UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

SUBJECTIVE EVALUATION OF THE IN-LINE PHASE-SENSITIVE IMAGING SYSTEMS IN BREAST CANCER SCREENING AND DIAGNOSIS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

FARID H OMOUMI

Norman, Oklahoma

2022

SUBJECTIVE EVALUATION OF THE IN-LINE PHASE-SENSITIVE IMAGING SYSTEMS IN BREAST CANCER SCREENING AND DIAGNOSIS

A DISSERTATION APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY THE COMMITTEE CONSISTS OF

Dr. Hong Liu, Chair

Dr. Zahed Siddique

Dr. Bin Zheng

Dr. Joseph P. Havlicek

Dr. Yuchen Qiu

This dissertation is dedicated to all individuals who helped me be myself, stood next to me, and made the world around me much more meaningful. I want to extend my special gratitude to my late parents, ***Dr. Hamid*** *and* ***Shahrzad****, whose words and attitude of encouragement will always remain with me lifelong! I would also like to express my special thanks to my brothers and their families, my uncle and his family, my parents-in-law, my brother-in-law, and my sister-in-law, who have never left my side and are very special to me.*

I finally, and most importantly, express my sincere love and appreciation to my wife, ***Dr. Mehri****, and my daughter,* ***Cadence****, for their unwavering support and patience when they needed me the most but endured my absences and assisted me in focusing on the completion of this degree, and cheered with me at every single step of this path.*

Acknowledgments

Prof. Dr. Hong Liu, words cannot express my boundless gratitude for you! I will be appreciated and beholden for what you have done and taught me in my academic research journey. Your advice and guidance were always essential and priceless, not only for my academic advancement but also for my personal life, without any exaggeration. I was extremely fortunate that I had this opportunity to be your student, learn from you, and choose you as my role model in my future professional career and personal life.

I would like to acknowledge Dr. Xizeng WU, Ph.D. from the University of Alabama, and Dr. Laurie Fajardo, M.D. from the University of Utah HSC, for their tremendous collaboration and encouragement. In addition, my special thanks go to my current and former research team members: Dr. Yuhua Li, Dr. Muhammad Ghani, Dr. Molly Wong, and Dr. Di Wu, for their unconditional help and support.

I would also like to thank Dr. Joseph Havlicek and Dr. Jennifer Holter, M.D., who guided, supported, and encouraged me to enter this program. Last but not least, I would like to express my special appreciation to the members of my Doctoral Committee; Dr. Zahed Siddique, Dr. Bin Zheng, Dr. Joseph Havlicek, and Dr. Yuchen Qiu, for their support and advice during my Ph.D. education at the University of Oklahoma.

Table of Contents

List of Tables

List of Figures

Abstract

Breast x-ray imaging remains the gold standard screening tool despite the various imaging modalities. The phase-sensitive breast imaging is an evolving technology that may provide higher diagnostic accuracy and potentially reduce the patient radiation dose. Many studies evaluate the performance of the In-line phase-sensitive breast imaging to improve this imaging modality further. Whereas radiologists are the end-users of this imaging technology, the primary goal of this dissertation project is to investigate the performance of human observers in varying conditions for further improvement of the in-line phase-sensitive x-ray imaging system.

A CDMAM phantom and an ACR mammography phantom are used in the observer performance study to compare the high-energy in-line phase-sensitive system with a mid-energy system as an alternative approach to balancing the attenuation-based image contrast with the accuracy of single-projection PAD-base phase-retrieval. Additionally, a series of ROC studies are designed by a contrast-detail phantom to evaluate the diagnostic accuracy of digital breast tomosynthesis (DBT) and the phase-sensitive prototype imaging system (PBT). The area under the ROC curves (AUC) and partial area under the ROC curves (pAUC) are estimated as a figure of merits in the two systems, delivering the equivalent radiation doses. A two-alternative-forced choice (2AFC) study is also designed to determine the preferred image in identifying the suspicious lesions within a heterogeneous pattern acquired by the DBT and PBT systems under an equivalent radiation dose.

The observer performance studies show that the mid-energy system has a potential advantage in providing a relatively higher image quality while the radiation dose is

reduced in the mid-energy system compared with a high-energy system. The ROC study shows that the diagnostic accuracy of observers is more significant in the prototype PBT system than in a commercial DBT system, delivering the same radiation dose. The 2AFC study also revealed that observers prefer the PBT system in detecting and distinguishing the conspicuity of tumors in the images with structural noise, and the results were statistically significant.

The dissertation also introduces a mathematical approach for estimating the half-value-layer (HVL) from measured or simulated x-ray spectra. The HVL measurement is expected to be less accurate or experimentally challenging in some clinical equipment or when a quick beam quality evaluation is needed. Additionally, the impact of varying thicknesses of external filtration is subjectively and objectively investigated to evaluate the feasibility of reducing the image acquisition time in a mid-energy system without compromising the observer's performance and detectability. The preliminary results from phase-contrast images suggest that an in-line phase-sensitive system operated at 59 kV shows a comparable image quality with the x-ray beams filtered by 1.3 mm and 2.5 mm-thick aluminum filters. This finding could help shorten the exposure time by 34% in the mid-energy system, where image blurring is a concern due to patient movement in a longer image acquisition time.

In summary, and as expected, the subjective analyses of the in-line phase-sensitive imaging system align with the previous findings. However, the PBT imaging system may benefit from further improvement in image processing algorithms and optimizing the system with the most appropriate x-ray beam quality, considering the acquisition time, breast glandular composition, breast thickness, and different x-ray energies.

Chapter 1. Introduction

1.1 Significance

The performance of any diagnostic or screening test in accurately identifying the diseased cases and ruling out the undiseased cases is always a crucial factor. The in-line phase-sensitive x-ray imaging, which could be translated into clinical applications, is a promising technique with multiple advantages. This dissertation research will focus on the subjective evaluation of in-line phase-sensitive x-ray imaging for breast cancer screening and diagnosis and evaluate the overall performance of the human observers in accurately identifying the suspicious structures in various conditions.

Breast cancer is still the third most common cancer worldwide and the leading cancer-related cause of disease burden for women.[1] American Cancer Society estimated approximately 284,000 new breast cancer cases and over 44,000 deaths related to breast cancer in 2021.[2] Early detection is generally considered a key element in patients' prognosis for all malignancies, including breast cancer. The combination of early detection and advanced treatments resulted in a significant increase in survival rate and minimized morbidity related to breast cancer.[3] From 1989 to 2012, breast cancer death rates decreased by 36%, which means approximately 250,000 fewer breast cancer mortality in the United States alone.[4]

There is no decisive curative protocol for breast cancer treatment thus far, and prevention through screening and early diagnosis offers the most practical and effective intervention to reduce breast cancer mortality or cancer-related morbidity. X-ray imaging is the standard imaging modality used for breast cancer screening in age-eligible and asymptomatic women, and many studies have proven that screening mammography can

decrease the cancer mortality rate.[5-9] However, in addition to the routine screening with Full-Field Digital Mammography (FFDM), high-risk women may benefit from adjunctive imaging modalities such as breast ultrasonography, MRI, Digital Breast Tomosynthesis (DBT), breast thermography, optical breast imaging, electrical impedance tomography (EIT), and the breast CT scan[10] or utilizing the diagnostic chemical biomarkers for early detection.[11]

Although the FFDM is currently the standard breast screening modality, it is less effective in women younger than 40 or with dense breasts. The mammographic screening does not offer any suggestive outcome for the detected disease.[12-13] Therefore, supplemental breast ultrasonography, breast MRI, and Positron Emission Tomography (PET scan) have been applied as additional imaging modalities to overcome the specified challenges. Breast sonography may provide relatively higher sensitivity in women with large dense breasts. However, it may result in an additional follow-up, which sometimes can be invasive and causes unwanted emotional stress in patients due to its low specificity.[14-15] Breast magnetic resonance imaging can identify the small lesions that cannot be detected by mammography; However, it results in overdiagnosis due to its low specificity in breast cancer screening. It is also expensive to be used as a first-line screening modality.[16-17] On the other hand, the PET scan is the most accurate imaging modality to investigate the spread of malignant cells in the body and the response to cancer treatments but has never been used for screening purposes.[18]

Despite the advantages and disadvantages of different imaging techniques, mammography remains the gold standard imaging modality for breast cancer screening. However, the superimposition of a three-dimensional (3D) structure in a 2-D image

would inevitably impact the sensitivity or specificity of screening mammography. Digital Breast Tomosynthesis (DBT) is an advanced form of digital mammography (DM) allowing 3D volumetric reconstruction of the breast from multiple low-dose 2D images. Although breast tomosynthesis began attracting the interest of many researchers as early as the 1990s, the Food and Drug Administration (FDA) approved Hologic's Selenia Dimensions DBT system (Hologic, Inc., Bedford, MA) in 2011. DBT has shown that it decreases the masking effect of 2D superimposition and increases lesion conspicuity compared to FFDM.[19-21] It is not unexpected that DBT will become the new gold standard in breast imaging, especially for women with dense breasts in the future.

Nevertheless, the relatively close x-ray attenuation coefficients for the non-cancerous glandular structure and the malignant lesions cause a significant challenge in cancer detection utilizing the FFDM and DBT imaging systems.[22] A few techniques, such as utilizing the low-energy x-ray beam or an anti-scatter grid, have been employed in breast x-ray imaging to increase the signal-to-noise ratio (SNR) and, consequently, the image quality. However, both resolutions ultimately contribute to a high radiation dose to one of the most sensitive organs in the body for ionizing radiation. Although the transition from film/screen to digital mammography practically reduced the patients' radiation dose between 15% and 40%, depending on the breast thickness,[23] due to the wide dynamic range of digital detectors and hence eliminating the need for repeated exposures, there might still be foreseeable resolutions for further dose reduction without compromising the image quality.

Recently, a newly evolved x-ray imaging technique has brought double layers of x-ray properties into play for breast imaging. The x-ray radiation is part of the

electromagnetic spectrum that attenuates by passing through the material while simultaneously experiencing the phase shift.[24-26] The constructive and destructive interferences of two waves emerging from the tissue will result in phase-induced contrast, as the x-ray absorption still produces the attenuation-based image contrast in this technique. The phase shift coefficients of soft tissue, including the various structures within the breast, are at least 2-3 orders of magnitude larger than their attenuation coefficients within the diagnostic x-ray energy range.[25, 27-28] Therefore, the phase-induced contrast may differentiate the various structures better than the attenuation-based image contrast. In phase-sensitive x-ray imaging, the phase-induced contrast can be added to the attenuation-based contrast and provide the radiologists with an image that offers superior detectability or better margin visibility.

Multiple phase-sensitive x-ray imaging techniques employ phase shift information. The x-ray interferometry and diffraction-enhanced imaging require the monochromatic x-ray beam and several special optical devices with complicated system configurations.[29-31] Contrarily, the in-line phase-sensitive x-ray imaging system utilizes the same system configuration as conventional radiography, albeit with a few modifications.[32-34] In this technique, a micro-focus x-ray tube provides a partially coherent x-ray beam, and an air gap between the object and the detector results in phase gradient production as x-ray beams propagate. The combined effects of phase-shift information and the attenuation-based contrast improve the image quality while the radiation dose could potentially be reduced.

Many studies have investigated the overall image quality of phase-sensitive imaging systems and the feasibility of dose reduction. Most of these studies mainly focus on the

objective evaluation of the imaging system, and the preliminary subjective evaluations have been done in preclinical settings by phantom studies.[35-44] Breast x-ray imaging requires a low-energy x-ray beam to preserve the attenuation-based image contrast. However, the in-line phase-sensitive imaging system usually operates at a higher range of diagnostic x-ray energy than conventional mammography. Due to the introduced distance between the object and the detector, we lose a large amount of signal-carrying photons, resulting in demands for a higher x-ray exposure to compensate for the signal loss. Additionally, the low-energy x-ray beam with extended exposure time to compensate for the photon loss on the detector plane is not a dose-efficient approach in clinical settings.

Moreover, to retrieve the phase information, either two projection images at the various object to detector distances are required -which again is not a dose-efficient approach- or a single-projection phase retrieval method must be utilized. There are four single-projection phase retrieval methods that have been introduced with adequate phase mapping.[45-50] The phase retrieval method using the principle of Phase-Attenuation Duality (PAD) is a practical and well-established single-projection phase retrieval method that extensively works when the effective atomic number of the object is low, and the x-ray energy is relatively higher than the utilized x-ray energy in traditional breast imaging.

Many objective and subjective investigations have been done on high-energy in-line phase-sensitive imaging systems. However, the attenuation-based image contrast is considerably affected by high-energy x-ray beams in soft tissue (e.g., breast imaging). However, with subsequent improvement in the PAD-base phase retrieval algorithm, it is believed that a minor error might be introduced in phase mapping, especially for very

light elements where the utilized x-ray energy is not too low. Recently, a preclinical phantom study has been published to investigate the feasibility of utilizing a mid-energy x-ray beam instead of a high-energy beam for in-line phase-sensitive breast imaging.[51]

To the best of my knowledge, there is limited research available for the subjective evaluation of the phase-sensitive imaging systems to evaluate the diagnostic accuracy of human observers -if it is not any- regardless of the utilized x-ray energy. Since the physicians are the end-users in radiology, the diagnostic capacity and the performance of interpreting physicians in identifying the pathology and confidently ruling out the suspicious features as a diseased case are the most critical aspects of a newly developed imaging system. Thus, this dissertation research investigates the detailed subjective performance of the in-line phase-sensitive imaging systems, evaluates the impact of multiple imaging acquisition parameters, and differentiates the capability of phase-sensitive breast tomosynthesis from the conventional digital breast tomosynthesis if the imaging system translates into clinical applications.

1.2 Objectives

This dissertation aims to extend the comprehensive investigation of the diagnostic performance of the in-line phase-sensitive breast tomosynthesis (PBT) and compare it with digital breast tomosynthesis (DBT). Previous investigations have explored the potential of the in-line phase-sensitive imaging systems mainly through objective evaluations in pre-clinical stages. It has been shown that the phase-sensitive imaging system improves the detectability and diagnostic sensitivity while the radiation dose can be reduced in phantom studies. The subjective performance of the system will be investigated in detail, and the diagnostic accuracy of a prototype PBT system under

different image acquisition parameters will be presented in this dissertation research for further improvement of the prototype system toward commercialization.

1.3 Organization of Dissertation

The organization of this dissertation is as follows. Chapter 2 describes the research background, including the principles of phase-sensitive medical imaging and the fundamental theory of the human observer performance. Chapter 3 compares the performance of high-energy and mid-energy phase-sensitive imaging systems, utilizing a dedicated imaging phantom in a pre-clinical setting. Chapter 4 investigates the performances of a prototype phase-sensitive breast tomosynthesis and the commercial digital breast tomosynthesis systems, employing the Receiver Operating Characteristic (ROC) paradigm. Chapter 5 investigates the competency of a prototype phase-sensitive breast tomosynthesis and the commercial digital breast tomosynthesis systems, employing the two-alternative forced-choice (2AFC) study. Chapter 6 defines a simplified mathematical approach to evaluate the x-ray beam quality and HVL estimation from the x-ray spectrum, where the speedy HVL estimation is required without tedious measuring experiments. Chapter 7 investigates the image quality impacted by varying x-ray beam qualities to evaluate the feasibility of reducing the image acquisition time without compromising the image quality. Ultimately, Chapter 8 provides the conclusion and the possible angles for future studies for further system optimization.

Chapter 2. Research Background

2.1 Breast Cancer Screening and Diagnosis

According to numerous observational cohort and case-control studies from 1989 to 2014, assessed by International Agency for Research on Cancer (IARC) Working Group, it is shown that women 50-69 years of age, who had regular mammographic screening, experienced approximately 40% reduction in mortality rate related to the breast cancer.[52] The American College of Radiology (ACR) recommends annual breast cancer screening for women starting at age 40 and continuing if life expectancy is equal to or greater than ten years.[53] On the other hand, the U.S. Preventative Services Task Force suggests biannual screening in women between 50-74 years of age.[54] Regardless of recommendations for the target group, mammography remains the gold standard for breast cancer screening with relatively higher sensitivity and specificity.[55] However, different imaging modalities have been proposed due to limitations of mammography, such as low sensitivity in dense breasts or radiation risks.

Breast Ultrasonography is an alternative primary imaging modality to diagnose and monitor the response to therapy in breast cancer.[56] MRI is another diagnostic imaging modality that monitors the high-risk population, detects the premalignant lesions, identifies the residual tumor in operated patients, assesses distant metastasis and cancer staging, and monitors the response to therapy.[57] PET and SPECT scans are other imaging modalities that could be used for breast cancer diagnosis or to evaluate the response to the treatment.[58-59] The magnification mammography is usually used to improve the diagnostic quality by increasing the contrast-to-Noise ratio (CNR), improving spatial resolution, and dose-efficient scatter rejection.[60-61] Moreover, the newly developed

imaging modalities, and biochemical markers are being introduced in breast cancer screening or diagnosis. The phase-sensitive breast x-ray imaging is also an evolving imaging modality that hypothesizes higher detectability with reduced radiation dose while still holding the advantages of breast x-ray imaging.

The diagnostic x-ray is electromagnetic radiation that carries high-energy photons-usually ranging between 20 keV and 150 KeV. As the x-ray traverse through the body organs, its energetic charged particles interact with various tissue via multiple mechanisms. When the x-ray photons pass through the tissue, they may penetrate without interaction or be scattered or absorbed by the matter. There are four major types of x-ray and matter interaction: a) Rayleigh scattering, b) Compton scattering, c) Photoelectric absorption, and d) pair production, which the first three contribute to diagnostic medical imaging.

However, the phase-sensitive x-ray imaging shall be described with wave-particle duality theory in quantum mechanics instead of classical Newtonian mechanics that only consider the particle-like properties of the x-ray beam. Four major phase-sensitive imaging techniques can be used to attain the phase shift information in the imaging sample. The experimental setup and the x-ray beam requirements make analyzer-based imaging, crystal interferometry, and grating-based interferometric imaging techniques challenging methods in clinical applications. On the other hand, a partially coherent x-ray beam can be utilized without additional tools in an in-line phase-sensitive imaging setup. The variations in thicknesses and refractive indices within the sample result in amplitude variation and phase shift in the projected x-ray.

When the detector is placed sufficiently far from the object, the x-ray phase shifts caused by various tissue with different electron densities result in destructive and constructive interference while the wave-front propagates over the introduced object-to-detector distance (R_2). Therefore, the x-ray amplitude and the phase shift depend on the tissue linear attenuation coefficients (μ) and tissue phase shift coefficients (φ), which the latter is two to three orders of magnitude larger than the first one in the diagnostic x-ray energy range for light elements such as human tissue. The tissue refractive index for the utilized x-ray beam is mainly responsible for the potential advantages of phase-contrast x-ray imaging.

The real part (δ) of the refractive index is mainly responsible for the x-ray phase shift, whereas the imaginary part (β) of the refractive index predominantly causes the x-ray amplitude shift. The tissue electron density impacts both components; however, the value for δ is usually much larger than β, and plays an enormous role in phase-contrast imaging compared to the imaginary part β. The modulus of the electron density can be obtained by the equation below:[62]

$$\rho = 2\pi \frac{\sqrt{(\delta-\delta_0)^2+(\beta-\beta_0)^2}}{\lambda^2 r_e} \quad (1)$$

where the subscript zero refers to the background material, λ is the wavelength of the x-ray, and r_e is the classical radius of the electron.

Breast cancer is classified into two broad categories: a) noninvasive and b) invasive breast cancer. The difference between invasive and non-invasive carcinoma is associated with the time of diagnosis. The malignant cells are confined to the breast tissue without spreading beyond the ductal walls or lobules if the cancer is diagnosed much earlier. However, the same pathology is clinically called invasive breast cancer if the cancer cells

have spread from the original site to other areas, like nearby breast tissue or beyond the ductal wall. The most common type of breast cancer -accounting for approximately 70% to 80% of all cases- starts from the milk ducts. If it is still in the non-invasive phase, the neoplastic change is called Ductal Carcinoma In-Situ (DCIS), and if it is transformed into an invasive type, it is called Invasive Ductal Carcinoma (IDC).

On the other hand, Invasive Lobular Carcinoma (ILC) is the second most common type of breast cancer (about 5% to 10% of all cases) and starts from the lobular site and spreads to nearby structures and, like IDC, potentially to the distant locations in the body through the bloodstream or lymphatic system. Despite the cancer type, all cancerous cells carry out specific traits: activated oncogenes or inactivated tumor suppressor genes, enhancing uncontrolled cell proliferation.

All the cell types in the body, with few exceptions, such as neurons and myocytes, keep the dividing capability lifelong for self-repair or to meet the organ's demand by the procedure called mitosis. The genes govern the behavior of each cell within a specific tissue type. The genes are the DNA sequences that can be assumed as the programming codes, dictating a specific type of protein production. A short sequence of these genetic codes is called proto-oncogenes, responsible for making a protein involved in cell growth and division. If a mutation occurs in a proto-oncogene, the mutated proto-oncogene is called an oncogene and could result in uncontrolled cell growth and high-frequency proliferation, ultimately causing the malignancy.

On the other hand, the tumor suppressor genes act as the brake, preventing the cell from dividing quickly. Any coding error in tumor suppressor genes, like a mutation in proto-oncogenes, may cause the cell to enter the mitosis phase more frequently,

potentially leading to malignancy. During mitosis, the genetic repository of the cells doubles in the *S phase* of the ″interphase″. DNA is one of the heaviest structural molecules inside the cell. As discussed earlier, the presence of an oncogene or the tumor suppressor gene malfunction forces the cell to enter mitosis. As a result, the malignant tissue contains many cells in the active mitotic phase compared to the non-malignant tissue. The mitotic cells contain two-fold DNA molecules as non-dividing cells, resulting in a higher electron density in malignant tissue.

Additionally, the cancerous cells in breast tissue make the vascular endothelial growth factor (VEGF) protein. The VEGF protein attaches to endothelial cells in arteries and triggers the blood vessels' regeneration to meet the nutritional demand of malignant cells. The broad network of blood vessels results in more blood per unit volume of malignant tissue, increasing the electron density of cancerous mass and neighboring structures, even to a greater extent.

The electron density of the entire breast in a premenopausal woman relative to the electron density of water equals 0.94±0.033.[62] Since the malignant tissue has high mitotic activity, the number of DNA molecules in the unit volume of the tissue is relatively higher than in normal tissue, in addition to the high number of blood cells circulating in expanded blood vessels within the malignant tissue, which result in a higher electron density in the malignant tissue. The average electron density of glandular tissue and tumor tissue is approximately estimated as 3.482E+23 and 3.564E+23, respectively, using a Compton scattering technique.[63] Undeniably, the electron density values for malignant lesions, glandular structures, and adipose tissue depend on age, diet, hormonal status, medication, and overall health status of the individual. For this reason, a broad

overlapping of values between the different individuals is expected. However, we expect a higher electron density in malignant tissue than in other healthy structures of the same tissue in any individual, regardless of age, hormonal status, and so forth.

The phase-sensitive x-ray imaging is highly sensitive to electron density variations in the imaging object. Therefore, phase-sensitive x-ray imaging could be an excellent modality for detecting pathological lesions, such as malignancies, in the breast tissue.

2.2 Signal Detection Theory (SDT)

All models in detection theory have at least two physiological processes: a) the sensory process, which transforms the physical stimulation into internal sensation, and b) the decision process, which comes up with a decision based on the output of the sensory process.[64] The decision process decides the response based on the internal representation of the sensory response whether to mark the image as *"Yes; the stimuli are present"* or *"No; the stimuli are not present"*. In the next layer of the decision process, a more detailed response is made on the confidence level of the observer, whether the signal is present or not.

The detection model evaluates a few performance measures: a) the conditional probability of a *"yes"* response when the signal is present (Hit Rate or HR), b) the conditional probability of a *"yes"* response when the signal is absent (False Alarm Rate or FAR), c) the conditional probability of *"no"* response when the signal is absent (Correct Rejection Rate or CRR), and d) the conditional probability of *"no"* response when the signal is present (Miss Rate or MR).

The High Threshold Model (HTM) defines the sensory threshold. When the stimulus is above the threshold, the decision process results in a *"yes"* response. The *"no"* response

is the outcome when the signal is below the threshold. However, the decision process might result in a *"yes"* response by a guess. The sensitivity of the sensory process and the guessing rate of the decision process are computed by the following equations:

$$p = \frac{HR-FAR}{1-FAR}, \tag{2}$$

$$q = FAR \tag{3}$$

where p is the probability that the signal exceeds the threshold and q is the decision criterion of the HTM. The extensive research has led to the rejection of the High Threshold Model.[65-66] The primary reason to reject the HTM is associated with the failing trait of this model in Receiver Operating Characteristics (ROC). The ROC curve plots the True Positive Rate (HR) against the False Positive Rate (FAR) when the observer changes the decision criteria. The expected function in HTM can be constructed by algebraic rearrangement of Equation 2, which will be a linear function for the High Threshold Model:

$$HR = (1 - p) \cdot FAR + p. \tag{4}$$

Nonetheless, the observed data in realistic detection experiments, using different degrees of response bias, produce an inverse exponential curve confined within the unit square, as shown in Figure 1.1.

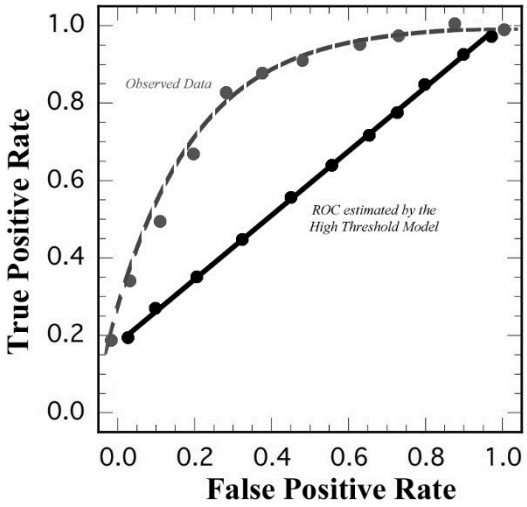


Figure 1.1: The Receiver Operating Characteristics (ROC) estimated by the High Threshold Model (solid line) and the realistic observed data (dashed line).

The alternative approach, Signal Detection Theory (SDT), was initially discussed in the 1950s. The SDT focuses on the ability of a detecting system to differentiate the information-bearing pattern (signal) under various conditions from the random pattern (noise) that distract the information of interest. When the detecting system is human, a wide range of characteristics such as user experience or training level, the signal-to-noise ratio in the receiver input, and the observer's psychophysiological state affect the observer's decision.

