Electron Beam X-Ray Computed Tomography for Multiphase Flows and An Experimental Study of Inter-channel Mixing

by

Seongjin Yoon

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Naval Architecture and Marine Engineering)
in the University of Michigan
2017

Doctoral Committee:

Professor Steven L. Ceccio, Co-chair
Professor Simo A. Mäkiharju, Co-chair
Professor Jeffrey A. Fessler
Professor Marc Perlin

Seongjin Yoon

seongjin@umich.edu

ORCID iD: 0000-0002-3487-3497

To Heeyoung and Michelle

Acknowledgments

I would like to thank the members of my dissertation committee, professor Steven Ceccio and professor Marc Perlin for offering me the opportunity to study in the University of Michigan, professor Simo Mäkiharju for his efforts to run this project and instructions how to run the experiment, and professor Jeffrey Fessler for his valuable lectures regarding the image processing and discussions. I also thank all the committee members for the meticulous reviews on this dissertation.

I would like to thank the Office of Naval Research and the Bettis Atomic Power Laboratory for sponsoring this research.

I would like to thank my family, my wife Heeyoung Kim for tolerating such a difficult time with me, my baby Michelle for her endless healthy, witty smiles, and my parents, Wolsuk Kwak and Geukno Yoon for their endless trust and support on me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

This thesis consists of two parts. In the first, a high speed X-ray Computed Tomography (CT) system for multiphase flows is developed. X-ray Computed Tomography (CT) has been employed in the study of multiphase flows. The systems developed to date often have excellent spatial resolution at the expense of poor temporal resolution. Hence, X-ray CT has mostly been employed to examining time averaged phase distributions. In the present work, we report on the development of a Scanning Electron Beam X-ray Tomography (SEBXT) CT system that will allow for much higher time resolution with acceptable spatial resolution. The designed system, however, can have issues such as beam-hardening and limited angle artifacts. In the present study, we developed a high speed, limited angle SEBXT system along with a new CT reconstruction algorithm designed to enhance the CT reconstruction results of such system. To test the performance of the CT system, we produced example CT reconstruction results for two test phantoms based on the actual measured sinograms and the simulated sinograms.

The second part examines, the process by which fluid mixes between two parallel flow channels through a narrow gap. This flow is a canonical representation of the mixing and mass transfer processes that often occur in thermo-hydraulic systems. The mixing can be strongly related to the presence of large-scale periodic flow structures that form within the gap. In the present work, we have developed an experimental setup to examine the single-phase mixing through the narrow rectangular gaps connecting two rectangular channels. Our goal is to elucidate the underlying flow processes responsible for inter-

channel mixing, and to produce high-fidelity data for validation of computational models. Dye concentration measurements were used to determine the time average rate of mixing. Particle Imaging Velocimetry was used to measure the flow fields within the gap. A Proper Orthogonal Decomposition (POD) of the PIV flow fields revealed the presence of coherent flow structure. The decomposed flow fields were then used to predict the time averaged mixing, which closely matched the experimentally measured values.

CHAPTER 1

Background and Motivation

The following thesis is composed of two main sections. The first reports on the development of an x-ray computed tomography system for the measurement of multiphase flows, and the second reports on an experiment that was conducted to examine the mixing between two flow streams that are connected with a narrow stream-wise channel. In this chapter, we will provide some background and motivation for this work

1.1 Electron Beam X-Ray Computed Tomography for Multiphase Flows

The advancement of Computational Fluid Dynamics (CFD) methods along with enhanced computer power make the simulation of three dimensional multiphase flows feasible at a scale of practical interest. However, the complexity of even canonical high Reynolds number multiphase flows still presents a challenge. The accurate modeling of such multiphase flows will rely on high fidelity experimental observations that can resolve the phase fraction and dynamics of the various flow components [1]. Ideally, we would be able to measure the phase distributions with high spatial and temporal resolution, and researchers have developed a variety of intrusive and non-intrusive measurement methods.

A principal goal of multiphase flow measurements is often the characterization of the temporally and spatially evolving phase distribution without intrusive probes. Optical methods have been used for some relatively low volume fraction of the dispersed phase. Zachos *et al.* [2] performed PIV for the solid-liquid flows with 1.3% glass sphere volume fraction in the water, and Lindken and Merzkirch [3] performed Particle Imaging Velocimetry (PIV) for the gas-liquid flows with 2.5% gas volume fraction. As can be seen, optical methods can only be used for flows with very low volume fraction of the dispersed phase.

While optical methods have provided some tremendous insights, in many instances, multiphase flows are opaque, and optical methods cannot be employed. Therefore, a num-

ber of non-optical methods have also been developed. Many of these are well summarized in Heindel [4]. They include electrical, magnetic, and acoustic methods. Heindel [4] focusses his review on the use of x-rays to probe multiphase flows. X-ray enable the non-invasive visualization of high volume fraction dispersed phase flows (greater than a few percent disperse phase fraction). And, x-ray based measurements can be made with high spatial resolution for time-averaged flows. Methods previously employed include radiography [5–9], stereography [10–13], and the Computed Tomography (CT) [14–16]. X-ray CT systems have been developed that can measure multiphase flows with excellent spatial resolution but poor temporal resolution, and they have been used to mainly measure time-averaged phase distributions. Recently, researchers have attempted to develop faster X-ray CT system for the measurement of multiphase flows. Such systems often employed multiple stationary x-ray sources and a number of fixed detectors [17, 18]. Such systems can achieve a high temporal resolution with a sacrifice of spatial resolution. Another approach to enhancing the temporal resolution of X-ray CT is the use of a scanning electron beam to create a rapidly moving source of x-rays, known as Scanning Electron Beam X-ray Tomography (SEBXT) [19–26]. SEBXT can significantly improve the temporal resolution of the X-ray CT. Instead of physically rotating a source and detector pair, SEBXT deflects the electron beam by applying electro-magnetic fields to generate X-rays in a sequence of specific positions along a tungsten target. A fixed array of detectors acquires the projection images at the same time as the beam sweeps the arc.

Although EBXT significantly enhances the temporal resolution that can be achieved compared to conventional CT, it still has several drawbacks. Because of its fixed source-detector configuration, full-angle CT data requires angular overlap between the source target and the detector. The conventional solution was to place an axial offset between the source and the detector [20]. However, this axial displacement can generate artifacts due to the misalignment between the source and detector in 2-D CT as each projection scans a different plane. Furthermore, 2-D array detectors are not easily usable for this configuration because the target can block the detectors. To solve the overlapping problem while enabling the use of 2-D array detectors, a tungsten coated graphite target was developed, enabling full-angle CT by making the X-ray source almost transparent to X-rays [22]. However, it is not yet available for commercial applications.

In the present study, our goal is to develop a high speed, limited angle SEBXT system along with the new CT reconstruction algorithm specifically designed to enhance the CT reconstruction result of such SEBXT system under development at the University of Michigan. The desired frame rate is of the order 100 to 1000 frames per second. This would allow us to resolve the time-variation of high Reynolds number multiphase flows, while maintain-

ing adequate spatial resolution. In the present study, we characterized the electron beam X-ray system under development, we created a new CT reconstruction algorithm and the numerical simulation that is adapted for the new system, and then we compared the initial experimentally developed images with companion numerical simulation results.

This portion of the thesis is presented in Chapters 2 through 5. In §2, we described the scanning electron beam X-ray tomography (SEXBT) system, and the testing methods used to identify the beam focal spot and the beam localization. In §3, we provided a CT reconstruction algorithm developed for the specific SEBXT system. In §4, we produced example CT reconstruction results using two test phantoms, and the results are then compared the experimentally determined CT reconstructions with the simulated CT reconstructions to assess the performance of the current SEBXT system. In §5, we summarized the conclusions and present ideas for future work.

1.2 An Experimental Study of Inter-channel Mixing

Mixing through narrow gaps connecting adjacent primary flow paths is an important mass and heat transfer process for many thermo-hydraulic applications, such as flows through nuclear reactor rod bundles and heat exchangers. In addition to the steady mixing due to any pressure gradients across the gap and mixing due to small scale turbulent flows (incoherent turbulent), mixing can result from the development of large-scale periodic flow structures (coherent turbulent) due to the shear flows on either side of the gap. The presence of these large-scale flow structures can drastically change the rate of mixing. Several researchers have studied the basic flow processes of the flow between parallel flow channels connected by gaps [27–34]. Meyer [34] offers a comprehensive review of past work on inter-channel mixing. He reported that researchers have gleaned significant insights into the underlying flow processes, including the observation of large-scale coherent structures in the narrow gaps between the channels.

With the advent of advanced computational fluid dynamics solvers, the flow in complex geometries such as rod bundles can be simulated with increasing resolution. However, given the complexity of these flows, it is important to validate such models with careful observations. Moreover, it is essential that simulations be compared against experimental observations only after the appropriate boundary conditions are matched between both computation and experiment. Inter-channel mixing has been numerically investigated by numerous researchers, including Chang and Tavoularis [35, 36], Home *et al.* [37], Derksen [38], and Home and Lightstone [39]. Merzari *et al.* [40] also used Proper Orthogonal Decomposition (POD) to better understand the underlying dynamics of the flow oscilla-

tions observed in the inter-channel flow. However, few studies have produced experimental data sets that are immediately suitable for the validation of high-fidelity Computational Fluid Dynamics (CFD) models. Since the resulting flows are sensitive to small changes in boundary conditions (see Ko *et al.* [41] for example of the sensitivity of similar flows), data sets that do not have accompanying characterization of the flow boundary conditions are not ideally suited for rigorous CFD code Verification and Validation (V&V). Derksen [38] performed computations and compared his results with experimental data, but also noted the lack of expected symmetry in the experimental data, again suggesting that there was an underlying issue with the flow geometry, the inlet conditions, the outlet flow conditions, or some combination of effects.

The present work is the extension of the previously published work, Mäkiharju *et al.* 2015 [42], where we have developed a canonical geometry that is analogous to the sub-channel flow within a rod bundle. We have examined the single-phase mixing through the narrow rectangular gaps connecting two channels with emphasis on measurement uncertainties, quantifying the inflow conditions, and closely coordinating each stage of the experiment to produce high-fidelity V&V data sets. The inlet flow conditions extensively surveyed via Laser Doppler Velocimetry (LDV) in order to verify that the flow was well-conditioned as it approached the test section gap. We determined the bulk mass transfer through the various gap openings from the mass flow rate and fluorescein tracer dye concentration measurements at the channel inlets and outlets. We preformed these measurements for seven channel flow rate combinations and eleven gap heights for both test sections.

Planar Particle Image Velocimetry (PIV) was employed to visualize the flow within the gat at select conditions to further examine the dynamics of the mixing phenomena. We used Proper Orthogonal Decomposition (POD) to decompose the flow velocity field into multiple linear modes, to identify the coherent structures in the gap. As far as we know, POD in fluid experiments is used mostly for the validation purpose, and there has not been an attempt to calculate the mass transfer directly from the POD analysis. Here, we used these POD results along with the time-averaged velocity field to compute the gross mixing rate. These results are favorably compared to the measured mixing determined from measurements of dye concentration.

The basic experimental setup and the inflow conditions are described in §6. The time-averaged results of integral mixing of both balanced and unbalanced inlet mass flow rates are presented in §7. We provided some additional details of the observed coherent structures based on POD applied to PIV data in §8, and the estimation of the mixing coefficients based on the PIV in §9. Finally, we summarized the conclusions and possible future work

in §10.

BIBLIOGRAPHY

[1] M. P. Dudukovic, "Opaque Multiphase Reactors: Experimentation, Modeling and Troubleshooting," *Oil and Gas Science and Technology*, vol. 55, no. 2, pp. 135–158, 2000.

[2] A. Zachos, M. Kaiser, and W. Merzkirch, "PIV measurements in multiphase flow with nominally high concentration of the solid phase," *Experiments in Fluids*, vol. 20, no. 3, pp. 229–231, 1996.

[3] R. Lindken and W. Merzkirch, "A novel PIV technique for measurements in multiphase flows and its application to two-phase bubbly flows," *DLR-Mitteilung*, vol. 33, no. 3, pp. 991–1003, 2001.

[4] T. J. Heindel, "A Review of X-Ray Flow Visualization With Applications to Multiphase Flows," *Journal of Fluids Engineering*, vol. 133, no. 7, p. 074001, 2011.

[5] J. G. Yates, D. J. Cheesman, and B. Engineering, "Experimental Observations of Voidage Distribution," *Chemical Engineering Science*, vol. 49, no. 12, pp. 1885–1895, 1994.

[6] I. Hulme and A. Kantzas, "Determination of bubble diameter and axial velocity for a polyethylene fluidized bed using X-ray fluoroscopy," *Powder Technology*, vol. 147, no. 1-3, pp. 20–33, 2004.

[7] S. Roels and J. Carmeliet, "Analysis of moisture flow in porous materials using micro-focus X-ray radiography," *International Journal of Heat and Mass Transfer*, vol. 49, no. 25-26, pp. 4762–4772, 2006.

[8] P. M. Jenneson and O. Gundogdu, "In situ x-ray imaging of nanoparticle agglomeration in fluidized beds," *Applied Physics Letters*, vol. 88, no. 3, pp. 1–3, 2006.

[9] S. A. Mäkiharju, "The Dynamics of Ventilated Partial Cavities over a Wide Range of Reynolds Numbers and Quantitative Two-Dimensional X-ray Densitometry for Multiphase Flow," Ph.D. dissertation, University of Michigan, 2012.

[10] A. Seeger, K. Affeld, U. Kertzscher, L. Goubergrits, and E. Wellnhofer, "Assessment of Flow Structures in Bubble Columns by X-ray Based Particle Tracking Velocimetry," *Image (Rochester, N.Y.)*, pp. 1–7, 2001.

[11] A. Seeger, U. Kertzscher, K. Affeld, and E. Wellnhofer, "Measurement of the local velocity of the solid phase and the local solid hold-up in a three-phase flow by X-ray based particle tracking velocimetry (XPTV)," *Chemical Engineering Science*, vol. 58, no. 9, pp. 1721–1729, 2003.

[12] S. J. Lee and G. B. Kim, "X-ray particle image velocimetry for measuring quantitative flow information inside opaque objects," *Journal of Applied Physics*, vol. 94, no. 5, pp. 3620–3623, 2003.

[13] T. B. Morgan and T. J. Heindel, "Qualitative Observations of Dense Particle Motion in a Vibration-Excited Granular Bed," *Proceedings of IMECE2007*, pp. 1–5, 2007. [Online]. Available: http://link.aip.org/link/abstract/ASMECP/v2007/i43017/p835/s1

[14] D. Toye, P. Marchot, M. Crine, and G. L'Homme, "Modelling of multiphase flow in packed beds by computer-assisted x-ray tomography," *Measurement Science and Technology*, vol. 7, no. 3, p. 436, 1996.

[15] C. E. Schmit, D. B. Cartmel, and R. Eldridge, "The experimental application of X-ray tomography to a vaporliquid contactor," *Chemical Engineering Science*, vol. 56, no. 11, pp. 3431–3441, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0009250901000367

[16] W. van der Merwe, W. Nicol, and F. de Beer, "Three-dimensional analysis of trickle flow hydrodynamics: Computed tomography image acquisition and processing," *Chemical Engineering Science*, vol. 62, no. 24, pp. 7233–7244, 2007.

[17] T. Kai, M. Misawa, and T. Kanda, "Observation of 3-D Structure of Bubbles in a Fluidized Catalyst Bed," *The Canadian Journal of Chemical Engineering*, vol. 83, pp. 113–118, 2005.

[18] R. Mudde, "Bubbles in a Fluidized Bed: A Fast X-Ray Scanner," *Particle technology and fluidization*, vol. 57, no. 10, pp. 2684–2690, 2010.

[19] M. Bieberle and U. Hampel, "Evaluation of a limited angle scanned electron beam x-ray CT approach for two-phase pipe flows," *Meas. Sci. Technol.*, vol. 17, no. 8, pp. 2057–2065, aug 2006.

[20] F. Fischer, D. Hoppe, E. Schleicher, G. Mattausch, H. Flaske, R. Bartel, and U. Hampel, "An ultra fast electron beam x-ray tomography scanner," *Meas. Sci. Technol.*, vol. 19, p. 094002, 2008.

[21] U. Hampel, F. Barthel, M. Bieberle, and T. Strzel, "Transparent target for ultrafast electron beam tomography," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 635, no. 1, pp. 8–12, 2011.

[22] T. Stürzel, M. Bieberle, E. Laurien, U. Hampel, F. Barthel, H.-J. Menz, and H.-G. Mayer, "Experimental facility for two- and three-dimensional ultrafast electron beam x-ray computed tomography." *Rev. Sci. Instrum.*, vol. 82, no. 2, p. 023702, feb 2011.

[23] M. Bieberle, F. Barthel, and U. Hampel, "Ultrafast X-ray computed tomography for the analysis of gas-solid fluidized beds," *Chem. Eng. J.*, vol. 189-190, pp. 356–363, may 2012.

[24] U. Hampel, F. Barthel, M. Bieberle, M. Schubert, and E. Schleicher, "Multiphase flow investigations with ultrafast electron beam x-ray tomography," *AIP Conf. Proc.*, vol. 167, pp. 167–174, 2012.

[25] U. Hampel, Y. Bartling, D. Hoppe, N. Kuksanov, S. Fadeev, and R. Salimov, "Feasibility study for mega-electron-volt electron beam tomography," *Rev Sci Instrum*, vol. 83, no. 9, p. 93707, 2012.

[26] M. Bieberle and U. Hampel, "Level-set reconstruction algorithm for ultrafast limited-angle X-ray computed tomography of two-phase flows," *Philos. Trans. R. Soc. London A Math. Phys. Eng. Sci.*, vol. 373, no. 2043, 2015.

[27] L. Meyer and K. Rehme, "Large-Scale Turbulence Phenomena in Compound Rectangular Channels," *Exp. Therm. Fluid Sci.*, vol. 8, no. 4, pp. 286–304, 1994.

[28] A. Mahmood, M. Rohde, T. H. J. J. van der Hagen, R. F. Mudde, T. van der Hagen, and R. F. Mudde, "Contribution of Large-Scale Coherent Structures towards the Cross Flow in Two Interconnected Channels," in *NURETH-13*, Kanazawa, Japan, 2009. [Online]. Available: http://inis.iaea.org/Search/search.aspx?orig{_}q=RN:43036722

[29] F. Baratto, S. Bailey, and S. Tavoularis, "Measurements of frequencies and spatial correlations of coherent structures in rod bundle flows," *Nucl. Eng. Des.*, vol. 236, no. 17, pp. 1830–1837, sep 2006. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0029549306001257

[30] W. Eifler and R. Nijsing, "Experimental investigation of velocity distribution and flow resistance in a triangular array of parallel rods," *Nucl. Eng. Des.*, vol. 5, pp. 22–42, 1967.

[31] T. Krauss, "Characteristics of turbulent velocity and temperature in a wall channel of a heated rod bundle," *Exp. Therm. Fluid Sci.*, vol. 12, no. 1, pp. 75–86, jan 1996. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/0894177795000763

[32] T. Krauss and L. Meyer, "Experimental investigation of turbulent transport of momentum and energy in a heated rod bundle," *Nucl. Eng. Des.*, vol. 180, pp. 185 – 206, 1998.

[33] S. V. Moller, "Single-Phase Turbulent Mixing in Rod Bundles," *Exp. Therm. Fluid Sci.*, pp. 26–33, 1992.

[34] L. Meyer, "From discovery to recognition of periodic large scale vortices in rod bundles as source of natural mixing between subchannels-A review," *Nucl. Eng. Des.*, vol. 240, no. 6, pp. 1575–1588, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.nucengdes.2010.03.014

[35] D. Chang and S. Tavoularis, "Convective Heat Transfer in Turbulent Flow Near a Gap," *J. Heat Transfer*, vol. 128, no. 7, p. 701, 2006. [Online]. Available: http://link.aip.org/link/JHTRAO/v128/i7/p701/s1{&}Agg=doi

[36] ——, "Numerical simulation of turbulent flow in a 37-rod bundle," *Nucl. Eng. Des.*, vol. 237, no. 6, pp. 575–590, mar 2007. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0029549306004705

[37] D. Home, G. Arvanitis, M. Lightstone, and M. Hamed, "Simulation of flow pulsations in a twin rectangular sub-channel geometry using unsteady Reynolds Averaged NavierStokes modelling," *Nucl. Eng. Des.*, vol. 239, no. 12, pp. 2964–2980, dec 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0029549309004403

[38] J. Derksen, "Simulations of lateral mixing in cross-channel flow," *Comput. Fluids*, vol. 39, no. 6, pp. 1058–1069, jun 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0045793010000162

[39] D. Home and M. F. Lightstone, "Numerical investigation of quasi-periodic flow and vortex structure in a twin rectangular subchannel geometry using detached eddy simulation," *Nucl. Eng. Des.*, vol. 270, pp. 1–20, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.nucengdes.2014.01.002

[40] E. Merzari, H. Ninokata, A. Mahmood, and M. Rohde, "Proper orthogonal decomposition of the flow in geometries containing a narrow gap," *Theor. Comput. Fluid Dyn.*, vol. 23, no. 5, pp. 333–351, sep 2009. [Online]. Available: http://www.springerlink.com/index/10.1007/s00162-009-0152-3

[41] J. Ko, D. Lucor, and P. Sagaut, "Sensitivity of two-dimensional spatially developing mixing layers with respect to uncertain inflow conditions," *Phys. Fluids*, vol. 20, no. 7, p. 077102, 2008. [Online]. Available: http://link.aip.org/link/PHFLE6/v20/i7/p077102/s1{&}Agg=doi

[42] S. A. Mäkiharju, S. L. Ceccio, J. R. Buchanan, A. G. Mychkovsky, K. J. Hogan, and K. T. Lowe, "Experimental Characterization of Interchannel Mixing Through a Narrow Gap," in *Int. Top. Meet. Nucl. React. Therm. Hydraul.*, 2015, pp. 4394–4409.

CHAPTER 2

Experimental Setup for X-ray Computed Tomography System

2.1 Introduction: Computed Tomography Using A Scanning X-Ray Source

A new X-ray tomography system is being developed at the University of Michigan for the investigation of multiphase flows. It is called the University of Michigan Scanning Electron Beam X-ray Tomography (SEBXT) system. This Computed Tomography (CT) system uses multiple X-ray projection images, in which we can extract the attenuation ratio taken between the source and a detector array as the X-rays pass through a measurement domain. The projected images are collected for various angles within a volume of the domain, and they are combined to form a reconstructed map of the attenuation distribution within the measurement volume.

The SEBXT creates X-rays through the impingement of a focused electron beam on a tungsten target. It differs from a conventional X-ray CT system in that the deflected electron beam (through the application of an electro-magnetic field) is used to generate X-rays at specific positions on a stationary target. This is in contrast to traditional systems where the X-ray source and detectors may be physically rotated around the domain in order to collect a wide range of projection angles. Because the electron beam can be rapidly moved along the target, the sweep speed of the SEBXT can collect projections much faster than a conventional CT (like the ones used for cardiac CT in medical application).

Fig. 2.1 shows a conceptual drawing of the SEBXT system with a single detector. Here the left vertical rectangular box represents the vacuum chamber of the electron beam gun, the gray object at the center is the object that rests in the measurement domain, and the circuit on the right represents a single detector of a detector array. Note that objects in Fig. 2.1 are not to scale. Once the X-ray beam is generated by the source, a fraction of the

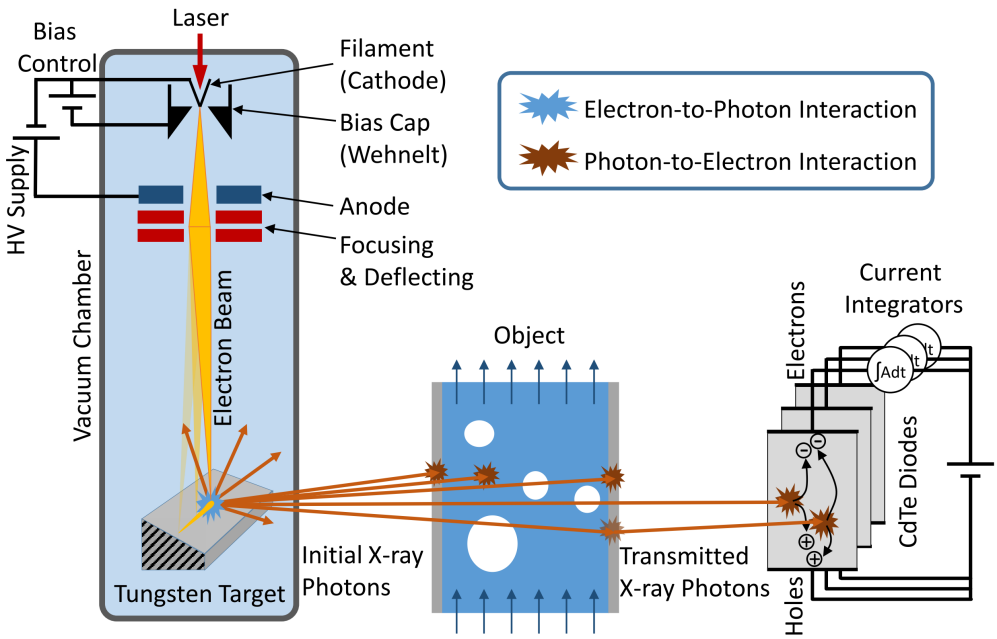


Figure 2.1: A schematic representation of the SEBXT-system showing the electron beam gun, the target and resulting X-rays, the X-rays passing through the measurement domain, and a single X-ray detector.

emitted photons pass through the object along a straight line from the source to a specific detector. In the SEBXT system, an array of detectors will be employed. These detectors will collect a fraction of the transmitted X-ray photons for multiple source positions at various angles, as the EB sweeps along the Tungsten target. We therefore collect one set of projections for each position of the X-ray source. A complete set of those projection data is called the sinogram. The amount of attenuation along a given path depends on the materials composition, density and thickness that the X-rays have to pass through, and this enables the reconstruction of a material map of the object in the measurement domain. The sinogram is then used to reconstruct the attenuation distribution in the domain and, hence, the material or phase distribution. The SEBXT is designed to measure the material distribution on a cross-section plane through the domain.

2.2 University of Michigan Scanning Electron Beam X-ray Tomography System (SEBXT)

The Scanning Electron Beam X-ray Tomography (SEBXT) system is comprised of a 150 kV, 20 kW scanning electron beam gun, a vacuum chamber with linear and semi-circular

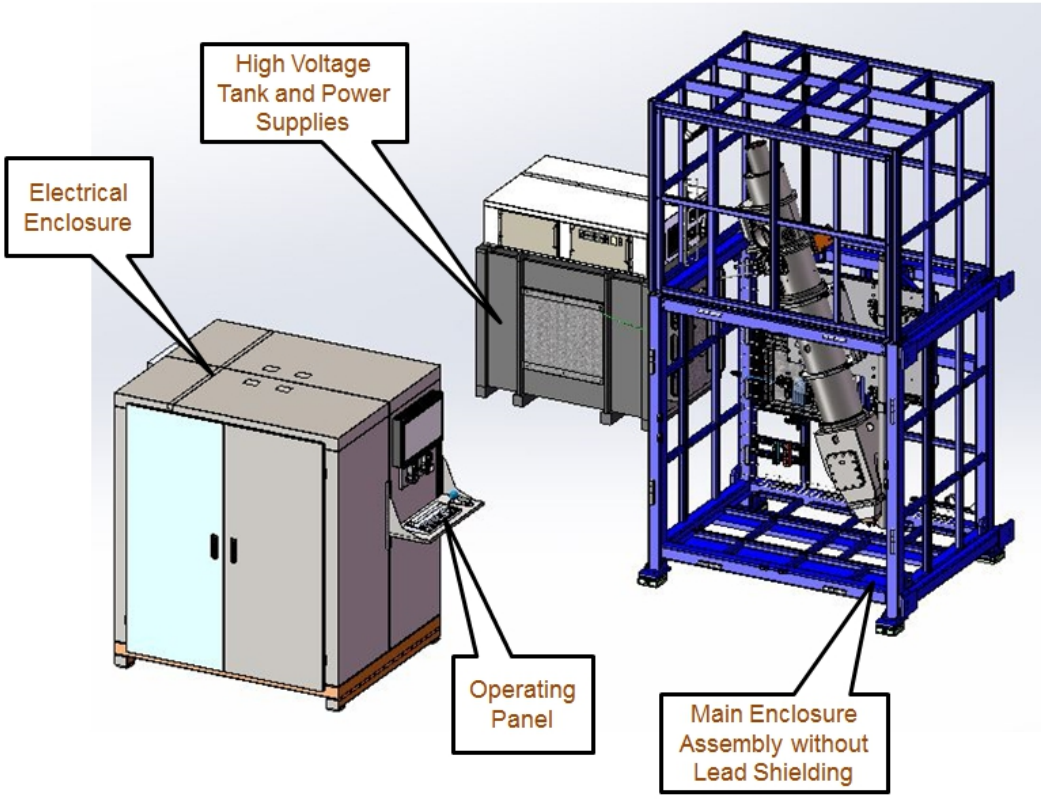


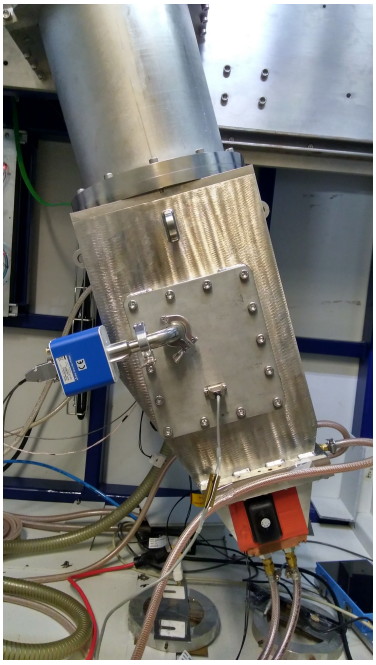
Figure 2.2: Major system components of the SEBXT. The electrical enclose houses the control system and the laser used to heat the cathode; the high voltage tank and power supplies provide the potential between the cathode and the tungsten target in the electron beam gun; the main enclosure contains the gun and the detector, and has a lead shielding around it which is omitted in this figure to show the interior.

tungsten targets, and a detector sub-system. A rendering of the major system components are shown in Fig. 2.2. These include the high-voltage tank that supplies the voltage bias between the filament (cathode) and the anode, the electronics enclosure, and the Electron Beam (EB) gun assembly. The cathode, bias cap, anode, and focusing/deflecting coils are enclosed in the top of the EB gun assembly, and the target is at the bottom of the assembly. The gun and chamber are maintained under high vacuum by a system of roughing and turbo-pumps.

2.2.1 The electron beam X-ray source

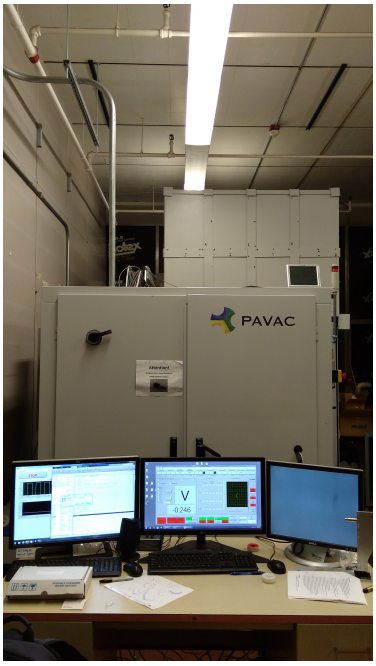
X-rays are produced with a 150 kV 20 kW electron beam gun system. To produce the electron beam, a tungsten filament is heated by a 50 W laser to emit electrons by the thermionic emission process [1, 2]. Laser heating of the filament is a novel feature of this system to

(a)



(b)

(c)



(d)

Figure 2.3: Photographs of the University of Michigan SEBXT. (a) upper part of the electron beam gun; (b) lower part of the electron beam gun; (c) high voltage supply; (d) electrical enclosure and operating panel.

minimize the unwanted beam deflection induced by the electricity to heat the filament. When not controlled by additional electro-magnetic force, the amount of emitted electrons from the heated tungsten filament is greater than the desired value. Therefore, a bias control circuit is utilized to block the over-production of electrons by applying a maximum of 5 kV bias between the cathode and the bias cap (formally known as the Wehnelt cylinder). Application of a bias voltage to the anode can either block or accelerate the passage of electrons from the filament to the target.

During beam operation, emitted electrons pass the bias cap aperture and are accelerated by the voltage bias between the cathode (filament) and the anode. Subsequently, they are focused and deflected by the focusing and deflecting coils. The focused electron beam then impinges on the tungsten target to generate X-rays. When the accelerated electrons interact with the tungsten target, they can produce a distribution of high energy photons. If an electron has energy higher than 10 keV, it can generate a "hard X-ray" photon that can effectively penetrate relatively highly attenuating matters, such as titanium.The chamber and electron beam gun assembly must be maintained in a high vacuum, less than 10^{-6} mbar (absolute). Without the high vacuum, destructive arcing can be produced across potential differences above 1.5 kV. Also, gas molecules in the air can scatter the electrons of the beam, lowering the energy of electrons before they strike the target and changing the focus of the beam.

Fig. 2.4 shows the cutout view zoomed-in at the bottom part of the chamber where the tungsten target is located. The circular tungsten target has inner diameter of 76 mm and 102 mm outer diameter, and the surface emitting the X-rays is at a 30-degree angle with respect to the measurement plane. Both tungsten targets are mounted on a water-cooled copper base to more effectively remove heat. It should be noted that the geometry was designed to accommodate replacing the targets with piecewise linear rotating tungsten targets, which is why the emitting surface was located such a distance from the copper base.

The electron beam that originates from the top of the gun is guided through slits in a molybdenum screen, which is used to prevent the beam from impinging on the aluminum or copper components of the base. When the electron beam passes through the slit it impinges on the tungsten target, generating a poly-energetic beam of X-rays. The highest energy of X-ray photons produced by the electron beam is limited by the maximum energy of the electrons.

Figure 2.4: Cutout showing the bottom part of the chamber with the molybdenum safety screen, tungsten targets, and the water-cooled copper base. Note that the aluminum is also cooled underneath the o-ring to prevent o-ring failure. This system has two tungsten targets, linear and arc. The linear target was included simply to enable testing of the damage threshold of the tungsten material during the safety testing to generate the X-rays, and for other testing of basic ideas.

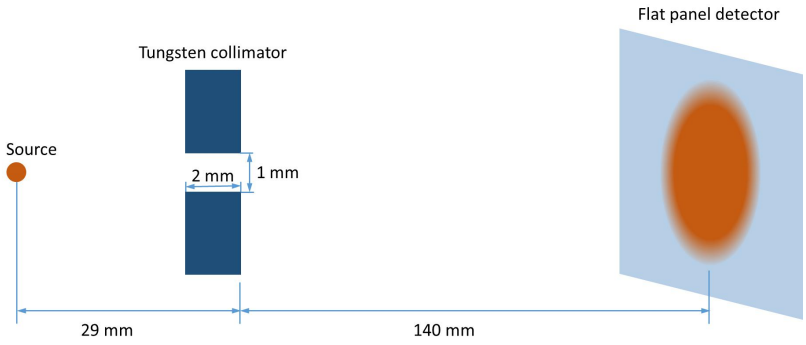


Figure 2.5: Conceptual drawing of the focal spot measurement configuration. The focal spot size of the beam is measured by taking the projected image of the tungsten collimator with the hole diameter 1 mm and the thickness 2 mm.

2.2.2 The electron beam focusing on the arc tungsten target

Along with the detector physical dimension, one of the important features that determines the resolution of the reconstructed image is the beam focal spot size on the tungsten target (i.e., the size of the source and its intensity distribution). To check the beam focal spot size, a 2 mm thick tungsten collimator with 1 mm diameter aperture is placed in front of the source, 29 mm from the approximate focal spot location, and 140 mm from the detector, with estimated uncertainties in relative positions less than 1 mm. The conceptual drawing of the focal spot measurement configuration is shown in Fig. 2.5.

A total of 30 projection images of the tungsten collimator are obtained by Dexela 1207CL flat panel detector, and a profile of the collimator aperture is taken from the ensemble average of the 30 images. The PerkinElmer Dexela 1207 CL is a 2-D flat panel X-ray detector with DRZ-standard scintillator film, and a CMOS panel with a total of 1536×864 pixels with 75×75 $\mu\mathrm{m}^2$ face area. This 2D flat panel detector was not used for the CT reconstruction. Fig. 2.6 shows the profile of the beam focal spot with various focus currents.

The beam focal spot sizes are computed by the deconvolution process based on the projected image. Let g be the measured projected image of the collimator aperture, and f the projected image of the collimator aperture with the ideally focused beam (infinitesimal focal spot). Because the aperture is much smaller than the distance between the source and the collimator, we assume the intensity of the initial X-ray beam is constant to simplify f. Let h be the ideal projected image of the beam profile through the ideal pin hole, i.e. a pin hole with infinitesimally small hole and thickness. We want h be unitless, with the integral

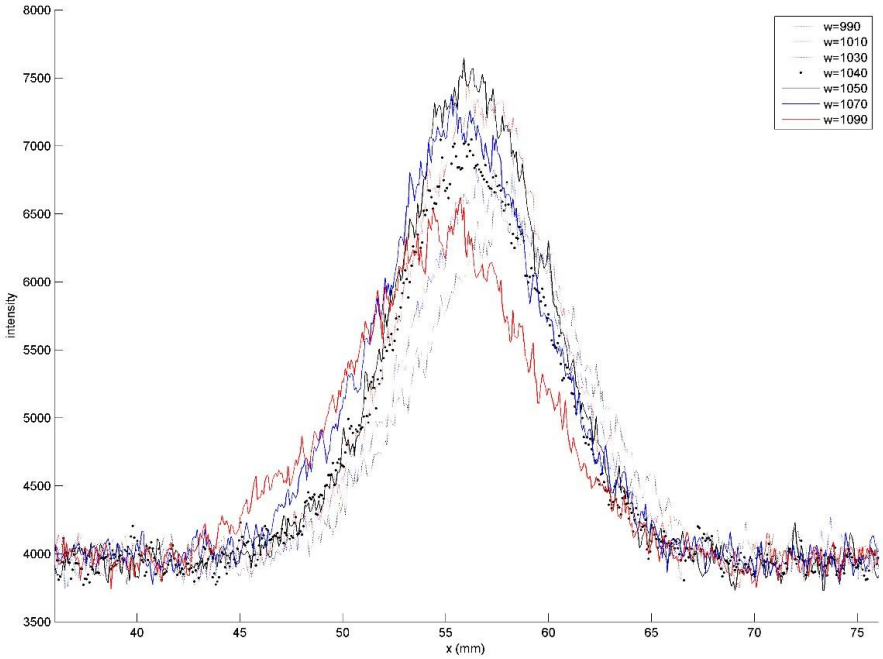


Figure 2.6: Comparisons of the profiles of the projected collimator with 1 mm aperture for beam focus currents ranging from 990 mA to 1090 mA.

of h being unity.

$$\int_{x=-\infty}^{\infty} h(x)\mathrm{d}x = 1. \tag{2.1}$$

The flat panel detector does not specify the unit or the gain of the sensor response. However, we can make h satisfying Eqn. (2.1) by adjusting the intensity of f so that $\int_x g(x)\mathrm{d}x = \int_x (f)\mathrm{d}x$.

Assume the X-ray collimating is a linear shift-invariant system, g can be expressed as a convolution between h and f

$$g(x) = (h * f)(x) \tag{2.2}$$

where $*$ denotes convolution in terms of x.

