Stony Brook University

Low-Contrast Lesion Detection in Tomosynthetic Breast Imaging

A Dissertation Presented

by

Lili Zhou

to

The Graduate School

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Electrical & Computer Engineering

Stony Brook University

May 2007

Stony Brook University

The Graduate School

Lili Zhou

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Dr. Gene Gindi
Dissertation Advisor
Associate Professor of Electrical & Computer Engineering and Radiology

Dr. Murali Subbarao
Chairperson of the Defense
Professor of Electrical & Computer Engineering

Dr. John Murray
Associate Professor of Electrical & Computer Engineering

Dr. Klaus Mueller
Associate Professor of Computer Science

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Low-Contrast Lesion Detection in Tomosynthetic Breast Imaging

by

Lili Zhou

Doctor of Philosophy

in

Electrical & Computer Engineering

Stony Brook University

2007

Conventional 2D mammography is currently the most effective approach to detecting early stage breast cancer. Tomosynthetic breast imaging is a potentially more valuable 3D technique for breast cancer detection. This technique acquires a limited number of noisy 2D projection images over a limited angular range and then mathematically reconstructs a 3D breast. In breast imaging, the contrast between cancerous lesions and background features is fairly small; furthermore, the presence of lesions is masked by normal anatomical structures. The low contrast, anatomical masking and effects of radiation noise all combine to make lesion detection a difficult task.

In this simulation study, we investigate the efficacy of three tomosynthetic reconstruction algorithms simple backprojection, algebraic, and statistical in the context of an especially difficult lesion detection task. This

is the detection of a very low-contrast mass embedded in a very dense fibro-glandular tissue background a clinically useful task for which tomosynthesis may be well suited. The project uses anatomically realistic 3D breast phantoms whose normal anatomic variability limits lesion conspicuity. In order to capture anatomical object variability, we generate an ensemble of 3D breast phantoms by using stochastic algorithms, each of which results in random instances of various breast structures. Power-law structural noise is added to simulate small-scale object variability. The irregular mass is simulated by a 3D random walk algorithm. Low-dose data are acquired using an isocentric geometry and simulated Poisson radiation noise is added. The data are then reconstructed via the three types of methods. Reconstructed slices through the center of the lesion are presented to human observers in a two-alternative-forced-choice test to measure lesion detectability by computing the area under the ROC curve. We conclude that backprojection algorithms perform significantly more poorly compared to the other two types of reconstruction methods.

In memory of my mother

Contents

List of Acronyms

2AFC	Two-Alternative-Forced-Choice
2D	Two-Dimensional
3D	Three-Dimensional
ART	Algebraic Reconstruction Technique
a-Se	amorphous Selenium
ASF	Artifact Spread Function
AUC	Area Under the ROC curve
A_{LROC}	Area Under the LROC Curve
A_{ROC}	Area Under the ROC curve
BKE	Background-Known-Exactly
BKS	Background-Known-Statistically
BP	Backprojection
CAD	Computer-aided Diagnosis
CC	Cranial Caudal
CCD	Charge Coupled Device
CHO	Channelized Hotelling
CNR	Contrast-to-Noise Ratio
COR	Center Of Rotation
COSEM	Complete-data Ordered Subset Expectation Maximization
CsI	Cesium Iodide
CT	Computed Tomography

DFT	Discrete Fourier Transform
DQE	Detective Quantum Efficiency
EM	Expectation-Maximization
FBP	Filtered Back-Projection
FE	Finite Element
FFT	Fast Fourier Transform
FNF	False Negative Fraction
FOM	Figure Of Merit
FPF	False Positive Fraction
keV	kilo electron Volts
kV	kilo Voltage
kVp	kilo Voltage peak
LROC	Localization Receiver Operating Characteristic
MAP	Maximum *A Posteriori*
MITS	Matrix Inversion Tomosynthesis
ML	Maximum Likelihood
MLO	Mediolateral Oblique
Mo	Molybdenum
NBP	Nonlinear Backprojection
OS	Ordered Subsets
OSEM	Ordered Subsets EM
OS-SPS	Ordered Subsets-Separable Paraboloidal Surrogate
PCG	Preconditioned Conjugate Gradient
PSF	Point Spread Function
Rh	Rhodium
ROC	Receiver Operating Characteristic

ROI	Region of Interest
SART	Simultaneous Algebraic Reconstruction Technique
SC	Scatter Correction
SIRT	Simultaneous Iterative Reconstruction Technique
SKE	Signal-Known-Exactly
SKS	Signal-Known-Statistically
SNR	Signal-to-Noise Ratio
SPECT	Single-Photon Emission Computed Tomography
SPR	Scatter-to-Primary Ratio
SPS	Separable Paraboloidal Surrogate
TACT	Tuned Aperture Computed Tomography
TDLU	Terminal Duct Lobule Unit
TNF	True Negative Fraction
TPF	True Positive Fraction
TRIOT	Transmission Incremental Optimization Transfer

List of Figures

Acknowledgments

I would like to express my deep thanks to my dissertation advisor, Gene Gindi. Without his patience, encouragement and guidance, I would not been able to finish this dissertation. He has been consistently generous with his time and enthusiasm, and excellent in teaching his students. I enjoy working with Gene.

I would also like to thank the rest of my dissertation committee - Murali Subbarao, John Murray and Klaus Mueller - for their precious time and valuable suggestions.

All the members of the Medical Imaging Processing Laboratory at Stony Brook University contributed in many ways to this dissertation: Santosh Kulkarni, Parmeshwar Khurd and Yuxiang Xing for introducing me to some software packages, especially Santosh Kulkarni for helping me solving some problems; Jorge Oldan for software package of 2AFC test; Chintu Pandya for technical support. I am specially thankful to those who volunteered as subjects for the human observer studies: Gene Gindi, Bo Zhao, Dan Li, Jun Zhou, Santosh Kulkarni, Parmeshwar Khurd and Ing-Tsung Hsiao. I deeply thank Paul Fisher for his valuable suggestions and extensive comments on the 3D breast phantom. I would like to thank Fang Xu and Klaus Mueller for the slicer and placer display tools; Jun Zhou, Jerome Liang, Wei Zhao, Jennifer Segui, Jing Wang, Bo Zhao and Su Wang for helpful discussions. I highly appreciate their contributions to this work.

I would like to thank Dan Li, Bo Zhao, Su Wang and Jing Wang for making my daily life in Stony Brook interesting and enjoyable.

Finally, I would like to thank my parents, Chunxia Wang and Aiqing Zhou for their support and encouragement in all my educational endeavors. I also deeply thank my brother, Xiaolong Zhou for his emotional support and his accompanying my father while I was pursuing my Ph.D. degree at Stony Brook. I greatly acknowledge Yongjun Jiao for his love and support during our seven years marriage.

Chapter 1

Introduction

Breast cancer is the second most common type of cancer in women and the second leading cause of cancer-related deaths among women [1]. One in eight women in the United States will develop breast cancer during her lifetime [2]. Approximately 200,000 women in the United States are diagnosed with breast cancer each year, and the disease causes about 40,000 deaths annually [2]. X-ray mammography is currently the most effective method of detecting early stage breast cancer, and has played an important role in reducing the breast cancer rate. It has been shown that the use of screening mammography has reduced mortality from breast cancer by 20%-40% [3, 4].

Conventional mammography is a two-dimensional (2D) breast imaging method relying on a pair of 2D x-ray images of the breast, which are obtained from two different directions: top-to-bottom and side-to-side. The breast is pulled away from the body, and compressed between two plastic plates. In regular screen film mammography, two x-ray views for each breast are recorded on film. It is suggested that screen-film mammography is and will continue to be a valuable tool for detection and diagnosis of breast cancer [5]. However, screen-film mammography has some limitations on displaying the finest features due to inadequate contrast, therefore resulting in a limited sensitivity for the detection of breast cancer with dense breast [6]. Film does not have a linear sensitivity to photon-flux [7], there is a narrow range over which it can detect small difference in contrast. Also film requires processing time and storage space. An-

other limitation is the effect of structural noise due to film granularity, degrading the visibility of micro-calcifications and other fine breast structures [5]. These limitations can be effectively overcome with a digital mammography system.

In digital mammography these two views of images (from top to bottom and from side to side) are recorded on a high-resolution digital detector whose size is about the same as a film-screen cassette ($18cm \times 24cm$ or $24cm \times 30cm$). With digital mammography, a electronic x-ray detector is used to record the breast image and to convert the image into a digital picture. After the digital mammogram is stored on a computer, it can be processed and displayed on a soft copy or hard copy device. One advantage of digital mammography is that its digital form allows the radiologist to alter the range and contrast of the image while viewing. Also, digital mammograms lend themselves to processing by Computer Aided Detection (CAD) systems that act to assist the radiologist in making diagnosis. The studies by Pisano *et al.* [6] suggested that the overall diagnostic accuracy of digital and film mammography for breast cancer was similar, but digital mammography is more accurate in a subpopulation of women with dense breasts. The improved performance in a digital mammography system is mainly associated with the x-ray detector and the display device. Therefore, despite the convenience of digital images obtained with digital mammography, and despite the superior performance of a digital detector, its clinical diagnostic efficacy is only a slight improvement on that obtained with conventional screen-film systems.

Mammographic features characteristic of breast cancer are masses, particularly ones with irregular margins [8], clusters of micro-calcifications, and architectural distortions of breast structures. In conventional 2D mammography using either screen-film or digital detectors, the cancer lesions present in one plane of the three-dimensional (3D) breast are sometimes difficult to visualize because of confounding anatomical structures in planes of above and below the one of interest. A major limitation in

conventional mammographic technique is that superimposed normal breast tissues generate a "structure noise" that obscures the breast cancer, especially in a dense breast, increasing the false-negative rate. Even without other sources of radiation or instrumentation noise, a cancer can be hidden by the anatomical background. It is this structure noise — also termed anatomical variability— that accounts for the only slight advantage of digital mammography over screen-film in spite of the superior detector performance of digital mammography. It has been shown that 30% of breast cancer may be missed by conventional (digital and screen-film) mammography [9]. On the other hand, the overlapped tissue structures may look like a cancerous tumor on a mammogram, and can be mistaken for abnormalities, causing false-positive callbacks for further diagnosis or even biopsy. In 2004, National Cancer Institute [10] reported that up to 12% normal breasts were mistakenly read as having breast cancer.

Three-dimensional mammographic imaging can make it possible to search through the interior of the breast plane-by-plane with minimal overlapping from the other tissues. One form of 3D mammographic imaging, digital breast tomosynthesis, is a new method that can reduce the tissue-overlap effect and retrieve depth information. This technique takes x-ray projection images of the breast at a limited number of views over a limited angular range. The final step in the tomosynthesis procedure is mathematically reconstructing the set of 2-D projection data to produce a 3D image of the breast, and hence showing the tissue structure in three dimensions. This dissertation focuses on reconstruction techniques and low-contrast lesion detection in digital breast tomosynthesis.

With the background information presented so far, we shall now state more precisely the central goal of the dissertation. Breast imaging, especially for detection of masses, is unique in radiography due to the very small amount of subject contrast in the breast [11]. In mammography, contrast is derived from the local differences

in composition and attenuation of breast tissues. The contrast between the mass and the background structure is fairly small, which presents a big challenge in lesion detection. Given the small contrast, a lesion can be mashed by anatomical variability. In order to capture anatomical variability, we generate an ensemble of realistic 3D breast phantoms, each of which comprises random instances of various breast structure components including adipose, fibrous connective tissue, ductal structures, pectoralis muscles and Cooper's ligaments. Low-contrast irregular masses are embedded in very dense fibro-glandular tissue background. In such a phantom, power-law structural noise is added as small scale object variability. With these phantoms, the 2D projection images are acquired using an isocentric tomosynthetic geometry at 11 views over 50^o as illustrated in Figure 8.1 (c), and noise is added to simulate radiation noise and detection noise. Several tomosynthetic methods are used to reconstruct 3D breast estimates. To investigate the efficacy of those reconstruction algorithms, a two-alternative-forced-choice (2AFC) test is adopted to measure lesion detectability by calculating the area under the ROC curve.

One limitation of our work is that we do not consider micro-calcifications. Some breast cancers are associated with micro-calcifications. Micro-calcifications are small (0.1mm-1mm) deposits of calcium that have a very high contrast relative to the surrounding normal or abnormal tissues. They can be present as either a very tiny single spot or a cluster of several spots. Three-dimensional breast imaging might help in some cases in better viewing of calcification clusters. Another form of 3D mammography is cone-beam mammographic computed tomography (CT). Here, special equipment is needed to view the breast from an unlimited 360^0 view as shown in Figure 2.10. Tomosynthesis produces limited angle reconstructions that are probably of lower quality than a cone-beam mammographic CT image, but construction of a breast tomosynthesis system requires only a small hardware modification of a conven-

tional mammographic system, while a mammographic CT system represents a major development project. Therefore, there is tremendous interest in breast tomosynthesis.

In summary, our objective is to create a realistic 3D breast phantom and use it to study the efficacy of different tomosynthetic reconstruction methods in the task of low-contrast lesion detection by a human observer.

This dissertation is organized as follows. Chapter 2 provides an overview of background information including breast anatomy, x-ray medical imaging physics, conventional mammography and 3D imaging mammography. A qualitative description of breast tomosynthesis is given. Chapter 3 addresses the problems and strategies in optimizing breast tomosynthesis systems. A mathematical imaging model is presented in chapter 4, which is used in subsequent chapters. In Chapter 5, I shall review various reconstruction methods used in transmission tomography. The construction of a realistic 3D breast phantom is described in Chapter 6. Chapter 7 provides some background on some topics related to task-based image quality assessment. Using 3D breast phantoms, we show how the projection data are acquired, and discuss three tomosynthetic methods used in our project, followed by the discussion of our experiment results in a low-contrast lesion detection task in Chapter 8. In Chapter 9, we discuss the future work and directions, and summarize our contributions.

Chapter 2

Background Information

In this chapter, I shall provide an overview of the general information about breast anatomy, X-ray imaging physics, conventional 2D mammography and 3D breast imaging.

2.1 Breast Anatomy: Normal and Abnormal

2.1.1 Breast Normal Composition

Many researchers, including us, build physical or digital "phantoms" to simulate the breast and use these phantoms in various imaging studies. A basic understanding of breast anatomy will help us in designing a digital breast phantom that we use in experiments described in subsequent chapters. The breast is attached to the chest wall. The skin of the breast is usually 0.5 to 2 mm in thickness [12]. The fascia layers lie beneath the skin. Blood vessels and lymphatics penetrate the fascial layers, running through the muscle and the other breast tissue. The pectoralis muscle is thick at the axilla, has a convex anterior border, spreads like a fan across the chest wall, and extends to the level of the nipple [13]. The pectoralis muscle permits breast traction when the breast is positioned and compressed.

In general, the breast is a grouping of glandular, fatty, and fibrous connective tissues positioned over the pectoral muscles of the chest wall and attached to the chest

wall by fibrous tissue called Cooper's ligaments [12]. A layer of fatty tissue surrounds the breast glands and extends throughout the breast. The fibrous connective tissues of the breast house the lobules (milk producing) and the ducts (milk passages). There are 8-20 major ducts that originate from the nipple [12]. Each of these ducts dilates into narrower and shorter branches until they form the terminal duct and its lobule (1-2mm) called the terminal duct lobule unit (TDLU), leading to a tree-like structure.

Adipose (fatty) and glandular tissues are the two predominant tissue types in the breast (Many researchers use a simple breast phantom composed of 50% adipose tissue (fat) /50% glandular tissue in their simulation models). The diameter of the major ducts varies between 2 and 4.5 mm [14], and the total length of the duct (the distance from the nipple to the TDLU) varies between 2 and 4.5 cm [15]. Cooper's ligaments are band-like connective tissues supporting the breast. The fat (subcutaneous adipose tissue) that covers the lobes gives the breast its size and shape. Each breast contains 15 to 20 lobes. Each lobe is composed of many lobules, at the end of which are tiny bulb-like glands[13]. Ducts connect the lobes, lobules, and glands. The ductal structure is an important component of the breast anatomy since practically most breast cancers originate in the ductal or lobular epithelium, with very few arising in the connective or adipose tissue [12]. Breast ducts also contribute significantly to the parenchymal pattern, the background texture (anatomical variability) seen in the mammogram. In chapter 6, we generate anatomically realistic 3D medium-sized breast phantoms, each of which comprises random instances of various breast structure components including adipose (fatty) tissue, fibrous connective tissue, ductal structures, pectoralis muscles and Cooper's ligaments.

2.1.2 Breast Abnormalities

Breast abnormalities can be divided into two major types: tumor masses and microcalcifications [16]. Tumor masses are groupings of abnormal cells with low-contrast

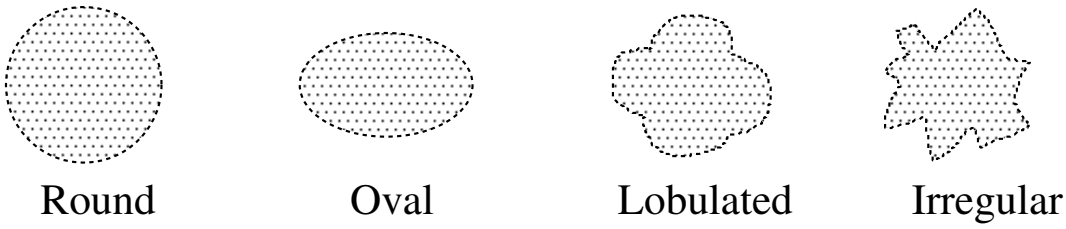


Figure 2.1: Mass shapes.

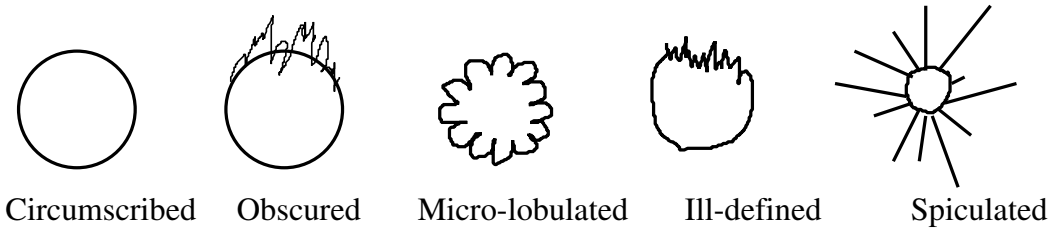


Figure 2.2: A mass can have one of five margins.

(relative to normal tissue) x-ray transmission attenuation, while microcalcifications are mineral deposits (calcium) with high attenuation. Both types can be differentiated as being malignant or benign. Microcalcifications have very high density and usually have a diameter range 0.1mm-1mm [12]. They can be present as either a tiny single spot or a cluster. Five or more calcifications, each with a diameter ≤ 0.5 mm, isolated in a cubic centimeter volume, have high probability of malignancy. Malignant calcifications are more irregular in shape, size and distribution [17].

Tumor masses vary in shape, size and density. The study by Kopans [12] suggested that $\geq 50\%$ benign tumor masses are $\leq$ 1cm in diameter. The American College of Radiology [18] classifies the mass shapes as round, oval, lobulated, irregular as shown in Figure 2.1. A mass can have one of five "margins" (small-scale boundary structure): circumscribed, microlobulated, obscured, ill-defined and spiculated, as illustrated in Figure 2.2. The more irregular in shape, the more possible in malignancy. The probability of malignancy is high in lesions with ill-defined margins [13]. A lesion that is round, oval, lobulated with sharply defined borders has a very high likelihood of being benign.

In chapter 6, we will describe the construction of 3D breast phantom in detail

based on the breast anatomy. Irregular tumor masses will be simulated as the breast abnormality in our project.

2.2 X-ray Production, Interaction and Detection

We now give an introduction to the medical physics of mammographic imaging. The whole process of x-ray production, interaction and detection can be described as an imaging chain. In this imaging chain, we'll emphasize some important factors.

2.2.1 X-ray Production

The imaging chain starts with x-ray production. The most common x-ray source is the standard rotating anode x-ray tube, as illustrated in Figure 2.3 [19]. The x-

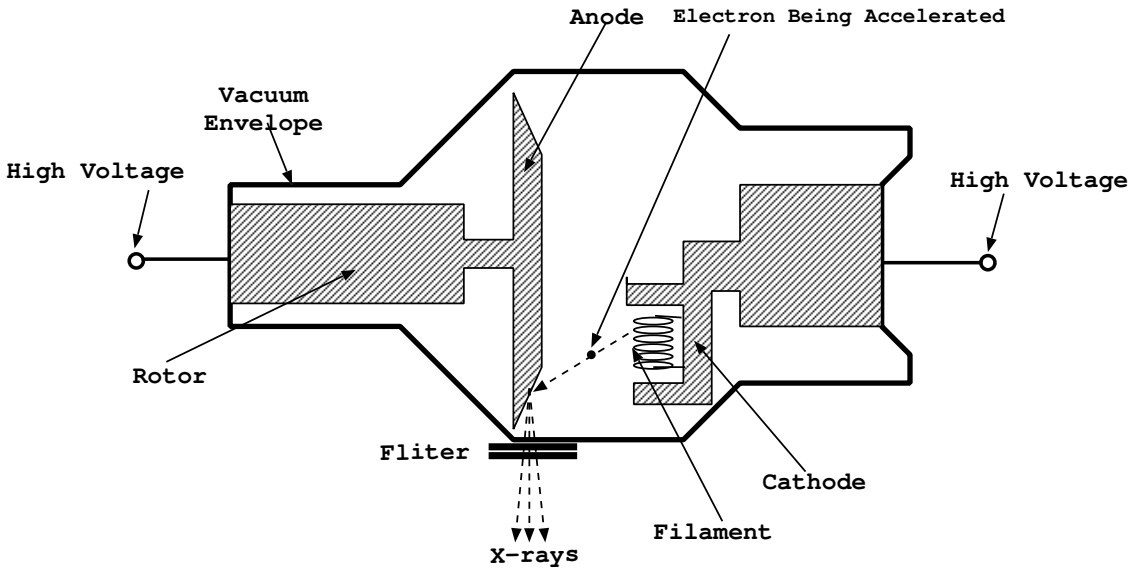


Figure 2.3: X-ray tube. See text for detailed explanations.

rays are produced from accelerated electrons in vacuum by applying a high voltage between the anode and the cathode. The applied potential is expressed in units kV (kilovolts). The peak kilo-voltage, kVp, is the maximum kV potential value applied

across the x-ray tube during the exposure [11]. A typical kVp for mammography is in a range 25-30 kVp, depending on the thickness of the breast. Tomosynthesis requires at least 2-3 kVp higher than that used in conventional mammography [20]. The electrons are emitted from a filament in the cathode. Emission occurs when the filament is heated by passing a current through it. When the filament is hot enough, some electrons obtain a thermal energy enough to overcome the binding energy. After being accelerated into the anode, most of the electrons' energy is converted into heating of the anode, but a small part is converted to x-rays. The energies of x-ray photons are expressed in electron volts (eV). The most common anode used for mammography are molybdenum, rhodium and tungsten.

Focal Spot

It is ideal to design the focal spot so that x-rays emanate from an approximate point source because any increase in source size will result in blurring of the final image. To avoid this blurring, the electrons must be focused to strike a small spot of the anode. However, there is a limit to the size of this focal spot because the anode material will melt if too much power is deposited into too small an area. This limit is improved by the use of a rotating beveled anode, where the anode target material is rotated about a central axis and new (cooler) anode material is constantly being rotated into place at the focal spot [19]. There is a compromise between the need to have a small spot size and the need to prevent the anode from melting. Incorporating a small anode beveled angle ($7^o - 15^o$) [11] allows the use of a large actual focal spot—actually a radial "stripe" along the anode, thus heat is dissipated over a large surface. The beam emerges at a shallow angle with respect to the anode, and so solid angle effects make the focal spot a very small point-like source. The effective focal spot size for x-ray tubes used in mammography is $0.15mm$ or $0.45mm$ [21].

X-Ray Spectrum

The emitted x-ray beam is characterized by its energy spectrum. The x-ray production is tuned to produce a spectrum that will result in good contrast in the 2D projection image. Below we describe contributing factors to this spectrum.

- Bremsstrahlung Radiation

  Part of the emitted spectrum is due to the fact that the deceleration of high-speed electrons upon hitting the anode results in emission of electromagnetic radiation, called ***bremsstrahlung radiation***, when the electrons interact with the nuclei of the anode. The bremsstrahlung spectrum is broad spectrum of low-energy radiation. The photons emitted due to ***bremsstrahlung radiation*** are generated with energies lower than the kVp.

- Characteristic X-rays

  Another component of the x-ray spectrum is due to ***characteristic radiation***. The production of characteristic x-rays begins when the accelerated electrons interact with atomic electrons in the anode, ejecting them from their electronic shells. Then, outer shell electrons fill in the empty shell, emitting characteristic x-rays at a few discrete energies. Since the energy of these characteristic x-rays depends on the material of the anode, we can choose specific materials to produce x-rays with desired energies. The most common anodes used for screen/film mammography are molybdenum, rhodium and tungsten. These also are used in breast tomosynthesis. The characteristic x-rays emitted by a molybdenum target are mostly at 17.4 keV energy with a smaller portion from 19.7 keV [12]. Those generated by rhodium target are mostly at 20.2 keV with a smaller portion at 22.8 keV. Those by tungsten are mostly at 59 keV with other very low energy peaks. The characteristic radiation of molybdenum and rhodium are

particularly valuable for breast imaging because their low-energy x-ray photons help enhance the low contrast between breast tissue and breast lesion.

X-Ray Filtration

The composition of the x-ray anode determines the x-ray spectrum. The spectrum can be further refined by using filters which allow photons of specific energies to pass through and block those unwanted with relatively high efficiency. The filter is made of sheets of metal, attached to the tube but not in the vacuum. If the appropriate filter is chosen, the x-ray spectrum can be narrowed down to a pretty small range to take advantage of better contrast from the low-energy photons. The filter material can be varied to match the anode material that has been selected. The filter material can, in fact, be the same as the anode material.

Composite Spectrum

The composite spectrum obtained from a typical mammography x-ray tube is shown in Figure 2.4. This spectrum is produced at 28 kVp with and without filtration. As one can see, the molybdenum anode spectrum is composed of bremsstrahlung and characteristic radiation. The filtered spectrum, using 25 μm rhodium (Rh) at 28 kVp, shows a distribution with no x-rays below about 5 keV, and a dramatic decrease of x-rays at below 17 keV and above 20 keV. The filtration eliminates the majority of low-and high-energy x-rays.

2.2.2 X-Ray Interaction with Matter

The next step along the imaging chain is the x-ray interaction with matter. When the produced x-ray beam passes through matter (i.e. the breast in mammography), it gets attenuated as photons are gradually removed from it. This attenuation takes place by two main processes: photoelectric absorption and Compton scattering [22].

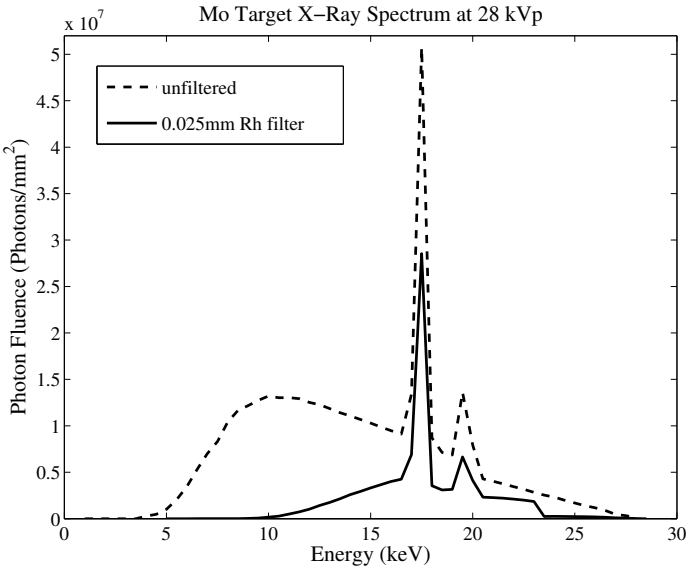


Figure 2.4: Energy spectrum of the X-rays emitted from a diagnostic mammography x-ray tube at 28 kVp. The anode is molybdenum. The peaks are due to characteristic radiation and the background is due to bremsstrahlung radiation. Adding 25 μm of a rhodium filter reduces much of the low-energy component of the x-ray spectrum, but also significantly attenuates x-rays in the unfiltered spectrum above 20 keV.

Photoelectric absorption is due to the interaction of an x-ray photon with a tightly bound inner electron in an atom. The result is that the x-ray photon imparts all of its energy to the inner electron, which uses some of this required energy to overcome the binding energy within its shell, the rest appearing as the kinetic energy of the freed electron [11]. Thus in photoelectric absorption, the x-ray photon is completely lost. On the other hand, Compton scattering is due to the interaction of the x-ray photon with either a free electron, or one that is only loosely bound in one of the outer shells of an atom [8]. As a result of this interaction, the x-ray photon is deflected onto a new path from its original direction with some loss of energy.

Subsequent detection of scattered x-ray photons clouds the image, reducing contrast and sharpness. In mammography the scattered photons degrade the visualization

of breast lesions. Barnes and Berzovich [23] found that the ratio of the scatter radiation to the primary beam might be 0.86 in a breast 6 cm thick.

Photoelectric absorption is the dominant attenuation mechanism in most x-ray interaction at mammographic energies. That is, more photons are absorbed than Compton scattered. The photoelectric absorption process can be used to amplify differences in attenuation between soft tissues, thereby improving image contrast. This differential absorption is exploited to improve image contrast by the selection of x-ray tube target material and filters in mammography [8].

These two main interactions discussed above combine to produce attenuation of the incident x-ray photon beam as it passes through matter. Attenuation is the removal of x-ray photons from the x-ray beam by either absorption or scattering events. Let us consider what happens when a beam of N_0 photons passes through a slab of thickness L. The unit of L in (2.3) is cm. Let's assume that a beam of n x-ray photons passes through a thin slab of matter with thickness dl with a probability of interaction μ (see Figure 2.5). Denote the decrease of photons from the beam by dn, Then,

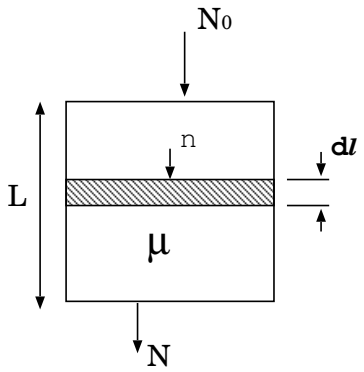


Figure 2.5: An X-ray beam with N_0 photons passes through a slab of matter with thickness L and linear attenuation coefficient μ.

$$dn = \mu n \, dl. \tag{2.1}$$

Taking integral on both sides, we get:

$$-\int_{N_0}^{N} \frac{dn}{n} = \mu \int_0^L dl. \tag{2.2}$$

where we put minus sign on the left side since dn is a decrease. Also since μ is constant, it can be brought outside the integral on the right side. Solving 2.2 gives rise to the Lambert-Beers Law:

$$N = N_0 e^{-\mu L}. \tag{2.3}$$

The symbol μ is called the ***linear attenuation coefficient*** and its unit is cm^{-1}. The value of μ gives the probability that an x-ray photon will be attenuated per centimeter of object [11] through absorption or scattering.

In this simple example, μ was independent on position. Attenuation coefficients μ depend on both energy ε and position r, so that $\mu = \mu(\varepsilon, r)$. In the diagnostic energy range, the linear attenuation coefficient decreases with increasing energy [11]. The goal of 3D reconstruction is to estimate an entity proportional to the object attenuation. In Chapter 4, we will generalize the simple Beer's law 2.3 to account for energy and position effects. A 2D mammogram is due to line integrals of μ through the breast as we shall see later.

2.2.3 X-Ray Detection

After the x-ray beam passes through the object, the surviving x-ray photons will be recorded. Our imaging chain now moves to x-ray detection. X-ray detection can be classified as direct or indirect [11]. The direct detection system transfers x-rays directly to electrical charge without intermediate state. Amorphous selenium (a-Se) has been usually used as the direct x-ray detector material. The indirect detection system transfers x-rays to visible light photons and then finally to electrical charge. A scintillator(such as cesium iodide (CsI))-based x-ray detector is used in an indirect detection system. In the indirect detection mechnism, x-rays interact with a phosphor,

causing it to emit light. The visible-light photons then propagate by optical diffusion to a photo-detector, such as a film emulsion or a silicon photo-diode. The photo-detector then records the pattern of visible light discharged by the phosphor as an image. In this process, x-rays are transfered to light photons and then to electrical charge. Since this dissertation is focused on breast tomosynthesis which requires digital detectors, we shall consider digital x-ray detectors, especially a flat-panel CsI indirect detector. Such a detection process chain can be described by a cascaded linear model [24] with the assumption of a stationary system. We will give the details of this cascaded model and noise effects caused by each stage in Section 3.6.

2.3 Conventional Mammography

The basic imaging chain has been described qualitatively. Here we show its use in conventional mammography. It is worth describing conventional mammography since tomosysthetic units are hardware modifications of conventional mammographic units. Conventional mammography can be classified as screen-film mammography and digital mammography. We discuss these two types of conventional mammography.

The common mammography unit is composed of an x-ray tube and an image recorder mounted on the opposite sides of a gantry (see Figure 2.6) . The system geometry is unusual since only on half of the field of the x-ray tube is used as illustrated in Figure 2.7 . As we can see, a compression paddle is attached to the mammography unit in order to hold and compress the breast. The thickness of the compressed breast is approximately 5 cm. Breast compression is an important step during mammography and used in order to [25]:

- Flatten the breast so there is less tissue overlap for better visualization of anatomy and potential abnormalities.

Figure 2.6: GE digital mammography equipment. The x-ray tube is housed in the box at the top. A gel phantom sits atop the detector (dark panel at bottom).

- Lessen overlapping normal shadows, which can appear as suspicious regions on the image receptor.
- Allow the use of a lower x-ray dose since a thinner amount of breast tissue is being imaged.
- Immobilize the breast to eliminate image blurring caused by motion.
- Reduce x-ray scatter which also degrades the image quality.
- Lower dynamic range requirements on the detector by creating a uniform thickness.

Typically, two views [25] of images are taken from each breast in conventional

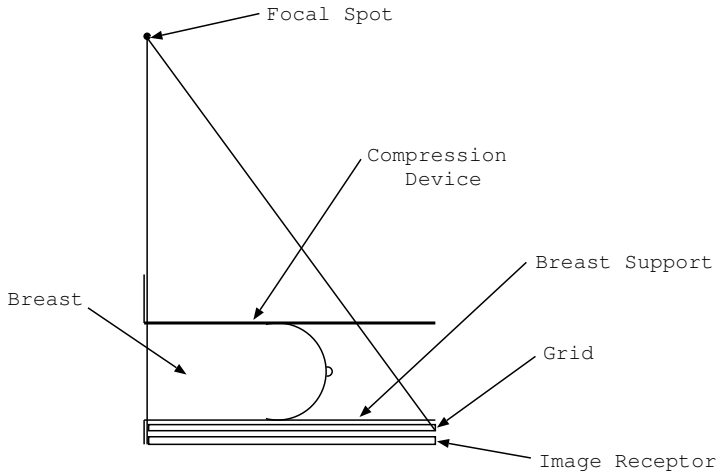


Figure 2.7: Geometric arrangement of system components in mammography. Geometry is not in relative scale.

mammography, namely from above (cranial-caudal view, CC) (from top to bottom) and from an oblique or angled view (mediolateral-oblique, MLO) (from side to side). The images are recorded on screen-film or digital detectors. The detected signal represents the summation of breast tissue attenuation along the transmission path.

2.3.1 Screen-Film Mammography

In screen-film mammography, the transmitted x-rays are recorded on a screen-film cassette under the breast after x-rays from a point source irradiate the breast. As the low-dose x-rays pass through the breast, they are attenuated to varying degrees by different tissues. The result is that fat appears as the black regions, others such as glandular tissue, tumor masses and micro-calcifications appears as varied levels of white regions on screen-film mammogram due to different attenuations at different breast tissue structures. Screen-film mammography is and will continue to be a valuable tool for detection and diagnosis of breast cancer given the following advantages [5]:

- The technology is relatively inexpensive and well established.

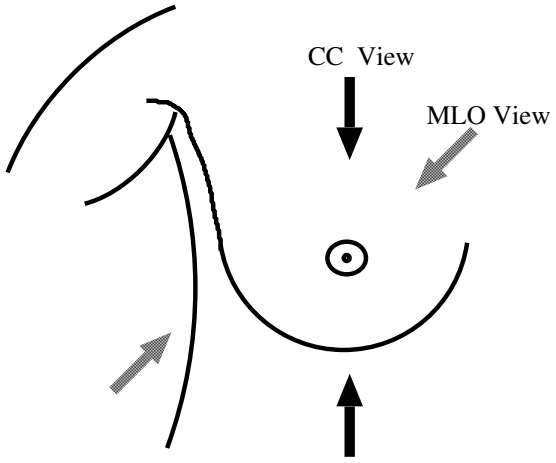


Figure 2.8: Solid arrows indicate the positioning of CC view. Dotted arrows indicate MLO view. The x-ray tube is positioned above the breast.

- The image receptor is capable of achieving very high limiting spatial resolution (at least 20 line-pairs/mm).
- Images are conveniently displayed using view-box technology.
- Film performs an inherent logarithmic compression of dynamic range onto the available optical densities of the film.

However, screen-film mammography has some limitations on displaying the finest features due to inadequate contrast, therefore resulting in a limited sensitivity for the detection of breast cancer with dense breast [6]. Film does not have a linear sensitivity to photon flux [7], there is a narrow range over which it can detect small difference in contrast. Also film requires processing time and storage space. Another limitation is the effect of structural noise due to film granularity, degrading the visibility of micro-calcifications and other fine breast structures [5]. More importantly, film-screen systems are not usable for tomosynthesis since tomosynthesis requires that several images must be acquired and read out in digital form within a very short time period,

which is not possible with film-screen systems. These limitations can be effectively overcome with a digital mammography system.

