Non-Gaussian statistical properties of breast images

Craig K. Abbey[a)]
Department of Psychology, University of California, Santa Barbara, California 93106

Anita Nosratieh
Department of Biomedical Engineering, University of California, Davis, California 95616 and Department of Radiology, University of California Medical Center, Sacramento, California 95817

Jascha Sohl-Dickstein
Department of Applied Physics, Stanford University, Stanford, California 94305

Kai Yang
Department of Radiology, University of California Medical Center, Sacramento, California 95817

John M. Boone
Department of Biomedical Engineering, University of California, Davis, California 95616 and Department of Radiology, University of California Medical Center, Sacramento, California 95817

Purpose: Several studies have shown that the power spectrum of x-ray breast images is well described by a power-law at lower frequencies where anatomical variability dominates. However, an image generated from a Gaussian process with this spectrum is easily distinguished from an image of actual breast tissue by eye. This demonstrates that higher order non-Gaussian statistical properties of mammograms are readily accessible to the visual system. The authors' purpose is to quantify and characterize non-Gaussian statistical properties of breast images as influenced by processing of a digital mammogram, different imaging modalities, and breast density.

Methods: To quantify non-Gaussian statistical properties, the authors consider histograms of filter responses from the interior of a breast image that have similar properties to receptive fields in the early visual system. They quantify departure from a Gaussian distribution by the relative entropy of the histogram compared to a best-fit Gaussian distribution. This entropy is normalized by the relative entropy of a best-fit Laplacian distribution into a measure they refer to as Laplacian fractional entropy (LFE). They test the LFE on a set of 26 patients recalled at screening for which they have available full-field digital mammography (FFDM), digital breast tomosynthesis (DBT), and dedicated breast CT (bCT) images as well as breast density scores and biopsy results.

Results: A study of LFE in FFDM comparing the raw "for-processing" transmission data from the device to log-converted density estimates and the processed "for-display" data shows that processing mammographic image data enhances the non-Gaussian content of the image. A check of the methodology using a Gaussian process with a power-law power spectrum shows relatively little bias from the finite extent of the region of interests used. A second study comparing LFE across FFDM, DBT, and bCT modalities shows that each maximized the non-Gaussian content of the image for different ranges of spatial frequency. FFDM is optimal at high spatial frequencies (>0.7 mm^{-1}), DBT is optimal at mid-range frequencies (0.3–0.7 mm^{-1}), and bCT is optimal at low spatial frequency (<0.3 mm^{-1}). A third study of breast density in FFDM and bCT shows that LFE generally rises slightly going from the low-to moderate density, and then falls considerably at higher densities.

Conclusions: Non-Gaussian statistical structure in breast images that is manifest in the responses of Gabor filters similar to receptive fields of the early visual system is dependent on how the image data are processed, the modality used to acquire the image, and the density of the breast tissue being imaged. Higher LFE corresponds with expected improvements from image processing and 3D imaging. © *2012 American Association of Physicists in Medicine.* [http://dx.doi.org/10.1118/1.4761869]

I. INTRODUCTION

The x-ray projection mammogram has been the clinical standard used to screen asymptomatic women for early signs of breast cancer. The technique is generally regarded as effective, although less than optimal performance has led to continued efforts to improve diagnostic accuracy through new technology or other means. One of the principle limitations of mammography is the masking of disease by normal anatomy. This general phenomenon has been recognized in medical imaging for many years,[1–5] and is often referred to as "object variability," "patient structured noise," or "anatomical noise."

More specific to mammography, research over the last ten years has shown masking by normal anatomy to be the limiting factor for detecting masses,[6–8] particularly for women with mammographically dense breasts.[9, 10]

The important role of normal breast tissue in limiting the detection of cancer has led to efforts to characterize the statistical properties of anatomy in x-ray mammograms. These have typically been in the form of a power spectrum, in which the anatomical component assumes a power-law form[6–8] with an exponent near −2.8 and with considerable variability across women. Images of the breast depart from this model at higher frequencies where the anatomical power continues to drop and rising acquisition noise power begins to contribute substantially to the total power in each frequency. Subsequently, it has been shown that the exponent of the anatomical spectrum is dependent on the nature of the image acquired. Metheany *et al.*[11] have shown that the exponent of the anatomical spectrum is increased by one (i.e., to approximately −1.8) in 3D images of the breast acquired with a dedicated high resolution breast CT scanner. Engstrom *et al.*[12] found that in tomosynthesis reconstructions, the exponent increased by approximately 0.2 across women. Chen *et al.*[13] have shown how the effective depth of a pixel influences the power-law exponent in a way that is consistent with the findings for radiographic projection, tomosynthesis, and tomographic images of the breast. Power-law power spectra (or closely related amplitude spectra) have been used frequently in the literature on the statistical properties of natural scenes.[14–16]

However, power-spectra alone do not fully characterize anatomical variability in breast images, as demonstrated in Fig. 1. Here, a region of interest (ROI) from a digital mammogram is displayed beside a synthesized Gaussian texture generated from a power-law process matched to the exponents of mammograms. It is very easy to distinguish a ROI from a mammogram such as this from a randomly generated Gaussian texture. And thus there must be significant higher order statistical structure in the images that is not captured by the Gaussian process. This has been demonstrated in mammograms through investigations of multiresolution wavelet expansions[17] and phase randomization.[2] Differences between mammograms and Gaussian processes have also been shown in observer performance studies. Bochud *et al.*[6] have shown that mammograms do not behave as a "pure noise," meaning that the mammographic background masks masslike targets less than its power spectrum would suggest. Burgess *et al.*[7] find that a simulated "nodule" is better discriminated against mammographic backgrounds than against a Gaussian process statistically matched to the mammograms, and that the difference gets larger as the target size increases. These studies suggest that there are higher order non-Gaussian statistics in mammogram images, and that human observers exploit these somehow to improve detection performance, which in turn motivates this effort to quantify and characterize non-Gaussian statistics in breast images. Our broader hope is that characterizing non-Gaussian statistical properties of breast images will allow us to better understand how they influence diagnostic accuracy.

The approach we take is based on research in natural scene statistics, where investigators in the fields of visual neuroscience and computer vision model visual receptive fields as filters,[18–20] and use statistics derived from histograms of filter responses to motivate models of the human visual system.[21–25] We use the information theoretic concept of relative entropy, or Kullback–Leibler divergence, between a response histogram from a Gabor filter function and a best-fit Gaussian as a measure of the non-Gaussian component of the distribution. To give the relative entropy a more meaningful interpretation, we normalize it by the relative entropy arising from a Laplacian distribution, to make a measure we refer to as the Laplacian fractional entropy (LFE).

