X-ray measurement model incorporating energy-correlated material variability and its application in information-theoretic system analysis

YIJUN DING,¹,∗ AND AMIT ASHOK,¹,2

¹Wyant College of Optical Sciences, University of Arizona, Tucson AZ 85721
²Department of Electrical and Computer Engineering, University of Arizona, Tucson AZ 85721
∗dingy@email.arizona.edu

Abstract: Extending our prior work, we propose a multi-energy X-ray measurement model incorporating material variability with energy correlations to enable the analysis and exploration of the performance of X-ray imaging and sensing systems. Based on this measurement model, we provide analytical expressions for bounds on the probability of error, \( P_e \), to quantify the performance limits of an X-ray measurement system for binary classification task. We analyze the performance of a prototypical X-ray measurement system to demonstrate the utility of our proposed material variability measurement model.

© 2020 Optical Society of America

1. Introduction

Imaging and sensing based on X-ray attenuation is commonly used to non-destructively discriminate materials in security screening, medical imaging and industrial inspection [1–3]. An X-ray measurement model with accurate statistics of the measurement data is desirable for many purposes such as, evaluating system performance, developing detection classification algorithms, and optimization of object-reconstruction algorithms.

The performance of attenuation-based X-ray imaging and sensing systems is limited by at least two fundamental factors in the measurement data: shot noise and inherent material variability [4]. The shot noise stems from the randomness in the generation, attenuation and detection of X-ray photons. The material variability, arising from the inherent fluctuations in material composition and density, limits the discrimination of materials. Therefore, a rigorous evaluation of an X-ray imaging or sensing system must take into account both shot noise and material variability.

An X-ray measurement model that considers only the shot noise has been analyzed in the context of material-discrimination applications by Huang et. at and Lin et. al. [5,6]. Recently, Masoudi et. at. [7] improved the model by incorporating material variability under the assumption of energy (statistical) independence. However, the energy-correlations are intrinsic in X-ray attenuation and cannot be ultimately ignored. In this paper, we propose an X-ray measurement model that considers energy-correlated material variability as well as shot noise.

Many imaging systems are used to perform binary-classification tasks. For example, a luggage scanner at an airport checkpoint is used to determine whether a luggage bag contains threat material or not; mammography is often used to classify tumor-present or tumor-absent; in industrial non-destructive evaluation, radiographs of parts are used to examine the existence of defects. An objective assessment of an system must take into account the task of the system [4,8].

Task-specific information (TSI) [8], which is the information content relevant to the task, is commonly used as an objective assessment metric. For classification tasks, Shannon mutual information (I_S) [9] is a natural choice as an information-theoretic metric, because it can be used to bound probability of classification error (\( P_e \)) [9–13]. Although I_S is expensive to compute for non-trivial distributions, we are able to derive closed-form expressions for bounds on I_S and...
bounds on $P_e$ for mixture distributions with the help of a recent work [14].

This paper is organized as follows. We derive the measurement model in Sec 2. Sec 3 reviews $I_S$, $P_e$ for binary-classification tasks and derives closed-form expressions for bounds on $I_S$ and bounds on $P_e$. In Sec 4, as an example application, we apply the model to a simplified luggage-scanner and present the simulation results. Sec 5 discusses the advantages and drawbacks of the model. Sec 6 provides a succinct conclusion.

2. Measurement model

The measurement model relates the data to the object and describes the statistical properties of the data. A general form of the measurement model can be written as

$$ g = \mathcal{H} f + n = \mathcal{H}(\bar{f} + \Delta f) + n, $$

where $g$ is the data, $\mathcal{H}$ describes the system, $f$ is the object, and $n$ is the system noise. $\bar{f}$ is the ensemble mean of the object, and $\Delta f$ describes the material variability.

2.1. X-ray attenuation coefficient $\mu$

In the energy range commonly used for X-ray transmission imaging, the interaction between X-ray photons and the medium can be categorized into the following three processes: photoelectric absorption, Compton scattering and coherent (Rayleigh) scattering. In photoelectric absorption, a photon disappears and the energy of the photon transfers to an electron in the material; in Compton scattering, an X-ray photon is deflected and transfers a portion of its energy to an electron; and in coherent scattering, an X-ray photon is deflected, but retains its energy. The strength of each interaction process can be characterized by the energy of the photon $E$, the atomic number of the medium $Z$ and the density of the medium $\rho$. For a material with fixed $Z$ and $\rho$, the X-ray attenuation coefficient $\mu(E)$ is a function of the X-ray energy $E$. Variations in $\mu(E)$ due to variability in $Z$ and $\rho$ demonstrate intrinsic energy correlation.

