Block-Iterative and String-Averaging Projection Algorithms in Proton Computed Tomography Image Reconstruction

Scott N. Penfold[1(a)], Reinhard W. Schulte[2], Yair Censor[3], Vladimir Bashkirov[2], Scott McAllister[4], Keith E. Schubert[4] and Anatoly B. Rosenfeld[1]

[1]Centre for Medical Radiation Physics, University of Wollongong, Wollongong, NSW, 2522, Australia

[2]Department of Radiation Medicine, Loma Linda University Medical Center, Loma Linda, CA, 92354, USA

[3]Department of Mathematics, University of Haifa, Haifa, 31905, Israel

[4]Department of Computer Science and Engineering, California State University San Bernardino, San Bernardino, CA, 92407, USA

[(a)]snp75@uow.edu.au

1 Introduction

Proton therapy is an advantageous form of radiotherapy because it allows for the placement of a high dose peak, the Bragg peak, at any desired depth by modulating proton energy. Currently, proton therapy treatment plans are carried out using data from X-ray CT scans, an imaging modality that generates tomographical maps of scaled photon linear attenuation coefficients, commonly known as Hounsfield units. However, to perform the treatment planning, one requires knowledge of the spatial distribution of *electron density* within the patient. In clinical practice Hounsfield units are converted to electron densities through an empirically derived relationship generated from measurements with tissue equivalent materials (Mustafa and Jackson 1983), (Schneider, Pedroni, and Lomax 1996). The end result of this conversion is a difference, typically ranging from several millimeters up to more than 1 cm, between the proton range calculated by the treatment planning software and the true proton range within the patient, depending on the anatomical region treated and the calibration method used (Schneider, Pedroni, and Lomax 1996). Thus, because the Bragg peak depth cannot be accurately predicted, the inherent advantages of proton therapy are partially negated in such an approach.

Proton computed tomography (pCT) is an imaging modality that has been suggested as a means for reducing the uncertainty of Bragg peak location in proton radiation treatments. In pCT, the spatial location of individual protons pre- and post-patient, as well as the energy lost along the path is recorded (Schulte et al. 2004). The spatial measurements are employed in a maximum likelihood proton path formalism that models multiple Coulomb scattering, i.e., multiple small-angle deflections of the proton path due to interaction with the Coulomb field of the nuclei of the medium within the patient (Schulte et al. 2008), maximizing the spatial resolution. The corresponding energy loss

measurements are converted to the integral relative electron density along this predicted path with the Bethe-Bloch equation, which describes the mean energy loss of a proton per unit track length as a function of density of the medium and the proton energy. By reconstructing many such events with an algebraic reconstruction technique (ART) (Gordon, Bender, and Herman 1970) capable of handling these nonlinear paths, three dimensional electron density maps can be generated without the need for any empirical conversion. These maps can then be used in the treatment planning system to accurately predict the proton dose distribution within the patient at treatment time.

It has been demonstrated by previous pCT studies (Li et al. 2006) that superior spatial resolution can be achieved by employing ART for reconstruction in comparison to transform methods, such as filtered back-projection. This is primarily because transform methods must assume the proton traveled along a straight path in the reconstruction volume. Algebraic techniques, however, are much more flexible, not only allowing proton paths to be nonlinear but also permitting the inclusion of *a priori* knowledge about the object to be reconstructed. This flexibility may come at the expense of computation time, however, which could be greater for some iterative techniques than that for transform methods.

If pCT is to be implemented in a clinical environment, fast image reconstruction is required. It has been suggested that the image reconstruction process should take less than 15 minutes for treatment planning images and less than 5 minutes for pre-treatment patient position verification images (Schulte et al. 2004). ART, recognized to be identical with the iterative projection algorithm of Kaczmarz (Kaczmarz 1937), has been implemented in previous pCT studies, displaying promising results (Li et al. 2006). However, ART carries out image updates after each proton history and is therefore inherently se-

quential, meaning that the speed of the reconstruction is dependent on the speed of the computer processing unit. As an example of the infeasibility of using ART for pCT in clinical practice we have recently observed that, using general purpose processing units, three dimensional images made up of a $256 \times 256 \times 48$ voxel reconstruction volume, reconstructed with 10 million proton histories will take approximately 1.5 hours to complete a single cycle, with the optimal image often being reached after 3-4 cycles.

With the development of parallel computing, work has been dedicated to developing iterative projection algorithms that can be executed in parallel over multiple processors to enable fast algebraic reconstructions. This paper compares the performance, in terms of image quality, of a number of parallel compatible block-iterative and string-averaging algebraic reconstruction algorithms with simulated pCT projection data. Quantitative assessment of image quality is based on the *normalized mean absolute distance measure* as described in (Herman 1980), and a qualitative note is made about image appearance. From these results recommendations are made on which image reconstruction algorithms should be used in future studies with pCT.

2 Methods and Materials

2.1 Reconstruction Algorithms

All of the algorithms discussed in this paper belong to the class of *projection methods.* These are iterative algorithms that use projections onto sets while relying on the general principle that when a family of (usually closed and convex) sets is present then projections onto the given individual sets are easier to perform than projections onto other sets (intersections, image sets under some transformation, etc.) that are derived from the given

individual sets. This is the case in pCT reconstruction, where the sets to be projected on in the iterative process are the hyperplanes H_i defined by the i-th row of the $m \times n$ linear system $Ax = b$, namely,

$$H_i = \{x \in \Re^n \mid \langle a^i, x \rangle = b_i\}, \quad \text{for } i = 1, 2, \ldots, m. \tag{1}$$

Here $\Re^n$ is the Euclidean n-dimensional space and a^i is the i-th column vector of A^T (the transpose of A), i.e., its components a^i_j occupy the i-th row of A. The right-hand side vector is $b = (b_i)_{i=1}^m$. In pCT, the a^i_j correspond to the length of intersection of the i-th proton history with the j-th voxel, x is the unknown relative electron density image vector and b_i is the integral relative electron density corresponding to the energy lost by the i-th proton along its path.

2.1.1 The Fully Sequential Algebraic Reconstruction Technique

ART is a sequential projections method for the solution of large and sparse linear systems of equations of the form $Ax = b$. It is obtained also by applying to the hyperplanes, described by each equation of the linear system, the method of successive projections onto convex sets. In the literature, the latter is called POCS (for "projections onto convex sets") or SOP (for "successive orthogonal projections") and was originally published by Bregman (Bregman 1965) and further studied by Gubin, Polyak, and Raik (Gubin, Polyak, and Raik 1967).

Given the control sequence $\{i(k)\}_{k=0}^{\infty}$ where $i(k) = k \bmod m + 1$ and m is the total number of proton histories used in the algorithm, the general scheme for the ART is as follows.

