous, and it would not be surprising if there is a significant variation in radiosensitivity throughout a single tumor. As an example, hypoxic cells are more radioresistant than nonhypoxic cells. In this section, we concentrate on intratumor variation in radiosensitivity. According to Webb (1997), intratumor variability primarily affects $D_{50}$. In this section, we assume that the value of α varies throughout the tumor. We also neglect β; that is, we assume $\beta d/\alpha \ll 1$.

Let us assume that the number of cells having a value of α between α and $\alpha + d\alpha$ is given by a Gaussian distribution, and therefore

$$dN = \frac{N}{\sigma_a\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\alpha-\bar{\alpha}}{\sigma_a}\right)^2} d\alpha, \tag{5.71}$$

where:

- $N$ is the total number of cells in the tumor
- $\bar{\alpha}$ is the average value of α
- $\sigma_a$ is the standard deviation of α.

Equation 5.71 is only valid when $\sigma_a/\alpha \ll 1$; otherwise, zero or negative values of α would have a significant probability of occurrence. The probability, δ$P$, that a dose $D$ will eradicate all of these $dN$ cells is $e^{-\mu}$, where $\mu = S_f\, dN$:

$$\delta P = \exp\left\{\frac{-N}{\sigma_a\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\alpha-\bar{\alpha}}{\sigma_a}\right)^2} e^{-\alpha D} d\alpha\right\}. \tag{5.72}$$

The TCP is the product of all such terms, and therefore the log is given by

$$\ln \overline{\overline{\mathrm{TCP}}} = -\frac{N}{\sigma_a\sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}\left(\frac{\alpha-\bar{\alpha}}{\sigma_a}\right)^2} e^{-\alpha D} d\alpha, \tag{5.73}$$

where $\overline{\mathrm{TCP}}$ is the tumor control probability for intratumor averaging over α.

The integral in Equation 5.73 is the Laplace transform of the Gaussian with respect to α (see Equation 5.87). The result of taking the Laplace transform is

$$\ln \overline{\mathrm{TCP}} = -\frac{1}{2} N e^{-\bar{\alpha}D\left(1-\frac{D\sigma_a^2}{2\bar{\alpha}}\right)} \left\{1 + \mathrm{erf}\left[\frac{\bar{\alpha} - D\sigma_a^2}{\sigma_a\sqrt{2}}\right]\right\}. \tag{5.74}$$

In the limit in which $\sigma_a \to 0$, this result reduces to $\ln \overline{\mathrm{TCP}} = -Ne^{-\bar{\alpha}D}$, as expected. The quantity $\gamma_{50}$ is given by

$$\gamma_{50} = \frac{\ln 2}{2} \bar{\alpha} D_{50} \left\{1 - \frac{\sigma_a}{\bar{\alpha}} \left[\sigma_a D_{50} - \frac{N}{\sqrt{2\pi}\ln 2} e^{-\frac{\bar{\alpha}^2}{2\sigma_a^2}}\right]\right\}. \tag{5.75}$$

When $\sigma_a \to 0$, this reduces to Equation 5.70.

Figure 5.23 shows the effect of intratumor variability. Even a relatively small standard deviation in $\alpha$ can lead to a significant increase in $D_{50}$ when $N$ is relatively large. The reason for this is that the spread in $\alpha$ introduces a radioresistant component of cells. Given the heterogeneity of tumors, it seems likely that this shift in $D_{50}$ will occur in real tumors. To put it another way, given a value of $D_{50}$, intratumor variability will make it appear as if the effective value of $\alpha$ is lower than $\bar{\alpha}$. Because of the sensitivity of $D_{50}$ to $\sigma_a$, realistic models of tumor control should include this effect when $N$ is large.

We can also look at the case where we have two different components, a component consisting of $N_1$ cells with normal radiosensitivity $\alpha_1$ and a radioresistant component with $N_2$ cells and sensitivity $\alpha_2$. The TCP for component $i$ is $\ln \mathrm{TCP} = -N_i e^{-\alpha_i D}$, and therefore for this two-component model,

$$\mathrm{TCP} = \exp\left\{-N\left(\frac{N_1}{N} e^{-\alpha_1 D} + \frac{N_2}{N} e^{-\alpha_2 D}\right)\right\}. \tag{5.76}$$

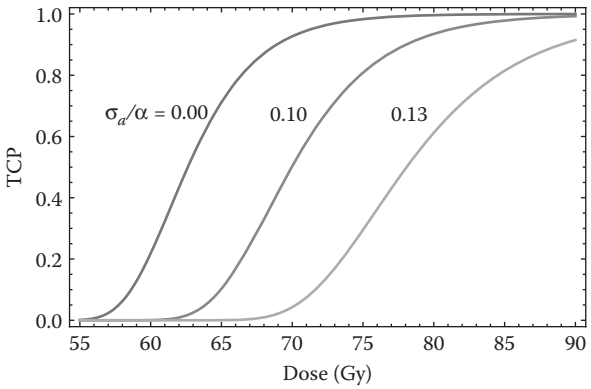



FIGURE 5.23 TCP vs. dose for intratumor variability. For all curves, $N = 10^8$, $\alpha = 0.30$ Gy$^{-1}$. The curve on the left has $\sigma_a = 0$. The value of $D_{50}$ is 62.6 Gy. The middle curve has $\sigma_a/\alpha = 10\%$. The value of $D_{50}$ shifts to 70.1 Gy. The curve on the right has $\sigma_a/\alpha = 13\%$. This causes a shift in $D_{50}$ to 80.1 Gy. The value of $\gamma_{50}$ is relatively insensitive to the value of $\sigma_a$.

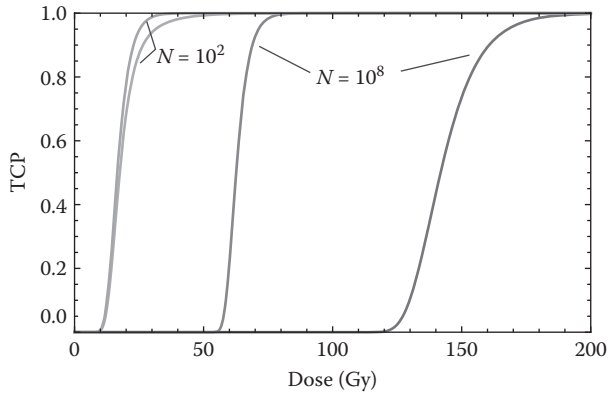



FIGURE 5.24 The effect of a two-component sensitivity model. We compare a homogeneous tumor ($\alpha = 0.30$ Gy$^{-1}$) with $N = 10^2$ cells (left-most curve) and $N = 10^8$ cells (left most curve labeled $N = 10^8$) to a two-component model containing just 1% of radioresistant cells having $\alpha = 0.10$ Gy$^{-1}$. The effect on the tumor having a small number of clonogens is slight. In contrast, the tumor having a large number of cells shows a dramatic increase in $D_{50}$ from 62 to 140 Gy.

Figure 5.24 shows an illustrative example of this for two different values of $N$ ($10^2$ and $10^8$). There is a dramatic difference in the way that a tumor responds to intratumor variability, depending on the number of clonogens. Tumors having a large number of clonogens are very sensitive to intratumor variations. A small number of radioresistant cells could effectively determine the value of $D_{50}$.

**5.5.2.2.2 Intertumor Variations** Intertumor variations pertain to variations in parameters between patients. The value of $\gamma_{50}$ depends on the log of the tumor volume. This certainly varies between patients, but because $\gamma_{50}$ depends on the slowly varying log of the volume, a population average will not reduce $\gamma_{50}$ significantly. Intertumor variations of radiosensitivity involve variations in the value of $\bar{\alpha}$, whereas intratumor variations involve variations in $\alpha$.

According to Moiseenko et al. (2005, p. 194) and references therein, "in-vitro measurements of radiosensitivity (for cell lines established from biopsy specimens) vary significantly between patients (the coefficient of variation is typically 30%–50%). It is not clear how one rules out intratumor variation (depending on the location from which the biopsy specimen is taken) as a confounding factor that could mimic intertumor variation. In addition, the biopsy specimens are presumably evaluated *in vitro*.

As explained by Wein et al. (2000), interpatient variability solves two problems:

1. The resulting radiosensitivity and clonogen number are more consistent with expectations from *in vitro* data.
2. The dose–response curves are more shallow and can match the observed clinical data.

One of the difficulties with this argument is that it is not clear that there is a problem to begin with. It is not obvious that *in vitro* values of α are closely correlated to *in vivo* values. In addition, as previously discussed, there is some evidence that the number of radioresistant clonogens is in fact quite small.

We can calculate the intertumor variation by averaging the TCP values over a Gaussian distribution of $\bar{\alpha}$ values with standard deviation $\sigma_e$:

$$\langle \text{TCP} \rangle = \frac{1}{\sigma_e \sqrt{2\pi}} \int_0^\infty e^{-\frac{1}{2}\left(\frac{\bar{\alpha}-\langle\bar{\alpha}\rangle}{\sigma_e}\right)^2} e^{-Ne^{-\bar{\alpha}D}} d\bar{\alpha}, \tag{5.77}$$

where the quantity $\langle\bar{\alpha}\rangle$ is the interpatient average of $\bar{\alpha}$. The value of $\sigma_e$ cannot be too large; otherwise, there will be a significant probability of values of $\bar{\alpha} \le 0$.

The parameter $\gamma_{50}$ is obtained from the definition (see Equation 5.1)

$$\gamma_{50} = \frac{ND_{50}}{\sigma_e \sqrt{2\pi}} \int_0^\infty \bar{\alpha} e^{-\bar{\alpha}D_{50} - Ne^{-\bar{\alpha}D_{50}}} e^{-\frac{1}{2}\left(\frac{\bar{\alpha}-\langle\bar{\alpha}\rangle}{\sigma_e}\right)^2} d\bar{\alpha}. \tag{5.78}$$

We have examined the case in which $N = 10^7$, $\alpha = 0.35$ Gy$^{-1}$ for various values of $\sigma_e$. We have integrated Equation 5.77 numerically using the NIntegrate function in Mathematica. We find that the value of $D_{50}$ hardly changes compared to the case in which $\sigma_e = 0$. For $\sigma_e = 0$, the value of $D_{50} = 47.2$ Gy. For $\sigma_e = 0.1$ Gy$^{-1}$, $D_{50} = 47.9$ Gy. The value of $\gamma_{50}$, however, drops substantially, from 5.7 to approximately 1.5. This is illustrated in Figure 5.25. The effect of intertumor variability is to primarily change the value of $\gamma_{50}$ without significantly affecting $D_{50}$. In contrast, $D_{50}$ is very sensitive to

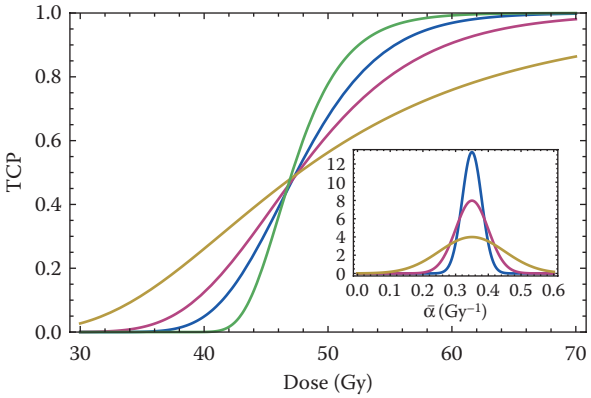



FIGURE 5.25 The effect of intertumor variation in radiosensitivity. The green curve shows the TCP for $N = 10^7$, $\langle\bar{\alpha}\rangle = 0.35$ $\sigma_e = 0$. The other curves are identical except that $\sigma_e = 0.03$, 0.05, and 0.1 (blue, violet, and olive green curves). The value of $D_{50}$ hardly changes, but the slope decreases dramatically as $\sigma_e$ increases. The graph inset shows the Gaussians that the TCP curves are averaged over.

intratumor variability (when $N$ is large), but $\gamma_{50}$ is insensitive. It may be that neither of these effects can be ignored.