For multi-element compounds and mixtures, the attenuation coefficient is

$$ \mu(E) = \rho \sum_c w_c \mu_c(E) $$

where $\rho$ represents the density of the medium, $w_c$ is the weight fraction of the $c^{th}$ element in the compound or mixture, $\mu_c(E)$ and $\rho_c$ are the attenuation and the density of the $c^{th}$ element. When there is variability inherent in the description of a material, which may stem from density fluctuations, composition variations and packaging differences, the attenuation coefficient $\mu(E)$ becomes a random process.

To consider material variation, we assume $\mu(E)$ is a Gaussian random process with covariance function $\Sigma_\mu(E, E')$. For any set of energy $E_1, E_2, ..., E_R$, the random variable $\mu(E_1), \mu(E_2), ..., \mu(E_R)$ are joint Gaussian random variables. If we denote the set of random variables $\mu(E_i)$ by the vector $\mu$, whose mean is $\mu_0$ and the covariance matrix is $\Sigma_\mu$, then the joint probability density is

$$ p_r(\mu) = N(\mu | \mu_0, \Sigma_\mu) = \frac{\exp[-\frac{1}{2}(\mu - \mu_0)^T \Sigma_\mu^{-1}(\mu - \mu_0)]}{\sqrt{2\pi^R |\Sigma_\mu|}}, $$

where $|\Sigma_\mu|$ is the determinant of $\Sigma_\mu$, $(\cdot)^T$ is the transpose of matrix $(\cdot)$, $R$ is the length of the vector $\mu$. When $\Sigma_\mu$ is singular, the Gaussian distribution turns into a Dirac delta function in the corresponding dimension.
2.2. X-ray attenuation $\tau$

Now consider a beam of X-ray penetrating multiple items, as illustrated in Figure 1, where each item contains one material. If there are a total of $N_{it}$ items between the X-ray source and the detector, the total attenuation along the path is

$$\tau(E) = \int \mu(E, l) \, dl = \sum_{t=1}^{N_{it}} \mu_t(E) l_t,$$

where $t$ is the index for items along the X-ray path, $l_t$ is the length of the material along the path, and $\mu_t(E)$ is the attenuation profile of the material in the $t^{th}$ item. In the following discussion, we refer to $\tau(E)$ as the total attenuation.

Under the assumption that the fluctuations of $\mu(E)$ in two different items are independent, $\tau(E)$ is a Gaussian random processes, since it is the sum of independent Gaussian random processes, $\mu_t(E)$. For a set of energies $E_1, E_2, ..., E_K$, the set of total attenuation $\tau$ follows a multivariate normal distribution

$$\text{pr}(\tau) = \mathcal{N}(\tau | \tau_0, \Sigma_\tau) = \frac{\exp[-\frac{1}{2}(\tau - \tau_0)^T \Sigma_\tau^{-1}(\tau - \tau_0)]}{\sqrt{2\pi^R | \Sigma_\tau |}},$$

where $R$ is the length of the vector $\tau$ and

$$\tau_0 = \sum_{t=1}^{N_{it}} \mu_{0,t} l_t,$$

$$\Sigma_\tau = \sum_{t=1}^{N_{it}} \Sigma_{\mu,t} l_t^2.$$

2.3. Beer's law

An illustration of the propagation of X-ray photons starting from the X-ray source, through the object, and ending with photon detection is shown in Figure 1. The X-ray attenuation follows beer’s law,

$$J(E) = \frac{N_0 S(E)}{t} e^{-\tau(E)},$$

where $J(E)$ is the mean spectral flux incident on a detector element at energy $E$, $N_0 S(E)/t$ is the source spectral flux, $\tau(E)$ is the total attenuation as a function of $E$, and $t$ is the exposure.
time. More specifically, $N_0$ is the number of photons emitted from the X-ray tube in the solid angle extended by a detector element over the exposure time $t$, and $S(E)$ is the normalized x-ray source spectrum. The units of $J(E)$ and $S(E)$ are $(\text{s} \cdot \text{keV})^{-1}$ and keV$^{-1}$, respectively. We denote $	au(E) = \tau_0(E) + \Delta \tau(E)$, where $\Delta \tau(E)$ is fluctuation or perturbation in attenuation around $\tau_0(E)$, and $J_0(E) = N_0 S(E) e^{-\tau_0(E)}/t$. When the material variance (perturbation) is small, or more specifically, when $\Delta \tau(E) \ll 1$, we can approximate $e^{-\Delta \tau(E)}$ with the first two terms in the Taylor expansion, resulting in:

$$J(E) = J_0(E) - J_0(E) \Delta \tau(E).$$

(9)

