Near optimal energy selective x-ray imaging system performance with simple detectors

Robert E. Alvarez^{a)}

Aprend Technology, Mountain View, California 94043

(Received 23 June 2009; revised 6 November 2009; accepted for publication 6 December 2009; published 27 January 2010)

Purpose: This article describes a method to achieve near optimal performance with low energy resolution detectors. Tapiovaara and Wagner [Phys. Med. Biol. **30**, 519–529 (1985)] showed that an energy selective x-ray system using a broad spectrum source can produce images with a larger signal to noise ratio (SNR) than conventional systems using energy integrating or photon counting detectors. They showed that there is an upper limit to the SNR and that it can be achieved by measuring full spectrum information and then using an optimal energy dependent weighting.

Methods: A performance measure is derived by applying statistical detection theory to an abstract vector space of the line integrals of the basis set coefficients of the two function approximation to the x-ray attenuation coefficient. The approach produces optimal results that utilize all the available energy dependent data. The method can be used with any energy selective detector and is applied not only to detectors using pulse height analysis (PHA) but also to a detector that simultaneously measures the total photon number and integrated energy, as discussed by Roessl *et al.* [Med. Phys. **34**, 959–966 (2007)]. A generalization of this detector that improves the performance is introduced. A method is described to compute images with the optimal SNR using projections in a “whitened” vector space transformed so the noise is uncorrelated and has unit variance in both coordinates. Material canceled images with optimal SNR can also be computed by projections in this space.

Results: The performance measure is validated by showing that it provides the Tapiovaara–Wagner optimal results for a detector with full energy information and also a conventional detector. The performance with different types of detectors is compared to the ideal SNR as a function of x-ray tube voltage and subject thickness. A detector that combines two bin PHA with a simultaneous measurement of integrated photon energy provides near ideal performance across a wide range of operating conditions.

Conclusions: Low energy resolution detectors can be used in energy selective x-ray imaging systems to produce images with near optimal performance. © 2010 American Association of Physicists in Medicine. [DOI: [10.1118/1.3284538](https://doi.org/10.1118/1.3284538)]

Key words: dual energy, energy selective, x-ray spectra, basis set, statistical detection theory, pulse height analysis, NQ detector

I. INTRODUCTION

The classic paper of Tapiovaara and Wagner¹ showed that a broad spectrum x-ray imaging system with optimal energy dependent weighting could provide improved signal to noise ratio (SNR) compared to a system using conventional energy integrating or photon counting detectors. As they stated, “If the energy spectrum of the signal to be detected differs from the energy spectrum of the background, a detector that only counts photons will lose some of the information available; the ideal detector must also consider the energy of the photons.” Later, Cahn *et al.*² and Giersch *et al.*³ showed that the optimal spectral weighting could be implemented using photon counting detectors with high energy resolution pulse height analysis (PHA). In addition, they showed simulations that demonstrated substantial improvement in the SNR by using the optimal spectral weights. Niederlohn *et al.*⁴ then showed that an improvement in SNR could be obtained using PHA with a small number of bins.

This paper shows a method to produce near Tapiovaara–

Wagner optimal SNR images with a variety of low energy resolution detectors. The method uses the representation of the attenuation coefficients with a two-dimensional linear space, followed by a solution of the nonlinear equations for the line integrals of the basis set coefficients as described in Alvarez–Macovski.⁵ The method then transforms the line integrals of the coefficients, considered to be points in an abstract vector space, to a “whitened” space with uncorrelated data and unit variance of each coordinate.

The results are derived by applying statistical detection theory^{6–8} with the abstract vector space data. The detection theory approach results in an optimal test that utilizes all of the energy dependent information. The approach is validated by showing that it predicts the same SNR with a detector with full energy resolution as the ideal Tapiovaara–Wagner result. In addition, it also gives the same results as Tapiovaara–Wagner with a conventional photon counting detector. The detection theory approach can, however, be used with any energy selective detector. It is applied to detectors with PHA and to detectors, as suggested by Roessl *et al.*⁹ and

Kraft *et al.*¹⁰ that simultaneously measure the total number of photons and their total energy. These are called NQ detectors in this paper, where N denotes the total number of photons and Q denotes their total energy. A generalization of this concept is a detector that combines PHA with measurement of the total energy of all the photons, called NKQ detectors, where K is the number of bins. This is shown to improve the performance substantially for higher tube voltage spectra.

The SNR of these low energy resolution detectors is compared to the ideal result as a function of the x-ray tube voltage and patient thickness. The NQ detector with no further energy resolution can provide a large fraction of optimal SNR with low x-ray tube voltage spectra. At higher voltages, the $N2Q$ detector with a single Q measurement and two bin PHA provides near ideal performance over a wide range of voltages and body thicknesses. Further, the optimal bin separation energy is easily computable by the system from the tube voltage. It is almost constant with variations in body thickness so it provides good results over a wide range of operating conditions. The $N2Q$ detectors are simpler to implement than high resolution PHA detectors and as a result have potential benefits such as higher count rate capability and lower cost.

A method to compute an image analogous to a conventional image but with optimal SNR is described. The method uses the whitened abstract vector space data. It forms a dot product of the data with a unit length vector in the direction from the mean of the background to the object data. The method is similar to the generalized projections discussed in Lehmann *et al.*¹¹ but operates in the whitened space that is optimized for noise. A similar method is described to produce improved SNR material canceled images. Simulated results with a Monte Carlo simulation are shown for both methods.

Although the results in this paper are derived for a projection imaging system, Appendix A shows that the performance parameter for a computed tomography (CT) system near the center of a large uniform object is a constant, which is independent of the energy spectral properties of the detector, times the value for the projection system. Therefore, the conclusions about the relative performance of different types of detectors are also applicable, in general, to CT.

II. METHODS

II.A. The Tapiovaara–Wagner optimal SNR

Tapiovaara and Wagner¹ modeled an x-ray imaging system as two parts: Detection and display. They then modeled the detection in three steps: (1) Detecting x-ray photons of different energies to produce signals proportional to their energy, (2) coding or weighting of the detected signals in a form more suitable for the image recorder, and (3) recording the information. With monoenergetic spectra, the weighting does not affect the results since all photons have the same energy. However, with broad spectrum x-ray data, the weighting does affect the results and can be optimized depending on the imaging task.

The imaging task they analyzed was a two-alternative forced choice, where the detector makes measurements of a field containing two areas, one with a signal and one without, and the task is to decide the area with the signal. When the signal is a photon number spectrum modeled as having Poisson statistics, they showed that the performance for the forced choice task depends on the signal to noise ratio squared

$$\text{SNR}^2 = A \frac{[\int (\overline{N_1(E)} - \overline{N_2(E)}) D(E) w(E) dE]^2}{\int (\overline{N_1(E)} + \overline{N_2(E)}) D(E) w^2(E) dE}, \quad (1)$$

where A is the signal area, $\overline{N_1(E)}$ and $\overline{N_2(E)}$ are the mean signals in the two areas of the image at x-ray energy E , $D(E)$ is the fraction of the photons detected, and $w(E)$ is the weighting or coding function. They then showed that for an ideal detector that maximizes SNR^2 , $D(E)=1$ and

$$w(E) = \frac{\overline{N_1(E)} - \overline{N_2(E)}}{\overline{N_1(E)} + \overline{N_2(E)}}. \quad (2)$$

With these parameters, the optimal SNR^2 is

$$\text{SNR}_{\text{ideal}}^2 = A \int \frac{(\overline{N_1(E)} - \overline{N_2(E)})^2}{\overline{N_1(E)} + \overline{N_2(E)}} dE. \quad (3)$$

Cahn *et al.*² then showed that for low contrast objects the optimal weighting function (2) is proportional to the difference in the attenuation coefficient functions of the object in the signal and the background areas

$$w(E) \sim \mu_1(E) - \mu_2(E),$$

and that, with this weighting, the SNR^2 is improved over either photon energy integrating or photon counting detectors. They commented that the weighting could be implemented in a system with high energy resolution PHA. Giersch *et al.*³ then described detailed Monte Carlo simulations that showed that the improved SNR^2 could be achieved in such a system.

The Tapiovaara–Wagner formulation does not take into account the low dimensionality of the attenuation coefficients of body materials. The following sections show that an approach that uses this fact can achieve near optimal SNR^2 with low energy resolution measurements.

II.B. Vector space descriptions of attenuation coefficients and line integrals

The vector space description of attenuation coefficients and their line integrals are widely used and will be described here mainly to introduce the notation that will be used in the rest of the paper.

We can accurately model the attenuation coefficient $\mu(E)$ as a linear combination of two functions of x-ray photon energy E

$$\mu(E) = a_1 f_1(E) + a_2 f_2(E). \quad (4)$$

The basis functions $f_i(E)$ can model the energy dependence of the mechanisms of interactions of x rays and matter⁵ or can be the attenuation coefficients of two chemically distinct

FIG. 1. Vector interpretation of the basis set coefficients of a mixture. Part (a) shows the total vector for a mixture with three components. See text for the description of the individual component vector lengths. Part (b) shows the line integral vector of an object consisting of two materials with thicknesses t_1 and t_2 .

materials.¹² The resultant coefficients $\mathbf{a} = [a_1 \ a_2]^T$ can be considered to be points in an abstract vector space, which will be abbreviated as $\mathbf{a}$ space. Note that vectors are represented as columns and the symbol T denotes a transpose. The linearity of the space results from the mixture rule (see Ref. 13 Chap. 25), which states that the mass attenuation coefficient of a mixture is related to the attenuation coefficients of its constituents by

$$\frac{\mu}{\rho} = \sum_{i=1}^N W_i \frac{\mu_i}{\rho_i}, \quad (5)$$

where ρ is the overall density, W_i is the fraction by weight of element i in the mixture, and μ_i/ρ_i is its mass attenuation coefficient. If the $\mathbf{a}$ vectors of the constituents of the material are

$$\mathbf{a}_{\text{material},i} = [a_{1i} \ a_{2i}]^T,$$

then the total vector has components

$$a_1 = \sum_{i=1}^N W_i \frac{\rho}{\rho_i} a_{1i}, \quad (6)$$

$$a_2 = \sum_{i=1}^N W_i \frac{\rho}{\rho_i} a_{2i}. \quad (7)$$

This can be considered to be a vector sum

$$\mathbf{a}_{\text{total}} = W_1 \frac{\rho}{\rho_1} \mathbf{a}_{\text{material},1} + \cdots + W_N \frac{\rho}{\rho_N} \mathbf{a}_{\text{material},N}$$

and has the graphical interpretation shown in Fig. 1(a).

The line integral of an inhomogeneous object is

$$\mathcal{L}(E) = \int_S \mu(\mathbf{r}, E) ds, \quad (8)$$

where $\mathbf{r}$ is the position vector of a point within the object. Introducing the vector space decomposition of the attenuation coefficient, the line integral is

FIG. 2. Geometry for hypothesis testing. The background is a slab of material with $\mathbf{a}$ vector $[0 \ 1]^T$ and thickness t_b . The feature has $\mathbf{a}$ vector $[1 \ 0]^T$ and thickness t_f . The feature is either present or not present under the two hypotheses.

$$\mathcal{L}(E) = A_1 f_1(E) + A_2 f_2(E), \quad (9)$$

where $A_j = \int_S a_j(\mathbf{r}) ds$ $j=1,2$ are the line integrals of the $\mathbf{a}$ vector components. The line integrals can also be considered as points in a linear vector space, the $\mathbf{A}$ space. If the object is composed of a single material with coefficient vector $\mathbf{a}_{\text{material}}$ and has a thickness t , its line integral vector will be $\mathbf{A} = [A_1 \ A_2]^T = t \mathbf{a}_{\text{material}}$. If the object is composed of two slabs of material with thicknesses t_1 and t_2 , then, by the linearity of Eq. (8), its line integral vector will be $\mathbf{A} = t_1 \mathbf{a}_{\text{material},1} + t_2 \mathbf{a}_{\text{material},2}$, which has the graphical representation as a vector sum shown in Fig. 1(b).

