Imaging, Detection, and Identification Algorithms for Position-Sensitive Gamma-Ray Detectors

by

Christopher G. Wahl

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
2011

Doctoral Committee:

Professor Zhong He, Chair
Professor Jeffrey A. Fessler
Emeritus Professor Glenn F. Knoll
Assistant Professor Clayton D. Scott

There once was a grad student named Chris,
Who sat down to write his thesis.
I hope you will agree,
When the next pages you see,
That it is better than his attempted poetry: this.

Acknowledgements

I am surrounded by a great cloud of people who have helped to make this dissertation possible, and to them I am grateful. First, to my parents, Glenn Wahl and Ruth Wahl, who taught me hard work and who have continuously encouraged me through the past 23 years of my schooling. And, to all the teachers and mentors who have shown me the path – Prof. James McLean, who advised me during undergraduate work and suffered through an entire summer while I put together my first journal paper; Prof. Ed 'Pogo' Pogozelski, who taught me how to speak in front of a group and so much more; and Dr. Leticia Pibida, who encouraged me to apply to the University of Michigan – just to name a few.

At Michigan, I am primarily indebted to my adviser, Prof. Zhong He, who provided an environment where I could stretch my wings. I have enjoyed my time in his group, and I feel lucky to have been a part of it. Beyond Prof. He's management, the Orion group would not have been what is was without those I have worked with over the years: the imaging group: Weiyi Wang, Dan Lingenfelter, Jason Jaworski, Sonal Joshi, and (previously) Crystal Thrall; those who constructed spectacular systems that made our imaging algorithms look great: Jim Berry, Dr. Feng Zhang, Yuefeng Zhu, Willy Kaye, Cedric Herman, Hao Yang, and Andy Boucher; and the other group members: Steve Anderson, Miesher Rodrigues, Dr. JaeCheon Kim, Dr. Burcin Donmez, and Suzanne Nowicki. I would especially like to thank Adam Higuera, my first summer student, who set down the basic structure to the UMImaging GUI and geometry classes and was a great help in the code's early days.

The rest of my committee has been influential. Prof. Glenn Knoll lived up to his reputation of being the benevolent god of radiation measurement, setting the standard I strove to emulate. Prof. Jeffrey Fessler is responsible for everything I know about model-based image reconstruction, and my lack of study in his classes is responsible for everything about which I am still ignorant. Without our occasional chats (what I called

"my drinks from the fire hose"), I would not have gotten far. I would also like to thank Prof. Clayton Scott for serving on my committee and providing a fresh perspective on my work.

In the department, I would like to thank Peggy Gramer, who was my first contact at Michigan and who had sage advice about everything from finding housing to scheduling a defense; and the fission guys – Dr. Greg Davidson, Dr. Troy Becker, Dr. Jeremy Conlin, and Nick Touran – who shared their programming tips daily, whether requested or not.

I am grateful for the gift of many friends during my time in Michigan – too many to list here – who welcomed me, trusted me, listened to me, included me, encouraged me, strengthened me, grew me, and put up with me. I am especially grateful to my friends from Graduate Christian Fellowship: the original Tuesday-night group of Alton, Amanda, Andrew, Anna, Dennis, Jennifer, Jenn, Jenny, Sara, and Stephen, the men's groups of John L., John T., Darren, and Jason, and later of Brittan, Chris, and Samir. You have meant so much to me. Time would fail me to tell of Melodie, Martha, Olivier, Cat, Dave, Rachel, Jen, Bill, Kira, Andy, and T; of Ashley and Lindsey; of Brad and Becky, and Josh and Becky; of north-campus-lunch accomplices Mark, Patrick, and Jim; of the classical pianists Matt, Joel, Richard, and Steven; of the physics girls Sarah, Mallory, and Nicole; of Crystal who listened to me for hours on end; of Katherine who taught me how to use compound-adjective hyphens; and of Jenn who taught me how to live. To these and so many others, I find myself in the happy position of gratitude.

A number of people besides my committee proofread and give suggestions on this manuscript: Dan Lingenfelter, Jason Jaworski, Willy Kaye, Crystal Thrall, Shikha Prasad, and Paul Stephens. I also need to thank the Department of Defense and their National Defense Science and Engineering Graduate Fellowship Program for my first three years of support, which allowed me to tackle what was most interesting to me without concern for funding.

Thank you to you all.

Table of Contents

List of Figures

List of Tables

Abstract

Three-dimensional-position-sensitive semiconductors record both the locations and energies of gamma-ray interactions with high resolution, enabling spectroscopy and imaging of gamma-ray-emitting materials. Imaging enables the detection of point sources of gamma rays in an otherwise extended-source background, even when the background spectrum is unknown and may share the point source's spectrum. The generalized likelihood ratio test (GLRT) and source-intensity test (SIT) are applied to this situation to detect one-or-more unshielded point sources from a library of isotopes in a spectrally unknown or known background when the background intensity varies spatially by a factor of two or less. In addition to estimating the number of sources present, their activities, isotopes, and directions from the detector are estimated. Experimental and some simulated results are presented for a single detector and an 18-detector array of 2 cm × 2 cm × 1.5 cm CdZnTe crystals and compared with the performance of spectral-only detection when the background and source are assumed to be spectrally different. Furthermore, the expected detection performance of the 18-detector array system is investigated statistically using experimental data in the case where the background is distinct spectrally from the point source and the possible source location and isotopic identity are known. Including imaging gave at least 7% higher SNR compared to ignoring the image dimension.

Also, imaging methods based on the maximum-likelihood expectation-maximization method are introduced to determine the spatial distribution of isotopes and to find the activity distributions within targets moving with known motion through a radioactive background. Software has also been developed to support the analysis of the data from 3D-position-sensitive spectroscopic systems, for a range of detector designs and applications. The software design and unique features that allow fast multidimensional data analysis are presented, along with parallel computing performance.

Chapter 1

Introduction

Police Officer (approaching a man slowly walking around the base of a streetlight looking down): *Are you looking for something?*
Man: *I lost my keys somewhere on this street.*
Police Officer: *Then why are you only looking under this streetlight?*
Man: *The light is better.*

Gamma rays cannot be felt or seen, so we must trust detectors – the syncopated ticking of a Geiger counter or the flashing lights of a roadside portal monitor. The menagerie of gamma-ray detectors includes detectors that can survive in rugged terrain with little power, quietly taking measurements, and others that need constant care, yet provide beautifully detailed data. The workhorses of this group, toiling with health physicists around reactors and engineers in industrial processes, all harvest either total count rates or, the best of them, muddy gamma-ray energy distributions. In the medical field, doctors have long ago learned how to put blinders on a flock of detectors simultaneously to learn the rays' spatial distribution within their patients.

Within this zoo of gamma-ray detectors, only a select few have one striking feature: three-dimensional position sensitivity combined with high energy resolution. Without having to move around, their unique senses map the gamma-ray energy distribution in all directions, and with a companion, they triangulate gamma-ray threats nearby. The sharp senses of these detectors have grown in recent years due to many development efforts. We harness the detectors as guides to the invisible world of gamma rays. Eventually, they may cultivate a number of fields from fundamental physics to health care to nuclear nonproliferation. The pages that follow describe several methods that, when grafted to these detectors, make full use of their capabilities and report their observations in meaningful ways.

Specifically, yoking detector output with data-analysis algorithms, the resulting pair more clearly images the gamma-ray field, more naturally understands moving sources, more completely makes use of the detector data, more accurately identifies gamma-ray sources, and more sensitively detects threats. These improved capabilities are valuable in, for instance, security applications. I will analyze the intrinsic capabilities of these detectors for detection of localized point sources and show the situations in which they have a distinct advantage over more standard workhorse systems. Such analysis informs users about which systems are best for specific circumstances.

Contributions of this Work

This work contributes to three areas of analysis for 3D-position-sensitive gamma-ray detectors, with the goal of better using detector output to obtain useful information about the radiation environment. First, in imaging, two new applications of the MLEM equation are proposed to deconvolve the contribution from a number of moving targets as a function of energy and to deconvolve the contribution from different isotopes as a function of direction. These methods are applicable to situations where one wishes to understand the spatial distribution of radiation emissions from moving sources or the isotopes that are responsible for the gamma-ray energy distribution from each direction. In addition, a large fraction of events in the detector are single-interaction events that are not imaged. Initial studies have been completed to use these single-interaction events for imaging based on attenuation in the detector.

Second, implementation of imaging, detection, and identification algorithms requires robust computer software that provides basic capabilities for dealing with detector output and algorithm results. Therefore, software has been developed to implement these algorithms in a modular, flexible manner. The software design aims to support all manner of data analysis for these detectors and to contribute to algorithm and hardware development.

Finally, an important application is the detection of threat sources in the environment. The prospects for automated detection of gamma-ray point sources in a background that is roughly spatially uniform are analyzed for a particular system. Statistical arguments show that when the expected point sources have emission lines

distinct from background, imaging contributes only a small improvement to detectability over just spectroscopy. However, when the expected point sources may not have emission lines distinct from background, spectroscopy alone is inadequate. The generalized likelihood ratio test (GLRT) and source-intensity test (SIT) are applied for automated detection in the latter case. The GLRT and SIT performance are studied in a number of cases where the source location, source identity, and background spectrum may or may not be known. Additionally, a modified GLRT is applied to estimate the number of point sources present.

Overview

In Chapter 2, I will introduce you, dear reader, to position-sensitive detectors and the specific species used in this study – pixelated CdZnTe. Although pixelated CdZnTe offers numerous advantages, I will also point out some current imperfections that complicate translation from raw measured data to images and detections. Chapter 3 offers an integrated overview of the imaging and analysis possible with position-sensitive detectors. In Chapter 4, one of these methods – maximum likelihood expectation maximization – is applied to a number of applications, some old and some novel. The design and implementation of software to perform these calculations is discussed in Chapter 5, with a focus on its distinctive features. This is followed by detailed analysis of the prospects for automated threat detection with the current array system. Chapter 6 argues statistically, and Chapter 7 applies two algorithms to detection of one or more point sources in an unknown or known background. Finally, Chapter 8 envisions, insofar as this is possible, future progress that will open these detectors to better performance in imaging, detection, and other applications.

Chapter 2

Detector Characteristics

> *We are like dwarfs on the shoulders of giants, so that we can see more than they, and things at a greater distance, not by virtue of any sharpness of sight on our part, or any physical distinction, but because we are carried high and raised up by their giant size.*
> – Bernard of Chartres (c. 1124), quoted by John of Salisbury (1159), made famous by Isaac Newton (1676), as found on Wikipedia (2011)

Throughout this work, I will use 3D-position-sensitive semiconductor detectors to image and test algorithm performance. It is thanks to many decades of development of these detectors that the information needed for imaging and detection is available. Without them, the rest of this dissertation probably would not exist.

This chapter describes the basic operation of these detectors and then discusses several complicating factors that may affect imaging and detection performance. The capabilities and limitations of these detectors inform algorithm development and analysis. For more details about these detectors and their operation, refer to other dissertations [1,2].

CdZnTe Detectors

CdZnTe (CZT) is a semiconductor with a bandgap of about 1.6 eV [3]. Therefore, the energy required to excite electrons from the valance to conduction bands is small, yet the probability of thermal excitation into the conduction band is extremely small at room temperature. Unlike HPGe detectors, which have a smaller bandgap and hence larger probability of thermal excitation to the conduction band at room temperature, the larger bandgap allows CZT systems to operate at room temperature.

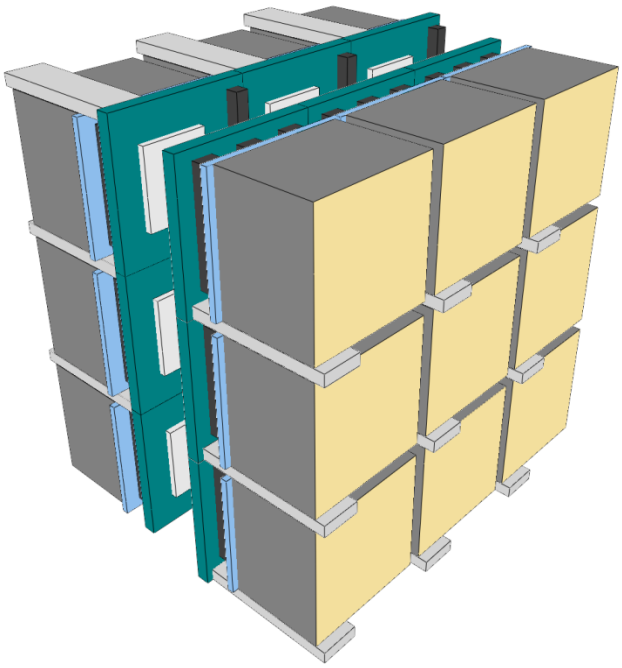

Figure 2.1. The 18-detector CZT system.

Radiation that deposits energy in the CZT crystal elevates a number of electrons proportional to the deposited energy into the conduction band. By applying a strong electric field through the crystal, these electrons (and to some extent the positively charged holes) will drift, inducing a signal on the anode and cathode electrodes. These signals indirectly indicate the interaction position and deposited energy.

The crystals used in these studies are CZT cuboids of 20 mm × 20 mm × 15 mm. The largest side is covered with a planar cathode electrode, and an array of 11 × 11 anode electrodes ('pixels') is attached to the opposite side, with a pixel pitch of 1.72 mm. Many detectors also include a thin, negatively biased, grid electrode between the anode pixels to improve charge collection. Each of the 123 electrodes is read out through a single application-specific integrated circuit (ASIC) developed by Gamma-Medica Ideas [4].

Recently, systems containing 18 of these crystals were constructed, with two layers of 3 × 3 crystals, as shown in Figure 2.1. The separation between neighboring crystals is nominally 2 mm, and the separation between the bottoms of the two planes is either 5.50 cm or 4.09 cm. Zhang *et al.* describe a similar system with smaller crystals [5].

Single-Carrier Charge Sensing

Electron mobility and lifetime is significantly higher in CZT than that of the holes. Therefore, for most purposes, the holes are assumed to be stationary and the entire signal is due to the electron movement. As the electrons in the charge cloud produced by the radiation interaction drift towards the anode, they induce charge on all the conducting electrodes consistent with that expected from electromagnetic theory. A simple method to calculate the magnitude of the induced charge is the Shockley-Ramo theorem, which describes the change in induced charge on an electrode due to charge motion by means of a 'weighting potential' that is independent of applied potential or trapped charge [6]. The change in the induced charge ΔQ_i on an electrode *i* is

$$\Delta Q_i = -q\,[\phi_i(x_1) - \phi_i(x_0)] \qquad (2.1),$$

where q is the moving charge and $[\phi_i(x_1) - \phi_i(x_0)]$ describes the difference in weighting potential for electrode *i* between the final and initial position of the charge. The weighting potential is calculated by imagining no space charge and a potential set to zero on all electrodes but the electrode of interest, which is set to unity. The resulting potential field – defined by Laplace's equation – is the electrode's weighting potential. This potential field is distinct from the physical potential field that causes charge drift, so the actual electric field in the detector is needed to understand the charge's path. Then the weighting potential along that path can be used to find the induced charge.

For the system described here, the weighting potential for the planar cathode electrode is a linear ramp from the anode side (weighting potential of zero) to the cathode (weighting potential of one) [3], with a slight curvature near the edges of the crystal due to edge effects [2]. This is consistent with the potential field expected from a planar device. Electron movement in the bulk then will induce a charge on the cathode proportional to both the charge and the distance traveled from the interaction position to the anode.

The weighting potential for each anode electrode is near zero in most of the bulk, including most everywhere on the anode side, but quickly rises to one very near the anode electrode of interest [3]. Because of this small-pixel effect, the anode electrodes register a signal proportional to the charge, but not interaction position, at the moment the

charge reaches the electrode. Reading out the collecting-anode signal tells the amount of charge and hence the deposited energy. In addition, the anode possessing the positive signal tells the lateral position of the charge.

The ratio of the cathode to the collecting-anode signal amplitudes gives the depth of interaction because the cathode signal is proportional to depth and charge and the anode signal is proportional to just charge. The detector hence provides the three-dimensional interaction position in the crystal, which is useful for imaging and for correcting the recorded energy for any charge trapping or non-uniformities in the crystal [7]. When multiple interactions occur in the crystal, the anode signals still record the charge in each pixel, but the cathode signal is the superposition of the signals from each interaction, so the cathode-to-anode ratios cannot uniquely determine the depths.

Fortunately, the depth of interaction can also be determined by electron drift time. As soon as the electron cloud begins to move, it begins to induce charge on the cathode electrode, establishing a start signal. The stop signal is from the quickly rising anode signal. Single-interaction events give the correlation between drift time and the cathode-to-anode ratio. Unlike the cathode-to-anode ratio, drift time can give the depth of multiple-pixel events because the start time on the cathode is the same for each interaction. The depth resolution depends on the amount of charge deposited, but usually is better than 1 mm.

Energy resolution of single-pixel events below 0.5% full width at half maximum (FWHM) at 662 keV is achievable in the most uniform detectors after applying depth and pixel corrections [8], although more common detectors attain about 1% FWHM. Due to complicating effects, the resolution of multiple-pixel events degrades slightly faster than that expected just due to the addition of noise, resulting in a few-percent FWHM at 662 keV [2].

Currently, signal amplitude and timing are read using analog shaping and peak hold circuitry in the ASIC. Work is in progress to use digital pulse shapes to learn more about the interactions. For instance, the small transient signals on neighboring pixels can help tell the sub-pixel position of the interaction [9], the shape of the anode signal can tell if there are multiple interactions within that pixel [10], and the shape of the cathode

signal can tell the drift times and therefore the depths of each interaction independently from anode timing [11].

Non-Ideal Characteristics

For best performance in imaging and detection, a detector should report the exact interaction positions and the energies deposited at each. Barring this, it should at least report these quantities with well-quantified uncertainty. The following section briefly lists a number of effects that confuse the system response beyond the idealized operation described above, sometimes leading to estimates of the interaction positions and energies that fall outside of what is expected from basic models. While these factors could be included in the models of the detector, some effects are hard to model in a computationally efficient way and others are understood only roughly.

Multiple Interactions in a Single Pixel

We generally assume that the charge collected at each anode pixel is due to a single interaction at one depth below that pixel. However, it is possible for multiple interactions to occur under one pixel. Depending on the depth difference between the interactions, the total recorded energy may be incorrect and the depth for at least one interaction will be wrong. Processing the digital waveform can overcome this problem, as long as the interactions occur at different depths, by observing the quick charge induction that occurs when each charge cloud reaches the anode [10].

Charge Sharing and Weighting-Potential Crosstalk

Conversely, our model assumes that when charge is recorded on multiple neighboring pixels there must have been an interaction in each of the pixels. However, there are two mechanisms that can cause a single radiation interaction to appear in multiple pixels: charge sharing and weighting-potential crosstalk.

Electrons traveling towards one anode electrode induce transitory charge on neighboring electrodes as they momentarily travel through a higher neighboring-pixel weighting potential before being collected by the collecting pixel. Shaped, this transitory neighbor pulse may appear like a positive induced charge, especially if the inducing

charge is large. Therefore, the signals appear to come from an event with one large interaction energy and one small interaction energy (the induced signal). This confusion is one aspect of weighting-potential cross talk [2]. This problem can also be overcome by processing the digital waveform [11].

In addition, because the radiation interaction creates not a point of charge but a cloud with extent roughly proportional to the energy deposition [1], there is some chance that the cloud could be collected by two – or even three – neighboring pixels. The signature of these charge-sharing events is that the true depth in each pixel is the same. Some work has been done to recognize these events and combine the two-pixel event into the true single-interaction event [12]. Throughout this work, I will use the phrase 'two-pixel event' when I am referring to the signals read from the detector and 'two-interaction event' when I am referring to those that have had charge-sharing events combined in this way. It should be clear from context when I instead am referring to the true interactions in the detector.

Dead Layers

There are some regions in the detector crystal where a radiation interaction will create such a small signal that it is undetectable. The first is in the small region next to the anode electrode. Here, the electrons drift only a very short distance, inducing very little charge on the cathode and, although the anode weighting potential increases sharply at the anode surface, inducing less than the full charge on the anode as well. Depending on the amount of drifting charge, the signal on both the anodes and cathode may be below the noise-discrimination threshold, missing the interaction. Therefore, the anode dead layer thickness is a function of the deposited energy and the noise discrimination level, and it may also depend on the sub-pixel interaction position since the anode weighting potential varies in amplitude with lateral position at the same depth. This complex dependence makes an exact model difficult.

The behavior of charge liberated near the sides of the detector is poorly understood. The anode electrodes cannot extend exactly to the edge, so the electric field may be weak or variable in the edge regions, making it difficult to collect charge or determine depth accurately [13]. A better understanding of this region is needed.

Pixel Jumping

Previous studies have shown that exposing the cathode side of some crystals to a flood field of photons creates a much more non-uniform distribution of interactions per pixel than expected [13]. This is hypothesized to be due to pixel jumping, where electric-field direction variations within the crystal cause charge in one pixel to drift into a nearby pixel. Collimator experiments are consistent with this hypothesis [2]. The degree of pixel jumping varies between different crystals. Shifting the recorded lateral interaction position by several millimeters can cause considerable errors in some imaging applications. This effect needs further investigation so that it can be corrected or included in the model used for imaging and detection.

Other Effects

Besides those described above, there are a multitude of other complicating effects that have already been addressed satisfactorily in other works [1,2], including electrode amplitude calibration, non-linearity, time-amplitude walk, weighting-potential cross talk skewing the energies of nearby interactions, and peak-hold drop in analog shaping. Corrections are regularly made for these effects when reading events from the detector to achieve interaction parameters that better match the true parameters, leading to better imaging and detection performance. In a sense, these detectors and our understanding of their operation are the giants on which we can stand as we analyze their data to better understand the radiation environment and detect threats. Fittingly, in Greek mythology, the giant on whom the dwarf stood was the hunter Orion.

Chapter 3
Imaging Methods

To them, I said, the truth would be literally nothing but the shadows of the images.
– Plato, The Republic, Book VII

Passive and Active Imaging

Imaging using gamma rays can be classified into two categories: passive imaging and active-interrogation imaging. In active interrogation, an interrogation system produces radiation (usually gamma rays or neutrons) and aims it at potential targets. By recording what returns or passes through, one can determine properties of the material [14]. In medical imaging, measurement of the un-collided radiation is the basis for X-ray and CT systems. In security applications, this method is used for the active interrogation of containerized cargo such as the "nuclear car wash" [15], and the scattered radiation is used for backscatter imagers [16,17,18]. The former can record material attenuation and look for products from nuclear reactions that identify materials. Interrogation of targets can produce strong signals even for materials that naturally emit little or no radiation, such as explosive materials [19]. Additionally, SNM responds distinctly to interrogation with high-energy particles [15]. Nevertheless, concerns remain about portability and the safety to stowaways or bystanders, who may be exposed to high doses. Also, active interrogation cannot be used in all situations, such as crowded venues. For long standoff distances, very strong interrogating beams are required since beam strength and resulting signal will both fall off with the square of the distance.

Instead, this work will focus on passive detection and imaging – relying on the signals produced intrinsically to detect and locate threatening materials or other sources. Passive methods therefore are limited to those objects that emit sufficient radiation to produce a signal at the detector. In medical imaging, this process can be promoted by

administering radioactive tracers to a patient, which will give a sufficiently strong signal at the detector, as in SPECT or PET scanning. But, in most passive imaging scenarios, it is not possible to radioactively tag targets of interest. Passive imaging is therefore limited to radiation-emitting materials, specifically those that emit gamma rays or neutrons that are neither too distant nor too heavily shielded. Signals of interest therefore are often weak, so detectors must be sensitive.

Types of Radiation

When designing a passive system to image a radiation environment, it is important to understand the expected emissions and their interaction methods. No radiation can be detected unless it interacts directly or indirectly with a detector. Of particles emitted from radioactive materials, gamma rays are the most common and easiest to observe. This work will focus on detection and imaging using gamma rays. The vast majority of possible threat sources emit gamma rays. Gamma rays have a long range in air and relatively high interaction probability in certain detector materials. It is, of course, possible to shield gamma rays, but effective shielding requires high-Z material.

Alpha particles and fission fragments have a very short range even in air [3], so are hard to measure directly unless the detector is nearly touching the emitting material. Some recent work has shown that secondary ultraviolet emissions from alpha-particle-induced ionizations of air can be imaged from long distances; however, this requires a direct line of sight to a bare source [20]. Similarly, the range of beta particles in materials is short, so most source-packaging geometries prevent beta-particle emission. Neutrinos or antineutrinos are also produced, but these rarely interact and therefore are not practically detectable for most applications.

A number of threat sources produce neutrons. Like gamma rays, neutrons can travel long distances in air and then interact with reasonable efficiency with some detector materials. The natural background of neutrons is low, so neutron-emitting threat sources are quite easily distinguishable from background. However, not all threat sources of radiation emit neutrons, limiting their use primarily to SNM. Still, many of the techniques used to detect and image gamma rays have analogs that have been applied to neutrons [21,22,23].

Gamma-Ray Imaging Concepts

A gamma ray entering a detector may interact in a number of ways: through the photoelectric effect, Compton (incoherent) scattering, or pair production. Each can be used to image the incident gamma-ray field. In addition, it is possible for low-energy photons to undergo coherent scattering, by Rayleigh, Thomson, or Bragg scattering, which changes the photon's direction, but not its energy [24]. The following sections describe how each type of interaction mechanism can be used to image with any 3D-position-sensitive detector system.

Imaging Using Coherent Scattering

Coherent scattering is not useful for detection directly since no energy is deposited in the detecting material, but Bragg scattering can be used to focus X-rays onto a detector to image their incident directions. This is the key mechanism behind some X-ray focusing mirrors [25,26]. Yet, this technology is primarily applicable, in reasonable geometries, to very low-energy photons. Generally, we ignore coherent scatters in our system models because they tend to be prevalent with large scatter angles only at lower energies than are commonly imaged.

Imaging Using Pair Production

In pair production, a photon with energy over twice the electron's rest mass can create an electron and positron in the electromagnetic field of a nucleus. The electron and positron share the photon's initial energy. The positron will then go on to annihilate with an electron in the material to produce two almost-back-to-back almost-511-keV photons (or, in rare situations, any number of photons that sum to 1022 keV [27,28]), which subsequently may also interact in the detector. Therefore, events with the signature of one, or two, 511-keV interactions (or a set of interactions that add up to 511 keV) suggest a pair-production event. If all but one interaction can be explained as undoubtedly from the 511-keV photons, the initial full energy can be determined unambiguously as the energy in that interaction plus 1022 keV.

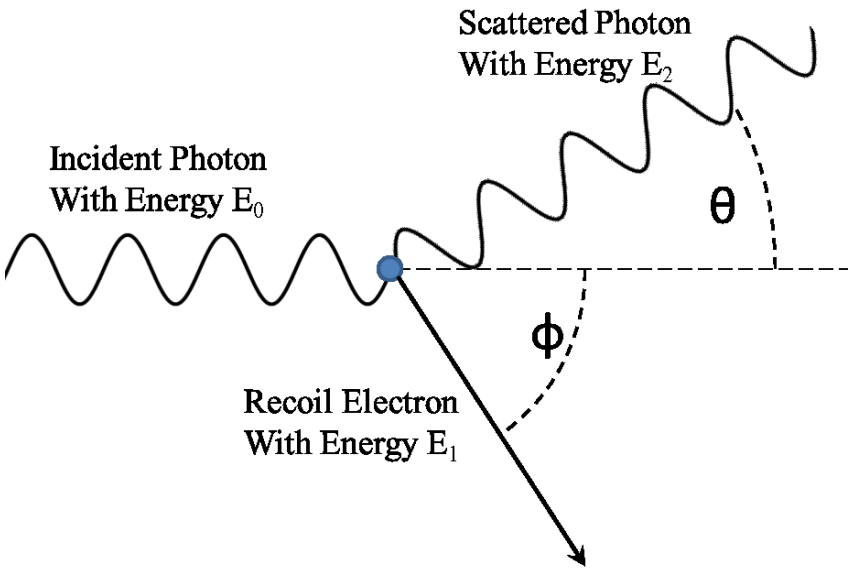

Figure 3.1. Compton scattering. A photon with energy E_0 transfers energy E_1 to an electron and scatters with energy E_2 at some angle θ from its original direction in the same plane as the electron scatter (neglecting Doppler broadening).

However, the direction of the back-to-back 511-keV pair is uncorrelated with the direction of the original incident photon, so, unless the initial direction of the electron and positron can be tracked, observing only a pair-production event gives no information about the spatial distribution of incident radiation besides from the attenuation distance to the pair-production location.

Imaging Using Compton Scattering

The Compton-scattering interaction is the basis for much of gamma-ray imaging that does not use collimation. If the incident gamma-ray energy is known, its incident direction can be determined to within the surface of one or two cones by recording the deposited energy at two interaction locations. With three or more interaction locations, it is theoretically possible to determine whether the full photon energy was deposited, and if it was, the photon's initial direction to within the surface of a cone.

In Compton scattering, a gamma ray elastically scatters with a weakly bound electron, as shown in Figure 3.1. Assuming the electron is free and at rest, conservation of energy and momentum uniquely define the relationship between the resulting energies

and scatter angle. The scatter angle θ as a function of incident, deposited, and scattered energies E_0, E_1, and E_2 respectively is

$$\theta = \cos^{-1}\left(1 - \frac{m_e c^2 E_1}{E_0 E_2}\right) \quad (3.1),$$

where $m_e c^2$ is the rest mass energy of an electron (511 keV). Alternatively, this can be expressed as the scattered photon energy in terms of the incident photon energy as

$$E_2 = \frac{E_0}{\frac{E_0}{m_e c^2}(1 - \cos\theta) + 1} \quad (3.2).$$

The distribution of scatter angles (or, equivalently, scattered energies) is predicted by the Klein-Nishina formula, and is more forward biased at high energies [3].

In practice, the weak binding of an outer-shell electron causes little deviation from these results [29]. However, the non-zero initial momentum of the electron causes an effect called Doppler broadening. Eqns. (3.1) and (3.2) describe an interaction in the electron's rest frame. In the lab frame, the scattered energies (or angles) are shifted due to the electron's initial motion. Since the initial electron momentum is unknown, it appears as a broadening in scattered energy at a particular angle. The effect roughly scales with atomic number and electron orbital structure and, at 662 keV in CZT, causes a few-degree FWHM in the scattered-angle distribution [30], with wide tails in the distribution [29].

One- and Two-Interaction Compton Events

A single Compton-scattering interaction is only useful for limiting the set of possible incident-photon directions if both the incident photon energy and the direction of the recoil electron are known. However, in most cases, it is hard to track the recoil electron direction, especially since it tends to change direction quickly as it slows. Very good detector spatial resolution is necessary to observe the initial electron direction [31,32].

If the detector records two interaction locations from a single incident photon, more information is available. In this case (except for the possibility of pair production), the first interaction must have been a Compton-scatter interaction and the second either a Compton-scatter or photoelectric interaction. For now, we will assume the interaction

sequence is known. With a recoil electron track measurement, the entire scatter collision can be reconstructed and the incident photon direction known. The plane of the two interactions and the electron track define the plane of the incident photon. The initial photon energy can be determined by the angle between the scattered photon direction (from the line between the interactions) and the electron track direction, which, together with energy deposited in the first interaction, gives the scatter angle.

Without a measurement of the recoil electron track, the incident direction is more ambiguous. Because the scattered photon has less energy than the original photon and the photoelectric probability increases with decreasing energy, it is likely that the second interaction will contain the entire scattered-photon energy. Therefore, it is often assumed that the total deposited energy is equal to the incident photon energy. In this case, the energy depositions can be used to calculate the scatter angle, using Eqn. (3.1). The plane of the scatter is still unknown, however, so the incident photon direction is known only to within the surface of a cone. It is, of course, possible that the second interaction could be a Compton-scatter interaction, in which case the assumed initial energy and scatter angle will be incorrect. Other energetically possible initial energies could be assumed as well.

Sequence Reconstruction

In most systems, the order of interaction is not directly measured since the time delay between interactions is only due to photon travel time. For small detectors, this is on the order of picoseconds. Therefore, there can be up to two cones, one for each possible sequence order. Often, one sequence is energetically impossible or can be eliminated based on assumptions of the direction to the source. For instance, when full-energy deposition is assumed and one interaction contains energy above the Compton edge at

$$E_{edge} = \frac{E_0}{\frac{m_e c^2}{2E_0} + 1} \qquad (3.3),$$

(the greatest amount of energy that can be deposited in a single Compton scatter when the photon backscatters), this interaction must be the second (photoelectric) interaction.

In many cases, however, the sequence order cannot be known definitively. There are two options. First, one could reconstruct both possible sequences, weighting each by

its probability to be correct, as described in the system model starting on page 21. Alternatively, one could choose one of these orders [29,33,34], as is commonly used in simple back-projection (see page 32). Methods of choose one of the orders usually fall into one of two categories: energy based or probability based. Energy-based sequence-reconstruction methods choose the best order based on which is most probable to occur based on the Klein-Nishina cross section, neglecting factors such as attenuation distance [33,34]. These tests are fast since they reduce to simple figure-of-merit calculations involving the recorded energies. One such test simply chooses the interaction with the highest energy as the first interaction [29]. Probability-based methods also include probability factors such as attenuation distance in hopes of choosing a more likely order given the exact detector geometry, although not much improvement has been observed for two-interaction events detectors of 15 mm × 15 mm × 10 mm [29].

Three- and Many-Interaction Events

When n interactions appear in the detector, at least $n-1$ of them must be Compton-scatter interactions (neglecting the possibility of pair production). If the order of interactions is known, the energy and scatter angle are known for all but the first and last interactions from the interaction locations. Therefore, Eqn. (3.2) can be solved for the initial energy, and the original photon's direction can be determined, again, to the surface of a cone. With observed electron tracks, the cone ambiguity is removed.

Sequence Reconstruction

When the interaction sequence is unknown, there are up to $n!$ possible sequences and as many possible cones. Again, some of these can be eliminated if they are energetically impossible. A number of methods can be applied to determine the best sequence of three or more interactions [34,35,36,37,38]. Since the only parameters needed for correct reconstruction of the photon direction are the initial energy and the first two interaction locations, the sequence of interactions past the second or third interaction usually do not need to be determined. Nevertheless, at best, sequence reconstruction still chooses the incorrect order a small fraction of the time, so it will not reconstruct intensity in the correct direction for all events.

The best sequence is often found by assuming that the photon was completely stopped in the detector and choosing the sequence with the best match between the middle angle(s) calculated by energy θ_E and by interaction geometry θ_R. That is, the sequence is chosen that has the biggest figure of merit (FOM)

$$FOM = \frac{1}{\sqrt{\sigma_{\theta_E}^2 + \sigma_{\theta_R}^2}} e^{-\frac{(\theta_E - \theta_R)^2}{2\left(\sigma_{\theta_E}^2 + \sigma_{\theta_R}^2\right)}} \quad (3.4),$$

where σ_{θ_E} and σ_{θ_R} are the modeled uncertainties in the scatter angles. With more interactions, the FOM is the products over all the middle angles. This is known as the minimum-squared-difference (MSD) method [34]. Other FOMs are possible [36].

Other Available Information in Compton Scattering

It should also be mentioned that apart from the interaction angles, the interaction locations can tell us something about the source direction of the photon based on the interaction probability in surrounding or detector material. This attenuation component is described more in the Imaging Methods section starting on page 21.

Imaging Using the Photoelectric Effect

In the photoelectric effect, a gamma ray interacts with an atom, ejecting an electron with energy equal to the incident gamma-ray energy less the electron's binding energy. An X-ray or Auger electron later carries away the binding energy when the vacancy is filled. The photoelectron emission direction tends to be at right angles to the incident photon at low energies and in the same direction as the incident photon at higher energies. Around 100 keV, the most likely emission angle is about 45°. However, the electron emission angle is not deterministic, but falls in a broad probability distribution, especially at low energies [24]. Therefore, even if the direction of the emitted electron can be determined, the incident photon direction can be known with a high confidence only at high energies.

As in Compton scattering, the electron track is often hard to observe. In addition, an interaction must be identified as a photoelectric interaction, so as not to confuse the

interaction with a Compton scatter. This can only be done some of the time – when an X-ray is emitted that deposits its energy significantly far from the original interaction site.

Though the measurement of each interaction may not specify the incident photon direction, the distribution of these interaction locations gives information about the flux distribution by considering the attenuation distance to each interaction location. This is the basis for coded-aperture imaging, the centroid method, and single-pixel imaging, described below (starting on pages 19, 20, and 60).

Coded-Aperture Imaging

Coded-aperture imaging uses the shadows cast by a known arrangement of attenuating material to estimate the distribution of incident photons. Like Plato's cave, the emissions cannot be directly observed and so we observe the shadows. The simplest coded aperture is a pinhole camera or its inverse. A pinhole camera consists of a block of attenuating material thick enough to block most gamma rays, with a single hole in it. A position-sensitive detector behind this hence knows each photon's origin by tracing the interaction location back through the hole, allowing the angular distribution of incident photons to be reconstructed. Because such a design is quite inefficient (most photons never make it to the detector), common coded-aperture designs increase the number of open regions in the mask until half of the mask is open [39]. Still, half of the photons are obstructed, so some designs use active masks – mask elements that are themselves detectors. While active mask elements may not change imaging performance, counting efficiency is improved. If the detector plane is itself a mask, coded-aperture imaging is possible in both directions from the detector planes [40].

Many authors have studied the optimal arrangement of the mask elements to most accurately reconstruct images of the source distributions [41,42,43], but it is still an open problem, especially if reconstruction will occur with probabilistic reconstruction methods [44]. In designing the mask structure, one chooses the properties of the mask for ease of reconstruction or image quality.

Coded-aperture imaging relies on significant attenuation within the mask material of the imaged gamma rays. Therefore, coded apertures find most application when the

gamma-ray energy is relatively low, so that the mask thickness is not large compared to the hole size, which will introduce edge effects [26].

Coded-aperture masks can also be combined with detectors designed for Compton imaging [45,46,47]. The mask improves Compton-imaging performance, at the cost of efficiency, by reducing the directions from which each event's photon was probable to originate. It also allows single-interaction events to be used for imaging, as a standard coded-aperture imager. Single-interaction events would otherwise not contribute to the image besides with the attenuation through the detector components, as described in the centroid method and single-pixel imaging sections (pages 20 and 60).

The Centroid Method

The distribution of the first interaction locations within a material follows an exponential distribution in distance x from the material surface that the photons strike. Mathematically,

$$N(x) = N_0 \mu \, e^{-\mu x} \tag{3.5}$$

with attenuation constant μ dependent on the material and energy, and a total number of interactions along a ray N_0. The attenuation constant μ generally increases with decreasing energy. If the interaction locations can be accurately recorded and the incident energies known, the interaction-location distribution could be used to determine the directional distribution of gamma rays striking the detector. The centroid method [48], and other similar methods for neutrons [22,49,50], make use of this fact. Comparing the first moment of the interaction distribution at a particular energy with the geometric first moment gives the direction to a single source at that energy. This method only works for a single source at each energy and will give erroneous results if multiple sources are present.

Several complications may arise. First, surrounding materials and non-active regions of the detector must be accurately modeled, and the interaction locations must be accurately measured. Dead layers, surrounding materials, or effects that shift the apparent interaction location in the detector will shift the centroid, resulting in an incorrect shift in the apparent source direction [48]. A more complicated model or experimental corrections are needed to correct for these detector effects.

Second, only the distribution of the first-interaction locations follows the exponential distribution. Later interactions are distributed exponentially radially around the first interaction location. Figure 5.8 shows this clearly for the 18-detector system. Although the distribution of later interactions may still be biased in the source direction, the distribution is smeared out compared to the distribution of first-interaction locations.

In general, it is not possible to identify the first interaction with certainty. Taking full-energy-deposition events that interacted only once guarantees that each must be the first (and only) interaction because all the energy is deposited in one interaction. Practically, these events are selected by choosing single-interaction events in the photopeak region. Still, this usually also includes continuum events under the photopeak, which are single-interaction events but not full-energy depositions. The distribution of such events tends to be biased towards the outer surfaces of the detector where it is most probable for the scattered photon to be lost. That is, first interactions near the center of the detector tend to become multiple-interaction events as the scattered photon interacts. Therefore, even taking just single-interaction events in the photopeak region may introduce bias in the centroid due to the detector geometry. Yet, for a strong source compared to background and a detector with large peak-to-total ratio, this bias is small. For multiple-interaction events, a sequence-reconstruction method, such as those described on page 15, can be used to estimate the most-probable first interaction, but usually the first interaction cannot be known with certainty.

