Advanced Imaging Algorithms with Position-Sensitive Gamma-Ray Detectors

by

Jiyang Chu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Nuclear Engineering and Radiological Sciences)
in The University of Michigan
2018

Doctoral Committee:

Professor Zhong He, Chair
Professor Jeffrey A Fessler
Professor Igor Jovanovic
Professor William R Martin

Jiyang Chu

chujy@umich.edu

ORCID iD: 0000-0003-4318-7103

To Yishu

ACKNOWLEDGEMENTS

First and foremost, I am most grateful to my advisor, Professor Zhong He, who provided an environment where I could enjoy the freedom of research and grew with firm support. I have learned so much from his insight of science and understanding of the world. I would like to express my special gratitude to Professor Jeffery Fessler, whose wisdom guided me through the sea of difficulties in my research and to the new continent of discoveries. Im also thankful to Professor Igor Jovanovic and Professor William Martin for serving on my committee and providing diversified perspectives to my thesis.

It is fortunate that I was surrounded by a group of smart and kind people during the graduate period. I enjoyed the brainstorm with Dr. Steven Brown, David Goodman and Daniel Shy, which always sparked innovations and clarified thoughts. Many thanks to the outstanding work by Dr. Yuefeng Zhu, Dr. Michael Streicher, Jiawei Xia, and Bennett Williams on the 3D position-sensitive detectors. The algorithms in this thesis would only be on paper without their help. I appreciate Dr. Will Koehler, Dr. Sean ONeal, and Charles Leak for bringing fresh ideas into the group and keeping me challenged by asking sharp questions. Many of them mentioned above did proofreading and gave suggestions on this thesis, and I am indebted to them. Thanks also go out to the H3D team, including Dr. Willy Kaye, Dr. Weiyi Wang, Dr. Feng Zhang, Dr. Christopher Wahl, Dr. Hao Yang, Dr. Jason Jaworski, and other members working on the CZT system. Their pioneer work inspired my research and their enthusiasm always encouraged me during my graduate time.

Five years are too long for me to remember every moment, and too short for me to forget any emotions. Particularly I would like to thank Yishu Wang, for her love when I knew nothing. I am very grateful to Shigeng Li and Jinyang Li for their brotherhood. Many thanks to Haoran Xu and Peizheng Ma for their friendship, as well as the other members of team *Special Fleet* for the joy they brought to my life. At last, I would like to give my deepest gratitude to my parents, Xiong Chu and Xiangmei Li, for their incredible support from the other side of the Earth.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF ABBREVIATIONS

ASIC Application-Specific Integrated Circuit

CAI Coded Aperture Imaging

CZT Cadmium Zinc Telluride

DSHT Discrete Spherical Harmonics Transform

EDID Energy Decremental Integrated Deconvolution

EIID Energy-Imaging Integrated Deconvolution

FBP Filtered Back-Projection

FFT Fast Fourier Transform

FT Fourier Transform

FWHM Full Width at Half Maximum

GUI Graphical User Interface

ML Maximum-Likelihood

MLEM Maximum-Likelihood Expectation-Maximization

OSEM Ordered-Subsets Expectation-Maximization

PET Positron Emission Tomography

PSF Point Spread Function

RMC Rotating Modulation Collimator

SBP Simple Back-Projection

SH Spherical Harmonics

SHT Spherical Harmonics Transform

SPECT Single Photon Emission Computed Tomography

UMIS University of Michigan Imaging System

WSS Wide-Sense Stationary

ABSTRACT

The 3D position-sensitive Cadmium Zinc Telluride (CZT) detector can provide position and energy information of gamma-ray interactions, enabling imaging capability. Previous imaging methods are able to reconstruct distinguishable source locations. However, there exists a major contradiction in imaging problem eternally: information rate and reconstruction speed. The information rate represents how much information can be extracted from a measurement, including resolution, variance, and noise level. The reconstruction speed relies on the computing hardware heavily, but can be greatly accelerated by sophisticated reconstruction algorithms.

To improve the combined performance of information rate and reconstruction speed, several imaging algorithms were investigated. Among them, the Simple Back-Projection (SBP) provides fastest reconstruction speed but least information due to its biased and blurred estimation. The Filtered Back-Projection (FBP) is almost as fast as SBP, but much more informative. However, better understanding of FBP is required to let it work for real data without too much manual intervention, thus the adjusted FBP and the adaptive FBP were developed in this thesis to make FBP practical. On the contrary, the Energy-Imaging Integrated Deconvolution (EIID) is a very informative reconstruction algorithm, but too slow to be applicable in many time-sensitive scenarios. The convergence rate of EIID should be accelerated without too much degradation to the information, thus the Energy Decremental Integrated Deconvolution (EDID) was developed in this thesis. Finally 3D image reconstruction is proposed, which provides a new dimension to interpret source distribution compared

to traditional 4π spherical image reconstruction, and is more natural for people to understand. The image reconstruction code University of Michigan Imaging System (UMIS) was designed and developed during the research period. The principle of UMIS is to keep the code fast, simple, reusable, and suitable for research.

CHAPTER I

Introduction

1.1 Principles of Compton Imaging

Gamma photons can be detected by interactions with matter. There are mainly three types of interaction mechanisms: photoelectric absorption, Compton scattering, and pair production. Among them the Compton scattering can provide angular information, which enables Compton imaging. As illustrated in Fig. 1.1, in a Compton scattering process, a photon with initial energy E_0 interacts with an electron, transfers part of its energy, and scatters by an angle θ with scattering energy E_2. The electron get deposited energy E_1 and is recoiled by an angle ϕ.

Given the initial energy E_0, the deposited energy E_1 and the scattering energy E_2, the scattering angle θ can be calculated by the well-known Compton scattering formula [1]

$$\theta = \cos^{-1}(1 - \frac{m_e c^2 E_1}{E_0 E_2}), \tag{1.1}$$

where $m_e c^2$ is the rest mass energy of an electron.

If the positions and the deposited energy of interactions are known, the incident photon direction can be constrained to the surface of a cone using the Compton scattering formula. As shown in Fig. 1.2, imagine the cone in a 3D space, the

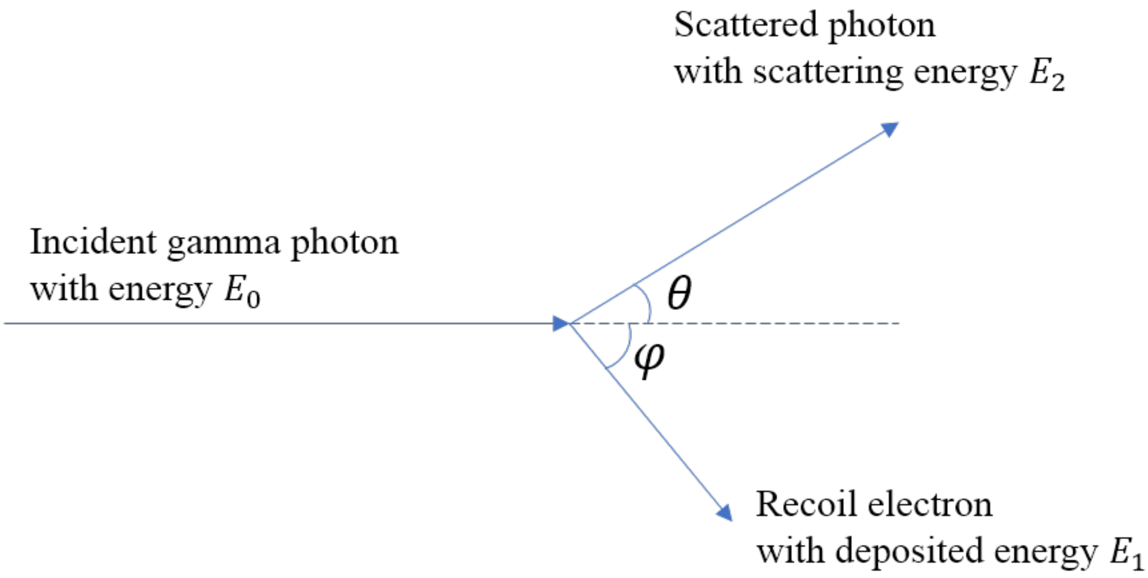


Figure 1.1: A gamma photon Compton scattered by an electron.

direction of the incident photon is constrained on the surface of the cone with some spacial uncertainties. Using the directional information, it is possible to reconstruct the source distribution or radiation image.

1.2 History of Compton Imaging

The detection of high-energy gamma photons is always a challenge due to the low attenuation coefficient and complicated interactions with matter. However, people still have interest in gamma photon detection in fields including nuclear medicine, nuclear safety, nuclear non-proliferation, and astrophysics. Various imaging systems have been developed, such as Single Photon Emission Computed Tomography (SPECT) [2], Positron Emission Tomography (PET) [3], Coded Aperture Imaging (CAI) [4], Rotating Modulation Collimator (RMC) [5], and Compton imager [6, 7]. Compton imager is quite attractive because it uses Compton scattering mechanics instead of mechanical collimators to provide angular information with high detection efficiency. Table 1.1 walks through the history of Compton imagers, from early 1970s

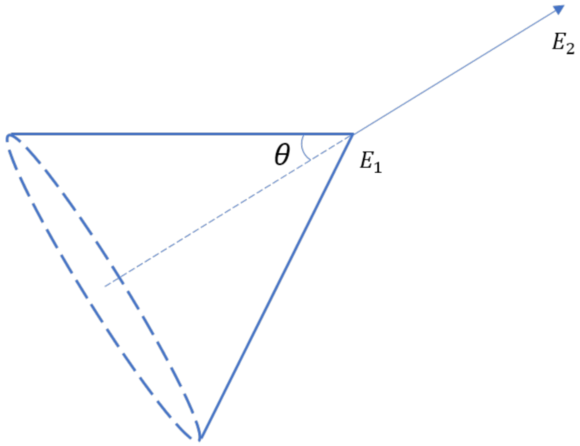


Figure 1.2: A gamma cone reconstructed from a double-interaction event, where the first interaction is a Compton scattering with deposited energy E_1 and the second interaction is a photon absorption with deposited energy E_2.

to recent 2010s.

Year	**Authors**	**References**	**Description**
1973	Schönfelder *et al.*	[8]	The concept of Compton imaging.
1974	Todd *et al.*	[9]	The concept of Compton imaging.
1983	Singh *et al.*	[6, 7]	The first Compton camera system for medical imaging.
1988	Kamae *et al.*	[10]	A Compton imager which replaced the HPGe detectors in Singhs design with many layers of Si strip detectors. The Si strip detectors were surrounded by a cylindrical CsI(Tl) scintillator.

1989	Piercey *et al.*	[11]	The first Compton telescope with HPGe detectors for both the front and back planes. A single HPGe detector was used as the front detector and an array of four HPGe detectors were used as the back plane.
1990	Dogan *et al.*	[12]	Reconstruct the image using multiple scattered gamma-rays based on Kamaes design.
1993	Martin *et al.*	[13]	A ring Compton scatter camera which consisted of a 4×4 array of HPGe detectors and a ring array of cylindrical NaI(Tl) scintillators.
1994	McKisson *et al.*	[14]	A Compton camera using two HPGe detector arrays for both the front and back planes.
1996	Phlips *et al.*	[15]	Replaced the HPGe detector arrays by two double-sided HPGe strip detectors. The front detector was a 25×25 strip detector with 2 mm pitch, and the back detector was a 5×5 detector with 9 mm pitch.

1998	LeBlanc *et al.*	[16, 17]	A prototype Compton camera system, C-SPRINT, for nuclear medicine as a Compton SPECT. Instead of using Si strip detectors, C-SPRINT consisted of $3 \times 3 \times 0.1$ cm^3 Si pad detector modules that were pixellated into 22×22 arrays.
2001	Schmid *et al.*	[18]	A Compton camera by employing a single large volume segmented coaxial HPGe detector. The position information was obtained by the outer contacts and digital pulse-shape analysis. By using a single crystal to detect both the Compton scattering and the photoelectric absorption, the separation between the front and back planes was eliminated, resulting in efficiency being increased by one order of magnitude. However, due to the large preamplifier noise, the experiments could not locate the source position.

2001	Du *et al.*	[19]	Used CZT as the crystal in a Compton camera for the first time. The CdZnTe detectors used in Dus work measured the 3D position information of single gamma-ray interactions in the detector. Since a Compton camera needs at least two gamma-ray interactions, two 1 cm^3 3D position-sensitive CdZnTe detectors were used in Dus experiment. The two detectors were separated by 5 cm, which limited the intrinsic efficiency at 511 keV to be only 1.5×10^4. The measured angular resolution was 5.2° at 662 keV.
2002	Kroeger *et al.*	[20]	Employed the concept of three-Compton telescope. The incident gamma-ray energy can be determined even if the gamma-ray is not fully absorbed, if the incident gamma-ray interacts three or more times in the detectors.

2002	Wulf *et al.*	[21]	Demonstrated that the depth information of a HPGe double-sided strip detector could be obtained by measuring the timing difference between the cathode and anode signals. Thus a Compton camera could be realized by using a single HPGe double-sided strip detector.
2003	Wulf *et al.*	[22]	Designed a Compton imaging system with two HPGe double-sided strip detectors and demonstrated that the source location could be reconstructed by the three-Compton technique even if gamma-rays did not deposit full energy in the detectors.
2004	Wulf *et al.*	[23]	Built a Compton camera using three layers of double-sided silicon strip detectors. Each silicon detector had an active volume of $57 \times 57 \times 2$ mm^3, and had 64 strips on each side. The Compton camera showed an angular resolution of 3.3° at 662 keV.

2004	Zhang *et al.*	[24]	Obtained the 3D position information of multiple gamma-ray interactions in a single detector by measuring the electron drift time with the improved design of the Application-Specific Integrated Circuit (ASIC). This advance greatly improved the gamma-ray detection efficiency because about 60% of the photopeak events are multiple interaction events at 662 keV in a 15 × 15 × 10 mm^3 detector.
2004	Lehner *et al.*	[25]	1.86% measured intrinsic imaging efficiency, and 17° measured angular resolution at 662 keV using a single 15 × 15 × 10 mm^3 CZT detector.
2005	Niedermayr *et al.*	[26]	Obtained about 5° angular resolution at 662 keV with a relative efficiency of 0.3% with the same design as Schmids.
2005	Burks *et al.*	[27]	A Compton telescope using a double-sided silicon strip detector and a double-sided germanium strip detector. The two detectors are parallel, aligned on their center axis and separated by 6 cm. An angular resolution of 3° to 4° was reported in the energy range from 150 keV to 3 MeV.

2011	Zhu *et al.*	[28]	Proposed the sub-pixel position sensing for Pixelated 3D Position Sensitive CZT detectors with a digital ASIC design, achieving 230 μm sub-pixel spatial resolution and less than 1% energy resolution.

Table 1.1: The history of Compton imagers.

Different reconstruction algorithms were investigated under the concept of Compton imaging. As most of the system responses of Compton imagers are very complicated in the sense of shift-variance and non-isotropic, probability-based iterative algorithms, such as Maximum-Likelihood Expectation-Maximization (MLEM), are widely used [29, 25, 30, 31]. Because iterative algorithms are usually computationally intensive, researchers continue to investigate direct inverse algorithms, such as FBP. FBP has advantages when real-time imaging is required and computational resources are limited.

In 2000, Parra [32] applied Spherical Harmonics (SH) in Compton imaging. The back-projection image in a 4π spherical space is transformed into SH domain, which is a generalized frequency domain, then the deconvolution uses the SH coefficients of an analytical Point Spread Function (PSF). In 2006, Xu *et al.* [33] developed a FBP algorithm in a 4π spherical space with SH and a simulated PSF, accounting for the effects of the scattering angle distribution due to the limited detector size. This algorithm is designed for a single cubic CZT detector or similar systems, assuming the PSF changes little for different incident directions. In 2015, Haefner *et al.* [34] proposed another filtering approach that embeds the 4π spherical space in a 3D Cartesian space and uses the 3D Fourier Transform (FT) and Radon transform.

Mathematically this method is equivalent to the SH method with an analytical PSF.

1.3 Objectives of This Work

The 3D position-sensitive CZT detector can provide the position and energy information of gamma-ray interactions, enabling imaging capability. Previous imaging methods are able to reconstruct distinguishable source locations. However, there exists a major contradiction in imaging problem eternally: information rate and reconstruction speed. The information rate represents how much information can be extracted from a measurement, including resolution, variance, and noise level. The reconstruction speed relies on the computing hardware heavily, but can be greatly accelerated by sophisticated reconstruction algorithms.

The main objective of this work is to develop imaging algorithms that improve the combined performance of information rate and reconstruction speed. As shown in Fig. 1.3, the ultimate goal is an ideal algorithm on the top right corner with the most information and fastest speed, which is the dream that everyone wants to catch. Among the realistic algorithms, SBP provides fastest reconstruction speed but least information due to its biased and blurred estimation. The FBP is almost as fast as SBP, but much more informative. However, better understanding of FBP is required such that it works for real data without too much manual intervention, thus the adjusted FBP and the adaptive FBP were developed to make FBP practical. On the contrary, the EIID method is a very informative reconstruction algorithm, but too slow to be applicable in many time-sensitive scenarios. It is necessary to accelerate the convergence rate of the EIID without too much degradation to the information, thus the EDID was proposed in this thesis. Finally 3D image reconstruction provides a new dimension to interpret source distribution compared to traditional 2D image reconstruction, and is more natural for human beings to understand. We use online updating algorithm to make it faster than the conventional MLEM in 3D.

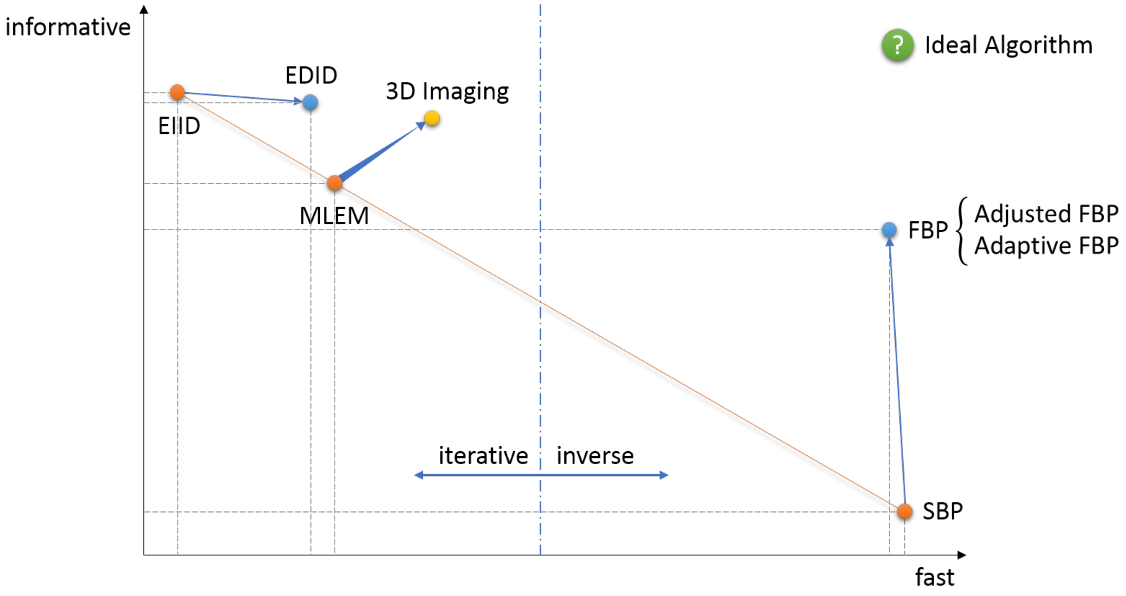


Figure 1.3: The performance of image reconstruction algorithms in the speed-information space. EIID, MLEM, and SBP lie on the conventional line, which is not necessarily straight. The thick arrow from MLEM to 3D imaging indicates that 3D imaging has significant different input and output compared to 2D imaging.

1.4 Overview

Chapter 2 shows the Shockley-Ramo theorem, simply describes the characteristics of pixelated CZT detector – the position-sensitive instrument used in this study, and analyzes the imaging model and system model that will be focused on. Chapter 3 studies two major direct reconstruction algorithms – SBP and FBP. Also two variations of FBP – adjusted FBP and adaptive FBP, are proposed. Chapter 4 will discuss the MLEM algorithm, apply regularization on it, extend its usage to energy-imaging integrated space to achieve EIID, and accelerate its convergence rate using EDID. Chapter 5 will show the recent research in 3D image reconstruction using moving detector system and its real-time capability. Chapter 6 will introduce the development of UMIS, a modern image reconstruction code for radiation detection, and its usages. Finally Chapter 7 will summarize all of the progress that has been achieved, and come up with some open problems that is worth to solve in the future.

CHAPTER II

From Detection to Imaging

2.1 The Shockley-Ramo Theorem

The Shockley-Ramo Theorem provides a way to calculate the induced signal on an electrode due to the motion of charge carriers [35].

The Shockley-Ramo Theorem states that the change of the induced charge ΔQ_L on an electrode L by the movement of charge q from position x_i to position x_f is

$$\Delta Q_L = \int_{x_i}^{x_f} q E_0 dx = -q[\phi_0(x_f) - \phi_0(x_i)], \tag{2.1}$$

where E_0 is the electric field, and ϕ_0 is the weighting potential of a unitless quantity. The equation is under the conditions that electrode L is at unit potential, all other electrodes are at zero potential, and all space charges are removed.

The change of the induced charge ΔQ_L is independent of the actual electric field and the space charges, but only determined by the starting and stopping positions of the charge carriers relative to the electrode L. Therefore if the weighting potential is known, the induced charge can be calculated on any specific electrode easily, even if the electrode configuration is complicated.

2.2 Detector Characteristics

This work is based on large-volume 3D position-sensitive semiconductor detectors, particularly pixelated CZT detectors. Thanks to the excellent performance of CZT detectors developed in the University of Michigan, many of the advanced algorithms can be implemented and tested using real data. Here the basic operation of these detectors is described, and the performance of them is discussed. More details about these detectors and their operation can be found in [36, 37, 38].

2.2.1 Hardware

CZT is a semiconductor material with a bandgap of about 1.6 eV [39]. The bandgap of CZT is larger than that of HPGe detectors, which implies that in CZT the energy required to excite electrons to the conduction band is greater compared to HPGe. So the probability of thermal excitation in CZT is much lower, allowing CZT systems to operate at room temperature.

The energy deposition in the CZT detector excites a number of electrons into the conduction band approximately proportional to the deposited energy, omitting the randomness. If a strong electric field is applied in the crystal, the excited electrons and the positively charged holes will drift and induce signals on the anode and cathode electrodes. It is possible to recover the position and energy information of this energy deposition interaction from the signals.

The CZT crystals used in recent detectors are cuboids of $20 \times 20 \times 15$ mm^3. The square side is covered with a planar cathode electrode, and the opposite side is attached to an array of 11×11 anode electrodes. The anode electrodes form pixels with a pixel pitch of 1.72 mm. To improve charge collection, many detectors have a negatively biased grid electrode between the anode electrodes. Each of the 123 electrodes in one crystal is read out through a single ASIC [40].

A detector system may contain multiple crystals to form a detector array. Fig.

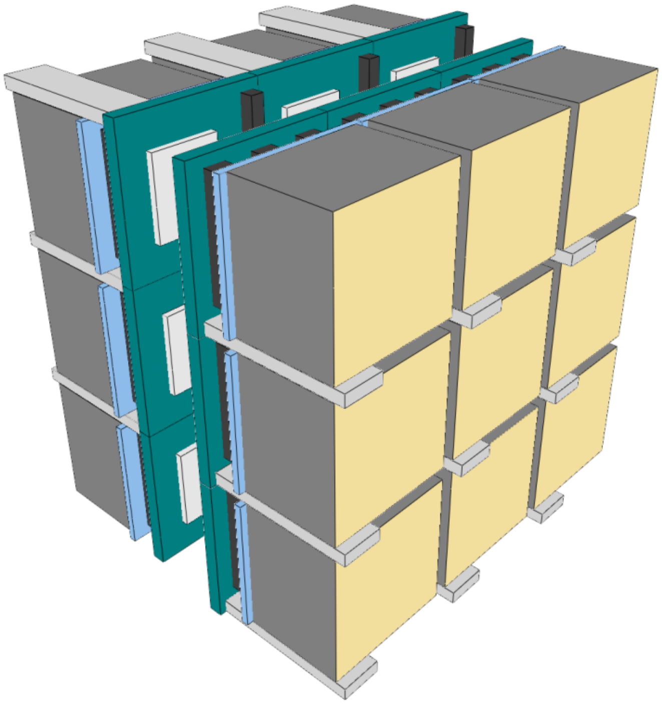


Figure 2.1: Polaris-II: The 18-crystal CZT detector system.

2.1 shows the diagram of Polaris-II, a CZT detector system containing two layers of 3×3 crystals [41]. There are also smaller array with only one crystal [42], and larger array with as many as 64 crystals named Polaris-J, manufactured by H3D, Inc. The modular design is an advantage of CZT detectors, making them adaptive to different measurement scenarios.

2.2.2 Performance

The performance and imaging capability of CZT detectors have improved steadily over the past decades. Around 1% Full Width at Half Maximum (FWHM) energy resolution at 662 keV was consistently demonstrated using $20 \times 20 \times 15$ mm^3 pixelated CZT detectors at University of Michigan. Recently less than 0.5% FWHM energy resolution at 662 keV was achieved [43]. Although there are more requirements on energy resolution from spectroscopy, it is basically sufficient for imaging.

However, for imaging purposes the spatial resolution is more critical than the energy resolution. The conventional analog readout ASIC system provides X/Y axis spatial resolution no better than the pixel pitch 1.72 mm, which is much worse than the 0.5 mm Z axis depth resolution. To improve the X/Y axis spatial resolution and make the spatial resolution more isotropic, Zhu [28] proposed sub-pixel position sensing technique, achieved through algorithms that compare the amount of transient charge induced on pixels that neighbor a charge-collecting pixel. Measurements of the charge induced on the non-collecting pixels are made through analysis of digitized preamplifier pulse waveforms using optimized digital signal processing algorithms. The sub-pixel resolution could reach as small as 300 μm at 662 keV, enabling much better angular resolution and imaging capability [43].

2.2.3 Limitations

In our model, the CZT crystal is abstracted as a cuboid with uniform attenuation coefficient and energy resolution. The X/Y axis spatial resolution is given by the sub-pixel resolution, and the Z axis spatial resolution is given by the depth resolution. Any interaction is defined as a tuple of deposited energy, spatial coordinates, and time, converted from the original waveforms. However, the internal limitations of CZT detectors bring model mismatch, potentially adding noise or artifacts to reconstructed images.

2.3 Imaging Model

It is necessary to define the imaging model before solving any imaging problems. Generally speaking, an image is a collection of parameters. However, it is usually impossible to get images directly in most detecting setups. Thus an image must be reconstructed from measurements or observations. The continuous noiseless relationship between the image f and the expected observation $\bar{g}$ can be abstracted as

$$\bar{g}(\boldsymbol{y}) = \int\limits_{\boldsymbol{x}} \mathcal{T}(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{x}) d\boldsymbol{x}, \tag{2.2}$$

where $\boldsymbol{x}$ is the image position vector, $\boldsymbol{y}$ is the observation position vector, and $\mathcal{T}$ is the system response function. If the image is divided into J bins and the observation is classified as I event types, a discrete representation of this relationship is

$$\bar{g}[i] = \sum_{j=1}^{J} t[i,j] f[j], i = 1, 2, ..., I, \tag{2.3}$$

where $f[j]$ is the value of the j-th image pixel, $\bar{g}[i]$ is the expected observation of the i-th event type, and $t[i,j]$ is the probability that a photon from the j-th image pixel is observed as the i-th event type. Another way to represent the imaging model is the matrix-form

$$\bar{\boldsymbol{g}} = \boldsymbol{T}\boldsymbol{f}, \tag{2.4}$$

where $\boldsymbol{f}$ is the length-J image vector with elements $f_j = f[j]$, $\bar{\boldsymbol{g}}$ is the length-I expected observation vector with elements $\bar{g}_i = \bar{g}[i]$, and $\boldsymbol{T}$ is the $I \times J$ system matrix with elements $t_{ij} = t[i,j]$. That is

$$\begin{bmatrix} \bar{g}_1 \\ \bar{g}_2 \\ \vdots \\ \bar{g}_I \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1J} \\ t_{21} & t_{22} & \dots & t_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ t_{I1} & t_{I2} & \dots & t_{IJ} \end{bmatrix} \cdot \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_J \end{bmatrix}. \tag{2.5}$$

2.3.1 Source Distribution and Imaging Space

The source distribution in a general continuous imaging space $\mathcal{X}$ is the expected density of photons emitted as a function of position $\boldsymbol{x}$ during the measurement time

$$f(\boldsymbol{x}) = \lim_{a(\boldsymbol{x}) \to 0} \frac{\bar{N}(\boldsymbol{x})}{a(\boldsymbol{x})}, \tag{2.6}$$

where $\boldsymbol{x} \in \mathcal{X}$, $a(\boldsymbol{x})$ is an convex area containing position $\boldsymbol{x}$, and $\bar{N}(\boldsymbol{x})$ is the expected number of photons emitted from area $a(\boldsymbol{x})$.

In real computation, it is convenient to use a discrete or pixelated imaging space. Denote $\mathcal{K} = \{1, 2, ..., K\}$ as the indices of pixels. The sampled density is

$$f[k] = f(\boldsymbol{x}_k), \tag{2.7}$$

where $k \in \mathcal{K}$, and $\boldsymbol{x}_k$ is the center of the k-th pixel. It is quite common to write the discrete density as vector form $\boldsymbol{f}$.

Another option is to use the average density

$$f'[k] = \frac{1}{|a_k|} \int\limits_{\boldsymbol{x} \in a_k} f(\boldsymbol{x}) d\boldsymbol{x}, \tag{2.8}$$

where $k \in \mathcal{K}$, and a_k is the area of the k-th pixel. If the pixel size is sufficiently small, the density within the pixel is almost invariant, and the sampled density is equivalent to the average density.

In some cases the total intensity $h[k]$ within a pixel is wanted,

$$h[k] = \int\limits_{\boldsymbol{x} \in a_k} f(\boldsymbol{x}) d\boldsymbol{x} = |a_k| f'[k] \approx |a_k| f[k], \tag{2.9}$$

where $k \in \mathcal{K}$, and $f'[k]$ and $f[k]$ are defined in Eq. 2.8 and Eq. 2.7 respectively.

In different imaging scenarios, different imaging spaces are needed. With far-field assumption, a 2D spherical space is usually preferred because the detector size is negligible compared to the distance between detector and source; with near-field assumption, a 2D Cartesian space is usually preferred because the detector size should be considered in the system model. If the parallax of the detector is sufficient, the source distribution can be reconstructed in a 3D Cartesian space. It is also sometimes useful to reconstruct images in a energy-imaging integrated space, if both full-energy

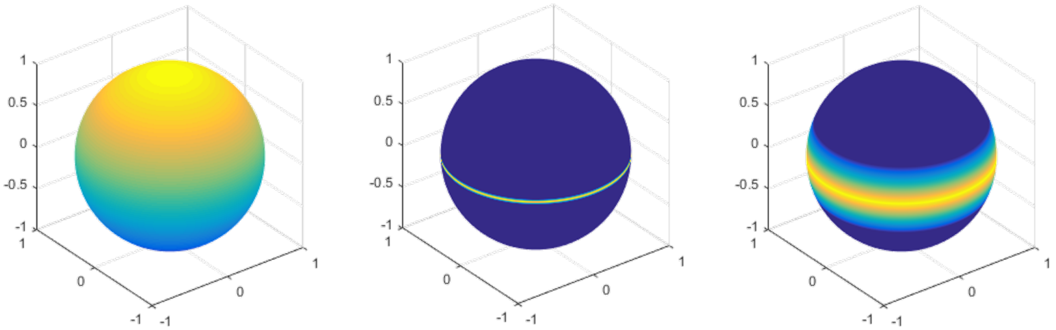


Figure 2.2: Left: An instance of the continuous space. Middle: A theoretical Compton ring in a continuous space, which is similar to a delta function. Right: A blurred Compton ring in a continuous space, which is commonly used in real reconstructions.

deposition and partial-energy deposition are considered in the system model.

2.3.1.1 2D Spherical Space

When the detector size is significantly smaller than the distance between detector and source, far-field assumption is valid. In this scenario, the detector size is assumed infinitely small. Hence the continuous imaging space is a 4π spherical space Ω surrounding the detector, and the source distribution is a function of incident direction $\boldsymbol{\omega} \in \Omega$. It is more convenient to use polar and azimuthal angles to represent incident direction

$$f(\boldsymbol{\omega}) = f(\theta, \phi), \tag{2.10}$$

where $\theta \in [0, \pi]$ is the polar angle, and $\phi \in [0, 2\pi]$ is the azimuthal angle, as shown in Fig. 2.2.

The discrete imaging space $\mathcal{K} = \{1, 2, ..., K\}$ is the indices of pixels, where the k-th pixel has solid angle a_k. If the grid is orthogonal,

$$f[k] = f[i, j] = f(\theta_i, \phi_j), \tag{2.11}$$

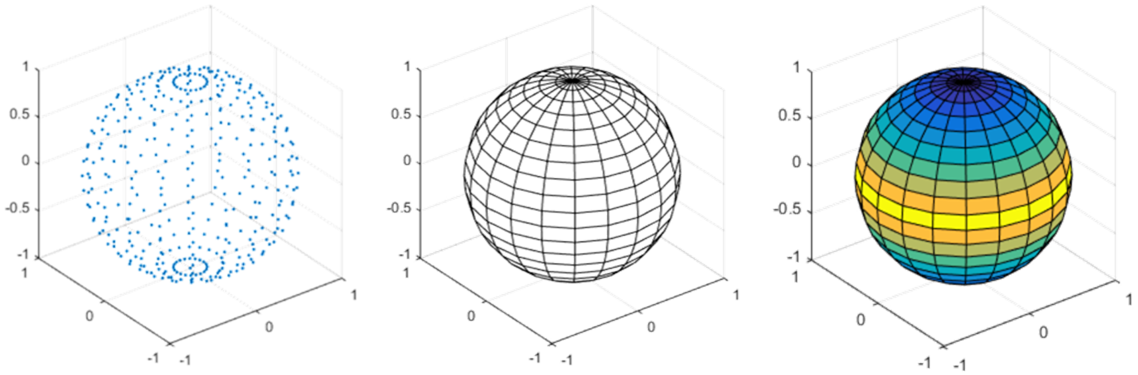


Figure 2.3: Left: An instance of a discrete space. Middle: The area of each pixel. Right: The intensity sensitivity for an isotropic detector in a discrete space.

where $k \in \mathcal{K}$, i and j are the indices of the pixel in polar direction and azimuthal direction, as shown in Fig. 2.3.

The total intensity within a pixel is defined as

$$h[k] = \int\limits_{a_k} f(\boldsymbol{\omega}) d\boldsymbol{\omega} \approx |a_k| f[k], \tag{2.12}$$

where a_k is the solid angle of the k-th pixel. If the grid is orthogonal and the pixel area is very small, the solid angle can be approximately calculated as

$$h[i,j] \approx f(\theta_i, \phi_j) \sin\theta \Delta\theta \Delta\phi = f[i,j] \sin\theta \Delta\theta \Delta\phi. \tag{2.13}$$

Although orthogonal sampling grid is convenient, its sampling density near the poles is higher than that near the equator. However, in many cases the sampling points are required to be as uniform as possible on the sphere, such that all of the pixels have almost the same area. As shown in Fig. 2.4, the grid is not necessary to be orthogonal, but can be arbitrarily selected and the sampling points are quasi-uniformly distributed on the sphere, providing almost the same pixel area.