Once discretized, the convolution can be re-written as a matrix multiplication as

$$\mathbf{g} = \mathbf{F}\,\mathbf{h}, \tag{2.3}$$

where $\mathbf{F}$ is the convolution matrix of f. To find h, we need to invert the linear system given in (2.3). However, (2.3) is not directly invertible if the size of $\mathbf{h}$ is not equal to the size of $\mathbf{g}$. Furthermore, naive deconvolution by matrix inversion amplifies noise significantly. Thus,

we use the regularized least-square method. Since the beam profile is generally smooth, we use the Tikhonov regularizer.

$$\hat{\mathbf{h}} = \underset{h_i>0}{\arg\min} \; ||\mathbf{F}\,\mathbf{h} - \mathbf{g}||_2^2 + \alpha||\mathbf{C}\,\mathbf{h}||_2^2, \tag{2.4}$$

where α is the regularizer strength parameter, and $\mathbf{C}$ is the first order finite difference matrix. α is set to be 10^{11} in this estimation. (We cannot guarantee that $\alpha = 10^{11}$ is the optimal value. We just choose the value so that the image looks neither too smooth nor too noisy.) Since the X-ray beam cannot have negative value, we employ a iterative constrained gradient method to solve (2.4).

$$\mathbf{h}^{(n+1)} = \mathcal{P}_0\left[\mathbf{h}^{(n)} - \nu\Big(\mathbf{F}^{\mathrm{T}}\left(\mathbf{F}\,\mathbf{h}^{(n)} - \mathbf{g}\right) + \alpha\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{h}^{(n)}\Big)\right], \tag{2.5}$$

where $\mathcal{P}_0$ is the element-wise projection operator defined as $\mathcal{P}_0[t] = \max[t, 0]$. Once $\mathbf{h}$ is computed, we can scale down the estimated projected profile to get the actual beam profile according to the distance ratio. Fig. 2.7 shows the measured projected image profile obtained by the Dexela 1207 flat panel detector and the simulated projected image profile with ideal infinitesimal focal spot beam. Fig. 2.8 shows the estimated beam profile.

Based on the experiments, best full-width at half-maximum (FWHM) of the focal spot diameter was estimated to be 1.18 mm at the focus current of 1050 mA. Later, we found that the actual focal spot size is possibly much wider than the estimated value, approximately 4 mm FWHM, by comparing the simulation and experiment sinogram data. The possible reason for this large focal spot is the electron beam machine malfunctioning due to the outgassing. We found that the electron beam machine took several seconds after the beam is on to initially set up the focal spot. The overheated cathode emits gas molecules, which may results in collisions between the emitted electrons and the gas molecules, causing the electrons to deflect. Another pumping system has been added to the bottom of the gun to reduce the gas molecules. When the focal spot size is larger than the size of the collimator aperture, the finite thickness greatly affects the projected image, makes the projection non-linear and shift-variant. Fig. 2.9 shows the examples showing the difference of the projected images of the collimator aperture for ideal, small, and large focal spots.

2.2.3 The electron beam positioning on the arc tungsten target

The location of the electron beam is controlled by the electro-magnetic forces induced by deflecting coils, which steer the beam laterally in a plane. The voltage readings in the control panel do not directly indicate the location of the beam, hence a voltage-to-

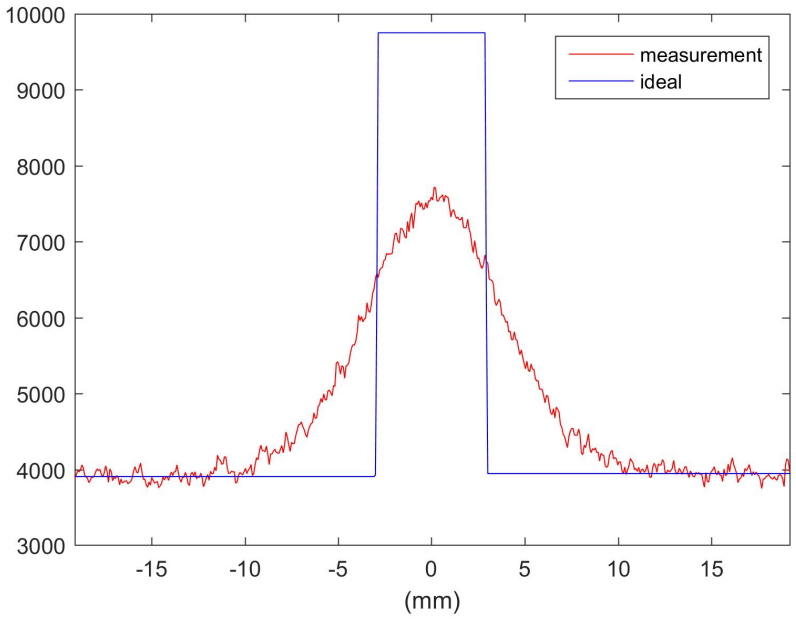


Figure 2.7: Measured projected image profile of the collimator aperture taken at the row with maximum intensity and simulated projected image profile of the collimator aperture with ideal infinitesimal focal spot beam. The measured profile has an unknown unit. We adjusted the intensity of the ideal beam profile so that the ideal profile has the same area under the curve as the measured profile.

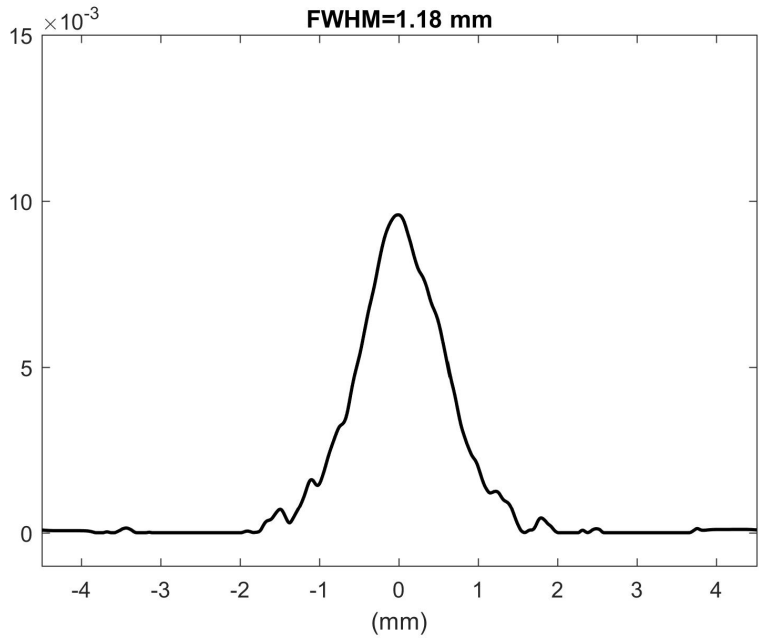


Figure 2.8: Estimated profile of the X-ray beam at the focusing current 1050 mA. Beam profile is adjusted so that the area under the profile is normalized to one.

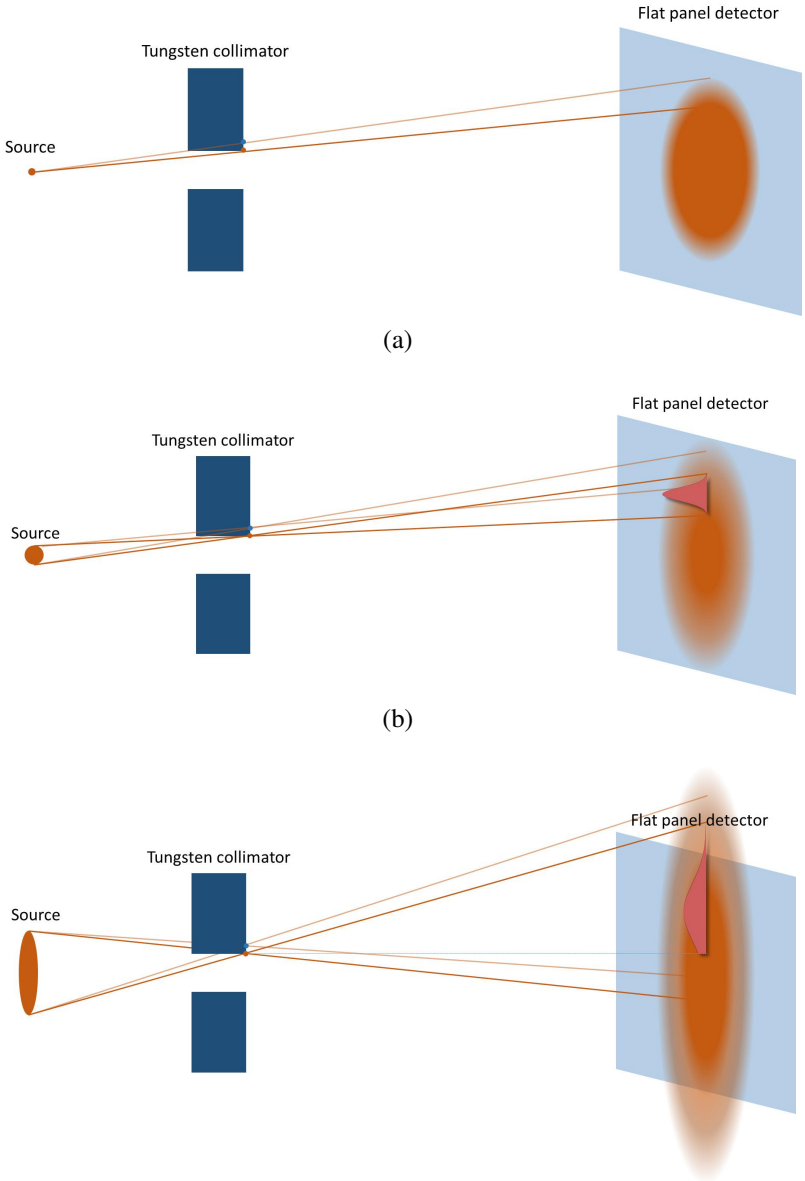


Figure 2.9: A schematic drawing of the focal spot size measurement experiment; (a) ideal infinitesimal focal spot, (b) small focal spot, and (c) large focal spot. Images are not in proportion to emphasize the effect of the focal spot size. A 2 mm thick tungsten collimator with 1 mm diameter aperture is placed 29 mm away from the approximate focal spot location, and 140 mm away from the detector, with estimated uncertainties in relative positions less than 1 mm. When the focal spot diameter is larger than the collimator aperture diameter, the projected image becomes significantly nonlinear, which violates the shift-invariant assumption.

coordinate correction factor must be obtained based on the projection images of a known object. The amount of deflection per voltage applied to coils is a function of the voltage bias between the cathode and anode. Therefore, to avoid damage to system components, one must first carefully calibrate the voltage at minimal beam current, around 0.3 mA, only after beam location is known increase the beam current to the operational level. (That is, the required deflection coil voltage is taken to be independent of the beam current, and is presumably linearly dependent on the bias voltage.)

A calibration phantom (Phantom 0) was used to determine the location of the electron beam on the semi-circular tungsten target. The calibration phantom consists of seven pairs of 1/32 inch (0.79 mm) diameter tungsten carbide balls that are placed in a 3 inch (76.2 mm) diameter acrylic disk along the straight lines that pass through the center of the phantom drawn at every 25 degrees, in a plane where the source and the detector are aligned. Refer to Appendix A for the detailed CAD drawing.

Fig. 2.10 shows the setup used to relate the deflection control voltage to the beam position. To produce the projection image, the calibration phantom was placed at the center of the domain, and the X-ray beam was generated from a spot on the tungsten target. Because the X-ray attenuation of the tungsten is much higher than that of the acrylic, the tungsten balls cast shadows on a flat panel X-ray detector. By examining the location of the shadows, we can determine the origin of the X-ray source. The flat detector is rotated to form the desired angle with respect to the line formed by a pair of tungsten balls. By changing the U and V voltages, the beam is deflected to a new position on the tungsten target. This then produces a different image on the flat panel. The voltages applied on the deflecting coils are adjusted until the shadows of two paired tungsten balls become concentric.

We recorded the X-ray projection image of the calibration phantom using the flat panel detector Dexela 1207 CL. Fig. 2.11 shows the image taken with the Dexela 1207 CL panel detector at four different angles, (a) 0 degree, (b) 25 degrees, (c) 50 degrees, and (d) 58.68 degrees, when the corresponding pair of tungsten balls are aligned with the X-ray source spot. The red circle in Fig. 2.11 indicates a pair of the tungsten balls to be aligned with the X-ray source spot in each configuration. The entire table for the deflection control voltage input for seven different angles is shown in Table 2.1. The beam location based on above described localization is assumed known within 1 mm.

We obtain the calibration matrix by least square fitting based on the angular position-voltage data shown in Table 2.1. As a result, the input deflection voltage U and V for any given physical coordinate x and y can be obtained from the following affine linear relation:

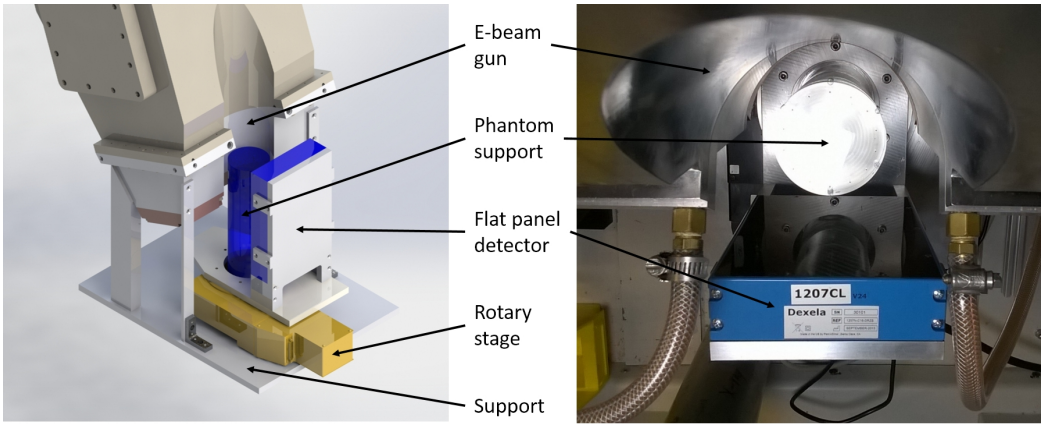


Figure 2.10: 3D drawing of the isometric view (left) and photograph of the top view (right) of the EB gun, target, domain, and flat panel detector configuration used for the beam positioning process

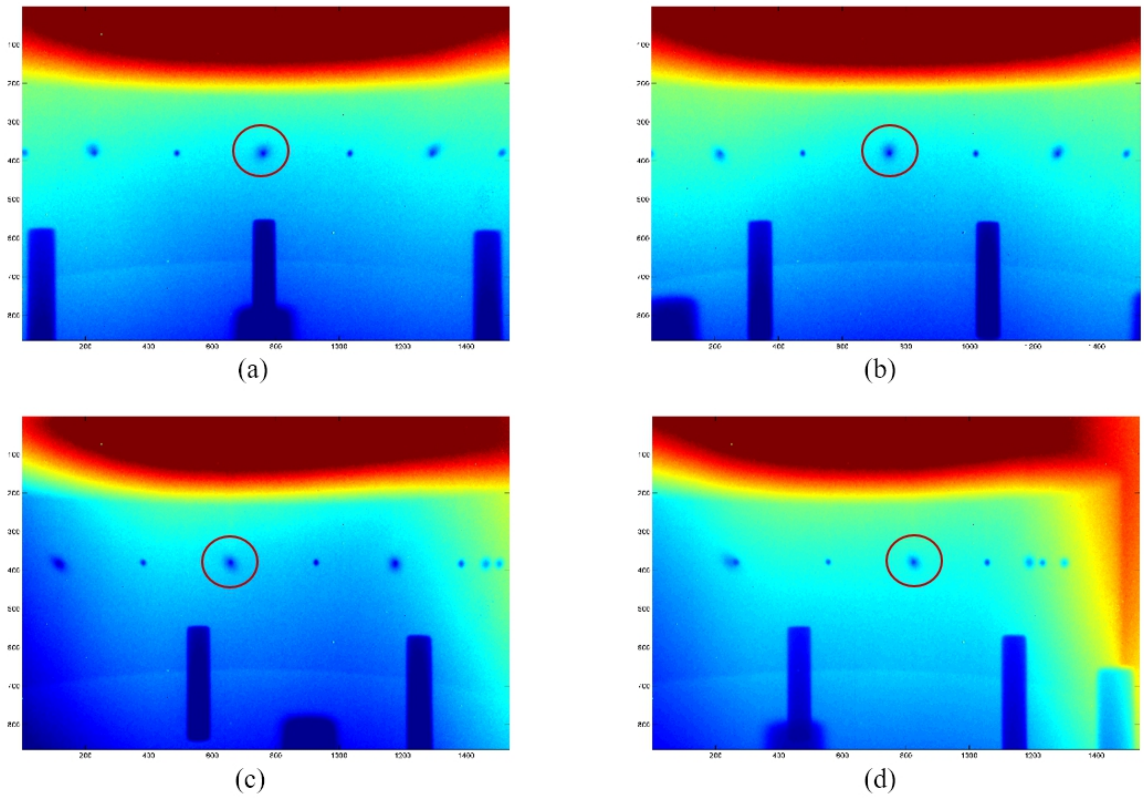


Figure 2.11: Alignment calibration images obtained by the Dexela 1207 CL flat panel X-ray detector. Images are shown here for alignment angles of (a) 0 degrees (b) 25 degrees (c) 50 degrees (d) 58.68 degrees. Red circles indicate where two tungsten balls are aligned. (Note: While the balls were located along the lines intersecting the center of the phantom every 25 degrees, other angles (e.g. 58.68 degrees as shown in (d)) could be formed when imaging balls along a line that does not intersects the center of the phantom.)

$$\begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} 0.6277 & -0.0191 & 1.6038 \\ -0.0292 & -0.6002 & -0.9018 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \tag{2.6}$$

2.2.4 The resulting X-ray spectrum

The spectrum of the X-ray source was measured using the Amptek XR-100T-CdTe detector. Due to the inability of the detector to cope with as high photon flux as the arc would be exposed to, greater than $O(10^6)$ counts per second (cps), we operate the source at 115 kV and 0.5 mA beam current and applied a tungsten collimator with a 100 μm diameter aperture on the detector to reduce the photon flux. The photon flux is then corrected for 100 mA beam with pixel face size 0.7×0.7 mm^2 detector, which is the actual condition for the SEBXT being operated in the future. The dead time ratio of the measured spectrum was 13.5 %, which means that the 13.5 % of total number of the photons were rejected by the pile-up detection. we compensate the dead time by multiplying the entire spectrum by 1/(1 - deadtime ratio$/100$). Note that we only applied this dead time correction to measure the spectrum. For the CT reconstruction, we acquired data without the pulse pile-up rejection, as we only need the integral of the energy, not the spectrum.

The measured spectrum was compared with the simulated spectrum produced using the X-ray source spectrum simulation software SpekCalc [3]. Fig. 2.12 shows the spectrum measured, and a comparison to the simulated spectrum. We can see that the measured spectrum has less low energy photons and more high energy photons. This might be due to the effect of pulse pile-ups. Note that not all pile-up can be removed by the circuit and some fast pile-ups are recorded as true energy of photons. The results show that the measurement spectrum is deformed by a Spectral Response Function (SRF), and it is important for the reconstruction algorithm to take the effect of the SRF into consideration in order to have an accurate quantitative result. The spectral response function will be further discussed in the reconstruction part in §3.2.4. Note that we recorded the measured spectrum data in Fig. 2.12 without a steel filter, which was used to obtain the actual sinogram data to follow. A tungsten collimator with a 100 μm diameter aperture was used instead, which is not usable in CT reconstruction due to the wide variation of incident angles.

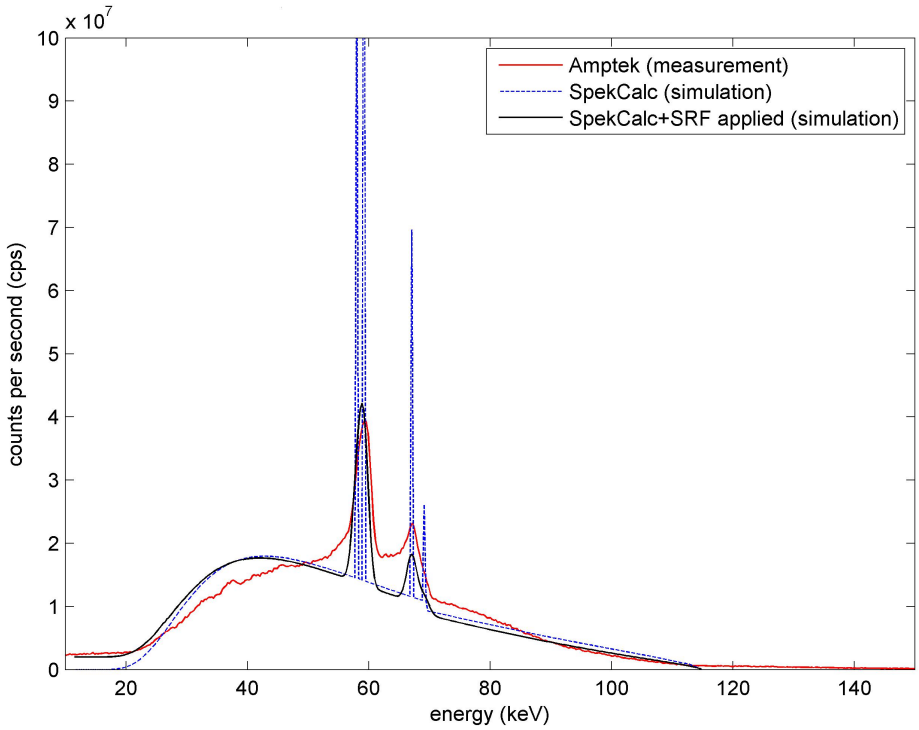


Figure 2.12: A comparison of the measured and computed photon energy spectrum for 115 kV electrons impinging on a tungsten target. The calculated spectrum was obtained using the simulation software (SpekCalc). The measured spectrum was recorded with the 100 μm diameter tungsten collimator, and the distance between the source and the detector was 168.69 mm. The results show that the measurement spectrum was deformed by the spectral response function, and that it is important for the reconstruction algorithm to take the effect of the spectral response function into consideration in order to have an accurate quantitative result. The shift of the Compton peak to the higher energy shows the occurence of pulse pile-ups.

Table 2.1: Voltages applied to the deflection coils at each aligning angle position with the electron beam cathode voltage set to 115 kV. The voltage input of the beam control software has the accuracy of three decimal places.

Angle (degree)	U (volt)	V (volt)
-58.68	0.510	0.994
-50.00	0.250	0.770
-25.00	-0.355	0.090
0.00	-0.59	-0.800
25.00	-0.428	-1.705
50.00	0.145	-2.450
58.68	0.406	-2.635

2.2.5 The X-ray detector array

A detector array has been developed for use on the SEBXT, designed and manufactured by the manufacturer Integrated Detector Electronics AS (IDEAS). The array consists of eight linear modules covering approximately 180 degrees. Each module has 32 pixels horizontally and 4 pixels vertically with a pixel pitch of 0.8 mm. Each pixel has a CZT semiconductor with the face size of 0.7×0.7 mm^2 connected to the charge-integrating circuit (current mode). Maximum design frame rate is 125 kfps for line scanning, 31.25 kfps for area scanning. Exposure time can be adjusted independent of the frame rate, with the minimum of 5 μs. For the present experiments, however, the X-ray arc detector was not available. Therefore, we created projection sets and the resulting sinograms by traversing the single point detector. An Amptek XR-100T-CdTe spectrometer with a PX5 multichannel analyzer (MCA) was utilized as a solitary detector that could be positioned along the location of the notional detector arc, and hence the data generated by this one detector at the locations of the detectors in the notional arc mimicked the data such an arc would yield. The detector is a thermo-electrically cooled X-ray and gamma-ray detector using a $3 \times 3 \times 1$ mm^3 CdTe diode, which can resolve the energy of the photon with a resolution of 1.2 keV for 122 keV 57 Co source (with actual resolution depending on photon energy, flux and detector settings).

The SEBXT system is designed to include an array of CZT energy-integrating detectors capable of $O(100)$kHz simultaneous data acquisition. With 100 projections per slice, this will enable $O(1)$kHz CT of moving objects or flow. Presently, only a single pixel spectrometer is mounted on the motorized rotary stage (RV160 PP) and rotated to mimic the arc detector array of radius 8 cm. Due to the inability of the detector to cope with as high photon flux as the arc would be exposed to, greater than $O(10^6)$ counts per second (cps), the source was operated at 115 kV and 1.0 mA beam current ($1/130^{\text{th}}$ of maximum) to reduce the photon flux. Additionally, a 0.5 inch (12.7 mm) thick steel filter was installed in front of the spectrometer to reduce the photon flux from $O(10^8)$ cps to $O(10^6)$ cps. Fig. 2.13 shows a photograph of the setup with the single pixel spectrometer with the steel filter mounted on the rotary stage.

Fig. 2.14 shows the input counts/output counts ratio of an Amptek Silicon Drift Detector (SDD) with the PX5 analyzer for different peaking time τ_p of the pulse generation, illustrating the nonlinear response due to pile-ups (multiple photons arriving at the detector too close in time to be distinguished), and the energy resolution of the same detector. (The type of the semiconductor used for this analysis is different from our CdTe detector. As CdTe has better efficiency than SDD, we expect that the CdTe detector will have more pile-up issues while it has better resolution.)

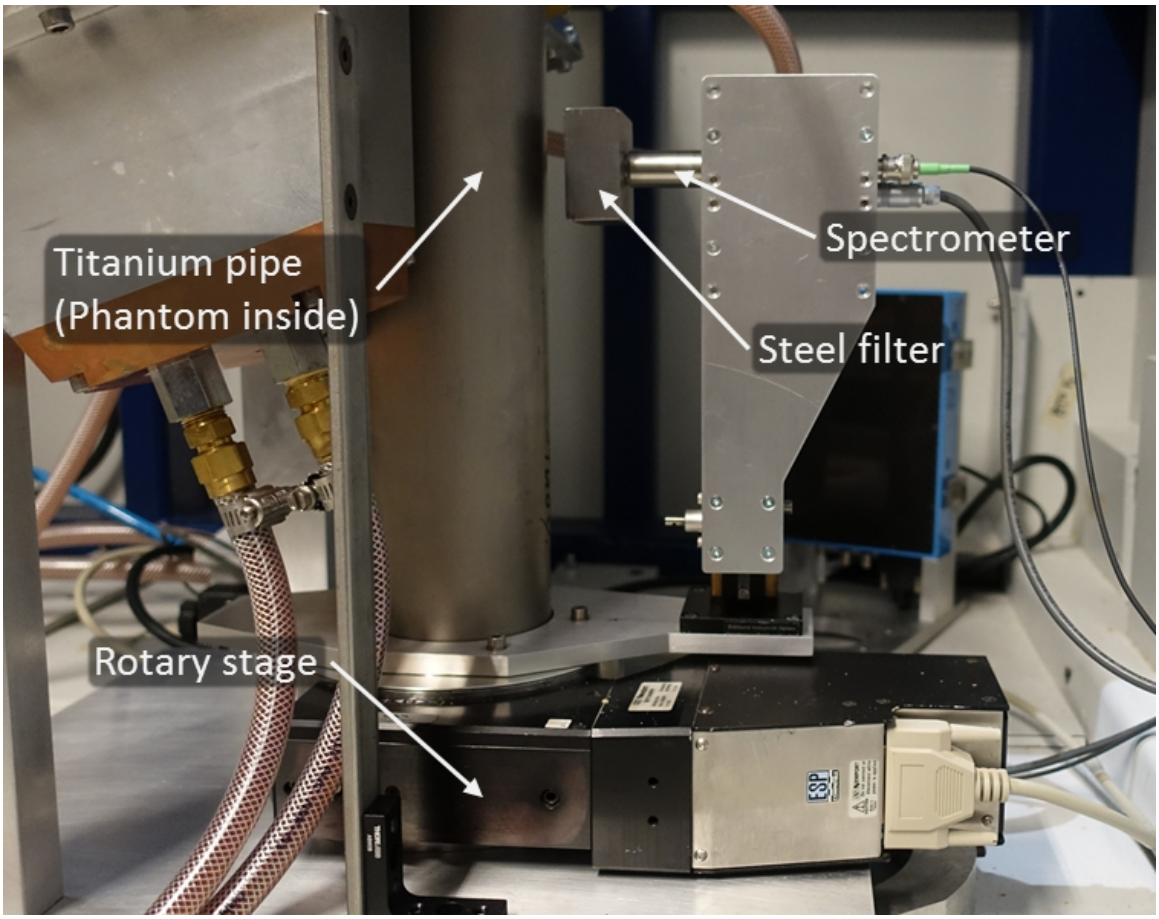


Figure 2.13: Single pixel spectrometer setup. A single pixel spectrometer is mounted on the motorized rotary stage (RV160 PP) and rotated to mimic the arc detector array of radius 8 cm. The tip of the spectrometer is covered by a 0.5 inch (12.7 mm) thick steel filter to reduce the X-ray photon flux down to $O(10^6)$ cps.

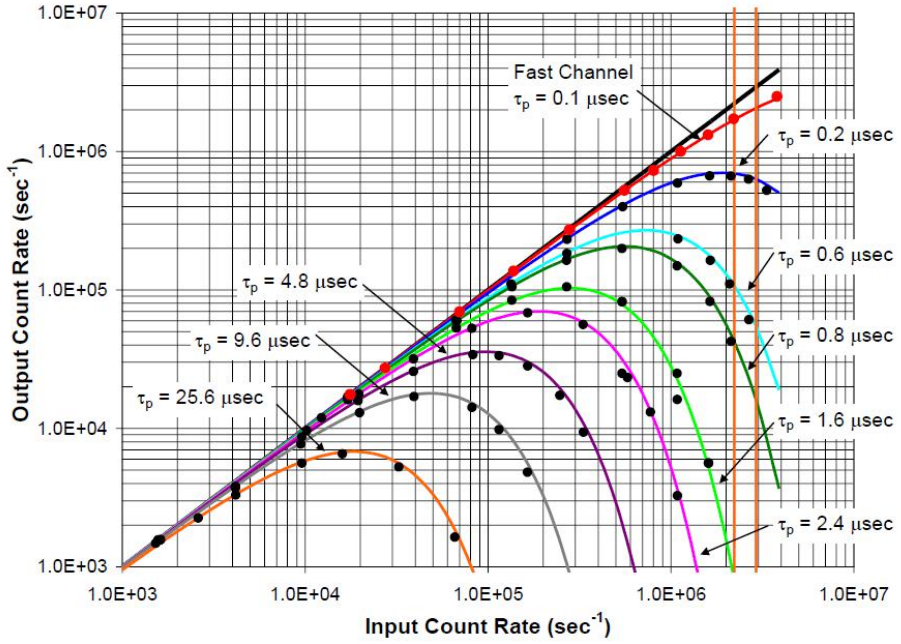


Figure 2.14: The input counts/output counts ratio of an Amptek Silicon Drift Detector (similar, but not identical to the one utilized) with the PX5 multi-channel analyzer for different peaking time τ_p, illustrating the nonlinear response due to the effect of pile-ups (multiple photons arriving at the detector too close in time to be distinguished). (Data from http://amptek.com/products/xr-100sdd-silicon-drift-detector/)

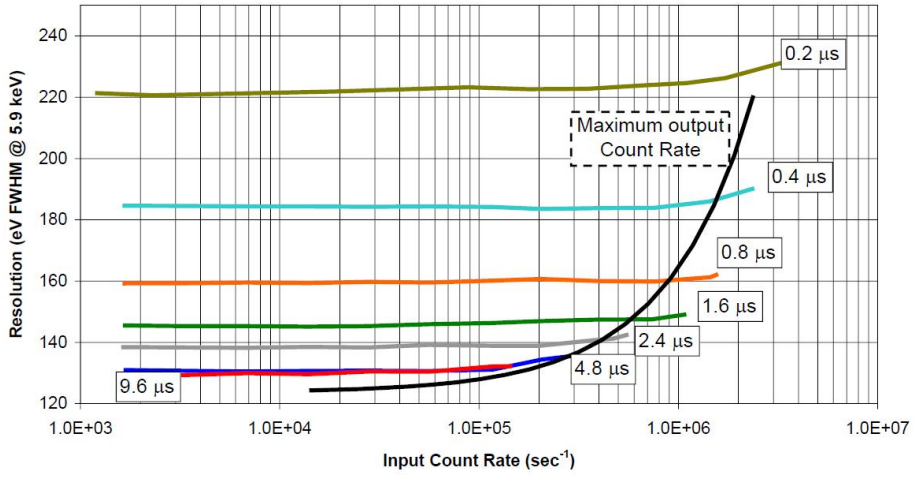


Figure 2.15: The energy resolution vs. input counts for different peaking time τ_p. As the peaking time is reduced, so is the energy resolution of the detector. However, lower peaking time is required to reduce the chance of pulse pile-ups. (Data from http://amptek.com/products/xr-100sdd-silicon-drift-detector/)

2.3 Description of the test phantoms

We created a total of three phantoms in order to evaluate the SEBXT system performance and to develop and test reconstruction algorithm. All phantoms are manufactured with a typical tolerance of ± 0.0005 inches (12.7 μm).

The first, Phantom 0, is used for testing the electron beam localization. Phantom 0 is made of clear acrylic with seven pairs of 1/32 inch (0.79 mm) diameter tungsten carbide balls. Phantom 0 is designed to calibrate the beam positioning, thus quantitative attenuation data are not necessary. Phantom 1 and Phantom 2 are used to test the CT reconstruction algorithms. The two latter phantoms were made out of Plastic water^{TM} LR [4], a water equivalent plastic which mimics the attenuation coefficient of water from 15 keV to 8 MeV within 1%. Both the plastic Phantom 1 and Phantom 2 were tightly encircled by a 3-inch (76.2 mm) diameter SCH-10 titanium pipe. The outer diameter of the pipe is 3.5 inches (88.9 mm), the inner diameter is 3.26 inches (82.8 mm), and the wall thickness of the pipe is 0.12 inches (3.0 mm).

2.3.1 Phantom 0: Calibration phantom for electron beam positioning

The calibration phantom (Phantom 0) was used to determine the location of the electron beam. Explanation on how this phantom was utilized are included in §2.2.3. Fig. 2.16 shows a picture of Phantom 0. Detailed CAD drawing is

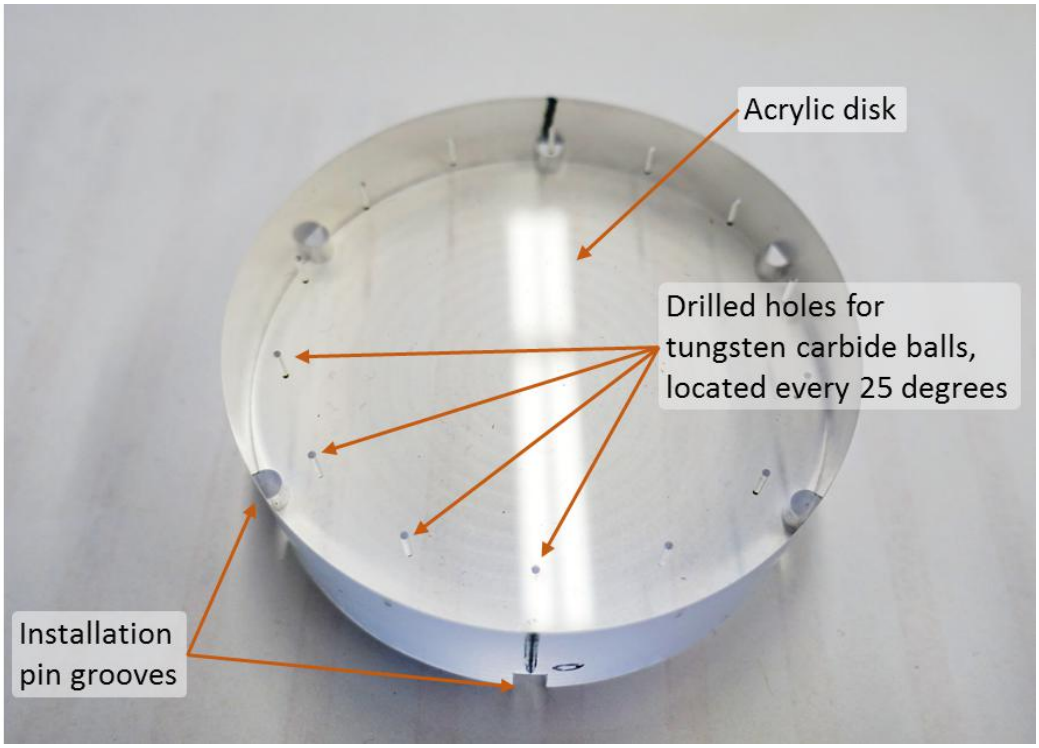


Figure 2.16: Picture of the calibration phantom, Phantom 0. This phantom incorporates seven pairs of 1/32 inch (0.79 mm) diameter tungsten carbide balls that are placed in a 3 inch (76.2 mm) diameter acrylic disk along radial lines that pass through the center of the phantom every 25 degrees. They determine when the X-ray source and detectors are aligned. Total of 14 tungsten carbide balls are used for seven different angles. The plastic component is encircled by a 3-inch SCH-10 titanium pipe. The outer diameter of the pipe is 3.5 inches (88.9 mm), the inner diameter is 3.26 inches (82.8 mm), and the wall thickness of the pipe is 0.12 inches (3.0 mm)

2.3.2 Phantom 1: Plastic rectangle in titanium pipe

To investigate the resolution of the reconstructed image with the current source-detector pair, a sinogram of a rectangular phantom that is identical with the one in [5] was recorded and reconstructed. The rectangle has dimensions of 18×22 mm^2. Fig. 2.17 shows a picture of Phantom 1.

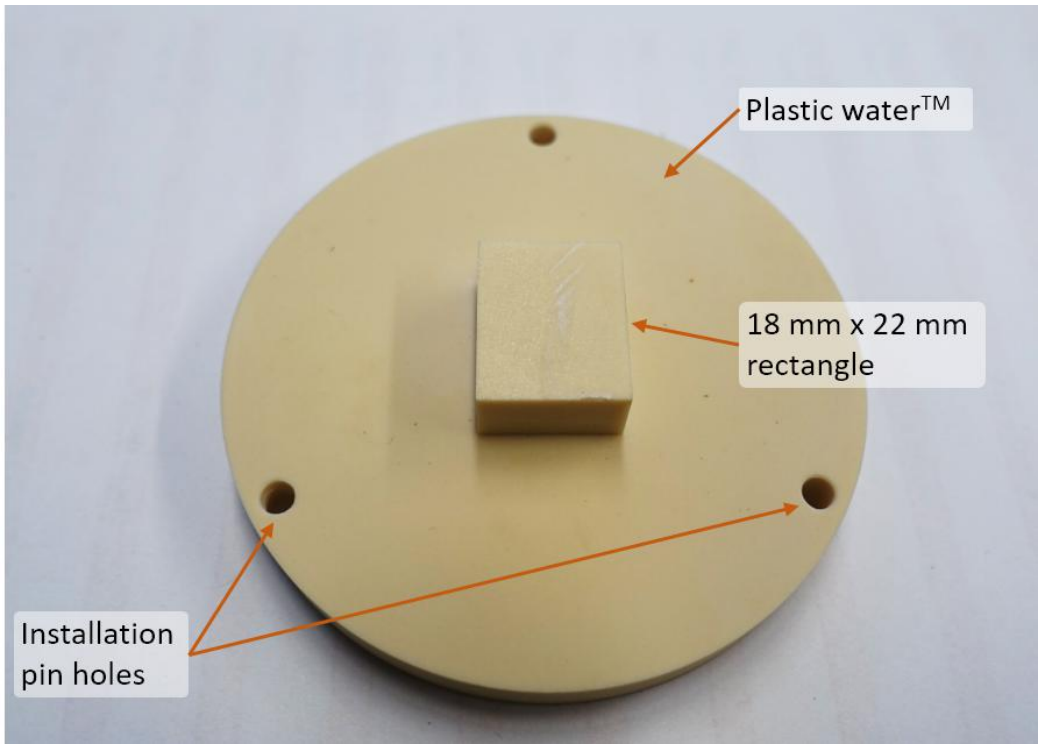


Figure 2.17: Picture of Phantom 1. Phantom 1 is a rectangular block placed on a circular disk substrate. The phantom was made of Plastic Water™ LR, a water equivalent plastic which mimics the attenuation coefficient of water from 15 keV to 8 MeV within 1%. The plastic component is encircled by a 3-inch SCH-10 titanium pipe. The outer diameter of the pipe is 3.5 inches (88.9 mm), the inner diameter is 3.26 inches (82.8 mm), and the wall thickness of the pipe is 0.12 inches (3.0 mm)

2.3.3 Phantom 2: "Bubbles in water" in a titanium pipe

Phantom 2 was designed to study the resolution of the SEBXT for circular objects, similar to gas bubbles in water. As it is desired to conserve the void fraction even when the SEBXT cannot resolve each bubble separately, this phantom contains regions of known void fraction with holes small enough that they were expected to be unresolvable. The phantom has circular openings of diameters ranging from 11.25 mm to 0.5 mm. All locations and sizes of the holes are summarized in Appendix B. Phantom 2 was made of Plastic Water^{TM} LR, a water equivalent plastic which mimics the attenuation coefficient of water from 15 keV to 8 MeV within 1%. The plastic component is encircled by a 3-inch (76.2 mm) diameter SCH-10 titanium pipe. Fig. 2.18 shows a picture of Phantom 2.

