

Scatter Corrections in X-Ray Computed Tomography: A Physics-Based Analysis

Zachary H. Levine¹, Timothy J. Blattner¹, Adele P. Peskin², and Adam L. Pinter¹

¹National Institute of Standards and Technology,
Gaithersburg, MD 20899 USA

²National Institute of Standards and Technology,
Boulder, CO 80305 USA

zlevine@nist.gov
timothy.blattner@nist.gov
adele.peskin@nist.gov
adam.pinter@nist.gov

Fundamental limits for the calculation of scattering corrections within X-ray computed tomography (CT) are found within the independent atom approximation from an analysis of the cross sections, CT geometry, and the Nyquist sampling theorem, suggesting large reductions in computational time compared to existing methods. By modifying the scatter by less than 1 %, it is possible to treat some of the elastic scattering in the forward direction as inelastic to achieve a smoother elastic scattering distribution. We present an analysis showing that the number of samples required for the smoother distribution can be greatly reduced. We show that fixed forced detection can be used with many fewer points for inelastic scattering, but that for pure elastic scattering, a standard Monte Carlo calculation is preferred. We use smoothing for both elastic and inelastic scattering because the intrinsic angular resolution is much poorer than can be achieved for projective tomography. Representative numerical examples are given.

Key words: computed tomography; elastic cross section; inelastic cross section; x-ray scattering

Accepted: April 19, 2019

Published: May 22, 2019

<https://doi.org/10.6028/jres.124.013>

1. Introduction

As originally conceived, X-ray computed tomography (CT) finds a material-dependent parameter from observations based on line integrals characterizing the absorption of X-rays in a sample, known in the field as projections [1]. One of the difficulties in CT arises from the fact that physical X-rays do not necessarily travel in straight lines. Instead, they can scatter into a detector. The problem has only gained in practical importance as large array detectors have been widely deployed commercially. Scatter estimation approaches were reviewed a few years ago [2] along with a review of the physics of iterative reconstruction in CT [3]. A sketch of typical CT data acquisition is shown in Fig. 1a.

The importance of scatter corrections was understood relatively early in the development of X-ray tomography [4, 5]. The importance of these corrections in diagnostic systems with flat-panel display was studied with phantoms simulating humans two decades ago [6], and the topic remains of current interest [7]. Solutions in both hardware and software are sought [8]. A software framework for studying low-dose CT was presented recently [9]. Scatter corrections in the context of multi-energy X-ray detectors have been

studied recently as well [10], and variance reduction techniques for scatter corrections have been presented [11].

Recently, a deep neural network approach has been used, motivated by scatter corrections being too slow for practical work [12, 13]. Such an approach suggests that the problem may be too difficult for analysis by conventional means unless a significant computational improvement is achieved.

Despite many years of serious attention to scatter corrections in X-ray tomography, we are not aware of a first principles analysis that attempts to assess the computational requirements from a fundamental point of view. The principal ingredients are the differential X-ray scattering cross sections of the elements and the Nyquist sampling theorem. For simplicity, we limit our analysis to axial tomography. As is common in X-ray tomography, we assume that the sample is viewed with evenly spaced angles through 180° about a single axis. More complicated patterns such as helical tomography, which is common in medicine, or advanced geometries, e.g., to satisfy Tuy's sufficiency condition [14], are excluded from the present analysis.

It is possible to treat the problem of diffusion tomography, as occurs, e.g., in the optical domain [15]. However, the fact that every point in the domain can send a signal to any detector pixel leads to algorithms that scale substantially worse than projective tomography [16]. Hence, there is an incentive to apply a scatter correction so that algorithms scale as in projective tomography.

2. Theory and Analysis

2.1 X-Ray Tomography and Projections

In the *vast majority* of cases, the X-ray tomography problem is formulated in terms of projections, *i.e.*, line integrals through a material [1]. As originally formulated, X-rays were taken either to be transmitted in a straight line to the detector or attenuated. However, the growing use of large pixelated detectors increased the possibility that an X-ray scattered from one ray would be detected by a pixel other than the one to which it was originally headed. In medical tomography, antiscatter grids are common. These are less common in industrial or microCT where cost and space constraints restrict their use. There are a number of more modern versions of X-ray tomography, including diffractive and phase-sensitive methods as well as tensor-based methods. Such advanced topics are out of the scope of this paper.

A sketch illustrating the fan beam geometry in two dimensions is shown in Fig. 1b. The reconstruction region is a circle of radius r , the center of which is a distance L from the source and R from the center of a plane detector. The geometric projection of the circle onto the detector is shown in Fig. 1b. Noting that the rays tangent to the circle determine the projected width W of the reconstruction region on the detector, it is simple to show that

$$W = 2r \frac{L+R}{\sqrt{L^2 - r^2}} \approx 2r \frac{L+R}{L}. \quad (1)$$

where the approximation holds only if $r \ll L$. The approximate result is often given as the geometric magnification $M = (L+R)/L$.

2.1.1 Single vs. Multiple Scattering

In this work, we will assume that the independent atom approximation is valid [17], *i.e.*, that the scattering cross sections are given by the sum of the atomic constituents [18]. While X-ray fine structure depends on the arrangement of the atoms in space [19], if a broad spectrum of X-rays is used to illuminate the target, e.g., from an X-ray tube, the fine structure tends to average out.

We are interested in determining the finest angular resolution that is relevant to the X-ray scattering problem. This depends on the X-ray photon energy, the element from which X-rays are being scattered, and

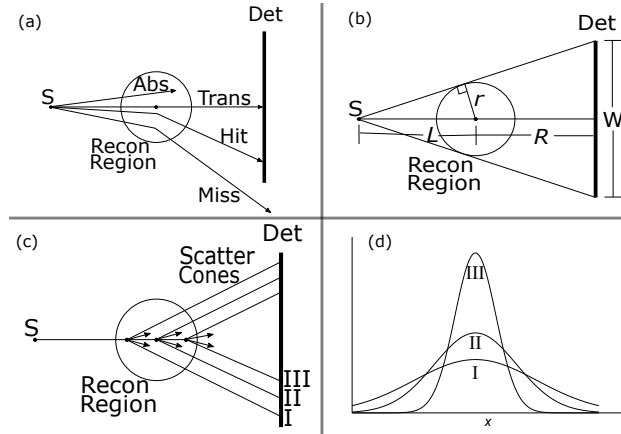


Fig. 1. Sketches of (a) interactions in CT, showing absorption, transmission, and scattering that hits the array detector and scattering that misses the array detector; (b) the system, including the X-ray source S which sends X-rays through the cylindrical reconstruction region with radius r , centered a distance L from the source, to the detector of minimum width W , a further distance R from the center of the reconstruction region; (c) the ways in which linear spatial resolution requirements are modified in the presence of a spatial offset characteristic of tomography, where each of the three scattering cones I, II, and III have the same angular distribution, but the spatial extent on the detector depends on the location of the scattering point within the reconstruction region; and (d) the ways in which the distribution on the detector is modified for a hypothetical distribution depending on the position of the scatterer. The position on the detector is denoted by x .

the geometry. When multiple scattering occurs, the angular distribution becomes less strongly peaked [20]. Intuitively, the multiple scattering distribution is a convolution of a function with itself, *i.e.* the distribution of a sum of independent and identically distributed random variables. Since the variances of a sum of independent random variables are additive, the distribution of the sum is more spread out than any of the individual distributions. Although the spherical geometry poses some technical complications, the solution is given in terms of the coefficients of a Legendre polynomial $P_\ell(x_L)$ expansion of the differential scattering cross section, where $x_L = \cos \theta$ and θ is the polar scattering angle. Sharp features are smoothed more rapidly than slowly varying ones. Mathematically, in this context the features of the scattering distribution associated with higher Legendre terms ℓ tend to zero faster than those with lower ℓ . The multiple scattering distributions are smoother than single-scattering distributions. Hence, going forward, in this article, we will restrict attention to sampling single-scattering distributions with the knowledge that any numerical grid sufficiently fine to sample a single-scattering distribution will also be sufficiently fine to sample the related multiple-scattering distribution.

2.1.2 CT Geometry and Sampling Requirements

The angular content of the differential cross sections represents one element of our analysis. If a sample is small compared to the space between the sample and the detector (as can arise in microCT), this is sufficient. Suppose, however, that the region to be reconstructed is a cylinder of radius r and that the minimum distance from the center of the reconstruction region to the detector is R , as shown in Fig. 1. For simplicity, we consider scattering in a plane and imagine that the detector has a number N pixels at radius R .

Suppose we have some criterion that provides an acceptable cut-off in ℓ for the differential cross section, namely ℓ_{\max} . Such a function has at most ℓ_{\max} wavelengths around a circle. Hence, by the Nyquist sampling theorem, it is sufficient to sample at $2\ell_{\max}$ points in one dimension, or one sample in a distance of $\frac{\pi R}{\ell_{\max}}$. The

use of flat-panel detectors does not make this estimate worse if R is interpreted as the minimum distance from the center of the reconstruction region to the detector.

The complication that the scatterer may be offset by a distance r implies that the minimum distance to the circular detector is $R - r$. Scattering about such a point leads to a requirement for more sampling points on the part of the detector closest to the scatterer. At the densest point, one sample per distance of $\frac{\pi(R-r)}{\ell_{\max}}$ is now required, *i.e.*, the worst-case sampling requirement has increased by a factor of $\frac{R}{R-r}$. Typically, $\frac{r}{R} \leq \frac{1}{2}$, so $\frac{R}{R-r} \leq 2$. On the other hand, if there is very small space between the sample and the detector, *i.e.*, if $R - r \ll R$, the sampling requirement diverges. Such a case does not correspond to the configuration of commercial tomographic instruments and must be excluded from the present analysis. In summary, the effect of the finite sampling region is to increase the sampling requirement by up to a factor of 2 in practical cases. The geometry is illustrated in Fig. 1c and its effect on a hypothetical scattering distribution is shown in Fig. 1d.