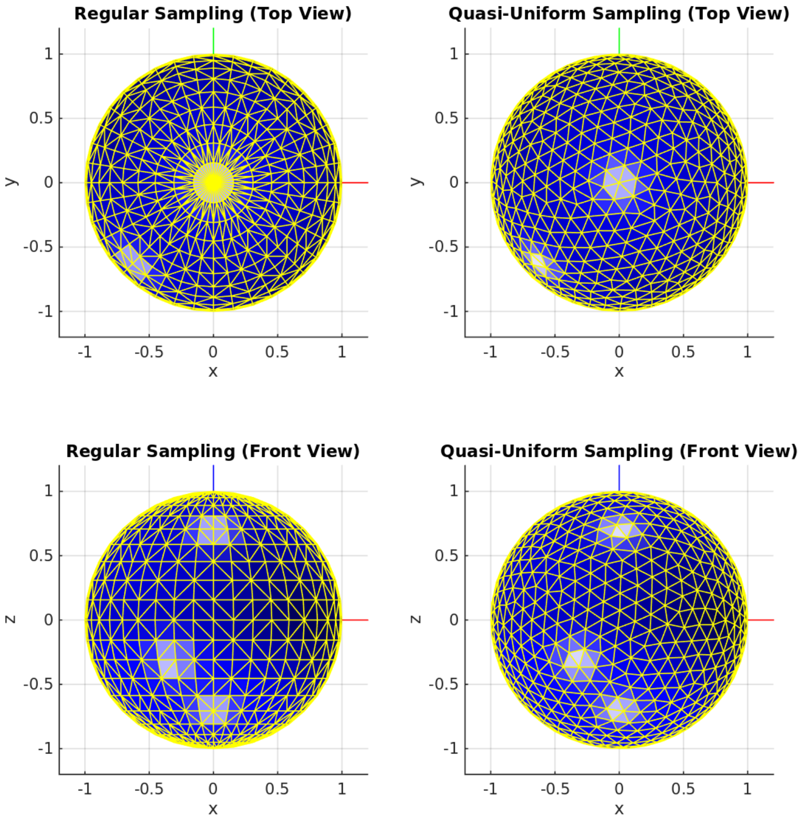


Figure 2.4: Left: Regular sampling, where poles have denser and smaller pixels than the equator. Right: Quasi-uniform sampling, where the pixel areas are approximately the same everywhere.

2.3.1.2 2D Cartesian Space

If the source is close to the detector and the distance between detector and source is comparable with the detector size, far-field assumption is invalid. Instead near-field assumption should be used in this scenario.

Define the source plane as a 2D plane $\mathcal{P}$, then the source distribution is a function of the location on the plane

$$f = f(\boldsymbol{x}) = f(x, y), \tag{2.14}$$

where $\boldsymbol{x}$ is the source location vector defined in the plane $\mathcal{P}$, x and y are dimensions of the plane $\mathcal{P}$.

The discrete imaging space $\mathcal{K} = \{1, 2, ..., K\}$ is the indices of pixels, where the k-th pixel has 2D area a_k. If the grid is orthogonal,

$$f[k] = f[i, j] = f(x_i, y_j), \tag{2.15}$$

where $k \in \mathcal{K}$, i and j are the indices of the pixel in x and y dimension.

The total intensity within a pixel is defined as

$$h[k] = \int\limits_{a_k} f(\boldsymbol{x}) d\boldsymbol{x} \approx |a_k| f[k], \tag{2.16}$$

where a_k is the 2D area of the k-th pixel. If the grid is orthogonal,

$$h[i, j] \approx f(x_i, x_j)\Delta x \Delta y = f[i, j]\Delta x \Delta y. \tag{2.17}$$

2.3.1.3 3D Cartesian Space

Similar to the computed tomography, if there is sufficient parallax from detectors, the source distribution can be reconstructed in a 3D Cartesian space. Possible methods to provide parallax include using multiple detectors, moving the detector, and

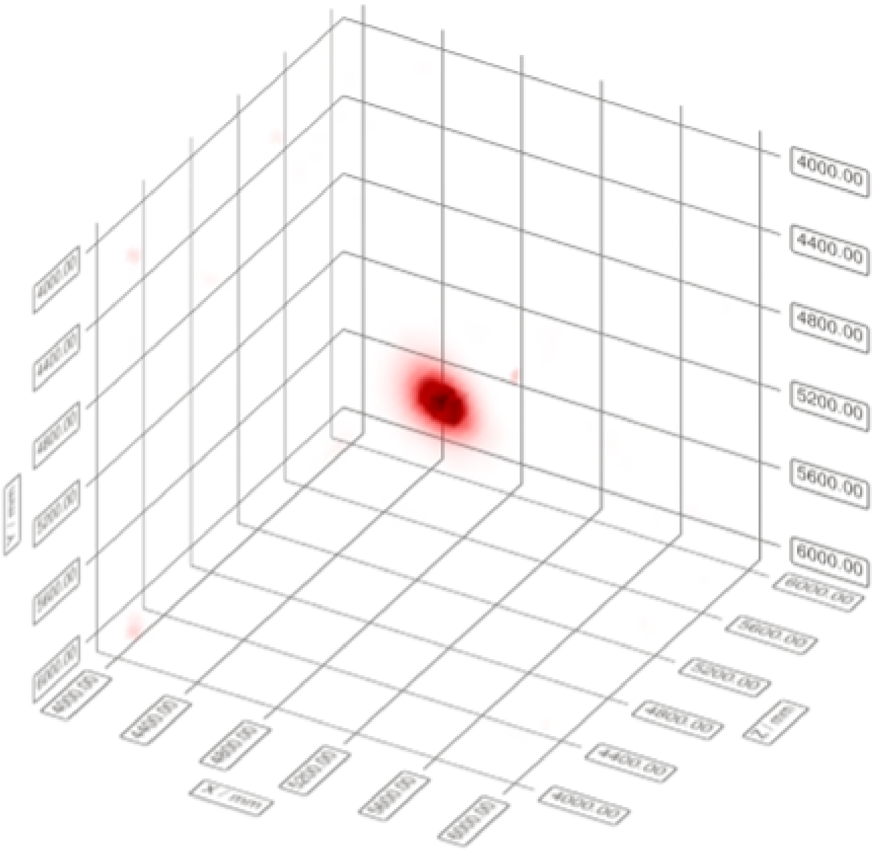


Figure 2.5: A 3D Cartesian space with a source at the center.

putting the detector very close to the source. A typical 3D Cartesian space is shown in Fig. 2.5.

Define the 3D Cartesian space as $\mathcal{P}$, then the source distribution is a function of the location in the space

$$f = f(\boldsymbol{x}) = f(x, y, z), \tag{2.18}$$

where $\boldsymbol{x}$ is the source location vector defined in the plane $\mathcal{P}$, x , y and z are dimensions of the space $\mathcal{P}$.

The discrete imaging space $\mathcal{K} = \{1, 2, ..., K\}$ is the indices of voxels, where the k-th pixel has 3D volume v_k. If the grid is orthogonal,

$$f[k] = f[i_1, i_2, i_3] = f(x_{i_1}, y_{i_2}, z_{i_3}), \tag{2.19}$$

where $k \in \mathcal{K}$, i_1, i_2 and i_3 are the indices of the voxel in x, y, and z dimension.

The total intensity within a voxel is defined as

$$h[k] = \int_{v_k} f(\boldsymbol{x}) d\boldsymbol{x} \approx |v_k| f[k], \tag{2.20}$$

where v_k is the 3D volume of the k-th voxel. The approximation is valid if $f(\boldsymbol{x})$ does not change too much within v_k, which is usually true if $|v_k|$ is very small. If the grid is orthogonal,

$$h[j_1, j_2, j_3] \approx f(x_{j_1}, y_{j_2}, z_{j_3}) \Delta x \Delta y \Delta z = f[j_1, j_2, j_3] \Delta x \Delta y \Delta z. \tag{2.21}$$

2.3.1.4 Energy-Imaging Integrated Space

In previous image spaces, there is no energy dimension. However, sometimes the reconstruction must consider both full-energy deposition and partial-energy deposition to extract most information from photon interactions. In this scenario an energy-imaging integrated space should be used, where the energy domain and the imaging domain are naturally orthogonal.

Define the energy domain as $\mathcal{E}$ and the imaging domain as $\mathcal{S}$. The continuous integrated domain is

$$\mathcal{E} \times \mathcal{S}, \tag{2.22}$$

and the source distribution is

$$f = f(e, \boldsymbol{x}), \tag{2.23}$$

where $e \in \mathcal{E}$ is the incident photon energy scalar, and $\boldsymbol{x} \in \mathcal{S}$ is the source location vector.

If the energy domain is discretized as energy bins $\mathcal{J} = 1, 2, ..., J$ and the imaging domain is discretized as imaging bins $\mathcal{K} = 1, 2, ..., K$, the discrete integrated domain

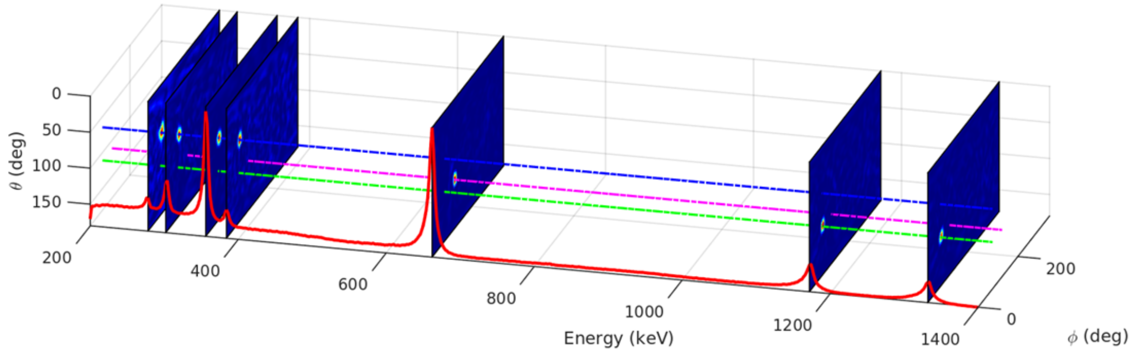


Figure 2.6: An energy-imaging integrated space with three point sources. Red curve shows the total spectrum.

is

$$\mathcal{J} \times \mathcal{K}, \tag{2.24}$$

and the source distribution as

$$f[j,k] = f(e_j, \boldsymbol{x}_k), \tag{2.25}$$

where $j \in \mathcal{J}$, and $k \in \mathcal{K}$.

With the energy-imaging integrated space, one can combine the spectral information and the spacial information of a photon, and do reconstruction generally. It is common to show the integrated space as a 3D space with the energy as the third dimension. As shown in Fig. 2.6, three point sources are distributed in the energy-imaging integrated space, where the imaging domain is a 4π spherical space. The green, pink, and blue dotted lines traversing the energy domain traces the spatial location of Co-60, Cs-137 and Ba-133, respectively. The slices in different energy bins show different spacial distributions.

2.3.2 Bin-Mode and List-Mode Observation Data

Using 3D position-sensitive detectors, the event reconstruction method generates the observation for each event from observed detector electrode signals. Each observed event is a collection of photon-mass interactions, where each interaction contains information such as 3D position and deposited energy. The observation g in the continuous event space $\mathcal{Y}$ is defined as

$$g = g(\boldsymbol{y}), \tag{2.26}$$

where $\boldsymbol{y} \in \mathcal{Y}$ is the event vector. The observation indicates the frequency of event type $\boldsymbol{y}$. If the event space is discretized as $\mathcal{I} = \{1, 2, .., I\}$,

$$g = g[i], \tag{2.27}$$

where $i \in \mathcal{I}$ is the index of event. The vector form of the discrete observations is $\boldsymbol{g}$.

Since the measurement time is limited, it is possible that only a fraction of event types are detected. If the number of detected events is much larger than the number of possible event types, bin-mode observation data is preferred; if the the number of detected events is much smaller than the number of possible event types, list-mode observation data is preferred.

2.3.2.1 Bin-Mode Observation Data

Theoretically the observation data is represented as a vector $\boldsymbol{g}$, which is called bin-mode data. Each element $g[i]$ indicates the frequency of the i-th event type. This form is useful in mathematical derivations and when the number of detected events is much larger than the number of possible event types.

2.3.2.2 List-Mode Observation Data

List-mode data is also known as sparse data. If the number of events is much smaller than the number of possible event types, it is not efficient or even not possible to record the frequency of all event types. Instead, individual events are usually recorded as a list and sorted by time. The event types not appearing in the list have frequency 0.

2.3.3 Statistics

To estimate the image parameters, statistics of the measurement must be considered because the number of photons detected is limited. From counting statistics it is known that the number of photons emitted from the j-th pixel and recorded as the i-th event type is a Poisson random variable. Since the photons emitted from different pixels are mutually independent, the total number of photons recorded as the i-th event is also a Poisson random variable.

The discrete linear imaging model without noise is

$$\bar{g}[i] = \sum_{j=1}^{J} t[i,j] f[j], \tag{2.28}$$

where $f[j]$ is the value of the j-th image pixel, $\bar{g}[i]$ is the expected observation of the i-th event type, and $t[i,j]$ is the probability that a photon from the j-th image pixel is observed as the i-th event type. There is also a matrix form

$$\bar{\boldsymbol{g}} = \boldsymbol{T}\boldsymbol{f}, \tag{2.29}$$

where $\boldsymbol{f}$ is the length-J image vector with elements $f_j = f[j]$, $\bar{\boldsymbol{g}}$ is the length-I expected observation vector with elements $\bar{g}_i = \bar{g}[i]$, and $\boldsymbol{T}$ is the $I \times J$ system matrix with elements $t_{ij} = t[i,j]$.

From Poisson statistics there is

$$g[i] \sim \text{Poisson}(\lambda = \bar{g}[i]), \tag{2.30}$$

and the probability to observe $g[i]$ given image $\boldsymbol{f}$ is

$$\mathbb{P}(g[i]|\boldsymbol{f}) = \frac{\bar{g}[i]^{g[i]} e^{-\bar{g}[i]}}{g[i]!}. \tag{2.31}$$

Assume $g[1], g[2], ..., g[I]$ are mutually independent random variables, which is true for most detector systems when deadtime is negligible, the probability to observe $\boldsymbol{g}$ given image $\boldsymbol{f}$ is

$$\mathbb{P}(\boldsymbol{g}|\boldsymbol{f}) = \prod_{i=1}^{I} \mathbb{P}(g[i]|\boldsymbol{f}) = \prod_{i=1}^{I} \frac{\bar{g}[i]^{g[i]} e^{-\bar{g}[i]}}{g[i]!}, \tag{2.32}$$

which is also called likelihood function.

Because the likelihood function is positive definite, the logarithmic likelihood function of $\boldsymbol{g}$ given image $\boldsymbol{f}$ can be calculated as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{g}|\boldsymbol{f}) &= \ln \mathbb{P}(\boldsymbol{g}|\boldsymbol{f}) \\
&= \sum_{i=1}^{I} \Big(-\bar{g}[i] + g[i] \ln \bar{g}[i] - \ln g[i]! \Big) \\
&= \sum_{i=1}^{I} \Big(-\sum_{j=1}^{J} t[i,j] f[j] + g[i] \ln \sum_{j=1}^{J} t[i,j] f[j] - \ln g[i]! \Big),
\end{aligned} \tag{2.33}$$

and one can get the maximum-likelihood estimation of $\boldsymbol{g}$ by maximizing the logarithmic likelihood function, which will be discussed in Chapter 4.

With a large number of counts, the Poisson statistics can be approximated by the Gaussian statistics,

$$g[i] \sim \text{Gaussian}(\mu = \bar{g}[i], \sigma^2 = \bar{g}[i]), \tag{2.34}$$

and the probability density to observe $g[i]$ given image $\boldsymbol{f}$ is

$$p(g[i]|\boldsymbol{f}) = \frac{1}{\sqrt{2\pi\bar{g}[i]}} e^{-\frac{(x-\bar{g}[i])^2}{2\bar{g}[i]}}. \tag{2.35}$$

Further analysis and applications of the Gaussian approximation will be in Chapter 3.

2.4 System Model

To build the system matrix, the physics and detector geometry have been analyzed, and a system model has been generated. Usually the electronic setup is ignored, and the detector output is treated as a set of interaction energies and locations with uncertainty. Xu [42] and Wahl [41] described the system model by complicated equations. Here another view of the problem is proposed, building the probability distribution from shallow to deep.

2.4.1 Ideal Point Detector

At first an ideal point detector is assumed. An ideal point detector has perfect spatial and energy resolution for interactions, and infinity small size such that the attenuation is ignored. For a full-energy deposition event with two interactions, denote the first interaction as $(\boldsymbol{x}_1, E_1)$, the second interaction as $(\boldsymbol{x}_2, E_2)$. Only the Compton scattering process for this event need to be considered. In a 4π spherical space,

$$f(\boldsymbol{\omega}) = \begin{cases} \delta, \text{if } \boldsymbol{\omega} \in \mathcal{W}, \\ \\ 0, \text{otherwise}, \end{cases} \tag{2.36}$$

where ω is the solid angle, and $\mathcal{W}$ is defined as

$$\mathcal{W} = \left\{ \boldsymbol{\omega} : \frac{\boldsymbol{\omega}(\boldsymbol{x}_1 - \boldsymbol{x}_2)}{|\boldsymbol{\omega}||\boldsymbol{x}_1 - \boldsymbol{x}_2|} = 1 - \frac{m_e c^2 E_1}{E_0 E_2} \right\} \tag{2.37}$$

which means that any $\omega \in \mathcal{W}$ satisfies Compton scattering formula. To keep the total conditional probability as 1,

$$\int\limits_{\boldsymbol{\omega}\in 4\pi} f(\boldsymbol{\omega})d\boldsymbol{\omega} = \int\limits_{\boldsymbol{\omega}\in \mathcal{W}} \delta d\boldsymbol{\omega} = 1, \tag{2.38}$$

which defines δ like a delta function.

2.4.2 Uncertainties and Gaussian Blur

However, a real detector has spatial uncertainty and energy uncertainty, which increases the complexity of the system model. The spatial uncertainty is not isotropic, because the uncertainty parallel to the anode plane is decided by the pixel or the sub-pixel resolution, and the uncertainty perpendicular to the anode plane is decided by the depth resolution. The energy uncertainty is a function of deposited energy, which has an influence of the scattering angle calculation. Other uncertainties include Doppler broadening and coherent scattering. The accurate modeling of these uncertainties could be very computationally intensive and unpractical.

Instead, a simplified way to approximate the uncertainties is adding a Gaussian blurring kernel to the Compton ring. Define the standard deviation of the blur as σ

$$\sigma = \sigma(\boldsymbol{x}_1, E_1, \boldsymbol{x}_2, E_2, ...), \tag{2.39}$$

which is a function of interactions. People usually fit the function empirically. The PSF for a Gaussian blurring kernel is

$$b(x, \sigma) = Ce^{-\frac{x^2}{2\sigma^2}}, \tag{2.40}$$

where C is the normalization factor such that the blurred image has a total intensity of 1.

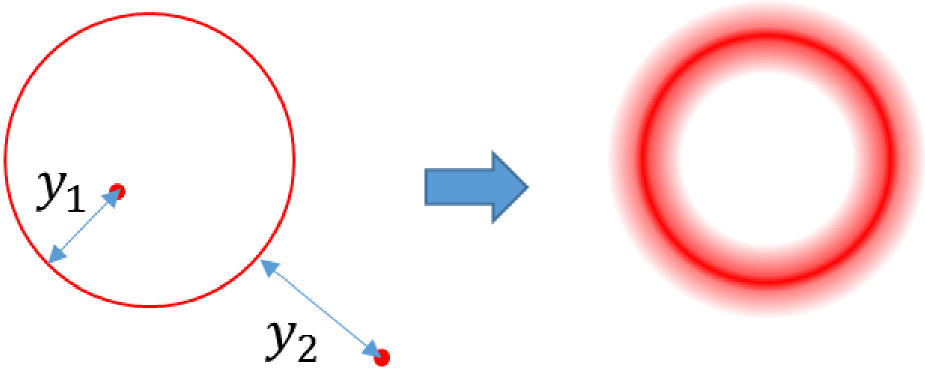


Figure 2.7: A Compton ring with Gaussian radius blur. y_1 and y_2 indicate the shortest distances between the two points and the ring.

2.4.2.1 Gaussian Radius Blur

In practice the Gaussian blurring kernel is calculated using a Gaussian radius blur. For any point $\boldsymbol{x}$ in the image, the distance $y(\boldsymbol{x})$ to the ring is defined as the shortest distance to any points on the ring, as shown in Fig. 2.7. The blurred ring is

$$f(\boldsymbol{x}) = Ce^{-\frac{y(\boldsymbol{x})^2}{2\sigma_r^2}}, \tag{2.41}$$

where C is the normalization factor, and σ_r is the standard deviation of the radius blur. It can be proved that if the radius of the ring is much greater than σ, the Gaussian radius blur approximates a Gaussian blurring kernel.

2.4.3 Geometry and Attenuation

A real detector has finite geometry and attenuation, affecting the probability of incident direction. According to Beer-Lambert law, the probability that a photon with energy E pierces a material with thickness x is

$$p = e^{-\mu(E)x}, \tag{2.42}$$

where $\mu(E)$ is the attenuation coefficient of the material under photon energy E. If there are multiple materials on the path of the photon, the probability becomes

$$p = \prod_n e^{-\mu_n(E)x_n}, \tag{2.43}$$

where n is the index of the material, $\mu_n(E)$ is the attenuation coefficient of the material n under photon energy E, and x_n is the total length of the material n on the path of the photon.

It is possible to include the attenuation information in the probability and model the system more accurately. However, this requires huge amount of computing resources to do ray-tracing, and make the system model too complicated to apply several fast algorithms. In Compton imaging, most information comes from the Compton rings, and adding attenuation can not significant affect the reconstruction results.

2.4.4 Sequencing

Since a Compton event has multiple interactions, it is critical to recover the sequence of interactions, otherwise the Compton ring is not properly back-projected. However, the time delay between interactions is on the order of picoseconds due to the small size of current CZT detectors and the ultra fast photon speed, so there could be ambiguities of the sequence of interactions.

In most probability-based reconstruction algorithms, such as MLEM, any possible sequences are considered in the system model, and the rings are weighted by the probability. In some direct reconstruction algorithms, such as SBP and FBP, sequencing error is allowed and considered in the PSF, so sequencing is done before image reconstruction. One can refer to [44, 25, 33, 41] for the existing sequencing methods. But for high energy sequencing, there are still many problems to solve.

CHAPTER III

Direct Reconstruction Algorithms

Since measured data from a detector does not directly represent the image, it is necessary to apply reconstruction algorithms. This chapter will discuss two direct reconstruction algorithms: Simple Back-Projection (SBP) and Filtered Back-Projection (FBP). These two algorithms have close-form solutions and can be applied on an event-by-event basis. However, there are problems in these two algorithms and significant improvement is required to make them practical [1].

3.1 Simple Back-Projection

Among all reconstruction algorithms, SBP may be the simplest, yet still widely used in research and industry due to its speed and robustness.

3.1.1 Imaging Model

The general discrete imaging model without noise in matrix-form is

$$\bar{\boldsymbol{g}} = \boldsymbol{T}\boldsymbol{f}, \tag{3.1}$$

---

[1] This chapter is based in part on [54].

where $\boldsymbol{f}$ is the length-J image vector, $\bar{\boldsymbol{g}}$ is the length-I expected observation vector, and $\boldsymbol{T}$ is the $I \times J$ system matrix with each row corresponding to a back-projected Compton cone. Based on this model, given actual observation vector $\boldsymbol{g}$, the SBP solution is

$$\hat{\boldsymbol{f}}_{\mathrm{SBP}} = \boldsymbol{T}'\boldsymbol{g}, \tag{3.2}$$

where $\boldsymbol{T}'$ is the conjugate transpose of $\boldsymbol{T}$. The expectation using this estimation is

$$\bar{\boldsymbol{f}}_{\mathrm{SBP}} = \boldsymbol{T}'\bar{\boldsymbol{g}} = \boldsymbol{T}'\boldsymbol{T}\boldsymbol{f} = \boldsymbol{B}\boldsymbol{f}, \tag{3.3}$$

where $\boldsymbol{B} \triangleq \boldsymbol{T}'\boldsymbol{T}$ is defined as the $J \times J$ point spread matrix. Since $\boldsymbol{T}'\boldsymbol{T}$ is rarely an identity matrix, $\bar{\boldsymbol{f}}_{\mathrm{SBP}}$ provides a biased estimation of the original image. Typically it is a very blurred version.

The SBP is mostly used in a 4π spherical space or a 2D plane space, and generates reasonable 2D images. Although in principle SBP can be used in a 3D space, the SBP images in 3D are usually too blurry to provide valid information. Thus, more advanced algorithms such as MLEM are required in 3D spaces.

3.1.2 Analysis of SBP

The SBP in a 4π spherical space is analyzed as an example. The SBP equation 3.2 can be rewritten as

$$\hat{\boldsymbol{f}}_{\mathrm{SBP}} = \boldsymbol{T}'\boldsymbol{g} = \sum_{i=1}^{I} g_i \boldsymbol{T}'_{i:}, \tag{3.4}$$

where g_i is the i-th element of the observation vector $\boldsymbol{g}$ indicating the number of observed events of type i, and $\boldsymbol{T}_{i:}$ is the i-th row of the system matrix $\boldsymbol{T}$.

The i-th row of the system matrix is the conditional probability distribution given the event type i. In Compton imaging with ideal point detectors of infinitely small size, the rows of system matrix are just Compton rings. So the SBP can be viewed as the sum of Compton rings for all the observed events, as shown in Fig. 3.1. As

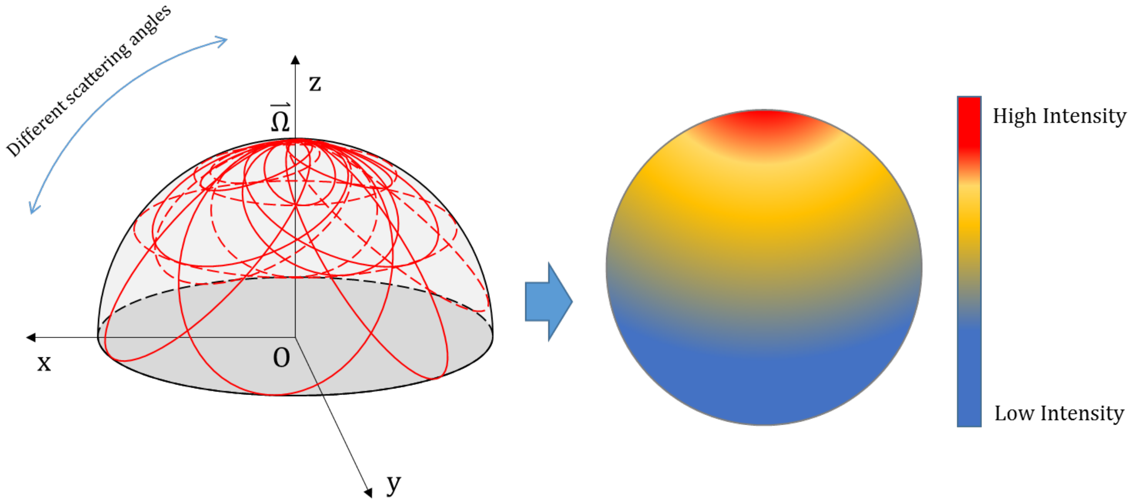


Figure 3.1: SBP can be viewed as the accumulation of Compton rings, inducing a hotspot for each point source.

more and more Compton rings are accumulated, a hotspot will present if the original image is a point source.

Theoretically the number of Compton rings is infinite, and the Point Spread Function (PSF) in SBP comes with ideal point detectors. The isotropic PSF is defined as $h_E(\cos\omega)$ which is related to the incident photon energy E. $h_E(\cos\omega)$ is decided by the scattering angle distribution $D_E(\theta)$,

$$D_E(\theta) = K(E, \theta)\sin\theta, \tag{3.5}$$

where $K(E, \theta)$ is the Klein-Nishina formula [45]. Fig. 3.2 shows some theoretical $D_E(\theta)$ with different energies.

The density of the Compton ring is defined as $\xi_E(\theta)$. As shown in Fig. 3.3, assuming there is a point source $\vec{\Omega}$ at the zenith of the sphere, the expected SBP image is just the PSF. Let's pick a sampling point at $\vec{\Omega}'$, and define the angle between $\vec{\Omega}$ and $\vec{\Omega}'$ as ω. All sampling points with angle ω build a sampling ring. Then the

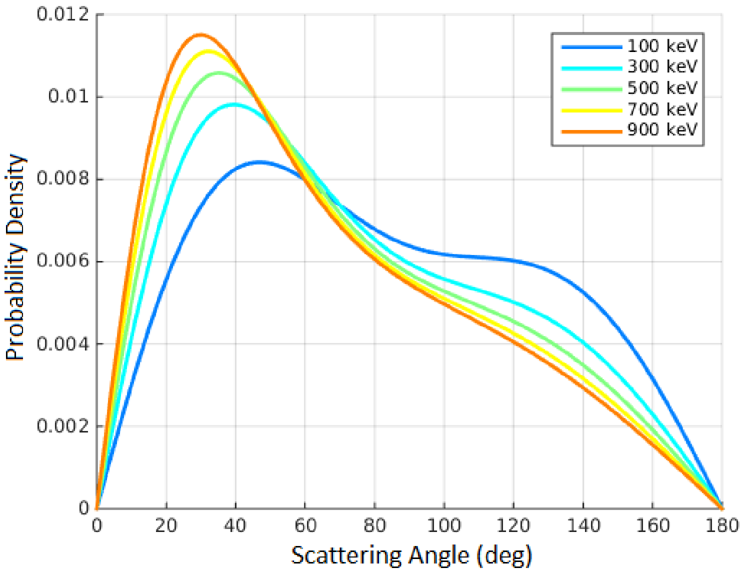


Figure 3.2: Theoretical $D_E(\theta)$ for different energies. High energy photons have a smaller probability of large angle scatter than low energy photons.

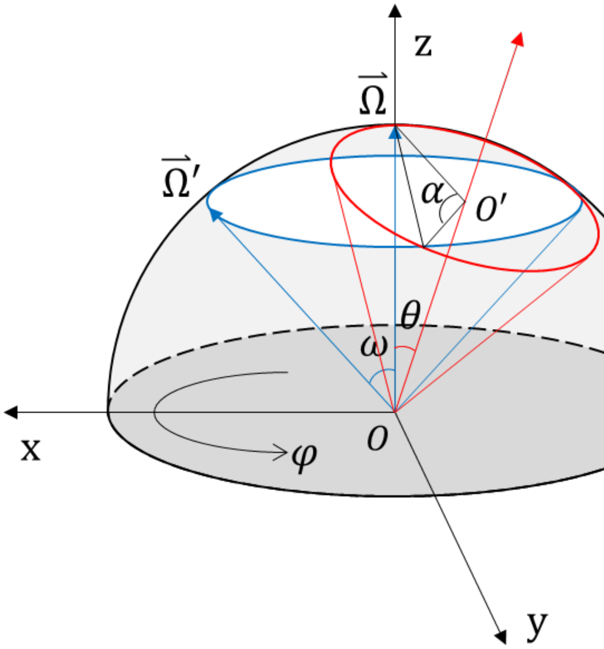


Figure 3.3: The geometry between a sampling (blue) ring and a scattering (red) ring with the point source at the zenith of the sphere.

PSF can be calculated with the equation

$$h_E(\cos\omega)d\omega \cdot 2\pi\sin\omega = \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \cdot 2D_E(\theta)\xi_E(\theta)d\alpha, \tag{3.6}$$

where the left side means the PSF at angle ω times the width of the blue ring $d\omega$ times the length of the blue ring $2\pi\sin\omega$; the right side means the red Compton ring with scattering angle θ intersects the blue ring twice and each intersection point contribute $D_E(\theta)\xi_E(\theta)d\alpha$ to the blue ring, and the integral is done over all possible θ and ϕ.

The black triangle indicates the relationship among α, ω, and θ

$$\cos\alpha = 1 - \frac{2\sin^2 0.5\omega}{\sin^2\theta}, \tag{3.7}$$

which leads to

$$\frac{d\alpha}{d\omega} = \frac{\cos 0.5\omega}{\sqrt{\cos^2 0.5\omega - \cos^2\theta}}. \tag{3.8}$$

So there is

$$h_E(\cos\omega) = \frac{1}{\sin 0.5\omega}\int_{0.5\omega}^{\pi-0.5\omega} \frac{D_E(\theta)\xi_E(\theta)}{\sqrt{\cos^2 0.5\omega - \cos^2\theta}}d\theta, \tag{3.9}$$

where the range of the integral is limited to $[0.5\omega, \pi - 0.5\omega]$ because the scattering ring must intersect with the sampling ring.

The density $\xi_E(\theta)$ should normalize the Compton ring such that the total intensity is 1. However, since SBP is already a biased estimation, arbitrary density values are allowed to enhance the image. There are several logical choices for $\xi_E(\theta)$ (but unlimited to them).

1. $\xi_E(\theta) = \frac{1}{\sin\theta}$: This is the standard choice assuming each Compton ring carries same information.

2. $\xi_E(\theta) = 1$: This is a widely used choice assuming each pixel on the Compton

ring carries same information.

3. $\xi_E(\theta) = \frac{\sin\theta}{D_E(\theta)}$: This choice makes the scattering angle distribution equivalent to a sin function, such that an analytical PSF is available:

$$\begin{aligned} h_E(\cos\omega) &= \frac{1}{\sin 0.5\omega} \int\limits_{0.5\omega}^{\pi-0.5\omega} \frac{\sin\theta}{\sqrt{\cos^2 0.5\omega - \cos^2\theta}} d\theta \\ &= \frac{\pi}{\sin 0.5\omega}. \end{aligned} \tag{3.10}$$

Since $\sin 0.5\omega = 0.5|\vec{\Omega} - \vec{\Omega}'|$,

$$h_E(\cos\omega) = \frac{2\pi}{|\vec{\Omega} - \vec{\Omega}'|} = 2\pi \sum_{l=0}^{\infty} P_l(\cos\omega), \tag{3.11}$$

where $P_l(\cos\omega)$ is the Legendre polynomial, and the PSF becomes a Newtonian potential.

In practice, the number of Compton rings is finite, and the Compton rings have limited width or Gaussian blur, so the shape of the PSF is like a peak with random noise, as shown in Fig. 3.4.

3.2 Filtered Back-Projection

SBP suffers from poor resolution. If Gaussian additive noise is assumed in the observation domain, which is approximately true if the number of events is large, then FBP can be used to deconvolve the PSF and improve the resolution of reconstructed images.

3.2.1 Imaging Model

Assume Gaussian additive noise on the SBP image,

$$\hat{\boldsymbol{f}}_{\text{SBP}} = \boldsymbol{T}'\boldsymbol{g} = \boldsymbol{T}'\bar{\boldsymbol{g}} + \boldsymbol{r} = \boldsymbol{T}'\boldsymbol{T}\boldsymbol{f} + \boldsymbol{r} = \boldsymbol{B}\boldsymbol{f} + \boldsymbol{r}, \tag{3.12}$$

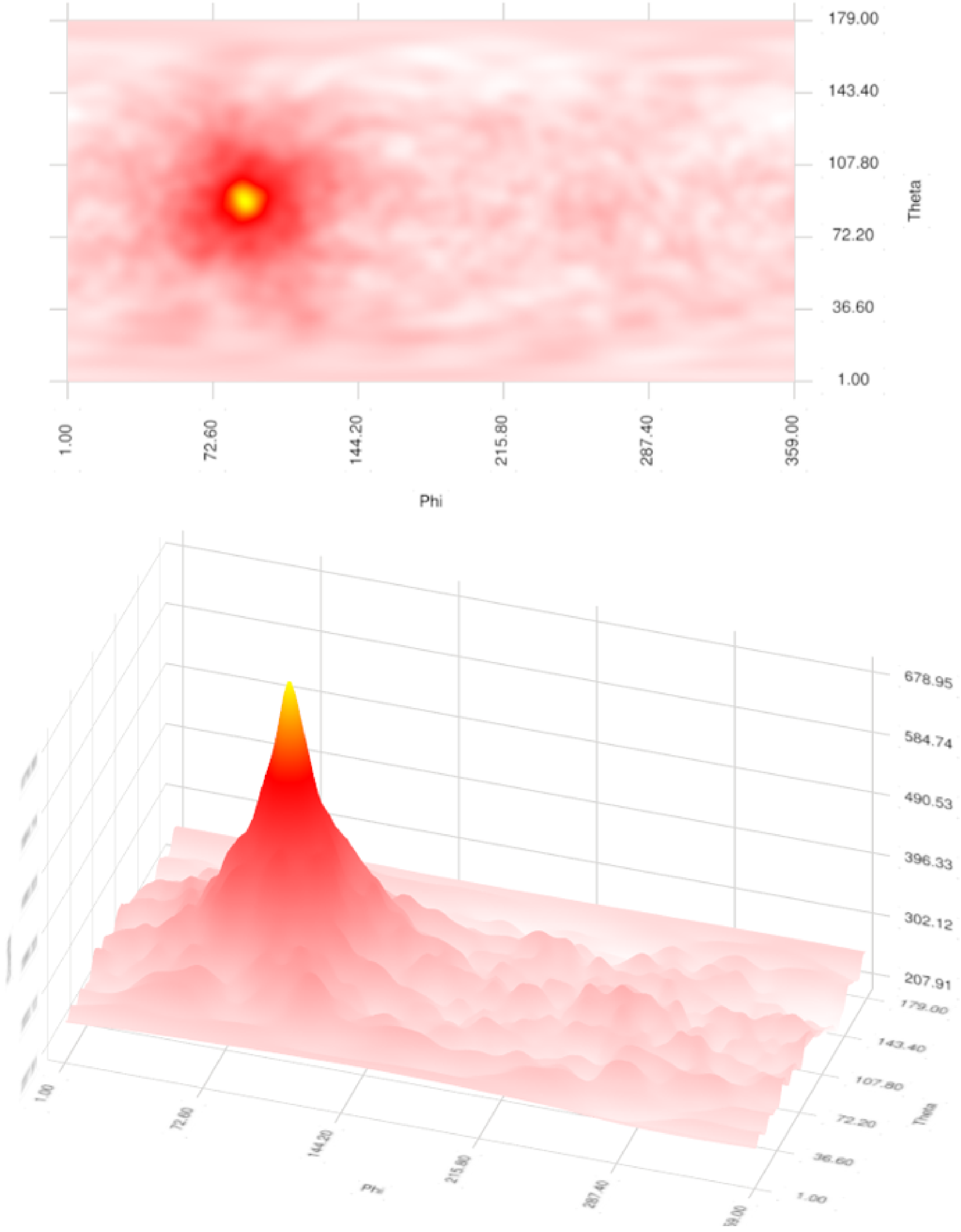


Figure 3.4: Left: SBP image of a point source shown as a heat map. Right: Same SBP image shown as a height map.

where $\boldsymbol{r} \sim \mathcal{N}(\bar{\boldsymbol{r}}, \boldsymbol{K})$ is the Gaussian noise vector with mean $\bar{\boldsymbol{r}}$ and covariance matrix $\boldsymbol{K}$. To simply the analysis, assume $\bar{\boldsymbol{r}} = \boldsymbol{0}$ and $\boldsymbol{K} = \sigma^2 \boldsymbol{I}$, which means that $\boldsymbol{r}$ is white Gaussian noise.

