A virtual trial framework for quantifying the detectability of masses in breast tomosynthesis projection data

Stefano Young

College of Optical Sciences, University of Arizona, Tucson, Arizona 85721

Predrag R. Bakic

Radiology Department, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Kyle J. Myers, Robert J. Jennings, and Subok Park^{a)}

Division of Imaging and Applied Mathematics, Center for Devices and Radiological Health, FDA, Silver Spring, Maryland 20993

(Received 12 September 2012; revised 26 February 2013; accepted for publication 22 March 2013; published 25 April 2013)

Purpose: Digital breast tomosynthesis (DBT) is a promising breast cancer screening tool that has already begun making inroads into clinical practice. However, there is ongoing debate over how to quantitatively evaluate and optimize these systems, because different definitions of image quality can lead to different optimal design strategies. Powerful and accurate tools are desired to extend our understanding of DBT system optimization and validate published design principles.

Methods: The authors developed a virtual trial framework for task-specific DBT assessment that uses digital phantoms, open-source x-ray transport codes, and a projection-space, spatial-domain observer model for quantitative system evaluation. The authors considered evaluation of reconstruction algorithms as a separate problem and focused on the information content in the raw, unfiltered projection images. Specifically, the authors investigated the effects of scan angle and number of angular projections on detectability of a small (3 mm diameter) signal embedded in randomly-varying anatomical backgrounds. Detectability was measured by the area under the receiver-operating characteristic curve (AUC). Experiments were repeated for three test cases where the detectability-limiting factor was anatomical variability, quantum noise, or electronic noise. The authors also juxtaposed the virtual trial framework with other published studies to illustrate its advantages and disadvantages.

Results: The large number of variables in a virtual DBT study make it difficult to directly compare different authors' results, so each result must be interpreted within the context of the specific virtual trial framework. The following results apply to 25% density phantoms with 5.15 cm compressed thickness and 500 μm^3 voxels (larger 500 μm^2 detector pixels were used to avoid voxel-edge artifacts): 1. For raw, unfiltered projection images in the anatomical-variability-limited regime, AUC appeared to remain constant or increase slightly with scan angle. 2. In the same regime, when the authors fixed the scan angle, AUC increased asymptotically with the number of projections. The threshold number of projections for asymptotic AUC performance depended on the scan angle. In the quantum- and electronic-noise dominant regimes, AUC behaviors as a function of scan angle and number of projections sometimes differed from the anatomy-limited regime. For example, with a fixed scan angle, AUC generally decreased with the number of projections in the electronic-noise dominant regime. These results are intended to demonstrate the capabilities of the virtual trial framework, not to be used as optimization rules for DBT.

Conclusions: The authors have demonstrated a novel simulation framework and tools for evaluating DBT systems in an objective, task-specific manner. This framework facilitates further investigation of image quality tradeoffs in DBT. © 2013 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4800501>]

Key words: virtual trials, breast tomosynthesis, task-specific, image quality, projections, model observers

I. INTRODUCTION

Digital breast tomosynthesis (DBT) is an attractive limited-angle tomographic alternative to conventional 2D screening mammography because it has potential to improve both sensitivity and specificity (call-back rates).^{1,2} However, this potential improvement is difficult to quantify because it depends not only on the imaging system, but also on the specific patient population and clinical task of interest. Even if we as-

sume that DBT improves screening, the challenge remains to quantitatively assess the merits of different DBT imaging systems.

Studies have shown that the optimal x-ray spectrum for full-field digital mammography (FFDM) imaging depends on patient characteristics.^{3,4} Some commercial FFDM systems are designed with selectable parameters in order to optimize the x-ray spectrum for a given patient at the time of the examination. The automatic-exposure-control subsystems of these

TABLE I. Key image quality factors in mammography and DBT.

Mammography	Digital breast tomosynthesis
Target/filter materials	Target/filter materials
Tube voltage	Tube voltage
Tube current-exposure product	Tube current-exposure product
Breast thickness	Breast thickness
Breast density	Breast density
Detector properties	Detector properties
Display device	Display device
	X-ray source scan angle
	X-ray source angular sampling
	Dose delivery scheme
	Reconstruction algorithm

devices automatically optimize over the space of mammography system parameters (e.g., tube voltage, tube current-exposure time product) based on pre-exposure measurements of patient characteristics (compressed breast thickness, breast attenuation measured by a low-level test exposure).⁵ However, these FFDM optimization parameters are only a subset of the factors affecting image quality in DBT (Table I). We expect that ongoing investigations into DBT will reveal improved algorithms for optimizing image quality on a patient-specific basis. Thus, DBT researchers continue to develop prototypes and explore the parameter space of this modality to find optimal imaging conditions.

Because physical prototype development is expensive and time-consuming, some groups are researching strategies for DBT system optimization based on simulation studies. Specifically, optimization of the x-ray source scan angle and angular sampling scheme is an area of interest in recent literature.^{6–17} These studies could reduce the number of physical prototype iterations and eventually lead to task- and patient-specific optimization of tomosynthesis. But to achieve these goals, we first need accurate, task-based strategies for evaluating key DBT system parameters in simulation. In practice, various assumptions are necessary to reduce the optimization space and design tractable simulations; this work is no exception. We sacrifice some realism in order to introduce a virtual framework for quantifying the tradeoffs between factors that influence DBT image quality. In this paper, we begin by developing a model for generating tomosynthesis data (Sec. II). Then, we apply a flexible, task-based assessment strategy using model observers (Sec. III). Finally, we show specific examples of how our approach can be used to assess image quality tradeoffs for different scan and sampling angle combinations in DBT (Sec. IV). Section V provides a direct comparison of our framework to other relevant publications, and details on the limitations of our framework are found in Sec. VI.

II. DBT SYSTEM MODEL

For DBT system evaluation, we need a computationally tractable x-ray imaging model that supports anthropomorphic digital breast phantoms, accurate radiation transport in multi-

projection geometries, a detector response/noise model, and a relationship to mammographic x-ray dose quantities. To meet these model requirements, we first generated a large ensemble of 3D voxelized breast phantoms (Sec. II.A). Then, we used open-source codes based on the PENELOPE Monte Carlo framework¹⁸ to simulate multiprojection x-ray scans (Sec. II.B). Next, we modeled the statistics of an idealized indirect-type x-ray detector (Sec. II.C). Finally, we related the simulated x-ray projection data to standard measures of dose in mammography (Sec. II.D).

II.A. The digital phantom ensemble

There are a number of software phantoms in the literature which seek to approximate the structural nature of breast tissue in digital or analytical form.^{19–25} We needed three specific features in a software phantom: (1) the ability to generate random breast parenchymal patterns, (2) projection image statistics that match mammogram statistics, and (3) the ability to generate a large batch of phantoms relatively quickly. Bakic’s voxelized breast phantom model was chosen to meet these requirements.^{19,26,27} Various studies have demonstrated that Bakic’s phantom can be used to simulate realistic breast parenchymal patterns.^{28–30} We created an ensemble of 1000 compressed breast phantoms having the same volume (450 mL) and approximately the same percent density (25% fibroglandular, 75% adipose). Based on recent results by Yaffe *et al.*,³¹ we chose a 25% density model as a more appropriate representation of the “standard” clinical breast than a 50–50 model.

Each phantom contained random numbers of fibroglandular and adipose compartments (drawn from Poisson distributions) randomly localized within the phantom volume. Figure 1 shows example sagittal cross-sections from three uncompressed breast phantoms with randomly-generated compartments, illustrating variations in the adipose (dark gray) and fibroglandular breast structures (white). Before simulating the forward problem, the phantoms were compressed via finite element methods.^{32,33} Material types were mapped to tissue density coefficients based on previously reported values,³⁴ but to simulate a more challenging detection task, we set the signal density to 1.0 g/cm³ instead of using the reported value for carcinoma density (1.044 g/cm³). Table II lists shared parameters of the breast phantom ensemble.

II.B. Multiprojection x-ray imaging

Full 3D x-ray simulation with scatter is challenging due to heavy computational requirements. To assess image quality, we have the additional need to produce hundreds or

TABLE II. Shared parameters of the virtual breast ensemble.