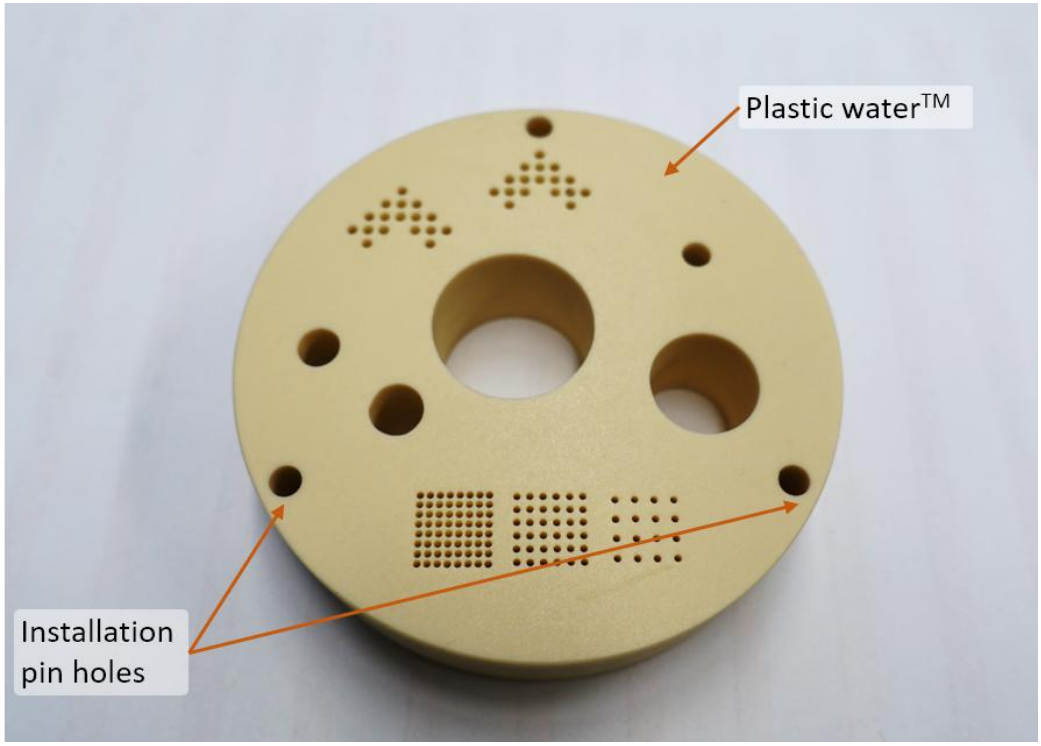


Figure 2.18: Picture of Phantom 2. The phantom has circular openings of diameters ranging from 1.125 cm to 0.05 cm. Locations and sizes of the holes are summarized in Appendix B. The phantom was made of Plastic Water^{TM} LR, a water equivalent plastic which mimics the attenuation coefficient of water from 15 keV to 8 MeV within 1%. The phantom is encircled by a 3-inch (76.2 mm) diameter SCH-10 titanium pipe. The outer diameter of the pipe is 3.5 inches (88.9 mm), the inner diameter is 3.26 inches (82.8 mm), and the wall thickness of the pipe is 0.12 inches (3.0 mm)

BIBLIOGRAPHY

[1] O. Richardson, *The emission of electricity from hot bodies*. London: Longmans, green and Co., 1916.

[2] J. Orloff, *Handbook of charged particle optics*. Boca Raton: CRC Press, 2009. [Online]. Available: https://mirlyn.lib.umich.edu/Record/005914013CN-QC372.2.D4H362009

[3] G. Poludniowski, G. Landry, F. DeBlois, P. M. Evans, and F. Verhaegen, "SpekCalc: a program to calculate photon spectra from tungsten anode x-ray tubes," *Phys. Med. Biol.*, vol. 54, no. 19, p. N433, 2009.

[4] R. Ramaseshan, K. Kohli, F. Cao, and R. K. Heaton, "Dosimetric evaluation of Plastic Water Diagnostic-Therapy." *J. Appl. Clin. Med. Phys.*, vol. 9, no. 2, p. 2761, 2008.

[5] F. Fischer, D. Hoppe, E. Schleicher, G. Mattausch, H. Flaske, R. Bartel, and U. Hampel, "An ultra fast electron beam x-ray tomography scanner," *Meas. Sci. Technol.*, vol. 19, p. 094002, 2008.

CHAPTER 3

Computed Tomography Reconstruction Algorithm

3.1 Introduction

In this chapter, we introduce a computed tomography (CT) reconstruction algorithm specifically designed to enhance the CT reconstruction result of the SEBXT system described in the previous chapter, §2. In theory, SEBXT can significantly enhance the temporal resolution of the CT image compared to the conventional CT. However, it still has several drawbacks. Full-angle CT data requires angular overlap between the source target and the detector. The SEBXT has limited angle settings, which inevitably generates artifacts due to the missing projection data to resolve the entire tangential line of the objects. Moreover, applying conventional CT reconstruction algorithms such as filtered back-projection (FBP) to the SEBXT system with a poly-energetic source induces an artifact due to the nonlinearity of the attenuation (i.e. low energy photons are attenuated more than high energy photons). With this condition, the reconstructed image has 1) cup-like global intensity distortion which is undesirable for quantative result, 2) dark and bright streaks between greater attenuation objects such as metals [1]. Therefore, we have endeavored to develop a custom CT reconstruction algorithm that is meant to produce the best reconstructions from the SEBXT projections.

For limited angle CT of two phase flows, there have been attempts to improve the resulting image by thresholding binarization [2] and level-set binarization [3]. However, binarization can only be applied to images where all objects of significance have size larger than the reconstruction resolution (i.e., voxel size). The size of objects (bubbles) in multiphase flows, particularly cavitating flows, ranges from microns to centimeters. Our objective is to develop an SEBXT system that can be used for any two-phase pattern of flows as long as the flow speed is within an acceptable range. In multi-phase fluid mechanics, a volume fraction approach is commonly used to account for smaller objects that cannot be fully

resolved, where the material property is volume-averaged over each voxel. With volume fraction approaches, each pixel can have any gray value (i.e., phase fraction) between 0 to 1. Applying binarization to the flows having void fractions varying continuously from 0 to 1, and being far from either extreme over a large part of the volume, can cause significant inaccuracy.

Numerous statistical reconstruction methods for reducing the beam hardening artifacts in poly-energetic source CT have been published. Among them, dual energy (DE) CT has been a popular method to account for the energy-dependent nonlinearity of the X-ray photon attenuation [4–30]. Although details vary, but most of the dual energy CT reconstructions use the dual material decomposition, which represents every material in the domain as a linear combination of two (soft and hard) materials to simplify the problem. DECT acquisitions can be classified into four categories: dual source [6, 19, 27], fast kVp switching [21, 22], dual layer detectors [31, 32] and energy binning detectors [9, 14]. Applying dual source or dual layer detector to the electron beam is difficult, as collimators cannot be used in SEBXT because the incident angle of the beam changes along the scan. Fast kVp switching in SEBXT requires synchronization between the sequences of beam voltages and deflection coils, and may not be achievable due to limitations of high voltage (HV) supply response time. Additionally, failed synchronization could cause the beam to hit the chamber walls, which can cause permanent damage in mere fractions of a second. To date, SEBXT has only been utilized for single energy CT, i.e., single source with energy-integrating detectors.

Achieving DECT-like results when using only the single bin energy-integrated data are not straightforward, especially when only limited angle projection data is available. Unlike the Poisson maximum likelihood of the photon counting detector, less attention has gone into finding an accurate probabilistic model for energy-integrating detectors, and many existing iterative methods use conventional weighted least square (WLS) method for data-fitting (likelihood) objective functions [11,28]. While WLS is a widely accepted method for high photon count rate, it could fail to accurately estimate parameters when photon count rates are low. Whiting *et al.* 2006 [33] showed that the compound Poisson distribution of the energy-integrating data match well with the measured probability density function, and it quickly becomes Gaussian-like when the number of photons becomes greater than 20. That work also showed that the mean and the variance of the energy-integrating data have a linear relation. Recently, Lasio *et al.* 2007 [34] proposed a simple Poisson log likelihood for energy-integrating detectors using equivalent photon counts, i.e., total energy divided by the equivalent mean energy of incident photons. Although the simple Poisson model was accurate enough for a certain problem [34], its linear approximation does not fully

reflect the nonlinear physics of the attenuation of photons.

In §3.2.2, we describe a way to take advantage of the dual material decomposition method similar to that used for the beam-hardening removal, while using energy-integrating detectors. To do this, we used the material decomposition method based on the volume fraction maps following [30], and a nonlinear Gaussian model as an approximation of the aggregated compound Poisson distribution. Then, we incorporate the spectral response model of the detector crystals into the probability model, by using a simple Gaussian and tail mixture model suggested by Srivastava *et al.* 2012 [35].

Estimating dual material volume fractions using only a single bin energy-integrated data leads to an ill-posed problem where the solution is not unique. Solving the ill-posed problem is challenging, and often it gives an unexpected result that is far from the actual solution. One way to help solving the ill-posed problem is applying *a priori* knowledge about the parameters to be estimated. For example, we may know that adjacent pixels of the images have similar values, or how the image values are bounded, and thus cannot exceed a certain value. Iterative CT reconstruction enables implementing these constraints. The Gradient Descent (GD) method is an iterative method that leads to the locally optimal CT solution where the given objective function is maximized (or minimized) by taking the gradient direction (or the negative gradient direction if it is minimization problem). The objective function is a vector-to-scalar function that measures how well the estimated data matches the measured data (data-fitting), or how well the parameters conform to the *a priori* knowledge (regularization). Classically, the X-ray CT reconstruction problem was approximated as linear, and Fourier transform based filtered-back projection (FBP) is used, which is faster than any other existing method. However, especially for sparse or limited angle CT, iterative reconstruction methods have proven to provide better results than the FBP. This is owing to its self-correcting characteristics and the freedom of choice for applying non-linear data-fitting functions, regularizers, or any other constraints that can improve the result.

Employing the volume fraction map has a benefit as its physical constraints, i.e., non-negativeness and the bounded sum, help to narrow the feasible domain. Even after the constraints are applied, the solution is still not unique. We employ two regularizers, an edge-preserving and ℓ_0 norm regularizer to determine the solution. Based on the dual material decomposition, different regularizers can be applied to the fraction maps depending on the characteristics of the materials.

In the present chapter, we employ a constrained and regularized, non-linear maximum likelihood iterative reconstruction. Maximum Likelihood (ML) estimation is a method to find parameters of a known probability density function that render the likelihood of the

measured data as great as possible [36]. For the proposed CT reconstruction, the parameters to be estimated are chosen as the material volume fraction maps, and the goal of the ML is to maximize the probability density function value of the energy-integrated detector data at the measured data point. To use ML estimation, a distribution model of the measurement data must be provided first. §3.2 will describe the modeling process to obtain the probability distribution. Once the probability distribution is defined, we need to convert it to a objective (cost) function and then find a set of parameters that minimizes the objective function. §3.3 describes the algorithm to find the locally optimal parameters that minimizes the objective function. CT reconstruction involves a large size data connected to a large size of parameters. To boost the speed of computation, a parallel computing is required. We will use a massively parallel GPU computation to accelerate the reconstruction process. GPU implementation of the reconstruction algorithm is shown in §3.4.

3.1.1 Basic principles of computed tomography reconstruction

We start our discussion of the computed tomography (CT) with the attenuation of X-ray photons under a 'narrow beam' or 'good geometry' assumption, in which X-ray rays are well-collimated such that only the straight beam can hit detectors [37, 38]. In other words, the X-ray photons that had gone through large angle scattering are assumed to be attenuated. With the narrow beam assumption, the attenuation ratio of monoenergetic photon flux passing through matter can be presented by a negative exponential function of the path length multiplied by material dependent factors. This is called the Beer's law or Beer-Lambert law:

$$\frac{y}{I^0} = e^{-\mu l}, \tag{3.1}$$

where y is the transmitted photon flux, I^0 is the initial photon flux, μ is the attenuation coefficient, and l is the path length. Each material has its own unique μ, which is dependent on the energy of the photons. In a more common case of multiple poly-energetic rays passing through mixture of matters, Eqn. (3.1) can be expressed as

$$\frac{y_i(\mathcal{E})}{I_i^0(\mathcal{E})} = e^{-\int_{\mathcal{L}_i} \mu(\mathbf{x},\mathcal{E})\mathrm{d}l}, \tag{3.2}$$

where i is the ray index, $\mathcal{E}$ is the energy level, and $\mathcal{L}_i$ is the photon path of the i^{th} ray assumed to be a straight line.

Energy-integrating detectors measure the integrated current over exposure time along the i^{th} source-detector pair. Our measurement data h_i are the total integrated X-ray photon energy obtained by linearly transforming the integrated current data. For an ideal 'large'

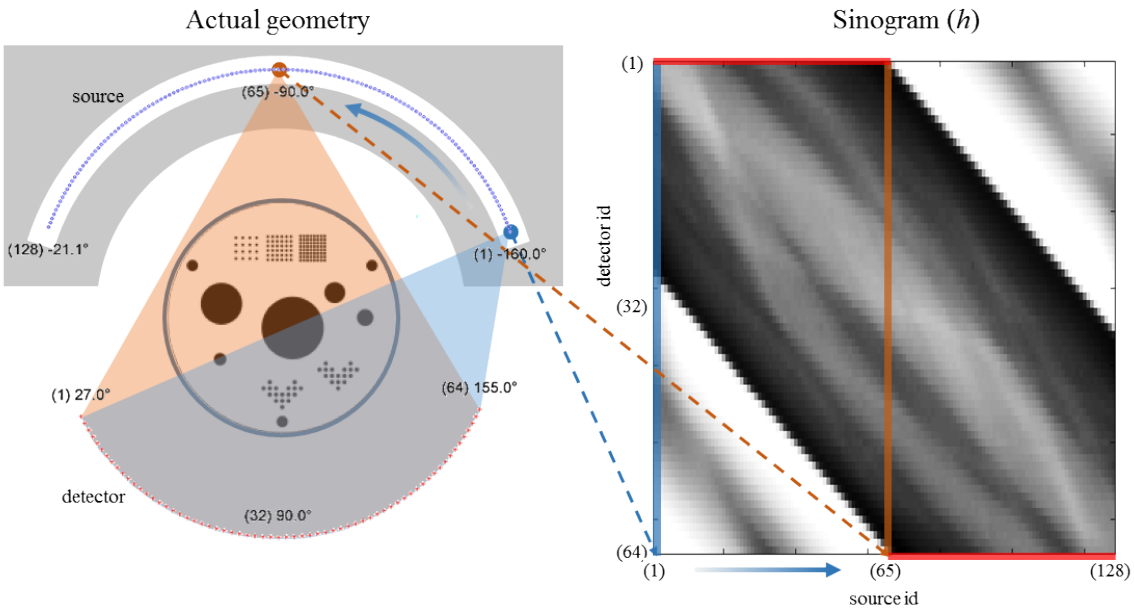


Figure 3.1: Example of a sinogram, i.e., map of the measurement data $\mathbf{h}$ plotted in a two dimensional array with the source and the detector indices as its axes. The thick red line indicates the region beyond which data are missing due to the use of a limited angle CT system (i.e. a system where every voxel in the domain is not imaged from every angle, from 0 to 180 degrees).

detector, which transfers the entire X-ray photon energy to electric energy, the measurement data (h_i) can be simply written as the integral of the energy of transmitted X-ray photons.

$$h_i = \int_0^\infty y_i(\mathcal{E})\,\mathcal{E}\,\mathrm{d}\mathcal{E}. \tag{3.3}$$

The map of h_i drawn with axes indicating the source and detector positions is called a sinogram. In reality, the process of X-ray transmission cannot be exactly described in a deterministic way. Not only are the generation and the attenuation process of the X-rays random, but X-ray photon energy is not fully transferred to the detectors. Those uncertainties require a more complicated statistical model to invert the measurement data. A statistical model which accounts for the efficiency of the detectors is discussed in §3.2.4. Fig. 3.1 shows an example of a simulated sinogram showing the relation between the source and detector geometry, and the axes of the sinogram.

Reconstructing an attenuation image $\mu(\mathbf{x}, \mathcal{E})$, or any other variations such as a volume fraction image, from the measurement data h is an inverse problem, as it has to invert the physical X-ray system to recover the object image from its projected data. CT reconstruction is the process by which many X-ray projection images are combined to create a

two-dimensional or a three-dimensional slice image. CT is usually ill-posed, i.e., a unique solution does not exist since there is not enough information to uniquely determine the solution. Moreover, the experimentally obtained projection data are corrupted by noise whose exact value is unknown. Even though over-determined data can improve the quality of reconstruction, acquiring more data means more cost (i.e. more detectors or projections, or a higher X-ray beam power).

Another challenge arises for sparse or limited angle projection datasets. The sparse angle problem occurs when the projection angles are too far apart, resulting in a loss of the connection between the projection images. The limited angle problem occurs when the projection dataset cannot cover the required angle of view. Generally, electron beam X-ray CT suffers either from sparse angles or the limited angle problem, unless a special X-ray transparent target is used [39].

3.2 Mathematical modeling of CT systems

3.2.1 Probability model

For current mode (energy-integrating) semiconductor type detectors with a polyenergetic source, the actual distribution of the energy transferred to the detector by photons falls into the sum of independent compound Poisson random variables, called the aggregated compound Poisson random variable [40], which is not practically applicable for the iterative maximum likelihood estimator due to its complicated form. As a common approach, the distribution of the measurement data H_i of the current mode detector is approximated by the normal distribution [40], [41], and [42]:

$$H_i \sim \mathcal{N}\left(\eta(\bar{y}_i), \sigma^2(\bar{y}_i)\right) , \tag{3.4}$$

where η is the mean of the measurement data H, σ^2 is the variance of H, i is the index of rays, and $\bar{y}_i$ is the ideal expected transmitted X-ray photon count at the i^{th} ray. After discretizing the continuous X-ray photon energy level $\mathcal{E}$ into N_k bins with an interval $\Delta\mathcal{E}$, we define $\bar{y}_i(\mathcal{E}_k)$ as the ideal expected photon counts at the i^{th} ray and the k^{th} energy bin based on the Beer's law:

$$\bar{y}_i(\mathcal{E}_k) \equiv \int_{\mathcal{E}_k - 1/2\Delta\mathcal{E}}^{\mathcal{E}_k + 1/2\Delta\mathcal{E}} I_i^0(\mathcal{E}) \exp\left(-\int_{\mathcal{L}_i} \mu(\mathbf{x}, \mathcal{E})\, \mathrm{d}l\right) \mathrm{d}\mathcal{E} , \tag{3.5}$$

where k is the energy level index, $\Delta\mathcal{E}$ is the interval between energy levels, $I_i^0(\mathcal{E})$ is the continuous initial emission spectrum, $\int_{\mathcal{L}_i} \cdot \, \mathrm{d}l$ is the line integral along the i^{th} ray, and $\mu(\mathbf{x}, \mathcal{E})$ is the attenuation coefficient at the spatial location $\mathbf{x}$ and energy $\mathcal{E}$. Further, assume $\Delta\mathcal{E}$ is small enough that we can approximate (3.5) as

$$\bar{y}_i(\mathcal{E}_k) \equiv I_i^0(\mathcal{E}_k) \exp\left(-\int_{\mathcal{L}_i} \mu(\mathbf{x}, \mathcal{E}_k)\, \mathrm{d}l\right) \Delta\mathcal{E}\,. \tag{3.6}$$

The main goal of the following section is to define the analytic expression of η and σ^2, then apply the maximum likelihood estimation, and lastly, solve it iteratively using a majorization-minimization algorithm specifically designed for the given likelihood function.

3.2.2 Basis material decomposition

In (3.5), the attenuation coefficient $\mu(\mathbf{x}, \mathcal{E})$ is a function of both space and energy. For the sake of simplicity, we first decompose it into an energy-dependent term and a space-dependent term. Sukovic and Clinthorne 1994 [8], 1999 [10] and 2000 [11] suggested a decomposition method to split the attenuation coefficient into the mass attenuation coefficient which depends on energy, and the local density which depends on space, as follows:

$$\mu(\mathbf{x}, \mathcal{E}) = \sum_{l=1}^{N_l} \beta_l(\mathcal{E}) \rho_l(\mathbf{x})\,, \tag{3.7}$$

where N_l is the total number of materials in the object, $\beta_l(\mathcal{E})$ is the known mass attenuation coefficient of the l^{th} material at energy level $\mathcal{E}$, and $\rho_l(\mathbf{x})$ is the local density at the spatial location $\mathbf{x}$. Elbakri and Fessler 2003 [17] further decomposed the density map into actual material density and fraction map,

$$\mu(\mathbf{x}, \mathcal{E}) = \sum_{l=1}^{N_l} \beta_l(\mathcal{E})\, \rho(\mathbf{x})\, f_l(\rho(\mathbf{x}))\,. \tag{3.8}$$

In this model, the material fraction is assumed to be a function of the density, where the function can be obtained by empirical curve fitting based on the known material data. While this approach can be used for the specific field where there is a strong relationship between the density and fraction, it cannot be applied to the more general case where the fraction is not a one-to-one function of the density. Instead, Long and Fessler 2014 [30] introduced a

modified volume fraction method in which the fraction is independent of the density:

$$\mu(\mathbf{x}, \mathcal{E}) = \sum_{l=1}^{N_l} \beta_l(\mathcal{E})\, \rho_l\, f_l(\mathbf{x})\,. \tag{3.9}$$

This type of decomposition can facilitate applying physical constraints. For example, if all materials in the object are considered in the reconstruction, we can set the constraint $\sum_l f_l(\mathbf{x}) = 1$ for any $\mathbf{x}$. This sum-to-one constraint is called volume conservation. Usually, X-ray attenuation due to the air is negligibly small, thus the attenuation of the air is assumed to be zero. Then, we can set the constraint $\sum_l f_l(\mathbf{x}) \leq 1$ for any $\mathbf{x}$, excluding the air fraction. This constraint significantly improves the quality of reconstructed images for limited angle CT, as shown later.

Finally, we can discretize $\mu(\mathbf{x}, \mathcal{E})$ using the piecewise constant basis in energy,

$$\mu(\mathbf{x}, \mathcal{E}_k) = \sum_{l=1}^{N_l} u_l(\mathcal{E}_k)\, f_l(\mathbf{x})\,, \tag{3.10}$$

where $u_l(\mathcal{E}_k) = u_{(k,l)} = \beta_l(\mathcal{E}_k)\rho_l$. After discretizing $f_l(\mathbf{x})$ in space, and substituting (3.10) into (3.5), the discretized average photon flux $\bar{y}_{(i,k)}$ can be written as:

$$\bar{y}_{(i,k)} = I^0_{(i,k)}\, \exp\left(-\sum_{l=1}^{N_l} u_{(k,l)} \left[\mathbf{A}\mathbf{f}_{(l)}\right]_i \right) \Delta\mathcal{E}\,, \tag{3.11}$$

where $I^0_{(i,k)} = I^0_i(\mathcal{E}_k)$, $[\mathbf{t}]_i$ is the i^{th} element of the vector $\mathbf{t} \in \mathbb{R}^{N_i}$, $\mathbf{A} \in \mathbb{R}^{N_i \times N_j}$ is the projection matrix with its element $a_{(i,j)}$ being the Radon transform of the j^{th} spatial basis along the i^{th} ray, and $\mathbf{f}_{(l)} = \left[f_{(1,l)}, \cdots, f_{(N_j,l)}\right]'$, where N_j is the number of image pixels in space.

3.2.3 Spatial discretization based on the Zwart-Powell box spline

In (3.11), we use the weight matrix $\mathbf{A}$ to discretize the volume fraction image in space. A box-spline basis interpolator is used to build $\mathbf{A}$. A box-spline is a generalized, non-separable version of the B-spline. While 2D B-spline can be computed by the tensor product of the two 1D B-spline, which is the multiple auto-convolution of 1D rectangular function, 2D box-spline is computed by convolving rotated rectangular functions defined in 2D that cannot be represented by a tensor product of two 1D functions. Separability in Cartesian coordinate is useful property for the uniform grid system, but it is not much helpful for Radon transform as it is performed in a cylindrical coordinate.

By sacrificing separability, box-spline attained some good properties compared to the B-spline. Among the group of box-spline elements, Zwart-Powell (ZP) element has been widely used, due to its advantages of symmetric and smooth shape, less aliasing but still covering high frequency region slightly better [43]. Fig. 3.2 and 3.3 show the shape of ZP element, and its frequency response compared with other typical B-spline elements, respectively. A good basis element should have a smaller support (faster computation), frequency response that covers the entire input within the Nyquist frequency square (high frequency restoration), but reject the input outside the Nyquist frequency square (anti-aliasing). In Fig. 3.2, we can see that ZP element has better response at the corners of the Nyquist frequency square compared to the bi-quadratic B-spline element, and much less aliasing (outside of the Nyquist frequency square) compared to the square box element. Also, ZP element has slightly smaller support compared with the bi-quadratic B-spline element.

Entezari *et al.* 2012 [44] developed a closed form solution to compute the Radon transform of the box-spline of any arbitrary order, and showed in noiseless reconstructed images that ZP box-spline outperforms linear B-spline. A box-spline element can be defined by a direction matrix Ξ, which is a set of the basis directional vectors ξ_i:

$$\Xi = [\xi_1 \cdots \xi_{N_b}] \ . \tag{3.12}$$

Each directional vector ξ_i indicates the direction of the impulse-rectangular functions (elementary box-splines), which form the box-spline by convolving each other.

$$M_{\Xi}(\mathbf{x}) = \left(M_{\xi_1} * \ M_{\xi_2} * \cdots * \ M_{\xi_{N_b}} \right) (\mathbf{x}) \, , \tag{3.13}$$

$$M_{\xi_i}(\mathbf{x}) = \mathrm{Rect} \left\{ \ \frac{1}{|\xi_i|_2} \xi_i' \mathbf{x} \right\} \ \delta \Big\{ \xi_{i_\perp}' \mathbf{x} \Big\} \, , \tag{3.14}$$

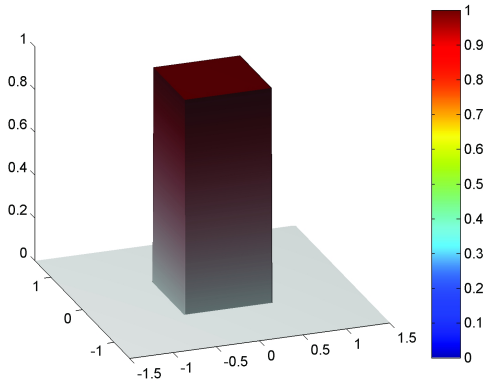
where $\delta[\cdot]$ is the Dirac delta function, $\xi_{i_\perp}$ is the unit vector orthogonal to ξ_i, and $\mathrm{Rect}[\cdot]$ is the scalar rectangular function:

$$\mathrm{Rect}\big[t\big] = \begin{cases} 1, & \text{if } |t| < \dfrac{1}{2} \\ 0, & \text{otherwise} \end{cases} \ . \tag{3.15}$$

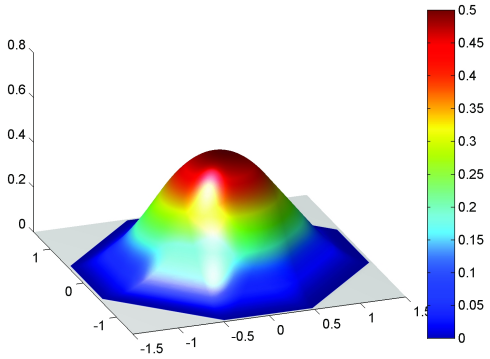
Zwart-Powell element is a special case of box-spline in 2-D, which has the direction matrix:

$$\Xi = \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \, . \tag{3.16}$$

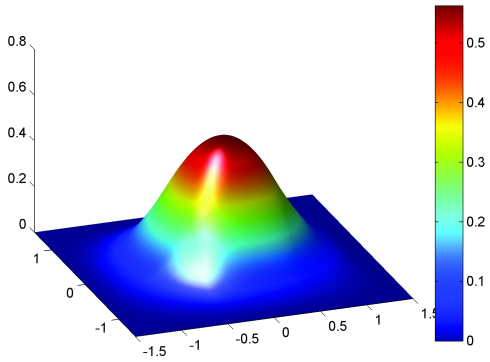
Estimating X-ray attenuation based on the Beer's law requires a line integral of the atten-



(a)

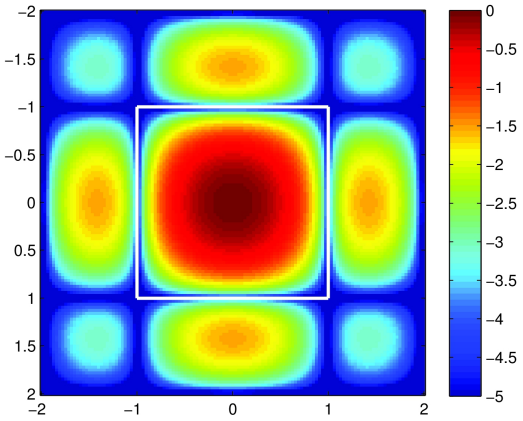


(b)

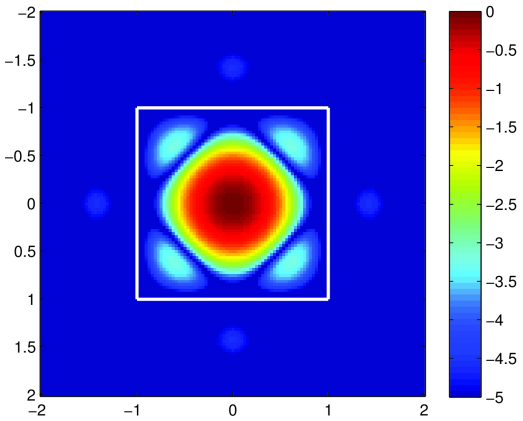


(c)

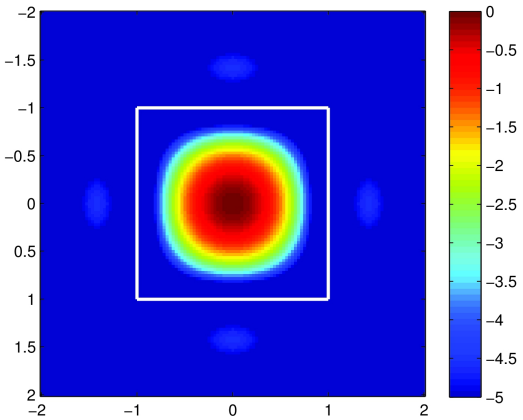
Figure 3.2: Shape of the square box (a), Zwart-Powell box spline (b), and the bi-quadratic B-spline elements (c).



(a)



(b)



(c)

Figure 3.3: Frequency response of the square box (a), Zwart-Powell box (b), and the bi-quadratic B-spline elements (c) shown in log scale.

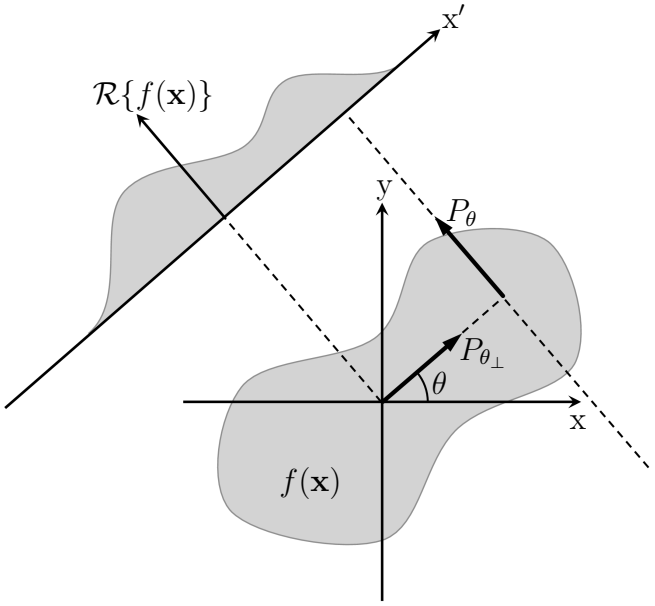


Figure 3.4: Radon transform coordinates

uation coefficient along the rays. This line integral can be represented by the projection along the direction of the ray θ and the corresponding hyperplane coordinate x', which is called the Radon transform. In two dimensional space, Radon transform of any spatial function(image) $f(\mathbf{x})$ can be written as

$$\begin{aligned}\mathcal{R}\{f(\mathbf{x})\}(\mathrm{x}',\theta) &= \iint_{-\infty}^{\infty} f(\mathbf{x})\delta(\mathrm{x}\cos\theta + \mathrm{y}\sin\theta - \mathrm{x}')\mathrm{dxdy} \\ &= \int_{-\infty}^{\infty} f(P_{\theta_\perp}\mathrm{x}' + P_\theta l)\mathrm{d}l\,, \end{aligned} \tag{3.17}$$

where the row vectors $P_{\theta_\perp} = [\cos\theta\ \sin\theta]$, and $P_\theta = [-\sin\theta\ \cos\theta]$. Radon transform has several useful linear properties. One of the convenient property for the computation of box-splines is the convolution-projection theorem:

$$\mathcal{R}\{(f_1 * f_2)(\mathbf{x})\} = \mathcal{R}\{f_1\} * \mathcal{R}\{f_2\}. \tag{3.18}$$

Using the theorem (3.18), the Radon transform of the box-spline element can be computed

as

$$\begin{aligned}\mathcal{R}\{M_{\Xi}(\mathbf{x})\}(\mathrm{x}',\theta) &= \left(\mathcal{R}\{M_{\xi_1}\} * \cdots * \mathcal{R}\{M_{\xi_{N_b}}\}\right)(\mathrm{x}',\theta) \\ &= \left(M_{\zeta_1} * \cdots * M_{\zeta_{N_b}}\right)(\mathrm{x}') \\ &= M_{\mathbf{Z}}(\mathrm{x}')\,, \end{aligned} \tag{3.19}$$

where $\mathbf{Z}$ is the projected direction matrix:

$$\mathbf{Z} = P_{\theta_\perp}\left[\xi_1 \cdots \xi_{N_b}\right] = \left[\zeta_1 \cdots \zeta_{N_b}\right]. \tag{3.20}$$

Convert the rectangular function into the difference of two step functions,

$$M_{\boldsymbol{\zeta}}(\mathrm{x}') = \Delta_{\zeta}\{u(\mathrm{x}')\}, \tag{3.21}$$

where $u(\mathrm{x}')$ is the unit step function, and $\Delta_{\zeta}\{f(\mathrm{x}')\}$ is the finite-difference operator, defined as

$$\Delta_{\zeta}\{f(\mathrm{x}')\} = \frac{f(\mathrm{x}') - f(\mathrm{x}'-\zeta)}{\zeta}. \tag{3.22}$$

Finally, by plugging (3.16) into (3.19), and applying auto-convolution property of step functions, Entezari *et al.* 2012 [44] found the solution of the Radon transform of the ZP element:

$$\mathcal{R}\{M_{\Xi}(\mathbf{x})\}(\mathrm{x}',\theta) = \frac{\Delta_{\cos\theta}\{\Delta_{\sin\theta}\{\Delta_{\cos\theta+\sin\theta}\{\Delta_{\cos\theta-\sin\theta}\{(\mathrm{x}'_+)^3\}\}\}\}}{3!}\,, \tag{3.23}$$

where x'_+ is the projection to the non-negative value, $x'_+ = \max\{\mathrm{x}', 0\}$.

One drawback of higher order box-spline elements including ZP element, is that it is computationally expensive when compared to commonly used constant box element. Furthermore, while strip integral of constant box basis can be achieved by an analytical expression, closed form of the strip integral of Zwart-Powell box-spline is not easy to be found. To reduce the excessive computation while slightly sacrificing the accuracy, first order interpolation with a pre-computed table is used along with the numerical integral scheme based on the six point Gauss-Legendre quadrature. A six point Gauss-Legendre quadrature can yield the exact integral value of polynomial functions up to 11 degrees. Gauss-Legendre quadrature integral of the square box basis suffers from the error due to its non-smooth corner, which cannot be accurately estimated by polynomial functions. Luckily, ZP element has a smooth shape which is preferable for the numerical Gauss-Legendre quadrature integral. Also, ZP element is four-axis symmetric, so mirrored boundary condition with 0

to 45 degree and positive distance data table suffices to cover the whole area. With texturing functionality of CUDA associated with the 2D CUDA array, this interpolation can be efficiently computed, which will be explained in §3.4.

3.2.4 Energy detection statistics for current mode detectors

To obtain statistical properties $\bar{\eta}$ and σ^2 for the total photon energy probability density function, assume that the event of the photo-electric interaction of each photon is independent, and number of X-ray photons at each discretized energy level follows independent Poisson distribution.

$$Y_{(i,k)} \sim \text{Poisson}\left(\bar{y}_{(i,k)}\right), \tag{3.24}$$

where the parameter (mean) $\bar{y}_{(i,k)}$ is defined in (3.11).