In Signal Detection Theory, there is no sensory threshold for the sensory process.[67] This model presumes a Gaussian signal distribution and Gaussian noise distribution, as shown in Figure 1.2. The mean of noise distribution (μ_n), equals 0, where the standard deviation, (σ_n), is believed to be 1.0. On the other hand, the mean (μ_s), and the standard deviation (σ_s) of signal-plus-noise distributions depend on the sensitivity of the sensory process and the intensity of the Signal-to-Noise Ratio (SNR).

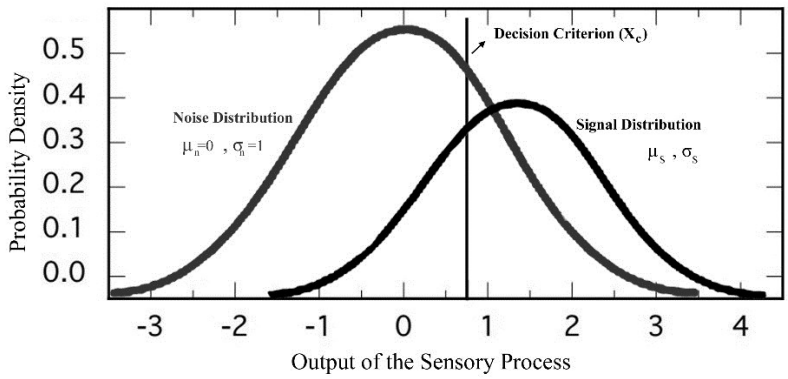


Figure 1.2: The Gaussian Probability Density functions for noise and signal-plus-noise distribution from the sensory process. When the output of the sensory process is higher than X_c, it results in a "yes" response, while the "no" response is the result of output lower or equal to X_c. The values for μ_s and σ_s depend on the strength of the SNR in the detection experiment.

In analyzing the performance of an imaging modality, two distinct, yet associated tasks can be performed based on the signal detection theory. The first type examines the observers' ability to distinguish the images with and without the stimuli. On the other hand, type-2 generally evaluates the confidence level of the observers in accurately classifying the images in the type-1 task. Moreover, the *"response-specific"* type-2 performance is specifically interested in investigating how confident a user performs **only** at *"signal present"* trials or **only** at *"signal absent"* trials.

In type-1 SDT, like the High Threshold Model, there are two possible objective stimuli identities for each trial: Signal Absent stimulus (SA) and Signal Present stimulus (SP). When the observer classifies the stimulus, he/she marks the trial as a Positive case (P) or a Negative case (N). Depending on the user's response and the trial identity, the experiment result could be categorized into four classes, as shown in Table 2.1.

Table 2.1: Possible outcomes in type 1 SDT analysis. SA and SP represent the signal absent and signal present, respectively.

	User's response	
Stimulus	Negative	Positive
SA	Correct Rejection (CR)	False Alarm (FA)
SP	Miss (M)	Hit (H)

In addition to the imaging system's ability to reveal the *"signal present"* cases among all trials, the various *non-system-specific* parameters may alter the system's performance in subjective evaluations. For instance, when the observer knows that incorrect *"SA"* will be penalized more heavily than incorrect *"SP"* or when the disease of interest is more prevalent in the general population, that may result in the observer's tendency to classify the suspicious trials as positive cases. By producing various manipulations that affect the observer's inclination in different circumstances, we collect multiple HR and FAR pairs and construct the ROC curve that presents the HR as a function of FAR. Therefore, the sensitivity is fairly a stable property of each sensory process.

However, the decision criterion utilized by each observer can widely diverge from task to task or from time to time. The instruction provided to the observer, the ratio of signal-carrying trials to non-signal trials in the experiment, and the payoff matrix for defining the benefits of correct decisions versus the cost of wrong decisions, play a tremendous role in the decision criterion. Therefore, a user may use a completely different decision criterion at any circumstance or time. Hence, the SDT has offered a robust methodology for analyzing the ability of an observer to discriminate stimuli in a signal-plus-noise image (sensitivity) while it considers the *"response bias"*.

The sensitivity of the sensory process (d_a or discriminability index) for the detection experiment can be defined as:[66]

$$d_a = \frac{(\mu_s - \mu_n)}{\sqrt{\frac{\sigma_s^2 + \sigma_n^2}{2}}} \quad (5)$$

where subscripts *s* and *n* represent signal-plus-noise and noise, respectively. The value for μ_s determines the discriminability index of the system. As we expect that μ_n equals zero, the $\mu_s = 0$ generates the straight diagonal ROC and zero discriminability index, indicating completely random discrimination. As μ_s increases, ROCs are being bowed toward (1,0), as shown in Figure 1.1.

On the other hand, the observer shall provide his confidence rating that characterizes his response's likelihood of correct response in the type-2 SDT model. The possible outcomes in the response-specific type-2 task and the principles for the analysis are similar to type-1 with minor variation. For the sake of simplicity (usually, there are more than two confidence levels), if we set the confidence rating as *"high"* and *"low"* for the responses given by the observer, four possible outcomes for the response-specific type-2 task will be as follow: a) high-confidence correct trials, b) low-confidence correct trials, c) high-confidence incorrect trials, and d) low-confidence incorrect trials.

By direct analogy with the type-1 analysis, we can establish four parameters of type-2 Hits (H_2), type-2 Misses (M_2), type-2 Correct Rejection (CR_2), and type-2 False Alarm (FA_2). Consequently, the classification of type-2 performance analysis could be subdivided into a response-specific analysis. When the observer reports that the image does not contain the signal (N), the accuracy of the observer's response might be correct if the trial is *"signal absent"* or incorrect if the trial is *"signal present"*. Based on the confidence rate that the observer provides, the response-specific (conditional ″N″ or ″P″) type-2 Miss, Hit, CR, and FA can be constructed as shown in Table 2.2.

Table 2.2 Response-Specific type 2 outcomes.

Response	Accuracy of observer	Confidence	
		LOW	HIGH
Negative	Incorrect (Type 1 M)	CR_2 | N	FA_2 | N
	Correct (Type 1 CR)	M_2 | N	H_2 | N
Positive	Incorrect (Type 1 FA)	CR_2 | P	FA_2 | P
	Correct (Type 1 H)	M_2 | P	H_2 | P

The type-2 performance analysis shall be evaluated for two different conditions, based on the user response only for the *"negative"* case response and *"positive"* case response as follow:

$$HR_2 \mid N = p(\text{High conf.} \mid \text{stimulus=SA, response=N}) = \frac{n(high\ conf.\ CR)}{n(CR)}, \quad (6)$$

$$FAR_2 \mid N = p(\text{High conf.} \mid \text{stimulus=SP, response=N}) = \frac{n(high\ conf.\ M)}{n(M)}, \quad (7)$$

$$HR_2 \mid P = p(\text{High conf.} \mid \text{stimulus=SP, response=P}) = \frac{n(high\ conf.\ H)}{n(H)}, \quad (8)$$

$$\text{and } FAR_2 \mid P = p(\text{High conf.} \mid \text{stimulus=SA, response=P}) = \frac{n(high\ conf.\ FA)}{n(FA)}. \quad (9)$$

From the equations above, the overall HR_2 and FAR_2 of the detecting system can be defined as weighted averages of the response-specific type-2 FARs and HRs, where the weights are determined by the proportion of correct and incorrect trials from each response type (N response or P response):

$$HR_2 = \frac{n(high\ conf.\ correct)}{n(correct)} = \frac{n(high\ conf.\ H)+n(high\ conf.\ CR)}{n(H)+n(CR)}$$

$$= \frac{n(H)\times HR_2|P+n(CR)\times HR_2|N)}{n(H)+n(CR)}$$

$$= p(H \mid correct) \times HR_2|P + [1 - p(H|correct)] \times HR_2|N \quad (10)$$

and similarly,

$$FAR_2 = p(FA \mid incorrect) \times FAR_2|P + [1 - p(FA|incorrect)] \times FAR_2|N. \quad (11)$$

To extend the confidence rating scheme into a discrete scale ranging from 1 to arbitrary number C, the observer can select a value c, where $1<c<C$. In this general rating scale, all confidence ratings greater than or equal to c shall be treated as *"high confidence"*, and the confidence rating smaller than c will be classified as *"low confidence"*. Each choice of c creates an individual FAR_2 and HR_2 pair. Therefore, the type-2 ROC curve can be constructed from those calculated points for varying c values.

2.3 The Performance of a Diagnostic Tool in Medicine

The signal detection theory is a broad scientific concept applied to all signal detecting systems regardless of the signal types or the detection processes. The diagnostic performance of an imaging modality in clinical settings also arises from the SDT, where a physician radiologist discriminates the diseased cases from the non-diseased cases. The clinicians are more familiar with different terms on the same notions described in Section 2.2. To lineup the widely accepted definition in medical imaging with the principles of SDT, the utilized conceptions in evaluating the diagnostic performance of an imaging modality are redefined as follows and will be cited throughout this work: a) True Positives (TP), b) True Negatives (TN), c) False Positives (FP), and d) False Negatives (FN). The confusion matrix can be used for dichotomous clinical tests, differentiating the individuals with the disease of interest from healthy individuals. Table 2.3 summarizes the observational outcomes in a dichotomous test and correlates it with universal concepts derived from the signal detection theory.

Table 2.3 Confusion Matrix for a dichotomous test in clinical settings

Interpretation		Patient's real diagnosis: Diseased (D+)	Patient's real diagnosis: Non-Disease (D-)	Total
Diseased	(R+)	**TP**	**FP**	m_1
Non-diseased	(R-)	**FN**	**TN**	m_2
Total		n_1	n_2	N

The statistical measures for the performance of an imaging modality as a diagnostic or a screening tool based on the confusion matrix stated above are defined as follows:

- ***"Sensitivity"* or True Positive Fraction (TPF)**: it is the conditional probability of a positive test (R+) given that the patient has the disease (D+):

$$TPF = p(R+ \mid D+) = \frac{TP}{TP+FN} = \frac{TP}{n_1}. \quad (12)$$

- ***"Specificity"* or True Negative Fraction (TNF)**: it is the conditional probability of a negative test (R-) given that the patient does not have the disease (D-):

$$TNF = p(R- \mid D-) = \frac{TN}{TN+FP} = \frac{TN}{n_2}. \quad (13)$$

- **False Positive Fraction (FPF)**, which is also called the type-I error and is the probability of a positive test (R+) given that the patient does not have the diseases (D-):

$$FPF = p(R+ \mid D-) = \frac{FP}{FP+TN} = \frac{FP}{n_2} = 1 - Specificity. \quad (14)$$

- **False Negative Fraction (FNF)**, which is also called the type-II error, is the probability of a negative test (R-) given that the patient has the diseases (D+):

$$FNF = p(R- \mid D+) = \frac{FN}{FN+TP} = \frac{FN}{n_1} = 1 - Sensitivity. \quad (15)$$

The intrinsic properties of a detecting system, such as SNR and target contrast, result in different confusion matrices for any given imaging modality. Additionally, the decision criterion provided to the observers as a reading rule and the personalized

characteristics in interpreting the diagnostic test generate the varying X_c. If X_c shifts to the left (decreased) in Figure 1.2, the sensitivity increases as specificity decreases. Contrarily, if the X_c shifts to the right, the sensitivity decreases as specificity increases. Therefore, the simple statistical measures are not suitable to uniquely define a *"good test"*, and a multifaceted measure such as the ROC study shall be employed to evaluate the overall performance of an imaging system.

Again, to line up the diagnostic ROC in medical imaging with SDT, the TPF and FPF at any given Xc are re-defined, and the function of the ROC curve ($ROC(\cdot)$) can be mapped as Equation 18:

$$TPF(X_c) = p(R \geq X_c | D+), \tag{16}$$

$$FPF(X_c) = p(R \geq X_c | D-), \tag{17}$$

$$ROC(\cdot) = \{FPF(X_c), TPF(X_c), X_c \in (-\infty, +\infty)\}. \tag{18}$$

Extensive phantom studies have been done to objectively characterize and evaluate the performance of the phase-contrast x-ray imaging (PCI) systems. Additionally, multiple phantom studies have focused on the detectability of breast PCI imaging and compared it with conventional mammography. As is previously stated, primitive measures such as a system's detectability determine the diagnostic performance of an imaging system relatively at a limited capacity and ignore the multi-dimensional nature of the observer performance study performed by a human observer. Thus, a complete subjective characterization of the PCI system that considers the human observation's compound nature is missing. The widely accepted scheme that integrates all fundamental aspects of the signal detection theory would be the ROC paradigm that considers the intrinsic characteristic of a detecting system and the user-dependent variables in image

interpretation to evaluate the overall performance of newly developed PCI systems in breast cancer diagnosis and screening.

2.3.1 The basics of the Receiver Operating Characteristic (ROC) study

Prior to ROC analysis, the diagnostic performance of an imaging system was being measured by the percentage of correct diagnoses in the clinical settings. On the other hand, in preclinical investigations, various objective and subjective evaluations such as system's spatial resolution, SNR, CNR, Contrast-Detail (CD) analysis, and the detection rate have been employed for performance comparison.

The diagnostic performance based on correct percentage encounters two significant limitations. It strongly depends on the proportion of persons in a population who have the disease (prevalence rate)[68] and does not reveal the relative frequencies of false-positive and false-negative errors. Due to stated limitations, it was necessary to evaluate the diagnostic performance of the imaging modalities with a more comprehensive method. The ROC analysis is a statistical tool widely accepted for a comprehensive investigation of medical diagnostic accuracy. It was initially developed during WWII to detect military weapons, but soon after, it was adopted into other research areas.

Each point on the ROC curve (operating points) represents the TPF (*y*-axis) and FPF (*x*-axis) associated with a given X_c to identify the positive and negative cases. To compare the accuracy of the two diagnostic tests, different types of summary measures (Figure of Merit) for the ROC studies are applicable. The commonly used Figure of Merit (FOM) is the *"area under the ROC curve"* (AUC).[69] The AUC correlates the sensitivity and specificity of the detecting system and is independent of X_c, the cost and benefit of varying thresholds, and the prevalence of the disease.

The AUC can be estimated either parametrically or non-parametrically. Parametric AUC estimation assumes that the test results follow a binormal distribution where non-parametric AUC drives from an empirical ROC curve without assuming any specific distribution. The non-parametric (empirical) AUC is the summation of the areas of the small trapezoids constructed under the empirical ROC curve between multiple *"operating points"*.[70] To estimate the non-parametric AUC, one should define the T_1 component ($V(T_{1i})$) of the ith subject and T_0 component ($V(T_{0j})$) of the jth subject as follow:

$$V(T_{1i}) = \frac{1}{m}\sum_{j=1}^{m}\Psi\left(T_{1i}, T_{0j}\right) \tag{19}$$

and

$$V\left(T_{0j}\right) = \frac{1}{n}\sum_{i=1}^{n}\Psi(T_{1i}, T_{0j}) \tag{20}$$

where

$\Psi(X,Y) = 0 \quad if\ Y > X,$

$\Psi(X,Y) = 0.5 \quad if\ Y = X,$

$\Psi(X,Y) = 1.0 \quad if\ Y < X.$

The subscript *"1"* and *"0"* denote the presence of disease and absence of disease in the population, respectively. *"m"* is the number of the diseased patients, and *"n"* is the number of patients without the disease of interest.

The following equations estimate the empirical (non-parametric) AUC_{emp} and its variance:

$$AUC_{emp.} = \sum_{i=1}^{m}\frac{V(T_{1i})}{m} = \sum_{j=1}^{n}\frac{V(T_{0j})}{n} \tag{21}$$

$$Var\left(AUC_{emp}\right) = \frac{1}{m(m-1)}\sum_{i=1}^{m}\left[V(T_{1i}) - AUC_{emp}\right]^2 + \frac{1}{n(n-1)}\sum_{j=1}^{n}\left[V\left(T_{0j}\right) - AUC_{emp}\right]^2. \tag{22}$$

The parametric AUC assumes that the values of the criterion variables in both the positive population (*X*) and the negative population (*Y*) are normally distributed:

$$X \sim N(\mu_x, \sigma_x^2) \& \quad Y \sim N\left(\mu_y, \sigma_y^2\right) \tag{23}$$

where μ_x and μ_y are the means and σ_x^2 and σ_y^2 are the variances of the positive and negative populations, respectively. The trajectory of the ROC curve can be defined as:[71]

$$\{FP(c), TP(c)\} = \left\{\Phi\left(\frac{\mu_x - c}{\sigma_x}\right), \Phi\left(\frac{\mu_y - c}{\sigma_y}\right)\right\} \quad -\infty < c < +\infty \tag{24}$$

where $\Phi(z)$ is the cumulative normal distribution function.

The following equation estimates the parametric AUC:

$$AUC = \int_{-\infty}^{\infty} \left[\Phi\left(\frac{\mu_y - c}{\sigma_y}\right) \Phi\left(\frac{\mu_x - c}{\sigma_x}\right)\right] dc. \tag{25}$$

To estimate the variance of the parametric AUC with an assumption of normal distribution, we can use the method of differentials. If μ_x and μ_y are the means of the positive and negative populations, respectively, σ_x^2 is the variance of positive population and σ_y^2 is the variance of the negative population, let

$\Delta = \mu_y - \mu_x$,

$a = \frac{\Delta}{\sigma_y}$, $b = \frac{\sigma_x}{\sigma_y}$,

and $E = exp\left(-\frac{a^2}{2(1+b^2)}\right)$

then:

$$Var_{AUC} = \left(\frac{\partial AUC}{\partial \Delta}\right)^2 V(\Delta) + \left(\frac{\partial AUC}{\partial \sigma_x^2}\right)^2 V(s_x^2) + \left(\frac{\partial AUC}{\partial \sigma_y^2}\right)^2 V\left(s_y^2\right) \tag{26}$$

where $V(\Delta) = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$ and $V(s_x^2) = \frac{2\sigma_x^4}{n_x - 1}$ & $V\left(s_y^2\right) = \frac{2\sigma_y^4}{n_y - 1}$.

To compare the diagnostic performance of the two imaging modalities, one can compare the AUC estimated from ROC analyses of these two independent samples using the z-test to either accept or reject the null hypothesis:[71]

$$z = \frac{AUC_1 - AUC_2}{\sqrt{V(AUC_1) + V(AUC_1) - 2Cov(AUC_1, AUC_2)}}. \quad (27)$$

We are interested in evaluating an imaging modality in *"clinically more relevant"* conditions in a few instances. For example, in a test with known higher sensitivity and lower specificity, we may want to focus on the performance of the test in the high sensitivity portion of the ROC curve. Analogously, there may be times that we know that a particular diagnostic test would not be beneficial in clinical settings, as its FPF is greater than a specified value, then we might need to focus on the restricted portion of the AUC that represents the desired FPF. Hence, a few alternative figures of merit have been suggested in ROC analyses, such as the *"partial area under the ROC curve"* (pAUC) to the left of a specified FPF, the pAUC between the two FPF values, a TPF value at a given FPF, or an FPF value at given TPF.[71]

2.3.2 The ROC rating Data Analyses

A variety of statistical analyses can be used to interpret and compare the result of ROC studies. For instance, the CORROC analysis, Swets/Pickett Analysis, and the Jackknife analysis are the most applied statistical methods in radiology experiments. Suppose we have the rating ROC data for *n* imaging modalities (*i=1, 2,..., n*). A group of *r* readers (*j=1, 2,..., r*) read a total number of *c* cases that are the same for all readers (*k=1, 2,..., c*). The rating ROC data for each modality can be arranged in a matrix M_i with *r* rows and *c* columns where X_{ijk} is the variable for the reading response for the modality *"i"* by reader *"j"* for the case *"k"*:

$$M_i = \begin{bmatrix} X_{i11} & X_{i12} & \dots & X_{i1c} \\ X_{i21} & X_{i22} & \dots & X_{i2c} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ X_{ir1} & X_{ir2} & \dots & X_{irc} \end{bmatrix} \quad (i = 1, 2, \dots, n). \tag{28}$$

2.3.2.1 CORROC analysis:

The CORROC model estimates the parameters of the test and the corresponding variance-covariance matrix by the Maximum Likelihood model.[72] This method is applicable when only one reader evaluates the performance of only two imaging modalities (*i=2, j=1*). The CORROC is a relatively easy-to-use model but has not been generalized for multiple readers and modalities. When the total number of cases is sufficiently large, the reading responses could be considered as normally distributed data, then:

$$H_\circ: A_2 = A_1,$$
$$\text{and } Z = \frac{\hat{A}_2 - \hat{A}_1}{\sqrt{\hat{A}_2 + \hat{A}_1 - 2Cov[\hat{A}_1, \hat{A}_2]}}. \tag{29}$$

The null hypothesis ($H_\circ$) can be either rejected or accepted by the z-test. If there are multiple readers *(i ≥ 2)*, the CORROC method shall be separately applied to each observer's data, and hence, the null hypothesis for any accuracy measure $''A''$ (e.g., AUC) should be tested for each observer separately. A few advanced statistical techniques allow combining the data from multiple readers to draw a single conclusion, if and only if an independent test is utilized during the data collection. However, in most radiologic observer performance studies, the observers read the same cases in two different imaging modalities for comparison purposes, so the assumption of mutual independence is violated, which results in an undesired limitation on the promptly applicable CORROC method.

2.3.2.2 Swets/Pickett Analysis:

Unlike CORROC analysis, this method can be generalized to a population of cases by multiple observers. Their method considers the random effects associated with cases, the observers, and the reading events. They proposed that the case correlation might be estimated by dividing the total case sample into subsamples, followed by the ROC analyses for the subsamples.[73] The suggested method runs into a few limitations: a) the maximum likelihood estimation may fail if the subsample size is not large enough due to limited sample size quantity, b) the subsample population might be formed with accidental bias, which ignores the impact of individual cases on the overall accuracy of the test, and c) subdividing the total sample size into different groups reduces the statistical power of the test. Hence, this method is not usually recommended unless the sample population is large enough to avoid stated limitations.

2.3.2.3 Jackknife analysis:

The Jackknife method is a predominantly accepted tool for statistical evaluation of the accuracy performance of a diagnostic test.[74] In 1984, McNeil and Hanley[75] utilized the Jackknife method in ROC analyses and suggested that this method can estimate the case variability and the reader variability (multiple-readers/multiple-cases). Later in 1992, Dorfman and colleagues introduced a method using the pseudo-values estimated by the jackknife method for accuracy measures offered by ROC studies.[76]

In the rating-assigned study, The Dorfman-Berbaum-Metz (DBM) multiple-reader multiple-case (MRMC) significance testing method that applies the jackknife procedures evaluates the individual contribution of each case on the overall performance of any given modality by generating the *"pseudovalues"*. Assume we want to compare the

performances of *i* modalities (*i=1, 2,..., I*) by reading the results of *j* readers (*j=1, 2,..., J*) and *k* cases (*k=1, 2,..., K*) utilizing the AUC. The pseudovalue Y_{ijk} is defined by:

$$Y_{ijk} = K \times A_{ij} - (k-1)A_{ij(k)} \tag{30}$$

where A_{ij} is the AUC for modality *i* and reader *j*, when all cases are included in AUC estimation, and $A_{ij(k)}$ is the AUC for modality *i* and reader *j* when the case *k* is excluded from the analysis. A 3-dimensional IJK matrix of pseudovalues will be formed by repeating the jackknife procedure for all modalities, readers, and cases.

The pseudovalues interpret the correctness of decisions made by multiple users. For instance, if a **diseased** case *k* has received a high-rating score (definitely diseased), removing this case from the initial dataset in the jackknife method results in a different data- set with one less **high-confidence correct** decision, which causes AUC for the new data- set ($A_{ij(k)}$) to slightly decrease, and therefore Y_{ijk} increases. Contrarily, if a **diseased** case k' has received a low-rating score (definitely non-diseased), removing k' from the dataset will cause a new set with one fewer **high-confidence incorrect** decision that results in an increased $A_{ij(k')}$ and a decreased $Y_{ijk'}$. Intuitively, the high-confidence *TP*s and high-confidence *TN*s yield higher than average Y_{ijk} while the high-confidence *FP*s and *FN*s yield lower than average pseudovalues.

Several variances for the pseudovalues are calculated from each dimension of the 3D matrix. The variances test the reader and case variabilities across the modalities or the reader variability across the cases. The magnitude of the variances can be evaluated by the Analysis of Variance (ANOVA) test to either accept the null hypothesis or reject it in comparing the multiple imaging modalities[76].

2.3.3 Multi-Reader Multi-Case Receiver Operating Characteristic Analysis:

The diagnostic performance of a clinical test heavily depends on the subjective interpretation of the trained users due to inherent variability in users' accuracies. Therefore, several trained observers (usually between 4 and 15) always evaluate the diagnostic performance of a potential test. The study is generally designed as a factorial experiment, where all diagnostic tests evaluate the same patients under the experiment, and the same readers interpret the results from all diagnostic tests.[77] There are at least five different methods that analyze the multi-reader ROC studies: a) ANOVA of Pseudovalues (DBM) model, b) ANOVA with Corrected F-test (OR) model, c) Multivariate WMW statistic model, d) Bootstrap of Components-of-Variance (BWC) model, and e) Bayesian Hierarchical Ordinal Regression (HROC) model.

The factorial experiment (paired-patient paired reader) design can introduce several sources of noise in estimating the diagnostic accuracy of the test. The variability among the patients (patient difficulty) and the variability among the readers (readers' skill and training competence) are the primary sources of noise in the factorial experiments. Additionally, the correlation between accuracy estimations across the tests and the correlation between accuracy estimations across the readers may impact the interpretation of system performance and model the test accuracy in several ways.

Most of the time, the difference between the means of readers' accuracy (e.g., AUC) is used and analyzed for test competence. However, there might be the conditions that the means for readers' accuracy are the same for two different diagnostic tests, but the variance of readings or how the confidence scale plays a role in patient management

impact our decision to select the suitable test in clinical settings. The brief review of each model and the key features of these five methods are as follow:

2.3.3.1 ANOVA of Pseudovalues or "Dorfman-Berbaum-Metz" (DBM) method:

The basic idea proposed by Dorfman and his colleagues is based on computing the jackknife pseudovalues[76] for i^{th} modality, j^{th} reader, and k^{th} patient, as discussed in section 2.3.2.3. In the DBM method, the accuracy tests can be characterized using various summary measures such as sensitivity, specificity, AUC, the partial area under the ROC curve (pAUC), and sensitivity at a fixed FPR by either parametric or non-parametric estimation. This model compares the means of accuracy measures. However, it does not provide the variances of accuracies.

