Title

Review of medical radiography and tomography with proton beams

Journal

Reports on Progress in Physics, 81(1)

ISSN

0034-4885

Author

Johnson, Robert P

Publication Date

2018

DOI

10.1088/1361-6633/aa8b1d

Peer reviewed

REVIEW OF MEDICAL RADIOGRAPHY AND TOMOGRAPHY WITH PROTON BEAMS

Robert P. Johnson
Department of Physics
University of California Santa Cruz
1156 High St.
Santa Cruz, CA 95064
rjohnson@ucsc.edu

1 Abstract

The use of hadron beams, especially proton beams, in cancer radiotherapy has expanded rapidly in the past two decades. To fully realize the advantages of hadron therapy over traditional X-ray and gamma-ray therapy requires accurate positioning of the Bragg peak throughout the tumor being treated. Already a half century ago suggestions were made to use protons themselves to develop images of tumors and surrounding tissue, to be used for treatment planning. The recent global expansion of hadron therapy coupled with modern advances in computation and particle detection has led several collaborations around the world to develop prototype detector systems and associated reconstruction codes for proton computed tomography (pCT) as well as more simple proton radiography, with the ultimate intention to use such systems in clinical treatment planning and verification. Recent imaging results of phantoms in hospital proton beams are encouraging, but many technical and programmatic challenges remain to be overcome before pCT scanners will be introduced into clinics. This review introduces hadron therapy and the perceived advantages of pCT and proton radiography for treatment planning, reviews its historical development, and discusses the physics related to proton imaging, the associated experimental and computation issues, the technologies used to attack the problem, the contemporary efforts in detector and computational development, and the status and outlook.

2 Charged-Particle Cancer Therapy and Treatment Planning

More than half a century ago Robert Wilson [1] recognized that charged particles held an intrinsic potential advantage for cancer therapy compared with gamma rays and X-rays, and that accelerators capable of accelerating protons or heavier ions to sufficiently high energy would soon be readily available. Since the specific ionization (energy loss per unit distance) of non-relativistic ions varies nearly inversely with the kinetic energy, the radiological dose is greatest near the end of the ion path, in the "Bragg Peak," named in honor of William Henry Bragg, who discovered the effect in 1903 [2]. Wilson proposed that with sufficient knowledge of the specific ionization, or "proton stopping power," of the tissue between skin and tumor, the ion energy could be tuned such that the ions stop in the tumor, resulting in minimal dose proximal to the tumor and nearly zero dose distal to the tumor. See Fig. 1 for a comparison of charge-particle irradiation with X-ray or gamma irradiation, for which the

depth of the ionizing interactions cannot be controlled. In the figure, the broad flat proton dose distribution in the tumor region is realized by superimposing Bragg peaks from protons of six different energies.

It is clear that in order to take full advantage of this treatment modality, not only does the size of the beam need to be controlled to conform to the morphology of the tumor, but also the incident proton kinetic energy must be controlled differently at each location in order to conform to both the depth of the cancerous tissue and the properties of the tissue in front of the tumor. Assuming a broad, diffuse proton beam, the lateral size can be controlled by collimators that are custom machined per patient, whereas the overall energy can be controlled by adjusting the accelerator (for a synchrotron or synchro-cyclotron) or by passing the beam though "degrader" material (for an isochronous cyclotron). The lateral variation in energy can be achieved by passing the beam through a custom-machined plastic "compensator," which is thicker in regions where less penetration is desired in the patient and thinner where deeper penetration is needed. See [3] for a recent thorough review of the physics behind proton therapy.

See [4] for a review of proton therapy itself, which is outside of the scope of this article. However, it is worth acknowledging here that proton therapy remains controversial, mainly because of its high cost relative to X-ray facilities. A cost-benefit analysis is difficult because despite the fact that from a purely physics standpoint one expects substantial benefits from proton therapy, due to its ability to localize better the radiation, those benefits in general have not been quantified and proven in large random trials. One difficulty in making comparisons with X-ray treatment is that the X-ray field is more mature, and therefore more advanced, than the field of proton treatment. In particular, it is not valid to compare "old fashioned" proton treatment, done as described above by use of a diffuse beam together with static collimators and compensators, to advanced Intensity Modulated Radiation Therapy (IMRT), a sophisticated X-ray treatment modality in which the radiation dose is made to conform more precisely to the three-dimensional shape of the tumor by modulating the intensity of the radiation beam in multiple small volumes. In fact, proton therapy is now moving away from static collimators and compensators toward greater use of Intensity Modulated Proton Therapy (IMPT), a new modality in which a "pencil beam"[1] is scanned across the patient while the beam energy and intensity are modulated according to the treatment program, which could include treatment from multiple directions. IMPT not only allows proton therapy to be applied to tumors that could not be accessed via the older proton-therapy treatment methods but also allows better control of the dose distribution and, therefore, greater success in sparing critical organs from harmful radiation dose. For example, X-ray irradiation of breast cancers is known to increase the risk of subsequent major coronary events [7, 8], but IMPT can deliver the needed irradiation with minimal dose to the heart. Proton therapy has to be evaluated not just in terms of the success of the irradiation in destroying the cancer but most importantly in terms of how well the effects of the radiation on the rest of the body are minimized. That is difficult to assess when the deleterious effects may not become apparent until years later [8], for example when the breast-cancer survivor has a heart attack. Fortunately, randomized trials are now getting underway to make direct

[1]The full width at half maximum of a "pencil beam" in air can be as low as about 7 mm at the highest beam energies, with a corresponding width of about 18 mm at the Bragg peak that is nearly independent of the beam energy [5]. However, sometimes beams as wide as 20 to 35 mm in air are called "pencil beams" [6].

comparisons between IMRT and IMPT. For example, see Ref. [9].

Even with the advent of IMPT the field of proton therapy has almost certainly not yet reached its full potential. One problem that remains to be satisfactorily resolved is the so-called "range problem." Clearly, for proton therapy to be successful, the range of the therapy protons in the tissue of the patient must be accurately known in advance in order to plan the treatment and then machine the compensator or program the IMPT. Up to now the proton "stopping power" (often called dE/dx by particle physicists) of the tissue has been estimated from X-ray CT images by converting the X-ray absorption (linear attenuation coefficient) measured in Hounsfield units (HU) to proton "relative stopping power" (RSP), the stopping power expressed relative to that of water.[2] The transformation is usually calibrated by direct measurement of both the Water-Equivalent Path Length of protons (WEPL, which is a path-integral of the RSP, or the length of a path in water that would produce the observed proton energy loss) and the photon attenuation in various tissue-equivalent materials. A problem with this approach is that X-rays interact very differently in material compared with protons, resulting in relations between HU and RSP that are not unique and can therefore be ambiguous. Artifacts in X-ray images can produce additional errors. Range errors of 5% in the abdomen [10] and up to 11% in the head [11] have been reported. More recent work predicts typical errors of 1.8% and 1.1% for bone and soft tissue respectively [12], although the presence of higher density materials, and the resulting beam-hardening artifacts, can result in larger errors for specific cases, depending on the position of the sample within the body and the size of the body. In general errors in the range prediction increase with the complexity of the geometry through which the protons must pass, but the use of Monte Carlo simulations can be a significant aid in reducing the uncertainties [13]. Reference [14] supports the estimates of [12] with in-vivo measurements of proton range based on MRI imaging of radiation-induced changes in spinal bone marrow. See the review [15] for more information on in-vivo range verification.

Various approaches are being taken, in addition to Monte Carlo simulations, to improve proton range predictions. One is called dual-energy CT (DECT) [16], which combines information from two images made using different X-ray energies in order to resolve the ambiguities in the HU to RSP conversion. Another approach is to measure the RSP directly by using protons. That can perhaps be satisfactorily accomplished through use of a proton radiograph to correct the RSP map obtained from X-ray CT, or by quality control of the compensators derived from X-ray CT measurements [17,18]. Or for the purpose of treatment planning one could replace the X-ray CT entirely in favor of a proton CT scan of the patient.

Whereas solving the range problem is a principal motivation of proton radiology and CT, it is not the only motivation. Early on it was realized [19] that proton radiography potentially provides better density resolution and tissue contrast compared to conventional X-ray imaging,[3] and more recent work appears to bear that out [20]. Proton radiography could also provide quick verification of patient setup in a cancer treatment facility, with very low radiological dose, and an appropriate radiography system might also be suited to real-time monitoring of the treatment beam [21]. Also, proton CT does not suffer from artifacts

[2]Sometimes RSP is referred to as "relative linear stopping power," or RLSP.

[3]X-ray images are commonly known for giving very high contrast to bones and teeth, due to the very strong atomic-number dependence of X-ray absorption, but in many cases good contrast between soft tissues is more important.

that often appear in reconstructions of X-ray CT scans, particularly as a result of metallic inserts (see Section 7 and Fig. 14).

Of course, to do a radiograph or a proton-CT scan of a patient, the protons necessarily have to pass completely through the patient. Therefore, they must be of significantly higher energy than the protons used in the cancer treatment itself, which may seem to detract from the advantage of making a direct measurement of the proton stopping power. In fact, proton stopping power does depend strongly on the proton energy, but fortunately that dependence is nearly material independent, such that the *relative* stopping power (RSP) is close to being energy independent over the relevant energy ranges. For example, from the NIST tables of proton stopping power [22], the RSP of polystyrene (its stopping power in MeV cm^2/g divided by the equivalent quantity for water) varies by only 0.6% from 10 MeV to 200 MeV, whereas the RSP of propane-based tissue-equivalent gas varies by 0.98% over the same energy range. Therefore, measurements of RSP made using fully penetrating protons still represent nearly direct measurements of the RSP needed for treatment planning. However, there does remain a problem that most existing proton therapy centers cannot provide a beam sufficiently high in energy to pass through the thickest parts of the body of many patients.

3 A Short History of Proton Radiography and CT

The idea of using pCT for the purpose of proton-therapy treatment planning was clearly outlined in Cormack's seminal 1963 paper on how to reconstruct images from sets of line integrals, as the last of three radiological applications that he proposed [23]. That decade also saw developing interest in proton radiography, exemplified by Andrew Koehler's 1968 publication in Science [19], which emphasized high contrast while acknowledging poor spatial resolution: "Energetic protons from an accelerator may be used to produce radiographs showing unusually high contrast but relatively poor spatial resolution." Cormack and Koehler also participated in the first laboratory implementation of pCT in the mid 1970s, using a 158 MeV pencil beam to image a phantom with small (0.5%) density variations [24]. Around the same time, Ronald Martin et al. at the Argonne National Laboratory proposed building proton accelerators dedicated to diagnostic work, including pCT [25], and in 1977 Martin, together with Ken Hanson and Bill Steward, a University of Chicago M.D., proposed to the NIH "Development of a prototype proton CAT scan system" [26]. With a 205 MeV beam from a proton synchrotron, the Argonne team of Stephen Kramer, Martin, Steward et al. demonstrated significant dose reduction and improved density resolution relative to conventional X-ray techniques [27]. During the 1970s Koehler and Steward's interest in proton radiography continued, with applications to detection of strokes [28] and tumors [29], diagnosis of breast carcinoma [30], and imaging of a brain tumor [31].

Through the 1980s work proceeded on development of pCT. In 1981 at Los Alamos National Lab Hanson used protons in a scanned pencil beam (1 mm FWHM minimum width and 6 milliradian divergence) to image a variety of phantoms and human organs [32].[4] In his studies of pCT, Hanson emphasized a possible dose advantage of a factor of 1/10 relative to X-ray CT [34] but did not consider possible advantages for treatment planning, probably

[4]LANL was not primarily interested in medical applications—proton radiography was employed as a tool for studying the behavior of materials in implosions driven by high explosives [33].

because charged-particle cancer therapy at that time was in its infancy and therefore rarely used. In 1987, Takada et al. at Tsukuba [35] were able to complete pCT scans in eight minutes using a system with 13 pencil beams and a large magnetic spectrometer to measure the residual energy. In 1999 Zygmanski et al. at the Massachusetts General Hospital (MGH) [36] implemented a very different method to accomplish proton CT, using a cone-beam modulated in energy, such that the integrated rate of protons entering a detector depended on the fraction ranging out in the phantom, giving a intensity-depth correlation similar to what is seen with X-ray absorption.

The latter two systems did not measure individual protons and necessarily assumed straight paths through the phantom. Hanson's apparatus did measure individual protons, using a multi-wire proportional chamber to measure individual proton exit locations, but it still had to assume straight paths through the phantom and was too slow to be clinically useful, requiring about 10 hours to complete a scan. The early proponents of pCT did realize, however, that measurement of individual proton trajectories both before and after the object being imaged could lead to major improvements (see Section 8.4 in Ref. [35], for example) and that future data acquisition systems would be able to accomplish a full scan in that way within a clinically useful time span. In Ref. [34] Hanson forecast a system capable of measuring 10^8 individual proton histories in 10 s, a rate that he wrote was possible with technology of the day, assuming sufficient parallel processing. Thirty years later our technology is certainly up to the task, as some large, expensive data acquisition systems in contemporary major particle-physics experiments acquire data at even higher rates. Nevertheless, measurement of 10^7 proton histories per second is still difficult to accomplish with the modest resources generally available in this field, although as shown below several existing pCT and particle radiography systems do successfully operate within a factor of 10 of that goal.

4 Physics and Technology of Particle Detection for Proton Imaging

Since the turn of the century the challenges and opportunities of proton CT and radiography have led to an increasing number of collaborations of experimental particle physicists with expertise in data acquisition and particle detection technologies working together with medical physicists on the subject, motivated by the rapidly expanding availability of medical centers for proton cancer therapy. The various ongoing efforts are summarized in Section 6, but first this section discusses the major technological requirements for proton imaging, how they can be met by current instrumentation technology, and what fundamental limitations there are to both spatial and WEPL resolution. Although there is ongoing work on development of proton radiography and CT systems that measure integrated dose in order to minimize detector complexity and cost, this review concentrates on systems that attempt to achieve optimal spatial resolution as well as excellent energy-loss resolution. That requires tracking and measurement of individual particles in order to be able to take into account the large multiple scattering that is inevitable in any thick phantom. The single-particle approach may also be the only way to achieve the lowest possible radiological dose while still producing good, useful images.

A contemporary pCT scanner generally takes the form illustrated in Fig. 2 [37]. Thin tracking detectors measure the particle trajectories both entering and exiting the phantom,[5] from which the "most likely path" (MLP) through the phantom can be estimated particle by particle. Note that a single layer in the front tracker would be sufficient if the incoming proton direction were well known. The second tracker is followed by an energy or range detector, usually in which the particles stop so that their residual energy or range can be measured. The beam is either fanned out into a cone to cover the instrument aperture (for example by scattering in a thin foil) or else scanned across the aperture by magnets. The instrument must rotate around the phantom, or more easily the phantom can be rotated within the instrument, by at least 180 degrees either continuously or in discrete steps. For each particle the instrument must provide the geometric path through the phantom together with the integrated energy loss, most conveniently in the form of the WEPL.

The second half of the pCT problem is to reconstruct a 3-dimensional image, or image slices, from the data, a process that is more complex than the equivalent reconstruction problem in X-ray CT, because of the curved paths. Early pCT experiments in which straight-line paths were assumed used efficient filtered-back-projection (FBP) techniques to reconstruct images. When the curved most-likely-path must be taken into consideration the FBP technique is still possible but is significantly more complicated [38]. Alternatively, algebraic reconstruction techniques (ART [39]) have the flexibility to incorporate curved paths, although at much greater computational expense. The image reconstruction problem is further discussed in Section 5.

