

Passive Beam Spreading in Proton Radiation Therapy

Bernard Gottschalk

Harvard High Energy Physics Laboratory
38 Oxford St., Cambridge, MA 01238, USA

gottschalk@huhepl.harvard.edu
<http://huhepl.harvard.edu/~gottschalk>

DRAFT

October 1, 2004

Contents

1	Introduction	5
1.1	Prerequisites	5
1.2	Accuracy of the Method	6
1.3	Mathematical Style	7
1.4	Software Platform	7
1.5	Acknowledgements	8
2	Symbol Definitions and Physical Constants	9
2.1	Acronyms	9
2.2	Symbol Definitions	9
2.3	Physical Constants	10
2.4	Conversions	10
3	Preliminaries	11
3.1	Areal Density: g/cm^2	11
3.2	MeV	11
3.3	Mass Stopping Power	12
3.4	Fluence	12
3.5	Physical Dose	12
3.6	Biologically Effective Dose	13
3.7	Computing Dose from Fluence	14
3.8	Computing Dose Rate from Proton Current Density	15
3.9	Dose Rate in a Therapy Beam	15
3.10	Air Filled Ionization Chambers	17
3.11	Proton Kinematics	18
4	Gaussians	19
4.1	Small-Angle Approximation	19
4.2	1D Gaussian Probability Density	19
4.3	2D Gaussian Probability Density	21
4.4	Implications for Single Scattering	21
4.5	Error Function	23
5	The Cubic Spline	26
5.1	Description	26
5.2	Cubic Spline Interpolation	26
5.3	Using the Cubic Spline	27
5.4	Cubic Spline Fit	28
5.4.1	Choosing Initial Points	28
5.4.2	Optimizing the Fit	30
5.4.3	Pathologies	30
5.5	The Broken Spline	31
5.6	The Spline as a General-Purpose Function	32

6	Stopping	33
6.1	Stopping Theory	33
6.1.1	Qualitative	33
6.1.2	Quantitative	34
6.2	Range	35
6.3	Range Straggling	36
6.4	Measuring Range	37
6.5	Clinical Depth Prescription	39
6.6	Interpolating Range-Energy Tables	40
6.7	Finding the Energy Loss in a Degradar	42
6.7.1	Using Range	42
6.7.2	Using Stopping Power	42
6.7.3	In a Computer Program	43
6.7.4	Exercise	43
6.7.5	Error Conditions	43
6.8	Water Equivalence	43
6.8.1	Water Equivalence of Thin Degradars	44
6.8.2	Equivalence in a Finely Divided Stack	45
6.8.3	Measured Water Equivalence	45
6.9	Does the Order of Degradars Matter?	45
6.10	Energy Spread of Degraded Beams	46
6.11	Energy Stability of an Accelerator	47
7	Scattering	49
7.1	Central Limit Theorem	51
7.2	Molière Theory	51
7.2.1	Thin Scatterer	52
7.2.2	Energy Loss	53
7.2.3	Compounds and Mixtures	54
7.2.4	Fano Correction	54
7.2.5	Evaluation of the Integrals	55
7.2.6	Numerical Results	55
7.3	Gaussian Approximations	59
7.3.1	Hanson's Formula	59
7.3.2	Highland's Formula	59
7.3.3	Lynch and Dahl's Formula	62
7.4	Fortran Function Theta0	62
7.5	Maximum Scattering from a Given Material	62
7.6	Adding Discrete Scatterers	63
7.7	Summary	64
8	Nuclear Interactions	65
8.1	Terminology	65
8.2	Elastic Scattering from Free Hydrogen	66
8.3	Overview of Nonelastic Reactions	67
8.4	Local Energy Deposition Model	67
8.5	Nuclear Buildup	68
8.6	Predicting P	72
8.7	Measuring P	72
8.7.1	Simple Faraday Cup	72
8.7.2	Multi Layer Faraday Cup	72
8.8	Summary	75

9	Binary Degraders	77
9.1	High-Z and Low-Z Materials	77
9.2	Notation to Describe Problems	77
9.3	‘Forward’ Problem	79
9.4	Single Degradar Inverse Problem	80
9.5	Binary Degradar Inverse Problem	81
9.6	Does the Order of Materials Matter?	82
9.7	Summary	83
10	Degradar Stacks and Beam Spreading	84
10.1	Beam Spreading	84
10.2	Experiment of Preston and Koehler	85
10.3	Effective Origin z_0	86
10.4	Fermi-Eyges Theory	89
10.5	Summary	89
11	The Bragg Peak	90
11.1	Understanding the Bragg Peak	90
11.1.1	Stopping Power, Straggling, Beam Energy Spread	90
11.2	Theoretical Models	91
11.2.1	Beam Size	92
11.2.2	$1/r^2$ Effect	93
11.2.3	Tank Wall and Other Degradars	93
11.2.4	Dosimeter	93
11.2.5	Low Energy Contamination	93
11.2.6	Nuclear Interactions	94
11.2.7	Electron and Nuclear Buildup	95
11.2.8	Pathlength Effect	95
11.2.9	Checklist for Measuring the Bragg Peak	95
11.3	Commercial Equipment	96
11.3.1	Tank Size	96
11.3.2	Current <i>vs.</i> Charge Measurement	97
11.3.3	Beam Monitor <i>vs.</i> Reference Chamber	97
11.3.4	Scanning <i>vs.</i> Dosimetry Mode	97
11.4	Fitting the Bragg Peak	98
11.4.1	Overlying Material	98
11.4.2	Extrapolation to Zero	99
11.5	Opening the BPK File	99
11.5.1	Removing the Fluence	100
11.5.2	Adjusting the Energy	100
11.5.3	Normalizing the Bragg Peak	100
11.5.4	A Closer Look at Normalization	101
12	Designing a Modulator	104
12.1	Introduction	104
12.2	Computing the Weights	106
12.3	Alternative Procedure for Weights	108
12.4	Defining Modulation	109
12.5	Step Size	110
12.6	Compensation	110
12.7	Useful Energy Range	112
12.8	Beam Gating	113
12.9	Beam Current Modulation	114
12.10	Physical Arrangement of Modulators	114

13 Single Scattering Energy Stacking Example	116
14 Double Scattering	117
14.1 Introduction	117
14.2 The Contoured Scatterer	119
14.3 Beam Imperfections	121
14.4 Solving the Forward Problem	121
14.5 Scaling and the Generic Solution	124
14.6 The Useful Radius	125
14.7 Optimization	125
14.8 Returning to the Real World	127
15 Double Scattering Upstream Modulator Example	129
16 Beam Steering	130
17 Capabilities of Double Scattering Systems	131
18 Transverse Penumbra	132
19 Distal Falloff	133
20 Collimators	134
List of Tables	135
List of Figures	136
Bibliography	139

Chapter 1

Introduction

This writeup is a guide to designing passive beam spreading systems from first principles using just three ingredients:

- EM stopping theory: the Bethe equation and its consequences¹
- EM multiple scattering theory: Molière theory and its consequences
- Bragg peaks measured in water

Nuclear effects and some others are accounted for by using measured, rather than computed, Bragg peaks. Designs based on these methods are usually close enough to specifications for clinical use, though they should be checked of course. They also provide reasonably accurate estimates of dose rates and beam monitor output factors, though the latter *must* be refined experimentally.

This writeup is a companion to our design programs. The Fortran source code for these programs may be downloaded from the Web site on the title page. Follow the instructions in the README file. Note the Disclaimer of Warranty.

1.1 Prerequisites

Our level of discussion aims at the college graduate in science or engineering. No background in particle physics is required. Indeed the kind of course that would help, an account of the basic interactions of the common particles with matter, is rarely taught nowadays. Algebra and the elements of calculus are required, of course. You must have a firm grasp of what is meant by a derivative or integral, but we shall only use the most elementary special functions and none of the advanced techniques of integration. If you can read Numerical Recipes [1] or the equivalent with good comprehension, you'll be fine.

We have tried to make this writeup self-contained. However, to keep it within bounds, we will discuss stopping and multiple scattering on an 'engineering' level without repeating the standard derivations. For these, you should go to the literature. In the same spirit, we will generally refer you to Numerical Recipes [1] for mathematical methods. Though there is a bibliography as a guide to further reading, a very few basic references will get you started:

- Numerical Recipes [1] unless you already know the material.
- A synopsis of stopping theory. The opening pages of ICRU49 [2] or Janni [3] are excellent.

¹ EM: electromagnetic. See chapter 2 for acronyms and symbol definitions.

- A synopsis of Molière’s multiple scattering theory. The HCL paper [4] is a good starting point. You will eventually need Bethe [5] if you really want to work through the derivation and formulas for the angular distribution. If you know German, by all means read the two papers by Molière [6, 7] which are beautifully written.

Concerning multiple scattering, be warned that the field is fairly tangled, and certain experimental and theoretical papers are not helpful [4].

The article by Chu et al. [8] is a good review of beam spreading techniques and instrumentation, broader in scope but with less emphasis on computation than the present report.

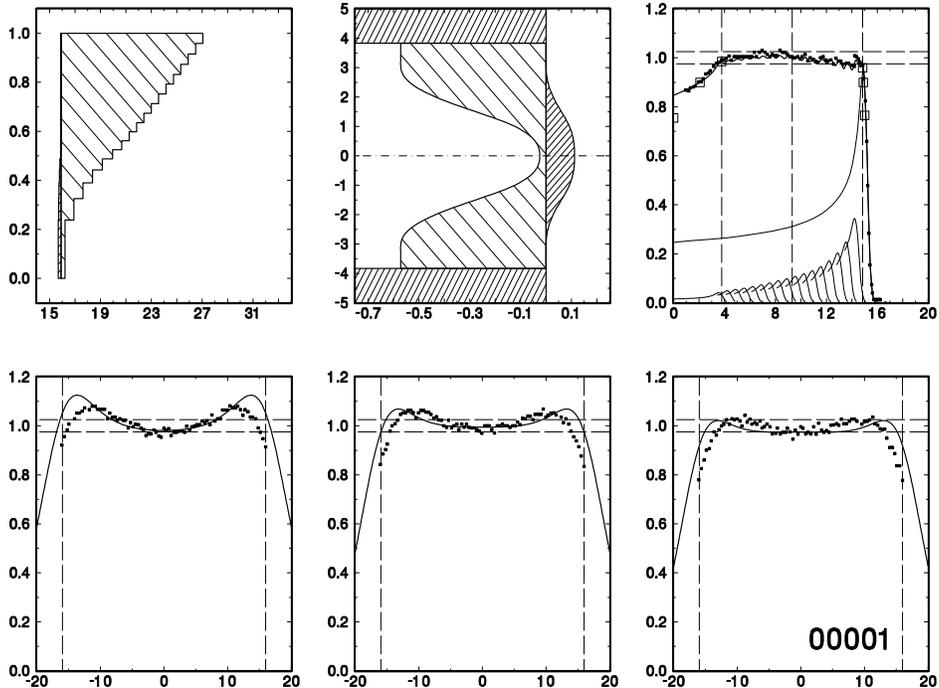


Figure 1.1: The first NEU design (1990), with experimental data.

1.2 Accuracy of the Method

How accurate are the methods described here? A double scattering system with a compensated upstream modulator and a compensated contoured second scatterer was designed, built and tested at the Harvard Cyclotron Laboratory (HCL) in 1990, using an early version of the program NEU. Figure 1.1 shows experimental data plotted against the prediction, using the 2004 version of NEU. A single normalization factor is used for all four panels. The depth-dose agrees extremely well with prediction. The transverse dose distributions agree fairly well.

Of course, this proof of principle experiment was not an optimum design. The long throw (distance between first scatterer and isocenter) precluded that. The modulator overscatters in its thicker parts, as evidenced in the design by the absence of compensating brass and in the transverse dose by the characteristic ‘dished’ appearance. (If you don’t understand the previous sentence, we hope

you will by the time you've finished this report.) One can do much better under ideal conditions. The essential point of Figure 1.1 is that the prediction agrees pretty well with experiment.

1.3 Mathematical Style

We will solve most problems by numerical techniques, rather than trying for answers in closed form. This reflects the general trend in applied science since computers took over. Though analytic results have a nice closed feeling and, if simple, may yield good insights, the numerical approach

- May allow us to avoid certain approximations;
- Does not require extensive knowledge of special functions;
- May lead to more efficient computation.

Regrettably, the description of a sequence of numerical procedures has a more amorphous feeling than a crisp sequence of equations.

Among the numerical techniques we will use are

- Cubic spline interpolation and fitting;
- Numerical integration (Simpson's Rule);
- Root finding (False Position or Bisection);
- Matrix inversion (Gauss-Jordan elimination);
- Optimization (Linear Least-Squares, Marquardt, Grid Search).

These are all covered by Numerical Recipes [1]. If you are already familiar with them you have a head start. The cubic spline is sufficiently central, and some of our applications sufficiently unusual, that we devote a chapter to it. Even here, we leave the mathematical details to Numerical Recipes.

1.4 Software Platform

We will frequently use program fragments in Fortran to describe procedures. We have adapted our programs, many quite old, to the Compaq Visual Fortran (VF) V6.6a compiler. You can use any recent PC running any recent version of Windows. Most applications, including the design of complicated double scattering systems, take just a few seconds. They are not Monte Carlo programs. They are written in Fortran 77 (fixed-format `.FOR`) as distinct from Fortran 90.

The programs are grouped into 'projects' (a project is a group of modules with a single `MAIN` program) all of which use the VF 'QuikWin' option. This offers simple but adequate graphics, which can be saved in bitmap `.BMP` format, and which we sometimes use to illustrate the text.

At this writing, we have made freely available certain end products in the form of executables (`LOOKUP` and `NEU` for example), as well as all of the source code except two routines from Numerical Methods. We do not plan to distribute the underlying project structure. You can use the Fortran source code as study material or as a basis for other projects. You may even be tempted to rewrite it in the computer language *du jour*, but it might be better to learn Fortran.

1.5 Acknowledgements

This report is a solo project but I owe my ability to write it to many persons associated with the Harvard Cyclotron Lab (HCL). I came there in 1956 as a graduate student, stayed till 1965 as a postdoc, and returned in 1981. By that time HCL was a full-time radiation therapy clinic. We treated our first patient on May 25, 1961 and our last on April 10, 2002: 9115 patients in 40 years. For several decades HCL patients accounted for more than half the world total.

Karl Strauch was my thesis advisor, a good friend who taught me the craft of experimental physics. Richard Wilson, another important teacher during my graduate years, became a staunch supporter of the clinical program and, most recently, wrote the history of HCL [9]. Bill Preston, an exceptionally kind man, was Director during most of the physics research years, then helped preserve the Cyclotron for medical use, and made important contributions to the applied physics of proton beams.²

Andy Koehler succeeded Bill Preston, but insisted on being ‘Acting’ Director for his entire long tenure. More than anyone else, he managed to promote local medical interest in proton beams and to pull the Cyclotron through its near-death experience in the late 1960’s. He made many important contributions to the technology of proton beams and his name will come up frequently in these notes. Above all, Bill Preston, Andy Koehler, and Miles Wagner, who followed them as Director, set a tone which made HCL a wonderful workplace. As Dick Wilson put it, HCL was a happy lab, and those who had the good fortune to work there will not forget the experience.

My friend Miles Wagner was my day-to-day colleague at HCL. We constantly bounced technical ideas off each other, and even though to this day our favorite solutions to any given problem tend to be diametrically opposed, each of us got the job done in his own way.

With Miles, Andy and me, Janet Sisterson completed the ‘gang of four’ that met from time to time to discuss the Big Picture and plan the future at HCL. Janet represented the physics research side of HCL during the medical years, collaborating in the measurement of many reaction cross sections of interest in space physics. She also served the particle therapy community for decades with her ‘Particles’ newsletter, providing a forum and listing particle therapy facilities worldwide. She has just now passed that job along.

Special thanks to Charles Mayo for his friendship and close collaboration, and to Rachel Platais, who built the first multi-layer Faraday Cup so well it worked right away. I also thank others at HCL: Gail Bradley, Jason Burns, Ethan Cascio, Lori Clarke, Alice Coggeshall, Dina Harp, Elliot Hammerman, Kris Johnson, Yefim Orsher, Jon Zoesman and Townsend Zwart. The Harvard Physics Department and the High Energy Physics Lab, personified by George Brandenburg, have given me a home and generous support post-HCL.

In memory of many coffee outings and some long days and late nights fixing ‘the machine’, I dedicate these notes to Miles and Andy.

² The rejection of his early manuscript with Andy Koehler [10] was a signal failure of the peer-review system.

Chapter 2

Symbol Definitions and Physical Constants

2.1 Acronyms

EM	electromagnetic
FC	Faraday Cup
FWHM	full width at half maximum
HCL	Harvard Cyclotron Laboratory (1949-2002)
IC	ionization chamber
ICRU	International Commission on Radiation Units and Measurements, Inc.
LHS	left-hand side of an equation
MLFC	Multi Layer Faraday Cup
MU	Monitor Unit(s)
NEU	Nozzle with Everything Upstream: a design program for scattering systems. (A play on the German for 'new'.)
NPTC	Northeast Proton Therapy Center
QA	quality assurance
RHS	right-hand side of an equation
RV	range verifier
SOBP	spread-out Bragg peak
VF	Visual Fortran

2.2 Symbol Definitions

A	cm^2	area
A		relative atomic mass
A_i	g	atomic weight, i^{th} constituent of compound
d	cm	ion chamber gap
D	Gy	physical absorbed dose
E	MeV	total energy
i_p	nA	proton current
i_{IC}	nA	ion chamber current
L	g/cm^2	scatterer thickness
L_R	g/cm^2	radiation length
N		number of protons
p	MeV/c	momentum
P	g/cm^2	pathlength

P		probability
P'	$\text{cm}^{-1}, \text{cm}^{-2}$	probability density
r_i		# atoms per molecule, i^{th} constituent of compound
R	g/cm^2	mean projected range
S	Mev/cm	stopping power = $-dE/dx$
t	g/cm^2	target thickness
T	MeV	kinetic energy
v	cm/sec	speed
W_m	MeV	largest possible energy loss to a free electron
x, y	cm	transverse coordinates
z		projectile charge number
z	cm	longitudinal coordinate
Z		target atomic number
β		proton v/c
ϵ		scattering system efficiency
ρ	g/cm^3	density
σ		<i>rms</i> deviation
σ_i	mB	reaction cross section, i^{th} constituent of compound
θ	radian	multiple scattering angle
θ_0	radian	characteristic multiple scattering angle
χ_c	radian	characteristic single scattering angle

2.3 Physical Constants

Taken from the Review of Particle Properties [11].

c	2.998×10^{10}	cm/sec	speed of light
	0.98	foot/nsec	
e	1.602×10^{-19}	Coulomb	quantum of charge
$e^2/\hbar c$	1/137.036		fine structure constant
$\hbar c$	197.327	Mev fm	conversion constant
$m_e c^2$	0.5110	MeV	electron rest energy
$m_p c^2$	938.27	MeV	proton rest energy
N_A	6.0221×10^{23}	mol^{-1}	Avogadro constant
r_e	2.818×10^{-13}	cm	classical electron radius
u	1.660×10^{-24}	g	unified atomic mass unit (mass ^{12}C atom)/12

2.4 Conversions

mb	millibarn	10^{-27} cm^2	reaction cross section
Gy	Gray	1 Joule/kg	physical absorbed dose

Chapter 3

Preliminaries

The purpose of this chapter is to introduce some basic quantities and ideas. Properly speaking, units should be expressed in the MKS (meter, kilogram, second) system but since we never compute in that system we'll express formulas in practical units. We'll assume the reader is familiar with the idea of *dimensional analysis* which is essential in checking and understanding equations.

3.1 Areal Density: g/cm²

The range of 160 MeV protons in water is 17.6 cm. In air it is 166 *meters*. The reason, of course, is that air is far less dense than water (830× less dense at standard conditions). To avoid such trivial density effects, and to see the deeper underlying trends in stopping power and scattering power, we almost always specify the thickness of degraders not as Δx (cm) but in terms of their *areal density*

$$\rho \times \Delta x \quad \text{g/cm}^2$$

where ρ (g/cm³) is the density of the degrader material. This takes some getting used to, but it will become second nature.

For three reasons, the areal density of a degrader is preferably measured by weighing it and dividing by its area, *not* by measuring thickness and multiplying by density. First, digital balances that will give us 5 significant figures are commonplace, whereas it's hard to measure thickness to comparable accuracy. Second, we don't have to rely on tabulated densities, which are unreliable for some materials. Last, we get the *average* value automatically, which is probably what we want. Of course it never hurts to measure the thickness as well and check the tabulated density.

3.2 MeV

In design work we almost always express energy in eV (electron volts) or MeV (million electron volts) rather than joules (the MKS unit of energy). An *electron volt* (eV) by definition is the energy gained by one electron in falling through a potential difference of 1 V (V \equiv volt). That energy is given by

$$\text{energy} = \text{charge} \times \text{change in potential.}$$

Let $e = 1.602 \times 10^{-19}$ C (C \equiv coulomb) be the magnitude of the electron charge. Then

$$1 \text{ eV} = e \times 1 \text{ V} = \left(\frac{e}{\text{C}}\right) \text{ C V} = \left(\frac{e}{\text{C}}\right) 1 \text{ J}$$

(J \equiv joule). Rearranging the last equation we find that, in any expression, we can replace e either by

$$1.602 \times 10^{-19} \text{ C} \quad \text{or by} \quad 1 \left(\frac{\text{eV C}}{\text{J}} \right) \quad (3.1)$$

whichever is more convenient. Alternatively, we can just use

$$1 \text{ MeV} = 0.1602 \times 10^{-12} \text{ J}$$

3.3 Mass Stopping Power

The central quantity of stopping theory is the proton's rate of energy loss dE/dx . E (MeV) is the energy and x (cm) is distance along the proton path.¹ dE/dx is negative: as x increases E decreases. One therefore defines the *stopping power* $S \equiv -dE/dx$ which is positive. However, S almost always appears in the combination

$$\frac{S}{\rho} \equiv - \frac{1}{\rho} \frac{dE}{dx} \quad \frac{\text{Mev}}{\text{g/cm}^2} \quad (3.2)$$

which is called the *mass stopping power*. The term *linear energy transfer* or LET, which you may encounter but which we won't use, is loosely equivalent to $-dE/dx$. See [12] for the precise definition.

3.4 Fluence

The *proton fluence* is defined by

$$\Phi \equiv \frac{dN}{dA}, \quad \frac{\text{protons}}{\text{cm}^2} \quad (3.3)$$

Technically dN is the number of protons traversing a sphere of cross-sectional area dA centered at the point of interest [12]. However, we will always be dealing with an essentially unidirectional beam of protons in which case dA is simply an area perpendicular to the beam. The *fluence rate* is

$$\dot{\Phi} \equiv \frac{d\Phi}{dt}, \quad \frac{\text{protons}}{\text{cm}^2 \text{ s}} \quad (3.4)$$

Sometimes the lower-case Greek letter ϕ is used for $\dot{\Phi}$. As above, we'll sometimes put in the word 'protons' to clarify an expression. 'Protons' is dimensionless. It is a substitute for the pure number 1.

3.5 Physical Dose

The *physical absorbed dose* at some point of interest in a radiation field is the energy absorbed per unit target mass:

$$D \equiv \frac{dE}{dm}, \quad \frac{\text{J}}{\text{kg}} \quad (3.5)$$

This is one place we *do* use MKS units since 1 Gy \equiv 1 J/kg (Gy \equiv gray) is the standard clinical unit of dose. However, old habits die hard and the *rad* (for

¹ We'll always call it dE/dx because that's standard, but later on we'll frequently use z as the direction of particle motion and x, y as transverse coordinates, more in keeping with the convention of particle accelerator physics.

‘radiation absorbed dose’) is also used frequently ($1 \text{ rad} \equiv 100 \text{ erg/g} = 0.01 \text{ Gy}$). Frequently, radiotherapists hedge by using the term ‘centiGray’ (cGy) for ‘rad’.

The qualifiers ‘physical’ and ‘absorbed’ require discussion. The latter is easier: the energy absorbed by the target material may be less than the energy lost by the radiation. Protons, for instance, have nuclear interactions which produce neutrons among other secondaries. The neutrons (being neutral) go a long way. Instead of depositing their energy in the target they will usually stop in the shield walls of the radiation therapy facility. This ‘lost’ energy typically amounts to a few percent. Chapter 11 will explain how we take it into account.

3.6 Biologically Effective Dose

‘Physical’ dose stands in contrast to ‘biologically effective’ and that is a broad subject, filling many books. Physical dose is not, in fact, a very good measure of biological effect but a) it’s the accepted standard, b) it’s the best we have and c) it’s relatively easy to define and measure. Although this report is concerned entirely with how to produce prescribed *physical* dose distributions we’ll discuss biological dose very briefly as background.

The biological effect of a given physical dose depends on many factors: the particular form of radiation, its LET, the target tissue type, the fraction of an organ exposed and the fractionation schedule (time structure of the dose delivery), to list just a few. The ‘relative biological effectiveness’ (RBE) is the ratio of biological effect to that of the same physical dose of a standard radiation, usually the $\approx 1.3 \text{ MeV } \gamma$ -rays from ^{60}Co . The RBE of protons is around 1.1. In other words their biological effect for a given physical dose is not greatly different from photons. That contrasts with the RBE of other heavy particles, for instance carbon ions (≈ 2.5) or neutrons (≈ 3) [13].

Biologically effective dose has its own units: sieverts (Sv) corresponding to gray and rem (‘roentgen equivalent man’) corresponding to rad. Thus 1 gray of carbon ions equals 2.5 Sv and 1 rad equals 2.5 rem.²

Compared to physical dose, biological effect is hard to define and measure, particularly *in vivo*.³ What kind of precision do we aim for in clinical practice? A well designed ‘dose escalation’ study can, in many cases, detect the change in outcome from a 10% change in dose. Standard quality assurance practice usually sets a upper limit of $\pm 3\%$ variation in dose delivered day to day. We usually try to achieve $\pm 1\%$.

To end this short detour we’ll list some typical doses. The average American receives an annual dose of $\approx 360 \text{ mrem}$ (millirem) about 80% of which is from natural sources [14]. A transcontinental roundtrip flight exposes you to $\approx 5 \text{ mrem}$. The permissible dose for an adult U.S. radiation worker is $\approx 100 \text{ mrem}$ per week (see [15] for the exact regulation). Reference [16] is an excellent review of the health effects of *low-level* radiation. At the other end of the scale, a *whole body* dose of 0.5-1.5 Gy will cause severe, frequently fatal, radiation disease. Nevertheless, 1-2 Gy to a *limited* target is a typical single fraction delivered in radiation therapy.

² The RBE values used here are only typical, because they depend on all the factors listed and others.

³ *In vivo* experiments are done in the living organism. *In vitro* experiments, using cell cultures in artificial environments, are usually easier to quantify but may be less relevant clinically.

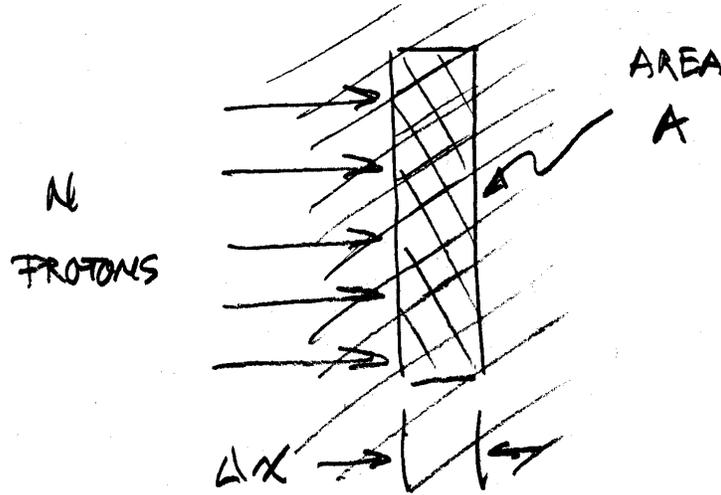


Figure 3.1: Drawing used in deriving Eq.3.6

3.7 Computing Dose from Fluence

The equation relating physical absorbed dose to fluence and stopping power is the starting point for all design work. Referring to Figure 3.1,

$$D \equiv \frac{\text{energy}}{\text{mass}} = \frac{(dE/dx) \times \Delta x \times N}{\rho \times A \times \Delta x}$$

that is

$$D = \Phi \frac{S}{\rho} \quad (3.6)$$

Used in this context, S/ρ refers to *absorbed energy*. When we calculate the dose in a water tank S/ρ will be an *effective mass stopping power* derived from a measured Bragg peak.

To write (3.6) in practical units let $\Phi = 1 \text{ gp/cm}^2$, where $\text{gp} \equiv \text{gigaproton} \equiv 10^9 \text{ p}$, and $S = 1 \text{ MeV}/(\text{g/cm}^2)$. Then

$$D = \frac{1 \text{ gp}}{\text{cm}^2} \times \frac{1 \text{ MeV}}{\text{g/cm}^2} \times \frac{10^3 \text{ g}}{\text{kg}} \times \frac{0.1602 \times 10^{-12} \text{ J}}{\text{MeV}} \times \frac{10^9 \text{ p}}{\text{gp}} = 0.1602 \text{ Gy}$$

so that

$$D = 0.1602 \Phi \frac{S}{\rho} \quad \text{Gy} \left(\frac{\text{gp}}{\text{cm}^2} \frac{\text{MeV}}{\text{g/cm}^2} \right)^{-1} \quad (3.7)$$

The dimensional part of (3.7) is equivalent to saying that D comes out in Gy if Φ is expressed in gp/cm^2 and S/ρ in $\text{MeV}/(\text{g/cm}^2)$. Since S/ρ is typically $5 \text{ MeV}/(\text{g/cm}^2)$ and a therapeutic dose per fraction is of the order of 1 Gy, Eq. (3.7) tells us that therapy fluences will be of the order of 1 gp/cm^2 . In other words, 10^9 protons is a convenient magnitude for therapy beams.

3.8 Computing Dose Rate from Proton Current Density

Taking the time derivative of (3.6) we have $\dot{D} = \dot{\Phi}S/\rho$. Current density is related to fluence rate by $i_p/A = \dot{N}e/A = e\dot{\Phi}$. Therefore

$$\dot{D} = \frac{1}{e} \frac{i_p}{A} \frac{S}{\rho} = \frac{i_p}{A} \frac{S}{\rho} \times \frac{\text{J}}{\text{eVC}}$$

using Eq. (3.1). To write this in practical units let $i_p/A = 1 \text{ nA/cm}^2$ and $S = 1 \text{ MeV/(g/cm}^2)$. Then

$$\dot{D} = \frac{1 \text{ nA}}{\text{cm}^2} \times \frac{1 \text{ MeV cm}^2}{\text{g}} \times \frac{\text{J}}{\text{eVC}} \times \frac{10^{-9}\text{C}}{\text{nAs}} \times \frac{10^6 \text{ eV}}{\text{MeV}} \times \frac{10^3 \text{ g}}{\text{kg}} = 1 \frac{\text{Gy}}{\text{s}}$$

so the desired formula is

$$\boxed{\dot{D} = \frac{i_p}{A} \frac{S}{\rho} \frac{\text{Gy}}{\text{s}} \left(\frac{\text{nA}}{\text{cm}^2} \frac{\text{MeV}}{\text{g/cm}^2} \right)^{-1}} \quad (3.8)$$

If the proton current density is 0.0033 nA/cm^2 and $S = 5 \text{ MeV/cm}^2$ we find $\dot{D} = 0.017 \text{ Gy/s}$ and the dose in 1 minute is 1 Gy.

We have assumed that i_p is uniform over A . If not, i_p/A should be interpreted as di_p/dA and the dose rate will also be non-uniform. Similarly, if the fluence in Eq. (3.7) is non-uniform the dose will be non-uniform.

Exercise: in a water tank, a test region of radius 5 cm is exposed to a uniform current of 2 nA of 160 MeV protons for 10 s. The mass stopping power of water for 160 MeV protons is $5.210 \text{ MeV/(g/cm}^2)$. Compute the dose, first with Eq. (3.8) and then Eq. (3.7). Answer: $A = \pi r^2 = 78.5 \text{ cm}^2$; $i/A = 0.0255 \text{ nA/cm}^2$; $\dot{D} = 0.132 \text{ Gy/s}$; $D = \dot{D}\Delta t = 1.32 \text{ Gy}$ from (3.8). The fluence is $\Phi = i\Delta t/(eA) = 1.59 \text{ gp/cm}^2$; the dose from (3.7) is 1.33 Gy.

3.9 Dose Rate in a Therapy Beam

Eqs. (3.7) and (3.8) by themselves do not answer the question most likely to be asked: assuming a complete beam spreading system (single- or double-scattering with a modulator), what is the dose rate in the spread-out Bragg peak (SOBP) given the proton current *entering the system*?

The derivation anticipates some later chapters, so you may want to put off this section for a while or revisit it later. Also, a properly written design program will give you essentially the same information—the incident gp needed for a given dose—automatically, and more accurately. Nevertheless, the formula we're about to derive is convenient for back of the envelope estimates absent a complete design.

Refer to Figure 3.2 (not to scale). The hypothetical scattering system has an upstream range modulator and a compensated contoured second scatterer. We assume it produces a flat dose distribution over a circular area A at the entrance to the water tank. The proton current i_p *entering the first scatterer* is assumed known (note the change in interpretation of i_p). We wish to compute the dose rate at the most distal peak of the SOBP (#1). The dose rate is the same everywhere else in the SOBP because the SOBP is flat. We start with Eq. 3.8 and introduce three dimensionless factors:

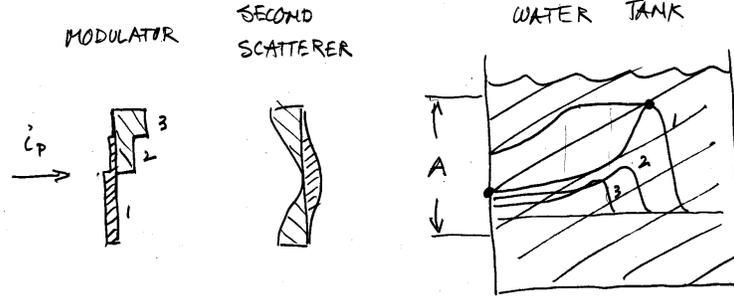


Figure 3.2: Drawing used in deriving Eq.3.9

1. Because (3.8) requires the current at A , we must multiply by the *efficiency* ϵ of the scattering system, defined as the ratio of current entering the useful area to current entering the scattering system. For a single scattering system ϵ is typically 0.05; for a double scattering system, 0.40.
2. The pristine Bragg peak can be considered as a curve of the effective stopping power of the actual proton beam *vs.* depth (see Ch. 11). The stopping power for step #1 at the *entrance* to the water tank is easy: ignoring small nuclear effects, it is just the table value of dE/dx in water at the energy of protons entering the tank (object 3) for step 1, $T_{3,1}$. Since, however, we are interested in the dose rate at the *top* of the Bragg peak we need to multiply by a factor f_{BP} defined as the peak/entrance ratio of the pristine Bragg peak (typically 3.5).
3. We now need to think about how a range modulator works. In its simplest form it's a propeller-shaped wheel that brings various thicknesses of plastic into the beam as it rotates.⁴ We assume the proton current $i_{p,1}$ entering the modulator (object 1) is constant. The dose rate at different parts of the SOBP is not constant, of course. The distal end, for instance, only receives dose while the modulator is on step 1. We wish to compute the dose rate $\langle \dot{D} \rangle$ *averaged over one modulator cycle* (or any integral number of cycles). We therefore need a dimensionless factor f_{MOD} defined as the dwell time for step 1 divided by the time for a full cycle. This depends on the modulator design, and mostly, on the amount of modulation. $f_{MOD} = 1$ for no modulation and $f_{MOD} \approx 0.3$ for full modulation.

At last we have

$$\langle \dot{D} \rangle = \epsilon f_{BP} f_{MOD} \frac{i_{p,1}}{A} \left(\frac{S}{\rho} \right)_{W,T_{3,1}} \frac{\text{Gy}}{\text{s}} \left(\frac{\text{nA}}{\text{cm}^2} \frac{\text{MeV}}{\text{g/cm}^2} \right)^{-1} \quad (3.9)$$

Eq. 3.9 also applies with *beam gating* and with *beam current modulation*. Both are used at NPTC to reduce the library of modulators required. Beam gating is a way of using a single modulator designed for full modulation to obtain anything less than full by simply turning the beam off during proximal steps. We must then take $i_{p,1}$ as the proton current while the beam is on. Beam current modulation allows us to trim the performance of a given physical modulator. In that case we must take $i_{p,1}$ as the current during the first modulator step. Gating can be thought of as an extreme form of beam current modulation.

⁴ Our modulator also has some lead: since it doubles as the first scatterer we want to keep the scattering constant as it rotates.

3.10 Air Filled Ionization Chambers

Though this report is not about ionization chambers (IC's) they are so important in proton therapy nozzles that a few formulas may be useful. They apply to *air filled plane-parallel* IC's under *standard conditions*. We follow Boag [17]. The current from an IC is

$$i_{IC} = q A d \left(\frac{1}{1 + \xi^2} \right) \quad (3.10)$$

where A is the effective area, d is the gap and q is the + or - charge liberated per unit volume per second. The factor in parentheses is the correction for ion recombination which we ignore for now (assume $\xi^2 = 0$). q is given by

$$q = \left(\frac{e \rho}{W} \right) \dot{D} \quad (3.11)$$

where $e = 1.602 \times 10^{-19} \text{ C}$ is the quantum of charge, $\rho = 0.00129 \text{ g/cm}^3$ is the density of air at standard conditions, $W = 34.3 \text{ eV}$ is the average energy expended per ion pair and \dot{D} is the dose rate. Combining these equations, assuming $A = 1 \text{ cm}^2$, $d = 1 \text{ cm}$ and $\dot{D} = 1 \text{ Gy/s}$, and working through the conversions we find in practical units

$$\boxed{i_{IC} = 37.6 A d \dot{D} \text{ nA} \left(\text{cm}^3 \frac{\text{Gy}}{\text{s}} \right)^{-1}} \quad (3.12)$$

This relation between dose rate and output current is convenient if we have a small active volume (such as a monitor pad) in a large beam. At the other extreme, we might have a large chamber with an unknown fluence and dose rate but we might know the total proton current i_p . Nonuniform dose will not matter if recombination is negligible. Using (3.8) in (3.12) we find the IC *multiplication*

$$\boxed{\frac{i_{IC}}{i_p} = 37.6 d \frac{S}{\rho} \left(\text{cm} \frac{\text{Mev}}{\text{g/cm}^2} \right)^{-1}} \quad (3.13)$$

Recombination of ions causes the measured current i_{IC} to be less than the ideal value given by Eqs.(3.12) and (3.13). It corresponds to non-zero ξ^2 in Eq. (3.10). Boag [17] gives, for standard air,

$$\xi^2 = 0.673 \times 10^{14} \left(\frac{\text{V}^2 \text{ s}}{\text{C m}} \right) \frac{d^4 q}{V^2}$$

Assuming this is not too large we can set $q \approx i_{IC}/Ad$. Rearranging, converting to practical units (nA, cm) and working through the conversions we find

$$\boxed{\xi^2 \approx 673 \frac{i_{IC}}{A} \left(\frac{d}{V} \right)^2 d \left(\frac{\text{nA}}{\text{cm}^2} \left(\frac{\text{cm}}{V} \right)^2 \text{ cm} \right)^{-1}} \quad (3.14)$$

One should use this only in the spirit of making sure it is sufficiently small for the design under consideration. If recombination is a problem, the most effective remedy by far is to decrease the gap. Of course that also decreases the signal. The next best is to increase the voltage. The proton current density i_p/A is usually a given, though i_{IC}/A also depends on d . Note that A may be either the active area of the chamber or the effective area of the beam, whichever is smaller. Recombination is usually detected experimentally by the dependence

of chamber output on V via Eq. 3.14. Boag [17] is the authoritative source and includes a review of measurements.

In addition to the explicit parameters of Eq. 3.14, recombination depends on the proton beam time structure via i_{IC} because one must use the value while the beam is on, not the long-term time average. Isochronous cyclotrons have fairly constant beam. The peak and average current are comparable, making it rather easy to avoid recombination. FM cyclotrons may have a duty factor around 2% in which case i_{IC} in (3.14) is $\approx 50\times$ the average value, and recombination is greater. Synchrotrons vary widely according to the method of beam extraction. The duty factor for slow extraction can be 1 s per 2 s or better. Single turn extraction, by contrast, may yield a duty factor on the order of $1 \mu\text{s}$ per 1 s. In that case IC's are virtually useless, and other means of monitoring the beam must be found.

3.11 Proton Kinematics

Occasionally we need to convert from one kinematic quantity, say the *kinetic energy* T to another, say pv , where p is proton's *momentum* and v its *speed*. Therapy protons are relativistic: their kinetic energy is not negligible compared to the *rest energy* $m_p c^2 = 938.27 \text{ MeV}$. Fortunately, single particle relativistic kinematics are trivial.⁵ We need only three equations. First, $\beta \equiv v/c$ (c is the speed of light) is given by

$$\beta = \frac{pc}{E} \quad (3.15)$$

where E is the proton's *total energy*. Second, E equals the kinetic energy plus the rest energy:

$$E = T + mc^2 \quad (3.16)$$

Finally,

$$E^2 = (pc)^2 + (mc^2)^2 \quad (3.17)$$

In proton therapy we normally deal with T : whenever we simply say 'energy' we mean T . Any kinematic quantity can be expressed in terms of T by using the previous three equations. For instance, the quantity pv , needed in multiple scattering theory, is found to be

$$pv = \left(\frac{T + 2mc^2}{T + mc^2} \right) T = \frac{E^2 - (mc^2)^2}{E} \quad (3.18)$$

As this example shows, it may be tidier to leave the expression in terms of E . It's always a good idea to check your work by taking the nonrelativistic limit $T \ll m_p c^2$. For instance, for Eq. 3.18 we find $pv \rightarrow 2T$, consistent with $T = mv^2/2$.

Exercise: it's instructive to compute the β of a therapy proton. From our three basic equations you should be able to show that

$$\beta = \sqrt{\left(1 + \frac{mc^2}{E}\right)\left(1 - \frac{mc^2}{E}\right)} \quad (3.19)$$

For instance, 160 MeV protons have $0.52 \times$ the speed of light. Show that Eq. (3.19) gives the correct answer $\beta = \sqrt{2T/mc^2}$ in the nonrelativistic limit.

⁵ If you are interested in much more complicated situations the authority is Hagedorn [18].

Chapter 4

Gaussians

Multiple scattering of protons is nearly Gaussian. Though we need not make that approximation in final calculations (there is no great advantage to be gained) it is nevertheless useful to discuss Gaussian scattering in one and two dimensions because the equations are simple and give us some good insights, particularly into the properties of single-scattering systems.

4.1 Small-Angle Approximation

Figure 4.1 defines the coordinate system. z is the beam direction, y is up, and x, y, z form a right-handed system. For protons, the *small-angle approximation*

$$\cos \theta \approx 1, \quad \sin \theta \approx \tan \theta \approx \theta, \quad (\text{radian})$$

is always valid, as can be argued two ways. First, we'll see in Chapter 7 that the *maximum* multiple scattering angle that can be obtained from lead under any circumstances is roughly 0.28 radians ($\approx 16^\circ$). Now $\sin(0.2800) = 0.2764$ so even here the approximation is pretty good. For a less extreme example, one can look at the angles we typically obtain with passive systems. In a proton gantry¹ with a throw of 270 cm a number of factors limit us to a field radius of ≈ 12 cm. In this case $\theta = 12/270 = 0.044$ radian (about 2.5°) where the small angle approximation is excellent.

4.2 1D Gaussian Probability Density

Figure 4.1 shows the common situation where we have a scatterer at z_0 and we are interested in the proton distribution on a *measuring plane* (MP) at z_{MP} a distance $L \equiv z_{MP} - z_0$ away. In Chapter 7 we'll see that multiple scattering from a simple scatterer is approximately Gaussian and in the present chapter we'll assume it's exactly Gaussian. Then the probability that the x component of the angle with the beam for a single proton falls in $d\theta_x$ about θ_x is

$$f(\theta_x) d\theta_x = \frac{1}{\sqrt{2\pi} \theta_{x0}} e^{-\frac{1}{2} \left(\frac{\theta_x}{\theta_{x0}} \right)^2} d\theta_x$$

where θ_{x0} is a characteristic multiple scattering angle. For now, we'll simply assume that θ_{x0} can be determined somehow. Exactly how we do that is the topic of Chapter 7.

¹ Numbers from NPTC.

4.3 2D Gaussian Probability Density

So far we have said nothing about the distribution in y . It can be anything. The protons can all lie along the x axis, or they can be uniformly distributed between two values of y . In the most common case, however, the multiple scattering is the same for x and y and the distribution in y is Gaussian with the same width parameter as x . Since the two scatters are independent, the probability that a proton falls in dx about x and dy about y is

$$f(x, y) dx dy = \frac{1}{\sqrt{2\pi} x_0} e^{-\frac{1}{2}\left(\frac{x}{x_0}\right)^2} dx \times \frac{1}{\sqrt{2\pi} y_0} e^{-\frac{1}{2}\left(\frac{y}{y_0}\right)^2} dy$$

If we now let $r_0 \equiv x_0 = y_0$ and transform to cylindrical coordinates $r^2 = x^2 + y^2$ and $dA = dx dy = r dr d\phi$ we obtain

$$\boxed{f(r) r dr d\phi = \frac{1}{2\pi r_0^2} e^{-\frac{1}{2}\left(\frac{r}{r_0}\right)^2} r dr d\phi} \quad (4.4)$$

$f(r)$ is a 2D probability density. $f(r) r dr d\phi$ is the probability that a single incident proton will hit the MP within dA . As before, the probability that a single proton lands somewhere in the MP is

$$\int_0^{2\pi} \int_0^\infty f(r) r dr d\phi = 1 \quad (4.5)$$

This integral, unlike (4.2), is straightforward, and is the indirect proof of Eq. (4.2) because the two describe the same physical situation. By contrast with (4.3), however

$$\sigma_r \equiv \langle r^2 \rangle^{1/2} \equiv \left(\int_0^{2\pi} \int_0^\infty r^2 f(r) r dr d\phi \right)^{1/2} = \sqrt{2} r_0 \quad (4.6)$$

The *rms* value of r is $\sqrt{2}$ times the *rms* value of x . We could, of course, have absorbed the $\sqrt{2}$ into the definition of the width parameter, thus using a different width parameter for the 1D Gaussian (4.1) and the 2D ‘cylindrical’ Gaussian (4.4), which would then read

$$f(r) r dr d\phi = \frac{1}{\pi \sigma_r^2} e^{-\left(\frac{r}{\sigma_r}\right)^2} r dr d\phi$$

Many writers including Molière [7] and Preston and Koehler [10] write it that way, and in reading the literature one must take note of which form is used.