With the vector space description, energy selective imaging systems can be considered to produce images of vector quantities. Single projection systems provide images of $\mathbf{A}$ vectors, while energy selective CT systems reconstruct images of $\mathbf{a}$ vectors in the cross section of the object from sinograms of $\mathbf{A}$ vectors.

II.C. Statistical detection theory applied to energy selective x-ray imaging

Statistical detection theory^{6,8} is based on hypothesis testing. To apply it to energy selective imaging systems, we can define the task as deciding whether a feature object, such as a tumor embedded in surrounding “background” tissue, is present (see Fig. 2). The two hypotheses are then: H_0 , the feature is not present, and H_1 , the feature is present. The theory derives optimal tests to decide which hypothesis is true based on the measured data, which in an energy selective projection system are the $\mathbf{A}$ vectors.

Since we are free to choose any suitable basis set for the expansion in Eq. (4), the comparison of the results with prior research, such as Tapiovaara–Wagner, will be facilitated by using the attenuation coefficients of the feature and background materials as the basis functions. We can use the attenuation coefficients of any two distinct materials as the basis set¹² and Roth¹⁴ found that attenuation coefficient

space is quite “flat” so the accuracy of the two function approximation is reasonably independent of the materials used. The analysis presented here is valid for any suitable set of basis functions and indeed can be generalized to more than two basis functions when required such as with high atomic number contrast materials. Using the feature/background basis set, the coefficients of the two materials in Fig. 2 are dimensionless vectors

$$\mathbf{a}_f = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{a}_b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

As will be discussed in subsequent sections, the basis set coefficient line integrals are modeled as multivariate normal random variables with different parameters under the two hypotheses. Under hypothesis H_0 , the expected value is $\mathbf{A}_0 = t_b \mathbf{a}_b$ and the covariance is C_0 , while under hypothesis H_1 , the expected value is $\mathbf{A}_1 = \mathbf{a}_b(t_b - t_f) + \mathbf{a}_f t_f$ and the covariance is C_1 .

There are two widely used approaches to deriving optimal detectors. The Bayes criterion minimizes the risk, that is, the expected value of the cost. The risk depends on costs that the user assigns to each course of action, the *a priori* probability of the occurrence of each hypothetical case, and on the probabilities of correct or incorrect decisions. The Neyman–Pearson criterion maximizes the true positive rate for a fixed false positive rate and does not depend on *a priori* probability or assigned costs. With both approaches, the optimal decision rule is computed by forming the likelihood ratio⁶ and comparing it against a threshold. The likelihood ratio $\Lambda(\mathbf{A})$ is the probability of the measured data given that hypothesis H_1 is true divided by the probability of the data given that H_0 is true,

$$\Lambda(\mathbf{A}) = \frac{p_{A|H_1}(\mathbf{A}|H_1)}{p_{A|H_0}(\mathbf{A}|H_0)}.$$

The decision rule is to decide on hypothesis H_1 if the ratio is greater than the threshold and H_0 if the ratio is less. The likelihood ratio is a function of the data so it is itself a random variable and its probability distribution function determines the performance of the estimator. In many cases it is simpler to use the logarithm of the likelihood ratio, called the log-likelihood ratio,

$$\begin{aligned} l(\mathbf{A}) &= \log[\Lambda(\mathbf{A})] \\ &= \log[p_{A|H_1}(\mathbf{A}|H_1)] - \log[p_{A|H_0}(\mathbf{A}|H_0)]. \end{aligned} \quad (10)$$

Since the logarithm is a one-to-one function, this is equivalent to the original test.

Assuming that A is bivariate normal random variable (see justification in the following section), the probability density function $p_{A|H_0}(\mathbf{A}|H_0)$ can be written as

$$p_{A|H_0}(\mathbf{A}|H_0) = \frac{1}{2\pi\sqrt{|\mathbf{R}_{A|H_0}|}} e^{-(\mathbf{A} - \overline{\mathbf{A}}_0)^T \mathbf{R}_{A|H_0}^{-1} (\mathbf{A} - \overline{\mathbf{A}}_0)},$$

where $\mathbf{R}_{A|H_0}$ is the covariance function and $\overline{\mathbf{A}}_0$ is the mean value under hypothesis H_0 . $|\mathbf{R}|$ denotes the determinant of

$\mathbf{R}$. There is a similar expression for $p_{A|H_1}(\mathbf{A}|H_1)$. Van Trees,⁶ Sec. 2.6, showed that, for a normal distribution, the log-likelihood ratio in the general case is the difference between two quadratic forms. If the covariances are the same under both hypotheses $\mathbf{R}_{A|H_0} = \mathbf{R}_{A|H_1} = \mathbf{R}_A$, then the log-likelihood function is a scalar normal random variable

$$l(\mathbf{A}) = (\delta\mathbf{A})^T \mathbf{R}_A^{-1} \mathbf{A},$$

where $\delta\mathbf{A} = \overline{\mathbf{A}}_1 - \overline{\mathbf{A}}_0$ is the difference in the mean values with the two hypotheses. In the equal covariance case, Van Trees showed that the performance of the test is completely determined by the generalized signal to noise ratio parameter (also known as the deflection coefficient)

$$d^2 = (\delta\mathbf{A})^T \mathbf{R}_A^{-1} \delta\mathbf{A}. \quad (11)$$

For application to energy-selective systems, the difference in the mean value vectors for the two hypotheses in Eq. (11) is

$$\delta\mathbf{A} = \overline{\mathbf{A}}_1 - \overline{\mathbf{A}}_0 = t_f(\mathbf{a}_f - \mathbf{a}_b) = t_f \begin{bmatrix} 1 \\ -1 \end{bmatrix}. \quad (12)$$

The d^2 parameter is a scalar quantity that summarizes the performance for equal covariance matrices based on all the available data. We can gain an intuitive understanding of its characteristics by considering the case of a scalar measurement. In that case, $R_A^{-1} = 1/\sigma_A^2$ and $(\delta A)^T \delta A = \delta A^2$ so $d^2 = \delta A^2/\sigma_A^2$, which is the classical definition of signal to noise ratio squared. The d^2 parameter can be used to compute the ROC curve and its area $\text{AUC} = \frac{1}{2}(1 + \text{erf}(d/2))$, where $\text{erf}(x)$ is the error function. Note that since $\text{erf}(x)$ is a monotonically increasing function, as d increases, the area increases and the detector performance improves.¹⁵

In x-ray imaging, equal covariance matrices imply a low contrast object where the difference in photon number flux between the two regions is small. While this is a somewhat restrictive assumption, it is widely used in analysis of x-ray imaging systems. Indeed, the equal covariance case is a widely used approximation in general signal and image processing because it provides useful insight to the general case where the covariance does vary. The usefulness of this approximation is noted in a standard textbook on statistical detection theory,⁸ “It should be emphasized that only for the mean-shifted Gauss–Gauss problem does the deflection coefficient completely characterize the detection performance. However, it is often used for approximate detection performance evaluation for other detection problems.” The statistical detection theory approach is still applicable if the low contrast assumption is not valid. In this case, the difference in quadratic forms approach discussed in Ref. 6 can be used to analyze the performance.

The formulation in this section is in terms of line integrals and is directly applicable to projection imaging systems. Appendix A shows that the d^2 parameter for a CT system near the center of a uniform circular object with equal incident exposure for all projections is proportional to the d^2 parameter for the projections

$$d_{CT}^2 = \frac{N_{\text{angles}}^2 w_{1/2}^3 \rho^2}{6.8 t_f^2} d_{\text{Proj}}^2.$$

The proportionality constant is independent of the energy spectral properties of the detector. Thus, the general conclusions based on a projection system about relative performance can also be applied to CT although specific results will have to be derived for particular objects and CT system parameters.

II.D. Noise in energy selective systems

The evaluation of the d^2 parameter requires the inverse covariance matrix of the A -space data, R_A^{-1} . This is determined by the x-ray source, the object, and the properties of the detector. In this section, I show that $R_A^{-1} = M^T R_L^{-1} M$, where M is a matrix of the effective attenuation coefficients of the basis materials for each spectrum. The matrix R_L is the covariance of the logarithm of the measured data for all spectra. The derivation is based on a linear approximation of the deviations of the data with noise from the mean. The approximation is valid when the standard deviation of the noise divided by the mean is small, which is required in medical systems to produce diagnostically useful images.

The measurement I with an x-ray system is

$$I = \int s_0(E) d(E) g(E) e^{-\int s \mu(r, E) ds} dE. \quad (13)$$

where the spectrum $s_0(E)$ can be either the number of photons $n(E)$ in the range E to $E+dE$ or their total energy $q(E) = En(E)$ depending on the type of detector. Detectors are not able to absorb and respond to all photons incident on them so the response will be modeled by an absorption fraction $d(E)$, which is a function of energy. In addition, the detector energy resolution will be described by the response function $g(E)$. The detector area will be factored into the definition of the x-ray source spectrum, $s_0(E)$. In order to separate the energy dependent performance from the resolution, the area will be taken to be large enough to average over the source and the detector resolving area.

Introducing the vector space description of the attenuation coefficient into the line integral (9), the measurement (13) is seen to be a function of the A vector,

$$I_k(A_1, A_2) = \int s_{k0}(E) e^{-A_1 f_1(E) - A_2 f_2(E)} dE. \quad (14)$$

An energy selective system uses measurements $I_k(A_1, A_2)$ with K different effective spectra to compute the A vector at each point in the image by inverting the system of Eq. (14). The spectra $s_{k0}(E)$, $k=1, K$ can be different source spectra, as with voltage switching, or effective spectra with energy resolving detectors, or both.

Since medical systems require low noise data with relatively small standard deviations, we can analyze the noise with a linear approximation to Eq. (14). Because of the inverse exponential relationship of the measured data and the line integrals, a vector Taylor's series of the logarithm of the measurements about the mean value will be used,

$$\begin{aligned} L(A_1, A_2) &= \begin{bmatrix} -\log[I_1(A_1, A_2)] \\ \vdots \\ -\log[I_K(A_1, A_2)] \end{bmatrix} \\ &= L(\bar{\mathbf{A}} + \delta \mathbf{A}) \\ &\approx L(\bar{\mathbf{A}}) + \frac{\partial L}{\partial \mathbf{A}} \delta \mathbf{A}, \end{aligned} \quad (15)$$

where $\mathbf{A} = [A_{1i} \ A_{2i}]^T$ and $\bar{\mathbf{A}}$ is the mean value vector in A space. Defining

$$M = \frac{\partial L}{\partial \mathbf{A}} = \begin{bmatrix} \frac{\partial L_1}{\partial A_1} & \frac{\partial L_1}{\partial A_2} \\ \vdots & \vdots \\ \frac{\partial L_K}{\partial A_1} & \frac{\partial L_K}{\partial A_2} \end{bmatrix}, \quad (16)$$

we can express the Taylor's series expansion as

$$\delta L = L(\bar{\mathbf{A}} + \delta \mathbf{A}) - L(\bar{\mathbf{A}}) = M \delta \mathbf{A}. \quad (17)$$

The coefficients of the M matrix are

$$\begin{aligned} M_{ij} &= -\frac{\partial \log(I_i)}{\partial A_j} = -\frac{1}{I_i} \frac{\partial I_i}{\partial A_j}, \\ M_{ij} &= \frac{\int f_j(E) s_i(E) e^{-A_1 f_1(E) - A_2 f_2(E)} dE}{\int s_i(E) e^{-A_1 f_1(E) - A_2 f_2(E)} dE}. \end{aligned}$$

Defining the normalized spectrum

$$\begin{aligned} \hat{s}_i(E) &= \frac{s_i(E) e^{-A_1 f_1(E) - A_2 f_2(E)}}{\int s_i(E) e^{-A_1 f_1(E) - A_2 f_2(E)} dE}, \\ M_{ij} &= \int f_j(E) \hat{s}_i(E) dE = \langle f_j(E) \rangle_i. \end{aligned} \quad (18)$$

That is, the coefficients of the M matrix are the average or effective values of the basis set functions over the spectra $\hat{s}_i(E)$ transmitted through the object.