This review concentrates on radiography and computed tomography with proton beams, with only brief mentions of the closely related field of radiography and tomography with heavy ion beams. Carbon ions, for example, suffer relatively little multiple scattering in the phantom but, on the other hand, do suffer from substantial nuclear spallation. Those distinctions result in substantially different requirements for an effective detector system. Helium ions require much less energy than carbon in order to fully penetrate a phantom. With less spallation they behave more like protons within the phantom, but with significantly reduced Coulomb scattering. Therefore, if helium ions were used in cancer therapy, using the beam also for imaging would be attractive, given the expected improved spatial resolution relative to proton CT. Proton-therapy centers vastly outnumber those using heavy ion therapy, and thus far in the United States there are none of the latter. Nevertheless, interest in heavy ion therapy is expanding, with a similar expansion of interest in heavy-ion imaging. See [40] for a recent review of heavy ion radiography and tomography. Recent reviews of proton radiography and computed tomography also exist. For example, see [41], [42], and [43].

4.1 Particle Energy and Spatial Resolution

In pCT all of the incident protons must pass completely through the phantom, so that their energy loss may be measured. Therefore, the incident proton energy is generally at or near the maximum energy of the medical accelerator, which typically is 230 MeV to 250 MeV kinetic energy. From the NIST range tables [22], the projected range of a 250 MeV proton in

[5]Since it has not yet been possible to deploy any pCT system in a clinic, and all existing systems are experimental prototypes, for simplicity we use throughout this paper the word "phantom" to refer to the object being imaged.

water is about 38 cm, not enough to pass in all directions through the hip region of a typical adult, and far short of the shoulder-to-shoulder distance through a human male [44]. But even 200 MeV is sufficient for scanning a human head, and 250 MeV is sufficient for the lung region of most people as long as the arms are raised out of the way. These facts, together with the typically small aperture of prototype pCT scanners, have led some groups working on pCT to concentrate on instruments that can image at least part of a human head.

Raising the proton energy above what is required to penetrate the phantom can improve the spatial resolution through reduced Coulomb scattering. The rms width of the angular distribution for multiple Coulomb scattering, projected onto a plane, can be approximated by Eqn. 30.15 in [45]:

$$\theta_0 = \frac{13.6\,\mathrm{MeV}}{\beta c p}\sqrt{x/X_0}\,[1 + 0.038\,\ln(x/X_0)] \ . \tag{1}$$

Here βc is the proton's speed and p its momentum, and x/X_0 is the number of radiation lengths of material traversed. For the non-relativistic protons of interest in proton therapy, the denominator is simply twice the kinetic energy, so the amount of scattering is approximately inversely proportional to the incoming proton kinetic energy. However, higher energies can result in more inelastic nuclear interactions, [46] and more importantly, higher energies generally result in reduced WEPL resolution. Therefore, there may exist an optimal energy below the accelerator's maximum, but the optimization depends on judgement of the relative importance of spatial versus WEPL resolution. One should keep in mind that if the image is to be used for treatment planning, then the objective is to map out the proton RSP on a useful spatial scale, not to make beautiful, high-resolution pictures.

For a two-layer tracking device, the error on the projection to the surface of the phantom can be characterized by a transverse displacement error and an angular error. The errors are limited by the point resolution of the detectors and by multiple scattering in the detector material. For simplicity of discussion, we assume that the tracking is two-dimensional, carried out in two independent orthogonal views, which corresponds to the most commonly used technologies, such as scintillating fibers and silicon-strip detectors. The intrinsic angular resolution of the detectors depends their point resolution divided by their spacing. The contribution from multiple scattering in the detectors depends mainly on the thickness of the detector layer closest to the phantom. For example, consider two silicon-strip layers, each 300 μm thick, with 250 μm strip pitch, and separated by 50 mm, with 50 mm distance between the second layer and the phantom. Assuming a simple binary readout of each strip (as opposed to pulse-height digitization that could be used to interpolate between strips), the rms point resolution is the strip pitch divided by $\sqrt{12}$. That results in an intrinsic angular resolution for the set of two detectors of $\sqrt{2} \times (0.25/\sqrt{12})/50 = 2.0$ milliradians. The contribution from multiple scattering of 200 MeV protons in the last silicon layer ($X_0 = 94\,\mathrm{mm}$) is, from Eqn. 1, 1.5 milliradians. For the common situation in which two silicon layers, one for each view, are needed at each measurement plane, the multiple scattering contribution is increased by $\sqrt{2}$ to 2.2 milliradians, nearly equal to the detector resolution contribution, giving a 3.0 milliradian total angular uncertainty. This results in a uncertainty in the 50 mm extrapolation to the phantom of 0.15 mm.

Systematic errors in tracking due to misalignment should also be considered. In the case that both front and rear trackers exist, corrections to the relative alignment of the

four detector layers in each projection can be readily accomplished by use of proton data taken with no phantom present and with the highest available beam energy. From millions of straight tracks, the alignment corrections can be adjusted to a sufficient precision that remaining errors are negligible. Then alignment of the tracking relative to the phantom isocenter is easily accomplished by imaging a narrow rod phantom.

Not surprisingly, the tracking uncertainties are small compared with the spatial uncertainties that result from the proton passage through a typical phantom, which for the purpose of this discussion we assume to be a cylinder of water with a diameter of 200 mm. The front and rear tracks both contribute to the prediction of a proton's path through the phantom. Consider for example trying to predict the location of the proton's path at the phantom center for a 200 MeV proton passing more-or-less along a diameter. From Eqn. 30.19 in [45], the positional uncertainty at the phantom center due to multiple scattering, considering just the front track projection, will be $100\,\text{mm} \times \theta_0/\sqrt{3} = 0.79\,\text{mm}$, but that is with the optimistic and incorrect assumption that the energy of the proton remains constant at 200 MeV, whereas it is expected to decrease to 150 MeV by the time it reaches the center of the phantom. In fact, the rear tracker tends to have significantly less predictive power than the front tracker, because the proton kinetic energy is reduced by the time it exits the phantom. Therefore the point on the path with the largest predictive uncertainty lies beyond the center, as can be seen in Fig. 3, where the maximum error on the prediction of the path of a 200 MeV proton passing through 200 mm of water, calculate from both front and rear track vectors, is between 0.5 mm and 0.6 mm and occurs about 115 mm from the phantom entrance.

Evidently, there is very little to be gained by improving the detector spatial resolution relative to the above silicon-strip example. Furthermore, given that the image spatial resolution near the phantom center will never be better than about 0.5 mm, the ideal voxel size in the image reconstruction will be in the range of 0.5 mm to 1.0 mm.

4.2 Detector Efficiency

Early optimistic predictions that proton CT and radiography would be imaging modalities with very low radiological dose were based on the expectation that detection of the protons would be highly efficient. Several technologies exist that with care can detect protons with nearly 100% efficiency while maintaining low noise levels. In fact, 200 MeV protons yield roughly double the ionization density of minimum ionizing particles typically of interest in particle physics, making the detection problem that much easier. However, near 100% efficiency of each detection layer is also more crucial in a typical pCT system than in a particle physics instrument with many redundant layers. In an instrument such as that described in Fig. 2, losing the signal from a single tracking layer can make the event unusable or at least significantly compromised. Even a seemingly respectable 95% single-plane efficiency results in a loss of 34% of the events when eight measurements are needed, aggravating the already difficult data acquisition problem one faces in trying to minimize the pCT scan time and correspondingly increasing the patient's radiological dose. And if a range detector is used to measure the WEPL, then it is crucial for each of the 60 or more layers to be close to 100% efficient in order to identify reliably the proton's stopping point, although achieving that result is aided by the increased ionization inherent in the Bragg peak. High efficiency

should by a high priority design goal.

Silicon strip detectors and some detectors that employ gas multiplication of ionization can readily yield particle detection efficiencies well above 99% with insignificant noise levels, at least until limited at high rates by signal pileup. Solid-state pixel detectors can do as well or better, at least in the case of those based on deeply depleted silicon diodes, and are far less susceptible to pileup. Scintillating fibers, while popular, have typically resulted in tracking systems with comparatively marginal signal-to-noise performance, even when employing relatively large fibers (e.g. 1 mm diameter) and some redundancy to cover gaps between fibers. As shown in more detail in Section 6, instruments based on scintillating fiber tracking for which efficiency figures have been published yielded low detection efficiencies compared with their silicon-strip based counterparts. However, most have been intended only for proton radiography, for which high efficiency is of reduced importance compared to pCT.

Other factors beyond detector-layer efficiency result in significant losses of events. Nuclear interactions are common, occurring in 20% or more of proton events, depending on how much material is traversed. Those events must be eliminated by data filters to the extent possible, as they contribute confusion or noise to the final images. Similarly, hard Coulomb scattering events should be filtered out, as they do not fit into the MLP framework used to analyze the data. Typically the filters rely on binning proton events that follow similar trajectories to identify and cut out tails in the WEPL and angle distributions. Overall at most half of the events from protons that pass through the regions of interest end up being useful for image reconstruction, and that fraction can easily be far smaller if care is not taken with the detector efficiency.

The fraction used of all protons that trigger the detector system is typically much smaller due to the simple fact that many miss the phantom entirely. Since human patients have not yet been involved in any of the tests of the systems discussed here, collaborations have not yet worried very much about this. A clinical system, however, will have to pay close attention to how well the beam conforms to the region of the patient being imaged, in order not to deliver unnecessary dose. Use of collimators in the beam nozzle will probably result in too many scattered, off-momentum protons that would confuse the image, so appropriate programming of a scanned beam will likely be the best choice.

4.3 WEPL Resolution

Although spatial resolution may have the greatest impact on how the image appears to the eye, WEPL resolution and accuracy in a pCT image are more important for the purpose of treatment planning. The WEPL resolution depends not only on the design and performance of the detector system responsible for measuring residual energy or range but also on the natural fluctuations in energy loss ("range straggling") in any degrader placed upstream of the detector, in the tracking detector, in the phantom and, for the case of a range detector, in the WEPL detector itself. Quality assurance assessment of treatment delivery is expressed in terms of of a percent dose difference (ΔD)) and a "distance to agreement" (DTA) in mm, and recent work advocates for a goal of $\Delta D/\mathrm{DTA} = 1\%/1\,\mathrm{mm}$ [47]. Therefore, to achieve the goal of improved treatment planning, we would like an RSP resolution and accuracy of one percent, to yield a range prediction at a typical depth of 100 mm with an error no worse

than 1 mm [20].

A calorimeter can measure directly the proton residual energy, but because its resolution is generally proportional to the energy deposition, it tends to achieve poor resolution on the WEPL when it is short, since in that case a large measured energy is subtracted from the known beam energy to yield a small result for the energy lost in the phantom. The most straightforward WEPL detector is a range detector, which detects where the proton stops, typically by means of many thin sensor layers interleaved with absorber material. Assuming that the sensors can be made sufficiently thin and numerous, then the WEPL resolution is independent of WEPL and is dominated by range straggling, which for a 200 MeV proton in water is about 3 mm, or 1.1% of its range.[6] Thus a small sample of measured protons along each path is sufficient in principle to achieve the needed resolution. Other detector variations exist. For example, the authors of Ref. [49] have built a hybrid of the calorimeter and range-detector concepts that employs five scintillation stages along the beam direction. The signal from each stage has to be digitized as in a calorimeter, but the advantage over a calorimeter is that protons with small WEPL in the phantom pass through several stages before stopping. In principle, the digitized signal is only of interest from the stage in which the proton stopped, as the previous stages contribute known amounts to the proton WEPL. Therefore, such a system does not suffer from poor resolution at short WEPL. For comparison, resolution models for the calorimetric, range counter, and hybrid concepts are illustrated in Fig. 4.

The hybrid multi-stage scintillator concept has an advantage of far fewer channels than the range counter, but each channel must be more complex, as precise pulse-height digitization is necessary. The major advantage of a simple range counter, however, is simplicity of calibration. In fact, each channel of a range counter needs only a single threshold, the setting of which can be extremely stable and non-critical if a high signal-to-noise detector technology is employed. In that case, a single calibration run could be not only very simple to execute but also sufficient for a long period of operation and possibly even the life of the instrument. By contrast, calibration of a high precision calorimeter or multi-stage hybrid is complex and demanding, requiring measurements of and corrections for many effects such as spatial variations in light collection, nonlinearity, not only of electronics but also of detectors (e.g. Birk's law [50]), time-varying gains and pedestals, aging of components such as scintillators, variations from channel to channel, and in the case of the multi-stage concept, threshold and noise-dependent complexities in dealing with protons that stop near a boundary of two stages. In general calibration errors will introduce visible artifacts into an image and reduce the accuracy of proton stopping-power information derived for use in treatment planning.

Other, very different types of detector could be used for the measurement of residual energy. Examples that come to mind are magnetic spectrometers and time-of-flight counters, although neither has been employed in pCT prototype systems designed to track and measure individual protons. These two examples could give an advantage over range counters by avoiding range fluctuations and nuclear interactions in the counters themselves. However, magnetic spectrometers are too expensive, heavy, and bulky (e.g. see Ref. [35]), and time-of-flight counters would also be very bulky unless the timing resolution could be pushed to the level of around 10 ps, nearly an order of magnitude beyond the capabilities of present

[6]The range straggling prediction follows from Monte Carlo simulations done using the GEANT4 toolkit [48], whereas the range follows from the NIST tables [22] as well as the simulation.

technology for any system with such a large aperture as needed for pCT.

4.4 Detection Rate

The time required for acquisition of a pCT image must be kept reasonably short for the obvious reasons of patient comfort and optimal use of the very-expensive proton treatment facilities. Scan times as short as six minutes have been demonstrated with prototype instrumenation that can measure a million protons per second and write 100 million bytes of data per second, to record between 300 and 400 million proton histories [51]. That is within the bounds of possible clinical use, but one would like to shorten the time by an order of magnitude. To do so while still measuring individual protons will require not only faster data acquisition but also detectors that either can make measurements spaced much more closely in time (which would not necessarily help much at a synchrotron if the protons remain tightly bunched) or else measure multiple protons simultaneously. Continuing with Ref. [51] as an example, the rate performance of that instrument's multistage scintillator detector is limited by the pulse shaping needed for successful digitization, and there is no lateral segmentation in the scintillators, ruling out simultaneous measurement of two or more protons. The instrument's silicon-strip detectors, which are used for tracking with fine (0.228 mm) spatial segmentation, can readily detect multiple protons simultaneously, but the use of two independent and orthogonal views results in unresolvable ambiguities when associating strips in one view with those in the other. Furthermore, silicon-strip detectors require pulse shaping (integration) of the strip signal to reduce effects of amplifier thermal noise, which limits the rate at which protons can be detected by a given strip without unacceptable pileup and inefficiency.

The problems posed by increased rate can probably best be solved by detectors able to measure multiple protons simultaneously, especially in applications at a synchrotron with tightly bunched beams. In principle that can be done by introducing lateral segmentation into existing detector technologies. However, in addition to increased readout channel count, segmentation tends to be associated with dead gaps, which can easily result in problematic artifacts in the image, something that significantly affected image reconstruction for the instrument of Ref. [52], due not only to segmentation of the calorimeter into 16 CsI crystals but also due to segmentation of the silicon-strip detector into two sensors per plane, as needed due to the limited size of available silicon-strip sensors. In that case the sensor edges were overlapped to avoid dead gaps, but that also tends to produce artifacts due to increased detector WEPL in the overlap regions.

As noted above, segmentation of tracking detectors in only one dimension, as done in the commonly used silicon-strip and scintillating fiber technologies, results in pattern-recognition ambiguities. Pixelated detectors would be preferable from that point of view but tend to be either too slow (e.g. CCDs and CMOS active pixel sensors [53, 54]) or else too complex and expensive, for example silicon-diode pixel detectors, which in existing systems also require readout electronics that add material within the sensitive field. See for example the TimePix chip [55], which is popular for photon detection but can also detect charged particles. The authors of [56] aim to achieve the ultimate tracking and rate performance for pCT by developing pixelated CMOS detectors with integrated high-rate readout. That may become the preferred technology of the future, but for now pattern recognition in multi-track events

can be done successfully with one-dimensional segmentation by adding additional "stereo" layers to give a third view, of course at the cost of more expense and more material. That approach is common in particle physics and is being taken by the Pravda pCT system [21] discussed below in Section 6.

