A statistical, task-based evaluation method for three-dimensional x-ray breast imaging systems using variable-background phantoms

Subok Park[a)] and Robert Jennings
Division of Imaging Applied Mathematics, OSEL/CDRH/FDA, Silver Spring, Maryland 20993

Haimo Liu
The University of Maryland, College Park, Maryland 20742

Aldo Badano and Kyle Myers
Division of Imaging Applied Mathematics, OSEL/CDRH/FDA, Silver Spring, Maryland 20993

Purpose: For the last few years, development and optimization of three-dimensional (3D) x-ray breast imaging systems, such as digital breast tomosynthesis (DBT) and computed tomography, have drawn much attention from the medical imaging community, either academia or industry. However, there is still much room for understanding how to best optimize and evaluate the devices over a large space of many different system parameters and geometries. Current evaluation methods, which work well for 2D systems, do not incorporate the depth information from the 3D imaging systems. Therefore, it is critical to develop a statistically sound evaluation method to investigate the usefulness of inclusion of depth and background-variability information into the assessment and optimization of the 3D systems.
Methods: In this paper, we present a mathematical framework for a statistical assessment of planar and 3D x-ray breast imaging systems. Our method is based on statistical decision theory, in particular, making use of the ideal linear observer called the Hotelling observer. We also present a physical phantom that consists of spheres of different sizes and materials for producing an ensemble of randomly varying backgrounds to be imaged for a given patient class. Lastly, we demonstrate our evaluation method in comparing laboratory mammography and three-angle DBT systems for signal detection tasks using the phantom's projection data. We compare the variable phantom case to that of a phantom of the same dimensions filled with water, which we call the uniform phantom, based on the performance of the Hotelling observer as a function of signal size and intensity.
Results: Detectability trends calculated using the variable and uniform phantom methods are different from each other for both mammography and DBT systems.
Conclusions: Our results indicate that measuring the system's detection performance with consideration of background variability may lead to differences in system performance estimates and comparisons. For the assessment of 3D systems, to accurately determine trade offs between image quality and radiation dose, it is critical to incorporate randomness arising from the imaging chain including background variability into system performance calculations. © *2010 American Association of Physicists in Medicine.* [DOI: 10.1118/1.3488910]

I. INTRODUCTION

For the last several years, there has been a lot of interest in the development of three-dimensional (3D) x-ray breast imaging systems, such as digital breast tomosynthesis (DBT) and computed tomography (CT), as well as in the optimization of these systems.[1–19] For example, Glick *et al.*[1] compared lesion detection accuracy between digital mammography, DBT, and cone-beam breast CT systems with use of an ensemble of simulated backgrounds with a power law spectrum and a receiver operating characteristic (ROC) study of five human observers. In another study, Glick *et al.*[2] employed a Fourier-domain based signal-to-noise ratio (SNR) that utilized the noise power spectrum (NPS) of the system and anatomical noise for investigating the impact of x-ray spectral shape on the image quality of a flat-panel breast CT system. Kwan *et al.*[3,4] studied x-ray scatter properties and spatial resolution properties of a cone-beam breast CT scan-

ner. Das *et al.*[5] investigated a variable dose DBT acquisition technique in terms of detection accuracy for breast masses and microcalcification clusters with a localization ROC (LROC) human observer study. Zhao and her colleagues[6–8] developed a cascaded linear system model for DBT to investigate the effects of detector performance, imaging geometry, and image reconstruction algorithms on the reconstructed image quality. Sechopoulos *et al.*[9,10] investigated scatter properties and glandular radiation dose in DBT systems using Monte Carlo simulation.

However, development of evaluation and optimization methods, as well as quality assurance and control (QA/QC) methods, for these systems has not caught up with advances in system development. Fourier-domain based quantitative measures, such as modular transfer function (MTF), detective quantum efficiency (DQE), NPS, and pixel SNR (pSNR), initially used for planar x-ray imaging systems are

currently being employed for evaluation of 3D breast imaging systems. Moreover, current QA/QC phantoms for planar mammography, such as the ACR mammography accreditation program (MAP) and CDMAM phantoms, do not probe 3D performance. These methods and phantoms work fairly well for evaluating systems when assumptions of the shift invariance of the system and the stationarity of the data statistics hold. However, these methods have limitations in that the system's shift invariance and the stationarity of the data statistics do not hold as well for the 3D systems.[20–22] Systematic and accurate incorporation of background variability into system evaluation is important for planar mammography,[23] and even more critical for the 3D systems because of the depth information arising from the 3D systems. Without adequate consideration of background variability and depth information in assessing the system's diagnostic performance, the 3D systems cannot be accurately evaluated and optimized for clinical tasks at hand. In addition, without further understanding what sorts of features for the background variability and 3D depth information need to be included in 3D evaluation methods, it would be difficult to develop QA/QC tests and phantoms that are more appropriate for the evaluation and calibration of the 3D imaging systems. Overcoming the limitations of the aforementioned 2D methods and developing appropriate 3D evaluation methods is an active area of current research in the breast imaging field.[7,8,12,14–16] In the meantime, human observer studies, while expensive and time consuming, are still the most relied upon method for a task-based assessment and optimization of the 3D systems. Unfortunately, it is simply infeasible to rely on human studies for an effective, rigorous system optimization over a large set of system as well as background and signal parameters. Therefore, the need for more comprehensive and effective system evaluation methods before a final clinical validation of system performance is significant.

In order to address the aforementioned issues, we advocate the use of a task-based approach to the assessment of image quality on the basis of statistical decision theory.[24] This approach requires a number of ingredients: (1) a relevant task of interest, (2) a model observer for performing the task, and (3) a clinically meaningful figure of merit for measuring observer (and hence system) performance. We note that it is important that the model observer used in this approach should make use of *necessary* and *sufficient* amount of statistical information in the data for a rigorous assessment and optimization of image quality. For rigorously performing the task-based, statistical assessment of image quality, a number of aspects need to be considered: (1) generation of realistic phantoms to be imaged, either virtual or physical, (2) accurate characterization and simulation of the imaging physics, and (3) improvement and appropriate use of task-based statistical assessment methods including the efficient implementation of model observers. For the past several years, the need for incorporating the task-based approach using model observers in the assessment of 3D breast imaging systems has been gaining ground in the medical imaging community.[25] However, the implementation of the model observer for use in the assessment of 3D x-ray imaging systems has not been sufficiently executed to address various scenarios in x-ray imaging due to the difficulty of accurately incorporating system geometry and physics as well as background and signal statistics into the model observer. Moreover, this difficulty gets exacerbated by the large data size problem inherent in digital x-ray imaging.

For improving the assessment of 3D breast imaging, much emphasis in the medical imaging community for the last several years has been in (1) generating anatomical virtual phantoms with patient data for use in system optimization via simulation[26,27] and (2) improving computational methods for accurate characterization of the x-ray imaging physics in simulation.[1,20,28–32] But, there is still much left to do in order to incorporate all necessary aspects of the imaging physics into the system assessment tools. In addition, creating virtual phantoms using patient data has a limitation in the sense that sample size is small, thus with these phantoms, it is difficult to address questions regarding different patient populations. However, if we are willing to sacrifice some of the realism of the resulting simulated phantom, it is feasible to create virtual phantoms without the sample size limitation.[33–35]

An alternative approach is to assess image quality using physical phantoms. In doing so, we have the ability to perform image acquisitions in reality rather than simplified models in silico. Therefore, it would be useful to develop physical phantoms that can produce randomly varying background images in the laboratory, of which the statistics resemble those in mammographic images. More importantly, the phantom should be designed so that it has the capability to create a large set of samples for a given patient class. With such a phantom, one can investigate various issues including how many samples of the mammographic background are required for a rigorous assessment of image quality in breast imaging. However, to date, such complex physical phantoms have not yet been utilized for 3D x-ray breast imaging, in particular, with the use of a task-based statistical assessment approach.

For examples of studies that made use of a task-based, statistical approach, Sechopoulos *et al.*[11] and Saunders *et al.*[13] employed a contrast-to-noise ratio (CNR), respectively, for the optimization of the acquisition geometry in DBT and the investigation of the impact of anatomical noise on breast compression in DBT. Both studies used different but randomly varying background phantoms simulated in computer in order to include more complex background statistics in their experiments than previous studies in the literature. However, their studies are limited in the sense that the background variability is not accounted for sufficiently in their calculation of system performance. For instance, Saunders *et al.*[13] only used three different phantoms to incorporate background variability, which is too small a sample size to draw any statistically significant conclusions. Moreover, the CNR is not a sufficient figure of merit for measuring the impact of background variability on the system's diagnostic performance because it does not sufficiently account for the background variability. By definition, the CNR is $\Delta c/(\sigma_b \bar{b})$,

where Δc is the contrast difference between the signal and the mean background, σ_b is the standard deviation of background image pixel values, and $\bar{b}$ is the mean background contrast. The contrast difference Δc in the numerator of the CNR can be an effective measure for indicating the contrast difference between the signal and the background in the case of uniform backgrounds. However, for the case of variable backgrounds, the contrast difference itself does not capture the difference between the signal and the background accurately due to the impact of anatomical noise on signal contrast even if it is normalized by the mean of the background image values. For instance, frequent occurrence of a concentration of anatomical features in a neighborhood of the signal location can increase the signal contrast, resulting in overestimated system performance using the CNR. In addition, measuring background complexity summarized in the form of the standard deviation of background image pixel values works well for backgrounds of which the statistics follow an independent and identically distributed (i.i.d.) Gaussian model. But this does not work well for realistic background statistics, which have more complex correlation between image pixels as well as different angular projections, such as in 3D breast imaging.

A better alternative is to calculate the performance of the Bayesian ideal observer, which makes use of all available statistical information in the data.[36] By design, the ideal observer sets an absolute upper bound for the performance of any observer, either human or model, and hence should be used for system optimization and evaluation whenever possible. However, it is difficult to estimate the performance of the ideal observer due to the high dimensionality and often unknown probability density functions of the data for many realistic applications including breast imaging. Developing computationally efficient algorithms for the estimation of the ideal observer is an active area of our research.[37–39] When no practical method to calculate the ideal observer is available, a good alternative is to use the linear ideal observer, which is also called the Hotelling observer, which uses the first and second order statistics of the data. High dimensionality of the data is often a bottleneck in accurately calculating the performance of the ideal observers. To reduce the dimension of the data but still approximate the performance of the Hotelling observer, a channelized Hotelling observer (CHO) can also be used with an *appropriate* choice of channels. A common figure of merit for the Hotelling and CHO observers is a task SNR, which makes use of the mean difference between the signal-present and signal-absent data as well as the inverse of the data covariance or a channelized version of the mean difference and data covariance. The Hotelling observer maximizes this SNR and hence is optimal among all linear observers in terms of this figure of merit.

Recently, Chawla *et al.*[12] implemented a CHO with Laguerre–Gauss channels, which has been shown to approximate the Hotelling observer in cases involving rotationally symmetric signals and stationary backgrounds[40] for the optimization of a DBT system with 25 different but fixed projection angles. They applied the CHO to the 25 individual pro-

jection cases, yielding 25 sets of decision variables (i.e., 25 decision variables per image) and 25 corresponding ROC curves. As an attempt to incorporate correlation between the 25 different angular projections, they combined the 25 decision variables in the following two ways and calculated the DBT system's detection performance. In their first approach, the 25 decision variables were linearly combined with weights defined with an assumption that a smaller angular separation from the center (zero-angle projection) provided a larger contribution to the correlation between the angular projections. However, this assumption is simply untrue. When an oblique-angle projection is closer to the central orientation, the backgrounds, as well as the signals, from both angles would appear more similar to each other, resulting in similar data statistics. This leads to a SNR similar to the SNR of the zero-angle projection case. This means that an oblique-angle projection *close* to the center may not add as much additional information for performing the detection task as a projection far away from the center. In the second approach, they assumed that the 25 decision variables per image were independent of each other, which implies that multiangle projections per patient were assumed to be at least uncorrelated. Therefore, with the second approach, spatial correlation between the angular projections was not at all incorporated in their SNR calculation.

