Proton Radiotherapy Uncertainties Arising from Computed Tomography

Daniel Warren

Brasenose College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2013

Abstract

Proton radiotherapy is a cancer treatment which has the potential to offer greater cure rates and/or fewer serious side effects than conventional radiotherapy. Its availability in the UK is currently limited to a single low-energy fixed beamline for the treatment of ocular tumours, but a number of facilities designed to treat deep-seated tumours are in development. This thesis focusses on the quantitative use of x-ray computed tomography (CT) images in planning proton radiotherapy treatments. It arrives at several recommendations that can be used to inform clinical protocols for the acquisition of planning scans, and their subsequent use in treatment planning systems.

The primary tool developed is a software CT scanner, which simulates images of an anthropomorphic virtual phantom, informed by measurements taken on a clinical scanner. The software is used to investigate the accuracy of the stoichiometric method for calibrating CT image pixel values to proton stopping power, with particular attention paid to the impact of beam hardening and photon starvation artefacts.

The strength of the method adopted is in allowing comparison between CT-estimated and exactly-calculated proton stopping powers derived from the same physical data (specified in the phantom), leading to results that are difficult to obtain otherwise. A number of variations of the stoichiometric method are examined, identifying the best-performing calibration phantom and CT tube voltage (kV_p). Improvements in accuracy are observed when using a second-pass beam hardening correction algorithm.

Also presented is a method for identifying the proton paths where stopping power uncertainties are likely to be greatest. Estimates of the proton range uncertainties caused by CT artefacts and calibration errors are obtained for a range of realistic clinical scenarios. The current practice of including planning margins equivalent to 3.5% of the range is found to ensure coverage in all but the very worst of cases. Results herein suggest margins could be reduced to <2% if the best-performing protocol is followed; however, an analysis specific to the CT scanner and treatment site in question should be carried out before such a change is made in the clinic.

Acknowledgements

Thank you:

- to my supervisors, Ken Peach and Mark Hill, for their advice and support, and particularly their comments as I put metaphorical pen to digital paper
- to my friends Tracy Underwood, Claire Timlin, Puthenparampil Wilson and Daniel Abler for the good times we had in and out of the office
- to Bleddyn Jones, Charlie Crichton, Steve Harris, Jan Jansen and Andrew Reilly for interesting and useful conversations along the way
- to the University of Oxford Department of Physics for funding my studentship and, alongside Brasenose College and the PARTNER project, funding my travels
- to the athletes and volunteers at KEEN (`http://www.keenoxford.org`) for helping me keep my sanity
- to my parents, Catherine and Alan, who have always been very supportive of me, and to my brother, Ed, for being my brother

Oxford
October 2013

Contents

List of Abbreviations

AF	adult female
AM	adult male
ASCII	American Standard Code for Information Interchange
BBHC	bone-based beam hardening correction
BREP	boundary representation
CODATA	Committee on Data for Science and Technology
CT	computed tomography
CTV	clinical treatment volume
DICOM	Digital Imaging and Communications in Medicine
DNA	deoxyribonucleic acid
FFT	fast Fourier transform
HU	Hounsfield Units
ICRP	International Commission on Radiological Protection
ICRU	International Commission on Radiation Units and Measurements
IDD	integral dose-depth
IQR	inter-quartile range
LAC	linear attenuation coefficient
LET	linear energy transfer
MAC	mass attenuation coefficient

MCNPX Monte Carlo N-Particle eXtended

MCS multiple Coulomb scattering

MDCT multi-detector computed tomography

MEX MATLAB executable

MLCs multi-leaf collimator

MRI magnetic resonance imaging

NHS National Health Service

NIST National Institute of Standards and Technology

NURBS non-uniform rational basis spline

OAR organ at risk

PSI Paul Scherrer Institute

PTV planning target volume

r.m.s. root-mean-square

RBE radiobiological effectiveness

ROCOCO Radiation Oncology Collaborative Comparison

RTP radiotherapy treatment planning

SOBP spread-out Bragg peak

SPR stopping power ratio

SRIM Stopping and Range of Ions in Matter

TPS treatment planning software

UK United Kingdom of Great Britain and Northern Ireland

Chapter 1

Introduction

Cancer is a major health risk: statistics show that almost 40% of the UK population will suffer from it at some point in their life [1], and in 2009 alone approximately 0.5% of the population were newly diagnosed [2]. Cancer comprises a wide range of diseases for which many treatment options exist. This thesis concerns radiotherapy, which, at the time of writing, constitutes 5% of the budget for cancer treatment in England yet contributes to 40% of cures [3] and, if properly utilised, could contribute to many more [4]. It specifically focusses on issues related to the burgeoning modality of proton radiotherapy whose availability in the UK is currently poor, but which has the potential to offer greater cure rates and/or fewer serious side effects for many treatment sites.

1.1 Radiotherapy

Radiotherapy involves exposing the region affected by cancer to ionising radiation. In the case of many well-localised tumours, cancer cells can be targeted with high selectivity and specificity through the use of external radiation beams with optimal geometry, intensity, constituent particle type and energy spectrum. This optimisation is the realm of radiotherapy treatment planning (RTP).

Radiotherapy treatments are prescribed using a quantity known as radiation dose, and more specifically a physical concept termed *absorbed dose* by the International Commission on Radiation Units and Measurements (ICRU) [5]. Absorbed dose, D is equal to $\frac{\mathrm{d}E}{\mathrm{d}m}$, where $\mathrm{d}E$ is the energy deposited by the incident radiation within a target mass $\mathrm{d}m$. The SI derived unit for absorbed dose is the gray (Gy), which corresponds to $1\,\mathrm{J{\cdot}kg^{-1}}$.

The response of biological tissue to radiation dose is a complex issue, being governed by myriad variables, and comprises a whole field of study in itself. Little radiation

biology will be discussed in this predominantly physics-based thesis, but it is of great clinical importance in determining radiotherapy prescriptions and schedules. It is generally accepted that most lethal effects of radiation on the cellular scale occur as a result of DNA damage produced, either by direct ionisation of the DNA helix or by the interactions of radiation-induced free radicals therewith. Such lethal damage may arise due to the build-up of injuries that are individually sub-lethal and repairable. Some have used this to argue a mechanistic basis for the widespread linear-quadratic model in which the fraction of cells surviving an acute radiation of dose D is equal to $\exp(-[\alpha D + \beta D^2])$ [6]. α and β must still be determined experimentally, being dependent on the cell type, cycle state and environment, as well as the radiation type and energy, at a minimum.

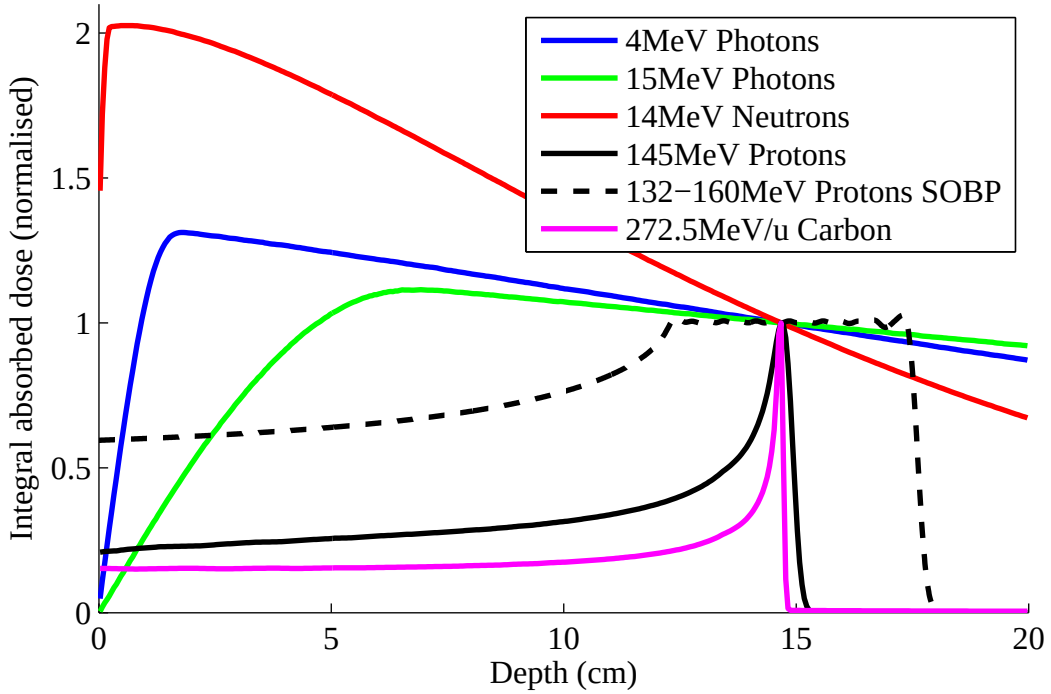
One particularly notable phenomenon is that many normal tissue cells are less prone to radiation damage than most cancer cells, a fact at least partially attributed to their greater ability to repair DNA damage [7]. The timescale for such repair is shorter than a day, so the differential radiosensitivity is capitalised upon by delivering radiotherapy in a number of daily sessions (typically around 30), termed 'fractions'[1]. When complete repair of sublethal damage is assumed, the treatment acts as n independent irradiations of dose $d = \frac{D}{n}$, so the linear quadratic model predicts a surviving fraction of $\exp(-n[\alpha d + \beta d^2])$.

1.2 Radiation types

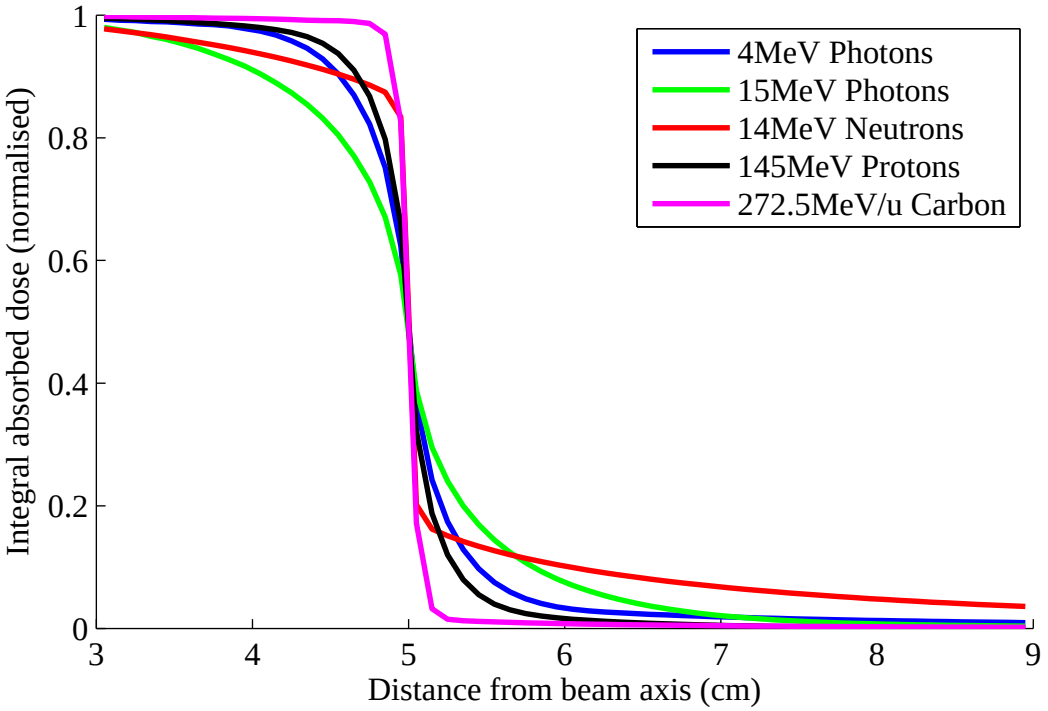
The relationship between radiation dose and depth is very important for radiotherapy, as it governs how effectively a tumour at a particular location can be irradiated and the degree to which surrounding tissues are also exposed. Dose-depth curves (and their lateral equivalents) vary a great deal for different types of radiation and as the energy of that radiation changes, which can be seen in figure 1.1. Two significantly different families of curve are apparent:

- The uncharged particles (neutrons, photons) display an initial build-up region in which dose increases from an initial value close to zero followed by an exponential fall-off. The build-up occurs because uncharged radiations are predominantly indirectly ionising – most of the energy is ultimately deposited by the electrons (in the case of photons) or ions (in the case of neutrons) set in motion by a

[1]Fractionation also exploits other time-dependent phenomena, including the supply of oxygen to hypoxic (and therefore radioresistant) regions of a tumour, and variations in the radiosensitivity of cells as they progress through the cell cycle.



(a) Dose-depth curves, integrated over the lateral plane and normalised to their value at the same point, approximately 15 cm in depth.



(b) Penumbral dose curves, integrated over the depth direction and normalised to the central dose.

Figure 1.1 – The spatial variation of physical dose deposited by a variety of clinical radiation types. Data from simulations of the energy deposited in water by beams that are parallel*, monoenergetic† and circular with a 5 cm radius, using the Monte Carlo particle transport code MCNPX. Secondary production of protons, neutrons, electrons, photons and ions were considered.

* In many clinical radiation sources, laterally spread beams emerge from a focal point and so inverse square effects are also observed. † Most therapy photons originate from Bremsstrahlung, which causes a broad spectrum of energies.

few catastrophic collisions of the projectile. These products deposit the bulk of their energy downstream and have a relatively predictable range. The peak dose occurs at the point electronic equilibrium is reached, which is to say that the number and energies of charged secondary particles generated upstream of that depth are equal to those travelling downstream. Broadly speaking, the position of the peak will be a function of the mean free path in water of the incident particles and the range of the secondaries. Beyond the peak, the curve exhibits an exponential falloff with the decay parameter determined by the mean free path of the projectiles in water. [8]

- The dose-depth characteristics of charged particles (e.g. protons, carbon ions) constitute an initial plateau that increases at a growing rate, culminating in a sharp peak (the Bragg peak) beyond which very little dose is deposited. Briefly, these features arise from the fact that charged particles undergo numerous glancing interactions with atomic electrons as they travel through matter. Each interaction slightly reduces the kinetic energy of the projectile and thus its speed, causing it to spend longer in the electrostatic field of the electrons in a given thickness of material; it consequently experiences a greater integrated decelerating force (impulse), deposits more energy locally (transferred to bulk electrons during deceleration) and eventually stops at a predictable depth when its kinetic energy is exhausted. The physics of proton dose deposition will be discussed in more detail in chapter 4. Given the sharp dose gradient, it is necessary to use charged particles at a number of different energies to accommodate targets of realistic thickness. Energy spectra can be designed to provide a homogenous dose over the region of interest, leading to a dose-depth curve known as a Spread Out Bragg Peak (SOBP). [9]

Photons are the most common particle to be used radiotherapeutically, particularly in the form of megavoltage x-rays generated by linear accelerators. A great deal of research and development has occurred since the first x-ray treatments took place in 1896 [10], with improvements in the technologies underlying x-ray generation, treatment planning, dosimetry and beam shaping, together with leaps in clinical understanding, greatly increasing efficacy and reducing morbidity. Nevertheless, the limited opportunities for modulation of the photon dose along the direction of incidence make it impossible to provide a curative dose at some tumour locations without causing intolerable levels of adverse side effects in other tissues irradiated by the beam.

The extra degree of freedom offered by variable energy ions was highlighted by Wilson in his 1946 paper postulating the radiotherapeutic use of protons [11]. At the time, a number of appropriately-specified accelerators were becoming available due to the ever-increasing energy frontier of nuclear and particle physics, and early human trials took place in physics laboratories[2]. Growth in clinical proton therapy capacity has been historically limited by the cost and scale of the acceleration equipment, resulting in its routine use for only a subset of particularly complex cases[3]. There has recently been a boom in the number of proton centres planned, partly fuelled by increased commercialisation of the necessary technology and an international market of well-funded patients who opt for the technique; 37 proton therapy centres are currently in operation worldwide [14] and it is expected that this statistic will have more than doubled by the end of this decade, contributed to by at least two new facilities in British hospitals.

In addition to the differences in macroscopic dose distribution, x-rays and protons display some differences in energy deposition on the microscopic level. The linear energy transfer (LET)[4], which quantifies the amount of energy deposited locally, of protons increases as they approach the end of their range, whilst it takes a smaller and broadly constant value for therapeutic photons. Increased ionisation densities are more likely to cause cell death at a particular dose because the probability of complex damage to a localised region of DNA is greater. Protons are thus said to have an increased radiobiological effectiveness (RBE)[5] for cell kill compared to photons, which is further enhanced at the end of their range. Another dose concept, RBE-weighted absorbed dose, has been introduced to facilitate isoeffect planning and the translation of clinical x-ray experience to proton treatments; denoted by D_{RBE}, it is the product of the absorbed dose D at a point and the corresponding RBE, with units given by

[2]Somewhat surprisingly, it was the sharp penumbra in the plateau region of 340 MeV protons, rather than the Bragg Peak, that John H Lawrence took advantage of in the first human trials of proton therapy. [12]

[3]In the UK, the NHS currently approves high-energy proton therapy to be delivered abroad for a range of rare childhood cancers, and a small number of adult cases where the tumour is located adjacent to crucial brain and spinal tissues [13]. A low-energy facility at the Clatterbridge Cancer Centre in Merseyside has been treating ocular tumours since 1989.

[4]$\mathrm{LET}_\Delta = \frac{\mathrm{d}E_\Delta}{\mathrm{d}x}$, where E_Δ corresponds to the energy lost in collisions that result in electron kinetic energies less than Δ [5]. In order to reflect ionisation density on the DNA scale, Δ should be such that the secondary electrons have a range of nanometres – a suitable value would be in the region of 100 eV. [15] Unrestricted LET, where $\Delta = \infty$, is often used in practice, and dose-averaged LET will typically be used when considering a volume.

[5]RBE for a particular radiation and biological effect is quantified as the dose of reference x-ray radiation needed to cause the effect divided by the dose of the test radiation that causes the same effect.

Gy (RBE). The current clinical approach to reporting RBE-weighted dose in proton therapy is described in ICRU Report 78 [16]. It notes that there is overwhelming evidence from in-vitro cellular experiments that the proton RBE for cell kill is greater than 1. For many years, a number of proton therapy centres have used a constant value of 1.1 irrespective of range effects; the ICRU believes that the present level of clinical evidence is consistent with this and recommend the use of a generic RBE of 1.1 in prescribing, recording and reporting doses[6].

1.3 Delivery of proton therapy

Proton therapy systems comprise a collection of individually-sophisticated components, a brief description of which follows.

1.3.1 Acceleration

In order for protons to be of therapeutic use, their initial kinetic energy must typically lie in the region of 60-250 MeV (range in water 3-39 cm). Two systems of particle acceleration are in use clinically: the cyclotron and the synchrotron. Both are cyclic accelerators, in which protons repeatedly revolve around a circuit, receiving a small accelerating kick every revolution. Once sufficient energy is attained, the protons are extracted from the accelerator and transferred to the treatment delivery head.

- Cyclotrons use a constant magnetic field to bend the protons along a circular path. In the non-relativistic limit[7], the Lorentz force results in a radius of curvature proportional to speed. In order to accommodate the accelerating particles, the cyclotron consists of an evacuated cylindrical space, only allowing extraction at the edge. Cyclotrons therefore produce therapeutic protons at a single energy and in a steady stream.

- Synchrotrons use a fixed beam pipe, keeping the trajectory of the protons at a constant radius, thus requiring the magnetic field and frequency of the electromagnetic waves used for acceleration to be increased as the particles increase in velocity. Relativistic effects do not limit the maximum energy

[6]Concerns related to an elevated RBE at the end of proton range are commonplace [17, 18, 19], however, and efforts have been made to assess the resulting impact on biological dose e.g. [20, 21]

[7]Therapeutic protons are actually moderately relativistic ($\gamma \lessapprox 1.3$) at maximum energy, but the seeming lag in progress this causes can be accommodated through modifications of the accelerating electric field.

attainable, as these can be taken into account when generating the waveforms in question. With a synchrotron, particles may thus be extracted at a range of different energies, albeit not in a continuous stream.

1.3.2 Beam spreading

The beam output by the accelerator has very small cross-sectional dimensions (millimetres) compared to the typical size of target (multiple centimetres). Two different classes of solution exist to ensure adequate lateral coverage:

- Scattering has a great deal of heritage, being the only beam spreading method available for proton treatments until 1996, and is still in use for the majority of patients. Here broadening is typically achieved using two scatterers: the first has uniform thickness[8] thus introducing a Gaussian distribution in angle, which is translated into lateral spread by a drift space; the second scatterer has a thickness profile calculated to convert the Gaussian to a distribution with a large uniform central region at the target distance, which is retained using a collimator [22]. Scattering inevitably causes some beam contamination, energy degradation and energy broadening (due to interactions in the scatterer) and considerable loss of fluence (due to the collimation), but irradiates the entire target plane simultaneously.

- Scanning modifies the lateral profile of the beam itself as little as possible, instead changing its trajectory as a function of time in order to ensure coverage of the target plane. Modern scanning systems use computer-controlled horizontal and vertical dipole magnets[9], and divide the target into a series of discrete spots (spot scanning). While scanning can be used to emulate scattering (a situation known as uniform scanning), a major driver in its adoption is the ability to modulate the dose delivered to each spot by adjusting the beam's dwell time. Such intensity modulation provides treatment planners with a welcome extra dimension for use in their optimisations[10]. The temporal structure of scanning presents additional challenges (and opportunities) when treating moving targets.

[8] often this will form the range shifter mentioned in 1.3.3

[9] The first clinical implementation at the Paul Scherrer Institute (PSI) used magnets to scan in one lateral direction, physically moving the patient in the other. [23]

[10] Although one can also conceive a form of intensity modulated proton therapy which uses scattered beams and actuated multi-leaf collimators (MLCs), for which motion interplay would be less of an issue.

1.3.3 Patient-specific beam modification

A number of devices may be used to modify the beam leaving the accelerator before it reaches the patient, the selection of which is one aspect of proton therapy treatment planning. In the case of scanned systems, the amount of material intersecting the beam path is best kept minimal in order to avoid broadening the narrow ('pencil') beam, so many of these devices may be absent, with spots instead being targeted by appropriate control of synchrotron output energy and scanning dipole strength.

- Range shifters are blocks of material with a constant thickness. They reduce the beam energy such that the maximum proton range matches the maximum depth of the target within the patient.

- Modulator wheels are discs of material which revolve along an axis parallel to (but offset from) the beam. The thickness profile varies along the circular track traced by the beam in such a manner as to produce an energy distribution equivalent to the Spread Out Bragg Peak weighting function. Wheels typically contain multiple tracks, allowing selection of a modulation depth which corresponds to the thickness of the target.

- Ripple filters are structures whose thickness varies periodically and rapidly with position in the lateral plane. The ripples are sufficiently small that variations in range with position are completely washed out by lateral scattering. Their purpose is to broaden excessively sharp Bragg peaks, improving SOBP homogeneity when the number of energies in use is small.

- Compensators are plastic blocks milled to have a varying thickness profile, used to conform the distal edge of the SOBP to the target region. A specific compensator will be manufactured for each radiation field used in every patient, the depth at each point in the lateral plane being derived from the distance between the patient's skin and the distal contour of the target.

- Apertures are metallic structures used to collimate the lateral cross section of the beam to that of the target. They are field-specific like compensators, but their composition makes them significantly more difficult to manufacture; consequently, the recently-commissioned Roberts Proton Therapy Centre in Philadelphia uses electronically actuated multi-leaf collimators for uniform scanning treatments.

1.4 Treatment planning

In almost all contemporary radiotherapy treatment planning, calculations of dose delivered to points within the patient are based upon a geometric model derived from medical imaging. In the current paradigm, this image originates from x-ray computed tomography (CT).

After a decision has been made to treat a patient with radiotherapy, a CT scan is performed in which the patient is positioned as they would be for treatment. The radiation oncologist will delineate the region that must be treated (CTV - clinical treatment volume[11]) as well as organs at risk of radiation injury (OARs) on the relevant slices of this 'planning CT'. A dose prescription will follow, which normally consists of a prioritised list of target dose or dose-volume criteria for these structures.

It is the role of medical physicists or dosimetrists to construct a radiotherapy treatment plan that meets the prescription using treatment planning software (TPS). Such a plan will include information on the particle type and energy, angles and positions of the treatment head relative to the patient, shapes of beam-modifying devices, and parameters that control intensity modulation such as pencil-beam spot positions and weightings. The TPS uses CT image pixel values in a quantitative manner to estimate parameters relating to radiation transport (such as mass density or proton stopping power) at each point within the patient. Radiation dose distributions are calculated in three dimensions using physical models that incorporate these parameters, allowing the operator to assess a plan against the prescription.

In the process of treatment planning, additional structures may be marked up on the scan. Traditionally this has included a planning target volume (PTV), usually a geometric expansion of the CTV, which is intended to account for the practical uncertainties involved in delivering dose to a particular point[12]. Examples of such uncertainties include deviations in the positioning of the patient relative to the treatment head from day to day and errors in the estimation of proton range within the patient. It is the latter category of uncertainties which will be considered in this

[11] In most solid cancers, the CTV consists of the gross tumour volume (GTV) visible on the CT scan, as well as a margin that takes account of the microscopic spread of malignant cells. Other imaging, such as MRI and PET, may also be used to help identify the CTV, and nearby lymph nodes may also be included.

[12] It should, however, be noted that the latest guidance (ICRU Report 78 [16]) acknowledges that use of a PTV may not be optimal when protons are incident from multiple directions, due to the asymmetric nature of dose uncertainties. Beam-specific PTVs in which margins are applied asymmetrically, being larger in the direction of beam travel than the transverse plane, have been proposed [24].

thesis, and specifically those which arise when pixel values in the planning CT deviate from those that would be expected under ideal circumstances.

Chapter 2

Computed tomography (CT) and image artefacts

X-ray computed tomography (CT) underpins the majority of contemporary radiotherapy treatment planning, providing both geometric data as to the position of the target and quantitative data used to predict how radiation will propagate within the body. CT is a three-dimensional imaging modality that maps a quantity known as the x-ray linear attenuation coefficient (LAC, conventionally denoted μ) within the imaging volume. The x-ray LAC at a point varies with the mass density and composition of the material located there, and is therefore useful for the identification of internal structures. Image data in CT has traditionally been collected as a stack of axial slices. If it were possible to take a standard projection x-ray image of an isolated slice along the cranial-caudal (head-to-toe) axis, the result would be equivalent to an idealised CT slice acquired at the same location.

Figures illustrating this Chapter were generated using the software described in chapter 3.

2.1 CT geometry

Computed tomography is made possible because mathematical transforms exist that allow approximate x-ray attenuation data for points in the $x - y$ plane to be derived from attenuation measurements made in other 2-dimensional coordinate systems. Most modern CT scanners use the third generation (or fan-beam) geometry shown in figure 2.1[1]. This is a cylindrical scheme in which the patient is located at the origin

[1]First generation scanners utilised a 'translate and rotate' approach in which a point source-detector pair moved in a straight line, obtaining attenuation profiles for a number of rotation steps. The second generation scheme was similar, but data acquisition time was decreased by use of a broad beam source and a detector array.

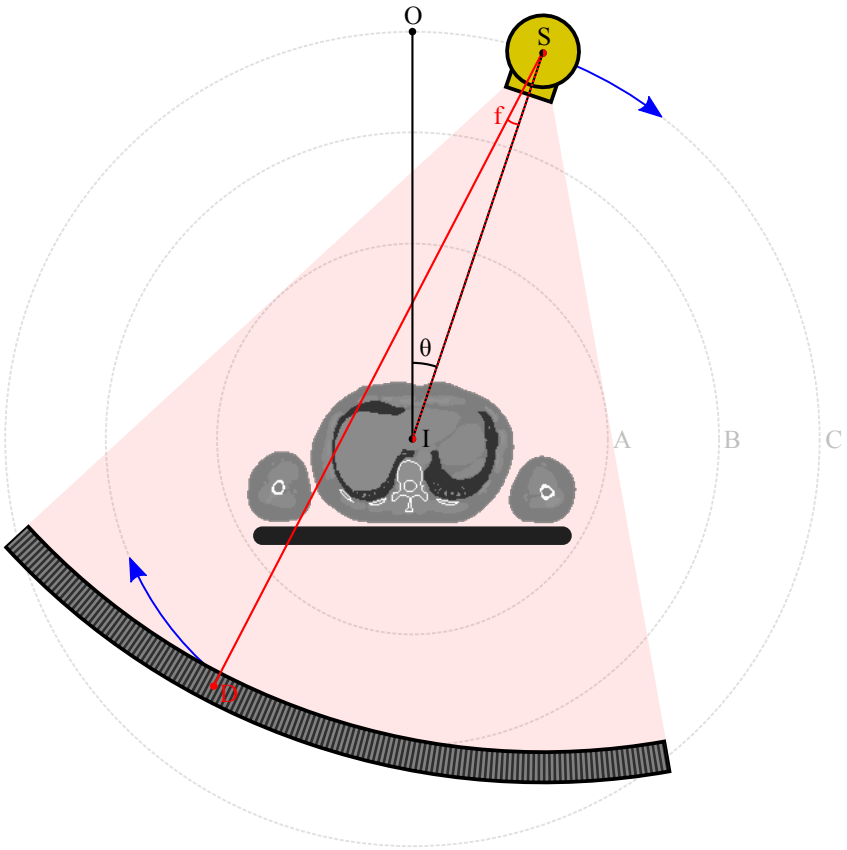


Figure 2.1 – The geometrical arrangement of a third generation CT scanner. Point I is the isocenter of the scanner, point S is the focal point of the x-ray source and point D is the position of the detector element corresponding to fan angle f at revolution angle θ. X-ray attenuation data is collected for many individual projections, represented by the line SD for each f and θ value sampled. Circle C shows the path travelled by the source during a complete revolution whilst circle B shows the corresponding path for the detector's central element. Circle A is the maximum area in which a point will contribute to a projection for all values of θ – it is the largest area in which objects may be reconstructed without the use of extrapolation techniques and a good indicator of bore size.

(termed the isocenter) about which a diametrically opposed x-ray source and detector revolve. The source is typically a high-vacuum x-ray tube operating around 100 kV whose output is shaped as a fan or cone beam, whilst the detector consists of a one- or two-dimensional array of elements in order to measure the attenuation along a number of lines at any one time. The raw measurements necessary to reconstruct an individual slice are thus labelled by the $\theta - f$ coordinate system.

2.2 X-ray attenuation

Attenuation of a primary x-ray beam is well described by the Beer-Lambert law [25]. It states that for a monochromatic pencil beam of x-rays with initial intensity I_0 passing through a homogenous medium, the central intensity at a depth x will be given by:

$$I(x) = I_0 \mathrm{e}^{-\mu x} \tag{2.1}$$

where μ is the x-ray linear attenuation coefficient (LAC) – a property particular to the medium and x-ray energy in question. This result is as would be expected for a fixed density of targets which present a constant interaction cross section. Attenuation of the primary beam in the 100 keV energy regime may occur due to the photoelectric effect or as a result of scattering (both elastic and inelastic) from electrons in the medium [26] – the relative contribution of each process to the attenuation cross section can be seen in figure 2.2.

Equation (2.1) can be extended to the case in which the medium varies along the path traversed in reaching depth x by considering discrete elements of infinitesimal thickness δx, each of which will independently affect the beam's intensity:

$$I(x) = I_0 \lim_{\delta x \to 0} \left[\prod_{n=0}^{x/\delta x} \mathrm{e}^{-\mu(x=n\cdot\delta x)\delta x} \right] = I_0 \exp\left[-\int_0^x \mu(x')\ \mathrm{d}x' \right] \tag{2.2}$$

An intensity-proportional detector located at point D would thus measure a fraction of the x-rays originally incident from point S given by:

$$\frac{I}{I_0} = \exp\left(-\int_{SD} \mu\mathbf{r} \cdot \mathrm{d}\mathbf{r} \right) \tag{2.3}$$

The set of line integrals $\zeta(\theta, f) = \int_{SD} \mu\mathbf{r} \cdot \mathrm{d}\mathbf{r} = -\ln\left[\frac{I(\theta,f)}{I_0}\right]$ is known as a sinogram and examples can be seen in figure 2.3.

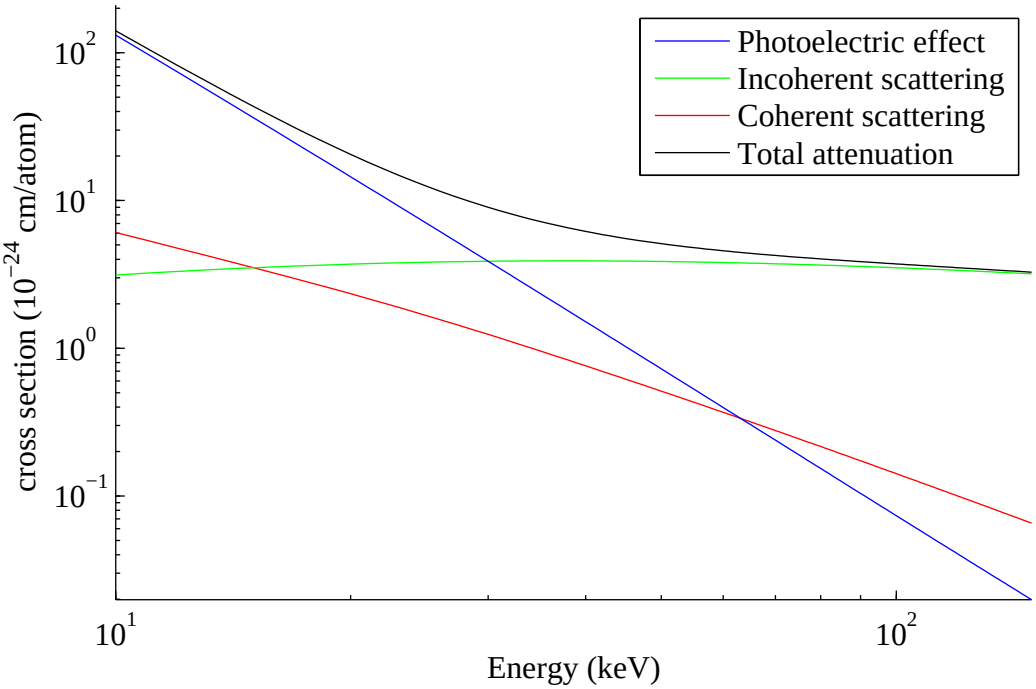


Figure 2.2 – Logarithmic plot of the cross sections for major x-ray attenuation processes in water in the energy regime of kilovoltage x-ray CT. No pair production occurs as x-ray energies are well below the threshold energy of 1.022 MeV. Data sourced from the XCOM database [27].

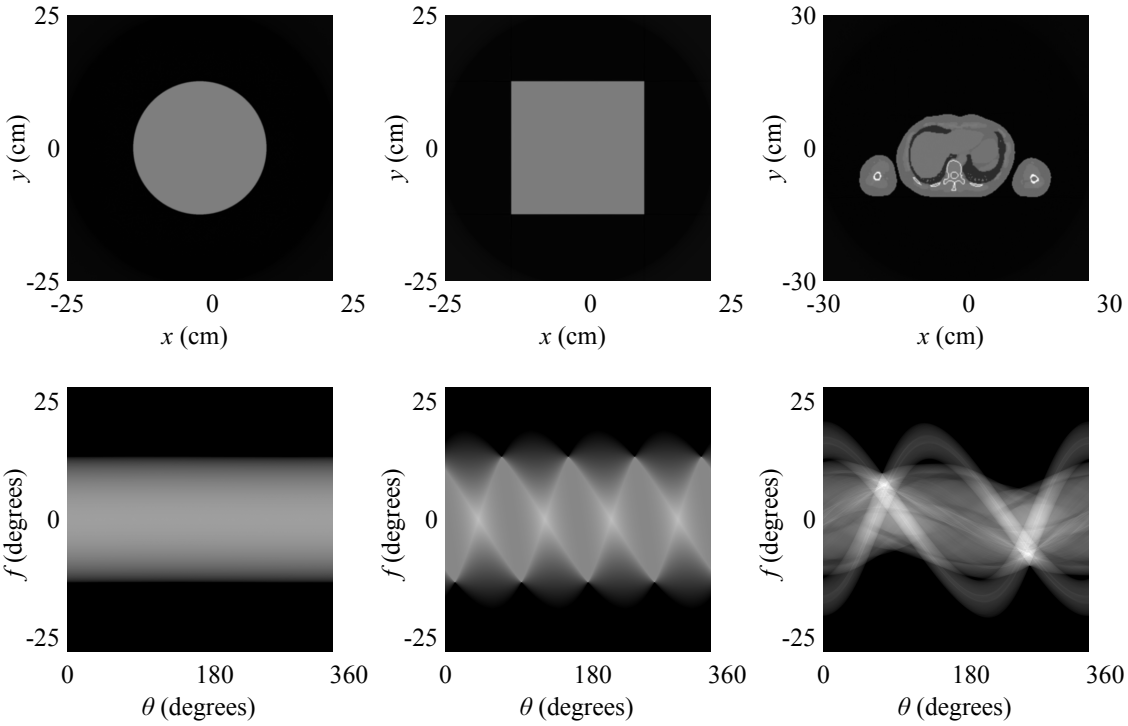


Figure 2.3 – X-ray attenuation maps for three virtual phantoms (top) and the corresponding attenuation sums for each projection used in CT (bottom). The latter are known as sinograms. No azimuthal variation is seen in the sinogram of the circular phantom due to its rotational symmetry, but a gradual reduction in intensity is seen in the fan-beam direction as the length of intersection of off-centre ray paths decreases. Rays which intersect the corners are particularly evident in the sinogram of the square because they provide the maximal intersection length for any particular value of θ. Global maxima occur at $\theta = 45°, 135°...$ when the central ray passes through both corners. The sinogram for the anthropomorphic phantom shows a broad band originating from the torso, orbited by two narrow sinusoidal bands that arise from the arms. Global maxima occur due to ray paths encompassing the rear of the ribcage, the spinal column and cortical bone in the arm, as well as a large amount of soft tissue.

2.3 Reconstruction

If a slice is considered as many discrete volume elements (voxels) labelled by i and j, a series of simultaneous linear equations links the sinogram to individual attenuation coefficients:

$$\zeta(\theta, f) = \sum_i \sum_j \mu_{i,j} \cdot d_{i,j}(\theta, f) \tag{2.4}$$

where $d_{i,j}(\theta, f)$ is the distance traversed inside voxel (i, j) by the projection at (θ, f). Such a system of equations can in general be solved if the number of unknowns to be determined is less than or equal to the number of independent equations. In most fan-beam geometries, this will be true if the total number of values sampled in the θ and f dimensions are slightly greater than the number of pixels to be reconstructed in the x and y directions.

There are numerous approaches to the reconstruction of images from sinograms for which the mathematical detail shall not be examined here – many good references exist, such as chapter 3 of Kak & Slaney [28]. The number of simultaneous equations defined by expression (2.4) in a single slice of a high-resolution scan typically approaches 10^6 – such systems can be solved by the inversion or decomposition of matrices with a number of elements equal to the square of this value. Even though the matrices in question are quite sparse[2], algebraic reconstruction is a significant computational burden. Approximate reconstruction methods are thus popular, in particular the back projection approach, which can be summarised as follows (illustrated in figure 2.4):

1. For each CT slice, a grid of pixels is defined in $x - y$ space, which will provide the geometry for the reconstructed image. All pixels begin with an intensity of zero.

2. For each sample in the measured sinogram, the corresponding line SD is projected through the image reconstruction grid. The intensity of each pixel i, j intersected by this line is increased by a value equal to $\zeta(\theta, f)/d_{i,j}(\theta, f)$. Geometric factors may also be necessary if the set of lines SD are not parallel – in the case of third-generation CT scanners this would be an inverse square correction.

Step 2 results in an image with considerable low spatial frequency smearing, caused by the equal distribution of projection intensity to all pixels along the line. This can

[2]The proportion of non-zero elements being less than $\frac{1}{n}$, where n is the number of pixels per side for a square reconstruction grid.

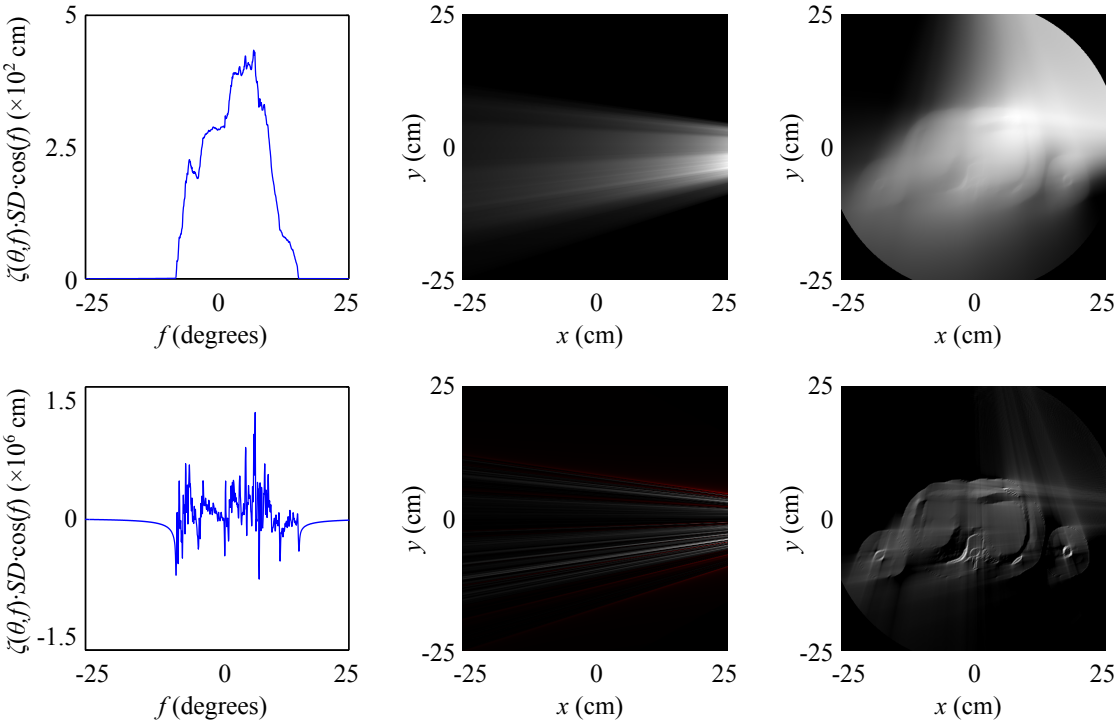


Figure 2.4 – Reconstruction using the back-projection method without (top) and with (bottom) application of a filter to remove low spatial frequency data from projections. Whilst the resulting projection values (left) differ considerably in amplitude, the integral values for a single view are comparable. Individual back-projections (middle, negative values shown in red) are summed to obtain the reconstructed image (right). Figure represents the state of the reconstruction at $\theta = 90°$ i.e. at 25% completion.

be avoided by pre-filtering the projections to remove redundant information. Typical filters operate in the angular spatial frequency domain and are constructed as the multiplication of a window function with the Fourier transform of the projection. Use of such functions causes the image to effectively be built-up in terms of its edges, leading to a much sharper representation. A sensible choice for the window function is a linear ramp between 0 and the Nyquist frequency (the maximum frequency that it is possible to represent), although this may be modified to smooth projection data thereby adjusting the character of spatial noise in the reconstructed image.

2.4 Hounsfield Units

The LACs calculated in the CT reconstruction procedure are typically rescaled such that they are in Hounsfield units (HU), at which point the quantity is referred to as *CT number*. The Hounsfield scale assigns a value of -1000 to air and a value of 0 to water. Thus in a scanner whose reference measurements are taken in air, the mapping

between μ and CT number is:

$$h\,[\mathrm{HU}] = 1000\left(-1 + \frac{\mu - \mu_{\mathrm{air}}}{\mu_{\mathrm{water}} - \mu_{\mathrm{air}}}\right) \tag{2.5}$$

The bulk of soft tissues and fluids in the body typically lie in the range ±100 HU, whilst lung tissue sits below −500 HU and bony tissues occupy the range 500 to 2000+ HU. As μ is a function of x-ray energy, so too is h.

CT images have a dynamic range greater than traditional 8-bit computer displays, which can show a maximum of 256 grey levels. Only a small part of the HU range is typically of interest for a particular examination and it is difficult to detect detail within low contrast images using the human eye[3]. As a result, images are typically viewed with a portion of the attenuation range expanded to the full dynamic range of the display. This process, known as *windowing*, was described in Geoffrey Hounsfield's original paper on CT [30].

The well-adopted DICOM medical imaging standard specifies a window by its width W and midpoint (or level) L [29, Part 3 C.11.2.1.2]. On an 8-bit display, the grey level Y of a pixel with CT number h is given by:

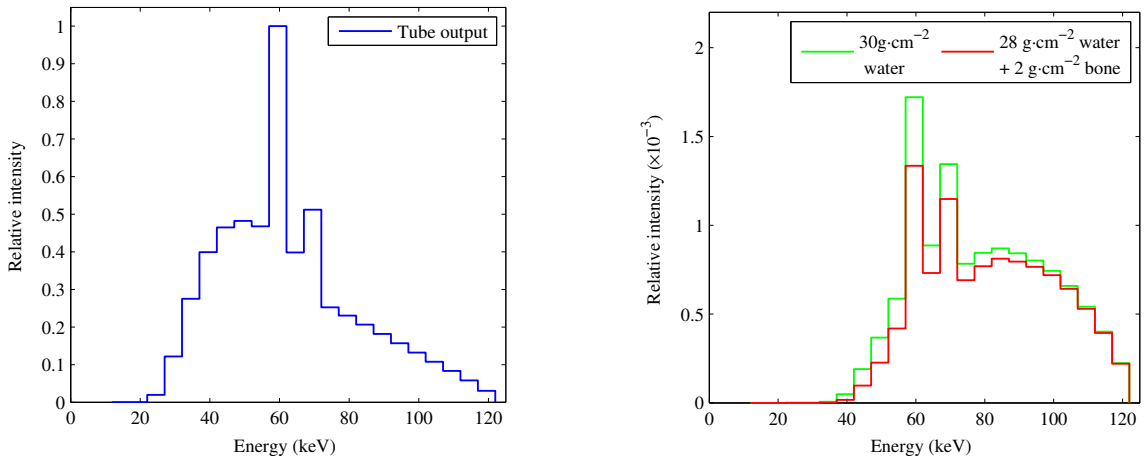
$$Y = \begin{cases} 0 & h \leq L - (W/2) \\ \left\lfloor 255 \cdot \left(\frac{h-L-0.5}{W-1}\right) + 0.5) \right\rceil & L - (W/2) < h < L + (W/2) \\ 255 & h > L + (W/2) \end{cases}$$

When viewing DICOM CT images, the window is set interactively by the viewer. Whilst only a small attenuation range may be visible to the viewer at a point in time, the whole range remains accessible.

2.5 Beam hardening

The treatment above describes the computed tomography process for a situation in which the incident x-ray source is monochromatic. Unfortunately, it is currently impractical and uneconomic to utilise near-monochromatic sources in clinical situations. Instead, the radiation emitted by a typical CT x-ray tube has a broad energy spectrum, arising from Bremsstrahlung and inverse photoelectric (characteristic x-ray) processes, as figure 2.5.

[3]The number of grey levels distinguishable by the human eye quoted in the radiology literature is often in the range 12-60. At typical display brightnesses (250 $\mathrm{cd/m^2}$), physical modelling of the human visual system suggests 700 shades can be discerned at a 'just noticeable' level [29, Part 14 Annex B].



(a) **(b)**

Figure 2.5 – Simulated x-ray energy spectra representative of a clinical CT scanner at 120 $\mathrm{kV_p}$ (a) before and (b) after traversing matter. It can be seen that the same radiological thickness of different materials can result in different levels of spectral attenuation. The beam leaving the phantom with bone is considered 'harder' than that leaving the water-only phantom as its mean energy is greater. The initial spectrum was generated with SpekCalc [31] and attenuated using coefficients tabulated by NIST (the US National Institute of Standards and Technology) [25].

In general, equations (2.1) to (2.4) only hold when I constitutes a specific spectral intensity, which is to say that μ is a function of energy, as are the scattering and photoelectric cross-sections that underlie its value. If this function and the spectrum are known, it is still possible to predict the response of an intensity-proportional detector to an attenuated beam:

$$\frac{I}{I_0} = \int P(E) \cdot \exp\left[-\int_{SD} \mu(\mathbf{r}, E) \cdot \mathrm{d}\mathbf{r}\right] \mathrm{d}E$$

The effect of objects comparable to the human body on the energy spectrum of the beam can be observed in figure 2.5. It is evident that the mean energy of the beam increases significantly between entering the object and leaving it, and even more so when some of the water in that object is replaced with an equivalent mass of bone. The implications for image formation in CT can be revealed by considering the behaviour of $\mu(E)$, the quantity which is imaged.

There are two classes of image artefacts that result from beam hardening effects in CT. It may be helpful to categorise them as:

- C-type artefacts (mostly observed as *cupping*), whose origin is a gross change in the beam energy spectrum with depth

- S-type artefacts (mostly observed as *streaking*), which occur due to the fact that the precise effect on the energy spectrum varies between materials

These artefacts are labelled by their predominant visual characteristics, but it is possible to contrive situations where C-type artefacts produce streaks in images and S-type artefacts result in a cupping effect. The distinction is nevertheless useful because different methods are used correcting for the two classes of artefact.

2.5.1 C-Type artefacts

Figure 2.6 shows the x-ray mass attenuation coefficients (MACs) as a function of energy in both water and cortical bone. The LAC for a material is the product of its MAC and mass density. Both materials display a decline of attenuation coefficient with energy in the region occupied by the energy spectrum, so one would expect a decrease in the pixel value at points which experience a harder beam (i.e. those at greater depth) when projections are naïvely reconstructed. This is indeed observed, as simulated in figure 2.7, and is known as a *cupping* artefact. Cupping is the only beam-hardening artefact one would expect for a cylindrically-symmetric object, regardless of its composition.

With no beam hardening correction, the magnitude of cupping artefacts observed has a strong dependence on the mass density of the material present and only a weak dependence on the composition of the material. For the purposes of this thesis, the C-type beam hardening artefact will be defined as that which would be observed if the scanned object consisted purely of water at different densities[4]. C-type artefacts can be corrected in a relatively simple manner, which will be discussed in Section 3.3.3.1.

2.5.2 S-Type artefacts

Figure 2.6 reveals that the MAC of bone can be somewhat different to that of water. If the MAC as a function of energy for bone were linked to that of water by a simple multiplicative constant c, it would be possible for bone to masquerade as water of density $c \cdot \rho_{\text{bone}}$ – indeed the two would be indistinguishable. In reality, however, the curves have different shapes, with bone displaying a proportionally greater attenuation coefficient at low energies. It is therefore impossible for any thickness of bone to have an equivalent effect on the x-ray spectrum as a given thickness of water. Any water

[4]There is no fundamental reason why the C-type artefact must be defined with reference to water. It is nevertheless useful to do so as most human tissues are sufficiently similar to water that they may be considered as perturbations.

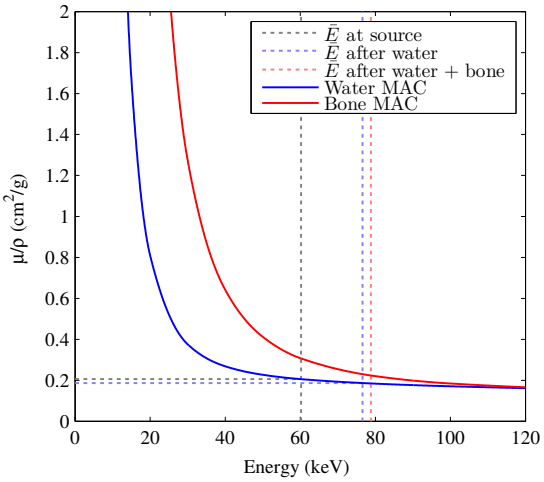


Figure 2.6 – X-ray mass attenuation coefficients as a function of energy for the materials used in the generation of figure 2.5.

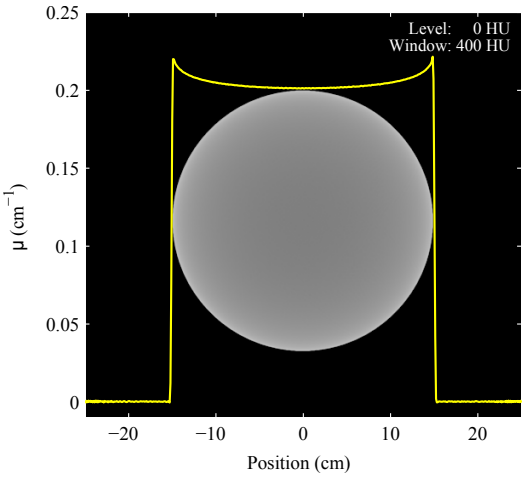


Figure 2.7 – Demonstration of cupping artefact in a simulated CT scan of a 30 cm water phantom at 120 kV_p. Image is overlaid by a central lateral profile.