Then, a mathematical model is needed to compute the probability density function of energy transferred to the detector for each photon of the discrete energy level $\mathcal{E}_k$, which is interpreted as the spectral response $\mathcal{B}(\mathcal{E}_k)$ of the detector. Detectors measure total integrated energy of X-ray photons by reading the current induced by the occurrence of electrons and holes when photons interact with the semiconductor material. For photons with energy less than 1 MeV, two major interactions convert the photon energy to the electron energy, i.e., photo-electric and Compton scattering interaction. While photo-electric interaction fully converts each photon energy to electron energy, Compton scattering only transfers small amount of energy to electrons and generate less energy photons. Unless the detector crystal is extremely small, photons are converted to electrons through either a single photo-electric interaction, or multiple Compton scattering interactions. Probability of multiple Compton scattering event depends on the size of the detector, and it is hard to compute analytically. Instead, Srivastava *et al.* 2012 [35] suggested a simplified empirical model for $\mathcal{B}(\mathcal{E})$ as a mixture of Gaussian photo-peak and uniform Compton continuum:

$$\mathcal{B}(\mathcal{E}';\mathcal{E}) = w_{\mathcal{E}}\mathcal{B}_G(\mathcal{E}';\mathcal{E},\sigma_{\mathcal{E}}) + (1-w_{\mathcal{E}})\,\mathcal{B}_C(\mathcal{E}';\mathcal{E}), \tag{3.25}$$

$$\mathcal{B}_G(\mathcal{E}';\mathcal{E}) = \frac{1}{\sqrt{2\pi}\sigma_{\mathcal{E}}}\exp\left(-\frac{(\mathcal{E}'-\mathcal{E})^2}{2\sigma_{\mathcal{E}}^2}\right), \tag{3.26}$$

$$\mathcal{B}_C(\mathcal{E}';\mathcal{E}) = \begin{cases} {}^1\!/_{\mathcal{E}} & \text{if } \ 0 \le \mathcal{E}' \le \mathcal{E} \\ 0 & \text{otherwise} \end{cases}, \tag{3.27}$$

$$\sigma_{\mathcal{E}} = k\sqrt{\mathcal{E}}. \tag{3.28}$$

Here, $w_{\mathcal{E}}$ is the mixture probability of photo-electric interaction, and k is the variance coef-

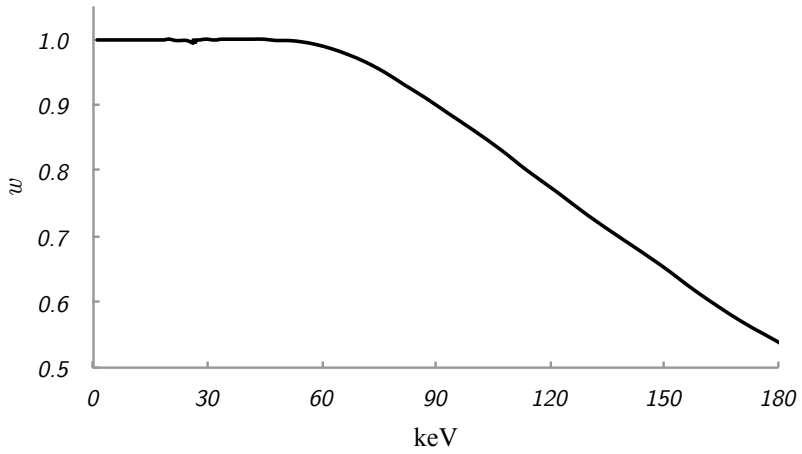


Figure 3.5: Mixture weight of photo-electric interaction $w_{\mathcal{E}}$ of 1 mm CdTe detector (based on the efficiency data from Amptek XR-100T-CdTe [45])

ficient. Without conducting experiments with multiple monoenergetic sources, estimating k and $w_{\mathcal{E}}$ of the detector by experiment is impossible. For the simulation purpose, we extracted $w_{\mathcal{E}}$ from the existing efficiency data of the commercial 1 mm thick CdTe detector provided by Amptek [45]. Fig. 3.5 shows estimated $w_{\mathcal{E}}$ for 1 mm thick commercial CdTe semiconductor detector. Note that $w_{\mathcal{E}}$ depends on both the material and the size of the detector. Estimated values of $w_{\mathcal{E}}$ in Fig. 3.5 are for the simulation only. For actual data reconstruction, $w_{\mathcal{E}}$ must be tuned to fit the spectral response of the specific detectors. In general, CdTe detector shows good efficiency at low energy, then gradually loses its efficiency after the energy higher than 100 keV.

Denote $E_m(\mathcal{E})$ as the energy transferred to the detector from m^{th} photon which originally had energy $\mathcal{E}$. As noted before, energy transferring events at the detector are assumed to be independent, then $E_m(\mathcal{E})$ can be seen as independent and identically distributed (i.i.d.) random variables that have the spectral response function as their probability density function,

$$E_m(\mathcal{E}) \overset{\text{i.i.d.}}{\sim} \mathcal{B}(\,\cdot\,;\mathcal{E})\,. \tag{3.29}$$

To get $\bar{\eta}$ and σ^2, we first need to compute the first, second, and third moment of $E_m(\mathcal{E})$ using (3.25) to (3.28),

$$m_E^{(1)} \equiv \mathbb{E}\big[E_m;\mathcal{E}\big] = \frac{w_{\mathcal{E}}+1}{2}\mathcal{E}\,, \tag{3.30}$$

$$m_E^{(2)} \equiv \mathbb{E}\big[E_m^2;\mathcal{E}\big] = w_{\mathcal{E}}\sigma_{\mathcal{E}}^2 + \frac{2w_{\mathcal{E}}+1}{3}\mathcal{E}^2\,, \tag{3.31}$$

$$m_E^{(3)} \equiv \mathbb{E}\big[E_m^3;\mathcal{E}\big] = 3w_{\mathcal{E}}\mathcal{E}\sigma_{\mathcal{E}}^2 + \frac{3w_{\mathcal{E}}+1}{4}\mathcal{E}^3\,. \tag{3.32}$$

Let T_i be the total energy transferred to the i^{th} detector from the monoenergetic beam with energy level $\mathcal{E}_k$, then T_i can be written as the compound Poisson random variable, i.e., sum of the i.i.d. random variables where the total number is Poisson random variable:

$$T_i(\mathcal{E}_k) = \sum_{m=1}^{Y_i} E_m(\mathcal{E}_k). \tag{3.33}$$

The mean, variance, and skewness of $T_i(\mathcal{E}_k)$ can be computed using the moments of $E_m(\mathcal{E}_k)$ in (3.30) to (3.32) as

$$\mathbb{E}\big[T_i(\mathcal{E}_k)\big] = \bar{y}_{(i,k)}\, m_{E_k}^{(1)}\,, \tag{3.34}$$

$$\mathrm{Var}\big[\mathrm{T_i}(\mathcal{E}_\mathrm{k})\big] = \bar{y}_{(i,k)}\, m_{E_k}^{(2)}\,, \tag{3.35}$$

$$\mathrm{Skew}\big[\mathrm{T_i}(\mathcal{E}_\mathrm{k})\big] = \frac{m_{E_k}^{(3)}}{\sqrt{\bar{y}_{(i,k)}\left(m_{E_k}^{(2)}\right)^3}}\,. \tag{3.36}$$

Details to obtain the statistical properties of the compound Poisson random variables can be found in [40]. For the polyenergetic beam, assume that all $T_i(\mathcal{E}_k)$ at different discretized energy levels are mutually independent. Then the sum $H_i = \sum_{k=1}^{N_k} T_i(\mathcal{E}_k)$ becomes the aggregated compound Poisson random variable [40], and its statistical properties are

$$\mathbb{E}\big[H_i\big] = \sum_{k=1}^{N_k} \bar{y}_{(i,k)} m_{E_k}^{(1)}\,, \tag{3.37}$$

$$\mathrm{Var}\big[\mathrm{H_i}\big] = \sum_{k=1}^{N_k} \bar{y}_{(i,k)} m_{E_k}^{(2)}\,, \tag{3.38}$$

$$\mathrm{Skew}\big[\mathrm{H_i}\big] = \frac{\sum_{k=1}^{N_k} \bar{y}_{(i,k)} m_{E_k}^{(3)}}{\left(\sum_{k=1}^{N_k} \bar{y}_{(i,k)} m_{E_k}^{(2)}\right)^{3/2}}\,. \tag{3.39}$$

where $\bar{y}_{(i,k)}$ is defined in (3.11).

Since skewness of H_i is non-zero, approximating probability density function of H_i requires at least a three-parameter distribution model to satisfy all the mean, variance and skewness at the same time. Shifted gamma distribution is known to be a good candidate to estimate the compound Poisson distribution [40], defined as

$$G^{tr}(h; h^{(0)}, a, b) = G(h - h^{(0)}; a, b)\,. \tag{3.40}$$

where G is the two-parameter gamma distribution. a and b are the shape and rate parameters

of the two-parameter gamma distribution, respectively. $h^{(0)}$ is the mean-shifting parameter. To find the parameters that satisfy the statistical properties described in (3.37), (3.38) and (3.39), first find a and b from the variance and skewness, then find $h^{(0)}$ to satisfy the mean by shifting the two-parameter gamma distribution. Then, we get the following analytic expressions for the shifted gamma approximation:

$$a_i = 4\,\frac{\left(\sum_{k=1}^{N_k} \bar{y}_{(i,k)} m_{E_k}^{(2)}\right)^3}{\left(\sum_{k=1}^{N_k} \bar{y}_{(i,k)} m_{E_k}^{(3)}\right)^2}\,, \tag{3.41}$$

$$b_i = 2\,\frac{\sum_{k=1}^{N_k} \bar{y}_{(i,k)} m_{E_k}^{(2)}}{\sum_{k=1}^{N_k} \bar{y}_{(i,k)} m_{E_k}^{(3)}}\,, \tag{3.42}$$

$$h_i^{(0)} = \sum_{k=1}^{N_k} \bar{y}_{(i,k)} m_{E_k}^{(1)} - \frac{a_i}{b_i}\,. \tag{3.43}$$

Even though the shifted gamma distribution can approximate the given compound Poisson approximation accurately, developing maximum likelihood for the X-ray CT which has massive amount of parameters to be estimated is not feasible due to the complicated digamma functions. Hence, we need some sort of simpler distribution model to design the estimator. As mentioned earlier, Gaussian distribution is chosen in the hope of the central limit theorem. Naive Gaussian approximation based on the mean (3.37) and the variance (3.38) are

$$H_i \sim \mathcal{N}\left(\eta(\bar{\mathbf{y}}_i), \sigma^2(\bar{\mathbf{y}}_i)\right)\,, \tag{3.44}$$

$$\eta(\bar{\mathbf{y}}_i) = \sum_{k=1}^{N_k} \bar{y}_{(i,k)} m_{E_k}^{(1)}\,, \tag{3.45}$$

$$\sigma^2(\bar{\mathbf{y}}_i) = \sum_{k=1}^{N_k} \bar{y}_{(i,k)} m_{E_k}^{(2)}\,, \tag{3.46}$$

where $\bar{\mathbf{y}}_i = \left[\bar{y}_{(i,1)}, \cdots, \bar{y}_{(i,N_k)}\right]'$. Note that the estimator based on (3.44) does not fall into WLS, as the variance is also a function of $\bar{\mathbf{y}}_i$. Due to the skewness error, maximum likelihood estimation of Gaussian approximation always give a biased result as the mean and the mode of the compound Poisson distribution are not in the same place. The difference between the mean and the mode can be seen in Fig. 3.6 where the mean is at the peak of the Gaussian distribution, and the mode is at the peak of the gamma or actual distribution. Thus, we want to move the mean of the Gaussian distribution close to the mode. For $\alpha_i > 1$, the difference between the mode and the mean can be found from the approximated

shifted gamma distribution:

$$\begin{aligned}\mathbb{E}\big[H_i\big] - \operatorname{mode}\big[H_i\big] \\ &= \left(\frac{a_i}{b_i} + t_{0i}\right) - \left(\frac{a_i - 1}{b_i} + t_{0i}\right) \\ &= \frac{1}{b_i}\,. \end{aligned} \tag{3.47}$$

In terms of maximum likelihood, shifting the mean of Gaussian distribution onto the mode of the actual distribution increases the accuracy of the estimation when the data is close to the mode, but it also decreases the accuracy when the data is away from the mode. As a trade off, we chose a parameter $\hat{\upsilon}$ that minimizes the mean-squared error between the actual distribution and the shifted Gaussian approximation within the full-width at half maximum (FWHM) region as:

$$H_i \sim \mathcal{N}\left(\eta(\bar{\mathbf{y}}_i) - \hat{\upsilon}\frac{1}{b(\bar{\mathbf{y}}_i)}, \sigma^2(\bar{\mathbf{y}}_i)\right)\,, \tag{3.48}$$

$$\hat{\upsilon} = \underset{\upsilon}{\arg\min} \int_{\mathrm{HM}+}^{\mathrm{HM}-} \left(P_{\mathrm{true}}(h) - \mathcal{N}\Big(h; \eta - \upsilon\frac{1}{b}, \sigma^2\Big)\right)^2 \mathrm{d}h\,, \tag{3.49}$$

where P_{true} is the actual probability density function, HM$-$ and HM$+$ indicates low and high FWHM bound, respectively. Finding optimal υ is difficult as the actual distribution P_{true} has a complicated form. Up to now, $\hat{\upsilon}$ could be found only from the numerical experiments. As a result of experiments, υ was found roughly at 0.8. Even though $\hat{\upsilon} = 0.8$ is a rough estimation and not the analytical solution of (3.49), this value will be sufficient as there are many other uncertainties that can exceed the error of $\hat{\upsilon}$.

To validate the approximated probability density function in (3.4), analytic expression of the total energy is also obtained from the following recursive auto-convolution equation [33]. For monoenergetic beam,

$$\begin{aligned} f_{T_i(\mathcal{E})}(t) &= \sum_{y=1}^{\infty} f_{T_i|Y_i(\mathcal{E})}\left(t|Y_i(\mathcal{E}) = y\right) \cdot \Pr\left(Y_i(\mathcal{E}) = y\right) \\ &= \sum_{y=1}^{\infty} \Big(f_{T_i|Y_i(\mathcal{E})}\left(t|Y_i(\mathcal{E}) = y-1\right) * f_{T_i|Y_i(\mathcal{E})}(t|Y_i(\mathcal{E}) = 1)\Big) \\ &\qquad \cdot \Pr(Y_i(\mathcal{E}) = y)\,. \end{aligned} \tag{3.50}$$

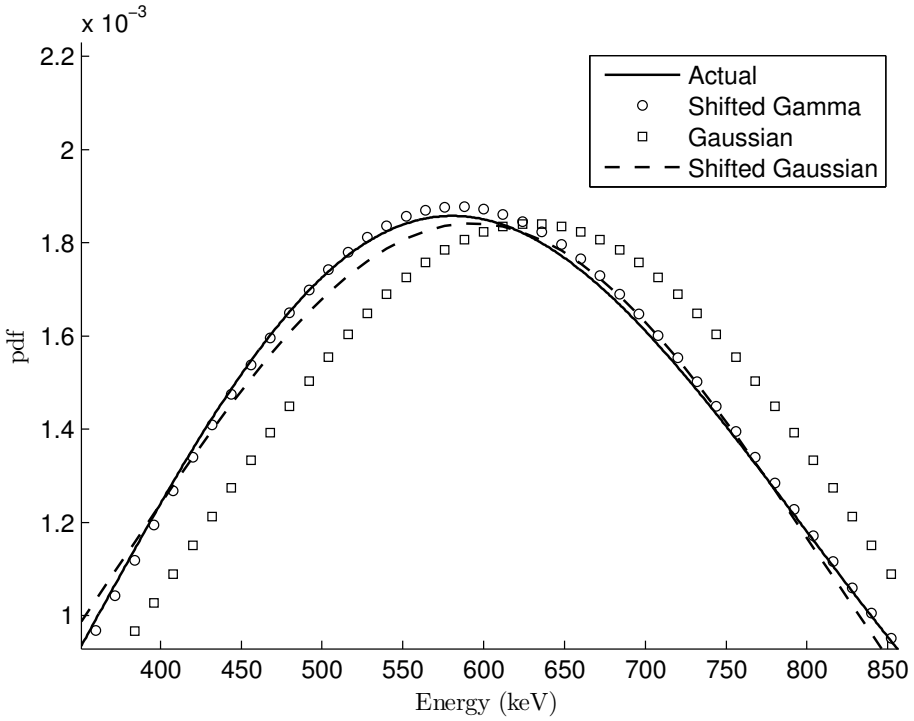


Figure 3.6: Comparison of the probability density functions for the energy transferred to the detector by bichromatic X-ray photons at FWHM region, $\bar{y}(\mathcal{E}_1) = 10$, $\bar{y}(\mathcal{E}_2) = 5$, $\mathcal{E}_1 = 20$, $\mathcal{E}_2 = 100$, $w_1 = w_2 = 0.8$, $k = 0.5$, $\upsilon = 0.8$

where $*$ denotes 1-D convolution with respect to t. For a polyenergetic beam,

$$\begin{aligned} f_{H_i}(t) &= f_{\sum_k T_i(\mathcal{E}_k)}(t) \\ &= \Big(f_{T_i(\mathcal{E}_1)} * f_{T_i(\mathcal{E}_2)} * \cdots * f_{T_i(\mathcal{E}_K)} \Big)(t) \,. \end{aligned} \tag{3.51}$$

Fig. 3.6 compares the analytic compound Poisson, shifted Gamma approximation, Gaussian approximation, and shifted Gaussian approximation for the case of bichromatic beam at FWHM region of the actual distribution. Note that the mean total number of photons are 15 in this case, which is equivalent to the worst case of the actual attenuated photon rate of our SEBXT.

Finally, assuming readout noises r_i are additive and independent, then η and σ^2 can be obtained by the linearity of the expectation and the variance of independent random variables:

$$H_i \sim \mathcal{N}\left(\eta(\bar{\mathbf{y}}_i) - \hat{\upsilon} \frac{1}{b(\bar{\mathbf{y}}_i)} + \bar{r}_i,\ \sigma^2(\bar{\mathbf{y}}_i) + \sigma_{r_i}^2 \right), \tag{3.52}$$

where $\bar{r}_i$ and $\sigma_{r_i}^2$ are the mean and variance of the readout noise r_i, respectively.

3.3 Reconstruction Algorithm

3.3.1 Maximum likelihood estimation

For simplicity, we drop readout noise terms hereinafter. We assume the detector signals are mutually independent, so that the jointly distributed probability density (likelihood) function can be derived from (3.48) as

$$f_H(\mathbf{h};\boldsymbol{\eta},\boldsymbol{b},\boldsymbol{\sigma}) = \frac{1}{\sqrt{2\pi^{N_i}\prod_{i=1}^{N_i}\sigma_i^2}} \prod_{i=1}^{N_i} \exp\left(-\frac{\left(h_i-\eta_i+\hat{v}/b_i\right)^2}{2\,\sigma_i^2}\right), \tag{3.53}$$

where N_i is the number of detector pixels, h_i is the measurement data at i^{th} detector pixel, $\eta_i = \eta(\bar{y}_i)$, $b_i = b(\bar{y}_i)$, $\sigma_i^2 = \sigma^2(\bar{y}_i)$, $\mathbf{h} = [h_1,\cdots,h_{N_i}]'$, $\boldsymbol{\eta} = [\eta_1,\cdots,\eta_{N_i}]'$, $\boldsymbol{b} = [b_1,\cdots,b_{N_i}]'$, and $\boldsymbol{\sigma} = [\sigma_1,\cdots,\sigma_{N_i}]'$, Since f_H is non-negative, finding the minimum of the negative log of the likelihood function is equivalent to finding the maximum of the likelihood function.

$$-\log f_H(\mathbf{h};\boldsymbol{\eta},\boldsymbol{b},\boldsymbol{\sigma}) = \frac{N_i}{2}\log 2\pi + \frac{1}{2}\sum_{i=1}^{N_i}\log\sigma_i^2 + \frac{1}{2}\sum_{i=1}^{N_i}\frac{(h_i-\eta_i+\hat{v}/b_i)^2}{\sigma_i^2}. \tag{3.54}$$

After discarding the constant term, the data-fitting objective function $\Psi_L(\bar{\mathbf{y}})$ is defined as

$$\Psi_L(\bar{\mathbf{y}}) = \frac{1}{2}\sum_{i=1}^{N_i}\psi_L(\bar{\mathbf{y}}_i) = \frac{1}{2}\sum_{i=1}^{N_i}\left[\log\sigma_i^2(\bar{\mathbf{y}}_i) + \frac{\left(h_i-\eta_i(\bar{\mathbf{y}}_i)+\hat{v}/b_i(\bar{\mathbf{y}}_i)\right)^2}{\sigma_i^2(\bar{\mathbf{y}}_i)}\right]. \tag{3.55}$$

Convexity of the data-fitting function is preferred for the sake of the uniqueness of the solution. To check the convexity, first see the second derivative of $\psi_L(\bar{y})$ for the case of monoenergetic beam:

$$\frac{\partial^2\psi_L}{\partial\bar{y}^2} = \frac{2h^2 - m_E^{(2)}\bar{y}}{m_E^{(2)}\bar{y}^3}. \tag{3.56}$$

According to (3.56), the function $\psi(\bar{y})$ is convex up to the critical point $\bar{y}_c = 2h^2/m_E^{(2)}$, and the rest is concave. To evaluate how significant this concavity is, we can check the order of $\bar{y}_c$ and the curvature of ψ_L. Let $\mathcal{Y}$ be the dimension of the photon counts, and $\mathcal{E}$ be the energy level. Practically, we can assume that $\mathcal{Y} \gg 1$ and $\mathcal{E} \gg 1$. Then, the order of each

variable is

$$\bar{y} \sim O(\mathcal{Y}), \tag{3.57}$$

$$h \sim O(\mathcal{E}\mathcal{Y}), \tag{3.58}$$

$$m_E^{(2)} \sim O(\mathcal{E}^2), \tag{3.59}$$

$$\bar{y}_c \sim O\left(\frac{\mathcal{E}^2\mathcal{Y}^2}{\mathcal{E}^2}\right) = O(\mathcal{Y}^2). \tag{3.60}$$

For $\bar{y} > 10$, the probability of the event $y > \bar{y}_c$ is $O(10^{-63})$, which is extremely unlikely. However, like all non-convex problems, iterative solution could be trapped in a local minimum, particularly if the initial estimate of $\bar{y}$ is larger than $\bar{y}_c$. In our case, we set the initial estimate to be the true titanium pipe image fully filled with water, which we found avoids the concavity region adequately in practice.

As shown in (3.11), $\bar{\mathbf{y}}$ is a function of the fraction image $\mathbf{f} = \left[\mathbf{f}_{(1)}, \cdots, \mathbf{f}_{(N_l)}\right]$. Consequently, we want to find $\mathbf{f}$ that minimizes the data-fitting function in (3.55).

$$\hat{\mathbf{f}} = \arg\min_{\mathbf{f}\in\mathbb{F}} \Psi_L\big(\bar{\mathbf{y}}(\mathbf{f})\big), \tag{3.61}$$

where $\mathbb{F}$ is the volume fraction constraint,

$$\mathbb{F} = \Big\{ \mathbf{f} \in \mathbb{R}^{N_j \times N_l} \;\Big|\; f_{(j,l)} \geq 0,\ \textstyle\sum_l f_{(j,l)} \leq 1 \\ \text{for } j = 1, \cdots, N_j \text{ and } l = 1, \cdots, N_l \Big\}. \tag{3.62}$$

Finding analytic solution of (3.61) is not feasible. In general, it must be solved iteratively. The gradient of Ψ_L is:

$$\frac{\partial \Psi_L}{\partial f_{(j,l)}} = \sum_{i=1}^{N_i} \left[\frac{1}{2}\left(\frac{1}{\sigma_i^2} - g_i^2\right)\frac{\partial \sigma_i^2}{\partial f_{(j,l)}} - g_i \frac{\partial \eta_i}{\partial f_{(j,l)}} - \hat{\upsilon} g_i \frac{1}{b_i^2}\frac{\partial b_i}{\partial f_{(j,l)}} \right] \tag{3.63}$$

$$\simeq \sum_{i=1}^{N_i} \left[\frac{1}{2}\left(\frac{1}{\sigma_i^2} - g_i^2\right)\frac{\partial \sigma_i^2}{\partial f_{(j,l)}} - g_i \frac{\partial \eta_i}{\partial f_{(j,l)}} \right]. \tag{3.64}$$

where g_i is defined as:

$$g_i = \frac{h_i - \eta_i + \hat{v}/b_i}{\sigma_i^2} \,. \tag{3.65}$$

To derive the approximated gradient (3.64) from the exact gradient (3.63), we need to check the order of each partial derivative term. One can verify that

$$\frac{1}{b_i^2}\frac{\partial b_i}{\partial f_{(j,l)}} \sim O(\mathcal{E}) \,, \tag{3.66}$$

$$\frac{\partial \eta_i}{\partial f_{(j,l)}} \sim O(\mathcal{E}\mathcal{Y}) \,, \tag{3.67}$$

$$\frac{\partial \sigma_i^2}{\partial f_{(j,l)}} \sim O(\mathcal{E}^2\mathcal{Y}) \,. \tag{3.68}$$

Since $\mathcal{Y} \gg 1$, we can safely remove the partial derivative of b_i from (3.63). This approximation also applies to the preconditioning matrix that is derived later. The remaining partial derivative terms in (3.64) can be further expanded by finding the partial derivative of $\bar{y}_{(i,k)}$ with respect to $f_{(j,l)}$ as

$$\begin{aligned}
\frac{\partial \bar{y}_{(i,k)}}{\partial f_{(j,l)}} &= -\, I^0_{(i,k)}\, u_{(k,l)}\, a_{(i,j)}\, \exp\left(-\sum_{l=1}^{N_l} u_{(k,l)} \left[\mathbf{A}\mathbf{f}_l\right]_i\right) \\
&= -\, a_{(i,j)}\, u_{(k,l)}\, \bar{y}_{(i,k)} \,, \qquad (3.69) \\
\frac{\partial \eta_i}{\partial f_{(j,l)}} &= \sum_{k=1}^{N_k} m^{(1)}_{E_k} \frac{\partial \bar{y}_{(i,k)}}{\partial f_{(j,l)}} \\
&= -\, a_{(i,j)} \sum_{k=1}^{N_k} \left(u_{(k,l)}\, m^{(1)}_{E_k}\right) \bar{y}_{(i,k)} \,, \qquad (3.70) \\
\frac{\partial \sigma_i^2}{\partial f_{(j,l)}} &= \sum_{k=1}^{N_k} m^{(2)}_{E_k} \frac{\partial \bar{y}_{(i,k)}}{\partial f_{(j,l)}} \\
&= -\, a_{(i,j)} \sum_{k=1}^{N_k} \left(u_{(k,l)}\, m^{(2)}_{E_k}\right) \bar{y}_{(i,k)} \,. \qquad (3.71)
\end{aligned}$$

Plugging (3.70) and (3.71) into (3.64), we get

$$\frac{\partial \Psi_L}{\partial f_{(j,l)}} \simeq \sum_{i=1}^{N_i} a_{(i,j)} \left[\frac{1}{2}\left(\frac{1}{\sigma_i^2} - g_i^2\right) n^{(2)}_{(i,l)} - g_i\, n^{(1)}_{(i,l)}\right] \,, \tag{3.72}$$

where

$$n_{(i,l)}^{(1)} = -\sum_{k=1}^{N_k} u_{(k,l)}\, m_{E_k}^{(1)}\, \bar{y}_{(i,k)}\,, \tag{3.73}$$

$$n_{(i,l)}^{(2)} = -\sum_{k=1}^{N_k} u_{(k,l)}\, m_{E_k}^{(2)}\, \bar{y}_{(i,k)}\,. \tag{3.74}$$

3.3.2 Preconditioned gradient descent (asymptotic majorization - minimization)

So far, we have obtained the data-fitting objective function (3.55) and its gradient (3.72) with respect to the parameters to be estimated, i.e., the fraction image of the l^{th} material $f_{(j,l)}$. Using those equations, we want to find the minimum by taking the direction of the negative gradient. The following iteration is called the preconditioned gradient descent method:

$$\mathbf{f}^{(t+1)} = \mathbf{f}^{(t)} - \mathbf{N}_L^{-1} \nabla \Psi_L(\mathbf{f}^{(t)}). \tag{3.75}$$

Here, preconditioning matrix $\mathbf{N}_L$ defines the step size. Depending on the characteristics of the objective function, if the eigenvalues of $\mathbf{N}_L^{-1}$ are too large, the iterates may not converge at all. On the other hand, if the eigenvalues of $\mathbf{N}_L^{-1}$ is too small, it will take a long time to converge. Thus finding appropriate step size is crucial for fast and stable iterative computation.

For quadratic objective functions, the Hessian matrix is the asymptotically optimal one-step (analytic) choice for $\mathbf{N}_L$:

$$\mathbf{N}_L = \mathcal{H}\left\{\Psi_L(\mathbf{f})\right\}, \tag{3.76}$$

where $\mathcal{H}\left\{\cdot\right\}$ is the Hessian operator, which is also called the curvature. However, computing $\mathbf{N}_L$ for non-linear problems is challenging. Furthermore, even when the system is quadratic, inverting a large matrix $\mathbf{N}_L$ can be computationally expensive. In this case, a surrogate function $\Psi_L^{(t)}$ at each iteration step t can be adopted, which is easier than the objective function to solve, and guaranteed to monotonically descend the objective function. Then, instead of minimizing the objective function, we minimize the surrogate function to move to the next iteration. This process is called majorization-minimization [46].

To decrease the objective function monotonically, we chose the majorizer $\Psi_L^{(t)}$ at each

iteration step t to satisfy the following majorization conditions:

$$\Psi_L^{(t)}(\mathbf{f}) \geq \Psi_L(\mathbf{f}) \quad \text{for all } \mathbf{f}\,, \tag{3.77}$$

$$\Psi_L^{(t)}(\mathbf{f}) = \Psi_L(\mathbf{f}) \quad \text{for } \mathbf{f} = \mathbf{f}^{(t)}\,. \tag{3.78}$$

There are many majorization tricks that satisfy (3.77) and (3.78). Here, we are going to use a quadratic upper bound. Since we have an analytic solution for the quadratic function, and the objective function is twice differentiable, we want to design a quadratic surrogate function that satisfies (3.77) and (3.78). If ψ_L is twice differentiable, then there exists $c \in (0, 1)$ such that

$$\begin{aligned}\psi_L(\mathbf{f}) = \; & \psi_L(\mathbf{f}^{(t)}) + \nabla\psi_L\big(\mathbf{f}^{(t)}\big)' \big(\mathbf{f} - \mathbf{f}^{(t)}\big) \\ & + \frac{1}{2}\big(\mathbf{f} - \mathbf{f}^{(t)}\big)' \mathcal{H}\Big\{\psi_L\big((1-c)\mathbf{f}^{(t)} + c\mathbf{f}\big)\Big\}\big(\mathbf{f} - \mathbf{f}^{(t)}\big)\,.\end{aligned} \tag{3.79}$$

Further, if there exists a constant matrix $\mathbf{N}_L$ such that $\mathbf{N}_L \succeq \mathcal{H}\{\psi_L(\mathbf{f})\}$ for any $\mathbf{f}$, i.e., if the curvature of $\psi_L(\cdot)$ is bounded, then we can design a quadratic surrogate function that satisfies (3.77) and (3.78) as

$$\begin{aligned}\psi_L(\mathbf{f}) &\leq \psi_L^{(t)}(\mathbf{f}; \mathbf{f}^{(t)}) \\ &= \psi_L(\mathbf{f}^{(t)}) + \nabla\psi_L\big(\mathbf{f}^{(t)}\big)' \big(\mathbf{f} - \mathbf{f}^{(t)}\big) \\ &\quad + \frac{1}{2}\big(\mathbf{f} - \mathbf{f}^{(t)}\big)' \mathbf{N}_L \big(\mathbf{f} - \mathbf{f}^{(t)}\big)\,.\end{aligned} \tag{3.80}$$

Then, the update finds the minimizer of $\psi_L^{(t)}(\mathbf{f}; \mathbf{f}^{(t)})$:

$$\mathbf{f}^{(t+1)} = \arg\min_{\mathbf{f}} \; \psi_L^{(t)}(\mathbf{f}; \mathbf{f}^{(t)}). \tag{3.81}$$

By repeating (3.80) and (3.81), we can monotonically descend to find a minimizer of the original objective function.

One of the problems applying (3.77) to the objective function (3.55) is that the gradient of the objective function is not Lipschitz continuous, i.e., the curvature goes to infinity as $\bar{y}_{(i,k)}$ approaches to zero, which can be seen from (3.56). For objective functions with unbounded curvatures, it is difficult to find a quadratic surrogate function that can descend the objective function monotonically.

To overcome this issue, we divide the objective function into two regions, $g_i(\mathbf{f}^{(t)}) > 0$ and $g_i(\mathbf{f}^{(t)}) \leq 0$, where $g_i(\cdot)$ is defined in (3.65). By numerical experiments, we found that the curvature of the objective function $\Psi_L(\mathbf{f})$ decreases almost monotonically as $g_i(\mathbf{f})$

approaches zero. In an ideal case, where y_i is infinitely large, finding $\bar{\mathbf{f}}$ that satisfies $g_i(\bar{\mathbf{f}}) = 0$ for all i leads to the optimal solution. This noiseless solution $\bar{\mathbf{f}}$ cannot be achieved in reality, but we can still expect that $\mathbf{f}$ tends toward the direction where $g_i(\mathbf{f}) \simeq 0$. In that case, we can find the maximum curvature at every iteration step t as

$$\mathbf{N}_L^{(t)} = \mathcal{H}\left\{\Psi_L(\mathbf{f}^{(t)})\right\}. \tag{3.82}$$

In addition, the curvature of the logarithmic term in (3.55) is of the order $O(1)$, while the other term is $O(\mathcal{Y})$. Thus, we neglect the curvature of the logarithmic term. The approximated majorization curvature can be computed as

$$\begin{aligned}\frac{\partial^2 \psi_{Li}}{\partial f_{(j_1,l_1)} \partial f_{(j_2,l_2)}} \simeq & \frac{2}{\sigma_i^2} \frac{\partial \eta_i}{\partial f_{(j_1,l_1)}} \frac{\partial \eta_i}{\partial f_{(j_2,l_2)}} \\ & + \frac{2g_i}{\sigma_i^2}\left(\frac{\partial \eta_i}{\partial f_{(j_1,l_1)}} \frac{\partial \sigma_i^2}{\partial f_{(j_2,l_2)}} + \frac{\partial \sigma_i^2}{\partial f_{(j_1,l_1)}} \frac{\partial \eta_i}{\partial f_{(j_2,l_2)}}\right) \\ & - 2g_i \frac{\partial^2 \eta_i}{\partial f_{(j_1,l_1)} \partial f_{(j_2,l_2)}} + \frac{2g_i^2}{\sigma_i^2} \frac{\partial \sigma_i^2}{\partial f_{(j_1,l_1)}} \frac{\partial \sigma_i^2}{\partial f_{(j_2,l_2)}} \\ & - g_i^2 \frac{\partial^2 \sigma_i^2}{\partial f_{(j_1,l_1)} \partial f_{(j_2,l_2)}}.\end{aligned} \tag{3.83}$$

Expanding all partial derivatives, and substituting, we have:

$$\begin{aligned}\frac{\partial^2 \psi_{Li}}{\partial f_{(j_1,l_1)} \partial f_{(j_2,l_2)}} \simeq & \, w_{(i,l_1,l_2)} \\ \equiv & \, 2a_{(i,j_1)} a_{(i,j_2)} \Bigg[\frac{1}{\sigma_i^2} n_{(i,l_1)}^{(1)} n_{(i,l_2)}^{(1)} \\ & + \frac{g_i}{\sigma_i^2}\left(n_{(i,l_1)}^{(1)} n_{(i,l_2)}^{(2)} + n_{(i,l_1)}^{(2)} n_{(i,l_2)}^{(1)}\right) \\ & - g_i n_{(i,l_1,l_2)}^{(1)} + \frac{g_i^2}{\sigma_i^2} n_{(i,l_1)}^{(2)} n_{(i,l_2)}^{(2)} - \frac{g_i^2}{2} n_{(i,l_1,l_2)}^{(2)}\Bigg],\end{aligned} \tag{3.84}$$

where $n_{(i,l)}^{(1)}$ and $n_{(i,l)}^{(2)}$ are defined in (3.73) and (3.74) respectively, and $n_{(i,l_1,l_2)}^{(1)}$ and $n_{(i,l_1,l_2)}^{(2)}$ are defined as

$$n_{(i,l_1,l_2)}^{(1)} = \sum_{k=1}^{N_k} u_{(k,l_1)} u_{(k,l_2)} \, m_{E_k}^{(1)} \, \bar{y}_{(i,k)}. \tag{3.85}$$

$$n_{(i,l_1,l_2)}^{(2)} = \sum_{k=1}^{N_k} u_{(k,l_1)} u_{(k,l_2)} \, m_{E_k}^{(2)} \, \bar{y}_{(i,k)}. \tag{3.86}$$

We can express $\mathbf{N}_L^{(t)}$ for each l^{th} material image in matrix form:

$$\mathbf{N}_{L(l)}^{(t)} = \mathbf{A}' \left(\sum_{l_2=1}^{N_l} \mathbf{W}_{(l,l_2)}^{(t)} \right) \mathbf{A} \,, \tag{3.87}$$

where $\mathbf{W}_{(l,l_2)}^{(t)} = \operatorname{diag}\Big[w_{(i,l,l_2)}^{(t)} \Big]_{i=1}^{N_i}$.

Most of the terms in (3.87) do not require additional computation as they are already found when computing the gradient, except $n_{(i,l_1,l_2)}^{(1)}$ and $n_{(i,l_1,l_2)}^{(2)}$. However, since the size of the matrix $\mathbf{N}_L$ is huge, inverting $\mathbf{N}_L$ is computationally expensive. Thus, a diagonal matrix $\mathbf{D}_L$ such that $\mathbf{D}_L \succeq \mathbf{N}_L$ (separable quadratic surrogate) is preferred [47], [48]:

$$\mathbf{D}_{L(l)}^{(t)} = \operatorname{diag}\left[\mathbf{A}' \left(\sum_{l_2=1}^{N_l} \overline{\mathbf{W}}_{(l,l_2)}^{(t)} \right) \mathbf{A}\mathbf{1} \right] \,, \tag{3.88}$$

where $\overline{\mathbf{W}}_{(l,l_2)}^{(t)} = \operatorname{diag}\Big[\Big| w_{(i,l,l_2)}^{(t)} \Big| \Big]_{i=1}^{N_i}$, and $\mathbf{1}$ is the all-ones vector. Forward-projection $\mathbf{A}\mathbf{1}$ can be pre-computed, thus additional overhead required at each step is mostly a single back-projection.

Finally, we need to address the physical constraints $\mathcal{F}$, where the set $\mathcal{F}$ is defined in (3.62). For the case of dual material decomposition, the feasible domain that satisfies (3.62) forms a triangle in the two dimensional plane. In general, the projection onto the simplex is moving the outlying points to the closest points on the simplex, which involves orthogonal projection, as shown in Fig. 3.7a. However, because the direction of the preconditioned gradient is almost orthogonal to the hypotenuse due to the ill-posed nature of the problem, the orthogonal projection cancels the update. Instead, we use a two-step projection scheme shown in (3.89) to (3.92). Fig. 3.7b shows the graphical representation of the two-step projection.

$$\mathbf{f}_{(1)}^{(t+1)} = \mathcal{P}_1 \left[\mathbf{f}_{(1)}^{(t)} - \mathbf{M} \left(\mathbf{D}_{L(1)}^{(t)} \right)^{-1} \nabla_{\mathbf{f}_{(1)}} \Psi_L^{(t)} \right] \tag{3.89}$$

$$\mathbf{f}_{(2)}^{(t+1)} = \mathcal{P}_2 \left[\mathbf{f}_{(2)}^{(t)} - \mathbf{M} \left(\mathbf{D}_{L(2)}^{(t)} \right)^{-1} \nabla_{\mathbf{f}_{(2)}} \Psi_L^{(t)};\ \mathbf{f}_{(1)}^{(t+1)} \right] \tag{3.90}$$

where $\mathcal{P}_1[t]$ and $\mathcal{P}_2[t;s]$ are element-wise projection operators defined as

$$\mathcal{P}_1[t] = \min\Big[\max\Big[t, 0\Big], 1\Big], \tag{3.91}$$

$$\mathcal{P}_2[t;s] = \min\Big[\max\Big[t, 0\Big], 1-s\Big], \tag{3.92}$$

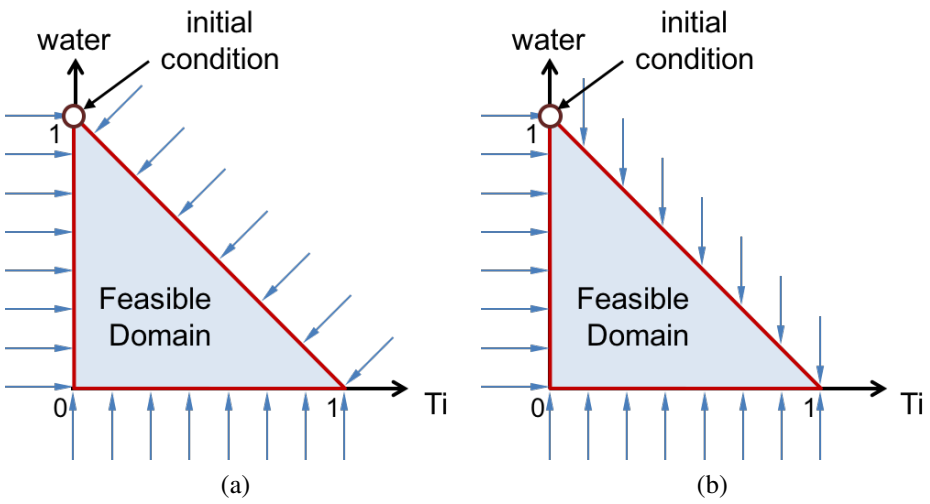


Figure 3.7: Simplex projection; (a) orthogonal projection, (b) two-step projection

and $\mathbf{M}$ is the masking matrix. The masking matrix $\mathbf{M}$ is essential for minimizing limited angle artifacts due to the lack of information by multiplying zeros to the pixel outside the 'known outline' of the object.

$$\begin{aligned} \mathbf{M} &= \text{diag}[m_j] \ , \\ m_j &= \mathbb{1}\{j \in \mathbb{Q}\}, \end{aligned} \tag{3.93}$$

where $\mathbb{1}\{t\}$ is the indicator function, 1 if t is true, and 0 otherwise. $\mathbb{Q}$ is the known bounded domain in the image where no objects are located outside of it. This approximated projection requires specific initial condition. To have this projection scheme work, material corresponds to f_1 needs to have higher attenuation coefficients than the other. As higher attenuation material images converge faster, lower attenuation material image can take the advantage of the faster converging higher attenuation material images.