The maximum likelihood estimation of this simplified statistical model is the least squares estimation [46]

$$\hat{\boldsymbol{f}}_{\text{LS}} = \boldsymbol{B}^{+} \hat{\boldsymbol{f}}_{\text{SBP}} = \boldsymbol{B}^{+} \boldsymbol{T}' \boldsymbol{g}, \tag{3.13}$$

where $\boldsymbol{B}^{+}$ is the pseudo-inverse of the point spread matrix $\boldsymbol{B}$. If $\boldsymbol{B}$ is invertible, $\boldsymbol{B}^{+} = \boldsymbol{B}^{-1}$ and the expectation of this estimation is

$$\bar{\boldsymbol{f}}_{\text{LS}} = \boldsymbol{B}^{-1} \boldsymbol{T}' \bar{\boldsymbol{g}} = \boldsymbol{B}^{-1} \boldsymbol{T}' \boldsymbol{T} \boldsymbol{f} = \boldsymbol{B}^{-1} \boldsymbol{B} \boldsymbol{f} = \boldsymbol{f}, \tag{3.14}$$

indicating that $\hat{\boldsymbol{f}}_{\text{LS}}$ is an unbiased estimation of the original image.

Since B is a invertible symmetric matrix, using the spectral theorem, the eigenvector decomposition [47] is

$$\boldsymbol{B} = \boldsymbol{Q}' \boldsymbol{\Gamma} \boldsymbol{Q}, \tag{3.15}$$

where the eigenvalues of $\boldsymbol{B}$ are on the diagonal of $\boldsymbol{\Gamma}$ and orthonormal eigenvectors of $\boldsymbol{B}$ are in the rows of $\boldsymbol{Q}$. The diagonal of $\boldsymbol{\Gamma}$ is also the frequency response, or spectrum, of the point spread matrix B. Applying $\boldsymbol{B}^{-1}$ is equivalent to do filtering in frequency domain, hence this method is named as FBP. So

$$\hat{\boldsymbol{f}}_{\text{FBP}} = \hat{\boldsymbol{f}}_{\text{LS}} = \boldsymbol{B}^{-1} \boldsymbol{T}' \boldsymbol{g} = \boldsymbol{Q}' \boldsymbol{\Gamma}^{-1} \boldsymbol{Q} \boldsymbol{T}' \boldsymbol{g}. \tag{3.16}$$

If the PSF is shift-invariant, the Fourier transform can be used in a 2D plane imaging space, and the SH can be used in a 4π spherical imaging space. Denote the number of Compton events as N, and the number of image pixels as J. Using the Fast Fourier Transform (FFT) or the fast Spherical Harmonics Transform (SHT), $\boldsymbol{Q}$ and $\boldsymbol{Q}'$ multiplications have $O(J^{1.5})$ time complexity. As $\boldsymbol{\Gamma}$ is diagonal, the multiplication of

$\mathbf{\Gamma}^{-1}$ has $O(J)$ time complexity. The multiplication of $\boldsymbol{T}'$, which is the back-projection step, has $O(NJ)$ time complexity. Since usually $N \gg J$, the computation totally has $O(NJ)$ time complexity, which is as fast as the SBP. Because $\hat{\boldsymbol{f}}_{\text{FBP}}$ is a linear transformation of $\boldsymbol{g}$, it is also additive and can be done on an event-by-event basis.

3.2.2 Spherical Harmonics

In a 4π spherical imaging space, eigenvector decomposition using Discrete Spherical Harmonics Transform (DSHT) matrix is a natural choice. Suppose the image is $\boldsymbol{f}$, then its j-th element

$$f_j = f(\boldsymbol{\omega}_j) = f(\theta_j, \phi_j). \tag{3.17}$$

The DSHT matrix is exactly $J \times J$ matrix $\boldsymbol{Q}$. The elements q_{kj} of $\boldsymbol{Q}$ are defined as

$$q_{kj} \triangleq c_j y_{kj}, k, j = 1, 2, ..., J, \tag{3.18}$$

where c_j is the weight of the sampling point f_j, and y_{kj} is defined as

$$y_{kj} \triangleq Y_l^{m*}(\theta_j, \phi_j), \tag{3.19}$$

which are the complex conjugates of the SH functions, $l \in [0, L]$, $m \in [-l, l]$, and $k = l^2 + l + m + 1$ such that there is a bijection between k and the degree-order pair (l, m). Ideally the maximum spherical harmonics degree L should satisfy $(L+1)^2 = K$. The complex conjugates of the SH function are

$$Y_l^{m*}(\theta_j, \phi_j) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} \cdot P_l^m(\cos\theta_j) e^{-im\phi_j}, \tag{3.20}$$

where $P_l^m(\cos\theta_j)$ are the associated Legendre polynomials. There exist several fast methods[48, 49, 50, 51, 52] to compute the DSHT and the inverse DSHT.

3.2.3 Filter Design

Rewrite the Gaussian additive noise model in matrix form

$$\hat{\boldsymbol{f}}_{\text{SBP}} = \boldsymbol{s} = \boldsymbol{B}\boldsymbol{f} + \boldsymbol{r}, \tag{3.21}$$

to function form

$$s = f \circledast b + r, \tag{3.22}$$

where $\circledast$ means circular convolution, s is the SBP image, f is the true image, b is the PSF, and r is zero-mean uncorrelated Gaussian noise, or white Gaussian noise, with variance σ^2. Assume Wide-Sense Stationary (WSS) and jointly WSS between h and f such that the Wiener filter can be applied [53].

The FBP estimation to this model is

$$\hat{f} = s \circledast h \tag{3.23}$$

where $\hat{f}$ is the estimation, and h is the deconvolution filter. Now define some image-frequency pairs

$$\begin{cases} f[j] \longleftrightarrow F(\lambda) \\ b[j] \longleftrightarrow B(\lambda) \\ s[j] \longleftrightarrow S(\lambda) \\ h[j] \longleftrightarrow H(\lambda) \\ \hat{f}[j] \longleftrightarrow \hat{F}(\lambda) \end{cases}, \tag{3.24}$$

where λ is frequency. Obviously

$$\hat{F}(\lambda) = S(\lambda)H(\lambda), \tag{3.25}$$

and the optimum Wiener filter is

$$H(\lambda) = \frac{B^*(\lambda)}{|B(\lambda)|^2 + R(\lambda)}, \ R(\lambda) \triangleq \frac{P_r(\lambda)}{P_f(\lambda)}, \tag{3.26}$$

where P_r and P_f are the power spectrum of r and f respectively, and R is the penalty parameter in frequency domain.

For Compton imaging in a 4π spherical space, the PSF can be simplified as

$$B(\lambda) = N(\lambda)G(\lambda), \tag{3.27}$$

where $N(\lambda)$ is the spherical harmonics kernel of Newtonian potential, $G(\lambda)$ is the spherical harmonics kernel of Gaussian blur, and

$$N(\lambda) = \frac{4\pi}{2\lambda + 1}, \tag{3.28}$$

$$G(\lambda) \longleftrightarrow g(\cos\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\omega^2}{2\sigma^2}}. \tag{3.29}$$

where λ is the spherical harmonics degree, ω is the angle to the center of the PSF, and σ is the standard deviation of the Gaussian blur. One can use empirical PSF calibrated from real measurements.

3.2.4 Reconstructions Using Simulated Data

Fig. 3.5 shows a simulation where four Cs-137 point sources were placed on the cathode side of the detector. The SBP image is too blurry to tell the point sources apart, but the FBP image with much better angular resolution is able to tell the boundary between sources. Compared to the SH coefficients of FBP, the SH coefficients of SBP are dominated by low-frequency components, hence the blurred image. The inverse filter is designed as a Wiener filter, where the medium-frequency region is amplified to improve resolution, and the high-frequency region is suppressed

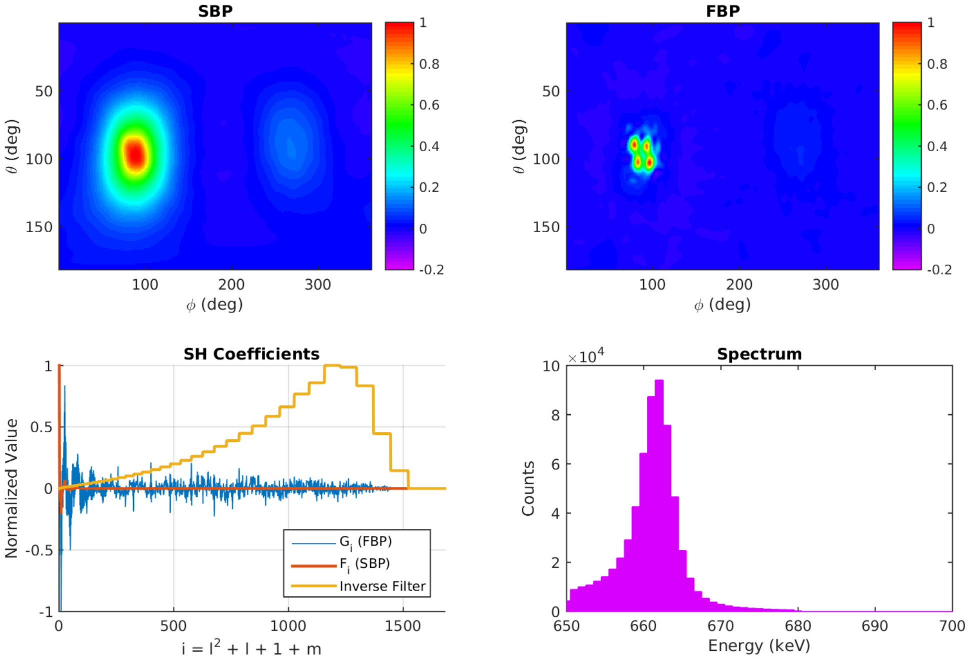


Figure 3.5: SBP and FBP of four Cs-137 point sources separated by 10 degrees, with their SH coefficients and the photopeak spectrum used for imaging. The ghost on the opposite direction of the sources is because of wrong sequenced events.

to denoise.

Fig. 3.6 compares the reconstructions of a single source with different statistics. From left to right the reconstructions use 10^3, 10^4, and 10^5 events. It is obvious that the SBP image with fewer events is noisier. The SH coefficients of SBP are dominated by the low-frequency components, which is indicated by the blurry images.

If the same inverse filter is used independently of measurement statistics, as shown in the second and third rows of Fig. 3.6, the FBP image with 10^3 events has many artifacts, and its SH coefficients have much stronger high-frequency noise compared to the SH coefficients of FBP with 10^5 events. The SH coefficients of FBP with 10^5 events show relatively clear shape in the high-frequency region. To suppress high-frequency noise, a low-pass filter is applied to the inverse filter, which is equivalent to a Wiener filter. The low-pass filter slightly blurs the FBP images but greatly reduces noises, as shown in the fourth and fifth rows of Fig. 3.6.

3.3 Adjusted FBP

Currently 3D position sensitive detectors provide advantages for Compton imaging, because a single monolithic detector serves as both the scattering and absorbing detector for photons from any incident direction simultaneously, providing higher detection efficiency and intrinsically better 4π imaging capability than traditional dual-plane Compton imagers. In addition the digital readout technique of recent systems allows sub-pixel position resolution and higher energy resolution, improving the accuracy of reconstructed images [28]. However, with the improvement of angular resolution, the influence of detector geometry on the PSF becomes considerable, especially in a detector array with a non-uniform spatial arrangement and when the number of events is sufficient.

As shown in Fig. 3.7, a 2×2 detector prototype array [43] has different system responses for different incident photon directions. For a source located on the y-axis, the scattering direction distribution is approximately axisymmetric about the incident direction. Contrastingly, for a source located on the x-axis with the same scattering angle, the photons scattered in the x-z plane have a higher probability of interacting with detectors than the photons scattered in the x-y plane, because the z-dimension of the array is thicker and provides more attenuation than the y-dimension. Therefore, the PSF is anisotropic and varies with incident photon direction. If a fixed PSF is used for FBP, the reconstructed images will have obvious bias because the measured PSF will never match the PSF used in the inverse filter. Fig. 3.8 shows the SBP images of a Cs-137 point source at the side of the 2×2 detector array with both simulated data and real data. It is clear that the PSF is anisotropic.

An intuitive solution would build a PSF for each incident photon direction. How-

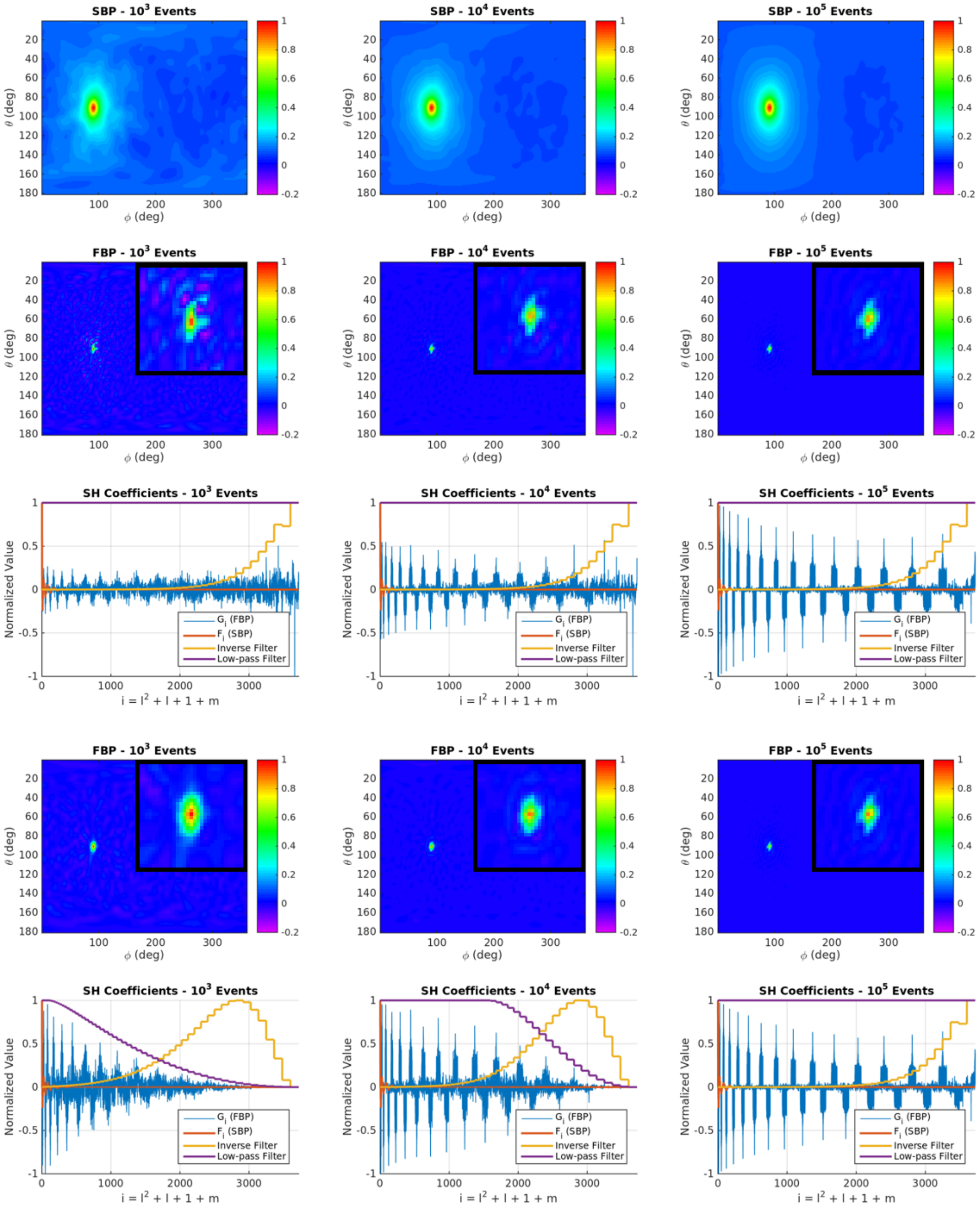


Figure 3.6: Left to right: reconstructions using 10^3, 10^4, and 10^5 events. First row: SBP images. Second row: FBP images without the low-pass filter, where a zoomed-in hotspot is at the top right corner. Third row: SH coefficients without the low-pass filter. Fourth row: FBP images with the low-pass filter, where a zoomed-in hotspot is at the top right corner. Fifth row: SH coefficients with the low-pass filter.

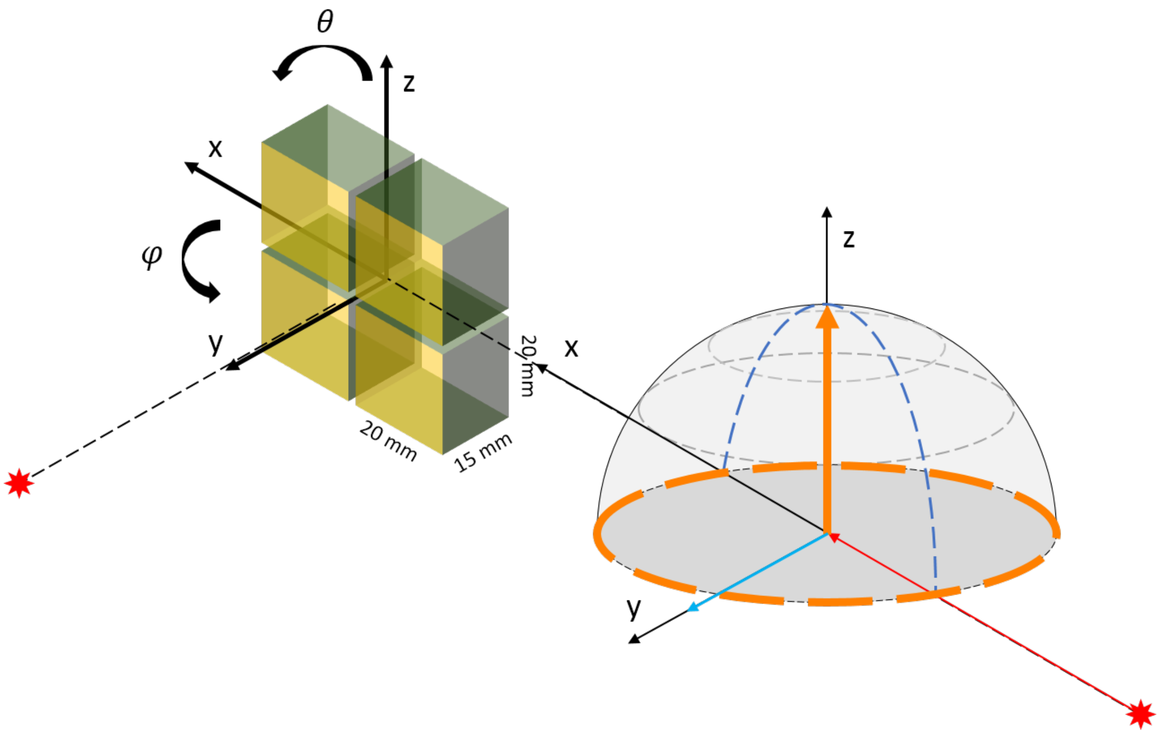


Figure 3.7: Non-isotropic detector array using a 2 × 2 Orion prototype.

ever, this would require too much computation and precludes the use of frequency spectrum based analysis. A method based on events grouping and system matrix decomposition, which separates the detector effects from the generalized PSF, was developed. With this method the geometric non-uniformity of detectors is considered and the use of frequency spectrum based analysis is preserved. Disregarding noise and discretization effects, an unbiased estimation can be achieved by FBP with a fast spherical harmonics transform. In contrast, references [32] and [34] did not focus on the detector effects but provided algorithms for filtering, and reference [33] incorporated the detector responses into the PSF.

To solve the problem of geometry non-uniformity, the adjusted FBP algorithm [54] was proposed that groups Compton events having the same type of Compton ring using a system matrix decomposition strategy. The algorithm produces more isotropic resolution than standard FBP, and preserves the capability to use frequency spectrum

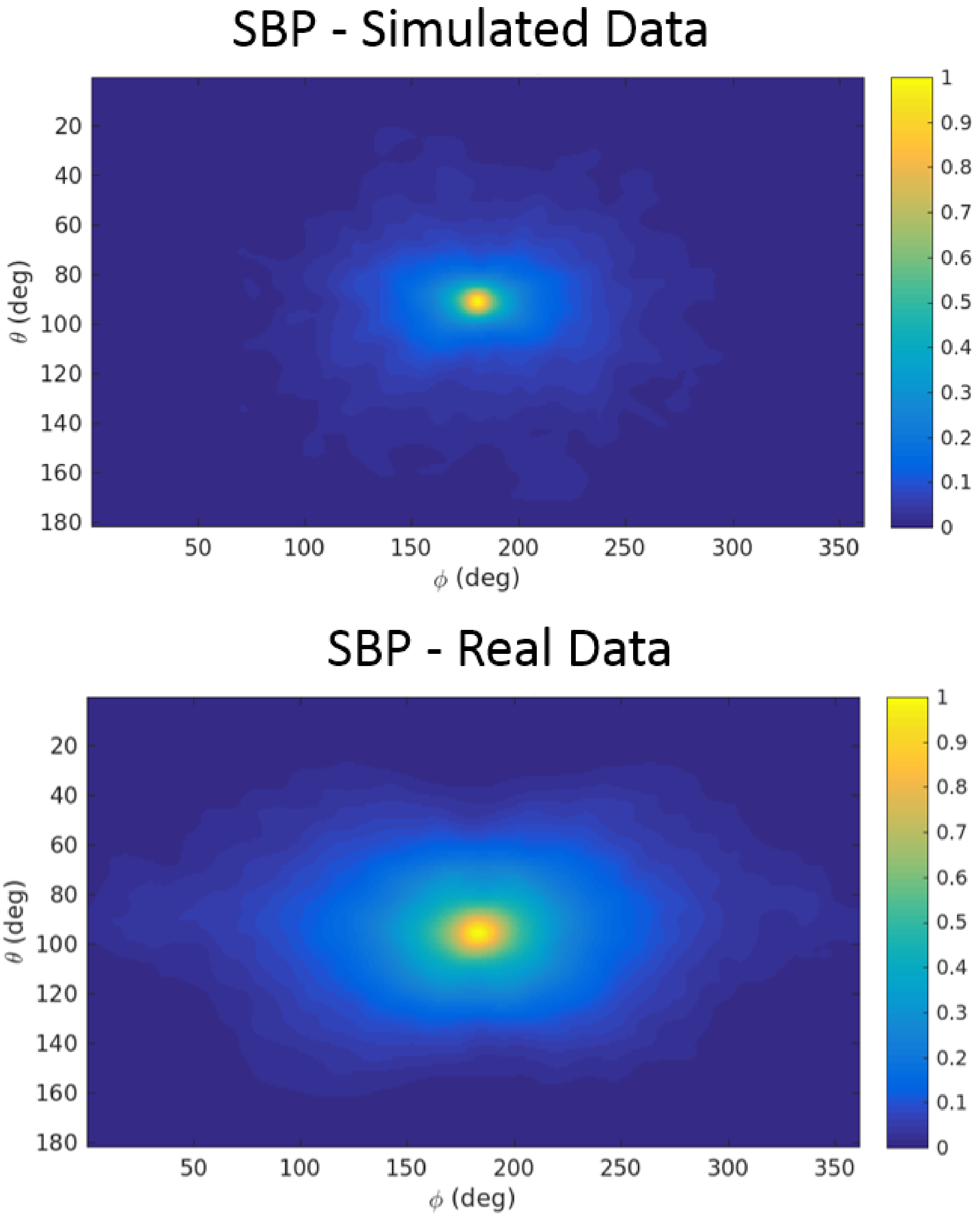


Figure 3.8: Top: SBP of a simulated Cs-137 point source at the side of the detector. Bottom: SBP of a measured Cs-137 point source at the side of the detector. Note the clear anisotropy.

based analysis. This algorithm has been applied to data from a 3D position-sensitive detector array with 4 crystals and a digital readout system.

3.3.1 Conditional Probability Chains

To derive an unbiased and computationally efficient FBP estimator, it is critical to analyze how the system matrix is generated. To simplify probability calculations, assume the incident photon energy equals the total energy deposited in the detector, and the source is monoenergetic in energy space. Split the probability elements $\{t_{ij}\}$ of system matrix T into conditional probability chains:

$$t_{ij} = \mathbb{P}(I_i^1 \mid D_j) \cdot \mathbb{P}(C_i \mid I_i^1, D_j) \cdot \mathbb{P}(E_i^1 \mid C_i, I_i^1, D_j) \cdot \mathbb{P}(I_i^2 \mid E_i^1, C_i, I_i^1, D_j), \tag{3.30}$$

where the symbols are defined in Table 3.1.

Symbol	Definition
D_j	photon from direction j
I_i^1	first interaction position of event bin i
C_i	scattering direction of event bin i
E_i^1	first deposited energy of event bin i
I_i^2	second interaction position of event bin i

Table 3.1: Symbols used in the probability chains.

3.3.2 Grouped Compton Events

Define $\{\mathbb{S}_k\}$ as the subsets of all Compton event bins, $k = 1, 2, ..., K$, where K is the number of all possible Compton rings with binned parameters. Compton events in the same subset have the same Compton ring, which means they share the same incident energy bin, the same scattering angle bin, and the same lever arm direction bin. Under the far-field assumption, the back-projected rings of these Compton events in the 4π spherical space are the same, but the interaction positions and the distances between two interactions can be different.

Because the initial photon direction will not affect the probability of the second interaction once the first interaction has been determined,

$$\mathbb{P}(I_i^2 \mid E_i^1, C_i, I_i^1, D_j) \equiv \mathbb{P}(I_i^2 \mid E_i^1, C_i, I_i^1), \tag{3.31}$$

Define

$$\eta_i \triangleq \mathbb{P}(I_i^2 \mid E_i^1, C_i, I_i^1). \tag{3.32}$$

In general, η_i is the probability that the scattered photon is absorbed at the second interaction position, which can be calculated from the exponential attenuation process in the detector.

For any i in set $\mathbb{S}_k$, as these events have the same Compton ring, denote

$$\psi_{kj} \triangleq \mathbb{P}(E_i^1 \mid C_i, I_i^1, D_j), \tag{3.33}$$

in which

$$\mathbb{P}(E_i^1 \mid C_i, I_i^1, D_j) = \begin{cases} 1, \text{ if the Compton formula is satisfied;} \\ 0, \text{ otherwise.} \end{cases} \tag{3.34}$$

It is possible to add uncertainties to blur the Compton ring, but binary form is used here for simplicity.

For any i in set $\mathbb{S}_k$, as the initial energy and the lever arm direction are determined by the subset $\mathbb{S}_k$, the Compton scattering cross section is a constant across the back-projected ring, so define

$$\sigma_k(j) \triangleq \mathbb{P}(C_i \mid I_i^1, D_j), \tag{3.35}$$

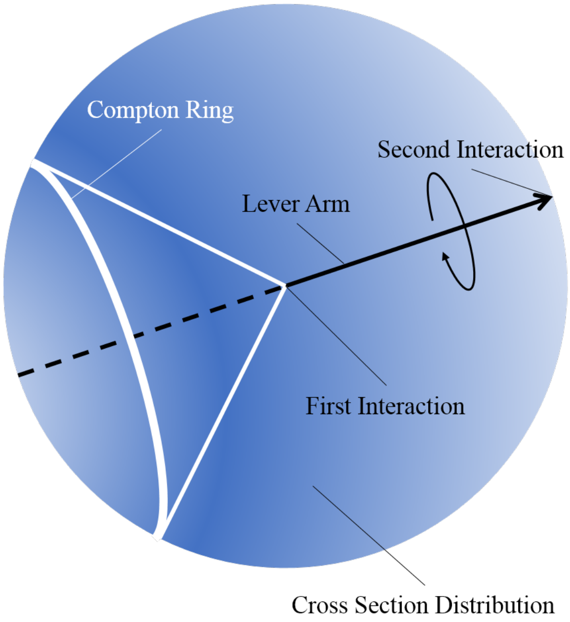


Figure 3.9: Compton ring.

$$\sigma_k(j)\psi_{kj} = \begin{cases} \sigma_k, \text{ if the Compton formula is satisfied;} \\ 0, \text{ otherwise.} \end{cases} \tag{3.36}$$

As shown in Fig. 3.9, for each event in one subset, the incident energy and the lever arm direction of the Compton ring have been determined. By the Klein–Nishina formula [45], it is known that the Compton scattering cross section is a function of incident angles, which is axisymmetric around the lever arm. In the image, darker shading represents a greater cross section, compared to a lighter shading. Given the Compton ring of this subset, the Compton scattering cross section on that ring is a constant value.

Since the incident photons usually have energy exceeding 500 keV for Compton scattering events, they have an approximately uniform probability distribution of interaction positions in the detector

$$\mathbb{P}(I_i^1 \mid D_j) \approx \alpha_j, \tag{3.37}$$

which is valid for the detector geometry and type being considered here, and is further analyzed in [54].

Consequently $\forall i \in \mathbb{S}_k$ the conditional probability chains can be written as

$$t_{ij} = \alpha_j \sigma_k \psi_{kj} \eta_i. \tag{3.38}$$

To group the probabilities by $\mathbb{S}_k$, denote:

$$u_{kj} \triangleq \sum_{i \in \mathbb{S}_k} t_{ij}, \;\; v_k \triangleq \sigma_k \sum_{i \in \mathbb{S}_k} \eta_i, \tag{3.39}$$

and from (3.38) and (3.39),

$$u_{kj} = \alpha_j \psi_{kj} \cdot \sigma_k \sum_{i \in \mathbb{S}_k} \eta_i = \alpha_j \psi_{kj} v_k = v_k \psi_{kj} \alpha_j. \tag{3.40}$$

3.3.3 System Matrix Decomposition

After grouping Compton events by the subsets $\{\mathbb{S}_k\}$, define a new aggregate measurement vector $\boldsymbol{h}$, having elements $\{h_k\}$ of $\boldsymbol{h}$ defined as

$$h_k \triangleq \sum_{i \in \mathbb{S}_k} g_i, \;\; k = 1, 2, ..., K. \tag{3.41}$$

The system model for the new observation vector is

$$\bar{\boldsymbol{h}} = \boldsymbol{U} \cdot \boldsymbol{f} \tag{3.42}$$

where $\bar{\boldsymbol{h}}$ is the expectation of $\boldsymbol{h}$, and $\boldsymbol{U}$ is the new $K \times J$ system matrix with elements $\{u_{kj}\}$. Using equation (3.40), the system matrix $\boldsymbol{U}$ is decomposed as

$$\boldsymbol{U} = \boldsymbol{V} \boldsymbol{\Psi} \boldsymbol{A}, \tag{3.43}$$

where $\boldsymbol{\Psi}$ is a $K \times J$ matrix with elements $\{\psi_{kj}\}$, $\boldsymbol{A}$ is a $J \times J$ diagonal matrix with elements $\{\alpha_j\}$, and $\boldsymbol{V}$ is a $K \times K$ diagonal matrix with elements $\{v_k\}$. $\boldsymbol{A}$ can be viewed as the correction of sensitivity, and $\boldsymbol{V}$ can be viewed as the correction of Compton scattering direction probability. Now the system model becomes:

$$\bar{\boldsymbol{h}} = \boldsymbol{V}\boldsymbol{\Psi}\boldsymbol{A}\boldsymbol{f}. \tag{3.44}$$

Note that the observation vector $\boldsymbol{h}$ still obeys a Poisson distribution with a mean vector $\bar{\boldsymbol{h}}$. Once the system model is available, different reconstruction algorithms can be applied such as SBP, FBP and MLEM. To show the benefits of the decomposed system matrix, $\boldsymbol{\Psi}'$ and $\boldsymbol{V}$ are used to do the back-projection:

$$\boldsymbol{\Psi}'\boldsymbol{\Psi}\boldsymbol{A}\boldsymbol{f} = \boldsymbol{\Psi}'\boldsymbol{V}^{-1}\bar{\boldsymbol{h}}, \tag{3.45}$$

from which a adjusted SBP estimator is derived as

$$\hat{\boldsymbol{f}}_{\text{ASBP}} = \boldsymbol{\Psi}'\boldsymbol{V}^{-1}\boldsymbol{h}. \tag{3.46}$$

Define the generalized point spread matrix $\tilde{\boldsymbol{B}}$ as

$$\tilde{\boldsymbol{B}} \triangleq \boldsymbol{\Psi}'\boldsymbol{\Psi} \tag{3.47}$$

where the jth column is the PSF of the jth pixel. This PSF is the sum of all possible Compton rings intersecting with the incident direction, so $\tilde{\boldsymbol{B}}$ is rotationally invariant and diagonalizable with the DSHT matrix

$$\tilde{\boldsymbol{B}} = \boldsymbol{Q}'\boldsymbol{\Gamma}\boldsymbol{Q} \tag{3.48}$$

in which $\mathbf{\Gamma}$ is a diagonal matrix. Finally the model becomes

$$\boldsymbol{f} = \boldsymbol{A}^{-1}\boldsymbol{Q}'\mathbf{\Gamma}^{-1}\boldsymbol{Q}\mathbf{\Psi}'\boldsymbol{V}^{-1}\bar{\boldsymbol{h}}, \tag{3.49}$$

and the adjusted FBP estimator is

$$\hat{\boldsymbol{f}}_{\text{AFBP}} = \boldsymbol{A}^{-1}\boldsymbol{Q}'\mathbf{\Gamma}^{-1}\boldsymbol{Q}\mathbf{\Psi}'\boldsymbol{V}^{-1}\boldsymbol{h}, \tag{3.50}$$

where $\boldsymbol{h}$ represents the grouped observation, $\boldsymbol{V}^{-1}$ represents the weighting for Compton rings, $\mathbf{\Psi}'$ represents SBP process, $\boldsymbol{Q}$ represents SHT, $\mathbf{\Gamma}^{-1}$ represents the inverse filter, $\boldsymbol{Q}'$ represents the inverse SHT, and $\boldsymbol{A}^{-1}$ represents the correction of sensitivity. As a comparison, the standard FBP estimator is

$$\hat{\boldsymbol{f}}_{\text{FBP}} = \boldsymbol{Q}'\mathbf{\Gamma}^{-1}\boldsymbol{Q}\mathbf{\Psi}'\boldsymbol{h}. \tag{3.51}$$

Since $\boldsymbol{A}^{-1}$ and $\boldsymbol{V}^{-1}$ are both diagonal matrices, their multiplications have $O(J)$ time complexity.

3.3.4 Experiments

The experiments used a 2×2 3D position-sensitive CZT detector array, with 2 mm gaps between crystals. Each crystal in the array was 2 cm × 2 cm × 1.5 cm, and had 11×11 pixels on the 2×2 cm^2 anode surface. With a digital readout technique [28, 37], the detector provides 0.3 mm sub-pixel resolution parallel to the anode surface, 0.3 mm depth resolution perpendicular to the anode surface. With a Cs-137 source, the energy resolution for all double-pixel events is 0.8% FWHM at 662 keV.