Algorithm 1 *Algebraic Reconstruction Technique (ART)*

Initialization: $x^0 \in \Re^n$ *is arbitrary.*

Iterative Step: Given x^k*, compute the next iterate* x^{k+1} *by*

$$x^{k+1} = x^k + \lambda_k \frac{b_{i(k)} - \langle a^{i(k)}, x^k \rangle}{\| a^{i(k)} \|^2} a^{i(k)}, \tag{2}$$

where $\{\lambda_k\}_{k=0}^{\infty}$ *is a sequence of user-determined relaxation parameters, which need not be fixed in advance, but could change dynamically throughout the iteration cycles.*

ART was used as a standard for comparison in this investigation.

2.1.2 Block-Iterative Algorithms

The block-iterative algebraic reconstruction technique was first published by Eggermont, Herman, and Lent (Eggermont, Herman, and Lent 1981). It can be viewed also as a special case of the block-iterative projections (BIP) method for the convex feasibility problem of Aharoni and Censor (Aharoni and Censor 1989). The BIP method allows the processing of blocks (i.e., groups of hyperplanes H_i) which need not be fixed in advance, but could change dynamically throughout the cycles. The number of blocks, their sizes, and the assignments of the hyperplanes H_i to the blocks may all vary, provided that the weights attached to the hyperplanes form a *fair* sequence, which is defined as follows.

Let $I = \{1, 2, \ldots, m\}$, and let $\{H_i | i \in I\}$ be a finite family of hyperplanes with nonempty intersection $H = \cap_{i \in I} H_i$. Denoting the nonnegative ray of the real line by R_+, introduce a mapping $w : I \rightarrow R_+$, called a weight vector, with the property $\sum_{i \in I} w(i) = 1$. A sequence $\left\{w^k\right\}_{k=0}^{\infty}$ of weight vectors is called *fair* if, for every $i \in I$, there exists infinitely many values of k for which $w^k(i) > 0$. The weight vectors must also obey the condition $\sum_{k=0}^{\infty} w^k(i) = +\infty$ for every $i \in I$, see (Aharoni and Censor 1989).

Given a fair weight vector w, define the convex combination $P_w(x) = \sum_{i \in I} w(i) P_i(x)$,

where $P_i(x)$ is the orthogonal projection of x onto the hyperplane H_i. The general scheme for the BIP technique for linear equations is as follows.

Algorithm 2 *Block-Iterative Projections (BIP)*

Initialization: $x^0 \in \Re^n$ *is arbitrary.*

Iterative Step: Given x^k*, compute the next iterate* x^{k+1} *by*

$$x^{k+1} = x^k + \lambda_k \left(P_{w^k}(x^k) - x^k \right), \tag{3}$$

where $\{w^k\}_{k=0}^{\infty}$ *is a fair sequence of weight vectors and* $\{\lambda_k\}_{k=0}^{\infty}$ *is a sequence of user-determined relaxation parameters.*

The block-iterative algorithmic structure stems from the possibility to have at each iteration k some (but, of course, not all) of the components $w^k(i)$, for some of the indices i, of the weight vector w^k equal to 0. A block-iterative version with fixed blocks is obtained from Algorithm 2 by partitioning the indices of I as $I = I_1 \cup I_2 \cup \cdots \cup I_M$ into M blocks and using weight vectors of the form

$$w^k = \sum_{i \in I_{t(k)}} w^k(i) e^i, \tag{4}$$

where e^q is the q-th standard basis vector (with 1 in its q-th coordinate and zeros elsewhere) and $\{t(k)\}_{k=0}^{\infty}$ is a control sequence over the set $\{1, 2, \ldots, M\}$ of block indices. In this case, and incorporating the expressions for the orthogonal projections P_i onto the hyperplanes H_i into the formula, the iterative step (3) of Algorithm 2 takes the form

$$x^{k+1} = x^k + \lambda_k \left(\sum_{i \in I_{t(k)}} w^k(i) \, \frac{b_i - \langle a^i, x^k \rangle}{\| \, a^i \, \|^2} a^i \right), \tag{5}$$

where $\{t(k)\}_{k=0}^{\infty}$ is a cyclic (or almost cyclic) control sequence on $\{1, 2, \ldots, M\}$. While the generality of the definition of a fair sequence of weight vectors permits variable block sizes and variable assignments of hyperplanes into the blocks that can be used, equal hyperplane weighting and constant block sizes were used in the implementation of BIP in the present investigation.

The block-iterative component averaging (BICAV) algorithm, introduced by Censor, Gordon, and Gordon (Censor, Gordon, and Gordon 2001), is a variant of Algorithm 2 that incorporates *component-related* weighting in the vectors w^k. BICAV also differs in the method of projection onto the individual hyperplanes, making use of generalized *oblique projections*, as opposed to orthogonal projections. For a detailed discussion of the consequences of this on the projection algorithm see (Censor, Gordon, and Gordon 2001). The iterative step in BICAV is defined in (6).

Algorithm 3 *Block-Iterative Component Averaging (BICAV)*

Initialization: $x^0 \in \Re^n$ *is arbitrary.*

Iterative Step: Given x^k*, compute the next iterate* x^{k+1} *by using, for* $j = 1, 2, \ldots, n$*,*

$$x_j^{k+1} = x_j^k + \lambda_k \sum_{i \in I_{t(k)}} \frac{b_i - \left\langle a^i, x^k \right\rangle}{\sum_{\ell=1}^{n} s_\ell^{t(k)} (a_\ell^i)^2} a_j^i, \tag{6}$$

where $\{s_\ell^t\}_{\ell=1}^n$ *is the number of non-zero elements* $a_\ell^t \neq 0$ *in the* ℓ*-th column of the* t*-th block of the matrix* A *given by*

$$A_t = \begin{pmatrix} a^{i_1^t} \\ a^{i_2^t} \\ \vdots \\ a^{i_{m(t)}^t} \end{pmatrix} \tag{7}$$

and $\{\lambda_k\}_{k=0}^{\infty}$ *is a sequence of user-determined relaxation parameters.*

Recently, Censor *et al.* (Censor et al. 2008) derived a component-dependent weighting technique that makes use of orthogonal projections onto hyperplanes rather than the generalized oblique projections employed in the BICAV algorithm. This method, called diagonally relaxed orthogonal projections (DROP), is outlined in Algorithm 4.

Algorithm 4 *Diagonally Relaxed Orthogonal Projections (DROP) (Censor et al. 2008, Algorithm 3.4)*

Initialization: $x^0 \in \Re^n$ *is arbitrary.*

Iterative Step: Given x^k*, compute the next iterate* x^{k+1}

$$x^{k+1} = x^k + \lambda_k U_{t(k)} \sum_{i \in I_{t(k)}} \frac{b_i - \langle a^i, x^k \rangle}{\|a^i\|^2} a^i, \tag{8}$$

where $U_{t(k)} = diag(\min(1, 1/s_\ell^t))$ *with* s_ℓ^t *as defined in Algorithm 3, and* $\{\lambda_k\}_{k=0}^{\infty}$ *is a sequence of user-determined relaxation parameters.*

Both the DROP and BICAV algorithms are computationally more expensive than the BIP method because of the need to calculate the s_ℓ^t's prior to any image updates. However, it is the goal of component-dependent weighting to markedly improve the initial convergence pattern of the algorithm, which may compensate for time spent on extra calculations.