The measured data vector L will be assumed to be a multivariate normal with covariance R_L . This assumption is examined in detail in the next section. With the linear approximation about the mean (15), the estimated line integrals are a linear combination of the measured data, which are normal random variables, so the estimates are themselves normal random variables. The inverse covariance required to compute the d^2 parameter is then [see Ref. 7, Eq. (4.26)]

$$R_A^{-1} = M^T R_L^{-1} M. \quad (19)$$

Substituting in the general expression for d^2 (11),

$$d^2 = (\delta \mathbf{A})^T M^T R_L^{-1} M (\delta \mathbf{A}). \quad (20)$$

II.E. Noise in photon counting and energy integrating detectors

The calculation of the d^2 parameter requires the covariance matrix R_L of the measured data. This matrix depends on

FIG. 3. Idealized models of the (a) NQ and (b) NKQ detectors. The signal from the sensing element that converts x-ray photons to electrical impulses is split into two circuits. In one, the impulses are integrated during the exposure time to produce the Q signal. In the second circuit, the number of photons is counted with a digital counter in the NQ detector or the impulses of different amplitudes are separated into K bins with a pulse height analyzer in the NKQ detector. See Ref. 19 for a detailed description of the NQ detector design.

the type of detector and this section introduces models for the noise with photon counting, energy integrating, NQ , and the NKQ detector.

Based on reasonable assumptions about the random properties of the arrival times of photons, the photon number measurement can be modeled as a Poisson random variable¹⁶ N with mean value parameter $\lambda=I$ from Eq. (14) with a photon number spectrum $s_0(E)=n(E)$. The energies of the photons can be modeled as a random variable E that describes the energy of each detected photon.^{17,18} The total energy Q is the sum of the random energies of a random number N of photons

$$Q = \sum_{n=1}^N E_n,$$

which is a compound Poisson random variable.¹⁶

In general, if measurements are made with different detectors, or with the same detector at different times or with nonoverlapping spectra, they are statistically independent. However, as discussed by Roessl *et al.*,⁹ the measurements from NQ detectors are correlated. This is evident when we consider the idealized model of the detector shown in Fig. 3(a). Based on this model, it is clear that as the number of photons N increases, we would expect their integrated energy

Q also to increase. The design of detectors is beyond the scope of this paper but see Ref. 19 for a detailed description of the NQ detector design.

X-ray imaging systems commonly utilize the logarithm of the data so the joint statistics of the logarithms of N and Q are derived in Appendix B. The result is that the mean values are

$$E_x \begin{bmatrix} \log(N) \\ \log(Q) \end{bmatrix} = \begin{bmatrix} \log(\lambda) \\ \log(\lambda \langle E \rangle) \end{bmatrix},$$

and the covariance matrix is

$$\text{Cov}(\log(N), \log(Q)) = \frac{1}{\lambda} \begin{bmatrix} 1 & 1 \\ 1 & F \end{bmatrix}. \quad (21)$$

The excess variance F in Eq. (21) is the ratio of the second moment $\langle E^2 \rangle$ of the x-ray energy E and the square of the first moment $\langle E \rangle$

$$F = \frac{\langle E^2 \rangle}{\langle E \rangle^2}. \quad (22)$$

The variance of E ,

$$\sigma_E^2 = \langle E^2 \rangle - \langle E \rangle^2 \geq 0 \quad (23)$$

is greater than or equal to zero and is equal to zero only for monoenergetic spectra with zero width. Dividing Eq. (23) by $\langle E \rangle^2$ shows that $F \geq 1$ with equality only for zero width spectra.

Figure 3(b) shows a generalization of the NQ detector, where, instead of simply counting the number of photons, a pulse height analyzer is used to count the number of photons whose energies fall in K nonoverlapping energy bins and simultaneously a single measurement of the overall integrated photon energy is made. This is the NKQ detector, where K is the number of bins. A detector with only K bin PHA and no Q measurement will be referred to as NK . Appendix B shows that the NKQ photon counts $\{N_k, k=1, \dots, K\}$ are independent of each other but each is positively correlated with the integrated photon energy Q . The mean value vector of the logarithm of the measurements is

$$E_x \begin{bmatrix} \log(N_1) \\ \log(N_2) \\ \vdots \\ \log(N_K) \\ \log(Q) \end{bmatrix} = \begin{bmatrix} \log(\lambda_1) \\ \log(\lambda_2) \\ \vdots \\ \log(\lambda_K) \\ \log(\lambda \langle E \rangle) \end{bmatrix}, \quad (24)$$

and the covariance matrix is

$$\begin{aligned} & \text{Cov}(\{\log(N_k), k=1, \dots, K\}, \log(Q)) \\ &= \frac{1}{\lambda} \begin{bmatrix} \lambda/\lambda_1 & 0 & \dots & 0 & \frac{\langle E \rangle_1}{\langle E \rangle} \\ 0 & \lambda/\lambda_2 & 0 & 0 & \frac{\langle E \rangle_2}{\langle E \rangle} \\ 0 & 0 & \dots & 0 & \vdots \\ 0 & 0 & 0 & \lambda/\lambda_K & \frac{\langle E \rangle_K}{\langle E \rangle} \\ \frac{\langle E \rangle_1}{\langle E \rangle} & \frac{\langle E \rangle_2}{\langle E \rangle} & \dots & \frac{\langle E \rangle_K}{\langle E \rangle} & F \end{bmatrix}. \end{aligned} \quad (25)$$

In Eq. (25), λ is the mean of the total number of photons, $\langle E \rangle$ is the first moment of the complete energy spectrum, F is as defined above, λ_k is the mean value of the number of photons, and $\langle E \rangle_k$ is the first moment of the effective spectrum of object k . Note that

$$\lambda = \sum_{k=1}^K \lambda_k.$$

The probability distribution functions of the measured data will be approximated by multivariate normal random variables with the mean and covariance given by these equations. Measurements in x-ray imaging systems are, of course, substantially more complex than these models and electronic noise, scatter, prepatient bow-tie filters and other factors affect the statistics.²⁰ Nevertheless, the normal distribution is widely used and has been experimentally validated. For example, Wang *et al.*²¹ used the Shapiro–Wilk test²² to show that the distribution of prereconstruction CT sinogram data is described by a normal distribution with high confidence. The Shapiro–Wilk test is based on the normal probability plot. This is a plot with nonlinear coordinate scales chosen so that data with a normal distribution fall on a straight line. The test derives a measure of the closeness to a straight line and is considered the most reliable indicator of whether a data set is normally distributed. Wang *et al.* used CT data with conventional energy integrating detectors, which, as shown above, can be modeled by a compound Poisson random variable.²³ This approaches a normal distribution for moderately large counts so the random data will be modeled as multivariate normal. The Shapiro–Wilk test also shows that the normal approximation can be applied to photon counting detectors for medical imaging applications. Applying the test shows that a normal approximation to a Poisson random variable is acceptable with a confidence level of better than 0.001 when the average number of counts is greater than 250. This is much smaller than the average photon number of data in medical imaging systems. Thus, we can apply the normal approximation using the mean and covariance derived in this section both with energy integrating, photon counting detectors and to detectors based on them such as the NQ and NKQ .

II.F. Performance of energy selective detectors

This section uses the statistical detection theory approach to derive formulas for the performance of different types of energy selective detectors. The results are compared to each other and to the Tapiovaara–Wagner optimal SNR².

II.F.1. Conventional counting detector

The first application of the statistical detection theory approach is to a conventional, that is, nonenergy selective counting detector. There is only one spectrum so the M matrix has a single row with elements computed using Eq. (18),

$$M = [\langle \mu_f(E) \rangle_N \langle \mu_b(E) \rangle_N]. \quad (26)$$

The notation $\langle \mu(E) \rangle_N$ indicates an average over the normalized photon number spectrum

$$\langle \mu(E) \rangle_N = \int \mu(E) \hat{n}(E) dE. \quad (27)$$

The spectrum transmitted through the object is $n_T(E) = n(E)e^{-t_b \mu_b(E)}$ so

$$\hat{n}(E) = \frac{n_T(E)}{\int n_T(E) dE}, \quad (28)$$

where the integral in the denominator is over the region, where $n_T(E)$ is nonzero so with PHA it would be over the energy bin. Recall that we use the background and feature material attenuation coefficients as basis functions so $\mu_b(E)$ and $\mu_f(E)$ are the attenuation coefficients of the background and feature materials.

Using the general results for the logarithm of a Poisson random variable (21),

$$R_L = \frac{1}{\lambda/2},$$

where $\lambda = \int n_T(E) dE$ and $\lambda/2$ is the average number of photons per measurement. A factor of 2 is used for ease of comparison with Tapiovaara–Wagner who assumed two measurements in the definition of SNR. Substituting R_L in the formula for the inverse of the covariance matrix (19)

$$R_A^{-1} = \frac{\lambda}{2} M^T M.$$

Substituting in Eq. (11) with $M = [M_{11} \ M_{12}]$, d^2 for the conventional photon counting detector is

$$\begin{aligned} d_{\text{conv},N}^2 &= \frac{\lambda t_f^2}{2} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} M_{11}^2 & M_{11}M_{12} \\ M_{11}M_{12} & M_{12}^2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \\ d_{\text{conv},N}^2 &= \frac{\lambda t_f^2}{2} (M_{11} - M_{12})^2 = \frac{\lambda t_f^2}{2} \langle \delta \mu \rangle_N^2, \end{aligned} \quad (29)$$

where $\delta \mu(E) = \mu_f(E) - \mu_b(E)$.

This can be compared to the Tapiovaara–Wagner expression for the SNR² with a conventional detector [Eq. (1)].