The maximum spherical harmonics degree L used in the DSHT was determined by the intrinsic angular resolution of the imaging system. For our system $L = 60$

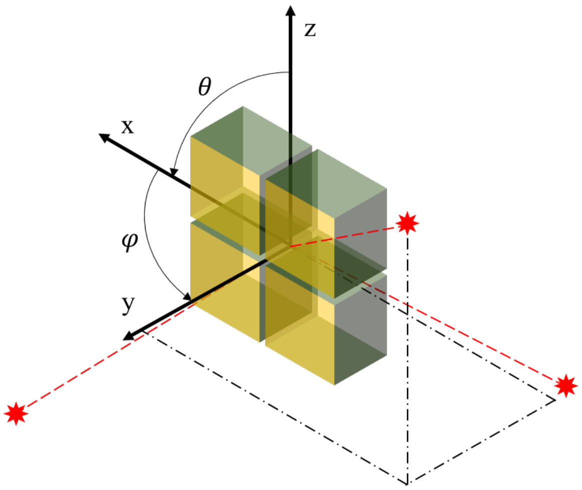


Figure 3.10: Three source locations used in the adjusted FBP experiment.

was large enough to represent all frequency information. The 4π image space was divided into a 60×120 uniform angular grid. Each sampling point on the grid has a weight based on its corresponding area. The Compton scattering angle was uniformly discretized into 180 bins, thus theoretically the number of subsets $K = 60 \times 120 \times 180 = 1,296,000$ for a given initial energy. Fortunately, as the list-mode data was used in the reconstruction, it was not necessary to calculate all subsets.

To suppress measurement noise, a DSHT-based Wiener filter was applied, as defined in (3.26) in the frequency domain. The parameter $R(\omega)$ was empirically selected as a constant 10^{-7}.

A 79 μCi Cs-137 point source was imaged from 3 different directions as shown in Fig. 3.10. The distance from the source to the detector was 1 m. The Cs-137 source was placed at $(\theta, \phi) = (90^\circ, 92^\circ)$, $(90^\circ, 185^\circ)$, and $(61^\circ, 146^\circ)$ in each measurement. The polar angle θ started from the +z axis, and the azimuth angle ϕ was counter-clockwise from the +x axis direction. The distance between the source and the detector was much larger than the size of the detector, hence the far-field condition is satisfied.

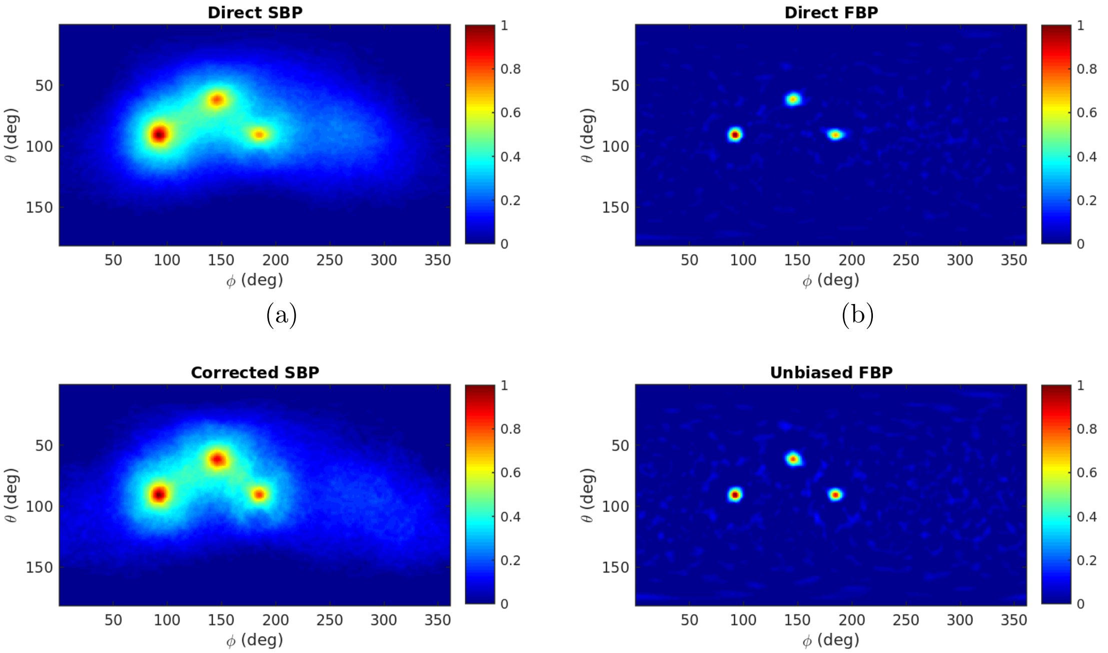
To limit events to the photopeak, an energy window of 650 - 670 keV was used, assuming full energy deposition for the summed energy from all interaction sites of a single event. Each measurement lasted 16 hours. The measurements for the sources at $(\theta, \phi) = (90^\circ, 92^\circ)$, $(90^\circ, 185^\circ)$, and $(61^\circ, 146^\circ)$ generated 276,543 multi-interaction events, 146,256 multi-interaction events, and 95,378 multi-interaction events in the energy window respectively. The detection efficiencies differed due to geometric non-uniformity and solid angle. For each measurement, only 90,000 multi-interaction events were used in reconstruction.

A simple comparison method [25] was applied to decide the sequence of interactions, namely the higher energy interaction was first as long as the Compton condition is satisfied. In each measurement, the direct and the adjusted methods were compared with the same dataset and filtering settings. The direct methods used analytical system matrix without geometric information of the detector, but the improved methods considered the geometric non-uniformity of the detector.

When the source was placed in front of the cathode side, the PSF in direct SBP was approximately isotropic and the direct filtering approach generated a round hotspot. But when the source was placed at the side of the detector array, the PSF in direct SBP was obviously anisotropic, thus direct filtering would lead to biased estimation. However, the PSF in adjusted SBP showed a rotational invariance, and the adjusted FBP was quite stable between different source directions.

Fig. 3.11 shows the reconstruction results where all measurements were combined and reconstructed simultaneously. This is similar to measuring all three sources at the same time. For each position, only 90,000 multi-interaction events were used.

In the image of direct SBP, the shapes of hotspots in different positions varied because of the geometric non-uniformity. The direct FBP could improve the FWHM of hotspots. For the hotspot at $(90^\circ, 185^\circ)$ or $(61^\circ, 146^\circ)$, the polar resolution was finer than the azimuth resolution. Thus the hotspot was biased with a non-uniform resolution. With the same data, the hotspots in the adjusted SBP became more isotropic.



(a) (b)

(c) (d)

Figure 3.11: Combined reconstruction of all three sources reconstructed simultaneously.

Compared with the direct FBP, the adjusted FBP provided a better performance in FWHM with the same filtering settings. The resolution was nearly uniform along different directions for each hotspot.

Fig. 3.12 shows the comparison between direct filtering and the new filtering approaches with contour images, where each measurement was reconstructed individually. The direct and the adjusted results were plotted in the same figure for comparison. In each image, the contours from inside to outside indicated 80%, 50%, and 20% of the peak value. Note that the SBP images and the FBP images had different scales. (a) and (b) are SBP and FBP images obtained for a source placed at $(90°, 92°)$. The images were all nearly isotropic, and no obvious improvement was observed from new methods. (c) and (d) are SBP and FBP images obtained for a source placed at $(90°, 185°)$. In both direct SBP and direct FBP, the impact of geometric non-uniformity is visible as the horizontal elongation of the source. With the same data, the adjusted SBP and the adjusted FBP resulted in more symmetric images and improved azimuthal resolution. (e) and (f) are SBP and FBP images obtained for a source placed at $(61°, 146°)$. In both direct SBP and direct FBP, the hotspots tilted to the left because of geometric non-uniformity. With the same data, the adjusted SBP and the adjusted FBP resulted in more symmetric images and improved azimuthal resolution.

3.4 Adaptive FBP

Another challenge in FBP is how to select the penalty parameter. For measurements with fewer counts, the Gaussian assumption is less valid, so there is more noise and stronger penalty parameter is desired. For measurements with more counts, the Gaussian assumption is closer to the real Poisson distribution, so resolution should be more focused and a weaker penalty is desired. However, it is neither convenient nor accurate to let the user decide the penalty parameter for each measurement. The adaptive FBP algorithm is proposed, which adaptively selects the penalty parameter basis iteratively, balancing the noise and resolution of the FBP image without manual intervention.

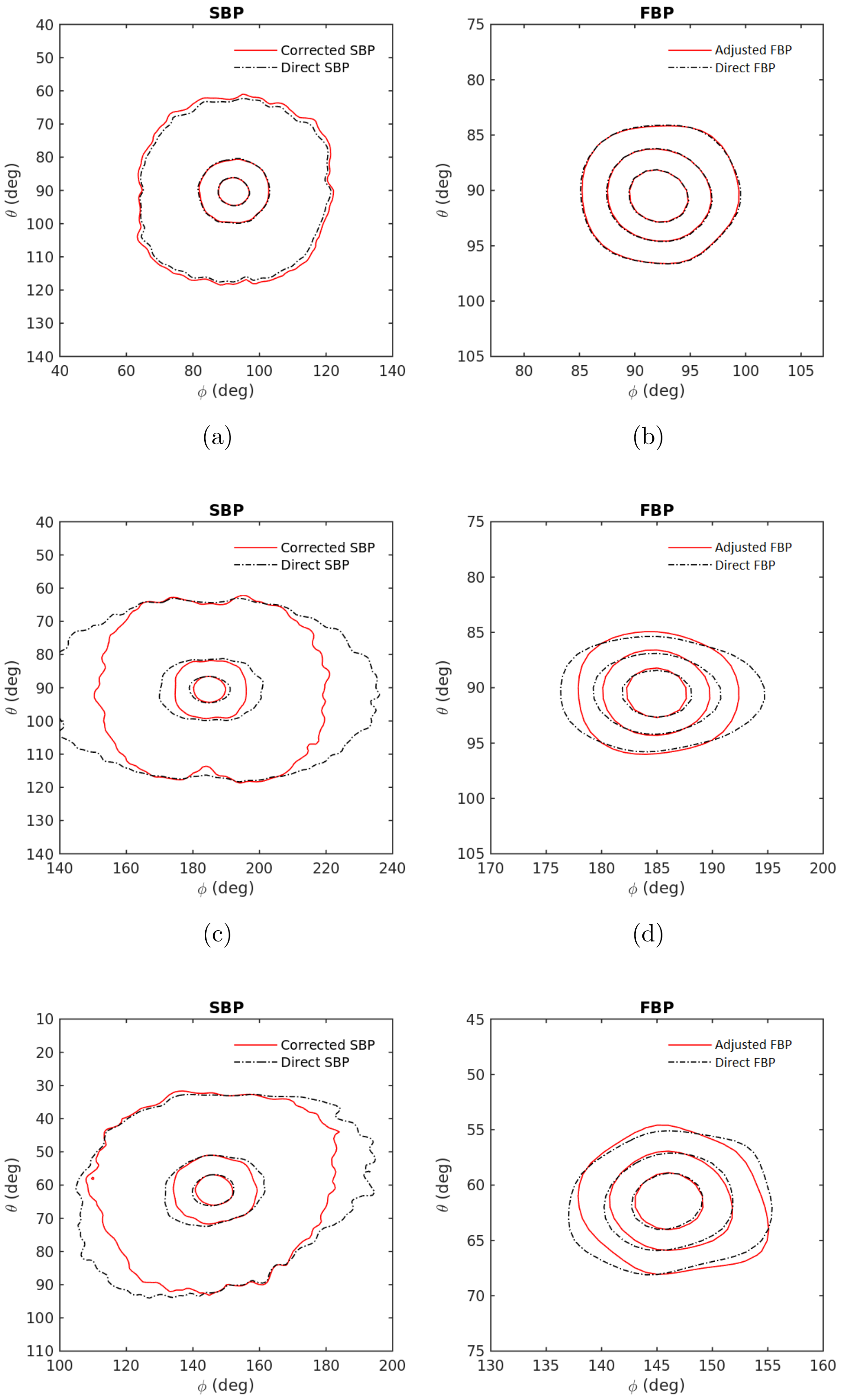
3.4.1 Penalty Parameter Computation

The power spectrum of the white Gaussian noise with variance σ^2 is simply σ^2. Since hotspots are of interest, if define

$$M \triangleq \max(\boldsymbol{f}), \tag{3.52}$$

then by the definition of power spectrum,

$$P_r(\lambda) = \sigma^2 \propto M, \tag{3.53}$$



(a) (b)

(c) (d)

(e) (f)

Figure 3.12: Zoomed in source profiles, illustrating the improvement of hotspots. Note that the axes differ.

$$P_f(\lambda) \propto M^2. \tag{3.54}$$

Thus the penalty parameter is

$$R(\lambda) = \frac{P_r(\lambda)}{P_f(\lambda)} = \frac{C}{M} = R, \tag{3.55}$$

where C is a parameter related to detector property and incident photon energy. Therefore the penalty parameter R can be selected adaptively from images.

Since M is a function of the true image $\boldsymbol{f}$, it is not possible calculate the penalty parameter R directly. However, it is reasonable to use FBP images instead of the true image, and calculate M iteratively as shown in the following

```
Require: $\hat{\boldsymbol{f}}^{(0)} = \boldsymbol{s}$
  for $k = 1$ to $K$ do
    $M^{(k)} \leftarrow \max(\hat{\boldsymbol{f}}^{(k-1)})$ (Update $M$ from previous FBP image)
    $\hat{\boldsymbol{f}}^{(k)} \leftarrow \hat{\boldsymbol{f}}_{\mathrm{FBP}}(R = \frac{C}{M^{(k)}})$ (Use $R = \frac{C}{M^{(k)}}$ as the parameter to do FBP)
  end for
  return $\hat{\boldsymbol{f}}^{(K)}$ (Resulting image)
```

3.4.2 Comparison with Empirical Penalty Parameters

Experiments were done to test the performance of the adaptive FBP. For the detector used in the measurement, $C = 10^{-5}$ was selected for 356 keV photopeak of Ba-133 point source. Fig. 3.13, Fig. 3.14 and Fig. 3.15 show the performance of the adaptive FBP in different statistics with fixed penalty parameter. In each figure, the top row shows the spectrum and the selected energy window, middle left shows the SBP image, middle right shows the adaptive FBP image, bottom left shows the FBP image with $R = 10^{-3}$, and bottom right shows the FBP image with $R = 10^{-7}$. The adaptive R is selected closer to 10^{-3} with poor statistics and to 10^{-7} with good statistics.

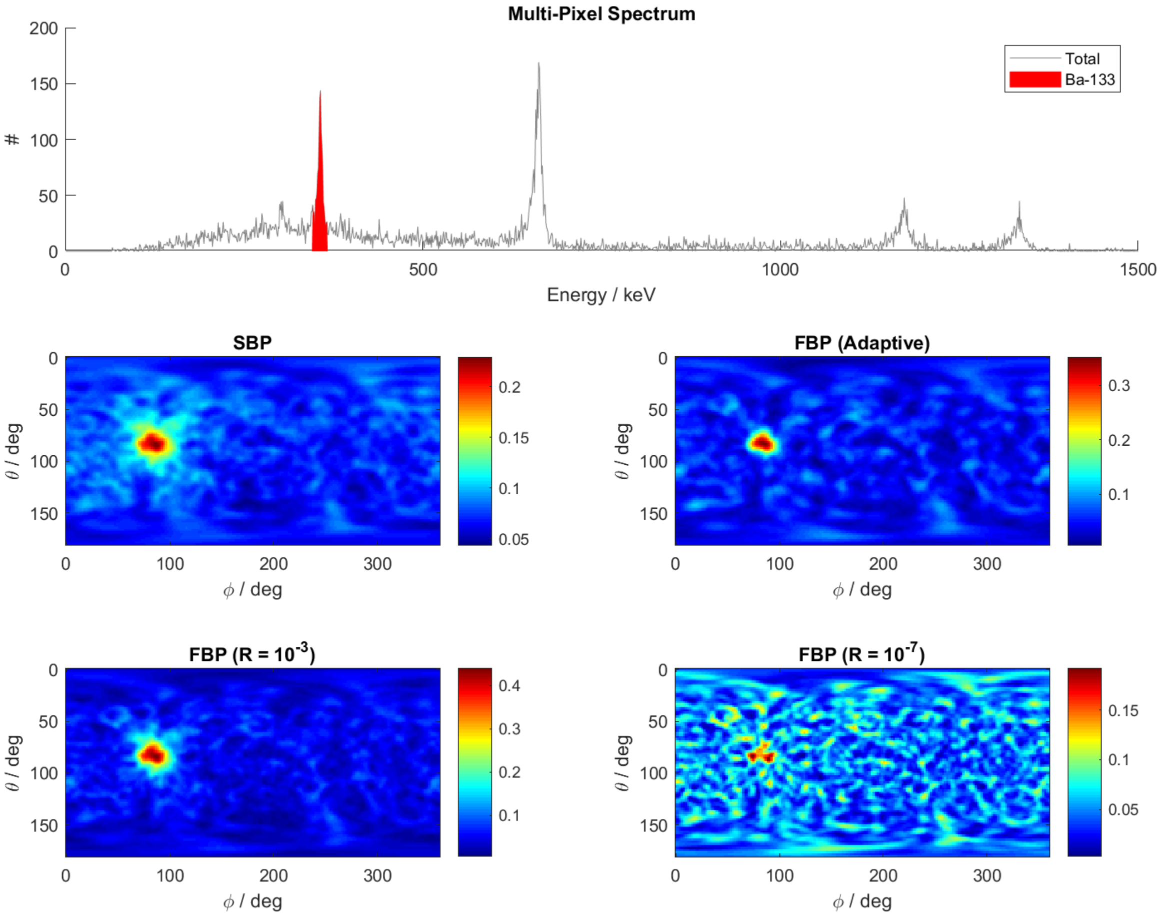


Figure 3.13: FBP with 10^3 counts in Ba-133 photopeak.

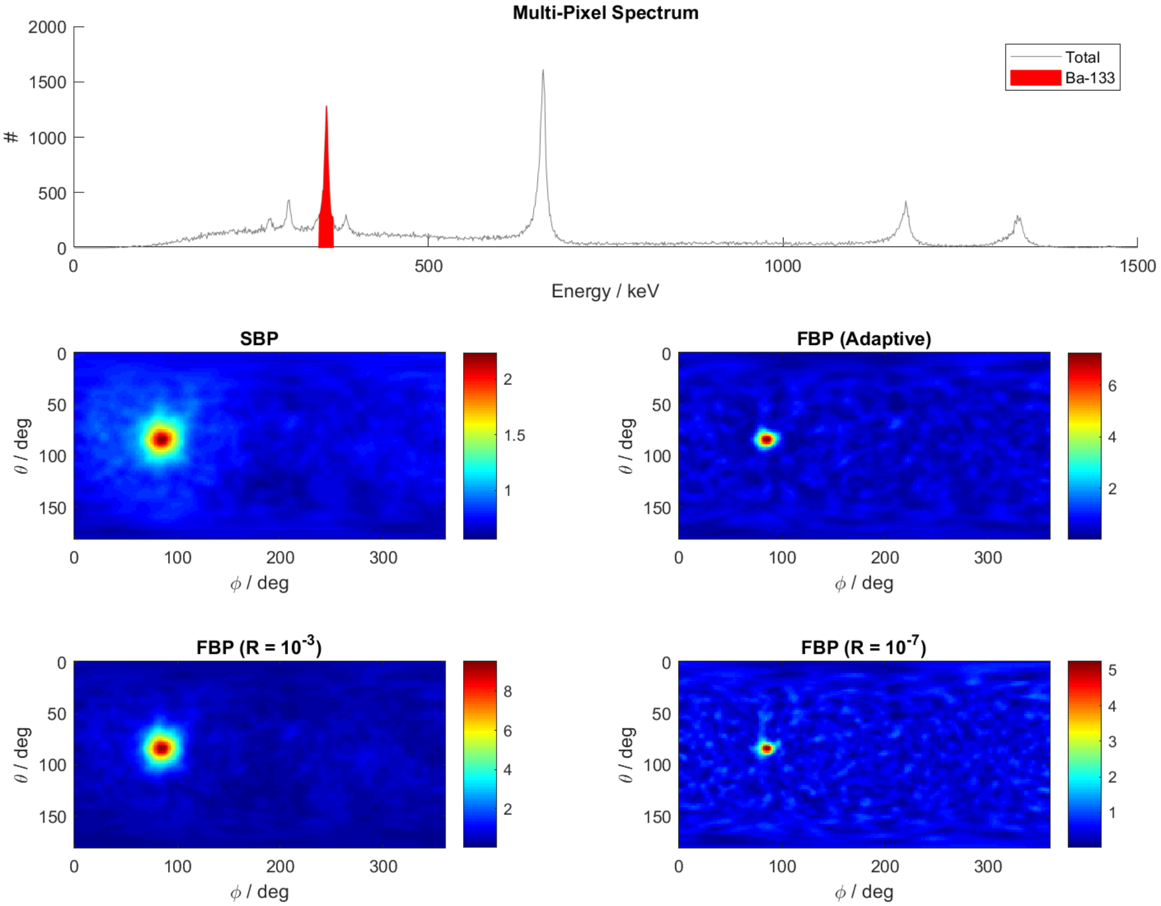


Figure 3.14: FBP with 10^4 counts in Ba-133 photopeak.

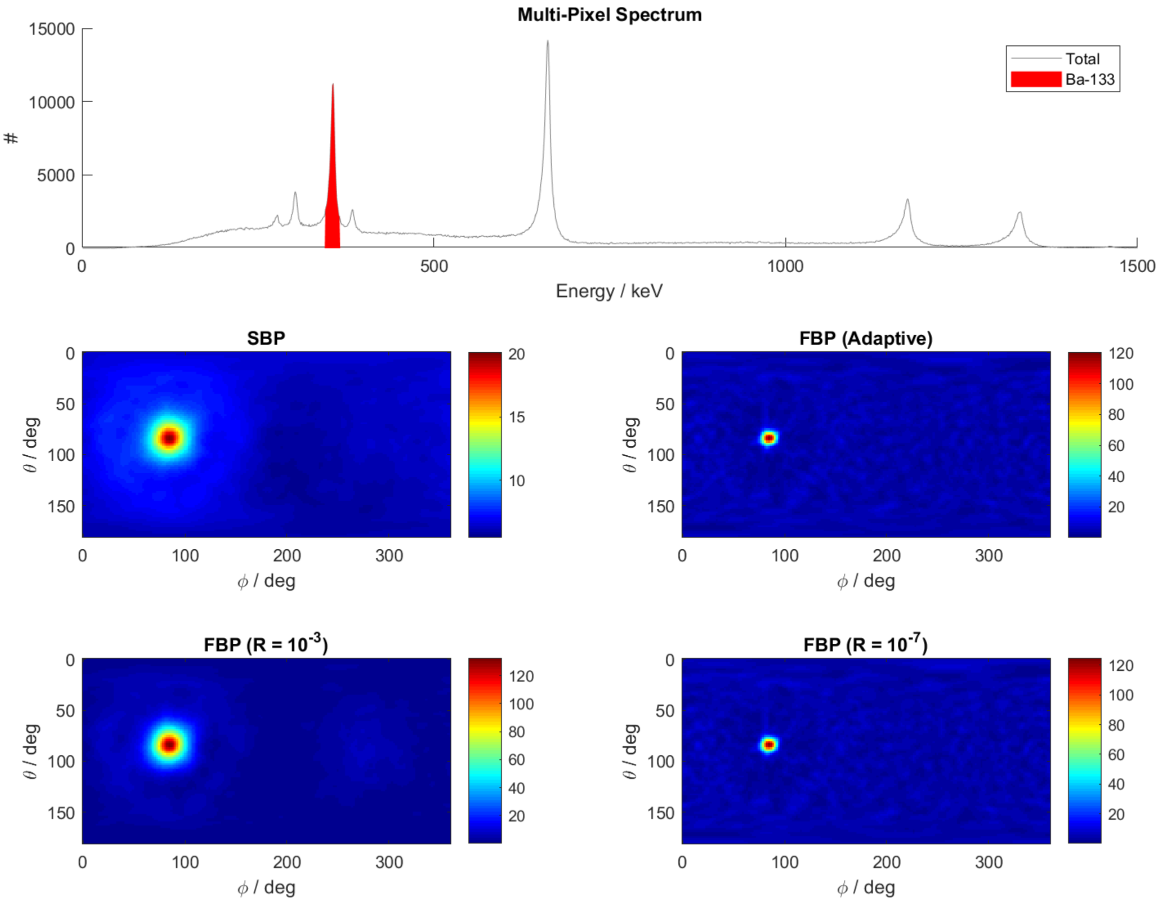


Figure 3.15: FBP with 10^5 counts in Ba-133 photopeak.

3.4.3 Adaptive Parameter Computation

Given a detector, the adaptive parameter C is a function of incident photon energy

$$C = C(E), \tag{3.56}$$

where E is the incident photon energy. It is possible to measure the properties of a detector system and generate a empirical function $C(E)$. Thus for a FBP using photons in a photopeak, $C(E)$ can be used to calculate penalty parameter R and apply the Wiener filter adaptively.

Define

$$w = -\log_{10} R, \tag{3.57}$$

$$X = \log_{10} M. \tag{3.58}$$

Because $R(E) = \frac{C(E)}{M}$,

$$-\log_{10} R = \log_{10} M - \log_{10} C(E), \tag{3.59}$$

so w and X have linear relationship. Define

$$w = a(E)X + b(E), \tag{3.60}$$

where $a(E)$ and $b(E)$ are parameters related to photopeak energy E. So for each photopeak energy E, as long as parameters $a(E)$ and $b(E)$ are known, it is possible to calculate the penalty parameter using X or M.

Fig. 3.16 shows the linear relationship between w and X empirically. In this experiment, a Cs-137 point source was placed in three positions around the detector, distinguished by different dot colors. Each data point was generated by resampling the dataset, then choosing the w that could maximize signal-to-noise ratio.

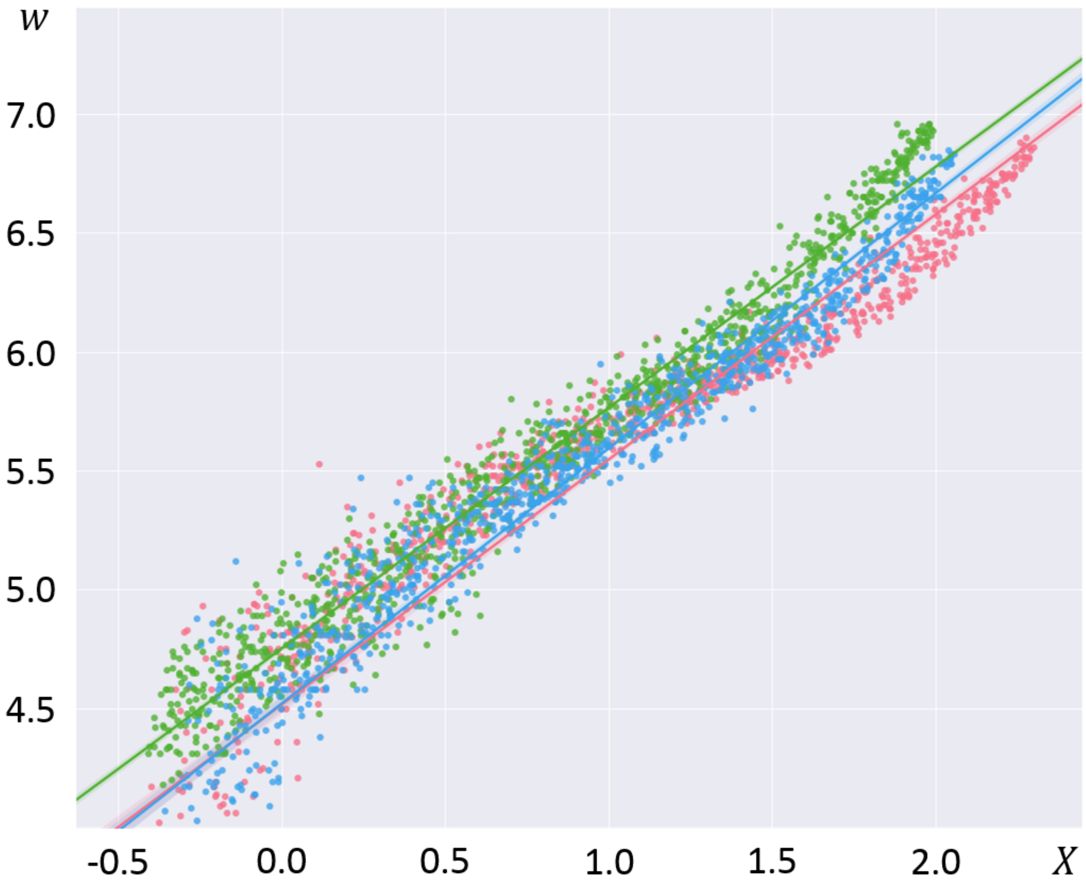


Figure 3.16: Linear relationship between w and X.

3.5 Summary

In this chapter, I propose the adjusted FBP and the adaptive FBP, making the FBP method more accurate and practical. Due to the detector non-isotropic geometry, the conventional FBP has shift-variant PSF, leading to a non-isotropic resolution. Therefore, I developed the adjusted FBP, which weights the rings based on the geometry and corrects the PSF to be shift-invariant. Another challenge for FBP is the selection of the denoising parameter. As the FBP uses Gaussian detection to approximate Poisson detection, a denoising filter should be applied whose parameter is related to the image statistics. Therefore, the adaptive FBP has been developed in this thesis, which considers the current statistics of the image and selects the Wiener parameter adaptively.

However, current adaptive method is based on the point source model, which may not be valid if the source is extended. The adaptive parameter selection in the scenarios with extended sources is an interesting topic and worth studying in future.

CHAPTER IV

Iterative Reconstruction Algorithms

Direct reconstruction algorithms are inapplicable in many imaging problems, such as the image estimation with Poisson measurements. Thus iterative reconstruction algorithms are required. This chapter will discuss the Maximum-Likelihood (ML) method for Poisson measurements, introduce its natural extension in energy-imaging integrated space as Energy-Imaging Integrated Deconvolution (EIID), and propose the accelerated method: Energy Decremental Integrated Deconvolution (EDID).

4.1 Maximum-Likelihood

Experimental measurements have Poisson distribution for detections, thus a simplified Gaussian noise model may introduce error in reconstruction if the statistics is not sufficient. To efficiently apply all information in the image reconstruction, the ML method is used to estimate image variables with Poisson observation. However, since there is no analytical solution for ML estimation problems in Poisson measurements, iterative algorithms are developed to estimate a solution given a set of observation.

4.1.1 Imaging Model

The general discrete imaging model in matrix-form is

$$\bar{\boldsymbol{g}} = \boldsymbol{T}\boldsymbol{f}, \tag{4.1}$$

where $\boldsymbol{f}$ is the image vector with length J , $\bar{\boldsymbol{g}}$ is the expected observation vector with length I, and $\boldsymbol{T}$ is the $I \times J$ system matrix with each row corresponding to a single back-projected Compton cone.

Assume the observations are mutually independent Poisson variables,

$$g_i \sim \text{Poisson}(\bar{g}_i), \tag{4.2}$$

where $i = 1, 2, ..., I$ is the index of observation vector. The logarithmic likelihood function is

$$\begin{aligned}
\mathcal{L}(\boldsymbol{g}|\boldsymbol{f}) &= \ln \mathbb{P}(\boldsymbol{g}|\boldsymbol{f}) \\
&= \ln \prod_{i=1}^{I} \mathbb{P}(g_i|\boldsymbol{f}) \\
&= \sum_{i=1}^{I} \ln \mathbb{P}(g_i|\boldsymbol{f}) \\
&= \sum_{i=1}^{I} (-\bar{g}_i + g_i \ln \bar{g}_i - \ln g_i!) \\
&\equiv \sum_{i=1}^{I} (-\bar{g}_i + g_i \ln \bar{g}_i) \\
&= \sum_{i=1}^{I} \phi_i(\bar{g}_i),
\end{aligned} \tag{4.3}$$

where $\phi_i(x) \triangleq g_i \ln x - x$, "$\equiv$" indicates that the constants independent of $\boldsymbol{f}$ are neglected, and

$$\bar{g}_i = \sum_{j=1}^{J} t_{ij} f_j. \tag{4.4}$$

The classical expectation-maximization algorithm for solving the ML estimation problem has the updating equation

$$f_j^{(k+1)} = \frac{f_j^{(k)}}{s_j} \sum_{i=1}^{I} \frac{t_{ij} g_i}{\bar{g}_i(\boldsymbol{f}^{(k)})}, \tag{4.5}$$

where $j = 1, 2, ..., J$ is the index of image pixel, k is the iteration number, $\bar{g}_i(\boldsymbol{f})$ is the expected observation defined as

$$\bar{g}_i(\boldsymbol{f}) = \sum_{j=1}^{J} t_{ij} f_j, \tag{4.6}$$

and s_j is the sensitivity defined as

$$s_j = \sum_{i=1}^{I} t_{ij}. \tag{4.7}$$

This is the famous Richardson-Lucy [55, 56] iterative procedure and is also known as the MLEM method for Poisson measurements.

4.1.2 Derivation

There are several possible derivations of the MLEM algorithm. Here one of them using the optimization transfer principles [57] is discussed. Although this is not how MLEM was derived originally, it may be the simplest rigorous derivation which illuminates the monotonicity properties of MLEM.

If $f_j^{(k)} > 0 \ \forall j = 1, 2, ..., J$, the expected observation is

$$\bar{g}_i(\boldsymbol{f}) = \sum_{j=1}^{J} t_{ij} f_j = \sum_{j=1}^{J} \frac{t_{ij} f_j^{(k)}}{\bar{g}_i^{(k)}} \cdot \frac{f_j}{f_j^{(k)}} \cdot \bar{g}_i^{(k)}, \tag{4.8}$$

where $\bar{g}_i^{(k)} \triangleq \bar{g}_i(\boldsymbol{f}^{(k)})$. $\boldsymbol{f}$ is strictly positive, thus the equation is well defined. Because $\phi_i(x) \triangleq g_i \ln x - x$ is concave, by the convexity inequality,

$$\begin{aligned}
\mathcal{L}(\boldsymbol{g}|\boldsymbol{f}) &= \sum_{i=1}^{I} \phi_i(\bar{g}_i) \\
&= \sum_{i=1}^{I} \phi_i\left(\frac{t_{ij} f_j^{(k)}}{\bar{g}_i^{(k)}} \cdot \frac{\bar{g}_i^{(k)} f_j}{f_j^{(k)}}\right) \\
&\geqslant Q(\boldsymbol{f}; \boldsymbol{f}^{(k)}) \triangleq \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{t_{ij} f_j^{(k)}}{\bar{g}_i^{(k)}} \phi_i\left(\frac{\bar{g}_i^{(k)} f_j}{f_j^{(k)}}\right) \\
&\equiv \sum_{j=1}^{J} Q_j(f_j; \boldsymbol{f}^{(k)}),
\end{aligned} \tag{4.9}$$

where

$$\begin{aligned}
Q_j(f_j; \boldsymbol{f}^{(k)}) &= \sum_{i=1}^{I} \frac{t_{ij} f_j^{(k)}}{\bar{g}_i^{(k)}} \left(g_i \ln \frac{\bar{g}_i^{(k)} f_j}{f_j^{(k)}} - \frac{\bar{g}_i^{(k)} f_j}{f_j^{(k)}}\right) \\
&= \sum_{i=1}^{I} \left(\frac{t_{ij} f_j^{(k)}}{\bar{g}_i^{(k)}} g_i \ln f_j + \frac{t_{ij} f_j^{(k)}}{\bar{g}_i^{(k)}} g_i \ln \frac{\bar{g}_i^{(k)}}{f_j^{(k)}} - t_{ij} f_j\right) \\
&\equiv \sum_{i=1}^{I} \frac{t_{ij} f_j^{(k)}}{\bar{g}_i^{(k)}} g_i \ln f_j - s_j f_j.
\end{aligned} \tag{4.10}$$

The maximization for the separable surrogate function Q reduces into the form

$$f_j^{(k+1)} = \underset{f_j \geqslant 0}{\arg\max} Q_j(f_j; \boldsymbol{f}^{(k)}). \tag{4.11}$$

Take the partial derivative

$$\frac{\partial}{\partial f_j} = \sum_{i=1}^{I} \frac{t_{ij} f_j^{(k)}}{\bar{g}_i^{(k)}} \frac{g_i}{f_j} - s_j, \tag{4.12}$$

equating to zero and considering the non-negativity constraint yields

$$f_j^{(k+1)} = \frac{f_j^{(k)}}{s_j} \sum_{i=1}^{I} \frac{t_{ij} g_i}{\bar{g}_i^{(k)}}, \tag{4.13}$$

which is the updating equation of MLEM for Poisson measurements.