We evaluated the LFE measure on patient data from 26 women. Each woman included in the set had bilateral scans with clinical mammography and tomosynthesis devices, as well as a dedicated breast CT scanner under development at UC Davis. The mammography images consist of both the raw "for processing" data as well as the final "for presentation" images generated by the image display software distributed with the device. We use the LFE measure to evaluate the effect of processing the mammography data. Under the hypothesis that non-Gaussian structure represents the anatomy, we would expect the LFE measure to increase with image processing. We also use LFE to make comparisons across imaging modality for mammography, tomosynthesis, and bCT. This allows us to investigate how the different imaging technologies impact the non-Gaussian structure of the resulting images. Finally, we investigate the effect of breast density, defined by the BIRADS density score from the FFDM images. This shows how the LFE measure varies with the most commonly used descriptor of breast texture.

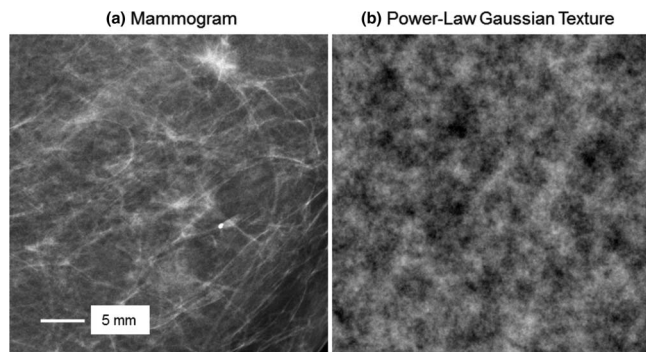(a) Mammogram   (b) Power-Law Gaussian Texture

5 mm

FIG. 1. Example of mammogram and Gaussian texture patches. A 3.6 cm ROI from a mammogram is shown next to a Gaussian texture generated using a power-law power spectrum that is matched to the average spectrum of mammograms (exponent = 2.8). While the images have similar mean intensity and power spectra, they clearly differ in terms of their texture, suggesting that higher order statistics must differ between the two.

II. THEORY

Statistical properties of images are essentially descriptive statistics that attempt to characterize the observed variability and spatial dependencies of image pixels. The approach we will use relies heavily on the assumption of ergodicity, which

means we can derive statistical properties by averaging over different locations within one image rather than independent realizations.[26] The ergodicity assumption allows us to interpret the results of a single breast image as an estimate of its statistical properties without having to specify a population of women or an ensemble of views.

II.A. Gabor filter responses

The quantity of interest in our approach is the response of a Gabor filter at locations in the breast interior. This functional form is chosen in this setting to represent receptive fields in the primary visual cortex, and hence characterizes components of statistical variability in the early stages of visual processing. The kernel of the filter is a sinusoidal plane wave attenuated by a 2D Gaussian envelope, which we will parameterize by it bandwidth in octaves, B, and center frequency in cycles/mm, f_c. Let σ_{Gab} be the standard deviation of the Gaussian envelope, which is related to the bandwidth and frequency parameters by

$$\sigma_{\text{Gab}} = \frac{\sqrt{2\ln(2)}}{2\pi f_c} \frac{2^B + 1}{2^B - 1}. \tag{1}$$

Note that this definition considers the passband of the filter to be the FWHM of the amplitude spectrum with the ratio of highest to lowest frequencies in the band to be 2^B. The resulting expression for a Gabor function that reflects these constraints is

$$g(x, y) = e^{-\frac{x^2 + y^2}{2\sigma_{\text{Gab}}^2}} \sin(2\pi f_c x), \tag{2}$$

with σ_{Gab} defined in Eq. (1). Some examples of Gabor filter kernels are seen in Fig. 2. Throughout this study, we use a fixed bandwidth of 1.4 octaves, which is representative of average bandwidths reported in studies of primate visual cortex.[18,20]

For generating filter responses, the Gabor function is sampled on a lattice matching the pixel size of the image. This

$f_c = 0.125$ mm^1
$f_c = 0.25$ mm^{-1}
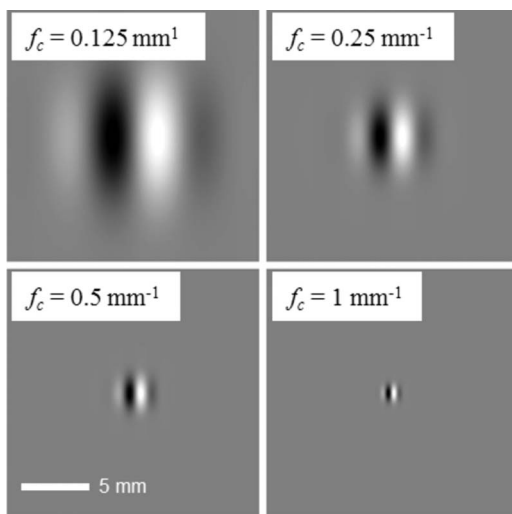$f_c = 0.5$ mm^{-1}
$f_c = 1$ mm^{-1}
5 mm

FIG. 2. Gabor functions. Each 2 cm patch shows the profile of a vertically oriented sine-phase Gabor function with 1.4-octave bandwidth.

step allows for the possibility of rotating the orientation of the Gabor function. For an N_x by N_y pixel image, let Δx and Δy be the horizontal and vertical pixels sizes, respectively. The filter kernel, for a given center frequency and orientation, is given by

$$k[n, m] = g(n\Delta x \cos(\theta) + m\Delta y \sin(\theta), n\Delta x \cos(\theta)$$
$$+ m\Delta y \sin(\theta)), \tag{3}$$

where $-N_x/2 \leq n \leq N_x/2 - 1$ and $-N_y/2 \leq m \leq N_y/2 - 1$. For an image, $I[n,m]$, a convolution operation is used to generate responses at every possible pixel location

$$r[n, m] = k^{**} I[n, m], \tag{4}$$

although only responses in a defined ROI corresponding to the breast interior (described below) are analyzed.

