The physics of proton therapy

To cite this article: Wayne D Newhauser and Rui Zhang 2015 Phys. Med. Biol. 60 R155

View the article online for updates and enhancements.

Related content

- Nuclear physics in particle therapy: a review
  Marco Durante and Harald Paganetti

- Neutron production from beam-modifying devices
  Angélica Pérez-Andújar, Wayne D Newhauser and Paul M DeLuca Jr

- Estimated risk from stray radiation for a pediatric patient receiving proton CSI
  Phillip J Tadder, Dragan Mirkovic, Jonas D Fontenot et al.

Recent citations

- Characterization of a multilayer ionization chamber prototype for fast verification of relative depth ionization curves and spread-out-Bragg-peaks in light ion beam therapy
  Alfredo Mirandola et al

- Intensity-modulated radiotherapy for whole pelvis irradiation in prostate cancer: A dosimetric and plan robustness study between photons and protons
  Ashley L.K. Ong et al

- Impact of irradiation setup in proton spot scanning brain therapy on organ doses from secondary radiation
  Oscar Ardenfors et al
Topical Review

The physics of proton therapy

Wayne D Newhauser\textsuperscript{1,2} and Rui Zhang\textsuperscript{1,2}

\textsuperscript{1} Medical Physics Program, Department of Physics and Astronomy, Louisiana State University, 202 Nicholson Hall, Baton Rouge, LA 70803, USA
\textsuperscript{2} Mary Bird Perkins Cancer Center, 4950 Essen Lane, Baton Rouge, LA 70809, USA

E-mail: newhauser@lsu.edu

Received 5 February 2014, revised 29 September 2014
Accepted for publication 28 October 2014
Published 24 March 2015

Abstract

The physics of proton therapy has advanced considerably since it was proposed in 1946. Today analytical equations and numerical simulation methods are available to predict and characterize many aspects of proton therapy. This article reviews the basic aspects of the physics of proton therapy, including proton interaction mechanisms, proton transport calculations, the determination of dose from therapeutic and stray radiations, and shielding design. The article discusses underlying processes as well as selected practical experimental and theoretical methods. We conclude by briefly speculating on possible future areas of research of relevance to the physics of proton therapy.

Keywords: proton therapy, interaction mechanism, transport, Monte Carlo, dosimetry, shielding

(Some figures may appear in colour only in the online journal)
1. Introduction

The history of proton therapy began in 1946 when Robert Wilson published a seminal paper in which he proposed to use accelerator-produced beams of protons to treat deep-seated tumors in humans (Wilson 1946). In that paper, he explained the biophysical rationale for proton therapy as well as the key engineering techniques of beam delivery. In 1954, the first human was treated with proton beams at the Lawrence Berkeley Laboratory (Lawrence et al. 1958). In 1962, specialized radiosurgical proton treatments commenced at the Harvard Cyclotron Laboratory (Kjellberg et al. 1962a, 1962b), followed in the mid 1970s by treatments for ocular cancers (Gragoudas et al. 1982) and larger tumors (Koehler et al. 1977). Physicists at Harvard, collaborating with clinical colleagues at Massachusetts General Hospital, the Massachusetts Eye and Ear Infirmary, and elsewhere, developed much of the physics and technology needed to treat patients with proton beams safely and effectively. Remarkably, the research and development program at Harvard continued for more than 40 years (Wilson 2004). During the same period, physicists elsewhere were developing other key technologies, including accelerators, magnetically scanned beams, treatment planning systems, computed tomographic imaging (CT), and magnetic resonance imaging.

The widespread adoption of proton therapy has been slow in comparison to, for example, intensity-modulated photon therapy. There are several reasons for this slow adoption of proton therapy, including technical difficulty, cost, and lack of evidence of cost-competitiveness. Commercial proton delivery systems had been contemplated for decades before they finally appeared in 2001 after overcoming considerable difficulties. The cost of proton therapy equipment remains much higher than that of comparable photon therapy equipment; the long-anticipated economies of scale have not, as yet, materialized. Even in times of relative prosperity, the allocation of scarce resources to proton therapy has been constrained by relatively

Despite these obstacles, much progress has been made. Today there are 16 proton therapy centers in operation in the United States and 46 centers worldwide (PTCOG 2014). The Particle Therapy Cooperative Group (PTCOG) reported that at least 105,743 patients had been treated worldwide by the end of 2013 (PTCOG 2014). The proton therapy community has stepped up efforts to conduct clinical trials that compare outcomes after proton therapy with those after other advanced technology radiation therapies (Duttenhaver et al 1983, Shipley et al 1995, Desjardins et al 2003, Zietman et al 2010).

The central rationale for proton therapy is its superior spatial dose distribution in the patient. In recent years, the advantage of protons over photons in providing a highly conformal and uniform dose to a tumor has been largely diminished by advances in photon therapies, such as intensity-modulated photon therapy and volumetric arc therapies (Weber et al 2009). However, the relative advantage of proton therapy in sparing normal tissues has never been more apparent or important; in the United States, approximately 65% of adults and 80% of children survive 5 years after their cancer diagnosis (Valdivieso et al 2012). About half of cancer patients receive radiotherapy as part of their treatment. Recent studies reported that the incidence of treatment-related morbidity, including second cancers, cardiovascular disease, fertility complications, and other late effects, is alarmingly high in long-term survivors of cancer (Wilson et al 2005, Carver et al 2007, Armstrong et al 2009, Merchant 2009, Sauvat et al 2009, Newhauser and Durante 2011, Olch 2013). Presently, about 3% of the US population are cancer survivors, corresponding to 11 million people, a figure projected to grow to 18 million by 2022 (de Moor et al 2013). For these reasons, there is increasing interest in exploiting the tissue-sparing capabilities inherent to proton therapy to reduce the burden of treatment-related complications on patients and the healthcare system. An understanding of the physics and biology of radiogenic late effects from proton therapy has started to emerge in the literature in the last decade.

This paper reviews the basic, essential physics underlying proton therapy. The literature includes excellent reviews of various aspects of proton therapy physics, most notably the comprehensive work of Chu et al (1993), and also those of Bonnett (1993), Pedroni et al (1995), Brahme (2004), Lomax (2009), Coutrakon (2007), and Schardt and Elsasser (2010) In addition, several reports (ICRU 1998, 2007) and textbooks (Scharf 1994, Breuer and Smit 2000, DeLaney and Kooy 2008, Linz 2012, Paganetti 2012, Ma and Lomax 2013, Moyers and Vatnitsky 2012) have covered various aspects of proton therapy physics. In recent years, many textbooks dealing with general radiation oncology have included relevant chapters on proton therapy (Van Dyk 1999, Halperin et al 2008, Pawlicki et al 2011). Many of the older works on proton therapy have withstood the test of time and remain excellent literature resources of continued relevance. However, in this review we mainly focus on the well-established basic physics of proton therapy and on selected advances from the last 15 years or so that are important in clinical proton therapy. In choosing advances to cover in this review, considerable selectivity was necessary because of the huge expansion of the proton therapy literature, especially in recent years. To the authors of the many studies that we were not able to mention in this review because of space limitations, we apologize and we appreciate your understanding.

2. Proton interaction mechanisms

In this section, we briefly review the predominant types of interactions of protons in matter and why they are important. Figure 1 illustrates several mechanisms by which a proton interacts with an atom or nucleus: Coulombic interactions with atomic electrons, Coulombic interactions
with the atomic nucleus, nuclear reactions, and Bremsstrahlung. To a first-order approximation, protons continuously lose kinetic energy via frequent inelastic Coulombic interactions with atomic electrons. Most protons travel in a nearly straight line because their rest mass is 1832 times greater than that of an electron. In contrast, a proton passing close to the atomic nucleus experiences a repulsive elastic Coulombic interaction which, owing to the large mass of the nucleus, deflects the proton from its original straight-line trajectory. Non-elastic nuclear reactions between protons and the atomic nucleus are less frequent but, in terms of the fate of an individual proton, have a much more profound effect. In a nuclear reaction, the projectile proton enters the nucleus; the nucleus may emit a proton, deuteron, triton, or heavier ion or one or more neutrons. Finally, proton Bremsstrahlung is theoretically possible, but at therapeutic proton beam energies this effect is negligible. Table 1 summarizes the proton interaction types, interaction targets, principal ejectiles, influence on the proton beam, and dosimetric manifestations. We review these interaction mechanisms, except proton Bremsstrahlung, in the following sections.

2.1. Energy loss rate

The energy loss rate of ions, or linear stopping power, is defined as the quotient of dE and dx, where E is the mean energy loss and x is the distance. It is frequently more convenient to

![Figure 1. Schematic illustration of proton interaction mechanisms: (a) energy loss via inelastic Coulombic interactions, (b) deflection of proton trajectory by repulsive Coulomb elastic scattering with nucleus, (c) removal of primary proton and creation of secondary particles via non-elastic nuclear interaction (p: proton, e: electron, n: neutron, γ: gamma rays)](image-url)
express the energy loss rate in a way that is independent of the mass density; the mass stopping power is defined as

$$\frac{S}{\rho} = -\frac{dE}{\rho dx},$$

(1)

where $\rho$ is the mass density of the absorbing material. Please note that stopping power is defined for a beam, not a particle.

The energy loss rate may be described by several mathematical formulae. The simplest, yet still remarkably accurate, formula is based on the Bragg–Kleeman (BK) rule (Bragg and Kleeman 1905), which was originally derived for alpha particles, and is given by

$$\frac{S}{\rho} = -\frac{dE}{\rho dx} \approx -\frac{E^{1-p}}{\rho a \alpha},$$

(2)

where $\rho$ is the mass density of the material, $\alpha$ is a material-dependent constant, $E$ is the energy of the proton beam, and the exponent $p$ is a constant that takes into account the dependence of the proton’s energy or velocity. Values of $\alpha$ and $p$ may be obtained by fitting to either ranges or stopping power data from measurements or theory.

A more physically complete theory, developed by Bohr (1915), is based on calculation of the momentum impulse of a stationary, unbound electron and the impact parameter. A more accurate formula, attributed to Bethe (1930) and Bloch (1933), takes into account quantum mechanical effects and is given by

$$\frac{S}{\rho} = -\frac{dE}{\rho dx} = 4\pi N_A r_e^2 m_e c^2 Z^2 z^2 \left[ \ln \frac{2\mu_c c^2 \gamma^2 \beta^2}{I} - \beta^2 - \frac{\delta}{2} - \frac{C}{Z} \right],$$

(3)

where $N_A$ is Avogadro’s number, $r_e$ is the classical electron radius, $m_e$ is the mass of an electron, $z$ is the charge of the projectile, $Z$ is the atomic number of the absorbing material, $A$ is the
atomic weight of the absorbing material, $c$ is speed of light, $\beta = v/c$ where $v$ is the velocity of the projectile, $\gamma = (1 - \beta^2)^{-1/2}$, $I$ is the mean excitation potential of the absorbing material, $\delta$ is the density corrections arising from the shielding of remote electrons by close electrons and will result in a reduction of energy loss at higher energies, and $C$ is the shell correction item, which is important only for low energies where the particle velocity is near the velocity of the atomic electrons. The two correction terms in the Bethe–Bloch equation involve relativistic theory and quantum mechanics and need to be considered when very high or very low proton energies are used in calculations. Figure 2 plots proton stopping power as a function of proton energy in water calculated by using equation (3) at high proton energies (above about 1 MeV) and other methods (not presented) at lower energies.

It is instructive to observe in equation (3) how the projectile’s characteristics govern its energy loss rate: energy loss is proportional to the inverse square of its velocity ($1/v^2$ classically and $1/\beta^2$ relativistically) and the square of the ion charge ($z = 1$ for protons), and there is no dependence on projectile mass. Similarly, equation (3) reveals that the absorber material can also strongly influence the energy loss rate. Specifically, the linear stopping power is directly proportional to the mass density. It is equivalent, but perhaps more physically meaningful, to state that the linear stopping power is proportional to the density of electrons in the absorber ($N_A \rho Z/A$), because the energy loss occurs by Coulombic interactions between the proton and atomic electrons. $Z/A$ varies by only about 16%, from 0.5 for biologic elements such as carbon and oxygen to 0.42 for high-$Z$ beamline components, such as lead. Hydrogen is an obvious exception to this; fortuitously, the concentration of hydrogen in the human body is low (only about 10%) and nearly uniform throughout the body. The stopping power also depends on a material’s $I$ value, and the $I$ value depends in a monotonic way on the $Z$ of the absorber, varying from about 19 eV for hydrogen to about 820 eV for lead. However, the stopping power goes with the logarithm of $I^{-1}$ value, so the dependence is diminished. Hence, putting these various dependencies in perspective, it is clear that the proton energy loss rate in the human body depends most strongly on the material density, which can vary by about...
three orders of magnitude, from air in the lung to cortical bone, and the ion velocity, which can cause the linear stopping power in water to vary by about a factor of 60 for proton energies between 1 and 250 MeV.


Thus far, we have described proton energy loss in an approximate way on the assumptions that a proton loses energy along a 2D line trajectory and that energy loss is continuous. Absorption of this same energy, however, occurs in a 3D volume. Furthermore, the ionization track of a proton has an irregular 3D structure caused by random fluctuations in the location and size of primary ionization events. This is caused mainly by proton-produced recoil electrons, some of which are sufficiently energetic to create small spur tracks of ionization emanating from the main track. Because the electrons are very much lighter than the protons, each interaction can reduce the proton energy only a little. The maximum possible energy transfer in a collision of an ion of mass $m$ with an unbound stationary electron is

$$\Delta E_{\text{max}} = \frac{2m_e\beta^2}{1 - \beta^2} \left[ 1 + 2 \frac{m_e}{M} \frac{1}{\sqrt{1 - \beta^2}} + \left( \frac{m_e}{M} \right)^2 \right],$$  \hspace{1cm} (4)$$

where $m_e$ is the mass of an electron, $M$ is the mass of the target material, $c$ is the speed of light, and $\beta = v/c$ where $v$ is the velocity of the projectile.

Even for very energetic protons, the secondary electrons do not acquire enough energy to travel more than a few millimeters from the proton track. For example, at 200 MeV proton energy, the maximum secondary electron energy is around 500 keV, which corresponds to an electron range of approximately 2 mm in water. The probability of producing secondary electrons may be calculated with various total or differential cross-sections; these were reviewed in ICRU Report 55 (1995). Track structure models may be used to estimate the radial properties of ions (Kraft et al 1999), although this has not found common application in clinical proton therapy. Regardless of the calculation method used, the spatial characteristics of secondary charged particles should, in principle, be taken into account near material interfaces (e.g. buildup effects in transmission beam monitoring instruments, skin, air–tumor interfaces in the lung) and in cases where the radiation quality is of interest (e.g. microdosimetric and nanodosimetric characterization of individual dose deposition events).

Finally, we note that in proton therapy water is considered an excellent tissue substitute because of its similar density, effective $Z/A$, and other properties. Furthermore, proton energy loss and residual range in various materials are often expressed in terms of their water-equivalent values. We discuss this in detail in section 3.1, but until then, for simplicity, we consider water as being perfectly equivalent to tissue.