Imaging Methods

Image-Formation Model

The general imaging problem is to reconstruct parameters based on measurements. In our case, this is reconstruction of the source distribution. Mathematically, we often assume a linear relationship between the source distribution and the measurement results, as

$$\bar{g}(\vec{r}) = \int_{\vec{s}} \mathcal{T}(\vec{s};\vec{r}) f(\vec{s})\, d\vec{s} \qquad (3.6),$$

which describes the expected count density $\bar{g}(\vec{r})$ produced around event attribute vector $\vec{r}$ in the detector due to a source distribution $f(\vec{s})$ around source-space parameter $\vec{s}$. That is, the mean number of counts expected in an attribute bin around $\vec{r}$ can be found by integrating the expected count density $\bar{g}(\vec{r})$ over the volume of the attribute bin and applying the appropriate measurement-time scaling. The transfer function $\mathcal{T}(\vec{s};\vec{r})$ gives the actual probability of a photon from near $\vec{s}$ to produce a recorded event near $\vec{r}$. It includes detector sensitivity and detector effects that may shift the reported attributes of an event from the true attributes to different measured attributes. Spurious events not arising from gamma-ray interactions are ignored.

Attributes measured by the detector may be three-dimensional position and energy of each of N interactions, in which case $\vec{r} = (N, x_1, y_1, z_1, E_1, x_2, y_2, z_2, E_2, \ldots x_N, y_N, z_N, E_N)$, and source distribution parameters may be direction and energy, in which case $\vec{s} = (\theta, \phi, E)$. More (or fewer) dimensions could parameterize the source distribution and detector as well. For instance, the source distribution may include polarization of gamma rays, a three-dimensional spatial distribution, or time. The detector may also be able to record other information such as the direction of the secondary electron in the interaction.

Any elements that cause count-rate-dependent transfer functions between the source distribution and the measurements will make this relationship between source and events non-linear. In our system, these include effects such as read-out dead time, coincidence events, and space-charge buildup in the detector. As long as operation occurs only in relatively low-count-rate environments, these factors are small.

In practice, the continuous functions in Eqn. (3.6) are approximated by discrete bins both in our measurements from the detector and in our reconstruction of the source distribution. Systematic uncertainties from calibration and detector performance are included in the system matrix $\mathcal{T}$. So,

$$\bar{g}_{[I\times 1]} = \mathcal{T}_{[I\times J]} f_{[J\times 1]} \tag{3.7}$$

where $f_{[J\times 1]}$ is the gamma-ray-source distribution in J source bins, $\bar{g}_{[I\times 1]}$ is the expected number of counts as reported by the detector in each of I detector bins, and $\mathcal{T}_{[I\times J]}$ describes the fraction of photons in each bin of the source distribution that go on to create

each event. One can think of $\mathcal{T}_{[I\times J]}$ as containing elements that describe the probability of event creation

$$\mathcal{T}_{ij} = p(\text{event in bin } i \text{ detected}|\text{photon from } j) \quad (3.8).$$

The number of counts reported by the detector $g_{[I\times 1]}$ follows a Poisson distribution with mean $\bar{g}_{[I\times 1]}$. We wish to estimate $f_{[J\times 1]}$ based on limited events in $g_{[I\times 1]}$. This matrix equation (3.7) may be under- or over-determined. We would like imaging to work in both cases.

System Model

The true system matrix $\mathcal{T}$ describes the actual probability of recording a particular event. We attempt to use a system matrix *t* that is as close to this as possible. However, a number of factors are difficult to include and a perfect match is not possible. Page 8 describes some of these model-mismatch factors in more detail for our system.

We choose to treat the detector output from our system as a set of interaction energies and locations with uncertainty in each determined by a simple equation. Fortunately, others' vast efforts to determine the energies and locations correctly and precisely (as introduced in Chapter 2) give a strong starting point for the system-matrix calculation. Starting with these data, interaction physics can be used to calculate the likelihood that the photon came from each image bin. Xu and He [51], and Wang *et al.* [52] have extensively described equations to calculate this approximate system matrix *t* for a 3D-position-sensitive spectrometer when the image is binned over directions and energy. Others have presented similar equations [53,54].

However, it should be noted that the detector output is actually a set of electronic signals that have undergone extensive processing to give interaction locations and energies that are as accurate as possible. In some cases, ambiguous events can be interpreted as one kind of event when the same signals are also consistent with another type of event. By considering the raw detector signals from an event and propagating probability from these, rather than from the reconstructed energies and locations, a better system model may be possible, at the expense of increased computation.

According to Wang *et al.* [52], uncertainty and pixelation in interaction positions and energy contribute significantly only to limited portions of the system matrix that

relates recorded interaction energies and locations with an image in space and energy. These terms that include uncertainty and pixelation describe whether the direction to the next interaction is along the cone originating at the previous interaction position and whether the energy of interest is equal to the total measured energy. Therefore, the system-matrix calculation primarily follows each step of the interaction physics without additional terms due to integrals over the uncertainty and pixelation regions. The event probabilities depend on the number of interactions that were recorded.

For single-interaction events,

$$t_{ij} \approx e^{-\int_{\vec{x}_j}^{\vec{x}_1} \mu_{E_j}(\vec{x})d\vec{x}} \left(\frac{\mu_p\left(E_j, \vec{x}_1\right)}{\sqrt{2\pi\sigma_{E_1}^2}} e^{-\frac{\left(E_j - \mathrm{E}_1\right)^2}{2\,\sigma_{E_1}^2}} + \frac{d\sigma_c\left(E_j, E_1, \vec{x}_1\right)}{d\mathrm{E}_{\mathrm{dep}}} e^{-\int_{\vec{x}_1}^{\infty} \mu_{E_j - E_1}(\vec{x})d\vec{x}} \right) \tag{3.9}$$

where

$$\begin{aligned} \frac{d\sigma_c\left(E_j, E_1, \vec{x}_1\right)}{dE_{dep}} &= \frac{2\pi m_e c^2}{\left(E_j - E_1\right)^2} \frac{R_0^2}{2} N_0(\vec{x}_1) \left[\frac{E_j - E_1}{E_j} \right. \\ &\quad - \left(\frac{E_j - E_1}{E_j} \right)^2 \left(1 - \left(1 - \frac{m_e c^2 E_1}{E_j\left(E_j - E_1\right)} \right)^2 \right) \\ &\quad \left. + \left(\frac{E_j - E_1}{E_j} \right)^3 \right] \end{aligned} \tag{3.10}$$

is the Compton attenuation coefficient for a photon of energy E_j to deposit energy E_1 at location $\vec{x}_1$. The other variables are defined in Table 3.1. All factors that will be the same for every image bin *j* for a particular event *i* have been removed from Eqn. (3.9) for computational efficiency since they will cancel when used in the MLEM algorithm described below (on page 34). Physically, the first term of Eqn. (3.9) is the probability for the photon to survive to the first interaction location. The parenthetical term is the sum of the probability for the photon to undergo photoelectric absorption such that the

Table 3.1. Description of variables used in this section.

Variable	Definition
$\mu_E(\vec{x})$	Total attenuation coefficient at energy E in material at position $\vec{x}$
$\mu_p(E,\vec{x})$	Photoelectric attenuation coefficient at energy E in material at $\vec{x}$
$E_j\,,\vec{x}_j$	Energy and direction (or position) associated with image bin j
E_n^s	Recorded energy deposited in n-th interaction when sequence is s
$\sigma_{E_n^s}$	Expected uncertainty in recorded energy in n-th interaction for sequence s
$\vec{x}_n^s$	Position of n-th interaction when sequence is s
$m_e c^2$	Rest-mass energy of the electron (511 keV)
R_0	Classical electron radius ($2.817940285 \times 10^{-13}$ cm)
$N_0(\vec{x})$	The material electron density at $\vec{x}$

deposited energy is equal to the assumed incident energy within the energy uncertainty, and the probability for the photon to undergo Compton scattering to deposit some of its energy and then exit the detector. The integral in the term for the attenuation leaving the detector is along an average path along the possible cone of directions leaving the detector. Figure 3.2 shows the directional portion of a row of the system matrix for a single-interaction event in the 18-detector array assuming full-energy deposition.

For two-interaction events,

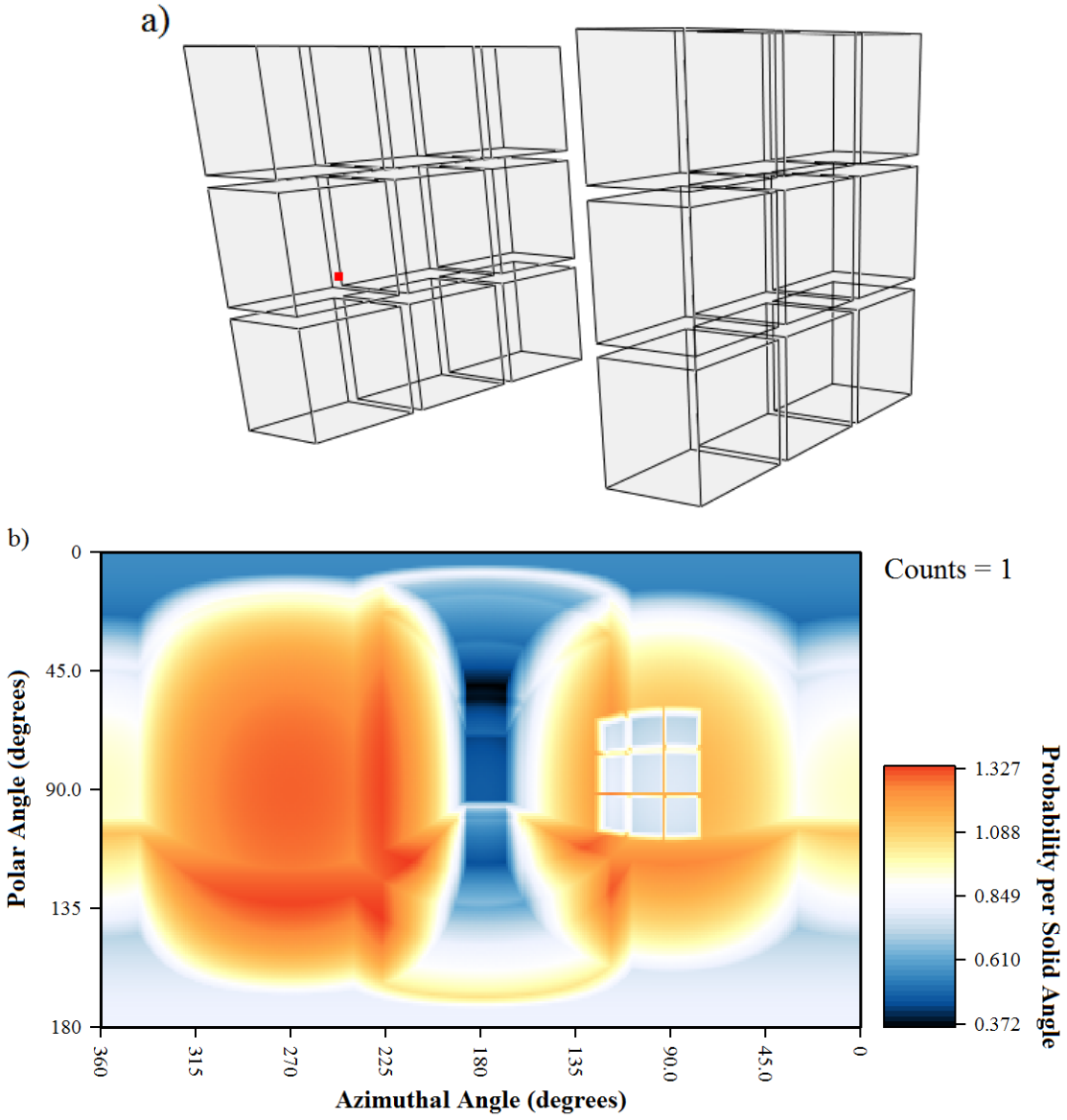

Figure 3.2. Example single-interaction system matrix elements. a) The 1565-keV interaction location, marked in red, in the 18-detector array. b) The spatial portion of the system matrix for this event, assuming full-energy deposition. Each system-matrix element corresponds to one pixel. The direction perpendicular to the faces of the planes corresponds to 90° and 270° in the azimuthal angle and 90° in the polar angle. One can see the attenuation due to the other detectors in the image.

$$t_{ij} \approx \sum_{s=1}^{2}\Bigg[e^{-\int_{\vec{x}_j}^{\vec{x}_1^s}\mu_{E_j}(\vec{x})d\vec{x}} \cdot \frac{d\sigma_c\left(E_j, E_1^s, \vec{x}_1^s\right)}{d\mathrm{E}_{\mathrm{dep}}} \cdot \frac{1}{\sigma_{\theta_1^s}\sin(\theta_1^s)} e^{-\frac{\left(\theta_j^s-\theta_1^s\right)^2}{2\,\sigma_{\theta_1^s}^2}} \cdot e^{-\int_{\vec{x}_1^s}^{\vec{x}_2^s}\mu_{E_j-E_1^s}(\vec{x})d\vec{x}} \cdot \Bigg(\frac{\mu_p\left(E_j-E_1^s, \vec{x}_2^s\right)}{\sqrt{2\pi\left(\sigma_{E_1^s}^2+\sigma_{E_2^s}^2\right)}} e^{-\frac{\left(E_j-\mathrm{E}_1^s-E_2^s\right)^2}{2\left(\sigma_{E_1^s}^2+\sigma_{E_2^s}^2\right)}} + \frac{d\sigma_c\left(E_j-E_1^s, E_2^s, \vec{x}_2^s\right)}{d\mathrm{E}_{\mathrm{dep}}} e^{-\int_{\vec{x}_2^s}^{\infty}\mu_{E_j-E_1^s-E_2^s}(\vec{x})d\vec{x}} \Bigg)\Bigg] \tag{3.11}$$

where the sum is over the possible interaction sequences, θ_j^s is the angle from the cone axis to image pixel j, θ_1^s is the scatter angle at the first interaction location, calculated with Eqn. (3.1), and $\sigma_{\theta_1^s}$ is the uncertainty in the cone angle relative to a fixed cone axis, as described by Xu et al. [33]. This uncertainty includes factors due to uncertainty in the recorded energies and interaction locations. Physically, Eqn. (3.11) combines the probabilities for the photon to survive to the first interaction location, Compton scatter, go in the right direction along the cone of ambiguity to reach the second interaction location, survive the material attenuation to that location, and then either be absorbed through the photoelectric effect or Compton scatter and leave the detector. The term describing the agreement between the image pixel and the scatter angle stands in for the probability that the calculated scatter angle is the correct size to hit the second interaction location. The geometric attenuation inversely proportional to the distance between the two interactions is independent of the source bin and sequence, so it is not included. Figure 3.3 shows the directional portion of a row of the system matrix for a two-interaction event in the 18-detector array assuming full-energy deposition.

For three or more interactions, the form is similar,

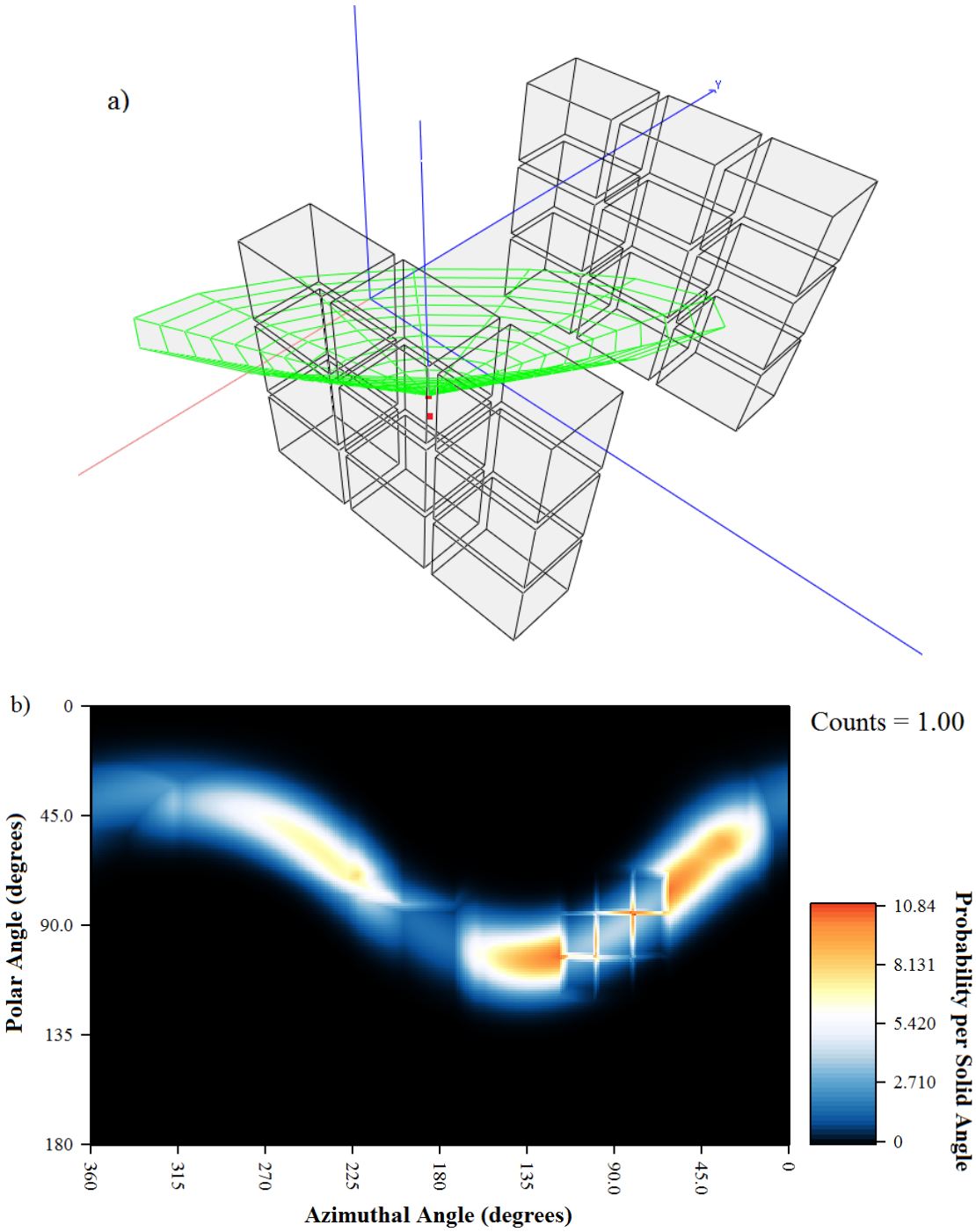

Figure 3.3. Example two-interaction system matrix elements. a) The 390.9-keV interaction locations, marked in red, in the 18-detector array. b) The spatial portion of the system matrix for this event, assuming full-energy deposition. Each system-matrix element corresponds to one pixel. The direction perpendicular to the faces of the planes corresponds to 90° and 270° in the azimuthal angle and 90° in the polar angle. One can see the possible directions along the cone and the attenuation due to the other detectors in the image.

$$
\begin{aligned}
& t_{ij} \\
& \approx \sum_{s=1}^{N!} \Bigg[e^{-\int_{\vec{x}_j}^{\vec{x}_1^s} \mu_{E_j}(\vec{x}) d\vec{x}} \cdot \frac{d\sigma_c\left(E_j, E_1, \vec{x}_1\right)}{d\mathrm{E}_{\mathrm{dep}}} \\
& \cdot \frac{1}{\|\vec{x}_2^s - \vec{x}_1^s\| \, \sigma_{\theta_1^s} \sin\left(\theta_{1,e}^s\right)} e^{-\frac{\left(\theta_j - \theta_{1,e}^s\right)^2}{2\,\sigma_{\theta_{1,e}^s}^2}} \cdot e^{-\int_{\vec{x}_1^s}^{\vec{x}_2^s} \mu_{E_j - E_1^s}(\vec{x}) d\vec{x}} \\
& \cdot \prod_{m=2}^{N-1} \Bigg(\frac{d\sigma_c\left(E_j - \sum_{n=1}^{m-1} E_n^s, E_m^s, \vec{x}_m^s\right)}{d\mathrm{E}_{\mathrm{dep}}} \\
& \cdot \frac{1}{\|\vec{x}_{m+1}^s - \vec{x}_m^s\| \, \sigma_{\theta_m^s} \sin\left(\theta_{m,e}^s\right)} e^{-\frac{\left(\theta_{m,e}^s - \theta_{m,g}^s\right)^2}{2\,\sigma_{\theta_m^s}^2}} \\
& \cdot e^{-\int_{\vec{x}_m^s}^{\vec{x}_{m+1}^s} \mu_{E_j - \sum_{n=1}^{m} E_n^s}(\vec{x}) d\vec{x}} \Bigg) \\
& \cdot \Bigg(\frac{\mu_p\left(E_j - \sum_{n=1}^{N-1} E_n^s, \vec{x}_N^s\right)}{\sqrt{2\pi \left(\sum_{n=1}^{N} \sigma_{E_n^s}^2\right)}} e^{-\frac{\left(E_j - \sum_{n=1}^{N} E_n^s\right)^2}{2\left(\sum_{n=1}^{N} \sigma_{E_n^s}^2\right)}} \\
& + \frac{d\sigma_c\left(E_j - \sum_{n=1}^{N-1} E_n^s, E_N^s, \vec{x}_N^s\right)}{d\mathrm{E}_{\mathrm{dep}}} e^{-\int_{\vec{x}_N}^{\infty} \mu_{E_j - \sum_{n=1}^{N} E_n^s}(\vec{x}) d\vec{x}} \Bigg) \Bigg]
\end{aligned}
\tag{3.12}
$$

where $\theta_{m,e}^s$ is the scatter angle at interaction *m* determined using the deposited energies (using Eqn. (3.1)), $\theta_{m,g}^s$is the scatter angle at interaction *m* determined through the geometry of the interactions, and $\sigma_{\theta_m^s}^2$is the quadrature sum of uncertainty in the two determinations of the scatter angle. Again, terms that are the same for each image pixel *j* have been removed. The terms are similar to those for two-interaction events, with the addition of the inverse of the distance between interactions to represent the probability that the next interaction falls on one portion of the cone that distance away. The term $\|\vec{x}_{m+1}^s - \vec{x}_m^s\| \sin\left(\theta_{m,e}^s\right)$ is the circumference of the cone where it passes the next

interaction; the larger the circumference the smaller the probability the measured interaction position was on the direction of scattered photon. In addition, interactions other than the first and last include a Gaussian-comparison term of the match between the scatter angle calculated by energy and that calculated by interaction locations. This is analogous to the Gaussian-comparison term comparing θ_j^s and θ_1^s for the first interaction.

Simple Back-Projection

Many techniques exist to solve the system of equations in Eqn. (3.7). If $I = J$, it may be possible to invert the system matrix t using one of a number of methods [55]. Then, neglecting statistical noise,

$$\hat{f}_{inv} = t^{-1}g \approx t^{-1}\bar{g} = t^{-1}\mathcal{T}f \approx f \tag{3.13},$$

and the estimate of the source distribution is exact, as long as the modeled system matrix t matches the true system matrix $\mathcal{T}$. Without infinite statistics, the statistical noise in $g_{[I\times 1]}$ skews the result.

However, the system matrix t may not always be invertible or well conditioned and certainly will not always be square. One can always apply the transpose of the system matrix to both sides of Eqn. (3.7). Think of applying the transpose to the event vector as, for every event, putting intensity in the image proportional to the chance that the event was produced by a photon from that image pixel, then summing over all events. Then, neglecting statistical noise

$$t^T g = t^T \mathcal{T} f \tag{3.14}.$$

The simple back-projection (SBP) method says to estimate the source distribution as

$$\hat{f}_{SBP} = t^T g = t^T \mathcal{T} f \tag{3.15},$$

and hence, it assumes that

$$t^T \mathcal{T} \approx I \tag{3.16},$$

where I is the identity matrix. Practically, since each element of $\mathcal{T}$ is nonnegative, Eqn. (3.16) means that each event can only be produced by a photon from one of the bins in the source distribution. A pinhole camera would give this type of response, for instance. The more possible source distribution bins which could be responsible for a given event, the more blur will be included in the estimated image.

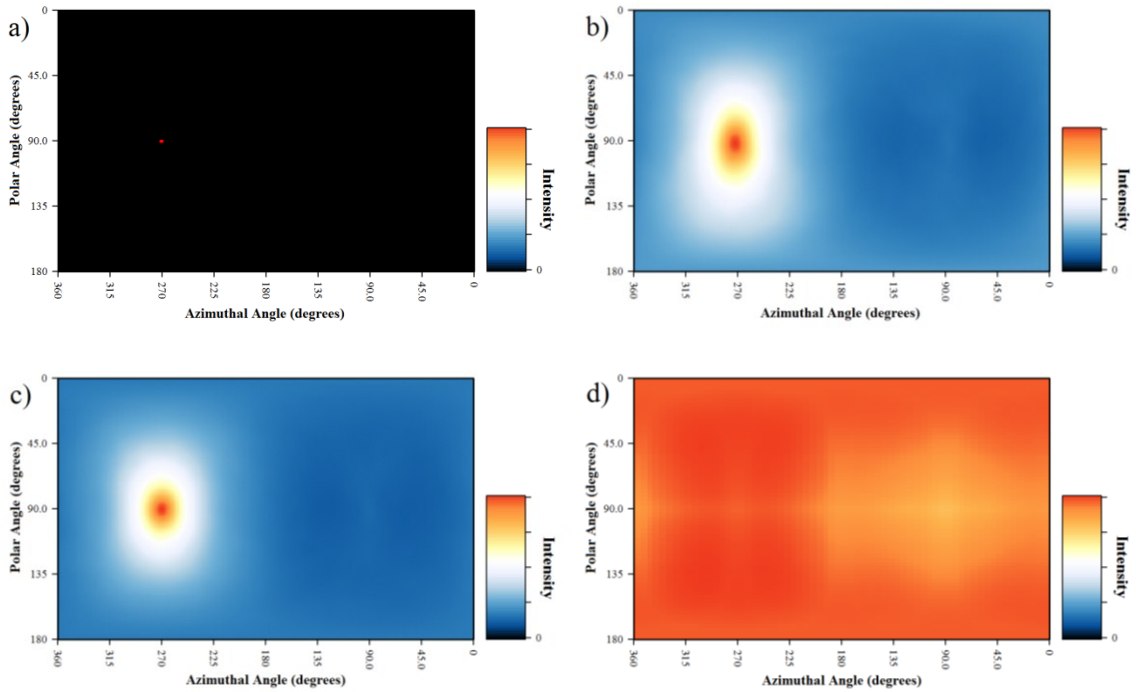

Figure 3.4. Examples rows of $t^T t$ and $t^T \mathcal{T}$ for a single 3D-position-sensitive CZT detector crystal. (a) shows the response from a system that obeys Eqn. (3.16). (b) represents a row of $t^T \mathcal{T}$ for a single point source near 90° and 270° in the polar and azimuthal directions using experimental multiple-interaciton events in the detector. (c) shows an approximation of the same row of $t^T t$ found by applying t^T to a simulated data set of multiple-interaction events. Finally, (d) shows the same row of $t^T t$ when using only single-interaction events, which supply less information about the source direction.

Figure 3.4 shows some examples of a row of $t^T t$ and $t^T \mathcal{T}$. These are known as point spread functions (PSFs) or point kernel functions (PKFs). It is clear in Figure 3.4b that our detector system adds blur beyond a detector satisfying Eqn. (3.16) shown in Figure 3.4a. Even eliminating most of the model mismatch in the true detector response as $t^T t$ in Figure 3.4c, there is significant blur due to reconstruction with SBP. The FWHM of the PSF depends on source energy, location, and the type of events used and can range from below 10° to over 50° for multiple-interaction events in the 18-detector system. Using less-informative events, such as single-interaction events, introduces even more blur, as shown in Figure 3.4d.

SBP System Matrix

For the best SBP angular resolution, one should form narrow PSFs to best approximate Eqn. (3.16). In this case, there is no reason why the approximate system matrix t needs to be similar to the true system matrix $\mathcal{T}$, and, in fact, the best images occurs when they are purposefully different. Intuitively, whenever back-projection intensities that will not contribute to the true direction are removed without greatly reducing the intensity in the true direction, the SBP image will improve, especially if the intensities are removed from bins far from the source location. The true system matrix, therefore, has not changed, but the back-projection matrix t^T becomes sparser, which, if the most important components are kept, better approximates Eqn. (3.16). However, care must be taken that certain directions or energies are not biased by consistently removing contributions to them from all events.

Omitting uncertainties in the SBP t^T is one way to form narrower PSFs. Sampling events which include measurement uncertainty, then back-projecting the possible source locations while also including this uncertainty tends to double-count the uncertainty contribution, making a more blurry image then necessary. Instead, during back-projection with sufficient counts, SBP angular resolution can be improved by just drawing a thin cone using the best estimates of the interaction positions and energies. The uncertainty contribution should not be included, except if the detector is incapable of recording events for which back-projection will contribute to the source bins in the omitted uncertainty region. In this case, weightings should be used to correct for efficiency, or a more advanced reconstruction method should be used. Nevertheless, well-designed imaging detectors rarely have such varying sensitivity over the region they are designed to image and they usually produce similar numbers and types of events for each image bin. Therefore, these are able to neglect uncertainty in SBP. However, it should be noted that with limited counts, including uncertainty in SBP offers smoothing that otherwise would be accomplished by simply recording more events.

Forming narrow PSFs is also the reason to use sequence reconstruction, as described starting on page 16. By back-projecting only the most-likely cone, significant intensity from (usually) the wrong direction is eliminated while not changing the intensity in the true direction as much. If the source truly was in the other direction, most other

events will (correctly) add intensity to that direction (again, assuming similar numbers and types of events for each image bin). Extending further to just back-projecting the most-likely direction on the cone is not used because some directions would be skewed hotter due to the detector geometry and dead layers. In addition, attenuation weights through the material are usually not included in the SBP cones. This is because, for many detector geometries, this tends to bias directions with a larger detector surface area more highly without significantly reducing intensity in the incorrect directions.

Filtered Back-Projection

Instead of making the assumption of Eqn. (3.16), one could instead invert the PSF $t^T\mathcal{T}$ ($\approx t^T t$) since it is square. As long as matrix t has linearly independent column vectors, $t^T t$ has linearly independent column vectors, and $t^T t$ is therefore invertible. In other words, the event distributions created by each possible source distribution must be different; it should not be possible to produce one row of $t^T t$ by simply combining other rows. Visually, each row of $t^T t$, as in Figure 3.4, should appear as a PSF around a different location. This is often satisfied. Then, ignoring noise

$$\begin{aligned}\hat{f}_{FBP} &= (t^T t)^{-1} t^T g \\ &\approx (t^T\mathcal{T})^{-1} t^T \bar{g} = (t^T\mathcal{T})^{-1}(t^T\mathcal{T}) f = f\end{aligned} \tag{3.17}.$$

The PSFs can be found through measurement, or through simulation of point sources in each image bin followed by application of Eqn. (3.15) [56,57]. Alternatively, mathematical methods can be used [58,59]. This is called filtered back-projection (FBP). One can think of FBP as first converting events back to image space using SBP, then applying a filter to transform the SBP image to a better estimate of the source distribution [57]. FBP could be applied to spatial and energy space as well. When it is applied to just the energy dimension, it is known as spectral deconvolution.

Generally, this shift-variant deconvolution is computationally difficult since $t^T\mathcal{T}$ can be large. Still, the inversion only needs to be performed once in preprocessing. If, instead, each PSF is simply shifted for each bin in the source space, the system is a linear shift invariant and significant simplifications can be made. Even when the PSF is not invariant, local invariance is often used to simplify the problem [57].

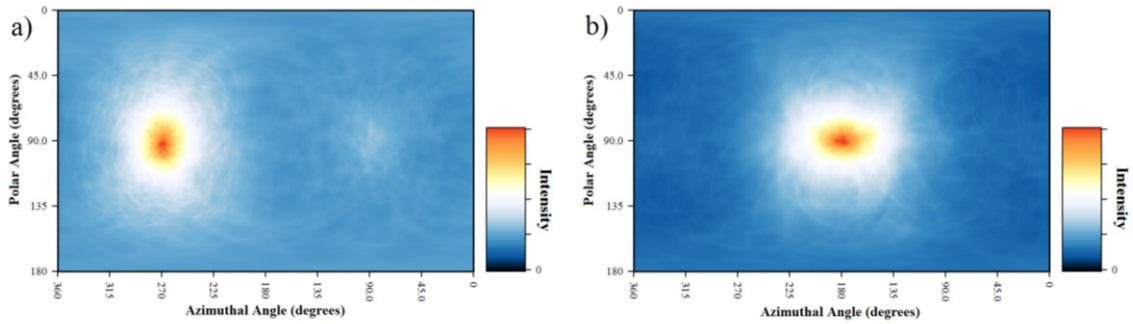

Figure 3.5. The experimental PSFs for two different locations around the 18-detector system using a Cs-137 source. It is clear that the PSFs between these two locations are quite different. a) The source at (270°, 90°). b) The source at (180°, 90°).

When the PSF is shift invariant, each row of $t^T\mathcal{T}$ is shifted relative to the row above it and the matrix multiplication in Eqn.(3.17) can be computationally simplified using

$$F_k = B_k G_k \tag{3.18}$$

where F_k is the Fourier transform of $\hat{f}_{FBP}$, B_k is the Fourier transform of $(t^T\mathcal{T})^{-1}$, and G_k is the Fourier transform of $t^T g$. Similar transforms using spherical harmonics are used in a spherical image space [58,60].

In a single-detector system, the PSF is approximately shift invariant [29]. However, this is not true for the whole 18-detector system. Two PSFs are shown in Figure 3.5 for the 18-detector system. It is clear that the shape of each is quite different, so assuming a shift-invariant PSF will produce distorted results.

One can show that Eqn. (3.17) is the least squares solution [55]. That is

$$\hat{f}_{FBP} = \underset{f}{\operatorname{argmin}} \ \|tf - g\|_2 \tag{3.19}.$$

Maximum Likelihood

Filtered back-projection attempts to find the image that best matches the measurements in the least-squares sense. However, this is just one possible choice; one could find the best match in the least-cubes sense or least-absolute-distances sense instead, for example. Maximum likelihood (ML) attempts to find the best match between the measurements expected from the estimated image and the actual measurements by equally penalizing any error over a tolerance level and assuming all parameter values are

equally probable. Visually then, when finding the best parameter estimate if the probability of the measurements is a complicated function of the parameter value, the least-squares estimate would be the mean of the parameter on the probability versus parameter-value plot, while the maximum-likelihood estimate would choose the parameter that contains the largest area in a small region around it on that plot.

Equally penalizing any parameter error implies that one should choose the maximum of the posterior density [61]. Mathematically,

$$\begin{aligned} \hat{f}_{MAP} &= \underset{f}{\operatorname{argmax}}\ p(f|g) \\ &= \underset{f}{\operatorname{argmax}} \frac{p(g|f)p(f)}{p(g)} \\ &= \underset{f}{\operatorname{argmax}}\ p(g|f)p(f) \end{aligned} \tag{3.20},$$

where the last step is possible because for each experiment, the set of events is fixed. This is known as the maximum a posteriori (MAP) estimator. If we assume no prior knowledge of the source distribution, the prior density of the image is uniform and Eqn. (3.20) reduces to the maximum-likelihood estimate

$$\hat{f}_{ML} = \underset{f}{\operatorname{argmax}}\ p(g|f) \tag{3.21}.$$

It is important to note that the MAP and ML estimates depend on the modeled distribution of statistical and random variation in the measured events, whereas these were ignored in SBP and FBP. If, for instance, the number of counts in each detector bin was modeled as Gaussian distributed with a constant variance σ^2 in each detector bin – a non-physical assumption – the conditional probability density function would be a joint Gaussian with independent parameters and the maximum-likelihood source estimate would be

$$\hat{f}_{ML} = \underset{f}{\operatorname{argmax}} \frac{1}{\left(\sqrt{2\pi}\right)^{\mathrm{I}} \prod_{i=1}^{I} \sigma} e^{-\sum_{i=1}^{I} \frac{1}{2\sigma^2} \left(\sum_{j=1}^{J} t_{ij} f_j - g_i\right)^2} \tag{3.22}.$$

Minimizing the negative logarithm of the right-hand side instead,

$$\begin{aligned} \hat{f}_{ML} &= \underset{f}{\operatorname{argmin}} \sum_{i=1}^{I} \left(\sum_{j=1}^{J} t_{ij} f_j - g_i \right)^2 \\ &= \underset{f}{\operatorname{argmin}} \|tf - g\|_2 \end{aligned} \tag{3.23}.$$

Therefore, the ML solution is equivalent to FBP (Eqn. (3.19)) when the number of counts in each detector bin is modeled as Gaussian with a constant variance.

However, in reality, the number of counts in each detector bin g_i are distributed as independent Poisson random variables [62] with mean and variance

$$\bar{g}_i = \sigma_i^2 = \sum_{j=1}^{J} t_{ij} f_j \tag{3.24},$$

because they represent the counts from a source with a small probability of creating an event at any time. Then, the maximum-likelihood source-distribution estimate is

$$\hat{f}_{ML} = \underset{f}{\operatorname{argmax}} \prod_{i=1}^{I} \frac{e^{-\sum_{j=1}^{J} t_{ij} f_j} \left(\sum_{j=1}^{J} t_{ij} f_j\right)^{g_i}}{g_i!} \tag{3.25}.$$

Again, maximizing the logarithm instead and ignoring the terms that are constant for a particular measurement,

$$\hat{f}_{ML} = \underset{f}{\operatorname{argmax}} \sum_{i=1}^{I} \left[-\sum_{j=1}^{J} t_{ij} f_j + g_i \ln\left(\sum_{j=1}^{J} t_{ij} f_j\right)\right] \tag{3.26}.$$

There are a number of ways to find the maximum of this equation using standard optimization methods. One method that has gained wide acceptance for this specific likelihood is the iterative maximum-likelihood expectation-maximization (MLEM) algorithm [63,64,65]. If we define the sensitivity of the detector to photons from a specific source bin as

$$s_j = \sum_{i=1}^{I} t_{ij} \tag{3.27},$$

the sum of the probabilities of causing and detecting each possible event in the detector, the next best estimate of the source distribution is

$$f_j^{k+1} = \frac{f_j^k}{s_j} \sum_{i=1}^{I} \frac{g_i t_{ij}}{\sum_{j'=1}^{J} t_{ij'} f_{j'}^k} \tag{3.28}.$$

The initial estimates of source intensity f^0 can be any positive values, usually a uniform image. The derivation of this equation is shown starting on page 39.

The MLEM equation is guaranteed to converge to a maximum-likelihood solution if the initial estimates f^0 are positive [29]. In addition, with a positive initial estimate, all intensity estimates stay non-negative, which is what is needed to describe physical source

densities. Finally, one can show [29] that during the iterations, counts are conserved in the sense that

$$\sum_{i=1}^{I} g_i = \sum_{j=1}^{J} f_j^k s_j \qquad \forall\, k \tag{3.29}$$

List-Mode MLEM

Eqn. (3.28) describes bin-mode MLEM in which the measurements are binned based on their values and the data from the measurement are the number of counts in each bin $\{g_i\}$. In the limit of a large number of very small bins, it makes more sense to keep track of the events themselves since the storage space to record the measured values of each event will be smaller than storing the number of counts in each and every bin. Also, with high probability, each bin will hold zero or one counts [66]. Then, one can redefine i in Eqn. (3.26) and Eqn. (3.28) to index just the observed events with $g_i = 1$. The rare case of repeated events can be treated as distinct events in different bins since in Eqn. (3.26)

$$g_i \ln\left(\sum\nolimits_{j=1}^{J} t_{ij} f_j\right) = \sum_{i'=1}^{g_i} \ln\left(\sum\nolimits_{j=1}^{J} t_{ij} f_j\right) \tag{3.30}$$

This is known as list-mode MLEM and is the way we implement MLEM. Note that the sum for the sensitivity in Eqn. (3.27) must still be over all possible events, or, since the bins are never explicitly defined in list-mode MLEM, another method must be used to find the sensitivity, as described next.