Some intertumor variation models appear in Table 5.5. All of these models assume $N = 10^8$, and they are based on simultaneous solution of Equations 5.77 and 5.78, with TCP = 0.5 at $D = D_{50}$. These solutions were obtained using the Mathematica commands "NIntegrate" and "FindRoot." The $D_{50}$ and $\gamma_{50}$ input values are based on the fit to Equation 5.64, and therefore these models do not predict the same $D_{95}$ that Equation 5.64 does. For this reason, these models are for illustrative purposes only. The values of $\langle\bar{\alpha}\rangle$ are closer to *in vitro* values than those of the models in Section 5.5.2.1.

The steepest possible patient-specific (as opposed to population) dose–response curve is a unit step function (or Heaviside function). The unit step will go from a value of zero to a value of 1.00 at a dose equal to

$$D_0 = D_{50} = -\frac{1}{\alpha}\ln\left(\frac{\ln 2}{N}\right).$$

Values of $\alpha$ will vary over the population, and therefore the location of the step will vary. Let us assume that the distribution in the patient-specific values of $D_0$ obeys a normal distribution with a standard deviation of $\sigma_D$. If we average over these values of $D_0$ and then take the derivative to calculate the population $\gamma_{50}$, it can be shown that[11]

$$\gamma_{50} = \frac{1}{\sqrt{2\pi}}\frac{\langle\bar{\alpha}\rangle}{\sigma_e}. \tag{5.79}$$

This is an upper limit to the slope (Moiseenko et al., 2005).

**5.5.2.2.3 Nonuniform Dose Distribution** The dose throughout a tumor is in reality never uniform. Let us assume that each voxel containing $N_i = \rho_i V_i$ cells receives a uniform dose, where $\rho_i$ is the clonogen density and $V_i$ is the voxel volume. The probability of eradicating a particular voxel $i$ is (see Equation 5.31)

$$P_i = \exp\left\{-\rho_i V_i e^{-\alpha D_i - \beta D_i^2/f}\right\}, \tag{5.80}$$

where $D_i$ is the total dose received by voxel $i$. The discussion in this section applies to individuals, not the patient population as a whole. Let us assume that the tumor is homogeneous ($\alpha$ and $\beta$ constant throughout the tumor). Let there be a total of $M$ voxels. Assume that the clonogenic density is constant throughout the tumor. In this case, the TCP is

$$\mathrm{TCP} = \prod_{i=1}^{M}\exp\left(-\rho V_i e^{-\alpha D_i - \beta D_i^2/f}\right). \tag{5.81}$$

If the dose is constant throughout the entire volume, this reduces to Equation 5.65, as expected. Let us look at the case in which the target is considered a continuum. The product in Equation 5.81 indicates that the exponentials are to be multiplied; this means that the exponents themselves can be added. The sum of the exponents in Equation 5.81 can then be written as

$$\sum_{i=1}^{M} \rho V_i e^{-\alpha D_i - \beta D_i^2/f} \rightarrow \rho \int_0^{D_{\max}} h(D) e^{-\alpha D - \beta D^2/f} dD. \tag{5.82}$$

The expression for the TCP is now given by

$$\ln(\text{TCP}) = -\rho \int_0^{D_{\max}} h(D) e^{-\alpha D - \beta D^2/f} dD. \tag{5.83}$$

In terms of the cDVH, this can be expressed as

$$\ln(\text{TCP}) = \rho V \left[ \int_0^{D_{\max}} H(D)\left(\alpha + 2\beta D/f\right) e^{-\alpha D - \beta D^2/f} dD - 1 \right]. \tag{5.84}$$

If we assume that β is zero, then Equation 5.84 becomes

$$\ln(\text{TCP}) = -\rho V \left[ 1 - \alpha \int_0^{D_{\max}} H(D) e^{-\alpha D} dD \right]. \tag{5.85}$$

Evaluation of Equation 5.85 simply requires a sum over the cDVH. The upper limit in the integral in Equation 5.85 can be made infinite; the function $H(D)$ will provide the cutoff so that the integral does not diverge. The log of the TCP is related to the Laplace transform of $H(D)$:

$$\ln(\text{TCP}) = \rho V \left[ \alpha L\left(H(D)\right) - 1 \right] \tag{5.86}$$

where

$$L\left(H(D)\right) = \int_0^{\infty} H(D) e^{-\alpha D} dD. \tag{5.87}$$

The uniform dose that is equivalent to the nonuniform dose distribution can be found by setting ln(TCP) given by Equation 5.83 equal to the ln(TCP) given by Equation 5.65. The result is

$$D_e = -\frac{1}{\alpha}\ln\left[\frac{1}{V}\int_0^{D_{\max}} e^{-\alpha D}h(D)\,dD\right]. \tag{5.88}$$

No similarity is apparent between this and Equation 5.19 for the EUD.

When the dose is relatively uniform, as it usually is by design, $D_e$ can be written in terms of the average dose and the standard deviation of the dose. We write $D = (D-\bar{D})+\bar{D}$, where $\bar{D}$ is the average dose. The quantity $(D-\bar{D})$ is assumed to be small. The exponential term in Equation 5.88 is expanded in a power series in $(D-\bar{D})$ up to the second order. No assumptions are necessary regarding the form of $h(D)$. The result is

$$D_e \approx \bar{D} - \frac{1}{\alpha}\ln\left[1+\frac{1}{2}\alpha^2\sigma_D^2\right], \tag{5.89}$$

where

$$\sigma_D^2 = \frac{1}{V}\int h(D)\left(D-\bar{D}\right)^2 dD.$$

Equation 5.89 shows that the equivalent dose is always less than the average dose. This equation allows a trivially easy computation of an approximate value of $D_e$, knowing only the average and standard deviation of the dose distribution. It is possible to carry out a similar expansion for the EUD (Equation 5.19), yielding

$$\text{EUD} \approx \bar{D}\left[1-\frac{n-1}{2n^2}\left(\frac{\sigma_D}{\bar{D}}\right)^2\right]^n. \tag{5.90}$$

Comparison of Equations 5.89 and 5.90 shows that (assuming the standard deviation is small)

$$n \approx \frac{-1}{\alpha\bar{D}-1}. \tag{5.91}$$

We see that the value of the volume index $n$ depends on the average dose. For $\alpha \approx 0.25\ \text{Gy}^{-1}$ and $\bar{D} = 70\,\text{Gy}$, $n = -0.06$. Accepted values of $n$ are on the order of –0.10.

The enhancement in the TCP due to a hot spot can be readily seen to be

$$\frac{\mathrm{TCP}}{\mathrm{TCP}_0} = \exp\left\{\rho V \alpha \int_{D_{\min}}^{D_{\max}} H(D) e^{-\alpha D} dD\right\} \tag{5.92}$$

where $\mathrm{TCP}_0$ is the TCP for a uniform dose of $D_{\min}$.

Consider the target cDVH in Figure 5.26, in which $(1 - \Delta v)$ represents the fraction of the volume receiving a dose of $D_{\max}$. This is an archetypal cDVH for a target. Real patient cDVHs are similar in form even though $\Delta v$ may be 1.0 and the portion of the graph between $D_{\min}$ and $D_{\max}$ may have curvature and wiggles. If the prescribed dose is $D_{\max}$, then $D_{\min}$ represents a cold spot and $1 - \Delta v$ represents the volume receiving the prescription dose. On the other hand, if $D_{\min}$ represents the prescription dose and $\Delta v = 1.0$, then $D_{\max}$ represents the hot spot dose. The TCP for the DVH in Figure 5.26 is given by

$$\mathrm{TCP} = \mathrm{Exp}\left\{-Ne^{-\alpha D_{\max}}\left[1 + \frac{\Delta v}{\alpha \Delta D}\left(e^{\alpha \Delta D} - 1 - \alpha \Delta D\right)\right]\right\}, \tag{5.93}$$

where $\Delta D = D_{\max} - D_{\min}$. Note that in the limit in which $\Delta D \to 0$ or $\Delta v \to 0$, the TCP reduces to the TCP for a constant dose $D_{\max}$.

Let us consider $D_{\min}$ to represent a cold spot. Let $\mathrm{TCP}_0 = \mathrm{Exp}\left[-Ne^{-\alpha D_{\max}}\right]$ represent the TCP in the absence of the cold spot (i.e., for uniform irradiation). We can then express TCP with the cold spot in terms of this:

$$\mathrm{TCP} = \mathrm{TCP}_0^{\left[1 + \frac{\Delta v}{\alpha \Delta D}\left(e^{\alpha \Delta D} - 1 - \alpha \Delta D\right)\right]}. \tag{5.94}$$

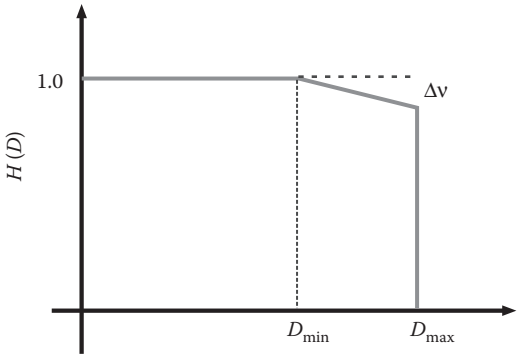



FIGURE 5.26 cDVH for a target volume. If $D_{\max}$ is the prescribed dose, then $1 - \Delta v$ is the fractional volume receiving this dose and $D_{\min}$ is the cold spot dose. If $D_{\min}$ is the prescribed dose and $\Delta v = 1$, then $D_{\max}$ represents the hot spot dose. These two scenarios are discussed in the text.

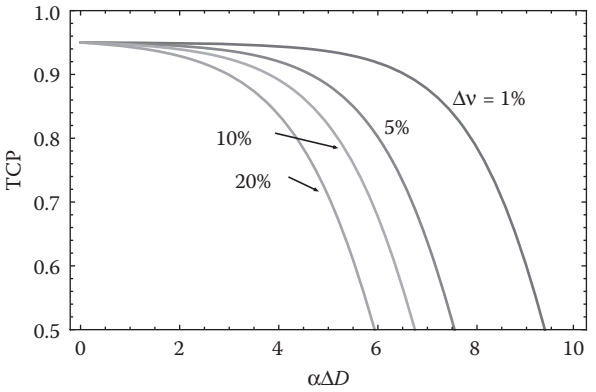



FIGURE 5.27 The effect of a cold spot on TCP. TCP is plotted as a function of $\alpha\Delta D$ for various values of $\Delta\nu$ based on Equation 5.94. The quantity $1 - \Delta\nu$ is the fraction of the volume receiving the prescription dose. All TCP values are 95% for $\Delta D = 0$.

The TCP in Equation 5.94 is independent of $N$ given a value of $TCP_0$. A graph of the TCP as given by Equation 5.94 is shown in Figure 5.27 for $TCP_0 = 95\%$ as a function of $\alpha\Delta D$ for various values of $\Delta\nu$. As an example, if $\Delta\nu = 10\%$, the cold spot is 80% of the prescription dose ($D_{min} = 0.8\ D_{max}$), $D_{max} = 70$ Gy, and $\alpha = 0.35\ Gy^{-1}$; then $\alpha\ \Delta D = 4.9$ and TCP drops from 95% to about 82%.

If we now consider $D_{min}$ to represent the prescription dose and $\Delta\nu = 1.0$, then $D_{max}$ represents the hot spot dose. Under these circumstances, Equation 5.93 can be written

$$\mathrm{TCP} = \exp\left\{-\frac{N}{\alpha\Delta D}e^{-\alpha D_{\min}}\left[1-e^{-\alpha\Delta D}\right]\right\} = \exp\left\{-\frac{2Ne^{-\alpha\bar{D}}}{\alpha\Delta D}\sinh\frac{\alpha\Delta D}{2}\right\}. \tag{5.95}$$

In the case in which $\alpha\ \Delta D \ll 1$,

$$\mathrm{TCP} = \exp\left\{-Ne^{-\alpha D_{\min}}\right\},$$

as expected. The uniform dose $D_e$ that is equivalent to this nonuniform dose is found by setting the TCP given by Equation 5.95 equal to the TCP given by Equation 5.65:

$$D_e = D_{\min} - \frac{1}{\alpha}\ln\left[\frac{1-e^{-\alpha\Delta D}}{\alpha\Delta D}\right]. \tag{5.96}$$

Note that $D_e$ is independent of $N$. When $\alpha\ \Delta D$ is small, this expression is (to the second order in $\Delta D$)

$$\frac{D_e}{D} \approx 1 - \frac{(\Delta D)^2}{24}\left(\frac{\alpha}{\bar{D}}\right). \tag{5.97}$$

This shows again that $D_e$ is always somewhat less than the average dose.