4.1.3 List-Mode MLEM

In real measurement, the size of observation vector $\boldsymbol{g}$ is extremely large so that I is much larger than the number of events and it is not efficient to build the whole system matrix $\boldsymbol{T}$ explicitly. Instead, a list-mode system matrix can be built to solve the list-mode likelihood problem.

Denote $\tilde{\boldsymbol{T}}$ as the $N \times J$ list-mode system matrix, where N is the number of events. With a list of index $\{i_1, i_2, ..., i_N\}$,

$$\tilde{t}_{nj} = t_{i_n j},\ n = 1, 2, ..., N, \tag{4.14}$$

where $\tilde{t}_{nj}$ is an element of $\tilde{\boldsymbol{T}}$. Similarly, define

$$\tilde{g}_n = \bar{g}_{i_n} = \bar{g}_{i_n}(\boldsymbol{f}) \tag{4.15}$$

as the expected observation for the n-th event. Since each row represent one event, the equivalent observation is always 1, so the updating equation can be rewritten as

$$f_j^{(k+1)} = \frac{f_j^{(k)}}{s_j} \sum_{i=1}^{I} \frac{t_{ij} g_i}{\bar{g}_i^{(k)}} = \frac{f_j^{(k)}}{s_j} \sum_{n=1}^{N} \frac{\tilde{t}_{nj}}{\tilde{g}_n^{(k)}}, \tag{4.16}$$

where all $g_i = 0$ items are removed, and $g_i > 0$ items are split into several events. Note that the sensitivity term $s_j = \sum_{i=1}^{I} t_{ij}$ is still the sum of rows of $\boldsymbol{T}$, not the sum of rows of $\tilde{\boldsymbol{T}}$.

4.1.4 Regularization

One variation of MLEM algorithm is to add regularization, which is a effective method to introduce prior information and enhance the quality of reconstructed images. The general form of cost function is

$$\Psi(\boldsymbol{f}) = -\mathcal{L}(\boldsymbol{g}|\boldsymbol{f}) + \beta\mathcal{R}(\boldsymbol{f}), \tag{4.17}$$

where $-\mathcal{L}(\boldsymbol{g}|\boldsymbol{f})$ is the negative log-likelihood, $\mathcal{R}(\boldsymbol{f})$ is the penalty function, and β is the hyper-parameter that determines the strength of penalty. Typical penalties include sparseness and smoothness.

The sparse penalty using the L-2 norm is

$$\mathcal{R}(\boldsymbol{f}) = ||\boldsymbol{f}||_2 = \big(\sum_{j=1}^{J} f_j^2\big)^{\frac{1}{2}}, \tag{4.18}$$

which is also called power penalty. Similarly, the sparse penalty using the L-1 norm is

$$\mathcal{R}(\boldsymbol{f}) = ||\boldsymbol{f}||_1 = \sum_{j=1}^{J} |f_j|. \tag{4.19}$$

The smooth penalty using the L-2 norm is

$$\mathcal{R}(\boldsymbol{f}) = ||\nabla_2 \boldsymbol{f}||_2^2, \tag{4.20}$$

where ∇_2 gives the 2D gradient of the image $\boldsymbol{f}$. Similarly the smooth penalty using L-1 norm is

$$\mathcal{R}(\boldsymbol{f}) = ||\nabla_2 \boldsymbol{f}||_1. \tag{4.21}$$

More sophisticated penalty functions [58] were investigated and probably have better performance.

4.1.5 Block Iteration

There are many variations of ML method to accelerate the convergence rate, including grid refinement [59, 60, 61], raise powers [62, 63, 64, 65, 66], over-relaxation [67, 68, 62, 69, 70, 71, 72, 73], and line search [59, 72, 74]. None of them gained wide popularity, either due to potential algorithm instabilities, inconvenience of implementation, and/or the paltry resulted gains [53]. However, block iterative method is the variation that gets widely used, and the Ordered-Subsets Expectation-Maximization (OSEM) [75] has the greatest impact to the medical imaging field.

The basic idea of block iterative method is to update the image in each sub-iteration using a batch of input data. Although the sub-iterations may not have the best performance, the convergence direction is roughly the same with the iteration using the whole input dataset. The key is that the convergence direction is not the most significant target in each iteration, but can be adjusted in future iterations. Thus, the block iterative method achieves convergence with a batch of input data in each sub-iteration, leading to higher convergence rate.

It is possible to apply the block iterative method in the list-mode MLEM for Compton imaging using the similar concept. For a given list of events with indices $\{1, 2, ..., N\}$, divide the indices into M subsets $\{\mathcal{S}_1, \mathcal{S}_2, ..., \mathcal{S}_M\}$ randomly and averagely. Then the updating equation becomes

$$f_j^{(k+m/M)} = \frac{f_j^{(k+(m-1)/M)}}{s_j/M} \sum_{n\in\mathcal{S}_m} \frac{\tilde{t}_{nj}}{\tilde{g}_n^{(k+(m-1)/M)}}, \tag{4.22}$$

for $m = 1, 2, ..., M$, where $\boldsymbol{f}^{(k+1)} = \boldsymbol{f}^{k+M/M}$. Note that the sensitivity s_j is divided by M because the size of each subset is only $\frac{1}{M}$ of the original list size.

In a real reconstruction, it is natural to divide the data by equal time intervals because the measurement in each time interval is basically equivalent, as all photon detections are random. With this property, the reconstruction can be accelerated using block iterative method and images can be updated online.

4.2 EIID and EDID

Previous algorithms focus on the reconstruction using photopeak events, which is based on the full-energy deposition assumption. However, in real measurements there always exist Compton escaping events, leading to partial-energy deposition and Compton continuum. Thus a system model considering the system response in both spatial domain and spectral domain has a better description of the physical process.

EIID, formalized by Dr. Dan Xu, aimed to solve the partial-energy deposition problem by the deconvolution in the spatial-spectral integrated domain [30]. However, due to the extremely large number of pixels in the spatial-spectral integrated domain, it is a great challenge to implement EIID in reasonable time. To solve this problem, the essence of EIID is analyzed and the EDID is proposed to accelerate the convergence rate.

4.2.1 Imaging Model

The radiation image is defined in the energy-imaging integrated domain, with the energy domain defined as $\mathcal{E}$ and the imaging domain defined as $\mathcal{X}$. The continuous integrated domain is

$$\mathcal{E} \times \mathcal{X}, \tag{4.23}$$

and the radiation image as

$$f = f(e, \boldsymbol{x}), \tag{4.24}$$

where $e \in \mathcal{E}$ is the incident photon energy, and $\boldsymbol{x} \in \mathcal{X}$ is the source location vector. Most of the time the energy domain is 1D, and the imaging domain is a 2D spherical space.

In practice, a discretized domain is used in computation. A general discrete imaging space is defined as $\mathscr{J} = \{1, 2, ..., J\}$, which are the indices of pixels. The pixelated image is

$$f[j] = \int\limits_{e \in r_j} \int\limits_{\boldsymbol{x} \in a_j} f(e, \boldsymbol{x}) de d\boldsymbol{x}, \tag{4.25}$$

where $j \in \mathscr{J}$, r_j is the energy range of the j-th pixel, and a_j is the solid angle area of the j-th pixel. The image vector can be represented as $\boldsymbol{f}$, where $f_j = f[j]$ and $j = 1, 2, ..., J$.

Since the detection space does not change, the observation vector is still $\boldsymbol{g}$ with length I, so the imaging model is

$$\bar{\boldsymbol{g}} = \boldsymbol{T}\boldsymbol{f}, \tag{4.26}$$

where the system matrix $\boldsymbol{T}$ has size $I \times J$ and elements $\{t_{ij}\}$.

4.2.2 System Model

To calculate the system matrix element t_{ij}, one need to consider the probability that a photon with energy in range r_j emitted from solid angle a_j is detected as event type i.

$$t_{ij} = \mathbb{P}(\text{event } i \mid \text{photon from pixel } j) = \mathbb{P}(\text{event } i \mid r_j, a_j). \tag{4.27}$$

If the deposited energy E_0 of event i is in the range r_j, then it is a photopeak event; if E_0 is below the range r_j, then it is a Compton continuum event; if E_0 is above the range r_j, then the probability is 0 since a photon cannot deposit more energy than its initial energy. Refer to [30] for detailed calculation of the probabilities in system matrices.

Fig. 4.1 shows the probability distribution in energy-imaging integrated domain given an event using ideal detectors. The deposited energy is 400 keV, and the probability of full-energy deposition (purple) is significantly higher than partial-energy deposition (blue). For each energy bin, the probability distribution is a Compton ring in polar-azimuth 4π spherical space. Consider the Compton formula where $m_e c^2$

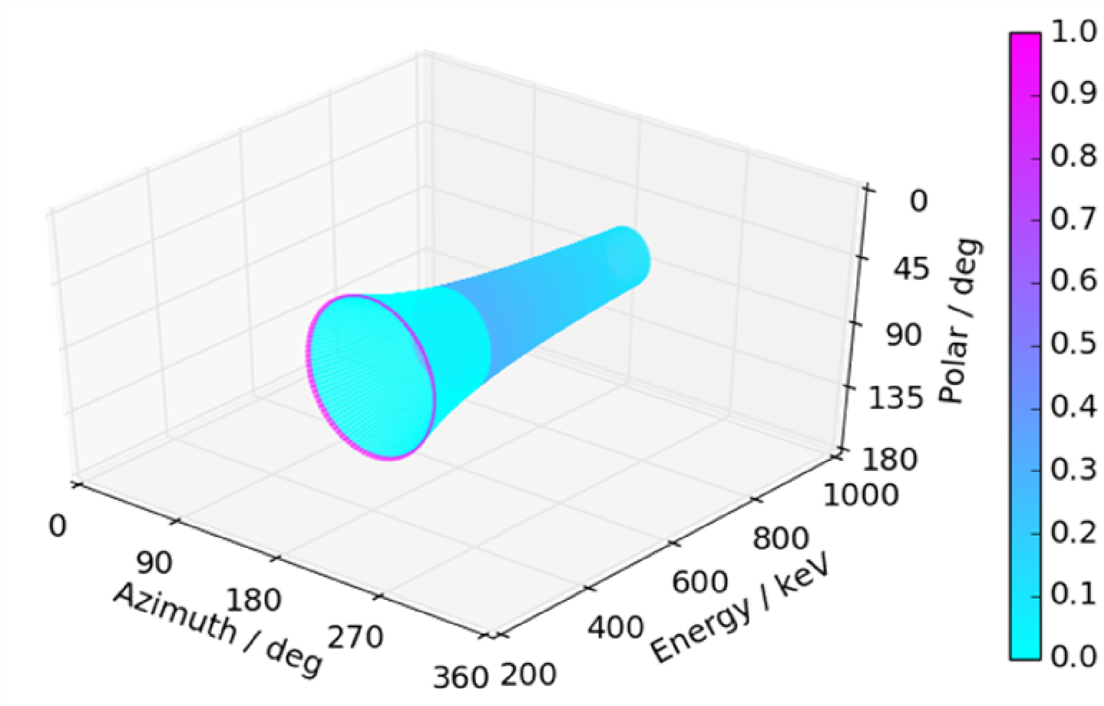


Figure 4.1: Normalized probability distribution in energy-imaging integrated domain for a 400 keV event in an ideal point detector.

is the rest mass energy of an electron, the opening angle for a 2-pixel event is

$$\theta = \cos^{-1}(1 - \frac{m_e c^2 E_1}{E_0 E_2}), \tag{4.28}$$

it is obvious that the higher the initial energy is, the smaller the ring radius is, which makes the probability distribution in energy-imaging integrated domain horn shaped.

If the probability distribution is projected to energy domain, there will be a peak of full-energy deposition and a low, but long, continuum of partial-energy deposition. Fig. 4.2 shows the probability distributions of different events in the energy domain. The true energy of incident photons is 662 keV from a Cs-137 source, but the recorded energy can be as low as 300 keV because of photon escaping. The events are sorted by their recorded energies. The full-energy deposition peaks dominate the probability distribution, but there is still information contained in the partial-energy deposition

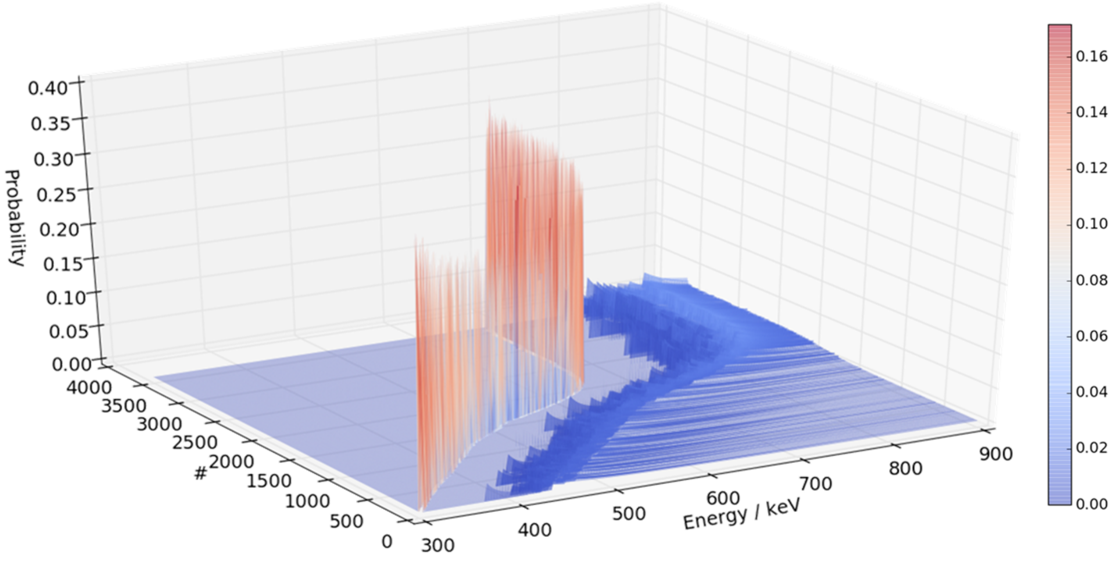


Figure 4.2: Probability distributions of different events, projected to energy domain.

tail.

4.2.3 Jiyang Bagel Phenomenon

An interesting phenomenon was observed using SBP imaging with full-energy deposition assumption.

Usually, photopeak events are selected for SBP, which have a high probability of full-energy deposition. If the original image is a point source, then a blurred hotspot should be reconstructed, indicating the direction of the source. However, it is still possible to get some information from the source using Compton continuum events. As shown in Fig. 4.3 and Fig. 4.4, several different energy windows are selected for a 4π image reconstruction of a simulated point source with 600 keV photon energy, assuming full-energy deposition. A hotspot is reconstructed using the 600 ± 25 keV energy window. As the energy of the window becomes lower and lower, the “hotspot ” becomes larger and larger and a hole appears at the center. This phenomenon was named as the “Jiyang Bagel” by Professor Zhong He. Although it becomes a bagel instead of a hotspot in low energies, the center of the hole still indicates the direction of the point source, and it is probable to extract information from partial-deposited events.

A simple explanation of the Jiyang Bagel is shown in Fig. 4.5. Define the incident photon direction as $\mathbf{\Omega}$, the deposited energy of the first Compton interaction as E_1, the scattering energy of the first Compton interaction as E_2, and the scattering angle as θ. The Compton formula gives

$$\cos\theta = 1 - \frac{m_e c^2 E_1}{(E_1 + E_2)E_2}. \tag{4.29}$$

However, the recorded energy may differ from the true energy. If the scattered photon energy is recorded as $E_2' = E_2$, then $\theta' = \theta$ and the reconstructed ring intersect with the true photon direction; if the scattered photon energy is recorded as $E_2'' < E_2$, then $\theta'' > \theta$ and the reconstructed ring is larger than the expectation. Under the full-energy deposition assumption, if photopeak events are used in SBP, then most events will intersect with the true photon direction and a hotspot will be generated. But if the partial-deposited events are used in SBP, then the back-projected rings will precess around the true photon direction and a bagel will be generated.

David Goodman did an experiment using a Cs-137 point source with 662 keV photopeak energy. Fig. 4.6 shows the SBP reconstructed image with 450 ± 25keV energy window, where the bagel is clearly observed.

4.2.4 Algorithm and Acceleration Using EDID

Given the imaging model equation (4.26) with Poisson measurement assumption, one can use MLEM iteration (4.5) to estimate the radiation image. However, a typical EIID reconstruction in 4π spherical space and 0-2 MeV energy range has a huge imaging domain. Due to the slow convergence rate of MLEM algorithm and the

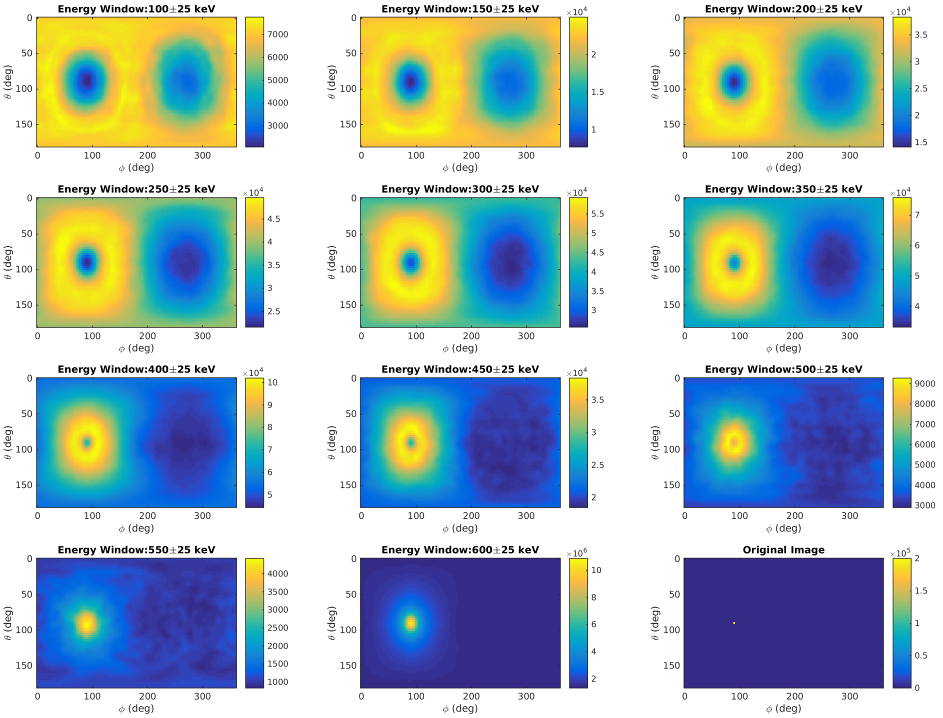


Figure 4.3: SBP in different energy windows.

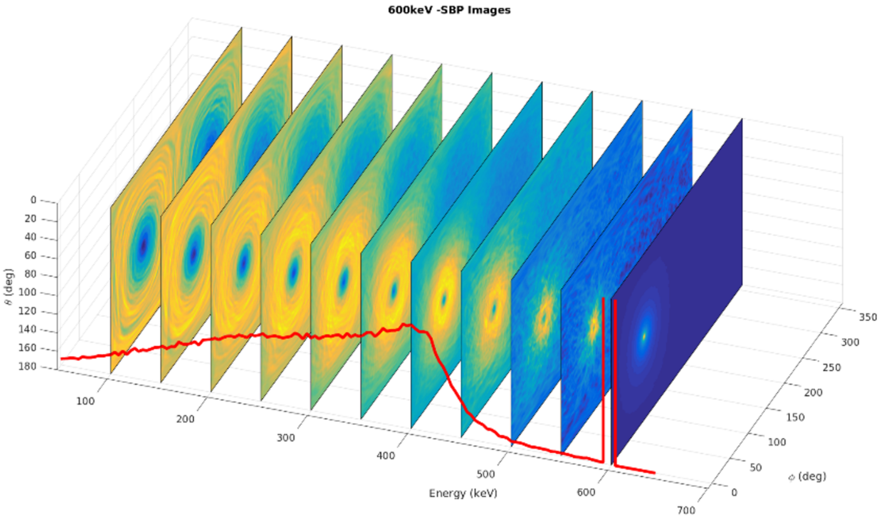


Figure 4.4: SBP in different energy windows shown in 3D.

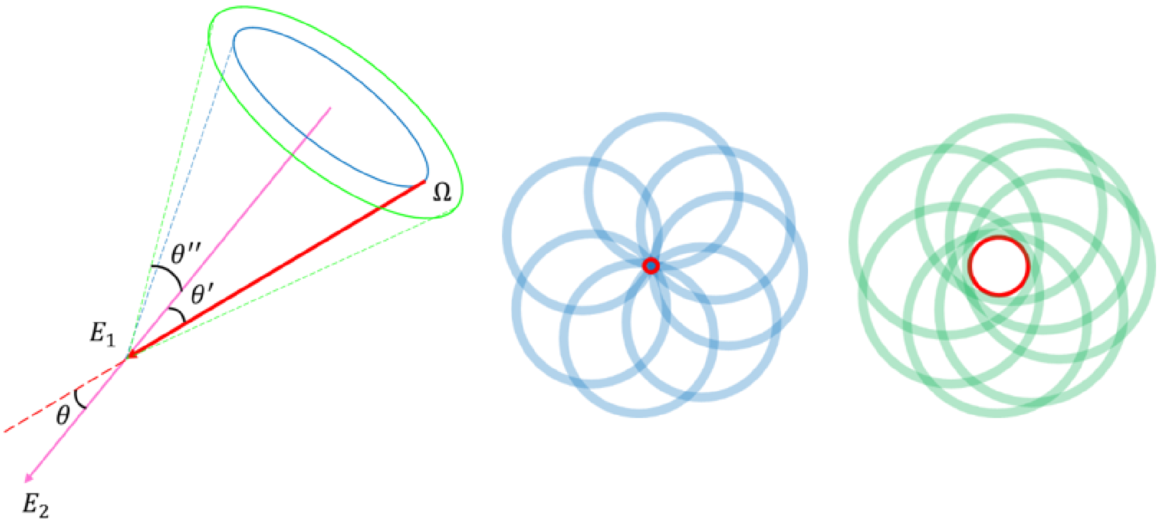


Figure 4.5: Left: The Compton cone with correct total energy and the Compton cone with underestimated total energy. Middle: SBP rings with correct total energies. Right: SBP rings with underestimated total energies.

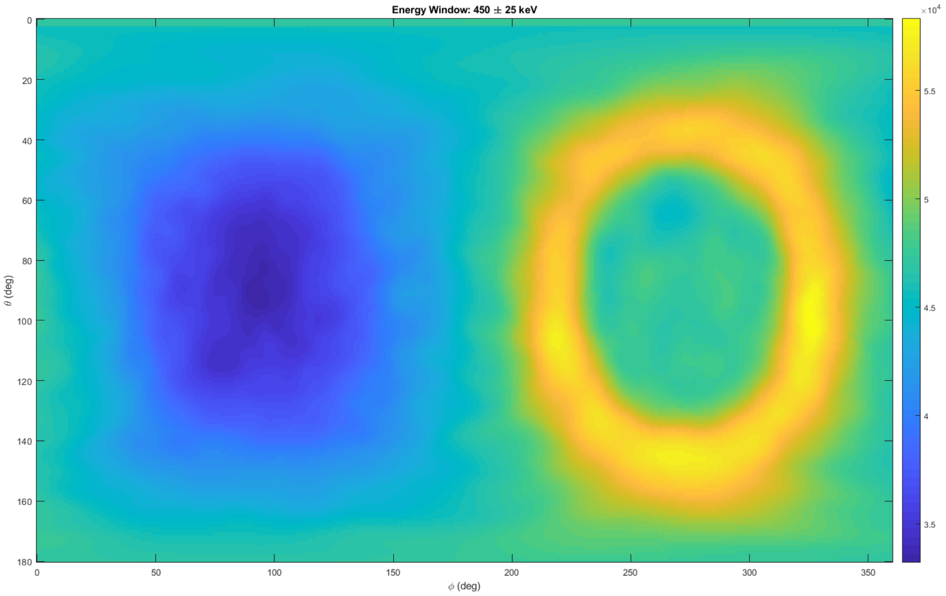


Figure 4.6: Jiyang Bagel observed in a real experiment using a Cs-137 point source.

large number of events, EIID is not practical.

4.2.4.1 Standard MLEM in Whole Space

Denote the number of energy bins as n_e, the number of imaging bins as n_x, the number of all bins as $J = n_e \times n_x$, the number of events as N, and the number of iterations as K, then the MLEM algorithm is

```
Require: $\boldsymbol{f}^{(0)}$: Non-zero initial image vector
  for $k = 1$ to $K$ do
    for $n = 1$ to $N$ do
      $\tilde{g}_n^{(k)} \leftarrow \sum_{j=1}^{J} \tilde{t}_{nj} f_j^{(k-1)}$
    end for
    for $j = 1$ to $J$ do
      $f_j^{(k)} \leftarrow \frac{f_j^{(k-1)}}{s_j} \sum_{n=1}^{N} \tilde{t}_{nj} \frac{\tilde{g}_n}{\tilde{g}_n^{(k)}}$
    end for
  end for
  return $\boldsymbol{f}^{(K)}$ (Resulting image)
```

where the time complexity for each iteration is $\mathcal{O}(NJ)$.

Assuming there are 90×180 imaging bins and 2000 energy bins, for a measurement with 1×10^4 events, the system matrix requires 2592 Gigabyte of memory using float number, which is obviously too large for a modern computer system. Although on-the-fly method is possible for the computation, it requires extra computing resource and much slower. Typically 20 iterations are required to guarantee convergence, so the system matrix has to be computed 20 times on-the-fly, which is not affordable in practice. Thus it is critical to find some methods to help simplify the system model and accelerate convergence.

4.2.4.2 MLEM in Each Energy Bin

Consider the property of the system matrix. For an event with deposited energy E_{dep}, define the the incident energy as E_{inc} and $\mathbb{P}(E_{\text{dep}} = E_{\text{inc}}) \gg \mathbb{P}(E_{\text{dep}} < E_{\text{inc}})$. In other words, given an event, full-energy deposition is much more probable than partial-energy deposition.

Fig. 4.7 shows the colored system matrix of EIID, where n_e is the number of energy bins, I_n is the number of events with deposited energy E_n, J_n is the number of bins with incident energy E_n, $\tilde{\boldsymbol{T}}_{n_1 n_2}$ is the block on the diagonal corresponding to deposited energy E_{n_1} and incident energy E_{n_2}, $\boldsymbol{f}_n$ is the image vector with incident energy E_n, and $\tilde{\boldsymbol{g}}_n$ is the observation vector with deposited energy E_n. The system matrix is divided to blocks based on the deposited energy and incident energy. It is known that the blocks on the diagonal, where deposited energy equals to incident energy, has much larger value than the blocks on the upper triangle, where deposited energy is smaller than incident energy. Since deposited energy can never be larger than incident energy, the lower triangle blocks always have zero value.

A naive approach to simplify the system matrix is to ignore the probability of partial-energy deposition on the upper triangle, and only consider the probability of full-energy deposition on the diagonal blocks. As shown in Fig. 4.8, it is obvious to have

$$\tilde{\boldsymbol{g}}_n = \tilde{\boldsymbol{T}}_{nn}\boldsymbol{f}_n, n = 1, 2, ..., n_e, \tag{4.30}$$

which means the images in different energy bins can be reconstructed separately, and sub-iteration is allowed in each energy bin. Assume M sub-iterations are done for each iteration, then the algorithm is

```
Require: $\boldsymbol{f}^{(0)}$: Non-zero initial image vector
  for $k = 0$ to $K - 1$ do
    for $n = 1$ to $n_e$ do
```

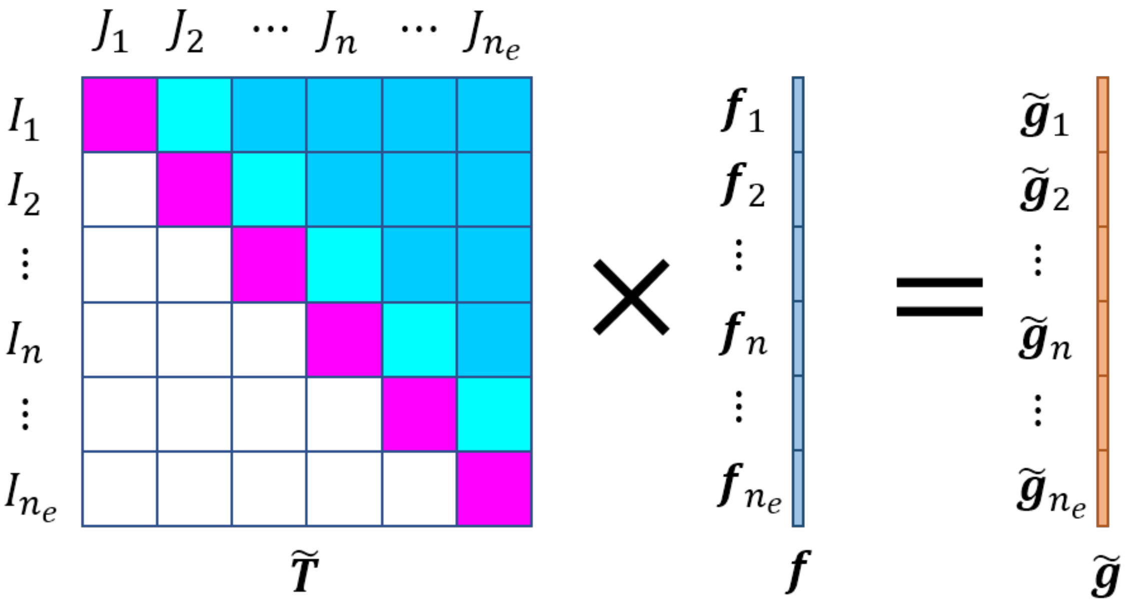


Figure 4.7: System matrix of EIID. The diagonal blocks colored as purple have much larger values than the upper triangle blocks colored as cyan.

```
    $\tau_{ij} \leftarrow$ $i$-row $j$-column element of $\tilde{\boldsymbol{T}}_{nn}$
    $f_j \leftarrow$ $j$-th element of $\boldsymbol{f}_n$
    $g_i \leftarrow$ $i$-th element of $\tilde{\boldsymbol{g}}_n$
    $s_j \leftarrow$ sensitivity of $f_j$
    for $m = 0$ to $M - 1$ do
      for $i = 1$ to $I_n$ do
        $g'_i \leftarrow \sum_{j=1}^{J_n} \tau_{ij} f_j^{(k+\frac{m}{M})}$
      end for
      for $j = 1$ to $J_n$ do
        $f_j^{(k+\frac{m+1}{M})} \leftarrow f_j^{(k+\frac{m}{M})}/s_j \times \sum_{i=1}^{I_n} \tau_{ij} \frac{g_i}{g'_i}$
      end for
    end for
  end for
end for
```

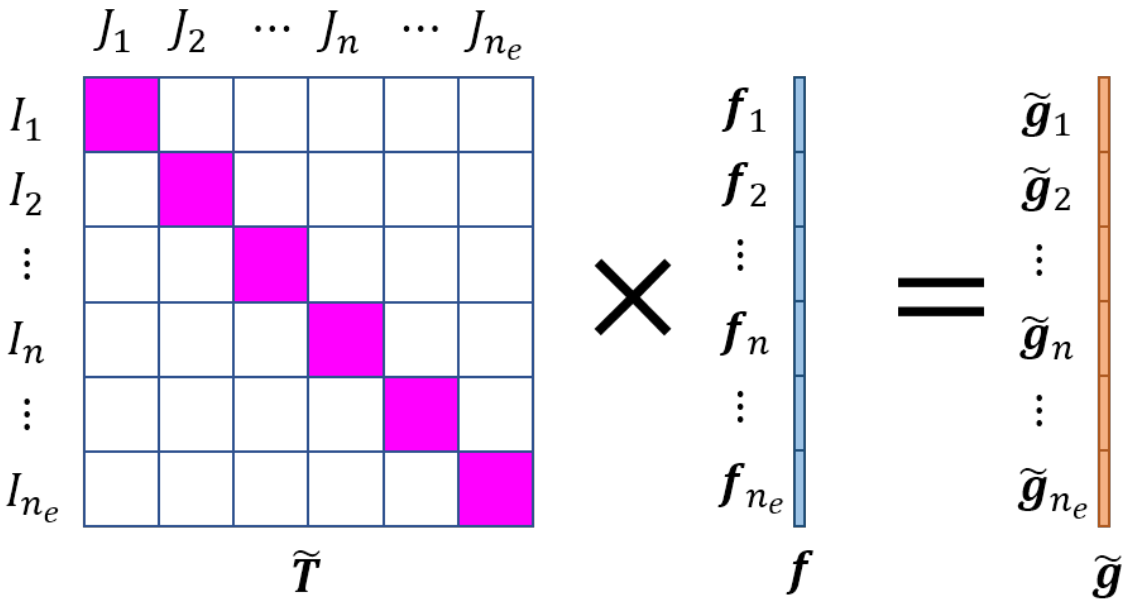


Figure 4.8: System matrix of EIID ignoring partial-energy deposition on the upper triangle.

return $\boldsymbol{f}^{(K)}$ (Resulting image)

where the time complexity for each iteration is

$$\mathcal{O}(\sum_{n=1}^{n_e} M I_n J_n) = \mathcal{O}(M n_x \sum_{n=1}^{n_e} I_n) = \mathcal{O}(N M n_x), \tag{4.31}$$

where $J_n = n_x$, $M \ll n_e$, so $M n_x \ll J$, and the time complexity of the simplified algorithm is much smaller than the standard MLEM in EIID. Although the simplified algorithm achieves much smaller time complexity, it loses information between different energies and leads to a biased estimation. Actually this algorithm is equivalent to doing image reconstructions in different energy bins without considering partial-energy deposition.

4.2.4.3 Energy-Decremental Integrated Deconvolution Algorithm

Is there a way to connect the information from different energy bins while preserve the simplicity? Here I propose the Energy Decremental Integrated Deconvolution (EDID) algorithm, which conducts sub-iterations decrementally from high energy bins to low energy bins, preserves most of the partial-energy deposition information, has the same time complexity as the standard MLEM with much faster convergence rate, and provides an unbiased estimation in energy-imaging integrated space.

Consider the property of the system matrix. Since the deposited energy can never be larger than the initial energy, the probabilities in lower triangle is always zero. Also

$$\mathbb{P}(E_{\text{dep}} = E_{\text{inc}}) \gg \mathbb{P}(E_{\text{dep}} < E_{\text{inc}}). \tag{4.32}$$

In another word, given an event, full-energy deposition is much more probable than partial-energy deposition.

Let's look into events with deposited energy E_{n_e} on the last row blocks. Because E_{n_e} is the largest energy defined in our integrated space, these events only have probability of full-energy deposition, so the reconstruction in energy bin E_{n_e} can be done directly using the events with deposited energy E_{n_e} and M sub-iterations. Although the events with lower deposited energy may contribute to the image in the energy bin E_{n_e}, the information within the events with deposited energy E_{n_e} dominates and the lower energy events are ignorable.