The remainder of the approximation can be bounded by

$$|R_2(E)| < \frac{J_0(E)|\Delta \tau(E)|^3}{6}.$$  

(10)

If a requirement on the remainder is to be less than 1% of $J_0(E)$, $|\Delta \tau(E)|$ should be less than 0.39. One can always guarantee such a requirement by splitting a material with large variations into multiple materials with smaller variations. For example, 40% sugar water with sugar content varying from 20% to 60% can be split into two materials with sugar content varying from 20% to 40% and 40% to 60%, respectively.

In the limit of small variability, the two-term approximation of $J(E)$, as defined by Equation (9), follows a normal distribution when $\tau(E)$ follows a normal distribution. Therefore, $J$, which is $J(E)$ at a set of energies, approximately follows a normal distribution:

$$\text{pr}(J) = \mathcal{N}(J|J_0, \Sigma_J).$$

(11)

The mean and covariance matrix are

$$J_0 = \frac{N_0 S}{t} \otimes e^{-\tau_0},$$

(12)

and

$$\Sigma_J = (J_0 J_0^T) \otimes \Sigma_\tau,$$

(13)

where $\otimes$ is element-wise multiplication.

2.4. Detector response and energy binning

If the energy response of the detector is linear, one can express the mean photon count collected in the $m^{th}$ energy bin as

$$J_m = t \int_0^\infty J(E) D_m(E) dE,$$

(14)

where $D_m(E)$ is the detector response of the $m^{th}$ energy bin to a photon with energy $E$ and $t$ is the exposure time.

With a total of $M$ energy bins, the mean photon count $\{J_1, J_2, ..., J_M\}$ after energy binning can be represented by a vector $J_d$, and

$$J_d = t \mathcal{D} J,$$

(15)

where $\mathcal{D}$ is the detector response operator and the subscript $d$ denotes detector.

When $\mathcal{D}$ is a linear operator and $J(E)$ is a Gaussian random process, $J_d$ follows a normal distribution. The mean and variance are

$$J_{d0} = t D J_0$$

(16)

and

$$\Sigma_{J_d} = t^2 D \Sigma_J D^T,$$

(17)

where $D$ is the matrix form of the operator $\mathcal{D}$ for a set of $R$ energies.
2.5. Shot noise

In attenuation-based X-ray imaging, the data collected is often the number of detected X-ray photons. Photon counting intrinsically introduces shot noise, hence the number of X-ray photons detected in one energy bin, \( g \), follows a Poisson distribution:

\[
\Pr(g | J_d) = \text{Poiss}(g | J_d) = \frac{(J_d)^g e^{-J_d}}{g!},
\]

where \( \text{Poiss} \) indicates a Poisson distribution and \( J_d \) is the mean photon count in the energy bin.

When the mean photon count is relatively large (i.e. more than say, 10 photons), the Poisson distribution can be approximated by a Gaussian/normal distribution. Denoting the continuous variable \( x = g \),

\[
\Pr(x | J_d) \approx N(x; J_d, J_d)
\]

\[
\exp\left[-\frac{(x - J_d)^2}{2J_d}\right]\sqrt{\frac{2\pi}{J_d}}.
\]

When the mean photon count \( J_d \geq 100 \), the error introduced by the Gaussian approximation is less than 4% for \( x = \bar{x} \pm \sigma(x) \), where \( \bar{x} \) and \( \sigma(x) \) are the mean and standard deviation of \( x \), respectively. A derivation of the percentage error is given in Appendix B.

2.6. Combined model

In our measurement model, we combine statistics of the shot noise and the statistics of the energy-correlated material variability. Thus, the covariance of the measurement data is a summation of the covariance matrices corresponding to the shot noise and the material variability. More specifically, if we denote the continuous data vector as \( x \), the probability density function (PDF) of \( x \) with mean photon counts, \( J_{d0} \), and a covariance matrix induced by material-variation, \( \Sigma_{J_d} \), is

\[
\Pr(x | J_{d0}, \Sigma_{J_d}) \approx N(x; J_{d0}, \Sigma_{J_d} + \text{diag}(J_{d0})).
\]

A detailed derivation is provided in Appendix C.