2.3.2 Digital Mammography

Digital mammography is similar to conventional screen-film mammography. Digital mammography uses essentially the same mammography system as conventional mammography, but the system (see Figure 2.9) includes a digital detector and a

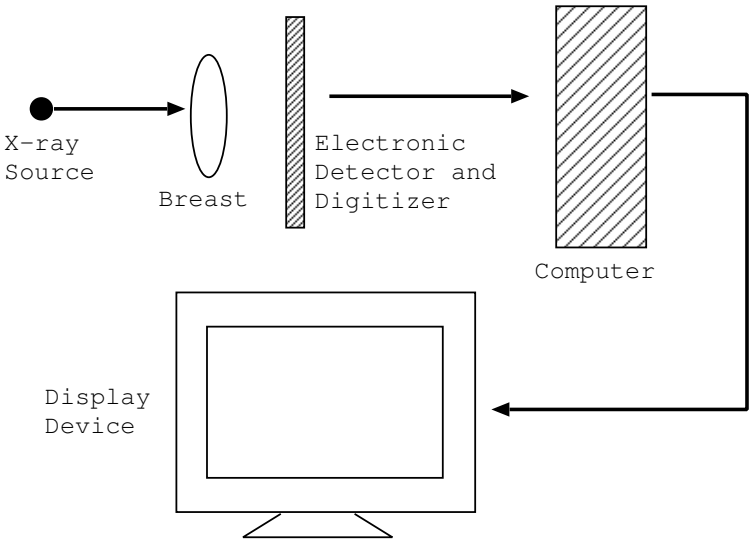


Figure 2.9: Schematic representation of a digital mammography system.

computer instead of the film cassette (Figure 2.6). With digital mammography a electronic x-ray detector is used to record the breast image and to convert the image into a digital picture. After the digital mammogram is stored on a computer, it can be processed and displayed on a soft copy or hard copy device [25]. Digital mammography can provide the following benefits over conventional screen-film mammography [26]:

- Improved contrast between dense and non-dense breast tissue.

- Faster image acquisition (less than a minute). Conventional screen-film mammography requires several minutes to develop the film.

- Easier image storage.

- Physician interaction with contrast and range of breast image display for more accurate detection of breast cancer.

- Ability to correct under or over exposure of films without having to redo mammograms.

- Transmittal of images over phone lines or a network for remote consultation with other physicians.

Digital mammography is starting to be important in clinical practice, but is still undergoing development. Studies by Pisano *et al.* [6] suggested that the overall diagnostic accuracy of digital and film mammography for breast cancer was similar, but digital mammography is more accurate in women with dense breasts. The improved performance in a digital mammography system is mainly associated with the x-ray detector and the display device. High resolution display monitors and hardcopy devices meet the demanding requirements of displaying digital mammography images. Current large-area flat-panel detectors can accommodate small and average-size breasts with an $18cm \times 24cm$ detector and a $24cm$ x $30cm$ detector for larger breasts, and achieve a spatial resolution of 50 μm per pixel. [27]. While full-field digital mammography may lack the spatial resolution of film, it can provide improved contrast resolution if using tomosynthesis, which may make abnormalities easier to see. Digital mammography can also act as the platform for digital breast tomosynthesis, but involves the application of computational procedures to the digitized images.

2.4 3D Breast Imaging-CT Breast Imaging and Tomosynthesis Mammography

Conventional mammography techniques, using either screen-film or digital detectors, are limited by the fact that the 3D breast is projected onto a 2D image, in which the overlap of normal tissue structure can be mistaken for abnormalities. This problem is well known clinically, and results in breast cancers that may be missed on screening mammograms. If one could remove the overlapping tissue from the image, better diagnostic accuracy may be obtained. This problem can be solved by a 3D breast imaging system, which is able to reduce the structure overlap and provide 3D visualization of breast tissue. Two types of 3D breast imaging methods have been investigated: cone-beam computed tomography (CT) and breast tomosynthesis. We shall introduce both, but our focus will on the latter in the remainder of the dissertation.

2.4.1 Cone-Beam CT Breast Imaging

The dedicated breast CT scanner (General Electric CT/M) was first introduced in 1978 [28]. However, little progress was made due to limitations of image quality and concerns of radiation dose and cost effectiveness [24]. With the advent of digital detectors, a number of research groups have begun to investigate the feasibility of cone-beam CT breast imaging using CsI based full field flat-panel detectors [29, 30, 31, 32, 33, 34]. The cone-beam CT breast imaging system is a major development project requiring new patient protocols.

A typical cone-beam CT breast imaging system [35, 33] is illustrated in Figure 2.10. In this system, the patient lies prone on a table with an opening cut through which the breast drops down. A cylindrical breast holder is used to stabilize the breast. The x-ray tube and the flat-panel detector are placed underneath the table,

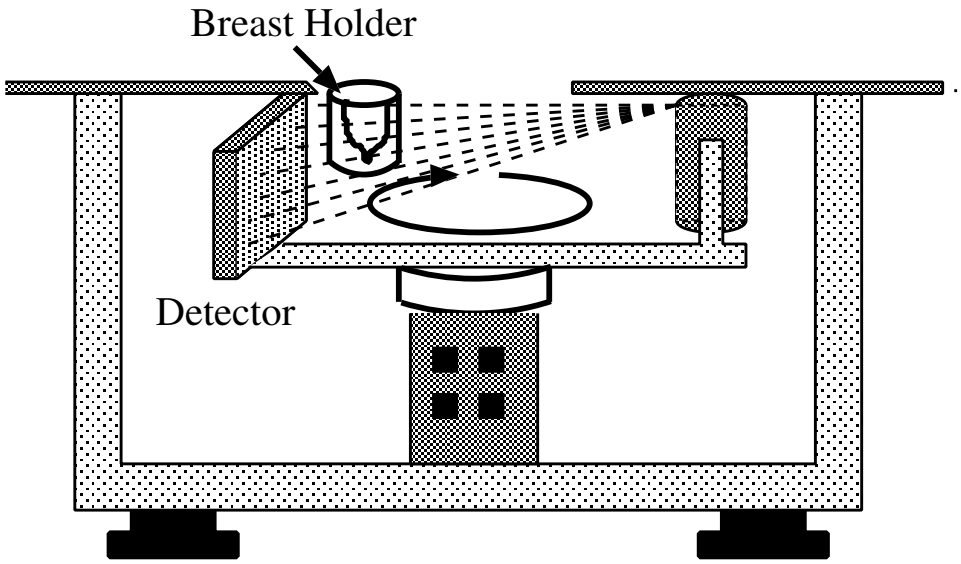


Figure 2.10: Cone beam breast CT system: the patient lies on the table with the breast hung through an opening into a cylindrical breast holder. The x-ray tube and detector are mounted on a rotating gantry and rotate around the breast during the scan. Only the patient's breast is shown inside the cylindrical breast holder.

and are rotated 360^o around the pendant breast. With such design, only the breast tissue is exposed to the x-ray radiation during each scan, and only the lower half of the cone-beam is needed to fully irradiate the breast to obtain the projection image. The acquired complete set of projection images are used to reconstruct the 3D breast. The mean glandular dose to the uncompressed breast is required to be same as that used for a two-view conventional mammography study. Therefore, the exposure per projection view is very low in CT breast imaging. The electronic noise of the flat-panel detector has more impact on image quality in CT breast imaging than in conventional mammography [36]. Since the pendant breast in CT imaging is uncompressed, it is expected that the higher kVp settings are required [37].

Glick *et al.* [38] investigated the optimal kVp settings of a tungsten anode for CT imaging of the uncompressed breast for a lesion detection task, given that an average

glandular dose is approximately equivalent to that of two-view planar mammography study. It was observed that the optimal kVp settings for the tungsten anode ranged over 30-50 kVp. This kVp setting is higher than that in tomosynthesis. The pendant-geometry breast CT system studied by Boone *et al.* [35, 33] showed good image quality with phantom studies as well as with patient images. Their studies demonstrated that a flat-panel cone-beam CT technique can detect relatively small tumors with a glandular dose less than or equal to that of conventional mammography. Ning *et al.* [39] compared conventional mammography and a prototype flat-panel CT breast imaging system using a compressed and an uncompressed physical breast phantoms with an acrylic "tumor". It was suggested that CT breast imaging increased tumor visualization compared to conventional mammography.

The prototype cone-beam CT breast systems are still under development and testing, and have not become clinically used yet. The improved lesion detectability provided by CT breast imaging system, compared to that of conventional mammography system, has driven the interest of many research groups in developing such systems.

2.4.2 Breast Tomosynthesis

Breast tomosynthesis is another type of 3D breast imaging. The hardware can be easily adapted from a conventional mammography system. There are two types of breast tomosythesis geometries available: complete isocentric illustrated in Figure 2.11 (a) and partial isocentric seen in Figure 2.11 (b). In a complete isocentric geometry, the x-ray tube and the detector are rotated over a limited angular range around a fixed central point while the compressed breast is stationary during a tomosynthesis scan. The partial isocentric system requires that the compressed breast and the detector remain stationary while the x-ray tube is rotated in an arc above the breast. With both geometries, the x-ray tube is pulsed to acquire different projection images at

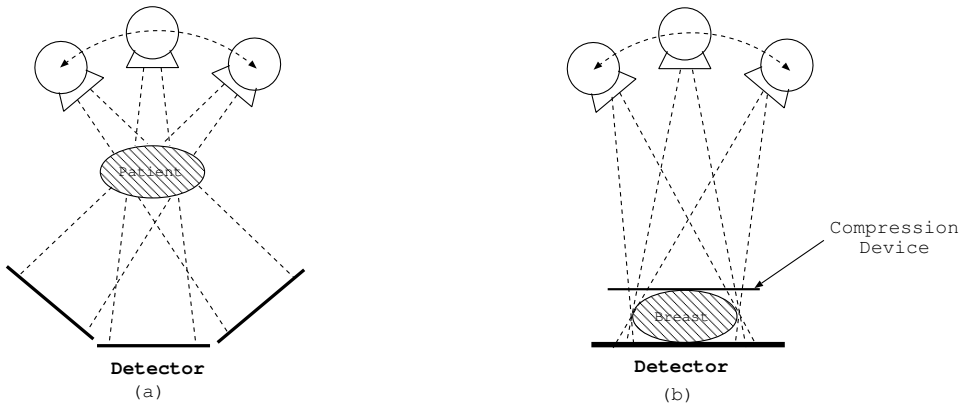


Figure 2.11: Tomosynthesis geometries (a) Complete isocentric motion in which both the X-ray tube and the detector rotate about a fixed central point. (b) Partial isocentric motion in which the detector is stationary and the X-ray tube rotates about some rotation point.

desired angles. The total radiation dose to the patient at N_{angle} views is equal to that used for the two projection views of a conventional mammography. Thus, only $1/N_{angle}$ of the conventional mammography dose is used at each tomosynthetic view. A typical number of angular views is 11-20. Since conventional screening mammography already uses a low dose, the dose for each tomosynthetic view is extremely low, yielding noisy data.

Reconstruction in the breast tomosynthetic procedure is the final step. From a set of projection images, a 3D x-ray attenuation distribution of the breast is computed in which the high resolution is retained in the two dimensions lying in the transverse plane (parallel to the plane of the compression paddle) as seen in Figure 2.12, while lower resolution is retained in the third dimension. (Note that we will describe isocentric tomosynthetic geometry in much more detail in Chapter 8. See for example, Figure 8.1) This reconstructed 3D object can be decomposed into a set of transverse slices. Each slice contains an image contribution due maily to the breast object in that slice, but also contains "out-of-focus" contributions from nearby slices.

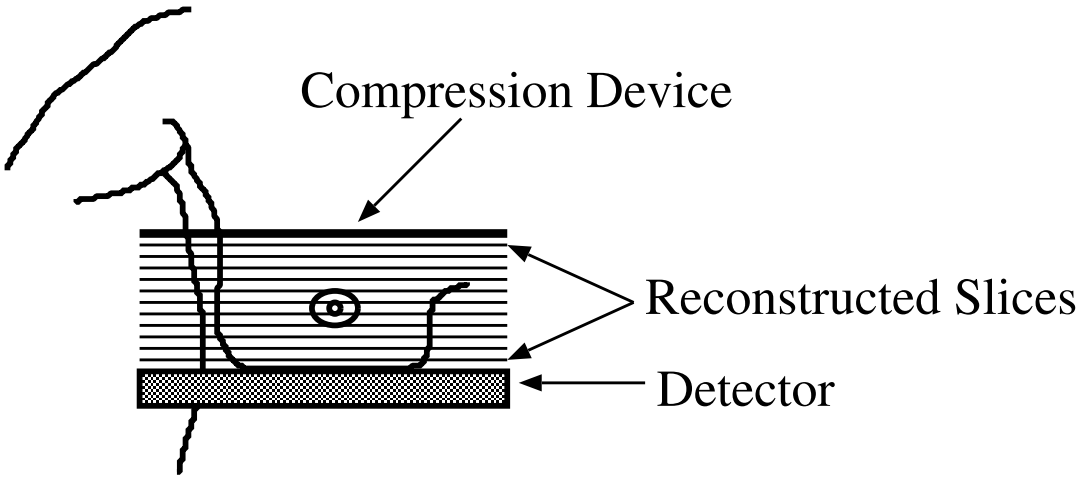


Figure 2.12: The reconstructed tomosynthesis slices are parallel to the detector image plane.

A simple form of the reconstruction step can be understood as generating a set of slice images from the summation of a set of shifted projection images acquired at different angles [40]. The structures in transverse plane at different heights above the detector undergo different degrees of position displacement as the tube moves, thus they are projected onto the detector at positions depending on the relative heights of the structures of the object as illustrated in Figure 2.13. In this example, the triangle and the ellipse overlay each other when the projection image is shot at 0^o; while at $\pm 15^o$ the object's shadows are shifted relative to each other in the image plane. In the isocentric geometry assumed in Figure 2.13, the ellipse will always fall on the same spot on the detector while the image of the triangle shifts with position for this particular case. Thus, if we take the images of Figure 2.13 and simply add them, the ellipse will be reinforced and the triangle blurred. This is seen in Figure 2.14 (a). On the other hand, we can in the computer appropriately shift each digital image as seen in Figure 2.14 (b). With the shift the triangle is now enhanced ("in focus") and the ellipse smeared out. Different shifts lead to focusing on different planes. While this simple "shift-and-add" process illustrates the basic idea of tomosynthesis, shift-and-add can

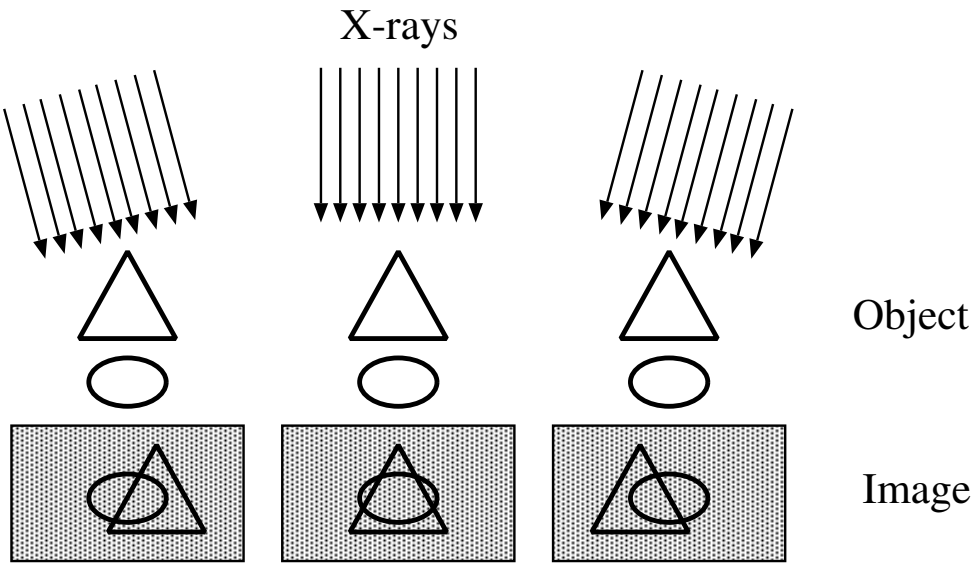


Figure 2.13: Tomosynthesis imaging. The triangle and the ellipse at different heights are seperated differently in the image plane when acquired at different angles $(-15^o, 0^o, +15^o)$

be replaced by more sophisticated algorithms that yield better focusing.

Recently, Gong *et al.* [24] conducted a computer simulation study comparing lesion detection accuracy with digital mammography, breast tomosynthesis, and cone-beam CT breast imaging. Their results indicated that for the same dose, a 5 mm lesion embedded in a structured breast phantom was detected by the two 3D breast imaging systems, breast tomosynthesis and cone-beam CT, with statistically higher confidence than with planar digital mammography, while the difference in lesion detection between breast tomosynthesis and cone-beam CT was not statistically significant.

Breast tomosynthesis offers better in-plane spatial resolution but worse contrast resolution than cone-beam CT.A number of commercial vendors (GE, Siemens, Hologic Co.) are currently testing breast tomosynthesis prototypes to get Food and Drug Administration (FDA) approval.

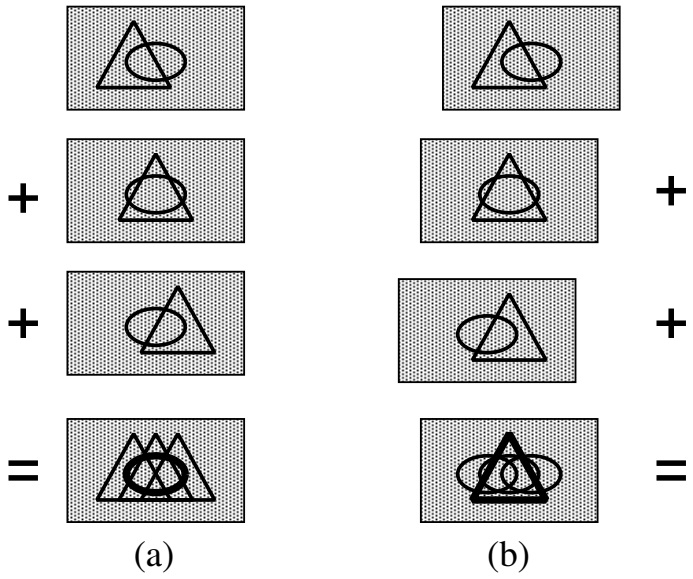


Figure 2.14: The acquired projection images are appropriately shifted and added to bring either triangle or ellipse into focus, structures outside the focal plane are spread across the image and blurred.

Chapter 3

The Optimization of Breast Tomosynthesis Systems

Although breast tomosynthesis only requires minor hardware modification of a conventional screening digital mammography system by incorporating motion of an x-ray tube gantry and pulsing of the x-ray exposure, optimizing imaging performance of the whole tomosynthesis imaging chain poses a tremendous challenge. For breast imaging, lesion detection would seem to be the most meaningful form of image quality. In that case, the ultimate goal of system performance optimization is to obtain optimal detectability.

We shall go through the entire imaging chain and address the problems in each stage. The problem has been attacked piecemeal i.e. many separate studies have been done to investigate one or two factors in the imaging chain, and the problem is too large to be addressed in one study. We shall start at the generation of the x-ray source, interaction with the body, detection, the reconstruction stage and end with image quality evaluation. The goal of this chapter is to qualitatively review each stage of the imaging system with an emphasis on

- What parameter at that stage can be selected and tuned to optimize image quality in breast tomosynthesis? What parameter cannot be adjusted but nevertheless must be accurately simulated?

- Why might the adjustment of this parameter affect image quality?

- Who has done previous work in the above?

Given the context of the massive optimization problem, we then describe what part of the problem we will be addressing.

As a digression, we point out that conventional mammography has its own imaging chain and its own optimization problem. However, for breast tomosynthesis, optimization could differ from that for conventional mammography because in conventional mammography, the observer looks at projection images and tries to detect abnormalities, while in tomosynthesis one uses many projection images to compute a 3D reconstruction which is then viewed and then one tries to detect a lesion in one of the slices.

3.1 Factors in X-ray Generation

The principle of x-ray production was given in Section 2.2.1. In this section, we shall discuss some factors involved in x-ray generation, which affects the image quality.

3.1.1 kVp Setting

The choice of kVp is an important operational factor affecting the final image quality in breast tomosynthesis. Recall that the kVp setting, together with target material and filtration, determines the energy spectrum of the x-ray beam. Higher kVp shifts the spectrum to higher energies. In general, the kVp setting depends on the density and thickness of the breast, and target/filter materials. For a given breast and target/filter combination, the kVp cannot be too low or too high. If the x-rays are too low in energy, they will be absorbed by the breast, contributing the dose but not increasing the contrast because they never reach the detector. If too high, they will pass through the breast without absorption, hence reduce the contrast. It is observed that thicker and denser (i.e. 100% glandular) breasts tend to require higher kVp

settings to reach maximal SNR. Since reduced breast compression is used in breast tomosynthesis, the x-ray kVp need to be raised in order to penetrate the thicker breast. So determining an optimal x-ray kVp is necessary.

Zhao *et al.* [20] presented a theoretical study of the choice of kVp using a-Se flat-panel digital detector. The detectability of a 200 μm microcalcification within the breast was calculated as a function of x-ray spectra in terms of kVp. Their results showed that for tomosynthesis with 11 projection views, the optimal kVp is at least 2-3 kVp higher than the optimum for screening mammography. Glick *et al.* [41] investigated the effect of the kVp setting using an indirect CsI based detector. The imaging task was defined as the detection of a isotropic lesion embedded in a compressed breast model of varying thickness and composition. The ideal observer SNR was used as a figure of merit under the assumption that the imaging system is linear and shift-invariant. Because kVp settings can affect detector noise (which in turn affects detectability), the purpose of this study was to investigate the impact of electronic noise on image quality by varying kVp settings. It was observed that the kVp setting that produced maximum SNR was consistent with the results in [20].

3.1.2 X-Ray Tube Target/Filter

Similar to kVp setting, target/filter combinations determine the energy spectrum and also can affect the image quality. Because less compression is used in breast tomosynthesis, the thickness of imaged breasts can vary, compared to the constant breast thickness (constrained by paddles) in conventional mammography. However, as pointed out earlier, even if breast thickness is not issue, optimization criteria for tomosynthesis differ from that for conventional mammography. Therefore, an examination of different target/filter combination is required. In a breast tomosynthesis system, there are four target-filter combinations: Mo-Mo, Mo-Rh, Rh-Rh and W-Rh. The right choice of target-filter combination can provide the optimal image quality.

Glick *et al.* [41] studied three different x-ray tubes (Mo-Mo,Mo-Rh and W-Rh) using two breast thickness (4 cm and 6 cm). The results suggested that W-Rh provided better performance for tomosynthesis system using indirect CsI detector systems than the other two target/filter combinations (Mo-Mo and Mo-Rh). Zhao *et al.* [20] also investigated these three combinations for a direct a-Se detector system. Their results showed that W-Rh was the optimal target-filter combination for all breast thickness (2-8 cm). Wu *et al.* [42] explored three target-filter combinations (Mo-Mo, Mo-Rh, Rh-Rh) for a-Se detector. Contrast-to-noise ratio was measured from projections acquired with three target-filter combinations. The results showed that Mo-Mo gave the best quality for small breast thickness (2cm) while Rh-Rh is more effective for large breast thickness ($\geq 4cm$).

3.1.3 Focal Spot Size

The image in each detector is blurred by a scaled version of the focal spot point spread function (PSF). As in conventional mammography, x-ray tube design requires as small a focal spot as possible. So though focal spot size affects the detected data, there seems to be no reason to optimize the focal spot size - one must just make it as small as possible.

3.1.4 Poisson Noise

The imaging chain introduces noise and here we describe one source of noise. In the x-ray production, x-ray emission like any electromagnetic radiation follows a Poisson distribution. That is, the number of photons emitted into 4π s.r. per unit time by the source follows a Poisson distribution. This noise is also known as quantum noise. Even after limited solid angle effects due to finite detector size, and even after attenuation, the number of x-ray photons entering a detector is still Poisson distributed. The mathematical derivation will be presented in Chapter 4. Note that

quantum noise is inherent in the x-ray radiation, and is distinct from "structural noise" or "anatomical noise" discussed later. Poisson noise cannot be eliminated, but must be simulated as accurately as possible.

3.2 Geometric Effects

There are two imaging geometries available in current breast tomosynthesis systems: complete isocentric and partial isocentric as illustrated in Figure 2.11. We shall discuss some effects involved in both geometries below.

3.2.1 Acquisition Angle

Typically, the angular range used in breast tomosynthesis is $\pm 15^o$ to $\pm 25^o$ with a limited number of views (typically 11 to 20). The limited angular range results in incomplete sampling in the Z direction, hence generates reconstruction artifacts. Therefore, tomosynthesis reconstruction usually uses larger voxel dimensions in the Z direction compared to the in-plane resolution. As discussed in Section 2.4.2, the total radiation dose to the patient at N_{angle} views is equal to that used for the two projection views of a conventional mammography. Thus, only $1/N_{angle}$ of the conventional mammography dose is used at each tomosynthetic view. The dose at each view is inversely proportional to the number of acquisition angles. Taking N_{angle} low-dose images results in decreased SNR in each image compared to taking a single image with the same total exposure [43].

In breast tomosynthesis, the x-ray tube/the detector have to be operated much faster than in screening mammography so that projection images can be acquired in a reasonable time (less than 10 seconds) without significant patient motion or discomfort. More views require more acquisition time, causing patient motion and artifacts in reconstructed images. Given that, the reasonable number of projection

views is considered to be 11 to 20. The angular schedule, i.e. the set of angles actually used, can affect image quality. Given a fixed number of angles, e.g. 20, is it better to make them equispaced over a wide range, equispaced over a narrower range? Is it better to cluster more angles near 0^o or to cluster more near the extreme $\pm\theta_{\max}$ angles? More angular samples at high θ result in better resolution in the Z direction but poorer resolution in the XY plane. More samples near 0^o results in better XY images but poorer Z resolution. The impact on lesion detectability is indirect, but the effect may be important. Optimizing this tradeoff is essential in breast tomosynthesis.

Chen *et al.* [44] explored different acquisition techniques where projection images were acquired with the following parameters: 13, 25, 49 projection views over $\pm 12.5^o$ and $\pm 25^o$ of angular ranges. The total exposures were same with different parameter settings. The results showed that the acquisition setting with 49 views over $\pm 25^o$ provided the best performance in terms of the noise power spectrum of the reconstructed plane. Wu et al. [43] investigated different acquisition protocols - limited but uniform angular sampling, uniform wide angle sampling, nonuniform wide angle sampling and uniform full sampling over 360^o. The experiments were conducted on a breast phantom containing one simulated mass sphere and simulated microcalcifications. The results showed that limited angle sampling results in good resolution of detail in the XY plane, but poor resolution in Z; A uniform full sampling protocol provided the best Z resolution, but poor resolution in XY plane. The non-uniform wide angle sampling is a compromise that yields moderate resolution in the Z direction while maintaining good image quality in the XY plane.

3.2.2 Tube Movement Modes

The x-ray tube can move in a continuous or step-and-shoot motion. In a step-and-shoot motion, the x-ray gantry must completely stop at each angular position before turning on x-rays in order to avoid blur caused by tube motion. The step-

and-shoot mode is challenging in mechanical design and can easily cause mechanical instability if scan time is short. For a practical breast tomosynthesis system, the set of projection images should be acquired in less than 10 s in order to avoid image blur from patient motion. If step-and-shoot is employed, a complete stop is required at each view, hence requiring much longer time. In that mode, patient motion is unavoidable and it causes artifacts in reconstructed images. Step-and-shoot poses a challenge in mechanical design, though it leads to better data if the patient does not move.

Currently, continuous tube motion with pulsed x-ray shots is more widely employed for practical use. The continuous tube motion causes focal spot movement during exposure, therefore introducing focal spot blur into the system. (This focal spot blur is to be distinguished from the blur caused by focal spot size discussed in Section 3.1.3.) A study in [45] quantified the blurring effects on the acquired data. These are dependent on the details of the geometry. The focal spot blur can not be removed. It is still under investigation how to reduce the effect of focal spot blur in breast tomosynthesis, perhaps by post processing. It would appear that if the breast could be immobilized for long period and mechanical problems associated with stop-and-shoot be solved, then step-and-shoot would be preferred.

3.3 Structural Noise in the Breast

Structural noise is often referred to as "anatomical variation" or "object variability" or "object clutter". It is an important influence on image quality. Normal anatomical structure including all normal breast components as discussed in Section 2.1, contribute to a loss of lesion detection accuracy by masking the lesion. Bochud *et al.* [46] conducted a study for conventional mammography to demonstrate the importance of variations in background anatomy. The results suggested that the anatomical structural noise is a dominating effect on the detection of a large signal (like a mass).

Hence, reducing the system noise by increasing the dose will not improve task performance. Burgess *et al.* [47] have shown that for conventional mammograms both quantum noise and normal mammographic background structural noise impeded visualization and reduce lesion detectability. It has been suggested that the degrading effect of normal mammographic structure on detection of lesions greater than 1mm diameter can be 30 to 60 times that of the quantum noise [46, 47]. Even if there is no other noise mechanism, the structural noise leads to the loss of lesion detection accuracy.

Structural noise is a given important source of noise that affects image quality. Since it is part of the breast anatomy, one can not adjust it, that is, one can not optimize it. Therefore, it is very important to accurately simulate this source of noise. Few research groups have done this. We have devoted considerable efforts to simulation of structural noise in a 3D breast phantom described in Chapter 6.

3.4 Scatter Effects

We first consider scatter in conventional mammography. As discussed in Sections 2.2.1 and 2.2.2, low x-ray energies (25 – 40 kVp) are used for breast imaging. At these energies, the probability of absorption via photoelectric interaction within the breast is significant. However, the probability of Compton scattering of x-rays within the breast is still quite high. Even though a breast is relatively small compared to most of the body, it is still the source of significant scattered radiation that reduces contrast. Around 33% to 50% of the total radiation photons would have experienced a scattering interaction within the breast [21]; i.e. the scatter-to-primary ratio (SPR) would be 0.5 to 1.0. It was estimated that the SPR is about 0.5 for a 5 cm 50% adipose 50% glandular breast at 31 kVp [48]. In addition to contrast reduction, the recording of scattered radiation uses up part of the dynamic range of the detector and adds

statistical noise to the image. This must be considered because in mammography we are attempting to see many low contrast structures where any reduction in contrast can be detrimental.

The SPR in conventional mammographic applications has been studied by several investigators [23, 49, 50, 51, 52]. Boone *et al.* [48] reported a comprehensive analysis of scattering properties covering a wide range of mammographically related parameters and investigated the magnitude and the effects of scatter on image quality intensively.

The most effective way to reduce scatter in 2D conventional mammography is using an anti-scatter grid. A grid is essentially a collimator placed atop the detector. The vanes are angled to point to the x-ray focal source. Since unscattered photons follow straight paths from focal spot to collimator, most pass through the grid. Scattered photons deviate from this straight line path and get absorbed by the collimator. A grid can attenuate primary radiation as well as scatter.

For most tomosynthetic geometries, grids are difficult to implement. For a fixed-detector/rotating-tube design for breast tomosynthesis, the detector and the breast stay stationary while the x-ray tube rotates around an axis during image acquisition. The total rotation angle ranges from 30^o to 50^o. It would be very difficult to design an anti-scatter grid for this geometry because most primary x-ray photons would be blocked by the standard grid orientation when the projection angle is large [24]. For an isocentric geometry, grids are not used in this imaging system to avoid grid line artifacts and high patient dose [53]. Since the exposure time for each projection in tomosynthesis is very short, it is difficult to avoid grid lines if a grid is used [53]. Zhao *et al.* [20] showed that the use of a grid is not beneficial for tomosynthesis acquisition because the grid will reduce the number of photons reaching the detector and then worsen the problem of electronic noise.

Instead of blocking scattered photons with a grid, one might attempt to allow

scattered photons to hit the detector, then try a digital scatter correction (SC) technique. The basic SC idea is as follows: First, obtain an initial estimate $\hat{\boldsymbol{\mu}}$. Then, use this $\hat{\boldsymbol{\mu}}$ to simulate, using a Monte Carlo packages, the flight of photons through the breast. From the Monte Carlo result, obtain an estimate of scatter and subtract it from the detector reading. The scatter corrected images can then be used to obtain an improved reconstruction.

Recently, Liu *et al.* [53] developed a breast specific scatter-correction method in breast tomosynthesis so that the attenuation coefficients reconstructed from projections will be more accurate. A phantom simulating a $5cm$ breast with uniform 70% adipose and 30% glandular tissue in the background was used in their study. Multiple masses, specks, fibers and line pairs are embedded in this phantom. The 3D breast was initially reconstructed from 15 projection images acquired from a GE prototype tomosynthesis system without SC. The voxelized digital breast model was entered into a Monte Carlo package simulating the prototype tomosynthesis system. One billion simulated photons were generated from the x-ray source for each projection and images of scattered photons were obtained. A primary-only projection image was then produced by subtracting the scatter image from the corresponding original projection image which contains contributions from both primary and scatter photons. The scatter free projection images were then used to reconstruct the 3D breast using the same algorithm. Compared with the uncorrected 3D image, the x-ray attenuation coefficients represented by the scatter corrected 3D image were closer to those derived from the measurement data. The Monte Carlo method is a mature and reliable technique for simulating radiation transport. However, it is not practical to use this technique in clinics due to the high computing power required. It took about 14 hours to generate, using Monte Carlo package, one projection image in the above study.

Gong *et al.* [24] simulated scatter radiation in breast tomosynthesis in the follow-

ing procedure. First, Monte Carlo simulation is performed to track x-ray transport through a uniformly dense breast model. For each breast phantom, a scatter-to-primary map was computed for each projection view. The scatter component of each projection image was then estimated by multiplying the SPR image by the primary x-ray incidence. Both scatter and primary components were then added. This method is similar to the one by Liu *et al.*. This method is also very time-consuming.

For scatter, optimization of image quality is controlled by SC methods. One might compare various SC methods to the case of no SC and see how it affects lesion detection. Because SC methods are based on Monte Carlo simulations, they are to date too computationally intense to be practical. There are no studies to date on the effects of SC on lesion detection in breast tomosynthesis.

3.5 Breast Compression in Tomosynthesis

In conventional mammography, breasts are highly compressed in order to reduce tissue overlap as discussed in Section 2.3. However, high compression pressure is not necessary for breast tomosynthetic imaging. Using the least possible compression for tomosynthesis could enhance separation of densities into distinct imaging slices, thus improving lesion detectability. Enough breast compression is needed to pull tissues out of the chest wall and to keep motion at a minimum [54]. Therefore, there is the possibility of less painful compression using tomosynthesis. The discomfort of full compression in conventional mammography discourage some women from undergoing the exam. On the other hand, if the breast compression is reduced, the x-ray energies need to increase in order that x-rays more efficiently penetrate the dense breasts as discussed in Section 3.1.1. We will discuss breast compression in more detail in Chapter 6.

The degree of breast compression (e.g. distance between paddle plates) affects

kVp setting, scatter events and the choice of target/filter combination. The use of less compression in breast tomosynthesis increases slice visibility, but tends to make motion more difficult to control. In addition, less compression increases effective breast thickness and interacts with kVp, while the dose must be kept constant. In this sense, the degree of compression (as measured for example by compression force on a paddle) is a variable that can be adjusted to optimize lesion detection. No specific studies for tomosynthesis have been done to date.

3.6 Detector Effects

In conventional mammography and in breast tomosynthesis, the effects of the large area flat-panel detector can play an important role in image quality. To aid the discussion of detector effects, we qualitatively summarize the behavior of one type of detector, a CsI based amorphous silicon flat-panel detector. The description below is based on work in [24].

The detector behavior is modeled as a linear cascade. At the entrance to the detector plane, the x-ray field is spatially varying and corrupted by Poisson noise. The x-ray photons first hit a CsI scintillator where a certain fraction of the photons are lost and the rest converted to optical photons. The absorption of lost photons can be modeled as a binomial process, and since a Poisson-binomial cascade is still Poisson, the x-ray quanta still surviving in the scintillator can be described by a Poisson distribution. Each remaining photon gets converted to numerous optical photons, with the conversion adequately described by a conversion gain factor. Since the Poisson variables are multiplied by a gain factor, they are no longer truly Poisson (variance no longer equals mean.). The optical photons are then passed to an amorphous silicon layer, again with a gain factor due to optical coupling efficiency. One electron is produced for every absorbed photon. So at this stage, the number of electrons are

represented by a scaled Poisson distribution. However, within the CsI scintillator, the photons spread, so that one x-ray /photon conversion results in a blob of optical photons at the interface of CsI and silicon. This spreading can be described as a convolution with a spread kernel. The electrons are now described as a scaled Poisson convolved with a PSF. The electrons are integrated over small pixel areas, and additive Gaussian readout noise is added to the signal. Since various gain factors are energy dependent, the effects of the above cascade must be summed in a weighted fashion over the x-ray spectrum. The result is a complex transformation of the entering x-ray photon field, described cleanly as a Poisson density, to a complex detector signal that is difficult to characterize analytically.

The above description was for a CsI based amorphous silicon detector, but another detector used in tomosynthesis, the a:Se direct detector, has a different physics. It is extensively described in [55].

The literature on detector performance is vast and beyond the scope of the present discussion. But two points are worth making: (1) Detector performance is always described in input-output metrics such as detective quantum efficiency (DQE) [56], not lesion detectability. (2) Detector performance is adjustable by altering bias voltages and fabrication dimensions. Thus, the detector stage is an important one and the user has some ability to control performance.

As we discuss at the end of this chapter, we will consider an idealized detector that simply counts x-ray photons and does not involve a complex conversion cascade.

3.7 Reconstruction Parameters

Image reconstruction is an inverse problem, and is an operation that acts on observed data $\mathbf{y}$ and transforms them to the reconstructed image $\hat{\boldsymbol{\mu}}$. Many image reconstruction algorithms leave a number of free parameters for the user to set for

a particular problem of interest. These parameters include initial estimates (IC), iteration stopping criteria (ISC), regularization parameter β, and apodization filters. It is often that these free parameters are themselves the subject of interest. Initial estimates are used in iterative reconstruction methods in order to speed up the convergence. Iteration stopping criteria, regularization weights and apodization filters are used to control the noise-resolution tradeoff in $\hat{\boldsymbol{\mu}}$. For example, for iterative methods, if the reconstruction stops at early iteration, it will result in low-contrast but less noisy estimates; if stopped later, the reconstructed image will be very noisy. There is a compromise between contrast and noise for these three parameters. We will describe these parameters in detail in Chapter 5.

Let O be the reconstruction operator. If $\boldsymbol{\zeta}$ is the free parameter vector, for example, $\boldsymbol{\zeta} = (\beta, \mathrm{IC}, \mathrm{ISC})$, then the reconstruction estimate $\hat{\boldsymbol{\mu}} = O(\mathbf{y}; \boldsymbol{\zeta})$. In the real world, data $\mathbf{y}$ are always corrupted by noise, e.g. Poisson noise. The noise is propagated through reconstruction operator O and affects the reconstructed image. The covariance matrix $K_{\hat{\boldsymbol{\mu}}}$ of the reconstructed image is an important statistical property, which affects lesion detectability. Since $\boldsymbol{\zeta}$ is under user control, it is easily adjusted to optimize lesion detectability.

3.8 Image Quality Evaluation

We will discuss this topic extensively in Chapter 7, but examine its role in the imaging chain here. Before image assessment, a task must be specified. In our study, we consider the detection task where the presence or absence of a signal (i.e. lesion) is of interest. One needs to evaluate the image quality. Given a task, an observer is required to perform the task. The observer might be human (i.e. radiologist) or a mathematical model observer. Having specified the task and the observer, a scalar figure of merit is needed to summarize the observer performance, for instance, in

order to compare different reconstruction algorithms or optimize the parameters of a reconstruction algorithm.