For the purpose of comparing rival treatment plans, it may be useful for the TCP of both plans to be the same, allowing a direct comparison of NTCP values. Otherwise, the comparison is an "apples to oranges" comparison. The question is, how do we accomplish this given two different dose distributions, $D_1(\vec{r})$ and $D_2(\vec{r})$, where $\vec{r}$ is the position vector? The dose is proportional to the total monitor units (MU) delivered. This may be adjusted up or down by adjusting the total MU without changing the *distribution* of dose. The problem then becomes one of adjusting the MU of plan 1 so that $\mathrm{TCP}_1 = \mathrm{TCP}_2$. If $D_1' = aD_1$ then $\mathrm{EUD}_1' = a\,\mathrm{EUD}_1$ and $D_{e1}' = aD_{e1}$; therefore, $\mathrm{MU}_1' = a\mathrm{MU}_1$. Plan 1 can be forced to have the same TCP as plan 2 by multiplying the MU for plan 1 by $a = \mathrm{EUD}_2/\mathrm{EUD}_1$ or $a = D_{e2}/D_{e1}$, depending on whether one favors the EUD model or the mechanistic model of this section.

## 5.6 PROBABILITY OF UNCOMPLICATED CONTROL

The overall goal of radiation therapy is to control the tumor without causing complications. The probability of uncomplicated control, $P_+$, is defined as the probability that no complication occurs *and* that the tumor is controlled:

$$P_+ = \mathrm{TCP} \times \prod_{i=1}^{M} (1 - \mathrm{NTCP}_i), \tag{5.98}$$

where $M$ is the number of OARs. If we assume that the target and the OAR receive the same uniform dose, we can plot TCP and NTCP versus dose on the same graph, as shown in Figure 5.28. For a reasonable chance of uncomplicated control, the NTCP curve needs to be to the right of the TCP curve, and it needs to be far enough to the right so that the product, $P_+$, in Equation 5.98 takes on a large value over some "therapeutic window" of dose. If the OAR receives less dose than the target, then the therapeutic window becomes larger. In some cases, there may only be a narrow window over which the probability of uncomplicated control exceeds a reasonably large value.

## 5.7 CONCLUSIONS/SUMMARY

The prospect of constructing mechanistic models for NTCP and TCP is seductive but daunting. With apologies to Winston Churchill, mechanistic models are based on uncertainties, piled on top of enigmas, wrapped up in ignorance. Ignorance and confusion persist; however, they now persist on a higher plane. It is difficult to reach any clinically useful conclusions based on mechanistic models. One of the reasons for this is that the parameters are so poorly known. *In vivo* values of α and β are unavailable and must be found from fits to clinical data. It is unlikely that mechanistic models

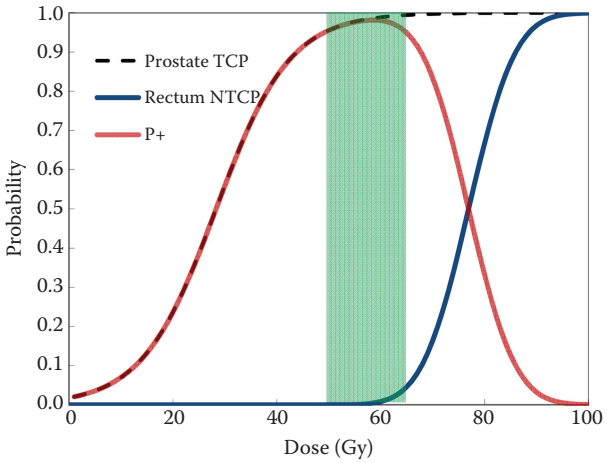



FIGURE 5.28 Probability of uncomplicated control for prostate cancer (T0 and T1) and rectal complications assuming that the rectum receives a uniform dose that is the same as for the prostate. This is based on data from Okunieff (prostate) and QUANTEC (rectal complications). The probability of uncomplicated control, *P*+ is 95% over a range in dose between 50 and 65 Gy (shaded region). This is the therapeutic window. The maximum value of *P*+ is 98% for a dose of 59 Gy.

will improve until better clinical data are available. The conclusion in the QUANTEC report is that more data are needed, not more models.

Models that fit observed data and have biology and probability built into them may be preferred over simple empirical data fitting models. At least there is some science built into these models, crude though it may be.

### 5.7.1 SERIAL OAR

It is expected that the maximum number of cells in an OAR is $N_c = N_0 k \sim 10^8$. We have performed calculations for simple serial models with no intra- or interpatient variability. The value of $\gamma_{50}$ is independent of $\alpha$ and $\beta$ and depends only on $N_0$ and $k$. When $N_0 \gg 1$, $\mathrm{NTCP} = 1 - 2^{-(D/D_{50})^k}$ and $\gamma_{50} \approx \frac{1}{2} k \ln 2$ (independent of $N_0$). If, in addition,

$$\left(\frac{\ln 2}{N_0}\right)^{1/k} \ll 1,$$

then

$$n \approx \frac{1}{k} \approx \frac{\ln 2}{2\gamma_{50}}.$$

The value of $D_{50}$ is sensitive to $\alpha + \beta d$, as one would expect. To reproduce observed values of $\gamma_{50}$, either $N_0$ or $k$ (or both) must be small.

We have solved for values of $N_0$, $k$, $\alpha$, and $\beta$ using as input observed values for the complication rates for cervical cord with $\alpha/\beta = 0.87$. For one model, we find $N_0 = 7.7 \times 10^6$, $k = 13$, $\beta = 1.7 \times 10^{-3}$ Gy$^{-2}$, $D_{50} = 70$ Gy, and $\gamma_{50} = 3.8$. The values of $\beta$ are considerably smaller than *in vitro* values. The predicted value of $n \approx 0.08$ and the observed value is 0.05.

Values of NTCP are relatively independent of volume except for $v \lesssim 0.1$.

For inhomogeneous irradiation, an analytic expression is derived for NTCP in terms of an integral over the cDVH. We have examined the effects of a hot spot for nonuniform irradiation.

The cervical cord serial model based on Equations 5.34 and 5.35 (1) fits observed complication data reasonably accurately; (2) predicts power law behavior for the volume effect with

$$n \approx \frac{\ln 2}{2\gamma_{50}};$$

and (3) the predicted equivalent dose (Equation 5.50) is a generalization of the EUD that takes account of fractionation effects associated with nonuniform irradiation.

## 5.7.2 PARALLEL MODELS

A model has been constructed by assuming that $L$ or more FSUs out of $N_0$ must be destroyed to cause a complication. This implies a fractional volume threshold of $v = L/N_0$. The value of $D_{50}$ depends only on the ratio $L/N_0$ and not on $L$ and $N_0$ separately. The value of $\gamma_{50}$ does not depend on $\alpha$ or $\beta$. Predicted values of $\gamma_{50}$ are orders of magnitude larger than observed values unless $L/N_0 \lesssim 10^{-6}$. Complication data are based on populations, not individuals. The value of $D_{50}$ may vary from patient to patient. If the model is correct for individuals, the NTCP curve for specific patients is essentially a step function of dose for which the NTCP goes from 0 to 1.0 at $D = D_{50}$. Averaging a step function over a Gaussian population distribution of $D_{50}$ with standard deviation $\sigma_{50}$ leads to a population average of

$$\langle \gamma_{50} \rangle = \frac{D_{50}}{\sigma_{50}\sqrt{2\pi}}.$$

It has been shown that the Lyman model is predicted by this averaging. An example model for kidneys assumes $D_{50} = 28$ Gy, $\langle \alpha + 2\beta \rangle = 0.25$ Gy$^{-1}$, 2 Gy per fraction, and $k = 1000$ cells per FSU. This results in a model for which $\sigma_{50}/D_{50} = 0.10$ and $L/N_0 = 0.40$. The models do not predict power law behavior for individuals for partial volume irradiation, except perhaps very crudely for $v \approx 1$.

### 5.7.3 TUMOR CONTROL PROBABILITY

Neglect of the β term in the linear quadratic cell survival probability is not justified for all tumors, particularly tumors of the prostate and breast.

The following results apply to a simple Poisson-based model without intra- or intertumor averaging. The dose for a particular value of the TCP depends on the logarithm of the volume. Values of $D_{50}$ and $\gamma_{50}$ depend on the tumor volume, but the ratio $\gamma_{50}/D_{50}$ depends only on $\alpha + \beta d$. The value of $\gamma_{50}$ depends only on the tumor volume and not on α or β. Observed values of $\gamma_{50} \approx 2$ imply that the number of clonogenic cells is ~$10^2$, eight orders of magnitude less than the number of cells in a small 10 $cm^3$ tumor. In addition, the model predicts unusually small values of $\alpha \lesssim 0.1$ $Gy^{-1}$. Two possible solutions are presented to address this dilemma. The first is to accept the values, arguing that the actual number of radioresistant clonogens to be destroyed by radiation is in fact as small as several hundred cells. The other possible solution involves variations in radiosensitivity, either intratumor or intertumor. Published dose–response curves apply to populations and not individuals.

Intratumor variations in α primarily affect $D_{50}$ and not $\gamma_{50}$. We have averaged the TCP over a Gaussian distribution with mean value $\bar{\alpha}$ and standard deviation $\sigma_a$. Even a small value of $\sigma_a$ leads to a significant decrease in the value of $D_{50}$ when $N$ is large. If $N$ is very large (~$10^8$), this effect cannot be ignored. For small values of $N$ (~$10^2$), however, the $D_{50}$ value is relatively insensitive to the value of $\sigma_a$.

Intertumor variations involve variation in the value of $\bar{\alpha}$. We have averaged TCP over a Gaussian distribution of $\bar{\alpha}$ values with an average value of $\langle\bar{\alpha}\rangle$ and a standard deviation of $\sigma_e$. Values of $D_{50}$ are insensitive to $\sigma_e$, but values of $\gamma_{50}$ are reduced when $\sigma_e > 0$. Assuming that $D_{50} = 50$ Gy, $\gamma_{50} = 2$, $\beta = 0$, and $N = 10^8$, we find that $\langle\bar{\alpha}\rangle = 0.379\ \text{Gy}^{-1}$ and $\sigma_e = 0.071\ \text{Gy}^{-1}$.

For inhomogeneous irradiation, ln(TCP) is related to the Laplace transform of the cDVH. Models with $N \sim 10^2$ are relatively insensitive to cold spots, whereas models with $N \sim 10^6 - 10^8$ have an TCP that is very sensitive to cold spots. An approximate value of the uniform equivalent dose for the mechanistic model may be calculated trivially from the value of α and the average and standard deviation of the dose for the target volume. The mechanistic model predicts an approximate value of the volume index $n$. The value of $n$ corresponds reasonably closely to values quoted in the literature. The index $n$ depends on the average dose, and the EUD may therefore not be a reliable quantity for predicting outcome.

## PROBLEMS

1. There are five links in a chain. The probability of a link breaking under a certain load is 0.01. Assume that this is the same for all the links.

a. What is the binomial probability that the chain will break?

b. Repeat the problem if the chain now has 100 links.

c. Repeat part (b) using the Poisson distribution

2. A thin rope consists of five strands. If two *or more* strands break, the rope itself will fail. If the probability of a strand failing is 0.20, calculate the probability that the rope will fail.