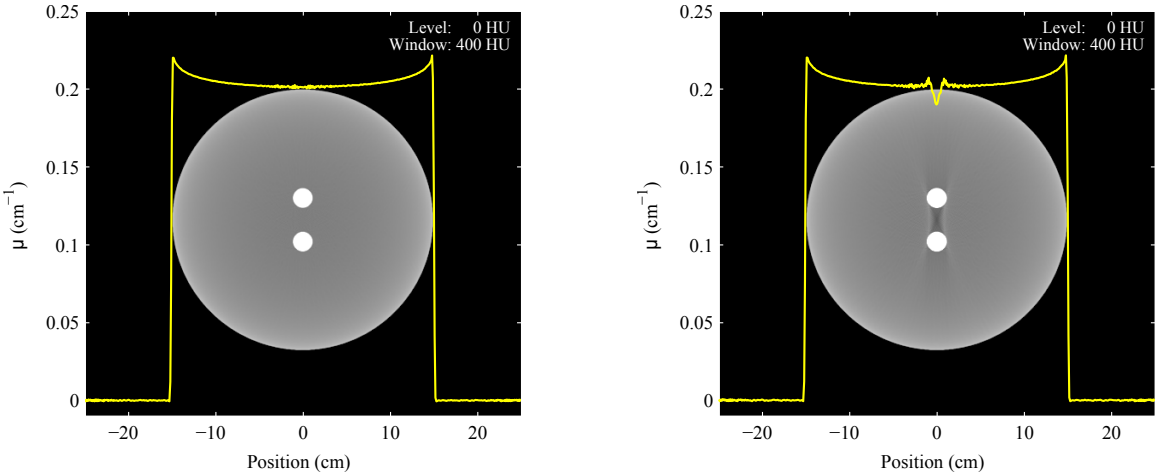
downstream of bone will see a harder beam than expected if the same mass of water were traversed and will thus display a lower attenuation coefficient. This is typically observed as a *streaking* artefact, as illustrated in figure 2.8.

Similar arguments to that in the previous paragraph apply whenever there is more than one material within the scanned object. Each material will have a slightly different effect on the x-ray spectrum and all whose MAC vs. energy is sufficiently different to water[5] will contribute some level of S-type artefact. S-type beam hardening artefacts will thus be defined as those that occur due to variations in material composition from that of water. S-type artefacts may be partially corrected, but assumptions must be made regarding the materials present in the scanned object – this will be discussed in more detail later.

2.5.3 Hounsfield Units revisited

It has been established that, in CT, the LAC associated with a material at a location in space is a function of the x-ray spectrum at that point. To a large extent, this spectrum is controlled by the voltage applied across the x-ray tube, the composition of its target and the characteristics of any filters present in the source. These factors vary between models of CT scanners and can even be specified differently for individual

[5]As before, the selection of reference in defining the S-type artefact is arbitrary, but water is a helpful choice.



(a) Rods composed of water at $1.92\,\mathrm{g/cm^3}$

(b) Rods composed of cortical bone at $1.92\,\mathrm{g/cm^3}$

Figure 2.8 – Simulated CT scans of a 30 cm water phantom containing a pair of 2 cm dense rods positioned with their centres 5 cm apart. Central lateral profiles are overlaid on the images. (a) displays only cupping as a single material is present whilst the presence of bone in (b) results in a streaking artefact. Source energy spectrum representative of a clinical scanner at $120\,\mathrm{kV_p}$.

protocols on the same scanner. Caution must clearly be exercised when using the Hounsfield scale for quantitative intercomparison.

However, the preceding brief overview of the effects of beam hardening effects in CT leads to the conclusion that the definition of Hounsfield Units in section §2.4 can not be entirely rigorous, even for a known source spectrum. A point within an object will be traversed by multiple projected rays and the spectrum it experiences is thus affected by the thickness and composition of preceding material along each. If μ_{water} and μ_{air} in equation (2.5) are intended to be constants, the geometry from which they are derived must be clearly stated. Even then, without complete correction for beam hardening, the CT number of a defined material within a CT slice is dependent on both the position of that material and all other surrounding material in that slice. Consequently, even within the same scan, it is necessary to critically assess any quantitative use of CT numbers.

Figure 2.8 shows an example of a region of water whose CT number drops by up to 5.7% as a result of an S-type artefact. Whilst the streak may be easily identifiable in this simple geometry, human anatomy is much more diverse and there are examples of beam hardening artefacts causing diagnostic challenges in the clinic. Chapter 5 will assess whether beam hardening may also have significant impact on the radiation transport calculations involved in planning proton therapy treatments.

2.6 Other sources of image artefacts

Whilst the discussion which follows in the next chapters is mostly limited to examining the impact of beam hardening artefacts, it is important to be aware of other potential sources of variations in CT number. Barrett and Keat [32] have produced a comprehensive summary with visual examples of the artefacts commonly observed in modern clinical scanners. A discussion of their origins, with the inclusion of an additional effect (scattering), follows. In reality many of these artefacts arise together in images which already contain complex patterns – their compounded effects can at times be indistinguishable from true anatomy, presenting challenges for diagnostics and radiotherapy treatment planning. Computer simulation of the CT process, which will be discussed in the following Chapter, allows artefacts to be examined in isolation and has been used to produce the Figures in this Section (as well as Figures 2.3, 2.4, 2.7 and 2.8).

2.6.1 Physics

The following image artefacts result from the fundamental physical processes governing x-ray attenuation.

2.6.1.1 Beam hardening

As described in section §2.5.

2.6.1.2 Photon starvation

Whilst the computed tomography process assumes that the x-ray linear attenuation coefficient is well-defined at any point in space, the laws of statistics determine the precision with which it may actually be measured. The Beer-Lambert Law (equation (2.1)) applies at the population-average level, but the interaction of an individual photon is a stochastic process so statistical variations in measured intensity should be expected for any finite number of source particles. The probability of successful traversal for each photon is independent of the others so, in the simplified case where detection efficiency η is independent of photon energy, the number of counts measured after the incidence of N_0 photons could be expected to follow a Poisson distribution with parameter $\lambda = \eta N_0 \frac{I}{I_0}$. As the variance of a Poisson distribution is known to be equal to λ, the standard deviation of counts measured will be proportional to the square roots of both N_0 and $\frac{I}{I_0}$. It is this variation that will cause statistical noise in the image.

Small intensities show the greatest fractional variation due to statistical effects. The reconstructed quantity is ζ, the negative logarithm of $\frac{I}{I_0}$, and $\frac{\mathrm{d}\zeta}{\mathrm{d}I}$ is asymptotic to $I = 0$. Consequently image noise increases drastically as intensity falls, in particular once the number of photons detected per detector element per view drops below about 1000 – this situation is termed *photon starvation*. A general lack of photon fluence results in noise throughout the image, as demonstrated in figure 2.9, although the magnitude of this noise is very much dependent on the specifics of the software filtering process that typically takes place during image reconstruction. If particularly high density regions exist within the object, photon starvation can also affect only a portion of the projections: images afflicted in this way display hypo- and hyper-intense stripes along the lines of greatest attenuation, which are demonstrated in figure 2.10.

Photon starvation can be a particular problem in large patients and slices containing much bone (or other highly-attenuating materials like metal implants). It can be avoided by increasing the radiation exposure for the scan, but a balance must be struck between image degradation and the radiation dose received by the patient, which has been associated with an increased risk of new cancer. The application of smoothing algorithms to the projection-level data can also assist in reducing the visual impact of statistical noise[6].

2.6.1.3 Scattering

The Beer-Lambert Law describes the attenuation of a primary x-ray beam, but attenuation does not imply absorption. Many of the photons which leave the primary beam do so in elastic or inelastic scattering events, which leave them following a different trajectory but remaining detectable. In fact, as can be seen by the comparison of the attenuation and energy-absorption coefficients in figure 2.11, high elasticity scattering is the dominant process at typical CT energies. Many of the scattered photons will not reach the detector: their intensity continues to be attenuated by the Beer-Lambert Law, whilst geometric factors also mitigate their detection probability. Although there is some forward bias to the scatter distribution, a single-slice detector subtends a small solid angle so the ratio of scattered to primary photons is lower than a naïve reading of figure 2.11 might suggest – a value of 10% or less might be expected for clinical scans of a 25 cm water phantom [33].

[6]Smoothing removes the high frequency content of a projection, reducing its weighting in the filtered back-projection. It is likely to affect the spatial resolution of the image along lines of high attenuation.

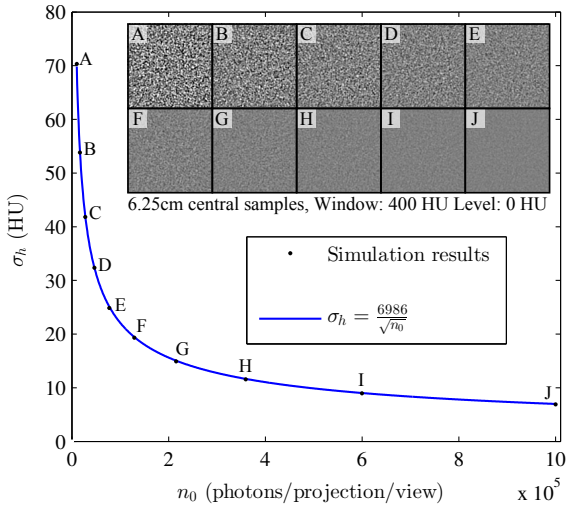


Figure 2.9 – Standard deviation of simulated CT numbers in a 30 cm diameter water phantom for varying numbers of incident photons. Monoenergetic source with energy 80 keV. HU measurement resolution is limited by the central limit theorem, as demonstrated in the fit to $\frac{1}{\sqrt{n_0}}$.

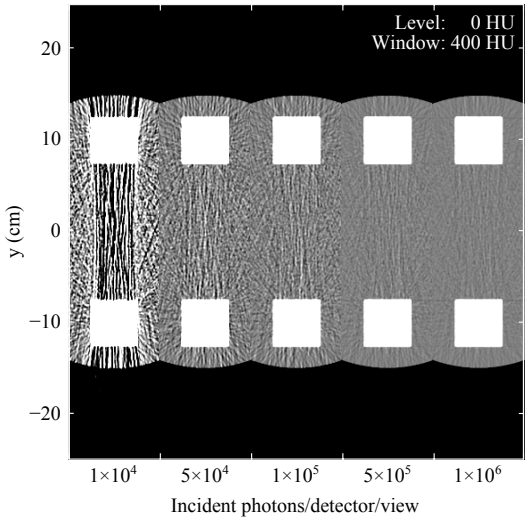


Figure 2.10 – Simulations of photon starvation noise for varying numbers of incident photons. Geometry consists of two 5 cm square blocks of water at 2.5 g/cm^3 within a 30 cm diameter water phantom at 1.0 g/cm^3. Monoenergetic source with energy 80 keV.

The artefacts produced by scattering have a tendency to mimic beam hardening in reconstructed images. The relatively isotropic distribution of scattered photons causes a background signal that is approximately isotropic. Due to the non-linear relationship between ζ and I, high attenuation projections are 'darkened' more than low attenuation ones, leading to streaking between high density objects. Cupping may be expected to occur in images of a cylindrical phantom for similar reasons.

The effects of scattering are minimised in clinical devices through software and hardware approaches. As the scatter distribution is of low spatial frequency, it can be adequately modelled with few parameters and the resulting calculations used to correct measured projections. To avoid measuring scatter in the first place, detectors feature anti-scatter grids – individual collimators for each detector element, which reduce their angular acceptance to a small range around the incoming trajectory of the primary beam. The amount of scatter occurring in each view can be minimised through use of a bowtie filter – a shaped beam attenuator, which is thicker at the extremities of a fan beam than the centre – after the x-ray source. This causes fewer photons to take short paths within the object, leading to less scatter, with signal-to-noise ratios that

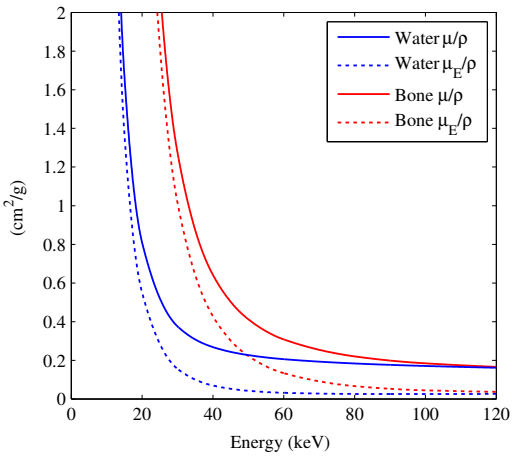


Figure 2.11 – Comparison of the primary x-ray beam attenuation and energy absorption coefficients for water and bone in the CT energy regime. Data from NIST [25].

are nevertheless equivalent to those in central projections[7].

2.6.2 Sampling

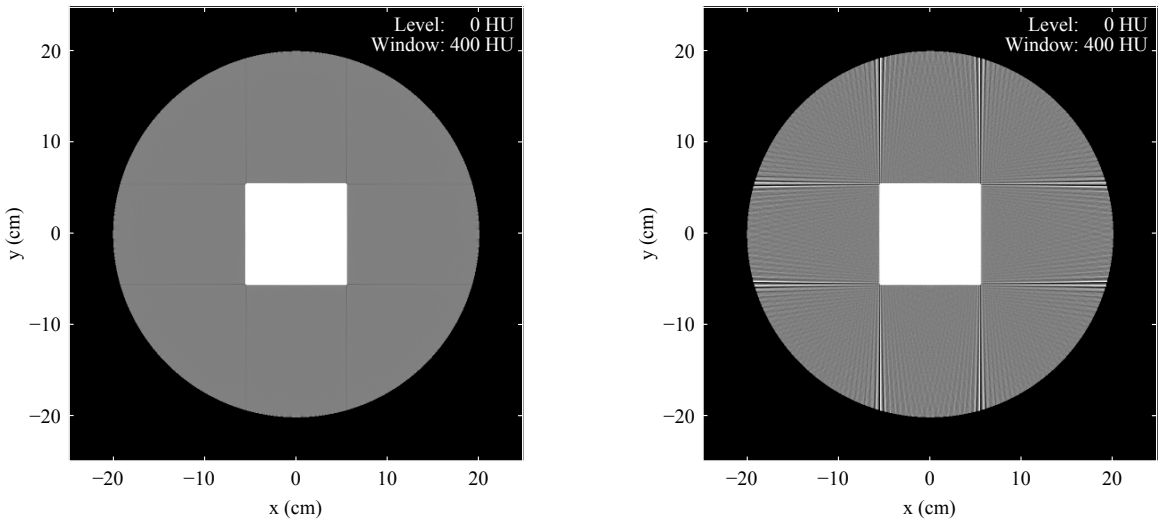
2.6.2.1 Aliasing

Technological and economic factors place limitations on the capabilities of commercial CT scanners, in particular on the number of detector elements and the speed at which the measured data can be processed. These constrain the number of views per image slice and the number of projections per view. The resulting sampling density is usually sufficient for millimetre-scale voxels, but sharp changes in attenuation parallel to projections or between views can result in image artefacts known as *aliasing*. Aliasing artefacts at the projection level appear as stripes that seem to be a 'continuation' of an object's edges beyond its extent. At the view level, aliasing results in a small fan of stripes emanating from the edge. The periodic nature of view aliasing artefacts originates in the filtration step of back-projection, in which projection data is convolved with a periodic function. Aliasing is illustrated in figure 2.12.

2.6.2.2 Incomplete projections

Circle A in figure 2.1 encompasses all points which can contribute to a projection at every view angle θ. If part of the object extends outside the circle, it will contribute to projections from some directions but not from others, leading to sinograms that

[7]Bowtie filters also have the secondary benefit of reducing patient dose due to this decrease in exposure at the extremities.



(a) 3,936 views with 228 projections/view

(b) 246 views with 3,648 projections/view

Figure 2.12 – Simulated CT scans of a square block of density $2.0\,\mathrm{g/cm^3}$ within a circular phantom of density $1.0\,\mathrm{g/cm^3}$, demonstrating aliasing artefacts. (a) shows ray aliasing and (b) shows view aliasing. In order to highlight only these artefacts, the simulated materials are composed of water and the source is monoenergetic.

are not logically consistent[8]. The effect in reconstruction is to produce a low quality representation of the portion of the object lying outside circle A, which is lacking in detail along the unsampled directions, surrounded by considerable smearing. The overall effect on the image is considerably worse if a contiguous region (e.g. the torso) extends beyond the area, compared to if a separate entity (e.g. an arm) is placed wholly outside; in the latter case, image degradation is mostly restricted to within that entity – although there will, of course, be an additional contribution to beam hardening and photon starvation artefacts in the main field of view.

2.6.2.3 Partial volume averaging

Partial volume averaging occurs along the line of the projections due to the additive nature of the reconstructed quantity ζ. If heterogeneities exist that are smaller than the reconstructed voxel, the CT number will be representative of the volume average of the voxel contents in the imaging plane[9].

Partial volume averaging can be minimised by using a reconstruction grid with a

[8]In an idealised situation with a monoenergetic beam encompassing the whole of the object, the sum of projections ζ over all fan angles f should be the same at each view angle θ.

[9]Notwithstanding beam hardening, whose effects will be negligible on a single voxel scale

larger number of voxels; however, without also increasing the sampling density in the f and θ directions, this also increases the degree of aliasing.

2.6.2.4 Non-linear partial volume artefacts

Complex non-linear artefacts occur if a heterogeneity partially intrudes into the imaging plane with its 'shadow' only covering part of the height of a detector element. In the situation that a fraction x of the slice thickness is occupied by material A and the rest by material B, the reduction in intensity at the detector as a result of an object of thickness d will be $x \cdot \mathrm{e}^{-\mu_A d} + (1 - x) \cdot \mathrm{e}^{-\mu_B d}$. As illustrated in figure 2.13, compared to the volume averaged situation, the effect of the change in CT number is disproportionately greater, peaking when the object intrudes approximately halfway into the slice. In general one would expect the intersection length d to vary between projections through the object, so the resulting back-projected image will show non-uniform CT numbers even if the region has uniform average density, as demonstrated in figure 2.14.

When using standard back-projection techniques, the magnitude of partial volume artefacts increases as the assumption of a thin two-dimensional slice becomes less valid. Improvements can be achieved by using thinner slices that are carefully positioned; however, as partial volume effects are essentially a result of quantisation, they will always be an issue for detectors comprising discrete elements.

Non-linear partial volume artefacts could theoretically occur in the f and θ directions; however, in practice the in-plane sampling density is very high and the effect is insignificant in this dimension.

2.6.2.5 Motion

Motion artefacts can be considered inadequate sampling in the temporal domain. The description of CT in this chapter is predicated on the existence of an attenuation distribution that remains constant during data acquisition. If this is not the case, the reconstructed image will be a linear superposition of the partial filtered back-projections at each timepoint along the path of the source. The visual characteristics of the artefacts vary depending upon the direction of the motion in question, but typically include streaking and some level of temporal averaging. Motion artefacts are minimised by discouraging movement whilst scanning and keeping scan times short, but involuntary motion remains an issue in clinical scans. Software corrections can assist, as well as gating – a practice in which data is only acquired in a certain phase of a motion cycle.

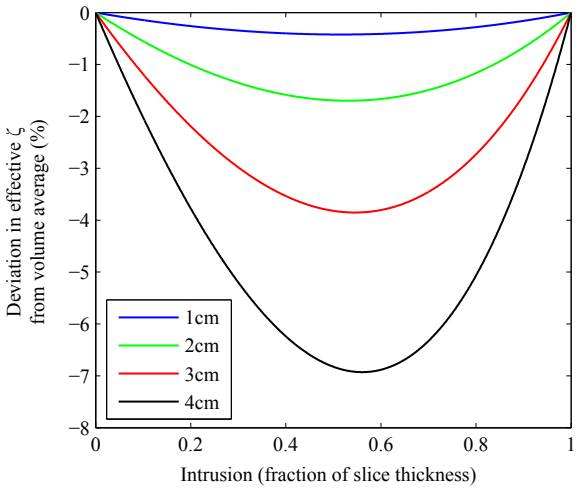


Figure 2.13 – Differences in the calculated effective projection values of slices containing high-density partial intrusions from that which would be expected by volume averaging. Lines represent differing lengths of water at $2\,\mathrm{g/cm^3}$ within 30 cm of water at $1\,\mathrm{g/cm^3}$. Monoenergetic source with energy 80 keV.

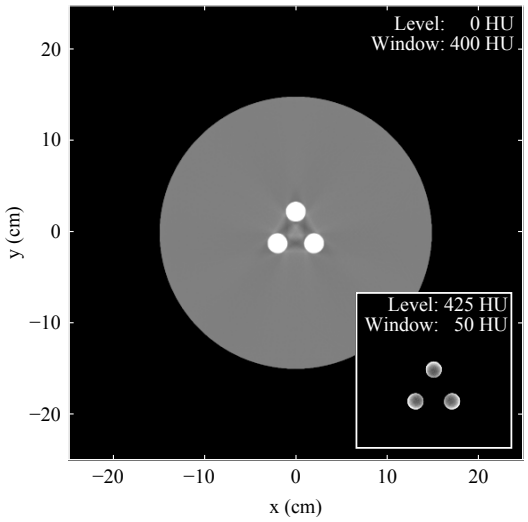


Figure 2.14 – A simulation of the partial volume artefacts observed when three 2 cm diameter rods of water at density $2\,\mathrm{g/cm^3}$ intrude halfway into a slice of a 30 cm diameter water phantom. Inset shows the rods with a window emphasising the non-linear nature of these artefacts. Monoenergetic source with energy 80 keV.

Four-dimensional CT aims to capture motion in a series of 3D frames, using the latest generation of scanners which can acquire multiple slices in a single rapid gantry rotation. It has utility in planning the radiotherapy of particularly mobile tumours (e.g. in the lung), as well as a number of diagnostic uses.

2.6.3 Source/detector geometry

Modern CT scanners include modes of operation that improve the resolution and speed of CT by acquiring multiple slices in a single gantry rotation. These modes can display the artefacts already discussed in a slightly more complex manner, but they also introduce their own issues.

2.6.3.1 Helical artefacts

In helical or spiral CT, the patient couch is constantly moving while projections are acquired – this means that the 3D volume is sampled in a helical manner rather than in slices as in conventional axial CT. Nevertheless, current radiotherapy treatment planning and diagnostic paradigms require stacks of 2D slices, so interpolation is utilised to produce conventional sinograms from this data. As changes in the axial cross-section

with z are not always entirely predictable, slices produced in this manner are liable to resemble weighted averages of data originating above and below the location of interest, with a weighting function that varies with θ. It is thus not uncommon to see abrupt changes in cross section (e.g. the diaphragm) appear gradated across a slice.

2.6.3.2 Cone-beam artefacts

Multi-detector CT (MDCT) scanners employ additional rows of detector elements, which they irradiate using a cone beam with a mild inclination in the z direction. Each row of detectors can provide an image slice when operating in axial mode, or a helical trajectory can be used to reconstruct an even greater number. Artefacts can occur in images derived from detector rows at the edge of cone beams because the geometry is sufficiently far removed from the 2D situation assumed in conventional reconstruction. When a ray is inclined, its length of intersection is greater than that in two-dimensions. If the object were sufficiently long and the detector row of infinitesimally-small height, correction would simply require adjusting ζ with a constant multiplicative factor. However, partial volume effects become apparent when the ray intersects the object's boundaries in the axial direction or when there is an appreciable difference in path length of rays reaching the top and bottom of the detector. After reconstruction, artefacts comparable to those in figure 2.14 may thus be observed. Cone-beam artefacts can be minimised by taking a 3D approach to reconstruction.

2.6.4 Equipment malfunction/miscalibration

Artefacts may also arise due to miscalibration or malfunction of components of the scanner. There are many possible failure modes, but Barrett and Keat specifically highlight the miscalibration of individual detector elements in the scanner, whose resulting artefacts are observed in reconstructed images as rings centred on the isocenter.

Chapter 3

A software CT scanner

In order to examine the impact of CT image artefacts on the calculations performed in proton radiotherapy treatment planning, the relevant artefacts must be identified. As previously discussed, real clinical images contain a great degree of intensity variation representative of the patient's anatomy, from which the contribution of artefacts can not always be distinguished. It is thus desirable to carry out the study in a situation where 'ideal' images can also be derived such that artefacts can be identified by direct comparison. Idealised images can only be produced with complete knowledge of the workings of the scanner and the patient's internal geometry and material composition. However, clinical CT scanners are complex commercial products that utilise proprietary components and methods, presenting a barrier to their full emulation, and the patient's internal anatomy will not be known in great detail – indeed it is the purpose of CT is to derive it. The solution adopted here is to construct a simplified computer simulation of a CT scanner whose output is comparable to that of a real scanner. Simulated images are then produced from a computer-based model of the anatomy of interest, known as a *virtual phantom*.

3.1 Virtual phantoms

Many virtual phantoms have been constructed to represent the human body, over one hundred of which are detailed in Xu and Eckerman's extensive handbook [34]. In order to use a phantom for the simulation of polychromatic CT, it must be specified in a manner that allows the x-ray LAC at the energies of interest to be derived at any point in space. An example set of sufficient information is mass density and chemical elemental composition, from which LACs may be calculated using tabulated data from NIST.

3.1.1 Classes of geometry

A phantom's geometry specification delineates various regions of space, each of which corresponds to a particular volume of interest (e.g. the external surface of the body, or an organ such as the left kidney). These regions can have properties (e.g. mass density, elemental composition) assigned to them through use of look-up tables. Whilst individual examples of virtual phantoms need not be examined at this point, it is necessary to consider the ways in which phantom geometry may be specified, as this affects the mathematical complexity of simulating CT.

- *Stylized geometric phantoms* were the first class of phantoms used in radiation dosimetry from the 1960s onwards. They are constructed from geometrical primitives, such as cones, spheres and planes, which can be defined using systems of second-order polynomial equations. Boolean combinations of these primitives are used to define the boundary of the phantom and its internal structures. This approach had great advantage at a time when memory capacity and processing power were limited, as relatively little information is needed to store them and exact intersection lengths are easily calculable. However, whilst some parameters of phantoms in this class can be based on measurements (e.g. total body mass), it is difficult to produce phantoms that accurately represent large portions of the anatomy using these primitives. First-generation geometric phantoms are thus stylized as in figure 3.1, aiming to replicate the gross positioning of organs within the given constraints. Given the lack of detail, the use of these stylized phantoms has declined as computing power has improved over the last two decades. Nevertheless, they continue to underpin much radiation protection knowledge, for example in the latest estimates of doses received by survivors of the atomic bombs at Hiroshima and Nagasaki [35].

- *Voxellised phantoms* began to be developed as computer memory capacity reached megabyte levels. The voxels in these phantoms, as in CT, are a 3D analogue of pixels and create a tessellating grid of cuboids[1] encompassing the entire volume of interest. These voxels are used as building blocks from which a model of external and internal anatomy is created. A benefit of working with voxels is that pre-existing voxellised imaging, such as CT and MRI, can easily be used as a reference leading to improvements in anatomical representation over the first-generation geometric phantoms (see figure 3.2). As technology develops,

[1]Other lattice systems could theoretically form a basis for voxels, but are not used in practice.

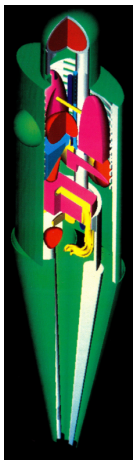


Figure 3.1 – The EVA phantom: an example of a stylized geometric phantom developed in the 1970s [37]

Figure 3.2 – The ICRP Adult Male reference computational phantom: an example of a voxellised phantom with moderate resolution (approx. 7 megavoxels) [38]

Figure 3.3 – Adult male from the UF family of reference hybrid phantoms, which combine NURBS (a form of BREP) and voxellised elements [39]

smaller voxels are being used and phantoms are now available with sub-millimetre resolution [36]. However, this accuracy comes at some cost - all regions must be specified at the same resolution, so much redundant information is stored in the case of large contiguous regions, leading to computational inefficiencies. Additionally, voxel geometries do not lend themselves well to deformation, which greatly increases the computational burden of 4D studies.

- *BREP* (boundary representation) *phantoms* are a new generation of analytically-specified phantoms whose geometry is given by methods that are much less constrained than early stylized techniques. These phantoms are modelled using modern 3D computer aided design concepts, defining the boundaries of regions by polygon meshes or NURBS[2] surfaces. Specifying a boundary, rather than the numerous points within it (as in the voxellised case), is memory efficient and facilitates fast geometric calculations. Motion can easily be incorporated into BREP phantoms using methods developed for computer animation. Figure 3.3 shows an example of a modern phantom containing BREP elements.

[2]Non-uniform rational basis splines, a method of defining curved surfaces by a series of control points and 'knot' vectors.

3.1.2 ICRP Publication 110 phantoms

The International Commission on Radiological Protection (ICRP) has a long history of producing authoritative publications pertaining to radiation interactions in the human body, including reference data on the composition of body tissues and anatomical models that have been the basis for virtual phantoms. Publication 110 [38] is the result of a collaboration with the ICRU to produce voxellised adult female (AF) and adult male (AM) virtual phantoms for use in radiation protection calculations.

The AF and AM phantoms are both based on pre-existing voxellised phantoms (known as Laura and Golem, respectively), whose anatomies were adjusted such that important metrics (for example height, weight, and the masses of individual organs and the skeleton) match the ICRP-advocated population averages for Caucasian adults. Laura and Golem were the result of a laborious manual organ segmentation process, utilising CT slices obtained in the course of whole-body irradiation treatment planning.

Whilst designed for radiation protection purposes, the ICRP 110 phantoms contain sufficient physical data to be of interest in simulation of computed tomography and proton radiotherapy. Figure 3.4 shows the format of the phantom information, available online as supplementary data to ICRP Publication 110. ICRP AF contains 14,255,124 voxels of size $1.78\,\text{mm} \times 1.78\,\text{mm} \times 4.84\,\text{mm}$; ICRP AM contains 7,161,276 voxels of size $2.14\,\text{mm} \times 2.14\,\text{mm} \times 8.00\,\text{mm}$. These phantoms were converted to a series of MATLAB data structures for use in later sections.

3.2 Existing CT simulation software

A number of software packages exist to simulate the computed tomography process. Some of the most prominent include:

- The free software packages CTSim [40] and learnCT [41], which perform forward- and back-projection in geometric phantoms, but do not simulate many artefacts seen in practice – as such, they are best suited for use as an educational tool.

- SINDBAD, a radiography simulation tool from France, whose phantoms may be specified in voxellised or geometric formats. It is able to model scatter through use of Monte Carlo methods, the effects of which have been assessed in a rasterised NURBS phantom [42]. Three-dimensional cone beam geometries and polyenergetic beams are possible. SINDBAD has not been released publicly.

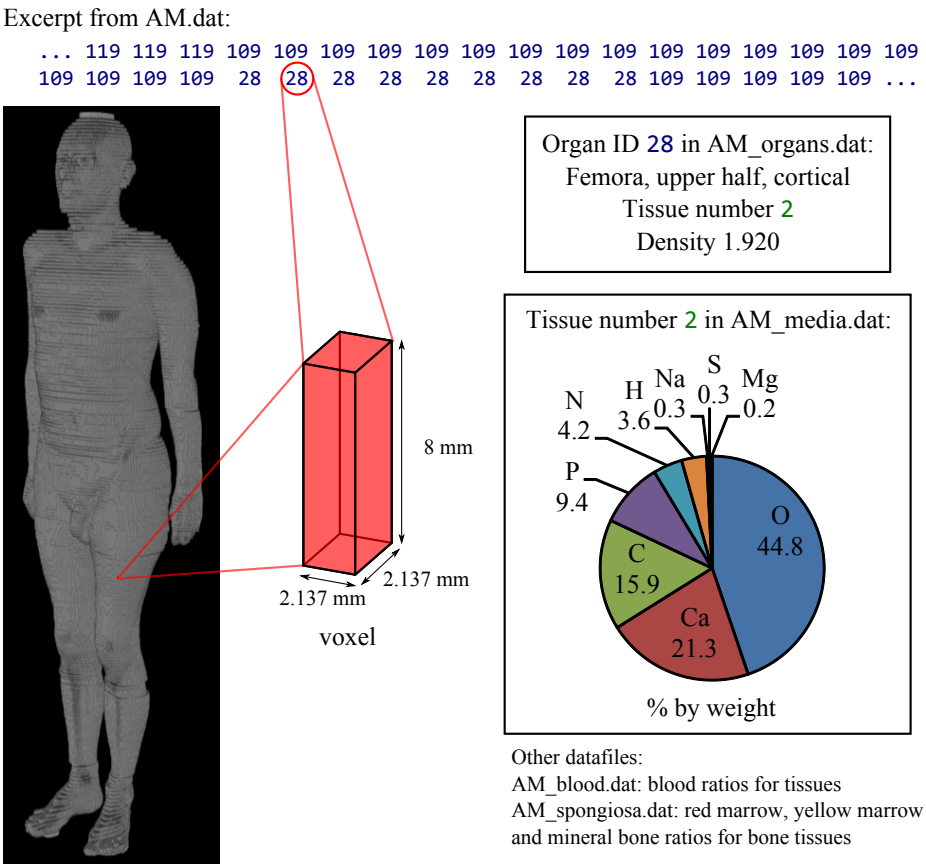


Figure 3.4 – Data format for the ICRP AM voxellised phantom. The phantom specification is stored in five ASCII text files: AM.dat contains a linear array of organ IDs for each voxel in the phantom; AM_organs.dat contains names, mass densities and tissue IDs for each organ ID; AM_media.dat contains names and percentage elemental composition by weight for each tissue ID; AM_blood.dat contains blood ratios for each tissue ID; AM_spongiosa.dat contains red marrow, yellow marrow and mineral bone ratios for each bone tissue ID. In order to construct a phantom, the number of voxels and spatial size of voxels in each dimension are also necessary, and these can be found in the publication itself. The AF phantom is specified identically, but has a different number and size of voxels.

- XRaySim [43] is an open-source C++ program that has been released in beta form. It can produce x-ray projections, including CT sinograms, from triangle-based (BREP) CAD models. However, it is currently mostly tooled for use in industrial non-destructive evaluation rather than medical imaging, and it is likely that the adaptation necessary for radiotherapy treatment planning investigations would involve significant work.

- SNARK09 [44] is a mature system, being a descendent of code used for CT research by Gabor Herman as early as 1983 [45]. It is open source and written in C++, with a bias towards the Linux operating system (some work is required to utilise its graphical features in Windows). SNARK works in the parallel and fan-beam source-detector geometries, with phantoms that are primarily specified by geometric primitives; however, these are rasterised internally in order to produce projections, so it is likely that little adaptation would be necessary to support voxellised phantoms. Photon statistics may be considered in the simulation process, as may energy-dependent LACs (with a spectrum containing a maximum of 7 energy bins). Whilst SNARK supports beam hardening correction for C- and S-type artefacts, the documentation suggests that there is little control over the process at the user level. SNARK09 showed a great deal of promise for investigations relevant to radiotherapy treatment planning, but it was felt that the learning curve would be rather steep, especially given that source-code modification would likely be necessary in order to gain sufficient levels of control, and additional software would need to be written in order to support its input and output formats.

- Jeff Fessler's Image Reconstruction Toolbox for MATLAB [46], which includes simulation and reconstruction functions for a wide range of medical imaging modalities. Its multiple functions relevant to CT include sophisticated iterative reconstruction methods that use the information inherent in photon statistics, and dual energy beam hardening correction. Unfortunately, it is not well-documented and familiarity with the source code would be necessary to modify it for use with custom scanner geometries and phantoms. The CT functions also require some binary components that are not distributed for Windows, and whose source code is only available with a licence.

Other relevant work has been carried out (for example, [47, 48, 49]) but none of the software used has had a general release. Recurring themes in assessing the pre-existing

options were insufficient documentation, inadequate scope and a lack of availability. A new piece of simulation software was ultimately written for the purposes of this thesis, allowing for complete control over the methods implemented, which are described in the following section.

3.3 Methods

The simulation code described here has been implemented in MATLAB [50], a flexible programming environment which allows for rapid iteration and debugging. MATLAB code is generally interpreted rather than compiled, which carries a speed penalty, but a variety of methods have been adopted to accelerate the simulation.

3.3.1 Phantom specification

Voxellised geometry forms the basis of the software CT scanner described in this chapter. The use of a single geometric primitive significantly simplifies the implementation and, with modern processing power, the speed penalty is tolerable. A large number of voxellised phantoms exist, some of which are publicly available, whilst the distribution of less mature BREP phantoms is currently limited. Adopting a voxellised approach to calculation does not completely preclude the use of BREP or geometric phantoms, as they may be converted to voxel maps of a chosen resolution by rasterisation.

In order to use a phantom in the simulation of CT, it must be possible to identify the x-ray LACs at all points within it. The voxellised phantom format used in this software is a MATLAB structure, containing an array of the material indices assigned to each voxel. A look-up table maps each material index to a percentage elemental composition and a mass density. This information enables a LAC (μ) to be calculated from the NIST tables of elemental MACs ($\frac{\mu}{\rho}$) using the mixture rule:

$$\frac{\mu}{\rho} = \sum_i w_i (\frac{\mu}{\rho})_i \tag{3.1}$$

where ρ is the mass density, i labels each element present and w_i is the fraction of an element in a material by weight.

The simulation software calculates a LAC for every material in the phantom at every energy in the input energy spectrum using equation (3.1), performing linear interpolation between the energies tabulated by NIST.

The mixture rule assumes that each atom acts as an independent absorber and that the effects of chemical bonds play no part in influencing μ. This assumption is

believed to be reasonable above energies of 10 keV (where the rule is correct within a few percent) except in the vicinity of absorption edges, whose position may exhibit minor shifts as the electronic environment is perturbed by chemistry [51]. Edge effects are unlikely to be an issue when considering the overall attenuation of a broad energy spectrum as in CT.

3.3.2 Projection algorithm

The MATLAB Image Processing Toolbox contains functions to perform forward-projection of pixel-based images within 2D CT-like geometries, in the form of the radon and fan-beam transforms. However, these functions have some very undesirable features when applied to simulating CT in a virtual phantom. In the standard MATLAB implementation of these functions, each image pixel is divided into four sub-pixels, and rays are traced through the centre of each sub-pixel. Unfortunately this presents a problem in the case of relatively low-resolution phantom data, as it seriously constrains the number of projections, causing unrealistic artefacts in reconstructed images. Additionally, the `fanbeam` function is implemented through a linear interpolation of parallel projection data to a fan-beam geometry, which will not be perfectly equivalent to direct fan-beam projection. These deficiencies are illustrated in figure 3.5. If a raytracing algorithm is to produce results comparable to a clinical scanner, regardless of phantom resolution, it is necessary to have full control over the number of rays used and the paths they take. The decision was thus taken to write a more flexible function to calculate $\zeta(\theta, f)$ within pixelmaps.

Unless otherwise stated, the simulations used to illustrate this section are created using the input parameters that will be determined to be representative of a clinical scanner in section §3.4.

3.3.2.1 Ray paths

The first stage of the simulation produces start and end coordinates for rays along the paths SD from figure 2.1. The values of θ and f to be sampled are calculated from a specific set of input parameters, such as those in table 3.2. For:

$$\Delta\theta = \frac{360^\circ}{\#\text{ views}},\ \theta_s = \{0, \Delta\theta, ..., 360^\circ - \Delta\theta\}$$

$$\Delta f = \frac{\text{Fan angle, } f_0}{\#\text{ detector cells}},\ f_s = \{-\frac{f_0}{2}, -\frac{f_0}{2} + \Delta f, ..., \frac{f_0}{2}\}$$

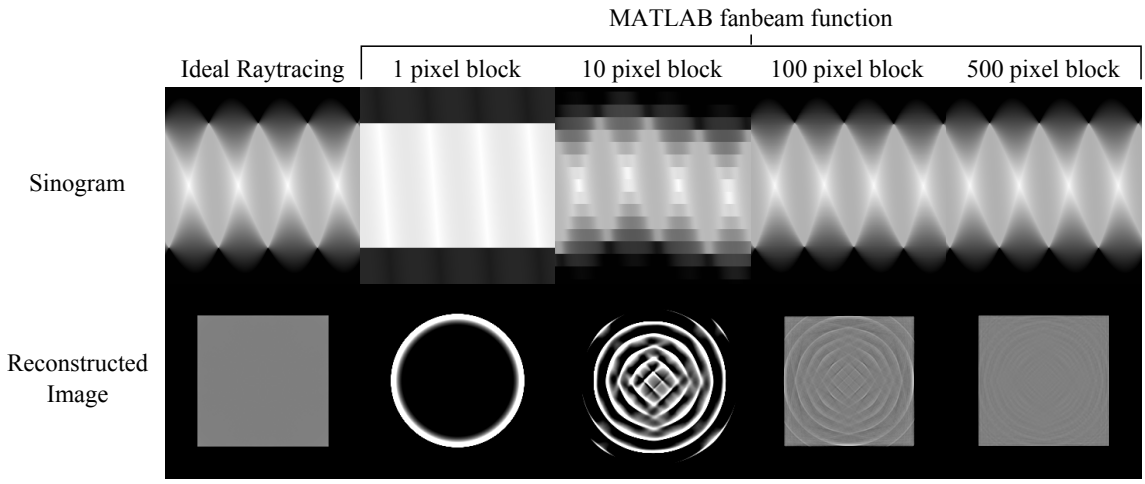


Figure 3.5 – The accuracy of the MATLAB Image Processing Toolbox `fanbeam` function for square water phantoms of various resolutions. All should appear identical (and match the ideal case), but significant quantisation is apparent in the f direction of the sinograms due to limitations imposed by the `radon` function. Reconstructions have Window 400 HU and Level 0 HU.

$$\begin{bmatrix} x_{start}(\theta_s, f_s) \\ y_{start}(\theta_s, f_s) \end{bmatrix} = \begin{bmatrix} -SI \cdot \cos(\theta_s) \\ -SI \cdot \sin(\theta_s) \end{bmatrix}$$

$$\begin{bmatrix} x_{end}(\theta_s, f_s) \\ y_{end}(\theta_s, f_s) \end{bmatrix} = \begin{bmatrix} -SI \cdot \cos(\theta_s) + SD \cdot \cos(\theta_s + f_s) \\ -SI \cdot \sin(\theta_s) + SD \cdot \sin(\theta_s + f_s) \end{bmatrix}$$

This is a minor simplification compared to real third-generation scanners, whose source and detector have a slight offset, but the effect on images will be insignificant.

In the final code, four evenly-spaced rays are simulated for each detector, with their mean value being used in the final sinogram. This is a better representation of the true scenario, in which each detector receives a narrow fan of x-rays, helping to reduce unrealistic aliasing. No oversampling is used in the θ direction as its impact was found to be minor, but real scanners do collect the data for a view over a few milliseconds.

3.3.2.2 Implementation

The projection code written for this simulation is in two parts:

- First, a geometrical raytracing function casts a ray r from a given starting position to a corresponding end position. It outputs the indices I_r and lengths of intersection $d_{r,i}$ for all voxels that the ray crosses, through use of the algorithm detailed in figure 3.6. The implementation used in the majority of this work

traces a single ray through a 2D grid – the function is thus called hundreds of thousands of times for every possible source-detector ray when simulating a single slice. In order to accelerate this crucial step, the MATLAB Compiler was used to produce a significantly-faster MEX binary, which was run on multiple cores using a third-party package [52]. It should be noted that this stage does not use the contents of the pixelmap, only its geometry. This means that the results of this step could be cached to disk for reuse if another pixelmap is defined upon the same grid and the source/detector geometry are unchanged.

- Then, for each ray r,the following sum is calculated over pixel indices i and energy bins j:

$$\zeta_r = -\ln\left[\sum_j P_j \cdot \exp\left(-\sum_{i \in I_r} \mu_{i,j} d_{r,i}\right)\right]$$

where P_j gives the post-filtration energy spectrum of the simulated scanner and sums to unity. This step is accelerated using vectorisation and also compiled to MEX. The calculated value ζ_r is the effective LAC-distance product for ray r in the monoenergetic case.

3.3.2.3 Photon starvation

Photon starvation is taken into account using Poissonian statistics, as described in section 2.6.1.2, replacing the ζ_r values with their randomly-sampled equivalent. Values of ζ_r correspond to the expected logarithmic attenuation at an individual detector element and view, so the number of photons surviving in ray r is sampled from a Poisson distribution with parameter $N_0 \cdot \exp(-\zeta_r)$. N_0 is the initial number of photons incident on an individual detector per view – here it is assumed to be the same for all detectors and views, but may vary in reality due to the effect of bowtie filters and automatic exposure control[3].

3.3.3 Post-processing

The projection data that exists after section 3.3.2.3 is the simulated equivalent of the physical measurements made in a CT scan. Once these have been acquired, a series of post-processing steps can be performed before reconstruction to generate the final

[3]Automatic exposure control modulates the beam current according to the level of attenuation signal received, reducing patient dose whilst ensuring a certain signal-to-noise ratio

Figure 3.6 – The raytracing algorithm used in the simulated CT scanner described in section §3.3. The inset figure shows each of the three classes of ray for a 3×3 pixelmap.

image. Beam hardening correction is described here but real scanners may include additional steps such as noise smoothing or scatter corrections.

3.3.3.1 First-pass (water) beam hardening correction

The issue of beam hardening was discussed in section §2.5. A reconstruction equivalent to the monoenergetic case is of greatest utility in both diagnostic and therapeutic radiology. Correcting the C-type artefact is the first step towards this in the traditional reconstruction paradigm. When only one material is present, the logarithmic attenuation ζ must be a one-to-one monotonically increasing function of radiological depth. When there is only one energy, the relationship is linear, but negative curvature is apparent in the polyenergetic situation as the LAC decreases at depth. Polynomial functions have been shown to be a good fit when mapping polyenergetic projections to monoenergetic ones [53].

It is necessary to consider the question of which single energy should be chosen as the reference value when correcting for beam hardening. There is no standardised recording of this in the DICOM medical imaging standard, and it is not publicly advertised by scanner manufacturers[4]. Herman and Trivedi [45] describe the choice of reference energy for beam hardening in water and bone as "to some extent arbitrary". They go on to determine the energy which minimizes the r.m.s. difference between mono- and polyenergetic projections for a 'typical' cross section before correction, arriving at a value close to two-thirds of the maximum energy in their spectrum, which informs their choice of reference energy. Such a procedure should not be carried out for every cross section because otherwise comparisons of CT numbers between images would be meaningless. $E_{ref} = {}^{2}/_{3}E_{max}$, where E_{max} is determined by the tube voltage in $\mathrm{kV_p}$, was selected as the reference energy in the simulations carried out here[5].

The polynomial water-correction coefficients will vary with input parameters and are derived as follows:

[4]The same can also be said of the x-ray tube energy spectrum, and the composition of filters introduced between the tube and the patient. In fact the only commonly-discussed parameter that is indicative of x-ray energy is tube voltage ($\mathrm{kV_p}$).

[5]Herman and Trivedi's rationale in minimising the correction size is to ensure that the initial reconstruction provides a good template for segmentation into soft and bony tissues, which is an input to second-pass beam hardening correction. Second-pass beam hardening (as described in Sec. 3.3.3.2) is an iterative process, and closer initial conditions lead to quicker convergence. The Tables in appendix B show that the CT numbers of soft tissues vary little over a wide range of energies, and whilst those measured in dense bone are more sensitive to changes in energy, the threshold between the two classes is relatively stable. This choice of E_{ref} remains "to some extent arbitrary", but benchmarking results that appear later suggest it is similar to that in clinical scanners.

- A simulated scan is performed within a water phantom. The water phantom is 30 cm in diameter for simulations representing a body protocol and 15 cm in diameter for those representing a head protocol. This provides the measured polyenergetic projection data ζ_p.
- A modified version of the projection algorithm is run, which returns only the distance of water traversed by each ray. From this the monoenergetic values are calculated by $\zeta_{ref,r} = \mu_w(E = E_{ref}) \cdot d_r$.
- A polynomial fit is performed to find the best-fit coefficients a_i in equation (3.2) using MATLAB's `polyfit` function. The fit is weighted such that the whole range from 0 to the maximum value of ζ_p is treated equally.

$$f_{\mathrm{wc}}(\zeta_p) = \sum_{i=0}^{4} a_i \zeta_p^i \tag{3.2}$$

The polynomial coefficients are cached such that they need not be recalculated in scans utilising the same geometry and source but a different object. Once the coefficients are known, the correction simply requires the calculation of equation (3.2) for each projection in the sinogram. This step is fast and computationally simple.

A fourth order fit was chosen based on the data displayed in figure 3.7. A quartic equation was the smallest order polynomial that had an r.m.s. error below 0.1% for $\zeta_P > 0.1$, below which small absolute deviations can lead to very large relative errors that are nevertheless inconsequential in most imaging scenarios. Correction functions for the energy spectra generated to represent a clinical scanner in this work (described in more detail in section §3.4) are shown in figure 3.8.

3.3.3.2 Second-pass (bone) beam hardening correction

Second-pass beam hardening correction aims to minimise the S-type artefacts that result from materials with different MACs being present within the scanned object. With only one set of projection data taken at one $\mathrm{kV_p}$ setting, little can be inferred regarding the object's chemical makeup. Second-pass beam hardening correction methods therefore need access to further information, or to make some assumptions regarding composition. Techniques in the first category include dual energy CT, which acquires projections with two different energy spectra simultaneously through use of an additional tube, enabling locations to be classified according to the energy-dependence

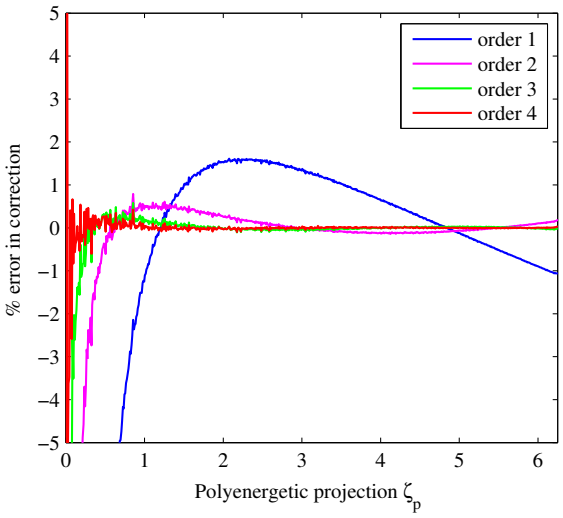


Figure 3.7 – Error in the first-pass beam hardening correction for polynomial fits of order 1 to 4, using a 30 cm water phantom, 120 kV_p energy spectrum and parameters in table 3.2.

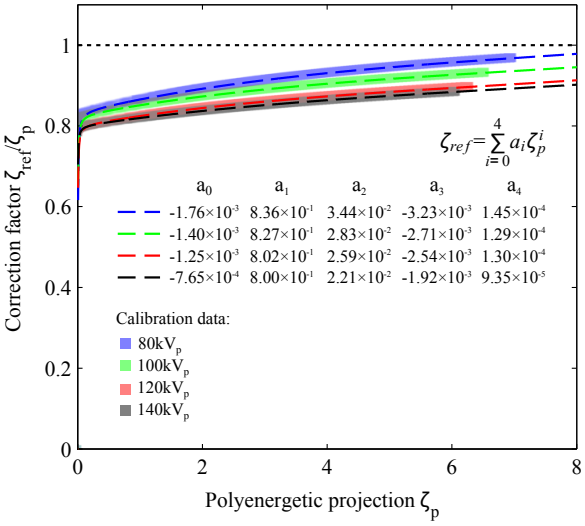


Figure 3.8 – First-pass beam hardening correction functions derived by simulating a 30 cm water phantom at four different energy settings, overlaid on the data used to obtain fits. The dotted-line shows the case when no correction is necessary (for monoenergetic $E = E_{ref}$).

of their LAC[6]. Techniques in the second category require reasonably detailed knowledge of the x-ray tube's energy spectrum but no additional hardware.

The assumption made in most software-based clinical beam hardening correction is that the patient contains n different materials whose LAC vs. energy dependence is known, each of which occupies different non-overlapping regions of the Hounsfield scale. In general-purpose cases, n is normally two[7] and the classes correspond to regions containing soft tissue and bone[8].

Two common methods for correcting the sinogram for beam hardening in bone are compared in a paper by Herman and Trivedi [45] – the first considers materials in the bone class to be a mixture of water and bone (labelled B_I), whilst the second considers those materials to be a mixture of air and bone (labelled B_{II}). Method B_I has been adopted in this CT simulation software. The intended result is a map that accurately

[6]Beam hardening correction is not a very common use of dual energy, however; it is more regularly used to identify between regions of similar attenuation but different composition (e.g. plaques, contrast media) in diagnostic situations.

[7]It may be larger when a known contrast medium or metal implant is present.

[8]Some historic methods assume the classes to be water at varying density and bone at varying density - whilst not completely unphysical, these classes do introduce an undesirable discontinuity in the mass of a material with HU.

reflects the linear attenuation coefficient within the scanned object at $E = E_{ref}$. The procedure is as follows:

1. Perform a preliminary reconstruction of the water-corrected sinogram, $p = f_{\text{wc}}(\zeta_p)$ (from equation (3.2)). In the ideal water-only situation, this would result in a map of μ at $E = E_{ref}$.

2. Forward-project the reconstruction to produce a degraded version of the original sinogram, m' – this will be used to derive an approximate correction for any information loss that occurs due to back-projection followed by forward-projection.