Andersen and Kak (Andersen and Kak 1984) developed a block-iterative technique called the simultaneous algebraic reconstruction technique (SART). They suggested the use of SART with blocks, which the authors called "subsets", made up of image projection rays from a single projection angle and in doing so, found that SART was able to deal well with noisy data. The algorithm was developed in such a way that it was equally applicable to subsets, or blocks, of any composition as it was to subsets comprised of rays from a

single projection angle. This block-iterative form, called ordered subsets simultaneous algebraic reconstruction technique (OS-SART) by Jiang and Wang in (Jiang and Wang 2001), is as follows.

Algorithm 5 *Ordered Subsets Simultaneous Algebraic Reconstruction Technique (OS-SART)*

Initialization: $x^0 \in \Re^n$ *is arbitrary.*

Iterative Step: Given x^k*, compute the next iterate* x^{k+1} *by using, for* $j = 1, 2, \ldots, n$*,*

$$x_j^{k+1} = x_j^k + \lambda_k \left(\frac{1}{\sum_{i \in I_{t(k)}} a_j^i} \right) \sum_{i \in I_{t(k)}} \frac{b_i - \left\langle a^i, x^k \right\rangle}{\sum_{j=1}^n a_j^i} a_j^i, \tag{9}$$

where $\{\lambda_k\}_{k=0}^{\infty}$ *is a sequence of user-determined relaxation parameters.*

2.1.3 String-Averaging Algorithms

In contrast to the block-iterative algorithmic scheme, the string-averaging scheme, proposed by Censor, Elfving, and Herman (Censor, Elfving, and Herman 2001), dictates that, from the current iterate x^k, sequential successive projections be performed along the strings and then the end-points of all strings be combined by a weighted convex combination. In other words, each operation *within* a string must be executed serially, but all string end-points can be calculated in parallel.

Firstly, let us introduce the string notation. For $t = 1, 2, \ldots, M$, let the *string* I_t be an ordered subset of $\{1, 2, \ldots, m(t)\}$ of the form

$$I_t = (i_1^t, i_2^t, \ldots, i_{m(t)}^t), \tag{10}$$

with $m(t)$ the number of elements in I_t. Suppose that there is a set $S \subseteq \Re^n$ such that

there are operators $R_1, R_2, \ldots, R_m$ mapping S into S and an operator R which maps S^m into S.

Algorithm 6 *String-Averaging Algorithmic Scheme*

Initialization: $x^0 \in S$ *is arbitrary.*

Iterative Step: Given x^k*, calculate, for all* $t = 1, 2, \ldots, M$*,*

$$T_t\left(x^k\right) = R_{i^t_{m(t)}} \ldots R_{i^t_2} R_{i^t_1}\left(x^k\right), \tag{11}$$

and then calculate

$$x^{k+1} = R(T_1\left(x^k\right), T_2\left(x^k\right), \ldots, T_M\left(x^k\right)). \tag{12}$$

For every $t = 1, 2, \ldots, M$, this algorithmic scheme applies to x^k successively the operators whose indices belong to the t-th string. This can be done in parallel for all strings and then the operator R maps all end-points onto the next iterate x^{k+1}. For recent references on the application of the string-averaging algorithmic scheme consult (Censor and Segal 2008).

In order to arrive at the iterative algorithmic structure for the string-averaging orthogonal projection algorithm, we must define the following. For $i = 1, 2, \ldots, m$, the operation $R_i(x) = x + \lambda_i(P_i(x) - x)$, where P_i is the orthogonal projection onto the hyperplane H_i and λ_i is an associated (user-determined) relaxation parameter. Then, to combine the strings we use $R(x^1, x^2, \ldots, x^M) = \sum_{t=1}^{M} w_t x^t$, with $w_t > 0$ for all $t = 1, 2, \ldots, M$, and $\sum_{t=1}^{M} w_t = 1$. This leads to the following algorithm.

Algorithm 7 *String-Averaging Projections (SAP)*

Initialization: $x^0 \in \Re^n$ *is arbitrary.*

Iterative Step: Given x^k*, for each* $t = 1, 2, \ldots, M$*, set* $y^0 = x^k$ *and calculate, for* $i = 0, 1, \ldots, m(t) - 1$*,*

$$y^{i+1} = y^i + \lambda_i \frac{b_i - \langle a^i, y^i \rangle}{\| a^i \|^2} a^i, \tag{13}$$

and let $y^t = y^{m(t)}$ *for each* $t = 1, 2, \ldots, M$*. Then, calculate the next iterate by*

$$x^{k+1} = \sum_{t=1}^{M} w_t y^t. \tag{14}$$

Similarly to the block-iterative algorithms, each string end-point was assigned equal weighting in the present investigation.

In a similar vain to the introduction of component-dependent weighting into block-iterative projection algorithms, Gordon and Gordon (Gordon and Gordon 2005) developed the component-averaged row projections (CARP) method for string-averaging algorithms. In the CARP algorithm, s_j^t is the number of strings which contain at least one equation with a nonzero coefficient of x_j. The algorithmic scheme for the CARP algorithm can be given as follows.

Algorithm 8 *Component-Averaged Row Projections (CARP)*

Initialization: $x^0 \in \Re^n$ *is arbitrary.*

Iterative Step: Given x^k*, for each* $t = 1, 2, \ldots, M$*, set* $y^0 = x^k$ *and calculate, for* $i = 0, 1, \ldots, m(t) - 1$*,*

$$y^{i+1} = y^i + \lambda_i \frac{b_i - \langle a^i, y^i \rangle}{\| a^i \|^2} a^i, \tag{15}$$

and let $y^t = y^{m(t)}$ *for each* $t = 1, 2, \ldots, M$*. Then, calculate the next iterate by*

$$x_j^{k+1} = \frac{1}{s_j^t} \sum_{t=1}^{M} y_j^t. \tag{16}$$

2.2 Proton CT Simulations

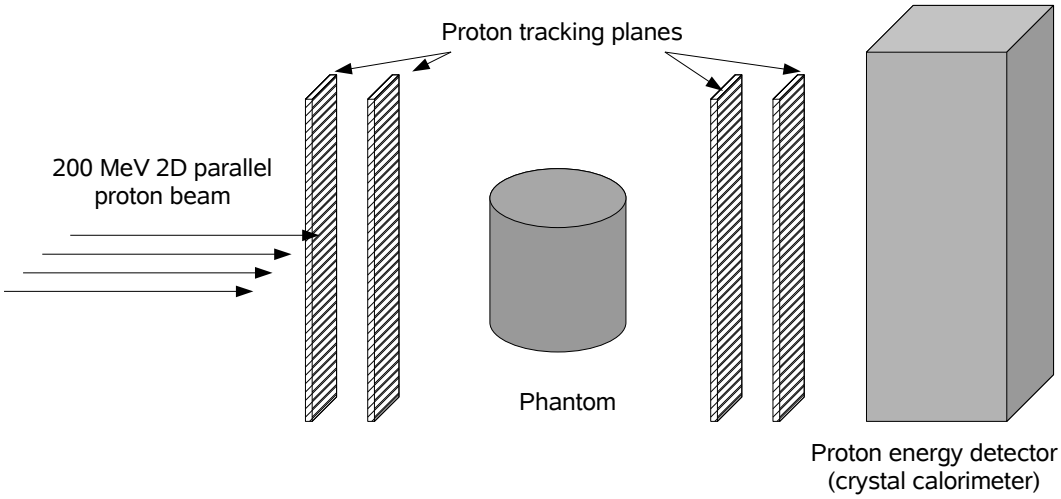


Figure 1: Schematic of the proton CT system modeled by the GEANT4 simulation.