3.3.3 Regularizer

Like any other data-fitting scheme, the data-fitting objective function in (3.55) tends toward overfitting, i.e., as the number of parameter grows, the objective function becomes overly sensitive and it tries to fit the parameters to the noisy data instead of its true underlying value. A statistical justification of the regularizer can be found from the Bayesian prior. First, find Maximum *A Posteriori* (MAP) estimator using the conditional probability

density function of $\mathbf{f}$ given the data $\mathbf{h}$ by applying the Bayes' rule,

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{f}\in\mathbb{F}} \mathbf{p}(\mathbf{f}|\mathbf{h}) = \arg\max_{\mathbf{f}\in\mathbb{F}} \frac{\mathbf{p}(\mathbf{h}|\mathbf{f})\,\mathbf{p}(\mathbf{f})}{\mathbf{p}(\mathbf{h})} = \arg\max_{\mathbf{f}\in\mathbb{F}} \mathbf{p}(\mathbf{h}|\mathbf{f})\,\mathbf{p}(\mathbf{f}) \tag{3.94}$$

Then, as usual, take the negative logarithm of the conditional density function:

$$\begin{aligned}\hat{\mathbf{f}} &= \arg\min_{\mathbf{f}\in\mathbb{F}} \left\{ -\log \mathbf{p}(\mathbf{h}|\mathbf{f}) - \log \mathbf{p}(\mathbf{f}) \right\} \\ &= \arg\min_{\mathbf{f}\in\mathbb{F}} \left\{ \Psi_L(\mathbf{f}) + \kappa \Psi_R(\mathbf{f}) \right\}, \end{aligned} \tag{3.95}$$

where κ is called the regularization parameter, which is the ratio between the constant gains in $\Psi_L(\mathbf{f})$ and $\Psi_R(\mathbf{f})$. In general, the true value of κ is neither known nor necessary, and it usually is tuned manually or automatically to have the best experimental result. The first term $\Psi_L(\mathbf{f})$ is called the data-fitting term, which was explained in §3.3.1. The second term $\Psi_R(\mathbf{f})$, which comes from the prior distribution, is called the regularizer, or the penalty function. Even though the role of regularizers can be explained in the probabilistic framework, the efficiency of the regularizer is evaluated based on the quality of the reconstructed images, rather than how accurately it can represent the prior, i.e., the actual probability density of $\mathbf{f}$.

To build the regularizer, we need to consider the characteristics of images. Most of the natural image pixels are continuous and correlated, i.e., the adjacent pixels likely have similar values, with some random variation. To build an approximate mathematical model, denote a 2-D image as a re-ordered 1-D array image f_j such that $j = \mathrm{x} + N_\mathrm{x}\mathrm{y}$. Then, assume f_j takes the independent and identically distributed Gaussian random walk process. Excluding boundary pixels, the Gaussian random walk process for a Cartesian grid can be expressed as:

$$\left.\begin{aligned} f_j - f_{j-1} \\ f_j - f_{j-N_\mathrm{x}} \end{aligned}\right\} \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma^2) \;\text{ for } f_j \text{ not at the boundary}\,. \tag{3.96}$$

In this case, $\Psi_R(\mathbf{f})$ falls into the L_2-norm regularizer, which is also called the generalized Tikhonov regularizer,

$$\Psi_R(\mathbf{f}) = \|\mathbf{C}\mathbf{f}\|_2^2\,, \tag{3.97}$$

where $\mathbf{C}$ is the first order finite difference matrix,

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_\mathrm{x} \\ \mathbf{C}_\mathrm{y} \end{bmatrix}\,. \tag{3.98}$$

Here $\mathrm{C_x}$ and $\mathrm{C_y}$ are the first order finite difference matrix in the x and y directions with appropriate boundary conditions, respectively. In the actual computation, finite difference matrix is not stored in the memory. Instead, matrix-equivalent functions are used, as it requires less memory and is faster computationally than the stored matrices computation. In general, this ℓ_2-norm regularization tends to smooth out the image excessively. If the actual image is not smooth, ℓ_1-norm is preferred as it preserves "edges",

$$\Psi_R(\mathbf{f}) = \|\mathbf{C}\mathbf{f}\|_1 . \tag{3.99}$$

This type of regularizer is called the ℓ_1-norm total variation regularizer, also known as the Least Absolute Shrinkage and Selection Operator (LASSO). Solving the minimization problem with the ℓ_1-norm regularizer requires additional effort as it is not differentiable. A simple "walk-around" is to round the tip of the ℓ_1-norm regularizer to make it differentiable,

$$\Psi_R(\mathbf{f}) = \sum_{p=1}^{N_p} \psi_R\left([\mathbf{C}\mathbf{f}]_p\right) , \tag{3.100}$$

where $[\mathbf{t}]_p$ is the pth element of an arbitrary vector $\mathbf{t} \in \mathbb{R}^{N_p}$, and $\psi_R(t)$ is a scalar function defined as

$$\psi_R(t) = \delta_R^2 \left(\sqrt{1 + (t/\delta_R)^2} - 1 \right) . \tag{3.101}$$

$\psi_R(t)$ is twice differentiable, thus much easier to solve using the gradient-based method. This smoothed corner version of the approximated ℓ_1-norm regularizer is called the hyperbola potential regularizer. Its first and second derivatives are

$$\dot{\psi}_R = \frac{t}{\sqrt{1 + (t/\delta_R)^2}} , \tag{3.102}$$

$$\ddot{\psi}_R = \frac{1}{(1 + (t/\delta_R)^2)^{3/2}} . \tag{3.103}$$

Using (3.102) and (3.103), the gradient and the Hessian of Ψ_R can be written as

$$\nabla \Psi_R = \mathbf{C}' \left[\dot{\psi}_R([\mathbf{C}\mathbf{f}]_p) \right]_{p=1}^{N_p} , \tag{3.104}$$

$$\mathcal{H}\left\{\Psi_R\right\} = \mathbf{C}' \operatorname{diag}\left[\ddot{\psi}_R([\mathbf{C}\mathbf{f}]_p) \right] \mathbf{C} . \tag{3.105}$$

A separable preconditioner for the regularization term can be found from the spectral

radius. Since the curvature of the hyperbola potential is bounded, i.e., $\ddot{\psi}_R \leq 1$,

$$\mathcal{H}\left\{\Psi_R\right\} \succeq \mathbf{C}'\mathbf{I}\mathbf{C} \succeq \mathbf{D}_R \triangleq c_R\mathbf{I}\,, \tag{3.106}$$

where c_R is the spectral radius of the matrix $\mathbf{C}$. Assuming matrix $\mathbf{C}$ is circulant, c_R can be found from the discrete Fourier transform of the point spread function, i.e., take one row of the matrix, then shift it left or right to make it properly centered.

$$\begin{aligned}
\mathcal{F}\{\mathbf{C}_{\mathrm{x}}\} &= 1 - e^{-iw}\,,\\
\mathcal{F}\{\mathbf{C}_{\mathrm{x}}'\mathbf{C}_{\mathrm{x}}\} &= \mathcal{F}^*\{\mathbf{C}_{\mathrm{x}}\}\mathcal{F}\{\mathbf{C}_{\mathrm{x}}\} = 2 - 2\cos(w)\,,\\
\max\left(\mathcal{F}\{\mathbf{C}_{\mathrm{x}}'\mathbf{C}_{\mathrm{x}}\}\right) &= \max\left(\mathcal{F}\{\mathbf{C}_{\mathrm{y}}'\mathbf{C}_{\mathrm{y}}\}\right) = 4\,.
\end{aligned} \tag{3.107}$$

As the circulant matrices of the same rank share the same eigenvector set (simultaneously diagonalizable), the spectral radius c_R can be found from the superposition of the maximum eigenvalues,

$$c_R = \max\left(\mathcal{F}\{\mathbf{C}_{\mathrm{x}}'\mathbf{C}_{\mathrm{x}}\}\right) + \max\left(\mathcal{F}\{\mathbf{C}_{\mathrm{y}}'\mathbf{C}_{\mathrm{y}}\}\right) = 8\,. \tag{3.108}$$

Finally, the transition equation for the preconditioned gradient descent in (3.89) and (3.90) can be rewritten after including the regularization terms as

$$\begin{aligned}
\mathbf{f}_{(l)}^{(t+1)} &= \mathcal{P}_l\left[\mathbf{f}_{(l)}^{(t)} - \mathbf{M}\left(\mathbf{D}_{(l)}^{(t)}\right)^{-1}\nabla_{\mathbf{f}_{(l)}}\Psi^{(t)}\right]\\
&= \mathcal{P}_l\left[\mathbf{f}_{(l)}^{(t)} - \mathbf{M}\left(\mathbf{D}_{L(l)}^{(t)} + \kappa_{(l)}\mathbf{D}_{R(l)}\right)^{-1}\left(\nabla_{\mathbf{f}_{(l)}}\Psi_L^{(t)} + \kappa_{(l)}\nabla_{\mathbf{f}_{(l)}}\Psi_{R(l)}^{(t)}\right)\right].
\end{aligned} \tag{3.109}$$

Note that the regularizer $\Psi_{R(l)}^{(t)}$ does not have to be the same for every l. For example, a steel fraction image has larger, flatter and sharp-edged objects than a water fraction image. We apply a larger value of κ to the steel fraction image to reduce the limited angle artifacts, and a lower κ to the water fraction image to maintain small objects in the water fraction image.

While edge-preserving regularizer is good for the denoising without much loss of high frequency details, it is not very effective to remove the limited angle artifacts. Furthermore, limited angle artifacts in a high attenuation material fraction map can greatly affect the quality of the lower attenuation material fraction map. For example, an attenuation coefficient of Fe at 100 keV is 17 times greater than that of the water, which means that a 5.9% error in iron fraction image can induce a 100% error in the water fraction image. By numerical experiments, we found that a combination of the edge-preserving hyperbolic regularizer and the sparsity-based ℓ_0-norm regularizer can effectively remove limited angle

artifacts in the hard material fraction image with minimal sacrifice of the small fraction values, while applying only the edge-preserving hyperbolic regularizer to the soft material fraction image.

A fraction image of the higher attenuation material $l = 1$, especially when the material is metal, is almost binary, in which the fraction value is either 0 or 1 except at the boundaries. In other words, pixels that have values other than 0 or 1 are "sparse", thus, encouraging the sparsity usually improves the quality of the image significantly. Since the fraction image requires a quantitative result, the ℓ_1-norm is not preferred as the corresponding soft-thresholding shrinkage operator decreases the original value everywhere. Instead, ℓ_0-norm regularizer is used, where the corresponding hard-thresholding shrinkage operator only affects a finite region.

$$\hat{\mathbf{f}} = \underset{\mathbf{f}\in\mathbb{F}}{\arg\min} \left\{ \Psi_L(\mathbf{f}) + \sum\nolimits_l \kappa_{(l)} \Psi_{R(l)}(\mathbf{f}_{(l)}) + \kappa_0 ||\mathbf{f}_{(1)}||_0 \right\}, \tag{3.110}$$

where $||\mathbf{t}||_0$ is defined as

$$||\mathbf{t}||_0 = \sum\nolimits_i \mathbb{1}\{t_i \neq 0\}. \tag{3.111}$$

Note that $||\mathbf{f}_{(1)}||_0$ is not differentiable. Thus it cannot be solved by the naive gradient descent method. Instead, the overall objective function (3.110) including the sparsity based regularizer can be solved using the iterative shrinkage-thresholding algorithm (ISTA), which is a member of the proximal gradient (PG) methods [49]. ISTA is an extension of the gradient descent method, where a shrinkage operator is iteratively applied after taking the gradient descent step without the sparsity regularizer. For ℓ_0 norms, the corresponding proximal operator is the element-wise hard-thresholding shrinkage operator, defined as

$$\mathcal{T}[t; s] = t \cdot \mathbb{1}\{|t| > s\}. \tag{3.112}$$

Then, (3.110) can be solved by taking the following iterative step:

$$\mathbf{f}_{(1)}^{(t+1)} = \mathcal{T}\left[\mathcal{P}_1\left[\mathbf{f}_{(1)}^{(t)} - \mathbf{M}\left(\mathbf{D}_{(1)}^{(t)}\right)^{-1} \nabla_{\mathbf{f}_{(1)}} \Psi^{(t)}\right]\right]. \tag{3.113}$$

Note that the ℓ_0 norm regularizer is only applied to the hard material fraction image $l = 1$, so there is no change for $l = 2$. (3.113) can be further simplified by merging the hard-thresholding operator $\mathcal{T}[\cdot]$ and the projection operator $\mathcal{P}_1[\cdot]$, which become the truncated

hard-thresholding operator:

$$\mathbf{f}_{(1)}^{(t+1)} = \mathcal{Q}\left[\mathbf{f}_{(1)}^{(t)} - \mathbf{M}\left(\mathbf{D}_{(1)}^{(t)}\right)^{-1} \nabla_{\mathbf{f}_{(1)}} \Psi^{(t)};\ \kappa_0 \left(\mathbf{D}_{(1)}^{(t)}\right)^{-1} \mathbf{1}\right], \tag{3.114}$$

where the element-wise operator $\mathcal{Q}\ [t; s]$ is defined as

$$\mathcal{Q}\ [t; s] = \begin{cases} 0 & \text{if } \ t < s \\ t & \text{if } \ s \leq t < 1\ . \\ 1 & \text{if } \ t \geq 1 \end{cases} \tag{3.115}$$

While ISTA can be applied to the gradient method without much effort, the convergence speed of ISTA is slow. To accelerate the convergence of the proximal gradient method, Nesterov's momentum-based method is widely used because of its simplicity. Beck and Teboulle [50] introduced a fast iterative shrinkage-thresholding algorithm (FISTA) that has a convergence rate $O(1/t^2)$ using Nesterov's linear momentum acceleration method [51]. While the original FISTA algorithm is designed for the Lipschitz continuous function, Zuo and Lin [52] further generalized the algorithm to the quadratic, preferably separable, surrogate problem that leads to the faster convergence. It is called the generalized accelerated proximal gradient method (GAPG). The convergence rate of the original Nestrov's method is guaranteed only when the objective function is strongly convex, i.e., there should be a quadratic function with a constant $\gamma > 0$ such that

$$\Psi(\mathbf{f}) \geq \Psi(\mathbf{g}) + \nabla\Psi(\mathbf{g})'(\mathbf{f} - \mathbf{g}) + \frac{\gamma}{2}||\mathbf{f} - \mathbf{g}||_2^2 \quad \forall\, \mathbf{f}, \mathbf{g} \in \mathbb{R}^n. \tag{3.116}$$

The data-fitting objective function in (3.55) does not satisfy the strong convexity condition in (3.116). As a result, overshoot can happen in the update step which slows down the convergence rate. O'Donoghue and Candès [53] suggested two possible conditions; the objective function or the gradient value to re-initialize the acceleration parameter to avoid overshoot. While O'Donoghue and Candès showed that there is no obvious difference in performance between those two options, we chose gradient value-based restart scheme combined with GAPG as an acceleration method in this study. The combined scheme is written in pseudocode as Algorithm 1.

3.3.4 Gauss-Legendre quadrature and dithering update

The line integral in Beer's law (3.6) is derived to compute the attenuation rate at an infinitesimal point. Since the detector has finite size pixels, we need to average the attenuation rate

Algorithm 1 Adaptive restart GAPG

$\mathbf{z}_{(1)}^{(0)} = \mathbf{f}_{(1)}^{(0)}, \mathbf{z}_{(2)}^{(0)} = \mathbf{f}_{(2)}^{(0)}, d^{(0)} = 1$

for $t = 0, 1, \cdots$ **do**

$\mathbf{z}^{(t)} = \left[\mathbf{z}_{(1)}^{(t)}, \mathbf{z}_{(2)}^{(t)}\right]$

$\mathbf{f}_{(1)}^{(t+1)} = \mathcal{Q}\left[\mathbf{z}_{(1)}^{(t)} - \mathbf{M}\left(\mathbf{D}_{(1)}^{(t)}\right)^{-1} \nabla_{\mathbf{f}_{(1)}} \Psi\left(\mathbf{z}^{(t)}\right);\ \kappa_0 \left(\mathbf{D}_{(1)}^{(t)}\right)^{-1} \mathbf{1}\right]$

$\mathbf{f}_{(2)}^{(t+1)} = \mathcal{P}_2\left[\mathbf{z}_{(2)}^{(t)} - \mathbf{M}\left(\mathbf{D}_{(2)}^{(t)}\right)^{-1} \nabla_{\mathbf{f}_{(2)}} \Psi\left(\mathbf{z}^{(t)}\right);\ \mathbf{z}_{(1)}^{(t)}\right]$

if $\sum_l \left(\nabla_{\mathbf{f}_{(l)}} \Psi\right)' \left(\mathbf{f}_{(l)}^{(t+1)} - \mathbf{f}_{(l)}^{(t)}\right) > 0$ **then**

$d^{(t)} = 1$

end if

$d^{(t+1)} = \left(1 + \sqrt{1 + 4(d^{(t)})^2}\right)/2$

$\mathbf{z}_{(1)}^{(t+1)} = \mathbf{f}_{(1)}^{(t+1)} + \frac{d^{(t)}-1}{d^{(t+1)}}\left(\mathbf{f}_{(1)}^{(t+1)} - \mathbf{f}_{(1)}^{(t)}\right)$

$\mathbf{z}_{(2)}^{(t+1)} = \mathbf{f}_{(2)}^{(t+1)} + \frac{d^{(t)}-1}{d^{(t+1)}}\left(\mathbf{f}_{(2)}^{(t+1)} - \mathbf{f}_{(2)}^{(t)}\right)$

$t = t + 1$

end for

along the span of the detector pixel for a more accurate model.

$$\bar{y}_i(\mathcal{E}_k) = \left[\frac{1}{\Delta s} \int_{s_{i1}}^{s_{i2}} I_i^0(s, \mathcal{E}_k) \exp\left(-\int_{\mathcal{L}_s} \mu(s, \mathcal{E}_k)\, \mathrm{d}l\right) \mathrm{d}s\right] \Delta\mathcal{E}\,, \tag{3.117}$$

where Δs is the width of the detector pixel, s_{i1} and s_{i2} are the end points of the detector pixel corresponding to the i^{th} ray.

Often, when the detector pixel is reasonably small, the negative exponential part is assumed to be linear along the span, so that we can move the integral along s inside the exponential function.

$$\bar{y}_i(\mathcal{E}_k) = I_i^0(\mathcal{E}_k)\, \exp\left(-\frac{1}{\Delta s} \int_{s_{i1}}^{s_{i2}} \int_{\mathcal{L}_s} \mu(\boldsymbol{s}, \mathcal{E}_k)\, \mathrm{d}l\, \mathrm{d}\boldsymbol{s}\right) \Delta\mathcal{E}\,. \tag{3.118}$$

Joseph and Spital 1981 [54] showed that this type of approximation can lead to a streak artifact called the 'exponential edge-gradient effect', when the detector size is not small and there exists a sharp transition of the attenuation in the object. This exponential edge-gradient effect is usually neglected as modern x-ray detectors have very small pixel size, which even achieved micrometer scale [55]. However, there can be an application where small pixel detectors are expensive, mainly due to the complexity of the electrical circuit

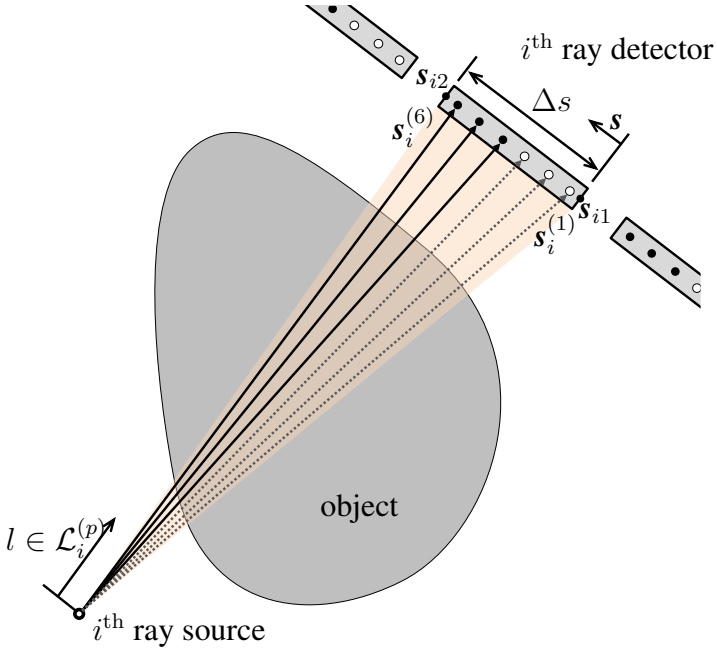


Figure 3.8: Source, ray, detector geometry and Gaussian-Legendre quadrature dithering points

and the data transferring latency. Moreover, large pixel detectors help to maximize the efficiency of the photo-electric interaction, while minimizing the charge sharing [56]. Instead of using this "internal" strip integral approach, we want to use a numerical integration to compute (3.117) directly, which will be called the "external" strip integral hereinafter. For the numerical implementation, assume $I_i^0(s, \mathcal{E})$ is constant along the span of the ith ray detector pixel, as the initial spectrum is less variant than the attenuation part. Assume the exponential part is an arbitrary polynomial function of s, and use the N_p point Gauss-Legendre quadrature for the integral over s, which can yield the exact value up to $2N_p - 1$ degree polynomial functions [57].

$$\begin{aligned}\bar{y}_i(\mathcal{E}_k) &= \frac{1}{\Delta s} I_i^0(\mathcal{E}_k) \sum_{p=1}^{N_p} \omega^{(p)} \exp\left(- \int_{\mathcal{L}_i^{(p)}} \mu(\boldsymbol{s}_i^{(p)}, \mathcal{E}_k)\, \mathrm{d}l \right) \frac{\Delta s}{2} \Delta\mathcal{E} \\ &= \frac{1}{2} I_i^0(\mathcal{E}_k) \sum_{p=1}^{N_p} \omega^{(p)} \exp\left(- \int_{\mathcal{L}_i^{(p)}} \mu(\boldsymbol{s}_i^{(p)}, \mathcal{E}_k)\, \mathrm{d}l \right) \Delta\mathcal{E}\,, \qquad (3.119) \\ \boldsymbol{s}_i^{(p)} &= \frac{\boldsymbol{s}_{i1} + \boldsymbol{s}_{i2}}{2} + \tau^{(p)} \frac{\boldsymbol{s}_{i2} - \boldsymbol{s}_{i1}}{2}\,, \qquad (3.120)\end{aligned}$$

where $\tau^{(p)}$ and $\omega^{(p)}$ are the abscissa and the weight of the Gauss-Legendre quadrature.

Lowan *et al.* 1943 [58] computed the abscissas and weights table up to 16 point Gauss-Legendre quadrature. Here, we use six point Gauss-Legendre quadrature, where $\tau^{(p)}$ and $\omega^{(p)}$ with 7 digit significant numbers are given by

p	$\tau^{(p)}$	$\omega^{(p)}$
1	-0.9324695	0.1713245
2	-0.6612094	0.3607616
3	-0.2386192	0.4679139
4	0.2386192	0.4679139
5	0.6612094	0.3607616
6	0.9324695	0.1713245

Table 3.1: Six point Gauss-Legendre quadrature coefficients

Compared to the internal strip integral (3.118), implementing the external strip integral (3.120) requires more computational cost as it requires computing the attenuation part N_p times, while the internal strip integral requires it only once. Instead, we propose a dithering update algorithm, which is similar to the ordered subset algorithm in that it only uses a part of the projection matrix at each iteration, but different as the separated data for each subset is not available.

Hudson and Larkin 1994 [59] suggested the ordered subset algorithm which takes a subset of the system matrix and corresponding data at each iteration, then compute the forward and back projection for the chosen set of detectors to update the parameters (image) to be estimated. To apply the ordered subset algorithm, each subset must satisfy the subset balance condition.

$$\nabla\Psi \simeq N_q \nabla\Psi^{(1)} \simeq \cdots \simeq N_q \nabla\Psi^{(N_q)}\,, \tag{3.121}$$

where N_q is the number of subsets, and $\Psi^{(q)}$ is the subset gradient. In a similar way, we can consider a set of Gauss-Legendre points as a separate detector pixel. The simplest way to achieve the subset balance for the Gauss-Legendre points is to split the Gauss-Legendre points into two equally weighted, self-balanced subsets, $\mathbb{P}^{(0)} = \{1, 2, 4\}$ and $\mathbb{P}^{(1)} = \{3, 5, 6\}$. With those subsets, based on experiments, the difference between the internal and the external strip integral within the subset is not significant, thus we can push the summation and weights into the exponential part, i.e., a partially internal strip integral. Then, define $\bar{y}_i^{(q)}(\mathcal{E})$ as the Beer's law equation corresponding to the q^{th} subset of

the Gauss-Legendre points:

$$\bar{y}_i^{(q)}(\mathcal{E}_k) = I_i^0(\mathcal{E}_k)\, \exp\left(- \sum_{p \in \mathbb{P}(q)} \omega^{(p)} \int_{\mathcal{L}_i^{(p)}} \mu(\boldsymbol{s}_i^{(p)}, \mathcal{E}_k)\, \mathrm{d}l \right) \Delta\mathcal{E} \quad \text{for} \quad q = 0 \text{ or } 1\,, \tag{3.122}$$

and define the projection matrix $\mathbf{A}_p$ as a line integral projection matrix with its element $a_{(i,j)}^{(p)}$ being the Radon transform of the jth spatial basis along the p^{th} Gauss-Legendre point of the i^{th} ray. Also, define $\mathbf{A}^{(q)}$ as

$$\mathbf{A}^{(q)} = \sum_{p \in \mathbb{P}(q)} \omega^{(p)} \mathbf{A}_p \quad \text{for} \quad q = 0 \text{ or } 1\,. \tag{3.123}$$

After applying the material decomposition and discretization written in §3.2.2, the discrete version of (3.122) can be written as

$$\bar{y}_{(i,k)}^{(q)} = I_{(i,k)}^0\, \exp\left(- \sum_{l=1}^{N_l} u_{(k,l)} \left[\mathbf{A}^{(q)} \mathbf{f}_l \right]_i \right) \Delta\mathcal{E} \quad \text{for} \quad q = 0 \text{ or } 1\,. \tag{3.124}$$

Let q be the remainder of $t/2$, and replace every instance of $\bar{y}_{(i,k)}$ with $\bar{y}_{(i,k)}^{(q)}$, and $\mathbf{A}$ with $\mathbf{A}^{(q)}$ in (3.42), (3.45), (3.46), (3.65), (3.73), (3.74), (3.86), (3.88), and (3.109), we get the transition equation for the dithering update:

$$\mathbf{f}_{(l)}^{(t+1)} = \mathcal{P}_l \left[\mathbf{f}_{(l)}^{(t)} - \mathbf{M} \left(\mathbf{D}_{L(l)}^{(t,q)} + \kappa_{(l)} \mathbf{D}_{R(l)} \right)^{-1} \left(\nabla_{\mathbf{f}_{(l)}} \Psi_L^{(t,q)} + \kappa_{(l)} \nabla_{\mathbf{f}_{(l)}} \Psi_{R(l)}^{(t)} \right) \right], \tag{3.125}$$

where

$$\mathbf{D}_{L(l)}^{(t,q)} = \mathrm{diag}\left[\left(\mathbf{A}^{(q)} \right)' \left(\sum_{l_2=1}^{N_l} \mathbf{W}_{(l,l_2)}^{(t,q)} \right) \mathbf{A}^{(q)} \mathbf{1} \right], \tag{3.126}$$

$\mathbf{W}^{(t,q)}$ and $\nabla \Psi_{L(l)}^{(t,q)}$ are the same as $\mathbf{W}^{(t)}$ and $\nabla \Psi_{L(l)}^{(q)}$ with all $\bar{y}_{(i,k)}$ and $\mathbf{A}$ being replaced by $\bar{y}_{(i,k)}^{(q)}$ and $\mathbf{A}^{(q)}$.

3.4 GPU Implementation (CUDA)

Due to the physical problem of modern central processing units (CPUs) such as heat dissipation, development of faster CPU is increasingly difficult. Instead, CPU manufacturers are trying to find a breakthrough from the multi processor. Currently, two types of multi

processors exist, one is the multi-core processor represented by Intel chipsets, the other is massively-parallel processor represented by Nvidia GPUs. While the former is still inherited from the conventional CPUs, the latter has totally different architectures.

GPU adopts the single instruction - multi data (SIMD) concept to control massively many but individually slow cores efficiently, and provides broad memory bandwidth to read and write data that are requested by massively many cores. This concept enables GPUs to compute a simple but enormously large scale data efficiently. These characteristics of GPUs also distinguish the GPU programming from the programming for CPUs. Also, GPU provides several hardware units which are originally designed for the graphic processing. For example, texture units (shown as Tex in Fig. 3.9), whose primary purpose was the texture mapping of a two dimensional image to the three dimensional surface, but can be used for any problem that requires two or three-dimensional interpolation. Nvidia GPUs also provide special function units (SFUs) to approximately compute CPU-intensive functions such as trigonometric functions.

In 2007, Nvidia released libraries, compilers, and toolboxes called Compute Unified Device Architecture (CUDA), which enables their GPUs to be used for the general purpose computing, in opposition to the graphic processing. This usage of GPU is called GPGPU (general-purpose GPU). GPU computation can benefit the CT reconstruction as its system is simple and its data size is massive. However, one problem with this approach is that CT reconstruction system is sparse. Naive immigration of CPU code to GPU is not efficient as too many cores become idle due to the SIMD restriction. Also, high order interpolation scheme such as Zwart-Powell box spline can make the CT system to be compute-intensive, which is inefficient system for GPU computing. In this chapter, we will explain how we can overcome those issues by using the simple compacting algorithm and hardware units in GPUs. Fig. 3.10 shows the flow diagram of the reconstruction algorithm.

Figure 3.9: Architecture of streaming multiprocessors in Nvidia Kepler GPUs [60]. 32 parallel cores are scheduled together as a warp. Special Function Units (SFUs) provide hardware computation of the CPU-bound functions such as trigonometric functions with slightly lower accuracy. Each thread can have up to 255 32-bit registers (fastest local memory). Texture memory provides its own read-only cache space and hardware unit to enable GPU to fetch continuously addressed data (linear interpolation) fast.

3.4.1 Texture unit and interpolation

Computation of Zwart-Powell elements in GPU is compute-bounded, i.e., computation clock cycles take more time than memory access clock cycles, thus latency of memory access is hidden by the compute latency. In this situation, improving memory access does not help to reduce the computation time. A common approach to boost the computation speed of the ZP element is to pre-compute a table of values, then interpolate to get the approximated values. However, fetching interpolated values at non-uniformly chosen points from a two dimensional array stored in a conventional C type linear array is inefficient due to the uncoalesced memory access. In CUDA, CUDA array and texturing are specifically designed to resolve this problem. CUDA array and texturing are originally designed for graphic processing to render a 2-D texture image on the non-uniform grid surface of the body, but it can be used for any type of data that requires localized interpolation. CUDA array uses special addressing technique called block-linear addressing which is faster in accessing spatially -localized data in multi-dimensional arrays. Unfortunately, Matlab does not provide built-in memory structure to store CUDA array in Workspace, thus data must be read from the Matlab GPU array then stored to CUDA array every time the GPU kernel is called. Once a memory is bound to texture, texture provides its own read-only cache space and hardware unit to enable GPU to fetch spatially-localized data rapidly [61]. For the bi-linear interpolation, four adjacent elements near the point to be estimated are required at every fetch operation. Along with fast uncoalesced memory access and linear interpolation, texture memory provides several useful hardware units such as automatic boundary addressing mode (clamp, wrap, and mirror), thus user can focus on the main algorithm without computing those vexing boundary conditions. As a result, ZP elements are computed mostly by hardware interpolation, and only angles and distances are computed by software. Based on the experiments, with only 1 Kbyte table of 8 angle intervals between 0 to 45 degrees and 32 distance intervals between 0 to 2.5 pixel units, the result of the single-precision, fourth order Gaussian -Legendre quadrature integral of the first-order interpolated ZP elements only differ by 0.002% in terms of the normalized absolute error from that of the double-precision, sixth order Gaussian-Legendre quadrature integral of the analytic ZP basis.

3.4.2 Compacting operations

The projection matrix for CT is sparse, for example, reconstruction of 128×128 $(N_x \times N_y)$ pixels by 192×128 (sources $\times$ detectors) rays with ZP basis requires six million nonzero

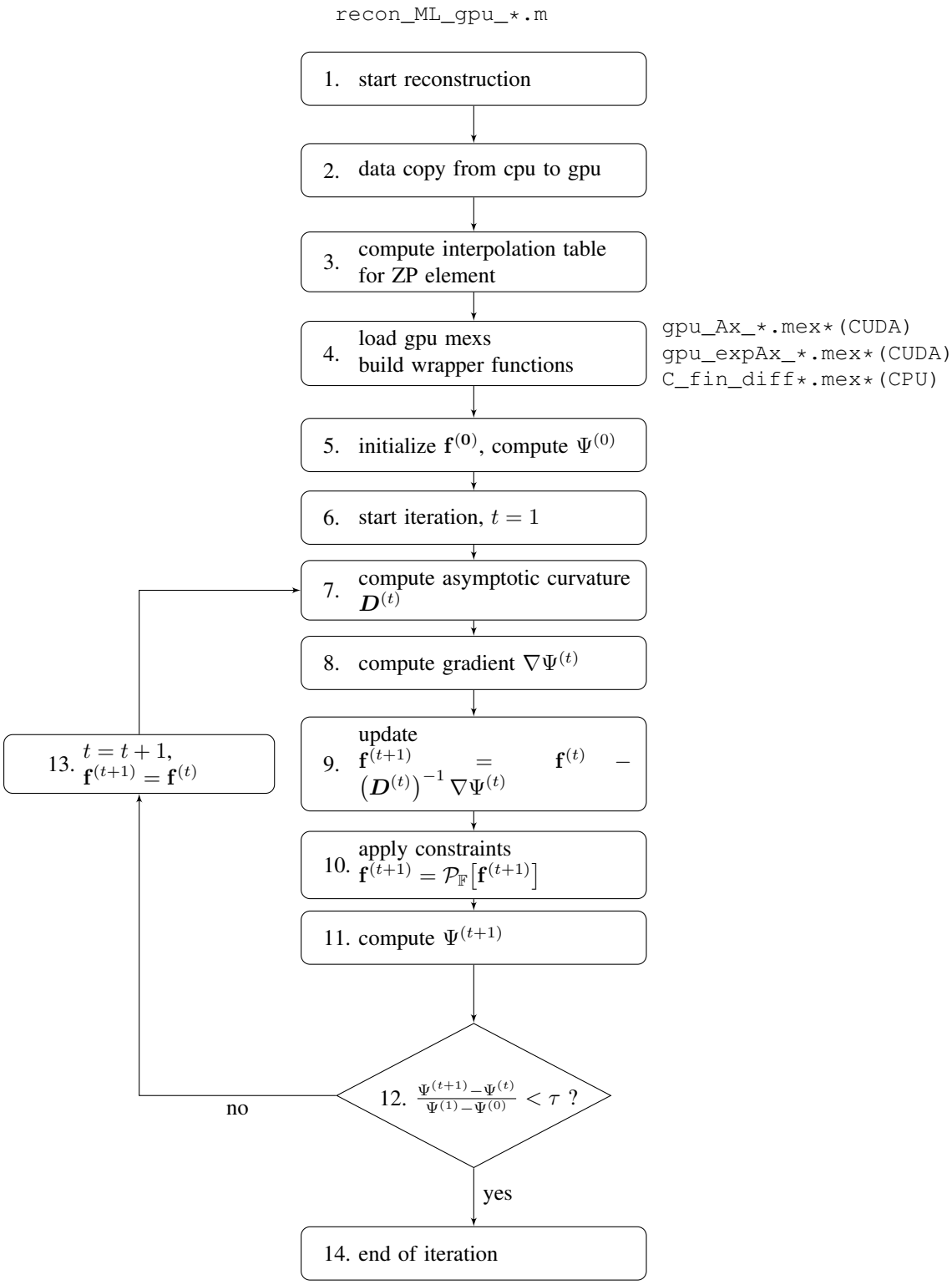


Figure 3.10: Flow diagram of the reconstruction algorithm

elements among the total of 400 million elements, which is about 1.5 percent of the total elements. To avoid unnecessary computation in a sparse system, one approach is to store the pointer array which indicates where the nonzero elements are, along with the non-zero data array. With this method, all processors in a warp can be fully utilized, but the size of the array is too big to be stored in a fast memory, thus it inevitably increases the number of access to the slowest global memory. Even worse, for three-dimensional CT, too much memory space is required which makes it less practicable. Another way to proceed is to loop over the whole elements, then compute the basis function when an element passes a conditional statement, i.e., compute the basis function only when the distance between the center of an image pixel and ray is within the maximum support size. This is the simplest approach, but is slow as it needs to compute the distances for every combination of image pixels and rays.

To utilize the threads efficiently, we introduce a simple search algorithm which does not involve additional computation other than already used distance and angle calculations to reduce unnecessary work.

For the forward projection, cosine approximation is used. Assume the image matrix is stored in a row-major sense, and let the angle between the orthogonal line of the ray and the horizontal line be θ_i. Define $d_{i,j}$ as the smallest distance between the center of image pixel j and ray i minus the maximum support length

$$d_{i,j} = \frac{\det\left\{ \left[\boldsymbol{q}_i - \boldsymbol{p}_i \, , \, \boldsymbol{r}_j - \boldsymbol{p}_i \right] \right\}}{||\boldsymbol{q}_i - \boldsymbol{p}_i||_2} - l_s \, , \tag{3.127}$$

where $\det\{\cdot\}$ is the determinant of a matrix, $\boldsymbol{r}_j = \left[r_j^{(x)}, r_j^{(y)} \right]'$ is the location vector of the jth image pixel, $\boldsymbol{p}_i = \left[p_i^{(x)}, p_i^{(y)} \right]'$ and $\boldsymbol{q}_i = \left[q_i^{(x)}, q_i^{(y)} \right]'$ are the location vector of the source and the detector pixel corresponding to the ith ray respectively, and l_s is the maximum support length from the center. For small angle θ, d can be approximated by a Taylor series expansion as

$$\begin{aligned} d_{i,j} &= \cos(\theta_i)\, l_{i,j} \\ &= l_{i,j} \left(1 - \frac{{\theta_i}^2}{2!} + \frac{{\theta_i}^4}{4!} - \cdots \right) \\ &\simeq l_{i,j}. \end{aligned} \tag{3.128}$$

Based on the approximation (3.128), if the image pixel j is not within the support range, the next image pixel to be searched, j^*, is chosen to be $j^* = j + \lfloor d \rfloor$. As $l \geq d$, jumping

$\lfloor d \rfloor$ pixels does not skip any pixels within the support, i.e. l is monotonically decreasing throughout the searching. Once the location of the pixel passes the support range of the ray, the next image pixel is chosen to be the next line. For example, if an image is in the $n \times m$ grid, the next image pixel is $n\,(\lfloor {}^{j}/_{n} \rfloor + 1)$. Distance d is required to check whether the pixel is within the support, thus this method does not need any additional computation. Fig. 3.11 shows a graphical explanation of the compacting scheme.