3. Identify the material class in each pixel by comparing its CT number with their HU thresholds. Given this material map (and assuming the pixel values are all valid at $E = E_{ref}$), estimate the density/mixture fractions or any other parameters necessary to calculate the LAC within a pixel for any energy.

4. Calculate and forward-project this new LAC map for a number of energy bins and combine according to the x-ray tube spectrum, producing the sinogram which would be expected if the map were correct, p'. This step is equivalent to the process described in Section 3.3.2, except with the phantom defined by the map in step 3. Perform water-based correction on the expected sinogram, giving $f_{\text{wc}}(p')$.

5. Sum the original sinogram and the difference between the degraded sinogram and the expected sinogram, giving $p - (f_{\text{wc}}(p') - m')$. Perform a reconstruction. If desired, the new image can be used in step 1 (retaining the original sinogram for use in this final step) and the process repeated for any number of iterations until the image is considered to have converged.

3.3.4 Reconstruction

The filtered back-projection approach to reconstruction was adopted, as introduced in Section 2.3. This decision was made on the basis that it is the technique most often used in clinical scans, but it should be noted that a series of promising maximum-likelihood methods exist and these have the potential to produce higher quality results from noisy or less well-sampled data. Some of these methods (also known as iterative reconstruction) have been integrated into commercial products, but processing time currently limits their routine use.

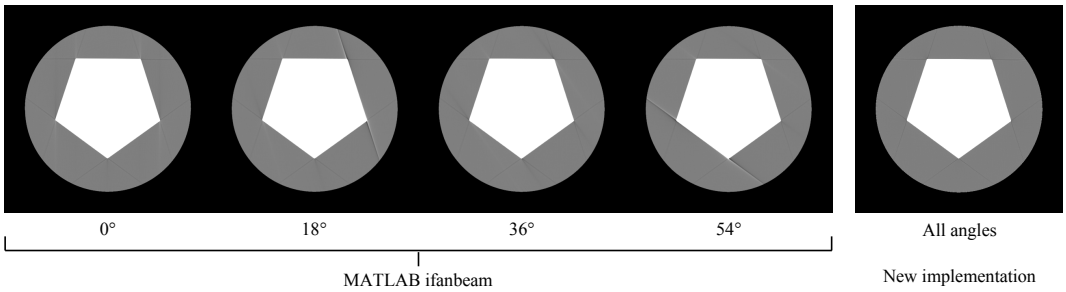


Figure 3.9 – Demonstration of variable streaking artefacts in images reconstructed from cyclic shifts of the same sinogram using MATLAB's `ifanbeam` function. Sinograms originated as monoenergetic projections through a pentagon of water at density $2\,\mathrm{g/cm^3}$ and maximum dimension 25 cm within a 40 cm cylindrical water phantom. Also shown is an example of the output of the reconstruction algorithm written for this project, which shows no significant changes as the sinogram is shifted.

As with forward-projection, the MATLAB Image Processing Toolbox provides functions to perform back-projection in parallel and fanbeam geometries. These inverse functions have significantly fewer limitations compared to their forward counterparts, although the parallel geometry is again the primary implementation with fanbeam related by linear interpolation. Nevertheless, directional streaking artefacts are evident from the output of `ifanbeam` and figure 3.9 demonstrates that this must be a result of its workings rather than deficiencies in the original projections. The explanation likely lies in the interpolation between fanbeam and parallel geometries; artefacts could, for example, arise if the procedure neglects the periodic nature of the sinogram. In order to avoid confusion between these reconstruction defects and realistic artefacts, it was necessary to use an alternative reconstruction function. A function was written specifically for the purposes of this simulation and is described below.

3.3.4.1 Algorithm

The algorithm adopted is essentially that described in Section 3.4.1 of Kak & Slaney [28]. For each slice, the following procedure is performed:

- Initialize a matrix R of the same size as the desired image and fill it with zeroes. Assign physical x and y positions (in cm) to the midpoints of each pixel.
- For each view in the sinogram (with rotation angle θ):
  - Filter the projections in the corresponding line of the sinogram (see Section 3.3.4.2) to produce $\zeta_{F,\theta}(f_s)$.

– Smooth high frequency data in the filtered sinogram to produce $\zeta_{G,\theta}(f_s)$ (see Section 3.3.4.3)

– Find the $f_{i,j}$ (fan angle) and $L_{i,j}$ (distance from source) coordinates for each pixel[9].

– Add the back-projection of this view to R by evaluating equation (3.3) for each pixel i, j, where $S_{F,\theta}(f)$ is a function that uses linear interpolation to estimate the projection value at any f from the discrete filtered sinogram $\zeta_{G,\theta}(f_s)$.

$$R_{i,j} = R_{i,j} + \Delta\theta \cdot \Delta f \cdot \frac{S_{F,\theta}(f_{i,j}) \cdot SI \cdot \cos(f_{i,j})}{L_{i,j}^2} \tag{3.3}$$

3.3.4.2 Filter

It is possible to avoid significant amounts of blurring in reconstructions of projections with an equiangular fanbeam geometry[10] using a convolution prior to back-projection. The implementation here used a direct discrete convolution, as given in equation (3.4). The filter function F used in the convolution takes the form of a linear ramp in angular spatial frequency space with a sinc^2 $\left(\equiv \frac{\sin^2 x}{x^2}\right)$ correction for finite sampling [55]. Equation 3.5 specifies the F used: it should be noted that the convolution window is twice the length of $\zeta(f_s,\theta)$ such that each element of ζ can affect all elements of ζ_F.

$$\begin{aligned} \zeta_{F,\theta}(f_s) &= \sum_{\frac{-f_0}{2} \leq (f_s - f'_s) \leq \frac{f_0}{2}} \zeta(f_s,\theta) \cdot F(f_s - f'_s), \\ \text{where } f'_s &= \{-N_{rays} \cdot \Delta f, ..., 0, \Delta f, ..., N_{rays} \cdot \Delta f\} \end{aligned} \tag{3.4}$$

$$F(f'_s) = \begin{cases} \frac{1}{8(\Delta f)^2} & n = 0 \\ \frac{-1}{2(\pi \sin(n \cdot \Delta f))^2} & \text{odd } n \\ 0 & \text{even } n \end{cases}, \text{ where } n = \frac{f'_s}{\Delta f} \tag{3.5}$$

3.3.4.3 Apodisation

Some aliasing artefacts can be reduced through the application of an apodisation filter, which improves the visual quality of reconstructed images. Such a process may also attenuate high-frequency stochastic noise, although it is at the expense of spatial

[9] f and L are orthogonal, so can uniquely specify any point in the image

[10] and exactly three other geometries [54]

resolution. In the CT reconstruction function written for this work, apodisation took the form of multiplication with a Hann window of form equation (3.6) in the angular spatial frequency domain. The Hann window is a well-known function in digital signal processing and was chosen as it has a smooth point spread function with rapid falloff and minimal sidelobes [56, p124]. The final smoothed and filtered projections, $\zeta_{G,\theta}(f_s)$, are found using the discrete fast Fourier transform (FFT) and its inverse, as in equation (3.6).

$$W(\tilde{f}) = \begin{cases} \frac{1}{2}(1+\cos(\frac{\pi\tilde{f}}{\tilde{f}_c})) & |\tilde{f}| \leq \tilde{f}_c \\ 0 & \text{otherwise} \end{cases} \quad \begin{array}{l} \text{where } \tilde{f} \text{ is the Fourier conjugate of } f \\ \text{and } \tilde{f}_c \text{ is a cutoff frequency} \end{array} \tag{3.6}$$

$$\zeta_{G,\theta}(f_s) = \mathrm{FFT}^{-1}\left\{ \mathrm{FFT}[\zeta_{F,\theta}(f_s)] \cdot W(\tilde{f}_s) \right\} \tag{3.7}$$

The apodisation equations require a cutoff frequency, $\tilde{f}_c$. Nyquist's theorem implies that the maximum frequency which may be represented in a regularly-sampled dataset is equal to half the number of samples (with units of reciprocal sample spacing). The Nyquist frequency $\tilde{f_N}$ might therefore be considered a logical choice for the cutoff frequency, but matters are complicated by the fact that the final reconstructed image is actually constructed in the x, y coordinate system, in which the linear spatial frequency represented by the spacing between projections changes with distance from the source. The best value of $\tilde{f}_c$ that balances aliasing and spatial resolution for a particular purpose should thus be found empirically.

Aliasing artefacts are most apparent towards the edge of images acquired using a fanbeam geometry, where the sampling density is lower and the effects of beam divergence are greatest. figure 3.10 shows reconstructions of an arm at the edge of the field-of-view when CT is simulated in an abdominal slice of the ICRP AM voxellised phantom (figure 3.2). A range of cutoff frequencies at intervals were examined and $0.5\tilde{f_N}$ was determined as the maximum frequency at which no stripes or ringing[11] (however minimal) were present. It should be noted that the 'texture' of real CT images bears more resemblance to that at higher cutoff frequencies. Some of this will result from natural density variations and stochastic noise, as well as aliasing, but it is likely scanner manufacturers consider radiologists to prefer sharper images and able to distinguish between the variety of causes.

As spatial variations in CT number are of great importance to radiotherapy treatment planning, it was also necessary to conduct an assessment of the limit that

[11] oscillatory behaviour in pixel values near edges, mostly evident as 'overshoots'

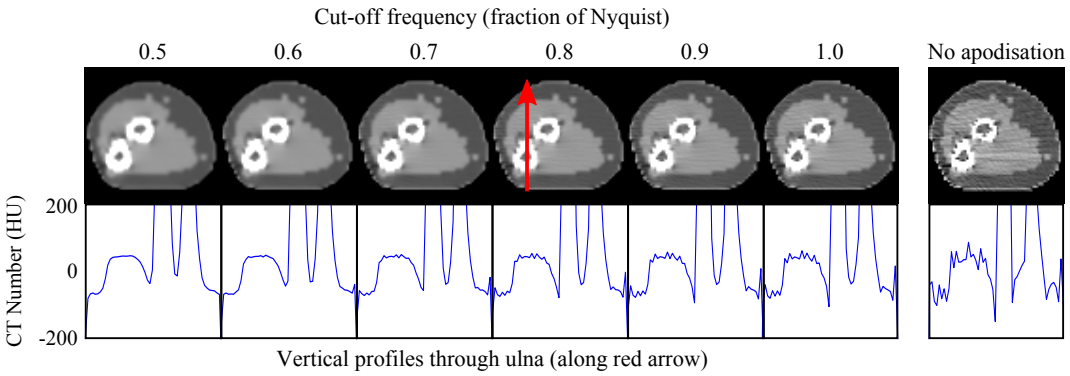
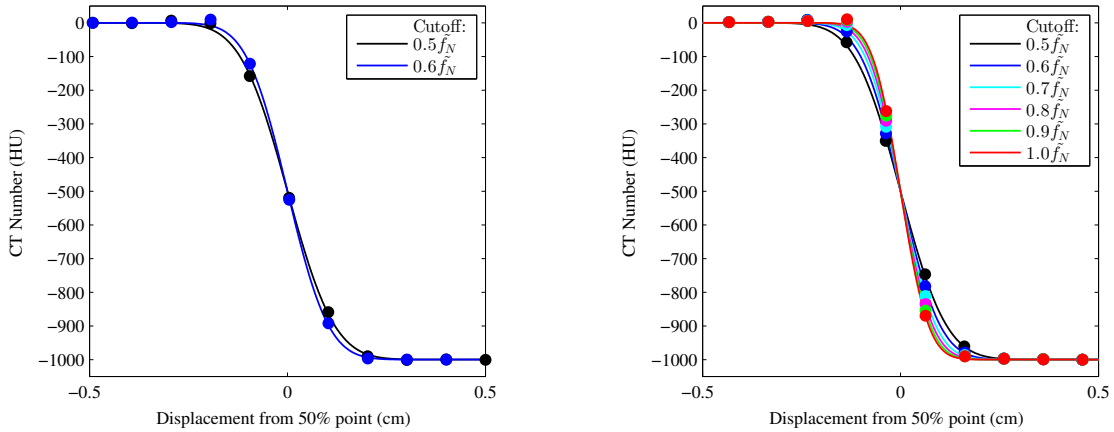


Figure 3.10 – Sections from simulated monoenergetic CT through the arm of the ICRP AM phantom when reconstructed with the new implementation of fanbeam filtered back-projection. It can be seen that apodisation using equation (3.7) causes a large reduction in aliasing, and that this is further enhanced by the use of smaller angular spatial frequency cutoffs, at the expense of spatial resolution.

the choice of cutoff frequency places on spatial resolution of the resulting image, providing a benchmark for the scale at which comparisons between images are useful. The point spread function of an imaging system is often used when considering resolution, but the discrete nature of the method used in these simulations does not allow for the required sinogram (which involves an infinitesimally sharp line). Instead, the edge spread function was examined, with the falloff-distance between a sharp water-air interface being considered a reasonable measure of image resolution.

Figure 3.11 and the data in table 3.1 characterise the profile of edges in a series of simulations containing water blocks of varying sizes. Reconstructions at a limited range of cutoff frequencies may be obtained using MATLAB's `ifanbeam` and, notwithstanding the significant shortcomings previously shown in figure 3.9, the image resolution is comparable to the inverse fanbeam procedure that was adopted in the remainder of the work. The data for the new implementation reveals a monotonically decreasing relationship between edge sharpness and cutoff frequency; the direction of this trend is intuitive, and its non-linearity may arise as a result of the diverging projection paths. Also notable is an increase in edge sharpness with distance from the image centre, occurring to a greater extent at large cutoff frequencies, which arises due to the reduction in sampling density towards the edge of the image. This apparent sharpening is more indicative of ringing than a real increase in image resolution. It can be seen that it takes approximately 5 mm for the CT number to fully drop from 0 to -1000 HU at the cutoff frequency of $0.5\tilde{f}_N$, compared to 3 mm at $1.0\tilde{f}_N$. These values are only true for this simulator, however, and should not be taken to be representative of all clinical scanners with parameters comparable to table 3.2, which have a variety



(a) MATLAB `ifanbeam` **(b)** New implementation

Figure 3.11 – Lateral profiles through simulated scans of a water-air interface located at 15 cm from the isocenter of the scanner, reconstructed with varying angular spatial frequency cutoffs, given as a fraction of the Nyquist frequency, $\tilde{f}_N$. Points show the pixel values in simulated images, averaged over the direction parallel to the edge, whilst lines show fits to the Gauss error function. Such fits do not show ideal agreement – whilst this may be due to some oscillatory behaviour arising from applying a cutoff in the frequency domain, it could also be a consequence of the small amount of data used in the fit; linear interpolation between points was thus used in the production of table 3.1.

of more advanced post-processing algorithms available to them.

3.3.5 Output

In order to create images comparable to those produced by clinical scanners, CT numbers were rounded to the nearest integer value with a low threshold of -1000 and stored as 16-bit integers in the DICOM medical imaging format. A mask with value -1000 HU was applied outside circle A in figure 2.1, removing a low intensity 'fog' that the reconstruction algorithm causes outside the sampling region.

3.4 Benchmarking

The simulator was benchmarked using scans acquired with a GE LightSpeed Plus scanner. Simulation parameters were selected to achieve results comparable to the output of this machine. The precise specifications of certain elements of the scanner in question were unclear as conflicting information exists, probably because machines manufactured later incorporate some revisions. Nevertheless, referencing multiple sources and the DICOM tags of the output images provided the complete parameter

	MATLAB `ifanbeam`*	
	$0.5\tilde{x_N}$	$0.6\tilde{x_N}$
r = 5 cm	0.142	0.129
r = 10 cm	0.139	0.124
r = 15 cm	0.141	0.128
r = 20 cm	0.138	0.123

	New implementation					
	$0.5\tilde{f_N}$	$0.6\tilde{f_N}$	$0.7\tilde{f_N}$	$0.8\tilde{f_N}$	$0.9\tilde{f_N}$	$1.0\tilde{f_N}$
r = 5 cm	0.140	0.125	0.115	0.108	0.103	0.099
r = 10 cm	0.135	0.119	0.107	0.098	0.090	0.084
r = 15 cm	0.135	0.118	0.106	0.098	0.090	0.084
r = 20 cm	0.133	0.115	0.103	0.093	0.085	0.081

Table 3.1 – Distance in cm between the 75% and 25% points of sharp edges in simulated scans reconstructed with varying angular spatial frequency cutoffs, given as a fraction of the Nyquist frequency, f_N. Data is derived from linear interpolation of lateral profiles through rectangular water blocks of size $2r$ in air and should be viewed in the context of a 0.099 cm pixel size. Monoenergetic projections without stochastic variation were used to avoid confounding the spatial sampling effects.

* `ifanbeam` calls `iradon` in order to reconstruct the image from parallel-beam projections, which it estimates from the fanbeam sinogram using linear interpolation. The transformation happens before apodisation, but it fixes the size of the intermediate laminogram (parallel-beam equivalent of a sinogram) to one projection per reconstructed image pixel, reducing the range of spatial frequencies represented. The cutoff frequencies provided to `ifanbeam` were set such that the linear spatial frequency cutoff at the centre of the image is equivalent to that in the new true fanbeam implementation, $\tilde{x_N}$ being the linear spatial frequency at the centre of the image when $\tilde{f_c} = \tilde{f_N}$. Cutoffs greater than $0.6\tilde{x_N}$ are above the maximum frequency that the laminogram for a 512×512 end-result may represent and are not accepted by `ifanbeam`.

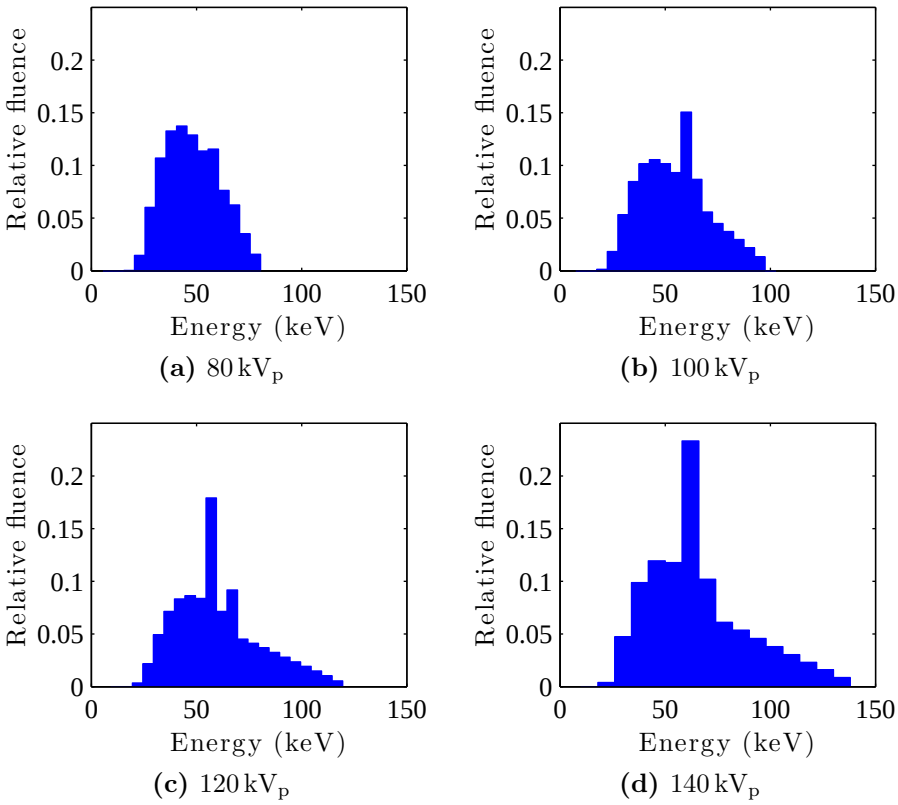


Figure 3.12 – X-ray energy spectra generated using SpekCalc for Section 3.4.

set in table 3.2. Representative energy spectra were generated using release 1.1 of the SpekCalc software [31], under the assumption that the x-ray beam was solely filtered by aluminium of thickness given in row 12 of table 3.2. The filter may contain other materials, but no information was available to confirm this[12]. Energy bins of 5 keV were used[13] and the two physics model parameters (as described in [58]) were kept at their default values. The spectra are shown in figure 3.12.

3.4.1 Reference images

The reference images used were kindly provided by Andrew Reilly, Helen Winter and Mark Hill. A series of blocks, composed of several different plastics, were stacked in two different arrangements and scanned at four kV_p levels on a GE LightSpeed

[12]The Aluminium equivalent thickness given in the table is a 'quality equivalent filtration' measured at 70 kV for radiation protection purposes. 'Quality equivalence' is often practically measured using two attenuation measurements [57] (in the form of half value layers) – it is quite possible for measurements from different thicknesses of different materials to agree well at this resolution.

[13]except for at 140 kV_p, where memory limitations in simulation necessitated the use of coarser 8 keV bins

#	Parameter	Value	Source
Geometric parameters			
1	Source → isocenter distance	541 mm	DICOM tag (0018,1111)
2	Source → detector distance	949 mm	DICOM tag (0018,1110)
3	Fan angle	56.4°	$(360°/2\pi) \times (\mathbf{4} \times \mathbf{5})/\mathbf{2}$
Detector parameters			
4	Number of cells in detector	912	DICOM tag (0019,1002)
5	Detector cell size	1.024mm	DICOM tag (0019,1004)
6	Geometric efficiency	0.7	[59]
7	Absorption efficiency	0.99	[59]
Rotation parameters			
8	Views per second	984 *	DICOM tag (0019,102F)
9	Gantry rotation period	1.0 s *	DICOM tag (0019,1027)
Source parameters			
10	Maximum energy (kV_p)	80, 100, 120 and 140	[59]
11	Target angle	7°	[59]
12	Filtration	Head: 4.75 mm Body: 5.65 mm (Al equivalent)	[60]
13	Exposure per slice	150 mAs *	DICOM tag (0018,1152)
Reconstruction parameters			
15	Image size	512×512 pixels	DICOM tags (0028,0010-11)
16	Cutoff frequency	0.7×Nyquist frequency	Determined in Section 3.4.2

Table 3.2 – Parameters used in the CT simulation software. Images from a GE LightSpeed Plus were used as a benchmark in order to ensure results were representative of a clinical CT scanner. Items marked with a * may take other values depending on scan protocol, and so are partially at the control of the user.

Material	Equivalent to	Density ($g \cdot cm^{-3}$)	Elemental composition (% by mass)*							Source
			H	C	N	O	Cl	Ca	Others	
LN10	Lung	0.27	8.4	68.0	2.3	18.9	0.1	2.4		[62]
RB2	Rib bone	1.29	5.7	50.6	1.7	28.2	0.1	13.17		[62]
SB5	Cortical bone	1.84	2.6	30.6	1.0	38.9	0.1	26.8		[62]
IB7	Inner bone	1.10	7.1	59.6	2.0	23.4	0.1	7.8		[62]
LV/L1	Liver	1.06	10.18	14.40	2.83	71.80	0.18	0.00	Na: 0.11, S: 0.24, K:0.29	[63]
BR12	Breast tissue	0.97	8.7	69.9	2.4	17.9	0.1	1.0		[64]
KD/L1	Kidney	1.05	10.40	11.35	2.74	74.50	0.28	0.00	Na: 0.18, P:0.19, K:0.25	[63]
WT1	Water	1.00	8.4	68.0	2.3	18.9	0.1	2.4		[62]
WTe	Water (for e^-)	1.03	7.39	61.99	2.01	22.18	0.11	6.32		[65]
AP7	Adipose tissue	0.95	8.4	69.1	2.4	16.9	0.1	0.0		[62]
Air	Air	0.001205	0.0	0.0	75.5	23.2	0.0	0.0	Ar: 1.3	[64]

Table 3.3 – Elemental compositions of the materials present in the reference images.

* Note that due to rounding errors, some of these compositions do not add up to 100% - in this situation, the values are rescaled such that their sum is 100%.

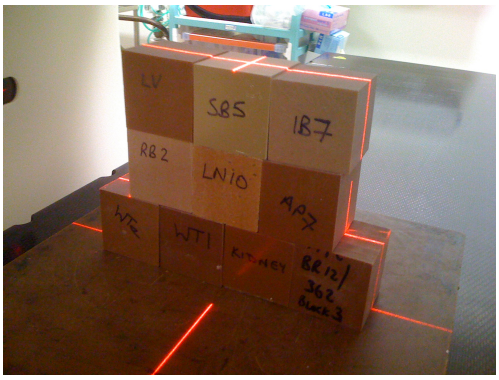
Plus scanner located at the Churchill Hospital, Oxford, using the protocol defined for use in the prostate region. The plastics in question are tissue and water equivalent[14] substances manufactured by the Tissue Substitute Section at Barts and The London NHS Trust. Photographs of the block stacks and example reference images are shown in figure 3.13.

The regular and rectangular nature of the block arrangements aided their translation into a voxellised phantom for use in the simulation software. As the geometry consists of cubes with side length 5 cm stacked on a 30 cm×5 cm slab, it could be adequately represented by a 14×6 array of pixels with dimension 2.5 cm×5 cm. Composition information for the tissue equivalent plastics was sourced from the literature and is tabulated in table 3.3. It should be noted that actual composition will differ somewhat between manufacturing batches – up to half a percent variation per constituent has been recorded for the A-150 tissue equivalent plastic [61].

3.4.2 Stochastic noise

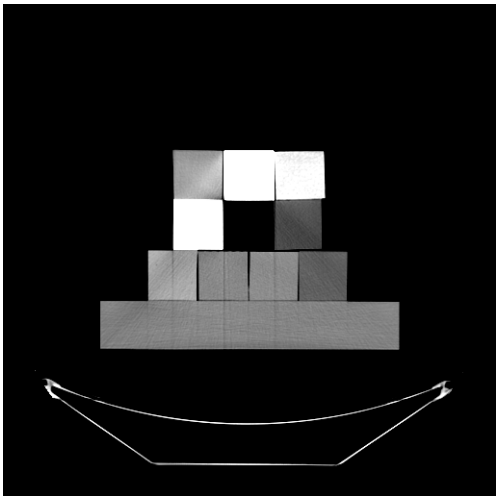
The sources used to obtain the parameter set in table 3.2 did not provide any information on the cutoff frequency, $\tilde{f}_c$ used in reconstruction, which very much affects the spatial characteristics of noise in the image. In order to compare CT number distributions (as opposed to average values) within a particular block, it was necessary to determine the value of the cutoff frequency for use in the simulator's reconstruction algorithm. Simulations were performed at intervals of $0.1\tilde{f}_N$ and the high spatial frequency content of a region of homogenous composition (inside the slab of WT1)

[14] in terms of x-ray attenuation for a specified range of energies

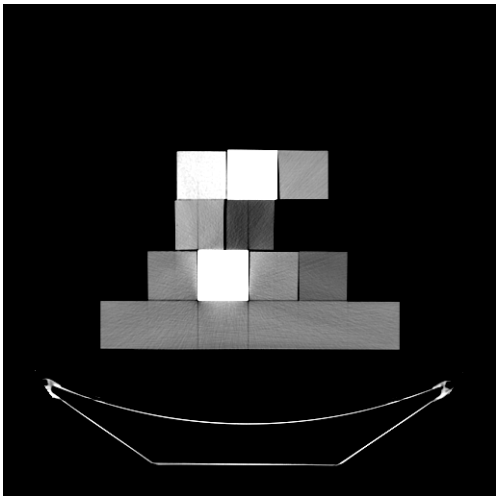


(a) Photograph of the 'Stack A' arrangement

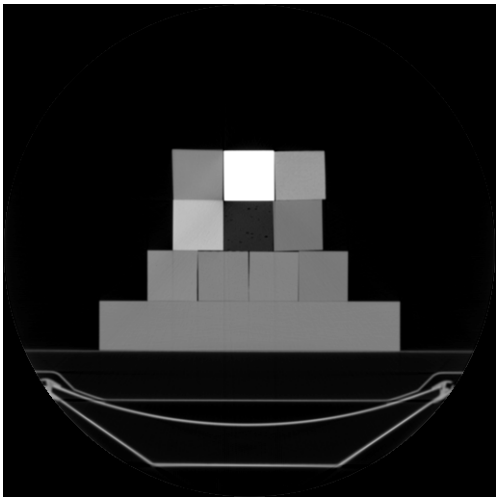
(b) Photograph of the 'Stack B' arrangement



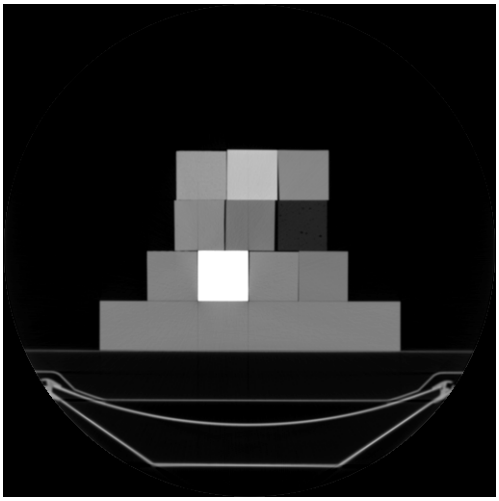
(c) Scan of Stack A at 120 kV_p. Level: 0 HU Window: 400 HU



(d) Scan of Stack B at 120 kV_p. Level: 0 HU Window: 400 HU



(e) Scan of Stack A at 120 kV_p. Level: 0 HU Window: 2000 HU



(f) Scan of Stack B at 120 kV_p. Level: 0 HU Window: 2000 HU

Figure 3.13 – The arrangements used to benchmark the simulation software.

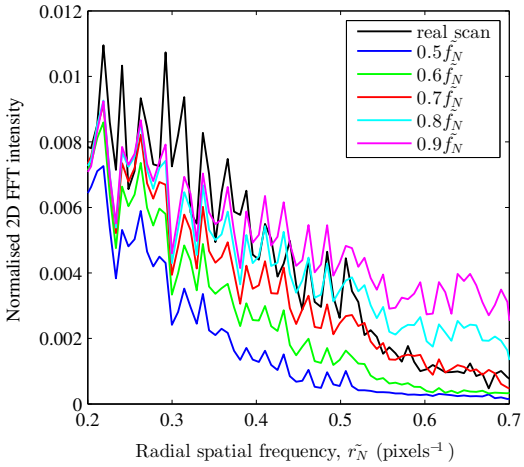


Figure 3.14 – Radial FFT results for a sample of WT1 in Stack A in simulated images (with a range of cutoff frequencies) and an actual scan. The maximum spatial frequency represented in the actual scan is most similar to that in the simulation when the cutoff is $0.7\tilde{f}_N$.

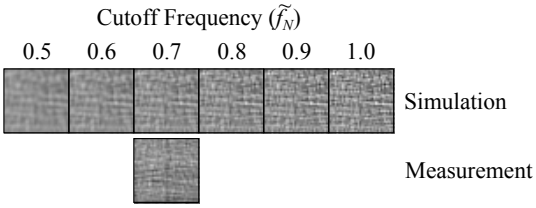


Figure 3.15 – Samples of WT1 pixel values from simulated images (at a number of cutoff frequencies) and an actual scan of Stack A. It can be seen that noise becomes smoother as the cutoff frequency decreases, and that noise in the simulation at $0.7\tilde{f}_N$ shows a similar granularity to the measured result. Level: 0 HU Window: 100 HU

in each reconstructed image was compared with the same area in the measured images. Radial spatial frequency spectra, produced by re-binning the two-dimensional fast Fourier transform (2D-FFT) of the images, are shown in figure 3.14. It can be seen that simulating with a cutoff frequency of $0.7\tilde{f}_N$ best reproduces the high frequency characteristics of the actual scan, as can be verified from a visual inspection of figure 3.15.

The intensity of stochastic noise is quite distinct from its resolution and, whilst figure 3.15 shows great similarity in the amount of CT number variation in the WT1 block, it was still necessary to verify this numerically in a wider range of materials and at a wider range of energies. Disagreement between the degree of spatial noise in simulated and measured images would be indicative of an incorrect value for the number of incident photons in the simulation. The initial number of photons per detector per view was calculated from the spectral fluence, given by SpekCalc in terms of $\phi_E = \frac{\text{number of photons}}{\text{keV}\cdot\text{cm}^2\cdot\text{mAs}}$ at a distance of 1 metre, by use of equation (3.8), in which bold numbers refer to the rows of table 3.2.

$$N_0 = \phi_E \cdot n_{E\,bins} \cdot \frac{\Delta z \times \mathbf{5} \times \mathbf{13}}{\mathbf{8} \times \mathbf{9}} \cdot \frac{2\pi \times \mathbf{3}}{360^\circ} \cdot \left(\frac{\mathbf{2}}{100\,\text{cm}}\right)^2 \tag{3.8}$$

As the relationship between tube current and rate of photon production used by SpekCalc is a function of the target composition, which is not quantitatively known here, it is quite feasible for there to be an additional multiplicative term η in equation (3.8), which could vary with the energy spectrum. This possibility was investigated by finding the standard deviation of CT numbers in every block, at each kV_p value and for both stacks, in simulations with a range of η values from 0.1 to 3.5. The results were fitted to equation (3.9), on the basis that those factors which are independent of photon statistics but produce pixel value variations (such as material inhomogeneity, aliasing and beam hardening) can be encompassed in σ_{other}, whilst stochastic noise can be characterised by an independent term with a variance proportional to the inverse of the number of incident photons. Clearly the parameters σ_{other} and a vary between stack, block and energy, but each can nevertheless be used to predict a value of η for the corresponding stack, block and energy in the measured scans.

$$\sigma_{\text{overall}} = \sqrt{\sigma_{\text{other}}^2 + \frac{a}{\eta}} \tag{3.9}$$

table 3.4 and figure 3.16 show the results of this investigation. It can be seen that there is little in the dataset to suggest that η deviates from unity. Whilst a number of fits failed and there is considerable variance in the final histogram, it should be noted that the procedure is very sensitive to small changes in HU values – a mere 1 HU difference in standard deviation (0.001% of the water LAC) can shift the best-fit η from 1 to 3, and a difference of 5 HU or less in standard deviation will cause the fit to fail completely. Such differences could easily be caused by inhomogeneity in the scanned block or any number of non-stochastic artefacts.

3.4.3 Overall assessment

The methods and parameters described in the preceding sections were brought together to produce a final set of simulated images, which could be compared with the reference images. The aim of this software was not to emulate the output of the scanner in question with high precision – a pursuit which would require a significantly larger amount of time and technical information – but to generate images which appear qualitatively similar, with artefacts (in particular those arising from beam hardening) whose magnitudes are quantitatively comparable. Numerical results used to assess the degree to which this objective has been achieved are given in table 3.5 and table 3.6, whilst visual comparisons for two of the reference situations may be seen in figure 3.17.

Stack A

	LV	SB5	IB7	RB2	LN10	AP7	WTe	WT1 (block)	KID	BR12	WT1 (slab)
80 kV_p	-	1.13	0.67	1.11	0.53	0.79	1.48	0.89	0.90	0.87	1.07
100 kV_p	-	-	0.59	2.23	0.27	0.85	0.76	0.78	0.87	1.06	1.08
120 kV_p	-	-	0.46	-	0.18	2.68	1.06	0.78	0.86	1.19	1.60
140 kV_p	-	-	0.28	-	0.11	2.54	0.87	0.77	0.67	1.06	1.40

Stack B

	LV	SB5	IB7	RB2	LN10	AP7	WTe	WT1 (block)	KID	BR12	WT1 (slab)
80 kV_p	0.67	1.42	1.28	1.18	0.97	0.53	1.92	1.00	0.73	1.27	0.85
100 kV_p	0.69	-	1.51	-	2.05	0.27	-	1.64	0.67	1.05	0.91
120 kV_p	0.70	-	1.73	-	-	0.15	-	-	0.94	1.15	1.00
140 kV_p	0.51	-	1.20	-	-	0.11	-	-	0.99	0.97	0.89

Table 3.4 – Best-fit values for η estimated using information from the reference scans and a series of simulations for each block, energy and arrangement. Cells containing no number correspond to situations in which no sane value of η was found, usually because the measured σ was less than that seen in the simulation with stochastic effects turned off. This is likely indicative of imperfections in the simulation's energy spectrum and block compositions. The data from this table is presented graphically in figure 3.16 and can be considered consistent with $\eta = 1$.

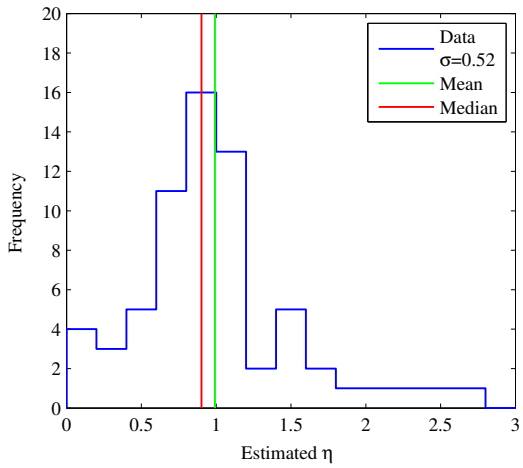


Figure 3.16 – Histogram of η values shown in table 3.4. Given the standard deviation and average values, there is no reason to reject the hypothesis that $\eta = 1$.

The insets of figure 3.17 show the simulator with no bone-based beam hardening correction to produce realistic-looking streaks in the region of dense blocks with high calcium content. This is particularly apparent in the top-left block of Stack A and the central block of stack B. Quantitative data supports the hypothesis that only beam hardening in water is corrected for: in all-but-one case[15], the no-BBHC simulation offers the best prediction of median and IQR (inter-quartile range) of the pixel values in the blocks, as assessed by the root-mean-square deviation from the measured values. Bone-based beam hardening correction is not mentioned in the technical reference manual for the GE LightSpeed Plus, so it is likely that it is indeed not enabled by default[16].

The simulator used here has some shortcomings. In particular, it makes no attempt to model scattering (which would be best done using computationally-intensive Monte Carlo particle tracking) and, being purely two-dimensional, cannot account for partial volume effects. Nevertheless, considering the complete set of reference data, the root-mean-square relative[17] error in CT number between the no-BBHC simulations and the measurements is 2.1%. The corresponding root-mean-square relative error in interquartile range of a block is 0.5% of its median. Especially given the small amount of information available regarding the x-ray tube energy spectrum, and the typical uncertainties in tissue equivalent material composition, this level of agreement is sufficient for the investigations of relative CT number variation in the following Chapters.

[15] the IQR of Stack B at $120\,\mathrm{kV_p}$

[16] Whilst the major vendors of CT scanners have implementations of BBHC, it seems that it is typically only applied in scans when a protocol designed for cranial scans is selected. (S-type artefacts are often particularly apparent in the region of the posterior fossa.)

[17] taking zero attenuation to be -1000 on the Hounsfield scale

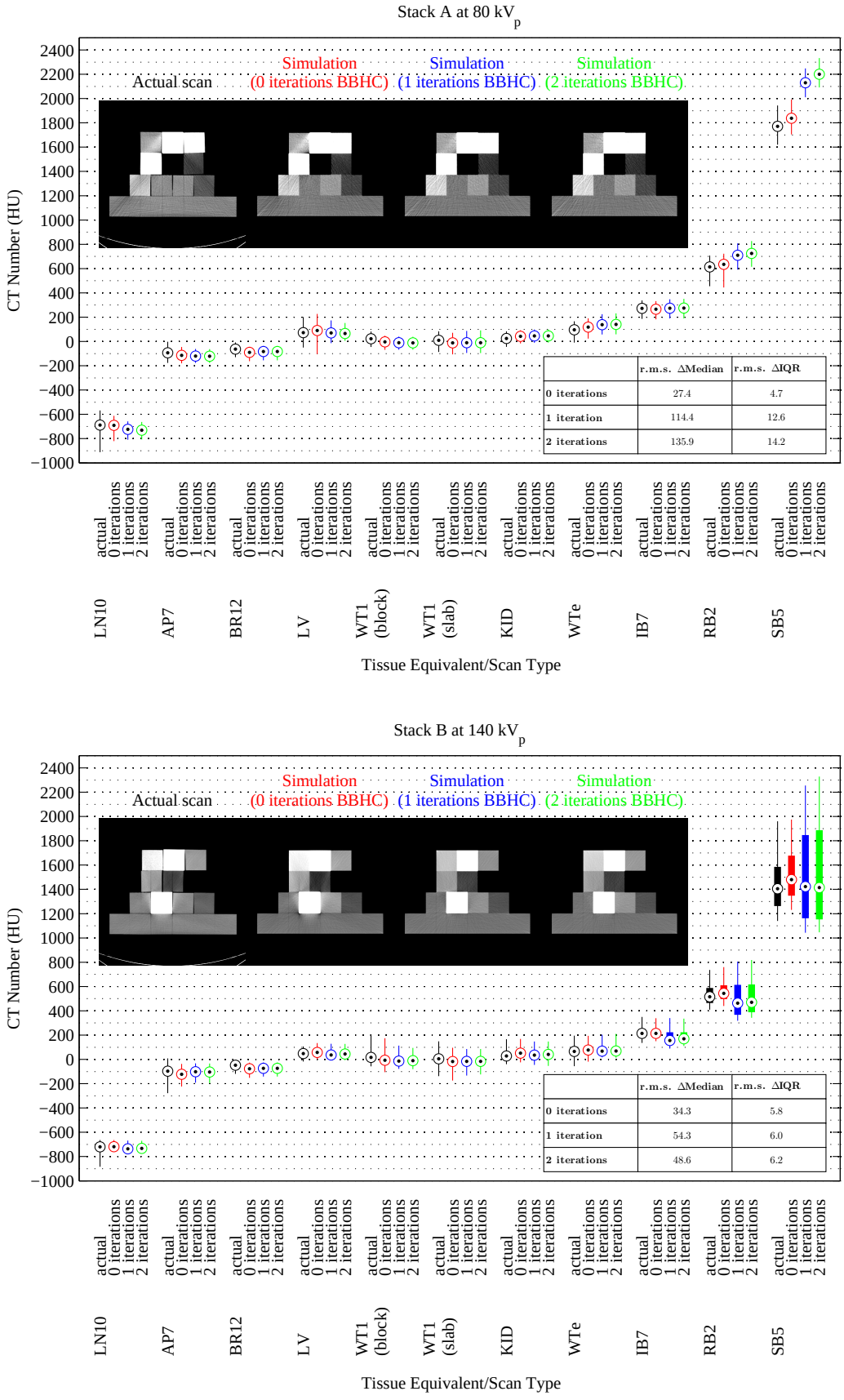


Figure 3.17 – Box plots of CT number for each material in two of the eight measured reference images, and in the corresponding images simulated with 0 to 2 iterations of bone-based beam hardening correction (BBHC). Points mark the median values, with broad lines representing the interquartile range (IQR) and narrow lines spanning the full range. Inset are the relevant images, displayed with a level of 0 HU and a window of 400 HU.

	Material	Measured		Sim, 0 iterations		Sim, 1 iteration		Sim, 2 iterations	
		Med	IQR	ΔMed	ΔIQR	ΔMed	ΔIQR	ΔMed	ΔIQR
Stack A, 80 kVp	AP7	-94	29	-20	-1	-26	-7	-28	-7
	BR12	-63	23	-27	-1	-20	-1	-20	-1
	IB7	272	30	-7	-1	2	-7	3	-7
	KID	25	20	17	-1	21	0	21	1
	LN10	-689	31	-2	0	-35	-6	-41	-8
	LV	73	56	17	-1	-3	-24	-8	-28
	RB2	614	61	21	-13	96	-23	111	-24.25
	SB5	1771	69	67	-6	359	-22	429	-26
	WT1 (block)	22	22	-25	1	-32	1	-34	1
	WT1 (slab)	9	28	-21	-4	-19	-5	-19	-4
	WTe	96	40.75	22	-3.75	42	-3	44	-1.75
Stack A, 100 kVp	AP7	-86	21	-21	-1	-26	-7	-26	-8
	BR12	-57	13	-26	1	-19	1	-19	0
	IB7	228	21	-6	-1	-32	-9	-32	-8
	KID	21	13	21	1	19	0	19	0
	LN10	-698	23.75	2	-1.75	-32	-9.75	-34	-9.75
	LV	67	40	18	8	-14	-20	-13	-24
	RB2	518	44	19	-2	13	-23	13	-24
	SB5	1500	45	68	12	124	-20	125	-21
	WT1 (block)	16	15	-24	-2	-31	-3	-31	-3
	WT1 (slab)	8	16	-23	0	-22	-2	-22	-2
	WTe	68	27	20	-2	26	-10	26	-9
Stack A, 120 kVp	AP7	-81	19.75	-21	1.25	-28	-9.75	-25	-8.75
	BR12	-52	10	-28	0	-20	1	-20	0
	IB7	201	20	-1	-3	-65	-11	-52	-9
	KID	18	10	24	0	13	-1	16	-1
	LN10	-700	20	3	2	-38	-8	-32	-8
	LV	63	36	22	11	-32	-21	-15	-23
	RB2	457	37	24	4	-62	-23	-52	-24
	SB5	1332	41	82	16	-81	-25	-98	-25
	WT1 (block)	13	11	-22	0	-32	-2	-30	-2
	WT1 (slab)	6	12	-24	2	-22	-2	-22	-2
	WTe	54	20	16	2	13	-7	12	-7
Stack A, 140 kVp	AP7	-78	17	-22	3	-27	-9	-24	-8
	BR12	-50	9	-27	-1	-20	-1	-20	-1
	IB7	183	18	-2	-4	-70	-10	-52	-7
	KID	17	8	25	0	12	-1	16	-2
	LN10	-703	17	1	2	-37	-7	-29	-7
	LV	60	33	22	9	-34	-20	-13	-20
	RB2	419	33	24	4	-78	-21	-61	-22
	SB5	1224	36	83	14	-133	-23	-139	-24
	WT1 (block)	10	9	-21	0	-30	-1	-28	-1
	WT1 (slab)	5	11	-24	1	-23	-2	-22	-2
	WTe	42	20	16	0	13	-10	11	-10

Table 3.5 – Median and interquartile range (IQR) of CT number in Hounsfield Units for each material in the measured reference images, and the difference between these metrics and those derived from simulations with 0 to 2 iterations of bone-based beam hardening correction. (Stack A)

	Material	Measured		Sim, 0 iterations		Sim, 1 iteration		Sim, 2 iterations	
		Med	IQR	ΔMed	ΔIQR	ΔMed	ΔIQR	ΔMed	ΔIQR
Stack B, 80 kVp	AP7	-113	44	-21	4	-5	-10	-5	-10
	BR12	-56.5	23	-29.5	1	-25.5	0	-25.5	1
	IB7	272	34	-1	-6	3	-7.5	3	-9
	KID	39	38	17	0	16	-5	14	-6
	LN10	-716	20	1	-1	-15	-2	-17	-2
	LV	51	24	9	4	4	6	4	6
	RB2	644	33	21	5	81	-3	90	-4
	SB5	1751	60	81	4	378	-3	449	-4
	WT1 (block)	26	41	-24	7	-28	-5	-31	-7
	WT1 (slab)	7	33	-21	-2	-18	-3	-18	-3
	WTe	106	39	6	4	21	-6	23	-8
Stack B, 100 kVp	AP7	-101	28	-23	10	-6	-7	-6	-8
	BR12	-51	16	-28	-2	-25	-2.75	-25	-3
	IB7	228	21	-1	-2	-34	-7	-32	-7
	KID	30	25.75	22	1.25	13	-7.75	14	-6.75
	LN10	-721	15	3	-2	-13	-5	-13	-4
	LV	49	14	9	3	-5	1	-3	1
	RB2	544	19	24	9	-3	-2	-5	-1
	SB5	1482	39	80	7	138	-10	140	-10
	WT1 (block)	18	27	-24	7	-28	-7	-28	-7
	WT1 (slab)	7	18	-24	1	-22	-1	-22	-1
	WTe	77	25	7	8	8	-7	8	-7
Stack B, 120 kVp	AP7	-94	22	-27	14	-5	-6	-8	-7
	BR12	-47	13	-29	-2	-25	-2	-25	-2.75
	IB7	202	18	1	-2	-70	-9	-52	-7
	KID	25	21	24	2	5	-8	13	-9
	LN10	-720	11	1	-1	-17	-3	-13	-3
	LV	48	11	8	2	-17	0	-6	0
	RB2	484	16	30	8	-84	-5	-73	-5
	SB5	1314	32	91	9	-69	-11	-83	-14
	WT1 (block)	14	21	-23	9	-31	-6	-26	-5
	WT1 (slab)	5	14	-24	1	-21	-1	-22	-1
	WTe	59	21	8	9	-2	-9	-1	-6
Stack B, 140 kVp	AP7	-89	20	-28	10	-5	-7	-8	-8
	BR12	-45	11	-28	-2	-24	-3	-25	-3
	IB7	183	17	1	-4	-74	-9	-51	-7
	KID	23	18	25	1	4	-7	14	-8
	LN10	-721	11	1	-2	-17	-4	-12	-5
	LV	46	9	10	1	-18	-1	-5	-1
	RB2	444	13	29	8	-101	-4	-80	-4
	SB5	1205	27	93	8	-119	-8	-122	-12
	WT1 (block)	11	19	-23	7	-30	-5.25	-25	-5
	WT1 (slab)	3	11	-24	2	-21	-1	-21	-1
	WTe	47	19	9	8	-2	-9	0	-5

Table 3.6 – Median and interquartile range (IQR) of CT number in Hounsfield Units for each material in the measured reference images, and the difference between these metrics and those derived from simulations with 0 to 2 iterations of bone-based beam hardening correction. (Stack B)

Chapter 4

CT calibration for proton radiotherapy treatment planning

4.1 Dose calculation for proton radiotherapy

The predominant phenomena underlying the radiation dose deposited in a known medium by a beam of protons are well understood and quantified; this can be seen for example in the Bethe-Bloch equation (equation (4.1)) and its many corrections, and in Molière's scattering theory. Unfortunately the use of such fundamental calculation methods in radiotherapy treatment planning (RTP) presents difficulties, not least because they require detailed physical characterisation of the treatment machine and the patient. Modern commercial proton treatment planning software (TPS) packages, such as Elekta's XiO or Varian's Eclipse, use semi-empirical approaches derived from the pencil beam method[1]. Monte Carlo methods have shown great promise for use in RTP, but their adoption has been limited by the processing power needed to produce results in the required timescale.

$$-\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{4\pi}{m_e c^2} \cdot \frac{nz^2}{\beta^2} \cdot \left(\frac{e^2}{4\pi\epsilon_0}\right)^2 \cdot \left\{ \ln\left[\frac{2m_e c^2 \beta^2}{I \cdot (1-\beta^2)}\right] - \beta^2 \right\}, \tag{4.1}$$

where $-\frac{\mathrm{d}E}{\mathrm{d}x}$ is the kinetic energy lost by the proton per unit distance in a medium with volumetric electron density n and mean excitation potential I. m_e is the mass of the electron, c is the speed of light, z is the charge of the proton in elementary units (1), β is the speed of the proton as a fraction of c, e is the elementary charge and ϵ_0 is the permittivity of free space.

[1]The pencil beams referred to here are distinct from those used in delivery of scanned treatments, and the method is just as applicable to scattered beams.

Whilst a range of possibilities exist in terms of physical models, the planning CT is a crucial input for all clinical applications. Two pencil beam algorithms that exist in commercial RTP software will be used to illustrate this fact: an early method described by Hong et al. in 1996 [66] and a more developed approach published by Soukup et al. in 2005 [67][2]. Both will be referred to below by the first author's name.

4.1.1 The pencil beam method

The pencil beam concept in radiotherapy dose calculation was popularised by Hogstrom in a 1981 paper on electron beam therapy [68].

$$D(x,y,z) = \iint S(x',y') \cdot d(x-x', y-y', z)\ \mathrm{d}x'\ \mathrm{d}y' \tag{4.2}$$

Equation 4.2 illustrates the approach for the special case of radiation incident parallel to the z direction. The dose distribution arising from such a radiation field is calculated in the patient coordinate system (x, y, z) by assuming that infinitesimally narrow beams of radiation (the 'pencil beams') originate from each point of the accelerator's final scatterer (x', y'), their initial fluence being defined by the function $S(x', y')$. The final dose distribution D results from a convolution with the function $d(x, y, z)$, which is the spatial distribution of dose deposited by a pencil beam originating at the origin, travelling along the z axis and impinging on semi-infinite target material $(z > 0)$[3]. The equation can easily by generalised for beams incident in other directions through coordinate system transformations.

It is often convenient to split the function $d(x, y, z)$ into depth-dose and scatter components for charged particles, DD and f_s respectively in equation (4.3), because models exist to characterise these two features independently. Inhomogeneities arising from anatomy within a real patient cause both components to vary in three dimensions, which is characterised using the CT number distribution.

$$d(x,y,z) = DD(x,y,z) \cdot f_s(x,y,z) \tag{4.3}$$

4.1.2 Depth-dose considerations

Significant computing power is necessary to calculate rapidly and repeatedly the results of the differential Bethe-Bloch equation coupled with a range straggling formulation in

[2]These algorithms have been implemented in Elekta's commercial RTP system XiO, for passively scattered and scanned proton treatments respectively.