Assuming that the detector area and absorption and response, $D(E)$ and $w(E)$, are factored into the definition of the spectrum, their result is

$$\text{SNR}_{\text{conv, Tapiovaara-Wagner}}^2 = \frac{[\int (\overline{N_1(E)} - \overline{N_2(E)}) dE]^2}{\int (\overline{N_1(E)} + \overline{N_2(E)}) dE}. \quad (30)$$

Referring to Fig. 2, this equation can be expressed using the variables in this paper. The Tapiovaara-Wagner $\overline{N_1(E)}$ is the spectrum of the x-ray photons transmitted through the background only region

$$\overline{N_1(E)} = n(E) e^{-t_b \mu_b(E)} = n_T(E)$$

and

$$\overline{N_2(E)} = n(E) e^{-t_b \mu_b(E) - t_f (\mu_f(E) - \mu_b(E))} = n_T(E) e^{-t_f \delta \mu(E)}.$$

Since the object is assumed to have low contrast, $t_f \delta \mu(E)$ is small and

$$\overline{N_1(E)} - \overline{N_2(E)} \approx n_T(E) t_f \delta \mu(E). \quad (31)$$

Also

$$\overline{N_1(E)} + \overline{N_2(E)} \approx 2n_T(E). \quad (32)$$

Substituting in Eq. (30) and using $N_1(E) = n_T(E)$,

$$\text{SNR}_{\text{conv, Tapiovaara-Wagner}}^2 = t_f^2 \frac{[\int n_T(E) t_f \delta \mu(E) dE]^2}{2 \int n_T(E) dE}.$$

Recalling that $\lambda = \int n_T(E) dE$ and using the definition of $\hat{n}(E)$ (28),

$$\begin{aligned} \text{SNR}_{\text{conv, Tapiovaara-Wagner}}^2 &= \frac{\lambda t_f^2}{2} \left[\int \hat{n}(E) t_f \delta \mu(E) dE \right]^2 \\ &= \frac{\lambda t_f^2}{2} \langle \delta \mu \rangle_N^2. \end{aligned}$$

This is the same as $d_{\text{conv}, N}^2$ (29) with the detection theory approach.

II.F.2. Conventional energy integrating detector

The M matrix with an energy integrating detector also has only a single row

$$M = [\langle \mu_f(E) \rangle_Q \langle \mu_b(E) \rangle_Q], \quad (33)$$

where

$$\langle \mu(E) \rangle_Q = \int \mu(E) \hat{q}(E) dE$$

and

$$\hat{q}(E) = \frac{E n_T(E)}{\int E n_T(E) dE}.$$

The variance of the logarithm of a compound Poisson random variable (21) is

$$R_L = \frac{F}{\lambda/2}.$$

Substituting in Eq. (20)

$$d_{\text{conv}, Q}^2 = \frac{\lambda t_f^2}{2F} \langle \delta \mu \rangle_Q^2. \quad (34)$$

Comparing this to the result for the counting detector [Eq. (29)], the energy integrating d^2 will be smaller for several reasons. First, F is greater than 1 with nonzero width spectra. Also, $q(E)$ has a larger average energy than $n(E)$ so $\langle \delta \mu \rangle_Q^2$ will be smaller than $\langle \delta \mu \rangle_N^2$ since attenuation coefficients decrease with energy above the K edge.

II.F.3. Photon counting detector with pulse height analysis

The opposite of a conventional detector is a detector that measures the full spectrum of the photons transmitted through the object. We can model this case as the limit of a photon counting detector with PHA when the number of bins becomes large. This section derives expressions for d^2 for a detector with a finite number of energy bins. This is of interest in its own right since it is a common configuration. The next section takes the limit as the number of bins approaches infinity.

A detector with PHA uses a set of K counters associated with bins for photons in the energy ranges $E_k = E_0 + k \Delta E$, $k = 1, K$. It measures the energy of each photon, computes the corresponding bin, and then increments the associated counter. In order to focus on the essential effects, the detector will be assumed to absorb all the energy of each incident photon, to have zero electronic noise, and that other effects such as pulse pileup are negligible so the response functions are “rect” functions with zero width transitions between bins. The counts accumulated during the x-ray exposure are then Poisson random variables with mean value $\lambda_k = \int_{E_0 + (k-1)\Delta E}^{E_0 + k\Delta E} n_T(E) dE$. Since the counts correspond to non-overlapping energy intervals, they are independent and the $K \times K$ covariance matrix of the logarithm of the data has nonzero values only on the diagonal. The diagonal values are equal to the variances of the logarithms of the counts for each bin

$$R_L = \begin{bmatrix} 1/\lambda_1 & & 0 \\ & \ddots & \\ 0 & & 1/\lambda_K \end{bmatrix}.$$

The M matrix has K rows and two columns

$$M = \begin{bmatrix} \langle \mu_f(E) \rangle_1 & \langle \mu_b(E) \rangle_1 \\ \vdots & \vdots \\ \langle \mu_f(E) \rangle_K & \langle \mu_b(E) \rangle_K \end{bmatrix},$$

where $\langle \rangle_k$ is the average value using the normalized spectrum of the energy bin k . Substituting in Eq. (20) and carrying out the matrix computations

$$d_K^2 = t_f^2 \sum_{k=1}^K \lambda_k \langle \delta \mu \rangle_k^2. \quad (35)$$

II.F.4. Ideal detector with full energy information

We can model the ideal detector using the results for K bin PHA (35) as $K \rightarrow \infty$ and the width of the bins approaches zero. Rewriting Eq. (35) in terms of the transmitted photon spectrum $n_T(E)$,

$$d_K^2 = \frac{t_f^2}{2} \sum_{k=1}^K \left[\int_{E_0+(k-1)\Delta E}^{E_0+k\Delta E} n_T(E) dE \right] \times \left[\int_{E_0+(k-1)\Delta E}^{E_0+k\Delta E} \hat{n}_k(E) \delta\mu(E) dE \right]^2.$$

Applying the mean value theorem²⁴ shows that the first integral in brackets is equal to $n_T(\xi_1)\Delta E$, where ξ_1 is in the range $[E_0+(k-1)\Delta E, E_0+k\Delta E]$. The mean value theorem for weighted means²⁴ implies that the second integral is $\delta\mu(\xi_2)$, where ξ_2 is in the same range as ξ_1 . Taking the limit as $K \rightarrow \infty$ and $\Delta E \rightarrow 0$, $n_T(\xi_1) \rightarrow n_T(E_k)$ and $\delta\mu(\xi_2) \rightarrow \delta\mu(E_k)$, where $E_k = E_0 + k\Delta E$ and the ideal result with full energy information predicted by the detection theory approach is

$$d_{\text{ideal}}^2 = \frac{t_f^2}{2} \lim_{K \rightarrow \infty} \sum_{k=1}^K n_T(\xi_1) [\delta\mu(\xi_2)]^2 \Delta E = \frac{t_f^2}{2} \int n_T(E) [\delta\mu(E)]^2 dE. \quad (36)$$

We can compare this to the ideal Tapiovaara–Wagner detector with optimal gain factor (3),

$$\text{SNR}_{\text{ideal, Tapiovaara–Wagner}}^2 = \int \frac{[\overline{N_1(E)} - \overline{N_2(E)}]^2}{\overline{N_1(E)} + \overline{N_2(E)}} dE.$$

Recall that the signal area has been incorporated into the definition of $\overline{N(E)}$. Utilizing the results from the previous section for the numerator (31) and the denominator (32),

$$\begin{aligned} \text{SNR}_{\text{ideal, Tapiovaara–Wagner}}^2 &= \frac{t_f^2}{2} \int \frac{n_T^2(E) [\mu_f(E) - \mu_b(E)]^2}{n_T(E)} dE \\ &= \frac{t_f^2}{2} \int n_T(E) [\delta\mu(E)]^2 dE. \end{aligned}$$

Thus, $d_{\text{ideal}}^2 = \text{SNR}_{\text{ideal, Tapiovaara–Wagner}}^2$.

We can also compare the ideal result to the result for the conventional photon counting detector (29) by writing the expression for d_{ideal}^2 (36) as

$$\begin{aligned} d_{\text{ideal}}^2 &= \frac{\lambda t_f^2}{2} \frac{\int n_T(E) [\delta\mu(E)]^2 dE}{\int n_T(E) dE} \\ &= \frac{\lambda t_f^2}{2} \langle [\delta\mu(E)]^2 \rangle_N, \end{aligned}$$

where $\lambda = \int n_T(E) dE$. Comparing to Eq. (29) and applying a similar argument used with F , the effective value of the square of $\delta\mu(E)$ is greater than or equal to the square of the effective value

$$\langle [\delta\mu(E)]^2 \rangle_N \geq \langle \delta\mu(E) \rangle_N^2.$$

Therefore $d_{\text{ideal}}^2 \geq d_{\text{conv}}^2$ with equality only if the spectrum has zero width or if $\delta\mu(E)$ is constant.

II.F.5. The NQ detector

The covariance of the logarithm of the NQ detector measurements is given by Eq. (21),

$$R_{L,NQ} = \frac{1}{\lambda} \begin{bmatrix} 1 & 1 \\ 1 & F \end{bmatrix}. \quad (37)$$

Roessl *et al.*⁹ noted that the positive correlation reduces the noise in the components of A and this section shows that this also increases d^2 . The M matrix with an NQ detector is 2×2 with the effective attenuation coefficients with the photon number spectrum in the first row and with the photon energy spectrum in the second

$$M_{NQ} = \begin{bmatrix} \langle \mu_f(E) \rangle_N & \langle \mu_b(E) \rangle_N \\ \langle \mu_f(E) \rangle_Q & \langle \mu_b(E) \rangle_Q \end{bmatrix}.$$

Substituting in Eq. (20) results in

$$d_{NQ}^2 = \frac{\lambda t_f^2}{2} \left[\langle \delta\mu \rangle_N^2 + \frac{(\langle \delta\mu \rangle_N - \langle \delta\mu \rangle_Q)^2}{F - 1} \right]. \quad (38)$$

Compared this to the equation for d_{conv}^2 [Eq. (29)], we see that

$$d_{NQ}^2 = d_{\text{conv}}^2 + \left(\frac{\lambda t_f^2}{2} \right) \frac{(\langle \delta\mu \rangle_N - \langle \delta\mu \rangle_Q)^2}{F - 1}.$$

Note that $F \geq 1$ so the second term is always positive except as the spectrum width approaches zero. In this case, the photon number and energy spectra are both delta functions so $F \rightarrow 1$ but $\langle \delta\mu \rangle_N \rightarrow \langle \delta\mu \rangle_Q$ and the second term is zero. Thus, the d^2 with an NQ detector is greater than or equal to the conventional detector value with equality only for zero width spectra.

II.F.6. The NKQ detector

As discussed in the previous section, a generalization of the NQ detector is to use PHA instead of simply counting the number of photons while still measuring the total energy of the photons. In this case, d^2 can be computed directly from the general expression (20) using the measurement covariance matrix (25).

II.G. Single material object d^2

In medical imaging the object is almost always heterogeneous but, in some cases, the object may be modeled as composed of a single material. The detection theory formulation described in Sec. II C used two materials chosen to be appropriate for the imaging task as the basis set. Since the decomposition of the attenuation coefficient requires two functions, the approach must be modified to apply to a single material case. In this case, the geometry is shown in Fig. 4. The object consists of a slab of a material with decomposition vector $\mathbf{a}_{\text{material}}$ of thickness t_b plus a feature region of the

FIG. 4. Geometry for single material object.

same material of thickness t_f . Under hypothesis H_0 the feature is not present so the basis set coefficient line integral vector is $\mathbf{A}_0 = t_b \mathbf{a}_{\text{material}}$, while under hypothesis H_1 , the object is present so the line integral vector is $\mathbf{A}_1 = (t_b + t_f) \mathbf{a}_{\text{material}}$. The difference in mean values is then

$$\delta \mathbf{A} = \mathbf{A}_1 - \mathbf{A}_0 = t_f \mathbf{a}_{\text{material}}.$$

We can use this in the general expression for d^2 (20). By using the material attenuation coefficient function as one of the basis set functions

$$\mathbf{a}_{\text{material}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

the calculations can then proceed as in the previous sections resulting in the same formulas with $\mu_b(E)=0$ so $\delta\mu(E) = \mu_f(E)$.