We define the combined data over all detector pixels as \( g \). The probability distribution function of \( g \) is the joint distribution of all \( x_n \),

\[
\Pr(g) = \Pr_{x_1, x_2, \ldots, x_N}(x_1, x_2, \ldots, x_N),
\]

where \( n \) is the detector-pixel index and \( N \) is the number of detector pixels. The joint distribution is determined by the geometry of the object and setup of the system.

Now, consider an ensemble of \( K \) objects that are examined by an X-ray system. Then the data representing the ensemble of objects can be described by the following mixture distribution:

\[
\Pr(g | a) = \sum_{i=1}^{K} a_i \Pr_i(X),
\]

where \( a_i \) is the probability of the \( i \)th object occurring in the ensemble, \( K \) is the number of objects in the ensemble, \( \sum a_i = 1 \), and \( \Pr_i(X) \) is the probability distribution function of measurement data \( X \) when the \( i \)th object is imaged. Note that we have used the discretized data \( g \) and the continuous variables \( \{x_1, x_2, \ldots, x_N\} \) interchangeably.

3. TSI for binary classification tasks

Shannon mutual information, \( I_S \), has long been used as a metric to quantify the task-specific fidelity of a measurement with respect to classification tasks [8]. This is because \( I_S \) is related to the error probability, \( P_e \) through Fano’s inequality [10] and Kovalevskij’s inequality [11–13].
In this section, we provide a brief summary of the relation between $I_S$ and $P_e$ and provide closed-form expressions for bounds on $I_S$ and $P_e$ defined on our X-ray measurement model.

The system performance is object dependent. Properties of the object, such as the size, the material and the geometry, affect the distribution of the data and hence the difficulty of the classification task. To reduce the dependence on test objects, a general assessment of a system should consider a large ensemble of objects. In the following discussion, we consider an ensemble of $K$ objects that consists of $K_1$ objects in the first class and $K_2$ objects in the second class. The probability of the $i^{th}$ object in the ensemble is $a_i$. The probabilities of the two class labels are $P_1$ and $P_2$, where $P_1 + P_2 = 1$ for binary classification.

### 3.1. Bounds on $I_S$

The $I_S$ is defined as

$$I_S(g; C) = H(g) - \sum_{c=1}^{2} P_c H(g|C = c),$$

where $H(g)$ is the Shannon entropy of the distribution of the measured data, and $H(g|C = c)$ is the entropy of the conditional distribution of the measured data given the class $C$.

When the data is a mixture distribution, the $I_S$ between data and class label has no closed-form expressions. However, $I_S$ can be bounded by bounds of the entropy [14],

$$I_S(g; C) \geq \hat{H}_{BD}(g) - \sum_{c=1}^{2} P_c \hat{H}_{KL}(g|C = c)$$

and

$$I_S(g; C) \leq \hat{H}_{KL}(g) - \sum_{c=1}^{2} P_c \hat{H}_{BD}(g|C = c),$$

where $\hat{H}_{BD}(g)$ and $\hat{H}_{KL}(g)$ are lower and upper bounds on entropy based on pair-wise Bhattacharyaa distance (BD) and pair-wise Kullback-Leibler (KL) divergence, respectively. More specifically, the bound on entropy based on either divergence is given by

$$\hat{H}_D(g) = \sum_{i=1}^{K} a_i H(p_i) - \sum_{i=1}^{K} a_i \ln \left( \sum_{j=1}^{K} a_j \exp(-D(p_i, p_j)) \right),$$

where $p_i$ is the PDF of the data $g$ if the $i^{th}$ bag is measured and $D(p, q)$ can be either Bhattacharyaa distance or KL divergence, which are defined by

$$BD(p, q) = -\ln \int dx \sqrt{p(x)q(x)},$$

and

$$KL(p, q) = \int dx p(x) \ln \frac{p(x)}{q(x)},$$

respectively.

An upper bound on $I_S$ based on pair-wise KL divergence and a lower bound based on pair-wise BD are provided in Appendix A. The minimum of the two upper bounds and the maximum of the two lower bounds can be used as the tighter version of the bounds on $I_S$.