The upper bound of the total iteration count for the searching can be computed from the Taylor expansion of the cosine function. For simplicity, drop the index i and j, and denote l_0 as the distance between the first image pixel of the row and the ray minus maximum support, and l_k as the distance between the image pixel after kth search and the ray minus maximum support.

$$\begin{aligned} l_k &\simeq l_{k-1}\big(1-\cos(\theta)\big) \\ &\simeq l_0\big(1-\cos(\theta)\big)^k \\ &= l_0\left(1-1+\frac{\theta^2}{2!}-\cdots\right)^k \\ &\simeq l_0\left(\frac{\theta^2}{2!}\right)^k \end{aligned} \tag{3.129}$$

For the distance normalized by the image pixel width, iteration k stops when $l_k \leq 1$, and the maximum value of l_0 is equal or smaller than n. Thus the upper bound of the iteration count for the searching per row is

$$n\left(\frac{\theta^2}{2!}\right)^k = 1 \tag{3.130}$$

$$k = \log_{2/\theta^2}(n) = O(\log n) \tag{3.131}$$

Since we need to find k for each row, total iteration counts for the searching is $O(m \log n)$.

Similarly, back projection can be done using the projection angle. For uniformly distributed circular detector arrays, the angle between detectors is constant. Define this angle as θ_c, and angle ψ_j as the angle between the line that connects center of the image pixel j and the source, and the line that connects the source and the first detector pixel. The first ray index i^* that has nonzero element can be found by $i^* = i + \lfloor {}^{\psi_j}/_{\theta_c} \rfloor$. The iteration count for the searching is $O(1)$ in this case. For detectors not uniformly distributed or not

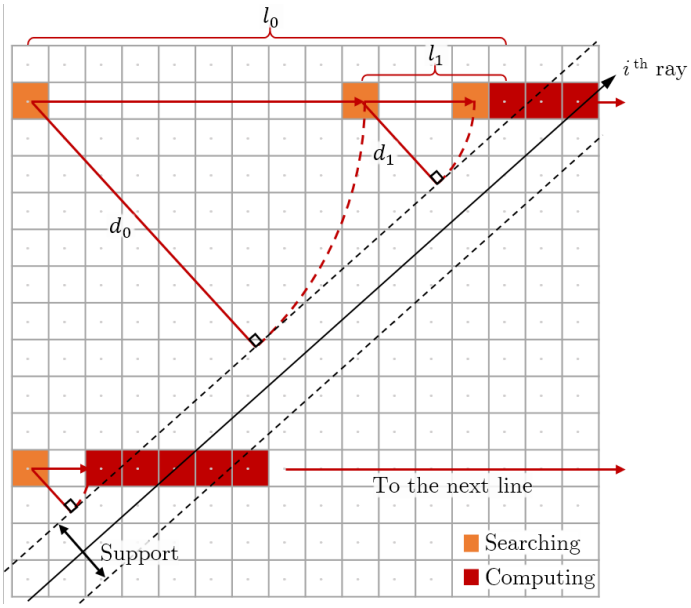


Figure 3.11: Schematic of the compacting operation for the forward projection

circular, we need to find the maximum angle difference (Lipshitz constant) θ_l such that

$$\theta_l = \max\,(\theta_i - \theta_{i-1}), \tag{3.132}$$

then find the next ray index i^* as $i^* = i + \lfloor {}^{\psi_j}/_{\theta_l} \rfloor$. This could fall into an exhaustive search with iterations $O(N_j)$, where N_j being the number of the detector pixels. However, if the detector array is distributed close to a circular shape, i.e. with piece-wise linear array with as many linear pieces as possible, it can be done in constant time. Fig. 3.12 and 3.13 show the graphical representation of the compacting operation for a circular detector and a piecewise linear detector, respectively

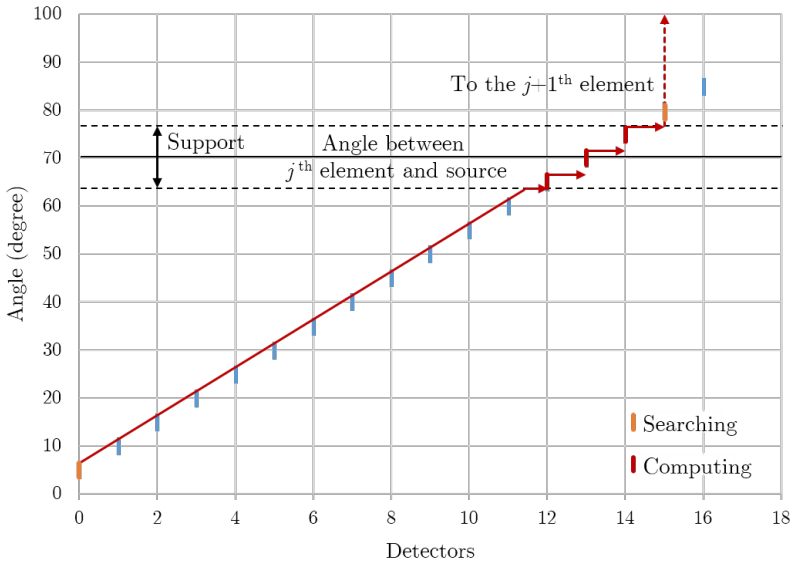


Figure 3.12: Schematic of the compacting operation for the back projection of the circular detector

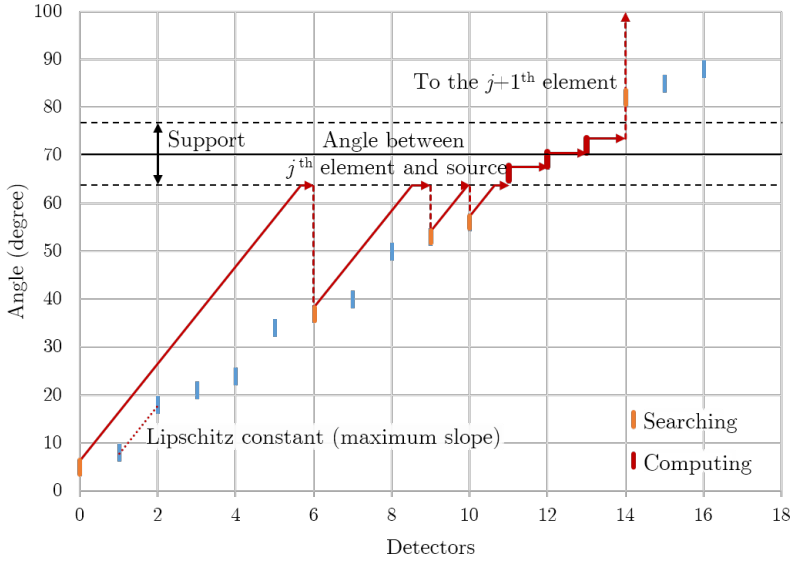


Figure 3.13: Schematic of the compacting operation for the back projection of the piecewise linear detector

BIBLIOGRAPHY

[1] J. F. Barrett and N. Keat, "Artifacts in CT: recognition and avoidance," *Radiographics : a review publication of the Radiological Society of North America, Inc*, vol. 24, no. 6, pp. 1679–1691, 2004.

[2] M. Bieberle and U. Hampel, "Evaluation of a limited angle scanned electron beam x-ray CT approach for two-phase pipe flows," *Meas. Sci. Technol.*, vol. 17, no. 8, pp. 2057–2065, aug 2006.

[3] ——, "Level-set reconstruction algorithm for ultrafast limited-angle X-ray computed tomography of two-phase flows," *Philos. Trans. R. Soc. London A Math. Phys. Eng. Sci.*, vol. 373, no. 2043, 2015.

[4] A. Macovski, R. E. Alvarez, J. L. H. Chan, J. P. Stonestrom, and L. M. Zatz, "Energy dependent reconstruction in X-ray computerized tomography," *Comput. Biol. Med.*, vol. 6, no. 4, pp. 325–336, 1976.

[5] P. M. Joseph and R. D. Spital, "A method for correcting bone induced artifacts in computed tomography scanners," *J. Comput. Assist. Tomogr.*, vol. 2, no. 1, p. 100, 1978.

[6] W. H. Marshall, R. E. Alvarez, and a. Macovski, "Initial results with prereconstruction dual-energy computed tomography (PREDECT)." *Radiology*, vol. 140, no. 2, pp. 421–30, 1981.

[7] J. P. Stonestrom, R. E. Alvarez, and A. Macovski, "A Framework for Spectral Artifact Corrections in X-Ray CT," *IEEE Trans. Biomed. Eng.*, vol. BME-28, no. 2, pp. 128–141, 1981.

[8] N. Clinthorne, "A constrained dual-energy reconstruction method for material-selective transmission tomography," *Nucl. Instruments Methods Phys. Res.*, vol. 353, no. 1-3, pp. 347–348, dec 1994.

[9] S. Gleason and H. Sari-Sarraf, "Reconstruction of multi-energy x-ray computed tomography images of laboratory mice," *IEEE Transactions on Nuclear Science*, 1999.

[10] P. Sukovic and N. Clinthorne, "Basis material decomposition using triple-energy X-ray computed tomography," *Instrum. Meas. Technol. Conf. 1999. IMTC/99. Proc. 16th IEEE*, pp. 1615–1618, 1999.

[11] P. Sukovic and N. H. Clinthorne, "Penalized weighted least-squares image reconstruction for dual energy X-ray transmission tomography." *IEEE Trans. Med. Imaging*, vol. 19, no. 11, pp. 1075–81, nov 2000.

[12] C. H. Yan, R. T. Whalen, G. S. Beaupré, S. Y. Yen, and S. Napel, "Reconstruction algorithm for polychromatic CT imaging: application to beam hardening correction." *IEEE Trans. Med. Imaging*, vol. 19, no. 1, pp. 1–11, jan 2000.

[13] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "An iterative maximum-likelihood polychromatic algorithm for CT." *IEEE Trans. Med. Imaging*, vol. 20, no. 10, pp. 999–1008, oct 2001.

[14] G. C. Giakos, S. Chowdhury, N. Shah, K. Mehta, S. Sumrain, A. Passerini, N. Patnekar, E. A. Evans, L. Fraiwan, O. C. Ugweje, and R. Nemer, "Signal evaluation of a novel dual-energy multimedia imaging sensor," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 5, pp. 949–954, 2002.

[15] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Trans. Med. Imaging*, vol. 21, no. 2, pp. 89–99, 2002.

[16] ——, "Efficient and Accurate Likelihood for Iterative Image Reconstruction in X-ray Computed Tomography," *Proc. SPIE 5032, Med. Imaging 2003 Image Process.*, vol. 5032, pp. 1839–1850, 2003.

[17] ——, "Segmentation-free statistical image reconstruction for polyenergetic x-ray computed tomography with experimental validation," *Phys. Med. Biol.*, vol. 48, no. 15, p. 2453, 2003.

[18] P. E. Kinahan, A. M. Alessio, and J. A. Fessler, "Dual energy CT attenuation correction methods for quantitative assessment of response to cancer therapy with PET/CT imaging," *Technol. Cancer Res. Treat.*, vol. 5, no. 4, pp. 319–327, 2006.

[19] T. G. Flohr, C. H. McCollough, H. Bruder, M. Petersilka, K. Gruber, C. Süβ, M. Grasruck, K. Stierstorfer, B. Krauss, R. Raupach, A. N. Primak, A. Küttner, S. Achenbach, C. Becker, A. Kopp, and B. M. Ohnesorge, "First performance evaluation of a dual-source CT (DSCT) system," *Eur. Radiol.*, vol. 16, no. 2, pp. 256–268, 2006.

[20] L. P. Nogueira, R. C. Barroso, C. J. G. Pinheiro, D. Braz, L. F. De Oliveira, G. Tromba, and N. Sodini, "Mapping lead distribution in bones by dual-energy computed microtomography with synchrotron radiation," *IEEE Nucl. Sci. Symp. Conf. Rec.*, pp. 3471–3474, 2009.

[21] W. Huh and J. A. Fessler, "Model-based image reconstruction for dual-energy x-ray CT with fast kvp switching," in *Proc. - 2009 IEEE Int. Symp. Biomed. Imaging From Nano to Macro, ISBI 2009*, 2009, pp. 326–329.

[22] W. Huh, J. A. Fessler, A. M. Alessio, and P. E. Kinahan, "Fast KVP-switching dual energy CT for pet attenuation correction," *IEEE Nucl. Sci. Symp. Conf. Rec.*, pp. 2510–2515, 2009.

[23] L. Yu, X. Liu, and C. H. McCollough, "Pre-reconstruction three-material decomposition in dual-energy CT," in *Proc. SPIE, Med. Imaging 2009 Phys. Med. Imaging*, vol. 7258, no. 8, 2009, p. 72583V.

[24] C. Maaß, R. Grimmer, and M. Kachelrieß, "Dual Energy CT Material Decomposition from Inconsistent Rays (MDIR)," *IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 1, no. 1, pp. 3446–3452, 2009.

[25] X. Liu, L. Yu, A. N. Primak, and C. H. McCollough, "Quantitative imaging of element composition and mass fraction using dual-energy CT: three-material decomposition." *Med. Phys.*, vol. 36, no. 5, pp. 1602–1609, 2009.

[26] J. Noh, J. A. Fessler, and P. E. Kinahan, "Statistical sinogram restoration in dual-energy CT for PET attenuation correction," *IEEE Trans. Med. Imaging*, vol. 28, no. 11, pp. 1688–1702, 2009.

[27] A. N. Primak, J. C. Ramirez Giraldo, X. Liu, L. Yu, and C. H. McCollough, "Improved dual-energy material discrimination for dual-source CT by means of additional spectral filtration." *Med. Phys.*, vol. 36, no. 4, pp. 1359–1369, 2009.

[28] C. Cai, T. Rodet, S. Legoupil, and A. Mohammad-Djafari, "A full-spectral Bayesian reconstruction approach based on the material decomposition model applied in dual-energy computed tomography." *Med. Phys.*, vol. 40, no. 11, p. 111916, 2013.

[29] C. O. Schirra, E. Roessl, T. Koehler, B. Brendel, A. Thran, D. Pan, M. A. Anastasio, and R. Proksa, "Statistical reconstruction of material decomposed data in spectral CT," *IEEE Trans. Med. Imaging*, vol. 32, no. 7, pp. 1249–1257, 2013.

[30] Y. Long and J. Fessler, "Multi-Material Decomposition Using Statistical Image Reconstruction for Spectral CT." *IEEE Trans. Med. Imaging*, vol. 0062, no. MMD, pp. 1–13, 2014.

[31] R. Carmi, G. Naveh, and A. Altman, "Material separation with dual-layer CT," *IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 4, pp. 1876–1878, 2005.

[32] A. Vlassenbroek, "Dual Layer CT," in *Dual Energy CT Clin. Pract.*, T. Johnson, C. Fink, O. S. Schönberg, and F. M. Reiser, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 21–34.

[33] B. R. Whiting, P. Massoumzadeh, O. A. Earl, J. A. O'Sullivan, D. L. Snyder, and J. F. Williamson, "Properties of preprocessed sinogram data in x-ray computed tomography." *Med. Phys.*, vol. 33, no. 9, pp. 3290–3303, 2006.

[34] G. M. Lasio, B. R. Whiting, and J. F. Williamson, "Statistical reconstruction for x-ray computed tomography using energy-integrating detectors." *Phys. Med. Biol.*, vol. 52, no. 8, pp. 2247–2266, 2007.

[35] S. Srivastava, J. Cammin, G. S. K. Fung, B. M. W. Tsui, and K. Taguchi, "Spectral response compensation for photon-counting clinical x-ray CT using sinogram restoration," *Proc. SPIE 8313, Med. Imaging 2012 Phys. Med. Imaging*, vol. 8313, no. 410, pp. 831 311–831 311–7, feb 2012.

[36] H. Cramér, *Mathematical Methods of Statistics*, ser. Princeton Mathematical Series. Princeton University Press, 1999. [Online]. Available: https://books.google.com/books?id=CRTKKaJO0DYC

[37] W. R. Hendee and E. R. Ritenour, *Medical imaging physics*. John Wiley & Sons, 2003.

[38] G. F. Knoll, *Radiation detection and measurement*. John Wiley & Sons, 2010.

[39] U. Hampel, F. Barthel, M. Bieberle, and T. Strzel, "Transparent target for ultrafast electron beam tomography," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 635, no. 1, pp. 8–12, 2011.

[40] C. D. Daykin, T. Pentikainen, and M. Pesonen, *Practical risk theory for actuaries*. CRC Press, 1993.

[41] H. Cramér, *Collective risk theory: A survey of the theory from the point of view of the theory of stochastic processes*. Nordiska bokhandeln, 1955.

[42] R. Seri and C. Choirat, "Comparison of Approximations for Compound Poisson Processes," *ASTIN Bull.*, vol. 45, no. 3, pp. 601–637, 2015.

[43] A. Entezari and T. Möller, "Extensions of the Zwart-Powell box spline for volumetric data reconstruction on the Cartesian lattice," *IEEE Trans. Vis. Comput. Graph.*, vol. 12, no. 5, pp. 1337–1344, 2006.

[44] A. Entezari, M. Nilchian, and M. Unser, "A box spline calculus for the discretization of computed tomography reconstruction problems," *IEEE Trans. Med. Imaging*, vol. 31, no. 8, pp. 1532–1541, 2012.

[45] R. Redus, "Efficiency and Attenuation in CdTe Detectors (Application Note AN-CdTe-001)," Amptek, Tech. Rep., 2010.

[46] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*. SIAM, 1970, vol. 30.

[47] J. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction." *IEEE Trans. Med. Imaging*, vol. 16, no. 2, pp. 166–175, 1997.

[48] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, p. 2835, 1999.

[49] A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal recovery," in *Convex Optim. signal Process. Commun.* Cambridge, UK: Cambridge Univ. Press, 2009, pp. 42–88.

[50] ——, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, jan 2009.

[51] Y. Nesterov, "A method of solving a convex programming problem with convergence rate O (1/k2)," in *Sov. Math. Dokl.*, vol. 27, no. 2, 1983, pp. 372–376.

[52] W. Zuo and Z. Lin, "A generalized accelerated proximal gradient approach for total-variation-based image restoration," *IEEE Trans. Image Process.*, vol. 20, no. 10, pp. 2748–2759, 2011.

[53] B. O'Donoghue and E. Candès, "Adaptive Restart for Accelerated Gradient Schemes," *Found. Comput. Math.*, no. February, jul 2013.

[54] P. M. Joseph and R. D. Spital, "The exponential edge-gradient effect in X-ray computed tomography," *Physics in Medicine and Biology*, vol. 26, no. 3, pp. 473–487, 1981.

[55] S. J. Schambach, S. Bag, L. Schilling, C. Groden, and M. a. Brockmann, "Application of micro-CT in small animal imaging," *Methods*, vol. 50, no. 1, pp. 2–13, 2010.

[56] W. C. Barber, E. Nygard, J. C. Wessel, N. Malakhov, G. Wawrzyniak, N. E. Hartsough, T. Gandhi, and J. S. Iwanczyk, "Fast photon counting CdTe detectors for diagnostic clinical CT: dynamic range, stability, and temporal response," *Proc. SPIE 7622, Med. Imaging 2010 Phys. Med. Imaging*, vol. 7622, pp. 76 221E–76 221E–8, mar 2010.

[57] J. Stoer and R. Bulirsch, *Introduction to numerical analysis*. Springer, 1993.

[58] A. N. Lowan, N. Davids, and A. Levenson, "Table of the zeros of the Legendre polynomials of order 1-16 and the weight coefficients for Gauss' mechanical quadrature formula," pp. 939–940, 1943.

[59] H. M. Hudson and R. S. Larkin, "Ordered Subsets of Projection Data," *IEEE Trans. Med. Imaging*, vol. 13, no. 4, pp. 601–609, 1994.

[60] NVIDIA Corporation, "NVIDIA's Next Generation CUDATM Compute Architecture: Kepler TM GK110/210," *Nvidia White Pap.*, 2014.

[61] N. Wilt, *The cuda handbook: A comprehensive guide to gpu programming*. Pearson Education, 2013.

CHAPTER 4

Synthetic Phantom Simulation and Experiment

4.1 Introduction

In this chapter, we produce example CT reconstruction results using two synthetic phantoms, and the results are then compared the experimentally determined CT reconstructions with the simulated CT reconstructions to assess the performance of the current SEBXT system. For both the measurement data and the simulation data, we use the volume-fraction based CT reconstruction algorithm introduced in §3. As mentioned in §2.2.5, the charge-integrating X-ray arc detector array was not available at the time of writing, so we used a single pixel spectrometer that can be positioned along the location of the notional detector arc to mimic the data such an arc detector would yield. Details of the measurement procedure are written in §4.2. A sinogram simulation was created to evaluate the best-expected performance of SEBXT. Details of the simulation procedure are written in §4.3 Phantom 1 and Phantom 2 are used as a reference to compare the CT reconstruction results based on the measurement data and the simulated data. The comparison results are shown in §4.4.

4.2 Collection of Measured Projection Set (Sinogram)

4.2.1 Measurement procedure

Projection sets were collected for Phantom 1 and Phantom 2 by traversing the single detector to a given location, and then moving the X-ray source position. We collected the data using the following procedure:

1. Position the point detector at 1 of 64 locations along the notional detector arc (Fig. 4.1).
2. Collect attenuation data for 128 X-ray source locations:

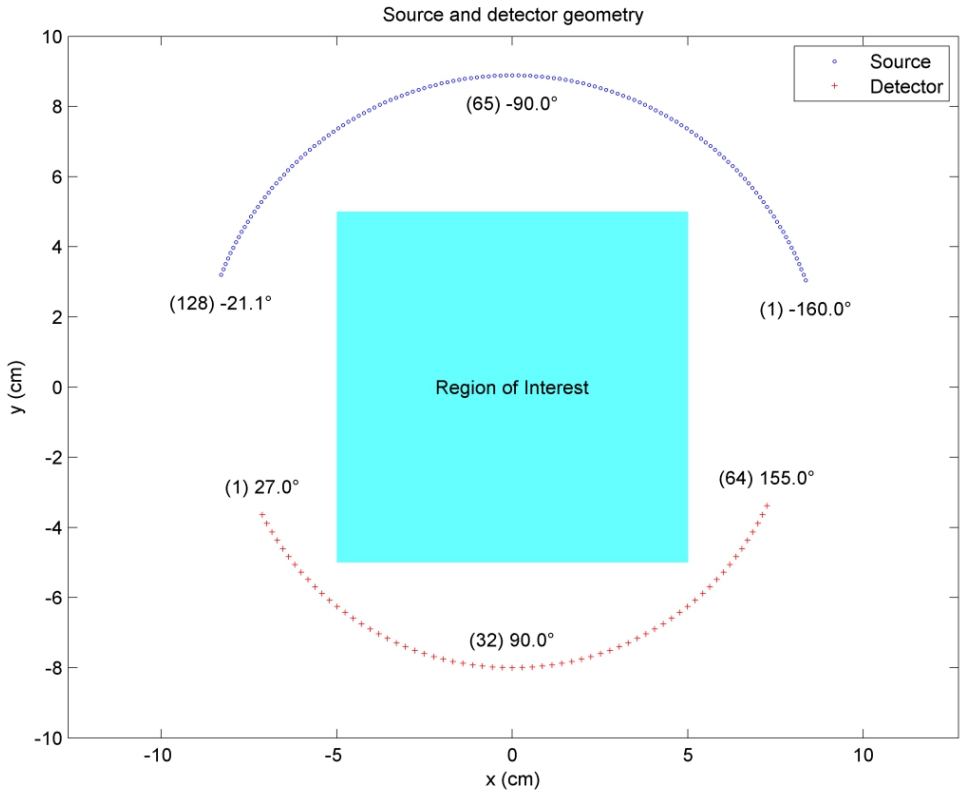


Figure 4.1: Geometry of the source points, detector points, and the region of interest

(a) Turn on the electron beam
(b) Move the beam to the discrete spot on the target
(c) Wait 0.05 sec (communication time between the server and the controller)
(d) Acquire data from the detector for 0.5 sec
(e) Wait 0.05 sec (communication time between the server and the controller)
(f) Move the beam to the next location on the tungsten target

3. Move the detector to the next location along the detection arc and repeat from step 2.

The voltage bias between the cathode and the anode was set to 115 kV, and the beam current was set to 1.0 mA. The beam current generally has a fluctuation of 0.1 mA. (Note: 0.1 mA fluctuation of the beam current is not insignificant considering that it is equivalent to 10% of the setting.) For the present study, the beam current could not be increased, because the spectrometer associated with the detector can be easily saturated. However, a current up to 133 mA may be supplied by the SEBXT, and hence with the full detector arc capable of handling a higher photon flux, the 0.1 mA fluctuation of the beam current is not expected to be a significant issue. The settings used to position the beam on the arc

tungsten target (the physical (x, y) coordinates) was related to the beam deflection voltage (U, V) after performing the alignment calibration described in §2.2.3.

4.2.2 Correction of measured data due to detector limitations

Fig. 4.2 and Fig. 4.3 show the raw data sinograms obtained from the measurement of Phantom 1 and Phantom 2, respectively. The raw measurement data obtained by the Amptek spectrometer cannot be used immediately for the reconstruction because 1) the measured data do not directly indicate energy values, but a relative energy intensity due to lack of accurate calibration, 2) the detector loses significant fraction of the photons when counting at a high count rate, as was seen in Fig. 2.14. Hence, while the spectrometer records the X-ray photon counts at each bin, the energy level of each bin must be calibrated by the user. Because of the distortion due to the high pulse pile-up events, energy calibration of the spectrometer could not be implemented. As a result, a relative energy intensity is used to record the result, where the relative energy intensity is defined as:

$$\bar{h}_i = \sum_{k=1}^{N_k} \bar{y}_i(\mathcal{E}_k)k. \tag{4.1}$$

Here k is the energy bin number (it is implicitly assumed photons counted, e.g., in bin 50 has twice the energy as those in bin 25), $\bar{y}_i(\mathcal{E}_k)$ is the number of photons counted in bin k for i^{th} ray and N_k is the total number of energy bins. The actual energy-integral is defined as:

$$h_i = \sum_{k=1}^{N_k} \bar{y}_i(\mathcal{E}_k)\mathcal{E}_k. \tag{4.2}$$

If the hardware response of the spectrometer is linear, then the actual energy integral can be found by affine-transforming the relative energy intensity. However, the detector system lost some of its input photon count due to pulse pile-ups and the amount of the energy lost due to the pile-up events increases with increased photon flux experienced by the detector. Consequently, the response of the detector becomes nonlinear, and a correction must be made to compensate for photons lost due to pile-up events. In order to convert the raw measurement data to the energy-integrated data, a correction of the nonlinear effect due to the saturation and pulse pile-ups is accomplished by fitting the radical-polynomial functions to the data. Guided by experiments, the following function is chosen as the objective fitting curve:

$$h_i = C_1\,\bar{h}_i^{2.5} + C_2\,\bar{h}_i^2 + C_3\,\bar{h}_i + C_4. \tag{4.3}$$

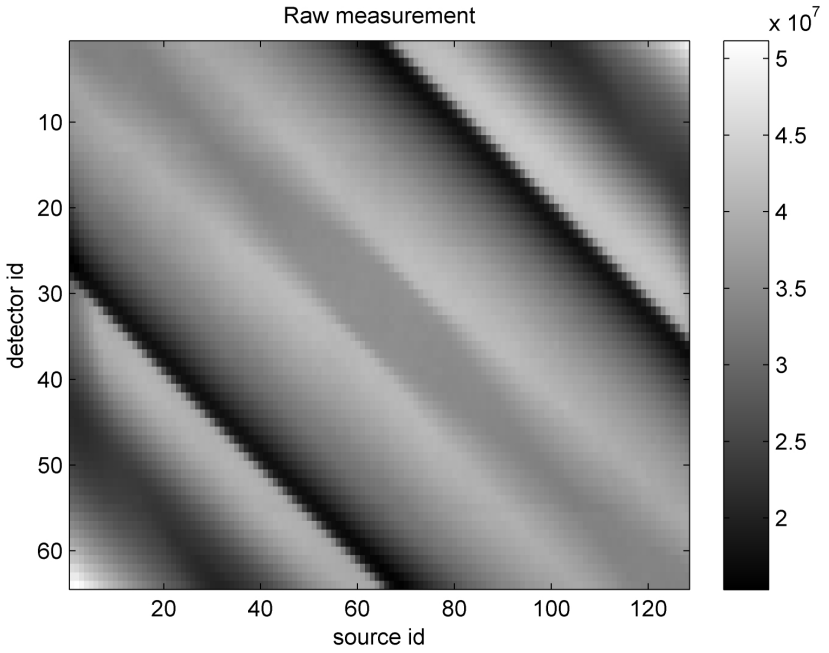


Figure 4.2: Raw measured sinogram data, Phantom 1. Source id and detector id indicate the location of the source and detector given in Fig. 4.1. The unit of the image intensity is arbitrary.

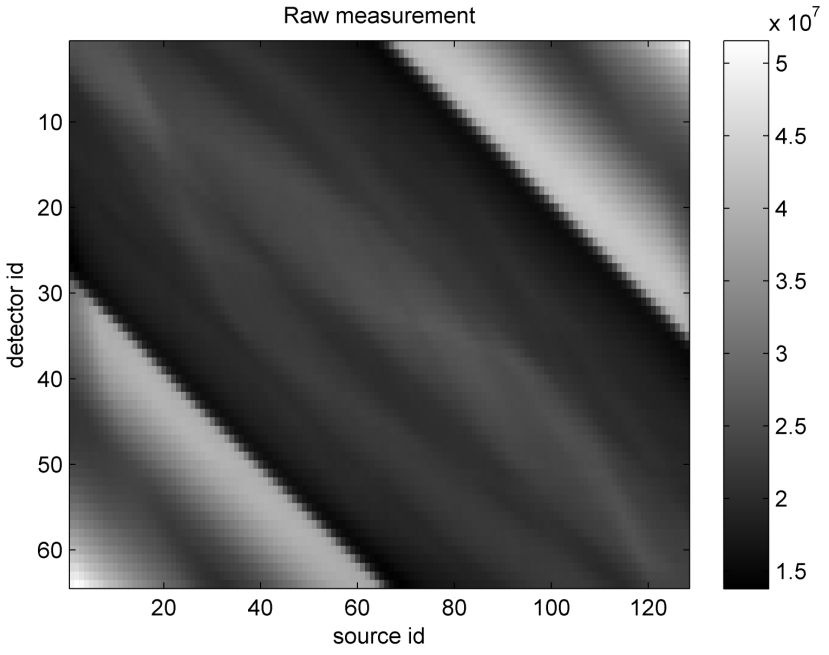


Figure 4.3: Raw measured sinogram data, Phantom 2. Source id and detector id indicate the location of the source and detector given in Fig. 4.1. The unit of the image intensity is arbitrary.

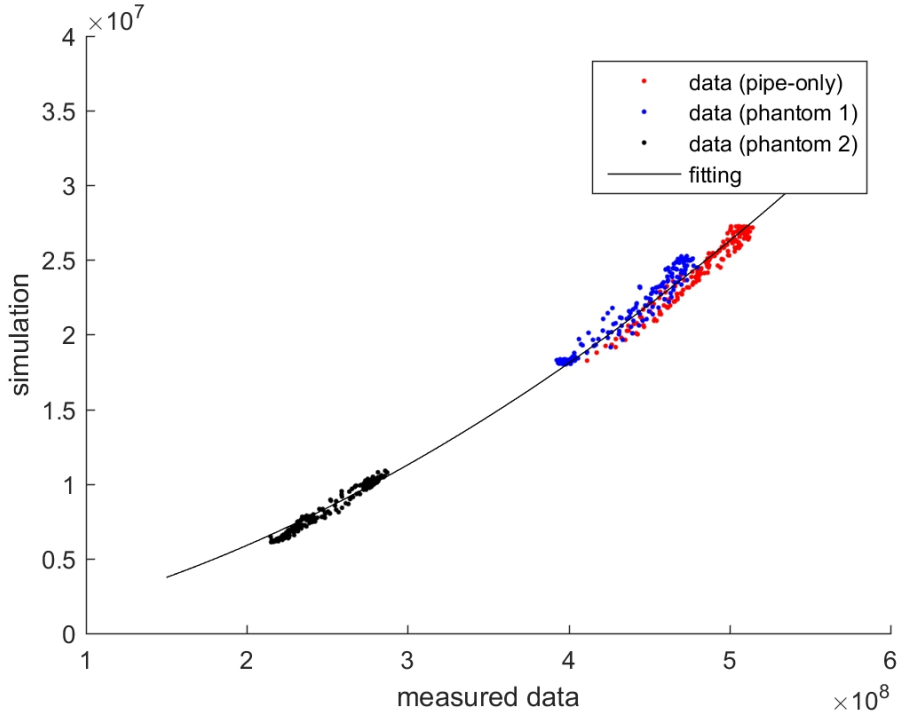


Figure 4.4: Energy data correction curve. The fitted curve is used to correct for the nonlinear response of the detector due to the pulse pile-up. Sampled data are taken from the center of the sinogram, detector location from 32 to 33 and source location from 34 to 94. Here horizontal axis has the unit of relative energy $\bar{h}_i$ described in (4.1) for the data measured by the Amptek XR-100T-CdTe, and vertical axis has the unit of the true integrated energy h_i described in (4.2) for the simulated data.

Samples for the estimation of coefficients in (4.3) are taken from the center location of the sinogram (detector id from 31 to 33, and source id from 34 to 94) of the empty pipe, Phantom 1 and the Phantom 2 sinograms for both measured data and the simulations, and the coefficients are computed based on a least-square fit, yielding

$$
\begin{aligned}
C_1 &= 8.2545\text{e}-14,\\
C_2 &= -8.1578\text{e}-14,\\
C_3 &= 1.8386\text{e}-6,\\
C_4 &= 7.6151\text{e}-19.
\end{aligned}
$$

The data points and the fitted curve for the energy correction are shown in Fig. 4.4.

Additionally, another correction is made using the difference of the simulated and measured sinograms for the empty pipe. The simulation does not entirely reflect the actual

environmental attenuation. For example, aluminum wall of the gun is assumed to be an exact half circle, while the actual aluminum wall is not perfectly circular and the wall thickness varies from center to side. Also, a steel filter was placed between the phantom and the detector. This steel filter is assumed to have an infinitely long width in the simulation, but it is only 2-inch-wide in the real system. Hence, the simulated environment deviates from the actual environment as the beam or the detector moves to the side. A correction map is obtained by dividing the values of the empty pipe simulation sinogram by those of the empty pipe measurement sinogram after corrected by the fitting curve described in (4.3). Fig. 4.5 shows the correction map obtained by comparing the empty pipe sinograms. This correction map is applied to each of Phantom 1 and Phantom 2 sinograms by element-wise multiplication.

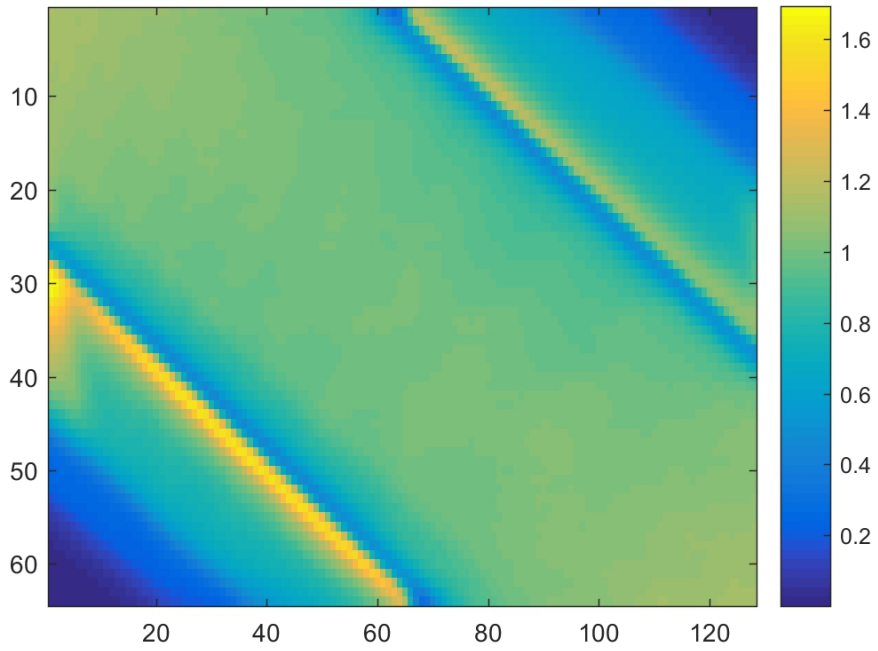


Figure 4.5: A correction map obtained from the empty pipe sinogram. This correction map is used to correct the unidentified environmental attenuation error between the simulation and the measurement. The correction factor of each pixel is obtained by dividing the simulated empty pipe sinogram by the measured empty pipe sinogram.

The comparisons between the corrected measurement sinograms and the simulated sinograms for Phantom 1 and Phantom 2 are shown in Fig. 4.6 and Fig. 4.7, respectively, for select detector locations. As can be seen in the comparison, corrected measurement data conform to the simulation data in the global aspect, but is blurred and lose detail compared to the simulation data. The noise level is much higher in the corrected measurement sinogram due to the unstable beam current at the low range. Unfortunately, the simulation does not include this type of control-related randomness of the beam current because the data acquisition with the SEBXT has been designed to employ much higher current in order to produce the lowest acquisition time.

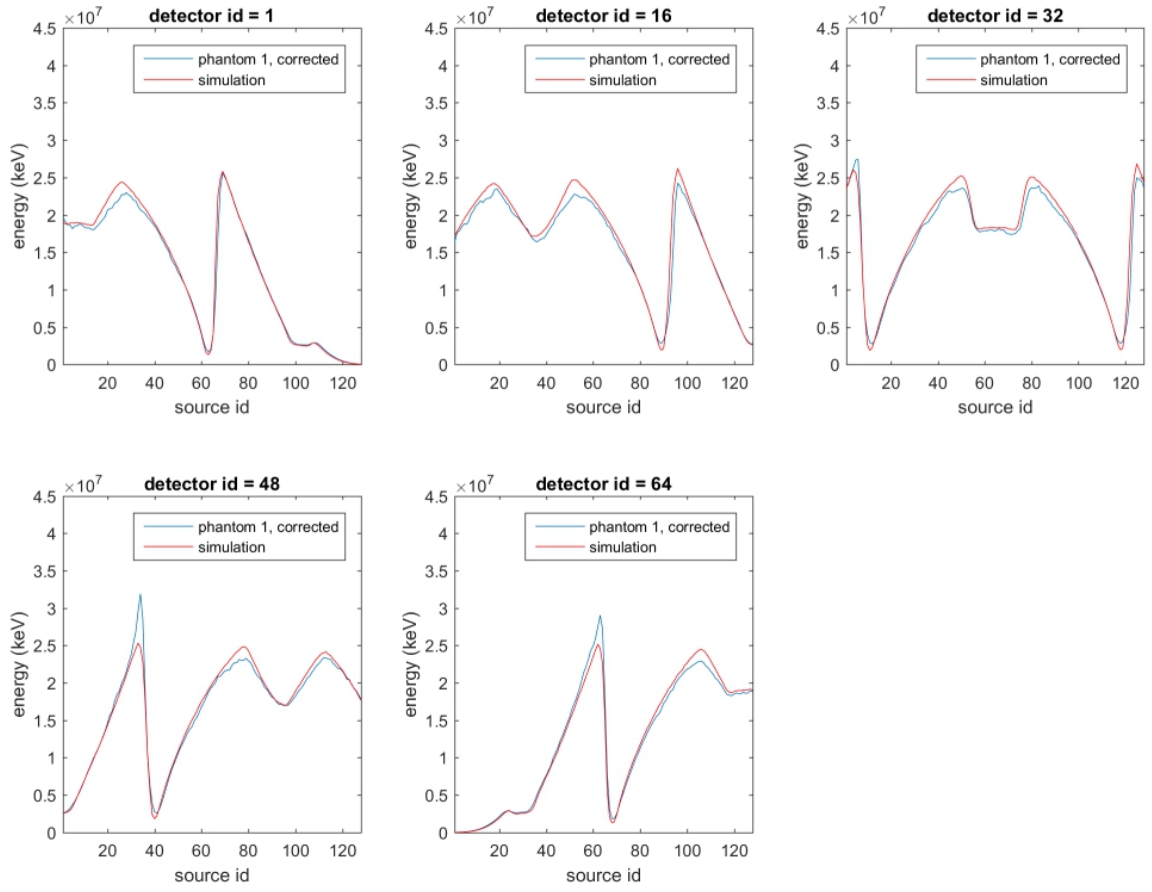


Figure 4.6: Comparisons between the corrected measurement and the simulated data from Phantom 1 at the detector locations 1, 16, 32, 48, and 64. Corresponding detector positions are shown in Fig. 4.1.