Solid state detectors, whether used as strips or pixels, will inevitably suffer problems with gaps introduced when tiling the large aperture needed for a practical device. Scintillating fibers have an advantage in avoiding this problem, as long as the gaps between fibers get covered by tiling with a second layer. Gap effects with silicon detectors can be ameliorated by minimizing the dead area along the edges [57] and by staggering the gaps from one layer to the next. In fact, the near 100% efficiency of the active areas of silicon-strip sensors can in some cases allow the gap to be used as a track coordinate when a track missing a hit on that layer projects to the vicinity of the gap [51].

Inefficiency from rate limitations of the tracking detectors can get severely aggravated when a narrow "pencil" beam is used during imaging. In fact, segmentation and stereo layers both become less effective as the beam width is decreased. The beam scan rate is very slow compared with the particle rate, so the detector system responds as though a static narrow beam were impinging always on the same spot. Silicon detectors with long strips have a distinct disadvantage in this regard, because of the relatively long signal shaping time needed to achieve low noise, and their efficiency will plummet if the same channel gets hit repeatedly at too high a rate. Again pixel detectors will be preferred if and when they become more practical, and scintillating fibers with appropriate signal processing can also regain some advantage. For now, silicon based tracking detectors that use strips have to operate at megahertz particle rates with relatively large beam spots or else employ relatively short strips and power-hungry electronics with pulse shaping that is as fast as possible.

Another common rate-related problem is that the low intensities required by proton CT and radiography in order to accommodate instrument rate limitations lie far below what existing treatment facilities were designed to deliver. As a result, the accelerator operators typically have no instrumentation built into their system that can detect the presence or indicate the quality of the beam when operating at such low current. In an integrated system in the future this could be resolved by using the pCT instrumentation to provide direct feedback to the operators. In the meantime verbal communication between the experimenters and beam operators usually has to suffice.

4.5 Instrument Simulation

Detailed Monte Carlo simulation of the detector system has long been an essential component of all particle physics experiments, and that practice has carried over into the field of proton CT and radiography. Standard packages from particle physics are used, such as GEANT4 [48] and FLUKA [58], but code developed specifically to support hadron therapy, such as TOPAS [59], is playing an increasingly important role. Simulations are important for instrument design, understanding detector performance, development of instrument software, and development and verification of data processing algorithms such as CT image reconstruction. It is important to simulate not just the instrument but also the beam that is delivered by the accelerator facility. A good example of such a simulation model has been recently published in [60].

4.6 Requirements Summary

In summary, pCT requires a very low intensity proton beam of about 200 MeV energy for the head and 250 MeV or more for larger parts of the body. The RSP resolution and accuracy of the image should be better than 1%, with spatial resolution of about 0.5 mm, the limit allowed by multiple scattering in the phantom.

The RSP requirement can be realized by a relatively simple range counter, which can measure WEPL with about 3 mm accuracy, limited by range straggling, and which can operate with stable calibration. Achieving the optimal spatial resolution requires a tracking system that can measure each proton trajectory before and after the phantom, but the required tracking resolution is easily met by various modern technologies, since the spatial resolution is limited by scattering in the phantom. High tracking efficiency is more important for ensuring good measurement of each proton.

The most difficult challenge is to reduce the pCT scan time to well below the 5 to 10 minutes that has already been achieved. To do so will require development of much faster detectors or else WEPL detectors capable of measuring several protons simultaneously, as well as faster data acquisition systems.

The requirements for proton radiography are easier to satisfy. The beam energy needs to be only high enough for the protons to pass through the phantom in one direction, whereas pCT requires full penetration all around in a 360 degree rotation. And present instruments are already sufficiently fast, being capable of producing a radiograph in less than 10 seconds. The RSP requirements are the same as for pCT, however, and if the best spatial resolution is desired, then it still is advantageous to track the protons before and after the phantom, in order to make use of the MLP.

5 Image Reconstruction for Proton CT

Computed tomography image reconstruction from X-ray projection data is a highly developed, mature field [61]. In general the X-ray data, whether in cone-beam CT, helical CT, or multi-detector CT, are amenable to fast, efficient reconstruction by way of filtered back projections, but more general, more computationally intensive methods such as ART have also been well developed and are recognized to hold some advantages such as yielding lower noise images with fewer artifacts, especially at low radiation dose levels and in regions with sharp changes in density, such as might be found around metallic inserts [62].

The filtered back projection is mathematically related to a close relative of the Fourier transform, the Radon transform, which relates a function f defined on the plane to a function $R(f)$ defined on the two-dimensional space of straight lines in the plane, where the value associated with a particular line is equal to the line integral of the function f along that line. Of course, a CT instrument does not measure the uncountably infinite space of lines in a plane but provides only a finite sampling of that space, together with appreciable noise and limited resolution.

A back projection in itself is a very simple concept: just take all the radiographs of the object, each acquired at a different orientation, and project them all back through the object. If the object were simply a small, dense sphere, for example, then each radiograph would have a sharp peak at the location of the sphere, and each back-projection would draw a line

back through the image space. The lines would intersect at the sphere's location, and their sum would greatly darken the region occupied by the sphere. But there would also be dim rays passing outward from the sphere's image in all directions, making a blur that falls off in intensity radially from the sphere's location. Thus a simple back-projection yields a very blurry image.

A two-dimensional ramp filter can eliminate most of the blur. To greatly increase the computational efficiency in the filter's implementation, in practice one takes advantage of the Fourier slice theorem, which states that the 1-D Fourier transform of the projection of an object is the same as the values of the 2-D Fourier transform of the object along a line drawn through the center of the 2-D Fourier transform frequency plane. Thus one can apply a 1-D ramp filter (high-pass filter) to the discrete Fourier transform of each projection, arrange the results properly in a 2-D Fourier transform plane (which requires some interpolation between the discrete radial projections), and then calculate the inverse discrete Fourier transform to obtain the filtered image, in the end essentially a discrete inverse Radon transform of the projection data [63]. Unfortunately, the high-pass ramp filter amplifies high-frequency noise that is always present at some level in experimental data, so in practice it is necessary to combine the ramp filter with a suitable low-pass filter. Further technical details are beyond the scope of this paper but are thoroughly discussed in textbooks as well as in the literature. What is most important to note here is that all of the discussion has assumed projections made along straight lines. Therefore, the FBP technique is not easily applicable to the curved MLPs inherent in pCT.

ART is conceptually simple, as it treats the image reconstruction as a solution to a large linear problem. The technique is very general and is readily applicable to pCT, as well as to X-ray CT. It is convenient to jump immediately to discussion of ART as an application for reconstruction of pCT data from a scanner in which the MLP and WEPL have been measured individually for each proton. Divide the space occupied by the phantom into n voxels, each of a size commensurate with the expected spatial resolution, typically around 1 mm cubed. Let $\vec{X}$ be the solution vector, giving the RSP of the material in each of the voxels, and let $\vec{Y}$ be the measurement vector, giving the WEPL of each of the m measured proton histories. From the proton MLPs we can construct a $m \times n$ matrix $A_{i,j}$ for which each element is the projection of the ith proton's MLP onto the jth voxel. The linear equation

$$y_i = \sum_{j=1\ldots n} A_{i,j} x_j \qquad i = 1 \ldots m \tag{2}$$

multiplies for each MLP the RSP in each voxel by the length of the proton's path through the voxel and sums over all of the voxels along the MLP. Clearly each such sum should give the WEPL for that proton history. The matrix A is enormous, because n will typically be the order of one million, while m, the number of proton histories measured, will be the order of at least a few hundred million. However, it is also sparse, since a given proton passes through only a small subset of the voxels.

The objective of ART is to invert Equation 2, a task that can only be accomplished approximately by an iterative procedure. The linear system is not only far too large for a straightforward solution, but it is also grossly overdetermined, since $m >> n$, and is noisy, due not only to random errors in the WEPL measurements but also due to the fact that in pCT the MLP is only an estimate of the true path. Therefore, one cannot search for

the solution, or even the best solution, but only for a good solution. Once a good solution is found, a heuristic methodology called "superiorization" is often used to look for another solution that better satisfies some criterion, such as minimizing "total variation" [64, 65].

The general method of solution by iterative projections was invented by Kaczmarz in 1937 [66], and the methodology used in ART, often called POCS for "Projections Onto Convex Sets," was first published by Bregman in 1965 [67] and then in 1970 employed in the field of image reconstruction by Gordon, Bender, and Herman [39]. The basic idea is to successively project the solution vector orthogonally onto the hyperplanes defined by each of the proton MLPs according to a recursion formula for the solution at iteration $k+1$ in terms of the solution at iteration k:

$$x^{k+1} = x^k + \lambda_k \frac{y_i - \langle a_i, x^k \rangle}{||a_i||^2} a_i \qquad i \equiv \mathrm{mod}(k, m) \tag{3}$$

where λ_k is a relaxation parameter and a_i is the ith row of the MLP matrix A. The λ_k is optimized empirically, typically by use of simulated data. While it can in principle vary with iteration k, for example a sequence converging to zero, most often it is taken for simplicity to be a constant, for example $\lambda = 0.002$ for the basic ART algorithm in Ref. [68]. A trivial example is a solution for finding the intersection of lines in a plane, as illustrated in Fig. 5, but the application to pCT differs in that the solution vector is of dimension $n \approx 10^6$ instead of 2, the lines are hyperplanes, the number of hyperplanes, m, is much greater than n, and there is no mathematical solution, only an approximate region in which the hyperplanes come close (hopefully!) to intersecting at a common solution.

Application of ART to pCT requires knowledge of the shape and extent of the phantom's surface, or its "hull," as well as a starting guess for the solution. Application of an FBP to the data, assuming straight-line paths for the protons, can provide both. Or a space-carving algorithm can be used to produce the hull, with the starting guess taken to be unity (RSP of water) for all interior voxels [69, 70]. To get an idea of the computational problem, consider that 6 hours were required in Ref. [68] to reconstruct an image from 130 million proton histories using a single CPU coupled to one GPU. With more computing power and optimization of code, however, image reconstruction from several hundred million proton histories has been accomplished in a clinically realistic time. Reference [71] presents several interesting examples, using recent data from the LLU/UCSC Phase-II Scanner [51]. The image reconstructions were done by a rack of 60 processors, each with 12 cores, 72 Gbyte RAM, and two GPUs [72]. With this system, an image can be reconstructed from 500 million proton histories in less than a minute [71]. Using different code B. Schultze demonstrated that a similar image could be reconstructed in 6.5 minutes on a single Xeon compute node using one NVIDIA K40 GPU [73]. Table 2 in Section 7 shows reconstruction results from Ref. [71] for two algorithms, from among many variants that exist, that differ in how the problem is broken up into parallel computations: DROP (Diagonally-Relaxed Orthogonal Projections) [74] and CARP (Component Averaged Row Projections) [75]. For the examples given in this reference, CARP achieved better accuracy than DROP, for which the smoothing due to superiorization tended to compress the RSP scale, especially when many "blocks" were used,

Inherent in the application of ART to pCT is prediction of the MLP, given the incoming and outgoing track vectors. The reconstruction of Ref. [71] employs a maximum-likelihood

formalism [76] derived from the initial work in Ref. [77] and the following work in Ref. [78]. A significant fraction of the computing time in the image reconstruction is devoted to calculating all of the necessary MLPs. More recent work in Ref. [79] presents an approach based on cubic spline approximations that could dramatically speed up the pCT reconstruction. Up to now pCT image reconstructions have assumed MLPs derived for homogenous media, but that could be generalized in an iterative reconstruction to take into account the heterogeneity of the phantom already known from previous iterations, something that the cubic-spline approach could make practical.

The community has not given up on the possibility of accomplishing the pCT reconstruction task with considerably less computing power that what is required by ART. Reference [38] presents a form of filtered back projection along most likely paths. Work is in progress to test that technique with existing data and simulations to compare the reconstruction quality, as well as speed, with ART.

6 Contemporary Instrumentation Efforts for Proton Radiography and CT

Several efforts are in progress around the world to build prototype pCT systems of the type described in Fig. 2, as listed in Table 1. All consist of a combination of a tracking detector and an energy or range detector, but not all use the same detector technologies. Some collaborations have built earlier, simpler prototypes that are not listed in the table, which is meant to reflect just the current state of the art. There are also recent and contemporary efforts to build similar systems that are restricted to proton radiography, as also listed in Table 1. Even when computed tomography is not needed, information on particle tracking before and after the phantom can be used to improve greatly the spatial resolution of a radiograph. Therefore, an optimal proton radiography detector system may closely resemble a system designed for pCT but will have the advantage of producing an image in seconds instead of minutes. None of the efforts has reached the stage at which the instrument could be considered to be ready for clinical or pre-clinical trials.

Integrating approaches, which do not require single-particle tracking, also continue to be studied but are only briefly mentioned here. Systems based on pencil beams were proposed and studied for pCT recently [80] as well as for carbon-ion imaging [81], in which case multiple scattering in the phantom is not nearly as important. Even emulsion-based detection has seen recent development [82]. A early approach at MGH based on energy modulation is already mentioned above [36], but the published images were of poor spatial resolution. Recently another clever approach has been tested at MGH [83] that is based on time-resolved dose measurements. A range modulator wheel placed upstream of the phantom produces a proton energy that varies with a 100 ms period, thus creating a periodic dose rate in a 2-D array of diode detectors placed behind the phantom. The resulting "dose rate function" depends on the WEPL and can be calibrated using suitable calibration phantoms. The spatial resolution of the published images were limited more by the course detector segmentation than multiple scattering, making it difficult to judge the power of the technique. Also, the reconstructed RSP values were quite far from the true values in most cases, so additional work is needed in order to prove this technique.

Some medical physicists believe that simple proton radiographs will be sufficient, when used together with X-ray CT based treatment planning, to solve the range problem. In that case, proton radiography may be used more as a quality assurance tool rather than being employed to develop the RSP map needed in the treatment planning. Or the WEPL map from a single proton radiograph may be compared with a corresponding X-ray image and the differences used to optimize the RSP map derived from X-ray CT [84]. Of course, any system that can accomplish pCT is also well suited for radiography, but radiography does not require such a high data acquisition speed or computing power as needed by pCT in order to complete an image in a clinically relevant time period.

The radiography instrument developed at PSI in the late 1990s at the Paul Scherrer Institute [85] is the most notable such instrument of that era. It can be viewed as a prototype for the concept being followed by contemporary groups to make pCT scanners that detect protons one by one (sometimes referred to as "list mode" as opposed to a signal integrating device). Although it apparently was never used for CT, the PSI instrument nevertheless follows almost completely the pCT scheme illustrated in Fig. 2 and was demonstrated to measure a million protons per second, a metric that most contemporary systems have as a goal but few have yet accomplished. Its main limitation relative to Fig. 2 is that each tracker is composed of only a single detector layer in each view, such that each provided just a point, not a vector, thus limiting the ability to define a most likely path of a proton through the phantom. This is a significant point even for radiography, because recent research has shown that the MLP formalism that is supported by tracking of individual protons before and after the phantom can significantly improve radiography relative to simpler approaches [86].

Each view of each tracking layer of the PSI instrument is provided by a scintillating fiber hodoscope composed of two overlapping layers of 2 mm square fibers, resulting in an effective detector pitch of 1 mm, and the fibers are read out by 16-channel photomultiplier tubes (PMT), each composed of a 4×4 array of independent photocathode and dynode chains. Each PMT views a bundle of 2 (in x) or 7 (in y) fibers, but the bundling is done in such a way that the combination of signals from the two overlapping layers uniquely identifies the x or the y coordinate. This scheme greatly reduces the number of readout channels and the data volume, facilitating the high speed readout, but it has the disadvantage that both layers must give a signal in order to yield an unambiguous coordinate location. The inactive material between scintillating fibers results in an overall efficiency of only 33% to get x, y coordinates from both tracking layers. Such a low efficiency could be a serious problem for a pCT instrument but is acceptable for simple radiography.