Sagittal dimensions	20.45 × 6.5 cm
Breast thickness (compressed)	5.15 cm
Voxel size	500 μm
Adipose tissue density	0.928 g/cm ³
Fibroglandular tissue density	1.035 g/cm ³
Signal mass density	1.0 g/cm ³

FIG. 1. Breast phantoms (before compression) in sagittal cross-section with randomly-generated compartments of adipose (gray) and fibroglandular tissues (white), showing example tissue structures for the simulated class of patients.

thousands of multiprojection scans in a reasonable amount of time. To simulate multiprojection scans, we began by generating scatter-free images using a voxelized version of the PENELOPE photon-electron transport code called penEasy.^{35,36} We used the penEasy package (v. 2010-09-02) in ray-tracing mode and ran simulations in parallel on a 190-node cluster of AMD Opteron and Intel Xeon servers in order to generate large image sets quickly. In penEasy ray tracing, a noise-free mean image is generated by casting a small number of deterministic rays from the focal spot to different positions inside each pixel to estimate the probability of x rays arriving at the pixel without interacting along the track. The ray-tracing image is scaled by the probability of emitting the x rays within the solid angle covering each pixel, assuming that the source is collimated to exactly cover the detector surface. The output is a normalized image with units eV/cm² per source photon.

We modeled an x-ray point source which radiates monoenergetic photons (20 keV), traces an arc of radius $R = 59$ cm through a maximum scan angle (θ_{scan}) of 96° at variable angular sampling intervals ($\Delta\theta$), and projects onto a 26×7 cm² detector region-of-interest (ROI) with $500 \mu\text{m}$ pixels (Fig. 2). $\theta_{\text{scan}} = 96^\circ$ corresponds to a wide-angle DBT system with acquisition parameters outside the range of published values for scanning-source DBT prototype systems (see Table II of Reiser and Nishikawa¹⁵). Wider angles were considered to allow for static-source systems where gantry speed is not an issue. We inserted a 1.5 cm air gap between the phantom and the detector, introducing a slight magnification. By translating the detector ROI with x-ray source angle, we modeled

FIG. 2. We modeled a partially-isocentric breast tomosynthesis acquisition geometry (schematic not to scale).

partially-isocentric motion with the center-of-rotation inside the breast phantom volume.

We generated a total of 97 000 projection images in penEasy corresponding to 1000 different phantom realizations with 97 projections per scan ($\Delta\theta = 1^\circ$). For the observer studies in Sec. III, we created a signal-known-exactly (SKE) task by inserting a small spherical signal mass (3 mm diameter) in half of the breast phantoms at the center-of-rotation of the x-ray source arc. The mass was inserted by changing the density coefficients of the voxels at the signal location to the signal mass density given in Table II. A 3 mm mass corresponds to an early-stage cancer with a favorable five-year prognosis,³⁷ and its small size makes it susceptible to anatomical masking—a good scenario for testing the benefits of DBT. Figures 2 and 3 together illustrate the forward problem for one of the breast phantoms. The projected signal was invisible to the human eye in a noise-free projection at $\theta = 0^\circ$ [Fig. 3 (center)]. For illustrative purposes only, we show that the signal is visible in a simple back projection reconstruction with $\theta_{\text{scan}} = 96^\circ$ and 25 projection angles [Fig. 3 (right)]. This corresponds to the kind of detection task where DBT may provide an advantage over conventional mammography. In Sec. IV, however, all results are for raw images without reconstruction.

For modeling scatter, we generated a uniform phantom with the same shape as the structured phantoms and a nominal density 0.955 g/cm^3 (for 75%–25% adipose-fibroglandular) based on results of Hammerstein *et al.*³⁸ We used the open-source MC-GPU code³⁹ with slight modifications to support partially-isocentric DBT geometries, to simulate one high-exposure (10^{12} source photons) acquisition at each projection angle. In MC-GPU, the source was defined as a fan beam with fixed polar (24°) and azimuthal (7°) aperture angles, producing a $27.6 \times 8.0 \text{ cm}^2$ rectangular field (at $\theta = 0^\circ$). Each Monte Carlo run with 10^{12} source photons took about 20 min to simulate using 8 NVIDIA GTX 580 GPUs in parallel.

MC-GPU stores separate images for the primary and scattered photons. The scatter-only image is normalized with

FIG. 3. From left-to-right: Sagittal slice through a compressed phantom ($z = 6$ cm) with 3 mm diameter signal (light gray), ray-tracing projection at $\theta = 0^\circ$, and simple back projection reconstruction from 25 equally-spaced ray-tracing projections with $\theta_{\text{scan}} = 96^\circ$ (refer to Fig. 2 for geometry).

units eV/cm^2 per source photon, which matches the output from penEasy ray tracing. At a given angle, we formed “noise-free” hybrid images by adding the high-exposure, scatter-only image of the uniform phantom to the penEasy ray-tracing images for each structured phantom. Thus, our scatter model depended on the projection angle, but not on the distribution of tissues in a given structured phantom. We used this “noise-free” hybrid model, instead of the total MC images, to allow for analyzing various noise levels in postprocessing.

II.C. Detector model

After storing the “noise-free” hybrid images from penEasy/MC-GPU in units of eV/cm^2 per source photon, we needed a practical detector noise model. We modeled an indirect-type detector with a CsI:Tl phosphor layer deposited onto a photodiode array. Relevant detector parameters are shown in Table III. There were three primary assumptions in the model: (1) noise-free gain in the phosphor material, (2) no depth-of-interaction effects, and (3) no energy-dependent effects. Also, the quantum efficiency of x-ray absorption was assumed to be independent of incident angle. In this idealized model, we can write the mean output voltage in pixel m as

$$\bar{g}_m = \frac{\eta A N_0 \bar{k} \Gamma}{P \bar{\mathcal{E}}} p_m, \quad (1)$$

where η is the quantum efficiency of the phosphor, A is the pixel area, N_0 is the total number of source photons summed over all projections, $\bar{k}$ is the mean optical gain (number of optical secondaries per x-ray primary), Γ is the electronic gain (photodiode output voltage per optical secondary), P is the number of projections in the DBT scan, $\bar{\mathcal{E}}$ is the mean x-ray energy, and p_m is the noise-free penEasy/MC-GPU pixel value for pixel m . Note that we assumed Γ is independent of

TABLE III. DBT model source, detector, and dose parameters.

Source Parameters:	
Monoenergetic source ($\bar{\mathcal{E}}$)	20 keV
Maximum scan angle (θ_{scan})	96°
Source-image distance (at $\theta = 0^\circ$)	65 cm
Detector Parameters:	
Detector type	Indirect (CsI:Tl)
Pixels per projection (M)	520×140
Pixel Area (A)	0.25 mm^2
Phosphor efficiency (η)	0.7
Optical gain ($\bar{k}$)	1050
Electronic gain (Γ)	5.13×10^{-9}
Dose Parameters:	
mGy per source photon (at 0°)	4.2×10^{-12}

m , which may not be true for real photodiode arrays. Also note that this model assumes the source photons are divided equally among the P projections. The variable g_m is used here for consistency with the $M \times 1$ image vector $\mathbf{g}$ described later in Sec. III.A. CsI:Tl has a gain of 55 optical photons per keV of incident x-ray energy.⁴⁰ For 20 keV photons, we assumed $\eta(20 \text{ keV}) \approx 70\%$ so $\eta \bar{k} \approx 800$ optical secondaries per x-ray primary. The electronic gain parameter Γ was calculated from Eq. (12.301) in Ref. 41, reproduced here:

$$\Gamma = -\frac{e \eta_{\text{pd}}}{C_{\text{pd}}}, \quad (2)$$

where e is the electron charge, η_{pd} is the quantum efficiency of the photodiode, and C_{pd} is the photodiode’s capacitance. For a $500 \times 500 \mu\text{m}$, $1\text{-}\mu\text{m}$ -thick a-Si:H photodiode, we used $\eta_{\text{pd}} = 0.8$ and $C_{\text{pd}} = 25 \text{ pF}$ (pp. 260–261 of Ref. 40) and found an effective gain $\Gamma = 5.13 \times 10^{-9}$ volts per optical secondary.