3. Show that the area under the DVH curve $H(D)$ is equal to the average dose $\bar{D}$ (see Equation 5.15).

4. Consider a sphere with dose that depends only radially on the distance from the center of the sphere, $r$; $h(D) = V/(D_{max} - D_{min})$ for $D_{min} < D < D_{max}$ and $h(D) = 0$ for $D < D_{min}$ and $D > D_{max}$, where $V$ is the volume of the sphere, $D_{max}$ is the dose at the center, $D_{min}$ is the dose at the surface, and $R$ is the radius of the sphere. Find the spatial distribution of dose $D = D(r)$.

5. Derive Equation 5.19 for the EUD in terms of the cDVH.

6. Show that the expression in Equation 5.20 approaches $D_{max}$ as $n$ approaches 0.

7. a. A uniform dose of about 30 Gy to the liver will result in approximately a 5% chance of radiation-induced liver disease. For a particular treatment plan, half of the liver receives a dose $D$ and the remainder receives no dose. What is the maximum value of $D$ if the probability of a complication is to remain at or below 5%? The value of $n$ for the liver is 0.86.

   b. Repeat part (a) for the kidneys if the uniform dose for a 5% complication probability is 15 Gy (for both kidneys irradiated) and the value of $n$ is 0.70. Note that this problem ignores the possibility of a threshold dose.

8. Set up a spreadsheet in Excel to calculate the NTCP using the LKB model for lungs. The complication is radiation pneumonitis and the parameters are $D_{50}(1) = 31$ Gy, $\gamma_{50} = 0.89$, and $n = 1.0$. Make an Excel "scatterplot" (with "smooth lines and markers") of NTCP versus the dose over a range from 0 to 70 Gy (use increments of 2 Gy). To calculate the error function, use the Excel function ERF. Make two plots on the same graph (labeling the axes and the plots), one for $\nu = 1.0$ and the other for $\nu = 0.5$.

   a. For uniform irradiation verify that NTCP = 0.5 at $D = D_{50}$.

   b. For uniform irradiation, what dose leads to a 10% complication rate? For partially uniform irradiation?

9. Derive Equation 5.38.

10. Show that Equation 5.54 for NTCP for a parallel OAR can be derived from the integral expression in Equation 5.53.

11. Use Equation 5.92 to calculate the enhancement in the TCP due to a 20% hot spot. Assume that $H(D) = 1$ for $D < D_{min}$ and that it is linear between $D_{min}$ and $D_{max}$. Calculate a numerical value for $D_{min} = 60$ Gy, $\alpha = 0.1$ Gy$^{-1}$, and $N = 10^2$.

12. The number of cells in an FSU for a serial organ is 20, and the number of FSUs in the organ is $1.0 \times 10^7$, $\alpha = 0.015$ Gy$^{-1}$, and $\alpha/\beta = 3$. If the dose distribution is uniform and the dose per fraction is $d = 2$ Gy,

    a. Calculate the fraction of the cells that survive a total dose of 22 Gy.

    b. Calculate the probability of killing all of the cells in an FSU.

    c. Calculate the probability of a complication.

13. A patient has a tumor with $N_0$ clonogenic cells. The patient delays the start of treatment for one doubling time (on the order of 60 days), during which the number of clonogenic cells increases to $2N_0$. If the treatment of the original tumor would have resulted in a TCP of $TCP_0$, what is the new TCP (expressed in terms of $TCP_0$) using the same total dose or fractionation scheme? If $TCP_0 = 95\%$, what is the new TCP?

## SYMBOLS

| | |
|---|---|
| $\bar{D}$ | Average dose in an OAR or target volume |
| $d$ | Dose per fraction |
| $D$ | Dose |
| $D(1)$ | Uniform tolerance dose for an entire volume |
| $D(\nu)$ | Tolerance dose for partial irradiation of volume $\nu$ |
| $D_{50}$ | Uniform dose leading to 50% complication probability |
| $\langle D_{50} \rangle$ | Interpatient average of $D_{50}$ |
| $D_e$ | Effective uniform dose that leads to the same NTCP or TCP |
| $D_i$ | Dose delivered to voxel $i$ |
| $D_{max}$ | Maximum dose in a structure |
| $D_{min}$ | Minimum dose in a structure |
| $D_t$ | Target dose |
| erf($t$) | Error function |
| EUD | Equivalent uniform dose |

| | |
|---|---|
| $f$ | Number of fractions |
| $f(x)$ | The probability of event $x$ |
| $h(D)$ | Differential dose volume histogram |
| $H(D)$ | Cumulative dose volume histogram (cDVH) |
| $k$ | Number of cells in an FSU *or* Okunieff parameter |
| $L$ | Critical number of FSU that must be damaged before a complication occurs |
| $m$ | Number of trials *or* a parameter related to $\gamma_{50}$ |
| $M$ | Number of different dose levels |
| $n$ | Dose volume index or parameter for partial volume irradiation |
| $\tilde{n}$ | Average value of $n$ for the parallel model |
| $N$ | Number of voxels in EUD equation *or* number of cells in a tumor |
| $N$ | Number of FSU irradiated |
| $N_0$ | Number of FSU in entire organ |
| $N_c$ | Total number of cells of the relevant type in an OAR |
| NTCP | Normal tissue complication probability |
| $\langle$NTCP$\rangle$ | Population averaged NTCP |
| $NTCP_0$ | NTCP for uniform irradiation |
| $NTCP_v$ | Normal tissue complication probability for fractional volume v |
| $p$ | Probability |
| $P$ | Probability of eradicating an FSU |
| $p$(A) or $p$(B) | Probability of event A or probability of event B |
| $P(\leq y)$ | Probability of less than or equal to $y$ successes |
| $P_+$ | Probability of uncomplicated control |
| $q$ | $1 - p$ |
| $r$ | Radial coordinate |
| $\vec{r}$ | Position vector |
| $R$ | Radius of sphere |
| $S$ | Fraction of cells surviving a dose $D$ |
| $S_f$ | Fraction of cells surviving a fractionated dose regimen with $f$ fractions |
| $t$ | Time or variable appearing in Lyman model |
| $T_2$ | Doubling time for tumor cell proliferation |
| TCP | Tumor control probability |
| $\overline{\text{TCP}}$ | Tumor control probability for an intratumor average over α |
| $\langle$TCP$\rangle$ | Tumor control probability based on an intertumor average over $\bar{\alpha}$ |
| $TCP_0$ | TCP for uniform irradiation |
| $V$ | Total volume of an OAR or target |
| $x$ | $L/N_0$ for the parallel model *or* number of successes in $m$ trials |
| $\bar{\alpha}$ | Intratumor average of α values |
| $\langle\bar{\alpha}\rangle$ | Intertumor average of $\bar{\alpha}$ values |
| α | Cell survival parameter in linear quadratic model, units $Gy^{-1}$ |

| | |
|---|---|
| $\beta$ | Cell survival parameter in linear quadratic model, units $Gy^{-2}$ |
| $\gamma_{50}$ | Logarithmic derivative of NTCP with respect to dose evaluated at $D_{50}$ |
| $\langle\gamma_{50}\rangle$ | Interpatient population average of $\gamma_{50}$ |
| $\mu$ | Average or mean value |
| $\rho$ | Number density of clonogenic cells in units of $cm^{-3}$ |
| $\sigma$ | Standard deviation |
| $\sigma_a$ | Intratumor standard deviation in the value of $\alpha$ |
| $\sigma_e$ | Intertumor standard deviation in the value of $\alpha$ |
| $\sigma_{50}$ | Standard deviation of $D_{50}$ |
| $\sigma_D$ | Standard deviation of the dose |
| $\rho$ | Number density of clonogenic cells in a tumor volume |
| $\eta$ | Exponent for cell proliferation |
| $\nu$ | Fraction of the total volume of an OAR or target |
| $L$ | Laplace transform |
| $\lfloor y \rfloor$ | Greatest integer less than or equal to $y$ |
| ${}_mC_x$ | Binomial coefficient |

# REFERENCES


Bertram, J.F., R.N. Douglas-Denton, B. Diouf, M.D. Hughson, and W.E. Hoy. 2011. Human nephron number: Implications for health and disease. *Pediatr. Nephrol*. 26(9), 1529–33.

Burman, C., G.J. Kutcher, B. Emami, and M. Goitein. 1991. Fitting of normal tissue tolerance to an analytic function. *Int. J. Radiat. Oncol. Biol. Phys.* 21, 123–135.

Dawson, L.A., B.D. Kavanagh, A.C. Paulino, S.K. Das, M. Mifton, X.A. Li, C. Pan, R.K. TenHaken, and T.E. Sdchultheiss. 2010. Radiation associated kidney injury. *Int. J. Radiat. Oncol. Biol. Phys*., 76(3, Suppl.), S108–15.

Deschavanne, P.J. and B. Fertil. 1996. *Int. J. Radiat. Oncol. Biol. Phys.* 34(1), 251–66.

Djajaputra, D. and Q. Wu. 2006. On relating the generalized equivalent uniform dose to the linear quadratic model. *Med. Phys*., 33(12), 4481.

Emami, B., J. Lyman, A. Brown, L. Coia, M. Goitein, J.E. Munzenrider, B. Shank, L.J. Solin, and M. Wesson. 1991. Tolerance of normal tissue to therapeutic irradiation. *Int. J. Radiat. Oncol. Biol. Phys.* 21, 109–22.

Goitein, M. 2008. *Radiation Oncology: A Physicist's-Eye View*. New York: Springer.

Hall, E.J. and A.J. Giaccia. 2012. *Radiobiology for the Radiologist*. Philadelphia: Lippincott, Williams and Wilkins.

Jackson, A., G.J. Kutcher, and E.D. Yorke. 1993. Probability of radiation-induced complications in normal tissues with parallel architecture under conditions of uniform whole or partial organ irradiation. *Med. Phys.* 20, 613–25.

Joiner, M. and A. van der Kogel (eds.). 2009. *Basic Clinical Radiobiology*. London: Hodder-Arnold.

Kallman, P.A., A. Agren, and A. Brahme. 1992. Tumour and normal tissue responses to fractionated non-uniform dose delivery. *Int. J. Radiat. Oncol. Biol. Phys.* 62, 249–62.

Kirkpatrick, J.P., A.J. van der Kogel, and T.E. Schultheiss. 2010. Radiation dose-volume effects in the spinal cord. *Int. J. Radiat. Oncol. Biol. Phys.* 76(3, Suppl.), S42–49.