This is a step in the image quality evaluation chain. But here, accurate simulation rather than optimization is required. If we use a numerical observer, then it must be designed to emulate human performance as well as possible. For any observer, the task itself must be as realistic as possible. Since many studies will be dealing with digital evaluation of detection, the step here is to get a realistic human emulating task. The task we use is detection of a lesion in a cluttered background. This could be made more realistic, for instance, by not specifying the lesion location or its XY plane and requiring the human or numerical observer to perform a search step. Thus, in summary, this stage requires accurate simulation (of the observer and/or task) rather than optimization.

3.9 Summary

As a summary, a flow diagram for the imaging chain is given in Figure 3.1, which serves as a road map for this chapter. In this flow chart, each box is a stage in the imaging chain. Each box contains parameters that can affect image quality. The parameter may be tunable, or it may not be. In Figure 3.1, the sign $*$ following the parameter indicates that parameter affects image quality and can be optimized and tuned; # indicates that parameter can not be optimized, but must be simulated as accurately as possible; $ denotes aspects that we will address.

The subset of the problem that we address in the rest of this dissertation includes

- Since structural noise is so important, we simulate it for signal (lesion) and object (breast).

- Since Poisson noise is unavoidable, we simulate it.

- We optimize reconstruction algorithm and parameters since this is an important contribution to image quality.

- For detection, we use a task of moderate complexity — signal-known-statistically, location-known-exactly, background-known-statistically, and human observer. Humans are the ultimate detectors since model observers only emulate them.

We choose the above subset of the vast tomosynthesis optimization problem for the following reasons. Poisson noise in the radiation source is fundamental and cannot be removed, and structural noise in the breast anatomy is similarly fundamental. On the other hand, improvements in detector technology can, in principle, end up eventually in a pure photon counter, thus, alleviating any detector effects. Hence we leave the big problem of detector optimization to others. One must reconstruct the breast, and the choice of reconstruction parameters is again unavoidable. So in summary, we have addressed fundamental limiting factors while ignoring other important aspects that are in principle, not fundamental. Perhaps the aspect of most interest is the role of anatomical noise (object clutter) in limiting lesion detectability.

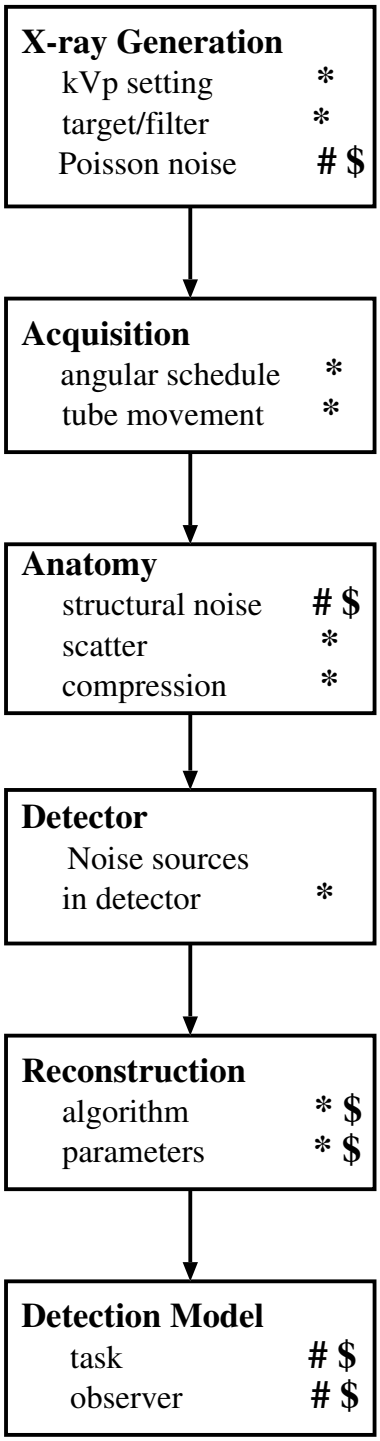


Figure 3.1: A diagram of the imaging chain. Parameters are listed at each stage. $*$ indicates that the parameter can be optimized; # indicates that the parameter cannot be optimized, but must be simulated as accurately as possible; $ denotes aspects that we will address.

Chapter 4

Mathematical Image Formation

From Chapter 2, we know that breast tomosynthesis is one type of transmission tomography. X-rays emanate from the x-ray tube, which is positioned on one side of the patient's body, then pass through the patient and are detected on the other side of the patient by the digital detector. A measurement at a specific location on the detector is determined by x-ray attenuation characteristics of the patient's anatomy along a straight line through the patient between the x-ray source and the corresponding location on the detector. The x-ray tube emits a relatively uniform x-ray distribution toward the patient. After the relatively spatially uniform x-rays interact with the patient, the altered x-ray distribution (unabsorbed x-ray photons) is recorded on a digital detector.

There are various geometries used in transmission tomography. If the x-ray tube rotates around the object over a full angular range (360^o or 180^o), one obtain a CT geometry as illustrated in Figure 4.1(a). If the x-ray tube rotates over the limited range of angles (e.g. $\pm 25^o$), one have a tomosynthesis geometry shown in Figure 4.1(b). In fact, CT, digital tomosynthesis and planar imaging are special cases in transmission tomography. In the most general case, the x-ray source and detector can move through arbitrary orbits. In many situations, the number of photons one can measure in a transmission scan is limited because of low dose, hence the low-count transmission data are very noisy. In modeling image formation, we must take this

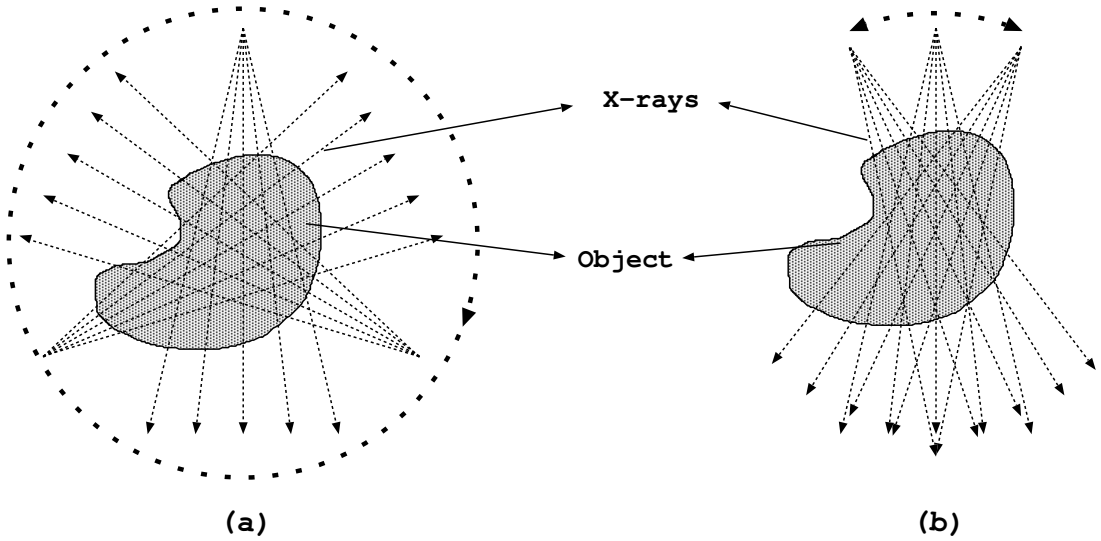


Figure 4.1: (a) a general CT geometry. (b) tomosynthesis geometry. In both geometries, the darker dash line denotes the trajectory in which the tube moves around the object. The arrows are rays from the source (X-ray tube) and the rays are detected by a suitable detector.

stochastic aspect into account.

4.1 Object and Projection Representations and Ideal Imaging Model

4.1.1 Image Formation for a General Imaging Geometry

We shall start with the general transmission scanning geometry. Since the point source and small detector element are connected by a narrow x-ray beam or "ray", we can use the symbol i ($i = 1, \ldots, N$) to index the ray, i.e. the source-detector pair. Note that one physical detector element can be associated with many rays if readings are taken in temporal sequence. For example, a stationary detector with 10^4 elements irradiated in turn by an x-ray source through 8 angles will result in $N = 8 \times 10^4$ rays.

Now consider the general geometry illustrated in Figure 4.2, where we show

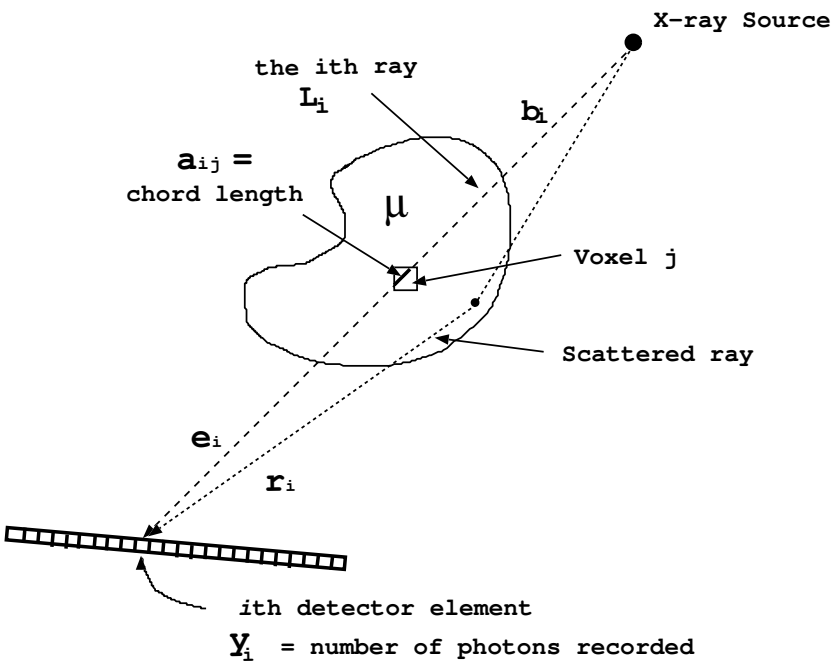


Figure 4.2: A beam of x-ray is shown propagating through a cross-section of the object. The detector is composed of many small pixel elements. One unscattered and one scattered rays are shown.

an object being illuminated by one unscattered x-ray and one scattered ray. For simplicity, we use a two dimensional object in Figure 4.2 to help understand the image formation. If we confine our attention to the cross-sectional plane drawn in the figure, we may consider μ to be a function of two space coordinates, x and y, and denoted it by $\mu(x, y)$ or $\mu(\mathbf{r})$, where $\mathbf{r} = (x, y)$. For the 3D case, $\mathbf{r} = (x, y, z)$. As discussed in Chapter 2, the units of μ are cm^{-1}. Let b_i denote the number of incident x-ray photons in the ith ray entering the object. Let y_i be the number of photons exiting through the ith ray and detected by a small detector element. Figure 4.2 is idealized in several ways. The x-ray source is not a point source, but has finite dimensions due to the finite focal spot. The detector is small but is not a point - it subtends a small solid angle as seen from the source. The object voxels have finite width. As a result, the number of x-ray photons that pass through a given voxel and hit a given detector

is the result of a complex geometrical calculation. Nevertheless, this simple depiction in Figure 4.2 will suffice. Suppose that the x-ray source is monoenergetic, the width of the ray is sufficiently small, and only one ray hits one detector element (pixel)(no scatter events), ideally one can derive the following relationship between the numbers b_i and y_i based on the Lambert-Beers law (2.3) for the ith ray:

$$y_i = b_i \exp\left(-\int_{L_i} \mu(\mathbf{r})\, dl\right), \quad i = 1, \ldots, N. \tag{4.1}$$

where L_i is the line between the source and detector for the ith ray and where the integration is made along the ith ray shown in this figure. We represent the measurement vector as the vector $\mathbf{y} = \{y_i; i = 1, \ldots, N\}$. Equation 4.1 will be later modified in several ways, but shows the basic relation of data y_i and unknown $\mu(\mathbf{r})$. In practice, the scatter effect must be taken into account. We shall mathematically explore the scatter event in the following section.

4.1.2 Object and Projection Representations

The quantity of interest in transmission tomography (i.e. breast tomosynthesis) is the linear attenuation coefficient $\mu(\mathbf{r})$, which is a continuous function of three-space coordinates on a object. We can represent $\mu(\mathbf{r})$ with a finite parameterization. For simplicity, the continuous function $\mu(\mathbf{r})$ is viewed on a digital display with a finite number of pixels, $\{\mu_j\}_{j=1,\ldots,M}$, by integrating over the jth pixel area,

$$\mu_j = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mu(\mathbf{r})\xi_j(\mathbf{r})\, d\mathbf{r}. \tag{4.2}$$

where M is the number of voxels in the object, and $\xi_j(\mathbf{r})$ is the basis function. There are many possible choices for the basis function. The conventional basis is just the "voxel" (3D) basis. Rect functions are used for voxel bases. The voxel basis $\xi_j(\mathbf{r})$ is 1 inside the jth voxel, and is 0 otherwise. The voxel basis function is given by

$$\xi_j(\mathbf{r}) = rect\left(\frac{x - x_j}{\triangle w_x}\right) rect\left(\frac{y - y_j}{\triangle w_y}\right) rect\left(\frac{z - z_j}{\triangle w_z}\right). \tag{4.3}$$

where

$$rect(x) = \begin{cases} 1 & |x| \leq \frac{1}{2} \\ 0 & |x| > \frac{1}{2} \end{cases} \tag{4.4}$$

In Equation 4.3, (x_j, y_j, z_j) is the center coordinates of the jth voxel and $\triangle w_x$,$\triangle w_y$, $\triangle w_z$ are the voxel widths in x,, y, z directions, respectively. The voxel could be cubic or rectanguloid.

Thus, the 3D continuous function $\mu(\mathbf{r})$ is transformed to 1D digital vector, we can represent this object vector as $\boldsymbol{\mu} = \{\mu_j; j = 1, \ldots, M\}$, where μ_j is the average value of $\mu(\mathbf{r})$ in the voxel j. In practice, the digitized object is displayed as a 3D digital matrix. A 3D discrete matrix can be represented as a vector by row-by-row and plane by plane scanning the matrix and then linking one row after another and one plane after another [57]. This is called lexicographic ordering.

Using the basis functions, $\mu(\mathbf{r})$ is approximated as follows:

$$\mu(\mathbf{r}) \approx \sum_{j=1}^{M} \mu_j \xi_j(\mathbf{r}). \tag{4.5}$$

Plugging (4.5) into the line integral in (4.1), we get the following equation:

$$\int_{L_i} \mu(\mathbf{r})\, dl \approx \int_{L_i} \sum_{j=1}^{M} \mu_j \xi_j(\mathbf{r})\, dl = \sum_{j=1}^{M} \mu_j \int_{L_i} \xi_j(\mathbf{r})\, dl = \sum_{j=1}^{M} a_{ij}\mu_j, \tag{4.6}$$

where

$$a_{ij} \equiv \int_{L_i} \xi_j(\mathbf{r})\, dl$$

is the line integral along the ith ray through the jth basis function. The units of a_{ij} are the length unit cm. In practice, a_{ij} can be simply understood as the chord length of the ith ray passing through the jth pixel. In Section 8.3, we shall describe calculation of a_{ij} using a ray-tracing method. One can represent $\{a_{ij}\}$ by the $N \times M$ matrix $\mathcal{A}$ whose ith row and jth column element is a_{ij}. For a breast tomosynthesis system with a $2000 \times 1000 \times 50$ breast and projection data acquired over 11 angles using a 1800×2400 ($18cm \times 24cm$) detector, the size of the system matrix in single-precision format is 19008000 GB ($2000 \times 1000 \times 50 \times 1800 \times 2400 \times 11 \times 4$ bytes). The

system matrix $\mathcal{A}$ is huge and it is impossible to directly use the system matrix during reconstruction. It must instead be computed on the fly.

With the system matrix, the ideal model in Equation 4.1 becomes the discrete one:

$$y_i = b_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i). \tag{4.7}$$

where

$$[\mathcal{A}\boldsymbol{\mu}]_i = \sum_{j=1}^{M} a_{ij}\mu_j \tag{4.8}$$

The system matrix $\mathcal{A}$ depends only on the system geometry. Any system geometry can be described by $\mathcal{A}$. Equation (4.8) can account for any transmission imaging system. Note that the imaging model Equation (4.7) is nonlinear in $\boldsymbol{\mu}$. The imaging model Equation (4.7) is ideal since it does not take into account any noise (i.e. data noise and electronic noise) or polyenergetic x-ray source effects, described below.

4.2 Statistical Forward Imaging Model

Building up a mathematical model of the physical imaging system - a forward model - is an important step in addressing the inverse problem of reconstruction. In this section, we intend to extend the simple imaging model Equation 4.7 to include the Poisson radiation noise and polyenergetic effects of the x-ray source, and also scatter events. A fourth component of the model is the noise associated with the detector. It is difficult to model the myriad detector effects, but we will present a simple model of additive electronic detector noise. For proper image reconstruction, one must incorporate these facts into the transmission imaging model for the measurements instead of simply using the ideal discrete model (4.7). These mathematical models presented in this section are then used in subsequent sections and chapters as a basis for the development of reconstruction algorithms and performance evaluation techniques.

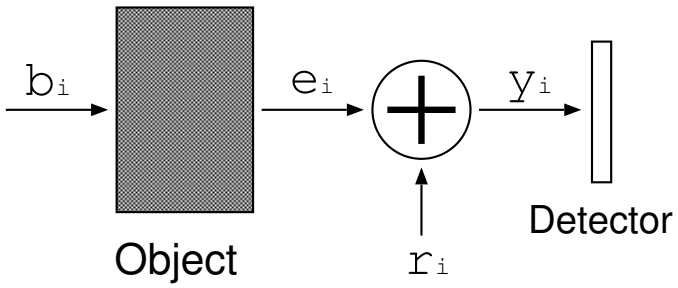


Figure 4.3: Pseudo-physical interpretation of the transmision data.

As mentioned in Chapter 2 we know that the x-ray source emits a random number of photons whose count distribution is Poisson. Hence b_i is a random variable, and the number of detected photons y_i is therefore also random. There often exist some additional counts recorded, primarily caused by Compton scattering, which also is Poisson [58]. A common way to model scatter events is using an additive term denoted by r_i even though the scatter effects can be reflected in the system matrix $\mathcal{A}$, which will make $\mathcal{A}$ more complicated and non-sparse.

Given the stochastic model, we shall redefine the terms used below: b_i denotes a Poisson random variable representing the number of photons emitted by the x-ray tube along the ith ray during the exposure time, y_i the random variable representing the number of photons recorded for the ith ray, r_i the Poisson random variable representing the number of photons from scatter events for the ith ray as seen in Figure 4.2, and e_i the random number of photons along the ith ray just before it hits the ith detector element. Therefore, one observes the following relationship as illustrated Figure 4.3:

$$y_i = e_i + r_i \tag{4.9}$$

We shall prove that both e_i and y_i follow a Poisson distribution. Suppose that the x-ray source emits b_i photons along the ith ray passing through an object toward a specified pixel on the detector. A radiative point source emanates a Poisson number

of photons/second into 4π s.r.. The number b_i emitted into a limited angle is still Poisson by virtue of the Bernoulli-Poisson property. Therefore, b_i is a Poisson random variable with mean $\bar{b}_i$:

$$Pr[b_i = n] = \frac{1}{n!} e^{-\bar{b}_i} \bar{b}_i^n.$$

Each of the b_i transmitted photons may either pass unaffected ("survive" passage) to the detector or may interact with the object (get attenuated) with the probability p. These are Bernoulli trials since the photons interact independently. From Beers' law (2.3) we know that the probability p of surviving passage using the discrete line integral expression is given by

$$p = \exp(-[\mathcal{A}\boldsymbol{\mu}]_i)$$

The number of photons e_i that pass unaffected through the object is also a random variable, then

$$Pr[e_i = m | b_i = n] = \left(\begin{array}{c} n \\ m \end{array} \right) p^m (1-p)^{n-m}, m = 0, \ldots, n.$$

Using total probability:

$$\begin{aligned} Pr[e_i = m] &= \sum_n P[e_i = m | b_i = n] P[b_i = n] \\ &= \sum_{n=m}^{\infty} \left(\begin{array}{c} n \\ m \end{array} \right) p^m (1-p)^{n-m} \frac{1}{n!} e^{-\bar{b}_i} \bar{b}_i^n \\ &= \frac{1}{m!} e^{-\bar{b}_i p} (\bar{b}_i p)^m \\ &= \frac{1}{m!} e^{-\bar{b}_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i)} \left(\bar{b}_i \exp\left(- [\mathcal{A}\boldsymbol{\mu}]_i \right) \right)^m, \quad m = 1, 2, \ldots. \end{aligned}$$

Therefore, the distribution of photons that survive passage is also Poisson with mean $E[e_i] = \bar{b}_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i)$.

Since e_i and r_i are independent Poisson random variables with the mean numbers $\bar{b}_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i)$ and $\bar{r}_i$, respectively, y_i is Poisson [59] with the mean number $\bar{y}_i =$

$\bar{b}_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i) + \bar{r}_i$. Thus, our forward model for general transmission measurements is:

$$y_i \sim Poisson\{\bar{y}_i(\boldsymbol{\mu})\}. \tag{4.10}$$

where the mean of the ith measurement is given by

$$\bar{y}_i(\boldsymbol{\mu}) = \bar{b}_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i) + \bar{r}_i \tag{4.11}$$

In general, this model can be used in any transmission geometry by changing $\mathcal{A}$ to reflect the geometry. In many papers, the $\bar{r}_i$'s are ignored. In our simulation, we shall assume that the $\bar{r}_i$'s are zero.

In medical centers, the $\bar{b}_i$'s are determined by periodic "blank scans": transmission scans with nothing but air (the attenuation coefficient of air is zero) in the scanner portal [60]. Since no patient is present, these scans can experience very long duration (typically a couple of hours, run automatically in the middle of the night). Thus the estimated $\bar{b}_i$'s obtained from such a long scan have much less variability than the transmission measurements y_i's. Therefore, we ignore the variability in these estimated $\bar{b}_i$'s, and use the symbol b_i as a known non-random parameter instead of $\bar{b}_i$. Therefore, combining Equations 4.10 and 4.11 result in

$$y_i \sim Poisson\{b_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i) + \bar{r}_i\} \tag{4.12}$$

Equation (4.12) includes the effects of photon noise and scatter effects, but it does not model polyenergetic effects and electronic detector noise (readout noise and dark noise). We shall incorporate these factors into our statistical forward model later.

The discussion above assume a monoenergetic x-ray source. In practice, the blank scan x-ray intensity b_i and the attenuation coefficient $\boldsymbol{\mu}$ are energy-dependent, thus the detected x-ray photon y_i are energy dependent. We rewrite Equation 4.12:

$$y_i(\varepsilon) \sim Poisson\{b_i(\varepsilon) \exp(-[\mathcal{A}\boldsymbol{\mu}(\varepsilon)]_i) + \bar{r}_i\} \tag{4.13}$$

where ε is x-ray energy, and the mean of $y_i(\varepsilon)$ is given as

$$\bar{y}_i\Big(\boldsymbol{\mu}(\varepsilon)\Big) = \bar{b}_i(\varepsilon)\exp\Big(-[\mathcal{A}\boldsymbol{\mu}(\varepsilon)]_i\Big) + \bar{r}_i \tag{4.14}$$

To model image blur attributed to the focal spot, a Gaussian function, with standard deviation modeling the focal spot size, can be then convolved with $\bar{y}_i$ given in Equation 4.14 [24].

The total of all detected x-ray photon for each pixel i can be obtained by integrating over all energies:

$$\begin{aligned} y_i &= \int y_i(\varepsilon)d\varepsilon \\ &= \int Poisson\left\{b_i(\varepsilon)\exp(-[\mathcal{A}\boldsymbol{\mu}(\varepsilon)]_i) + \bar{r}_i\right\}d\varepsilon \end{aligned} \tag{4.15}$$

Interchanging the operations $\int$ and $Poisson$ does not alter the result, therefore, one can rewrite Equation 4.15 as:

$$y_i = Poisson\left\{\int b_i(\varepsilon)\exp(-[\mathcal{A}\boldsymbol{\mu}(\varepsilon)]_i)d\varepsilon + \bar{r}_i\right\} \tag{4.16}$$

In Chapter 3, we discussed the complex effects of detector noise. Models for this noise depend on the type of detector used, but any such model is extremely complex. Instead of likelihood models for detectors, most researchers have used linear system models and unrealistic stationary assumptions in modeling detector effects. For reasons outlined in Chapters 3 and 9, we ignore detector effects in our modeling.

In our project discussed in Chapter 8, we shall use monoenergentic x-ray source and ignore electronic noise and the background events ($\bar{r}_i$). Thus, Equation 4.12 is simplified to

$$y_i \sim Poisson\left\{b_i\exp(-[\mathcal{A}\boldsymbol{\mu}]_i)\right\} \tag{4.17}$$

Chapter 5

Image Reconstruction in Transmission Imaging

As we point out in earlier chapter, the goal of transmission tomography, including tomosynthesis breast imaging, is to obtain an estimate of the 3D attenuation coefficients $\boldsymbol{\mu}$ of the imaged object (patient). Given the measured noisy projection data $\mathbf{y}$, a reconstruction step is needed to compute this estimate.

There are many reconstruction methods available, and they can be classified into two categories: deterministic and statistical. The image model includes Poisson photon noise, electronic noise (detector noise), scatter and a polyenergetic x-ray source. Different reconstruction methods handle these differently.

The deterministic reconstruction methods ignore Poisson photon noise and detector noise since they do not use a noise model. Various deterministic methods are available. They can be divided into two categories: analytical and linear algebraic. Analytical methods such as the filtered backprojection (FBP) algorithm try to invert the imaging operator in the continuous domain based on simplified imaging models. Linear algebraic methods such as the algebraic reconstruction technique (ART) and simultaneous algebraic reconstruction technique (SART) can incorporate sophisticated imaging models via the $\mathcal{A}$ matrix. In tomosynthetic mammography, the necessities of a short-scan and low-dose result in noisy projection data due to a low number of collected counts. Poisson noise is a concern, therefore the reconstruction problem naturally becomes a statistical problem. The statistical reconstruction methods in-

corporate the Poisson noise into its model (4.10) and can also incorporate detector noise. Statistical methods (methods using a noise model) of image reconstruction often outperform deterministic algorithms in terms of image quality [61]. The major impediment to their implementation transmission tomographic systems has been their heavy computational burden. We shall discuss the mathematical formulations for some popular image reconstruction algorithms in the two categories of deterministic and statistical algorithms and emphasize their applications and problems in breast tomosysnthesis.

5.1 Deterministic Reconstruction Methods

Deterministic approaches to transmission tomography begins with estimating the line integrals from the ideal model Equation 4.1 and then applies algorithms to the collection of line-integral estimates:

$$l_i = \int_{L_i} \mu(\mathbf{r})\, dl \quad i = 1, \ldots, N.$$

where l_i denotes the line integral of attenuation coefficients of the voxels along the ith ray path. One obtains the estimate g_i of l_i by log-transform as follows:

$$g_i = \log\left(\frac{b_i}{y_i - \bar{r}_i}\right) \tag{5.1}$$

Note that y_i's are noisy measurement data. Inspection of Equation 4.1 reveals that the g_i in Equation 5.1 are identical to l_i if there is no noise present. Since Poisson photon noise is ignored in deterministic methods, ideally we have

$$g_i = l_i \tag{5.2}$$

One then reconstructs estimates $\hat{\boldsymbol{\mu}}$ from $\{g_i\}$ using some deterministic algorithm.

For the discrete case with Equation (4.6) and Equation (4.8), we rewrite Equation (5.2):

$$g_i = [\mathcal{A}\boldsymbol{\mu}]_i = \sum_{j=1}^{M} a_{ij}\mu_j, \quad i = 1, \ldots, N, \quad j = 1, \ldots, M. \tag{5.3}$$

Since (5.3) is a series of linear equations, Equation (5.3) can be represented as

$$\mathcal{A}\boldsymbol{\mu} = \mathbf{g} \tag{5.4}$$

where $\mathbf{g}$ is a line integral vector with N elements. This linear system model is the basis for deterministic methods. The reconstruction is a inverse problem and its goal is to solve this linear system, where $\mathcal{A}$ and $\mathbf{g}$ are known and $\boldsymbol{\mu}$ is unknown. In tomosynthesis, Equation (5.4) is $\mathcal{A}_{tomo}\boldsymbol{\mu} = \mathbf{g}$. In CT geometry, Equation (5.4) is $\mathcal{A}_{CT}\boldsymbol{\mu} = \mathbf{g}$. When the angle difference $\triangle\theta$ between the adjacent angular positions of X-ray source is infinitely small, $\mathcal{A}_{CT}\boldsymbol{\mu}$ becomes the line-integral in continuous space. One can use generalized FBP (Filtered Backprojection) methods to reconstruct the attenuation coefficients $\boldsymbol{\mu}$ perfectly [57] for certain system geometries.

5.1.1 Backprojection methods

Simple backprojection (BP) is a special case of deterministic methods. The BP method is attractive because it is easy to implement and its result can be used as initial values for more sophisticated iterative methods to achieve efficient reconstruction. In tomosynthesis, the BP method is essentially equivalent to a "shift-and-add" algorithm [40]. The shift-and-add algorithm is valid only when the motion of the x-ray source is parallel to the detector (e.g., a linear motion at a fixed height above the detector) [62]. A BP algorithm accurately incorporates the imaging geometry no matter how the x-ray tube is moved. Backprojection is performed by smearing a set of 2D projection images uniformly back into the 3D breast space along the incident x-ray paths. In this operation, the estimated attenuation coefficient of the voxel is proportional to the weighted summation of all the values of detector pixels whose rays intersect this voxel. The attenuation coefficient updates of voxels can be performed in parallel and are given by:

$$\hat{\mu}_j = \frac{\sum_{i=1}^{N} a_{ij} g_i}{\sum_{i=1}^{N} a_{ij}} \tag{5.5}$$

The denominator in Equation 5.5 is a normalization factor, and is itself a backprojection operated on a unit image. Note that BP can be accomplished during data acquisition by simply backprojecting data from each angle as it is obtained. The BP algorithm has its problems, but it is fast. Other researchers [63] have variations of Equation 5.5 with different normalizations.

Tuned aperture computed tomography (TACT) - backprojection [64, 65] has been developed to address some of practical issues in tomosynthesis imaging, and has been applied to breast tomosynthesis imaging [66]. TACT involves an interesting backprojection step. TACT backprojection is similar to shift-and-add except that TACT uses knowledge derived from the projections themselves to establish the geometric relationships required for reconstruction. The projection geometry used in TACT is not required to be fixed (or even specified) before reconstruction. In other words, TACT can be used with an arbitrary geometry. In TACT, a set of reference points are used to determine the precise orientation of the imaging geometry and projection images prior to the shift-and-add process. The alignment of images is based on the reference points. The projection of a reference point is used as an anchor and each projection image is shifted proportionally toward the reference point by an amount determined by the desired relative section position [66]. Shifted projections are added, yielding a composite tomographic image with structures in the reconstruction plane in sharp focus and structures outside the reconstruction plane blurred. This process can be adjusted to bring any arbitrary plane into focus and all other planes blurred by an amount proportional to their distance from the focal plane. While we have discussed TACT in a BP context, other researchers [67, 66] have used it within iterative reconstruction formulations.

Recently, a nonlinear backprojection (NBP) for emission tomography [68] was proposed, which consists of nonlinear operations on the sinogram, followed by back-

projection. NBP is composed of a training phase and an application phase. The projection data is modeled by a mixture model. The goal of the training phase is to infer parameters of the mixture model of the data and to optimize a set of sinogram filters. One can design the filters by minimizing the mean square error between the training images and the images reconstructed by backprojecting the nonlinearly filtered sinogram samples. After this training phase, the inferred parameters and filters are used to reconstruct the images from sinograms. This method does not need to specify a statistical model for the forward process and for the images to be reconstructed. NBP shows promising results relative to both filtered backprojection and maximum *a posteriori* methods. However, training of the filters is still computationally expensive.

5.1.2 Matrix Inversion Techniques

Equation 5.4 is expressed in a matrix form. The goal of reconstruction is to solve this equation. The natural way to solve such a equation is to use matrix inversion techniques. There are some variants in this category such as algebraic reconstruction techniques (ART) [69] and matrix inversion tomosynthesis (MITS) [70]. We shall discuss these methods in detail, but not offer any mathematical derivations since these are beyond the scope of the discussion. Please see the appropriate citations for derivations.

Algebraic Reconstruction Technique (ART)

As an example of deterministic reconstruction methods, algebraic methods with the system model (5.4) are often used. In algebraic reconstruction techniques, the tomographic inverse problem is to solve the large-scale system of linear equations of Equation 5.4. To explain computational steps involved in these methods, we first write

Equation 5.4 in an expanded form:

$$\begin{aligned} a_{11}\mu_1 + a_{12}\mu_2 + \cdots + a_{1M}\mu_M &= g_1 \\ a_{21}\mu_1 + a_{22}\mu_2 + \cdots + a_{2M}\mu_M &= g_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{N1}\mu_1 + a_{N2}\mu_2 + \cdots + a_{NM}\mu_M &= g_N \end{aligned} \tag{5.6}$$

Under ideal conditions, such simultaneous equations could be solved exactly by matrix inversion. However, such a direct matrix inversion approach is not feasible due to the following facts [22]:

- The numbers of M and N are prohibitively large.
- A unique solution will not exist if the equations are under-determined ($N < M$).
- No solution may exist when $N > M$, and $\{g_i\}$ are corrupted by noise.

In such circumstance, iterative methods are applied to solve these equations without matrix inversion. To implement such a method, one can first make the initial guess at the solution. This initial guess, denoted by $\hat{\mu}_1^{(0)}, \hat{\mu}_2^{(0)}, \cdots, \hat{\mu}_M^{(0)}$, is represented by the vector $\hat{\boldsymbol{\mu}}^{(0)}$ in the M-dimensional space. In most cases, one can simply set all the initial $\hat{\mu}_j$'s to be zero. This initial guess is projected on the hyperplane represented by the first equation in (5.6) giving $\hat{\boldsymbol{\mu}}^{(1)}$. Then $\hat{\boldsymbol{\mu}}^{(1)}$ is projected on the hyperplane represented by the second equation in (5.6) to yield $\hat{\boldsymbol{\mu}}^{(2)}$ and so on. When $\hat{\boldsymbol{\mu}}^{(i-1)}$ is projected on the hyperplane represented by the ith equation in (5.6) to yield $\hat{\boldsymbol{\mu}}^{(i)}$, the process can be mathematically described by

$$\hat{\boldsymbol{\mu}}^{(i)} = \hat{\boldsymbol{\mu}}^{(i-1)} - \frac{\hat{\boldsymbol{\mu}}^{(i-1)} \cdot \mathbf{a}_i - g_i}{\mathbf{a}_i \cdot \mathbf{a}_i}\mathbf{a}_i \tag{5.7}$$

where $\mathbf{a}_i = [a_{i1}, a_{i2}, \cdots, a_{iM}]$ and $\mathbf{a}_i \cdot \mathbf{a}_i$ is the dot product of $\mathbf{a}_i$ with itself.

In applications requiring a large number of views (large N) and large-sized (large M) reconstructions, the difficulty with using Equation 5.6 can be in the calculation,

storage, and fast retrieval of w_{ij}. To get around the implementation difficulties, several algebraic approaches have been suggested, many of which are approximations to (5.7). To discuss some of the more implementable approximations, one can rewrite Equation 5.7 in a different form [69]:

$$\hat{\mu}_j^{(i)} = \hat{\mu}_j^{(i-1)} + \frac{g_i - q_i}{\sum_{k=1}^{M} a_{ik}^2} a_{ij} \tag{5.8}$$

where

$$q_i = \hat{\boldsymbol{\mu}}^{(i-1)} \cdot \mathbf{a}_i = \sum_{k=1}^{M} \hat{\mu}_k^{(i-1)} a_{ik} \tag{5.9}$$

These equations say that when the $(i-1)$th solution is projected onto the ith hyperplane (the ith equation in Equation 5.6), the new value of the jth parameter is obtained by correcting its current value $\hat{\mu}_j^{(i-1)}$ by $\triangle\mu_j^{(i)}$, where

$$\triangle\mu_j^{(i)} = \hat{\mu}_j^{(i)} - \hat{\mu}_j^{(i-1)} = \frac{g_i - q_i}{\sum_{k=1}^{M} a_{ik}^2} a_{ij}. \tag{5.10}$$

While g_i is the measured line-integral along the ith ray, q_i can be understood to be the computed line-integral for the same ray based on the $(i-1)$th solution. The correction $\triangle\mu_j$ to the jth voxel is obtained by first calculating the difference between the measured line-integral and the computed line-integral, normalizing this difference by $\sum_{k=1}^{M} a_{ik}^2$, and then assigning this value to all voxels in the ith ray, each assignment being weighted by the corresponding a_{ij} [71]. Combining Equations 5.8, 5.9, 5.10 yields:

$$\hat{\mu}_j^{(i)} = \hat{\mu}_j^{(i-1)} + \frac{g_i - \sum_{j=1}^{M} \hat{\mu}_j^{(i-1)} a_{ij}}{\sum_{j=1}^{M} a_{ij}^2} a_{ij} \tag{5.11}$$

The linear attenuation coefficients are updated at each projection view, therefore ART converges fast to a least squares solution which can be very noisy for severely ill-posed inverse problem such as limited-angle tomosynthetic reconstruction [63]. There are some variations on its computer implementation. ART has been modified to other methods such as simultaneous algebraic reconstruction technique (SART) [69] and simultaneous iterative reconstruction technique (SIRT) [69]. In SART, the correction

terms are simultaneously applied for all the rays in one projection, and the linear attenuation coefficient of each voxel is updated after all rays passing through this voxel at one projection view are processed; while in SIRT, all projection rays are calculated and applied to a correction factor, and the update is performed after all rays in all projection views are processed. SIRT converges slowly because its update is averaged over all projection rays and the reconstruction could be overly smoothed. We shall describe the SART method in detail.

As mentioned above, for SART, the value update of each voxel is performed after all rays at one projection view are processed. The number of updates in one full iteration is equal to the number of projection views K, and also is called as the number of subiterations. Let $\hat{\mu}_j^{n,k}$ denote the estimated linear attenuation coefficient of the jth voxel at the end of the k subiteration of the nth iteration. The initial and final update values at one iteration are assigned as follows:

$$\hat{\mu}_j^{n,1} = \hat{\mu}_j^{n-1}, \quad \hat{\mu}_j^{n} = \hat{\mu}_j^{n,K} \tag{5.12}$$

where $\hat{\mu}_j^n$ is the estimate at the end of the nth iteration, which is equal to the estimate after all K projection views are processed. Let $\mathcal{G}_j^k$ denote the set of the measured line integrals passing through the jth voxel at the kth projection angle. The update of the linear attenuation coefficient at the jth voxel is defined as follows:

$$\hat{\mu}_j^{n,k+1} = \hat{\mu}_j^{n,k} + \lambda \frac{\sum_{g_i \in \mathcal{G}_k} \left(a_{ij} \frac{g_i - \sum_{j=1}^{M} a_{ij} \hat{\mu}_j^{n,k}}{\sum_{j=1}^{M} a_{ij}} \right)}{\sum_{g_i \in \mathcal{G}_k} a_{ij}} \tag{5.13}$$

where λ is a relaxation factor ranged over $(0, 1]$. The relaxation factor is used to reduce the noise during reconstruction. In some cases, this parameter is chosen as a function of the iteration number. That is, λ decreases as the number of iterations increases.