The conceptual pCT system considered in the current work produces maps of relative electron density through individual proton spatial and energy loss measurements (Schulte et al. 2004), (Schulte et al. 2005). Single protons are tracked pre- and post-patient with 2D sensitive silicon strip detectors (SSDs), providing information about proton position and direction at the boundaries of the image space. This allows the effects of multiple Coulomb scattering within the object to be accounted for in a most likely path (MLP) estimation (Schulte et al. 2008). The advantage of this in terms of spatial resolution of the reconstructed image has been shown in a previous study (Li et al. 2006).

As well as tracking the position of individual protons, the energy lost by each proton after traversal of the image space is also recorded. Substituting these measurements into (17), the integral relative electron density along the estimated proton path can be calculated by

$$\int_L \eta_e d\mathbf{r} = \int_{E_{out}}^{E_{in}} \frac{dE}{S(I(\mathbf{r}), E(\mathbf{r}))}. \tag{17}$$

In (17), E_{in} and E_{out} are the measured entry and exit proton energies at the image space

boundaries respectively, η_e is the relative electron density at spatial location $\mathbf{r}$, and L is the estimated proton path through the image space. The stopping power $S(I(\mathbf{r}), E(\mathbf{r}))$ is given by the following Bethe-Bloch equation

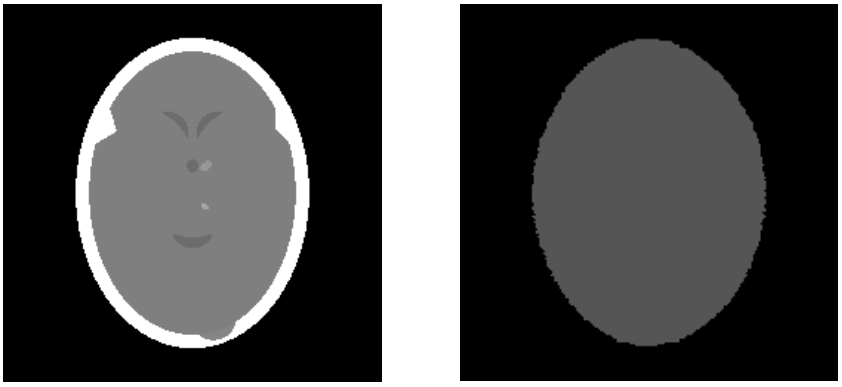
$$S(I(\mathbf{r}), E(\mathbf{r})) = K \frac{1}{\beta^2(E(\mathbf{r}))} \left[ln \left(\frac{2m_e c^2}{I(\mathbf{r})} \frac{\beta^2(E(\mathbf{r}))}{1-\beta^2(E(\mathbf{r}))} \right) - \beta^2(E(\mathbf{r})) \right]. \quad (18)$$

Here, $K = 0.170$ MeV/cm contains various physical constants, m_e is the mass of the electron, β is the velocity of the proton relative to the speed of light c, $E(\mathbf{r})$ is the proton kinetic energy at $\mathbf{r}$, and $I(\mathbf{r})$ is the mean excitation energy of the material, which can also vary with $\mathbf{r}$. In pCT reconstructions, we use a constant mean excitation energy ($I(\mathbf{r}) = I_{water}$) when converting energy loss to integral relative electron density. Therefore, we are reconstructing an object consisting of water with varying density, that results in the same energy loss as with the imaged object. This conforms to the current proton treatment planning practice (see, for example, (Hong et al. 1996)).

Figure 1 illustrates the geometry of a GEANT4 (Agostinelli et al. 2003) application created to model an ideal pCT system. The proton beam consisted of a 200 MeV monoenergetic 2D parallel geometry. To record proton position and direction at the entry and exit planes of the reconstruction volume, two upstream and two downstream 2D sensitive silicon tracking planes 30cm × 30cm × 0.04 cm in size were located at -30 cm, -25 cm, 25 cm and 30 cm along the axis of the beam, relative to the center of the phantom. All tracking sensitive volumes were allocated a pitch of 0.2 mm. To accurately record proton exit energy a 32 cm × 32 cm × 12 cm cesium iodide (CsI) crystal calorimeter was placed downstream of the tracking modules. The face of the crystal was positioned 5 cm behind the second exiting tracking module.

A cylindrical phantom with an elliptical cross-section, based on the head phantom

design of Herman (Herman 1980), was located at the center of the imaging system. The major axis of the phantom cross-section was set to 17.25 cm and the minor axis to 13 cm. A cross-section of the phantom can be seen in Figure 2(a). The different regions have the same chemical composition (water) but varying physical density. Because of the use of I_{water} in (18), simulating with a water phantom means that the reconstructed values can be directly compared to the true phantom values, simplifying the analysis of image quality. Image quality would not be changed if a varying chemical composition was used, however analysis of image reconstruction accuracy would not be as straightforward. In this case, a conversion of phantom electron densities to water equivalent electron densities, which are calculable from the known chemical composition and physical density, would be required.



(a) (b)

Figure 2: (a) Cross-section of the phantom used in the GEANT4 simulation. (b) Object boundary definition by the direct summation method.

A total of 180 proton beam projections were carried out at two degree intervals with the first 20,000 protons that were found to traverse the geometry and deposit energy in the CsI calorimeter being recorded in each projection angle. Protons with an exit angle or exit energy falling more than three standard deviations from the respective means were excluded from the simulation, the motivation for which is described elsewhere (Schulte et al. 2008). The low energy electro-magnetic and low energy hadronic physics processes

were used as the basis for the interactions to be considered in the simulation (Urban et al. 2007).

2.3 Proton CT Reconstruction

The algorithmic structure of the iterative steps to be investigated in the various algebraic methods of reconstruction are but one ingredient of the overall pCT reconstruction process. The overall procedure can be broken into the sub-routines listed below.

1. Load the measured proton data (energy loss, entry and exit coordinates and directions).

2. Bin the individual proton histories based on their exit location for each projection angle.

3. Analyze exit angle and exit energy of protons within each bin and exclude protons in which exit angle or energy is beyond three standard deviations from the mean (Schulte et al. 2005), (Li et al. 2006), and (Schulte et al. 2008).