Sensitivity Calculation

The direct calculation of Eqn. (3.27) is difficult for two reasons. First, this sum is dependent on the event-wise scaling of each system-matrix element, and these scale factors have not been implemented in our system matrix. Second, the sum should be over all possible events, which is a very large set. A distinct event consists of a set of interaction locations and energies, so even conservatively binning into 20 depths and 400 energy bins, an 11-by-11-pixel detector will have about 10^6 possible single-pixel events, 10^{12} possible two-pixel events, and 10^{18} possible three-pixel events. Even if a teraflop computer had instant access to the elements, the sum itself would take many years.

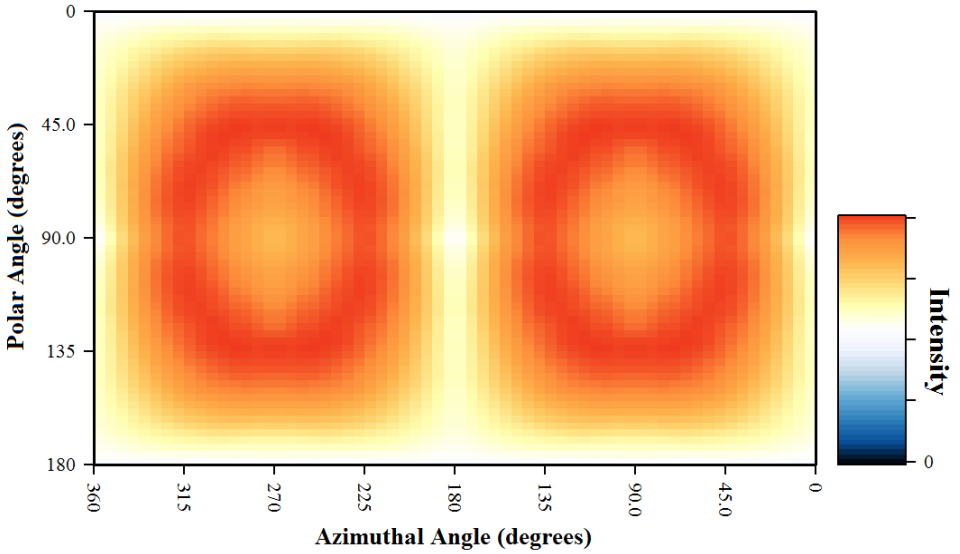

Figure 3.6. The sensitivity of the 18-detector array as a function of source direction at 662 keV for two-, three-, and four-interaction events. The direction perpendicular to the faces of the planes corresponds to 90° and 270° in the azimuthal angle and 90° in the polar angle.

Therefore, the sensitivity as a function of direction and energy is calculated using a Monte Carlo method proposed by Xu [29,67]. First, gamma rays are simulated uniformly emitted from all directions around the detector and at all energies up to the maximum that will be reconstructed. Then, since, with the appropriate sensitivity image, spectral and image reconstruction of this situation should reconstruct back to uniform in space and energy, the efficiency can be deduced from the first iteration of the reconstructed energy-dependent image. Put another way, the forward projection during the simulation samples events according to their probability to occur, then the back-projection during the first reconstruction iteration yields the relative probability to create this event from different directions and energies. The sum of these gives the efficiency. Realistic energy and position resolution are included in the simulation, but other detector factors such as surrounding materials, dead layers, and detector-pixel non-uniformities were not included in the sensitivity calculation. Neglecting these factors causes the sensitivity for single-interaction events to be skewed the most – making single-interaction events difficult to image, as discussed on on page 60. Figure 3.6 shows the two-, three-, and four-interaction sensitivity at 662 keV as a function of direction for the 18-detector array. The direction directly above the planes corresponds to 90° and 270° in the azimuthal angle and 90° in the polar angle. The most sensitive directions are about

halfway between directly above one of the planes and to the side of the planes, and the sensitivity varies by about a factor of two between different directions.

For a more direct measure of sensitivity that is independent of the modeled system matrix, it is also possible to simulate a source emitted from each bin in the image space sequentially, or even to approximately do this experimentally. Counting the number of usable events from each trial relative to the number emitted will give the sensitivity.

MLEM Derivation

MLEM can be derived a number of ways [29,63,64,65]. One way is through optimization transfer [68]. The likelihood to minimize from Eqn. (3.26) is

$$-L(f) = \sum_{i=1}^{I}\sum_{j=1}^{J} t_{ij}f_j - \sum_{i=1}^{I} g_i \ln\left(\sum_{j=1}^{J} t_{ij}f_j\right) \quad (3.31).$$

By replacing

$$\sum_{j=1}^{J} t_{ij}f_j = \sum_{j=1}^{J} \frac{\beta_{ij}^k t_{ij}f_j}{\beta_{ij}^k} \quad (3.32),$$

for $\beta_{ij}^k > 0$ and

$$\sum_{j=1}^{J} \beta_{ij}^k = 1 \quad (3.33),$$

Eqn. (3.31) becomes

$$-L(f) = \sum_{i=1}^{I}\sum_{j=1}^{J} t_{ij}f_j - \sum_{i=1}^{I} g_i \ln\left(\sum_{j=1}^{J} \beta_{ij}^k \frac{t_{ij}f_j}{\beta_{ij}^k}\right) \quad (3.34).$$

Now, the negative natural log function is a convex function, which means that

$$-\ln\left(\sum_i a_i x_i\right) \leq -\sum_i a_i \ln(x_i) \quad \forall\, x_i, \qquad a_i > 0, \qquad \Sigma_i a_i = 1 \quad (3.35),$$

except where

$$\sum_i a_i x_i = 0 \quad (3.36).$$

Therefore,

$$-L(f) \leq \sum_{i=1}^{I}\sum_{j=1}^{J} t_{ij}f_j - \sum_{i=1}^{I}\sum_{j=1}^{J} g_i\beta_{ij}^k \ln\left(\frac{t_{ij}f_j}{\beta_{ij}^k}\right) = \sum_{j=1}^{J} \phi_j^k(f_j) \qquad (3.37),$$

except where intensity estimates are zero for every source bin with non-zero probabilities for a specific event. But, an estimate of the source distribution for which a particular observed event is completely impossible is a poor estimate anyway. We define

$$\Phi^k(f) \equiv \sum_{j=1}^{J} \phi_j^k(f_j) \qquad (3.38)$$

and

$$\phi_j^k(f_j) \equiv \sum_{i=1}^{I}\left[t_{ij}f_j - g_i\beta_{ij}^k \ln\left(\frac{t_{ij}f_j}{\beta_{ij}^k}\right)\right] \qquad (3.39).$$

Eqn. (3.37) through Eqn. (3.39) describe a surrogate optimization function that is always above the function we wish to optimize $-L(f)$, except where it is equal for a specific selection of β_{ij}^k. Unlike the original likelihood, the surrogate function $\Phi^k(f)$ is separable in f_j, making minimization significantly easier since each source-bin intensity estimate can be minimized individually.

Then, it may be possible to optimize $-L(f)$ by starting at an initial source-distribution guess, choosing the β_{ij}^k terms so that the surrogate function matches the original function at this point, and then minimizing the surrogate. We are guaranteed to find a point on the original function that is the same or smaller since the surrogate is equal in value at the starting point and always above the original function. By repeating this iteration, convergence will occur.

First, to find the selection for β_{ij}^k to make the two functions equal, set

$$-L(f) = \Phi^k(f) \qquad (3.40),$$

and equality occurs when

$$\beta_{ij}^k = \frac{t_{ij}f_j^k}{\sum_{j'=1}^{J} t_{ij'}f_{j'}^k} \qquad (3.41),$$

where f^k is the source intensity estimate at the k-th iteration. Next, the minimum of Eqn. (3.39) occurs when

$$\frac{d\phi_j^k(f_j)}{df_j} = 0$$

$$\sum_{i=1}^{I} t_{ij} - \sum_{i=1}^{I}\left[\frac{g_i\beta_{ij}^k}{f_j}\right] = 0 \qquad (3.42).$$

$$f_j = \sum_{i=1}^{I}\left[\frac{g_i\beta_{ij}^k}{\sum_{i=1}^{I} t_{ij}}\right]$$

Combining Eqn. (3.42) with Eqn. (3.27) and (3.41), and identifying f_j as the next iteration estimate of f_j^{k+1},

$$f_j = \sum_{i=1}^{I}\left[\frac{g_i t_{ij} f_j^k}{s_j \sum_{j'=1}^{J} t_{ij'} f_{j'}^k}\right] \qquad (3.43),$$

which is equivalent to Eqn. (3.28).

Regularization

The above discussion of MLEM assumes that there is no prior knowledge of the source distribution. However, if properties of the image are expected, this prior knowledge can be built into the model as a prior in Eqn. (3.20), which manifests itself as a penalty added to the log likelihood. This is called regularization [68]. Penalty functions can encourage image properties such as sparseness [69], smoothness, and flat regions with limited number of edges, for example [68].

Summary

Imaging with gamma rays is rarely a deterministic process. In almost all modalities, a single observed event is consistent with a gamma ray incident from a number of possible directions and energies. Therefore, it is necessary to use imaging methods that analyze a collection of events to estimate the radiation distribution. In SBP, each event contributes intensity to some directions that seem most probable to have produced the event, yet the selection of these directions is an art, balancing using limited directions, correctly guessing the true direction, and not overly biasing some directions. In FBP, the PSF is deconvolved, yet because the problem is solved in the least-squares sense, counting statistics are (incorrectly with few counts) assumed to be Gaussian. In

MLEM, Poisson statistics are used to find the source distribution that makes the observed events most probable, yet the solution can only be found iteratively. The model relating the source distribution and observed events is used, to some extent, in all these reconstruction methods, yet it appears explicitly only in the MLEM reconstructions. This model is useful in any reconstruction problem and will be used later in other imaging applications in Chapter 4 and detection methods in Chapter 7.

Chapter 4
Applications of Maximum-Likelihood Expectation-Maximization

> *Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.*
> – George E. P. Box and Norman R. Draper, Empirical Model-Building and Response Surfaces, 1987

MLEM is a powerful tool for reconstructing intensities in a source distribution from event measurements when there is no prior expectation of the source-distribution properties. It is applicable to many different types of reconstructions depending on the choice of basis functions used to describe the unknown source distribution and the type of measurements collected. As long as the correct system response is used, best performance should occur when the input is the actual data recorded by the detector and the output is what you ultimately wish to learn with no other weightings. The fewer source-distribution parameters that must be estimated, the better.

The next few sections describe different choices for the unknown source distribution parameters that can be used for various applications. The Spatial MLEM and Energy-Spatial MLEM methods have been reported by previous authors. The Isotope-Spatial MLEM and Energy-Target MLEM methods are proposed for the first time here to support applications in which the user is interested in either the isotopes responsible for emissions as a function of direction or the spatial distribution of emissions in moving objects. Such algorithms may be beneficial for security applications using position-sensitive detectors, for instance. This chapter ends with the first results of using an 18-detector array system for single-interaction imaging, which is important for improving the fraction of recorded events used for imaging.

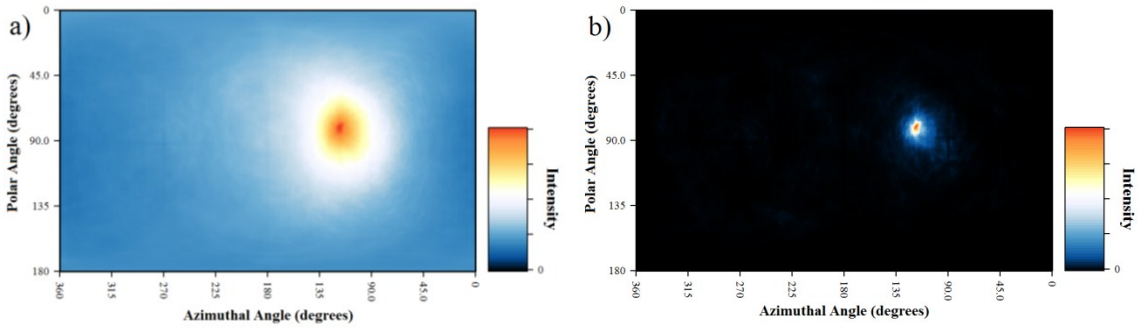

Figure 4.1. Spatial MLEM of a Cs-137 source in the energy window covering the full-energy peak from 620 keV to 700 keV, using 2×10^4 events. a) The image at the first iteration after a uniform image. b) The image at the 25-th iteration after a uniform image.

Spatial MLEM

One of the most basic types of reconstruction is reconstruction to an image space consisting of the 2D or 3D space around the detector. Two modes can be distinguished: when the energy of the incident radiation is known or can be assumed, and when the incident radiation may be from any energy. The former is prevalent in the literature [29,33,34,53,70,71]. Often in that case, full-energy deposition is assumed for each event; the last interaction is assumed to be a photoelectric interaction [29,33,34,53]. Then Eqns. (3.9), (3.11), and (3.12) will not contain the second part of the sum in the innermost parenthesis that describes escape probability. Neglecting the small spread due to energy uncertainty, each energy bin is independent and MLEM can be applied to each separately, reducing computation time since each event only need be used to update the image bins at its own energy. From the overall spectrum, one can often know the energies of the sources that are present and only events from these energy bins can be used to simplify the calculation further. According to this model, non-full-energy-deposition events have scattered out of the useful energy range and no longer contribute to the sensitivity of the detector, so the sensitivity factor tallies only full-energy-deposition events.

Figure 4.1 shows a MLEM reconstruction of a Cs-137 source in the energy window from 620 keV to 700 keV. Full energy deposition was assumed and 2×10^4 events were used. The sensitivity scales the intensity to a number emitted per solid angle at some distance. Figure 4.1a shows the first iteration past a uniform image; it is similar

to a SBP image with azimuthal and polar FWHM of about 45° and 47°, respectively. Figure 4.1b shows the result after 25 additional iterations. The number of iterations are chosen to produce a clean image by stopping before the image has completely converged to the slightly more noisy maximum-likelihood solution. The image in Figure 4.1b has better than 6° and 8° FWHM in the azimuthal and polar directions, however because the image is the result of an iterative process, this may not represent the true angular resolution of discriminating neighboring points. To do this, an image with incrementally closer point sources is needed. Angular resolution for this system using MLEM is presented by Wang [72].

Assuming full-energy deposition means that MLEM will always reconstruct some events – Compton-continuum events – incorrectly. This is not a problem if the energy range of interest only covers the highest photopeak in the spectrum where there is no continuum from partial-energy depositions, but at other energies, artifacts appear that depend on the location of the source responsible for the continuum events. Nevertheless, distributed background continuum will be spread more evenly over all directions and may not skew the image much.

MLEM reconstruction to just the spatial domain allows simpler computation because only limited physics is modeled. The way to remedy artifacts due to the limited physics is to model the physics of Compton scattering. To model Compton scattering while reconstructing to just space, one could either assume an initial energy for all events, or integrate probability over energies assuming a prior on the energy spectrum to include the probability each event was from any energy up to some maximum expected. However, the first is problematic when photons of multiple initial energies are present, and, while the second includes correct probabilities, it throws away information about the true energy distribution.

Energy-Spatial MLEM

Instead, to include the physics of Compton scattering in the model when the incident energy distribution is unknown, deconvolution must take place over both energy and 2D or 3D space around the detector. This is because the PSF in space depends on the

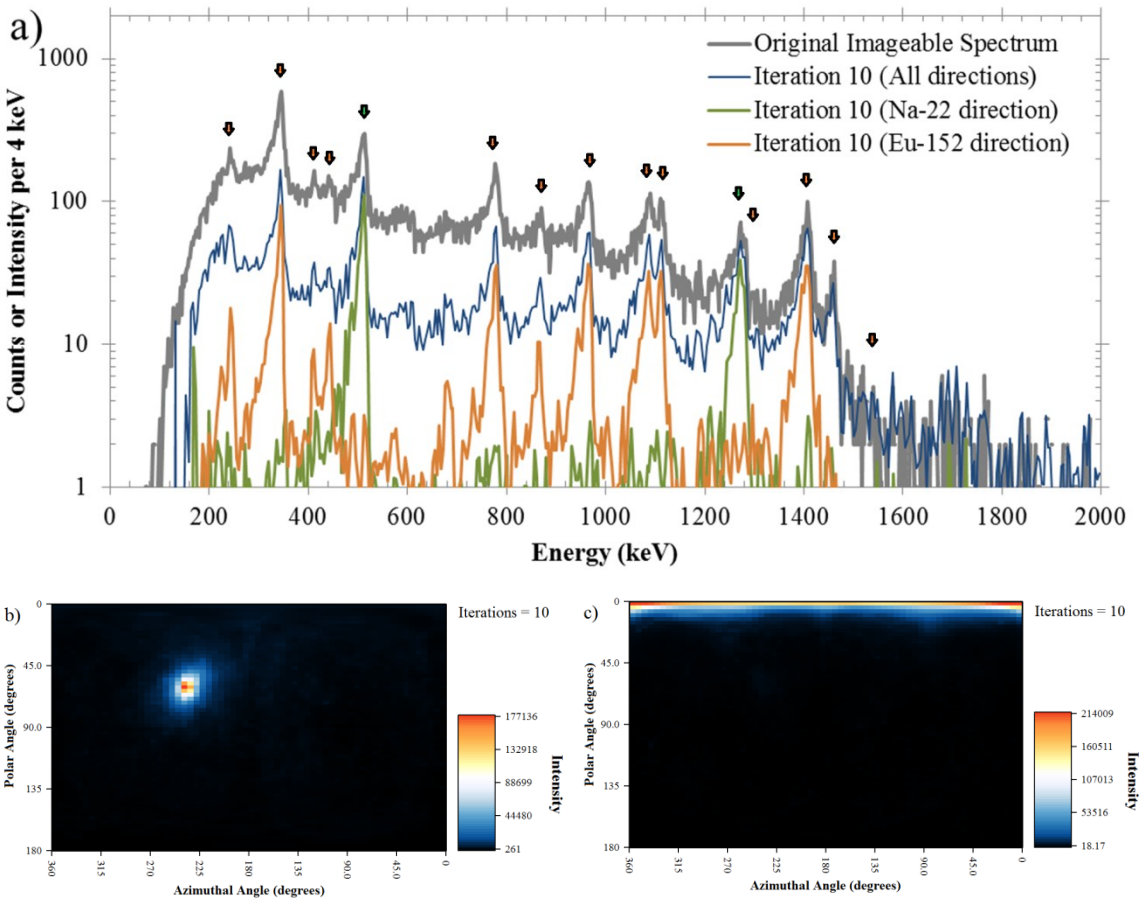

Figure 4.2. An example energy-spatial MLEM deconvolution. a) The raw and deconvolved spectrum from multiple-interaction events. The directional spectra cover 0.5 steradians around the source direction. The arrows denote expected emission lines with relative branching ratios over 8%. b) The image in the energy windows covering the expected Eu-152 emission lines. c) The image in the energy windows covering the expected Na-22 emission lines. Note that this image shows a point source located exactly on the pole.

energy of the incident radiation and the spectral response depends on the source position of the photons. Therefore, deconvolution must be done in both simultaneously. Here, the MLEM image space consists of bins, each describing a particular (small) energy range and region in space. This image space was introduced by Xu and He [51] and later used by Maltz *et al.* [73] and is known as energy-imaging integrated deconvolution (EIID). The system-response functions and sensitivity image are described starting on page 23 and page 37.

Figure 4.2 shows an example of a measurement with an 18-detector array in which two sources, Na-22 and Eu-152, were placed in different directions. The system collected 5×10^4 multiple-interaction events and 10 iterations of MLEM were performed,

starting with an initial back-projection image, at which the energy-image changed little during each iteration. Figure 4.2a shows the raw spectrum of multiple-interaction events and the deconvolved spectrum over all directions. Arrows over the raw spectrum denote expected emission lines with relative branching ratios of over 8%. The vertical scale of the deconvolved spectrum is in emission rate at some distance and can be scaled by a constant depending on the units chosen. The spectrum in a 0.5-steradian region around the direction of each source shows the significant lines for each source. While there is some noise in the deconvolved spectrum, one can imagine making a cut below an intensity of 4 to eliminate most of it. The deconvolved spatial distributions in Figure 4.2b and Figure 4.2c show a small hot spot for each source and very little intensity in directions not from either source.

Isotope-Spatial MLEM

One way to reduce the size of the problem in EIID is to choose basis functions in the energy domain other than the standard equally spaced bins. In almost all applications, the set of possible isotopes that may be present is smaller than about 50. The recorded energy spectrum will be made up of the emissions from isotopes within this set. Therefore, we can reduce the number of energy parameters that need to be estimated from several hundred energy bins to fewer than 50 isotopes that may be present. This approach may help performance with limited statistics. In addition, reporting isotope contributions instead of an energy spectrum may be more useful for end users. Table 4.1 shows a set of possible isotopes common in background or laboratory sources that are used in this study.

If there were never material between the emitting atom and the detector, this set of isotope basis functions would be sufficient to describe any energy distribution observed in normal operation. However, much of the energy distribution due to background sources has undergone significant scattering in the background material. This produces a broad continuum in the background energy spectrum. Such a distribution is hard to predict without a detailed model of surrounding materials and the natural radiation distribution within them [74].

Table 4.1. Library of background and common isotopes.

Nuclide	Energy (keV)	Emission Fraction
Na-22	511.0	0.6667
	1274.5	0.3333
K-40	1460.8	1
Mn-54	834.8	1
Fe-59	1099.2	0.5667
	1291.6	0.4333
Co-60	1173.2	0.4997
	1332.5	0.5003
Ga-68	511.0	0.9823
	1077.3	0.0177
Ru-103	497.1	0.9405
	610.3	0.0595
Cs-137	661.7	1
Tl-208	510.7	0.1033
	583.2	0.3864
	860.6	0.0568
	2614.5	0.4534
Tl-209	465.1	0.4926
	1567.1	0.5074
Bi-214	609.3	0.4240
	768.4	0.0454
	934.1	0.0279
	1120.3	0.1389
	1238.1	0.0533
	1377.7	0.0368
	1408.0	0.0198
	1509.2	0.0194
	1729.6	0.0269
	1764.5	0.1416
	1847.4	0.0194
	2204.2	0.0467
Ac-228	338.3	0.1616
	463.0	0.0631
	794.9	0.0609
	911.2	0.3700
	964.8	0.0716
	969.0	0.2266
	1588.2	0.0462

The distribution of photon energies after mono-energetic photons scatter one time follows the Klein-Nishina distribution [3]. To first order, this can be approximated as a rect function extending from the initial energy down to the Compton cutoff associated

with back-scattered photons. Each photon may then scatter again to an energy between the Compton cutoff from backscatter and its initial energy. So, a possible set of continuum basis functions would be rect functions extending from each possible energy down to its Compton cutoff. Note, however, that from the physics of backscatter these Compton cutoffs are always less than 256 keV, which is near the lower limit for detection using two-pixel events on our system. Since the system sensitivity is zero below about 200 keV (the approximate Compton cutoff for 1000-keV photons), we simply extend these basis functions down to 0 keV for ease of implementation without change in response. To reduce the number of parameters needed to estimate, the initial energies are spaced only every 100 keV and every 500 keV above 1 MeV. In practice, this does not seem to be a bad approximation since the set of step functions is still similar to the shape of a continuum. Adding more continuum bins above 1 MeV does not greatly change reconstruction results.

One of the advantages of reconstructing to this new set of energy basis functions is that it directly gives quantities of interest. End users are more interested in the isotopes present than the emission spectrum. Rather than having to convert from spectrum to isotope identification in post-processing, a spatial source distribution is given directly for each isotope in the library.

Method Implementation

It is possible to use MLEM for reconstruction to the isotope and image domain by identifying the source distribution bins as a set of direction bins for each isotope in our isotope library. With 2D direction bins, each pixel may be differently sized, so the solid angle Ω_j of each direction j must be considered. For clarity, if we let j index the J image directions and n the N isotope bins, the MLEM equation in Eqn. (3.28) turns into

$$f_{jn}^{k+1} = \frac{f_{jn}^{k}}{s_{jn}} \sum_{i=1}^{I} \frac{t_{ijn}}{\sum_{j'=1}^{J} \left(\Omega_{j'} \sum_{n'=1}^{N} t_{ij'n'} f_{j'n'}^{k} \right)} \tag{4.1}$$

where f_{jn}^{k} is the estimate of the fluence of photons (photons/steradian) for isotope n and direction j at iteration k, and s_{jn} is the sensitivity of the system to observe photons from isotope n and direction j.

The sensitivity for each isotope-direction bin can be calculated from the sensitivity calculated for EIID in each energy-direction bin by combining energy bins since the set of energy bins *{m}* is a non-overlapping neighboring string of narrow rect functions that cover the energy dimension. The sensitivity for isotope *n* in direction *j* can be calculated as

$$s_{jn} = \sum_{m=1}^{M} s_{jm}\Gamma_{\mathrm{mn}} \tag{4.2}$$

where Γ_{mn} is the normalized branching ratio (the fraction of emissions of isotope *n* that fall in energy bin *m*) and the sum is over all *M* small standard energy bins. For isotopes, most Γ_{mn} are zero, so the sum only has as many terms as emission lines from the source. For generality, the extra continuum basis functions are treated like isotopes with

$$\Gamma_{mn} = \begin{cases} \dfrac{1}{M_n}, & m \le M_n \\ 0, & m > M_n \end{cases} \tag{4.3}$$

where M_n is the number of standard energy bins below the maximum energy in the continuum function *n*.

The system matrix elements t_{ijn} are calculated by taking the integral of the system-matrix elements for regular energy bins over the energy distribution of each isotope. This is implemented as

$$t_{ijn} = \sum_{m=1}^{M} t_{ijm}\Gamma_{mn} \tag{4.4}$$

where t_{ijm} is the system matrix element calculated for the standard energy binning case for energy bin *m* and direction *j*.

One of the most computationally intensive aspects of this reconstruction is the calculation of the system matrix elements. When only reconstructing to a set of isotopes without the continuum bases, the calculation time is greatly reduced, since elements at most energies will not need to be evaluated. However, when including the continuum bases, all t_{ijm} elements must be calculated, so this method will not reduce computation time for the system matrix over using standard energy bins. Still, because there are fewer parameters to estimate, each iteration can be computed faster and fewer iterations may be needed.

Performance

Performance was tested in simulation and experiment using a single 3D-position-sensitive 20 mm × 20 mm × 15 mm CZT detector described in Chapter 2, using only the well-studied two-interaction events.

Simulation

A simulation was performed with Geant4 [75] of a single radiation source at 662 keV in parallel-beam geometry. Realistic position and energy resolution were included, as were effects of the finite charge-cloud size and Doppler broadening. Since the system model used in the reconstruction only approximates effects due to position resolution and does not fully consider the finite charge-cloud size and Doppler broadening, we expect some artifacts in the reconstructed image due to some degree of model mismatch.

A set of 10^4 two-interaction events were used in the algorithm to reconstruct the source's spatial and energy distribution. Similar results were obtained using ten times more events. The 4π field of view was pixelated into 36-by-36 pixels over the polar and azimuthal spherical coordinates. The resulting energy distribution over all directions from reconstruction over 500 standard energy bins and from reconstruction with continuum basis functions and the library of isotopes in Table 4.1 are shown in Figure 4.3. The raw two-interaction spectrum for these 10^4 events recorded by the detector is also shown in Figure 4.4 to illustrate the degree of deconvolution that is necessary.

One can see that in simulation both methods reconstruct primarily a Cs-137 source at 662 keV. In the standard energy bin reconstruction, an intensity of 269 (in arbitrary units) is in the bin covering 660 keV to 664 keV and an intensity of 313 is in the 5 bins from 652 keV to 672 keV that cover the primary photopeak region. There is a small escape peak in the simulation that appears because of threshold differences with experiment, which has an estimated intensity of 27 over the region of 632 keV to 652 keV. Compare this with the slightly larger reconstructed intensity of 320 when isotope bins are used.

Very similar degrees of deconvolution are achieved in both methods. For an informed user, deconvolution to the standard energy bins may be clearer since the small

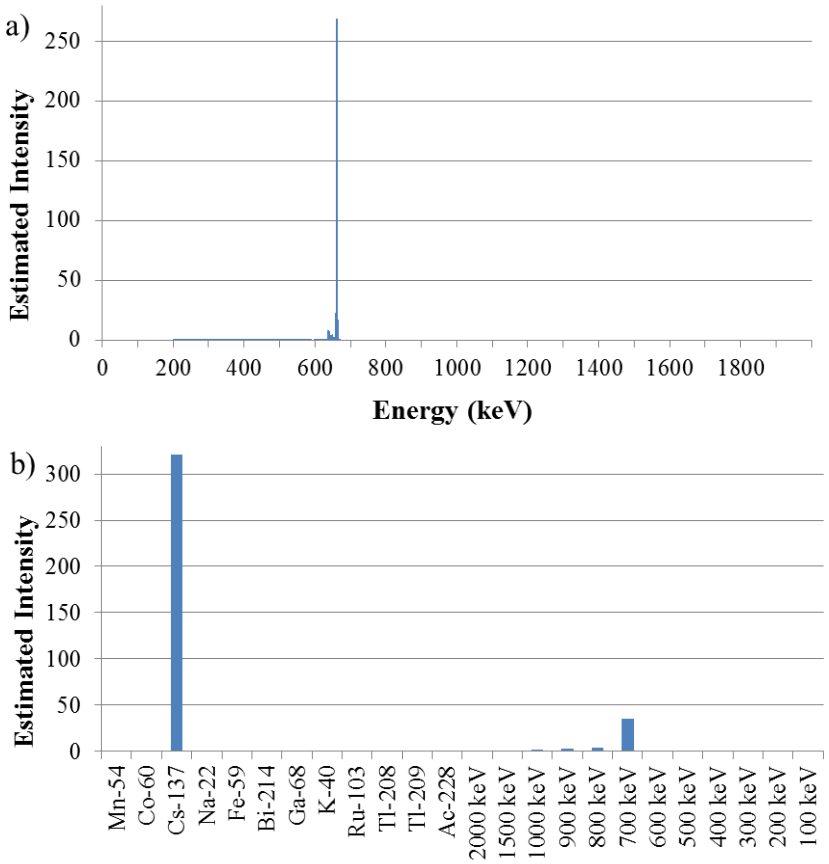

Figure 4.3. The reconstructed energy distributions over all directions with 10^4 simulated events from Cs-137 after 25 iterations near convergence. (a) Using 500 standard energy bins as basis functions over the energy range 0 to 2 MeV. (b) Using a set of isotopes and continuum functions as basis functions over the same range. The labels with energies on (b) represent the upper limit of a continuum basis function. Both plots use a 36-by-36 spherical image mesh. The reconstructed intensities have arbitrary units, but are comparable between methods.

escape peak can immediately be identified to belong with the Cs-137 peak. However, more novice users may prefer the direct deconvolution to the isotope domain where Cs-137 is identified and the escape peak is classified as continuum.

It is interesting to note how a small continuum component, besides the escape peak, appears in both reconstructions even though the simulation had no such continuum. This is attributed to model mismatch and is discussed in the next section.

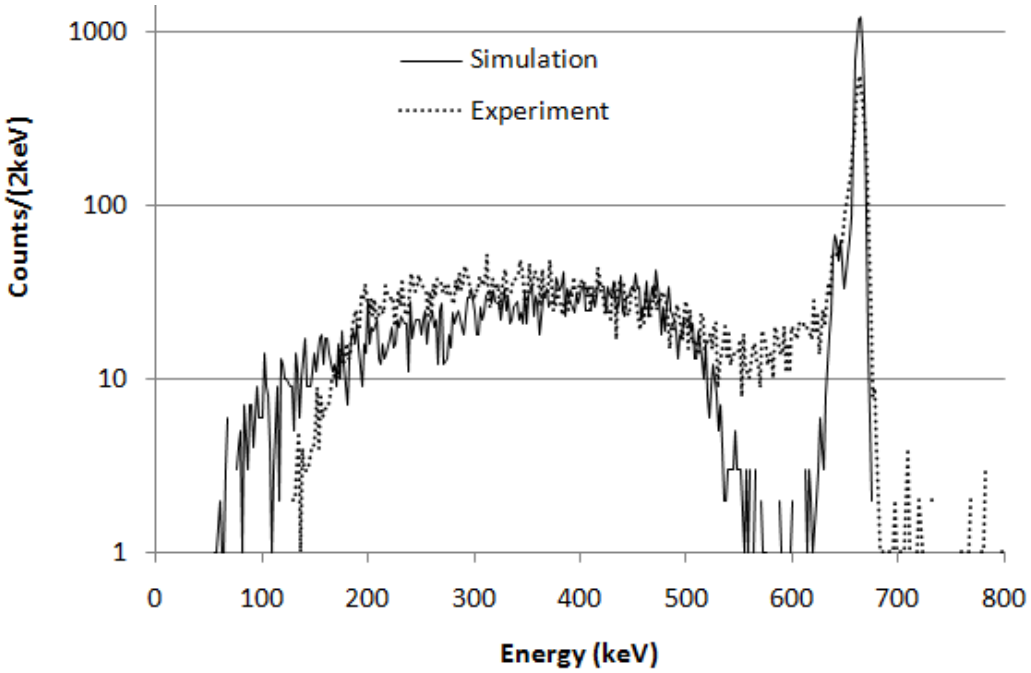

Figure 4.4. The recorded energy spectrum over all directions for 10^4 simulated and experimental two-interaction events from Cs-137. Compare the difference in spectral shape due to experimental factors not included in the simulation.

Experiment

An experiment similar to that of the simulation was performed of a single Cs-137 point source. Again, 10^4 two-interaction events were collected and reconstruction was performed with the same methods as in the simulation case. The estimated energy distribution over all directions is shown in Figure 4.5a for standard energy bins and in Figure 4.5b for isotope and continuum basis functions. The comparison between methods is similar to simulation. The peak channel in the reconstruction with standard energy bins has an intensity of 89 and the region from 652 keV to 672 keV contains an intensity of 169. Compare this to the estimated Cs-137 intensity of 148 using isotope bins. The degree of deconvolution is similar. Again, it may be easier for an inexperienced user to use a system that reports Cs-137 rather than the deconvolved spectrum.

Small components of Bi-214, K-40, and Ac-228 show up in the isotope reconstruction. There is actually some evidence that these background isotopes are present in small quantities in this experiment. For Bi-214, significant detectable lines should be at 609 keV and 1120 keV. Because of extra events between the Compton edge

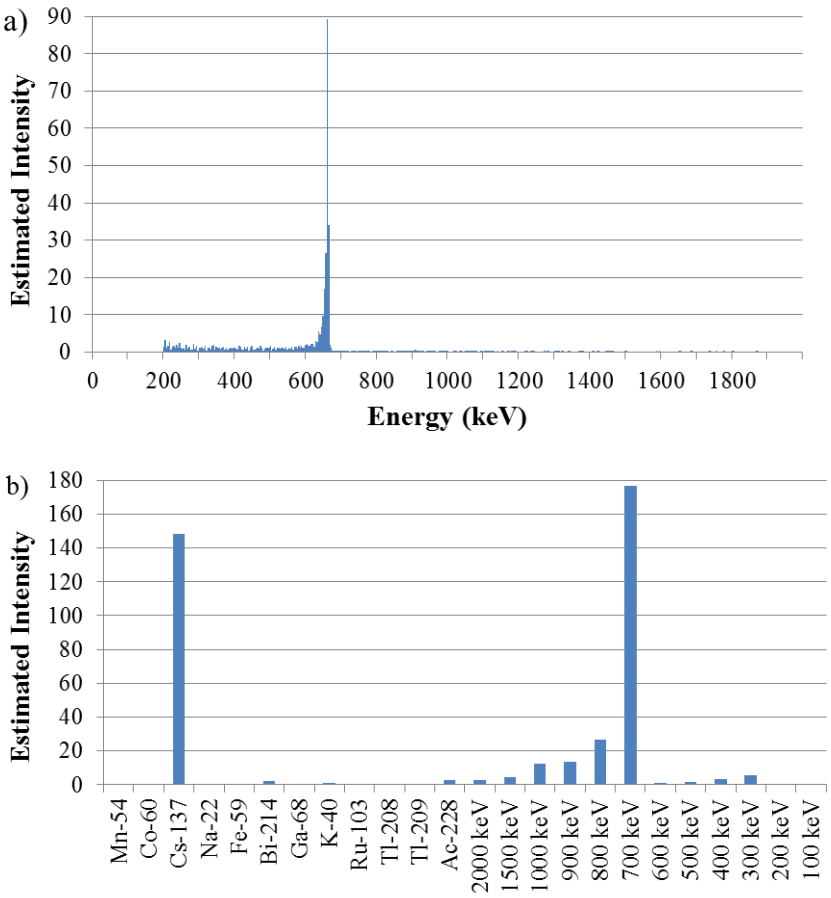

Figure 4.5. The reconstructed energy distributions over all directions with 10^4 experimental events from Cs-137 after 25 iterations near convergence. (a) Using 500 standard energy bins as basis functions over the energy range 0 to 2 MeV. (b) Using a set of isotopes and continuum functions as basis functions over the same range. The labels with energies on (b) represent the upper limit of a continuum basis function. Both plots use a 36-by-36 spherical image mesh. The reconstructed intensities have arbitrary units, but are comparable between methods.

and full-energy peak around 600 keV in the original spectrum, a small Bi-214 component at 609 keV is estimated in the isotope deconvolution since such a shape could be due to a small peak at 609 keV. A small peak of three counts is present at 1113 keV in the raw spectrum as well, which is near the expected 1120 keV energy. The raw spectrum also has small (about three-count) peaks around 1460 keV and 911 keV, which suggest the presence of K-40 and Ac-228. From this experiment, it seems that reconstruction in the isotope domain may be able to separate these weak signatures from continuum. The presence of these sources is not obvious in the reconstruction to the standard energy bins.

Model Mismatch

Figure 4.4 shows the experimental spectrum, and it is clear that the shape is significantly different from simulation. This is due to effects such as scatter in surrounding electronics and supports, crystal non-uniformities, coincident events, Doppler broadening, and cross talk. None of these effects are included in the system matrix for reconstruction, and most are not included in the sensitivity image. The more distorted shape of the spectrum from what the system model expects causes degradation in reconstruction performance. In Figure 4.5a, a larger continuum than in simulation appears from 200 keV to 662 keV since the experiment contains more of these events than expected. This also shows up in the isotope reconstruction as significantly more continuum components, especially in the range of 0 to 700 keV.

The effect of this mismatch between the true system and the system model used for reconstruction is also present, but to a lesser degree, in the simulation. Mismatch in the simulation is due to simulated factors that do not appear in the reconstruction's system model.

In simulated or experimental reconstructions, model mismatch appears to create somewhat of a continuum extending from the full-energy peak down to the threshold for recorded events. Since this also matches one of the continuum basis functions used in the reconstruction to isotope bins, the effect of model mismatch is, to some degree, moved to a continuum bin without greatly skewing which isotopes are estimated. Much of the time, these continuum bins may be ignored as either model mismatch or background, but the presence of large amounts of continuum can also be indicative of shielding around a source. So, best performance can obviously be obtained by including factors in the system model that reduce this mismatch.