In general, the SART procedure implementation can be summarized as follows:

1. Take an initial guess $\boldsymbol{\mu}^0$.

2. Compute the estimated line integrals $\mathbf{g^{n,k}}$ at kth angle and nth iteration.

3. Subtract the estimated line integrals $\mathbf{g^{n,k}}$ from $\mathbf{g}$.

4. Perform backprojection of correction terms at kth angle and update $\boldsymbol{\mu}^{n,k}$ values.

5. Repeat from step 2 with next view until all $\mathcal{K}$ views are completed, and update $\boldsymbol{\mu}^{n}$ values.

6. Repeat from step 2 with next new iteration or stop with final breast estimate.

The choice of the initial guess is very important. This is a general problem for iterative reconstruction methods. A good choice of initial condition can speed up the convergence. An initial condition chosen to the final solution will speed up convergence. In some cases, a simple BP reconstruction serves as a good initial condition. Another problem associated with iterative reconstruction approaches is the choice of a stopping point. An appropriate stopping point determines the image quality of reconstruction. Earlier termination results in a low-contrast reconstruction, while more iterations yields more noisy estimates. Given this compromise, the "optimal" number of iterations is often predetermined based on visual comparison and image quality analysis. Because iterations are terminated before convergence, the iterative methods do not exactly invert Equation 5.4.

Matrix Inversion Tomosynthesis (MITS)

Matrix inversion tomosynthesis technique [40] is mainly used in tomosynthetic imaging such as breast tomosynthesis. In MITS, linear algebra is applied to solve for the relative blurring in each plane. Given that the geometry of the imaging system is known, the amount of blurring f at each plane can be calculated when every other plane is tomosynthesized. The reconstructed image t_i by the shift-and-add method

can be represented as the convolution of the structures s_i in the ith plane and the blurring function f_{ij} from the structures in plane j:

$$\begin{aligned} s_1 \otimes f_{11} + s_2 \otimes f_{12} + \cdots + s_n \otimes f_{1n} = t_1 \\ s_1 \otimes f_{21} + s_2 \otimes f_{22} + \cdots + s_n \otimes f_{2n} = t_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ s_n \otimes f_{n1} + s_2 \otimes f_{n2} + \cdots + s_n \otimes f_{nn} = t_1 \end{aligned} \tag{5.14}$$

The convolutional formalism in Equation 5.14 is limited to those acquisition geometries, such as a linear tube motion, where the blur function is space invariant. If the Fourier transform is taken, the above equations can be expressed as:

$$\begin{pmatrix} F_{11} & F_{12} & \cdots & F_{1n} \\ F_{21} & F_{22} & \cdots & F_{2n} \\ \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots \\ F_{n1} & F_{n2} & \cdots & F_{nn} \end{pmatrix} \begin{pmatrix} S_1 \\ S_2 \\ .. \\ S_n \end{pmatrix} = \begin{pmatrix} T_1 \\ T_2 \\ .. \\ T_n \end{pmatrix} \tag{5.15}$$

One can rewrite Equation 5.15 in matrix form: $F.S = T$, where S is the matrix of Fourier transform of the structures and F is the matrix of Fourier transform of the blurring functions. Therefore, the structures s can be obtained by taking inverse Fourier transform over $F^{-1}T$. The MITS is a good technique to deal with the mid and high spatial frequency components of the image, while the low frequency noise tends to be present in the resulting reconstructions [40]. Other methods have developed to reduce low-frequency noise.

5.1.3 Filtered Backprojection (FBP) Methods

In the simple BP method as discussed above, one takes the line integral at each pixel in the projection and smears it back along the path of the incident ray. If this process is performed for each pixel in a projection and for all projection views, one can obtain a simple backprojected estimate of the object. Consider a 2D object with

parallel-beam projections. If we had a delta function object, the backprojected reconstruction appear as a "spoke pattern" and its Fourier transform falls off linearly as the radial spatial frequency increases. Filtering the 2D "spoke pattern" image by a 2D ramp filter will restore the backprojection to the true object. By the central slice theorem [22], one could equivalently do this filtering in the projection domain. Filtering the projection image by multiplying its Fourier transform by a ramp function (proportional to radial spatial frequency) and then backprojecting them, one can reconstruct the point, and by linearity, an entire object. This process is referred to as filtered backprojection (FBP). The 2D FBP algorithm for complete parallel-beam angular sampling is well known and described in many textbooks [22, 69] and we will not give the mathematical description here.

The 2D FBP methods for parallel and fan-beam projections are frequently used in transmission and emission CT [72], in which a large number of projection images acquired over $360^o/180^o$ are used to reconstruct cross-sectional images. With a large number of projections, the information in the object is well sampled and the corresponding spatial frequency domain is well sampled, so the object can be restored by combing the information from all projections. In 2D FBP methods, the Fourier central slice theorem [22] is a fundamental key. With a parallel-beam approximation, the Fourier transform of a projection yields a plane through the 2D Fourier space of the object along the direction perpendicular to the x-ray beam [22]. As the x-ray source and the detector are rotated around the object, a set of data is swept out in 2D Fourier space. The central slice theorem holds for a projection at any angle. For limited angle acquisitions, the theorem tells us which parts of Fourier space are not being sampled.

FBP methods also exist for 3D data, and these can be applied to the tomosynthesis problems. As in 2D, 3D FBP reconstructions obey a form of central slice theorem. In tomosynthetic reconstruction, only a limited angular range is swept during acqui-

sition, not all of the volume of 3D Fourier space is sampled. The main limitation for tomosynthesis is the incomplete angular sampling of the object. The cone-beam geometry is used in breast tomosynthesis. The relation between the cone-beam projections and the Radon transform was presented by Smith [73], and the solution to the general cone-beam reconstruction also was provided. For 3D cone-beam breast tomosynthesis, the Feldkamp method [74], an approximation of the cone-beam FBP algorithm, was explored by Wu *et al.* [62] but their reconstruction results were very noisy, along with artifacts. The result of the Feldkamp algorithm indicates that the low-frequency contrast restoration should be considered. A prototype digital tomosynthesis mammography system using a modified FBP method was reported in [45], but no details were given.

In our study, we do not choose the FBP algorithm since FBP is handicapped by limited-angle problem in breast tomosynthesis where the object spatial frequency domain is far from completely sampled. The image quality of the reconstruction strongly depends on the filter design, and other researchers continue to investigate analytical FBP methods for limited-angle problems.

So far we have discussed some commonly used deterministic methods. They have several deficiencies. First, these reconstruction approaches do not take the Poisson nature of noise statistics into account. Second, Equation (5.1) produces the biased line-integral estimates due to the nonlinearity of the logarithm. Specifically, it overestimates the line-integral on the average: $E[g_i] \geq l_i$ [75, 76]. Therefore, the log-transform is unsuitable for low-count transmission scans. In addition, the FBP method is not suited to nonstandard transmission imaging geometries, such as truncated fan-beam or cone-beam scans [77, 78, 79, 80, 81, 82]. To eliminate these problems, one can use statistical methods. These methods use the observed "raw" measurements y_i's rather than log-transformed data g_i's. We shall discuss some statistical methods

in the following section.

5.2 Statistical Reconstruction Methods

5.2.1 Maximum Likelihood Reconstruction

As discussed in Section 4.2, the Poisson noise distribution is modeled in Equation 4.12. Since the Poisson nature of x-ray generation implies that the various projections are independent, the measurement joint probability function given $\boldsymbol{\mu}$ is:

$$P[\mathbf{y}|\boldsymbol{\mu}] = \prod_{i=1}^{N} P[y_i|\boldsymbol{\mu}] = \prod_{i=1}^{N} \frac{e^{-\bar{y}_i}\bar{y}_i^{y_i}}{y_i!} \tag{5.16}$$

where the mean of the ith measurement is given by (4.11):

$$\bar{y}_i(\boldsymbol{\mu}) = b_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i) + \bar{r}_i$$

Equation 5.16 is a likelihood. A likelihood $P(b|a)$ is a measure of the likelihood that the presence of a causes the observed b to occur. For reasons made obviously below, usually, it is its logarithm that people work with, which is referred to as *log likelihood*. The log-likelihood function for the independent projection data in transmission tomography is

$$\log P[\mathbf{y}|\boldsymbol{\mu}] = \sum_{i=1}^{N} y_i \log \bar{y}_i - \bar{y}_i - \log(y_i!)$$

$$= \sum_{i=1}^{N} y_i \log(b_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i) + \bar{r}_i) - (b_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i) + \bar{r}_i) - \log(y_i!)$$

Ignoring constant terms independent of $\boldsymbol{\mu}$, then one can rewrite the log-likelihood as:

$$L(\boldsymbol{\mu}) = \sum_{i=1}^{N} y_i \log(b_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i) + \bar{r}_i) - (b_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i) + \bar{r}_i) \tag{5.17}$$

The maximum likelihood (ML) is a classical estimation approach. One seeks to find a statistical estimate of the attenuation map $\boldsymbol{\mu}$ which most likely agrees with the given

acquired projection data. Therefore, the ML estimate $\hat{\boldsymbol{\mu}}$ is defined as follows:

$$\begin{aligned}
\hat{\boldsymbol{\mu}} &= \arg\max_{\boldsymbol{\mu}\geq\mathbf{0}} P[\mathbf{y}|\boldsymbol{\mu}] \\
&= \arg\max_{\boldsymbol{\mu}\geq\mathbf{0}} L(\boldsymbol{\mu}) \\
&= \arg\max_{\boldsymbol{\mu}\geq\mathbf{0}} \sum_{i=1}^{N} y_i \log(b_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i) + \bar{r}_i) - (b_i \exp(-[\mathcal{A}\boldsymbol{\mu}]_i) + \bar{r}_i) \quad (5.18)
\end{aligned}$$

The second equality is true because the logarithm is a monotonic function. The ML method seeks the nonnegative parameter vector $\boldsymbol{\mu}$ (the object) that maximizes the probability of having observed the particular measurements that were recorded [60]. The ML principle applied to transmission tomography was first proposed by Rockmore and Macovski [83] in 1977. Lange and Carson proposed the first practical algorithm — a transmission EM algorithm [84] in 1984.

ML-EM Algorithms

The expectation maximization (ML-EM) algorithm is one iterative technique for optimizing the maximum-likelihood estimate. The EM algorithm is a general method for obtaining ML estimates and has been applied specifically to medical image reconstruction in transmission and emission [85] imaging. We first discuss an EM algorithm for transmission proposed in [84]. We shall refer to this EM algorithm as the Lange-Carson EM [84].

This EM method incorporates the observed incomplete data $\mathbf{y}$ into a complete data space. The complete data is composed of the number of photons entering and leaving each voxel along each x-ray. With this complete data, it is easier to find a ML solution. Let U_{ij} and V_{ij} denote the numbers of photons entering and leaving each voxel j along each projection ray i. Briefly speaking, EM is an iterative procedure that proceeds in two steps at each iteration: an expectation step (E-step) and a maximization step (M-step). In the E-step, one computes the expectations of the

complete data given the incomplete data $\mathbf{y}$ and the current estimate $\boldsymbol{\mu}^n$, denoted by $M_{ij} = E(U_{ij}|y_i, \boldsymbol{\mu}^n)$ and $N_{ij} = E(V_{ij}|y_i, \boldsymbol{\mu}^n)$, where E denotes expectation. The complete data log-likelihood function $Q_0(\boldsymbol{\mu}|\boldsymbol{\mu}^n)$ for the Lange-Carson EM is given by

$$Q_0(\boldsymbol{\mu}|\boldsymbol{\mu}^n) = \sum_i \sum_j [-N_{ij} a_{ij} \mu_j + (M_{ij} - N_{ij}) ln(1 - e^{-a_{ij}\mu_j})] \tag{5.19}$$

where

$$M_{ij} = y_i + b_i \exp\left(-\sum_{k \in S_{ij}} a_{ik}\mu_k^n\right) - b_i \exp\left(-[\mathcal{A}\boldsymbol{\mu}]_i\right) \tag{5.20}$$

$$N_{ij} = y_i + b_i \exp\left(-\sum_{k \in S_{ij} \cup \{j\}} a_{ik}\mu_k^n\right) - b_i \exp\left(-[\mathcal{A}\boldsymbol{\mu}]_i\right) \tag{5.21}$$

where S_{ij} is the set of voxels between the x-ray source and the voxel j along the ray i. In the M-step, the log-likelihood function $Q_0(\boldsymbol{\mu}|\boldsymbol{\mu}^n)$ is maximized with respect to $\boldsymbol{\mu}$ assuming the expected complete data from the E step are correct. Taking the partial derivative of $Q_0(\boldsymbol{\mu}|\boldsymbol{\mu}^n)$ with respect to μ_j, setting it to zero and then solving the equation for μ_j yields the final form for the Lange-Carson EM algorithm [84]:

$$\mu_j^{n+1} = \frac{\sum_i (M_{ij} - N_{ij})}{\frac{1}{2}\sum_i (M_{ij} + N_{ij}) a_{ij}} \tag{5.22}$$

Thus Equations 5.20, 5.21 and 5.22 constitute an EM algorithm for transmission tomography.

The transmission ML-EM algorithm has its virtues: (1) It automatically includes the positivity constraint on all parameters. (2) It has been proved to globally convergent [86]. However, the ML-EM has drawbacks that limit its use in routine practical applications. One is the low speed of convergence and the expensive computational load due to the large number of exponentiations required in each iteration. Researchers have proposed some strategies to accelerate the convergence [87, 88, 89]. Another drawback is that it generates noisy images at high iteration. Particularly, when the counts are low, the signal-to-noise ratio is low for the Poisson distributed data, and the ML reconstruction is extremely noisy since the ML estimate is ill posed.

To overcome this problem, the commonly used solutions are stopping rules, post-smoothing and regularizers. The method incorporating a regularizer (or a prior) is a powerful extension of the ML reconstruction method. It is termed a Bayesian reconstruction method, and one type of Bayesian estimate is a maximum *a posteriori* (MAP) estimate, because it maximize the posterior in Bayesian theory. EM solutions exist for the MAP problem. We shall discuss MAP in next section.

An improvement on the original transmission Lange-Carson ML-EM algorithm is convex ML-EM algorithm [90]. It requires $\bar{r}_i = 0$, otherwise the likelihood is not convex and the algorithm does not work. The convex ML-EM still is guaranteed to monotonically increase $L(\boldsymbol{\mu})$ each iteration. Assuming that $\bar{r}_i = 0$ and dropping the constant term independent of $\boldsymbol{\mu}$, the likelihood $L(\boldsymbol{\mu})$ in Equation 5.17 can be rewritten in the following form:

$$L(\boldsymbol{\mu}) = \sum_{i=1}^{N} f_i([\mathcal{A}\boldsymbol{\mu}]_i) \tag{5.23}$$

where

$$f_i(l) = y_i \log(b_i e^{-l} + \bar{r}_i) - (b_i e^{-l} + \bar{r}_i) \tag{5.24}$$

is a strictly concave function.

By using convexity tricks applied to f as proposed in [91, 92], one obtains

$$\begin{aligned} L(\boldsymbol{\mu}) &= \sum_{i=1}^{N} f_i(\sum_{j=1}^{M} \frac{a_{ij}\mu_j^n}{[\mathcal{A}\boldsymbol{\mu}^n]_i} \frac{\mu_j}{\mu_j^n} [\mathcal{A}\boldsymbol{\mu}^n]_i) \\ &\geq \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{a_{ij}\mu_j^n}{[\mathcal{A}\boldsymbol{\mu}^n]_i} f_i(\frac{\mu_j}{\mu_j^n} [\mathcal{A}\boldsymbol{\mu}^n]_i) \\ &\equiv Q(\boldsymbol{\mu}|\boldsymbol{\mu}^n) \end{aligned} \tag{5.25}$$

When $\boldsymbol{\mu} = \boldsymbol{\mu}^n$, the inequality above becomes an equality, namely, $L(\boldsymbol{\mu}^n) = Q(\boldsymbol{\mu}^n|\boldsymbol{\mu}^n)$. The difference $L(\boldsymbol{\mu}) - Q(\boldsymbol{\mu}|\boldsymbol{\mu}^n)$ has its minimal point at $\boldsymbol{\mu} = \boldsymbol{\mu}^n$, thus,

$$L(\boldsymbol{\mu}^{n+1}) - Q(\boldsymbol{\mu}^{n+1}|\boldsymbol{\mu}^n) \geq L(\boldsymbol{\mu}^n) - Q(\boldsymbol{\mu}^n|\boldsymbol{\mu}^n) \tag{5.26}$$

Now $\boldsymbol{\mu}^{n+1}$ is chosen to maximize $Q(\boldsymbol{\mu}|\boldsymbol{\mu}^n)$, thus,

$$Q(\boldsymbol{\mu}^{n+1}|\boldsymbol{\mu}^n) \geq Q(\boldsymbol{\mu}^n|\boldsymbol{\mu}^n) \tag{5.27}$$

With inequalities 5.26 and 5.27, $L(\boldsymbol{\mu}^{n+1})$ satisfies

$$\begin{aligned} L(\boldsymbol{\mu}^{n+1}) &= L(\boldsymbol{\mu}^{n+1}) - Q(\boldsymbol{\mu}^{n+1}|\boldsymbol{\mu}^{n}) + Q(\boldsymbol{\mu}^{n+1}|\boldsymbol{\mu}^{n}) \\ &\geq L(\boldsymbol{\mu}^{n}) - Q(\boldsymbol{\mu}^{n}|\boldsymbol{\mu}^{n}) + Q(\boldsymbol{\mu}^{n}|\boldsymbol{\mu}^{n}) \\ &= L(\boldsymbol{\mu}^{n}) \end{aligned} \tag{5.28}$$

Therefore, $L(\boldsymbol{\mu})$ is monotonically increased at each iteration.

To maximize $Q(\boldsymbol{\mu}|\boldsymbol{\mu}^n)$, one sets its partial derivative to zero:

$$\begin{aligned} \frac{\partial}{\partial \mu_j} Q(\boldsymbol{\mu}|\boldsymbol{\mu}^{n}) &= -\sum_{i=1}^{N} a_{ij} f_i^{'}(\frac{\mu_j}{\mu_j^n}[\mathcal{A}\boldsymbol{\mu}^n]_i) \\ &= -\sum_{i=1}^{N} a_{ij}[-b_i e^{-\frac{\mu_j}{\mu_j^n}[\mathcal{A}\boldsymbol{\mu}^n]_i} + y_i] \\ &= 0 \end{aligned} \tag{5.29}$$

Equation 5.29 can not be solved directly. One can apply Newton's method to this equation. The second-order partial derivative of $Q(\boldsymbol{\mu}|\boldsymbol{\mu}^n)$ is given by

$$\begin{aligned} \frac{\partial^2}{\partial \mu_j^2} Q(\boldsymbol{\mu}|\boldsymbol{\mu}^{n})|_{\boldsymbol{\mu}=\boldsymbol{\mu}^n} &= -\sum_{i=1}^{N} \frac{a_{ij}}{\mu_j^n}[\mathcal{A}\boldsymbol{\mu}^n]_i f_i^{''}(\frac{\mu_j^n}{\mu_j^n}[\mathcal{A}\boldsymbol{\mu}^n]_i) \\ &= -\sum_{i=1}^{N} \frac{a_{ij}}{\mu_j^n}[\mathcal{A}\boldsymbol{\mu}^n]_i b_i e^{-[\mathcal{A}\boldsymbol{\mu}^n]_i} \end{aligned}$$

and its first-order partial derivative at $\boldsymbol{\mu} = \boldsymbol{\mu}^n$ is given by

$$\frac{\partial}{\partial \mu_j} Q(\boldsymbol{\mu}|\boldsymbol{\mu}^{n})|_{\boldsymbol{\mu}=\boldsymbol{\mu}^n} = \frac{\partial}{\partial \mu_j} L(\boldsymbol{\mu}^n) \tag{5.30}$$

Therefore, one step of Newton's method yields the approximate solution [90]:

$$\mu_j^{n+1} = \mu_j^n + \mu_j^n \frac{\sum_{i=1}^{N} a_{ij} \left(1 - \frac{y_i}{b_i e^{-[\mathcal{A}\boldsymbol{\mu}^n]_i}}\right) b_i e^{-[\mathcal{A}\boldsymbol{\mu}^n]_i}}{\sum_{i=1}^{N} a_{ij}[\mathcal{A}\boldsymbol{\mu}^n]_i b_i e^{-[\mathcal{A}\boldsymbol{\mu}^n]_i}} \tag{5.31}$$

The summation over rays in the numerator represents a backprojection of the difference image between the current projection estimate and the observed projection, normalized by the backprojected volume in the denominator. We note that each iteration requires one forward projection (to compute $[\mathcal{A}\boldsymbol{\mu}^n]_i$) and two backprojections

(two summations over i in the numerator and denominator), and provides parallel voxel updates. The non-negativity constraint is easy to impose in each iteration. Compared to the Lange-Carson ML-EM method which requires a large amount of exponentiations, this convex EM algorithm requires less computation per iteration. We will adopt the convex ML-EM method in our project.

5.2.2 Maximum *a posteriori* (MAP) Reconstruction

In the maximum *a posteriori* (MAP) estimation approach we choose $\hat{\boldsymbol{\mu}}$ to maximize the posterior PDF $P(\boldsymbol{\mu}|\mathbf{y})$:

$$\hat{\boldsymbol{\mu}} = \arg\max_{\boldsymbol{\mu} \geq \mathbf{0}} P(\boldsymbol{\mu}|\mathbf{y}) \tag{5.32}$$

By Bayes rule:

$$P(\boldsymbol{\mu}|\mathbf{y}) = \frac{P(\mathbf{y}|\boldsymbol{\mu})P(\boldsymbol{\mu})}{P(\mathbf{y})} \tag{5.33}$$

or, equivalently,

$$\log P(\boldsymbol{\mu}|\mathbf{y}) = \log P(\mathbf{y}|\boldsymbol{\mu}) + \log P(\boldsymbol{\mu}) - \log P(\mathbf{y}) \tag{5.34}$$

where $P(\boldsymbol{\mu})$ is the prior in MAP. Ignoring the term independent of $\boldsymbol{\mu}$, one can redefine the objective function to be maximized as:

$$\Phi(\boldsymbol{\mu}) = \log P(\mathbf{y}|\boldsymbol{\mu}) + \log P(\boldsymbol{\mu}) \tag{5.35}$$

$$\equiv L(\boldsymbol{\mu}) - \beta R(\boldsymbol{\mu}) \tag{5.36}$$

where $L(\boldsymbol{\mu})$ was defined earlier in Equation 5.17 and $R(\boldsymbol{\mu})$ is a "penalty" function, discussed below, that tends to smooth the reconstruction. The parameter $\beta \geq 0$ is a scalar that controls the tradeoff between spatial resolution and noise. Larger values of β result in less noisy or smoother reconstructed images at the expense of reduced spatial resolution. Equation 5.36 is referred to as a penalized-likelihood objective

function. Thus, the MAP estimator is given as:

$$\hat{\boldsymbol{\mu}} = \arg\max_{\boldsymbol{\mu}\geq\mathbf{0}} L(\boldsymbol{\mu}) - \beta R(\boldsymbol{\mu}) \tag{5.37}$$

A typical penalty function is given below [93]:

$$R(\boldsymbol{\mu}) = \frac{1}{2}\sum_{j=1}^{M}\sum_{k\in\mathcal{N}_j} w_{jk} v(\mu_j - \mu_k) \tag{5.38}$$

where $\mathcal{N}_j$ is a local neighborhood of the jth voxel, and $w_{jk} = w_{kj} \geq 0$. Typically, w_{jk} is inversely proportional to the distance between the centers of voxel j and voxel k. The potential function v is symmetric, twice differentiable and strictly convex. There are many possible choices for the potential function v satisfying these conditions. One simple form is $v(t) = \frac{1}{2}t^2$ [94]. This choice tends to smooth local noise fluctuations , but deters the formation of discontinuities. More sophisticated revisions of v can promote the formation of discontinuities, sometimes desired, but Equation 5.38 leads to more tractable analysis.

There are various techniques to optimize the penalized-likelihood objective function $\Phi(\boldsymbol{\mu})$. The paraboloidal surrogates algorithm [88] is one example. Such methods easily enforce the non-negativity constraint, improve the convergence rate relative to less recent methods and ensure monotonicity even for the non-concave objective function that arises due to background events $\bar{r}_i$. We shall discuss the paraboloidal surrogates algorithms in detail below.

Paraboloidal Surrogate Algorithm

Since the objective function Φ is difficult to maximize, at each iteration we can use the optimization transfer idea proposed by De Pierro [91, 92]. This solves many of the problems such as speed and imposition of a positivity constraint that gradient ascent algorithms have. The basic idea is to replace, at each iteration, Φ with a "surrogate" function ϕ that is easier to maximize. We summarize as follows. Let $\boldsymbol{\mu}^n$ be

the attenuation map estimate at the nth iteration. One would like to find a surrogate function $\phi(\boldsymbol{\mu}; \boldsymbol{\mu}^n)$ that is easier to maximize than $\Phi(\boldsymbol{\mu})$. This approach transforms the optimization problem into a simpler problem at each iteration as illustrated in Fig. 5.1 . If one chooses the surrogates appropriately, the sequence of estimates should eventually converge to a maximizer [95].

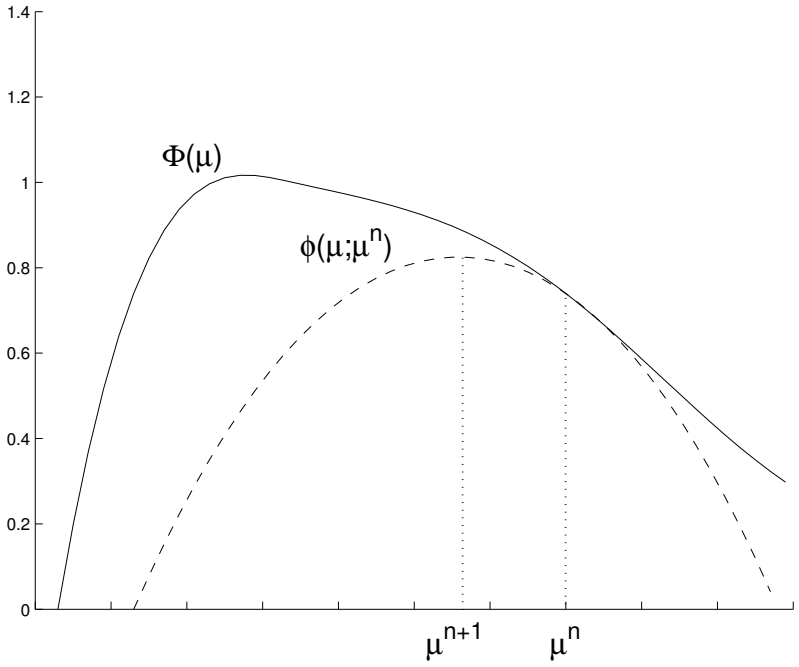


Figure 5.1: One-dimensional illustration of the optimization transfer principle. Instead of maximizing $\Phi(\mu)$, we maximize the surrogate function $\phi(\mu; \mu^n)$ at the nth iteration to obtain μ^{n+1}. Then a new surrogate $\phi(\mu; \mu^{n+1})$ is constructed to find μ^{n+2} and so on.

The following monotonicity condition on the surrogate function is sufficient to ensure that the iterates $\{\boldsymbol{\mu}^n\}$ monotonically increase Φ:

$$\Phi(\boldsymbol{\mu}) - \Phi(\boldsymbol{\mu}^n) \geq \phi(\boldsymbol{\mu}; \boldsymbol{\mu}^n) - \phi(\boldsymbol{\mu}^n; \boldsymbol{\mu}^n), \quad \forall \boldsymbol{\mu} \geq 0 \tag{5.39}$$

The surrogate functions are restricted to be differentiable. Any differentiable surrogate function that satisfies the following conditions will satisfy (5.39).

1.

$$\phi(\boldsymbol{\mu}^n; \boldsymbol{\mu}^n) = \Phi(\boldsymbol{\mu}^n) \tag{5.40}$$

2.

$$\frac{\partial}{\partial \mu_j}\phi(\boldsymbol{\mu}; \boldsymbol{\mu}^n)\Big|_{\boldsymbol{\mu}=\boldsymbol{\mu}^n} = \frac{\partial}{\partial \mu_j}\Phi(\boldsymbol{\mu})\Big|_{\boldsymbol{\mu}=\boldsymbol{\mu}^n}, \; j = 1, \ldots, M \tag{5.41}$$

3.

$$\phi(\boldsymbol{\mu}; \boldsymbol{\mu}^n) \leq \Phi(\boldsymbol{\mu}), \; \forall \boldsymbol{\mu} \geq 0 \tag{5.42}$$

Fig. 5.1 illustrates a surrogate function that "touch" Φ at the current estimate (condition 1), is tangent to the original objective at the current estimate (condition 2) and lies below it for all feasible values of the parameters (condition 3).

One way to derive $\phi(\boldsymbol{\mu}; \boldsymbol{\mu}^n)$ is to separately construct simpler surrogates for the likelihood part $L(\boldsymbol{\mu})$ and the penalty part $R(\boldsymbol{\mu})$. This leads to a simultaneous update algorithm called separable paraboloidal surrogates (SPS) algorithm [88] in which all pixels can be updated in parallel. Parallel update algorithms in transmission tomography are more computationally efficient per iteration than sequential update algorithm in which all pixels are updated in sequence, however they tend to converge slowly, but there are ways to speed them up. We shall discuss this below.

The log-likelihood function (5.23) is the sum of functions f_i defined in Equation 5.24 and depends on the parameter $\boldsymbol{\mu}$ only through $[\mathcal{A}\boldsymbol{\mu}]_i$. A 1-D surrogate function for each of the 1-D f_i functions is chosen in terms of $l_i = [\mathcal{A}\boldsymbol{\mu}]_i$. The overall sum of these individual 1D functions will be an appropriate surrogate for the likelihood part $L(\boldsymbol{\mu})$ of the objective $\Phi(\boldsymbol{\mu})$. For convenience, let $l_i^n = [\mathcal{A}\boldsymbol{\mu}^n]_i$ denote the ith line integral of the attenuation coefficient estimate at the nth iteration. One can choose the following quadratic function for the surrogate functions h_i:

$$h_i(l; l_i^n) = f_i(l_i^n) + \dot{f}_i(l_i^n)(l - l_i^n) - \frac{1}{2}c_i(l_i^n)(l - l_i^n)^2 \tag{5.43}$$

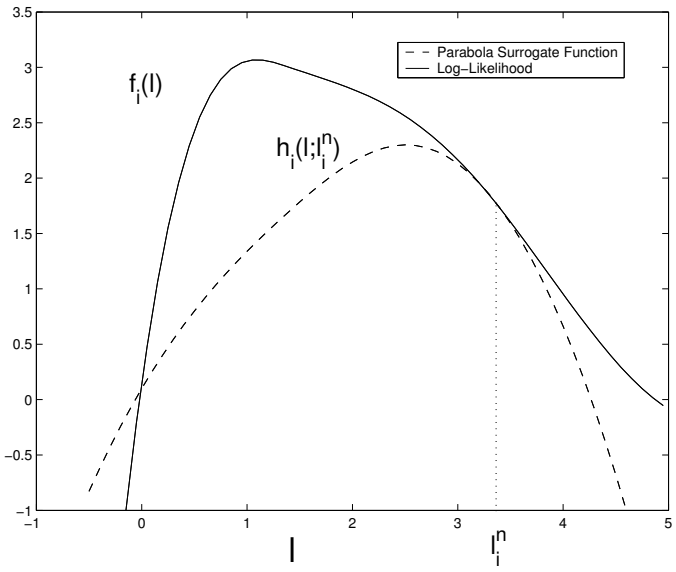


Figure 5.2: 1D parabola surrogate function $h_i(l; l_i^n)$ for log-likelihood $f_i(l)$.

where $c_i(l_i^n) \geq 0$ is the curvature of the parabola h_i and $\dot{f}$ denotes first derivative of f. Fig. 5.2 shows a parabola surrogate function $h_i(l; l_i^n)$ for $f_i(l)$. This construction ensures that $h_i(l_i^n; l_i^n) = f_i(l_i^n)$ and $\dot{h}_i(l_i^n; l_i^n) = \dot{f}_i(l_i^n)$, which satisfy conditions 1 and 2. To ensure monotonicity, one must choose the curvatures to satisfy the following inequality at each iteration:

$$f_i(l) \geq h_i(l; l_i^n), \, \forall l \geq 0 \tag{5.44}$$

The optimum curvature that provides the fastest convergence rate while preserving monotonicity [96] is given below:

$$c_i(l_i^n) = \begin{cases} \left[-2 \frac{f_i(0) - f_i(l_i^n) + \dot{f}_i(l_i^n) l_i^n}{(l_i^n)^2} \right]_+, & l_i^n \geq 0 \\ \left[\ddot{f}_i(0) \right]_+, & l_i^n = 0 \end{cases} \tag{5.45}$$

where $[x]_+$ is x if $x > 0$ and zero otherwise. Erdogan and Fessler [94] show that the parabola that :

1. tangent to f_i at the current line integral l_i^n, and

2. intersects f_i at $l = 0$,

is guaranteed to lie below f_i, $\forall l > 0$. This claim is true only when the curvature $c_i(l_i^n)$ is nonnegative. If the curvature obtained is negative, then $c_i(l_i^n)$ is set to zero. When $c_i(l_i^n) = 0$, the h_i function is the line which is tangent to the f_i curve at the current line integral value l_i^n. The curvature in (5.45) is shown to yield the fastest convergence rate while still guaranteeing monotonicity [88]. The non-negativity constraint plays an important role. If non-negativity is not enforced, the line integrals at an iteration may go negative and the curvature (5.45) will not guarantee monotonicity anymore.

With all of these considerations, the parabolas $h_i(l; l_i^n)$ are summed up to obtain an overall paraboloidal surrogate function for the log-likelihood $L(\boldsymbol{\mu})$ as below:

$$H_1(\boldsymbol{\mu}; \boldsymbol{\mu}^n) \equiv \sum_{i=1}^{N} h_i([\mathcal{A}\boldsymbol{\mu}]_i; l_i^n) \leq L(\boldsymbol{\mu}), \; \forall \boldsymbol{\mu} \geq \mathbf{0} \tag{5.46}$$

However, (5.46) is nonseparable in terms of the parameter μ_j, and can not yield a simultaneous update algorithm. For implementations in which the system matrix is not precomputed (e.g. software that uses projector/backprojector subroutines which compute the a_{ij} on the fly), algorithms that update all the parameters simultaneously are preferable. A simultaneous update algorithm can be obtained from the above paraboloidal surrogate by finding yet another separable surrogate function that lies below it for all possible feasible parameter values $\boldsymbol{\mu}$. The additive separability of the surrogate enables decoupling of the optimization problem for each parameter, so each parameter μ_j can be updated independently from the others. To obtain this separable function, Lange [90] applied De Pierro's convexity tricks [91, 92] to transmission tomography to get a separable function that is tangent to the log-likelihood and lies below it everywhere when it is concave. The line integral $[\mathcal{A}\boldsymbol{\mu}]_i$ is rewritten as:

$$[\mathcal{A}\boldsymbol{\mu}]_i = \sum_{j=1}^{M} a_{ij}\mu_j = \sum_{j=1}^{M} \theta_{ij} \left\{ \frac{a_{ij}}{\theta_{ij}} (\mu_j - \mu_j^n) + [\mathcal{A}\boldsymbol{\mu}^n]_i \right\} \tag{5.47}$$

where

$$\sum_{j=1}^{M} \theta_{ij} = 1, \forall i \quad \text{and} \quad \theta_{ij} \geq 0 \tag{5.48}$$

Since the second-order derivative $\ddot{h}_i(l; l_i^n) = -c_i(l_i^n) \leq 0$ for all l, h_i is concave and the surrogate function $H_1(\boldsymbol{\mu}; \boldsymbol{\mu}^n)$ is also concave. One can apply the concavity of h_i:

$$h_i([\mathcal{A}\boldsymbol{\mu}]_i; l_i^n) \geq \sum_{j=1}^{M} \theta_{ij} h_i(\frac{a_{ij}}{\theta_{ij}}(\mu_j - \mu_j^n) + [\mathcal{A}\boldsymbol{\mu}^n]_i; l_i^n) \tag{5.49}$$

The left hand side and right hand side of (5.49) have the same function value and gradient at the current iterate $\boldsymbol{\mu} = \boldsymbol{\mu}^n$. One possible choice for θ_{ij} is given in [97]:

$$\theta_{ij} = \frac{a_{ij}}{\gamma_i}, \qquad \gamma_i = \sum_{k=1}^{M} a_{ik}$$

where γ_i is the forward projection of an image with all ones. Using the inequality (5.49) with these θ_{ij}'s, a separable surrogate function for $H_1(\boldsymbol{\mu}; \boldsymbol{\mu}^n)$ is formed as follows:

$$\begin{aligned} H_2(\boldsymbol{\mu}; \boldsymbol{\mu}^n) &= \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{a_{ij}}{\gamma_i} h_i(\gamma_i(\mu_j - \mu_j^n) + [\mathcal{A}\boldsymbol{\mu}^n]_i; l_i^n) \\ &\leq \sum_{i=1}^{N} h_i([\mathcal{A}\boldsymbol{\mu}]_i; l_i^n) \\ &= H_1(\boldsymbol{\mu}; \boldsymbol{\mu}^n) \leq L(\boldsymbol{\mu}), \, \forall \boldsymbol{\mu} \geq 0 \end{aligned} \tag{5.50}$$

Since the function $H_2(\boldsymbol{\mu}; \boldsymbol{\mu}^n)$ is separable in j and quadratic, it is easily maximized. The exact maximization is reduced to maximization of M 1D parabolas each of which depend on μ_j only. The curvature of the surrogate $H_2(\boldsymbol{\mu}; \boldsymbol{\mu}^n)$ is

$$-\frac{\partial^2}{\partial \mu_j^2} H_2(\boldsymbol{\mu}; \boldsymbol{\mu}^n) \Big|_{\boldsymbol{\mu}=\boldsymbol{\mu}^n} = \sum_{i=1}^{N} a_{ij} \gamma_i c_i(l_i^n) \tag{5.51}$$

The Penalty Part

The above section derived the surrogate function for the log-likelihood $L(\boldsymbol{\mu})$. A separable surrogate for the penalty function $R(\boldsymbol{\mu})$ is needed to form a simultaneous update for the penalized-likelihood objective function $\Phi(\boldsymbol{\mu})$.