2.3.3.2 "Obuchowski-Rockette" (OR) method

Unlike the DBM method, which utilizes the pseudovalues for modality-reader-case, the OR model is the modality-by-reader ANOVA test with correlated errors. However, this method also characterizes any index for the summary measures, utilizing the parametric and non-parametric estimations.[78] Analogous to the DBM method, the OR method only compares the means of accuracy measures.

If the study is designed as a factorial experiment, the OR model can be defined as:

$$\theta_{ij} = \mu + \alpha_i + R_j + (\alpha R)_{ij} + \varepsilon_{ij} \tag{31}$$

where θ_{ij} denotes the accuracy measure (e.g., AUC) for the i^{th} modality and j^{th} reader, μ is the overall population mean, α_i is the fixed effect of diagnostic modality i, R_j is the random effect of reader j. $(\alpha R)_{ij}$ denotes the random effect of interaction between diagnostic modalities and readers and ε_{ij} is the error term with zero mean and the variance of σ_ε^2. Note that each reader reads the same cases from each diagnostic modality

in a factorial study, and therefore, the error terms are not independent. However, both random effects are assumed independent and normally distributed and are assumed independent of ε_{ij}. Hence, three possible covariances of the error terms can be defined as follow:

$$Cov\left(\varepsilon_{ij}, \varepsilon_{i'j'}\right) = \begin{cases} Cov_1 & i \neq i' \; and \; j = j' \\ Cov_2 & i = i' \; and \; j \neq j' \\ Cov_3 & i \neq i' \; and \; j \neq j' \end{cases} \quad (32)$$

2.3.3.3 "Multivariate Wilcoxon-Mann_Whitney" (WMW) method

This method only evaluates the Area under the ROC curve as an accuracy measure of several diagnostic tests. The WMW method is the extension of the non-parametric approach introduced by DeLong et al., which is applicable only when one observer interprets each diagnostic test.[70] The later model extends the previous model to the multivariate situation from multiple readers in contrast with single observer studies.[79] This model only measures the accuracy index of AUC and statistically compares the means of the accuracies among the diagnostic tests.

2.3.3.4 "Bootstrap of Component-of-Variance" (BWC) method

The BWC model uses the same approach for component-of-variance as which used by the DBM model. The difference is that BWC analysis estimates uncertainties of the performance by the DBM method and decomposes these uncertainties into contributions in a DBM method.[80] In realistic clinical investigations, the population size and the readers are always limited. The BWC model replaces the set of population experiments with bootstrap resampling experiments from the available finite data set.

The bootstrap resampling forms a system of linear equations which can be solved to estimate the weight of each certainty (i.e., randomness caused by the reader or case)

within the population. Estimating components of the variances allows the investigators to define the required sample size for the pivotal study from a pilot study. The test provides any ROC index and statistically compares the means and variances of estimated indices.

2.3.3.5 "Hierarchical Ordinal Regression" (HROC) method

The HROC model generates the ROC curves for each observer who reads multiple cases. The average curve is defined from individual ROC curves, and the variabilities across the multiple observers and the consistency of the readers' interpretation over the multiple cases are analyzed utilizing the Bayesian hierarchical ordinal regression model.[81] The model is constructed based on the unobserved *"latent variables"* inferred from the observable variables. This model provides any ROC index and compares the means and variances of the accuracy indices.

2.3.4 Location-Specific ROC study

In the ROC paradigm, the reader is asked to identify the suspicious lesion and rate the identified feature, but correctly localizing the suspected lesion is not the focus of the investigation. However, this approach is most appropriate when the disease of interest shows a diffuse pathologic characteristic in the body, and then the single ratings offer the needed information for the ROC studies. When the disease is manifested by a localized lesion, such as breast adenocarcinoma, the investigator cannot rule out that the observer missed the actual pathology (False Negative outcome) and erroneously report a non-diseased feature as a suspected lesion (False Positive outcome). Therefore, marking the correct location for the suspected pathology rewards the correct localization, and incorrect localization penalizes the observer's wrong decision rather than ignoring both FP and FN outcomes when a suspected lesion is identified in an incorrect location.

2.3.4.1 Location Receiver Operating Characteristics (LROC) paradigm

Like the ROC paradigm, the observers provide the rating that the lesion is present, but unlike the ROC, the observers also mark the most suspicious location of the pathology in the LROC paradigm.[82] The experimenter should set an acceptance radius based on the clinical significance of correct localization, and any localization within that radius is considered correct localization (CL) or otherwise.

Contrarily to the ROC curve, in which the x-coordinate and y-coordinate respectively represent the FPF (x_c) and TPF (x_c) at the observer's decision threshold (X_c), the LROC utilizes the different variables in curve plotting. The fraction of diseased images that are correctly identified and localized at the observer's decision threshold ($TPF_{CL}(x_c)$) plots along the y-axis. The LROC curve contains the left-hand endpoint (0,0) because

$$\lim_{X_c \to \infty} TPF_{CL}(X_c) = 0.$$

However, unlike the ROC, the LROC paradigm may not be constrained to the right-hand endpoint (1,0) as $X_c \rightarrow -\infty$ because the LROC study has some finite fraction of diseased images with incorrect localization. This fraction of less than unity will constrain the ordinate for the uppermost endpoint. The area under the LROC curve (AUC_L) can be served as one figure of merit, precisely like the AUC for ROC, with the only difference that the ideal system may not have the AUC_L equal to 1.

2.3.4.2 Region of Interest Receiver Operating Characteristic (ROIROC) paradigm

This method is somewhat like the LROC paradigm, with the difference that the experimenter divides each image into multiple regions based on clinical considerations.[83] For example, the breast might be divided into five ROIs: a) superior- lateral, b) Inferior lateral, c) superior medial, d) inferior medial, and e) retro-areolar. The defined ROIs must

be the same in all images. Afterward, the observer is asked to grade each ROI based on his/her confidence in whether the lesion exists. The experimenter treats each ROI as one observational unit (instead of each patient as an observational unit) and performs the usual ROC analysis on the obtained data.

2.3.4.3 Free-Response Receiver Operating Characteristic (FROC) paradigm

Unlike the LROC paradigm, where the observer is forced to localize an abnormality (even in non-diseased cases), the observer is free to mark and rate as many abnormalities as he thinks above X_c, according to his clinical judgment. There is one crucial difference in the FROC rating system compared to the usual ROC. In ROC, the lowest rating denotes that the reader is highly confident of the absence of a lesion on the image, but in FROC, the lowest rating represents low confidence of the observer for the presence of the lesion in a marked location (otherwise, the reader must fill up the entire area of the image with the lowest rating possible).

An additional distinction for location base ROC (such as FROC), as discussed earlier, is the acceptance radius. The experimenter defines the radius tolerance to classify the identified lesions by the readers as *"lesion localization"* (LL) and *"non-lesion localization"* (NL). The expected tolerance depends on the clinical applications and the significance of the stringent localization in patients' outcomes. However, the marked location for the center of the lesion should generally not be greater than the diameter of the lesion. The minimum acceptance tolerance should be greater than 3 mm to avoid marking small lesions as NL due to hand jitter.

The *"Lesion Localization Fraction"* at threshold X_c ($LLF(x_c)$) is defined by the number of correctly localized lesions divided by the total number of lesions in the

dataset. Similarly, the *"Non-lesion Localization Fraction"* at X_c (NLF(x_c)) is defined as the number of NLs divided by the total number of images.[84] The FROC curve is the plot of the LLF on the Y coordinate vs. the NLF on the X coordinate. The FROC curve, like LROC, is also not constrained within the unit-square because the X coordinate can be extended to infinity as the NLF can theoretically be any value smaller or greater than 1.0. Chakraborty et al. have suggested an alternative way[85] to represent FROC data to constrain the curve within the unit square. The AFROC2 (Alternative FROC) plots LLF vs. FPF and limits the X coordinate to the interval [0,1] as FPF is always less than unity.

2.3.5 The FROC rating Data Analysis

The most common FOM used in FROC analysis is the area under the AFROC2 curve. The area under the AFROC2 curve rewards a higher-rated *"lesion localization"* (LL) more than a lower-rated LL. Similarly, it penalizes a higher-rated *"false positive"* (FP) more than a lower-rated FP. The noticeable difference between FROC and ROC paradigms is that the *"true negatives"* are unclassified in FROC investigations, and consequently, the system's specificity is unknown.

However, one can indirectly assess the system's ability to identify the true negatives as the observers who avoid marking non-diseased regions would also tend to have smaller FPF. The area under the AFROC2 curve for these observers would be slightly higher than those that may cause lower specificity due to high false negatives. Hence, the figure of merit for the area under the AFROC2 curve could credit observers who have not marked the non-diseased regions.

2.3.5.1 Jackknife Alternative FROC2 (JAFROC2) analysis

In the JAFROC2 method, the area under the AFROC2 curve serves as a figure of merit, and the DBM MRMC ANOVA method is used to evaluate the statistical significance. The pseudovalues are determined analogously to the Y_{ijk} in the ROC paradigm with a slight difference where the pseudovalues are calculated from the trapezoidal area under the AFROC2 curves. Moreover, the area under the ROC curve has an intuitive context and is always bounded between 0.5 and the unity. However, the area under the AFROC2 curve is expected to range from zero to unity because, intuitively, it is possible that one at least marks all non-diseased cases once where none of the lesions are correctly localized.

2.3.5.2An alternative method to JAFROC2 analysis: JAFROC1

The previously discussed FOM for the area under the FROC2 curve (JAFROC2) only considers the LL and does not utilize the NL on diseased cases. Therefore, an alternative FOM has been introduced, which uses a different definition for the x-axis and is called JFROC1[85]. In this method, the total number of highest-rated NL is divided by the total number of cases instead of simple FPF, in which the denominator is the non-diseased cases. Since the FPF is estimated using the total number of cases, rather than NL divided by non-diseased cases, it is more stable, and therefore, JAFROC1 provides a higher statistical power than JFROC2.[86]

2.4 The Observer Preference Study in Medicine

The terminology of ″preference″ in medical imaging is entirely different from most visual preferences in other fields. The purpose of imaging in medicine is to find the abnormal anatomy or functionality in the human body; hence, the aesthetic property of

the acquired images is not a point of judgment. This argument becomes even more dominant when medical imaging utilizes harmful ionizing radiation. According to SDT theory, we know that observers could prefer the images with higher SNR over the images with lower SNR. However, the signal-producing radiation shall always be limited due to health concerns among the patients. Therefore, the fundamentals of observer preference studies are still applicable in clinical observer preference studies with some variation in definition from an aesthetic criterion to the preference in diagnostic capabilities.

2.4.1 Two-Alternative Forced Choice (2AFC) studies

The 2AFC study is usually considered the most optimal preference study in medical imaging. The average probability of choosing an imaging modality/technique over another provides the relative preference of that system. The concerning drawback of the 2AFC paradigm is the number of required trials to measure the preference when more than two systems are compared. When *n* systems are being compared, the total trials that are required will be $n^2 - n$, which might be expensive and time-consuming.

2.4.2 Rank Ordering

The rank-ordering task presents available images from all imaging modalities/techniques, and the observer is asked to rank them from most to least preferred images. Ultimately, the average rank order for each imaging system is calculated as the measure of relative preference. Unlike the 2AFC, the rank-order paradigm requires only one trial for the *n* imaging system, but it might be impractical to simultaneously present the k images from a different system to the observers.

2.4.3 Rating

When the number of stimuli is large or the presented image contains complex details, the rating paradigm might be an appropriate alternative for the preference study. The rating could be either the Likert scale or continuous rating scales. Analogous to two previous paradigms, the average rating for the modality-case-observer is acquired to measure relative preference. Like the rank-ordering paradigm, only n trials are necessary to measure the relative preference. However, consistent ratings across the trials could be challenging, specifically at the beginning of the experiment. Therefore, a representative sample of cases is initially presented to observers alongside detailed instruction from the investigator to train the observers to make synchronized responses throughout the experiment.

Chapter 3. The Comparison of the High-energy and Mid-Energy Phase-Sensitive Imaging Systems: Pre-Clinical Observer Performance Study

3.1 Introduction

The x-ray beam is part of the electromagnetic (EM) spectrum. The EM radiation, including the x-ray beam, can be characterized by the theory of Wave-Particle Duality. The x-ray beam shows the wave characteristic when it interacts with an object with a similar dimension as its wavelength, and it shows the particle characteristic when interacting with an object much smaller than its wavelength.

Conventional attenuation-based radiography mainly utilizes the particle-like behavior of x-ray photons, where the primary x-ray photons are either scattered or absorbed by the various part of the body at different magnitudes. Contrarily, the phase-contrast imaging (PCI) systems utilize both wave-particle characteristics of x-ray energy by detecting the x-ray attenuation and its phase shift after it passes through the object. Several PCI techniques exist that can be classified into five major categories: a) propagation-based imaging systems (PBI),[87] b) Analyzer-based imaging systems (ABI),[88] c) interferometric methods based on crystal utilization,[89] d) the grating interferometric (GI) method,[90] and e) grating non-interferometric method.[91]

The propagation-based phase-contrast imaging system is the most straightforward and practical PCI technique in clinical settings, which does not require the highly monochromatic plane-wave x-ray beam and sophisticated x-ray optics. When we utilize the polychromatic x-ray source with relatively high spatial coherence and place the digital detector far enough from the object, the propagated beam has interferences of the

sub-waves and will result in image contrast, rising from both the x-ray attenuation and phase.

The schematic illustration of the in-line phase-contrast imaging system is shown in Figure 3.1. The R_1 denotes the source to object distance (SOD) and the R_2 is the sample to detector distance ($SID - SOD$), and the system utilizes the microfocus x-ray tube to provide a partial coherent x-ray beam.

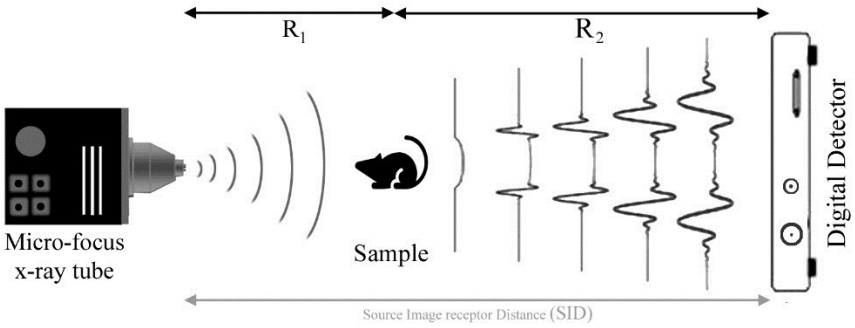


Figure 3.1: The Schematic diagram of in-line PCI experimental setup. The object is placed in a magnification geometry, M= $\frac{R_1+R_2}{R_1}=\frac{SID}{SOD}$.

In the acquired image under the experimental setup shown in Figure 3.1, the x-ray absorption and the phase shift are tangled. One can obtain the phase-contrast image by utilizing either the digital detector or film/screen detector. In the phase-contrast imaging system, the qualitative information from the tissue structures provides contrast and edge enhancement. On the other hand, the phase-retrieved image, acquired only by a digital detector, may also offer quantitative information about the tissue electron densities that are quite different in normal and cancerous breast tissues. In the phase-contrast image, the edge enhancement increases with the Laplacian, and the gradient of the phase shift, $\phi(x,y)$,[32] and only one x-ray exposure is required for the image acquisition.

The general x-ray transport equation can be written as follows:[46]

$$I(\overrightarrow{r_D})=\frac{I_0}{M^2}\left\{cos\left(\frac{\lambda R_2}{4\pi M}\nabla^2\right)\left[A^2(\vec{r})-\frac{\lambda R_2}{2\pi M}(\nabla.(\emptyset(\vec{r})\nabla A^2(\vec{r})))\right]-2\,sin(\frac{\lambda R_2}{4\pi M}\nabla^2)[A^2(\vec{r})\emptyset(\vec{r})]\right\} \quad (33)$$

where λ is the average x-ray wavelength, I_0 is the entrance intensity at the object plane, M is the geometric magnification, $\emptyset(\vec{r})$ and $A^2(\vec{r})$ are the phase shift and attenuation, respectively.

According to Equation 33, two different x-ray exposures with varying R2 are theoretically required to quantitate both the $A^2(\vec{r})$ and $\emptyset(\vec{r})$. However, this might be an undesirable solution in patient studies due to the health concern of unnecessary additional radiation. Hence, a significant challenge would be finding the phase retrieval method that does not need a double exposure but estimate the phase shift map with adequate precision. One of the most pragmatic approaches to this challenge is utilizing the principles of Phase-shift Attenuation Duality (PAD) theory.

In light elements, the X-ray-Matter interaction for the x-ray beam of approximately 60 kV to 500 kV is dominated by incoherent (Compton) scattering, where the photoelectric absorption and coherent scattering considerably diminish. Therefore, the x-ray attenuation and the phase shift are governed by the electron density of the object. This correlation between $A^2(\vec{r})$ and $\emptyset(\vec{r})$ is known as phase-attenuation duality in light elements.[50,92]

The air gap in in-line phase-sensitive imaging eliminates the grid requirement in conventional mammography. Low radiation doses can be delivered to the patient by grid removal without compromising the imaging quality due to noise related to the scattered photons. Additionally, the phase shift coefficient decreases much slower than the attenuation coefficient by increasing the x-ray energy.[29] Therefore, the x-ray energy might be increased to the spectrum we do not use in conventional breast imaging. Higher x-ray energy would indirectly enable further dose reduction because the signal-

contributing photons relatively increase compared to the non-signal-contributing photon, which deposited all their energies in the breast tissue.

The phase attenuation duality applies to the soft tissue at x-ray energy greater than 60 KeV. However, the x-ray tubes utilized in clinical applications are polychromatic sources, and the effective x-ray energy depends on the kVp and the external filtration. The higher the polychromatic x-ray energy is, the more reliable the phase retrieval would be based on the single projection PAD-based method. The in-line high-energy phase-sensitive imaging has thoroughly been investigated to determine the performance and detectability of the system in breast-phantom studies. The results showed improved detectability with reduced radiation dose.[39,43,93]

Nonetheless, the increased kVp compromises the image contrast given by the attenuation differences. Therefore, it was hypothesized that utilizing the mid-energy x-ray beam would increase the attenuation-based contrast with a partial impact on the single-projection PAD-based phase retrieval method. The introduced error for the phase mapping is expected to be trivial in low atomic number materials (i.e., breast tissue with Z_{eff}=7.4[94] and when the x-ray energy is not too low. This study has been published, and details can be found here.[51]

3.2 Materials and Methods

3.2.1 Imaging prototype:

A prototype imaging system based on the in-line phase-contrast principle has been employed in this study. A Tungsten target microfocus x-ray tube that can provide a current of up to 300 μA was used to produce a partially coherent x-ray beam. The image acquisition is performed under mid-energy and high-energy beams of 60 kV and 90 kV to

evaluate the quantitative and qualitative performance of the in-line phase-contrast system by a phantom study. A flat panel digital detector, providing a pixel pitch of 50 μm, is placed at R2 equals 32.5 inches. The Source to object distance (SOD) was 27 inches, which renders the geometric magnification of 2.2. Table 3.1 summarizes the key characteristic of the in-line phase-contrast prototype system. The phase contrast and phase retrieved images are evaluated both quantitively and qualitatively.

Table 3.1: The key characteristic of the mid-energy and high-energy prototype systems.

Component	Characteristic
X-ray source	Target: Tungsten. Entrance Window: 0.5 mm Beryllium. Output power: 18W and 27W for 60Kv and 90kV, respectively. The focal Spot size of ≈27 and ≈37 μm for mid-energy and high-energy beams, respectively. External filtration: 1.2 mm and 2.5 mm of Aluminum for 60 kV and 90 kV beams, respectively.
Detector	Direct Flat Panel Detector with Gadolinium oxysulfide (GOS) scintillator. Pixel pitch: 50 μm. X-ray tube voltage: 20 to 150 kV. The number of active pixels: 2316(H)× 2316 (V). Top cover material: 1.0 mm Carbon fiber. Resolution: 8 lp/mm at spatial frequency of CTF = 5%.
Geometry	R_1: 27″ (68.6 cm), R_2: 32.5″ (82.55 cm), M: 2.2.
Image Acquisition	Mid-Energy: 60 kVp, 7.5 mAs, 0.9 mGy. High-Energy: 90 kVp, 6.5mAs, 1.2 mGy. PAD-based phase-retrieval.

3.2.2 Dose Calculation:

The x-ray spectra for 60 kV and 90 kV filtered by1.2 mm and 2.5 mm aluminum filters, respectively, were measured by the CdTe x-ray detector. The normalized glandular dose coefficient values (D_{gN}) were approximated using the Monte-Carlo simulation discussed in detail previously,[95-96] considering multiple parameters (e.g., x-ray spectrum, HVL, breast thickness, and composition). The entrance exposure (X_{ESE}) at the

object plane is measured using the Leakage and Low-Level Measurements Ionization Chamber (10X6-180, Radcal®, Monrovia, California). The following equation estimates the Average Glandular Dose (AGD):

$$AGD = DgN \cdot X_{ESE}. \tag{34}$$

The hypothesis for this study was the outperformance of the mid-energy system, as the attenuation-based contrast would be more enhanced at the lower x-ray energies. Therefore, the delivered radiation dose in the mid-energy system was slightly less than the delivered dose in the high-energy system.

3.2.3 Imaging Objects:

This study aims to evaluate the quantitative and qualitative differences in pre-clinical settings. The phase retrieved image provides the phase shift map, rising from the electron density within the object. Contrarily, the phase contrast image shows the edge enhancement that increases with the Laplacian and the gradient of phase shift by a relatively coherent x-ray beam. A four mm-thick laser-cut acrylic slab with a sharp edge is used to investigate the impact of the two imaging systems on the Edge Enhancement Index (EEI). The EEI measures the degree of edge enhancement relative to the absolute change in the pixel intensity of the acquired image.[42] The EEI is given as:

$$\mathrm{EEI} = \frac{(\mathrm{P} - \mathrm{T})/(\mathrm{P} + \mathrm{T})}{(H - L)/(H + L)} \tag{35}$$

where P and T are the peak and trough intensities at the edge, and H and L are the average intensities at the high and low-intensity regions next to P and T.

The CDMAM phantom (Type 3.2) is used to perform the objective and subjective analyses to evaluate the detectability of the two imaging systems. The CDMAM phantom was initially developed for quality control in conventional mammography systems.

Unlike other mammographic phantoms, which are primarily used for quality control, observation bias is eradicated in this phantom due to the unknown location of the disk in each cell. Additionally, the numerous sizes and thicknesses of gold discs allow the investigator to evaluate signal detection more extensively. The CDMAM phantom has a Plexiglass equivalent thickness of 10 mm. Therefore, two 50-50 glandular-adipose equivalent CIRS slabs of two cm are added distal to the CDMAM phantom to simulate the five cm-thick compressed breast tissue. Due to magnification, the detector does not cover the entire area on a single projection. For this reason, 14 images are acquired to thoroughly scan the entire area of the phantom with varying disc diameters and thicknesses. The size of gold discs embedded in this phantom varies between 0.10 to 3.20 mm, where the contrast arises from disc thicknesses of 0.05 μm up to 1.60 μm.

In addition to the CDMAM phantom, the Gammex 156D Stereotactic Mammographic Accreditation Phantom is also used to evaluate the impact of the two imaging systems on detecting different abnormal anatomic structures in the breast tissue like neoplastic masses and microcalcifications. The phantom consists of four fibers (0.4 to 0.93 mm) mimicking the breast fibrous structure, four speck groups (0.2 to 0.54 mm) mimicking the microcalcifications, and four tumor simulating masses with varying diameters of 0.25 to 1.0 mm.

3.2.4 Observer study:

Three experienced observers are asked to review the images of CDMAM and ACR accredited phantom. All images are randomly mixed and unlabeled to establish a blind study. The reader study uses a high-resolution monochrome mammography display (Sony LMD-DM50) with a pixel size of 165 μm.

The correction scheme explained in the phantom manual[97] is applied after I record the reading results for each observer. For the correction scheme, there are two main rules: a) A *True* observation needs two or more correctly indicated nearest neighbors to remain a *True*, and b) A *False* or *Not indicated* disc will be considered as a *True* when it has 3 or 4 correctly indicated nearest neighbors. Three exceptions are also adopted to the correction scheme as follows: a) A *True,* which has only two nearest neighbors at the edge of the phantom, only needs *one* correctly indicated nearest neighbor to remain a *True*, b) A *False* or *Not indicated* disc that has only two nearest neighbors will be considered a *True* if both nearest neighbors are correctly identified, and c) the absent corners of phantom (0.05 µm/3.20 mm and 1.60 µm/0.10 mm) will be marked as a *True* when both nearest neighbors are correctly indicated. The readers are also asked to follow the rules defined by the Mammography Quality Standard Act (MQSA) to report the image interpretation for the images of the Gammex 156D phantom.

For each imaging system, a Contrast-Detail (CD) curve is obtained based on the subjective study, and two evaluating figures of Correct Observation Ratio (COR) and Image quality Figure (IQF) are compared statistically. The specified figures are defined as follows:[97]

$$\text{COR}=\frac{\text{Correct Observation}}{Total\ number\ of\ cells} \times 100 \tag{36}$$

$$\text{and } IQF = \sum_{1}^{16} C_i \text{ x } D_{i,\,min} \tag{37}$$

where C_i is the corresponding thickness (contrast) of column i and $D_{i,\,min}$ is the minimum detectable disc diameter in the contrast column i. One can conclude the higher performance of a system when the correctly identified discs have smaller diameters and thicknesses compared with another system. Therefore, the smaller *IQF* would indicate

higher performance, but the inverse image quality figure (IQF_{inv}) is commonly used for intuitive conception. The imaging system with greater IQF_{inv} can be considered more desirable in terms of detectability.