### 3.2. Bounds on $P_e$

Starting from $I_S$, the Fano’s inequality [10] provides a lower bound on $P_e$ for binary classification, as following

$$P_e \geq h^{-1}_q[H(C) - I_S(g; C)],$$

where $h_q$ is the q-th entropy function.
where \( h_b(x) = -x \log_2(x) - (1 - x) \log_2(1 - x) \) is the binary entropy function, \( h_b^{-1}(\cdot) \) is the inverse function of \( h_b(\cdot) \), and \( H(C) = h_b(P_C) \) is the Shannon entropy of the class label \( C \). More specifically, one can calculate \( P_e \) by placing the value \( H(C) - I_s(g; C) \) on the left side of the binary entropy function and solving for \( x \). When \( P_e < 1 \), \( H(C) - I_s(g; C) \approx -P_e \log P_e \) and hence is on the same order of magnitude as \( P_e \).

An upper bound on binary classification errors \( P_e \), which is tighter than Kovalevskij’s inequality, has been reported recently [13],

\[
P_e \leq \min \left\{ P_{\min}, f_{ub}^{-1}[H(C) - I_s(g; C)] \right\},
\]

(28)

where \( P_{\min} = \min\{P_1, P_2\} \), and \( f_{ub}(x) \) is an upper bound function defined by

\[
f_{ub}(x) = -P_{\min} \log_2 \frac{P_{\min}}{x + P_{\min}} - x \log_2 \frac{x}{x + P_{\min}}.
\]

(29)

and \( f_{ub}^{-1}(\cdot) \) is the inverse function of \( f_{ub}(\cdot) \).

### 3.3. Closed-from expressions for pair-wise BD and KL divergence

If we assume that the measurement data at different pixels are statistically independent with each other, the PDF of the measurement data becomes a product of the PDFs of the data measured at all pixels:

\[
pr(g) = \prod_{n=1}^{N} pr(x_n | J_{d0,n}, \Sigma_{j,i,n}) \approx \prod_{n=1}^{N} N(x_n ; J_{d0,n}, \Sigma_{j,i,n}),
\]

(30)

where \( n \) is the detector-pixel index, and \( N \) is the total number of detector pixels.

Calculation of bounds on \( I_s \) and \( P_e \) require pair-wise Bhattacharyya distance and pair-wise KL divergence. To simplify notation, we define

\[
\Delta J_n = J_{d0,n,i} - J_{d0,n,j},
\]

(31)

\[
\Sigma_{n,i} = \Sigma_{j,i,n,i} + \text{diag}(J_{d0,n,i}),
\]

(32)

and

\[
\Sigma_n = \Sigma_{n,i} + \Sigma_{n,j}.
\]

(33)

The analytical form of Bhattacharyya distance can be expressed as

\[
\text{BD}(pr_i, pr_j) = \sum_{n=1}^{N} \left[ \frac{\Delta J_n^T \Sigma_n^{-1} \Delta J_n}{4} - \frac{\ln |\Sigma_{n,i}\Sigma_{n,j}|}{4} + \frac{\ln |\Sigma_n|}{2} - \frac{M \ln 2}{2} \right];
\]

(34)

and the analytical form of KL divergence can be expressed as

\[
\text{KL}(pr_i, pr_j) = \frac{1}{2} \sum_{n=1}^{N} \Delta J_n^T (\Sigma_{n,j})^{-1} \Delta J_n - \ln |\Sigma_{n,i}| + \ln |\Sigma_{n,j}| + \text{tr}(\Sigma_{n,j}^{-1} \Sigma_{n,i}) - M,
\]

(35)

where \( \text{tr}(\cdot) \) is the trace of the matrix.

### 4. Illustrative System Study and Results

In this section, we apply our measurement model to study a simple X-ray measurement system for the task of material-based threat detection (i.e. a binary classification problem). The X-ray system, as illustrated in Figure 2, has 10 X-ray sources that produce parallel pencil-beams and 10 corresponding photon-counting energy-sensitive detector elements. A source with a tungsten
target operating at 160 kVp is assumed, and the corresponding source spectrum was generated with SpekCalc [?]. The energy-sensitive detector can have 1, 2 and 3 energy bins. The bin edges are determined by balancing the photon count after attenuation. More specifically, the bin edges are [30, 160] keV for one bin, [30, 70, 160] keV for two bins, and [30, 60, 85, 160] keV for three bins. The objects under inspection contain 10 vials of materials, and the location of each vial is along one parallel-beam X-ray path. Each vial contains 4 materials, which is randomly sampled from a library of materials. The lengths of the vials are randomly chosen between 0.5 cm to 20 cm.