Multiple Sources

Figure 4.6 shows the raw spectrum from an experiment with 8434 events from a Na-22 point source and a Cs-137 point source. Figure 4.7 shows the estimated energy distribution over all directions using standard energy bins in Figure 4.7a and isotope and continuum basis functions in Figure 4.7b. As before, the intensity in the energies or isotopes that contain a true source is large and there is continuum present. In the isotope

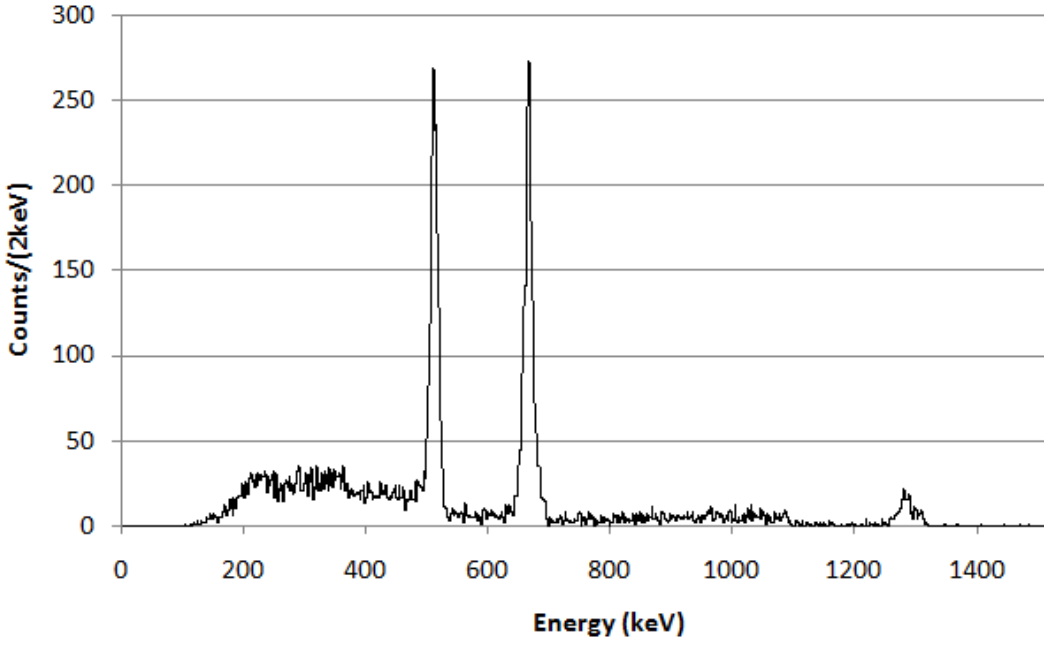

Figure 4.6. The recorded energy spectrum over all directions for 8434 experimental two-interaction events from a Na-22 point source and a Cs-137 point source in different directions.

reconstruction, the two largest contributions, besides continuum, are from the two sources present.

However, there is also large estimated intensity for Fe-59 and some intermediate estimated intensity for Co-60, Ga-68, K-40, Ru-103, and Ac-228. The significant intensity in the misidentification of Fe-59 is troubling until we look closely at the 1274-keV peak in the raw spectrum. Because of calibration error, this peak has shifted to higher energies, so according to the model, only the left half is consistent with a Na-22 source. The right half matches perfectly with the higher-energy line from Fe-59. The other line for Fe-59 falls at the upper edge of the Compton edge for 1274 keV. As we have seen before, experimental data tends to have more counts than expected in this region, which supplies the evidence for the other Fe-59 line. Because of these two effects, the reconstruction may see a large Fe-59 component. In addition, detector sensitivity at these higher energies is small, so even a few counts provide evidence of large source intensity. Co-60 also may occur due to high counts past the Compton edge in this region. Ga-68 is easily explained by overlap with the Na-22 line and the high

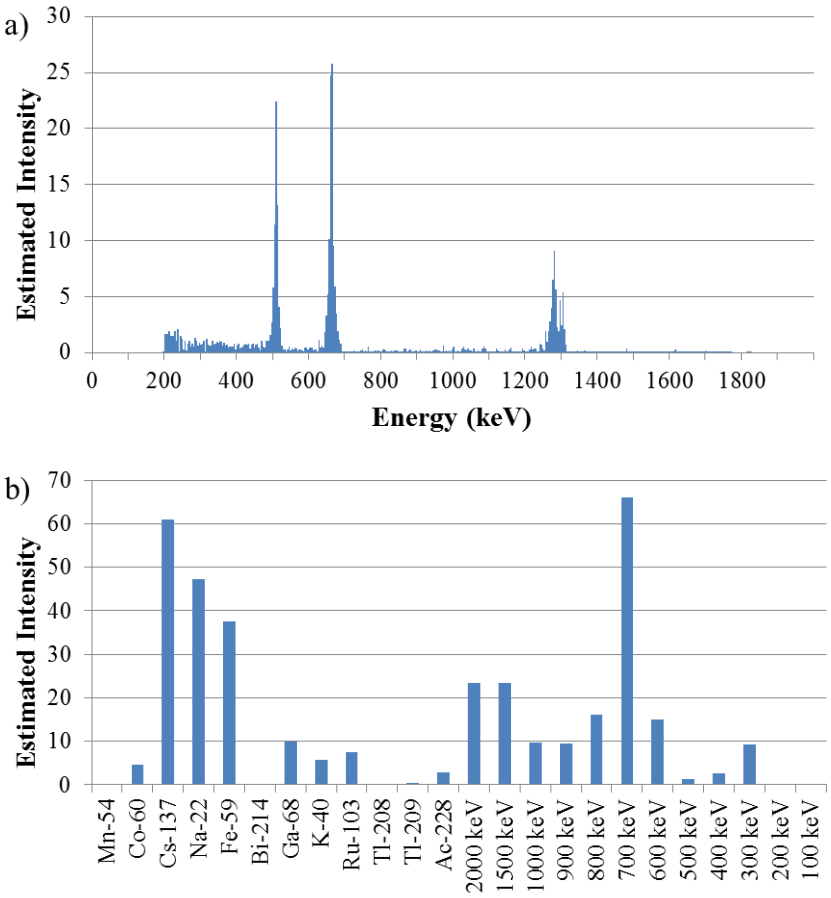

Figure 4.7. The reconstructed energy distributions over all directions with 8434 experimental events from a Na-22 point source and a Cs-137 point source after 25 iterations near convergence. (a) Using 500 standard energy bins as basis functions over the energy range 0 to 2 MeV. (b) Using a set of isotopes and continuum functions as basis functions over the same range. The labels with energies on (b) represent the upper limit of a continuum basis function. Both plots use a 36-by-36 spherical image mesh. The reconstructed intensities have arbitrary units, but are comparable between methods.

Compton edge. The 511-keV peak is explained better by Ga-68 than Na-22 since the entire 1274-keV peak is not associated with Na-22. Based on their location in the spatial part of deconvolution, K-40, Ru-103, and Ac-228 all are associated with misidentification of photons from the Na-22 source. All show up in a band near the Na-22 source location.

Conclusions

In this study, we have presented basis functions for the energy dimension of emission imaging that consist of a set of common isotopes and step continuum functions. These functions may allow for faster computation and the need for fewer measurements

since fewer parameters need to be estimated. In addition, the reconstruction result directly reports isotope spatial distributions that are important to the end user. To some degree, the continuum functions can keep the most obvious effects of model mismatch from qualitatively skewing isotope intensity estimates, but effects such as non-linearity in the raw event spectrum can still affect the resulting set of identifications and intensities, as was shown when a Na-22 peak was shifted to higher energies. Nevertheless, the degree of deconvolution achieved is similar to using standard energy bins.

In the future, the model mismatch leading to incomplete deconvolution to the source that is present should be investigated. Currently, continuum bins are ignored since they often contain just the effects of model mismatches. However, the continuum bins could describe other effects such as the amount of shielding if the amount of model mismatch were reduced.

In addition, if one were concerned with situations where one expects only a few sources present, some form of sparsity regularization could be included. However, such regularization could be undesirable in other situations where a large set of spatially distributed sources are present.

Energy-Target MLEM

Another choice of basis functions for the spatial dimension is useful when there are objects with known motion in the field of view. By choosing basis functions as a regular mesh over stationary space with additional bins that follow moving targets, intensity can be reconstructed for each of the moving targets and the stationary space (the backdrop) simultaneously. This allows the user to locate both moving and stationary sources even when sources pass the same direction concurrently. This is called energy-target integrated deconvolution and is represented graphically in Figure 4.8. Theory and performance is presented in detail by Jaworski *et al.* [76,77].

Mathematically, the intensities that must be estimated are f_{jm}, where we define m as the energy bin, and j as the image bin

$$j \in 1, \dots, J, J+1, \dots, J+J_1, \dots, \mathrm{J} + \sum_{r=1}^{R-1} J_r + 1, \dots, J + \sum_{r=1}^{R} J_r \qquad (4.5),$$

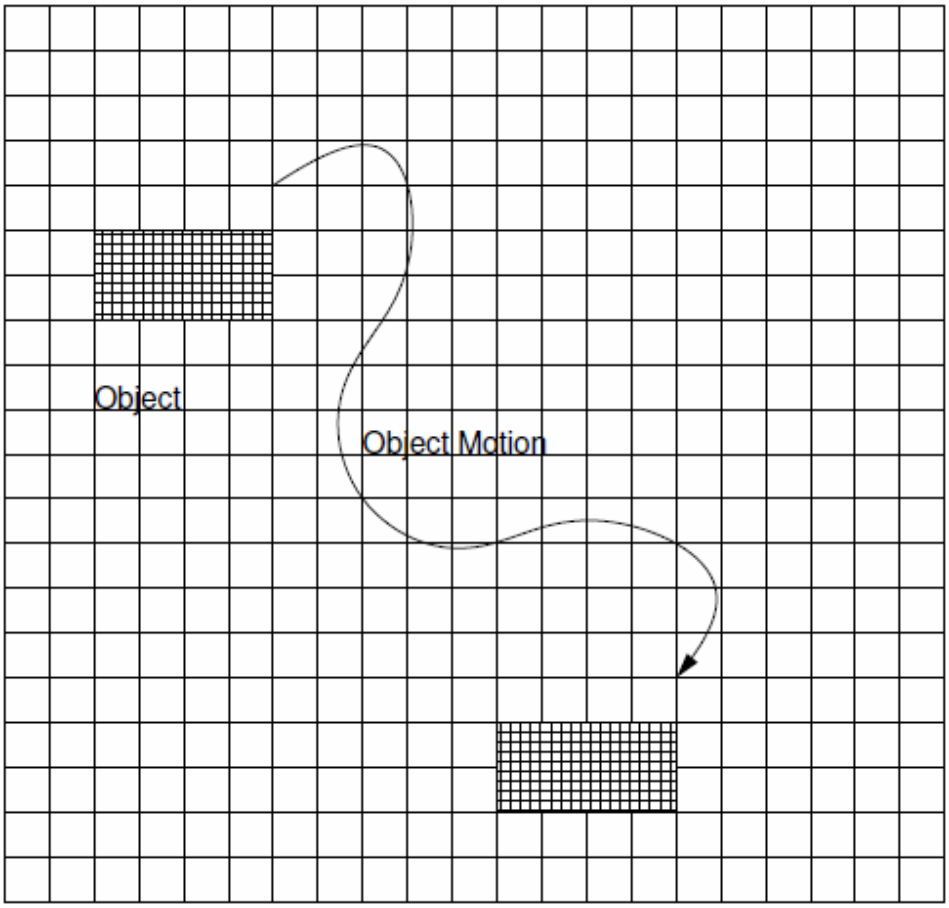

Figure 4.8. The idea behind energy-target integrated deconvolution. A stationary mesh covers the backdrop and a moving object mesh covers each of the objects with known motion.[1]

where there are R moving objects and the number of image bins around the r-th moving object is J_r. The definition of the system matrix elements is the same, but for $j > J$, the incident photon direction is set to the direction from the bin at the time of the event. The sensitivity terms, however, are now a function of time, and hence event, because the direction to some image bins relative to the detector geometry change as a function of time. Therefore, the MLEM equation in Eqn. (4.1) must be further modified to

$$f_{jm}^{k+1} = f_{jm}^{k} \sum_{i=1}^{I} \frac{t_{ijm}}{s_{ijm} \sum_{j'=1}^{J+\sum_{r=1}^{R} J_r} \left(\Omega_{j'} \sum_{m'=1}^{M} t_{ij'm'} f_{j'm'}^{k} \right)} \tag{4.6}$$

with the solid angles constant over time (determined by the mesh structure chosen for each object) and the sensitivity calculated from the direction to the image bins at the time

[1] Image from Jeffrey Fessler, University of Michigan, Ann Arbor.

of event *i*. It is, of course, also possible to use isotope bins for the energy dimension and modify the elements appropriately.

This energy-target integrated deconvolution should be able to determine the spectrum and intensity from each moving object as well as each direction in the stationary space. There will not be crosstalk even between objects that pass through the same region of space simultaneously – as long as they do not share their entire paths – if the system model is accurate and enough events are collected. To understand why, consider standard spatial-only deconvolution. A single event has a positive probability to be from a number of spatial bins, but the ML solution ascertains which of these directions is genuine using the other collected events. Even with multiple sources present, in which a number of events have probability to be from the direction of either source, the entire collection of events allows MLEM to find the correct energies and intensities for the two sources. In the same way, when the direction dimension contains a stationary space and target space, it is inconsequential that events that occur during the overlap of two targets have probability to be from either target. The other events collected when the targets are not co-directional allow MLEM to deconvolve the energies and intensities that were emitted from each target.

Single-Interaction Imaging

Single-interaction events make up a large fraction of events for gamma-rays below a few MeV; therefore, it is useful to be able to include them for imaging. MLEM can image photons that only interact once in the detector by defining the system matrix and sensitivity similarly as with events that have multiple interactions [78]. Mathematically, single-interaction events are no different from other events, so they can be inherently combined with other events to create the image due to all events, with no event weighting necessary. The system-matrix elements are defined for single-interaction events in Eqn. (3.9) and the system matrix for one example event with assumed full-energy deposition was shown in Figure 3.2. Compared to the system matrix for a multiple-interaction event, as shown in Figure 3.3, the system matrix for single-interaction events is much less sparse and less concentrated at a particular direction or

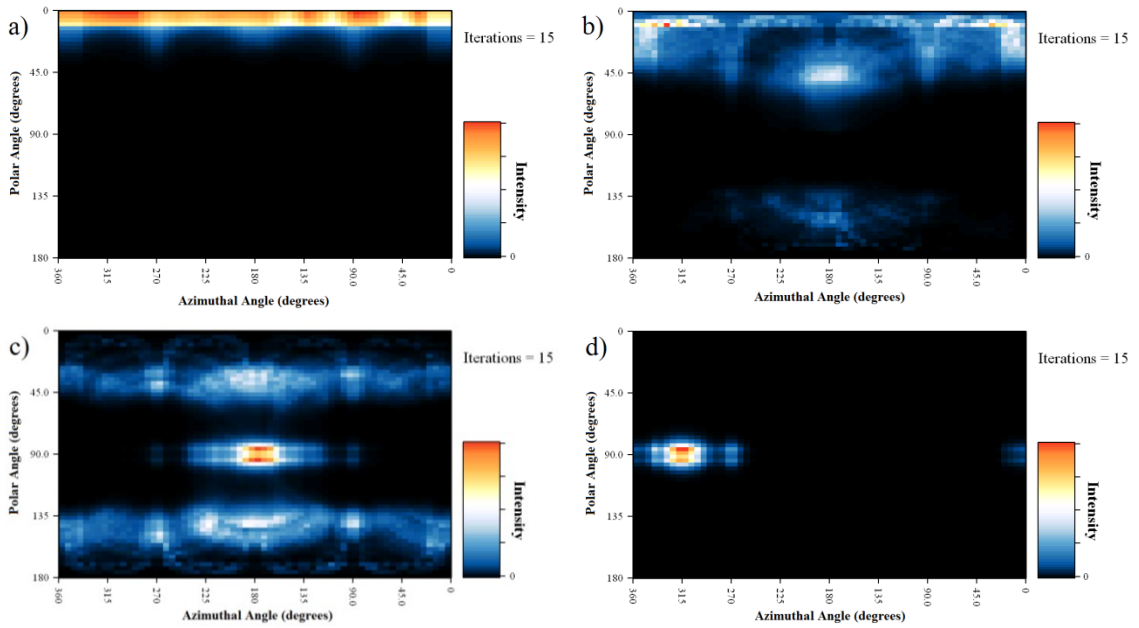

Figure 4.9. The single-interaction images after 15 iterations of MLEM for simulated 122-keV parallel-beam sources coming from a) a polar angle of 0°, b) a polar angle of 45° and azimuthal angle of 180°, c) a polar angle of 90° and azimuthal angle of 180°, and d) a polar angle of 90° and azimuthal angle of 315°.

energy. Nevertheless, with the correct system model and sufficient statistics, deconvolution is possible.

A parallel beam of photons at 122 keV was simulated striking the 18-detector array system with realistic position and energy resolution. The photopeak sensitivity was calculated through direct individual simulation of 122-keV point sources at 18 × 18 directions around the detector. Using 1000 single-interaction events, rough deconvolution is possible, as shown in Figure 4.9 for a number of directions. Because there is little Compton scatter at these energies, full-energy deposition was assumed. In most cases, the hottest direction appears in the direction of the source, and there is always a hot region in the direction of the source. Even with simulation, there is significant model mismatch that skews the image. For instance, the probability calculation does not include the finite size of the detector voxels. This is a significant effect at low energies where the almost all interactions occur in just the surface pixels.

Whereas imaging multiple-interaction events is most sensitive to the direction between the interactions, imaging single-interaction events depends on the distribution of event positions within the detector, especially near the detector edges where most

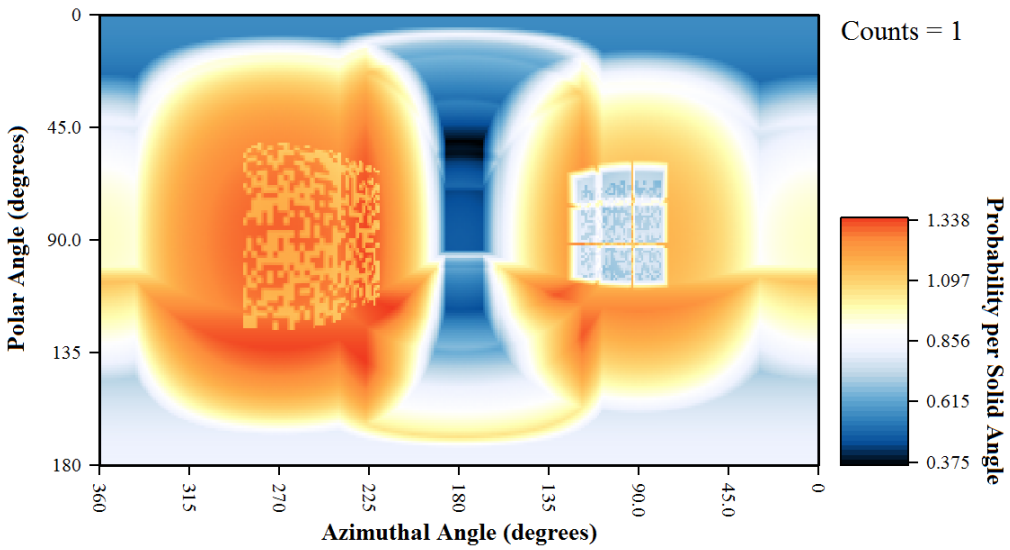

Figure 4.10. A row of the system matrix for the same event as in Figure 3.2 when coded apertures are placed in front of the detector planes.

interactions occur. However, this is the region of the detector that is also least understood. The distribution of single interactions in experiment, therefore, is dissimilar to what the model expects. This large model mismatch skews the estimated source distributions.

A better understanding of the detector effects that contribute to this model mismatch are needed before single-interaction events can be combined with multiple-interaction events for MLEM imaging. Single-interaction events are beneficial because they make up a large fraction of all events, especially at low energies. Below a few-hundred keV, all events are single-pixel events. Therefore, using these events can extend imaging to lower energies, even without a coded-aperture mask. The additional modulation provided by the presence of a coded-aperture mask and mask attenuation in the system model may help improve angular resolution for single-interaction imaging at the expense of efficiency. Figure 4.10 shows a row of the system matrix when masks are present on both sides of the array system (as described by Joshi [13]).

Chapter 5
UMImaging Software

UMImaging
It slices.
It dices.
And, it entices
Those who ask "Radiation? Where?"

Introduction

Three-dimensional-position-sensitive detectors offer a multitude of unique capabilities due to their ability to image a radiation field. Several position-sensitive detector systems have previously been developed to record gamma-ray interaction locations and energies in an uncollimated detector [5,79,80,81,82]. Such systems offer a multitude of analysis possibilities for count-rate analysis, spectral analysis, imaging, automated detection, source localization, verification of correct system performance, and other categories. Each type of calculation encompasses a wide range of possible algorithms. For instance, in order to create an image of the radiation field, algorithms can range from computationally simple techniques such as back-projection (SBP) [33] to more computationally demanding methods such as maximum-likelihood (ML) estimation [51] or regularized reconstruction [69].

Instead of independently implementing each possible analysis method, it makes sense to combine an assortment of these analysis methods into a library. UMImaging attempts to provide a framework for this analysis in a general manner that will allow use by different detector designs and applications. These analysis methods are supported by a range of unique features that allow fast, accurate, and multidimensional data analysis. UMImaging also incorporates the newest algorithms originally developed for imaging and detection on 3D-position-sensitive CZT detector arrays [5], but implemented for any position-sensitive detector.

Other software for some of these imaging tasks already exists, such as for Compton-telescope simulation and image reconstruction [83]. Our design starts from the ground up, looking at what features are required for position-sensitive-detector data analysis, then designing a flexible, modular framework to handle this analysis in a general manner.

Currently, the main components of UMImaging have been implemented. Additional modules and functionality are added continuously as they are needed. The library currently stands at 125 C++ classes with over 30,000 lines of code.

This paper describes the design and unique features of UMImaging. The first section describes the software architecture design and descriptions of the implemented reconstruction methods. This is followed by discussions of features of the code and example usage. Finally, other distinctive features are mentioned.

Design

UMImaging is designed to allow a diverse set of analysis methods to operate in the same framework with maximal reuse of code. The structure centers around an interconnect class which holds a number of reconstruction methods and directs work, as shown in Figure 5.1. The interconnect acts as a buffer between the user interface and calculation classes, allowing different types of user interfaces for different types of applications, such as GUIs for different applications, command-line interfaces, or scripts. This also allows the addition and modification of reconstruction methods without having to modify the user interface. A graphical user interface is implemented using the Qt library [84].

Below the interconnect, design centers around reconstruction methods and supporting classes. A reconstruction method is any algorithm that performs any type of analysis. Such methods may ask for events and geometry information, perform calculations based on user specified options, and/or return or save results. Multiple reconstruction methods of the same or different types can run simultaneously. A wide variety of reconstruction methods are available in UMImaging, each implemented in its own class, with inheritance from related classes as necessary.

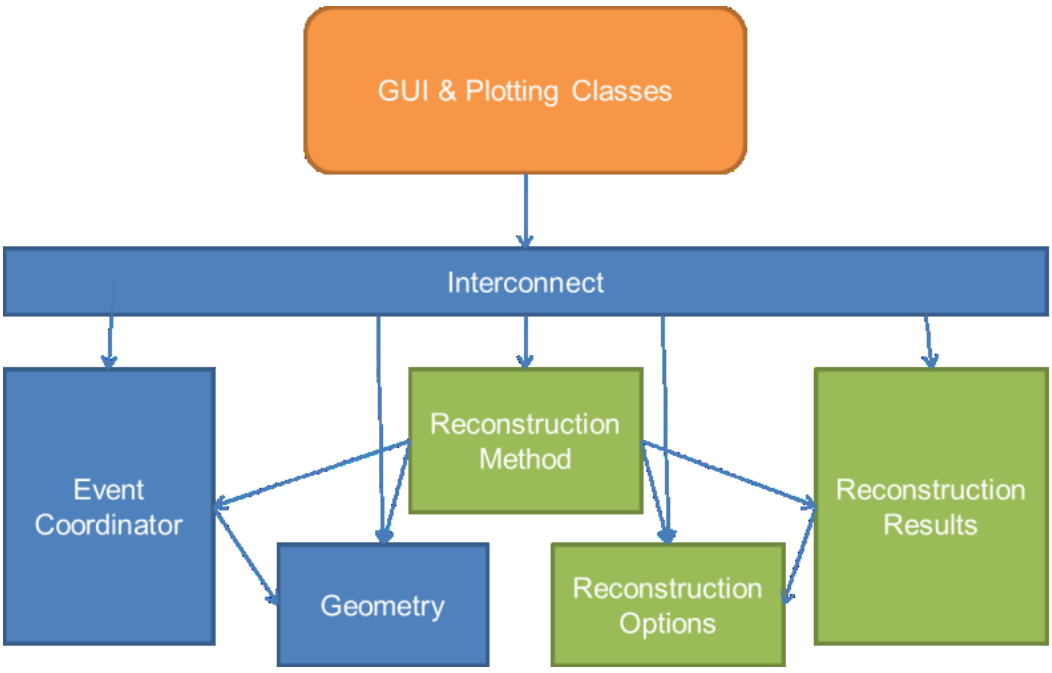

Figure 5.1. The basic design of the software library. Boxes represent classes or collections of classes and arrows represent the direction of communication. Only one instance of the interconnect, event coordinator, and geometry are created, but there may be multiple instances of the reconstruction classes.

Each reconstruction method points to an event coordinator that provides the reconstruction method with events to analyze. An event is a collection of coincident signals that have already been calibrated to give energy deposition for each detector voxel. The event coordinator seamlessly provides events from different sources (different format files or internet connections) to the reconstruction method without the reconstruction method needing to know the source. The event coordinator provides all active reconstruction methods with events when they are requested.

There is also a time-dependent global geometry that is available to supply the reconstruction methods with detector location, orientation, and specifications. It also includes functions for geometry calculations, such as ray tracing. Currently, this geometry includes an arbitrary arrangement of rectangular box detectors with corresponding information about their material and resolutions. These detectors are collected into a detector array tied to each file or internet-connection event stream. Time-dependent geometry can be passed in by the internet connection or read in from a file in XML format [85]. An example of such a geometry file is show in Figure 5.2.

```
<geometry>
  <state time="0">
    <array arraynumber="0">
      <newaxes>
        <position p1="0" p2="0" p3="0"></position>
        <xaxis x1="1" x2="0" x3="0"></xaxis>
        <yaxis y1="0" y2="1" y3="0"></yaxis>
        <righthanded truth="1"></righthanded>
        <dimensions xlength="2" ylength="2"
          zlength="1.5"></dimensions>
      </newaxes>
      <detector material="CZT" pixelpitch="0.172"
        pixelmap="eVProducts" depthuncertainty = "0.1">
        <position p1="0" p2="0" p3="0"></position>
        <xaxis x1="1" x2="0" x3="0"></xaxis>
        <yaxis y1="0" y2="1" y3="0"></yaxis>
        <righthanded truth="1"></righthanded>
        <dimensions xlength="2" ylength="2"
          zlength="1.5"></dimensions>
      </detector>
    </array>
  </state>
</geometry>
```

Figure 5.2. An example XML file of a geometry specification. In this example, a single 2 cm × 2 cm × 1.5 cm CZT detector is placed at the origin at time zero.

Analysis methods often have a number of free parameters for reconstruction, such as what subset of events to use or how much to smooth. Therefore, each reconstruction method has a set of reconstruction options that are stored in a tree structure grouped by category. Reconstruction options, ranging from the number of events to analyze to what technique to use, can be added at any point in this tree. Each option class has its own methods to validate the input or limit it to a certain range. The tree coordinates inter-relations between options. The option tree can be read/written to XML and converted automatically to a dialog window in the GUI for ease of changing parameters and options. Figure 5.3 shows an example of such a dialog.

Finally, analysis methods have output that is often multidimensional. Reconstruction results and helper classes are responsible for managing this multidimensional database as it is output from the reconstruction method. Reconstruction results can hold any binned output from the results of reconstruction methods and currently has multidimensional data as a function of direction, space voxel, moving target, energy, detector array, detector, type of event, locations and geometry of

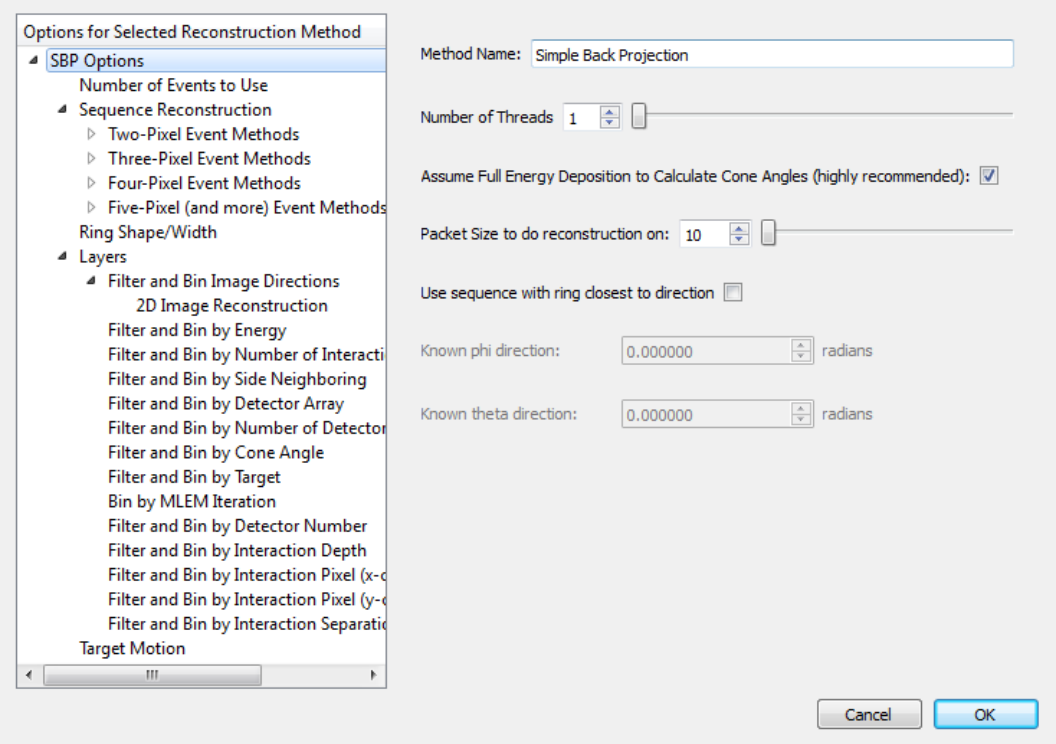

Figure 5.3. An example options dialog for simple back-projection.

interactions, reconstructed cone angle, and iteration. Plotting classes can query the reconstruction-results object, by way of the interconnect, to obtain a reduced data set with binning over a limited number of the original dimensions with filters applied to other dimensions.

Reconstruction Methods

Figure 5.4 shows the reconstruction methods in UMImaging and the relationships between them. All methods inherit from 'Reconstruction Method', which provides general functionality such as keeping track of whether the reconstruction is running and how much progress it has made. It also stores basic options and keeps track of where results will be stored. 'Simple Back-Projection (SBP) Reconstruction' inherits from 'Reconstruction Method' to get all of its functionality. 'SBP Reconstruction' also serves as a base class for 'Probability-Based Reconstruction'; methods to perform back-projection from 'SBP Reconstruction' are needed in 'Probability-Based Reconstruction'. 'Probability-Based Reconstruction' cannot run on its own. Rather, it provides functions

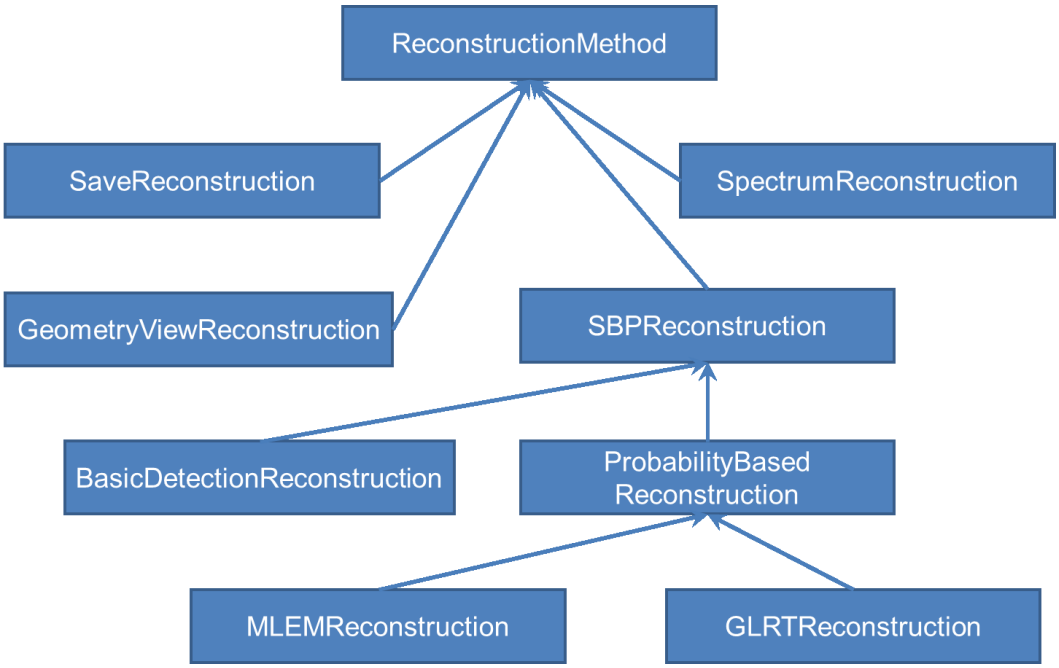

Figure 5.4. The inheritance structure of current reconstruction methods. The direction of the arrow defines the "inherits from" relationship.

to calculate and manage a list-mode system matrix for the collected events. These functions are used by other algorithms that perform probability-based reconstruction.

All reconstruction methods run in real time; events are processed as soon as they are available. If the event rate is too high to keep up computationally, even with parallel computation, events will be processed in the order of arrival, with a jump to more recent events after a set event lag has built up.

Traditionally, MLEM is a post-processing algorithm in that a set of data is collected, then analyzed all at once. In real-time imaging, it is useful to do as much of this as possible while events are being collected. So, in UMImaging, a row of the list-mode system matrix and the contribution to the initial iteration is calculated when each event becomes available. Iterations of MLEM are performed whenever the algorithm has excess computation time while waiting for the next event it will analyze. Events that occur during this iteration stage are stored and analyzed as soon as the iterations finish. In this way, the MLEM image can be calculated in real time for reasonable count rates. For long measurements, it is important to realize that since the number of collected events is large, each iteration can require significant time. However, by only performing iterations when there is excess computational time, the time between iteration

calculations will also increase since the events collected during the previous iteration stage must first be added to the system matrix. This self-limiting behavior allows iterations to occur only as often as the computation load can be sustained. In addition, to reduce iteration time, the system matrix could be shrunk to include only the most recent or best events after a long measurement has occurred. For high count rates, the iteration stage will occur rarely if ever. Instead, the iterations may be calculated after a predetermined interval of collected events set by the user.

Multidimensional Data

One of the features of UMImaging is its support for highly multidimensional data analysis. This allows the user to easily view and filter complex data. For instance, one could choose to reconstruct a spectrum just using events with three or more recorded interaction locations, or one could choose to reconstruct a spectrum for all types of events but bin by the number of interactions in the event. In the latter case, the user could then look at the spectrum due to all events or the spectrum only from events having a certain number of interactions. The options for each reconstruction method contain options to set the multidimensional filters and binning for the reconstruction results. Before the reconstruction begins, these binning and filter parameters are set for each dimension either by the user or from defaults. When reconstruction begins, the data structure is set to hold this multidimensional histogram, as shown in Figure 5.5. A contiguous memory space holds the database with each dimension nested into the one above it. During reconstruction, each event, or the image associated with each event, can be added to the appropriate multidimensional bin if it passes all of the filters.

For simple linear binning, as would be performed for spectrum reconstruction or SBP, for instance, each event can simply be summed to the data already present. First, each event is checked to make sure it passes the filters for each dimension. For instance, if a filter is applied to just a small energy range, events with total energy above and below this will be thrown out. For the remaining events, the algorithm next finds the bin location in each dimension with which this event is associated. For multivalued parameters, such as interaction location, which may have several values for a single

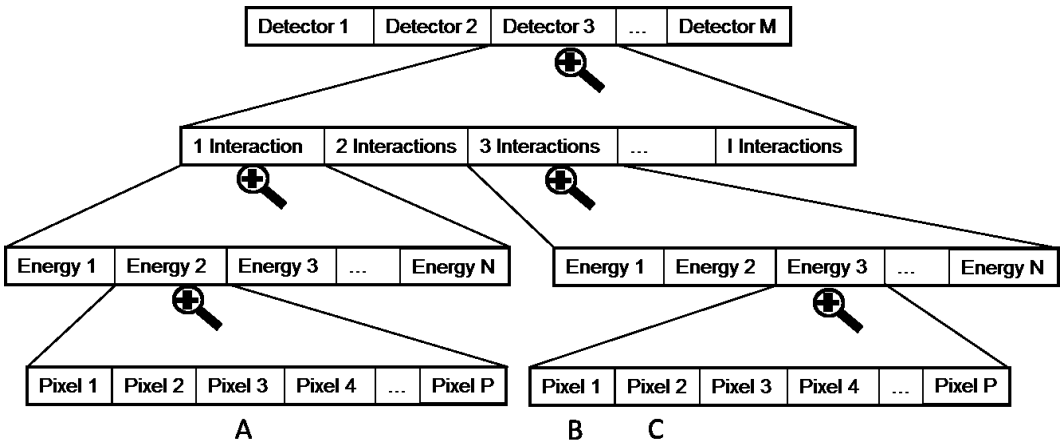

Figure 5.5. An example multidimensional data structure implemented in UMImaging. (A) represents the reconstructed intensity for single-interaction events in the third detector that have total energy in the second energy bin and reconstruct to the third pixel of the image. (B) and (C) represent the fraction of reconstructed intensity for three-interaction events due to events from the third detector that have total energy in the third energy bin and reconstruct to image pixel 1 or 2, respectively. The length of each dimension is *M*, *I*, *N*, and *P*, respectively for detector, interactions, energy, and pixel. The index of location (A) is $2 \cdot PNI + 0 \cdot PN + 1 \cdot P + 2$. The index for location (B) is $2 \cdot PNI + 2 \cdot PN + 2 \cdot P + 0$.

event, all possible locations are returned. From these, the multidimensional index *I* is calculated as

$$I = x_0 + \sum_{d=1}^{D-1} \left(x_d \prod_{i=0}^{d-1} N_i \right) \tag{5.1}$$

where x_i and N_i are the zero-indexed bin location and the number of total bins in dimension *i*, and *D* is the number of total dimensions. This is repeated for each possible combination of bin locations for the event (to allow for multiple parameter values). Finally, a fraction of the intensity is added to each unique multidimensional index location. The fraction guarantees that an integral over all bins will produce the total intensity. For instance, for a spectrum reconstruction of a two-interaction event with one interaction in detector module 1 and one interaction in detector module 3, 0.5 counts will be added to the multidimensional index location for detector 1 and 0.5 counts will be added to the multidimensional index location for detector 3.

When displaying data from this database, the software must be able to select subsets of events and convert the multidimensional data set into reduced dimensional data for display. This is performed with a recursive function that works through each

dimension of the data and "folds over" those dimensions that will not be included in the final data. For instance, in the example in Figure 5.5, if the user wishes to see the total image from two- or three-interaction events that involved the first detector, regardless of the total deposited energy, the sum

$$vNewData[p] = \sum_{m=0}^{0} \sum_{i=1}^{2} \sum_{n=0}^{N-1} vData[m \cdot PNI + i \cdot PN + n \cdot P + p] \qquad (5.2)$$

is calculated, where P, N, and I are the lengths for the pixel, energy, and interaction dimensions.

This is implemented in the general case with a recursive algorithm. First, the bin locations i in each dimension p that pass the new filters are identified. These are stored into a two-dimensional vector called *vBinLoc*. These may be a continuous or discontinuous range of bin locations. Also, flags are set for those dimensions that the final data needs to be binned over and stored to the one-dimensional vector, *vBinByParm*. Finally, a new vector associated with the original multidimensional data is defined where

$$\text{vBinsBelow}[p] \equiv \begin{cases} \prod_{i=0}^{p-1} N_i, & p \neq 0 \\ 1, & p = 0 \end{cases} \qquad (5.3).$$

A similar vector, *vNewBinsBelow*, is defined for the down-binned data set. The vectors holding the original data and the down-binned data are named *vData* and *vNewData*, respectively. Then, the recursive function pseudo-code in Figure 5.6 can be called starting with *origLoc=newLoc=0* and p equal to one fewer than the number of total dimensions. This will fill the new vector that holds the reduced data set. This can also contain integration constants to include factors such as solid angles, depending on the units of the original data vector.