To model the quantum noise in a pixel, we invoked the noise-free gain assumption so that the quantum noise variance of g_m may be written as the mean number of absorbed x rays times the total gain squared (pp. 1110–1111 of Ref. 41). This model can be considered as a “special case” of the Burgess variance theorem.⁴¹ In the notation of Eq. (1), we have:

$$\text{Var}(g_m) = \frac{\eta A N_0}{P \bar{\mathcal{E}}} (\bar{k} \Gamma)^2 p_m, \quad (3)$$

where again we emphasize the assumption that no additional variance is introduced during the x-ray-to-voltage conversion process in the detector. For a CsI:Tl detector, the Swank factor is close to unity (p. 250 of Ref. 40), so we could reasonably assume a noise-free gain mechanism in the phosphor. This model makes the noise more computationally tractable, because we can model the quantum noise in a pixel by sampling from a Gaussian distribution with the appropriate variance. We modeled multiprojection scans by dividing the total number of photons N_0 equally among the projections, so scans with more projections suffered from more quantum noise (Fig. 4).

In addition to modeling quantum noise, we added Gaussian noise with variance σ_{elec}^2 to approximate the effects of electronic noise in the detector. Reported empirical values for electronic noise in indirect x-ray detectors are variable

FIG. 4. From left-to-right: Zero-angle projection ROIs (128×128 pixels) from scans with $P = 1, 5$, and 15 total projections. Applying Eq. (3) with fixed total exposure ($N_0 = 10^{10}$ source photons) and increasing P leads to increased quantum noise per projection.

and highly dependent on the measurement conditions, but one author estimated the overall electronic noise to be about 1000 electrons.⁴² Another group estimated the standard deviation at 2200 electrons.⁶ For a theoretical reference point with units matching Eq. (3), we assumed a simple model of electronic noise given by $\sigma_{\text{elec}}^2 = k_B T / C_{\text{pd}}$ where k_B is Boltzmann's constant, and T is the temperature. This corresponds to the reset noise for a gated integrator (p. 722 of Ref. 41). At room temperature, this noise model produces a voltage variance $\sigma_{\text{elec}}^2 = 1.67 \times 10^{-10}$ (volts)², corresponding to a standard deviation of ≈ 2500 electrons ($\sigma_{\text{elec}} / \Gamma$). Note that we used larger pixels than the references above. Also, σ_{elec}^2 may underestimate the variance in real detectors because we neglected other noise sources such as dark current and $1/f$ noise. Still, we report a value to give context to the results in Sec. IV.A.

II.D. Dose considerations

Working in simulation, it is important to relate our study conditions to standard measures of breast dosimetry. Using the dose tally feature in MC-GPU version 1.2, we computed the mean glandular dose (MGD) as a function of projection angle for all 500 phantoms in the signal-absent phantom set (Fig. 5). We assumed that the MGD was symmetric about $\theta = 0^\circ$. In MC-GPU, we simulated $N_0 = 10^{10}$ source photons at each projection angle. At each angle, we recorded the average dose tally for each phantom material type in units of electron volts per gram per source photon and converted the average glandular dose to milliGray per source photon. At $\theta = 0^\circ$, the mean glandular dose was $\approx 4 \times 10^{-12}$ mGy per source photon. The 1997 MQSA/NEXT survey average in NCRP Report 149 for mean glandular dose in mammography was 1.6 mGy, corresponding to $N_0 \approx 4 \times 10^{11}$ source photons in our model.

III. TASK-BASED IMAGE QUALITY ASSESSMENT

In Secs. III.A–III.D, we provide background on task-based assessment and key relations for evaluating DBT systems. Section III.A frames the signal detection task. Section III.B describes channels and considerations for designing a robust observer model. Section III.C gives details about the figure of merit we used to evaluate systems. Finally, Sec. III.D introduces covariance decomposition in the context of DBT as a tool for accelerating system evaluation and visualizing trade-offs between various parameters.

FIG. 5. The mean glandular dose (in milliGray per source photon) declined slightly with projection angle θ (in degrees). This trend qualitatively agrees with results from Sechopoulos *et al.* (Ref. 11) for a phantom with similar chest-to-nipple distance (e.g., Figs. 8(c) and 9(c) of Ref. 44). Error bars are $\pm 2\sigma$ based on a set of 500 phantoms.

III.A. Detection of small masses

In mammography, one important task is to detect small, low-contrast masses in the breast. To understand DBT's potential advantages for performing the same task, we started with the task of detecting small signals in the raw multiprojection data. It is important to first quantify signal detectability in the raw data because different postprocessing and reconstruction algorithms may lead to differences in task performance.⁴³

We represented the discrete-to-discrete x-ray imaging process with the imaging equation:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n}, \quad (4)$$

where $\mathbf{f}$ is the voxelized phantom object ($N \times 1$ voxels), $\mathbf{H}$ is a nonlinear operator ($M \times N$) which approximates the physical interactions that occur between x-ray source and detector plane, $\mathbf{n}$ is the measurement noise, and $\mathbf{g}$ is the resulting discrete projection image ($M \times 1$ pixels). In a multiprojection image acquisition with P projections, $\mathbf{H}$ changes with each projection angle. If we stack these $\mathbf{H}$ operators to form a multiprojection $\mathbf{H}$ operator ($M' \times N$) and concatenate the projections from a DBT scan into a single vector, Eq. (4) still holds with $M' = PM$. In Sec. II, we describe a simulation framework for Eq. (4), modeling $\mathbf{n}$ with a combination of quantum and electronic noise. In practice, the multiprojection $\mathbf{H}$ operator was too large to save (~ 40 terabytes for $P = 25$), so only $\mathbf{f}$ and $\mathbf{g}$ were stored.

Consider a detection task with n_0 signal-absent and n_1 signal-present DBT scans given by

$$\mathbf{g}_i = \mathbf{H}\mathbf{f}_i + \mathbf{n}, \quad i = 1, \dots, n_0 \quad (5)$$

$$\mathbf{g}_j = \mathbf{H}(\mathbf{f}_j + \mathbf{s}_j) + \mathbf{n}, \quad j = 1, \dots, n_1 \quad (6)$$

where we have used different subscripts to indicate that the $\mathbf{f}_i$ and $\mathbf{f}_j$ phantoms come from independent phantom sets. The vector $\mathbf{s}_j$ is the signal in the j th phantom. As described in

Sec. II.B, s_j corresponds to *changes* in the voxel density coefficients inside the 3 mm signal volume. Therefore s_j depends on the pre-existing density coefficients in $\mathbf{f}_j$ at the signal location.

III.B. Dimensionality reduction using channels

Linear discriminant analysis (LDA) is a common tool for measuring the separability of two data classes, as in our detection task. In the medical imaging community, the ideal linear discriminant is known as the Hotelling observer (HO).⁴⁴ It involves computing a score t for each data set $\mathbf{g}$ using a linear template as follows:

$$t = \Delta \bar{\mathbf{g}}^t \mathbf{K}_{\mathbf{g}}^{-1} \mathbf{g}, \quad (7)$$

where $\Delta \bar{\mathbf{g}} = \bar{\mathbf{g}}_1 - \bar{\mathbf{g}}_0$ is the mean difference in the data under the signal-present and signal-absent classes ($PM \times 1$, superscript t for transpose) and $\mathbf{K}_{\mathbf{g}}^{-1}$ is an inverse intraclass covariance matrix ($PM \times PM$). Standard LDA depends on estimating the data covariance matrix from samples, which is usually not possible when the data dimension PM exceeds the number of data samples $n = n_0 + n_1$. To circumvent this problem, we performed dimensionality reduction (aka feature transformation) using channels which extract salient features from the data for the given task. For example, with $P = 25$ projections and $M = 72\,800$ pixels per projection, standard LDA would require estimation and inversion of an exceedingly large matrix ($P^2 M^2 > 10^{12}$ elements). We would need *at least* as many phantoms as matrix elements to estimate $\mathbf{K}_{\mathbf{g}}$ by standard methods. Instead, we first channelized each projection image by applying Laguerre-Gauss (LG) channels (Fig. 6). We chose LG channels for two reasons: (1) our signal is approximately rotationally-symmetric, and (2) Witten *et al.* showed that for both circular and elliptical signals in non-Gaussian, randomly-varying backgrounds, channelized-ideal observer performance approached the ideal observer's performance with relatively few LG channels (Fig. 6 of Ref. 45). Barrett *et al.* originally proposed using LG channels in similar types of detection tasks since they form a basis on the space of rotationally-symmetric functions.⁴⁶ The channelization step is given by

$$\mathbf{v} = \mathbf{U}' \mathbf{g}, \quad (8)$$

where $\mathbf{U}$ is a channel matrix with dimensions $PM \times PC$ and C is the number of channels. Projecting onto the LG basis reduces the size of each multiprojection data vector from $PM \times 1$ to $PC \times 1$. Reformulated in terms of channel outputs, Eq. (7) becomes

$$t = \Delta \bar{\mathbf{v}}^t \mathbf{K}_{\mathbf{v}}^{-1} \mathbf{v}, \quad (9)$$

where the covariance $\mathbf{K}_{\mathbf{v}}$ now has dimensions $PC \times PC$. This is the channelized-Hotelling observer (CHO) score – the score computed in the LG feature space.