Next look into the events with deposited energy E_i, assuming the events with deposited energy higher than E_i have been used and the images with initial energy higher than E_i have been reconstructed. As shown in Fig. 4.9, the images $\{\boldsymbol{f}_{i+1}, \boldsymbol{f}_{i+2}, ..., \boldsymbol{f}_{n_e}\}$ have been reconstructed, and the image $\boldsymbol{f}_i$ is expected to reconstruct using the I_i events with deposited energy E_i and M sub-iterations. Since $\{\boldsymbol{f}_{i+1}, \boldsymbol{f}_{i+2}, ..., \boldsymbol{f}_{n_e}\}$ have been determined by previous steps, only $\boldsymbol{f}_i$ gets updated in

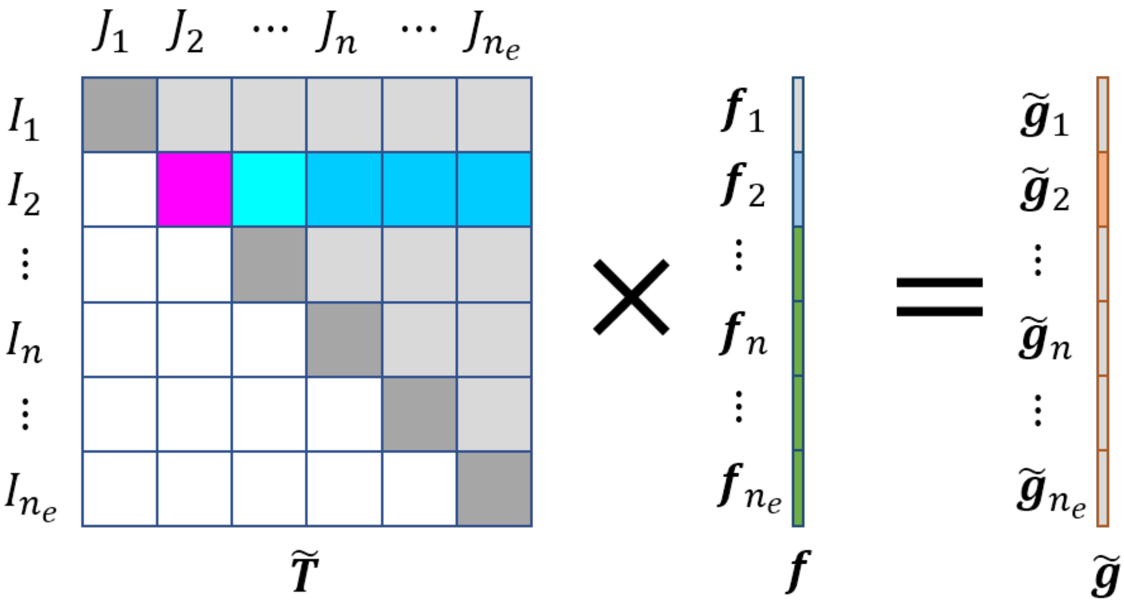


Figure 4.9: System matrix used in EDID. In each energy bin, only consider the influence from higher energy bins, and update the image under the current energy bin.

each sub-iteration.

In summary, the EDID algorithm conducts reconstructions from high energy bins to low energy bins. In each energy bin, the images with energy higher than or equal to current energy are considered in the forward-projection step, but only the image with current energy is updated in each sub-iteration. The algorithm is

```
Require: $\boldsymbol{f}^{(0)}$: Non-zero initial image vector
  for $k = 0$ to $K - 1$ do
    for $n = 1$ to $n_e$ do
      $\tau_{ij} \leftarrow$ $i$-row $j$-column element of $[\tilde{\boldsymbol{T}}_{n,n}, \tilde{\boldsymbol{T}}_{n,n+1}, ..., \tilde{\boldsymbol{T}}_{n,n_e}]$
      $f_j \leftarrow$ $j$-th element of $[\boldsymbol{f}_n; \boldsymbol{f}_{n+1}; ...; \boldsymbol{f}_{n_e}]$
      $g_i \leftarrow$ $i$-th element of $\tilde{\boldsymbol{g}}_n$
      $s_j \leftarrow$ sensitivity of $f_j$
      $U_n \leftarrow \sum_{l=n}^{n_e} J_l = (n_e - n + 1)n_x$
```

```
            for $i = 1$ to $I_n$ do
                $w_{ni}^{(k)} \leftarrow \sum_{j=J_n+1}^{U_n} \tau_{ij} f_j^{(k)}$
            end for
            for $m = 0$ to $M - 1$ do
                for $i = 1$ to $I_n$ do
                    $g'_i \leftarrow \sum_{j=1}^{U_n} \tau_{ij} f_j^{(k+\frac{m}{M})} = \sum_{j=1}^{J_n} \tau_{ij} f_j^{(k+\frac{m}{M})} + w_{ni}^{(k)}$
                end for
                for $j = 1$ to $J_n$ do
                    $f_j^{(k+\frac{m+1}{M})} \leftarrow f_j^{(k+\frac{m}{M})} / s_j \times \sum_{i=1}^{I_n} \tau_{ij} \frac{g_i}{g'_i}$
                end for
            end for
        end for
    end for
    return $\boldsymbol{f}^{(K)}$ (Resulting image)
```

Note that g'_i is split into two parts: the first part is updated in each sub-iteration, and the second part $w_{ni}^{(k)}$ is fixed given n. For each iteration, computing g'_i requires $\mathcal{O}(\sum_{n=1}^{n_e} M I_n J_n) = \mathcal{O}(NMn_x)$ time complexity. Updating f_j requires the same time complexity $\mathcal{O}(NMn_x)$. Computing w_{ni} requires

$$\mathcal{O}(\sum_{n=1}^{n_e} \sum_{j=J_n+1}^{U_n} I_n) = \mathcal{O}(\sum_{n=1}^{n_e} I_n(n_e - n)n_x) \leq \mathcal{O}(\sum_{n=1}^{n_e} I_n n_e n_x) = \mathcal{O}(NJ). \tag{4.33}$$

Since $M \ll n_e$, $Mn_x \ll n_e n_x = J$, so the total time complexity is $\mathcal{O}(NJ)$, which is the same as the standard MLEM. However, M sub-iterations are conducted in every k-th iteration, therefore a much faster convergence rate is achieved.

The negative log likelihood is defined as

$$\mathcal{N} = -\mathcal{L}(\boldsymbol{g}|\boldsymbol{f}) \equiv -\sum_{i=1}^{I} \phi_i(\sum_{j=1}^{J} \tilde{t}_{ij} f_j), \tag{4.34}$$

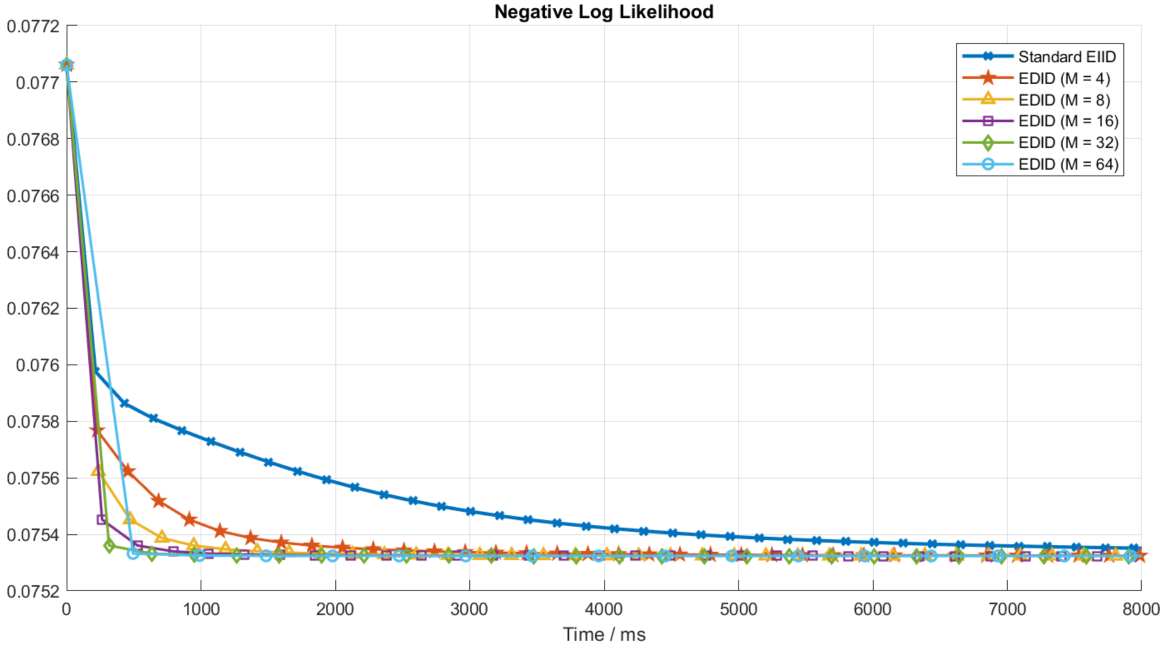


Figure 4.10: Negative log likelihood of the iterations in 8000 ms.

where $\phi_i(x) \triangleq g_i \ln x - x$. A smaller negative log likelihood indicates a better convergence to the desired image.

A convergence rate comparison was done among the standard EIID and the EDID with different numbers of sub-iterations using the negative log likelihood of each reconstructed image. All algorithms started from the same initial SBP image, as shown in Fig. 4.10 and Fig. 4.11. In the plots, each data point represents one iteration with its completion time and negative log likelihood. The standard EIID is equivalent to EDID with $M = 1$. When an appropriate M is selected, such as 32 in this experiment, the EDID method can converge 10 times faster than the standard EIID method.

4.2.5 Experiments

Several experiments were done to show the performance of EIID. As shown in Fig. 4.12, a Ba-133 point source, a Na-22 point source, and a Cs-137 point source

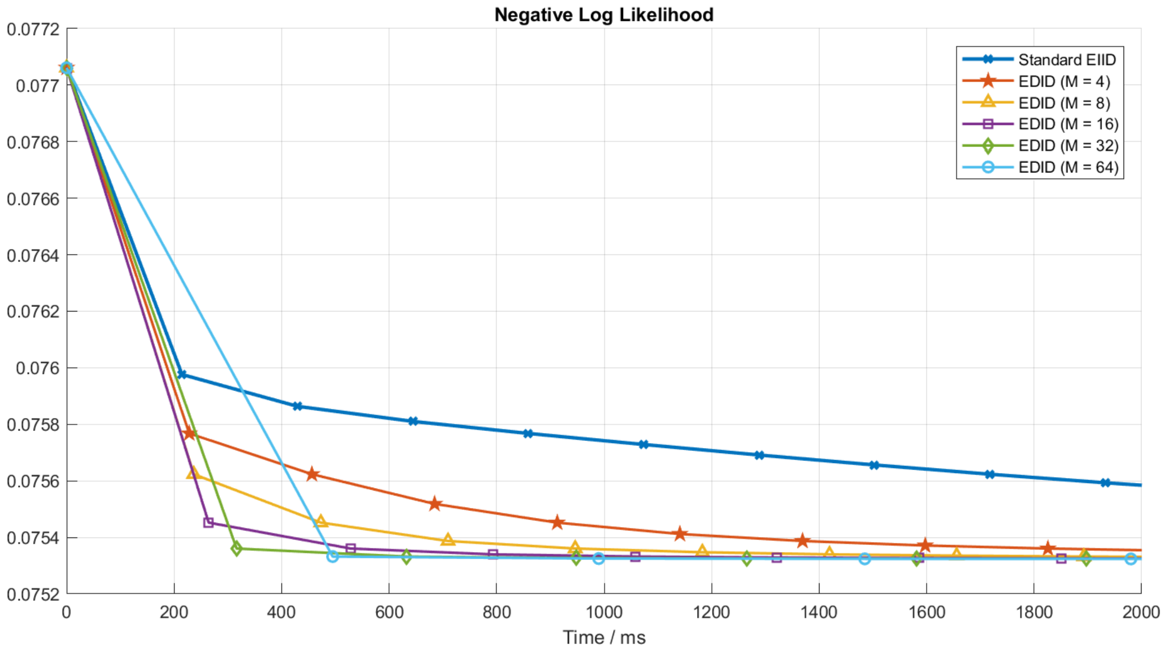


Figure 4.11: Negative log likelihood of the iterations in 2000 ms, showing more details of the early iterations.

were placed around the detector. The raw spectrum had large continuum, especially in the low energy region. After the deconvolution in the energy-imaging integrated space, the directional spectrum was generated, and the three isotopes were clearly identified. The spatial distribution of the sources were also reconstructed and normalized. Standard EIID and EDID gave basically the same results, but EDID was nearly ten times faster than standard EIID.

4.3 Summary

To reconstruct with more accurate statistics, iterative reconstruction algorithms were studied. As a natural extension of the conventional MLEM in every energy bin, the EIID is the ultimate solution for Compton imaging in the energy-imaging integrated space. However the speed of EIID significantly limits its usage, thus EDID

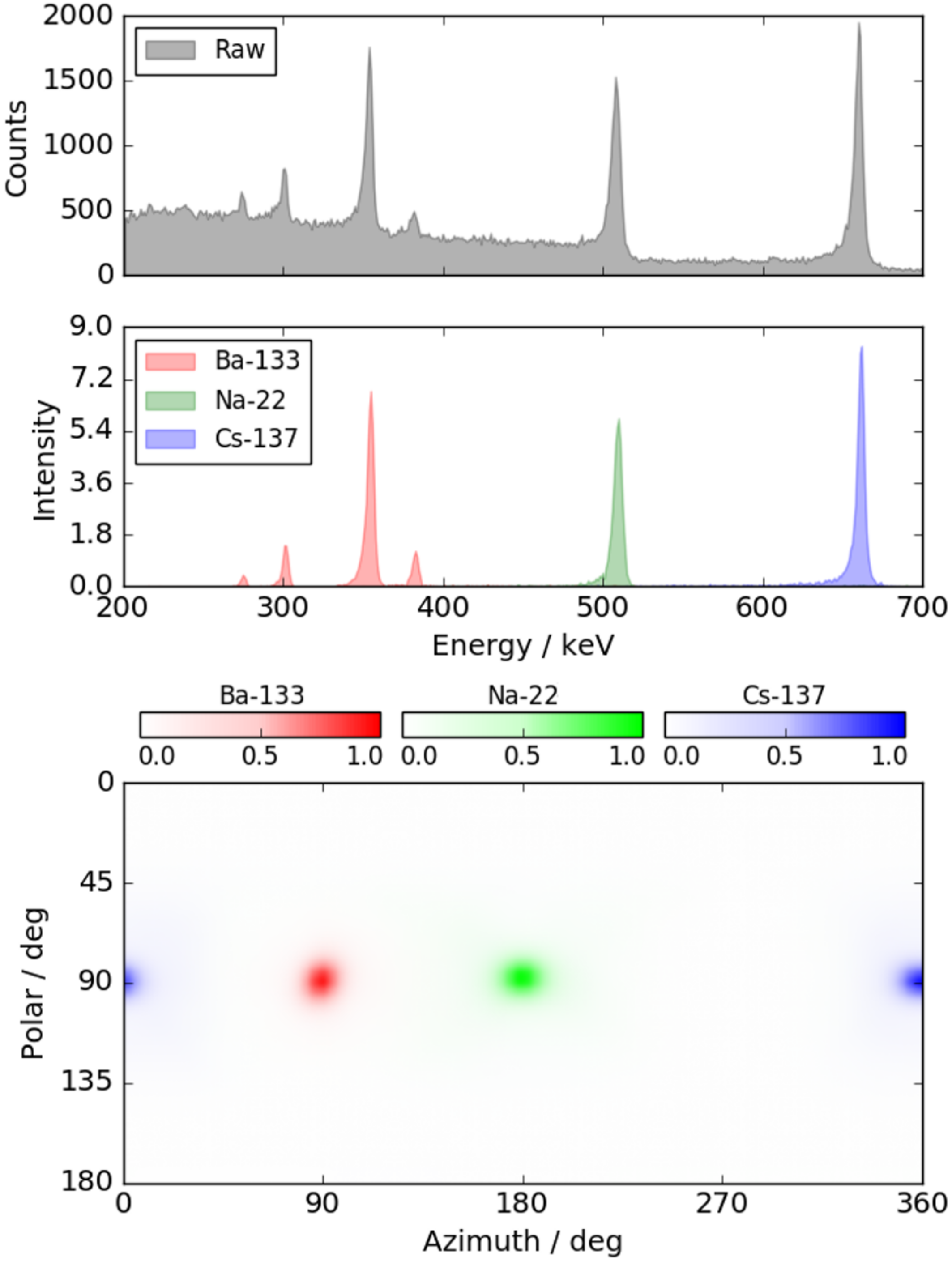


Figure 4.12: Top: Raw spectrum. Middle: Directional spectrum identified as different isotopes. Bottom: Spatial distribution of different isotopes.

has been developed in this thesis to accelerate the convergence rate. The essence of EDID is to use the conditional probability distribution in energy space, such that the images in high-energy bins are firstly reconstructed, and the images in low-energy bins are gradually reconstructed, considering the possible sources in current energy bin and in higher energy bins. The EDID can converge 10 times faster than the EIID, and generate roughly the same images in the energy-imaging integrated space.

CHAPTER V

3D Image Reconstruction Algorithms

The 3D space is a natural extension of the 4π spherical space in Compton imaging. The 4π spherical space cannot distinguish the intensities in the same angular direction. Therefore for a complicated environment, it is not sufficient to reveal the source location only with 2D reconstruction. Thus 3D reconstruction is required to eliminate the ambiguities. Also it is more intuitive to map sources in a 3D space to the real space where human-beings live. So people always have great interest in 3D image reconstruction. Although 3D Compton image reconstructions have been previously studied [76, 77, 78, 79], there are still many misunderstandings to clarify and problems to solve.

A single Compton event is not able to locate the source in a 3D space directly. Hence parallax must be provided, which is the most significant difference between 3D imaging and 4π imaging. This chapter will discuss the model of 3D imaging, solve its major problems, and show the reconstructed results using the real data from a CZT detector mounted on a robot.

5.1 Model of 3D Imaging

5.1.1 Imaging Model

Usually a Cartesian coordinate system is used for a 3D space. Define the 3D Cartesian space as $\mathcal{P}$, then the continuous source distribution is

$$f = f(\boldsymbol{x}) = f(x, y, z), \tag{5.1}$$

where $\boldsymbol{x}$ is the location vector defined in $\mathcal{P}$, x , y and z indicate the three dimensions of the space $\mathcal{P}$. Practically, discrete imaging space is used. Generally define the the indices of voxels as $\mathcal{J} = \{1, 2, ..., J\}$, where the j-th pixel has 3D volume v_j and centers at position $\boldsymbol{x}_j$. The total intensity within a voxel is defined as

$$f[j] = \int\limits_{v_j} f(\boldsymbol{x}) d\boldsymbol{x} \approx |v_j| f(\boldsymbol{x}_j), \tag{5.2}$$

where the approximation is valid if $f(\boldsymbol{x})$ does not change rapidly within v_j, which is usually true if $|v_j|$ is sufficiently small. The 3D space is not necessarily cuboid but can be any arbitrary shape, as long as there are finite number of voxels and the voxels do not intersect.

However, the cuboid 3D space with the orthogonal voxel grid is still most widely used because of its simplicity and memory efficiency. In such case

$$\mathcal{J} = \mathcal{J}_x \times \mathcal{J}_y \times \mathcal{J}_z, \tag{5.3}$$

where $\mathcal{J}_x = \{1, 2, ..., J_x\}$, $\mathcal{J}_y = \{1, 2, ..., J_y\}$, and $\mathcal{J}_z = \{1, 2, ..., J_z\}$.

$$f[j] = f[j_x, j_y, j_z] \approx f(x_{j_x}, y_{j_y}, z_{j_z}) \Delta x \Delta y \Delta z, \tag{5.4}$$

where $j_x \in \mathcal{J}_x$, $j_y \in \mathcal{J}_y$, $j_z \in \mathcal{J}_z$, with fixed voxel dimensions Δx, Δy, and Δz.

The general discrete imaging model is still

$$\bar{\boldsymbol{g}} = \boldsymbol{T}\boldsymbol{f}, \tag{5.5}$$

where $\boldsymbol{f}$ is the length-J image vector, $\bar{\boldsymbol{g}}$ is the length-I expected observation vector, and $\boldsymbol{T}$ is the $I \times J$ system matrix with each row as a Compton back-projection cone.

5.1.2 System Model

The system model of 3D imaging needs to consider the absolute location of each interactions in one event. Compared to the 4π imaging, which only need to consider the relative location of interactions in one event, 3D imaging requires more information and hence more complicated. Also because the number of voxels in a 3D space is usually much larger than the number of pixels in a 4π spherical space, 3D imaging is more computationally intensive.

For a sufficiently small voxel far away from the detector, it is reasonable to assume a point voxel and a point detector. Also Gaussian radius blur is assumed for the Compton cones to simplify the analysis. The system matrix $\boldsymbol{T}$ has elements

$$t_{ij} = \mathbb{P}(\text{event } i \mid \text{photon from voxel } j). \tag{5.6}$$

Denote the distance between the voxel j and the detector as r,

$$t_{ij} \propto \frac{1}{r^2}, \tag{5.7}$$

because for a photon isotropically emitted from voxel j, the probability that the photon will intersect with the detector is $\frac{\sigma_c}{4\pi r^2}$, where σ_c is the cross section of the detector.

The event i has a Compton cone with scattering angle θ. Define the angle between the lever arm of the Compton ring and voxel-vertex line as α,

$$\beta = |\alpha - \theta|, \tag{5.8}$$

where β represents the angular distance from the voxel to the Compton cone, and

$$t_{ij} \propto e^{-\frac{\beta^2}{2\sigma^2}}, \tag{5.9}$$

which is similar to the Gaussian radius blur used in the 4π imaging, and σ is the standard deviation of the blurring kernel, as shown in Fig. 5.1. The unnormalized probability can be calculated as

$$t_{ij} = \frac{1}{r_{ij}^2} e^{-\frac{|\alpha_{ij} - \theta_i|^2}{2\sigma_i^2}}, \tag{5.10}$$

where r_{ij} and α_{ij} are determined by i and j, θ_i and σ_i are determined by i. Fig. 5.2 and Fig. 5.3 show a probability distribution of a Compton scattering event in a 3D space. The r^{-2} effect is shown in Fig. 5.2, and the radius blur is shown in Fig. 5.3.

It is also possible to include the attenuation and other effects in the system model. However, the distance parameter and the Compton cone dominate the information containing in a Compton event, hence a simplified model is used in our research. The more accurate model is left to study in the future.

5.1.3 Sensitivity

In many cases, multiple detectors or moving detectors are used to provide sufficient parallax in a 3D space. In the ideal point detector model, the total sensitivity can be calculated by the convolution between the stationary sensitivity and the integral of detector's position over time.

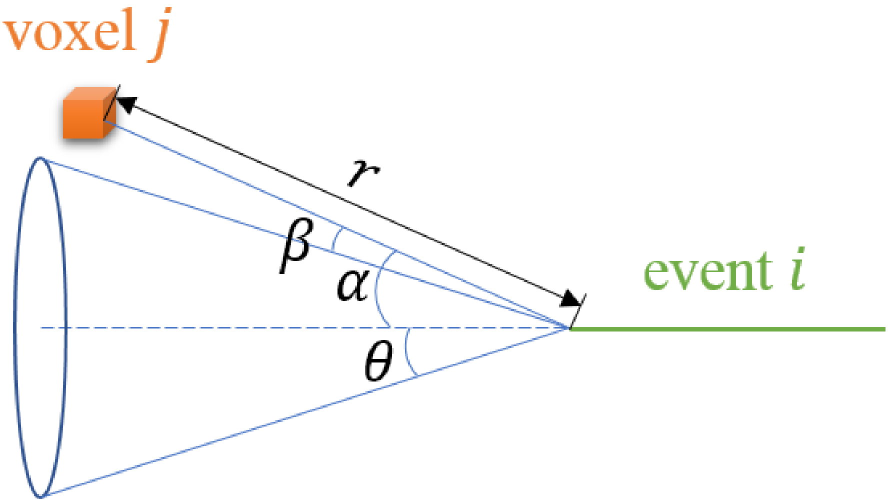


Figure 5.1: The angular distance from the voxel to the Compton cone.

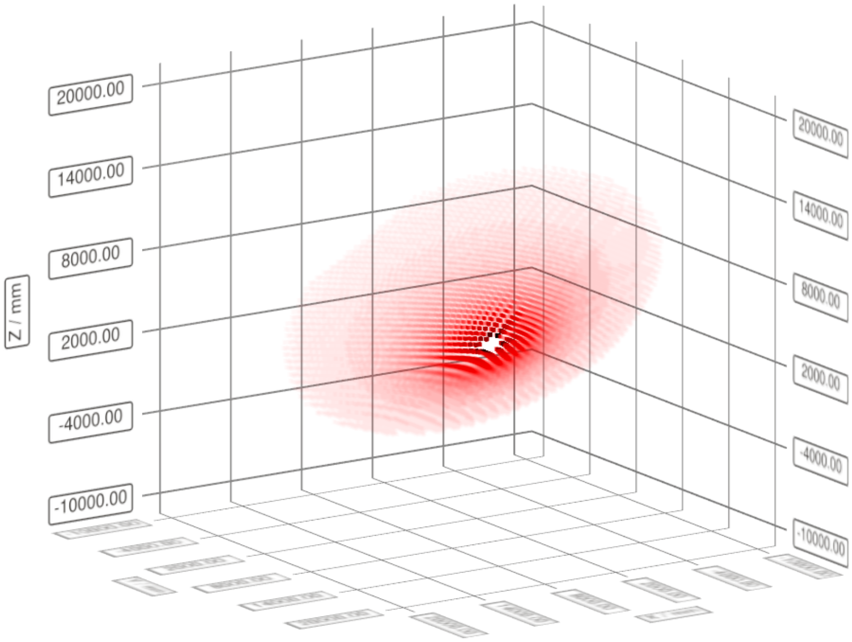


Figure 5.2: Probability distribution of a Compton scattering event in a 3D space. The vertex is removed to avoid extremely large value.

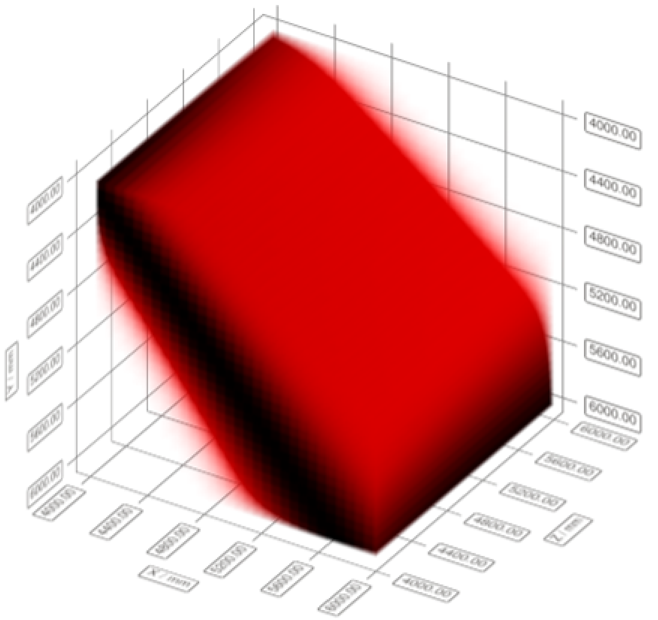


Figure 5.3: A small piece of the Compton cone showing the radius blur.

For a fixed energy window, the stationary sensitivity of a point detector is

$$b(\boldsymbol{x}) = \frac{\sigma_c \epsilon}{|\boldsymbol{x}|^2}, \tag{5.11}$$

where σ_c is the cross section of the detector, and ϵ is the absorption efficiency of the detector. The stationary sensitivity map looks like a ball cloud as shown in the left of Fig. 5.4. Since the stationary sensitivity is isotropic, the FFT is used to compute the total sensitivity quickly.

Denote the path weight of the detector d at time t is $p_d(\boldsymbol{x}, t)$, then the total path weight of all D detectors is

$$p(\boldsymbol{x}) = \sum_{d=1}^{D} \int_{0}^{T} p_d(\boldsymbol{x}, t) dt, \tag{5.12}$$

where the measurement starts at time 0 and ends at time T. In real measurement, the total path weight is the 3D histogram of the detector locations in different time.

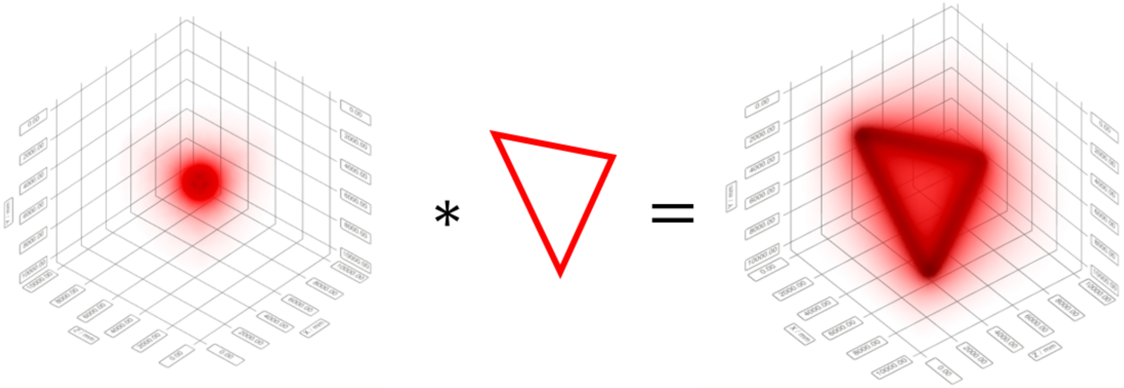


Figure 5.4: The stationary sensitivity (left) convolve with the total path weight (middle) and generate the total sensitivity (right).

The total sensitivity is the convolution between the stationary sensitivity and the total path weight

$$s(\boldsymbol{x}) = b(\boldsymbol{x}) * p(\boldsymbol{x}), \tag{5.13}$$

where the FT pairs are

$$\begin{cases} s(\boldsymbol{x}) \longleftrightarrow S(\boldsymbol{\omega}) \\ b(\boldsymbol{x}) \longleftrightarrow B(\boldsymbol{\omega}) \\ p(\boldsymbol{x}) \longleftrightarrow P(\boldsymbol{\omega}) \end{cases}, \tag{5.14}$$

and

$$S(\boldsymbol{\omega}) = B(\boldsymbol{\omega}) \times P(\boldsymbol{\omega}). \tag{5.15}$$

FFT can be used to accelerate the computation. The convolution is shown in Fig. 5.4.

5.2 Solving 3D Imaging Problems

Theoretically 3D imaging problems and 2D imaging problems can be solved using same methods. However, there are unique issues in 3D imaging, and they need to be treated carefully. First of all, since the imaging space is much larger for 3D imaging, more events are required to provide sufficient statistics, and the reconstruction is more computationally intensive. Secondly, serious corner artifacts may appear if the imaging space is not selected properly.

5.2.1 Online Updating Algorithm

The 3D image reconstruction is computationally intensive, because of its extremely large imaging space and huge number of events. However, real-time or pseudo real-time 3D imaging is required in many scenarios. Thus the online updating algorithm is proposed to do image reconstruction in real-time. There are two concepts in the online updating algorithm: batching and weighting.

5.2.1.1 Batching

The observation data stream comes in with time, and more accumulated data leads to better statistics. However, it is inefficient to wait for the data to accumulate. Instead, reconstructions can be conducted whenever there is a new batch of data coming in.

The algorithm has a similar essential idea as the ordered-subsets methods [75]. However, the subsets of data are not ordered by viewing angle but ordered by timestamps. Another difference is that traditional ordered-subsets methods use one subset in each iteration, but the online updating algorithm uses all accumulated dataset in each iteration to achieve as best statistics as possible. In each iteration, K sub-iterations are done to help converge. The batching-only version of the online updating algorithm is

```
Require: $\boldsymbol{f}^{(0)}$: Non-zero initial image vector
Require: $\{\boldsymbol{T}_1, \boldsymbol{T}_2, ..., \boldsymbol{T}_N\}$: batches of system matrix ordered by time
  for $n = 1$ to $N$ do
```

```
    $\boldsymbol{T} = [\boldsymbol{T}_1; \boldsymbol{T}_2; ...; \boldsymbol{T}_n]$
    for $k = 1$ to $K$ do
        $\boldsymbol{f}^{(n-1+\frac{k}{K})} \leftarrow \mathrm{MLEM}(\boldsymbol{f}^{((n-1+\frac{k-1}{K}))}, \boldsymbol{T})$
    end for
end for
return $\boldsymbol{f}^{(N)}$ (Resulting image)
```

5.2.1.2 Weighting

In the batching-only version, older datasets are iterated more times than newer datasets, which may cause the problem that the new data is not updated into the image timely. Another problem is that the algorithm becomes slower and slower as more and more data is accumulated. To solve these problems, weighting method is introduced, which defines the life cycle and weight of each event.

The weighting factor should be a function of time and event quality. For event i, its weighting factor w_i at time t is

$$w_i(t) = w_i(t - t_i, l_i) = w_i(\Delta t_i, l_i), \tag{5.16}$$

where t_i is the timestamp of the event i, $\Delta t_i \triangleq t - t_i$, and l_i is the lever arm length of the event i. As a naive approach, the lever arm length is chosen as the representation of event quality for simplicity, and there could be more accurate ways to represent the event quality. Typically the weighting factor should penalize the events with long life cycles and short lever arms, as shown in Fig. 5.5. Since the weighting factor is monotonically decreasing with Δt, when an event gets a weighting factor 0, it can be removed from the dataset forever.

The online updating algorithm with batching and weighting is

Require: $\boldsymbol{f}^{(0)}$: Non-zero initial image vector

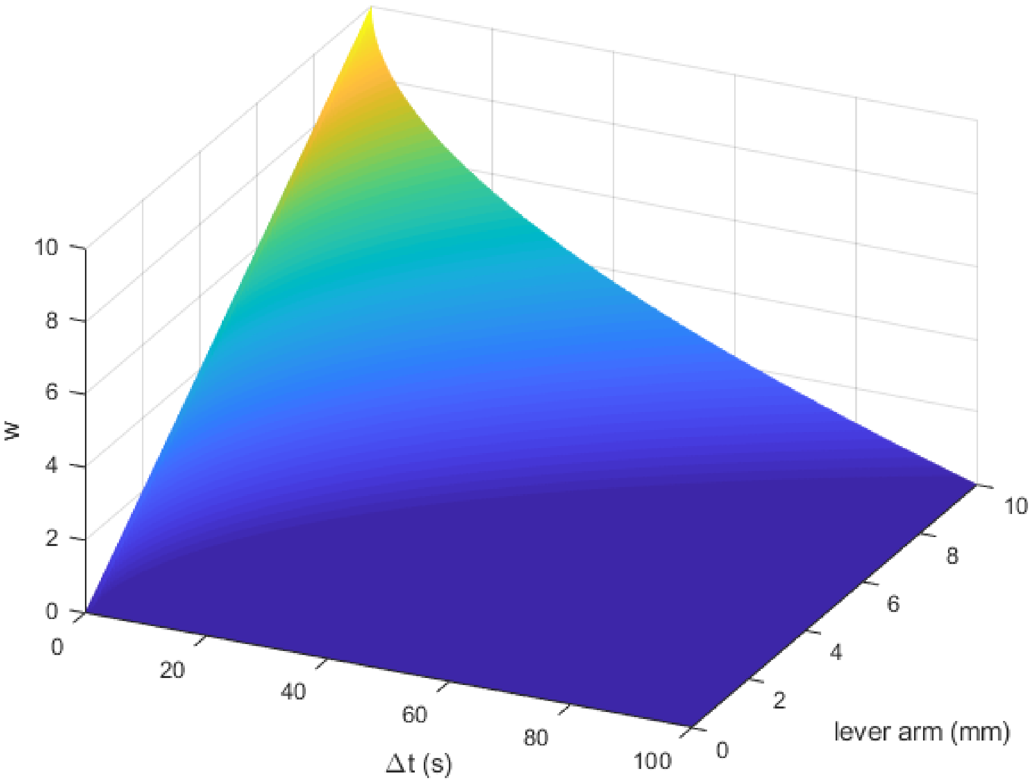


Figure 5.5: A possible weighting factor function. The fresher events with longer lever arm have larger weighting factors.

```
Require: $\{\boldsymbol{T}_1, \boldsymbol{T}_2, ..., \boldsymbol{T}_N\}$: batches of system matrix ordered by time
for $n = 1$ to $N$ do
    $\boldsymbol{T} = [\boldsymbol{T}_1; \boldsymbol{T}_2; ...; \boldsymbol{T}_n]$
    $\boldsymbol{w} = [\boldsymbol{w}(n-1, \boldsymbol{T}_1); \boldsymbol{w}(n-2, \boldsymbol{T}_2); ..., \boldsymbol{w}(0, \boldsymbol{T}_n)]$
    for $k = 1$ to $K$ do
        $\boldsymbol{f}^{(n-1+\frac{k}{K})} \leftarrow \mathrm{MLEM}(\boldsymbol{f}^{((n-1+\frac{k-1}{K}))}, \boldsymbol{T}, \boldsymbol{w})$
    end for
end for
return $\boldsymbol{f}^{(N)}$ (Resulting image)
```

5.2.2 Corner Artifacts

Corner artifacts were first observed in reconstructions in a spherical space. If the field of view is not 4π, then the image is very likely to converge to the corners of the imaging space, which are strong artifacts that affect the detection of real sources. The image reconstructions in a 3D space also suffer from the corner artifacts. As shown in Fig. 5.6, although the point source is at the center of the imaging space, the corners of the space become hottest after several iterations, leading to an unexpected image.