As stated in Chapter 2, the object visualization by the human eye is directly related to the quantitate qualities in the acquired images by SDT principles. For objective comparisons, the Contrast to Noise Ratio (CNR) is calculated for both systems according to the Rose model defined as below:[98]

$$CNR = \frac{\bar{I}_A - \bar{I}_B}{\sqrt{\frac{(\sigma_A^2 + \sigma_B^2)}{2}}} \times \sqrt{\frac{a_D}{a_p}} \tag{38}$$

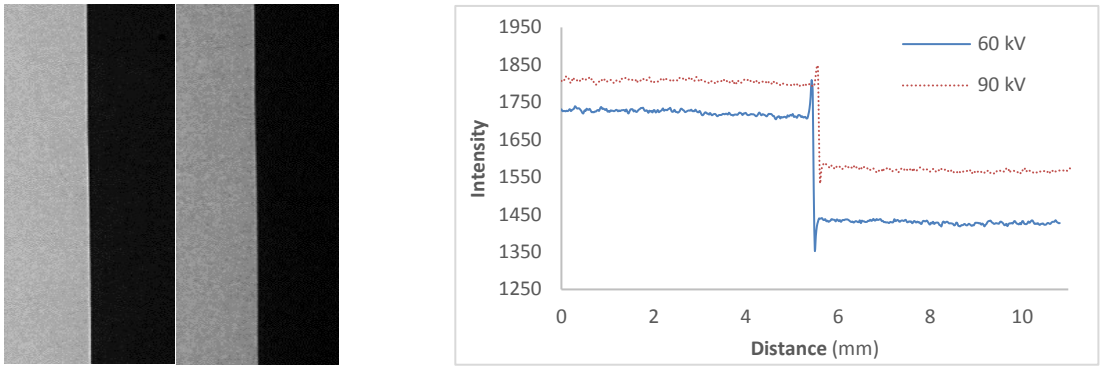
where $\bar{I}_A$ and $\bar{I}_B$ are the mean pixel intensities over the target region and the background over the same size area,σ_A^2 and σ_B^2 are the corresponding pixel value variances, a_D and a_p are the target region area and the pixel area, respectively.

Because the radiation dose was reduced in the mid-energy imaging system compared with the high-energy system, the dose-independent Figure Of Merit (FOM) shall be employed for fair quantitative comparisons. The CNR in quantum-limited x-ray detectors is related to the number of x-ray photons (N) by the noise term $\sqrt{N}$ and the radiation dose is linearly proportional to $"N"$. Therefore, by squaring the numerator, which indirectly depends on $\sqrt{N}$. The influence of *N* would be eliminated in comparative analyses. The FOM for quantitative comparisons is then defined as follows:[99]

$$FOM = \frac{CNR^2}{Dose} \tag{39}$$

3.3 Results:

Figure 3.2 shows the images of the laser-cut acrylic edge acquired by both mid-energy and high-energy systems and the edge profile at the acrylic-air boundary.



(a) (b)

Figure 3.2: a) *Phase contrast image of Acrylic edge acquired at 60 kV (left) and 90 kV (right). b) Edge profiles of two different kVs are shown on the right.*

From visual inspection, the higher overshooting in the edge profile for the mid-energy imaging system will result in a higher degree of edge enhancement on the phase contrast image acquired at an x-ray energy of 60 kV. The quantities analysis for the edge enhancement index (EEI) also confirmed the higher edge enhancement by x-ray energy of 60 kV compared with the image acquired at 90 kV. The estimated EEIs for the images of acrylic edge phantom were 1.53 and 1.32 for mid-energy and high-energy systems, respectively.

Figure 3.3 compares the threshold contrast detection for the phase-contrast images acquired at x-ray energies of 60 kV and 90 kV. The average contrast threshold detected by three observers shows that the mid-energy system detects lesions with less contrast than the high-energy system at any specified lesion size.

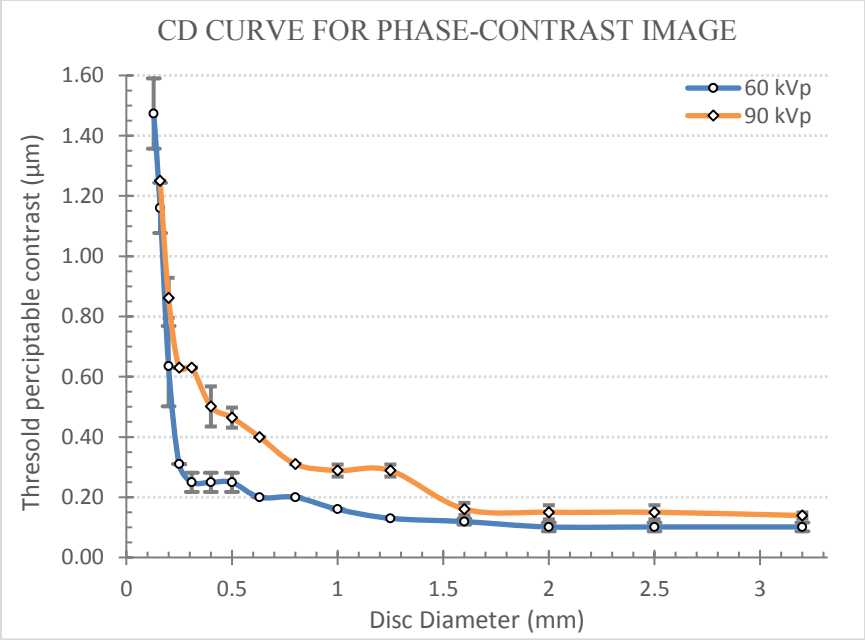


Figure 3.3: C-D curve generated from the reading results of CDMAM phantom by three experienced observers. Both 60 kV and 90 kV images are phase-contrast images. The error bars in the figure represent 95% CI with df=2.

Similarly, the phase-retrieved image of the CDMAM phantom acquired by the mid-energy system reveals the discs with less contrast at any given diameter, except in the region with a disc size smaller than 0.2 mm. Higher radiation dose, and consequently the higher SNR, and the higher accuracy of the PAD-base phase-retrieval method at higher x-ray energies might justify this crossed observation. However, the C-D curve for the phase-retrieved image shows higher performance for the dose-saving mid-energy system, albeit within the clinically significant spatial resolution range. Figure 3.4 shows the C-D curve generated for the phase-retrieved images.

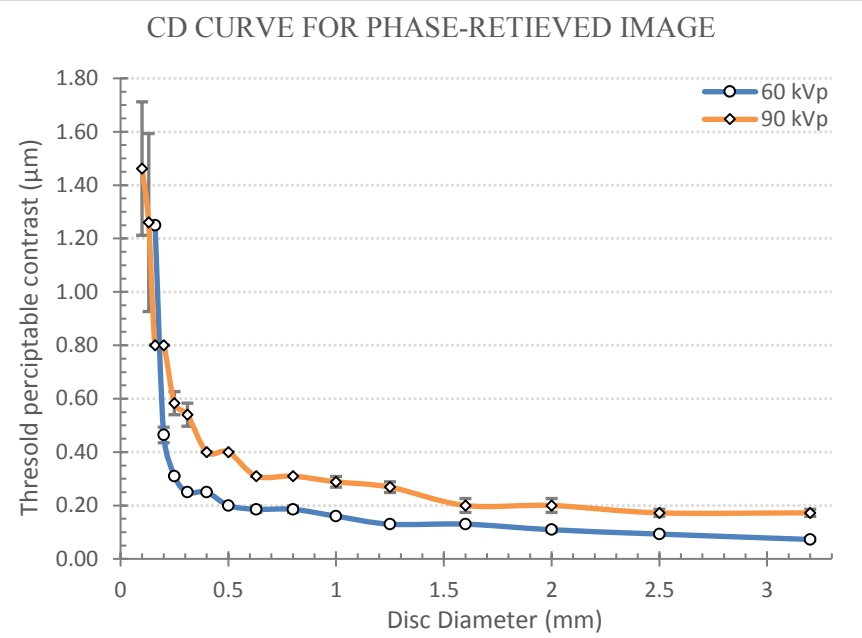


Figure 3.4: C-D curve generated from the reading results of CDMAM phantom by three experienced observers. Both 60 kV and 90 kV images are phase-retrieved images. The error bars in the figure represent 95% CI with df=2.

The average identified disc numbers on phase-contrast images are 116.66 and 84.0 with 60 kV and 90 kV, respectively. The calculated COR for each system reveals the statistically significant difference between the two imaging systems (COR_{60kV}= 56.91%, COR_{90kV}=40.97%, df=2, p-value=0.002). Similarly, there was a significant difference between the two phase-retrieved images acquired by mid-energy and high-energy systems (COR_{60kV}= 57.88%, COR_{90kV}=43.41%, df=2, p-value=0.0007). The calculated IQF_{inv} for both phase-contrast and phase-retrieved images are greater in the mid-energy imaging system, and the differences are statistically significant. Table 3.2 tabulates the outcomes of the observer performance study performed by three observers on the CDMAM phantom.

Table 3.2: The quantitative figures of observer performance study on CDMAM phantom.

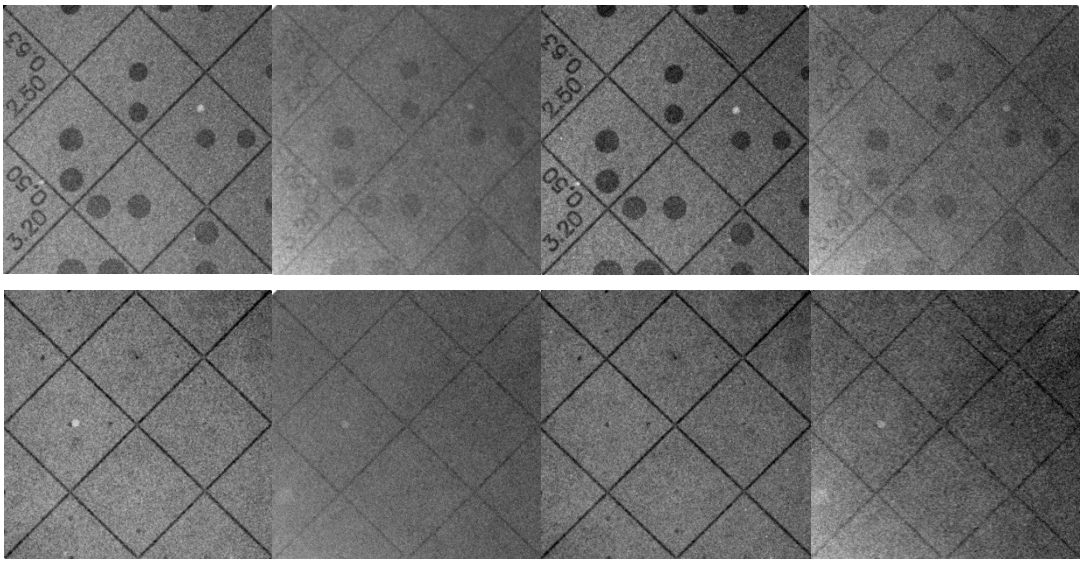
	60kV (0.9 mGy)	**90kV** (1.2 mGy)	**P-value**
COR $_{\text{Phase-contrast}}$	56.91	40.97	0.002
IQF_{inv} Phase-contrast	0.370	0.239	0.002
COR $_{\text{Phase-retrieved}}$	57.88	43.41	0.0007
IQF_{inv} Phase-retrieved	0.389	0.244	0.0004

As discussed earlier, human perception indirectly follows the quantitative image qualities such as CNR or SNR. Although the C-D curve provides the concept of the spatial and contrast resolutions in an imaging system, quantitative analyses like CNR is independent of human perception and related inter/intra-observer variabilities. Two gold discs are arbitrarily selected for quantitative analyses. One of them is a larger disc with relatively lower contrast, and another is a smaller disc with higher contrast than the other one. Due to dose differences, equation seven is used to compare the results independent of the delivered radiation dose. Table 3.3 shows the calculated CNR and FOM for each disc acquired by mid-energy and high-energy imaging systems.

Table 3.3: quantified CNRs and FOMs from two discs with varying diameters/thicknesses in the CDMAM phantom calculated for the mid-energy and high-energy images.

	CNR_{60kV}	CNR_{90kV}	FOM_{60kV}	FOM_{90kV}
Disc Size 2.5 mm×0.5μm	3.47	1.60	13.38	2.13
Disc Size 0.8 mm×0.63μm	4.06	1.8	18.32	2.70

Figure 3.5 shows the phase-contrast and phase-retrieved images of two ROIs from the CDMAM phantom, containing the gold discs, used for quantitative CNR and FOM analyses.



(a) (b) (c) (d)

Figure 3.5: The phase-contrast images of two ROIs from CDMAM phantom: a) acquired by the mid-energy system and b) acquired by the high-energy system. The phase-retrieved images of the same ROIs: c) acquired at 60 kVp and d) acquired at 90 kVp.

An ACR accredited mammography phantom is also used in this study to evaluate the two imaging systems objectively and subjectively. Table 3.4 shows the average reading scores for the specific features embedded in the phantom.

Table 3.4: Reading scores for each lesion type in the ACR accredited mammography phantom. Three experienced observers (df=2) scored the images based on the MQSA scoring guidelines. Each category has four lesions with varying sizes and contrasts.

		60kV *(0. 9 mGy)*	**90kV** *(1.2 mGy)*	**P- value**
Masses	Phase-Contrast	3.50	3.33	0.41
	Phase- Retrieved	3.66	3.33	0.23
Fibers	Phase- Contrast	3.66	3.33	0.22
	Phase-Retrieved	3.66	3.0	0.06
Specks	Phase- Contrast	3.16	2.33	0.11
	Phase- Retrieved	3.16	3.0	0.11

The phase-contrast images of the ACR accredited mammography phantom, acquired at x-ray energies of 60 kV and 90 kV, are shown in Figure 3.6. Like CDMAM image

acquisition parameters, the mid-energy system delivers less radiation dose (0.9 mGy) than the high-energy system (1.2 mGy).

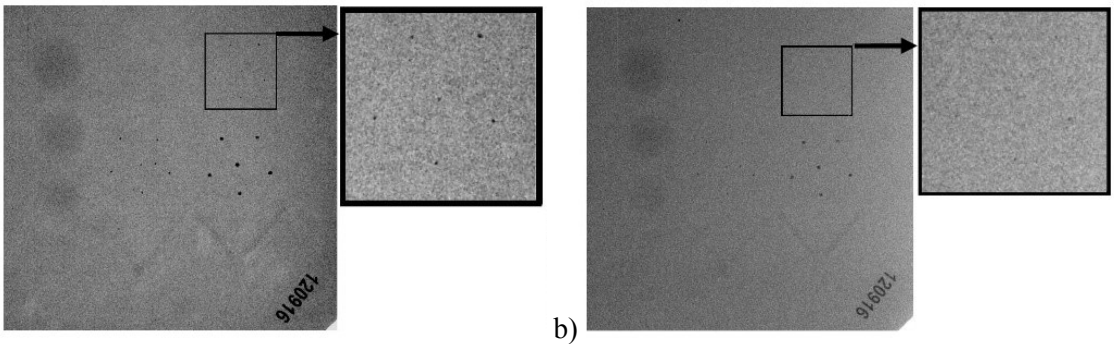


Figure 3.6: Phase-contras images of ACR-156D Phantom acquired by a) the mid-energy system and b) the high-energy system.

3.4 Discussion:

The radiation dose is a big concern in clinical imaging because the breast is one of the most sensitive organs to ionizing radiation. The phase-contrast imaging modality can reduce the radiation dose, and unlike conventional mammography, it is not necessary to utilize the x-ray in the lower-end energy spectrum. However, phase retrieval shall be performed to map the phase changes when the x-ray beam passes through different structures with varying electron densities to maximize the contrast information. The PAD-base phase-retrieval is one of the most practicable approaches that can be done by only one projection in non-grating-base phase-sensitive x-ray imaging systems. PAD-base phase-retrieval method can be applied in higher x-ray energies where the phase-attenuation duality exists.

On the other hand, utilizing the high x-ray energy obliterates the attenuation-base image contrast. Many pre-clinical studies have shown the potential advantages of the in-line high-energy phase-sensitive imaging system. However, there was a hypothetical expectancy that if we utilize relatively lower x-ray energy, we might be able to increase

the attenuation-base image contrast while the PAD-base retrieval method does not get impacted extensively. The idea for the mid-energy phase-sensitive imaging instead of the high-energy system arose because the introduced error in phase shift mapping might be negligible in light elements such as soft tissue, especially when the x-ray energy is not too low.

The results show that mid-energy phase-contrast images of the CDMAM and ACR phantoms provide higher detection scores and image quality than images acquired at higher x-ray energy. Although the phase retrieval in a higher x-ray energy system slightly performs better than in the mid-energy system, the results showed that the detectability and quantitative figures are still higher in phase-retrieved images acquired by the mid-energy system. It is important to note that although the accuracy of quantitative values for phase shift, determined by the PAD-base method in mid-energy systems (effective x-ray energy less than 60 kV), carry slightly erroneous estimation, the objective and subjective results still demonstrate the potential advantage of mid-energy phase-sensitive breast x-ray imaging.

This important finding may lead to employing the mid-energy x-ray beam in clinical pilot studies rather than higher x-ray energies, studied more in the pre-clinical studies.

As shown in Figure 3.2, overshooting is more prominent in the image acquired by the mid-energy system for the phase-contrast image. The intra-tissue heterogeneity is usually an indicator of possibilities of malignancy within the breast tissue. Therefore, the higher EEI will expand the capability of the radiologists to detect the potential intra-tissue heterogeneity.

Interestingly, the image of CDMAM phantom acquired at 60 kV with AGD equivalent to 0.9 mGy resolves smaller and thinner discs than high energy x-ray system, both in phase-contrast and phase-retrieved images. The calculated inverse Image Quality Figure (IQF_{inv}) confirms this statement. Similar scores (if not higher) for the mid-energy system are also recorded for all lesion types in the ACR accreditation phantom compared to the high-energy system's image. It is essential to acknowledge that the high-energy system delivers a slightly higher radiation dose than the mid-energy system.

This study mainly focuses on the sensitivity of two imaging systems, but the performance of an imaging system depends not only on the sensitivity but also on the specificity of the imaging system. In the observer study, the readers either located the gold disc in each CDMAM phantom cell or could not identify the gold discs with small diameters or lesser thicknesses. Due to the lack of false-positive reading, specificity evaluation was not feasible. Therefore, this study only investigates the improved performance merely in the system’s detectability aspect. The earlier studies have clearly shown the advantages of the high-energy in-line phase-contrast imaging systems over conventional imaging. This study has added a claim to the phase-contrast imaging techniques that a minor error in the PAD-base phase-retrieved images is not significant enough to adversely impact the detectability of the system operating at the mid-energy range. In fact, the acquired images by mid-energy x-ray beam in phantom studies showed the potential advantages compared with the high-energy system. The ACR phantom has a limited level of lesion size and contrast.

Consequently, extensive analyses of the system’s detectability might be challenging using these phantoms. On the other hand, the CDMAM phantom provides several layers

of spatial resolutions and contrast resolutions, but the embedded discs in the CDMAM phantom are made of gold with a higher atomic number. The accuracy of the PAD-based phase retrieval method might be impacted further when the imaging object has a relatively higher atomic number compared with breast tissue. Therefore, it is expected that the mid-energy system would be even more preferred in soft tissue imaging. Further investigations with a more suitable phantom that does not have high Z materials but has the same design as the CDMAM phantom will be helpful in future studies.

Chapter 4. The Comparison of the In-line Phase-Sensitive and Conventional Breast Tomosynthesis, utilizing the ROC paradigm: An Observer Performance Study.

4.1 Introduction

Unlike the objective analyses, which are reader-independent, the subjective evaluations are reader-dependent investigations. The physician is the end-user in medical imaging, and a human decides on the appropriate diagnosis. One may expect that the radiologists will better diagnose the disease of interest if the acquired image provides a higher Contrast-to-Noise Ratio. However, in clinical settings, patient-related conditions also play a role in diagnostic accuracy. Therefore, the observer studies could provide valuable insight into the performance of an imaging system and potentially be more realistic than a pure objective comparison.

The simplest form of a human observer study in radiology began in the late 1940s using Rose-Burger test phantoms, designed to generate the contrast-detail curves.[100] The CD diagram plots the contrast signal as a function of the signal corresponding to the size of the object of interest. However, decades later, medical physicists became interested in applying the signal detection theory, which was used to evaluate radar signals' detectability in the presence of various random noises. The Receiver Operating Characteristic (ROC) might be the most appropriate method to evaluate the observer detection performance based on SDT.

Goodenough et al. at the University of Chicago Radiology Department showed in the early 1970s that the ROC method might be suitable for evaluating signal detection in a

quantum-noise-limited image.[101] Nonetheless, in clinical images, the tissue structural variations from patient to patient (patient noise) dominate over the quantum noise and do not have random noise properties. The SDT analyses could be theoretically compromised due to these challenges in clinical settings.

However, on an empiric basis, it was demonstrated that ROC analysis is valid for comparing relative performance and could be utilized to compare the observers' performance for different imaging modalities.[102] The variety of summary measures, including the Area Under the ROC Curve (AUC) and the detectability index (d_A) are used for the comparisons. d_A is a value obtained from AUC when transformed into a domain similar to the SNR.

The clinical ROC study can compare the variability in the radiologists' performance[103] or compare the diagnostic performance of multiple imaging modalities.[104] In an SDT analysis, the primary requirements are well-defined and rigorously controlled conditions. It should be noted that the Signal-Known-Exactly (SKE) and Background-Known-Exactly (BKE) decision tasks that are produced by a random noise with Gaussian probability density function are necessary controlled conditions. The phantom studies might be challenging because the imaging phantoms are mainly designed to evaluate the presence of a signal rather than the absence of the signal. Noting this challenge, I designed a ROC study utilizing a Contrast-Detail phantom to compare the performances of high-energy and mid-energy phase-sensitive imaging systems.

4.2 Materials and Methods

4.2.1 Imaging Phantom:

A custom-made 1-cm-thick CD phantom containing a matrix of 6×6 lesions in various diameters and depths is utilized to design the ROC study. The diameter of the disc in each column varies from 4.25 mm to 0.25 mm (4.25, 3, 2, 1, 0.5, and 0.25 mm), while six different contrasts for each diameter are produced by varying depths of drilled discs from 1 mm to 0.1 mm (1, 0.8, 0.6, 0.4, 0.2, and 0.1 mm) in each row. The phantom is placed between the heterogeneous tissue-equivalent slabs in two different experiments. In the first experiment, the CD phantom is placed between two 50% glandular-50% adipose tissue-equivalent heterogenous slabs, with thicknesses of 2 cm each, to simulate the 5cm compressed breast. In the second experiment, three 70% glandular-30% adipose tissue-equivalent heterogenous slabs, with thicknesses of 2 cm each, are used to simulate the 7 cm compressed hyperglandular breast. The heterogeneous breast tissue-equivalent slabs are fixed in each experiment as the pattern over the drilled discs stays the same in both DBT and PBT images.

4.2.2 Imaging Acquisition Systems:

The commercial Digital Breast Tomosynthesis device (*Selenia Dimension*) acquired the DBT images in Automated Exposure Control (AEC) mode. The anode target was Tungsten (W), and an external Aluminum filter of 0.7 mm was utilized for the beam hardening. Both thinner and thicker phantom images are acquired at an x-ray energy of 31 kV. On the other hand, the prototype PBT system with Tungsten anode and fixed external filtration of 2.5 mm-thick aluminum is used to acquire the phase-sensitive imaging. The PBT image of a 5-cm-thick phantom is acquired at 59 kV, while the PBT

system is operated at 89 kV to acquire the image of a 7-cm-thick dense phantom. The delivered radiation doses for thinner and thicker phantoms are 1.58 and 2.43 mGy by the commercial DBT system, respectively. The DgN values are estimated based on the x-ray spectra of 59 and 89 kVs and the object thicknesses to estimate the AGD in the PBT system. The entrance exposures are measured at any tube voltage, and the AGD is set to a similar value as the DBT system in each experiment using Equation 34. Therefore, the delivered radiation doses in both experiments for DBT and PBT images are identical. Figure 4.1 demonstrates the PBT prototype system utilized for phase-sensitive image acquisition under specified parameters.

Figure 4.1: Phase Contrast Breast Tomosynthesis system (Hologic, Inc. MAN-05666, Revision 002, Marlborough, MA, U.S.A.)

4.2.3 Observer Study:

Five experienced observers participated in this study. The observers are non-medical professionals but have extensive experience reading pre-clinical x-ray imaging. The phase-retrieved image of the 2×2 binning mode acquired by the PBT prototype system is used for comparison. The in-focus slices from DBT and PBT images are used to create the multiple ROIs for the study. As previously discussed, the CD phantom has 36 discs of varying sizes and thicknesses. 36 ROIs are cropped from the location where the discs are present, so each ROI has one signal-simulating disc. Also, 36 ROIs are cropped from the areas on the images where the discs are not present to simulate the non-diseased samples. The cropping location for the non-diseased sample is fixed for both DBT and PBT images.

The target disc in each cropped ROI could be located anywhere, and the location of the discs is unknown to the observers. The non-diseased samples are prepared from the exact locations in both images to guarantee that any false positive report due to heterogeneity of the non-diseased ROI has the same possibility of being reported as a false positive in both imaging systems.

Thus, 288 ROIs are produced and randomly arranged to conduct the blind observer study. The observers are unaware of the utilized imaging system or phantom type throughout the study. The observers are asked to review the images in a dark room on a 5 MP grayscale radiology monitor (SONY LMD DM50) and identify the diseased and non-diseased cases. Each observer records the results using a 10-scale reporting system. The observers are asked to report a score of 10 when the disc is entirely visible in ROI with a clear and complete circular margin. The observer shall report the score of 1 if he/she is

confident that the ROI does not contain the disc. The scores between the lowest and the highest score depend on the observers' interpretation. The scores lower than five tend to be interpreted as negative cases, and a score equal to or greater than five is considered a positive case.

4.2.4 Analysis of the Observer performance Study:

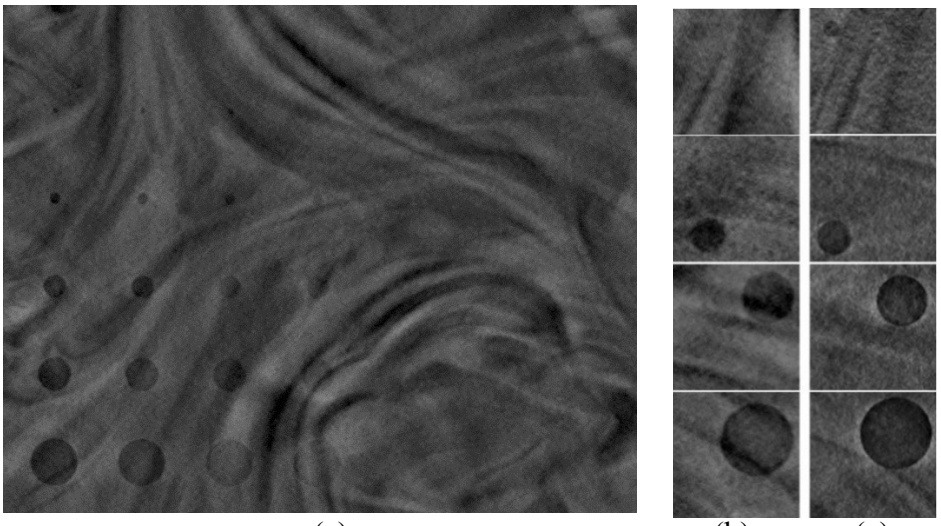
The reading results are evaluated by ROCKIT 1.1B2 Beta version (Developed by Dr. Metz, Department of Radiology at the University of Chicago), employing the Multi-Reader Multi-Case (MRMC) DBM method. The statistical analysis is performed by a univariate z-score test of the difference between the area under the two ROC curves and a univariate z-score test of the difference between the TPFs on the two ROC curves at a selected FPF. A fully paired (correlated) test is performed on the acquired dataset as the reviewed ROIs are extracted from the exact location and pattern in both DBT and PBT images.

