ALTERNATIVE HULL DETECTION TECHNIQUES FOR PREPROCESSING IN PROTON COMPUTED TOMOGRAPHY RECONSTRUCTION

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Computer Science

---

by

Blake Edward Schultze

June 2013

ALTERNATIVE HULL DETECTION TECHNIQUES FOR PREPROCESSING IN PROTON COMPUTED TOMOGRAPHY RECONSTRUCTION

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

---

by

Blake Edward Schultze

June 2013

Approved by:

Keith Evan Schubert, Advisor, School of Computer Science and Engineering	Date
Ernesto Gomez	
Dick Botting	
Reinhard Schulte	

ABSTRACT

Proton computed tomography (pCT) is a novel imaging modality developed for patients receiving proton radiation therapy. Proton CT reconstruction requires solving a large sparse linear system of the form $Ax = b$, where A is the system matrix constructed using the estimated paths of a large number of protons intersecting the discretized object, the unknown object vector x of relative stopping power (RSP) values, and b is a vector containing the water-equivalent path length (WEPL) measurements from each registered proton. Determining the elements of the matrix A requires accurate knowledge of the object boundary in reconstruction space. Previous work has used filtered backprojection (FBP) to determine this boundary and produce an approximate object hull, but FBP is a sophisticated image reconstruction algorithm itself, so this an overly complicated and time consuming method for simply determining the hull of an object. In addition, it is unclear if FBP will be capable of satisfying the requirements of the image reconstruction algorithms currently being developed to exploit the inherent sparsity in the linear system. Therefore, the purpose of the work presented here was to develop computationally efficient hull detection techniques appropriate for image reconstruction using sparse matrices. The hull detection techniques investigated were space carving (SC), modified space carving (MSC), and space modeling (SM) and these were compared to the cone-beam version of the filtered backprojection (FBP) algorithm in terms of their computation time and the quality of the object hull they produced. The input data used for these comparisons included simulations of a digital head phantom and experimental data obtained from scans of a pediatric head phantom and a live rat under anesthesia.

ACKNOWLEDGEMENTS

I would like to thank my committee members for all of their help throughout the course of my research, especially Dr. Schubert and Dr. Schulte for all of the late nights they spent with me. I would also like to thank the students and faculty at Loma Linda University and University of California, Santa Cruz, especially Ford Hurley and Vladimir Bashkirov for their valuable input and suggestions. I must also acknowledge the sacrifices that spouses have made so that these people could be available to me, especially Kym Schubert for all of the times Dr. Schubert spent hours with me at school and at their home.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Cancer, known medically as a malignant neoplasm, is one of the biggest health related issues in our society, especially in modern times as the average life span increases. It is the second largest cause of death in the United States behind major cardiovascular diseases with more than one million new cases diagnosed every year [2]. There are various treatments for cancer including surgery, radiation therapy, and chemotherapy with about half of all cancer patients receiving definitive radiation therapy as either their primary treatment or in combination with chemotherapy or surgery. Overall, approximately two thirds of all cancer patients will receive radiation therapy at some point during their treatment [3]. Radiation therapy is a very dynamic research field which is being driven by new technology developments but the majority of radiation treatments are still performed with linear accelerators (linacs) that generate energetic electron beams and x-rays.

However, in recent decades particle beam therapy using proton and carbon ion beams has rapidly evolved into a new and exciting frontier in cancer therapy [4]. One of the biggest challenges to proton and ion therapy is the uncertainty in the range prediction of particle beams in tissue, which arises due to the fact that x-rays and particle beams interact with tissue differently. Range predictions are currently derived from x-ray CT reconstructions, which provide attenuation coefficients of each object

voxel, so the effectiveness of particle beam therapy is limited by how accurately these attenuation coefficients predict the behavior of a particle inside a patient. Since this behavior actually depends on the stopping power of the tissue the proton encounters and xCT provides the attenuation coefficients of each object voxel, particle ranges are predicted by converting attenuation coefficients into stopping powers. Unfortunately, the error incurred with this conversion is 3-4% of the proton's range on average, though they can be larger in locations with large tissue inhomogeneities, such as the interface between bone and soft tissue [5]. This can sometimes result in a 10-15 mm error in a proton's range inside a human head, thereby partially offsetting the benefits of particle beam therapy and eliminating it as a viable treatment for tumors located near sensitive brain structures, so there is substantial incentive to develop alternative methods of range prediction. Proton CT has emerged as a particularly appealing alternative because the range uncertainty can be reduced by measuring the water-equivalent path length (WEPL) of protons traversing the patient and determining the relative stopping power (RSP) of each object voxel directly using tomographic image reconstruction techniques [1].

The solution of the reconstruction problem requires solving a large linear system of the form $Ax = b$, where A is the system matrix constructed using the estimated paths of a large number of protons intersecting the discretized object, represented by the unknown object vector x of relative stopping power (RSP) values, and b is a vector containing the measured water-equivalent path length (WEPL) values of each registered proton [6]. Proton CT is a challenging imaging technique because of the complicated physical interactions occurring as the proton travels through the patient,

resulting in proton paths that deviate from a straight line. However, this challenge can be overcome by using the most likely path (MLP) when calculating the system matrix A [7]. An additional challenge arises from the statistical variations in the WEPL measurements, which can be reduced by increasing the number of protons. The most important feature of proton CT is the absence of the systematic error introduced by the conversion of Hounsfield values from x-ray CT, thus reducing the range uncertainty in proton treatment planning.

Preprocessing is a very important aspect of proton CT as this phase of reconstruction identifies and discards unsuitable proton data that are caused by inelastic nuclear interactions and large-angle elastic scattering events. Preprocessing also constructs the system matrix A by approximating the path of each proton with the MLP and produces the initial iterate of the object vector x for subsequent iterative image reconstruction. Therefore, preprocessing is perhaps the most important aspect of image reconstruction and the research contained in this thesis focuses on improving its effectiveness. In particular, methods for improving the accuracy of object hull-detection are explored since this directly affects the accuracy of the system matrix A and object vector x and preconditions image reconstruction.

1.1 Physical Interactions of Protons

For the theory of pCT, it is important to understand the physical interactions of protons as they interact with matter. In principle these can be divided into two main interactions: (1) Scattering due to Coulomb interactions with the nuclei of the material, which leads to changes in the direction of the proton usually without energy

loss (multiple Coulomb scattering) and (2) energy loss due to Coulomb interactions with the outer electrons of the material atoms, which does not change the proton direction [8]. In addition, there are occasional collisions with the atomic nuclei which either lead to large-angle elastic scattering events or inelastic interactions with the nucleus, which leads to a large energy loss and angular deflection of the proton.

When analyzing individual proton data during preprocessing, it is important to determine which physical interactions occurred so appropriate actions can be taken to prevent inaccurate data from affecting the quality of the reconstructed image. Therefore, we must define the various types of physical interactions and how they affect the proton's behavior so we can identify when they occur and what action is appropriate to take when they do.

1.1.1 Scattering

The trajectory of the proton is determined mostly by multiple Coulomb scattering (MCS) events and occasionally by large-angle elastic or inelastic nuclear scattering events [8]. MCS events arise due to the electrostatic forces of the positive nuclei of the traversed material on the positively charged proton. Because electric fields are conservative, these are elastic scattering events so the proton does not lose energy. MCS events deflect the proton slightly, but the amount of deflection depends on the proton's energy (i.e. velocity), increasing in average magnitude as the proton energy decreases . Elastic and inelastic nuclear scattering events arise due to collisions between the proton and atomic nuclei, potentially producing large angle scattering. The inelastic scattering events only affect a minority of protons but if they occur, the

proton is frequently lost in the interaction. If it retains sufficient energy it may emerge from the scanned object and contribute to the data that deviates from the typical behavior. These proton data have to be recognized and discarded. The MCS events occur repeatedly as the proton travels downstream, adding a statistically random component to the proton's trajectory and producing a zigzag path. Fortunately, the angular dispersion resulting from the sum of these small deviations can be modelled fairly well by a Gaussian distribution and, therefore, protons that underwent one or more nuclear interactions can be recognized.

The probability of a scattering event depends on the mean distance between the atomic nuclei in the medium the proton travels through, so a medium with a low density of atoms, like air, will produce minimal scattering. On the other hand, although the silicon strip detectors used for tracking are reasonably dense, the proton spends very little time inside them because they are thin, so there is minimal scattering in the tracking detectors as well. In fact, because scattering is far more likely to occur inside the object than anywhere else, it is assumed that scattering occurs exclusively in the object. Although this simplifying assumption is not strictly true, since it is wrong so infrequently and introduces such insignificant error, it would not be worthwhile to attempt to account for such events.

Aside from simply deflecting the trajectory of the proton, there are several additional consequences of scattering and each impacts the accuracy of measurements to a varying degree, depending on where they occur in the scanner system. The two most significant effects of scattering are range straggling and energy straggling.

1.1.2 Energy Loss

When protons traverse a medium, they undergo (quasi-) continuous slow down due to multiple interactions with the Coulomb field of the outer electrons of the medium. The mean energy loss per path length is called the stoppling power and the stopping power relative to water (RSP) is the quantity that is reconstructed in pCT. The stopping power depends mostly on the electron density of the medium and additional quantum mechanical effects that are characterized by the medium-dependent mean excitation energy I. The energy loss is well described by the Bethe formula [8].

1.1.3 Range Straggling

An important quantity to consider for proton treatment is the *range* of the proton, which is how far a proton with a specific initial energy will penetrate into a material before losing all of its energy and coming to rest. This is important because a proton releases the bulk of its energy just prior to coming to a stop, so the goal of proton treatment is to stop the proton at specific locations inside the patient so that radiation is delivered to the desired location while sparing nearby sensitive structures. This is accomplished by choosing the initial energy of the proton such that its range coincides with the depth of the desired delivery site. If a proton lost energy definitively as it penetrated an object, its range would be a well defined quantity that depended on the proton's initial energy and the target medium. Unfortunately, a proton's energy loss is a statistical process which depends on the number of collisions with atomic nuclei. Therefore, the range of identical protons will be distributed around some mean range, which to a first approximation, resembles a Gaussian distribution [8]. Range

straggling is of importance because the width of its distribution limits how accurately we can measure the WEPL values in pCT.

1.1.4 Energy Straggling

Energy straggling is essentially the same phenomenon as range straggling but viewed from a different perspective; energy straggling is the statistical distribution of energies possessed by protons after penetrating a fixed distance through an object [8]. For an absorber of reasonable thickness, like an object for example, this too can be modelled by a Gaussian distribution. Although this distribution is also the result of the variable number of collisions occurring while a proton passes through a specific thickness of material, this is closely related to the residual energy of protons leaving the object rather than residual range. Thus it is important when WEPL is determined by measuring residual energy with a calorimeter rather than residual range with a range counter.

1.2 Proton Therapy

Proton therapy is a modern form of radiation therapy that can be applied more precisely than the conventional radiation therapy with x-rays. The following section briefly describes the two main components of a proton therapy system, such as the one installed at Loma Linda University Medical Center, where a proton CT scanner is being developed.

1.2.1 Proton Accelerator

The proton accelerator used at the Loma Linda University facility is a synchrotron, which is a variation of the classical cyclotron particle accelerator [9]. A cyclotron accelerate particles inside a vacuumed disk cavity using a constant magnitude magnetic field (for guiding) and a constant frequency electromagnetic field (for accelerating), causing the particle to spiral outward as it accelerates. A synchrotron varies these two fields in time so the path of the particle can be held constant as it accelerates inside a torus shaped vacuum cavity, allowing the acceleration, guidance, and focusing of the particle beam to be performed by separate components and producing a more cost effective design solution.

By adjusting the strength of the magnetic field and frequency of the electromagnetic field, the accelerator can produce protons with various kinetic energies, but the two typical energies used at this facility are 200 MeV and 100 MeV. Although the time between successive protons being produced by the accelerator is not constant, the average rate at which protons are produced by the accelerator can be varied. However, it is not necessarily possible, or even helpful, to produce protons at the accelerator's highest possible rate. Proton production is approximately a Poisson process, so although we would like one proton to be produced at a time, there is some probability that two or more protons will be delivered at one time and this increases with the rate of proton production. In addition, there is a limit on the rate at which the detectors and readout electronics can be operated, so this places a limit on the rate of proton production.

The accelerator feeds several rooms at the facility, with each line offering different

delivery and scanning options based on the intended applications (e.g. pCT/radiography research, proton treatment, etc.) in the particular room. Large magnets are used to guide the proton beam from the accelerator to the beam delivery location, which is a fairly large distance, so the proton beam is not a perfect point source once it reaches the delivery location. Fortunately, a perfect point source is not required for the scan and most references to the source location have been removed from preprocessing calculations. However, the source position is still required to calculate the cone/fan beam angle during filtered backprojection, but it can safely be considered a point source for these calculations and this filter is likely to be replaced in the future anyways. Similarly, the energy of the proton has some variation due to the proton selection process at the accelerator and interactions along the path from the accelerator to the delivery location, but this variation is quite small and can be ignored.

1.2.2 Proton Gantry

The purpose of the proton gantry is to delivery a proton beam from any angle in the plane perpendicular to the rotational axis of the gantry. This allows to target a tumor from many different directions but can also be utilized for a pCT scan, which requires penetration of the object from many different angles covering a full circle. There is some uncertainty in the position of the gantry due to the torque produced by the large weight 100 tons) of the gantry, but this discrepancy, while being important for treatment accuracy, can be ignored for pCT because protons are tracked independently from the gantry position. It is possible to rotate the gantry continuously throughout the scan, and this may be explored in the future, but the

current approach is to to scan the object on a fixed horizontal beam line by rotating it in discrete angular steps during the scan.

2. PROTON COMPUTER TOMOGRAPHY (PCT)

Classically, treatment planning for proton therapy has used the tomographic images produced by scanning the object with x-rays in a process known as x-ray computed tomography (xCT). Part of the appeal of x-rays is that they do not scatter significantly, so their path through the object is well approximated by a straight line, greatly simplifying the reconstruction process. The effectiveness of proton treatment depends on how accurately protons can be delivered to a particular location inside the object, which requires accurate knowledge of the relative stopping power (RSP) throughout the object. However, xCT images provide information about the attenuation coefficients (Hounsfield units) of various portions of the object, so attenuation coefficients must be converted into approximate RSP values for treatment planning. Unfortunately, this conversion alone can introduce up to 3% error, which combined with the other measurement errors and uncertainties, limits the accuracy of proton delivery and reduces the benefits of proton treatment over traditional radiation treatments. To overcome these issues and eliminate the need to convert from Hounsfield units to RSP values, the object can be scanned using protons instead of x-rays (called pCT), allowing RSP values to be obtained directly. The pCT data can then be used to predict the behavior of a proton inside an object, significantly improving proton treatment planning.

2.1 Design of the LLUMC Proton CT Scanner

The proton CT scanner currently in use at Loma Linda University Medical center uses a synchrotron proton accelerator to produce a proton beam with a desired energy which is then guided to the beam delivery room using magnetic fields. A schematic representation of the Phase I scanner is shown in Figure 2.1 and an image showing a side view of the actual scanner is shown in Figure 2.2 [10]. The initially pencil like beam is directed toward a thin scattering foil, thereby producing a cone beam and providing full scanning coverage of the target object. Individual protons are tracked using silicon strip detectors (SSDs), which record the position of a proton as it passes through it. By placing two tracking detector pairs before and after the object, the trajectory of the proton before and after the object can also be determined. The phase I scanner uses a CsI calorimeter, which is composed of an array of individual CsI crystals and photodiodes, to determine the residual energy of the proton as it exits the object.