Li, X.A., et al. 2012. The use and QA of biologically related models for treatment planning. AAPM TG 166. College Park, MD: American Association of Physicists in Medicine.
Lyman, J.T. 1985. Complication probability as assessed from dose volume histograms. *Radiat. Res.* 104, S13–19.
Marks, L.B., S.M. Bentzen, J.O. Deasy, F.M. Kong, J.D. Bradley, I.S. Vogelius, I. El Naga, et al. 2010a. Radiation dose volume effects in the lung. *Int. J. Radiat. Oncol. Biol. Phys.* 76(3, Suppl.), S70–76.
Marks, L.B., E.D. Yorke, A. Jackson, R.K. TenHaken, L.S. Costine, A. Eisbruch, S.M. Bentzen, J. Nam, and J.O. Deasy. 2010b. Use of normal tissue complication probability models in the clinic. *Int. J. Radiat. Oncol. Biol. Phys.* 76(3, Suppl.), S10–19.
Marks, L.B., R.K. Ten Haken, and M.K. Martel. 2010c. Guest editors introduction to QUANTEC: A users guide. *Int. J. Radiat. Oncol. Biol. Phys.* 76(3, Suppl.), S1–S2.
Metcalfe, P., T. Kron, and P. Hoban. 2007. *The Physics of Radiotherapy X-Rays and Electrons*, chap. 14. Madison, WI: Medical Physics Publishing.
Michalski, J.M., H. Gay, A. Jackson, S.L. Tucker, and J.O. Deasy. 2010. Radiation dose-volume effects in radiation induced rectal injury. *Int. J. Radiat. Oncol. Biol. Phys.* 76(3, Suppl.), S123–129.
Miralbell, R., et al. 2012. Dose-fractionation sensitivity of prostate cancer deduced from radiotherapy outcomes of 5969 patients in seven institutional datasets: $\alpha/\beta = 1.4$ (0.9-2.2) Gy. *Int. J. Radiat. Oncol. Biol. Phys.* 82(1), e17–24.
Moiseenko, V., J.O. Deasy, and J. Van Dyk. 2005. Radiobiological modeling for treatment planning. In *The Modern Technology of Radiation Oncology*, ed. J. Van Dyk, vol. II, 185–220. Madison, WI: Medical Physics Publishing.
Nahum, A.E. and Tait. 1992. Presentation at Proceedings of ART91, Munich, 1991. In *Advanced Radiation Therapy: Tumor Response Monitoring and Treatment Planning*, ed. A. Briet, 84. Berlin: Springer.
Niemierko, A. 1997. Reporting and analysing dose distributions: A concept of equivalent uniform dose. *Med. Phys.* 24(1), 103–10.
Ochs, M., J.R. Nyengaard, A. Jung, L. Knudsen, M. Voigt, T. Wahlers, J. Richter, and H.J.G. Gundersen. 2004. The number of alveoli in the human lung. 2004.. *Am. J. Crit. Care Med.* 169(1), 120–24.
Okunieff, P., D. Morgan, A. Niemierko, and H.D. Suit. 1995. Radiation dose-response of human tumors. *Int. J. Radiat. Oncol. Biol. Phys.* 32(4), 1227–37.
Pan, C.C., B.D. Kavanagh, L.A. Dawson, X.A. Li, S.K. Das, M. Miffen, and R.K. TenHaken. 2010. Radiation associated liver injury. *Int. J. Radiat. Oncol. Biol. Phys.* 76(3, Suppl.), S94–100.
Qi, X.S., J. White, and X.A. Li. 2011. Is $\alpha/\beta$ for breast cancer really low? *Radiother. Oncol.* 100(2), 282–88.
Schultheiss, T.E. 2008. The radiation dose-response of the human spinal cord. *Int. J. Radiat. Oncol. Biol. Phys.* 71(5), 1455–59.
Spence, A.P. and E.B. Mason. 1992. *Human Anatomy and Physiology*. Egan, MN: West Publishing.
Trott, K.R. and J. Kummermehr. 1985. What is known about tumor proliferation rates to choose between accelerated fractionation or hyperfractionation? *Radiother. Oncol.* 3(1), 1–9.
Viswanathan, A.N., E.D. Yorke, L.B. Marks, P.J. Eifel, and W.U. Shipley. 2010. Radiation dose-volume effects of the urinary bladder. *Int. J. Radiat. Oncol. Biol. Phys.* 76(3, Suppl.), S116–22.
Webb, S. 1993. *The Physics of Three Dimensional Radiation Therapy.* Bristol: Institute of Physics Publishing Egan, Minnesota.
Webb, S. 1997. *The Physics of Conformal Radiotherapy*, chap. 5. Bristol: Institute of Physics Publishing.

Wein, L.M., J.E. Cohen, and J.T. WU. 2000. Dynamic optimization of a linear-quadratic model with incomplete repair and volume-dependent sensitivity and repopulation. *Int. J. Radiat. Oncol. Biol. Phys.* 47, 1073.

Wigner, E. 1960. The unreasonable effectiveness of mathematics in the natural sciences. In *Communications in Pure and Applied Mathematics*, vol. 13, no. I. New York: John Wiley.

Yorke, E.D. 2003. Biological indices for evaluation and optimization of IMRT. In *Intensity Modulated Radiation Therapy: The State of the Art*, ed. J.R. Palta and T.R. Mackie. AAPM Medical Physics Monograph No. 29, 77–114. Madison, WI: Medical Physics Publishing.

Yorke, E.D., J.G. Mechalakos, and K.E. Rozenzweig. 2013. Dose-volume considerations: An update for use in treatment planning. In *The Modern Technology of Radiation Oncology*, ed. J. Van Dyk, vol. III, 59–90. Madison, WI: Medical Physics Publishing.

# ENDNOTES

1. There is an older definition of EUD (Niemierko, 1997) as the dose that "when distributed uniformly across the target volume causes the survival of the same number of clonogens" as the actual dose distribution.
2. This whole discussion presumes that complications are related to cell death only and not some type of cell injury or other predisposing factor.
3. This appears to be consistent with the data quoted in the paper by Kilpatrick et al. (2010). These authors used $\alpha/\beta = 0.87$ to calculate the 2 Gy equivalent total dose.
4. Parameters describing the goodness of fit are not quoted here because (a) the author is too lazy and (b) he doesn't want you to take this too seriously.
5. Dawson et al. (2010), however, indicate that unilateral kidney irradiation may not be without risk.
6. It is reported by Marks et al. (2010) that there is no observed threshold for radiation pneumonitis.
7. Many references refer to the quantity $a = 1/n$.
8. The symbol $k$ appearing in this formula should not be confused with the number of cells per FSU as used throughout the rest of this chapter.
9. The doubling time may decrease significantly after initial radiation treatment. This is known as accelerated repopulation (Zips, in Joiner and van der Kogel, 2009).
10. The user must make a decision whether to regard the volume as a GTV, CTV, or PTV.
11. The derivative of the step function is the delta function and $\sigma_{D_0} = \sigma_\alpha \bar{D}_0 / \langle \bar{\alpha} \rangle$; $N$ is assumed not to vary.
12. A more general result including the $\beta$ term may be found in Djajaputra and Wu (2006).

# APPENDIX: PROBLEM SOLUTIONS

## CHAPTER 1

1. Substitute the expression for $B_r$ from Equation 1.20b into Equation 1.19a:

$$\frac{1}{r}\frac{\partial E_z}{\partial \theta} \mp ikE_\theta = \pm\frac{\omega}{k}\left[\frac{\partial B_z}{\partial r} - \frac{i\omega}{c^2}E_\theta\right].$$

Collect terms involving $E_\theta$ on one side:

$$\pm ik\left[\frac{\omega^2}{k^2c^2} - 1\right]E_\theta = \pm\frac{\omega}{k}\frac{\partial B_z}{\partial r} - \frac{1}{r}\frac{\partial E_z}{\partial \theta}.$$

Recall that $\gamma^2 = \omega^2/c^2 - k^2$; therefore,

$$E_\theta = \frac{\pm ik}{\gamma^2}\left(\frac{1}{r}\frac{\partial E_z}{\partial \theta} \mp \frac{\omega}{k}\frac{\partial B_z}{\partial r}\right).$$

2. $k = 2\pi/3d$ and $c = \omega/k$; therefore, $d = c/3\nu = 3.3$ cm.
3. Substitution of Equation 1.67 into Equation 1.68 yields

$$\mathbf{H}_c = \frac{-i}{\sigma\omega\mu_c}\mathbf{n}\times\frac{\partial^2}{\partial\xi^2}(\mathbf{n}\times\mathbf{H}_c).$$

Using the BAC-CAB rule for the vector cross product leads to

$$i\sigma\omega\mu_c\mathbf{H}_c = \mathbf{n}\left(\mathbf{n}\cdot\frac{\partial^2\mathbf{H}_c}{\partial\xi^2}\right) - \frac{\partial^2\mathbf{H}_c}{\partial\xi^2}. \tag{A.1}$$

Now take the dot product of **n** with Equation A.1:

$$i\sigma\omega\mu_c\mathbf{n}\cdot\mathbf{H}_c = \mathbf{n}\cdot\mathbf{n}\left(\mathbf{n}\cdot\frac{\partial^2\mathbf{H}_c}{\partial\xi^2}\right) - \mathbf{n}\cdot\frac{\partial^2\mathbf{H}_c}{\partial\xi^2} = 0.$$

Thus, we arrive at Equation 1.69b: $\mathbf{n} \cdot \mathbf{H}_c = \mathbf{0}$. To derive Equation 1.69a, take the cross product of $\mathbf{n}$ with Equation A.1:

$$i\sigma\omega\mu_c \mathbf{n} \times \mathbf{H}_c = \mathbf{n} \times \mathbf{n}\left(\mathbf{n} \cdot \frac{\partial^2 \mathbf{H}_c}{\partial \xi^2}\right) - \frac{\partial^2}{\partial \xi^2}(\mathbf{n} \times \mathbf{H}_c).$$

The first term on the right-hand side of this equation is zero. Using the definition of the skin depth δ (Equation 1.70), we arrive at Equation 1.69a:

$$\frac{\partial^2}{\partial \xi^2}(\mathbf{n} \times \mathbf{H}_c) + \frac{2i}{\delta^2}(\mathbf{n} \times \mathbf{H}_c) \cong 0.$$

We can show that Equation 1.71 is a solution of the differential equation by substituting it in Equation 1.69a. First, evaluate

$$\frac{\partial^2}{\partial \xi^2}(\mathbf{n} \times \mathbf{H}_c) = (\mathbf{n} \times \mathbf{H}_t)\frac{\partial^2}{\partial \xi^2} e^{-\xi/\delta} e^{i\xi/\delta} = (\mathbf{n} \times \mathbf{H}_t)\left(\frac{1-i}{\delta}\right)^2 e^{-\xi/\delta} e^{i\xi/\delta} = -\frac{2i}{\delta^2}(\mathbf{n} \times \mathbf{H}_t).$$

If this is added to the second term on the left-hand side of Equation 1.69b, the sum is zero, and therefore the equation is satisfied.

4. First rearrange Equation 1.97 as follows: $E(z) = (E_0 + Ir_s)e^{-\alpha z} - Ir_s$. Now integrate this expression to get

$$V_e = \int_0^L E(z)\,dz = \frac{L}{\tau}(E_0 + Ir_s)\left(1 - e^{-\tau}\right) - Ir_s L.$$

Now substitute

$$E_0 = \sqrt{\frac{2\tau \langle P_0 \rangle r_s}{L}}.$$

and rearrange to obtain Equation 1.101.

5. The absence of beam loading means that $I = 0$. In this case,

$$V_e = \left(\frac{1 - e^{-\tau}}{\tau}\right) E_0 L.$$

If there are no wall losses, then $\tau \to 0$.
Taking the limit of the expression for $V_e$ (using L'Hospital's rule) gives $V_e = \mathrm{E}_0 L$.

6. a. The beam power is $IV_e = IV_0 - FI^2$. This is a maximum when $d(IV_e)/dI = 0 = V_0 - 2FI$ or $I = V_0/2F$, and thus $V_e = V_0/2$.

   b. The dose rate is given by $X = BIV_e^3 = BI\left(V_0 - FI\right)^3$. This will be a maximum when $dX/dI = 0$. Taking the derivative, setting it equal to zero, and solving for $I$ yields $I = V_0/4F$. Substituting this back into the equation for $V_e$ gives $V_e = 3/4\ V_0$.

7. a. The definition is

$$\alpha = \frac{-1}{2\langle P\rangle}\left\langle \frac{dP_w}{dz}\right\rangle.$$

   Using Equations 1.82 and 1.84, we get $\alpha^{-1} = \sigma\delta\mu_0 R v_g = 0.1\ \text{m}^{-1}$.

   b. $\tau = \alpha L = 0.2$

   c. $\langle P\rangle = \langle P_0\rangle e^{-2\alpha L} = 0.67\langle P_0\rangle$.

8. a. There are many ways to see this. In the absence of beam current,

$$V_0 = \sqrt{\left(1 - e^{-2\tau}\right)\langle P_0\rangle r_s L}.$$

   The definition of $r_s$ is $r_s = -E^2/(dP_w/dz)$. In the absence of beam loading, $dP_w/dz = dP/dz = -\langle P_0\rangle\ (1 - e^{-2\tau})/L$. Substituting this into the expression for $V_0$ gives $E_0 L$.

   b. For the constant-impedance guide in the absence of beam loading,

$$V_0 = \sqrt{2\tau}\,\frac{1 - e^{-\tau}}{\tau}\sqrt{\langle P_0\rangle r_s L}.$$

   We also have the relation $\langle P_0\rangle = E_0^2\ /\left(2\alpha r_s\right)$, where $\alpha = \tau/L$. Solve this for $\langle P_0\rangle$ $r_s$ and substitute into the expression for $V_0$ to obtain

$$V_0 = \frac{1 - e^{-\tau}}{\tau} E_0 L.$$

   The value of $\tau \approx 0.4$, and therefore $V_0 = 0.82\ E_0 L$. The unloaded beam energy is about 80% of that for a constant-gradient guide.