In this paper, we present a statistical assessment framework for 3D breast imaging on the basis of statistical decision theory, in particular, using the Hotelling observer, and demonstrate the method for comparing laboratory mammography and three-angle DBT systems in signal detection tasks. In the following sections, we will describe a physical phantom that can produce an ensemble of variable-background images in the laboratory, of which the statistics are more complex than those of stationary backgrounds often used in the assessment of image quality in breast imaging. While this phantom does not provide short- and long-range variabilities of an actual 3D breast object, we employ this phantom to demonstrate the effect of object variability on image quality when using physical acquisitions instead of using simplified simulations. Previously, with use of this physical phantom and a uniform phantom of the same size filled with water, we investigated the impact of background variability on the performance of a laboratory mammography system using the Hotelling observer.[16]

In the current paper, we will describe an expanded mathematical framework for performing a task-based statistical assessment of 3D imaging systems. With the expanded framework, spatial correlation between angular projections as well as between image pixels within each projection can be incorporated into the system's diagnostic performance measures. We will demonstrate this evaluation framework by comparing the performance of the laboratory mammography and three-angle DBT systems for two different x-ray exposures and as a function of signal size and intensity. Our results will illustrate that when background variability is incorporated, the system's diagnostic performance yields different performance trends from the case when such information is not incorporated.

II. MATHEMATICAL BACKGROUND

II.A. Image formation and binary detection

In this work, we focus on binary signal detection for the assessment of mammography and DBT systems. For binary detection tasks, signal-present and signal-absent hypotheses are considered here, given by

$$H_1: \mathbf{g} = \mathbf{b} + \mathbf{s} + \mathbf{n}, \tag{2.1}$$

$$H_0: \mathbf{g} = \mathbf{b} + \mathbf{n}, \tag{2.2}$$

where the vectors $\mathbf{b}$ and $\mathbf{s}$ represent the noiseless background and signal images, respectively, $\mathbf{n}$ represents measurement noise, and $\mathbf{g}$ represents the resulting data vector. All the vectors are M-dimensional, i.e., M pixels per image. In our work, the noise is not additive, but the signal $\mathbf{s}$ is implemented to be additive and independent of the background.

II.B. Task performance with the Hotelling observer

To measure the performance of a given system, the Hotelling observer[24] that maximizes the task SNR can be employed for performing binary detection tasks. This observer uses the mean and covariance of the data to form its template $\mathbf{w_g}$ via

$$\mathbf{w_g} = \mathbf{K_g}^{-1} \Delta \mathbf{s}, \tag{2.3}$$

where $\mathbf{w_g}$ is an $M \times 1$ vector and $\mathbf{K_g}$ is an $M \times M$ matrix that is the average of the covariance matrices of the signal-absent and signal-present image data. In this work, we focus on a signal-known-exactly (SKE) task with the additive signal, so $\mathbf{K_g} = \mathbf{K}_{\mathbf{g}|H_0}$. The vector $\Delta \mathbf{s}$ is an $M \times 1$ vector that is the difference between the mean signal-present and signal-absent image data. That is,

$$\Delta \mathbf{s} = \bar{\mathbf{g}}_1 - \bar{\mathbf{g}}_0, \quad \bar{\mathbf{g}}_j = \langle \mathbf{g}|H_j \rangle, \quad j = 0, 1, \tag{2.4}$$

$$\mathbf{K_g} = \langle (\mathbf{g} - \bar{\mathbf{g}}_0)(\mathbf{g} - \bar{\mathbf{g}}_0)^t \rangle_{\mathbf{g}|H_0}, \tag{2.5}$$

where t denotes the transpose operator.

For the SKE binary detection case, in theory, the mean difference signal $\Delta \mathbf{s}$ is the same as the true signal $\mathbf{s}$. In practice, the mean signal is estimated and suffers from noise due to finite sample size effects. But, in this work, we have full knowledge of $\Delta \mathbf{s}$ since this is a SKE task, there is not uncertainty in the signal. In addition, we want to use a model observer that is as optimal as possible, so we use the true signal $\mathbf{s}$ for the mean difference signal $\Delta \mathbf{s}$.

With the Hotelling template $\mathbf{w_g}$, the signal-present and absent test statistics can be computed using

$$t_j(\mathbf{g}) = \mathbf{w_g}^t [\mathbf{g}|H_j], \tag{2.6}$$

where $[\mathbf{g}|H_j]$ is sampled from the hypothesis H_j. For measuring observer performance, the area under the ROC curve (AUC) can be estimated via the Wilcoxon statistic using the signal-absent and present test statistics, t_0 and t_1. Then the $\mathrm{SNR_{AUC}}$ can be computed via[24]

$$\mathrm{SNR_{AUC}} \equiv 2 \, \mathrm{erf}^{-1}(2\mathrm{AUC} - 1), \tag{2.7}$$

for two-alternative forced-choice (2AFC) signal-detection experiments, where erf^{-1} is the inverse of the error function. This quantity, $\mathrm{SNR_{AUC}}$, is also called the detectability index d_A. With signal intensity, AUC approaches 1, which yields infinite $\mathrm{SNR_{AUC}}$. To avoid this problem when the signal is strong, another definition of the SNR can be used, given by[24]

$$\mathrm{SNR}_t = \frac{\langle t_1 \rangle - \langle t_0 \rangle}{\sqrt{(\sigma_0^2 + \sigma_1^2)/2}}, \tag{2.8}$$

where σ_j, $j = 0, 1$, indicates the standard deviation of t_j. The SNR_t and $\mathrm{SNR_{AUC}}$ are equivalent when each of the test statistics, t_1 and t_0, follows a normal distribution. In our simulation study, SNR_t was used to produce observer-performance maps as a function of signal size and intensity.

When the dimensionality of the data $\mathbf{g}$ is large, estimation of the data covariance and its inverse can be a difficult problem, leading to unstable performance estimates. In such a situation, a linear transformation $\mathbf{T}$, which consists of N_c rows of channels, can be applied to $\mathbf{g}$ to reduce the dimension of the data,

$$\mathbf{v} = \mathbf{Tg}, \tag{2.9}$$

where $\mathbf{T}$ is an $N_c \times M$ matrix and $\mathbf{v}$ is an N_c-dimensional vector. A linear observer that uses the mean and the covariance of $\mathbf{v}$ is called a CHO. For the CHO, all the formuli given in this section can be used with replacement of all $\mathbf{g}$ related quantities with the corresponding $\mathbf{v}$ related vectors. We note that $\mathbf{T}$ can be chosen so that the CHO approximates either the Hotelling observer or the human observer, in which cases channels are, respectively, referred to as *efficient* and *anthropomorphic*. For system design and optimization, we advocate the use of efficient channels for approximating the Hotelling (or nonlinear ideal) observer whenever possible. However, this is an active area of research in the field, and efficient channels for mammographic images and signals have not yet been fully identified. Thus, in this work, we will make use of the Hotelling observer and a reasonably sized region of interest (ROI) to reduce instability in the estimation of the covariance and its inverse.

II.C. Variability and bias in observer-performance estimates

In human-observer studies, case variability arises when a different image set is used, and reader variability arises in the following two fashions: intra and inter. Intra-reader variability arises when the observer repeats the experiment, whereas inter-reader variability arises when different observers perform the same experiment. Similarly, variability in the performance of a model observer is caused by a number of sources such as the estimation of the template and the test statistic. The former and the latter can be regarded, respectively, as the *reader and case variabilities*. The intra-reader variability of a model observer is zero since the model observer is a computer program that can always perform the experiment exactly. The variance estimate of observer per-

formance can generalize to a similar experiment in which we draw a new set of readers and a new set of cases. This variance is often referred to as the *multiple-reader and multiple-case variance*. The closer the template and the test statistic estimates approximate the true template and the true population of the test statistic, respectively, the closer the performance estimate to the truth. In order to achieve this, a sufficient number of samples, which well represent the true population, should be used for calculating the template as well as the test statistic, which we call *training and testing*, respectively.

Bias in observer-performance estimates can be caused by ways used in estimating the covariance and the test statistic. Mainly, bias in the estimation of either the covariance or the test statistic occurs when a finite set of samples, which does not adequately represent the true population, is used for estimating either of the two quantities. Other sources of bias in observer-performance estimates include the size of a chosen ROI and the choice of channels if a CHO is used for performing the task. If the ROI size does not account for sufficient background statistics required for estimating the data covariance, bias in observer-performance estimates in comparison with the true performance is likely to be larger than the case when the ROI is large enough to incorporate complete statistical information regarding the background for the task. Similarly, if channels do not capture sufficient features in the data for performing the task (to approximate either the Hotelling or the human), bias in the performance of a CHO using these features is likely to be larger than the case when all necessary features are utilized in the CHO.

III. METHODS AND MATERIALS

III.A. Constructing physical phantoms

To incorporate background variability into the estimation of detection performance of our mammography and DBT systems, we employed the concept of the bead phantom developed by Hesterman *et al.*[41] and developed a physical phantom that consists of spheres of different sizes and densities for simulating tissue compositions and textures similar to those of the breast. In particular, we constructed a 9.5 cm(height) × 9.9 cm(width) × 5 cm(thickness) container filled with 35%, 25%, 31%, and 9%, by volume, of water and spheres of polyethylene, polymethyl methacrylate (PMMA), and nylon, respectively. The container walls were made of PMMA of thickness of 0.45 cm. The volume fraction for PMMA includes the two 9.5 cm(h) × 9 cm(w) container walls. The diameters of the spherical balls in the container ranged from about 6.4 to 16 mm. Table I summarizes characteristics of the different spherical balls used to construct this phantom. Figure 1 shows the linear attenuation coefficients for the different materials, including the phantom itself with the assumption of uniform structure, compared to the adipose and glandular tissue materials. These coefficients were obtained from the database of the National Institute of Standards and Technology.[42]

With this phantom, 200 configurations of the random background structure were realized by stirring the contain-

TABLE I. Specifications of the spheres used to construct the variable phantom.

Individual spheres	Diameter (cm)	Quantity	Weight (g)
Polyethylene (CH_2)	1.59	5	1.89
Polyethylene	1.91	4	3.23
Polyethylene	0.95	98	0.41
Polyethylene	0.79	96	0.25
Polyethylene	0.64	94	0.12
Nylon66 ($C_6H_{11}NO$)	0.95	9	0.52
Nylon66	1.59	3	2.35
Nylon66	0.79	83	0.30
Nylon66	0.64	44	0.15
PMMA ($C_5H_8O_2$)	0.64	42	0.16
PMMA	0.95	68	0.54
PMMA	1.3	17	1.28

Phantom summary	Volume (%)	Density (g/cm³)	Overall density (g/cm³)
Polyethylene (CH_2)	25	0.9114	1.04814
Water (H_2O)	35	1.0	
PMMA ($C_5H_8O_2$)	31	1.189	
Nylon66 ($C_6H_{11}NO$)	9	1.13	

er's contents and imaging each configuration. See Fig. 2 for two different tissue configurations of the same phantom and their 370 × 370 projection images. In addition, in order to compare the random background case to the uniform background case, we filled the same phantom container with water and imaged the phantom 35 times without stirring the contents. Throughout this paper, we call the phantoms of spheres and water the *variable* and *uniform* phantoms, respectively.