2.2 Cross Sections of Typical Elements

It is common to write the differential cross section as $\frac{d\sigma}{d\Omega}$, where $d\Omega = dx_L d\phi$ is the differential of the solid angle, $x_L = \cos \theta$, θ is the polar scattering angle, and ϕ is the azimuthal angle. Here, we are restricted to cases in which the differential cross section depends only on θ but not on ϕ . In such a case, given that a scattering event has occurred, we write the probability density function of the cosine of the polar scattering angle as

$$p(x_L) = \frac{1}{\sigma} \frac{d\sigma}{dx_L} \equiv \frac{2\pi}{\sigma} \frac{d\sigma}{d\Omega} \quad (2)$$

where p has the probability normalization

$$1 = \int_{-1}^1 dx_L p(x_L). \quad (3)$$

Total cross sections from the photon cross section database XCOM [21] in the range of photon energies typically used in tube-based X-ray tomography are shown in Fig. 2 for typical elements, namely C, O, Si, Ca, and Fe. Also shown are the three principal contributions to the cross section, namely elastic scattering, inelastic scattering, and photoabsorption. For carbon, inelastic scattering is seen in Fig. 2a to be the dominant process. As the atomic number increases, the minimum photon energy at which inelastic scattering dominates shifts to higher photon energies. Already for iron, scattering is not the dominant contrast mechanism in the range of tube voltages that are commonly used in tomography. This remains true for higher Z elements.

2.3 Transfer of Cross Section for Smoothing

The differential elastic (e) and inelastic (i) cross sections for X-ray scattering are expressed as [18]

$$\frac{d\sigma^{(e)}}{d\Omega} = |F(q)|^2 \frac{d\sigma^{(T)}}{d\Omega} \quad (4)$$

$$\frac{d\sigma^{(i)}}{d\Omega} = S(q) \frac{d\sigma^{(KN)}}{d\Omega} \quad (5)$$

where $\hbar q$ is the momentum transfer, F and S are the elastic and inelastic form functions, and the Thomson (T) and Klein-Nishina (KN) cross sections are:

$$\frac{d\sigma^{(T)}}{d\Omega} = \frac{r_e^2}{2} (1 + \cos^2 \theta) \quad (6)$$

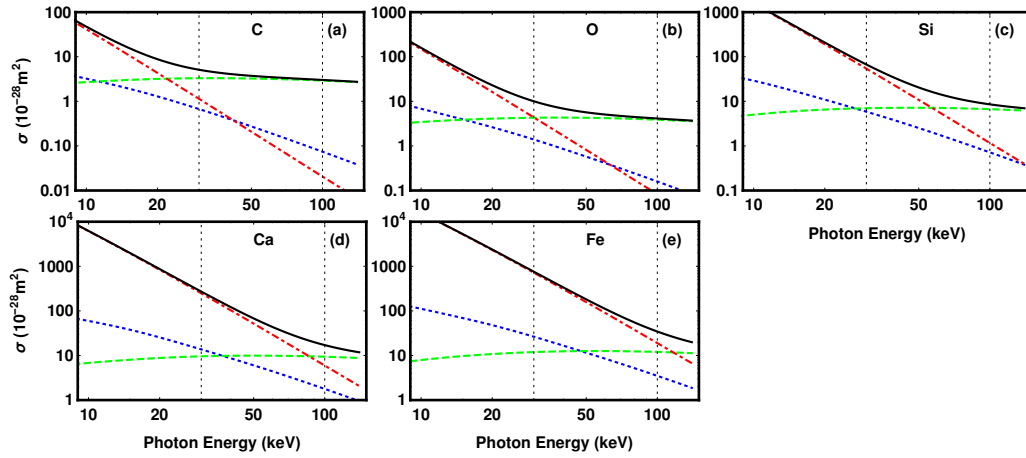


Fig. 2. Cross sections from XCOM [21] for (a) C, (b) O, (c) Si, (d) Ca, and (e) Fe in the range 10 keV to 140 keV. The region between 30 keV and 100 keV, bounded by vertical dashed lines, is particularly important for tube-based X-ray tomography. The contributions of elastic scattering (blue, dotted), inelastic scattering (green, dashed), and photoabsorption (red, dot-dashed) as well as the total (black, solid) are shown.

$$\frac{d\sigma^{(KN)}}{d\Omega} = \frac{r_e^2}{2} [P_{KN}^3 + P_{KN} - P_{KN}^2 \sin^2 \theta] \quad (7)$$

with r_e being the classical electron radius. Here

$$P_{KN} = \left[1 + \frac{E_\gamma}{m_e c^2} (1 - \cos \theta) \right]^{-1} \quad (8)$$

is the fractional energy loss. In Eq. (8), E_γ is the photon energy, and $m_e c^2$ is the rest energy of the electron. The functions $F(q)$ and $S(q)$ are the elastic and inelastic atomic form factors, respectively. Typical differential scattering cross sections are shown in Fig. 3, from the Evaluated Photon Data Library (EPDL) [22].

Next, we motivate a transfer of cross section from the elastic to the inelastic channel in order to smooth the inelastic function. The elastic scattering is largely in the forward direction. Hence, for pure elastic scattering, a significant number of events will end by hitting the detector. It is appropriate to use conventional Monte Carlo scattering in this case. Inelastic scattering occurs through much larger angles on average and tends to miss the detector, so the method of fixed forced detection (FFD) has been introduced to improve the efficiency of the simulation [23].

For inelastic scattering up to about 100 keV, the KN factor has a rapidly convergent Legendre series [24]. The total differential cross section is given by the sum of elastic and inelastic contributions, namely,

$$\begin{aligned} \frac{d\sigma}{d\Omega} &= \frac{d\sigma^{(e)}}{d\Omega} + \frac{d\sigma^{(i)}}{d\Omega} \\ &= \frac{1}{2\pi} \left(|F(q)|^2 \frac{d\sigma^{(T)}}{dx_L} + S(q) \frac{d\sigma^{(KN)}}{dx_L} \right) \\ &\approx \frac{1}{2\pi} \left([|F(q)|^2 + S(q) - Z] \frac{d\sigma^{(T)}}{dx_L} + Z \frac{d\sigma^{(KN)}}{dx_L} \right). \end{aligned} \quad (9)$$

However, the factor $Z - S(q)$ in this equation has a short angular range, and so its presence dominates the convergence of the Legendre series of the differential cross section, which in turn leads to a requirement

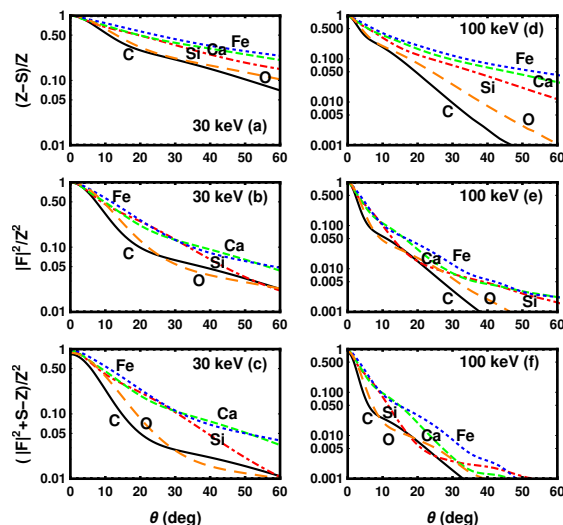


Fig. 3. Atomic form factors for C (solid black), O (long dashed orange), Si (dot-dashed red), Ca (dashed green), and Fe (dotted blue) from the EPDL as a function of the polar scattering angle θ . Curves are given for (a,d) the complement of inelastic scattering form factor S , scaled to unity at $\theta = 0$ (where S itself goes from 0 at $\theta = 0$ to at most Z), (b,e) the elastic scattering form factor scaled to unity at $\theta = 0$, and (c,f) the effective elastic scattering form factor for photon energies of (a–c) 30 keV and (d–f) 100 keV.

for high-spatial-frequency sampling. This requirement, in turn, necessitates a large number of fixed forced detection points. For each photon being detected, it must be projected from a position in the sample to all of the fixed forced detection points, so this step can easily dominate the total computational time for scatter corrections. If we could somehow fill in the hole created by the factor $Z - S(q)$, many fewer FFD points would be required.

We take advantage of an approximate equality between inelastic and elastic scattering cross sections for scattering in the forward direction. Figure 3 shows $S(q)$ differs from Z only for small values of the relevant parameter in Eq. (8), $\frac{E_\gamma}{mc^2}(1 - \cos \theta)$.

Because of the Pauli principle, $|F(q)|^2 \geq Z - S(q)$, a relation from the early days of quantum mechanics [18, 25], with equality holding only for hydrogen, the two terms on the right side of Eq. (9) are positive individually. We verified this point numerically by examining functions from the EPDL [22]. Hence, accepting a very small error, we may define effective elastic and inelastic cross sections according to these two terms. As shown in Fig. 4, the cross sections differ by less than 1 % in the relevant range defined by the strongly-peaked functions shown in Fig. 3. Furthermore, there is a less than 1 % fractional energy loss by the scattered photon at the angles relevant to the transfer of the cross section. This picture is reaffirmed in Fig. 5 where the effect of the transfer is shown on the inelastic and elastic differential cross sections and their sum. The inelastic differential cross section is much smoother after the transfer, whereas the sampling requirements for the elastic differential cross are little changed.