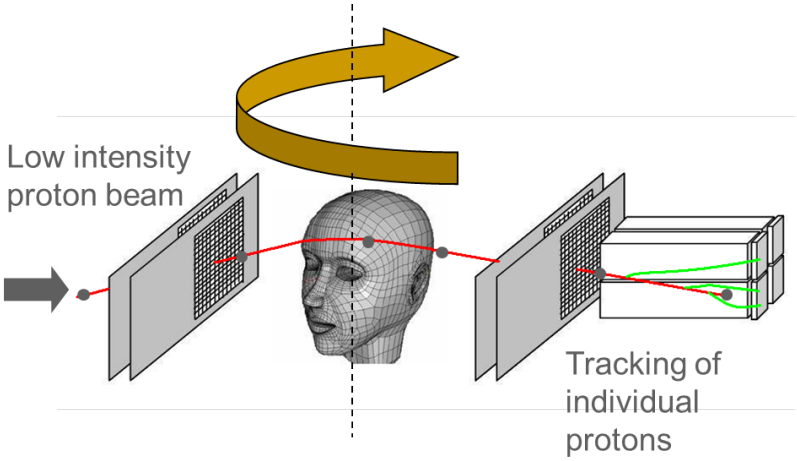


Fig. 2.1: Schematic representation of the phase I proton CT scanner at Loma Linda University Medical Center

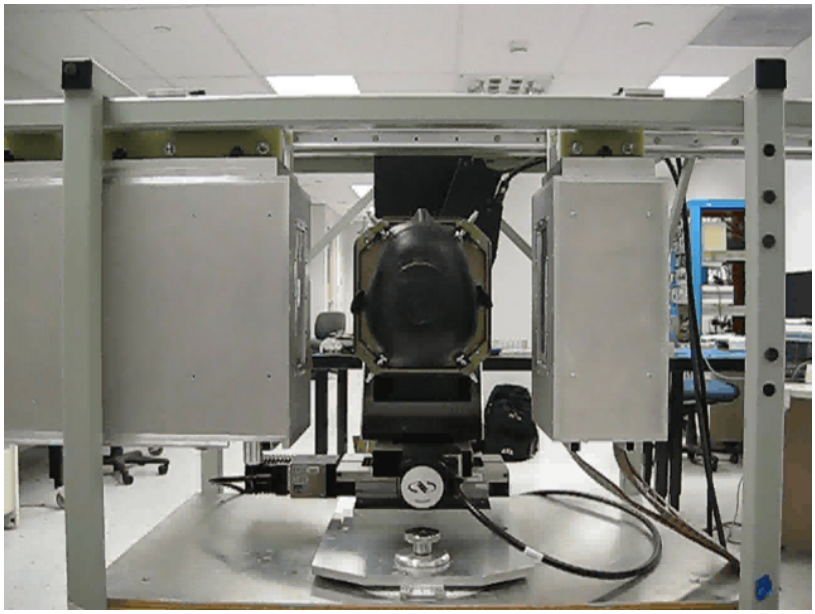


Fig. 2.2: Image showing a side view of the actual phase I proton CT scanner at Loma Linda University Medical Center

2.1.1 Tracking Detectors

The tracking system on Loma Linda University's Phase I proton scanner uses silicon strip detectors (SSDs) to determine the path of a proton as it travels through the scan region. Each silicon strip detector is composed of n-type bulk silicon with a thin layer of aluminum on the back face and p-type strips implanted in the front face. The n-type and p-type materials are produced by adding impurity atoms to a semiconductor material such as silicon in a process called *doping*. As a proton traverses the detector, its kinetic energy excites electrons in the n-type region and frees electrons for conduction, thereby ionizing the n-type region. The electron-hole pairs produced by this ionization migrate toward the positive and negative terminals and accumulate on the edges of the n-type and p-type regions, respectively. The

negative charges accumulate on the surface of the p-strips imbedded in the top face of the SSD and produce a current that can be detected by the readout electronics and used to determine the position of the proton as is passed through. The design of an SSD is shown schematically in Figure 2.3

The aluminum strips in an SSD are all aligned in the same direction so a single SSD can only provide the position of the proton in a single direction. To determine the position of the proton in both the horizontal and vertical directions, two silicon strip detectors (SSDs) are placed back to back and rotated by ninety degrees with respect to each other so the position of the proton can be determined in both directions. There are double sided silicon strip detectors (DSSDs) which place $n-$type strips on the backside of the SSD's $n-$type region, rotated by ninety degrees with respect to the $p-$type strips, thereby exploiting the simultaneous migration of positive and negative charge carriers to acquire the proton's position in both the horizontal and vertical directions. This decreases the overall thickness of the tracking detectors and therefore, the probability that the proton is scattering in the detector, so it may be beneficial to replace the current tracking detectors by these in the future. Associated

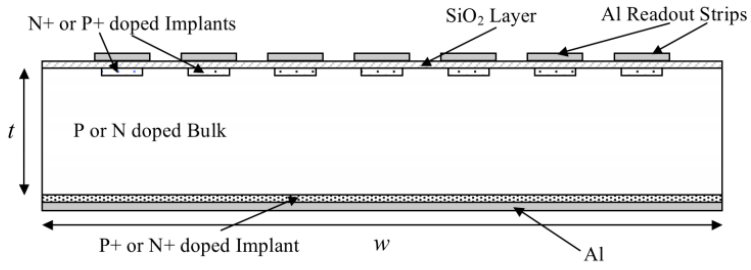


Fig. 2.3: Diagram depicting the mechanism of charged particle detection in a silicon strip detector (SSD)

with every physical measurement is some level of measurement uncertainty related to the precision of the measurement device. The statistical nature of the proton's behavior contributes an additional uncertainty to the various measurements, some of which can be accounted for mathematically, but sometimes this requires data to be discarded. These uncertainties propagate through subsequent calculations and have a varying effect on accuracy depending on the mathematical relationship between the particular measurement and the value being sought.

2.1.2 Energy Detector

The detector used to measure the residual energy of the proton in the phase I scanner is a CsI (Cesium Iodide) crystal calorimeter [11]. For practical purposes and to increase resolution, the calorimeter is comprised of a 6×3 array of CsI crystals separated by a thin layer of mylar, resulting in a sensitive area of 9 cm$\times$ 18 cm. A scintillating material like CsI luminesces when it is excited by ionizing radiation. Therefore, the energy deposited in the crystal by incoming protons excites electrons in the crystal to a higher energy level and this excess energy is released via photon emission as electrons return to a neutral energy state. These photons strike the photodiode attached to the back of the crystal, producing an electrical signal whose strength depends on the number of photons emitted. The proton slows as it passes through the crystal, so the depth at which is stops, and thus the number of photons emitted, depends on its incoming energy. The readout electronics then transmit the signal to be analyzed to approximate the WEPL value for this response based on the calibration measurements performed before the scan [1]. Measurements are

synchronized using the tracking detector events to trigger the readout electronics and because the higher energy states have a very short lifetime (about 16 ns), these crystals are suitable for scanning with a high rate of incoming protons.

Energy straggling produces random fluctuations in residual energy and contributes to the uncertainty in WEPL measurements. In addition, the proton can enter the calorimeter at an angle, so it may pass through multiple crystals or even escape the calorimeter all together. If the proton passes through multiple crystals, the combined response can be used to approximate a WEPL value by calculating a weighted sum, but these situations can be difficult to identify; it is not uncommon for several crystals to produce a measurable response even when the proton only passes through a single crystal, most likely resulting from light leaking or reflecting from one crystal into another. Various techniques have been developed to handle these issues and improve calibration, but these remain a significant source of uncertainty in WEPL measurements.

2.2 Proton CT Data Acquisition and Event Builder

During a scan, protons are produced at an adjustable rate, accelerated up to the desired energy, and steered with magnetic fields to the beam delivery location. The beam is then directed towards the target object and mounted to a rotating gantry so that the direction of the proton beam can be adjusted, allowing the object to be be scanned from different angles. A pair of detector planes on either side of the object record the position of the proton before and after the object and a calorimeter records the residual energy of the proton after is passes through the object. The

detector planes and calorimeter rotate with the gantry so they remain stationary with respect to each other and the object remains between the two pairs of detector planes and between the incident beam and calorimeter. Instead of recording the detector responses continuously throughout the scan, the passage of the proton through the SSDs triggers the readout electronics to perform data acquisition.

2.2.1 Triggering

A threshold is placed on each of the detectors so that random fluctuations (such as thermal fluctuations) do not trigger accidental measurements. When a proton passes through the tracking planes, their electrical response rises above this threshold and the readout electronics are signaled that a measurement should be taken. This also signals the readout electronics of the calorimeter so WEPL measurements corresponding to the proton can also be obtained. Timing measurements is a challenging task because there is some delay between an event and the triggering of a measurement and although this delay is quite small, so is the propagation time of the proton. Fortunately, using the tracking detector events to trigger the calorimeter readings helps synchronize the various measurements.

2.2.2 Readout Electronics: ASICs and FPGAs

The experimental data acquired by the four silicon strip detectors (SSDs) and calorimeter constitute the information associated with each proton and are collectively known as a proton *history*. For each detector there is an application specific integrated circuit (ASIC) that drives data collection and a field programmable gate array (FPGA)

that transmits this data to another FPGA called the *event builder*. The event builder receives data from each detector and performs the task of associating the individual measurements with a particular proton and combining them into a single proton history. This is a nontrivial task since a proton can miss or fail to trigger one or more detectors, resulting in incomplete data for a particular proton history and requiring this history to be discarded. This means that a proton history must first be determined to be complete, so data cannot simply be assigned to proton histories in the order it is received.

There is a time stamp associated with each measurement and this can be helpful in discerning data from different protons, but the data rates desired from scans produce very short time intervals between consecutive protons and the chronological order of protons can only be guaranteed through a particular detector since the response time of the detectors and their associated communication electronics are inhomogeneous and prevent reliable system synchronization. Despite these complications, measurements can be assigned to the correct proton history most of the time and simply be discarded if not. Discarding data is infrequent enough that it does not present a serious issue to data collection.

2.2.3 Data Format and File Naming Scheme

When pCT is ready to be used in a clinical environment, experimental data will be transmitted directly to the computers performing image reconstruction. However, at the current stage of development, the data acquisition, preprocessing, and image reconstruction phases are being developed by separate groups and consequently, are

not integrated into a single process or program. Therefore, the experimental data is written to file and stored on the collaboration server so others can access and use the data off site. However, to simplify integration later, the collaboration has agreed upon a data format which makes sense both for storage and computing now and for direct transmission later. The data format has changed over time as necessary but the changes that have been made to it recently have all been to improve performance. The file naming scheme also needs to be standardized since there are multiple files for a single data set and these need to be read automatically by the preprocessing program. Currently, the data format and file naming scheme is as follows:

1. Data is stored in binary (.bin) file format

2. There is a separate file for each gantry angle and the end of the file name includes three numbers which indicate the particular gantry angle this file corresponds to. For example, gantry angle 0 would end in *000.bin*, gantry angle 4 would end in *004.bin*, and gantry angle 356 would end in *356.bin*.

3. Sometimes the object is scanned once and then moved (i.e. translated) and scanned again. Therefore, each file name includes an indicator of which translation it is from. This is indicated using the abbreviation "*trans*" and then the number of the translation before the gantry angle number and separated from it by an underscore (_). In other words, the name of the file corresponding to the 1^{st} translation and gantry angle 0 ends with *trans1_000.bin* and the name of the file name corresponding to the 2^{nd} translation and gantry angle 12 ends with *trans2_012.bin*.

4. The name of the data set precedes the translation number and gantry angle number. For example, if a water phantom is scanned and we want to call the data set "*water*", then the name of the file corresponding to the 1^{st} translation and gantry angle 0 is *water_trans1_000.bin* and the name of the file name corresponding to the 2^{nd} translation and gantry angle 12 is *water_trans2_012.bin*.

5. For each file, the values of a particular measurement are listed for each proton history sequentially. For example, each file begins with the measurements from the 1^{st} detector ($v\_in\_1(i)$) for each of the N proton histories; these are listed sequentially ($0 \leq i \leq N$) and separated by a space, as in:
   $v\_in\_1(1)\, v\_in\_1(2)\, v\_in\_1(3) \cdots v\_in\_1(N)$

6. The 14 measurements are listed in the following order:

   (a) v_in_1

   (b) v_in_2

   (c) v_out_1

   (d) v_out_2

   (e) t_in_1

   (f) t_in_2

   (g) t_out_1

   (h) t_out_2

   (i) u_in_1

   (j) u_in_2

(k) u_out_1

(l) u_out_2

(m) $WEPL$

(n) $gantry\_angle$

Therefore, for a file containing measurements from N proton histories, the file begins with all of the N v_in_1 measurements in order, followed by all of the N v_in_2 measurements in order, $\cdots$ and ending with the N $gantry\_angle$ measurements in order.

Until recently, the measurements corresponding to a proton history were written together such that each line in the file contained the fourteen measurements from a single proton history. Although this was a perfectly sensible way to write the data, the new format makes reading the data a more efficient process. Each of the fourteen measurements are stored in their own array, which are obviously quite large, so the difference between the first memory address of two arrays is greater than or equal to the number of proton histories. By writing all of the data for a particular measurement sequentially, the data is read in the same order as it is stored in memory, so unlike the previous format, a large jump in memory address is not required prior to each write. Although this will not produce a sizeable difference (it may not even be noticeable) but the reading process is slightly faster and certainly more efficient with the new data format and it is easier to write the data to file in this format, so there is no reason not to use the new format.

2.3 Measurement Errors and Uncertainties

2.3.1 Tracking Uncertainties

When a proton passes through a tracking detector, it ionizes the $n-$type material and the biasing electric field forces charge separation. The negative charges migrate towards the aluminium strips and trigger the readout electronics. Since an entire region of the $n-$type material is ionized, the numerous negative charge carriers repel each other as they migrate towards so a measurable charge accumulates on multiple aluminum strips. The average index of the triggered strips is then calculated and used to assign a position to the proton at that detector. Therefore, the uncertainty in the position of the proton at each tracking detector depends on the number of strips triggered, which increases as the number of ionized atoms increases, so this uncertainty is larger with higher energy protons. Fortunately, since the strips are thin, the mean strip index is a relatively accurate method for determining the position of the proton as it passed through the SSD.

2.3.2 WEPL Measurement Uncertainties

A crystal calorimeter is typically used to measure the residual energy of ionizing radiation based on the amount of light received by the attached light detector, since these are proportional, but we are interested in WEPL measurements and these cannot be determined directly. Accurate WEPL measurements require calibration routines to be performed prior to a proton scan so the relationship between detector response and WEPL value can be determined. Since the calorimeter is constructed using an array of CsI crystals and each responds slightly differently, this relationship must

be determined for each individual crystal. In addition, the statistical behavior of the proton causes the response of the calorimeter to fluctuate slightly, so numerous measurements are taken and the average response is calculated for each WEPL value.

Known WEPL values are produced by inserting a homogenous absorber of known thickness and RSP, which corresponds to a specific WEPL, and numerous measurements are taken to determine the relationship between this WEPL value and the average detector response of each crystal. Repeating this procedure with several absorbers of varying thickness, a calibration curve can be interpolated using the relationship between the average detector response of each crystal and WEPL value for several different WEPL values. As you can see in Figure 2.4, the relationship between WEPL value and average calorimeter response is nonlinear [1]. Since the calibration curve is then used to convert calorimeter responses into a WEPL measurements during a scan, the precision of each WEPL measurement is limited by the magnitude of the fluctuations in calorimeter response and by the response differences between each crystal. However, there are various other sources of measurement uncertainty which are unrelated to the precision of the calorimeter itself.