III.B. Tissue compositions of physical phantoms

To understand how the physical phantoms relate to real breast tissue compositions using adipose and glandular tissue
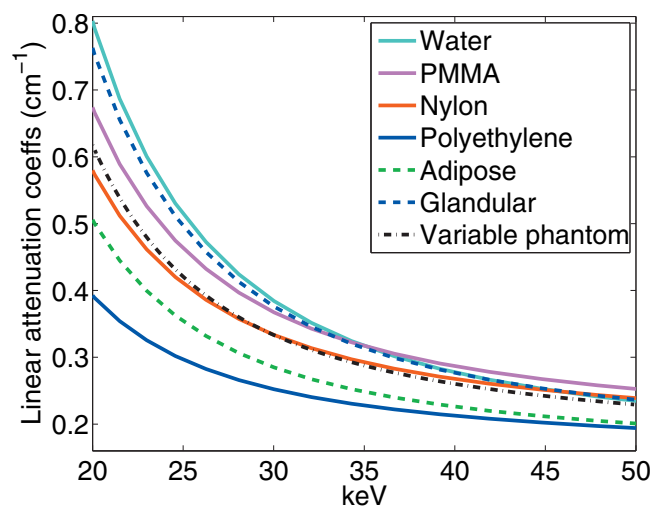

FIG. 1. Linear attenuation coefficients of individual phantom materials compared to those of the glandular and adipose tissues as well as the variable phantom.
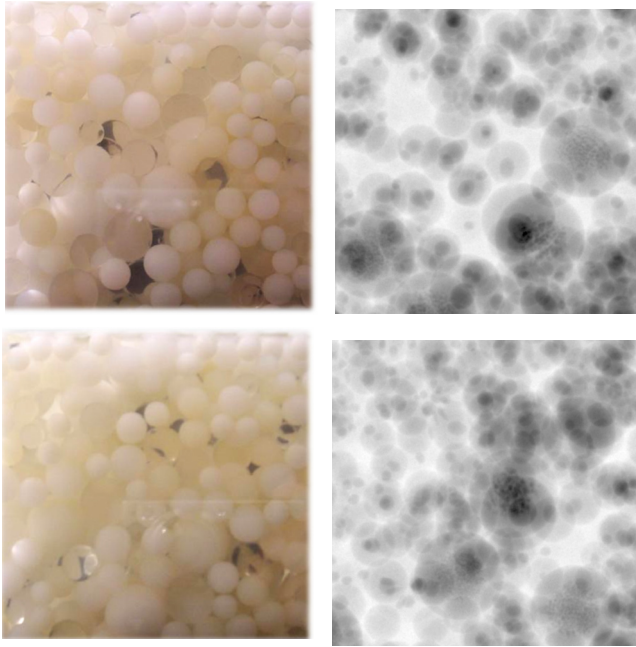
FIG. 2. The left column shows two different configurations of the variable phantom and the right shows the negative logarithm of their 370 ×370 (pixels) projection images.
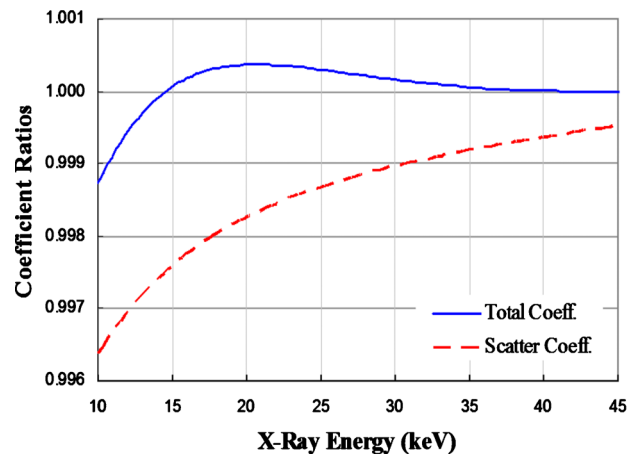

FIG. 3. The plot shows the ratios of x-ray attenuation coefficients of the variable phantom over a simulated phantom consisting of glandular and adipose blocks (33: 76), which translates to 30% glandular and 70% adipose tissue composition.

materials, we followed the method proposed by Jennings[43–45] for determining the thicknesses of adipose and glandular tissue layers that provide the closest match, in a minimum root-mean-square error sense, to the narrow-beam transmission of the variable phantom over the energy range 10–40 keV. For this analysis, the variable phantom was treated as a phantom that consists of a uniform mixture of the two PMMA container walls and the variable-phantom contents between the two walls. For others who are interested in reproducing the variable phantom and matching the breast tissue compositions, Table II summarizes the weight fractions of atomic materials of the uniform and variable phantoms as well as those of the glandular and adipose tissues, taken from Ref. 46, used in the analysis.

The variable phantom used in this work was found to be equivalent to a phantom of glandular and adipose blocks simulated by the aforementioned method, which we call the *simulated* phantom. The unnormalized ratios of the thicknesses for the glandular and adipose tissue blocks relative to

the variable phantom were found to be 0.33 and 0.76, which translated to a 30% glandular and 70% adipose tissue composition, by volume. The total thickness of the simulated phantom was about 1.09 times that (5 cm) of the variable phantom used in this work, which resulted in a 5.45 cm thick simulated phantom. Figure 3 shows the ratios of x-ray (total and scatter) attenuation coefficients of the simulated phantom over the variable phantom. The coefficients for the simulated phantom were adjusted by the ratio of the thickness of the simulated and variable phantoms to facilitate the comparison. This figure reveals that the x-ray attenuation properties of the two phantoms match well. In summary, our variable phantom of thickness of 5 cm has similar x-ray attenuation properties to those of the simulated phantom of 30% glandular and 70% adipose tissue composition and thickness of 5.45 cm.

The uniform phantom used in this study was found to be equivalent to a phantom of only glandular tissue with the same thickness as the uniform phantom. As shown in Fig. 1, there is a gap between the x-ray attenuation coefficients of the variable phantom and this glandular tissue only phantom although the gap decreases with x-ray energy. This affects the resulting gray levels in the x-ray projections of the uniform and variable phantoms and hence detection performance calculations using these projections. This aspect of the work will be further discussed in Sec. III D 3.

TABLE II. Overall density ρ (g/cm³) as well as weight fractions of atomic materials for the uniform phantom, variable phantom, glandular, and adipose tissues.

Z	Uniform phantom	Variable phantom	Glandular tissue	Adipose tissue
1	0.1053	0.1089	0.1020	0.1120
6	0.1260	0.4573	0.1840	0.6190
7	0.0000	0.0111	0.0320	0.0170
8	0.7687	0.4226	0.6770	0.2510
15	0.0000	0.0000	0.0050	0.0010
ρ	1.034	1.048	1.040	0.930

III.C. Experimental setup

The detector in this work was a Varian 4030CB (Varian Corp., Salt Lake City US) with 2048×1596, 195 μm pixels pixels, and a 600 μm thick columnar CsI (Tl) phosphor. The x-ray beam was generated at 40 kVp with a Varian B180 x-ray tube (with 0.6 mm focal spot and Be window) using a tungsten anode, Be window, and 1 mm Al added filtration.
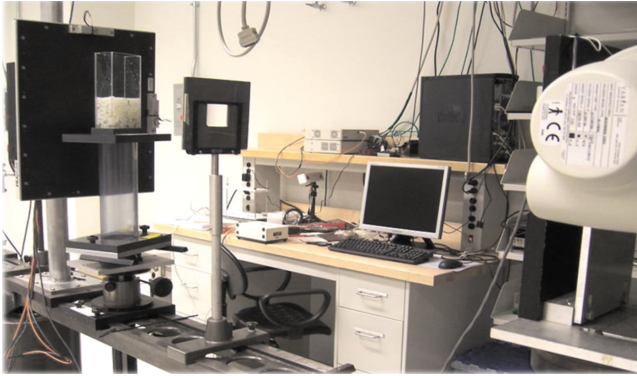
FIG. 4. Experimental setup for imaging the phantom. To produce angular projections, the phantom was rotated, and the detector and the source were fixed.

See Fig. 4 for the experimental setup. For the system geometry, the source-to-detector, detector-to-center of rotation, and center of rotation-to-source distances were 123, 14, and 109 cm. Note that the surface of the x-ray tube was used as the location of the source for the system geometry measurement. To create a simplified DBT system, a phantom rotator was used, fixing the detector and the source, for producing three different angular projections ($-20°$, $0°$, $20°$). The rotator was built in our laboratory, and it consisted of the rotator stage (Newport 481-A series) and an aluminum plate top, which was connected to the stage. The rotator allowed for $360°$ rotation, with an adjustment range accurate to $1°$. The height of the rotator was also adjustable in order to allow the x-ray beam to pass through the center of the phantom. The phantom to be imaged was placed on top of the aluminum plate. See Fig. 5 for a closeup of the phantom rotator.

The exposure was taken at the entrance surface of the phantom using an RTI Piranha 657 probe. In our experiment, two different dose conditions were used: (1) keeping the same dose per projection, which we call the *isodose per projection condition,* and (2) the same dose for both the systems, i.e., one-third of mammography dose per DBT projection,
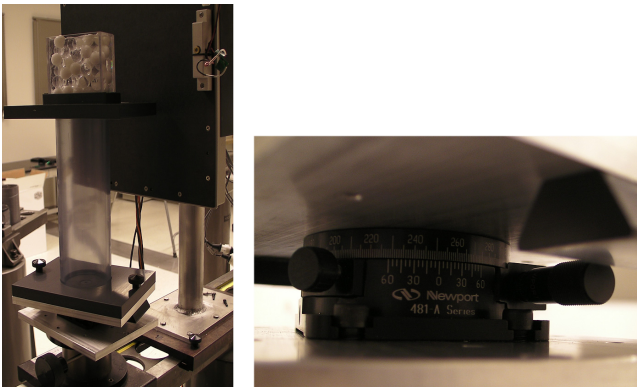

FIG. 5. On the left, the phantom rotator for producing angular projections and its closeup view on the right.

which we call the *isodose per modality condition.* This way we were able to compare the performance of mammography to that of the three-angle DBT system using the same as and triple the dose of the mammography system in the experiment.

The projection measurements for the isodose per modality condition were performed at a later time than the measurements for the isodose per projection condition. Therefore, there were slight differences in the two sets of the system parameters. In terms of exposures, for the isodose per projection condition, all projections, either DBT or mammography, were taken with 200 mA, 63 ms, and 33 mR. For the isodose per modality condition, DBT projections were taken with 63 mA, 63 ms, and 10 mR, which yields a total of 30 mR for the DBT system, and mammography projections for comparison were taken with 200 mA, 63 ms, and 33 mR, which is a little above the DBT dose.

In addition, for the isodose per modality condition, the source-to-detector, detector-to-center of rotation, and center of rotation-to-source were 123, 12, and 111 cm. To facilitate stirring the variable phantom contents, another 20 cm tall phantom container was used. This phantom container had 0.3 cm thick PMMA walls. The resulting dimensions of the variable phantom were 9.3 cm(height)$\times$10.2 cm(width) $\times$5.1 cm(thickness). We believe that these differences were not significant enough to change observer-performance trends.

All images were flat-field corrected via

$$g_m = (r_m - d_m)\rho_m, \tag{3.1}$$

where ρ_m, g_m, r_m, and d_m are the *m*th elements of the vectors $\boldsymbol{\rho}$, $\mathbf{g}$, $\mathbf{r}$, and $\mathbf{d}$. The vector $\boldsymbol{\rho}$ is the gain calculated by

$$\rho_m = \frac{\sum_{m=1}^M (f_m - d_m)}{f_m - d_m}, \tag{3.2}$$

where $\mathbf{d}$ is the mean dark field image and $\mathbf{f}$ is the mean flat field image. Lastly, the vectors $\mathbf{r}$ and $\mathbf{g}$ are the raw data taken in the laboratory and the resulting flat-field corrected image vector, respectively. In this work, for estimating the mean dark field and mean flat field, 100 dark field and 50 flat field images, and 35 dark field and 35 flat field images were used, respectively, for the isodose per projection and modality conditions. For the uniform phantom case, 35 flat field images were considered as the raw data, $\mathbf{r}$.