After transfer, the effective inelastic cross section is dramatically smoother, and so it may be calculated with a limited number of FFD points. The number may be estimated the rapidity with which the Legendre coefficients of the KN distribution approach zero. The formulas for these coefficients with numerical values were published over 50 years ago [24]. For 100 keV, the series is rapidly convergent, dropping off by about a factor of 10 for each ℓ . Fewer than 10 points in one dimension would suffice for sampling the function over the full range of polar angles, *i.e.*, 0 to π . However, the situation is even more favorable, because we only

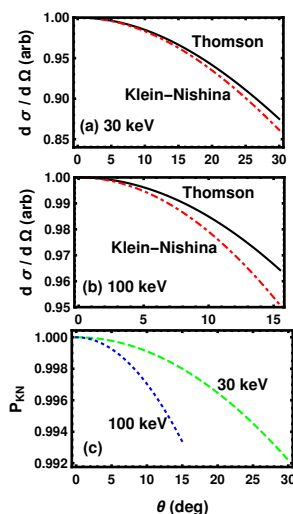


Fig. 4. Thomson (black, solid) and Klein-Nishina (red, dot-dashed) angular distributions, in the range of the transfer of cross section, relative to Thomson cross section 1 at $\theta = 0$ for photon energies of (a) 30 keV and (b) 100 keV. The Thomson and Klein-Nishina differential cross sections are equal at $\theta = 0$. (c) Fraction of energy retained by the photon after inelastic scattering P_{KN} .

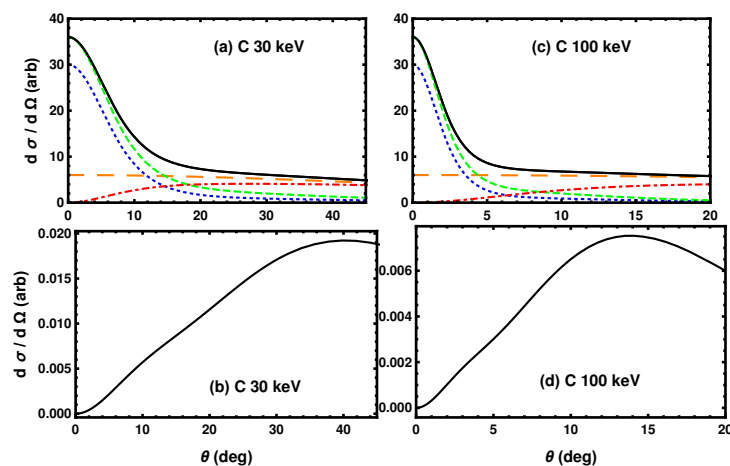


Fig. 5. The effect of the transfer approximation on the cross section for the case of carbon, for (a,b) 30 keV photons and (c,d) 100 keV photons. For parts (a,c), the elastic and inelastic differential cross sections (green, dashed green line / dot-dashed red line) sum to the total differential cross section (solid black line). The corresponding terms after transfer is elastic (dotted blue line) and inelastic (long dashed orange line). The sum is also given, but it is indistinguishable to within the width of the line plotted. (b,d) The differences of the differential cross section before transfer and after are shown using the same scale between parts (a) and (b) and between parts (c) and (d).

need to sample the Klein-Nishina function over the detector which is typically less than 1 sr. A low order polynomial suffices. Hence, the number of FFD points may be limited to the square of some small integer, e.g., 3^2 to 7^2 points. This is dramatically fewer than the number of points required if $Z - S$ were being sampled.

The elastic cross section is little changed by the transfer, except that the effective elastic cross section for

hydrogen vanishes. Since elastic scattering is largely in the forward direction, such scattering tends to hit the detector rendering the FFD procedure unnecessary. In our work, we handle the elastic scattering through an ordinary forward-propagating radiation transport Monte Carlo calculation. Mixed elastic and inelastic scattering is treated as inelastic scattering because the angular distribution is dominated by inelastic scattering. The transfer plays a large role in reducing the computational needs.

2.4 Angular Variation of the Cross Sections

To learn about the angular variation of the elastic scattered function, we consider the cases of carbon and water at 100 keV. Referring back to Fig. 3, the lower Z elements require the best angular resolution to sample the effective elastic scattering distribution. There is an exception for hydrogen: It vanishes. Because the elements helium, lithium, beryllium, and boron are not commonly studied in tomography, we consider carbon be a critical case, with water to be of similar interest. Recall also from Fig. 2 that the low Z elements tend to have scattering as their dominant contrast mechanism. Also, the higher photon energy requires higher angular resolution, and 100 keV is a practical upper bound for tube-based sources. Although higher tube voltages are used, e.g., 140 kV, most of the X-ray spectrum will be below 100 keV, even in this case. Although we only present one case, it is more or less a bound for tube-based X-ray tomography as it is normally practiced. Lower-energy photons require less spatial resolution to achieve the same degree of approximation to Nyquist sampling.

The Legendre expansion of the effective elastic scattering form factor is shown in Fig. 6. The Thomson cross section of Eq. (6), which has an exact two-term Legendre expansion proportional to $P_0 + \frac{1}{2}P_2$, would add little angular content if included. The fit to the signal, given in the caption of Fig. 6, may be reparameterized as a Gaussian with a standard deviation of $\sigma_\ell = 5.73$ centered on $\ell = -1.78$. If 99 % of the expansion is included, $\ell_{\max} = 14$. Spherical harmonics provide uniform resolution over the sphere, suggesting $(\ell_{\max} + 1)^2 = 225$ polynomials are sufficient for 4π sr. Given that two samples per wavelength will represent the Nyquist sampling limit, we can estimate that one sample per $\frac{2(4\pi)^{1/2}}{15\pi} \approx 0.15$ rad (8.6°) is sufficient to sample the effective elastic scattering form factor.

As discussed earlier, the spatial resolution required by the detector depends on the location of the scatterer in the reconstruction region. Simple geometrical considerations lead to a variation by up to a factor of $\frac{R+r}{R-r}$. Although this factor can be arbitrarily large in principle, in practice the distance from the center of the sample to the detector R is at least twice the radius of the sample r , leading to a bound of a factor of 3 variation in the factor relating angular scattering to the region affected on the detector plane. Given that we expect the detector to subtend an angle larger than 0.15 rad, a large fraction of the elastically scattered events will hit the detector, obviating the need for FFD in this part of the problem. Instead, we collect these events in a virtual detector which is somewhat larger than the physical detector, and then we use Fourier smoothing with a bandwidth dictated by the estimate given here.

In Fig. 7a, we show the scatter for 10^8 photons through a 1 nm^3 cube of water with a density of 10^{10} kg/m^3 . The scatter is nearly circularly symmetric, and it has an isotropic one-dimensional (1D) standard deviation of 45.5 mm. We chose a 150 kV tube spectrum with a 5 mm Al filter and a 0.25 mm Cu filter using TASMICS [26]. The filter was chosen to match the M150 beam quality that is used to calibrate X-ray detectors at NIST [27].

We choose the signal and noise model to make a Wiener filter [28]

$$W(k) = \frac{|S(k)|^2}{|S(k)|^2 + |N(k)|^2}. \quad (10)$$

Our choice of the noise model is a spatial-frequency-independent constant N_0 , which is suggested by theory

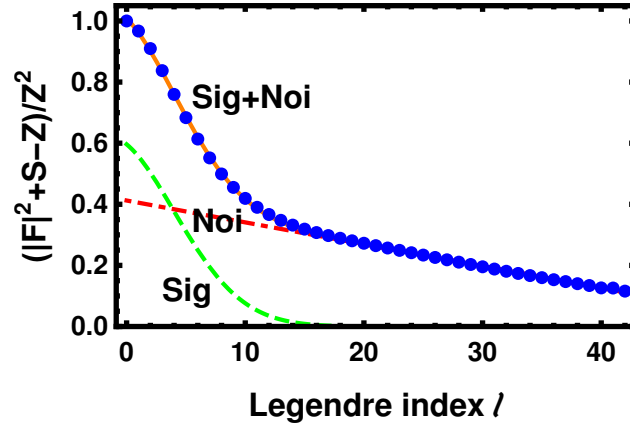


Fig. 6. Legendre expansion of effective elastic scattering form factor $|F|^2 + S - Z$ for carbon at 100 keV (blue dots), as fit by a signal model (Sig, dashed green line), $0.596218 \exp(-0.0539594\ell - 0.0151884\ell^2)$, and a noise model (Noi, dot-dashed red line), $0.412347 - 0.00715297\ell$. The sum (Sig+Noi, solid orange line) is also shown.