Figure 5.7 through Figure 5.9 show examples of multidimensional binning for an 18-detector 3D-position-sensitive CZT detector array. It has two planes of 3 × 3 detectors, each of 2.0 cm × 2.0 cm × 1.5 cm with 11 × 11 anode pixels. Interaction depth is found from the cathode-to-anode ratio or, for multiple-pixel events, drift time. The system is capable of using multiple-pixel events for imaging, either from a single detector module or from distinct modules. The system is similar to that described in [5].

```
recursiveDownbin(origLoc, newLoc, p)
{
   if p == -1 //base case
   {
      vNewData[newLoc] += vData[origLoc];
      return;
   }
   for each element i in vBinLoc[p]
   {
      int nextOrigLoc = origLoc + vBinLoc[p][i]*vBinsBelow[p];
      int nextNewLoc = newLoc;
      if vBinByParm[p]
         nextNewLoc += vBinLoc[p][i]*vNewBinsBelow[p];
      recursiveDownbin(nextOrigLoc, nextNewLoc, p-1);
   }
}
```

Figure 5.6. The recursive algorithm for reducing the dimension of a data set. The variables are defined in the text.

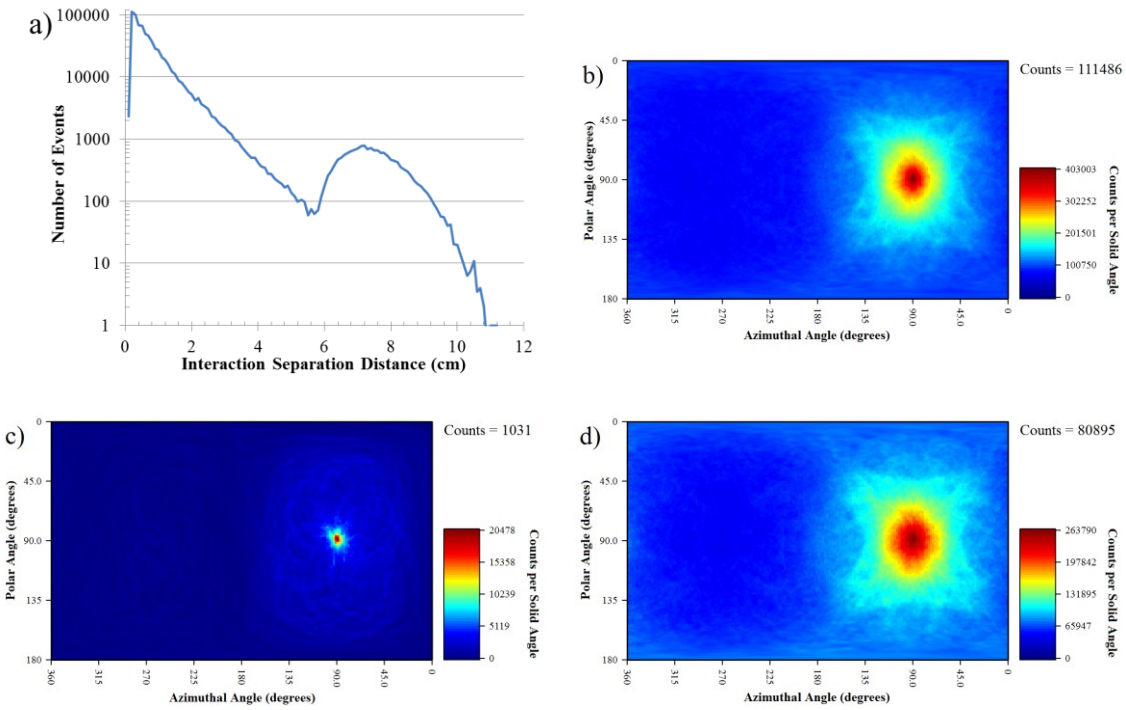

Figure 5.7. Images of a Cs-137 point source measured with the 18-detector system using various cuts on interaction separation distance. (a) The distribution of the interaction separation distance for two-interaction events. (b) The image of all two-interaction events. (c) The image of two-interaction events with separation distance of 7.5 cm or greater. (d) The image of two-interaction events with separation distance of 1 cm or less.

Figure 5.7 shows a series of images from a simple experiment of a single Cs-137 point source on one side of the 18-detector system. Simple back-projection Compton

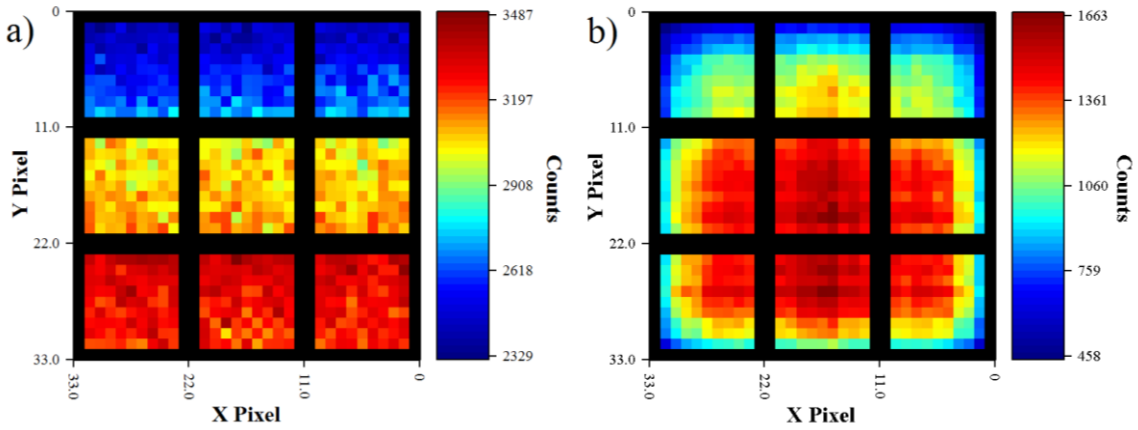

Figure 5.8. The distribution of interaction locations over one plane of the 18-detector system for a simulated parallel beam coming from the bottom of the figure. Only full-energy deposition events are kept. There are 3 × 3 detectors, each with 11 × 11 pixels, separated from the other detectors by 2 mm. The detector edge pixels are excluded since they appear hot due to slightly larger area. (a) includes only single-interaction events and (b) includes only three-interaction events.

imaging of two-interaction photopeak events produces the image in Figure 5.7b. However, one can bin the images by the separation distance between the interaction locations. The separation-distance distribution is shown in Figure 5.7a. Most interactions happen very close together, but because of the 5.5-cm separation between the two planes, there is a collection of events that interact many centimeters apart. Imaging just those interactions that are farther than 7.5-cm apart produces Figure 5.7c. This has significantly better angular resolution (below 10° FWHM) because the long average separation distance reduces uncertainty in the cone axis. Conversely, imaging just those interactions closer than 1 cm produces Figure 5.7d. This has near 50° FWHM because of large cone-axis uncertainty.

With multidimensional binning, one can investigate the distribution of interactions among and within detectors as a function of the number of interactions that occurred. Figure 5.8 shows the results of a simulation of a parallel beam of 662-keV photons incident at one end of the planes. Figure 5.8a shows the locations of single-interaction events depositing full-energy and Figure 5.8b shows the locations of three-interaction events depositing full-energy. Each shows the relative number of total interactions which occurred in each pixel in the nine detectors in one plane. Each

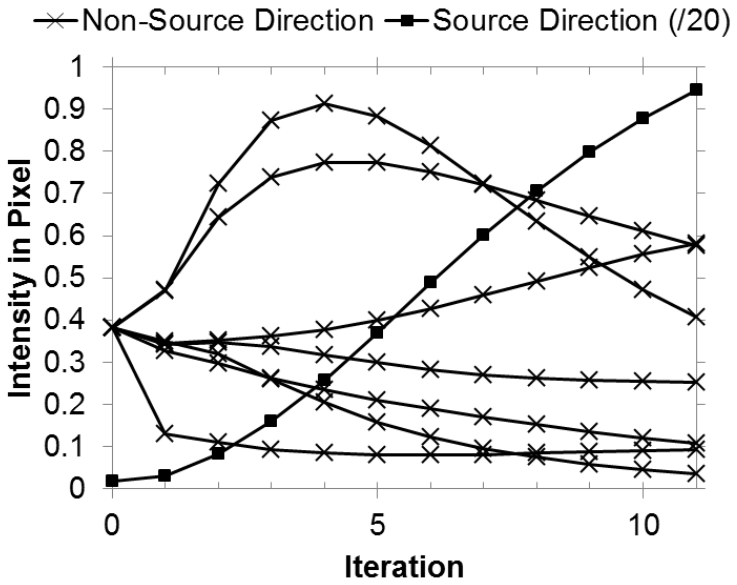

Figure 5.9. The intensity in selected pixels of a 64 × 64 image summed over 500 energy bins during iterations of the energy-imaging integrated deconvolution algorithm. The source-direction pixel scale is twenty times that of the other pixels. All pixels start with the same average value. The data are 1000 multiple-interaction events from a Cs-137 measurement.

complete event contributes one count, so each interaction in a three-interaction event contributes 0.33 in these figures.

It is clear that there is a significantly different distribution of counts in these two cases. Full-energy single-interaction events are most common near the entrance surface (y-pixel of 33) and decrease slightly due to attenuation towards the other side (y-pixel of 0). Full-energy three-interaction events are most common near the center of the array, slightly shifted towards the photon entrance side. This is because those photons scattering near the center of the array are more likely to be captured elsewhere in the array, becoming a multiple-interaction event. Similarly, interactions from multiple-interaction events occur more often at intermediate depths in the detector than at either surface. In this figure, detector edge pixels are ignored since they are slightly larger in area and hence receive more events and make the distribution of counts harder to observe.

When performing iterative reconstructions, it may be interesting to observe the value of parameters as iteration progresses. Figure 5.9 shows the value of selected pixels in a 64-pixel-by-64-pixel image of 4π space of a point source measurement through each

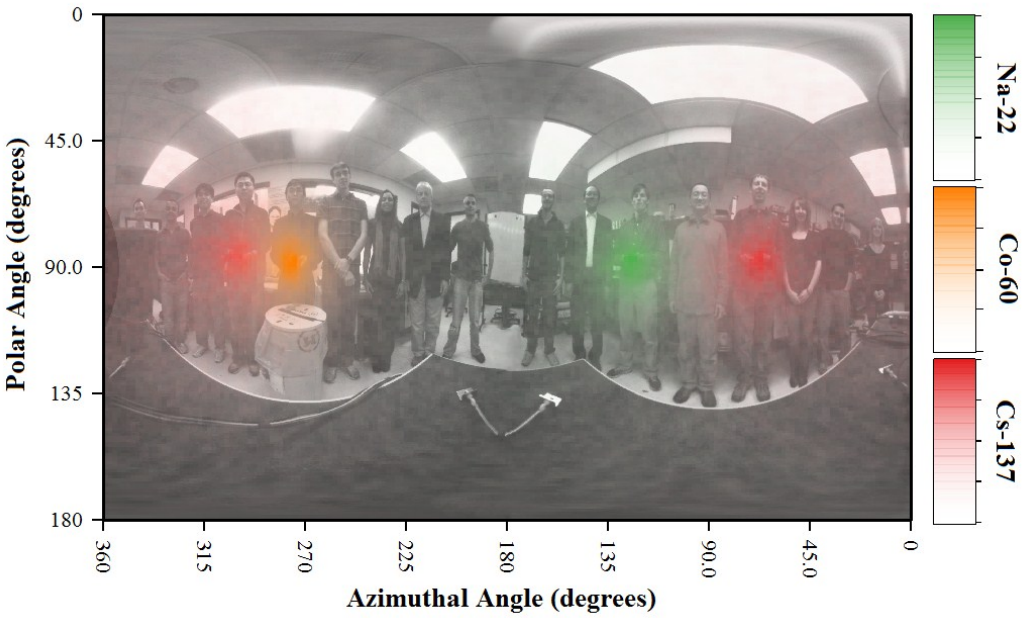

Figure 5.10. Isotope-binned image for Cs-137, Na-22, and Co-60. Each isotope is displayed as a different color in an intensity scale from transparent to opaque, and a current photograph of the environment is included as a backdrop. All can be created in a single reconstruction. The arbitrary-region binning can include multiple, non-contiguous regions of energy space into a single bin.

iteration of the energy-imaging integrated deconvolution algorithm [51]. Through the iterations, the source direction pixel becomes hottest while most other directions decrease in intensity.

For parameters with continuous values, it is possible to define bin regions as arbitrary unions of ranges. This allows, for instance, the energy dimension to act alternatively as an isotope dimension. Events with energy falling in any of the photopeak regions of a particular isotope can be added to a single bin for that isotope. Events with an energy near photopeaks from multiple isotopes can contribute to each isotope. One application of this it to separate an image by the components due to each library isotope instead of the component due to each energy range. Figure 5.10 shows an image of three isotopes in our laboratory using this binning.

Parallel Performance

Since increases in computing power are expected to occur more through increased parallelization than serial clock speed, support for parallel algorithms is important to

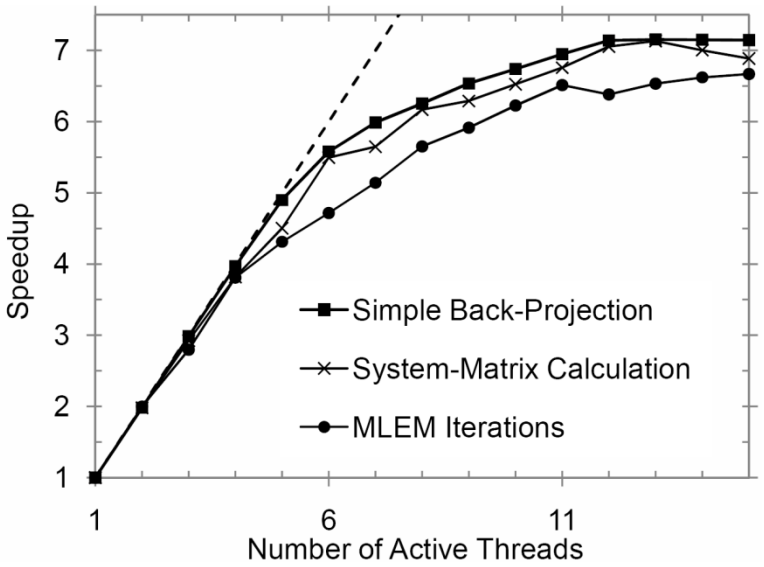

Figure 5.11. The speedup in calculation time due to running in parallel on six CPU cores with hyper-threading enabled as a function of the number of active threads for three types of calculations. Parallel implementations of these algorithms allow a near seven-fold increase in execution time over a single processor (with or without hyper-threading technology) by efficiently using all available processing resources.

allow reconstruction to occur as close to real time as possible. Many image reconstruction algorithms are embarrassingly parallel, in that the reconstruction for each set of events can occur separately and be combined at the end. We expect large speedup from such methods, so parallel implementations are desirable. Most classes are therefore thread safe – multiple processors can access the same functions simultaneously to do things like get events or perform geometry calculations. For parallel implementation, we chose to use OpenMP [86] because of its relative simplicity and presumed future support. This restricts use to a single shared-memory machine. However, since each reconstruction method only interacts with other data structures through a handful of simple commands, expanding to reconstructions using a distributed memory machine could be done on the level of the reconstruction method without significantly changing the rest of the architecture.

Figure 5.11 shows the speedup obtained for three types of calculations on a system with six processor cores. Simple back-projection, working on data from a CdZnTe pixelated detector, reconstructs 2×10^4 events to a 3.24×10^4-pixel mesh in 104 s using a single processor, but it only requires 14 s when running in parallel on six

processors. More than a factor-of-six speedup is possible because the processor's hyper-threading ability allows each core to be used more efficiently. Very similar speedup was also obtained for iterative image reconstruction using the maximum-likelihood expectation-maximization (MLEM) method [51] on 1000 events with a fine energy and spatial mesh. MLEM requires extensive geometry calculations during the creation of the system matrix and extensive RAM access while performing iterations. Figure 5.11 shows the speedup for both steps of the calculation.

Other Design Features

Besides those described above, UMImaging has a number of other features that support analysis of events from 3D-position-sensitive detectors. Some of these are listed below.

- Many different operating systems are used commonly in data analysis, including multiple flavors of Microsoft Windows, Linux, and Mac. UMImaging is written in standard C++ to be platform independent with minimal use of outside libraries. This allows the same version of UMImaging to be used on all systems. The GUI and parallelization libraries are currently supported (and expected to be supported in the future) on all three systems.
- The software should be accessible to new users, both in the use of current libraries and in how to extend them for other purposes. The code is therefore highly documented with over 5000 header comment lines parse-able using the Doxygen documentation generator [87].
- Optical images overlaid with gamma-ray images can be useful to locate source positions accurately. UMImaging supports overlaid 4π optical images streamed over the internet or read from file.

Conclusions

A software package has been designed and implemented to support a wide range of data-analysis tasks from 3D-position-sensitive detectors. The software is platform independent and easily extensible. The use of an interconnect allows the user interface to

be specialized to the application and include graphical and other user interfaces such as scripts and command-line operation. A modular framework allows advanced reconstruction methods to be developed more quickly using existing tools to set options, get events, perform geometry calculations, manage output, and display results. Options are stored in a tree structure, which can easily be converted to a GUI interface. Event managers transparently pass events from files or internet connections to each reconstruction method as the events are available. Geometry classes can store any arbitrary geometry (currently arrays of boxes) and perform geometry calculations needed in image reconstruction.

The framework supports high-dimensional data output. Each dimension of the data can have an arbitrary binning structure, which, for instance, allows isotope binning in the energy dimension. Multidimensional data analysis requires functions to bin events quickly and down-binning functions to view the data. Examples have been presented for several situations, including the distribution of interaction locations within a detector array and the image quality as a function of interaction separation distance.

A range of reconstruction methods have been implemented through the course of our work and are able to process events from arbitrary detector geometries in real time. MLEM achieves pseudo-real-time by applying extra processing time to perform iterations. Throughout, classes and functions are designed to work in parallel. Parallel implementations of reconstruction methods show near seven-fold speedup on a six-processor computer with hyper-threading. New features and algorithms continue to be added to this software package to support new analysis methods for 3D-position-sensitive radiation detectors.

Chapter 6

Detection Analysis

> *If you hired me to walk through the cellars of Washington to see whether there were atomic bombs, I think my most important tool would be a screwdriver to open the crates and look. I think that just walking by, swinging a little gadget would not give me the information.*
> – J. Robert Oppenheimer, 1946 *[88]*

The automated detection of gamma-ray-emitting threat sources within a gamma-ray background is of interest in many applications, such as security and non-proliferation, where a threat source, such as special nuclear material or a radiological dispersal device, may be hidden in some environment. Although well-trained operators are often able to observe the detector output and make informed decisions based on unexpected features in the count rate, spectrum, or image, automating detection decisions allows inexperienced operators to use the system and may give better performance than the most expert of operators. While the best way to discover a threat source – considering economic consequences, international cooperation, privacy, expense, and other factors – is a matter of debate, this chapter will investigate the more limited subject of automated passive detection of gamma-ray-emitting threat sources using spectroscopic 4π-imaging detectors.

In this chapter, we are primarily interested in understanding the conditions under which imaging improves detection and quantifying those improvements. This chapter will give some background on non-imaging systems and contrast their common detection methods with what may be achievable using an imaging system. This is followed by new analysis of the contribution of imaging to the automated detection of point sources on the 18-detector array system. The steps of this analysis can be replicated in future systems to quantify their imaging performance and inform comparisons between imaging systems.

Because much of this analysis only relies on the simplest imaging algorithms, the performance measures can be more algorithm independent than if the performance of advanced detection algorithms were compared for a number of specific situations. In addition, the results of the analysis report the contribution to detection from each type of event, which can be used to improve the system design.

It is useful to classify automated detection approaches by the type of information they work on: count rate, spectrum, image, or a combination of these. In some situations, prior background measurements may be available, but in many others, the gamma-ray background may be relatively unknown. Similarly, the identity of this source may or may not be known. In addition, the search may occur over a wide area, or may focus on standoff detection in a known target region. The best approach to each of these situations may be different. The following sections will give an overview of current detection technologies, and then present an analysis of detection ability for many of these situations.

Overview of Detection Systems

Gross-Counting Systems

Among the most basic of radiation-detection systems is the gross-counting system. This reports a number of received counts, so automated detection occurs through the count rate. A Geiger-Mueller counter is the canonical example of this type of system. Polyvinyl-toluene plastic scintillators may also operate as gross-counting detectors in portal monitoring [89,90]. The fundamental operating principle follows the equations standardized by Currie [91], and sketched much earlier, in which detection occurs when the measured count rate is above the rate that is probable to be produced by just background [3]. Gross-counting systems hence need an estimate of the background intensity from a prior measurement and a model of background variation. For instance, in gross-counting portal monitors, numerous studies have attempted to better track the background levels in order to better estimate the background intensity [89,90,92,93,94].

Gross-counting detection methods are therefore susceptible to performance degradation due to spatial and temporal shifts in the background that reduce the quality of

the background estimate [89,90,95,96]. This may occur, for instance, when moving the sensor to a new environment containing a hidden source. Then, as a new background is estimated, it will likely incorrectly include a contribution from the hidden source that the system is trying to detect. The erroneously high background estimate will decrease sensitivity to additional counts introduced by the source. In gross-counting portal monitors, there is a concern that the slow approach of a source will skew the background estimate, allowing even a strong source to pass through undetected if moved slow enough [97]. Background variation can also occur over time due to weather conditions and time of day, and some of these changes can be quite rapid [90,98].

It should be noted that, in the hands of an experienced operator, handheld gross-counting systems can reveal the spatial distribution of activity. As the operator moves the system around the region, she mentally converts the scalar count-rate field due to geometric and material attenuation into an activity distribution. Operated as such, the system becomes a non-spectroscopic imaging detector and so may not suffer the same problems as a stationary gross-counting system. In addition, if targets are moving, the time-varying count rate may be used to identify the target distribution of radioactivity [95,99].

Spectroscopic Systems

Other systems attempt to reduce the need for an estimated background by observing the energy spectrum of incident photons. The full-energy peaks expected in most backgrounds are known, so automated detection occurs by observing the presence of other peaks associated with non-background isotopes [100,101]. Many software libraries have been written to automate this spectral peak finding and isotope identification [102,103,104,105,106,107,108], leading to a number of comparison studies [103,109,110,111], yet none has gained a widespread following and myriad new peak-search algorithms continue to be proposed such as [103,112,113,114,115,116,117,118,119,120] and [121,122,123,124]. Other spectral analysis algorithms rely on template matching – instead of searching just for spectral peaks, the entire spectral shape can be used to deconvolve or unfold the incident energy spectrum and hence the isotopic sources [125,126,127,128,129,130,131,132,133,134].

This method allows a greater fraction of events to contribute to the signal, but it is highly dependent on a good spectral-shape model or prior measurement [135]. Varying source direction or surrounding materials can skew the spectral shape.

One complicating factor in spectral detection and identification is the presence of significant shielding between the source and detector. Besides greatly reducing the signal from the source, shielding scatters gamma rays from the photopeak into the continuum, hampering peak detection and skewing template matching. In the later analysis, we will assume no shielding is present.

As long as the target isotopes have lines distinct from background lines, spectral-based detection can perform well. However, when the emission energies from a threat source overlap those found in background or naturally occurring radioactive materials (NORM), peak presence is not sufficient for detection. The background intensity and spectrum must be estimated in the same manner as in the gross-counting case, leading to similar performance degradation from spatial and temporal background shifts. Furthermore, the peaks expected from background in contaminated regions may not be known or may be at the same energy as threat sources, so detection must rely on background estimates.

Spectroscopic Imaging Systems

Since threatening sources are often point-like, other studies have developed methods to differentiate point sources from benign distributed sources by extracting the unique signatures of point sources on detector systems with spatially varying responses [95,96,99,136,137]. It has been shown that the addition of imaging – reporting the spatial or angular activity distribution – can improve the background estimates in the local region near the point source [96,136]. Therefore, a detector with imaging capability in an unknown background performs at least as well as a non-imaging detector in approximately known local background. Even in the case of exactly known background, Lingenfelter *et al*. have shown that a detector with uniform detection efficiency in direction with imaging capability always has better detection performance than a similar system without imaging capability [138]. With imaging, detection occurs by looking for

localized hot spots in an otherwise smooth background image, rather than by looking at deviations from a previously estimated background intensity or spectrum [139].

A detector that can produce an image in a single measurement has one other significant advantage over systems that require multiple measurements: localization of the possible source direction can occur in a single measurement. The resulting vector field that is traced as the system moves is more informative than a scalar count-rate field and hence benefits the *search* for localized sources. The improvement in search performance – being able to locate a threat source – is distinct from the improvement in detection performance – being able to know a threat source exists. The analysis that follows focuses on the detection, rather than search, problem.

Summary

Each system described above considers different types of information to make a detection decision, and each makes different assumptions about the environment. Table 6.1 summarizes some of these assumptions for different algorithms. Parameters labeled 'Will estimate' are parameters that the algorithm does not need to know before making a decision, but can be estimated in the process of making the decision. If the parameter value is known exactly or approximately a priori, it can be given to the algorithm to improve detection performance.

Statistics of Detection

Spectrum Only

In the following section, we will investigate the performance of the 18-detector array system in terms of the signal-to-noise ratio (SNR) for detection. Although the SNR may not capture all the subtleties of some detection algorithms, it is suitable for comparing the relative ease of detectability using different systems and signals.

First, consider operating the 18-detector array as a spectrometer to detect an isotope with known emission lines that has at least one emission line not present in the background. We will assume that the system knows neither the background spectrum nor the background count rate and is stationary. In this case, the unique signatures of the

Table 6.1. Types of information used and estimated by different types of standard detection systems. Variations on these systems exist. The last column describes the algorithms described later in this chapter.

		Gross-Counting Detection	Spectroscopic Detection	Gross-Counting Imaging Detection	Spectroscopic Imaging Detection	Current GLRT or SIT Algorithms
Background	**Count Rate**	Prior measure-ment or estimate	Will estimate	Will estimate	Will estimate	Will estimate
	Spectral Distribution	Not used	Will estimate	Not used	Will estimate	Will estimate
	Spatial Distribution	Not used	Not used	Will estimate	Will estimate (assumes smooth)	Assumes uniform
Source	**Emission Lines**	Not used	Cannot be in background	Not used	Will estimate	From source in library
	Number of sources	Not used	Estimate only if different isotopes	Will estimate	Will estimate	Will estimate
	Location	Not used	Not used	Will estimate	Will estimate (assumes point sources)	Will estimate (assumes point sources)
	Count rate	Will estimate	Will estimate	Will estimate	Will estimate	Will estimate
	Shielding	Not used	Can estimate	Not used	Will estimate	Can confuse

source are the full-energy peaks in the spectrum at the source-emission energies that are distinct from background lines. A template-matching scheme would also use the full-energy peaks (as features in the template) to determine the presence of the isotope since the background spectral shape is unknown. In a strong background environment, the continuum is quite uninformative even to a template-matching algorithm since down-scatter in surrounding materials can produce a range of continuum shapes. For optimal detection of these full-energy peaks when there is known source intensity and known background continuum, the likelihood ratio test (LRT) [61] says to threshold on the test statistic

$$\Lambda = \sum_i y_i \ln\left(\frac{C_{s_i}}{C_B} + 1\right) \tag{6.1},$$

where y_i is the number of recorded counts in energy bin i, C_B is the number of counts in each energy bin of the background in the region around the peaks, and C_{S_i} is the number of counts expected in the i-th energy bin of the source distribution. Poisson statistics are assumed in each energy bin. For small SNR, when the peak size is small compared to the background under the peak, this acts like a matched filter – i.e., a dot product is formed between the measured and expected spectral shapes. Here I have assumed, and will continue to assume, a single emission line distinct from background, but these equations can be generalized to look for multiple lines as well.

However, since both the source and background intensities are unknown, they first need to be estimated. A test that uses the estimated source and background intensities (rather than the true intensities) is the generalized likelihood ratio test (GLRT), which may no longer be optimal [61]. Instead of applying a GLRT here, for simplicity and consistency with future sections, I will simply count the events falling above the background continuum in the peak region for the test statistic. That is, the filter is square rather than matched, and the number of counts in the peak region is compared to the number of counts in neighboring regions. Such a test lends itself to easy SNR analysis without straying far from the optimal test.

A practical application of such a test would involve determining the continuum contribution in the spectrum on either side of the full-energy peak and subtracting this from the number of counts in the peak region. Usually, the continuum is approximately linear in the region around the peak, so the sum of the average number of counts in equal-sized regions to either side of the peak can approximate the continuum contribution under the peak. Then, if, for instance, the nearby regions for determining the continuum contribution together are as wide as the peak region, the number of counts in the continuum region B may be subtracted from the number of counts in the peak region P to estimate the source counts:

$$S = P - B \tag{6.2}.$$

The uncertainty in the calculated source counts is then

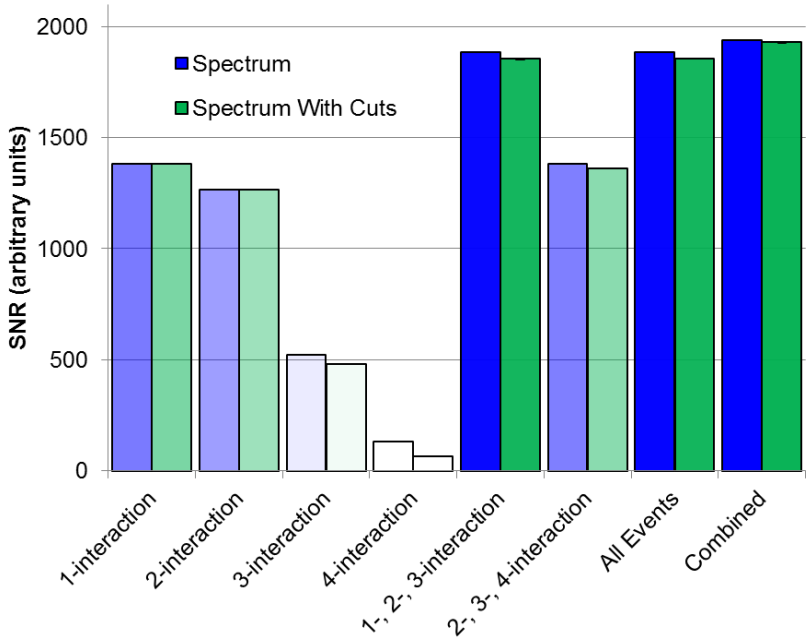

Figure 6.1. The relative experimental SNR for a Cs-137 source in background using the spectrum of various sub-sets of events recorded in 18-detector-array experiments. The 'combined' category has a different peak region width for each type of event and combines the counts from each into a single sum. The 'Spectrum with Cuts' series throws out three- and four-interaction events believed to be in the continuum. The transparency of each bar is proportional to the signal relative to the signal in the combined spectrum case.

$$\sigma_S = \sqrt{\sigma_P^2 + \sigma_B^2} = \sqrt{P + B} \approx \sqrt{2B} \tag{6.3}$$

where the last approximation holds if the source is weak or missing. Note that when the continuum level under the peak is known exactly, the variance in S is just

$$\sigma_S^2 = P \approx B \tag{6.4}.$$

Again, the last approximation holds when the source is weak or missing. Therefore, whether the continuum is known or unknown,

$$SNR \propto \frac{S}{\sqrt{B}} \tag{6.5}$$

since the uncertainty (noise) is proportional to $\sqrt{B}$. In the following sections, I take the peak region to have the width that gives the optimal SNR. For a Gaussian approximation of the peak and in practice, this is 1.2 times the FWHM of the peak.

Figure 6.1 shows the results of applying this test to a long measurement with particular background and Cs-137-source strengths using the 18-detector system. Each

bar shows the SNR achievable for different subsets of events. However, because different signal or background strengths will scale the vertical axis by the same fraction for all subsets of events, the relative SNR comparison is applicable for any signal and background strengths. The 'Spectrum' series uses all events that are thought to interact a certain number of times. It is clear that events that interact the fewest number of times give the highest SNR. This is because these events offer better energy resolution and they are more common. With better energy resolution, the signal in the full-energy peak is situated in a narrower region, so the contribution of the background continuum under the peak is therefore reduced, improving the SNR. More events reduce the relative variation in the background continuum, also improving the SNR. On the other hand, if we exclude events that only interact once or twice in the detector, we preferentially exclude Compton-continuum events. Therefore, the limited set of events that have interacted many times will have a larger peak-to-total ratio, which should improve the SNR by reducing the continuum level under the peak. However, it is clear from Figure 6.1 that the additional decrease in peak statistics and increase in peak width for multiple-interaction events overwhelms any improvement.

In the 'Spectrum with Cuts' series, three- and four-interaction events in which there is poor agreement between the observed scatter angles and the scatter angles calculated through energies are eliminated. The calculated scatter angles assume full-energy deposition, so agreement only occurs when the gamma ray deposits its entire energy in the system. Therefore, the preponderance of the remaining events are full-energy deposition. This further accentuates the large peak-to-total ratio [140], but the inadvertent reduction in true peak events reduces the signal significantly enough to reduce the SNR as well.

The tradeoff between statistics, resolution, and peak-to-total ratio for detection of Cs-137 is easily seen by combining events in different combinations. All events together give better counting statistics than only single-interaction events, with only some loss of resolution, resulting in better SNR. The four-interaction events contribute little to increased statistics and degrade the overall resolution, therefore the addition of four-interaction events may even slightly decrease SNR in this system. The wider peak region

needed to cover the peak will incorporate a larger amount of background while contributing only a small portion to the signal.

It should be noted that it is possible to include all events together in a single test by accurately modeling the uncertainty in each. The best SNR in this example can be achieved by taking all events and splitting them into spectra by the number of recorded interactions. Then, taking the number of events within the optimal width of the peak on each spectrum and summing contributes the largest signal relative to the background continuum noise, resulting in slightly better SNR, as shown in the 'Combined' category in Figure 6.1.

For isotopes other than Cs-137, the energy of the emissions will produce different fractions of each type of event, in which case, the tradeoff between statistics and spectral shape may be different.

Spectrum and Image

Now, consider a simple detection scenario involving an imaging detector, in which the user knows both the target direction and source identity before commencing the measurement and attempts to detect the source. As before, the target emits at least one energy line not present in the background, but the background count rate and spectrum are otherwise unknown. In this case, the signatures of the source are a hot spot in the direction of the target and peaks in the spectrum at the source-emission energies. An appropriate algorithm could use all events, but for ease of analysis, the signal could be those events (1) that fall above the extended background in the image in the direction of the target or (2) that fall above the background continuum in the spectrum at the expected energies of the source. For the 18-detector array, the events in the second set are generally larger than those in the first set since, in most cases, accurately estimating the correct total energy is simpler than estimating the possible direction. Put another way, all correctly imaged events usually appear in the full-energy peak, but not all full-energy peak events will be reconstructed through the correct direction. For this reason, it makes sense to identify the signal as the number of events that fall in the full-energy peaks of the spectrum after filtering out events that do not pass the source direction,

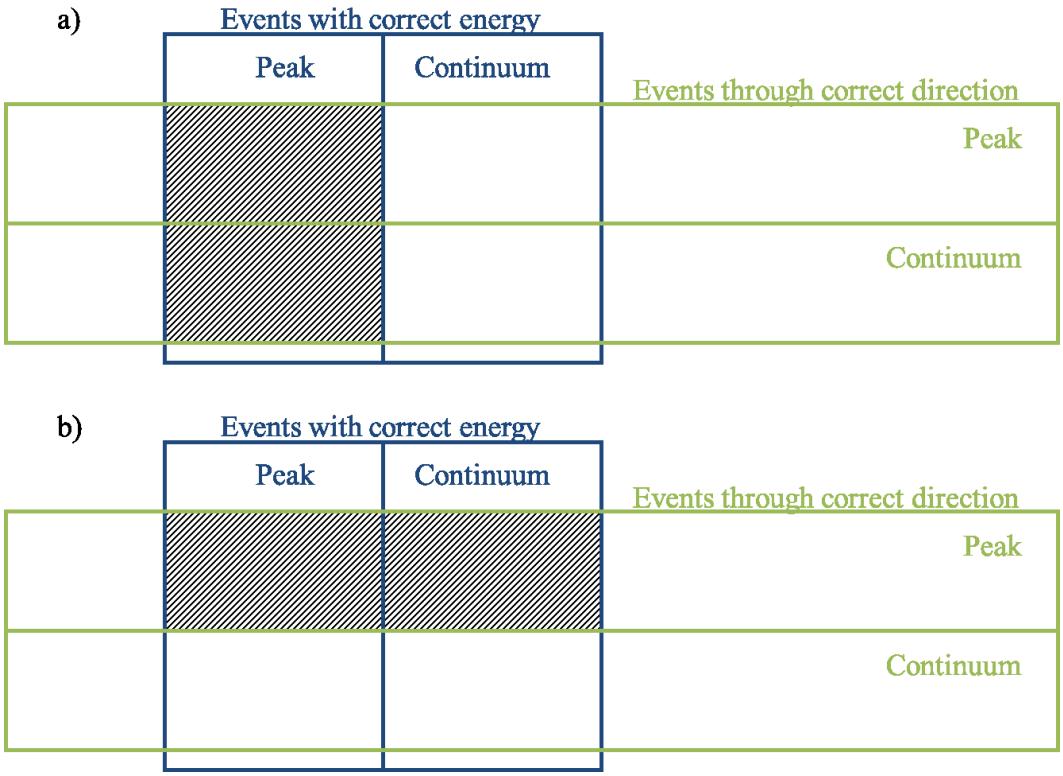

Figure 6.2. The set of events (a) that fall in the full-energy peak of the spectrum after filtering out events that do not pass the source direction, and (b) the set of events that fall in the hot spot of the image after filtering out events not in the correct energy regions.

rather than the number of events that fall in the hot spot of the image after filtering out events not in the expected energy regions (see Figure 6.2).

Note that the original spectrum information still exists; it is just spread over the extra image dimensions. This is shown in Figure 6.3. Two parameters exist, derived from the original spectrum and the spectrum in the source direction. An algorithm could simply ignore the image dimensions and have the same performance as the spectrum-only case, as along the horizontal axis in Figure 6.3. Alternatively, an algorithm can rely completely on the spectrum in the source direction in the image, as along the vertical axis in Figure 6.3. A suitable algorithm could therefore always perform at least as well as just spectral detection by appropriately combining these two parameters using the (possibly paltry) image dimension as a check for its spectral detection. However, these two parameters are quite correlated, as shown in Figure 6.3, therefore, the improvement gained from their combination is quite small. For instance, transforming the two-

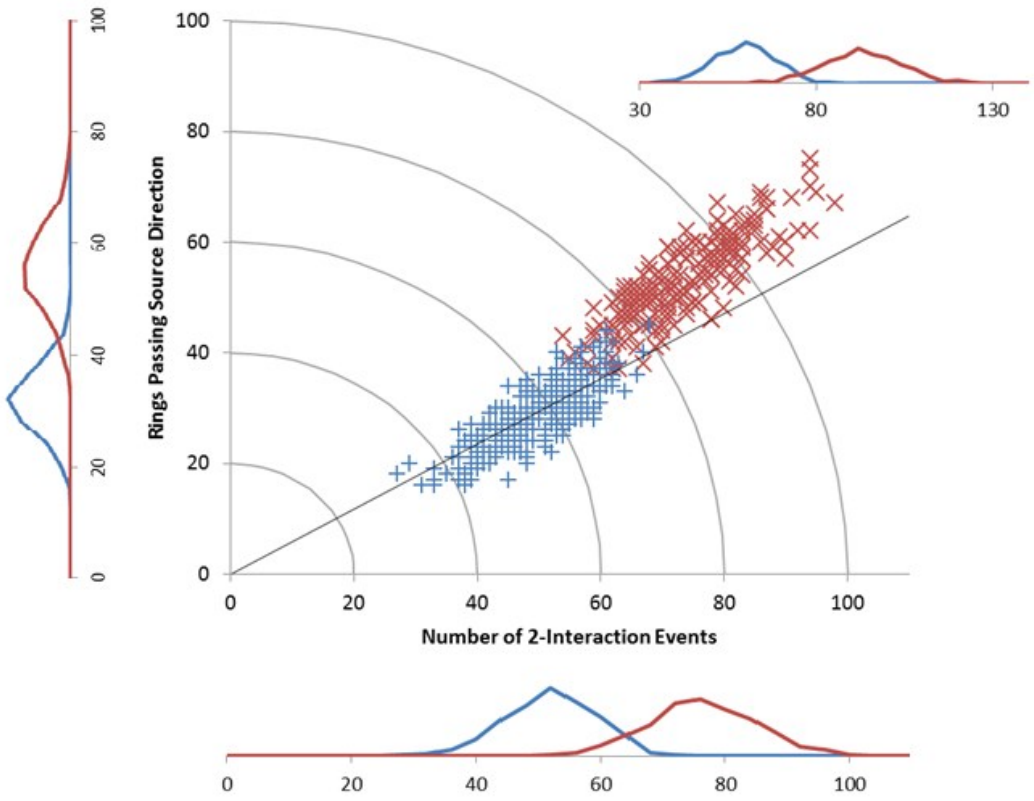

Figure 6.3. The correlation between the number of two-interaction events in the photopeak region around Cs-137 and the number of rings passing the source direction in the image. The blue is background and the red includes the source. The SNR is not significantly improved by using the distance from the origin instead of the number of rings passing the source direction.

dimensional data to a one-dimensional data set by the distance of each sample from the origin does not improve SNR over the SNR of the best of the axes (the vertical axis, in this case). Therefore, I will simply choose to use the axis that provides the better SNR, remembering that slight improvement is possible by intelligently using both together.