### 2.2. Range

Range is defined as the depth at which half of protons in the medium have come to rest, as shown in the range-number curve plotted in figure 3. There are small variations in the energy loss of individual protons (an effect called range straggling, which is discussed in the following section). Consequently, the range is inherently an average quantity, defined for a beam and not for individual particles. By convention, this means half of the protons
incident on the absorber are stopped, although in some cases this is taken instead to mean
half of the protons survived to near the end of range (i.e. neglecting protons removed by
nuclear reactions).

The path of most protons in matter is a nearly straight line. On average, the proton’s path-
length is very nearly equal to its projected pathlength and range. This simple but important
fact renders many proton range calculations tractable with relatively simple numerical or ana-
lytical approaches.

Let us first consider a simple numerical calculation of proton beam range. We use proton
stopping power data and perform a 1D proton pathlength transport calculation on the assump-
tions that the ions travel only straight ahead (negligible lateral scattering) and that the protons
lose energy in a continuous matter. (These assumptions are reasonable for many clinical calcu-
lations, but we examine then relax these assumptions in later sections.) In this case, the range
($R$) may be calculated as

$$R(E) = \int_0^E \left( \frac{dE'}{dx} \right)^{-1} dE' \approx \sum_0^E \left( \frac{dE'}{dx} \right)^{-1} \Delta E',$$

where $E$ is the ion’s initial kinetic energy. The summation denotes that the continuous trans-
port is approximated by calculations of discrete steps. In fact, as discussed above, this equa-
tion actually gives the pathlength, which is an excellent approximation of range in most
clinical situations. Figure 2 plots proton range in water calculated by using equation (5).

Next, we calculate the proton range using an analytical approach, which may be faster and
more practical than the numerical approach for many clinical calculations. We begin by noting

\[\text{Figure 3. Relative fraction of the fluence } \Phi \text{ in a broad beam of protons remaining}
\text{as a function of depth } z \text{ in water. The gradual depletion of protons from entrance to}
\text{near the end of range is caused by removal of protons from nuclear reactions. The}
\text{rapid falloff in the number of protons near the end of range is caused by ions running}
\text{out of energy and being absorbed by the medium. The sigmoid shape of the distal}
\text{falloff is caused by range straggling or by stochastic fluctuations in the energy loss}
\text{of individual protons.} \]
that the interval of proton range of interest typically extends from 1 mm (about the size of a voxel in an anatomic image of patient anatomy) to about 30 cm (about the midline of a large adult male’s pelvis, the deepest site in the human body). These ranges correspond to 11 MeV and 220 MeV, respectively, as seen in figure 2. More importantly, figure 2 reveals that the relationship between the logarithm of range and logarithm of energy is almost linear. This is fortuitous because this means that the range follows a very simple power law, as realized by Bragg and Kleeman (1905) and others early in the last century. Thus, the range of a proton may be calculated using the Bragg–Kleemann rule, or

$$ R(E) = \alpha E^p, $$

where, as before, $\alpha$ is a material-dependent constant, $E$ is the initial energy of the proton beam, and the exponent $p$ takes into account the dependence of the proton’s energy or velocity.

The uncertainty in proton range depends on many factors. For example, the uncertainty in a range measurement depends on the precision and accuracy of the measurement apparatus and, in some cases, on the experimenter’s skill. A common concern in clinical proton therapy is the uncertainty in calculated range, e.g. in calculating the settings of the treatment machine for a patient’s treatment. The uncertainty in range may depend on the knowledge of the proton beam’s energy distribution and on properties of all range absorbing materials in the beam’s path. These properties include elemental composition, mass density, and linear

---

**Figure 4.** Energy loss PDFs are plotted for various thicknesses of water absorbers, where the thickness is expressed in units of mean free path (mfp). For visual clarity, the energy-loss PDFs have been scaled on both the abscissa and ordinate. The single event energy loss is expressed as a fraction of the mean energy lost in the entire absorber thickness, or $(I - I_{av})/I_{av}$. Each PDF was scaled so that the integral over all energy-loss values yields unit value. For thin absorbers (curves a–e), the PDFs are broader and asymmetric and are modeled with the Vavilov (1957) or Landau (1944) theories. For thick absorbers (curve f), the PDFs are symmetric and well-approximated with Bohr’s theory (1915), i.e. a Gaussian distribution (reproduced with permission from ICRU 1993).
stopping power. The linear stopping powers deduced from computed tomography scans have many additional sources of uncertainty, including imperfections in the calibration of CT scanners (i.e. the method used to convert image data from Hounsfield Units to linear stopping power), partial volume effects, motion artifacts, and streaks artifacts.

2.3. Energy and range straggling

In the preceding sections, we approximated the energy loss rate by assuming that the slowing of ions occurs in a smooth and continuous manner. In fact, we considered the mean energy loss rate and neglected variations in the energy loss rates of individual protons. For many clinical calculations, these assumptions are valid and lead to a reasonably good first-order approximation. However, the accumulation of many small variations in energy loss, termed energy straggling or range straggling, is one of the physical processes that strongly governs the shape of a proton Bragg curve, a subject that will be covered in section 3.2. Thus, an understanding of range straggling is key to understanding the characteristics of proton dose distributions.

Figure 4 plots the relative energy loss probability density functions (PDFs) for protons transmitted through water absorbers of various thickness. The curves have been normalized to enhance visual clarity. Apparently, thick absorbers result in a symmetric distribution of energy losses, whereas thin absorbers yield curves that are asymmetric with modes less than the mean and long tails of large energy losses. In principle, straggling PDFs may be calculated numerically from first principles, but usually theoretical approaches are used. Later in the section, three of the most commonly used straggling theories are described.

Having examined the mean energy loss rate in modest detail (section 2.1), and having conceptually introduced energy straggling, it is instructive to understand how these effects are related mathematically before delving into straggling theory. To understand these relationships, let us consider the moments of the ion energy PDF, or

\[ M_n = \int_0^{\Delta_{\text{max}}} \Delta^n f(\Delta) d\Delta, \]  

where \( \Delta \) is the energy loss of an ion in traversing an absorber, \( f(\Delta)d\Delta \) is the probability of energy loss occurring in the interval from \( \Delta \) to \( \Delta + d\Delta \), \( n \) is the order of the moment, and \( \Delta_{\text{max}} \) is from equation (14). The zeroth moment is the total collision cross section, the first moment is the (mean) electronic stopping power, the second moment corresponds to the width (variance) of the energy straggling distribution, the third moment to its asymmetry (skewness), and the fourth to its kurtosis. The variance, sometimes denoted by \( \sigma_2 \), or second moment of the straggling distribution \( f(\Delta) \), is

\[ M_2(\Delta) = \sigma_2^2 = \bar{v} \int_0^{\Delta_{\text{max}}} \Delta^2 f(\Delta) d\Delta, \]  

where \( \bar{v} \) is the average number of primary collisions per proton traversal. A more detailed discussion of the straggling moments was given by Rossi and Zaider (1996).

Next we examine theories for calculating energy straggling proposed by Bohr (1915), Landau (1944), and Vavilov (1957). These theories are valid for thick, intermediate, and thin absorbers, respectively. The criterion for selecting a valid theory for a given absorber thickness is based on a reduced energy loss parameter,
where $\xi$ is the approximate mean energy loss and $\Delta_{1 \text{max}}$ is the maximum energy loss possible in a single event.

According to Bohr’s theory, the energy straggling distribution behaves according to a Gaussian PDF, or

$$f(\Delta)d\Delta = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left(-\frac{\Delta - \bar{\Delta}}{2\sigma_1^2}\right),$$

where for non-relativistic ions the variance is given by

$$\sigma_1^2 = 2\pi^2 \rho e^2 \bar{x} \beta^2 \Delta_{1 \text{max}} x^2,$$

where $\rho$ is the mass density and $x$ is the thickness of the absorber. Bohr’s theory assumes that the absorber is thick enough that there are many individual collisions (i.e. the central limit theorem holds), that the ion velocity does not decrease much in crossing the absorber, and that the absorber is made of unbound electrons. For most applications in proton therapy, Bohr’s theory provides adequate accuracy. Several authors have reported convenient power law approximations to estimate sigma as a function of the proton beam range (Chu et al. 1993), or

$$\sigma_1 \approx kR_0^m,$$

where $R_0$ is the range in water in centimeters for a mono-energetic proton beam, $k$ is a material-dependent constant of proportionality, and the exponent is empirically determined. For protons in water, $k = 0.012$ and $m = 0.935$ (Bortfeld 1997).

Landau’s theory relaxed the assumption that the central limit holds, i.e. there are relatively fewer individual collisions in an intermediate thickness absorber, and used an approximate expression for $\Delta_1$. In this case, we have

$$f(\Delta, \rho x)d\Delta = \frac{1}{\xi} \varphi_L(\Delta_L),$$

where the parameter $\varphi_L(\Delta_L)$ roughly corresponds to the deviation from the mean energy loss and was defined by Landau as

$$\varphi_L(\Delta_L) = \frac{1}{\pi} \int_0^\infty e^{-\nu (\nu - \Delta_L)^2} \sin(\nu x) d\nu.$$

Evaluation of the integral in equation (14) is straightforward and the reader is referred to the literature for details (Seltzer and Berger 1964).

Vavilov’s theory is in essence a generalization of Landau’s theory that utilizes the correct expression for $\Delta_1$ and is given by

$$f(\Delta, \rho x)d\Delta = \frac{1}{\xi} \varphi_\lambda(\lambda_0, k, \beta^2) d\lambda_0,$$

where

$$\varphi_\lambda(\lambda_0, k, \beta^2) = \frac{1}{\pi} e^{k(1 + \beta^2)} \int_0^\infty e^{-\nu \lambda_0 \nu + k \nu} d\nu.$$
and
\[ \lambda_v = \frac{\Delta - \Delta_{\text{max}}}{\Delta_{\text{max}}} - k(1 + \beta^2 - \gamma), \]  

(17)

where \( \gamma \) is Euler’s constant. The evaluation of Vavilov’s theory is computationally more expensive than that of Bohr’s or Landau’s. For additional details, the reader is referred to Vavilov’s original work (Vavilov 1957).

2.4. Multiple Coulomb scattering

As illustrated in figure 5, a proton passing close to the nucleus will be elastically scattered or deflected by the repulsive force from the positive charge of the nucleus. While the proton loses a negligible amount of energy in this type of scattering, even a small change in its trajectory can be of paramount importance. In fact, it is necessary to take into account Coulomb scattering when designing beamlines and treatment heads (Gottschalk 2010b) and in calculations of dose distributions in phantoms or patients, e.g. with treatment planning systems (Hong et al 1996, Pedroni et al 2005, Schaffner 2008, Koch and Newhauser 2010).

To characterize the amount that a beam is deflected by scattering, we use the quantity of scattering power, which is defined as
\[ T = \text{d}<\theta^2>/\text{dx}, \]

(18)

where \(<\theta^2>\) is the mean squared scattering angle and \(x\) is the thickness of absorber through which the proton has traveled. Similarly, the mass scattering power is simply \(T/\rho\), where \(\rho\) is the mass density of the absorber material. Notice that the definition of scattering power utilizes the mean square of the scattering angle; scattering is symmetric about the central axis, and therefore the mean scattering angle is zero and contains no useful information. Also, note that the value of the scattering power depends on the material properties and dimensions of the absorber being considered.

Predictions of elastic Coulomb scattering are commonly classified according to the number of individual scattering events \(N_s\) that occur in a given absorber. For single scattering \((N_s = 1)\), Rutherford scattering theory applies. Plural scattering \((1 < N_s < 20)\) is difficult to model theoretically and is not discussed further here. For multiple Coulomb scattering (MCS; \(N_s \geq 20\)), the combined effect of all \(N_s\) scattering events, may be modeled by using a statistical approach to predict the probability for a proton to scatter by a net angle of deflection, commonly denoted by \(\theta\) (figure 5).
It is instructive to briefly examine Rutherford’s theory of single scattering (Rutherford 1911). The differential cross section \( \frac{d\sigma}{d\Omega} \) for scattering into the solid angle \( d\Omega \) is given by

\[
\frac{d\sigma}{d\Omega} = z_1^2 z_2^2 e^2 \frac{(m_e c / \beta p)^2}{4 \sin^4(\theta/2)},
\]

where \( z_1 \) is the charge of the projectile, \( z_2 \) is the atomic number of the absorber material, \( r_e \) is the classical electron radius, \( m_e \) is the mass of the electron, \( c \) is the speed of light, \( \beta = v/c \), \( p \) is the ion momentum, and \( \theta \) is the scattering angle of the proton. The angular dependence is governed by the term \( 1/\sin^4(\theta/2) \), i.e. in individual scattering events, the proton is preferentially scattered in the forward direction, at very small angles.

In clinical proton therapy, most objects of interest are thick enough to produce a great many scattering events. Thus, for clinical proton therapy, we are usually more interested in the net effect that many small-angle scattering events have on many protons, e.g. how beamline scattering in the treatment head influences the spatial distribution of dose in a patient.

Rigorous theoretical calculations of MCS are quite complex. The most complete theory was proposed by Molière (1948). Assuming that scattered particles are emitted at small deflection angles (i.e. the small-angle approximation in which \( \sin(\theta) \approx \theta \)),

\[
P(\theta) d(\Omega) = \eta d\eta \left( 2 \exp(-\eta^2) + \frac{F_1(\eta)}{B} + \frac{F_2(\eta)}{B^2} + \ldots \right),
\]

where \( \eta = \theta / (\theta_1 B^{1/2}) \), \( \theta_1 = 0.3965 (Z/z_2) \rho \delta x / A \) and \( d(\Omega) \) is the solid angle into which the particles are scattered. The functions \( F_k(\eta) \) are defined as

\[
F_k(\eta) = \frac{1}{k!} \int J_0(\eta y) \exp \left( -\frac{y^2}{4} \right) \left[ \frac{y^2}{4} \ln \left( \frac{y^2}{4} \right) \right]^k dy,
\]

where \( J_0(\eta y) \) is a Bessel function. Values of \( F_k \) have been tabulated in the literature by several authors. Parameter \( B \) in equation (20) may be found by numerical methods, by solving

\[
g(B) = \ln B - B + \ln \gamma - 0.154 = 0,
\]

where \( \gamma = 8831 q z^2 \rho \delta x / (\beta^2 A) \) and \( A = 1.13 + 3.76 (Z/z_2) \) (\( Z/z_2 \) is the thickness, and \( \rho \) is the mass density. The proton momentum is denoted by \( p, \beta = v/c \), and \( z = 1 \) is the proton charge. Even though we have not presented all of the details, clearly the evaluation of Molière’s theory is indeed complex. Consequently, considerable attention has been paid in the literature to developing simpler formulae; the simplicity usually comes at the cost of reduced accuracy in modeling scattering at large angles. Gottschalk et al (1993) discussed several of these methods in detail. One approximation method takes the form of a Gaussian distribution,

\[
P(\theta) \approx \frac{2\theta}{\langle \theta^2 \rangle} \exp \left( -\frac{\theta^2}{\langle \theta^2 \rangle} \right) d\theta,
\]

where \( \langle \theta^2 \rangle^{1/2} \) is the root mean square (rms) scattering angle or the width parameter of the Gaussian distribution.