9. $Ir_e L/(2V_0) = 0.42$. The assumption that $Ir_e L/(2V_0) \ll 1$ is therefore dubious.

10. The shunt impedance is given by

$$r_s = \frac{\sigma\delta c^2\mu_0^2}{\pi J_1^2\left(x_{01}\right)}\,\frac{l/R}{R\left(1 + l/R\right)}.$$

The skin depth is given by

$$\delta = \left( \frac{2}{\mu\omega\sigma} \right)^{\frac{1}{2}}.$$

For a frequency of 11.4 GHz, the skin depth is $6.21 \times 10^{-7}$ m. $R = cx_{01}/\omega = 1.01$ cm and $l/R = \pi/x_{01} = 1.31$ (for $w = \pi/2$); therefore, $r_s = 340$ MΩ/m.

# CHAPTER 2

1. The kinetic energy of a relativistic particle is $T = mc^2 - m_0 c^2 = (\gamma - 1)m_0 c^2$, where $\gamma = (1 - v^2/c^2)^{-1/2}$,

$$\gamma = \frac{T}{m_0 c^2} + 1 = \frac{250}{940} + 1 = 1.26. \left( \frac{v}{c} \right)^2 = 1 - \frac{1}{\gamma^2} = 0.373,$$

and therefore $v/c = 0.6$.

2. The de Broglie wavelength is $\lambda = h/p$, where $h$ is Planck's constant and $p$ is the momentum of the particle. The momentum is $p = \sqrt{2mT}$, where $T$ is the kinetic energy. The de Broglie wavelength is therefore $\lambda = 2 \times 10^{-15}$ m, about the same as the range of the strong force. It is expected that the strong force will begin to play an important role at this energy.

3. Refer to Equation 2.6. This equation can be turned into an equation for the dose rate by putting dots over $D$ and $\Phi$:

$$\dot{D}(x) = 1.6 \times 10^{-10} \dot{\Phi}(x) \left( \frac{dT}{\rho dx} \right) \left[ \text{MeV cm}^2 \text{ g}^{-1} \right].$$

The fluence rate is given by the number of protons/s/area = $i/(e \times \text{area}) = 1.3 \times 10^9$ cm$^{-2}$ s$^{-1}$, where $i$ is the beam current and $e$ is the charge on the proton. Substitution into the expression for $\dot{D}$ yields 4.2 Gy/s = 2500 cGy/min.

4. a. The total distance traveled is roughly $2\pi RN$, where $N$ is the number of round-trips. If we assume that the particles are accelerated twice on each round-trip,

$$N = 2.5 \times \frac{10^8}{2} \times 10^5 = 1.25 \times 10^3.$$

The time required is $= 2\pi RN/v$. The deflection is therefore $\Delta z = 2\pi RN(v_z/v)$ and

$$\frac{v_z}{v} = \frac{\Delta z}{2\pi NR} = 2.5 \times 10^{-8}!$$

b. Focusing provided by the sectors prevents large excursions in the vertical direction.

5. a. The kinetic energy is $T = mc^2 - m_0c^2 = (\gamma - 1)m_0 c^2$. From the equation of motion (2.15),

$$v = \dot{\theta}R = \frac{qB_0R}{\gamma m_0}.$$

and

$$\gamma = \frac{1}{\sqrt{1-(v/c)^2}}.$$

Solving the equation of motion for $v$ and substituting this into the expression for $\gamma$ yields

$$\gamma = \sqrt{1+\left(\frac{qB_0R}{m_0c}\right)^2}.$$

and thus

$$T = m_0c^2\left[\sqrt{1+\left(\frac{qB_0R}{m_0c}\right)^2} - 1\right].$$

b. Substitution of the parameters given into the equation above yields $T = 264$ MeV.

6. Derivation of the perturbation equations. Substitute the perturbation variables (Equations 2.28 through 2.30) into the equations of motion (2.15 through 2.17). Start with Equation 2.15 for the radial acceleration:

$$\frac{m}{q}\left[\ddot{r}' - (r' + r_0)\left(\dot{\theta}' - \omega_0\right)^2\right] = (r' + r_0)\left(\dot{\theta}' - \omega_0\right)\left(B_0 + b_1r' - \frac{b_1}{2r_0}z'^2\right).$$

We only retain quantities that are linear in the perturbation variables $r'$, $\theta'$, $z'$, and $b_1$. This leaves

$$\frac{m}{q}\left[\ddot{r}' - r'\omega_0^2 + 2r_0\dot{\theta}'\omega_0 - r_0\omega_0^2\right] = -\omega_0r'B_0 - \omega_0r_0B_0 + r_0\dot{\theta}'B_0 - \omega_0r_0b_1r'.$$

The "zeroth-order" terms on both sides of this equation cancel because they are equal. Likewise for the $r'$ terms. Combining remaining terms leads to $\ddot{r}' + \omega_0 r_0 \dot{\theta}' + n\omega_0^2 r' = 0$.
We now perturb Equation 2.16:

$$\frac{m}{q}\left(2\dot{r}\dot{\theta} + r\ddot{\theta}\right) = \dot{z}B_r - \dot{r}B_z$$

$$\frac{m}{q}\left[2\dot{r}'\left(\dot{\theta}' - \omega_0\right) + \left(r' + r_0\right)\ddot{\theta}'\right] = \dot{z}'b_1 z' - \dot{r}'\left(B_0 + b_1 r'\right).$$

Discarding all quantities that are not linear in the perturbation variables yields $r_0\ddot{\theta}' - \dot{r}'\omega_0 = 0$.
Finally, we derive Equation 2.33 by perturbing Equation 2.17:

$$\frac{m}{q}\ddot{z}' = -\left(r + r_0\right)\left(\dot{\theta}' - \omega_0\right)\frac{nz'B_0}{r_0}.$$

Neglecting all terms that are not linear in the perturbation variables yields $\ddot{z}' + n\omega_0^2 z' = 0$.

7. $\dfrac{\Omega_z}{\omega_0} = \sqrt{n} = 0.45.$

8. a. $n = -\dfrac{r}{B_z}\dfrac{\partial B_z}{\partial r} \sim -\dfrac{R}{B_0}\dfrac{\Delta B}{\Delta R} \sim -\dfrac{\Delta B}{B} \approx \dfrac{0.7}{5.0} = 0.14.$

b. $\omega_0 = \dfrac{qc^2 B_0}{E}$ and $\nu = \dfrac{\omega_0}{2\pi}$.

$E$ = 940 MeV + 230 MeV, and therefore ν = 61 MHz.

9. $W = 13\dfrac{\text{mSv}}{\text{h}\cdot\text{nA}} \times 10\ \text{nA} \times 500\dfrac{\text{h}}{\text{yr}} = 6.5\times10^4\ \dfrac{\text{mSv}}{\text{yr}},$

#TVL = 2.2/0.65 = 3.39, $B = 10^{-\#TVL} = 4.13 \times 10^{-4}$, $d$ = 6.5 m, and $H$ = 0.16 mSv. This is below the NRC annual limit for members of the public by about a factor of 6.

# CHAPTER 3

3. a. For $x' < x$, $|x - x'| = x - x'$, and for $x' > x$, $|x - x'| = x' - x$; therefore, the dose is

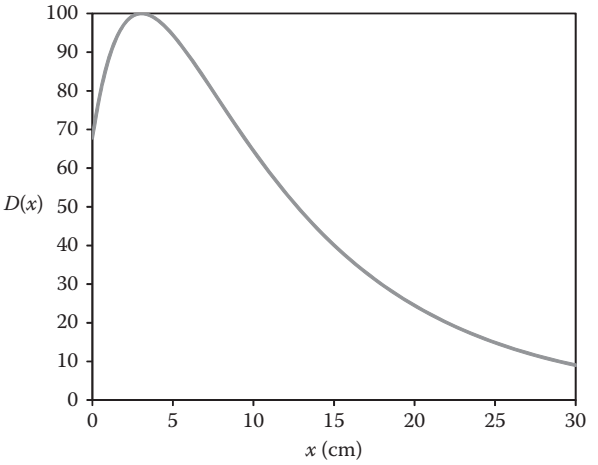
$$D(x) = \int_0^\infty T(x')A(x - x')dx' = A_0T_0\left[\int_0^x e^{-\mu_0 x'}e^{-\mu(x-x')}dx' + \int_x^\infty e^{-\mu_0 x'}e^{-\mu(x'-x)}dx'\right]$$

$$= \frac{A_0T}{\mu - \mu_0}\left[\frac{2}{1 + \mu_0/\mu}e^{-\mu_0 x} - e^{-\mu x}\right].$$

b. Take the derivative and set it equal to 0 and solve for $x$:

$$x_{\max} = \frac{1}{\mu_0 - \mu} \ln\left[\frac{2\mu_0/\mu}{1+\mu_0/\mu}\right].$$

c.

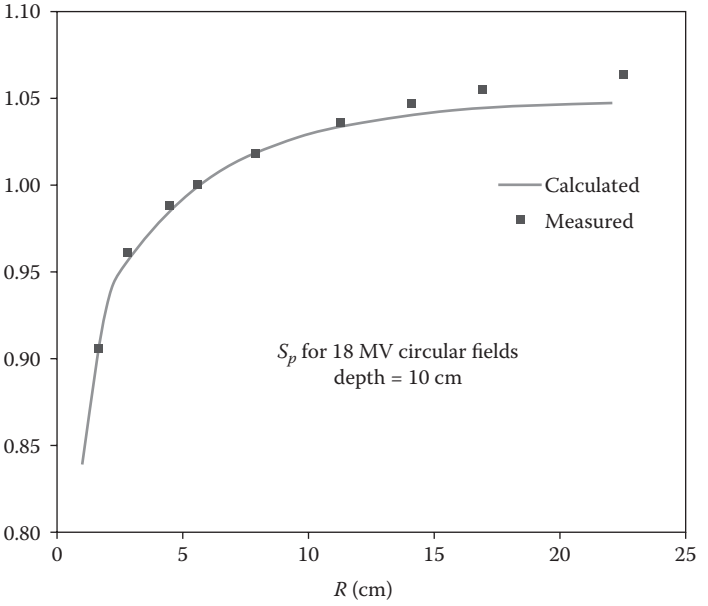



8. a. The dose for a circular field of radius $R$ for the pencil beam of the form of Equation 3.60 is given by Equation 3.58:

$$D(x,y,z) = \int\int_{\text{area}} \Psi(x',y',z_0) P_z(x-x',y-y')\,dx'dy'$$

$$= 2\pi\Psi\int_0^R P_z r_c\,dr_c = 2\pi\left[\frac{A_z}{a_z}\left(1-e^{-a_z R}\right)+\frac{B_z}{b_z}\left(1-e^{-b_z R}\right)\right]. \qquad \text{(A.2)}$$

b.



9. The average energy is given by Equation 3.15. The integrals can be evaluated on a spreadsheet using the trapezoidal rule:

$$\int_a^b f(x)dx \approx \sum_{i=1}^{N} \frac{1}{2}\left[x_{i+1} - x_i\right]\left[f(x_i) + f(x_{i+1})\right],$$

where $x_1 = a$ and $x_{N+1} = b$.

**Nominal Accelerating Potential and Average Energy**

| NAP (MV) | $\overline{E}$ (MeV) | NAP/3 | $\overline{E}$/(NAP/3) |
|---|---|---|---|
| 6 | 2.4 | 2.0 | 1.2 |
| 10 | 3.3 | 3.3 | 1.0 |
| 18 | 5.2 | 6.0 | 0.9 |

The rule of thumb that the average energy is equal to NAP/3 appears to be correct to within about 20%.

12. Computation time $\propto MN^3$ and $N \rightarrow (2N)$; therefore, computation time goes up by a factor of 8.
13. Computation time $\propto MN^3$ and $N \rightarrow (3N/2)$; therefore, computation time goes up by a factor of $(3/2)^3 \approx 3.5$.