Assume that the convex potential function $v(t) = t^2/2$ in (5.38). By using the convexity of v, one can obtain the surrogate for $v(t)$ as follows [92, 90]:

$$\begin{aligned} v(\mu_j - \mu_k) &= v\left[\frac{1}{2}(2\mu_j - \mu_j^n - \mu_k^n) + \frac{1}{2}(-2\mu_k + \mu_j^n + \mu_k^n)\right] \\ &\leq \frac{1}{2}v(2\mu_j - \mu_j^n - \mu_k^n) + \frac{1}{2}v(-2\mu_k + \mu_j^n + \mu_k^n) \\ &\equiv \hat{v}_{jk}(\boldsymbol{\mu}; \boldsymbol{\mu}^n) \end{aligned} \tag{5.52}$$

Using this inequality, a separable surrogate function for the penalty $R(\boldsymbol{\mu})$ is formed as follows:

$$\begin{aligned} R_1(\boldsymbol{\mu}; \boldsymbol{\mu}^n) &= \frac{1}{2}\sum_{j=1}^{M}\sum_{k\in\mathcal{N}_j} w_{jk}\hat{v}_{jk}(\boldsymbol{\mu}; \boldsymbol{\mu}^n) \\ &\geq \frac{1}{2}\sum_{j=1}^{M}\sum_{k\in\mathcal{N}_j} w_{jk} v(\mu_j - \mu_k) \\ &= R(\boldsymbol{\mu}),\ \forall \boldsymbol{\mu} \geq 0 \end{aligned} \tag{5.53}$$

The surrogate R_1 satisfies Conditions 1 and 2:

$$R_1(\boldsymbol{\mu}^n; \boldsymbol{\mu}^n) = R(\boldsymbol{\mu}^n; \boldsymbol{\mu}^n)$$

$$\frac{\partial}{\partial \mu_j} R_1(\boldsymbol{\mu}; \boldsymbol{\mu}^n)\Big|_{\boldsymbol{\mu}=\boldsymbol{\mu}^n} = \frac{\partial}{\partial \mu_j} R(\boldsymbol{\mu})\Big|_{\boldsymbol{\mu}=\boldsymbol{\mu}^n},\ j = 1, \ldots, M$$

When $v(t) = t^2/2$ is used as the potential function, the curvature of the surrogate R_1 at the current iterate $\boldsymbol{\mu}^n$ is twice that of the original penalty R given below:

$$\begin{aligned} \frac{\partial^2}{\partial \mu_j^2} R_1(\boldsymbol{\mu}; \boldsymbol{\mu}^n)\Big|_{\boldsymbol{\mu}=\boldsymbol{\mu}^n} &= 2\frac{\partial^2}{\partial \mu_j^2} R(\boldsymbol{\mu})\Big|_{\boldsymbol{\mu}=\boldsymbol{\mu}^n} \\ &= 2\sum_{k\in\mathcal{N}_j} w_{jk},\ j = 1, \ldots, M \end{aligned}$$

Separable Paraboloidal Surrogate Algorithm

Given the surrogate functions for the log-likelihood and the penalty part, combining H_2 in (5.50) and R_1 in (5.53) yields the overall separable surrogate function for Φ

$$\phi(\boldsymbol{\mu}; \boldsymbol{\mu}^n) = H_2(\boldsymbol{\mu}; \boldsymbol{\mu}^n) - \beta R_1(\boldsymbol{\mu}; \boldsymbol{\mu}^n) \tag{5.54}$$

Since $H_2(\boldsymbol{\mu};\boldsymbol{\mu}^n) \leq L(\boldsymbol{\mu})$ in (5.50), $R_1(\boldsymbol{\mu};\boldsymbol{\mu}^n) \geq R(\boldsymbol{\mu})$ in (5.53) and $\beta \geq 0$, $\phi(\boldsymbol{\mu};\boldsymbol{\mu}^n) \leq L(\boldsymbol{\mu})) - \beta R(\boldsymbol{\mu}) = \Phi(\boldsymbol{\mu})$, $\forall \boldsymbol{\mu} \geq 0$, which satisfies Condition 3. At the current iterate $\boldsymbol{\mu}^n$, $\phi(\boldsymbol{\mu};\boldsymbol{\mu}^n)$ has the following properties:

$$\phi(\boldsymbol{\mu}^n;\boldsymbol{\mu}^n) = \Phi(\boldsymbol{\mu}^n), \quad \nabla\phi(\boldsymbol{\mu}^n;\boldsymbol{\mu}^n) = \nabla\Phi(\boldsymbol{\mu}^n)$$

Hence, Conditions 1 and 2 are satisfied.

In [96], it is shown that increasing the surrogate function $\phi(\boldsymbol{\mu};\boldsymbol{\mu}^n)$ also increases the original objective function $\Phi(\boldsymbol{\mu})$. If a value $\boldsymbol{\mu}^{n+1}$ can be found such that $\phi(\boldsymbol{\mu}^{n+1};\boldsymbol{\mu}^n) \geq \phi(\boldsymbol{\mu}^n;\boldsymbol{\mu}^n)$, then by (5.39) one obtains

$$\Phi(\boldsymbol{\mu}^{n+1}) - \Phi(\boldsymbol{\mu}^n) \geq \phi(\boldsymbol{\mu}^{n+1};\boldsymbol{\mu}^n) - \phi(\boldsymbol{\mu}^n;\boldsymbol{\mu}^n) \geq 0$$

$$\Longrightarrow \Phi(\boldsymbol{\mu}^{n+1}) \geq \Phi(\boldsymbol{\mu}^n)$$

Hence, this algorithm is intrinsically monotonic.

With the surrogate function $\phi(\boldsymbol{\mu};\boldsymbol{\mu}^n)$, the optimization problem at each iteration becomes

$$\begin{aligned}
\boldsymbol{\mu}^{n+1} &= \arg\max_{\boldsymbol{\mu}\geq 0} \phi(\boldsymbol{\mu};\boldsymbol{\mu}^n) \\
&= \arg\max_{\boldsymbol{\mu}\geq 0} [H_2(\boldsymbol{\mu};\boldsymbol{\mu}^n) - \beta R_1(\boldsymbol{\mu};\boldsymbol{\mu}^n)] \\
&= \arg\max_{\boldsymbol{\mu}\geq 0} \sum_{j=1}^{M} \left\{ \sum_{i=1}^{N} \frac{a_{ij}}{\gamma_i} h_i(\gamma_i(\mu_j - \mu_j^n) + [\mathcal{A}\boldsymbol{\mu}^n]_i; l_i^n) + \frac{1}{2} \sum_{k\in\mathcal{N}_j} w_{jk}\hat{v}_{jk}(\boldsymbol{\mu};\boldsymbol{\mu}^n) \right\}
\end{aligned}$$

Since $\phi(\boldsymbol{\mu};\boldsymbol{\mu}^n)$ is separable in terms of μ_j, the update for each parameter depends on the parameter itself and the current estimate $\boldsymbol{\mu}^n$ only. And since $\phi(\boldsymbol{\mu};\boldsymbol{\mu}^n)$ is quadratic in each parameter μ_j, it is easily maximized by zeroing the first partial derivative of $\phi(\boldsymbol{\mu};\boldsymbol{\mu}^n)$ with respect to μ_j given below

$$\frac{\partial}{\partial\mu_j}\phi(\boldsymbol{\mu};\boldsymbol{\mu}^n) = \sum_{i=1}^{N} a_{ij}\dot{f}_i(l_i^n) - d_j^n(\mu_j - \mu_j^n) - \beta \sum_{k\in\mathcal{N}_j} w_{jk}(2\mu_j - \mu_j^n - \mu_k^n) = 0 \quad (5.55)$$

where

$$d_j^n = \sum_{i=1}^{N} a_{ij}\gamma_i c_i(l_i^n)$$

and

$$\dot{f}_i(l) = \left[1 - \frac{y_i}{b_i e^{-l} + \bar{r}_i}\right] b_i e^{-l}$$

Solving this equation finally yields the following update for the separable paraboloidal surrogates algorithm (SPS):

$$\mu_j^{n+1} = \left[\mu_j^n + \frac{\sum_{i=1}^N a_{ij} \dot{f}_i(l_i^n) - \beta \sum_{k \in \mathcal{N}_j} w_{jk}(\mu_j^n - \mu_k^n)}{d_j^n + 2\beta \sum_{k \in \mathcal{N}_j} w_{jk}}\right]_+ \tag{5.56}$$

where $[x]_+$ is x if $x > 0$ and zero otherwise. The non-negativity of the parameter μ_j is enforced because the attenuation coefficients of matter are always nonnegative. The denominator term $2\beta \sum_{k \in \mathcal{N}_j} w_{jk}$ in (5.56) is the second-order derivative of $R_1(\boldsymbol{\mu}; \boldsymbol{\mu}^n)$ with respect to μ_j, and can be precomputed before iteration since it is a constant. This algorithm is highly parallel.

Equation 5.56 has the advantages (1) Easy to compute. (2) Preserving positivity. (3) Parallel update., and is superior to a direct method, such as a positivity constrained gradient algorithm, applied to $\Phi(\boldsymbol{\mu})$. However, SPS converges slowly due to separable surrogates. This can be overcome by the application of ordered subsets. We discuss the ordered subsets technique below .

5.2.3 Ordered Subset (OS) Acceleration

The ordered-subset principle can be applied to any algorithm which requires a sum over projection view indices. Due to the assumed statistical independence of tomographic data, such sums arise in statistical tomographic reconstruction problems including ML or MAP. The sum over all projection views are replaced by sums over subsets of the data and an ordered subsets version of the algorithm is formed. In ordered subsets methods projection data are partitioned into a set of subsets and each subset is processed in sequence. Hudson and Larkin [98] reported that the best method to group the subsets is to choose the projections corresponding to angles with maximal angular distance from previously used angles at each step. This means that the subsets

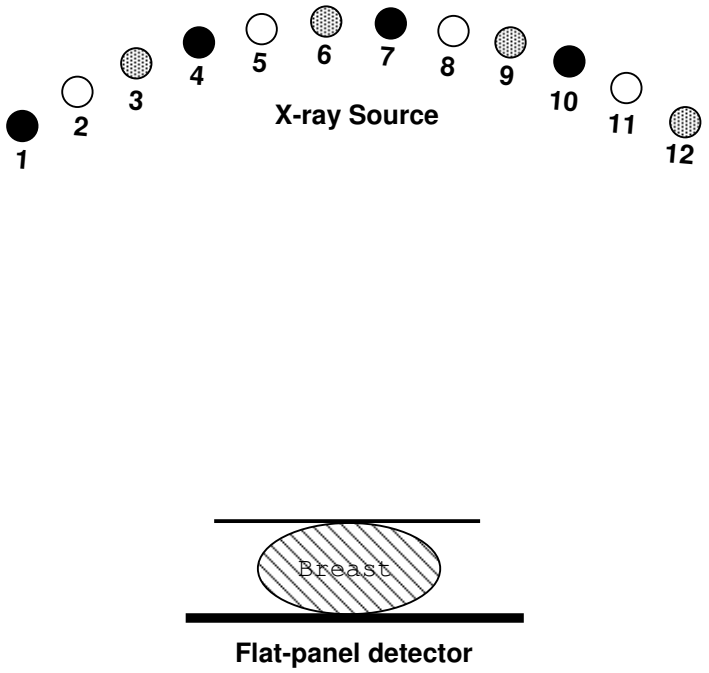


Figure 5.3: the partition of the subsets. Angles 1,4,7,10 are in the first subset. 2,5,8,11 in the second subset. 3,6,9,12 in the third subset.

are chosen to be as orthogonal as possible to previously used subsets. Figure 5.3 shows a total of 12 projections views which are partitioned into 3 subsets. The projection views represented with the same pattern belongs to the same subset. For example, The projection data in views 1, 4, 7, 10 comprise the first subset. Radically different forms of subsets can be used. In PET, for example, a subset can be defined as the sequential accumulation of a specified number of events. But here, we stick with the angular interpretation since it is natural for tomosynthesis.

This technique has been efficiently used in algorithms that update the estimates simultaneously at each iteration. The ordered subsets principle has been applied to transmission ML algorithms. Manglos *et al.* [80] investigated the application of the ordered subsets idea in the transmission EM method for the ML problem. The ordered subset EM (OSEM) algorithm produces visually appealing images fairly quickly and

hence has become very popular. Kamphuis and Beekman [99] applied the ordered subsets principle to Lange and Fessler's convex EM algorithm to accelerate ML transmission reconstruction. Ordered subsets techniques have also been extensively used in emission tomography, but we do not discuss that here.

Erdogan and Fessler [88] applied the ordered subset idea to the SPS algorithm described earlier. The ordered subsets accelerates the initial speed of the original SPS algorithm in terms of increasing the objective function. However, ordered subsets - SPS (OS-SPS) is not guaranteed to be monotonic and does not converge to the true optimum. In fact, this property is true for all direct OS techniques applied in transmission and emission reconstruction algorithms. Very recent modifications, to ensure convergence, are mentioned below. It is desirable for any algorithm to converge to the optimal point of the objective function. Recently, Ahn and Fessler [89] proposed a convergent OS type method called transmission incremental optimization transfer (TRIOT) that accelerate convergence speeds and ensure global convergence. It is parallelizable. The non-negativity constraint is enforced. Their results show that convergent TRIOT is faster in increasing the penalized likelihood objective than OS-SPS algorithm. We note that our own group has proposed a convergent ordered-subset algorithm, COSEM, for ML and MAP emission tomography [100].

In breast tomosynthesis, only 11-15 projection views are used at a limited angular range. Although OS methods are quite flexible, there is much work to be done in designing one that is good for tomosynthetic breast imaging. It is an open question whether whether OS type algorithms can achieve significant speedups without performance degradation for the breast tomosynthesis system.

So far we have discussed reconstruction algorithms (deterministic and statistical) used in transmission tomography and emphasized some applications in breast

tomosynthesis. In our experiment we will use three algorithms: simple BP, SART and convex ML-EM.

Chapter 6

Construction of the 3D Breast Phantom

From Section 2.1, we have some basic understanding about breast anatomy. The breast is inherently a low-contrast organ and highly variable in density and size, and presents tremendous challenges for breast imaging [101]. In order to evaluate reconstruction algorithms for tomosynthesis breast imaging, we need to generate a 3D realistic breast phantom with complex and variable breast structures, that are consistent with the parenchymal pattern in a 2D mammogram. Much work has been done on the simulation of 2D mammogram texture with different statistical methods [102, 103, 104] and on the simulation of 2D high-contrast micro-calcifications [105, 106, 107]. Although they are realistic looking 2D images, such models do not reflect the 3D anatomical properties of the breast, nor can they produce projection images of one breast from different views [108].

Recently, Bakic *et al.* [108] proposed a 3D breast phantom containing two ellipsoidal regions of large scale tissue elements: predominantly adipose tissue and predominantly fibro-glandular tissue. The internal tissue structures of these regions, namely the adipose compartments and the breast ductal network, are approximated by realistically distributed medium phantom elements: spherical shells, blobs and simulated ductal trees. A statistical model of ductal trees generation was proposed. A crescent-shape fibro-glandular region was formed along with a sharp and regular margin in the resulting synthetic mammogram. This breast phantom has a number of nice prop-

erties: however, it is questionable whether simulated tumors embedded within the regular shapes (i.e., spheres and shells) of the phantom can accurately model the task of detecting tumors embedded within breast tissue with irregular texture [24]. The borders between the adipose tissue region and fibro-glandular tissue region in the projection images appeared as a clear, geometrically regular separation degrading the subjective perception of reality.

Bliznakova *et al.* [109] presented a methodology for 3D breast modeling. The simulated breast includes some major tissue features. Their model consists of overall breast shape, adipose tissue, duct system, Cooper's ligaments, abnormalities, pectoralis muscle, and mammographic background. Their model included random generation of variable structures.

The complicated structure of breast tissue make it very difficult to accurately model the 3D breast. Below we describe the mathematical generation of our own medium-sized phantoms in detail. Many of related parameters, which we will use for breast simulation, are approximations justified by Dr. Fisher, who is an experienced mammographer in the Breast Cancer Center, Stony Brook University Medical Center.

6.1 Generation of Breast Volume and Nipple

We use $360 \times 230 \times 400$ voxel array with 0.4 mm cubic voxels as the 3D breast space. Therefore, the physical size of our breast phantoms is $14.4cm \times 9.6cm \times 16cm$ as shown in Figure 6.1. This size is consistent with an actual medium breast size. It is somewhat coarser than the 3D reconstructed resolution obtainable from modern high-resolution digital detectors, but it can be improved. High-resolution breast phantoms take longer to create and reconstruct. For now, we are restricted to low-resolution breast phantoms.

In what follows, we will give mathematical formulas and functions for generating

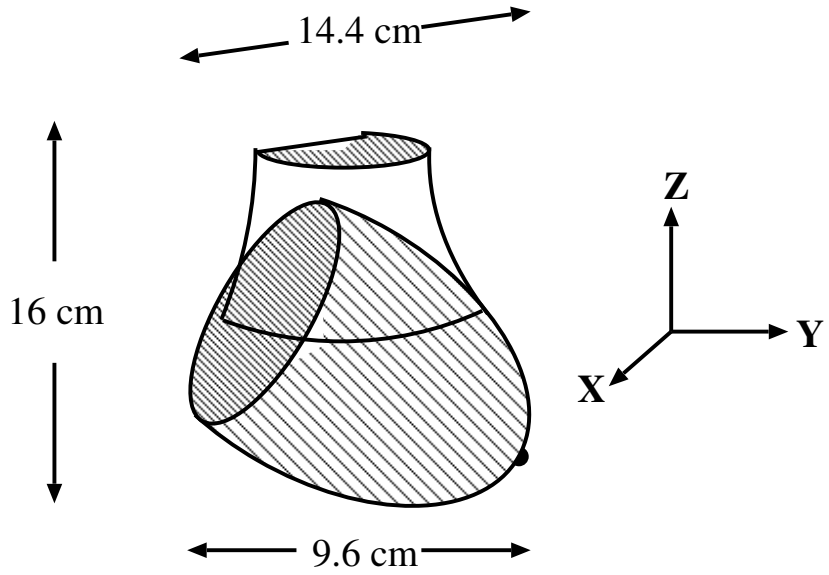


Figure 6.1: External breast shape modeling.

various anatomical structures. It is to be understood that all units will be in voxel edge lengths ($0.4mm$).

The breast shape characterizes the breast volume. The breast volume comprises a rotated half-ellipsoid combined with a half-hyperboloid [109]. Our implementation starts with defining the Cartesian coordinate system as shown in Figure 6.1: the direction from the left breast to the right breast indicates the positive X-axis; the direction from the chest wall to the nipple indicates the positive Y-axis; the direction from the lower breast to the upper breast indicates the positive Z-axis.

The voxels in the rotated half-ellipsoid around the x-axis satisfy the following inequality:

$$\begin{aligned} \frac{(x-x_{ec})^2}{x_{es}^2} &+ \frac{[(y-y_{ec})\cos\theta(z-z_{ec})\sin\theta]^2}{y_{es}^2} \\ &+ \frac{[(y-y_{ec})\sin\theta(z-z_{ec})\cos\theta]^2}{z_{es}^2} \leq 1, \quad y \geq y_{ec} \end{aligned} \tag{6.1}$$

where (x_{ec}, y_{ec}, z_{ec}) is the center of the half-ellipsoid. We choose semi-axes $x_{es} = 170, y_{es} = 200, z_{es} = 170$ to fill the 3D breast space. The rotation angle of the ellipsoid

$\theta = \pi/6$. The voxels in the half-hyperboloid satisfy the following inequality:

$$\frac{(x-x_{hc})^2}{x_{hs}^2} + \frac{(y-y_{hc})^2}{y_{hs}^2} - \frac{(z-z_{hc})^2}{z_{hs}^2} \leq 1, \quad y \geq y_{hc} \quad and \quad z \geq z_{lower} \tag{6.2}$$

where (x_{hc}, y_{hc}, z_{hc}) is the center of the half-hyperboloid. We set semi-axes $x_{hs} = 81, y_{hs} = 81, z_{hs} = 89$. The quantity z_{lower} is carefully chosen to make half-ellipsoid volume and half-hyperboloid volume join each other seamlessly (i.e. a continuous surface). Finally, a half-spherical nipple with semi-axis 9 is attached to the breast shape along the positive y-axis. The breast volume is enveloped between the breast surface and the chest wall. We also incorporate a skin layer of thickness 7.5 mm that sheaths the breast except for a tapered region about the nipple. Figure 6.2 sort of shows this tapered region around the breast margin. All internal structures (except major ducts and Cooper's ligaments) are bounded by the chest wall and this skin layer. Figure 6.2 shows slice images of the breast volume from the CC and ML views.

6.2 Generation of Fibrous Connective Tissue

In Section 2.1.1, we addressed the basic anatomical characteristics of the fibrous tissue. This information can be used as our platform in modeling the fibrous connective tissue. We know that fibrous connective tissue is dispersed within the breast, and supports the ductal, glandular and fat tissues of the breast. In our simulation model, the fibers are represented by a set of linear segments oriented approximately toward the nipple, randomly placed inside the breast volume. We choose the number of fibers to follow a Poisson distribution with mean 1200. This number, as well as many of the parameters we will describe for breast simulation, are approximations justified by Dr. Fisher, who is an experienced mammographer in Breast Cancer Center, Stony Brook University Medical Center. This number 1200 is determined by evaluating the number of fibers per unit volume. The generation of one fiber is described: For a given fiber, the starting point of the fiber (x_{fs}, y_{fs}, z_{fs}) is uniformly chosen within the breast

volume. The connection between this starting point and the center of the nipple gives rise to the initial direction of this fiber indicated by colatitude angle θ_0 and azimuth angle ϕ_0. The final colatitude angle θ_1 and azimuth angle ϕ_1 vary randomly but uniformly within $\pm 5^o$ around the initial direction θ_0 and ϕ_0, respectively. Given the starting point and final direction of the fiber, we can draw the segment with thickness 1 and the length uniformly sampled in the range [40, 55]. Figure 6.2 (a) and (d) show two views of slice images with fibers only.

6.3 Generation of Cooper's Ligaments

As discussed in Section 2.1.1, Cooper's ligaments are the supportive fibrous structure of the breast. They incompletely sheath the peripherally located lobules of the breast, extending outward and attaching to the skin [110]. We simulate Cooper's ligaments with linear segments attaching to the skin. The surface area of the breast shape is around 350 cm^2. We estimate two Cooper's ligaments per cm^2, so about 700 are needed for our breast model. The number of Cooper's ligaments N_c follows a Poisson distribution with mean 700. This number, 700, is justified by Dr. Fisher. We start with the construction of one Cooper's ligament: A reference point $(refX, refY, refZ)$ is picked near the chest wall, and used for all Cooper's ligaments. A starting point of the Cooper's ligament (x_{cs}, y_{cs}, z_{cs}) is uniformly sampled on the breast surface. The connection between (x_{cs}, y_{cs}, z_{cs}) and $(refX, refY, refZ)$ defines the initial growth direction of this Cooper's ligament. With the starting point and the growth direction, we plot the segment with the thickness uniformly sampled in the range [1, 3] and the length sampled in the range [20, 30]. These ranges are consistant with anatomical knowledge. A total of N_c Cooper's ligaments are constructed, and lie beneath the skin. We note that in the report by Bliznakova *et al.* [109], Cooper's ligaments were simulated by ellipsoidal shells placed within the breast volume, but no algorithm was

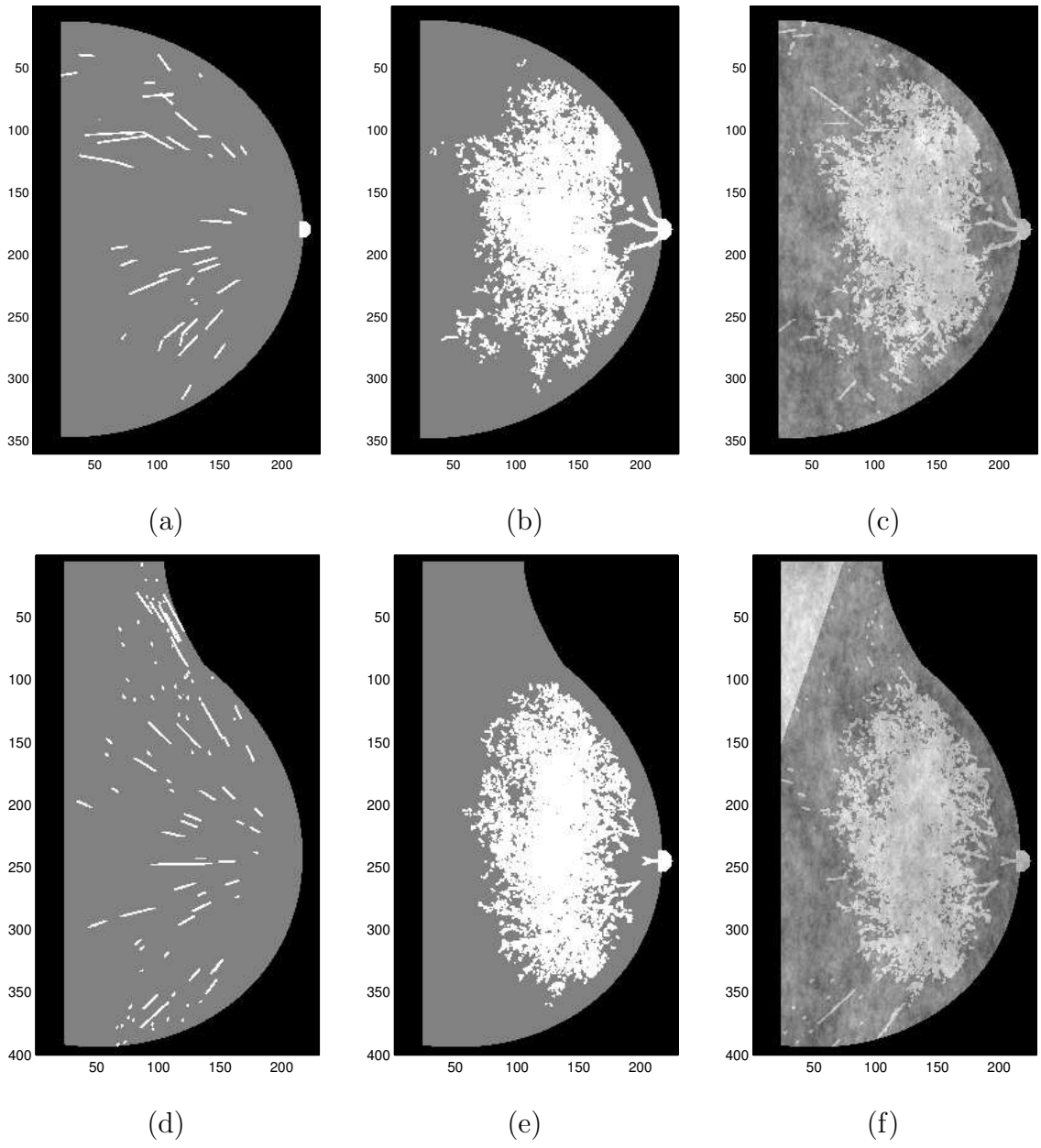


Figure 6.2: (a) and (d) Slice images of one phantom with fibers only at CC view and ML view, respectively. (b) and (e) Slice images of one phantom with ductal tree only at CC view and ML view, respectively. (c) and (f) Slice images of the breast phantom including all anatomically variable and fixed components through a slice containing the nipple at CC view and ML view, respectively. Note that there is no lesion shown in those slices. Also note that (a) (b) (d) (e) are plotted at a common grey scale, and (c) and (f) at a different common grey scale.

given.

6.4 Generation of Ductal Structure

A ductal structure of the breast has a tree-like pattern. Various investigators report that 8 to 20 major ducts originate in the nipple, extending back into the breast in a branching network of smaller and narrower segmental ducts culminating in the terminal ducts and the ending lobules [12]. We choose 11 major ducts to make parenchyma relatively dense in our breast phantom. Taylor *et al.* [111] presented an algorithm that simulated the growth of duct systems based on dichotomous branching and self-similarity theory. They used the topological branching pattern of a perfectly balanced binary tree (each duct branches uniformally into two identical new segments) for tree simulation, resulting in a very regular image texture. Bakic *et al.* [112] proposed a method to simulate the breast ductal network based on branching morphology by ramification matrices andbinary trees. We follow the duct growth algorithm as described by Bliznakova *et al.* [109] albeit with different parameters.

In this method, each of the 11 major ducts of fixed length (6.5) and radius (2.5), but random orientation, originates at the nipple. The major duct branches into a random number of lactiferous ducts, each with a fixed length ($h_0 = 17$) and radius ($r_0 = 2.0$). The number of major duct branches is uniformly distributed in $[2, 9]$. The random orientations of the ducts are given by:

$$\phi_n = \phi_{n-1} + rand * 360 \tag{6.3}$$

$$\theta_n = \frac{\theta_{n-1}}{2} + \theta_{n-1} * rand \tag{6.4}$$

where "rand" is a random number chosen from a uniform distribution on the interval [0.0, 1.0], and (ϕ_n, θ_n) are the child azimuth angle and the child colatitude angle relative to the parents $(\phi_{n-1}, \theta_{n-1})$ in a 3D coordinate system. The length and radius of the ducts decrease as the branch level increases. The parameters of the child branch

are calculated with respect to ones of its parent branch. Each child branch is simulated by a cylinder with radius r_n and length h_n given as:

$$h_n = \frac{h_{n-1}}{2}(1 + rand) \tag{6.5}$$

$$r_n = h_n \frac{r_0}{h_0} \tag{6.6}$$

The process of branching into further lactiferous ducts repeats recursively. The recursion terminates when the ducts reach the skin boundary (7.5 mm beneath the skin) or the maximal branch level. Each duct ends with spherical lobules with radius uniformly sampled in the range [1.0, 2.5]. The structure of a tree is controlled by branching variations—the direction of a major duct, the number and directions of child branch ducts, the length and radius of a child branch duct, and the dimension of a terminal spherical lobule. Even with the same initial parameters, we can generate various patterns of duct trees because of the randomness involved with branch growth. Figure 6.2 (b) and (e) shows two slices of one ductal structure with 11 major ducts.

6.5 Generation of Pectoralis Muscle

We can observe the rounded appearance of the pectoralis muscle from craniocaudal (CC) view mammograms, and the triangular shape from mediolateral oblique (MLO) view mammograms. It is reasonable that the pectoralis muscle is simulated by a cone shape given by

$$(x-x_{mus})^2/x_a^2+(y-y_{ec})^2/y_a^2 <= (z_1-z)^2/(z_2-z_1)^2,\ y \geq y_{ec},\ z \geq z_1\ and\ z \leq z_2 \tag{6.7}$$

where (x_{mus}, y_{ec}, z_1) is the center of the cone, $y_{ec} = 24$ is defines as the same as in Section 6.1:, z_2 is the upper limite in z direction, x_a is the semi-axis in x direction, and y_a is the semi-axis in y direction. We set $x_{mus} = 180$, $z_1 = 250$, $z_2 = 395$ and $x_a = y_a = 50$. The pectoralis muscle itself is highly variable based on body habits and muscle mass and is sensitive to variation in positioning. The presented appearance of

our (fixed) pectoralis is within commonly observed pectoralis morphology. A portion of the pectoralis muscle is seen near the upper chest wall in Figugre 6.2 (f). The 3D cone cuts into Figure 6.1 although not shown in Figure 6.1

6.6 Generation of Masses

Modeling realistic lesions and structured breast tissue is an important factor in considering studies in breast imaging. Previous work in the literature have used spherical or cylindrical lesions in a uniform background. Modeling the lesion with a regular shape may artifically make lesion detection easier. As described in Section 2.1.2, the shapes of masses are divided into round with circumscribed margin, oval with microlobulated margin, lobulated with obscured margin, and irregular with indistinct or spiculated margin. The more irregular a lesion in shape, the more likely the malignancy. The x-ray attenuation and size of the masses is variable. Burgess *et al.* [113] used a 3D random walk algorithm to simulate micro-calcifications. Bliznakova *et al.* [109] used a 3D random walk algorithm to simulate abnormalities (masses and micro-calcifications).

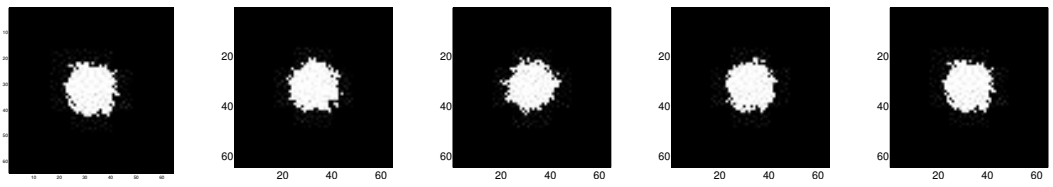


Figure 6.3: The slices through the center of different lesion instances generated with the same parameters.

Here, we describe our own implementation of a 3D random walk algorithm to simulate fixed-location irregular masses with size 7-8 mm. The input parameters include the dimension of the 3D lesion space ($64 \times 64 \times 64$), the center of the lesion $(32, 32, 32)$ within the lesion space, the number of steps in one walk (20), and the

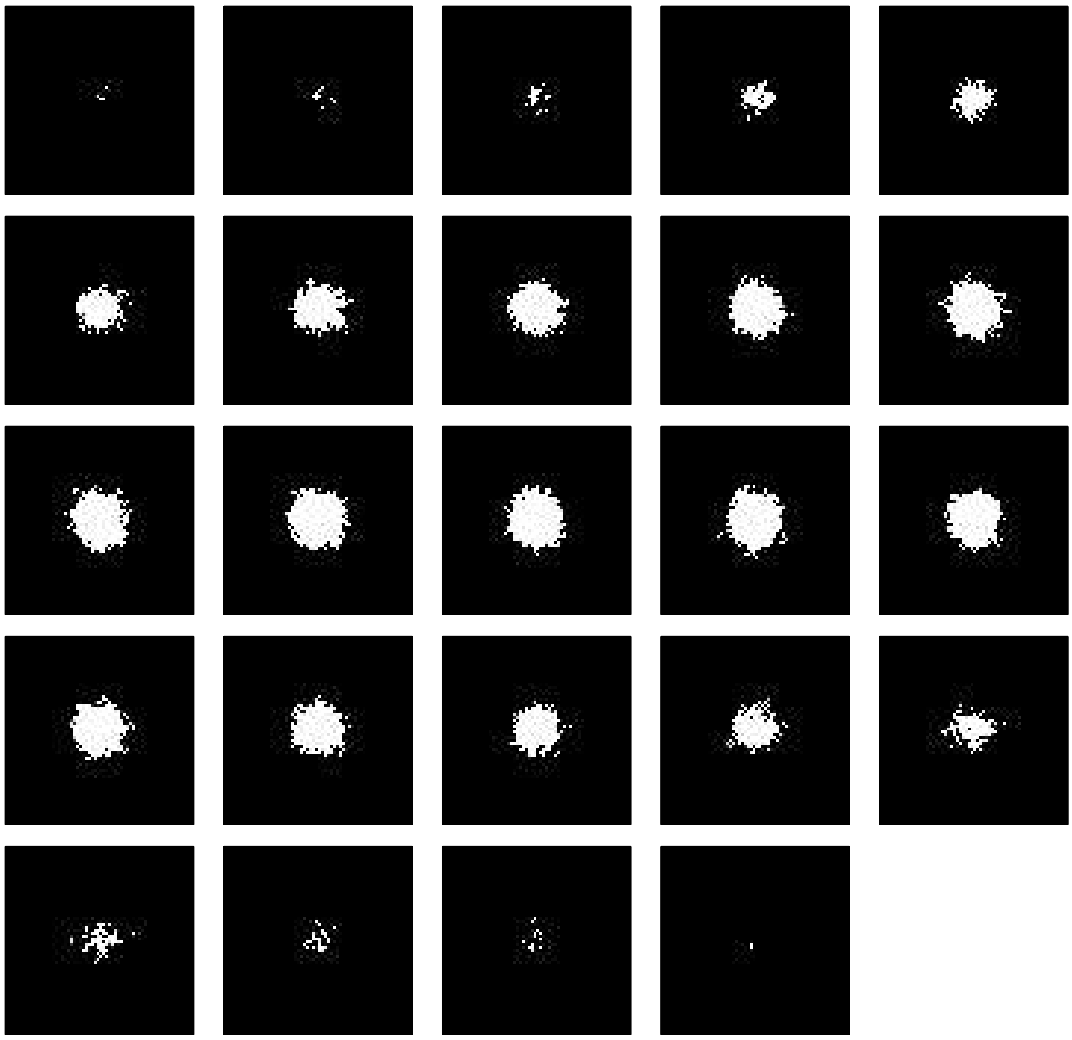


Figure 6.4: Transverse slices through the simulated lesion in Figure 6.3 (a).

number of walks (10000). The step number of one walk mainly controls the size of the lesion. The number of walks controls the irregularity of the lesion. The less the number of walks, the more irregular the lesion. To obtain the irregular shape of a mass, we use "logical" random walks: If one voxel is visited, the value of this voxel is set to 1 no matter whether it was visited before or not; otherwise 0. The center voxel of the mass is the starting point for all walks, and its value is always 1. In the 3D space, the current visited voxel has 26 neighbors (six of them are 1 unit away from the center of this voxel, twelve $\sqrt{2}$ units away from the center and eight $\sqrt{3}$ units away from the center). Without counting the eight furthest neighbors, the next visited voxel can be randomly picked from the nearest 18 neighbors of the current visited voxel with equal probability. The purpose of cutting 8 corner voxels is to obtain a more spherical lesion with irregular boundary. With the same parameters, we can generate different instances of lesions as illustrated in Figure 6.3. Figure 6.4 shows the transverse slices through the simulated lesion in Figure 6.3 (a).

6.7 Generation of Structural Background Noise

Even without other sources of radiation or instrumentation noise, a lesion can be hidden by the anatomical background. We use 3D power-law noise as a model of this background variability. Spectral analysis of mammograms by Burgess [114] demonstrated that mammograms have an approximately isotropic 2D power-law spectrum of the form:

$$F(\rho) = A^2/\rho^\beta \tag{6.8}$$

where ρ is the 2D radial spatial frequency, and the parameter A controls the noise amplitude. and β controls the spectrum slope in the range $[2, 4]$, and its mean value for mammograms was reported to be 3 [114, 47]. We used $\beta = 3$. The value $A = 85.0$ corresponding to the most realistic projection mammogram (as chosen by our mam-

mographer) is the one used. We carefully choose it via visual inspection of projection images by Dr. Fisher. By the central slice theorem [22], the power spectrum of projection at a specific angle is equivalent to the power spectrum along a radial line at the same angle. Hence, if the power spectrum of a 2D projection image is radially symmetric and power-law distributed, then the 3D power spectrum of the object will also be radially symmetric with a characteristic radius determined by the 2D power spectral radius and is also power-law distributed. We can use Eq. 6.8 in 3D object space, where ρ is a 3D radial coordinate in Fourier space.