2.3.3 Unresolvable Proton Histories

As it was mentioned earlier, there is some probability that the proton accelerator will produce more than once proton simultaneously and this causes several problems. The first challenge is to identify the occurrence of these events. This can typically be accomplished using the tracking detector information because the mutual repulsion between protons usually causes the protons to pass through different portions of the

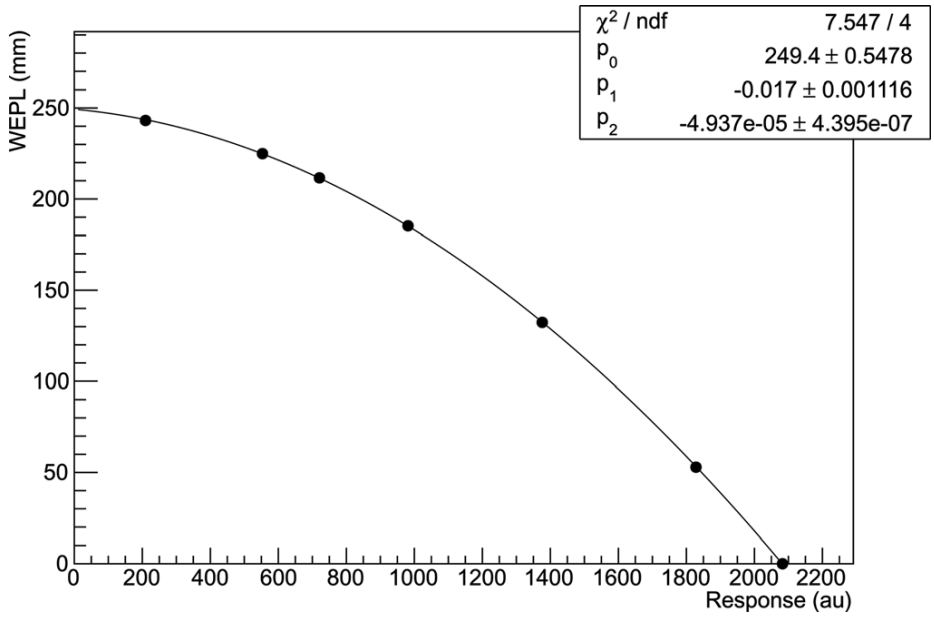


Fig. 2.4: Graph relating the calorimeter response to a WEPL value [1]

SSDs, allowing the position of each proton to be determined individually. Likewise, this can continue to be done in the other tracking detectors and calorimeter, but these positions must also be assigned to the correct proton history. Most of the time this can be inferred by the trajectory of the proton, but when it cannot, these proton histories must be discarded. The most significant issue arising from simultaneous protons is with WEPL measurements because it becomes difficult to discern between the responses of each proton. In the worst case scenario, the sum of the responses is assigned to a single proton history. Although this procedure is challenging and introduces the possibility that measurements are assigned to the wrong proton history, multiple proton events occur much too frequently to simply discard them every time, so thy should be retained whenever possible.

Somewhat related to the issue of simultaneous protons is the issue of detector

pile-up and *after glow*. The response of the calorimeter does not return immediately to zero, there is a decay time which depends on the magnitude of the previous measurement. If two protons are separated by a small amount of time, it is possible for the second proton to enter the calorimeter before its response from the first proton has decayed back to zero. This produces an inflated detector response for the second proton, resulting in an over estimation of the WEPL measurement for this proton. There are various strategies which could be used to account for the residual response of the calorimeter prior to each measurement, but these have not been incorporated into the scanning routine at this point in development.

3. PREPROCESSING

The purpose of preprocessing is to define and precondition the linear system $Ax = b$ for subsequent iterative image reconstruction, where each row of the system matrix A represents the path of a proton through the image vector x and the corresponding row of the vector b contains the $WEPL$ measurement associated with this proton. It is important to produce a well-conditioned linear system, since the condition number directly affects the convergence rate of iterative reconstruction and the quality of the reconstructed image. A well-conditioned linear system is primarily dependent on how accurately the object hull can be determined during preprocessing since this information is used to construct the system matrix A and precondition the linear system. This is a nontrivial task as this represents a circular problem: the linear system must be well-conditioned to ensure iterative reconstruction produces an accurate image of the object, but at the same time, an accurate image of the object is required to construct a well-conditioned linear system. Fortunately, an approximate object hull is sufficient as long as it correctly identifies all voxels belonging to the object, even if it includes some voxels that do not actually belong to the object. Therefore, the goal of preprocessing is to quickly identify the voxels belonging to the object and the various steps of preprocessing have been designed to aid this effort and improve the accuracy of the object hull approximation.

3.1 Configuring Preprocessing Parameters and Counting Histories

Accompanying every experimental data set is a configuration file (.cfg file format) which defines scanning parameters relevant to preprocessing, such as the number of translations, the angular interval between gantry positions, the data format, and the position of the tracking detectors. The number of translations notifies the program of how many scans of an object were performed. Sometimes an object is scanned more than once, typically with the object placed at different positions for each scan, in an attempt to either provide better resolution or to overcome asymmetric behavior somewhere in the scanning system (such as a defective portion of a tracking detector).

The gantry is rotated in discrete angular intervals ($\Delta\theta$) during a scan and the size of these intervals determines the number of projection angles (projection angles $= {}^{360^\circ}/_{\Delta\theta}$) since a complete scan rotates around the object completely. There is a separate file associated with the data from each projection angle so the number of files that need to be read can now be determined:

$$\#\text{ of files to read} = \text{translations} \times \text{projection angles} = \text{translations} \times \left({}^{360^\circ}/_{\Delta\theta}\right).$$

A typical scan can easily contain $100,000,000$ proton histories, each of which has coordinates (i.e. (u, t, v)) for the four tracking detector intersections, a WEPL measurement, and a gantry angle associated with it. This data is stored as either a *float*

or an *integer*, 4 bytes each, so this corresponds to:

$$\begin{aligned}\text{size of data} &= \#\text{ proton histories} \times \#\text{ of data values} \times \text{size of float/integer} \\ &= 100,000,000 \times 14 \times 4 \text{ bytes} = 5,600,000,000 \text{ bytes} \\ &= 5,468,750 \text{ kilobytes} = 5,340.58 \text{ megabytes} \\ &= 5.21541 \text{ gigabytes}\end{aligned}$$

The sheer volume of experimental data alone requires the data to be processed iteratively. While it is not uncommon for a computer to have 16GB+ of host RAM, especially one that is to be used in a hospital for biomedical imaging, a typical GPU will not have enough RAM to store the entire experimental data set as well as any additional arrays required for intermediate data storage. To reduce the computation time required to produce reconstructed images, pCT reconstruction utilizes GPGPU programming as much as possible in an attempt to exploit the inherent parallelism present in the process. To account for the possibility that the GPU cannot maintain the entirety of the experimental data simultaneously, the data is read iteratively, with each iteration processing a user definable number of proton histories. However, if the limit on the number of simultaneous proton histories allowed exceeds the total number of proton histories, the entire data set is read and processed in one iteration. To facilitate the iterative data reading process, all the data files are scanned to determine how many proton histories each contains prior to reading the data.

The data format specifies the way in which the data is stored in files. For example, a common data format is structured such that each row in the file corresponds to a single proton history. The last parameter provided by the configuration file is the location of the tracking detectors along the proton beam direction. The tracking

detectors remain stationary with respect to the proton beam source since they rotate with the gantry, therefore these locations remain constant throughout the entire scan. However, the detectors are moved prior to each scan to accommodate the size of the target object so their location does not remain constant for every scan and this information must be specified in the configuration file.

3.2 Iteratively Reading the Data From File

Once the configuration file has been read and the histories have been counted and recorded for each file, the iterative data reading can begin. Prior to reading the data each iteration, the files that will be read in this iteration is determined by calculating how many files can be read in entirety without exceeding the limit on simultaneous proton histories imposed by the amount of RAM on the GPU. If a file cannot be read in its entirety, reading this file is postponed until the next iteration. Once the files to be read have been identified, they are read according to the particular file format and the data is transferred to the GPU so it can be used to determine which protons passed through the reconstruction volume and discard those that did not.

3.3 Determining if and Where a Proton Passed Through the Reconstruction Volume

An unavoidable fact about the data acquired by a proton scan is that it contains errant data, both random and systematic. Physical interactions, such as scattering, can cause significant variation between measurements, even for those that pass through the same portion of the object. In addition, there is some level of uncertainty

associated with any physical measurement due to the precision achievable with the measurement device. In the case of proton position and energy measurements, an additional source of uncertainty is the statistical nature of the physical interactions occurring along its path. Fortunately, because these physical interactions are fairly well described by statistical distributions, errant data can be identified and discarded if a sufficient number of measurements are taken.

In order to perform statistics, however, protons that follow a similar path must be grouped together so the typical behavior of protons passing through approximately the same portion of the object can be determined [12]. Identifying protons with similar paths through the object requires accurate knowledge of where the protons entered and exited the object. Unfortunately, at this point we don't have any specific knowledge of the object or its boundaries, making it difficult to determine proton paths. However, approximate dimensions of the target object can be determined either directly or inferred based on the object being scanned (e.g. adult torso, pediatric head, rat, etc.). This can then be used to define the smallest volume of space guaranteed to contain the entire object; this volume is called the *reconstruction volume* and is defined to be a cylinder of radius r and height Z centered on the gantry's rotation axis. Proton paths are then calculated assuming that the boundaries of the object coincide with the surfaces of the cylinder, so the reconstruction volume is ideally just large enough to fully enclose the object.

The pair of tracking detector planes before the object record the (t, v) position of the proton as it passes through each plane. Although the detector planes rotate along with the gantry, the distance (u_i) between the axis of rotation $(v - axis/z -$

axis) and each detector plane remains the same and the (u, t, v) coordinates can be used to calculate the trajectory of the proton in the utv-coordinate system (detector coordinate system). Defining the coordinates of the proton through the first and second detectors as (u_1, t_1, v_1) and (u_2, t_2, v_2), respectively, the angle of the proton before the object in the utv-coordinate system is given by:

$$\theta_{ut_{entry}} = \arctan\left(\frac{t_2 - t_1}{u_2 - u_1}\right) \tag{3.1}$$

$$\theta_{uv_{entry}} = \arctan\left(\frac{v_2 - v_1}{u_2 - u_1}\right) \tag{3.2}$$

Similarly, the pair of tracking detector planes after the object record the (t, v) position of the proton as it passes through each plane. Defining the coordinates of the proton through the third and fourth detectors as (u_3, t_3, v_3) and (u_4, t_4, v_4), respectively, the angle of the proton after the object in the utv-coordinate system is given by:

$$\theta_{ut_{exit}} = \arctan\left(\frac{t_4 - t_3}{u_4 - u_3}\right) \tag{3.3}$$

$$\theta_{uv_{exit}} = \arctan\left(\frac{v_4 - v_3}{u_4 - u_3}\right) \tag{3.4}$$

We are technically interested in the path of the proton in the global coordinate system (also called the image coordinate system) since the detector coordinate system rotates around the object. This can be determined using the angle of the gantry with respect to the global coordinate system at the time the corresponding measurements were made. The global coordinate axes are defined as (x, y, z) so that at $0°$, $(x = u, y = t, z = v)$. The angle of the proton path in the global coordinate system before and

after the object are then given by:

$$\theta_{xy_{entry}} = \theta_{ut_{entry}} + \theta_{gantry} \tag{3.5}$$

$$\theta_{xz_{entry}} = \theta_{uv_{entry}} + \theta_{gantry} \tag{3.6}$$

$$\theta_{xy_{exit}} = \theta_{ut_{exit}} + \theta_{gantry} \tag{3.7}$$

$$\theta_{xz_{exit}} = \theta_{uv_{exit}} + \theta_{gantry} \tag{3.8}$$

The positions of the proton in the global coordinate system are calculated via coordinate transformation. Coordinate transformations are performed by multiplying the position vector by the appropriate rotation matrix, depending on the axis of rotation. Since the equation of a line can be defined uniquely by a point and slope, only one of the two points before and one of the two points after the object need to be transformed. The three dimensional rotation matrix corresponding to a positive rotation angle around the z-axis is given by:

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.9}$$

Therefore, choosing to transform the coordinates (u_2, t_2, v_2) and (u_3, t_3, v_3) yields:

$$\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} \cos\theta_{xy_{entry}} & -\sin\theta_{xy_{entry}} & 0 \\ \sin\theta_{xy_{entry}} & \cos\theta_{xy_{entry}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_2 \\ t_2 \\ v_2 \end{pmatrix}$$

$$\boxed{\begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} u_2\cos\theta_{xy_{entry}} - t_2\sin\theta_{xy_{entry}} \\ u_2\sin\theta_{xy_{entry}} + t_2\cos\theta_{xy_{entry}} \\ v_2 \end{pmatrix}} \tag{3.10}$$

$$\begin{pmatrix} x_3 \\ y_3 \\ z_3 \end{pmatrix} = \begin{pmatrix} \cos\theta_{xy_{exit}} & -\sin\theta_{xy_{exit}} & 0 \\ \sin\theta_{xy_{exit}} & \cos\theta_{xy_{exit}} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_3 \\ t_3 \\ v_3 \end{pmatrix}$$

$$\boxed{\begin{pmatrix} x_3 \\ y_3 \\ z_3 \end{pmatrix} = \begin{pmatrix} u_3\cos\theta_{xy_{exit}} - u_3\sin\theta_{xy_{exit}} \\ u_3\sin\theta_{xy_{exit}} + u_3\cos\theta_{xy_{exit}} \\ v_3 \end{pmatrix}} \tag{3.11}$$

Given a point (x, y) and an angle θ_{xy}, we can then define the equation of the line in the xy-plane defining the path of the proton before entering and after exiting the

object by $y = mx + b$:

$$m_{y_{entry}} = \tan\left(\theta_{xy_{entry}}\right)$$

$$b_{y_{entry}} = y_2 - m_{y_{entry}} x_2$$

$$y_{SSD_2} = m_{y_{entry}} x + b_{y_{entry}} = x\tan\left(\theta_{xy_{entry}}\right) + \left[y_2 - x_2\tan\left(\theta_{xy_{entry}}\right)\right]$$

$$\boxed{y_{SSD_2} = (x - x_2)\tan\left(\theta_{xy_{entry}}\right) + \left(u_2\sin\theta_{xy_{entry}} + t_2\cos\theta_{xy_{entry}}\right)} \tag{3.12}$$

$$m_{y_{exit}} = \tan\left(\theta_{xy_{exit}}\right)$$

$$b_{y_{exit}} = y_3 - m_{y_{exit}} x_3$$

$$y_{SSD_3} = m_{y_{exit}} x + b_{y_{exit}} = x\tan\left(\theta_{xy_{exit}}\right) + \left[y_3 - x_3\tan\left(\theta_{xy_{exit}}\right)\right]$$

$$\boxed{y_{SSD_3} = (x - x_3)\tan\left(\theta_{xy_{exit}}\right) + \left(u_3\sin\theta_{xy_{exit}} + t_3\cos\theta_{xy_{exit}}\right)} \tag{3.13}$$

Similarly, given a point (x, z) and an angle θ_{xz}, we can then define the equation of the line in the xz-plane defining the path of the proton before entering and after exiting the object by $z = mx + b$:

$$m_{z_{entry}} = \tan\left(\theta_{xz_{entry}}\right)$$

$$b_{z_{entry}} = z_2 - m_{z_{entry}} x_2$$

$$z_{SSD_2} = m_{z_{entry}} x + b_{z_{entry}} = x\tan\left(\theta_{xz_{entry}}\right) + \left[z_2 - x_2\tan\left(\theta_{xz-entry}\right)\right]$$

$$\boxed{z_{SSD_2} = (x - x_2)\tan\left(\theta_{xz_{entry}}\right) + v_2} \tag{3.14}$$

$$m_{z_{exit}} = \tan\left(\theta_{xz_{exit}}\right)$$

$$b_{z_{exit}} = z_3 - m_{z_{exit}} x_3$$

$$z_{SSD_3} = m_{z_{exit}} x + b_{z_{exit}} = x\tan\left(\theta_{xz_{exit}}\right) + \left[z_3 - x_3\tan\left(\theta_{xz_{exit}}\right)\right]$$

$$\boxed{z_{SSD_3} = (x - x_3)\tan\left(\theta_{xz_{exit}}\right) + v_3} \tag{3.15}$$

Using the path of the proton before and after the object, these paths are then pro-

jected forward and backward, respectively, to determine where these intersect the reconstruction volume, if they do at all. Ideally we would like to determine where these paths intersect the object itself, but the object boundaries are unknown at this point so the reconstruction volume is an acceptably close approximation.