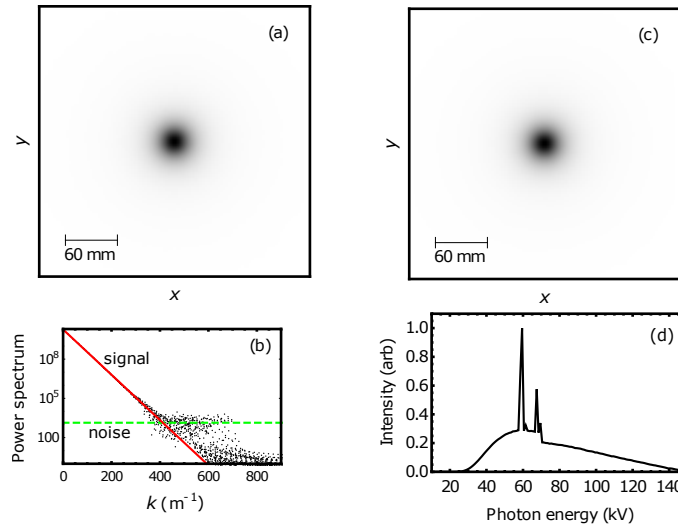


Fig. 7. Scatter from a 1 nm^3 cube of water including (a) the unfiltered scatter, (b) the signal and noise models of the power spectrum (arbitrary units), and (c) the filtered scatter. The simulation uses (d) the M150 source spectrum.

for counting statistics and is also used in Tikhonov regularization [29]. The signal model is taken to be

$$S(k) = S_0 \exp(-k/k_0). \quad (11)$$

We give here $S_0/N_0 = 3672$, which we round to 4000 for the program and $k_0 = 50 \text{ m}^{-1}$, with k_0 being -2 times the slope of the fit shown in Fig. 7b. Figure 7c shows the sample after the filter is applied. The filtered spot strongly resembles the original. For example, the angular average of 1D standard deviations is 45.5 mm in both cases. In practice, we collect scatter over a hypothetical detector that is twice as large per dimension as the actual detector, apply Fourier filtering on the larger detector, and then crop the filtered result to our physical detector. In this way, we avoid the potential problem of an artifact due to the periodicity assumption inherent in Fourier analysis, which appears at the edge of the domain.

Because the effective inelastic scattering leads to particle trajectories that usually miss the detector, FFD [23] is an effective way to gain efficiency. Using FFD, when a scattering event occurs, the probability of the particle propagating to fixed points on the detector is found. Attenuation of the particle on the way to the detector is taken into account, as are solid angle factors. In order to do this, a particle must be propagated to each of the points on the FFD grid. Because this occurs for every particle sampled by Monte Carlo, it requires a relatively large computational effort. To minimize this effort, we select Chebyshev collocation points and smooth the resulting counts using Chebyshev polynomials over the detector.

As discussed above, for the effective elastic scattering, the scattering angles are modest and the scattered particles tend to hit the detector, so FFD is not required. This keeps the cost per photon down. Because there may be a substantial number of effective bins, smoothing in the Fourier domain is an effective strategy. Fourier smoothing has a problem with boundaries if the function is not periodic. For elastic scattering, we address that issue by collecting scattered points on a virtual detector that extends past the physical detector by at least a couple of smoothing lengths. In this case, boundary effects of smoothing are confined to a region that does not intersect the physical detector. Fourier smoothing followed by truncation are then effective strategies.

The fact that the inelastic scattering is smoother than the elastic scattering suggests intuitively that it does not need so many Monte Carlo photons to sample. A quantitative statement of this principle is given in the Appendix. However, in this work, we retain the natural odds of elastic and inelastic sampling.

The purpose of studying the cut-off for the Wiener filter is as follows. Nyquist's theorem says that a band-limited function may be sampled at two points per period. The parameters of the Wiener filter, combined with a willingness to neglect a small fraction of the power spectrum at high spatial frequency, lead to the number of angular samples required. Determining this number is a major goal of the present study.

2.5 Reconstruction Method for Multi-Energy, Multimaterial Projective Tomography

The reconstruction method is presented here. We use a Bayesian approach based on the Bouman-Sauer formulation [30]. However, unlike the Bouman-Sauer formulation, we use a constant for the prior distribution, which does not express a preference for geometric structure, *a priori*, like the Bouman-Sauer prior choice. Such an assumption is reasonable in many situations including random-dot patterns. We include scatter corrections iteratively using the moving, expanding window method of Levine and Pinter [31]. In this method, the calculated Monte Carlo scatter from the most recent 50 % of the iterations values is included, balancing the need to keep the number of Monte Carlo trials moderate while omitting scatter corrections from early reconstruction iterations which are no longer useful at a given level of precision.

The problem — and our reconstruction code — is formulated for the multi-material case [32]. In this paper, we only reconstruct systems with one material. However, we do account for beam hardening. We assume that we have a beam composed of photons of various energies. In projective tomography, when the photons are attenuated, it is assumed that they do not hit the detector pixel hit by the projection. In this case, Beer's law holds, energy by energy, albeit not for the beam as a whole. The non-exponential nature of the sum is known as beam hardening. The role of scatter corrections is to make the assumption of projective tomography valid.

Suppose there are several spectra, $I_j^{(0)}(E)$, generated by a set of X-ray sources, assumed to be used at different times. Suppose there are several materials in the sample, the attenuation parameters of which are given by $\alpha_i(E)$. Let the detector sensitivity be $D(E)$. Although a single function is assumed here, as we will see, only the dose-detector product $D(E)I_j(E)$ enters the theory. This recognition implies that systems with different values of $D(E)$ can be included in the present formulation with a small change of notation. The

observations at several viewpoints indexed by ψ are given by

$$I_{j\psi} = \int dE D(E) I_j^{(0)}(E) \exp \left(- \sum_{i\vec{r}} f_{\vec{r}i} \alpha_i(E) A_{\vec{r}\psi} \right) \quad (12)$$

where $f_{\vec{r}i}$ represents the amount of material of type i present in a voxel indexed by \vec{r} , and $A_{\vec{r}\psi}$ is the distance a ray, indexed by ψ , travels going through a voxel at \vec{r} . Here, \vec{r} is a discrete index. The matrix $A_{\vec{r}\psi}$ is sparse because a single ray will intersect $O(N)$ of N^3 voxels as it traverses the reconstruction domain. We will need the derivatives

$$\frac{\partial I_{j\psi}}{\partial f_{\vec{r}i}} = - \int dE D(E) I_j^{(0)}(E) \exp \left(- \sum_{i'\vec{r}'} f_{\vec{r}'i'} \alpha_{i'}(E) A_{\vec{r}'\psi} \right) \alpha_i(E) A_{\vec{r}\psi}. \quad (13)$$

The objective function, which combines information from the detectors with prior information, is given by [30]

$$L_{MAP}(\vec{n}|\vec{f}) = L_{ML}(\vec{n}|\vec{f}) + g(\vec{f}) \quad (14)$$

where \vec{n} represents the counts in each channel (with an assumed Poisson distribution), \vec{f} is the proposed reconstruction, and $g(\vec{f})$ is the prior distribution. The first term is the likelihood function derived assuming that each observation n_J obeys the Poisson distribution with mean I_J . Maximizing Eq. (14) gives the maximum *a posteriori* (MAP) estimate whereas maximizing L_{ML} alone is the Maximum Likelihood (ML) method.

$$L_{ML}(\vec{n}|\vec{f}) = \sum_J \left[\ln n_J! - n_J \ln I_J(\vec{f}) + I_J(\vec{f}) \right] \quad (15)$$

where $J = (j\psi)$ is a joint index. Note that if the intensities I_J were each allowed to vary independently, the condition $\partial L / \partial I_J = 0$ implies $n_J = I_J$.

Choosing $g(\vec{f})$ for this problem is a research topic, although there is a good deal of published work, such as the q -GGMRF method on the single-material problem [33]. In this work, we simply take $g = 0$, making the MAP estimate the ML estimate.

The function L_{ML} or L_{MAP} may be minimized using the L-BFGS-B algorithm [34]. The BFGS method is a quasi-Newtonian method for the minimization of high-dimensional functions. The L refers to the use of limited memory. The B refers to the use of bounds. In our case, it is natural to insist that there be at least none of any given material in any given voxel, *i.e.*, $f_{\vec{r}i} \geq 0$. In some cases, there may be plausible upper bounds as well. In performing the minimization, a useful expression is

$$\frac{\partial L_{ML}}{\partial f_{\vec{r}i}} = \sum_J \frac{\partial I_J}{\partial f_{\vec{r}i}} \left(1 - \frac{n_J}{I_J} \right). \quad (16)$$

The L-BFGS-B method uses values of the function and its gradient. It is possible to find these together with considerable reuse of intermediate results.

Much of the time in the algorithm is devoted to calculating projections through the voxels. The integral over energy may be performed after the projection has been calculated. If the number of voxels is large compared to the number of energies, the exponential and the sum over energies are not the dominant contributions to the run time.

As an example of the program applied without scatter corrections, we take the case of a binary pattern with $128 \times 128 \times 1$ voxels using the tube spectrum of Fig. 7d for the expression $I_j^{(0)}(E)$ and only a single value of j . The average energy is 74 keV. The detector response $D(E)$ is taken to be proportional to E . The

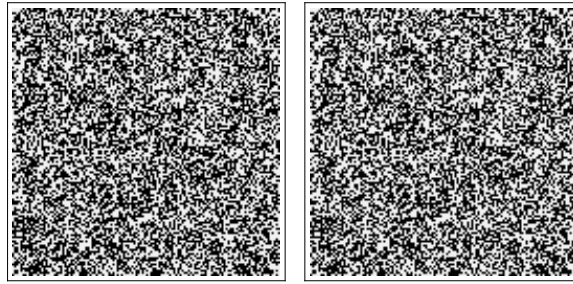


Fig. 8. Tomographic reconstruction of a $128 \times 128 \times 1$ random dot pattern (left) and its reconstruction (right) within the purely projective model. The snow-and-ice model is used, so the low values represent water at a density of 50 kg/m^3 and the high values represent water at a density of 900 kg/m^3 . The M150 spectrum is used. The detector has 256×1 pixels, and 192 angles are sampled.

product is normalized to a probability and sampled with the Walker Alias method [35]. Our code for Walker Alias sampling is publicly available [32].

The observations that are used in Eq. (15) refer to projective tomography, and, as such, they represent values after the scatter corrections are made. We obtain scatter corrections with Monte Carlo radiation transport, as described above. Inelastic energy loss is accounted for through “Russian roulette” [36]; *i.e.*, if there is energy loss, the probability of propagating the photon is proportional to the energy retained. In this way, we retain Poisson statistics among the detected photons. Moreover, energy loss is a proxy for a photon heading away from the detector.