[3]These are known as the pencil beam boundary conditions.

a heterogeneous medium. Clinical TPS developers have preferred to work with closed-form approximations of the depth-dose curve, which are calibrated using measurements performed in a water tank. A common approach is to interpolate measured or Monte Carlo simulated depth-dose data. This is computationally simple but, as it is not instructed by any of the underlying physics, it is hard to predict the effects of changing beam properties or the target material. Many sets of measurements must thus be taken in order to build up a library of dose-depth curves covering the useful energy range. Alternatively, it is possible to use a smaller amount of data to inform the same range of predictions by following Bortfeld's more mechanistic approach [69], the results of which are summarised in equation (4.4).

$$DD(z) = \Phi_0 \frac{e^{-\zeta^2/4}\sigma^{1/p}\Gamma(1/p)}{\sqrt{2\pi}\varrho p\alpha^{1/p}(1+\beta R_0)} \left[\frac{1}{\sigma}\mathcal{D}_{-1/p}(\zeta) + \left(\frac{\beta}{p} + \gamma\beta + \frac{\epsilon}{R_0}\right)\mathcal{D}_{-1/p-1}(\zeta)\right] \quad (4.4)$$

where $R = \alpha E_0^p$, $\zeta = \frac{z-R_0}{\sigma}$ and $\sigma = \sqrt{\sigma_{mono}^2 + \sigma_{E_0}^2 \alpha^2 p^2 E_0^{2p-2}}$, Γ is the gamma function and $\mathcal{D}_\nu$ is the parabolic cylinder function with parameter ν. For beams of therapeutic energy, a good approximation is $\sigma_{mono} = \alpha' \frac{p^2\alpha^{2/p}}{3-2/p} R_0^{3-2/p}$.

Nine parameters in equation (4.4) must be fixed in order to generate dose-depth curves for a particular situation. Three of these are beam-specific parameters: nominal energy E_0 MeV, width of Gaussian energy spectrum σ_{E_0} MeV and low energy tail fluence fraction ϵ. The remaining six parameters are properties of the material and energy regime in question: mass density ϱ, range-relation constants α and p, inelastic nuclear interaction parameters β and γ, and range straggling parameter α'. Appropriate values have been determined for water at therapeutic energies, allowing the necessary beam parameters to be extracted by a fit of the function to broad beam water tank depth-dose measurements.

4.1.2.1 Stopping power ratios

Whilst soft tissue is similar in density to water, it should be noted that its chemical composition differs. Some materials within the body significantly differ from water in both density and composition, for example bone. Traditionally, the reference dose-depth curves used by a TPS are calculated or measured in water, with the differences in tissue accounted for through use of a 'stopping power ratio' (SPR). A particular tissue is treated as water with a 'water equivalent thickness' related to true thickness by multiplication with the mass density and the inverse of the SPR. The sum of the water equivalent thickness along a particular line is known as radiological path length.

Hong's algorithm uses a dose-depth function $DD(x, y, z)$ which pertains to the central axis of the beam; thus, when water measurements are scaled for use in a material with SPR $\neq 1$, it is necessary to make an inverse square correction of magnitude $\left(\text{radiological path length}\backslash\text{physical path length}\right)^2$. In Soukup's algorithm this function is the dose-depth integrated over the lateral plane (IDD) and so no correction is required, geometric divergence being accounted for in the lateral scatter component $f_s(x, y, z)$.

Adequate determination of SPR for all irradiated tissues is essential for the accurate prediction of dose distributions, especially given that they affect the estimated proton range, beyond which the dose deposited rapidly falls to zero. The SPRs are calculated from the planning CT scan, using models that have been calibrated for a specific protocol on a specific CT scanner. In order to avoid clinically significant side effects from over- or under-irradiation, there must be an awareness of the magnitude of uncertainties inherent in this process.

4.1.2.2 Dose-to-water versus dose-to-tissue

Radiotherapy treatments have traditionally been prescribed in terms of dose-to-water. If a region is sufficiently small that it has negligible impact on the beam's particle and energy spectra, dose-to-water can be considered equivalent to the dose that would be deposited in that region if it was instead composed of water. It is important to have such a standard for physical dosimetry as the sensitive volume of dosimeters can be constructed from a variety of materials. Converting between dose-to-material and dose-to-water requires a minimum of two corrections: one for the change in relative density and another for the stopping power differences due to differing composition. The clinical acceptance of dose-to-water is convenient for the use of pencil beam algorithms as scaling and interpolation of dose-depth data through the water equivalent thickness method implicitly produces exactly that.

Paganetti gives a detailed review of the relative merits of dose-to-water versus dose-to-tissue [70]. If the choice is simply between two possible reporting conventions, for which there is sufficient information to perform inter-conversion, the decision is a matter of which provides the most convenient reference standard. In the context of reporting treatment dose, there is likely merit in the status quo; as Paganetti says "all clinical experience is based on [dose-to-water], and tumor cells in the human body consist of mostly water". In practice, however, the calculation methodology is typically linked to the desired output quantity: Monte Carlo calculations do not typically retain

sufficient information to enable post-hoc conversion[4], and modelling the patient's composition as water of varying density may lead to large differences in nuclear effects and low-energy stopping power variation when compared with a full treatment of elemental composition. The results in Paganetti's paper suggest that, in bone regions, these calculations can lead to relative dose differences of 10% and Bragg Peak range differences approaching 3 mm.

4.1.3 Lateral spread

Lateral spreading of a proton beam mostly arises from multiple Coulomb scattering (MCS) of primary protons by the electrons in a medium. A small amount of dose is deposited considerably further from the axis of incidence than would be expected by MCS alone, as a result of inelastic nuclear interactions in which secondary protons or other ions are produced.

At clinical energy and length scales, MCS is modelled well by a Gaussian distribution whose standard deviation increases with thickness of material traversed. Gottschalk et. al. have reviewed a number of approximations which may be adopted to calculate the r.m.s. MCS angle at discrete steps along the beam's path [71]. Equation 4.5 shows Rossi's expression for the r.m.s. angular spread θ_0 that results when protons with velocity βc and momentum p undergo MCS in a thin[5] slab of material with thickness Δz and radiation length X_0 [72]. Rossi derives $E_s = m_e c^2 \cdot \sqrt{\frac{2\pi}{\alpha}}$, where m_e is the mass of an electron and α is the fine structure constant[6]. In the Soukup algorithm, θ_0 is evaluated for every step up to the desired depth z_M and the resulting spreads are combined in quadrature as equation (4.6). The user may also select MCS formulations attributed to Highland or Lynch and Dahl.

$$\theta_0^2(z) = \left(\frac{E_s}{\beta \cdot p}\right)^2 \cdot \frac{\Delta z}{X_0(z)} \tag{4.5}$$

$$\sigma_{MCS}^2(z_M) = \sum_{m=1}^{M} \theta_0(z_m)^2 \cdot z_m^2 \tag{4.6}$$

Hong's algorithm uses the Highland approximation of MCS. In all of the methods described, radiation length is a necessary material parameter. Hong works around this

[4] A full treatment would require incident energy and dose deposition on a single particle level.

[5] such that energy loss in negligible

[6] $E_s = 14.99$ MeV using the latest CODATA constants [73], but Soukup et. al. report that use of a smaller value leads to better agreement with simulations.

requirement by calculating the MCS spread as that expected in a thickness of water equal to the radiographic path length. However, the inverse relationship between SPR and radiation length that this implies agrees poorly with measurements. Soukup instead estimates radiation length from the planning CT[7] for which, once again, a calibration process must be carried out.

Nuclear effects are not considered in the lateral direction in Hong's algorithm. Soukup introduces a second Gaussian with an amplitude and deviation that increases with depth whilst the amplitude of the primary particle Gaussian correspondingly decreases. The precise depth-dependences are determined by fitting data from Monte Carlo simulations in water, and are mapped to other materials using the radiographic path length.

4.2 CT calibration

4.2.1 Stoichiometric method

The purpose of a CT calibration curve is to provide a lookup function which links CT number with the desired physical property y for all tissues within a patient[8]. There are a number of practical issues with the direct use of human tissue for such calibration. An alternative which requires only inert and ethically-unencumbered samples is the stoichiometric method, first proposed by Schneider in 1996 [74] but built on by a number of other workers [75, 76]. A generalised summary is as follows:

1. CT scan a selection of tissue equivalent materials, with known density and composition.

2. Fit the resulting CT numbers to a function of the density and elemental composition, determining coefficients that reflect the strength of each major x-ray attenuation process for the energy spectrum used in the scan.

3. Use the coefficients from step 2 to predict CT number (the independent variable in the calibration curve) for a range of real biological tissues, using reference data for their density and elemental composition. For each of these, also calculate the desired dependent variable (the choice of which is examined in Section 4.2.2).

[7]This will be estimated using CT values from the central axis of the pencil beam: predictions can thus deviate from measurements if significant density variations occur parallel to the axis. Appropriate selection of the sampling interval used to discretise the convolution in equation (4.2) will mitigate the issue somewhat.

[8]In proton RTP, y may be the mass density or SPR, but the method is more general and y could represent other properties for other radiation modalities.

4. Use the points from step 3 to specify a one-to-one lookup function linking CT number and y, through interpolation or a fitting process.

The specifics of each of these steps vary between implementers, can significantly affect the resulting curve, and will be explored further in later Sections. The following Section discusses an important point that must first be established, namely the physical property y to which the CT scanner's scale is calibrated.

4.2.2 Mass density vs. SPR

As is evident from the Bethe-Bloch formula, there is an energy dependence to stopping power. Electron density and mean excitation energy for a particular material will generally differ from those of water, so the SPR will also vary with the proton's residual energy. Figure 4.1 demonstrates this: it is notable that the majority of variation occurs below 10 MeV, where the proton's range in water is no more than 1.3 mm. Soukup et al. summarise the relationship between SPR and energy as "weak but non-negligible", and highlight their earlier finding [77] that SPR[9] in a wide range of human tissues[10] can be predicted within 1% by a formula of two variables: energy and mass density. In the same paper, radiation length[11] is also related to mass density with similar precision. They consequently elect to use a mass density map as the input data for modelling both stopping and scattering within the patient, requiring the calibration process to establish a mapping between CT number and mass density. In contrast, the algorithm of Hong et al. derives the SPR at points within the patient using a calibration curve that directly maps CT number and SPR, making the assumption that SPR is independent of the properties of the incident proton. It would only require a small modification to Soukup's algorithm in order to accommodate a similar input under the same assumption.

There is a question as to whether SPR or mass density constitutes the best variable for the y-axis of a CT calibration curve. Whilst associating HU values with a fixed SPR ignores the energy dependence and shifts the onus of determining stopping powers and radiation lengths on to the user, the empirical formulae linking mass density with these properties are valid only for tissues whose composition and density are sufficiently similar to those specified in ICRU 46.

[9] as calculated by the PSTAR tool [78]

[10] all those specified in ICRU Report 46 [79], except gallstone, protein, carbohydrate and urinary stone

[11] as calculated using Tsai's equations [80], assuming no molecular effects

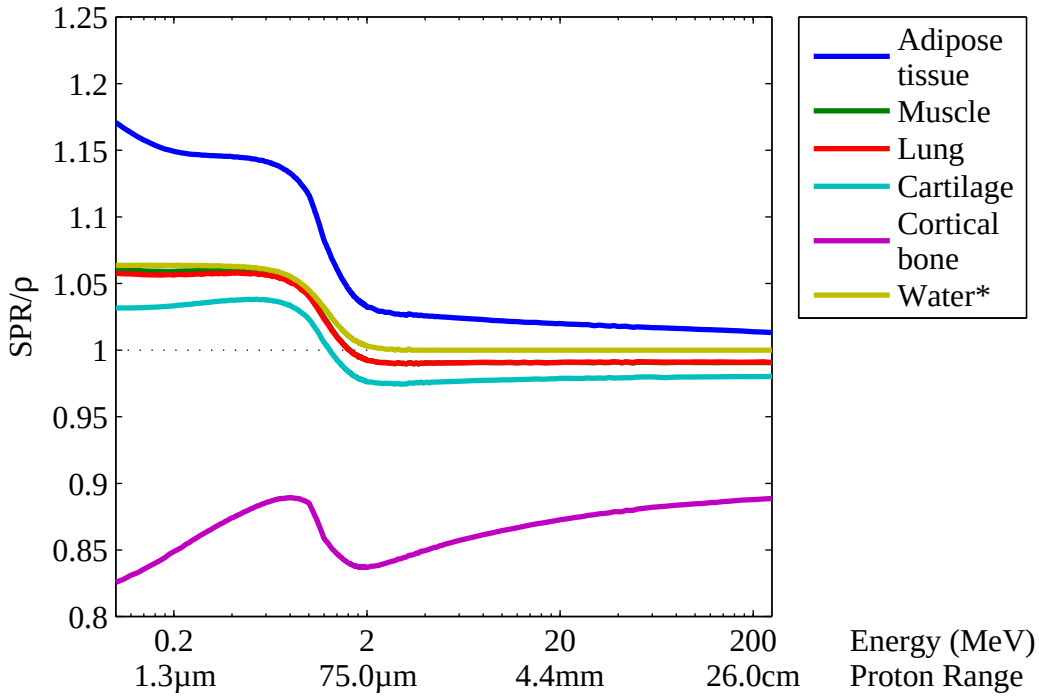
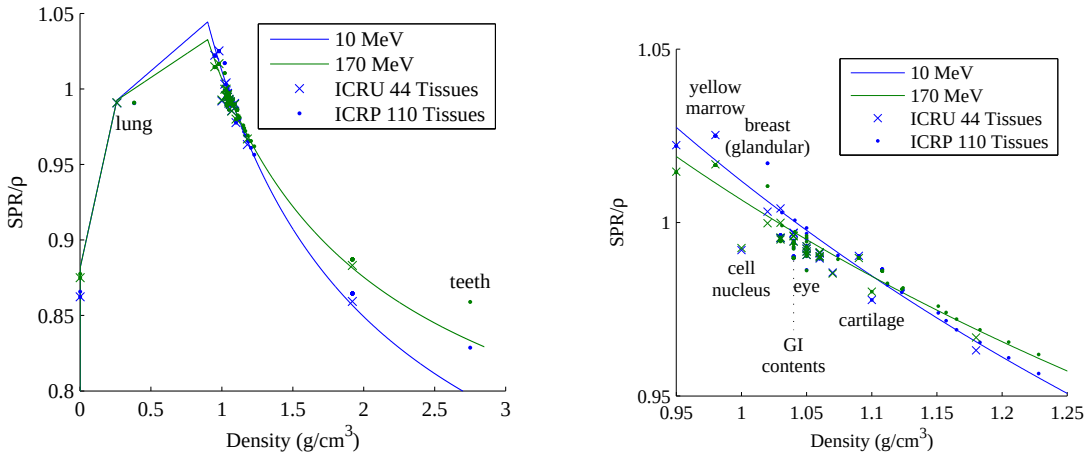


Figure 4.1 – Mass stopping power ratios as a function of energy for four tissues in the ICRP Adult Male reference phantom. Stopping powers for these tissues were calculated using the SRIM code (version 2013 [81]), treating the materials as solid mixtures of independent atoms (no compound corrections). It is known that this assumption leads to an underestimation of the mean excitation energy for water, affecting stopping power estimates less than ˜2 MeV, as shown by the line labelled Water*. The water stopping power in the SPR denominator uses SRIM's recommended compound correction parameter of 0.94, which is interpreted as a reduction of up to 6% in stopping power at low energies. Proton range estimates are in water, using R from equation (4.4) with $\alpha = 0.0022$ and $p = 1.77$.



(a) All tissues **(b)** Soft tissue region

Figure 4.2 – Line plots of Fippel's SPR-density relation [77], reproduced in equation (4.7), at 10 MeV and 170 MeV. Also shown as points are SRIM-calculated SPR values (no compound corrections) at each energy, for tissues with elemental compositions and densities as specified in ICRU Report 44 [64] and ICRP Publication 110. It should be noted that Fippel used PSTAR, rather than SRIM, to calculate stopping powers. Points lying furthest from the line are labelled with the corresponding tissue.

4.2.2.1 Significance of SPR energy dependence

The major benefit of calibrating CT numbers with mass density is that the TPS may then consider the energy dependence of SPR internally, allowing the user to work with a material-specific constant; however, doing so also requires the assumption that the tissue in question belongs to the family of materials for which the algorithm's SPR function was derived. The extent to which energy variation and alternative compositions for the same tissue affect proton range calculations can be investigated by comparing Monte Carlo and pencil beam dose-depth data, such as in the short study which follows.

Longitudinal aspects of Soukup's pencil beam algorithm were implemented in a function for the MATLAB programming language. A modification was introduced to allow SPRs to be calculated by two different methods: the formula used in Soukup's algorithm (from a paper by Fippel and Soukup [77], hereafter referred to by the first author's name), which is reproduced in equation (4.7) and illustrated in figure 4.2, and stopping power calculation with the SRIM[12] code.

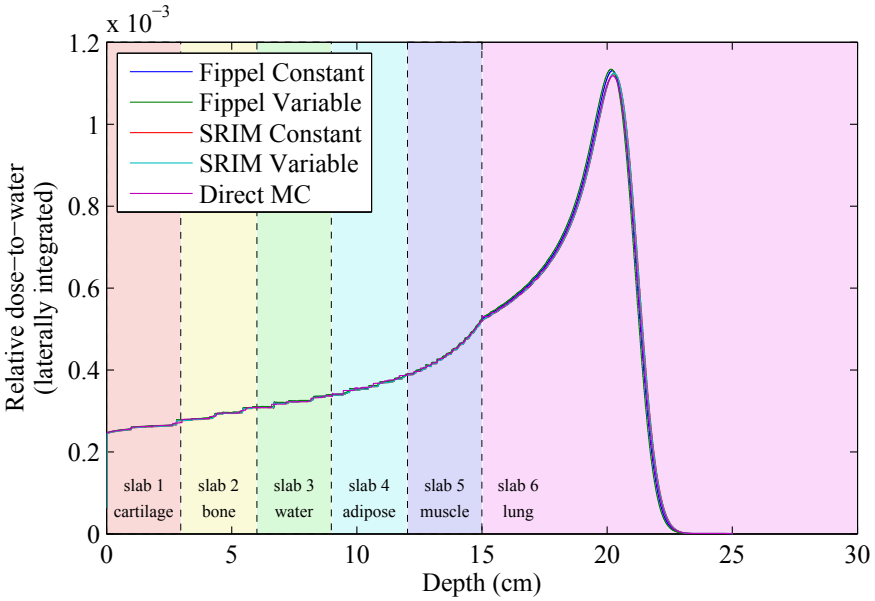
[12]Stopping Power and Range In Matter, version 2013 [81]

$$
\begin{aligned}
SPR(\rho, E) &= \frac{\rho}{1.000\,\mathrm{g}\cdot\mathrm{cm}^{-3}}\, f_S(\rho, E) \\
f_S(\rho, E) &= \begin{cases}
1.0123 - 3.386\cdot 10^{-5}E + ... & \rho \geq 0.9 \\
\quad 0.291\cdot(1 + E^{-0.3421})\cdot(\rho^{-0.7} - 1) & \\
0.9925 & \rho = 0.26 \text{ (lung)} \\
0.8815 & \rho = 0.0012 \text{ (air)} \\
\text{interpolate} & \text{all other } \rho < 0.9
\end{cases}
\end{aligned}
\tag{4.7}
$$

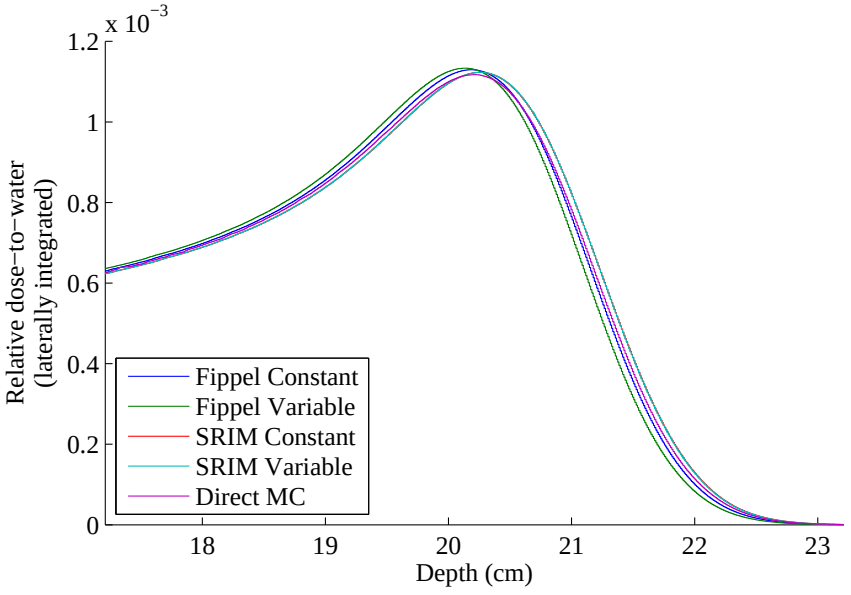
It can be seen that the only material parameters necessary for Fippel's formula are mass density, whilst SRIM also requires elemental composition data. Both methods allow for a full treatment of the energy variation, or the adoption of a single SPR derived at high-energy (170 MeV) for each material. The reference pencil beam data was generated by the Monte Carlo particle transport code MCNPX v2.7.0 [82] with the default LA150 cross section dataset [83]. A collimated beam of protons with a Gaussian lateral fluence profile with $\sigma = 1\,\mathrm{mm}$ and a Gaussian energy profile with $\mu = 170\,\mathrm{MeV}$ and $\sigma = 1\,\mathrm{MeV}$ was generated to be incident perpendicular to the centre of a face on a $40\,\mathrm{cm} \times 40\,\mathrm{cm} \times 40\,\mathrm{cm}$ water cube. The energy deposited was recorded using a cylindrical mesh tally[13] with radial bins of 1 mm and depth bins of 0.1 mm within a region measuring of diameter 12 cm and depth 30 cm. Only protons (including secondary protons arising from elastic recoil) were transported to a minimum energy of 100 keV, at which point the remainder of their energy was deposited. The energy of all other particles was deposited at the point of generation.

Variation in SPR was introduced by considering a geometry comprising five neighbouring $3\,\mathrm{cm} \times 40\,\mathrm{cm} \times 40\,\mathrm{cm}$ slabs, each composed of a different material, followed by a final $25\,\mathrm{cm} \times 40\,\mathrm{cm} \times 40\,\mathrm{cm}$ slab. The following six materials were considered, in this order and its cyclic permutations: adipose tissue, muscle, lung, cartilage, cortical bone and water. The arrangement is illustrated in figure 4.3a. MCNPX simulations were carried out in each geometry with parameters that were otherwise identical to the water tank simulation. No molecular perturbations have been included in either the Monte Carlo results or SRIM calculations, as the standard release of MCNPX is unable to consider them. Monte Carlo dose-to-water in arbitrary materials was obtained by dividing energy deposition per unit volume by the mass density and the SRIM-calculated energy-dependent SPR.

[13] type CMESH3, discussed in detail in the MCNPX documentation [82]



(a) Illustration of the slab-based geometry used for the study described in Section 4.2.2.1. Each slab is homogenous and has a cross-section measuring 40 cm × 40 cm. Superimposed is exemplar dose-depth data for the arrangement that ends with lung, using ICRP Publication 110 tissue definitions.



(b) A reproduction of the curves seen in (a) with x-axis encompassing just the Bragg Peak and distal penumbra. Even at this scale, the lines corresponding to SRIM Constant and SRIM Variable are indistinguishable.

Figure 4.3

ICRU Report 46 Tissue Definitions

Final material:	Water		Cortical bone		Cartilage		Lung		Muscle		Adipose	
distances in mm	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$
Fippel, Constant	-0.3	0.0	-0.4	0.0	-0.2	0.0	-0.6	-0.1	-0.1	0.0	0.0	0.0
Fippel, Variable	-0.3	0.0	-0.2	+0.3	-0.4	0.0	-1.2	-0.1	-0.3	0.0	-0.2	-0.1
SRIM, Constant	0.0	0.0	-0.2	0.0	+0.2	-0.1	+0.8	-0.1	+0.2	0.0	+0.4	+0.1
SRIM, Variable	+0.2	0.0	+0.1	+0.2	+0.2	0.0	+0.8	-0.1	+0.3	0.0	+0.4	-0.2

ICRP Publication 110 Tissue Definitions

Final material:	Water		Cortical bone		Cartilage		Lung		Muscle		Adipose	
distances in mm	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$
Fippel, Constant	-0.1	0.0	-0.3	-0.1	0.0	-0.1	-0.4	-0.1	0.0	-0.1	0.0	0.0
Fippel, Variable	-0.1	-0.1	-0.1	0.0	-0.2	0.0	-0.9	-0.1	-0.2	-0.1	-0.1	-0.1
SRIM, Constant	+0.1	0.0	-0.2	0.0	+0.2	0.0	+0.3	+0.01	+0.2	0.0	+0.3	+0.1
SRIM, Variable	+0.2	0.0	+0.1	0.0	+0.2	0.0	+0.3	0.0	+0.3	-0.1	+0.2	-0.2

Table 4.1 – Differences in proton range R and $d_{80\text{-}20}$ between pencil beam and direct Monte Carlo dose calculations for 170 MeV protons in the geometry described in Section 4.2.2.1 and illustrated in figure 4.3. Stopping power ratios have been calculated using the empirical formula of Fippel (a function of mass density and proton energy) and the SRIM code (which also uses the full elemental composition), considering the cases where the high-energy (170 MeV) SPR is assumed to apply for all energies ('Constant') and where the SPR is allowed to vary with energy ('Variable'). $d_{80\text{-}20}$ is a measurement of the distal penumbra size, corresponding to the distance between the 80% and 20% distal isodose lines. Direct Monte Carlo calculations each ran for 100 minutes on one core of an Intel Xeon E5-2650 CPU, leading to relative uncertainties below 0.25% in each longitudinal bin.

Recommended values for the elemental composition and density for the tissues in question differ between publications, and table 4.1 shows the variation in Bragg Peak characteristics obtained when using those from ICRU Report 46 (as in the derivation of Fippel's formula) and ICRP Publication 110. As may be expected, Fippel's formula performs best when using tissue specifications from the former, but the difference is slight – range and distal falloff agree to at least 1 mm in all of the arrangements considered, regardless of the family of tissue compositions. It is evident that in this simple geometry, which nevertheless includes a range of tissues, the scale of variation in longitudinal dose profile resulting from SPR energy dependence is unlikely to be of clinical significance. Differences between the SRIM and MCNPX results in table 4.1 can reasonably be attributed to alternative stopping power calculation methods/cross-section data used in MCNPX and SRIM, to effects relating to the size of calculation steps in the pencil beam algorithm[14] and to quantisation error in scaling the reference dose-depth data[15].

4.2.2.2 Fippel's model in the lung

The greatest variations in range calculated with the pencil beam method and MCNPX occur when the final slab consists of ICRP lung. As the density of lung is much lower, the spatial scale of the range uncertainties is correspondingly increased. However, it should be noted from figure 4.2a that the calculated f_S for ICRP lung is one of the furthest outliers from the predictions of the model. ICRP lung has an identical composition to ICRU lung but a greater density (0.382 vs 0.260). As the region $\rho < 0.9\,\mathrm{g \cdot cm^{-3}}$ is poorly populated by tissues, the predicted f_S is obtained by linear interpolation between the single point for lung and the bottom end of the soft tissue region. The least dense soft tissues contain a high proportion of fat, leading to compositions that are markedly different to that of lung. This results in an overestimation of the SPR in ICRP lung by approximately 1% at 170 MeV and 1.25% at 10 MeV: the observed 1 mm undershoot corresponds to a similar proportion of the thickness of lung traversed.

[14]Efforts have been made to limit step-size effects by using multiple calculation steps per dose voxel in the pencil beam implementation here. For each step, material and average proton energy are sampled at the start point and assigned to the neighbouring region equal to 20% of the final voxel size. SPR is calculated from these parameters and the energy deposition expected given the reference dose-depth curve is assigned to the region. The material is then sampled again for the next step. Finally, the resulting distribution is resampled to the dose grid and scaled to dose-to-water.

[15]Whilst the depth resolution of the reference data was 0.1 mm in water, the water equivalent thickness of a dose voxel with $\rho < 1\,\mathrm{g \cdot cm^{-3}}$ will be smaller.

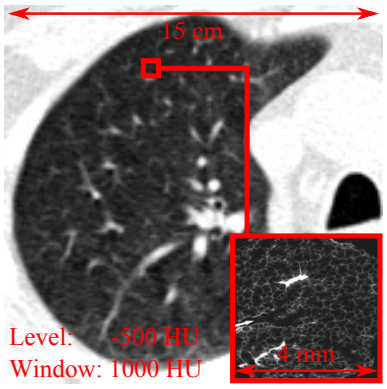


Figure 4.4 – Heterogeneity in lung tissue shown in a CT slice from the ROCOCO dataset and, inset, a micro-CT with area equivalent to sixteen CT voxels. The micro-CT originally appeared in Watz et. al Radiology 2005;236:1053-1058 [84] and is reproduced with permission from the RSNA.

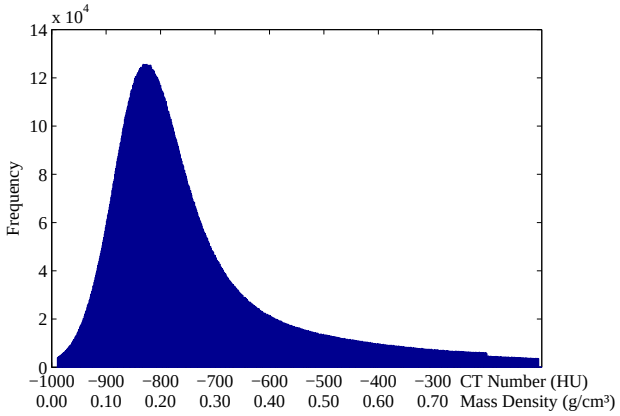


Figure 4.5 – Aggregate CT number distribution within the regions contoured as lung for 25 non-small cell lung cancer patients whose data was included in the ROCOCO in-silico planning trial [85]. Representative mass densities are calculated assuming an x-ray energy of 80 keV and ICRP/ICRU lung composition (median/mean = $0.22/0.27\,\mathrm{g\cdot cm^{-3}}$). Voxels below -990 HU and above 100 HU (2.3% of the total number) were not included.

When large volumes of lung are present, the range uncertainty described above may approach clinically significant scales. Moreover, regardless of the ideal bulk average for lung tissue density, CT scans of the lung do not show regions of great homogeneity. The lung has a complex structure on a range of scales, as illustrated in figure 4.4, resulting in CT voxels with a range of average densities. In the approximation that CT voxels with $\rho < 0.9\,\mathrm{g\cdot cm^{-3}}$ contain only ICRP/ICRU lung tissue appearing at different densities[16], it would seem inadvisable to perform linear interpolation of f_S between air at $0.0012\,\mathrm{g\cdot cm^{-3}}$ and lung at $0.26\,\mathrm{g\cdot cm^{-3}}$, and between lung at $0.26\,\mathrm{g\cdot cm^{-3}}$ and soft tissue at $0.9\,\mathrm{g\cdot cm^{-3}}$, as recommended by equation (4.7). f_S is purely dependent on elemental composition and independent of mass density, which is accounted for separately in the estimation of SPR.

Figure 4.5 shows a histogram of CT numbers within lung tissue for a dataset of 25 patients, which have been mapped to mass densities using equation (2.5) and equation (3.1) for ICRP/ICRU lung at 80 keV[17]. These data were used to investigate

[16]This approximation should be valid if they contain a lung-air mixture, the change in composition arising from the presence of air being negligible. ICRU Report 44 gives the density of air as $0.0012\,\mathrm{g\cdot cm^{-3}}$ and deflated lung as $1.05\,\mathrm{g\cdot cm^{-3}}$.

[17]Estimated mass densities change by no more than 1% over the range 50-150 keV

the possible errors in range introduced by the linear interpolation of f_S in variable density lung. The CT number distribution was sampled 300 times at intervals corresponding to one-third-percentiles and the resulting densities assigned to neighbouring $1\,\text{mm} \times 40\,\text{cm} \times 40\,\text{cm}$ slabs of ICRP/ICRU lung in four arrangements: with density ramping-up, ramping-down, ramping-up then ramping-down (peaked) and alternating high and low densities which converge on the median. Monte Carlo and pencil beam methods were used to calculate dose-depth curves for 60 and 90 MeV beams of protons in an identical manner to section 4.2.2.1. The results are tabulated in table 4.2 and depicted in figure 4.6.

In this one-dimensional lung model with a single material at multiple mass densities, the range prediction errors introduced by using equation (4.7) to estimate SPR are approaching a clinically significant scale, ranging from 0.7 mm to 3.7 mm depending on mass distribution and proton energy. When considered in terms of water-equivalent path length, as in the second half of table 4.2, the same errors correspond to 0.2-3.1% of the 90 MeV proton range. With one exception, the water-equivalent deviations for SRIM-derived SPRs are less than 0.25% and vary little with density arrangement, implying that most of the differences in the Fippel case can be attributed to SPR calculation rather than the pencil beam implementation. As in Section 4.2.2.1, the changes resulting from allowing SPR to vary with energy were relatively insignificant.

The tabulated results are best explained with reference to the the ideal form of figure 4.2 for this situation, which is a horizontal line between $0.011\,\text{g}\cdot\text{cm}^{-3}$ (-990 HU) and $0.9\,\text{g}\cdot\text{cm}^{-3}$ that passes through the points marked as lung. The function actually used quickly departs from the ideal situation both above and below the threshold of $0.26\,\text{g}\cdot\text{cm}^{-3}$, as may be seen in the histograms of figure 4.7. Approximately 40% of the total mass in the distribution is below the threshold, where the SPR is underestimated by up to 12%, whilst the SPR may be overpredicted by up to 5% in the remaining 60% of mass. The net result on pencil beam range predictions depends on the proportion of the mass traversed by protons (˜40% at 60 MeV and ˜80% at 90 MeV) and the density at which that mass appears. In the alternating and peaked cases at 90 MeV, protons experience the majority of the density spectrum and the opposing effects of range errors on both sides of the distribution result in a range deviation close to zero. For the same arrangements at 60 MeV, the skewness of the mass distribution has slightly different consequences – an undershoot is seen in the alternating case, where most of the mass has been traversed at high density, and an overshoot in the peaking case, where lower densities dominate as the distribution's peak is never reached. For the same reasons, range is underestimated in the ramp-down case and overestimated

At 60 MeV

Density arrangement:	Ramp up		Ramp down		Peaked		Alternating	
distances in mm	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$
Fippel, Constant	+3.7	-0.2	-2.1	-0.1	+0.7	0.0	-1.8	-0.1
Fippel, Variable	+3.7	-0.2	-2.1	-0.2	+0.7	0.0	-1.8	-0.1
SRIM, Constant	-0.3	-0.1	-0.1	-0.1	+0.3	+0.1	+0.0	0.0
SRIM, Variable	-0.3	-0.1	-0.1	0.0	+0.3	0.0	+0.0	0.0
RPL as % of range	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$
Fippel, Constant	+3.1	-0.1	-2.7	-0.1	+1.2	+0.1	-1.6	+0.1
Fippel, Variable	+3.1	-0.1	-2.7	-0.2	+1.2	+0.1	-1.6	+0.1
SRIM, Constant	-0.2	-0.1	-0.1	-0.1	+0.5	+0.2	0.0	+0.1
SRIM, Variable	-0.2	-0.1	-0.1	0.0	+0.5	0.0	0.0	+0.1

At 90 MeV

Density arrangement:	Ramp up		Ramp down		Peaked		Alternating	
distances in mm	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$
Fippel, Constant	+1.4	-0.1	-3.3	+0.2	-0.7	-0.1	-0.7	+0.1
Fippel, Variable	+1.3	-0.1	-3.3	+0.1	-0.7	-0.1	-0.7	0.0
SRIM, Constant	+0.2	0.0	+0.3	0.0	+0.5	0.0	+0.3	0.0
SRIM, Variable	+0.2	0.0	+0.3	0.0	+0.5	0.0	+0.3	0.0
RPL as % of range	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$	ΔR	$\Delta d_{80\text{-}20}$
Fippel, Constant	+1.3	0.0	-1.0	+0.1	-0.3	0.0	-0.2	0.0
Fippel, Variable	+1.2	0.0	-1.0	+0.1	-0.3	0.0	-0.2	0.0
SRIM, Constant	+0.2	0.0	+0.1	0.0	+0.2	0.0	+0.1	0.0
SRIM, Variable	+0.2	0.0	+0.1	0.0	+0.2	0.0	+0.1	0.0

Table 4.2 – Differences in proton range R and distal falloff $d_{80\text{-}20}$ between pencil beam and direct Monte Carlo dose calculations for 60 and 90 MeV protons in the lung geometries described in Section 4.2.2.2 and illustrated in figure 4.6. Stopping power calculation methods and the definition of $d_{80\text{-}20}$ are as the caption of table 4.1. The constant and variable energy lines are almost indistinguishable in all cases. Also shown are the same quantities measured in radiological path length, reported as a percentage of the proton range in water at the energy in question. Direct Monte Carlo calculations each ran for 10 CPU hours on Intel Xeon E5-2650 processors, leading to relative uncertainties below 0.3% in each longitudinal bin.

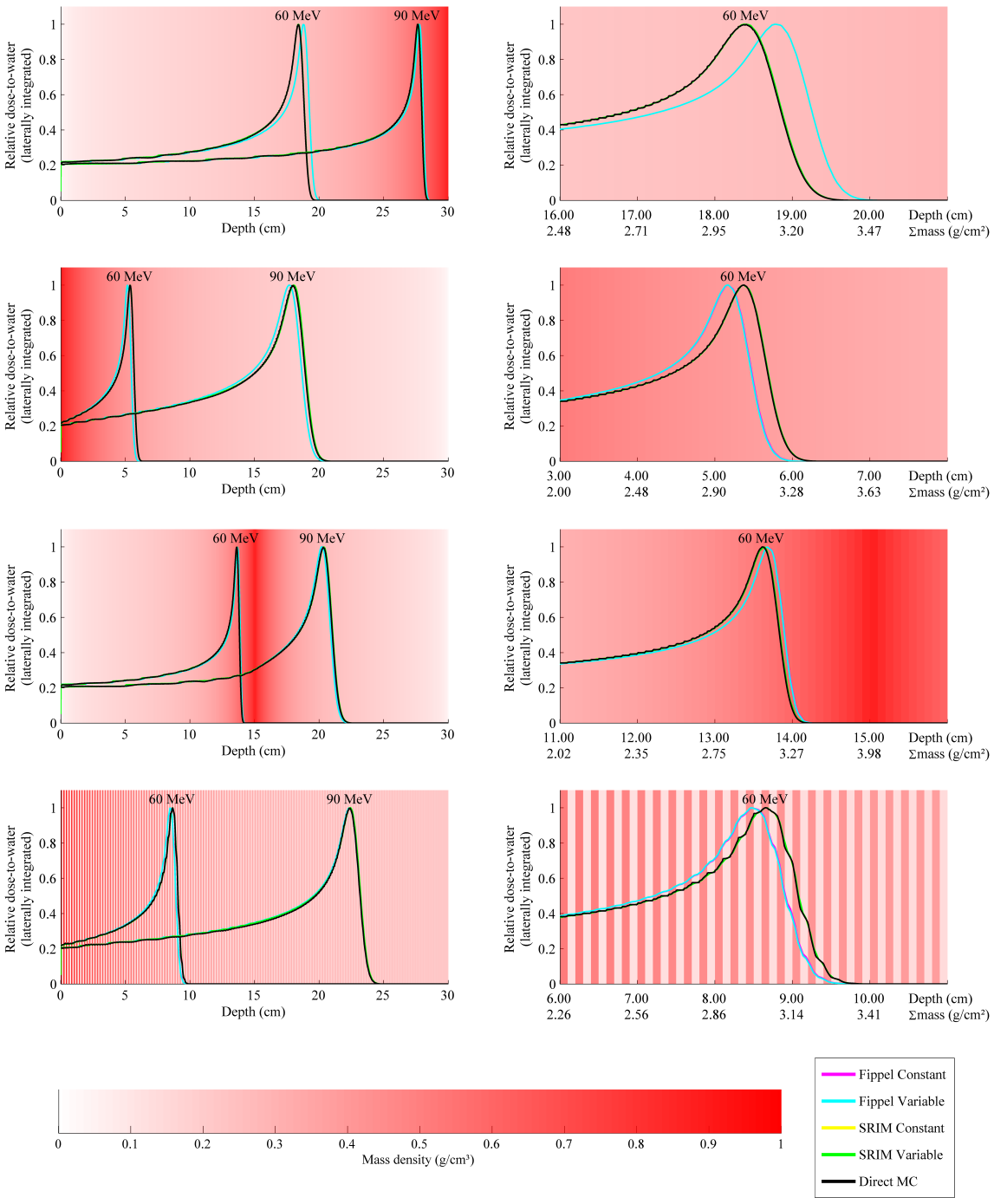
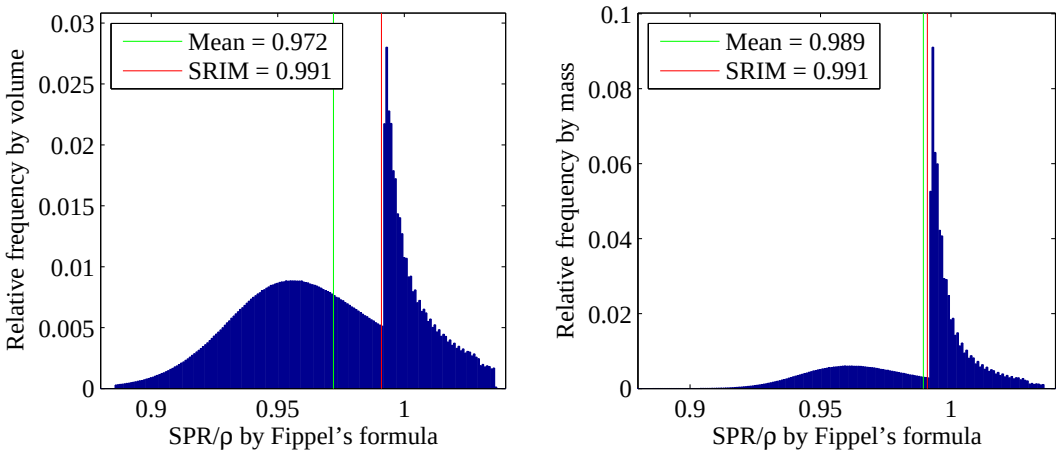


Figure 4.6 – Left: Longitudinal dose calculations for 60 and 90 MeV protons in 30 cm of lung tissue with mass density distribution as given in figure 4.5. Results are shown for four spatial distributions of mass, each with pencil beam calculations using the SPR-determination methods described in Section 4.2.2.1 and a direct Monte Carlo simulation. Right: The 60 MeV data enlarged to show the region of the Bragg peak, with a secondary axis displaying the longitudinal integral of mass density. In each case two groups of overlapping lines can be seen: the 'Fippel Constant' and 'Fippel Variable' lines coincide, as do the lines for 'SRIM Constant', 'SRIM Variable' and 'Direct MC'.



(a) Distribution by volume **(b)** Distribution by mass

Figure 4.7 – Histograms of mass stopping power ratio as estimated by equation (4.7) for the density distribution in figure 4.5, with frequencies weighted by a) volume and b) mass. In the mass-weighted case, the distribution mean is almost equal to the SRIM calculated SPR for ICRU/ICRP lung; in the volume-weighted case the figures differ by approximately 2%.

in the ramp-up case at both energies. In general, the relative size of the errors (in water-equivalent terms) decreases as the density histogram within the traversed mass approaches that assigned to the whole volume: serendipitously, for the distribution utilised here, the mass average of the SPR calculated by Fippel's method is within 0.2% of the SRIM-calculated value.

In clinical proton radiotherapy of the lung, uncertainties in physical characterisation of the patient are taken into account by expanding the (internal[18]) clinical treatment volume. Typical proximal margins are 3.5% of the tumour's proximal water equivalent depth and typical distal margins are 3.5% of the tumour's distal water equivalent depth [86] (both excluding 3 mm setup uncertainties). Such margins would be sufficient to encompass the largest deviations detailed in table 4.2, but it is conceivable that this may not be the case for even lower energies. It should be acknowledged that such large relative uncertainties are only likely to occur when lung at particularly low or high densities constitutes a significant proportion of the beam's water equivalent path length, and that the adequacy of the one-material-multiple-density model for in-vivo lung can only be determined by physical measurement. This example nevertheless highlights a benefit of calibrating CT numbers to stopping power rather than mass

[18]For moving targets, volumes that encompass the entire range of motion of the tumour are denoted 'internal'.

density, in that it allows the physicists commissioning a proton therapy system to be fully aware of, and to take full responsibility for, the methods by which such an important physical property is determined.

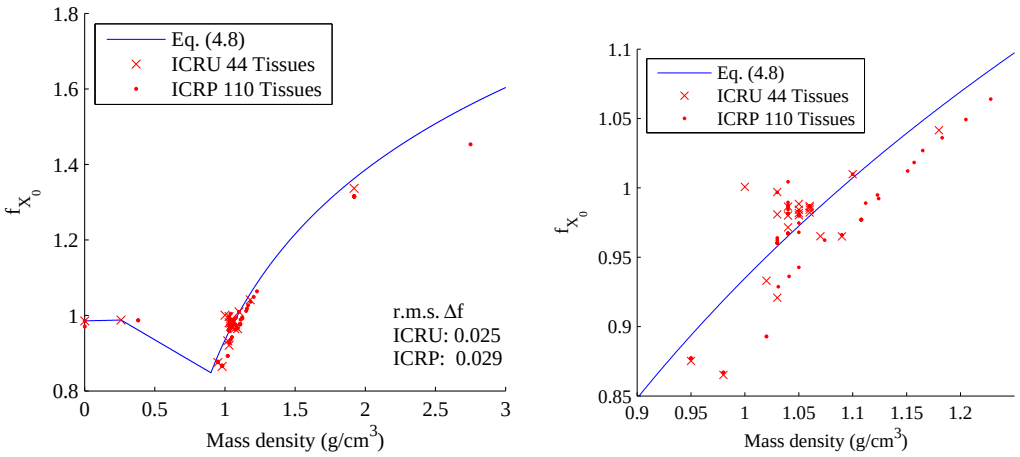
4.2.2.3 Radiation length considerations

In the Soukup algorithm, radiation length is used to calculated how the lateral penumbra broadens as the pencil beam axis passes through a particular voxel. The input calibration curve is between CT number and mass density, so the algorithm uses a formula linking radiation length and mass density. As with SPR, the relationship is determined in a paper by Fippel and Soukup [77] through fits to ICRU Report 46 tissues; it is reproduced below as equation (4.8).

$$
\begin{aligned}
X_0(\rho) &= \frac{1.000\,\mathrm{g\cdot cm^{-3}}}{\rho}\,\frac{X_w}{f_{X_0}(\rho)},\ \text{where } X_w = 36.0863\,\mathrm{cm} \\
f_{X_0}(\rho) &= \begin{cases} 1.19 + 0.44\ln(\rho - 0.44) & \rho \geq 0.9 \\ 0.9879 & \rho = 0.26\ \text{(lung)} \\ 0.9857 & \rho = 0.0012\ \text{(air)} \\ \text{interpolate} & \text{all other } \rho < 0.9 \end{cases}
\end{aligned}
\tag{4.8}
$$

This equation is plotted against the ICRU Report 44 and ICRP Publication 110 tissues in figure 4.8a. The method for calculating radiation length is described in appendix A. As interpolation is once again used between air, lung, and the lower end of the soft tissue region, arguments similar to those in previous section could be used to hypothesise that the lateral penumbra may be over- or underestimated in lung tissue appearing at densities other than $0.26\,\mathrm{g\cdot cm^{-3}}$. This shall not be investigated in further detail as the absolute size of deviation in question is unlikely to be clinically significant.

A similar fitting procedure to that carried out by Fippel can be used to link radiation length with SPR. Figure 4.9 shows the results of a fit to ICRP Publication 110 tissues when the independent variable is SRIM-calculated SPR at 170 MeV instead of mass density. The quality of the fit to equation (4.9), as measured by the root-mean-square deviation of points from the line in the f direction, is comparable to that in the mass density fit. It is thus evident that lateral scattering considerations should not impede the use of stopping power as the dependent variable in CT calibration. Additional accuracy might be obtained by performing a separate stoichiometric calibration for



(a) All tissues **(b)** Soft tissue region

Figure 4.8 – Plot of equation (4.8), which links radiation length with mass density in the proton dose calculation algorithm of Fippel and Soukup. Also shown are points corresponding to tissue compositions recommended by the ICRP and ICRU.

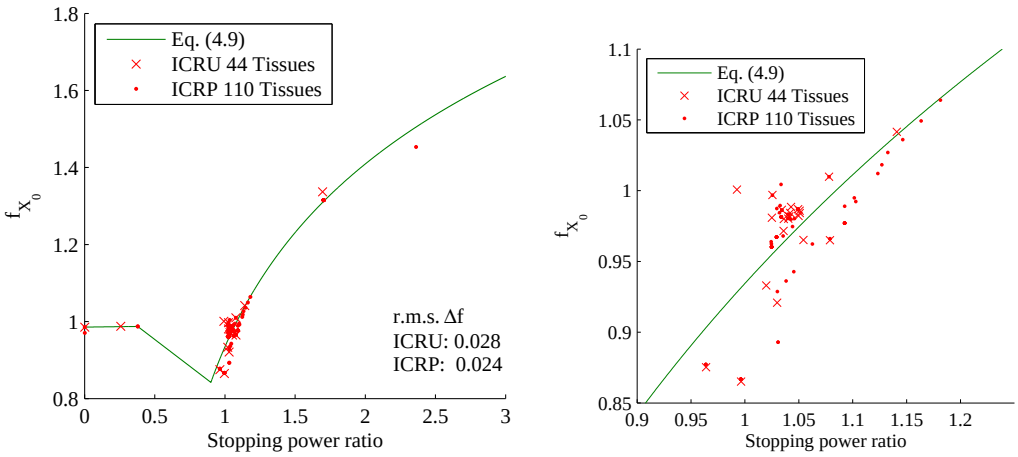
radiation length with CT number, at little extra effort since the measurements necessary to characterise the scanner will already have been made for SPR calibration.

$$f_{X_0}(SPR) = \begin{cases} 1.211 + 0.455\ln(\rho - 0.455) & SPR \geq 0.9 \\ 0.9873 & SPR = 0.382 \text{ (lung)} \\ 0.9857 & SPR = 0.0011 \text{ (air)} \\ \text{interpolate} & \text{all other } SPR < 0.9 \end{cases} \tag{4.9}$$

4.2.3 Characterising attenuation processes

In introducing the stoichiometric method [74], Schneider, Pedroni and Lomax adapted a formula of Rutherford et. al. [26] for the material-dependence of x-ray linear attenuation coefficients – the equation used to calibrate the scanner in Step 2 of Section 4.2.1. It is given below as equation (4.10) for a single element labelled by i, where Z is the atomic number, A is the relative atomic mass, N_A is Avogadro's constant, ρ is the mass density and K_{ph}, K_{coh}, K_{KN} are energy-dependent coefficients described as representing the strength of the photoelectric effect, coherent scattering and incoherent scattering respectively.

$$\begin{aligned} \mu_i &= \rho N_{g,i}(K_{ph}Z_i^{3.62} + K_{coh}Z_i^{1.86} + K_{KN}) \\ N_{g,i} &= N_A\frac{Z_i}{A_i} \end{aligned} \tag{4.10}$$



(a) All tissues **(b)** Soft tissues

Figure 4.9 – Plot of equation (4.9), which links radiation length with proton SPR. Also shown are points corresponding to tissue compositions recommended by the ICRP and ICRU. The fit is of comparable quality to that in figure 4.8, suggesting radiation length considerations should not be an impediment to the use of SPR as the calibration variable in proton radiotherapy treatment planning.