The inverse operation in Eq. (9) can still present difficulties, particularly when the covariance matrix is ill-conditioned. In this work, we used MATLAB's truncated SVD pseudoinverse algorithm (`pinv` function) to perform the in-

FIG. 6. Zoomed ROIs (only the central 24×24 pixels shown) from the first six (of C) LG channels used to reduce the dimensionality of $\mathbf{g}$. The LG width parameter $a = 6$, and the full ROIs were 128×128 to match the image dimensions in Fig. 4.

verse. Singular values below a default tolerance are set to zero.

An important question for robust system evaluation is how to choose the LG width parameter a and the number of channels C . For a rotationally-symmetric signal, Gallas and Barrett argued that the CHO's performance should asymptote to the performance of the ideal observer as the number of channels increases.⁴⁷ Then, the optimal strategy is to choose a and C to achieve the asymptotic limit of the ideal observer with as few channels as possible. Gallas' simulated signals and backgrounds differ from ours, but we followed a similar line of reasoning. Our problem is complicated by the fact that the signal loses rotational symmetry at larger projection angles. The signal diameter varies from 3.3 mm in the 0° projection to 3.45 mm in the 48° projection (≈ 6 –7 pixels). To address this variation, we started with $a = 6$ pixels (the signal diameter before magnification/angular effects). Then, we computed the area under the receiver operating characteristic curve (AUC) as a function of C for the smallest ($\theta = 0^\circ$) and largest ($\theta = 48^\circ$) projection angles in our study (Fig. 7). There was no electronic noise, and the simulation parameters were set as in Table III. Since the signal shape changes with angle, the optimal C could also change with θ . Based on Fig. 7, we chose $C = 6$ as a reasonable compromise between achieving asymptotic AUC at all θ and maintaining a small, consistent number of channels for practical figure-of-merit calculations.

III.C. Figures of merit

To calculate figures of merit for signal detectability, we first rearranged the channel data into training and testing subsets of equal size ($n_{\text{train}} = n_{\text{test}} = 500$ scans). In each subset, half of the phantoms were signal-absent and half contained embedded signals. From Eq. (9), the score for each testing scan takes the form

$$t = \widehat{\Delta \bar{\mathbf{v}}}^t \widehat{\mathbf{K}_{\mathbf{v}}}^{-1} \mathbf{v}, \quad (10)$$

FIG. 7. AUC versus the number of channels for $\theta = 0^\circ$ and $\theta = 48^\circ$ projection angles. We chose $C = 6$ for approximately asymptotic performance. $N_0 = 5 \times 10^{11}$ and $\sigma_{\text{elec}}^2 = 0$.

where $\widehat{\Delta\bar{v}}$ is the mean difference in the channelized multi-projection *training* data under the signal-present and signal-absent classes ($PC \times 1$) and $\widehat{\mathbf{K}}_v^{-1}$ is the inverse intraclass covariance matrix ($PC \times PC$). The hats indicate quantities estimated from samples. The first term ($\widehat{\Delta\bar{v}}$) is just a sample mean, and we will discuss covariance estimation in Sec. III.D. After computing scores for all 500 testing scans, we used the `perfcurve` function in MATLAB's Statistical Toolbox v7.6 to estimate the area under the ROC curve (AUC) along with bootstrap confidence intervals.

III.D. Covariance decomposition

One way to estimate AUC is to generate noisy sample projections, channelize the data, compute their scores by Eq. (10), and then compute an ROC curve. Generating thousands of noisy sample projections is time-consuming, however, so we sought faster alternatives to estimate AUC more efficiently across a range of system parameters.

Under broad assumptions, we can write the total covariance in Eq. (7) as a sum of components:⁴⁴

$$\mathbf{K}_g = \mathbf{K}_g^{\text{poiss}} + \mathbf{K}_g^{\text{gain}} + \mathbf{K}_g^{\text{elec}} + \mathbf{K}_g^{\text{obj}}. \quad (11)$$

The “poiss” term is the Poisson photon-counting noise, the “gain” term describes the noise in the gain processes, the “elec” term lumps together all of the electronic/excess noise effects, and the “obj” term is the covariance in the data due to anatomical variations in the breast phantoms. The full decomposition includes other terms, but we ignored them in this work. Also, this model assumes that the signals from the mass lesions are weak, so the covariance components are effectively unperturbed by the presence of masses (i.e., $\mathbf{K}_{g|0} \approx \mathbf{K}_{g|1}$). Note that although Eq. (11) is written as a sum, the terms contributing to the total covariance are not all statistically-independent. For example, changing the characteristics of the breast phantom ensemble could change both the “poiss” and “obj” terms in the decomposition. Expanding the first three terms of Eq. (11) according to the detec-

tor model in Sec. II.C, we can write (in the channelized data space):

$$\widehat{\mathbf{K}}_v = \frac{\eta A N_0}{P \bar{\mathcal{E}}} (\bar{k} \Gamma)^2 \mathbf{U}^t \text{Diag}(\widehat{\bar{\mathbf{p}}}) \mathbf{U} + \mathbf{U}^t \mathbf{K}_g^{\text{elec}} \mathbf{U} + \widehat{\mathbf{K}}_v^{\text{obj}}, \quad (12)$$

where $\text{Diag}(\widehat{\bar{\mathbf{p}}})$ is a $PM \times PM$ diagonal matrix with the mean noise-free multiprojection vector $\bar{\mathbf{p}}$ along the diagonal; $\mathbf{K}_g^{\text{elec}}$ is a $PM \times PM$ electronic noise matrix; and $\widehat{\mathbf{K}}_v^{\text{obj}}$ is a $PC \times PC$ matrix. As before, hats indicate the quantities that we estimate from sample data. We assumed independent, identically-distributed electronic noise, so we can replace $\mathbf{U}^t \mathbf{K}_g^{\text{elec}} \mathbf{U}$ with $\sigma_{\text{elec}}^2 \mathbf{I}$ since $\mathbf{U}$ is unitary for LG channels. Note that due to the noise-free gain assumption in Sec. II.C, the first two terms in Eq. (11) reduced to a single, diagonal “quantum noise” matrix. We found sparse matrix tools helpful for computing this term.

Next, we modified the object covariance $\mathbf{K}_v^{\text{obj}}$ to express it in terms of noise-free channelized projection data $\mathbf{v}_p$, where $\mathbf{v}_p = \mathbf{U}^t \mathbf{p}$. This introduces a quadratic dependence on the total number of source photons N_0 , and the total covariance becomes:

$$\begin{aligned} \widehat{\mathbf{K}}_v = & \frac{\eta A N_0}{P \bar{\mathcal{E}}} (\bar{k} \Gamma)^2 \mathbf{U}^t \text{Diag}(\widehat{\bar{\mathbf{p}}}) \mathbf{U} + \sigma_{\text{elec}}^2 \mathbf{I} \\ & + \left[\frac{\eta A N_0 \bar{k} \Gamma}{P \bar{\mathcal{E}}} \right]^2 \widehat{\mathbf{K}}_{v_p}^{\text{obj}}. \end{aligned} \quad (13)$$

This equation shows how the DBT data covariance matrix in our model depends on key system and object parameters. It contains two diagonal parts and a structured object covariance part. Using Eq. (13), we can estimate the total covariance for systems having different mean x-ray energies ($\bar{\mathcal{E}}$), quantum efficiencies (η), gain parameters ($\bar{k}$, Γ), total numbers of source photons (N_0), numbers of projections (P), or electronic noise variances (σ_{elec}^2). Only two terms – $\bar{\mathbf{p}}$ and $\widehat{\mathbf{K}}_{v_p}^{\text{obj}}$ – are estimated from the training set of n_{train} noise-free DBT scans. Due to the weak-signal assumption, we estimated $\bar{\mathbf{p}}$ from the signal-absent scans only and computed $\widehat{\mathbf{K}}_{v_p}^{\text{obj}}$ as the average of the signal-absent and signal-present sample covariances. The same n_{train} scans were used to estimate the sample mean ($\widehat{\Delta\bar{v}}$). Then, by Eq. (10), it is straightforward to compute scores and AUC. This saves time compared to the alternative of simulating noisy projections for every possible combination of the above system parameters. Note, however, that in this study we limited the phantom ensemble to one breast volume class from a specific phantom-generating algorithm, so $\Delta\bar{v}$, $\widehat{\mathbf{K}}_{v_p}^{\text{obj}}$, and AUC should be re-computed for different phantom sizes and phantom-generating models.