3. Results

3.1 Verification

As discussed in the introduction, the purpose of scatter corrections is to bring the problem back to one that can be solved by projective tomography. Accordingly, we begin by illustrating our code based on the algorithm given above is capable of reconstructing complex patterns. In Fig. 8, we show the reconstruction of a randomly chosen two-dimensional (2D) binary pattern of size $128 \times 128 \times 1$ voxels. The reconstruction and the original pattern are visually identical, with a root mean square deviation of 2.8 kg/m^3 . The reconstruction parameters are shown in Table 1, although no Monte Carlo photons were used in this case. The pattern is in an ice-and-snow model, consisting of water with densities of either 50 kg/m^3 or 900 kg/m^3 chosen by an independent Bernoulli trial in each voxel with probability $\frac{1}{2}$. We sample in our standard geometry, with the source 300 mm from the center of the pattern, with a pattern width of 100 mm, and with the detector also 300 mm from the center of the pattern. The number of angles is $\frac{3}{2}$ times the number of voxels in 1D (*i.e.*, 192) and the number of detector elements is twice the number of voxels in 1D (*i.e.*, 256). In this case, the detector is 256×1 pixels, but it will be chosen as a square in the scattering examples. Both the simulated data and the reconstruction treat all interactions as purely attenuating.

Table 1. Throughout the paper, we principally use the parameters shown in this table N_{vox} is the number of voxels across 1 edge of the cube, N_{det} is the number of pixels across 1 edge of the square detector, N_{ϕ} is the number of angles, and N_{phot} is the number of photons used in the Monte Carlo analysis per angle. In some cases, N_{phot} is reduced by a factor of 10 or 100.

N_{vox}	N_{det}	N_{ϕ}	N_{phot}
16	32	24	655 360 000
32	64	48	655 360 000
64	128	96	655 360 000
128	256	192	655 360 000

We used a high density of hydrogen located at the origin as a test case. Hydrogen was chosen because it has an analytic scattering cross section [37], namely,

$$F_H = \left[1 + \left(\frac{k s a_0}{2} \right)^2 \right]^{-2} \quad (17)$$

for the elastic scattering form factor and

$$S_H = 1 - F_H^2 \quad (18)$$

for the inelastic scattering form factor, where a_0 is the Bohr radius, $k = 2\pi/\lambda$ for an X-ray with wavelength λ , and $s = 2 \sin(\theta/2)$ is a scattering parameter.

For a numerical test, we simulated 10^7 photons of 30 keV. The hydrogen is concentrated in a cube 1 nm on a side with a density of 10^8 kg/m^3 . The transmission probability is $\exp(-3.57 \times 10^{-3})$, so multiple scattering can be neglected because it is second order in the argument of the exponential. We take a $1 \text{ m} \times 1 \text{ m}$ square detector with 100×100 pixels 2 m from the source. The intensity of particles at a given detector pixel includes an inverse square factor as well as an obliquity factor and the KN energy loss factor in the case of inelastic scattering. The pixel-by-pixel results of the elastic test are shown in Fig. 9a along with theoretical predictions. A χ^2 test was performed among the 4420 pixels that had at least 5 counts, a validity condition for the test [38]. A value of 4457 was observed, leading to a p value of 0.34, therefore, the model does not exhibit lack-of-fit.

The fourth degree Chebyshev polynomial fit is shown. A second degree fit shows deviations at the level of 5 %, whereas sixth- and eighth-degree fits show no noticeable improvement. The inelastic test results are shown in Fig. 9b.

In addition to providing a test of the program, it is helpful to ensure that the EPDL functions are being read and interpreted properly. Another verified aspect is that the integrated differential cross sections of EPDL are very nearly equal to the total cross sections given by XCOM. By matching to quantum mechanical theory and a second data base, we have verified the underlying validity of our radiation transport simulation.

3.2 A Critical Example

To avoid extensive examples, we select only one which is, however, relatively challenging for scatter corrections.

1. We chose water as the material. Water is commonly found in tomographic subjects, including all living things. Oxygen is one of the more challenging elements for elastic scattering, because its angular distribution is nearly as tight as that of carbon, but the elastic cross sections are much higher.

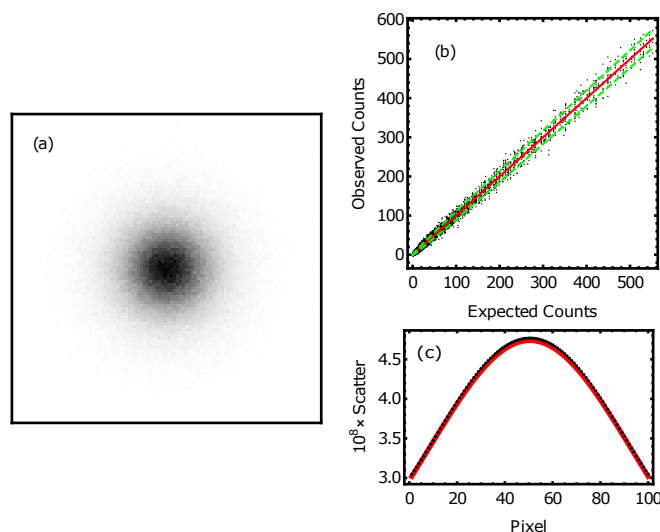


Fig. 9. Scattering of 30 keV photons through hydrogen as described in the text. (a) Elastic scattering intensity values over the 100×100 pixels, with a full angular range of 27° . (b) Observed counts from part (a) for 10^7 photons of 30 keV (black dots) vs. per pixel results vs. analytically expected results (red, solid) with ± 1 standard deviation bounds (green, dashed lines) based on the Poisson distribution. (c) Inelastic scatter per photon into each pixel across a central slice from theory (red) vs. numerical simulation (black), as described in the text. The numerical data were fit to a 2D cross product of fourth degree Chebyshev polynomials.

2. We considered spheres embedded in a light background. Spheres remain of current industrial interest for verification and validation of image segmentation software [39]. We call the two materials “ice” and “snow.” The ice has a density of 900 kg/m^3 (compared to a physical density of 917 kg/m^3 [40]), and the snow has a density of 50 kg/m^3 , which is at the low end of the physical range [41]. We concentrated on the results from a central slice both to simplify the visualization and in recognition of the existence of cone beam artifacts [14], which we do not address in this work to keep the focus on scattering physics.
3. We chose the simulated M150 beam shown in Fig. 7d. This tube voltage is on the high side for use in practical tomography, which leads to tighter patterns in the scattering as discussed above.
4. Noting that single scatter is the most challenging, and the proportion of photons undergoing single scatter is maximized when the system diameter is one interaction length, we chose a system where the interaction lengths are roughly that size. The XCOM cross section for water at 74 keV is $0.1888 \text{ cm}^2/\text{g}$, leading to an interaction length of 58.9 mm for a density of 0.9 g/cm^3 . Hence, we choose one or two spheres with a diameter of 50 mm to be placed in a cube having a side given by $a = 100 \text{ mm}$.

In the example, we spanned the extremes of the photon energies likely to be encountered in practice. We argued above that carbon is the most challenging case for scatter corrections among commonly studied elements. Water is the traditional subject in scattering studies. Water, dominated by the scattering by oxygen, is seen in Fig. 3 to have scattering properties very similar to carbon. With thicker samples, we would have more multiple scattering, which is smoother. With thinner samples, the scattering distributions would be only marginally tighter, but the Monte Carlo calculation would become increasingly inefficient as fewer samples would be scattered at all.

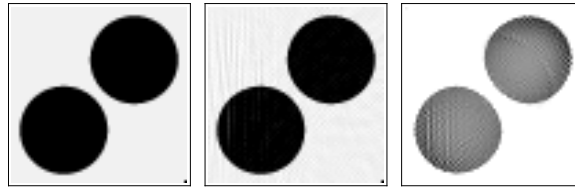


Fig. 10. A central slice from the 64x64x64 two sphere ice-and-snow model with white at 0 kg/m³ and black at 900 kg/m³. (left) The ideal object with white at 0 and black at 900 kg/m³; (middle) the sinograms are generated and reconstructed treating all interactions as absorptive; and (right) the sinograms are generated with scatter corrections but reconstructed without scatter corrections. To ensure uniform scaling of the gray scales, the upper-right and lower-left voxels have been set to the minimum and maximum values, respectively, in each image. The “ice” appears as dark balls and the “snow” appears as a light gray background.

Our detector geometry is somewhat arbitrary, but it is representative of microtomography instruments. The source is taken to be a point 300 mm from the center of the sample, and 600 mm from the detector. The detector width may be determined using Eq. (1) with $L = R = 300$ mm, and $r = a/\sqrt{2} = 71$ mm, so $W \geq 291$ mm is required. We pick $W = 300$ mm. In the study, we scaled the number of voxels in the reconstruction region and the number of pixels in the detector and held the overall physical sizes (*i.e.*, a and W) fixed.

Our reconstruction is shown in Fig. 10. Only a central slice is given. The ideal spheres are shown in Fig. 10a. Partial volume effects were accounted for by making a $256 \times 256 \times 256$ voxel binary pattern with the snow-and-ice densities then averaging over $4 \times 4 \times 4$ voxel blocks to create a $64 \times 64 \times 64$ voxel target structure. The reconstruction with scatter corrections is shown in Fig. 10b. If the scatter corrections are omitted in the reconstruction but are present in the simulated sinogram, the reconstruction of Fig. 10c results. Although the basic result of two spheres is present in the figure, the spheres are not uniform, and their overall level is off by about 25 %.