In the independent atom approximation, this is easily extended to a composite material using the mixture rule (equation (3.1)), where $N_g = \sum_i w_i N_{g,i}$, as equation (4.11). The convention has been to incorporate the sum of the powers of Z into 'effective' atomic numbers with respect to the photoelectric effect ($\tilde{Z}$) and coherent scattering ($\hat{Z}$)[19].

$$\mu = \rho N_g (K_{ph} \tilde{Z}^{3.62} + K_{coh} \hat{Z}^{1.86} + K_{KN}) \tag{4.11}$$

The origins of these equations should be examined carefully before opting to apply them in stoichiometric calibration. When doing so, it is more useful to consider the elemental linear attenuation coefficient as the product of an atomic cross section and the volumetric density of atoms. The three physical processes included in equation (4.12) are the only significant contributors to x-ray attenuation below the pair-production threshold of 1.022 MeV. Jackson and Hawkes state unequivocally that only with the simplest theory are expressions for any of the three cross sections separable to a function of energy and an exponentiation of Z [87]. Nevertheless, the total cross section formula of Rutherford et. al. is a sum in three powers of Z: representing the photoelectric effect ($\sigma_{ph} \propto Z^m$), idealised incoherent scattering ($Z \cdot \sigma_{KN}$, where σ_{KN} is the Klein-Nishina cross section given in equation (4.14)), and a correction term

[19]These effective atomic numbers are sometimes imaged in multiple-energy CT studies.

which both accounts for coherent scattering and adjusts for binding energy effects in incoherent scattering ($\delta \propto Z^n$). The two unknown exponents, m and n, were each determined through a logarithmic fit to Veigele's tables of individual cross sections [88], the scattering portions of which were calculated from theory and the photoelectric portions of which were derived from measurements[20]. Rutherford et. al. performed their fit in "the region of energy and atomic number of interest": whilst not explicitly stated in their paper, it may be inferred from the Figures therein that the best-fit exponents pertain to eight biologically important elements[21] with $Z > 6$ at the energy of 60 keV.

$$\mu_i = \frac{\rho N_A}{A}\left[\sigma_{coh,i}(Z,E) + \sigma_{incoh,i}(Z,E) + \sigma_{ph,i}(Z,E)\right] \tag{4.12}$$

$$= \frac{\rho N_A}{A}\left[Z \cdot \sigma_{KN}(E) + \delta_i(Z,E) + \sigma_{ph,i}(Z,E)\right] \tag{4.13}$$

$$\sigma_{KN} = \frac{e^4}{8\pi^2\epsilon_0^2 m_e^2 c^4}\left\{\frac{1+\varepsilon}{\varepsilon}\left[\frac{2+2\varepsilon}{1+2\varepsilon} - \frac{\ln(1+2\varepsilon)}{\varepsilon}\right] + \frac{\ln(1+2\varepsilon)}{2\varepsilon} - \frac{1+3\varepsilon}{(1+2\varepsilon^2)}\right\},$$
$$\text{where } \varepsilon = \frac{E}{m_e c^2} \tag{4.14}$$

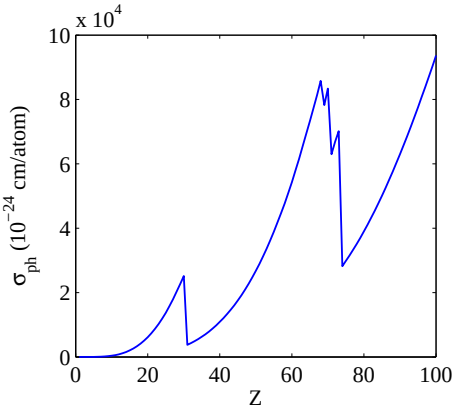
Many further tabulations of x-ray attenuation cross sections have been published since the 1970s, incorporating a greater number of measurements and calculations according to more modern theory. NIST have collated coherent scattering, incoherent scattering, photoelectric and pair-production cross-sections for elements $Z \leq 100$ and x-ray energies 1 keV to 100 GeV in a database called XCOM [27]. It is prudent to re-examine Rutherford's equations with reference to this contemporary dataset. Therefore cross sections for each element $Z \leq 100$ were taken from XCOM at 1 keV intervals in the region 10-150 keV, covering the complete typical energy range in clinical kilovoltage CT[22]. Equation 4.14 was evaluated for the same situations, and subtracted from the sum of the coherent and incoherent scattering cross sections to arrive at correction terms δ.

Figure 4.10 shows the photoelectric cross section and scattering correction as a function of atomic number. The latter is monotonic (a necessary condition for

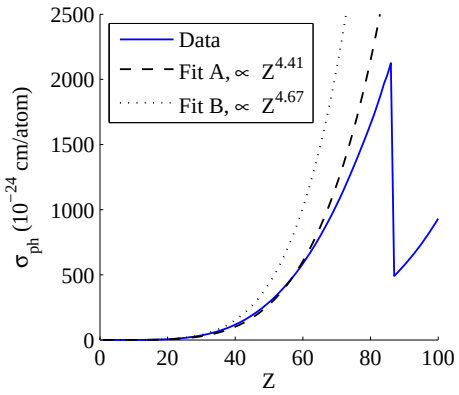
[20]In the energy region below 10 keV, which has very limited relevance to CT, the photoelectric cross sections are also from calculation.

[21]C, N, O, Na, P, Cl, K and Ca

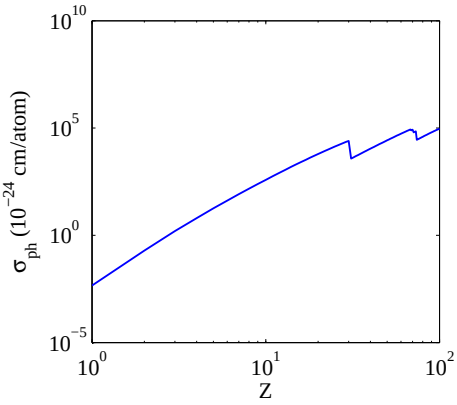
[22]The XCOM frontend provides results for arbitrary energies, which are obtained by (non-linear) interpolation of its base data.



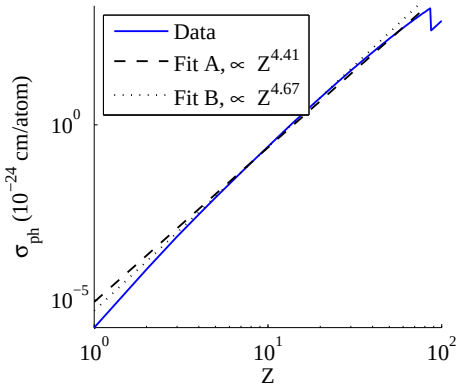
(a) photoelectric cross section, 10 keV, linear scale



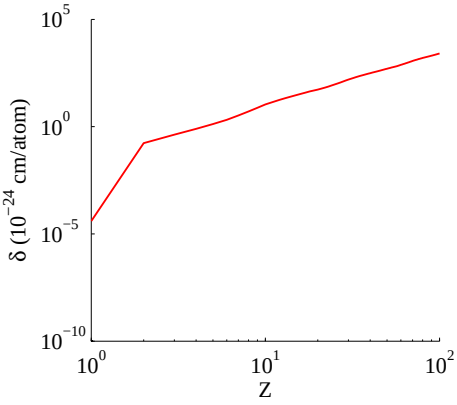
(b) photoelectric cross section, 100 keV, linear scale



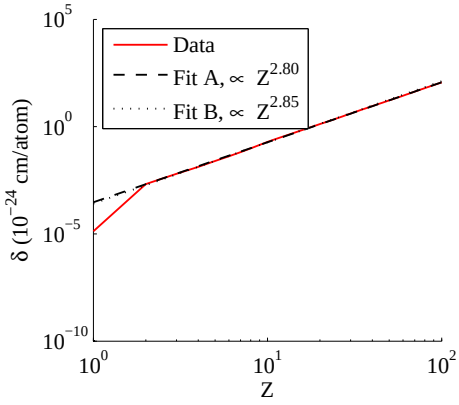
(c) photoelectric cross section, 10 keV, log-log scale



(d) photoelectric cross section, 100 keV, log-log scale



(e) scattering correction, 10 keV, log-log scale



(f) scattering correction, 100 keV, log-log scale

Figure 4.10 – Photoelectric cross sections and Klein-Nishina scattering corrections from the XCOM data for elements with atomic numbers 1 to 100 at x-ray energies of 10 and 100 keV. Black lines show power law fits at 100 keV: Fit A uses all elements $2 < Z < 86$ and Fit B uses data for the eleven elements present in the ICRP Publication 110 phantoms with $2 < Z < 30$ (C, N, O, Na, Mg, P, S, Cl, K, Ca, Fe).

exponentiation to be a reasonable model) in Z for all elements considered, but the former only varies monotonically while the energy in question is above the element's K absorption edge. As x-ray energy increases, so does the maximum Z of elements whose K-edge is exceeded, therefore the range of elements whose absorption can be characterised well using a formula like equation (4.11) will vary with energy: at 10 keV, a power in Z is only suitable for representing the photoelectric cross sections when $Z < 30$, whilst at 100 keV it can be expected to be reasonable for $Z < 86$[23]. It is more difficult to give precise thresholds in the polyenergetic situation of clinical CT, unless beam hardening correction is used and the energy of reconstruction is known, but caution should clearly be exercised when characterising the scanner using, or predicting the attenuation of, materials containing significant proportions of heavy elements.

The two exponents m and n, encompassing the Z-dependence of the photoelectric cross section and scattering correction respectively, were derived by linear regression analysis of the logarithm of the data against $\log(Z)$. As can be seen in the log-log plots of figure 4.10, the fits are somewhat sensitive to the elements included: the photoelectric cross-section does not exactly obey a power law below the K-edge, and hydrogen deviates significantly from the general trend for the scattering correction. The best-fit lines at 100 keV show how the fit differs when considering all elements above hydrogen with a K-edge greater than 100 keV (Fit A), compared to a selection of elements (excluding hydrogen) that constitute the majority of mass in the human body (Fit B). The exclusion of hydrogen should not affect results unduly as, whilst hydrogen accounts for ˜10% of the body's mass (as represented by the ICRP Publication 110 phantoms), its contribution to x-ray attenuation is very much less than that of heavier elements. The best-fit exponents for both processes are plotted in figure 4.11 as a function of energy. In order to verify the method's equivalence to that of Rutherford, it was benchmarked using Veigele's cross sections at 60 keV as an input. The Fit B results (which include Rutherford's elements, and five more, and are most applicable to the imaging of biological tissue) for Veigele's data were $m = 2.82 \pm 0.03, n = 4.66 \pm 0.03$, comparing favourably with Rutherford's exponents of 2.86 and 4.62. At the same energy, with XCOM input data, the Fit B results are $m = 2.84 \pm 0.03, n = 4.58 \pm 0.03$.

As each x-ray attenuation process varies differently with energy, the effective energy of a CT scanner's polyenergetic beam in an unknown object for each process can also

[23]These thresholds are for monotonicity. As they are approached, the true value of $\frac{\Delta\sigma_{ph}}{\Delta Z}$ increases less steeply than is predicted by an exponentiation of Z, so it it inadvisable to apply the same power law to all elements whose K-edge has been exceeded.

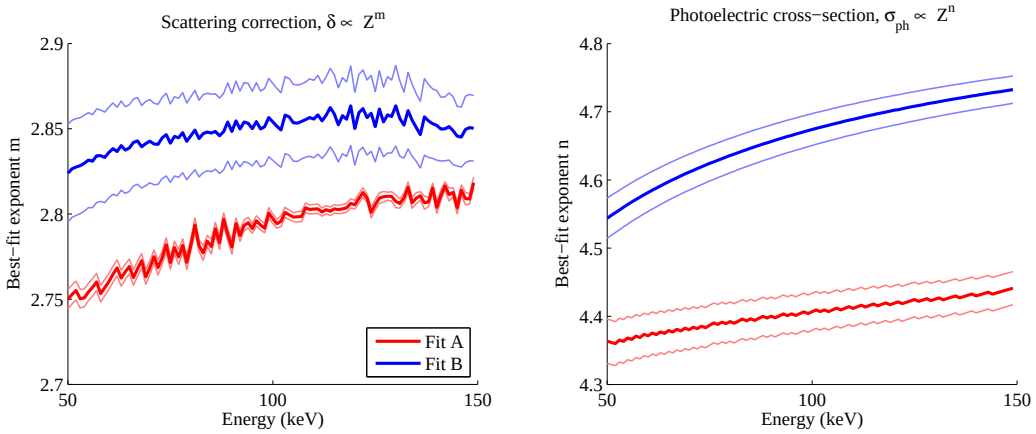


Figure 4.11 – Best-fit exponents of Z for the scattering correction (δ in equation (4.13), left) and the photoelectric cross section (σ_{ph}, right) for XCOM data in the energy range $50 - 150\,\mathrm{keV}$. Powers determined by a log-log least squares fit, for all elements with $Z > 2$ that are below their K-edge, and for the eleven elements present in the ICRP Publication 110 phantoms with $2 < Z < 30$ (C, N, O, Na, Mg, P, S, Cl, K, Ca, Fe). Thin lines show $\pm 1\sigma$ error bounds.

be expected to differ. It is thus impossible to determine the ideal energy for which the exponent should be chosen. In the case of the software CT scanner described in chapter 3, the reference energies for beam hardening correction lie in the the range 53-93 keV for tube voltages of 80-140 $\mathrm{kV_p}$. There is only a mild departure from 60 keV best-fit values of m and n for this energy range and good agreement was obtained with Rutherford's values, so following sections will not break from convention and will use $m = 2.86$ and $n = 4.62$. It should, however, be reiterated that these powers have been determined to apply for the subset of biologically-abundant elements and that they may be less appropriate in the presence of high-Z elements.

4.2.3.1 Application to Hounsfield Units

Equation 4.11 has been adapted to Hounsfield Units through two different approaches. Both ignore the very minor contribution of the air LAC to CT number. Uwe Schneider et. al. implicitly fold the LAC of water into the three existing spectrum-dependent coefficients, resulting in a formula that closely resembles the original equation, such as in equation (4.15). Wilfried Schneider et. al. also use a sum over the elemental composition for the LAC of water in addition to the calibration material, which reduces the number of coefficients to two when the ratio is taken, as in equation (4.16).

$$h_U + 1000 = \frac{\rho}{1.00\,\mathrm{g \cdot cm^{-3}}} \sum_i \frac{w_i}{A_i} \left[k_1 \cdot Z_i + k_2 \cdot Z_i^{2.86} + k_3 \cdot Z_i^{4.62} \right] \tag{4.15}$$

$$h_W + 1000 \;=\; \frac{1000 \cdot \rho}{1.00\,\mathrm{g \cdot cm^{-3}}} \frac{\sum\limits_i \frac{w_i}{A_i}\left[Z_i + k_A \cdot Z_i^{2.86} + k_B \cdot Z_i^{4.62}\right]}{\sum\limits_{j \text{ in water}} \frac{w_j}{A_j}\left[Z_j + k_A \cdot Z_j^{2.86} + k_B \cdot Z_j^{4.62}\right]} \tag{4.16}$$

It is often the case that increasing the degrees of freedom results in a fit that is a poorer predictor, as there will be an increased sensitivity to noise in the input data. However, that may not be the case here if the coefficients that best fit water vary significantly from those for the calibration materials (for example, if there are sufficiently large differences in elemental composition). One possible method for assessing whether h_U or h_W is the better relationship for fitting is to compare their predictions (after coefficients have been determined) to the LACs for a set of materials of known composition. Possible sets of calibration materials will be considered at the same time.

4.2.4 Choice of tissue equivalents

The sole purpose of scanning tissue equivalent materials in the stoichiometric method is as a set of calibration materials with which to characterise the strength of each x-ray attenuation process when the scanner is operated with a particular energy spectrum. It is thus not necessary for these materials to be 'equivalent' to tissue in terms of proton stopping power. However, as explored in the previous section, there is a desire for the calibration materials to have a similar elemental composition to biological tissues and to exhibit similar levels of x-ray attenuation in the CT tube's output energy range ($\sim 10 - 150\,\mathrm{keV}$) in order to ensure adequate sampling of the relevant parameter space.

Three possible sets of tissue equivalent materials are to be considered under idealised circumstances in which the incident x-rays are monoenergetic, with the aim of determining which provides the best calibration for the tissues of the ICRP Publication 110 phantoms. One set is the ten materials from the Tissue Substitute Section at Barts and the London NHS Trust, previously detailed in table 3.3. The other two sets are the six materials proposed by Uwe Schneider et. al. (table 4.3a) when developing the stoichiometric method, and a set of sixteen materials adopted in the paper of Wilfried Schneider et. al (table 4.3b).

Tables 4.4 and 4.5 compare the root-mean-square deviations between directly-calculated CT numbers [24] and those predicted by the stoichiometric method for

[24] according to the methods of chapter 3 in the ideal monoenergetic case

Material	Equivalent to	Density ($g \cdot cm^{-3}$)	Elemental composition (% by mass)										
			H	C	N	O	Cl	Ca	Na	Mg	P	S	Others
AP6	Adipose tissue	0.91	8.69	69.14	2.36	16.94	0.14	0	0	0	0	0	F: 3.07
Water	Water	1.00	11.19	0	0	88.81	0	0	0	0	0	0	
MS/SR 4	Muscle	1.07	9.5	70.25	3.48	15.15	0.12	0.01	0.08	0.02	0.18	0.50	K: 0.30
IB/SR 1	Inner bone	1.15	8.73	63.19	2.36	17.83	0.12	5.09	0.06	0	2.62	0	
TSK/SR 1	Skeleton	1.32	6.4	46.4	2.8	26.4	0.1	10.0	0.3	0.1	7.0	0.2	
HB/SR 4	Hard bone	1.48	4.45	29.09	3.88	31.93	0.06	19.99	0.06	0.21	10.0	0.32	

(a) Tissue equivalent compositions used in U. Schneider et. al. 1996 [74]

Material	Density ($g \cdot cm^{-3}$)	Elemental composition (% by mass)									
		H	C	N	O	Mg	Si	Cl	Ca	Ti	Sn
Compact bone	1.84	3.1	31.26	0.99	37.57	0	0	0.05	27.03	0	0
Muscle	1.05	8.1	67.17	2.42	19.85	0	0	0.14	2.32	0	0
Fat	0.92	8.5	72.88	2.24	16.25	0	0	0.13	0	0	0
Lung	0.30	8.36	60.43	1.67	17.33	11.36	0.72	0.13	0	0	0
Spongiosa	1.14	7.9	63.79	4.23	9.88	0	0	0	14.2	0	0
Solid water	1.035	8.09	67.22	2.40	19.84	0	0	0.13	2.32	0	0
RW-3	1.045	7.59	90.41	0	0.80	0	0	0	0	1.20	0
H-800	0.23	7.96	64.21	0	16.29	0	11.48	0	0	0	0.06
H-500	0.47	8.04	45.93	0	19.41	0	26.48	0	0	0	0.14
H+200	1.04	7.7	31.5	0	22.99	0	35.66	0	1.96	0	0.19
H+400	1.12	6.35	28.07	0	27.38	0	29.4	0	8.64	0	0.16
H+700	1.36	4.59	23.63	0	33.09	0	21.28	0	17.33	0	0.06
H+900	1.43	3.72	21.41	0	35.93	0	17.21	0	21.64	0	0.09
H+1200	1.65	2.4	18.07	0	40.21	0	11.1	0	28.16	0	0.06
PMMA	1.19	8	60	0	32	0	0	0	0	0	0
Polyethylene	0.94	14.4	85.6	0	0	0	0	0	0	0	0

(b) Tissue equivalent compositions used in W. Schneider et. al. 2000 [76]

Table 4.3

Prediction

Calib.	B	U	W	I
B	0.5	4.3	18	5.7
U	4.8	0.6	22.6	3.8
W	7.5	11.6	13.8	14.6
I	3.6	1.4	21.6	3.7

(a) 50 keV

Prediction

Calib.	B	U	W	I
B	0.5	1.3	5.4	1.6
U	1.5	0.3	6.7	1.1
W	2.4	3.5	4.3	4.2
I	1.1	0.4	6.5	1.2

(b) 75 keV

Prediction

Calib.	B	U	W	I
B	0.4	0.6	2.4	0.7
U	0.7	0.4	2.7	0.5
W	1.1	1.4	1.6	1.8
I	0.6	0.4	2.6	0.5

(c) 100 keV

Prediction

Calib.	B	U	W	I
B	0.3	0.4	1.3	0.4
U	0.4	0.2	1.5	0.4
W	0.5	0.6	1	1.1
I	0.4	0.2	1.6	0.4

(d) 125 keV

Prediction

Calib.	B	U	W	I
B	0.3	0.3	0.8	0.4
U	0.3	0.2	0.8	0.3
W	0.4	0.4	0.8	0.5
I	0.3	0.2	0.9	0.3

(e) 150 keV

Table 4.4 – Root-mean-square deviations in predicted vs. calculated CT numbers (in HU) using the stoichiometric method with the 2-parameter attenuation model h_W given in equation (4.16). Four sets of materials have been used to determine best-fit coefficients k_A and k_B (calibration) and for assessing the accuracy of the model (prediction) at five energies. Material abbreviations are **B** Barts Water (table 3.3, $n = 10$), **U** U. Schneider's paper (table 4.3a, $n = 6$), **W** W. Schneider's paper (table 4.3b, $n = 16$), **I** ICRP Publication 110 organs ($n = 141$).

Prediction

Calib.	B	U	W	I
B	0.4	4.3	18	5.5
U	4.8	0.6	22.6	3.8
W	13.2	21.3	10.6	29.9
I	3.2	1.9	21.4	3.7

(a) 50 keV

Prediction

Calib.	B	U	W	I
B	0.5	1	5.4	1.6
U	1.5	0.3	6.7	1.1
W	4.1	6.6	3.1	9.1
I	1.1	0.4	6.5	1.1

(b) 75 keV

Prediction

Calib.	B	U	W	I
B	0.4	0.6	2.4	0.7
U	0.7	0.4	2.7	0.5
W	1.7	2.6	1.3	3.6
I	0.5	0.4	2.6	0.5

(c) 100 keV

Prediction

Calib.	B	U	W	I
B	0.3	0.4	1.3	0.4
U	0.4	0.2	1.6	0.4
W	1	1.4	0.8	2
I	0.4	0.3	1.6	0.4

(d) 125 keV

Prediction

Calib.	B	U	W	I
B	0.3	0.3	0.8	0.4
U	0.3	0.2	0.8	0.4
W	0.7	0.7	0.6	1.4
I	0.3	0.2	0.8	0.3

(e) 150 keV

Table 4.5 – Root-mean-square deviations in predicted vs. calculated CT numbers (in HU) using the stoichiometric method with the 3-parameter attenuation model h_U given in equation (4.15). Four sets of materials have been used to determine best-fit coefficients k_1, k_2 and k_3 (calibration) and for assessing the accuracy of the model (prediction) at five energies. Material abbreviations are **B** Barts Water (table 3.3, $n = 11$), **U** U. Schneider's paper (table 4.3a, $n = 6$), **W** W. Schneider's paper (table 4.3b, $n = 16$), **I** ICRP Publication 110 organs ($n = 141$).

models with two (h_W) and three (h_U) parameters respectively. The models have been calibrated and tested with every combination of the three material sets described in the previous paragraph, as well as the tissues of the ICRP 110 phantoms. Best-fit parameters were determined by least-squares methods: for h_U by multiple linear regression in MATLAB using the `regress` function, and for h_W by multiple non-linear regression with MATLAB's `nlinfit` function[25].

As should be expected, the highest quality predictions of CT numbers for each set of calibration materials occur when the model is tested with that same set of materials. Of greatest interest for clinical radiotherapy is how well each set of materials predicts the CT numbers of ICRP tissues, which are taken to be representative of the human body, and so close attention should be paid to the rightmost column of the Tables. Of the three tissue equivalent families, the best calibration is offered by the set denoted U and the worst by the set denoted W, for both models at all five energies[26]. This can be explained by reference to the composition of the materials in question: sets B and U contain only elements that were used in deriving the exponents of Z in the attenuation equation (as Fit B of figure 4.11); set W, on the other hand, additionally contains a significant proportion of silicon and small amounts of titanium and tin. Silicon has an atomic number of 14 – by this point on the Z axis, the photoelectric cross section at 100 keV according to Fit B is underestimated by ˜3% compared to the XCOM data, thereby affecting the minimum deviation attainable in a least-squares fit. The magnitude of deviation observed with set W at lower energies should be of some concern, demonstrating the potential to introduce errors of up to 3% of water's attenuation into some regions of the calibration curve even before polychromaticity and other artefacts have been considered. It is apparent that predictions are universally better above 100 keV, despite the fact that the exponents in the attenuation power laws were determined at 60 keV, which is justified by the fact that the photoelectric cross-section is well-approximated by a power law until a greater Z value as energy increases.

In general the expressions with two and three parameters offer comparable deviations between expected and observed values when calibrated with sets B and U, but not when set W is used. In the case of set W, the root-mean-square deviations are reduced by approximately 50% when using the two-parameter model (h_W) compared

[25]which uses the Levenberg-Marquardt algorithm, and was initialised with W. Schneider's values of $k_A = 0.0124$ and $k_B = 3.06 \times 10^{-5}$

[26]There is one case (125 keV with three parameters) where set B performs marginally better than set U.

to the three-parameter model (h_U). The reduction can be attributed to the extra information incorporated in equation (4.16) (in the form of the composition of water), providing fewer degrees of freedom, leading to smaller confidence intervals for its parameters after fitting. In contrast, fitting the less-constrained three-parameter model of equation (4.15) results in relatively larger confidence intervals for parameters that reflect the physical reality less well. In many situations where set W is used for calibration, the best-fit k_2 value is in fact negative, and systematic offsets are observed in predicted HU values for all ICRP organs (˜20 HU at 50 keV and ˜3 HU at 100 keV).

The CT numbers of most ICRP organs are predicted to within ±2 HU for both formulae at all energies considered if the calibration is carried out with sets B and U. An exception occurs when calibrating with set B and predicting in the teeth or cortical bone, which is underestimated by 11 HU at 50 keV but well-predicted at 100 keV. It is not readily apparent why this offset occurs with set B but not with set U. There is one other outlier, which is the thyroid. The thyroid is a special case as it is the only organ in the ICRP AM and AF phantoms having any constituents with $Z > 26$, containing 0.1% iodine ($Z = 53$). As the exponents are optimised for much lower atomic numbers (by the Fit B approach), the photoelectric cross section for iodine will be significantly overestimated compared to reality, as is evident in figure 4.10. At 50 keV, calibration of either model with both sets B and U leads to overestimation of attenuation in excess of 40 HU, and by 5 HU at 100 keV.

From these results, it is clear that stoichiometric calibration can introduce minimal variation when predicting x-ray attenuation of realistic human tissues in the idealised monoenergetic case. It is equally clear that an appropriate set of materials must be used in the calibration process, and that those detailed in table 4.3a may be suitable. However, if it were possible to produce monoenergetic CT scans with a known energy, it would not be necessary to perform stoichiometric calibration as the expected CT numbers would be directly-calculable. Real scanners, of course, emit x-rays with a spectrum of energies and the resulting images exhibit beam hardening artefacts. The following Section will discuss some of the issues involved by performing stoichiometric calibration of the software CT scanner presented in chapter 3.

4.2.5 Calibration scans

Beam hardening and other imaging artefacts present a complication in the stoichiometric calibration of CT scanners. As discussed earlier, the default post-processing procedures in CT typically only include first-pass (water-based) beam hardening correction, so Type S artefacts, arising from differences in material composition within

the scan volume, may be present. Thus, under real conditions, the CT number for a particular material is liable to vary based on its position and surroundings. In order to obtain stable reference measurements, many authors recommend measuring the CT number of each calibration material individually by performing separate scans of samples located at the centre of a cylindrical water tank. Nevertheless, without the complete correction of beam hardening artefacts in both calibration and measurement, the pixel values observed may vary compared to those seen if the sample were located within a real patient[27]. A number of physical phantoms also exist whose express purpose is to provide a reference object for the quality assurance and calibration of CT scanners. It should be noted that such phantoms are typically constructed with the aim of producing calibration curves in photon radiotherapy, rather than for stoichiometric calibration in the case of proton therapy. In this section, stoichiometric calibration will be performed for the software CT scanner using scans of individual materials and a multi-material phantom, in order to determine the impact that beam hardening artefacts may have on the results.

The materials used for calibration are those in table 4.3a, which have been shown to enable good characterisation of monoenergetic x-ray attenuation processes in the previous Section. Parameters for the software CT scanner were those determined in the benchmarking procedure of Section 3.4, except with a 20% increase in all linear dimensions[28]. The effects of varying tube voltage will be investigated through simulation at $80\,\mathrm{kV_p}$, $100\,\mathrm{kV_p}$, $120\,\mathrm{kV_p}$ and $140\,\mathrm{kV_p}$, and the impact of second-pass beam hardening correction (off, and with 1 and 2 iterations) and stochastic noise will be considered. Five arrangements of the calibration materials will be compared. Illustrated in figure 4.12 are four of the arrangements, which use a multi-material virtual phantom based on the geometry of the Gammex RMI Tissue Characterization Phantom (Model 467) [89]. The multi-material phantom consists of a 33 cm diameter cylinder of water equivalent plastic (WT1 is used here) with sixteen 2.8 cm diameter cylindrical spaces for calibration materials, which are arranged in two concentric rings at distances of 5.5 cm and 10.5 cm from the phantom's axis. The fifth arrangement

[27]for example, if the typical water equivalent depth of a tissue in vivo is different to the radius of the water phantom

[28]This increase was necessary in order for the broadest parts of the anthropomorphic ICRP phantoms to fit within the reconstruction volume, for which the calibrations described here will be applied later. The change is not expected to affect the relevance of any conclusions, given that the size of clinical scanners varies, and mean CT numbers for blocks in the benchmarking scans of 3.4 differed by no more than $\pm 1.5\,\mathrm{HU}$ when it was implemented.

involves scanning each material as a single 3 cm cylindrical sample at the centre of a 30 cm cylindrical water phantom, as shown in figure 4.13.

In order to cover all combinations of scan properties and tissue equivalent arrangements, 240 separate simulations were performed, the results of which will not be presented individually. CT numbers for each material were sampled by taking the mean[29] value of pixels in the image whose centre locations are within that material, as defined by the phantom geometry, but excluding a two pixel border to account for 'bleeding' at boundaries. The measured CT numbers were used to determine parameters for (i.e. calibrate) the 2- and 3-parameter models h_W and h_U following the methods of previous Section.

The impact of each parameter on the quality of calibration was investigated using two metrics.

- The self-consistency of the resulting model as measured by the r.m.s. deviation between input and predicted CT number, given as Δ_C in equation (4.17), where i labels the calibration materials.

$$\Delta_C = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(h_{\mathrm{meas},i} - h_{\mathrm{model},i})^2} \tag{4.17}$$

- The accuracy of the model in predicting in-situ CT numbers for organs of the ICRP AM phantom under the same scan conditions. In order to assess this, simulated scans had to be performed for every slice of the phantom with the relevant parameter sets – the resulting images will be examined in detail in the next Chapter. Measurements were taken by calculating the mean of voxels located wholly[30] within each organ, which were compared with the predictions of the calibrated model for the elemental composition of the organ in question. The prediction accuracy is reported by means of the volume-weighted[31] r.m.s. deviation between model and measurement, given as Δ_P in equation (4.18), where j labels the organs of the ICRP AM phantom, V_j is the volume of organ j, and V is the total volume of all the voxels considered.

[29]The mean was chosen because the multi-material phantom has two equally-sized samples of each material, which may lead to a bimodal distribution for which the median is a poor sample estimator.

[30]The pixel grid for reconstruction is different from that upon which the phantom is defined, causing some image pixels to straddle the boundaries of phantom pixels.

[31]Many organs of the ICRP phantoms have only a small number of voxels therefore their mean CT number may be strongly influenced by beam hardening or other artefacts. Volume-weighting the r.m.s. deviation ensures that small organs do not unduly affect the metric, so it is more representative of the expected deviation of an individual CT voxel.

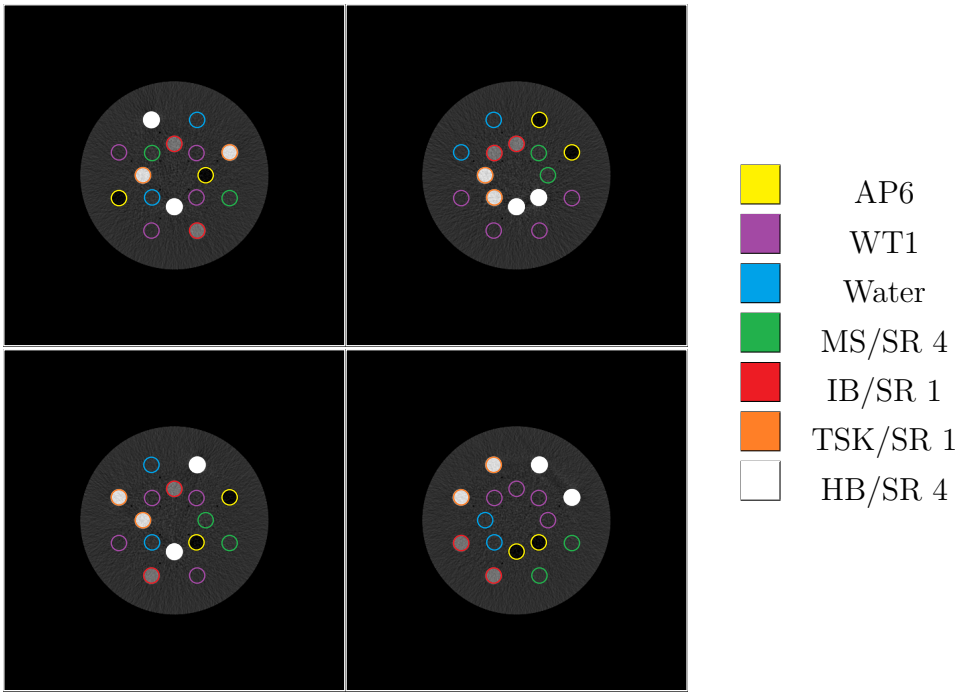


Figure 4.12 – Simulated images of the multi-material virtual phantom used in simulations of stoichiometric calibration. Coloured outlines mark the tissue equivalent materials, whose composition can be found in table 4.3a. Arrangements are numbered 1 to 4 starting in the top left and continuing in a clockwise direction. Simulated scan parameters: 80 kV_p, stochastic noise equivalent to 150 mAs exposure, no second-pass beam hardening correction, 60 cm field of view. Level: 300 HU, Window: 1000 HU.

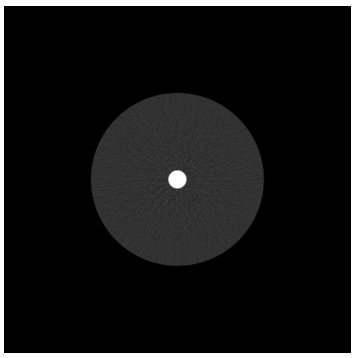


Figure 4.13 – Example of the single material virtual phantom used in simulations of stoichiometric calibration. Scan and window parameters as figure 4.12. The material in this image is HB/SR 4.

$$\Delta_P = \sqrt{\frac{1}{V}\sum_j V_j \cdot (h_{\mathrm{meas},j} - h_{\mathrm{model},j})^2} \tag{4.18}$$

The values of Δ_C found ranged from 0.75 HU to 9.13 HU with a mean of 4.06 HU, whilst values of Δ_P ranged from 6.19 HU to 22.91 HU with a mean of 12.6 HU.

4.2.5.1 2 vs. 3 parameter model

In Section 4.2.4, the 2- and 3- parameter models were found to display similar performance when calibration was performed in ideal monoenergetic conditions using the materials considered here. The histograms in figure 4.14 compare their suitability when calibrated with CT numbers from the more realistic, polyenergetic simulated scans. The size of the differences in Δ_C and Δ_P between models are relatively small, being less than 10% of the larger value in most cases. The largest difference in prediction accuracy corresponds to a volume-weighted mass density difference of no more than $0.002\,\mathrm{g}\cdot\mathrm{cm}^{-3}$, meaning the choice of model is unlikely to have clinical significance for these calibration materials and scan properties. Nevertheless, for both metrics, it can be seen that the 3-parameter model is favoured in a greater number of cases. However, the difference in Δ_C between models for a certain set of inputs is not a particularly good predictor of the difference in Δ_P (Spearman's r= 0.33, which is nevertheless statistically significant at a level $> 99.9\%$). In all cases the best-fit parameters were positive, in keeping with their physical interpretations in the monoenergetic case.

As the 3-parameter model h_U displayed the best self-consistency and prediction accuracy more frequently, it was adopted for subsequent comparisons.

4.2.5.2 Stochastic noise

The potential impact of stochastic noise (as photon starvation artefacts) on CT images was discussed in Section 2.6.1.2. In addition to intensity variations on the individual pixel level, high attenuation materials have the potential to introduce significant streaking outside their boundaries. If the x-ray exposure is limited, such streaking could be sufficiently intense to affect the reliability of calibration measurements. Scans at lower $\mathrm{kV_p}$ are the most susceptible to photon starvation due to the increase in attenuation at low x-ray energies. The simulated images shown in figure 4.12 are thus those with the greatest level of stochastic noise, and show no obvious starvation streaking, indicating it is not a problem for the x-ray exposure of 150 mAs considered.

Figure 4.15 shows quantitatively the effect of noise on the values of Δ_C and Δ_P. Again the differences in question are relatively small compared to the absolute size of

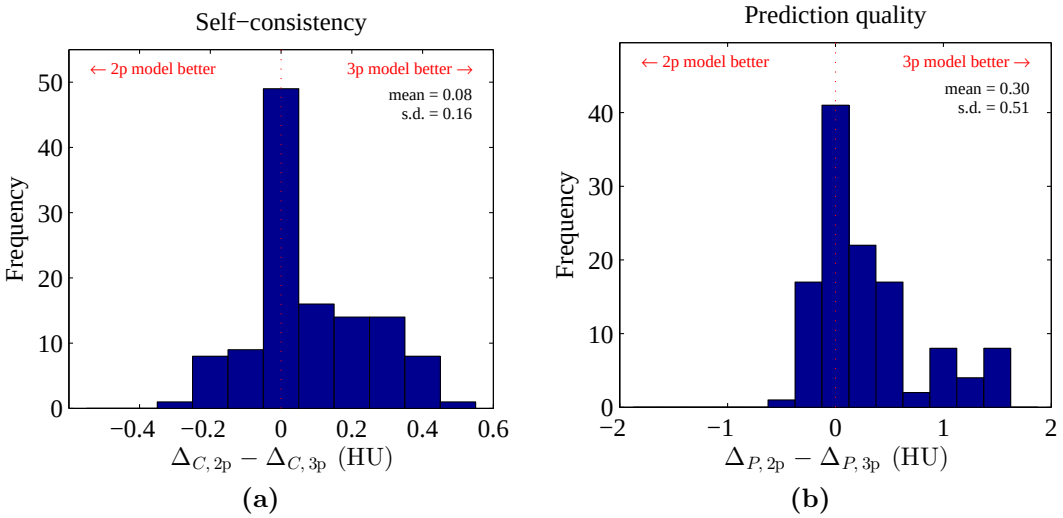


Figure 4.14 – Histograms showing differences in performance between the 2- and 3-parameter models for each set of calibration data, as assessed by the self-consistency and prediction accuracy metrics.

the metrics. No systematic shift is evident in the self-consistency of the calibration. The histogram of Δ_P shows two sub-groups – one for which stochastic noise has no effect on average, and one for which stochastic noise markedly degrades the quality of predictions. The second sub-group comprises all five arrangements at 80 $\mathrm{kV_p}$ with second-pass beam hardening correction activated. As this split is only seen for Δ_P, the reason likely lies with the simulated scans of the ICRP phantom rather than those of the blocks. The HU variation introduced by stochastic noise is at its most intense at 80 $\mathrm{kV_p}$ – a small number of voxels in the phantom are likely therefore misidentified as bone by the BBHC algorithm, causing their HU number to be unrepresentative of the underlying material after correction and therefore leading to an increased Δ_P.

4.2.5.3 Second-pass (bone) beam hardening correction

Whilst Type C artefacts are relatively well corrected for by the first-pass (water-based) beam hardening correction, Type S artefacts are regularly apparent in clinical images. The simulated images in figure 4.12 are displayed with a wide window, yet some streaks are visible between the sample of HB/SR 4 in arrangements 2 and 3. More subtle artefacts are also present, especially within the bone samples whose attenuation tends to be overestimated before second-pass beam hardening correction (BBHC). The histograms of 4.16 show that the magnitude of improvement in stoichiometric calibration accuracy when using BBHC approaches significant levels. A mean decrease

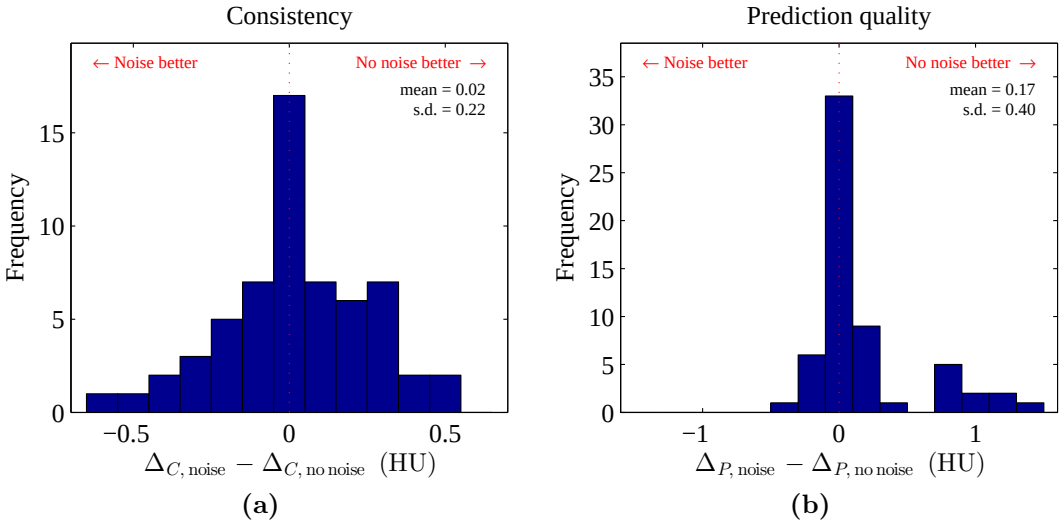


Figure 4.15 – Histograms showing the differences in the self-consistency and prediction accuracy metrics between sets of calibration data with stochastic noise enabled and disabled. The noise in question is equivalent to that in a 150 mAs exposure.

in Δ_P of 9 HU is observed when using two iterations of BBHC instead of none, which corresponds to roughly 1% of the density of water, and would map to a similar reduction in proton range uncertainty. A smaller decrease is also seen in Δ_C. The improvements can be explained by the fact that the mean LACs in both the calibration and ICRP phantom scans are closer to their values in a monoenergetic scan (the situation for which the model was derived). It is evident that using BBHC in stoichiometric calibration (and therefore necessarily also in planning scans) may have benefits that are of some interest in the clinic. The implications for particular anatomical locations will be investigated in the next Chapter.

4.2.5.4 Arrangement of tissue equivalents and energy spectrum

The self-consistency and prediction accuracy of models calibrated using each arrangement of materials are plotted in figure 4.17 for all four energy spectra, with and without BBHC. A number of trends are apparent. Firstly, regarding energy, higher $\mathrm{kV_p}$ values result in calibrations that perform better with respect to both metrics when BBHC is disabled. An explanation is that the variations in CT numbers due to beam hardening (and photon statistics) are less profound at higher tube voltages because a larger proportion of photon energies are above the sharp drop in attenuation seen in figure 2.11, so the calibration and phantom images appear closer to monoenergetic scans. The same trend is seen in Δ_P if two iterations of BBHC are performed, but

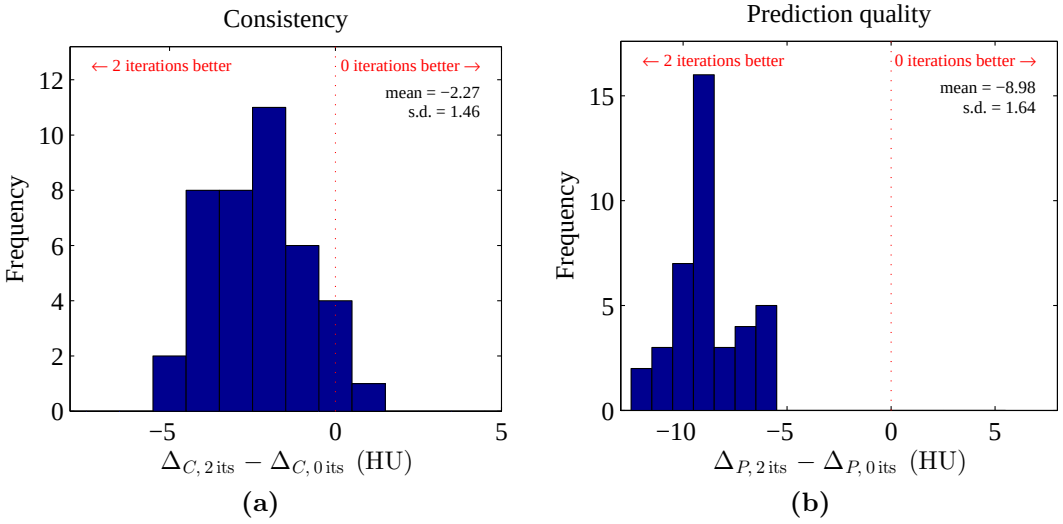


Figure 4.16 – Histograms showing the differences in the self-consistency and prediction accuracy metrics between sets of calibration data with 2 iterations of bone-based beam hardening correction and those with BBHC deactivated.

$\mathrm{kV_p}$	E_{ref} (keV)	$\Delta_{C,\,\mathrm{ref}}$ (HU)
80	53.3	0.52
100	66.6	0.29
120	80	0.32
140	93.3	0.31

Table 4.6 – Δ_C for models calibrated with ideal monoenergetic CT numbers at the beam hardening correction reference energies. Calculations by identical means to U-U cells in table 4.5.

not in Δ_C (for which there is no simple pattern in the monoenergetic values, given in table 4.6, likely as a result of the small number of materials).

When no second-pass beam hardening is performed, the arrangement of tissue equivalent materials affects the self-consistency of the calibration much more than the quality of its predictions in the ICRP phantoms. However, it is clear that performing individual scans of a single-material phantom produces marginally better predictions than the multi-material phantom, with which Arrangement 3 performs somewhat worse than other layouts. The data in table 4.7 shows ±2% variation in the CT number of HB/SR 4 at these two extremes. It is suggested that the average water equivalent depth in the ICRP phantom is closer to that of samples in the single-material phantom (where it is ˜13.5 cm) than the multi-material phantom (where the observed CT numbers imply it is greater). Beam hardening artefacts are particularly prominent between adjacent bone samples located in the outer ring of Arrangement 3, as seen

Material	Arrangement				
	1	**2**	**3**	**4**	**Indiv.**
AP 6	-162.5	-162.3	-161.9	-160.2	-162.8
Water	2.0	-1.3	-2.1	0.2	0.6
MS/SR 4	7.3	5.4	11.9	6.7	6.4
IB/SR 1	249.6	252.4	251.3	250.9	258.7
TSK/SR 1	659.6	663.1	651.2	656.4	678.7
HB/SR 4	1251.8	1254.5	1232.9	1254.8	1280.2

Table 4.7 – Mean CT numbers in HU for the calibration scans at 80 $\mathrm{kV_p}$, with stochastic noise and no second-pass beam hardening correction.

in figure 4.12. Placing these high attenuation materials at a greater distance from the isocenter results in an energy spectrum that is harder over a larger area causing the mean CT numbers to be lower. After two iterations of BBHC, all arrangements perform within the 1 HU quantisation error with respect to both metrics.

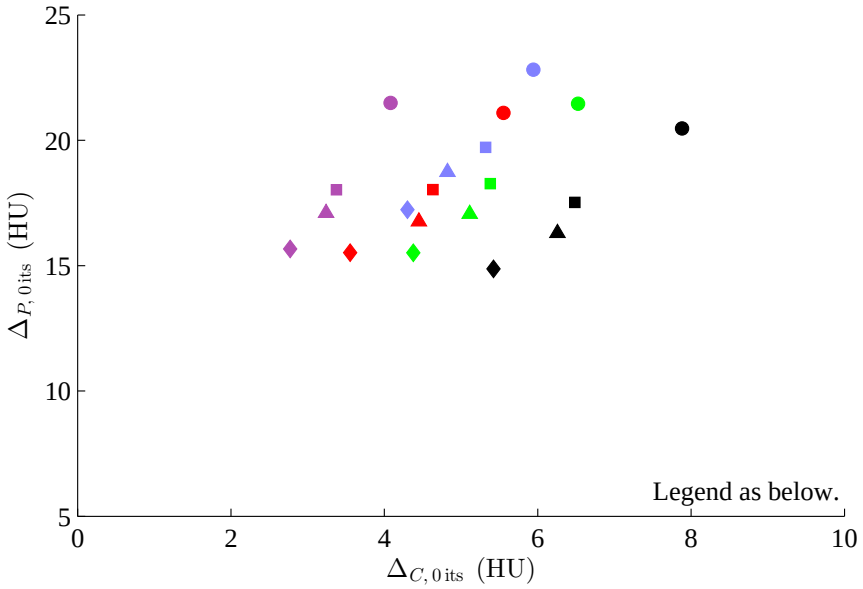
4.2.5.5 Parameters

Table 4.8 gives the best-fit values for parameters k_1, k_2 and k_3 of equation (4.15) for four situations, include the conditions that achieve the best and worse Δ_P figures. CT calibration curves will be defined for these parameter sets in the next Section.

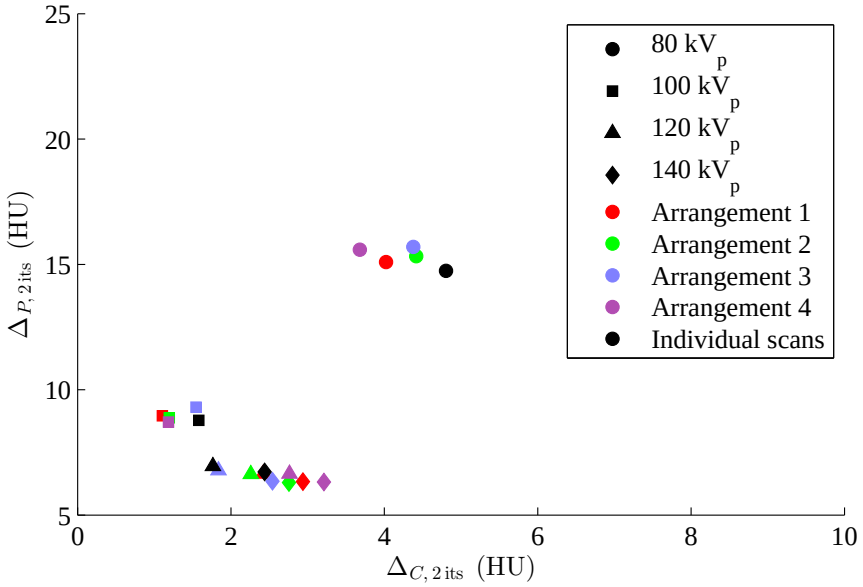
4.2.6 Producing the calibration curve

Once it is possible to predict the CT number and SPR for a material of known composition in a particular scanner, sufficient information is available to produce the calibration curve required by proton radiotherapy TPS packages. It is necessary for calibration curves to be defined by a single-valued function whose domain is CT numbers, but in practice the same CT number might be measured in materials of different composition and proton stopping powers. Some points of the calibration curve may therefore require a compromise that takes into account the number of materials that may appear at a particular CT number. In idealised conditions, the act of creating a calibration curve involves deciding how to 'join the dots' defined by the calculated SPRs and CT numbers of human tissues, as illustrated in figure 4.18. Beam hardening and stochastic noise artefacts will further confound the issue.

As with many other aspects of the stoichiometric calibration, the papers lead-authored by Uwe Schneider and Wilfried Schneider differ on their approach to curve construction. The former used three linear fits to data corresponding to 35 tissues from ICRP Publication 23 [90], in the HU regions between −1000 and −150, 23 and



(a) No second-pass beam hardening correction



(b) 2 iterations of second-pass beam hardening correction

Figure 4.17 – Self-consistency and prediction accuracy for stoichiometric calibrations performed with four different x-ray energy spectra (indicated by shape of marker) using the tissue equivalent arrangements illustrated in figure 4.12 and figure 4.13 (indicated by colour of marker). Photon statistics were considered for an exposure of 150 mAs.

Layout	kV_p	BBHC	k_1	k_2	k_3
Indiv.	80	none	1562 ± 238	2.68 ± 9.40	0.094 ± 0.049
Indiv.	100	none	1619 ± 200	2.00 ± 7.88	0.072 ± 0.041
Indiv.	120	none	1649 ± 185	1.73 ± 7.31	0.059 ± 0.038
Indiv.	140	none	1672 ± 160	1.46 ± 6.34	0.050 ± 0.033
Indiv.	80	2 its.	1549 ± 148	2.26 ± 5.85	0.116 ± 0.031
Indiv.	100	2 its.	1590 ± 47	3.00 ± 1.87	0.064 ± 0.010
Indiv.	120	2 its.	1600 ± 59	4.05 ± 2.32	0.029 ± 0.012
Indiv.	140	2 its.	1620 ± 73	3.97 ± 2.90	0.018 ± 0.015
Arr. 3	80	none	1594 ± 180	1.72 ± 7.11	0.093 ± 0.037
Arr. 3	100	none	1639 ± 160	1.49 ± 6.30	0.069 ± 0.033
Arr. 3	120	none	1670 ± 152	1.18 ± 5.99	0.057 ± 0.031
Arr. 3	140	none	1688 ± 128	1.08 ± 5.06	0.047 ± 0.026
Arr. 3	80	2 its.	1576 ± 133	1.41 ± 5.25	0.117 ± 0.027
Arr. 3	100	2 its.	1605 ± 50	2.62 ± 1.97	0.064 ± 0.010
Arr. 3	120	2 its.	1613 ± 58	3.68 ± 2.30	0.029 ± 0.012
Arr. 3	140	2 its.	1629 ± 76	3.72 ± 3.01	0.017 ± 0.016

Table 4.8 – Parameters from stoichiometric calibration of h_U with individual calibration scans and Arrangement 3 of the multi-material phantom, with 0 and 2 iterations of BBHC, including photon statistics. Errors relate to the 95% confidence interval of the multiple linear regression.