The range telescope of the PSI instrument is a stack of 64 three-millimeter thick plastic scintillator tiles read out by wavelength shifting fibers connected to PMTs. The proton range is measured according to the last pair of tiles that produced a signal. Figure 6 shows a typical range spectrum from a collection of protons all of the same energy. The location of the peak can be clearly identified despite there being some protons of shorter range, caused by nuclear collisions in the tracking layers, the phantom, or the range counter. The error on the mean range is dominated by unavoidable range straggling and is about $2.6\,\mathrm{mm}/\sqrt{N}$, where N is the number of protons in the peak for the given pixel.

The PSI data acquisition system is based upon 16-channel VME discriminator modules followed by VME-based custom ECL (emitter-coupled-logic) logic modules that both define a trigger and also reduce the data volume down to only 32 bits per event. Then, with a

megahertz proton rate a radiograph can be acquired in about 20 s, including about 10 s of dead time. This device holds the distinction of being the only system of this kind to date that has acquired radiographs of live animals, a dog patient being treated for a nasal tumor [87]. Unfortunately, there have been no publications of new results from this instrument since then.

A few other collaborations have followed the PSI lead and developed similar scintillating-fiber trackers. The "OFFSET" collaboration has built and tested one intended for medical applications with very small 0.5 mm fibers of square cross section [88]. A multianode photomultiplier tube is used for the light detection, but in a scheme reminiscent of that employed by PSI to reduce significantly the number of channels needed. Both ends of the fibers are read out, but with fibers bundled in groups of 10 in such a way that a coincidence between the two ends uniquely identifies which fiber was hit. From tests with 62 MeV photons, the group reports an efficiency for a single plane of 62%. Such a low efficiency would be ill suited for a device such as is described in Fig. 2, in which at least 6, preferably 8, independent detector planes must register a hit in order to measure adequately a proton history, but it could be sufficient for other purposes. In fact, more recently the same authors have presented a beam monitoring and radiography system called "QBeRT" based on similar technology, but read out by Multi-Pixel Photon Counters (MPPC, often referred to as silicon photomultipliers, or SiPM). It includes a two-layer tracker and a range detector composed of scintillating fibers [89, 90]. It is intended to handle fluences of 10^9 particles per second when operated in beam monitoring and verification mode and 10^6 particles per second in radiography imaging mode. The existing system is small and of limited range depth but could be scaled up. Test results have shown good results in measuring beam profiles and beam range versus energy (see Fig. 7), and some radiographs of simple high-contrast phantoms have been published.

Another proton radiography instrument development effort that is more recent than the PSI effort is the Advanced Quality Assurance project (AQUA) at the TERA foundation [91, 92]. The tracking is based on 30 cm × 30 cm GEM detectors [93], a pixelated detector that operates with gain from electron multiplication by a gas, which are placed only immediately in front of the range detector. Addition of another set in front of the phantom is being investigated, but the team prefers use of the beam scanning system to define the incoming proton trajectory in order to enhance the portability of the system within a clinic. The range detector is a stack of 48 three-mm thick (not including wrapping) scintillator planes, each coupled by a wave-shifting fiber to an SiPM. The signals are individually digitized by 12-bit pipelined ADCs, and the data acquisition system can handle a throughput of about a million events per second. Up to this time, results have only been published from the 10-cm square prototype system [92], which had a similar construction but a different electronic readout system. The large 30-cm square system was ready for testing as of March 2015 [94], but no published results have yet appeared.

With the smaller system the proton range was measured event by event in two ways: by using a simple threshold to locate the last scintillator hit and by detailed fitting of the ADC signals to a function shaped similar to the expected Bragg peak. Both methods gave the same range precision, limited by range straggling as well as measurement errors, suggesting that the readout could be simplified by eliminating the ADCs. From analysis of data taken with a 2 cm thick acrylic phantom with cylindrical holes of different depths, the range resolution was about $1.7\,\mathrm{mm}/\sqrt{N}$ for N protons passing through a given pixel. Figure 8

shows a radiograph of that phantom. As expected, the spatial resolution was limited by multiple scattering together with lack of knowledge of the incoming proton trajectory in the broad, disperse beam used in the test.

Scintillating fiber tracking was employed by a collaboration between Northern Illinois University (NIU) and the Fermi National Accelerator Laboratory (FNAL) to build a complete pCT system with the largest aperture of any pCT system completed or planned to date [95]. In this case both front and rear trackers were designed to measure vectors as in Fig. 2, not just points. They are based on 0.5 mm diameter scintillating fibers read out by SiPMs. To achieve sufficient efficiency despite the inactive cladding on each fiber, the round tracking fibers are arranged in close-packed double layers and are routed to SiPMs in triplets of touching fibers, yielding an effective detector pitch of 0.94 mm. In the tracking layers the fibers themselves detect the protons, whereas in the range detector wavelength-shifting fibers collect light from 3.2 mm thick plates of plastic scintillator and route the light to the SiPMs. An advantage of the scintillating fiber tracking is that large area tracking planes can be built up without any gaps between sensor elements. A potential disadvantage, especially when reading the fibers by room-temperature photodetectors, is that even with a double layer the signal-to-noise is poor compared with what is easily achieved using silicon-strip or GEM detectors. As discussed in more detail in Section 4.2, good signal-to-noise is critical in this kind of system, in which there is little redundancy in the tracking, in order to achieve the necessary near-100% proton detection efficiency while keeping random noise at a low level. Another potential disadvantage is limited spatial resolution, since making the fibers smaller in cross section rapidly reduces the signal-to-noise, but that is greatly mitigated by the fact that in most cases the spatial resolution is limited by multiple scattering in the phantom, not by the intrinsic tracking resolution.

The NIU/FNAL data acquisition is unconventional in that it is data driven, rather than being triggered. The signal from each SiPM is digitized at 75 MHz by pipelined ADCs, and each readout board continuously sends digitized data, including time stamps, through a 1 Gbit/s Ethernet link to a rack of 6 Linux-based workstations that build events based on the time stamps. The publication [95] shows an clean Bragg peak measured by just the range detector in a 200 MeV proton beam, and work was completed on simulating and calibrating the range detector [96]. However, there have not yet been any results or images presented from operation of the entire scanner, due to problems with malfunctioning SiPMs, difficulties with the data acquisition firmware, and expiration of funding needed for resolution of those problems (personal communication, August 2016).

Several other pCT efforts employ silicon-strip detectors for proton tracking. The raw sensors are expensive relative to scintillating fibers, especially if purchased in small quantities, but they offer significant advantages that can mitigate the cost impact in a complete system. They are solid-state devices that can be integrated into circuit boards using modern electronics assembly techniques, such as automated wire bonding, to produce compact assemblies, and the resulting system can operate with relatively low voltage (compared with GEM detectors). Scintillating fiber trackers, by comparison, require labor intensive work to position and fix in place the fibers and to integrate them with photodetectors, resulting in relatively bulky assemblies. The greatest advantage over scintillating fibers, however, is the very high signal-to-noise ratio of 20 or greater that is readily realized for silicon-strip detectors to detect non-relativistic protons, as well as minimum-ionizing particles, which can

result in a system with near 100% hit efficiency and very simple and stable calibration.

Such a system is in development in Italy by the PRIMA collaboration [97], using eight layers of silicon-strip sensors arranged to give two orthogonal views with two measurements in each view both before and after the phantom. The silicon strips are read out by a 32-channel ASIC (application-specific integrated circuit) that provides a binary readout for each channel. The energy detector is a calorimeter composed of a 2×7 array of cerium-activated yttrium aluminum garnet (YAG:Ce) crystals, which is a scintillator with a fast (70 ns) decay time. Each crystal is $3 \times 3\,\mathrm{cm}^2$ in cross section and 10 cm long, and their scintillation light is detected by silicon photodiodes. Only results from a relatively slow $5 \times 5\,\mathrm{cm}^2$ prototype with 4 crystals have been reported, with some simple phantoms reconstructed from data taken in Uppsala Sweden with a 175 MeV beam [98]. For example, see Fig. 9 for two slices from a CT image of a cylindrical phantom. In that prototype, the photodiode signal was shaped with a 1 μs time constant [99], and the data acquisition operated at a modest rate of 10,000 events per second. The goal for the full-scale system is to operate at a rate of a million events per second, but there are no published results from it yet available.

The "Phase-II Scanner" by a collaboration of Loma Linda University (LLU) and the University of California at Santa Cruz (UCSC) (with collaborators from Baylor University and NIU contributing to image reconstruction), is the most complete in terms of realizing its original project goals [51] and is a good representation of the state of the art. It builds upon experience with an earlier, smaller, and much slower scanner built by a collaboration of LLU, UCSC, and NIU [52]. The Phase-II Scanner has been demonstrated to achieve a sustained acquisition rate of 1.2 million proton histories per second in beam tests at both the LLU synchrotron and the isochronous IBA cyclotron of the Northwester Medicine Chicago Proton Center, and numerous CT scans have been completed in as little as six minutes each at the Chicago Proton Center by continuously rotating the phantom on a rotation stage during the run (scans at the synchrotron take longer, due to the dead time between proton spills). The high rate was achieved using a triggered, fully-custom hierarchal data acquisition system based on ASICs at the tracking detector front end, 14 FPGAs operating in parallel on the front-end boards, and a Xilinx Vertex-6 FPGA acting as an event builder, which sends event packets over a gigabit-per-second Ethernet link to a computer.

Similar to the PRIMA instrument, each of the scanner's front and rear tracking detectors use four layers (two in each orthogonal view) of standard single-sided silicon-strip sensors left over from fabrication of the NASA Fermi-LAT orbiting gamma-ray telescope [100], read out by a 64-channel ASIC designed by UCSC specifically for the pCT project [101]. The aperture of the scanner corresponds to the size of four of the sensors placed end-to-end and is more than large enough to encompass the width of a human head, although two separate scans are required in order to include the full height of a head.

The scanner's "5-stage" energy/range detector is a unique design that measures a combination of energy and range [49], as discussed in Section 4.3. The residual energy measurement is taken from only the stage in which the proton stops, allowing good WEPL resolution to be obtained from fast plastic scintillators with only five readout channels. The disadvantage relative to a range detector is complexity of calibration, which is achieved using data in each beam test taken with a special stepped calibration phantom having 41 different known thickness of polystyrene. Preliminary images presented in Ref. [51] show significant ring-like artifacts that resulted from calibration deficiencies, which have more recently been

greatly reduced by further work, as shown below. Nevertheless, the first CT scans published from this device in Ref. [51] do demonstrate good reconstruction of the RSP of inserts in a CATPHAN 404 phantom as well as impressive image slices from a scan of a pediatric head phantom, reproduced here in Fig. 10.

A Japanese group has also presented results from a small-aperture prototype that operates at a low data acquisition rate of 30 Hz, along with plans to build a ten times larger system that can operate at a rate of 100 kHz to 1 MHz [102]. The tracking detector employs the same Fermi-LAT sensors as employed in the LLU/UCSC Phase-II Scanner, although with an aperture corresponding to a single 9×9 cm^2 wafer. The calorimeter consists of a single 3-inch thallium-doped cesium-iodide crystal read out by a photomultiplier tube. To reach higher rates they are developing a binary readout system for the tracker, similar to what is used in the PRIMA and LLU/UCSC instruments, and are studying a range-counter concept with organic scintillators to replace the slow crystal calorimeter.

A group in Korea has also proposed a similar system based on silicon-strip tracking plus a cesium-iodide calorimeter [103], but the work that they have presented thus far is restricted to Monte Carlo simulations of the concept, including full reconstruction of tomographic images from the simulated data. Each simulated tracking layer consisted of a single 10-cm square silicon-strip detector with two orthogonal sets of strips and a relatively course strip pitch of 1 mm. A sensor thickness of 200 μm and a spacing between sensors of 10 cm were found by simulation to be optimal. From the publication it is unclear whether the group intends to proceed to build a prototype device. In fact, the specified calorimeter technology would not be capable of reaching the high rates necessary for a practical system.

A British collaboration named PRaVDA (Proton Radiotherapy Verification and Dosimetry Applications) has ambitious plans for a pCT system based entirely on silicon detectors [104]. The tracking system is more advanced than the LLU/UCSC Phase-II Scanner in several respects [21]:

- The detectors and electronics are designed to withstand high-radiation environments, to support a verification mode in which a therapy-intensity beam is used.
- A double threshold on each channel also supports the treatment mode as well as allowing tracking of two particles per channel per readout frame.
- Each tracker layer provides three views, with strips oriented 60° apart, instead of just two orthogonal views, in order to resolve pattern recognition ambiguities in multi-track events.
- The tracking readout ASIC and data acquisition have a special mode of operation in which the tracker hits are read out at regular intervals, every 100 μs, and beam profile histograms are accumulated.

A prototype is operational [105], although it has a limited aperture corresponding to only a single silicon-strip wafer and is not yet paired with an energy detector or range detector. From data obtained using a 125 MeV proton beam, some radiographs and CT images of a small phantom (7.5 cm diameter) have been published [105], for which scattering power in the phantom was used in place of measurements of energy loss. The reference does not state at what rate the proton histories were acquired, but this was the first time CT

images were produced in this manner (although for radiographs the technique has been demonstrated before [106]). Whereas using scattering power allows rather noisy images to be made in the absence of an energy or range detector, the dependence of multiple-scattering on the phantom material properties is quite different than the dependence of energy loss. Therefore, a measurement of scattering power alone would not be sufficient for use in treatment planning, although it is in itself useful information for understanding the beam penumbra.

The PRaVDA collaboration does plan to implement a range detector stack. Their concept is unique in that, in place of the usual monolithic slabs of plastic scintillator, it is to use a stack of 24 solid-state detector elements with very high spatial segmentation [107] interleaved with absorber. When used together with the PRaVDA tracker, the resulting pCT system in principle could be capable of operating at very high proton rates, at least an order of magnitude faster than the LLU/UCSC Phase-II Scanner, by tracking multiple protons in each readout frame, assuming that the range detector can be read out fast enough.

However, the PRaVDA range telescope is intended to employ CMOS Active-Pixel-Sensor (APS) technology, for which sufficiently fast readout is a major challenge. Reference [54] gives results of testing such sensors, of size 12.8×12.8 cm^2, in 36 MeV and 200 MeV proton beams. They are read out using a rolling shutter, in which rows are read out in sequence to build up a frame. A full-frame readout can be done at a rate of only 6.5 Hz, but region-of-interest readout can run faster, for example at up to 1.4 kHz for a readout of 10 rows. The tests on the existing sensors did demonstrate detection of individual protons, but the readout is much too slow for a pCT system. Reference [54] gives a design specification of 1 kHz for full-frame readout for future pCT sensors. In that case, if it were run at 10 times the proton rate as the LLU/UCSC Phase-II Scanner, there would be 10,000 proton hits per frame per layer in the range stack, leading to a formidable pattern recognition task in order to correlate those hits with the high-rate information from the silicon-strip tracker. The reference does not explain how that would be accomplished. Note, however, that the more expensive but far faster silicon-strip readout technology could also be used in such a range detector and would match well with the tracking system.