II.H. Achieving optimal d^2 in an imaging system

We can think of the parameter d^2 as the square of a signal to noise ratio, which intuitively is the distance in A space between the mean values under the two hypotheses divided by the standard deviation of the noise. With this interpretation, we could visualize the ability to distinguish between materials by plotting the mean values on a graph and measuring the distance between them. However, we cannot directly apply this approach because the noise in A space, or a space for CT, is highly negatively correlated. This is shown in Fig. 5(a), a scatterplot of the simulated image noise of an object consisting of soft tissue with an air background. With correlated noise, the performance depends on the angle of the vector from the mean values of the background to the object. In this section, a “whitening transform” is derived that produces uncorrelated transformed data with equal standard deviations on both axes. This not only allows us to gauge the performance based in the distance between the mean values but, by using projections in the whitened A space, we can derive a method to produce images with the optimal signal to

FIG. 5. Scatterplots of noise in A -space. The scatterplot shows the values of the A vectors of pixels in a region containing air, i.e., no object or a uniform slab of soft tissue. The vertical and horizontal lines intersect at the origin. The bottom graph shows the whitened data computed by transforming with the Φ_w matrix (40). The components of the transformed data are uncorrelated and have unit variance. The distance between the centers of the mean values is the signal to noise ratio d .

noise ratio d^2 . The projection in whitened A -space method is then extended to produce material canceled images.

II.H.1. The whitening transformation

In general, we can implement any linear transformation of coordinates by multiplying by a square matrix of dimension equal to the dimensionality of the space. If we multiply by a matrix Φ whose columns are the eigenvectors ϕ_i of the covariance matrix R_A , the transformed data in each coordinate will be uncorrelated. This can be seen as follows. By the definition of an eigenvector, the matrix Φ satisfies

$$R_A \Phi = \Phi V, \quad (39)$$

where V is a diagonal matrix with the eigenvalues v_i on the diagonal.

$$V = \begin{bmatrix} v_1 & & 0 \\ & \ddots & \\ 0 & & v_K \end{bmatrix}.$$

If $A' = \Phi^T A$ are the transformed data, the covariance of A' is

$$\begin{aligned} R_{A'} &= E_x[(\delta A')(\delta A')^T] = E_x[\Phi^T \delta A \delta A^T \Phi] = \Phi^T R_A \Phi \\ &= \Phi^T \Phi V = V, \end{aligned}$$

where the last step follows because the eigenvectors are orthonormal so $\Phi^T \Phi = \mathbf{I}$, the identity matrix. All the off-diagonal elements of V are 0 so the coordinates are uncorrelated.

The components of A' are uncorrelated but they do not have equal variance. We can accomplish by transforming with a related matrix Φ_w whose columns are the eigenvectors divided by the square root of their eigenvalues

$$\Phi_w = \begin{bmatrix} \frac{\phi_1}{\sqrt{v_1}} & \dots & \frac{\phi_K}{\sqrt{v_K}} \\ \vdots & \ddots & \vdots \end{bmatrix} = \Phi V^{-1/2}. \quad (40)$$

This definition relies on the fact that the matrix V is diagonal so

$$V^{-1/2} = \begin{bmatrix} 1/\sqrt{v_1} & & 0 \\ & \ddots & \\ 0 & & 1/\sqrt{v_K} \end{bmatrix}.$$

With this definition, $A_w = \Phi_w^T A$ has uncorrelated components with variance one,

$$\begin{aligned} R_{A_w} &= E_x[(\delta A_w)(\delta A_w)^T] \\ &= E_x[\Phi_w^T \delta A \delta A^T \Phi_w] \\ &= \Phi_w^T R_A \Phi_w \\ &= V^{-1/2} \Phi^T R_A \Phi V^{-1/2} \\ &= V^{-1/2} \Phi^T \Phi V V^{-1/2} \\ &= V^{-1/2} V^{1/2} = \mathbf{I}. \end{aligned}$$

We can show that the d^2 parameter is unchanged in this new coordinate system by substituting in the general expression (11),

$$d_w^2 = (\delta A_w)^T R_{A_w}^{-1} \delta A_w = (\delta A_w)^T \delta A_w,$$

where the last step follows because $R_{A_w} = \mathbf{I} = R_{A_w}^{-1}$. By definition, $\delta A_w = \Phi_w^T \delta A$ so

$$d_w^2 = (\delta A)^T [\Phi_w \Phi_w^T] \delta A. \quad (41)$$

Applying the definition of Φ_w (40), the term in brackets in Eq. (41) is

$$[\Phi_w \Phi_w^T] = \Phi V^{-1/2} (V^{-1/2})^T \Phi^T = \Phi V^{-1} \Phi^T,$$

where the last step follows since V is diagonal and $(V^{-1/2})^T = V^{-1/2}$. Since Φ is orthonormal, $\Phi^T = \Phi^{-1}$. Therefore, by Eq. (39), $V^{-1} \Phi^T = \Phi^{-1} R_A^{-1}$ and

$$[\Phi_w \Phi_w^T] = \Phi \Phi^{-1} R_A^{-1} = R_A^{-1}.$$

Substituting in Eq. (41) shows that $d_w^2 = (\delta A)^T R_A^{-1} (\delta A) = d^2$.

Since the noise in the transformed coordinate system has variance equal to one for both axes, the performance parameter d^2 is simply the square of the distance between the trans-

formed mean values irrespective of the angle of the vector. Figure 5(b) shows the result of applying the whitening transformation to the simulated data with noise.

II.H.2. Grayscale images with optimal d^2 in an imaging system

Medical imaging systems typically display scalar grayscale image data. This section describes a method to compute a grayscale image analogous to a conventional image but with the optimal d^2 . We can do this by forming a scalar dot product, $g = \mathbf{A}_w \cdot \hat{\delta}$, of the two-dimensional whitened A -space data with a vector $\hat{\delta}$ with unit length in the direction from the mean value of the background data to the mean value of the feature data

$$\hat{\delta} = \frac{\langle \mathbf{A}_{w,f} \rangle - \langle \mathbf{A}_{w,b} \rangle}{\| \langle \mathbf{A}_{w,f} \rangle - \langle \mathbf{A}_{w,b} \rangle \|},$$

where the brackets $\langle \dots \rangle$ indicate a mean value and the braces $\| \dots \|$ indicate the length of a vector. The data in the feature region of the image are

$$g_f = \mathbf{A}_{w,f} \cdot \hat{\delta} = A_{w,f,1} \hat{\delta}_1 + A_{w,f,2} \hat{\delta}_2$$

and, since the components of $\mathbf{A}_w$ are uncorrelated and have variance equal to one, the variance of g_f is

$$\sigma_{g_f}^2 = \hat{\delta}_1^2 \sigma_{A_{w,f,1}}^2 + \hat{\delta}_2^2 \sigma_{A_{w,f,2}}^2 = \hat{\delta}_1^2 + \hat{\delta}_2^2 = 1.$$

By the same argument, the variance of the data in the background region $\sigma_{g_b}^2$ is also one. As derived in the previous section, the distance between the mean values in A_w space is d so this is also the difference between the projected values

$$\langle g_f \rangle - \langle g_b \rangle = (\langle \mathbf{A}_{w,f} \rangle - \langle \mathbf{A}_{w,b} \rangle) \cdot \hat{\delta} = \| \langle \mathbf{A}_{w,f} \rangle - \langle \mathbf{A}_{w,b} \rangle \| \| \hat{\delta} \| = d$$

since $\| \hat{\delta} \| = 1$. The signal to noise ratio squared of the scalar data is then

$$\text{SNR}_g^2 = \frac{(\langle g_f \rangle - \langle g_b \rangle)^2}{\sigma_g^2} = d^2.$$

II.H.3. Material canceled images in whitened A space

The energy selective system not only provides a grayscale image equivalent to a conventional image with the optimal signal to noise ratio but it can also compute material canceled images. For example, computing a dot product of the A -space data with a unit vector $\hat{\delta}_\perp$ perpendicular to $\hat{\delta}$ will cancel the contribution of the average soft tissue material from the data. Figure 6(b) shows that if there is another material such as the adipose tissue, denoted by circles in the figure, with data at a different angle from the soft tissue material, it can be distinguished in the canceled image. The noise will not be canceled and the signal to noise ratio will be different from the maximum signal to noise ratio data computed in the previous section. However, the canceled data represent a different physical quantity so the signal to noise ratios are not directly comparable.

FIG. 6. Computing images with maximum d^2 and material canceling in whitened A space. The maximum d^2 image is computed by forming a dot product of the A -space data with a unit vector δ from the mean of the background to the material data. To cancel the material, use a unit vector $\delta_{\perp}$ that is perpendicular to δ . The histograms of the scalar projected data are superimposed on the plots.

II.I. Computer simulation of performance

The formulas derived in this section were used with a computer simulation to compare detector d^2 as a function of operating parameters. In general, the d^2 parameter depends on the total flux so it varies with factors such as detector area, exposure time, and distance from the source. In order to focus on the spectral properties, the results are presented as the ratio of d^2 to the Tapiovaara–Wagner optimal SNR^2 . All the expressions for d^2 derived in the previous section can be factored into a product of the total flux λ and a factor that depends only on the normalized spectrum. We also can rewrite the expression for the optimal value (36) as

$$d_{\text{ideal}}^2 = \lambda \frac{t_{\text{int}}^2}{2} \int \hat{n}(E) [\delta \mu(E)]^2 dE$$

by using the definition of the normalized spectrum (28). Thus, d_{ideal}^2 is also proportional to λ and it factors out of the ratio d^2/d_{ideal}^2 , leaving a quantity that only depends on the normalized spectrum. In addition, of course, the ratio measures how closely the detectors are able to approach the optimal result.

For a particular set of operating conditions such as tube voltage and object transmission, the same spectrum was assumed to be incident on all detectors. As a result, the detectors were compared with equal patient dose.

The formulas for d^2 require the effective values of the difference of attenuation coefficients with the spectra incident on the detector. The incident spectrum is the product of the source spectrum times the transmission of the object at each energy so the simulation requires the computation of these two quantities.

The computer simulation assumed that the source was an x-ray tube. The tube spectra at a given voltage were computed with a computer model that generates the bremsstrahlung and fluorescent radiation separately. The bremsstrahlung x-ray spectrum data are calculated from Kramer's model.²⁵ The fluorescent radiation is then added at the characteristic energies of the anode material if the tube voltage is sufficiently high. The intensities of the fluorescent radiation are based on Tothill's results.²⁶ The simulations assumed a tungsten anode tube with a 3 mm aluminum window thickness. The software based on Kramer's model produces an energy spectrum $q(E)$, which is the energy of the x-ray photons per bin normalized so the sum of the bin values is the total energy of the x-ray photons emitted at the tube surface in joules for 1 mA s integrated current. The photon number spectrum $n(E)$ was computed by dividing the energy spectrum for each bin by the photon energy E in keV and multiplying by the number of keV/J.

This model was chosen because it allows computation of spectra with arbitrarily fine energy resolution, which was found to be required to give accurate numerical results when comparing to the Tapiovaara–Wagner optimal value. In the simulations, the spectrum was computed at 200 points from 10 keV to the electron charge e times the tube voltage.

The attenuation coefficients were computed based on the NIST Hubbell–Seltzer database.²⁷ This database provides attenuation coefficients of elements as a function of energy that were then log-log interpolated between absorption edges. The attenuation coefficients of the body materials were computed using the elemental composition and fraction by weight tables from International Commission on Radiological Protection (ICRP) available on the NIST ESTAR database.²⁸

III. RESULTS

This section shows the results of the computer simulation of detector performance as a function of operating parameters.