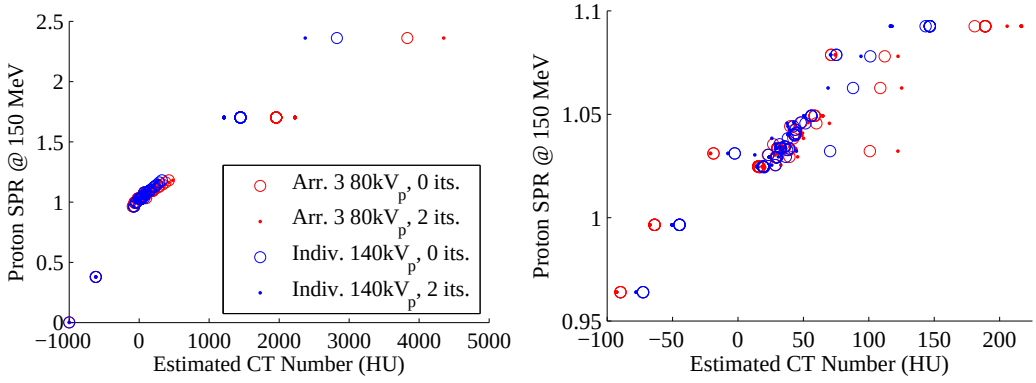


Figure 4.18 – CT numbers and SRIM-calculated stopping power ratios for the organs of the ICRP Publication 110 phantoms, as predicted by model h_U with the parameters given in table 4.8. The whole tissue range is shown on the left, with a close-up of the soft tissue region on the right.

			r.m.s. Δ_{SPR}			v.r.m.s. Δ_{SPR}		
Layout	**kV_p**	**BBHC**	**US**	**BS**	**WS**	**US**	**BS**	**WS**
Arr. 3	80	none	0.016	0.016	0.015	0.009	0.009	0.009
Arr. 3	80	2 its.	0.018	0.018	0.016	0.010	0.010	0.010
Indiv.	140	none	0.013	0.013	0.012	0.007	0.007	0.006
Indiv.	140	2 its.	0.011	0.010	0.011	0.005	0.005	0.005

Table 4.9 – Root-mean-square differences in SPR between the points and lines for figure 4.19 (US), figure 4.20 (BS) and figure 4.21 (WS). 'v.r.m.s' is the volume-weighted r.m.s. in which the contribution of each point is weighted according to the number of voxels the corresponding organ occupied in the ICRP AM phantom, using a similar formula to equation (4.18).

60, and above 60 HU, the first two regions being connected through the point for fat. The latter introduces a new interpolation method, which estimates a chemical composition for each CT number and will be described shortly. Also of note is the paper of Schaffner and Pedroni [75], in which the model h_U was used with Bethe-Bloch SPR calculations for materials in ICRU Report 44 and Woodard and White [91], with the resulting calibration curve formed from five fitted linear segments.

SRIM has been used to calculate all SPRs in this Section, regardless of the method used by the original authors of a paper.

4.2.6.1 Linear fits

Curves produced by the method of Uwe Schneider are shown in figure 4.19 for a subset of the calibrations given in table 4.8. Numerical data on the average deviation of organs in the ICRP AM phantom can be found in table 4.9. Errors of up to 2% of the stopping power of water can be seen when organs are weighted evenly in the calculation, dropping to 1% if volume-weighted. It is important to emphasize that the figures in question only account for the differences between the points and the fitted curve, and not for any deviations of the points from 'reality', which was discussed in Section 4.2.5 and will be explored further in the next Chapter.

An alternative approach involving linear fits is that of Schaffner. The domains for each fit were informed by the analysis of CT number distributions for different anatomical sites. The implementation used here works as follows. All unique body tissues in ICRU Report 44 and Woodard and White [91] are identified[32], and classified into six categories (organs and muscle, adipose tissue, breast, cartilage, bone marrow

[32]The latter was a source for some of the compositions in the former, and it was necessary to prevent duplication of tissues.

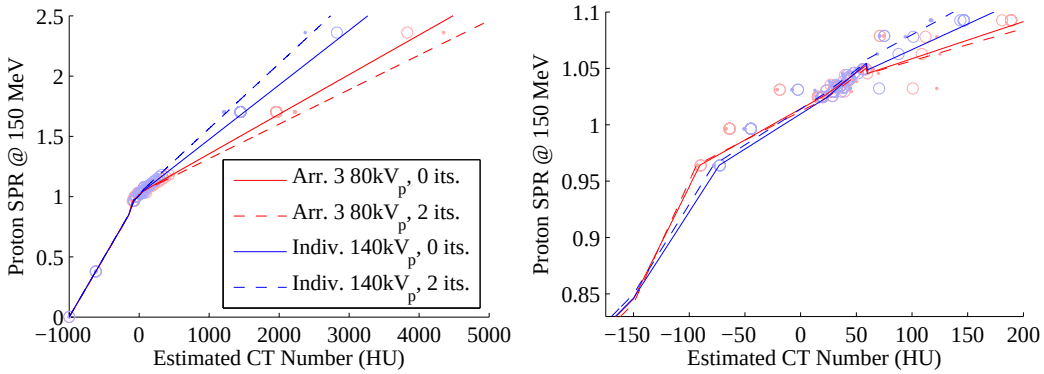


Figure 4.19 – Curves linking CT number and stopping power constructed from three linear fits as described by Uwe Schneider et. al. for stoichiometric calibrations with parameters given in table 4.8. Curves are overlaid on points from figure 4.18.

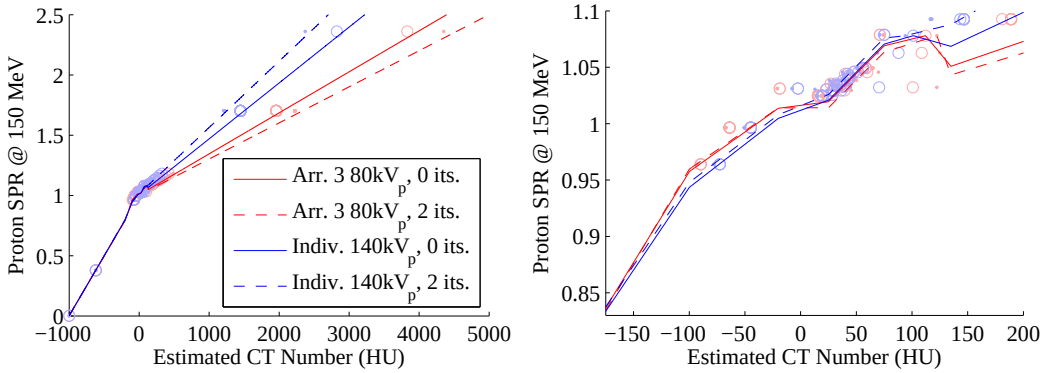


Figure 4.20 – Curves linking CT number and stopping power constructed from linear fits as described by Schaffner and Pedroni for stoichiometric calibrations with parameters given in table 4.8. Curves are overlaid on points from figure 4.18.

and bone). In the region below −200 HU, the fit is to the organs and muscle category; between −100 and −20 HU, it is to adipose tissue; between 25 and 75 HU, it is again to organs and muscles; and above 135 HU, the fit is to bone tissues. The gaps between −200 and −100 HU −20 and and 25 HU are filled by connecting the two neighbouring segments with a straight line. Between 75 and 135 HU, the line segments are joined through the point for cartilage. Figure 4.20 illustrates the results of the procedure, with r.m.s. deviations of the ICRP points from the curves given in table 4.9. The quality of fit to the phantom data is essentially equivalent to that of the previous method, although the undesirable discontinuity at −60 HU is avoided.

4.2.6.2 Material interpolation

Instead of performing a series of linear fits to SPRs and CT numbers calculated for a range of tissues, the method of Wilfried Schneider generates a chemical composition

and mass density for each HU value. Monte Carlo simulations, for which that data is a necessity, provided the original motivation, but it can also provide a basis for SPR calculations. Tissues were categorised by their major elemental components, and a number of regions were defined in which mixing two base tissues with varying ratios provided a reasonable approximation. The formulae equation (4.19) and equation (4.20), in which material X is a composite of materials 1 and 2 of unknown ratio, were then used to derive mass density ρ and weight fraction w_i for elements labelled by i, from CT number h. The Equations are a logical consequence of the mixture rule equation (3.1).

$$\rho_X = \frac{\rho_1 \cdot h_2 - \rho_2 \cdot h_1 + (\rho_2 - \rho_1) \cdot h_X}{h_2 - h_1} \tag{4.19}$$

$$w_{X,i} = w_{2,i} + \frac{\rho_1 \cdot (h_2 - h_X) \cdot (w_{1,i} - w_{2,i})}{\rho_1 \cdot h_2 - \rho_2 \cdot h_1 + (\rho_2 - \rho_1) \cdot h_X} \tag{4.20}$$

Tissues in Woodard and White provided the basis for W. Schneider's analysis, which can be extended to produce a calibration curve as follows. Below the CT number of Woodard and White's adipose tissue #3, the material is assumed to be lung of varying density. Interpolation then occurs between adipose tissue #3 and adrenal gland, between adrenal gland and small intestines, and between small intestines and connective tissue. Above the CT number of connective tissue, the interpolation is between a 50:50 red-yellow bone marrow mixture and cortical bone.

Results of the material interpolation method are displayed in figure 4.21 and table 4.9, as for the previous cases. Very minor improvements in the fit of the line to the ICRP points are evident, but errors remain in the range of 0.5-1.6% of the stopping power of water. Two discontinuities occur in the curves illustrated. It is suggested that such features should be avoided given that CT number resolution is far from ideal in the presence of artefacts.

4.2.6.3 Comparison

Figure 4.22 compares the results of the three approaches to defining calibration curves for two sets of input parameters. Variations of up to 5% of the stopping power of water can be seen between curves at some CT numbers. In the low-energy case, the SPR calculated for a number of tissues differs from that predicted by the line by at least 1%. Some of the deviation may be explained by the differing definitions of tissues in the ICRP phantom compared to ICRU/Woodard and White.

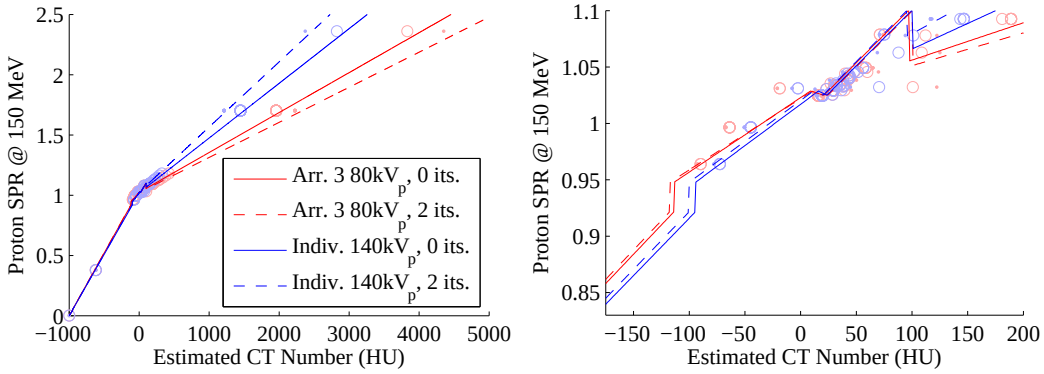
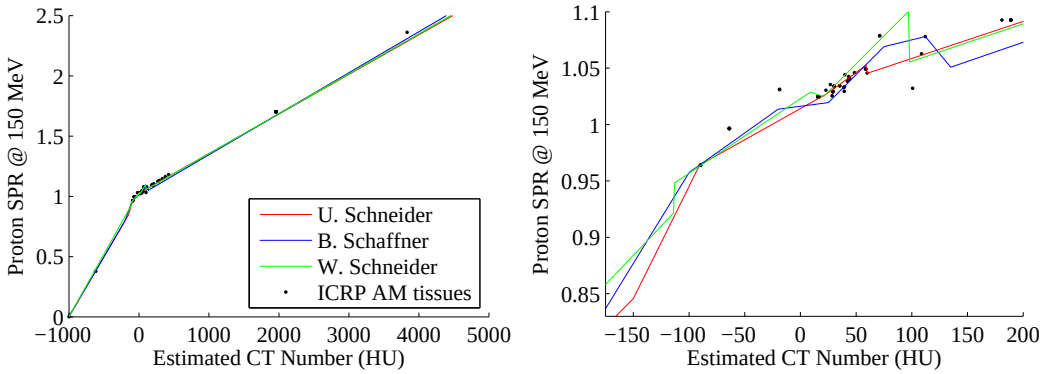
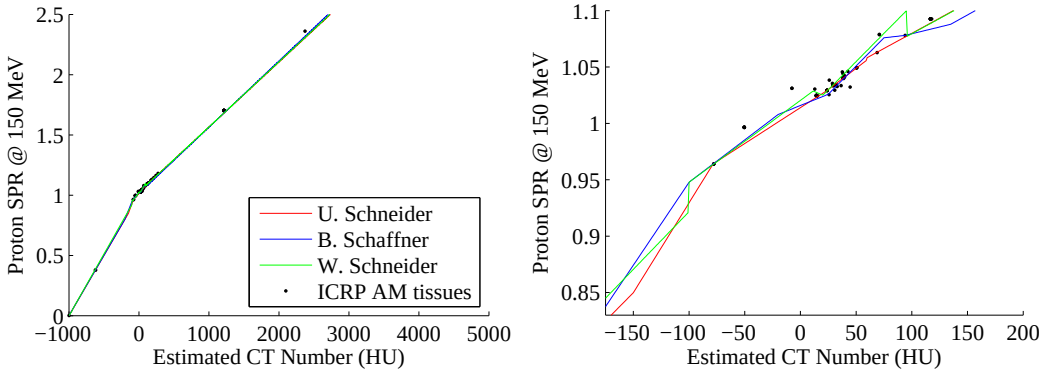


Figure 4.21 – Curves linking CT number and stopping power constructed by the material interpolation approach described by Wilfried Schneider et. al. for stoichiometric calibrations with parameters given in table 4.8. Curves are overlaid on points from figure 4.18.



(a) Arrangement 3, 80 kV_p, no BBHC



(b) Individual scans, 140 kV_p, 2 iterations BBHC

Figure 4.22 – A comparison of the three methods of generating CT-SPR curves from stoichiometric calibration parameters.

4.2.7 Overall assessment

This Chapter has shown that there is potential for clinically significant errors in the stopping power estimated from CT scans. Some of this may occur due to issues with the mapping of mass density to proton stopping power in the TPS, and a scenario that may result in overestimation of range in the lung by almost 4 mm was devised in Section 4.2.2.2. Provided that materials within the patient have no component with very high Z, the power laws traditionally used in stoichiometric calibration were found to be reasonable. However, the best-fit calibration parameters were found to vary based on the materials used to calibrate the scanner even in the ideal monoenergetic case. These were only significant at low energies and ruled out the use of characterisation materials whose composition were not sufficiently similar to biological tissues; with the right set of materials, it would be reasonable to expect CT numbers in human tissues to be predicted within ±4 HU for a single energy in the clinical range. Considering a polyenergetic source introduces complication due to the influence of beam hardening, which varies based on the energy spectrum in question and the position within the scan. At 150 mAs, photon statistics did not significantly affect the calibration. Volume-weighted r.m.s. CT number prediction errors of up to 23 HU were seen in the most challenging calibration scenario, but these were reduced by around 10 HU when 2 iterations of bone-based beam hardening correction were employed. At the highest tube voltage, with bone-based beam hardening correction, such errors were reduced to approximately 6 HU. These errors in predicted HU govern the accuracy with which CT numbers for the tissues that underlie the calibration curve may be estimated. For a particular calibration, common methods in defining the curve varied by up to 5% in the worst case, but more typically by about 1%.

The above assessment might lead one to conclude that typical uncertainties in the estimation of stopping powers from CT are in the region of 1-3%, depending on energy spectrum and use of beam hardening, assuming the calibration is to SPR as opposed to mass density. Nevertheless, the relationship between some of the uncertainties mentioned is complicated, and these numbers provide no information as to how errors vary with anatomy. Direct SPR calculations are possible for organs of the ICRP Publication 110 phantoms, for which the elemental composition is fully specified. Comparisons with the predictions made by mapping a calibration curve on to a simulated CT scan can reveal the anatomical locations at which stopping power uncertainty may be greatest. This will be the focus of the next Chapter.

Chapter 5

Estimating proton range uncertainties in anthropomorphic phantoms

5.1 Simulating CT scans of the ICRP 110 phantoms

This Chapter discusses features of the images resulting from simulation of CT in the ICRP Publication 110 voxellised phantoms (introduced in Section 3.1.2). The simulations were performed using the software presented in chapter 3, with the parameter set used for benchmarking, except for a 20% increase in the scanner's radius. This increase was necessary to accommodate the full width of the phantoms, which include arms, and was also adopted in the previous Chapter (see Section 4.2.5). In order to obtain results for the 200+ slices of the phantoms at multiple energies in a reasonable timeframe, the code was compiled for use outside of MATLAB, results of common calculations were cached, and runs were performed in parallel on the Oxford Particle Physics Linux batch cluster. Table 5.1 gives representative data on the computing resources required. Final reconstructed images were saved as DICOM files with sufficient metadata to allow their use in commercial TPS packages.

Different filtration is typically used in scanning the head compared to the body, and this results in a different energy spectrum. Slices 190-222 of the AM phantom and 306-348 of the AF phantom were thus considered separately, with an energy spectrum and a beam-hardening calibration data appropriate to the head.

5.1.1 Phantom resolution

When assessing the intensity of artefacts that arise from scan parameters, it is important to consider if any features of the phantoms used have influenced their

Stage of simulation	Output file size	Processing time
Pre-calculations: ICRP AM grid (254×127)	1.28 GB	200 s
Pre-calculations: image grid (512×512)	2 GB	160 s
Raytracing (forward-projection)	N/A	85 s
1 iteration of second-pass beam hardening correction	N/A	240 s
Reconstruction (back-projection)	1 MB	15 s

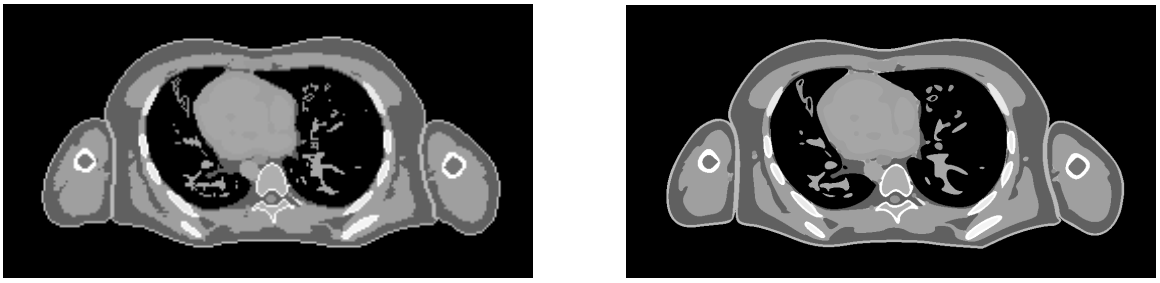
Table 5.1 – Approximate computational demands for stages of the CT simulation software when applied to the ICRP AM phantom. Pre-calculation steps only need to be performed once for each set of pixel coordinates – for these the times quoted are using all four cores of an Intel i5-2400 desktop processor. The remaining stages must be run for each slice and have been accelerated by the existence of pre-calculated data – for these the times are using one core of an Intel i5-2400, and may be shorter on the high performance cluster. Given the large size of the pre-calculation results, there are some circumstances in which speed can be limited by I/O performance rather than the CPU. The initial processing time for the ICRP AM grid is greater than that for the image grid due to the 4× oversampling employed in the first raytracing, as described in Section 3.3.2.1.

strength. Boundaries between tissues in the human body form continuous curves. In contrast, the voxellised nature of the phantom means that abrupt changes in attenuation coefficient occur at regular intervals along the x and y axes, on a grid that is larger than that of the reconstructed image. As seen in figure 2.12, aliasing artefacts are particularly intense in the presence of straight boundaries between materials of very different attenuation and these could be expected to occur more frequently when voxels are larger than the reconstruction grid. In order to ascertain if there is an appreciable effect, comparisons were made between CT images simulated for the same phantom at different resolutions.

A method of upsampling[1] and introducing additional curvature in the ICRP phantoms was developed through use of intermediary vector images, created from organ voxel maps by the open source bitmap tracing tool 'potrace'. This was used to create an upsampled version of the AM phantom with an in-plane voxel size of 0.53 mm, which is smaller than the corresponding figure of 1.17 mm in the reconstruction grid. The results of upsampling can be seen in figure 5.1, and show that, whilst the method has a tendency to 'fill in' very small heterogeneities within otherwise uniform tissues (for example the solitary voxels of adipose tissue within the muscle of the left arm), the phantom geometries are very similar.

In order to isolate the effects of aliasing and the raytracing algorithm, the simulations discussed in the situation are monoenergetic, and comparison is made with maps

[1] increasing the resolution



(a) Original phantom **(b)** 4× upsampled

Figure 5.1 – Mass density maps of one slice of the original and upsampled ICRP AM phantoms. Window: 0.4 g/cm^3 Level: 1.0 g/cm^3

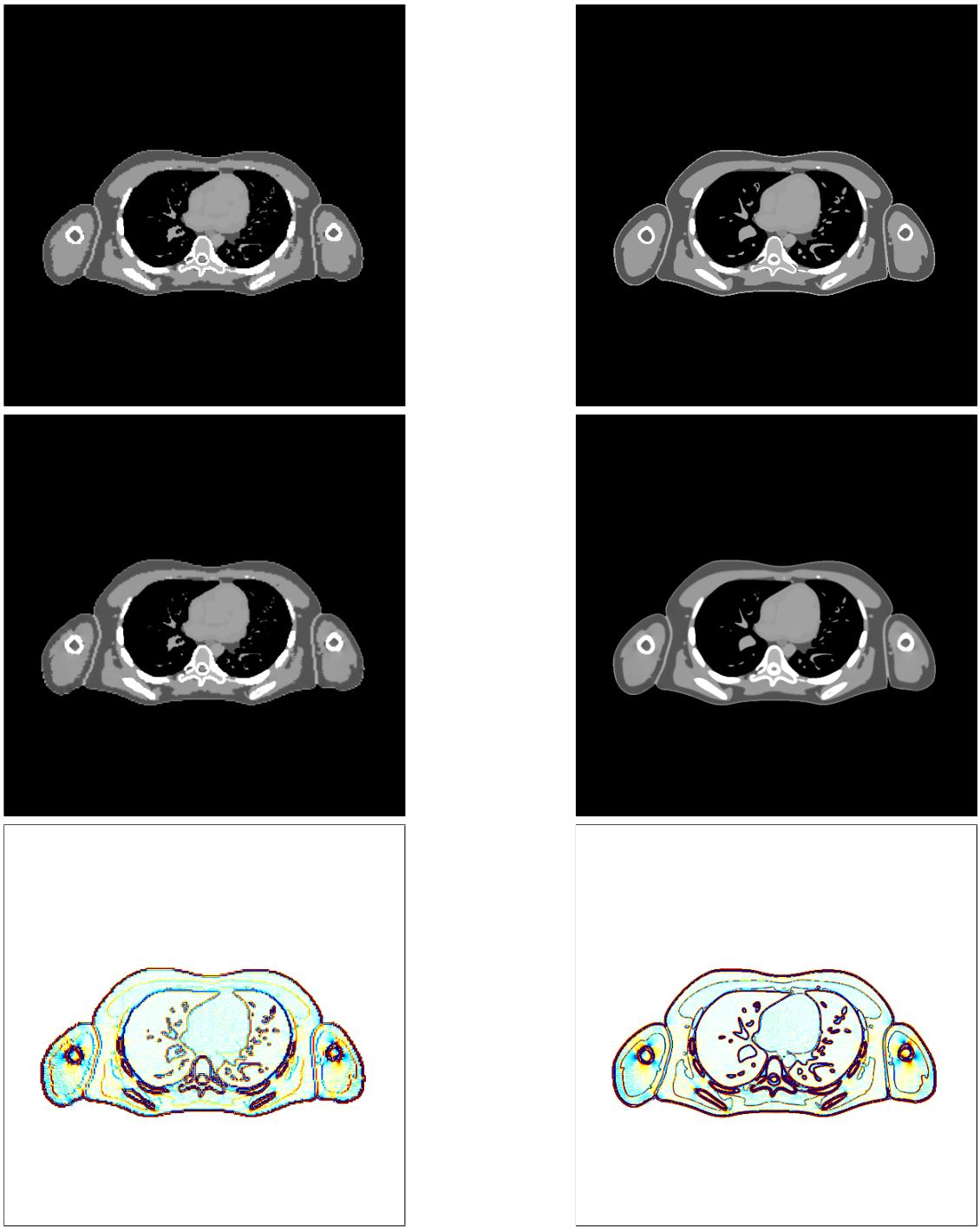
of CT numbers that are calculated directly from the elemental composition of the phantom's organs. A visual comparison of one slice is shown at 80 keV in figure 5.2, which is illustrative of differences seen throughout the entire phantom volume. There are some notable common features of both difference maps. Firstly, differences are apparent at boundaries between regions of radically different attenuation, appearing as dark borders surrounding the torso and arms in particular, as could also be expected when the edge spread function extends over multiple pixels[2]. However, inexact alignment of the coordinate systems may also be a contributing factor, and high-contrast edges occur less frequently in the human body than the phantom, so boundary pixels will be excluded from further consideration.

There are few discernible differences between simulations of phantoms defined at the lower and higher resolution. Whilst the noise has a slightly different spatial character, the mean and standard deviation of simulation artefacts within homogenous tissues are changed by less than 1 HU as a result of the change in resolution. Using the ICRP phantoms as originally specified should thus result in no increased image artefacting compared to higher resolution representations, but will carry significant benefit in the form of shorter computation time. They will therefore be used in the remaining Sections.

5.1.1.1 Removal of boundary pixels

As discussed in the previous Section, large deviations in CT number may occur between direct calculations and simulated CT scans in pixels that straddle organ boundaries. The decision was taken to exclude boundaries from comparisons between simulated and ideal images. The pixels to be excluded are those whose centres lie either side of a

[2]as it does for this simulator, illustrated earlier in figure 3.11



(a) Original phantom

(b) 4× upsampled

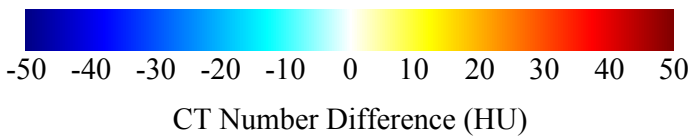


Figure 5.2 – Directly calculated CT numbers (top), monoenergetic simulated CT (middle) and difference between the two images (bottom) for one slice of the original and upsampled ICRP AM phantom at 80 keV. The top and middle images have a level of 0 HU and a window of 400 HU. A colour scale bar for the difference maps is shown above.

boundary (as determined by nearest neighbour interpolation of the organ map), plus one additional pixel either side account for the falloff in the edge transfer function. This removal procedure is quite stringent, retaining only 60% of all internal voxels. Nevertheless it is necessary to ensure that the regions being compared are entirely homogenous.

It is not expected that the removal of these pixels will have significant impact on conclusions drawn in soft tissue, but it should be noted that bone, in which beam hardening effects can be appreciable, is more likely to be excluded[3]: the mean and standard deviation mass density within the phantom change by $-0.03\,\mathrm{g/cm^3}$ and $-0.06\,\mathrm{g/cm^3}$ as a result of the exclusion process.

5.1.2 Aliasing

Figure 5.3 shows the r.m.s. deviation between simulated and ideal CT numbers in non-boundary voxels by slice within the patient. The situation depicted is monoenergetic at the reference energy for each $\mathrm{kV_p}$ value, thus revealing the noise which is inherent in the raytracing and reconstruction process, much of which could be termed aliasing. It is apparent that the noise is more prominent in broader regions of the phantom, especially the abdomen and pelvis slices in which the arms are well-separated from the body. Significantly more aliasing is visible in the arms of figure 5.2, which are on the edge of the reconstructible field-of-view and consequently have a lower sampling density – in fact, as displayed in figure 5.4, the maximum distance of the phantom surface from the isocenter is a good predictor for the observed CT number deviation. A secondary consideration is the existence of high-contrast boundaries, particularly visible in the vicinity of the teeth and skull-base: the artefacts in question can be seen around the bones of the arm in figure 5.2, taking the form of hypo- and hyper-intense regions that differ by $\pm 15\,\mathrm{HU}$. Whilst visually similar to beam hardening, their origin in this monoenergetic case is likely in-plane partial volume effects due to a sampling density which is insufficient to capture the sharp edges. The trend with energy is in the same direction as the CT number for bone, although more moderate.

It is clear that aliasing is a minor consideration in CT number accuracy for the scan parameters considered here, except at the very edge of the reconstructible region. When examining beam hardening and stochastic noise, polyenergetic scans will be compared against the monoenergetic simulations performed in this Section, thereby

[3]most bone regions are small in size, so a higher proportion of their voxels are classified as boundary voxels

separating the effect of aliasing and allowing direct voxel-by-voxel comparison without needing to consider organ boundary issues.

5.1.3 Beam hardening

Chapter 2 described the cupping and streaking artefacts that arise as the result of beam hardening when a realistic energy spectrum is emitted by the x-ray source. The anthropomorphic ICRP phantoms are more complex than the simple geometries illustrated there, having many more materials in any individual slice with irregular spatial arrangements. Differences in CT numbers that are introduced by replacing a monoenergetic source with a polyenergetic one therefore deserve particular attention, especially as beam hardening may result in systematic shifts in the measured attenuation over large distances.

Figure 5.5 shows examples of CT slices at sites of possible clinical interest before second-pass beam hardening correction is performed. Dark streaks are clearly visible emanating from the mineral bone in the ribs, arms and femora, being particularly prominent along lines that contain a large proportion of highly-attenuating material. The BBHC algorithm removes visible artefacting in two iterations, as evidenced in figure 5.6. However, it is quantitative rather than qualitative aspects of the image that are particularly important for the purposes of proton radiotherapy treatment planning. Thus, a comparison was made between the pixel values observed in polyenergetic simulations and in monoenergetic simulation at the relevant reference energy, the slice-by-slice results of which are given in figure 5.8. Summary statistics are shown in table 5.3.

Before second-pass beam hardening correction, the amount of bone in a slice modulates both the average and the spread of $h - h_{ref}$. This is particularly evident in the periodic nature of the lines in figure 5.8 around the ribs, and the increased deviations in the pelvis. It is interesting that the skew of the distributions within each slice varies with the peak energy of the x-ray spectrum, being slightly negative at $80\,\mathrm{kV_p}$, approximately symmetric at $100\,\mathrm{kV_p}$ and somewhat positive for $120\,\mathrm{kV_p}$ and $140\,\mathrm{kV_p}$. The relationship between energy spectrum and bone attenuation coefficients (plotted in figure 5.7) is such that, at $80\mathrm{kV_p}$, the beam is attenuated more by bone than would be expected at the reference energy, whilst the converse is true at $120\,\mathrm{kV_p}$ and above. It is not generally possible to choose a reference energy that would avoid this situation without prior knowledge of the contents of the imaging slice, and doing so would inhibit inter-comparison of CT numbers between slices. The trend in IQR is similar, with the least variation being observed at $100\,\mathrm{kV_p}$ and the most at $140\,\mathrm{kV_p}$.

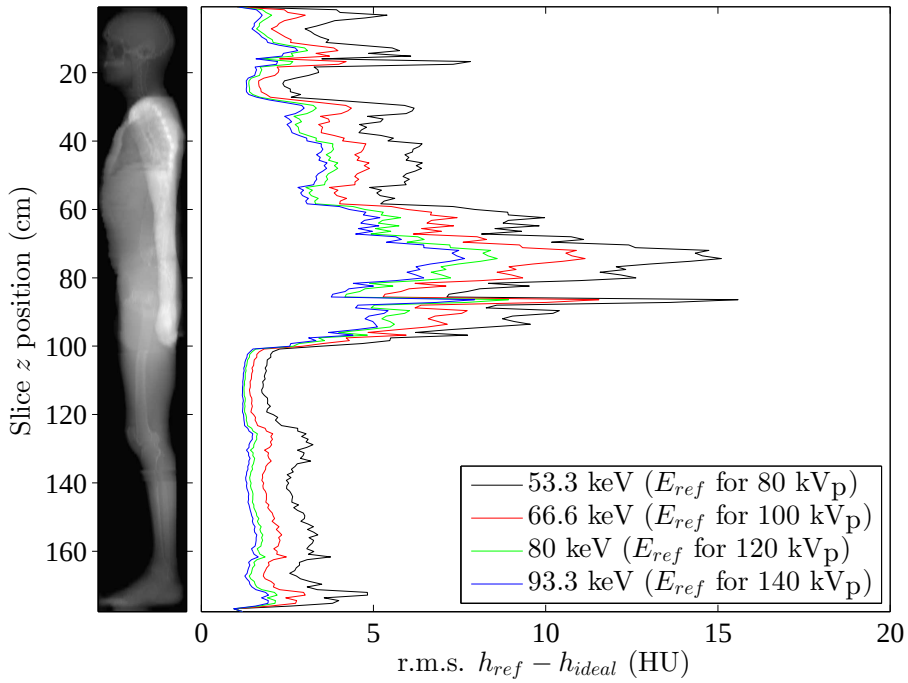


Figure 5.3 – Root-mean-square difference in the pixel values found in monoenergetic simulations compared to directly calculated, idealised maps of CT numbers at the reference energies. Boundary voxels have been excluded as described in Section 5.1.1.1.

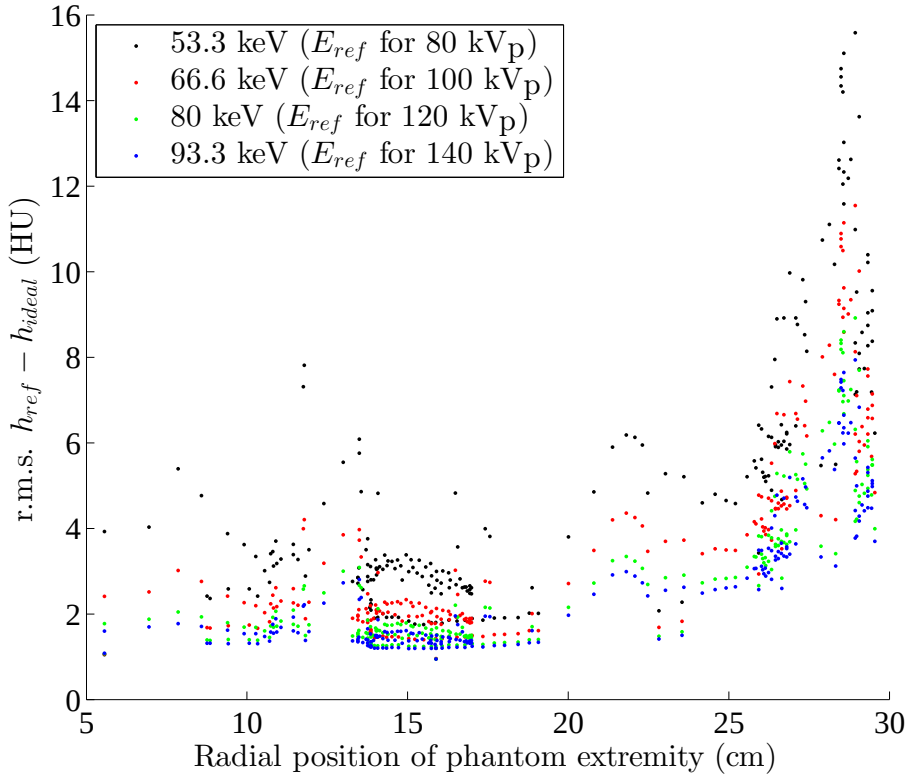
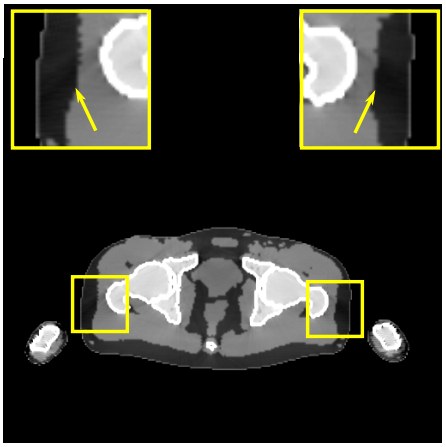
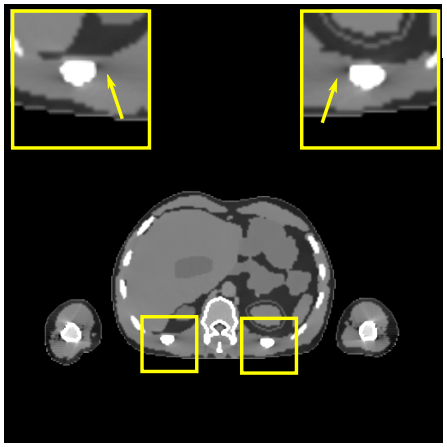


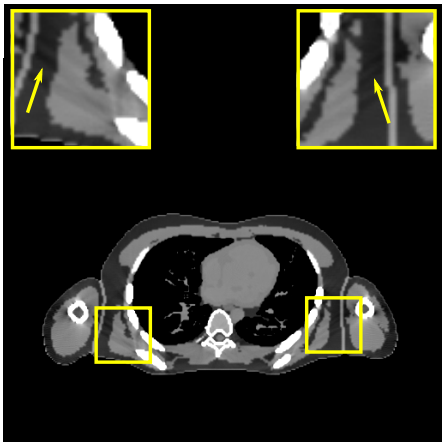
Figure 5.4 – The data in figure 5.3 plotted as a function of the displacement of the furthest point of the phantom from the isocenter.



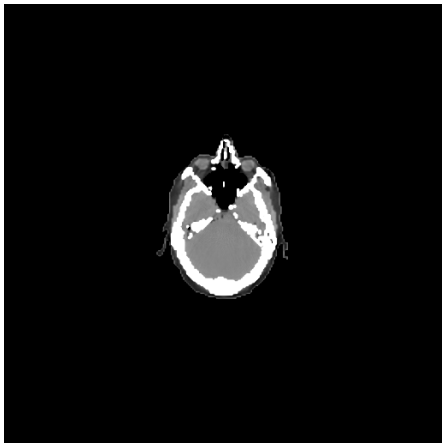
(a) slice 116: pelvis, prostate region



(b) slice 155: abdomen, liver region

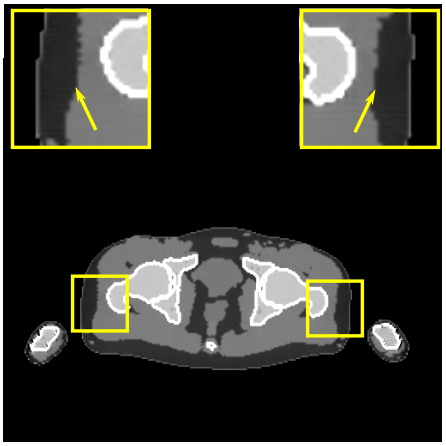


(c) slice 170: thorax, containing lung and heart

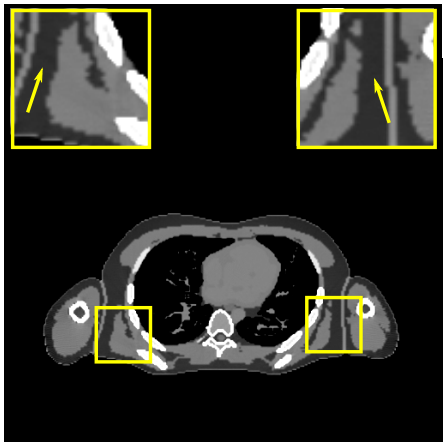


(d) slice 208: head (ear level)

Figure 5.5 – Example slices of polyenergetic simulated CT scans with an energy spectrum corresponding to 120 $\mathrm{kV_p}$. Only first-order (water-based) beam hardening correction has been carried out. Level: 0 HU, Window: 400 HU. Prominent beam hardening artefacts have been magnified and marked.



(a) slice 116: pelvis, prostate region



(b) slice 170: thorax, containing lung and heart

Figure 5.6 – Slices from figure 5.5 after two iterations of second-pass (bone-based) beam hardening correction. Level: 0 HU, Window: 400 HU. Annotations are as figure 5.5.

# BBHC Iterations	Pelvis		Abdomen		Thorax	
	median	IQR	median	IQR	median	IQR
0	3	38	1	14	1	18
1	-10	11	-8	9	-8	15
2	-7	11	-6	7	-5	13
3	-7	10	-6	7	-6	13
4	-7	11	-6	6	-6	13
5	-7	11	-7	7	-6	13
6	-8	11	-7	6	-7	13
7	-9	12	-7	6	-7	13
8	-11	12	-8	6	-9	14
9	-15	12	-10	8	-12	14
10	-23	14	-13	9	-18	16

Table 5.2 – Median and interquartile ranges for the difference in pixel values between CT slices simulated at 120 kV_p and the monoenergetic reference energy of 80 keV, with varying degrees of BBHC.

This may be somewhat unexpected, as beam hardening artefacts in soft tissues are more intense at lower tube voltages, but that is more than counteracted by the large changes seen in the smaller number of voxels containing mineral bone (which can be more than 200 HU).

Applying one iteration of beam hardening correction immediately reduces the overall spread in CT number deviations as the attenuation coefficients converge on monoenergetic values. There is a tendency for the first iteration to 'overshoot' to some degree, with average deviations for the head in particular being consistently negative and more so at higher kV_p values. Given bone's increased MAC compared to water for at least part of each energy spectrum, the mass of bone in the image will be overestimated from the initial reconstruction and the magnitude of the correction will be too great. A better estimate of bone content is obtained from the output of the first iteration, causing average deviations in the second iteration to be closer to zero. No more than two iterations are considered here: the lossy nature of repeated forward- and back-projection is seen to introduce noise into images that outweighs any residual benefit, especially in the face of increased computation time, and this is reflected in the data of table 5.2. It is notable that systematic shifts in the corrected vs. ideal CT number (ie. differences in median) are present at convergence. They will have no overall effect on the accuracy of SPR calibration techniques as they will also be present in calibration scans, but can be understood as a consequence of the number of materials in the phantom being greater than the three (air, water and mineral bone)

80 kV$_p$		$h - h_{ref}$ (HU)		
BBHC		AM, torso	AM, head	AF, torso
none	median	0	4	0
	IQR	15	28	17
1 it.	median	4	2	3
	IQR	8	15	11
2 its.	median	4	1	3
	IQR	7	13	11

100 kV$_p$		$h - h_{ref}$ (HU)		
BBHC		AM, torso	AM, head	AF, torso
none	median	0	8	-2
	IQR	18	34	21
1 it.	median	-1	-7	-3
	IQR	8	11	13
2 its.	median	-1	-4	-2
	IQR	7	11	13

120 kV$_p$		$h - h_{ref}$ (HU)		
BBHC		AM, torso	AM, head	AF, torso
none	median	0	8	-3
	IQR	23	45	27
1 it.	median	-9	-20	-12
	IQR	11	18	15
2 its.	median	-6	-7	-8
	IQR	11	14	16

140 kV$_p$		$h - h_{ref}$ (HU)		
BBHC		AM, torso	AM, head	AF, torso
none	median	0	7	-4
	IQR	22	44	27
1 it.	median	-11	-23	-13
	IQR	10	18	14
2 its.	median	-7	-7	-9
	IQR	10	14	15

Table 5.3 – Median and interquartile ranges of differences in CT numbers between monoenergetic and polyenergetic simulated scans, with varying amounts of second pass beam hardening correction. The head and torso regions refer to all voxels included in the upper and lower halves of figure 5.8 respectively.

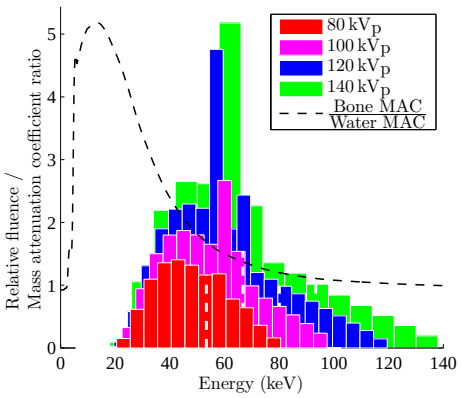


Figure 5.7 – The energy spectra used for the simulated scans of the torso, overlaid by the energy-dependence of the magnitude of the bone x-ray mass attenuation coefficient relative to that of water. White dashes show the position of the reference energy for each spectrum.

assumed by the BBHC algorithm.

5.1.4 Stochastic noise

If the x-ray exposure in a CT scan is small enough, the standard deviation in the number of photons arriving at a detector element can be significant relative to the mean[4], which causes stochastic noise as introduced in Section 2.6.1.2. Figure 5.9 shows the distribution of additional differences in pixel number arising from stochastic noise effects, after the beam hardening and aliasing described above have been considered.

As was seen in the benchmarking scans of tissue equivalent blocks, the spread in CT numbers decreases markedly as the the peak energy in the spectrum increases. The trend is an expected result: the x-ray attenuation coefficients of biological materials in the kilovoltage regime increase as energy decreases, so a higher proportion of the incident photons would be expected to be absorbed when spectra have a more significant low energy component.

Whilst the median difference in observed CT numbers is zero in most cases, for the remainder it is universally greater than zero. Two phenomena may explain this, both of which are only appreciable when the number of detected photons is small. Firstly, the Poisson distribution is positively skewed such that its median is almost always greater than its mean[5]. Secondly, even if the noise in the detected signal is symmetric, the corresponding variations in the ray attenuation sum ζ will not be as

[4] $\frac{\sigma}{\mu} \propto \frac{1}{\sqrt{n_{detected}}}$ in the Gaussian approximation

[5] This line of reasoning can get somewhat complicated due to the discrete nature of the distribution.

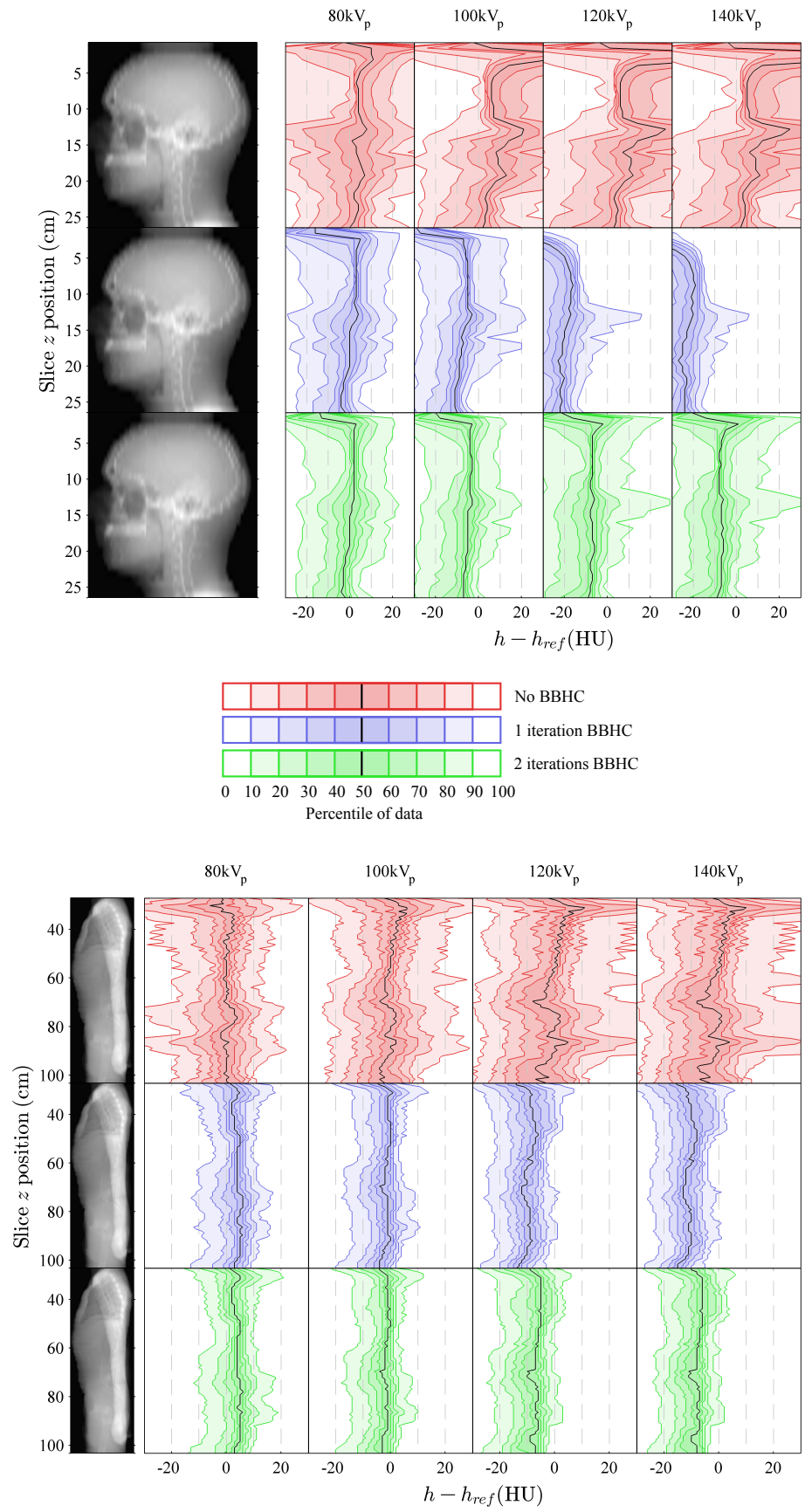


Figure 5.8 – The distribution of differences in pixel values between CT slices simulated with polyenergetic source spectra and monoenergetic reference scans, as a function of the slice position in the cranial-caudal direction. If beam hardening correction were perfect, the median difference would be zero and no spread in differences would be seen. Voxels external to the phantom (ie. air) have been excluded from this analysis.

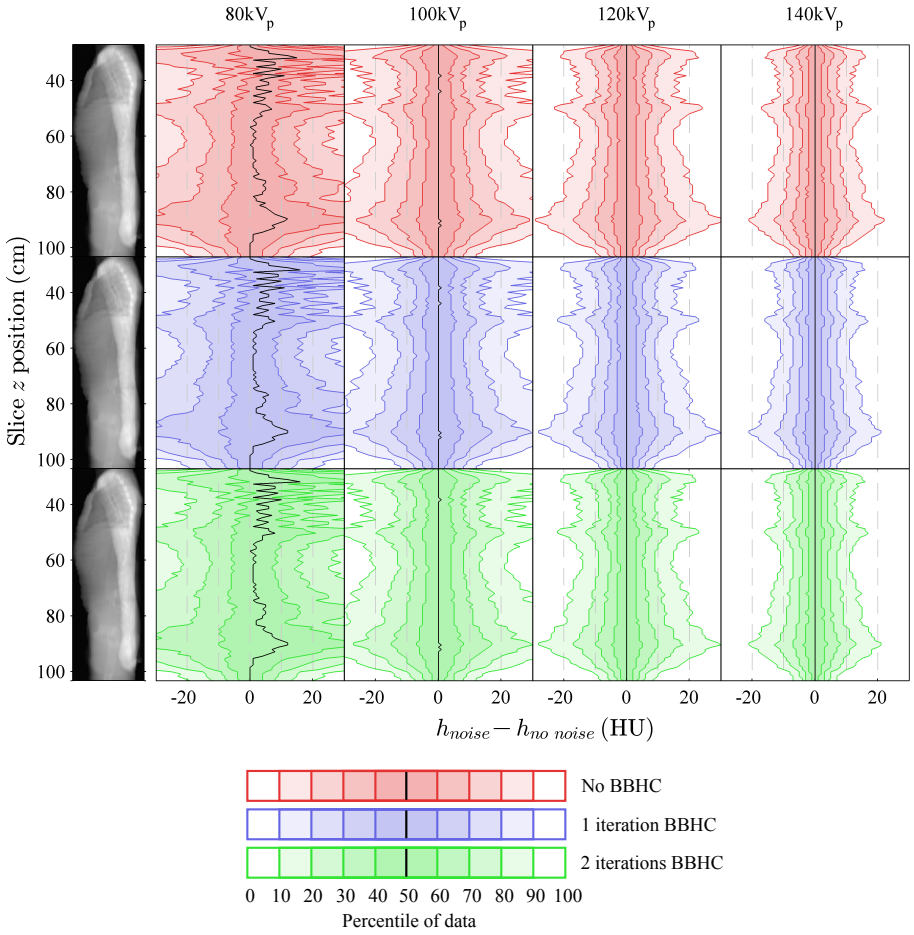


Figure 5.9 – The distribution of differences in pixel values between CT slices simulated with and without stochastic noise equivalent to an 150 mAs exposure, as a function of the slice position in the cranial-caudal direction. Voxels external to the phantom (ie. air) have been excluded from this analysis.

they are related by a logarithm. The gradient of the logarithm is greater when its argument is smaller, so statistical undercounts (ie. overestimation of attenuation) will affect the ray sum to a greater degree than overcounts.