The iMPACT project [108], with a tracking component called "ProXY" [56], is the most ambitious of all in terms of planned high-rate performance in pCT. The plan is to develop CMOS monolithic active pixel sensors (MAPS) with an innovative compression readout scheme that could provide efficient tracking at a very high rate over a large area [109]. The pixel geometry would provide simultaneous tracking of multiple particles. Coupled with a fast readout, the authors envisage completing an entire CT scan in a few seconds. While it is arguable whether it is essential, as the authors state in [108], to achieve tracking speeds of a billion protons in one or a few seconds in order to deliver a practical system, it is true that modern X-ray CT systems are able to do complete scans in a fraction of a second. The authors claim that sensors of their technology could be produced as large as 10×10 cm^2 with 20 μm pixels, which is as large as a silicon-strip detector from a 6-inch wafer, but the near term goal is a 4×4 cm^2 sensor that could handle up to 250 million proton hits per second per square centimeter, with only 4 hits per frame.

Of course, a successful pCT system requires more than just particle tracking, so the authors propose a range detector in which each of the 48 to 64 detection planes is segmented into 16 fingers of plastic scintillator, with the fingers rotated by 90° from one plane to the

Table 1: List of some contemporary efforts on prototype pCT systems and particle radiography (pRad) systems with particle tracking. For operational systems the rate is approximately what has been demonstrated to date. Earlier prototypes from the same collaborations are not listed.

Collaboration	Type	Aperture (cm^2)	Tracking Technology	WEPL Detector Technology	Rate	Comment
AQUA [91]	pRad	10×10	GEM	scint. range counter	10 kHz	1 MHz planned
LLU/UCSC Phase-II [51]	pCT	36×9	Si strip	5 scint. stages	1.2 MHz	Operational
Niigata [102]	pCT	9×9	Si strip	NaI calorimeter	30 Hz	Larger, faster instr. planned
NIU, FNAL [95]	pCT	24×20	Sci Fi	scint. range counter	2 MHz	Not operational
PRaVDA [104]	pCT	4.8×4.8	Si strip	CMOS APS telescope	2.5 MHz	Only tracker operating
PRIMA [97]	pCT	5.1×5.1	Si strip	YAG:Ce calorimeter	10 kHz	20×5 cm^2 1 MHz instr. planned
PSI [85]	pRad	22.0×3.2	Sci Fi	scint. range counter	1 MHz	Program completed
QBeRT [89]	pRad	9×9	Sci Fi	Sci Fi range counter	1 MHz	Also a beam monitor

Table 2: Performance of pCT image reconstruction by ART algorithms for a CATPHAN 404 sensitometry phantom [71]. The data are from the LLU-UCSC Phase-II Scanner [51].

		DROP (10 blocks)		CARP	
Material	RSP	Mean ± Std Dev	Error	Mean ± Std Dev	Error
Teflon	1.790	1.774 ± 0.003	-0.89%	1.782 ± 0.019	-0.44%
PMP	0.883	0.897 ± 0.005	1.59%	0.882 ± 0.016	-0.11%
LDPE	0.979	0.986 ± 0.003	0.72%	0.971 ± 0.018	-0.82%
Polystyrene	1.024	1.031 ± 0.006	0.76%	1.017 ± 0.018	-0.68%
Delrin	1.359	1.349 ± 0.003	-0.74%	1.353 ± 0.015	-0.44%
Air	0.00113	0.060 ± 0.003		0.029 ± 0.010	

next, and they propose a rate goal of 200 MHz per finger. The current status is that a small prototype CMOS MAPS sensor of size 3×3 mm^2 and 256×256 pixels has been fabricated, although no test results have yet been published for it. Some results of successful testing of an earlier 32×32 pixel prototype with radiation sources have been published [56].

7 Proton CT Performance

The best illustrations of the performance of a proton-CT instrument at this time are from the LLU-UCSC Phase-II scanner. Reference [71] presents results from it that are more recent than those found in the instrument paper [51], and with much improved calibration. For example, Fig. 11 shows a slice from the pCT image of the sensitometry phantom used to obtain the results in Table 2. The ring artifacts seen in Ref. [51] are no longer evident. The calibration improvements were obtained by means of a new calibration phantom consisting of a smooth wedge, rather than discrete steps. It provides a continuous range of WEPL values, which was found to be particularly important for protons that stop near the boundaries of the five scintillator stages. Perhaps the most important result demonstrated in Table 2 is that proton CT can achieve an RSP resolution and accuracy of better than one percent.

The spatial resolution of the LLU/UCSC Phase-II scanner was thoroughly studied by a UCSC graduate student and published in Ref. [110]. She evaluated the resolution in terms of the instrument's modulation transfer function (MTF), using data acquired with a

custom phantom that she designed specifically for this purpose. It is a plastic cylinder with embedded cubes that simulate enamel, cortical bone, lung, and air. Figure 12 shows a slice from a pCT image of the phantom, and Fig. 13 shows the measured spatial resolution, which depends strongly, of course, on the amount of material through which the proton passed. The results show that the best spatial resolution that one can expect for objects embedded deeply within a phantom is about 2 mm. This is limited by multiple scattering, which cannot be avoided, but it is about double the maximum theoretical MLP uncertainty for protons passing along a 25 cm WEPL [110] (see also Fig. 3), suggesting that it might be possible for the image reconstruction algorithm itself to improve in terms of approaching the inherent resolution limit. Note that while this measured MTF represents how fuzzy an edge will look in an image, of course the position of the edge could be localized to much greater accuracy, depending on how many protons pass through the region. Also, it is worth mentioning again that the principal purpose of pCT and proton radiography is not to compete with X-ray CT in terms of image sharpness but rather to improve knowledge of the proton RSP throughout the tissue.

Another point of interest for pCT, however, is its very low susceptibility to production of artifacts (assuming, of course, that the calibration is correct). Figure 14 shows an interesting comparison of pCT and X-ray CT images of a pediatric head phantom in a slice that passes through a gold tooth [71]. The X-ray image shows extremely large artifacts radiating from the gold tooth location, whereas nothing of the sort is visible in the pCT image. Proton CT and radiography do not suffer from beam hardening, a phenomenon inherent in X-ray CT in which a high density, high-atomic-number object absorbs preferentially the lower energy photons in a necessarily multichromatic beam, leaving a "harder" beam to pass through the remaining material [111]. Such effects are difficult to account for and correct in the image reconstruction.

As discussed in Section 3, low radiological dose was emphasized in the early treatments of pCT, motivated by the very high detection efficiency possible with protons. Proving the point is not trivial, because there are many ways to lose efficiency in a detector system. Protons can be lost because they arrive too closely in time, because an essential detector element is inefficient, because the data acquisition buffers are full, or because they pass through a crack or dead channel in the detector system. They might also be rejected because they scattered too much, did not leave a reasonable signal in the range or energy detector (possibly due to nuclear interactions), or scattered out of the acceptance. During the image reconstruction additional proton histories will be rejected because the measured WEPL lies in the tails of the distribution from protons passing along nearly the same path. Overall, even with the best system less than half of the protons that contribute to the dose will end up being used to reconstruct the image. Additional dose might also be incurred from neutrons produced in nuclear interactions in the patient as well as the detector system.

With respect to dose performance in pCT, the LLU/UCSC Phase-II system and the PRIMA system are good examples of the state of the art. For the Phase-II scanner, the dose was measured during a 6-minute pCT scan by placing an ionization chamber at the center of a 16-cm diameter acrylic cylinder (Catphan CTP554 dose phantom). The tracking system measured a flux of about 1.2 million protons per cm^2 passing through the phantom center, more than sufficient to produce a good CT image, while a dose of 1.4 mGy was measured by the ionization chamber [112]. The PRIMA collaboration published a dose measurement for

their pCT scan of a metallic phantom with inserts of various densities (Fig. 9). They inferred the dose from the observed tracks, using Monte Carlo simulations to correct for unobserved effects, and found that with a 1.3 mGy dose they achieved better than 1% density resolution. Such a dose is small compared with the 30–50 mGy dose typically delivered by an X-ray CT scan of the head [113]. Note that a single radiograph can be expected to deliver a dose of only around 0.01 mGy. The corresponding effective dose will most likely be in the μSv range, but a careful radiobiological evaluation will be needed to confirm that.

8 Proton Radiography and Computed Tomography Outlook

Thus far proton and ion radiography has seen very little use in the clinic, even for range verification and quality control, two decades after the paper by Schneider and Pedroni [17]. Nevertheless, based on the number of instrumentation efforts that are underway this decade, interest appears to be growing along with the rapid worldwide growth in the number of proton and ion therapy centers. Particle integrating systems could see the first widespread clinical use due to their advantages in terms of expense and simplicity, especially for carbon ions, but they will never provide the good spatial resolution that can be afforded by particle tracking, and they also are unlikely to be able to operate with as low a dose as a particle tracking based system. Furthermore, a tracking based system can be made compact and simple. Systems like the LLU/UCSC Phase-II scanner have demonstrated very high tracking reliability and high data acquisition throughput via compact, inexpensive FPGAs that stream data over a single Ethernet cable to a standard PC, and they can acquire a quality radiograph in a few seconds with a dose the order of 0.01 mGy. The particular Phase-II scanner five-stage energy/range detector requires daily, relatively complex calibration, but such a tracking system paired with a true range detector could operate day after day with essentially no recalibration, since all of the signal processing would be based on simple thresholds. Thus existing technology is ready for the task.

For proton radiography to move toward widespread clinical use, however, will require industrial involvement to manufacture and market a standardized device ready for use. That will probably best be done by or in collaboration with one of the existing proton-therapy facility vendors, because to be useful the system will have to be well integrated into the control system of the facility. For example, for the very low beam currents envisaged, the accelerator operators will likely need feedback from the radiography scanner in order even to know whether beam is being delivered. The stage is now set for this to happen, as not only does appropriate technology exist, but also research already long ago demonstrated convincing advantages to having such an instrument available for range verification and image guidance [17, 114]. Already at least one private firm, ProtonVDA Inc., is engaged in designing a proton radiography/CT system, with a goal of measuring 10^7 protons per second [115]. Their concept appears to be somewhat different from Fig. 2, but not enough information is yet public to ascertain in detail how it functions or how effective it will be.

The outlook for pCT is less certain. Research still needs to be completed demonstrating significant gains in treatment planning quality relative to, for example, dual-energy X-ray CT. Even given some advantage, however, one has to consider the considerable added expense

of bringing the patient into the proton beam line prior to treatment planning as well as for the treatment itself. An additional problem is the limited beam energy (e.g. 230 MeV) available in most proton therapy centers, which would preclude doing pCT in some parts of the body, especially for larger patients. A dedicated very-low-intensity, poor-emittance (i.e. potentially inexpensive) proton accelerator dedicated to diagnostic and planning work would be a game changer, as was recognized as far back as 1975 [25], but nothing of the sort is on the horizon. Also, the best existing pCT system still requires the order of six minutes to complete a scan and a similar time, employing considerable computing resources, to generate an image. More development can almost certainly shrink those times but may also increase the expense. Nevertheless, it is interesting to note that an optimal system for radiography would look and perform very similar to one of the existing tracking-based scanners designed for pCT. Hence in the near future such a device located in a treatment room could serve the needs of pCT in special cases for which the extra expense and time are warranted while more frequently serving the needs of proton radiography for range verification and image guidance.

The ultimate pCT system would be mounted on a gantry, making use of the gantry to rotate 180 degrees or more about the patient during the scan, but to get there would be a huge, unlikely step at this point, requiring major redesign efforts of proton therapy facilities. A much more practical starting point would be to rotate the patient while he or she is sitting in a chair custom designed for that purpose. It is not difficult to envisage for the near future, then, a fast tracking-based proton radiography system in a fixed-beam treatment room that could also be used on occasion with a rotating chair for pCT, particularly for the head, for which limited beam energy is normally not a problem.

References

[1] Robert R. Wilson. Radiological use of fast protons. *Radiology*, 47:487–491, 1946.

[2] W.H. Bragg. The reflexion of x-rays by crystals. *Proc. R. Soc. Lond. A*, 88:428–438, 1913.

[3] Wayne D. Newhauser and Rui Zhang. The physics of proton therapy. *Phys. Med. Biol.*, 60:R155–R209, 2015.

[4] Alfred R. Smith. Proton therapy. *Phys. Med. Biol.*, 51:R491–R504, 2006.

[5] L. Lin, C.G. Ainsley, and J.E. McDonough. Experimental characterization of two-dimensional pencil beam scanning proton spot profiles. *Phys. Med. Biol.*, 58:6193–6204, 2013.

[6] H.M. Kooy et al. A case study in proton pencil-beam scanning delivery. *Int. J. Radiation Oncology Biol. Phys.*, 76:624–630, 2010.

[7] D.J. Brenner, I. Shuryak, G. Jozsef, and et al. Risk and risk reduction of major coronary events associated with contemporary breast radiotherapy. *JAMA Intern. Med.*, 174:158–160, 2014.

[8] S.C. Darby et al. Risk of ischemic heart disease in women after radiotherapy for breast cancer. *New England J. Med.*, 368:987–998, 2013.

[9] Steven J. Frank. Study #2012-0825, phase II/III randomized trial of intensity-modulated proton beam therapy (IMPT) versus intensity-modulated photon therapy (IMRT) for the treatment of oropharyngeal cancer of the head and neck. `https://www.mdanderson.org/patients-family/diagnosis-treatment/clinical-trials/clinical-trials-index/clinical-trials-detail.ID2012-0825.html`. Accessed: 2016-11-28.

[10] G. Chen, R. Singh, J. Castro, J. Lyman, and J. Quivey. Treatment planning for heavy ion radiotherapy. *J. Rad. Oncol. Biol. Phys.*, 5:1809–1819, 1979.

[11] E. Alpen, D. Saunders, A. Chatterjee, J. Llacer, G. Chen, and J. Scherer. A comparison of water equivalent thickness measurements: CT method versus heavy ion beam technique. *Br. J. Radiol.*, 58:542–548, 1985.

[12] 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. *Phys. Med. Biol.*, 43:1579–1592, 1998.

[13] H. Paganetti. Range uncertainties in proton therapy and the role of monte carlo simulations. *Phys. Med. Biol.*, 57:R99–R117, 2012.

[14] M.R. Gensheimer et al. In vivo proton beam range verification using spine MRI changes. *Int. J. Rad. Oncol. Biol. Phys.*, 78:268–275, 2010.

[15] A.C. Knopf and A. Lomax. In vivo proton range verification: a review. *Phys. Med. Biol.*, 58:R131–R160, 2013.

[16] Thorsten R.C. Johnson. Dual-energy ct: General principles. *Am. J. Roentgenology*, 199:S3–S8, 2012.

[17] U. Schneider and E. Pedroni. Proton radiography as a tool for quality control in proton therapy. *Med. Phys.*, 22:353–363, 1995.

[18] Seyjoon Park et al. Proton-radiography-based quality assurance of proton range compensator. *Phys. Med. Biol.*, 58:6511–6523, 2013.

[19] A.M. Koehler. Proton radiography. *Science*, 160:303–304, 1968.

[20] R.W. Schulte et al. Density resolution of proton computed tomography. *Med. Phys.*, 32:1035–1046, 2005.

[21] J.T. Taylor et al. A new silicon tracker for proton imaging and dosimetry. *Nucl. Instrum. Meth. A*, 831:362–366, 2016.

[22] National Institute of Standards and Technology (NIST). pstar: stopping-power and range tables for protons. `http://physics.nist.gov/PhysRefData/Star/Text/PSTAR.html`. Accessed: 2016-11-29.

[23] A.M. Cormack. Representation of a function by its line integrals, with some radiological applications. *J. Applied Phys.*, 34:2722, 1963.

[24] A.M. Cormack and A.M. Koehler. Quantitative proton tomography: preliminary experiments. *Phys. Med. Biol.*, 21:560–569, 1976.

[25] R. Martin, M. Foss, J. Moenich, and R. Lari. The proton diagnostic accelerator. *IEEE Trans. Nucl. Sci.*, 22:1802–1804, 1975.

[26] Christopher L. Vaughan. *Imagining the Elephant, A Biography of Allan MacLeod Cormack*. Imperial College Press, London, 2008.