Determining whether the proton passed through the reconstruction volume amounts to satisfying two separate conditions. First, the projection of the proton path into the xy-plane must pass through the boundary of the reconstruction cylinder in the xy-plane, which is defined by a circle of radius r. This manifests as determining if there is a solution to the equation $f_{proton}(x) = f_{cylinder}(x)$, which proceeds as follows:

$$\begin{aligned}
f_{proton}(x) &= f_{cylinder}(x) \\
mx + b &= \sqrt{r^2 - x^2} \\
(mx + b)^2 &= r^2 - x^2 \\
m^2x^2 + 2mbx + b^2 &= r^2 - x^2 \\
m^2x^2 + x^2 + 2mbx + b^2 - r^2 &= 0 \\
\left(m^2 + 1\right)x^2 + 2mbx + \left(b^2 - r^2\right) &= 0 \\
Ax^2 + Bx + C &= 0 \\
\left.\begin{aligned}
\implies A &= m^2 + 1 \\
B &= 2mb \\
C &= b^2 - r^2
\end{aligned}\right\}
\end{aligned} \tag{3.16}$$

The roots of a second order polynomial $Ax^2 + Bx + C = 0$ like this can be found

using the quadratic formula:

$$
\begin{aligned}
x &= \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \\
x &= \frac{-2mb \pm \sqrt{(2mb)^2 - 4\left(m^2+1\right)\left(b^2 - r^2\right)}}{2\left(m^2+1\right)} \\
x &= \frac{-mb \pm \sqrt{(mb)^2 - \left(m^2+1\right)\left(b^2 - r^2\right)}}{m^2+1} \\
x &= \frac{-mb \pm \sqrt{(mb)^2 - (mb)^2 + (mr)^2 + \left(r^2 - b^2\right)}}{m^2+1} \\
\boxed{x &= \frac{-mb \pm \sqrt{r^2\left(m^2+1\right) - b^2}}{m^2+1}}
\end{aligned}
\tag{3.17}
$$

If a proton passes through the reconstruction volume, the formula above must produce two solutions, corresponding to the two points where the proton intersected the boundary of the reconstruction volume. The quadratic formula will produce 0, 1, or 2 solutions depending on the value of the discriminant; it will produce 0 solutions if $B^2 < 4AC$, 1 solution if $B^2 = 4AC$, or 2 solutions if $B^2 > 4AC$. Therefore, if the discriminant is evaluated and $B^2 \leq 4AC$, then this proton history is discarded. Otherwise, the two solutions are calculated, yielding the x-coordinate of the points where the proton enters (x_{entry}) and exits (x_{exit}) the reconstruction volume. The two solutions for x are compared to x_2 and x_3 to determine which is x_{entry} and which is x_{exit}; the x-coordinate closest to x_2 is defined to be x_{entry} and the x-coordinate closest to x_3 is defined to be x_{exit}. The linear equations $y(x)$ and $z(x)$ that define the path of the proton prior to entering and after exiting the reconstruction volume are then used to determine the y-coordinate and z-coordinate corresponding to each of these

two x-coordinates as follows:

$$y_{entry} = m_{y_{entry}} x_{entry} + b_{y_{entry}}$$

$$\boxed{y_{entry} = (x - x_{entry}) \tan\left(\theta_{xy_{entry}}\right) + \left(u_2 \sin\theta_{xy_{entry}} + t_2 \cos\theta_{xy_{entry}}\right)} \tag{3.18}$$

$$z_{entry} = m_{z_{entry}} x_{entry} + b_{z_{entry}}$$

$$\boxed{z_{entry} = (x - x_{entry}) \tan\left(\theta_{xy_{entry}}\right) + v_3} \tag{3.19}$$

$$y_{exit} = m_{y_{exit}} x_{exit} + b_{y_{exit}}$$

$$\boxed{y_{exit} = (x - x_{exit}) \tan\left(\theta_{xy_{exit}}\right) + \left(u_2 \sin\theta_{xy_{exit}} + t_2 \cos\theta_{xy_{exit}}\right)} \tag{3.20}$$

$$z_{exit} = m_{z_{exit}} x_{exit} + b_{z_{exit}}$$

$$\boxed{z_{exit} = (x - x_{exit}) \tan\left(\theta_{xy_{exit}}\right) + v_3} \tag{3.21}$$

Even if the quadratic formula yielded two solutions, satisfying the first condition, this does not guarantee that the proton passed through the reconstruction volume. The projection of the cylinder onto the xy-plane (i.e. the circle $x^2 + y^2 = r^2$) contains no reference to the height z, so it is possible for all or part of the proton path to pass above or below the reconstruction volume. If $|z| \geq {}^{Z}/_{2}$ at the point where the proton enters the reconstruction volume (x_{entry}), then the proton did not enter through the radial surface of the reconstruction cylinder and this history is discarded, regardless of the value of z at the exit; if $|z| < {}^{Z}/_{2}$ upon entry, then this history is retained. However, if the proton exits the reconstruction volume through the top or bottom face of the cylinder, then the proton did not exit the cylinder at the point we calculated and its coordinates (i.e. $(x_{exit}, y_{exit}, z_{exit})$) must be recalculated.

We can determine if the proton exited through the top or the bottom face of the reconstruction cylinder, and hence the $z-$coordinate at this point, based on the

sign of z_{exit}. If $z_{exit} > 0$, then the proton exited though the top face at $z = {}^{Z}/_{2}$; if $z_{exit} < 0$, then the proton exited though the bottom face at $z = {}^{-Z}/_{2}$. We can then use similar triangles to determine the coordinates of the point where the proton exited the cylinder. For example, consider the situation in Figure 3.3(a) where the proton exits through the top face of the cylinder. For simplicity and clarity, we view this situation with the reconstruction cylinder projected into the $xz-$plane. Figure 3.3(b) shows the

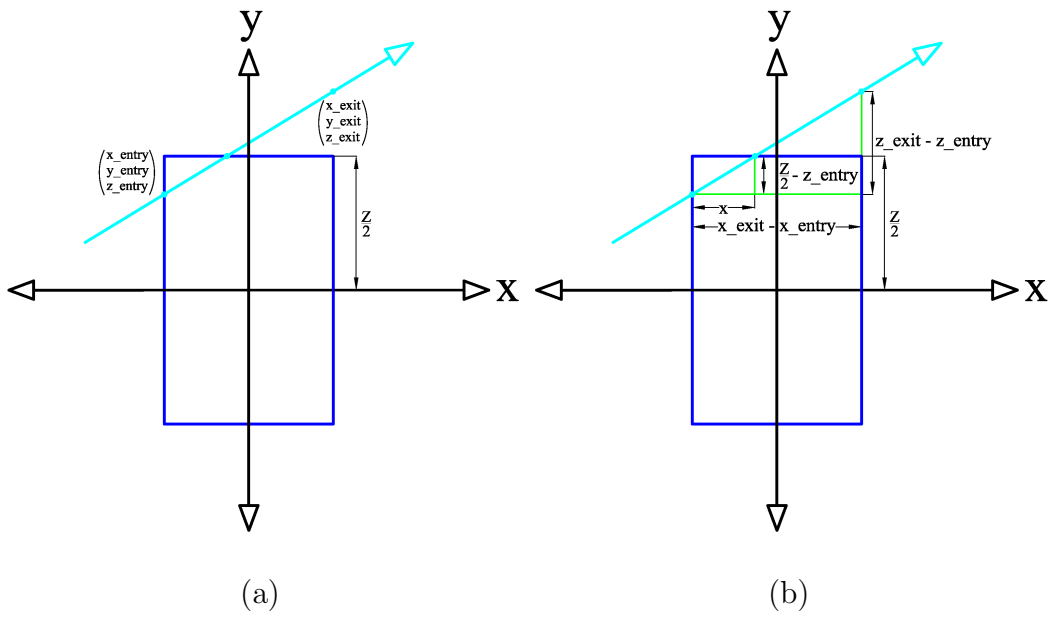


Fig. 3.1: (a) Proton exiting through the top face of the reconstruction cylinder as viewed in the $xz-$plane. (b) Visualizing the similar triangles used to determine the fraction of the path passing through the interior of the reconstruction volume.

similar triangles used to determine the value of Δx and calculate the $x-$coordinate of the point where the proton exits through the top face of the cylinder. From this drawing it is clear that Δx is given by:

$$\frac{\Delta x}{x_{exit} - x_{entry}} = \frac{{}^{Z}/_{2} - z_{entry}}{z_{exit} - z_{entry}} \implies \Delta x = (x_{exit} - x_{entry}) \left(\frac{{}^{Z}/_{2} - z_{entry}}{z_{exit} - z_{entry}} \right) \tag{3.22}$$

Therefore, the $x-$coordinate of the point where the proton exits the reconstruction

cylinder (x) is given by:

$$x = x_{entry} + \Delta x = x_{entry} + (x_{exit} - x_{entry}) \left(\frac{Z/2 - z_{entry}}{z_{exit} - z_{entry}} \right) \tag{3.23}$$

The geometry is no different when viewing the cylinder in the $yz-$plane, so we can determine the $y-$coordinate of the point where the proton exits the cylinder (y) in exactly the same way, yielding:

$$\frac{\Delta y}{y_{exit} - y_{entry}} = \frac{Z/2 - z_{entry}}{z_{exit} - z_{entry}} \tag{3.24}$$

$$\Delta y = (y_{exit} - y_{entry}) \left(\frac{Z/2 - z_{entry}}{z_{exit} - z_{entry}} \right) \tag{3.25}$$

$$\implies y = y_{entry} + \Delta y = y_{entry} + (y_{exit} - y_{entry}) \left(\frac{Z/2 - z_{entry}}{z_{exit} - z_{entry}} \right) \tag{3.26}$$

Therefore, the exit intersection point is redefined to be the point where the proton exits the top face of the reconstruction cylinder, which is given by $(x, y, Z/2)$:

$$x_{exit} = x = x_{entry} + \Delta x = x_{entry} + (x_{exit} - x_{entry}) \left(\frac{Z/2 - z_{entry}}{z_{exit} - z_{entry}} \right) \tag{3.27}$$

$$y_{exit} = y = y_{entry} + \Delta y = y_{entry} + (y_{exit} - y_{entry}) \left(\frac{Z/2 - z_{entry}}{z_{exit} - z_{entry}} \right) \tag{3.28}$$

$$z_{exit} = Z/2 \tag{3.29}$$

In general, the proton can exit through the top or bottom face of the cylinder, but the only difference is $Z/2$ is replaced by $-Z/2$ if the proton exits through the bottom face of the cylinder. Therefore, the general formula for the exit point when the proton

exits through the top or bottom face of the reconstruction cylinder is given by:

$$x_{exit} = x = x_{entry} + \Delta x = x_{entry} + (x_{exit} - x_{entry}) \left(\frac{{}^{\pm Z}/_{2} - z_{entry}}{z_{exit} - z_{entry}} \right) \tag{3.30}$$

$$y_{exit} = y = y_{entry} + \Delta y = y_{entry} + (y_{exit} - y_{entry}) \left(\frac{{}^{\pm Z}/_{2} - z_{entry}}{z_{exit} - z_{entry}} \right) \tag{3.31}$$

$$z_{exit} = {}^{\pm Z}/_{2} \tag{3.32}$$

$$\boxed{\begin{pmatrix} x_{exit} \\ y_{exit} \\ z_{exit} \end{pmatrix} = \begin{pmatrix} x_{entry} + (x_{exit} - x_{entry}) \left(\frac{{}^{\pm Z}/_{2} - z_{entry}}{z_{exit} - z_{entry}} \right) \\ y_{entry} + (y_{exit} - y_{entry}) \left(\frac{{}^{\pm Z}/_{2} - z_{entry}}{z_{exit} - z_{entry}} \right) \\ {}^{\pm Z}/_{2} \end{pmatrix}} \tag{3.33}$$

3.4 Binning the Histories for Statistical Cuts and FBP

Once all proton histories have been processed, protons that did not pass through the reconstruction volume have been discarded and the entry and exit points of the reconstruction volume have been determined for those that did. Protons that pass through approximately the same portion of the object can now be identified and grouped together so that the typical behavior of protons in these groups can be determined and used to discard errant data. Proton histories are binned (i.e. grouped) according to the $xy-$angle (θ_{bin}) of their path through the reconstruction volume, its lateral displacement (t_{bin}), and its vertical position (v_{bin}); essentially, protons are grouped according to their behavior in the $xy-$plane and their position in the vertical (z) direction, known as its slice [12]. Cumulatively, the histories are grouped into slices, where each slice contains proton histories from a range of z values with varying behavior in the $xy-$plane.

The size of the bins ($\theta_{binsize}, t_{binsize}, v_{binsize}$) are defined by the user and have an

impact on the quality of the statistics and reconstructed image. The resolution of the image increases as the bin size is decreased, but this also decreases the number of histories in each bin and can cause problems with the statistics calculations if the number of histories drops too low. Therefore, if we want to increase the resolution by decreasing the size of the bins, this requires a large enough increase in the total number of histories so that each bin will still contain enough histories to produce meaningful statistics. Since we are concerned with both the length of a scan and the amount of radiation the patient is exposed to, some compromise must be established for the desired resolution which keeps the required number of histories to a reasonable amount.

Figure 3.2 gives a graphical representation of the binning process. The path is defined to be the straight line connecting the entry and exit points of the reconstruction volume (see Figure 3.4; notice that $(x_{entry}, y_{entry}, z_{entry})$ and $(x_{exit}, y_{exit}, z_{exit})$ have been relabeled (x_i, y_i, z_i) and (x_o, y_o, z_o) in the figure (corresponding to $i = in$ and $o = out$), respectively, to save space. Binning the histories proceeds by first calculating the $xy-$angle (θ_{path}) of the path through the reconstruction volume and determining which angular bin (θbin) is closest to this angle based on the size $\theta_{binsize}$ of the angular bins:

$$\theta_{path} = \arctan\left(\frac{y_{exit} - y_{entry}}{x_{exit} - x_{entry}}\right) \tag{3.34}$$

$$\boxed{\implies \theta_{bin} = \left\lfloor \frac{\theta_{path}}{\theta_{binsize}} + 0.5 \right\rfloor} \tag{3.35}$$

The angular bin (θ_{bin}) is then used to define the $utv-$coordinate system corresponding to this proton history. A lateral displacement (t_{path}) must then be assigned to this

path and used to determine the lateral bin (t_{bin}) for this history (see Figure 3.4. However, since the angular bin (θ_{bin}) is not equal to the actual angle (θ_{path}) of the path (i.e. $\theta_{bin} \neq \theta_{path}$), the lateral displacement is not constant along the path, so the distance which best describes the lateral displacement (t_{path}) of the path must be determined. The arithmetic mean lateral displacement ($\bar{t}$) produces the smallest root-mean-square (RMS) deviation between the two paths, which could be calculated directly, but this would require considerable effort (e.g. calculus). Fortunately, an equivalent and computationally inexpensive method is to calculate the arithmetic mean coordinate along the path in each direction ($(\bar{x}, \bar{y}, \bar{z})$), which turns out to be the midpoint ($(x_{mid}, y_{mid}, z_{mid})$) of the path, and determine the lateral displacement (t_{path}) at this point (see Figure 3.4:

$$x_{mid} = \frac{x_{entry} + x_{exit}}{2} \tag{3.36}$$

$$y_{mid} = \frac{y_{entry} + y_{exit}}{2} \tag{3.37}$$

$$z_{mid} = \frac{z_{entry} + z_{exit}}{2} \tag{3.38}$$

The lateral displacement (t_{path}) of the midpoint is equivalent to determining the length of the line between the midpoint ($(x_{mid}, y_{mid}, z_{mid})$) and the $u-$axis and is perpendicular to the $u-$axis. While this could be calculated by determining the equation of the line perpendicular to the $u-$axis through the midpoint and then finding its length, it is much easier to rotate the midpoint ($(x_{mid}, y_{mid}, z_{mid})$) into the $utv-$coordinate system since the lateral displacement (t_{path}) is simply the t-coordinate in this coordinate system. The coordinate transformation rotating points in the $xyz-$coordinate system into the $utv-$coordinate system is accomplished using the rotation matrix that rotates through a negative angle about the $z-$axis. However,

this same operation can be performed using the rotation matrix that rotates through a positive angle about the $z-$axis, which we used earlier and is given by 3.9, by simply using a negative angle in the calculations. Additionally, since we are only interested in the $t-$coordinate from this transformation, we can extract the row corresponding to this coordinate and avoid the other unnecessary calculations as follows:

$$t_{path} = x \sin(-\theta_{bin}) + y \cos(-\theta_{bin}) \tag{3.39}$$

$$\boxed{t_{path} = y \cos(\theta_{bin}) - x \sin(\theta_{bin})} \tag{3.40}$$

The proton histories must now be binned according to their vertical position, corresponding to the particular slice (i.e. v_{bin}) they belong to. Since each slice corresponds to a plane with thickness defined by $v_{binsize}$, we must determine the $z-$coordinate which best describes each path through the reconstruction volume. The vertical displacement which produces the smallest RMS deviation between the path of the proton and the $ut-$axis is the arithmetic mean z along this path is the $z-$coordinate of the midpoint of the path, i.e., z_{mid}. We can now determine the lateral displacement bin (t_{bin}) and the vertical slice (v_{bin}) using the lateral displacement (t_{path}) and the vertical position (z_{mid}), respectively, as follows:

$$t_{bin} = \left\lfloor \frac{t_{path}}{t_{binsize}} + 0.5 \right\rfloor \tag{3.41}$$

$$v_{bin} = \left\lfloor \frac{z_{mid}}{v_{binsize}} + 0.5 \right\rfloor \tag{3.42}$$