III.A. Performance with two material object

For the performance as a function of tube voltage, the background material was assumed to be 3 cm soft tissue and the object cortical bone. For the variable thickness results, the material was also soft tissue and the tube voltage was fixed at 120 kV.

FIG. 7. Performance vs bin separation energy for two bin PHA. The photon counts only detector is labeled $N2$, while the detector that also measures the total energy is labeled $N2Q$. The results are shown with x-ray tube spectra for 40, 70, 90, and 120 kV transmitted through a 10 cm water object. The optimal values for each type of detector are indicated by vertical lines.

III.A.1. Optimizing the bin separation energy with two bin PHA

The performance with PHA depends on the bin separation energy and the results for two bin PHA in this section are derived by selecting the value of the bin separation energy that maximizes d^2 . Figure 7 shows plots of d^2 for two bin PHA only ($N2$) and two bin PHA plus total energy ($N2Q$) detectors for four tube voltages.

Figure 8 shows the optimal bin separation energy versus the tube voltage and soft tissue object thickness.

FIG. 8. Optimal threshold energy for detectors with two bin PHA (a) vs tube voltage and (b) object thickness. The photon counts only detector is labeled $N2$, while the detector that also measures the total energy is labeled $N2Q$. A soft tissue object is assumed.

FIG. 9. Signal to noise ratio squared for energy selective detectors as a function of x-ray tube voltage. The results are normalized by dividing by the SNR^2 of the ideal detector at each voltage. The legend indicates the type of detector: (1) $N2Q$ detector, (2) NQ detector, (3) two energy bin photon counting detector ($N2$), (4) conventional photon counting ($convN$), and (5) conventional energy integrating ($convQ$). The ideal value was computed using Eq. (36).

III.A.2. d^2 vs tube voltage

The tube voltage has a strong effect on the spectral characteristics so the effective signal to noise ratio squared was computed as a function of the voltage for five types of detectors: (1) Conventional energy integrating ($convQ$), (2) conventional photon counting ($convN$), (3) two energy bin photon counting detector ($N2$), (4) NQ detector, and (5) $N2Q$ detector (see Fig. 9). The spectrum incident on the detector was the same for all types so the patient dose was the same in all cases. As discussed in the previous section, the detectors were assumed to absorb all incident radiation and those with PHA to have a zero width transition between each energy bin. The optimal bin separation energy was recomputed for each tube voltage. The d^2 values were divided by the ideal detector value giving a relative value. Note that the distance to the object and the detector area do not affect these relative value results.

III.A.3. d^2 vs object thickness

The results for the same set of detector types as a function of object thickness at a fixed tube voltage is shown in Fig. 10. For detectors with two bin PHA, the separation energy was optimized for each object thickness value although, as shown in Fig. 8(b), the optimal value does not vary substantially.

III.B. Performance with other materials and single material objects

This section compares the relative performance of each of the detector types with different object compositions. The basis materials are chosen to best model a particular imaging task. The results of the previous section assume a bone and soft tissue object, which might be used to model a chest image. The is an example of materials with a relatively large

FIG. 10. The d^2 parameter as a function of soft tissue object thickness. The incident spectrum was from a 120 kV x-ray tube. The values are normalized by dividing by the ideal value.

difference in the atomic number composition. In other applications, such as mammography, two relatively low atomic number materials such as soft and adipose tissues may be more appropriate. In other cases, objects composed of a single material should be considered. The results with a single hypothetical material composed of the element, hydrogen, oxygen, or calcium, are also presented in this section.

III.B.1. Two material case: Soft and adipose tissue object

Figure 11 shows the performance as a function of tube voltage for an object composed of soft and adipose tissues.

III.B.2. Single material case: Effect of atomic number

The performance relative to d^2_{ideal} depends on the variation in $\mu_f(E)$. For low atomic number materials, the attenuation coefficient is dominated by Compton scattering so the variation is relatively small, while for higher atomic number ma-

FIG. 11. Relative performance with a soft and adipose tissue object as a function of tube voltage. The soft tissue thickness was 10 cm.

FIG. 12. Relative d^2 as a function of tube voltage for a single material object. The materials are (a) hydrogen, (b) oxygen, and (c) calcium.

terials, the attenuation is dominated by the photoelectric effect and the variation is relatively large. Figure 12 illustrates the effect of the material atomic number with objects composed of three pure elements: (a) hydrogen, (b) oxygen, and (c) calcium.

III.B.3. $\delta\mu(E)$ with different object materials

Since d^2 depends on the weighted average value of $\delta\mu(E)$ over the normalized x-ray spectrum transmitted through the object, the plots of this function in Fig. 13 for the soft/adipose tissue object and the three single element objects give us some insight into the results in the previous section.

IV. DISCUSSION

The approach presented here and the Tapiovaara–Wagner approach give the same results for the SNR of a conventional photon counting detector (29) and the optimal SNR² (36). The two methods use different models of the object x-ray attenuation so the concordance of the theoretical formulas and also the numerical results is a good indication of the validity of the approach.

Figure 9 shows the performance of different types of detectors as a function of x-ray tube voltage. In general, all of the detector types with energy selectivity perform substantially better than either conventional energy integrating or photon counting detectors.

For all detectors, the ratio d^2/d^2_{ideal} is larger for lower tube voltage spectra than at higher voltages. This may be due to the fact that at lower voltages the spectra are more narrow and therefore have less energy dependent information. The NQ detector SNR is near ideal for low voltage spectra but it decreases at higher tube voltage although it remains comparable to two bin PHA at higher voltages. The N2Q detector

FIG. 13. $\delta\mu(E)$ and the normalized spectrum. The d^2 parameter depends on the integral of the product of these two functions. The values of $\delta\mu(E)$ for each object material were normalized to have the same maximum value. Note the similar shape of the bone/soft, soft/adipose, and bone only $\delta\mu(E)$, while the soft tissue only function is more constant over the tube spectrum range. The curve for hydrogen (H) and the spectrum are both drawn with solid lines.

with two bin PHA and the integrated photon energy has near ideal performance at all voltages. It therefore represents a good trade-off between complexity and performance. Finally, as is well known, the conventional photon counting detector gives better results than an energy integrating detector for all voltages.

Figure 10 shows that d^2/d_{ideal}^2 does not vary substantially with object thickness and that the relative ranking of the detector types does not change except that the NQ and $N2$ detector performance become almost equal. The $N2Q$ detector again gives near optimal performance at all thicknesses.

The performance vs bin separation energy results in Fig. 7 show an interesting difference between $N2$ and $N2Q$ detectors. The $N2$ curves are unimodal with a peak at the optimal trade off between the noise and energy separation of the two effective spectra. The $N2Q$ detector has a more complex bimodal curve representing the additional factor of the correlations of the photon count data and Q . The optimal bin separation energy with either detector type vs tube voltage, shown in Fig. 8(a), is a smooth function that can be easily computed by the system to adjust the threshold for the known tube voltage of a particular examination. At low voltage, the object attenuation is high so there is significant beam hardening and the optimal threshold is close to the maximum energy in the spectrum. The optimal bin separation energy vs object thickness, shown in Fig. 8(b), is approximately con-

stant. This is a necessary property of detectors used in medical imaging since subjects may have a wide range of attenuation.

The object material has an effect on the relative performance. Figures 9, 11, and 12(c) show that two material objects or high atomic number objects all have nearly the same d^2/d_{ideal}^2 . A single relatively low atomic number object such as hydrogen has a markedly different performance, as shown in Fig. 12(a). The performance depends in the difference $\delta\mu(E)$ in the attenuation coefficients of the basis materials, shown in Fig. 13. Since Compton scattering is approximately proportional to mass density, the difference in the mass attenuation coefficient is approximately only the photoelectric component, which has a common functional form of $1/E^3$ for elements in body materials. As a result, the relative performance with two materials, such as the bone-soft tissue case in Fig. 9 and the adipose-soft tissue case in Fig. 11, is very similar. The attenuation coefficient of calcium is also dominated by the photoelectric component so its results are similar to the two material cases. However, the attenuation coefficient of hydrogen is dominated by the Compton scattering component particularly at higher energies. The Compton scattering cross section has less variation with energy than the photoelectric cross section and therefore less energy dependent information. As a result, in Fig. 12(a), d^2/d_{ideal}^2 is approximately constant and near one for all detector types except energy integrating. With this detector, the F factor increases the noise variance and, since it increases with the spectral width at higher voltages, it shows a decrease in performance. The performance with oxygen is intermediate between hydrogen and calcium.

Section II H shows that the improved SNR with low resolution detectors can be implemented in an imaging system by using the projections in whitened A-space approach. The method is applicable with any type of energy selective data acquisition and not limited just to detectors with PHA. It can be used, for example, with systems with voltage switching although noise level will depend on the characteristics of the data acquisition.

The A-space information from the energy selective system can also be used to produce material canceled images with optimal SNR, as shown in Sec. II H 3. This shows that energy selective systems extract more information for same dose as conventional systems since both increased SNR and material-specific information are extracted simultaneously. Material canceled images are fundamentally different than conventional so they have different noise characteristics. We may have to use more dose to produce images suitable for diagnosis but that is due to imaging different quantities, not increased noise with the energy selective system.

The results presented in this paper assume idealized detectors but, since the principal conclusion is that low resolution detectors can provide good performance, we would not expect that the degradation of resolution with experimental detectors used in medical imaging systems would change this conclusion. The statistical detection theory methods described in this paper provide an approach to comparing ex-

perimentally measured energy selective system performance with the ideal Tapiovaara–Wagner value comparable to detective quantum efficiency with conventional detectors. Experimental methods to implement this comparison are under study.

The derivation of the d^2 parameter assumed that the covariance matrix is the same under both hypotheses, which is valid for low contrast objects. For objects with higher contrast, the quantitative results for d^2 are an approximation as discussed previously. With high contrast objects, the underlying vector space approximation of the attenuation coefficient and the normal probability distribution models of the noise remain valid but the inversion of the attenuation equation (14) requires nonlinear methods. Once the nonlinear equations are inverted, the results are points in the A space regardless of whether the object is low or high contrast, so the general conclusions for low contrast objects remain valid.

The solution of the nonlinear equations has been discussed extensively in the literature for two spectral measurements but the case with more than two measurements, such as from systems with PHA, is the subject of ongoing research. Roessl and Proksa²⁹ described an iterative likelihood maximization method to solve the equations in this case. The method is applicable with a wide variety of detector types and they show simulations of systems with measurements with five spectra. As with any iterative method, it may suffer problems with convergence and speed but their results give an existence proof of the solution with multiple measurements.

As shown in Appendix A, the general conclusions about the relative energy selective detector performance with a projection system are also applicable with CT for a simplified model of an object. In general, noise in CT is nonstationary but lower noise in the line integrals results in lower noise in the reconstructed image. In addition, the A -space data can be used to eliminate beam hardening artifacts. Giersch *et al.*³ noted an increased “cupping” artifact with optimal Tapiovaara–Wagner weighting. Niederlohner *et al.*⁴ and Schmidt³⁰ observed that doing separate reconstructions of the data from each energy bin reduces the artifacts. However, the A -space method can theoretically eliminate the artifacts completely. It provides the framework for material canceled images and other simple displays of the energy dependent information¹¹ and only requires two reconstructions.