3.3 Angular Variation of the Scatter Corrections

Scatter corrections for selected detector pixels as a function of viewing angle are shown in Fig. 11a for the $16 \times 16 \times 16$ voxel case and Fig. 11b for the $64 \times 64 \times 64$ voxel case. There are two points to observe: (1) The angular variation for each pixel is a relatively gentle function of angle. (2) The number of detector pixels increases, but the patterns do not become more complex.

In Fig. 12a–d, we show the power spectrum, *i.e.*, the squared magnitude of the Fourier transform from the azimuthal angle ϕ to its conjugate m . (We have a Fourier series with basis functions $e^{im\phi}$.) Using the data in Fig. 12a–d, we construct a signal model

$$|S(m)|^2 = 10^{-\beta[\exp(-m/m_0)-1]} - 10^{-\beta} \quad (19)$$

and a noise model

$$|N(m)|^2 = 10^{-\beta}. \quad (20)$$

We use these to construct a second Wiener filter, in analogy with Eq. (10),

$$\Phi(m) = \frac{|S(m)|^2}{|S(m)|^2 + |N(m)|^2}. \quad (21)$$

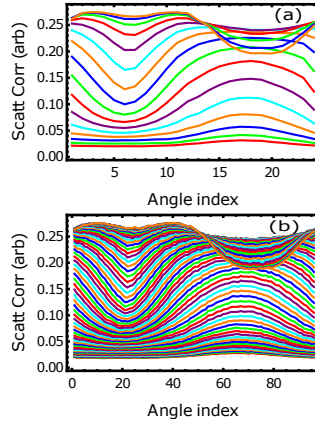


Fig. 11. Scatter corrections for detector pixels for a middle column (16 of 32) and the lower half of the rows as a function of angle index for the (a) $16 \times 16 \times 16$ voxel example and (b) $64 \times 64 \times 64$ voxel example. The corresponding angles run from 0° to (a) 174.4° in steps of 5.625° and to (b) 178.6° in steps of 1.406° .

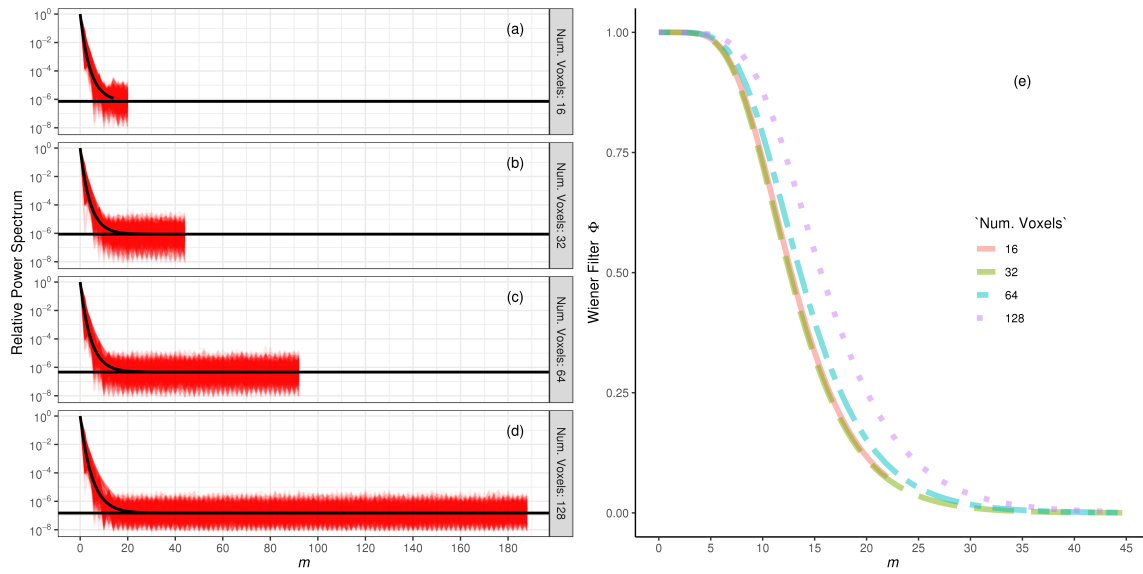


Fig. 12. The power spectrum (*i.e.*, square of the Fourier transform) for 1024 representative detector pixels is shown for (a) the $16 \times 16 \times 16$ voxel case, (b) the $32 \times 32 \times 32$ voxel case, (c) the $64 \times 64 \times 64$ voxel case, and (d) the $128 \times 128 \times 128$ voxel case, along with a signal and noise model described in the text with parameters from Table 2. Here, m is the conjugate variable to ϕ and is known in quantum mechanics as the azimuthal quantum number. (e) The Wiener filter derived from the signal and noise models of parts (a-d) and Table 2.

Note that $S(k)$ and $S(m)$ are unrelated, and $N(k)$ and $N(m)$ are also unrelated. The characteristic size may be parameterized by the value

$$m_{1/2} = -m_0 \ln \left[\frac{\log_{10} 2}{\beta} \right] \quad (22)$$

which obeys $\Phi(m_{1/2}) = 1/2$ by definition.

Table 2. Parameters β and m_0 in Eq. (19) and Eq. (20), for the fit to the signal and noise models of Fig. 12. The parameter $m_{1/2}$, determined by β and m_0 as shown in Eq. (22), is defined by $\Phi(m_{1/2}) = 1/2$ in the Wiener filters shown in Fig. 12e. The standard error of each fitting parameter, determined by bootstrap resampling [42], is given in parentheses.

N_{vox}	10β (dB)	m_0	$m_{1/2}$
16	61.41(19)	4.229(55)	12.75(17)
32	60.57(7)	4.195(23)	12.59(7)
64	63.35(4)	4.464(13)	13.60(4)
128	68.26(2)	5.000(7)	15.61(2)

As anticipated by the similarity between Fig. 11a and Fig. 11b, as well as the similarity among parts Fig. 12a–d, the Wiener filters, shown in Fig. 12e, with parameters tabulated in Table 2, do not show much dependence on the size of the system. As N_{vox} increases, we see the standard errors of the fit decrease and the parameter β , the $m = 0$ signal to noise in bells, increases. The m_0 and $m_{1/2}$ parameters increase modestly as well. While the main message is that these parameters are roughly constant we attribute the small changes to the fact that we have a fixed number of photons per angle, so there are more total samples as N_{vox} increases. Recall from Table 1, $N_\phi = \frac{3}{2}N_{\text{vox}}$.

3.4 Timing

In Table 3, we show the time required to perform reconstructions for the various size systems. The evaluation machine we used has two Intel Xeon E5-2680 v4 Broadwell Processors (28 physical cores, 56 logical cores) with 512 GB of DDR4 memory.¹ The time to create the simulated data is dominated by the scatter corrections, and these are linear in the number of Monte Carlo photons as seen in column 3. For the reconstruction, adding more photons has decreasing importance as the system size increases, as seen in column 4. In our implementation, we used the same number of Monte Carlo photons for each view. The previous section suggests that it would be sensible to use the same number of Monte Carlo photons per reconstruction, which would reduce the scaling of the time to make the scatter correction by one power of the linear system size.

Although the scatter corrections take most of the time, if the voxel size of the reconstruction were decreased, the time for projective tomography would increase, perhaps as the inverse fourth power of the voxel size, whereas the time to compute scatter corrections would increase much less quickly. Moreover, if the scatter corrections were made on a coarse-resolution version of the reconstruction, their computational time could be independent of the voxel size.

¹Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment are necessarily the best available for the purpose.

Table 3. Timing for runs. The number of photons is given per view, using the parameters from Table 1. The value of 1 for $N_{\text{phot}}^{(\text{rel})}$ corresponds to N_{phot} given in Table 1. $N_{\text{phot}}^{(\text{rel})} = 0$ means no scatter corrections were applied in either the creation of the sinogram or its reconstruction. For reference, 3600 s = 1 h and 86 400 s = 1 d.

N_{vox}	$N_{\text{phot}}^{(\text{rel})}$	Sinogram Time (s)	Reconstruction Time (s)
1D			
16	0	0.3	1.4
16	0.01	13	65
16	0.1	128	162
16	1	1 198	1 263
32	0	0.9	16
32	0.01	39	846
32	0.1	367	1 113
32	1	3 664	4 249
64	0	3.2	333
64	0.01	150	17 153
64	0.1	1 139	17 997
64	1	11 123	25 083
128	0	15.5	4 738
128	0.01	992	297 479
128	0.1	4 910	309 908
128	1	40 312	311 514

4. Concluding Remarks

The main question that we addressed in this work was how to set up a sampling scheme for scatter corrections in X-ray CT that will be close to its necessary and sufficient bounds. We do not claim to have achieved rigorous bounds, but merely to suggest approximately where these bounds lie.

We presented an analysis of the sampling requirements for CT scattering based on the underlying physics, assuming the independent atom approximation. We first showed that these would be determined by single scattering events — as opposed to multiple scattering which is smoother. We considered both the total and differential cross sections as tabulated for the range of photon energies relevant to X-ray tubes. We discussed the geometric effects associated with translating angular ranges to spacing on the detector. We identified a highly accurate approximation to make the inelastic scattering smoother by transferring a portion of the elastic scattering cross section to it. Atomic physics guarantees such a transfer is always possible without creating negative differential cross sections in the untransferred portion. Although this work concentrated on the smoothing in 2D, smoothing in 3D is possible, as shown earlier [23]. Here, we develop principles for choosing the angular spacing. We used fixed forced detection (FFD) like other authors [23], but we greatly reduced the number of FFD collection points through our analysis and through the use of fitting functions that span the whole detector. FFD, as implemented, samples at the spatial resolution of the elastic scattering over the whole detector. This scales badly if the detector has a large solid angle. Our FFD sampling was based on the inelastic scattering angular resolution that is about 3 orders of magnitude larger in solid angle. Moreover, we avoided the high spatial resolution which elastic scattering would require by using Monte Carlo, which was appropriate because the small elastic scattering angles imply that the detector was hit in a large proportion of the cases, so FFD was not necessary. Moreover, it

was not desirable because a given scattered photon will have very little weight on most of the detector. The heart of our proposal to achieve numerical efficiency is to recognize that the elastic scattering and inelastic scattering (after the transfer approximation) have different angular scales and should be treated with different numerical methods.