We have investigated the spectral axis above. Here we investigate the image axis by throwing out events that do not pass the known source direction. This acts to decrease both the background and signal contributions. Therefore, the modified SNR is

$$SNR \propto \frac{f_S S}{\sqrt{f_B B}} \tag{6.6}$$

with two factors f_S and f_B that decrease the number of source and background counts. We expect the decrease in background counts to be considerably more than the fractional decrease in source counts if we have a good imaging system. As long as

$$\frac{f_S}{\sqrt{f_B}} > 1 \tag{6.7}$$

the SNR will be larger by filtering of events based on the image. Therefore, to see the relative SNR improvement due to the addition of imaging, one can calculate this term with different subsets of events, analogous to the procedure in the previous section. This acts as a figure of merit for the imaging system.

As in the last section, I will use a simple count of events over the region. Considerably better performance is possible using a more complex function to weight events based on the distance of the ring from the center of the region. However, for simplicity in understanding the statistics and comparing data, the simple sum is used. Assuming a Gaussian-shaped hot spot, the optimal width of the cut for the sum should be rings that pass within 1.2 times the hot spot's FWHM as in the spectral case above. To achieve this, I widen the rings and take the value in the direction of the source rather than using narrow rings and integrating over a region in the image. I also use the best-performing sequence-reconstruction methods to choose the most-probable sequence for each event before performing SBP. These are the simple comparison method for two-interaction events and the MSD method for three- and four-interaction events. This gives better results than using all possible sequences.

Figure 6.4 shows the SNR improvement factor as a function of the ring width (from center to edge) for different subsets of events. It is clear that the best improvement factor occurs for two-interaction events, although the improvement is only about 43% at the best width. As shown in Table 6.2, at the optimal width, a surprisingly small fraction of events contribute to the signal in the target direction. For four-interaction events, only 40% of source events are reconstructed through the correct direction. The ring width needed to produce the largest SNR improvement matches well with the expected width based on the hot-spot FWHM, as shown in Table 6.2.

If we now apply this improvement to the spectral results in Figure 6.1, we can see the improvement in the SNR for all events with the imaging information available in multiple-interaction events. This is shown in Figure 6.5. When more than one type of event is combined, the width of the ring in the image of each is still the optimal width for that type of event. Using just multiple-interaction events, the SNR is improved by 40%

Table 6.2. Summary of SNR improvements due to imaging.

Case	Image FWHM (degrees)	Expected Optimal Ring Width (degrees)	Optimal Ring Width (degrees)	Fraction of background rings through target direction	Fraction of source rings through target direction
2-interaction	30-40	18-24	25	0.30	0.79
3-interaction	24-26	14-16	15	0.17	0.52
4-interaction	20-30	12-18	10	0.09	0.40

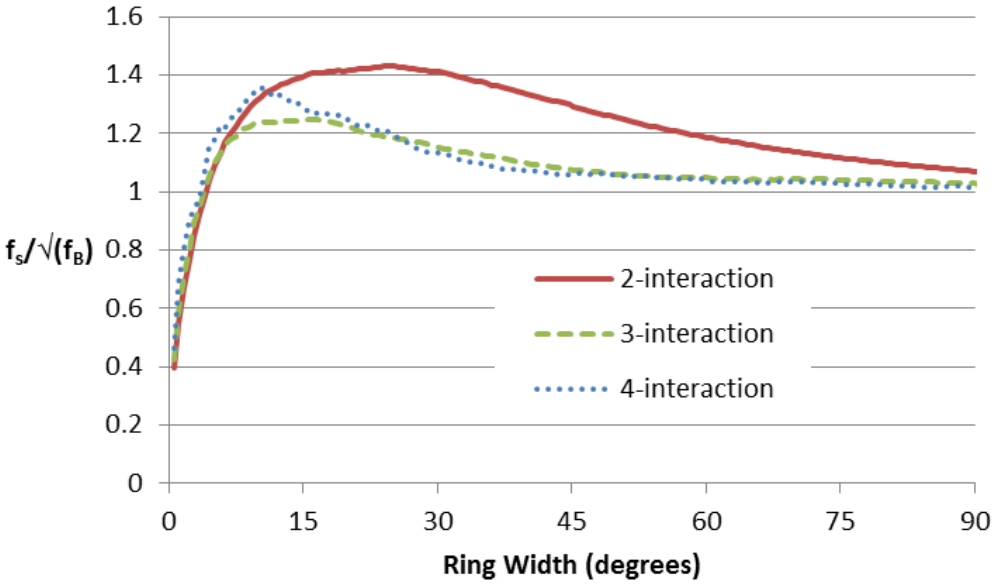

Figure 6.4. The SNR improvement due to filtering by events that only pass the known source direction as a function of ring width (or equivalently integration-area radius) for different subsets of events. Factors great than unity show improvement in the SNR using imaging. The optimal ring width agrees well with that expected based on the FWHM observed in the image.

with the addition of imaging. However, since single-interaction events constitute a majority of the events, the best SNR is achieved when combining all types of events with their own optimal energy windows, and imaging only improves the SNR by 7% in this case. The improvement may be different for sources other than Cs-137 that emit different energies and produce different fractions of each type of event.

Spectrum and Image with Unknown Target

In the previous section, imaging information was used to improve the SNR of the spectrum when a target direction and source identity were known before the measurement

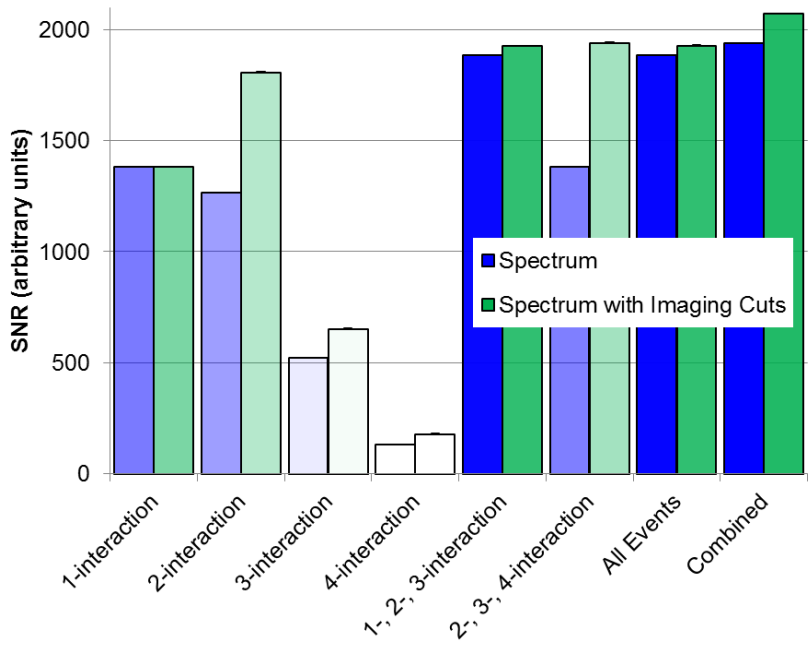

Figure 6.5. The relative experimental SNR for a Cs-137 source in background for different subsets of events. The 'Combined' category has a different spectral peak region width for each type of event and combines the counts from each into a single sum. The 'Spectrum with Imaging Cuts' series throws out events that do not pass the source direction. The transparency of each bar is proportional to the signal relative to the signal in the combined spectrum case. Each series' transparency is scaled separately.

commenced. Consider now another situation in which neither is known, as may occur when looking for a threat source in the field. For now, we will assume that the source identity is among those in a library and each possible isotope emits at least one line that is absent from the background. This is a multiple-hypothesis testing case. Detection methods in this case might involve estimating the possible direction and isotope that may be present and then applying the analysis from the last section. However, the existence of this extra estimation step hinders performance. Statistical fluctuation of the background may cause a small fraction of background-only trials to appear as if they contained a particular isotope in a particular direction. As long as the test focuses on a certain isotope and certain direction where the target may be, there is only a small probability for such false-positive reports. However, when the test looks for signals from a number of isotopes and directions, these small probabilities build, resulting in a much higher false-positive fraction – and lower specificity – at the same sensitivity level. Therefore, the sensitivity must be reduced so that the probability for any background

statistical fluctuation to look like any library isotope anywhere is kept low. In other words, high specificity requires lower sensitivity when there are multiple hypotheses. The Bonferroni Correction describes the sensitivity decrease that must accompany multiple hypotheses to keep a constant false-positive rate.

Because each event may have come from multiple directions and multiple incident energies, there is some degree of correlation between the reconstructed signal in different directions and energies. If one pixel in the image is high, the others near it tend to be high as well, and if one pixel in the image is low, the others near it tend to be low. This acts to reduce the severity of the Bonferroni Correction needed compared to the same number of independent pixels. Still, one can imagine an equivalent number of independent bins that is less than or equal to the original number of bins, which can be used to describe the distribution of the maximum in the image. Dubey, Armitage, and Palmer suggest the approximation for the number of independent regions as

$$n_{independent} = n_{original}^{1-r} \tag{6.8}$$

where r is the average correlation between bins [141]. However, in the 18-detector array system, the average correlation between image pixels tends to be independent of the number of image pixels, so this approximation does not work. Instead, to get an idea of the number of independent regions, one can plot the distribution of the image maximum for a background measurement as the number of pixels in the image is increased. This is shown in Figure 6.6. The important part of these curves is the high-side tail at which the threshold will be set to limit the false-positive rate. It is clear that the distribution of the maximum changes little above a mesh of 15 × 15, so an upper bound on the number of independent regions is 15^2=225, although a better estimate is probably significantly lower, since signs of convergence occur after a mesh of just 3 × 3.

Figure 6.7 shows what we would expect if there were no correlation between each image pixel. As expected, the distribution with one image pixel matches well, but the distribution in the 9-pixel image is significantly sharper than observed because of correlation between pixels in the actual image. Nevertheless, the high-side tail is similar. The case without correlation in Figure 6.7 continues to increase past the 7 × 7 mesh since each region is independent.

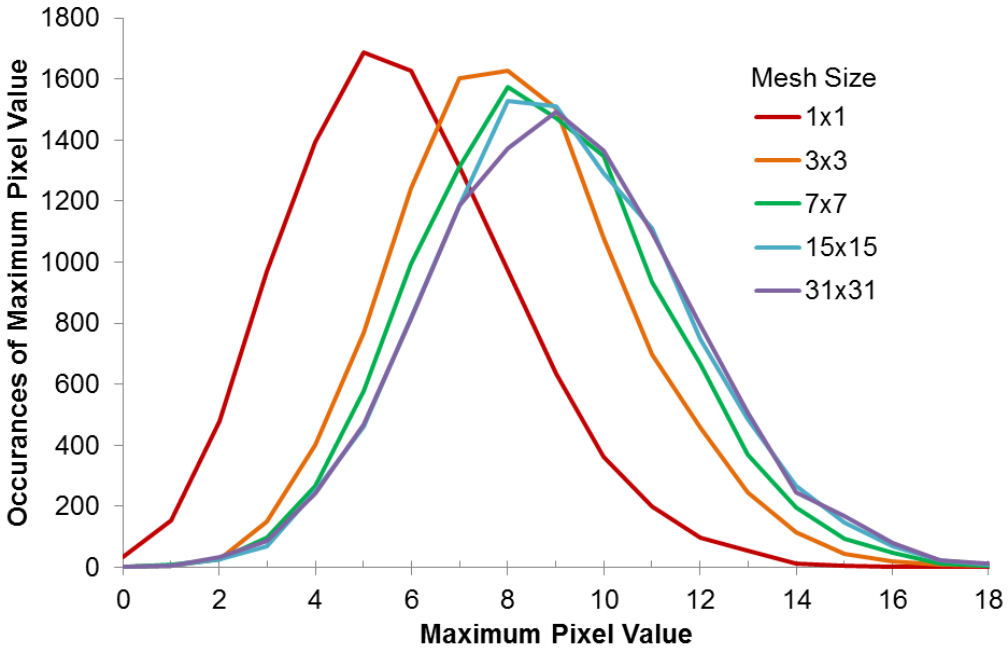

Figure 6.6. The distribution of maximum values in an image of background using two-interaction events with an average of 10 events per image.

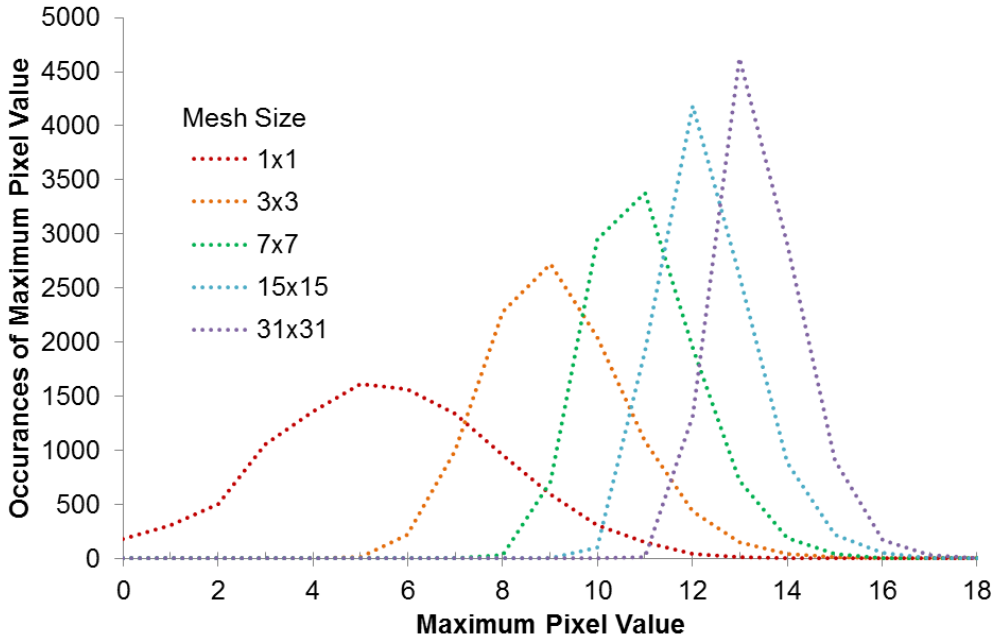

Figure 6.7. The expected distribution from completely independent image regions. Compare to Figure 6.6.

With multiple independent regions of the image, the false-positive fraction in each pixel must be reduced to

$$f_{pixel} = 1 - (1 - f_{total})^{\frac{1}{n_{independent}}} \tag{6.9}$$

where f_{total} is the required false-positive rate [141]. For low false-positive rates, this is approximately

$$f_{pixel} \approx \frac{f_{total}}{n_{independent}} \tag{6.10}.$$

The Bonferroni Correction when the target direction and isotope are unknown significantly impacts detection performance. Although the directional spectrum may have 7% higher SNR, the requirement to increase the threshold in each pixel may trounce this advantage. The degree of the trouncing, of course, depends on the number of equivalent independent regions in the image, but also on the required false-alarm rate. Because of the tail behavior of the maximum-in-the-background distribution, each fractional decrease in false-positive probability requires a corresponding smaller increase in SNR to achieve similar detection probability. Therefore, at some requirement of low false-positive probability, the 7% SNR enhancement is enough to counter the large threshold increase required on each pixel no matter the number of independent regions. However, when a large false-positive probability is specified, the fractional increase in SNR for imaging is not enough to overcome the Bonferroni Correction.

This is shown in Figure 6.8. The SNR in the original all-event spectrum required to achieve 50% detection is plotted as a function of the false-positive probability specified by the application. The 'Without Imaging' series is what is expected with standard Gaussian-distributed counting data. The 'With Imaging' series takes into account both the Bonferroni Correction and the 7% SNR improvement above the spectral SNR achievable when including imaging. At high false-positive fractions, the 7% SNR improvement due to imaging is not enough to offset the higher threshold required in imaging. At low false-positive fractions, the imaging SNR improvement is enough to offset the Bonferroni Correction. A lower raw SNR in the original spectrum can still produce a certain detection fraction when imaging is employed. For a set fractional improvement in SNR between an image pixel and the overall spectrum, this improvement is largest when there are fewer independent regions in the image. However, we expect a larger fractional SNR improvement in images with more independent regions, leading to the best performance when there are many independent regions in the image. Figure 6.8 shows the expected improvements when there are between 2 and 49 independent image

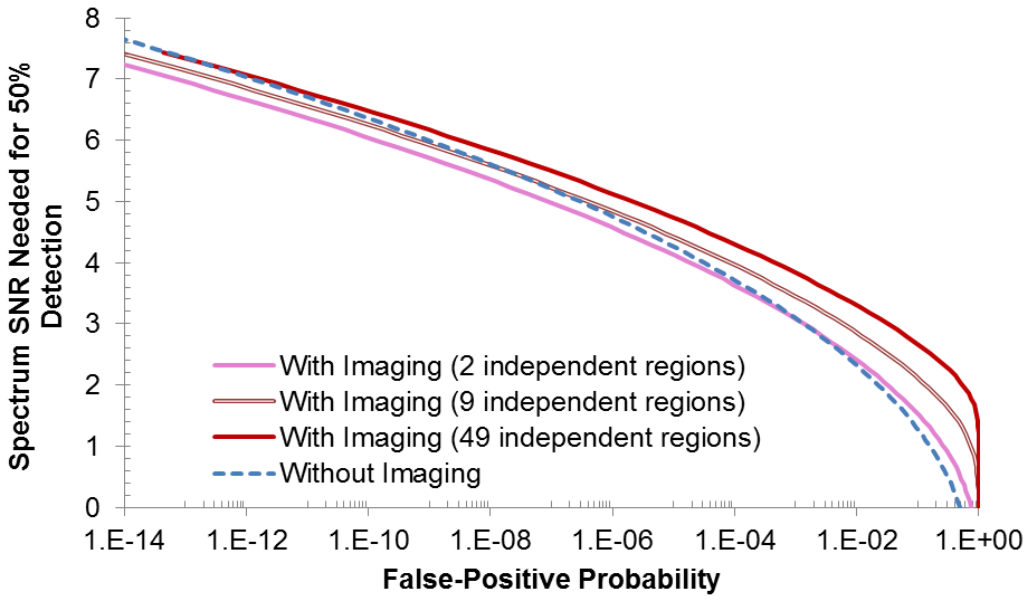

Figure 6.8. The SNR in the spectrum required for 50% detection of Cs-137 as a function of required false-positive fraction. This assumes a normal distribution of values in independent image regions with a 7% improvement in imaging SNR over the spectrum alone. At low false-positive rates, imaging can perform slightly better than simply using the spectrum alone.

regions. As described above, the actual number of regions lies somewhere between 9 and 225.

The improvement in detection performance of Cs-137 using imaging only seems to occur in Figure 6.8 at very low false-positive probabilities, and even then, the improvement is minimal. This is not to say that an imaging system is useless, however. In the worst case, an imaging system still provides all the same information as a spectroscopic or gross-counting detector, so the spectral axis in Figure 6.3 can be used instead and the pure detection performance will never be worse than similar-designed spectroscopic or gross-counting systems (with appropriate algorithms). The imaging system also provides an image. Although this may not improve detection in all cases, it can help guide the user to the source and characterize the source distribution. It can answer questions like "How many sources are present?" and "Which source is in which direction?" that are more challenging without an image.

It is also important to note that when the target direction or isotope is not known exactly, narrowing the possibilities can reduce the magnitude of the Bonferroni Correction. For instance, outside, it might be a reasonable assumption to assume that no

sources will be above you. Looking in fewer directions reduces the number of independent pixels that must be tested.

In addition, imaging is important when the other signals (spectrum and gross counts) cannot be used for detection. For instance, if one cannot assume that each possible target isotope will emit at least one line that is absent from the background and the background spectrum and intensity are unknown, the recorded spectrum provides no clues to the presence of a source. It is impossible to distinguish the emissions from the background and source. In this case, a hot spot in the image is the only signature of a source, so imaging is required for any detection. The SNR in this case is the SNR between the events passing the source direction and those in all other directions that describe the background continuum.

Chapter 7
Detection Methods

Grant me to know and understand, Lord, which comes first: ... to know you or to call upon you? Must we know you before we can call upon you? Anyone who invokes what is still unknown may be making a mistake. Or should you be invoked first, so that we may then come to know you?
– St. Augustine, The Confessions, Book I

The previous chapter described algorithms for detection with gross-counting or spectroscopic detector systems. There exists a need for detection algorithms that include imaging information, especially for situations in which the threat source may not have emission lines distinct from the background. In this chapter, I will apply two statistical algorithms, the generalized likelihood ratio test (GLRT) and the source-intensity test (SIT), to detect point threat sources [142]. These algorithms can be applied to a range of situations, but I will focus on a case for which only imaging detectors can be used – when the source possibly shares emission lines with the background and the background energy spectrum is unknown. The GLRT and SIT use imaging information to distinguish between a smooth background and a background with an additional point source while making minimal other assumptions about the situation. The GLRT and SIT algorithms are applicable to any radiation-imaging system, but will be tested using an 18 detector, 3D position sensitive, CdZnTe, Compton imaging system. They will make use of the list-mode system matrix previously derived for deconvolving the incident energy spectrum in each direction around the detector as described starting on page 23.

Imaging detectors have an inherent ability to locate point sources in a single measurement, which non-imaging detectors do not. Therefore, all other things being equal, one would expect imaging detectors to find the direction to a source more quickly than non-imaging detectors and hence perform better in source-search applications. The

algorithms I will present, on the other hand, focus on the detection performance: the ability to discriminate between an environment of just background and an environment containing a source while operating at a fixed location. The sources and detector are therefore assumed to be stationary. Future work could extend the algorithms to detectors and sources with known or unknown motion trajectories.

I will compare these statistical detection methods with Genie 2000 spectral analysis software. While not directly comparable in terms of assumptions and information required (see Table 6.1), Genie can give a baseline for the performance of other methods under the same conditions.

First, I will introduce the comparison algorithm, Genie 2000 [101] and show some sample results. Then I will discuss the statistical detection algorithms, the GLRT and SIT, and describe the model for the background and source. Then, I will show their performance for sample cases with various sets of known and unknown parameters. I also comment on performance degradation due to model mismatch. Finally, I will show performance for detection of multiple sources simultaneously. Most cases will benchmark performance using Genie 2000 commercial identification software for spectrometers.

Genie 2000

Genie 2000 is a spectral analysis software by Canberra [101] that finds spectral peaks and identifies the nuclides that are present by comparing peak locations with a library. When Genie is presented with a spectrum, it first finds all peaks larger than a threshold value (the 'peak significance') set by the user. Then, it finds peak areas, correcting for efficiency, as provided by the user. The output is an estimated source activity associated with each peak and a confidence associated with each isotope in the library with matching peaks. Efficiency as a function of energy was calculated by finding the average photopeak size from a simulation of point sources distributed around the array. All events in the detector were used, and the spectrum had 1-keV energy bins.

Genie was given the library in Table 7.1, which consists of a number of isotopes with energies in the Compton-imageable range using multiple-interaction events for the

Table 7.1. The nuclide library used in this section.

Nuclide	Energy (keV)	Emission Fraction
Cs-137	661.7	1
Mn-54	834.8	1
Co-60	1173.2	0.4997
	1332.5	0.5003
Na-22	511.0	0.6667
	1274.5	0.3333
Fe-59	1099.2	0.5667
	1291.6	0.4333
Cu-64	511.0	1
Ga-68	511.0	0.9823
	1077.3	0.0177
Sr-85	514	1
Ru-103	497.1	0.9405
	610.3	0.0595
Ag-110m	446.8	0.0119
	620.4	0.0088
	657.8	0.3095
	677.6	0.0347
	687.0	0.0211
	706.7	0.0536
	744.3	0.0158
	763.9	0.0743
	818.0	0.0241
	884.7	0.2386
	937.5	0.1123
	1384.3	0.0817
	1475.8	0.0138

system. Isotopes with lines outside the Compton-imageable energy range were not included since these would not be possible to detect using the statistical imaging methods described later. The isotopes near the end of the list are included to expand the library and give a more realistic impression of false-alarm rate. For instance, the Ag-110m source suggests the false-alarm rate that might occur for similar sources with many lines.

No background sources are explicitly included in the library, with the possible exception of the 511-keV line labeled Cu-64. We expect that if these sources were included, Genie would detect them with high activity and confidence. Yet, the goal of

this study is to see if Genie can detect possible threat sources, not background. Therefore, background isotopes were not included in the library. The absence of these isotopes will not affect detection of the other sources because Genie reports the confidence for each isotope independently of the other isotopes in the library. That is, if a peak at some energy is detected and one source in the library emits just that line, the addition of another isotope with that same line will not decrease the confidence or activity estimates of the first. Rather, both will be estimated with the same activity and confidence as if they were the only isotope in the library.

I have investigated using a test statistic for detection with Genie that is either the largest confidence or the largest activity reported for each of the isotopes. I determine the isotope activity by finding the smallest reported activity of the lines detected that were associated with the isotope. In every trial, the nuclide with the largest estimated confidence or source intensity was chosen as the identification for that trial. The optimal test statistic depends on the application and the choice of peak significance [143].

For the types of sources, library isotopes, and activity ranges in this study the optimal test statistic seems to be the source activity, using a peak significance of 3.5 for the 18-detector array and a peak significance of 0.1 for a single detector. The first case is shown for a trial Cs-137 source measured with an 18-detector array in Figure 7.1. Similar results were found using a Na-22 source. Measurements were taken as explained below (on page 115) to obtain 5-minute trials of background or background with Cs-137. All recorded events are used for Genie, although only multiple-interaction events will be used in the statistical methods. The plots show the probability to alarm and report each of the sources in the library when the Cs-137 is present versus the probability to false alarm as each isotope in the library. In the array system, the highest correct alarm rates occur when using peak significance of 0.1, but these are accompanied by high false-alarm rates, giving worse ROC performance. The performance using confidence or activity is similar, but activity generally performs slightly better in terms of correct detections for low false-alarm rates. Genie will be used as a benchmark for statistical detection methods.

It is important to note that Genie assumes that the emission lines it is trying to detect from the threat sources are not present in the background. If they were, the

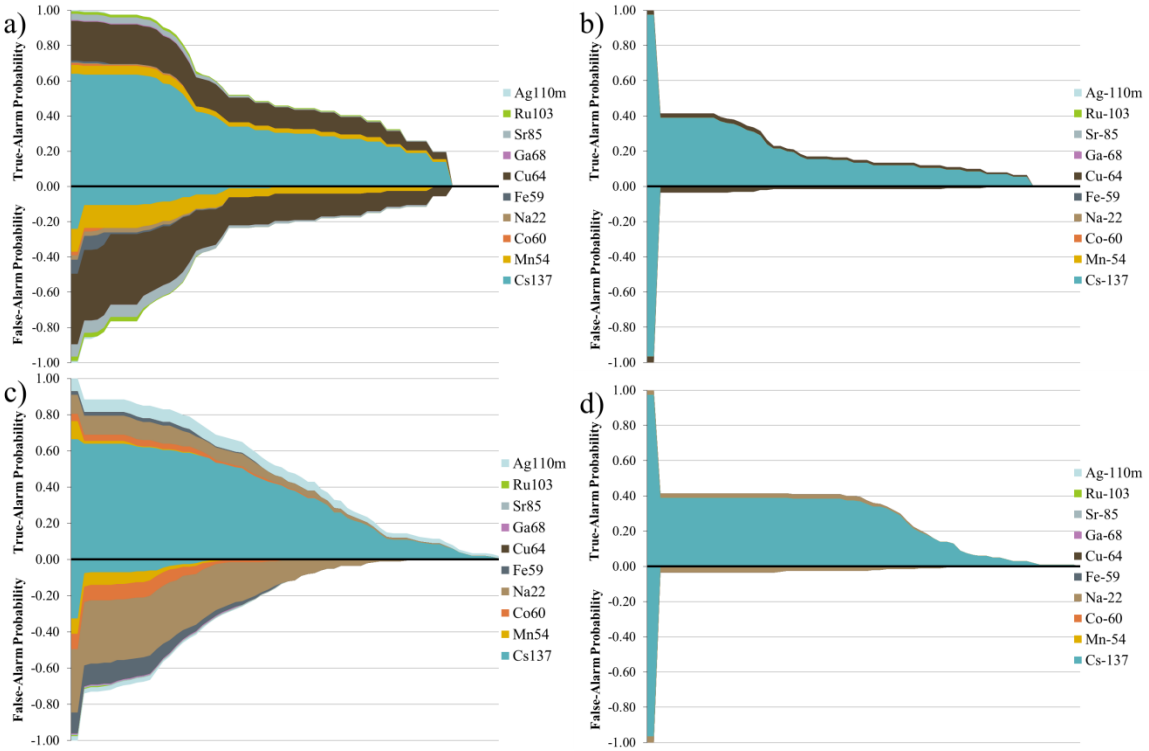

Figure 7.1. The detection and identification probability as a function of threshold using four different settings on Genie 2000. a) Confidence with 0.1 peak significance, b) Confidence with 3.5 peak significance, c) Activity with 0.1 peak significance, d) Activity with 3.5 peak significance. Trials with no source present gives the false-alarm probabilities and trials with a Cs-137 source present give the true-alarm probabilities. An average of 4.5×10^4 background events were used per trial, and the source contributed an average of 2000 counts.

presence of a peak would not be sufficient evidence for detection of the source. Therefore, comparison with the GLRT and SIT is only to benchmark performance, and is not mean to imply that the algorithms are applicable to the same situations.

Statistical Detection Methods

Generalized Likelihood Ratio Test

The generalized likelihood ratio test (GLRT) test statistic, Λ, is the ratio of the probability to see the observed events under the hypothesis that a source is present and the probability to see the observed events under the hypothesis that only background is present. To calculate each probability, the most-likely values of unknown parameters are first estimated from the data. Detection of a threat occurs if the test statistic falls above a detection threshold γ [144]. The probability in the denominator uses the maximum-

likelihood background intensities found independently when the source intensity is set to zero.

Mathematically, the algorithm reports detection when

$$\Lambda = \frac{\max_{\alpha,\phi,H,\lambda} p(r \cap D_I; \alpha, \phi, H, \lambda)}{\max_{\lambda} p(r \cap D_I; \alpha = 0, \lambda)} > \gamma \quad (7.1),$$

where $p(\cdot)$ denotes probability, r is the data vector, D_I is the event of detection of these I recorded events, α is the source intensity, ϕ is the source direction, H is the source isotope, and λ is the vector of background intensity in each energy bin. The maximization of the numerator and denominator are performed separately, as described below on page 111.

Source-Intensity Test

The source-intensity test (SIT) also takes the most-likely parameters, but uses them directly to form the test statistic

$$\Omega = \frac{\hat{\alpha}\, s_m}{\sqrt{J}} > \gamma' \quad (7.2)$$

where s_m is the sensitivity to record photons from the estimated source m, $\hat{\alpha}$ is the estimated source intensity, and J is the total number of recorded events, which reports detection when the test statistic Ω falls above the threshold γ'. A similar test was used by Bethel and Bell [145]. This function was developed because it was observed that using the source-intensity estimate alone gave good separation between the source-present and background-only cases when the background count rate was held constant, but as more background counts were added, the source-intensity estimate increased proportionally to the square root of the number of counts. Therefore, by normalizing by the square root of the total number of counts, this test was observed to have a constant false-alarm rate for all background count rates and measurement times. The efficiency factor converts the source-intensity estimate into observed counts for better performance with multiple-source libraries.

Model

I will consider applications where the background energy spectrum and intensity may both be known, unknown only in intensity, unknown only in energy distribution, or completely unknown in both intensity and distribution. I will neglect any possible prior information about common background spectral shapes. One of the consequences of assuming a completely unknown background energy spectrum is that a point source with emission energies also significantly present in the background appears no differently to the algorithm than a source with independent emission lines. Therefore, the algorithm can handle sources emitting lines also emitted by the background. I will use a library to inform the algorithms of possible point-source emission energies. Without an explicit library, a list of energy bins could be formed to represent the possible emission energies of a source. Identification would then identify the energies rather than the isotope of the source. Nevertheless, if available, a known set of emission energies can give the best detection performance.

Spatially, the model will assume a uniform intensity background. Since this is often not true, the effect of the spatial non-uniformities in experimental background will be presented in the analysis section on page 127. Finally, for simplicity, the algorithms will assume unshielded point sources so the true branching ratios reach the detector. If this is not the case, the model mismatch will cause performance degradation. To include the effect of shielding, the library could be extended to include each line separately, or one could add a shield term to the source model.

The first step in the GLRT and SIT is the estimation of the maximum-likelihood (ML) estimates of the unknown parameters in the model. The model used here is a spatially uniform background with arbitrarily varying intensity in the energy dimension. This appears as an independent background intensity λ_k in each of K=200 energy bins k from 0 to 1.5 MeV. There are M radiation point sources present, indexed by m, each with a source identity H_m (from the nuclide library), source direction ϕ_m (in 4π space), and source intensity α_m. Each library source has an associated intensity as a function of energy of

$$\Gamma_{m,k} \equiv \Gamma_{H_m,k} \tag{7.3}$$

In most cases, this is the branching ratio distribution of the isotope.

Mathematically, the maximization that must be performed when no prior knowledge is assumed is

$$\hat{\theta} = \left(\hat{\alpha}, \hat{\phi}, \hat{H}, \hat{\lambda}\right) = \text{argmax}_{\alpha,\phi,H,\lambda} p(r \cap D_I; \alpha, \phi, H, \lambda) \tag{7.4}$$

where we define θ as the vector of unknown parameters to be estimated, $p(r \cap D_I; \theta)$ is the likelihood of θ based on the data vector of discrete *recorded* attributes of photon-interaction events r such that all I have been detected, and α, ϕ, H, and λ are vectors of parameters for each source or energy bin.

The likelihood of the parameter vector θ is calculated as the product of the likelihood of the parameter vector given the attributes for each separate event and the likelihood of the parameter vector given the number of *recorded* events I [66] as

$$p(r \cap D_I; \theta) = p(D_I; \theta) \prod_{i=1}^{I} p(r_i | D_I; \theta) = {e^{-\bar{I}}\, \bar{I}^I}/{I!} \prod_{i=1}^{I} p(r_i | D_I; \theta) \tag{7.5}$$

where r_i is the recorded attributes of photon event i. For instance, r_i contains the detector pixel, recorded depth, and recorded energy for each interaction in the event. Eqn. (7.5) used the fact that the recorded attributes for each event are independent of the other events and total number of events (assuming a low enough count rate that detector dead time, space-charge buildup, and coincident events are small contributions). Each event only requires that at least that many total events have been recorded, so

$$p(r_i | r_{i+1}, \dots, r_I, D_I; \theta) = p(r_i | D_I; \theta) = p(r_i | D; \theta) \quad i \leq I \tag{7.6},$$

where D is the detection of this event. Here the total number of events I is modeled as a Poisson distribution with mean

$$\bar{I} = \sum_{m=1}^{M} \alpha_m s_m + \sum_{k=1}^{K} \lambda_k r_k \tag{7.7}$$

where

$$s_m \equiv \sum_{k=1}^{K} \Gamma_{m,k} s_{k,\phi_m} \tag{7.8}$$

$$r_k \equiv \frac{1}{4\pi} \sum_{j=1}^{J} s_{k,j} \Omega_j \tag{7.9}.$$

With slight abuse of notation, s_{k,ϕ_m} is the system efficiency for detection of a photon emitted from energy bin k and the direction of source m, while $s_{k,j}$ is the system efficiency for detecting a photon emitted from energy bin k and the direction of pixel j. The solid angle subtended by pixel j is represented by Ω_j. The sensitivity terms, s_m and r_k, hence represent the detection efficiency for photons from the source and the spatial average detection efficiency at a particular energy, respectively.