V. CONCLUSION

Simple detectors, particularly a two bin PHA detector that simultaneously measures integrated energy data, can provide near ideal performance over a wide range of operating conditions. Images with SNR equal to the optimal hypothesis test d^2 can be implemented in an imaging system using the projections in whitened A -space approach. A similar approach can be used to compute material canceled images with optimal SNR.

ACKNOWLEDGMENTS

The author would like to acknowledge useful discussions with Dr. John Boone and Dr. Anthony Seibert and the reviewers for helpful comments and suggestions on the manuscript.

APPENDIX A: RELATIONSHIP BETWEEN PROJECTION AND CT d^2

Assuming we reconstruct the vector a -space data with a convolution, backprojection reconstruction algorithm for each component of the line integrals $\vec{A}$, the reconstructed $\vec{a}$ vector is

$$\vec{a} = \sum_{l=1}^{N_{\text{angles}}} \sum_{k=-N/2}^{N/2} \vec{A}(k,l)h(k),$$

where l designates one of N_{angles} angles, the k index designates the detector number and $h(k)$ is the impulse response of the convolution filter. With a disk object with diameter t_b and a region of interest (ROI) at the center that is larger than the point spread function, the line integrals $\vec{A}$ are essentially constant across the ROI, and since the noise is independent for each separate detector, the covariance matrix at the center is

$$R_a = R_A N_{\text{angles}} \sum_{k=-N/2}^{N/2} h^2(k).$$

As shown in Ref. 31, Vol. 2, Sec. 10.5.3,

$$\sum_{k=-N/2}^{N/2} h^2(k) = \frac{2\pi^2 f_m^3}{N_{\text{angles}}},$$

where f_m is the cutoff frequency of the convolution filter. This is related to the full width at half maximum of the point spread function by $w_{1/2} = 0.7/f_m$ so

$$R_a = \frac{6.8}{N_{\text{angles}}^2 w_{1/2}^3} R_A.$$

For CT, we can use general expression for d^2 (11) with δa instead of δA ,

$$\delta a = \rho \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

where ρ is a dimensionless constant equal to the ratio of the density of the object divided by the density of the basis material. Substituting in Eq. (11)

$$d_{\text{CT}}^2 = \frac{N_{\text{angles}}^2 w_{1/2}^3 \rho^2}{6.8 t_f^2} d_A^2.$$

The d_{CT}^2 parameter is therefore a multiplicative constant independent of the detector energy spectral properties times the value with projection imaging for the object assumed.

APPENDIX B: THE STATISTICS OF X-RAY DETECTOR MEASUREMENTS

In this appendix I derive the formulas for the full statistics including covariance of a compound Poisson random variable and its logarithm and also the statistics of NKQ detector data.

1. The Poisson random variable

A Poisson random variable N is defined for non-negative integers $n=0, 1, \dots$ and has the probability density function

$$f_N(n) = \text{Prob}(N=n) = e^{-\lambda} \frac{\lambda^n}{n!},$$

where λ is the parameter. The (inverse) Fourier transform of $f_N(n)$ is called the characteristic function³² and is

$$\phi_N(u) = \int_{-\infty}^{\infty} f_N(x) e^{iux} dx = e^{\lambda(e^{iu}-1)}. \quad (\text{B1})$$

The characteristic function is useful for computing the moments of a random variable such as the mean and the variance. The m th moment is given by³²

$$\langle N^m \rangle = (-i)^m \left. \frac{d^m \phi}{du^m} \right|_{u=0}. \quad (\text{B2})$$

Applying this to the Poisson random variable gives the moments and the variance,

$$\langle N \rangle = \lambda, \quad (\text{B3})$$

$$\langle N^2 \rangle = \lambda + \lambda^2, \quad (\text{B4})$$

$$\text{var}(N) = \lambda. \quad (\text{B5})$$

2. The compound Poisson random variable

Suppose a random variable Q is the sum of a random number of independent and identically distributed random variables $\{E_k\}$

$$Q = \sum_{n=1}^N E_n, \quad (\text{B6})$$

where N is a Poisson random variable and Q is defined to be 0 when $N=0$. This is called a compound Poisson random variable and its characteristic function is (see Ref. 16, Sec. 4.2)

$$\phi_Q(u) = e^{\lambda(\phi_E(u)-1)}, \quad (\text{B7})$$

where $\phi_E(u)$ is the characteristic function of E . Using this function, we can compute the mean and variance. Differentiating using the chain rule and noting that $\phi_E(0)=1$ for any characteristic function

$$\begin{aligned} \langle Q \rangle &= -i\lambda \phi_E'(0) e^{\lambda(\phi_E(0)-1)} \\ &= -i\lambda \phi_E'(0) = \lambda \langle E \rangle. \end{aligned}$$

Differentiating again gives the second moment and, combining with the first moment, the variance

$$\text{var}(Q) = \lambda \langle E^2 \rangle.$$

The random variables Q and N are not independent. Indeed, they are positively correlated as is clear from the definition (B6). The covariance was derived by Roessl *et al.*⁹ but here I give a closed form expression. The covariance is defined as

$$\text{cov}(N, Q) = E_x[(N - \langle N \rangle)(Q - \langle Q \rangle)]. \quad (\text{B8})$$

Expanding the product

$$\text{cov}(N, Q) = E_x[NQ] - \langle N \rangle \langle Q \rangle. \quad (\text{B9})$$

Expanding the first term

$$\begin{aligned} E_x[NQ] &= E_x[E_x[NQ|N=n]] \\ &= \sum_{n=0}^{\infty} n E_x \left[\sum_{k=1}^n E_k \right] \text{Prob}\{N=n\}, \end{aligned}$$

The E_k have same probability distribution so

$$E_x \left[\sum_{k=1}^n E_k \right] = n \langle E \rangle$$

and

$$E_x[NQ] = \langle E \rangle \sum_{n=0}^{\infty} n^2 \text{Prob}\{N=n\} = \langle E \rangle \langle N^2 \rangle = \langle E \rangle (\lambda + \lambda^2).$$

Applying the previous results $\langle N \rangle = \lambda$ and $\langle Q \rangle = \lambda \langle E \rangle$, so $\langle N \rangle \times \langle Q \rangle = \lambda^2 \langle E \rangle$ gives the final result

$$\text{cov}(N, Q) = \lambda \langle E \rangle.$$

Summarizing these results in the matrix form, the covariance matrix is

$$\text{Cov}(N, Q) = \lambda \begin{bmatrix} 1 & \langle E \rangle \\ \langle E \rangle & \langle E^2 \rangle \end{bmatrix}.$$

3. The logarithm of a Poisson and compound Poisson

The logarithm of the data is often used in x-ray imaging systems but is mathematically ill-defined because a Poisson random variable can have a value of zero with a nonzero probability. In hardware implementations, a value of 1 is added to the data or some other method is used to avoid the problem. Here, I will assume that this has been done so the logarithm of the data is mathematically well behaved. To derive expressions for the mean and variance of $y = \log(x)$, we can use a general result³² based on a series expansion of a function $y = g(x)$ about the mean value $\eta = E_x[x]$

$$g(x) = g(\eta) + g'(\eta)(x - \eta) + g''(\eta) \frac{(x - \eta)^2}{2} + \dots \quad (\text{B10})$$

Using this series,

$$E_x[g(x)] \approx g(\eta) + g''(\eta) \frac{\sigma_x^2}{2},$$

$$\text{var}[g(x)] \approx [g'(\eta)]^2 \sigma_x^2.$$

If $g(x)=\log(x)$, then $g'(x)=1/x$ and $g''(x)=-1/x^2$. Combining these results with those of the previous section gives the following for the log of the Poisson random variable $\log(N)$:

$$\begin{aligned} E_x[\log(N)] &= \log(\lambda) - \frac{\lambda}{2\lambda^2} + \dots \\ &\approx \log(\lambda), \\ \text{var}[\log(N)] &= \frac{1}{\lambda^2} \lambda + \dots \approx \frac{1}{\lambda}. \end{aligned} \quad (\text{B11})$$

Similarly,

$$\begin{aligned} E_x[\log(Q)] &= \log(\lambda \langle E \rangle) - \frac{\lambda \langle E^2 \rangle}{2\lambda^2 \langle E \rangle^2} \dots \\ &\approx \log(\lambda \langle E \rangle), \\ \text{var}[\log(Q)] &\approx \frac{1}{\lambda^2 \langle E \rangle^2} \lambda \langle E^2 \rangle = \frac{\langle E^2 \rangle}{\lambda \langle E \rangle^2}. \end{aligned} \quad (\text{B12})$$

We also require the covariance. Similarly to Eq. (B9)

$$\begin{aligned} \text{cov}[\log(N), \log(Q)] &= E_x[\log(N)\log(Q)] \\ &\quad - E_x[\log(N)]E_x[\log(Q)]. \end{aligned} \quad (\text{B13})$$

To evaluate the first term in Eq. (B13), we can use

$$\begin{aligned} \log(N) &= \log \left[\langle N \rangle \left(1 + \frac{\delta N}{\langle N \rangle} \right) \right] \\ &\approx \log(\langle N \rangle) + \frac{\delta N}{\langle N \rangle} \end{aligned}$$

with a similar expression for $\log(Q)$. With these two expansions, the first term becomes

$$\begin{aligned} E_x[\log(N)\log(Q)] &\approx \log(\langle N \rangle)\log(\langle Q \rangle) + \frac{\text{cov}(N, Q)}{\langle N \rangle \langle Q \rangle} \\ &= \log(\langle N \rangle)\log(\langle Q \rangle) + \frac{1}{\lambda}. \end{aligned}$$

Using Eqs. (B11) and (B12), the second term in Eq. (B13) is $E_x[\log(N)]E_x[\log(Q)] \approx \log(\langle N \rangle)\log(\langle Q \rangle)$. Combining these, Eq. (B13) becomes

$$\text{cov}[\log(N), \log(Q)] = \frac{1}{\lambda}.$$

Defining

$$F = \frac{\langle E^2 \rangle}{\langle E \rangle^2},$$

the covariance matrix of the logarithm of the compound Poisson variables is

$$\text{Cov}[\log(N), \log(Q)] = \frac{1}{\lambda} \begin{bmatrix} 1 & 1 \\ 1 & F \end{bmatrix}.$$

4. Covariance of NKQ detector measurements

With the NKQ detector, we use PHA to separate the random variable instances into K groups depending on their values and simultaneously measure the total energy. The numbers of photons in the individual bins, $\{N_1, N_2, \dots, N_K\}$, are Poisson random variables with parameters λ_k , $k = 1, \dots, K$. The total number of photons, N , is the sum

$$N = \sum_{k=1}^K N_k.$$

The counts are independent since they correspond to non-overlapping intervals of E values. Since the sum of independent Poisson random variables is also a Poisson random variable with parameter equal to the sum of the parameters of the terms, N is Poisson with parameter

$$\lambda = \sum_{k=1}^K \lambda_k.$$

The NKQ detector also measures the total energy Q of all the photons as defined in Eq. (B6).