It can be seen that the effect of second-pass beam hardening correction on the distributions is negligible.

5.1.5 CT number resolution

From the above assessment of CT number variation due to aliasing, beam hardening and stochastic noise, it is clear that a region containing homogenous material may show a degree of variation in pixel values in a realistic CT scan. The amount of variation in question is dependent on the scanner parameters, and the geometry and composition

of the surrounding areas. Given that the definition of CT calibration curves in proton therapy typically assumes that a material may be represented by a single CT number, it is interesting to observe the spread in pixel values within homogenous organs when simulated scans of the ICRP phantoms are performed. Table 5.4 gives the median and interquartile ranges of some of the organs of the AM phantom at 120 kV_p with and without noise, and after varying degrees of second-pass beam hardening correction.

The presence of stochastic noise sets a lower limit on the spread of CT numbers seen in any organ, and the variations arising from beam hardening appear to act independently of those caused by photon statistics. Some gains in CT number resolution occur as a result of the application of second-pass beam hardening correction in particular sensitive regions such as bone (e.g. femora, where the IQR is decreased by 45 HU) and organs close to bone (e.g. spinal cord, prostate, where the IQR decreases are 13 HU and 11 HU respectively). Similar improvements are seen at other kV_p settings. A reasonable summary may be that, in the absence of stochastic noise, the largest attenuation variations in regions of an appreciable size are ±2% before second-pass beam hardening correction and ±1% after. However, given that these are based on estimates of the interquartile range, it should be recognised that small areas may display deviations well beyond these limits.

5.1.5.1 Effect of the arms

The ICRP phantoms represent an individual lying supine with their arms to the side but, when performing CT scans, it is in fact typical to avoid having arms within the field of view. Unless there is a diagnostic reason, the additional attenuation of the arms and bones therein unnecessarily exacerbate beam hardening and photon starvation. For radiotherapy, the positioning of the patient must be reproduced for each treatment session – an extra source of positional uncertainty is removed if the arms are kept outside the scan volume.

An indicative estimate of the improvement in CT number resolution was obtained by removing the arms from the phantom. In every slice of the AM phantom where the arms were separate from the body, they were replaced with air. Table 5.5 gives the differences in the median and interquartile range of pixel values measured in the organs that were previously examined. Only small differences are apparent in the median and IQR without noise, suggesting that the difference in beam hardening is relatively insignificant. With noise, on the other hand, the IQR decreases by over 10 HU for some organs in the pelvic region, corresponding to a decreased spread in

	CT number without noise (HU)						CT number with noise (HU)					
# BBHC iterations	*0*		*1*		*2*		*0*		*1*		*2*	
Organ Name	**median**	**IQR**	**median**	**IQR**	**median**	**IQR**	**median**	**IQR**	**median**	**IQR**	**median**	**IQR**
Ascending colon contents	28	1.75	25	1	25	1	28	15	24	14.75	24	15
Ascending colon wall	32	5	26	2	27	2	32	16	25	14	26	14
Descending colon contents	27	2	24	2	23	1	27	16	24	15	23	15
Descending colon wall	29	3	25	2	25	2	30	14	25	12	26	13
Femora - upper half - cortical	1686	56	1317	13	1346	11	1685	60	1317	17	1348	15.5
Femora - upper half - medullary cavity	-24	10	-61	3	-48	4.25	-24	14	-61	11	-48	11
Femora - upper half - spongiosa	190	12	148	5	155	4	190	28	148	25	155	24
Kidney - left - cortex	45	4	37	2	37	2	44	20	37	19	37	19
Kidney - left - medulla	44	4	37	1	37	2	44	17	38	15	37	15
Liver	44	4	36	3	37	2	44	18	35	16	37	16
Lung - right - blood	58	8	47	6	48	4	57	19	47	16	48	17
Lung - right - tissue	-616	6	-621	3	-620	3	-616	14	-622	13	-621	13
Muscle - trunk	46	11	34	7	36	4	46	24	33	20	36	19
Pancreas	42	3	34	3	35	3	43	16	35	15	36	15
Prostate	18	16	15	5	16	5	21	37	16	33	17	32
Rectum wall	36	9	29	4	29	3	35	28	28	29	29	28
Residual tissue - trunk	-89	13	-84	8	-86	9	-89	20	-86	18	-87	18
Small intestine contents	29	2	25	2	25	2	29	15	25	14	25	14
Small intestine wall	33	4	26	3	27	2	33	14	25	13	27	13
Spinal cord	26	19	13	10	17	6	26	25	9	18.5	18	14
Spleen	43	6	31	3	33	2	44	29	32	26	34	27

Table 5.4 – Median and interquartile range for a selection of organs in simulated CT scans of the ICRP AM phantom at 120 kV_p. Values are given with and without stochastic noise at a level equivalent to an exposure of 150 mAs, and with varying degrees of second-pass (bone) beam hardening. Boundary voxels, as described in Section 5.1.1.1, were excluded from the statistics.

attenuation of around 1%. It would therefore seem to be worthwhile avoiding placing arms in the field of view when possible, given the potential ease of doing so.

5.2 Accuracy of proton path lengths

When considering the size of the CT number deviations observed alongside the calibration curves shown in figure 4.22, it is entirely apparent that there will be some degree of error in the pixels of the stopping power map, but it is less obvious how these errors will accumulate along the paths taken by protons in proton therapy. By applying the calibration curves of the previous Chapter to the phantom scans simulated in the first half of this Chapter, it is possible to produce estimated SPR maps that resemble those used internally by radiotherapy treatment planning systems, which can be compared to directly-calculated SPR maps for the same phantom. Figure 5.10 shows the pixel-level differences for one particular set of parameters: whilst many pixels show a predicted stopping power that differs from the reference value by more than 5% of water's stopping power, the integrated value over a number of pixels may 'regress to the mean', resulting in predicted range errors that are less dramatic. Therefore, the comparisons carried out in this Section will be in terms of the water equivalent path length along distances much greater than one pixel in length.

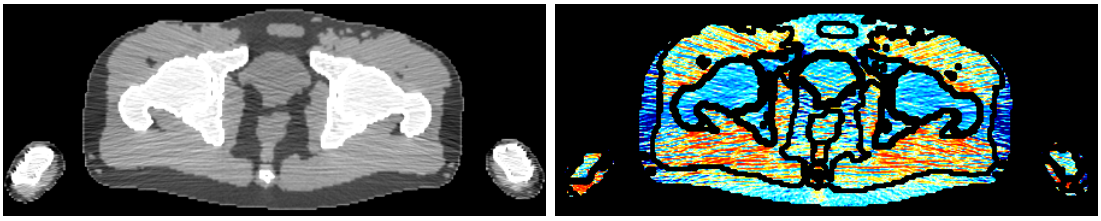
5.2.1 Comparison methodology

Proton radiotherapy typically uses an isocentric approach, which is to say that the central axes of multiple radiation fields converge on a single point in space (the isocenter). The position of the isocenter will depend upon clinical factors, in particular the location of the tumour. In order to avoid the selection of an isocenter but remain realistic, the paths used to investigate proton range uncertainty will be defined by the radon transform. The radon transform is the parallel-beam equivalent of the fanbeam forward-projection process studied extensively in Section 3.3. The radon geometry is illustrated in figure 5.11. MATLAB's `radon` function will be used to perform the radon transforms.

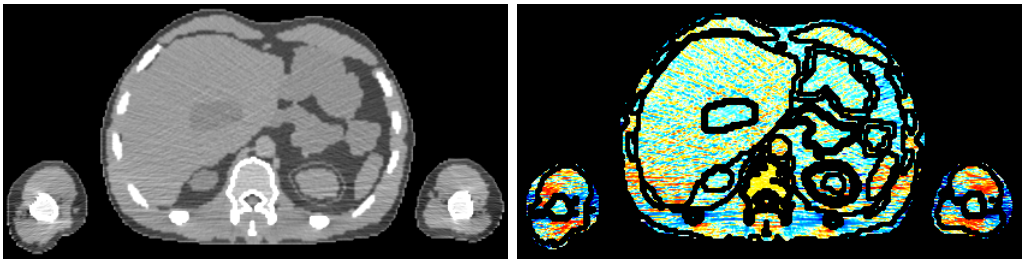
For each calibration curve and simulated CT slice, an estimated SPR map has been created by lookup of the value assigned to the CT number of each pixel. Linear interpolation and extrapolation are employed for unspecified values. True SPR maps were generated by calculating the SPR for each organ of the AM phantom at 150 MeV using SRIM; linear interpolation was then employed to transform the resulting 3D SPR map from the phantom grid to the image grid. The difference of the estimated

Table 5.5 – Change in the values given in table 5.4 when the phantom's arms are removed.

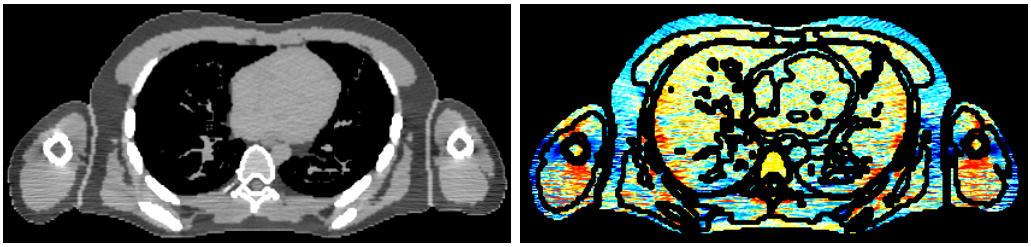
	Without noise (HU)						With noise (HU)					
# BBHC iterations	***0***		***1***		***2***		***0***		***1***		***2***	
Organ Name	**Δmed**	**ΔIQR**	**Δmed**	**ΔIQR**	**Δmed**	**ΔIQR**	**Δmed**	**ΔIQR**	**Δmed**	**ΔIQR**	**Δmed**	**ΔIQR**
Ascending colon contents	+1	+0	-2	0	-1	0	+1	-5	-1	-6	0	-6
Ascending colon wall	0	-1	-2	0	-1	0	+1	-4	-1	-4	0	-4
Descending colon contents	+1	+1	-2	-1	0	0	+1	-6	-2	-6	0	-6
Descending colon wall	+2	0	-2	+1	-1	0	+1	-4	-2	-3	-2	-4
Femora - upper half - cortical	0	-1	0	0	0	-1	0	-1	0	+1	-1	+1
Femora - upper half - medullary cavity	0	0	0	0	0	+1	0	0	0	-1	0	0
Femora - upper half - spongiosa	+5	-1	-3	0	0	0	+5	-6	-3	-6	0	-6
Kidney - left - cortex	0	-1	-3	0	-1	0	0	-9	-2	-10	-1	-10
Kidney - left - medulla	0	-1	-3	0	-1	0	0	-6	-4	-6	-1	-5
Liver	0	-1	-2	-1	-1	0	0	-7	-1	-6	-1	-6
Lung - right - blood	-3	-2	-1	0	-1	0	-2	-8	-2	-6	-2	-8
Lung - right - tissue	+1	0	-1	+1	-1	0	+1	-2	0	-3	0	-4
Muscle - trunk	+1	-2	-3	0	0	0	+1	-8	-2	-6	0	-7
Pancreas	0	+1	-1	-1	0	0	-1	-4	-2	-5	-1	-4
Prostate	-2	0	0	-1	0	-1	-4	-12	-1	-11	-1	-10
Rectum wall	0	-1	-2	-1	0	0	+1	-10	-1	-13	0	-12
Residual tissue - trunk	0	0	0	0	0	0	0	0	0	0	0	0
Small intestine contents	+1	0	-5	+3	-3	+2	0	-5	-4	-5	-2	-5
Small intestine wall	0	0	-1	-1	-1	-1	0	-5	-1	-5	-1	-5
Spinal cord	+1	0	-2	0	0	+1	+1	-4	-1	-4	-1	-4
Spleen	-2	-3	-1	-1	-1	0	0	-7	+2	-4	0	-2



(a) slice 116: pelvis, prostate region



(b) slice 155: abdomen, liver region



(c) slice 170: thorax, containing lung and heart

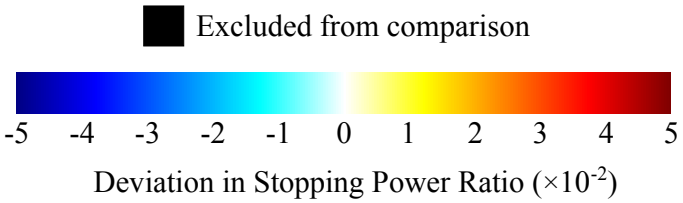


Figure 5.10 – Pixel-level difference between the SPR estimated from a simulated CT scan (120kV$_p$, no BBHC, stochastic noise) using a CT calibration curve (Schaffner method, sample arrangement 2 in figure 4.12), and that calculated directly from the AM phantom's composition data using SRIM. Pixels that are black meet the exclusion criteria given in Section 5.1.1.1.

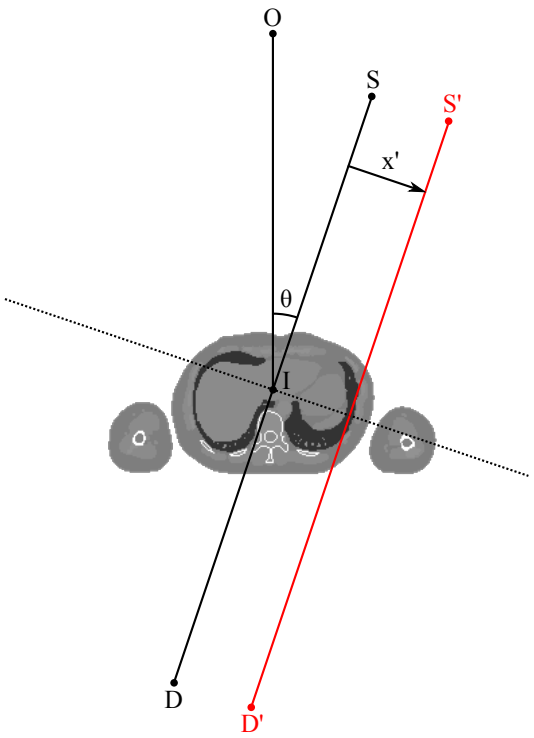


Figure 5.11 – The parallel geometry used in the radon transform. The radon transform of a map at the coordinate (θ, x') returns the line integral of the map along the line S'D' defined by that coordinate.

and true maps reveals the error introduced by the process of CT scanning followed by application of a CT-to-SPR calibration curve. The radon transform of the difference map gives the errors in water equivalent path length along all straight lines which traverse a slice.

The nature of the radon transform means that the paths examined traverse the entire cross section of the phantom. In reality, it would be rare for protons used in treatment to traverse much more than half the thickness of the body. When treating a patient, the path length of protons is typically kept minimal on the basis that it is beneficial to irradiate less tissue, assuming no other constraints[6].

Paths with an actual water equivalent length less than 5 cm were not included in the comparison, as they were considered to be unlikely to be used clinically. In addition, organ boundaries were again excluded from comparison by disregarding errors arising from voxels meeting the criteria given in Section 5.1.1.1. In the following Sections, the difference between water equivalent path length derived from the CT-estimated and true SPR maps for straight line paths in the transverse plane, after excluding short paths and boundary effects, will be denoted δ.

[6]The desire to avoid critical radiosensitive structures often provides an additional constraint.

The comparisons below look at the distribution of δ within the phantom as a whole (ie. over straight-line paths in the transverse plane, with any value of z) and within three clinically-interesting slices (pelvic, abdominal and thoracic, in the paths examined are those with a particular value of z).

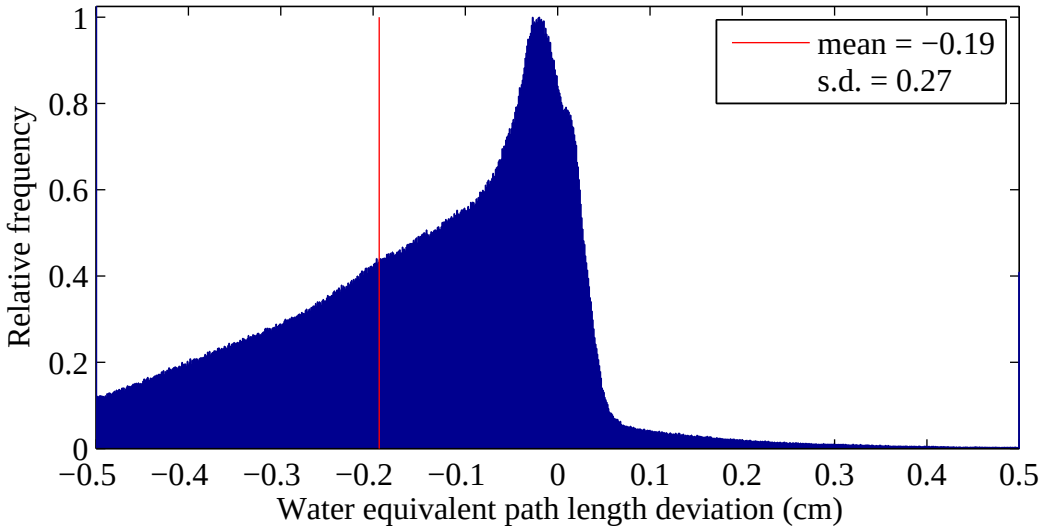
5.2.2 Results

Taking into account the full range of scanning parameters and calibration curves considered, the maximum r.m.s. value of δ seen in the full phantom was 3.0 mm. This was observed at 80 kV_p with stochastic noise and no BBHC, with calibration performed using the Uwe Schneider method [74] and calibration materials (tabulated in table 4.3a) being scanned individually in the center of a water phantom. The size of this deviation should be viewed with the caveat that on average 13% of the voxels along a path are excluded due to meeting the organ boundary criteria – in reality, the relative uncertainty within these voxels could be similar or somewhat larger (due to the increased magnitude of beam hardening effects in bone), so the r.m.s. deviation may be underestimated by a similar percentage[7].

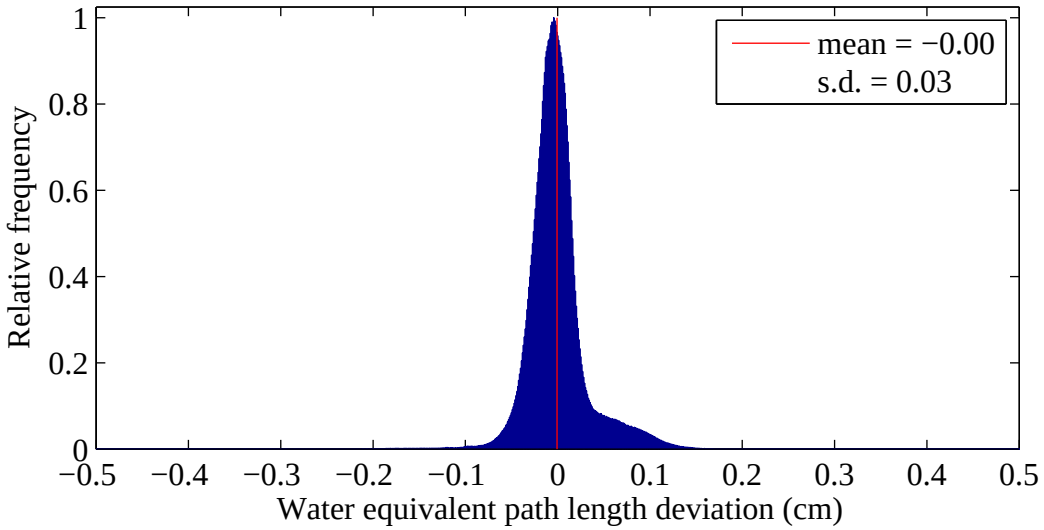
δ distributions for the full phantom are shown as histograms in figure 5.12. The two examples shown cover the best and worst combinations of scanning parameters and calibration curves: some level of skew is apparent in both distributions, and is also seen in the majority of other combinations. Therefore, r.m.s. deviation has not been used in the remainder of this discussion; instead, the 'shift' inherent in δ will be represented by its median and the 'spread' by the positions of the quantiles corresponding to 2.5% and 97.5% of the data. Together, the three statistics can provide a 95% confidence interval for water equivalent path length determination by CT, which constitutes a worst case estimate of the 95% confidence interval for range estimation in proton radiotherapy treatment planning.

The impact of each scan simulation parameter and calibration curve definition method on the confidence interval for δ is discussed below. The peak tube voltage is considered as a secondary parameter in each case; it is important to note that the metric considered here is not the only factor to consider when selecting the tube voltage. In particular, target volumes and organs at risk must be outlined on the planning scan by clinicians – soft tissue contrast is generally worse at high kV_p, making

[7] If the SPR is consistently underestimated by 2% in all excluded voxels (a relative deviation comparable to that found for the CT number of bone in table 5.4, and seen in the ribs in figure 5.10b), the r.m.s. value of δ for this simulated scan and calibration curve increases to 4.0 mm. In the extremely unlikely event that the error in SPR is underestimated by 5% in all excluded voxels, it increases to 6.3 mm.



(a) 80 kV_p, stochastic noise, no BBHC. Calibration with the Uwe Schneider method, calibration materials scanned individually.



(b) 140 kVp, no stochastic noise, 2 iterations of BBHC. Calibration with the Wilfried Schneider method, calibration materials scanned using Arrangement 2 of the multi-material phantom.

Figure 5.12 – Histograms of the difference between water equivalent path lengths calculated from a simulated CT scan and calibration curve, and those calculated from 'true' stopping power maps, for paths defined by the radon transform in all slices of the ICRP AM phantom. The two examples shown cover the best and worst cases.

	δ_{median} (mm)				$\delta_{97.5} - \delta_{2.5}$ (mm)			
kV_p	80	100	120	140	80	100	120	140
With noise	-0.7	-0.3	-02	-0.2	7.4	3.5	3.0	2.7
Without noise	-0.3	-02	-0.2	-0.1	4.2	3.2	2.9	2.7
Difference	-0.4	-0.1	-0.1	0.0	3.2	0.3	0.1	0.0

Table 5.6 – Parameters of the 95% confidence interval for δ in the whole AM phantom with and without noise equivalent to a 150 mAs exposure. No second-pass beam hardening correction was employed. Calibration curve was generated using the Uwe Schneider method and arrangement 2 of the multi-material calibration phantom.

this task more difficult. The radiation dose that the patient will receive should also be taken into account, generally increasing as the tube voltage increases, as this has been associated with an increased risk of secondary cancers and other negative side effects.

5.2.2.1 Stochastic noise

Table 5.6 gives summary statistics for δ throughout the whole phantom in the case of the worst-performing calibration curve with and without stochastic noise. The data suggests that, for the CT scan parameters and calibration approaches investigated, the worst case 95% confidence interval for proton range predictions within the ICRP AM phantom is -0.7 ± 7.4 mm. However, the confidence intervals are significantly narrower in most cases.

It is apparent that stochastic noise introduces much greater spread in the δ distribution at 80 kV_p than for any of the other energy spectra. This is simply due to the much greater average attenuation coefficient seen by x-rays with that spectra, as may be inferred from figure 5.7, leading to smaller numbers of photons traversing the phantom and thus a larger relative variance in the number of counts measured at each detector element. At 100 kV_p and above, the spread attributable to stochastic noise is sub-millimetre. Noise is seen to have a small effect on the median of δ, which is a result of the logarithmic relationship between the number of photons detected and the attenuation coefficient. Without noise, the median water equivalent path length deviations are well below 1 mm.

After accounting for the effects of stochastic noise, switching between energy spectra nevertheless causes variation in the width of the confidence interval of over 50%. The same trend is seen in CT number deviations in figure 5.9, and can therefore be attributed to changes in beam hardening effects with energy.

	δ_{median} (mm)				$\delta_{97.5} - \delta_{2.5}$ (mm)			
kV_p	80	100	120	140	80	100	120	140
No BBHC	-0.3	-0.2	-0.2	-0.1	4.2	3.2	2.9	2.7
1 iteration	-0.3	-0.3	-0.3	-0.3	2.9	1.7	1.4	1.7
2 iterations	-0.3	-0.3	-0.3	-0.2	2.8	1.7	1.2	1.1

Table 5.7 – Parameters of the 95% confidence interval for δ in the whole AM phantom for varying degrees of second-pass (bone) beam hardening correction. No stochastic noise was included. Calibration curve was generated using the Uwe Schneider method and arrangement 2 of the multi-material calibration phantom.

5.2.2.2 Second-pass (bone) beam hardening correction

Table 5.7 gives summary statistics for δ throughout the whole phantom in the case of the worst-performing calibration curve for varying degrees of second-pass (bone) beam hardening correction. After just one iteration of BBHC, the width of the 95% confidence interval can be reduced to less than half of its size in the no-BBHC case. In the case of the best-performing calibration curve[8], the width of the 95% confidence interval can be reduced to 1.0 mm by employing 2 iterations of BBHC, even with realistic levels of noise.

The changes in the distribution of water equivalent path length deviations with energy and number of BBHC iterations very much mirror those seen in CT number deviations, evident when comparing table 5.7 and figure 5.8. It can therefore be inferred that most of the narrowing of the confidence interval arises from correction of streaking within the phantom CT images, rather than from errors in the calibration process, which are more likely to be of a systematic nature.

5.2.2.3 Calibration curve definition

The calibration curves tested were generated by three variations of the stoichiometric method, with a number of different arrangements of the calibration materials. The manner in which the confidence interval changes with calibration curve is very consistent at all energies and degree of BBHC. In all cases, the difference in the median of the δ distribution as a result of changing the calibration curve is less than 1 mm (0.5 mm if 80 kV_p is disregarded), as seen in the data of table 5.8.

The width of the 95% confidence interval of δ shows differences of greater than 1 mm between the three variations of the stoichiometric method. Table 5.9 compares the performance of the Schaffner [75], U. Schneider and W. Schneider [76] approaches

[8]Schaffner method with calibration performed using arrangement 2 of the multi-material phantom

	no BBHC				2 its. BBHC			
kV_p	80	100	120	140	80	100	120	140
Schaffner	-0.38	-0.26	-0.19	-0.18	-0.50	-0.38	-0.34	-0.30
U. Schneider	-0.35	-0.26	-0.20	-0.17	-0.32	-0.29	-0.29	-0.27
W. Schneider	+0.25	+0.18	+0.17	+0.13	+0.21	+0.02	-0.04	-0.07

Table 5.8 – Median value of δ in the whole AM phantom, given in millimetres, for three variations of the stoichiometric method. No stochastic noise was incorporated. Tabulated values are mean averages for the five different arrangements of calibration materials.

	no BBHC				2 its. BBHC			
kV_p	80	100	120	140	80	100	120	140
Schaffner	3.3	2.3	2.1	1.9	2.0	1.3	1.1	1.0
U. Schneider	4.2	3.2	2.8	2.6	2.8	1.7	1.2	1.1
W. Schneider	4.2	3.4	3.1	2.8	3.1	2.1	1.5	1.3

Table 5.9 – Width of the 95% confidence interval for δ (ie. $\delta_{97.5} - \delta_{2.5}$) in the whole AM phantom, given in millimetres, for three variations of the stoichiometric method. No stochastic noise was incorporated. Tabulated values are mean averages for the five different arrangements of calibration materials.

to stoichiometric calibration – it can be seen that in all cases the predictions of the Schaffner method are the most consistent, and those of the W. Schneider method are the least consistent. The greater magnitude of the errors introduced by the W. Schneider approach may be explained by reference to figure 4.22, in which large changes in estimated SPR occur over changes in CT number of 1 HU. It has already been shown that CT number resolution is generally well above 1 HU for most organs of the phantom – some voxels of organs whose ideal CT number is close to these abrupt changes will therefore have significant errors in their estimated SPR, especially if situated in a region where beam hardening artefacts are appreciable.

5.2.2.4 Anatomical location

The data shown so far in this Section has pertained to δ distributions for the whole phantom, but there is clearly significant variation in the distribution of δ for individual CT slices because of the differing anatomy present in that slice. In particular, errors in estimating water equivalent path length might be expected to be greater in slices containing relatively high amounts of bone due to a correspondingly higher prevalence of beam hardening artefacts.

Distributions of δ were determined for the three clinically-interesting slices illus-

trated in figure 5.10 (pelvis, abdomen, thorax). Summary statistics for the 95% confidence interval are given in table 5.10, and samples of paths that lie outside the confidence interval at 120 kV_p for a particular calibration curve are shown visually in figure 5.13. Both stochastic noise and beam hardening artefacts are present in the images used here, and it is clear that their effects have a greater impact on some slices than others.

The values in table 5.10 suggest that, of the three slices considered, the greatest errors in proton range prediction are likely to be located in the pelvic slice (containing the prostate, a common target region in clinical radiotherapy). A high proportion of the slice in question is composed of bone, due to the presence of the femoral heads, and this causes both beam hardening and photon starvation artefacts. The width of the δ distribution is especially great at 80 kVp, due to the extreme photon starvation resulting from the increased x-ray attenuation at low energies. However, even at 140 kV_p, when photon statistics are of minor concern, predicted range uncertainties are at least 50% larger than in the other slices due to the effects of beam hardening. Of particular note in figure 5.13a is the path running from left to right, running through the femoral heads and very close to the prostate gland itself. This path is commonly used in proton radiotherapy of the prostate, by parallel opposed beams [92], which traverse approximately half the patient's width. Therefore, if the CT artefacts in reality are of similar magnitude to those seen in this simulation, range uncertainties resulting from image acquisition and stoichiometric calibration alone may exceed 2 mm.

Examining the trajectories of paths in all three images of figure 5.13 leads to the conclusion that path lengths are typically underestimated if they go through large amounts of adipose tissue, which accounts for much of the uncertainty seen in the thoracic slice. The physical size of range uncertainties in lung tissue should be expected to be three to four times the water equivalent size due to the difference in density, therefore some of the paths in that slice (which might realistically be chosen when treating a lung tumour) could cause over- or under-shoots of greater than 1 mm.

Soft tissue is most prevalent in the abdominal slice, with the result that beam hardening and photon stochasticity-related artefacts are mild, and a narrower δ distribution is also observed.

Figures 5.14 to 5.19 illustrate how the deviations between CT-estimated and true water equivalent path length build up for a 120 kV_p simulation with and without BBHC and stochastic noise, for a calibration curve determined by the Schaffner method. In the case of the prostate slice without BBHC and with stochastic noise, an average

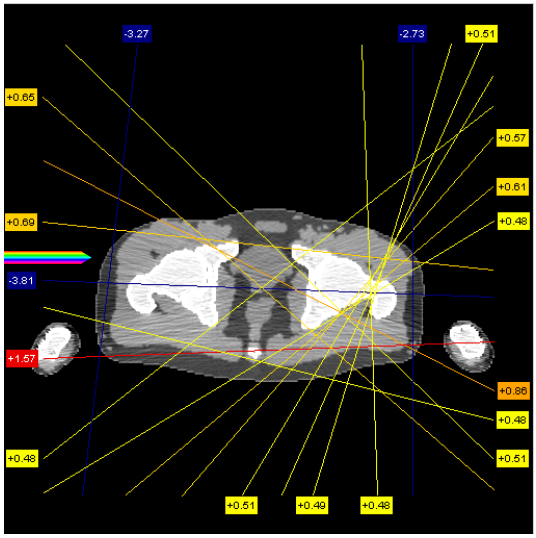
	δ_{median} (mm)				$\delta_{97.5} - \delta_{2.5}$ (mm)			
kV_p	80	100	120	140	80	100	120	140
slice 116: pelvis, prostate region	-2.2	-0.5	-0.3	-0.3	6.9	3.3	3.1	2.8
slice 155: abdomen, liver region	-0.5	-0.2	-0.1	-0.1	2.4	1.4	1.3	1.2
slice 170: thorax, containing lung and heart	-1.0	-0.1	-0.1	-0.1	5.2	2.4	2.0	1.9
Whole phantom	-0.6	-0.2	-0.1	-0.1	7.2	2.6	2.2	2.1

Table 5.10 – Parameters of the 95% confidence interval for δ in three individual slices of the ICRP AM phantom from simulated scans at four tube voltages, with stochastic noise equivalent to 150 mAs and no second-pass beam hardening correction. Calibration curves were generated using the Schaffner approach and arrangement 2 of the multi-material phantom.

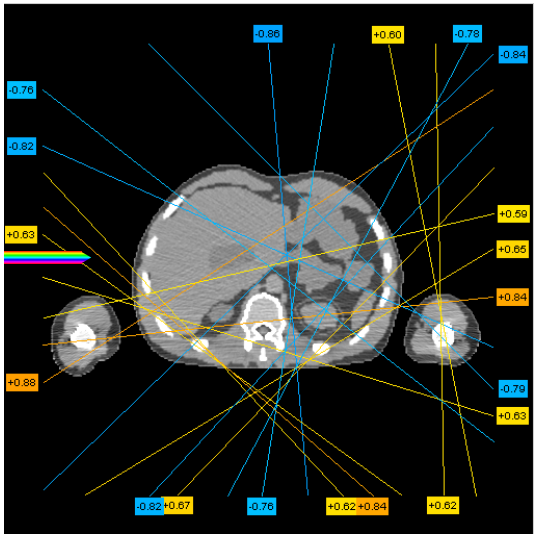
of 1.5 mm error in predicted range will have built up by the time a depth of 15 cm within the patient has been reached. There is also approximately 1 mm spread in that deviation, which will mean that the distal falloff in delivered dose distributions would be correspondingly spread out compared to that calculated in a treatment planning system based on CT. Similar spreading out is apparent in the abdominal slice, but there the water equivalent path length is well-estimated on average without BBHC. It is apparent that the bulk of the variation seen between each of these paths is a result of the statistical noise in the image, whilst beam hardening correction has an effect on the average deviation observed at depth. The Figures also show the same path length deviation data as a percentage of the true path length, under the assumption that the relative deviation remains constant in the excluded boundary voxels. For the same pelvic slice, with no BBHC and with stochastic noise, the spread in relative deviations between paths converges to approximately 1% once a sufficient amount of tissue has been traversed (˜10 cm). In the corresponding abdominal case, the final median deviation is close to zero, but in the pelvic slice a systematic underestimation of 1.25% is apparent. Data is also shown for the thoracic slice. It should be acknowledged that, whilst results are generally better with BBHC and without stochastic noise, the abdominal slice shows a deterioration in the quality of range estimation by 0.1% on average when using BBHC; however, an effect of this magnitude is washed out by that of photon statistics.

5.3 Summary

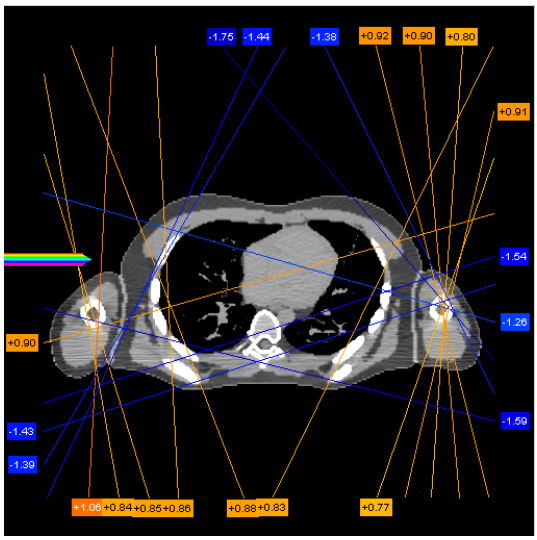
This Chapter has shown that variations of 2-4% may be expected in the CT numbers observed within organs of the ICRP AM phantom, under realistic scanning conditions. It has examined how the size of the errors in measured CT numbers changes with a



(a) slice 116 - pelvis, prostate region



(b) slice 155 - abdomen, liver region



(c) slice 170 - thorax, containing heart and lungs

Figure 5.13 – Examples of paths lying outside the 95% confidence interval for δ in three slices of the ICRP AM phantom, with CT simulated at 120 kV_p with stochastic noise equivalent to 150 mAs and no second-pass beam hardening correction. The calibration curve used was generated using the Schaffner to the stoichiometric method and arrangement 2 of the multi-material calibration phantom. Each path is labelled with the corresponding deviation between predicted and true water equivalent path lengths (ie. the value of δ) in millimetres. Rainbow arrows in a) and b) show the starting points of the paths examined in Figures 5.14 to 5.19.

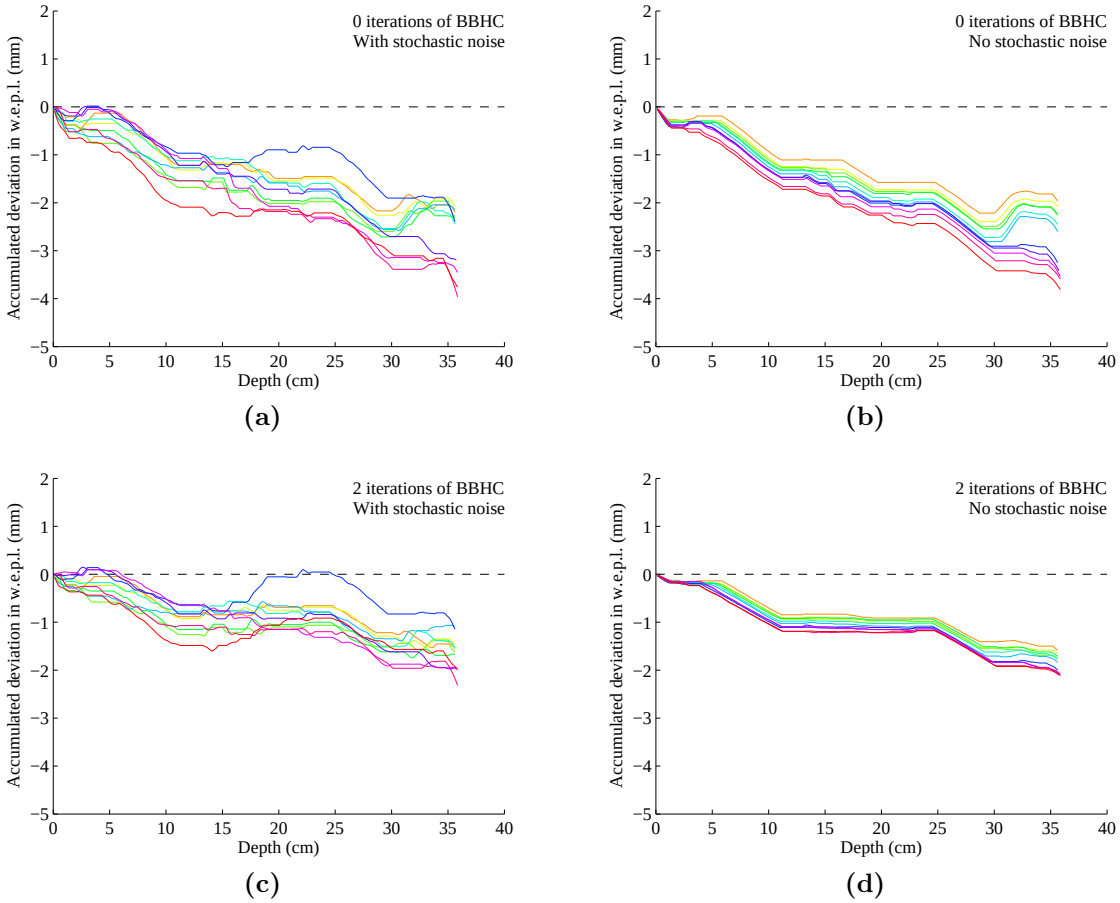


Figure 5.14 – a) Accumulation with depth of the deviation between CT-estimated and exactly-calculated water equivalent path length over 11 paths within the pelvic CT slice (number 116) shown in figure 5.13. b), c) and d) show the equivalent metric when the simulated CT scanner employs 2 iterations of BBHC and/or does not include stochastic noise.

Paths run parallel to the x axis, beginning at points shown by the rainbow arrow in figure 5.13. Regions in which no change in deviation is observed mostly correspond to boundary voxels which have been excluded from consideration.

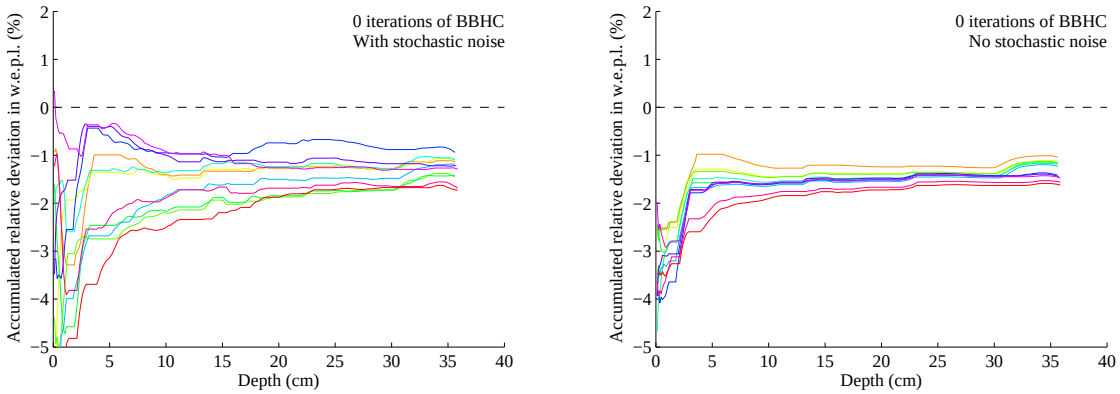


Figure 5.15 – Accumulation with depth of the water equivalent path length deviation along lines through the pelvic slice as shown in figure 5.14, but in relative (%) terms.

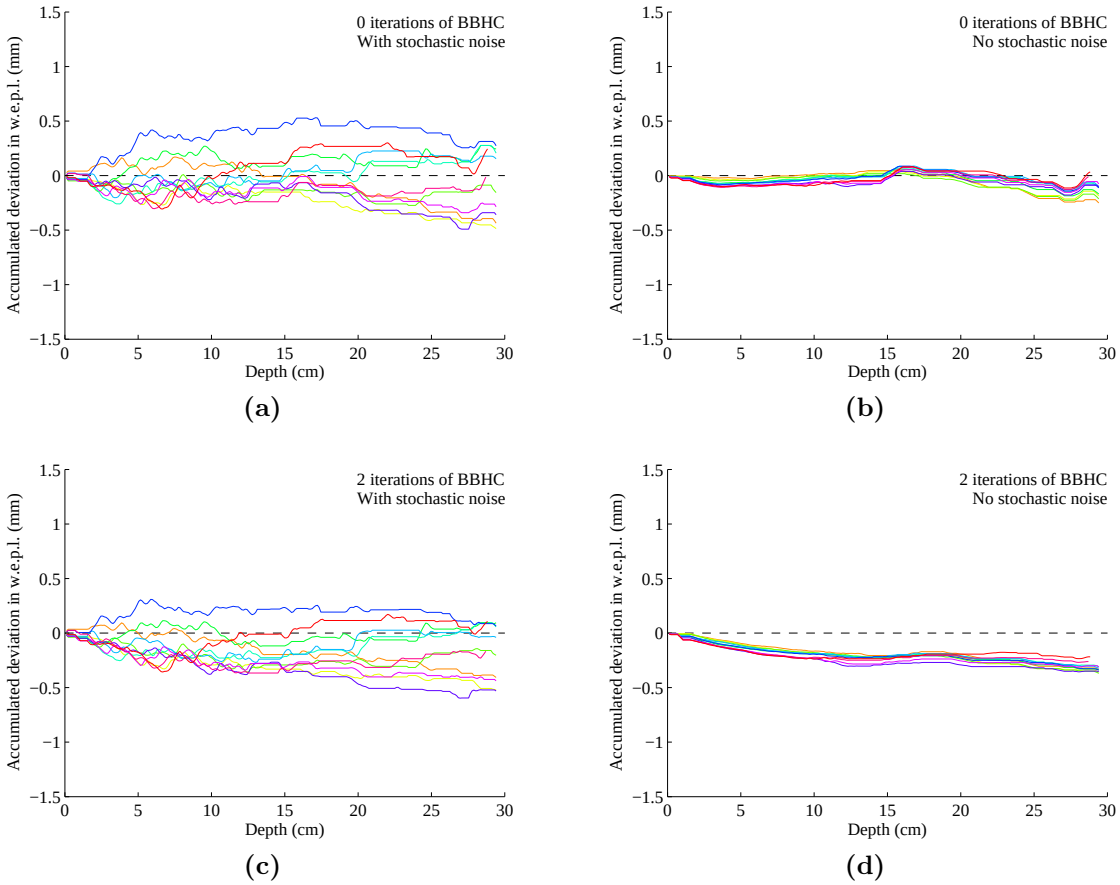


Figure 5.16 – Accumulation with depth of water equivalent path length deviation as in figure 5.14, but for the abdominal slice (number 155).

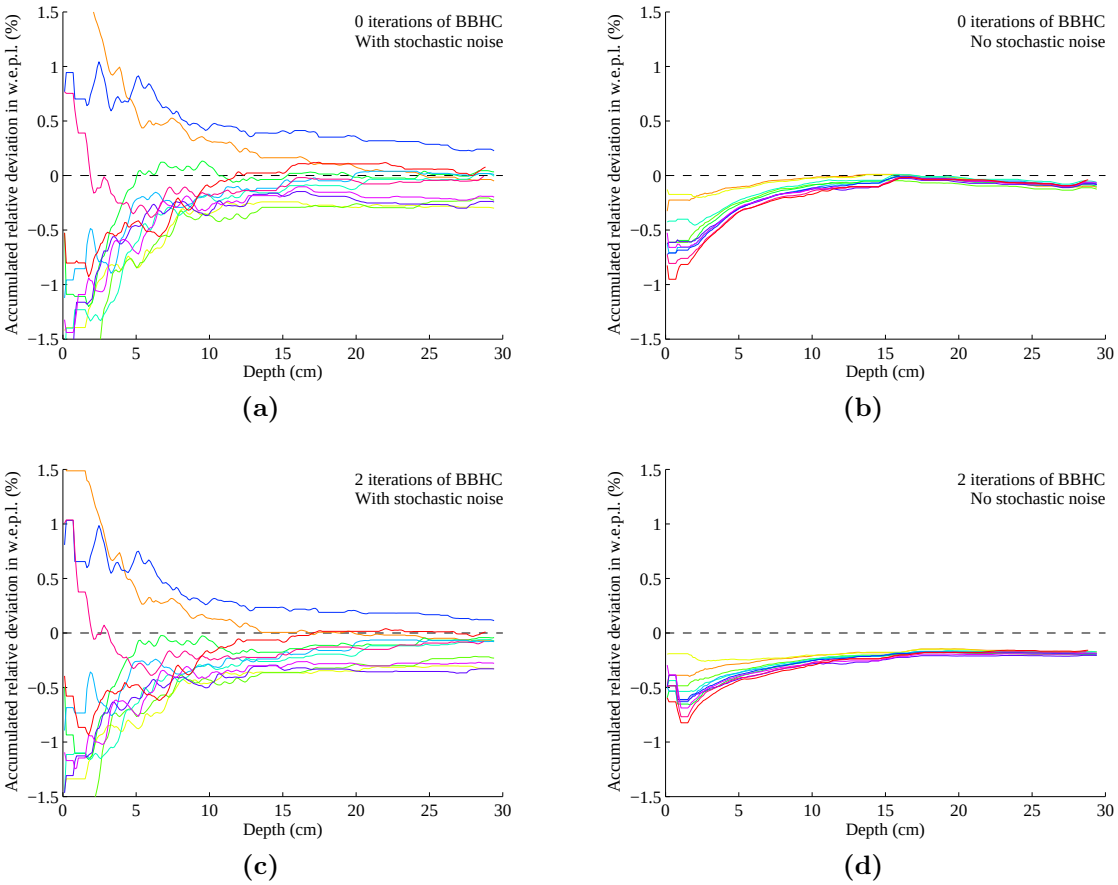


Figure 5.17 – Accumulation with depth of the water equivalent path length deviation along lines through the abdominal slice as shown in figure 5.16, but in relative (%) terms.

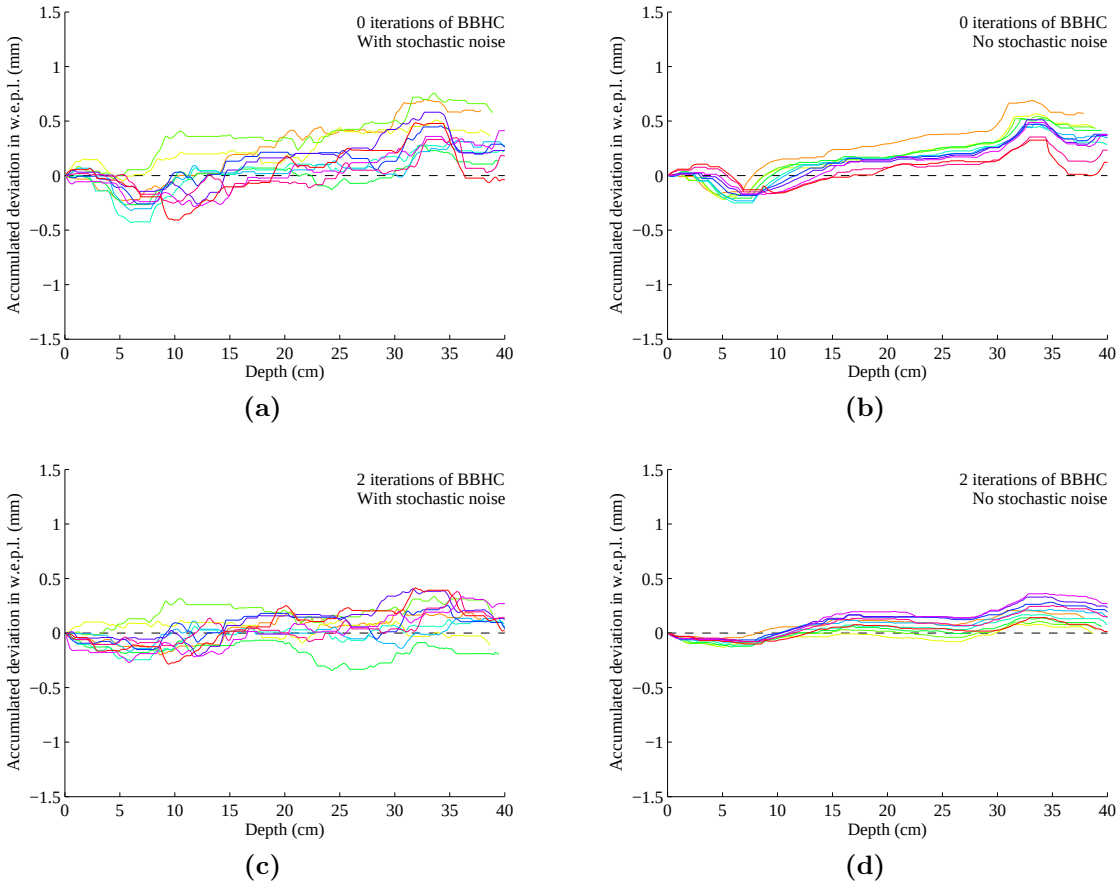


Figure 5.18 – Accumulation with depth of water equivalent path length deviation as in figure 5.14, but for the thoracic slice (number 170).

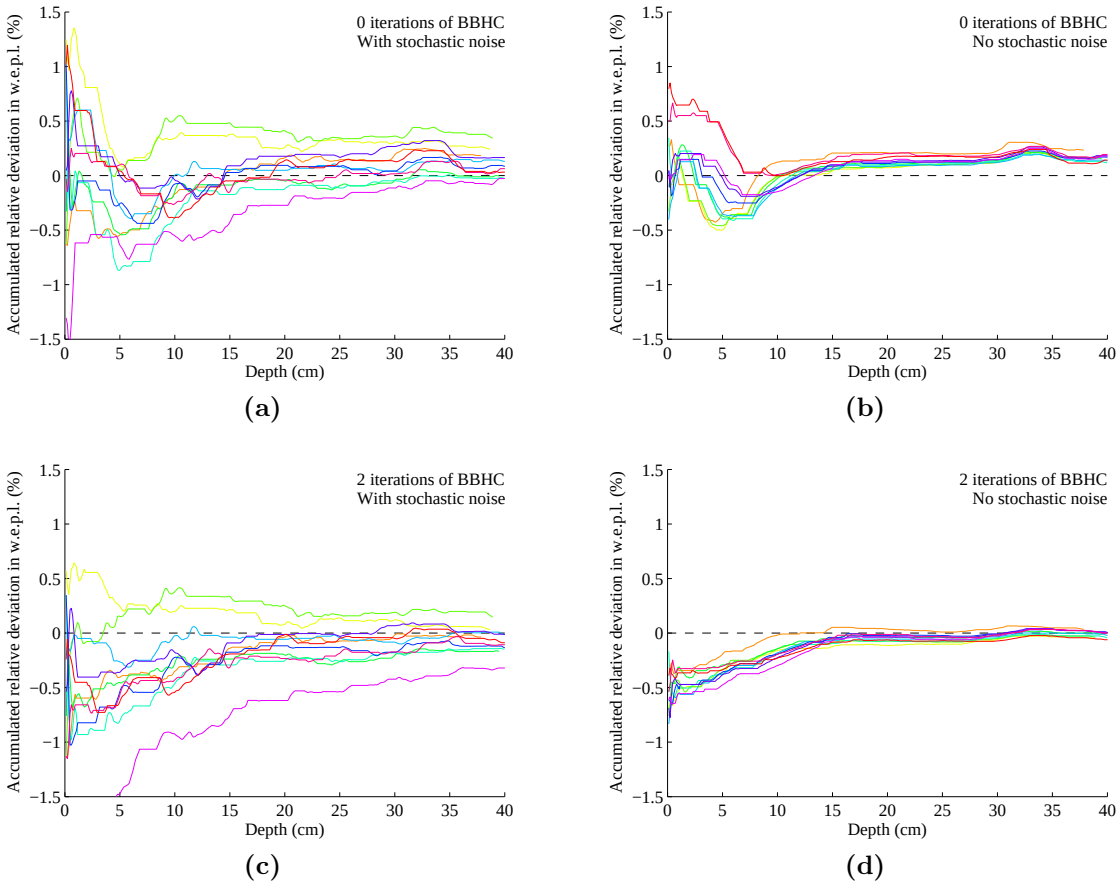


Figure 5.19 – Accumulation with depth of the water equivalent path length deviation through the thoracic slice as shown in figure 5.18, but in relative (%) terms.

number of CT scan-related parameters, leading to the conclusion that incorporating second-pass beam hardening correction can improve pixel-level consistency in CT numbers by a factor of 2. The compounded effect of mapping CT numbers to proton stopping power ratios was also investigated, in terms of the deviations between the estimated and true water equivalent path length over a number of paths. In general, the magnitude of 95% confidence intervals for measuring the water-equivalent path length through the ICRP phantom decreased from 5 mm to 3 mm over the range 80 kV_p to 140 kV_p (in the absence of extreme photon starvation), and improvements of 50-250% were seen as a result of using second-pass beam hardening correction. For paths of over 10 cm, typical uncertainties in estimation of water equivalent path lengths due to the calibration process at 120 kV_p amount to random variations of 1% and systematic shifts with anatomical location of up to 1.25%.