III.D. System performance analysis

III.D.1. Task performance for the 3D case

For assessing the DBT system, the ROIs of the angular projections can be concatenated, which was previously demonstrated by Young *et al.*,[14–16] to form a single data vector $\mathbf{g}$ for use in our performance analysis,

$$\mathbf{g} = \{\mathbf{g}_k\}_{k=1}^{N_p}, \tag{3.3}$$

$$\mathbf{g}_k = \{\mathbf{g}_{k,i}\}_{i=1}^M, \tag{3.4}$$

where N_p is the number of angular projections and M is the number of pixels in each ROI. In this work, $N_p = 3$ and M
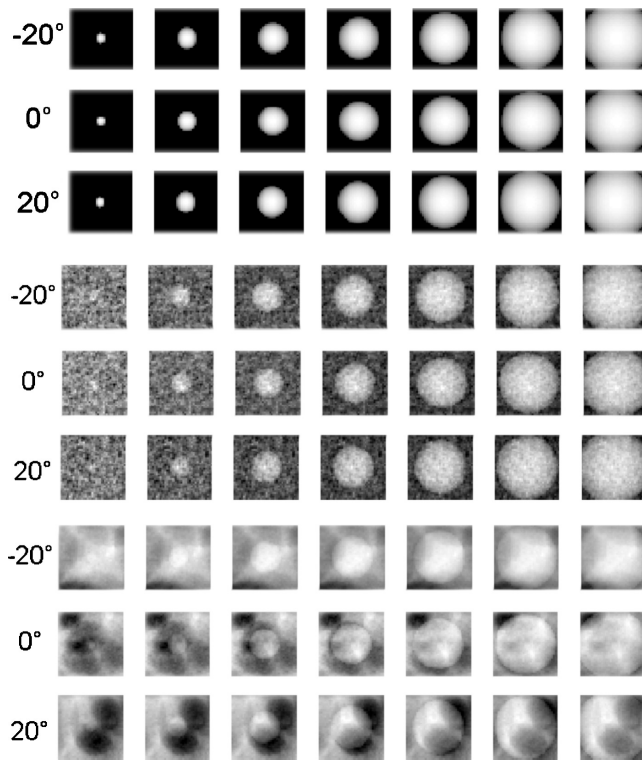
FIG. 6. The 2D projections of 3D spheres of different diameters used in this work (top) embedded in the background projections using the uniform (middle) and variable (bottom) phantoms. The projection angles are indicated in front of each row of the 2D projections. For either of the phantom cases, three different backgrounds were realized for the three different angles. For display, the signal intensity in the variable backgrounds was twice that in the uniform backgrounds.

$=37^2$. Here we call the Hotelling observer using the resulting concatenated vector the *3D projection (3Dp) Hotelling observer*. When appropriately used, this approach allows for incorporation of spatial correlation information between pixels in different angular projections as well as pixels within each projection.

Alternatively, channels that can extract spatial correlation

between angular projections can be applied to a collection of projections per patient to produce a channelized data, **v**, and reduce the dimension of the problem,

$$\mathbf{v} = \sum_{k=1}^{N_p} \mathbf{T}_k \mathbf{g}_k, \tag{3.5}$$

where **v** is an N_c-dimensional vector and each $\mathbf{T}_k$ is an $N_c \times M$ matrix consisting of appropriate channels for the given projection data $\mathbf{g}_k$. We call this observer the *3Dp CHO*. Then, the test statistic and SNR formuli using **v** given in Sec. II B can be used. Choosing appropriate channels for 2D as well as 3D background images to approximate the unchannelized ideal observer by a channelized observer is still one of our ongoing research activities,[37–39] some of which will be further discussed in Sec. V. Thus, in this work, we chose to use the approach with the 3Dp Hotelling observer for incorporating spatial correlation between different angular projections. In our simulation study, to simplify the process of choosing ROIs while obtaining sufficient statistics, ROIs from the projection images were used as independent background images, resulting in considering the spatial correlation more random than it actually is. Further discussion on this subject will also be presented in Sec. V.

III.D.2. Detection task study setup

For producing noisy background images of the variable phantom, we collected 200 $[370 \times 370 \ (\text{pixels})]$ projections of the phantom using the aforementioned imaging protocol. Then, we extracted 100 $[37 \times 37 \ (\text{pixels})]$ ROIs from each projection and hence obtained a total of 20000 ROIs for use in training and testing of the Hotelling observer. We note that the $[37 \times 37 \ (\text{pixels})]$ ROIs translated to 7.2 mm $\times 7.2$ mm ROIs, which may not be a sufficient size for clinically meaningful applications in breast imaging. However, the method in this work is stylized to demonstrate the use of the expanded evaluation framework and investigate the impact of background variability on system performance in

TABLE III. Diameters (mm) of the 3D spheres and their corresponding 2D projections. For each elliptical projection, σ_x and σ_x represent the lengths of the major and minor axes in mm. The 2D signal $\mathbf{s} = a_s \mathbf{p}$ and $\mathbf{p}_k$ indicates projections of the seven different 3D spheres.

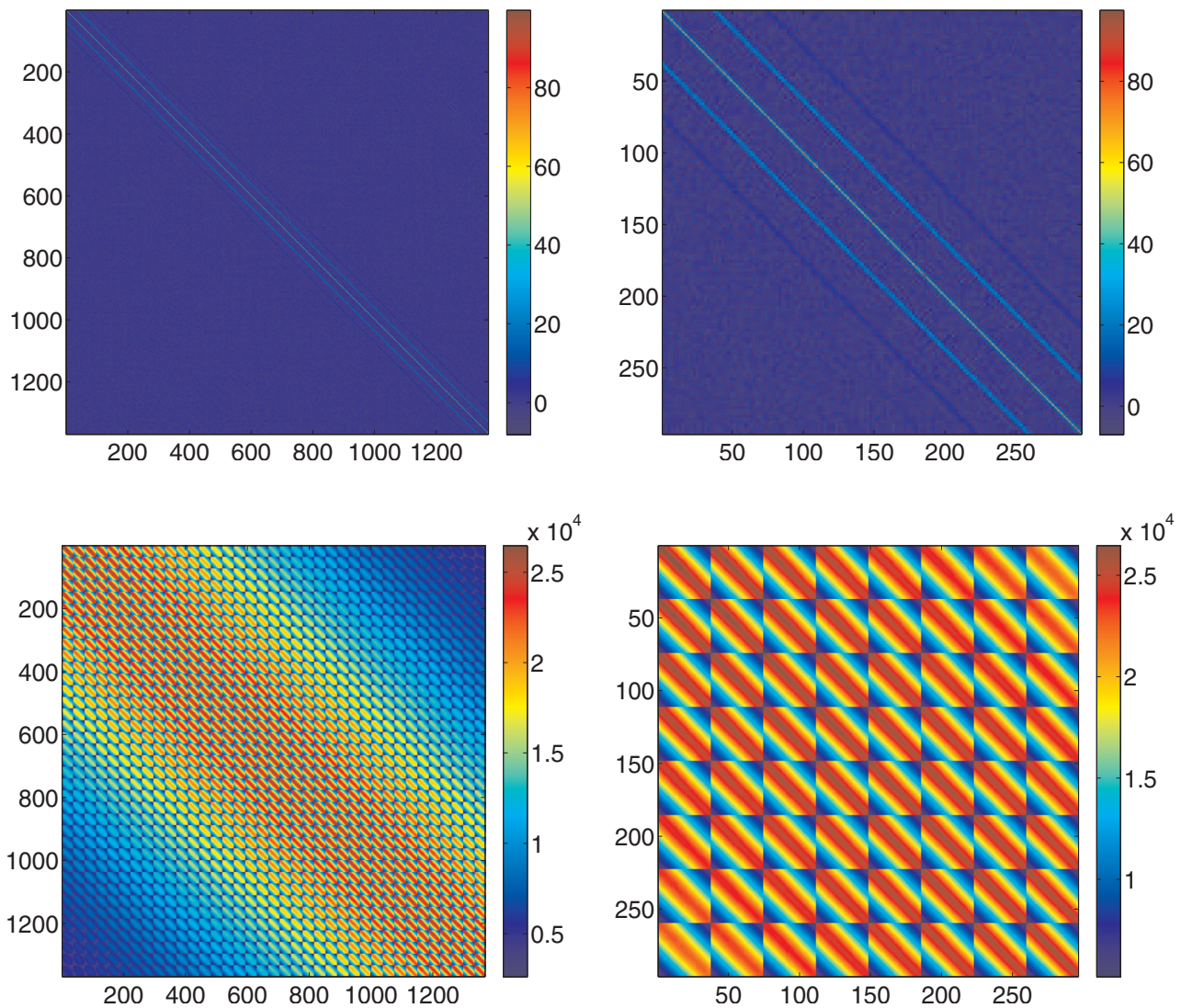| **p** | 0° projection | | ±20° projection | | 3D sphere |
	σ_x (mm)	σ_y (mm)	σ_x (mm)	σ_y (mm)	Diameter (mm)
$\mathbf{p}_1$	0.98	0.98	0.98	1.37	1.0
$\mathbf{p}_2$	2.15	2.15	2.15	2.54	2.0
$\mathbf{p}_3$	3.32	3.32	3.32	3.71	3.0
$\mathbf{p}_4$	4.49	4.49	4.49	4.88	4.0
$\mathbf{p}_5$	5.66	5.66	5.66	6.05	5.0
$\mathbf{p}_6$	6.83	6.83	6.83	7.22	6.0
$\mathbf{p}_7$	8.0	8.0	8.0	8.34	7.0

FIG. 7. Mammography: (top) 1369×1369 covariance estimate of projection images using the uniform phantom and the first 296×296 zoomed-in version; (bottom) 1369×1369 covariance estimate using the variable phantom and the first 296×296 zoomed-in version.

simple cases. For the case of the uniform phantom, 35 $[370 \times 370 \text{ (pixels)}]$ projections were collected, which translated to 3500 $[37 \times 37 \text{ (pixels)}]$ ROIs.

To generate signal-present images, first the 2D x-ray projections of 3D spherical lesions were generated in the form of a sum of the x-ray attenuation coefficients times the x-ray path lengths in the computer using the same geometry as our laboratory DBT system geometry. We set $\mathbf{s} = a_s \mathbf{p}$, where the vector $\mathbf{p}$ represents the 2D projection of a 3D sphere normalized by the maximum pixel value in each projection and varied a_s to create different signal magnitudes. Then the resulting signals ($\mathbf{s}$) were added to the negative logarithm of the projection images ($\mathbf{b}+\mathbf{n}$) measured in the laboratory. We note that the negative logarithm was taken so that both background and signal projections were in the same data format.

Note that the 2D signal projection was assumed to be statistically independent of the background projection images. See Fig. 6 for the 2D projections of the 3D spheres used in the work as well as those embedded in the back-

grounds produced using the uniform and variable phantoms. For either of the uniform and variable phantom cases, three different background images were, respectively, used for the three different angle cases shown in Fig. 6.