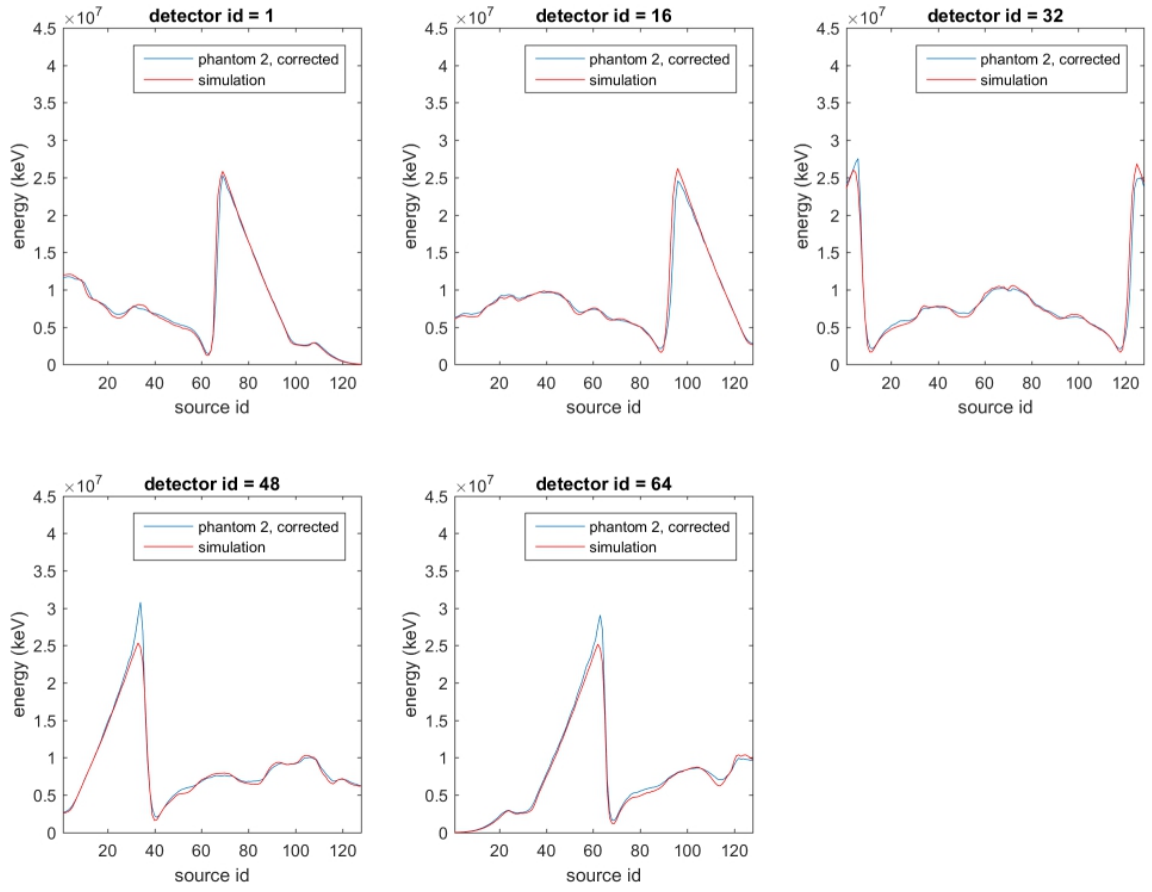


Figure 4.7: Comparisons between the corrected measurement and the simulated data from Phantom 2 at the detector locations 1, 16, 32, 48, and 64. Corresponding detector positions are shown in Fig. 4.1.

4.3 Simulation of projection set (sinogram)

We created a simulation code to get the best-expected CT reconstruction result based on the physics of the X-ray source, the propagation physics of the X-ray photons through the domain, and the known performance of the detectors. By doing this, we can validate the simulation's performance with measurement data. The requirement is that the simulation must adequately represent reality.

In order to simulate the performance of an X-ray Computed Tomography system, we would like to predict the sinogram that will be created for a given object with the system measurement domain. The sinogram will be strongly related to both the attenuation of materials (or equivalently, volume fraction of materials) within the domain as well as the system parameters and characteristics (e.g. the X-ray source spot size, the number and size of the detectors, the detector arc angle). The simulation program therefore will model several important physical features of the CT system, including models of the following system elements:

1. The X-ray source spot size, photon energy distribution, and the uncertainties in the number of X-ray photons and their energies based on the input control parameters to the EB gun.
2. The physical range of the X-ray projections for the given control inputs.
3. The X-ray attenuation through the measurement domain for each expected material within the domain for the given X-ray photon energy spectrum.
4. The response of the detector for each level of X-ray photon energy given its finite surface area and thickness.

Once these elements are modeled, a simulation of the noiseless sinogram can be created given Beer's law. Then, we added the gamma-approximated system-specific random noise to the sinogram as described in §3.2.4. The final result of the simulation is a noise-corrupted sinogram. This noise-corrupted sinogram will be used for the CT reconstruction

4.3.1 Flow diagram of the sinogram simulation

The steps performed in the simulation are described in Fig. 4.8. First, the detector and source positions are computed based on the input parameters. Then, the attenuation coefficients are obtained at each discrete energy bin by interpolating the National Institute of Standards and Technology (NIST) dataset of the X-ray attenuation coefficients [1]. For each source-detector pair, the attenuation coefficients are integrated along the line formed by the X-ray source point and 20 evenly spaced sampling points on the surface of

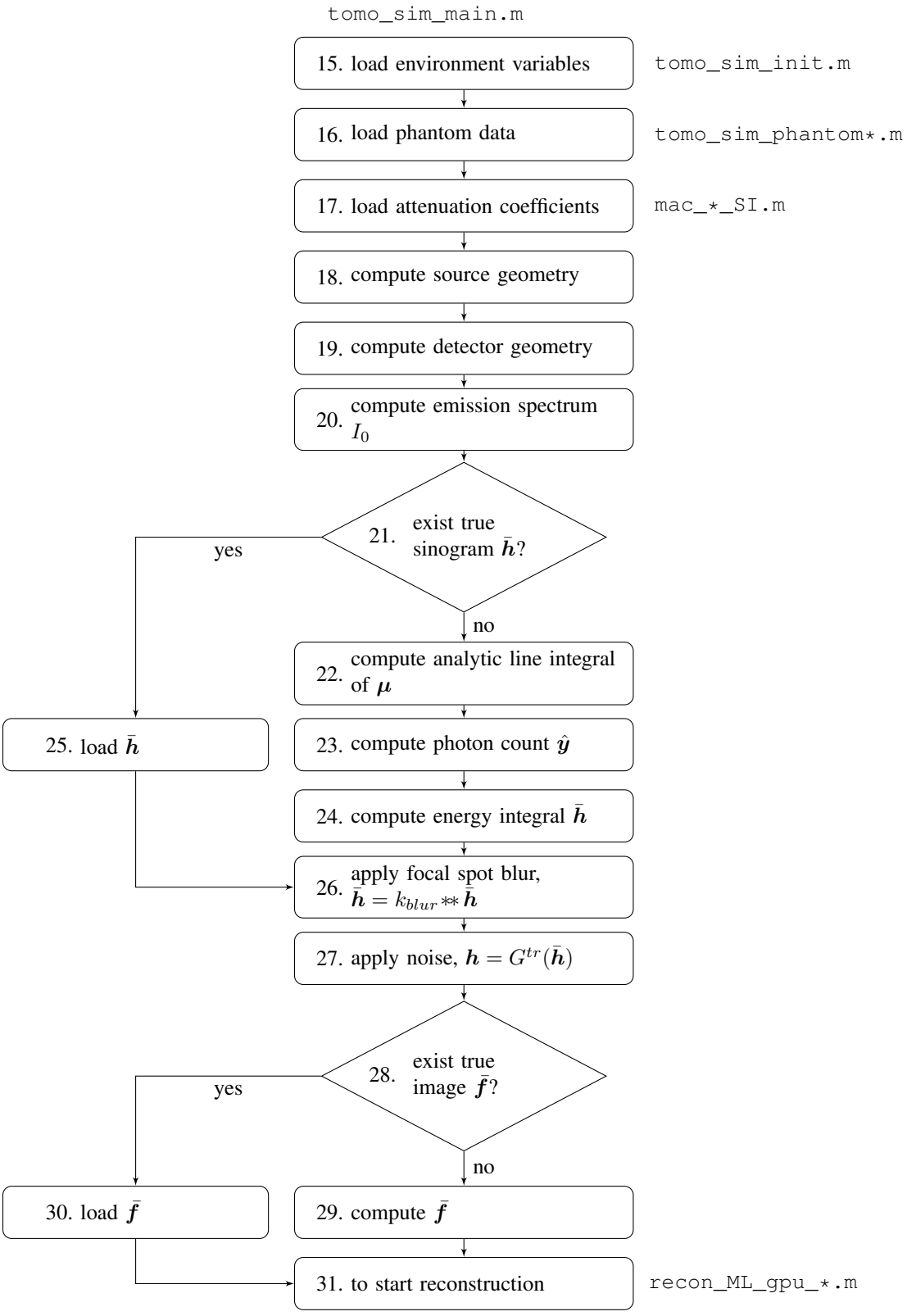


Figure 4.8: Flow diagram of the sinogram simulation

the detector, which is then averaged over the 20 sampled points. Analytic solutions for the X-ray projection of rectangular and elliptical objects are used to compute the line integrals. Scattering is assumed to be small and is therefore neglected in the simulation. Once the analytic attenuation is computed for each source-detector pair, a noise is applied to the data using the shifted gamma distribution. A brief explanation of each node is presented below.

1, 2, 4, 5. load environment/phantom variables, compute source/detector geometry

Most of the geometries for the source/detector and the phantoms are parameterized, and those parameters are stored in "`tomo_sim_init.m`" (source and detector) and "`tomo_sim_phantom XXX.m`", where XXX denotes the identification number of the phantom.

3. load attenuation coefficients

For every energy bins, mass attenuation coefficients are obtained by interpolating the data tables given by NIST [1]. Since the data points for the interpolation are spread evenly in the logarithmic scale, linear interpolation fails to achieve a good estimation. To get a better estimation, Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) is used. (See http://www.mathworks.com/help/matlab/ref/pchip.html for more details.)

6. compute emission spectrum

The emission X-ray photon spectrum obtained by the SpekCalc software [2] account for the case when the attack angle is 0 degrees, detector face size is 1 cm^2 and the distance between the source and the detector is 1 m. Thus, the emission spectrum must be corrected for each beams angle of attack and the distance between the source and the detector. The following formula has been used to correct the emission X-ray photon spectrum.

$$I_0^{(i)} = I_s(\mathcal{E}) A_d \left(\frac{100}{d_t^{(i)}} \right)^2 \cos\left(a_t^{(i)} \right) \tag{4.4}$$

where $I_0^{(i)}$ is the corrected emission spectrum of the i^{th} ray, I_s is the emission spectrum obtained by the SpekCalc software, $\mathcal{E}$ is the energy level, A_d is the detector face area in cm^2, $d_t^{(i)}$ is the distance between the source and the detector of the i^{th} ray in cm, $a_t^{(i)}$ is the angle of attack of the i^{th} ray as defined in Fig. 4.9.

7. exist sinogram?

The code saves sinograms with the name convention "`y_data kW=XXX kV=XXX`

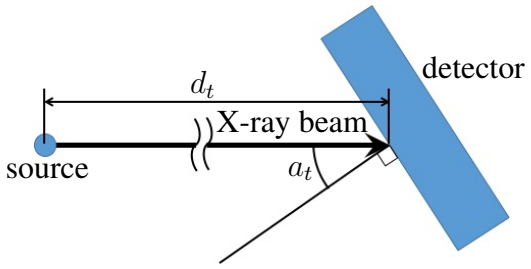


Figure 4.9: Schematic diagram of X-ray beam attack angle and distance

`kHz=XXX Ns=XXX Nd=XXX type=XXX phantom=XXX dyn=XXX.mat`", where XXX are values from left to right, beam power, beam voltage, scan frequency, number of source location, number of detectors, source type, phantom ID, and flow speed. If the sinogram exists, go to 12. Otherwise, generate a new sinogram for the given condition.

8. compute analytic line integral of μ

   a) Generate the $i_{\text{up}}{}^{\text{th}}$ source - detector pair after up-sampling the detector position by a factor of 20, $i \to i_{\text{up}}$

   b) Compute $t_{i_o}^{i_{\text{up}}}$, the thickness of $i_o{}^{\text{th}}$ object along the $i_{\text{up}}{}^{\text{th}}$ ray. The current code can only compute thickness of two shapes, 3-D ellipses and 2-D rectangles, and the corresponding Matlab codes are "`sol_intersect3.m`", and "`rect_intersect.m`".

   c) Compute the line integral of μ along the $i_{\text{up}}{}^{\text{th}}$ ray

$$\mu_{\text{tot}}^{(i_{\text{up}})}(\mathcal{E}_k) = \sum_{i_o} \left\{ \mu_{i_o}(\mathcal{E}_k) \int_{\mathcal{L}^{i_{\text{up}}}} t_{i_o}^{i_{\text{up}}} \, \mathrm{d}l \right\} \tag{4.5}$$

9. compute mean photon count $\hat{y}$

   a) Down-sample $\mu_{\text{tot}}^{(i_{\text{up}})}$ by a factor of 1/20, $i_{\text{up}} \to i$

   b) Compute $\hat{y}$ for each discrete energy level $\mathcal{E}_k$

$$\hat{y}^{(i)}(\mathcal{E}_k) = I_o^{(i)}(\mathcal{E}_k) \exp\left(-\mu_{\text{tot}^{(\text{i})}}\right) \Delta\mathcal{E}. \tag{4.6}$$

10. compute mean energy integral $\hat{h}$

Total energy acquired by a detector can be obtained from the first moment of $\hat{(y)}$ with

respect to $\mathcal{E}_k$,

$$\hat{h}^{(i)} = \sum_k \mathcal{E}_k \; \hat{y}^{(i)}(\mathcal{E}_k). \tag{4.7}$$

12. apply blur

To apply the effect of the blurring due to the beam focal spot size, the following steps occur:

a) Assuming the object is small, the blur kernel is approximated by the blurring effect of the beam passing through the center point of the object (In reality, blurring effect due to the beam profile significantly changes depending on the location. The method applied here is a rough approximation in the average sense.)

b) Assume the beam profile is Gaussian curve-shaped with known FWHM, and the transmitted X-ray beam intensity is constant throughout the surface of each detector, which can be represented by a rectangular function. For example, a 3 mm FWHM beam profile and the beam intensity on the surface of 3 mm width detector is shown in Fig. 4.10a and 4.10b, respectively.

c) A continuous blur kernel is computed by convolving the transmitted X-ray beam intensity with the beam profile. An example of continuous blur kernel with the condition described in the item b) is shown in Fig. 4.10c. Each colored square represents the range of each detector.

d) To get the discrete blur kernel, the value of the continuous blur kernel is numerically integrated over the range of each detector. For the support size of the blur kernel, total of 9 detectors (4 on the left, center, and 4 on the right) are considered.

e) The discrete kernel values are normalized so that the sum of the discrete kernel values becomes unity. An example of discrete blur kernel with the condition described in the item b) is shown in (d) of Fig. 4.10. Each colored circle represents the discrete value corresponding to the detector with the same color.

f) Finally, the columns (detector axis) of the original sinogram are convolved with the discrete blur kernel obtained at the item e). Fig. 4.11 shows the comparison between the measurement data and simulated data with different focal spot Full-Width at Half-Maximum (FWHM), to show the effect of the beam focal spot to the projected image. Note that the object size in the simulation differs depending on the actual location of the object.

For the simulations presented in this work, FWHM of the beam focal spot is chosen to be 4 mm, which is much worse case than the experimentally obtained FWHM of 1.18 mm. However, by comparing sinograms, we found that FWHM 4 mm is more adequate as shown in Fig. 4.11. Possible reason for this deviation is explained in §5.

13. apply noise

Considering the physics of the X-ray photons interacting with the detector crystal, the noise follows the aggregated compound Poisson distribution, in which photon energy follows the spectral response distribution of the detector, while the total number of photons follows the Poisson discrete distribution. However, since this approach takes too much time to simulate, an approximated three-parameter distribution (i.e., shifted gamma distribution) is used to apply the noise. An approximation method to convert the aggregated compound Poisson distribution to shifted Gamma distribution can be found in [4]. Detailed process to acquire parameters for the shifted Gamma distribution is explained in §3.2.4.

14, 15. exist true image? If not, compute true image

Same as the sinogram, true image is obtained after up-sampling the image grid by a factor of 20, draw the binary fraction image, and then down-sample by a factor of 1/20. The true image is stored in the name formatted "`actual NxXXX NyXXX phantomXXX.mat`", where XXX are number of pixels in x, number of pixels in y, and phantom ID, from left to right. If the true image already exists, load the image and start reconstruction. Otherwise, build the true image.

4.4 CT Reconstruction of Measured and Simulated Sinograms Compared to the Physical Phantom

The reconstruction images are achieved by applying the Matlab code, which is the implementation of the algorithm described in Chapter 3. Using the energy-integrated measurement sinogram data obtained by synthesizing the slow energy-resolved detector data, two volume fraction images (titanium and water) are obtained as an output based on the material decomposition scheme. The quality of the reconstructed image is represented by the root mean square errors (RMSE) of each volume fraction image given by:

$$\mathrm{RMSE}_{(l)} = \sqrt{1/N_j \sum_j \left(\hat{f}_{(j,l)} - \bar{f}_{(j,l)}\right)^2}. \tag{4.8}$$

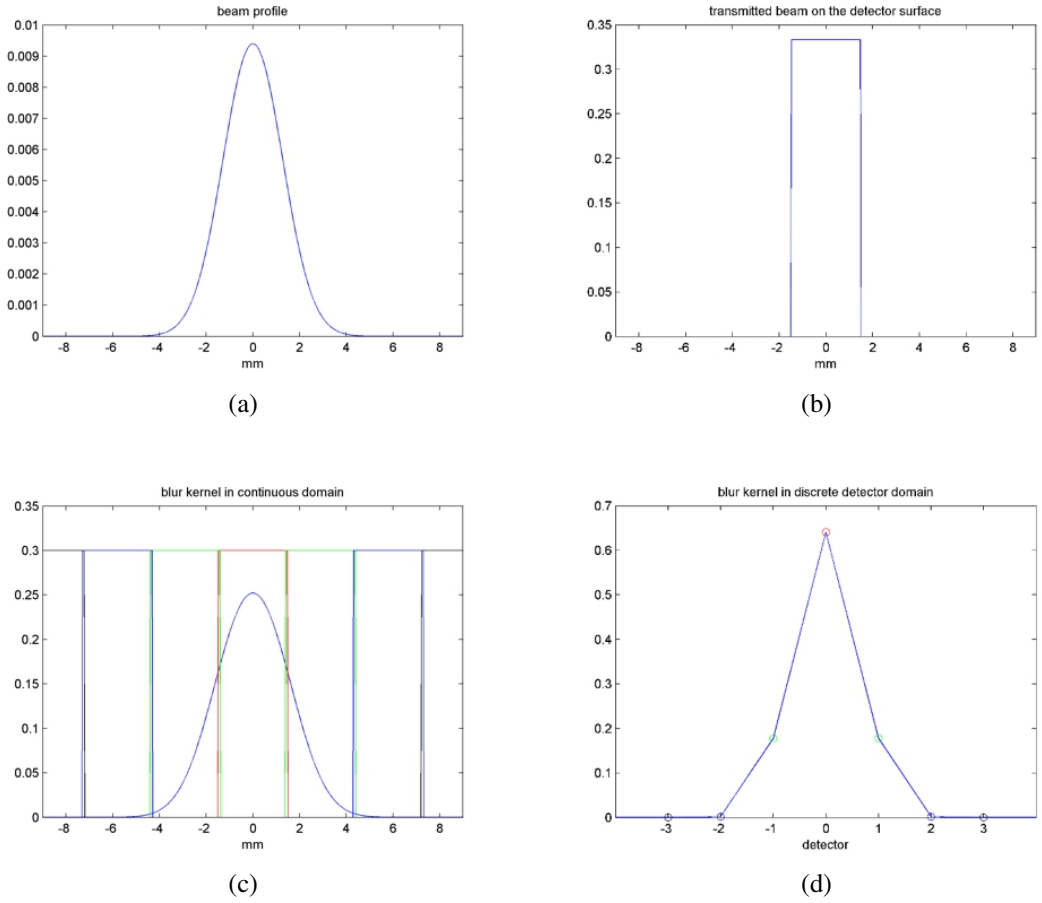


Figure 4.10: Generation of the discrete blur kernel for 3 mm FWHM beam profile and 3 mm detector as detailed in "step 12" of the algorithm. All computations are based on 2D, small, and centered objects; (a) the electron beam profile on the tungsten target, assumed Gaussian shaped; (b) transmitted beam intensity on the detector surface when zero-area focal spot, assumed constant; (c) continuous blur kernel, obtained by the convolution between (a), (b), and the sensitive area of adjacent detectors (colored squares); and (d) the discrete blur kernel, obtained by integrating the continuous blur kernel over each sensitive area of adjacent detector, and then normalized.

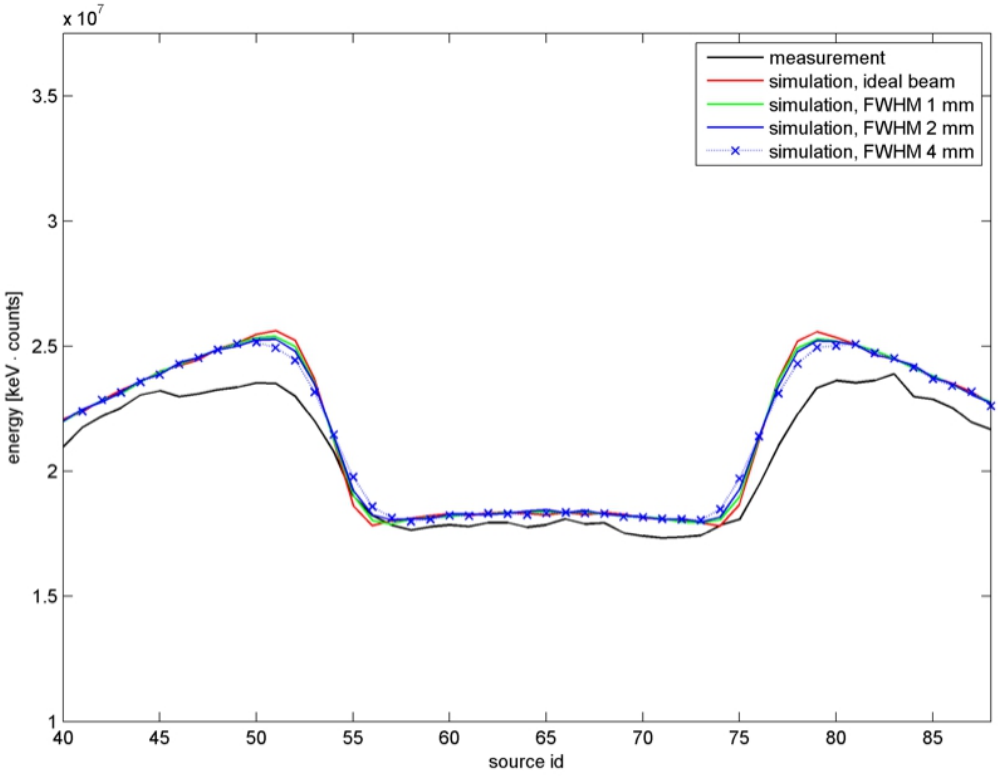


Figure 4.11: Comparisons of the blurring effect due to the beam focal spot FWHM taken for Phantom 1 at the center detector location. In the simulated sinograms, the blurring effect induced by the finite focal spot is applied using the discrete blur kernel explained in §4.3, Item 12).

4.4.1 Phantom 1

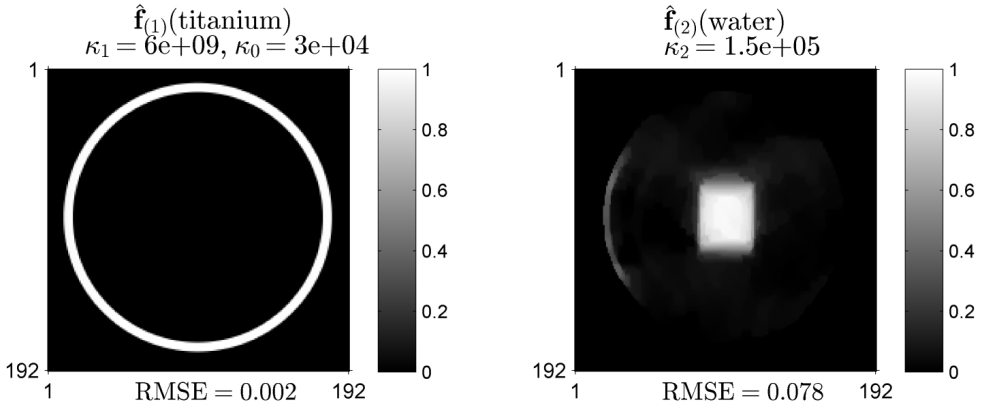
Fig. 4.12 shows the reconstructed images based on the corrected measurement and simulated sinogram data, and true volume fraction images of Phantom 1. Fig. 4.13 shows the volume fraction deviation of the reconstructed Phantom 1 images compared with the true images, and the volume fraction deviation between the reconstruction images based on the corrected measurement and the simulation sinogram data. Each RMSE of the reconstructed image is shown below the corresponding figure.

To quantify the resolution of the CT system, a Modulation Transfer Function (MTF) is obtained from the sampled edge of Phantom 1 reconstruction in the same manner described in Fischer et al. (2008) [3]. Fig. 4.14 presents the steps to compute the modulation transfer function obtained from Phantom 1, (a) edge sampling location, (b) edge spread function, (c) line spread function, and (d) modulation transfer function at the sampled edge. Based on the MTF results, the current system has a maximum spatial resolution of 0.52 lp/mm (line pair per millimeter) at 10% MTF, which is comparable to the result of 0.51 lp/mm in [3]. For reference, the heart-shaped bubble clusters at the bottom of the Phantom 2 have frequency components of 0.42 lp/mm and higher. Considering the present UM SEXBT is a limited angle system with effectively 64 3 mm$\times$3 mm detector pixels, the resolution result is promising. However, the modulation transfer function of Phantom 1 does not form a smooth curve due to the structured noise. This structured noise is suspected to be induced by the unstable current and the focal spot uncertainty of the beam, combined with the effect of the edge-preserving regularizer.

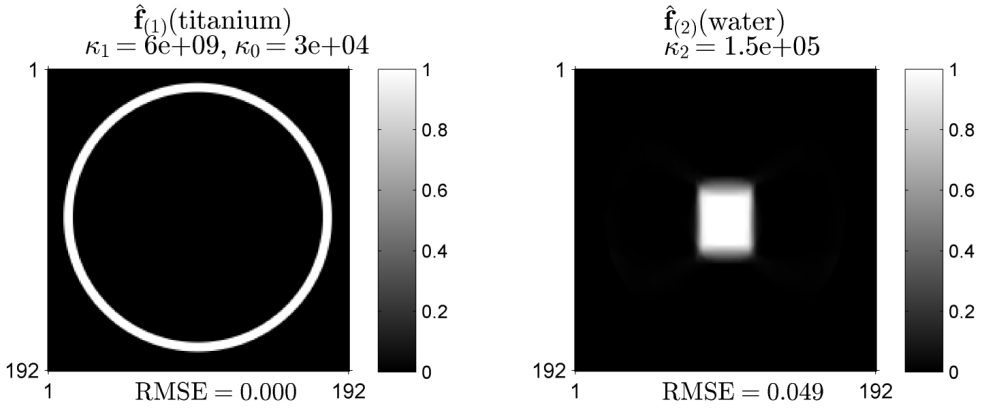
4.4.2 Phantom 2

Fig. 4.15 shows the reconstructed images based on the corrected measurement and simulated sinogram data, and true volume fraction images of Phantom 2. Fig. 4.16 shows the volume fraction deviation of the reconstructed Phantom 2 images compared with the true images, and the volume fraction deviation between the reconstruction images based on the corrected measurement and the simulation sinogram data. Same as Phantom 1, each RMSE defined in (4.8) of the reconstructed image is shown below the corresponding figure.

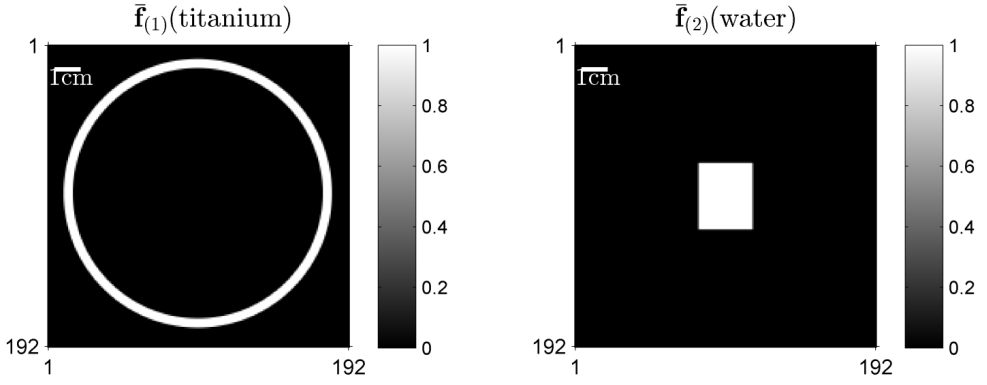
Average void fractions of the sampled partial images are compared with those of true image with the same resolution. Samples taken from the true image and the reconstructed images based on the corrected measurement and simulation are shown in Fig. 4.17 for comparison. The averaged void fraction results for all sampled areas are summarized in Table 4.1.



(a) Real data, corrected

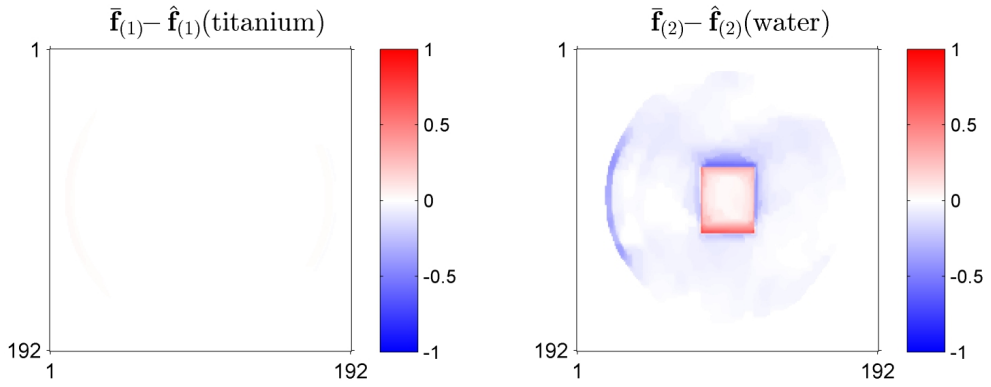


(b) Simulated data

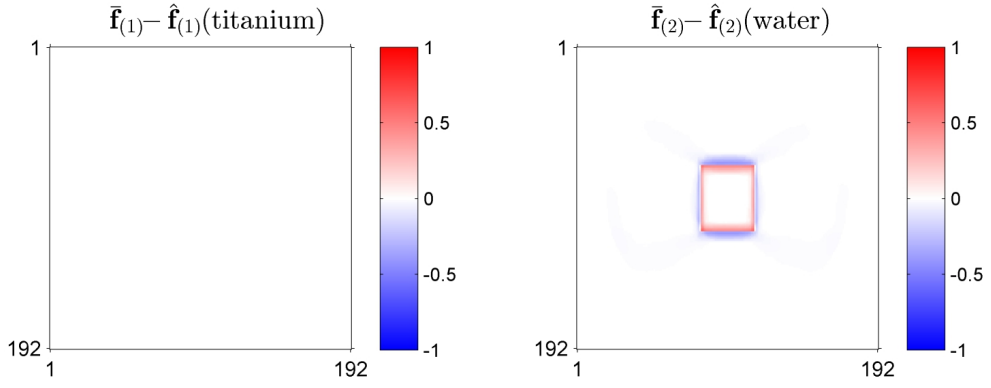


(c) True image

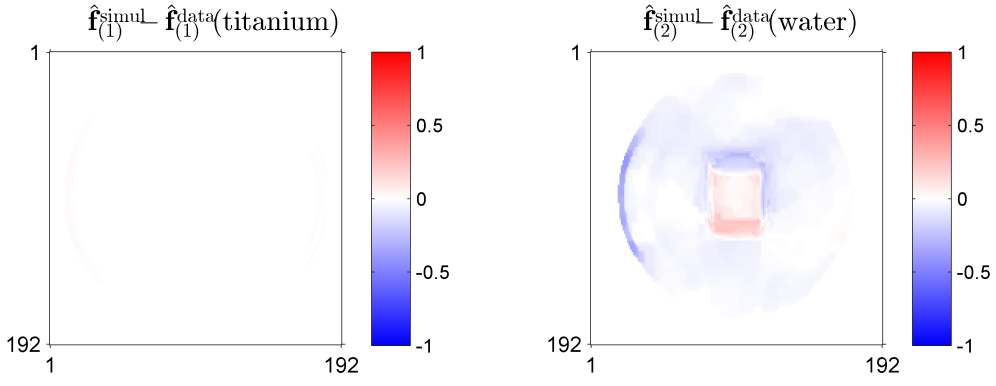
Figure 4.12: CT reconstructed volume fraction image based on (a) the corrected measurement sinogram and (b) the simulated sinogram for Phantom 1. True volume fraction image of Phantom 1 is shown in (c).



(a) Real data, corrected



(b) Simulated data



(c) Simulated data - real data

Figure 4.13: Volume fraction deviation of the reconstructed images; (a) error image of the CT reconstructed image based on the corrected measurement data; (b) error image of the CT reconstructed image based on the simulated data; (c) deviation between the corrected measurement CT result and the simulated CT result for Phantom 1.

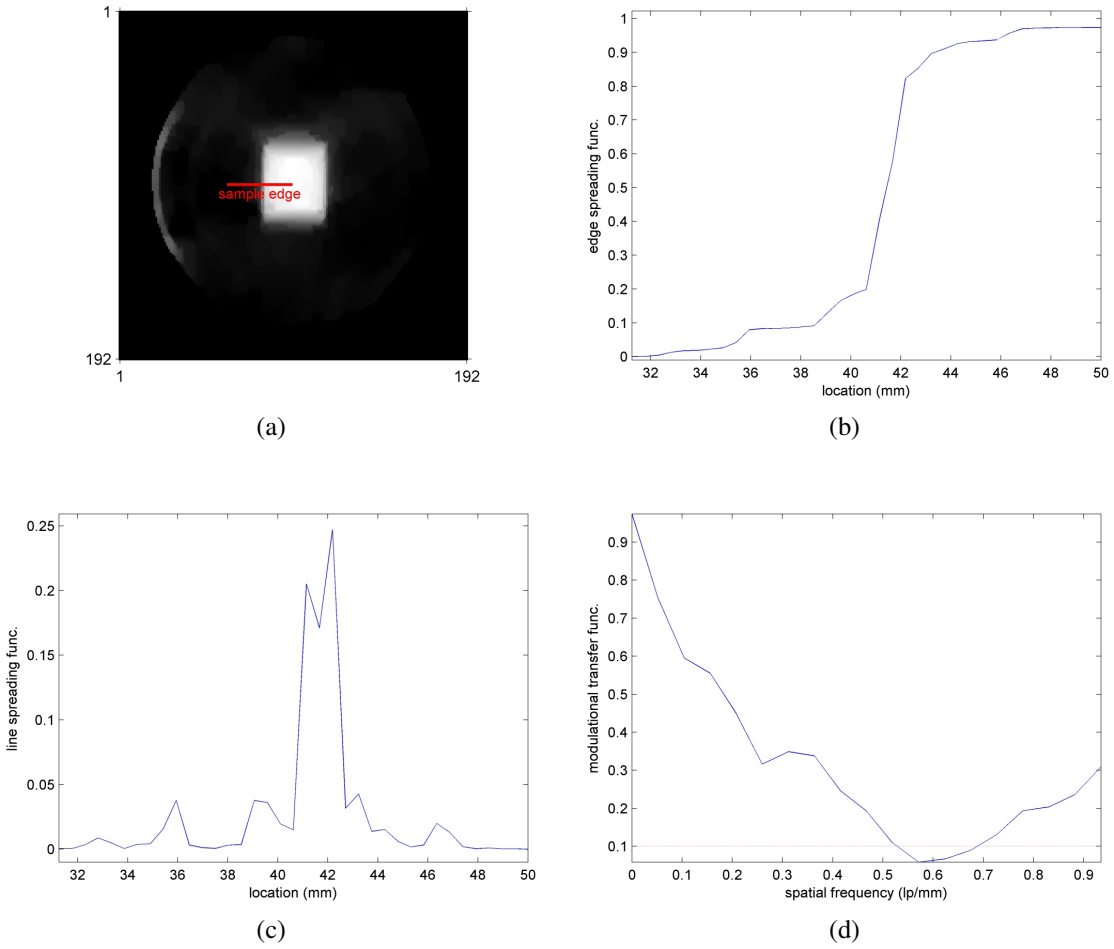
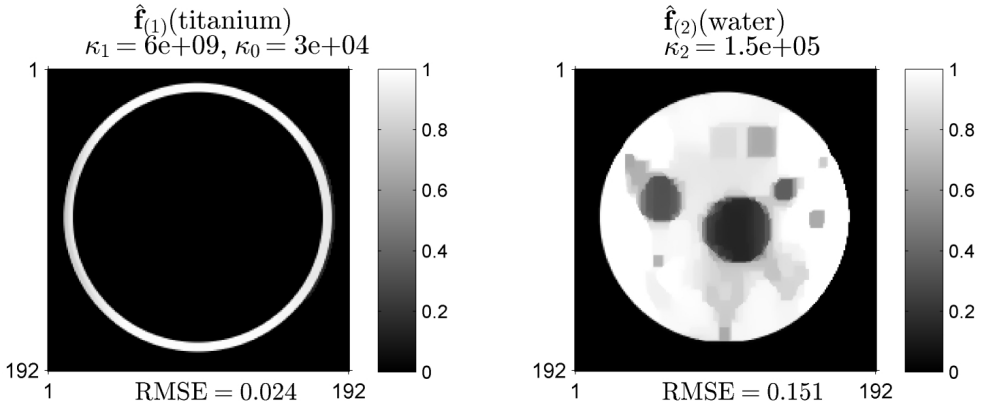


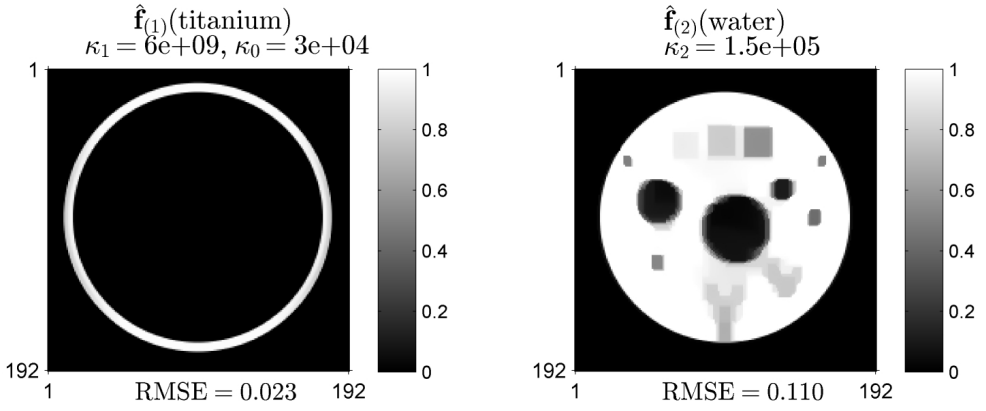
Figure 4.14: Modulation transfer function obtained from Phantom 1; (a) sample edge location, (b) sampled edge spreading, (c) line spreading, and (d) modulation transfer function at the sampled edge. Image pixel size is 0.520 mm, and corresponding Nyquist frequency is 0.96 lp/mm^2.