4.4 Implications for Single Scattering

A single scattering system is simply the one shown in Figure 4.1, a homogeneous scatterer followed at some distance by a measuring plane. There may for various reasons be more than one scatterer, but they can be combined into a single effective scatterer (Chapter 10), and the net scattering is still Gaussian. For now, assume the effective location z_0 and the net characteristic radius r_0 are known. Then $f(r)$ in Eq. 4.4 is just the *fluence per proton* at the MP. In other words, if N protons are incident on the scatterer, the fluence at the MP is $N \times f(r)$. In particular, the fluence on axis is

$$N f(0) = \frac{N}{2\pi r_0^2} = \frac{N}{\pi \sigma_r^2}$$

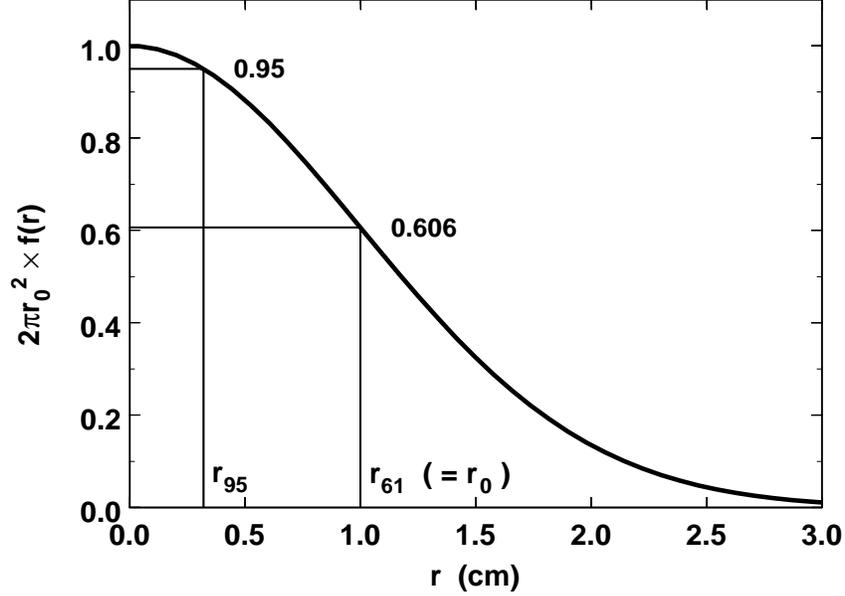


Figure 4.2: Gaussian radial distribution assuming $r_0 = 1$ cm.

Off axis the fluence falls (Figure 4.2), reaching 61% at $r = r_0$. With a single scattering system, we simply use as much of the Gaussian as we can and still stay within the prescribed dose uniformity. For instance, if $\pm 2.5\%$ is required we can use the Gaussian out to the 95% point r_{95} (Figure 4.2).² Designing a single scattering system is basically simple. The useful radius (r_{95} in our example) is given by the clinical requirement. We must use it to find the required r_0 and θ_0 . We must then use multiple scattering theory to get that much scattering.

In the Gaussian approximation the first step, finding θ_0 , can be done exactly. Continuing with our 95% example, the exponential factor of Eq. 4.4 is 1 at $r = 0$ and therefore we wish to find r_0 such that

$$0.95 = e^{-\frac{1}{2}\left(\frac{r_{95}}{r_0}\right)^2}$$

given r_{95} . The solution is

$$\boxed{r_0 = r_{95} (-2 \ln 0.95)^{-1/2}, \quad \theta_0 = r_0/L} \quad (4.7)$$

with the obvious generalization to the case where the ‘95’ is something else. For example, suppose we want $\pm 2.5\%$ uniformity out to 3 cm and the throw L is 150 cm. We find

$$r_0 = 3/\sqrt{-2 \ln 0.95} = 9.37 \text{ cm}, \quad \theta_0 = 0.0624 \text{ radian}$$

At 160 MeV incident, that will take about 8 mm of lead scatterer and the outgoing energy will be 135 MeV—but that’s for later.

The *efficiency* ϵ of the scattering system is of considerable interest. Of the N protons, what fraction falls within the useful radius? Preston and Koehler [10] pointed out a simple and exact answer for single scattering in the Gaussian

² Consistent with the notation r_{95} we should rename r_0 ‘ r_{61} ’ but we won’t.

approximation. ϵ is the probability that a particle falls within a radius R in the MP, namely

$$\epsilon = \int_0^{2\pi} \int_0^R \frac{1}{2\pi r_0^2} e^{-\frac{1}{2}\left(\frac{r}{r_0}\right)^2} r dr d\phi = 1 - e^{-\frac{1}{2}\left(\frac{R}{r_0}\right)^2}$$

or

$$\boxed{\epsilon = 1 - f(R)/f(0)} \quad (4.8)$$

In our example, we capture only $1 - 0.95 = 5\%$ of the incident protons. If we wanted $\pm 5\%$ uniformity, we would capture 10% and so on. Single scattering systems have low efficiency and, as we saw above, they use up a lot of energy. To do better we need double scattering—but that, too, is for later.

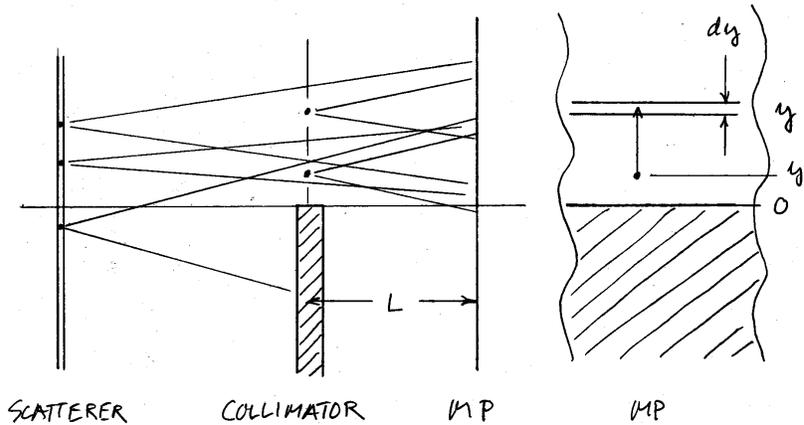


Figure 4.3: Geometry of transverse penumbra calculation.

4.5 Error Function

As our final topic related to Gaussians we discuss the error function. This arises when computing transverse penumbra, an important topic which we will treat more fully in Chapter 18. Here we'll consider only the simplest model, a beam line consisting of a scatterer, a collimator (whose single edge we put at $y = 0$ for convenience) and, at some distance L from the collimator, a measuring plane (Figure 4.3). L corresponds to the unavoidable *air gap* between the patient-specific collimator and the patient's skin. In most cases we would like as sharp a dose distribution at the skin as we can get.³ In other words, we would like the collimator to cast a sharp shadow. This desire is frustrated by the combination of the source size, angular distribution of protons in the beam, and the inevitable air gap.

Assume that a uniform parallel beam of infinite transverse extent (our model is 1-dimensional) falls on the scatterer. Each point of the scatterer can then be considered as emitting a Gaussian angular distribution characterized by some θ_{y0} . Now let us translate that whole plane of emitters to the plane of the collimator. The angular distribution is unchanged (there is no more scattering) but we now see that any emitters below $y = 0$ cannot contribute. They are

³ Less frequently, a fuzzier dose edge is desired and we deliberately increase the penumbra.

blocked by the collimator. Now take the second step of projecting the remaining proton trajectories onto the MP. Each one can be represented by an arrow starting at whatever y the proton had in the collimator plane, call it y' , and ending at some y in the MP. y' must be greater than 0 but y can be anything. The distribution of the length of these arrows is Gaussian with a characteristic length $y_0 = L\theta_{y0}$.

The relative number of protons falling in dy about y is gotten by integrating over all arrows starting anywhere and ending in dy , weighting each by the probability of its length $|y - y'|$:

$$p(y) dy = \left[\frac{1}{\sqrt{2\pi} y_0} \int_0^\infty e^{-\frac{1}{2} \left(\frac{y' - y}{y_0} \right)^2} dy' \right] dy \quad (4.9)$$

We now introduce the *error function*⁴

$$\operatorname{erf} x \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (4.10)$$

which has the properties

$$\operatorname{erf}(0) = 0, \quad \operatorname{erf}(\infty) = 1, \quad \operatorname{erf}(-x) = -\operatorname{erf}(x) \quad (4.11)$$

See [19] for a numerical table, [1] for a Fortran subroutine based on the fact that the error function is a special case of the incomplete gamma function, and our own subroutine `ERF.FOR` which uses series expansions [20] directly. If we need to evaluate an arbitrary definite integral of a standard 1D Gaussian, having width parameter x_0 and mean \bar{x} , the following formula is easily derived from (4.10):

$$\frac{1}{\sqrt{2\pi} x_0} \int_{x_1}^{x_2} e^{-\frac{1}{2} \left(\frac{x - \bar{x}}{x_0} \right)^2} dx = .5 \left(\operatorname{erf} \left(\frac{x_2 - \bar{x}}{\sqrt{2} x_0} \right) - \operatorname{erf} \left(\frac{x_1 - \bar{x}}{\sqrt{2} x_0} \right) \right) \quad (4.12)$$

Applying this to the penumbra equation (4.9) we obtain

$$p(y) = 0.5 + 0.5 \operatorname{erf} \left(\frac{y}{\sqrt{2} y_0} \right) \quad (4.13)$$

which is plotted in Figure 4.4 for $y_0 = 1$ mm.

For clinical purposes we usually want to know the width of the penumbra measured from the 80% point to the 20% point, the ‘80/20 width’, rather than y_0 which is what we will obtain from scattering theory. The two parameters are easy to relate. y at the 80% point, for instance, is found from

$$0.8 = 0.5 + 0.5 \operatorname{erf} \left(\frac{y_{80}}{\sqrt{2} y_0} \right)$$

whose solution is

$$y_{80} = \sqrt{2} y_0 \operatorname{erf}^{-1}(0.6)$$

By repeating this for y_{20} , or by symmetry, we find

$$\boxed{y_{80} - y_{20} = 2 \sqrt{2} \operatorname{erf}^{-1}(0.6) \times y_0 = 1.683 y_0} \quad (4.14)$$

as illustrated by Figure 4.4. That, incidentally, would be a fairly good penumbra in air. The best obtained in any beam at HCL was $y_{80} - y_{20} \approx 0.9$ mm.

⁴ The error function, favored by mathematicians, is closely related to the *normal probability integral*, favored by statisticians. Either one could be used.

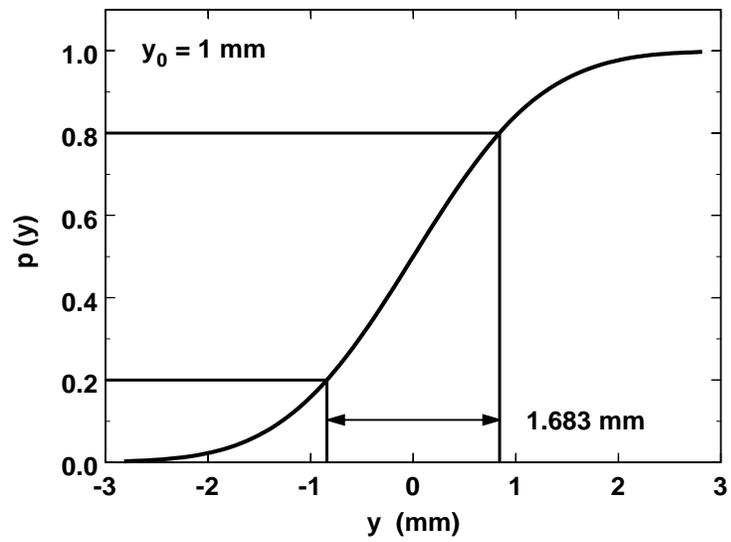


Figure 4.4: Penumbra profile showing 80/20 width.

Chapter 5

The Cubic Spline

One idiosyncrasy of our programs is the very extensive use of the cubic spline.¹ You may already know it as an option offered by graphics packages to join a set of points smoothly, cubic spline interpolation. We will use it that way, but also as a fit function and as a generalized smooth function whose parameters are adjusted to meet some mathematical goal.

5.1 Description

Assume we have a set of N points x_i, y_i . The cubic spline is a smooth line which passes through all the points. In each interval *e.g.* $x_4 \leq x \leq x_5$ it is a cubic polynomial in x , but the four polynomial coefficients change from one interval to the next. They are computed [1] in such a way that the first and second derivatives are continuous at the interval boundaries, that is, at the given points. That is what makes the cubic spline smooth. It has no corners, that is, no places where the first derivative jumps.

It turns out [1] that requiring the curve to pass through all the points and to have continuous first and second derivatives is not quite enough to determine the $4(N - 1)$ polynomial coefficients. We also need to specify two boundary conditions. At each end, we must either a) specify the *first* derivative $y' \equiv dy/dx$ or b) require that the *second* derivative y'' be 0. The latter is called the ‘natural’ boundary condition. The conditions can be mixed: natural at one end and a specified y' at the other. Our routine `SPLINE.FOR` is a slightly reorganized version of that given in [1] and uses the same mechanism (a non-physically large y') to signal that the ‘natural’ boundary condition is to be used at one or the other end.

A natural cubic spline through just two points is simply the straight line joining them.

5.2 Cubic Spline Interpolation

As described so far, the cubic spline is an *interpolation* function. It passes through all the data points. We will use it this way to interpolate range-energy tables. However, cubic spline interpolation (or any other sophisticated method) may not work well if the data are noisy.

Figure 5.1 shows a set of points (solid squares) generated by picking points uniformly along a smooth curve, then adding some Gaussian noise in x and y

¹ The name comes from a draftsman’s tool, a sort of bendable ruler, used to draw arbitrary smooth curves.

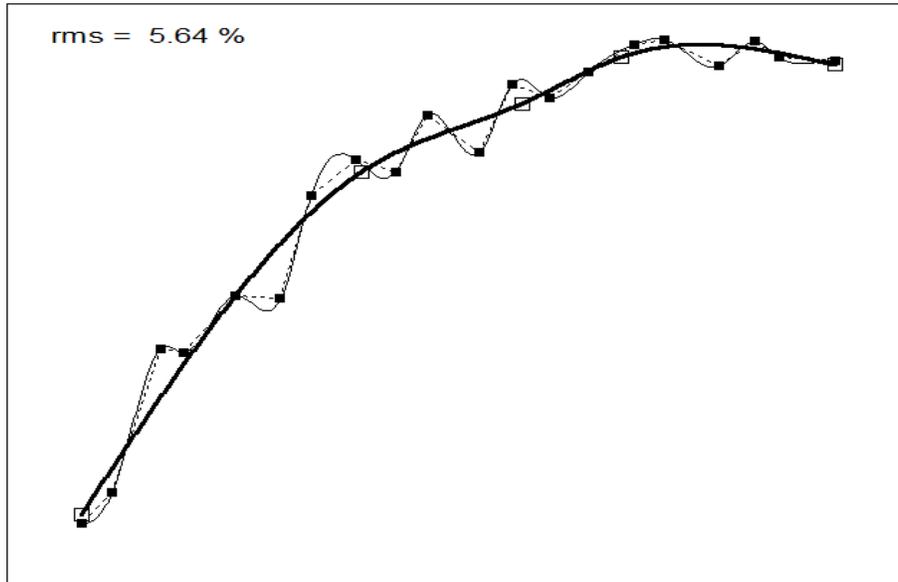


Figure 5.1: Cubic spline fit. Light line: cubic spline used to interpolate noisy data (full squares). Dashed line: linear interpolation of the same data. Bold line: cubic spline fit to the data, using 5 auxiliary points (hollow squares).

to simulate measurement error. The natural cubic spline through these points (light wavy line) is indeed smooth, but it does not represent the underlying function particularly well. In its effort to pass smoothly through all the data points it makes excursions that actually amplify the noise in the data. Linear interpolation (dashed line) is both simpler and better.

5.3 Using the Cubic Spline

Before we continue, a word about how a calling program uses the cubic spline. Module `SPLINE.FOR` has two entries `InitSpline` and `Spline`, used as follows:

```

DIMENSION x(n),y(n),s(n)
DATA yp1,ypn/0.,1.E30/
...
i = InitSpline(n,x,y,s,yp1,ypn,xx,klo,khi)
...
ya = Spline(n,x,y,s,yp1,ypn,xa,klo,khi)
yb = Spline(n,x,y,s,yp1,ypn,xb,klo,khi)

```

Initialization is only done once, or whenever the x, y table changes. It causes an auxiliary array s , for which space was reserved by the calling program, to be set up. In fact, s is an array of second derivatives y'' . The computation of s is fairly lengthy [1]. `yp1` and `ypn` specify the boundary conditions, here $y' = 0$ at x_1 and natural ($y'' = 0$) at x_N . `xx` is a placeholder. After initialization, `Spline` may be called as often as needed. For each x , here xa , it returns the interpolated y , here ya .

When called with some xa `Spline` first finds its *table position*, that is, indices `klo` and `khi` such that $x_{KLO} \leq xa < x_{KHI}$. It does this by bisection. Because we frequently work our way systematically up a table, `KLO` and `KHI` are preserved. If on entering `Spline` we find that the table position is already correct, we

can skip that step. Once the table position is known, ya is found by a short calculation using the arrays x , y and s .

In addition to the bisection routine, a *linear* interpolation routine is included in `SPLINE.FOR`. It also has two entries, used in the same way just described. In this case s is an auxiliary array of precalculated slopes y' .

5.4 Cubic Spline Fit

In contrast to an interpolation function, which passes through the data points, a *fit function* passes *near* the data points in some mathematically defined sense. Most common by far is the *least-squares* fit which seeks to minimize

$$\chi^2 \equiv \sum_1^N (y_i - y(x_i))^2,$$

the sum of squared vertical distances of the fit function from the data points. The mathematical form of $y(x)$ is suggested either by the trend of the data or by some underlying physical theory we may have in mind. Its precise shape depends on a set of *fit parameters*. It is these parameters we adjust to minimize χ^2 . There is an extensive literature on the subject [1, 21].

For instance, a common fit function is the *polynomial*

$$y(x) = a_1 + a_2x + a_3x^2 + a_4x^3 + a_5x^4,$$

where for illustration we have used a polynomial of the fourth degree having five parameters a_i .² Polynomials belong to a special class of functions which are *linear* in their fit parameters and lead to a *linear least-squares* fit, for which special techniques apply [1, 21].

The cubic spline is well adapted for use as a fit function since it can fit any smooth curve. Pick a set of auxiliary points (the hollow squares in Figure 5.1), far fewer in number than the data points, and adjust their positions so that the cubic spline through the auxiliary points minimizes χ^2 with respect to the data points. The x and y coordinates of the auxiliary or ‘spline points’ as we shall call them are the adjustable parameters. However, we must lock the x coordinates of the first and last points or the fit will collapse.

The cubic spline is nonlinear in its adjustable parameters even though it consists of a set of polynomials. We therefore have a *non-linear least-squares fit* and must use methods that differ from the linear case [1, 21]. Fundamentally, all of these consist of making an educated guess at initial values of the parameters and then improving them by trial and error. In the case of the cubic spline fit the educated guess is easy because the parameters are simply new points, which must lie very close to the data set we are trying to fit.

The bold line of Figure 5.1 is a spline fit to the data set. We have chosen to use five spline points, picked their initial values using the density function method described below, and optimized the fit with a grid search, also described. The fit represents the underlying trend of the data better than either spline or linear interpolation. In this particular case, a low-order polynomial fit would do as well. However, polynomials do not work well for Bragg peaks.

5.4.1 Choosing Initial Points

The simplest way to choose initial points is to pick x uniformly and obtain y values by linear interpolation of the data set. That would work sufficiently

² From the programmer’s point of view it is the number of parameters that matters more.

well in our artificial example, Figure 5.1. It does not work for Bragg peaks, whose curvature varies greatly. The cubic spline can, in each segment, mimic something as complicated as a skewed parabola. We waste this power if we use the same density of points in the nearly straight entrance region as we need in the peak. The fit function will have needlessly many parameters and will begin chasing experimental noise. We require an automatic procedure for picking points more efficiently.

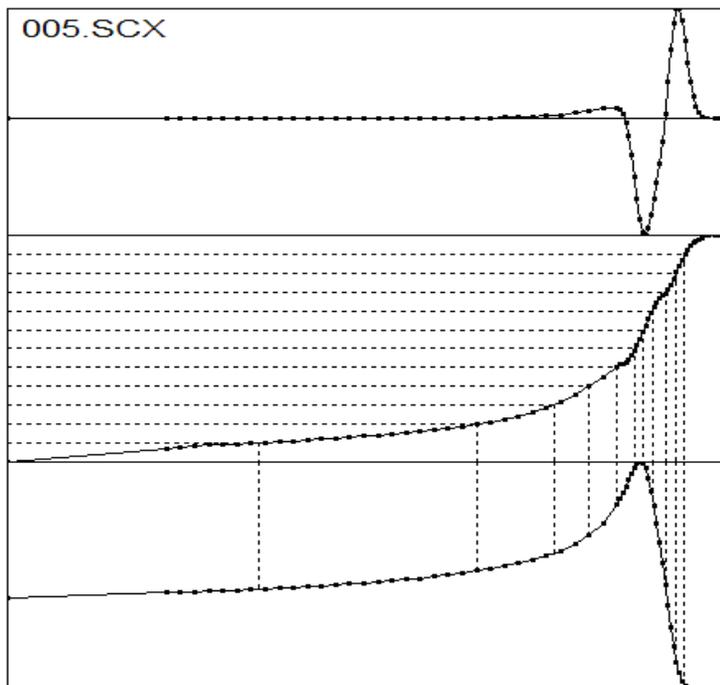


Figure 5.2: Procedure for picking initial spline points. Bottom frame: Bragg peak to be fitted. Top frame: smoothed second derivative of the Bragg peak. Middle frame: density function (see text).

Figure 5.2 uses a measured Bragg peak (bottom frame) as a target function to illustrate a convenient procedure. We input the total number of points desired. We space these uniformly in y and project them onto a *density function* which a) increases monotonically with x and b) increases most rapidly where the target function curves the most. A reasonable choice for the density function (middle frame) is

$$\int_0^x |d^2y/dx'^2|^p dx'$$

namely the integral of the absolute value of the smoothed second derivative, raised to some fractional power p . Most of this is obvious. ‘Curvature’ suggests the second derivative, the absolute value is taken because we don’t care about the sign of the curvature, and the integral ensures a monotonically increasing function. With that alone, however, we find that we pick too few points in the entrance region because the curvature is simply too small. We therefore introduce the fractional power p to provide some dynamic compression of the y scale. $p = 0.5$ (square root) appears to work well for many cases. Smaller values give still more points in the flat region.

5.4.2 Optimizing the Fit

The gold standard for nonlinear least-squares fits is the Marquardt method [1, 21]. Spline fits seem, however, to be one case where a simple grid search with inverse parabolic interpolation [1] is not only more robust but actually faster. This method (see `GridParab` in `BSFIT.FOR` assumes that χ^2 is an approximately parabolic function of any of the parameters near a minimum. We simply slog through each parameter in order, skipping the two x_1, x_N that must be held constant. (Those two initial increments are set to 0 as a signal to `GridParab`.) For each parameter, we look for a neighboring value where the overall χ^2 is smaller, going back and forth and reducing the increment each time to avoid an infinite loop. If we find such a value we go further in the same direction until χ^2 increases again. We have now bracketed a minimum of the parabola and we use the formula given in [1] to locate it. This is a simpler version of the Brent procedure [1] and seems fast enough.

The reason the grid search works so well is probably that the fit parameters are fairly independent in their effect. Changing one of them has very little effect a few spline points away. The entire process (say 15 spline points and 5 passes) takes hundredths of a second on a high-end PC.

Figure 5.3 shows the final result for a typical measured Bragg peak. Roughly 50 points are reduced to 13 and experimental noise is averaged out.

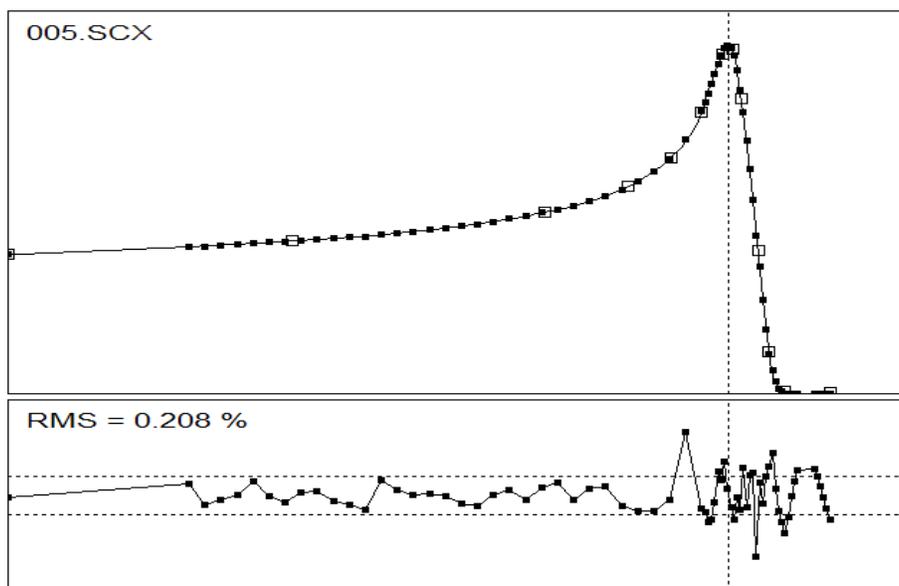


Figure 5.3: Upper frame: Bragg peak with spline fit. Lower frame: deviation from fit. The dashed lines are at \pm the *rms* deviation.

5.4.3 Pathologies

Once in a while it will happen that two of the points cross and the x_i no longer increase monotonically. This could probably be fixed by simply swapping the points in question but we have not tried that. Some trivial change in the startup file (decreasing the number of spline points by 1 or slightly changing p) usually solves the problem.

5.5 The Broken Spline

The cubic spline is averse to corners because it insists on a continuous first derivative. We can put this to good use if we have a function, for instance a spread-out Bragg peak (SOBP), whose corners we wish to locate. We decide on the number of segments (3) in advance and use a separate natural cubic spline for each, with different numbers of spline points as appropriate. We need only a crude corner-finding procedure to predefine the initial range of each spline. We then assign initial points to each spline as described above. When we turn the fit loose, locking the outermost x 's as before, the inner spline endpoints will gravitate towards the corners precisely because a corner is not well fit *inside* a spline.

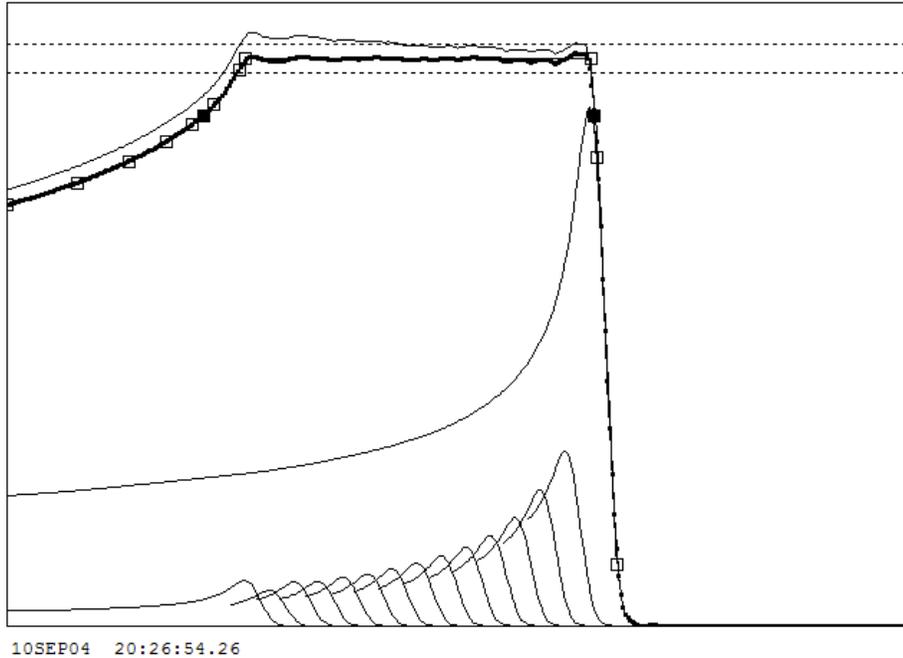


Figure 5.4: Spread-out Bragg peak fit by 3-segment broken spline.

Figure 5.4, generated by the single scattering design program LAMINATE, has many features we will skip over at present. The entrance region is fit by a spline of 8 points, the top by a spline of 2 points (a straight line) and the distal edge by 3 points. The spline points are the open squares. The full squares merely mark the proximal and distal 90% dose points. Finding these is a one-liner once the entire function has been ‘splinified’.

A broken spline fit program must deal with the fact that the inner spline endpoints appear in the lists for *two* splines each, but each represents just *one* fit parameter. Therefore the example has $8 + 2 + 3 = 13$ spline points but only $(13 - 2) \times 2 = 22$ adjustable parameters, two of which are locked. You can examine BSFIT.FOR to see how this is handled. Incidentally, BSfit is the only spline fitting program we use. A simple spline fit, as for a pristine Bragg peak, is handled as a single segment broken spline.

Not too long ago we advocated using three polynomials to do the same job [22]. The basic idea of that paper was the same as here: the analysis of an arbitrary SOBP should concentrate on finding the corners. In [22] the corners were defined as the intersections of three polynomials. Unfortunately,

the proximal and top polynomials sometimes fail to intersect. Then we must change gears and find the point of closest approach. Also, that part of the fit which finds the corners must be handled as an outer loop around the linear fit which determines each polynomial. All of this makes for messy code. The broken spline fit is much more elegant, and absolutely guarantees a continuous function representing the entire SOBP.

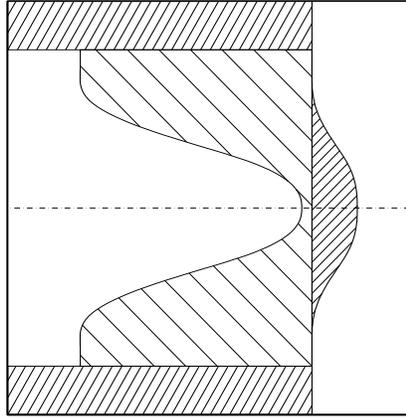


Figure 5.5: Compensated contoured second scatterer.

5.6 The Spline as a General-Purpose Function

Instead of using the cubic spline to fit a set of data points we may use it as a general smooth function and adjust its parameters to accomplish some other mathematical goal. It is used that way in designing contoured second scatterers (Figure 5.5). A quick outline of the method (details in Chapter 14) is as follows. First, a generic scatterer shape defined by spline points at four fixed radii is designed by varying the four scattering strengths to obtain, for a given first scatterer strength, a flat dose distribution at the patient. Given the physical requirements of a specific problem (incident energy, throw, radius and depth) we can convert this generic mathematical shape to a specific profile for the lead and plastic³ by using the laws of stopping and scattering. We again use the spline, this time as an interpolation function, to express those profiles. If it is physically impossible to meet the requirements (for example, if we ask for too large a radius at the given depth, energy and throw) the conversion to physical profiles will fail one way or another.

³ The plastic serves to make proton energy loss independent of radius. That is what we mean by 'compensated'.

Chapter 6

Stopping

We discuss stopping before multiple scattering because multiple scattering has almost no effect on stopping. It only enters via the difference, negligible at therapy energies, between pathlength and projected range. By contrast, we need stopping theory to compute multiple scattering in thick degraders.

The introduction to [2] will convince you to leave the actual computation of range-energy tables to the experts. We'll only outline the theory. After that, we'll take the existence of range-energy tables as a given and discuss how to use them. A range-energy table is a tabulation of stopping power and mean projected range (which we'll usually just call 'range') as a function of incident energy. Sometimes other things are included. For instance, Janni [3] tabulates range straggling and nuclear interaction probability.

6.1 Stopping Theory

6.1.1 Qualitative

Protons lose their energy by myriad electromagnetic (EM) interactions with atomic electrons. These interactions ionize and excite the target atoms and impart kinetic energy to the electrons. The *maximum* kinetic energy that can be imparted to an electron by a proton of kinetic energy T in a head-on collision is approximately

$$W_m \approx 4 \frac{m_e c^2}{m_p c^2} T = 0.35 \text{ MeV (at 160 MeV)}$$

(See (6.4) for the exact equation.) Most of the knock-on electrons have far lower energies than that. The more energetic ones are called δ rays. Knock-on electrons have further collisions and the net result is that all of the energy the proton loses in each electronic collision is deposited nearby in the stopping material. You can think of a track surrounded by a very short-range electron cloud.

Protons can also lose energy by elastic EM collisions with atomic nuclei. This gives rise to a 'nuclear stopping power' S_{nuc} which is only important at incident energies far lower than the ones we're interested in.

Protons also undergo *nuclear* (as distinct from EM) interactions with atomic nuclei of the stopping material. These are infrequent but not negligible. At 160 MeV in water, $\approx 20\%$ of the primary protons will have a nuclear interaction before they stop. Unlike the energy lost by EM interactions, some of the energy lost to nuclear interactions is *not* deposited in the material because it is carried by neutrons which can go a long way before interacting. Traditionally, these

nuclear interactions are treated separately. They do *not* enter into the tabulated stopping power (or range) that we are discussing here. Nuclear interactions will be taken up in chapter 8.

The most important thing about S is that it increases, down to very small proton energy, as the proton slows down:

$$S \propto \frac{1}{v^2}$$

where v is the proton speed. This gives rise to the Bragg peak of ionization which is one of the advantages of protons for radiation therapy.

6.1.2 Quantitative

We will not actually use the following formulas, which are only meant to give the flavor of the theory and convince the reader to leave the calculation of range-energy tables to the experts. The discussion follows [2]. (No two authors state the Bethe and ancillary equations exactly the same way.) See ICRU49 [2], Janni [3], the Review of Particle Physics [11] or Bethe and Ashkin [23] for formulas for the various corrections and discussions of their range of validity. The Bethe equation for mass stopping power is

$$(1/\rho)S_{col} = -(1/\rho)(dE/dx)_{el} = \frac{4\pi r_e^2 m_e c^2}{\beta^2} \frac{1}{u} \frac{Z}{A} z^2 L(\beta) \quad (6.1)$$

The function $L(\beta)$ is called the *stopping number*. It takes into account the fine details of the energy loss process whereas the factor preceding it gives the gross features. L can be written

$$L(\beta) = L_0(\beta) + zL_1(\beta) + z^2L_2(\beta) \quad (6.2)$$

zL_1 is the ‘Barkas correction’ and z^2L_2 , the ‘Bloch correction’. The main term L_0 is given by

$$L_0(\beta) = \frac{1}{2} \ln \left(\frac{2m_e c^2 \beta^2 W_m}{(1-\beta^2) I^2} \right) - \beta^2 - \frac{C}{Z} - \frac{\delta}{2} \quad (6.3)$$

W_m is the largest possible energy loss in a single collision with a free electron, given by

$$W_m = \frac{2m_e c^2 \beta^2}{1-\beta^2} \times [1 + 2(m_e/m_p)(1-\beta^2)^{-1/2} + (m_e/m_p)^2]^{-1} \quad (6.4)$$

The simpler version given earlier is a non-relativistic approximation. The factor in square brackets is very nearly unity. C/Z in (6.3) is the ‘shell correction’ and $\delta/2$, the ‘density-effect correction’. All of these corrections are made in recent tables, but since they are difficult to calculate it is of interest to know how important they are in the proton therapy regime.

To help decide what proton energy range we’re interested in, here’s a very short range-energy table in water with numbers from [2]:

energy	1	3	10	30	100	300	MeV
range	0.002	0.014	0.123	0.885	7.718	51.45	cm H ₂ O

For therapy, we certainly lose interest in a proton once its residual range is 0.14 mm. At the other end, we don’t really need to consider depths above 51 cm. Therefore the range of interest for radiation therapy is about 3-300 MeV. Happily, protons of this energy fall in a very favorable region for the theory.

The complicated low-energy region discussed at length in [2] and [3], where the Bethe equation does not apply at all, does not need to be considered. The Bloch, Barkas, shell and density effect corrections are all small. The detour factor (ratio of projected range to pathlength) due to multiple scattering is very close to 1, and S_{nuc} is only a tiny fraction of S .

In fact, in the clinical range the largest uncertainty by far comes in via the *mean excitation energy* I . In principle, I can be calculated from atomic structure but this is not sufficiently accurate. In practice, I is adjusted to agree with experimental data for elements where data exist, and is interpolated, with theory as a guide, where it does not. Figure 2.4 of [2] shows that this is not a trivial task. When two modern range-energy tables such as [2] and [3] differ, it is most likely because of their choice of I .

Another issue is the assumption of *additivity* made when the stopping power of a chemical compound is computed as the sum of the stopping powers of its atomic constituents acting independently. Janni [3] discusses this at some length, noting that the experimental data are ‘... contradictory and somewhat inconsistent.’ Fortunately, once again, non-additivity is only important at low energies.

To summarize this very brief discussion: a lot of complicated things can happen, but they are all minor in the clinical energy range 3-300 MeV. The main uncertainty is the choice of I , which must be made separately for each element. It is entirely possible for a given set of range-energy tables to be accurate on the whole but poor for one or two elements or compounds.

The error in range associated with the uncertainty of I is roughly 1-2%. This is how much you might expect sets of tables to differ one from the other. Unfortunately, at a typical clinical range of 15 cm in water this translates to 1.5-3 mm, not quite what we would like. We can get around this when necessary by using range-energy tables for most of our design, then using *measured* water-equivalent thicknesses to polish the answer (see Chapter 13).

As an exercise, you might try using the equations of this section, ignoring all the named corrections, to compute the stopping power for 150 MeV protons in amorphous carbon ($I = 81$ eV). A few intermediate results to help:

$$(4\pi r_e^2 m_e c^2 / u) = 0.3072 \text{ MeV}/(\text{g}/\text{cm}^2)$$

$$(Zz^2/A\beta^2) = 1.946 \quad W_m = 0.353 \text{ MeV}$$

and don't forget to convert I to MeV. You should find a value within $\approx 0.2\%$ of the ICRU49 value, 4.844 MeV/(g/cm²), showing that the corrections are indeed small at this energy.

6.2 Range

Once the stopping power as a function of energy has been calculated over the range T_0 to T_f , the computed pathlength is found by numerical integration:

$$P(T_0) = P(T_f) + \int_{T_f}^{T_0} \left(\frac{1}{\rho} \frac{dE}{dx} \right)^{-1} dE \quad \text{g}/\text{cm}^2 \quad (6.5)$$

T_f is some very small final kinetic energy and $P(T_f)$ is the corresponding pathlength. Janni [3] takes these as 10 eV and 0 respectively. The *mean projected range* is a bit smaller than the wiggly pathlength. It is given by

$$R(T_0) = R(T_f) \langle \cos \theta \rangle_f + \int_{T_f}^{T_0} \langle \cos \theta \rangle \left(\frac{1}{\rho} \frac{dE}{dx} \right)^{-1} dE \quad \text{g}/\text{cm}^2 \quad (6.6)$$

where θ is the space angle between the nominal beam direction and the proton direction at any instant. Its average value $\langle \cos \theta \rangle$ is calculated by means of multiple scattering theory. However, as already stated, the resulting correction (the difference between P and R) is negligible over the clinical energy range: about 0.14% for 100 MeV protons in water.

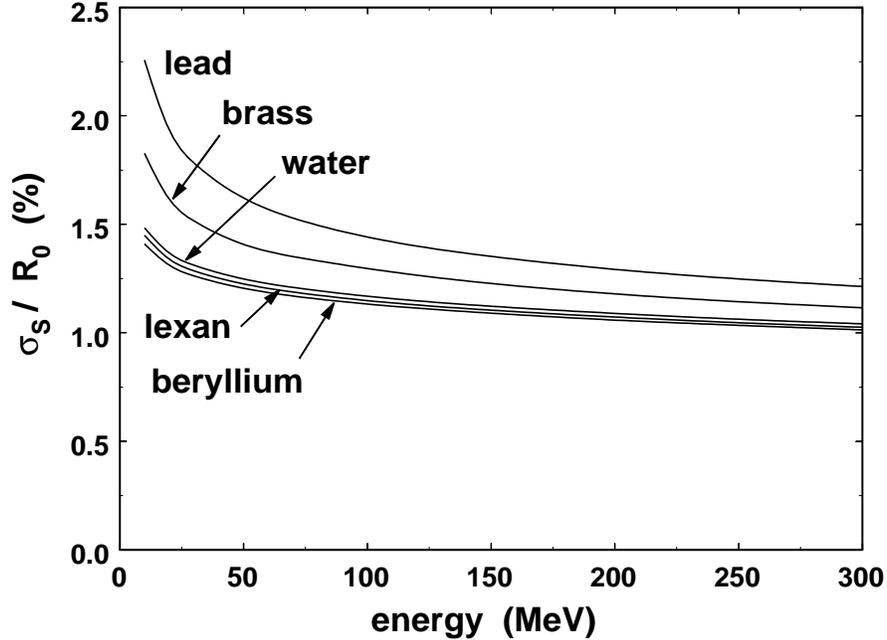


Figure 6.1: Range straggling in several materials, from Janni [3].

6.3 Range Straggling

Energy loss occurs in discrete interactions and thus has statistical fluctuations. Protons passing through a degrader will lose slightly different amounts of energy. The energy loss distribution is of interest to persons designing particle detectors, and there is a considerable literature on the subject, associated with the names Landau and Vavilov. Another consequence of the fluctuations is that monoenergetic protons stopping in some material will not all stop at exactly the same depth. This effect, called ‘range straggling’, means that the number *vs.* depth distribution, and therefore the Bragg peak of ionization, will have some minimum width even if the incident beam has zero energy spread.

In designing passive beam spreading systems, range straggling raises only one issue of importance. Suppose we have an incoming beam of fixed energy. We may wish to decrease its depth in the patient (or in the usual proxy, a water tank) by adding material upstream. It is usually more convenient to use plastic rather than water for this extra degrader. If the beam spreading system also calls for more scattering, we may want to add some lead. When we do all this, we would like to make the approximation that the Bragg peak is simply ‘pulled back’ by the water equivalent of the extra material: that its width, in particular, is unaffected.

We therefore need to know how range straggling depends on the material. Janni’s tables [3] include range straggling. He considers fluctuations in energy

loss to electrons as well as two effects that only contribute at low energies (energy loss to nuclei and fluctuations in the effective charge state of the projectile). The formulas are complicated and we shall not reproduce them here. He computes the standard deviation σ_S of a 1D Gaussian representing the distribution of stopping depths, and tabulates σ_S/R_0 , straggling relative to range, expressed in percent. His results for five materials of interest are shown in Figure 6.1.

Lexan and beryllium, materials we might use to reduce the range, are barely distinguishable from water. Brass and lead, which we might use for scattering, have somewhat more straggling than water: lead/water= 21% at 150 MeV. If we use a little lead and a lot of lexan (the usual situation for a range modulator) it seems extremely unlikely that there would be a detectable increase in the Bragg peak width in the water tank. Incidentally, a small increase in the width does no harm anyway. It is only a significant *decrease* in the width that would cause trouble, by introducing wiggles in the SOBP.

The fact that our designs make the ‘pullback’ assumption and produce flat SOBP’s is indirect evidence that the width does not change significantly. If it did, that would be accompanied by an opposite change in height, which (since we do not correct for it) would lead to a non-flat SOBP.

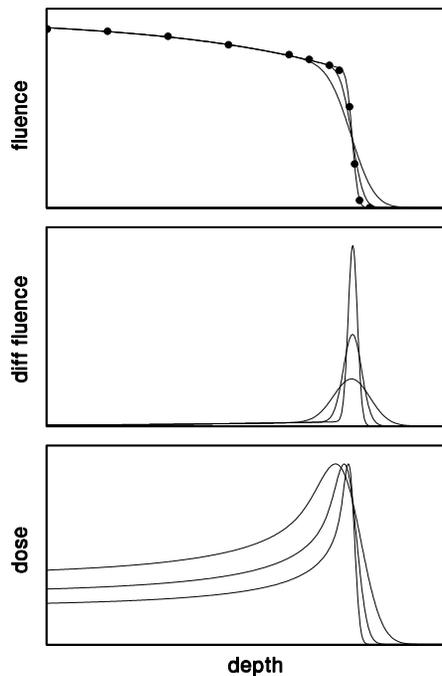


Figure 6.2: Fluence, differential fluence and dose as a function of depth for proton beams of a given range and different energy spreads, illustrating $r_0 = d_{80}$.

6.4 Measuring Range

So much for the *computation* of mean projected range R_0 , which we’ll just call ‘range’ from now on. How do we *measure* the range of a beam? We need to find that depth where the proton (number *vs.* depth) distribution falls to half its initial value, correcting for nuclear interactions. To do this we need a *fluence meter*. However, in many cases it is more convenient to use a *dose meter* (such as a small ionization chamber) to measure the Bragg peak in a water tank.

The range can also be found from that kind of measurement, but it does *not* correspond to the peak value! Figure 6.2 illustrates data from three possible measurements.