An explanation for the corner artifacts in a 3D imaging space is that the detector is out of the image space or the image space is too sparse, such that the Compton cone is not fully represented. If the estimated ring width uncertainty is smaller than the true ring width uncertainty, the back-projected cone may not well intersect with the source position. When the detector is in the image space, the whole cone will always be preserved and the weights of wrong pixels are trivial. However, when the detector is out of the image space, the cone may have some extremely large weights at the corners / edges of the image space. This contributes to the corner artifacts.

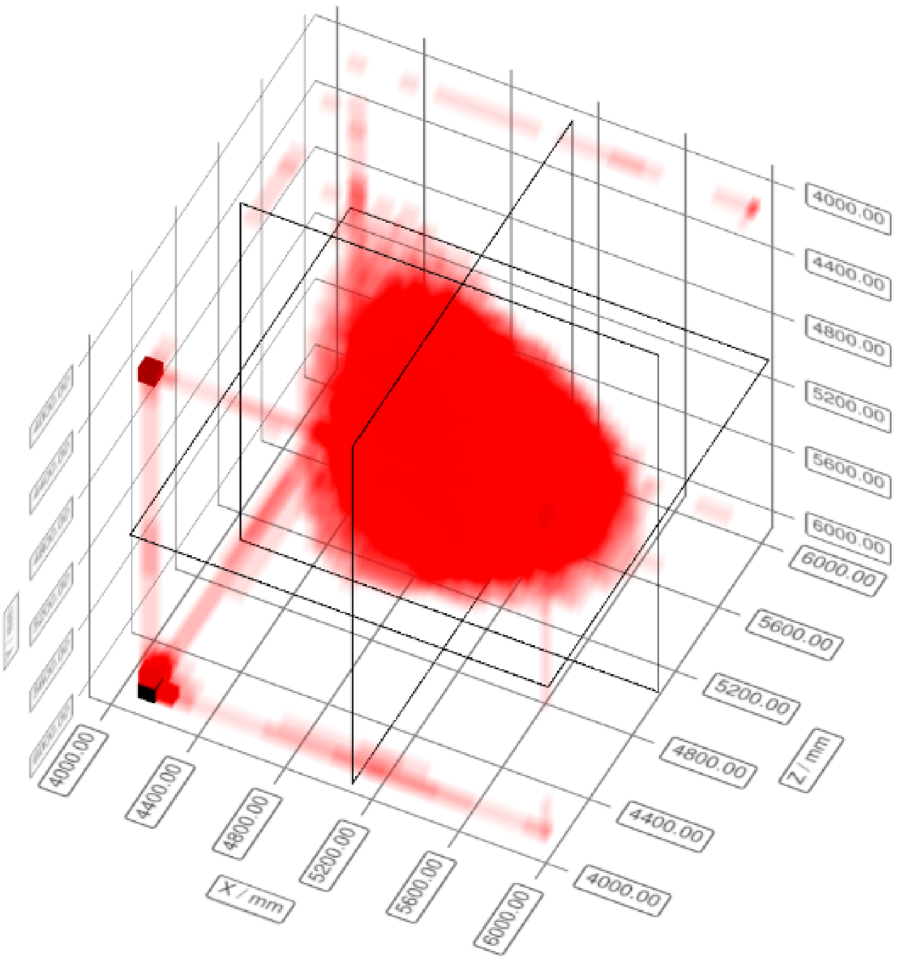


Figure 5.6: Corner artifacts in a 3D imaging space.

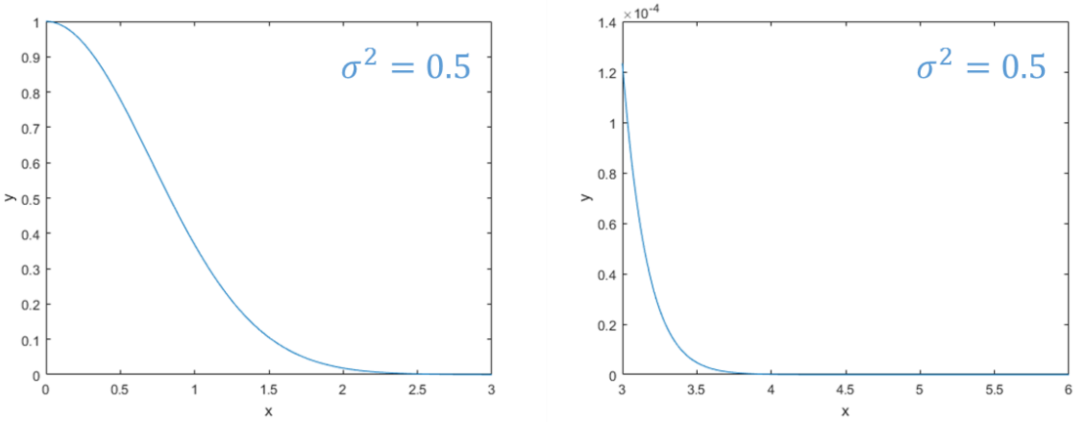


Figure 5.7: Gaussian distribution with $\sigma^2 = 0.5$. Left: x axis from 0 to 3. Right: x axis from 3 to 6.

The Compton cone is blurred with a Gaussian kernel

$$y = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}, \tag{5.17}$$

where σ is the standard deviation, and

$$|\frac{y'}{y}| = \frac{x}{\sigma^2}, \tag{5.18}$$

which means that the relative decreasing speed is faster when x goes larger, as shown in Fig. 5.7.

Fig. 5.8 shows several Compton cones that only partially intersect with the imaging space and have large values at corners and edges, contributing most to the corner artifacts.

A simple way to avoid the corner artifacts is to extend the imaging space, such that the detector is within it and the Compton cones are fully intersect with it. If this is not possible, one may add penalty in the cost function and suppress the partial Compton cones.

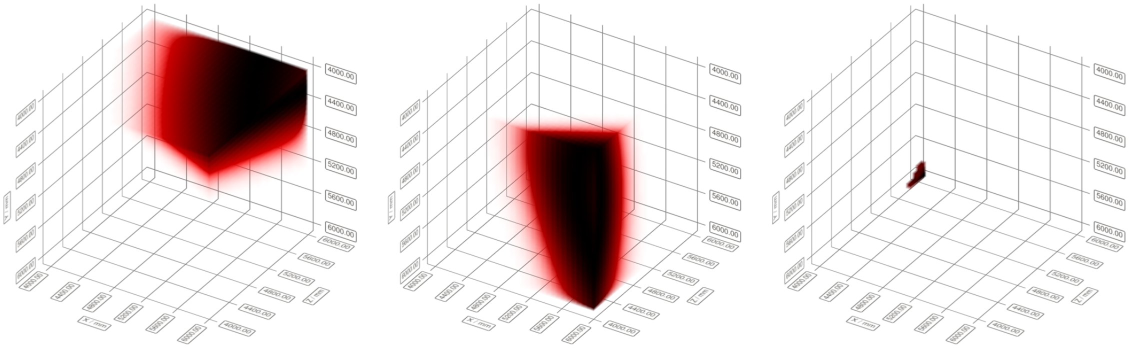


Figure 5.8: The Compton cones only partially intersect with the imaging space and have large values at corners and edges.

5.3 Experiments Using Moving Detectors

Joint experiments were done with the group from Carnegie Mellon University. A Polaris-H [80] detector with one $20 \times 20 \times 15$ cm^3 CZT crystal was mounted on a Lilred robot, as shown in Fig. 7. The robot can be controlled by a joystick or do path search autonomously. The position and orientation information is provided by the robot, while the radiation information is provided by the CZT detector. Data packets were received and combined by the robot in real-time, then saved for further reconstruction while the robot is moving. Several measurements with various settings were done, and 3D source images were reconstructed using the collected data.

Fig. 5.10 shows the reconstruction of four sources in a small room: Ba-133 (purple), Cs-137 (red), Co-60 (orange), and Na-22 (green). The path of the robot is shown as the red curve. The room is shown as a 3D point cloud map in green. The true source locations are marked as spheres, and the reconstructed source locations are marked as cubes. All three sources are successfully reconstructed with no more than 10 cm uncertainty.

Fig. 5.11 shows the reconstruction of three sources in a large room. Co-60 (green) and Na-22 (red) are successfully detected, but the Cs-137 data has too poor statistics and is not able to reveal the source location.

Figure 5.9: The Lilred robot with a Polaris-H mounted on its shoulder.

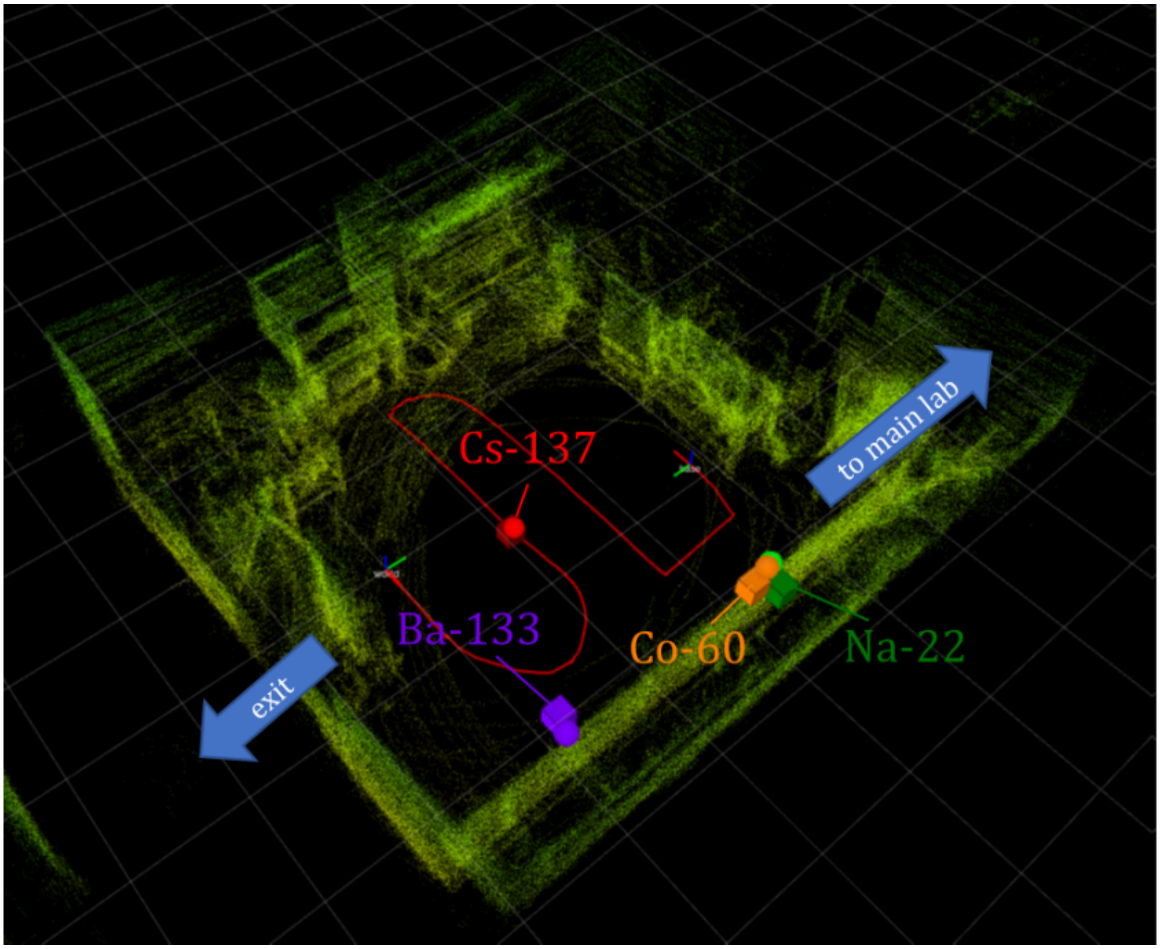


Figure 5.10: Reconstruction of four sources in a small room.

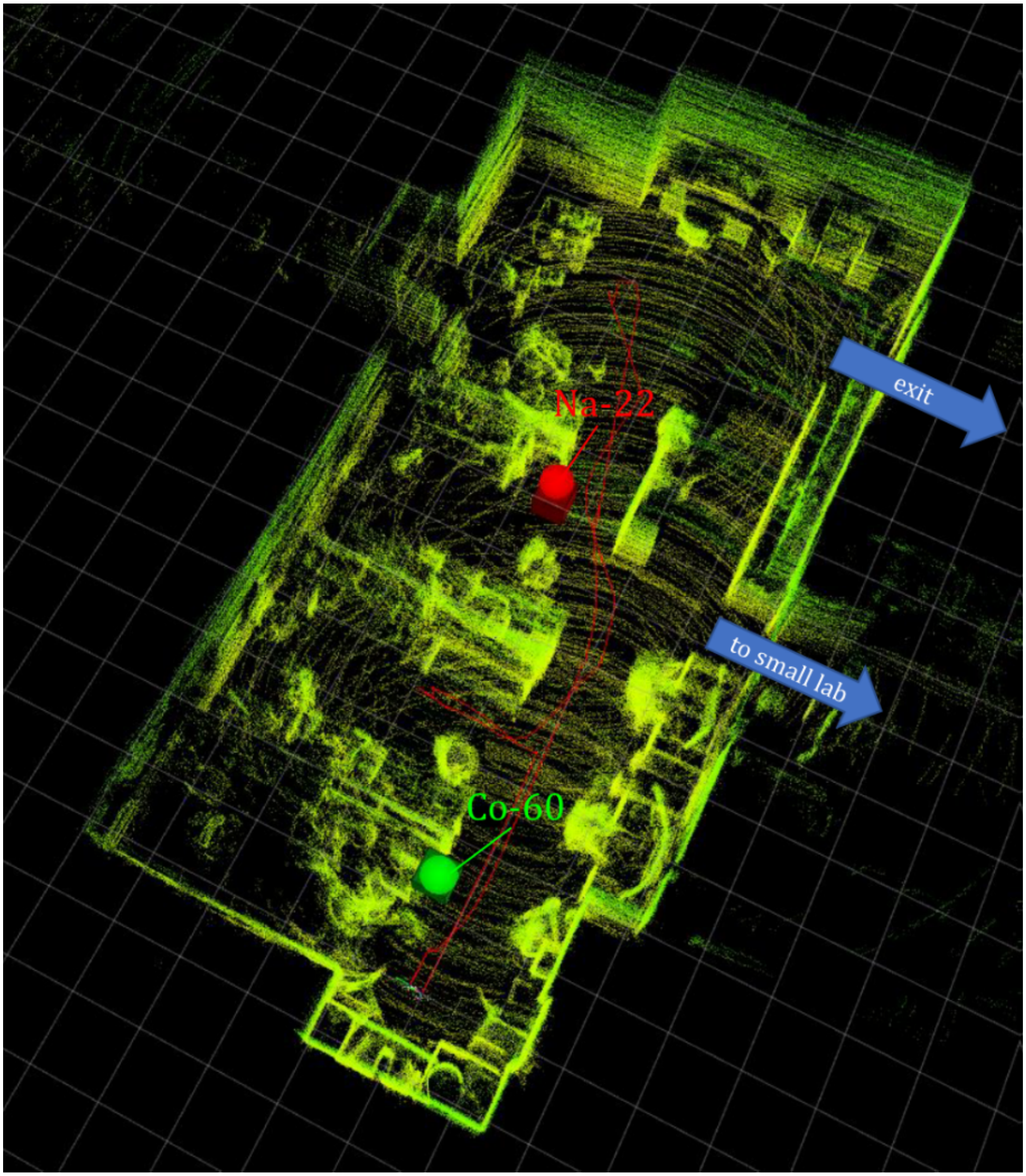


Figure 5.11: Reconstruction of three sources in a large room.

5.4 Summary

Current work is trending to 3D image reconstruction. Compared to 4π spherical imaging, 3D imaging has an imaging space with a huge number of voxels, and relies on the parallax of different viewing angles to the sources, hence there are more challenges in reconstruction speed. To implement 3D imaging in pseudo real-time, the online updating algorithm has been proposed. By batching the incoming data the algorithm will update the image gradually with time. By weighting the events by quality and time, the algorithm always has a memory of previous data and is able to generate the image in the whole measurement period. A joint experiment has shown the feasibility and capability of 3D imaging using moving detectors.

However, there is still a lot of work to be done in 3D imaging. Current sensitivity assumes an ideal point detector, which could be improved by integrating the attenuation into the sensitivity map. The optical camera could also help by excluding some voxels, such as air, thus reducing the size of the imaging space. In addition, it could be very interesting to combine the source detection and imaging with the autonomous control, plan the searching route based on known information and increase searching efficiency.

CHAPTER VI

UMIS Development

6.1 Introduction

The research and applications of radiation imaging require tools with more and more capabilities. For example, in some scenarios people need extremely high reconstruction speed to support real-time imaging, in some scenarios people attempt to compare different reconstruction algorithms using the same raw data, and in some scenarios people want to develop their own algorithms by adapting existing code. Thus a well-designed imaging software is necessary.

Previously UMImaging [81] was developed as the toolkit for research, which attempts to provide a framework for the analysis in a general manner that will allow use by different detector designs and applications. However, UMImaging has a top-down design that integrates all libraries in its main structure, which increases the difficulty of extension and reimplementation. Also some functions in UMImaging are over-designed such that it is very complicated to achieve fast development iteration.

To conquer these problems, University of Michigan Imaging System (UMIS) was designed and developed. The principle is to keep the code fast, simple, reusable, and suitable for research. Unlike UMImaging, UMIS has a bottom-up design, and focuses on the data flow instead of the code components. To track the development history and help collaboration, Git was used for version control. This chapter will introduce

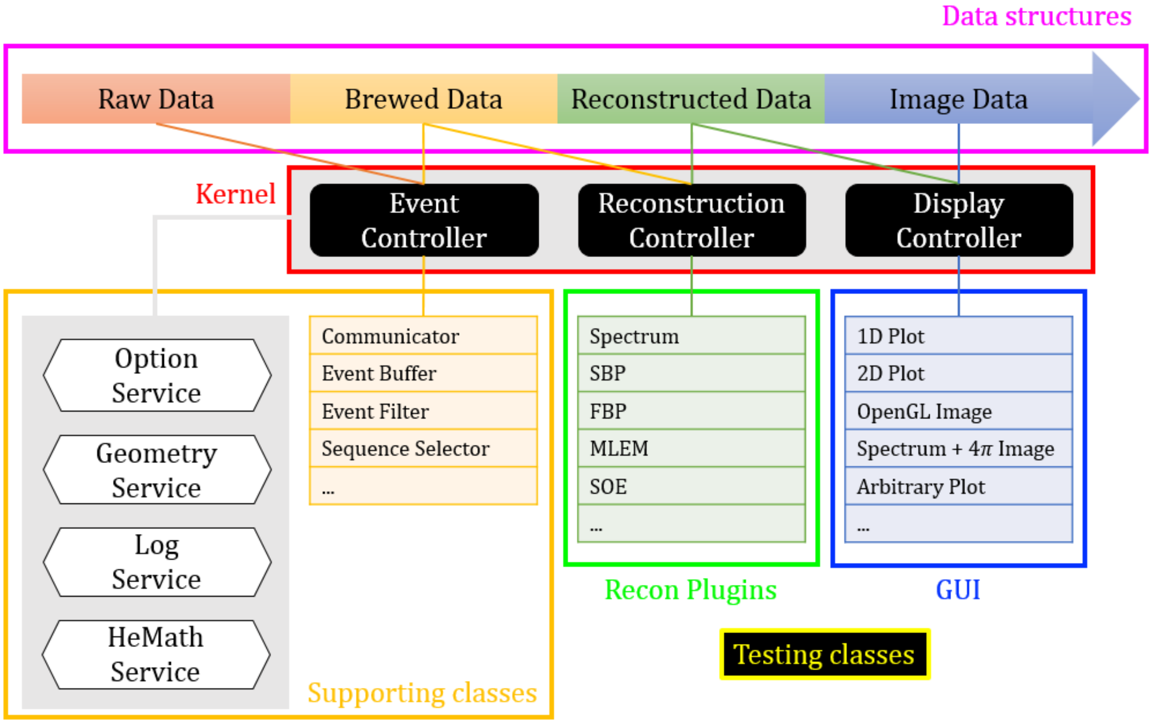


Figure 6.1: Code architecture of UMIS.

the architecture, libraries and kernel of the code, and discuss some reconstruction algorithms implemented, which are followed by example usages.

6.2 Architecture Overview

The code architecture is dataflow driven, as shown in Fig. 6.1. Raw data is passed from the detector side and turned into brewed data, reconstructed data and image data sequentially. The whole process is controlled by the event controller, reconstruction controller display controller, which have their own threads and form the kernel of the program. There are also several services providing global functions. To improve robustness, every class has its own testing class. The testing classes are used to examine the time complexity of algorithms, the boundary conditions and the responses to error inputs and any internal problems.

A detailed architecture is shown in Fig. 6.2. Gray cells represent classes. Anything under a class cell is contained by it. The table within a cell represents a subclass derived from the base class. The keywords are explained in Table 6.1.

Keyword	Explanation
thread	This class has its own thread.
data	A memory consuming dataset.
ptr	A pointer.
cptr	A constant pointer.
var	A variable instance within a class.
func	A function within a class.
virtual	A virtual function which will be implemented by a subclass.

Table 6.1: Explanation of the keywords in Fig. 6.2.

6.3 Code Kernel

Event controller, reconstruction controller and display controller are the kernel of the code. Each controller has its own thread and communicates with each other by signal/slot mechanism provided by Qt. Actually each controller does not know the existence of other controllers and information can only be passed indirectly, which is designed to keep these controllers decoupled. Thus the code architecture is clearer.

As shown in Fig. 6.3, user commands are sent to event controller. After the event controller brews the data, a ready signal will trigger the reconstruction controller. When the data is reconstructed or any image settings are changed a ready signal will trigger the display controller, then the image will be updated. Note that event controller has full access to raw data and brewed data; reconstruction controller has full access to reconstructed data but only read access to brewed data; display controller has full access to image data but only read access to reconstructed data.

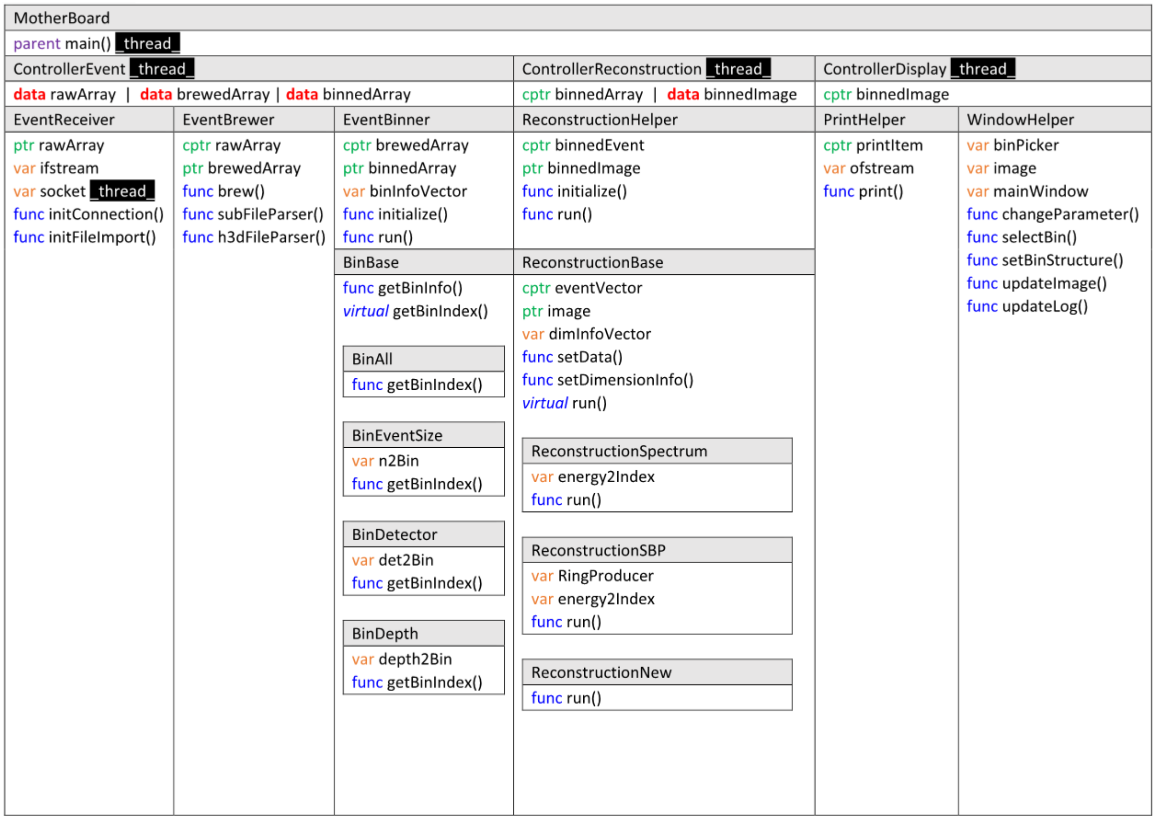


Figure 6.2: Detailed code architecture of UMIS.

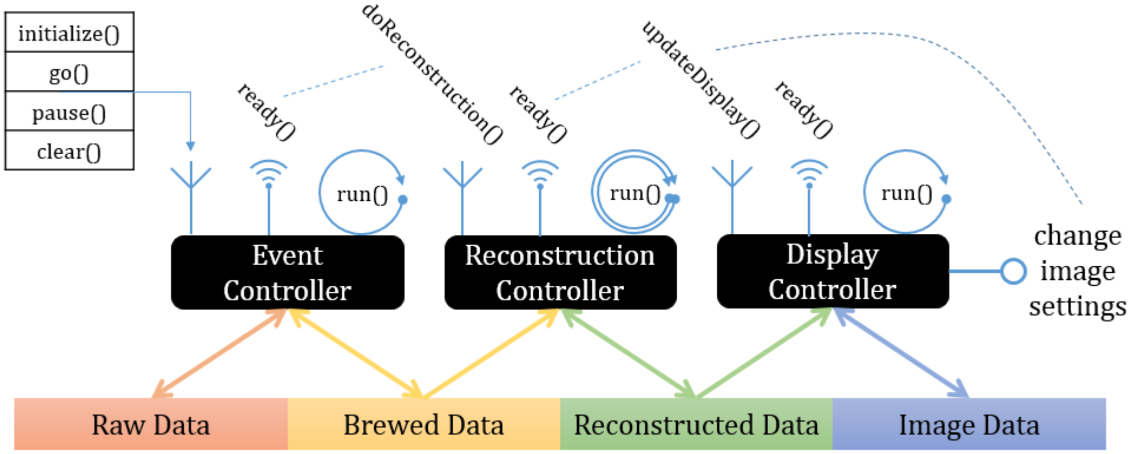


Figure 6.3: Signal/slot mechanism in the kernel.

6.3.1 Motherboard

The mother board is the entrance and carrier of the code. The motherboard is initialized at the start of the program, then loads the options and connects all of the controllers.

6.3.2 Event Controller

The event controller receives data via Internet or reads data from disks, then cleans the data before sending it to the next stage.

The incoming data is first received by the event receiver. If the data comes from Internet, the event receiver will establish the TCP/IP connection and handle any possible errors. Each data package is processed immediately and passed to the next stage. If the data comes from disks, the event receiver will treat the data file as one data package, read all of them, and pass it to the next stage. If any error happens, such as losing connection or invalid file, the event controller will stop data feeding and send a message to the user.

The next stage is the event brewer. When the raw data is passed from the event receiver to the event brewer, it will be parsed with the given method and turned to brewed data. For example, some raw data only contains detector numbers and local coordinates of events, so the event brewer needs to transfer the local coordinate system to the global coordinate system according to the predefined geometry file. After the event brewer, the data becomes standard format, where each event is made of one interaction or several interactions, and each interaction has its global coordinate, deposited energy, time, and other useful information.

Finally the data is passed to the event purifier. In some scenarios not all the data is necessary, so only a subset of data will be used in reconstruction. The condition is set in the option file, read by the event purifier, and is used to filter the data. Also the event purifier will check the integrity and validity of data, remove corrupted events, and send a message to the user if any error occurs.

6.3.3 Reconstruction Controller

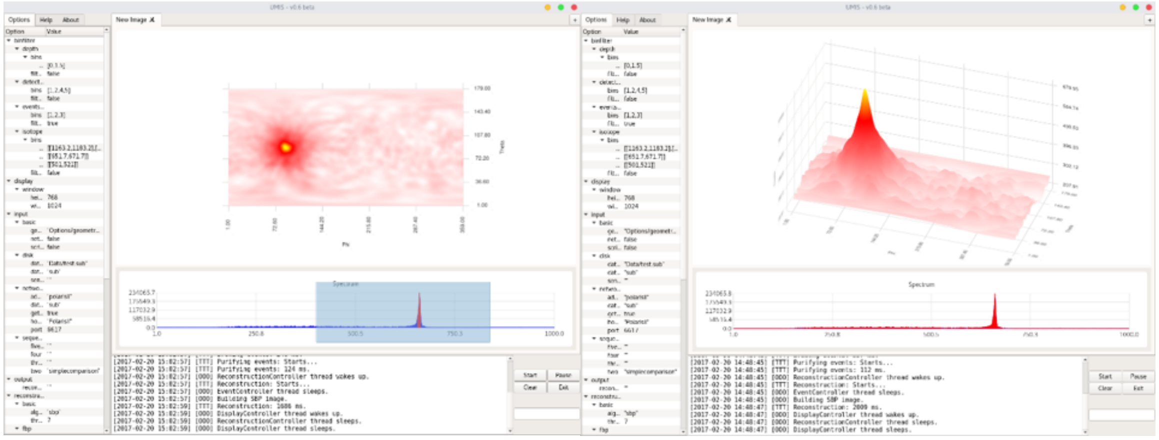
The reconstruction controller is the core of the code, which controls the reconstruction algorithm and handles its input data and output data.

There are three layers between data and the algorithm. The first layer is the reconstruction controller itself. It grabs the data from the event controller, instantiates the output datacube, and communicates with the display controller. The second layer is the reconstruction helper. It initializes the algorithm instance based on the option file, coordinates the data references, and controls the algorithm. The third layer is the algorithm. The protocol between algorithm and data is well-defined, and the algorithm only needs to deal with its own work. The purpose of the three-layer design is to decouple the code and increase the reusability.

6.3.4 Display Controller

The display controller reads reconstructed images from the reconstruction controller and displays or prints the images to users with the given methods. It also controls the Graphical User Interface (GUI), and sends the commands from the GUI to the motherboard.

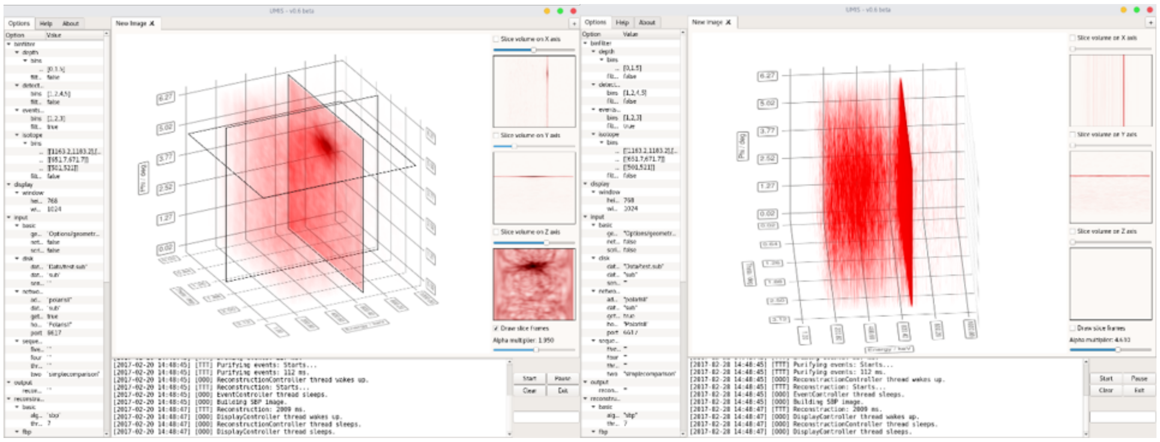
The GUI has been developed, tested, and integrated into the code. In the main window, the left column shows options and selectable bins. The bottom area contains the logs and a simple control panel. The image and the spectrum are shown as a tab. One can add more tabs to visualize the datacube. Currently it supports spectrum, surface view, and volume view of datacube. Options are shown at the left of the main window, logs are shown at the bottom, and the images are shown at the center as tabs. See Fig. 6.4 and Fig. 6.5 for examples.



```
[2017-02-20 15:02:57] [TTT] Purifying events: Starts...
[2017-02-20 15:02:57] [TTT] Purifying events: 124 ms.
[2017-02-20 15:02:57] [000] ReconstructionController thread wakes up.
[2017-02-20 15:02:57] [TTT] Reconstruction: Starts...
[2017-02-20 15:02:57] [000] EventController thread sleeps.
[2017-02-20 15:02:57] [000] Building SBP image.
[2017-02-20 15:02:59] [TTT] Reconstruction: 1686 ms.
[2017-02-20 15:02:59] [000] DisplayController thread wakes up.
[2017-02-20 15:02:59] [000] ReconstructionController thread sleeps.
[2017-02-20 15:02:59] [000] DisplayController thread sleeps.
```

```
[2017-02-20 14:48:45] [TTT] Purifying events: Starts...
[2017-02-20 14:48:45] [TTT] Purifying events: 112 ms.
[2017-02-20 14:48:45] [000] ReconstructionController thread wakes up.
[2017-02-20 14:48:45] [TTT] Reconstruction: Starts...
[2017-02-20 14:48:45] [000] EventController thread sleeps.
[2017-02-20 14:48:45] [000] Building SBP image.
[2017-02-20 14:48:47] [TTT] Reconstruction: 2009 ms.
[2017-02-20 14:48:47] [000] DisplayController thread wakes up.
[2017-02-20 14:48:47] [000] ReconstructionController thread sleeps.
[2017-02-20 14:48:47] [000] DisplayController thread sleeps.
```

Figure 6.4: Left: Spectrum and surface view. Right: Tilted surface view. One can select different energy windows to show the sum of image slices.



```
[2017-02-20 14:48:45] [TTT] Purifying events: Starts...
[2017-02-20 14:48:45] [TTT] Purifying events: 112 ms.
[2017-02-20 14:48:45] [000] ReconstructionController thread wakes up.
[2017-02-20 14:48:45] [TTT] Reconstruction: Starts...
[2017-02-20 14:48:45] [000] EventController thread sleeps.
[2017-02-20 14:48:45] [000] Building SBP image.
[2017-02-20 14:48:47] [TTT] Reconstruction: 2009 ms.
[2017-02-20 14:48:47] [000] DisplayController thread wakes up.
[2017-02-20 14:48:47] [000] ReconstructionController thread sleeps.
[2017-02-20 14:48:47] [000] DisplayController thread sleeps.
```

```
[2017-02-28 14:48:45] [TTT] Purifying events: Starts...
[2017-02-28 14:48:45] [TTT] Purifying events: 112 ms.
[2017-02-28 14:48:45] [000] ReconstructionController thread wakes up.
[2017-02-28 14:48:45] [TTT] Reconstruction: Starts...
[2017-02-28 14:48:45] [000] EventController thread sleeps.
[2017-02-28 14:48:45] [000] Building SBP image.
[2017-02-28 14:48:47] [TTT] Reconstruction: 2009 ms.
[2017-02-28 14:48:47] [000] DisplayController thread wakes up.
[2017-02-28 14:48:47] [000] ReconstructionController thread sleeps.
[2017-02-28 14:48:47] [000] DisplayController thread sleeps.
```

Figure 6.5: Left: volume view. Right: volume view with changed transparency. One can show the datacube in 3D space and get slices in all dimensions. The slices are shown on the right of the 3D image, or can be displayed directly in 3D space. To see the details inside the cube or show shapes more clearly, one can modify the alpha multiplier to change transparency.

6.4 Data Structure

There are several major data structures used in reconstruction.

6.4.1 EventArray

The EventArray is used to keep events. The events are stored as a vector. Initially all events are sorted by time created, and are capable to be sorted on any given parameter as long as the parameter is sortable. Quick sort is used, which provides an average time complexity $O(n \log n)$, where n is the number of events.

6.4.2 DenseData

Most of the reconstructions will generate an nD datacube, which has orthogonal grid and should be stored in the DenseData. Technically the data is stored as a 1D vector in memory, but we should be able to access data by its nD coordinate or 1D index, as shown in Fig. 6.6. We have

$$\text{index} = \sum_{d=0}^{D-1} n_d \prod_{i=1}^{d-1} N_i, \tag{6.1}$$

where D is the number of dimensions, N_i is the number of samples in dimension i, and n_d is the coordinate in dimension d. Obviously it takes $O(D)$ time complexity to do this transform.