II.B. Modeling response histograms

The general approach to creating a response histogram is shown graphically in Fig. 3. An initial ROI, shown in Fig. 3(a), is used to identify interior regions of the breast. A secondary ROI, the white region shown in Fig. 3(b), indicates the region from which filter responses come entirely from the breast interior. A more detailed description of how these ROIs are defined for different imaging modalities is given in Sec. III.C below. Responses to an example filter are shown in Fig. 3(c), with the histogram shown in Fig. 3(d). We find that histograms of responses in many cases have long tails of extreme responses that extend out with very low frequency of occurrence. The far reaches of these tails are somewhat problematic since they become noisy, and we also find that they are often driven by calcifications in the image (note arrows in Fig. 3). Since we are more interested in the variability arising from soft-tissue contrast, we will constrain our analysis to the central 99% of the histogram where responses from soft tissue will dominate other processes such as calcifications.

The histogram is generated by sorting all filter responses from the ROI, a total of N_R responses, which we will denote r_i, $i = 1, \ldots, N_R$. To find the boundary of the central 99% of the histogram, we set the limits of the histogram to be the response values that define the central 99% of the sorted list. To determine r_{lo}, we find the index $0.5\% \times N_R$ rounded to the nearest integer and use the response value associated with this index. A similar approach is used to find r_{hi}, except that the index is $99.5\% \times N_R$ rounded to the nearest integer. For N_{Bin} histogram bins between r_{lo} and r_{hi}, we set the bin size to be

$$\Delta h = \frac{r_{\text{hi}} - r_{\text{lo}}}{N_{\text{Bin}}}. \tag{5}$$

The histogram counts in the nth bin, h_n ($n = 1, \ldots, N_{\text{Bin}}$), are the number of filter responses between $r_{\text{lo}} + (n - 1)\Delta h$ and $r_{\text{lo}} + n\Delta h$. We include one extra bin that contains all counts outside the range r_{lo}–r_{hi}, for a total of $N_{\text{Bin}} + 1$ bins in each histogram.
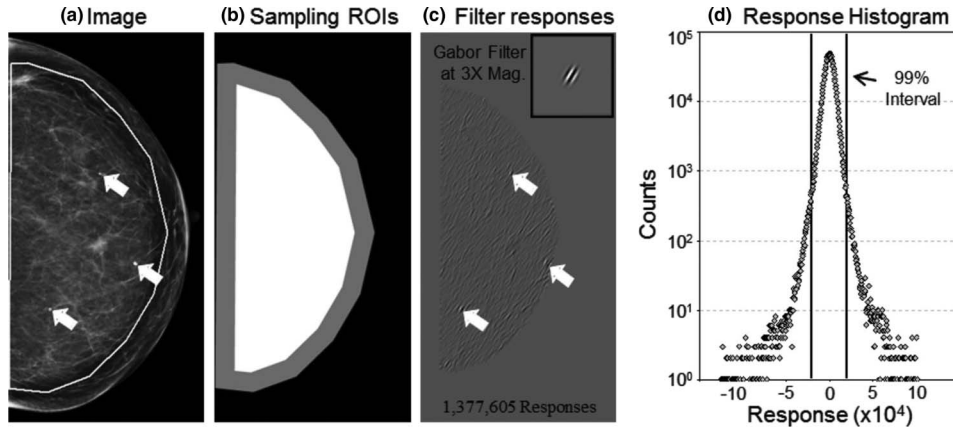
FIG. 3. Analysis of mammograms. An example mammogram is shown with manually drawn border (a), and three large calcifications (arrows). Sampling regions (b) show the ROI for the interior (gray) and the ROI for centers of the Gabor profiles (white). Filter responses from convolution (c) with the Gabor profile show extreme responses near the three calcifications (arrows). The resulting histogram is plotted (d) with the central 99% or responses between the vertical lines.

The histogram counts are converted to a probability estimate by dividing each one by the total number of responses

$$p_n = \frac{h_n}{N_R}. \tag{6}$$

It is these probability distributions that are evaluated for their non-Gaussian character. If the counts in a response histogram actually came from a Gaussian distribution, with mean value of μ and a standard deviation σ, then we would expect them to assume the form

$$p_{\text{Gauss},n}(\mu,\sigma) = \Phi\left(\frac{h_{\text{lo}} + n\Delta h - \mu}{\sigma}\right) - \Phi\left(\frac{h_{\text{lo}} + (n-1)\Delta h - \mu}{\sigma}\right), \tag{7a}$$

for $n = 1$ to N_{Bin}, and

$$p_{\text{Gauss},N_{\text{Bin}}} + 1(\mu,\sigma)$$
$$= 1 - \left(\Phi\left(\frac{h_{\text{lo}} + N_{\text{bin}}\Delta h - \mu}{\sigma}\right) - \Phi\left(\frac{h_{\text{lo}} - \mu}{\sigma}\right)\right), \tag{7b}$$

up to sampling error.

The entropy relative to the best-fit Gaussian (or Kulback–Leibler divergence between them), defined as

$$E_{\text{Rel}} = \underset{\mu,\sigma}{\text{Min}}\left(\sum_{n=1}^{N_{\text{Bin}}+1} p_n \log_2\left(\frac{p_n}{p_{\text{Gauss},n}(\mu,\sigma)}\right)\right). \tag{8}$$

Note that if the p_n are actually derived from a Gaussian distribution, then the ratio inside the log can be made 1 for some choice of μ and σ, thereby achieving a relative entropy value of 0.

The relative entropy can be interpreted as the excess bits needed to code the p_n distribution using $p_{\text{Gauss},n}$ instead of the true distribution.[27] However, it is not clear how "non-Gaussian" one bit of excess coding is. So we have adopted a different approach to normalizing E_{Rel} based on comparison to a Laplacian distribution. The two-sided Laplacian distribution is a common model of non-Gaussian scene statistics.[28,29]

Its cumulative distribution function, parameterized by location and spread parameters, a and b respectively, is given by

$$P_{\text{Lap}}(x) = \begin{cases} \frac{1}{2}e^{(x-a)/b} & x \leq a \\ 1 - \frac{1}{2}e^{-(x-a)/b} & x \geq a \end{cases}. \tag{9}$$

The binned probability distribution based on a Laplacian is given by

$$p_{\text{Lap},n}(a,b) = P_{\text{Lap}}\left(\frac{h_{\text{lo}} + n\Delta h - a}{b}\right) - P_{\text{Lap}}\left(\frac{h_{\text{lo}} + (n-1)\Delta h - a}{b}\right) \tag{10a}$$

for $n = 1$ to N_{Bin}, and for events outside the binned range

$$p_{\text{Lap},N_{\text{Bin}}+1}(a,b)$$
$$= 1 - \left(P_{\text{Lap}}\left(\frac{h_{\text{lo}} + N_{\text{bin}}\Delta h - a}{b}\right) - P_{\text{Lap}}\left(\frac{h_{\text{lo}} - a}{b}\right)\right). \tag{10b}$$

We can determine the relative entropy of the Laplacian with respect to the Gaussian by solving

$$E_{\text{Lap}} = \underset{a,b}{\text{Min}}\left(\sum_{n=1}^{N_{\text{Bin}}+1} p_{\text{Lap},n}(a,b) \log_2\left(\frac{p_{\text{Lap},n}(a,b)}{p_{\text{Gauss},n}(\mu_{\min},\sigma_{\min})}\right)\right), \tag{11}$$

where $\mu_{\min}$ and $\sigma_{\min}$ are the values minimizing Eq. (8).