In general, the 3D power-law background noise is generated by filtering 3D white noise with a power-law kernel. The whole process involves four stages:

1. Generate $720 \times 460 \times 800$ white Gaussian noise $\mathbf{n}$ by a Gaussian random number generator. To avoid the "wrap around" effect of the Fourier transform, we construct a $720 \times 460 \times 800$ noise instead of $360 \times 230 \times 400$ [47].

2. Take the 3D FFT of white noise $\mathbf{n}$.

3. Filter the result with a low-pass filter of the form $A/(1+\rho^{\beta/2})$. We put 1 in the denominator to prevent the Fourier transform of the filter from going to infinity as the spatial frequency approaches zero.

4. Take the inverse 3D FFT and pick the central $360 \times 230 \times 400$ region.

The procedure above can be easily done in Matlab. Figure 6.5 (b) shows a slice image realization of power law noise.

6.8 Construction of Final Breast Phantom

After construction of random instances of connective fibers, ductal structure, Cooper's ligaments and masses, and a fixed pectoralis muscle, the final step is to assign X-ray attenuation coefficients at 30 keV to them as follows: $\bar{\mu}_{fiber} = \bar{\mu}_{ligament} =$

$\bar{\mu}_{duct} = 0.037/\text{mm}$; $\bar{\mu}_{muscle} = 0.038/\text{mm}$ and $\bar{\mu}_{mass} = 0.04045/\text{mm}$. The lesion is placed within the dense ductal structures and its location is fixed. The purpose of this lesion placement is to make lesion detection task difficult as well as realistic. After filling the breast volume with these components described above, the remaining unfilled voxels comprise adipose (fat) tissue. These voxels are assigned $\bar{\mu}_{fat} = 0.025/\text{mm}$. The contrast $C = (\bar{\mu}_{mass} - \bar{\mu}_{duct})/\bar{\mu}_{duct} \approx 9.32\%$ is very low. To complete the phantom, we add an instance of power law noise within the breast volume. Figure 6.5 (a) displays one instance of the phantom before the addition of power law noise, Figure 6.5 (b) one instance of power law noise, and Figures 6.5 (c) the final (slice) phantom in which both are added. Figure 6.2 (c) and (f) also display slice images of one final phantom realization at two views.

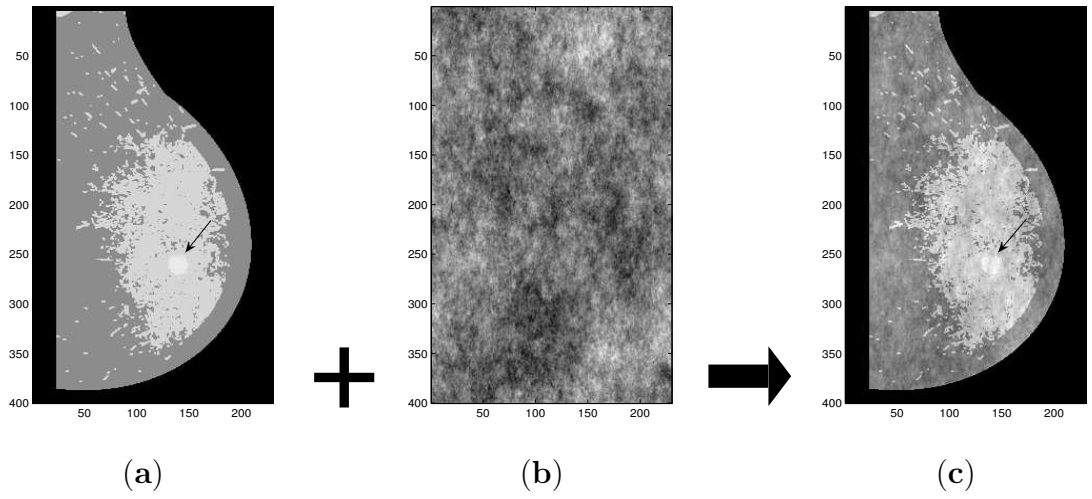


Figure 6.5: (a) A slice image of one phantom with a lesion before power law noise addition. (b) A slice image of one power law noise realization. (c) A slice image of the final phantom with the lesion. Arrow indicates lesion. (a) and (c) are at a common grey scale and (b) is scaled over the full dynamic range.

We note that our simulation model does not include the other anatomical components such as the blood vessels, nerves and lymphatics. Since our task is low-contrast

lesion detection, we do not take into account high-contrast microcalcifications in our model.

We also note that our isocentric simulation geometry does not include breast compression. If we examine the reasons for possibly improved lesion conspicuity in conventional mammography following compression, we find that some of these advantages may be moot for tomosynthesis. In conventional mammography, compression may enhance lesion conspicuity, but in principle, using the least possible compression for tomosynthesis could enhance separation of densities into distinct imaging slices, thus improving detectability. Conventional mammography immobilizes the breast with paddles, but for tomosynthesis there may be techniques that can immobilize the breast without the full compression of mammography. We note that the discomfort of full compression in mammography dissuades some women from undergoing the exam. The smaller force of compression afforded by tomosynthesis results in less pain and discomfort for women. Techniques include the application of a damping paddle along the free (lateral) surfaces of the breast, or improved coupling agents between the paddle and the breast skin. Paddles in conventional mammography create uniformly thick tissue and help reduce the exposure range in film-screen mammography, but given the wide latitude and linear response of digital receptors, this benefit from compression is unnecessary. Compression, along with grids, reduce scatter in conventional mammography. The additional scatter associated with uncompressed (or lightly compressed) and gridless tomosynthetic geometries, or of isocentric geometries, may lead to some image quality degradation, but the advantage of tomosynthesis may more than overcome this degradation. We note that Wu *et al.* [43] have advocated less compression for tomosynthesis than for conventional mammography. Smith [54] suggested that it is adequate to use enough breast compression to pull tissues out of the chest wall and keep motion at a mininum.

We nevertheless note that, especially for conventional mammographic simulation, that several efforts to mathematically model breast compression have been proposed. The compression model used in [108] is based on tissue elasticity properties and a deformation model, which takes advantage of the regular shapes of simulated spherical shells and blobs. It is not clear to the reader how then to go back to a breast shape after deformation. Samani *et al.* [115] presented a biomechanial compression model of the breast deformation using a finite element formulation. Emphasis is given to the modeling of breast tissue deformation which takes place in breast imaging procedures. Their mechanical method using a finite element formulation is very time-consuming.

Chapter 7

The Evaluation of Image Quality for A Lesion Detection Task

In our project presented in Chapter 8, we explored three tomosynthetic reconstruction algorithms—BP which we referred to as method A, SART referred to as method B and EM referred to as method C. How can we determine which method produces the best images? What do we mean by "best"? These are the problems of the objective measurement of image quality.

To answer these questions, we have to specify two things: the task and the observer [116]. One oft-considered clinically important task is the detection of a signal (i.e. a lesion) in a noisy background. In our case, the task is to observe the noisy reconstructed images and decide whether or not a lesion (tumor) is present. The most rigorous test would involve a large set of real tomosynthetic mammographic reconstructions (reconstructed by methods A, B, C) presented to real mammographers. It would be known *a priori* (from a followup biopsy for example) whether each mammogram contained a tumor, and the test could be conducted by having each mammographer read the A, B, C mammogram sets and decide if they detect a tumor. (The mammographer would not be told of the true state - lesion present or absent - of the data set.) Such a test is realistic, but the labor in obtaining such dataset is fantastic, so we resort to *simulated* images.

Our task is a simple binary decision task [117]. In mammography the most interesting task is to detect the abnormality (the evidence of a tumor in the breast).

The determination of whether a lesion or tumor is present in an image is a signal-detection task [117]. The observer is the entity that will make use of the images to perform the task. In the binary decision task, the observer determines which of two states the image belongs to. Usually the observer is the mammographer or a medically untrained graduate student. Alternatively, the observer could be a computer program, such as a commercially available CAD (computer aided detection) package in use in breast centers.

The detection task involves detecting a signal in a background. We consider different classes of signals and backgrounds. We shall discuss several often-investigated detection tasks — signal known exactly/background known exactly (SKE/BKE), signal known statistically/background known statistically (SKS/BKS), SKE/BKS and SKS/BKE. In SKE tasks, the signal is assumed to be a known deterministic entity, i.e., all parameters characterizing the signal location, size, shape, amplitude, etc. are assumed to be known to the observer. The signal function is completely specified and the only uncertainty is whether or not it is present. The SKE detection tasks are convenient to analyze because their simplicity sometimes allows for analytical computation of figures of merit. Such tasks also serve as a good starting point for investigating more complex signal-detection tasks that incorporate signal uncertainty. In BKE tasks, one may assume the underlying object (the attenuation coefficients) or background is known to the observer and fixed. There is thus no object variability or object randomness. In SKE/BKE task, an error in detection could be due to the effects of radiation Poisson noise or other image artifacts. However, such simplified tasks are not clinically realistic. One needs to introduce randomness in both signal and background (i.e. SKS/BKS). The stochastic background variation is also termed "clutter", "object variability" or "background variability". It is essential that the background variability be as realistic as possible. One can achieve background

variation by statistically defined background such as i.i.d noise and complex lumpy background [118]. Those statistically defined backgrounds are easy to treat analytically, however, it may not be sufficiently realistic for some applications. Sometimes a signal is not known exactly but can be described by a function with a number of unknown parameters. Typical parameters characterizing randomness in the signal are amplitude, size, shape, contrast and location. When the signal location is unknown, the more realistic task extends to a detection *plus localization*.

In this study, we actually have a SKS/BKS task with signal location known since a randomly generated lesion is embedded in a complex randomly generated background clutter. The observer is human, in this case, members of several medical imaging labs at Stony Brook University.

7.1 Figures of Merit

7.1.1 The ROC Curve

We shall restrict our discussion to the binary decision case, in which the image belongs to two hypotheses: H_0 (signal absent) and H_1 (signal present). We first consider a very general case with an abstract observer. Let g denote an image that may or may not contain a signal (i.e. a lesion or a tumor). An observer interprets the images to distinguish "positive" (signal present) images from "negative" (signal absent) images and determines a scalar decision variable (test statistic) which is then compared to a threshold to classify the image into H_0 or H_1. Since the test statistic is related [119] to the data through a discriminant function $T(\mathbf{g}) = t$, and it is a random variable, sometimes denoted as $t(\mathbf{g})$. Roughly speaking $t(\mathbf{g})$ is a measure of confidence that the signal is present. The probability density function on $t(\mathbf{g})$ depends on the hypothesis. Let $p(t|H_0)$ and $p(t|H_1)$ denote probability density functions conditioned on the underlying hypothesis H_0 and H_1, respectively, and $P(t|H_0)$ and

$P(t|H_1)$ denote cumulative functions conditioned on the underlying hypothesis H_0 and H_1, respectively. The overlap of the density functions $p(t|H_0)$ and $p(t|H_1)$ is shown in Figure 7.1. Image quality is determined by the degree of separation/overlap of these two density function [117].

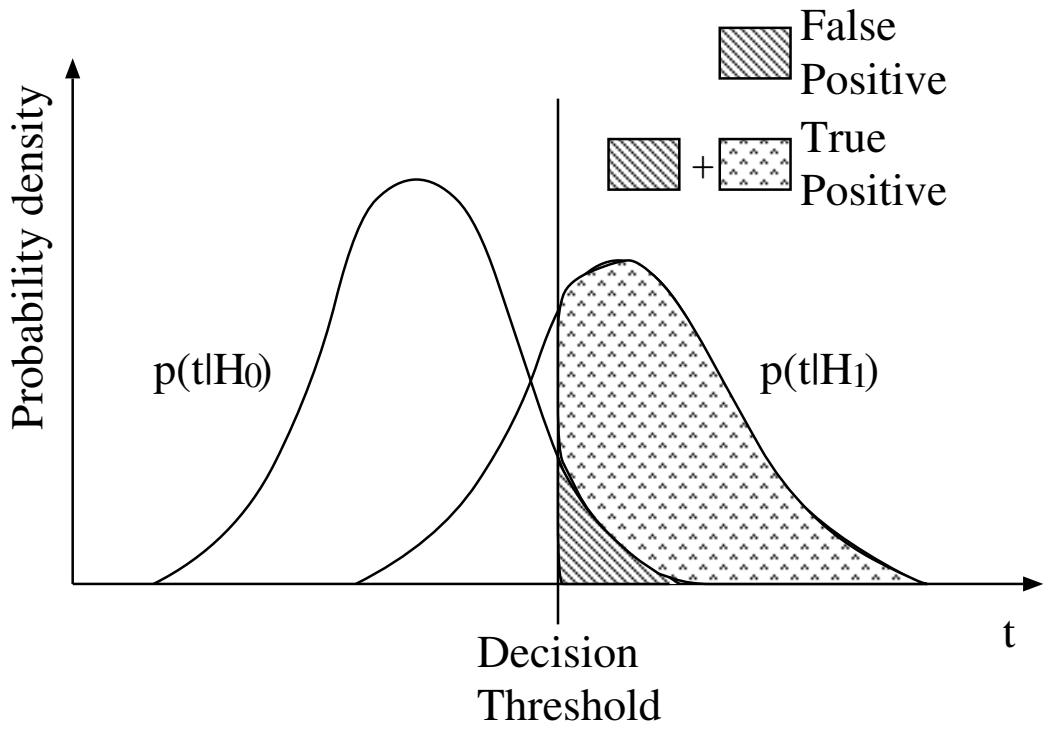


Figure 7.1: The decision making diagram. As indicated in the legend, the shaded areas under the curves represent the false positive and true positive decision probabilities for the threshold shown.

Once the test statistic t is determined for a given image, the observer compares it to a threshold τ and decides signal present if $t > \tau$, otherwise signal absent. There are four possible decision outcomes for a given threshold τ [119]:

- True positive (TP): H_1 is true; observer decides that H_1 is true. In other words, an actually-positive image will be diagnosed correctly as "positive".
- False positive (FP): H_0 is true; observer decides that H_1 is true. In other words,

an actually-negative image will be diagnosed falsely as "positive".

- True negative (TN): H_0 is true; observer decides that H_0 is true, which means that an actually-negative image will be diagnosed correctly as "negative".

- False negative (FN): H_1 is true: observer decides that H_0 is true, which means that an actually-positive image will be diagnosed falsely as "negative".

We notice that the observer correctly determines the underlying hypothesis in two of decision outcomes (TP and TN), and also makes two types of errors (FP and FN). The probability of a true positive decision is called as *true positive fraction* (TPF), which is also known as the *sensitivity* of the imaging system [120]. The probability of deciding signal present when the signal is not is called as *false positive fraction* (FPF). The system's *specificity*, also called *true negative fraction* (TNF) is the probability of deciding the absence of a signal when it is absent. The *false negative fraction* (FNF) is the probability of deciding the absence of a signal when it is present. Let N be the number of decisions made by an observer, and N_{TP} be the number of true positive decisions made by the observer. Similar notations apply to N_{FP}, N_{TN} and N_{FN}. Since a decision made by the observer always falls in one of four decision outcomes, $N = N_{TP} + N_{FP} + N_{TN} + N_{FN}$. Four probability functions are mathematically defined as [119]:

$$TPF = \langle \frac{N_{TP}}{N_{TP} + N_{FN}} \rangle = \lim_{N \to \infty} \frac{\text{Number of true positive decisions}}{\text{Number of actually positive cases}} \quad (7.1)$$

$$TNF = \langle \frac{N_{TN}}{N_{TN} + N_{FP}} \rangle = \lim_{N \to \infty} \frac{\text{Number of true negative decisions}}{\text{Number of actually negative cases}} \quad (7.2)$$

$$FPF = \langle \frac{N_{FP}}{N_{TN} + N_{FP}} \rangle = 1 - TNF \quad (7.3)$$

$$FNF = \langle \frac{N_{FN}}{N_{TP} + N_{FN}} \rangle = 1 - TPF \quad (7.4)$$

These four fractions range from zero to one. As seen in Figure 7.1, these four fractions

are the areas under the corresponding PDFs on t for a given threshold τ:

$$TPF(\tau) = \Pr(t \geq \tau | H_1) = \int_{\tau}^{\infty} p(t|H_1)dt = 1 - P(\tau|H_1) \quad (7.5)$$

$$FPF(\tau) = \Pr(t \geq \tau | H_0) = \int_{\tau}^{\infty} p(t|H_0)dt = 1 - P(\tau|H_0) \quad (7.6)$$

$$FNF(\tau) = 1 - TPF = \int_{-\infty}^{\tau} p(t|H_1)dt = P(\tau|H_1) \quad (7.7)$$

$$TNF(\tau) = 1 - FPF = \int_{-\infty}^{\tau} p(t|H_0)dt = P(\tau|H_0) \quad (7.8)$$

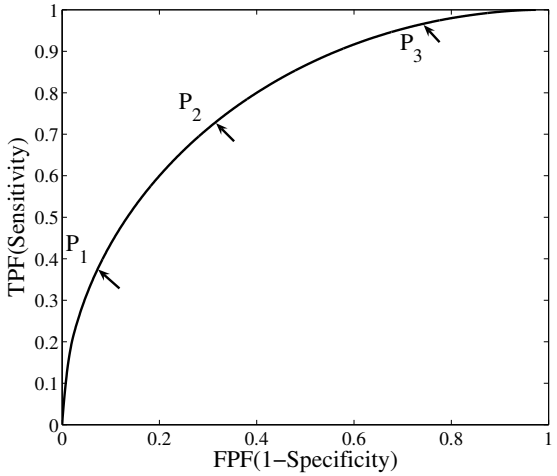


Figure 7.2: An ROC curve.

As we can see, TPF and FPF depend on the threshold value τ. The decision made by the observer is based upon the threshold value. For example, if an observer wants to be cautious (conservative), he/she might designate a suspicious region as "cancer" even though they think it unlikely to actually be "cancer". If they were less cautious, the same image might lead to "no cancer" decision. Changing the threshold value changes both an imaging system's sensitivity and the system's specificity, and results in a different set of decision fractions. When varying the decision threshold τ, we obtain a curve of TPF versus FPF which is known as a *receiver operating characteristic* (ROC) curve [120]. Figure 7.2 displays a typical ROC curve. Three operating points are shown in Figure 7.2. Point P_1 corresponds to use of a strict

threshold so that the observer make very few true positive decisions and very few false positive ones. Point P_2 indicates the use of a moderate threshold level. Point P_3 corresponds to use of a conservative threshold where an observer makes many true positive response as well as false positive decisions.

7.1.2 The Area Under the ROC Curve (AUC)

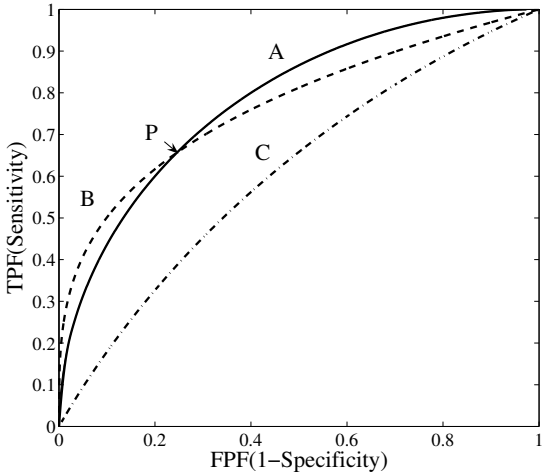


Figure 7.3: A graphy showing three ROC curves.

The the area under the ROC curve (AUC) is a commonly used figure of merit to measure detection performance, which is independent of any decision threshold since the AUC is the average TPF over all FPF [119] and integrated over all operating points. Below we will show how to compute AUC without computing the ROC curve itself. Figure 7.3 shows three ROC curves obtained from three different tests (e.g. these could be reconstruction algorithms A, B and C in our study). Curve A and curve B with the same area cross over each other at point P. In this case, one would conclude that tests A and B provide overall equal detectability of the lesion, but they are not identical. In order to determine the optimal test system from A and B, one needs to know the clinical requirements of the test system [120]. For example, if the

system needs a high sensitivity, then the user would pick test B. Thus the ROC curve provides more information than the AUC, but the AUC is an acceptable figure of merit for initial studies where optimal thresholds are possibly unknown.

The AUC can be mathematically related to the distributions of the test statistic under the signal-present and signal-absent hypotheses in two different ways as shown below:

$$\begin{aligned} AUC &= \int_0^1 TPF d(FPF) \\ &= \int_\infty^{-\infty} \left[1 - P(t|H_1)\right] \left(-p(t|H_0)dt\right) \\ &= 1 - \int_{-\infty}^{\infty} P(t|H_1)p(t|H_0)dt \end{aligned} \tag{7.9}$$

$$\begin{aligned} AUC &= \int_0^1 (1 - FPF) d(TPF) \\ &= \int_\infty^{-\infty} P(t|H_0) \left(-p(t|H_1)dt\right) \\ &= \int_{-\infty}^{\infty} P(t|H_0)p(t|H_1)dt \end{aligned} \tag{7.10}$$

7.1.3 Signal-To-Noise Ratio (SNR) and Detectability d_A

The separability of two classes in the binary detection problem is determined by the degree of overlap of the density functions of the test statistic [119] (see the overlap of two lumps of the density functions in Figure 7.1). It is this overlap that controls the detectability of the signal. The AUC is one metric of this overlap. Another available FOM for detection tasks is the observer signal-to-noise ratio (SNR) related to t:

$$\text{SNR} = \frac{\langle t|H_1\rangle - \langle t|H_0\rangle}{\sqrt{\frac{1}{2}var(t|H_1) + \frac{1}{2}var(t|H_0)}} \tag{7.11}$$

where $\langle t|H_1\rangle$ and $\langle t|H_0\rangle$ are means of t under hypotheses H_1 and H_0, $var(t|H_1)$ and $var(t|H_0)$ are variances of t under hypotheses H_1 and H_0. The SNR is one way to gauge the separability of two lumps of the probability density functions (refer to Figure 7.1). From Equation 7.11, we can see that SNR is proportional to the difference between their means and is inversely proportional to their variance, that is, the bigger the

difference between their means and the smaller their variance is , the more separable the two lumps are. If they are highly separable, then the likelihood of a correct decision increases. Notice that SNR only requires the information of the first and second moments of $t|H_1$ and $t|H_0$, thus is less representative but easier to quantify than the AUC.

When $t|H_0$ and $t|H_1$ are normally distributed, the SNR is found to have a monotonic relationship with AUC [119]

$$\text{AUC} = \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{\text{SNR}}{2}\right) \tag{7.12}$$

where the error function erf(z) is defined by

$$\text{erf} = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-x^2)dx \tag{7.13}$$

There are potential advantages in using the AUC as an FOM rather than the SNR when the test statistic does not follow a Gaussian distribution under one or two hypotheses. This is because the SNR only considers the first- and second- order statistics of the test statistic, whereas the AUC incorporates higher-order statistics as well [121].

Another figure of merit for signal detection is the detectability d_A. This FOM is defined simply as a nonlinear transformation of the AUC. It is found written in three different ways in the literature [119]

$$d_A = 2\text{erf}^{-1}(2\text{AUC} - 1) = 2\text{erfc}^{-1}\left(2(1 - \text{AUC})\right) = \sqrt{2}\Phi^{-1}(\text{AUC}) \tag{7.14}$$

where erf^{-1} and erfc^{-1} are the inverses of the error function defined in Equation 7.13 and the complementary error function, respectively. The function Φ^{-1} is the inverse of the standard cumulative normal distribution. The detectability provides a relationship between the AUC and the observer SNR. If the observer response has a Gaussian distribution under each hypothesis, then $d_A = \text{SNR}$ [119].

7.2 Two-Alternative-Forced-Choice (2AFC) Methods

The ROC curve provides a complete description of observer performance and can be applied to any two-state classification task (i.e. signal present or absent). In situations where the full ROC is not required, a two-alternative forced-choice (2AFC) technique provides an alternative approach to evaluate the performance of human observers on a detection task by measuring the AUC.

In a 2AFC test, an observer is presented with two images simultaneously, and one is drawn from the group with signal present and another is drawn from the group with signal absent. The observer is required to choose which image contains the signal present. The probability of a correct decision, referred to P_C, ranges from 0 to 1 and measures the performance of the observer-system combination on the detection task. The observer does not need to adopt any threshold; Instead he/she just compare images with each other.

We shall show the expected P_C in this 2AFC test equals the expected AUC ($P_C = AUC$) in a conventional ROC experiment [119] below. From Equations 7.5 and 7.6, one can rewrite Equation 7.9

$$\begin{aligned} \text{AUC} &= \int_{\infty}^{-\infty} (-p(t|H_0)dt)\, TPF(t) \\ &= \int_{-\infty}^{\infty} p(t|H_0)dt \int_{t}^{\infty} p(\tau|H_1)d\tau \end{aligned} \tag{7.15}$$

By using the step function to rewrite Equation 7.15, one can obtain another form for AUC:

$$\text{AUC} = \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} p(t|H_0)p(\tau|H_1)step(\tau - t)d\tau \tag{7.16}$$

The AUC derived above requires knowing the densities $p(t|H_0)$ and $p(t|H_1)$. Since the test statistic t is a random variable through its dependence on the M-dimensional data $\mathbf{g}$, the density functions $p(t|H_0)$ and $p(t|H_1)$ can be expressed as:

$$p(t|H_0) = \int_{\infty} p(t|\mathbf{g})p(\mathbf{g}|H_0) = \int_{\infty} \delta(t - T(\mathbf{g}))p(\mathbf{g}|H_0)d^M g \tag{7.17}$$

$$p(t|H_1) = \int_\infty p(t|\mathbf{g})p(\mathbf{g}|H_1) = \int_\infty \delta(t - T(\mathbf{g}))p(\mathbf{g}|H_1)d^M g \tag{7.18}$$

Substituting Equations 7.17 and 7.18 into Equation 7.16, one has

$$\begin{aligned} \text{AUC} &= \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\tau \int_\infty \delta(t - T(\mathbf{g}))p(\mathbf{g}|H_0)d^M g \\ &= \int_\infty \delta(\tau - T(\mathbf{g}^+))p(\mathbf{g}^+|H_1)step(\tau - t)d^M g^+ \\ &= \int_\infty d^M g \int_\infty p(\mathbf{g}|H_0)p(\mathbf{g}^+|H_1)step(T(\mathbf{g}^+) - T(\mathbf{g}))d^M g^+ \end{aligned} \tag{7.19}$$

In the 2AFC method, the data $\mathbf{g}$ and $\mathbf{g}^+$ are drawn from $p(\mathbf{g}|H_0)$ and $p(\mathbf{g}^+|H_1)$, respectively. The observer's task is to choose image from H_1. To make a decision, the observer computes two test statistics $T(\mathbf{g})$ and $T(\mathbf{g}^+)$ and the data vector that results in higher value is then decided to be signal present. This decision is correct if $T(\mathbf{g}^+) > T(\mathbf{g})$. Thus, the probability of a correct decision is

$$P_c = p(T(\mathbf{g}^+) > T(\mathbf{g})) = \int_\infty d^M g \int_\infty p(\mathbf{g}|H_0)p(\mathbf{g}^+|H_1)step(T(\mathbf{g}^+) - T(\mathbf{g}))d^M g^+ \tag{7.20}$$

which is equal to AUC in Equation 7.19.

There is another way to proof that the AUC and P_C are mathematically identical. Using Equation 7.8, one can rewrite Equation 7.10 as

$$\begin{aligned} \text{AUC} &= \int_{-\infty}^{\infty} P(t|H_0)p(t|H_1)dt \\ &= \int_{-\infty}^{\infty} p(t|H_1)dt \int_{-\infty}^{t} p(\tau|H_0)d\tau \end{aligned} \tag{7.21}$$

In a 2AFC trial, the observer is presented with two images and asked to identify which image contains the signal. This decision process is modeled by assuming that the observer forms a response to each image and then choose the image with the largest response. Let t^+ be a random variable drawn from $p(t|H_1)$ and let t^- be an independent random variable drawn from $p(t|H_0)$. Then the probability of a correct decision is defined as

$$P_C = p(t^+ > t^-) \tag{7.22}$$

We can see that P_C is simply the probability that a random sample drawn from $p(t|H_1)$ will be greater than an independent sample drawn from $p(t|H_0)$. The equivalence between Equations 7.21 and 7.22 is not obvious. We shall derive that. The probability that $t^+ > t^-$ for a fixed value of $t^+ = t$ is given by

$$p(t^+ > t^-|t^+ = t) = \int_{-\infty}^{t} p(\tau|H_0)d\tau \tag{7.23}$$

Using the equation above, the P_C can be derived as [122]

$$\begin{aligned} P_C &= p(t^+ > t^-) \\ &= \int_{-\infty}^{\infty} p(t^+ = t)p(t^+ > t^-|t^+ = t)dt \\ &= \int_{-\infty}^{\infty} p(t|H_1) \int_{-\infty}^{t} p(\tau|H_0)d\tau \\ &= \text{AUC} \end{aligned} \tag{7.24}$$

In a 2AFC experiment, P_C is commonly estimated with the percentage of correction decisions. The estimate of AUC is typically obtained by computing the area under an estimated ROC curve, while P_C is estimated from 2AFC experiments with no reference to an ROC curve. Therefore, if only the total area under the ROC curve is of interest, the AUC can be measured directly by the 2AFC test. We have conducted such tests for nuclear medicine imaging systems [123], and we shall adopt the same methodology here to compare three tomosynthetic algorithms in a low-contrast lesion detection task. The 2AFC testing procedure is described in detail in Section 8.4.1.

7.3 Alternative Figures of Merit

For breast tomosynthesis, besides AUC and SNR, there are other FOMs used by researchers—contrast-to-noise ratio (CNR) and artifact spread function (ASF). The CNR measures the detectability of a feature in a reconstruction plane as well as AUC, while the ASF is the FOM for evaluation between in-focus plane and off-focus plane.

7.3.1 Contrast-To-Noise Ratio (CNR)

The contrast-to-noise ratio is a quantity used for comparing low-contrast detectability, also called the signal difference-to-noise ratio [62] and is defined by

$$\mathrm{CNR} = \frac{\bar{\mu}_{feature} - \bar{\mu}_{BG}}{\sigma_{BG}} \tag{7.25}$$

where $\bar{\mu}_{feature}$ and $\bar{\mu}_{BG}$ are the average pixel intensities measured in the mass and and its background area of the tomosynthetic reconstruction slice, respectively, and σ_{BG} is the standard deviation of the pixel intensity in the background area. The average pixel intensity of a mass $\bar{\mu}_{feature}$ can be calculated in a small region of interest (i.e. 35×35-pixel region) which is enclosed within the relatively uniform central area of the mass [63]. To estimate the background noise $\bar{\mu}_{BG}$ and σ_{BG}, the same size region of interest (i.e. 35×35-pixel region) is chosen from the same slice as the the mass under consideration, which is far from the mass and the boundaries of the imaged volume.

7.3.2 Artifact Spread Function (ASF)

In tomosynthetic reconstruction, the Z resolution is much lower than the in-plane resolution which is determined from the detector resolution due to the limited angular range in image acquisition. Artifacts exists in the reconstruction as a result of the limited Z resolution. Ghosting artifacts from a real feature located in an image plane are usually observed in other image planes, with an appearance similar to that of the real feature. The artifacts are spread to other adjacent planes, and stronger in image planes that are closer to the plane in which the real feature is located. The artifact spread function (ASF)is used to evaluate image blur in the Z direction (perpendicular to the XY detector plane) of reconstructions and the artifact effects of features in the adjacent off-focus planes, which is defined as the ratio of contrasts [62]

$$\mathrm{ASF}(z) = \frac{\bar{\mu}_{AF}(z) - \bar{\mu}_{BG}(z)}{\bar{\mu}_{feature}(z_0) - \bar{\mu}_{BG}(z_0)} \tag{7.26}$$

where z_0 is the slice location of the in-focus plane of the feature, z is the location of an off-focus plane,$\bar{\mu}_{feature}(z_0)$ and $\bar{\mu}_{BG}(z_0)$ are the average pixel intensities of the feature and the image background in the in-focus plane, respectively, $\bar{\mu}_{AF}(z)$ and $\bar{\mu}_{BG}(z)$ are the average pixel intensities of the artifact and the image background in the off-focus plane, respectively. The values of $\bar{\mu}_{feature}(z_0)$ and $\bar{\mu}_{BG}(z_0)$ can be measured in two small regions with the same size located in the center of the mass and in the background of the in-focus plane respectively; $\bar{\mu}_{AF}(z)$ and $\bar{\mu}_{BG}(z)$ can be measured in two small regions at the same off-focus slice plane.

A similar ASF was used to evaluate the artifacts along Z direction in a breast tomosynthesis study [63], and defined as the ratio of the CNRs between the off-focus plane and the in-focus (feature) plane:

$$\mathrm{ASF}(z) = \frac{\mathrm{CNR}(z)}{\mathrm{CNR}(z_0)} \tag{7.27}$$

The CNR is used to define the ASF since the noise levels are expected to be different on the reconstructed images at different depths and the detectability of a feature is dependent more on the CNR than on the contrast.

Finally, we mention another commonly used figure of merit, DQE (detective quantum efficiency). This FOM does not apply to signal detection, but is instead used to measure the physical performance of a detector. Therefore, we do not consider it in this work.

Chapter 8

Experimental Results

Given the 3D breast phantom described in Chapter 6, we can investigate various problems. Ours is to evaluate the performance of three reconstruction methods in a low-contrast lesion detection task. Before doing that, we first need to collect 2D projection data and then reconstruct them to a 3D breast estimate. The following is our project.

8.1 Data Acquisition in Breast Tomosynthesis

For our study, we generate 100 signal-present phantom realizations and 100 signal-absent phantom realizations. Each realization has variability in power-law noise, connective fibers, ductal trees and lesions (for signal-present phantoms). To reduce computation time, we use 50 ductal patterns distributed amongst the 200 breast realizations. A specific realization of a ductal pattern may or may not obscure a (fixed location) lesion. We select these 50 ductal patterns so that they all obscure the lesion (when a lesion was present) in the breast phantom. We recommend variable patterns of ductal trees in all experimental phantoms. Ductal tree realizations are constructed by the stochastic branching process. For each realization, we need to decide whether this realization is good or not (i.e. is sufficiently obscuring) for our purposes in terms of the fixed lesion location in the dense area of the breast. The reject rate is around 50%. The density of the area surrounding the lesion is a dominant factor for the

lesion detection task. The more dense the lesion surrounding area, the more difficult the lesion detection. One instance of Cooper's ligaments is used in all 200 phantoms since it does not affect lesion detectability that much.

In this experiment, an isocentric half cone-beam geometry is used to collect 11 low-dose projection images over a limited angular range from -25^o to $+25^o$ at 5^o equal increments, as shown in Figure 8.1. Figure 8.1 defines our $X-Y-Z$ coordinate system. Figure 8.1 (c) shows the front view (drawn to scale) of the isocentric geometry in X-Z plane. The x-ray tube and the detector (666x400 pixels, 0.3 mm pixel size) are rotated in an arc about a common axis while the breast phantom is stationary. The distance between the X-ray source and the center of rotation is 110 cm, and the distance between the center of rotation and the detector 20 cm. The center of rotation lies within the breast.

This isocentric geometry can be used if some method of stabilizing breast motion, such as a cup arrangement could be used. An easier arrangement to construct is to lightly compress the breast with a conventional paddle and use a conventional mammographic hardware while the tube is allowed to rotate. The isocentric geometry is easier to simulate though we are able to simulate any geometrical arrangement. The detection experiments results are not expected to change much with these variations in geometry.

As discussed in Section 4.2, the number of photons y_i recorded on the ith detector pixel is an independent Poisson random variable distributed as Equation 4.17:

$$y_i \sim Poisson\left(b_i\, e^{-[\mathcal{A}\boldsymbol{\mu}]_i}\right),\ i = 1, \ldots, N$$

where the blank scan b_i is the photon number recorded on the ith detector element with the breast phantom absent. Though b_i should vary with i, we set $b_i = 30125$ for all rays. This number was chosen so that the Poisson noise per one of our detector bins is consistent with the dose requirement for acquisition (i.e. the total dose over

all views equals the dose for a conventional mammography study.) This number was obtained from data provided by Jun Zhou also working on mammography at Stony Brook University. Let $\boldsymbol{\mu}$ denote the vector whose element μ_j is the unknown linear attenuation coefficient (in mm^{-1}) at the jth voxel ($j = 1, \ldots, M$), where M is the number of voxels (M=360x230x400). Figure 8.2 shows noisy ML and CC projection images of two breast phantom realizations. Scatter events and electronic noise effects for a flat-panel detector were not simulated in this study.

8.2 Reconstruction Algorithms

In our simulation study, we investigate the efficacy of three different tomographic reconstruction methods, convex ML-EM, SART and Backprojection (BP), in the context of an especially difficult mammographic detection task. The task is the detection of a low-contrast mass embedded in very dense fibro-glandular tissue — a clinically useful task for which tomosynthesis may be well suited. The choice of algorithms is motivated as follows: BP is simple to implement and fast. SART is an example of a deterministic linear inversion technique, and the convex ML-EM is a statistical reconstruction technique that models one component of the maximum likelihood, the Poisson noise. We chose convex ML-EM instead of MAP and OS type of algorithms because MAP and OS were developed concurrently with our own project. In our study, we wanted to use one basic example of BP, one basic linear algebraic algorithm, and one basic statistical algorithm.

As discussed in Section 2.4.2, in tomosynthetic digital mammography, we are interested in reconstructed slices with high resolution in the XY plane and with lower resolution along the Z direction, since projection data is over a limited angle. This can be done by acquiring projection data from 0.4 mm cubic phantom voxels and reconstructing the breast estimates onto rectanguloid voxels of longer length along

the Z direction. Hence, each layer of the reconstructed breast is a "thick" slice. The length in the Z direction of the rectanguloid voxel is variable according to the required resolution. Here we use 3.2 mm, equivalent to a rectanguloid of relative dimension $1 \times 1 \times 8$. The reconstructed breast size is thus $360 \times 230 \times 50$ (This for reconstruction, $M = 360 \times 230 \times 50$).

Deterministic methods require computing the logarithm of the measured projection data to obtain the line integral estimate prior to estimating the tomosynthetic slice images. Statistical methods use the measured projection data rather than their logarithm, based on the Poisson measurement statistics.

8.2.1 Simple Backprojection Method

The tomosynthetic BP method is attractive because of its simplicity and speed of implementation. As discussed in Section 5.1.1 for certain acquisition geometries, it is essentially equivalent to a "shift-and-add" algorithm [40]. Backprojection is performed by smearing a set of 2D projection images uniformly back into the 3D breast space along the incident x-ray paths. In this operation, the estimated attenuation coefficient of the voxel is proportional to the weighted summation of all the values of detector pixels whose rays intersect this voxel. The attenuation coefficient updates of voxels are parallel as given by Equation 5.5: The BP algorithm leads to low contrast blurred slice images due to artifacts caused by out-of-plane features. Figure 8.3 (b) shows one BP reconstruction with lesion present.