For the zero projection angle, the diameters of the 2D signal projections used in this study ranged from 0.98 to 8.0 mm, which was approximately equivalent to the diameters of the 3D spherical lesions ranging from 1 to 7 mm. Table III summarizes all the signal diameters for the three different angle cases. The size of the 3D sphere was limited by the size of the ROI chosen in this study, and the ROI size was chosen to use reasonable numbers of training and testing sets for having sufficient statistical power in the estimation of the task SNR. In addition, the largest 3D sphere diameters, 6 and 7 mm, were included to show the impact of background variability on observer performance using the uniform and variable phantom methods. Further discussion on this subject can be found in Sect. V. We note that the object was discretized before transfer through the simulated system mimicking the
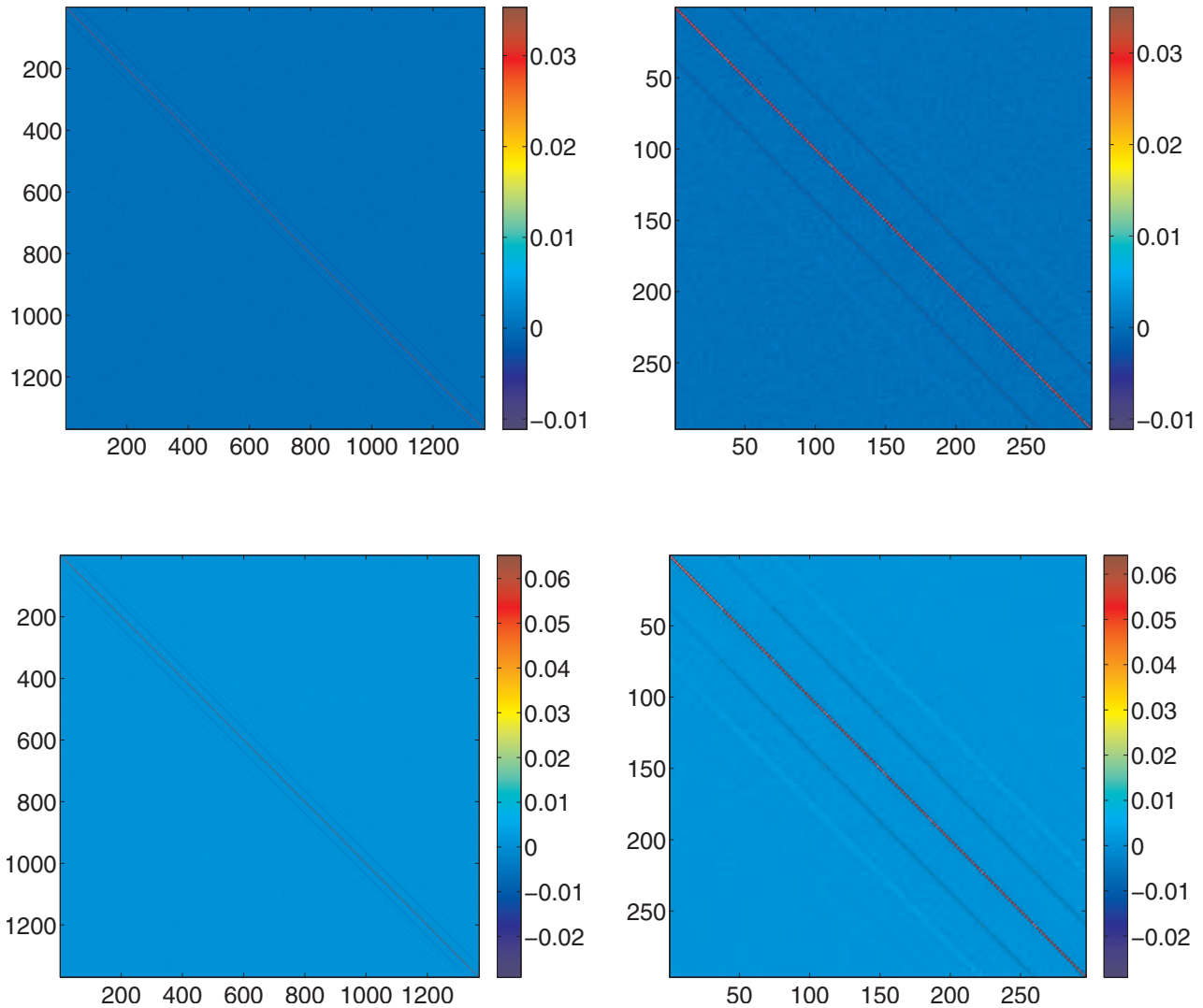
FIG. 8. Mammography: (top) inverse of the 1369×1369 covariance estimate of projection images using the uniform phantom and the first 296×296 zoomed-in version; (bottom) inverse of the 1369×1369 covariance estimate using the variable phantom and the first 296×296 zoomed-in version.

laboratory DBT system, so there is some discrepancy between the diameters of the 3D sphere and its resulting 2D projection.

III.D.3. Training and testing the observer

For training the 3Dp Hotelling observer, the covariance K_g for the variable phantom case was estimated using an independent set of 19000 signal-absent ROIs, whereas 3000 ROIs were used for the uniform phantom. See Figs. 7–9 for intensity plots of the covariance and inverse matrices of both the uniform and variable phantoms for the mammography and DBT systems. In this work, since we had full knowledge of the signals and intended to estimate the upper bound for the performance of the model observer, the true signal image (**s**) was used for Δ**s**. For testing the observer, in order to calculate t_1 and t_0 given in Eq. (2.6), the remaining 500 pairs of signal-present and signal-absent ROIs were used for the variable phantom case, whereas the remaining 250 ROI pairs

for the uniform phantom case. Signal intensity and diameter were varied to compare signal detectability for each system and background type.

In this work, as discussed in Sec. III B, the overall x-ray attenuation properties of the uniform and variable phantoms were slightly different, i.e., x rays going through the uniform phantom would attenuate more than through the variable phantom. This resulted in the mean gray levels of the negative logarithm of their projection images being slightly different, e.g., 7.51 and 7.67, respectively, for mammography projections of the uniform and variable phantoms. In our simulation, the signal intensity values used for adding the signal to both the variable and uniform backgrounds were kept the same. However, in practice, if an actual 3D signal was inserted in both the uniform and variable phantoms, and then the phantoms were imaged, the 2D signals in the resulting variable-phantom projections would have had higher contrast than those in the uniform-phantom projections. In fact, if we were to assume that the variable phantom has a
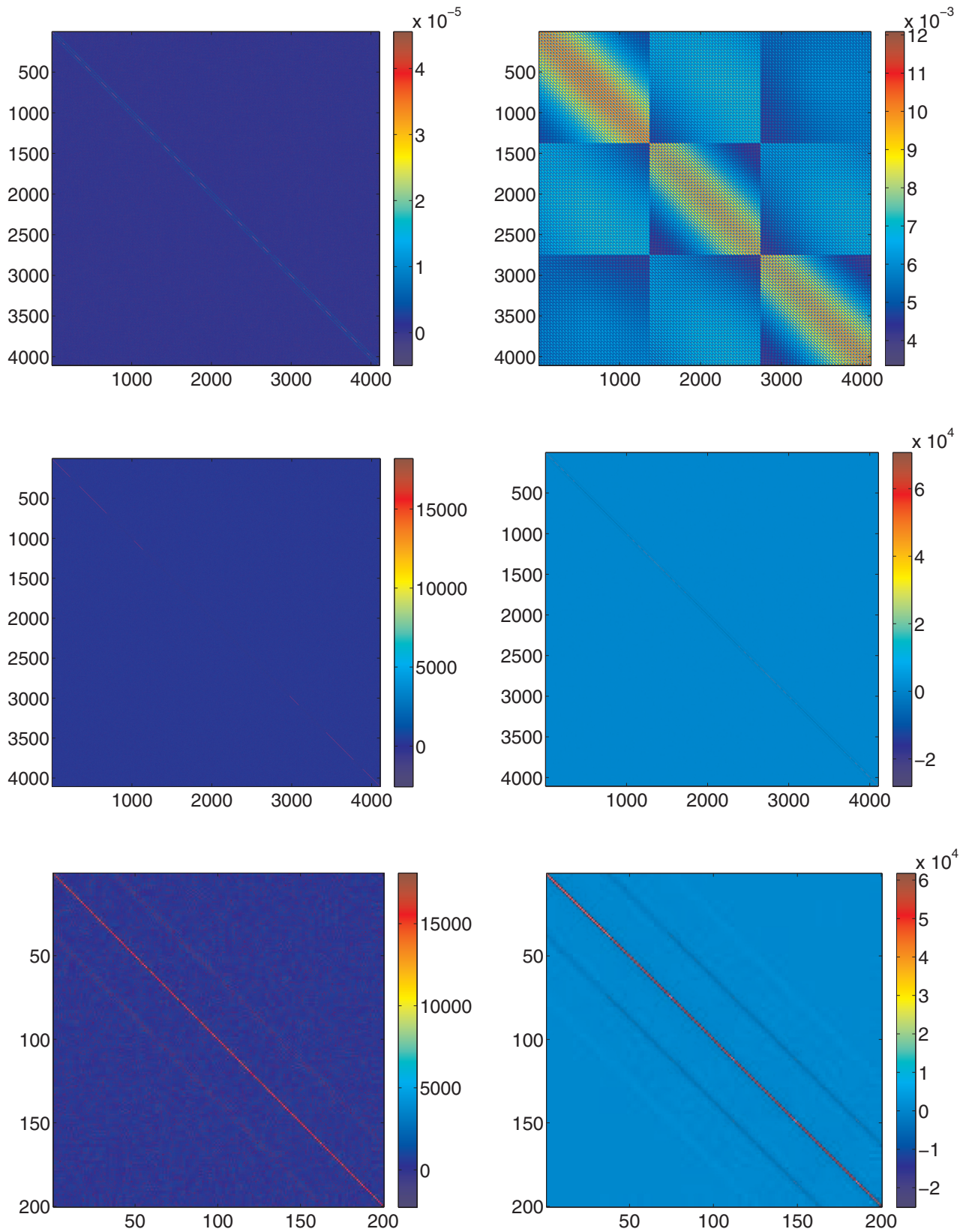
FIG. 9. Breast tomosynthesis: (top) 1369×1369 covariance estimates of projection images using the uniform (left) and variable (right) phantoms; (middle) inverse of the covariances of the uniform (left) and variable (right) phantom cases; (bottom) the first 296×296 zoomed-in version of the inverse covariances in order.

uniform mixture of the phantom materials, the intensity of the signal should have the following relationship:

$$a_s(\text{variable}) \approx \frac{\overline{g}_{\text{variable}}}{\overline{g}_{\text{uniform}}} a_s(\text{uniform}), \tag{3.6}$$

where $\overline{g}_{(\cdot)}$ and $a_s(\cdot)$ are, respectively, the mean gray level and signal intensity for the projection data from each phantom case. This consideration was incorporated in estimating the SNR ratios for comparing the uniform and variable phantom methods for assessing the mammography and DBT systems. That is, to calculate the ratio of the SNR estimates, denoted by γ, of the uniform phantom over the variable phantom, the adjusted SNR for the variable phantom case using the same signal intensity as for the uniform phantom, SNR*, was calculated as the initial SNR, SNR(variable), of the variable phantom times the ratio of the aforementioned two mean gray levels,

$$\text{SNR}^* = \frac{\overline{g}_{\text{variable}}}{\overline{g}_{\text{uniform}}} \text{SNR(variable)}, \tag{3.7}$$

$$\gamma = \frac{\text{SNR(uniform)}}{\text{SNR}^*}, \tag{3.8}$$

where SNR$(\cdot)$ is the SNR for each phantom case and SNR* denotes the adjusted SNR.

IV. RESULTS

IV.A. Comparison of uniform and variable phantom methods

As discussed in Sec. II C, observer-performance estimates vary depending on the sample size involved in the training and testing of the observer. To account for such uncertainty in our observer-performance measures, we calculated a single-observer but multiple-case variance for the AUC before transforming AUC estimates for a set of signal intensities and diameters to the SNR estimates. We observed that the AUC trends were statistically significant and hence so were the SNR$_{\text{AUC}}$ trends. We also observed that the performance trends using SNR$_t$ remained the same as the trends using SNR$_{\text{AUC}}$. In particular, for the uniform phantom case, two standard errors for the AUC were less than 0.05, while they were less than 0.035 for the variable phantom case. These upper bounds are for the AUC values close to .5, which is not surprising as the lower AUC means the task is more difficult, and hence larger variability. As the task incorporating background variability is a lot harder than one without it, the performances of both mammography and DBT systems using the uniform phantom method were much better than those using the variable phantom. For this reason, we faced the problem of AUC approaching or equaling 1, which yielded infinite SNR$_{\text{AUC}}$. Therefore, we chose to use SNR$_t$ instead of SNR$_{\text{AUC}}$ for all comparisons discussed in the following sections. This enabled us to set signal intensity the same for all different cases.