[27] S.L. Kramer, D.R. Moffett, R.L. Martin, E.P. Colton, and V.W. Steward. Proton radiography application to medical imaging. In *Proc. Ann. Mtg. Radiological Soc. N. America*, Chicago, Illinois, 1978.

[28] V.W. Steward and A.M. Koehler. Proton radiographic detection of strokes. *Nature*, 245:38–40, 1973.

[29] V.W. Steward and A.M. Koehler. Proton beam radiography and tumor detection. *Science*, 179:913–914, 1973.

[30] V.W. Steward and A.M. Koehler. Proton radiography in the diagnosis of breast carcinoma. *Radiology*, 110:217–221, 1974.

[31] V.W. Steward and A.M. Koehler. Proton radiography of a human brain tumor within the skull: a preliminary report. *Surg. Neurology*, 2:283–284, 1974.

[32] K.M. Hanson et al. Proton computed tomography of human specimens. *Phys. Med. Biol.*, 27:25–36, 1982.

[33] C. Morris, J.W. Hopson, and P. Goldstone. Proton radiography. *Los Alamos Science*, 30:32–45, 2006.

[34] Kenneth M. Hanson. Proton computed tomography. *IEEE Trans. Nucl. Sci.*, NS-26:1635–1640, 1979.

[35] Y. Takada, K. Kondo, T. Marume, K. Nagayoshi, I. Okada, and K. Takikawa. Proton computed tomography with a 250 MeV pulsed beam. *Nucl. Instrum. Meth. Phys. Res. A*, 410:410–422, 1988.

[36] P. Zygmanski, K.P. Gall, M.S.Z. Ragin, and S.J. Rosenthal. The measurement of proton stopping power using proton-cone-beam computed tomography. *Phys. Med. Biol.*, 45:511–528, 2000.

[37] V.A. Bashkirov, R.P. Johnson, H.F.-W. Sadrozinski, and R.W. Schulte. Development of proton computed tomography detectors for applications in hadron therapy. *Nucl. Instrum. Meth. A*, 809:120–129, 2016.

[38] S. Rit, G. Dedes, N. Freud, D. Sarrut, and J.M. Letang. Filtered backprojection proton CT reconstruction along most likely paths. *Medical Physics*, 40:031103, 2013.

[39] Richard Gordon, Robert Bender, and Gabor Herman. Algebraic reconstruction techniques (ART) for threedimensional electron microsocopy and X-ray photography. *J. Theor. Biol.*, 29:471–481, 1970.

[40] K. Parodi. Heavy ion radiography and tomography. *Physica Medica*, 30:539–543, 2014.

[41] G. Poludniowski, N.M. Allinson, and P.M. Evans. Proton radiography and tomography with application to proton therapy. *Br J Radiol.*, 88:20150134, 2015.

[42] Martina Bucciantonio and Fabio Sauli. Proton computed tomography. *Mod. Phys. Lett. A*, 30:1540024, 2015.

[43] R. Amblard, V. Floquet, G. Angellier, J.M. Hannoun-Lévi, and J. Hérault. Imageri protonique pour la protonthérapie: état de l'art. *Cancer/Radiothérapie*, 19:139–151, 2015.

[44] First in Architecture. Average male and female dimensions / heights. `http://www.firstinarchitecture.co.uk/average-male-and-female-dimensions`. Accessed: 2016-11-29.

[45] J. Beringer et al. (Particle Data Group). Review of particle physics. *Phys. Rev. D*, 86:323–338, 2012.

[46] John R. Letaw, R. Silberberg, and C.H. Tsao. Proton-nucleus total inelastic cross sections: an empirical formula for $E > 10$ MeV. *Ap.J. Supplement*, 51:271–276, 1983.

[47] S.B. Crowe et al. Relationships between gamma criteria and action levels: Results of a multicenter audit of gamma agreement index results. *Med. Phys.*, 43:1501–1506, 2016.

[48] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[49] V.A. Bashkirov et al. Novel scintillation detector design and performance for proton radiography and computed tomography. *Medical Physics*, 43:664, 2016.

[50] J.B. Birks. Scintillations from organic crystals: Specific fluorescence and relative response to different radiations. *Proc. Phys. Soc. A*, 64:874, 1951.

[51] R.P. Johnson et al. A fast experimental scanner for proton CT: technical performance and first experience with phantom scans. *IEEE Trans. Nucl. Sci.*, 63:52, 2016.

[52] H.F.-W Sadrozinski et al. Detector development for proton computed tomography (pCT). *Proc. IEEE NSS/MIC Conf., Valencia, Spain*, pages 4457–4461, 2011.

[53] Joao Seco and Nicolas Depauw. Proof of principle study of the use of a cmos active pixel sensor for proton radiography. *Med. Phys.*, 38:622–623, 2011.

[54] G. Poludniowski et al. Proton-couting radiography for proton therapy: a proof of principle using CMOS APS technology. *Phys. Med. Biol.*, 59:2569–2581, 2014.

[55] X. Llopart, R. Ballabriga, M. Campbell, L. Tlustos, and W. Wong. Timepix, a 65k programmable pixel readout chip for arrival time, energy and/or photon counting measurements. *Nucl. Inst. and Meth. A*, 581:485–494, 2007.

[56] S. Mattiazzo et al. Advanced proton imaging in computed tomography. *Rad. Protection Dosimetry*, 166:388–392, 2015.

[57] M. Christophersen et al. Alumina and silicon oxide/nitride sidewall passivation for p-and-type sensors. *Nucl. Instrum. Meth. A*, 699:14–17, 2013.

[58] G. Battistoni et al. The FLUKA code: An accurate simulation tool for particle therapy. *Front. Oncol.*, 6:16, 2016.

[59] J. Perl et al. TOPAS: an innovative proton monte carlo platform for research and clinical applications. *Medical Physics*, 39:6818–6837, 2012.

[60] V. Giacometti et al. Software platform for simulation of a prototype proton ct scanner. *Medical Physics*, 44:1002–1016, 2017.

[61] Gabor T. Herman. *Fundamentals of Computerized Tomography.* Springer-Verlag, London, 2009.

[62] A. Padole, R. Khawaja, M. Kalra, and S.Singh. Ct radiation dose and iterative reconstruction techniques. *American J. Roentgenology*, 204:W384–W392, 2015.

[63] Steven W. Smith. *The Scientist and Engineer's Guide to Digital Signal Processing.* California Technical Publishing, 2011.

[64] S.N. Penfold, R.W. Schulte, Y. Censor, and A.B. Rosenfeld. Total variation superiorization schemes in proton computed tomography image reconstruction. *Med. Phys.*, 37:5887–5895, 2010.

[65] D. Butnariu, R. Davidi, G.T. Herman, and I.G. Kazantsev. Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems. *IEEE J. Sel. Topics Signal Proc.*, 1:540–547, 2007.

[66] Stefan Kaczmarz. Angenäherte auflösung von systemen linearer gleichungen. *International Bulletin of the Polish Academy of Science and Letters*, 35:355–357, 1937.

[67] L.M. Bregman. The method of successive projections for find a common point of convex sets. *Soviet Mathematics Doklady*, 6:688–692, 1965.

[68] Scott N. Penfold. *Image Reconstruction and Monte Carlo Simulations in the Development of Proton Computed Tomography for Applications in Proton Radiation Therapy.* PhD thesis, University of Wollongong, 2010.

[69] Kiriakos Kutulakos and Steven Seitz. A theory of shape by space carving. *Int. J. Computer Vision*, 38:199–218, 2000.

[70] Blake Schultze, Micah Witt, Yair Censor, Reinhard Schulte, and Keith Schubert. Performance of hull-detection algorithms for proton computed tomography reconstruction. *Contemporary Mathematics*, 636:211–224, 2015.

[71] Caesar Ordoñez et al. A real-time image reconstruction system for particle treatment planning using proton computed tomography (pCT). In *Proc. 24th Conf. on Applications of Accelerators in Research and Industry, Fort Worth, TX*, 2016. arXiv:1707.05014.

[72] Nicholas T. Karonis et al. Distributed and hardware accelerated computing for clinical medical imaging using proton computed tomography (pCT). *J. Parallel Distrib. Comput.*, 73:1605–1612, 2013.

[73] B. Schultze et al. Reconstructing highly accurate relative stopping powers in proton computed tomography. In *Proc. IEEE NSS/MIC Conf., San Diega, CA.*, 2015.

[74] Y. Censor, T. Elfving, G.T. Herman, , and T. Nikazad. On diagonally-relaxed orthogonal projection methods. *SIAM J. Sci. Computing*, 30:473–504, 2008.

[75] D. Gordon and R. Gordon. Component-averaged row projections: A robust block-parallel scheme for sparse linear systems. *SIAM J. Sci. Computing*, 27:1092–1117, 2005.

[76] R.W. Schulte, S.N. Penfold, J.T. Tafas, and K.E. Schubert. A maximum likelihood proton path formalism for application in proton computed tomography. *Med. Phys.*, 35:4849–4856, 2008.

[77] D.C. Williams. The most likely path of an energetic charged particle through a uniform medium. *Phys. Med. Biol.*, 49:2899–2911, 2004.

[78] R.W. Schulte, R.W. Penfold, J.T. Tafas, and K.E. Schubert. A maximum likelihood proton path formalism for applications in proton computed tomography. *Medical Physics*, 35:4849, 2008.

[79] C-A Collins Fekete, P. Doolan, M.F. Dias, L. Beaulieu, and J. Seco. Developing a phenomenological model of the proton trajectory within a heterogeneous medium required for proton imaging. *Phys. Med. Biol.*, 60:5071–5082, 2015.

[80] Regina Rescigno, Cécile Bopp, Marc Rousseau, and David Brasse. A pencil beam approach to proton computed tomography. *Medical Physics*, 42:6610, 2015.

[81] I. Rinaldi, S. Brons, J. Gordon, R. Panse, O. Jäkel, and K. Parodi. Experimental characterization of a prototype detector system for carbon ion radiography and tomography. *Phys. Med. Biol.*, 58:413–427, 2013.

[82] S. Braccini et al. First results on proton radiography with nuclear emulsion detectors. *J. Instrum.*, 5:P09001, 2010.

[83] M. Testa et al. Proton radiography and proton computed tomography based on time-resolved dose measurements. *Phys. Med. Biol.*, 58:8215–8233, 2013.

[84] P.J. Doolan, M. Testa, G. Sharp, E.H. Bentefour, G. Royle, and H-M Lu. Patient-specific stopping power calibration for proton therapy planning based on single-detector proton radiography. *Phys. Med. Biol.*, 60:1901–1917, 2015.

[85] P. Pemler, J. Besserer, J. de Boer, M. Dellert, C. Gahn, M. Moosburger, U. Schneider, F. Pedroni, and H. Stäuble. A detector system for proton radiography on the gantry of the paul-scherrer-institute. *Nucl. Instrum. Meth. Phys. Res. A*, 432:483–495, 1999.

[86] C.A. Collins-Fekete, S. Brousmiche, S.K.N. Portillo, and J. Seco. A maximum likelihood method for high resolution proton radiography/proton CT. *Phys. Med. Biol.*, 61:8232–8248, 2016.

[87] U. Scneider et al. First proton radiograph of an animal patient. *Med. Phys.*, 31:1046–1051, 2004.

[88] D. Lo Presti et al. Optical fiber folded scintillating extended tracker. *Nucl. Instrum. Meth. A*, 737:195–202, 2014.

[89] D. Lo Presti et al. Design and characterization of a real time proton and carbon ion radiography system based on scintillating optical fibres. *Physica Medica*, 32:1124–1134, 2016.

[90] G. Gallo et al. QBeRT: an innovative instrument for qualifcation of particle beam in real time. *J. of Instrumentation*, 11:C11014, 2016.

[91] M. Bucciantonio, U. Amaldi, F. Sauli, and D. Watts. Development of a fast proton range radiography system for quality assurance in hadrontherapy. *Nucl. Instr. Meth. A*, 732:564–567, 2013.

[92] U. Amaldi et al. Construction, test and operation of a proton range radiography system. *Nucl. Instr. Meth. A*, 629:337–344, 2011.

[93] F. Sauli. Gem: A new concept for electron amplification in gas detectors. *Nucl. Instrum. Meth. A*, 386:531–534, 1997.

[94] Martina Bucciantonio. Proton range radiography. `http://indico.cern.ch/event/365380/contributions/1780230/attachments/726455/996903/Bucciantonio_PRR_2015_03_20.pdf`. Accessed: 2016-12-29.

[95] Md. Naimuddin et al. Development of a proton computed tomogaphy detector system. *J. Instrum.*, 11, 2016.

[96] S.A. Uzunyan et al. Calibration and GEANT4 simulations of the Phase II proton computed tomography (pCT) range stack detector. *arXiv:1601.00249*, 2016.

[97] M Scaringella et al. The PRIMA (PRoton IMAging) collaboration: Development of a proton computed tomography apparatus. *Nucl. Instr. Meth. A*, 730:178–183, 2013.

[98] M. Bruzzi et al. Proton computed tomography images with algebraic reconstruction. *Nucl. Instrum. Meth. A*, page http://dx.doi.org/10.1016/j.nima.2016.05.056, 2016.

[99] M. Scaringella et al. A proton computed tomography based medical imaging system. *J. Instrum.*, 9, 2014.

[100] W. Atwood et al. Design and initial tests of the tracker-converter of the gamma-ray large area space telescope. *Astroparticle Physics*, 28:422–434, 2007.

[101] R.P. Johnson, J. DeWitt, C. Holcomb, S. Macafee, H.F.-W. Sadrozinski, and D. Steinberg. Tracker readout asic for proton computed tomography data acquisition. *IEEE Trans. Nucl. Sci.*, 60:3262–3269, 2013.

[102] Y. Saraya, T. Izumikawa, J. Goto, T. Kawasaki, and T. Kimura. Study of spatial resolution of proton computed tomography using a silicon strip detector. *Nucl. Instrum. Meth. A*, 735:485–489, 2014.

[103] C.Y. Lee, H. Song, C.W. Park, Y.H. Chung, J.S. Kim, and J.C. Park. Optimization of proton CT detector system and image reconstruction algorithm for on-line proton therapy. *PLoS ONE*, 11, 2016.

[104] J.T. Taylor et al. Proton tracking for medical imaging and dosimetry. *J. Instrum.*, 10(2):C02015, 2015.

[105] J.T. Taylor et al. An experimental demonstration of a new type of proton computed tomography using a novel large-area silicon tracking detector. *Medical Physics*, 2016.

[106] T. Plautz et al. 200 MeV proton radiography studies with a hand phantom using a prototype proton CT scanner. *IEEE Trans. Med. Imaging*, 33:875–881, 2014.

[107] M. Esposito et al. Cmos active pixel sensors as energy-range detectors for proton computed tomography. *j. Instrum.*, 10, 2015.

[108] P. Giubilato et al. iMPACT: innovative pCT scanner. In *Proc. IEEE NSS/MIC Conf.*, San Diego, CA, 2015.

[109] P. Giubilato et al. Lepix, a high resistivity, fully depleted monolithic pixel detector. *Nucl. Inst. and Meth. A*, 732:91–94, 2013.

[110] T.E. Plautz et al. An evaluation of spatial resolution of a prototype proton CT scanner. *Med. Phys.*, 43:6291, 2016.

[111] R.A. Brooks and G. Di Chiro. Beam hardening in x-ray reconstructive tomography. *Phys. Med. Biol.*, 21:390–398, 1976.

[112] R.P. Johnson et al. Results from a prototype proton-CT head scanner. In *Proc. 24th Conf. on Applications of Accelerators in Research and Industry, Fort Worth, TX*, 2016. arXiv:1707.01580.

[113] F.A. Mettler Jr. et al. Effective doses in radiology and diagnostic nuclear medicine: a catalog. *Radiology*, 248:254–263, 2008.