We use the relative entropy of a best-fit Gaussian relative to the Laplacian as a measure of how non-Gaussian the observed histogram is

$$\text{LFE} = \frac{E_{\text{Rel}}}{E_{\text{Lap}}} \times 100\%, \tag{12}$$

which we refer to as the LFE to reflect the role of the Laplacian in normalizing the relative entropy of the observed distribution of responses.
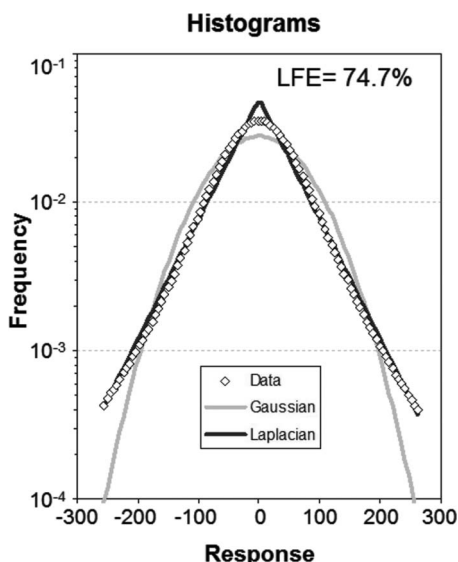
Histograms

FIG. 4. Components of the Laplacian fractional entropy measure. Data from an example histogram is shown along with fitted models based on Gaussian and Laplacian density functions. The relative entropy of the histogram to the Gaussian is scaled by the relative entropy of the Laplacian to the Gaussian to form the LFE measure of departure from a Gaussian model.

The magnitude of the LFE measure is readily interpreted: LFE = 100% means that the distribution exhibits as much non-Gaussian behavior (i.e., Kullback–Leibler divergence) as a Laplacian distribution. Figure 4 shows a plot of an observed histogram, along with the fitted Gaussian and Laplacian distributions after binning. In this case, the observed distribution function clearly diverges from the Gaussian, with a sharper peak near zero and more slowly decaying tails. The tails are well matched by the Laplacian distribution, which is even more strongly peaked with a cusp near zero. The resulting LFE value of 74.7% reflects this intermediate level of divergence between the Gaussian and Laplacian.

III. MATERIALS AND METHODS

III.A. Patient images

We have applied the Laplacian fractional entropy analysis derived in Eq. (11) to three areas in breast imaging: (1) processing of digital mammograms, (2) comparisons of breast imaging technologies, and (3) evaluating the effect of breast density. All of these studies utilized a dataset of images from 26 women with positive mammographic findings (BIRADS 4 and 5), and biopsy verification of disease status (benign or malignant) subsequent to the imaging exams as part of an ongoing clinical trial. Mammographic findings included both masses (in 17 cases) and calcifications (in 10 cases), with only 2 lesions noted as palpable. There were no cases of bilateral disease. Pathology findings showed that the cases were evenly split between malignant and benign findings (13 each).

Additional data were available for each woman, including the radiologists' BIRADS breast density scores. All data were acquired with consent under a IRB approved protocol in the Department of Radiology at the UC Davis Medical Center in

Sacramento. Each woman had full-field digital mammography (FFDM) images, digital breast tomosynthesis (DBT) images, and dedicated breast CT (bCT) images of both breasts available for this investigation. We note that the 26 cases used for this study came from a larger set of 30 patients, but 4 of these were excluded because of incomplete data.

III.B. Characteristics of scans

Mammograms and DBT scans were acquired on a Hologic Dimensions system (Hologic Inc.), the bCT images were acquired on a system developed at UCDMC.[30–33] The FFDM data included both the raw "for processing" data as well as the final "for display" images produced after processing with the display software distributed by the manufacturer. Tomosynthesis images were reconstructed over a 15° angular range using filtered backprojection software developed by the manufacturer. Each woman had images of both breasts, and each breast had craniocaudal (CC) and mediolateral oblique (MLO) views. The mammograms had 0.07 mm pixel dimensions. The DBT images averaged 0.12 mm pixel dimensions in the reconstructed planes (range: 0.09–0.23 mm) with a distance of 1 mm between tomographic planes. The bCT images were acquired at 80 kV at a dose equivalent to two-view mammography.[31–34] The 3D images were reconstructed by filtered backprojection with a Shepp-Logan filter.[35, 36] The resulting CT images consisted of 512×512 pixels in the coronal plane, with a variable number of coronal sections that depended on the size of the breast. Within each coronal plane, the pixel size averaged 0.36 mm (range: 0.29–0.38 mm), with a section thickness of 0.19 mm (range: 0.16–0.19 mm). The distance between coronal sections averaged 0.23 mm (range: 0.18–0.28 mm).

III.C. Image processing

Each of the FFDM, DBT, and bCT image modalities required some preparation of the image data for analysis. Modality specific details are given below. In all modalities, a ROI was selected from which we would accept filter responses. The purpose of this was to confine analysis to the interior of the breast, away from chest wall, muscle, nipple, and skin line where filter responses would not represent parenchymal texture. In the case of FFDM, we also created additional images to better understand the process and check the effects of a finite-sized ROI on the LFE measure.