We implemented these ideas in the context of a Bayesian iterative reconstruction algorithm. We found that the scatter correction can be found at low resolution. Going forward, this suggests a multi-grid approach may lead to an effective algorithm, wherein the scatter corrections are always found in an appropriately sparse volume. The method of Levine and Pinter [31] may be used to improve the scatter corrections iteratively given further knowledge of the ultimate reconstruction. If a multi-grid approach is taken, the scatter corrections need to change only to the extent that high-resolution changes lead to changes at low spatial frequencies. The hope is that, using the methods of this paper, the time to compute an iterative reconstruction will ultimately be dominated by the time required to solve the projective problem iteratively.

We introduced an approximation based on the underlying physics to split the scattering into two components: the inelastic scattering, which can be made quite smooth, and the elastic scattering which is much more strongly peaked. The traditional approach to X-ray scattering has two terms that have about the same degree of angular variation. After the transfer from the elastic cross section to the inelastic cross section, the inelastic part is much smoother. Since inelastic scattering is dominant, this leads to both conceptual and algorithmic simplifications. We used a derivation from the early days of quantum mechanics [18, 25] to guarantee that the transfer is always permitted without generating negative quantities in the now-reduced elastic channel.

In our implementation, we treated the two types of scattering differently: The inelastic scatter, which tends to miss the detector, was found using FFD. We chose our FFD points to permit Chebyshev interpolation. After the transfer approximation, the inelastic scattering is smooth. This permitted us to use a small number of FFD points. Because inelastic scattering is the dominant type, the time to compute scatter corrections is greatly reduced with this separation.

However, the elastic scattering, which tends to hit the detector, was found using ordinary Monte Carlo calculation followed by Fourier smoothing. We introduced an extended detector to avoid having to deal with end effects introduced by the artificial periodicity of the Fourier domain representation. These are not the only schemes available, but in general, we recommend acquiring the inelastic and elastic scatter on different grids. In some geometries, the elastic scattering may also be better handled with FFD points. Even if these are used, we nevertheless recommend that elastic and inelastic scatter be collected on different grids, since the inelastic grid can have many fewer FFD points.

We provide a theorem to enhance convergence based on the idea that scatter with more angular variation needs to be sampled more than the natural odds ratio allows in the Appendix. Future work might generalize the theorem to account for the cost of calculating one type of scattering event vs. the other.

The angular dependence of the scatter corrections appears to be largely independent of the resolution in the range we considered. We believe that this is tied to the underlying physics of scattering: The functions have a certain smoothness, which is defined by the differential cross sections and the sampling geometry. Looking ahead to algorithms, this suggests that it may be possible to obtain the scatter corrections at low spatial resolution and use them to correct a high spatial resolution projective algorithm. The finding that $m = 14$ is sufficient to capture the angular resolution would, for 2D Nyquist sampling, require 28 samples around the circle, which, in turn could be represented by a cube $(2/\pi)28 \approx 20$ voxels on a side. Given that applications typically generate reconstruction regions that are 512 voxels on a side, there is a huge potential computational savings by calculating the scatter corrections at a fixed coarse-resolution that does need not to vary as the spatial resolution of the reconstruction increases. Topics for future research include exactly how to communicate the information from the coarse to the fine grids, and whether any calculation of the scatter

correction on the fine grid needs be done. It may also be useful to define an algorithm in which the inelastic scatter is calculated on a smaller (and hence cheaper) set of voxels than the elastic scatter. In our algorithm, we have two loops, an inner loop given by the L-BFGS-B algorithm, and an outer loop where scatter corrections were given using the moving, expanding window method [31]. We converged the solution to the same level of accuracy at each iteration. This probably is not necessary, and an improved algorithm could gradually impose stricter convergence criteria as the scatter corrections are accumulated blending the two iterative processes. Under the Moving Expanding Window method, the scatter correction is changed slowly. Given that the BFGS method keeps a limited history, it may be possible that the adjustment to the scatter correction is small enough that the BFGS method may be used without change.

The broader challenge ahead is to design an algorithm for which scatter corrections are not a dominant part of the computation for a typical reconstruction (e.g., $512 \times 512 \times 512$ voxels). We hope that the present work contributes towards that goal.

5. Appendix: Monte Carlo Sampling Requirements for Scattering Processes with Unequal Angular Resolution

The inelastic scattering cross sections are much larger than the elastic scattering cross sections in the photon energy range of interest in X-ray tomography, as shown in Fig. 3. After the transfer approximation discussed in Sec. 4, the inelastic scattering is much smoother in angle. The transfer approximation involves a relatively small amount of the total cross section, so the values in Fig. 3 are still valid within the linewidths on the graph. Intuitively, we expect the increased smoothness to lower the sampling requirement. Here, we give a formula for the sampling requirement.

As discussed already, we limit the analysis to single scattering, because single scattering requires the highest angular resolution. Suppose we are doing a radiation transport Monte Carlo calculation and we choose the number of photons to sample from a Poisson distribution with mean N . Of these, on average Np scatter inelastically, and $N(1-p)$ scatter elastically.

The random number of inelastically scattered photons, Y , follows a Poisson distribution with mean Np . We model the difference in angular resolution as follows. We imagine that there is a section of the detector wherein the inelastic scattering is more or less uniform, but for elastic scattering, we must divide the detector into N_d bins wherein the elastic scattering is more or less uniform. The relative frequencies of elastically scattered photons in the N_d bins are given by μ_j , where $\sum_j^{N_d} \mu_j = 1$. Hence, the random number of elastically scattered photons arriving at bin j , Z_j , is Poisson distributed with mean $N(1-p)\mu_j$.

Let T_j be the distribution of the total number of photons arriving in bin j . T_j is Poisson distributed with mean $\frac{Np}{N_d} + N(1-p)\mu_j$. Said differently, $T_j = \frac{Y}{N_d} + Z_j$. By definition, T_j is an unbiased estimator of its mean, $\frac{Np}{N_d} + N(1-p)\mu_j$.

If we need to estimate the mean of T_j for all j , can we use importance sampling to reduce the total variance given that the contribution from the inelastic scattering is constant over the bins? Let p_1 be the probability of an inelastic event. Let Y_1 be a Poisson-distributed random variable with mean Np_1 and let Z_{1j} be a Poisson distributed random variable with mean $N(1-p_1)\mu_j$. The sum process

$$T_{1j} = \frac{p}{p_1} Y_1 + \frac{1-p}{1-p_1} Z_{1j} \quad (23)$$

has the same mean as T_j , i.e., $\frac{Np}{N_d} + N(1-p)\mu_j$. The variance of T_{1j} is given by

$$\text{Var}(T_{1j}) = \left(\frac{p^2}{p_1^2 N_d^2} + \frac{(1-p)^2}{1-p_1} \mu_j \right) N. \quad (24)$$

Minimizing $\sum_j \text{Var}(T_{1j})$ with respect to p_1 will yield the answer.

As $p_1 \rightarrow 0$ or $p_1 \rightarrow 1$, $\text{Var}(T_{1j})$ approaches infinity. Since $\text{Var}(T_{1j})$ is differentiable in the interval $(0, 1)$, it must reach a minimum in that interval. The minimizing point is given by

$$\left(\frac{p_1}{1-p_1}\right)^2 = \frac{1}{N_d} \left(\frac{p}{1-p}\right)^2 \quad (25)$$

which says that the best odds ratio of inelastic to elastic scattering for the program to use is the natural odds ratio scaled down by the number of bins over which the inelastic scattering is roughly constant.

Equation (25) may be solved explicitly as

$$p_1 = \left(1 + N_d^{1/2} \frac{1-p}{p}\right)^{-1}. \quad (26)$$

For our application, we may estimate N_d as the ratio of the angular resolution of inelastic scattering to elastic scattering squared (for two dimensions). Averaging over different views will increase N_d further, perhaps at the $3/2$ power.

We may estimate typical values of p and p_1 from the differential cross sections given by the EPDL and the KN formulas. At 100 keV, a Legendre expansion up to $\ell = 3$ is sufficient to contain 99 % of the sum of the coefficients vs. $\ell = 15$ for the effective elastic scattering coefficients. Hence, an resolution ratio of $4 = (15 + 1)/(3 + 1)$ is a reasonable estimate. We expect the estimate to apply both in-plane and over angles, so $N_d = 4^3 = 64$ is a reasonable expectation. In the limit in which $1 - p \ll 1$ the fraction of resources going into elastic scattering may be increased by a factor of $N_d^{1/2}$, or 8 in the estimate. Assuming that the sampling requirement is dominated by the requirement of resolving the highest spatial frequency features, the overall number of samples could be reduced by about a factor of 8.

Acknowledgments

This work was supported by NIST Information Technology Laboratory Building the Future funding.