A material library and a variability model of the material composition has been previously developed [7]. The material library contains 25 threat materials and 33 non-threat materials. Examples of threat materials are ammonium nitrite, hydrogen peroxide and gun powder; examples of non-threat materials are milk, toothpaste and polyethylene. The composition and variance of weight fractions of each material were determined based on industrial standards. The density variation was folded in either by varying the density $\rho$ or by adding air as a component. For each material, 1000 composition realizations were randomly generated; and for each material realization, an X-ray attenuation profile was computed based on the NIST XCOM database [?]. For each material, the mean and covariance of the 1000 attenuation profiles were used as the mean and the covariance of $\mu(E)$ in our model. The number of energy samples in $\mu$ and the source spectrum is $R = 180$.

We simulate objects, where each object has 10 vials, in pairs and each pair consists of one object containing one threat material and one object containing no threat material. A threat object and a non-threat object in a pair share the same geometry (aka. vial lengths and materials) and are thus different by only one material. An illustration of an object pair is shown in Figure 2. We simulate objects in pairs because of the following two reasons: (1) the system performance is dominated by objects that are located close to the class boundary in the data space; and (2) in general, the distance in the data space between two objects in a pair is closer than that between two random objects.

An ensemble of 160 bag-pairs was simulated. Equal prevalence of each class and equal probability of occurrence of each bag-pair were assumed for the calculation of the TSI measure.

There is no spatial correlation between data measured from different detector elements in this prototypical example. The closed-form expressions derived under the assumption of spatial independence were used to calculate the bounds on $I_S$ and $P_e$. First, we study the behavior of bounds on $I_S$ and bounds on $P_e$ for one, two and three energy bins. Second, we compare the following three measurement models in terms of bounds on $P_e$: (1) shot noise only, (2) material variation only, and (3) combined model. Lastly, we compare the energy-correlated measurement model and material-variation model with the energy-uncorrelated models in terms of lower bound on $P_e$. 

![Fig. 2: Illustration of the simulated X-ray luggage scanner.](image)
4.1. Behavior of TSI metrics

In this section, we present results calculated with the energy-correlated measurement model that incorporates both shot noise and material variation and illustrate the behavior of the TSI metrics.

Fig. 3: Upper bound (blue) and lower bound (red) on $I_S$, $H(C) - I_S$ and detection error probability $P_e$ for one (top), two (middle) and three (bottom) energy bins. The maximum possible value of $P_e$, which is 0.5 for equal prior, is plotted in black dashed line.

Figure 3 presents bounds on $I_S$ and $P_e$ as functions of source photon budget $N_0$ for one, two and three energy bins, respectively. The upper bound (blue solid line) and lower bound (red dashed line) of $I_S$, $H(C) - I_S$, and $P_e$ are shown from left to right in each figure. Note that for binary classification tasks with equal prevalence, $H(C) = 1$. The $H(C) - I_S$ is plotted, because it
is in the same order of magnitude with $P_e$ when $P_e \ll 1$.

For a binary classification task with equal prevalence of the two classes, $I_S \in [0, 1]$ and $P_e \in [0, 0.5]$. The task-specific performance improves when $I_S$ increases and $P_e$ decreases. When the source photon budget $N_0$ increases, both upper bound and lower bound of $I_S$ increases, which means that, as source count increases, the imaging system captures more information for the binary classification task. However, both $I_S$ and $P_e$ saturates at high photon regime, which is due to material variability.

![Fig. 4: Bounds on detection error rate $P_e$ for 1, 2, and 3 energy bins.](image)

To show $P_e$ for all three energy binning occasions, Figure 4 presents the bounds on $P_e$ for 1, 2, and 3 bins. At a fixed source count $N_0$, bounds on $P_e$ decrease as the number of energy bins increases. In fact, for systems with only one energy bin, the upper bound on $P_e$ is 0.5 and the lower bound saturates at around 0.25, which means that binary classification with such a system is close to random guess. The specific number of bounds on $P_e$ changes with the material library and other simulation setups, but the general trend that more energy bins provides higher task-specific information is valid.