The distribution of the recorded attributes for a particular event is a mixture of the distribution due to photons from the source and the distribution due to photons from the background, with mixture coefficients determined by the relative source and background intensity parameters. Mathematically,

$$\begin{aligned} p(r_i|I;\theta) &= p(r_i|D;\theta) \\ &= \sum_{m=1}^{M} p(S_m|D;\theta)p(r_i|S_m;\theta) \\ &\quad + \sum_{k=1}^{K} p(B_k|D;\theta)p(r_i|B_k;\theta) \\ &= \sum_{m=1}^{M} \frac{\alpha_m s_m}{\bar{I}} p(r_i|S_m;\theta) + \sum_{k=1}^{K} \frac{\lambda_k f_k}{\bar{I}} p(r_i|B_k;\theta) \end{aligned} \tag{7.10}$$

where S_m represents a photon originating from source m and being detected, and B_k represents a photon originating from energy bin k of the background and being detected. The probabilities of the occurrence of event i when a detected photon originated from either a source or a background energy bin can be calculated by noting

$$p(r_i|S_m;\theta) = \sum_{j=1}^{J} p(r_i|\phi_j \cap S_m;\theta)p(\phi_j|S_m;\theta) \tag{7.11}$$

and

$$p(r_i|B_k;\theta) = \sum_{j=1}^{J} p(r_i|\phi_j \cap B_k;\theta)p(\phi_j|B_k;\theta) \tag{7.12},$$

where j parameterizes the image pixels, so $p(\phi_j|X;\theta)$ is the probability for a photon to be emitted from pixel j under the conditions of X. Since the sources are point sources, the

photons are only emitted from one pixel, so the last term in the sum of Eqn. (7.11) is zero except when $\phi_j = \phi_m$. Therefore

$$\begin{aligned} p(r_i|S_m;\theta) &= p\left(r_i\middle|\phi_j = \phi_m \cap S_m;\theta\right) \\ &= \frac{\sum_{k=1}^{K}\Gamma_{m,k}p(r_i \cap D|\phi = \phi_m, E_k;\theta)}{\sum_{k=1}^{K}\Gamma_{m,k}p(D|\phi = \phi_m, E_k)} \\ &= \frac{\sum_{k=1}^{K}\Gamma_{m,k}t_{i,k,\phi_m}}{\sum_{k=1}^{K}\Gamma_{m,k}s_{k,\phi_m}} \end{aligned} \tag{7.13}$$

where E_k represents an initial photon energy from energy bin k, and t_{i,k,ϕ_m} is the system-matrix element for the detection of event i in direction ϕ_m and energy k:

$$t_{i,k,\phi_m} \equiv p(r_i|\text{photon from energy bin } k \text{ and direction } \phi_m) \tag{7.14}.$$

Since the model for the background is for a spatially uniform distribution of emission,

$$p\left(\phi_j\middle|B_k;\theta\right) = \frac{s_{k,j}\Omega_j}{4\pi\, r_k} \tag{7.15}$$

and

$$p\left(r_i\middle|\phi_j \cap B_k;\theta\right) = \frac{p\left(r_i \cap D\middle|\phi_j, E_k;\theta\right)}{p\left(D\middle|\phi_j, E_k;\theta\right)} = \frac{t_{i,k,j}}{s_{k,j}} \tag{7.16}.$$

So, combining Eqns. (7.10), (7.12), (7.15), and (7.16),

$$p(r_i|B_k;\theta) = \sum_{j=1}^{J}\frac{t_{i,k,j}}{s_{k,j}}\frac{s_{k,j}\Omega_j}{4\pi\, r_k} = \frac{1}{4\pi}\sum_{j=1}^{J}\frac{t_{i,k,j}\Omega_j}{r_k} = \frac{\sum_{j=1}^{J}t_{i,k,j}\Omega_j}{\sum_{j=1}^{J}s_{k,j}\,\Omega_j} \tag{7.17}.$$

The probability of an event recorded from either a source or the background is therefore the ratio of the system matrix elements and the system efficiency in either the direction of the source or the integral over all directions, respectively. Finally, combining Eqns. (7.4), (7.5), (7.10), (7.13), and (7.17), the parameter maximization is in the form

$$\begin{aligned}\hat{\theta} &= \left(\hat{\alpha}, \hat{\phi}, \hat{H}, \hat{\lambda}\right) = \text{argmax}_{\alpha,\phi,H,\lambda} p(r \cap D_I; \alpha, \phi, H, \lambda) \\ &= \text{argmax}_{\alpha,\phi,H,\lambda} \, e^{-\bar{I}} \, \bar{I}^I / I! \prod_{i=1}^{I} \left[\sum_{m=1}^{M} \frac{\alpha_m s_m}{\bar{I}} \frac{\sum_{k=1}^{K} t_{i,k,\phi_m} \Gamma_{m,k}}{\sum_{k=1}^{K} s_{k,\phi_m} \Gamma_{m,k}} \right. \\ &\qquad \left. + \sum_{k=1}^{K} \frac{\lambda_k r_k}{\bar{I}} \frac{\sum_{j=1}^{J} t_{i,k,j} \Omega_j}{\sum_{j=1}^{J} s_{k,j} \, \Omega_j} \right] \\ &= \text{argmax}_{\alpha,\phi,H,\lambda} \, e^{-\bar{I}} / I! \prod_{i=1}^{I} \left[\sum_{m=1}^{M} \alpha_m \sum_{k=1}^{K} \Gamma_{m,k} t_{i,k,\phi_m} \right. \\ &\qquad \left. + 4\pi \sum_{k=1}^{K} \lambda_k \sum_{j=1}^{J} t_{i,k,j} \Omega_j \right]\end{aligned} \tag{7.18}$$

Taking the natural logarithm and removing constant terms,

$$\begin{aligned}\hat{\theta} &= \text{argmax}_{\alpha,\phi,H,\lambda} L(\theta) \\ &= \text{argmax}_{\alpha,\phi,H,\lambda} \log p(r \cap D_I; \alpha, \phi, H, \lambda) \\ &= \text{argmax}_{\alpha,\phi,H,\lambda} \left[- \sum_{m=1}^{M} \alpha_m s_m - \sum_{k=1}^{K} \lambda_k r_k \right. \\ &\qquad + \sum_{i=1}^{I} \log \left(\sum_{m=1}^{M} \alpha_m \sum_{k=1}^{K} \Gamma_{m,k} t_{i,k,\phi_m} \right. \\ &\qquad \left. \left. + 4\pi \sum_{k=1}^{K} \lambda_k \sum_{j=1}^{J} t_{i,k,j} \Omega_j \right) \right]\end{aligned} \tag{7.19}$$

It is instructive to investigate the behavior of this function when no sources are assumed to be present. Then,

$$\hat{\theta} = \text{argmax}_{\alpha,\phi,H,\lambda} \left[- \sum_{k=1}^{K} \lambda_k r_k + \sum_{i=1}^{I} \log \left(4\pi \sum_{k=1}^{K} \lambda_k \sum_{j=1}^{J} t_{i,k,j} \Omega_j \right) \right] \tag{7.20}$$

This is maximized when

$$\frac{d}{d\lambda_k} L(\theta) = -r_k + \sum_{i=1}^{I} \frac{\sum_{j=1}^{J} t_{i,k,j} \Omega_j}{\sum_{q=1}^{K} \lambda_q \sum_{j=1}^{J} t_{i,q,j} \Omega_j} = 0 \tag{7.21}$$

Consider the special case where full-energy deposition is assumed in the system model. In this simplified system model for Compton imaging, a two-interaction event is assumed to be due to a Compton-scatter in one pixel followed by a photoelectric interaction in another pixel. The possibility for photons to be scattered out of the second pixel that was included in the original complete system model is ignored, causing incorrect modeling of non-full-energy events. Nevertheless, the same set of events is used in either model. In the simplified system model, each event has non-zero probability only for possible photons coming from one of the energy bins, so most elements in the system matrix are zero. Therefore, the denominator for each event in Eqn. (7.21) only depends on the intensity estimate at the energy of the event. So,

$$r_k = \sum_{i=1}^{I} \frac{\sum_{j=1}^{J} t_{i,k,j}\Omega_j}{\lambda_k \sum_{j=1}^{J} t_{i,k,j}\Omega_j} = \sum_{i=1}^{I} \frac{1}{\lambda_k} = \frac{I_k}{\lambda_k} \qquad (7.22).$$

Hence, in this special case, the best estimate of the background contribution at each energy is the number of recorded events at that energy. The same argument can be used when a source is present but has energies different from a particular background energy. Therefore, in the full-energy assumption case, the best estimate of the background energy distribution with no source present is the recorded spectrum. When a source is present, all background energy bins other than those overlapping with the source emission lines can be trivially estimated as the recorded counts at that energy. Furthermore, terms involving these independent background components will cancel in any likelihood ratio, allowing faster computation. This greatly reduces the computational scale of the problem and decreases computation time by almost two orders of magnitude for the same number of events.

In the particular 18-detector array system used for testing, only multiple-interaction events are used. This is because single-interaction events, as described on page 60, lack a sufficiently accurate model to contribute meaningfully to the image. The energy spectrum of these events is not useful if the background spectrum is unknown and the source could share emission lines with the background, which is the situation in which we are most interested. Including the spectral information from single-interaction events would be useful if the background spectrum were known or were expected to have

```
For each possible source m
{
   For each possible source identity (isotope)
   {
      For each possible source location
      {
         Estimate intensity for background and all sources so far
         Calculate likelihood
      }
   }
   Find and store most likely identity and location for source m
   Set identity and location for this source as constant for future
}
Choose number of sources present
```

Figure 7.2. The algorithm for estimating the most-likely parameters in Eqn. (7.19).

distinct emission lines from all threat sources. In addition, further constraints could be added to the spectral estimate if the background were known to have emission lines distinct from all threat sources.

Parameter Estimation

Next, we must find a way to maximize this likelihood in Eqn. (7.19). The method we use depends, to some extent, on the types of known information. For simplicity, when necessary, we will estimate source identity and position with an exhaustive search. Since computation increases quickly with the number of exhaustive search parameters, the parameters for multiple sources are found using sequential exhaustive searches. For each possible source position and identity, we will estimate the background and source intensities with either coordinate descent or maximum likelihood expectation maximization (MLEM). Algorithmically, this is shown in Figure 7.2.

In coordinate descent, a step of Newton's method is applied sequentially for each unknown parameter while holding the others fixed, and this cycle is repeated until the estimates converge. Source and background intensities are forced to be non-negative at each step. With initial guesses from the raw background shape, convergence usually occurred in fewer than tens of iterations. The steps of coordinate descent are

$$\widehat{\alpha_m}^{(n+1)} = \widehat{\alpha_m}^{(n)} - \frac{\frac{d}{d\alpha_m} L(\theta)}{\frac{d^2}{d{\alpha_m}^2} L(\theta)} \Big|_{\alpha_m = \widehat{\alpha_m}^{(n)}}, \qquad m = 1, \ldots, M \tag{7.23}$$

and

$$\hat{\lambda}_k^{(n+1)} = \hat{\lambda}_k^{(n)} - \frac{\frac{d}{d\lambda_k} L(\theta)}{\frac{d^2}{d\lambda_k^2} L(\theta)} \Big|_{\lambda_k = \hat{\lambda}_k^{(n)}}, \qquad k = 1, \ldots, K \tag{7.24},$$

where

$$\frac{d}{d\alpha_m} L(\theta) = -s_m + \sum_{i=1}^{I} \frac{\sum_{k=1}^{K} \Gamma_{m,k} t_{i,k,\phi_m}}{\sum_{q=1}^{M} \alpha_q \sum_{k=1}^{K} \Gamma_{q,k} t_{i,k,\phi_q} + 4\pi \sum_{k=1}^{K} \lambda_k \sum_{j=1}^{J} t_{i,k,j} \Omega_j} \tag{7.25}$$

$$\frac{d^2}{d\alpha_m^2} L(\theta) = -\sum_{i=1}^{I} \left(\frac{\sum_{k=1}^{K} \Gamma_{m,k} t_{i,k,\phi_m}}{\sum_{q=1}^{M} \alpha_q \sum_{k=1}^{K} \Gamma_{q,k} t_{i,k,\phi_q} + 4\pi \sum_{k=1}^{K} \lambda_k \sum_{j=1}^{J} t_{i,k,j} \Omega_j} \right)^2 \tag{7.26}$$

$$\frac{d}{d\lambda_k} L(\theta) = -r_k + \sum_{i=1}^{I} \frac{4\pi \sum_{j=1}^{J} t_{i,k,j} \Omega_j}{\sum_{m=1}^{M} \alpha_m \sum_{k=1}^{K} \Gamma_{m,k} t_{i,k,\phi_m} + 4\pi \sum_{q=1}^{K} \lambda_q \sum_{j=1}^{J} t_{i,q,j} \Omega_j} \tag{7.27}$$

$$\frac{d^2}{d\lambda_k^2} L(\theta) = -\sum_{i=1}^{I} \left(\frac{4\pi \sum_{j=1}^{J} t_{i,k,j} \Omega_j}{\sum_{m=1}^{M} \alpha_m \sum_{k=1}^{K} \Gamma_{m,k} t_{i,k,\phi_m} + 4\pi \sum_{q=1}^{K} \lambda_q \sum_{j=1}^{J} t_{i,q,j} \Omega_j} \right)^2 \tag{7.28}.$$

These can be significantly simplified when the full-energy assumption is used.

In the case where the background spectral shape but not intensity is known, a simple scale factor can be included in the likelihood and the updates for coordinate descent computed in a similar way. If other parameters are known exactly or within a range, the exhaustive search space can be modified and known values can be used in place of the intensity estimates.

Another method to estimate the intensity parameters is to use the maximum likelihood expectation maximization (MLEM) method [66]. We can group terms so that the negative log likelihood is in the form

$$-L(\theta) = \sum_j a_j x_j - \sum_{i=1}^{J} \log\left(\sum_j a_{ij} x_j\right) \tag{7.29}$$

where

$$x_j = \begin{cases} \alpha_j, & j = 1, \ldots, M \\ \lambda_{j-M}, & j = M+1, \ldots, M+K \end{cases} \tag{7.30}$$

$$a_j = \begin{cases} s_j, & j = 1, \ldots, M \\ r_j, & j = M+1, \ldots, M+K \end{cases} \tag{7.31}$$

$$a_{ij} = \begin{cases} \sum_{k=1}^{K} \Gamma_{j,k} t_{i,k,\phi_j}, & j = 1, \ldots, M \\ 4\pi \sum_{n=1}^{J} t_{i,j,n} \Omega_n, & j = M+1, \ldots, M+K \end{cases} \tag{7.32}.$$

The intensities x_j represent the means of Poisson random variables, so the MLEM method can be applied. The MLEM iterative solution to minimize (7.29) is

$$x_j^{(n+1)} = \frac{x_j^{(n)}}{a_j} \sum_{i=1}^{I} \frac{a_{ij}}{\sum_{k=1}^{M+K} a_{ik} x_k^{(n)}}, \quad j = 1, \ldots, M+K \tag{7.33},$$

where *n* is the index for the iteration and one may start with any positive initial guesses. Note that similar computation is required for each step of coordinate descent and MLEM. The MLEM has the advantage that total counts are always conserved and the intensity estimates are always non-negative.

When the background is known exactly, the same equation can be used, but the background estimates should not be updated; that is, in Eqn. (7.33), $j = 1, \ldots, M$. In the case where the background spectral shape, but not intensity is known, Eqns. (7.30), (7.31), and (7.32) can be replaced with

$$x_j = \begin{cases} \alpha_j, & j = 1, \ldots, M \\ \lambda_{bkgd}, & j = M+1 \end{cases} \tag{7.34}$$

$$a_j = \begin{cases} s_j, & j = 1, \ldots, M \\ \sum_{k=1}^{K} \lambda_k r_k, & j = M+1 \end{cases} \tag{7.35}$$

$$a_{ij} = \begin{cases} \sum_{k=1}^{K} \Gamma_{j,k} t_{i,k,\phi_j}, & j = 1, \ldots, M \\ 4\pi \sum_{k=1}^{K} \lambda_k \sum_{j=1}^{J} t_{i,k,j} \Omega_j, & j = M+1 \end{cases} \quad (7.36),$$

where λ_{bkgd} is the intensity scale factor to the known background λ_k.

Model Order Selection

If the detection test aims to determine whether no sources or some known number of sources M are present, the standard form of the GLRT with estimated parameters for each of the sources provides the needed test statistic. But, if the number of sources is unknown, the number of sources – and hence the model order – must first be estimated. When choosing the number of sources, the maximum likelihood estimate will always choose the maximum number since this will effectively increase the model order. Therefore, the GLRT must be modified to penalize the addition of multiple sources. Kay [146] suggests the test statistic

$$\begin{aligned} T(\boldsymbol{\theta}) = \max_{1 \le m \le M} \Bigg\{ \Bigg[& 2 \ln \Lambda(\boldsymbol{\theta}) \\ & - \big((m-1)\beta + 1\big) \left(\ln \left(\frac{2 \ln \Lambda(\boldsymbol{\theta})}{\big((m-1)\beta + 1\big)} \right) + 1 \right) \Bigg] \\ & \cdot u \left(\frac{2 \ln \Lambda(\boldsymbol{\theta})}{\big((m-1)\beta + 1\big)} - 1 \right) \Bigg\} \end{aligned} \quad (7.37)$$

where $u(\cdot)$ is the step function that is 1 for inputs greater than 0. Kay has a scale factor β of unity that works when there is no model mismatch, but we introduce a larger scale factor that helps to penalize the addition of multiple sources further.

Speedup Using Guessed Direction

The exhaustive search over direction can take significant computational power, especially if the image space uses a fine mesh. Instead of searching over all directions, the algorithm can quickly guess a probable direction to use as a stand-in for a known direction. One such direction can be obtained by choosing the hottest point in the image obtained from the first iteration of MLEM at the energies of the estimated source isotope,

but with a sensitivity factor for the isotope as a whole so as not to magnify statistical noise at energies with low sensitivities. Mathematically,

$$\phi_{est,m} = \text{argmax}_j \sum_{k=1}^{K} \frac{\Gamma_{m,k}}{s_{m_j}} \sum_{i=1}^{I} \frac{t_{i,k,j}}{\sum_{k=1}^{K}\sum_{j=1}^{J} t_{i,k,j}\Omega_j} \quad (7.38),$$

where s_{m_j} is the efficiency of the detector to photons from isotope m in direction j.

Experiment, Results, and Analysis

Experiment

In order to investigate performance of these algorithms, experimental data were collected of background events and events due to Cs-137, Co-60, and Na-22 samples. Two different systems were used: a single 2.0 cm × 2.0 cm × 1.5 cm CZT detector, and an array of 18 of those detectors. The case of unknown background with a single source was investigated with a single detector while cases of known background with a single source and multiple sources were investigated with the 18-detector array.

To produce data consistent with a challenging detection scenario, events from background and source measurements were combined with various frequencies. For the single detector, we took a measurement of just background within the laboratory over 8.6 days and another of a 6.58-μCi Cs-137 sample at 23 cm from the detector. Similar measurements were performed for Co-60 and Na-22. The stored events from both measurements were combined with a frequency to produce, on average, a background of 5000 two-interaction events and 79 two-interaction source events per trial, to give intermediate detection performance. Approximately 50 trials of the background-only case and the source-present case were performed. Both background and source events were accepted from 0 to 1.5 MeV, so Compton-continuum source events (non-full-energy-deposition events) were counted. This number of background and source counts corresponds to about a 4-hour measurement with a 2-mCi Cs-137 source at 100 m. The number of background and source counts were chosen from a Poisson distribution, but had no additional background variation as one might see due to background variation in the environment. Since we are only using the well-studied two-interaction events, we are likely understating performance compared to using all multiple-interaction events over

the same measurement time. Single-interaction events will not improve performance unless the background spectrum is known. Genie, on the other hand, uses all events in the detector, including single-interaction events, since it assumes that the presence of source peaks proves the presence of the isotope.

For the 18-detector array, 4050 to 8100 multiple-interaction background counts were combined with an average of 235 multiple-interaction source counts from a Cs-137 source. Although the background rate within the lab is steady, time-variable background intensity was used in the known background case to simulate background in a search scenario. The background count-rate fluctuation is similar to that observed in experiments in the field [5]. The number of counts is consistent with a 5- to 10-minute background measurement in our lab, and a 7.5-minute measurement of a 0.5-μCi Cs-137 source at 1 m, or, alternatively, a 13-mCi Cs-137 source at 100 m.

A representative trial for the 18-detector array is shown in Figure 7.3 and Figure 7.4. The spectrum in Figure 7.3 shows all the events used in one trial, including single-interaction events. The blue region contains those events contributed from the background measurement and the red region also contains those events contributed from the Cs-137 measurement. There is a small photopeak due to Cs-137 at 662 keV in the source-present case. The simple back-projection images in the photopeak region are shown in Figure 7.4 for a background trial and for a trial using the same background but with additional source events. The source is present at approximately 90º polar and 150º azimuthal, although this is not immediately obvious from these low-statistic trials. In actual measurement, there is, of course, no distinction between background and source events.

For the study of multiple-source detection, 4050 multiple-interaction background counts were combined with an average of 235 multiple-interaction source counts from Cs-137, Co-60, or Na-22. This is consistent with a 5-minute background measurement.

The library of ten sources shown in Table 7.1 on page 101 was used to inform the algorithms of possible point-source emission energies. These were selected to be sources with emission lines in the Compton-imageable range for the system. Some are likely to be seen in practice, one has many lines, and some have similar emission energies to each other.

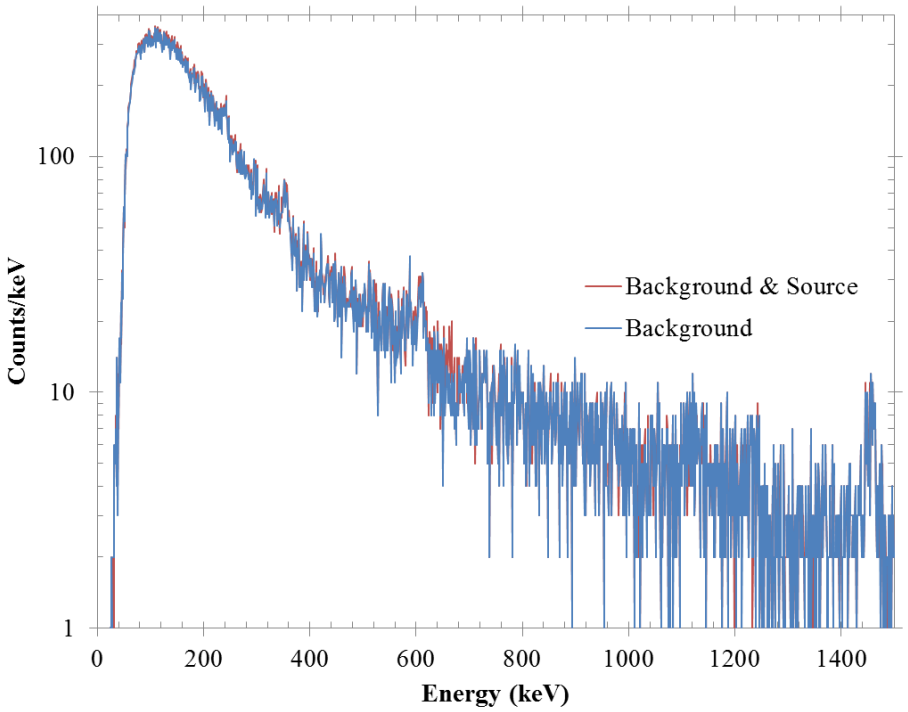

Figure 7.3. A representative spectrum from one detection trial with the 18-detector array. The contribution to the spectrum from background and source events is shown, however this distinction is not passed to the detection algorithms.

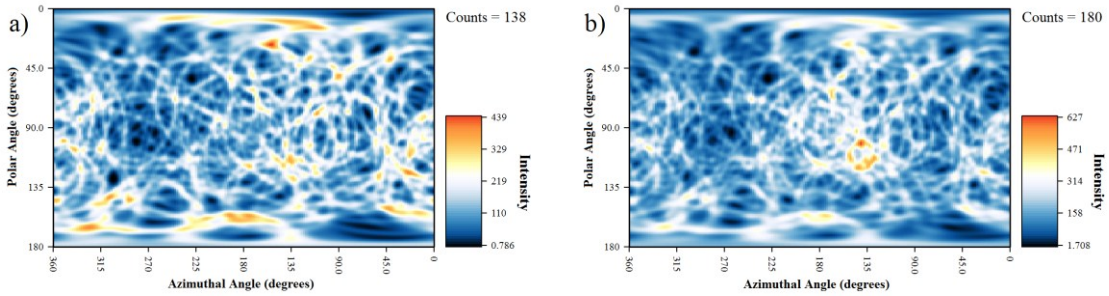

Figure 7.4. Representative images from the Cs-137 photopeak region from one experimental detection trial. Image (a) shows just events originating from the background and (b) includes the contributions from a Cs-137 source placed at 150° azimuthal and 90° polar.

Unknown Background and Unknown Source Direction

In a sample experiment, we tried to detect and identify a Cs-137 point source in standard background in our laboratory with data collected on a single detector. Using the complete system model required approximately 8 hours of computation for each trial on a

4-core PC, but using the simplified system model that assumes full-energy deposition (described on page 110) required less than 3 minutes.

For experimental data, the SIT performed slightly better using the simplified system response compared to the complete system model, so the simplified response will be used for the tests with the SIT, unless stated otherwise. In simulation, the best performance using the SIT occurred when the more realistic complete system model was used, as expected. The difference may be because of the additional model mismatch introduced by the detector performance in experiment and the need to non-trivially estimate more parameters when using the complete system model. This small effect was not observed for the GLRT, however.

The ability of the algorithms to differentiate trials of just background events from trials of background and source events are shown in Figure 7.5a in the form of an ROC curve, which plots the probability of detection as a function of the probability of false alarm. I define detection probability as the probability of detecting any source when one is truly present, and the false-alarm probability as the probability of detecting any source when one is not present. For the best performance, the GLRT used the complete system model, while the SIT used the simplified system model. The curve also shows benchmark performance using Genie 2000 using its optimal settings. For all three methods, the solid curves show the performance with known source identity and the dashed curves show the performance using the library in Table 7.1. The source location is unknown.

Both the GLRT and SIT perform as well or better than using the maximum reported source activity from Genie 2000, even though the GLRT and SIT do not assume that the emission lines from the sources are distinct from background. This may be because imaging information is being used in these statistical algorithms to discriminate point sources from extended background sources, whereas Genie uses neither statistical algorithms nor imaging information. The SIT is able to estimate source intensity better than Genie's peak-fitting algorithm, at least to the extent that it can discriminate between zero counts from a source and a positive number of counts from a source. It is not immediately clear whether the GLRT or SIT is better in all cases. The SIT seems to perform best with a small library size.

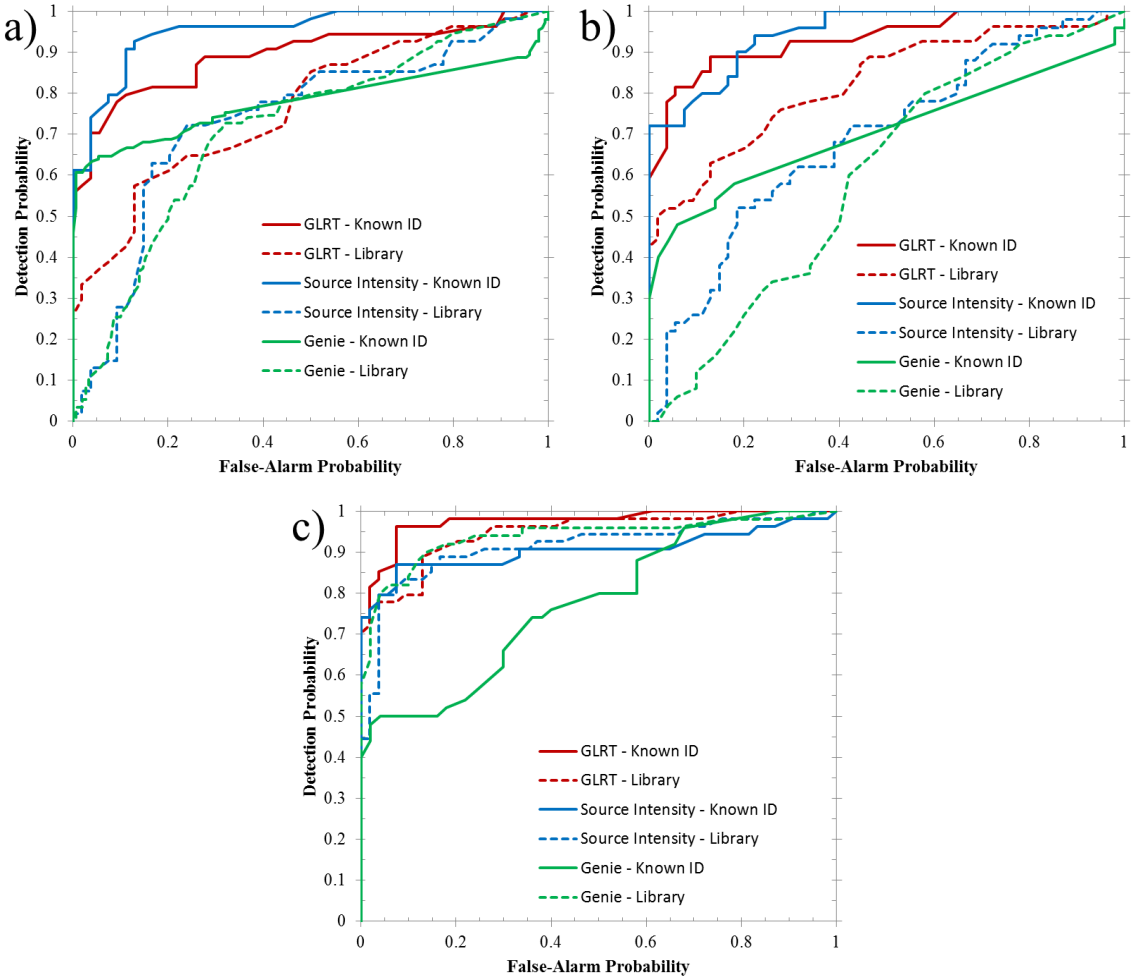

Figure 7.5. ROC curves of the experimental detection performance for a background of 5000 events, Poisson distributed, and (a) a Cs-137 source with an average of 79 events, (b) a Co-60 source with an average of 158 events, or (c) a Na-22 source with an average of 158 events, Poisson distributed. Approximately 50 trials of the background-only and each source-present case were performed using a single detector module. The performance of the three methods are shown in the case of known source identity (solid lines) and using the source library in Table 7.1 (dashed lines). In all cases, the source direction is unknown.

We performed similar experiments with a Co-60 source and a Na-22 source. Poor detection was observed for all methods with an average of 5000 two-interaction background and 79 two-interaction source events, probably because of the lower full-energy efficiency and multiple emission lines of these sources. Figure 7.5b and c show the performance with twice the source strength, 158 two-interaction source events, with the same methods used in Figure 7.5a for Cs-137 and 50 trials. Again, Compton-

continuum events were included. Relative performance between the three detection methods is similar to the Cs-137 case.

Interestingly, Genie has better detection performance for Na-22 when a larger library is used. This is likely because of incorrect source-isotope estimates causing inflated source-intensity estimates. When the Na-22 source is present, the 511-keV peak from background and Na-22 is obvious whereas the 1274.5-keV peak from Na-22 is very small due to low detector efficiency at that energy. Therefore, Genie may estimate a Na-22 source intensity near zero. However, the estimated source intensity for Cu-64 or Ga-68 may be relatively large, since the primary line for these isotopes, at 511 keV, is strong. In the source-present case then, we have observed that identification skews towards Cu-64 and Ga-68 and the source-intensity estimate is larger than if Na-22 were the only isotope in the library. When only background is present, we observed that the intensity estimates for Cu-64 and Ga-68 rarely were larger than those for Na-22. Therefore, when the only library source is Na-22, the difference in estimated source intensity between the background and source-present case is small, but when the full library is used, the difference is larger, due to the larger values for Cu-64 and Ga-68. This results in improved detection performance, but poor identification performance.

Figure 7.6 shows how often the correct isotope is estimated for the point source – the identification performance – for the described algorithms and for Genie. Identification in the described algorithms occurs at the parameter-estimation stage, and is therefore the same for the GLRT and SIT. However, the estimated parameters are different when using the simplified system response rather than the complete system response. As stated before, the complete system response gave the best detection performance in the GLRT and the simplified model gave the best detection performance in the SIT for experimental data, so these are used for detection. However, the best detection performance does not imply the best identification performance. Figure 7.6a shows the performance using the complete system response matrix and Figure 7.6b shows the performance when full-energy deposition is assumed. Both show the fraction of trials that are estimated to most likely contain a point source with that identity, when the algorithm is forced to choose from the nuclide library. The first column for each nuclide is for trials with just background present and the next three columns are for trials

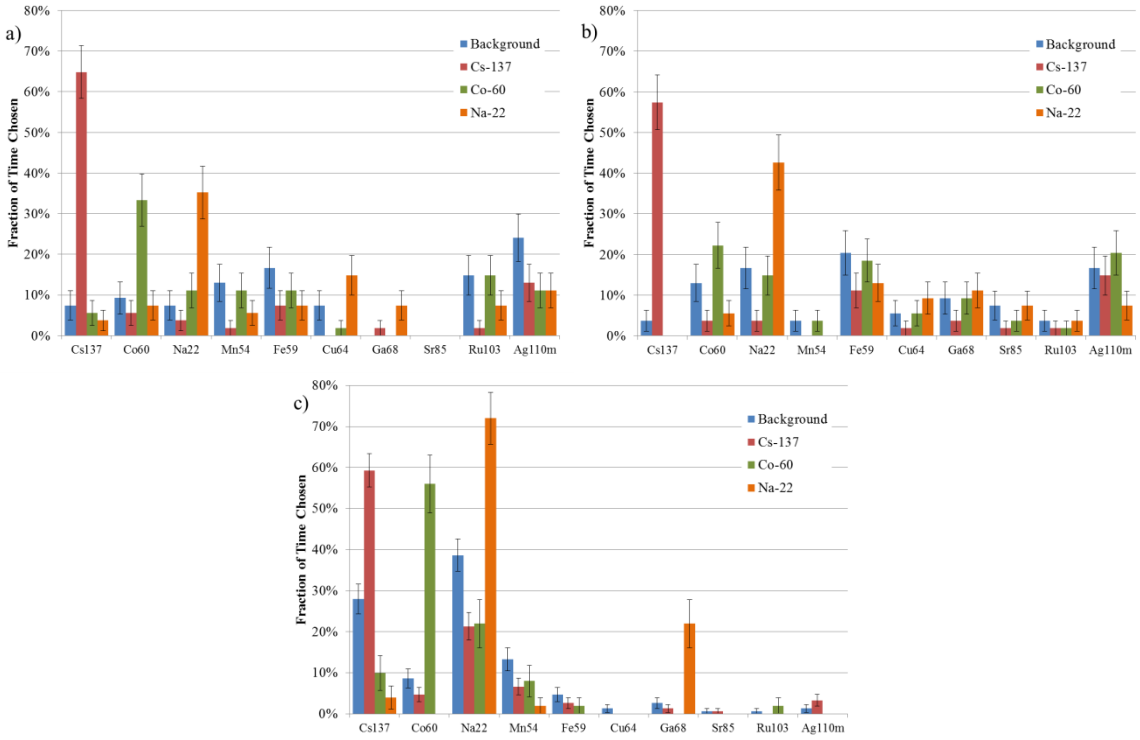

Figure 7.6. The experimental identification performance using (a) the complete system model, (b) the simplified system model, or (c) Genie 2000. The bars show the fraction of time each nuclide on the horizontal axis was reported for the background-only trials and for each set of source-present trials for each of the three isotopes. Background was an average of 5000 events and the source contribution was an average of 79 events, Poisson distributed, using a single detector module. The difference between the identification fraction for a source with only background present and the identification fraction when the true source is present gives a measure for identification performance. In all cases, the source direction is unknown.

with either a Cs-137, a Co-60, or a Na-22 source present with the background. In these figures, each source contributed an average of 79 two-interaction events and the background contributed an average of 5000 two-interaction events.

It is important to note that for each trial, even if it is most likely for no sources to be present, the algorithms are forced to choose the most-likely single nuclide. The maximum-likelihood estimate of the source nuclide is needed to compute the GLRT and SIT. The nuclide it chooses is shown on the identification performance plots, while the detection performance was shown in Figure 7.5.

In Figure 7.6a and b, a larger fraction of the trials are identified as a source when that source is present than when only background is present, as expected. The difference between the fraction of trials identified as a specific source when just the background is

present and when the true source is present tells how well the algorithm can correctly identify the present nuclide. In addition, a flatter distribution of nuclides for the background-only case shows that the algorithm does not preferentially identify any nuclide in the background spectrum. One can see that the identification performance is better when the complete system model is used compared to the simplified system model. This is probably because the complete system model also makes use of the Compton-continuum events to better determine which source is present. In the best case, 65% of trials were identified correctly as Cs-137 when it was present.

Figure 7.6c shows the same measure of identification performance using Genie 2000. In the Co-60 and Na-22 cases, identification is better than the statistical methods, probably because all events were used, rather than only two-interaction events. However, using only two-interaction events in Genie 2000, identification performance is much weaker than the GLRT and SIT in all cases because the full-energy peaks in the spectrum due to the sources are small. An interesting plot feature is the large fraction of trials identified as Na-22 in the background. This is due to a natural background peak at 511 keV, an emission line of Na-22 as well. So, identification of Na-22 might be expected due to background lines since this line was included in Genie's library for Na-22. In the algorithms developed here, however, Na-22 is not identified as often, especially using the complete model, because, although a peak may be present in the spectrum, it can be recognized as background because the intensity is not localized in space like a point source. This shows how imaging can be used to find the presence of a point source with emission lines also present in the background.

Unknown Background and Known Source Direction

Figure 7.7 shows the experimental detection performance when the possible source direction is known for the Cs-137 source scenario described above. Here, the source direction is known within a cone half-angle of 10°. The possible source direction may be known in applications where one is interested in the presence of a source within a container, for instance.

The performance in this case is compared to the performance when both the possible source direction and identity are unknown. The difference between these shows

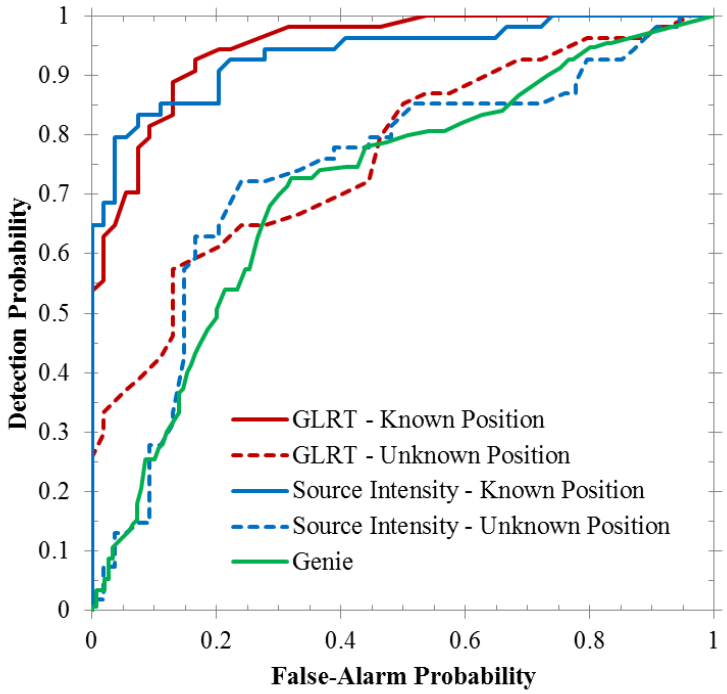

Figure 7.7. An ROC curve of the experimental detection performance for an average background of 5000 events and an average Cs-137 source of 79 events, Poisson distributed, for the cases of known (solid lines) and unknown source direction (dashed lines) using the source library in Table 7.1. Approximately 50 trials of background-only and source-present cases were performed using a single detector module. The performance of each of the algorithms described is shown. Genie 2000 does not benefit from known source position.

the large detection improvement available to an imaging system when it can focus on a specific direction or object.

The performance improves because at a constant test-statistic threshold, the rate of false-positive detections decreases by approximately the fraction of the solid angle subtended by the possible source directions to the solid angle in all directions, as described in Eqn. (6.9) and Eqn. (6.10). This allows a lower detection threshold while keeping the same false-alarm rate and, hence, the ability to detect less obvious sources.

Figure 7.8 shows the identification performance for the same experiment, when the simplified system model is used. The performance using the complete system model is similar. Identification performance is better when the possible source location is known compared to when it is unknown, reaching 76% and 81% correct identification respectively when using the complete system model and when using the simplified system model. Interestingly, the simplified system model performs better in this case,

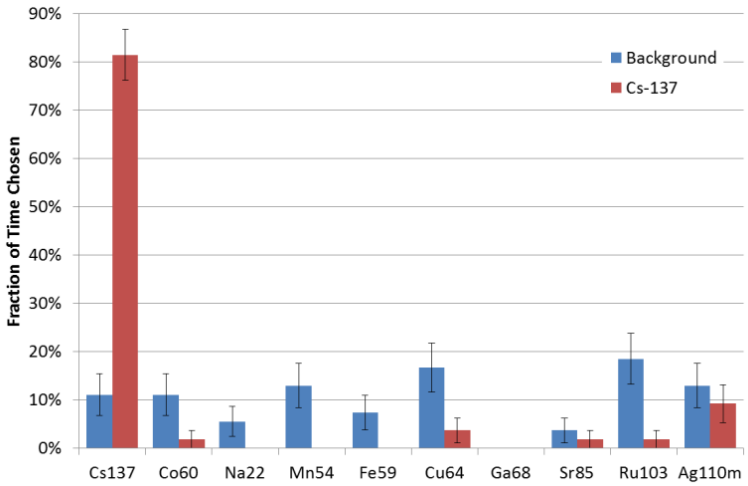

Figure 7.8. The identification performance using the simplified system model when the possible source direction is known. The bars show the fraction of time each nuclide on the horizontal axis was reported for the background-only trials and for the Cs-137-present trials. Background was an average of 5000 events and the source contribution was an average of 79 events, Poisson distributed, using a single detector module. The difference between the identification fraction for a source with only background present and the identification fraction when the true source is present gives a measure for identification performance.

primarily because there is a larger fraction of Ag-110m identifications in the source case when using the complete system model. This may be due to forcing the algorithm to choose a source position that may not appear especially hot. The fraction of trials identified as each isotope in the background and source-present cases when using the simplified system model is shown in Figure 7.8. Compare this to the background and Cs-137 cases in Figure 7.6b when the source direction is unknown.

Counting Time and Unknown Background

Thus far, experiments have looked at situations with a weak signal-to-noise ratio with a large number of background counts. In this section, we investigate the effect of increasing the measurement time (or equivalently the detector-system efficiency) as the source-to-background ratio is held constant. Figure 7.9 shows the detection performance for each algorithm as the measurement time increases from that required to collect an average of 10 background events to that required to collect an average of 1000 background events. The source is Cs-137 with a strength of 11% of the background intensity, and at least 200 trials of each case were performed.