# CHAPTER 4

1. Use the divergence theorem to convert the surface integral in Equation 4.26 to

$$\oint dS \cdot \hat{\Omega}\Phi_{\Omega,E} = \int \nabla \cdot \left(\hat{\Omega}\Phi_{\Omega,E}\right) d^3r.$$

If the vector $\hat{\Omega}$ is independent of the position coordinates, then

$$\nabla \cdot \left(\hat{\Omega}\Phi_{\Omega,E}\right) = \hat{\Omega} \cdot \vec{\nabla}\Phi_{\Omega,E}.$$

All the terms in Equation 4.26 may be collected on one side of the equation as

$$\int \left\{ \int EdE \int d\Omega \left(\hat{\Omega} \cdot \vec{\nabla}\Phi_{\Omega,E} - S_0\right) + \frac{d\overline{\varepsilon}}{dV} - \frac{dQ}{dV} \right\} d^3r = 0.$$

As this equation must be satisfied for all volumes, the integrand itself must equal zero. The term

$$\frac{d\overline{\varepsilon}}{dV} = \rho D(\vec{r}),$$

and therefore we may write

$$D(\vec{r}) = \frac{1}{\rho}\int dE \int d\Omega E\left[S_0\left(\vec{r},E,\hat{\Omega}\right) - \hat{\Omega}\cdot\vec{\nabla}\Phi_{\Omega,E}\left(\vec{r},E,\hat{\Omega}\right)\right] + \frac{dQ}{dm}.$$

2. Integration over angle φ inserts a factor of 2π into Equation 4.50. If $\sin\theta' \approx \theta'$, then $\sin^{-1}(R/r) \approx R/r$ and the integral for θ′ may be written as

$$\int_0^{(R/r)} \theta' \sqrt{1 - \left(\frac{r}{R}\right)^2 \theta'^2}\, d\theta'.$$

This integral is of the form $u^n\, du$, where $u$ is the term inside the square root. The integral is equal to

$$\frac{1}{3}\left(\frac{R}{r}\right)^2,$$

and thus

$$\Phi_E = \frac{N(E)}{4\pi r^2},$$

as in Equation 4.51.

3. The source term is the total number of particles emitted per unit volume per solid angle. For the disk,

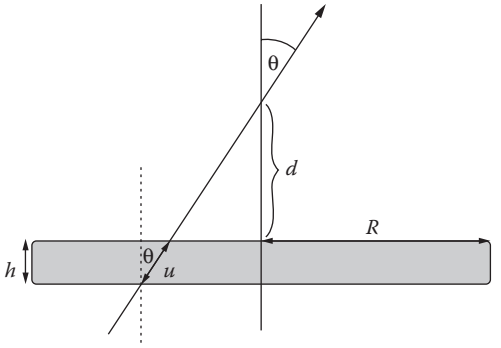
$$S_0 = \frac{N(E)}{4\pi^2 R^2 h}.$$

Equation 4.47 may be used to calculate the differential fluence. In this case, $\Phi_{\Omega,E} = S_0 u$, where $u$ is the distance shown in the figure below: $u = h/\cos\theta$. We ignore edge effects near the edge of the disk. The differential fluence is therefore $\Phi_{\Omega,E} = S_0 h \sec\theta$. The differential energy fluence is

$$\Phi_E = \int_{4\pi} \Phi_{E,\Omega}\, d\Omega = 2\pi S_0 h \int_0^{\theta_{max}} \tan\theta\, d\theta,$$

where $\tan\theta_{max} = R/d$. Carrying out the integral, we find $\Phi_E = \pi S_0 h \ln[1 + (R/d)^2]$. When $(R/d) \ll 1$, $\ln[1 + (R/d)^2] \approx (R/d)^2$, and therefore

$$\Phi_E = \frac{N(E)}{4\pi d^2}.$$




4. a. We wish to show that

$$\hat{\Omega}\cdot\vec{\nabla}\Omega_{\Omega,E}=\cos\theta'\frac{\partial\Phi_{\Omega,E}}{\partial r}-\frac{\sin\theta'}{t}\frac{\partial\Phi_{\Omega,E}}{\partial\theta'}=0.$$

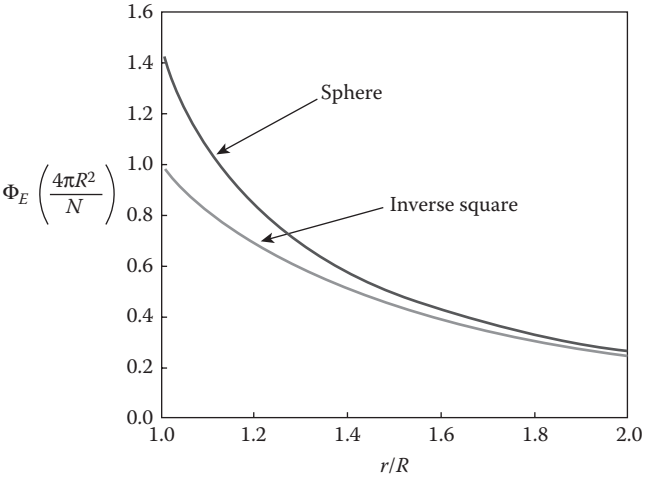
$$\frac{\partial\Phi}{\partial\theta'}=-\frac{3N}{8\pi R^2}\left[1-\left(r/R\right)^2\sin^2\theta'\right]^{-1/2}\left[\left(r/R\right)^2\sin\theta'\cos\theta'\right] \tag{A.2}$$

and

$$\frac{\partial\Phi}{\partial r}=-\frac{3N}{8\pi R^2}\left[1-\left(r/R\right)^2\sin^2\theta'\right]^{-1/2}\left[\left(r/R^2\right)\sin^2\theta'\right].$$

Substitution into Equation A.2 shows that this equation is satisfied.

b.



5. Substitute Equations 4.59 and 4.58 into the right-hand side of the BTE (Equation 4.34):

$$\int dE' \int_{4\pi} d\Omega' \mu_s \Phi_{\Omega',E'}$$

$$= \int dE' \int d\Omega' \left[ \frac{1}{\Delta x} \delta\left( E' - E - \frac{dE}{dx}\Delta x \right) \delta\left( \hat{\Omega}' - \hat{\Omega} \right) + \mu_s\left( \hat{\Omega}' \to \hat{\Omega} \right) \delta\left( E' - E \right) \right] \Phi_{\Omega',E'}$$

$$= \int d\Omega' \left[ \frac{1}{\Delta x} \delta\left( \hat{\Omega}' - \hat{\Omega} \right) \Phi_{\Omega',E+\frac{dE}{dx}\Delta x} + \mu_s\left( \hat{\Omega}' \to \hat{\Omega} \right) \Phi_{\Omega',E} \right].$$

Now,

$$\Phi_{\Omega',E+\frac{dE}{dx}\Delta x} \approx \Phi_{\Omega',E} + \frac{\partial \Phi_{\Omega',E}}{\partial E} \frac{dE}{dx} \Delta x,$$

and therefore

$$\int dE' \int_{4\pi} d\Omega' \mu_s \Phi_{\Omega',E'} = \frac{1}{\Delta x} \Phi_{\Omega,E} + \frac{dE}{dx} \frac{\partial \Phi_{\Omega,E}}{\partial E} + \int d\Omega' \mu_s\left( \hat{\Omega}' \to \hat{\Omega} \right) \Phi_{\Omega',E}.$$

6. First, compute the derivatives of some of the terms that will be needed later:

$$\frac{dA_n}{dz} = nA_{n-1}$$

and

$$\frac{dA_0}{dz} = \frac{1}{4} T$$

from Liebnitz's rule:

$$\frac{dB}{dz} = \frac{1}{4} A_2 T.$$

From these derivatives, it follows that

$$\theta_x \frac{\partial \Phi}{\partial x} = \frac{1}{8\pi B^{3/2}} \left[ A_1 \theta_x^2 - A_0 x \theta_x \right] \exp\{\ \},$$

where exp{ } is shorthand for

$$\exp\left\{ \frac{-A_2 \theta_x^2 + 2A_1 x \theta_x - A_0 x^2}{4B} \right\}.$$

The right-hand side of Equation 4.88 is

$$\frac{1}{4}T\frac{\partial^2\Phi}{\partial\theta_x^2}=\frac{1}{8\pi B^{3/2}}\cdot\frac{1}{4}T\left[-A_2+\frac{1}{2B}\left(A_1x-A_2\theta_x\right)^2\right]\exp\{\ \}.$$

The term

$$\frac{\partial\Phi}{\partial z}=\frac{1}{8\pi B^{3/2}}$$
$$\times\left\{-\frac{1}{4}A_2T+\frac{1}{2B}\left[\frac{1}{4}A_2T\left(A_2\theta_x^2-2A_1x\theta_x+x^2\frac{A_1^2}{A_2}\right)+B\left(2x\theta_xA_0-2\theta_x^2A_1\right)\right]\right\}\exp\{\ \}.$$

Adding the two terms on the left-hand side of Equation 4.88 yields the expression for

$$\frac{1}{4}T\frac{\partial^2\Phi}{\partial\theta_x^2}$$

given above.

7. Equations 4.93 define $\sigma_{\theta_x}^2$ and $\theta_p$. When $T = T_0$ is a constant,

$$A_n=\frac{1}{4}\frac{T_0}{n+1}z^{n+1}.$$

In this case, Equation 4.90 gives

$$B=\frac{1}{12\cdot16}T_0^2z^4.$$

This leads to

$$\sigma_{\theta_x}^2=\frac{1}{8}T_0z$$

and

$$\theta_p=\frac{3x}{2z}.$$

8. When $T_0$ is a constant,

$$A_n=\frac{1}{4}\frac{T_0}{n+1}z^{n+1}.$$

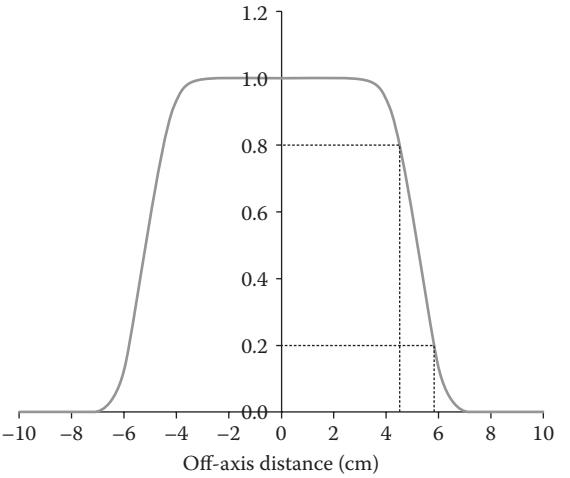
Under these circumstances,

$$B = \frac{1}{16} \cdot \frac{1}{12} T_0^2 z^4,$$

and therefore

$$\Phi(z, x, \theta_x) = \frac{2\sqrt{3}}{\pi T z^2} \exp\left\{ -\frac{4}{Tz} \left( 3\left(\frac{x}{z}\right)^3 - 3\left(\frac{x}{z}\right)\theta_x + \theta_x^2 \right) \right\}.$$

9.




The penumbra width is approximately 1.3 cm.

10. a. Use Equation 4.102 in polar coordinates to calculate the dose:

$$D(0,0,z) = \frac{S_\rho(z)\Phi_T(0)}{4\pi A_2} \int_0^R e^{-\frac{r^2}{4A_2}} 2\pi r dr = S_\rho(z)\Phi_T(0)\left[ 1 - e^{-\frac{R^2}{4A_2(z)}} \right].$$

b. For small values of $z$, $A_2(z) \ll 1$, and the dose will therefore be constant independent of the field size. For large values of $z$, the dose will depend on the field size, and it will decline toward zero as $z \rightarrow$ infinity. There is no absorption built into the Fermi–Eyges theory, and therefore the dose declines with increasing depth at large depths because of out-scatter. The beam becomes more dilute. For real electron beams, the dose rises to a maximum and then declines to approximately zero at a depth of $z = R_p$. Above a field size of about $R = 5$ cm, the depth dose does not depend on field size.