[114] J.L. Romero et al. Patient positioning for proton therapy using a proton range telescope. *Nucl. Instrum. Meth. A*, 356:558–565, 1995.

[115] E. DeJohngh, F. DeJohngh, V. Rykalin, J. Welsh, and M. Pankuch. A fast monolithic system for proton imaging. In *55th conf. Particle Therapy Co-operative Group, Prague, Czech Republic*, 2016. Accessed 5 Jan. 2017.

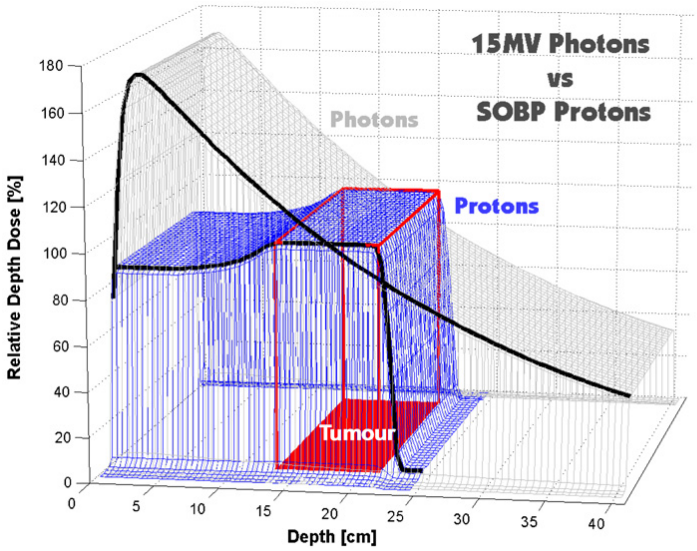

Figure 1: A comparison of depth doses for 15 MV photons and range/intensity modulated protons of variable energy [4].

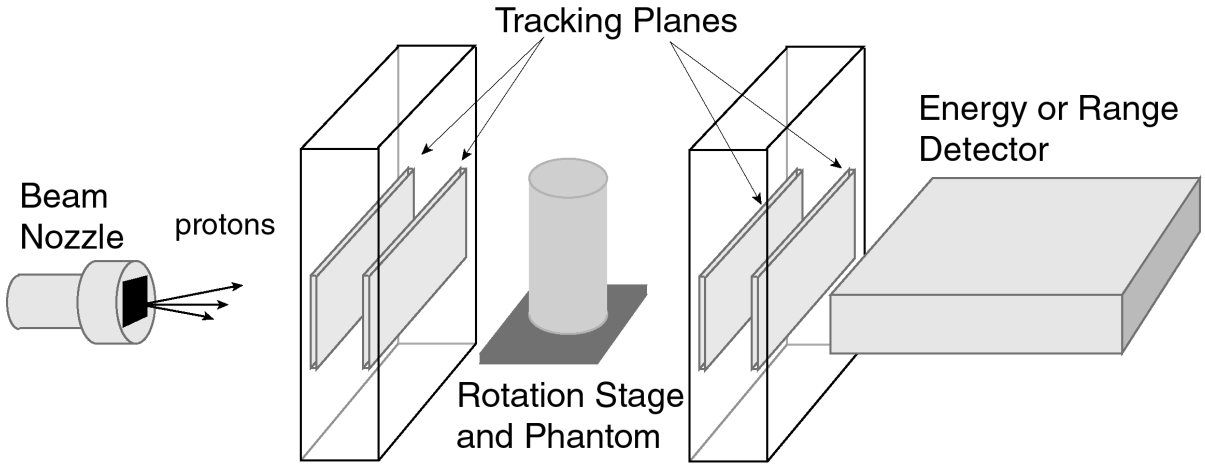

Figure 2: Schematic of a typical contemporary pCT scanner designed to measure individual proton histories.

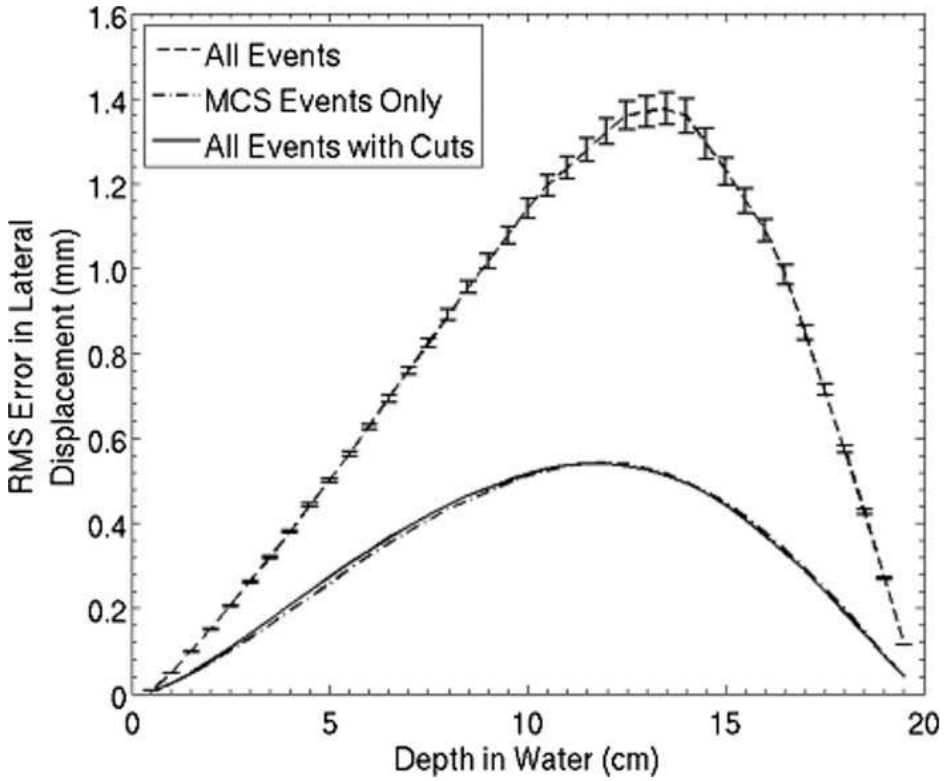

Figure 3: The Monte Carlo predicted deviation of the most likely path (MLP) of a proton from its actual trajectory, for 200 MeV protons passing through 200 mm of water. Taken from [78]. Most large deviations due to simulated nuclear interactions and hard scattering were easily eliminated by cutting out events with large exit angles, leaving events affected mainly by multiple Coulomb scattering.

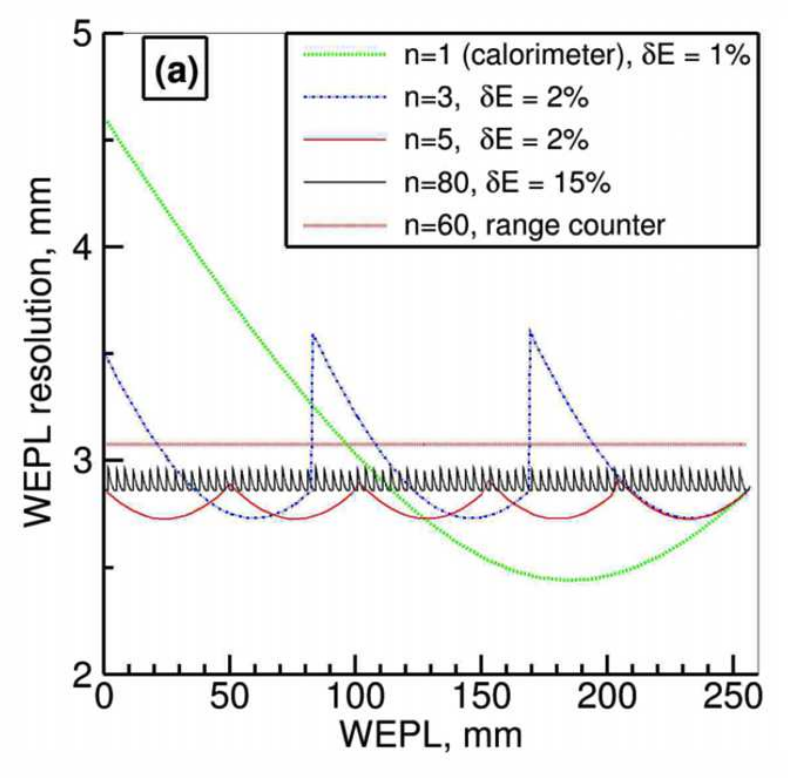

Figure 4: Models of WEPL resolution for various detector concepts, from Reference [49], Figure 2(a). The parameter n is the number of independent detector elements in depth, each of which measures the energy deposition, except for the case of the range counter, for which a simple threshold is applied to each detector signal to indicate the presence of the particle.

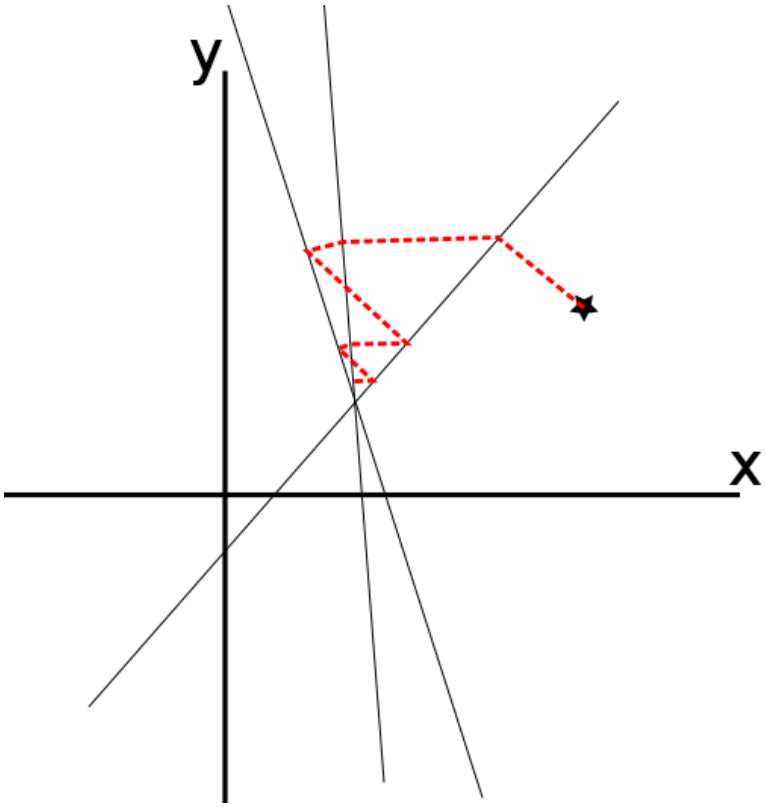

Figure 5: A trivial example of using an iterative sequence of othogonal projections to solve a linear problem. Starting from the star, the successive orthogonal projections onto the three lines converge toward their intersection.

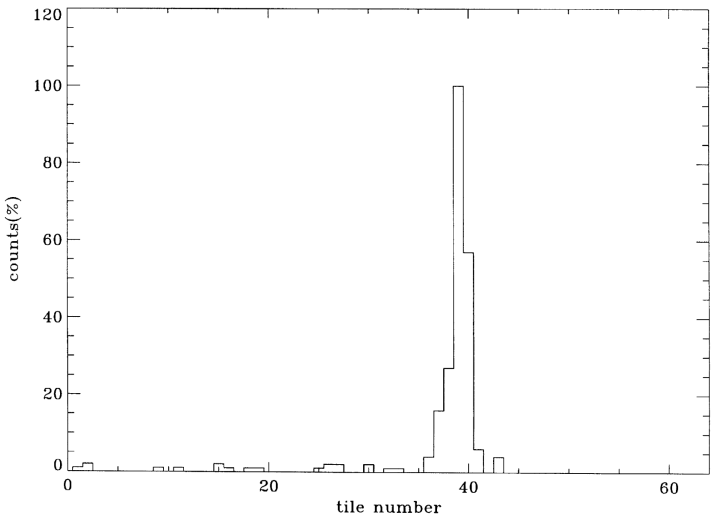

Figure 6: A typical range spectrum from the PSI range detector [85], collected from a monoenergetic sample of protons. The counts in front of the peak result from protons that undergo nuclear interactions.

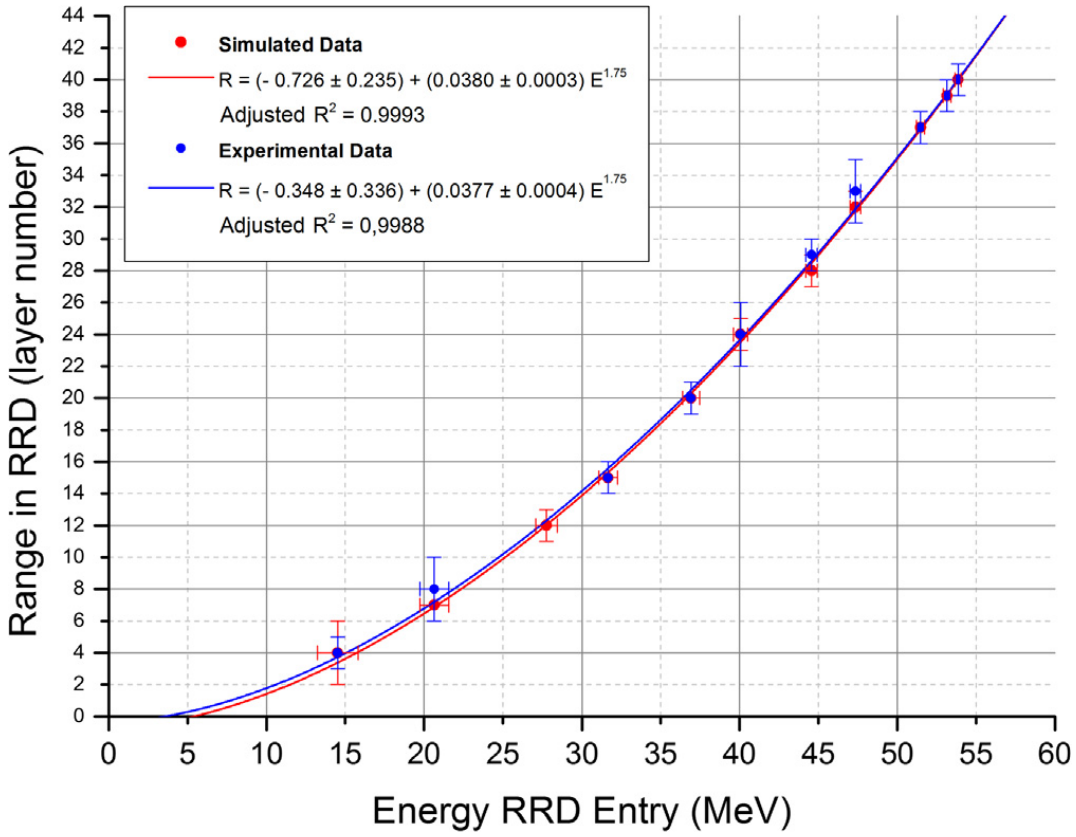

Figure 7: Experimental and simulated range measurements using the residual range detector of the QBeRT radiography system [89]. The acronym RRD refers to the "residual range detector."