In Fig. 7, the top plot shows that the covariance of the uniform phantom data from the mammography system is not diagonal, which indicates that there still exits some spatial correlation between pixels even in the uniform projection data. These short-range correlations are the result of correlations in the detector output caused by correlated x-ray transport processes in the detector phosphor. The amount of correlation increases greatly when the projection data incorporate background variability, which is shown in the bottom plots in Fig. 7. Figure 8 shows the inverse of these covariances for both uniform and variable phantom cases. A similar trend is observed for the data covariances and their inverses of the DBT system, as shown in Fig. 9. Note that there is little correlation between the angular projections of the water phantom, whereas there is significant spatial correlation between the angular projections of the variable phantom even when the spatial correlation was partially incorporated with the independence assumption of ROIs in this work. We also note that the projections for the isodose per projection condition were used to produce the plots in Fig. 9. But, we observed that using the data covariance from the isodose per modality case yielded similar SNR and SNR ratio plots to those using the covariance in this figure, which will be further discussed in the following sections. With use of the aforementioned covariance matrices, the impact of background variability and spatial correlation information on observer performance is shown in the SNR plots of Figs. 10–12, and discussed in the following sections.

IV.A.1. Assessing mammography

Figure 10 shows the 3Dp Hotelling SNR maps for assessing the mammography system with the use of the uniform and variable phantoms. The top and middle plots in this figure show the 3Dp Hotelling observer's SNR maps, respectively, for the uniform and variable phantom cases for detecting the signal as a function of signal intensity (y-axis) and the diameter of the 3D signal in mm (x-axis). These plots indicate that the SNR for the uniform phantom case increases with signal intensity and signal diameter (as expected), but this is not true for the variable phantom case. With the variable phantom, the SNR of the mammography system does not increase with signal diameter while it does with signal intensity. In particular, the SNR for the variable phantom case fluctuates with a decreasing trend with signal diameter. We believe that this fluctuation is caused by the relationship between the projected signal parameters (shape and size) and the projection background statistics, which are affected by the spheres of different diameters used in our work. For instance, when the diameter of the projected signal is similar to those of the projections of the background spheres, it would be more difficult to detect the projected signal than when the diameter of the projected signal is largely different from those of the background spheres. In this work, the shape of the signal was the same as those of the background spheres, but differently sized spheres were used, resulting in different diameters of the projected background spheres. In addition, even the diameter of the projection of the same background sphere can vary depending on the sphere's location in the phantom, resulting in some of the projected spheres obscur-
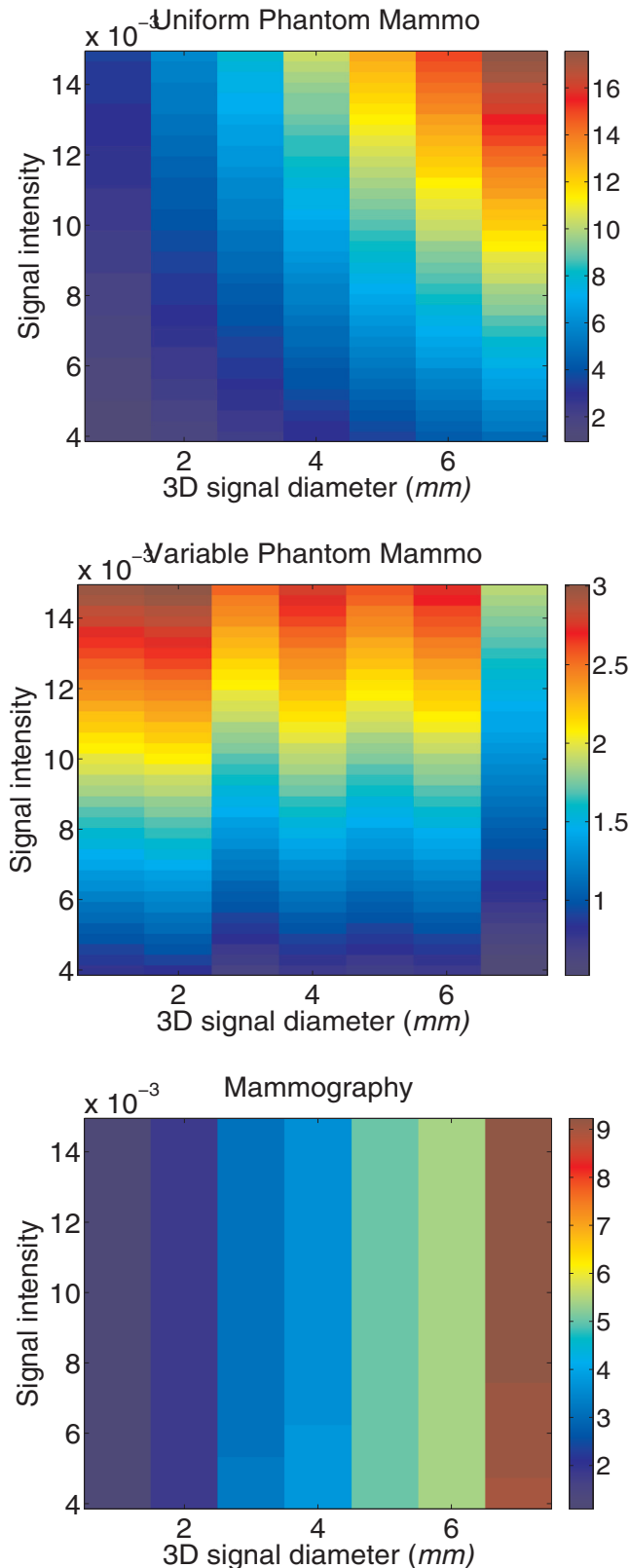
FIG. 10. Mammography: the first two plots show the Hotelling observer's SNR as a function of signal intensity and signal diameter (mm) for detecting a 2D signal in projection images of the uniform and variable phantoms, respectively. The bottom plot shows the SNR ratio of the uniform phantom over the variable phantom as a function of signal intensity and diameter (mm).

ing the signal more than others. However, with respect to signal intensity, the SNR trend is smooth because the SNR estimated using the Hotelling observer is linear with respect to signal intensity. Our SNR results are consistent with Burgess' finding:[47] in order to obtain the same level of detectability, signal intensity must increase with signal size for mammographic backgrounds, whereas for uniform backgrounds, the required signal intensity decreases with signal size for achieving the same level of detectability.

The bottom plot in Fig. 10 shows the resulting SNR ratio map as a function of signal intensity and diameter in mm. This plot indicates that using the uniform phantom method in evaluating mammography results in higher SNR than is obtained in the presence of a structured background. In addition, for a given signal diameter, the amount of the difference in SNRs with signal intensity remains about the same. However, with signal diameter, the difference in SNR levels between the uniform and variable phantoms appears to increase.

IV.A.2. Assessing breast tomosynthesis

Figures 11 and 12, respectively, show the 3Dp Hotelling SNR maps of the DBT systems using the isodose per modality and per projection conditions. The top and middle plots in each figure, respectively, present the 3Dp Hotelling SNRs for the uniform and variable phantom methods. The bottom plot in each figure shows the SNR ratio using the SNR estimates from the first two plots. In Figs. 11 and 12, while the actual values of the SNR are different for the two different dose conditions, the SNR trends remain similar in the sense that the SNR for the variable phantom case appears to decrease with signal diameter, whereas the SNR for the uniform phantom case tends to increase. For both the phantom cases, the SNR tends to increase with signal intensity, which was the expected outcome. However, with signal diameter, the SNR trend tends to increase (as expected) for the uniform phantom case but decrease for the variable phantom case. We note that the SNR trend with signal diameter for the variable phantom case is smoother with a decreasing trend than what is seen in Fig. 10 for the mammography system. This is because with some 3D information from the multiangle projections, the relationship between the diameters of the 3D signal and the 3D spheres influenced the detection task more than the relationship between the projected signal and background sphere diameters. As discussed in Sec. IV A 1, the latter case influenced the detection task for the mammography system. Therefore, given the diameters of the background spheres and 3D signal used in this work, the detection task for the DBT system became more difficult with signal diameter. The SNR ratio plots in the bottom of both Figs. 11 and 12 reveal higher estimates of the system's diagnostic performance when using the uniform phantom method that did not incorporate background variability. The trend in performance differences is similar to that of the mammography case discussed in Sec. IV A 1 in that with signal diam-
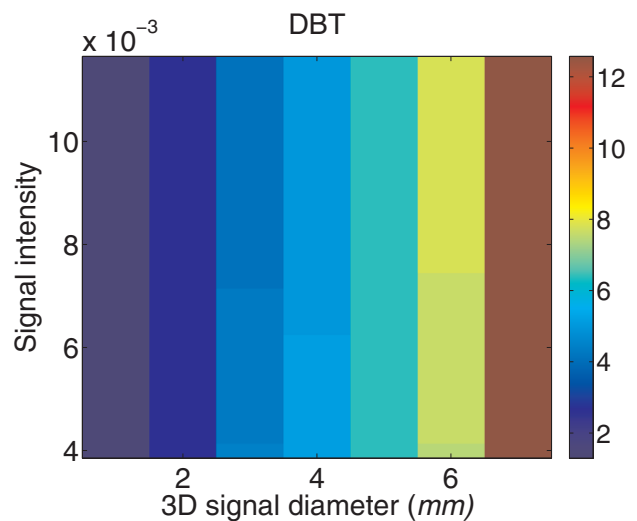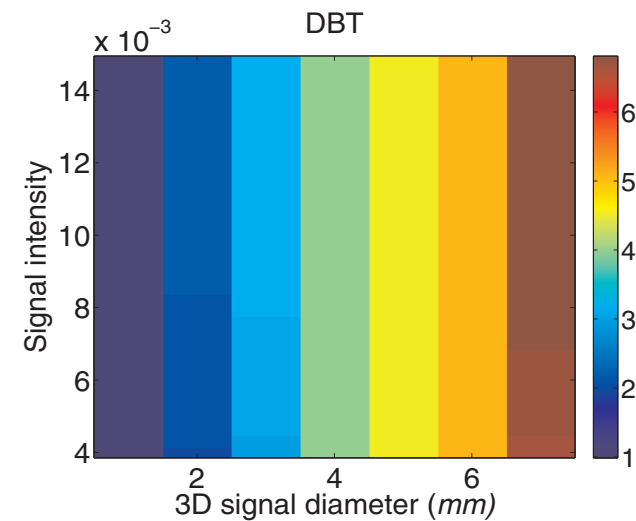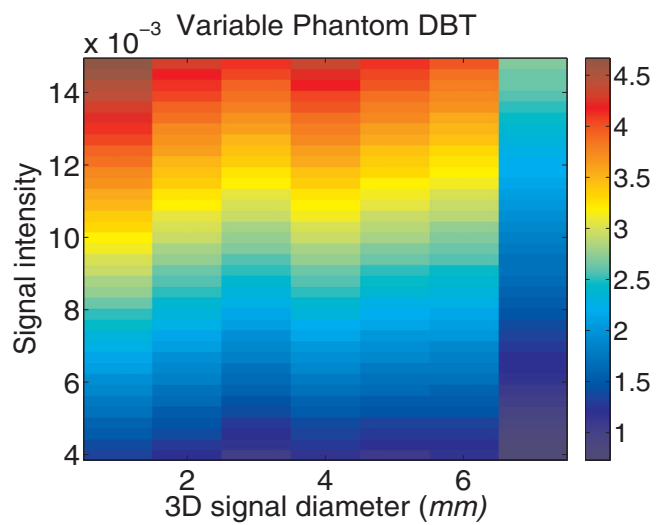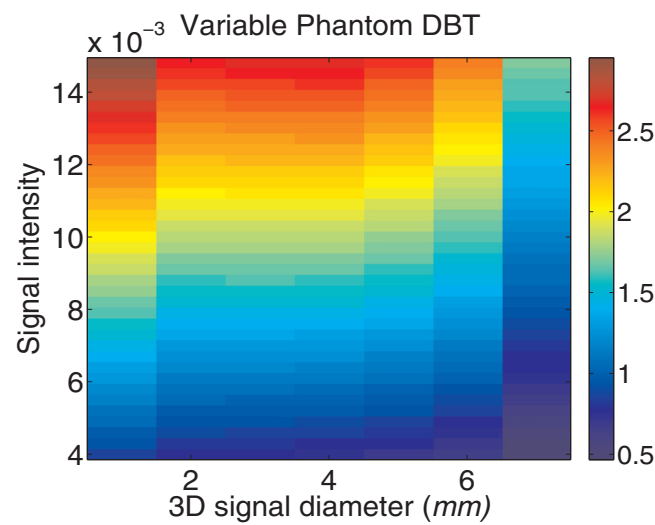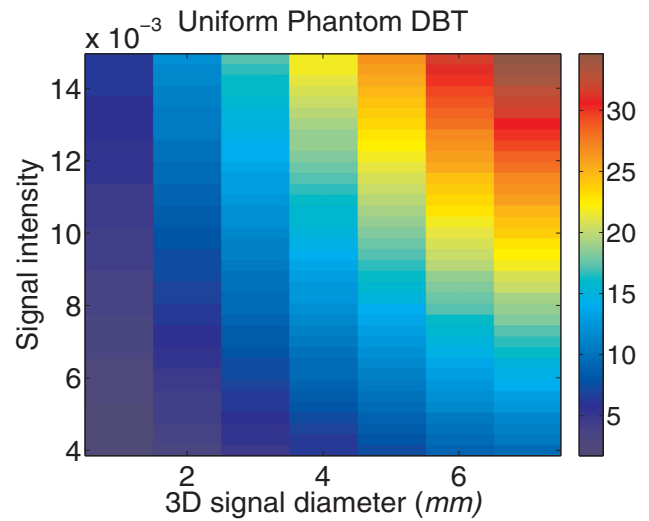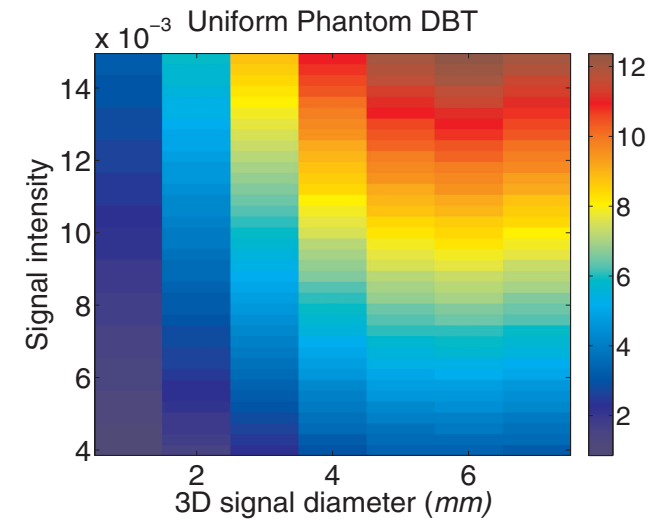
FIG. 11. Breast tomosynthesis with mammography dose: the first two plots show the Hotelling observer's SNR as a function of signal intensity and signal diameter (mm) for detecting a 2D signal in x-ray projection images of the uniform and variable phantoms, respectively. The bottom plot shows the SNR ratio of the uniform phantom over the variable phantom as a function of signal intensity and diameter (mm).