# CHAPTER 5

1. The probability of $x$ successes in $m$ trials is $f(x) = {}_mC_x\, p^x q^{m-x}$.
   a. In this case, we need one success (actually a failure) in five trials (links); therefore, $f(1) = {}_5C_1\, p^1\, q^4 = 5(0.01)^1(0.99)^4 = 0.048$. Thus, there is a 5% chance the chain will break.
   b. For 100 links, $f(1) = {}_{100}C_1\, p^1 q^{99} = 0.370$. For 100 links, there is a 37% chance the chain will break.
   c. Using the Poisson distribution,

$$f(x) = \frac{\mu^x e^{-n}}{x!}.$$

   and $\mu = mp = 1.0$; therefore, $f(1) = e^{-1} = 0.368$, close to the answer for part (b), as expected.
2. In this case, we need the probability that two or more strands break. The word *or* is a tip-off that we need to add probabilities. We need to add the probabilities that two break, that three break, that four break, and that five break, that is,

$$\sum_{x=2}^{5} {}_5C_x p^x q^{5-x} = 0.205 + 0.051 + 0.006 + 0.0003 = 0.262,$$

   or 26%, or more easily, calculate the probability that one or no strands break (0.738) and subtract from 1.00.
3. $\bar{D} = \dfrac{1}{V}\displaystyle\int_0^{D_{\max}} Dh(D)\,dD$

   and

$$h(D) = -V\frac{dH}{dD};$$

   therefore,

$$\bar{D} = -\int_0^{D_{\max}} D\frac{dH}{dD}\,dD.$$

   Integration by parts yields

$$-\left\{(DH(D)\Big|_0^{D_{\max}} - \int_0^{D_{\max}} H(D)\,dD\right\} = \int_0^{D_{\max}} H(D)\,dD,$$

   as $H(D_{\max}) = 0$.

4. The volume having a dose between $D$ and $D + dD$ is $dV = h(D)dD$. The volume of a spherical shell is $dV = 4\pi r^2 dr$. Equating the expressions for $dV$, we get

$$\frac{dD}{dr} = -\frac{4\pi r^2}{h},$$

where the minus sign indicates that dose is decreasing as $r$ increases. Integration of this expression for $D$ yields

$$D(r) = D_{\max} - (D_{\max} - D_{\min})\left(\frac{r}{R}\right)^3.$$

5. The

$$\mathrm{EUD}^{1/n} = \frac{1}{N}\sum_{i=1}^{N} D_i^{1/n}$$

is the average value of $D^{1/n}$. If $D$ varies continuously, then the average value of $D^{1/n}$ is

$$\left(\overline{D^{1/n}}\right) = \frac{1}{V}\int_0^{D_{\max}} h(D)D^{1/n}dD$$

Substitute the expression

$$h(D) = -V\frac{dH}{dD}$$

into this integral, and then integrate by parts to obtain

$$\left(\overline{D^{1/n}}\right) = -\left(D^{1/n}H\right)\Big|_0^{D_{\max}} + \frac{1}{n}\int_0^{D_{\max}} H(D)D^{\frac{1}{n}-1}dD,$$

and therefore

$$\mathrm{EUD}^{1/n} = \frac{1}{n}\int_0^{D_{\max}} H(D)D^{\frac{1}{n}-1}dD.$$

6. $\mathrm{EUD} = \left[ \dfrac{D_{\max}^{1/n+1} - D_{\min}^{1/n-1}}{(1+1/n)\Delta D} \right]$;

when $n$ becomes very small,

$$\mathrm{EUD} \to \left[ \frac{D_{\max}^{1/n} - D_{\min}^{1/n}}{\Delta D / n} \right]^n \to \left[ \frac{nD_{\max}^{1/n}}{\Delta D} \right]^n \to D_{\max} \left[ \frac{n}{\Delta D} \right]^n$$

Now take the log of both sides:

$$\ln(\mathrm{EUD}) = \ln D_{\max} + n \ln \left[ \frac{n}{\Delta D} \right].$$

Write the second term as

$$\frac{\ln(n / \Delta D)}{n^{-1}}$$

and use L'Hospital's rule:

$$\lim_{x \to c} \frac{f(x)}{g(x)} = \lim_{x \to c} \frac{f'(x)}{g'(x)}$$

to show that the second term above goes to zero, and therefore

$$\lim_{n \to 0} \mathrm{EUD} = D_{\max}.$$

7. We are asked to find the dose delivered to half the liver, which will lead to the same NTCP as a uniform dose of 30 Gy. We use the expression for the EUD:

$$\mathrm{EUD} = \left[ \frac{1}{N} \sum_{i=1}^{i=N} D_i^{1/n} \right]^n.$$

If half the OAR receives a uniform dose $D$ and the rest receives a dose of zero, then

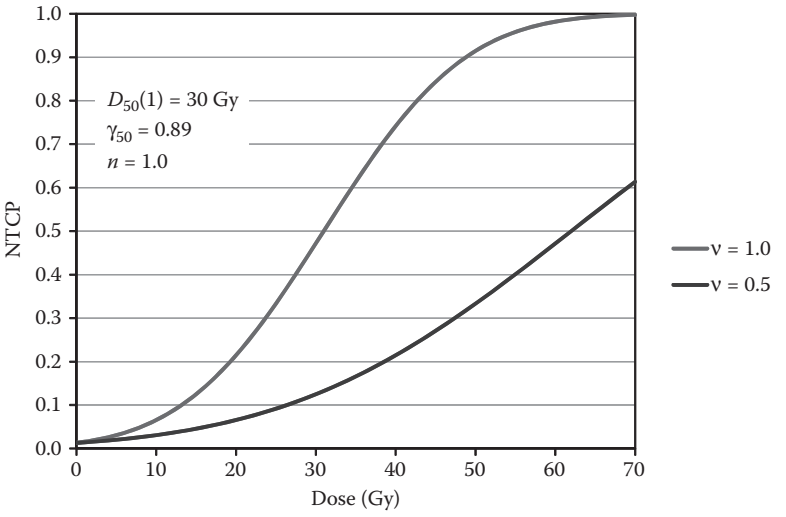
$$\mathrm{EUD} = \left[ \frac{1}{N} \times \frac{N}{2} D^{1/n} \right]^n = \left( \frac{1}{2} \right)^n D,$$

and therefore $D = (\text{EUD}) \times 2^n$.

a. For liver, $D = 54$ Gy.

b. For kidneys, $D = 24$ Gy.

8.




9. Solve Equation 5.37 for $\alpha + \beta d$ and substitute this into the equation for NTCP (Equation 5.34). At the same time, use the approximation that

$$1 - 2^{-1/N} \approx \frac{\ln 2}{N}.$$

This results in

$$\text{NTCP} \approx 1 - \left\{ 1 - \left( 1 - \left[ 1 - \left( \frac{\ln 2}{N} \right)^{1/k} \right]^{D/D_{50}} \right)^k \right\}^N.$$

Continue by expanding the brackets from the innermost outward:

$$\left[ 1 - \left( \frac{\ln 2}{N} \right)^{1/k} \right]^{D/D_{50}} \approx 1 - \frac{D}{D_{50}} \left( \frac{\ln 2}{N} \right)^{1/k},$$

and therefore

$$\left( 1 - \left[ 1 - \left( \frac{\ln 2}{N} \right)^{1/k} \right]^{D/D_{50}} \right)^k \approx \left( \frac{D}{D_{50}} \right)^k \frac{\ln 2}{N},$$

which leads to

$$\text{NTCP} \approx 1 - \left\{1 - \left(\frac{D}{D_{50}}\right)^k \frac{\ln 2}{N}\right\}^N.$$

We are now faced with the problem of finding a limit of the form

$$\lim_{x\to\infty}\left(1 - \frac{a}{x}\right)^x.$$

This limit looks somewhat like

$$\frac{1}{e} = \lim_{n\to\infty}\left(1 - \frac{1}{n}\right)^n.$$

We leave it to the reader's ingenuity to show that

$$\lim_{x\to\infty}\left(1 - \frac{a}{x}\right)^x = e^{-a}.$$

This leads to the final result:

$$\text{NTCP} \approx 1 - 2^{-\left(\frac{D}{D_{50}}\right)^k}.$$

10. Let

$$y = \frac{x - \mu}{\sigma\sqrt{2}},$$

then

$$dy = \frac{1}{\sigma\sqrt{2}}\,dx$$

and

$$\int_0^L e^{-(x-\mu)^2/2\sigma^2}\,dx = \sigma\sqrt{2}\int_{\frac{\mu}{\sigma\sqrt{2}}}^{\frac{L-\mu}{\sigma\sqrt{2}}} e^{-y^2}\,dy.$$

This integral can be written as

$$\int_{-\frac{\mu}{\sigma\sqrt{2}}}^{0} e^{-y^2}\,dy + \int_{0}^{\frac{L-\mu}{\sigma\sqrt{2}}} e^{-y^2}\,dy.$$

Due to the fact that $e^{-y^2}$ is an even function, the limits of the first integral can be changed to 0 to $+\mu/(\sigma\sqrt{2})$.
Recall that the error function is defined as

$$\text{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-y^2}\,dy.$$

In terms of the error function, we have

$$\int_0^L e^{-(x-\mu)^2/2\sigma^2}\,dx = \sigma\frac{\sqrt{2\pi}}{2}\left\{\text{erf}\left(\frac{\mu}{\sigma\sqrt{2}}\right) + \text{erf}\left(\frac{L-\mu}{\sigma\sqrt{2}}\right)\right\}.$$

We presume that $N_0P \gg 1$, and therefore the argument of the first erf term in the previous integral is much larger than 1. Hence, the value of the error function is approximately 1. Substitution into Equation 5.53 leads to

$$\text{NTCP} \approx \frac{1}{2}\left[1 + \text{erf}\left(\frac{\mu - L}{\sigma\sqrt{2}}\right)\right].$$

11. For $D < D_{\min}$, $H(D) = 1.0$; for $D_{\min} < D < D_{\max}$, $H(D) = 1 - (D - D_{\min})/\Delta D$; and for $D > D_{\max}$, $H(D) = 0$. The integral

$$\int_{D_{\min}}^{D_{\max}} H(D)e^{-\alpha D}\,dD = \frac{e^{-\alpha D_{\min}}}{\alpha^2 \Delta D}\left[e^{-\alpha\Delta D} + \Delta D - 1\right],$$

and thus

$$\frac{\text{TCP}}{\text{TCP}_0} = \exp\left\{\frac{N}{\alpha\Delta D} e^{-\alpha D_{\min}}\left[e^{-\alpha\Delta D} + \alpha\Delta D - 1\right]\right\}.$$

For $D_{\min} = 60$ Gy, $\Delta D = 12$ Gy, $\alpha = 0.1$ Gy$^{-1}$, and $N = 100$, $\text{TCP}/\text{TCP}_0 = 1.11$, a modest enhancement. This is in part because TCP is already high (78%).

12. a. The fraction of the cells surviving is $S = \exp(-D(\alpha + \beta d))$. For $D = 22$ Gy, $d = 2$ Gy, $\alpha = 0.015$ Gy$^{-1}$, and $\alpha/\beta = 3.0$, $S = 0.577$.
    b. The probability of killing all cells in an FSU is $P = (1 - S)^k = 3.37 \times 10^{-8}$ for $k = 20$.
    c. The probability of a complication is NTCP $= 1 - (1 - P)^N$. For $N = 10^7$, NTCP = 29%.
13. The original $TCP_0 = \text{Exp}[-N_0 e^{-D(\alpha+\beta d)}]$; the new TCP is given by $TCP' = \text{Exp}[-2N_0 e^{-D(\alpha+\beta d)}]$ and $TCP'/TCP_0 = \text{Exp}[-N_0 e^{-D(\alpha+\beta d)}] = TCP_0$; therefore, $TCP' = TCP_0^2$. If $TCP_0 = 0.95$, then $TCP' = (0.95)^2 = 0.90$.

# INDEX