8.2.2 SART Algorithm

SART is one type of iterative algebraic reconstruction method operated on a set of measured line integrals. The breast estimate is generated by iteratively performing projections of intermediate estimates and correction backprojections. SART is well suited for the tomosynthesis cone-beam geometry with a few projection views. The

goal of the SART method is to seek solutions $\boldsymbol{\mu}$ to the following linear system of equations using an iterative technique: The attenuation coefficient updates of voxels are given in Equation 5.13:

$$\mu_j^{n,k+1} = \mu_j^{n,k} + \lambda \frac{\sum_{g_i \in \mathcal{G}_k} \left(a_{ij} \frac{g_i - \sum_{j=1}^{M} a_{ij} \mu_j^{n,k}}{\sum_{j=1}^{M} a_{ij}} \right)}{\sum_{g_i \in \mathcal{G}_k} a_{ij}}$$

One iteration requires traversing all projection views ($\mathcal{K} = 11$). The denominator term $\sum_{g_i \in \mathcal{G}_k} a_{ij}$ in the above equation is a backprojection operation. The term $\sum_{j=1}^{M} a_{ij}$ is the forward projection of the unit volume. Both terms are one-time computations, and can be precomputed and stored. Therefore, the update needs only one forward projection $\sum_{j=1}^{M} a_{ij} \mu_j^{n,k}$ in the numerator and one backprojection — the summation over g_i (the numerator term). We note that the attenuation coefficient update of voxels is parallel for a given subiteration. Typically, the iterations are terminated before convergence to control noise growth. Figure 8.3 (c) shows one SART reconstruction with lesion present.

8.2.3 ML-EM Algorithm

As discussed in Section 5.1, deterministic methods do not take into account the statistical nature of noisy data. Accurate statistical modeling may lead to improved detection performance in high-noise reconstructions. In this study we employ the convex ML-EM algorithm proposed by Lange and Fessler [90] rather than the basic EM algorithm in[84] because the basic EM converges very slowly and requires a large number of exponential computations as discussed in Section 5.2.1. This ML-EM method has been applied to breast tomosynthesis reconstruction by Wu *et al* [43].

Using the convex ML-EM method, the update of the voxel attenuation coefficient at each iteration is given by Equation 5.31:

$$\mu_j^{n+1} = \mu_j^{n} \frac{\sum_{i=1}^{N} a_{ij} \left[b\, e^{-[\mathcal{A}\boldsymbol{\mu}^n]_i} \left(1 + [\mathcal{A}\boldsymbol{\mu}^n]_i\right) - y_i \right]}{\sum_{i=1}^{N} a_{ij} [\mathcal{A}\boldsymbol{\mu}^n]_i b\, e^{-[\mathcal{A}\boldsymbol{\mu}^n]_i}}$$

The numerator of the second term is proportional to the attenuation estimate μ_j^n, therefore the update is multiplicative. To avoid a zero estimate, one needs to choose non-zero initial condition. The summation over rays in the numerator is a backprojection of the difference between the current projection estimate $b\,e^{-[\mathcal{A}\boldsymbol{\mu}^n]_i}$ and the observed projection y_i. This difference term is backprojected into the voxel, then weighted by a backprojection factor in the denominator.

The optimal number of iterations for two iterative algorithms—SART and ML-EM is determined by series of 2AFC tests with small sample sizes (60 pairs). See Section 7.2 for more theoretical detail. The involved test procedure will be presented in next section. We perform reconstructions at varying numbers of iterations and conducted 2AFC tests with 4 human observers to choose the iteration number that maximized the percentage of correctness. For EM, this is 20 iterations; for SART, it is 5. Note that for iterative SART and ML-EM, we take the initial estimate equal to a constant object. Once we have chosen the stopping criteria, we use the reconstruction results from these three algorithms to conduct 2AFC tests in Section 8.4.

8.3 Implementation of Reconstruction

In tomosynthetic mammography, we are most interested in reconstructed object with high resolution in the XY 2D plane and lower resolution in the Z direction because there are a limited number of projection views. Less information about feature detail in the Z direction is needed. For implementation, we propose using the anisotropic (rectanguloid) voxel whose length in the Z dimension is much longer than in the X and Y dimensions and whose XY planes are square. Hence, each layer of the reconstructed object is a "thick" slice. The length in the Z direction of the rectanguloid voxel is variable according to the required resolution. We use $0.4mm \times 0.4mm \times 3.2mm$

voxels.

Typically, The reconstruction process involves projector/backprojector subroutines. Recall that the projector is basically the line integral of attenuation coefficient distributions as given in Equation 5.3:

$$g_i = [\mathcal{A}\boldsymbol{\mu}]_i = \sum_{j=1}^{M} a_{ij}\mu_j, \quad i = 1, \ldots, N, \quad j = 1, \ldots, M.$$

where μ_j is the attenuation coefficient at voxel j and a_{ij} is the intersecting chord length of the ray i through the voxel j. The key is how to calculate the chord length a_{ij}. Now, we describe our implementation based on the modified Siddon's ray-tracing method [124] by Han [125]. Figure 8.4 illustrates the isocentric cone-beam geometry at 0^o, which is used in our calculation. The 3D coordinate system has its origin at the center of rotation (COR), though the X and Y axes are displayed in a displaced manner for clarity. The source (x_0, y_0, z_0) and the detector are rotated around the the COR synchronously. The source and detector rotate synchronously about the fixed breast. The rotation angle θ is shown in Figure 8.4. The angular movement is perpendicular to the plane of the figure.

The user need to define the following input parameters: the breast size ($ObNo_x$, $ObNo_y$, $ObNo_x$) in the X, Y, Z directions in terms of the number of voxels, the physical voxel sizes ($Vsize_x$, $Vsize_y$, $Vsize_z$) with units if mm, the detector size ($BinNo_x$, $BinNo_y$) in the X, Y directions in terms of the number of pixels, the physical pixel sizes ($Bsize_x, Bsize_y$) with units if mm, the number of angles, the starting angle (i.e. -25^o), the ending angle (i.e. $+25^o$), the distance $SourceR$ between the source (x_0, y_0, z_0) and the COR with unit mm, the distance $DetectorR$ between the COR and the detector with unit mm, the physical position of the breast (x_{b0}, y_{b0}, z_{b0}) with units mm and the physical position of the detector (x_{d0}, y_{d0}). The COR is the origin $(0, 0, 0)$ of this coordinate system in Figure 8.4. Note that the physical voxel size in the Z direction $Vsize_z$ can be different from $Vsize_x$ or $Vsize_y$, thus resulting

in rectanguloid voxels. The physical positions (x_{b0}, y_{b0}, z_{b0}) and (x_{d0}, y_{d0}) are relative to the COR.

Before performing any calculation, we need to make sure that the detector width with given pixel size and given number of pixels can capture all information of the 3D breast with given voxel size and given number of voxels, that is, no rays passing through the breast can fall off the detector. This can be done by using a "placer" visualization software developed by Fang Xu, a research assistant in Computer Science Department, SUNY at Stony Brook. Figure 8.6 illustrates the display of the placer. The placer software allows one to specify a bounding volume for the breast, an acquistion geometry and detector size. It then checks that all projection views at all angles do not "fall off" the detector.

The source position (x_0, y_0, z_0) varies with the acquisition view θ, given by:

$$\begin{aligned} x_0 &= SourceR * \sin\theta \\ z_0 &= SourceR * \cos\theta \end{aligned} \tag{8.1}$$

Note that $y_0 = 0$ does not vary with the angle θ. For a given angle θ, the position (x_1, y_1, z_1) of the detector bin indexed by (bin_x, bin_y) is calculated as:

$$\begin{aligned} x_1 &= (x_{d0} + (bin_x + 0.5) * Bsize_x) * \cos\theta - DetectorR * \sin\theta \\ y_1 &= y_{d0} + (bin_y + 0.5) * Bsize_y \\ z_1 &= -(x_{d0} + (bin_x + 0.5) * Bsize_x) * \sin\theta - DetectorR * \cos\theta \end{aligned} \tag{8.2}$$

where the bin index $bin_x = 1, \cdots, BinNo_x$ and $bin_y = 1, \cdots, BinNo_y$. The projection ray i is given by the line connecting the x-ray source (x_0, y_0, z_0) denoted by P_0 and the detector bin (x_1, y_1, z_1) denoted by P_1.

Then we can follow the modified Siddon's ray-tracing method to compute the indices of the voxels that the given ray i passes through and to calculate the chord lengths intersected voxels along the ray. The indices of intersecting voxels and their

corresponding chord lengths are stored in user-defined structure variable *raytbl*, which is defined as follows:

```
struct  raytbl3D{
            short   *x; /* voxel index in x direction */
            short   *y; /* voxel index in y direction */
            short   *z; /* voxel index in z direction */
            float   *length; /* corresponding chord length */
            short   size; /* the number of voxels along a given ray i */
        }   raytbl
```

For simplicity, a 2D geometry in Figure 8.5 is used to illustrate the intersections of chords with voxel boundaries inside the breast volume. In this ray-tracing algorithm, a ray from the source P_0 to a specific detector pixel P_1 is represented as:

$$x(\lambda) = x_0 + \frac{\lambda}{L}(x_1 - x_0) \tag{8.3}$$

$$y(\lambda) = y_0 + \frac{\lambda}{L}(y_1 - y_0) \tag{8.4}$$

$$x(\lambda) = z_0 + \frac{\lambda}{L}(z_1 - z_0) \tag{8.5}$$

where λ is the distance from the source P_0 or the starting point of the ray, and L is given by:

$$L = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2} \tag{8.6}$$

The parameter λ is computed at each intersection of the ray with each set of parallel planes and and then merged in the tracing process. Then the intersecting length can be computed as

$$l(m) = \lambda(m) - \lambda(m-1) \tag{8.7}$$

where $\lambda(m)$ is the distance from P_0 to the current intersecting voxel, and $\lambda(m-1)$ is the distance from P_0 to the previous neighboring intersecting voxel. The chord length

of the current intersecting voxel $l(m)$ is stored as the mth element in the array *length* of structure variable *raytbl*. The indices of the current intersecting voxel is computed as

$$v_x(m) = \begin{cases} v_x(m-1)+1, & \text{if} \quad x_0 < x_1 \\ v_x(m-1)-1, & \text{if} \quad x_0 > x_1 \end{cases} \tag{8.8}$$

$$v_y(m) = \begin{cases} v_y(m-1)+1, & \text{if} \quad y_0 < y_1 \\ v_y(m-1)-1, & \text{if} \quad y_0 > y_1 \end{cases} \tag{8.9}$$

$$v_z(m) = \begin{cases} v_z(m-1)+1, & \text{if} \quad z_0 < z_1 \\ v_z(m-1)-1, & \text{if} \quad z_0 > z_1 \end{cases} \tag{8.10}$$

The indices of the current intersecting voxel, $v_x(m)$, $v_y(m)$, $v_z(m)$, are stored as the mth element in arrays x, y, z of variable *raytbl*, respectively. In Figure 8.5, a given ray i passes through seven voxels, $r_n, n = 1, \cdots, 7$ are corresponding chord lengths. Once a_{ij} in Eq. 8.3 are known, we can use these values to perform forward projection g_i on the fly. A similar approach is applied to the backprojector subroutine. The tomosynthetic reconstruction routines are iterative operations of projector/backprojector. Recall that all iterative reconstruction algorithms require forward projections and backprojections at each iteration from Chapter 5. The operations both $\mathcal{A}\boldsymbol{\mu}$ for forward projection and $\mathcal{A}^T\mathbf{y}$ for backprojection involve the system matrix $\mathcal{A}$, that is, need its elements a_{ij}'s, which can be obtained by running this described ray-trace routine above.

We have described in some detail the projection/backprojection software for an isocentric geometry. The software is quite easily adapted to a variety of geometries including a stationary detector geometry favored in some tomosynthesis systems. (We have experimented with such geometries but results are not presented here.)

When the 3D breast object or its reconstruction needs to be viewed, we use a "slicer" tool developed by Fang Xu in the Stony Brook Computer Science Department

to display the dense 3D object one slice at a time. The slicer is able to simultaneously display slices from two objects (i.e. the breast and its reconstruction), both at the same orientation. The slicer has capabilities to control contrast, zoom, rotation and other parameters. Figure 8.7 illustrates one example of the slicer display.

8.4 2AFC Test in Low-Contrast Lesion Detection Task

8.4.1 Testing Procedure

Recall that our goal is to estimate AUC via the 2AFC method are described in Chapter 7. And recall from earlier section that 100 pairs (signal present and signal absent) of 3D breast estimates are reconstructed through three methods—BP, SART and EM, respectively. The 2D slice images of interest (the center of the lesion) is extracted from each of 100 pairs of reconstructions as our 100-image test set for each algorithm. Thus the total number of images is 600. For each reconstruction method, we used a 25-image training set to familiarize human observers with the image set's characteristics. Each observer underwent a total of 3 sessions, each for each algorithm. The sequence of three algorithms for each observer is variable. For example, observer 1 may choose algorithm A as his/her first test session, algorithm B as the second, algorithm C as the third, while observer 2 may choose algorithm B as his/her first test session, algorithm C as the second, algorithm A as the third. The observer took a significant rest after each session. This strategy, along with the choice of the number of training and test images, avoided observer fatigue. The sets corresponding to different reconstruction methods were presented in different pairs and different orders to different observers so as to avoid learning effects. There were a total of $N_{obs} = 5$ observers, all graduate students and postdocs from our lab (excluding the author) or other labs.

To implement the 2AFC, we used the software modified by a member of our

group, Jorge Oldan, and the source of the software is from the "Test2AFC" MATLAB scripts provided by the University of Arizona's Center for Gamma-Ray Imaging (an NIH resource). Our modification of this software package can be found at

www.mil.sunysb.edu/publications.html

For each reconstruction method, the 2AFC tests were conducted as follows. As seen in each row of Figure 8.8, each person was presented with three black and white 360×230 images on a standard CRT monitor. The central panel always displayed the signal in order to help the human observer localize the signal locations in the test images. The left and right panels display two noisy reconstructions on the sides. One panel contained a noisy reconstruction with the lesion absent, the other a noisy reconstruction with the lesion present. The panel with the lesion was randomly determined via a randomization function that consulted the system clock to be either on the right or the left side. The size of all three pictures as measured together on the monitor was 20.32 cm wide by 10.0 cm high. Each individual picture is rectangular, measuring 6.43 cm wide by 10.0 cm high.

The observer had to choose with a right or left mouse-button click which side they thought the picture (with lesion) was on. The observer has as much time as they needed to observe the images, and was free to move their head position and viewing distance as they saw fit. The average viewing distance was about 38 cm and we noted that actual viewing distances did not dramatically differ from this average. The correctness of their answer was immediately reported (i.e. "correct" or "incorrect"), as was their total percent correct (P_C) when they finished the experiment. Room lights were turned off and the blinds are used to block the light from outside through the window.

8.4.2 Results Discussion

After the test results from 2AFC expriments were obtained for all five observers, we calculated the averaged P_C over five human observers for each algorithm, $\bar{P}_C$, given as follows:

$$\bar{P}_C = \frac{1}{N_{obs}} \sum_{r=1}^{N_{obs}} P_{C,r} \tag{8.11}$$

where $N_{obs} = 5$ is the number of observers. Let r index the observer, so that $P_{C,r}$ is the percent correct (as measured by 2AFC) of the rth observer (Recall that $P_C = AUC$ as was shown in Section 7.2). We also calculated error bars for P_C, σ_{P_C}, using the standard error of the mean-squared percent correct:

$$\sigma_{P_C} = \left[\frac{1}{N_{obs}(N_{obs}-1)} \sum_{r=1}^{N_{obs}} (P_{C,r} - \bar{P}_C)^2 \right]^{\frac{1}{2}} \tag{8.12}$$

where σ_{P_C} is the standard error associated with inter-observer variability. With 100

Algorithm	$\bar{P}_c$	σ_{P_c}
ML-EM	0.888	0.0114
SART	0.828	0.0102
BP	0.661	0.013

Table 8.1: Human observer results for EM, SART and BP

2AFC trials per algorithm, the error in P_C is determined far more by inter-observer variability than intro-observer variability. That is, with 100 Bernoulli trials in the 2AFC test, the sample error for one observer is small compared to the inter-observer variability.

The numbers of iterations for both ML-EM and SART methods were experimentally determined in this study. In order to obtain the optimal number of iterations, we run additional 2AFC tests with small sample sizes (60 pairs). The reconstructions are

performed at varying numbers of iterations and 2AFC experiments are conducted with 4 observers. We choose the iteration number that maximizes $\bar{P}_C$ as our optimal number of iterations. For ML-EM, this optimal number is 20 iteration for this particular task; for SART, it is 5. In fact, the SART method even with one iteration can provide tomosynthesized images with much detail structures. The SART method converged faster than the ML-EM method. From this point of view, the SART method is more effective than the ML-EM method in reconstruction time. A possible reason for the faster SART convergence is this: the linear attenuation coefficients are updated at each projection view in the SART method; while for the ML-EM method, the update is performed after all projection views are processed.

Table 8.1 summarizes performance results of three algorithms. It is seen from the preliminary results that for this difficult low-contrast lesion detection task, the iterative methods ML-EM and SART perform better than simple BP in terms of percent correctness. As seen in Figure 8.8, The BP method resulted in low-contrast features and vey smooth reconstruction images. In the BP images, it is very difficult to observe low-contrast masses. The SART and ML-EM methods provided the better conspicuity of low-contrast tissues by enhancing contrast and edge sharpness, and separated the overlapping tissue structure very well. However, both SART and ML-EM iterate from the constant initial condition and amplify the noise. One can see that both SART and ML-EM reconstructions were much noiser than BP reconstruction. For the experiments we conducted, the statistical method (ML-EM) has a slight advantage in performance relative to the deterministic one.

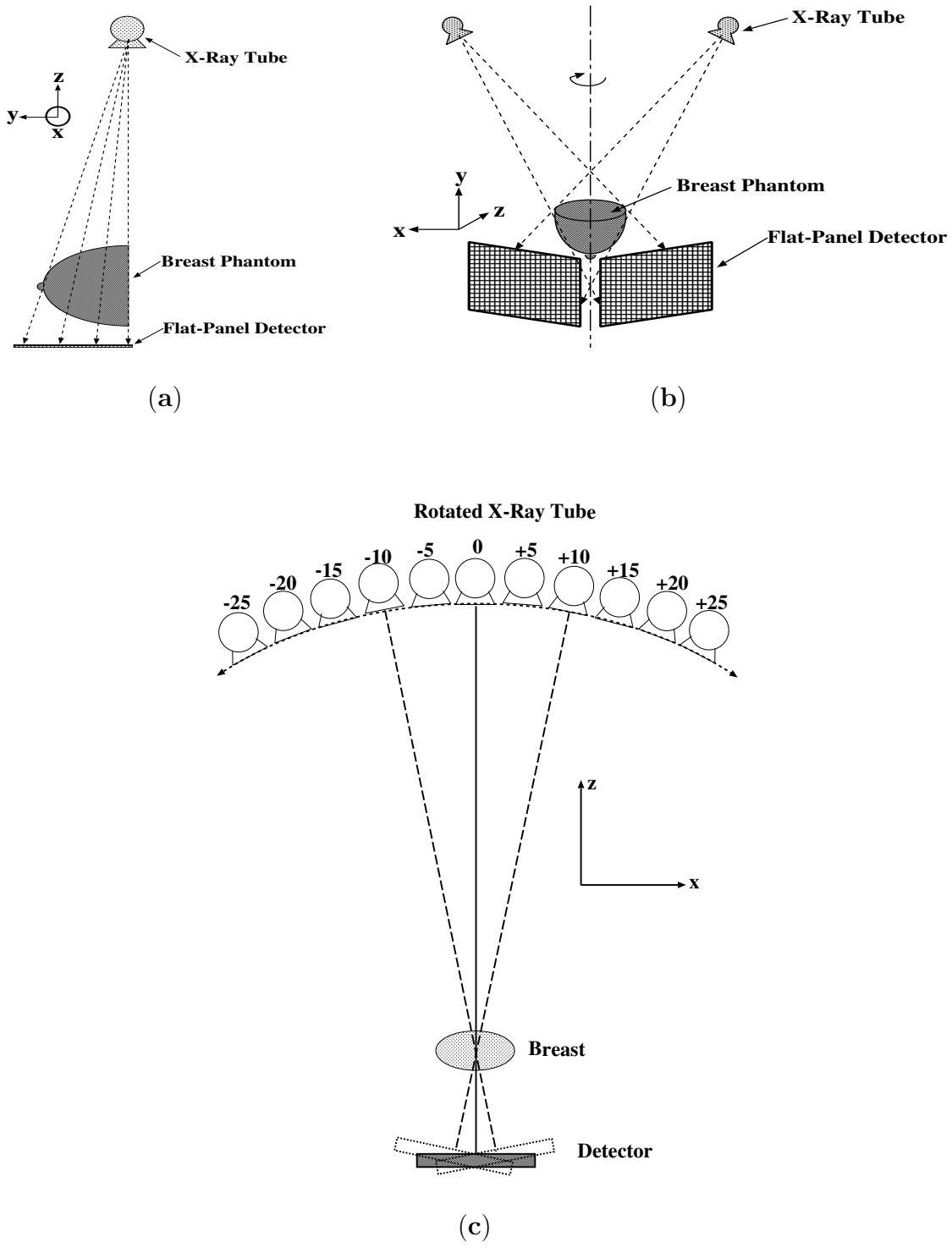


Figure 8.1: (a) side view of tomosynthesis breast imaging geometry. (b) 3D perspective view showing acquisition at only 2 of the 11 angles. (c) front view of the geometry. The detector is rotated around the central axis synchronously with the detector. The drawing in (c) is scaled to the dimensions used in the simulation.

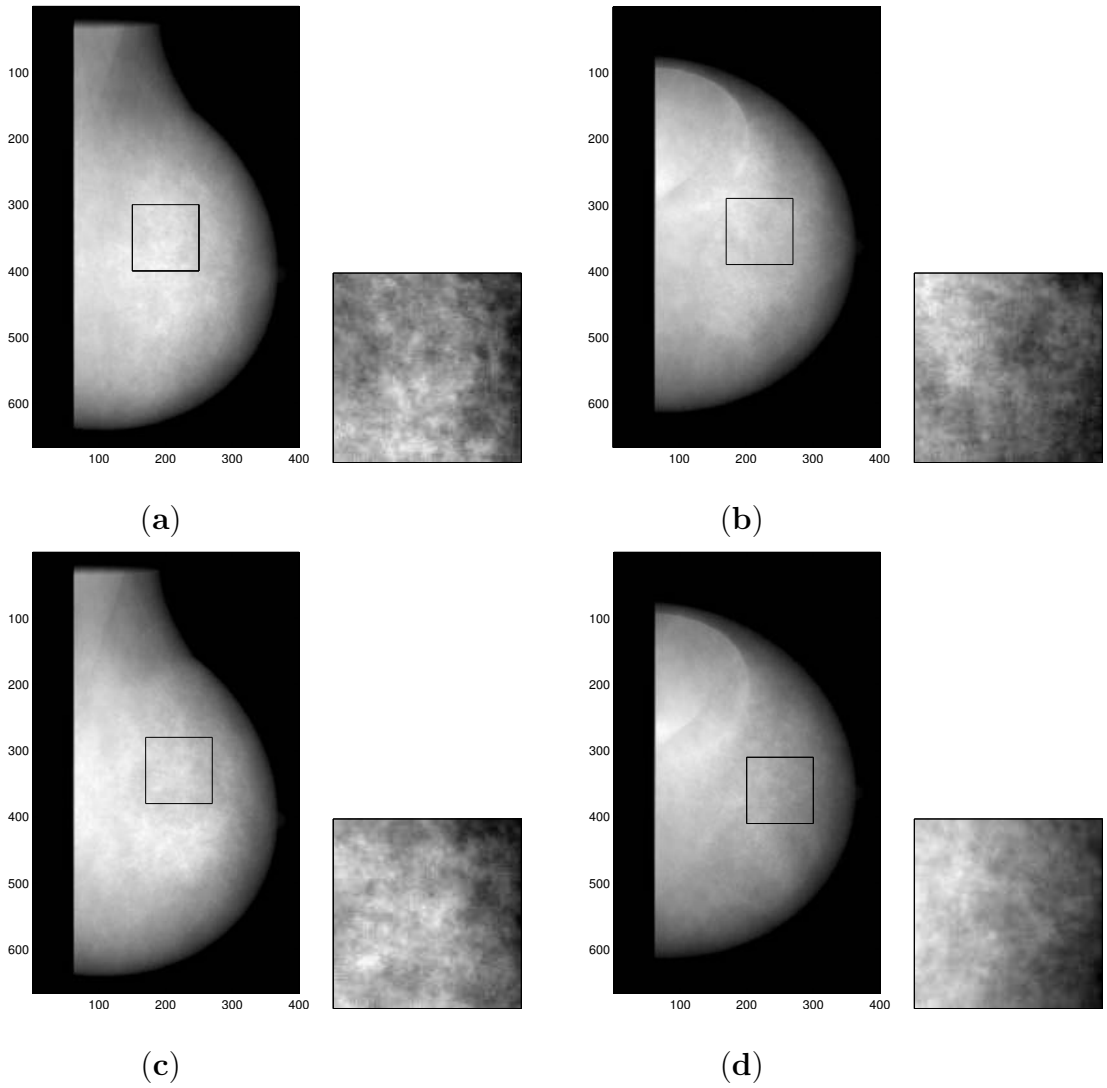


Figure 8.2: (a) and (c) noisy projection images of two breast phantoms at ML views corresponding to 90^o (not used in our simulations). (b) and (d) noisy projection images of two breast phantoms at 25^o, which are used for reconstructions. The small figures display magnified regions of their corresponding left-hand images, delineated by the black squares in the left-hand, and show complex structures in the projection images.

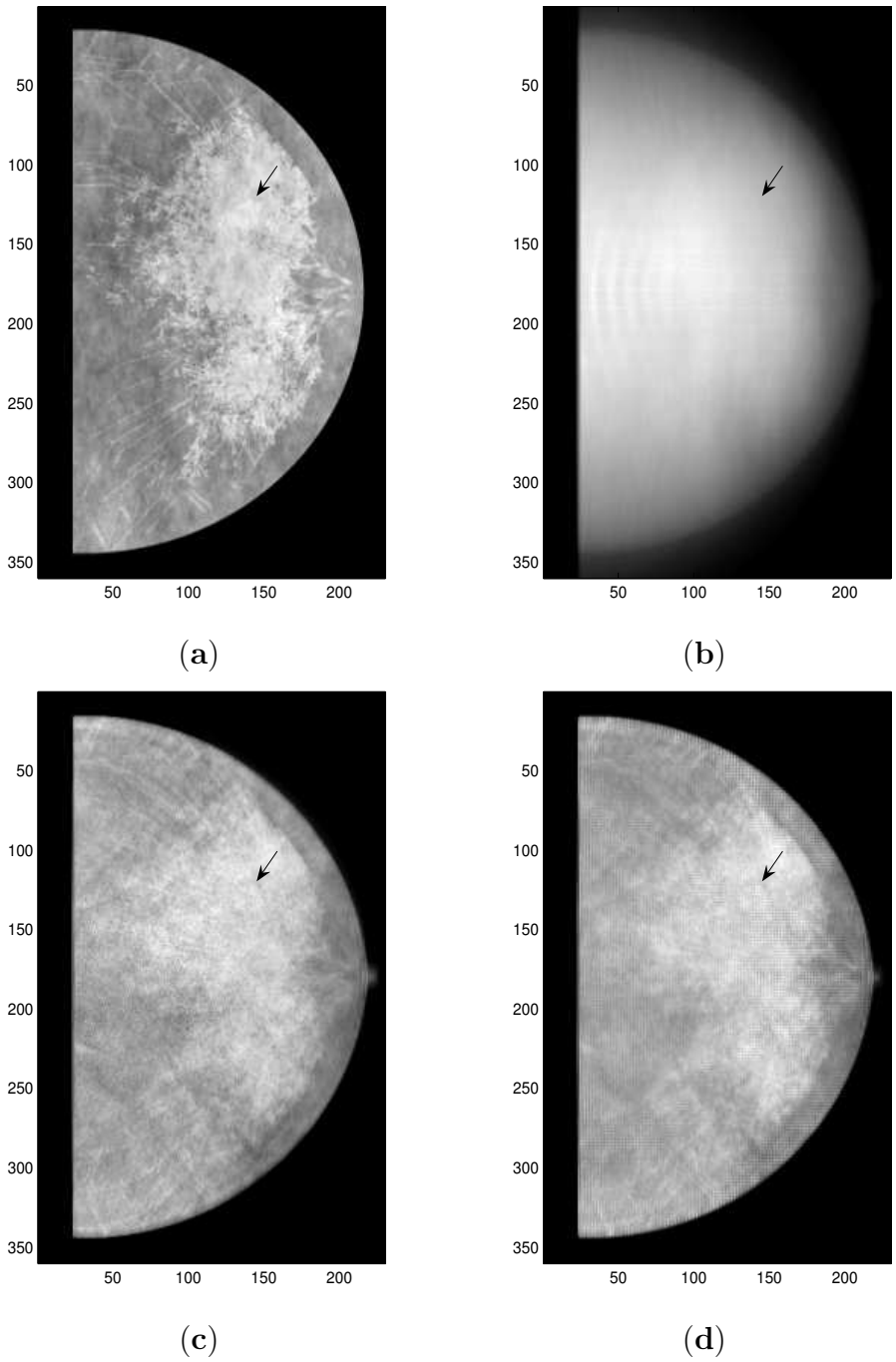


Figure 8.3: (a) Averaged slice image over eight consecutive slices of the original phantom with lesion present. (b), (c) and (d) reconstructed slice images with lesion present from one set of projection data by BP, SART, and EM, respectively. Arrows indicate lesion positions. Note that the averaged phantom and reconstructions have a 3.2 mm slice thickness, whereas the original phantom slice is 0.4 mm thick.

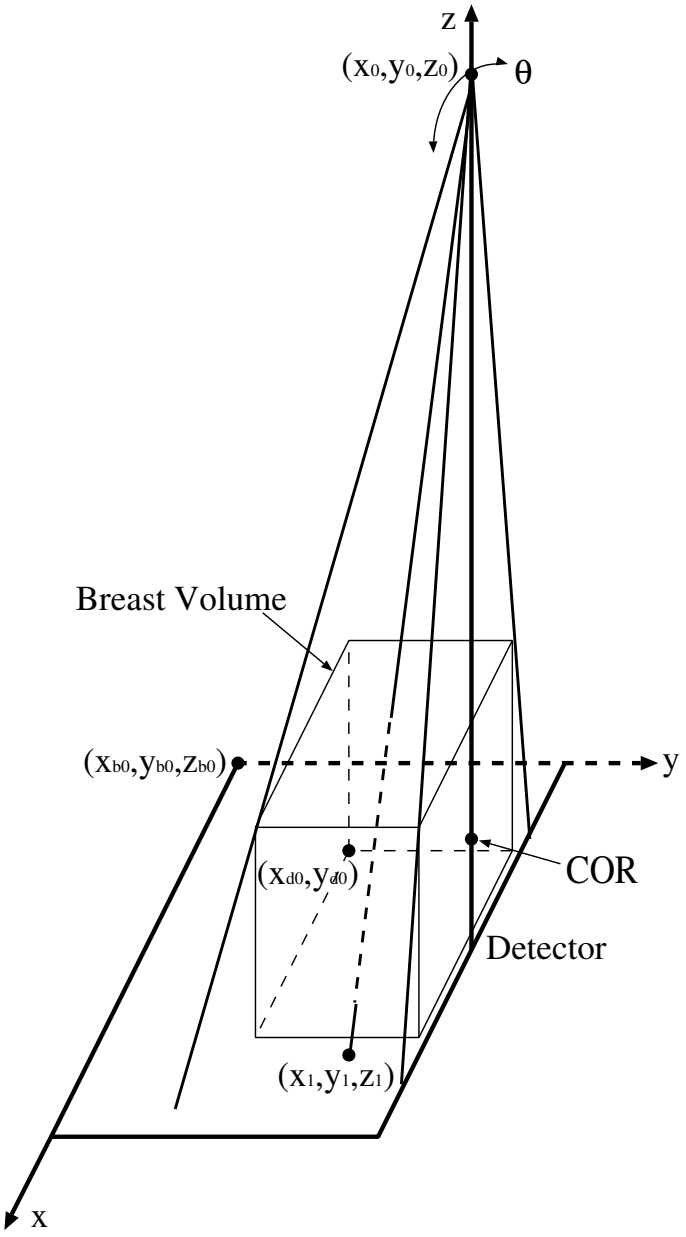


Figure 8.4: Cone-beam tomosysthesis geometry used in our calculation. COR is the center of rotation. See text for detailed explanation.

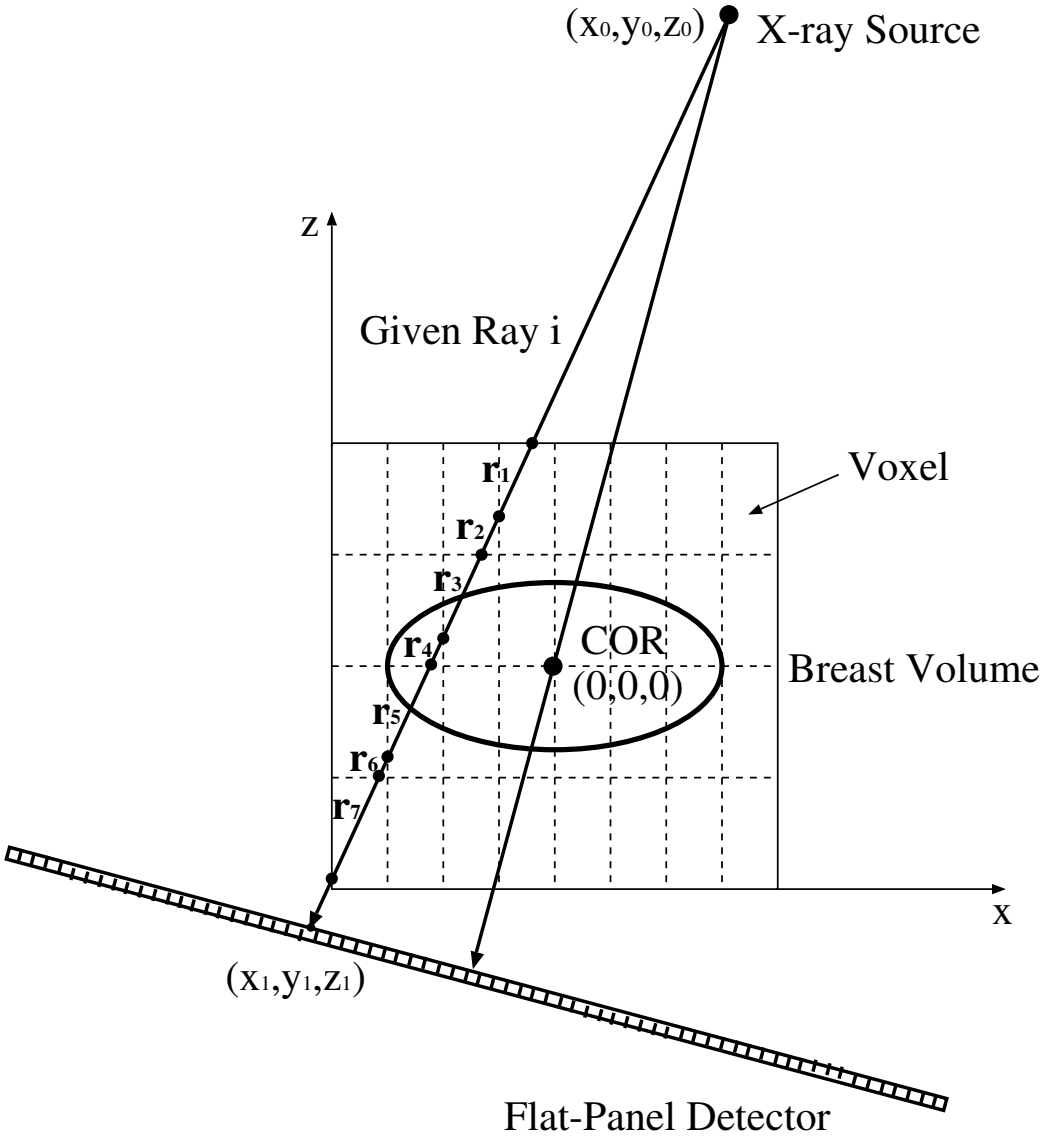


Figure 8.5: Geometry for calculation of quantities for rectanguloid voxel. Only x, z axises are displayed. See text for detailed explanation.