6. References

- [1] Herman GT (1980) *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography* (Academic, New York). <https://doi.org/10.1007/978-1-84628-723-7>
- [2] Rührnschopf EP, Klingenberg K (2011) A general framework and review of scatter correction methods in x-ray cone-beam computerized tomography. Part 2: Scatter estimation approaches. *Medical Physics* 38:5186–5199.
- [3] Nuyt J, De Man B, Fessler JA, Zbijewski W, Beekman FJ (2013) Modelling the physics in the iterative reconstruction for transmission computed tomography. *Physics in Medicine & Biology* 58:R63–R96. <https://doi.org/10.1088/0031-9155/58/12/R63>
- [4] Johns PC, Yaffe M (1982) Scattered radiation in fan beam imaging systems. *Medical Physics* 9:231–239. <https://doi.org/10.1118/1.595076>
- [5] Glover GH (1982) Compton scatter effects in CT reconstructions. *Medical Physics* 9:860–867. <https://doi.org/10.1118/1.595197>
- [6] Siewerdsen JH, Jaffray DA (2001) Cone-beam computed tomography with a flat-panel imager: Magnitude and effects of x-ray scatter. *Medical Physics* 28:220–231. <https://doi.org/10.1118/1.1339879>
- [7] Maslowski A, Wang A, Sun M, Wareing T, Davis I, Star-Lack J (2018) Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter - Part I: Core algorithms and validation. *Medical Physics* 45:1899–1913.
- [8] Stankovic U, Ploeger LS, Herk Mv, Sonke JJ (2017) Optimal combination of anti-scatter grids and software correction for CBCT imaging. *Medical Physics* 44:4437–4451. <https://doi.org/10.1002/mp.12385>
- [9] Mason JH, Nailon WH, Davies ME (2017) A Monte Carlo framework for low dose CT reconstruction testing. *Lecture Notes in Computer Science* 10557:79–88. https://doi.org/10.1007/978-3-319-68127-6_9
- [10] Sossin A, Rebuffel V, Tabary J, Letang JM (2016) Experimental validation of a multi-energy x-ray adapted scatter separation method. *Physics in Medicine & Biology* 61:8625–8639.
- [11] Mainegra-Hing E, Kawrakow I (2010) Variance reduction techniques for fast Monte Carlo CBCT scatter correction calculations. *Physics in Medicine & Biology* 55:4495–4507. <https://doi.org/10.1088/0031-9155/55/16/S05>

- [12] Maier J, Sawall S, Knaup M, Kacheiriess M (2018) Deep scatter estimation (DSE): Accurate real-time scatter estimation for x-ray CT using a deep convolutional neural network. *Journal of Nondestructive Evaluation* 37:57. <https://doi.org/10.1007/s10921-018-0507-z>
- [13] Maier J, Eulig E, Voth T, Knaup M, Kuntz J, Sawall S, Kachelriess M (2019) Real-time scatter estimation for medical ct using the deep scatter estimation: Method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation. *Medical Physics* 46:238–249. <https://doi.org/10.1002/mp.13274>
- [14] Tuy HK (1983) An inversion formula for cone-beam reconstruction. *SIAM Journal on Applied Mathematics* 43:546–552. <https://doi.org/10.1137/0143035>
- [15] Durduran T, Choe R, Baker WB, Yodh AG (2010) Diffuse optics for tissue monitoring and tomography. *Reports on Progress in Physics* 73:076701. <https://doi.org/10.1088/0034-4885/73/7/076701>
- [16] Oh S, Milstein AB, Bouman CA, Webb KJ (2005) A general framework for nonlinear multigrid inversion. *IEEE Transactions on Image Processing* 14:125–140. <https://doi.org/10.1109/TIP.2004.837555>
- [17] Poludniowski G, Evans PM, Webb S (2009) Rayleigh scatter in kilovolt x-ray imaging: Is the independent atom approximation good enough? *Physics in Medicine & Biology* 54:6931–6942. <https://doi.org/10.1088/0031-9155/54/22/012>
- [18] Hubbell JH, Veigele WJ, Briggs EA, Brown RT, Cromer DT, Howerton RJ (1975) Atomic form factors, incoherent scattering functions, and photon cross sections. *Journal of Physical and Chemical Reference Data* 4:471–538. <https://doi.org/10.1063/1.555523>
- [19] Rehr JJ (2000) Theoretical approaches to x-ray absorption fine structure. *Reviews of Modern Physics* 72:621–654. <https://doi.org/10.1103/RevModPhys.72.621>
- [20] Goudsmit S, Saunderson JL (1940) Multiple scattering of electrons. *Physical Review* 57:24–29. <https://doi.org/10.1103/PhysRev.57.24>
- [21] Berger MJ, Hubbell JH, Seltzer SM, Chang J, Coursey JS, Sukumar R, Zucker DS, Olsen K (1998) XCOM: photon cross sections database. Standard Reference Database 8 (XGAM), National Institute of Standards and Technology. <https://doi.org/10.18434/T48G6X>
- [22] Cullen DE, Hubbell JH, Kissel L (1997) Epdl97: The evaluated photon data library: '97 version (Lawrence Livermore National Laboratory, Livermore, CA), UCRL-LR-50400, Vol. 6, Rev. 5. Available at <https://www.osti.gov/servlets/purl/295438>.
- [23] Poludniowski G, Evans PM, Hansen NV, Webb S (2009) An efficient Monte Carlo-based algorithm for scatter correction in keV cone-beam CT. *Physics in Medicine & Biology* 54:3847–3864. <https://doi.org/10.1088/0031-9155/54/12/016>
- [24] Renken JH (1967) Legendre polynomial expansion for the Klein-Nishina formula. *Journal of Applied Physics* 38:4925–4927. [https://doi.org/10.1016/0021-9991\(74\)90003-5](https://doi.org/10.1016/0021-9991(74)90003-5)
- [25] Waller I, Hartree DR (1929) On the intensity of total scattering of x-rays. *Proceedings of the Royal Society of London* 124:119–142. <https://doi.org/10.1098/rspa.1929.0101>
- [26] Hernandez AM, Boone JM (2014) Tungsten anode spectral model using interpolating cubic splines: Unfiltered x-ray spectra from 20 kV to 640 kV. *Medical Physics* 41:042101. <https://doi.org/10.1118/1.4866216>
- [27] Lamperti PJ, O'Brien M (2013) NIST measurement services: Calibration of x-ray and gamma-ray measuring instruments (National Institute of Standards and Technology, Gaithersburg, MD), NIST Special Publication (SP) 250-58. <https://doi.org/10.6028/nist.sp.250-58>
- [28] Press WH, Teukolsky SA, Vetterberg WT, Flannery BP (2007) *Numerical Recipes* (Cambridge University, New York) 3rd Ed. Sec. 13.3.
- [29] Golub GH, Hansen PC, O'Leary DP (1999) Tikhonov regularization and total least squares. *SIAM Journal on Matrix Analysis & Applications* 21:185–194. <https://doi.org/10.1137/S0895479897326432>
- [30] Bouman C, Sauer K (1993) A generalized gaussian image model for edge-preserving map estimation. *IEEE Transactions on Image Processing* 2:296–310. <https://doi.org/10.1109/83.236536>
- [31] Levine ZH, Pintar AL (2015) A fixed-memory moving, expanding window for obtaining scatter corrections in x-ray CT and other stochastic averages. *Computer Physics Communications* 196:455–459. <https://doi.org/10.1016/j.cpc.2015.05.019>
- [32] Levine Z (2018) The Walker Alias Method in modern Fortran. National Institute of Standards and Technology. <https://doi.org/10.18434/T4/1432807>
- [33] Thibault JB, Sauer KD, Bouman CA, Hsieh J (2007) A three-dimensional statistical approach to improved image quality for multislice helical CT. *Medical Physics* 34:4526–4544. <https://doi.org/10.1118/1.2789499>
- [34] Zhu C, Byrd RH, Nocedal J (1997) Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software* 23:550–560. <https://doi.org/10.1145/279232.279236>
- [35] Walker AJ (1977) An efficient method for generating discrete random variables with general distributions. *ACM Transactions on Mathematical Software* 3:253–256. <https://doi.org/10.1145/355744.355749>
- [36] Kahn H (1955) *Use of Different Monte Carlo Sampling Techniques* (Rand Corporation, Santa Monica, CA) p 766.
- [37] Pirene MH (1946) *The Diffraction of X-rays and Electrons by Free Molecules* (Cambridge University Press, Cambridge, UK).
- [38] Fisher RA (1941) *Statistical Methods for Research Workers* (G. E. Stechert & Co.) p 82.
- [39] Horner M, Genc KO, Luke SM, Pietila TM, Cotton RT, Ache BA, Levine ZH, Townsend KC (2016) Towards estimating the uncertainty associated with 3D geometry reconstructed from medical image data. *2016 BMES/FDA Frontiers in Medical Devices*

Conference, (BMES/FDA, College Park, MD). Available at

<https://www.nist.gov/publications/towards-estimating-uncertainty-associated-3d-geometry-reconstructions-medical-image>

[40] Weast RC (1981) *Handbook of Chemistry and Physics* (Chemical Rubber Company, Cleveland, OH) 61st Ed. pp F–1.

[41] Cuffey K, Patterson WSB (2010) *The Physics of Glaciers* (Academic) 4th Ed. Sec. 2.4.

[42] Efron B, Tibshirani RJ (1993) *An Introduction to the Bootstrap* (Chapman & Hall, Boca Raton, FL) p 170ff.

About the authors: Zachary Levine is a physicist in the Quantum Measurement Division at NIST and a co-inventor on two tomography-related US patents, 6,389,101 and 7,967,507. Timothy Blattner and Adele Peskin are computer scientists in the Software and Systems Division at NIST. Adam Pintar is a statistician in the Statistical Engineering Division at NIST. The National Institute of Standards and Technology is an agency of the U.S. Department of Commerce.