### 4.2. Three measurement models

Figure 5 shows the lower bound and the upper bound of $P_e$ for three measurement models as a function of source photon number ($N_0$). The three measurement modes consist of a model that considers only the shot noise (approximated as Gaussian, dot-dashed lines), a model that considers only the energy-correlated material variation (dashed lines), and a model that considers both the shot noise and the energy-correlated material variation (solid lines). Results for detectors with 1, 2 and 3 energy bins are presented in different colors. At low count region, the system performance is limited by the quantum noise as the bounds on $P_e$ of the combined model is similar to that of the model considering only quantum noise; at high count region, the system performance is limited by the material variation as the bounds on $P_e$ of the combined model approaches that of the model considering only material variation.

### 4.3. Energy-correlated model vs. energy-uncorrelated model

We compare the energy-correlated models with the energy-uncorrelated models in Figure 6. From top to bottom, results of detectors with 1, 2 and 3 energy bins are presented. In each figure, the results calculated from the material-variation model are also presented for reference (red dashed line). When there is one energy bin, the bound on $P_e$ of the energy-uncorrelated model is
lower than that of the correlated model. When the number of energy bins equals to 2, the bound on $P_e$ of the uncorrelated model is close to that of the correlated model. When the number of energy bins equals to 3, the bound on $P_e$ of the uncorrelated model is much higher than that of the correlated model.

The difference in bound on $P_e$ calculated from the energy-uncorrelated model and the energy-correlated model is due to the difference in the material variation considered in the two models, as indicated by the difference in the red-dashed lines. When energy correlation in the data is considered, $\Sigma_J$, which is the covariance matrix of $J(E)$, is calculated based on Equation (13); in contrary, when energy correlation is not considered, the off-diagonal elements of $\Sigma_J$ are ignored. Since the energy-binning process is an integration over energy, each element of $\Sigma_J$ is the sum of a block matrix in $\Sigma_J$. For example, when there is one energy bin, $\Sigma_{J\delta}$ is the sum of all elements in $\Sigma_J$ when energy correlation is considered; while in the uncorrelated model, $\Sigma_{J\delta}$ is the sum of only the diagonal elements of $\Sigma_J$. Therefore, $\Sigma_{J\delta}$ of the energy-correlated model is larger than that of the uncorrelated model for the one energy bin scenario.

When there are more than one energy bin, the comparison is more complex and we introduce a method to quantitatively compare the material variability model with energy correlation and that without energy correlation. The size of the material variability can be quantified by the volume of the noise bubble induced by the material variability. If we define the volume of the noise bubble as the volume of a $M$-dimensional ellipsoid determined by the covariance matrix $\Sigma_{J\delta}$, where $M$ is the number of energy bins. The eigenvector of the covariance matrix from singular value decomposition provides the principle axes of the ellipsoid and the square root of the eigenvalues are half the length of the principle axes. For example, when energy correlation is not considered, $\Sigma_{J\delta}$ is diagonal with each element equals to the variance of detected photon count in the corresponding energy bin; the square root of an eigenvalue is the standard deviation of the detected photon count; and the volume of the ellipsoid determined by the covariance matrix is the product of the standard deviation in each energy bin.

To compare the material variability estimated by the energy-correlated model and uncorrelated model, we can calculate the ratio of the two volumes determined by the two ellipsoids. Denote this ratio as $r = V_{corr}/V_{uncorr}$. In the dataset used in Figure 6, $\log_{10}(r)$ is $1.00 \pm 0.07$ (mean $\pm$
Fig. 6: Comparison of the energy-correlated model with energy-uncorrelated model in terms of lower bound on $P_e$. The material-variation-only model (blue lines) and the full measurement model (red dashed lines) are considered.
standard deviation) for one energy bin, 0.98 ± 0.45 for two energy bins and −1.08 ± 1.05 for three energy bins. For the set of bags we studied, the uncorrelated model always underestimates the material variability for one energy bin; and with three energy bins, the uncorrelated model often severely overestimates the material variability.

5. Discussion

To incorporate material variation, a linear perturbation of the beer's law is used in the derivation from Equation 8 to Equation 9. This perturbation is equivalent to approximating the sum of correlated log-normal random variables with a normal distribution, which has been justified in Section 2 for small material variations. One direction of future work is to examine other approximations of the sum of log-normal random variables and developing a measurement model that accommodates large material variations. Such a model could further reduce the number of materials in the material library and hence reduce the number of objects needed in a study.

At low count region, the measurement model should be applied with caution, due to the break down of using Gaussian distribution to approximate the statistics of the shot noise. In comparison, Masoudi et. at. [7] incorporated the Poisson distribution in an energy-uncorrelated model. Their model performances better than our model, when shot noise is dominating and material variation is not important. However, many X-ray systems are not photon starving and the effect of material variation should be considered. To address the concern about the accuracy of the Poisson-Gaussian approximation, we provide an error analysis in Appendix B.