III.C.1. Preparation of FFDM images

The acquired mammograms had both "for processing" and "for display" images saved in DICOM formats. The for processing images consisted of the raw detector outputs at 14-bit depth corrected for detector inhomogeneities, and can, therefore, be thought of as a measure of the counts transmitted to the detector at each detector element location. The for display image was generated from the "for processing" image at 12-bit depth by Hologic's proprietary display processing algorithm, which is based on unsharp masking.[37]

We generated two other images for comparison. The first was an approximate density image generated by log-converting the "for processing" transmission image (neglecting effects of scattered radiation). In principle, log-conversion makes the superposition of tissues in a mammogram approximately linear. The central limit theorem would then suggest that to the degree that tissue at different depths in the x-ray path is independent; the resulting observations should be closer to a Gaussian distribution.

As a check of our methodology, we also generated a random Gaussian texture on the same dimensions as each mammogram. The Gaussian process that created these images had a power-law power spectrum except at the DC component, which was arbitrarily set to the value at the first harmonic. The exponent of the power-law was set to -2.8, the estimated value for mammograms.[7] Since this process rigorously conforms to the default assumptions of a stationary Gaussian process, nonzero LFE values highlight any bias that comes from the limited spatial extent of averaging.

The four images we have for each scan (for processing, for display, log-converted, and Gaussian texture) all have the same spatial layout, and hence the same ROI is used for all four in a given scan. These ROIs were created in a two-step process. An initial ROI was created inside of a set of manually selected connected points that were inside of the chest wall, approximately 1 cm from the skin line, and inside of the nipple. An example of this region can be seen in the outline of Fig. 3(a). Image values within this region are considered acceptable for contributing to filter responses. In order to ensure that the Gabor filters do not extend outside the ROI, responses are only taken from the final ROI an additional 1 cm inside this initial ROI. Initial and final ROIs can be seen in Fig. 3(b). ROIs were generated for all images (right, left, MLO, and CC views). Gabor responses are generated for each image by convolving a Gabor filter with the image, and then using responses inside the final ROI to build a histogram for LFE analysis. FFDM histograms consisted of 2.1×10^6 responses on average (range: 2.5×10^5–5.6×10^6).

III.C.2. *Preparation of DBT images*

For the DBT images, ROIs were constructed similar to those in FFDM, with a manually selected region designed to avoid chest wall, nipple and skin lines, and an additional 1 cm interior to avoid contamination of filter responses from points outside this area. The DBT images consisted of multiple planar images through the breast volume, with 66 such images on average (range = 34–99). We manually selected an interior range of planar images so that planes from the skin were not considered. The same spatial boundary was used on this range of interior images. For each Gabor filter function, responses were generated by convolving each interior image with the Gabor kernel and then assembling responses from all interior images into a single histogram. Response histograms for DBT images consisted of 1.8×10^7 filter responses on average (range: 8.7×10^5–1.2×10^8).

III.C.3. *Preparation of dedicated bCT images*

The bCT images were analyzed as coronal sections as this has been radiologists' preferred orientation for viewing them. In addition to the images themselves, where voxels in the image represented the density of the tissue in Hounsfield units, we also had segmented images categorized into air, skin, adipose, and glandular tissue regions. A manual procedure was used to select sections between the chest wall and nipple, and with the initial ROI defined using the voxels labeled as adipose or glandular tissue. The final ROI was defined as all voxels in this slice range that were 1 cm in distance to the edge of the initial ROI. Response histograms for bCT images consisted of 1.5×10^7 filter responses on average (range: 2.3×10^6–3.0×10^7).

IV. RESULTS AND DISCUSSION

The LFE was calculated at each of the 26 patients in each modality, view, filter center frequency, and orientation. For the FFDM images, which had four modalities ("for processing," log-converted, "for display," and Gaussian texture), 4 views (LCC, LMLO, RCC, and RMLO), 11 spatial frequencies (from 0.125–4.0 cyc/mm), and 6 orientations, a total of 27 456 histograms were generated and then fit. A total of 4992 histograms were fit for the DBT modality since they had only one modality (the DBT reconstruction) and only 8 spatial frequencies were used (ranging from 0.125–1.4 cyc/mm) due to the larger pixel sizes of the images. A total of 1872 histograms were fit for the bCT modality, since only 2 images were available for each patient (right and left), and only 6 spatial frequencies were used (ranging from 0.125–0.707 cyc/mm).

We discuss each of the three studies below in terms of their consistency with the idea that LFE is capturing the structural content of the data for a given spatial frequency. This suggests that images with greater LFE are generally better diagnostic images. However, it is clear to us that LFE is not a general surrogate for image quality, and we discuss the limitations of the measure below in Sec. IV.D.

IV.A. Processing of mammograms

The main results of the FFDM study are shown in Fig. 5. Examples of an image patch for each condition are shown in the panel. Note that the "for processing" (i.e., transmission) image is inverted relative to the log-converted and "for display" images. The inversion itself should not affect LFE, since it is unchanged if the x axis of the histogram is reversed. An example of the Gaussian texture used to check the methodology is also shown in this panel [Fig. 5(a) bottom right].

For each patient, we average LFE across views and orientation, and these are then averaged across patients to make plots with standard errors as a function of spatial frequency. Two-way ANOVA with spatial frequency and image processing stage as factors finds a significant effect of the image processing ($p < 0.0001$). All of the mammographic images have their highest values of LFE (70%–90%) at the lowest frequency evaluated (0.125 cyc/mm), and then decay as
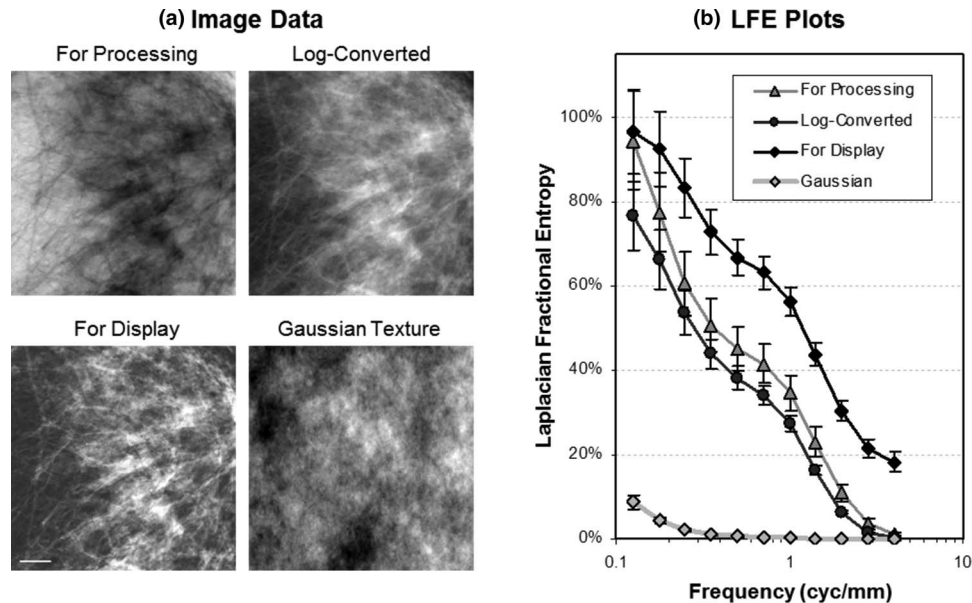
FIG. 5. Effect of processing in mammograms. The image panel (a) shows three different levels of data processing in the generation of a "for display" mammogram as well as a power-law Gaussian texture included as a check. The plots (b) show the Laplacian fractional entropy averaged over all views of both breasts, and over the six different orientations evaluated. Error bars represent one standard error across the 26 patient datasets used in the study.