Slicing and projection are very important functions required by datacube. Theoretically it can be done in $O(m)$, where m is the number of values involved in the operation, as shown in Fig. 6.7 and Fig. 6.8.

6.4.3 SparseData

In many cases we need to store data sparsely. In SparseData, data points are stored as coordinate-value pairs $(x, y, z, ...) \to v$. SparseData can be easily converted

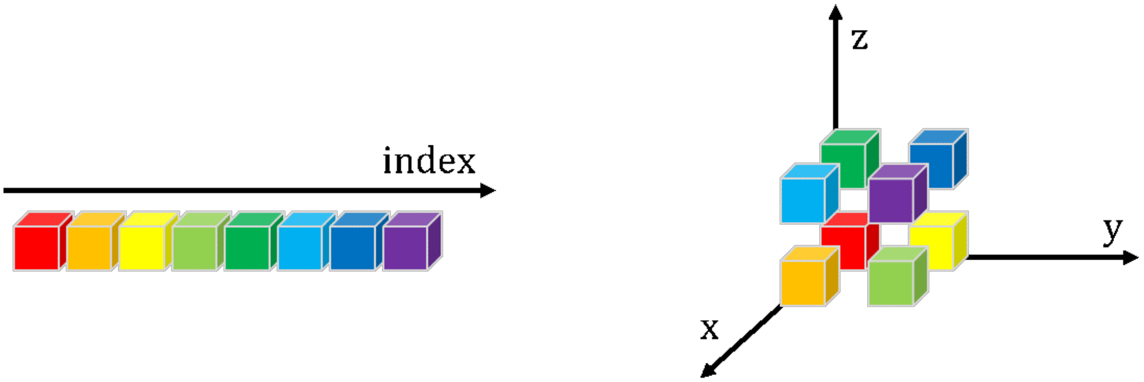


Figure 6.6: Left: Data is stored as a 1D vector in memory, which can be accessed by its index. Right: Data is represented as an nD cube, which can be accessed by its coordinate.

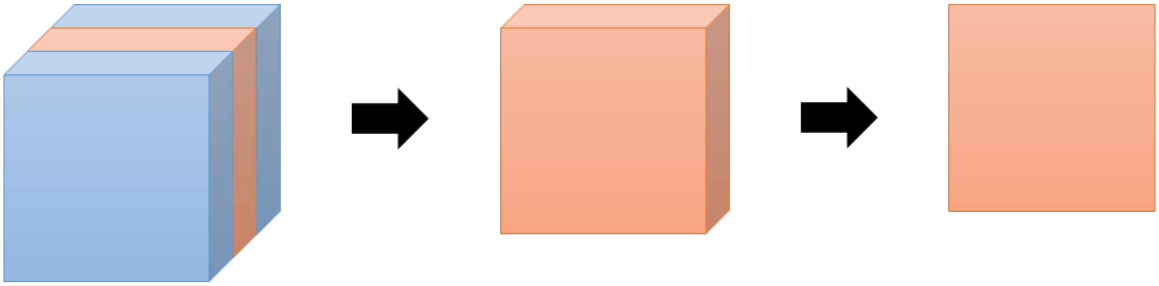


Figure 6.7: A data slice or a data block can be taken from a datacube, then we can project to a low-dimensional space.

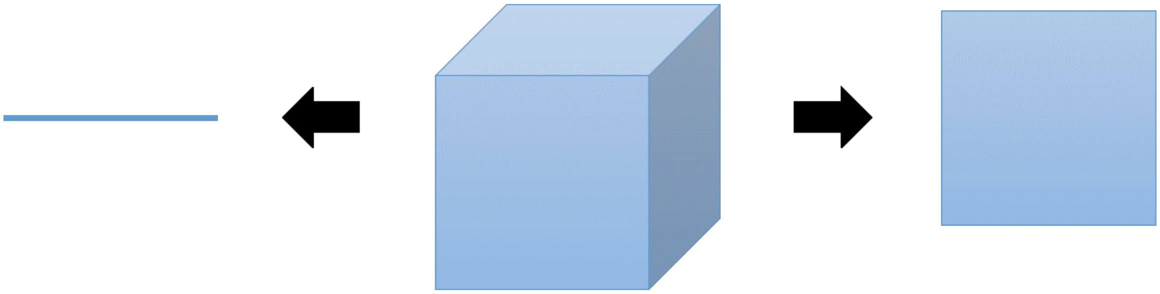


Figure 6.8: Projection to 1D or 2D space is widely used and optimized in the code.

from DenseData, but not reverse. Just like what we do in EventArray, which can be viewed as a special SparseData, we can quick sort the discrete points in SparseData on a given dimension with an average time complexity $O(n \log n)$. SparseData can also be sliced or projected in $O(n)$.

6.4.4 Data Organization

In UMIS, the whole dataset is divided into several packages on time series. This design is used because the incoming events are delivered as pulse packages. In each data package, a bin structure is defined to collect and bin events. Note that this bin structure is only based on the parameters before reconstruction. Any reconstruction is performed within one bin. In this bin, events can be sorted by a given parameter, and a datacube can be generated by a given reconstruction algorithm. A datacube is a multi-dimensional array. Finally, the binned datacubes are provided as the reconstructed result. Fig. 6.9 shows the diagram of the data organization.

6.5 Arguments and Options

A command line arguments parser was implemented, which is helpful in running a script. The usage is shown in Table 6.2.

`-h`	`--help`	Displays help.
`-v`	`--version`	Displays version information.
`-a`	`--auto`	Runs automatically.
`-o`	`--option <file-path>`	Assigns `<file-path>` as the option file.
`-d`	`--default`	Assigns `DefaultOptions.json` as the option file.

Table 6.2: Usage of UMIS under command line.

If the option file is not assigned, an option selector dialog will be opened to help

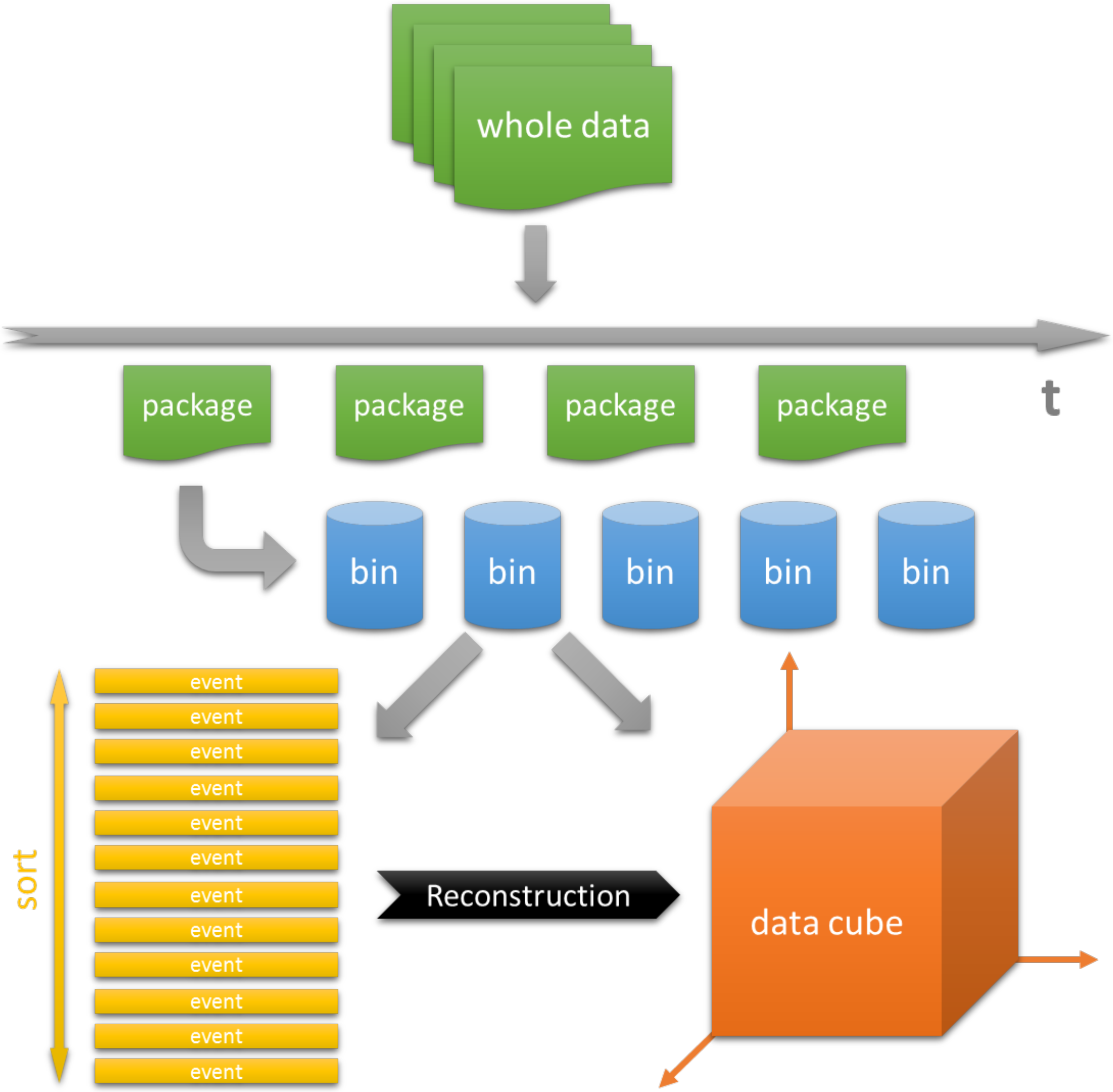
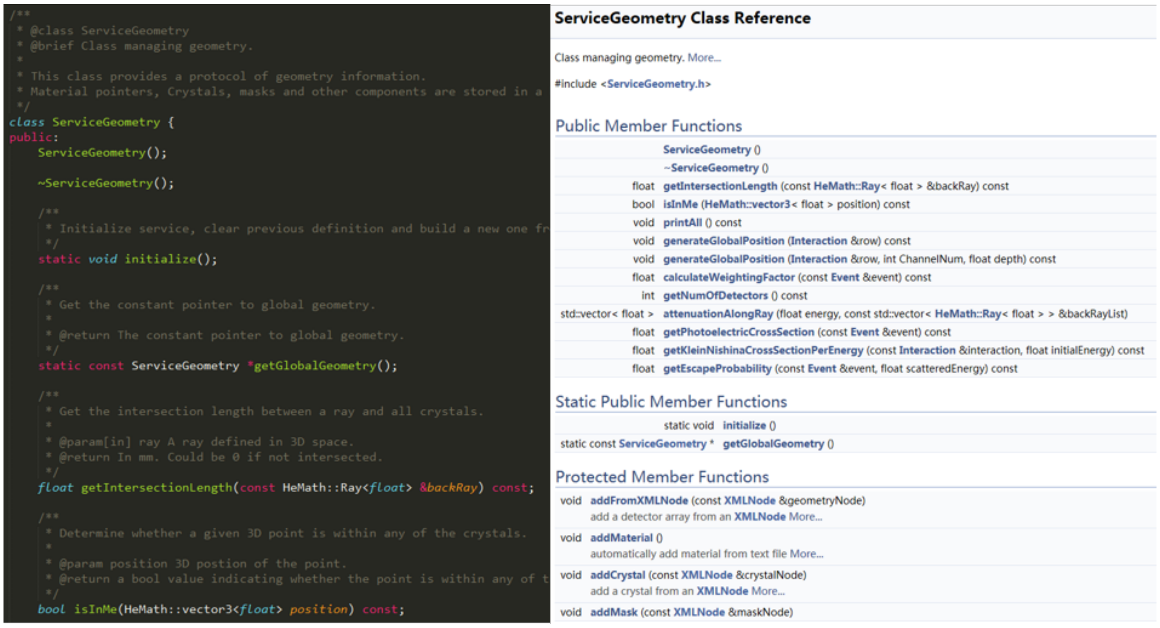


Figure 6.9: Diagram of the data organization.



```cpp
/**
 * @class ServiceGeometry
 * @brief Class managing geometry.
 *
 * This class provides a protocol of geometry information.
 * Material pointers, Crystals, masks and other components are stored in a
 */
class ServiceGeometry {
public:
    ServiceGeometry();

    ~ServiceGeometry();

    /**
     * Initialize service, clear previous definition and build a new one fr
     */
    static void initialize();

    /**
     * Get the constant pointer to global geometry.
     *
     * @return The constant pointer to global geometry.
     */
    static const ServiceGeometry *getGlobalGeometry();

    /**
     * Get the intersection length between a ray and all crystals.
     *
     * @param[in] ray A ray defined in 3D space.
     * @return In mm. Could be 0 if not intersected.
     */
    float getIntersectionLength(const HeMath::Ray<float> &backRay) const;

    /**
     * Determine whether a given 3D point is within any of the crystals.
     *
     * @param position 3D postion of the point.
     * @return a bool value indicating whether the point is within any of t
     */
    bool isInMe(HeMath::vector3<float> position) const;
```

ServiceGeometry Class Reference

Class managing geometry. More...

#include <ServiceGeometry.h>

Public Member Functions

	ServiceGeometry ()
	~ServiceGeometry ()
float	getIntersectionLength (const HeMath::Ray< float > &backRay) const
bool	isInMe (HeMath::vector3< float > position) const
void	printAll () const
void	generateGlobalPosition (Interaction &row) const
void	generateGlobalPosition (Interaction &row, int ChannelNum, float depth) const
float	calculateWeightingFactor (const Event &event) const
int	getNumOfDetectors () const
std::vector< float >	attenuationAlongRay (float energy, const std::vector< HeMath::Ray< float > > &backRayList)
float	getPhotoelectricCrossSection (const Event &event) const
float	getKleinNishinaCrossSectionPerEnergy (const Interaction &interaction, float initialEnergy) const
float	getEscapeProbability (const Event &event, float scatteredEnergy) const

Static Public Member Functions

static void	initialize ()
static const ServiceGeometry *	getGlobalGeometry ()

Protected Member Functions

void	addFromXMLNode (const XMLNode &geometryNode) add a detector array from an XMLNode More...
void	addMaterial () automatically add material from text file More...
void	addCrystal (const XMLNode &crystalNode) add a crystal from an XMLNode More...
void	addMask (const XMLNode &maskNode)

Figure 6.10: Left: Annotated header file of a class. Right: generated documentation for this class in HTML format. PDF format is also available.

the user select an option file. Option files use JSON format, describing most of the options used in the reconstruction. Refer to `DefaultOptions.json` contained in the source code of UMIS for instance.

6.6 Documentation

Most of the classes have been annotated and documented through Doxygen, which is the de facto standard tool for generating documentation from annotated C++ sources, as shown in Fig. 6.10.

CHAPTER VII

Summary and Future Work

7.1 Summary

3D position-sensitive large-volume pixelated semiconductor detectors demonstrated advantages in radiation imaging over other systems, including good energy resolution, sub-pixel spatial resolution, 4π field of view, and portability while at room temperature. Thus better image reconstruction algorithms are requested to make full use of the information provided by detector systems. This dissertation has presented a number of new image reconstruction algorithms that improve the combined performance of information efficiency and reconstruction speed. Also, software has been designed and developed to help research and implement the algorithms.

The essence of image reconstruction is to solve the inverse problem: given the detected interactions, how can the image parameters be generated accurately, effectively, and efficiently? The inverse problem needs a predefined imaging space, which should be carefully selected depending on the problem we want to solve. In addition, a reasonable system model is necessary for the inverse problem, which should neither be too complicated to be efficient, nor too simple to be practical.

Since people always need real-time imaging, direct reconstruction algorithms were studied. Although SBP is quite simple to implement and robust to noise, its poor resolution limits its usage in many scenarios, making FBP a better choice than SBP.

Due to the detector non-isotropic geometry, the conventional FBP has a shift-variant PSF, leading to a non-isotropic resolution. Therefore, I developed the adjusted FBP, which weights the rings based on the geometry and corrects the PSF to be shift-invariant. Another challenge for FBP is the selection of the denoising parameter. As the FBP uses Gaussian detection to approximate Poisson detection, a denoising filter should be applied whose parameter is related to the image statistics. Therefore, the adaptive FBP has been developed in this thesis, which considers the current statistics of the image and selects the Wiener parameter adaptively. These modifications make FBP more applicable in real measurements.

To reconstruct with more accurate statistics, iterative reconstruction algorithms were studied. As a natural extension of the conventional MLEM in every energy bin, the EIID is the ultimate solution for Compton imaging in the energy-imaging integrated space. However the speed of EIID significantly limits its usage, thus EDID has been developed in this thesis to accelerate the convergence rate. The essence of EDID is to use the conditional probability distribution in energy space, such that the images in high-energy bins are firstly reconstructed, and the images in low-energy bins are gradually reconstructed, considering the possible sources in current energy bin and in higher energy bins. The EDID can converge 10 times faster than the EIID, and generate roughly the same images in the energy-imaging integrated space.

Current work is trending to 3D image reconstruction. Compared to 4π spherical imaging, 3D imaging has an imaging space with a huge number of voxels, and relies on the parallax of different viewing angles to the sources, hence there are more challenges in reconstruction speed. To implement 3D imaging in pseudo real-time, the online updating algorithm has been proposed. By batching the incoming data the algorithm will update the image gradually with time. By weighting the events by quality and time, the algorithm always has a memory of previous data and is able to generate the image in the whole measurement period. A joint experiment has shown the feasibility and capability of 3D imaging using moving detectors.

To implement the algorithms and help research, UMIS has been designed and developed. The principle of design is to keep the code fast, simple, reusable, and suitable for research. The fundamental libraries have been made as decoupled as possible. The structure of the code is designed to focus on the data stream rather than the code itself, so that people have more control on the data storing and processing. The kernel of the code has three controllers: event controller, reconstruction controller, and display controller. In event controller, data is received and cleaned. In reconstruction controller, the events are turned into image data using an algorithm, which is easy to develop and plug into the code. In display controller, the image data is printed to file or displayed in the GUI. With the help of UMIS, the research of image reconstruction algorithms is accelerated, and the usage of the algorithms is greatly simplified.

7.2 Future Work

There are several topics not addressed in this work, and future work may make incremental improvements in the combined performance of information efficiency and reconstruction speed.

7.2.1 Model Improvements

Many of the algorithms use a simplified system model to conduct mathematical analysis or achieve faster reconstruction speed, at a sacrifice of model accuracy. One possible direction of future work is to model more system parameters, including geometry attenuation, electronics, and environment, while still preserving a reasonable reconstruction speed.

7.2.2 Filtered Back-Projection

Currently the adaptive FBP focuses on the hottest source in the field of view, thus the denoising parameter may not be optimized for the weaker sources. A possible way to solve the problem is to do filtering using different parameters in different spatial regions.

7.2.3 3D Imaging

Obviously there is a lot of work to be done in 3D imaging. Current sensitivity assumes an ideal point detector, which could be improved by integrating the attenuation into the sensitivity map. The optical camera could also help by excluding some voxels, such as air, thus reducing the size of the imaging space. In addition, it could be very interesting to combine the source detection and imaging with the autonomous control, plan the searching route based on known information and increase searching efficiency.

7.2.4 UMIS

The development of UMIS is not yet finished. There is much space for code optimization and refactoring. Based on the code structure, more algorithms could be coded into the system, and more plotting methods could be developed. Since the code is designed to be usable under the command line interface, more complicated script can be made to do image reconstruction automatically and continuously. In fact, UMIS can be viewed as a combination of library, engine, and GUI, where the code can be easily reused and generate more productivities.

BIBLIOGRAPHY

[1] Arthur H. Compton. A Quantum Theory of the Scattering of X-rays by Light Elements. *Physical Review*, 21(5):483–502, 1923.

[2] Mark W Groch and William D Erwin. Single-photon emission computed tomography in the year 2001: instrumentation and quality control. *Journal of nuclear medicine technology*, 29(1):12–18, 2001.

[3] John M Ollinger and Jeffrey A Fessler. Positron-emission tomography. *IEEE Signal Processing Magazine*, 14(1):43–55, 1997.

[4] Edward E Fenimore and Thomas M Cannon. Coded aperture imaging with uniformly redundant arrays. *Applied Optics*, 17(3):337–347, 1978.

[5] D M Smith, G J Hurford, and S E Boggs. Rotating modulation collimator imagers. *New Astronomy Reviews*, 48(1-4):209–213, 2004.

[6] Manbir Singh. An electronically collimated gamma camera for single photon emission computed tomography. Part I: Theoretical considerations and design criteria. *Medical Physics*, 10(4):421–427, 1983.

[7] Manbir Singh and David Doria. An electronically collimated gamma camera for single photon emission computed tomography. Part II: Image reconstruction and preliminary experimental measurements. *Medical Physics*, 10(4):428–435, 1983.

[8] V. Schönfelder, A. Hirner, and K. Schneider. A telescope for soft gamma ray astronomy. *Nuclear Instruments and Methods*, 107(2):385–394, 1973.

[9] R. W. Todd, J. M. Nightingale, and D. B. Everett. A proposed γ camera. *Nature*, 251(5471):132–134, 1974.

[10] T. Kamae, N. Hanada, and R. Enomoto. PROTOTYPE DESIGN OF MULTIPLE COMPTON GAMMA-RAY CAMERA. *IEEE Transactions on Nuclear Science*, 35(1), 1987.

[11] R. B. Piercey, A. G. Weisenberger, J. E. McKisson, and C. Girit. Preliminary characteristics of a germanium-based, compton-scatter telescope. *IEEE Transactions on Nuclear Science*, 36(1):887–890, 1989.

[12] N. Dogan, D. K. Wehe, and G. F. Knoll. Multiple Compton scattering gamma ray imaging camera. *Nuclear Inst. and Methods in Physics Research, A*, 299(1-3):501–506, 1990.

[13] J. B. Martin, G. F. Knoll, D. K. Wehe, N. Dogan, V. Jordanov, N. Petrick, and M. Singh. A Ring Compton Scatter Camera for Imaging Medium Energy Gamma Rays. *IEEE Transactions on Nuclear Science*, 40(4):972–978, 1993.

[14] J. E. McKisson, P. S. Haskins, G. W. Phillips, S. E. King, R. A. August, R. B. Piercey, and R. C. Mania. Demonstration of Three-Dimensional Imaging with a Germanium Compton Camera. *IEEE Transactions on Nuclear Science*, 41(4):1221–1224, 1994.

[15] B. F. Phlips, S. E. Inderhees, R. A. Kroeger, W. N. Johnson, R. L. Kinzer, J. D. Kurfess, B. L. Graham, and N. Gehrels. Performance of a Compton Telescope using Position-Sensitive Germanium Detectors. *IEEE Transactions on Nuclear Science*, 43(3):1472–1475, 1996.

[16] J.W. LeBlanc, N.H. Clinthorne, C.-H. Hua, E. Nygard, W.L. Rogers, D.K. Wehe, P. Weilhammer, and S.J. Wilderman. C-SPRINT: a prototype Compton camera system for low energy gamma ray imaging. *IEEE Transactions on Nuclear Science*, 45(3):943–949, 1998.

[17] J. W. Leblanc, N. H. Clinthorne, C. H. Hua, E. Nygard, W. L. Rogers, D. K. Wehe, P. Weilhammer, and S. J. Wilderman. Experimental results from the C-SPRINT prototype compton camera. *IEEE Transactions on Nuclear Science*, 46(3 PART 1):201–204, 1999.

[18] G. J. Schmid, D. A. Beckedahl, J. E. Kammeraad, J. J. Blair, K. Vetter, and A. Kuhn. Gamma-ray Compton camera imaging with a segmented HPGe. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 459(3):565–576, 2001.

[19] Y. F. Du, Z. He, G. F. Knoll, D. K. Wehe, and W. Li. Evaluation of a Compton scattering camera using 3-D position sensitive CdZnTe detectors. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 457(1-2):203–211, 2001.

[20] R. A. Kroeger, W. N. Johnson, J. D. Kurfess, B. F. Phlips, and E. A. Wulf. Three-Compton telescope: Theory, simulations, and performance. *IEEE Transactions on Nuclear Science*, 49 I(4):1887–1892, 2002.

[21] E. A. Wulf, J. Ampe, W. N. Johnson, R. A. Kroeger, J. D. Kurfess, and B. F. Phlips. Depth measurement in a germanium strip detector. *IEEE Transactions on Nuclear Science*, 49 I(4):1876–1880, 2002.

[22] Eric A. Wulf, Bernard F. Phlips, W. Neil Johnson, Richard A. Kroeger, James D. Kurfess, and Elena I. Novikova. Germanium strip detector compton telescope using three-dimensional readout. In *IEEE Transactions on Nuclear Science*, volume 50, pages 1182–1189, 2003.

[23] Eric A. Wulf, Bernard F. Phlips, W. Neil Johnson, James D. Kurfess, and Elena I. Novikova. Thick silicon strip detector compton imager. *IEEE Transactions on Nuclear Science*, 51(5 I):1997–2003, 2004.

[24] Feng Zhang, Zhong He, Dan Xu, Glenn F. Knoll, David K. Wehe, and James E. Berry. Improved resolution for 3-D position sensitive CdZnTe spectrometers. *IEEE Transactions on Nuclear Science*, 51(5 I):2427–2431, 2004.

[25] Carolyn E. Lehner, Zhong He, and Feng Zhang. 4pi Compton Imaging Using a 3-D Position-Sensitive CdZnTe Detector Via Weighted List-Mode Maximum Likelihood. *IEEE Transactions on Nuclear Science*, 51(4):1618–1624, 2004.

[26] T. Niedermayr, K. Vetter, L. Mihailescu, G. J. Schmid, D. Beckedahl, J. Blair, and J. Kammeraad. Gamma-ray imaging with a coaxial HPGe detector. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 553(3):501–511, 2005.

[27] Morgan Burks, Dan Chivers, Chris Cork, Mark Cunningham, Lorenzo Fabris, Donald Gunter, Ethan Hull, David Lange, Hugh Manini, Lucian Mihailescu, Karl Nelson, Thomas Niedermayr, John Valentine, Kai Vetter, and Doug Wright. Imaging performance of the Si/Ge hybrid compton imager. In *IEEE Nuclear Science Symposium Conference Record*, volume 1, pages 6–10, 2005.

[28] Yuefeng Zhu, Stephen E. Anderson, and Zhong He. Sub-Pixel Position Sensing for Pixelated, 3-D Position Sensitive, Wide Band-Gap, Semiconductor, Gamma-Ray Detectors. *IEEE Transactions on Nuclear Science*, 58(3):1400–1409, 2011.

[29] S.J. Wilderman, Jeffrey A. Fessler, Neal H. Clinthorne, J.W. LeBlanc, and W.L. Rogers. Improved modeling of system response in list mode EM reconstruction of Compton scatter camera images. *IEEE Transactions on Nuclear Science*, 48(1):111–116, 2001.

[30] Dan Xu and Zhong He. Gamma-ray energy-imaging integrated spectral deconvolution. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 574(1):98–109, 2007.

[31] Weiyi Wang, Christopher G. Wahl, Jason M. Jaworski, and Zhong He. Maximum-likelihood deconvolution in the spatial and spatial-energy domain for events with any number of interactions. *IEEE Transactions on Nuclear Science*, 59(2):469–478, 2012.

[32] Lucas C. Parra. Reconstruction of cone-beam projections from Compton scattered data. *IEEE Transactions on Nuclear Science*, 47(4):1543–1550, 2000.

[33] Dan Xu. *GAMMA-RAY IMAGING AND POLARIZATION MEASUREMENT USING 3-D POSITION-SENSITIVE CdZnTe DETECTORS*. PhD thesis, University of Michigan, 2006.

[34] Andrew Haefner, Donald Gunter, Ross Barnowski, and Kai Vetter. A Filtered Back-Projection Algorithm for 4pi Compton Camera Data. *IEEE Transactions on Nuclear Science*, 62(4):96–99, 2015.

[35] Zhong He. Review of the Shockley-Ramo theorem and its application in semiconductor gamma-ray detectors. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 463(1-2):250–267, 2001.

[36] Feng Zhang. *Events Reconstruction in 3-D Position Sensitive CdZnTe Gamma Ray Spectrometers*. PhD thesis, University of Michigan, 2005.

[37] Hao Yang. *Applications of Digital ASIC Array System for Noise Analysis , Non-linearity Correction , Event Classification and Reconstruction*. PhD thesis, University of Michigan, 2013.

[38] Yuefeng Zhu. *Digital Signal Processing Methods for Pixelated 3-D Position Sensitive Room-Temperature*. PhD thesis, University of Michigan, 2012.

[39] G Knoll. *Radiation detection and measurement*. 2005.

[40] Feng Zhang, Zhong He, and Carolyn E. Seifert. A prototype three-dimensional position sensitive CdZnTe detector array. *IEEE Transactions on Nuclear Science*, 54(4):843–848, 2007.

[41] Christopher Wahl. *Imaging, Detection, and Identification Algorithms for Position-Sensitive Gamma-Ray Detectors*. PhD thesis, University of Michigan, 2011.

[42] Dan Xu, Zhong He, and Feng Zhang. Detection of gamma ray polarization using a 3-D position-sensitive CdZnTe detector. *IEEE Transactions on Nuclear Science*, 52(4):1160–1164, 2005.

[43] Michael W Streicher. *Applications of Digitized 3-D Position-Sensitive CdZnTe Spectrometers for National Security and Nuclear Nonproliferation by*. PhD thesis, University of Michigan, 2017.

[44] Carolyn E Lehner, Zhong He, and Glenn F Knoll. Intelligent Gamma-Ray Spectroscopy Using 3-D Position-Sensitive Detectors. *IEEE Transactions on Nuclear Science*, 50(4):1090–1097, 2003.

[45] Oskar Klein and Yoshio Nishina. Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac. *Zeitschrift für Physik*, 52(11-12):853–868, 1929.

[46] Charles M. Stein. Estimation of the Mean of a Multivariate Normal Distribution. *The Annals of Statistics*, 9(6):1135–1151, 1981.

[47] James H. McClellan and Thomas W. Parks. Eigenvalue and Eigenvector Decomposition of the Discrete Fourier Transform. *IEEE Transactions on Audio and Electroacoustics*, 20(1):66–74, 1972.

[48] J.R. Driscoll and D.M. Healy. Computing Fourier Transforms and Convolutions on the 2-Sphere. *Advances in Applied Mathematics*, 15(2):202–250, 1994.

[49] Reiji Suda and Masayasu Takami. A fast spherical harmonics transform algorithm. *Mathematics of Computation*, 71(238):703–716, 2001.

[50] Stefan Kunis and Daniel Potts. Fast spherical Fourier algorithms. *Journal of Computational and Applied Mathematics*, 161(1):75–98, 2003.

[51] Martin Reinecke. Libpsht - algorithms for efficient spherical harmonic transforms. *Astronomy & Astrophysics*, 526(A108):9, 2011.

[52] Nathanaël Schaeffer. Efficient spherical harmonic transforms aimed at pseudospectral numerical simulations. *Geochemistry, Geophysics, Geosystems*, 14(3):751–758, 2013.

[53] Jeffrey A. Fessler. *Image Reconstruction: Algorithms and Analysis*. Unpublished, 2017.

[54] J Chu, M Streicher, J A Fessler, and Z He. Unbiased Filtered Back-Projection in 4π Compton Imaging With 3D Position Sensitive Detectors. *IEEE Transactions on Nuclear Science*, 63(6):2750–2756, 2016.

[55] William Hadley Richardson. Bayesian-based iterative method of image restoration. *JOSA*, 62(1):55–59, 1972.

[56] Leon B Lucy. An iterative technique for the rectification of observed distributions. *The astronomical journal*, 79:745, 1974.

[57] Kenneth Lange, David R Hunter, and I. Yang. Optimization Transfer Using Surrogate Objective Functions. *Journal of Computational and Graphical Statistics*, 9(1):1–20, 2000.

[58] Daniel J Lingenfelter. *Source Detection and Image Reconstruction with Position - Sensitive Gamma - Ray Detectors*. PhD thesis, University of Michigan, 2012.

[59] Linda Kaufman. Implementing and Accelerating the EM Algorithm for Positron Emission Tomography. *IEEE Transactions on Medical Imaging*, 6(1):37–51, 1987.

[60] M V Ranganath, A P Dhawan, and N Mullani. A multigrid expectation maximization reconstruction algorithm for positron emission tomography. *IEEE transactions on medical imaging*, 7(4):273–8, 1988.

[61] T S Pan and A E Yagle. Numerical study of multigrid implementations of some iterative image reconstruction algorithms. *IEEE transactions on medical imaging*, 10(4):572–88, 1991.

[62] E Tanaka. A fast reconstruction algorthm for stationary positron emission tomography based on a modified em algorithm. *IEEE transactions on medical imaging*, 6(2):98–105, 1987.

[63] S Vishampayan, J Stamos, R Mayans, K Koral, N Clinthorne, and W L Rogers. Maximum likelihood image reconstruction for SPECT. *J. Nucl. Med.;(United States)*, 26(CONF-850611-), 1985.

[64] Alfredo N. Iusem. Convergence analysis for a multiplicatively relaxed EM algorithm. *Mathematical Methods in the Applied Sciences*, 14(8):573–593, 1991.

[65] T. S. Zaccheo and R. A. Gonsalves. Iterative maximum-likelihood estimators for positively constrained objects. *Journal of the Optical Society of America A*, 13(2):236–242, 1996.

[66] H. Lantéri, M. Roche, O. Cuevas, and C. Aime. A general method to devise maximum-likelihood signal restoration multiplicative algorithms with nonnegativity constraints. *Signal Processing*, 81(5):945–974, 2001.

[67] R M Lewitt and G Muehllehner. Accelerated iterative reconstruction for positron emission tomography based on the em algorithm for maximum likelihood estimation. *IEEE transactions on medical imaging*, 5(1):16–22, 1986.

[68] H Hart and Z Liang. Bayesian image processing in two dimensions. *IEEE transactions on medical imaging*, 6(3):201–8, 1987.

[69] Peter Schmidlin, Matthias E. Bellemann, and Gunnar Brix. Iterative reconstruction of PET images using a high-overrelaxation single-projection algorithm. *Physics in Medicine and Biology*, 42(3):569–582, 1997.

[70] P Schmidlin, G Brix, M E Bellemann, and W J Lorenz. Computation of high overrelaxation parameters in iterative image reconstruction. *IEEE Trans. Nuc. Sci.*, 45(3):1737–1742, 1998.

[71] DoSik Hwang and Gengsheng L. Zeng. Convergence study of an accelerated ML-EM algorithm using bigger step size. *Physics in Medicine and Biology*, 51(2):237–252, 2006.

[72] Johan Nuyts, Luc Mortelmans, and Paul Suetens. Acceleration of Maximum Likelihood Reconstruction, Using Frequency Amplification and Attenuation Compensation. *IEEE Transactions on Medical Imaging*, 12(4):643–652, 1993.

[73] N Rajeevan and K Rajgopal. Vector-extrapolated fast maximum likelihood estimation algorithms for emission tomography. *Medical Imaging, IEEE Transactions on*, 11(1):9–20, 1992.

[74] Timothy J. Holmes and Yi-Hwa Liu. Acceleration of maximum-likelihood image restoration for fluorescence microscopy and other noncoherent imagery. *Journal of the Optical Society of America A*, 8(6):893, 1991.

[75] H Malcolm Hudson and Richard S Larkin. Accelerated Image Reconstruction Using Ordered Subsets of Projection Data. *IEEE transactions on medical imaging*, 13(4):601–609, 1994.

[76] a.C. Sauve, Iii Hero, a.O., W.L. Rogers, S.J. Wilderman, and N.H. Clinthorne. 3D image reconstruction for a Compton SPECT camera model. *IEEE Transactions on Nuclear Science*, 46(6), 1999.

[77] Lucian Mihailescu, Kai Vetter, and Daniel Chivers. Standoff 3D gamma-ray imaging. *IEEE Transactions on Nuclear Science*, 56(2):479–486, 2009.

[78] Van-Giang Nguyen, Soo-Jin Lee, and Mi No Lee. GPU-accelerated 3D Bayesian image reconstruction from Compton scattered data. *Physics in medicine and biology*, 56(9):2817–36, 2011.

[79] Jason M. Jaworski. *Compton Imaging Algorithms for Position-Sensitive Gamma-Ray Detectors in the Presence of Motion.* PhD thesis, University of Michigan, 2013.

[80] Christopher G. Wahl, Willy R. Kaye, Weiyi Wang, Feng Zhang, Jason M. Jaworski, Alexis King, Y. Andy Boucher, and Zhong He. The Polaris-H imaging spectrometer. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 784:377–381, 2015.

[81] Christopher G. Wahl, Jason M. Jaworski, and Zhong He. UMImaging: A software package for image reconstruction from 3D-position-sensitive gamma-ray detectors. *IEEE Transactions on Nuclear Science*, 59(4 PART 3):1672–1680, 2012.