Table 4.1: Comparison of true, measured and simulated void fraction data at the sampled locations shown in Fig. 4.17

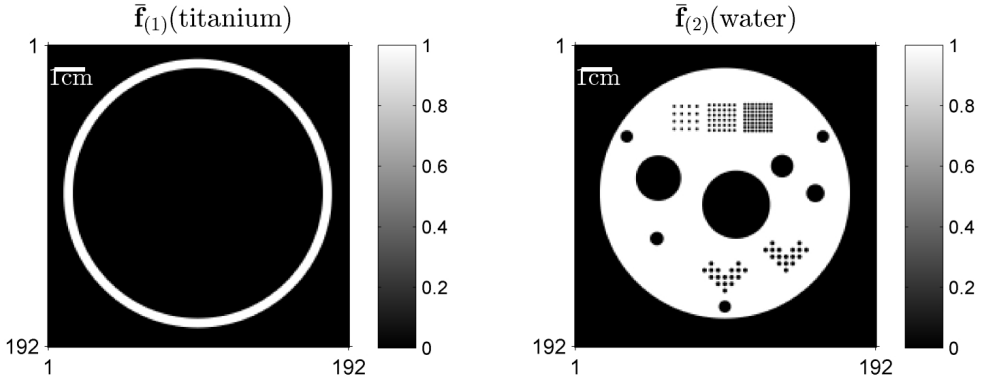
Region	true	measured	simulated	true - measured	true - simulated
(a)	1.000	0.837	0.952	0.16	0.05
(b)	0.100	0.036	0.041	0.06	0.06
(c)	0.226	0.117	0.161	0.11	0.07
(d)	0.401	0.258	0.338	0.14	0.06
(e)	0.112	0.113	0.082	0.00	0.03
(f)	0.236	0.163	0.168	0.07	0.07



(a) Real data, corrected

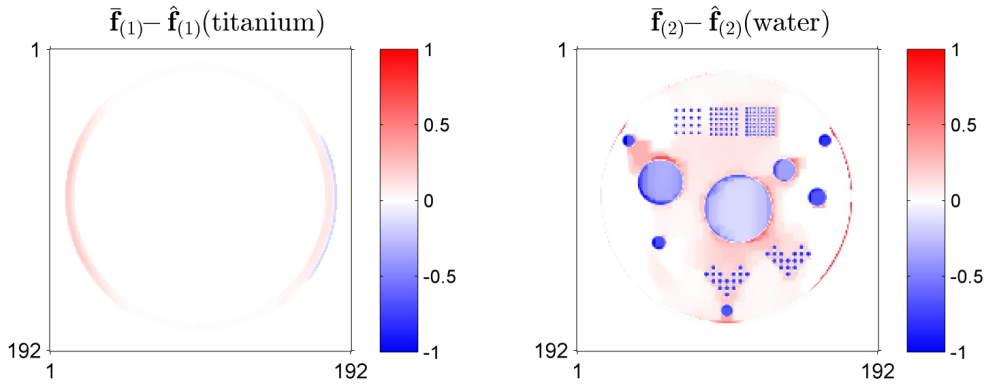


(b) Simulated data

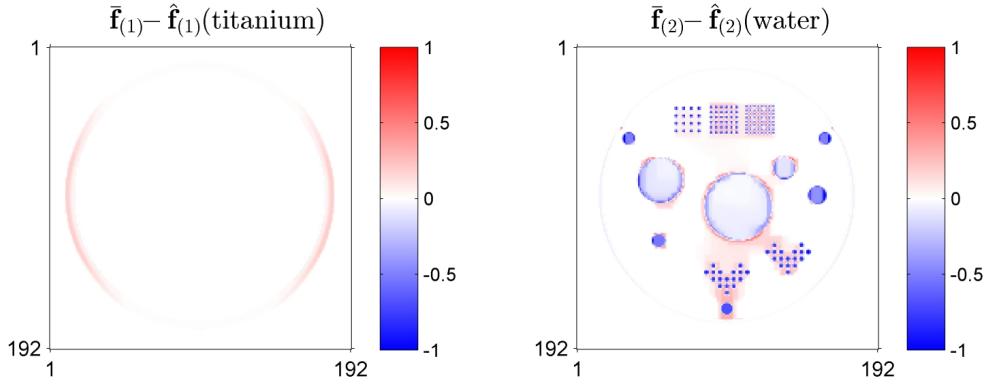


(c) True image

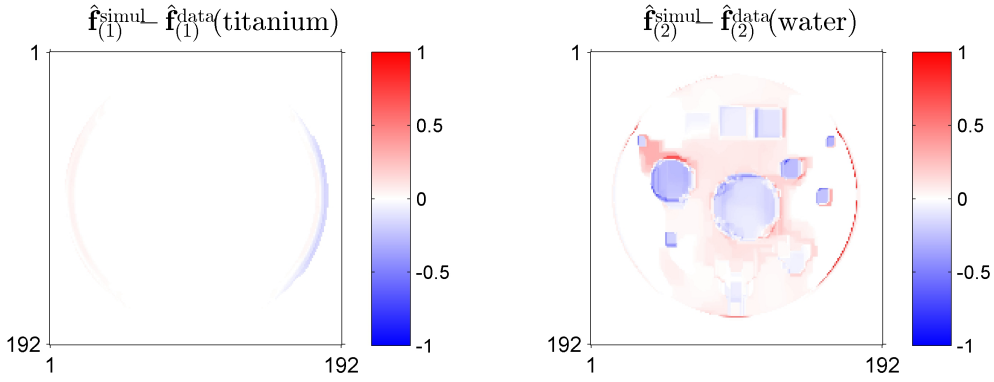
Figure 4.15: CT reconstructed volume fraction image based on (a) the corrected measurement sinogram and (b) the simulated sinogram for Phantom 2. True volume fraction image of Phantom 2 is shown in (c).



(a) Real data, corrected



(b) Simulated data



(c) Simulated data - real data

Figure 4.16: (a) error image of the CT reconstructed image based on the corrected measurement data; (b) error image of the CT reconstructed image based on the simulated data; (c) deviation between the corrected measurement CT result and the simulated CT result for Phantom 2.

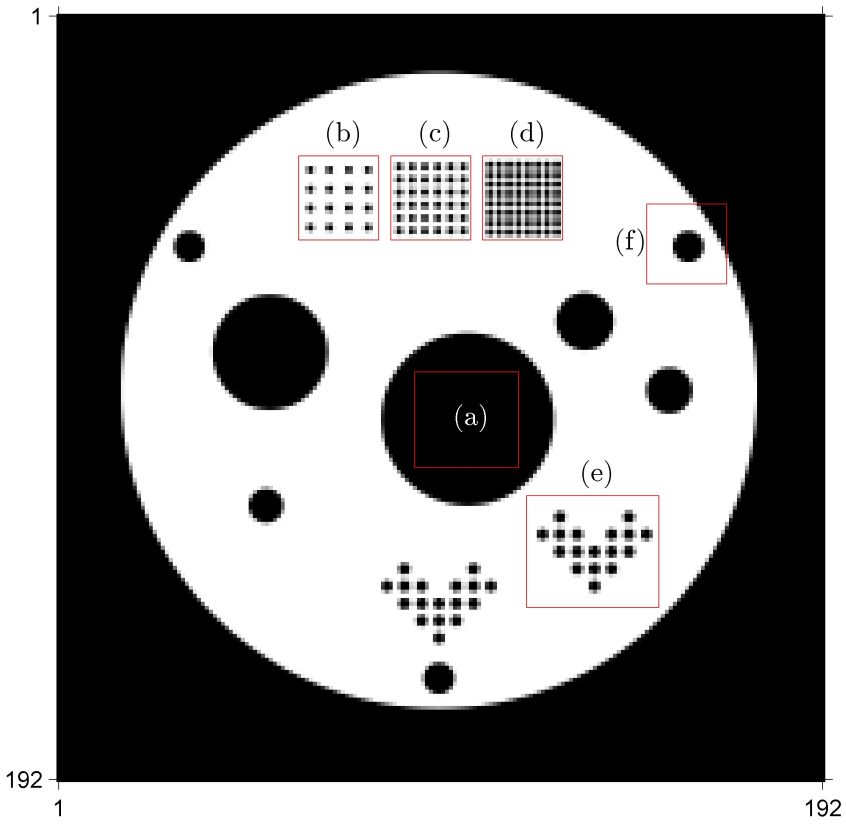
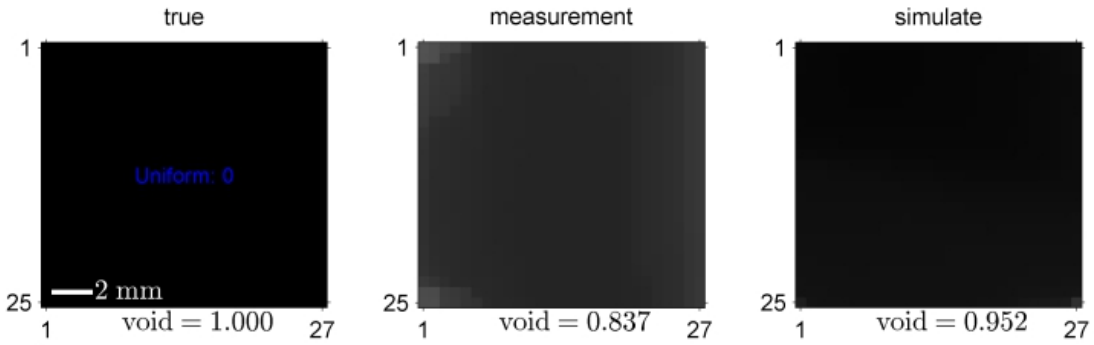
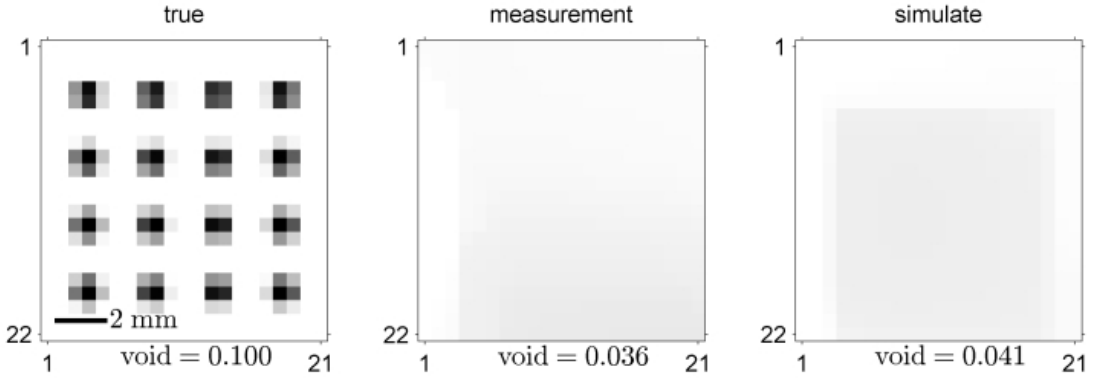


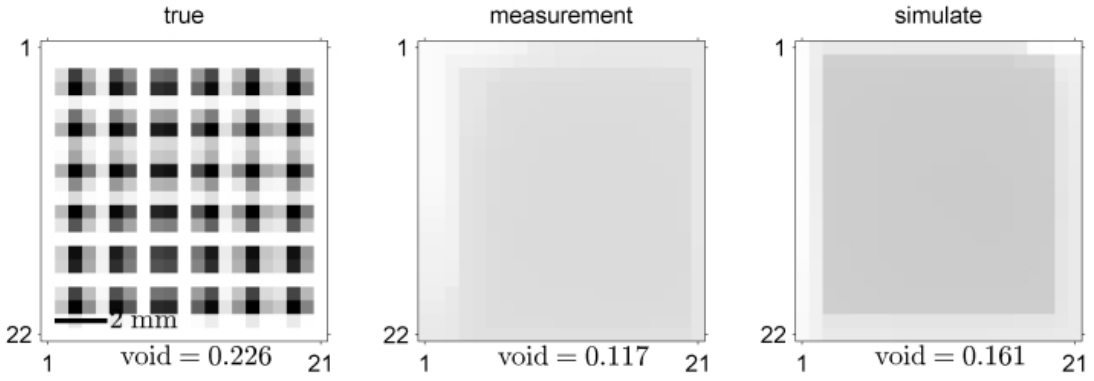
Figure 4.17: Sampled areas to test the accuracy of feature size void fraction shown in Fig. 4.18.



(a)

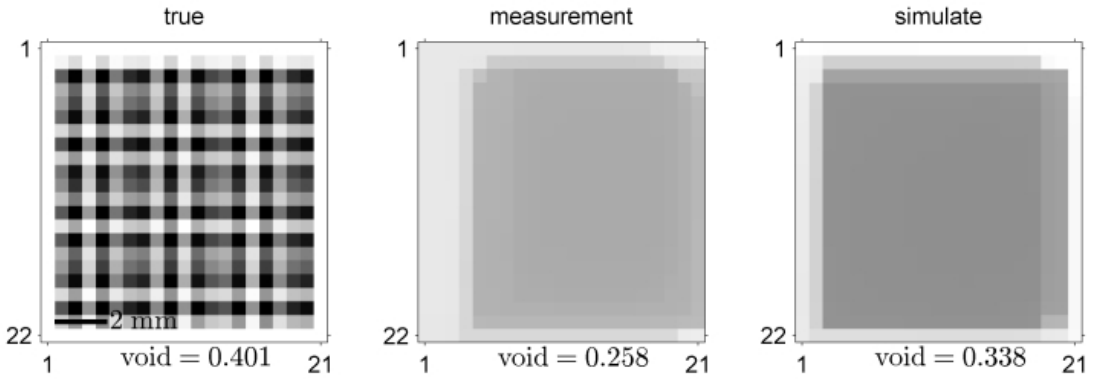


(b)

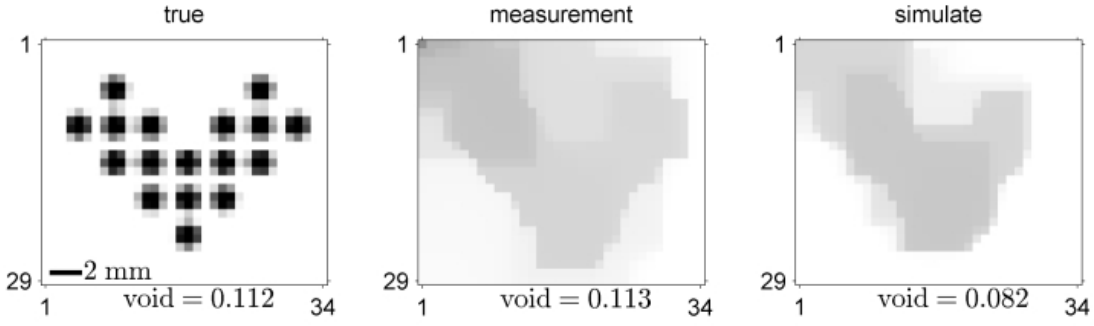


(c)

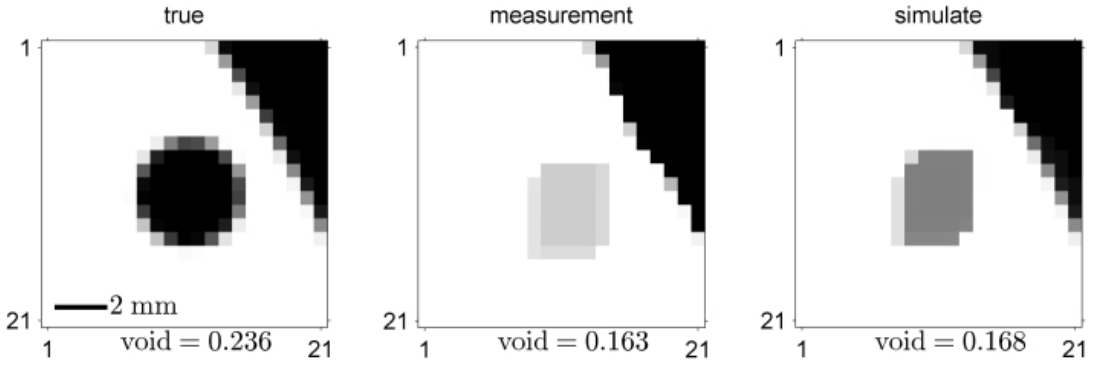
Figure 4.18: Sampled images and the resulting average void fractions of the true (left), measured (middle) and simulated (right) images. The locations of the sampled areas for (a) to (f) are shown in Fig. 4.17



(d)



(e)



(f)

Figure 4.18: (cont.) Sampled images and the resulting average void fractions of the true (left), measured (middle) and simulated (right) images. The locations of the sampled areas for (a) to (f) are shown in Fig. 4.17

4.5 Effect of the focal spot and detector sizes on the quality of CT reconstruction evaluated based on the simulations

In X-ray CT, there are many factors affecting the resolution of the CT reconstructed images, number of projected images, number of detector pixels, focal spot size, detector size, and X-ray scattering, to give a few examples. In this section, we evaluate the effect of the focal spot and detector sizes on the resolution of the CT reconstruction using a series of simulation results with different settings.

We consider two different acquisition modes, low speed and high speed acquisition modes. Low speed acquisition mode represents the current measurement condition, where we used the slow response photon counting detector with a 0.5 inch steel filter to reduce the X-ray photon flux down to $O(10^{5\sim6})$ cps. Maximum total photon counts of the meaningful projection (meaningful projection means a set of beam attenuated by the phantom so that it is actually used for the CT reconstruction) is approximately 6×10^4 when the detector size is 3×3 mm. For the low speed acquisition mode, the source has a 115 kV, 1 mA beam with 0.5 inch steel filter on the detector side with integral time 0.5 seconds.

Fast acquisition mode is what the SEBXT is designed for, but cannot be used because the charge-integrating detector was not available. For the fast speed acquisition mode, the source has 115 kV, 133 mA beam without any filter. Integral time is only 5 μs. The cps of high speed acquisition mode is a lot higher than the low speed acquisition mode, amount to $O(10^{11\sim12})$ cps. However, maximum total photon counts of the meaningful projection is only 1×10^4 for high speed acquisition mode when the detector size is 3×3 mm, which is 6 times lower than the low acquisition mode. Thus, the noise level will be higher in high speed acquisition mode, and the optimal detector size can be different in low and high acquisition modes.

For each mode, we ran simulations for four different focal spot sizes, FWHM 1 to 4 mm, and three different detector sizes 0.5×0.5, 1×1, and 3×3 mm^2. Additionally, we added the simulation results with a detector size 0.7×0.7 mm^2 for the fast acquisition mode, which is the size of the pixel in the charge-integrating arc detector that we are going to use in the future. Note that the scattering effect is not considered in the simulated sinograms. Actual image quality might depend on the scattering as we cannot apply collimators due to the varying incident angles.

4.5.1 Low speed acquisition mode

Fig. 4.19 and 4.20 show the RMSE values of the reconstructed image of Phantom 1 and 2 at low speed acquisition setting based on the simulated sinograms, respectively. Intuitively, smaller focal spot size is always better in terms of the image resolution (although it is easier to damage the tungsten target physically when the beam is more concentrated). However, detector size is a trade-off between the resolution and the noise. Smaller detector size means less blurriness, but the photon flux also decreases, inducing more noise.

To quantify the resolution of the reconstructed image, we define f_{10} be the frequency in line pairs per millimeter (lp/mm) where the value of MTF crosses 10% of the maximum value (maximum value is 1 in our case because the reconstructed image is non-dimensional). The higher f_{10} is, the better the CT reconstruction can resolve the smaller object. Fig. 4.21 shows f_{10} with detector size 3×3 mm^2. For detector size smaller than or equal to 1×1 mm^2, f_{10} is greater than the Nyquist frequency even with the maximum focal spot of FWHM 4 mm.

4.5.2 High speed acquisition mode

Fig. 4.22 and 4.23 show the RMSE values of the reconstructed image of Phantom 1 and 2 at high speed acquisition setting based on the simulated sinograms, respectively. The increase in error in the case of detector size 0.5×0.5mm is more prominently shown in the high speed acquisition mode, because noise level is higher than that of the slow acquisition mode. Among the detector sizes we considered, 1mm^2 had the best performance in terms of RMSE. However, we can further increase the resolution of the reconstructed image by decreasing the pixel size when the detector face size is small. In that case, smaller detector might have more benefit to resolve the small objects.

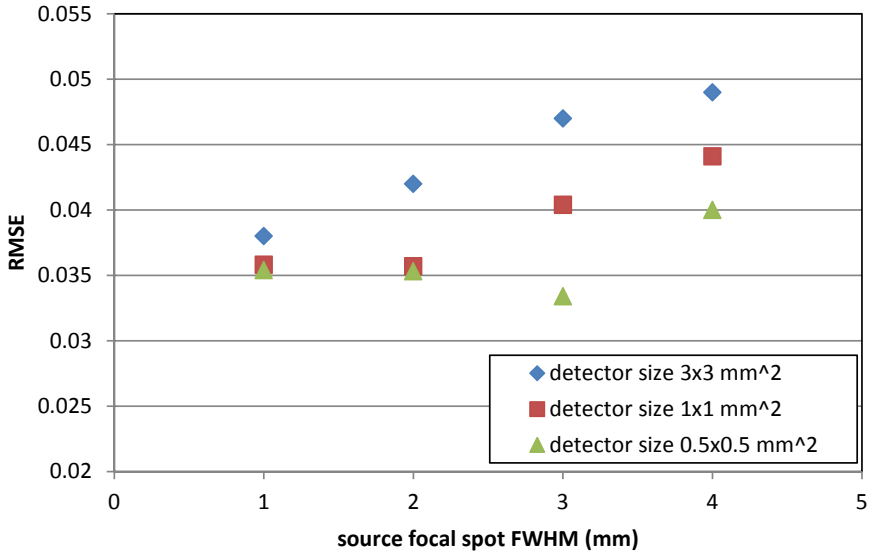


Figure 4.19: RMSE of the simulated CT reconstruction results for Phantom 1 at the low speed acquisition setting.

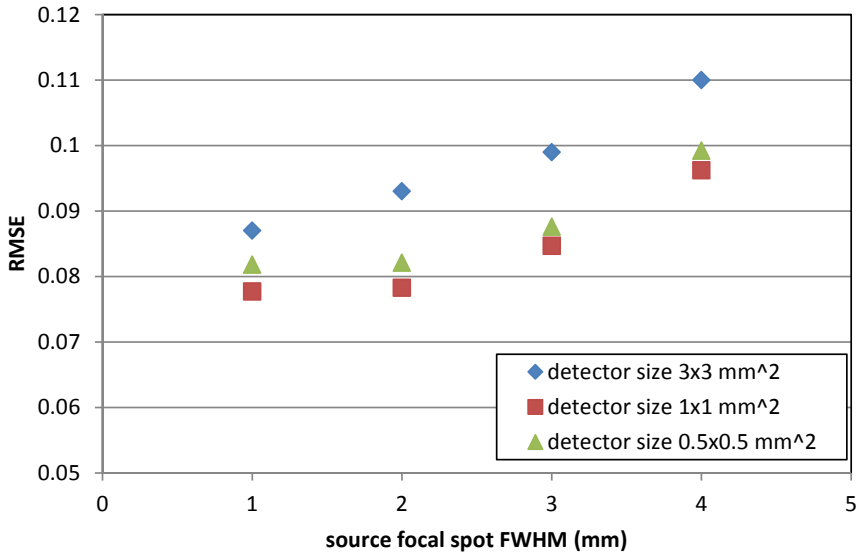


Figure 4.20: RMSE of the simulated CT reconstruction results for Phantom 2 at the low speed acquisition setting.

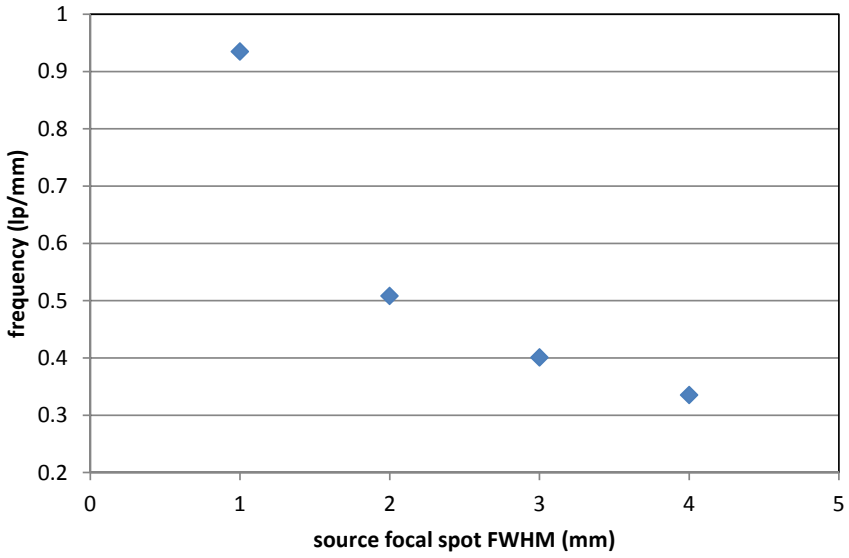


Figure 4.21: Frequency at 10% of the maximum MTF of the simulated CT reconstruction results for Phantom 1 with the detector face size 3×3 mm^2 at the low speed acquisition setting.

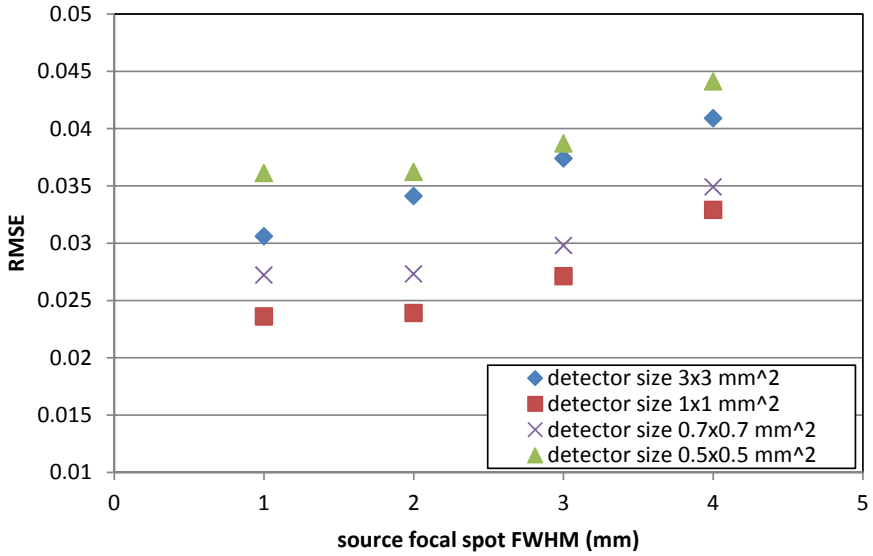


Figure 4.22: RMSE of the simulated CT reconstruction results for Phantom 1 at the high speed acquisition setting.

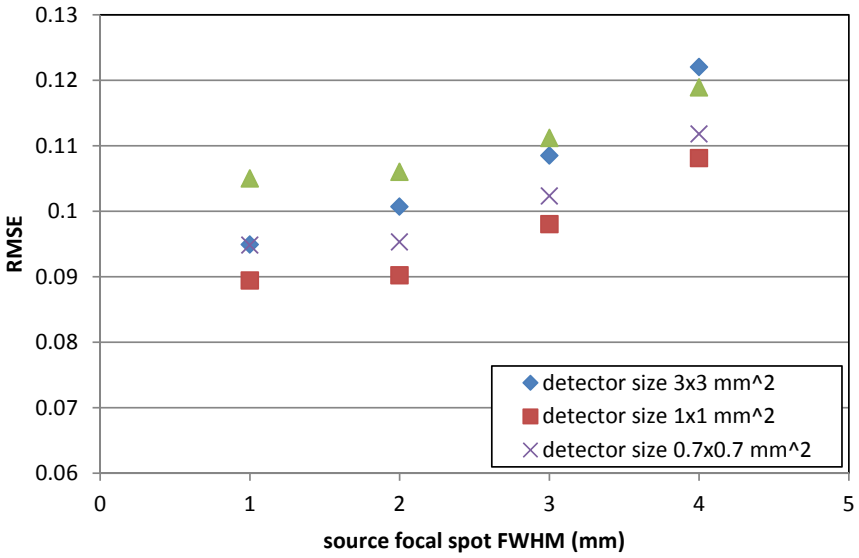


Figure 4.23: RMSE of the simulated CT reconstruction results for Phantom 2 at the high speed acquisition setting.

BIBLIOGRAPHY

[1] J. H. Hubbell and S. M. Seltzer, "Tables of X-ray mass attenuation coefficients and mass energy-absorption coefficients 1 keV to 20 MeV for elements Z= 1 to 92 and 48 additional substances of dosimetric interest," National Inst. of Standards and Technology, Tech. Rep., 1995.

[2] G. Poludniowski, G. Landry, F. DeBlois, P. M. Evans, and F. Verhaegen, "SpekCalc: a program to calculate photon spectra from tungsten anode x-ray tubes," *Phys. Med. Biol.*, vol. 54, no. 19, p. N433, 2009.

[3] F. Fischer, D. Hoppe, E. Schleicher, G. Mattausch, H. Flaske, R. Bartel, and U. Hampel, "An ultra fast electron beam x-ray tomography scanner," *Meas. Sci. Technol.*, vol. 19, p. 094002, 2008.

CHAPTER 5

Conclusions: Electron Beam X-Ray Computed Tomography for Multiphase Flows

We reviewed the electron beam X-ray system that is being developed at the University of Michigan and described the basic principles of its design and operation. As the energy-integrating X-ray arc detector was not available, we described how we use a single pixel spectrometer to generate data similar to the data generated by the energy-integrating arc detector array. Therefore, we measured the intact energy spectrum using the single pixel spectrometer and a tungsten collimator with 100 μm diameter aperture. The measured intact energy spectrum of the X-ray matches well with the simulated X-ray spectrum when the spectral response of the detector is applied.

To characterize the beam spot size and location, we estimated the focal spot profile using the iterative regularized and constrained least square method based on the images of the tungsten collimator with 1 mm diameter aperture. As a result, we get the minimum full-width-half-maximum of the beam profile being 1.18 mm at the beam current 1050 mA.

We characterized the relationship between the deflection voltages applied to the electron beam and the ultimate location of the focal spot on the target. This data was not provided by the manufacturer, so we performed a localization test for seven different angles. We used an acrylic phantom with seven pairs of tungsten carbide balls to determine the position where the electron beam meets the tungsten target. Based on those data, we estimated the mapping function between the deflection voltages and the physical position of the beam as a form of a linear matrix system based on the least square estimation.

Finally, we introduced three phantoms that was used for the beam alignment (Phantom 0), and will be used to test the performance of the CT reconstruction (Phantom 1 and Phantom 2).

We developed a CT reconstruction method for the Scanning Electron Beam Tomography (SEBXT) system characterized by a limited angle, single bin energy integrating measurement. For the data-fitting term, the method involved a nonlinear

Gaussian model for the energy-integrating statistics combined with the volume fraction material decomposition method.

The reconstruction method employed several assumptions that may affect the quality of the result in practice. Those include no charge sharing, no scattering of X-ray photons, ideally pulsed and focused electron beam. Without collimators, the scattering effect is an important factor as the power of scattered beam can reach up to 78% of the primary beam [1]. Unfortunately, collimator cannot be used in SEBXT due to the wide variation of incident angles. Those scattering would show up as a blurred sinogram, and eventually a blurred reconstructed image. The blurring induced by the X-ray scattering is not linear shift-invariant, and it depends on all the materials that the beam has passed through, which makes it more difficult to estimate and invert. The spectral response depends on the type of the electric circuits combined with the detector crystal. One example is the incomplete charge collection. Holes are slower than electrons, a portion of holes may not be collected due to their limited mobility. Those electric circuit effects are not considered in the spectral response model. In reality, an experiment, ideally with a monochromatic source, to acquire the spectral response of the specific detector is required. This issue will be considered when the detector array is installed in the system.

The ℓ_0 norm and the edge-preserving hyperbola regularizers are applied accordingly depending on the characteristics of the materials. The role of regularizers is to provide additional removal of small local volume fraction errors and limited angle artifacts. As two regularization schemes are combined, the ℓ_0 norm and the edge-preserving hyperbola potential regularizer are connected by the curvature, and competing with each other. As a result, actual strength of the ℓ_0 norm regularizer depends on the strength parameter of the edge-preserving regularizer indirectly, as well as its own strength parameter. Finding optimal values for the strength parameters may require combining the strength parameters into a single parameter.

We used an asymptotic curvature along with the separable quadratic surrogate functions to simplify the minimization step while maintaining a reasonable convergence speed. However, this asymptotic curvature has not been proven to monotonically decrease the objective function for all cases. A further study regarding the convergence may be required in the future.

Interpolation based spatial basis computation and compacting algorithm are used to increase the GPU computing speed. The compacting algorithm is effective when the register spilling is the main factor of the bottleneck as it does not require any register space other than registers used for the Radon transform computation. However, performance of the compacting has not been compared with other methods, and the result can be different

when there are enough registers in each thread. More extensive comparison will be required to assess the efficiency of the compacting.

The reconstruction results shown in Fig. 4.12 and 4.15 showed that the present SEBXT system captures the object between the source and the detector. However, the accuracy of detecting local structures was much lower than expected, as shown in Table 4.1. Below is the list of problems identified during the beam calibration and CT reconstruction process that cause the low accuracy.

1. Electron beam focal spot

   The source's focal spot was not focused adequately, causing blurry images. It is possibly due to the interaction between electrons and gas molecules in the chamber. Two possible sources of gas molecules exist. First, the turbo-molecular pump installed on the upper part of the chamber cannot effectively remove the gas molecules at the bottom of the chamber, thus it is possible that the pressure at the lower part of the gun is much higher than the pressure level indicated by the upper chamber pressure gauge. Second, slow acquisition in low power resulted in heating the cathode for too long. Overheated cathode can cause the outgassing, another source of the gas molecules. To improve the focal spot, additional turbo pump will be installed at the lower chamber in the future. Cathode overheating problem can be solved once the machine can be operated in high speed acquisition mode.

   The effect of focal spot varies with the relative location of each object, source and detectors, thus the linear shift-invariant blur kernel applied to the simulation does not reflect the actual effect of the focal spot. If focal spot size is still a significant factor after the installation of the additional turbo pump, the effect of the finite focal spot can be added as a matrix form in the system matrix, then solved iteratively along with the CT reconstruction.

2. Scattered X-ray

   Conventionally, detectors have collimators to block scattered X-rays. However, detectors on the SEBXT cannot use collimators because the angle between the source and detector is variable due to the fixed configuration of the detectors. The code does not directly include the effect of scattered photons that may reach the detector. This is expected to cause a notable discrepancy near the high gradient regions. A software approach similar to the item 1 may be considered to reduce the blurring due to the scattered X-ray, i.e., build a matrix that represents the scattering, then solve iteratively along with the CT reconstruction.

3. Pulse pile-ups

   The SEBXT is not designed to be used along with photon-counting detectors. The single pixel spectrometer used for sinogram acquisition was not capable of fully collecting all charges, and the count rate often exceeded the capacity of the detector to measure accurate spectrum due to pulse pile-ups. The presently utilized photon-counting detector can count the photons with reasonable accuracy up to $O(10^5)$ counts per second (cps), while the source at 1 mA generates $O(10^6)$ cps after 0.5 inch steel filtering, and this can cause up to 30% loss of total counts due to pile-up. This effect makes the response of the detector nonlinear. A correction has been made to linearize the response, but there are uncertainties related to the pile-up events. (Note: We had to operate the source at relatively higher current than the detector capacity, due to instability in beam current with 0.1 mA fluctuation.) In the future, charge-integrating detectors will be used, which give almost linear response in the operating cps range. The correction function may not be required when the new detector is fully operational.

4. Beam current instability

   The beam current fluctuations were of the order of 10% for the 1 mA beam current. This leads to higher noise compared to the simulation, in which the beam current is assumed to be constant and only photon production uncertainty is accounted for. However, the effect of the power fluctuation would relatively decrease by factor of 1/133, less than 0.1% of the signal, if the source was operated at full power.

5. Low void fraction

   In Table 4.1, void fraction is always lower than the true value for both the corrected measurement and simulation data. Despite of its edge-preserving characteristics, strong regularizer tends to make the image flat. Regularizer strength parameter in this chapter was unusually large to correct the error due to the beam current uncertainty, which was not considered in the probability model. This low void fraction problem will be disappeared by lowering the regularization strength once we are able to operate the source at full power.

6. Acquiring intact emission spectrum

   To get the best correction image, we need an empty scan data set. (Empty scan means that scanning without any objects, including the pipe, in the measurement domain.) We cannot acquire the empty scan data as the photon energy spectrum was distorted too much due to the excessive pile-up events. Thus, we used an empty pipe image to correct the unknown obstacles such as the actual source's aluminum wall thickness. The correction map had errors due to the uncertainty of the pipe location.

The uncertainty of the pipe location is less than 1 mm, but that is enough to generate a significant error.

7. Large detector face size
   The detector that we used has the face area of $3 \times 3\text{mm}^2$, which is not designed for the CT reconstruction. Part of the blurriness in the reconstructed image comes from the large size detector as shown in §4.5. This can be easily improved by adopting smaller size detectors. The forthcoming charge-integrating arc detector has face size of 0.7×0.7mm, As shown in the numerical simulations in §4.5, we can potentially improve the reconstructed images once the arc detector is available.

In conclusion, we have characterized the X-ray source for the SEBXT, we have developed a novel CT reconstruction algorithm, and we experimentally implemented the CT reconstruction method for two phantoms. By doing this, we have readied the system for the installation of the fast detector array. Future work will include the integration of the array into the SEBXT control system, characterization of the array response with calibrated X-ray sources, and the characterization of the interaction of the SEBXT X-ray beam with the detectors. The new detector array has gaps between modules. We may need a new simulation code to account for the effect of those gaps. Once this is done, we will return to measurement of the two phantoms.

BIBLIOGRAPHY

[1] M. Endo, S. Mori, T. Tsunoo, and H. Miyazaki, "Magnitude and effects of x-ray scatter in a 256-slice CT scanner," *Med. Phys.*, vol. 33, no. 9, pp. 3359–3368, 2006.