The proton histories can now be placed in the bin corresponding to their angular bin (θ_{bin}), lateral displacement (t_{bin}), and vertical bin/slice (v_{bin}). Ideally there will be 100 or more histories in each bin, but this might not be the case for all bins, but the bins that have less than 100 histories are typically those associated with paths

towards the edge of the image and therefore, have little to no effect on the ultimate quality of the reconstructed image. The histories in each bin can now be compared to each other to determine what range of behavior is acceptable and how errant data is identified and discarded. Now that the angular and lateral bins have been determined,

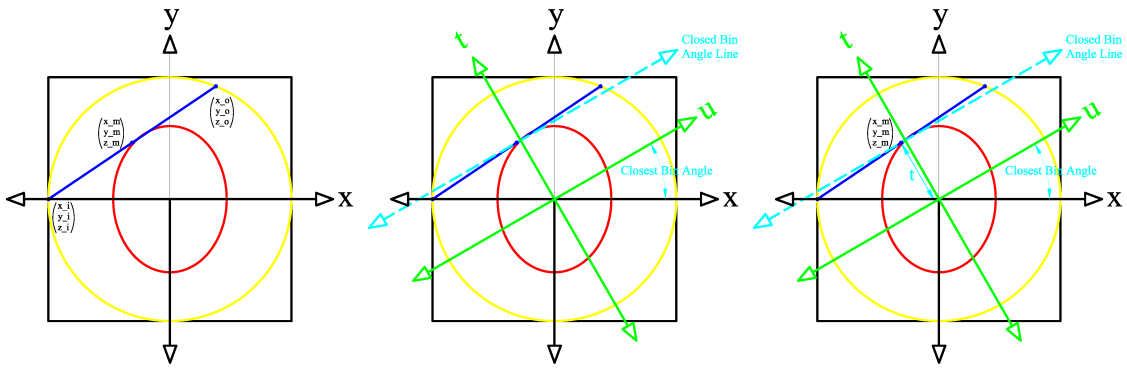


Fig. 3.2: Graphical representation of the process by which proton histories are assigned to angular (θ) and lateral (t) bins; the square represents the boundary of the image, the yellow circle represents the boundary of the reconstruction volume, and the red ellipse represents the boundary of an idealized phantom.(a) Approximating the path of the proton through the object by connecting the entry (x_i, y_i, z_i) and exit (x_o, y_o, z_o) points of the reconstruction volume with a straight line (in cyan) and calculating the midpoint (x_m, y_m, z_m) along this line.(b) Calculating the angle of the approximate proton path and assign this proton history to the closest angular bin. This angle is then used to define the $utv-$coordinate axes and the line parallel to the $u-$axis through the midpoint of the line. (c) Determine the distance between the $u-$axis and the parallel line through the midpoint of the line to define the $t-$value for this proton history.

the path has now been defined in the $xy-$plane.

3.5 Performing WEPL and Relative Angle Statistics for Each Bin

Performing statistical calculations is a reasonably straight forward process now that proton histories traversing similar portions of the reconstruction volume have been grouped together. The assumption is that protons that pass through the same portion of the object encounter roughly the same structures and material regardless of which direction it was headed and therefore behave similarly. An additional assumption is that a proton that goes through the object from left to right will behave the same as a proton passing through the same part of the object in the opposite direction. Although this is not entirely true because scattering is slightly dependent on the proton's energy, the energy dependence is small enough that this will not significantly affect the accuracy of the calculations.

The proton histories we would like to discard are those that either yielded a significantly different WEPL measurement than the other histories in its bin experiences or experienced a large angle scattering event [12]. Since WEPL measurements are used to assign a relative stopping power (RSP) to each voxel in the reconstructed image, it is important to discard histories that experienced some type of measurement error. Large angle scattering events typically result in considerable energy loss, but this is not the reason these histories are discarded since these histories will be discarded by the WEPL cuts. Recall that we approximate proton paths using a straight line, so if a proton collides with a nucleus and is scattered at a large angle, its path will be significantly different than a straight line. Therefore, the voxels this proton actually passes through are not correctly identified, so to avoid its WEPL measurement from being used to assign RSP values to the wrong voxels, these histories are simply

discarded.

To determine the typical behavior of protons in each bin, we calculate the arithmetic mean WEPL measurement and its scattering angle in both the horizontal and vertical directions. The scattering angle is defined to be the difference between the trajectory of the proton before and after passing through the object, i.e., the relative angle $\Delta\theta = \theta_{exit} - \theta_{entry}$. Since the scattering in the horizontal and vertical directions can be considered independent statistical processes, these are analyzed separately. There are N_{bin} proton histories in each bin, so the mean (μ) and standard deviation (σ) of each bin's WEPL, relative angle in the $xy-$plane, and relative angle in the $xz-$plane are calculated as follows:

$$\mu_{WEPL} = \frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} WEPL_i \tag{3.43}$$

$$\mu_{\Delta\theta_{xy}} = \frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} \Delta\theta_{xy_i} \tag{3.44}$$

$$\mu_{\Delta\theta_{xz}} = \frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} \Delta\theta_{xzi} \tag{3.45}$$

$$\sigma^2_{WEPL} = \frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} \left(\mu_{WEPL} - WEPL_i\right)^2 \tag{3.46}$$

$$\sigma^2_{\Delta\theta_{xz}} = \frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} \left(\mu_{\Delta\theta_{xy}} - \Delta\theta_{xy_i}\right)^2 \tag{3.47}$$

$$\sigma^2_{\Delta\theta_{xz}} = \frac{1}{N_{bin}} \sum_{i=1}^{N_{bin}} \left(\mu_{\Delta\theta_{xz}} - \Delta\theta_{xzi}\right)^2 \tag{3.48}$$

The mean (μ_{bin}) and standard deviation (σ_{bin}) of the three measurements associated with each bin can then be used to identify and discard proton histories that behave significantly differently than the other histories in that bin.

3.6 Performing Statistical Cuts

When a proton experiences a large angle scattering event or some detector related error occurs, the measurements acquired for that proton history will be inaccurate and should be removed from the data set. There is no obvious method for identifying these histories directly, so we compare histories with similar paths to each other to infer which of these likely encountered some source of error and should be discarded. Outliers are discarded by comparing the measurement to the mean of that bin and discarding those with an unacceptably large difference [12]. For our purposes, an unacceptably large difference is three or more standard deviations. Therefore, a proton history is discarded if any of the following three relations are true:

1. $(\mu_{WEPL} - WEPL_i) \geq 3\sigma_{WEPL}$
2. $\left(\mu_{\Delta\theta_{xy}} - \Delta\theta_{xy_i}\right) \geq 3\sigma_{\Delta\theta_{xz}}$
3. $(\mu_{\Delta\theta_{xz}} - \Delta\theta_{xzi}) \geq 3\sigma_{\Delta\theta_{xz}}$

Although this process may not identify every proton history that encountered some type of measurement error, those that remain will have little impact on the accuracy of the reconstructed image since its measurements were still within an acceptably large range around the mean. Therefore, at this point the errant data has been removed and we are now in a position to approximate the boundary of the actual object and use this to construct the A matrix needed for reconstruction.

3.7 Constructing the Sinogram

We currently use a common image reconstruction technique called filtered backprojection (FBP) to produce an initial approximation of the object and its boundaries. FBP proceeds by first constructing images called a *sinogram*, which is a frequency space representation of the proton history data. Fortunately, the sinogram has the exact same structure as the bins we have already constructed to perform statistical cuts; i.e., a sinogram is a two dimensional plot of projection angle (θ_{xy}) versus displacement (t) for a particular range of vertical displacements (z/v) where each point represents the value of the integral (i.e. WEPL) along the line defined by that projection angle and displacement. Therefore, we must construct several sinograms, one for each slice, to represent the entire reconstruction volume. To produce these sinograms, we need only recalculate the mean WEPL for each bin now that the errant data has been discarded. Figure 3.3 shows an example of a sinogram.

3.8 Performing Filtered Backprojection (FBP)

The sinograms can now be used to reconstruct an image of each slice of the reconstruction volume, where the value of each voxel in the image represents the RSP of this portion of the object [12]. FBP is a common and popular image reconstruction technique which exploits the properties of the *Radon transform* to determine the properties of the object using the path and WEPL measurements [13, 14]. The Radon transform (in two dimensions) is an integral transform which takes as input a continuous function $f(x, y)$ and produces as output the continuous function $Rf(L)$ defined on the space of all possible lines $L(\theta, t) \in \mathbb{R}^2$ by the line integral of $f(x, y)$

Fig. 3.3: Example of a sinogram

along each such line. Likewise, the inverse Radon transform takes as input the continuous function $Rf(L)$ on the space of all possible lines $L(\theta, t) \in \mathbb{R}^2$ and produces the continuous function $f(x, y)$ as output. Each WEPL measurement corresponds to the line integral of the function $RSP(x, y)$ along a particular proton path $L(\theta, t)$, so a sinogram represents the Radon transform ($Rf(L)$) of the target object $RSP(x, y)$ for a particular slice of the reconstruction volume. Therefore, the goal of FBP is to reconstruct the object image $RSP(x, y)$ from each sinogram (i.e. Radon transform), which corresponds to performing the inverse Radon transform.

However, notice that the Radon transform is theoretically a continuous function of integrals of the function $f(x, y)$ along *all* possible lines $L(\theta, t) \in \mathbb{R}^2$, but the WEPL measurements obtained during a scan correspond to only a finite number of proton paths, so the sinogram yielded from the experimental data is not continuous like the

theory requires. Consequently, the sinogram is a discrete Radon transform of the object and the inverse Radon transform will need to be modified so it can be applied to this discrete data. Since these calculations will be performed on a computer, there will also be some approximations that will need to be made to produce calculable equations. The reconstructed image yielded from the discrete inverse Radon transform will still be a good approximation as long as the number of proton histories is fairly large, the proton paths are well distributed spatially, and the size of the bins and voxels are relatively small. In fact, as the bins gets smaller, the sinogram approaches a continuous function and the maximum image resolution increases. Unfortunately, smaller bins will contain fewer proton histories and this produces less reliable statistics, so the bins can not be made arbitrarily small. For practical diagnostic medical imaging purposes, voxel sizes rarely need to be smaller than 1 mm $\times$ 1 mm $\times$ 1 mm, so high quality images can be obtained using horizontal/vertical bin sizes of 1 mm and angular bin sizes of $1^\circ - 4^\circ$.

Another consequence of a discrete sinogram is that the binning procedure produces discontinuities in the WEPL profile between each bin and this can cause artifacts in the reconstructed image. This is equivalent to high frequency contamination and removing it can be accomplished with the same filtering techniques used in analog electrical circuit design. The most common filtering technique uses the Ram Lak filter (also known as the ramp filter) or some variation of it, such as the Shepp-Logan filter, which multiplies the Ram Lak filter by the sinc function. By filtering the sinogram prior to backprojection, the high frequency artifacts can be avoided and the quality of the reconstructed image improves significantly.

The image produced by FBP is then used to determine which voxels are part of the object and which are not based on each voxel's RSP value. A voxel with an RSP above some threshold value is assumed to be part of the object and those below the threshold are assumed to lie outside the object [12]. There is no analytical method for determining an appropriate threshold on RSP, but experience has shown that a threshold of $RSP = 0.6$ is an effective choice for accurately defining the object hull and its boundaries. Accurately determining the boundaries of the object is important for determining the most likely path (MLP) of each proton because MLP calculations assume that the path of the proton only deviates from a straight line inside the object. Consequently, the path of the proton can only be determined accurately if we can correctly identify where the proton entered and exited the object, which requires accurate knowledge of the object and its boundaries.

The purpose of the MLP calculations is to determine which voxels were intersected by a particular proton and the distance this proton traveled through each of these voxels, since this is the information required to construct the system matrix A. In addition, the FBP image is also used to define the initial iterate of the image vector x for iterative image reconstruction process. Since the FBP image is used to construct the system matrix A and image vector x, thereby preconditioning the linear system $Ax = b$, accurately determining this image is perhaps the most important aspect of preprocessing because its accuracy directly affects the accuracy and convergence properties of iterative image reconstruction. Although FBP is only being used here to determine the boundary of the object and precondition the linear system, it should be noted that FBP is actually a full image reconstruction technique itself and can

be applied iteratively to improve the quality of reconstructed images. However, the importance of this aspect of preprocessing is the reason why a sophisticated technique like FBP is being used simply to determine the boundary of the object.

3.9 Calculating the MLP for Each Valid Proton History

Up until this point the path of the proton has been approximated by a straight line, but now that the object hull and its boundaries have been determined, we are now in a position to determine the most likely path (MLP) of each proton through the object [6]. The path of the proton deviates from a straight line as a result of the various types of scattering events that occur as it travels downstream. Although most of these scattering events only deflect the proton's trajectory slightly, the sum of all of these deflections produces a zig zag path with a random angular dispersion and lateral/vertical displacement; the angular dispersion is a statistical process which can be modeled as a Gaussian. The probability that a proton will encounter an atom and be deflected depends on the density of the medium it passes through. Therefore, the total amount of scattering depends on the density of the medium and its thickness. Since air has a very low density of atoms, there is a low probability that a proton will be scattered as it travels through air. On the other hand, although the SSD tracking detectors have a reasonably high density, the proton will not be deflected much inside the tracking detectors because they are quite thin. Consequently, nearly all of the proton's scattering will occur inside the object itself and the proton's path can be approximated by a straight line outside of the object. Therefore, the first objective of the MLP procedure is to determine where the proton entered and exited the object.