frequency increases. Log-conversion of the "for processing" image reduces LFE across the spectrum before both decay nearly to 0 by 4.0 cyc/mm. This is consistent with the idea that log-conversion makes the tissue superposition approximately additive, and by the central limit theorem results in response distributions that are closer to a Gaussian. By contrast, the display processing software distributed with the device increases LFE across the range of spatial frequencies. The images generated from a Gaussian process give LFE values near 0, except at the lowest frequencies where they remain below 10%. This suggests that the limited size of the ROIs are not causing a substantial bias in LFE.

The FFDM data we show are consistent with the idea that the goal of mammographic image processing is to enhance the structural content of the images. Log-conversion, while implementing the necessary inversion of intensity, results in an approximately linear superposition of tissues, and thereby reduces the non-Gaussian structure of the data because of the central limit theorem. By contrast, a more involved display algorithm that selectively and adaptively modulates spatial frequencies results in an image with a higher non-Gaussian structural content.

IV.B. Breast imaging technologies

Results of LFE across different breast imaging modalities are shown in Fig. 6. The panel of images shows example patches from the same region of the breast for FFDM and DBT, and approximately the same region for bCT. The plots show the FFDM "for-display" results described above along with LFE assessed in the DBT and bCT images from the same set of patients. The different modalities are plotted out to different frequencies reflecting the larger pixel sizes in the DBT and bCT.

Two-way ANOVA with spatial frequency (over the common range of frequencies from 0.13 to 0.71 cyc/mm) and imaging modality as factors finds a significant interaction between the two ($p < 0.0001$), suggesting that the three imaging modalities have significantly different frequency profiles. Average LFE in the DBT images peaks slightly above the FFDM images in midrange frequencies from 0.25–0.5 cyc/mm, but decays more rapidly to 0 above 0.7 cyc/mm. Average LFE in the bCT images is about a factor of 2 larger than either FFDM or BCT at the lowest spatial frequencies (<0.2 cyc/mm), and it also decays rapidly to 0 as frequencies increase.

These data suggest that the different breast imaging technologies emphasize structure at different spatial-frequency ranges. FFDM images, which have the highest resolution, maximize LFE at the highest spatial frequencies. At lower spatial frequencies, super-position of tissues may be the dominant effect, in which case techniques to reduce or remove superposition should improve the representation of structure. Thus the limited-angle tomography employed in DBT is able to make a limited increase in LFE at midrange spatial frequencies, and the full-range tomography employed in bCT makes a substantial increase in LFE at low frequencies.

IV.C. Effect of breast density

Figure 7 shows the effects of breast density on FFDM and bCT images. We use the radiologist's visual assessment of breast density as recorded in their BIRADS density scores from the FFDM images to define the breast density on a patient-by-patient basis. The four-point scale ranges from lowest to highest density. Of the 26 patients with complete data, 5 had a density score of 1 (adipose), 4 had a density score of 2 (scattered density), 11 had a score of 3 (heterogeneous density), and 7 had a score of 4 (dense).
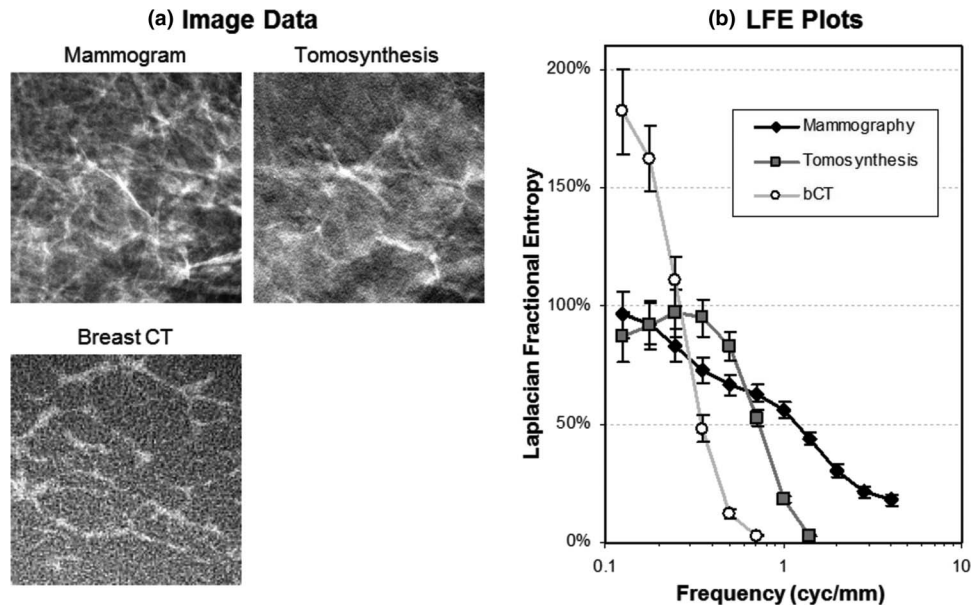
FIG. 6. Effect of different imaging modalities. The image panel (a) shows 5 cm example regions from a ("for display") mammogram, tomosynthesis image, and coronal breast CT section. The images are all from the same patient in roughly the same area of the breast. Laplacian fractional entropy (b), averaged over different views and the six filter orientations, is plotted as a function of spatial frequency for each modality. Error bars represent one standard error across the 26 patient datasets used in the study.