4. Determine the object boundary location. In this work, the object boundary location was calculated by performing an initial run through of the data with the direct summation method described by Herman and Rowland (Herman and Rowland 1973), and by simplifying the proton path to a straight line. This initial image is used for the object boundary only, as the actual pixel values calculated with this method are quite erroneous. See Figure 2(b) for an example of how well the object boundary is defined with the direct summation method.

5. Calculate the path of the accepted proton histories. If a straight line between proton entry and exit location was found to intersect the object, the MLP formalism

(Schulte et al. 2008) was employed, if not, a straight line was used. By modeling multiple Coulomb scattering within the object boundary, the MLP formalism predicts the proton path of maximum likelihood given the entry and exit tracking measurements.

6. Calculate integrated relative electron density along each proton path and apply the iterative reconstruction algorithm.

Although the relaxation parameters λ_k in the iterative algorithms mentioned above may vary dynamically with cycle number, in this study we considered only the case of constant λ. The data was subdivided into 180, 60, and 12 subsets of equal sizes, arranged such that each subset contained an equal number of proton histories from each projection angle. Other block structures and assignments may be useful but this needs further study. The optimal relaxation parameter, which was defined to be the value that returned the best image quality within ten complete cycles, was found for each algorithm and subset size. Note that an iteration refers to the update of the image while a cycle is a complete run through m proton histories.

All our computations with the reconstruction algorithms were done on a single processor. Further clock-time gains could thus be achieved for those algorithms that enjoy a greater degree of parallelism in their structure. We analyzed images up until the completion of the tenth cycle, as any more iterations than this will likely result in an image reconstruction time too large for clinical practicality.

2.4 Evaluation of Image Quality

Image quality is not a well-defined concept. It depends on the purpose for which the image is generated. In the case of pCT images, visual appearance is important so that

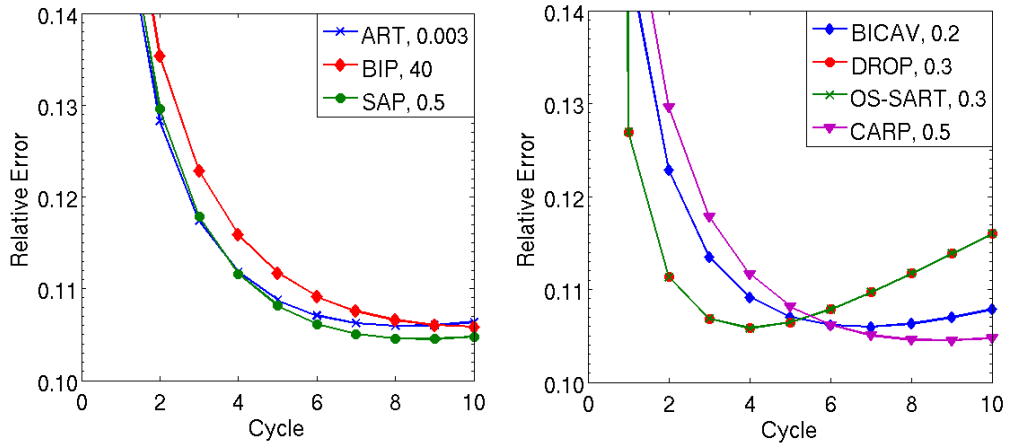
structures can easily be identified in the treatment planning process. Also, the actual values of the digitized picture are of equal importance as these values are used to calculate dose deposition by the treatment planning software. Since it is difficult to quantitatively evaluate image appearance, we based our analysis of image quality on how close the values of the reconstructed images are to the test phantom. For this purpose we have employed the normalized mean absolute distance measure, which is defined (see, e.g., Herman (Herman 1980)) as

$$\epsilon_k = \frac{1}{\sum_{j\in S} |\tilde{x}_j|} \sum_{j\in S} \left| x_j^k - \tilde{x}_j \right|, \tag{19}$$

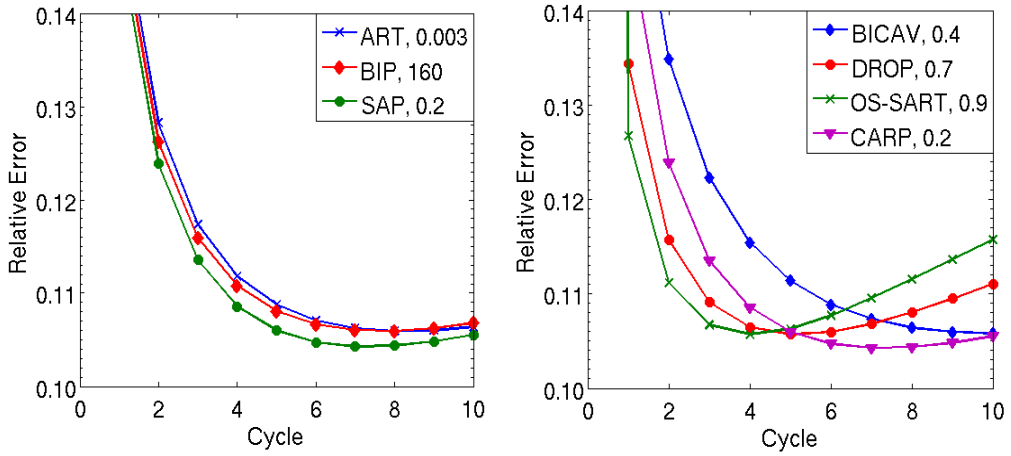
where $\tilde{x}_j$ is the relative electron density of the phantom in pixel j, x_j^k is the j-th pixel value of the reconstructed image after the k-th cycle, and S is the set of indexes j of pixels which are in the region of interest. We selected the region of interest to be the set of those pixels that were part of the phantom object (see Figure 2(a)). Therefore, ϵ_k, hereby referred to as the relative error, is a measure of how close the relative electron density values of the reconstructed images are to the true values of the test phantom.