4.3 Results:

The 288 cropped ROIs from a contrast detail phantom are read by five observers. One-half of the ROIs had a disc with varying sizes and contrasts, and one-half were extracted from the background with no disc present in the ROI. The location of the disc on each ROI is unknown to the observers, and so are the imaging modality and image acquisition parameters.

The observers report the likelihood of disc being detected on each ROI on a ten-scale reporting system. The 288 ROIs are randomly presented in one session, and the observers could zoom the ROI in/out without other image modifications like window/level operation.

Figure 4.2 shows the image of the CD phantom acquired by the commercial DBT system and a few extracted ROIs from both DBT and the PBT systems, reviewed by the observers for the ROC analysis.



(a) (b) (c)

Figure 4.2: The images of the five cm-thick CD phantom with 50% Glandular and 50% Adipose tissue composition. a) the Full Field DBT view on the focal plane, b) four ROIs extracted from the DBT image, and c) the corresponding ROIs extracted from the image acquired by the prototype PBT system.

The parametric Area Under the ROC Curves estimated for the 50% glandular phantom of 5-cm-thickness for DBT and PBT systems are 0.816 and 0.881, respectively. The non-parametric ROC evaluation reveals the AUCs of 0.802 and 0.867 for DBT and PBT imaging systems, respectively. Figure 4.3 shows the parametric and non-parametric ROC curves, comparing the diagnostic performance of the DBT and PBT systems.

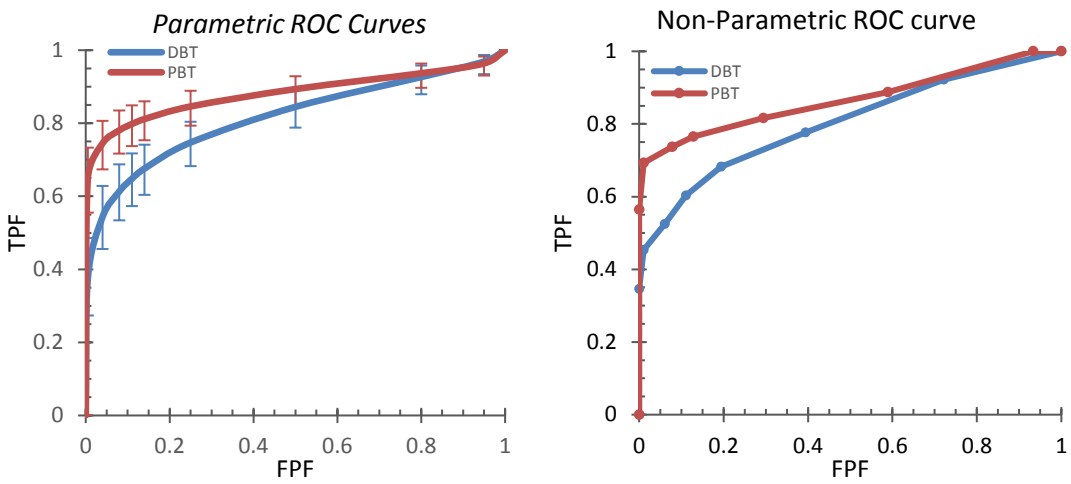
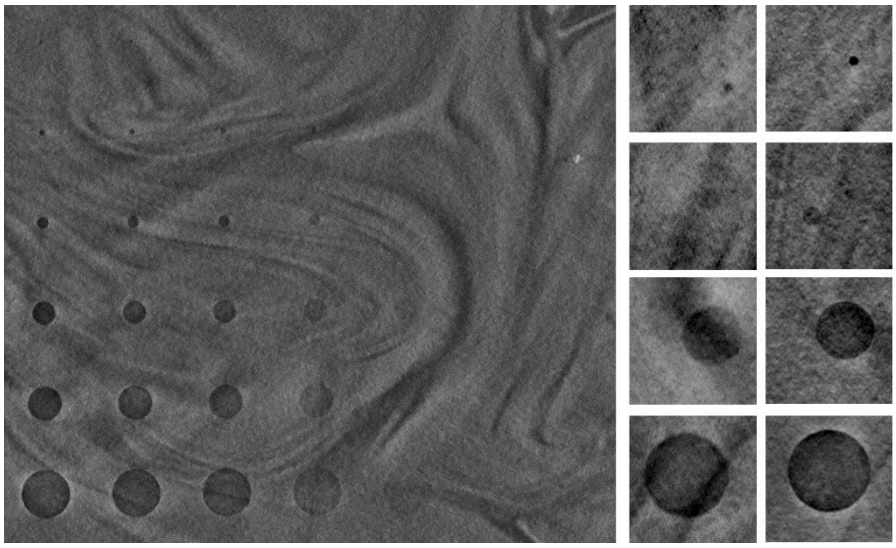


Figure 4.3: The parametric and non-parametricROC curves generated from the reading data of a 5-cm-thick phantom with 50% glandular composition. AGD=1.58 mGy in both systems, and the x-ray energies are set to 31 kV and 59 kV in DBT and PBT systems, respectively. The error bars show the 95% Confidence Interval for parametric curves, df=4.

The analysis of the difference between the two correlated ROC curves, utilizing the Binormal ROC parameters, shows the statistical significance (two-tailed P-value= 0.001, approximate 95% CI for AUC_{DBT}-AUC_{PBT} = [-0.103, -0.026], df=4).

Similarly, the results from the ROC study, performed on the thick and hyperglandular phantom, showed that the PBT system outperformed the conventional mammography system. Figure 4.4 shows the full-field image of the 7-cm-thick phantom with 70% glandular composition and a few extracted ROIs from the acquired images by the DBT and PBT systems, reviewed by the observers for the ROC analysis.



(a) (b) (c)

Figure 4.4: The images of the 7-cm-thick CD phantom with 70% Glandular and 30% Adipose tissue composition. a) the Prototype PBT image of CD phantom on the focal plane, b) four ROIs extracted from the DBT image, and c) the corresponding ROIs extracted from the PBT image.

Figure 4.5 shows the parametric and non-parametric ROC curves generated from the reading data collected from the five observers. The parametric AUC for the PBT system (0.840) is greater than the parametric AUC for the DBT system (0.821). However, the statistical evaluation of the diagnostic performance is insignificant (two-tailed p-value=0.288, 95% CI for AUC_{DBT}-AUC_{PBT} = [-0.053, 0.016], df=4).

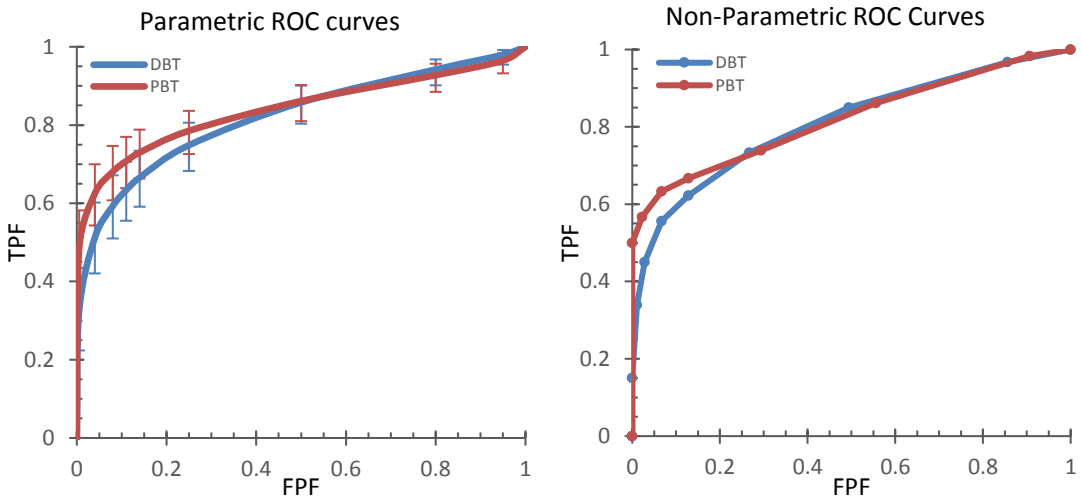


Figure 4.5: The parametric and non-parametricROC curves generated from the reading data of a 7-cm-thick phantom with 70% glandular composition. The images are acquired at AGD=2.43 mGy in both systems. The x-ray energies are set to 31 kV and 89 kV in DBT and PBT systems, respectively. The error bars show the 95% Confidence Interval for parametric curves, df=4.

Table 4.1 summarizes the results of the ROC studies performed on randomized ROIs, acquired by the mid-energy and high-energy phase-sensitive prototype system, compared with the conventional digital tomosynthesis. The estimated AGD on each condition is considered the same, while the commercial DBT system benefits from the advanced image processing and reconstruction algorithm, contrarily to the PBT system which the image processing techniques are not matured yet.

Table 4.1: The estimated key parameters of the ROC study and the statistical significance related to the performance of the clinical digital breast tomosynthesis and prototype phase-sensitive tomosynthesis, evaluated by the two imaging phantoms with different breast thicknesses and compositions.

	DBT image 5cm, 50G/50A	**PBT image** 5cm, 50G/50A	**DBT image** 7cm, 70G/30A	**PBT image** 7cm, 70G/30A
AUC	0.8162	0.8812	0.8212	0.8400
AUC (Wilc.)	0.8017	0.8667	0.8093	0.8207
Std. Err. AUC	0.0237	0.0200	0.0226	0.0221
Std. Err. AUC (Wilc.)	0.0233	0.0194	0.0229	0.0222
95% CI for AUC	(0.766, 0.858)	(0.837, 0.915)	(0.773, 0.862)	(0.793, 0.879)
Correlation of AUCs	0.6032		0.6871	
Two-tailed p-value	0.001		0.288	
95% CI for difference	(-0.1036, -0.0263)		(-0.0535, +0.0159)	

The partial areas under the ROC curve (pAUC), corresponding to the clinically relevant specificity value at 90% ranges for 5-cm-thick phantom, are estimated at 0.055 and 0.073 for DBT and PBT systems, respectively (F-value=2.19, df=4, p-value=0.14). Similarly, the pAUC values at a specificity of 90% for 7cm-thick phantom are estimated as 0.051 and 0.062 (F-value=6.50, df=4, p-value=0.011) for DBT and PBT systems, respectively. Table 4.2 summarizes the results for pAUC evaluation at the clinically significant territory of 90% True-Negative report (FPF=0.1) using the statistical analysis by random-reader and random-case.

Table 4.2: The estimated pAUCs (0,0.1) for clinical DBT and prototype PBT using two imaging phantoms with different breast thicknesses and compositions.

	DBT image 5cm, 50G/50A	**PBT image** 5cm, 50G/50A	**DBT image** 7cm, 70G/30A	**PBT image** 7cm, 70G/30A
pAUC	0.0546	0.0732	0.0514	0.0622
Std. Err. AUC	0.0084	0.0087	0.0090	0.0086
95% CI for AUC	(0.038, 0.071)	(0.056, 0.090)	(0.034, 0.069)	(0.045, 0.079)
Two-tailed p-value	0.0110		0.1406	
95% CI for difference	(-0.0329, -0.0043)		(-0.0252, +0.0036)	

4.4 Discussion:

The detectability of the system is the first-line evaluation to determine the performance of a newly developed imaging system. Either objective or subjective studies can investigate any medical imaging system's contrast or spatial resolutions. Although the system characterization takes different approaches to determine the system's performance, the Signal Detection Theory could correlate these two segregated analyses.

Nevertheless, the subjective evaluation might be slightly superior because the physicians are the end-users in medical imaging, at least for the time being. The patient noise and the background structures may result in an altered outcome in the detection capability of a human that cannot be justified or correlated to the SDT.

An additional concern in all diagnostic tests, including medical imaging, is the system's accuracy vs. the sensitivity alone. Identifying the suspicious structure as a potential origin of a disease is very important. Likewise, dismissing an individual who shows a questionable radiologic feature but is otherwise healthy might be as substantial as finding the diseased person so that the costly and invasive diagnostic procedures can be avoided. Therefore, the evaluation of the medical imaging system’s accuracy is also crucial as the contrast or spatial resolutions of the system are.

In real-world clinical settings, the participants in clinical trials may present a realistic radiologic feature that might be mistaken as a potential pathology. However, the imaging phantoms are mainly designed to evaluate the contrast or spatial resolution of the interested imaging device. Hence, to perform the ROC study that impartially examines the accuracy of a diagnostic test, some modification must be employed for system characterization in a pre-clinical stage. This study employs such a modification in a

phantom study to model a ROC study and compare the diagnostic performance of the newly developed phase-contrast breast tomosynthesis device with a clinical DBT system as the standard of care in breast cancer screening.

The results from the ROC studies, derived from the heterogeneous phantom study, revealed that the PBT system offers a significantly higher diagnostic performance than the clinical system when operated at the mid-energy x-ray range. On the other hand, the PBT system, operating at a higher x-ray energy range, results in a larger area under the ROC curve for the PBT system, but the difference between the two imaging systems is not statistically significant. The outcome is aligned with the conclusion drawn in the previous chapter of this dissertation. The mid-energy phase-sensitive imaging system better balances the attenuation-based image contrast and the precision of the PAD-based phase retrieval algorithm by the effective x-ray energy below 60 kV.

Many published articles show the potential advantages of phase-sensitive breast x-ray imaging in pre-clinical stages. However, it is vital to discern an unfair comparison in this investigation as the attenuation-based system acquires the images in fully optimized conditions, exploiting the rigorously advanced 3D image reconstruction, image processing, and artifact removal algorithms. In contrast, the prototype system's reconstruction algorithm, artifact removal, and image processing technique are not fully optimized yet. Nevertheless, the diagnostic accuracy of the PBT system at both operating x-ray energies yields promising results.

Further investigation is necessary to arbitrate whether the high-energy system would yield a statistically significant difference in diagnostic accuracy after the PBT system undergoes software and hardware advancement. Regardless of the utilized x-ray energy

beam, the PBT system with less-matured image processing results in desirable detection and relatively a higher accuracy. The future direction would be designing an observer performance study that investigates how far the radiation dose to the patients can be reduced without any adverse impact on the diagnostic accuracy of breast cancer screening compared with the attenuation-based breast tomosynthesis.

Chapter 5. The Comparison of the In-line Phase-Sensitive and Conventional Breast Tomosynthesis, utilizing the 2AFC paradigm: An Observer Preference Study

5.1 Introduction

Forasmuch as a human is involved in radiographic image interpretation, the observer's decision-making capability is crucial, and the visual perception of the images plays a reasonable role in the diagnostic performance of the imaging system. The ROC paradigm provides a comprehensive insight into the sensitivity and specificity of a diagnostic test. However, performing the ROC study is time-consuming and relatively expensive compared to conducting the observer preference studies such as alternative force choice studies.[105] Nevertheless, the purpose of the system evaluation would determine the most appropriate approach to perform an observer study.

For instance, the decision-making capability of an imaging system can be appraised by quantitative evaluations like SNR, CNR, Modulation Transfer Function (MTF), and Detective Quantum Efficiency (DQE). Also, qualitative investigations like the Contrast-Detail curve would reveal the system's potential in identifying the suspected lesions. However, we sometimes need to evaluate how comfortable/confident the observers are in image interpretation. Additionally, the *nAFC* studies are gaining interest in assessing the human observer performance of image-based decision tasks parallel to well-known and widely accepted ROC studies.[106]

The most applied n-alternative forced choice study would be designed by one alternative, containing a signal plus noise, and *n-1* alternatives, containing only noise.

The reading results will be analyzed based on the fraction of signal-containing images among the signal-free alternatives selected by multiple observers.

On the other hand, we compared the two different x-ray imaging systems operating by different principles. Therefore, the 2AFC study with a slightly tailored design would be applicable to the purpose of the current investigation.

5.2 Materials and Methods

5.2.1 Imaging phantom:

The same imaging phantom used for the ROC study in Chapter 4 is used for the two-alternative forced-choice study. The one-centimeter-thick CD phantom contains 36 discs, with varying sizes and thickness, placed in a 6×6 lesions matrix. The diameter of the discs in each row stays the same while the contrast decreases from left to right. Contrarily, the contrast of the discs stays the same in each column while the diameter increases from top to bottom. The varying diameters are 4 mm, 3 mm, 2 mm, 1 mm, 0.5 mm, and 0.25 mm, where the lesion contrast is produced by varying drilled depths of 1 mm, 0.8 mm, 0.6 mm, 0.4 mm, 0.2 mm, and 0.1 mm. In the first experiment, the CD phantom is placed between two heterogenous, two-centimeter-thick 50% glandular tissue-equivalent slabs. The images of the second experiment were acquired when the CD phantom was sandwiched among three heterogenous 70% glandular tissue-equivalent slabs. Hence, the first experiment simulates the 50% glandular breast tissue with a Compressed Breast Thickness (CBT) of 5 cm, while the second mimics a relatively large and dense breast (70G/30A composition) with a CBT of 7 cm. The heterogeneous breast tissue equivalent slabs are fixed on both DBT and PBT image acquisition, so the pattern

over the drilled discs that intimate the patient noise caused by the breast structures remains the same on each alternative image presented to the observers.

5.2.2 Imaging acquisition systems:

The commercial *"Selenia Dimension"* mammography device is utilized to acquire the DBT images in both experiments. The DBT system generates the x-ray with the anode target of Tungsten (W), and the beam is hardened by the external Aluminum filter of 0.7 mm. The DBT system emits the x-ray energy of 31 kV in each experiment by AEC mode and the estimated AGD in the first and second experiments are 1.58 and 2.43 mGy, respectively. The prototype PBT system with Tungsten anode and an external filtration of 2.5 mm-thick aluminum is used to acquire the phase-sensitive imaging. The PBT image in the first experiment is acquired at 59 kV, while the second experiment utilizes the x-ray energy of 89 kV. The AGD for each experiment is set to an equivalent radiation dose as the DBT system by the estimated DgN for the x-ray spectra of 59 and 89 kVs and the object thicknesses, measured entrance exposure, and manual mA.s settings on the PBT system using Equation 34

5.2.3 Observer study:

Eight observers participated in this study. All observers who participated in this study are non-medical professionals. Three observers are highly experienced in reading the phantom images, while five other observers are in the medical imaging field, with relatively less experience than the first group. The phase-retrieved image of the 2×2 binning mode acquired by the PBT prototype system is compared with the images of attenuation-based digital breast tomosynthesis. The in-focus slices from DBT and PBT images are used to create the multiple ROIs for the study. A total of 12 ROIs are cropped

from each modality in a way that 6 ROIs contain the discs with the same diameters and varying contrast levels, and 6 ROIs contain the discs with the same contrast but varying diameters.

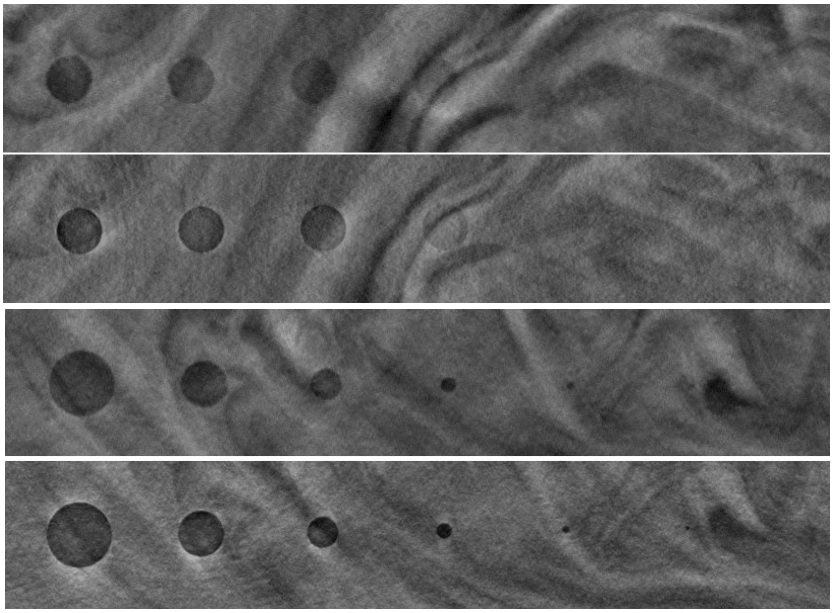
The two corresponding ROIs, extracted from the two imaging modalities, are presented on the 5 Megapixel grayscale radiology monitor (SONY LMD DM50) in a dark room, side by side. The PBT and DBT images are randomly placed on the right or left side of the screen at each trial. Although the observers must choose one image over another, there is no time constraint on the selection process, and observers could spend any desired time deciding which image is the preferred one. Each experiment is repeated two times, focusing on one image characteristic at a time. The evaluated characteristics are the total number of discs that can be detected on each ROI and the conspicuity of the discs.

According to the investigator's preference, the images are adjusted at the optimal window/level. The observers were not allowed to modify the window/level, but digital magnification was permitted. The reading results correlated to the pool of observers and the highly experienced group are analyzed independently for each image quality characteristic.

5.3 Results:

Figure 5.1 shows the two trials in the observer preference study. The first pair is an example of a diameter-fixed trial, and the second represents the contrast-fixed trial in the 2AFC study.

(a)



(b)

Figure 5.1: the images of 5-cm-thick CD phantom with 50/50 glandular-adipose tissue composition with a heterogeneous background. a) an example of displayed pair in a single trial, containing the discs of 4 mm in diameter and varying drilled depth. b) an example of a displayed pair with varying discs size but fixed contrasts of 1 mm drilled depth.

The delivered doses to 5-cm-thick 50% glandular breast phantom and 7-cm-thick 70% glandular breast phantom are estimated as 1.58 and 2.43 mGys, respectively. The first round of the 2AFC study only focuses on the detectability of both systems. The second experiment asks the observers to mark the preferred image on each trial in terms of the lesion conspicuity.

The PBT imaging system outperforms the DBT system in both experiments as the total number of PBT images selected by the pool of eight observers during the study is significantly higher than the selected images acquired by the DBT system. Table 5.1 indicates the result of the 2AFC study for each independent experiment among all observers.

Table 5.1: The overall percentage of DBT and PBT selected ROIs at two conditions and dual experiments to evaluate the detectability and the conspicuity of the detected objects on each modality. The PBT image of the 50G/50A phantom is acquired at 59 kV, while the PBT image of the 70G/30A phantom is acquired at 89 kV. The DBT image in either condition is acquired at an x-ray energy of 31 kV.

		DBT (%)	PBT (%)	p-value	Cohen's size
50G/50A phantom CBT = 5cm	Detectability	7.29	92.71	<0.001	15.99
	Conspicuity	6.3	93.7	<0.001	14.85
70G/30A phantom CBT = 7 cm	Detectability	17.7	82.3	<0.001	4.29
	Conspicuity	14.6	85.4	<0.001	7.30

The results show that the PBT imaging system is also the preferred modality, evaluated by only the experienced observers for detectability and conspicuity, and the difference is statistically significant. Table 5.2 shows the results analyzed from the data recorded by the experienced observers.

Table 5.2: The percentage of selected DBT and PBT ROIs by only experienced observers at two different conditions and dual experiments to evaluate the detectability of the imaging system and the conspicuity of the detected objects.

		DBT (%)	PBT (%)	p-value	Cohen's size
50G/50A phantom CBT = 5cm	Detectability	11.2	88.8	<0.001	16.17
	Conspicuity	8.3	91.7	<0.001	10.0
70G/30A phantom CBT = 7 cm	Detectability	25.0	75.0	0.0021	3.0
	Conspicuity	13.89	86.11	0.006	4.16

5.4 Discussion:

It has been extensively investigated over the decade, and phase-sensitive x-ray imaging shows promising potential in breast cancer screening and diagnosis. The phase-contrast image resolves the objects with relatively lower contrast than the attenuation-based radiography regardless of the utilized tube potential. The presumed reasoning can be listed as follow: a) the phase-shift coefficient is three magnitudes of order higher than

the attenuation coefficient in breast tissue or tissue equivalent phantoms, and b) the edge enhancement in phase-contrast image contributes to superior detection by the human eye. The phase-retrieved images also provide superior contrast because the image contrast is generated from both x-ray absorption and the phase-ship information, directly translated from the electron density differences between the object of interest and the surrounding background. The accuracy of phase retrieval depends on the utilized x-ray energy, and the effective energy shall not be significantly low because the phase attenuation duality does not hold.

Moreover, due to the geometric magnification in the in-line method, the imaging system's spatial resolution may also increase and contribute to higher detection capability. Analogous to the investigations previously discussed in this dissertation, the 2AFC study confirms that the mid-energy and high-energy phase-sensitive systems outperform the attenuation-based radiography in breast imaging. The mid-energy system shows slightly higher performance due to higher attenuation-based contrast without sacrificing the PAD-base phase retrieval to a great extent.

Again, it is essential to note that image processing algorithms in the PBT prototype system are not fully optimized. The 2AFC study is extremely sensitive to the overall image quality, such as noise reduction or artifact removal in 3D reconstruction. The observers in the 2AFC study reviewed a pair of DBT and PBT images on each trial and were asked to mark one image as the preferred image, even if there was no difference between the two images. The study design would benefit the DBT images to be selected because they are acquired by fully optimized, commercial, and advanced equipment. However, the results favor the PBT system with a statistically significant difference.

The analyses also revealed that the conspicuity of the object is well-established in the phase-sensitive system. The PBT detectability is still outperforming the DBT system, but the signal of the 2AFC study is not as strong as the 2AFC signal for conspicuity. The margin of the two signals is considerable, specifically for the experienced observer group. This observation can be justified as the experienced eye is trained to detect the subtle lesions on a radiographic image. Currently, the PBT images are noisier than the DBT images. Therefore, the 2AFC signal is not as strong as the inexperienced group.

Although the PBT system can resolve smaller objects with lower contrast, the difference between the two system performances would not be massive and considerable in the clinical application because the highly experienced and well-trained mammographer may still be able to perform similarly on DBT and PBT image reading. The noted observation also suggests that the commercialization of the imaging system requires advanced software optimization to see a considerable improvement in system accuracy among highly trained radiologists. They can detect the potential abnormalities when the acquired images are well-processed, but when the advanced image processing is adopted into the PBT system, a considerable difference in clinical performance might be on the horizon, according to the pre-clinical results.