Gottschalk (2010b) proposed a differential approximation to Molière’s theory to predict the scattering power according to

\[
T_{AM} = f_{AM}(p v, p v_1) \left( \frac{E_S}{p v} \right)^2 \frac{1}{X_S},
\]
where $E_S = 15$ MeV, $p$ is the momentum and $v$ is the velocity of the proton, $p_1$ and $v_1$ are the initial momentum and velocity, $X_S$ is scattering length of the material, and $f_{dM}$ is a material-independent nonlocal correction factor given by

$$f_{dM} = 0.5244 + 0.1975 \log \left(1 - \frac{p v}{p_1 v_1}\right) + 0.2320 \log(p v) - 0.0098 \log(p v) \log \left(1 - \frac{p v}{p_1 v_1}\right).$$

The factor $f_{dM}$ takes into account the accumulation of scattering that occurs as the proton slows from $v_1$ to $v$ and is material independent. The scattering length is given by

$$\frac{1}{\rho X_S} = aN_r^2 Z^2 \left(2 \log(33219(AZ)^{1/3}) - 1\right),$$

where $a$ is the fine structure constant, $N$ is Avogadro’s number, $r_e$ is electron radius, and $Z, A,$ and $\rho$ are the atomic number, atomic weight, and density of the target material.

It is sometimes convenient to project the expected scattering angle $\langle \theta \rangle$ in 3D space to the corresponding value in a plane, denoted $\langle \theta_x \rangle$, according to

$$\langle \theta_x \rangle = \frac{1}{2} \langle \theta \rangle.$$

Also, in proton therapy the lateral displacement ($r$) of a proton beam is typically calculated from the scattering angle. Under the Gaussian approximation, we have

$$P(r) \, dr = 6r \left(\langle \theta^2 \rangle r^2 \right)^{-1} \exp \left(-3r^2/\langle \theta^2 \rangle r^2\right) \, dr,$$

where $t = x/L_{\text{rad}}$, and $L_{\text{rad}}$ is the radiation length of the material, which can be calculated easily or looked up in tables. The mean square lateral displacement is given by

$$\langle r^2 \rangle = \frac{\langle \theta^2 \rangle r^2}{3}.$$
A power law approximation for rms displacement of protons is

$$r_{\text{rms}} = aR^b,$$  \hspace{1cm} (30)
where $a$ is a unitless material constant, $R$ is the water-equivalent proton beam range in cm, and the exponent $b$ governs the range dependence. For protons in water, $a = 0.0294$ and $b = 0.896$ (Chu et al 1993). Koehler and Preston derived a convenient expression to calculate $r_{rms}$ as a function of depth in an absorber and knowledge of its maximum values, $r_{max}$, at the end of range (from an unpublished manuscript; some portions of their work were reported by Gottschalk (2010b)).

In proton therapy, MCS in the treatment head (i.e. in the scattering foils) is helpful because it allows the beam to be spread laterally to useful dimensions, e.g. to make a beam laterally large and flat so that a tumor may be completely covered with a uniform dose. Scattering foils in the nozzle are carefully designed to utilize MCS and energy loss to produce clinically useful beams (Koehler et al 1977, Gottschalk 2004) However, MCS in the treatment head and in the patient blurs lateral penumbral sharpness. This is manifested as penumbral growth at the edge of collimated beams and/or the growth of the lateral spot size of a scanned beam (figure 6). Understanding and preserving penumbral sharpness is key to realizing the full benefit of proton therapy for sparing healthy tissue.

Recent studies have revealed that MCS plays an important role in proton dose distribution around small implanted metal objects. Specifically, implanted fiducial markers for image-guided patient alignment have been used in proton therapy for many years (Gall et al 1993, Welsh et al 2004, Newhauser et al 2007a, 2007c, Ptaszkiewicz et al 2010, Matsuura et al 2012). Substantial recent improvements in on-board imaging systems, patient positioning, and patient immobilization have led to increased use of radiopaque implanted fiducial markers in proton therapy to many disease sites, with the goal of improving target coverage and/or normal tissue sparing. However, recent studies revealed that some commonly used markers, even those less than 1 mm in size, can cause severe cold spots (figure 7), compromising target coverage (Newhauser et al 2007a, 2007c, Carnicer et al 2013). The severity of the cold spots varies with fiducial size, material composition, and mass density. These parameters, in turn, determine the amount of MCS and energy loss in the fiducial and, hence, perturbations to the dose distribution in surrounding tissue. In essence, MCS in the fiducial causes a downstream dose shadow that may be partially or fully filled in by MCS in the surrounding tissue. While MCS is important, energy loss in the fiducial (or its water-equivalent thickness), the proximity of the fiducial to the end of the proton beam range, and of course its size and orientation with respect to the beam also should be considered. The physics of dose perturbations is explained in detail elsewhere (Newhauser et al 2007a), and several subsequent studies (Giebeler et al 2009, Lim et al 2009, Cheung et al 2010, Huang et al 2011) have shown that it is possible to achieve good radiographic visibility using novel markers that do not significantly perturb the therapeutic dose distribution in tissue.

Others have examined the effects of MCS in larger metal objects on clinical proton beams and characterized the suitability of approximate methods to predict MCS in practical clinical applications (Herault et al 2005, Stankovskiy et al 2009, Newhauser et al 2013).

2.5. Nuclear interactions

In addition to the mechanisms already described, protons may interact with the atomic nucleus via non-elastic nuclear reactions in which the nucleus is irreversibly transformed, e.g. a reaction in which a proton is absorbed by the nucleus and a neutron is ejected, denoted by (p.n). The main effect of nuclear reactions within a therapeutic region of a proton field is a small decrease in absorbed dose due to the removal of primary protons, which is compensated to a large extent by the liberation of secondary protons and other ions. In this section, we discuss this and several other important aspects of nuclear reactions.
Before discussing reaction mechanisms, it is instructive to examine a range-number curve (figure 3), which plots the remaining number of protons versus depth in an absorber as a beam comes to rest. The gradual depletion of protons from entrance to near the end of range is caused by removal of protons by nuclear reactions. The rapid falloff in the number of protons near the end of range is caused by ions running out of energy and being absorbed by the medium. The sigmoid shape of the distal falloff is caused by range straggling or by stochastic fluctuations in the energy loss of individual protons.

To enter the nucleus, protons need to have sufficient energy to overcome the Coulomb barrier of the nucleus, which depends on its atomic number. The total non-elastic cross-section for proton-induced nuclear reactions has a threshold, on the order of 8 MeV in the atomic nuclei of biologically relevant elements, which rises rapidly to a maximum at around 20 MeV, then asymptotically declines to about half the maximum value by about 100 MeV (figure 8). Tabulated and graphical nuclear data may be obtained conveniently online from the Evaluated Nuclear Data File (ENDF) (IAEA 2013). ICRU Report 63 also provides extensive nuclear data for hadron therapy and radiation protection (ICRU 2000).

Several nuclear reactions are particularly important to clinical proton therapy and proton therapy research. In a proton therapy beam, proton-induced reactions can produce energetic protons, deuterons, tritons, 3He, 4He, and other ions. Secondary protons comprise as much as about 10% of the absorbed dose in a high-energy proton treatment beam; they have a small but non-negligible impact on the spatial dose distribution in a patient (Medin and Andreo 1997, Boon 1998, ICRU 1998, Paganetti 2002, Wroe et al 2005). Deuterons and heavier ions are present in much smaller proportions; collectively they comprise about 1% or less of the therapeutic absorbed dose (ICRU 1998). Their energy and range are very small and they deposit their kinetic energy locally, i.e. very near their point of creation.

Relatively high-current protons beams are incident on certain beam production and delivery equipment and on some patients. These proton beams produce neutrons that create significant potential safety hazards. Great care must be taken to limit exposures of personnel to
neutrons (NCRP 1971, 1990, 1993, 2005, Newhauser and Durante 2011). Some electronic systems must also be hardened, shielded, or located so that neutron radiation does not cause soft upsets or permanent damage to semiconductor components. Attention must also be paid to neutron activation of beamline components, air, groundwater, and other materials (IAEA 1988).

Neutrons are produced in copious quantities: they span 10 orders of magnitude in neutron energy, their energy distributions depend strongly on the proton beam energy and direction, they are extremely penetrating, and their relative biologic effectiveness is as much as about 20 times higher than that of proton radiation (ICRP 2007). Thus they potentially increase the risk of radiogenic late effects (Hall 2006, Brenner and Hall 2008, Newhauser and Durante 2011). Several specific aspects of neutron exposures are considered in a later section of this paper.

Nuclear reactions inside the patient may provide a non-invasive approach to measure a variety of beam and patient properties, such as proton beam range, elemental composition of tissues, and even intra- or inter-fraction physiology. The basic approach is to detect gamma rays from proton-induced nuclear reactions, such as neutron capture reactions, denoted by \((n, \gamma)\). Approaches are under development that detect photons from positron annihilation, prompt gammas, and delayed gammas. Gamma ray detection approaches have included positron emission tomography camera (Parodi et al 2007, Moteabbed et al 2011, Cho et al 2013, Min et al 2013), Compton camera (Peterson et al 2010, Smeets et al 2012), 1D detector arrays (Min et al 2012), and photon counting systems (Kim et al 2012). These techniques are in various stages of research and development; none is routinely used in clinical practice. There remain many challenges to overcome, including instrument sensitivity and calibration; interpretation of measurements, including an understanding of managing measurement artifacts; and competition from alternative methods, e.g. magnetic resonance imaging techniques (Krejcarek et al 2007, Raaymakers et al 2008, Gensheimer et al 2010).

3. Proton transport calculations

In this section, we review several aspects of proton transport physics that are encountered frequently in clinical and research situations. We describe the 1D water-equivalent thickness of an arbitrary material, the shapes of a pristine Bragg curve and a spread out Bragg peak (SOBP) curve, and stray radiation exposures.
As we mentioned previously, in proton therapy water closely mimics the properties of human tissues in terms of energy loss, MCS, and nuclear interactions. As such, water is a recommended phantom material for dose and range measurements and reference material for reporting corresponding calculated quantities (ICRU 1998, IAEA 2000). For example, it is a common and convenient clinical practice to specify the penetrating power of a proton beam by its range in water (ICRU 1998, 2007). In this way, range losses in various beamline objects and the patient may be easily added or subtracted from one another in a physically consistent and intuitive way. Viewed another way, it is also convenient to specify the range-absorbing power of various objects in the beam path, e.g. beam transmission monitors and immobilization devices, by their equivalent thickness if they were made of water.

Water-equivalent thickness (WET) is often used to characterize the beam penetration range; figure 9 schematically illustrates the concept of WET and how it can be calculated or measured. For treatment sites with nearby critical structures, e.g. an optic nerve, the range of the planned and delivered beams must agree within a few millimeters. To accomplish this, treatment planning systems are commonly configured with the WET values of all items not included in the planning CT images, such as components in the treatment head, immobilization devices not present during the CT scan, or a treatment couch (Newhauser et al 2007b). In this way, range losses in various beamline objects and the patient may be easily added or subtracted from one another in a physically consistent and intuitive way. Viewed another way, it is also convenient to specify the range-absorbing power of various objects in the beam path, e.g. beam transmission monitors and immobilization devices, by their equivalent thickness if they were made of water.

Water-equivalent thickness (WET) is often used to characterize the beam penetration range; figure 9 schematically illustrates the concept of WET and how it can be calculated or measured. For treatment sites with nearby critical structures, e.g. an optic nerve, the range of the planned and delivered beams must agree within a few millimeters. To accomplish this, treatment planning systems are commonly configured with the WET values of all items not included in the planning CT images, such as components in the treatment head, immobilization devices not present during the CT scan, or a treatment couch (Newhauser et al 2007b). In this way, range losses in various beamline objects and the patient may be easily added or subtracted from one another in a physically consistent and intuitive way. Viewed another way, it is also convenient to specify the range-absorbing power of various objects in the beam path, e.g. beam transmission monitors and immobilization devices, by their equivalent thickness if they were made of water.

### Table 2. Common materials (lung substitute plastic, high-density polyethylene (HDPE), water, polystyrene, polymethylmethacrylate (PMMA), polycarbonate resin (Lexan), bone substitute plastic, aluminum, titanium, stainless steel, lead and gold) used in heavy charged particle beams, with their mass densities \( \rho \), values of \((Z/A)_{\text{eff}}\), mole fractions and fitting parameters of \( \alpha \) and \( p \) for these materials when applying the BK rule. (The energies used in the fit were from 10 to 250 MeV).

<table>
<thead>
<tr>
<th>Material</th>
<th>( P ) (g cm(^{-3}))</th>
<th>((Z/A)_{\text{eff}})</th>
<th>Mole fraction (%)</th>
<th>( \alpha )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung substitute</td>
<td>0.3</td>
<td>0.537</td>
<td>H 55.577, C 32.738, N 0.927, O 7.508, Cl 0.019, Si 0.184, Mg 3.048</td>
<td>8.994 \times 10^{-3}</td>
<td>1.735</td>
</tr>
<tr>
<td>HDPE</td>
<td>0.964</td>
<td>0.570</td>
<td>H 66.717, C 33.283</td>
<td>2.541 \times 10^{-3}</td>
<td>1.737</td>
</tr>
<tr>
<td>Water</td>
<td>1.0</td>
<td>0.555</td>
<td>H 66.667, O 33.333</td>
<td>2.633 \times 10^{-3}</td>
<td>1.735</td>
</tr>
<tr>
<td>Polystyrene</td>
<td>1.06</td>
<td>0.538</td>
<td>H 49.851, C 50.149</td>
<td>2.545 \times 10^{-3}</td>
<td>1.735</td>
</tr>
<tr>
<td>PMMA</td>
<td>1.185</td>
<td>0.539</td>
<td>H 53.333, C 33.333, O 13.333,</td>
<td>2.271 \times 10^{-3}</td>
<td>1.735</td>
</tr>
<tr>
<td>Lexan</td>
<td>1.20</td>
<td>0.527</td>
<td>H 42.424, C 48.485, O 9.091</td>
<td>2.310 \times 10^{-3}</td>
<td>1.735</td>
</tr>
<tr>
<td>Bone substitute plastic</td>
<td>1.829</td>
<td>0.516</td>
<td>H 35.215, C 29.592, N 0.803, O 26.695, Cl 0.16, Ca 7.679</td>
<td>1.666 \times 10^{-3}</td>
<td>1.730</td>
</tr>
<tr>
<td>Aluminum</td>
<td>2.698</td>
<td>0.482</td>
<td>Al 100</td>
<td>1.364 \times 10^{-3}</td>
<td>1.719</td>
</tr>
<tr>
<td>Titanium</td>
<td>4.519</td>
<td>0.459</td>
<td>Ti 100</td>
<td>9.430 \times 10^{-4}</td>
<td>1.710</td>
</tr>
<tr>
<td>Stainless steel</td>
<td>7.85</td>
<td>0.466</td>
<td>C 0.045, N 0.045, Si 0.450, Cr 18.150, Mn 1.250, Fe 71.460, Ni 8.550, Mo 0.050</td>
<td>5.659 \times 10^{-4}</td>
<td>1.706</td>
</tr>
<tr>
<td>Lead</td>
<td>11.322</td>
<td>0.396</td>
<td>Pb 100</td>
<td>6.505 \times 10^{-4}</td>
<td>1.676</td>
</tr>
<tr>
<td>Gold</td>
<td>19.311</td>
<td>0.401</td>
<td>Au 100</td>
<td>3.705 \times 10^{-4}</td>
<td>1.677</td>
</tr>
</tbody>
</table>
Our discussion of WET measurement methods is very brief, mainly because they are relatively simple and obvious. In practice, however, WET measurements remain very important (Andreo 2009, Gottschalk 2010a, Newhauser and Zhang 2010, Zhang et al. 2010b, Besemer et al. 2013, Moyers et al. 2010).