Chapter 6

Conclusions

6.1 Summary of results

The preceding Chapters have discussed a number of sources of uncertainty related CT-based treatment planning in proton radiotherapy. Chapter 2 presents a series of imaging artefacts that are inherent in traditional computed tomography, of which beam hardening, stochastic noise and aliasing/reconstruction-related noise have been examined in further detail using the software CT scanner developed and benchmarked in chapter 3.

The pencil beam dose calculation method was introduced at the beginning of chapter 4, leading to a discussion of which property of a material constitutes the most useful 'response' variable when calibrating pixel values of images from a CT scanner. Proton stopping power displays an energy dependence, but range predictions for 170 MeV protons in a heterogeneous geometry using a fixed high-energy SPR for each material showed less than 1 mm discrepancy when compared to the same predictions using an energy-dependent SPR. In contrast, a plausible scenario was demonstrated in which calibration using a clinically-adopted mass-density-based model in the lung may introduce range errors of multiple millimetres, even before CT artefacts are considered, and similar concerns could be raised for materials that are not biological in origin. This provides the rationale for recommending that calibrations utilise a high-energy SPR as the dependent variable, allowing the user to have confidence that the physical parameter they provide is being used directly in the pencil beam calculation.

The physical basis of the stoichiometric method of CT calibration was investigated, concluding that the power laws adopted hold only when the 'equivalence' of tissue equivalent relates to the chemical elements present as well as their x-ray attenuation properties. At 50 keV (close to the mean of an 80 kV_p spectrum), attempting stoichiometric calibration with a set of tissue equivalents containing significant amounts

of silicon led to errors in CT number prediction corresponding to greater than 2% of the attenuation of water – a factor of five larger than either of the sets of materials that contained biologically-present elements. The performance of 2- and 3-parameter formulations of the stoichiometric method were evaluated in polyenergetic situations and differences were found to be insignificant. Similarly, stochastic noise at a level equivalent to 150 mAs exposure did not affect the calibration accuracy in any meaningful way for any of the energy spectra or calibration geometries examined. The impact of second-pass beam hardening correction was more notable, with improvements of over 10 HU in the quality of CT number predictions when two iterations were performed. Surprisingly, the differences in the prediction quality between scanning tissue equivalents individually and using a multi-material phantom (with an appropriately-balanced arrangement of samples) were minor. Variations of 1% (average) to 5% (maximum) were seen between calibration curves specified using three different methods from the literature for the same set of input data.

The first Section of Chapter 5 investigates the impact of the CT artefacts previously discussed in the context of an anthropomorphic phantom (specified in ICRP Publication 110). Complete stacks of whole-body CT images were simulated for a range of scanner parameters. The largest x-ray attenuation variations in organs of an appreciable size were ±2% before second-pass beam hardening correction and ±1% after. Including the phantom's arms was seen to cause negligible changes in beam hardening, but a significant increase in photon starvation, and so it was suggested that the arms should be kept out of the scanning volume whenever possible. The combined impact of errors in calibration curve definition and CT artefacts were examined in the second section of Chapter 5, by calculating the error estimated water equivalent path lengths through the phantom for many possible paths. Here, the Schaffner approach to stoichiometric calibration was seen to be superior to those suggested by U. Schneider and W. Schneider. The width of the 95% confidence intervals in question decreased from 5 mm to 3 mm over the range 80 kVp to 140 kVp (in the absence of extreme photon starvation), and improvements of 50-250% were seen as a result of using second-pass beam hardening correction. Greater uncertainty was seen in regions containing more bone, and it was suggested that range uncertainties of around 2 mm may be expected in each of the beams typically used for proton radiotherapy of the prostate.

6.2 Comparisons with other results in the literature

A number of authors have previously offered assessments of the impact of CT number variations on uncertainties in proton therapy treatment planning. One of the most relevant pre-existing studies is the theoretical analysis of single- and dual-energy CT calibration methods carried out by Yang and colleagues at the M D Anderson Cancer Centre [93]. The virtual scanner constructed for their study considered none of the spatial aspects of image acquisition, consisting of energy spectrum calculation with SpekCalc followed by x-ray LAC determination using the XCOM database – both tools are used earlier in this thesis. The approach to beam hardening issues involved calculating the LAC averaged over the energy spectrum after 16 cm of water. Stoichiometric calibration was performed following U. Schneider's method and comparisons of SPRs were made using the Bethe-Bloch equation for tissues in ICRU 44 and White [94]. Maximum and RMS deviations of 3.24% and 0.89% were observed for the 34 tissues in question at 120 kV_p. Somewhat smaller values of 1.00% and 0.26% are found with their dual-energy method, which derives an electron density ratio and effective atomic number for each material from the two measured linear attenuation coefficients. The paper also includes an interesting sensitivity analysis, showing that variations in tissue composition and density (known to occur both within and between patients) do not necessarily result in responses of equivalent size in SPR values calculated by either the stoichiometric or Bethe-Bloch methods.

The work performed in this thesis differs from that of Yang et. al. in that there has been no assumption that a single CT number exists for a particular organ within the human body and with a specified x-ray energy spectrum. In reality, CT artefacts can cause complex spatial variations in observed CT numbers within materially-homogenous regions. It has been shown here that realistic simulations of those artefacts may be performed in silico, and that the results can be used to inform estimates of proton range uncertainty specific to a particular anatomical location. Such estimates are of direct relevance in the clinic, and they should be considered carefully when deciding upon the margins to be applied to the clinical treatment volume in order to generate the target volume.

In the same paper that introduced the stoichiometric method, U. Schneider et. al. assess the accuracy of its predictions against measurements of the radiographic path length of protons through a sheep's head fixed in water equivalent material [74]. They observed the deviations in integrated stopping power over a number

of paths to follow a Gaussian distribution, with the stoichiometric method seen to overestimate path lengths by 0.5% on average with a standard deviation of 1.4%. Calibration curves defined by direct measurements of CT numbers and SPRs of tissue equivalent materials, which were commonplace at the time, performed markedly worse. Schaffner and Pedroni studied the proton stopping power uncertainties introduced by their approach to calibration curve definition, through comparison of predicted and measured water-equivalent path lengths in ten animal tissues [75]. The standard deviation of CT numbers found in the partially-homogenised tissue samples ranged from 0.5-2%. Agreement between calculations and measurements was at a level better than 1%, although the authors raise the point that non-biological materials (including bolus materials common in radiotherapy) may show greater deviations. The conclusion of the work is that a maximum of 1.8% error may be expected in stopping power determination as a result of stoichiometric calibration and beam-hardening-related positional variations in CT numbers.

W. Schneider's version of the stoichiometric method was derived primarily for use with Monte Carlo simulations of photons and electrons [76], so the publication offers no error estimates relevant to proton therapy. However, a later paper from Massachusetts General Hospital [95] applies it to proton Monte Carlo treatment planning, contrasting the results with those of a simple assignment to one of four materials (air, lung, soft tissue and bone, as in the software BEAMnrc) by CT number thresholds alone, and a pencil beam method with a five-point calibration curve whose origin is unspecified. Comparisons were made between the doses observed when calculations were performed by each method for two clinical treatment plans, and dose differences of over 5% were seen in a non-negligible number of voxels within the target, with the W. Schneider results showing better agreement with pencil beam results than the BEAMnrc approach. No reference measurements were performed, however, so this paper simply highlights the dose variability that may be expected when the manner in which a material specification is determined from CT changes.

Pixel level variations in CT numbers and estimated SPR values of a similar magnitude to those found in the papers of U. Schneider and Schaffner have been found in this thesis. The size of uncertainties in the integrated water equivalent path length also agree with those reported by those authors for a typical clinical scan (120 kVp, with no second pass beam hardening correction). The additional benefit of the work performed in Chapters 4 and 5 is that the dependence of the uncertainty distribution on a large number of scan parameters has been investigated, including that of second pass beam hardening correction, which was been shown to result in appreciable

improvement in proton range estimates. The side-by-side comparison of the differing approaches of Schaffner, U. Schneider and W. Schneider to stoichiometric calibration in the same Chapters provides useful information on their relative performance for a wide range of scenarios, which it would otherwise be difficult to infer.

Stochastic noise was examined in a note by Chvetsov and Paige [96], which compared the proton range calculated by the continuous slowing down approximation along lines in water whose pixel values took Gaussian distributions with standard deviation 2.5%, 5% and 7.5%. The x-ray tube current and other scanner parameters that would lead to noise at these levels are not discussed[1], and the approach is clearly a simplification given the streaking nature of noise that is observed in heterogeneous geometries. Nevertheless, the resulting 95% confidence intervals for range calculation are of interest, amounting to 1-3% of the range at clinical energies in the case of 2.5% noise and a 3 mm calculation grid. The issue of calculation grid size has not been explicitly discussed in this thesis. As a concept which is related to the exposure, it is much more likely to affect the random, rather than systematic, portion of range uncertainties.

Lomax [97] carried out a study of IMPT treatment plan robustness in the presence of global changes in pixel values of $\pm 3\%$, presented as a realistic estimate of the combined uncertainty in stoichiometric calibration, beam hardening and reconstruction artefacts. Systematic dose differences of up to 5% were seen within the CTV for plans using the distal edge tracking method, whilst dose distributions for the same target volume lay largely unaffected by HU variation if fully-3D IMPT planning was performed. Doses to organs at risk were also affected at a level of approximately 2-4%, and again 3D planing was favoured. Dose-response curves can be steep for both tumour cure probability (TCP) and normal tissue complication probability (NTCP) and dose variations on this scale can easily increase or decrease risks by a factor of ten in certain circumstances. It is likely that dose differences resulting from CT artefacts discussed in this thesis will in general be more moderate than in the case investigated by Lomax, as most artefacts are well-localised. However, there will be locations at which the pixel value may differ by more than 3% from the ideal value, and these may also result in localised dose differences above those reported for large volumes.

[1] To observe stochastic noise with $\sigma = 25$ HU in a 30 cm water phantom at 120 kV_p, the simulated scanner of chapter 3 uses an x-ray exposure of only 10 mAs, much below typical values used in the clinic.

6.3 Proposals for further work

Stoichiometric calibration has been the focus of proton range uncertainty assessments in this work, due to its prevalence, but other methods for the estimation of proton stopping power exist which deserve similar attention. One such method is the polybinary method of Kanematsu [98], which claims a calibration accuracy of 1% using just four materials (water, air, ethanol and potassium phosphate solution). Of similar interest is a technique discussed by Yang et. al. [93] which uses dual-energy CT, providing two values (effective electron density and effective atomic number) for each pixel instead of a CT number. Because their dual-energy approach involves the fusion of two single-energy scans, the artefacts already discussed could manifest themselves in different ways, and a new class of artefacts may arise from the fusion process itself. Other approaches to dual-energy or spectral information may also be of interest, for example maximum likelihood methods of reconstruction [99] in which beam hardening is considered intrinsically. Megavoltage CT has been used to collect radiotherapy planning CT scans when a patient has metal implants, reducing beam hardening and stochasticity artefacts. Stoichiometric calibration is still possible in this case [100], although strictly speaking the pair-production cross section should be incorporated into the attenuation model, and an assessment of the accuracy of stopping power determination would be valuable. As higher energy accelerators become more common, proton tomography has the potential to provide routine physical measurements of the difference between planned and delivered proton ranges: a physical analogue of the values reported in Section 5.2.

In order to further develop the conclusions of this thesis, the dosimetric impact of calibration errors could be assessed. The simulated CT images and calibration curves created in earlier Chapters could be input into a clinical treatment planning system. Whilst no tumours are currently defined within the ICRP phantoms, likely target locations can be identified with sufficient experience of clinical treatment planning, and realistic beam arrangements can be applied. DICOM-RT contours created from the phantom data can aid this process. Once a treatment plan exists and a voxellised 'planned' dose distribution has been calculated, the same beam parameters may be applied to a phantom whose pixel values represent the true stopping power in order to produce the 'delivered' dose distribution. The difference between 'planned' and 'delivered' dose distributions will be an estimate of the typical errors that result in delivery of a treatment plan due to the CT calibration process: it can be quantified using clinical quality assurance metrics such as the gamma criterion [101].

The general approach presented in this thesis could be applied to a number of other artefacts that may occur in computed tomography. The code used to simulate CT has been further developed into a modular package that can perform simulations in three and four dimensions[2]. This package could be used together with the ICRP phantoms to assess the impact of artefacts that only occur in three dimensions, such as those which occur due to helical scanning and the use of conical beams of x-rays, as well as partial volume effects. With a four-dimensional phantom (such as 4D VIP-man [102]), the interplay between physiological motion (eg. breathing) and the movement of the x-ray source and detector could be examined. Assessment of the errors arising from movement is of particular importance in the treatment of lung cancer, although movements in the lung and bowel also cause changes in many other sites.

6.4 Final assessment

The typical sizes of planning margins currently used clinically at a number of proton therapy centres are given in table 6.1. These margins are added to account for setup uncertainty, CT number variation and stopping power calibration errors. Figure 6.1 shows all the range uncertainty distributions found at 120 $\mathrm{kV_p}$ in this work, which are attributable to the latter two sources of error. Whilst medical physicists should always carefully consider the potential effects of CT artefacts and stopping power calibration methods on their treatment plans, it can be seen that in only the rarest of cases examined here (<1%) are the planning margins used insufficient. This thesis therefore provides a strong basis for accepting current practice relating to the size of planning margins. It also provides evidence (as in figure 6.2) that the use of second-pass beam hardening correction and the highest tube voltages may allow a reduction of 33% or more in the size of these margins, thereby sparing healthy tissue from unnecessary irradiation.

[2]It should, however, be noted that each extra dimension adds further computational complexity to the raytracing procedure and requires more data to be stored in memory. In order to acquire the same number of images that were generated for this thesis using fully-3D or 4D simulations, significant effort would need to be made to improve the code's efficiency, perhaps through use of GPUs to trace hundreds of rays in parallel and/or programming languages that are lower-level than MATLAB.

Institution	Typical planning margin
Massachusetts General Hospital	3.5% of range + 1 mm
MD Anderson Proton Therapy Center	3.5% of range + 3 mm
Loma Linda University Medical Center	3.5% of range + 3 mm
Roberts Proton Therapy Center	3.5% of range + 3 mm
University of Florida Proton Therapy Institute	2.5% of range + 1.5 mm

Table 6.1 – Water-equivalent size of margins added to account for proton range uncertainty at a number of current proton therapy centres [103]

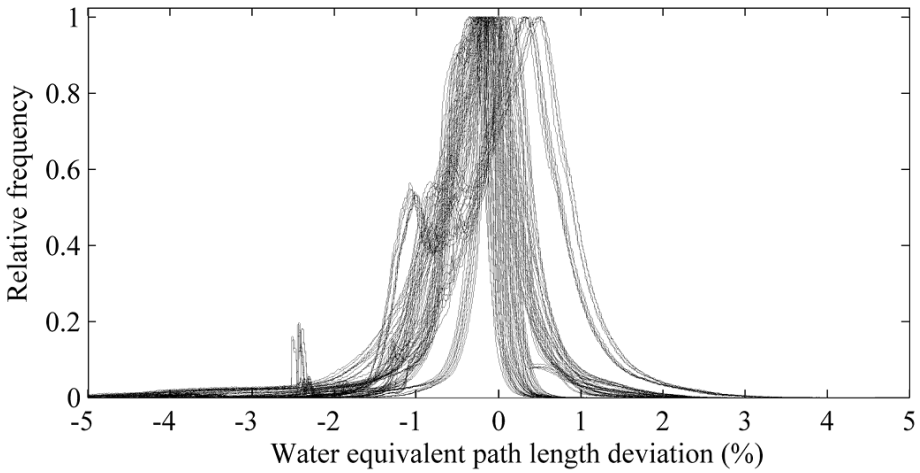


Figure 6.1 – Distributions of percentage deviation in water equivalent path length at 120 kV_p for every calibration curve and set of scan parameters tested in Chapter 5. The histograms plotted are relative versions of the absolute histograms of δ seen in Section 5.2.

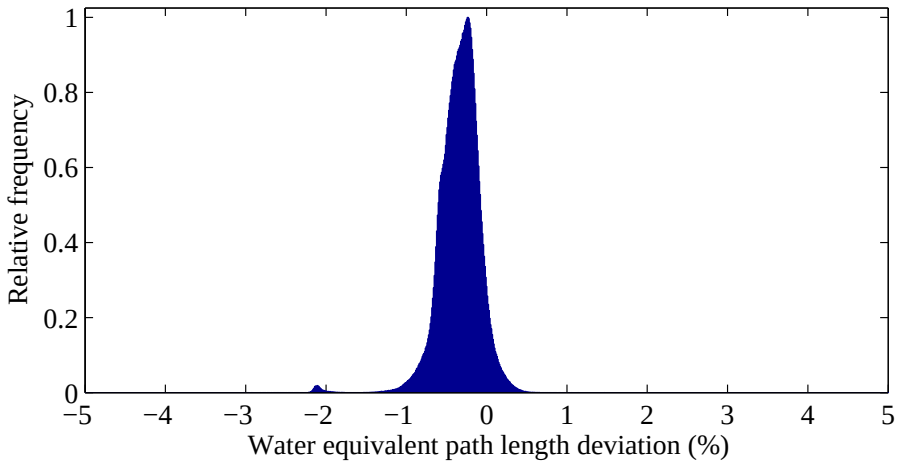


Figure 6.2 – Distribution of percentage deviation in water equivalent path length at 140 kV_p for the Schaffner approach to stoichiometric calibration with two iterations of second-pass (bone-based) beam hardening correction. Stochastic noise has been included. This is a relative version of the absolute histograms of δ seen in Section 5.2.

Appendix A

Radiation length calculations

The radiation lengths in Section 4.2.2.3 are calculated using a method described by Tsai [80]. Compound materials are considered as a mixture of their constituent elements, each contributing independently to the radiation length i.e. the possibility of molecular effects is neglected[1]. The radiation length of a material M, comprising elements labelled by i with weight fractionw_i, is given by equation (A.1).

$$X_M = \left(\sum_i \frac{w_i}{X_0(Z_i, A_i)} \right)^{-1} \tag{A.1}$$

Tsai calculates the elemental radiation length X_0 for atomic number Z and atomic mass A using equation equation (A.2), where the constant $k = \frac{m_e^2 c^2}{4\hbar^2 \alpha^3 N_A} \approx 716.407\,\mathrm{g} \cdot \mathrm{cm}^{-2}$.

$$X_0(Z, A) = \frac{k \cdot A}{Z^2 \cdot (L_{rad} - f) + Z \cdot L'_{rad}} \tag{A.2}$$

The terms L_{rad} and L'_{rad} are known as the radiation logarithms for interactions with the nuclear and electron fields respectively. As these fields vary with the atomic number, so do the radiation logarithms – the values adopted by Tsai are as equation (A.3) and equation (A.4).

[1]This assumption is not perfect: for example, in molecular H_2 a 2.4% reduction in radiation length has been observed compared to the free atom case, which agrees with calculations of the molecular screening effect within error [104].

$$L_{rad}(Z) = \begin{cases} 5.31 & Z = 1 \\ 4.79 & Z = 2 \\ 4.74 & Z = 3 \\ 4.71 & Z = 4 \\ \ln(184.15 \cdot Z^{-1/3}) & Z \geq 5 \end{cases} \tag{A.3}$$

$$L'_{rad}(Z) = \begin{cases} 6.144 & Z = 1 \\ 5.621 & Z = 2 \\ 5.805 & Z = 3 \\ 5.924 & Z = 4 \\ \ln(1194 \cdot Z^{-2/3}) & Z \geq 5 \end{cases} \tag{A.4}$$

f is the Coulomb correction to the radiation logarithm in the nuclear field, for which Tsai gives a polynomial expansion of an expression by Davies, Bethe and Maximon [105]. Modern numerical computing methods allow this to be calculated directly to high precision using equation (A.5), where ψ is the digamma fuction (the logarithmic derivative of Euler's gamma function). Tsai's expansion is accurate to 0.05% for all elements below uranium.

$$f(Z) = \alpha^2 Z^2 \cdot \sum_{n=1}^{\infty} \frac{1}{n(n^2 + \alpha^2 Z^2)} = \frac{1}{2}\left[\psi(1 - \mathrm{i}\alpha Z) + \psi(1 + \mathrm{i}\alpha Z) - 2\psi(1)\right] \tag{A.5}$$

Appendix B

CT numbers, stopping power ratios and radiation lengths for ICRU 44 and ICRP 110 tissues

ICRU Report 44 Tissues

Tissue	**Density**	**Stopping Power Ratio**				**CT Number (HU)**					X_0/X_w
	$(g \cdot cm^{-3})$	10 MeV	50 MeV	100 MeV	200 MeV	50 keV	75 keV	100 keV	125 keV	150 keV	
Air	0.001	0.001	0.001	0.001	0.001	-1000	-1000	-1000	-1000	-1000	0.001
Lung	0.260	0.258	0.258	0.258	0.258	-741	-742	-743	-743	-743	0.263
Adipose Tissue	0.950	0.971	0.967	0.965	0.963	-111	-74	-60	-57	-54	1.085
Yellow marrow	0.980	1.005	1.000	0.998	0.996	-88	-46	-31	-27	-23	1.133
Cell nucleus	1.000	0.992	0.993	0.993	0.993	23	5	-1	-3	-4	0.999
Breast	1.020	1.023	1.021	1.021	1.020	-19	1	8	10	12	1.093
Lymph	1.030	1.025	1.026	1.026	1.026	33	29	27	27	27	1.033
GI tract	1.030	1.025	1.025	1.025	1.025	22	23	24	24	24	1.050
Red marrow	1.030	1.034	1.032	1.031	1.030	-3	12	18	19	21	1.119
Testes	1.040	1.034	1.035	1.034	1.034	39	36	35	35	34	1.054
Brain	1.040	1.037	1.037	1.036	1.036	43	38	36	36	35	1.061
Pancreas	1.040	1.037	1.037	1.036	1.036	30	32	33	33	34	1.070
Ovaries	1.050	1.043	1.043	1.043	1.043	50	46	44	44	44	1.062
Thyroid	1.050	1.042	1.043	1.042	1.042	96	66	53	50	46	1.067
Kidneys	1.050	1.041	1.041	1.041	1.041	47	44	42	42	42	1.069
Muscle	1.050	1.040	1.041	1.040	1.040	47	43	41	41	41	1.071
Blood	1.060	1.049	1.050	1.049	1.049	64	56	53	52	51	1.074
Spleen	1.060	1.051	1.051	1.051	1.051	60	55	53	52	52	1.075
Heart	1.060	1.051	1.052	1.051	1.051	63	56	53	53	52	1.077
Liver	1.060	1.050	1.050	1.050	1.050	58	53	51	51	51	1.079
Eye lens	1.070	1.055	1.055	1.054	1.054	42	49	52	53	54	1.109
Skin	1.090	1.080	1.079	1.079	1.079	69	74	76	77	77	1.130
Cartilage	1.100	1.075	1.078	1.078	1.078	130	102	91	89	86	1.089
Spongiosa	1.180	1.137	1.141	1.141	1.141	488	283	207	189	169	1.133
Cortical bone	1.920	1.650	1.683	1.691	1.697	2592	1497	1088	995	889	1.436

ICRP Publication 110 Tissues

Tissue	Density	Stopping Power Ratio				CT Number (HU)					X_0/X_w
	($g \cdot cm^{-3}$)	10 MeV	50 MeV	100 MeV	200 MeV	50 keV	75 keV	100 keV	125 keV	150 keV	
Internal air	0.001	0.001	0.001	0.001	0.001	-1000	-1000	-1000	-1000	-1000	0.001
Lung tissue	0.382	0.378	0.379	0.379	0.379	-619	-621	-621	-622	-622	0.387
Residual tissue	0.950	0.971	0.967	0.965	0.963	-110	-74	-60	-57	-53	1.083
Breast (adipose tissue)	0.950	0.971	0.967	0.965	0.963	-110	-74	-60	-57	-53	1.083
Medullary cavity (humeri, ulnae, radii, femora, tibiae, fibulae, patellae)	0.980	1.005	1.000	0.998	0.996	-87	-46	-31	-27	-23	1.131
Breast (glandular tissue)	1.020	1.037	1.033	1.032	1.030	-37	-4	9	12	15	1.142
Lymphatic nodes	1.030	1.025	1.026	1.026	1.026	33	29	27	27	26	1.033
Oesophagus	1.030	1.026	1.025	1.025	1.024	16	20	21	21	22	1.069
Adrenals	1.030	1.026	1.025	1.025	1.024	15	19	21	21	22	1.070
Airway walls (nasal passages, trachea, bronchi)	1.030	1.026	1.025	1.025	1.024	14	19	21	21	22	1.073
Gall bladder (wall, contents)	1.030	1.026	1.025	1.025	1.024	14	19	21	21	22	1.073
Pituitary gland	1.030	1.026	1.025	1.025	1.024	14	19	21	21	22	1.073
Prostate	1.030	1.026	1.025	1.025	1.024	14	19	21	21	22	1.073
Salivary glands	1.030	1.026	1.025	1.025	1.024	14	19	21	21	22	1.073
Spinal cord	1.030	1.026	1.025	1.025	1.024	14	19	21	21	22	1.073
Thymus	1.030	1.026	1.025	1.025	1.024	14	19	21	21	22	1.073
Tonsils	1.030	1.026	1.025	1.025	1.024	14	19	21	21	22	1.073
Ureter	1.030	1.026	1.025	1.025	1.024	14	19	21	21	22	1.073
Uterus	1.030	1.026	1.025	1.025	1.024	14	19	21	21	22	1.073
Spongiosa (sacrum)	1.031	1.034	1.033	1.031	1.030	21	22	23	23	23	1.110
Urine	1.040	1.033	1.033	1.033	1.034	45	39	37	36	36	1.035
Bladder wall	1.040	1.033	1.033	1.033	1.033	44	38	35	34	34	1.051

ICRP Publication 110 Tissues

Tissue	Density	Stopping Power Ratio				CT Number (HU)					X_0/X_w
	($g \cdot cm^{-3}$)	10 MeV	50 MeV	100 MeV	200 MeV	50 keV	75 keV	100 keV	125 keV	150 keV	
Spleen	1.040	1.029	1.030	1.029	1.030	45	36	33	32	31	1.053
Ovaries	1.040	1.035	1.035	1.034	1.034	39	36	35	35	34	1.055
Testes	1.040	1.035	1.035	1.034	1.034	39	36	35	35	34	1.055
Thyroid	1.040	1.032	1.033	1.032	1.032	86	55	43	40	36	1.056
GI tract walls (stomach, intestines, rectum)	1.040	1.034	1.034	1.034	1.034	31	32	33	33	33	1.060
GI tract contents	1.040	1.030	1.030	1.029	1.029	31	29	29	29	28	1.075
Spongiosa (sternum)	1.041	1.042	1.041	1.039	1.038	45	37	34	34	33	1.112
Liver	1.050	1.040	1.041	1.040	1.041	48	43	42	41	41	1.068
Kidneys	1.050	1.041	1.041	1.041	1.041	48	44	42	42	42	1.068
Oral mucosa (outer tongue, lips, cheeks)	1.050	1.040	1.041	1.040	1.040	47	43	41	41	41	1.071
Muscle	1.050	1.040	1.041	1.040	1.040	47	43	41	41	41	1.071
Tongue (inner part)	1.050	1.040	1.041	1.040	1.039	47	43	41	41	41	1.071
Brain	1.050	1.047	1.047	1.046	1.046	53	48	46	46	45	1.071
Heart wall	1.050	1.043	1.043	1.043	1.043	46	44	43	43	42	1.072
Pancreas	1.050	1.045	1.045	1.044	1.044	41	42	43	43	43	1.077
Eyes	1.050	1.036	1.036	1.036	1.036	25	31	34	34	35	1.085
Spongiosa (cervical spine)	1.050	1.048	1.047	1.046	1.045	67	51	45	43	42	1.114
Blood	1.060	1.049	1.050	1.049	1.049	64	56	53	52	51	1.074
Spongiosa (thoracic spine)	1.074	1.064	1.064	1.063	1.062	130	88	72	68	64	1.116
Skin	1.090	1.080	1.079	1.079	1.079	70	75	77	77	77	1.128
Cartilage	1.100	1.075	1.078	1.078	1.078	130	102	91	89	86	1.089

ICRP Publication 110 Tissues

Tissue	Density	Stopping Power Ratio				CT Number (HU)					X_0/X_w
	$(g \cdot cm^{-3})$	10 MeV	50 MeV	100 MeV	200 MeV	50 keV	75 keV	100 keV	125 keV	150 keV	
Spongiosa (lower humeri, ulnae radii, wrists, hand bones, lower femora, tibiae, fibulae, patellae, ankles, foot bones)	1.108	1.093	1.094	1.093	1.092	232	147	115	108	100	1.134
Spongiosa (lumbar spine)	1.112	1.092	1.093	1.093	1.092	221	144	115	109	101	1.124
Spongiosa(pelvis)	1.123	1.100	1.102	1.102	1.101	258	165	130	122	113	1.129
Spongiosa (upper femora)	1.124	1.102	1.103	1.103	1.103	264	167	131	123	114	1.133
Spongiosa (clavicles)	1.151	1.121	1.123	1.123	1.123	332	208	162	152	140	1.137
Spongiosa (cranium)	1.157	1.124	1.127	1.127	1.127	350	218	169	158	146	1.136
Spongiosa (ribs)	1.165	1.129	1.133	1.132	1.133	362	227	177	166	153	1.134
Spongiosa (scapulae)	1.183	1.142	1.146	1.146	1.146	416	258	199	186	170	1.142
Spongiosa (upper humeri)	1.205	1.158	1.163	1.163	1.164	470	291	224	209	192	1.148
Spongiosa (mandible)	1.228	1.175	1.180	1.181	1.181	531	327	251	234	214	1.154
Cortical bone	1.920	1.660	1.691	1.699	1.705	2491	1459	1074	986	887	1.460
Teeth	2.750	2.279	2.338	2.353	2.365	4881	2849	2090	1917	1721	1.892

Bibliography

[1] Cancer Research UK, 2011. Cancer risk
http://www.cancerresearchuk.org/cancer-info/cancerstats/incidence/risk

[2] Cancer Research UK, 2012. Cancer incidence for all cancers combined
http://www.cancerresearchuk.org/cancer-info/cancerstats/incidence/all-cancers-combined/

[3] Department of Health, 2012. Radiotherapy Services in England 2012
https://www.wp.dh.gov.uk/publications/files/2012/11/Radiotherapy-Services-in-England-2012.pdf

[4] M. V. Williams and K. J. Drinkwater, 2009. Radiotherapy in England in 2007: modelled demand and audited activity. *Clinical oncology (Royal College of Radiologists (Great Britain))* **21**(8) 575–90. doi:10.1016/j.clon.2009.07.003

[5] International Commission on Radiation Units and Measurements, 2011. Dosimetry. *Journal of the ICRU* **11**(1) 23–28. doi:10.1093/jicru/ndr003

[6] R. K. Sachs and D. J. Brenner, 1998. The mechanistic basis of the linear-quadratic formalism. *Medical Physics* **25**(10) 2071–2073

[7] E. J. Hall and A. J. Giaccia, 2006. Time, Dose and Fractionation in Radiotherapy. In *Radiobiology for the radiologist*, Lippincott Williams & Wilkins, chapter 22, 378–397. ISBN 0781741513

[8] P. Mayles, A. E. Nahum and J.-C. Rosenwald, 2007. *Handbook of radiotherapy physics: theory and practice*. Taylor & Francis, London. ISBN 9780750308601

[9] B. Gottschalk, 2012. Proton Therapy Physics. In H. Paganetti (editor) *Proton and Charged Particle Therapy*, CRC Press, Boca Raton, FL, USA, chapter 2, 19–59

[10] R. F. Mould, 2011. X-rays in 1896-1897. *NOWOTWORY Journal of Oncology* **61**(6) 100e–109e

[11] R. R. Wilson, 1946. Radiological use of fast protons. *Radiology* **47**(5) 487–91

[12] J. H. Lawrence, 1957. Proton irradiation of the pituitary. *Cancer* **10**(4) 795–798. doi:10.1002/1097-0142(195707/08)10:4<795::AID-CNCR2820100426>3.0.CO;2-B

[13] National Specialised Commissioning Team, 2011. Guidance for the Referral of Patients Abroad for NHS Proton Treatment - Version 2.3 `http://www.specialisedservices.nhs.uk/library/23/Guidance_for_referral_of_patients_abroad_for_NHS_Proton.pdf`

[14] Particle Therapy Co-Operative Group (PTCOG). Particle therapy facilities in operation (incl. patient statistics) `http://ptcog.web.psi.ch/ptcentres.html`

[15] J. Meesungnoen, J.-P. Jay-Gerin et al., 2002. Low-Energy Electron Penetration Range in Liquid Water. *Radiation Research* **158**(5) 657 – 660

[16] 2007. *ICRU Report 78: Prescribing, Recording, and Reporting Proton-Beam Therapy. Technical Report 2*, International Commission on Radiation Units and Measurements, Bethesda, Maryland

[17] R. A. Britten, V. Nazaryan et al., 2013. Variations in the RBE for cell killing along the depth-dose profile of a modulated proton therapy beam. *Radiation research* **179**(1) 21–8. doi:10.1667/RR2737.1

[18] P. Bettega, P. Calzolari and D. Chauvel, 2000. Radiobiological studies on the 65MeV therapeutic proton beam at Nice using human tumour cells. *International Journal of Radiation Biology* **76**(10) 1297–1303. doi:10.1080/09553000050151565

[19] A. Courdi, N. Brassart et al., 1994. The depth-dependent radiation response of human melanoma cells exposed to 65 MeV protons. *British Journal of Radiology* **67**(800) 800–804. doi:10.1259/0007-1285-67-800-800

[20] A. Carabe-Fernandez, R. G. Dale and B. Jones, 2007. The incorporation of the concept of minimum RBE (RBE min) into the linear-quadratic model and the potential for improved radiobiological analysis of high-LET treatments. *International Journal of Radiation Biology* **83**(1) 27–39. doi:10.1080/09553000601087176

[21] M. C. Frese, J. J. Wilkens et al., 2011. Application of constant vs. variable relative biological effectiveness in treatment planning of intensity-modulated proton therapy. *International journal of radiation oncology, biology, physics* **79**(1) 80–8. doi:10.1016/j.ijrobp.2009.10.022

[22] B. Gottschalk, 2008. Treatment Delivery Systems. In T. F. Delaney and H. M. Kooy (editors) *Proton and Charged Particle Therapy*, Lippincott Williams & Wilkins, Philadelphia, PA, USA, chapter 5, 33–40

[23] E. Pedroni, 1995. The 200-MeV proton therapy project at the Paul Scherrer Institute: Conceptual design and practical realization. *Medical Physics* **22**(1) 37. doi:10.1118/1.597522

[24] P. Park, X. Zhu et al., 2010. WE-A-BRA-02: A Novel Beam-Specific PTV Design to Account for Setup Error and Range Uncertainties for Scanning Beam Proton Therapy. *Medical Physics* **37**(6) 3409. doi:10.1118/1.3469323

[25] J. Hubbell and S. Seltzer, 2004. Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients
`http://physics.nist.gov/xaamdi`

[26] R. A. Rutherford, B. R. Pullan and I. Isherwood, 1976. Measurement of effective atomic number and electron density using an EMI scanner. *Neuroradiology* **11**(1) 15–21. doi:10.1007/BF00327253

[27] M. J. Berger, J. H. Hubbell et al., 1998. XCOM: Photon Cross Sections Database [Version 3.1]
`http://www.nist.gov/pml/data/xcom/index.cfm`

[28] A. C. Kak and M. Slaney, 2001. *Principles of computerized tomographic imaging*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA. ISBN 9780898714944

[29] 2011. *Digital Imaging and Communications in Medicine (DICOM). Technical report*, National Electrical Manufacturers Association, Rosslyn, Virginia, USA

[30] G. N. Hounsfield, 1973. Computerized transverse axial scanning (tomography): Part 1. Description of system. *British Journal of Radiology* **46**(552) 1016–1022. doi:10.1259/0007-1285-46-552-1016

[31] G. Poludniowski, G. Landry et al., 2009. SpekCalc: a program to calculate photon spectra from tungsten anode x-ray tubes. *Physics in medicine and biology* **54**(19) N433–8. doi:10.1088/0031-9155/54/19/N01

[32] J. F. Barrett and N. Keat, 2004. Artifacts in CT: recognition and avoidance. *Radiographics : a review publication of the Radiological Society of North America, Inc* **24**(6) 1679–91. doi:10.1148/rg.246045065

[33] G. H. Glover, 1982. Compton scatter effects in CT reconstructions. *Medical Physics* **9**(6) 860. doi:10.1118/1.595197

[34] X. G. Xu and K. F. Eckerman (editors) 2009. *Handbook of Anatomical Models for Radiation Dosimetry*. Taylor & Francis. ISBN 9781420059793

[35] 2005. *Reassessment of the Atomic Bomb Radiation Dosimetry for Hiroshima and Nagasaki - Dosimetry System 2002 (DS02). Technical report*, Radiation Effects Research Foundation, Hiroshima, Japan

[36] X. G. Xu, T. C. Chao and A. Bozkurt, 2000. VIP-Man: an image-based whole-body adult male model constructed from color photographs of the Visible Human Project for multi-particle Monte Carlo calculations. *Health physics* **78**(5) 476–86

[37] L. Casalini and G. Gualdrini, 1998. Use of the SABRINA code for MCNP geometry checking and displaying of the particle tracks. In *Use of MCNP in Radiation Protection and Dosimetry*, ENEA, Rome, 311–330. ISBN 88-8286-000-1

[38] ICRP, 2009. *Adult Reference Computational Phantoms. ICRP Publication 110. Technical Report 2*, International Commission on Radiological Protection, Ottawa, Ontario

[39] C. Lee, D. Lodwick et al., 2010. The UF family of reference hybrid phantoms for computational radiation dosimetry. *Physics in medicine and biology* **55**(2) 339–63. doi:10.1088/0031-9155/55/2/002

[40] K. M. Rosenberg. CTSim
`http://www.ctsim.org/`

[41] D. Farrell. learnCT
`http://homepage.ntlworld.com/jfarrell/javactapp/`

[42] J. Tabary, S. Marache-Francisco et al., 2009. Realistic X-ray CT simulation of the XCAT phantom with SINDBAD. In *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*. IEEE, 3980–3983. ISBN 978-1-4244-3961-4. doi:10.1109/NSSMIC.2009.5401942

[43] K. Viswanathan. XRaySim
`http://xraysim.sourceforge.net/index.htm`

[44] J. Klukowska, R. Davidi and G. T. Herman, 2013. SNARK09 - A software package for reconstruction of 2D images from 1D projections. *Computer methods and programs in biomedicine* **110**(3) 424–40. doi:10.1016/j.cmpb.2013.01.003

[45] G. T. Herman and S. S. Trivedi, 1983. A comparative study of two postreconstruction beam hardening correction methods. *IEEE transactions on medical imaging* **2**(3) 128–35. doi:10.1109/TMI.1983.4307626

[46] J. Fessler. Image reconstruction toolbox
`http://web.eecs.umich.edu/%7Efessler/code/`

[47] J. Zhu, S. Zhao et al., 2005. Computed tomography simulation with superquadrics. *Medical Physics* **32**(10) 3136. doi:10.1118/1.2040727

[48] W. P. Segars, M. Mahesh et al., 2008. Realistic CT simulation using the 4D XCAT phantom. *Medical Physics* **35**(8) 3800. doi:10.1118/1.2955743

[49] B. De Man, S. Basu et al., 2007. CatSim: a new computer assisted tomography simulation environment. 65102G–65102G–8. doi:10.1117/12.710713

[50] The MathWorks Inc., 2012. MATLAB R2012a

[51] D. F. Jackson, 1982. Chemical effects in X-ray transmission measurements. *Nuclear Instruments and Methods in Physics Research* **193**(1-2) 387–389. doi:10.1016/0029-554X(82)90729-7

[52] M. Buehren, 2011. Multicore - Parallel processing on multiple cores
`http://www.mathworks.co.uk/matlabcentral/fileexchange/13775`

[53] G. T. Herman, 1979. Correction for beam hardening in computed tomography. *Physics in Medicine and Biology* **24**(1) 81–106. doi:10.1088/0031-9155/24/1/008

[54] G. Besson, 1996. CT fan-beam parametrizations leading to shift-invariant filtering. *Inverse Problems* **12**(6) 815–833. doi:10.1088/0266-5611/12/6/002

[55] A. C. Kak and M. Slaney, 2002. Errata for Principles of Computerized Tomographic Imaging
`https://engineering.purdue.edu/%7Emalcolm/pct/pct-errata.html`

[56] M. Bertero and P. Boccacci, 1998. *Introduction to Inverse Problems in Imaging.* Institute of Physics Publishing, Bristol, UK

[57] 2004. *IEC TR 60788: Medical electrical equipment - Glossary of defined terms. Technical report*, International Electrotechnical Commission, Geneva, Switzerland

[58] G. G. Poludniowski, 2007. Calculation of x-ray spectra emerging from an x-ray tube. Part II. X-ray production and filtration in x-ray targets. *Medical Physics* **34**(6) 2175. doi:10.1118/1.2734726

[59] GE Healthcare, 2005. Lightspeed 2.X Technical Reference Manual
`http://apps.gehealthcare.com/servlet/ClientServlet?REQ=RAA&DIRECTION=2281207-100&FILENAME=2281207-100r7.pdf&FILEREV=7&DOCREV_ORG=7`

[60] ImPACT, 2004. *Four slice CT scanner comparison report version 11. Technical Report June*

[61] J. B. Smathers, 1977. Composition of A-150 tissue-equivalent plastic. *Medical Physics* **4**(1) 74. doi:10.1118/1.594380

[62] G. Poludniowski, P. M. Evans et al., 2011. Removal and effects of scatter-glare in cone-beam CT with an amorphous-silicon flat-panel detector. *Physics in medicine and biology* **56**(6) 1837–51. doi:10.1088/0031-9155/56/6/019

[63] Y. Watanabe and C. Constantinou, 2006. Phantom Materials in Radiology. In J. G. Webster (editor) *Encyclopedia of medical devices and instrumentation Vol. 5*, Wiley-Interscience, Hoboken, N.J., USA, 252–269. 2nd edition. ISBN 9780470040706

[64] 1989. *ICRU Report 44: Tissue Substitutes in Radiation Dosimetry and Measurement. Technical report*, International Commission on Radiation Units and Measurements, Bethesda, Maryland

[65] N. Wellock, 2012. Personal communication

[66] L. Hong, M. Goitein et al., 1996. A pencil beam algorithm for proton dose calculations. *Physics in Medicine and Biology* **41**(8) 1305–1330. doi:10.1088/0031-9155/41/8/005

[67] M. Soukup, M. Fippel and M. Alber, 2005. A pencil beam algorithm for intensity modulated proton therapy derived from Monte Carlo simulations. *Physics in medicine and biology* **50**(21) 5089–104. doi:10.1088/0031-9155/50/21/010

[68] K. R. Hogstrom, M. D. Mills and P. R. Almond, 1981. Electron beam dose calculations. *Physics in medicine and biology* **26**(3) 445–59

[69] T. Bortfeld, 1997. An analytical approximation of the Bragg curve for therapeutic proton beams. *Medical physics* **24**(12) 2024–33

[70] H. Paganetti, 2009. Dose to water versus dose to medium in proton beam therapy. *Physics in medicine and biology* **54**(14) 4399–421. doi:10.1088/0031-9155/54/14/004

[71] B. Gottschalk, A. Koehler et al., 1993. Multiple Coulomb scattering of 160 MeV protons. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **74**(4) 467–490. doi:10.1016/0168-583X(93)95944-Z

[72] B. Rossi and K. Greisen, 1941. Cosmic-Ray Theory. *Reviews of Modern Physics* **13**(4) 240–309. doi:10.1103/RevModPhys.13.240

[73] P. J. Mohr, B. N. Taylor and D. B. Newell, 2012. CODATA Recommended Values of the Fundamental Physical Constants: 2010. *Journal of Physical and Chemical Reference Data* **41**(4) 043109. doi:10.1063/1.4724320

[74] U. Schneider, E. Pedroni and A. Lomax, 1996. The calibration of CT Hounsfield units for radiotherapy treatment planning. *Physics in Medicine and Biology* **41**(1) 111–124. doi:10.1088/0031-9155/41/1/009

[75] B. Schaffner and E. Pedroni, 1998. The precision of proton range calculations in proton radiotherapy treatment planning: experimental verification of the relation between CT-HU and proton stopping power. *Physics in Medicine and Biology* **43**(6) 1579–1592. doi:10.1088/0031-9155/43/6/016

[76] W. Schneider, T. Bortfeld and W. Schlegel, 2000. Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions. *Physics in Medicine and Biology* **45**(2) 459–478. doi:10.1088/0031-9155/45/2/314

[77] M. Fippel and M. Soukup, 2004. A Monte Carlo dose calculation algorithm for proton therapy. *Medical Physics* **31**(8) 2263. doi:10.1118/1.1769631

[78] National Institute of Standards and Technology. Stopping-power and range tables for protons
`http://physics.nist.gov/PhysRefData/Star/Text/PSTAR.html`

[79] 1992. *ICRU Report 46: Photon, Electron, Proton and Neutron Interaction Data for Body Tissues. Technical report*, International Commission on Radiation Units and Measurements, Bethesda, Maryland

[80] Y.-S. Tsai, 1974. Pair production and bremsstrahlung of charged leptons. *Reviews of Modern Physics* **46**(4) 815–851. doi:10.1103/RevModPhys.46.815

[81] J. F. Ziegler, M. Ziegler and J. Biersack, 2010. SRIM - The stopping and range of ions in matter (2010). *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **268**(11-12) 1818–1823. doi:10.1016/j.nimb.2010.02.091

[82] D. B. Pelowitz, J. W. Durkee et al., 2011. MCNPX 2.7.0 Extensions
`http://mcnpx.lanl.gov/opendocs/versions/v270/v270.pdf`

[83] M. B. Chadwick, P. G. Young et al., 1999. Cross-Section Evaluations to 150 MeV for Accelerator-Driven Systems and Implementation in MCNPX. *Nuclear Science and Engineering* **131**(3) 293–328

[84] H. Watz, A. Breithecker et al., 2005. Micro-CT of the human lung: imaging of alveoli and virtual endoscopy of an alveolar duct in a normal lung and in a lung with centrilobular emphysema–initial observations. *Radiology* **236**(3) 1053–8. doi:10.1148/radiol.2363041142

[85] E. Roelofs, M. Engelsman et al., 2012. Results of a multicentric in silico clinical trial (ROCOCO): comparing radiotherapy with photons and protons for non-small cell lung cancer. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer* **7**(1) 165–76. doi:10.1097/JTO.0b013e31823529fc

[86] J. Y. Chang, X. Zhang et al., 2006. Significant reduction of normal tissue dose by proton radiotherapy compared with three-dimensional conformal or intensity-modulated radiation therapy in Stage I or Stage III non-small-cell lung cancer. *International journal of radiation oncology, biology, physics* **65**(4) 1087–96. doi:10.1016/j.ijrobp.2006.01.052

[87] D. F. Jackson and D. Hawkes, 1981. X-ray attenuation coefficients of elements and mixtures. *Physics Reports* **70**(3) 169–233. doi:10.1016/0370-1573(81)90014-4

[88] W. Veigele, 1973. Photon cross sections from 0.1 keV to 1 MeV for elements Z = 1 to Z = 94. *Atomic Data and Nuclear Data Tables* **5**(1) 51–111. doi: 10.1016/S0092-640X(73)80015-4

[89] Gammex Inc. Gammex Products Catalog `http://www.gammex.com/ace-files/Gammex_Catalog.pdf`

[90] ICRP, 1975. *Report on the Task Group on Reference Man. ICRP Publication 23. Technical report*, International Commission on Radiological Protection, Ottawa, Ontario

[91] H. Q. Woodard and D. R. White, 1986. The composition of body tissues. *British Journal of Radiology* **59**(708) 1209–1218. doi:10.1259/0007-1285-59-708-1209

[92] J. Meyer, J. Bluett et al., 2010. Spot scanning proton beam therapy for prostate cancer: treatment planning technique and analysis of consequences of rotational and translational alignment errors. *International journal of radiation oncology, biology, physics* **78**(2) 428–34. doi:z

[93] M. Yang, G. Virshup et al., 2010. Theoretical variance analysis of single- and dual-energy computed tomography methods for calculating proton stopping power ratios of biological tissues. *Physics in medicine and biology* **55**(5) 1343–62. doi:10.1088/0031-9155/55/5/006

[94] D. R. White, H. Q. Woodard and S. M. Hammond, 1987. Average soft-tissue and bone models for use in radiation dosimetry. *British Journal of Radiology* **60**(717) 907–913. doi:10.1259/0007-1285-60-717-907

[95] H. Jiang, J. Seco and H. Paganetti, 2007. Effects of Hounsfield number conversion on CT based proton Monte Carlo dose calculations. *Medical Physics* **34**(4) 1439. doi:10.1118/1.2715481

[96] A. V. Chvetsov and S. L. Paige, 2010. The influence of CT image noise on proton range calculation in radiotherapy planning. *Physics in medicine and biology* **55**(6) N141–9. doi:10.1088/0031-9155/55/6/N01

[97] A. J. Lomax, 2008. Intensity modulated proton therapy and its sensitivity to treatment uncertainties 1: the potential effects of calculational uncertainties. *Physics in medicine and biology* **53**(4) 1027–42. doi:10.1088/0031-9155/53/4/014

[98] N. Kanematsu, N. Matsufuji et al., 2003. A CT calibration method based on the polybinary tissue model for radiotherapy treatment planning. *Physics in Medicine and Biology* **48**(8) 1053–1064. doi:10.1088/0031-9155/48/8/307

[99] J. A. Fessler, I. A. Elbakri et al., 2002. Maximum-likelihood dual-energy tomographic image reconstruction. In M. Sonka and J. M. Fitzpatrick (editors) *Medical Imaging 2002: Image Processing*. SPIE, San Diego, CA, USA, 38–49. doi:10.1117/12.467189

[100] W. Newhauser, J. Fontenot et al., 2007. SU-FF-T-25: A Monte-Carlo Based Dose Engine for Proton Radiotherapy Treatment Planning. *Medical Physics* **34**(6) 2406. doi:10.1118/1.2760670

[101] D. A. Low, W. B. Harms et al., 1998. A technique for the quantitative evaluation of dose distributions. *Medical physics* **25**(5) 656–61

[102] X. G. Xu and C. Shi, 2005. Preliminary Development Of A 4D Anatomical Model For Monte Carlo Simulations. In *American Nuclear Society. The Monte Carlo Method: Versatility Unbounded In A Dynamic Computing World.* Chattanooga, TN, USA

[103] H. Paganetti, 2012. Range uncertainties in proton therapy and the role of Monte Carlo simulations. *Physics in medicine and biology* **57**(11) R99–117. doi:10.1088/0031-9155/57/11/R99

[104] D. Bernstein and W. Panofsky, 1956. Bremsstrahlung Yield of High-Energy Electrons in Hydrogen. *Physical Review* **102**(2) 522–527. doi:10.1103/PhysRev.102.522

[105] H. Davies, H. Bethe and L. Maximon, 1954. Theory of Bremsstrahlung and Pair Production. II. Integral Cross Section for Pair Production. *Physical Review* **93**(4) 788–795. doi:10.1103/PhysRev.93.788