Chapter 6. Evaluation of the x-ray beam quality by swift mathematical approach to estimate the average glandular radiation dose in breast x-ray imaging.

6.1 Introduction

The standard quality assurance and the concerns on the ionization radiation delivered on each x-ray imaging to the patients strictly depend on the x-ray beam quality. The x-ray energy and beam hardening impact the image quality for many reasons, such as the extent of the x-ray absorption in varying photon energy and detector quantum efficiency. Moreover, any x-ray photon deposits an expected energy load and bears a different risk level in altering the live organism's DNA, depending on the carrying energy.

The most applicable path to evaluate the x-ray beam quality is the Half Value Layer (HVL) thickness measurement of the utilized x-ray beam. However, the HVL measurement and the Quarter Value Layer (QVL) or Tenth Value Layer (TVL) measurements are time-consuming and tedious. Additionally, the measurement might be an impractical or hard-to-perform task. For instance, in Computed Tomography (CT) equipment, the x-ray source is not stationary in operating conditions. In this circumstance, if one needs to perform the HVL measurement, a service engineer must disable the rotating nature of the x-ray source in the CT scanner for this purpose. Similarly, modern fluoroscopy equipment utilizes the Automated Exposure Control (AEC) mode. Placing the aluminum foil in the x-ray beam path automatically triggers the tube potential or current adjustments where the HVL measurement should be performed under constant x-ray exposure and potential.

On top of all those procedural challenges, the HVL measurement shall be performed under stringent requirements, such as a narrow x-ray beam and minimal scattering rate from the surrounding objects. Understandably, controlling those requirements might be challenging in some operational sites if it is not infeasible. If one overcomes all those hurdles, the HVL measurement correlates to the calibration accuracy of the ionizing chambers. At the best operation settings, the ionization chambers usually carry ±4% tolerance in calibration accuracy and ± 5% tolerance in x-ray energy and exposure rate dependence. However, the increasing thicknesses of high attenuating materials like aluminum or copper during the experiment alters the accuracy of the calibrated ionization chambers. Therefore, a non-experimental estimation of HVL thickness for the utilized x-ray energy would be an alternative approach if it carries acceptable veracity.

Bremsstrahlung x-rays are produced when the electrons strike an anode target in the x-ray tube. Many techniques and algorithms have been produced to simulate the x-ray spectra based on the tube potential, target material, and other source characteristics to avoid the complexity of x-ray spectroscopy.[107-111] Although an authentic x-ray spectrum can always be measured, the advanced simulation algorithms also accurately provide the simulated x-ray spectra and can be used with high certainty. Unlike conventional mammography, the higher x-ray energy is usually applied in the phase-sensitive imaging (PCI) system. There has been significant progress in PCI System optimization, and this method has recently been employed in a first-ever pilot study in clinical settings. However, there is still an anticipated improvement in optimal x-ray beam quality in varying conditions.

Any system optimization which involves the uncommon x-ray energy demands the exposure re-adjustment to fulfill the regulatory health concern related to the patients' radiation dose. It is plausible to investigate the x-ray beam quality and dose estimation by HVL measurements for various kVp and external filtration substitutes during the foreseeable system optimization. However, the author of this dissertation has applied a mathematical approach and introduced an algorithm to analytically estimate the HVL, QVL, and TVL thicknesses without going over the demanding technical challenges and time-consuming procedural tasks for multiple alternatives for potential x-ray energies and external filtering by changeable materials. This original work has been published in the journal of applied medical physics.[112]

6.2 Materials and Methods

The X-ray-Matter interaction is determined by the number of atoms per volume unit of the object due to photons absorption or scattering probability. The outcome is often known as the linear attenuation coefficient of the object.[113-114] The x-ray exposure after passing through the object would be estimated by the Beer-Lambert equation as follows:

$$I = I_0 e^{-\mu x} \tag{40}$$

where, I_0 is the measured x-ray exposure prior to X-ray-Matter interaction, I is the measured x-ray exposure after passing through the object of thickness x, and μ is the object's linear attenuation coefficient. However, the linear attenuation coefficient dramatically depends on the energy of the interacting x-ray photon. Therefore, in a polychromatic x-ray beam, the Beer-Lambert equation can be extended as follows:

$$\int I \, dE = \int I_{0(E)} e^{-\mu_E \cdot x} \, dE \tag{41}$$

where, μ_E is the linear attenuation coefficient at the photon energy of E.

It must be acknowledged that the x-ray photons might be absorbed/scattered by the air molecule, yet, the amount is extremely negligible in experimental conditions with a short source-to-object distance. Hence, to precisely HVL estimation, the x-ray mass energy-absorption coefficient of the air shall be used to determine the x-ray exposure right before the object is placed. Equation 41 can be rewritten, considering the X-ray-Air interaction over the photon trajectory as follow:

$$I_1 = \int E.S(E).\left(\frac{\mu}{\rho}\right)_{E(en-Air)} dE \tag{42}$$

where, $S(E)$ is the normalized x-ray spectrum and is the mass-energy absorption coefficient of the air for the x-ray photon energy of E. The HVL thickness of aluminum is widely accepted to evaluate the x-ray beam quality in diagnostic radiology. Assume that the x-ray intensity right before the aluminum slab equals to I_1. According to the Beer-Lambert equation, the x-ray exposure after passing through an aluminum slab with a thickness of "x" can be determined as:

$$I_2 = \int E.S(E).\left(\frac{\mu}{\rho}\right)_{E(en-Air)} . e^{-\mu_E \cdot x}\, dE \tag{43}$$

The HVL is the object's thickness that reduces the x-ray exposure by one-half after x-ray photons pass through that object. Therefore, by the definition of HVL, the measured value for I_2 should be one-half of I_1. Accordingly, the HVL of aluminum or any other material can be mathematically estimated as follow, using Equations 42 and 43:

$$\int E.S(E).\left(\frac{\mu}{\rho}\right)_{E(en-Air)} . e^{-\mu_{E(AL)} \cdot x}\, dE = \frac{1}{2}\int E.S(E).\left(\frac{\mu}{\rho}\right)_{E(en-Air)} dE \tag{44}$$

Similarly, the QVL and TVL are computable by Equation 44 if the constant on the right side of the equation is changed to $\frac{1}{4}$ or $\frac{1}{10}$, respectively. However, the measured x-ray spectra were recorded as a discrete quantity rather than a continuous dataset.

Therefore, equation five might be rewritten as a discrete equation by any desired keV interval:

$$\sum E.S(E).\left(\frac{\mu}{\rho}\right)_{E(en-Air)}.e^{-\mu_E \cdot x} = \frac{1}{2}\sum E.S(E).\left(\frac{\mu}{\rho}\right)_{E(en-Air)}. \qquad (45)$$

The KeV interval is an arbitrary choice and depends on the desired accuracy we are expecting from the calculation. For simplicity, I estimate the HVL, QVL, or TVL using the normalized spectrum by 1 KeV interval, but the method can always be modified based on one's wish. To proceed with the mathematical estimation, the linear attenuation coefficients of aluminum (μ_E) and the air mass energy-absorption coefficients $\left(\frac{\mu}{\rho}\right)_{E(en-Air)}$ at any energy level should be available.

The XCOM program provided by the National Institute of Standard and Technology (NIST) is utilized to extract the aluminum mass attenuation coefficients at desired x-ray energy levels.[115] The air mass energy-absorption coefficients for the x-ray energies of 1 KeV to 20 MeV are also provided by NIST.[116] However, the data provided by the NIST do not cover all needed energies and only presents 11 data points within the diagnostic x-ray energy range. For interpolation accuracy, the air mass energy-absorption coefficients for interested x-ray photon energies were individually fit utilizing the least-square technique.[117-118]

The HVL, QVL, and TVL thicknesses can be mathematically estimated either by x-ray spectra measurement or by obtaining the simulated spectra from trusted sources. This method facilitates quick HVL assessment without complicated experiments for any alternative kVp/filter combination during the system optimization. However, to validate the trustworthiness of the introduced approach, the HVL, QVL, and TVL are measured by a traditional method for comparison purposes.

First, the x-ray spectra at various tube potentials and filtration are measured with the cadmium telluride (*CdTe*) x-ray detector (X123-CdTe Complete X-ray & Gamma Ray detector, Amptek, Bedford, MA). The device has a detector area of 25 mm^2, 25 μm thick graphite plus 100 μm thick Beryllium window, and it provides a channel resolution of up to 8K. The multiple-layer collimator (Amptek, Bedford, MA) with stainless steel housing, Brass spacer, and two Tungsten Collimator discs are used to limit the x-ray photon reaching the active area of the spectrometer.

The Tungsten discs are made of alloy HD17 (90% W, 6%Ni, and 4% Cu) with 2 mm thicknesses and central holes of 1000 μm and 200 μm in diameter. The number of x-ray photons reaching the spectrometer is counted in an energy resolution of 0.04 kV. The x-ray tube utilized in this study is a micro-focus X-ray source (Hamamatsu Photonics, Model L9181-06). The x-ray source is made of a tungsten (W) anode target, a 0.5 mm-thick Beryllium output window, and provides a focal spot size of 50 μm or smaller, depending on the tube output power. The x-ray spectrum at any tube potential is measured five consecutive times with a resolution of 4096 channels, and the final photon count comes from the average values of five measurements. The x-ray spectrum is rearranged in the desired KeV interval of 1 KeV.

Two physical effects may alter the recorded raw spectrum: (a) escape of secondary x-ray photon from the *Cd* or *Te* atoms and (b) loss of efficiency due to attenuation in the Beryllium window and the transmission through the detector. The secondary x-ray photons, developed in the detector by interaction with *Cd* or *Te*, escape the detector and reduce the measured energy. The difference between the actual number and the detected number of photons, $N_{\Delta}(E)$, at the energy of E, can be determined by:

$$N_{\Delta}(E) = \sum_p f_p(E + E_p) N_r(E + E_p) - \sum_p f_p(E - E_p).\, N_r(E) \qquad (46)$$

where p represents the four escape lines with energies E_p,$f_p(E)$ is the probability of characteristic photon escaping at energy E, and $N_r(E)$ is the actual number of photons at the energy level of E. The escape peak correction is performed by stripping procedure which is discussed in detail in the literature[14] on a channel-by-channel basis

The correction software (XRS-FP, Amptek) applies the described algorithm to correct the escape peak and the efficiency loss in the recorded raw spectrum. Since the attenuation and mass energy-absorption coefficients are energy-dependent, the median point for each x-ray energy level is used to roughly cancel out the variations in any given one keV range.

The experiment is performed by using a narrow beam geometry, calibrated ionization chambers, and varying thickness of aluminum slabs.[119] The Dedicated Mammography Chamber (10X6-6M, Radcal®, Monrovia, California) and the Leakage and Low-Level Measurements Chamber (10X6-180, Radcal®, Monrovia, California) are used for the low energy beam, and the mid-energy/high-energy x-ray beams, respectively. The source-to-slab distance is 69 cm, and the x-ray exposures are measured ten consecutive times, and then the average value is recorded as I_1. The varying thickness of aluminum foils are added to the x-ray path one after another, and the average values for the traversed x-ray exposures are recorded as I_2. The HVL, QVL, and TVL are estimated from the data by the curve fitting method.

The simulated x-ray spectra are obtained by using the software SpekCalc Pro version. This software provides the simulated x-ray spectra emitted from a tungsten anode. It operates between 0-300 kV and offers filtration options of W, Cu, Al, Sn, Ta, Be, air, and

water. The software's theoretical approach to calculating the emitted x-ray spectra was developed at the Institute of Cancer Research, UK.[120-121]

Solving Equation 6 for "x" could be impractical by the analytical method. A proprietary multi-paradigm programming language and numerical computing environment (MATLAB, MathWorks, Natick, MA) is used to solve the equation by a numerical method, using the midpoint algorithm by iteration task. Figure 6.1 shows the algorithm which is used to solve the equation numerically.

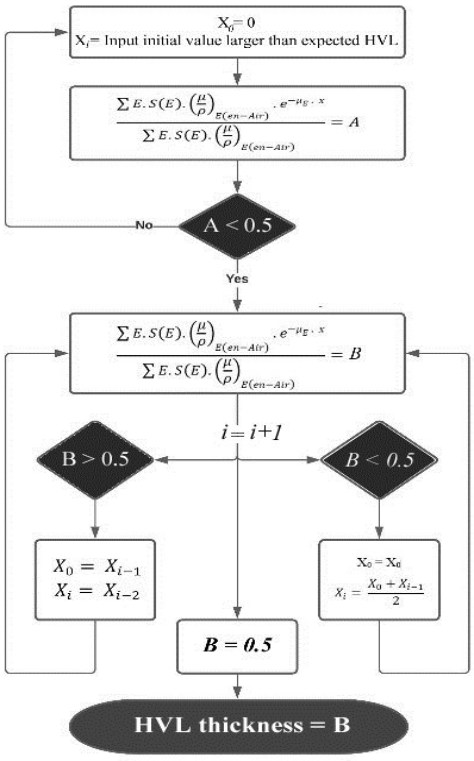


Figure 6.1: The numerical algorithm applied to solve Equation 6 for HVL thicknesses. The Value of 0.50 is replaced by 0.25 or 0.10 to estimate the QVL and TVL thicknesses.

6.3 Results

Figure 6.2 shows four normalized x-ray spectra measured by the Amptek x-ray spectrometer and the corresponding spectrum simulated by SpekCalc software. The x-ray

exposures under specified conditions at the distance of 69 cm from the source are measured consequently with varying thickness of aluminum slabs. The HVL, QVL, and TVL thickness are computed using the average recorded exposure values.

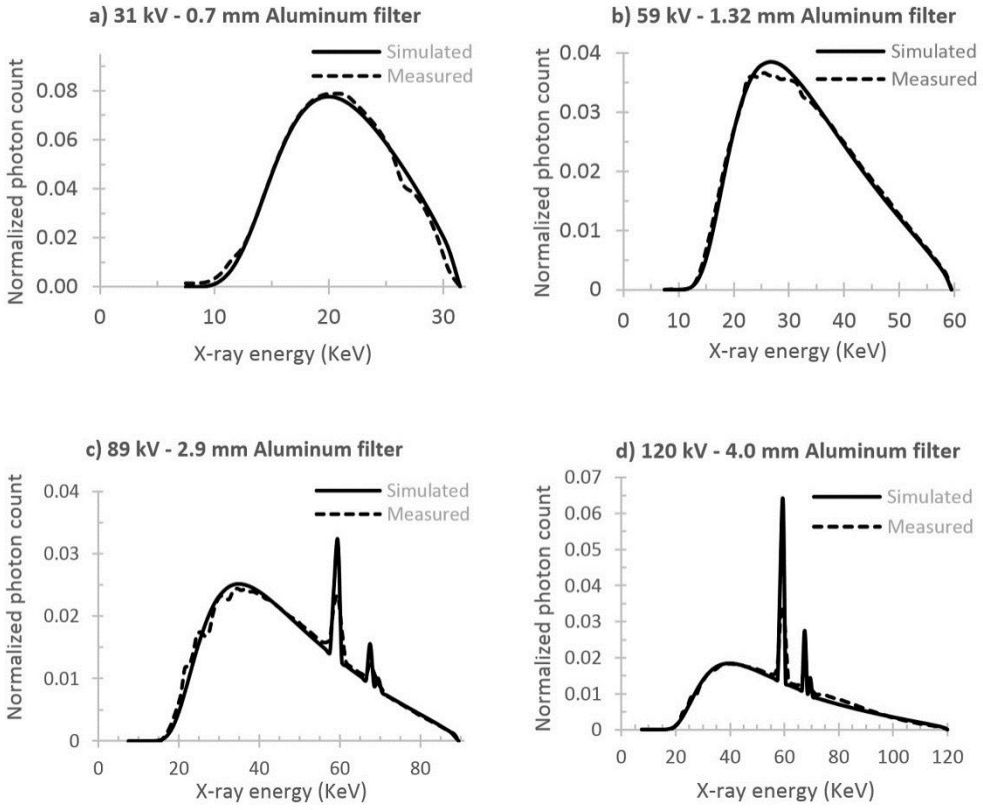


Figure 6.2: The measured x-ray spectra at four different tube potentials a) 31kV, b) 59kV, c) 89kV, and d) 120kV (Dashed line), and the corresponding simulated spectra (Solid line).

The calculated HVL, QVL, and TVL derived from measured x-ray spectra and the corresponding values measured by a calibrated ionization chamber in the experiment are presented in Table 6.1. The disagreement between the measured HVL and computed HVL from the measured x-ray spectra for 31 kV and 59 kV x-ray beams is 0.01 mm of aluminum, while the disagreements between these two values for 89 kV and 120 kV x-ray beams are 0.05 mm and 0.09mm of aluminum, respectively.

Table 6.1: The measured HVL, QVL, and TVL (in millimeters of aluminum) Vs. the corresponding estimated values by mathematical calculation. The percentage differences between measured and calculated values are shown inside the parentheses.

		31 kV *0.7 mm Al*	59 kV *1.3 mm Al*	89 kV *2.9 mm Al*	120 kV *4.0 mm Al*
Measured Values	HVL	0.51	1.31	2.99	5.05
	QVL	1.23	3.38	7.66	12.21
	TVL	2.47	7.36	16.21	25.23
Calculated by Measured Spectra	HVL	0.50 ***(-1.96%)***	1.30 ***(-0.76%)***	2.94 ***(-1.67%)***	4.96 ***(-1.78%)***
	QVL	1.22 ***(-0.81%)***	3.33 ***(-1.48%)***	7.57 ***(-1.17%)***	12.35 ***(+1.15%)***
	TVL	2.43 ***(-1.62%)***	7.28 ***(-1.09%)***	16.10 ***(-0.68%)***	24.80 ***(-1.70%)***
Calculated by Simulated Spectra	HVL	0.50 ***(-1.96%)***	1.30 ***(-0.76%)***	2.93 ***(-2.0%)***	4.97 ***(-1.58%)***
	QVL	1.22 ***(-0.81%)***	3.33 ***(-1.48%)***	7.57 ***(-1.17%)***	12.30 ***(+0.73%)***
	TVL	2.42 ***(-2.02%)***	7.28 ***(-1.09%)***	16.10 ***(-0.68%)***	24.65 ***(-2.30%)***

Table 6.1 also shows the computed HVL, QVL, and TVL derived from simulated x-ray spectra compared to measured values by a calibrated ionization chamber for the same experimental settings. Similarly, the disagreement between the measured HVL and computed HVL from the simulated x-ray spectra for 31 kV and 59 kV x-ray beams is 0.01 mm of aluminum. The disagreements between these two values for 89 kV and 120 kV x-ray beams are 0.06 mm and 0.08mm of Aluminum, respectively.

6.4. Discussion and Conclusion

The x-ray beam quality or penetration ability of utilized radiation is usually characterized by illustrating the thickness of Aluminum or copper that reduces the intensity of x-ray to one-half. A polychromatic x-ray beam is generally used in diagnostic or therapeutic radiology, and x-ray photons at various energies are absorbed or scattered differently. Hence, the measurement of HVL under certain experimental conditions is widely accepted by medical physicists. However, the narrow beam implementation, the position and the distance of the ionization chamber from the source and the attenuating material, the presence of scattering material in the vicinity of the chamber, the calibration accuracy of the utilized detector, the energy dependence, or the exposure rate dependence of the ionization chambers, and the ability of equipment to emit stationary and constant x-ray exposure during the experiment sometimes introduce technical challenges to readily measuring the HVLs.

On the other hand, the HVL measurement might be questionable due to the energy dependence and the calibration accuracy of ionization chambers. At the same time, the x-ray spectra continually become harder by adding additional aluminum foils in the HVL measuring procedure. The results show that the calculated values for HVL, QVL, and TVL from the x-ray spectra are usually smaller than the measured ones. This might be expected due to various reasons, including but not limited to unknown scattering rate during the experiment or issues with calibration in the ion chamber when the beam continually hardens during the experiment by adding the additional aluminum filters. Another concern in experimental measurement is the robustness of the curve fitting model to estimate the exact values. Theoretically, the reduction rate for the x-ray

exposure shall follow the Beer-Lambert equation, but the calibration issue for various x-ray spectra and the scattered x-ray photons slightly distort the exponential curve. Hence, the applied fitting model for the generated curve may also impact the HVL, QVL, and TVL estimation. For instance, in Table 6.1, all the measured values are greater than the calculated values, except QVL for the 120 kV beam. The reason could be the fitting model for the imperfect exponential curve caused by relatively thicker filtration for the 120 kV beam during the experiment.

The proposed method in this article provides a consistent approach to calculate the HVL, QVL, and TVL in a variety of situations such as high-energy x-ray breast imaging, CT scan studies, the radiographic equipment with fixed AEC feature, inevitable x-ray scattering from the surrounding objects in diagnostic radiology or radiation therapy sites with adequate precision. If the measured x-ray spectra are being used to calculate the HVL, the accuracy of the measured spectrum must be confirmed as the spectroscopy itself is a sensitive procedure, and a well-experienced operator is a fundamental requirement. Correspondingly, if the simulated x-ray spectra are used to estimate the HVL by this method, the robustness of the x-ray simulating method shall be upheld to avoid imprecise results.

Needless to point out, the conventional HVL measurement is widely accepted in the field, and the scope of this work is not to undermine the current practice or to suggest replacing it with the mathematical HVL estimation. Additionally, the novel solid-state diagnostic dosimeters are increasingly being used in clinical practice. The significant advantage of these dosimeters over the traditional ionization chambers is their capability to identify the air kerma, tube voltage, exposure time, and HVL from single irradiation

and accommodate the problems associated with the backscatter radiation. However, a relatively noticeable error is expected when measuring the x-ray beams filtered by various materials, such as Copper.[122] Nevertheless, this straightforward method could be an appropriate solution and provides a fast and convenient estimation of HVL, QVL, or TVL with decent precision, when the HVL measurement is challenging due to given conditions or expresses the possibility of disputable values if the measured or simulated x-ray spectra are available.

Chapter 7. The impact of the x-ray spectral filtration on image quality and detectability of the mid-energy in-line phase-sensitive x-ray imaging

7.1 Introduction

This chapter aims to evaluate the impact of varying external filtration on the image quality and investigate the feasibility of reducing the image acquisition time while the delivered dose to the patient remains the same and the anticipated image quality is preserved. X-ray sources provide either monochromatic or polychromatic x-ray beams. Almost all of the utilized x-ray sources in medical imaging generate the polychromatic x-ray beam by striking the high-energy electrons to a target material. The Bremsstrahlung radiation produced by an x-ray tube consists of a continuous spectrum of radiation wavelengths, and its characteristic is governed by both the target material and the potential of the striking electrons.

Chapter 6 of this dissertation discusses that the x-ray beam quality can be assessed by a half-value layer (HVL) thickness measurement. The beam's hardness degree directly impacts the patients' absorbed radiation dose and image quality. Therefore, regulatory organizations such as FDA have set a minimum requirement for the x-ray beam quality by defining the least acceptable HVL of aluminum (CFR Title 21 §§ 1020.30) for any specified x-ray energy and detector type.

The minimum requirements are mainly because a small segment of the x-ray energies within the emitted spectrum tends to increase the absorbed radiation dose with less contribution to image quality if there is no impact at all.[122] Therefore, inherent filtration is usually utilized on x-ray tubes, and depending on the clinical application, additional filtration is usually applied to remove the remaining undesired low-energy x-ray photons.

Although the x-ray beam hardening by external filtration is a valuable and widely accepted approach for reducing the radiation dose, the signal production, system's response to the filtered x-ray beam, and image quality at delivered radiation dose may vary from case to case, depending on the imaging object and the detector characteristics.

The beam quality becomes more significant in breast imaging as the breast is one of the most sensitive tissue in the body to ionizing radiation.[123] Additionally, conventional breast x-ray imaging is acquired at the lower band of diagnostic x-ray energy to yield the highest attenuation-base image contrast possible due to minor differences between the tumor and glandular tissue in terms of x-ray attenuation coefficients. However, the low-energy beam might be absorbed by breast tissue at a higher rate than a high-energy beam, primarily used in other x-ray imaging applications. The additional x-ray exposure is required to offset the low-energy photon absorption to ensure enough signal-to-noise ratio.

Although the system's Modulation Transfer Function (MTF) would not be impacted significantly by varying degrees of external filtration, the noise power spectrum (NPS), consequently, the detective quantum efficiency (DQE) of any digital imaging system may show a notable change. The external filtration would considerably influence the Normalized glandular dose (DgN) coefficient that determines the absorbed radiation dose. Under constant radiation dose, the observer's ability to detect the suspicious lesions and the objective imaging parameters, such as contrast-to-noise ratio (CNR), might be remarkably impacted, depending on the system's response to varying beam quality.

The in-line phase-sensitive breast x-ray imaging is an evolving and promising technique that exploits both the x-ray photon absorption and the wave-front phase shift,

traversing through the imaging object. The attenuation-base image contrast decreases swiftly while the x-ray energy increases. Contrarily, the phase-induced contrast decreases much slower than attenuation-base contrast by increasing the x-ray photon energy.[50]

Hence, we can increase the x-ray energy and restore the attenuation-based lost information by retrieving the object phase map. Throughout this dissertation, I investigated both high-energy and mid-energy phase-sensitive imaging potentials. The high-energy system offers more accurate PAD-base phase retrieval than the mid-energy system.[48,93,125] However, it has been shown that the phase-attenuation duality may sufficiently exist in light elements for the PAD-base phase retrieval method when the system operates the mid-energy x-ray beam.[50-51] Chapter 3 of this dissertation research assessed the plausible advantages of the mid-energy system by subjective studies. Chapters 4 and 5 subjectively investigated the performance of conventional breast x-ray imaging and the phase-sensitive imaging systems utilizing the mid-energy x-ray beams, and the subsequent challenges in further improvement of the system are identified.

A critical challenge in in-line mid-energy phase-sensitive imaging systems would be a prolonged exposure time to deliver the expected radiation to generate the desired SNR. In phantom studies, a longer image acquisition time is not much concern. However, in the clinical translation of phase-sensitive imaging, the extended image acquisition means a highly unpleasant experience in patients while the breast is compressed during the image acquisition, and most importantly, added image blurring due to patient movement.

The microfocus x-ray sources cannot provide a higher tube current as a conventional x-ray source provides because the high-amp current might damage or melt the target by excessive heat. The manufacturing technology of the microfocus tubes is advancing, and

the higher output power tubes could slowly launch to the market. Nevertheless, until the high-output power tubes are not widely available, it might be essential to investigate the feasibility of utilizing varying external filtration to reduce exposure time if it does not substantially impact the image quality and users' performance.