The IAEA (2000) proposed that WET can be approximated by
\[
t_{\text{w}} = t_{\text{m}} \frac{\rho_{\text{w}}}{\rho_{\text{m}}} c_{\text{m}},
\]
where the depth scaling factor, \(c_{\text{m}}\), can be calculated, to a good approximation, as the ratio of the continuous slowing down approximation (CSDA) range (in g cm\(^{-2}\)) in water to that in the target:
\[
c_{\text{m}} = \frac{\rho_{\text{w}} R_{\text{w}}}{\rho_{\text{m}} R_{\text{m}}}. \tag{32}
\]

Because the ranges in equation (32) correspond to complete loss of ion energy, this approach is strictly valid only for stopping-length targets. An exact equation for WET that is applicable to thin targets was reported by Zhang and Newhauser (2009):
\[
t_{\text{w}} = t_{\text{m}} \frac{\rho_{\text{w}} \overline{S}_{\text{w}}}{\rho_{\text{m}} \overline{S}_{\text{m}}}, \tag{33}
\]
where \(\rho_{\text{w}}\) and \(\rho_{\text{m}}\) are the mass densities of water and material, respectively, and \(\overline{S}_{\text{m}}\) and \(\overline{S}_{\text{w}}\) are the mean proton mass stopping power values for the material and water, respectively, defined by

![Figure 10. Calculated WER values as a function of proton beam energy. This plot illustrates the dependence of WER value on the target material, the beam energy, and the target thickness (reproduced with permission from Zhang and Newhauser 2009).](image-url)
For thin targets, where the proton loses a negligible fraction of its energy in the absorber material, we have

$$S = \frac{\int S dE}{\int dE}.$$  \hspace{1cm} (34)

For thin targets, where the proton loses a negligible fraction of its energy in the absorber material, we have

$$t_w \approx t_m \left( \frac{\rho_m}{\rho_w} \right) \left( \frac{S_m}{S_w} \right) = t_m \frac{\alpha_n P_w}{\alpha_n P_m} E^{p_n - p_w},$$  \hspace{1cm} (35)

where the reader will recognize $\alpha$ and $p$ from the discussion of the BK rule (see section 2.1). Values of $\alpha$ and $p$ for commonly encountered materials in proton therapy are provided in table 2. Zhang and Newhauser (2009) also reported a slightly more complex analytical formula to calculate WET for targets of arbitrary thickness.

As can be seen in the curves of the water-equivalent ratio (WER = $t_w/t_m$) plotted in figure 10, taking into account the target thickness in calculating WER is most important for absorbers that are thick and made of high Z materials, e.g. lead scattering foils, and for protons that are of comparatively low energy when impinging on the target. For low-Z materials, such as tissue and plastic, WER depends only very weakly on the target thickness and initial proton beam energy, and the approximate (thin and stopping length) analytical methods provide sufficient accuracy for most clinical applications.
Figure 12. Comparison of depth-dose curves from proton SOBPs and from electron beams. Because the proton mass is nearly 2000 times that of an electron, proton scattering interactions (individual angular deflections and variations in collisional energy losses) are much smaller, leading to sharper lateral and distal falloff distances (reproduced with permission from Koehler and Preston 1972).

Figure 13. Absorbed dose $D$ as a function of depth $z$ in water from an unmodulated (pristine) proton Bragg peak produced by a broad proton beam with an initial energy of 154 MeV. The various regions, depths, and lengths that are labeled are defined in the text. (The electronic buildup is not visible in this plot.) This type of dose distribution is useful clinically because of the relatively low doses delivered to normal tissues in the sub-peak and distal-falloff regions relative to the target dose delivered by the peak.
3.2. Features of pristine and spread-out Bragg curves

The spatial dose distribution from clinical proton therapy beams is quite similar to those from photon and electron beams. The lateral profiles are generally quite flat in the central high-dose region, then fall off rapidly in the penumbral regions, where the penumbra width increases with depth in the patient. The central-axis depth-dose curve from protons is somewhat similar to that from electrons, but with a sharper distal falloff. Figures 11 and 12 compare the central-axis depth-dose curves from several radiation therapy beams, revealing the main dosimetric properties that are clinically advantageous in many cases, namely, relatively low entrance dose, large and uniform dose to cover the tumor, and rapid falloff of dose near the end of range to spare normal tissues. These properties, together with a uniform lateral dose profile and a sharp lateral penumbral width, allow proton beams to treat a wide variety to tumor sizes and locations while providing superior sparing of normal tissues in many cases.

Having casually inspected the shape of proton depth-dose curves, we next examine their structure in greater detail, pointing out nomenclature and the physical processes that govern the shape of various regions. Figure 13 shows a pristine proton peak along with labels identifying several regions. In order of increasing depth, these are the regions of electronic buildup, protonic buildup, sub-peak, peak, and distal falloff. The figure also shows several characteristic depths (e.g. the depth zBP at which the peak occurs) and various characteristic lengths (e.g. the 80%-to-20% distal-falloff length l_{80-20} and the proximal-80%-to-distal-80% pristine-peak width).

The anatomic definitions of an SOBP are, in many ways, similar to those of a pristine Bragg curve, as seen in figure 14. However, there are several unique difficulties in characterizing SOBPs because of their sometimes unusual shape. For example, SOBPs with two or more
discrete pristine Bragg curves may have multiple dose maxima in the modulated-peak region (e.g. the ripple shown in figure 15). Because of such problems we introduce a few additional quantities that are defined only for SOBPs. To a large extent, however, we have defined quantities and terminology that are common to both modulated and pristine Bragg curves.

We have not yet mentioned how MCS affects the shape of the depth–dose curves. In fact, near the central region of a laterally ‘large’ beam, or more correctly well inside the periphery of a large beam, there is an equilibrium in which lateral scattering away from the central axis is exactly compensated by scattering towards it. This effect is described in figure 16, which is adapted from Koehler and Preston (1972). As the field size shrinks to the dimension of the rms lateral displacement due to MCS, lateral equilibrium is lost and MCS progressively depletes the proton fluence and dose along the central axis. Small proton beams have been investigated in several studies, including those by Takada (1996), Moyers et al (1999), Vatnitsky et al (1999b), Bednarz et al (2010), and Gottschalk (1999), as well as others, especially in the context of scanned beams and pencil beam dose algorithms.

Here, we use a Cartesian coordinate system with the \( z \) axis parallel to and centered about the proton beam central axis. The \( x \) and \( y \) axis are mutually orthogonal and perpendicular to the \( z \) axis. The coordinate system origin is located at the front face of the absorber, e.g. the extended medium in which we consider the absorbed dose distribution.

**Pristine Bragg curve**: A depth-dose distribution in an absorber irradiated with a monoenergetic or nearly-monoenergetic proton beam. In other words, no device or technique has been intentionally deployed for modulating the proton fluence or spectral fluence.

**Spread-out Bragg curve**: A depth-dose distribution in an absorber irradiated with a beam that has been intentionally modified to increase the axial dimension of the peak region. This is accomplished by modulating the range and the fluence of the beam. Clinical systems accomplish this by combining multiple quasi-monoenergetic beams or with a continuously modulated beam.

Figure 15. Absorbed dose \( D \) as a function of depth \( z \) in water from a spread-out Bragg peak (SOBP) (uppermost curve) and its constituent pristine Bragg peaks (lower curves; for clarity, all but the deepest pristine Bragg peak are only partly drawn). In many cases, the clinical target volume is larger than the width of a pristine Bragg peak. By appropriately modulating the proton range and fluence of pristine peaks, the extent of the high-dose region can be widened to cover the target volume with a uniform dose.
Electronic buildup region: A small region near the surface of the absorber where the proton beam is incident. As discussed in section 2.1, high-energy proton beams liberate delta rays with sufficient kinetic energy to travel several millimeters in tissue. Under some circumstances, this region exhibits an increase of dose with increasing depth, asymptotically approaching absorbed dose in the sub-peak region within the depth corresponding to the range of the most penetrating recoil electron. In some cases, electronic buildup is not observed. There are several possible reasons for this: the presence of some material just upstream of the surface (e.g. an immobilization device or a range compensator) may provide partial or full electronic charged particle equilibrium in the absorber, it may occur in combination with protonic buildup, it may be masked by changes in the proton energy loss rate near the end of range, or the wall of a cavity dosimeter may be sufficiently thick to present electronic equilibrium to the dosimeter’s sensitive volume.

Protonic buildup region: A region near the surface of the absorber where the absorbed dose increases with depth because of the buildup of secondary protons that are attributable to proton-induced non-elastic nuclear interactions (e.g. $^{16}$O(p, xp) reactions). As with electronic buildup, the protonic buildup may not be observed in some cases, particularly at low incident proton beam energies.

Sub-peak region: The region extending from the surface of the absorber to the depth just proximal of the peak. The physical processes involved here are, in decreasing order of importance, the stopping power’s dependence on the inverse-square of the proton velocity, the removal of some protons by nuclear reactions, the liberation of secondary particles from nuclear reactions, and for very small fields, the accumulation of lateral deflections from MCS leading to lateral protonic disequilibrium and reduction of the proton fluence on the central axis. The distal extent of the sub-peak region can be calculated from $z_m - 2\sigma$, where $z_m$ is the depth at the pristine Bragg peak and $\sigma$ is the width of the peak. The width parameter $\sigma$ can be estimated from the incident proton beam spectral fluence and the range straggling accumulated in the absorber, as discussed in section 2.3.

Pristine Bragg peak: The pristine Bragg peak is simply the maximum (or mode) dose near the end of range, and is located at $z_{BP}$, which is defined next. The physical processes governing
the location and/or height of the peak are mainly the proton stopping power and energy straggling, nuclear reactions to a much lesser extent and, for very small fields, MCS.

**Pristine Bragg peak depth:** The depth near the end of range of the primary protons at which the protons produce the maximum dose rate, denoted by _z_BP_. Although small proton beams are not yet widely used, it is helpful to define the location of _z_BP_ in a way that is compatible with large and small beams. Figure 16 shows that the maximum dose for beams of diameter larger than 6 mm is clearly single valued and located near the end of range. For smaller beams, however, the dose at the peak near the end of range may be less than the dose in the proximal regions, creating multiple maxima to choose from. Hence, the definition of _z_BP_ restricts it to exist in the region of the _R ± 4σ_, where _σ_ is the distal falloff width, thereby preventing possible ambiguities, and makes _z_BP_ conceptually independent of the beam cross-sectional area.

**Distal falloff region:** This region extends from depths greater than that of the pristine Bragg peak depth, _z_BP_. The width of this region is not restricted. In many practical situations, however, the distal falloff region can be truncated can be truncated at a depth where the dose falls below a threshold value, e.g. 1% of the dose at the Bragg peak, _D_(z_BP).

**Distal-50% depth:** The distalmost depth, denoted by _z_d50_, at which the absorbed dose is equal to half of the absorbed dose at the pristine Bragg peak depth, or _D_(z_BP)/2. For an SOBP, it is defined as the distalmost depth at which the absorbed dose is equal to half of the absorbed dose at the SOBP dose in the modulated peak region. In many cases, the dose in the modulated peak region varies with depth (perhaps by design), making selection of an absorbed dose value in the modulated peak region arbitrary. Since this clearly hinders an unambiguous definition of _z_d50_, we instead define _z_d50_ for an SOBP as being equal to _z_b_. If the dose in the modulated peak region varies with depth, _z_b_ may have to be determined with numerical methods. Physically, the value of _z_b_ is closely related to the _R_CSDA_ value of the mean proton energy corresponding to the most energetic peak in the SOBP. Definitions for various other distal depths are similarly defined, e.g. the distal-90% depth _z_d90_ and _z_d20_.

**Proximal-50% depth:** The second most distal depth, denoted by _z_p50_, at which the absorbed dose is equal to half of the absorbed dose at the pristine Bragg peak depth, or _D_(z_BP)/2, provided that occurs within the absorber. (For very low-energy pristine Bragg curves, the entrance dose may be greater than half the peak dose). In such cases, and for SOBPs, it is defined as the most proximal depth at which the absorbed dose is equal to half of the absorbed dose in the peak region. Because of problems that are analogous to those described in the _z_d50_ definition above, _z_p50_ is defined as the location at which the dose is equal to half the value at depth _z_a_ + _Δ_2, where _Δ_ is the width of the transition from the peak region to the sub-peak region. _Δ_ is not a critical parameter and it may be estimated as the amount of range straggling at _z_a_ or it can be determined from measured Bragg curves. Conceptually, _z_a_ typically corresponds to the shallowest location that is expected to receive the maximum dose. Physically, the value of _z_a_ is closely related to the _R_CSDA_ value of the least energetic pristine peak in the SOBP curve (or the only peak in the Bragg curve for a pristine peak). In many cases, particularly for SOBPs with large modulated-peak regions, the absorbed dose throughout the sub-peak region exceeds 50% of the value at _z_a_ + _Δ_, in which case proximal 50% depth is undefined. In such cases, proximal depths at higher dose percentages, such as _z_p95_ and _z_p90_, may be used. If the Bragg peak occurs at zero depth, as is commonly the case for treating superficial tumors, one may simply use _z_p100_ = 0.

**80% to 20% distal-falloff length:** The distance between the distal-80% and distal-20% depths, denoted by _l_d80-d20_. Other distal-falloff lengths are similarly defined, e.g. the 90%-to-10% distal-falloff length _l_d90-d10_.

**Proximal-80% to proximal-80% pristine-peak width:** The distance between the proximal-80% depth and the distal-80% depth, denoted by _l_p80-d80_, where the proximal-80% depth is defined as the proximal-80% dose rate, denoted by _D_(z_p80), which is the dose rate at the proximal-80% depth.
and distal-80% depths are defined analogously to the distal-50% depth, \( z_{d80} \), already defined. Other pristine peak widths are similarly defined, e.g. the 80%-to-80% pristine peak width. In cases where the Bragg curve does not include a proximal 90 dose point, the methods described in ‘Proximal-50% depth’ may be used.

**Proximal-80%-to-distal-80% modulated peak width:** The width of the modulated peak region is defined as \( l_{mod, d80-p80} = z_{d80} - z_{p80} \). In cases where the Bragg curve does not include a proximal 80% dose point, the methods described in ‘Proximal-50% depth’ may be used.

**Modulated-peak region:** The region extending from \( z_a \) to \( z_b \). In general, the values of \( z_a \) and \( z_b \) are most reliably determined using iterative numerical fitting methods. Conceptually, they are closely related to the proton ranges of the most and least penetrating pristine peaks in the SOBP.

These definitions may initially appear pedantic and overly precise for a conceptual understanding of proton Bragg curves. However, experience has shown that these definitions facilitate quantitative analysis and reporting of the characteristics of a wide variety of Bragg peaks in clinical and research settings. The definitions were developed from experience in manual and algorithmic analyses of measured clinical pristine and spread-out Bragg curves (Newhauser 2001a, 2001b, Newhauser et al 2002a, 2002b) and for developing and commissioning proton dose algorithms for treatment planning purposes (Koch and Newhauser 2005 2010, Newhauser et al 2007b). Finally, we note that in some clinical situations, the practicing medical physicist may need to define and use additional parameters, e.g. \( z_{p98} \) and \( z_{d98} \), as appropriate to a particular clinical situation or protocol. Regardless of the particular parameters chosen, it is difficult to overstate the paramount importance of using the parameters in a consistent way and being clear about their meaning when reporting results.