Fortunately, the points where the proton intersected the reconstruction volume and its trajectory at these points have already been determined, so we need only use these to calculate the path of the proton prior to entering ($y_{forward}$ and $z_{forward}$) and after exiting ($y_{backward}$ and $z_{backward}$) the reconstruction volume and then project these paths towards the object boundary until a nonzero voxel is encountered in the thresholded FBP image. Approximating these paths as straight lines (i.e. $f(x) = m_x x + b$) yields:

$$
\begin{aligned}
m_{y_{entry}} &= \tan\left(\theta_{xy_{entry}}\right) \\
b_{y_{entry}} &= y_{entry} - m_{y_{entry}} x_{entry} \\
y_{forward} &= m_{y_{entry}} x + b_{y_{entry}} = m_{y_{entry}} x + \left(y_{entry} - m_{y_{entry}} x_{entry}\right) \\
y_{forward} &= m_{y_{entry}} \left(x - x_{entry}\right) + y_{entry} \\
\boxed{y_{forward} = \tan\left(\theta_{xy_{entry}}\right) \left(x - x_{entry}\right) + y_{entry}} & \qquad (3.49)
\end{aligned}
$$

$$
\begin{aligned}
m_{z_{entry}} &= \tan\left(\theta_{xz_{entry}}\right) \\
b_{z_{entry}} &= z_{entry} - m_{z_{entry}} x_{entry} \\
z_{forward} &= m_{z_{entry}} x + b_{z_{entry}} = m_{z_{entry}} x + \left(z_{entry} - m_{z_{entry}} x_{entry}\right) \\
z_{forward} &= m_{z_{entry}} \left(x - x_{entry}\right) + z_{entry} \\
\boxed{z_{forward} = \tan\left(\theta_{xz_{entry}}\right) \left(x - x_{entry}\right) + z_{entry}} & \qquad (3.50)
\end{aligned}
$$

$$m_{y_{exit}} = \tan\left(\theta_{xy_{exit}}\right)$$

$$b_{y_{exit}} = y_{exit} - m_{y_{exit}} x_{exit}$$

$$y_{backward} = m_{y_{exit}} x + b_{y_{exit}} = m_{y_{exit}} x + \left(y_{exit} - m_{y_{exit}} x_{exit}\right)$$

$$y_{backward} = m_{y_{exit}} \left(x - x_{exit}\right) + y_{exit}$$

$$\boxed{y_{backward} = \tan\left(\theta_{xy_{exit}}\right)\left(x - x_{exit}\right) + y_{exit}} \tag{3.51}$$

$$m_{z_{exit}} = \tan\left(\theta_{xz_{exit}}\right)$$

$$b_{z_{exit}} = z_{exit} - m_{z_{exit}} x_{exit}$$

$$z_{backward} = m_{z_{exit}} x + b_{z_{exit}} = m_{z_{exit}} x + \left(z_{exit} - m_{z_{exit}} x_{exit}\right)$$

$$z_{backward} = m_{z_{exit}} \left(x - x_{exit}\right) + z_{exit}$$

$$\boxed{z_{backward} = \tan\left(\theta_{xz_{exit}}\right)\left(x - x_{exit}\right) + z_{exit}} \tag{3.52}$$

Now that the equations defining the paths of the proton outside the object have been derived, these are projected towards the object boundary to determine where they intersect the object and where the proton path begins to deviate from a straight line [12]. We do not have an equation defining the boundary of the object like we did for the reconstruction volume boundary, so we must walk along the paths until a nonzero voxel is encountered. To make sure the *edge* of the *first* nonzero voxel along each path is found, the voxel walk is not composed of constant length steps, which could potentially miss the first nonzero voxel (if the intersection length of this voxel is smaller than the step) or fail to identify the edge of the first voxel as the intersection point (if the step lands inside the first voxel). This variable step length voxel walk routine is a variation of the digital difference analyzer (DDA) incorporating aspects of Bresenham's line algorithm to extend the routine to three dimensions and determine

the coordinates (x, y, z) of the points where the line intersects each voxel. Since we know the object lies inside the reconstruction volume, we can begin walking along the paths at the points (x_{entry} and x_{exit}) where each path intersects the reconstruction volume. Starting at these points, the voxel walk routine is used to follow each path and determine the entry and exit points of the object. MLP calculations can then be performed to determine the most likely path of the proton between these two points inside the object and use these paths to construct the A matrix.

The derivation of the MLP calculations proceeds by assuming that the object is composed of a homogenous material and use the trajectory of the proton prior to entering and after exiting the object to determine the most likely path of the proton through the object [7]. These calculations do not produce a closed form solution for the equation of the proton's path, they take a depth u inside the object as input and produce the t or v coordinate corresponding to this depth as output. Therefore, the path of the proton is determined by taking small steps along the proton's path and using the u coordinate at each step to calculate the corresponding t and v coordinates. The scattering in the t and v directions are statistically independent processes and are calculated separately, but it is important that these calculations are performed in a coordinate system in which the proton's depth u increases monotonically. This is important because because the path of the proton is curved, so there can be several points along the proton's path which have the same x coordinate. Since MLP calculations are deterministic, an input depth u will always produce the same t and v coordinates even if there are multiple points along the proton's path with that same $u-$coordinate. The easiest way to avoid these complications is to perform the MLP

calculations in a rotated coordinate system which is defined by the trajectory of the proton as it enters the object. This guarantees that there are no two points along the proton's path with the same u coordinate and each point can be determined uniquely.

$$\sigma_{t_1}^2 = E_0^2 \left[1 + 0.038 \log\left(\frac{u_1 - u_0}{X_0}\right)\right]^2 \int_{u_0}^{u_1} \frac{(u_1 - u)^2}{\beta^2(u)p^2(u)} \frac{du}{X_0} \tag{3.53a}$$

$$\sigma_{t_1\theta_1}^2 = E_0^2 \left[1 + 0.038 \log\left(\frac{u_1 - u_0}{X_0}\right)\right]^2 \int_{u_0}^{u_1} \frac{u_1 - u}{\beta^2(u)p^2(u)} \frac{du}{X_0} \tag{3.53b}$$

$$\sigma_{\theta_1}^2 = E_0^2 \left[1 + 0.038 \log\left(\frac{u_1 - u_0}{X_0}\right)\right]^2 \int_{u_0}^{u_1} \frac{1}{\beta^2(u)p^2(u)} \frac{du}{X_0} \tag{3.53c}$$

$$\Sigma_1 = \begin{pmatrix} \sigma_{t_1}^2 & \sigma_{t_1\theta_1}^2 \\ \sigma_{t_1\theta_1}^2 & \sigma_{\theta_1}^2 \end{pmatrix} \tag{3.54}$$

$$\sigma_{t_2}^2 = E_0^2 \left[1 + 0.038 \log\left(\frac{u_2 - u_1}{X_0}\right)\right]^2 \int_{u_1}^{u_2} \frac{(u_2 - u)^2}{\beta^2(u)p^2(u)} \frac{du}{X_0} \tag{3.55a}$$

$$\sigma_{t_2\theta_2}^2 = E_0^2 \left[1 + 0.038 \log\left(\frac{u_2 - u_1}{X_0}\right)\right]^2 \int_{u_1}^{u_2} \frac{u_2 - u}{\beta^2(u)p^2(u)} \frac{du}{X_0} \tag{3.55b}$$

$$\sigma_{\theta_2}^2 = E_0^2 \left[1 + 0.038 \log\left(\frac{u_2 - u_1}{X_0}\right)\right]^2 \int_{u_1}^{u_2} \frac{1}{\beta^2(u)p^2(u)} \frac{du}{X_0} \tag{3.55c}$$

$$\Sigma_2 = \begin{pmatrix} \sigma_{t_2}^2 & \sigma_{t_2\theta_2}^2 \\ \sigma_{t_2\theta_2}^2 & \sigma_{\theta_2}^2 \end{pmatrix} \tag{3.56}$$

Notice that each term of the scattering matrices depend on the values of $\beta(u)$ and $p(u)$, which both depend on the velocity of the proton at a depth u inside the object. There is no method for determining the velocity of the proton at intermediate depths

inside the object, so GEANT4 simulations are used to produce a fifth order polynomial fit to the common term $\beta^2(u)p^2(u)$ as follows:

$$\frac{1}{\beta^2(u)p^2(u)} = a_0 + a_1 u + a_2 u^2 + a_3 u^3 + a_4 u^4 + a_5 u^5 \tag{3.57}$$

The coefficients of this polynomial fit for a 200 MeV proton are listed in Table 3.1:

The current depth u_1 is calculated after each step and used to calculate the corresponding t_1-coordinate using the following equation:

Tab. 3.1: Coefficients of polynomial fit

Coefficients, a_i	Value
a_0	7.457×10^{-6}
a_1	4.548×10^{-7}
a_2	-5.777×10^{-8}
a_3	1.301×10^{-8}
a_4	-9.228×10^{-10}
a_5	2.687×10^{-11}

$$\boxed{y_{MLP} = \begin{pmatrix} t_1 \\ \\ \theta_1 \end{pmatrix} = \left(\Sigma_1^{-1} + R_1^T \Sigma_2^{-1} R_1\right)^{-1} \left(\Sigma_1^{-1} R_0 y_0 + R_1^T \Sigma_2^{-1} y_2\right)} \tag{3.58}$$

where the entry point (y_0), exit point (y_2), and R matrices are defined as:

$$y_0 = \begin{pmatrix} t_0 \\ \theta_0 \end{pmatrix} \tag{3.59}$$

$$y_2 = \begin{pmatrix} t_2 \\ \theta_2 \end{pmatrix} \tag{3.60}$$

$$R_0 = \begin{pmatrix} 1 & u_1 - u_0 \\ 0 & 1 \end{pmatrix} \tag{3.61}$$

$$R_1 = \begin{pmatrix} 1 & 0 \\ u_2 - u_1 & 1 \end{pmatrix} \implies R_1^T = \begin{pmatrix} 1 & u_2 - u_1 \\ 0 & 1 \end{pmatrix} \tag{3.62}$$

Similarly, the $v-$coordinate after each step is calculated by replacing the various $t-$coordinates in the MLP formulas with the corresponding $v-$coordinates:

$$\boxed{z_{MLP} = \begin{pmatrix} v_1 \\ \theta_1 \end{pmatrix} = \left(\Sigma_1^{-1} + R_1^T \Sigma_2^{-1} R_1\right)^{-1} \left(\Sigma_1^{-1} R_0 z_0 + R_1^T \Sigma_2^{-1} z_2\right)} \tag{3.63}$$

where the entry point (z_0) and exit point (z_2) are defined as:

$$z_0 = \begin{pmatrix} v_0 \\ \theta_0 \end{pmatrix} \tag{3.64}$$

$$z_2 = \begin{pmatrix} v_2 \\ \theta_2 \end{pmatrix} \tag{3.65}$$

Once the points along the path of the proton have been calculated, they can be used to determine which voxels were intersected and the length of the path (i.e. chord length) through each of these voxels so the system matrix A can be constructed.

3.10 Construct $\boldsymbol{A}$, $\boldsymbol{x}$, and $\boldsymbol{b}$ Matrices

The A matrix (3.66) contains the path information from each proton history, where each row i corresponds to a particular proton history and each column j corresponds

to a single voxel in the reconstruction volume [12]. Values are assigned to the A matrix such that, if the proton associated with proton history i passes through voxel j, then the element a_{ij} of A is assigned the chord length of this intersection [6]. Although the proton intersects voxels outside of the object, these do not need to be included in the A matrix, which saves some memory and reduces the number of calculations that need to be performed in the subsequent iterative image reconstruction. These can be excluded because $RSP = 0$ for every voxel outside the object, so the multiplication Ax performed during image reconstruction is unaffected by the value of a_{ij} since $a_{ij}xj = 0$ for all $x_j = 0$, regardless of the value of a_{ij}. Since there is no additional information provided by assigning values to the a_{ij} associated with intersections outside the object, these can simply be ignored, thereby increasing preprocessing and image reconstruction performance. In addition, since the image reconstruction process converts these matrices into sparse matrix format to decrease time and space complexity, excluding these elements from the A matrix also reduces the number of elements that need to be maintained in memory and consequently, the number of calculations that need to be performed during reconstruction.

$$A = \begin{pmatrix} a_{00} & a_{01} & a_{02} & \cdots & a_{0N} \\ a_{10} & a_{11} & a_{12} & \cdots & a_{1N} \\ a_{20} & a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{M0} & a_{M1} & a_{M2} & \cdots & a_{MN} \end{pmatrix} \tag{3.66}$$

Notice that in constructing the linear system $Ax = b$, the three dimensional image array x is represented by a one dimensional vector instead, so the elements of x must

be mapped appropriately. A visual representation of this mapping is shown in 3.67, where element x_{mno} is mapped according to the relation $x_{mno} \rightarrow x_{n+mN+oMN}$. Each box corresponds to a particular slice o and encloses the elements corresponding to that slice.

$$\left(\begin{array}{|cccc|}\hline a_{0,0} & a_{0,1} & \cdots & a_{0,(X-1)} \\ a_{1,0} & a_{1,1} & \cdots & a_{0,(X-1)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{(Y-1),0} & a_{(Y-1),1} & \cdots & a_{(Y-1),(X-1)} \\ \hline\end{array}\right.$$
$$\left.\begin{array}{|cccc|}\hline a_{Y,0} & a_{Y,1} & \cdots & a_{Y,(X-1)} \\ a_{(Y+1),0} & a_{(Y+1),1} & \cdots & a_{(Y+1),(X-1)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{(2Y-1),0} & a_{(2Y-1),1} & \cdots & a_{(2Y-1),(X-1)} \\ \hline\end{array}\right)$$
$$\left(\begin{array}{|cccc|}\hline a_{2Y,0} & a_{2Y,1} & \cdots & a_{2Y,(X-1)} \\ a_{(2Y+1),0} & a_{(2Y+1),1} & \cdots & a_{(2Y+1),(X-1)} \\ \vdots & \vdots & \ddots & \vdots \\ a_{(3Y-1),0} & a_{(3Y-1),1} & \cdots & a_{(3Y-1),(X-1)} \\ \hline\end{array}\right.$$
$$\left.\begin{array}{|cccc|}\hline \vdots & \vdots & \ddots & \vdots \\ a_{\{(Z-1)Y,0\}} & a_{\{(Z-1)Y,1\}} & \cdots & a_{\{(Z-1)Y,X-1\}} \\ a_{\{(Z-1)Y+1,0\}} & a_{\{(Z-1)Y+1,1\}} & \cdots & a_{\{(Z-1)Y+1,X-1\}} \\ \vdots & \vdots & \ddots & \vdots \\ a_{\{ZY-1,0\}} & a_{\{ZY-1,1\}} & \cdots & a_{\{ZY-1,X-1\}} \\ \hline\end{array}\right)$$
$$\xrightarrow{x}$$
$$\left(\begin{array}{|c|}\hline a_0 \\ a_1 \\ \vdots \\ a_{(Y-1),(X-1)} \\ \hline\end{array}\right.$$
$$\left.\begin{array}{|c|}\hline a_{Y,0} \\ a_{Y,1} \\ \vdots \\ a_{(2Y-1),(X-1)} \\ \hline\end{array}\right)$$
$$\left(\begin{array}{|c|}\hline a_{2Y,0} \\ a_{2Y,1} \\ \vdots \\ a_{(3Y-1),(X-1)} \\ \hline\end{array}\right.$$
$$\left.\begin{array}{|c|}\hline \vdots \\ a_{\{(Z-1)Y,0\}} \\ a_{\{(Z-1)Y,1\}} \\ \vdots \\ a_{\{ZY-1,X-1\}} \\ \hline\end{array}\right) \quad (3.67)$$

Once the path information has been processed and the A matrix has been constructed, the linear system $Ax = b$ has now been defined since the initial iterate of the image vector x corresponds to the image produced by FBP (pre thresholding) and the vector b corresponds to the WEPL measurements associated with each proton history, where element b_i corresponds to the WEPL measurement from proton history i. These three matrices are then provided as input to the image reconstruction routine, which requires writing them to file at this point in pCT development since the data acquisition, preprocessing, and image reconstruction phases of pCT have not been incorporated into a single routine yet. In the future, these three phases will be combined into a single program, removing the need to write intermediate data to file and improving computational performance.

4. NEW PREPROCESSING DEVELOPMENTS

Preprocessing is an extremely important part of pCT and there are several aspects of this process which have considerable room for improvement. Perhaps the most important step in preprocessing is the process of hull detection, which is currently performed by thresholding the images produced by FBP. Although FBP has proved to be a reasonably accurate and effective tool for determining the object and its boundaries in the past, it is quite computationally expensive and may no longer be an appropriate hull detection technique for use in the sparse matrix reconstruction algorithms which are currently being developed. In an effort to increase the computational efficiency and accuracy of image reconstruction, recent efforts have focused on exploiting the memory and computational benefits of storing matrices in sparse matrix format.

Since there are a large number of voxels in the image vector x which do not correspond to the object and a proton only passes through a small number of voxels, the system matrix A and image vector x both contain a significant number of zeros add although these zeros add no information to reconstruction, they have still been maintained in memory in the past. However, storing A and x in sparse matrix format allows these elements to be ignored, thereby reducing the number of elements that need to be maintained in memory and consequently, the number of calculations

that need to be performed during reconstruction. An important consequence is that every voxel of the object must be identified during preprocessing because elements with value zero are not maintained in memory and do not enter into reconstruction calculations, so these elements cannot be recovered during reconstruction and will remain zero. Therefore, if hull detection fails to identify a voxel of the object during preprocessing, this voxel will remain zero after reconstruction and will appear as a whole in the reconstructed image. If voxels outside the object are maintained during reconstruction, although they actually have an RSP of zero, some will be assigned a nonzero RSP. Since RSP values are assigned by assuming the voxels along a particular proton path each contribute equally to the associated WEPL measurement, including voxels which actually did not contribute underestimates the contributions of the other voxels and decreases their RSP value. Therefore, ignoring voxels outside the object during reconstruction prevents these voxels from being assigned a nonzero value and assigns RSP values to voxels inside the object more accurately.