3 Results

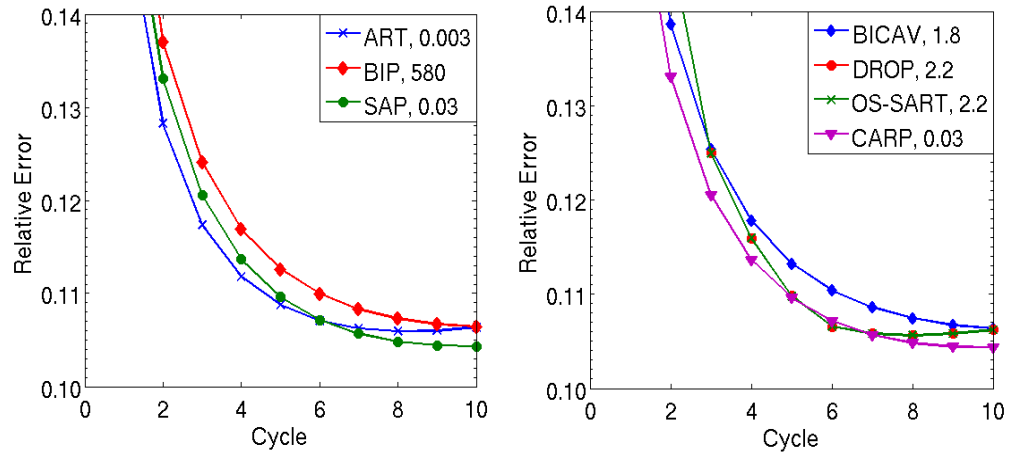
In figure 3 the relative error with optimal relaxation parameter is plotted as a function of cycle number for each algorithm with the data partitioned into 180, 60, and 12 subsets of equal sizes (with the exception of ART which is fully sequential). The left-hand column contains ART and the component-*independent* block-iterative and string-averaging algorithms (Algorithms 1, 2, and 7), while the right-hand column contains the component-*dependent* algorithms (Algorithms 3, 4, 5, and 8). These results are also summarized in



(a)



(b)



(c)

Figure 3: Relative error as a function of cycle number for all tested algorithms. The left-hand column contains ART and the component-independent algorithms BIP and SAP, while the right-hand column contains the component-dependent algorithms BICAV, DROP, OS-SART and CARP. The data was divided into (a) 180, (b) 60, and (c) 12 subsets. In each case ART is plotted for comparative purposes and was not divided into the aforementioned subsets. The number next to each algorithm in the legends corresponds to the relaxation parameter that resulted in the smallest relative error within ten cycles.

Table 1, which contains the minimum relative error and cycle number at which this was reached with the various reconstruction algorithms using an optimal relaxation parameter.

It can be seen that, for all subset sizes, the component-independent methods (ART, BIP, and SAP) are very similar in their convergence pattern with an asymptotic approach to a minimum relative error between 7 and 10 cycles. Of these, SAP achieves the smallest relative error in all subset sizes, however, the minimum relative error of all algorithms are within $\pm 2\%$ of each other.

For the component-dependent algorithms, DROP and OS-SART have an advantage in terms of initial speed of convergence, in particular for a large number of subsets (60 and 180), however, there is a rapid increase in error after achieving the minimum relative error. Again, the minimum relative errors are relatively close to each other (within 2%), and are also within 2% of the errors achieved with the component-independent weighted techniques.

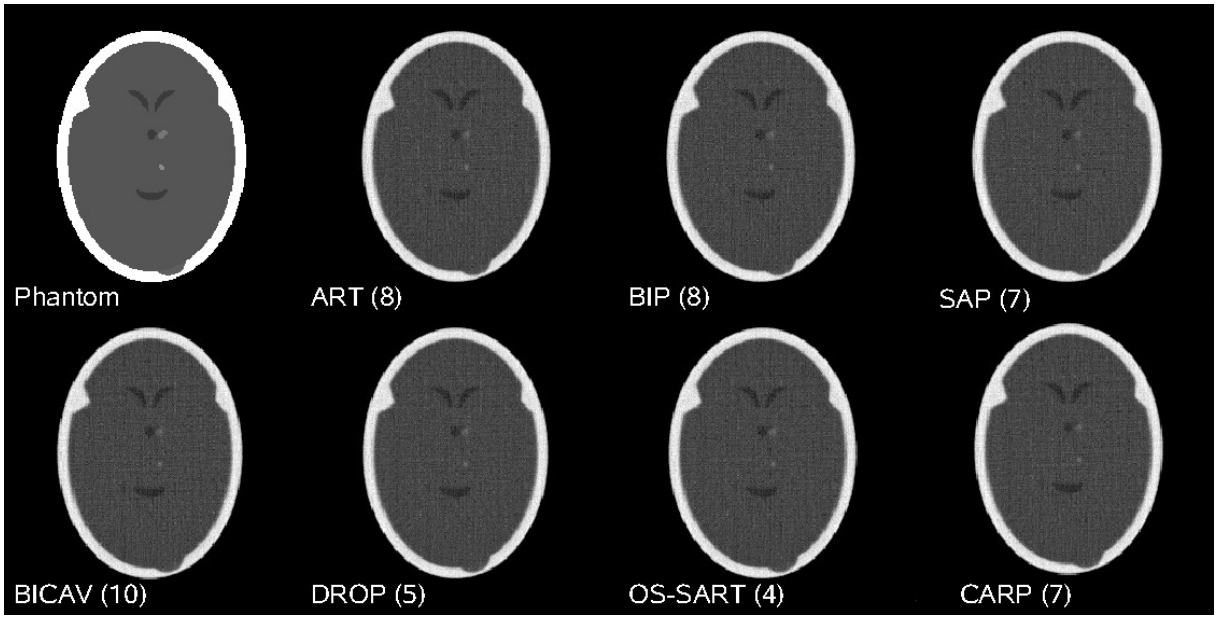


Figure 4: Reconstructed images with optimal relaxation parameter and 60 subsets (with the exception of the fully sequential ART) corresponding to the cycle at which the minimum relative error was found. The cycle number for each algorithm is shown in parentheses.

It was also observed that extreme over-relaxation was required for the BIP algorithm to achieve a competitive initial convergence rate. This is due to the weighting factor in Algorithm 2 being far less than 1 when equal weighting is assigned to each proton history. It is also apparent that with a smaller number of subsets (e.g., 12), the initial convergence rate of all algorithms is reduced in comparison to that when a greater number of subsets is used.

The images corresponding to the cycle at which the minimum relative error was produced by each reconstruction algorithm with 60 subsets and optimal relaxation parameter are shown in Figure 4. It can be seen that, qualitatively, the images are similar in appearance, which is to be expected considering the relatively small difference in minimum relative error achieved by the different algorithms.

The effect of iterating beyond the cycle at which the minimum relative error is achieved can be seen in Figure 5. Here, the image corresponding to the cycle of the minimum relative error is compared to that produced after 10 cycles for the DROP and OS-SART algorithms. The increased relative error is reflected in the noise level of the image.

4 Discussion

The goal of pCT image reconstruction is to produce accurate electron density maps in the shortest possible time. Parallel compatible projection algorithms that can be simultaneously executed over multiple processing units provide a means of *computationally* accelerating the image reconstruction process. Acceleration of these algorithms can also be achieved with the use of a component-dependent weighting scheme, several of which were investigated in this work.