IV. RESULTS

IV.A. Covariance decomposition

For a given task, we want to know which noise source is the limiting factor in observer performance. The covariance decomposition described in Sec. III.D makes it straightforward to visualize noise tradeoffs in our model (Fig. 8). Each data point in the figure was calculated by changing the values of

FIG. 8. AUC map as a function of total MGD for the whole scan and electronic noise variance (σ_{elec}^2) illustrates the tradeoffs in our covariance decomposition noise model. In this case $P = 5$ and $\theta_{\text{scan}} = 64^\circ$, and the dotted lines are approximate boundaries between different performance-limiting factors.

N_0 and σ_{elec}^2 in the covariance decomposition and recomputing AUC. We then computed the corresponding MGD using the normalized dose coefficients from Fig. 5.

By visual inspection of Fig. 8, we can divide the AUC map into approximate zones corresponding to ranges of MGD and σ_{elec}^2 where the performance-limiting factor is anatomical, quantum, or electronic in nature. In the specific case shown ($P = 5$ and $\theta_{\text{scan}} = 64^\circ$), the dominant “noise” is anatomical variability for doses higher than about $40 \mu\text{Gy}$. Above $40 \mu\text{Gy}$, the AUC appears to *saturate* at a value close to 0.95. In this regime, more dose does not lead to better image quality – the image quality is limited by the anatomical backgrounds present in our phantoms. Below $40 \mu\text{Gy}$, we can distinguish the quantum and electronic regimes by drawing a vertical dividing line approximately where the slope of the rainbow band changes from zero to positive. This is the point where σ_{elec}^2 is large enough that electronic noise begins to dominate.

For a typical single-view mammography dose of about 1.6 mGy , we are clearly operating in the Anatomical regime in Fig. 8. However, the behavior of Fig. 8 may depend on the scan angle and number of projections. Section IV.B focuses on AUC tradeoffs at a clinically-relevant dose point in the Anatomical regime, but we also investigate two other test cases in the MGD- σ_{elec}^2 space (Table IV). The total MGD for the whole scan varies with P , so the cases are described in terms of a fixed N_0 and an approximate MGD value.

TABLE IV. Parameters of the test cases in Figs. 9 and 10.

Abbr.	N_0	Approx. MGD	σ_{elec}^2	Limiting factor
A	5×10^{11}	2 mGy	0	Anatomical variability
Q	10^8	$0.4 \mu\text{Gy}$	0	Quantum noise
E	10^8	$0.4 \mu\text{Gy}$	10^{-9}	Electronic noise

IV.B. Image quality tradeoffs: scan angle and number of projection angles

The DBT scan angle and number of projections are key image quality parameters that separate DBT from conventional mammography. In this section, we isolate these two parameters and evaluate AUC as a function of each. First, we analyzed AUC versus scan angle for four different numbers of projections (Fig. 9). For each test case in Table IV, AUC was computed using covariance decomposition and MATLAB LDA tools as described in Secs. III.C–III.D. The projection angles were evenly-spaced, so the number of points in each plot varies depending on whether projection data were available at the required sampling angles for a given scan angle. The error bars are 95% confidence intervals computed from 100 bootstrap samples.

Figure 9 shows that in the Anatomical (A) regime, AUC does not appear to depend on scan angle. The trends suggest a *slight* increase in AUC at larger scan angles for $P > 5$, but the error bars are overlapping so it is difficult to conclude that increasing the scan angle improves detectability. In the Q and E regimes, the trends look similar except at the largest scan angles where AUC appears to decrease slightly relative to the A regime curves.

To evaluate the impact of the number of projections, we fixed the scan angle and evaluated different subsets of equally-spaced projection angles. For example, with $\theta_{\text{scan}} = 96^\circ$, $P = 3$ gives the subset $\theta = 0, \pm 48^\circ$ and $P = 5$ gives the subset $\theta = 0, \pm 24, \pm 48^\circ$. $P = 1$ corresponds to single-view mammography, so the AUC value is the same for all scan angles. We computed AUC versus P at three different scan angles for the three regimes in Table IV (Fig. 10). For $\theta_{\text{scan}} = 24^\circ$ and 60° , AUC clearly increases from $P = 1$ to $P = 3$ in the A regime, but there appears to be minimal improvement when $P > 3$. For $\theta_{\text{scan}} = 96^\circ$, however, the AUC in the A regime seems to increase more gradually and reach a threshold around $P = 7$ projections. For all three scan angles, the A and Q regime curves trend upward, while the E regime curves generally decrease with P . Taken as a whole, Figs. 9 and 10 suggest that to maximize the AUC for our task, we can operate at any of the three scan angles; but the optimal number of projections depends on the chosen scan angle.

The $\theta_{\text{scan}} = 60^\circ$ and 96° curves do not appear to represent smooth functions. At $P = 5$ in particular, something causes the AUC to deviate from neighboring data points. We suspect that this is caused by certain angles having relatively stronger/weaker signal content than others due to the characteristics of the phantom model. For example, $\theta_{\text{scan}} = 60^\circ$ and $P = 5$ corresponds to $\theta = 0^\circ, \pm 15^\circ, \pm 30^\circ$. For the same scan angle, $P = 7$ corresponds to $\theta = 0^\circ, \pm 10^\circ, \pm 20^\circ, \pm 30^\circ$. The AUC “deviation” at $P = 5$ should then be related to the differences between the projections at $\theta = \pm 10^\circ, \pm 15^\circ, \pm 20^\circ$. This appears to be a consequence of using a nonuniform phantom model. While many studies rely on uniform phantoms, our results suggest that it may be clinically important to investigate nonuniform phantom effects to understand the effects of patient variability on DBT design choices.

FIG. 9. In the anatomical-variability-limited (A) regime, AUC appears to be independent of scan angle. In the Q and E regimes, the trends appear similar to the A regime trends except at large θ_{scan} . Error bars are 95% confidence intervals computed from 100 bootstrap samples.

V. LITERATURE ANALYSIS

Several groups have published simulation studies for evaluating DBT systems in a virtual environment. In this section, we summarize a number of studies in this vein. Figure 11 is not exhaustive, as it does not include studies involving physical prototype DBT systems. We have limited the comparison to DBT studies where the patient population and image acquisition are entirely virtual, as in our study. Two of the studies in Fig. 11 used human readers,^{6,48} but we still consider those studies to be consistent with a virtual trial paradigm. Also, in the Das and Sechopoulos studies, some details of their noise models are cross-referenced, so we have included relevant reference numbers^{49–54} in the appropriate boxes in Fig. 11.

VI. DISCUSSION

Image quality metrics for assessing DBT systems abound in the literature. Some metrics are related to each other by simple transformations (e.g., AUC and SNR_{CHO}), while others come from entirely different interpretations of image quality. In this work, we used a spatial-domain observer model (in projection space) with the AUC figure of merit because

this approach offers distinct advantages over other spatial and Fourier domain approaches. Most importantly, AUC is a task-specific metric. It depends on all the parameters of the detection task, such as signal size, detector noise correlations, breast phantom variability, pixel size, and observer strategy. By contrast, a task-generic metric like CNR assumes pixel-independent noise in the data (no correlations).

Working in projection space is both a limitation and an advantage. It is a limitation because human observers only look at reconstructions. However, it is advantageous to separate the analysis of DBT systems into analysis of the raw projection images (with model observers) and analysis of the reconstruction algorithms we use to visualize the data (with human or model observers). Take, for example, the result in Fig. 10 which says that for $\theta_{\text{scan}} = 24^\circ$, $P = 3$, and $P = 9$ give about the same AUC. While a reconstruction from 3 projections would generally be considered unacceptable due to artifacts, this indicates a failure in the reconstruction algorithm, not necessarily a lack of information in the raw data. Our results suggest that some information gets lost in the reconstruction process for $P = 3$. The only way to measure this detectability loss is with a projection-space reference point.