The mean value vector of the measurements is

$$E_x \begin{bmatrix} N_1 \\ N_2 \\ \dots \\ N_K \\ N_Q \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_K \\ \lambda \langle E \rangle \end{bmatrix}.$$

To derive the joint statistics, we only need to consider the two bin case and derive the results for $\text{cov}(N_1, Q)$. These results can be immediately generalized to the K bin case because, from the point of view of the analysis, all the other measurements can be summed to a single Poisson random variable that is independent of N_1 , which is equivalent to the two bin case. Also, without loss of generality, the results for N_1 are applicable to any of the other individual measurements. The covariance is defined as

$$\text{cov}(N_1, Q) = E_x[(N_1 - \langle N_1 \rangle)(Q - \langle Q \rangle)]. \quad (\text{B14})$$

Expanding the product

$$\text{cov}(N_1, Q) = E_x[N_1 Q] - \langle N_1 \rangle \langle Q \rangle. \quad (\text{B15})$$

Expanding the first term

$$\begin{aligned} E_x[N_1 Q] &= E_x[E_x[N_1 Q | N_1 = n]] \\ &= \sum_{n_2=0}^{\infty} \sum_{n_1=0}^{\infty} n_1 E_x \left[\sum_{i=1}^{n_1} E_i + \sum_{i=1}^{n_2} E_i \right] \\ &\quad \times \text{Prob}\{N_1 = n_1\} \text{Prob}\{N_2 = n_2\}. \end{aligned}$$

The E random variables have same probability distribution so

$$E_x \left[\sum_{i=1}^{n_1} E_i \right] = n_1 \langle E \rangle_1,$$

where $\langle E \rangle_1$ denotes the mean value of the instances of E in the first bin. Similarly,

$$E_x \left[\sum_{i=1}^{n_2} E_i \right] = n_2 \langle E \rangle_2,$$

$$E_x[N_1 Q] = \sum_{n_2=0}^{\infty} \sum_{n_1=0}^{\infty} [n_1^2 \langle E \rangle_1 + n_1 n_2 \langle E \rangle_2] \\ \times \text{Prob}\{N_1 = n_1\} \text{Prob}\{N_2 = n_2\},$$

$$E_x[N_1 Q] = \langle N_1^2 \rangle \langle E \rangle_1 + \lambda_1 \lambda_2 \langle E \rangle_2$$

where the last step uses the fact that N_1 and N_2 are independent. The second moment of a Poisson random variable is

$$\langle N_1^2 \rangle = \lambda_1^2 + \lambda_1.$$

The final result for the first term in Eq. (B15) is

$$E_x[N_1 Q] = (\lambda_1^2 + \lambda_1) \langle E \rangle_1 + \lambda_1 \lambda_2 \langle E \rangle_2.$$

In the second term in Eq. (B15), $\langle N_1 \rangle = \lambda_1$ and $\langle Q \rangle = (\lambda_1 + \lambda_2) \langle E \rangle$. The overall mean value $\langle E \rangle$ can be expressed in terms of the bin mean values as

$$\langle E \rangle = \frac{\lambda_1 \langle E \rangle_1 + \lambda_2 \langle E \rangle_2}{\lambda_1 + \lambda_2}$$

so the second term is

$$\langle N_1 \rangle \langle Q \rangle = \lambda_1^2 \langle E \rangle_1 + \lambda_1 \lambda_2 \langle E \rangle_2.$$

Combining the first and second terms gives the final result

$$\text{cov}\{N_1, Q\} = \lambda_1 \langle E \rangle_1.$$

By the argument stated above, this can be generalized to

$$\text{cov}(N_k, Q) = \lambda_k \langle E \rangle_k.$$

As previously derived, the variance of Q is $\sigma_Q^2 = \lambda \langle E^2 \rangle$, so the covariance matrix is

$$\text{Cov}(\{N_k, k = 1, \dots, K\}, Q) \\ = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & \lambda_1 \langle E \rangle_1 \\ 0 & \lambda_2 & 0 & 0 & \lambda_2 \langle E \rangle_2 \\ 0 & 0 & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_K & \lambda_K \langle E \rangle_K \\ \lambda_1 \langle E \rangle_1 & \lambda_2 \langle E \rangle_2 & \dots & \lambda_K \langle E \rangle_K & \lambda \langle E^2 \rangle \end{bmatrix}.$$

5. Covariance of logarithm of NKQ measurements

The same arguments used in the previous section are also applicable to the logarithms of the measurements, so I will derive the covariance of $\log(N_1)$ and $\log(Q)$ for the two group case. Using the general result

$$\text{cov}[\log(N_1), \log(Q)] = E_x[\log(N_1) \log(Q)] \\ - E_x[\log(N_1)] E_x[\log(Q)].$$

As with the derivation for the NQ detector, the first term is

$$E_x[\log(N_1) \log(Q)] \approx \log[\langle N \rangle] \log[\langle Q \rangle] + \frac{\text{cov}(N_1, Q)}{\langle N_1 \rangle \langle Q \rangle}.$$

Inserting the results from the previous section,

$$\frac{\text{cov}(N_1, Q)}{\langle N_1 \rangle \langle Q \rangle} = \frac{\langle E \rangle_1}{\langle Q \rangle}.$$

Combining these

$$\text{cov}[\log(N_1), \log(Q)] = \frac{\langle E \rangle_1}{\langle Q \rangle}.$$

Using the result from the NQ detector for the variance of $\log(Q)$ and recalling that $F = \langle E^2 \rangle / \langle E \rangle^2$ and $\langle Q \rangle = \lambda \langle E \rangle$, the covariance matrix is

$$\text{Cov}(\{\log(N_k), k = 1, \dots, K\} \log(Q)) \\ = \frac{1}{\lambda} \begin{bmatrix} \lambda/\lambda_1 & 0 & \dots & 0 & \frac{\langle E \rangle_1}{\langle E \rangle} \\ 0 & \lambda/\lambda_2 & 0 & 0 & \frac{\langle E \rangle_2}{\langle E \rangle} \\ 0 & 0 & \dots & 0 & \vdots \\ 0 & 0 & 0 & \lambda/\lambda_K & \frac{\langle E \rangle_K}{\langle E \rangle} \\ \frac{\langle E \rangle_1}{\langle E \rangle} & \frac{\langle E \rangle_2}{\langle E \rangle} & \dots & \frac{\langle E \rangle_K}{\langle E \rangle} & F \end{bmatrix}.$$

^{a)}Electronic mail: ralvarez@aprendtech.com

¹M. J. Tapiovaara and R. Wagner, "SNR and DQE analysis of broad spectrum x-ray imaging," *Phys. Med. Biol.* **30**, 519–529 (1985).

²R. N. Cahn, B. Cederstrom, M. Danielsson, A. Hall, M. Lundqvist, and D. Nygren, "Detective quantum efficiency dependence on x-ray energy weighting in mammography," *Med. Phys.* **26**, 2680–2683 (1999).

³J. Giersch, D. Niederlohnner, and G. Anton, "The influence of energy weighting on x-ray imaging quality," *Nucl. Instrum. Methods Phys. Res. A* **531**, 68–74 (2004).

⁴D. Niederlohnner, J. Karg, J. Giersch, M. Firsching, and G. Anton, "Practical aspects of energy weighting in x-ray imaging," Nuclear Science Symposium Conference Record, 2004 IEEE, 2004, Vol. 5, pp. 3191–3194.

⁵R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerized tomography," *Phys. Med. Biol.* **21**, 733–44 (1976).

⁶H. L. Van Trees, *Detection, Estimation, and Modulation Theory. Part I: Detection, Estimation, and Linear Modulation Theory* (Wiley, New York, 1968).

⁷S. M. Kay, *Fundamentals of Statistical Processing, Volume I: Estimation Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1993).

⁸S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1998).

⁹E. Roessl, A. Ziegler, and R. Proksa, "On the influence of noise correlations in measurement data on basis image noise in dual-energy-like x-ray imaging," *Phys. Med. Biol.* **34**, 959–966 (2007).

¹⁰E. Kraft, P. Fischer, M. Karagounis, M. Koch, H. Kruger, I. Peric, N. Wermes, C. Herrmann, A. Nascetti, M. Overdick, and W. Rutten, "Counting and integrating readout for direct conversion x-ray imaging concept, realization and first prototype measurements," IEEE Nuclear Science Symposium Conference Record, 2005, Vol. 5, pp. 2761–2765 (unpublished).

¹¹L. A. Lehmann, R. E. Alvarez, A. Macovski, W. R. Brody, N. J. Pelc, S.

- J. Riederer, and A. L. Hall, "Generalized image combinations in dual KVP digital radiography," *Med. Phys.* **8**, 659–667 (1981).
- ¹²R. E. Alvarez and E. J. Seppi, "A comparison of noise and dose in conventional and energy selective computed tomography," *IEEE Trans. Nucl. Sci.* **NS-26**, 2853–2856 (1979).
- ¹³R. D. Evans, *The Atomic Nucleus* (McGraw-Hill, New York, 1955).
- ¹⁴E. Roth, "Multiple-energy selective contrast material imaging," Ph.D. dissertation, Stanford University, 1984.
- ¹⁵H. H. Barrett and K. Myers, *Foundations of Image Science* (Wiley, Hoboken, NJ, 2003).
- ¹⁶E. Parzen, *Stochastic Processes* (Holden-Day, San Francisco, CA, 1999).
- ¹⁷D. L. Snyder, *Random Point Processes* (Wiley, New York, 1975).
- ¹⁸R. E. Alvarez, "Extraction of energy dependent information in radiography," Ph.D. dissertation, Stanford University, 1976.
- ¹⁹E. Kraft, "Counting and integrating microelectronics development for direct conversion x-ray imaging," Ph.D. dissertation, Bonn University, 2008.
- ²⁰B. R. Whiting, P. Massoumzadeh, O. A. Earl, J. A. O'Sullivan, D. L. Snyder, and J. F. Williamson, "Properties of preprocessed sinogram data in x-ray computed tomography," *Med. Phys.* **33**, 3290–3303 (2006).
- ²¹J. Wang, H. Lu, Z. Liang, D. Eremina, G. Zhang, S. Wang, J. Chen, and J. Manzione, "An experimental study on the noise properties of x-ray CT sinogram data in radon space," *Phys. Med. Biol.* **53**, 3327–3341 (2008).
- ²²R. M. Heiberger and B. Holland, *Statistical Analysis and Data Display* (Springer, Berlin, 2004).
- ²³B. R. Whiting, L. E. Antonuk, and M. J. Yaffe, "Signal statistics in x-ray computed tomography," *Medical Imaging 2002: Physics of Medical Imaging* [Proc. SPIE 4682, 53–60 (2002)].
- ²⁴W. Fulks, *Advanced Calculus: An Introduction to Analysis* (Wiley, New York, 1969).
- ²⁵N. A. Dyson, *X-Rays in Atomic and Nuclear Physics*, 2nd ed. (Cambridge University Press, Cambridge, 1990).
- ²⁶P. Tothill, "The ratio of k characteristic to total radiation emitted from a tungsten target x-ray tube," *J. Phys. D: Appl. Phys.* **1**(9), 1093–1107 (1968).
- ²⁷J. H. Hubbell and S. M. Seltzer, Tables of X-Ray Mass Attenuation Coefficients and Mass Energy-Absorption Coefficients, National Institute of Standards and Technology (online), available at <http://physics.nist.gov/PhysRefData/XrayMassCoef/cover.html>.
- ²⁸M. Berger, J. Coursey, M. Zucker, and J. Chang, Stopping-Power and Range Tables for Electrons, Protons, and Helium Ions, National Institute of Standards and Technology (online), available at <http://physics.nist.gov/cgi-bin/Star/compos.pl?matno=001>.
- ²⁹E. Roessl and R. Proksa, "K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors," *Phys. Med. Biol.* **52**, 4679–4696 (2007).
- ³⁰T. G. Schmidt, "Optimal image-based weighting for energy-resolved CT," *Med. Phys.* **36**, 3018–3027 (2009).
- ³¹H. H. Barrett and W. Swindell, *Radiological Imaging* (Academic, New York, 1981).
- ³²A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, New York, 1965).