Breast density affects LFE in all three imaging modalities, as seen in Fig. 7. Two-way ANOVA with spatial frequency (over the entire range for each modality) and breast density as factors finds a significant main effect of density (FFDM: $p < 0.001$, DBT: $p < 0.0001$, bCT: $p < 0.03$). The spatial frequency profiles are relatively similar across breast density within each modality, but the magnitude changes with the density. LFE increases in FFDM and bCT images going from BIRADS density scores of 1–2. DBT images have approximately equal LFE between the two. LFE generally decreases as density scores increase above 2, with the exception of high spatial frequencies (>1 cyc/mm) in the FFDM images where the order is reversed.

The general finding of decreased LFE with increased density (above a BIRADS density score of 2) is still consistent with the idea of LFE as a measure of structural content, even though our intuition might suggest that greater density implies more structure in the breast. In images with scattered density, the filter responses lead to more pronounced tails, which heighten the departure from normality.

IV.D. Limitations of the LFE measure

We have defined the LFE measure as a way to quantify non-Gaussian statistical properties of images. Because the measure operates on Gabor filter outputs which are similar to receptive fields in the early visual cortex, we believe that the measure may be particularly relevant to visually perceived structure in images. However, the approach makes a number of assumptions that need to be considered when interpreting the measure, and particularly when interpreting it as a
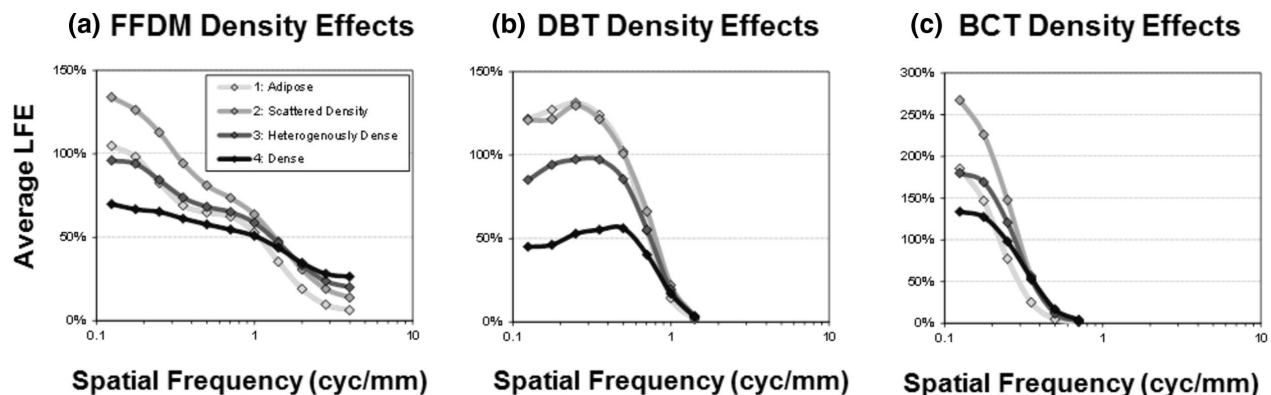

FIG. 7. Effects of breast density. LFE averaged across BIRADS density categories for FFDM images (a) is plotted as a function of spatial frequency. Note that LFE increases going from adipose to scattered density, the two lowest categories, and then decreases as density increases (legend applies to all three plots). Plots for DBT (b) and BCT (c) show similar effects.

measure of structural content of an image. We will review these briefly before concluding.

At the most fundamental level, we are assuming that the appearance of the image in the interior regions of the breast is a realization of an ergodic stationary process, which allows us to approximate the response probability distribution from a histogram. A nonstationary process can give rise to highly non-Gaussian response histograms, even if the underlying image were Gaussian. For example, an image with the left side being Gaussian white noise with a standard deviation of 1 and the right side being Gaussian white noise with a standard deviation of 100 will produce LFE values that are nonzero even though the underlying image is Gaussian. Similar effects can be induced by adaptive processing of an image. This may explain some of the higher LFE values seen in the for-display FFDM images at high spatial frequencies in Fig. 5(b).

In the results above, we find that higher LFE corresponds remarkably well with a more interpretable image diagnostic image. However, the results from the very lowest density images serve as a caution against making too much of these findings as well. It is easy to imagine dual energy or other subtraction techniques that remove only the masking structural component of the image, thereby increasing diagnostic performance in many applications, but likely reducing LFE because the structural background has been subtracted. These issues suggest caution in interpreting LFE as a surrogate for diagnostic performance.

V. CONCLUSIONS

The study described here demonstrates and characterizes the statistical properties of breast images that go beyond what can be modeled by a multivariate Gaussian distribution parameterized by a mean and covariance matrix. Of the unlimited number of possible statistics to consider in this regard, we have chosen to analyze responses to Gabor filter functions because of their relation to receptive fields in the early visual system. The structure of parenchymal breast tissue results in approximately Laplacian tails on the response histograms out to extreme values influenced by calcifications. After excluding extreme points, we use the entropy relative to a Gaussian as a general measure of non-Gaussian statistical properties. In order to give the measure a meaningful scale, we make it relative to the magnitude of a fitted Laplacian distribution. Thus, a value of 100% means the histogram has the same entropy relative to a Gaussian as the histogram arising from a fitted Laplacian, and hence we refer to this quantity as the Laplacian fractional entropy. The LFE can be thought of as representing the non-Gaussian structural content of an image (or ROI) at the frequency, bandwidth, and orientation of the Gabor filter, using the Laplacian distribution as a yardstick.

We have used LFE to investigate x-ray images of the breast using a set of images from 26 patients. In all cases, the LFE of breast images decays at higher spatial frequencies as the effects of resolution and noise limit the structural content of the images. We find that processing of raw projection mammography data to a final "for-presentation" image increases LFE across the tested range of spatial frequencies, and this

cannot be simply attributed to a nonlinear conversion to integrated density. Different modalities for breast imaging emphasize LFE at different spatial frequencies. Coronal breast CT slices have the highest LFE at spatial frequencies below 0.3 cyc/mm. Tomosynthesis images achieve the highest LFE in a middle range for spatial frequencies from 0.3 to 0.6 cyc/mm. Processed projection mammograms have the highest LFE above 0.6 cyc/mm. When breast density is considered, we find that higher density—assessed either by BIRADS density scores from the mammograms or by segmented volumes in breast CT—reduce LFE. While this may seem counterintuitive given that we think of a dense breast as having more glandular structure, it is consistent with our argument that dense breasts lack sparsity of structure making them more difficult to diagnose accurately.