For the top panel, we have put a Faraday Cup (FC) into the beam and we measure the change in charge collected by the FC as degraders are placed in front of it. Because the proton charge is constant and well known the FC is a ‘proton counter’: it measures proton current. For the degrader (here CH₂) we use whatever material we want to measure the range in. As we start out we see only a gradual decrease in flux. That is due to the removal of primary protons by nuclear interactions in the CH₂.¹ Eventually we reach a sharp decrease in the proton count when the primary protons range out. The slope of this ‘cliff’ comes from range straggling (irreducible) and the energy spread of the incident beam (depends on the accelerator). The measured value of R_0 is the degrader thickness at which the proton count is half its value at the top of the cliff.²

Another fluence-based measurement, shown in the middle panel, uses a multi-layer Faraday Cup (MLFC), a stack of thin degraders isolated from each other electrically. We measure the charge collected by each. That way we get the entire previous measurement in one shot, at the cost of needing a lot of charge-measuring devices (current integrators) instead of just one. The MLFC measures *differential* fluence. The *maximum* value or, more accurately, the *mean position* of the sharp peak, corresponds to R_0 .

MLFC’s can be built two ways. The sheets can be a conductor such as copper or aluminum, separated by thin insulators, with each sheet connected directly to a current integrator. Or the sheets can be insulator, separated by thin conducting charge collection plates which are connected to current integrators. In the latter case it is the charge electrostatically induced on the collectors that is counted. The two kinds work equally well, and allow us to measure beam stopping in a large variety of materials. MLFC’s are frequently used as *range verifiers* to monitor the stability of accelerators and beam lines, because they yield an accurate measurement in seconds at typical therapy beam intensities.

What happened to the nuclear events in the MLFC? They are in the barely visible buildup region before the EM peak. All charged nuclear secondaries have shorter projected ranges than primaries, so all the nuclear events are in the buildup and all the primary protons are in the peak. MLFC’s allow us to measure nuclear events in detail [24, 25] as do well-designed experiments of the first kind (simple FC with external degraders). This will be discussed in greater detail in Chapter 8.

Lastly, suppose we measure depth-dose in a water tank using a dosimeter such as a small ion chamber. Dosimeters measure the *energy deposition* of protons rather than the *number* of protons. We obtain a Bragg peak as shown in the lowest panel. (The peak in the middle panel is *not* a Bragg peak.) Some point on the Bragg peak must correspond to R_0 , but which one? That question was investigated in the 1960’s by Andy Koehler. He created a Bragg peak numerically by folding the variation of dE/dx (which increases as the proton slows down) with the Gaussian spread introduced by range straggling and the energy spread of the incident beam. Comparing this computed Bragg peak with the proton range he assumed at the beginning of the calculation he found

$$R_0 = d_{80} \tag{6.7}$$

where d_{80} is the distal 80% point of the Bragg peak. This result has since been

¹ There are charged nuclear secondaries that would count in the FC but almost all of them either range out in the degrader because of their low energies, or miss the FC geometrically because of their large angles.

² Not half the entrance value!

confirmed several times [26, 27]. Note that this is a strictly *numerical* result. In fact, the real number is not exactly 80%, but 80% is close enough, and easy to remember.

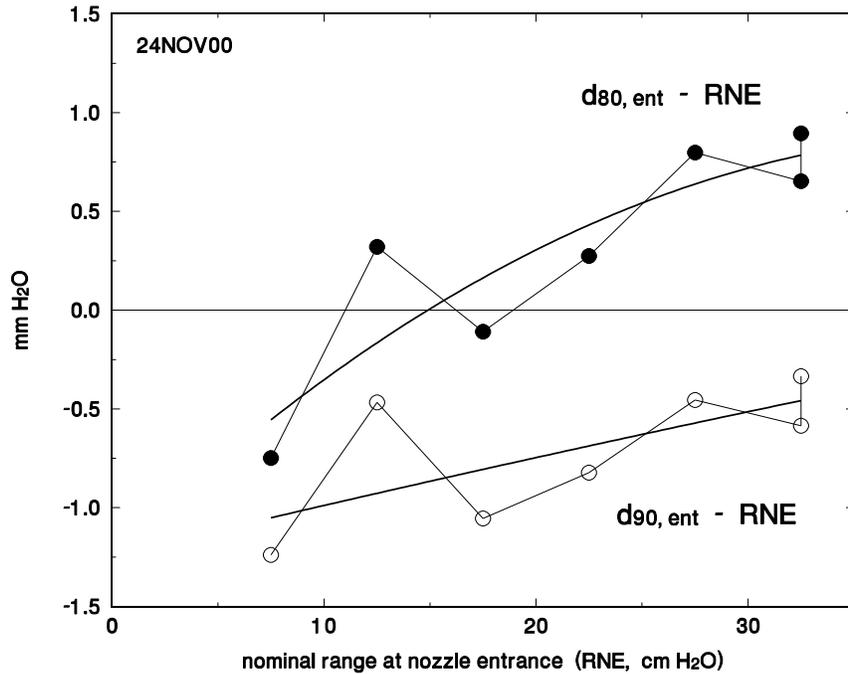


Figure 6.3: d_{80} and d_{90} vs. range at nozzle entrance' measured at NPTC.

Andy also found that (6.7) held *independent of the energy spread* he assumed. This confirms it as the true measure of mean projected range, because a mean value should not depend on the width of the underlying distribution. Figure 6.2 shows what happens in each of the three measurements when we increase the energy spread of the incident beam.³ d_{80} is the invariant point of the Bragg peak.

6.5 Clinical Depth Prescription

Usually, that the *clinical* specification refers to some level other than d_{80} , such as d_{90} . A prescription might require that the distal 90% point be at exactly 10.5 cm in water ($d_{90} \equiv 10.5$ cm). To fill this prescription we would need to

1. Measure $(d_{80} - d_{90})$ of a Bragg peak at some range close to the specified d_{90} . Suppose $(d_{80} - d_{90}) = 0.6$ mm.
2. Add 0.6 mm to the prescribed value to get $d_{80} = 10.56$ cm.
3. Supply a beam of that range.

Often, as here, the correction is barely significant. Obviously, if we need to do this frequently we should make a table or graph of premeasured values. d_{90} will depend very slightly on the energy width of the beam; d_{80} will not. Figure 6.3

³ In the top two panels, we kept the total number of incident protons the same. In the lowest one, to illustrate $R_0 = d_{80}$, we renormalized the curves to the peak value.

shows such a set of measurements. The smooth curves are polynomial fits to d_{80} and d_{90} and the difference between them is the correction we need.

In a computer program for beam design we will always have access to a measured Bragg peak, so we can compute any necessary correction on the fly:

```
CALL InitBragg(t1,.8,dmin,dl80,dpk,du80,dmax,bfn)
ypk = Peak(Bragg,dpk,dmin,dmax)
du90 = Rtfp(Bragg,dpk,dmax,.9*ypk,kErr)
dCorr = du80 - du90
```

The first line reads a measured Bragg peak file, setting up a function called `Bragg(x)` and returning d_{80} (called `du80` to distinguish it from the *lower* 80% point which we need for other things). The second line finds the peak value of `Bragg(x)` between `dmin` and `dmax`. The third line finds, by the Method of False Position [1], the point between `dpk` and `dmax` at which `Bragg(x)` crosses $.9*ypk$, and calls it `du90`.

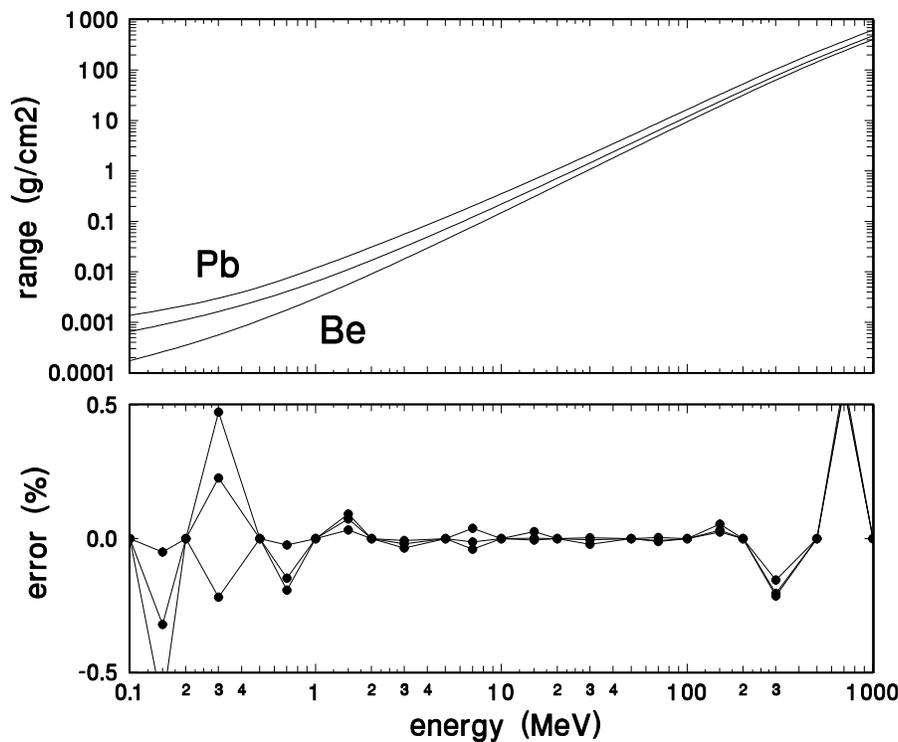


Figure 6.4: Range-energy in Be, Cu and Pb. Lower frame: test of cubic spline interpolation of points at .1, .2, .5, 1 ... MeV.

6.6 Interpolating Range-Energy Tables

Assuming, as we will from now on, that we are not in the business of computing range-energy tables but plan to use someone else's, how do we find a range between two tabulated values? *Linear* interpolation is not accurate unless the table is very closely spaced. For a table like ICRU49 [2] one should use power-law interpolation. For a not-too-wide energy interval R is given approximately by

$$R(T) \simeq a T^b \quad (6.8)$$

Fitting this to the ICRU49 [2] table for water, Bortfeld [26] found that $b = 1.77$ and $a = 0.0022$, assuming R is in cm ($= \text{g/cm}^2$ for water) and T is in MeV, gave the range to ± 1.5 mm over the range $0 < T < 200$ MeV. Incidentally, (6.8) leads immediately to

$$T(R) \simeq \frac{1}{a^{1/b}} R^{1/b} \quad (6.9)$$

and

$$S(R) \equiv -dT/dz \simeq \frac{1}{b a^{1/b}} R^{(1/b)-1} \quad (6.10)$$

by the rules of differentiation.⁴ For power-law interpolation, use adjacent values $R_1(T_1)$ and $R_2(T_2)$ to determine a and b for that interval, then use (6.8) to compute the desired R .

We want a method which can be used for any range-energy table and is more accurate than a single power-law approximation over the clinical range. Log-log plots of $R(T)$ are nearly straight lines (Figure 6.4). Using surprisingly few points $T = .1, .2, .5, 1, \dots$ we can join them with a *cubic spline* interpolation function [1] which gives us $\ln R$ vs. $\ln T$. That's equivalent to approximating R by T raised to a cubic polynomial instead of a linear one. We use natural logarithms since there's no reason not to. The accuracy of this method is shown in the lower frame of Figure 6.4, where the error for intermediate points is plotted along with that (zero, of course) for the given points. The error over the clinical range 3 – 300 MeV is less than 0.1% except for 300 MeV where it is 0.25%.⁵

This is a good time to introduce the program LOOKUP, a general-purpose proton desk calculator which computes range, energy, multiple scattering and other useful things. If you have LOOKUP you will never need to get out your range-energy tables, much less interpolate them. LOOKUP includes, among others, a module RANGE.FOR which is used as follows:

```
i = InitRange('BGWARE\DATA\ICRU49.RET')
...
rnge = Range(engy,mat1)
...
engy = Energy(rnge,mat1)
```

The call to `InitRange` reads the given range-energy table which is an ASCII file of range-energy data plus other data that will be required later for multiple scattering. You can easily add elements or compounds to this file, though ICRU49 and MIXED already have most of what you'll need for proton therapy. `InitRange` sets up the cubic spline interpolation table by constructing a list of second derivatives at the given points [1]. This is done only once, at the start of the program. Afterwards, lines like the second and third will return range and energy values. LOOKUP is merely a driver for these and similar routines, providing a convenient keyboard dialog.

For many calculations it is important that $T(R)$ be the *exact* inverse of $R(T)$. For this reason, instead of setting up a cubic spline table $\ln T(\ln R)$ as a companion to $\ln R(\ln T)$ (which would be exactly correct only at the data points), we solve the equation

$$\ln R(\ln T) = \ln R_a$$

where R_a is the range whose corresponding energy we desire and $\ln R(\ln T)$ is our cubic spline parametrization. We solve it numerically using the Method of

⁴ In (6.10) you can connect z and R by noting that a proton at depth z in a water tank has residual range $R = R_0 - z$ if the initial range was R_0 . This gives $dz = -dR$.

⁵ In earlier work we used a polynomial fit to $\ln R(\ln T)$ (see [4] Figure 1). The cubic spline is better.

False Position [1], obtaining $\ln T_a$. The root finder will converge quickly because $\ln R(\ln T)$ is nearly a straight line. Finally, $T_a = \exp(\ln T_a)$. The code reads

```

rlog = LOG(r)                ! take logarithm of R_a
x = Rtfp(RangeLog,xa,xb,rlog,k) ! find logarithm of T_a
Energy = EXP(x)             ! raise e to that power

```

6.7 Finding the Energy Loss in a Degradar

The most basic range-energy problem is to find the proton energy emerging from a degrader of known thickness if the incoming energy is given. For instance: find the energy out of 10 g/cm² of lexan assuming 160 MeV incident protons. Since we have a single degrader we'll call the material $m1$, its areal density (g/cm²) $g1$, the incoming energy (MeV) $t1$ and the outgoing energy (which is the incoming energy for the next object, whatever that is) $t2$. Formally, we wish to find $t2$ given $m1$, $g1$ and $t1$:

$$m1, g1, t1 \rightarrow t2$$

6.7.1 Using Range

The exact method is straightforward:

1. Find the proton range in material $m1$ at energy $t1$.
2. Subtract $g1$ to obtain the residual range.
3. Find the energy in $m1$ for that range. That equals $t2$.

The numbers for this method are given in the next section.

6.7.2 Using Stopping Power

Less accurate method, but slightly easier:

1. Find the stopping power S in material $m1$ at energy $t1$.
2. Multiply by $g1$ to obtain the energy loss.
3. Subtract the energy loss from $t1$ to obtain $t2$.

The numbers for each method are:

method	$t1$ MeV	S MeV/(g/cm ²)	$r1$ g/cm ²	$r2$ g/cm ²	$t2$ MeV
range-energy	160		18.523	13.523	133.66
dE/dx	160	4.962			135.19

The second method will always give too high an answer because we're looking up S at the incoming energy, and S increases as t decreases during the proton's passage through the degrader. Therefore the effective value of S is always higher than $S(t1)$. In our example the discrepancy is fairly serious because the degrader is fairly thick. In computer programs there is no real advantage in doing it this way, so we won't.

6.7.3 In a Computer Program

Assume we have at our disposal the two subprograms ('functions' in the Fortran sense) mentioned in Section 6.6:

```
Range(energy, material)
Energy(range, material)
```

We can implement the exact method as

```
r1 = Range(t1,m1)    ! total range in material m1
rr = r1 - g1         ! residual range in m1
t2 = Energy(rr,m1)  ! energy corresponding to that range
```

or condense the code to

```
t2 = Energy(Range(t1,m1)-g1,m1)
```

6.7.4 Exercise

As an example of another kind of problem, suppose that 30 MeV protons emerge from 10 g/cm² of lead (Pb). What was their energy going in? (a) Write down the formal problem, (b) write down the formal solution and (c) find the answer using LOOKUP. Do the problem both ways: exactly, and using dE/dx . The answers for the exact method are

```
m1, g1, t2 → t1

rr = Range(t2,m1)    ! residual range in material m1
r1 = rr + g1         ! total range going into m1
t1 = Energy(r1,m1)  ! energy corresponding to that range

t1 = Energy(Range(t2,m1)+g1,m1)

83.4 MeV
```

Make up little problems like this and make sure you can do them. It is helpful to organize your numbers in a little sketch of the problem.

6.7.5 Error Conditions

When we specify an outgoing energy we can always find an incoming energy no matter what the degrader thickness (unless we exceed the available range-energy table). When we specify an incoming energy and a thick degrader, the protons could range out in the degrader and never emerge. The first problem is always well posed, the second may not be. A good program will test for such error conditions and take appropriate action. Programmers note: usually the 'appropriate action' takes a good deal more thought than the test itself. If this is part of a big program, we don't necessarily want to bring everything to a halt!

6.8 Water Equivalence

We rarely use water equivalence in our programs, preferring to switch back and forth between $R(T)$ in different materials, using T as the common coin between degraders. However, water equivalence is frequently used, so we'll briefly discuss it and point out the pitfalls. Here we'll occasionally refer to actual thickness (cm) rather than (g/cm²) so read carefully. When we're talking about water it makes no difference because $\rho_{water} = 1$ g/cm² by definition.

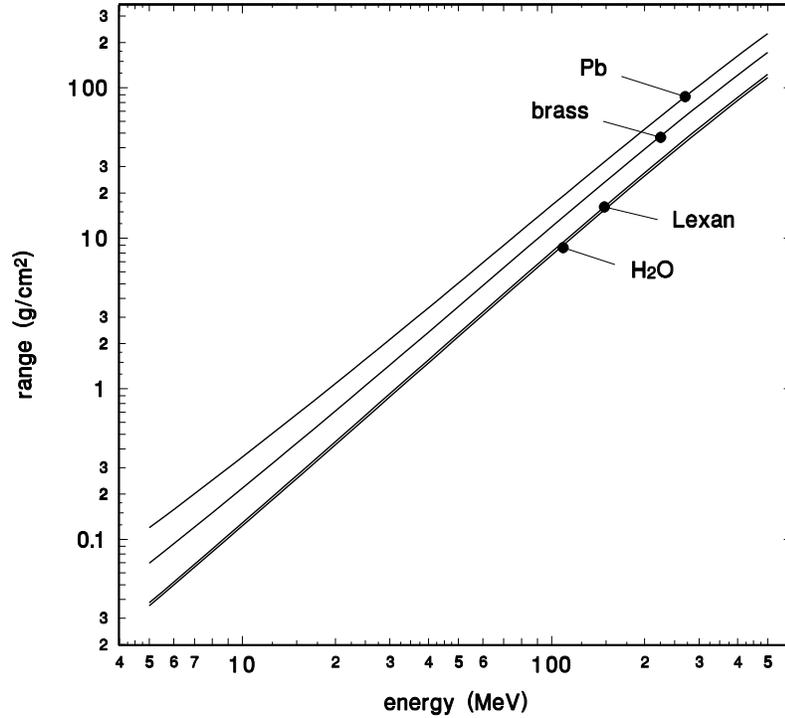


Figure 6.5: Range-energy over the clinically useful energy range for 4 useful materials.

6.8.1 Water Equivalence of Thin Degraders

The *water equivalent* of a degrader is the thickness of water that has the same energy loss. More generally, a degrader of material 2 is said to be equivalent to a degrader of material 1 if the energy loss is the same. From $S \equiv -(1/\rho)(dE/dx)$ we obtain

$$S_2(T) \rho_2 \Delta x_2 = S_1(T) \rho_1 \Delta x_1$$

which gives, for the lexan/water case

$$\Delta x_W = \frac{S_L(T)}{S_W(T)} \frac{\rho_L}{\rho_W} \Delta x_L \quad (6.11)$$

We already know enough to be wary of equations like (6.11) if the degraders are not thin. But there can be problems even if they are. The answer can depend on energy, which makes it much less useful. The predicted water equivalent of 1 cm of lexan is 1.143 cm at 200 MeV and 1.144 cm at 70 MeV: no problem. But the water equivalent of 0.62 cm of lead (the thickness of one of the second scatterers in the NPTC nozzle) is 3.556 cm at 200 MeV and 3.355 cm at 70 MeV. The 2 mm difference is just enough to be troublesome, and the scatterer is in fact exposed to that range of energies in clinical operations.

When we're safe, and when not, can be seen from Figure 6.5, a log-log plot of $R(T)$ for four useful materials. The slopes are slightly different as we move from low- Z to high- Z materials. That means the stopping power ratios in equations like (6.11) will depend on energy. 'Equivalence' is a good concept for two materials if they're close in the periodic table (water and plastic) but not if they're far apart (water and lead).

6.8.2 Equivalence in a Finely Divided Stack

In the case of a multi-layer Faraday Cup (MLFC), frequently used as a range verifier (RV), we have a finely-divided stack of two materials. For instance the range verifier at NPTC is a stack of 2 mm brass sheets separated from each other by 0.05 mm of polyethylene. We would like to characterize measurements with this RV as ‘range in brass’. How do we correct for the polyethylene? While the protons stop, they encounter brass and polyethylene at all energies from the incident energy T to zero. What energy should we use to find the average stopping power ratio? The answer, which can be found numerically with range-energy tables, is $0.52 \times T$. Ranges in any two materials are related by

$$\left. \frac{R_2}{R_1} \right|_T = \left. \frac{S_1}{S_2} \right|_{0.52T} \quad (6.12)$$

If we have a stack of equal plates of brass, each B g/cm², separated by equal sheets of polyethylene, each P g/cm², the average brass equivalent $\langle B_P \rangle$ of a polyethylene sheet is

$$\langle B_P \rangle = \left. \frac{S_P}{S_B} \right|_{0.52T} \times P \quad (6.13)$$

This rule works for any finely-divided stack independent of the materials and their relative thicknesses. The stopping power ratio depends on energy if the materials are far apart on the periodic table. However, if the minor material is thin compared to the other, that effect is second-order.

6.8.3 Measured Water Equivalence

Though we are not in general fond of water equivalence, it has one very useful property. We can *measure* it with equipment always found at proton therapy centers, and the measurement is trivial. Simply measure d_{80} of a Bragg peak in a water tank, then put the desired degrader in front of the tank and remeasure d_{80} . With good instrumentation you can do this to ± 0.1 mm. The measured water equivalence found in that way can be used in the final stages of a design instead of range-energy tables, which we have already seen are not quite good enough because of their dependence on I . At HCL we measured the water equivalence of lexan and lucite (plexiglas) a number of times and found

$$\text{water equivalent of 1 cm lexan or lucite} = 1.15 \text{ cm}$$

which disagrees very slightly with the computed value given earlier.

If you measure the water equivalence of lead this way, you’ll find it depends on energy as we have already noted. Sometimes it happens (see Chapter 13) that the design conditions are such that the lead always sees the same, or nearly the same, energy. In that case, improving the depth accuracy of the design by using a measured water equivalent thickness for the lead can still be a useful thing to do.

6.9 Does the Order of Degraders Matter?

The answer to this question is based on considerations of the previous section. The second degrader in a stack sees a lower energy than the first. If that doesn’t matter, that is, if the stopping powers in the two materials vary the same way with energy, then the order doesn’t matter. Putting it another way, the order of two materials doesn’t matter if their $R(T)$ are parallel on a log-log plot (Figure

6.5). Therefore, the order matters only if the stack mixes high-Z and low-Z materials. Check this using LOOKUP with lexan and lead. The effect will be largest if both degraders are fairly thick (a significant fraction of the range).

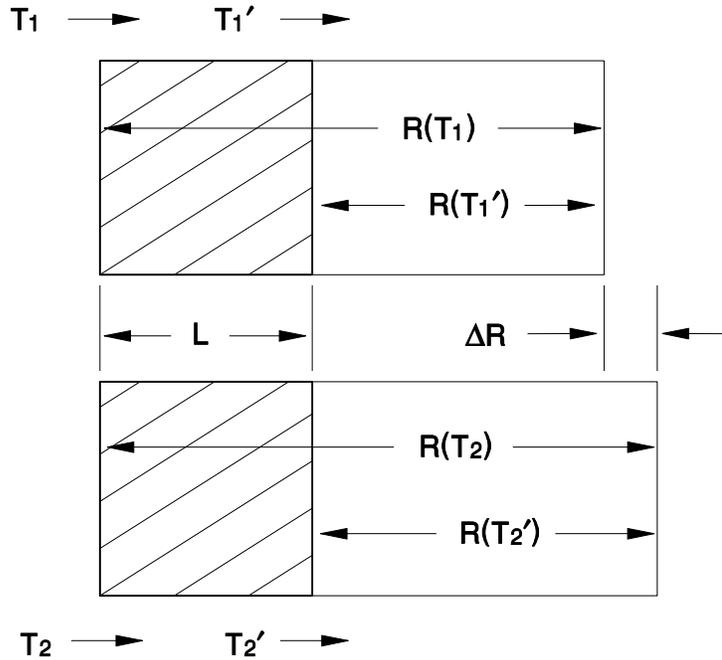


Figure 6.6: Drawing for proof of Eq. (6.14).

6.10 Energy Spread of Degraded Beams

What is the energy distribution coming out of a degrader if the input energy spread is ΔT ? Normally we're not interested in energy distributions. In most design problems the effect of a degrader is entirely taken into account by saying that it pulls back the number *vs.* depth distribution, or the Bragg peak, by the water equivalent of the degrader. But suppose we are setting up a beam for a non-clinical user who wants to explore the radiation sensitivity of a semiconductor device. Such a user *is* interested in the energy and usually, in the energy spread as well. This problem too was solved by Andy Koehler who found

$$\frac{\Delta T'}{\Delta T} \simeq \frac{S(T')}{S(T)} \quad (6.14)$$

where unprimed stands for initial, primed for final, and ΔT stands for some consistent measure (the standard deviation of a Gaussian or the FWHM of the energy distribution) of width. To prove 6.14 consider Figure 6.6.⁶ The cross-hatched part represents the degrader and the remainder shows the residual range for two nearly equal incoming energies T_1 and T_2 . From the definition of range

$$S = (T_2 - T_1)/(R(T_2) - R(T_1))$$

⁶ This proof and that of the following section can also be found in [28].

or

$$R(T_2) - R(T_1) = (T_2 - T_1)/S(T) \quad (6.15)$$

For protons *leaving* the degrader we have, similarly,

$$R(T'_2) - R(T'_1) = (T'_2 - T'_1)/S(T') \quad (6.16)$$

But the left-hand sides of (6.15) and (6.16) are just two measures of the same quantity ΔR . Therefore the right-hand sides are equal and 6.14 follows. Note that we have only used the fact that range is the integral of inverse stopping power and have not assumed any particular form for $R(T)$.

The approximation sign in (6.14) reflects the fact that, as the output energy is made lower and lower, the energy distribution is skewed by the nonlinear range-energy relation. In a recent experiment [29] at NPCTC the energy spread of degraded beams was measured in a sodium iodide scintillation spectrometer.⁷ The results are shown in Figure 6.7. Eq. (6.14) holds to a good approximation, but [29] includes a more detailed analysis of the growth in energy width.

6.11 Energy Stability of an Accelerator

As a final example of range-energy calculations, consider the following practical problem. We wish to irradiate a structure at depth L in some material ν , holding a tolerance ΔL . Before entering the target material the beam passes through other degraders $1, 2 \dots, (N-1), N$ of materials $\alpha, \beta \dots \mu, \nu$ (Figure 6.8). How tight a tolerance ΔT must we hold on the incident energy? This is a typical situation in passive beam spreading. Letting T_i and T'_i be two closely spaced energies and T_N be their mean entering the i^{th} degrader we have, starting from the end of the stack

$$\Delta L_N \equiv R_\nu(T'_N) - R_\nu(T_N) = \frac{dR_\nu(T)}{dT} \Delta T_N = \frac{\Delta T_N}{S_\nu(T_N)}$$

We now use (6.14) as often as necessary, working backwards through the stack, to relate the energy spread ΔT at each degrader to the energy spread into the previous degrader, obtaining

$$\Delta L_N = \frac{S_\alpha(T_2)}{S_\alpha(T_1)} \times \frac{S_\beta(T_3)}{S_\beta(T_2)} \times \dots \times \frac{S_\mu(T_N)}{S_\mu(T_{N-1})} \times \frac{\Delta T_1}{S_\nu(T_N)} \quad (6.17)$$

This is as far as we can go without assuming an approximate form for $R(T)$. Let, as usual, $R(T) \simeq aT^b$ which implies $S(T) \simeq T^{1-b}/ab$. Make the additional approximation that b is the same for all materials (the lines in Figure 6.5 are approximately parallel), substitute for each S in (6.17) and simplify to obtain

$$\Delta L \simeq b R_\nu(T_1) \frac{\Delta T_1}{T_1}, \quad b \simeq 1.77 \quad (6.18)$$

This is an interesting result. To the extent that stopping power ratios at two energies are independent of the material, the particular materials overlying the target are irrelevant. What spoils the precision (that is, increases ΔL for a given $\Delta T_1/T_1$) is the extra energy consumed by the overlying materials.

⁷ To do this the beam intensity has to be greatly reduced from clinical levels, but there is good evidence that this does not affect the beam energy spread.

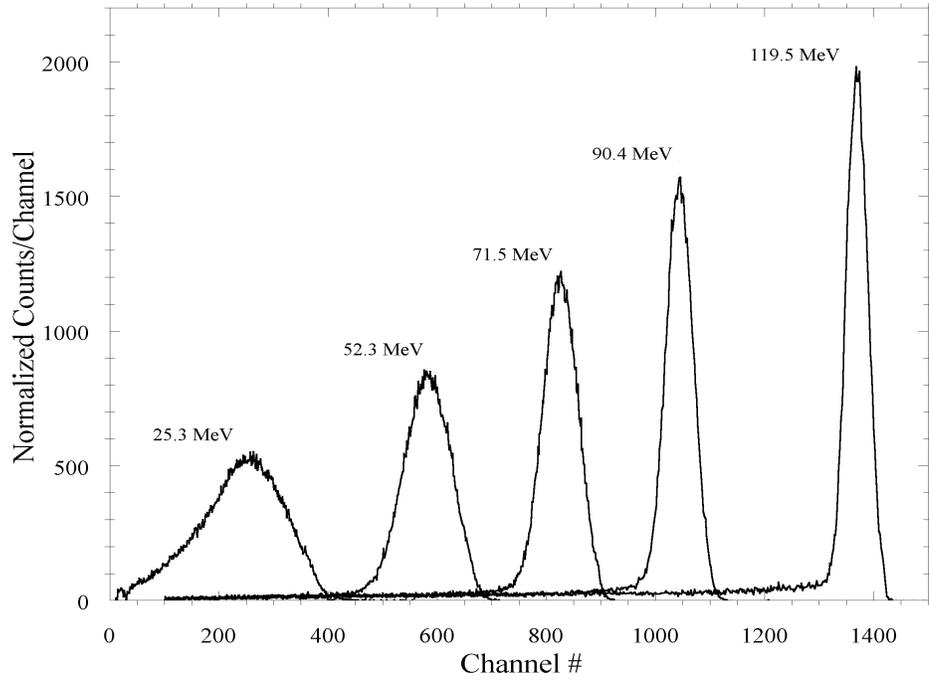


Figure 6.7: Energy spectra of degraded proton beams at NPTC. Energy spread and skewness increase as the energy is reduced.

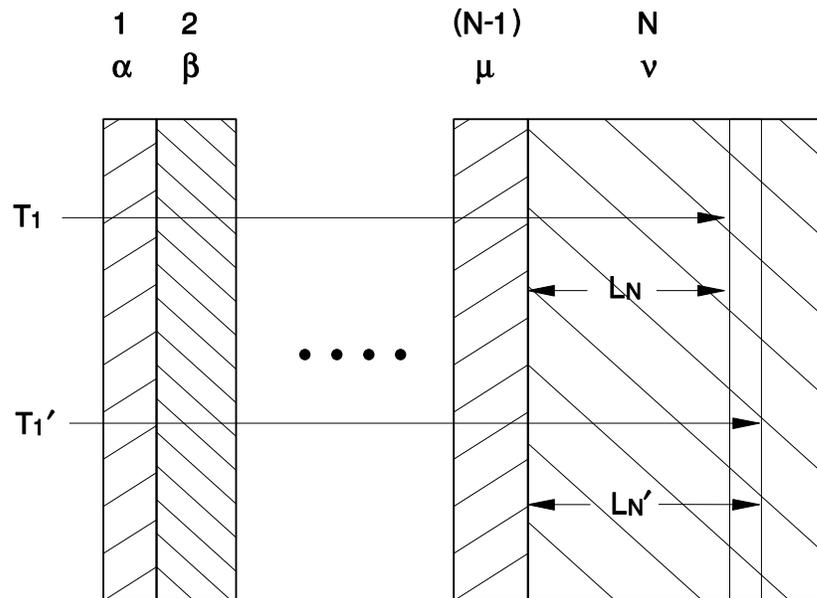


Figure 6.8: Drawing for proof of Eq. (6.18).

Chapter 7

Scattering

As protons slow down they also scatter, mainly by interaction with atomic nuclei rather than electrons, because collisions with electrons hardly deflect the massive proton. Most single deflections, even by nuclei, are very small, so it is the net result of very many such events that we must compute. Because the interaction is electromagnetic, through the Coulomb force, the phenomenon is called *multiple Coulomb scattering*.

The subject received wide attention in the 1930's and 40's. It was ultimately laid to rest in two papers by Molière,¹ whose solution to the problem is now considered definitive. In the first paper [6] Molière investigates single scattering in the Coulomb field of the nucleus, which is screened by the atomic electrons. One needs a good model of single scattering before one can address multiple scattering. In the second [7] he applies this to plural scattering (just a few scatters) and multiple scattering (very many scatters). His elegant and comprehensive treatment includes the effect of energy loss by the primary particle, and allows the stopping medium to be a chemical element or a compound or mixture. Both extensions of the theory are very important in proton therapy design, as we often deal with thick scatterers or compounds.

In 1993 an HCL group compared our and all other existing proton data with Molière's theory [4]. You should read [4] and some of the references therein if you wish to pursue the subject beyond these notes. By then the literature had become somewhat tangled. There were at least four competing 'exact' theories with different domains of validity, whose relationship to each other is discussed by Bethe [5]. Some experimental groups were unaware that Molière had treated energy loss² and used 'effective energy' approaches to fill the perceived gap. Others found that Molière's predictions were incorrect. At least one theoretical paper [30] claimed to correct flaws in Molière's theory.

The HCL paper reviewed and in some cases reanalyzed all the experimental data and found that Molière theory was correct. Figure 7.1 reproduced from [4] is a grand summary of the deviation of experimental data from Molière theory for seven experiments spanning 1 MeV to 200 GeV incident energy. Data from each experiment have been averaged over everything but target material. The distribution of 39 independent measurements is normal with a mean value $-0.3 \pm 0.5\%$ and an *rms* spread of 3%.

The sense in which 'discrepancy from experiment' is defined in Figure 7.1

¹ Despite the name Molière was German. He may have been a student of Heisenberg, who is thanked in both papers. The second paper was his 'Habilitationsschrift' at the University of Tübingen, a publication which had to be formally accepted before one was admitted to the faculty. He went on to make fundamental contributions in several fields of physics.

² Perhaps because Bethe's paper [5], more accessible to the English speaker than Molière's, omitted that aspect of the theory.

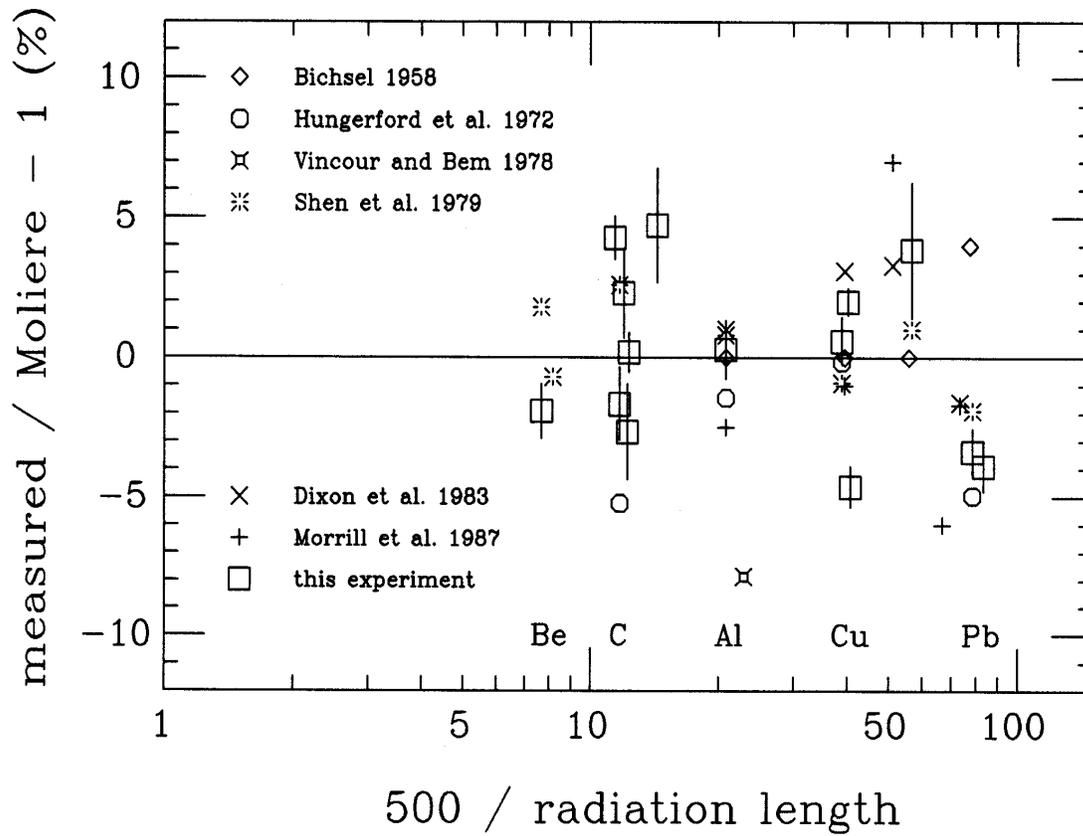


Figure 7.1: Grand summary of proton results reviewed in [4]. The x coordinate is arbitrary and the target names merely serve as landmarks. Data from each experiment are averaged over everything but target material.

means that Molière is high by 0.3%, but this is statistically consistent with 0. In other words, one cannot say how good Molière theory actually is for protons because experimental errors, typically a few percent, dominate the comparison. Future experiments to test the theory more rigorously need to be considerably more accurate, which will not be easy.

The only substantial theoretical improvement to Molière theory was the Fano correction [31] for scattering by atomic electrons. This is appreciable only for very light elements whose nuclei have low charge. This correction had been approximated earlier by Bethe by just substituting $Z(Z + 1)$ for Molière's Z^2 in the formula for χ_c^2 . That is much simpler and works almost as well [4].

Molière theory has no material dependent parameters, in fact no adjustable parameters at all. It is true that Molière introduces an approximate numerical expression (7.4) in evaluating the Thomas-Fermi model of the atom in order to compute screening. However, that result has since been confirmed by computer calculations [4, 32] and is not material specific.

There are no ‘scattering tables’ analogous to range-energy tables. One must compute the characteristic multiple scattering angle from Molière's equations or use one of the (far simpler) approximate formulas we will give. Our Fortran function `Theta0` incorporates all of Molière theory as well as the approximate theories. It computes the equations given below, and was used to calculate the numbers in [4] and in Table 7.2.5.

7.1 Central Limit Theorem

Let θ be the angle between the particle direction and the longitudinal axis and let θ_0 be a characteristic angle. If we assume that multiple scattering can be computed by taking the quadratic sum of small angular kicks we will find that the angular distribution is Gaussian

$$f(\theta, \phi) \theta d\theta d\phi = \frac{1}{2\pi \theta_0^2} e^{-\frac{1}{2}\left(\frac{\theta}{\theta_0}\right)^2} \theta d\theta d\phi \quad (7.1)$$

and that the width parameter θ_0 is proportional to \sqrt{t} where t is the scatterer thickness. Both are consequences of the Central Limit Theorem, which states that the distribution of the sum of small random kicks tends toward a Gaussian as the number of kicks is increased. However, a condition for the theorem to apply is that all the random kicks must be *small* in a sense that can be made mathematically precise [33]. This condition is not met for the EM scattering of protons by the nucleus (Rutherford scattering). To put it another way, as θ increases, Rutherford scattering falls off as $1/\theta^4$, much more slowly than a Gaussian. Therefore if (7.1) were to hold we would reach the absurd conclusion that, at some angle, multiple scattering is less than single scattering. Multiple scattering must approach $1/\theta^4$ at ‘large’ angles and be greater at small angles. This means that the *full* angular distribution cannot be Gaussian, though it will turn out to be nearly Gaussian at small angles, and that its characteristic width cannot be exactly proportional to \sqrt{t} .

7.2 Molière Theory

What follows is the bare bones of Molière's algebra. For fuller discussions see the HCL paper [4], Bethe [5], the extensive review by Scott [34] and, of course, Molière [6, 7]. In place of the Gaussian's single parameter θ_0 , Molière theory has

two, χ_c (an angle) and B (a dimensionless quantity). A characteristic multiple scattering angle θ_M can be defined in terms of these as

$$\theta_M = \frac{1}{\sqrt{2}} (\chi_c \sqrt{B})$$

and this plays the same role in Molière theory as θ_0 does in the Gaussian. The Molière angular distribution depends strongly on θ_M and weakly on B , which carries the dependence on target thickness.

7.2.1 Thin Scatterer

Following Molière we first assume a thin scatterer (so kinematic factors can be considered constant) made of a single chemical element. Later we'll add energy loss, compounds and mixtures, and finally the Fano correction at low Z .

The *characteristic single scattering angle* χ_c is given by

$$\chi_c^2 = c_3 t / (pv)^2 \quad (7.2)$$

where t is the target thickness (g/cm²), p and v the particle momentum and speed, and

$$c_3 \equiv 4 \pi N_A \left(\frac{e^2}{\hbar c} \right)^2 (\hbar c)^2 \frac{z^2 Z^2}{A} \quad (7.3)$$

where Z and A are the atomic number and atomic weight of the target material, z is the charge number (charge/ e) of the incident particle, N_A is Avogadro's number, $(e^2/\hbar c)$ is the fine structure constant and $(\hbar c)$ is the usual conversion constant.³ χ_c has the physical meaning that, on the average, a particle suffers exactly one single scatter greater than χ_c in its traversal of the target [7, 4].

To find B we must next compute a *screening angle* χ_a from

$$\chi_a^2 = \chi_0^2 (1.13 + 3.76 \alpha^2) \quad (7.4)$$

where

$$\chi_0^2 = c_2 / (pc)^2 \quad (7.5)$$

and the *Born parameter* α is given by

$$\alpha^2 = c_1 / \beta^2 \quad (7.6)$$

The constants are defined by

$$c_1 \equiv \left[\left(\frac{e^2}{\hbar c} \right) z Z \right]^2 \quad (7.7)$$

and

$$c_2 \equiv \left[\frac{1}{0.885} \left(\frac{e^2}{\hbar c} \right) (m_e c^2) Z^{1/3} \right]^2 \quad (7.8)$$

β stands for v/c of the incident particle and $m_e c^2$ is the electron rest energy. Next we compute a quantity

$$b = \ln \left(\frac{\chi_c^2}{1.167 \chi_a^2} \right) \quad (7.9)$$

b is interpreted as the natural logarithm of the effective number of collision in the target. Finally the *reduced target thickness* B is defined as the root of the equation

$$B - \ln B = b \quad (7.10)$$

³ Chapter 2 gives the values of physical constants.

which can be solved by standard numerical methods (a root finder). It turns out that B is almost proportional to b in the region of interest. At last, we can compute the *characteristic multiple scattering angle*

$$\theta_M = \frac{1}{\sqrt{2}} (\chi_c \sqrt{B}) \quad (7.11)$$

We should note here that standard Molière theory speaks instead of $\chi_c \sqrt{B}$, which is the *rms* value of the space angle θ . We have introduced the $\sqrt{2}$ to put θ_M on the same footing as the Gaussian parameter θ_0 .

We are now positioned to compute the distribution of θ . Define a reduced angle variable θ' :

$$\theta' \equiv \frac{\theta}{\chi_c \sqrt{B}} \quad (7.12)$$

Molière approximates the desired distribution function $f(\theta)$ by a three term power series in $1/B$:

$$f(\theta) = \frac{1}{2\pi \theta_M^2} \frac{1}{2} \left[f^{(0)}(\theta') + \frac{f^{(1)}(\theta')}{B} + \frac{f^{(2)}(\theta')}{B^2} \right] \quad (7.13)$$

where the $f^{(n)}$ are given by

$$f^{(n)}(\theta') = \frac{1}{n!} \int_0^\infty y dy J_0(\theta' y) e^{y^2/4} \left(\frac{y^2}{4} \ln \frac{y^2}{4} \right)^n \quad (7.14)$$

$f^{(0)}$ is simply a Gaussian:

$$f^{(0)}(\theta') = 2 e^{-\theta'^2} \quad (7.15)$$

Molière and Bethe [5] give further formulas and tables for $f^{(1)}$ and $f^{(2)}$. We have used Bethe's tables in our function THETA0.

Eq.(7.13) differs from equivalent equations in Bethe and Molière by the factor $(1/2\pi)$ which we have included to obtain

$$\int_0^{2\pi} \int_0^\infty f(\theta) \theta d\theta d\phi = 1 \quad (7.16)$$

which is frequently more convenient than Molière's normalization and leads to the conventional equation in the Gaussian approximation. Since Molière theory only holds for small θ_M and θ , ' ∞ ' in (7.16) means some suitable upper limit.

The foregoing equations, with Bethe's tables [5], permit one to compute the characteristic multiple scattering angle θ_M and to evaluate the scattering probability density $f(\theta)$ under any conditions provided the target consists of a single chemical element and the energy loss is small. Except for rearrangements of physical constants to conform to modern usage, and the normalization of $f(\theta)$, the equations are identical to Molière's.