Although it is essential to identify every voxel corresponding to a part of the object, it is also important that hull detection produces an image with as few voxels from outside the object as possible since maintaining unnecessary voxels reduces the benefits of sparse matrix reconstruction. Therefore, the goal of the research presented in this thesis was to develop alternative hull detection techniques and compare these to FBP in terms of their computational performance and their ability to correctly identify voxels inside the object and ignore voxels outside the object. The data used to perform these comparisons included both simulated and experimental data

4.1 Alternative Hull Detection Techniques: Theory and Implementation

One of the requirements for a clinically viable imaging technique is its ability to produce a reconstructed image in a reasonably small amount of time. Since pCT is a challenging and computationally demanding imaging process and some aspects of reconstruction have little to no room for decreasing computation time, it is absolutely essential to identify aspects of preprocessing which can be simplified or performed more efficiently. Although FBP is a well developed and robust imaging technique, because it is designed to produce a fully reconstructed image, it is an overly complex and computationally expensive technique for simply determining the boundary of an object. Recall that FBP also requires a substantial amount of preliminary calculations, such as binning and statistical analysis. Although many of these calculations are also used for identifying and removing errant data, it has been suggested that these processes could be simplified and may not be necessary in the future. Therefore, the hull detection techniques presented here have been designed to operate independently and without any of the supplemental data required for FBP. While these techniques could potentially benefit from this information, the goal of this research was to demonstrate proof of concept for several techniques since it is unclear what approach is most appropriate for satisfying the various hull detection requirements, especially those arising from the new sparse matrix reconstruction algorithms. Three hull detection techniques were developed, each with the minimum possible complexity, although one of these is a variant of another. These three hull detection techniques are space carving (SC), modified space carving (MSC), and space modeling (SM).

4.1.1 Space Carving (SC)

Space carving (SC) is a simple and computationally inexpensive hull detection technique which produces a convex hull by identifying and excluding voxels lying outside of the object, similar to the way a sculpture is chiseled from a block of material [15, 16, 17]. The block of material in this case is the reconstruction volume, so space carving begins with the assumption that the object fills the reconstruction volume completely. Protons that miss the object entirely, which are identified from their WEPL measurements, are then used to exclude voxels along the proton's path through the reconstruction volume [18]. Repeating this for all such proton histories gradually excludes all of the voxels outside the object and eventually reveals the object and its boundaries. This process is shown schematically in Figure 4.1 for a couple example proton paths; voxels in red have already been excluded, voxels in yellow have just been identified for exclusion, and the remaining voxels are green. It is essential to correctly identify protons that did not pass through the object because misidentifications can cause voxels from the object to be removed, an error that cannot be remedied through sparse matrix reconstruction as previously noted. Protons that do not pass through the object do not experience significant scattering or energy loss so their path is closely approximated by a straight line. Therefore, angular deviation and WEPL measurements can be used, in principle, to identify which protons missed the object entirely.

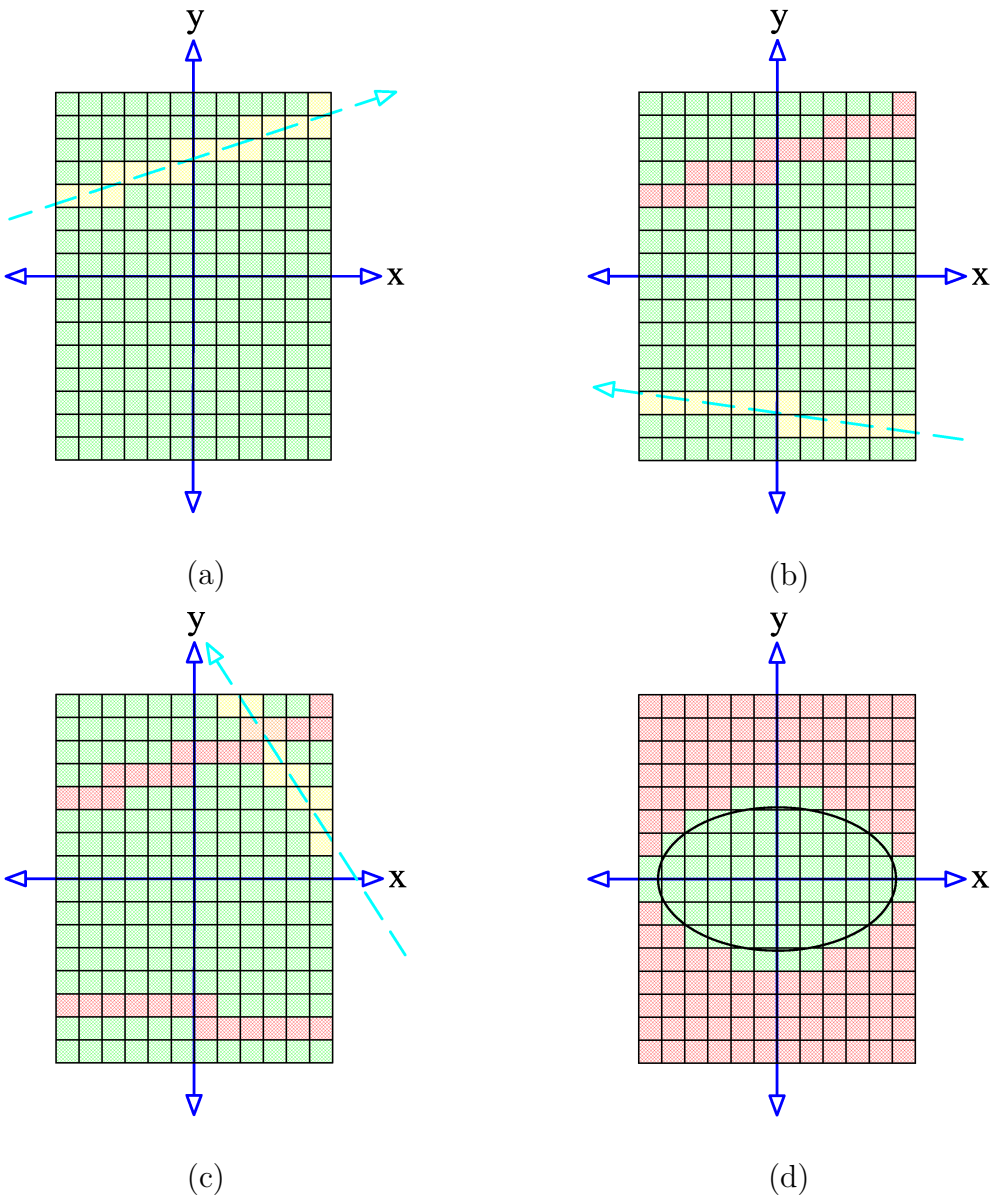


Fig. 4.1: Graphical representation of the process by which voxels are excluded from the object by space carving (SC). The voxels intersected by the straight line approximating the path of the proton (in cyan) are excluded ((a)-(c)), revealing the object after all voxels outside the object have been excluded (d) The voxels remaining after processing all proton histories define the object hull (green voxels). Notice that the voxels that are completely or partially inside the boundary of the object (represented by the solid black line) are not excluded.

4.1.2 Modified Space Carving (MSC)

Modified space carving (MSC) is a variation of SC designed to be more insensitive to the unavoidable errant data present in an experimental data set. Each voxel will be intersected by the path of several protons during a scan, so a voxel lying outside the object will be identified by MSC several times, proceeding exactly the same way as SC. However, SC does not record the number of times, N, a voxel was determined to lie outside the boundary of the object. If a voxel is mistakenly identified as lying outside the object, it will be excluded from the object hull and it cannot be recovered by the reconstruction process. Since WEPL measurements are used to identify voxels outside the object and physical interactions and experimental error produce random measurement error, voxels can be misidentified by the SC process, especially for voxels near the boundary of the object. Protons passing through these voxels do not lose much energy so they are already difficult to discern from protons that miss the object based on their WEPL measurements.

To prevent misidentified voxels from being excluded from the object hull, MSC places a threshold on the number of times a voxel must be identified as lying outside the object before being excluded. However, the number of correctly identified and misidentified voxels varies depending on the total number of proton histories, so the same threshold cannot be applied to every data set, it must be determined for each data set individually. Although it is possible for voxels in the interior of the object to be mistakenly identified as lying outside the object, since these voxels should have a relatively low value of N compared to voxels outside the object, the difference, ΔN, in N between neighboring voxels should be small for these voxels. However, ΔN should

increase sharply at the boundary of the object and continue to increase the further a voxel is from the boundary of the object. This behavior can be exploited to determine an appropriate threshold value, N_T, using a variation of a basic steepest decent edge detection technique. This approach, chosen for its simplicity, proceeds by calculating the largest difference in N between neighboring voxels and defines the threshold using the larger value of N of the two voxels at this edge. This is theoretically an effective approach because the difference in N between neighboring voxels should be largest at the edge of the object, making its boundary relatively obvious.

4.1.3 Space Modeling (SM)

Space modeling (SM) is a technique that only uses protons passing through the object to identify voxels belonging to the object. Unlike SC and MSC, which reveal the object by determining the voxels that do not belong to it, SM essentially solves the inverse of this and identifies the voxels of an object directly. Aside from this, SM is performed using a similar approach as MSC, except now N represents the number of times a voxel is determined to lie inside the object. Unfortunately, a threshold cannot be determined using a steepest decent edge detection approach because the difference ΔN in N between neighboring voxels is no longer largest at the edge of the object. However, N is relatively constant inside the object and begins to increase at the edge of the object and rapidly increases the further a voxel is from the boundary of the object. Therefore, the edge of the object is determined by locating the edges where the difference in N begins to increase monotonically and this . This can be accomplished by identifying the edges where the difference in N between neighboring

voxels is larger than some small threshold value, N_T; a threshold $N_T = 50$ should be acceptable for now because, although ΔN is not largest at the boundary of the object, it begins increasing rapidly starting at the edge of the object.

4.2 Input Data

To assess these new hull detection techniques and compare them to FBP, they were performed on two types of input data: simulated and experimental data. Simulated data provides several distinct advantages over experimental data and offers insight into the theoretical capabilities and limits of each hull detection technique. Simulated data is produced by performing a simulated scan of a digital phantom, so the accuracy of the object hull produced by each hull detection technique can be determined by comparing it to the digital phantom. In particular, the accuracy of an object hull is assessed in terms of how many object voxels were not identified and how many voxels it contains from outside the object. Unfortunately, this comparison cannot be performed on an object hull produced from experimental data since it; since experimental data represents the measurements acquired from a scan of an unknown object, there is nothing to compare each object hull to.

Another advantage of simulating a scan of a digital phantom is the error in the simulated data can be controlled. A simulator has various options and parameters which allow the error in the simulated data to be controlled by turning various physical interactions and noise on or off; the specific options and parameters available depend on the particular simulator. The more sophisticated simulators (like GEANT4) are capable of accurately modeling physical interactions, but because there is no method

for predicting the errors that might occur with an actual scanner, it is important to assess the effectiveness of each hull detection technique using experimental data as well. Although the object hull produced by each technique cannot be compared to an exact phantom, they can be compared qualitatively and the amount of time it took to produce them can still be measured.

The capabilities and deficiencies of each hull detection technique can ultimately be determined by performing a thorough analysis of the results from multiple simulated and experimental data sets. The results from the simulated data, which will not include effects from physical interactions, will be used to determine the theoretical accuracy of each hull detection technique and how sensitive each technique is to noise, thereby providing a proof of concept. The experimental data will be used to identify the challenges introduced by physical interactions and the various inaccuracies produced by the components of an actual scanner, so this analysis will help guide future developments.

4.2.1 Simulated Data and Digital Phantom

Particle physics simulations are typically performed with a simulator like GEANT4, which is capable of modeling nearly every type of physical interaction. However, the simulated data is only being used to assess the theoretical effectiveness and viability of the new hull detection techniques, so the simulated data was produced using a proton CT simulator specifically designed for algorithm analysis [19]. The simulator generates a uniform, parallel proton beam with random lateral and vertical displacement and adds an additional angular displacement to simulate a cone-beam-like scattering

of the beam in two position-sensitive detection planes upstream of the phantom. It also generates bivariate normal random variables for exiting angle and displacement to simulate the effect of proton scattering inside the object. For simplicity, the intersection lengths for the voxels that lie in a proton path are assumed to be 1 mm so a noiseless WEPL measurement can be obtained by summing the RSP of each voxel intersected by the path of the proton, which is assumed to be a straight line between the energy and exit points of the phantom.

The simulator provides the user with the ability to construct nonhomogeneous digital head phantoms of various sizes and allows simplified representations of anatomical features, such as, ventricles, frontal sinus, ears or a nose to be included as well. The digital phantom used in this work is shown in Figure 4.2; the phantom had an isotropic voxel size of 1 mm^3 and was comprised of an outer elliptical region representing skull bone enclosing brain and two inner elliptical sections representing fluid-filled ventricles [18]. These regions were assigned realistic RSP values (1.6 for bone, 1.04 for brain, and 0.9 for ventricles). Two data sets (11,796,480 proton histories each) were produced for this digital head phantom, one without noise and one with noise from energy straggling. The noiseless data set was produced first, as this is the default operation of the simulator. Instead of performing a second simulation with energy straggling, which would produce completely different paths through the object, the data from the first simulation was duplicated and used to calculate a noisy WEPL value. The noisy WEPL values are created by first converting the noiseless WEPL value into an exit energy, generating a normally distributed noisy energy value with a standard deviation described by Tschalar's energy straggling theory [20], and then

Fig. 4.2: Digital head phantom used in this work.

converting that noisy energy value back to a final WEPL value. The conversion of exit energy to WEPL and vice versa is based on the Bethe theory of proton energy loss.

4.2.2 Experimental Data

Two experimental data sets were used, both of which were acquired using the phase I proton scanner at Loma Linda University Medical Center. One data set contains 50,897,953 proton histories from a scan of a anthropomorphic pediatric head phantom, Model 715-HN, CIRS; the second data set contains 65,128,961 proton histories obtained from a scan of a living rat held in restraints and under anesthesia. To maintain equivalence between the simulated and experimental data, only the data from a single translation was used. To eliminate inconsistencies in WEPL measurements, both data sets have been calibrated to account for the response differences between

the CsI crystals.

5. HULL DETECTION IMPLEMENTATION AND RESULTS

5.1 Implementation of Hull Detection Techniques

There are several implementation details which affect the performance of these hull detection techniques and these differ for simulated and experimental data. Each of the techniques presented here require various parameters to be defined and these must be chosen wisely to optimize their performance. All hull detection methods are programmed in C++/CUDA to exploit the inherent parallelism present in the calculations and improve computational performance. These were all executed on the same computer system to maintain consistency for comparisons.

5.1.1 Space Carving

In an effort to avoid the effects of the statistical variations in WEPL measurements associated with experimental data, a proton is assumed to have missed the object only if its WEPL$\leq$ -2.0 mm, thereby decreasing the probability that the uncertainty resulted in an artificially low WEPL measurement. However, the noiseless data produced by the simulator is always greater than or equal to zero, so there would be no voxels excluded by SC if the same threshold was used here. Therefore, a WEPL threshold of WEPL$\leq$1.0 mm was used for both of the simulated data sets. The selection and rejection process was performed for all proton histories in the data set,

and the voxels in the space not carved out were assigned an RSP value of 1.

5.1.2 Modified Space Carving

Like SC, a WEPL threshold of $N_T \leq$ -2.0 mm was used for MSC to avoid voxels from inside the object from being excluded. Since N drops more sharply at the boundary of the object than in any other location, the edge with the largest gradient in each slice was located and the voxel on this edge with the largest N was used to set the threshold, N_T, for that slice. Any voxel in that slice with $N \geq N_T$ was then assumed to be part of the object and the process was repeated for each slice to determine the entire object hull. Again, voxels inside the object are assigned an RSP value of 1.