FIG. 10. Multiprojection imaging ($P > 1$) improved AUC for all three scan angles (in the A regime), though the optimal number of projections appears to depend on scan angle. For $\theta_{\text{scan}} = 96^\circ$, there is a threshold number of projections around $P = 7$ for maximum AUC. Error bars are 95% bootstrap confidence intervals.

A logical next step is to apply this framework to detecting microcalcifications. In this study, we were limited by the phantom voxel size – $500 \mu\text{m}$ was state-of-the-art at the time of this work. This leads to voxel-edge artifacts in the images when the pixel size is less than the voxel size. By using larger $500 \mu\text{m}$ pixels, we avoided voxel-edge artifacts at the cost of somewhat distorting the task; larger pixels may explain why the AUC remained high even at an exposure level of $N_0 = 10^9$ source photons, which corresponds to less than 1% of typical mammography dose. In the task of detecting a 3 mm mass, anatomical noise appears to be the dominant noise source, since we only saw AUC degradation for ultra-low dose levels. However, for the task of detecting microcalcifications, quantum noise could start to dominate at a higher dose threshold (i.e. the “Quantum” zone in Fig. 8 could shift upwards). Based on a recently updated method for the generation of Bakic software phantoms,⁵⁵ it may be possible to efficiently simulate breast anatomy with smaller voxel sizes and investigate microcalcification detection within our framework.

We used a homogeneous phantom to model scattering in this study. There is some precedent for using the scatter from a homogeneous phantom (e.g., Refs. 6 and 11), though we have not seen a DBT study quantifying the impact of randomly-varying scatter from an ensemble of heterogeneous phantoms on lesion detectability. While breast compression was used in the phantom model, there was no breast support plate in this work. Since the breast support plate affects scattering in an angle-dependent fashion, we expect that it could impact the magnitude of our AUC values and the AUC trends versus scan angle. However, these effects are likely small compared to the effects of breast compression. In this work, we used a binarized version of the UPenn phantom model for simplicity (i.e., fewer tissue material definitions). However, the penEasy and MC-GPU codes are general and they can accept any voxelized phantom model. Non-binary phantom models are worth exploring to increase the clinical relevance of our simulated detection tasks with little change in total computation time.

First author Reference #	Young this work	Gong 6	Reiser 15	Das 48	Sechopoulos 11
Virtual Population	1000 region-growing phantoms	120 power-law phantoms	20 power-law phantoms	8 phantoms derived from mastectomy specimens	50 power-law phantoms
Compressed breast thickness (cm)	5.15	5	5	5	5
Fibroglandular fraction	25%	50%	75%	Not specified	50%
Task	Detect a 3-mm-diameter lesion.	Detect a 5-mm-diameter, spiculated lesion.	Detect lesions ranging from 0.5-8 mm in diameter.	Detect and localize lesions 1 cm ³ or less in size (also microcalcs).	Detect a 6-mm-diameter lesion (also 0.4-mm cubic microcalcs).
X-ray spectrum	20 keV (monoenergetic)	Mo/Rh at 28 kVp	20 keV (monoenergetic)	Mo anode at 30 kVp	Mo/Mo at 26 kVp
Focal spot size (mm)	Point source	0.3	Point source	0.3	Not specified
Geometry type	Partially-isocentric	Isocentric	Partially-isocentric	Isocentric	Partially-isocentric
Compression plate	Not modeled	Not specified	Not specified	Not specified	2 mm PMMA
Source-detector distance at 0° (cm)	65	66	66	Not specified	66
Maximum scan angle	96°	30°	60°	60°	60°
Maximum number of views	97	11	41	21	61
Ray tracing algorithm	Siddon's	Siddon's	Siddon's	Siddon's	Siddon's
Ray tracing code	penEasy_Imaging	In-house	In-house	In-house	GEANT
Scatter model	MC simulation of a 25% glandular uniform phantom at each view angle	MC simulation of a 50% glandular uniform phantom at each view angle	Not modeled	Not modeled	Assumed a constant SPR at each angle, determined previously.
Scatter code	MC-GPU (adapted for DBT)	GEANT	N/A	N/A	GEANT
Pixel size (μm)	500	100	100	100	100
Quantum noise model	Scaled Poisson distribution	Scaled Poisson distribution	Poisson distribution	Cross-referenced from [50-52]	Cross-referenced from [53-55]
Scintillator blur	Not modeled	Modeled using measured MTFs	Not modeled	Modeled using measured MTFs	Modeled using measured MTFs
Scintillator thickness (μm)	N/A	100	N/A	100	Not specified
Electronic noise model	Additive Gaussian process	Additive Gaussian process	Not modeled	Cross-referenced from [50-52]	Cross-referenced from [53-55]
Reconstruction algorithm	Not modeled	Backprojection	MLEM (8 iterations)	Filtered-backprojection	MLEM (4 iterations)
Reconstructed voxel size (mm)	N/A	0.1 × 0.1 × 1	0.1 × 0.1 × 1	0.2 × 0.2 × 0.2	0.1 × 0.1 × 1
MGD (mGy)	2, 1.6, 0.0004	4	Not specified	4	2

FIG. 11. Comparison of five virtual DBT trial frameworks (continued on the next page).

First author	Young	Gong	Reiser	Das	Sechopoulos
Reference #	this work	6	15	48	11
Readers	Spatial-domain CHO (mpCHOC variant)	5 medical physicists	Fourier-domain PW observer	4 medical physicists	Not specified
Study paradigm	500 training cases (half with signal), 500 testing cases (half with signal)	20 training cases (half with signal), 100 testing cases (half with signal)	20 cases with resubstitution (i.e. no training/testing distinction)	10 training cases (half with signal), 82 testing cases (half with signal) created from 8 phantoms	150 cases (50 with mass, 50 with microcalc) with resubstitution
Figure of merit	AUC	AUC	Detectability index	Area under the LROC curve	Quality factor
Error analysis	Bootstrapping (for AUC error bars)	Fitting scores with ROCKIT software and using ANOVA tests to compare systems	Bootstrapping (for detectability index error bars)	Fitting scores with Swensson's LROC software and using two-way ANOVA tests to compare uniform and variable-dose schemes	Not specified
Abridged DBT image quality findings (and their p-values, if specified)	1. AUC is constant or slightly increasing with scan angle.	1. AUC is improved versus digital mammography. ($p=0.003$)	1. Detectability increases with scan angle for all signal sizes.	1. For masses, observers performed better with UD-DBT than VD-DBT. ($p=0.22$, not significant)	1. Maximizing the scan angle always increases vertical resolution.
	2. AUC increases asymptotically with the number of projections.		2. While detectability of larger signals is primarily affected by scan angle, detectability of smaller signals is limited by both quantum noise and number of views.	2. For microcalcifications, observers did better with UD-DBT than VD-DBT. ($p=0.038$)	2. Increasing the number of projections beyond a relatively low threshold does not increase the vertical resolution further.

GLOSSARY (in alphabetical order)

ANOVA:	Analysis of variance
AUC:	Area under the (receiver operating characteristic) curve
CHO:	Channelized-Hotelling observer
Isocentric:	Source and detector rotate about a common point
LROC:	Localization receiver operating characteristic
MC:	Monte Carlo
MGD:	Mean glandular dose
MLEM:	Maximum-likelihood expectation maximization
mpCHOC:	A multiprojection variant of the Channelized-Hotelling observer (see Ref. [55])
MTF:	Modulation transfer function
N/A:	Not applicable to the study in question
Partially-isocentric:	Source rotates about a point, while detector plane is fixed
PMMA:	Polymethyl methacrylate
Quality factor:	The normalized contrast-to-noise ratio of a mass divided by the artifact spread function width of a microcalci
UD-DBT:	Uniform-dose DBT (equal dose in each projection)
VD-DBT:	Variable-dose DBT (half the dose in the central projection, the rest divided equally)

FIG. 11. (Continued).

A strength of the framework is the use of large sets of anthropomorphic phantoms. There is no standard patient/phantom ensemble size in the DBT assessment literature for demonstrating that image quality gains are statistically significant. Anatomical noise may be the limiting factor in many imaging situations, so it is important to model anatomical variability as realistically as possible. We treated this often overlooked issue by using a large phantom ensemble and computing bootstrap confidence intervals on our AUC metrics, but we discovered that we may need even larger phantom ensembles to reduce the variance to acceptable levels. Fukunaga and Hayes⁵⁶ showed that the number of data samples required for a reliable estimate of the covariance of a classifier grows quickly with the size of the data vector. For example, their criteria suggest that a 128-element data vector requires $n = 9396$ samples, or a factor of 73 times the vector size. In this work, we were limited to $n = 500$ scans for training the model observer. This corresponds to a factor of about 28 times the smallest vector size $PC = 18$ and a factor of only 3.5 times the largest channel vector size $PC = 150$. Therefore, an area of future work is to improve the reliability of our covariance estimates by increasing the size of our phantom sets or using more efficient channels.