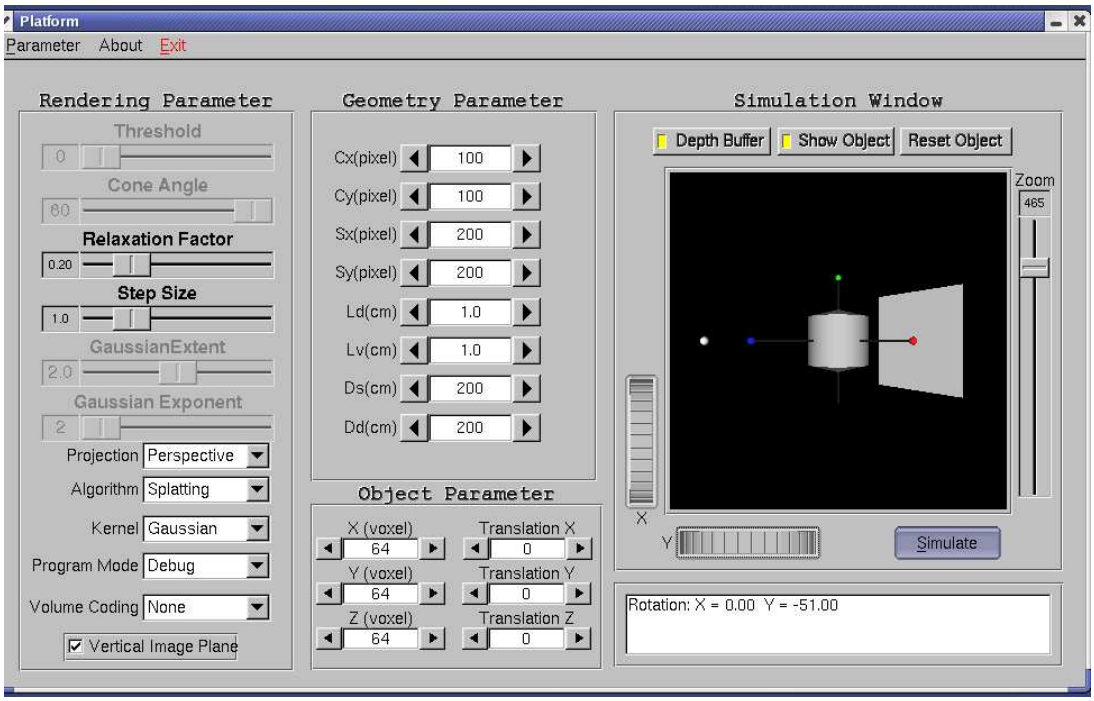
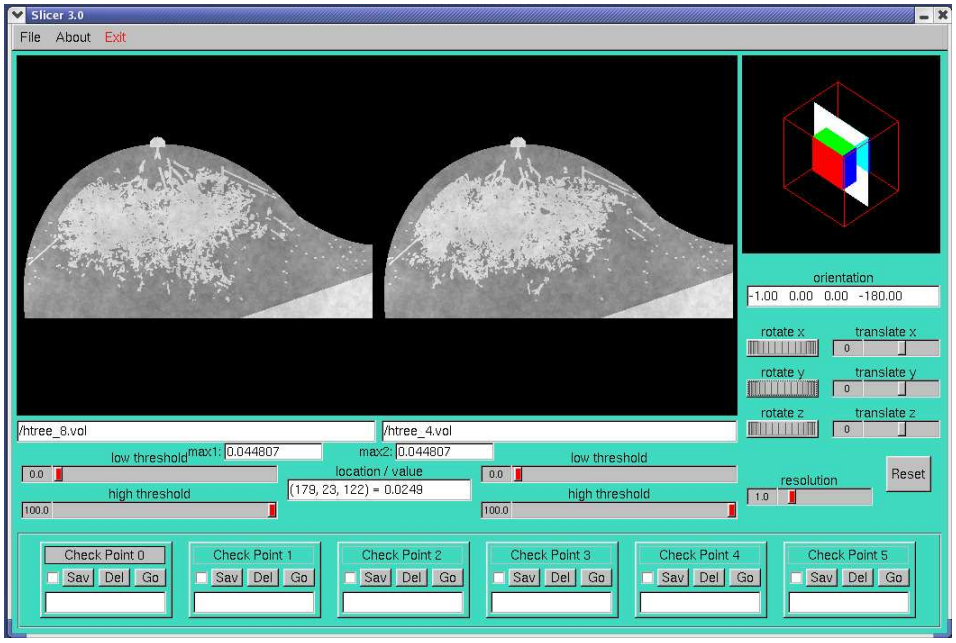
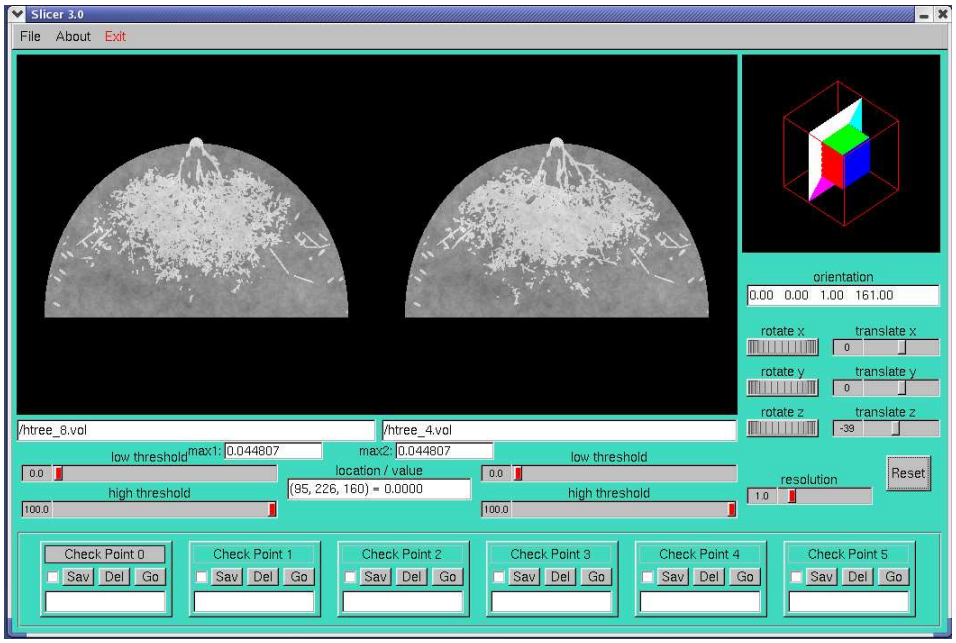


Figure 8.6: the placer display.



(**a**)



(**b**)

Figure 8.7: (a) the slicer display of ML view in a breast object (b) the slicer display of CC view in a breast object.

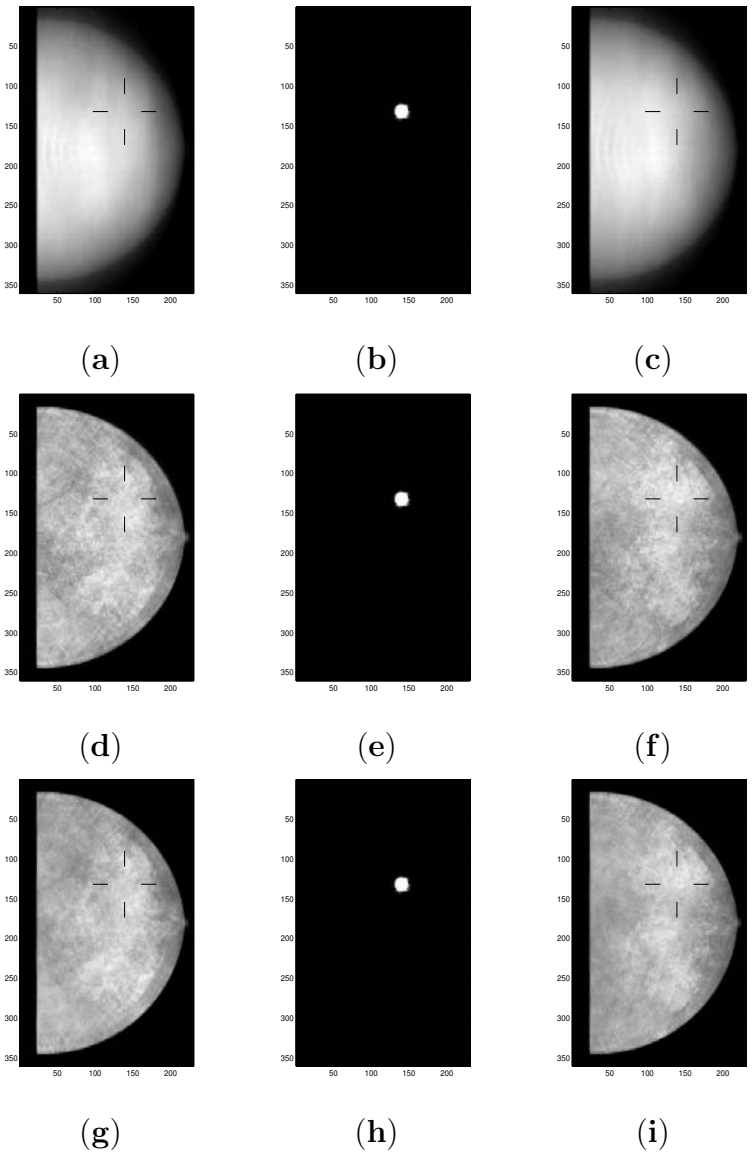


Figure 8.8: 2AFC displays. Each row shows a typical 2AFC display. The lesion instance is shown in the central panel to help the human observer. The crosshairs localize the potential lesion locations. The left and right panels display two noisy reconstructions. The observer chooses which reconstruction (left or right) contains a lesion. All slices are in the CC orientation. (a) (b) (c) for BP. (d) (e) (f) for SART. (g) (h) (i) for EM.

Chapter 9

Future Work

9.1 Improvements in the Breast Phantom

One contribution of this dissertation is to simulate a realistic 3D breast phantom with various anatomical components. However, our phantom is still crude and needs more refinements. The blood vessels, nerves and lymphatic vessels are parts of the breast anatomy; they also represent a disturbance in detecting the lesion and become a form of anatomical noise. We shall take into account the blood vessels, nerves and lymphatic vessels in the future. Modeling such anatomical components can be done via a method similar to that used in modeling the fibrous connective tissue as discussed in Section 6.2. To simulate them, the random size, length, and thickness of each component have to be specified via some simulation procedure.

In this dissertation, we only considered low-contrast lesion detection task. However, a high-contrast microcalcification is another type of abnormality. As discussed in Section 2.1.2, microcalcifications appear as either a tiny single spot or a cluster. It is trivial to simulate a single spot. We can adopt the method described in [109] to model a cluster of microcalcifications defined by the location in the breast, the size and the number of microcalcifications. The location of the cluster is randomly chosen to lie within the region of the ductal system. The user needs to specify the radius of the cluster and the mean number of microcalcifications. The number of microcalcifications

follows Poisson distribution. Each microcalcification is defined by its size and location in the cluster. The location of each microcalcification is uniformly sampled inside the cluster. The microcalcification could be formed as an ellipsoid with semi-axes that are uniformly sampled in a user-defined range.

Our simulated breast phantom lacks compression. As discussed in Section 3.5, mild compression is needed to keep the breast stabilized in breast tomosynthesis. Modeling breast compression is necessary to make our project more realistic, and a possible branch in our future work. One possible way to simulate deformation can be based on elasticity properties of various breast components. However, even a crude deformation model is very difficult to form, and involves lots of efforts for mechanical deformation modeling.

Low resolution ($0.4mm$ cubic voxels) is another limitation of our current phantom. Generating high resolution ($0.1mm$ cubic voxels) is a project for future work. Because real detector resolution is quite good and conventional mammograms have a good resolution (50 μm), we'd like the XY resolution of a tomosynthetic slice to approach that of a real mammogram, and so we need correspondingly small voxels in the phantom. Using smaller voxels leads to severe computational complexity for reconstruction in the projection and backprojection steps since these steps are $\mathcal{O}(N^3)$, where N is the number of voxels. Such problems are addressable in part by partitioning the problem onto a high-speed cluster. (We currently have a 16-node cluster that could be used for this.) Another speedup for projection/backprojection is to use GPU hardware [126, 127].

9.2 Other Improvements

Our goal had been to explore the limitations of detectability due to the fundamental restrictions of Poisson and anatomical noise. Other contributions to detectability

loss, such as electronic noise, were not considered. Though it would be quite a bit of work, it would be interesting to see what portion of detectability loss is due to Poisson and anatomical noise, and what portion is due to other sources including detector effects, moving focal spot, scatter, patient motion, and x-ray source characteristics. This tells us how much improvement to expect in the difficult problem of optimizing these other sources.

Our task of detection of a mass in an XY plane that intersects the mass center is very simple. A more realistic test, whether using human observers or numerical observers, is to force the observer to locate the lesion within the 3D reconstruction. Our group has done extensive research into numerical observer implementations that emulate the search and detection task [128, 129, 130].

References

[1] American Cancer Society, "Cancer Facts and Figures 2005", www.cancer.org, 2005.

[2] Oncology Channel, "Breast Cancer", www.oncologychannel.com, 2006.

[3] S. Feig, "Decreased Breast Cancer Mortality Through Mammographic Screening: Results of Clinical Trials", *Radiology*, 167, pp. 659–665, 1988.

[4] S. Feig, "Estimation of Currently Attainable Benefit From Mammographic Screening of Women Aged 40-49 Years", *Cancer*, 75, pp. 2412–2419, 1995.

[5] M. J. Yaffe, "Digital Mammography", In J. Beutel, H.L. Kundel, and R.L. V. Metter, Eds., *Handbook of Medical Imaging: volume 1. Physics and Psychophysics*, SPIE-The International Society for Optical Engineering, 2000.

[6] E. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J. Baum, S. Acharyya, E. Conant, L. Fajardo, L. Bassett, C. D'Orsi, R. Jong, and R. Rebner, "Diagnostic Performance of Digital Versus Film Mammography for Breast-Cancer Screening", *The New England Journal of Medicine*, 353(17), pp. 1773–1783, 2005.

[7] A. Smith, "Fundamentals of Digital Mammography: Physics, Technology and Practical Considerations", Hologic, Inc., March 2005.

[8] J. T. Bushberg, J. A. Seibert, E. M. Leidholdt Jr., and J. M. Boone, *"The Essential Physics of Medical Imaging"*, Lippincott Williams and Wilkins, 2002.

[9] L.T. Niklason, B. T. Christian, L. E. Niklason, D. B. Kopans, D. E. Castleberry, B. H. Opsahl-Ong, C. E. Landberg, P. J. Slanetz, A. A. Giardino, R. Moore, D. Albagli, M. C. Dejule an dP. F. Fitzgerald, D. F. Fobare, B. W. Giambattista, R. F. Kwasnick, J. Liu, S. J. Lubowski, G. E. Possin, J. F. Richotte, C. Wei, and R. F. Wirth, "Digital Tomosynthesis in Breast Imaging", *Radiology*, pp. 399–406, 1997.

[10] www.cancer.gov.

[11] J.M. Boone, "X-ray Production, Interaction, and Detection in Diagnostic Imaging", In J. Beutel, H.L. Kundel, and R.L. V. Metter, Eds., *Handbook of Medical Imaging: volume 1. Physics and Psychophysics*, SPIE-The International Society for Optical Engineering, 2000.

[12] D. B. Kopans, *Breast Imaging*, Lippincott-Raven, 1998.

[13] G. Cardenosa, *"Breast Imaging"*, Lippincott Williams & Wilkins, 2004.

[14] R. L. Egan, *"Breast Imaging: Diagnosis and Morphology of Breast Diseases"*, 1988.

[15] L. C. Junqueria, J. Carneiro, and R. O. Kelley, *"Basic Histology "*, Norwalk: Appleton and Lange, 1989.

[16] E. Shaw De Paredes, *"Atlas of Film-Screen Mammography"*, Urban and Schwarzenberg, Baltimore, MD, 1989.

[17] A. Fandos-Morera, M. Prats-Esteve, J. M. Tura-Soteras, and A. Traveria-Cros, "Breast Tumors: Composition of Microcalcifications", *Radiology*, 169, pp. 325–327, 1988.

[18] www.acr.gov.

[19] Shroy. R, M. S. Van Lysel, and M. J. Yaffe, "X-ray", In *the Biomedical Engineering Handbook*, CRC Press, Inc., 1995.

[20] W. Zhao, R. Deych, and E. Dolazza, "Optimization of Operational Conditions for Direct Digital Mammography Detectors for Digital Tomosynthesis", In *Proc. SPIE Med Imaging*, volume 5745, pp. 1272–1281, San Diego, CA, 2005.

[21] M. J. Yaffe, "Mammography", In *the Biomedical Engineering Handbook*, CRC Press, Inc., 1995.

[22] H. H. Barrett and W. Swindell, *The Theory of Image Formation, Detection, and Processing*, volume 1, Academic Press, 1981.

[23] G. T. Barnes and I. A. Berzovich, "The Intensity of Scattered Radiation in Mammography", *Radiology*, pp. 126–243, 1978.

[24] X. Gong, S. Glick, B. Liu, A. Vedula, and S. Thacker, "A Computer Simulation Study Comparing Lesion Detection Accuracy with Digital Mammography, Breast Tomosynthesis, and Cone-Beam CT Breast Imaging", *Med. Phys.*, 33(4), pp. 1–12, April 2006.

[25] http://imaginis.com.

[26] www.radiologyinfo.org.

[27] J. A. Rowlands and J. Yorkston, "Flat Panel Detectors for Digital Radiography ", In J. Beutel, H.L. Kundel, and R.L. V. Metter, Eds., *Handbook of Medical Imaging: volume 1. Physics and Psychophysics*, SPIE-The International Society for Optical Engineering, 2000.

[28] C. H. Chnag, J. L. Sibala, S. L. Fritz, J. H. Gallagher, S. J. Dwyer III, and A. W. Templeton, "Computed Tomographic Evaluation of The Breast", *American Journal of Roentgenology*, 131(3), pp. 459–464, 1978.

[29] B. Chen and R. Ning, "Cone-Beam Volume CT Breast Imaging: Feasibility Study", *Med. Phys.*, 29(5), pp. 755–770, 2002.

[30] A. A. Vedula, S. J. Glick, and X. Gong, "Computer Simulation of CT Mammography Using A Flat-Panel Imager", In *Proc. SPIE Physics of Medical Imaging*, volume 5030, pp. 349–360, San Diego, CA, 2003.

[31] J. M. Boone, T. R. Nelson, K. K. Lindfors, and J. A. Seibert, "Dedicated Breast CT: Radiation Dose and Image Quality Evaluation", *Radiology*, 221, pp. 657–667, 2001.

[32] L. Chen, C. C. Shaw, S. Tu, M. Altunbas, T. Wang, C. Lai, X. Liu, and S. Kappadath, "Cone-Beam CT Breast Imaging with A Flat Panel Detector-A Simulation Study", In *Proc. SPIE Med Imaging*, volume 5745, pp. 943–951, San Diego, CA, 2005.

[33] J. M. Boone, A. L. C. Kwan, T. R. Nelson, N. Shah, G. Burkett, J. A. Seibert, K. K. Lindfors, and G. Roos, "Performance Assessment of A Pendant-Geometry CT Scanner for Breast Cancer Detection", In *Proc. SPIE Med Imaging*, volume 5745, pp. 319–323, San Diego, CA, 2005.

[34] M. P. Tornai, R. L. McKinley, C. N. Brzymialkiewicz, P. Madhav, S. J. Cutler, D. J. Crotty, J. E. Bowsher, E. Samei, and C. E. Floyd Jr, "Design and Development of A Fully-3D Dedicated X-Ray Computed Mammotomography System", In *Proc. SPIE Med Imaging*, volume 5745, pp. 189–197, San Diego, CA, 2005.

[35] J. M. Boone and T. R. Nelson, "The Case for Dedicated CT Screening for Breast Cancer", *Radiology Suppl.*, S, pp. 217–601, 2000.

[36] S. Glick, S. Thacker, and X. Gong, "The Importance of Modeling Normal Mammographic Structure in Optimizing Flat-Panel CT Breast Imaging Systems", In *Proc. SPIE Med Imaging*, volume 5745, pp. 847–859, San Diego, CA, 2005.

[37] X. Gong, A. Vedula, S. Thacker, and S. Glick, "A Comparison of Lesion Detection Accuracy Using Digital Mammography and Flat-Panel CT Breast Imaging", In *Proc. SPIE Med Imaging*, volume 5745, pp. 860–869, San Diego, CA, 2005.

[38] S. J. Glick, S. Vedantham, and A. Karellas, "Investigation of Optimal kVp Settings for CT Mammography Using A Flat-Panel Imager", In *Proc. SPIE Med Imaging*, volume 4682, pp. 392–402, San Diego, CA, 2002.

[39] R. Ning, Y. Yu, D. L. Conover, X. Lu, H. He, Z. Chen, L. Schiffhauer, and J. Cullinan, "Preliminary System Characterization of Flat-Panel-Detector-Based Cone-Beam CT for Breast Imaging", In *Proc. SPIE Med Imaging*, volume 5368, pp. 292–303, 2004.

[40] J. T. Dobbins III and D. J. Godfrey, "Digital X-ray Tomosynthesis: Current State of the Art and Clinical Potential", *Phys. Med. Biol.*, 48, pp. R65–R106, Sept. 2003.

[41] S. Glick and X. Gong, "Optimal Spectra for Indirect Detector Breast Tomosynthesis", In *Proc. SPIE Med Imaging*, volume 6142, San Diego, CA, 2006.

[42] T. Wu, B. Liu, R. Moore, and D. Kopans, "Optimal Acquisition Techniques for Digital Breast Tomosynthesis Screening", In *Proc. SPIE Med Imaging*, volume 6142, San Diego, CA, 2006.

[43] T. Wu, A. Stewart, M. Stanton, T. McCauley, W. Phillips, D. B. Kopans, R. H. Moore, J. W. Eberhard, B. Opsahi-Ong, L. Niklason, and M. B. Williams, "Tomographic Mammography Using A Limited Number of Low-dose Cone-beam Projection Images", *Med. Phys.*, 30(3), pp. 365–380, 2003.

[44] Y. Chen, J. Lo, and J. Dobbins III, "Noise Power Spectrum Analysis for Several Digital Breast Tomosynthesis Reconstruction Algorithms", In *Proc. SPIE Med Imaging*, volume 6142, pp. 6142591–6142598, San Diego, CA, 2006.

[45] B. Ren, C. Ruth, J. Stein, A. Smith, I. Shaw, and Z. Jing, "Design and Performance of the Prototype Full Field Breast Tomosynthesis System with Selenium Based Flat Panel Detector", In *Proc. SPIE Med Imaging*, volume 5745, pp. 550–561, San Diego, CA, 2005.

[46] F. O. Bochud, F. R. Verdun, and J. F. Valley, "The Importance of Anatomical Noise in Mammography", In *SPIE. Medical Imaging*, volume 3036, pp. 74–80, San Diego, CA, Feb. 1997.

[47] A. E. Burgess, F. L. Jacobson, and P. F. Judy, "Human Observer Detection Experiments with Mammograms and Power-Law Noise", *Med. Phys.*, 28(4,), pp. 419–437, April 2001.

[48] J. Boone, K. Lindfors, V. Cooper III, and J. Seibert, "Scatter/Primary in Mammography: Comprehensive Results", *Med. Phys.*, 27(10), pp. 2408–2416, Oct. 2000.

[49] G. Barnes, "Contrast and Scatter in X-ray Imaging", *Radiographics*, 11, pp. 307–323, 1991.

[50] D. Dance and G. Day, "The Computation of Scatter in Mammography by Monte Carlo Methods", *Phys. Med. Biol.*, 29, pp. 237–247, 1984.

[51] S. Fritz, C. Chang, and W. Livingston, "Scatter/Primary Ratios for X-ray Spectra Modified to Enhance Iodine Contrast in Screen-Film Mammography", *Med. Phys.*, 10, pp. 866–870, 1983.

[52] Z. Jing, W. Huda, and J. Walker, "Scattered Radiation in Scanning Slot Mammography", *Med. Phys.*, 25, pp. 1111–1117, 1998.

[53] B. Liu, T. Wu, R. Moore, and D. Kopans, "Monte Carlo Simulation of X-ray Scatter Based on Patient Model from Digital Breast Tomosynthesis", In *Proc. SPIE Med Imaging*, volume 6142, pp. 61421N1–61421N9, San Diego, CA, 2006.

[54] A. Smith, "Full Field Breast Tomosynthesis", Hologic, Inc., March 2003.

[55] W. Zhao and J. Rowlands, "Digital Radiology Using Active Matrix Readout of Amorphous Selenium: Theoretical Analysis of Detective Quantum Efficiency", *Med. Phys.*, 24(12), pp. 1819–1833, Dec. 1997.

[56] I. A. Cunningham, "Applied Linear-Systems Theory", In J. Beutel, H.L. Kundel, and R.L. V. Metter, Eds., *Handbook of Medical Imaging: volume 1. Physics and Psychophysics*, SPIE-The International Society for Optical Engineering, 2000.

[57] A. Rosenfeld and A. C. Kak, *"Digital Picture Processing"*, volume 1, Academic Press, New York, 2nd edition, 1982.

[58] B. Chan, M. Bergstrom, M. R. Palmer, C. Sayre, and B. D. Pate, "Scatter Distribution in Transmission Measurements with Positron Emission Tomography", *J. Comp. Assisted Tomo.*, 10, pp. 296–301, Mar. 1986.

[59] A. Macovski, *"Medical Imaging Systems"*, Prentice-Hall, Englewood Cliffs, 1983.

[60] J. A. Fessler, "Statistical Image Reconstruction Methods for Transmission Tomography", In Milan Sonka and J. Michael Fitzpatrick, Eds., *Handbook of Medical Imaging: Volume 2. Medical Image Processing and Analysis*, SPIE-The International Society for Optical Engineering, 2000.

[61] J. A. Fessler, "Statistical Image Reconstruction, Short Course Notes", In *Conf. Rec. IEEE Nuc. Sci. Symp. Med. Imaging Conf.*, 2002.

[62] T. Wu, R. H. Moore, E. A. Rafferty, and D. B. Kopans, "A Comparison of Reconstruction Algorithms for Breast Tomosynthesis", *Med. Phys.*, 31(9), pp. 2636–2647, Sept. 2004.

[63] Y. Zhang, H. Chan, B. Sahiner, J. Wei, M. Goodsitt, L. M. Hadjiiski, J. Ge, and C. Zhou, "A Comparative Study of Limited-Angle Cone-Beam Reconstruction Methods For Breast Tomosynthesis", *Med. Phys.*, 33(10), pp. 3781–3785, Oct. 2006.

[64] R. L. Webber, "Self-Calibrated Tomosynthetic, Radiographic-Imaging System, Method and Device", Patent #5,668,844 (United States), 1997.

[65] R. L. Webber, R. A. Horton, D. A. Tyndall, and J. B. Ludlow, "Tuned-Aperture Computed Tomography (TACT-TM): Theory and Application for Three-Dimensional Dentoalveolar Imaging", *Dentomaxillofac. Radiol.*, 26, pp. 53–62, 1997.

[66] S. Suryanarayanan, A. Karellas, S. Vedantham, S. J. Glick, C. J. D'Orsi, S. P. Baker, and R. L. Webber, "Comparison of Tomosynthesis Methods Used With Digital Mammography", *Acad. Radiol.*, 7, pp. 1085–1097, 2000.

[67] U. E. Ruttiman, A. J. Roelf, and R. L. Webber, "Restoration of Digital Multiplane Tomosynthesis by A Constrained Iteration Scheme", *IEEE Trans Med Imaging*, MI-3, pp. 141–148, 1984.

[68] B. Andia, K. Sauer, and C. Bouman, "Nonlinear Backprojection for Tomographic Reconstruction", *IEEE Trans. Nucl. Sci.*, 49(1), pp. 61–68, Feb. 2002.

[69] A.C. Kak and M. Slaney, *"Principles of Computerized Tomographic Imaging"*, IEEE Press, 1988.

[70] J. T. Dobbins III, A. O. Powell, and Y. K. Weaver, "Matrix Inversion Tomosynthesis", In *RSNA 73rd Scientific Assembly*, 1987.

[71] G. T. Herman, *"Image Reconstruction From Projections: the Fundamentals of Computerized Tomography"*, New York: Academic Press, 1980.

[72] Z.H. Cho, J.P. Jones, and M. Singh, *Foundations of Medical Imaging*, John Wiley & Sons, Inc, 1993.

[73] D. B. Smith, "Image Reconstruction From Cone-Beam Projections: Necessary and Sufficient Conditions and Reconstruction Methods", *IEEE Trans. Med. Imaging*, M1-4, pp. 14–25, 1985.

[74] L. A. Keldkamp, L. C. Davis, and J. W. Kress, "Practical Cone-Beam Algorithm", *J. Opt. Soc. Am. A.*, 1, pp. 612–619, 1984.

[75] J. A. Fessler, "Hybrid Poisson/Polynomial Objective Functions for Tomographic Image Reconstruction from Transmission Scans", *IEEE Tr. Im. Proc.*, 4, pp. 1439–1450, Oct. 1995.

[76] D. S. Lalush and B. M. W. Tsui, "MAP-EM and WLS-MAP-CG Reconstruction Methods for Transmission Imaging in Cardiac SPECT", In *Conf. Rec. IEEE Nuc. Sci. Sym. Med. Imaging Conf.*, volume 2, pp. 1174–1178, 1993.

[77] S. H. Manglos, "Truncation Artifact Suppression in Cone-Beam Radionuclide Transmission CT Using Maximum Likelihood Techniques: Evaluation With Human Subjects", *Phys. Med. Biol.*, 37, pp. 549–562, Mar. 1992.

[78] E. P. Ficaro and J. A. Fessler, "Iterative Reconstruction of Truncated Fan Beam Transmission data", In *Conf. Rec. IEEE Nuc. Sci. Sym. Med. Imaging Conf.*, volume 3, 1993.

[79] J. A. Case, T. S. Pan, M. A. King, D. S. Luo, B. C. Penney, and M. S. Z. Rabin, "Reduction of Truncation Artifacts in Fan Beam Transmission Imaging Using A Spatially Varying Gamma Prior", *IEEE Trans. Nucl. Sci.*, 42, pp. 2260–2265, Dec. 1995.

[80] S. H. Manglos, G. M. Gagne, A. Krol, F. D. Thomas, and R. Narayanaswamy, "Transmission Maximum-Likelihood Reconstruction with Ordered Subsets for Cone Beam CT", *Phys. Med. Biol.*, 40, pp. 1225–1241, July 1995.

[81] T. S. Pan, B. M. W. Tsui, and C. L. Bryne, "Choice if Initial Conditions in the ML Reconstruction of Fan-Beam Transmission with Truncated Projection Data", *IEEE Trans. Med. Imaging*, 16, pp. 426–438, Aug. 1997.

[82] G. L. Zeng and G. T. Gullberg, "An SVD Study of Truncated Transmission Data in SPECT", *IEEE Trans. Nucl. Sci.*, 44, pp. 107–111, Feb. 1997.

[83] A. J. Rockmore and A. Macovski, "A Maximum Likelihood Approach to Transmission Image Reconstruction from Projections", *IEEE Trans. Nucl. Sci.*, 24, pp. 1929–1935, June 1977.

[84] K. Lange and R. Carson, "EM Reconstruction Algorithms for Emission and Transmission Tomography", *Journal of Computer Assisted Tomography*, 8(2), pp. 306–316, April 1984.

[85] L. A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography", *IEEE Tr. Med. Im.*, 1, pp. 113–122, Oct. 1982.

[86] C. Wu, On the Convergence Properties of the EM Algorithm, *Annls. Statist.*, 11(1), pp. 95–103, 1983.

[87] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, "Grouped-Coordinate Ascent Algorithms for Penalized-Likelihood Transmission Image Reconstruction", *IEEE Tr. Med. Im.*, 16, pp. 166–175, April 1997.

[88] H. Erdogan and J. A. Fessler, "Monotonic Algorithms for Transmission Tomography", *IEEE Trans. Med. Imaging*, 18(9), pp. 801–814, Sep. 1999.

[89] S. Ahn, F. Fessler, D. Blatt, and A. Hero, "Convergent Incremental Optimization Transfer Algorithms: Application to Tomography", *IEEE Trans. Med. Imaging*, 25(3), pp. 283–296, March 2006.

[90] K. Lange and J. A. Fessler, "Globally Convergent Algorithms for Maximum A Posterior Transmission Tomography", *IEEE Tr. Im. Proc.*, 4(10), pp. 1430–1438, 1995.

[91] A. R. De Pierro, "On the Relation Between the ISRA and the EM Algorithm for Positron Emission Tomography", *IEEE Trans. Med. Imaging*, 12(2), pp. 328–333, June 1993.

[92] A. R. De Pierro, "A Modified Expectation Maximization Algorithm for Penalized Likelihood Estimation in Emission Tomography", *IEEE Trans. Med. Imaging*, 14(1), pp. 132–137, Mar. 1993.

[93] S. Geman and D. E. McClure, "Bayesian Image Analysis: An Application To Single Photon Emission Tomography", In *Proc. of Stat. Comp. Sect. of Amer. Stat. Assoc.*, pp. 12–18, 1985.

[94] H. Erdogan and J. A. Fessler, "Ordered Subsets Algorithms for Transmission Tomography", *Phys. Med. Biol.*, 44(11), pp. 2835–2851, Nov. 1999.

[95] J. A. Fessler and A. O. Hero, "Penalized Maximum-Likelihood Image Reconstruction Using Space-Alternating Generalized EM Algorithms", *IEEE Tr. Im. Proc.*, 4, pp. 1417–1429, Oct. 1995.

[96] H. Erdogan and J. A. Fessler, "Accelerated Monotonic Algorithms for Transmission Tomography", In *Proc. IEEE Intl. Conf. on Image Processing*, volume 2, pp. 680–684, 1998.

[97] J. A. Fessler, "Grouped Coordinate Decent Algorithms for Robust Edge-Preserving Image Restoration", In *Proc. SPIE 3071, Image Reconstruction and Restoration II*, pp. 184–194, 1997.

[98] H. M. Hudson and R. S. Larkin, "Accelerated Image Reconstruction Using Ordered Subsets of Projection Data", *IEEE Trans. Med. Imaging*, 13(4), pp. 601–609, 1994.

[99] C. Kamphius and F. J. Beekman, "Accelerated Iterative Transmission CT Reconstruction Using An Ordered Subsets Convex Algorithm", *IEEE Trans. Med. Imaging*, 17, pp. 1001–1005, 1983.

[100] I-T. Hsiao, A. Rangarajan, P. Khurd, and G. Gindi, "An Accelerated Convergent Ordered Subset Algorithm for Emission Tomography", *Phys. Med. Biol.*, 49(11), pp. 2145–2156, 2004.

[101] D. Monticciodo, "Breast Disease: Anatomy, Pathology, and Diagnosis", In RSNA *Categorical Course in Diagnostic Radiology Physics: Advances in Breast Imaging-Physics, Technology, and Clinical Applications*, pp. 9–11, 2004.

[102] H. Li, M. Kallergi, L. Clarke, V. Jain, and R. Clark, "Markov Random Field for Tumor Detection in Digital Mammography", *IEEE Trans. Med. Imaging*, 14, pp. 565–576, 1995.

[103] J. Heine, S. Deans, R. Velthuizen, and L Clarke, "On the Statistical Nature of Mammograms", *Med. Phys.*, 26, pp. 2254–2265, 1999.

[104] F. Bochud, C. Abbey, and M. Eckstein, "Statistical Texture Synthesis of Mammographic Images with Clustered Lumpy Backgrounds", *Med. Phys.*, 29, pp. 2131–2139, 1999.

[105] F. Lefebvre, H. Benali, R. Gilles, and R. Paola, "A Simulation Model of Clustered Breast Calcifications", *Med. Phys.*, 21, pp. 1865–1874, 1994.

[106] M. Lado, P. Tahoces, M. Souto, A. Mendez, and J. Vidol, "Real and Simulated Clustered Microcalcifications in Digital Mammograms: ROC Study of Observer Performance", *Med. Phys.*, 24, pp. 1385–1394, 1997.

[107] J. Nappi and P. Dean, "Mammographic Feature Generator for Evaluation of Image Analysis Algorithms", In *SPIE. Medical Imaging*, pp. 911–918, Newport Beach, 1997.

[108] P. R. Bakic, M. Albert, D. Brzakovic, and A. D. Maidment, "Mammogram Synthesis Using a 3D Simulation. 1. Breast Tissue Model and Image Acquisition Simulation", *Med. Phys.*, 29(9), pp. 2131–2139, Sept. 2002.

[109] K. Bliznakova, Z. Bliznakov, V. Bravou, Z. Kolitsi, and N. Pallikarakis, "A Three-Dimensional Breast Software Phantom for Mammography Simulation", *Phys. Med. Biol.*, 48, pp. 3699–3719, Nov. 2003.

[110] V. Andolina, S. Lillé, and K. Willison, *"Mammographic Imaging: A Practical Guide"*, J. B. Lippincott Company, Philadelphia, 1992.

[111] P. Taylor, R. Owens, and D. Ingram, "Simulated Mammography Using 3D Breasts", In *Digital Mammography Nijmegen 1998, Proceedings of the 4th International Workshop on Digital Mammography*, pp. 283–290, Nijmegen, The Netherlands, June 1998.

[112] P. Bakic, M. Albert, D. Brzakovic, and A. Maidment, "Mammogram Synthesis Using A 3D Simulation. III. Modeling and Evaluation of the Breast Ductal Network", *Med. Phys.*, 30, pp. 1914–1925, 2003.

[113] A. E. Burgess and S. Chakraborty, "Producing Lesions for Hybrid Mammograms; Extracted Tumors and Simulated Microcalcifications", In *SPIE. Medical Imaging*, volume 3663, pp. 316–322, San Diego, CA, Feb. 1999.

[114] A. E. Burgess, "Mammmographic Structure: Data Preparation and Spatial Statistics Analysis", In *SPIE. Medical Imaging*, volume 3661, pp. 642–653, San Diego, CA, Feb. 1999.

[115] A. Samani, J. Bishop, M. Yaffe, and D. Plewers, "Biomechanical 3D Finite Element Modeling of the Human Breast Using MRI Data", *IEEE Tran. Medical Imaging*, 20(4), pp. 271–279, April 2001.

[116] E. Clarkson and H. H. Barrett, "An Approach to the Task-Based Assessment of Image Quality", *SIAM Activity Group on Imaging Science Siagis Newsletter*, 1(1), April 2002.

[117] K. J. Myers, "Ideal Observer Models of Visual Signal Detection", In J. Beutel, H. L. Kundel, and R. L. Van Metter, Eds., *Handbook of Medical Imaging: Volume 1. Physics and Psychophysics*, SPIE-The International Society for Optical Engineering, 2000.

[118] J. P. Rolland and H. H. Barrett, "Effect of Random Background Inhomogeneity on Observer Detection Performance", *J. Opt. Soc. Am. A*, 9(5), pp. 649–658, 1992.

[119] H. H. Barrett and K. J. Myers, *"Foundations of Image Science"*, John Wiley & Sons, Inc., 2004.

[120] C. E. Metz, "Basic Principles of ROC Analysis", *Seminars in Nuclear Medicine*, 8(4), pp. 283–298, Oct. 1978.

[121] P. Khurd, *"On Reconstruction Methods and Image Quality in Emission Tomography"*, PhD Thesis, Department of Electrical and Computer Engineering, SUNY at Stony Brook, Stony Brook, NY, 2005.

[122] C.K. Abbey, *"Assessment of Reconstructed Images"*, PhD Thesis, The University of Arizona, Tucson, Arizona, 1998.

[123] J. Oldan, S. Kulkarni, P. Khurd, and G. Gindi, "Channelized Hotelling and Human Observer Study of Optimal Smoothing in SPECT MAP Reconstruction", *IEEE Trans. Nucl. Sci.*, 51(3), pp. 733–741, June 2004.

[124] R. Siddon, "Fast Calculation of the Exact Radiological Path for a 3D CT Array", *Med. Phys.*, 12, pp. 252–255, 1985.

[125] G. Han, *"Image Reconstruction in Quantitative Cardiac SPECT with Varying Focal-Length Fan-Beam Collimators"*, PhD Thesis, Department of Physics, SUNY at Stony Brook, Stony Brook, NY, 2000.

[126] F. Xu and K. Mueller, "Accelerating Popular Tomographic Reconstruction Algorithms on Commodity PC Graphics Hardware", *IEEE Trans. Nucl. Sci.*, 52(3), pp. 654–663, 2005.

[127] F. Xu and K. Mueller, "Real-Time 3D Computed Tomographic Reconstruction Using Commodity Graphics Hardwar", *Phys. Med. Biol.(to appear)*, 2007.

[128] P. Khurd and G. Gindi, "Decision Strategies That Maximize the Area Under the LROC Curve", *IEEE Trans. Med. Imaging*, 24(12), pp. 1626–1636, Dec. 2005.

[129] P. Khurd and G. Gindi, "Fast LROC Analysis of Bayesian Reconstructed Emission Tomographic Images Using Model Observers", *Phys. Med. Biol.*, 50(7), pp. 1519–1532, 2005.

[130] P. Khurd and G. Gindi, "Rapid Computation of LROC Figures of Merit Using Numerical Observers (for SPECT/PET Reconstruction)", *IEEE Trans. Nucl. Sci.*, 52(3), pp. 618–626, 2005.