### 3.3. Model of pristine Bragg curves

In the preceding sections, we described all the major physical processes that govern the shape of proton dose distributions. Pristine Bragg curves may be calculated using a wide variety of techniques, from look-up tables of measured data to Monte Carlo simulations to analytical models. In the early days of proton therapy, indeed through the 1990s, dose algorithms in most proton treatment planning systems included very few, if any, physical models in their dose calculation algorithms. However, in the last 15 years of so, much progress has been made in modeling proton dose distributions with increasing physical completeness, realism, and accuracy. In this section, we review a method to calculate a pristine Bragg curve using a physics-based analytical model. For brevity, we present only one model, a model that describes many of the important physical processes, is computationally straightforward and fast, and is of considerable practical value in a clinical proton setting.

Bortfeld (1997) proposed an analytical equation to calculate the Bragg curve for proton energies between 10 and 200 MeV, as follows:

\[
D(z) = \frac{\phi_0}{\Gamma(1/p)} \frac{e^{-\zeta^2/\sigma^2}}{2\pi \sigma^p} \left[ 1 - D_{1+p}(-\zeta) - \frac{1}{\sigma} D_{1+p}(-\zeta) + \frac{\lambda}{1 + \lambda} D_{1+p}(-\zeta) \right],
\]

where \( D(z) \) is the depth dose, \( z \) is the depth, \( \phi_0 \) is the primary fluence, \( R_0 \) is the range of the proton beam, \( \sigma \) is the standard deviation of the Gaussian distribution of the proton depth, \( \zeta = (R_0 - z)/\sigma \), \( \alpha \) and \( p \) are material-dependent constants (defined in equation (2)), \( \varepsilon \) is the fraction of low-energy proton fluence to total proton fluence, \( \Gamma(x) \) is the gamma function, and \( D_\lambda(x) \) is the parabolic cylinder function. Very good agreement is found between measured and
calculated curves, and this approach has been used for the analysis and characterization of clinical proton beams (figure 17) (Newhauser et al 2002b) and for treatment planning dose calculations (Szymanowski and Oelfke 2003, Soukup et al 2005, Rethfeldt et al 2006, Li et al 2008).

3.4. Model of spread-out Bragg curves

Spread-out Bragg curves may be designed by combining multiple pristine Bragg curves. This approach was described in Wilson’s seminal paper and its implementation was described by Koehler et al (1977). Figure 15 plots several pristine Bragg curves and their resultant sum, which was similar to the SOBPs used at Harvard from the early 1970s onwards. The fluence is also modulated, which can be seen in the same figure as variations in the relative peak dose of each pristine Bragg curve. The range is typically modulated in steps that are small compared with the width of the Bragg peak, so that the SOBP contains little or no ripple. The modulation step size may be fixed or variable; generally the smallest step increments are needed for the deepest peaks, e.g. the amplitude of the ripple decreases with depth because an increasing fraction of dose is from the sub-peak region of the Bragg curves. Figure 18 plots several typical clinical SOBPs from a contemporary proton therapy system.

SOBPs can also be modeled with analytical methods. Bortfeld and Schlegel (1996) proposed a model of the form

\[
D(z) = \begin{cases} 
\frac{3}{4} + \frac{\sqrt{3}}{4\pi} \ln \left( 1 + \frac{\hat{z}^2}{1 - \hat{z}^2} \right) - \frac{3}{2\pi} \arctan \frac{2\hat{z} - 1}{\sqrt{3}} & : \text{for } z < z_a \\
1 & : \text{for } z_a \leq z \leq z_b \\
0 & : \text{for } z > z_b
\end{cases} \quad (37)
\]

where the dimensionless depth \( \hat{z} \) is given by

\[
\hat{z} = \frac{z}{z_a}
\]
and where $z_a$ and $z_b$ denote the depths of the proximal and distal extents of the SOBP (see preceding section for details of their definitions) and $z$ is the depth in water. The model was derived with several simplifying assumptions and approximations that limit its ability to predict real SOBPs. In particular, the proton energy loss rate is based on the CSDA (i.e. energy straggling is neglected) and the converted energy per unit mass approximation (where the proton energy loss and the absorbed dose in the water are assumed proportional), and non-elastic interactions of the protons with the absorber nuclei were not considered. Consequently, it does not reproduce real SOBPs well.

To attain a more realistic model, one of us (WN) introduced terms for the finite distal-falloff distance associated with range straggling, protonic buildup effects, arbitrary slope of the modulated-peak region, and a term for the ripple in the modulated peak. In addition, the piecewise use of the function for various regions was modified to allow for a transition region between the sub-peak and modulated-peak regions, which eliminated a pronounced discontinuity there. With these additional terms and modified regions, the model becomes

$$D(z) = c_0 A_P \begin{cases} c_1 + \frac{\sqrt{c_5}}{4\pi} \ln \frac{1 + \xi^2}{1 - \xi^2} - \frac{c_3}{2\pi} \arctan \frac{c_5}{\sqrt{c_3}} & : \text{for } z \leq z_a - \Delta \\ m(z + b) + Y & : \text{for } z_a - \Delta < z < z_a + \Delta \\ (mz + b) + Y & : \text{for } z \geq z_a + \Delta \end{cases}$$

(39)
where $\xi$ is defined as before, $c_0$ is a proportionality constant, $\Lambda$ takes into account the dependence on the dose due to inverse-square-law reductions in the proton fluence, $f_{BU}$ is the secondary-particle buildup term, $c_1$ through $c_5$ are coefficients that influence the shape of the sub-peak region, and $m$ and $b$ are the slope and intercept of the modulated-peak region, respectively. The ripple in the modulated-peak region is described by the term $Y$, e.g., an underdamped harmonic oscillator. The distal falloff term is denoted by $\gamma$ and is modeled with a cumulative normal PDF. The transition region is centered about $z_a$ and is of width $\Delta$, which is a non-critical parameter that can be estimated from the amount of range straggling expected at depth $z_a$ (see previous section for detailed definition) or it may be empirically deduced from measured Bragg curves. The extended model agrees well with measured SOBPs, several of which are shown in figure 18. This work was previously unpublished.

3.5. Analytical transport of therapeutic proton beams

Analytical models to predict therapeutic radiation are generally simple, fast, and accurate. Early treatment planning systems, many of which were still in use in the 1990s, used dose algorithms that looked up measured profiles of absorbed dose and/or used empirical formulas. Over the last 15 years, advances in our understanding of proton therapy physics, as well as huge gains in computer speed and memory capacity, have spurred the development and clinical use of increasingly realistic and complete physical models to predict absorbed dose in patients.

The basic approach underlying most analytical dose algorithms is that the proton beam’s energy loss and lateral scattering may be modeled independently from one another. The physical basis for this assumption lies in the principle of conservation of momentum and energy. Lateral scattering, which is predominantly caused by MCS, occurs at small angles (see section 2.4). Hence, the energy loss associated with the vast majority of these small scattering events is negligible. Conversely, the predominant proton energy-loss mechanism, namely inelastic Coulomb interactions with atomic electrons, only negligibly deflects the protons because the proton rest mass is 1832 times greater than that of an electron (see section 2.1). Consequently, for an excellent approximation, the 3D distribution of absorbed dose in a phantom or patient may be cast in the form of

$$D(x, y, z) = D(z) \times \text{OAR}(x, y, z),$$

where $x$ and $y$ are lateral coordinates, $z$ is the depth coordinate along the beam axis, $D(z)$ is the large-field Bragg curve (i.e., with lateral protonic equilibrium, described in section 3.2), and $\text{OAR}(x, y, z)$ is a term that takes into account the lateral properties of the dose distribution. Broad beam algorithms use equation (40) with simple ray-casting techniques (Siddon 1985) to determine the energy loss and range along each ray. Similarly, to determine the lateral extent of the therapeutic field, the rays are terminated in target-shaped final collimators by approximating them as black-body absorbers (Hong et al 1996). Carefully designed broad beam algorithms are simple, fast, and accurate in media that are of relatively uniform mass density and material composition. Recently, Koch and Newhauser (2010) developed an analytical broad beam algorithm of the form

$$\frac{D(x, y, z)}{\text{MU}} = \sum_{i=1}^{N} \frac{DD(z_{\text{eff}})}{\text{MU}} \omega_{i}^{\text{SAF}} \omega_{i}^{\text{RMW}} \left( \frac{\text{SSD} + \xi_{\text{eff}}}{\text{SSD} + z} \right)^{2} \text{OAR}_{i} \cdot \Gamma_{i},$$

where $N$ is the number of range modulation steps, $\omega^{\text{SAF}}$ is the scatter and attenuation fluence weighting factor, $\omega^{\text{RMW}}$ is the range modulator wheel (RMW) fluence weighting factor,
OAR is off-axis ratio, and $\Gamma$ is collimation modifier. This algorithm explicitly takes into account edge-scattered protons and range straggling, predicts absolute dose per monitor unit ($D / MU$) values, and provides excellent agreement with measurements in homogeneous media, e.g. for ocular treatments. In highly heterogeneous media, however, the dose at some locations may depend strongly on the material properties of the heterogeneities in the overlying tissue, and therefore the approach of casting a single ray does not, in general, provide sufficient dosimetric accuracy.

To improve the dosimetric accuracy in heterogeneous media, the dose may be calculated at a point as the superposition of multiple pencil beams. A pencil beam is, in essence, a mathematical construct that has no exact and direct analogy in the physical world, but it resembles a beam of infinitesimally small lateral size. By superposing enough pencil beams, the physical dose distribution can be modeled. Fortunately, as we shall see, many of the physical principles and algorithmic aspects from the broad beam approach are directly applicable to the pencil beam algorithm.

Hogstrom et al (1981) originally proposed a pencil beam algorithm for electron therapy. Basically, an ion beam is divided into a number of finite pencil beams and the dose distribution of each pencil beam is described by the multiplication of a central-axis term and an off-axis term. The final relative absorbed dose to any point in the patient ($D_p(x, y, z)$) is the summation of the dose contributions from all the pencil beams:

\begin{figure}
\centering
\includegraphics[width=\textwidth]{Figure19.png}
\caption{(Upper) Geometric model of proton therapy unit and the voxelized phantom oriented for the superior spinal proton field. The beam delivery system includes a vacuum window (A), a beam profile monitor (B), a range modulator wheel (C), a second scatterer (D), a range shifter assembly (E), backup and primary monitors (F), the snout (G), the range compensator (H), the treatment couch (I), and the patient (J) (reproduced with permission from Zhang et al 2013a). (Lower) Simulated proton transport through the proton therapy unit and the patient.}
\end{figure}
\[
D_p(x, y, z) = \int dx' dy' S(x', y') \left( \frac{\text{PDD}(z_{\text{eff}}) (\text{SSD} + z_{\text{eff}})}{\text{SSD} + z} \right)^2 \frac{1}{2\pi \sigma_{\text{tot}}^2(z)} \exp \left( -\frac{(x' - x)^2 + (y' - y)^2}{2\sigma_{\text{tot}}^2(z)} \right),
\]

(42)

where \( S(x', y') \) is the relative intensity of the pencil beam at \( x', y' \), \( \text{PDD}(z) \) is the central-axis percentage depth dose, \( \text{SSD} \) is the source-to-surface distance, \( z_{\text{eff}} \) is the effective depth of the point, and \( \sigma_{\text{tot}}^2 \) describes the final lateral distribution of a pencil beam by taking into account the MCS contributions from each beam-modifying device and from the patient. Many other groups have subsequently adapted and extended this concept for application in proton therapy (Petti 1992, Hong et al 1996, Schaffner et al 1999, Szymanowski and Oelfke 2002, Ciangaru et al 2005, Schaffner 2008, Westerly et al 2013). Contemporary proton pencil beam algorithms provide excellent accuracy (Newhauser et al 2007b), especially in homogeneous media, superior to that of broad beam algorithms in heterogeneous media. The improvement in accuracy comes mostly at the cost of greater computation times. Because of approximations in the pencil beam algorithm, it may not provide sufficient accuracy in extremely heterogeneous media, at material and/or density interfaces, or in other complex situations.

Currently, most proton therapy clinics utilize commercial treatment planning systems that contain pencil beam algorithms for calculation of therapeutic dose distributions. Computation speeds are generally sufficiently fast for most routine planning.

### 3.6. Monte Carlo transport of therapeutic proton beams

Monte Carlo transport technique has been used increasingly in radiation therapy (Seco and Verhaegen 2013). It is more accurate than analytical models in that it takes into account the physical processes during particle transport. The most commonly used general purpose Monte Carlo codes for proton therapy research are MCNPX (Pelowitz 2011), Geant 4 (Agostinelli et al 2003), and FLUKA (Ferrari et al 2005). Several groups have developed special purpose Monte Carlo codes, e.g. a treatment planning dose engine (Tourovycky et al 2005, Yepes et al 2009a), typically with faster computation speed but fewer features than general purpose codes. Here we focus mainly on applications that use the MCNPX code because it is representative of most general purpose codes, we are familiar with it, and constraints on journal space do not allow discussion of all codes.

The main components of a Monte Carlo simulation model are a source of protons, a treatment head, and a patient or phantom (figure 19). The proton source is typically taken as being incident on the treatment head (beamline transport is not usually simulated), where the source parameters are deduced from dose profile measurements (Newhauser et al 2005) or taken from other beam properties, e.g. from the manufacturer (Newhauser et al 2007b). The model of a treatment head typically includes all major mechanical components, e.g. a vacuum window, a beam profile monitor, a range modulator wheel, a second scatterer, shielding plates, a range shifter assembly, backup and primary monitors, the snout and the brass collimator, and a patient-specific range compensator (Zheng et al 2007b, Perez-Andujar et al 2009). Models of scanned beams include magnetic fields and transport the proton beam deflection (Peterson et al 2009, Dowdell et al 2012), or the proton source may be pointed, e.g. using the ‘thin lens’ approximation to increase computation speed (Lee 2004). The patient may be represented in voxels whose material composition and mass density are deduced from 3D CT images of the patient (Taddei et al 2009). The matrix of Hounsfield unit values in these voxels may be converted into corresponding matrices of material composition and mass density values based
on a tri-linear calibration curve (Newhauser et al 2008a). A simulation capability to accommodate time-dependent geometry was reported by Paganetti et al (2005).

Some useful applications of the Monte Carlo method are described here:

(a) Alternative or complimentary source of dosimetric data for developing, configuring, and validating analytical dose algorithms in clinical treatment planning systems. At least two studies revealed that, after a significant development effort, a Monte Carlo model of a proton therapy apparatus is sufficiently accurate and fast for prospective commissioning of an ocular treatment planning system (Koch and Newhauser 2005, Newhauser et al 2005) and a large-field treatment planning system (Newhauser et al 2007b). This prospective capability reduced the time required for preclinical work by several months, resulting in substantial cost savings. Studies about scanning proton beam configuration have also been reported (Tourovsy et al 2005, Sawakuchi et al 2010, Dowdell et al 2012).