7.2.2 Energy Loss

To account for energy loss Molière generalized (7.2) to

$$\chi_c^2 = c_3 \int_0^t \frac{dt'}{(pv)^2} \quad (7.17)$$

and in place of (7.4) defined an effective value $\bar{\chi}_a^2$ by

$$\ln \bar{\chi}_a^2 = \frac{c_3}{\chi_c^2} \int_0^t \frac{dt'}{(pv)^2} \ln \chi_a^2(t') \quad (7.18)$$

where $\chi_a^2(t')$ is found at each depth t' in the target by means of (7.4), (7.5) and (7.6) using the values of p and β appropriate at that depth. These generalizations preserve the form of (7.9) so that it and subsequent equations through (7.16) remain valid. For instance (7.9) becomes

$$b = \ln\left(\frac{\chi_c^2}{1.167 \bar{\chi}_a^2}\right) \quad (7.19)$$

and so forth.

7.2.3 Compounds and Mixtures

For compounds and mixtures Molière introduced a sum over constituents in addition to the integral over depth. Now

$$\chi_c^2 = \left(\sum_i c_{3i}\right) \int_0^t \frac{dt'}{(pv)^2} \quad (7.20)$$

and

$$\ln \bar{\chi}_a^2 = \frac{1}{\chi_c^2} \sum_i \left[c_{3i} \int_0^t \frac{dt'}{(pv)^2} \ln \chi_{ai}^2(t') \right] \quad (7.21)$$

where i runs over the number of elementary constituents in the target. χ_{ai} is obtained from (7.4), (7.5) and (7.6) with the target constants c_{3i} , c_{2i} and c_{1i} defined as before with the replacements $Z \rightarrow Z_i$ and $(1/A) \rightarrow (f_i/A_i)$ where f_i is the fraction by weight of the i^{th} constituent of the compound or mixture. Eqs.(7.9) through (7.16) still hold with these new definitions of χ_c^2 and χ_a^2 .

Implicit in the integrals involving pv is the assumption that energy is a well-defined function of depth in the target. This fails for targets so thick that the energy spread from range straggling is comparable to the mean energy. For these ‘very thick’ targets we cannot expect Molière theory or any similar analytic theory to apply. As we shall see, the data suggest that this occurs when the target thickness exceeds $\approx 97\%$ of the mean proton range.

7.2.4 Fano Correction

Fano [31] derived the correction to Molière theory for scattering of the incident particle by atomic electrons in addition to the screened Coulomb field of the nucleus. The expedient recommended by Bethe [5], replacing Z in formulas for χ_c^2 by $Z(Z+1)$, only works as long as the scattering by atomic electrons is well approximated by the Rutherford formula. Therefore it is not exactly correct even for incident electrons, and less so for incident heavy particles. For these Fano recommends using Molière’s original definition of χ_c^2 (that is, retaining Z^2) and adding to Molière’s b a quantity D/Z where

$$D = \ln\left[\frac{1130}{Z^{4/3}((1/\beta^2) - 1)}\right] - u_{in} - \frac{1}{2} \beta^2 \quad (7.22)$$

and $-u_{in}$ is an integral over the incoherent scattering function. Fano quotes several calculations of $-u_{in}$ and suggests using the value $-u_{in} = 5$. Eq.(7.22) only covers the thin elementary target case because D is a function of incident particle energy as well as target species. Fortunately Scott [34] fills this gap, showing how to include Fano’s correction in the theory for thick compound targets. The formula for χ_c^2 is unaffected, while (7.21) becomes

$$\ln \bar{\chi}_a^2 = \frac{1}{\chi_c^2} \sum_i \left[c_{3i} \int_0^t \frac{dt'}{(pv)^2} \left(\ln \chi_{ai}^2(t') - \frac{D_i}{Z_i} \right) \right] \quad (7.23)$$

(From here on we'll drop the bar over $\bar{\chi}_a^2$ because there is no risk of confusion.) Scott uses $-u_{in} = 5.8$, the value Fano quotes for the Thomas-Fermi model, instead of 5, the value recommended by Fano and the value we shall use.

In summary, what we call ‘Molière theory’ will be defined by equations (7.20) for χ_c^2 and (7.23) for χ_a^2 along with the equations of the first section with kinematic and target dependences included as necessary. Although the full generality of this approach is required only for thick, compound, low- Z targets it is valid as well in the case of thin, elementary high- Z targets so we will use it throughout.

7.2.5 Evaluation of the Integrals

We will evaluate the integrals in Eqs.(7.20) and (7.23) numerically using the cubic spline parametrization of the range-energy relation described in Chapter 6. The integrals can be approximated as sums over finite slabs of the target using *Simpson's Rule* [1]

$$\int_{x_a}^{x_c} f(x) dx \approx \frac{1}{6}(f_a + 4f_b + f_c) \Delta x \quad (7.24)$$

where $\Delta x \equiv x_c - x_a$ is the total thickness of the slab and x_b is the midpoint. Simpson's Rule [1] is exact if $f(x)$ is a cubic or lower polynomial. It minimizes function calls because f_c can be recycled as f_a for the next interval.

Because of the $1/(pv)^2$ kinematic factor the integrand increases rapidly as the proton slows down. Therefore it is inefficient to divide the target into slabs of equal thickness. Instead, we divide it equally in the logarithmic sense, that is, by equal thickness *ratios*. Each slab will be (say) $1.6\times$ as thick as the previous one. To accomplish this let R_1 be the range, in the target material, of the proton entering the target and let R_2 be its residual range leaving the target. We wish to divide t into N slabs t_i such that the ratio $\rho > 1$ of each thickness to the previous one will be the same. This can be done by letting

$$\rho = (R_1/R_2)^{1/N}$$

and

$$t_1 = (\rho - 1)R_2, \quad t_2 = \rho(\rho - 1)R_2, \quad \dots, \quad t_N = \rho^{N-1}(\rho - 1)R_2.$$

As a check

$$t_1 + \dots + t_N = (1 + \rho + \dots + \rho^{N-1})(\rho - 1)R_2 = (\rho^N - 1)R_2 = R_1 - R_2$$

by Dwight 26 [20]. Rather than fixing N , leading to needless subdivision of thin targets, we let the computer choose N for each case so that ρ will not exceed some given ρ_{max} . Then $N = 1$ for any thin target.⁴ With $\rho_{max} = 1.6$ we obtain better than 0.1% computational accuracy in θ_M , and this leads to a mere 8 slabs for $R_2 \approx 0.03R_1$ beyond which, as noted earlier, the theory no longer applies.

7.2.6 Numerical Results

Table 7.2.5 lists Molière parameters as well as the final result θ_M for 158.6 MeV protons incident on Be, Cu and Pb targets of normalized thicknesses $t' \equiv t/R_0$ ranging from 0.0001 to 0.97. It is an abridged and slightly altered version of Table 1 in [4]. Slight differences in the results are probably due to the use

⁴ Using Simpson's Rule (7.24) even for a single slab already guarantees we do much better than picking a single value for pv .

mt'l	t/R	t g/cm ²	T_2 MeV	α	χ_0 millirad	χ_a millirad	χ_c millirad	b	B	θ_M millirad	N
Be	0.0001	0.00213	158.59	0.0564	0.0118	0.00406	0.0828	5.874	7.947	0.165	1
	0.0010	0.02129	158.51	0.0564	0.0118	0.00406	0.2618	8.177	10.531	0.601	1
	0.0100	0.21290	157.69	0.0565	0.0118	0.00407	0.8298	10.481	13.050	2.120	1
	0.1000	2.12899	149.32	0.0577	0.0122	0.00414	2.6913	12.798	15.541	7.502	1
	0.5000	10.64497	107.00	0.0662	0.0145	0.00463	6.9757	14.479	17.332	20.535	2
	0.9000	19.16094	43.71	0.0989	0.0231	0.00629	13.1010	15.129	18.021	39.325	5
	0.9700	20.65124	22.49	0.1357	0.0324	0.00795	16.5681	15.131	18.022	49.735	8
Cu	0.0001	0.00263	158.59	0.4086	0.0228	0.02705	0.2510	4.301	6.111	0.439	1
	0.0010	0.02626	158.51	0.4087	0.0228	0.02706	0.7938	6.603	8.775	1.663	1
	0.0100	0.26258	157.67	0.4095	0.0229	0.02711	2.5163	8.907	11.335	5.990	1
	0.1000	2.62578	149.14	0.4187	0.0236	0.02764	8.1651	11.222	13.851	21.487	1
	0.5000	13.12892	106.08	0.4819	0.0282	0.03124	21.2363	12.889	15.639	59.383	2
	0.9000	23.63205	42.22	0.7290	0.0455	0.04497	40.3613	13.445	16.232	114.984	5
	0.9700	25.47010	21.14	1.0137	0.0646	0.06130	51.6584	13.319	16.097	146.556	8
Pb	0.0001	0.00361	158.59	1.1553	0.0322	0.07749	0.4605	3.410	5.024	0.730	1
	0.0010	0.03606	158.51	1.1556	0.0322	0.07751	1.4566	5.712	7.762	2.869	1
	0.0100	0.36057	157.65	1.1580	0.0323	0.07769	4.6176	8.015	10.353	10.506	1
	0.1000	3.60573	148.96	1.1844	0.0333	0.07963	14.9909	10.321	12.877	38.038	1
	0.5000	18.02864	105.27	1.3671	0.0401	0.09304	39.1116	11.928	14.610	105.709	2
	0.9000	32.45156	40.97	2.0906	0.0653	0.14718	75.0959	12.315	15.025	205.830	5
	0.9700	34.97557	19.95	2.9477	0.0941	0.21646	97.2268	12.060	14.752	264.054	8

Table 7.1: Molière parameters for 158.6 MeV protons incident on Be, Cu and Pb using the ICRU49 [2] range-energy tables. N is the number of target slabs used for numerical integration. See text for other definitions.

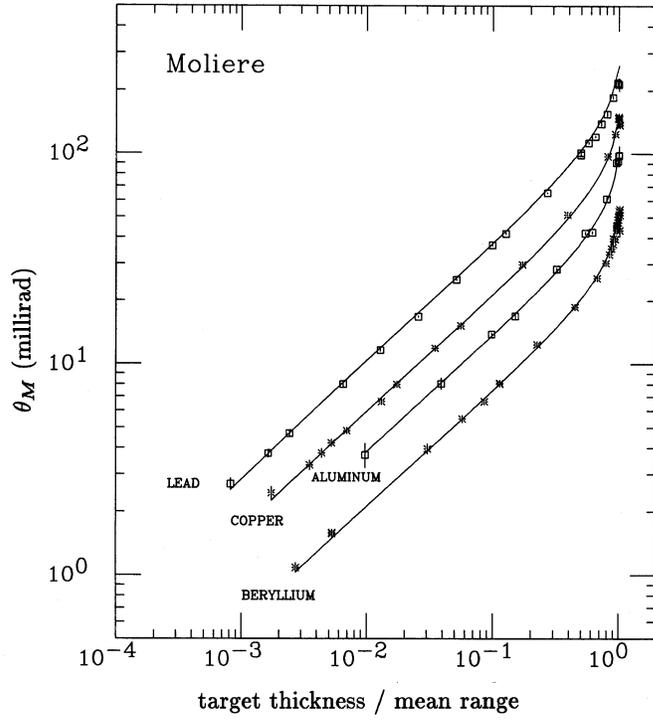


Figure 7.2: Molière/Fano theory *vs.* experiment for four metals [4]

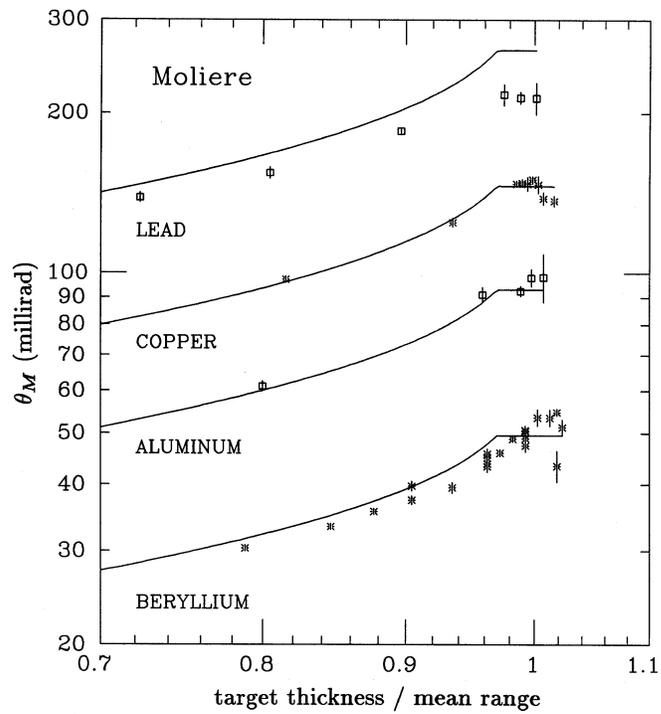


Figure 7.3: Expanded view of Figure 7.2 near end of range.

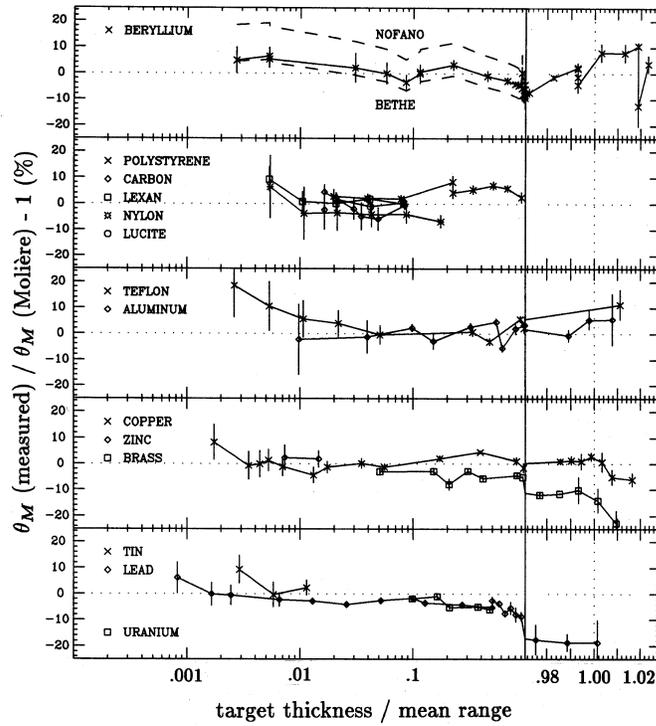


Figure 7.4: Deviation from Molière/Fano theory as a function of normalized target thickness for HCL data [4].

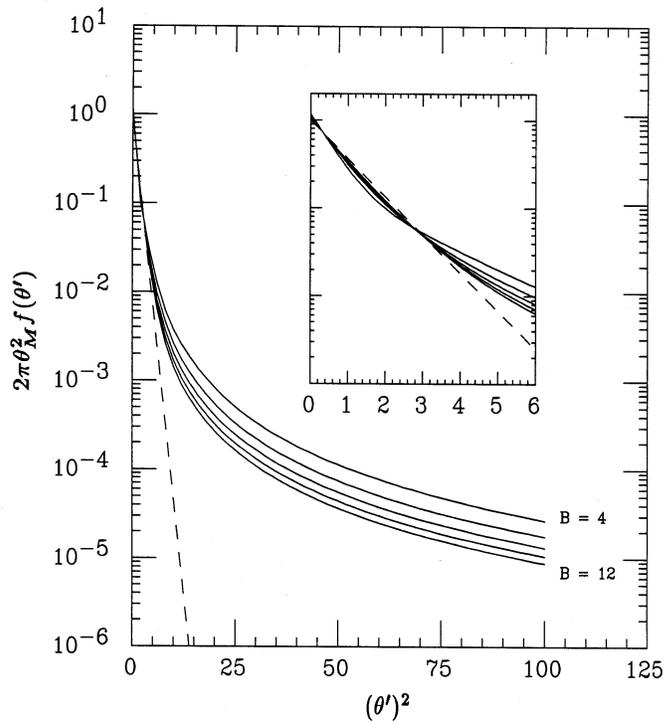


Figure 7.5: Molière angular distribution. Dashed line: $f^{(0)}$ term (Gaussian) only. Inset: expanded view near origin.

of different range-energy tables and the cubic spline instead of the polynomial parametrization of the range-energy relation.

A study of this table reveals that Molière’s reduced target thickness B is very nearly proportional to the logarithm of the normalized target thickness, with a constant of proportionality that depends upon the material.

Figure 7.2, from [4], shows θ_M for four metals over three decades of normalized target thickness. The data points are from the HCL experiment. The range of this parameter-free theory is astounding, and its accuracy holds up under closer scrutiny (Figure 7.4). See [4] for details of the experiment and analysis.

Figure 7.3 is an expanded view of Figure 7.2 near end-of-range, showing a leveling off of the multiple scattering angle for protons that just make it out of the target. There is no theoretical model in this region, though one can imagine that the degrader itself filters out protons with larger angles. We approximate this region in THETA0 by allowing θ_M to saturate at $t/R_0 = 0.97$ as shown. This region is unimportant in the design of beam spreading devices.

Figure 7.4 is a linear plot of the deviation from Molière/Fano theory as a function of normalized target thickness for the HCL data [4]. The window on the right is an expanded view of the ‘very thick’ target region and should not be used in judging the accuracy of the theory. As just stated, in this region we have set θ_M to its value at $t/R_0 = 0.97$.

7.3 Gaussian Approximations

Figure 7.5 shows the Molière angular distribution plotted in the form $2\pi\theta_M^2 f(\theta')$ (equation 7.13) *vs.* $(\theta')^2$ in which representation a Gaussian appears as a straight line. The inset is an expanded view near the origin. The dashed line shows the result of using only the Gaussian term $f^{(0)}$. This plot illustrates two properties of the Molière distribution. First, it depends only weakly on target thickness: recall that the B range corresponds to a factor $\approx 1000\times$ in target thickness (Table 7.2.5). Second, $f(\theta)$ for small θ (on the order of the characteristic angle θ_M) is nearly Gaussian, but *not* the Gaussian obtained by merely dropping $f^{(1)}$ and $f^{(2)}$.

7.3.1 Hanson’s Formula

Hanson *et al.* [35] first pointed out that $f(\theta)$ is best approximated by a somewhat narrower Gaussian, one obtained by replacing $\chi_c\sqrt{B}$ by $\chi_c\sqrt{B-1.2}$. We define

$$\theta_H \equiv \frac{1}{\sqrt{2}} (\chi_c \sqrt{B-1.2}) \quad (7.25)$$

The *Hanson approximation* consists of going through the Molière calculation, computing θ_H from equation 7.25 and using it as the characteristic width of a Gaussian distribution, that is, replacing θ_0 in (7.1) by θ_H .

7.3.2 Highland’s Formula

In the Hanson approximation $f(\theta)$ is far simpler but we still have to go through the complete Molière calculation to find the characteristic angle. Highland [36] developed a formula which is vastly easier to compute but quite accurate:

$$\theta_0 = \frac{14.1 \text{ MeV}}{pv} z \sqrt{\frac{L}{L_R}} \left[1 + \frac{1}{9} \log_{10} \left(\frac{L}{L_R} \right) \right] \quad \text{rad} \quad (7.26)$$

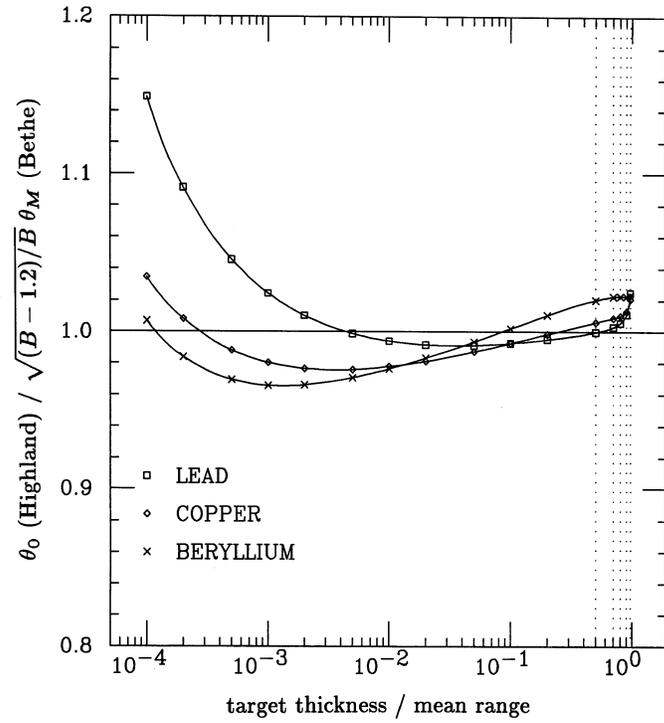


Figure 7.6: Test of Highland's formula as a proxy for Molière theory.

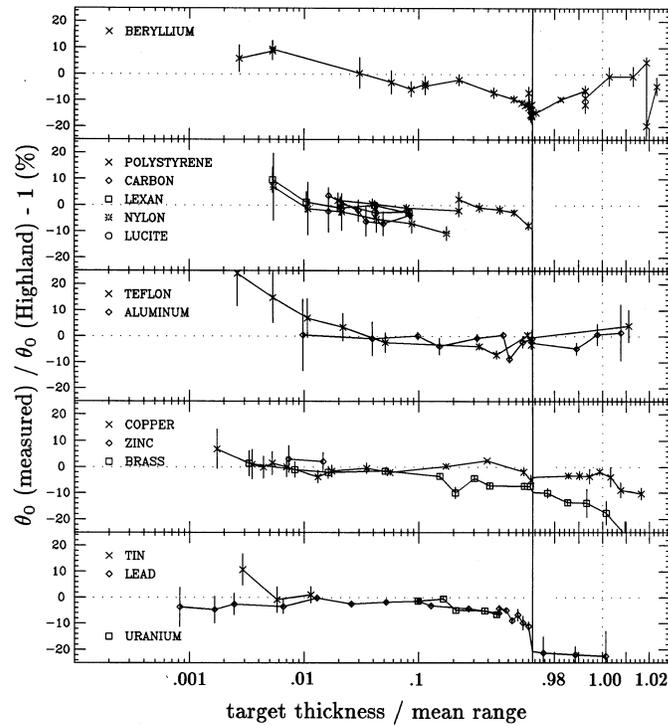


Figure 7.7: Test of Highland's formula (7.27) against experimental data.

L_R is the *radiation length* of the target material and L is the target thickness in the same units. You can look up L_R for many materials in [11]. Highland obtained this formula by fitting the Bethe variant of Molière theory ($Z(Z+1)$, no Fano correction) over a range of target thicknesses and materials. Figure 7.6 shows that his parametrization is extremely effective. Except for the lowest points for Pb, which fall outside the range of validity claimed by Highland, this simple formula has a *peak* error of $\pm 4\%$ and a typical error perhaps half that. It is definitely the formula of choice for pocket calculator work.

A brief discussion of formulas of the general type of (7.26) is in order because they occur widely in the early literature. The part preceding the square brackets is the ‘Rossi’ formula [37], except that Rossi’s constant was 21 MeV instead of 14.1. Most of this difference is in the usual $\sqrt{2}$: putting that in, Rossi’s constant is 15.0 MeV. The Rossi formula was widely used for rough estimates. It is typically 15% high, not good enough for design work. Over the years a number of *ad hoc* material- and thickness-dependent corrections to the Rossi formula were published. They have all been superseded by Highland’s elegant result. Highland’s formula makes explicit the point we raised earlier: because the Central Limit Theorem does not strictly apply to multiple scattering, the characteristic scattering angle cannot be exactly proportional to \sqrt{L} .

The kinematic factor in (7.26) raises the usual problem of thick targets, which Highland [36] did not address. We have generalized (7.26) to thick targets by regarding it as the formula for an infinitesimal slab, then integrating over the target, adding contributions in quadrature. We obtain

$$\theta_0 = 14.1 z \left[1 + \frac{1}{9} \log_{10} \left(\frac{L}{L_R} \right) \right] \times \left[\int_0^L \left(\frac{1}{pv} \right)^2 \frac{dt'}{L_R} \right]^{1/2} \quad \text{rad} \quad (7.27)$$

The integral should be evaluated in similar fashion to the Molière integrals discussed above. Use Simpson’s Rule to convert it to a finite sum and divide the target by equal ratios.

We have taken the correction factor out of the integral thus regarding it as applying to the *entire* target thickness. Without this step the integral would not exist since the integrand would become ever smaller with dt' . It would appear that we could have avoided this problem by using a sum in the first place and never attempting to write down an integral. But then the answer would change discontinuously as we changed the number of slabs. Thus we would have traded one arbitrary step (fixing the number of terms in a sum) for another (taking the logarithmic correction out of the integral). This is another example of the impossibility of obtaining the results of any *exact* multiple scattering theory simply by adding contributions in quadrature.

As an alternative to integration for moderately thick targets you can apply (7.26) just once, with pv replaced by the geometric mean of initial and final quantities, $\sqrt{p_1 v_1 p_2 v_2}$. The error in this shortcut, compared to integration, is negligible for $L/R_0 < 0.2$ and only $\approx 3\%$ for $L/R_0 \approx 0.7$.

We have tested (7.27) against experiment [4] and it is only slightly less accurate than Molière theory. The deviation from experiment for 115 points is normally distributed with a mean of $-2.6 \pm 0.4\%$ and an *rms* spread of 6%. In other words, Highland’s formula is $\approx 2.6\%$ higher than experiment, averaged over all HCL experimental points having $L/R_0 \leq 0.97$.

7.3.3 Lynch and Dahl's Formula

Lynch and Dahl [38] have revised the constants in Highland's formula thus:⁵

$$\theta_L = \frac{13.6 \text{ MeV}}{pv} z \sqrt{\frac{L}{L_R}} \left[1 + 0.088 \log_{10} \left(\frac{L}{L_R} \right) \right] \quad \text{rad} \quad (7.28)$$

Because this has the same form as Highland's equation our method of treating thick scatterers carries over directly. As with Highland's equation, θ_L is to be used as the characteristic angle of a Gaussian distribution, Eq.7.1.

Unlike Highland, Lynch and Dahl did not fit Molière theory directly. Their constants were obtained [38] by a maximum likelihood fit to the central 98% of scatters generated by a Monte Carlo program which used Molière theory. We compare the Lynch and Dahl formula with experiment in [4]. It is about 2.1% higher than experiment, slightly better than Highland's formula.

7.4 Fortran Function Theta0

Theta0 incorporates all the formulas and procedures of this Chapter. Before calling it you must initialize the range-energy functions:

```
i = InitRange('BGWARE\DATA\ICRU49.RET')
      . . .
mode = 'MOLIERE '
thetaM = Theta0(mode,tgm,t2,matl,bb)
```

Input parameters are MODE which is a code (8 characters) for the scattering theory; TGM, the target g/cm²; T2, the *outgoing* kinetic energy (MeV) and MATL which is a code for the target material (20 characters). Output, in addition to the function value, is BB, Molière's B . There is also a COMMON block

```
COMMON/Extras/nn,alfasq,chi0sq,chiasq,chicsq,b
```

which allows the calling program to access the Molière parameters. It need not be used otherwise. NN is the number of target slabs used for integration; the others are obvious. Theta0 has extensive comments: see these for details.

In addition to computing the characteristic angle for Molière theory and its variants and approximations, THETA0.FOR also contains a subprogram for computing the multiple scattering angular distribution. This routine, BetheF, uses Bethe's tables [5]. To see how it is used consult InitFscat in CSI.FOR. Note that Fscat is initialized by MAIN only once, using typical target parameters *viz.* LEAD, 1 g/cm², 160 MeV. These are only used to compute B , on which $f(\theta)$ depends very weakly (Figure 7.5).

7.5 Maximum Scattering from a Given Material

As an example of the use of Theta0, LOOKUP was used to create data for Figure 7.8 which illustrates an interesting fact about multiple scattering. Because multiple scattering depends on $1/T$ it might be thought that one could get a larger θ_0 (in whatever approximation) for a given material by lowering the incident energy. Not so. Figure 7.8 shows that the *maximum* scattering angle for a given material is very nearly independent of incident energy. Each case was run with TGM equal to the range at that incident energy and T2 = 0 which is a legal value and causes Theta0 to return the maximum angle. The maximum

⁵ The constant 0.038 in [11] appears to be a typographic error.

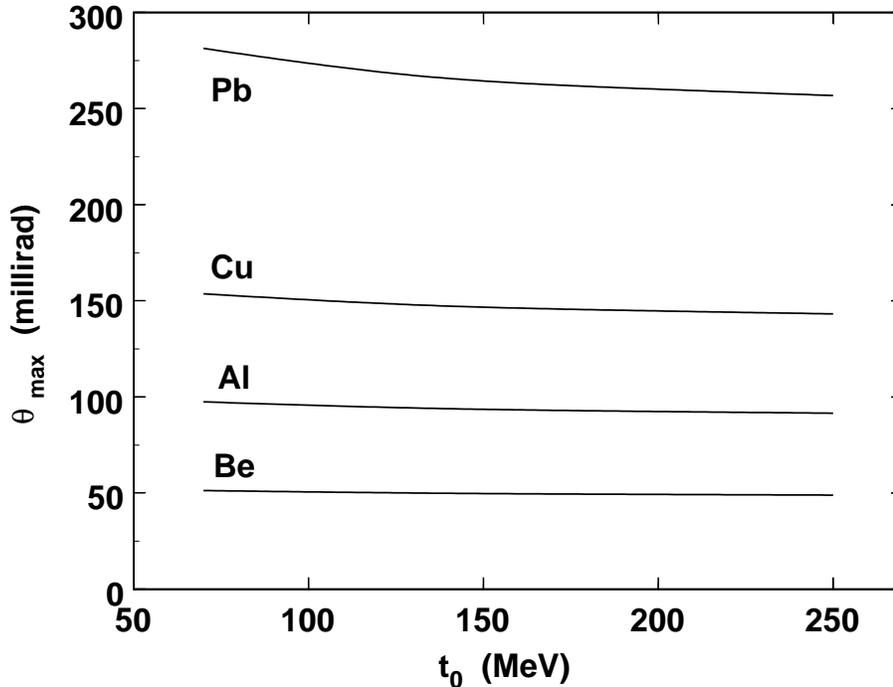


Figure 7.8: Maximum θ_M over the clinical energy range for four materials.

scattering angle for lead changes by only 10% over the clinical energy range. The other materials change less. To put it another way, Figure 7.2 would look nearly the same at any incident energy.

7.6 Adding Discrete Scatterers

Molière theory and the approximations to it provide an unambiguous answer for single scatterers, elemental or otherwise. One could even work out the answer for a finely divided mixture of lead and plastic. *Not* covered is (say) a plastic degrader followed by one of lead and so forth: a *stack* of scatterers. However, this case is very common in design work. We recommend computing the characteristic scattering angle (θ_{M_i} for example) separately for each scatterer, taking energy loss into account of course. Then add the results in quadrature:

$$\theta_M = \sqrt{\theta_{M_1}^2 + \cdots + \theta_{M_N}^2} \quad (7.29)$$

This is the procedure used, for instance, by function `TOUT` of module `TOUT.FOR`. Something of a consistency check can be obtained by computing, say, 2 g/cm^2 of Pb compared with $2 \times 1 \text{ g/cm}^2$. When we do this, using task `STACK` of `LOOKUP`, we find the result doing it the second (wrong) way is 3.1% low for θ_M and 3.5% low for θ_0 . This is measurable, but would require a rather careful experiment.

We justify using (7.29) by the fact that it seems to yield adequate designs and by the lack of a simple alternative. It is directly opposed to the recommendation of Lynch and Dahl [38] and the authors of [11]. While we agree that our method gives slightly too low an answer, the procedure detailed in [38] seems arbitrary, the improvement is not demonstrated, and it would be exceedingly difficult to incorporate in a general design program.

7.7 Summary

We have presented the formulas for the Molière/Fano calculation of the characteristic multiple scattering angle θ_M , the reduced target thickness B and, with the aid of Bethe's tables, the angular distribution $f(\theta)$. θ_M must always be used with the Molière angular distribution. It is incorrect to use it with a Gaussian. In design programs the Gaussian approximation carries little advantage since, for efficiency, one will tabulate $f(\theta)$ once and for all during initialization.

Nevertheless, it is almost certain that an equivalent Gaussian is sufficiently accurate since $\approx 98\%$ of the protons are contained in the nearly Gaussian part of the distribution. To compute the width parameter of the equivalent Gaussian one may use Hanson's prescription (essentially the full Molière calculation), Highland's formula, or Lynch and Dahl's formula. The last two are easy to compute and only slightly less accurate than Molière/Fano.

Our programs usually invite the user either to run in **MOLIERE** mode (full Molière/Fano formalism) or in **HANSON**, **HIGHLAND** or **LYNCH** mode. For any choice, you will get flat depth- and transverse dose distributions, assuming the problem posed is solvable with physical materials. That is merely a test of the mathematical self-consistency of the program: it should work even if the underlying scattering theory is seriously wrong! The only way to find out which option is better is to build the system both ways—the prescriptions for lead and plastic will be slightly different—and see which comes closer to the prediction. Since differences in the characteristic angle are $\approx 2\%$ that experiment will have to be done rather carefully. We generally prefer **MOLIERE** mode because we found the exact theory was, after all, in slightly better agreement with experimental data [4].

Chapter 8

Nuclear Interactions

In addition to electromagnetic (EM or Coulomb) interactions (stopping and multiple scattering) protons suffer nuclear interactions. These are not rare. At therapy energies, some 20% of the protons suffer a nuclear reaction before stopping. Even so, we appear to ignore nuclear processes when designing beam-spreading systems. Effectively, we include them by the simple expedient of using measured rather than theoretical Bragg peaks for the depth-dose. The design problem would be far harder if we had to evaluate nuclear effects explicitly.

That said, the rest of this chapter can be skipped on a first reading, since nothing said here will affect how we actually design devices. However, some understanding of nuclear interactions is essential to a complete picture of the Bragg peak: how they affect its shape, and how their effects can be observed directly. Key to all this is the ‘local energy deposition’ model. Before getting to that we’ll introduce some terminology and look at the kinds of reactions that go on, the secondary particles they produce, and their characteristic energies. Then we’ll introduce the model and discuss how the functions that define it, P and f , may be computed or measured. At our present state of knowledge, the model does not seem to work as well as it should, an issue that will be taken up again in chapter 11.

However, as just stated, that only means our understanding of nuclear processes is incomplete. It does not compromise our ability to design devices.

8.1 Terminology

ICRU63 [39] defines an *elastic* nuclear reaction as

... a reaction in which the incident projectile scatters off the target nucleus, with the total kinetic energy being conserved (the internal state of the target nucleus and of the projectile are unchanged by the reaction).

while *nonelastic*

... is a general term referring to nuclear interactions that are not elastic (*i.e.* kinetic energy is not conserved). For instance, the target nucleus may undergo breakup, it may be excited into a higher quantum state, or a particle transfer reaction may occur.

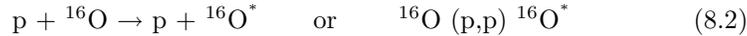
and *inelastic*

... refers to a specific type of nonelastic reaction in which the kinetic energy is not conserved, but the final nucleus is the same as the bombarded nucleus.

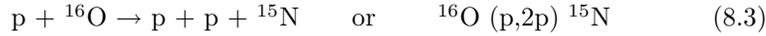
Thus



is elastic (${}^{16}\text{O}$ is in its ground state),



is inelastic (* denotes an excited state), and

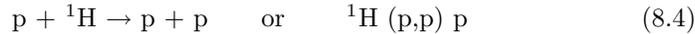


(quasi-free p,p scattering [40] in oxygen) is nonelastic even if the N nucleus recoils in its ground state, because it took energy ('binding energy') to remove the target proton from the nucleus.

Consider a proton beam passing through various absorbers and scatterers and stopping in a water tank. Some of the protons undergo nuclear interactions on the way, either in the absorbers and scatterers or in the water. At any given point we call those particles *primaries* which have only suffered EM interactions: stopping by interactions with atomic electrons and scattering by screened atomic nuclei. Particles resulting from nuclear reactions are *secondaries*. Both final state protons from reactions like (8.3) are secondaries even though one (we cannot tell which) was the incident proton. All primaries are protons, of course, and all neutrons must be secondaries.

8.2 Elastic Scattering from Free Hydrogen

Water (H_2O) and many plastics of interest (*e.g.* CH_2) contain free hydrogen. Janni [3] and Berger [27] correctly note that the nonelastic cross section for hydrogen is negligible. The proton has no internal degrees of freedom at therapy energies. Therefore interactions with H nuclei are traditionally omitted from nuclear effects. We think this is inappropriate. Consider



which is simply elastic scattering by a very light nucleus, a classic billiard-ball collision. The final protons emerge at right angles to each other¹ and share the incident kinetic energy. The more 'forward' proton carries more of the energy and the other, less. These protons greatly resemble those from reactions like (8.3) though they are slightly more energetic because there was no nuclear binding energy. If we are interested in the changing composition of the beam, rather than the classification of reactions, we ought to include (8.4).

Having made this point we will not belabor it. From here on we'll take the conventional position that only nonelastic reactions diminish the primary fluence.

The bremsstrahlung reaction



does occur at therapy energies. Though measurable [41], its cross section is tiny by comparison with (8.4).

¹ Actually somewhat less because of relativistic effects: 87.5° at 160 MeV incident.

8.3 Overview of Nonelastic Reactions

Possible secondaries from nonelastic reactions at therapy energies are protons, neutrons, heavy fragments such as alphas, γ rays, and the recoiling residual nucleus. Heavy fragments other than alphas are in fact quite rare. For 160 MeV protons stopping in CH_2 [25], per incident proton

^1H (p)	^2H (d)	^3H (t)	^3He	^4He (α)
0.24	0.09	0.01	0.03	0.13

Per nonelastic nuclear reaction induced by primaries or secondaries

p	d	t	^3He	α	n
1.16	0.46	0.06	0.13	0.71	0.41

By either reckoning the number of alphas is striking. Alphas are very stable and exist as preformed clusters in light nuclei [42]. The numbers quoted are from the precompound model in the GEANT4 Monte Carlo, not from direct experiment. The precompound model is ultimately, however, a model of experimental results.

An alternative measure of the relative importance of secondaries is the *fraction of initial energy* carried away by each. For nonelastic interactions of 150 MeV protons with ^{16}O nuclei Seltzer [43] finds

p	d	t	^3He	α	recoils	n
0.57	0.016	0.002	0.002	0.029	0.016	0.20

The total imparted to charged particles is 0.64 . Photons, not listed, presumably make off with 0.16 . These numbers are also from a Monte Carlo, not directly from experiment. The general picture is the same. Most of the final kinetic energy is in protons, neutrons and photons. Alphas stand out among the heavy fragments, but still have only 2.9% of the total energy.

In elastic scattering from free hydrogen, *all* of the secondary energy is carried by protons, strengthening our argument for including them.

Though little of the energy goes into heavy fragments, they might have a significant relative biological effect (RBE) because of their high ionization density. This has been investigated by Seltzer [43], Paganetti [44, 45] and others. It turns out that RBE enhancement due to heavy fragments and recoils is small. Their extremely high ionization density is outweighed by the fact that little energy goes into those channels. What RBE enhancement there is comes from the abundant low energy protons: secondaries and, at the distal end of the Bragg peak, primaries. In any case, the present work is concerned with *physical* dose, so we will speak no more of RBE.

Secondaries usually make large angles with the beam direction (think billiard balls) unlike primaries which, even after multiple scattering, rarely exceed a few degrees. This is important. It means that secondaries produced upstream in scatterers or absorbers will clear out of the beam for purely geometric reasons before entering the volume of interest.

8.4 Local Energy Deposition Model

The *local energy deposition model* seeks to model the effects of nuclear reactions on the depth-dose without using the full-bore Monte Carlo approach (generating and tracking all secondaries). It can be summarized as follows:

- Primary fluence is converted into secondary fluence.
- Energy of primaries is converted into energy of secondaries.
- Energy of charged secondaries is deposited on the spot.
- Energy of neutral secondaries is lost.

The energy of neutral secondaries is not destroyed, of course. It is deposited in the shielding. Also, energy of charged secondaries is not *really* deposited on the spot: some secondaries have projected ranges of a few cm. Nevertheless, this approximation works in many situations because of longitudinal equilibrium, which will be discussed shortly.

Mathematically, the local energy deposition model requires two functions, P and f . Imagine protons stopping in a water tank. $P(z; r_0)$ is defined as the probability that a proton of initial range r_0 will have a nonelastic nuclear interaction by the time it reaches depth z in the tank. Alternatively we can speak of $dP/dz \equiv P'(z; r_0)$. $P'(z; r_0) dz$ is the probability of an interaction in dz about z . We'll find that P and P' are fairly easy to predict or measure.

$f(z; r_0)$ is the fraction of primary energy deposited locally. We have seen above that f , averaged over all z , is ≈ 0.64 at 150 MeV ($r_0 = 15.8 \text{ g/cm}^2$). f is difficult to measure. Microdosimetry sheds some light, or the detailed shape of the Bragg peak can be compared with the local energy deposition model to infer f [27]. However, we mainly have to rely on Monte Carlo estimates.

Instead of picking r_0 and z as independent variables, we could have used t_0 and t , the incident and residual kinetic energy as the proton traverses the tank. Either way works. Using depth and range instead of energy gives a more immediate connection with the Bragg peak.

8.5 Nuclear Buildup

A brief experimental interlude. During the commissioning of NPTC some depth-doses were taken with a vertical beam, that is, a wall-less water tank [46]. Very few nuclear secondaries enter the water because the density of air is $\approx 1000\times$ less than water. Secondaries produced in large volumes far upstream clear out of the beam because of their large angles, and there is not enough air mass near the tank to create many secondaries that hit the region of interest.

The dose measurement (Figure 8.1) shows a small but definite defect at the entrance², which we interpret as follows. At $z = 0$, nuclear reactions happen at the full rate, but their energy is deposited a bit downstream so the stopping power here is just S_{EM} . As we move downstream, charged secondaries begin to range out and the stopping power acquires an additional term $S_{nuclear}$. Soon (certainly by 4 g/cm^2) we reach *longitudinal equilibrium* where there are as many charged secondaries entering any Δz as leaving, because the number and composition of secondaries changes very slowly with depth (energy). 'Local energy deposition' does not mean that the energy is literally deposited on the spot where the secondaries were produced. It comes from volume elements further upstream, but those secondaries are almost indistinguishable from those produced on the spot.

Of course, longitudinal equilibrium cannot hold at a discontinuity in density such as the tank entrance. That's where we see buildup effects.

² This is not observed in horizontal water tank scans because the usual tank wall introduces enough nuclear buildup to obscure it. It would, however, be easy to build a sufficiently thin-walled tank to show the buildup in horizontal scans.

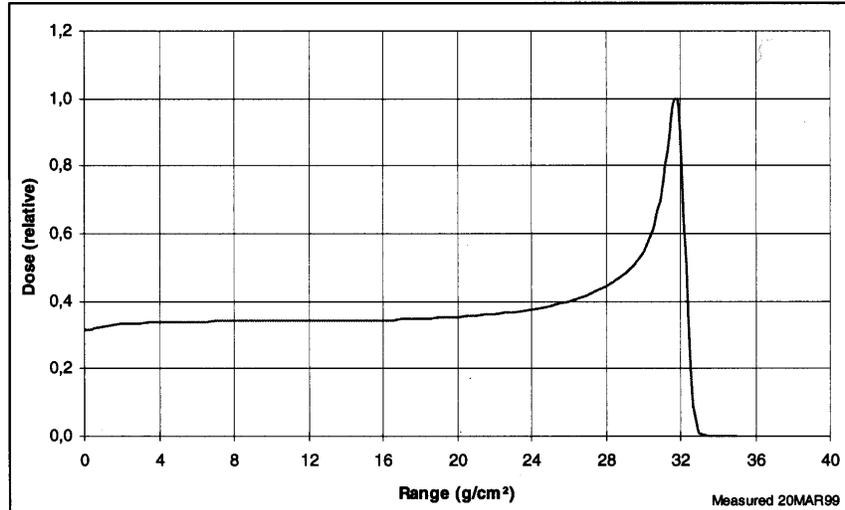


Figure 8.1: Nuclear buildup in a Bragg peak.

Though this model works qualitatively, it doesn't quite hold up numerically. Let's compare energy absorption in Δz about $z = 0$ and $z = 4$ g/cm² where longitudinal equilibrium is fully established. The primary range, energy, EM stopping power, P and P' at the two positions are

z g/cm ²	r g/cm ²	t MeV	S_{EM} MeV/(g/cm ²)	P %	P' %/(g/cm ²)
0	32	226	4.15	26.7	0.571
4	28	209	4.36	24.4	0.600

P and P' come from the next section. Also, we know that $f \approx 0.64$. Therefore at $z = 4$, accounting for the decrease in primary fluence from $z = 0$, we have

$$\Delta E = 4.36 \times (1 - 0.023) \times \Delta z + 0.006 \times 0.64 \times 209 \times \Delta z$$

$$\Delta E/\Delta z = 4.26 + 0.802 = 5.06 \text{ MeV}/(\text{g}/\text{cm}^2)$$

whereas at $z = 0$

$$\Delta E/\Delta z = S_{EM}(226) = 4.15 \text{ MeV}/(\text{g}/\text{cm}^2)$$

giving a defect of $4.15/5.06$ or $\approx 18\%$. The observed defect (Figure 8.1) is only $\approx 8\%$.

The measurement of nuclear buildup, and that it is smaller than expected, dates back at least to 1977 when Carlsson and Carlsson [47] measured dose buildup with 185 MeV protons incident on a stack of $(\text{C}_2\text{F}_4)_n$ sheets interleaved with thermoluminescent detectors. They observed a nuclear buildup of $\approx 7\%$ and characteristic length ≈ 2 g/cm². They commented that it was 'somewhat lower' than expected, and cited dosimeter nonlinearity as a possible cause. Their observations in $(\text{C}_2\text{F}_4)_n$ agree well with ours in water.