7.2 Method and Materials

7.2.1 Imaging prototype:

A similar imaging geometry to the previous investigations is employed to identify the most suitable external filtration for the mid-energy in-line phase-sensitive system to reduce the image acquisition time under fixed radiation dose without a significant compromise in image quality. A microfocus x-ray source with a tungsten anode target and 0.5 mm-thick beryllium inherent filtration (Hamamatsu Photonics, Model L9181-06) provides the x-ray beams at 59 kVp and 89 kVp. A high-resolution x-ray sensitive CCD camera (Photonic Science X-ray ImageStar 9000, Hastings UK), with a pixel pitch of 22 microns, 300 μm-thick column-structured Cesium Iodide scintillator, and 66×66 mm active area is used for image acquisition. The R_1 distance of 27 inches (≈68.6 cm) and R_2 distance of 33 inches (≈83.8 cm) are applied to provide a magnification (M) factor of approximately 2.2, followed by the principle of the in-line phase-sensitive imaging systems.[45,125-126] The varying thicknesses of aluminum slabs are placed in front of the output window to remove the low-energy x-ray photons by different proportions. The combinations of x-ray tube potentials and external aluminum filters for the acquired images are shown in Table 7.1

Table 7.1: different x-ray exposure times for varying beam qualities, delivering the equivalent average glandular dose of 1.3 mGy.

x-ray energy	External filtration	Estimated Radiation Dose	Exposure Time (sec)
59 kV	1.0 mm Al	≈1.3 mGy	38
	1.3 mm Al		46
	1.5 mm Al		47
	2.0 mm Al		57
	2.5 mm Al		67
89 kV	1.5 mm Al	≈1.3 mGy	20
	2.0 mm Al		23
	2.5 mm Al		26
	2.9 mm Al		29
	3.3 mm Al		32

7.2.2 Dose calculation:

The x-ray spectra were measured by a Cadmium-Telluride (CdTe) spectrometer (X123-CdTe complete X-ray & Gamma-ray detector, Amptek) at each operating x-ray energy. The correction tool ((XRS_FP, Amptek) is used to correct the loss of efficiency and escape of secondary x-ray photons from Cadmium and Telluride atoms utilized in the spectrometer. The normalized glandular dose coefficients are estimated with computer-based Monte-Carlo simulation using the varying x-ray spectra for a breast phantom of 50% glandular composition.[94,127] The entrance exposure rates at R1 distance are measured by the leakage and low-level measurement ionizing chamber (10X6-180, Radcal®), and the exposure time and tube current are set to deliver the specified radiation dose during the image acquisition using Equation 34.

7.2.3 Imaging object

Three modular homogenous phantoms (CIRS Inc, Norfolk, VA) are used to acquire the images. The phantoms are epoxy resins, simulating the x-ray attenuation properties similar to 50% glandular and 50% adipose tissue breast. The total thickness of the three

phantoms equals 50 mm compressed breast thickness. One of the phantom slabs included the 6×6 matrix of a contrast-detail test pattern. The cylindrical holes are drilled to provide the contrast discs with diameters of 0.25, 0.5, 1.0, 2.0, 3.0, and 4 mm and depths of 0.1, 0.2, 0.4, 0.6, 0.8, and 1.0 mm. The image of the 6×6 CD test pattern cannot be scanned by a single image acquisition with the CCD camera, which has a smaller input active area, and the object is divided into four ROIs to scan the contrast-detail pattern entirely.

7.2.4 Human observer study

Three experienced observers randomly review the acquired images and identify the minimum perceptible disc for each diameter. The images of the observer study are all acquired at a fixed radiation dose (1.3 mGy at 59 and 89 kVps). The observers were blind to the image acquisition parameters, and the images were displayed on a grayscale high-resolution mammography LCD monitor (Sony LMD-DM50) in a dark room. The readers could adjust the window/level, and there was no time limit for interpretation. The C-D curves are generated from combined ROIs for each image according to the average reading scores of the observers to evaluate the minimum perceptible contrast that is visible at any specific diameter based on the signal-detection theory (SDT).

7.2.5 Objective studies

According to SDT, the contrast-detail curve carries the concepts of spatial and contrast resolutions. However, the number of observers and cases was limited in this study. Therefore, quantitative analyses are also added to compare and confirm the subjective results. The Contrast-to-Noise Ratios (CNR) of the two target discs (diameters/drilled depths of 2.0mm/1.0mm and 1.0mm/0.8mm) are calculated by the following equation.[128-129]

$$CNR = \frac{\bar{I}_A - \bar{I}_B}{\sqrt{\frac{(\sigma_A^2 + \sigma_B^2)}{2}}} \quad (47)$$

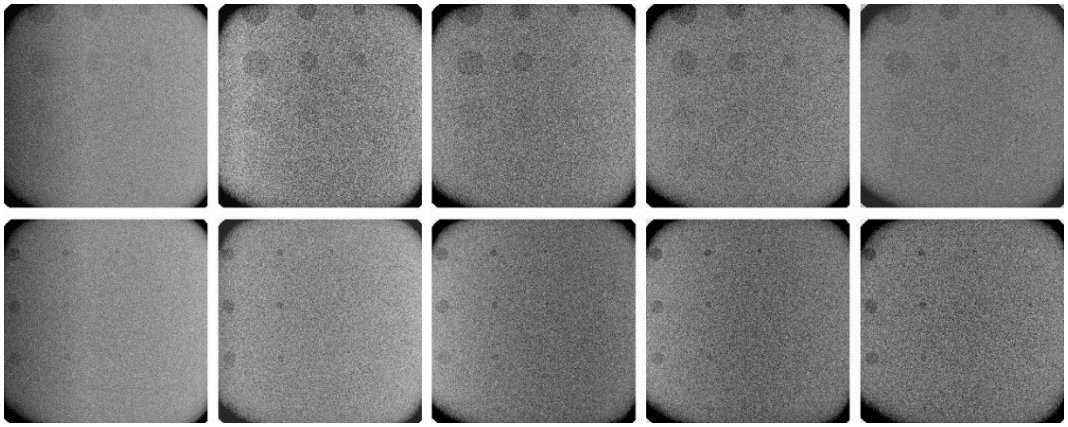
where $\bar{I}_A$ and $\bar{I}_B$ are the mean pixel intensities over the target region and the background over the same size area, and σ_A^2 and σ_B^2 are the corresponding pixel value variances.

The background intensities and variances are estimated by an average of 5 ROIs around the target disc with the same size as each corresponding target disc. The same imaging geometry acquires all the images.

7.3 Results

7.3.1 Human observer study:

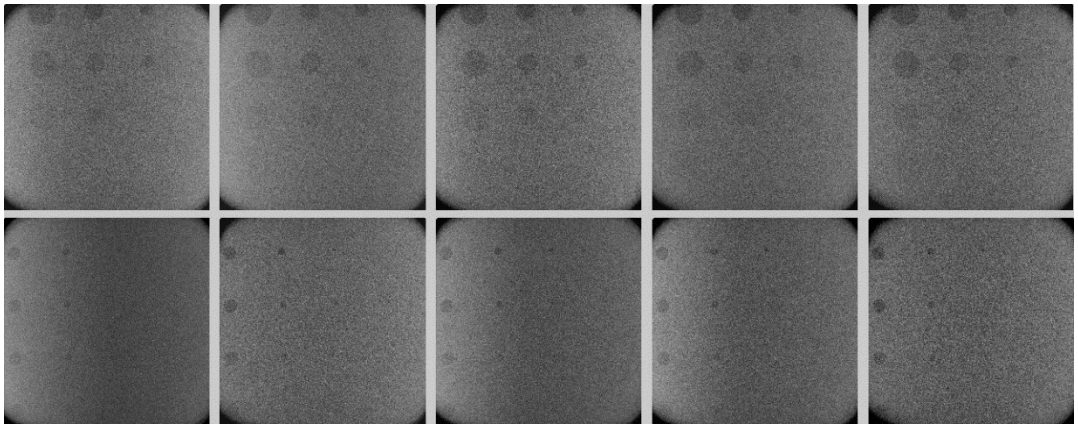
Figure 7.1 shows two ROIs out of four, where the perception thresholds are located. The contrast-detail images of the homogenous five-cm-thick phantoms were acquired at an x-ray energy of 59 kV by varying external filtration stated in Table 7.1 and the equivalent average glandular radiation dose of 1.3 mGy.



(a) *(b)* *(c)* *(d)* *(e)*

Figure 7.1: The images of C-D phantoms acquired at an x-ray energy of 59 kV by varying x-ray beam filtration a) 1.0 mm Al, b) 1.3 mm Al, c) 1.5 mm Al, d) 2.0 mm Al, and e) 2.5 mm Al filters.

Figure 7.2 shows the images of the same homogenous five-cm-thick Contrast-Detail phantoms acquired at an x-ray energy of 89 kV and a similar AGD of 1.3 mGy.



(a) (b) (c) (d) (e)

Figure 7.2: The images of C-D phantoms acquired at an x-ray energy of 89 kV by varying x-ray beam filtration a) 1.5 mm Al, b) 2.0 mm Al, c) 2.5 mm Al, d) 2.9 mm Al, and e) 3.3 mm Al filters.

Three observers viewed the acquired phase-contrast images independently and were blind to the imaging parameters. The contrast-detail curves are generated from the average scores of three readings. Figures 7.3 and 7.4 show the C-D curves generated for the in-line phase-sensitive imaging systems operating at 59 kVp and 89 kVps.

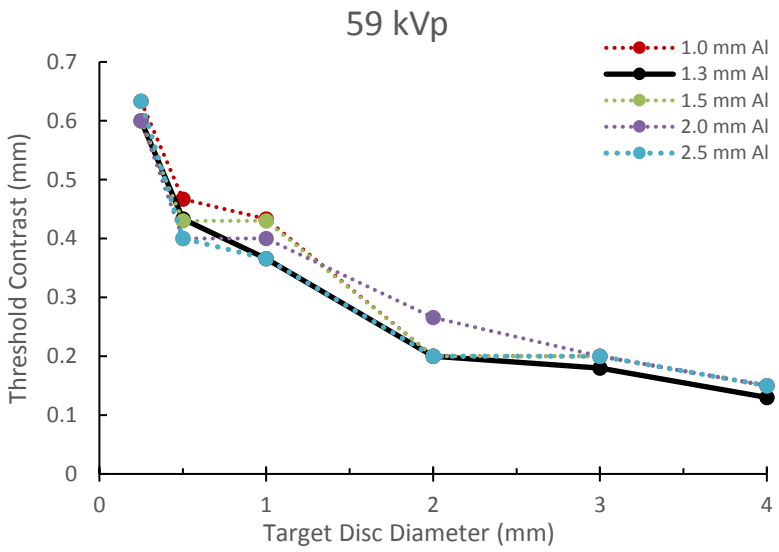


Figure 7.3: The C-D curve generated from the phase-contrast images acquired at 59 kVp and a varying thickness of aluminum filtration.

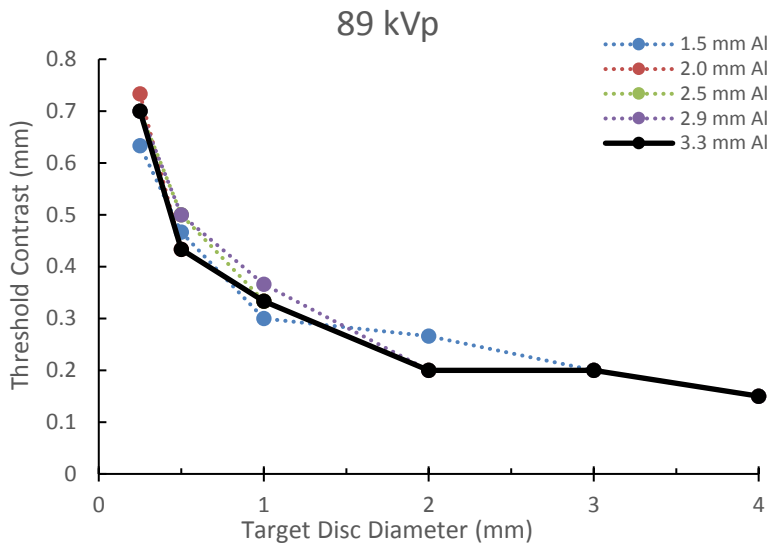


Figure 7.4: The C-D curve generated from the phase-contrast images acquired at 89 kVp and a varying thickness of aluminum filtration.

As shown in Figure 7.1, the observer performance is comparable when the image is acquired at a tube potential of 59 kV, and the x-ray beam is filtered by either 2.5 mm or 1.3 mm-thick aluminum. The exposure time can be reduced from 67 seconds to 44 seconds which is 34% shorter than the image acquired by 2.5 mm aluminum filtration. The imaging system operated at 89 kVp shows that observer performance is relatively comparable by varying filtration with a 3.3 mm aluminum filter slightly performs better. Intuitively, the imaging system generally performs better when the calculated area under the C-D curve shows a smaller value as the system distinguishes the lesions with less contrast. The areas under the CD curves are calculated by the trapezoid model and shown in Table 7.2.

Table 7.2: The calculated area under the CD curves for two imaging systems operated at 59 and 89 kVps and varying x-ray beam qualities by different filtrations.

Tube potential	Filter thickness (mm of Al)	Area under the CD curve
59 kV	1.0	1.1333
	1.3	0.9568
	1.5	1.0337
	2.0	1.0660
	2.5	0.9786
89 kV	1.5	1.0201
	2.0	0.9788
	2.5	0.9997
	2.9	1.0245
	3.3	0.9747

As shown in the table above, the system providing the x-ray energy of 59 kV overall resolves the target discs with less contrast where the system operated at 89 kVp slightly performs better when the aluminum thickness of 3.3 mm filters the x-ray beam.

7.3.2 Objective analyses:

The observer performance study might be suffering from the low statistical power due to the limited number of participating observers. The objective analysis has been added to evaluate and confirm the results from the subjective analyses. Table 7.3 represents the calculated CNRs in the image acquired by different acquisition parameters. The CNR_1 represents the estimated contrast-to-noise ratio on the target disc of 2.0 mm in diameter and 1.0 mm drilled depth, whereas the CNR_2 serves as calculated contrast-to-noise ratios on the target disc of 1.0 mm in diameter and drilled depth of 0.8 mm.

Table 7.3 The calculated CNRs on two arbitrary target discs on the images acquired by 59 kV and 89 kV and varying aluminum thicknesses as external filtration.

	Filter (mm Al)	**CNR_1**	**CNR_2**
59 kV	1.0 mm Al	1.45	0.89
	1.3 mm Al	1.50	1.00
	1.5 mm Al	1.47	0.95
	2.0 mm Al	1.42	0.90
	2.5 mm Al	1.51	1.00
89 kV	1.5 mm Al	1.48	0.90
	2.0 mm Al	1.51	0.90
	2.5 mm Al	1.53	1.00
	2.9 mm Al	1.53	1.02
	3.3 mm Al	1.57	1.02

The normalized glandular dose coefficient increases as the x-ray energy increase at fixed external filtration. Similarly, the glandular dose coefficient increases by beam hardening at the same tube potential. Furthermore, we expect a nearly linear decrease in normalized glandular radiation dose by a softer beam at fixed kVp, where the x-ray exposure increases exponentially. Therefore, reducing the aluminum thickness reduces the image acquisition time. The currently available prototype PBT system, designed for the clinical study, runs by a fixed external aluminum filter of 2.5 mm thick. Table 7.4 shows the deviation from the current condition regarding the required image acquisition time to deliver the intended radiation dose.

Table 7.4 The deviations of the image acquisition time from the required time in the prototype PBT system to deliver the intended radiation dose.

	Filter (mm Al)	**Time/1mGy**(sec)	**Deviation**(%)	**CNR_1**
59 kV	1.0 mm Al	29.3	-43.3	1.45
	1.3 mm Al	35.5	-35.0	**1.50**
	1.5 mm Al	36.3	-29.8	1.47
	2.0 mm Al	43.5	-15.8	1.42
	2.5 mm Al	51.7	0.0	**1.51**
89 kV	1.5 mm Al	15.3	-24.2	1.48
	2.0 mm Al	17.7	-12.1	1.51
	2.5 mm Al	20.1	0.0	1.53
	2.9 mm Al	22.4	+11.0	1.53
	3.3 mm Al	24.5	+21.8	1.57

7.4 Discussion and Conclusion

This part of the dissertation investigates the feasibility of reducing the image acquisition time by utilizing a potentially thinner external filtration where the delivered radiation dose is not changed. A low-energy x-ray beam is usually utilized in breast x-ray imaging due to insignificant differences between the attenuation coefficients of glandular structure and the malignant lesions. For the same reason, a low-energy system was first established for the in-line phase-sensitive x-ray imaging system. However, there are two main challenges in a low-energy in-line phase-sensitive imaging system. At least, a partial coherent x-ray beam is an essential requirement for the in-line phase-sensitive technique that mandates utilizing the microfocus polychromatic x-ray tubes. The microfocus tubes usually have a limited tube current to protect the target material from the excessive heat that results in low output power. Hence, lower x-ray energy in a microfocus tube with the current in the range of micro-amp will result in a longer exposure time to deliver the desired radiation.

Moreover, the phase-attenuation duality is more prominent at slightly higher x-ray energy than the x-ray beam utilized in conventional mammography. Thus, the in-line phase-sensitive imaging technique soon migrated to a high-energy system to reduce exposure time and ensure that the PAD-base retrieval algorithm yields a more accurate phase mapping. As is discussed in Chapter 3, the mid-energy has the potential to balance the attenuation-base, and phase-induced image contrasts, without a significant compromise in PAD-base phase retrieval but slightly increased exposure time.

In pre-clinical studies, the image acquisition time is not a big concern. However, in clinical studies, that would be a significant concern for the two following reasons: a) the

longer image acquisition would cause a higher degree of image blur because the patient is not able to hold her breath throughout the image acquisition, and b) the mammographic procedures are done by breast compression, enabling us to deliver a reduced radiation dose, and the breast compression is not a pleasant experience in patients, specifically when the acquisition time is longer by a mid-energy system.

The prototype PBT system used for the observer performance studies in this dissertation research has a fixed aluminum filter of 2.5 mm thick. Further optimization of the PBT system requires the changeable external filter, specific for the utilized x-ray energy. The image acquisition time is considerably longer when the system is operated at a tube potential of 59 kV. The response of the digital detector to different x-ray beam qualities would not be the same as the NPS and DQE slightly vary as the x-ray beam hardens. Additionally, the external filtration removes more low-energy photons from the beam, and the normalized glandular dose coefficient and the entrance x-ray exposure change. Thus, the AGD becomes another variable that should take into consideration.

This preliminary investigation revealed that we could use a thinner aluminum filter, such as 1.3 mm-thick aluminum for a 59 kV system, that reduces the image acquisition time by 35%, where the imaging system's performance is comparable to 2.5 mm filtration if it is not even better. The small increase in filter thickness in the system operated at an x-ray energy of 89 kV (2.9 mm or 3.3 mm Al) slightly improves system performance. The image acquisition time might not be a significant concern in non-dense standard breast-size imaging by a higher x-ray energy system. However, when a higher dose is needed for dense and large breast imaging, the 20% increase in imaging time could be an issue and requires detailed studies in clinical settings.

This preliminary study only investigated the observer performance and the quantitative parameter of the image quality on the phase-contrast images. The PAD-based phase retrieval method operates more precisely if the utilized x-ray spectrum has higher effective x-ray energy. This investigation should be expanded to the phase-retrieved images to ensure that a softer x-ray beam, filtered by 1.3 mm, would not affect the competency of the PAD-base phase-retrieved method.

Lastly, according to some technical challenges in image acquisition by a flat panel detector, this study was conducted by a less noisy CCD detector. I do not expect a significant variation in the trend of image quality while the x-ray beam quality is altered. This investigation aims only to evaluate the feasibility of reducing the exposure time in a mid-energy system. However, I should acknowledge that response of each detector will slightly change depending on the scintillator type, scintillator thickness, and the corresponding DQE by interacting photon energy. The results might be independently confirmed in the future in the prototype PBT system.

Chapter 8. Conclusions

8.1. Summary

The idea for in-line phase-sensitive x-ray imaging in breast cancer application started over two decades ago. Soft tissue x-ray imaging, such as mammography, is usually performed at a lower x-ray energy range to increase the attenuation-bases image contrast. Breast x-ray imaging might still be challenging in women with relatively denser or larger breasts because there are minor differences in x-ray attenuation coefficients between the glandular structure and the pathologic lesions, regardless of the utilized x-ray energy.

The idea behind the potential advantages of breast x-ray imaging, utilizing the phase-sensitive method, arises from the differences between the malignant and healthy tissue characteristics and, more importantly, the disparity between the real and imaginary parts of x-ray refraction indices for light elements within the diagnostic x-ray energy range. The malignant tissue shows higher physiologic and biochemical activities. Additionally, the blood vessels are promoted in malignant tissue and neighboring structure (by a mechanism called angiogenesis) to meet malignant cells' substantial energy demand. The alterations in cancerous tissue result in higher electron density and, consequently, considerable phase shift for the x-ray beam passing through the malignant tissue inside the breast.

Accordingly, the initial investigations in breast phase-sensitive imaging began to evaluate the validity of such a hypothesis. The most applicable phase-sensitive x-ray imaging technique in clinical applications could be the in-line phase-sensitive system that does not need grating-bases tools or monochromatic x-ray beam.

However, the fundamental requirement for this technique is at least a partially coherent beam provided by a microfocus x-ray tube. The optimal x-ray energy for soft tissue imaging falls in the low-energy spectra to increase the attenuation-base image contrast. However, the tube current is limited in the micro-focus x-ray sources that are currently available. Additionally, to retrieve the tissue phase map, one should acquire two images with varying R_2 distances using this imaging technique. Due to radiation dose concerns, a two-projection approach is not desired for patient imaging. Thus, the low-energy in-line phase-sensitive x-ray imaging has been replaced with the high-energy prototype to overcome the challenges related to image acquisition time and phase mapping by applying the PAD-base retrieval method.

The PAD-base phase retrieval employs the principle that the x-ray interaction with light elements is mainly dominated by Compton scattering in the higher end of the diagnostic x-ray spectra. Therefore, the single-projection PAD-based phase retrieval method could be employed with negligible error in phase mapping if the x-ray energy is higher than what is usually utilized for breast x-ray imaging. The increased x-ray energy also shortens the image acquisition time compared to a low-energy x-ray beam, delivering the same radiation dose. Since the phase shift coefficients are three orders of magnitude greater than attenuation coefficients, the increased x-ray energy is not expected to diminish the image quality as it would happen in attenuation-based techniques, even if we lose attenuation-base image contrast by higher x-ray energy to some extent.

For the stated reasons, the vast majority of pre-clinical investigations have been performed by utilizing the higher x-ray energies in the range of 100 kV up to 130 kV. On

the other hand, PCI imaging is not merely phase mapping, and the information obtained from the x-ray absorption still provides valuable insight into the structure and potential abnormalities within the breast. Utilizing the mid-energy x-ray spectrum as a balancing trade-off between the attenuation-based x-ray imaging and admissible precision in phase mapping by the PAD-base retrieval method emerged based on the speculation mentioned above.

The quantitative analyses such as NPS and DQE showed that the mid-energy system would potentially outperform the high-energy system, and the recent improvement in the PAD-based algorithm allowed shifting PCI from high-energy into the mid-energy x-ray spectra. Like the ones discussed in this dissertation, a few pre-clinical investigations have been performed to evaluate the possibility of adopting the mid-energy PCI in clinical applications. The mid-energy system shows a promising result in identifying the lesion with lower contrast than a PCI system operating at relatively higher energy.

The concern was the performance of the PAD-based phase retrieval in the mid-energy range and the inevitable longer exposure time in a mid-energy system that results in image blur due to the patient's movement. This dissertation research shows that a mid-energy system provides a higher image quality than a high-energy system in phase-contrast images. The PAD-base phase retrieval algorithm in the mid-energy system (59 kVp and above) could still provide an acceptable and adequate assessment of objects' phase mapping resulting in comparable phase-retrieved images acquired by both mid-energy and high-energy systems under equivalent radiation doses.

The observer performance and preference studies also showed that the prototype PBT system operated at either 59 kVp or 89 kVp outperformed a commercial attenuation-

based breast tomosynthesis system in the pre-clinical studies. However, the length of image acquisition does not influence the image quality in stationary objects like simulating breast phantoms. On the other hand, the impact of a relatively longer image acquisition time in human imaging is unknown, as the patient would not be able to hold her breath during prolonged image acquisition. Therefore, the last chapter focused on the x-ray beam quality and investigated the impact of varying beam filtration on image quality to assess the feasibility of reducing exposure time at a fixed radiation dose while the quantitative and qualitative element of the acquired image is preserved.

The preliminary investigation showed that reducing the acquisition time at 59 kVp is conceivable by at least 35% at equivalent doses in the prototype PBT system if the external aluminum filter is switchable from 2.5 mm to 1.3 mm. The objective and subjective studies showed slightly better system performance with a 1.3 mm aluminum filter than with a 2.5 mm-thick filter, and the acquisition time can also be cut considerably for the clinical translation of the system. However, this preliminary investigation is performed on the phase-contrast images on a bench-top setup where the prototype PBT system provides the phase-retrieved images. Although an extensive change is not expected in how the PAD-base phase-retrieval algorithm would perform by varying beam quality at any tube potential, the impact of less low-energy photon removal from an x-ray beam on the phase-retrieved images has not been investigated yet.

8.2. Future Research Direction

The results and the analyses on the observer performance and preference study presented in this dissertation further encourage future investigation and system improvement. In the pre-clinical phase, high-energy in-line phase-sensitive x-ray imaging

has been investigated in detail. The mid-energy in-line phase-sensitive system is somehow a new approach for breast phase-sensitive imaging. The pre-clinical studies confirm the advantage of the mid-energy beam in balancing between the attenuation-based image contrast and slight offset in the PAD-based phase retrieval method. However, the ultimate destination in the phase-sensitive x-ray imaging project is translating the well-established theory into clinical applications. Hence, the subjective evaluation of the system performance would be necessary for system characterization in the translational stage.

The investigations presented here utilized the prototype system, and the acquired images were compared with the images acquired by the commercial systems. Expectedly, the image quality and noise reduction strategies would impact human observers' ability to identify the subtle pathologies and, consequently, the expected outcome in an observer preference study.

Nevertheless, my investigations by observer performance and observer preference studies revealed the advantages of the mid-energy system over both the high-energy phase-sensitive and the attenuation-based imaging systems in phantom studies. The results were encouraging, considering that the PBT prototype system still employs the evolving algorithms in image processing rather than matured ones.