FIG. 12. Breast tomosynthesis with triple mammography dose: the first two plots show the Hotelling observer's SNR as a function of signal intensity and signal diameter (mm) for detecting a 2D signal in x-ray projection images of the uniform and variable phantoms, respectively. The bottom plot shows the SNR ratio of the uniform phantom over the variable phantom as a function of signal intensity and diameter (mm).
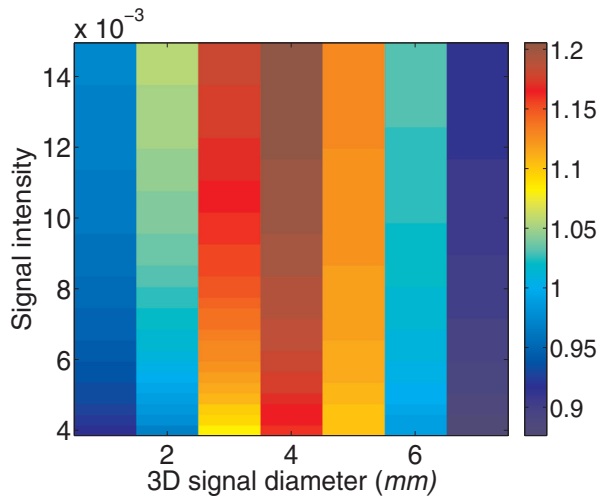
FIG. 13. Isodose per modality: plots show the SNR ratio of breast tomosynthesis over mammography with the uniform phantom (top) and variable phantom (bottom) methods.
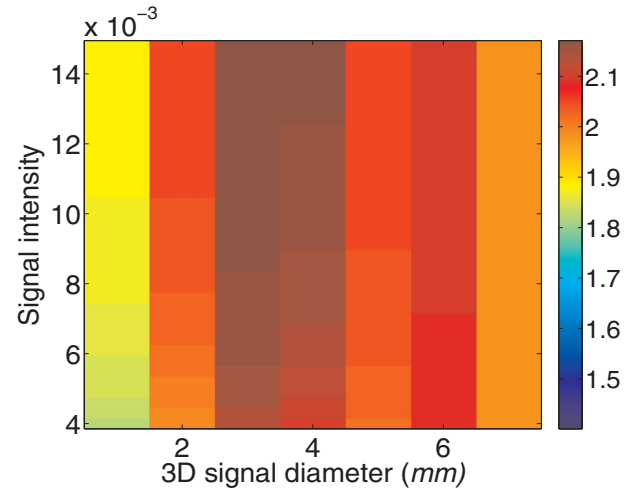

FIG. 14. Isodose per projection: plots show the SNR ratio of breast tomosynthesis over mammography with the uniform phantom (top) and variable phantom (bottom) methods.

eter, the difference in performance levels between the uniform and variable phantoms tends to increase, whereas it does not change much with signal intensity.

IV.B. Comparison of mammography and breast tomosynthesis

Figures 13 and 14 show the SNR ratios of the DBT system over the mammography system, respectively, for the isodose per modality and per projection conditions. In each figure, the left plot shows the SNR ratio using the uniform phantom, whereas the right plot shows the results using the variable phantom method. For the isodose per modality condition, the right plot in Fig. 13 shows that with the variable phantom, the three-angle DBT system does not give much advantage over mammography as the SNR ratio is less than or equal to 1 for most of the signal parameters. However, with the uniform phantom method as shown in the left plot of Fig. 13, the DBT system appears to be slightly better than the mammography system for certain parameters of the 3D lesion such as diameters of 3, 4, and 5 mm. For the isodose

per projection condition, as shown in Fig. 14, the left plot for the uniform method indicates that the DBT system is slightly better than the mammography system as the SNR ratio is above 2.0 for the majority of the signal parameters. However, we note that for the DBT system to achieve twice the SNR, three times the mammography dose was used. When background variability is incorporated in the SNR estimation, as shown in the right plot of Fig. 14, the advantage of the DBT system with triple the mammography dose is even further reduced because the SNR ratio is reduced to below around 1.7. These results indicate that when background variability was incorporated in the SNR estimation, the DBT system had less advantage over the mammography system, which was more pronounced in the isodose per projection case, i.e., more x-ray dose for the DBT system.

V. DISCUSSIONS

We employed the method of the Hotelling observer that uses the first and second orders of the data statistics and maximizes the task SNR for comparing the performance of

the mammography and DBT systems. To incorporate "3D" spatial correlation information between angular projections into our observer model (the 3Dp Hotelling observer), we made use of an ensemble of vectors, each of which concatenates the angular projections per phantom configuration and estimated the covariance for use in calculating the performance of the 3Dp Hotelling observer for the DBT systems. To have sufficient statistical power, i.e., sufficient numbers of ROIs for use in estimating the covariance for the 3Dp Hotelling observer as well as calculating the observer performance, we utilized all possible ROIs from each angular projection instead of limiting the ROI to the area where the 2D projected lesion is expected to appear. Due to this simplification, spatial correlation between pixels in different angular (background) projections may have been realized to be less than it actually is. In principle, using relevant local ROIs with respect to the signal location and size instead of using all ROIs throughout the FOV may lower or increase signal detectability, depending on the relationship between the local and overall background statistics. For instance, if the local background statistics have shorter-range correlations than the overall background statistics, then signal detectability estimated using the local ROIs can be improved in comparison to the SNR estimated using all ROIs. Full understanding of the interplay between the local and overall statistics and its impact on signal detectability is beyond the scope of this work. But, in our current work, we believe that 37×37 ROIs extracted from the center of the FOV have a similar distribution to that of 37×37 ROIs extracted from the whole FOV because of the way the phantom is designed. That is, when the contents of the phantom are stirred, different spheres can move around within the phantom, yielding similar local textures in the x-ray projections of one phantom configuration to another. The signal diameter in our work was chosen so that it takes some range of values within the given ROI used in the work. In addition, the ROI size was chosen so that a sufficient number of samples can be provided experimentally and the impact of a range of signal diameter values on system performance can be studied. We kept the largest signal diameter, which slightly goes over the boundaries of the square ROI, because using such a signal can show the impact of background variability on signal detectability better when the uniform and variable phantom methods are compared. More specifically, in a uniform background, having a larger signal simply means there is more signal energy and hence always higher signal detectability. But in a structured background, signal detectability can either improve or degrade because it is affected by the relationship between the background statistics and signal characteristics.

For the investigation of the aforementioned problem, one method is to try to avoid the sample size problem by constraining the Hotelling observer to a set of *efficient* channels that can extract salient statistical information from the data and to approximate the true observer performance by the *channelized* observer.[24,37,39,40,48] However, this alternative requires the knowledge of what kinds and how many channels are necessary for approximating the performance of the ideal observer, which depend on the specific background data statistics. One approach to implement this alternative is to apply two-dimensional (2D) channels that are efficient for each of the angular projections, yielding an N_c-dimensional channelized projection vector, and concatenating N_p channelized projection vectors into a vector of size $N_c \times N_p$, where N_c is the number of channels and N_p is the number of projections. This observer is called the *3Dp CHO* (Ref. [14] and [15]) as discussed in Sec. III D 1. However, if N_c and N_p are large enough, this approach can again have a problem of needing large numbers of images for training and testing.[14] Investigation of trade offs between the number of training images and the accuracy and stability of covariance estimation is one of our ongoing research activities. Another approach is to use the 3D CHO by applying 3D efficient channels, which can fully incorporate spatial correlation between angular projections into an N_c-dimensional channelized data vector and allow the 3D CHO to approximate the 3D Hotelling observer. In this case, a set of angular projections can be regarded as a 3D-like object in a multiangle projection space, and each of the N_c number of 3D efficient channels can be applied to the whole set of the angular projections via the dot product, yielding the N_c-dimensional channelized data vector. This way, the dimension of the data vector for use in observer performance and variance analyses is only N_c, which is much smaller than M, the dimension of the original data set, or $N_c \times N_p$ of the 3Dp CHO approach. This aspect is currently under investigation, for example, to extend the method of choosing efficient channels using the partial least squares algorithm, as investigated by Witten *et al.*,[39] to the 3D case.

In this work, we considered only one breast tissue composition. We are interested in extending our method by using different mixtures of spheres to represent different classes of breast tissue compositions in order to investigate the impact of breast density on observer-performance trends. We are also interested in investigating more realistic tissue texture models with the use of objects of different shapes. To incorporate the impact of breast density on the intensity of the 2D signal projection and hence detection performance, we plan to expand our work to more realistically model the lesion projections incorporating the effects of the 3D lesions embedded in the variable phantom instead of adding a computer simulated signal to the phantom projection data with the assumption that the signal projection is independent of the background projection. For a more accurate comparison of the uniform and variable-background phantoms, instead of using the uniform phantom filled with water, we are interested in employing Jennings' approach[45] to create a phantom consisting of a couple of different fluids, which would give closer x-ray attenuation properties to the variable phantom. Lastly, we are currently investigating important features and locations of features in order to simplify the variable phantom to a fixed phantom of salient features while capturing important information for use in the evaluation and calibration of the 3D systems.