Incidentally, Carlsson and Carlsson also measured the electron buildup which is greater but has a much shorter characteristic distance ($\ll 1$ mm). Electron buildup with incident protons can also be observed using very thin-walled ionization chambers. In contrast with photon therapy, it does not yield useful skin sparing in clinical practice because the buildup length is so short.

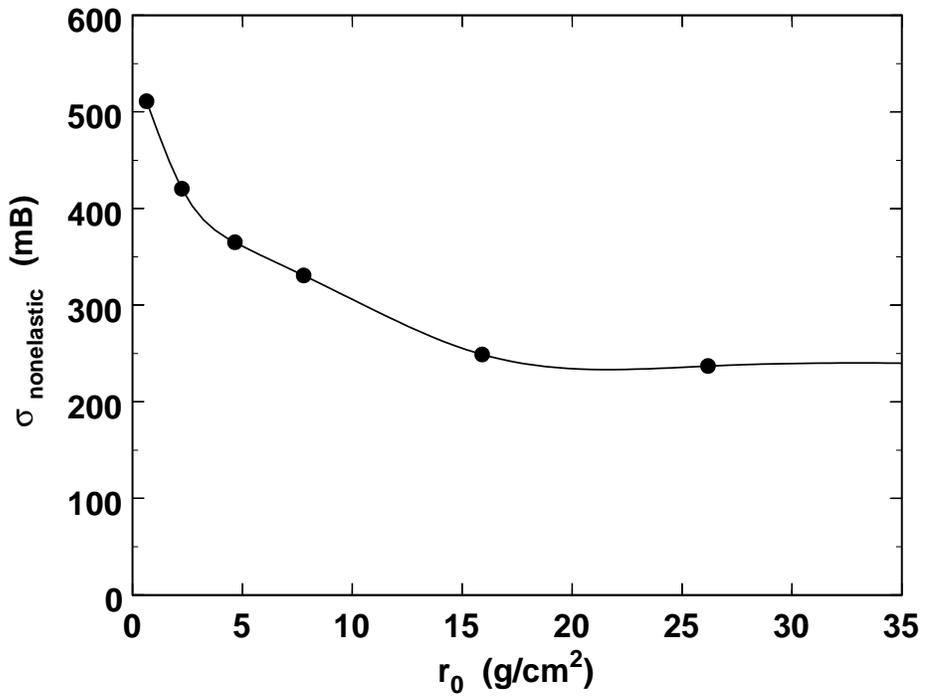


Figure 8.2: Total nonelastic cross section of oxygen versus proton range.

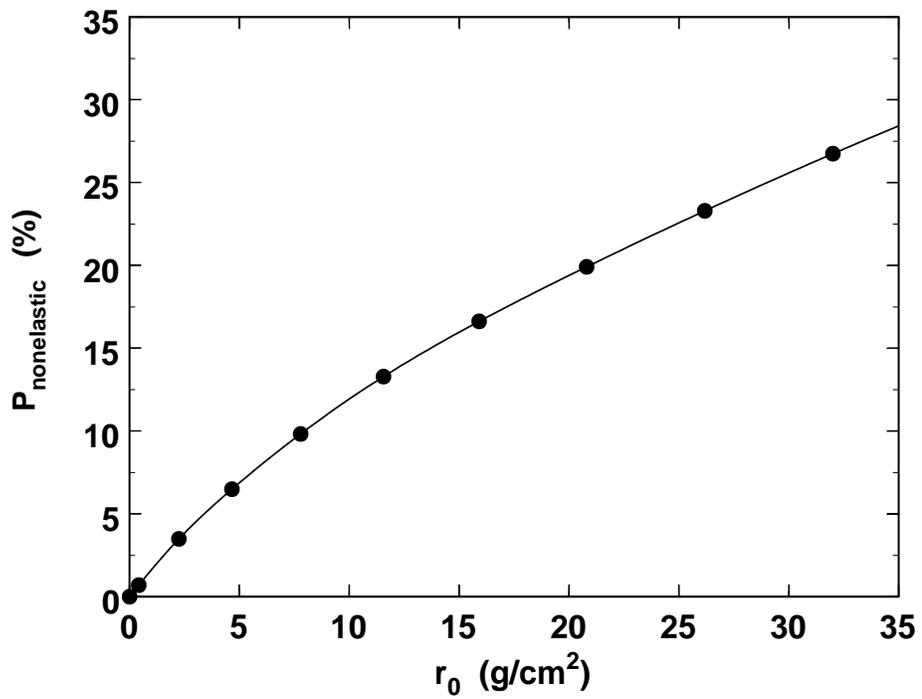


Figure 8.3: Probability of a nonelastic interaction versus proton range in water.

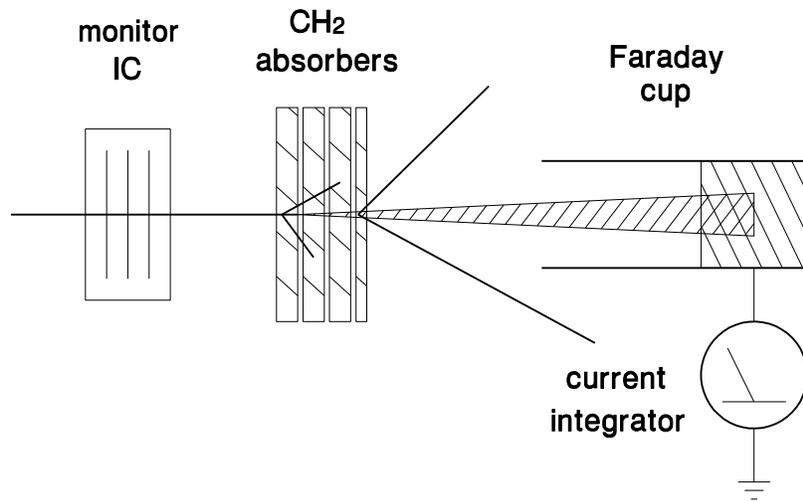


Figure 8.4: Experiment to measure reduction of primary fluence.

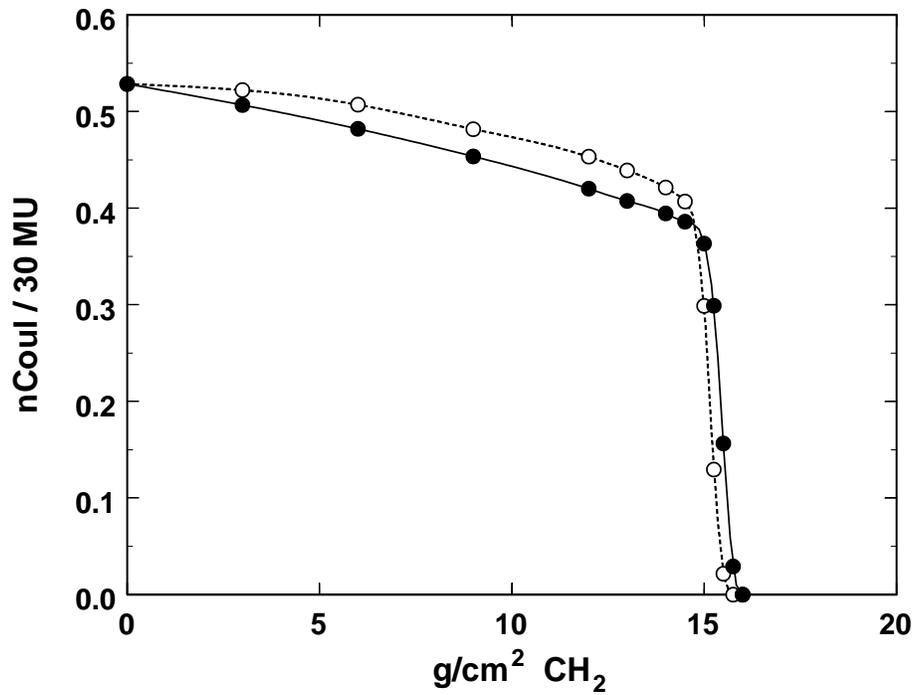


Figure 8.5: Reduction of primary fluence observed with Faraday cups.

8.6 Predicting P

To predict P we need the nonelastic cross section for O. Janni [3] mixes theory with experimental data to obtain a smooth and self-consistent cross section, shown in Figure 8.2 as a function of proton range. Others [43, 27] use slightly different procedures and experimental data, but also find that the nonelastic cross section is nearly constant at therapy energies except in the last few cm where it roughly doubles before falling to zero.

The total nonelastic interaction probability for a stopping proton, $P(r_0; r_0)$, is the integral of the nonelastic cross section over the entire range. Figure 8.3 shows Janni's result for water [3]. For example, a 209 MeV proton, with a range of 28 g/cm², has a 24% probability of a nonelastic reaction before it stops. P' is the slope of this curve. At high energies (as the proton enters the water tank) P' is nearly constant because of the constancy of the cross section. Our 209 MeV proton has an initial P' of 0.60 %/(g/cm²). As the energy and residual range decrease and the cross section increases, P' increases. The value of 1.2 %/(g/cm²) adopted by Bortfeld [26] is an average over therapy energies.

8.7 Measuring P

8.7.1 Simple Faraday Cup

P and P' can be measured using a Faraday Cup (FC) in appropriate geometry (Figure 8.4). The size and position of the FC should be such that primary protons are counted even when multiple scattering in the absorber stack is taken into account, but almost all secondaries either range out in the absorber or have angles too large to enter the FC.

A trial at HCL used 159 MeV protons incident on CH₂ absorbers with two different FCs and geometries. The first was a classical vacuum FC [48]. Its stopping block was at some distance from the absorber stack as shown in Figure 8.4. The results are shown as solid circles in Figure 8.5. The initial slope is $P' = 1.4$ %/(g/cm²) and the total decline to the edge of the EM 'cliff' is $P = 27\%$.

The second FC was a non-vacuum 'Poor Man's Faraday Cup' consisting of a brass block wrapped in insulator and electrostatic shielding. This detector was larger, and was placed up against the CH₂ absorbers. The results are shown as empty circles. The large angular acceptance meant that this detector counted a significant fraction of the secondary fluence. Therefore it did not measure P' accurately. However, the total decline to the edge of the cliff is still a valid measurement of P because, for a thick absorber stack, almost all the secondaries range out. The result is $P = 23\%$.

We have described this experiment merely to illustrate a simple technique that should work. Properly executed, it should allow one to measure P' and P in any material. The present results, particularly $P' = 1.4\%$ for CH₂, seem larger than expected even when the contribution of free hydrogen is taken into account.

8.7.2 Multi Layer Faraday Cup

P can also be measured with a Multi Layer Faraday Cup. MLFCs come in two varieties. Conducting plates, each connected to a current integrator, can be separated by thin insulating sheets. Alternatively, insulating plates can be separated by thin conducting sheets, each connected to a current integrator. In

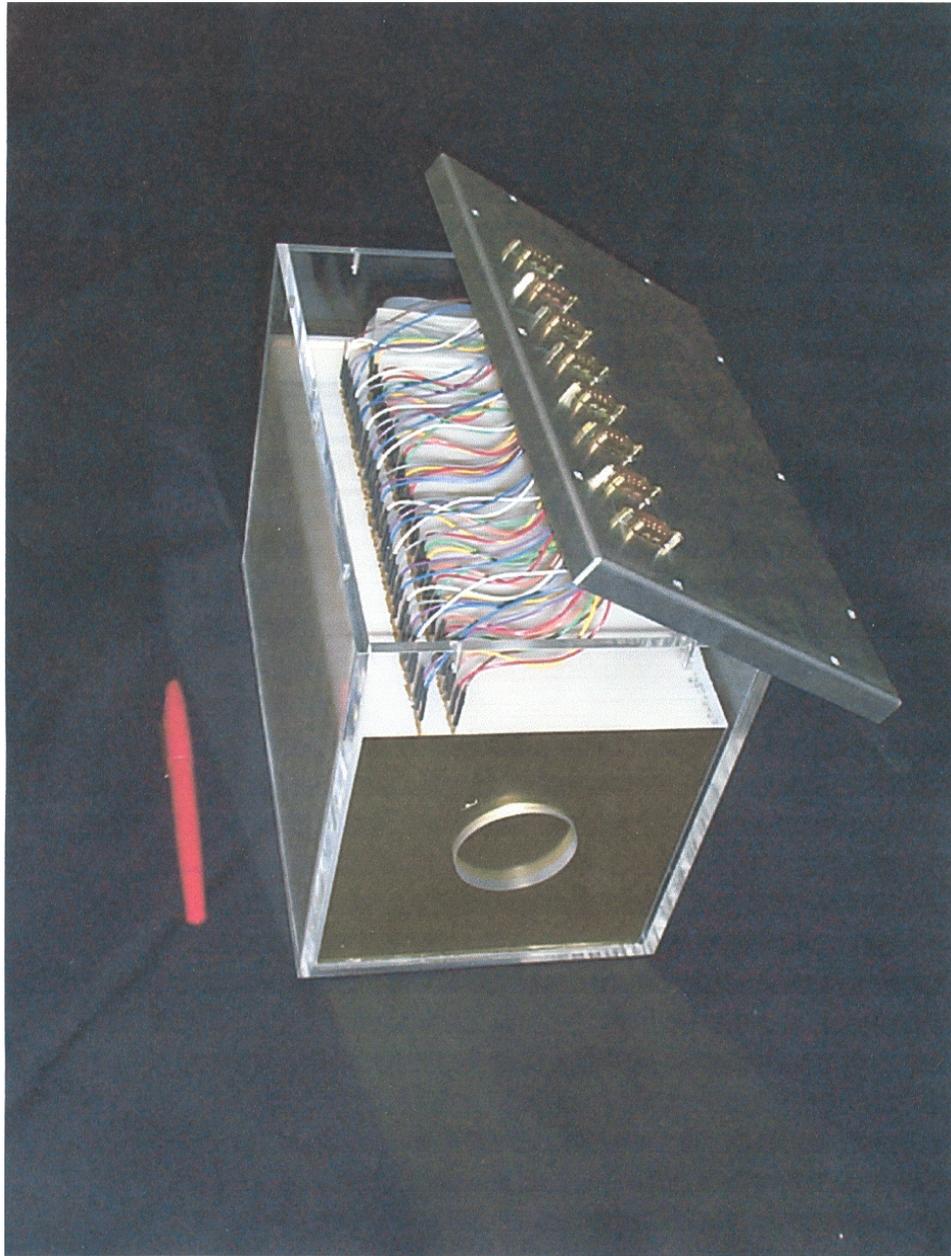


Figure 8.6: A multi-layer Faraday Cup with CH_2 plates (aluminum shield not shown).

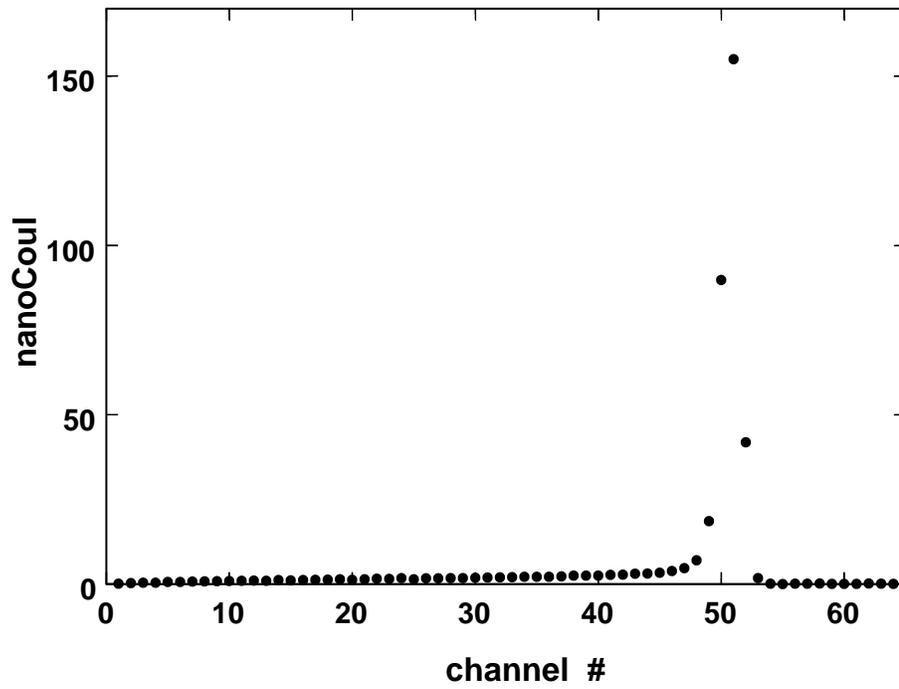


Figure 8.7: Differential fluence observed in the CH₂ MLFC.

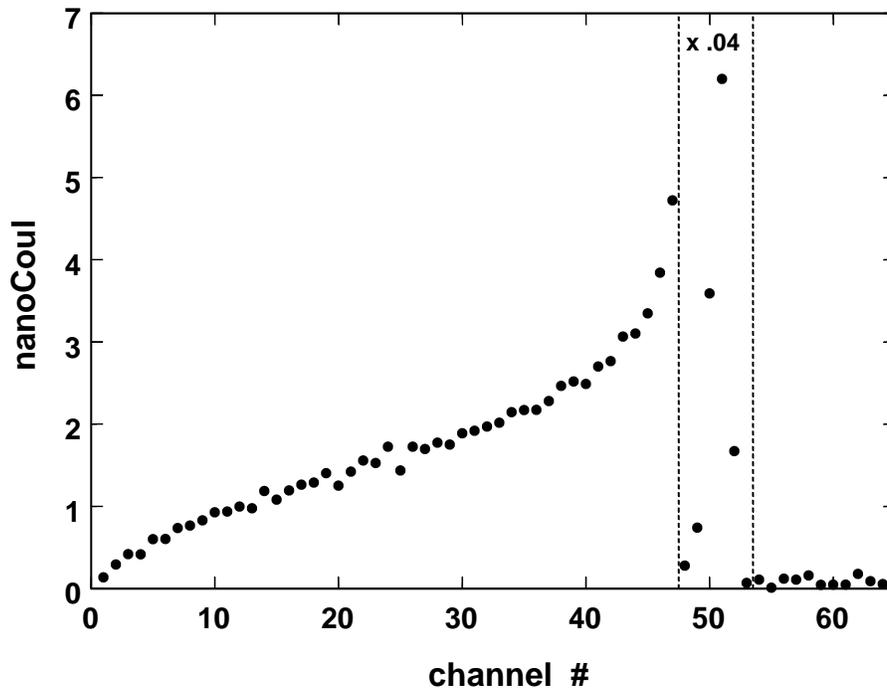


Figure 8.8: Nuclear interactions observed in the CH₂ MLFC.

the latter case it is the charge *induced* on the conducting sheets that is detected. Both types have been built and operated at HCL.

We show data from an ‘insulating’ MLFC of CH₂ sheets separated by thin brass charge collectors (Figure 8.6). Figure 8.7 shows the charge distribution when a proton beam stops. MLFCs measure *differential* fluence as a function of depth. A pronounced peak from primary protons stopping by EM processes corresponds to the ‘cliff’ in Figure 8.5. It is preceded by a barely visible buildup region. If we zoom in (Figure 8.7) we see a well-defined curve, the projected range distribution of charged secondaries. Its shape is a test of Monte Carlo models [24, 25, 49].

Here, however, we’ll focus on its area. The net charge deposited for each reaction (including the recoil) is $+e$ (the charge of the incoming proton) and all secondaries have shorter projected ranges than primaries. Therefore

$$P = \frac{\text{buildup area}}{\text{buildup area} + \text{EM peak area}}$$

We find $P = 20 \pm 1\%$. The error comes from estimating where the buildup ends and the EM peak begins. Janni [3] gives $P = 18 \pm 4\%$.

8.8 Summary

ICRU63 categorizes nuclear reactions as elastic, nonelastic and inelastic. It is common practice [3, 27, 43, 26] to consider only nonelastic nuclear reactions in computing the reduction of primary fluence. We have argued that elastic scattering from free hydrogen has a very similar effect and should be included.

We looked at the breakdown of secondaries and the energies they carry. Most of the secondaries are protons, neutrons and recoiling nuclei. Charged secondaries share typically 60% of the initial energy and have projected ranges much shorter than primaries. Their energy can be considered to be deposited on the spot. The remaining 40% of energy is carried by neutrals and is lost to the region of interest. This picture breaks down at discontinuities, for instance when the beam leaves air and enters a water tank. Here nuclear buildup can be observed and longitudinal equilibrium takes a couple of g/cm² to establish itself. There are at least two measurements, one of them published [47], that show nuclear buildup. In both cases the buildup distance is reasonable but the magnitude is about half that expected from the nuclear model.

The local energy deposition model is described by two functions. $P(z; r_0)$ is the probability that a primary of range r_0 will have a nuclear interaction before it gets to z . $f(z; r_0)$ is the fraction of primary energy in charged secondaries, considered to be deposited on the spot. P in water can be computed from the interaction cross section for oxygen (and, we would suggest, hydrogen). f must be found by Monte Carlo methods. P and f have been tabulated by Berger [27] and Seltzer [43].

P is fairly easy to measure. All one needs is a fluence meter that somehow separates primaries from secondaries. Either a Faraday Cup with appropriate angular acceptance, or a multi-layer Faraday Cup, can be used. f is difficult to measure, except perhaps by microdosimetry.

We have spent an entire chapter on nuclear reactions mainly to round out our understanding of how protons interact with matter. In designing passive beam spreading systems we will constantly do computations with stopping theory and scattering theory. These theories of EM processes are accurate and well tested. By comparison, nuclear theory is exceedingly complicated and not as well tested. What saves us is that we do not need to use it explicitly. Using experimental

Bragg peaks in our designs seems to be a good enough way of including nuclear effects. On the rare occasion where we might want to estimate nuclear effects, the local energy deposition model should be a pretty good proxy for the full model. Of course we can always resort to a full Monte Carlo simulation [50].

Chapter 9

Binary Degradors

In this chapter we learn how to design degraders with specified scattering¹ ('scatterers'), specified energy loss ('absorbers') or both. We'll assume we have at our disposal four functions

```
Range(energy, material)
Energy(range, material)
Dedx(energy, material)
Theta0(mode, target thickness, outgoing energy, material)
```

where `material` is a code *e.g.* 'WATER', `mode` *e.g.* 'HIGHLAND' specifies the scattering theory to be used, `range` and `target thickness` are in g/cm² and energies are kinetic energies in MeV. `Dedx` is the mass stopping power S in Mev/(g/cm²). Note that input to `theta0` is the *outgoing* energy; that way the problem is always well posed.

If you wish to follow the examples, we'll be using LOOKUP with the MIXED range-energy table (mostly ICRU49) in HIGHLAND mode when applicable.

9.1 High-Z and Low-Z Materials

Figure 9.1 shows scattering (top curve) and energy loss (bottom curve) of 160 MeV protons incident on 1 g/cm² degraders of various materials. The x coordinate is the material's radiation length L_R . We have used this for x first, because it figures in Highland's formula and second, because, unlike atomic number Z and mass A , L_R is defined [11] for compounds and mixtures.

We see that materials low in the periodic table, which we'll call low-Z, have somewhat more stopping power and a great deal less scattering power than high-Z materials. We can combine them to make a *binary degrader* with any desired energy loss and scattering angle, within limits.

It goes without saying that, when we want energy loss with low scattering, we should use a low-Z material. When we want scattering with low energy loss, we should use a high-Z material.

9.2 Notation to Describe Problems

Different problems can be posed for one or two degraders. It is convenient to have a compact notation. Let

¹ 'Scattering' should always be understood as 'characteristic multiple scattering angle'.

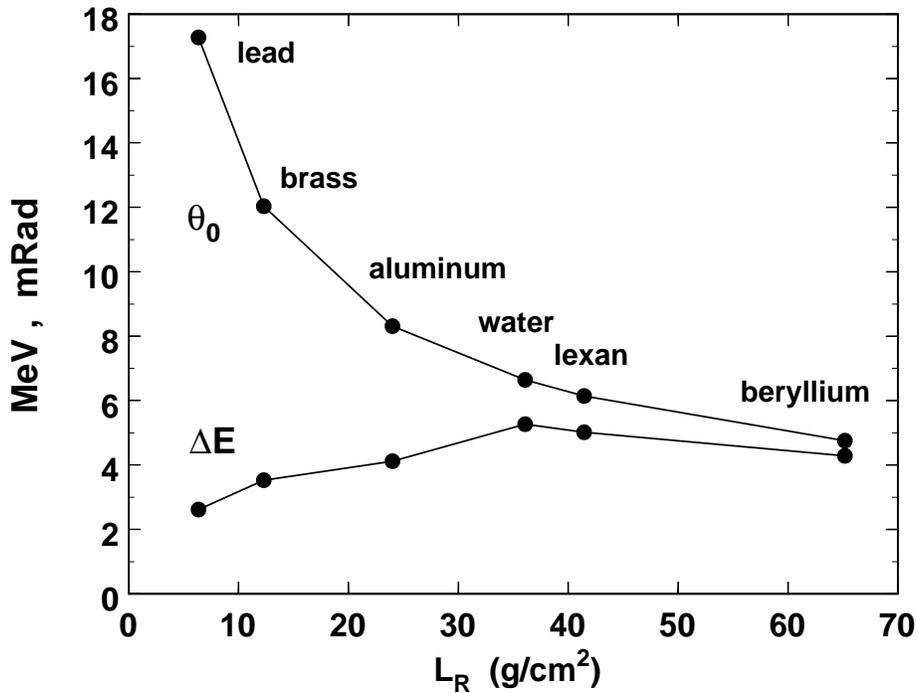


Figure 9.1: Scattering and stopping in various materials.

a denote scattering angle (rad)
 g thickness (g/cm²)
 m material
 t kinetic energy (MeV)
 1 'entering the first element'
 2 'entering the second element' ...

Then

$$a_1, a_2, t_2 \rightarrow g_1, t_1$$

is the single degrader problem 'Given the incoming and outgoing angles and the outgoing energy, find the thickness and the incoming energy.' while

$$a_1, a_3, t_1, t_3 \rightarrow g_1, g_2, t_2$$

is the binary degrader problem 'Given the incoming and outgoing angles and energies, find the required thicknesses and the energy going into 2.' Note that a) materials are *always* given and therefore not usually mentioned, b) the index 2 ('single' problem) or 3 ('binary' problem) represents 'the next element' whatever that is and c) the unknowns (or output variables if we're talking subroutines) may include things you don't happen to be interested in.

File `BDG.FOR` has subroutines for solving single degrader problems

1 : $a_1, a_2, t_2 \rightarrow g_1, t_1$
 2 : $a_1, a_2, t_1 \rightarrow g_1, t_2$

and binary degrader problems

- 1 : $a1, a3, t1, t3 \rightarrow g1, g2, t2$
- 2 : $a1, a3, g1, t3 \rightarrow g2, t1, t2$
- 3 : $a1, a3, g1, t1 \rightarrow g2, t2, t3$
- 4 : $a1, a3, g2, t1 \rightarrow g1, t2, t3$

all of which are needed at one time or another. We have also listed the jump index which tells BDG or SDG which problem is to be solved. LOOKUP will remind you of these if you enter a nonsense index *e.g.* 9. The rest of this chapter explains how these routines work.

9.3 ‘Forward’ Problem

The problems listed are ‘inverse’ problems: at least one thickness is not known. The ‘forward’ single degrader problem is: ‘Given degrader material, thickness, incoming energy and incoming scattering angle, find the outgoing energy and angle.’ that is

$$m1, g1, t1, a1 \rightarrow t2, a2$$

This is fairly straightforward:

1. Find the proton range in material $m1$ at energy $t1$.
2. Subtract $g1$ to obtain the residual range.
3. Find the energy in $m1$ for that range. That equals $t2$.
4. Use **Theta0** to find the scattering.
5. Add it to $a1$ in quadrature to find $a2$.

You can repeat this for as many degraders as you have. Each time, there is a range-energy calculation and an angle calculation, which must be done in that order. There’s a shortcut for the range-energy calculation if the degrader is thin:

1. Find the stopping power S in material $m1$ at energy $t1$.
2. Multiply by $g1$ to obtain the energy loss.
3. Subtract the energy loss from $t1$ to obtain $t2$.

Example: find the energy and angle out of 10 g/cm² of lexan assuming 160 MeV incident protons with no angular divergence. The two procedures are

$t1$	S	$r1$	$r2$	$t2$	θ_0
MeV	MeV/(g/cm ²)	g/cm ²	g/cm ²	MeV	mrad
160		18.523	13.523	133.66	16.09
160	4.962			135.19	15.94

The S method is slightly off because the degrader is fairly thick. S increases significantly through the degrader, but we are using the entrance value.

For general design problems it is very convenient to have a subroutine that applies the forward solution iteratively to an arbitrary stack of absorbers. Such a routine, TOUT.FOR, can be found in source file TOUT.FOR. One can also run the procedure backwards to obtain the input energy and angle given the output energy and angle. See TINP.FOR in the same file. The simple logic of both procedures is somewhat obscured by tests.

9.4 Single Degraded Inverse Problem

Consider

$$a1, a2, t1 \rightarrow g1, t2$$

namely ‘How thick a scatterer do I need? what will the output energy be?’

Before trying to solve an inverse problem we should test for conditions that may rule out a solution. In this case the input energy may be so low that the desired scattering cannot be had before the proton ranges out. We can find the maximum available scattering by computing the range in $m1$ at $t1$, then calling `Theta0` with that degrader thickness and zero output energy, which causes `Theta0` to return the maximum scattering available. If that is less than the desired $a2$ a sensible subroutine will return $a2$ reassigned to that value, assuming the user will test $a2$ if failure is at all likely. Otherwise, we can proceed.

We want $a2$ to equal some specified value. Let’s construct a function $a2 = F(g2)$. F is not an algebraic function of $g2$ in closed form. Even if we could, with some thought, write down a sufficiently good approximation, that’s not how we want to proceed. A numerical method is far more general and requires far less effort. F is simply a subprogram (‘function’ in the Fortran sense) which, given $g1$ and whatever else is needed, finds $a2$. We illustrate this below by listing the subprogram for this problem. The procedure is exactly that of the previous section for solving the forward problem. We have made the logic clearer by deleting some safety tests (*e.g.* for negative argument of the square root).

```
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
C
C      FUNCTION Asdg1(xx,m1,m2,g1,g2,t1,t2,t3,a1,a2,a3,mode)
C
C      Angle out of single degrader as a function of xx = g1.
C
C      CHARACTER*(*) m1,m2,mode
C
C      rr = Range(t1,m1) - xx
C      t2 = Energy(rr,m1)
C      Asdg1 = SQRT(a1**2 + Theta0(mode,xx,t2,m1,bb)**2)
C
C      END
C
C%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Now let $g2$ be called x and the desired value of $a2 = F(x)$ be called c . We need to find that value of x for which $F(x) = c$, that is, we need to solve

$$F(x) - c = 0$$

or find the root of $F(x) - c$. $F(x)$ is not an algebraic expression so we need a numerical root finder. These are discussed in Numerical Methods [1]. For most of our work the Method of False Position (*regula falsa*) offers the right combination of speed and robustness. Formally, we construct a Fortran function, call it `RootFalse`, such that the call