(b) Alternative dose engine to calculate absorbed dose in phantoms or patient. In clinical practice, there frequently arise challenging cases where the accuracy of analytical models is unknown or suspected to be insufficient. In many such cases, measurements are not possible or practical, e.g. intracranial in vivo range verification. Multiple groups have developed Monte Carlo treatment planning engines (Tourovsy et al 2005, Newhauser et al 2008b, Perl et al 2012, Mairani et al 2013) based on various Monte Carlo codes, independently proving the technical feasibility of this approach. The physics of general purpose Monte Carlo codes for this purpose has been exhaustively validated against benchmark measurements in therapeutic proton beam data (Polf 2007, Kimstrand et al 2008, Titt et al 2008a, Randeniya et al 2009). Just 10 years ago, the slow execution times, memory constraints, unknown accuracy, and unknown technical feasibility of a full-blown Monte Carlo engine were serious research challenges. Today, the only major remaining obstacle to routine use of Monte Carlo engines for treatment planning is their slow execution times. For example, a proton craniospinal treatment simulation took more than 10³ CPU hours (Taddei et al 2009). Encouragingly, however, algorithmic improvements have increased speeds dramatically; Zhang et al (2013a) found that using lattice tally in MCNPX can speed up dose simulations by one order of magnitude compared to mesh tally, especially for dose reconstructions involving large numbers of voxels. The physics of the radiation transport are identical for the lattice and mesh tallies, i.e. there was no loss of accuracy. Yepes et al (2009b) recently reported on a fast dose calculator, a Monte Carlo track-repeating algorithm that uses a database of pre-computed proton trajectories in water and applies these trajectories to heterogeneous media by scaling the proton range and scattering angles. This approach reduced the computation time for dose reconstruction in voxelized patient geometry by more than two orders of magnitude compared to MCNPX and GEANT4 (Yepes et al 2009a, 2010a). The use of low-cost alternative parallel computer architectures also appears promising, including the use of graphics processing units (Yepes et al 2010b, Jia et al 2012) and grid technologies (Vadapalli et al 2011).

(c) Assessment of beam perturbation due to objects implanted in patients and development of mitigating strategies. The proportion of cancer patients with implanted devices is probably less than 10%, but it is increasing because of the aging of the population, increasing life expectancies, and increasing use of implanted devices. Examples include cardiac pacemakers, defibrillators, fiducial markers for radiotherapy localization, permanent radioactive seed implants, stents, prostheses, surgical reconstructive devices, and foreign bodies such as gunshot and shrapnel. Analytical dose algorithms in contemporary clinical treatment planning systems lack the capability to reliably assess dose perturbations from

(d) Investigation of characteristics of therapeutic beams and beamlines. For example, edge-scattered protons can degrade the dose distribution, and the effect is difficult to model with current analytical models. Accurate predictions of D / MU require taking into account the dose from protons scattered from the edge of the patient-specific collimator, particularly for small field sizes and large depths (Titt et al 2008b). Currently available spot scanning system only offer few options for adjusting beam spot properties like lateral and longitudinal size, and Monte Carlo simulations was used to optimize scanned beam spot (Titt et al 2010).

Because the Monte Carlo simulation technique is inherently general, it can be applied to virtually any problem involving the transport of protons or secondary radiation. Space constraints prevent us from reviewing more than a few illustrative examples. In later sections we briefly mention roles for Monte Carlo methods in research on stray radiation exposures to patients (section 5) and shielding barriers to protect staff (section 6).

4. Therapeutic absorbed dose determination

4.1. Reference dosimetry

By reference dosimetry, we mean the determination of absorbed dose in a manner that allows it to be directly related or referred to an accurate and uniform standard of absorbed dose. Clinical reference dosimetry comprises the measurement of absorbed dose in a clinic, which is related to the absorbed dose at a primary or secondary standards laboratory. This approach ensures that clinical reference dosimetry is accurate and uniform across participating institutions. Typically, clinical reference dosimetry is established by first calibrating a clinic’s dosimeter at a standards laboratory, then ‘transferring’ the calibration to the clinic’s treatment beams. To minimize systematic errors introduced by this transfer process, both irradiations are made with the same dosimeter and under identical (or very similar) ‘reference conditions’. Therefore, reference conditions must be reproducible and clinically relevant. However, national or international calibration laboratories do not yet produce proton calibration beams of relevance to proton therapy (presumably due to prohibitively high costs, low demand, and limited resources). Because of this limitation, the proton therapy community developed alternative methods for proton reference dosimetry.

For many decades, the reference fields for calibrating proton dosimeters were characterized using a Faraday cup (Verhey et al 1979) to measure the proton fluence in air. Today, most proton therapy institutions implement reference proton dosimetry utilizing an ionization chamber technique to measure the absorbed dose to water. With the latter technique, an ionization chamber is calibrated using reference conditions for photon therapy (i.e. 60Co radiation fields that are widely available at calibration laboratories) and a correction factor is applied that corrects the differences in the chamber’s response to 60Co and proton beams. When implemented properly, the techniques agree within uncertainties (Newhauser et al 2002a, 2002b).

Several advisory bodies have published dosimetry protocols for reference dosimetry, such as the American Association of Physicists in Medicine (AAPM 1986), the European Clinical Heavy Particle Dosimetry Group (ECHED) (Vynckier et al 1991, 1994), the International Commission on Radiation Units and Measurements (ICRU 1998, 2007), and the International Atomic Energy Agency (IAEA 2000). Vatnitsky et al (1999a) performed an international
proton dosimetry inter-comparison based on the ICRU Report 59 protocol (ICRU 1998) and reported that absorbed dose to water can be delivered within 3%.

4.2. Patient field–specific dosimetry

The determination of absolute absorbed dose per monitor unit for each patient and treatment field, denoted by \( \frac{D}{MU} \), rests upon the foundation of reference dosimetry. Recall that reference dosimetry is performed under simple and reproducible conditions, e.g. using a dosimeter in a water-box phantom. In contrast, dosimetry in a patient should take into account the full complexity of a patient’s anatomy, e.g. irregular surface shapes, heterogeneous elemental composition and mass density, and treatment field size and shape. These additional complexities call for a conceptual framework and measurement techniques for performing routine \( \frac{D}{MU} \) determinations for patient treatment fields.

For several decades, \( \frac{D}{MU} \) values for passively scattered treatment beams were typically measured at depth in a water-box phantom the range compensator present. However, a systematic study by Fontenot et al (2007) that measuring \( \frac{D}{MU} \) without the range compensator present provided more reliable results because use of the range compensator increased the severity of dose gradients near the calibration point and this resulted in larger overall uncertainties in \( \frac{D}{MU} \). It is not known if this finding holds for special cases, e.g. small diameter treatment beams, large air gaps, and moving targets. In many practical situations, it is convenient or necessary to predict the absolute dose in a patient based on the corresponding dose under reference conditions (e.g. a water phantom). Because an exact theoretical approach is not possible, all approaches have utilized approximate methods, numerical calculations, measurements, or a combination of these. On the basis of previous studies, Newhauser et al (2014a, 2014b) reported the following method to calculate proton dose in a patient \( \frac{D}{MU}_p \):

\[
\left( \frac{D}{MU} \right)_p = \left( \frac{D}{MU} \right)_w \cdot F_{\text{CPS}}
= \left( \frac{D}{MU} \right)_w^{\text{ref}} \cdot F_{\text{OF}} \cdot F_{\text{RS}} \cdot F_{\text{SOBP}} \cdot F_{\text{InvSq}} \cdot F_{\text{FS}} \cdot F_{\text{ColS}} \cdot F_{\text{CPS}},
\tag{43}
\]

where \( \left( \frac{D}{MU} \right)_w \) is the \( \frac{D}{MU} \) in a water phantom under patient-specific field conditions; \( F_{\text{CPS}} \) is the compensator scatter and patient scatter factor, which takes into account differences in the scattering and attenuation of the patient and range compensator together relative to that of water and no range compensator; \( \left( \frac{D}{MU} \right)_w^{\text{ref}} \) is the \( \frac{D}{MU} \) in a water phantom under reference conditions and is typically 1 cGy/MU; \( F_{\text{OF}} \) corrects for proton beam energy spectrum relative to the reference condition; \( F_{\text{RS}} \) corrects for range shifter effects; \( F_{\text{SOBP}} \) corrects for SOBP effects relative to the reference field; \( F_{\text{InvSq}} \) corrects for beam divergence relative to the reference condition; \( F_{\text{FS}} \) corrects for field size effect; and \( F_{\text{ColS}} \) corrects for differences in scatter from the reference aperture (e.g. 10 cm x 10 cm) to the patient-specific aperture. They found that for prostate treatment fields, uncertainty in \( F_{\text{CPS}} \), which is the least well understood factor, was clinically acceptable for prostate treatment (Newhauser et al 2014b). In a companion study, the same group reported the variability in \( F_{\text{CPS}} \) was clinically significant for some lung treatment fields (Newhauser et al 2014a).

For scanning proton beams, it is important to calibrate individual spot beams because patient specific collimating apertures or range compensators are not typically used in the beam line. First the number of protons per MU for each beam energy is calibrated by Faraday-cup measurements or Monte Carlo simulations. Then, either physical model or Monte Carlo simulation can be used to calculate absolute dose normalized to the number of incident protons. A
simple nuclear halo model was developed at Paul Scherrer Institute (PSI) (Pedroni et al 2005) to include secondary dose around the primary pencil beam and can predict precisely the dose directly from treatment planning without renormalization measurements.

In addition to the physical effects already discussed above, there are many additional effects that contribute to the uncertainty in $D / MU$ that arise from imperfections in anatomic images used for treatment planning, organ motion, and in the case of scanned beams, the interplay of organ motion and the moving beam, etc. Such uncertainties may depend strongly on anatomical site, approximations utilized by the treatment planning system, beam delivery system, and experimenter’s skill. Ideally, clinical treatment planning systems would provide accurate estimates of $(D / MU)_b^{ref}$ and $(D / MU)_b$ with known and small uncertainties, methods for determining and reporting $(D / MU)_b$ would be standardized, and secondary methods (e.g. hand calculations similar to equation (43) or fast Monte Carlo simulations) would be available to independently verify $(D / MU)_b$ values determined with the primary method. In principal, there appears to be an adequate knowledge of basic proton interaction physics and transport physics to reach or approach these clinical ideals. In addition, recent research studies report encouraging results for a variety of special cases, e.g. for a particular treatment technique, treatment planning system, or treatment delivery systems. As mentioned in section 3.5, Koch and Newhauser (2010) developed an analytical broad beam algorithm to predict absolute dose per monitor unit $(D / MU)$ values with good accuracy in water for ocular proton treatments, but the ability to predict $D / MU$ in the heterogeneous patient body still needs improvement. To date, the methods for determining $(D / MU)_b$ in the literature have not yet been validated for general application and they have not been standardized.

5. Stray radiation

As with other forms of external-beam radiation therapies, proton beams unavoidably produce stray radiation that impinges on the patient’s entire body (figure 20). In the 1990s, very little was known about the exposure of patients to stray radiation produced by therapeutic proton beams. By the mid 2000s, concerns about stray radiation had become a matter of considerable speculation and controversy, particularly regarding the suitability of proton beams for treating children with cancer because of concerns about the risks from stray neutron exposures (Hall 2006, Brenner and Hall 2008). By the late 2000s, research on stray radiation exposures from proton and other radiation therapies had intensified and some questions have been partially or fully answered. However, many questions remain open. In this section, we review selected developments and discuss a few open questions.

In a study using Monte Carlo simulations, Agosteo et al (1998) reported that neutron exposures were the principal concern among the various types of stray radiation from proton therapy. Comprehensive measurements in clinical proton therapy beams were reported by Yan et al (2002), including neutron spectral fluences and dose-equivalent data for large-field, radiosurgery, and ocular beamlines. That study included measurements of neutron energy spectra and neutron dose-equivalent values. Several other groups have measured neutron exposures, including Roy and Sandison (2004), Tayama et al (2006), Mesoloras et al (2006), Schneider et al (2002) and Wroe et al (2007). The early measurements were important because they suggested that the neutron exposures, while generally small in comparison to therapeutic doses, should not be neglected.

Several groups developed neutron dose reconstruction systems based on general purpose Monte Carlo codes for a wide variety of clinical proton beamlines (Siebers 2000, Jiang et al 2005, Polf and Newhauser 2005, Fontenot et al 2005, Newhauser et al 2007b, 2008b, Zheng...
et al 2007b, Moyers et al 2008). A wide variety of simulation approaches have been used, with varying degrees of clinical realism, assumptions, and approximation, and not surprisingly, rather disparate results. Some of the literature on measurements and simulations of neutron exposures was reviewed in NCRP Report 170 (2011). In the remainder of this section, we emphasize progress subsequent to that report.

Research methods of relevance to stray radiation estimation have progressed dramatically since Agosteo et al (1998) published their pioneering study using Monte Carlo simulation to study neutron fluence and dose in air or in a simple water box phantom. One key advance was the validation of Monte Carlo predictions against benchmark neutron measurements, for example the studies of Tayama et al (2006), Fontenot et al (2005a), and Polf et al (2005).

High-performance computing techniques, such as using massively parallel computing clusters, allows fast simulations of whole-body dose reconstructions involving complex treatment techniques, e.g. craniospinal irradiation (CSI) (Newhauser et al 2009, Taddei et al 2009 2010b, Perez-Andujar et al 2013a, Zhang et al 2013c), which is widely considered one of the most challenging treatments to simulate. Another key advance was increasing the level of realism in modeling anatomy. Methods have progressed in 1998 from simple water-box phantoms to stylized anthropomorphic phantoms to generic voxelized phantoms in the mid 2000s to patient-specific personalized phantoms by the late 2000s (i.e. by retrospectively using treatment planning CT scans). In fact, this year a study was completed comprising a 17-patient in silico clinical trial comparing proton CSI and photon CSI, according to the contemporary standards of care, including full Monte Carlo simulations of stray and therapeutic exposures (Zhang et al 2014).
Another research result of considerable clinical relevance is that it is possible and indeed feasible to dramatically reduce exposures from stray neutrons that leak out of the treatment apparatus. For example, Taddei et al (2008) found that modest modifications of the treatment unit (adding shields near the patient, using a tungsten-alloy collimator, and adding an upstream collimator near the range shifter assembly) substantially reduced the neutron dose (by 66%) for patients receiving passively scattered proton therapy for prostate cancer. Similarly, Brenner et al (2009) investigated various pre-collimator/collimator combinations with different geometries and materials and concluded that an optimized design can be achieved to significantly reduce the stray neutron dose.