We assumed the attenuation profiles, \( \mu(E) \), follow normal distributions and derived that the measurement data \( g \) following normal distributions under the following two assumptions: (1) The material variation \( \Delta \tau \) is small, and (2) the distribution of the shot noise can be approximated by a Gaussian. Although our derivation relied heavily on the normality of data, the mean and covariance matrix of the data \( g \) are independent to the shape of the distributions, as long as the above assumption (1) is still valid.

The main goals of this work include: presenting the framework, justifying the approximations used in the measurement model, and studying the effect of energy correlation. In future work, we will incorporate spatial correlation and other sources of variations, such as source fluctuation and detector variation.

6. Conclusion

In this work, we have presented an energy-correlated X-ray measurement model that incorporates both material variation and shot noise. Energy correlations are inherent in a material’s attenuation profile and affect the system performance. Therefore, it is important to consider the energy correlations. We successfully modeled the shot noise and energy-correlated material variation by a multivariate Gaussian model. Furthermore, under the assumption of no spatial correlation, we provided analytical forms for TSI metrics, including bounds on \( I_S \) and bounds on \( P_e \), for binary classification tasks. Spatial correlation of the X-ray measurement data is not explored in this work and will be a focus of future study.

Funding

The authors gratefully acknowledge the support of the US Department of Homeland Security. The research for this project was conducted under contract with the U.S Department of Homeland Security (DHS) Science and Technology Directorate (S&T), contract HSHQDC-16-C-B0014. The opinions contained herein are those of the contractors and do not necessarily reflect those of DHS S&T.
Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. K. Wells and D. Bradley, “A review of x-ray explosives detection techniques for checked baggage,” Appl. Radiat. Isot. 70, 1729–1746 (2012).
2. J. Hsieh, Computed Tomography Principles, Design, Artifacts, and Recent Advances (John Wiley & Sons, 2009).
3. R. Hanke, T. Fuchs, and N. Uhlmann, “X-ray based methods for non-destructive testing and material characterization,” Nucl. Instruments Methods Phys. Res. Sect. A: Accel. Spectrometers, Detect. Assoc. Equip. 591, 14–18 (2008).
4. H. H. Barrett, “Objective assessment of image quality: effects of quantum noise and object variability,” JOSA A 7, 1266–1278 (1990).
5. J. Huang and A. Ashok, “Information optimal compressive x-ray threat detection,” in Computational Optical Sensing and Imaging, (Optical Society of America, 2015), pp. CTh2E–4.
6. Y. Lin, G. G. Allouche, J. Huang, A. Ashok, Q. Gong, D. Coccarelli, R.-I. Stoian, and M. E. Gehm, “Information-theoretic analysis of x-ray photoabsorption based threat detection system for check-point,” in Anomalous Detection and Imaging with X-Rays (ADIX), vol. 9847 (International Society for Optics and Photonics, 2016), p. 98470F.
7. A. Masoudi, J. Voris, D. Coccarelli, J. Greenberg, M. Gehm, and A. Ashok, “X-ray measurement model and information-theoretic system metric incorporating material variability (conference presentation),” in Anomalous Detection and Imaging with X-Rays (ADIX) III, vol. 10632 (International Society for Optics and Photonics, 2018), p. 106320H.
8. M. A. Neifeld, A. Ashok, and P. K. Baheti, “Task-specific information for imaging system analysis,” JOSA A 24, B25–B41 (2007).
9. T. M. Cover and J. A. Thomas, Elements of information theory (John Wiley & Sons, 2012).
10. R. M. Fano and D. Hawkins, “Transmission of information: A statistical theory of communications,” Am. J. Phys. 29, 793–794 (1961).
11. V. Kovalevskij, “The problem of character recognition from the point of view of mathematical statistics,” Character Readers Pattern Recognit. (1967).
12. D. Tebbe and S. Dwyer, “Uncertainty and the probability of error (corresp.),” IEEE Transactions on Inf. Theory 14, 516–518 (1968).
13. B.-G. Hu and H.-J. Xing, “An optimization approach of deriving bounds between entropy and error from joint distribution: Case study for binary classifications,” Entropy 18, 59 (2016).
14. A. Kolchinsky and B. D. Tracey, “Estimating mixture entropy with pairwise distances,” Entropy 19, 361 (2017).