```
x = RootFalse(F,a,b,c)
```

returns x as the root of $F(x) - c = 0$. In addition to c , the desired value of F , we must supply a and b , values of x which bracket the root, thus instructing the routine to look for a solution between a and b . In the problem at hand the root finder is invoked with

`g1 = Rtf(Asdg1,0.,gg,a2,m1,m2,g1,g2,t1,t2,t3,a1,a2,a3,mode)`

`Asdg1` is our mnemonic for this particular F . `0` and `gg` (the proton range in material $m1$ at incident energy $t1$) bracket the $g1$ we seek. We know this from our preliminary test. $c = a2$ is the desired value of F . Many other arguments are also passed because we want just one routine `Rtf` for all our single and binary degrader problems. Therefore any quantity that might be needed as input or output is passed. In any given case, many are ignored.

Some closing remarks:

- The method always works if we can write a procedure for the forward problem. The programming task is to do this and make sure the necessary quantities are supplied.
- We haven't mentioned $t2$. It is computed as an intermediate step in `Asdg1`. Because of the way variables are passed, it will have assumed its correct value when `Rtf` returns without our having to lift a finger.
- The Method of False Position always stays within the initial brackets. Therefore routines like `Range` are never called with illegal arguments as `Rtf` does its work.
- In many problems there is more than one choice for the 'key' variable, and more than one will work. In our example we used $g1$ because its brackets are obvious: the answer must lie somewhere between 0 and the range of the proton.

The other single degrader problem is handled the same way. It's simpler because there is always a solution. See `BDG.FOR` for details.

9.5 Binary Degraded Inverse Problem

We'll use as an example

$$a1 \ a3, \ t1, \ t3 \ \rightarrow \ g1, \ g2, \ t2 \tag{9.1}$$

or 'Given some lead and lexan, the incoming energy and the desired outgoing energy and scattering, what thicknesses do I use?' The basic method is the same. Focus on one of the unknowns ($t2$ is convenient), write down the forward solution $a3(t2)$ in terms of it, then use a root finder to locate the value of $t2$ that yields the desired $a3$. However, this problem is somewhat complicated so we'll first study a concrete case. Suppose 160 MeV protons enter a lead/lexan degrader. We want to reduce the energy to 140 MeV while obtaining exactly 40 milliradians of scattering. What combination of lead and lexan do we need? Over what range of output energies can the problem be solved? What happens outside that range?

Figure 9.2 shows a family of solutions for 20, 30 and 40 mrad. Let's start at the lower right and work down the energy scale. Initially we want scattering with minimum energy loss so we use lead only. As we increase the lead, the output energy decreases and the scattering increases. Eventually the scattering reaches 40 mrad, the desired value. To add more lead would be to exceed this, so we add lexan, moving further down the energy scale. There is some scattering even in the lexan, so as we add lexan we remove lead to stay on the 40 mrad line. Eventually, all the lead is gone and the scattering in lexan is 40 mrad. To reduce the energy further we have to use more lexan and accept more scattering. Finally, the curve flattens at very low $t3$. Here the lexan thickness is greater

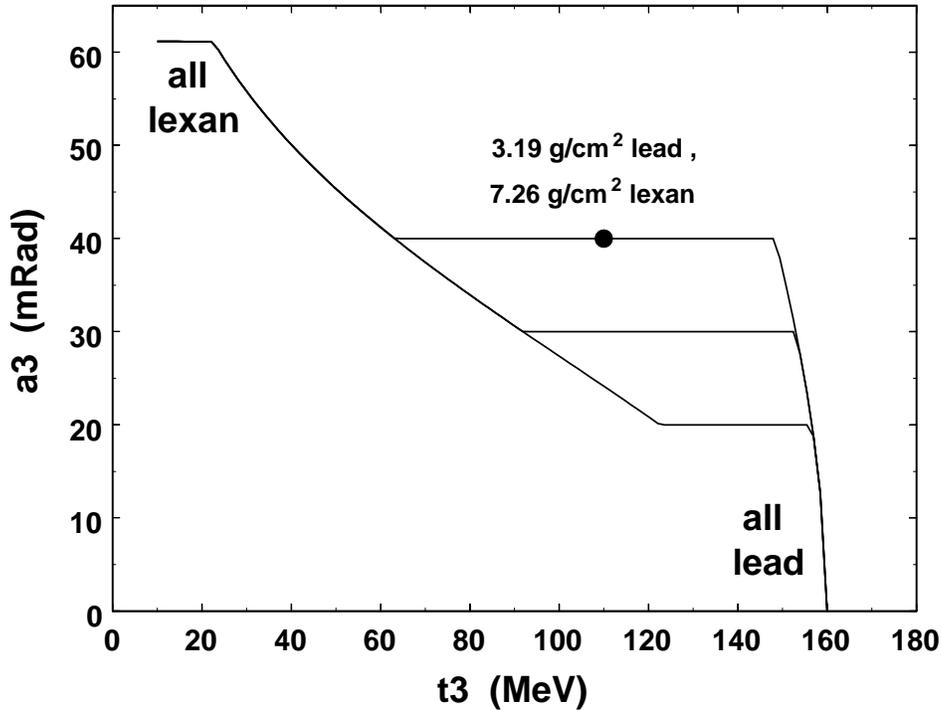


Figure 9.2: Scattering v. output energy for a lead/lexan degrader.

than 97% of the proton range. Multiple scattering saturates [4] and `Theta0` treats it as constant.

Program `(BDG(1...))` solves (9.1). It accepts any two materials in either order but you'll only get the desired effect if one is higher-Z and the other lower-Z. Before attempting a solution `BDG` makes two tests and takes appropriate action:

1. If the desired scattering exceeds what can be obtained with all higher-Z, satisfy the energy requirement, return $g(\text{higher-Z})$, the scattering $a3$ actually obtained, and $g(\text{lower-Z}) = 0$.
2. If the desired scattering is less than what can be obtained with all lower-Z material, satisfy the energy requirement and return $g(\text{lower-Z})$, whatever $a3$ is obtained, and $g(\text{higher-Z}) = 0$.

Otherwise, we solve the problem as described in section 9.4 using $t2$, whose brackets are obviously $t1$ and $t3$, as the primary variable. See `BDG.FOR` for details. If $a3$ returned equals its initial value, all specifications were met.

The three other binary degrader problems are solved in the same general way. We will meet them as we design modulators whose scattering is independent of modulator step ('compensated absorbers') or contoured scatterers whose energy loss is independent of radius ('compensated scatterers').

9.6 Does the Order of Materials Matter?

Definitely. If the lead is second it sees a lower energy ($t2$) and scatters more. Therefore less lead is needed. If we run our previous example (160 MeV in, 110 MeV and 40 mrad out) through `LOOKUP` we find

order	t_2 MeV	lead g/cm ²	lexan g/cm ²
lead/lexan	151.6	3.186	7.261
lexan/lead	116.0	1.837	7.985

The range over which a solution exists does not change, however. We can't do any better by changing the order. If a degrader is all lead, it doesn't matter whether it's first or last! In the present example the t_3 range over which we can hold 40 mrad is 64 to 148 MeV (Figure 9.2).

9.7 Summary

We have seen that high-Z materials do more scattering and low-Z materials, more stopping. We have learned how to solve the forward problem for arbitrary stacks of degraders, finding the scattering angle and output energy at the end of the stack (TOUT.FOR). The same procedure can be reversed to find the input energy and angle of a stack given the output energy and angle (TINP.FOR).

Inverse problems, where one or both degrader thicknesses are unknown, can be solved with a root finder as long as we can write down a procedure that solves the forward problem. We discussed an inverse single degrader problem and an inverse binary degrader problem in detail. Binary *e.g.* lead/lexan degraders can be designed to satisfy an energy loss requirement and a scattering requirement simultaneously. The energy range over which this works does not depend on the order of the two materials but, within this range, the proportion of each material does.

In this chapter we have introduced a style of problem solving which we will use throughout. It attaches little or no value to writing down functions or solutions in closed form. This usually takes considerable effort and may require undesirable approximations. We regard a function simply as a numerical procedure for finding one variable given another, and we use numerical root finders to solve inverse problems.

Chapter 10

Degrader Stacks and Beam Spreading

In the previous chapter we learned how to find the output energy and scattering angle for a stack of degraders, but that's not the whole story. If we want to treat an extended stack as a single degrader we must find out where the protons seem to be coming from, their effective origin. To compute that we need to consider beam spreading in a single degrader or a stack.

10.1 Beam Spreading

Preston and Koehler [10] used the method of projection onto a measuring plane (MP) to compute beam spreading in a thick degrader. Their method is easily generalized to a stack of n degraders. In the Gaussian approximation (which we use throughout this chapter) the multiple scattering is Gaussian and, because the angles are small, its projection on an arbitrary measuring plane is also Gaussian. Therefore the 2D fluence per proton at the MP is

$$P'(r, \phi) r dr d\phi = \frac{1}{2\pi r_0^2} e^{-\frac{1}{2}\left(\frac{r}{r_0}\right)^2} r dr d\phi \quad (10.1)$$

and we need to compute r_0 .

Consider the characteristic multiple scattering $d\theta_0$ in a layer dz (Figure 10.1). The characteristic transverse kick on the MP is $dr_0 = (Z - z)d\theta_0$. Consistent with the Gaussian approximation we want to add such kicks in quadrature to find r_0 . According to Highland's formula (7.26) $(d\theta_0)^2$ is proportional to $\rho dz/L_R$. For a stack of n scatterers we obtain

$$r_0^2(Z) = \sum_{j=1}^n \left[\left(1 + \frac{1}{9} \log_{10} \frac{L_j}{L_{Rj}}\right) \int_{z_j}^{z_j+L_j} \left(\frac{14.1(Z-z)}{pv}\right)^2 \frac{\rho_j}{L_{Rj}} dz \right] \quad (10.2)$$

The limits of integration are valid as shown if the MP is downstream of all the degraders. The power of the method is that it need not be. It can be anywhere downstream of z_1 , and by computing r_0 at various Z 's we find the envelope of the beam. The limits of integration must, of course, be appropriate to the position of the MP.

In (10.2) we have arbitrarily taken Highland's logarithmic correction factor outside the integral, evaluating it for each degrader *as a whole* rather than for each infinitesimal layer. Some such step is necessary. Otherwise the integral

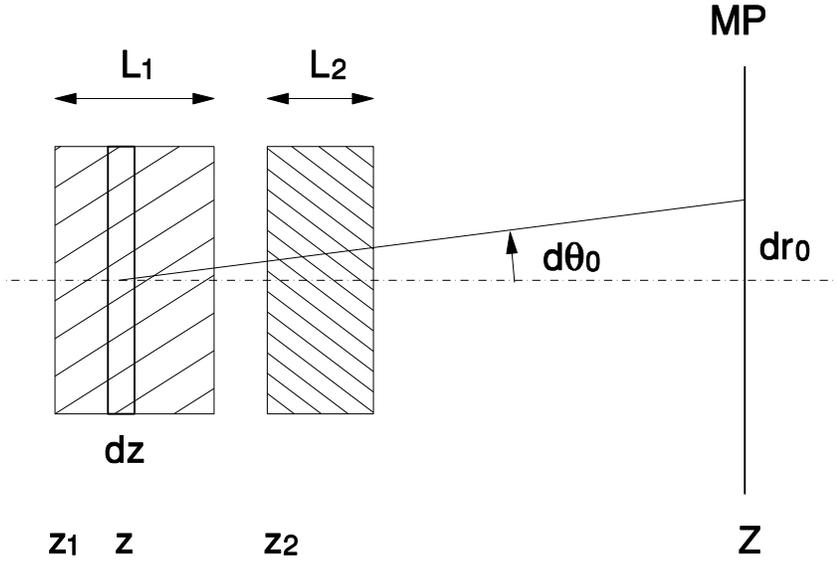


Figure 10.1: Displacement dr_0 in MP from scattering in a target layer dz .

would not reach a limit as a function of dz . The answer would keep getting smaller. We could, instead, freeze the number of intervals in the integral, but that too would be arbitrary. The ultimate justification of (10.2) is that it agrees with experiment, as we will show.

Function `Rch` in file `TOUT.FOR` computes r_0 according to (10.2) using Simpson's Rule (7.24) to evaluate the integral. pv decreases as the proton travels down the stack. It must be updated whenever the integrand is evaluated. `Rch` accepts arbitrarily many degraders of arbitrary materials. Before computing r_0 it checks that the degraders do not overlap in z .

10.2 Experiment of Preston and Koehler

Preston and Koehler [10] measured beam spreading for 127 and 134 MeV protons incident on water and 112 and 158 MeV incident on aluminum. They showed that $\sigma(T)/\sigma(R_0)$ is a universal function of $y = T/R_0$, R_0 being the initial range in whatever material and T is the depth at which σ is measured (their notation). They found a closed expression for that function using the multiple scattering theory of Bethe and Ashkin [23]:

$$\sigma(T)/\sigma(R_0) = \left[2(1-y)^2 \ln \frac{1}{1-y} + 3y^2 - 2y \right]^{1/2} \quad (10.3)$$

Figure 10.2 reproduces their Figure 17 which shows excellent agreement between theory and experiment.

Much later we replotted their data (Figure 5a of [4]) against calculations from an early version of `Rch` and adding a measured point that had been omitted in [10]. The four curves (water and aluminum at two energies each) were identical:

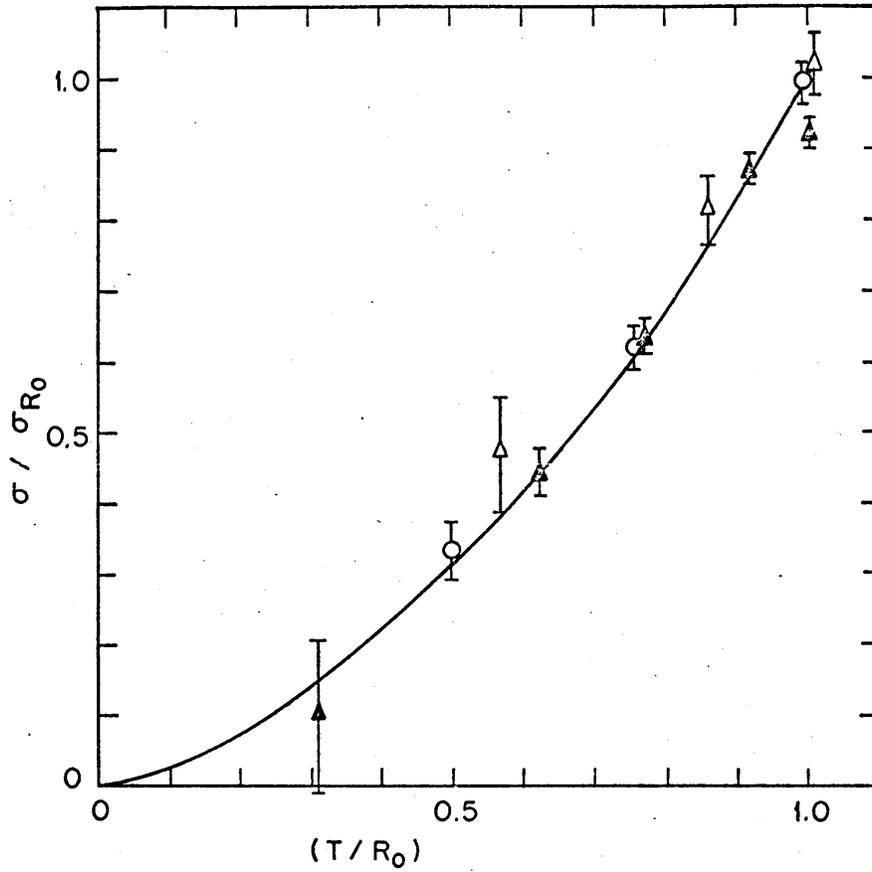


Figure 10.2: (caption from [10]) Dimensionless plot of the standard deviation due to scattering versus depth of penetration. The curve is a plot of Eqn. (15) of Sec. 2 (author: eq. 10.3 of the present report). Open triangles are experimental results for 112 MeV protons on aluminum; solid triangles for 158 MeV protons on aluminum; open circles for 127 MeV protons on water.

beam spreading is indeed universal in Preston and Koehler's scaled variables. A direct comparison showed $\pm 2\%$ differences between Rch and (10.3). Rch is therefore well tested in single homogeneous degraders.

10.3 Effective Origin z_0

Far downstream of a thick target or degrader stack, protons will again travel in straight lines. The point at which those back-projected lines cross the z axis is z_0 , the *effective origin* of scattering. Figure 10.3 illustrates the idea.¹ z_0 obviously bears some relation to the degrader thickness, which may be ≈ 25 cm if the degrader happens to be a range modulator at 220 MeV. The variation of z_0 as the modulator turns will cause an appreciable variation of the fluence ($1/r^2$ effect)² particularly in a short beamline (gantry).

¹ The solid lines curve even outside the target: they are envelopes of r_0 , not the trajectories of protons.

² There are few if any *explicit* $1/r^2$ corrections in our programs. The $1/r^2$ effect is included in the fluence calculation, which requires z_0 .

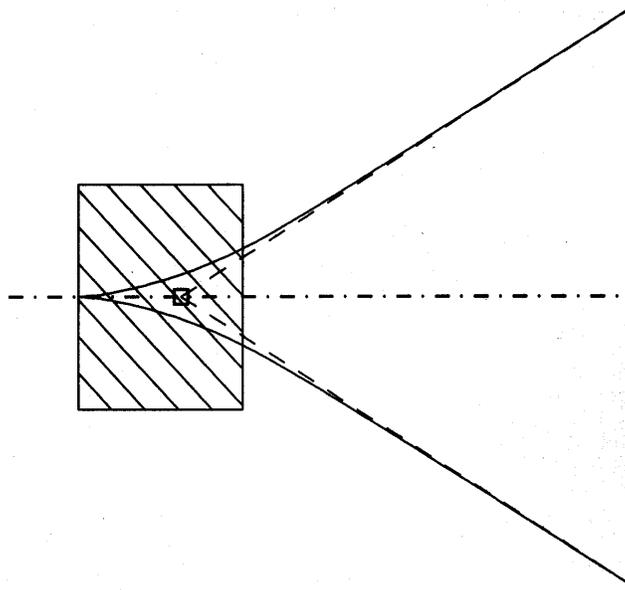


Figure 10.3: Effective origin of scattered particles in a thick target.

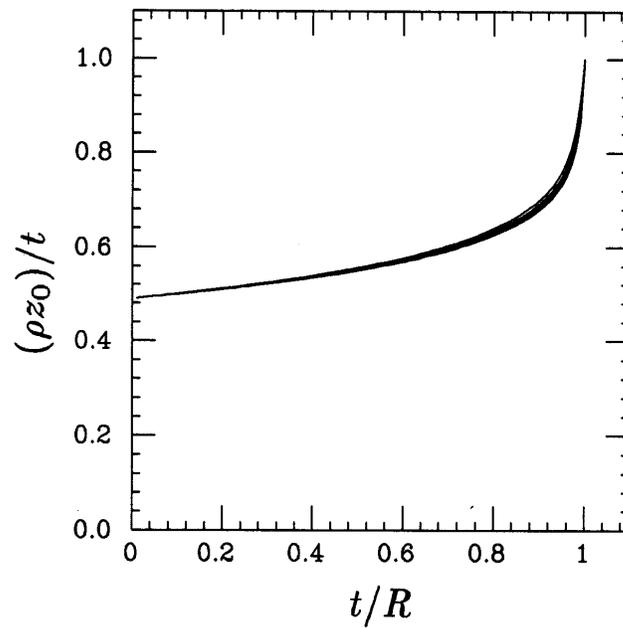


Figure 10.4: Effective origin of scattering ρz_0 divided by target thickness t as a function of normalized target thickness. The lines represent nine separate calculations for 50, 150 and 250 MeV protons on Be, Cu and Pb.

In learning how to compute beam spreading we have already done the hard part of finding z_0 . Proceed by finding the characteristic radius r_{01} and the characteristic scattering angle θ_{01} at the exit of the stack, $Z_1 = z_n + L_n$. These set the displacement and angle scales for the problem at hand and can be used to compute two locations sufficiently far downstream, say

$$Z_2 = Z_1 + 10 \times r_{01}/\theta_{01} \quad \text{and} \quad Z_3 = Z_1 + 20 \times r_{01}/\theta_{01}$$

Evaluate r_{02} and r_{03} at these points. By geometry

$$z_0 = (r_{03}Z_2 - r_{02}Z_3)/(r_{03} - r_{02}) \quad (10.4)$$

The asymptotic slope should equal the characteristic angle at the exit of the stack, that is, you should find

$$(r_{03} - r_{02})/(Z_3 - Z_2) = \theta_{01}$$

This tests the consistency of the calculation and the accuracy of the numerical integration. It will not quite check if you use the `MOLIERE` option to find θ_{01} because the Gaussian approximation using `HIGHLAND` is built into `Rch`.

The effective origin as a fraction of target thickness, $\rho z_0/L$, is a nearly universal function of normalized target thickness L/R_0 as shown in Figure 10.4 where it is plotted for three widely spaced materials each at three widely spaced energies (see caption). The effective origin is halfway through the target for thin targets. It moves downstream as the target gets thicker. For very thick targets, low energy protons appear to originate from the downstream face.

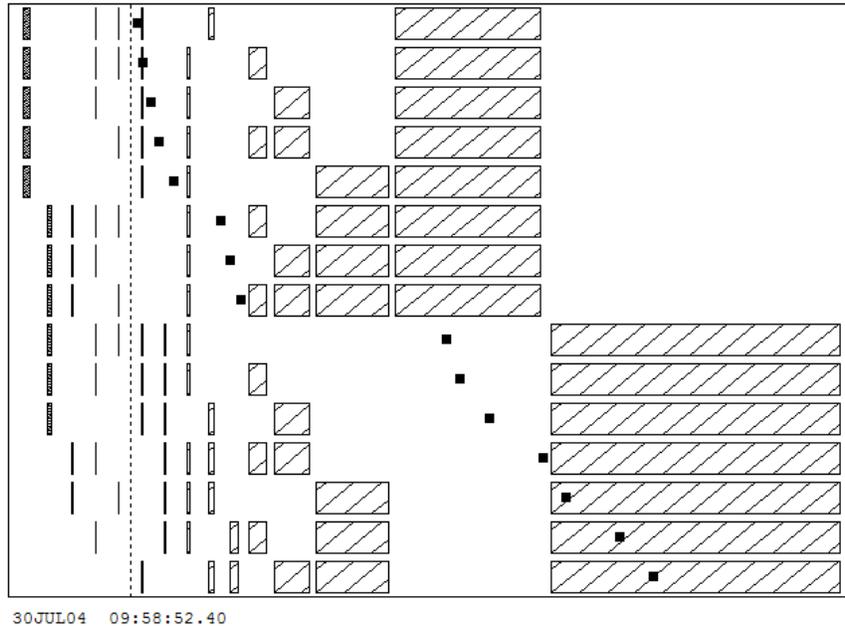


Figure 10.5: Variation of z_0 in a lead/lexan binary stack.

Modules `Zeff1`, `Zeff6` and `ZeffN` of file `TOUT.FOR` compute z_0 for various cases. As mentioned above, `Rch` does the hard work. Figure 10.5 shows the variation of z_0 as a stack consisting of lead and lexan degraders in binary increments is used to create a spread-out Bragg peak (program `LAMINATE`). Note the jump in z_0 at major binary transitions.

10.4 Fermi-Eyges Theory

We shall not use Fermi-Eyges theory. It addresses the same problem as this chapter and is often cited, particularly for electron beams, so we will try to put it into perspective. The author is not expert in this theory so the following comments should be taken with a grain of salt.

Fermi wrote down the joint probability distribution for multiple scattering in the Gaussian approximation: the probability that the particle acquires a given angle (projected onto a plane) *and* a given displacement (also projected). Evidently he never published this but a University of Chicago lecture was quoted extensively by Rossi and Greisen in their review article [37]. Fermi ignored energy loss but Eyges [51] generalized the theory to include it.³ Formally, Fermi-Eyges theory is an elegant solution to the problem we have been discussing. However, it embodies the inaccurate Rossi formula for θ_0 , and it seems to be restricted to a single homogeneous degrader.

Let's take the first point. Eyges defines a function $A_0(t)$ which is essentially the characteristic multiple scattering angle. t is depth in units of the radiation length. Working through the differences in notation we find, in the case of no energy loss, $2A_0 = \theta_0^2$ (Rossi). Referring to the second equation in the footnote, we see that Eyges' angular distribution is just the familiar Gaussian with Rossi's formula for θ_0 :

$$\theta_0 = \frac{15.0 \text{ MeV}}{p v} \frac{\sqrt{L}}{L_R} \quad (\text{Rossi})$$

which is not good enough, at least for protons. For 160 MeV protons into 1 g/cm² Pb, which gives $\approx 1^\circ$ of scattering, $\theta_0(\text{Rossi}) = 1.17 \times \theta_0(\text{Highland})$, and we know that Highland is just about right. It is not clear how intimate the connection with Rossi's formula is. Substituting a Highland-type formula would clearly cause problems with the integral over t .

Eyges does not discuss the second point directly. If the theory applied to stacks of degraders of different materials, possibly with air gaps, one would expect a discussion of how the scaled depth t is generalized and how the integrals defining A_0 , A_1 and A_2 are broken up. [51] is written very much as though it applies to a single material, though this is not stated. It may be that later authors have extended the theory.

10.5 Summary

We learned how to compute beam spreading r_0 by adding characteristic kicks in quadrature on an imaginary measuring plane. Numerical results agree well with measurements and with earlier computations [10]. Using two measuring planes, sufficiently far downstream, we can project the beam envelope back to find the effective origin z_0 . This can be computed for stacks of arbitrary complexity. We discussed the relationship of Fermi-Eyges theory to our method.

³ Eyges' Eqns. (14) and (15) have typographic errors, corrected by Zerby and Keller [52]. They should read respectively

$$F(t, y, \theta) = \frac{1}{4\pi B^{1/2}} \exp\left(-\frac{\theta^2 A_2 - 2y\theta A_1 + y^2 A_0}{4B}\right)$$

$$\int_{-\infty}^{\infty} F(t, y, \theta) dy = \frac{1}{2(\pi A_0)^{1/2}} \exp\left(-\frac{\theta^2}{4A_0}\right)$$

Chapter 11

The Bragg Peak

The Bragg peak (BP) as we shall use the term is the depth-dose distribution, in water, of a monoenergetic proton beam.¹ A carefully measured Bragg peak is essential to designing a range modulator. In this chapter we will discuss factors that influence the shape of the Bragg peak, precautions needed in measuring it, and how to convert the measurements into a mathematically convenient form for general use.

11.1 Understanding the Bragg Peak

Figure 11.1 shows a set of Bragg peaks measured at IBA during the cyclotron acceptance tests for NPTC.² Each curve is normalized so that its entrance value corresponds to the tabulated mass stopping power in water at that incident energy. Therefore the curves may be regarded roughly as curves of effective stopping power *vs.* depth averaged over all the protons at each beam energy.

11.1.1 Stopping Power, Straggling, Beam Energy Spread

The increase of stopping power as the proton slows down, followed by its plunge to zero when the proton stops, dominates the overall shape of the BP. If we could measure the stopping power S/ρ of an *individual* proton we would see a very high and sharp spike. Because of range straggling, protons stop at slightly different depths. The distribution of endpoints is Gaussian with some σ_S that can be found from straggling theory and has been tabulated for many materials by Janni [3]. The depth-dose of a large cohort of stopping protons is therefore the convolution of S/ρ with a Gaussian describing the range straggling.

If the incident beam is not strictly monoenergetic but has some energy spread, we must add that in quadrature to σ_S after converting it to an equivalent range spread.

Range straggling expressed as a percentage of range falls slightly with energy, but that means that *absolute* straggling increases strongly with energy [3]:

MeV	straggle %	straggle g/cm ²
70	1.210	0.050
150	1.123	0.179
250	1.063	0.407

¹ We may use ‘Bragg peak’ either for the entire curve or for the peak in the curve.

² We are indebted to Damien Priels of IBA for these data.

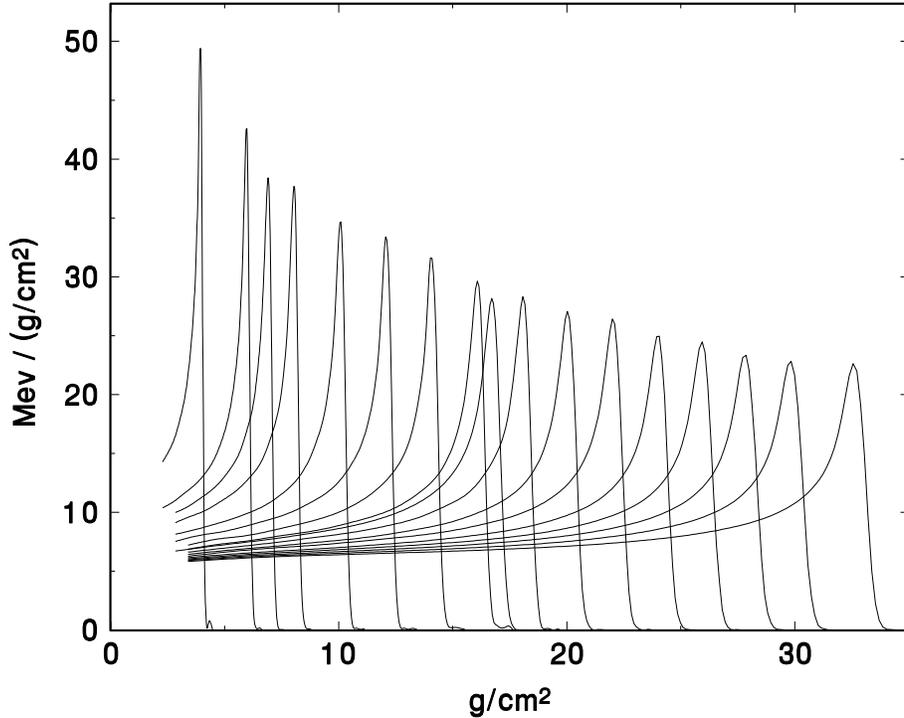


Figure 11.1: Measured Bragg peaks from 69 to 231 MeV.

Therefore the Bragg peak grows wider as energy increases (Figure 11.1). This is an important factor in the design of range modulators.

11.2 Theoretical Models

We briefly interrupt the physics discussion to comment on theoretical models of the Bragg peak *vs.* direct measurement. Many writers have modeled the Bragg peak mathematically, performing the convolution graphically [53], analytically [26], numerically [54] or by means of Monte Carlo calculations [55]. The citations are just samples from a large literature. All three approaches work, though the beam energy spread is invariably reserved as an adjustable parameter to improve the agreement. Usually there are a few other free parameters such as a rough model of nuclear interactions.

In other words, we can take it as proven that the observed Bragg peak can be rather accurately predicted from the underlying physics. Our approach to design, however, does not rely on this or make use of it. Instead, we will measure the Bragg peak relevant to our treatment setup, parametrize the measurement in a model independent way, and use that directly. If we are only doing a rough design and do not yet have the ability to measure Bragg peaks, we can borrow one that was taken under similar conditions. Before doing a final design and building actual modulators, it is essential to remeasure the Bragg peak under the prevailing conditions using the guidelines given below.

It would appear that a good theoretical model would make remeasurement unnecessary. In practice this is not the case. All available models have free parameters that are not known *a priori*.

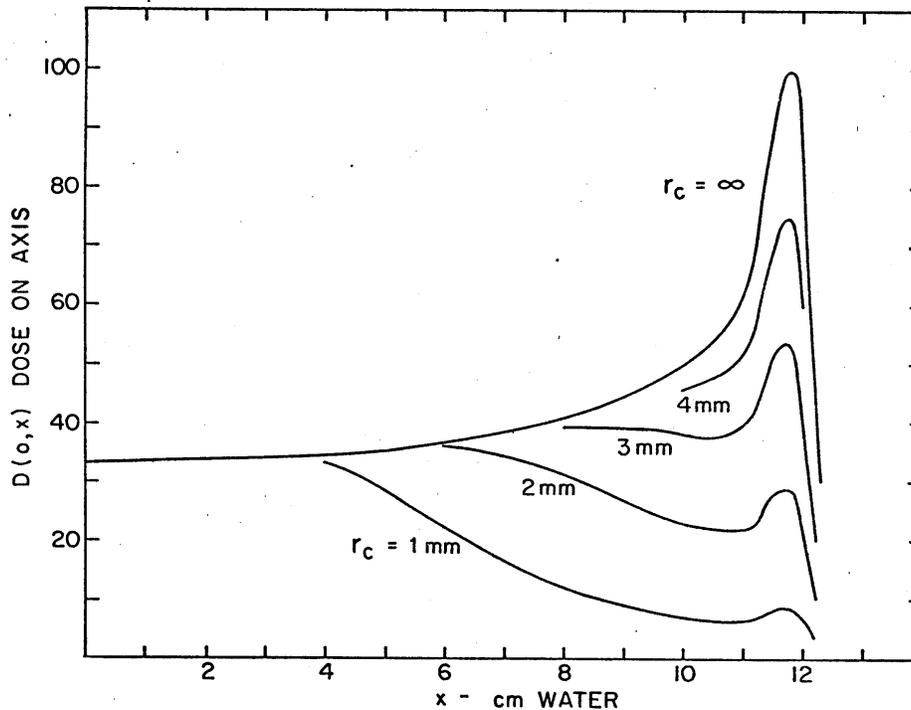


Figure 11.2: The relative dose on the axis of a uniform circular proton beam of initial range 12 cm of water and radius r_c at the collimator. The curve for $r_c = \infty$ is experimental, the others are calculated.

11.2.1 Beam Size

Consider the depth-dose distribution on the axis of a proton beam of large cross section, a so-called ‘broad’ beam. As the beam penetrates the water, protons scatter out of volume elements lying along the axis. However, there is no net loss. Because the beam is broad, an equal number of protons scatter *into* those volume elements. This condition is called *transverse equilibrium*. It obtains at all depths if the beam radius entering the tank is much larger than the final radius of an entering pencil beam would be (see Chapter 10).

If the beam is not broad, there is a net loss of protons from on-axis volume elements, increasing with depth because the cumulative scattering increases. In other words, the fluence on axis decreases with depth because of multiple scattering. Therefore the dose also tends to decrease with depth, counteracting its normal tendency to increase. We lose the Bragg peak.

Figure 11.2 from [10] is a calculation of this effect by Preston and Koehler for beams of circular cross section. Of course we must use a small dosimeter to measure this, that is, we must truly measure the dose on axis. In fact if we use a dosimeter, say a plane parallel ion chamber, that covers the whole beam we will not see the effect at all. The Bragg peak will be the same as one measured with a small chamber in a broad beam. But that would not reflect the dose received by a small lesion on axis.

11.2.2 $1/r^2$ Effect

Dose equals fluence times stopping power. If the incident beam is broad and parallel the fluence will be the same at all depths (ignoring nuclear interactions). In practice, the protons usually diverge in a cone from some effective source. They are therefore denser at the entrance to the tank than at depth. The fluence decreases approximately as $1/r^2$ where r is the distance from the effective source to the depth of interest. This tends to decrease the peak to entrance ratio of the Bragg peak, more so the shorter the source distance. Therefore it is a significant effect in proton gantries, whose throw is short for practical reasons.

In design programs we must be careful to correct for $1/r^2$ but, by the same token, not to double-correct for it. The way we will handle this is to take $1/r^2$ out of the measured Bragg peak, that is, to convert it to what we would have seen in a parallel beam. To do this we must of course know what r was at the time of measurement. Later, when calculating with this modified Bragg peak, the finite source distance correction will come in automatically when we compute the fluence in the given geometry.

11.2.3 Tank Wall and Other Degradors

The ideal Bragg peak for design purposes would be one measured at the nozzle entrance in an ideal (wall-less) tank with an ideal dosimeter. That is impractical. In practice, the water tank will be at isocenter. We'll need one or two scatterers to spread the beam. The water tank has a wall to keep the water in, and the dosimeter has a wall to keep the water out. Therefore our measurements start at some finite water-equivalent depth rather than 0.

Like $1/r^2$, the presence of all this 'dead' material is not really a problem as long as we keep careful notes on the measurement. What we do with the information later is a bookkeeping problem that depends on the nature of the design program. We'll discuss it in Section 11.4.

11.2.4 Dosimeter

Small (Markus-type) plane parallel air filled ionization chambers have the best geometry for heavy particle beams. They are usually assumed [12] to be ideal dosimeters, having no energy dependence at the percent level. While arguments can be made for this, it is worth pointing out that it is an assumption. To our knowledge there are no experiments to prove it. By contrast, thimble chambers do have small energy effects, geometric in origin [56]. A Bragg peak taken with a thimble chamber will exhibit a wider peak and a smaller peak/entrance ratio than one taken with a Markus chamber under the same conditions.

Some semiconductor diodes have rather large energy effects. For instance, Koehler [57] reported a peak/entrance ratio $\approx 8\%$ higher for a certain silicon diode than for a nitrogen filled plane parallel IC. Other diodes show little or no difference from plane parallel IC's.

The choice of dosimeter in a clinical program depends on a number of factors: size, cost, familiarity of the staff with that particular dosimeter and so on. Our point here is merely that the dosimeter used to measure Bragg peaks for design purposes should be the same as the one that will eventually be used to confirm the spread-out Bragg peaks. Otherwise, discrepancies may be expected.

11.2.5 Low Energy Contamination

If there is a collimator near the water tank there will be some protons that enter the front collimator material but scatter back out [58, 59]. These are directed

inwards (otherwise they would have stayed in the material), b) have a wide angular distribution with a mean of $\approx 6^\circ$ (for brass) and c) have a wide energy distribution. They converge towards the axis, then cross over and diverge, rapidly becoming quite dilute. The contamination is most pronounced for small collimators because the circumference is larger compared to the area, and for small air gaps. It appears as a large proximal enhancement in the depth-dose in Figure 11.3 taken from Hong et al [60], admittedly a rather extreme case. Incidentally, this is also a nice experimental illustration of the disappearing Bragg peak for small beams.

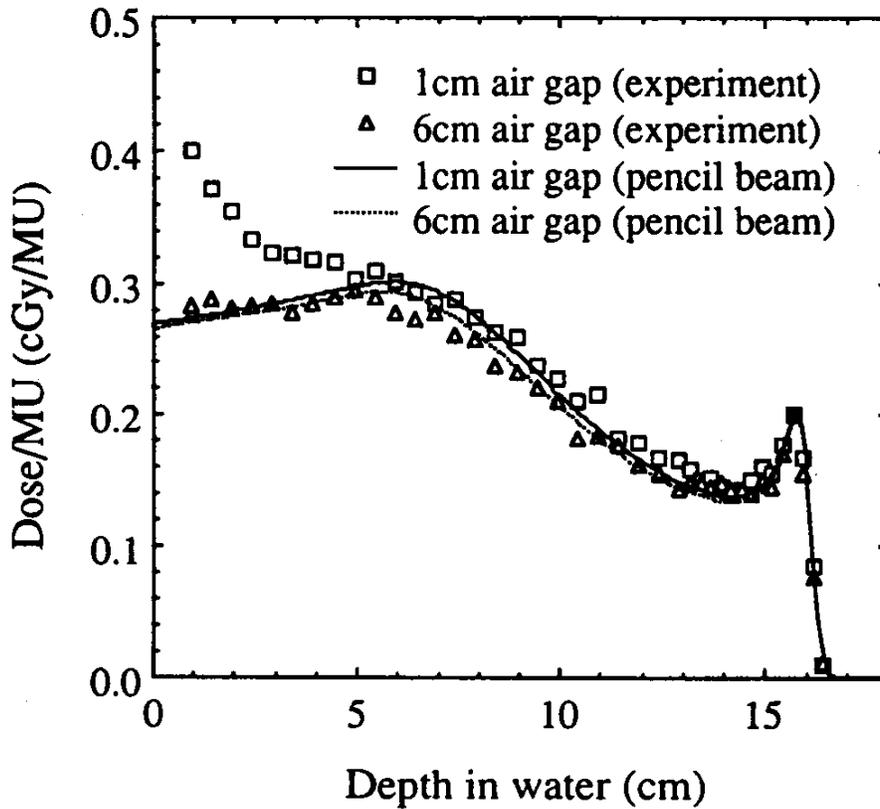


Figure 11.3: (From Hong et al. [60]) Central-axis depth-dose measured with a diode. The beam is collimated by a 2.4mm radius aperture. The air gap between the aperture and the water tank is 1 or 6 cm. Lines: computations using pencil beam algorithm.

11.2.6 Nuclear Interactions

As protons stop they suffer nuclear interactions at the rate of $\approx 1.2\% / (\text{g}/\text{cm}^2)$. The primary protons acquire a companion fluence of short-range charged secondaries (which receive about 60% of the primary energy) and long-range neutrals. The energy of charged secondaries is deposited pretty near the interaction point. The remaining 40% is 'lost', that is, deposited far from the volume of interest. The net effect is to remove energy from the peak, some of which is redistributed to shallower points and some of which is lost. Any theoretical model of the Bragg peak must account for this at least roughly in order to get the right

peak/entrance ratio. Chapter 8 discusses this in much more detail.

11.2.7 Electron and Nuclear Buildup

If the Bragg peak is measured with little or no solid material just before the water (no tank wall and a very thin dosimeter wall) you will see electron and nuclear buildup in the entrance region. The only published measurement of both that we know of is that of Carlsson and Carlsson [47], already discussed in Chapter 8. To understand electron buildup, assume an uncontaminated beam enters the water from vacuum. (In practice, air is good enough.) The protons will not yet have their accompanying sheath of low energy electrons caused by EM interactions. It takes ≈ 1 mm for this sheath to build up. Because all the absorbed dose to water is due to these secondary electrons, the *true* entrance dose is zero. However, it builds up to the equilibrium value very quickly, so you have to work hard to see the effect. For instance, the standard cap supplied with a Markus chamber is already thick enough to establish electron buildup.

As an aside, we'll comment that the rapid electron buildup is actually one of the *disadvantages* of protons relative to high-energy photon beams. The latter (being neutral) have considerably longer buildup which leads to *skin sparing* in clinical situations. Protons, because of their short buildup and the almost inevitable presence of some buildup material just before the skin, have little or no skin sparing. This must be taken into account in treatment plans.

Nuclear buildup is like electron buildup except that we are concerned with the sheath of *nuclear* secondaries. Nuclear buildup is easier to see because the buildup distance is of the order of a centimeter, but it is smaller because nuclear secondaries account for only a small fraction of the absorbed dose. You will see it if you do a vertical scan in a gantry (the easiest way to get a wall-less water tank) or if your water tank has a thin entrance window.

Do you want to measure it? Probably not. In the clinical application of your modulator you are very likely to have some buildup material or low-energy beam contamination from the patient specific collimator, so including the little entrance dip in your design Bragg peak is probably a worse approximation of reality.

11.2.8 Pathlength Effect

The EM peak in the number *vs.* depth curve which underlies the Bragg peak is not exactly Gaussian but slightly skewed to the shallower side. This is because it refers to *projected* range, which can be smaller than, but never greater than, the pathlength. This is a very small effect for protons where the pathlength and projected range are nearly the same [3]. Nevertheless it is included in the best theoretical models [54].

11.2.9 Checklist for Measuring the Bragg Peak

This checklist summarizes the previous discussion.

- Be sure you have a *broad beam*. If you already have a beam spreading system, use it. Otherwise, you can use a single lead scatterer at some distance upstream from the tank. By now you should be able to compute the scatterer required. As a check, make sure the measured Bragg peak does not change when you move your measurement axis a centimeter or two to the side.
- Be sure you know your *effective source distance* and make it as large as practical.

- Use the same *dosimeter* as you plan to use later in QA.
- Be sure you know the *tank wall* thickness, the dosimeter wall thickness and the amount of water between the wall and the dosimeter when the scanning system reads ‘zero’. Record the beamline geometry including the thickness of everything in the beam: scatterers, monitor IC and so on.
- If the beam *energy spread* is adjustable by momentum selection slits, be sure these are set to the value you plan to use for therapy.

11.3 Commercial Equipment

You will probably use commercial ‘water phantom’ hardware and software to measure the Bragg peak. Unfortunately, in our experience equipment designed for conventional photon therapy is not as convenient for proton beams as it might be, either for preliminary measurements or day-to-day QA. This situation is unlikely to change because of the tiny market for particle beam therapy equipment.

11.3.1 Tank Size

Water tanks for conventional photon therapy are very large. They come with their own cart and a pump to fill and empty the tank. Setting these up can be quite burdensome in the schedule of a busy facility. Though NPTC owns such devices they are almost never used.

Miles Wagner had the important insight that a water tank for heavy charged particles can be quite small in cross section because protons and charged nuclear secondaries do not move very far transversely to the beam. If we have a large water tank we can insert any object we like a few cm off the scan axis and it will have no effect on a depth-dose measurement. We can even go so far as to eliminate the water starting a few cm from the scan axis. Miles placed a high-Z scatterer along a water tank wall and then found out how close to it he had to scan before it had any effect on the depth-dose, concluding that a transverse distance of 5 cm between the dosimeter and the nearest tank wall would be ample at 160 MeV. This led to the computer controlled ‘Oilcan’ dosimeter used for many years at HCL to map dose fields in 3D.³ A small detector or detector array moves along the centered depth axis of a small water tank. Transverse distributions are mapped by moving the entire tank.

Even if you don’t go to all the effort of building a computer controlled device, you will probably want to replace your large commercial tank by a small one which you can set up by hand in a minute. The length should correspond to whatever depth-dose you want to measure plus a few cm. Of course it will only have room for a single (depth) scan direction. To measure transverse doses you will have to use it sideways or scan the patient positioner on which it is mounted.

The tank wall of commercial tanks is much thicker than it needs to be, so a potentially important part of the depth-dose cannot be measured. It is relatively trivial to furnish a smaller water tank with a large thin window of (say) 0.005" Kapton. If it bulges a little you need merely move the dosimeter to touch it and reset the scanner to zero.

³ The original concept called for a water tank similar in size and shape to the cylindrical tin can in which one used to buy motor oil. In time the tank became rectangular, which is probably easier to fabricate anyway.

11.3.2 Current *vs.* Charge Measurement

There are two ways of mapping dose fields, electronically speaking. You can scan the dosimeter continuously, sampling the current from the dosimeter and a reference IC simultaneously at frequent intervals. The CRS scanner [61] works that way. Or, you can move to a point, stop, measure dosimeter and reference IC charges with a current integrator for some fixed period, then move to another point and repeat. The Scanditronix and Wellhöfer tanks work that way. The current sampling method tends to be noisy and some mathematical smoothing is needed. Although both ways work and have been used successfully at NPTC, we somewhat favor the integrator method for conventional scans.⁴ Neither method has a clear speed advantage if they are used intelligently.

11.3.3 Beam Monitor *vs.* Reference Chamber

Conventional water tanks assume the user will place a *reference chamber* in some fixed location. They plot the ratio of the ‘field’ chamber to the reference. That is not the most convenient arrangement in a proton therapy center. Particle beams frequently have large transverse dose gradients. Also, we may be interested in changing beam line parameters during a series of measurements. Therefore it may be difficult to find a good fixed location for the reference chamber.

There is always a beam monitor IC somewhere in the beam line, if only because we need to turn off the beam when the prescribed dose is reached. This monitor IC is the most convenient reference. If the measuring system keeps track of the absolute monitor output as well as the absolute field IC output, each measurement of a Bragg peak or SOBP automatically becomes a measurement of the monitor output factor for that beam line configuration. One difficulty is that the beam monitor output current already goes elsewhere in the control system. We must find a way of either a) sharing it in such a way that the fraction going to the control system does not depend on whether the scanner is connected or not or b) tapping into the electronic output of the monitor current integrator and convincing the scanner to use it as a reference.⁵ At NPTC we have generally found it difficult to trick commercial scanning tanks into using the beam monitor as a reference.

11.3.4 Scanning *vs.* Dosimetry Mode

Proton facilities require QA (confirming SOBP shapes) as well as calibrations (confirming the monitor output factor). If stable current integrators are used and absolute charges are recorded, a point by point SOBP measurement doubles as a measurement of the output factor for that scatterer and modulator system. One has only to use a calibrated dosimeter and record the temperature and pressure at the time of measurement.

Unfortunately, commercial water phantom software usually has completely separate ‘scanning’ and ‘dosimetry’ modes because of the conventions of photon therapy. In scanning mode the measured absolute charge tends to disappear in a welter of renormalizations. Inability to use scans for absolute measurements leads to duplication of effort.

⁴ The CRS technique is used at NPTC for time-correlated depth-dose measurements, which are useful in diagnosing range modulator performance.

⁵ One method is to convert the pulse train from the monitor integrator back into a current and use that for the reference [62].

11.4 Fitting the Bragg Peak

Having measured our Bragg peak we wish to convert it into a standard form. We fit it with a cubic spline as explained in Chapter 5. It is convenient to do this in an auxiliary program BPW (for Bragg Peak Width). BPW reads the data, fits the spline, produces graphic output and writes a .BPK file which contains all the information needed by a subsequent programs. Figure 11.4 shows a typical run, whose output file RNE175.BPK is listed below the figure.

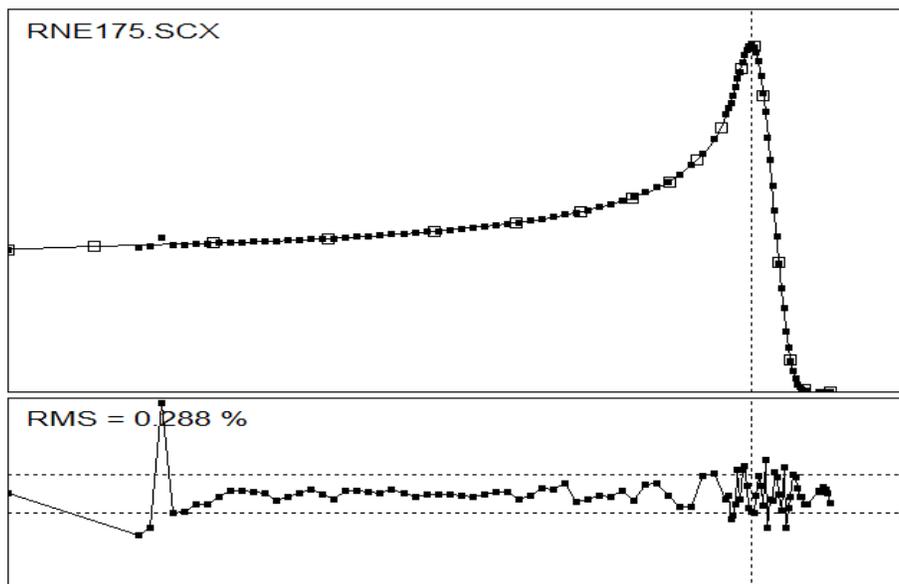


Figure 11.4: Bragg peak extrapolated to zero and fit with a cubic spline by program BPW. The BPK file is listed below.

```
put a comment here ...
280.0 cm from source to Bragg peak
18  0.10000E+31  0.10000E+31  # pts, yp1, ypn
  0.0000  1.8703  4.4415  6.9359  9.2388  11.0088  12.4067  cm H2O
13.5244  14.3358  14.9433  15.4579  15.8944  16.1706  16.3812
16.7241  16.9718  17.2745  17.8250
  0.4105  0.4196  0.4290  0.4418  0.4630  0.4877  0.5186  rel dose
  0.5599  0.6061  0.6689  0.7616  0.9325  0.9978  0.8550
  0.3724  0.0908  0.0057  0.0014
```

11.4.1 Overlying Material

For any Bragg peak measurement there will be material in the beam, some associated with the tank and measurement process ('wall correction') and some associated with the beam line (scatterers, monitor IC's, air ...). We need to handle this in a consistent fashion. Specifically, the scanning software usually calls the depth of the first point zero, since it has no way of knowing the true depth. We need to add a water equivalent depth to each recorded depth to make the depth-dose the same as what we would have measured with an ideal tank and dosimeter.

The wall correction should always be made and usually has three parts: the

tank wall, some finite water gap when the scanner reads zero, and the water equivalent thickness of the dosimeter cap or wall.

Material further upstream requires more thought since its treatment depends on how we plan to use the Bragg peak. At one limit, the prospective computer program may account explicitly for everything in the beam line at some future time and this may not be the same as it was at the time of measurement. In that case we want to conceptually ‘move’ the tank to nozzle entrance before we fit the Bragg peak, by adding to each point the water equivalent of everything between nozzle entrance and the tank at the time of measurement. If that includes high-Z materials such as lead we must be careful because their water equivalence depends on energy. Note also that air is not negligible: a meter of air corresponds to about a millimeter of water.

Other computer programs treat explicitly only some of the objects in the beam. For instance, NEU ‘knows’ about the first scatterer or modulator, pre- or post-scatterers, and the second scatterer, but not about air and monitor IC’s. It assumes that the ignored things are the same ‘now’ as they were during the Bragg peak measurement. Therefore the depth data should *not* be corrected for air and monitor IC, but *should* be corrected for scatterers in place at the time of measurement.

11.4.2 Extrapolation to Zero

Once we have done the depth correction we will have a Bragg peak that no longer starts at zero. Should we leave it that way (*cf.* Figure 11.1) or generate a fake point at zero, recognizing that what lies between it and the first measured point is a guess (*cf.* Figure 11.4)?

This question is less important than it may seem for the following reason. If a prospective program *really* needs to know the Bragg peak upstream of the first true point, we have not done a good enough job measuring it. We should use less upstream scatterer, a thinner tank wall or less ‘dead’ water. But usually the application will have its own upstream material which guarantees that the Bragg peak is only accessed at depths where it is truly known. As a safety belt, SPLINE.FOR has some extrapolation capability. You can call it with a value somewhat less than the first table value and still get a sensible answer.

That said, I prefer to add a fake point so that the ‘measured’ Bragg peak starts mathematically at zero. BPW does this using an input parameter (typically 0.1) to control the fraction of data points used in a linear extrapolation, or zero to skip it. The procedure usually works well except for low energy Bragg peaks which may have appreciable curvature in the entrance region. But remember: you’re not really supposed to use that part of the Bragg peak.

11.5 Opening the BPK File

The BPK file is opened by using