With the use of GEANT4 simulated pCT data, it was found that these block-iterative

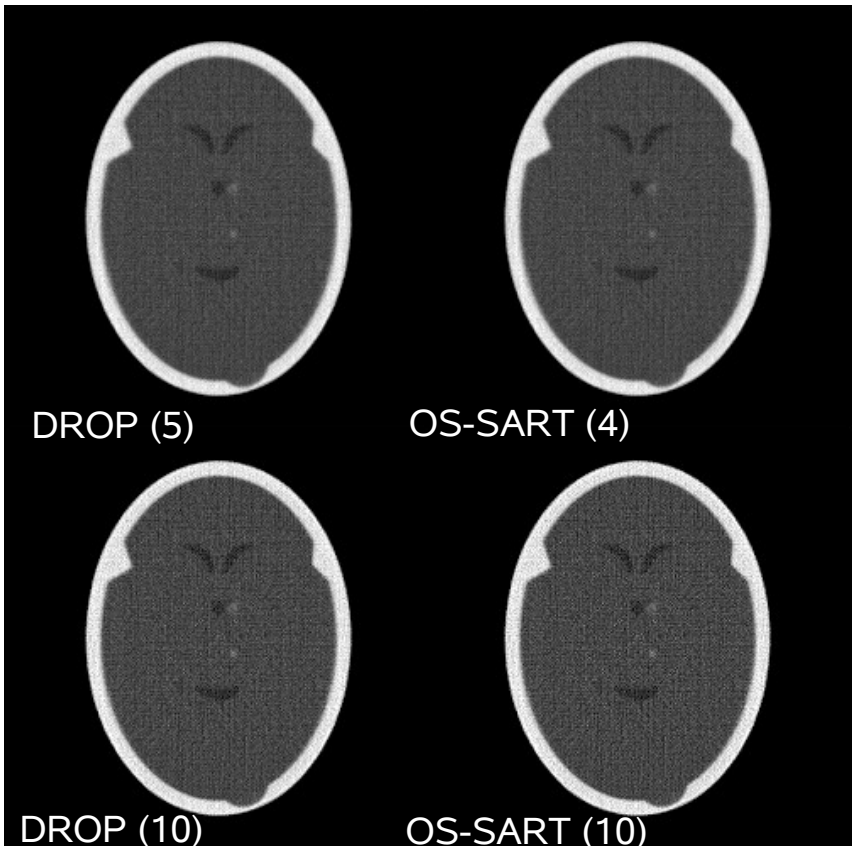


Figure 5: Reconstructed images with optimal relaxation parameter and 60 subsets for DROP and OS-SART at the cycle at which the minimum relative error was found and also after 10 cycles. Iterating beyond the optimal stopping point amplifies noise in the pCT data.

and string-averaging algorithms perform as well as the currently used ART algorithm from the point of view of image quality. This can be appreciated from the figures presented above. The results in Table 1 also show that some methods arrived at the same minimal relative error value in fewer computational cycles then ART (e.g., DROP and OS-SART). Although our work is not yet presenting any statistical results of experiments with ensembles of test images (phantoms), we believe that when we come to parallel processing and to exploiting hardware advantages, the inherent parallelism embodied already in the mathematical formulation of the block-iterative and string-averaging algorithms might become more useful than the (fully-sequential) ART algorithm.

It should be noted though that for sparse problems, the sequential row-action ART

method can also be parallelized by simultaneously projecting the current iterate onto a set of mutually orthogonal hyperplanes (obtained by considering equations whose sets of nonzero components are pairwise disjoint). For the case of image reconstruction from projections, such sets of equations can be obtained by grouping rays that are sufficiently far apart so as to pass through disjoint sets of pixels. The relative efficacy of this as compared to the parallelism possible for a block-iterative method depends on the number of equations that can be grouped in the above-mentioned fashion and the number of available processors.

The results suggest that the string-averaging algorithms are able to produce images of smaller relative error than block-iterative methods for the reconstruction problem in the current study. The results also show that the choice of subset size is important to obtain better image quality in the smallest number of cycles. We have demonstrated that when partitioning the data for the string-averaging algorithm, one should choose a string size that is not so large that the number of sequential operations is so numerous that noise becomes an issue, but not so small that the initial convergence suffers. This choice of course depends on the number of histories that are to be used in the reconstruction process.

It can also be seen from our results that component-dependent weighting has little effect on the block-iterative BICAV and string-averaging CARP algorithms. Indeed, SAP and CARP display identical results in terms of relative error. This is because the method of weighting suggested in (Gordon and Gordon 2005) and implemented here is based on the number of strings in which the particular pixel was intersected by a proton history. Since there are a huge number of proton histories in each string, all corresponding to an equation in the linear system $Ax = b$ of the imaging problem (far more equations in

pCT than in X-ray CT), nearly all pixels are intersected in each string. This reduces the weighting systems of SAP and CARP to be approximately identical.

From the results, it appears that noise and error are not monotonically decreasing as a function of cycle number. This is most noticeable in the DROP and OS-SART results, but we believe similar observations would be made for the other algorithms if further iterations were performed. A similar "semi-convergence" characteristic has been observed in iterative X-ray CT reconstructions (see, for example, (Censor, Gordon, and Gordon 2001)). This effect is probably due to amplification of noise in the pCT projection data with an increasing number of cycles. In iterative X-ray CT cases, it has been found that regularization with priors and implementation of stopping rules can result in more stable systems (Elfving and Nikazad 2007). Similar applications to iterative pCT reconstruction may have similar effects, but was not investigated in the present work.

A draw-back of all the projection algorithms discussed in this study is the need to find an optimal relaxation parameter, λ. In this study it was possible to determine the "best" λ because the true density distribution of the phantom was known, but in a realistic scenario, this will not be the case. We are currently investigating the implementation of the Dos Santos scheme (Dos Santos 1987) into block-iterative and string-averaging algorithms. Here, the optimal λ is calculated at each iterative step, and in doing so also accelerates the initial convergence to minimum relative error.

Furthermore, we believe that some of the parallel compatible algorithms discussed here can be modified to further improve the handling of noisy pCT data. The primary factors that contribute to the noise in pCT data are:

1. The statistical nature of proton energy loss when traversing an object and noise associated with the detector system itself, leading to inaccurate values of the elements

of the vector b.

2. The statistical variations of the paths of the protons, leading to inaccurate values of the elements of the matrix A.

These factors contribute to spatial blurring and image noise in the reconstructed data in a complex way and differently for the different algorithms as we have shown. We are investigating incorporation of the method of projections onto *hyperslabs* (Herman 1975), as opposed to hyperplanes, for string-averaging and block-iterative projection algorithms. This method provides a means for modeling the uncertainties in the b vector but not with those in the A matrix. The latter may be approached with more accurate proton path estimation algorithms.

The potential of clock-time savings of the parallel compatible algorithms tested in this study was not demonstrated here. However, execution of the algorithms, found to provide the best image quality in this study, on general purpose graphical processing units (GPGPU) is an active area of research that we currently work in.

5 Conclusion

Image reconstruction in proton CT aims at efficient computation and provision of accurate electron density maps. The block-iterative and string-averaging projection algorithms investigated in this paper provide an algorithmic platform for achieving both goals. The parallel compatible nature means that execution on a computer cluster or parallel GPGPUs would speed up the image reconstruction process considerably, producing images in clinically practical amounts of time. Also, the combination of simultaneous and sequential operations should lead to initial convergence rates that are superior to those of fully

simultaneous algorithms and to better handling of noisy data than that of fully sequential methods. The results of this paper suggest that the string-averaging methods can achieve more accurate electron density maps in comparison to the block-iterative algorithms. Further, component-dependent weighting was found to have minimal effect in the string-averaging approach, meaning that, in our application, there is little advantage in using the computationally more expensive CARP algorithm in comparison to SAP. The block-iterative OS-SART and DROP algorithms displayed the most rapid initial convergence.This was, however, at the expense of increased image noise with increasing number of iterations.