On the other hand, both mid-energy and high-energy approaches require a substantially more prolonged exposure due to x-ray tube power limitation. More powerful x-ray tubes that would not face high-temperature challenges because of the high current at the target site are required in the transitional phase. The expected alternative would be the liquid metal jet x-ray tubes or carbon nanotubes. Additionally, the improvement in the

mechanical stability of the system (continuous gantry movement instead of stop and shoot acquisition) further shortens the acquisition time, which will further reduce the image blurring and improve image quality in the PBT system.

In the meanwhile, and until the tube output power will not be a challenge in clinical translation, the comprehensive investigations on the impact of utilizing different materials, such as Rhodium, Molybdenum, and Silver with absorption K-edge at 20-25 KeV on image quality and required exposure time to deliver the desired radiation dose could be helpful in further improvement of the imaging system.

Additionally, the ancillary advantage of the phase-sensitive imaging systems is reduced patient dose. The current dissertation research has not profoundly investigated how much dose reduction is achievable with the high or mid-energy in-line phase-sensitive system while the diagnostic or screening performance of the new technology is preserved or confined within the accepted range of the sensitivity and specificity for breast cancer among women with different breast compositions, breast sizes, and from different age groups.

During the observer preference study, it was noted that the PBT system has more impact on identifying the suspected lesions in the less experienced observer than on the experienced users. The interpretation of breast x-ray imaging is a challenging task and is highly dependent on the skill and experience of the interpreting physicians. The sample size in the phantom study and observer pool was limited in this investigation to draw a solid conclusion. Future clinical studies might be necessary, combining the highly experienced mammographers and relatively less experienced radiologists with basic breast image interpretation skills to confirm the stated finding. If it turns out to be the

case, phase-sensitive imaging will offer great potential in early cancer detection, where a significant portion of screening tests are interpreted by less experienced radiologists, especially in developing countries.

Reference

1. Fitzmaurice C, Abate D, et al. "Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: a systematic analysis for the global burden of disease study," *JAMA Oncology*, **5**(12): pp. 1749-68, 2019.
2. Siegel RL, Miller KD, et al. "Cancer Statistics, 2021," *CA: A Cancer Journal for Clinicians,* pp. 7-33, 2021.
3. Wang L, "Early diagnosis of breast cancer," *Sensors,* **17**(7): pp. 1572, 2017.
4. DeSantis CE, Fedewa SA, et al. "Breast cancer statistics, 2015: Convergence of incidence rates between black and white women," *CA: A Cancer Journal for Clinicians,* **66**(1): pp. 31-42, 2016.
5. Tabár L, Dean PB, et al. "The impact of mammography screening on the diagnosis and management of early-phase breast cancer," *Breast Cancer,* New York City: Springer, pp. 31-78, 2014.
6. Heath M, Bowyer K, Kopans D, et al. "Current status of the digital database for screening mammography," *Digital Mammography*, Netherlands: Springer, pp. 457-460, 1998.
7. Dromain C, Balleyguier C. "Contrast-enhanced digital mammography," *Digital Mammography*, Berlin Heidelberg: Springer, pp.187-198, 2010.
8. Pisano ED, Hendrick RE, Yaffe MJ, et al. "Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST," *Radiology*, **246**(2): pp. 376-83, 2008.

9. US Preventive Services Task Force. Screening for breast cancer: US Preventive Services Task Force recommendation statement. Annals of internal medicine, **151**(10): pp. 716, 2009.

10. Sree SV, Acharya RU, et al. "Breast imaging: a survey," *World Journal of Clinical Oncology,* **2**(4): pp. 171-178, 2011.

11. Levenson VV. "Biomarkers for early detection of breast cancer: what, when, and where?" *Biochimica et Biophysica Acta (BBA)-General Subjects*, **1770**(6): pp. 847-856, 2007.

12. Hellquist BN, Czene K, et al. "Effectiveness of population-based service screening with mammography for women ages 40 to 49 years with a high or low risk of breast cancer: socioeconomic status, parity, and age at birth of first child," *Cancer*, **121**(2): pp. 251-258, 2015.

13. Onega T, L Goldman E, et al. "Facility Mammography Volume in Relation to Breast Cancer Screening Outcomes," *Journal of Medical Screening*, **23**(1): pp. 31-37, 2016.

14. Ozmen N, Dapp R, et al. "Comparing Different Ultrasound Imaging Methods for Breast Cancer Detection," *IEEE TRANSACTIONS ON ULTRASONICS, FERROELECTRICS, AND FREQUENCY CONTROL*, **62**(4): pp. 637-646, 2015.

15. Niell BL, Freer PE, et al. "Screening for Breast Cancer," *Radiologic Clinics of North America*, **55**(6): pp. 1145-1162, 2017.

16. Hassan AM, El-Shenawee M. "Review of electromagnetic techniques for breast cancer detection," *IEEE Reviews in Biomedical Engineering*, **4**: pp. 103-118, 2011.

17. Roganovic D, Djilas D, et al. "Breast MRI, digital mammography and breast tomosynthesis: comparison of three methods for early detection of breast cancer," *Bosnian Journal of Basic Medical Sciences*, **15**(4): pp. 64-68, 2015.

18. Xu P, et al. "SU-E-I-81: Targeting of HER2-Expressing Tumors with Dual PET-MR Imaging Probes," *Medical Physics*, **42**(6): pp. 3260, 2015.

19. Andersson I, Ikeda DM, et al. "Breast tomosynthesis and digital mammography: a comparison of breast cancer visibility and BIRADS classification in a population of cancers with subtle mammographic findings," *European Radiology*, **18**(12): pp. 2817-2825, 2008.

20. Houssami N, Macaskill P, et al. "Breast screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading–evidence to guide future screening strategies," *European Journal of Cancer*, **50**(10): pp.1799-1807, 2014.

21. Greenberg JS, Javitt MC, et al. "Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice," *American Journal of Roentgenology*, **203**(3): pp. 687-693, 2014.

22. Law J "The development of mammography," *Physics in Medicine and Biology*, **51**: pp. R155-R167, 2006.

23. Gennaro G, Di Maggio C. "Dose comparison between screen/film and full-field digital mammography," *European Radiology*, **16**(11): pp. 2559-66, 2006.

24. Wilkins SW, Gureyev TE, Stevenson AW., et al. "Phase-contrast imaging using polychromatic hard X-rays," *Nature*, **384**(6607): pp. 335-338, 1996.

25. Arfelli F, Assante M, Bonvicini V, Bravin A, Pani S., et al. "Low-dose phase contrast x-ray medical imaging," *Physics in Medicine and Biology*, **43**(10): pp. 2845, 1998.

26. Wilkins SW, Gureyev TE, Stevenson AW., et al. "Phase-contrast imaging using polychromatic hard X-rays," *Nature*, **384**(6607): pp. 335-338, 1996.

27. Wu X, Dean A, Liu H. "X-ray diagnostic techniques," *Biomedical Photonics Handbook*. Ed. Tuan Vo-Dinh. 2. Vol. 2. Tampa: CRC Press, pp. 1-34, 2003.

28. Lewis RA "Medical phase-contrast x-ray imaging: current status and future prospects." *Physics in Medicine and Biology*, **49**(16): pp. 3573-3583, 2004.

29. Fitzgerald R. "Phase-Sensitive X-Ray Imaging: Fitzgerald R. Phase-sensitive X-ray imaging; new approaches that can detect X-ray phase shifts within soft tissues show promise for clinical and biological applications," *Physics Today*, **53**: pp. 23-28, 2000.

30. Kiss MZ, Sayers DE, Zhong Z. "Measurement of image contrast using diffraction enhanced imaging," *Physics in Medicine and Biology*, **28**: pp. 325-340, 2003.

31. Takeda T, Momose A, Itai Y., et al. "Human Carcinoma: Early Experience with Phase-Contrast X-ray CT with Synchrotron Radiation-comparative Specimen Study with Optical Microscopy," *Radiology*, **214**(1): pp. 298-301, 2000.

32. Wu X, Liu H "Clarification of aspects in in-line phase-sensitive x-ray imaging," *Medical Physics*, **34**(2): pp.737-743, 2007.

33. Bravin A, Coan P, Suortti P. "X-ray phase-contrast imaging: from pre-clinical applications towards clinics," *Physics in Medicine and Biology*, **58**(1): pp. R1-R35, 2012.

34. Gureyev TE, Mayo SC, Myers DE, Wilkins SW., et al. "Refracting Röntgen's rays: propagation-based x-ray phase contrast for biomedical imaging," *Journal of Applied Physics*, **105**(10): pp. 102005, 2009.

35. Donnelly EF, Price RR, Pickens DR, "Quantification of the effect of system and object parameters on edge enhancement in phase-contrast radiography," *Medical Physics*, **30**(11): pp. 2888-2896, 2003.

36. Diemoz PC, Bravin A, Reiser M, et al. "A method for high-energy, low-dose mammography using edge illumination x-ray phase-contrast imaging," *Physics in Medicine and Biology*, **61**(24): pp. 875, 2016.

37. Zhang D, Donovan M, Liu H, et al. "Preliminary feasibility study of an in-line phase-contrast x-ray imaging prototype," *IEEE Transactions on Biomedical Engineering*, **55**(9): pp. 2249-2257, 2008.

38. Zhao Y, Brun E, Coan P, Bravin A, et al. "High-resolution, low-dose phase contrast X-ray tomography for 3D diagnosis of human breast cancers," *Proceedings of the National Academy of Sciences*, **109**(45): pp. 18290-18294, 2012.

39. Ghani MU, Wong MD, Liu H, et al. "Detectability comparison of simulated objects within a dense breast phantom using high energy x-ray phase-sensitive and conventional imaging systems," Ed. SPIE BiOS. *International Society for Optics and Photonics*, pp. 100650K-100650K, 2017.

40. Donnelly EF, Lewis KG, Price RR, et al. "Characterization of the phase-contrast radiography edge-enhancement effect in a cabinet x-ray system," *Physics in Medicine and Biology*, **51**(1): pp. 21, 2005.

41. Wu D, Yan A, Wu X, Liu H, et al. "Characterization of a high-energy in-line phase contrast tomosynthesis prototype," *Medical Physics*, **42**(5): pp. 2404-2420, 2015.

42. Matsuo S, Katafuchi T, Fujita H, et al. "Evaluation of edge effect due to phase-contrast imaging for mammography," *Medical Physics*, **32**(8): pp. 2690-2697, 2005.

43. Wong MD, Yan A, Liu H, et al. "Dose and detectability improvements with high energy phase-sensitive x-ray imaging in comparison to low energy conventional imaging," *Physics in Medicine and Biology*, **59**(9): pp. N73, 2014.

44. Hammonds JC, Price RR, Donnelly EF, Pickens DR. "Phase-contrast digital tomosynthesis," *Medical Physics*, **38**(5): pp. 2353-2358, 2011.

45. Wu x, Liu H. "X-Ray cone-beam phase tomography formulas based on phase-attenuation duality," *Optic Express*, **13**(16): p. 6000-6014, 2005.

46. Wu X, Liu H. "A general theoretical formalism for X-ray phase-contrast imaging," *Journal of X-Ray Science and Technology*, **11**(1): pp. 33-42, 2003.

47. Beltran MA, Paganin DM, et al. "2D and 3D X-ray phase retrieval of multi-material objects using a single defocus distance," *Optic Express*, **18**(7): pp. 6423-6436, 2010.

48. Wu X, Yan A. "Phase retrieval from one single phase-contrast X-ray image," *Optic Express*, **17**(13): pp. 11187-11196, 2009.

49. Groso A, Abela R, Stampanoi M. "Implementation of a fast method for high-resolution phase-contrast tomography," *Optic Express*, **14**(18): pp. 8103-8110, 2006.

50. Wu X, Liu H, Yan A. "X-ray phase attenuation duality and phase retrieval," *Optic Letters*, **30**(4): pp. 379-381, 2005.

51. Omoumi F, Ghani MU, Liu H, et al. "The Potential of Utilizing Mid-Energy X-Rays for In-Line Phase Sensitive Breast Cancer Imaging," *Biomedical Spectroscopy and Imaging*, **9**(3-4): pp. 89-102, 2020.

52. Lauby-Secretan B, Scoccianti C, et al. "Breast-cancer screening—viewpoint of the IARC Working Group," *New England Journal of Medicine*, **372**(24): pp. 2353-2358, 2015.

53. Lee CH, Dershaw DD, Kopans D, et al. "Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer," *Journal of American College of Radiology*, **7**(1): pp. 18-27, 2010.

54. Siu AL. "Screening for breast cancer: US Preventive Services Task Force recommendation statement," *Annals of Internal Medicine*, **164**(4): pp. 279-296, 2016.

55. Wellings E, Vassiliades L, Abdalla R. "Breast cancer screening for high-risk patients of different ages and risk-which modality is most effective?" *Cureus*, **8**(12), 2016.

56. Ozmen N, Dapp R, Zapf M, et al. "Comparing different ultrasound imaging methods for breast cancer detection," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **62**(4): pp. 637-646, 2015.

57. Sardanelli F, Boetes C, et al. "Magnetic resonance imaging of the breast: recommendations from the EUSOMA working group," *European Journal of Cancer*, **46**(8): pp. 1296-1316, 2010.

58. Ahmad A, Barrington S, et al. "Use of positron emission tomography in evaluation of brachial plexopathy in breast cancer patients," *British Journal of Cancer*, **79**(3): pp. 478-482, 1999.

59. Greene LR, Wilkinson D. "The role of general nuclear medicine in breast cancer," *Journal of Medical Radiation Sciences*, **62**(1): pp. 54-65, 2015.

60. Koutalonis M, Delis H, et al. "Contrast-to-noise ratio in magnification mammography: a Monte Carlo study," *Physics in Medicine and Biology*, **52**(11): pp. 3185-3199, 2007.

61. Funke M, Breiter N, Hermann KP, et al. "Magnification survey and spot view mammography with a new microfocus X-ray unit: detail resolution and radiation exposure," *European Radiology*, **8**(3): pp. 386-390, 1998.

62. Kalef-Ezra JA, K. A. "Electron density of tissues and breast cancer radiotherapy: a quantitative CT study," *International Journal of Radiation Oncology* Biology* Physics,* **41**(5): pp. 1209-1214, 1998.

63. Al-Bahri JS, S. N. "Electron density of normal and pathological breast tissues using a Compton scattering technique," *Applied Radiation and Isotopes,* **49**(12): pp. 1677-1684, 1998.

64. Krantz DH. "Threshold theories of signal detection," *Psychological Review*, **76**(3): pp. 308-324, 1969.

65. Wickens TD. *Elementary signal detection theory*. U.S.S: Oxford University Press, 2002.

66. Macmillan NA, Creelman CD. *Detection theory: A user's guide*. Psychology press 2004.

67. Tanner WP, Swets JA. "A decision-making theory of visual detection," *Psychological Review*, **61**(6): pp. 401-409, 1954.

68. Gelfand DW, Ott DJ. "Methodologic considerations in comparing imaging methods," *American Journal of Roentgenology*, **144**(6): pp. 1117-1121, 1985.

69. Obuchowski NA. "Receiver operating characteristic curves and their use in radiology," *Radiology*, **229**(1): pp. 3-8, 2003.

70. DeLong ER, DeLong DM, Clarke-Pearson DL. "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," *Biometrics*, **44**(3): pp. 837-845, 1998.

71. McClish DK. "Analyzing a portion of the ROC curve," *Medical Decision Making*, **9**(3): pp. 190-195, 1989.

72. Metz CE, Wang PL, Kronman HB. "A new approach for testing the significance of differences between ROC curves measured from correlated data," *Information Processing in Medical Imaging*, Dordrecht: Springer, pp. 432-445, 1984.

73. Swets JA, Pickett RM. "Statistical Design of a Performance Test," *Evaluation of Diagnostic Systems*. New York City: Academic Press, pp. 68-79, 1982,

74. Efron, B. *The jackknife, the bootstrap, and other resampling plans*. Philadelphia: Society for industrial and applied mathematics, 1982.

75. McNeil BJ, Hanley JA. "Statistical approaches to the analysis of receiver operating characteristic (ROC) curves," *Medical Decision Making*, **4**(2): pp. 137-150, 1984.

76. Dorfman DD, Berbaum KS, Metz CE. "Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method," *Investigative Radiology*, **27**(9): pp. 723-731, 1992.

77. Obuchowski NA, "Multi-reader ROC studies: a comparison of study," *Academic Radiology*, **2**(8): pp. 709-716, 1995.

78. Obuchowski NA, Rockette HE. "Hypothesis testing of the diagnostic accuracy for multiple diagnostic tests: an ANOVA approach with dependent observations," *Communications in Statistics-simulation and Computation*, **24**(2): pp. 285-30, 1995.

79. Song HH. "Analysis of correlated ROC areas in diagnostic testing," *Biometrics*, **53**(1): pp. 370-382, 1997.

80. Beiden SV, Wagner RF, Campbell G. "Components-of-variance models and multiple-bootstrap experiments: an alternative method for random-effects, receiver operating characteristic analysis," *Academic Radiology*, **7**(5): pp. 341-349, 2000.

81. Ishwaran H, Gatsonis CA. "A general class of hierarchical ordinal regression models with applications to correlated ROC analysis," *The Canadian Journal of Statistics*, **28**(4): pp. 731-750, 2000.

82. Starr SJ, Metz CE, Lusted LB, Goodenough DJ. "Visual detection and localization of radiographic images," *Radiology*, **116**(3): pp. 533-538, 1975.

83. Obuchowski NA, Lieber ML, Powell KA. "Data analysis for detection and localization of multiple abnormalities with application to mammography," *Academic Radiology*, **7**(7): pp. 516-525, 2000.

84. Chakraborty DP, Yoon HJ. "Operating characteristics predicted by models for diagnostic tasks involving lesion localization," *Medical Physics*, **35**(2): pp. 435-445, 2008.

85. Chakraborty DP, Winter LH. "Free-response methodology: alternate analysis and a new observer-performance experiment," *Radiology*, **174**(3): pp. 873-881, 1990.

86. Chakraborty DP. "Validation and statistical power comparison of methods for analyzing free-response observer performance studies," *Academic Radiology*, **15**(12): pp. 1554-1566, 2008.

87. Wilkins SW, Gureyev TE, et al. "Phase-contrast imaging using polychromatic hard X-rays," *Nature*, **384**(6607): pp. 335-338, 1996.

88. Ingal VN, Beliaevskaya EA. "X-ray plane-wave topography observation of the phase contrast from a non-crystalline object," *Journal of Physics D: Applied Physics*, **28**(11): pp. 2314, 1995.

89. Momose, A. "Phase-sensitive imaging and phase tomography using X-ray interferometers," *Optics Express,* **11**(19): pp. 2303-2314, 2003.

90. Weitkamp T, Diaz A, David C, Pfeiffer F, et al. "X-ray phase imaging with a grating interferometer," *Optics Express*, **13**(16): pp. 6296-6304, 2005.

91. Olivo A, Arfelli F, Cantatore G, et al. "An innovative digital imaging set-up allowing a low-dose approach to phase contrast applications in the medical field," *Medical Physics*, **28**(8): pp. 1610-1619, 2001.

92. Wu X, Liu H, Yan A. "Robustness of a phase-retrieval approach based on phase-attenuation duality," *Journal of X-Ray Science and Technology*, **15**(2): pp. 85-95, 2007.

93. Ghani MU, Wong MD, et al. "Detectability comparison between a high energy x-ray phase-sensitive and mammography systems in imaging phantoms with varying glandular-adipose ratios," *Physics in Medicine and Biology*, **62**(9): pp. 3523-3538, 2017.

94. Berggren K, Eriksson M, et al. "In vivo measurement of the effective atomic number of breast skin using spectral mammography," *Physics in Medicine and Biology*, **63**(21): pp. 215023-215031, 2018.

95. Boone JM. "Normalized glandular dose (DgN) coefficients for arbitrary x-ray spectra in mammography: Computer-fit values of Monte Carlo derived data," *Medical Physics*, **29**(5): pp. 869-875, 2002.

96. Wu X, Gingold EL, et al. "Normalized average glandular dose in molybdenum target-rhodium filter and rhodium target-rhodium filter mammography," *Radiology*, **193**(1): pp. 83-89, 1994.

97. K.R. Bijkerk KR, et al., "Manual CDMAM Phantom," pp. 10-11, 1995.

98. Liu H, Karellas A, et al. "Lesion detectability considerations for an optically-coupled CCD x-ray imaging system," *IEEE Transactions on Nuclear Science*, **41**(4): pp. 1506-1509, 1994.

99. Boone JM, Seibert JA. "A figure of merit comparison between bremsstrahlung and monoenergetic x-ray sources for angiography," *Journal of X-ray Science and Technology*, **4**(4): pp. 334-45, 1994.

100. Burger, G C. "The perceptibility of details in roentgen examinations of the lung," *Acta Radiologica*, **31**(3): pp. 193-222, 1949.

101. Goodenough DJ, Rossmann K, Lusted LB. "Radiographic applications of signal detection theory," *Radiology*, **105**(1): pp. 199-200, 1972.

102. Metz, C E. "ROC analysis in medical imaging: a tutorial review of the literature," *Radiological Physics and Technology*, **1**(1): pp. 2-12, 2008.

103. Beam CA, Layde PM, Sullivan DC. "Variability in the interpretation of screening mammograms by US radiologists: findings from a national sample," *Archives of Internal Medicine*, **156**(2): pp. 209-213, 1996.

104. Hendrick RE, Cole EB, Pisano ED, et al. "Accuracy of soft-copy digital mammography versus that of screen-film mammography according to digital manufacturer: ACRIN DMIST retrospective multi-reader study," *Radiology*, **247**(1): pp. 38-48, 2008.

105. Burgess, A E. "Visual Perception Studies and Observer Models in Medical Imaging," *Seminars in Nuclear Medicine*, **41**(6): pp. 419-436, 2011.

106. Burgess, A E. "Comparison of receiver operating characteristic and forced choice observer performance measurement methods," *Medical Physics*, **22**(5): pp. 643-655, 1995.

107. Kramers H A. "XCIII. On the theory of X-ray absorption and of the continuous X-ray spectrum," *The London Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **46**(275): pp. 836-871, 1923.

108. Birch R., Marshall M. "Computation of bremsstrahlung x-ray spectra and comparison with spectra measured with a Ge (Li) detector," *Physics in Medicine and Biology*, **24**(3): pp. 505, 1979.

109. J.M, Boone. "The three parameter equivalent spectra as an index of beam quality," *Medical Physics*, **15**(3): pp. 304-310, 1988.

110. Tucker, D.M., Barnes G.T., Chakraborty D.P. "Semiempirical model for generating tungsten target x-ray spectra," *Medical Physics*, **18**(2): pp. 211-218, 1991.

111. Ay M.R., et al. "Monte Carlo simulation of x-ray spectra in diagnostic radiology and mammography using MCNP4C," *Physics in Medicine and Biology*, **49**(21): pp. 4897, 2004.

112. Omoumi FH, Wu X, et al. "Mathematical estimation of half-value layer thicknesses," *Journal of Applied Clinical Medical Physics*, **22**(10): pp. 320-328, 2021.

113. J.H., Hubbell. "Review of photon interaction cross-section data in the medical and biological context," *Physics in Medicine and Biology*, **44**(1): pp. R1, 1999.

114. Henke BL, Gullikson EM, Davis JC. "X-ray interactions: photoabsorption, scattering, transmission, and reflection at E= 50-30,000 eV, Z= 1-92," *Atomic Data and Nuclear Data Tables,* **54**(2): pp. 181-342, 1993.

115. Berger M.J., Hubbell J. H. "XCOM: Photon cross-sections on a personal computer," National Bureau of Standards, Washington, D.C.: Center for Radiation Research, 1987.

116. https://physics.nist.gov/PhysRefData/XrayMassCoef/ComTab/air.html

117. Bevington PR, Robinson DK. *Data Reduction and Error Analysis*. New York: McGraw-Hill, 2003.

118. Press WH, Teukolsky SA, et al. *Numerical Recipes with Source Code CD-ROM*. Ed. 3rd. The Art of Scientific Computing. Cambridge University Press, 2007.

119. Mathieu K.B., et al. "An empirical model of diagnostic x-ray attenuation under narrow-beam geometry," *Medical Physics*, **38**(3): pp. 4546-4555, 2011.

120. Poludniowski G.G., Philip M.E. "Calculation of x-ray spectra emerging from an x-ray tube. Part I. Electron penetration characteristics in x-ray targets," *Medical Physics*, **34**(6): pp. 2164-2174, 2007.

121. Poludniowski, G G. "Calculation of x-ray spectra emerging from an x-ray tube. Part II. X-ray production and filtration in x-ray targets," *Medical Physics*, **34**(6): pp. 2175-2186, 2007.

122. Tse J, McLean D. "Investigation of the clinical performance of a novel solid-state diagnostic dosimeter," *Journal of Applied Clinical Medical Physics*, **16**(4): pp. 244-253, 2015.

123. Hupfer M, Kolditz D, Nowak T, Eisa F, Brauweiler R, Kalender WA. "Dosimetry concepts for scanner quality assurance and tissue dose assessment in micro-CT," *Medical Physics*, **39**(2): pp. 658-70, 2012.

124. Golubicic I, Borojevic N, Pavlovic T. "Risk factors for breast cancer: is ionizing radiation among them," *Journal of BUON,* 2008. **13**(4): pp. 487-94, 2008.

125. Ghani MU, Wong MD, Omoumi FH, et al. "Detectability comparison of simulated tumors in digital breast tomosynthesis using high-energy X-ray inline phase-sensitive and commercial imaging systems," *Physica Medica,* **47**: pp. 34-41, 2018.

126. Wu X, Liu H. "Clinical implementation of x-ray phase-contrast imaging: Theoretical foundations and design considerations," *Medical Physics*, **30**(8): pp. 2169-79, 2003.

127. Wu X, Liu H. "An experimental method of determining relative phase-contrast factor for x-ray imaging systems," *Medical Physics*, **31**(5): pp. 997-1002, 2004.

128. Sobol WT, Wu X. "Parametrization of mammography normalized average glandular dose tables," *Medical Physics*, **24**(4): pp. 547-54, 1997.

129. Rose, Albert. Vision: Human and Electronic. United States, Springer US, 2013.

130. Hendrick, R. Edward. *Breast MRI: fundamentals and technical aspects*. Vol. 2. New York: Springer, 2008.

131. Ghani MU, Fajardo LL, Omoumi F, Yan A, Jenkins P, Wong M, Li Y, Peterson ME, Callahan EJ, Hillis SL, Zheng B. "A phase-sensitive x-ray breast tomosynthesis system: Preliminary patient images with cancer lesions," *Physics in Medicine and Biology*, **66**(21): pp. 21LT01, 2021.