```
CALL InitBragg(t1,fNucl,bLevel,dmin,d180,dpk,du80,dmax,bfn)
```

where `bfn` is a file name. `InitBragg` sets up a function `Bragg(x)` which returns the *effective mass stopping power* of the proton beam at depth x , as we will now explain. We assume that the file contains a ‘splined’ depth-dose measurement where the dose is relative but the depth in cm H_2O is known exactly.

11.5.1 Removing the Fluence

Eq. (3.7) relating dose to fluence and stopping power in practical units can be rewritten

$$\frac{D}{\Phi} = 0.1602 \frac{S}{\rho} \frac{\text{Gy}}{\text{gp/cm}^2} \left(\frac{\text{MeV}}{\text{g/cm}^2} \right)^{-1} \quad (11.1)$$

which is to say that if the stopping power is expressed in MeV/(g/cm²) then the left-hand side, dose divided by fluence, will be in Gray/(gp/cm²). The first step is to divide by the fluence. We do not know the *absolute* fluence but we know that it varies as 1/*r*² where *r* is given by the second line of the BPK file. Accordingly we divide by the relative fluence, in effect correcting for 1/*r*². The new numbers will have a larger peak to entrance ratio than before.

11.5.2 Adjusting the Energy

`InitBragg` gives us the option of either using the Bragg peak as it stands or of linearly shrinking or stretching the depth scale to adapt the Bragg peak to some desired energy. The latter case is useful for preliminary designs where we happen to have a Bragg peak at nearly, but not exactly, the right energy. The first option is exercised by calling `InitBragg` with `t1=0`. in which case `t1` will be returned at the value deduced from `d80` and a range-energy table. The second option is triggered by any non-zero `t1`.

11.5.3 Normalizing the Bragg Peak

The reason to normalize the Bragg peak (that is, convert it from a relative value to absolute Gy/(gp/cm²)) is that subsequent calculations using it will give us the actual number of protons required for the prescribed dose rather than just the relative number at each modulator step. This in turn allows the design program to make a reasonably good estimate of the incident beam current needed for the prescribed dose rate.

We do not know the stopping power at the peak, where the proton fluence is a complicated mix of low energies. We do, however, know it at the entrance. Recall that the Bragg peak is the convolution of the stopping power *vs.* depth curve for a single proton with a Gaussian representing the proton number *vs.* depth curve. At the entrance this convolution has a negligible effect because the stopping power varies slowly. Therefore, once we have deduced from `d80` that our Bragg peak corresponds to (say) 150 MeV protons at the first measured point, it follows that we can identify the relative dose at that point with the tabulated EM mass stopping power of water at 150 MeV. Eq. 11.1 says that if we multiply that by 0.1602 the resulting function will be the effective stopping power of our beam expressed in Gy/(gp/cm²). This argument ignores nuclear reactions.

However, by an equally good argument, once we have assigned that value to the first measured point, the *area* or integral under the Bragg curve starting with that point should equal 150 MeV, the average energy of a beam proton. In fact we will invariably find that the area is less than 150 MeV. That is because the area argument also ignores nuclear reactions.

So far we have given two different ways of normalizing the Bragg peak, which should be consistent in some sense once we take nuclear reactions into account. The next section will go into that more deeply. We will decide to use the entrance normalization for want of something better. In the larger scheme of therapy beam line design this is a minor issue. After all, what is at stake is merely the accuracy of our estimate of the required beam current. However, the

problem does appear to raise some concerns about our understanding of nuclear reactions. The reader who does not care would do well to stop here.

11.5.4 A Closer Look at Normalization

‘Entrance’ normalization to the tabulated EM mass stopping power ignores the absorbed dose contribution from nuclear secondaries which is truly zero only at the entrance, if the beam has not passed through solid or liquid matter. Otherwise, the nuclear contribution builds up in ≈ 1 cm to a value which can be estimated if we believe the ‘local energy deposition’ model of nuclear reactions (Chapter 8). Assume for the sake of argument we have 150 MeV protons at some depth where nuclear buildup is already established. The model says that $P' = 1.2\%$ of the primary fluence interacts every cm. For each reaction the entire 150 MeV goes into secondaries and the model says that $f \approx 60\%$ of this energy is deposited locally. Therefore, ignoring second order effects, $0.012 \times 0.6 \times 150 = 1.08$ MeV per average proton is deposited in the next cm so we must add $1.08 \text{ MeV}/(\text{g}/\text{cm}^2)$ to the $5.45 \text{ MeV}/(\text{g}/\text{cm}^2)$ EM stopping power. Of course we have also seen in Chapter 8 that this computation overestimates the observed nuclear buildup by a factor of two.

‘Area’ normalization also ignores nuclear interactions if it assumes that the area equals the average incident energy, because we know some of the energy that goes into nuclear secondaries is lost to the volume of interest. This calculation is slightly more complicated because the primary energy, hence the amount lost, varies from the initial energy to zero as the protons stop.

We can relate entrance and area normalization rather easily if we make the approximations that P' and f are constant, independent of energy. These are not very good approximations, but will give us a general idea of how things should go. Let N_0 be the number of protons incident and $N(z)$ the surviving number of primaries at depth z . To avoid constant factors of $\rho = 1$ let’s express z in g/cm^2 . Let $B(z)$ be the energy absorbed by water per *incident* primary per g/cm^2 , the *effective stopping power* of an average proton. Let S be the tabulated EM mass stopping power (normally S/ρ) and let $T(z)$ be the kinetic energy. Then at some depth z in dz we have

$$N_0 B dz = N S dz - f T dN$$

noting that dN is negative, or

$$B = \frac{N}{N_0} S - \frac{1}{N_0} \frac{dN}{dz} f T$$

Let $P(z; T_0)$ be the probability that an incident proton of energy T_0 will have a nuclear reaction by the time it reaches z .⁶ In other words

$$P(z) \equiv 1 - \frac{N(z)}{N_0} \tag{11.2}$$

and

$$P' \equiv \frac{dP}{dz} = -\frac{1}{N_0} \frac{dN}{dz}$$

and

$$B(z) = (1 - P(z)) S(z) + P'(z) f(z) T(z) \tag{11.3}$$

At $z = 0$, if nuclear equilibrium is already established,

$$B(0) = S(0) + P'(0) f(0) T(0) \tag{11.4}$$

⁶In our notation $P(r_0; T_0)$ is the function tabulated by Janni [3].

which, in symbolic form, is just the argument we made above. Now consider the energy obtained by finding the area of the entire Bragg peak:

$$T_A \equiv \int_0^\infty B dz = \int_0^{R_0} (1 - P) S dz + \int_0^{R_0} P' f T dz \quad (11.5)$$

There is some sleight-of-hand with limits here. We want to take the integral over the *entire* measured Bragg peak on the LHS, because the act of folding the stopping power with a Gaussian does not change the area. It merely spreads it out. On the RHS, however, we have mathematical functions that are only defined between 0 and R_0 , the mean range.

So far we have no assumptions about the z dependence of the quantities on the RHS of Eq. 11.3. We do not need to: we can always do the integral numerically. However, to get a simple result in closed form we approximate the range-energy relation by a power law

$$R = aT^b, \quad b = 1.77$$

so that

$$T(z) = \left(\frac{R_0 - z}{a} \right)^{1/b}$$

and

$$S(z) \equiv -\frac{dT}{dz} = \frac{1}{b a^{1/b}} (R_0 - z)^{(1/b)-1}$$

We also assume that

$$P'(z) \approx \text{constant} = 0.012 \text{ (g/cm}^2\text{)}^{-1}$$

so that

$$P(z) = P' z$$

and further, that

$$f(z) \approx \text{constant} = 0.6$$

The assumptions concerning P' and f amount to simplifications of the local energy deposition model of nuclear reactions. We can now do the integrals in Eq. 11.5 and we find

$$\boxed{\text{absorbed fraction} \equiv \frac{T_A}{T_0} \approx 1 - 0.639 P' (1 - f) R_0} \quad (11.6)$$

Let's briefly review the derivation. It starts with the assumption that, once we have determined normalized the measured Bragg peak at some point, its area (which we are then in a position to compute, and which has the dimensions of energy) should represent the total energy absorbed by water for a average proton. That can be estimated from the local energy deposition model, and we make a couple of not too bad approximations to get the result in closed form.

To test this result we can take the set of Bragg peaks shown in Figure 11.1, compute the area T_A of each starting with the first measured point, compute the energy T_0 at that point from d_{80} and plot T_A/T_0 vs. R_0 . The integral is made easy because we have fit the Bragg peaks with a cubic spline. Simpson's rule is exact for a cubic, so the integral is exact (that is, as good as the fit) so long as we use the spline points to delimit each integration interval. The assumption of full nuclear buildup at the 'entrance' point is validated by the appearance of Figure 11.1. Each scan starts after a good deal of material.

Figure 11.5 shows the results. We have used two extremes to normalize the Bragg peak. The open circles set the first measured point at S , completely

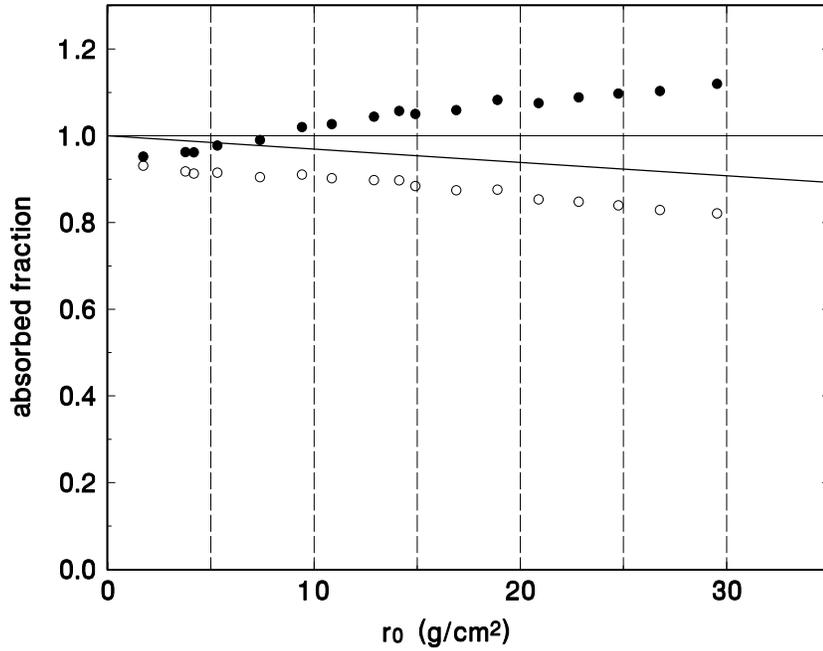


Figure 11.5: Absorbed energy fraction computed from the scans of Figure 11.1. Open circles: EM normalization. Full circles: EM + nuclear. Line: theoretical prediction for $P' = 0.012 / (\text{g}/\text{cm}^2)$, $f = 0.6$.

ignoring nuclear effects. The filled circles use Eq. 11.4, assuming full nuclear buildup. Note that it is the *filled* circles that should follow the line, Eq. 11.6. Instead, they rise to a nonphysical value great than 1. At present we are at a loss to account for this result. It is almost certainly the same effect noted earlier [47], that the observed nuclear buildup is $\approx 2\times$ less than predicted by nuclear models. Reducing the nuclear component at the entrance improves the agreement, and at least brings T_A/T_0 below 1.

Understanding nuclear effects in the Bragg peak quantitatively should be an interesting topic for further investigation. However, as noted before, it is not a serious problem in designing of beam spreading systems. For the time being we shall use S normalization, ignoring nuclear effects.

Chapter 12

Designing a Modulator

12.1 Introduction

A *range modulator* is a rotating device that moves graduated thicknesses of plastic through the proton beam, thus scanning the Bragg peak in and out to create a spread-out Bragg peak (SOBP). In its simplest form it can be fabricated by stacking sheets of plastic. Figure 12.1 shows a modulator wheel that was used for many years at HCL. The thinnest part has no plastic at all; it simply lets the beam through. This modulator was used downstream in the proton nozzle. It was quite large and heavy since it needed to cover the treatment field. Proton scattering in the modulator had little effect because of the short throw to the patient. We'll call this a 'downstream' modulator.

Alternatively, it is possible to put a modulator upstream where, in addition to creating the SOBP, it acts as an element of the beam spreading system. Scattering then becomes very important. We can control it, up to a point, by adding brass or lead to each step in such a way that the water equivalent pullback is what we want for that step and the multiple scattering angle is also what we want. We'll call this an 'upstream' or 'compensated' modulator. Figure 12.2 shows an early upstream modulator made of brass and plastic on a numerically controlled milling machine. An upstream modulator can either be used by itself in a single scattering system or as the first scatterer in a double scattering system.

Upstream, the modulator can be much smaller because it only needs to cover the incident beam, not the treatment field. There is also a second, less obvious, advantage. Moving the modulator upstream improves the transverse penumbra. The shadow cast by the patient specific collimator becomes sharper, allowing us to create sharper dose distributions. This is because a downstream modulator (or any beam modifying material placed near the patient) behaves like a diffusing screen, enlarging the effective proton source. We'll discuss this in detail in Chapter 18.

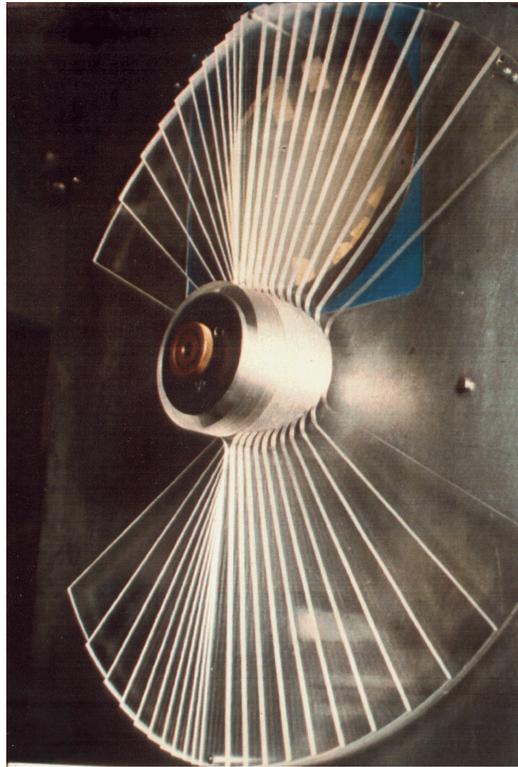


Figure 12.1: A downstream modulator.

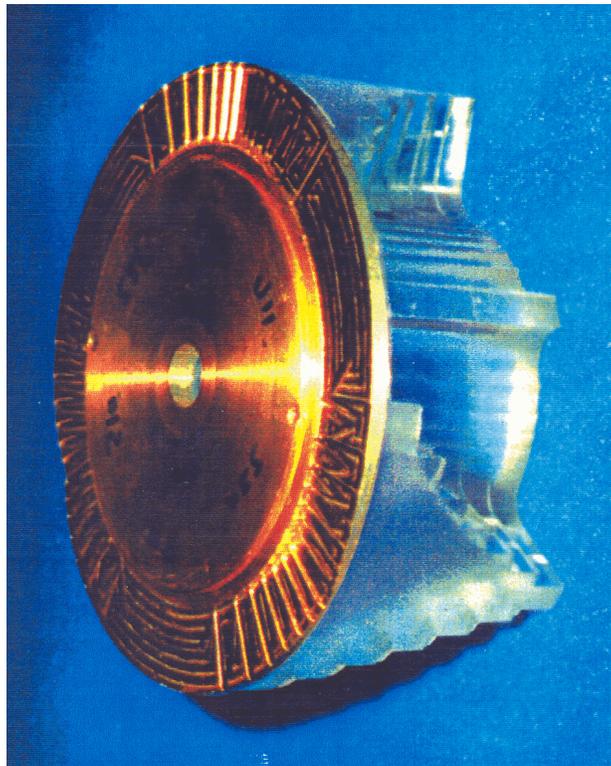


Figure 12.2: A compensated upstream modulator.

Historically, the idea of a range modulator was already proposed by R.R. Wilson in his seminal paper [63]. Design techniques, construction and experimental results were presented by Koehler et al. [64] in 1975. They computed the amplitude of each step (the amount of time each thickness spends in the beam) by iteration. The hands-off method of simultaneous linear equations we will use was developed by the author, and independently by a number of others. The realization that an upstream modulator was possible and desirable, and some proof of principle experiments, came at HCL in the late 1980's [65]. The NPTC proton nozzles make use of upstream modulator single and double scattering systems.

Because the range modulator is a rotating device, beam time structure must be considered. If the beam duty factor is fairly large and the treatment time is on the order of a minute, there is no problem rotating the modulator fast enough so that a treatment corresponds to many modulator cycles. In that case, any minor fluctuations in the beam average out. However, if the duty factor is very small and beam bursts are very far apart, the beam effectively takes a series of 'snapshots' of the modulator, and a rotating device becomes virtually useless. A case in point is the ITEP synchrotron in Russia, whose fast extraction system produces something like a microsecond of beam every one or two seconds. The *ridge filter* is a static device consisting of a degrader whose thickness varies spatially. Ridge filters are discussed in detail by Chu et al. [8], and Nichiporov et al. [66] have recently published an elegant variant of the idea. It is difficult to fabricate ridge filters for very large modulation (say more than half the depth), and of course the technique of beam gating, which allows one modulator to be used for any modulation, is inapplicable. Though ridge filters are certainly useful in some situations, we will not discuss them further.

In either the upstream or downstream location, it is important to realize that the modulator need not be so large that only one step is illuminated! In other words, beam can pass through more than one step at a time. Downstream, that simply means that different depths of the target volume will be treated at slightly different times depending on their transverse position. Upstream, not even that happens, because the beam tends to get scrambled by multiple scattering before it reaches the target.

Designing a modulator involves

- Computing the *step size*, the amount of pullback for each segment.
- Computing the *weight* of each segment, the time spent in the beam.
- For an upstream modulator, computing the mixture of high-Z and low-Z material to be used.

The last has been covered in Chapter 9. As to the first, one cannot go far wrong if one uses a fixed step size whose water equivalent is equal to the width of the Bragg peak at the 80% level. Assuming this for the moment, we will begin by computing the weights. Some fine points regarding step size will be discussed later.

12.2 Computing the Weights

We assume that, at the time of modulator design, the scattering system (single or double) has been designed. In other words, the scattering required from the modulator is known (it may be zero in the case of a downstream modulator) and the fluence per proton Φ can be calculated on axis at any depth in the water tank.

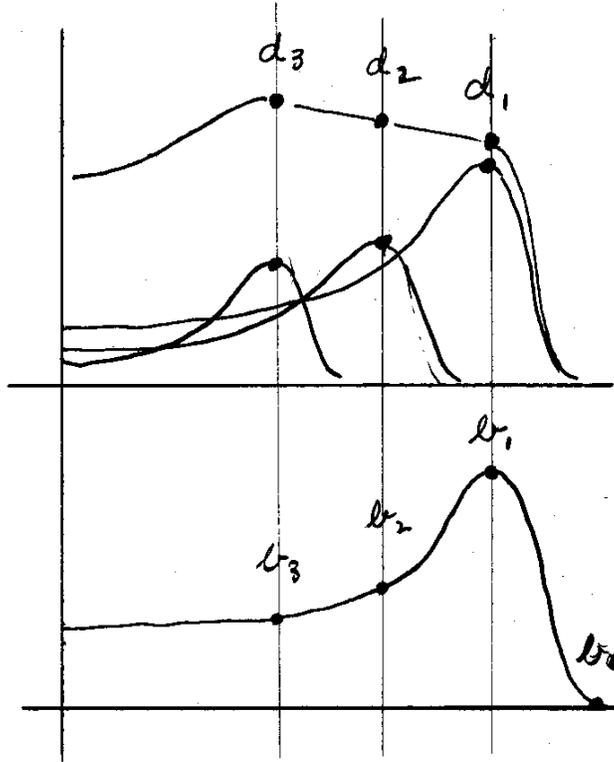


Figure 12.3: Figure illustrating modulator weight calculation.

We will use the following notation:

M	# modulator steps = # z positions
$i = 1 \dots M$	position index
$j = 1 \dots M$	modulator step index (1 is deepest step)
z_i	a measuring plane position along beam axis
b_i	value of pristine Bragg peak at position z_i (effective stopping power in Gy/(gp/cm ²))
Φ_{ij}	fluence per proton on axis; depends on i via location of measuring plane and on j via scattering effects (modulator thickness, proton energy)
d_i	desired dose at position z_i
N_j	# gp (gigaprotons) required for step j

A 3-step modulator will suffice to illustrate the math. Refer to Figure 12.3. The desired doses d_i are usually the same (say 1 Gy) but they need not be. The dose at each point is the sum of contributions from three Bragg peaks:

$$\begin{aligned}
 d_1 &= (b_1 \Phi_{11}) N_1 + (b_0 \Phi_{12}) N_2 + (b_{-1} \Phi_{13}) N_3 \\
 d_2 &= (b_2 \Phi_{21}) N_1 + (b_1 \Phi_{22}) N_2 + (b_0 \Phi_{23}) N_3 \\
 d_3 &= (b_3 \Phi_{31}) N_1 + (b_2 \Phi_{32}) N_2 + (b_1 \Phi_{33}) N_3
 \end{aligned}$$

b_0 , the value of the Bragg peak one step up from the peak, is small and b_{-1} is almost certainly zero but it does no harm to include them. The known quantities in parentheses are the elements of a matrix \mathbf{G} and the system of linear equations can be written formally

$$\vec{d} = \mathbf{G} \vec{N}$$

where \vec{d} is a known vector and \vec{N} is a vector to be found. The formal solution

$$\vec{N} = \mathbf{G}^{-1} \vec{d}$$

can be computed by a subroutine such as `GaussJordan`. See Numerical Recipes for a description of ‘Gauss-Jordan elimination with full pivoting’ as well as some slightly faster techniques that could be used.

Although we traditionally call the N_j ‘weights’ implying a relative quantity, note that by virtue of the definition of Φ as the fluence at z_i *per proton incident on the scattering system* and of b_i as the effective stopping power expressed in Gy/(gp/cm²), the N_j are the actual *numbers of incident gigaprotons for each step* required to obtain the d_i . In other words, the design program tells us not only the shape of the modulator but (by a trivial additional calculation) the incident proton current required for the prescribed dose rate. This estimate is, however, only as good as the normalization of b (Chapter 11).

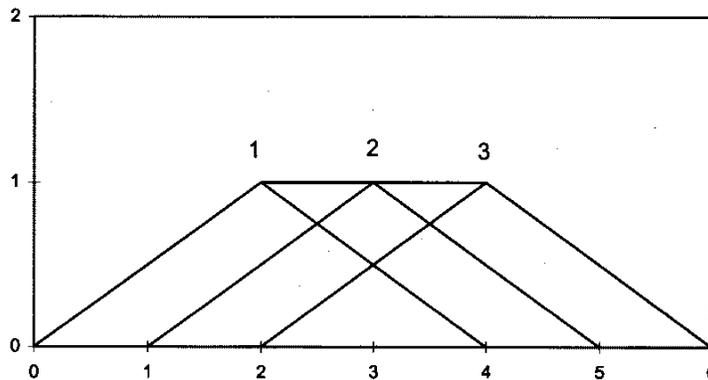


Figure 12.4: Prieels’ argument for finite step size.

12.3 Alternative Procedure for Weights

It might reasonably be argued that we could achieve a better result by requiring the dose to be flat¹ everywhere or at a large number of points rather than only M points. This, and the closely related question of whether we want a ‘stepped’ or ‘continuous’ modulator, was investigated by Prieels [67] using an optimization procedure instead of solving a set of linear equations. Prieels suppressed nonphysical solutions by adding a penalty for negative weights to the objective function. This works better than a hard constraint. He was able, for a given problem, to obtain two solutions depending on the starting point, one corresponding to a continuous modulator and the second to a stepped modulator.

Surprisingly, the SOBP was slightly smoother for the stepped case. Prieels gave a simple heuristic argument for this. Figure 12.4 from [67] shows that, if we try to create a flat SOBP from triangular ‘Bragg peaks’, we will get the desired result by using peaks 1 and 3 and that any intermediate peak such as 2 must be rejected. If we use real Bragg peaks and the algebraic method of the previous

¹ From here on we’ll assume unless explicitly stated that the goal is a *constant* dose over some extent in depth.

section, a too small step size will yield negative N 's and a too large step size will yield undulations or 'wiggles' in the SOBP.²

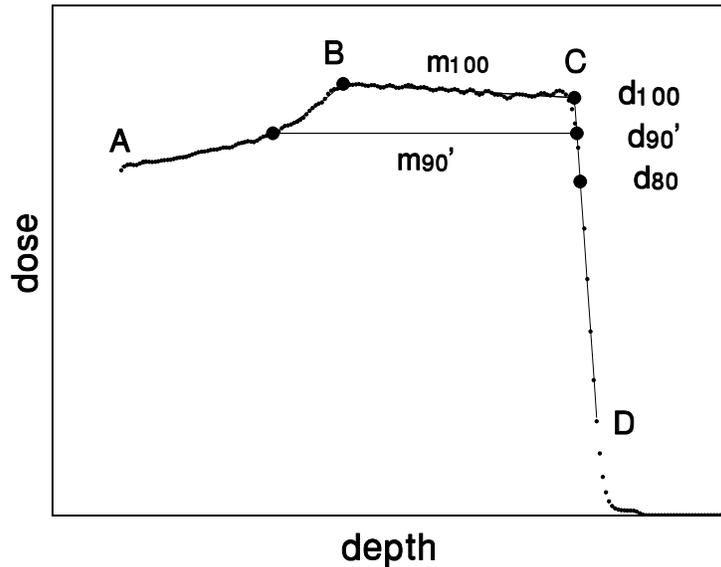


Figure 12.5: Model SOBP defining measures of depth d and modulation m .

12.4 Defining Modulation

We have gone this far without defining what we mean by the ‘modulation’ m of a given SOBP and, for that matter, by the ‘depth’ d . d and m are two of the quantities transmitted from the treatment planner (dosimetrist) to the physicist or operator setting up the hardware. We must not expect too much from their definitions. The planner needs to know far more about the dose distribution than those two numbers: the falloff in every direction, the entrance dose and so forth. Nevertheless it is worth trying to come up with the most convenient definition.

We have discussed this at some length in a recent paper [22]. Figure 12.5 shows the gist of that paper, which argues that the corners B and C are the most important features of the SOBP. We explain a way of finding those corners mathematically by fitting the SOBP with three polynomials describing segments AB, BC and CD.³ The 100% level is then defined as the mean dose between B and C, and m'_{90} is defined as the distance between the intersections, with the SOBP, of a horizontal line drawn at the 90% dose level. The prime in our notation means that the fiducial line is horizontal. m_{90} would be measured using a line that follows the slope (if any) of the SOBP. The quantity we have called m'_{90} is used *de facto* in most treatment centers as ‘the modulation’.

The problem is that m'_{90} is undefined when the SOBP is wide enough that the entrance dose exceeds 90%. In [22] we advocate using m_{100} , which is defined for any SOBP. However, m_{100} has a problem of its own for very narrow SOBP's, which are frequently used in radiosurgery: for instance, early pituitary

² When the motor generator set at HCL finally quit and was replaced by a solid state power supply, the energy spread of the beam decreased slightly. Suddenly all our SOBP's had larger wiggles because the Bragg peak was a bit narrower compared to the step size.

³ We have since replaced the polynomial fit by a simpler ‘broken spline’ fit (Chapter 5).

treatments at HCL used converging pristine Bragg peaks. The only sensible mathematical value we can assign to the m_{100} of a pristine Bragg peak is zero. If ‘modulation’ is intended to convey the extent in depth of the target volume that receives clinically significant dose, calling it zero for a pristine Bragg peak does not do very well. m_{90} is better in this case.

Perhaps the solution is to measure the distance between the distal 90% point and a proximal level that varies from 90% to 100% according to some predefined function, presumably a quadratic. Then the modulation would be defined for any SOBP but would have to be characterized by two numbers *e.g.* $m_{97/90}$ much as we characterize penumbra or distal falloff. On the other hand, this may strike some as too complicated.

Eventually this matter may be settled by agreement between particle therapy centers, and we will not belabor it further here. One step forward that we will take, and urge upon others, is to use notation that says precisely what we mean. In the rest of this chapter we will use m_{100} because the number of steps is far more simply related to it than any other parameter. If m'_{90} is specified clinically it is a simple matter to generate a lookup table or curve that gives us the corresponding m_{100} or to perform the conversion in the design program.

12.5 Step Size

An empirical rule we have already mentioned is that the optimum step size equals the width of the Bragg peak at the 80% point, call it s_{80} . If it is greater, you will get wiggles in the SOBP. If smaller, you will increase the distal falloff and, if much smaller, you will begin to see negative weights in the modulator design.

That said, there is a good reason for making it *slightly* smaller if we want to match the prescribed modulation very closely. The problem is that the number of steps is integral. If $s_{80} = 8\text{mmW}$ we will not be able to match a prescribed modulation to better than $\pm 4\text{mmW}$ in general. The solution is to design a modulator with s_{80} choosing the number of steps M such that the modulation will come out a little too large. We then decrease the step size proportionately to the error in modulation, keep M the same, and repeat the design. This procedure yields m equal to its desired value $\pm 1\text{mmW}$.

We would need a large library of prefabricated modulator wheels to cover any modulation to $\pm 1\text{mmW}$. On the other hand, that is probably not needed clinically, particularly for large modulation. NPTC uses two complementary schemes. In the gantry rooms we use beam gating, which has infinite resolution in m . In the planned neurosurgery beam we will, for simplicity, not use beam gating. However, in that beam the SOBP is made by ‘lamination’ or ‘energy stacking’. Instead of using a modulator, we generate the SOBP one layer at a time by selecting plastic and lead degraders from binary weighted sets. That yields a resolution in m and d equal to the water equivalent of the thinnest plastic degrader. We compute the plastic and lead by designing a virtual modulator, then converting each step into plastic and lead (details in Chapter 13).

12.6 Compensation

Assume that the modulator is the first scatterer in a double scattering system. Further, assume the input energy is given, the scattering system is designed (you know the scattering angle required of the modulator), and the modulator is designed (you know the water equivalent thickness of each step). You also know the number of protons to be delivered through each step but that is

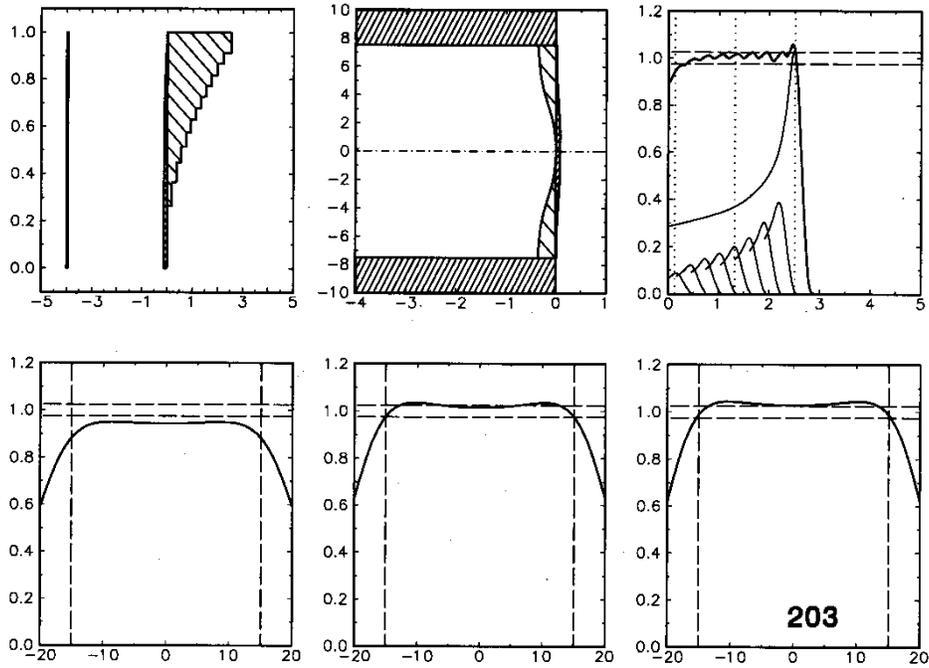


Figure 12.6: Performance of a 70-82 MeV system at 70 MeV.

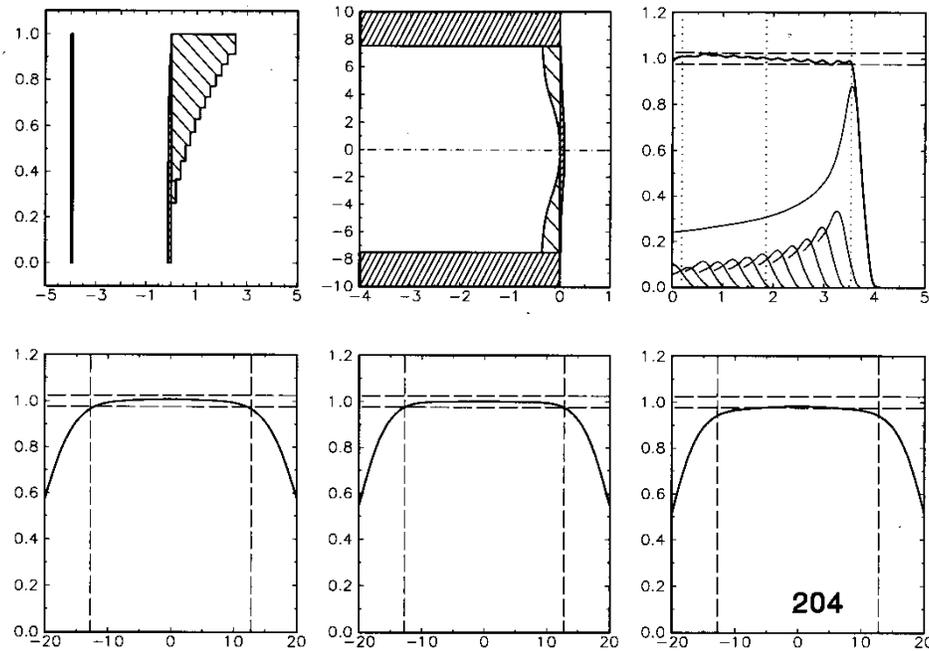


Figure 12.7: Performance of a 70-82 MeV system at 82 MeV.

irrelevant at this time. Your task is to find the combination of lead and lexan (for instance) that will work properly. Do this one modulator step at a time, using the methods of Chapter 9, as follows:

- From d_{80} of that step, find the proton energy T into the tank.
- Backtrack through the second scatterer using its lead and lexan thicknesses on axis to find T into the second scatterer.
- Ditto through any ‘post absorber’ material between the modulator and the second scatterer.
- You now know the energy out of the modulator. With this, the energy into the modulator, and the scattering angle, you can compute the lead and the lexan.

Having done this for each step you will find, in the ideal case, that step 1 (the thinnest) is all lead, step M is all lexan, and every other step has some lead and some lexan. That is the design goal. We call it a ‘minimal’ modulator. Far more likely, you will see one of these pathologies:

- (Harmless.) The modulator carries extra lexan. Remove lexan from the track or wheel and use a fixed absorber or, better, reduce the beam energy.
- (Harmless.) The modulator carries extra lead. Remove lead from the track or wheel and use a fixed scatterer.
- (Serious.) Steps before the end (before step M) have no lead.

Regrettably, the last is common. By implication it means that the required lexan already overscatters. (You now recall that you got an error message ‘ N failures in compensation’.) You can fix this by decreasing the throw (which may not be an option) or increasing the prescribed radius (which is undesirable) because either move *requires* more scattering. Fortunately, you can always put up with a few uncompensated steps. The depth-dose on axis will be correct, because it is constrained to be, and the effect off-axis may be tolerable because the relative weights of the last few steps are small. Too much of this, however, and you will see the dished transverse scan characteristic of too strong a first scatterer.⁴ You have reached the natural limit of the upstream modulator.

12.7 Useful Energy Range

A physical modulator wheel, when it is the first scatterer in a double scattering system, will only work satisfactorily over a certain range of incident energies. Below the design energy, the modulator overscatters and the SOBP acquires a positive slope. Above the design energy, the modulator underscatters and the SOBP acquires a negative slope. Figures 12.6 and 12.7 from [68] illustrate these effects. The energy range, 70-82 Mev, can be taken as typical of what can be achieved, though the exact range will depend on your tolerances. Beam current modulation (Section 12.9) can extend the useful range greatly, at the expense of a more complicated system and safety tests.

A second energy effect is the narrowing of the Bragg peak at lower energy. If the modulator is designed at the high end using s_{80} , the SOBP will have wiggles at the low end.

⁴ For mild examples of what we mean by ‘dished’ and ‘domed’ see Figures 12.6 and 12.7 respectively.

If a modulator is to cover a finite energy range, at what energy within that range should it be designed? The wiggle problem would be solved by designing the modulator at the low end with s_{80} . However, we would then not have full modulation at the high end and would have to come up with some procedure for adding steps by hand. It seems better to us to design at the high end using a step reduction factor so as to ensure a small enough step at the low end.

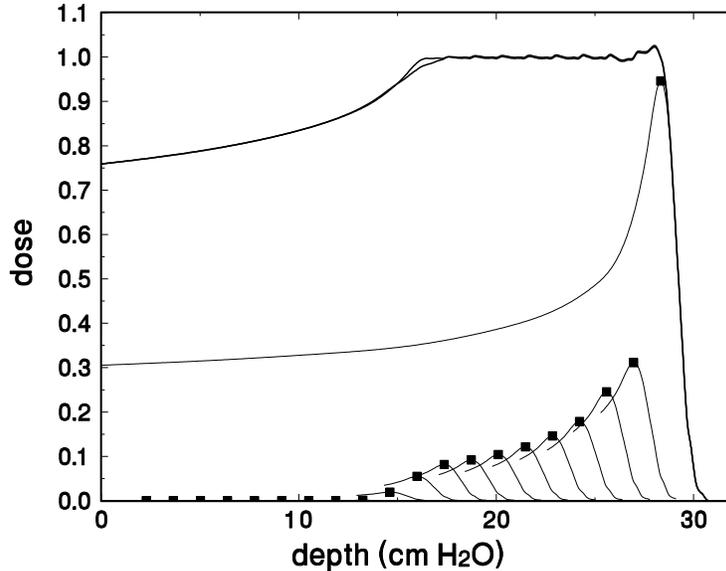


Figure 12.8: Beam gated SOBP showing rounding of proximal corner.

12.8 Beam Gating

If you have a variable energy proton generator you will want to adjust the beam energy into the nozzle to suit the prescribed depth of each field. You will therefore need a library of modulators, one to deal with each energy range.⁵ Beam gating allows you to avoid a second dimension of this library by obtaining any prescribed modulation from a single modulator designed for full modulation. You simply cut off the cyclotron beam when the desired proximal depth is reached.

A modulator wheel or ‘track’ designed for gating must have a *beam block* section, as was first pointed out by D. Prieels. The reason is that, for any practical sized design, the beam spot will inevitably illuminate more than one step. Specifically, there will be a time during each revolution when appreciable beam is on both the thickest and the thinnest step. If we cut off the beam at this time to trim the most proximal step we will inevitably cut into the most distal step, spoiling that end of the SOBP. The beam block cleanly separates the two ends of the modulator.

The fact that more than one step is illuminated at a time produces a proximal corner that is somewhat rounded. This is illustrated by Figure 12.8 from [69] which compares a simple and a gated modulator using beam and track sizes typical of NPTC.

⁵ The original NPTC design called for eight modulators to cover the range 70-230 MeV. This can be cut to five or so by using beam current modulation to correct the SOBP.

With gating, the average dose rate for a given beam current is lower than would be gotten with a simple modulator because the beam is off part of the time. This is exacerbated by the need for a beam block, but is still a minor disadvantage.

A more important disadvantage is the fact that, with gating, we have a somewhat complicated system with significant potential for delivering the wrong SOBP if something fails. We need redundant, radiation resistant means of knowing the modulator angle with sufficient resolution. Timing should be done in hardware because millisecond level timing in software is unreliable on a large multi-tasking computer. With a simple modulator, we need merely confirm that it is rotating at something like the right speed. By contrast, beam gating requires a higher level of engineering.

12.9 Beam Current Modulation

At a still higher level of complexity, we can change the cyclotron output current in synchronism with modulator rotation to correct defects in the SOBP. At NPTC this is done [70] under the control of a beam current modulation (BCM) file which specifies the current at each modulator angle. We will not go into the details of how the BCM file is computed from the SOBP to be corrected.

Cyclotron output current is measured by an ionization chamber (in which recombination and radiation damage are potential problems) and controlled by means of the cyclotron ion source current. The relation between ion source and output is non-linear and changes from day to day depending on machine settings. A stand-alone feedback circuit, which needs to be ‘trained’ from time to time, is used to linearize it over some range. That range determines how much correction can be obtained.

12.10 Physical Arrangement of Modulators

In a gantry based nozzle we want to change modulators remotely, for convenience and to minimize radiation exposure to staff. Various layouts can be imagined. At one extreme we might have a number of independent modulators arranged on the periphery of a large wheel which is turned to bring the desired modulator into play. Or, one might have a single large wheel with (say) nine modulator tracks cut into it. That turns out to be very large indeed when one considers the sideways motion required, and rather inflexible. Miles Wagner proposed the elegant ‘3 × 3’ scheme used by IBA (Figure 12.9). Three small wheels, each cut with three tracks and driven by its own motor, are mounted on a large wheel. A simple rotation of the large wheel can bring any desired track into the beam. The small wheels are modular and fairly easy to change, though that has not proved necessary at NPTC.

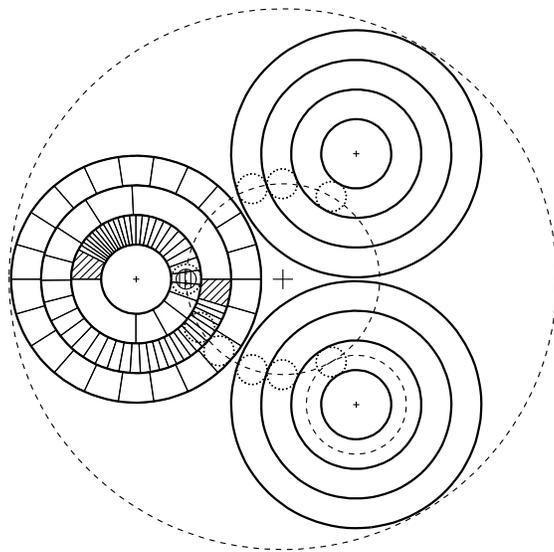


Figure 12.9: The 3×3 arrangement of modulator wheels.

Chapter 13

Single Scattering Energy Stacking Example

This chapter will describe how LAMINATE uses single scattering and a set of binary weighted lead and lexan degraders to obtain a prescribed SOBP (depth and modulation). It also explains how to estimate the beam monitor output factor.

Chapter 14

Double Scattering

14.1 Introduction

A double scattering system does *not* merely consist of two uniform scatterers, since any number of uniform scatterers can be combined into a single effective scatterer, which still produces a Gaussian scattering distribution. A double scattering system combines a uniform first scatterer with a second scatterer that is non-uniform in such a way as to produce a flat or nearly flat dose at the patient. Double scattering systems have much better efficiency than single scattering: they put more protons into the flat dose region. At the same time, they waste less proton energy. Three basic schemes have been devised for the second scatterer.

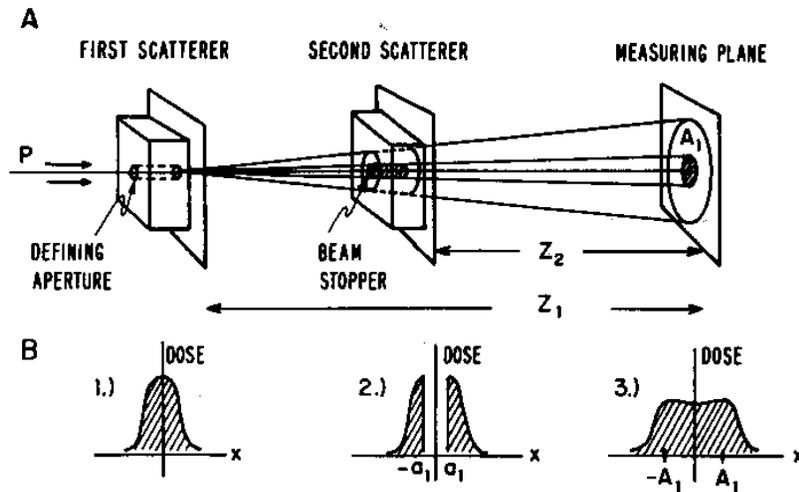


Figure 14.1: Double scattering with a beam stopper [64].

The *occluding ring* method was first used by Koehler et al. in 1973 [64].¹ As Figure 14.1 from [64] shows, a uniform second scatterer is immediately preceded by a cylindrical beam block centered on a cylindrical defining aperture. The beam block punches a hole in the Gaussian, which is filled in by the second scatterer so that, at the patient, the dose distribution is fairly flat.

¹ Interestingly, results from the first magnetic 'wobbling' system are also described in this paper. A rotating permanent magnet was used to sweep the proton beam in a circle. This technique was, however, never used for treatments at HCL.

The central beam block is merely the first of a family of similar methods: block, annulus, block plus annulus and so on, of which [64] describes the first two in detail. With more occluding rings, the second scatterer can be made thinner, preserving energy, but the point of diminishing returns is rapidly reached and the rings become difficult to fabricate and align. The design problem is to specify the first and second scatterer as well as the inner and outer radii of all the open zones in front of the second scatterer: two zone radii for a block, three for an annulus, four for a plug plus annulus and so on. Although [64] fully solves the forward problem (computing the dose distribution given all the design parameters) it is silent on how the parameters were finally optimized.

Reference [64] gives details and experimental results for three proton ‘nozzles’ using the annulus method. Suffice it to say here that the efficiency for $\pm 2.5\%$ uniformity into a 30 cm diameter field (the ‘large field’ nozzle) is $\approx 30\%$ (six times what it would be for a single scattering system) and the range loss is 0.6 cm out of 17.2 cm, far less than it would be for single scattering. The large throw (≈ 6 m) of the beam line is very helpful in preserving range, because less scatterer is needed no matter what the method.

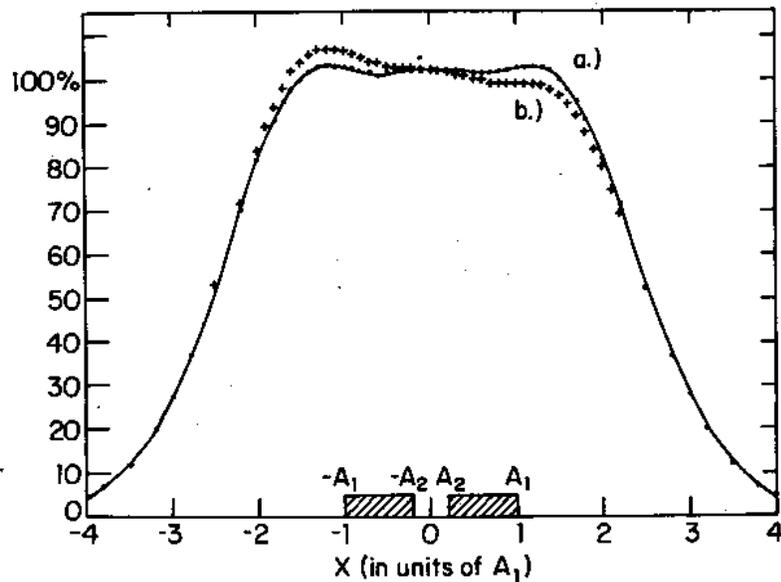


Figure 14.2: The effect of beam steering on double scattering.

Beam misalignment is a potential problem for all double scattering systems. It is the *centering* of the beam on the second scatterer, regardless of the beam direction, that matters most. Figure 14.2, taken from [64], shows the computed effect of a beam offset equal to 5% of the annulus outer radius (2 mm for the large field nozzle). It agrees with experiment.

Energy compensation is not a problem with the occluding ring system. All protons either stop or pass through the same amount of scatterer.

A *dual-ring* double scattering system was described by Takada in 1994 [71], and in 2002 he went on to include the effects of a non-ideal beam [72]. Figure 14.3, taken from [72], shows the method. The second scatterer is formed of two rings designed for equal energy loss, which takes care of energy compensation. The inner, stronger, scatterer produces a central Gaussian fluence in the measuring plane while the outer scatterer produces a ring-shaped fluence. The

two combine to form a nearly uniform total fluence.

The dual ring can be thought of as a two-step discretization of the compensated contoured scatterer to be described next. One could probably extend the approach to three or more rings made of appropriate materials, obtaining a slightly flatter dose. However, it is not clear there is any advantage over the contoured scatterer.

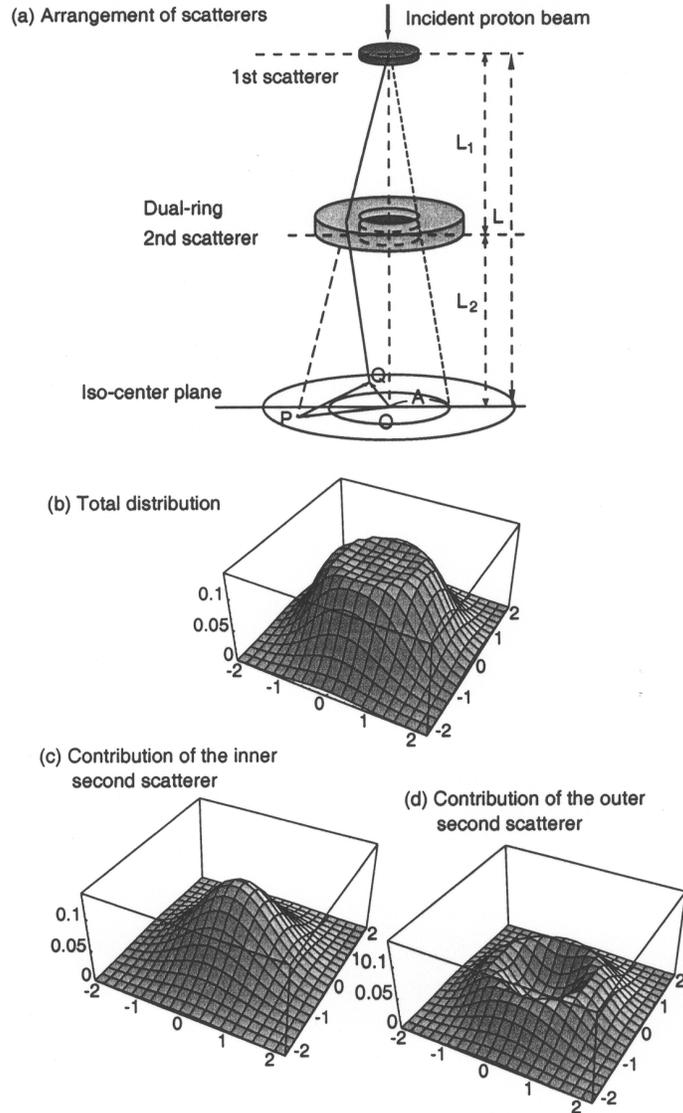


Figure 14.3: Double scattering with dual-ring scatterer [72].

14.2 The Contoured Scatterer

The contoured scatterer was devised at Uppsala in 1988 by Montelius and Brahme [73] as an offshoot of flattening techniques in electron radiotherapy, and independently by us at HCL. The Uppsala group published a long account [74] in 1994. Figure 14.4, taken from [74], shows the principle. The second scat-

terer scatters the central beam more strongly and outside beam less so, yielding a flat fluence at the measuring plane if the contour is designed appropriately.

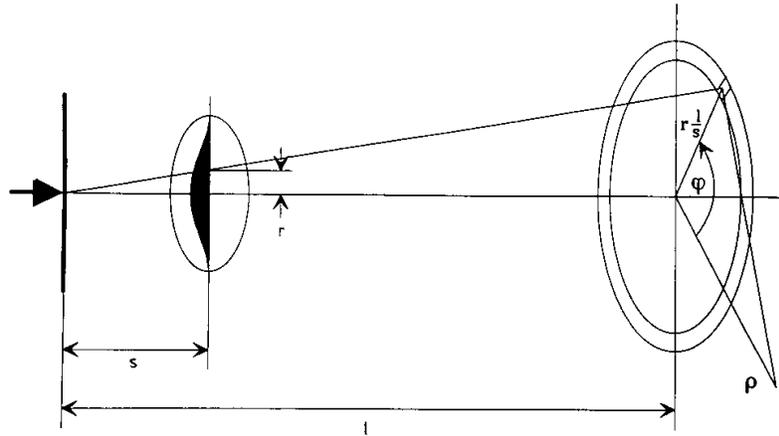


Figure 14.4: Double scattering with contoured scatterer [74].

Though reference [74] reads quite differently from our description below, the differences are mainly superficial. The authors use the method of *projection onto the measuring plane* to derive their equations (as did Koehler et al. [64]). They speak of a *universal radial profile* (we will call it the ‘generic shape’) which needs only to be scaled in order to apply to a given physical setup (defined by incident energy, throw and desired radius of the flat field). Their choice of scaling variable (strength of the first scatterer) is, however, different from ours (radius over which the optimization is carried out). Their universal profile is a nine-step function based on a deformed cosine, whereas we will use a cubic spline with three or four points adjustable in y .

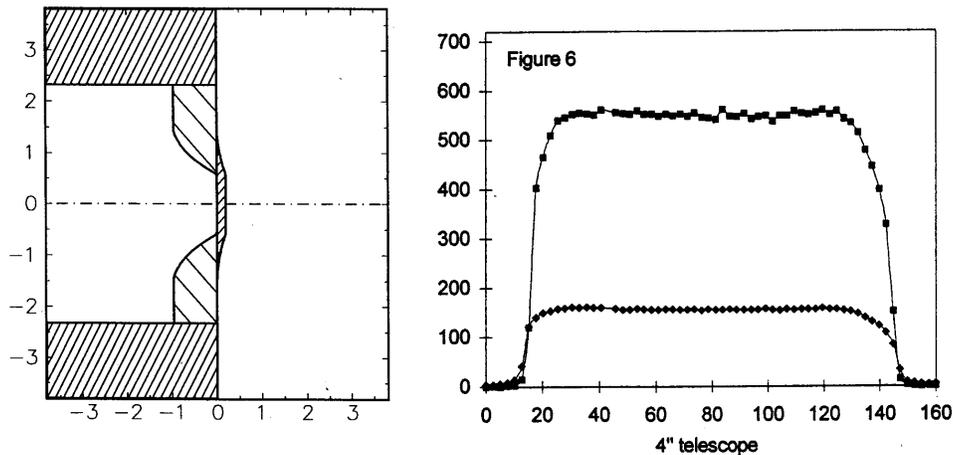