Despite the rapid growth of research on stray neutron radiation exposures to patients, the literature is mainly limited to a few case studies and anatomical sites. In general, the knowledge of stray radiation from proton therapy is still largely incomplete. The gaps in knowledge received considerable attention following the cautionary paper by Hall (2006), who suggested that the use of advanced radiotherapy modalities like proton therapy may not be justified because of incomplete knowledge of stray neutron exposures and second cancer induction. From subsequent studies, a coherent picture is gradually emerging: second cancer risk following proton therapy is lower than that after photon therapy, the risk differential depends strongly on anatomic site and other host and treatment factors, the risk posed by stray radiation is small but not negligible, and the largest reductions in risk are achieved by using scattered- or scanned-proton beams instead of photon beams. Furthermore, the majority of risk of radiogenic second cancer is from in-field radiation, not out-of-field radiation. Consequently, the difference in predicted total risk (from both in-field and out-of-field radiation) from scattered versus scattered proton treatments is small and clinically negligible. At the risk of oversimplification, the theoretical risk advantage of proton therapy derives mainly from its ability to spare healthy tissues that are distal and lateral to the target volume. With judicious treatment techniques (e.g. beam orientations, location of field junctions, etc.), it is possible in many cases to utilize the rapid distal and lateral dose falloff of proton beams to reduce the dose to sensitive organs and tissues to below levels achievable with technologically comparable photon treatments.

It is worth underscoring that our knowledge and understanding of this topic is incomplete and likely to evolve substantially in the future. Current dose and risk assessments are mostly based on standard-of-care treatment plans, i.e. using clinical planning and research systems that do not attempt to algorithmically minimize risk of radiation side effects. As clinical treatment planning systems are incrementally extended with these capabilities, the standards of care for proton and photon therapies will continue to evolve. This evolution will likely render our current understanding of radiation risks obsolete.

5.1. Monte Carlo simulations

During the transport of proton beams through the treatment unit to the patient, not only is therapeutic radiation deposited in the patient but also stray radiation is generated because of nuclear interactions (as explained in section 2.4). In proton therapy, secondary neutrons are dominant among this stray radiation and are a major concern (Agosteo et al 1998, Fontenot et al 2008). Estimation of stray neutron dose can be challenging and time consuming, and the variations in measurement or calculation results can be large (Xu et al 2008).

The inherent accuracy of Monte Carlo simulation makes it an irreplaceable tool for stray neutron dose estimation, and indeed this technique has been used by many investigators studying stray neutrons generated during proton therapy (Agosteo et al 1998, Jiang et al

Polf and Newhauser (2005) found that neutron dose equivalent per therapeutic absorbed dose \((H/D)\) at different locations around a passively scattered proton therapy unit increased with increasing range modulation and that the major source of neutrons shifted from the final collimator to the range modulation wheel. Zheng et al (2007a, 2007b) reported that \(H/D\) generally increased with decreasing collimating aperture size, increasing proton beam energy, and increasing SOBP width, while it decreased with snout distance from the isocenter and increasing distance from the treatment unit. More than 50% of the neutron dose at all locations was from neutrons with energies higher than 10 MeV. Zheng et al (2008) also reported that neutron spectral fluence contained two pronounced peaks, a low-energy peak around 1 MeV and a high-energy peak that ranged from around 10 MeV up to the proton energy. The mean neutron radiation weighting factors varied only slightly, from 8.8 to 10.3, with proton energy and location for a closed-aperture configuration.

5.2. Analytical model

Because of the complexity associated with Monte Carlo simulation and measurement, a simple analytical model to predict stray neutron dose in proton therapy is desirable. Zheng et al (2007b) proposed an exponential decay model to predict \(H/D\) in air with good accuracy. This largely empirical model was extended by Zhang et al (2010a) to predict \(H/D\) both in air and in the patient’s body by taking into account off-axis effect, phantom attenuation effect, and low/high energy peaks observed in neutron spectral fluence. Perez-Andujar et al (2013b) subsequently enhanced the model by replacing empirical components with physics-based components, obtaining

\[
(H/D)_{\text{iso}} = (H/D)_{\text{iso}} (d/d_{\text{iso}})^\rho \left[ C_1 e^{-(\sigma_1 + s_1) d_{\text{iso}}} + C_2 e^{-(\sigma_2 + s_2) d_{\text{iso}}} + \ldots \right],
\]

where \((H/D)_{\text{iso}}\) is the neutron equivalent dose per therapeutic absorbed dose at isocenter. \(C_1, C_2, C_3,\) and \(C_4\) apportion the \(H/D\) contributions from intranuclear cascade, evaporation, epithermal, and thermal neutrons, respectively. \((d/d_{\text{iso}})^\rho\) represents the power law that governs the neutron dose falloff as a function of distance from the effective neutron source, which is independent of the proton beam energy, \(d\) represents the distance from the neutron source in the treatment nozzle to the detecting volume and \(d_{\text{iso}}\) is the distance from the neutron source to isocenter. \(\sigma_1, \sigma_2, \sigma_3,\) and \(\sigma_4\) denote the attenuation coefficients of intranuclear cascade, evaporation, epithermal, and thermal neutrons, respectively. \(d_{\text{iso}}\) is the distance from the phantom surface to the isocenter, and \(d\) is the distance from the phantom surface to the detecting volumes. The lateral distribution of neutrons is governed by \(s_1\) for intranuclear cascade, \(s_2\) for evaporation, \(s_3\) for epithermal, and \(s_4\) for thermal neutrons. \(z\) is the longitudinal coordinate for the neutron dose receptor and is used to scale the width parameters.

The accuracy of this model is comparable to the accuracy of typical Monte Carlo simulations or measurements of neutron dose (figure 21). Taddei et al (2013) modified the model reported by Zhang et al (2010a) and used a simplified double-Gaussian model to calculation
Figure 21. Results from Monte Carlo simulations and analytical model predictions of equivalent per therapeutic dose ($H/D$) for proton beams with various energies in the water phantom as a function of vertical depth ($a$) and lateral distance ($b$) (reproduced with permission from Perez-Andujar et al 2013b).
out-of-field dose in photon therapy and they also had good agreement between measurement data and model-based values.

With additional research and development on stray dose algorithms, it will be possible to implement them in treatment planning systems and use them for routine clinical purposes.

6. Shielding design

In this section, we will briefly review the design of shielding to protect humans from stray radiation. In particular, we will focus on the bulk shielding barriers (e.g. walls, ceilings) and mazes. Bulk shielding is an extremely important aspect of facility design because proton therapy equipment is capable of producing lethal levels of stray radiation. In addition, bulk shielding intrudes on space available for equipment and personnel and it is expensive. In many ways, the design of bulk shielding is a classical engineering problem; one develops a solution that comprises an acceptable balance of the competing attributes of safety, utility, and cost.

Broadly, to design shielding one must have knowledge of the stray radiation production, transport, and attenuation in shielding barriers. In addition, the shielding design goals, i.e. the permissible exposure rate at an occupied location, depend on the fraction of time an area is occupied, its designation as a controlled or uncontrolled area, and the type of individuals present, i.e. patients, staff (radiation workers) and the general public. There are typically multiple design goals to be satisfied, e.g. one for short term exposures (averaged over one hour) and one for long term exposures (averaged over one year). In general, one does not know a priori which design goal will ultimately govern the shielding design, necessitating shielding calculations that represent multiple scenarios.

Before discussing the basic physics of shielding design, we digress briefly to point out that shielding barriers are one of several necessary components of an effective radiation protection program. Shielding barriers alone do not ensure safety. It is impractical, not to mention prohibitively expensive, to build shields that by themselves provide adequate protection for unrestricted usage of contemporary proton therapy systems. Administrative controls on beam usage are therefore necessary and, for obvious reasons, the anticipated beam usage must be taken into account in the shielding design process. Knowledge of the proton beam usage is often difficult to predict, particularly for first-of-kind treatment units where neutron production in the beam production and delivery systems are poorly known. In addition, the future outcome of a clinical trial or a change in health care policy may dramatically change proton beam usage. Leaving these uncertainties aside, one needs to know not only the beam usage (energies, charges and currents, and directions of travel) at each of the treatment locations, but also all of the corresponding proton loss rates in the accelerator, energy selection system (if present), beam transport system, and treatment head. The loss proton losses and neutron production vary strongly with the proton beam energy and other factors (see figure 22), necessitating separate calculations for several proton beam energies (Newhauser et al 2002d). Many aspects of shielding of proton therapy facilities are similar to those for photon therapy facilities, which have been reviewed in detail elsewhere (NCRP 2008).

On the assumption that the beam usage, occupancy factors, and other aspects are known, one may calculate the neutron dose equivalent rate at a point of interest behind a slab shielding barrier (e.g. wall, ceiling, etc.) that is parallel to and at a distance $a$ the proton beam according to

$$H(E_p, \theta, \frac{d}{\lambda(\theta)}) = \frac{H_0(E_p, \theta)}{r^2} \exp \left[ -\frac{d}{\lambda(\theta) g(\alpha)} \right],$$

(45)
where $H(E_p, \theta, d/\lambda)$ is the ambient dose equivalent beyond the shield, $E_p$ is the proton beam energy, $\theta$ is the emission angle from the proton loss point and the point of interest, $r$ is the distance between the proton loss point and the point of interest, $d$ is the thickness of the shield.
The attenuation length of the shield material, $\lambda(\theta)$, is the source term, $\alpha$ is the angle of incidence of the radiation impinging on the shield, and $g(a)$ is the corresponding obliquity factor. This semi-empirical approach, based on the work of Moyer (1962), has many attractive attributes, including the inclusion of major dependencies, simplicity, and computational speed. Five decades on, in spite of its limitations, the Moyer model still is widely used, especially in cases where speed and convenience are more important than accuracy.

Personnel must be able to ingress to and egress from access shielded vaults quickly, e.g. for efficient routine clinical operation, in emergencies such as attending to a sick patient, and to maintain and repair the accelerator and beam transport system. This necessitates large openings in the bulk shielding walls. Obviously such openings result in a reduction in attenuation that must be restored. Typically the required attenuation is restored by fitting the opening with a long and curved shielding tunnel, commonly referred to as a maze or labyrinth. The maze attenuates radiation that is incident upon it in two ways: its walls attenuate the deeply penetrating radiation that is incident on the maze walls, and it attenuates the less-penetrating radiation that enters the opening of the maze by a combination of scattering and absorption processes as the radiation propagates through of the maze.

Typically the mazes of treatment rooms are not equipped with massive shielded doors because they can significantly increase access times. On the other hand, infrequently accessed vaults (e.g. for the accelerator and beam transport system) commonly utilize a maze equipped with a massive shielding door.

The dose equivalent rate outside the maze (without a door) from scattered radiation that enters the maze entrance (inside the vault) is given by

$$H(r_1, r_2, r_3, ..., r_n) = \frac{2\alpha^2 H_o(a)}{r_1^2} \prod_{i=2}^{n} f_i(r_i),$$

(46)

where $r_1$ is the line of sight distance from the source to the point of interest in the first (closest to the source) leg of the maze maze, $H_o(a)$ is the dose equivalent at distance $a$ from the source, $a$ is distance from the source to the entrance of the first leg, and $f_i(r_i)$ are attenuation factors for the subsequent legs of the maze (Agosteo 2009). These are given by

$$f_i(r_i) = \frac{e^{-r_i/0.45} + 0.022A_i^{1.3}e^{-r_i/2.35}}{1 + 0.022A_i^{1.3}}.$$

(47)

where $r_i$ is the distance from entrance of the $i$th leg to the point of interest in that leg, and where $A_i$ is the cross sectional area of the $i$th leg of the maze (Agosteo 2009). In addition, one must still (separately) consider attenuation of radiation impinging upon the outside of the maze walls (e.g. using the Moyer model) and the attenuation in maze doors, in cases where they are used. The attenuation of the maze increases with the number and length of the legs, the thickness of the maze walls, and decreases with cross sectional area. It would be difficult to overemphasize the high relative importance of the design of the mazes. Fortunately, there are available several design methods that are well understood, accurate, and that have been validated with measurements in clinical proton therapy facilities.

Materials used for the bulk shielding barriers and mazes vary somewhat, depending on the cost and availability of the shielding materials and the cost of space occupied by the shielding barriers. Typically, ordinary concrete with steel reinforcement is utilized because of its large hydrogen content and mass density (2.35 g cm$^{-3}$), high strength, good availability, and low cost. Contemporary proton therapy centers use concrete shielding barriers of up to several meters thick (Newhauser et al 2002, Titt and Newhauser 2005). The neutron attenuation
length, or the thickness required to reduce the incident fluence by a factor of $1/e$, is plotted in figure 23 for ordinary concrete. At the high-energy limit, the attenuation length is approximately 50 cm of ordinary concrete (Moritz 2001). The density and attenuating properties of concrete can be increased substantially by utilizing high-density aggregate material, but its higher cost prohibitive in most cases. Some facilities have utilized comparatively small ‘local’ shields, e.g. to shield beamline components that produce copious quantities of neutrons, to lessen the attenuation requirements of larger bulk shields. In some cases, higher density materials such as alloys of iron (7.9 g cm$^{-3}$) or tungsten (19.25 g cm$^{-3}$) may be advantageous for shielding against high-energy neutrons because they can be made more compact than with concrete or other bulk shielding materials. Proton beam stoppers, which prevent the primary proton beam from impinging on the bulk shielding, have been made from a variety of materials, including plastic and steel.

The shielding design methods most often used for proton therapy facilities are analytical formulas and Monte Carlo simulations. The knowledge of the uncertainties in predictions from both of these methods is incomplete. Newhauser et al (2002c) measured, calculated (using semi-empirical analytical models that were developed for slab and maze shielding geometrics), and simulated neutron dose rates at a 235 MeV proton therapy center, and they found that the analytical model overestimated neutron dose at most positions compared to measurements, while Monte Carlo simulations agreed more closely with the measurements. However, the Monte Carlo method is considerably more challenging because of the need to obtain convergence of solutions; generalized and automated variance reduction techniques are lacking. Titt and Newhauser (2005) evaluated the uncertainty associated with Monte Carlo shielding technique by comparing the analytical predictions with detailed Monte Carlo simulations of neutron equivalent doses at various receptor locations. They found the optimum rejection criterion to be 10% statistical uncertainty for the Monte Carlo simulations. This rejection criterion provided additional confidence because virtually all accepted simulated results had converged.

In an unpublished study, one of us (WN) estimated the cost of concrete and steel used for shielding a typical multi-room proton therapy at approximately 6 M USD, with the potential

Figure 23. The attenuation length of neutrons in concrete versus proton beam energy (reproduced with permission from Agosteo et al 2007).
to reduce shielding costs by almost one third through improved neutron shielding designs. Evidently there is considerable potential to achieve cost savings and other benefits by developing improved shielding design methods and to optimize shield designs.

Although we have not discussed small local shields of beamline components in this section, there is considerable overlap with the material presented in the previous section, for brevity we shall not consider stray radiation inside of a treatment room here.

7. Challenges and future of proton therapy

The future utilization of proton therapy is difficult to predict. There are some tentative indications that it will continue to become more widely available in developed countries, perhaps owing to the clear theoretical dosimetric advantages associated with proton beam dose distributions. Perhaps the strongest argument in favor of protons is actually an indirect counter argument: there is little evidence of beneficial effect from unnecessary irradiation of healthy tissues (i.e. from the exit dose of photon beams) (Terasawa et al 2009). Other arguments in favor include an excellent record of treatment safety and efficacy. Furthermore, over the past 15 years, many arguments against proton therapy have been largely or fully disproved, e.g. it is too complex and too difficult for most clinical organizations, manufacturers cannot be counted upon to deliver systems on schedule, the range uncertainty is too large, and the risks from stray neutron doses are too great.