Choosing appropriate channels to reduce images to their salient features is an important and active area of research.^{45,57} Efficient channels reduce the data dimensions as much as possible while yielding approximately the same AUC as the original (non-reduced) data. In other words, $AUC_{CHO}/AUC_{HO} \approx 1$ for an efficient CHO. We used 2D Laguerre-Gauss channels because LG channels seem to be robust for capturing statistical information and approximating the HO for circular and elliptical signals in non-Gaussian, randomly-varying backgrounds.⁴⁵ Diagnostic findings correspond to three-dimensional lesions, however, so 2D LG channels may not be efficient for real clinical tasks. In addition, the effects of using rotationally-symmetric LG channels on asymmetric signals have not been thoroughly investigated, to our knowledge. New 3D observer model and channel designs, like those proposed by Platiša *et al.*,⁵⁷ could improve our understanding of the theoretical limits of detectability.

The aim of this work was to provide a framework for determining the upper bounds of signal detectability in projection space for different DBT system configurations. Abbey and Boone have shown – using Markov-chain Monte Carlo (MCMC) – that the ideal observer outperforms the HO in a binary model of breast parenchymal tissue.⁵⁸ However, their work required a parametric, correlated Gaussian phantom model and smaller data dimensions. For the large phantom and data dimensions in our framework, it is infeasible to estimate the true covariance required by the MCMC approach. Thus we were limited in using the channelized-HO as an approximation to the ideal observer.

In Sec. III.D, we introduced covariance decomposition in the context of DBT as a tool for isolating the effects of different noise sources and accelerating DBT evaluation over a range of system parameters. With assumptions about the form of the data covariance matrix as in Eq. (13), we can quickly evaluate systems at different dose levels as we did in Fig. 8,

for example. We can characterize the impact of additional noise sources simply by adding terms to the decomposition in Eq. (11). We hope to extend the covariance decomposition approach to more realistic noise models in future work.

For the purposes of framework development, we neglected some additional detector characteristics such as spatial blur, excess noise in the gain mechanism, and the effect of reabsorbed Compton-scattered and characteristic x rays. In future work, we can improve the detector model by incorporating energy-, depth-, and angle-dependent information.^{59–62} As the simulations grow faster and more efficient, it should also become practical to investigate polyenergetic source models and scatter models that incorporate the effects of nonuniform phantoms. More realistic physics models combined with computationally efficient model observers can lead to faster and more accurate assessment of image quality in DBT.

VII. CONCLUSIONS

We developed a simulation framework for quantifying DBT signal detectability from projection images. The framework incorporates anthropomorphic voxelized phantoms, open-source tools for modeling the forward problem and estimating dose, and a channelized-Hotelling observer model with covariance decomposition for computing figures of merit. We studied detectability tradeoffs as a function of the DBT scan angle and number of projections, and compared our findings with relevant conclusions from other authors. We believe our framework is a viable tool for further investigation of image quality tradeoffs in DBT through task-based objective assessment.

ACKNOWLEDGMENTS

The author would like to acknowledge Dr. Matthew Kupinski for helpful discussions and reading early drafts, Drs. Jonathan Boswell and Andreu Badal for their computing expertise and help in the use of the FDA in-house compute cluster and Monte Carlo codes, Dr. Harrison Barrett for his thoughts on noise, and Mr. Kyle Anderson for his early work on PenVT. The authors would also like to acknowledge funding support from the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Oak Ridge Institute for Science Education (ORISE) Research Participation Program at the Center for Devices and Radiological Health (CDRH) of the FDA, and the FDA Office of Women's Health.

^{a)}Author to whom correspondence should be addressed. Electronic mail: subok.park@fda.hhs.gov

¹J. T. Dobbins III, "Tomosynthesis imaging: At a translational crossroads," *Med. Phys.* **36**, 1956–1967 (2009).

²J. A. Baker and J. Y. Lo, "Breast tomosynthesis: State-of-the-art and review of the literature," *Acad. Radiol.* **18**, 1298–1310 (2011).

³R. J. Jennings, R. J. Eastgate, M. P. Siedband, and D. L. Ergun, "Optimal x-ray spectra for screen-film mammography," *Med. Phys.* **8**, 629–639 (1981).

⁴R. Fahrig, A. D. A. Maidment, and M. J. Yaffe, "Optimization of peak kilovoltage and spectral shape for digital mammography," *Proc. SPIE* **1651**, 74–83 (1992).