Acknowldgements

The authors thank Gabor Herman and an anonymous reviewer for helpful comments on an earlier version of the paper. This work was supported in part by Award Number R01HL070472 from the National Heart, Lung, And Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, And Blood Institute or the National Institutes of Health.

References

Agostinelli S. *et al.* (2003). "GEANT4-a simulation toolkit." *Nuclear Instruments and Methods in Physics Research A* 506:250–303.

Aharoni R. and Y. Censor. (1989). "Block-iterative projection methods for parallel computation of solutions to convex feasibility problems." *Linear Algebra and its Applications* 120:165–175.

Andersen A.H. and A.C. Kak. (1984). "Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm." *Ultrasonic Imaging* 6:81–94.

Bregman L.M. (1965). "The method of successive projections for finding a common point of convex sets." *Soviet Mathematics Doklady* 6:688–692.

Censor Y., T. Elfving, and G.T. Herman. (2001). "Averaging strings of sequential iterations for convex feasibility problems." In: *Inherently Parallel Algorithms in Feasibility*

and Optimization and Their Applications, D. Butnariu, Y. Censor and S. Reich (Eds.), Elsevier Science Publications, Amsterdam, The Netherlands, 101–114.

Censor Y., T. Elfving, G.T. Herman, and T. Nikazad. (2008). "On diagonally-relaxed orthogonal projection methods." *SIAM Journal on Scientific Computing* 30:473–504.

Censor Y., D. Gordon, and R. Gordon. (2001). "BICAV: A block-iterative parallel algorithm for sparse systems with pixel-related weighting." *IEEE Transactions on Medical Imaging* 20:1050–1059.

Censor Y. and A. Segal. (2008). "Iterative projection methods in biomedical inverse problems." In: *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT)*, Y. Censor, M. Jiang and A.K. Louis (Eds.), Edizioni della Normale, Pisa, Italy, 65–96.

Dos Santos L.T. (1987). "A parallel subgradients projection method for the convex feasibility problem." *Journal of Computational and Applied Mathematics* 18:307–320.

Eggermont P.P.B., G.T. Herman, and A. Lent. (1981). "Iterative algorithms for large partitioned systems, with applications to image reconstruction." *Linear Algebra and its Applications* 40:37–67.

Elfving T. and T. Nikazad. (2007). "Stopping rules for Landweber-type iteration." *Inverse Problems* 23:1417–1432.

Gordon R., R. Bender, and G.T. Herman. (1970). "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography." *Journal of Theoretical Biology* 29:471–481.

Gordon D. and R. Gordon. (2005). "Component-averaged row projections: A robust block-parallel scheme for sparse linear systems." *SIAM Journal on Scientific Computing* 27:1092–1117.

Gubin L., B. Polyak, and E. Raik. (1967). "The method of projections for finding the common point of convex sets." *USSR Computational Mathematics and Mathematical Physics* 7:1–24.

Herman G.T. (1975). "A relaxation method for reconstructing objects from noisy X-rays." *Mathematical Programming* 8:1–19.

Herman G.T. *Image Reconstruction From Projections: The Fundamentals of Computerized Tomography*. New York, NY: Academic Press, 1980.

Herman G.T. and S.W. Rowland. (1973). "Three methods for reconstructing objects from x-rays: a comparative study." *Computer Graphics and Image Processing* 2:151–178.

Hong L., M. Goitein, M. Bucciolini, R. Comiskey, B. Gottschalk, S. Rosenthal, C. Seragoy, and M. Urie. (1996). "A pencil beam algorithm for proton dose calculations." *Physics in Medicine and Biology* 41:1305–1330.

Jiang M. and G. Wang. (2001). "Development of iterative algorithms for image reconstruction." *Journal of X-ray Science Technology* 10:77–86.

Kaczmarz S. (1937). "Angenäherte Auflösung von Systemen linearer Gleichungen." *Bulletin de l'Académie Polonaise des Sciences et Lettres,* A35:355–357.

Li T., Z. Liang, J.V. Singanallur, T.J. Satogata, D.C. Williams, and R.W. Schulte. (2006). "Reconstruction for proton computed tomography by tracing proton trajectories: A Monte Carlo study." *Medical Physics* 33:699–706.

Mustafa A.A. and D.F. Jackson. (1983). "The relation between x-ray CT numbers and charged particle stopping powers and its significance for radiotherapy treatment planning." *Physics in Medicine and Biology* 28:169–176.

Schneider U., E. Pedroni, and A. Lomax. (1996). "The calibration of CT Hounsfield units for radiotherapy treatment planning." *Physics in Meidicine and Biology* 41:111–124.

Schulte R., V. Bashkirov, T. Li, Z. Liang, K. Mueller, J. Heimann, L.R. Johnson, B. Keeney, H.F.-W. Sadrozinski, A. Seiden, D.C. Williams, L. Zhang, Z. Li, S. Peggs, T. Satogata, and C. Woody. (2004). "Conceptual design of a proton computed tomography system for applications in proton radiation therapy." *IEEE Transactions on Nuclear Science* 51:866–872.

Schulte R.W., V. Bashkirov, M.C. Loss Klock, T. Li, A.J. Wroe, I. Evseev, D.C. Williams, and T. Satogata. (2005). "Density resolution of proton computed tomography." *Medical Physics* 32:1035–1046.

Schulte R.W., S.N. Penfold, J.T. Tafas, and K.E. Schubert. (2008). "A maximum likelihood proton path formalism for application in proton computed tomography." *Medical Physics* 35:4849–4856.

Urban L., M. Maire, D.H. Wright, and V. Ivanchenko. "GEANT4 v9.1 Physics Reference Manual: Hadron and Ion Ionisation." [Online] Available: http://geant4.web.cern.ch/geant4/UserDocumentation/UsersGuides/PhysicsReferenceManual/html/PhysicsReferenceManual.html [14 December 2007].

Table 1: Minimum Relative Error and Cycle Number

Algorithm	**Subsets**	**Min. Rel. Error.**	**Cycle**
ART	NA	0.1059	8
BIP	180	0.1058	10
	60	0.1059	8
	12	0.1064	10
SAP	180	0.1045	9
	60	0.1043	7
	12	0.1043	10
BICAV	180	0.1060	7
	60	0.1058	10
	12	0.1064	10
DROP	180	0.1058	4
	60	0.1057	5
	12	0.1056	8
OS-SART	180	0.1058	4
	60	0.1057	4
	12	0.1056	8
CARP	180	0.1045	9
	60	0.1043	7
	12	0.1043	10