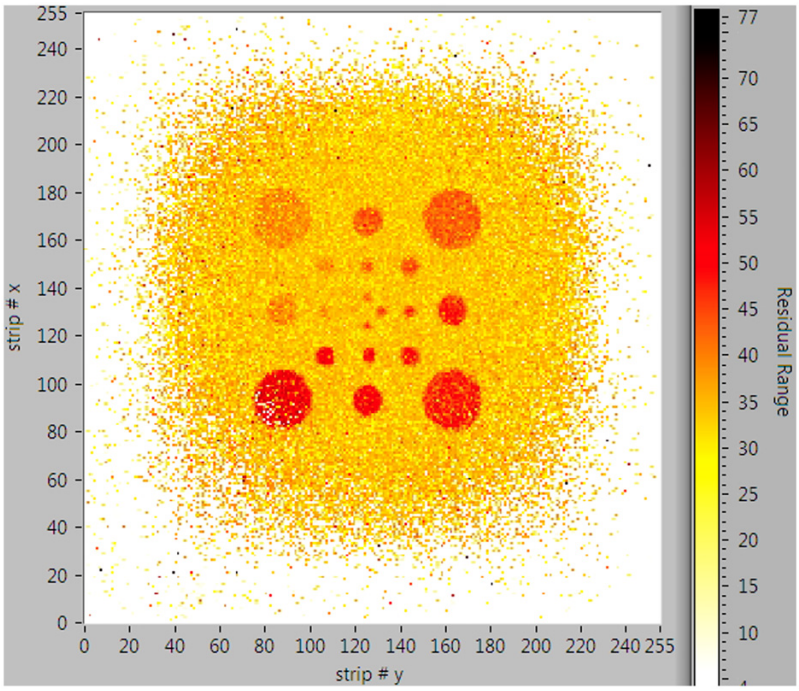

Figure 8: Radiograph of a 2 cm thick acrylic plate with holes of various size and depth drilled in it, made by the AQUA program of the TERA Foundation [92]. The strip pitch in the GEM tracking detectors is 0.4 mm, and units of mm are used for the residual range color scale.

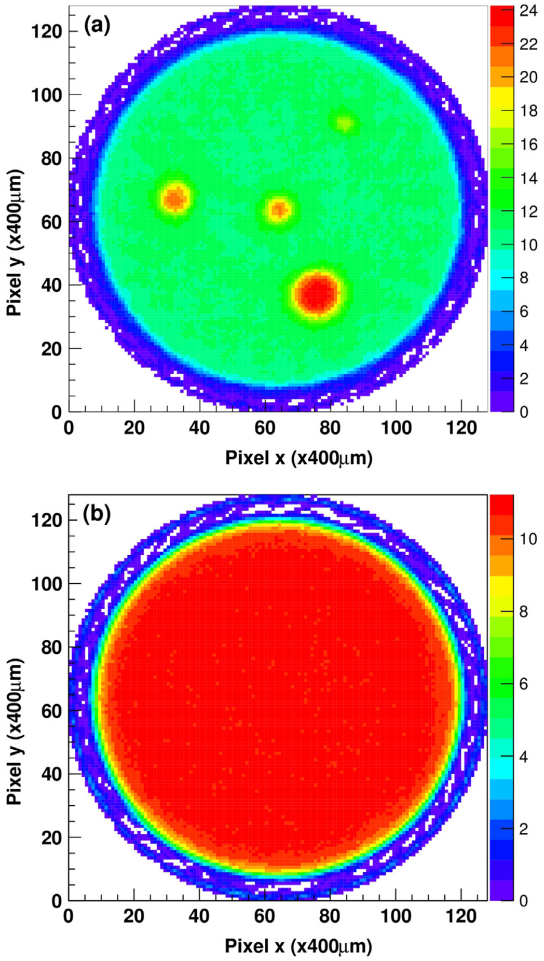

Figure 9: Tomographic reconstructions of a phantom imaged by the PRIMA prototype proton-CT scanner [98]. The top figure is a slice passing through copper and iron inserts, and the bottom figure is a slice passing through a homogenous cylinder. The data were accumulated using a 175 MeV beam of protons at a rate of 10 kHz. The color palette represents the measured stopping power in units of MeV/cm.

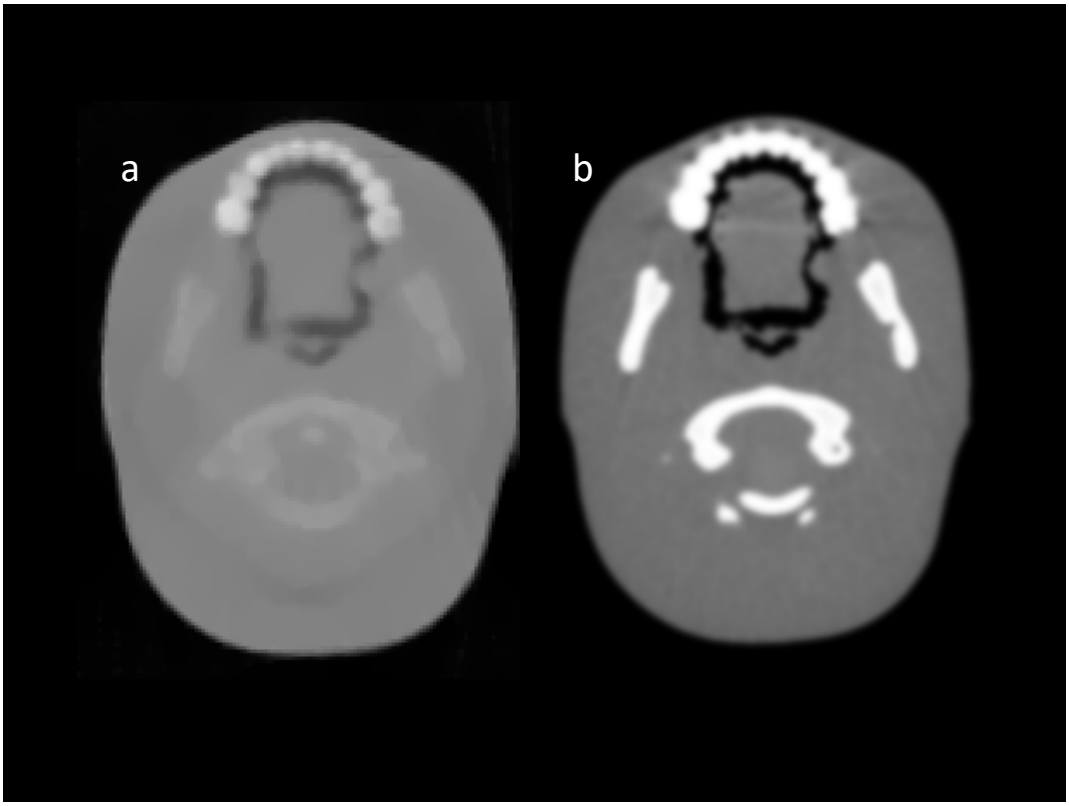

Figure 10: Slices from pCT (a) and X-ray CT (b) scans of a pediatric head phantom, from Ref. [51].

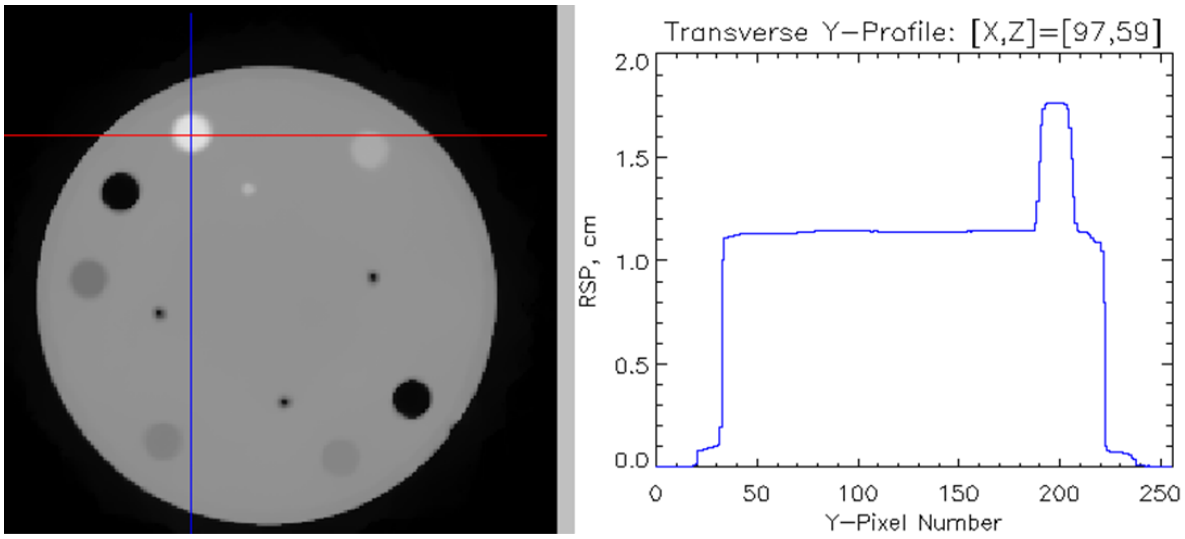

Figure 11: A slice from the pCT of the CATPHAN 404 sensitometry phantom used to obtain the RSP results given in Table 2, from Ref. [112]. The profile on the right was taken along the vertical line plotted over the slice.

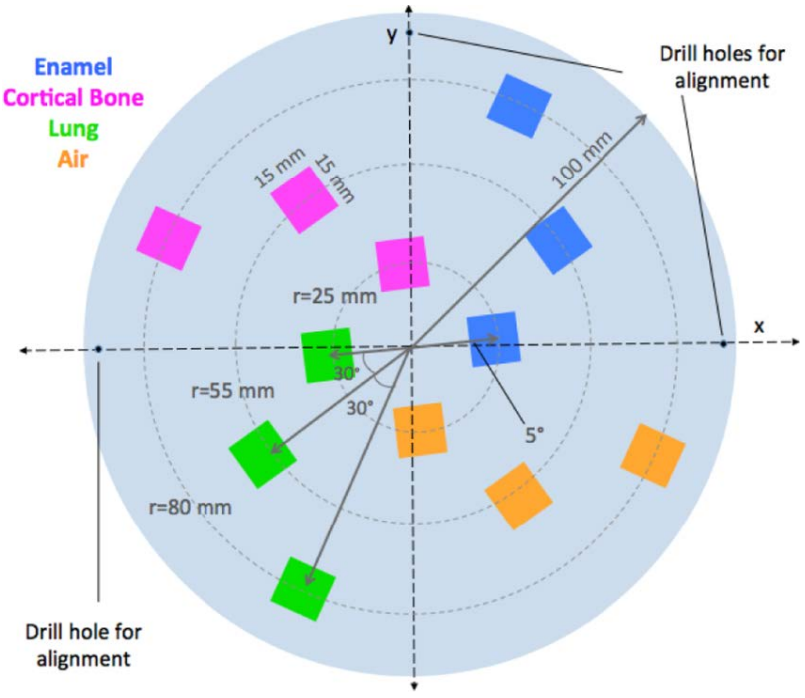

Figure 12: The phantom designed for measurement of pCT spatial resolution. Taken from [110].

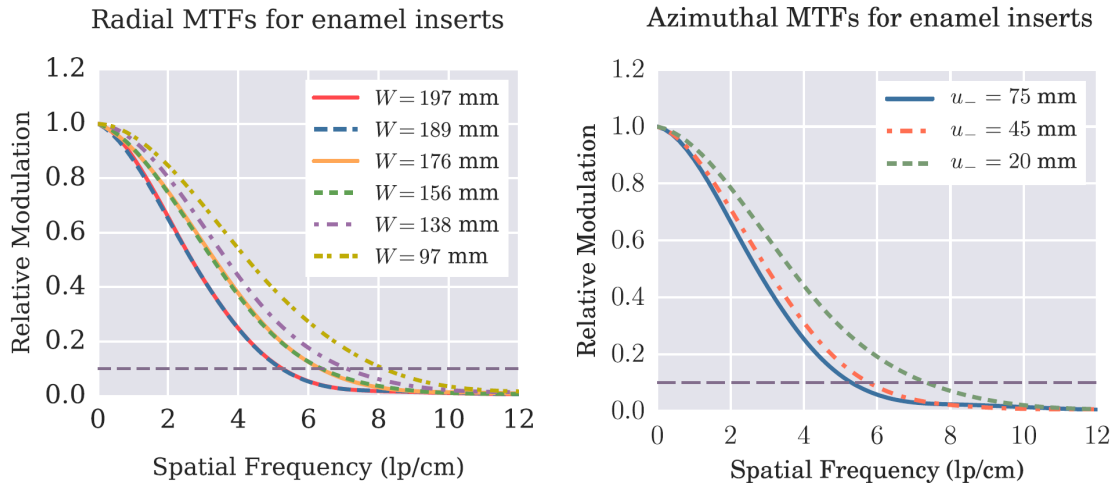

Figure 13: The measured modulation transfer function (spatial resolution) for the LLU/UCSC Phase-II scanner versus the water equivalent path length (WEPL) of the protons through the phantom material. The results were derived from a pCT image (see Fig. 12) reconstructed from experimental data using 0.625 mm pixels and 2.5 mm slices. The two separate graphs represent measurements from edges of cubes embedded in the plastic cylinder that are respectively parallel to a chord ("radial") or parallel to a radius ("azimuthal"). "MTF10%" represents the line spacing for which the instrument would show a 10% modulation of the signal, thus representing the minimum spacing that would be readily visible in an image. Taken from [110].

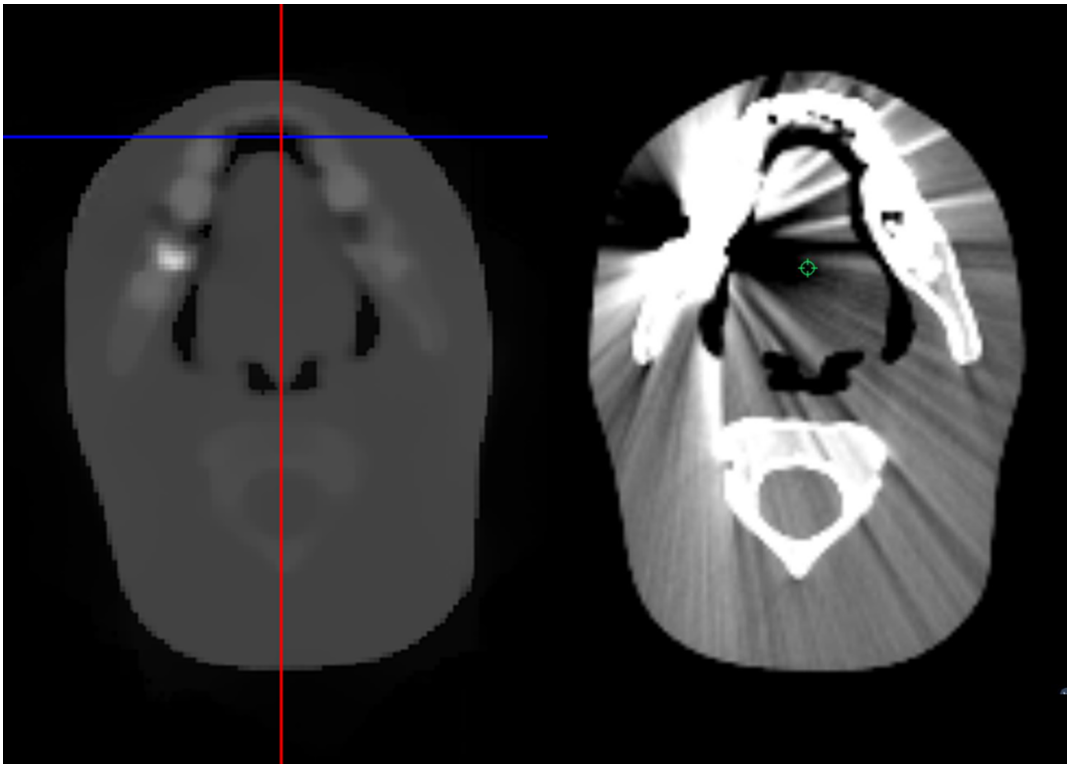

Figure 14: A comparison of pCT (left) and X-ray (right) CT images of a pediatric head phantom in which a gold tooth is visible in the slice displayed. The X-ray image, but not the pCT image, shows severe artifacts due to beam hardening caused by the gold [71].