5.1.3 Space Modeling

The path of a proton that passes through the object also passes through several voxels outside the object, so the number N of times a voxel is determined to belong to the object decreases the further from the object the voxel is. Since N remains relatively constant inside the object, the difference ΔN in N between neighboring voxels inside the object is quite low. This continuously increases as the distance from the object increases voxels with N larger than its neighbors N by ≥ 50 were excluded from the object and this threshold was found to be insensitive to both the number of proton histories and the particular data set. Again, voxels inside the object are assigned an RSP value of 1.

5.1.4 Filtered Backprojection

In this work, we used the Feldkamp Davis Kress (FDK) algorithm [21], a cone-beam variant of the FBP algorithm, assuming that all proton paths through the object are straight and follow a cone-beam pattern. The FDK algorithm was performed with 4° angular bin spacing, a 0.1 cm lateral bin size, and a 0.5 cm vertical bin size. Each slice of the reconstruction volume was defined to be 20 cm × 20 cm and 0.3 cm thick. With this thickness and a reconstruction volume height of 9.6 cm, a total of 32 slices were produced. A Shepp-Logan filter was used prior to back-projection. The resulting image was thresholded to produce an object hull. Any voxel with $RSP \geq 0.6$ was assumed to belong to the object and was assigned an RSP value of 1. Voxels with RSP values below this threshold were assigned an RSP value of 0.

5.2 Hull Detection Results

5.2.1 Results from Simulated Data

The results obtained from the simulated data using each technique are shown in Figure 5.1 (noiseless data) and Figure 5.2 (noisy data) and their corresponding computation times are shown in Table 5.1 and Table 5.2, respectively [18]; the digital head phantom is included in each figure for comparison. All images are 200 voxels × 200 voxels and represent an area of 200 mm × 200 mm. The tables show the number of non-zero voxels in the phantom that are missing from the object hull, the number of voxels in the object hull that are not actually part of the phantom, and the amount of time it took to perform each technique.

The simulated data results for SC are quite accurate and executed in under 100 ms,

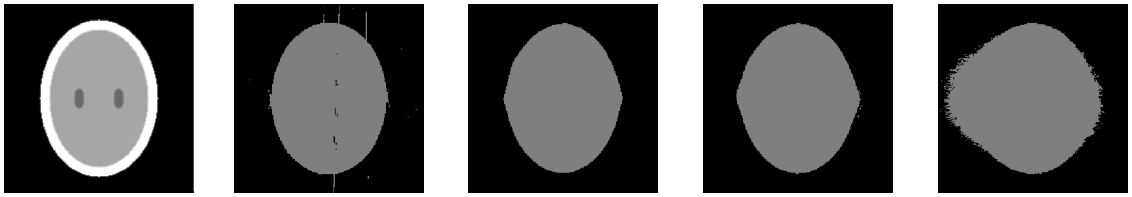
nearly 1000 times faster than any other technique. The MSC results were slightly more accurate and although it did not execute as quickly as SC, it was still about 3 times faster than FBP. These results also show that SC and MSC are relatively insensitive to noise, while the quality of the FBP results declined considerably with the introduction of noise. Notice that although the SC and MSC object hulls look very similar to the digital phantom, FBP was able to identify more of the object voxel. The voxels that SC and MSC failed to identify are from a thin layer around the perimeter of the phantom, whereas the voxels that FBP failed to identify lie primarily in the interior of the phantom and increased with the introduction of noise. On the other hand, SM exhibited significantly different behavior. Although it required slightly less time to produce an object hull than MSC and was able to identify more voxels in the object than the other techniques, the object hull contained far more voxels from outside the object than the other techniques.

Tab. 5.1: Comparison: Noiseless Simulated Data

	FBP	SC	MSC	SM
Computation Time	16.70s	<100ms	5.95s	5.52s
Missing Voxels	79	692	433	57
Extra Voxels	180	66	256	3598

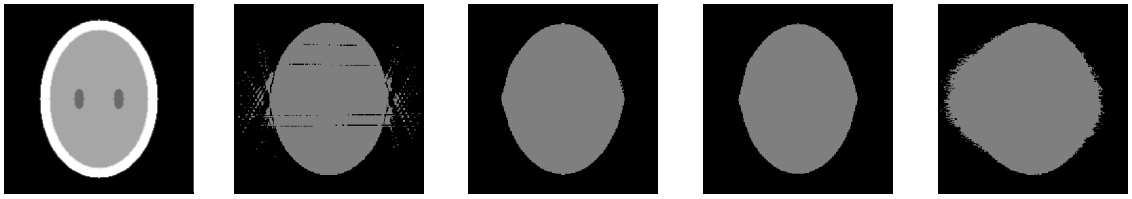
5.2.2 Results from Experimental Data

The object hulls produced by each technique for are shown in Figure 5.3 (pediatric head phantom) and Figure 5.4 (rat) and their corresponding computation times are shown in Table 5.3 [18]. Although there is no object to compare these images to,



(a) Simulated Phantom (b) FBP (c) SC (d) MSC (e) SM

Fig. 5.1: The simulated phantom alongside the object hulls determined by each technique using the noiseless simulated data



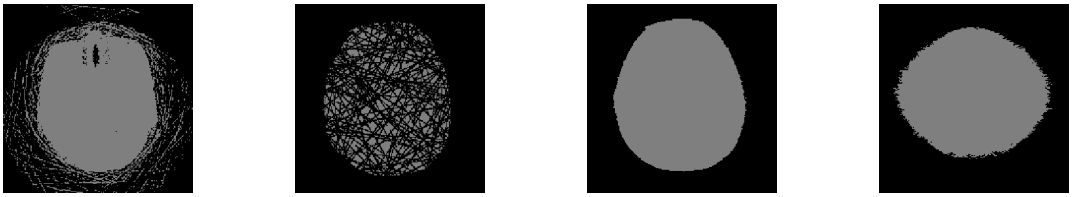
(a) Simulated Phantom (b) FBP (c) SC (d) MSC (e) SM

Fig. 5.2: The simulated phantom alongside the object hulls determined by each technique using the noisy simulated data

Tab. 5.2: Comparison: Noisy Simulated Data

	FBP	SC	MSC	SM
Computation Time	16.72	<100ms	6.14s	5.86s
Missing Voxels	264	373	688	139
Extra Voxels	897	277	66	3166

it is clear that FBP and MSC were far more effective than SC and SM, though FBP again produced object hulls with missing voxels in the interior. Notice that a significant portion of the object was removed by SC with both data sets but the thresholding process performed by MSC prevented any voxels from the object from being removed. SM performed similarly with the experimental data as it did with the simulated data, clearly including more voxels from outside the object than any other



(a) FBP (b) SC (c) MSC (d) SM

Fig. 5.3: The object hulls determined for the pediatric head phantom

(a) FBP (b) SC (c) MSC (d) SM

Fig. 5.4: The object hulls determined for the rat

technique. However, by including these additional voxels, it is likely that SM identified more voxels from inside the object than any other technique again. Although the experimental data sets were substantially larger than the simulated data sets, the computational performance of these techniques were relatively unaffected and their order of from fastest to slowest remained the same: SC, SM, MSC, FBP.

Tab. 5.3: Computation Times

	FBP	SC	MSC	SM
Pediatric Head	18.872 s	<100 ms	7.87	7.46
Rat	19.912 s	<100 ms	8.53	8.14

6. CONCLUSIONS AND FUTURE WORK

As can be seen from the results of this work, no single technique was able to satisfy all criteria. The most important requirement is a viable hull detection technique must be able to identify *all* voxels from inside the object. Although filtered backprojection (FBP) produces a relatively accurate object hull, it consistently produces an object hull with holes in the interior of the object, whereas space carving (SC), modified space carving (MSC), and space modeling (SM) only fail to identify voxels on the perimeter of the object. This is a meaningful distinction because a thin layer can easily be added to an object hull to recover voxels on the perimeter, but the streaks of extra voxels outside an FBP object hull make recovering the voxels in the interior more challenging. Typically these holes can be filled in by analyzing the neighboring voxels, but the streaks outside the object make identifying the interior of the object problematic and prone to error. Streaks are commonly introduced as a result of the back-projection process, as energy loss is assumed to have happened smoothly along each proton path, including voxels outside the object since its boundaries are unknown at this point. Since the amount of energy deposited in voxels outside the object is then overestimated, the amount of energy deposited in voxels in the interior of the object, and therefore their RSP value, is underestimated. These unnecessary voxels in the object vector x are therefore maintained throughout the reconstruction

process, thereby increasing memory usage and computation time as well as slightly reducing the accuracy of the reconstructed image.

Although using finer binning or a different filter could improve FBP results slightly, the streaks and holes present in each object hull is an inherent consequence of the thresholding process and can only be removed completely by increasing the threshold RSP value. Unfortunately, this will result in a drastic increase in the number of extra voxels and a highly inaccurate object boundary, resulting in poor accuracy in MLP calculations. In addition to these accuracy issues, FBP was by far the slowest hull detection technique and it has already been programmed for optimal performance, so any modifications introduced to overcome its deficiencies are most likely going to be too computationally expensive. Therefore, future work should focus on improving the techniques presented here.

SC, MSC, and SM each possess attractive properties as preconditioners, but improvements will need to be made to take full advantage of the sparse matrix format since this will provide iterative reconstruction algorithms with greater efficiency. Future work will focus on obviating each technique's deficiencies so an efficient preconditioning technique can be developed which consistently and quickly produces an object hull free of missing voxels and with as few extra voxels as possible. SC was by far the fastest technique, nearly 1000 times faster than FBP, and although there were no voxels missing from the interior of the objects and only a thin layer of missing voxels from the perimeter of the objects using the simulated data, but protons passing through the object were mistakenly excluded from the object hull using the experimental data, resulting in a several lines of voxels being removed from the ob-

ject hull [18]. Although this happens infrequently, a data set with tens of millions of proton histories will have enough of these events to remove significant portions of the object hull. MSC was designed to reject events like these that cause protons passing through the object to be identified as having missed the object entirely, which is most likely the result of pile-up in the energy detector [18]. On the other hand, MSC produced similar results as SC for the simulated data, as desired, though it was considerably slower and still had a thin layer of missing voxels. However, since the object boundaries are smooth and there are no stray voxels, a thin layer could easily be added to account for the missing voxels, but modifications to the MSC thresholding technique may prevent the outside layer of the phantom from being removed in the first place.

SM was able to identify more object voxels than any other technique but it also included more voxels from outside the object than any other technique, which partially offsets the computational benefits of sparse matrix representations. Like SC and MSC, the voxels that SM failed to identify were located on the perimeter of the object. However, replacing the simple steepest slope edge detection method used here with a more sophisticated edge-detection technique might improve or even solve these boundary issues, though this will be at the cost of computational efficiency. However, since MSC and SM are currently 3 times faster than FBP and this included the current edge detection technique, this cost might not be so prohibitive. Although there are not a lot of changes that can be made to SC, it might be possible to first identify and remove the spurious events that cause voxels in the interior of the object to be removed and perform SC after. Since computational performance is a major

consideration, future work will focus on implementing these changes before resorting to more drastic actions, but because the performance benefits of sparse matrix image reconstruction is far greater than any performance losses incurred in hull detection, ultimately the goal moving forward is to produce the most computationally efficient technique that satisfies all requirements.

REFERENCES

[1] R. F. Hurley, R. W. Schulte, V. A. Bashkirov, A. Ghebremedhin A. J. Wroe, H. F.-W. Sadrozinski, V. Rykalin, G. Coutrakon, P. Koss, and B. Patyal, "Water-equivalent path length calibration of a prototype proton ct scanner," *Med. Phys.*, vol. 39, pp. 2438–2446, 2012.

[2] American Cancer Society, "Cancer facts and figures 2013," Tech. Rep., American Cancer Society Atlanta, 2013.

[3] American Society for Therapeutic Radiology (ASTRO), "Fast facts about radiation oncology," `https://www.astro.org/News-and-Media/Media-Resources/FAQs/Fast-Facts-About-Radiation-Therapy/Index.aspx`, November 2012.

[4] Ute Linz, *Ion Beam Therapy: Fundamentals, Technologies, and Clinical Applications*, vol. 320 of *Biological and Medical Physics, Biomedical Engineering*, Springer, 2012.

[5] B. Schaffner and E. Pedroni, "The precision of proton range calculations in proton radiotherapy treatment planning: experimental verification of the relation between ct-hu and proton stopping power," *Physics in Medicine and Biology*, vol. 43, pp. 15791592, 1998.

[6] S. N. Penfold, A. B. Rosenfeld, R. W. Schulte, and K. E. Schubert, "A more accurate reconstruction system matrix for quantitative proton computed tomography," *Journal of Medical Physics*, vol. 36, no. 10, pp. 4511–4518, October 2009.

[7] R. Schulte, S. Penfold, J. Tafas, and K.E. Schubert, "A maximum likelihood proton path formalism for application in proton computed tomography," *Journal of Medical Physics*, vol. 35, no. 11, pp. 4849–4856, November 2008.

[8] William R. Leo, *Techniques for Nuclear and Particle Physics Experiments*, Springer, 2nd edition, 1994.

[9] G. Coutrakon, J. Hubbard, J. Johanning, G. Maudsley, T. Slaton, and P. Morton, "A performance study of the loma linda proton medical accelerator," *Med. Phys.*, vol. 21, pp. 16911701, 1994.

[10] H. F.-W. Sadrozinski, R. P. Johnson, S. Macafee, A. Plumb, D. Steinberg, A. Zatserklyaniy, V. Bashkirov, F. Hurley, and R. Schulte, "Development of a head scanner for proton ct," *Nucl.Instrum.Meth. A*, vol. 699, pp. 205–210, 2013.

[11] Missaghian J., Hurley F., Bashkirov V., Colby B., Rykalin V., Kachigiun S., Fusi D., Schulte R., Martinez Mckinney F., Sadrozinski H., and Penfold S, "Beam test results of a csi calorimeter matrix element," *JINST*, vol. 5, pp. P06001, 2010.

[12] S.N. Penfold, *Image Reconstruction and Monte Carlo Simulations in the Development of Proton Computed Tomography for Applications in Proton Radiation Therapy*, Ph.D. thesis, University of Wollongong, Australia, 2010.

[13] A.V. Lakshminarayanan, "Reconstruction from divergent ray data," Tech. Rep., Dept.Computer Science, State University of New York at Buffalo, 1975.

[14] G.N. Ramanchandran and A.V. Lakshminarayanan, "Three dimensional reconstruc-tions from radiographs and electron micrographs: Application of convolution instead of fourier transforms," *Proceedings of the National Academy of Sciences*, vol. 68, pp. 2236 – 2240, 1971.

[15] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," in *Proc. Seventh International Conference on Computer Vision (ICCV)*, 1999, pp. 307–314.

[16] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *International Journal of Computer Vision*, vol. 38, no. 3, pp. 199–218, Marr Prize Special Issue 2000.

[17] S. Seitz S. Vedula, S. Baker and T. Kanade, "Shape and motion carving in 6d," in *Proc. Computer Vision and Pattern Recognition Conf. (CVPR)*, 2000.

[18] B. Schultze, M. Witt, K. E. Schubert, and R. Schulte, "Space carving and filtered back projection as preconditioners for proton computed tomography reconstruction," in *Proceedings of the IEEE Nuclear Science Symposium Medical Imaging Conference*, 2012.

[19] M. Witt, R. Schulte, and K.E. Schubert, "A proton simulator for testing implementations of proton ct reconstruction algorithms on gpgpu clusters," in *Proceedings of the IEEE Nuclear Science Symposium Medical Imaging Conference*, 2012.

[20] C. Tschalar, "Straggling distributions of extremely large energy losses," *Nuclear Instruments and Methods*, vol. 61, pp. 141156, 1968.

[21] Feldkamp LA, Davis LC, and Kress JW, "Practical cone-beam algorithms," *J Opt Soc Am A*, vol. 6, pp. 612–619, 1984.