- ⁵R. Klausz and N. Shramchenko, "Dose to population as a metric in the design of optimised exposure control in digital mammography," *Radiat. Prot. Dosim.* **114**, 369–374 (2005).
- ⁶X. 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," *Med. Phys.* **33**(4), 1041–1052 (2006).
- ⁷A. S. Chawla, E. Samei, R. Saunders, C. Abbey, and D. DeLong, "Effect of dose reduction on the detection of mammographic lesions: A mathematical observer model analysis," *Med. Phys.* **34**(8), 3385–3398 (2007).
- ⁸H. C. Gifford, C. S. Didier, M. Das, and S. J. Glick, "Optimizing breast-tomosynthesis acquisition parameters with scanning model observers," *Proc. SPIE* **6917**, 69170S-1–69170S-9 (2008).
- ⁹B. Zhao, J. Zhou, Y.-H. 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).
- ¹⁰T. Mertelmeier, J. Ludwig, B. Zhao, and W. Zhao, "Optimization of tomosynthesis acquisition parameters: Angular range and number of projections," in *Digital Mammography*, Lecture Notes in Computer Science Vol. 5116, edited by E. Krupinski (Springer, Berlin, 2008), pp. 220–227.
- ¹¹I. Sechopoulos and C. Ghetti, "Optimization of the acquisition geometry in digital tomosynthesis of the breast," *Med. Phys.* **36**, 1199–1207 (2009).
- ¹²S. Young, S. Park, S. K. Anderson, A. Badano, K. J. Myers, and P. Bakic, "Estimating breast tomosynthesis performance in detection tasks with variable-background phantoms," *Proc. SPIE* **7258**, 72580O-1–72580O-9 (2009).
- ¹³A. S. Chawla, J. Y. Lo, J. A. Baker, and E. Samei, "Optimized image acquisition for breast tomosynthesis in projection and reconstruction space," *Med. Phys.* **36**(11), 4859–4869 (2009).
- ¹⁴G. 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**(5), 1948–1965 (2010).
- ¹⁵I. S. Reiser and R. M. Nishikawa, "Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise," *Med. Phys.* **37**(4), 1591–1600 (2010).
- ¹⁶B. Ren, C. Ruth, T. Wu, Y. Zhang, A. Smith, L. Niklason, C. Williams, E. Ingal, B. Polischuk, and Z. Jing, "A new generation FFD/tomosynthesis fusion system with selenium detector," *Proc. SPIE* **7622**, 76220B (2010).
- ¹⁷S. Young, P. Bakic, K. J. Myers, and S. Park, "Performance tradeoffs in a model breast tomosynthesis system," in *Digital Image Processing and Analysis* (Optical Society of America, Washington, DC, 2010), p. DTuA3.
- ¹⁸F. Salvat, J. M. Fernandez-Varea, and J. Sempau, "PENLOPE, A Code System for Monte Carlo Simulation of Electron and Photon Transport," *Workshop Proceedings Issy-les-Moulineaux*, France 7–10 July 2003 (OECD Nuclear Energy Agency, Paris, France, 2003).
- ¹⁹P. R. Bakic, M. Albert, D. Brzakovic, and A. D. A. Maidment, "Mammogram synthesis using a 3D simulation. I. Breast tissue model and image acquisition simulation," *Med. Phys.* **29**(9), 2131–2139 (2002).
- ²⁰K. Bliznakova, Z. Bliznakov, V. Bravou, Z. Kolitsi, and N. Pallikarakis, "A three-dimensional breast software phantom for mammography simulation," *Phys. Med. Biol.* **48**, 3699–3719 (2003).
- ²¹C. Hoeschen *et al.*, "A high-resolution voxel phantom of the breast for dose calculations in mammography," *Radiat. Prot. Dosim.* **114**, 406–409 (2005).
- ²²I. Reiser, E. Sidky, R. Nishikawa, and X. Pan, "Development of an analytic breast phantom for quantitative comparison of reconstruction algorithms for digital breast tomosynthesis," in *Digital Mammography*, Lecture Notes in Computer Science Vol. 4046, edited by S. Astley, M. Brady, C. Rose, and R. Zwigelaar (Springer, Berlin, 2006), pp. 190–196.
- ²³J. Michael 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**, 691315 (2008).
- ²⁴K. Hussein, C. L. Vaughan, and T. S. Douglas, "Modeling, validation and application of a mathematical tissue-equivalent breast phantom for linear slot-scanning digital mammography," *Phys. Med. Biol.* **54**, 1533–1553 (2009).
- ²⁵C. M. Li, W. P. Segars, G. D. Tourassi, J. M. Boone, and J. T. Dobbins III, "Methodology for generating a 3D computerized breast phantom from empirical data," *Med. Phys.* **36**, 3122–3131 (2009).
- ²⁶C. 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-1–69180V-10 (2008).
- ²⁷P. R. Bakic, C. Zhang, and A. D. A. Maidment, "Development and characterization of an anthropomorphic breast software phantom based upon region-growing algorithm," *Med. Phys.* **38**, 3165–3176 (2011).
- ²⁸P. R. Bakic, M. Albert, D. Brzakovic, and A. D. A. Maidment, "Mammogram synthesis using a 3D simulation. II. Evaluation of synthetic mammogram texture," *Med. Phys.* **29**(9), 2140–2151 (2002).
- ²⁹O. Tischenko, C. Hoeschen, D. R. Dance, R. A. Hunt, A. D. A. Maidment, and P. R. Bakic, "Evaluation of a novel method of noise reduction using computer-simulated mammograms," *Radiat. Prot. Dosim.* **114**, 81–84 (2005).
- ³⁰D. Kontos, C. Zhang, N. Ruiter, P. Bakic, and A. Maidment, "Evaluating the effect of tomosynthesis acquisition parameters on image texture: A study based on an anthropomorphic breast tissue software model," in *Digital Mammography*, Lecture Notes in Computer Science Vol. 5116, edited by E. Krupinski (Springer, Berlin, 2008), pp. 491–498.
- ³¹M. J. Yaffe, J. M. Boone, N. Packard, O. Alonzo-Proulx, S.-Y. Huang, C. L. Peressotti, A. Al-Mayah, and K. Brock, "The myth of the 50–50 breast," *Med. Phys.* **36**, 5437–5443 (2009).
- ³²N. V. Ruiter, R. Stotzka, T.-O. Muller, H. Gemmeke, J. R. Reichenbach, and W. A. Kaiser, "Model-based registration of x-ray mammograms and mr images of the female breast," *IEEE Trans. Nucl. Sci.* **53**, 204–211 (2006).
- ³³N. V. Ruiter, C. Zhang, P. R. Bakic, A.-K. Carton, J. Kuo, and A. D. A. Maidment, "Simulation of tomosynthesis images based on an anthropomorphic software breast tissue phantom," *Proc. SPIE* **6918**, 69182I (2008).
- ³⁴P. C. Johns and M. J. Yaffe, "X-ray characterisation of normal and neoplastic breast tissues," *Phys. Med. Biol.* **32**, 675–695 (1987).
- ³⁵J. Sempau, A. Badal, and L. Brualla, "A PENLOPE-based system for the automated Monte Carlo simulation of clinacs and voxelized geometries—Application to far-from-axis fields," *Med. Phys.* **38**, 5887–5895 (2011).
- ³⁶A. Badal, "Development of advanced geometric models and acceleration techniques for Monte Carlo simulation in medical physics," Ph.D. dissertation, Univ. Politecnica de Catalunya, Catalunya, Spain (2008).
- ³⁷C. L. Carter, C. Allen, and D. E. Henson, "Relation of tumor size, lymph node status, and survival in 24740 breast cancer cases," *Cancer* **63**, 181–187 (1989).
- ³⁸G. 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).
- ³⁹A. Badal, I. Kyrianiou, D. P. 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).
- ⁴⁰J. A. Rowlands and J. Yorkston, "Flat panel detectors for digital radiography," *Handbook of Medical Imaging: Volume 1. Physics and Psychophysics* (Society of Photo-Optical Instrumentation Engineers, Bellingham, WA, 2000), Chap. 4, pp. 223–328.
- ⁴¹R. E. Burgess, "Homophase and heterophase fluctuations in semiconducting crystals," *Discuss. Faraday Soc.* **28**, 151–158 (1959).
- ⁴²H. Wiecek, "Physical aspects of detector design," *Radiat. Meas.* **33**, 541–545 (2001).
- ⁴³D. Van de Sompel, Sir M. Brady, and J. Boone, "Task-based performance analysis of FBP, SART, and ML for digital breast tomosynthesis using signal CNR and channelised Hotelling observers," *Med. Image Anal.* **15**, 53–70 (2011).
- ⁴⁴H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley, New York, 2004).
- ⁴⁵J. M. Witten, S. Park, and K. J. Myers, "Partial least squares: A method to estimate efficient channels for the ideal observers," *IEEE Trans. Med. Imaging* **29**, 1050–1058 (2010).
- ⁴⁶H. H. Barrett, C. K. Abbey, B. D. Gallas, and M. P. Eckstein, "Stabilized estimates of Hotelling-observer detection performance in patient-structured noise," *Proc. SPIE* **3340**, 27–43 (1998).
- ⁴⁷B. 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).
- ⁴⁸M. 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).
- ⁴⁹A. A. Vedula, S. J. Glick, and X. Gong, "Computer simulation of CT mammography using a flat-panel imager," *Proc. SPIE* **5030**, 349–360 (2003).
- ⁵⁰X. Gong, A. A. Vedula, and S. J. Glick, "Microcalcification detection using cone-beam CT mammography with a flat-panel imager," *Phys. Med. Biol.* **49**, 2183–2195 (2004).

- ⁵¹J. H. Siewerdsen, L. E. Antonuk, Y. El-Mohri, J. Yorkston, W. Huang, J. M. Boudry, and I. A. Cunningham, "Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (AMFPIs) for diagnostic radiology," *Med. Phys.* **24**, 71–89 (1997).
- ⁵²C. Ghetti, A. Borri, O. Ortenzia, R. Rossi, and P. L. Ordóñez, "Physical characteristics of ge senographe essential and ds digital mammography detectors," *Med. Phys.* **35**, 456–463 (2008).
- ⁵³R. S. Saunders and E. Samei, "A method for modifying the image quality parameters of digital radiographic images," *Med. Phys.* **30**, 3006–3017 (2003).
- ⁵⁴J. 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).
- ⁵⁵D. D. Pokrajac, A. D. A. Maidment, and P. R. Bakic, "Optimized generation of high resolution breast anthropomorphic software phantoms," *Med. Phys.* **39**, 2290–2302 (2012).
- ⁵⁶K. Fukunaga and R. R. Hayes, "Effects of sample size in classifier design," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 873–885 (1989).
- ⁵⁷L. Platiša, B. Goossens, E. Vansteenkiste, S. Park, B. D. Gallas, A. Badano, and W. Philips, "Channelized Hotelling observers for the assessment of volumetric imaging data sets," *J. Opt. Soc. Am. A* **28**, 1145–1163 (2011).
- ⁵⁸C. K. Abbey and J. M. Boone, "An ideal observer for a model of x-ray imaging in breast parenchymal tissue," in *Proceedings of the 9th international workshop on Digital Mammography, IWDM'08* (Springer-Verlag Berlin, Heidelberg Germany, 2008), pp. 393–400.
- ⁵⁹B. 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**(11), 3132–3149 (2004).
- ⁶⁰A. Badano, I. S. Kyprianou, R. J. Jennings, and J. Sempau, "Anisotropic imaging performance in breast tomosynthesis," *Med. Phys.* **34**(11), 4076–4091 (2007).
- ⁶¹I. S. Reiser, R. M. Nishikawa, and B. A. Lau, "Effect of non-isotropic detector blur on microcalcification detectability in tomosynthesis," *Proc. SPIE* **7258**, 72585Z (2009).
- ⁶²M. 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).