Figure 14.5: Left: second scatterer with Siebers-type profile [75]. Right: cross-field scans [76] at entrance and at Bragg peak.

The claim [74] of a universal profile which is globally optimum for minimum total energy loss should be taken with a grain of salt. One reason for saying this is that, in 1992, Siebers and Miller [75] suggested chopping the nose off the bell-shaped profile. Figure 14.5 shows a compensated contoured scatterer

based on this idea and used in the HCL radiosurgery beam [76]. Because of the discontinuity, the computed fluence (not shown) shows some ripples, but these are so slight as not to show up in the measured cross-field scan.

The second reason has to do with compensating the physical second scatterer for constant energy loss independent of radius, which is mentioned by [74] only in passing. Actually, it is essential in most applications, but we are limited by physical materials. It is easy (with the aid of LOOKUP) to show that lead and lexan scatterers, each of which has the same energy loss in the few-MeV range, will differ in scattering by a little over $4\times$. Therefore the dynamic range of a universal profile must be less than $4\times$ if we are to compensate it, and that proposed in Figure 3 of [74] is considerably greater.

Incidentally, Figure 14.5 shows the close relation of the contoured scatterer to Takada's scheme, the only difference (which is fairly important) being the continuous transition between the two rings, which makes for a flatter dose.

14.3 Beam Imperfections

All the authors cited thus far consider the effect of beam imperfections. Koehler et al. [64] compute the effect of mis-centering of the second scatterer (equivalent to mis-centering of the beam on the second scatterer) on a pre-designed system. Grusell et al. [74] and Takada [72] go further, actually allowing beam parameters to influence the design. Takada's treatment is comprehensive and the resulting formulas are rather complicated.

Our approach, following Koehler, will be to ignore beam imperfections in the design stage. The beam will be regarded as a single ray of protons perfectly aligned with the nozzle axis. The effect of beam imperfections (finite emittance and misalignment) is best studied separately in a Monte Carlo program, and we will devote a separate chapter to that. The most serious imperfection is mis-centering on the second scatterer, which leads to a tilted dose distribution (Figure 14.2) in any variety of double scattering system. It can only be corrected by moving the second scatterer or the beam. The only imperfection that can easily be compensated in the design is the angular divergence of the beam at the first scatterer, which can be fixed by reducing the first scatterer strength or moving the first scatterer slightly downstream.

14.4 Solving the Forward Problem

Our first task is to find the fluence function in an arbitrary measuring plane (MP) *given* the parameters of a double scattering system. Only then can we hope to adjust the parameters to produce a favorable fluence function. Figure 14.6 shows the gross geometry. z is the coordinate along the beam line. The first scatterer is located at z_0 . If it is a compound scatterer, z_0 is its effective location (Chapter 10). If it is a modulator, z_0 will depend on the step number and the following calculation must be done for each step. The second scatterer is located at z_2 , and the MP at z_m .

We now shift our attention to the *projection* of the double scattering process on the MP (Figure 14.7). We shall set up the problem so as to handle both the occluding ring system and the contoured scatterer. \vec{t} is a vector arising from the displacement of the beam from the axis. Its precise relationship to the actual beam misalignment is confusing, but need not concern us here, because we only include it for compatibility with subprogram CSI.FOR (Contoured Scatterer Integral). We will set $\vec{t} = 0$ for all design work.

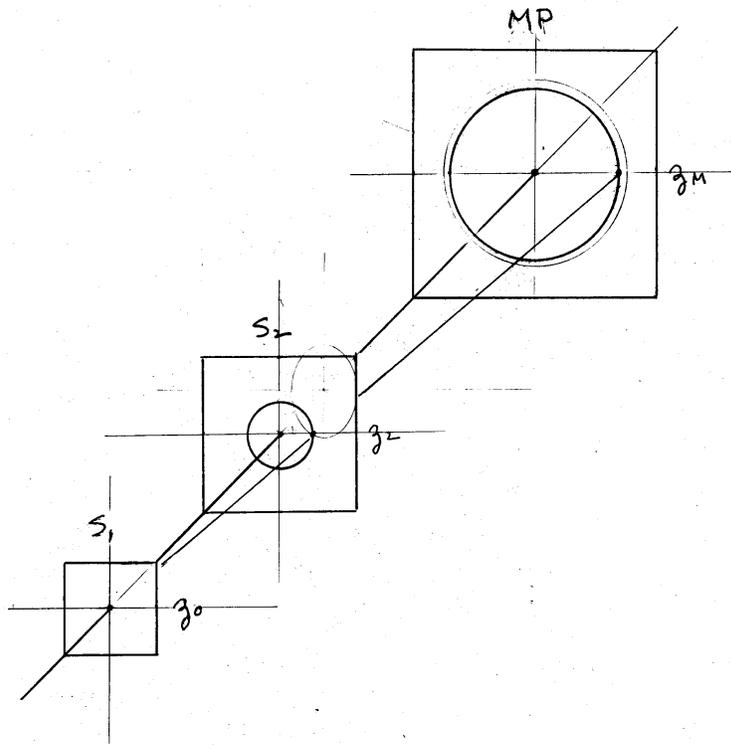


Figure 14.6: Geometry of double scattering system.

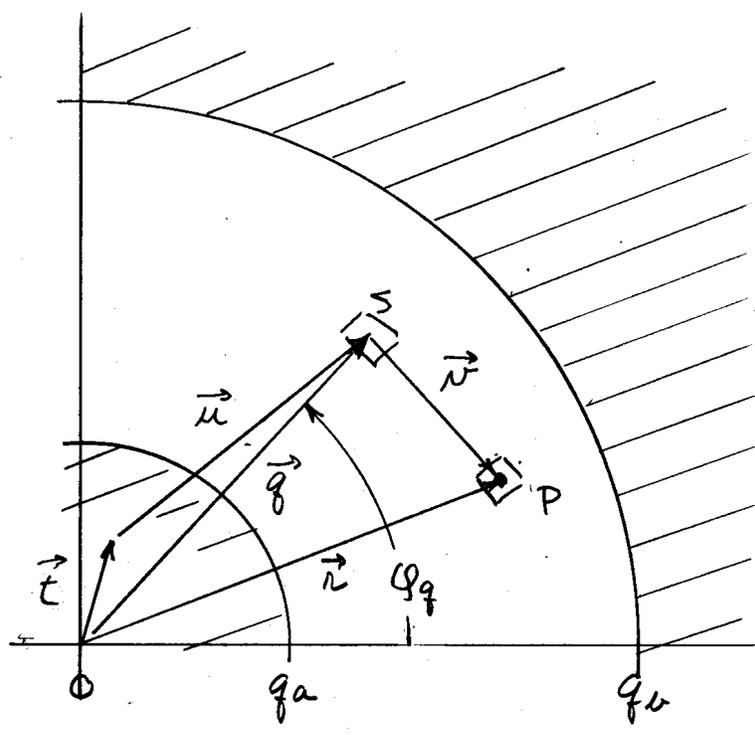


Figure 14.7: Double scattering projected on the measuring plane.

\vec{u} is a vector representing the projected first scatter. It is distributed in 2D according to a Gaussian or Molière function with a constant characteristic width u_0 , and related to the actual first scattering angle θ_1 by

$$u = (z_m - z_0) \theta_1, \quad u_0 = (z_m - z_0) \theta_{01} \quad (14.1)$$

\vec{v} is a vector representing the projected second scatter. It is distributed in 2D according to a Gaussian or Molière function with a characteristic width which may either be constant (occluding ring case) or a function of the radius q of the projected second scattering point. \vec{v} is related to the actual second scattering angle θ_2 by

$$v = (z_m - z_2) \theta_2, \quad v_0(q) = (z_m - z_2) \theta_{02}(q) \quad (14.2)$$

There will also be one or more significant radii at the second scatterer that project to radii in the MP. For instance, if r_{max} is the radius of the hole in the second scatterer holder, its projection q_b is related by a magnification factor

$$q_b = \frac{z_m - z_0}{z_m - z_2} r_{max} \quad (14.3)$$

as though produced by a point light at the first scatterer. Similarly, q_a in Figure 14.7 could be the projection of the central plug in an occluding ring setup. Nor need such radii relate to physical objects. Later we will define the second scatterer thickness contour by its value at certain radii, and these too project according to (14.3).

Finally, S , described by the vector \vec{q} , is the projected point where the second scatter takes place and P , described by the vector \vec{r} , is the *field point* at which we wish to compute the fluence. The fluence is the probability for a scatter through \vec{u} into the area element at S followed by a scatter through \vec{v} into the area element at P , integrated over all possible positions of S . P can be anywhere in the plane but S can only be in the non-shaded area. The joint probability is the product of individual probabilities. Noting that $\vec{u} = \vec{q} - \vec{t}$ and $\vec{v} = \vec{r} - \vec{q}$ we obtain

$$\Phi(\vec{r}) = \frac{1}{(2\pi u_0)^2} \int_{q_a}^{q_b} dq \int_0^{2\pi} d\phi_q \frac{q}{v_0^2(q)} f(\vec{q} - \vec{t}) f(\vec{r} - \vec{q}) \quad (14.4)$$

where f is either the Gaussian (Eq. 4.4) or the Molière (Eq. 7.13) probability density without their normalizing factors, which we have written explicitly.

If Eq. (14.4) is applied to the contoured scatterer, $q_a = 0$ and $v_0(q)$ is some known function. The central plug version of the occluding ring is exactly as shown, with $v_0(q) = \text{constant}$. The annulus version would have two open zones and would require two integrals, one from 0 to some q_a and the next from q_b to q_c , with $v_0(q) = \text{constant}$. The generalization to more occluding rings is obvious. The dual ring scheme could be accommodated by letting $v_0(q)$ be a two step function.

Assuming f is Gaussian the double integral in Eq. (14.4) can be reduced to a single integral involving the modified Bessel function [64, 74, 72]. There is no real advantage in this as it is perfectly straightforward to evaluate the double integral by nested applications of Simpson's Rule (Eq. 7.24). One strategy to speed things up is to tabulate the sine and cosine functions and the scattering probability density (Gaussian or Molière). Another is to notice that the ϕ_q integral will be maximum for $q \approx r$ and start there, working first inwards and then outwards until, in each case, the ϕ_q integral vanishes. See `CSI.FOR` for details.

14.5 Scaling and the Generic Solution

By transferring our attention to projections on the MP we have converted all quantities of interest to lengths. For instance, if the characteristic scattering angle of the first scatterer is $\theta_{01} = 0.030$ radian and the throw is 250 cm, these are represented by $u_0 = 0.03 \times 250 = 7.5$ cm. Point P can be anywhere in the MP, and $\Phi(\vec{r})$ can, so far, be any positive function whatsoever. Now we wish to adjust the parameters $u_0, v_0(q), q_b \dots$ so that $\Phi(q)$ is as constant as possible over some region. To define the problem mathematically, we need to define the region. For the present, we'll assume that the region of interest is a circle centered on the origin² and call its radius (not shown in Figure 14.7) the *design radius* r_d .

We now detour briefly to discuss scaling. Different physical beamlines will lead to very different values for $u_0 \dots$. For instance, if the throw were 500 cm instead of 250, u_0 in the previous example would be 15 cm instead of 7.5. Now suppose we have a procedure that adjusts $u_0, v_0(q), \dots$ so that Φ is constant out to $r_d = 10$ cm for a throw of 250 cm. It should be obvious that, in order to find the optimum $u_0, v_0(q), \dots$ for 500 cm, it is not necessary to go through the procedure again. We need merely double all the physical quantities $u_0, v_0(q), \dots$ and we will have an optimum solution for the second case.

To view it another way, all configurations whose $u_0, v_0(q), \dots$ stand in the same proportion to each other represent the same basic mathematical problem, and it is not necessary to solve it over and over. The MP and its associated quantities may be viewed at any magnification without changing the fundamental problem. That said, we might as well assign to one of the quantities the value 1. Then all the others will be of order 1. The choice of this scaling variable is a matter of convenience. There is no 'right' answer. For instance, Koehler et al. [64] use the outer radius of the plug or annulus. This choice seems too restricted to occluding ring systems. Grusell et al. [74] use the projected first scatterer strength u_0 . We will make a different choice, setting $r_d = 1$ cm and calling the corresponding set of optimum values of $u_0, v_0(q), \dots$ a *generic solution*.

A generic solution is a set of MP parameters optimized on a 1 cm circle.

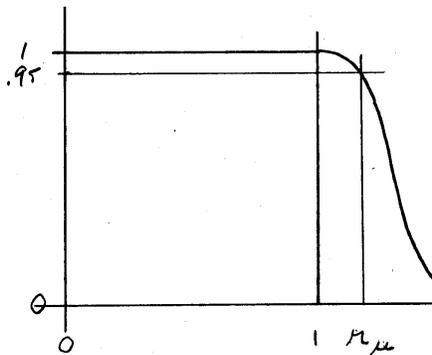


Figure 14.8: Definition of the useful radius.

² The assumption of cylindrical symmetry is not required, and it would of course be very useful to generate elliptical fields. We have tried this without much success, presumably because the physics of scattering in ordinary materials is cylindrically symmetric.

14.6 The Useful Radius

We need to introduce still one more radius to the MP, the *useful radius* r_u characteristic of a given generic solution (Figure 14.8). The optimization process may well yield a fluence function Φ that is totally flat out to 1 cm, but that usually exceeds the specification. If, for instance, we require $\pm 2.5\%$ uniformity, we are able to use everything out to the 95% isodose. $r_u = 1.2$ for the generic solution is typical, depending, of course, on the uniformity requirement and the exact shape of Φ .

14.7 Optimization

In the general case covering both occluding rings and contoured scatterers the parameters to be adjusted are u_0 , $v_0(q)$ and at least one radius r . Finding a generic solution consists of adjusting these by trial and error to obtain as flat as possible a function Φ in the unit circle. There are other considerations in addition to dose uniformity:

- Efficiency: the fluence fraction inside r_u should be as large as possible.
- Energy loss: the sum of u_0 and v_0 (max) should be as small as possible.
- Energy compensation: $v_0(q)$ should be realizable with physical materials.

Some may weigh more heavily than others depending on circumstances. To make the first quantitative it is useful to define the *efficiency* ϵ as

$$\epsilon \equiv \int_0^{2\pi} \int_0^{r_u} \Phi(r) r dr d\phi \quad (14.5)$$

recalling that the integral to $r_u = \infty$ is 1^3 because of the normalization of the scattering functions. As to the third, we have already noted that it requires $v_0(\text{max})/v_0(\text{min}) \leq 4$ if we use lead and lexan.

For contoured scatterers we need to choose a trial function whose parameters we will vary. Grusell et al [74] use a modified cosine which is later discretized to nine steps (it is not clear why this last step is necessary). We will choose a cubic spline because it is guaranteed to be smooth and because we can fix the first derivative at either end if we wish to. Figure 14.9 illustrates how we do this. In the top frame, we have defined four radii, the first being 0 and the last, that of the hole in the scatterer holder. The y values, that is to say, the $v_i \equiv v(r_i)$, are the adjustable parameters of the fit. Requiring $y' = 0$ at $r = 0$ certainly seems to make sense, but the other end is less clear. We have found interesting solutions either by requiring $y' = 0$ or letting it float (natural spline). In the latter case we obtain secondary ‘wings’ in the contour and a slightly larger ϵ , but the gain is not great and the physical scatterer is harder to fabricate. The lower frame illustrates a Siebers-type trial function. The design program will use this form if we enter $r_{min} > 0$.

In finding the best parameters we will usually vary some ‘by hand’ and others ‘by machine’, turning some computer program loose on them. The following procedures work reasonably well:

- Occluding rings: vary u_0 , v_0 and r_{max} by hand, the other r_i by machine.
- Contoured: vary u_0 , v_0 (max) and r_{max} by hand, the other v_i by machine.

³ Because our integration procedure is numerical and there are numerous sources of error or inaccuracy, it is extremely useful for the design program to check this fact numerically.

The machine procedure can either be a grid search or the Marquardt procedure (faster but less robust) [1]. Whatever machine procedure is used, it is essential to be able to ‘lock’ and ‘unlock’ selected parameters while studying their effect.

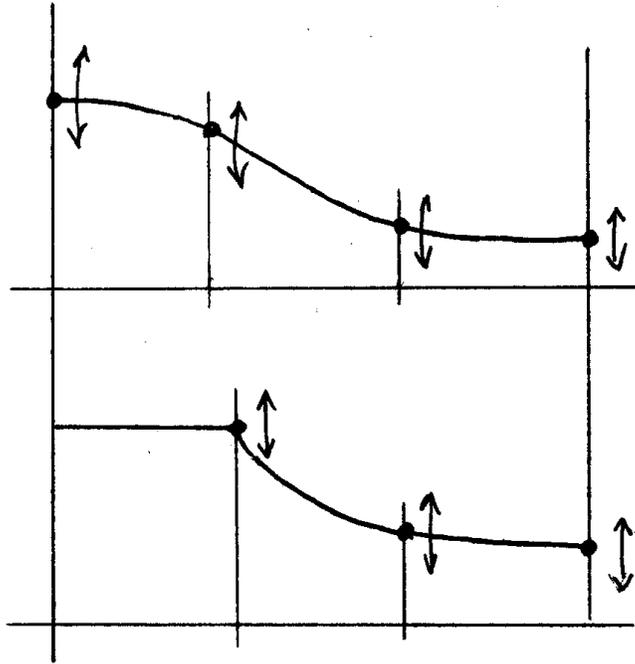


Figure 14.9: Spline trial functions. Upper: domed. Lower: Siebers type.

Like the literature cited so far, we have been vague on optimization. In our opinion, no fully satisfactory method has yet been found. To be specific, let’s consider the optimization of a simple 4-parameter ‘domed’ contoured scatterer by NEU, that is, finding v_2 , v_3 and v_4 given u_0 , v_1 and the r_i . If you try this, you will always get a perfectly flat fluence over some finite region but, otherwise, the procedure suffers from every ill known to non-linear optimization. The solution depends on starting values and even the method (grid or Marquardt). That means there are many minima of the objective function (which is just the *rms* deviation of the fluence from a constant). In addition, we have had a lot of trouble steering the solution by hand to one or another objective: maximum ϵ or minimum energy loss, for instance, or in assuring that the solution will satisfy $v_0(\max)/v_0(\min) \leq 4$. The general behavior suggests that the contour is over-parametrized, yet we have not had much luck cutting back to three v_i instead of four.

That said, we have come up with some pretty good solutions that have worked well in practice, and we doubt (but cannot prove) that anything very much better is waiting in the wings. Some reasonable generic profiles can be found in the NEU examples on the website. One can get an efficiency ϵ around 45%, significantly better than anything we have seen for occluding rings, and the dose is dead flat over most of the useful radius. Depth of penetration is better, within limits, if we load more scattering onto the first scatterer (it has a longer lever arm) but of course the limit of that is single scattering which has exceedingly poor depth of penetration. There may also be secondary reasons

for adjusting the ratio of first to second scatterer. For instance, more lead in the first scatterer will make it easier to compensate if it is also a range modulator. The Siebers profile gives, in practical cases, a few mm better penetration.

In summary, global optimization of the generic profile to meet certain goals is still an unsolved problem. Perhaps one could add to the objective function penalty terms related to efficiency, total scatterer and $v_0(\max)/v_0(\min) \leq 4$, analogous to Prieels' [67] procedure in the modulator problem.

14.8 Returning to the Real World

Assuming we have a reasonable generic solution, how do we convert it to lead and lexan? The procedure is lengthy but straightforward. Let's assume we are not only designing a second scatterer but also an upstream modulator (first scatterer) for a specified SOBP (depth and modulation).

1. The ratio of the actual radius desired (say 12 cm) to the useful radius of the generic solution (say 1.2 cm) is key. Multiply everything in the generic solution by 10.
2. Convert the new u_0 and v_i back to angles by dividing by the appropriate distances, $(z_m - z_0)$ and $(z_m - z_2)$ respectively. Convert the new r_i into radii at the second scatterer by multiplying them by $(z_2 - z_0)/(z_m - z_0)$. You'll need to assume a reasonable z_0 (see comment below).
3. Use the depth of penetration and the modulator step number to find the kinetic energy T_9 into the water tank, that is, out of the second scatterer.⁴
4. If this is the first modulator step, and assuming the central point of the second scatterer is all lead, find the thickness of lead required for that output T_9 and the desired scattering. Find the scatterer input energy T_7 . If you have decided to keep a given amount of lexan even at the center (convenient for machining) you can still find the lead and T_7 . It's a binary degrader problem (Chapter 9).
5. If this is the first modulator step, convert the scattering angle at each radius, along with T_7 and T_9 , to thicknesses of lead and lexan as a function of radius.⁵
6. You now know the energy T_5 out of the first scatterer, and of course you know the energy in, T_3 (the beam energy, corrected for any pre-absorber or scatterer). These, and the desired angle, define a binary degrader problem which allows you to find the lead and lexan for that modulator step.
7. Unless this is the last modulator step, go to step 3.

We have illustrated only one of the many design problems that can be posed. In a way it is the easiest. Provided the beam energy is large enough to provide the desired d_{80} along with the necessary scattering, it can always be solved, usually at the expense of excess lexan on the modulator. If we had specified a

⁴ Here we use NEU indexing, where object 9 is the water upstream of the SOBP and 10 is the SOBP itself. Object 3 is the high-Z part of the first scatterer, 4 is the low-Z part of the first scatterer (if used), 7 is the low-Z part of the second scatterer and 8 is the high-Z part. NEU allows for pre- and post- absorbers and scatterers, objects 1,2 and 5,6 which must be *given*. For now we ignore them.

⁵ To improve accuracy, we do this not only at the four or so design radii, but also at intermediate radii.

simple first scatterer rather than a modulator, the problem would be overdetermined. We cannot specify a beam energy, a scattering system and at the same time, a depth of penetration. NEU is set up to handle many different problems. Here, we have only tried to show one of many paths through the woods.

We have also simplified the explanation by ignoring the dependence of z_0 on modulator step. This has two ramifications. First, the physical second scatterer can only be designed exactly for one modulator step. It makes sense to use the first (thinnest) step which has by far the greatest weight. Second, the entire process just described needs to be iterated over z_0 . As usual, you should look at the program (NEU and its subroutines) for the exact logic.

It may happen that compensation fails for some steps. Second scatterer compensation will fail if the generic solution calls for too little scattering. Then the lexan required for the given energy loss (which always takes priority) already over-scatters the protons. The same applies to the modulator. Compensation will fail if the amount of lexan required for that step already over-scatters the protons. In either case, there will be error messages. Neither case is necessarily fatal to an acceptable design. NEU automatically finishes in a playback mode where the dose distribution is computed for the devices (modulator and second scatterer) *as designed* rather than the ideal devices. The dose distributions will show you whether the failures in compensation are acceptable or not.

Chapter 15

Double Scattering Upstream Modulator Example

This chapter will explain how NEU solves various problems involving a first scatterer or modulator and a second scatterer (none or occluding ring or contoured). A key feature is automatic *playback* of the design using components *as designed* rather than ideal components.

Chapter 16

Beam Steering

This chapter will explain the sensitivity of double scattering to beam imperfections, in particular misalignment at the second scatterer. It will discuss dynamic beam steering to correct misalignment, assuming we have a quality control chamber either near the second scatterer or near the patient (skewness and all that).

Chapter 17

Capabilities of Double Scattering Systems

This chapter will survey the global capabilities of double scattering systems. For instance, if we have a fixed beam energy but can vary the throw, what combinations of treatment depth and field radius are possible? How do these compare with clinical requirements?

Chapter 18

Transverse Penumbra

This chapter will discuss how to compute transverse penumbra at some depth of interest in the patient, given the nozzle configuration (*e.g.* position and thickness of range shifter and modulator). Transverse penumbra is minimized by putting beam modifying materials as far upstream as possible.

Chapter 19

Distal Falloff

This short chapter will discuss how to sharpen distal falloff by minimizing excess material in the beam.

Chapter 20

Collimators

This chapter will discuss collimator scatter and the wise and unwise use of collimators.

List of Tables

7.1	Molière parameters for 158.6 MeV protons incident on Be, Cu and Pb using the ICRU49 [2] range-energy tables. N is the number of target slabs used for numerical integration. See text for other definitions.	56
-----	--	----

List of Figures

1.1	The first NEU design (1990), with experimental data.	6
3.1	Drawing used in deriving Eq.3.6	14
3.2	Drawing used in deriving Eq.3.9	16
4.1	Coordinates and definitions for scattering.	20
4.2	Gaussian radial distribution assuming $r_0 = 1$ cm.	22
4.3	Geometry of transverse penumbra calculation.	23
4.4	Penumbra profile showing 80/20 width.	25
5.1	Cubic spline fit. Light line: cubic spline used to interpolate noisy data (full squares). Dashed line: linear interpolation of the same data. Bold line: cubic spline fit to the data, using 5 auxiliary points (hollow squares).	27
5.2	Procedure for picking initial spline points. Bottom frame: Bragg peak to be fitted. Top frame: smoothed second derivative of the Bragg peak. Middle frame: density function (see text).	29
5.3	Upper frame: Bragg peak with spline fit. Lower frame: deviation from fit. The dashed lines are at \pm the <i>rms</i> deviation.	30
5.4	Spread-out Bragg peak fit by 3-segment broken spline.	31
5.5	Compensated contoured second scatterer.	32
6.1	Range straggling in several materials, from Janni [3].	36
6.2	Fluence, differential fluence and dose as a function of depth for proton beams of a given range and different energy spreads, illustrating $r_0 = d_{80}$	37
6.3	d_{80} and d_{90} vs. range at nozzle entrance' measured at NPTC.	39
6.4	Range-energy in Be, Cu and Pb. Lower frame: test of cubic spline interpolation of points at .1, .2, .5, 1 . . . MeV.	40
6.5	Range-energy over the clinically useful energy range for 4 useful materials.	44
6.6	Drawing for proof of Eq. (6.14).	46
6.7	Energy spectra of degraded proton beams at NPTC. Energy spread and skewness increase as the energy is reduced.	48
6.8	Drawing for proof of Eq. (6.18).	48
7.1	Grand summary of proton results reviewed in [4]. The x coordinate is arbitrary and the target names merely serve as landmarks. Data from each experiment are averaged over everything but target material.	50
7.2	Molière/Fano theory vs. experiment for four metals [4]	57
7.3	Expanded view of Figure 7.2 near end of range.	57
7.4	Deviation from Molière/Fano theory as a function of normalized target thickness for HCL data [4].	58

7.5	Molière angular distribution. Dashed line: $f^{(0)}$ term (Gaussian) only. Inset: expanded view near origin.	58
7.6	Test of Highland's formula as a proxy for Molière theory.	60
7.7	Test of Highland's formula (7.27) against experimental data.	60
7.8	Maximum θ_M over the clinical energy range for four materials.	63
8.1	Nuclear buildup in a Bragg peak.	69
8.2	Total nonelastic cross section of oxygen versus proton range.	70
8.3	Probability of a nonelastic interaction versus proton range in water.	70
8.4	Experiment to measure reduction of primary fluence.	71
8.5	Reduction of primary fluence observed with Faraday cups.	71
8.6	A multi-layer Faraday Cup with CH ₂ plates (aluminum shield not shown).	73
8.7	Differential fluence observed in the CH ₂ MLFC.	74
8.8	Nuclear interactions observed in the CH ₂ MLFC.	74
9.1	Scattering and stopping in various materials.	78
9.2	Scattering v. output energy for a lead/lexan degrader.	82
10.1	Displacement dr_0 in MP from scattering in a target layer dz	85
10.2	(caption from [10]) Dimensionless plot of the standard deviation due to scattering versus depth of penetration. The curve is a plot of Eqn. (15) of Sec. 2 (author: eq. 10.3 of the present report). Open triangles are experimental results for 112 MeV protons on aluminum; solid triangles for 158 MeV protons on aluminum; open circles for 127 MeV protons on water.	86
10.3	Effective origin of scattered particles in a thick target.	87
10.4	Effective origin of scattering ρz_0 divided by target thickness t as a function of normalized target thickness. The lines represent nine separate calculations for 50, 150 and 250 MeV protons on Be, Cu and Pb.	87
10.5	Variation of z_0 in a lead/lexan binary stack.	88
11.1	Measured Bragg peaks from 69 to 231 MeV.	91
11.2	The relative dose on the axis of a uniform circular proton beam of initial range 12 cm of water and radius r_c at the collimator. The curve for $r_c = \infty$ is experimental, the others are calculated.	92
11.3	(From Hong et al. [60]) Central-axis depth-dose measured with a diode. The beam is collimated by a 2.4 mm radius aperture. The air gap between the aperture and the water tank is 1 or 6 cm. Lines: computations using pencil beam algorithm.	94
11.4	Bragg peak extrapolated to zero and fit with a cubic spline by program BPW. The BPK file is listed below.	98
11.5	Absorbed energy fraction computed from the scans of Figure 11.1. Open circles: EM normalization. Full circles: EM + nuclear. Line: theoretical prediction for $P' = 0.012 / (\text{g/cm}^2)$, $f = 0.6$	103
12.1	A downstream modulator.	105
12.2	A compensated upstream modulator.	105
12.3	Figure illustrating modulator weight calculation.	107
12.4	Prieels' argument for finite step size.	108
12.5	Model SOBP defining measures of depth d and modulation m	109
12.6	Performance of a 70-82 MeV system at 70 MeV.	111
12.7	Performance of a 70-82 MeV system at 82 MeV.	111
12.8	Beam gated SOBP showing rounding of proximal corner.	113

12.9	The 3×3 arrangement of modulator wheels.	115
14.1	Double scattering with a beam stopper [64].	117
14.2	The effect of beam steering on double scattering.	118
14.3	Double scattering with dual-ring scatterer [72].	119
14.4	Double scattering with contoured scatterer [74].	120
14.5	Left: second scatterer with Siebers-type profile [75]. Right: cross-field scans [76] at entrance and at Bragg peak.	120
14.6	Geometry of double scattering system.	122
14.7	Double scattering projected on the measuring plane.	122
14.8	Definition of the useful radius.	124
14.9	Spline trial functions. Upper: domed. Lower: Siebers type. . . .	126

Bibliography

- [1] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, “Numerical Recipes: the Art of Scientific Computing,” Cambridge University Press (1986).
- [2] Berger et al., ‘Stopping Powers and Ranges for Protons and Alpha Particles,’ ICRU Report 49 (1993).
- [3] J.F. Janni, ‘Proton Range-Energy Tables, 1KeV - 10 GeV,’ Atomic Data and Nuclear Data Tables **27** parts 1 (compounds) and 2 (elements) (Academic Press, 1982).
- [4] Gottschalk et al., ‘Multiple Coulomb scattering of 160 MeV protons,’ Nucl. Instr. Meth. **B74** (1993) 467-490. We have discovered the following errors: Eq.(2) should read
- $$\Xi(\chi) = \frac{1}{\pi} \frac{\chi_c^2}{(\chi^2 + \chi_a^2)^2}$$
- and in Table 1 the heading α should read α^2 and $\times 10^9$ under χ_c^2 should read $\times 10^6$.
- [5] H.A. Bethe, ‘Molière’s theory of multiple scattering,’ Phys. Rev. **89** (1953) 1256-1266. Four entries in the second column (the Gaussian) of Table II are slightly incorrect (A. Cormack, priv. comm.) but the error (corrected in our programs) is at worst 1%.
- [6] G. Molière, ‘Theorie der Streuung schneller geladenen Teilchen I Einzelstreuung am abgeschirmten Coulomb-Feld,’ Z. Naturforschg. **2a** (1947) 133-145.
- [7] G. Molière, ‘Theorie der Streuung schneller geladenen Teilchen II Mehrfach- und Vielfachstreuung,’ Z. Naturforschg. **3a** (1948) 78-97.
- [8] W.T. Chu, B.A. Ludewigt and T.R. Renner, ‘Instrumentation for treatment of cancer using proton and light-ion beams,’ Rev. Sci. Instr. **64** (1993) 2055-2122.
- [9] Richard Wilson, “A Brief History of the Harvard University Cyclotrons,” Harvard University Department of Physics (2004) (distributed by Harvard University Press, Cambridge, Massachusetts).
- [10] W.M. Preston and A.M. Koehler, ‘The effects of scattering on small proton beams,’ unpublished manuscript (1968), Harvard Cyclotron Laboratory.
- [11] Eidelman et al., ‘Review of Particle Physics,’ Phys. Lett. **B592** (2004). Tables, listings, reviews and errata are also available at the Particle Data Group website <http://pdg.lbl.gov>.

- [12] Verhey et al., ‘Clinical Proton Dosimetry Part I: Beam Production, Beam Delivery and Measurement of Absorbed Dose,’ ICRU Report 59 (1998).
- [13] M.R. Raju, “Heavy Particle Radiotherapy,” Academic Press (1980).
- [14] NCRP Report #93, “Ionizing Radiation Exposure of the Population of the United States,” (1987). For an excerpt, Google ‘average annual exposure’.
- [15] U.S. Nuclear Regulatory Commission 56FR23396 (Google ‘occupational dose limits’).
- [16] A.C. Upton, ‘Health effects of low-level ionizing radiation,’ *Physics Today* **44/8** (1991) 34-39.
- [17] J.W. Boag, ‘Ionization chambers,’ in “The Dosimetry of Ionizing Radiation,” ed. K.R. Kase, B.E. Bjärngard and F.H. Attix, Academic Press (1987) 169-243.
- [18] R. Hagedorn, “Relativistic Kinematics,” W.A. Benjamin (1963).
- [19] M. Abramowitz and I.A. Stegun, ‘Handbook of Mathematical Functions,’ Dover, New York (Ninth printing, 1970).
- [20] H.B. Dwight, ‘Tables of Integrals and Other Mathematical Data,’ Macmillan (1952).
- [21] Philip R. Bevington, “Data Reduction and Error Analysis for the Physical Sciences,” McGraw-Hill (1969).
- [22] B. Gottschalk, ‘On the Characterization of Spread-Out Bragg Peaks and the Definition of Depth and Modulation,’ technical report 28APR03 available at <http://huhepl.harvard.edu/~gottschalk>.
- [23] H.A. Bethe and J. Ashkin, ‘Passage of radiations through matter,’ in *Experimental Nuclear Physics*, E. Segrè (ed.), Wiley, New York (1953): Part II, page 283 et seq.
- [24] B. Gottschalk, R. Platais and H. Paganetti, ‘Nuclear interactions of 160 MeV protons stopping in copper: a test of Monte Carlo nuclear models,’ *Med. Phys.* **26(12)** (1999) 2597-2601.
- [25] H. Paganetti and B. Gottschalk, ‘Test of GEANT3 and GEANT4 nuclear models for 160 MeV protons stopping in CH₂,’ *Med. Phys.* **30(7)** (2003) 1926-1931.
- [26] Thomas Bortfeld, ‘An analytical approximation of the Bragg curve for therapeutic proton beams,’ *Med. Phys.* **24(12)** (1997) 2024-2033.
- [27] Martin J. Berger, ‘Status of proton transport calculations,’ unpublished NIST technical note (June 1992).
- [28] G. Arduini, R. Cambria, C. Canzi, F. Gerardi, B. Gottschalk, R. Leone, L. Sangaletti and M. Silari, ‘Physical specifications of clinical proton beams from a synchrotron,’ *Med. Phys.* **23(6)** (1996) 939-951.
- [29] E.W. Cascio, J.M. Sisterson, B. Gottschalk and S. Sarkar, ‘Measurements of the Energy Spectrum of Degraded Proton Beams at NPTC,’ *Proc. IEEE Radiation Effects Data Workshop* (2004) (in press).

- [30] B.P. Nigam, M.K. Sundareshan and Ta-You Wu, ‘Theory of multiple scattering: second Born approximation and corrections to Molière’s work,’ *Phys. Rev.* **115** (1959) 491-502.
- [31] U. Fano, ‘Inelastic collisions and the Molière theory of multiple scattering,’ *Phys. Rev.* **93** (1954) 117-120.
- [32] M.J. Berger and R. Wang, in “Monte Carlo transport of electrons and photons,” eds. T.M. Jenkins, W.R. Nelson and A. Rindi (Plenum, 1989).
- [33] C. Kittel, “Elementary Statistical Physics,” Wiley (1967).
- [34] W.T. Scott, ‘The theory of small-angle multiple scattering of fast charged particles,’ *Rev. Mod. Phys.* **35** (1963) 231-313.
- [35] A.O.Hanson, L.H. Lanzl, E.M. Lyman and M.B. Scott, ‘Measurement of multiple scattering of 15.7-MeV electrons,’ *Phys. Rev.* **84** (1951) 634-637.
- [36] V.L. Highland, ‘Some practical remarks on multiple scattering,’ *Nucl. Instr. Meth.* **129** (1975) 497-499 and Erratum, *Nucl. Instr. Meth.* **161** (1979) 171.
- [37] B. Rossi and K. Greisen, ‘Cosmic-Ray Theory,’ *Rev. Mod. Phys.* **12** (1941) 240-309.
- [38] G.R. Lynch and O.I.Dahl, ‘Approximations to multiple Coulomb scattering,’ *Nucl. Instr. Meth.* **B58** (1991) 6.
- [39] Chadwick et al., ‘Nuclear Data for Neutron and Proton Radiotherapy and for Radiation Protection,’ ICRU Report 63 (2000).
- [40] B. Gottschalk, K.H. Wang and K. Strauch, ‘Distorted momentum distributions from quasi-elastic proton-proton scattering in carbon at 160 MeV,’ *Nucl. Phys.* **A90** (1) (1967) 83-100.
- [41] B. Gottschalk, W.J. Schlaer and K.H. Wang, ‘Proton-proton bremsstrahlung at 158 MeV,’ *Nucl. Phys.* **75** (3) (1965) 549-560.
- [42] B. Gottschalk and S.L. Kannenberg, ‘Quasi-elastic scattering of 160 MeV protons by alpha clusters in light nuclei I. Carbon,’ *Phys. Rev.* **C2** (1) (1970) 24-40.
- [43] Stephen M. Seltzer, ‘An assessment of the role of charged secondaries from nonelastic nuclear interactions by therapy proton beams in water,’ NIST technical note NISTIR 5221 (1993) (available from the National Technical Information Service, NTIS, U.S. Department of Commerce, Springfield, VA 22161).
- [44] H. Paganetti and M. Goitein, ‘Radiobiological significance of beam line dependent proton energy distributions in a spread-out Bragg peak,’ *Med. Phys.* **27** (2000) 1119-1126.
- [45] H. Paganetti and M. Goitein, ‘Biophysical modeling of proton radiation effects based amorphous track models,’ *International Journal of Radiation Biology* **77** (2001) 911-928.
- [46] We thank Damien Prieels of Ion Beam Applications s.a. (IBA) for providing us with figure 8.1.
- [47] C.A. Carlsson and G.A. Carlsson, ‘Proton dosimetry with 185 MeV protons: dose buildup from secondary protons and recoil electrons,’ *Health Phys.* **33** (1977) 481-484.

- [48] Lynn J. Verhey, Andreas M. Koehler, Joseph C. McDonald, Michael Goitein, I-Chang Ma, Robert J. Schneider and Miles Wagner, ‘The determination of absorbed dose in a proton beam for purposes of charged-particle radiation therapy,’ *Radiation Research* **79** (1979) 34-54.
- [49] A. Mascia, J. DeMarco, P. Chow and T. Solberg, ‘Benchmarking the MC-NPX nuclear interaction models for use in the proton therapy energy range,’ *Proc. XIVth Intl. Conf. on the Use of Computers in Radiation Therapy*, Seoul, May 10-13 (2004) 478-481.
- [50] H. Paganetti, H. Jiang, S.-Y. Lee and H.M. Kooy, ‘Accurate Monte Carlo simulations from nozzle design, commissioning and quality assurance for a proton radiation therapy facility,’ *Med. Phys.* **31** (7) (2004) 2107-2118.
- [51] L. Eyges, ‘Multiple scattering with energy loss,’ *Phys. Rev.* **74** (1948) 1534-1535.
- [52] C.D. Zerby and F.L. Keller, ‘Electron transport theory, calculations, and experiments,’ *Nucl. Sci. Eng.* **27** (1967) 190-218.
- [53] A.M. Koehler, priv. comm.
- [54] H. Bichsel and T. Hiraoka, ‘Energy spectra and depth-dose curves for 70 MeV protons,’ *Intl. Journ. Quant. Chem.* **23** (1989) 565-574.
- [55] Martin J. Berger, ‘Penetration of proton beams through water I. Depth-dose distribution, spectra and LET distribution,’ NIST technical note NISTIR 5226 (1993) (available from the National Technical Information Service, NTIS, U.S. Department of Commerce, Springfield, VA 22161).
- [56] H. Bichsel, ‘Calculated Bragg curves for ionization chambers of different shapes,’ *Med. Phys.* **22**(11) (1995) 1721-1726.
- [57] A.M. Koehler, ‘Dosimetry of proton beams using small silicon diodes,’ *Rad. Res. Suppl.* **7** (1967) 53-63.
- [58] E.J. Burge and D.A. Smith, ‘Theoretical study of slit scattering,’ *Rev. Sci. Inst.* **33** (1962) 1371-1377.
- [59] P. van Luijk, A.A. van’t Veld, H.D. Zelle and J.M. Schippers, ‘Collimator scatter and 2D dosimetry in small proton beams,’ *Phys. Med. Biol* **46** (2001) 653-670.
- [60] Hong et al., ‘A pencil beam algorithm for proton dose calculations,’ *Phys. Med. Biol.* **41** (1996) 1305-1330.
- [61] Computerized Radiation Scanners Inc., 140 Sopwith Drive, Vero Beach, FL 32968, tel (772) 562 0405, email crs5@juno.com.
- [62] B. Gottschalk, ‘Logarithmic audible ratemeter for NPTC,’ technical note HCL 3/17/00.
- [63] R.R. Wilson, ‘Radiological use of fast protons,’ *Radiology* **47** (1946) 487-491.
- [64] A.M. Koehler, R.J. Schneider and J.M. Sisterson, ‘Flattening of proton dose distributions for large-field radiotherapy,’ *Med. Phys.* **4**(4) (1977) 297-301.
- [65] B. Gottschalk, A.M. Koehler, J.M. Sisterson and M.S. Wagner, ‘The case for passive beam spreading,’ *proc. Proton Therapy Workshop*, Paul Scherrer Institut (1991) 50-53 (PSI report 111).

- [66] V. Kostjuchenko, D. Nichiporov and V. Luckjashin, 'A compact ridge filter for spread out Bragg peak production in pulsed clinical beams,' *Med. Phys.* **28(7)** (2001) 1427-1430.
- [67] D. Prieels, 'A new technique for the optimization of a range modulator,' IBA technical note 88.17.52.003 (1997).
- [68] B. Gottschalk, 'Preliminary design of passive beam spreading for the NPTC proton radiotherapy nozzle,' technical note HCL 04APR95.
- [69] B. Gottschalk, 'Gated modulators for the NPTC nozzle,' technical note HCL 18SEP95 (IBA 88.17.52.004).
- [70] D. Prieels, 'Clinical performance in double scattering,' IBA technical note 88.17.53.011 Rev. A (1999).
- [71] Y. Takada, 'Dual-ring double scattering method for proton beam spreading,' *Japan J. Appl. Phys.* **33** (1994) 353-359.
- [72] Y. Takada, 'Optimum solution of dual-ring double-scattering system for an incident beam with given phase space for proton beam spreading,' *Nucl. Instr. Meth.* **A485** (2002) 255-276.
- [73] Anders Montelius and Anders Brahme, 'Charged particle beam flattening using an optimized dual scattering foil technique,' unpublished manuscript (1988).
- [74] Grusell et al., 'A general solution to charged particle beam flattening using an optimized dual scattering foil technique, with application to proton therapy beams,' *Phys. Med. Biol.* **39** (1994) 2201-2216.
- [75] J.V. Siebers and D.W. Miller, 'Passive scattering system design optimization for proton radiation therapy,' AAPM Annual Meeting (1992) abstract V5.
- [76] J. Burns, 'Introducing the new neurosurgery telescope,' HCL technical note February 1998.