As discussed above, there are still a number of improvements to be made for our evaluation method, and many questions to be addressed and many different parameters to be investigated for a complete evaluation and optimization of

3D breast imaging systems. This may be infeasible to do so through the use of experimental data from the laboratory setting due to the large space of parameters to be investigated. Therefore, while it is useful to have the physical phantom proposed in this work, it is also essential to make use of sophisticated simulation and computation tools in these efforts toward developing a statistical, task-based evaluation method and evaluating the 3D system accordingly,[15,30,35,49] and gaining the knowledge necessary to improve the current status of not only system optimization methods but also QA/QC phantoms and tests. When such further investigations are performed via simulation and hence a range of optimal system parameters for the 3D imaging is found, the method and the phantom presented in this work can be useful for validating the optimal parameters found via simulation studies as well as improving the current status of the QA/QC phantoms and tests.

VI. CONCLUSION

We presented a statistical, task-based assessment method for evaluating planar and 3D x-ray breast imaging systems and demonstrated our evaluation method for comparing the laboratory mammography and three-angle DBT systems in signal detection tasks. With use of the variable and uniform phantoms, we investigated the impact of background variability on this system performance comparison. For the phantoms tested in our study, when background variability and correlation between the multiangle projections were incorporated into our observer performance calculations, the SNR trends were found to be different from those without background variability. This outcome implies that it is important to take into account background variability and spatial correlation information between any two pixels in angular and single projection images in estimating the system's diagnostic performance for accurate evaluation of 3D breast imaging systems.

ACKNOWLEDGMENTS

a)Electronic mail: subok.park@fda.hhs.gov.

[1] X. Gong, S. J. Glick, B. Liu, A. A. Vedula, and S. Thacker, "A computer simulation study comparing lesion detection accuracy with digital mammography, breast tomosynthesis, and cone-beam CT breast imaging," Med. Phys. **33**, 1041–1052 (2006).

[2] S. J. Glick, S. Thacker, X. Gong, and B. Liu, "Evaluating the impact of x-ray spectral shape on image quality in flat-panel CT breast imaging," Med. Phys. **34**, 5–23 (2007).

[3] A. C. Kwan, J. M. Boone, and N. Shah, "Evaluation of x-ray scatter properties in a dedicated cone-beam breast CT scanner," Med. Phys. **32**,

2967–2975 (2005).

[4] A. C. Kwan, J. M. Boone, K. Yang, and S. Huang, "Evaluation of the spatial resolution characteristics of a cone-beam breast CT scanner," Med. Phys. **34**, 275–281 (2007).

[5] M. Das, H. C. Gifford, J. M. O'Connor, and S. J. Glick, "Evaluation of a variable dose acquisition technique for microcalcification and mass detection in digital breast tomosynthesis," Med. Phys. **36**, 1976–1984 (2009).

[6] J. Zhou, B. Zhao, and W. Zhao, "A computer simulation platform for the optimization of a breast tomosynthesis system," Med. Phys. **34**, 1098–1109 (2007).

[7] B. Zhao and W. Zhao, "Three-dimensional linear system analysis for breast tomosynthesis," Med. Phys. **35**, 5219–5232 (2008).

[8] B. Zhao, Y. Hu, T. Mertelmeier, J. Ludwig, and W. Zhao, "Experimental validation of a three-dimensional linear system model for breast tomosynthesis," Med. Phys. **36**, 240–251 (2009).

[9] I. Sechopoulos, S. Suryanarayanan, S. Vedantham, C. J. D'Orsi, and A. Karel-las, "Scatter radiation in digital tomosynthesis of the breast," Med. Phys. **34**, 564–576 (2007).

[10] I. Sechopoulos, S. Suryanarayanan, S. Vedantham, C. J. D'Orsi, and A. Karel-las, "Computation of the glandular radiation dose in digital tomosnthesis of the breast," Med. Phys. **34**, 221–232 (2007).

[11] I. Sechopoulos and C. Chetti, "Optimization of the acquisition geometry in digital tomosynthesis of the breast," Med. Phys. **36**, 1199–1207 (2009).

[12] A. S. Chawla, E. Samei, R. S. Saunders, J. Y. Lo, and J. A. Baker, "A mathematical model platform for optimizing a multiprojection breast imaging system," Med. Phys. **35**, 1337–1345 (2008).

[13] R. S. Saunders, E. Samei, J. Y. Lo, and J. A. Baker, "Can compression be reduced for breast tomosynthesis? Monte Carlo study on mass and microcalcification conspicuity in tomosynthesis," Radiology **251**, 673–682 (2009).

[14] S. Young, S. Park, S. K. Anderson, K. J. Myers, A. Badano, and P. Bakic, "Estimating breast tomosynthesis performance in detection tasks with variable-background phantoms," Proc. SPIE **7258**, 72580O1–72580O9 (2009).

[15] S. Young, P. Bakic, K. Myers, and S. Park, "Performance tradeoffs in a model breast tomosynthesis system," in Digital Image Processing and Analysis, OSA Technical Digest (CD) (Optical Society of America, 2010), paper DTuA3.

[16] S. Park, H. Liu, R. Jennings, R. Leimbach, I. Kyprianou, A. Badano, and K. J. Myers, "A task-based evaluation method for x-ray breast imaging systems using variable-background phantoms," Proc. SPIE **7258**, 72581L1–72581L8 (2009).

[17] S. Richard and E. Samei, "Quantitative imaging in breast tomosynthesis and CT: Comparison of detection and estimation task performance," Med. Phys. **37**, 2627–2637 (2010).

[18] G. J. Gang, D. J. Tward, J. Lee, and J. H. Siewerdsen, "Anatomical background and generalized detectability in tomosynthesis and cone-beam CT," Med. Phys. **37**, 1948–1965 (2010).

[19] I. Reiser and R. M. Nishikawa, "Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise," Med. Phys. **37**, 1591–1600 (2010).

[20] J. G. Mainprize, A. K. Bloomquist, M. P. Kempston, and M. J. Yaffe, "Resolution at oblique incidence angles of a flat panel imager for breast tomosynthesis," Med. Phys. **33**, 3159–3164 (2006).

[21] A. Badano, I. S. Kyprianou, R. J. Jennings, and J. Sempau, "Anisotropic imaging performance in breast tomosynthesis," Med. Phys. **34**, 4076–4091 (2007).

[22] A. Badano, I. S. Kyprianou, M. Freed, R. J. Jennings, and J. Sempau, "Effect of oblique x-ray incidence in flat-panel computed tomography of the breast," IEEE Trans. Med. Imaging **28**, 696–702 (2009).

[23] A. E. Burgess, F. L. Jacobson, and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise," Med. Phys. **28**, 419–437 (2001).

[24] H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley, New York, 2004).

[25] American Association of Physics in Medicine Focused Research Meeting on Model Observers for Tomosynthesis and CT of the Breast: Theoretical and Practical Considerations, Chicago, IL, March 2009.

[26] C. M. Li, W. P. Segars, G. D. Tourassi, J. M. Boone, and J. T. Dobbins, "Methodology for generating a 3D computerized breast phantom from empirical data," Med. Phys. **36**, 3122–3131 (2009).

[27] M. J. O'Connor, M. Das, C. Didier, M. Mah'D, and S. J. Glick, "Using mastectomy specimens to develop breast models for breast tomosynthesis

and CT breast imaging," Proc. SPIE **6913**, 6913151–6913156 (2008).

[28] A. Badal and A. Badano, "Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel Graphics Processing Unit," Med. Phys. **36**, 4878–4880 (2009).

[29] A. Badal, I. S. Kyprianou, D. Banh, A. Badano, and J. Sempau, "penMesh–Monte Carlo radiation transport simulation in a triangle mesh geometry," IEEE Trans. Med. Imaging **28**, 1894–1901 (2009).

[30] M. Freed, S. Park, and A. Badano, "A fast, angle-dependent, analytical model of CsI detector response for optimization of 3D x-ray breast imaging systems," Med. Phys. **37**, 2593–2605 (2010).

[31] B. D. Gallas, J. S. Boswell, A. Badano, R. M. Gagne, and K. J. Myers, "An energy- and depth-dependent model for x-ray imaging," Med. Phys. **31**, 3132–3149 (2004).

[32] G. Hadjok and I. A. Cunningham, "Penalty on the detective quantum efficiency from off-axis incident x rays," Proc. SPIE **5368**, 109–118 (2004).

[33] K. Bliznakova, Z. Bliznakova, V. Bravou, Z. Kolitsi, and N. Pallikarakis, "A three-dimensional breast software phantom for mammography simulation," Phys. Med. Biol. **48**, 3699–3719 (2003).

[34] P. R. Bakic, M. Albert, D. Brzakovic, and A. D. Maidment, "Mammogram synthesis using a 3D simulation. I. breast tissue model and image acquisition simulation," Med. Phys. **29**, 2131–2139 (2002).

[35] C. Zhang, P. R. Bakic, and A. D. A. Maidment, "Development of an anthropomorphic breast software phantom based on region growing algorithm," Proc. SPIE **6918**, 69180V (2008).

[36] H. L. Van Trees, *Detection, Estimation, and Modulation Theory (Part I)* (Academic, New York, 1968).

[37] S. Park, J. M. Witten, and K. J. Myers, "Singular vectors of a linear imaging system as efficient channels for the Bayesian ideal observer," IEEE Trans. Med. Imaging **28**, 657–668 (2009).

[38] S. Park and E. Clarkson, "Efficient estimation of ideal-observer performance in classification tasks involving high-dimensional complex backgrounds," J. Opt. Soc. Am. A **26**, B59–B71 (2009).

[39] J. Witten, S. Park, and K. J. Myers, "Partial least squares: A method to compute efficient channels for the Bayesian ideal observers," IEEE Trans. Med. Imaging **29**, 1050–1058 (2010).

[40] B. D. Gallas and H. H. Barrett, "Validating the use of channels to estimate the ideal linear observer," J. Opt. Soc. Am. A **20**, 1725–1738 (2003).

[41] J. Y. Hesterman, M. A. Kupinski, E. Clarkson, and H. H. Barrett, "Hardware assessment using the multi-module, multi-resolution system (M^3R)—A signal-detection study," Med. Phys. **34**, 3034–3044 (2007).

[42] X-ray attenuation tables found at the National Institute of Standards and Technology website (http://physics.nist.gov/PhysRefData/FFast/html/form. html).

[43] R. J. Jennings, "A method for comparing beam-hardening materials for diagnostic radiology," Med. Phys. **15**, 588–599 (1988).

[44] R. J. Jennings, "Spectrally precise phantoms for quality assurance in diagnostic radiology," Br. J. Radiol., Suppl. **18**, 90–93 (1984).

[45] R. J. Jennings, "Computational methods for the design of test objects and tissue substitutes for radiologic applications," Radiat. Prot. Dosim. **49**, 327–332 (1993).

[46] G. R. Hammerstein, D. W. Miller, D. R. White, M. E. Masterson, H. Q. Woodard, and J. S. Laughlin, "Absorbed radiation dose in mammography," Radiology **130**, 485–491 (1979).

[47] A. E. Burgess and P. F. Judy, "Signal detection in power-law noise: Effect of spectrum exponents," J. Opt. Soc. Am. A **24**, B52–B60 (2007).

[48] K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal observer model," J. Opt. Soc. Am. A **4**, 2447–2457 (1987).

[49] K. S. Anderson, S. Park, A. Badal, I. Kyprianou, and A. Badano, *Accurate Simulation of Breast Tomosynthesis Projections Through Complex Phantoms Using Monte Carlo* (RSNA, Chicago, IL, 2008).