Nonetheless, skepticism and controversy remain regarding the ultimate role of proton therapy in radiation oncology. Especially in the last few years, the debate seems to focus on cost-effectiveness and cost-competitiveness. Basically, the argument goes, the cost and value of proton therapy has not been proven with evidence of improved patient outcomes, which are presumed to offset some or all of the higher costs of proton therapy systems. If the price differential between proton and photon therapies were to substantially shrink or disappear, e.g. due to economies of scale, many clinics would replace at least some photon treatment units with proton units.

For all these reasons, there are many urgent research questions of relevance to proton therapy. Some of these are enumerated here, with an emphasis on those questions that will require physics research and development to reduce cost, improve treatment quality and efficiency, and create previously new treatment capabilities of clinical importance.

(a) Can novel techniques, such as proton arc therapy (Sandison et al 1997, Sengbusch et al 2009, Rechner et al 2012), be developed to improve the quality of treatment, reduce treatment time, and increase cost-competitiveness and -effectiveness?
(b) Can cost-competitiveness or treatment capability be increased significantly through incremental improvements to existing accelerator technologies, e.g. fixed-field alternating gradient synchrotron (Johnstone et al 1999) and superconducting cyclotron accelerators (Blosser et al 1997), or novel linear accelerators, e.g. wakefield laser accelerators (Schwoerer et al 2006) and dielectric wall accelerators (Caporaso et al 2008)?
(c) Can ultra-compact low-energy proton accelerators provide hand-held or robotically mounted radiation sources for treating superficial tumors or intra-operative treatment of deep-seated tumors?
(d) How can range uncertainties be quantified and minimized? In cases where uncertainties are presently too large, what roles will prospective imaging play, including proton CT (Schulte et al 2004) and megavoltage photon CT (Langen et al 2005, Newhauser et al 2008a, De Marzi et al 2013)? Similarly, what role will real-time or post-treatment imaging play, including prompt gamma imaging (Peterson et al 2010, Gueth et al 2013),
positron emission tomography (Parodi et al 2007, Cho et al 2013, Min et al 2013), proton radiography (Schneider and Pedroni 1995), and magnetic resonance imaging (Krejcarek et al 2007)?

(e) In the future, what out-of-field dose algorithms should be developed for treatment planning systems used for research or routine clinical practice? Techniques for the calculation and measurement of therapeutic dose are reasonably well established (with a few exceptions mentioned below). In most cases, intensity-modulation techniques have enabled the clinician to provide essentially identical dose distributions to the target volumes using either proton or photon therapy. Consequently, it appears that comparative treatment planning studies of the future will mainly focus on the radiation exposure of healthy tissues and on the feasibility of delivering novel treatments. Knowledge of the physics of stray radiation exposures is incomplete, and most commercial treatment planning systems neglect or very crudely approximate the out-of-field dose.

(f) What are the optimal strategies for treating moving tumors, especially in the thorax and abdomen? For example, will scanned beam treatments with target tracking be beneficial? How should organ motion be measured during treatment? Do the benefits of incremental margin shrinkage justify increased risk of interplay dose defects? Are there cases where passively scattered beams are superior?

(g) As the use of on-board imaging is expected to increase, how will additional radiation exposures to patients be monitored and minimized? Are the radiogenic risks of future imaging schemes justified by their benefit to the patient?

(h) What is the role of future multimodality radiation therapies that include protons? The objectives of such protocols include reducing skin dose, sharpening penumbral widths and distal falloff lengths, and reducing dose delivery errors. Important sources of errors include temporal interplay of the organ motion and beam location and range uncertainties due to image artifacts caused by metal implants. Possible implementations could include a source of protons and other particles, e.g. photons, on a single rotational gantry.

(i) Is there a need for a national or international primary standards laboratory to provide reference proton beams to support a standard for dose absorbed to water?

In the last 15 years or so, much progress has been made toward a complete understanding of the physics of proton therapy. In particular, today there exist analytical models and simulation techniques to design and model most of the major aspects of clinical proton therapy systems currently in operation, several of which we have reviewed in this paper. However, even with today’s state of the art knowledge of proton therapy, the answers to many questions of scientific interest and economic importance remain tantalizingly beyond the reach of current research capabilities. As in the past, clinical needs and economic forces will likely define many of the future research frontiers in proton therapy.

To fully exploit the advantages of proton beams to improve patient outcomes, it is clear that additional research is needed to optimize the current generation of proton therapy systems, to make new discoveries and translate breakthroughs into novel prototype research systems, to obtain a deeper and clinically applicable understanding of the relevant aspects of radiation biology, to improve the efficiency clinical operations (e.g. industrial and process engineering), and to generate theoretical and observational evidence to assess the comparative effectiveness and cost of proton therapy relative to other comparable treatments for a wide variety of diseases, anatomical sites, and outcome endpoints. The relative priority of these goals is a matter of subjective judgment and speculation, which we shall leave to the reader.

Much of the important research will require experts and specialists from disciplines such as accelerator physics, imaging physics, various engineering specialties, oncology, biology,
biostatistics and epidemiology, and computer science. Many basic and applied research studies are well suited to purely academic or clinical research environments and research teams. Other research studies will benefit from the formation of collaborative teams comprising members drawn from the academy, medicine, and industry.

Acknowledgments

This work was supported in part by the National Cancer Institute (award 1R01 CA131463-01A1), Department of Defense (award W81XWH-10-1-0170), and the Bella Bowman Foundation.

References

Bethe H 1930 Zur theorie des durchgangs schneller korpuskularstrahlen durch materie Ann. Phys. 5 324–400
Bohr N 1915 On the decrease of velocity of swiftly moving electrified particles in passing through matter Phil. Mag. 30 581–612
Boon S 1998 Dosimetry and quality control of scanning proton beams PhD Thesis University of Groningen, the Netherlands
Bragg W and Kleeman R 1905 On the alpha particles of radium and their loss of range in passing through various atoms and molecules Phil. Mag. 10 318–40
Breuer H and Smit B J 2000 Proton Therapy and Radiosurgery (Berlin: Springer)
Carnicer A et al 2013 Quantification of dose perturbations induced by external and internal accessories in ocular proton therapy and evaluation of their dosimetric impact Med. Phys. 40 061708
Chadwick M B et al 2011 ENDF/B-VII.1 Nuclear data for science and technology: cross sections, covariances, fission product yields and decay data Nuclear Data Sheets Vol 112 pp 2887–996
Desjardins L et al 2003 Treatment of uveal melanoma with iodine 125 plaques or proton beam therapy: indications and comparison of local recurrence rates J. Fr. Ophtalmol. 26 269–76
Duttenhaver J R et al 1983 Protons or megavoltage X-rays as boost therapy for patients irradiated for localized prostatic carcinoma. An early phase II/II comparison Cancer 51 1599–604
ENDF 2011 Evaluated Nuclear Data File (ENDF)/B-VII.1 (Upton, NY: National Nuclear Data Center) (www.nds.iaea.org/endf)
Goitein M and Jermann M 2003 The relative costs of proton and x-ray radiation therapy Clin. Oncol. 15 S37–50
Gottschalk B 1999 Physical parameters of small proton beams Handout at PTCOG 31st Conf. (Bloomington, IN)
Gragoudas E S et al 1982 Proton beam irradiation of uveal melanomas. Results of 5 1/2 year study Arch. Ophthalmol. 100 928–34
Halperin E C, Perez C A and Brady L W 2008 Perez and Brady’s Principles and Practice of Radiation Oncology 5th edn (Philadelphia, PA: LWW)
IAEA 1988 Radiological Safety Aspects of the Operation of Proton Accelerators (Vienna, Austria: IAEA)
Johnstone C, Wan W and Gurren A 1999 Fixed field circular accelerator designs Particle Accelerator Conf. (New York, USA) pp 3068–70
Kjellberg R N, Koehler A M, Preston W M and Sweet W H 1962a Stereotaxic instrument for use with the Bragg peak of a proton beam Confinia Neurol. 22 183–9
Landau L 1944 On the energy loss of fast particles by ionization *J. Phys. (USSR)* **8** 201–5
Lawrence J H *et al* 1958 Pituitary irradiation with high-energy proton beams: a preliminary report *Cancer Res.* **18** 121–34
Lee S Y 2004 *Accelerator Physics* 2nd edn (Singapore: World Scientific)
Li H S, Romeijn H E, Fox C, Palta J R and Dempsey J F 2008 A computational implementation and comparison of several intensity modulated proton therapy treatment planning algorithms *Med. Phys.* **35** 1103–12
Matsuura T *et al* 2012 Biological effect of dose distortion by fiducial markers in spot-scanning proton therapy with a limited number of fields: a simulation study *Med. Phys.* **39** 5584–91
Merchant T E 2009 Proton beam therapy in pediatric oncology *Cancer J.* **15** 298–305
Moteabbed M, Espana S and Paganetti H 2011 Monte Carlo patient study on the comparison of prompt gamma and PET imaging for range verification in proton therapy *Phys. Med. Biol.* **56** 1063–82
Moyer B J 1962 Method of calculation of the shielding enclosure for the Berkeley Bevatron *Proc. 1st Int. Conf. on Shielding Around High Energy Accelerators* (Saclay, France) pp 65–70
NCRP 1971 Protection against neutron radiation NCRP Report 38 National Council on Radiation Protection and Measurements, Bethesda, MD
NCRP 1990 Implementation of the principle of As Low As Reasonably Achievable (ALARA) for medical and dental personnel NCRP Report 107 Bethesda, MD
NCRP 1993 Limitation of exposure to ionizing radiation (supersedes NCRP Report 91) NCRP Report 116 Bethesda, MD
NCRP 2005 Radiation protection for particle accelerator facilities NCRP Report 144 Bethesda, MD
NCRP 2008 Structural shielding design and evaluation for megavoltage x- and gamma-ray radiotherapy facilities NCRP Report 151 Bethesda, MD
NCRP 2011 Second primary cancers and cardiovascular disease after radiation therapy NCRP Report 170 Bethesda, MD
Newhauser W D, Titt U and Dexheimer D T 2002d A perspective on shielding design methods for future proton therapy facilities Shielding Aspects of Accelerators, Targets and Irradiation Facilities-SATIF 6 (Stanford, CA: Nuclear Energy Agency)
Newhauser W D 2001a Dosimetry for the gantry beams at the Northeast Proton Therapy Center (NPTC): I. Dimensions and geometric relationships NPTC Internal Report
Newhauser W D 2001b proton field calibration measurements and analysis with the CalibrationSetUp Program (CSU) Version 1.1 Internal Report
Newhauser W, Giebeler A, Guo B and Zhang R 2014a Uncertainty in dose per monitor unit estimates for passively scattered proton therapy: II. The role of compensator and patient scatter in thoracic cases in preparation
Newhauser W D and Zhang R 2010 In reply to Dr. Gottschalk’s comment on ‘Calculation of water equivalent thickness of materials of arbitrary density, elemental composition and thickness in proton beam irradiation’ Phys. Med. Biol. 55 L31–2

Olch A J 2013 Pediatric Radiotherapy—Planning and Treatment (Boca Raton, FL: CRC Press)


Paganetti H 2012 Proton Therapy Physics (Boca Raton, FL: CRC Press)


Parodi K et al 2007 PET/CT imaging for treatment verification after proton therapy: a study with plastic phantoms and metallic implants Med. Phys. 34 419–35


Parodi K et al 2007 PET/CT imaging for treatment verification after proton therapy: a study with plastic phantoms and metallic implants Med. Phys. 34 419–35


Parodi K et al 2007 PET/CT imaging for treatment verification after proton therapy: a study with plastic phantoms and metallic implants Med. Phys. 34 419–35


Parodi K et al 2007 PET/CT imaging for treatment verification after proton therapy: a study with plastic phantoms and metallic implants Med. Phys. 34 419–35


Parodi K et al 2007 PET/CT imaging for treatment verification after proton therapy: a study with plastic phantoms and metallic implants Med. Phys. 34 419–35


Parodi K et al 2007 PET/CT imaging for treatment verification after proton therapy: a study with plastic phantoms and metallic implants Med. Phys. 34 419–35


Parodi K et al 2007 PET/CT imaging for treatment verification after proton therapy: a study with plastic phantoms and metallic implants Med. Phys. 34 419–35


Parodi K et al 2007 PET/CT imaging for treatment verification after proton therapy: a study with plastic phantoms and metallic implants Med. Phys. 34 419–35


Parodi K et al 2007 PET/CT imaging for treatment verification after proton therapy: a study with plastic phantoms and metallic implants Med. Phys. 34 419–35


Parodi K et al 2007 PET/CT imaging for treatment verification after proton therapy: a study with plastic phantoms and metallic implants Med. Phys. 34 419–35


Parodi K et al 2007 PET/CT imaging for treatment verification after proton therapy: a study with plastic phantoms and metallic implants Med. Phys. 34 419–35

Roy S C and Sandison G A 2004 Scattered neutron dose equivalent to a fetus from proton therapy of the mother Radiat. Phys. Chem. 71 997–8
Rutherford E 1911 The scattering of α and β particles by matter and the structure of the atom Phil. Mag. 21 669–88
Sauvat F, Binart N, Poiriat C and Sarnacki S 2009 Preserving fertility in prepubertal children Hormone Res. 71(suppl. 1) 82–6
Sawakuchi G O et al 2010 An MCNPX Monte Carlo model of a discrete spot scanning proton beam therapy nozzle Med. Phys. 37 4960–70
Schadt D and Elsasser T 2010 Heavy-ion tumor therapy: physical and radiobiological benefits Rev. Mod. Phys. 82 383–425
Scharf W H 1994 Biomedical Particle Accelerators (Woodbury, NY: AIP Press)
Schulte R et al 2004 Conceptual design of a proton computed tomography system for applications in proton radiation therapy IEEE Trans. Nucl. Sci. 51 866–72
Schwoerer H et al 2006 Laser-plasma acceleration of quasi-monoenergetic protons from microstructured targets Nature 439 445–8
Seco J and Verhaegen F 2013 Monte Carlo Techniques in Radiation Therapy (Boca Raton, FL: CRC Press)
Takada Y 1996 Study of small proton beam field Particles 18 A35
Tesch K 1982 The attenuation of the neutron dose equivalent in a labyrinth through an accelerator shield Part. Accel. 12 169–75
Vadapalli R, Yepes P, Newhauser W D and Lichten R 2011 Grid-enabled treatment planning for proton therapy using Monte Carlo simulations Nucl. Technol. 175 16–21
Vavilov P V 1957 Ionization losses of high-energy heavy particles Sov. Phys. JETP 5 749–51
Weber D C et al 2009 RapidArc, intensity modulated photon and proton techniques for recurrent prostate cancer in previously irradiated patients: a treatment planning comparison study Radiat. Oncol. 4 34
Wilson R 1946 Radiological use of fast protons Radiology 47 487–91
R209

Wroe A J, Cornelius I M and Rosenfeld A B 2005 The role of nonelastic reactions in absorbed dose distributions from therapeutic proton beams in different medium Med. Phys. 32 37–41
Zhang R et al 2013c Predicted risks of radiogenic cardiac toxicity in two pediatric patients undergoing photon or proton radiotherapy Radiat. Oncol. 8 184
Ziegler J F 1999 Stopping of energetic light ions in elemental matter J. Appl. Phys. 85 1249–72
Ziegler J F, Biersack J P and Ziegler M D 2008 SRIM—The Stopping and Range of Ions in Matter (Chester, MD: SRIM Co.)