This last point warrants some caution in interpretation. The measure does not make any distinction between normal and abnormal structures, and it can potentially be misled by adaptive processing and other procedures with complex statistical properties. Nonetheless, we believe that further investigation into the structure of breast images may overcome these limitations, and characterizing the structural content of breast images may provide insights into more effective breast imaging technologies.

a)Author to whom correspondence should be addressed. Electronic mail: abbey@psych.ucsb.edu

[1] H. H. Barrett, "Objective assessment of image quality: Effects of quantum noise and object variability," J. Opt. Soc. Am. A **7**, 1266–1278 (1990).

[2] F. O. Bochud, F. R. Verdun, C. Hessler, and J. F. Valley, "Detectability of radiological images: The influence of anatomical noise," Proc. SPIE **2436**, 156–164 (1995).

[3] M. P. Eckstein, C. K. Abbey, and F. O. Bochud, "Visual signal detection in structured backgrounds. IV. Figures of merit for model performance in multiple-alternative forced-choice detection tasks with correlated responses," J. Opt. Soc. Am. A Opt. Image Sci. Vis. **17**, 206–217 (2000).

[4] H. L. Kundel and G. Revesz, "Lesion conspicuity, structured noise, and film reader error," Am. J. Roentgenol. **126**, 1233–1238 (1976).

[5] G. Revesz, H. L. Kundel, and M. A. Graber, "The influence of structured noise on the detection of radiologic abnormalities," Invest. Radiol. **9**, 479–486 (1974).

[6] F. O. Bochud, J. F. Valley, F. R. Verdun, C. Hessler, and P. Schnyder, "Estimation of the noisy component of anatomical backgrounds," Med. Phys. **26**, 1365–1370 (1999).

[7] 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).

[8] J. J. Heine, S. R. Deans, R. P. Velthuizen, and L. P. Clarke, "On the statistical nature of mammograms," Med. Phys. **26**, 2254–2265 (1999).

[9] J. M. Boone and K. K. Lindfors, "The effect of breast density on cancer detection performance in mammography," J. Womens Imaging **3**, 122–128 (2001).

[10] N. F. Boyd, H. Guo, L. J. Martin, L. Sun, J. Stone, E. Fishell, R. A. Jong, G. Hislop, A. Chiarelli, and S. Minkin, "Mammographic density and the risk and detection of breast cancer," N. Engl. J. Med. **356**, 227–236 (2007).

[11] K. G. Metheany, C. K. Abbey, N. Packard, and J. M. Boone, "Characterizing anatomical variability in breast CT images," Med. Phys. **35**, 4685–4694 (2008).

[12] E. Engstrom, I. Reiser, and R. Nishikawa, "Comparison of power spectra for tomosynthesis projections and reconstructed images," Med. Phys. **36**, 1753–1758 (2009).

[13] L. Chen, C. K. Abbey, A. Nosratieh, K. K. Lindfors, and J. M. Boone, "Anatomical complexity in breast parenchyma and its implications for optimal breast imaging strategies," Med. Phys. **39**, 1435–1441 (2012).

[14] A. P. Pentland, "Fractal-based description of natural scenes," IEEE Trans. Pattern Anal. Mach. Intell. **6**, 661–674 (1984).

[15] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," Phys. Rev. Lett. **73**, 814–817 (1994).

[16] A. van der Schaaf and J. H. van Hateren, "Modelling the power spectra of natural images: Statistics and information," Vision Res. **36**, 2759–2770 (1996).

[17] J. J. Heine, S. R. Deans, D. K. Cullers, R. Stauduhar, and L. P. Clarke, "Multiresolution statistical analysis of high-resolution digital mammograms," IEEE Trans. Med. Imaging **16**, 503–515 (1997).

[18] R. L. De Valois, D. G. Albrecht, and L. G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex," Vision Res. **22**, 545–559 (1982).

[19] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," J. Opt. Soc. Am. A **4**, 2379–2394 (1987).

[20] D. L. Ringach, M. J. Hawken, and R. Shapley, "Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences," J Vision **2**, 12–24 (2002).

[21] J. J. Atick and A. N. Redlich, "What does the retina know about natural scenes?," Neural Comput. **4**, 196–210 (1992).

[22] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," J. Opt. Soc. Am. A **2**, 1160–1169 (1985).

[23] B. A. Olshausen, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," Nature (London) **381**, 607–609 (1996).

[24] D. L. Ruderman, "The statistics of natural images," Network Comput. Neural Syst. **5**, 517–548 (1994).

[25] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," Annu. Rev. Neurosci. **24**, 1193–1216 (2001).

[26] H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley, New York, 2003).

[27] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," IEEE Trans. Inf. Theory **44**, 2743–2760 (1998).

[28] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," IEEE Trans. Image Process. **14**, 2117–2128 (2005).

[29] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," Appl. Comput. Harmon. Anal. **11**, 89–123 (2001).

[30] J. M. Boone and K. K. Lindfors, "Breast CT: Potential for breast cancer screening and diagnosis," Future Oncol. **2**, 351–356 (2006).

[31] J. M. Boone, T. R. Nelson, K. K. Lindfors, and J. A. Seibert, "Dedicated breast CT: Radiation dose and image quality evaluation1," Radiology **221**, 657–667 (2001).

[32] K. K. Lindfors, J. M. Boone, T. R. Nelson, K. Yang, A. L. Kwan, and D. F. Miller, "Dedicated breast CT: initial clinical experience," Radiology **246**, 725–733 (2008).

[33] W. T. Yang, S. Carkaci, L. Chen, C. J. Lai, A. Sahin, G. J. Whitman, and C. C. Shaw, "Dedicated cone-beam breast CT: Feasibility study with surgical mastectomy specimens," AJR, Am. J. Roentgenol. **189**, 1312–1315 (2007).

[34] I. Sechopoulos, S. Vedantham, S. Suryanarayanan, C. J. D'Orsi, and A. Karellas, "Monte Carlo and phantom study of the radiation dose to the body from dedicated CT of the breast," Radiology **247**, 98–105 (2008).

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

[36] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," IEEE Trans. Nucl. Sci. **21**, 21–43 (1974).

[37] E. D. Pisano, E. B. Cole, B. M. Hemminger, M. J. Yaffe, S. R. Aylward, A. D. A. Maidment, R. E. Johnston, M. B. Williams, L. T. Niklason, and E. F. Conant, "Image processing algorithms for digital mammography: A pictorial essay1," Radiographics **20**, 1479–1491 (2000).