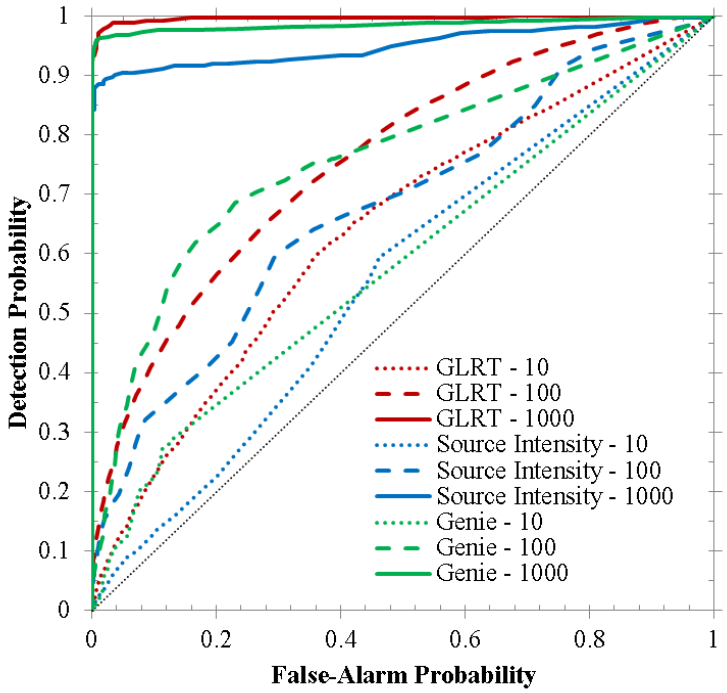

Figure 7.9. An ROC curve of the experimental detection performance on a single detector module for an average background of between 10 and 1000 events, Poisson distributed, as noted in the legend, and an average Cs-137 source intensity of 11% of the background strength, using the source library in Table 7.1. At least 200 trials of background-only and source-present cases were performed for each curve. The performance of each of the algorithms described is shown.

As the measurement time increases, detection performance improves for all algorithms, at approximately the same rate as one would expect due to a signal-to-noise-ratio argument, although for even longer counting times, performance may improve faster than this as enough events are collected to begin to produce smooth images and spectra. At any particular measurement time, the algorithms presented here perform similarly to using all events with Genie. However, using the same set of two-interaction events, these algorithms perform better than using a test statistic from Genie, with the GLRT's performance with a background of only 10 events near that for Genie's after 100 background events.

Identification performance is shown in Figure 7.10 for the complete system model. When few total events are present, sometimes the likelihood ratio is unity for each possible source, so all are equally unlikely. In this case, no source can be identified. However, after collection of 100 events, there is always a most-likely source. The figure

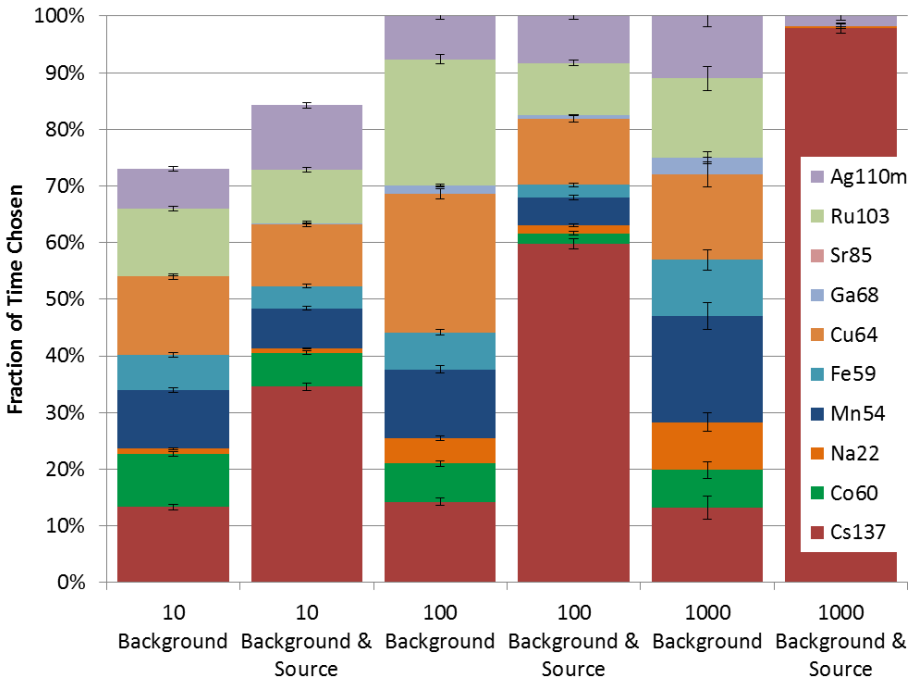

Figure 7.10. The identification performance using the complete system model as the source-to-background ratio is held constant and the collection time is increased. The bars show the fraction of time each nuclide was reported for the background-only trials and for the source-present trials. The order of the isotopes in each column matches that of the legend. In a few cases, all nuclides were equally unlikely, so none was reported. The source was Cs-137 with a strength of 11% of the background. The difference between the identification fraction for a source with only background present and the identification fraction when the true source is present gives a measure for identification performance. A single detector module was used.

shows that when no source is present, Cs-137 is chosen as the most-likely source about 13% of the time, no matter how many events have been collected. When Cs-137 is truly present, it is identified as such a much larger fraction of the time. Even when receiving 1.1 events on average from Cs-137 for the shortest measurement time, the algorithm chooses Cs-137 as the most-likely source 35% of the time. A similar trend is present using the simplified system model, with comparable background performance and correct identifications in the source-present case occurring in about 10% fewer of all trials. Genie identifies Cs-137 in between 1% and 12% of trials in the background case and produces correct identifications in the source-present case similar to those in Figure 7.10.

Unknown Background Simulation Performance

So far, only experimental performance has been shown. The system model includes a number of assumptions that may degrade performance in experiment, including assumptions of a spatially uniform background, no surrounding materials, uniform detector performance, and simplifications in probability calculations [51]. All but the last of these is absent in simulation, so a simulation can give an idea of the performance degradation due to model mismatch within experimental results. In addition, the performance degradation due to the assumption of spatially uniform background can be quantified by comparing simulations of spatially uniform and non-uniform background.

A simulation with the same source and background intensities as the experiment was performed using Geant4 [75]. The background was simulated as either uniform in space or with one hemisphere hotter than the other, with an emission spectrum similar to the measured background spectrum, developed by Robinson *et al*. [74]. In the non-uniform case, one hemisphere was either twice as hot or eight times as hot as the other. Experimentally, in our laboratory, the background image is about a factor of two non-uniform with extra non-uniformity due to detector response. Other environments may be more or less uniform. A 662-keV parallel-beam flood source was used as the source and was placed in either the hot or the cold hemisphere. In experiment, the source was on the boundary between the hemispheres. Realistic energy and position resolution were simulated.

The detection performance using the simulation is shown in Figure 7.11 for approximately 50 trials of the background-only and source-present cases. It appears that model mismatch of background non-uniformity has little effect on the simulated performance when the background is a factor of two hotter in one hemisphere, no matter where the source is located. That is, the background events still appear more probable to be spatially uniform than to be spatially uniform with a single point source matching the energy spectrum of a library source, whereas the source events still appear to come from a localized direction with a spectrum in the library even when mixed with the background events. This may be because many background events are equally likely to be caused by photons from one hemisphere as the other, but the events are not as likely to be due to

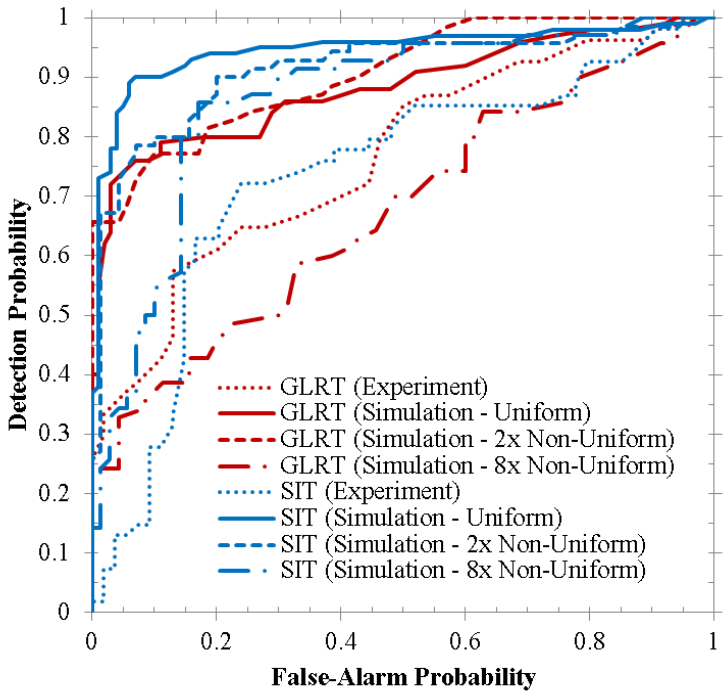

Figure 7.11. An ROC curve of the detection performance for an average background of 5000 events and an average Cs-137 source of 79 events, Poisson distributed, using the source library in Table 7.1 and a single detector module. Simulated performance with spatially uniform background (solid lines) and background with a hemisphere twice as hot as the other (dashed lines) or 8 times as hot as the other (dot-dashed lines) is compared to experimental performance (dotted lines). In the simulation, the source position was always in the cold hemisphere where detection performance is worst. Approximately 50 trials of background-only and source-present cases were performed. The two colors correspond to two different detection algorithms.

photons from the exact same place as other events. The assumption of spatially uniform background may therefore be sufficient for backgrounds that are a factor of two non-uniform between two hemispheres.

When the background is a factor of eight non-uniform, there is no performance degradation when the source is in the hot hemisphere (not shown in Figure 7.11), but the performance can be significantly worse if the source is in the cold hemisphere, especially for the GLRT. Because the background modeled in the algorithms is spatially uniform, point sources in the cold hemisphere can appear more like a uniform background than a uniform background with single hot spot. Point sources in the hot hemisphere stand out more since they fall on top of a hot region.

Also included for comparison is the performance using experimental data. Performance is always better in simulation than experiment except for the case of background with factor-of-eight non-uniformity and a source in the cold region. Although a portion of the performance degradation may be due to local non-uniformities in the background, the significant difference between simulation and experiment in the case with non-uniform background suggests that a large portion is due to other model mismatch. The achievable performance with model or detector improvements is seen in the simulated results.

Genie performance is not included on this figure since the simulated background includes a significant peak at 662 keV (an emission line of Cs-137). Genie therefore detects Cs-137 presence whether or not a point source is present. With unknown background, Genie therefore cannot differentiate between detection of the 662-keV line from background and a Cs-137 point source.

Known Background

In some situations, one might know something about the spectral distribution of the background – either the spectral shape or the shape and intensity. In this study, an 18-detector system was used to study the improvement achievable when the background is known. First, the GLRT and SIT performance in an unknown background is shown for this system using multiple-interaction events when the background count rate can vary by up to a factor of two. This is compared with the performance of Genie 2000 using its optimal parameters, when it is allowed to use all events from the detector. Finally, known background is included and the performance shown.

GLRT and SIT with unknown background

Performance of the GLRT and SIT are shown in Figure 7.12 and Figure 7.13 for an unknown background using an 18-detector array. This is the baseline performance when nothing is known about the background or sources. In this case, as above, the background spectrum must be estimated along with the source information. Using the complete model for both the GLRT and SIT, detection and identification performance is better than using Genie 2000, as shown in Figure 7.12 and Figure 7.13. The full-energy

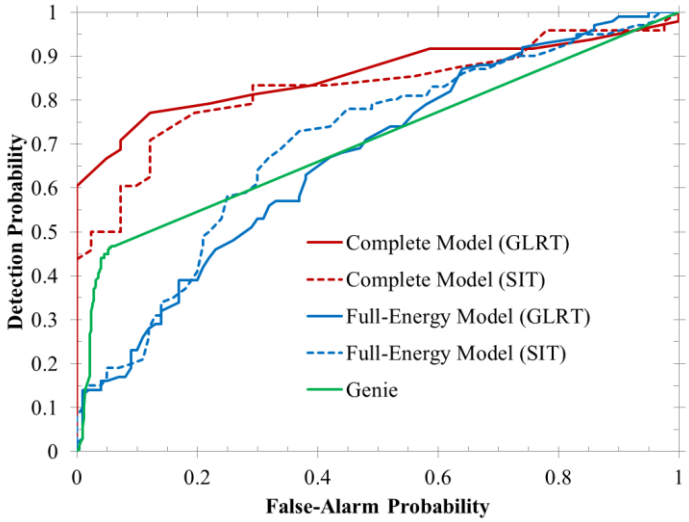

Figure 7.12. Detection performance for the GLRT and SIT using either the complete system model or the computationally simpler full-energy model for a Cs-137 source around the 18-detector array system when the background spectrum is unknown. A trial is considered a detection if the test statistic was above a threshold. In addition, Genie 2000 performance is shown for the same measurement time.

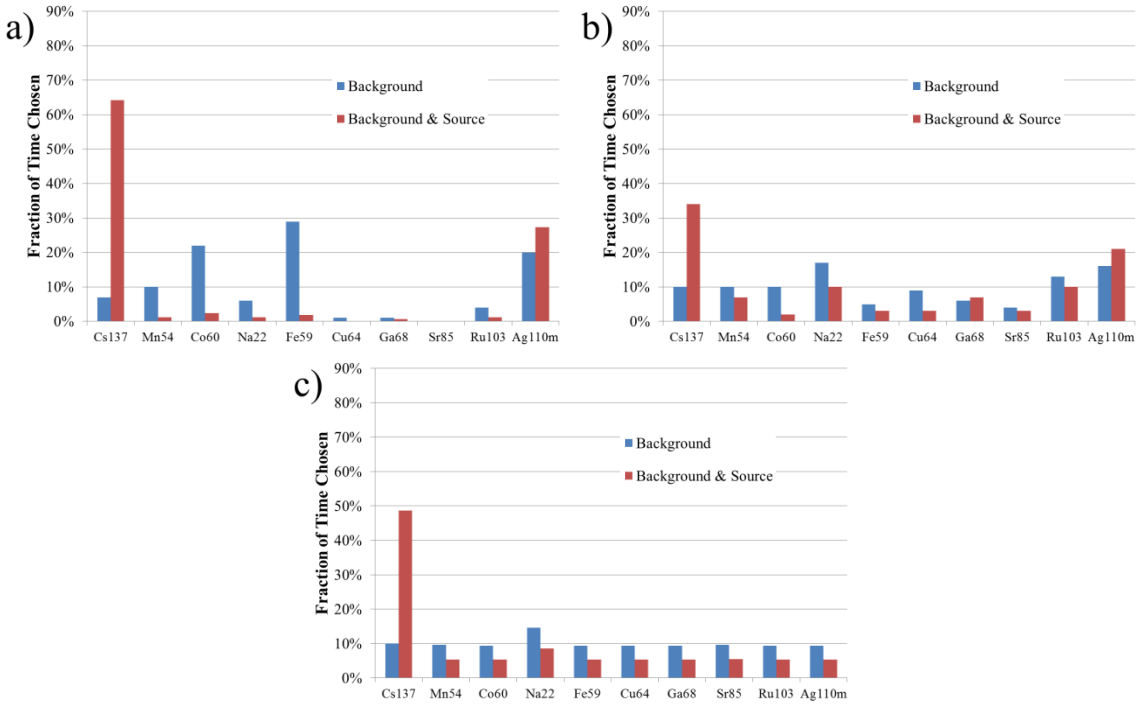

Figure 7.13. Identification performance for the GLRT or SIT using either a) the complete system model or b) the computationally simpler full-energy model for a Cs-137 source around the 18-detector array system when the background spectrum is unknown. Since both the GLRT and SIT use the same parameter estimates, identification performance is the same for both detection methods. c) The identification performance for Genie 2000 using its optimal settings. If any sources were equally likely, one of the sources was choosen randomly.

model presumably performs less well than the complete model because of greater mismatch between the experiment and the model. Since the GLRT and SIT make no assumptions of background spectral shape but use imaging, they are also able to detect the presence of point sources with emission lines also present in background and locate the direction of sources, which is not possible using Genie 2000.

GLRT and SIT with known background

The performance for the GLRT and SIT are shown in Figure 7.14 and Figure 7.15 when a background spectral shape, but not intensity, is provided. This spectral shape was from a prior long background measurement. Detection and identification significantly improve over an unknown background. Known background spectral shape produces more accurate likelihood calculations since the background is known rather than estimated, and it also produces better estimates of other parameters, such as the source intensity and source direction. Therefore, if the background spectrum is known, it can be used to improve detection and identification performance.

Again, the model assuming full-energy deposition does not perform as well as the more complete system model, presumably because of model mismatch. However, when the background shape is known, the performance degradation in using this more computationally efficient method is less severe.

Performance in Genie changes very little if the background is known. Subtracting background even when the background count rate is known results in a very similar curve to the unknown background case [147].

Multiple Source Detection

Unknown number of sources

When the number of possible sources is unknown, the number of sources must be estimated, as described on page 114. The modified GLRT has two parameters that affect the false- and true-alarm rates: the threshold on the test statistic T and the amount to penalize each additional source β. The latter describes the sensitivity to the presence of multiple sources, and the former describes the total-alarm probability.

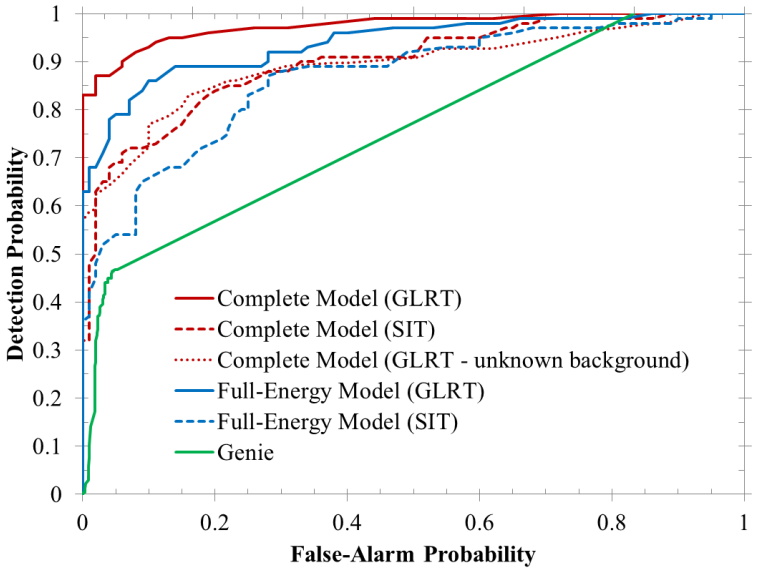

Figure 7.14. Detection performance for the GLRT and SIT using either the complete system model or the computationally simpler full-energy model a Cs-137 source around the 18-detector array system when the background spectrum is known. A trial is considered a detection if the test statistic was above a threshold. In addition, Genie 2000 performance is shown for the same measurement time.

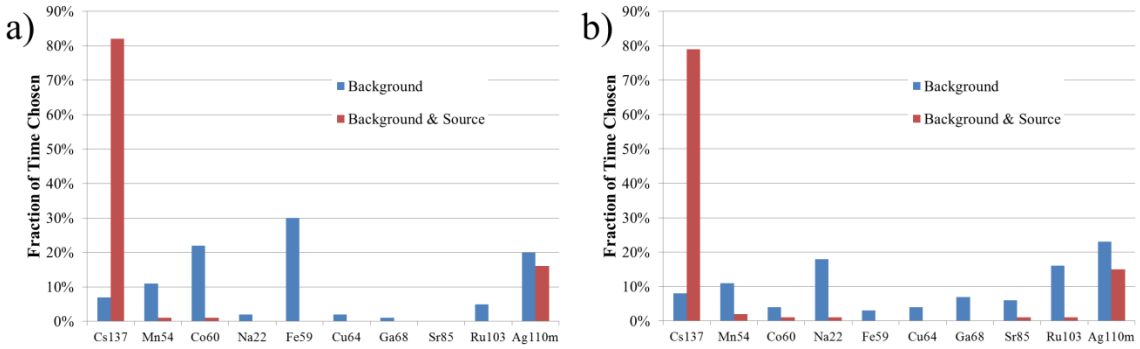

Figure 7.15. Identification performance for the GLRT or SIT using either a) the complete system model or b) the computationally simpler full-energy model for a Cs-137 source around the 18-detector array system when the background spectrum is known. Since both the GLRT and SIT use the same parameter estimates, identification performance is the same for both detection methods. Genie performance is similar to that shown in Figure 7.13c.

Figure 7.16 shows the probability to estimate the presence of two sources when two Cs-137 sources are present and when no sources are present as a function of the model-order parameter β. Each Cs-137 source contributes an average of 235 multiple-interaction events in a background of 4050 multiple-interaction events. Full-energy deposition was assumed, the background was unknown, and the sources were assumed to

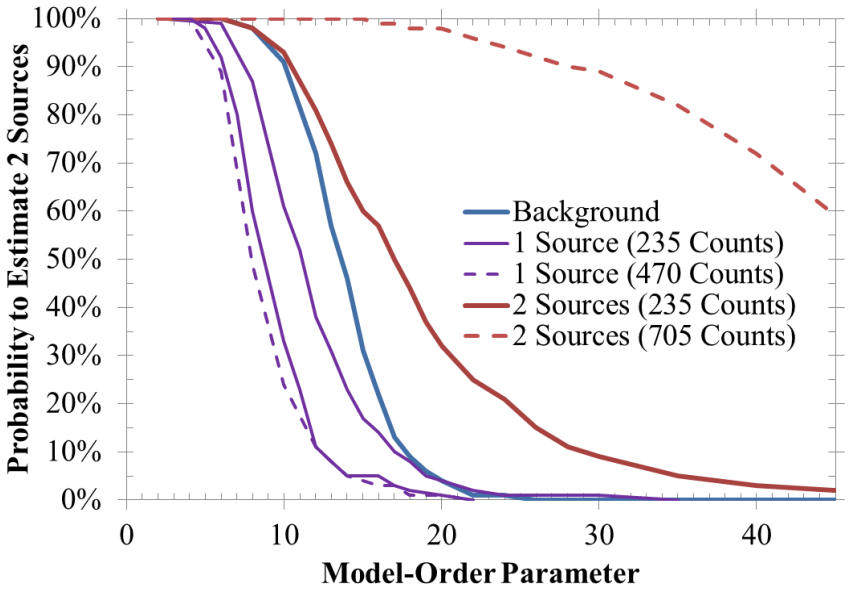

Figure 7.16. The probability to estimate two sources instead of one as a function of the model order parameter β for the case of just background, one true Cs-137 source, and two true Cs-137 sources with two different source strengths. 200 trials of each case were performed.

be from the library in Table 7.1. If the model-order parameter is set very low, two sources are estimated to be present in every single trial, even if one source or no sources are present. As the parameter increases and more highly penalizes multiple sources, the trials with the two sources are estimated to have two sources a larger fraction of the time than the background or single-source trials. Increasing the source strength or decreasing the background gives a larger separation; the stronger the single source, the lower a model-order parameter is needed to estimate it as a single source and the stronger each of the two sources, the higher the model-order parameter can be to still estimate two sources. This is shown in the figure for a single source with twice the strength and for two Cs-137 sources with thrice the strength. A similar curve exists for trials with two different isotopes or when using the complete radiation-interaction model.

Due to model mismatch, the model-order parameter must be significantly larger than unity. The choice of operating location depends on the consequence of incorrectly alarming for two sources when one or none is present and the consequence of missing a second source that is truly present. We choose to operate, rather arbitrarily, with a β value of 16 that gives just over 50% correct estimation of two sources for the weakest

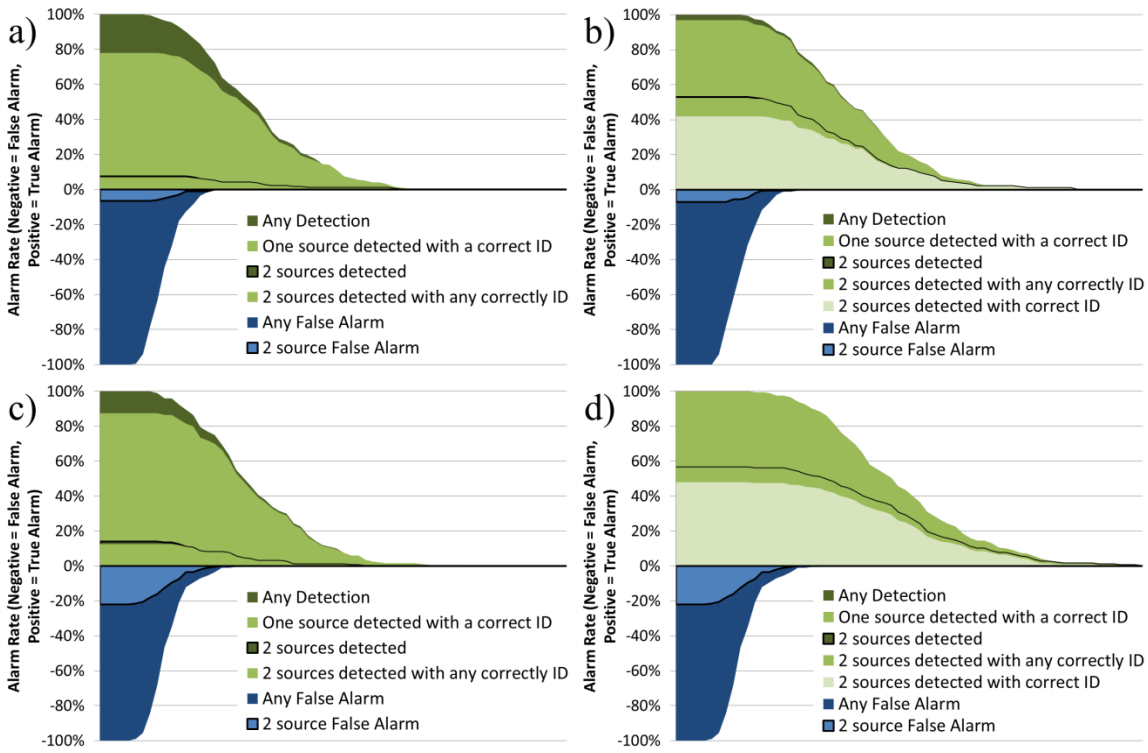

Figure 7.17. The types of alarms as a function of test-statistic threshold when the model-order parameter is set to 16. The complete system model is used in a) and b), and the full-energy model is used in c) and d). Source trials in a) and c) have a single Cs-137 source present and in b) and d) have two Cs-137 sources present.

case studied without estimating two sources a large fraction of the time in the background or single-source case.

Figure 7.16 also shows that when only one Cs-137 source is present, it is correctly estimated as a single source even a larger fraction of the time than for background. This result is reasonable because the test statistic for detection for the background trials is generally lower than for any of the source-present trials. That is, the background trials see little evidence that any source exist, but if forced to say a source or sources are present, the data can often be explained better by estimating two sources rather than one.

To understand the detection statistic suppressed in this plot better, consider Figure 7.17. Figure 7.17 shows the types of detections as a function of the threshold on the test statistic for the single- and two-Cs-137 case when the model-order parameter is set to 16. The false-alarm rate recorded from the background trials is shown as negative alarm rates and true positive alarms recorded from the source-present trials are shown as positive alarm rates. The solid line shows the fraction of multiple-source detections as a function

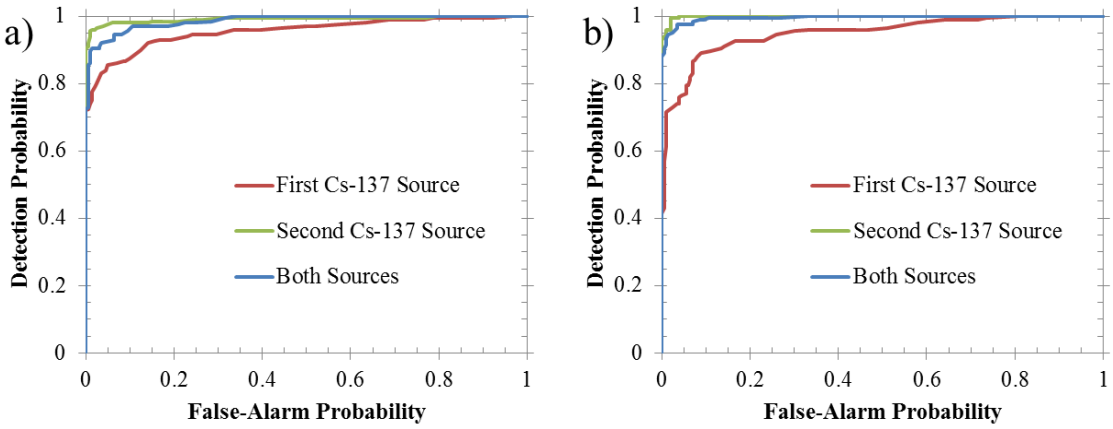

Figure 7.18. The ROC curve detection performance for each Cs-137 source individually in background and when both sources are present in the same background. a) uses the complete sytem model and b) uses the full-energy model. Any test statistics over the threshold is considered a detection or false-alarm irrespective of the identification or number of sources estimated. The model-order parameter β is set to 16.

of threshold. It is important to remember that the fraction of multiple-source detections can be increased by decreasing the model-order penalty parameter. Then, a larger fraction of two-source trials will have both sources detected and correctly identified, at the expense of incorrectly estimating two sources when only one is present.

As in Figure 7.16, Figure 7.17 shows that a larger fraction of trials are identified as containing two sources when two sources are actually present, compared to background-only or one-source trials. This is true whether the full-energy model or the complete system model is used. In addition, we see that in this case the overall detection performance – if an alarm goes off when at least one source is present and does not when all sources are absent – is significantly better when there are two sources. This is not always the case, especially if there are multiple sources of the same isotope. Depending on their locations, multiple sources can look like a uniform spatial background (if they are spread evenly in space), or a single hot spot (if they are close to one another). The second case gives much better detection. In this example, better detection can be explained by the second source being located in a direction – directly in front of a plane – that is more easily detectable. Figure 7.18 shows the ROC curve for each Cs-137 source individually and for both together. It is clear that detection in the two-source case is similar to that when just the second source is present.

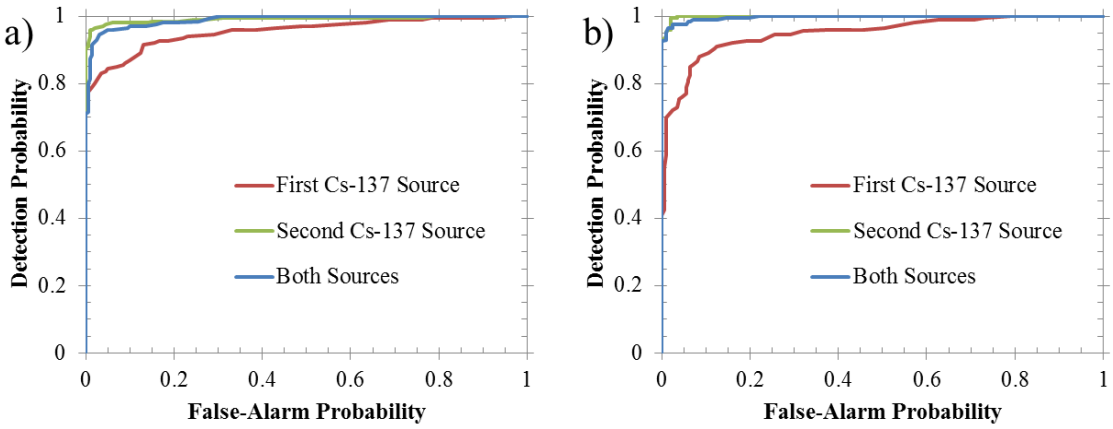

Figure 7.19. The ROC detection performance for each Cs-137 source individually in background and when both sources are present in the same background when the number of sources present is known for each set of trials. a) uses the complete sytem model and b) uses the full-energy model. Any test statistics over the threshold is considered a detection or false-alarm irrespective of the identification or number of sources estimated.

Figure 7.17 also shows identification performance. The background trials estimate all isotopes about evenly, with no isotope being estimated in more than 17% of the trials, except for Ag-110m, which tends to be preferentially estimated as the second source, appearing as the second source 22% of the time. In the source-present trials, identification as Cs-137 is almost always correct, especially when two sources are present. Those trials in which the second source does not exist, or is incorrectly identified, the identification is spread evenly between all sources, but with a preference for Ag-110m, especially when the complete system model is used.

Known number of sources

When the number of sources that could possibly be present together is known a priori, the standard GLRT can be used for detection just by adding new parameters to estimate the position, identity, and intensity of each extra source.

Figure 7.19 shows the detection performance when the number of sources is known for each case in Figure 7.18. Detection performance is very similar no matter if the number of sources is known. For the case of two sources, each trial will give two identifications. Figure 7.20 shows the results of identifications reported for each trial regardless of whether they were detected. The distribution of isotopes chosen as the first source when there is just background is the same whether there is thought to be a single

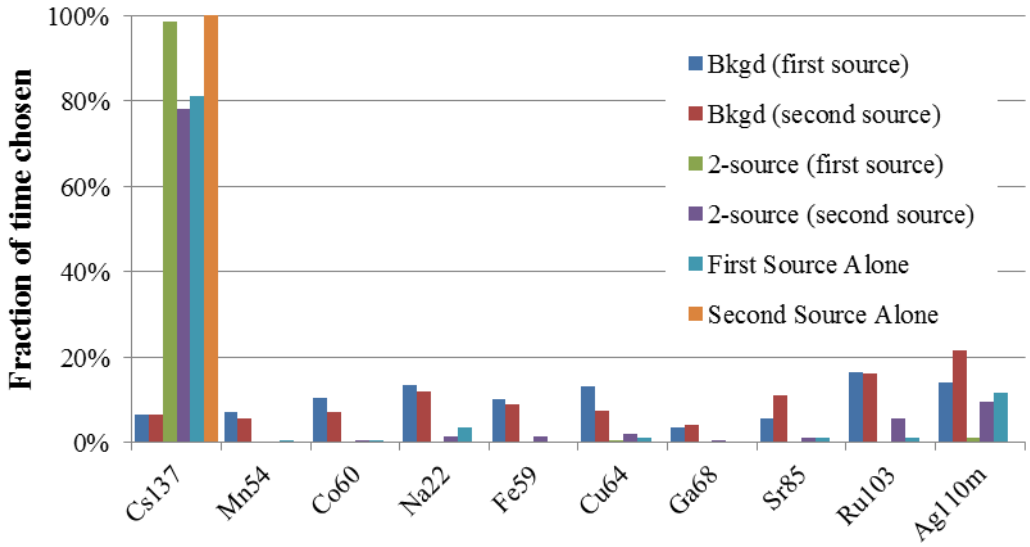

Figure 7.20. The identification performance when the number of sources present is known for each set of trials with one source present and with trials in which both sources are present.

source or two sources. When asked to find a second source, very similar isotopes are estimated, as shown in the second series of Figure 7.20. When there is a source present, the identification performance depends on the source position. In these trials, the second source is in a position with very good imaging performance and therefore has better detection and identification even though the number of recorded counts is similar. The last two series of Figure 7.20 show the identification performance for each source alone. When two sources are present together, the algorithm is correct a larger fraction of the time for the first source it estimates since this is the easier source to find. This is consistent with other source pairs tested. Still, the fraction of trials with correct identifications for both sources is similar to the fraction correct when each source appears individually. The same is true for the estimated source locations. For these sources and locations, there does not appear to be any confusion due to the presence of multiple sources.

Conclusions

This study investigated the use of direction and energy information to detect and identify point sources of radiation in an unknown or known background. Two algorithms were applied that both use maximum-likelihood estimates of the background intensity and spectrum and the source identity, intensity, and direction. In the case of a weak

single source possibly present in a somewhat spatially non-uniform, unknown background, the generalized likelihood ratio test (GLRT) and the source-intensity estimate scaled by efficiency and the square root of the total counts (the SIT) both performed as well or better than using a test statistic of the maximum reported source activity from the commercial spectral-analysis software Genie 2000. In addition, the GLRT and SIT, unlike Genie, made no assumption that the source lines were absent from background. The better performance of the GLRT and SIT may be because they are statistical algorithms and can discriminate the presence of a point source from distributed background. Using the methods described above, identification performance was also observed to be similar with Genie, and the algorithms presented in this work were correctly able to reject a possible source with emission line also present in the background.

Knowing the source direction was shown to improve detection performance by decreasing the false-alarm rate. This has implications for applications in stand-off detection on a specific target or, with a more efficient detector system, at chokepoints where the possible source is known to be in a moving vehicle or container. By visually tracking the object, one can assume a possible source direction, improving detection.

Performance degradation due to some aspects of model mismatch also was presented. However, no significant performance degradation was observed due to background that had one hemisphere twice the intensity of the other when the model expected a spatially uniform background.

Performance was also investigated using an array system consisting of 18 position-sensitive CdZnTe detectors. The GLRT and SIT have improved detection and identification over standard spectroscopic software running on the same set of events when the system model including partial-energy deposition is used, even when the background is unknown. With known background spectral shape, detection and identification performance greatly improved. Therefore, if information about the background spectrum is available, it should be included in the detection and identification methods.

Finally, performance was investigated when multiple sources are possibly present, both when the number of sources is known a priori and when only the maximum number

is known. In the first case, detection and identification was found to perform similarly to when a single source is present. In the second case, a penalizing function was used to control the sensitivity to additional sources. This gave some degree of separation between trials with two Cs-137 point sources and a single Cs-137 source. Background trials estimated the presence of two sources more often than when a single source was present, although with a low detection test statistic. Although estimation of the number of sources was not always reliable, detection and identification performance of the sources that were estimated to be present was similar to looking for each source individually.

The GLRT and SIT are useful for detection and identification of threat sources when the background is unknown and may contain the same emission lines as a threat source. In this situation, Genie and standard spectroscopic methods cannot function. Yet, even when the source has distinct emission lines from background, the GLRT and SIT often have better detection performance than Genie. Therefore, even for situations in which the source emission lines are known to be distinct from background or the background spectrum is known, the GLRT and SIT may improve detection over implementations of standard spectroscopic methods such as Genie. However, since the types of algorithms in Genie and in the GLRT and SIT are quite different, the performance difference may be explained by differences in the capability of the algorithms to detect a weak signal along with differences in the utility of the information each uses. Nevertheless, the GLRT and SIT have not been designed for the situation of unique source-emission lines; better performance is still possible by including constraints on the background spectral shape so as not to contain emission lines similar to a source and by using the spectral information from single-interaction events.

Chapter 8
Summary and Future Work

There are more things in heaven and earth, Horatio,
Than are dreamt of in your philosophy.
– Shakespeare, Hamlet, Act 1, Scene 5

Summary

Three-dimensional position sensitive, room temperature, semiconductor detectors offer a number of advantages over other systems, including good energy resolution, quick start-up time after calibration, and the ability to image. This dissertation has presented a number of new imaging and detection algorithms to use the data from these systems for practical applications. In addition, software has been designed to implement these algorithms.

Imaging gamma-rays with position-sensitive detectors can take many forms, using the different types of gamma-ray interactions. In most all methods, the inverse problem must be solved – finding the gamma-ray distribution from a set of measurements. Simple back-projection solves this roughly by multiplying the count distribution by the transpose of the modeled forward-projection matrix, with heuristic modifications so that it better approximates a true inverse. If the forward-projection matrix is known well, better images can be obtained by estimating the maximum likelihood gamma-ray source distribution using the MLEM method. The dimensions of the reconstruction can be chosen to match the situation. The new method presented here of reconstructing to an isotope-direction domain instead of an energy-direction domain can give the isotopes present in each direction or location around the system. In addition, the smaller number of parameters may improve estimates, especially with few events. Current results are complicated by model mismatch and energy calibration. When moving targets are present, a new method is now available to reconstruct the image portion to a combined

backdrop and target space to find the energy distribution and intensity in each direction of the backdrop and each moving target.

To perform these imaging algorithms efficiently with a number of options, software called UMImaging was designed. UMImaging is platform independent, with a modular structure that allows easy redesign of the interfaces and easy addition of reconstruction methods and options. A number of basic geometries are included to perform calculations for the reconstruction methods. It supports highly multidimensional data analysis with arbitrary binning in each dimension, which informs the user about detector and algorithm performance. It also supports parallel computation.

The extra dimension of imaging available on position-sensitive gamma-ray detectors can help improve the detection of point sources in an extended background. Analysis has been performed to quantify this improvement. Using only the spectrum for an 18-detector array, best SNR for detection of a known source with emission lines not in the background occurred using all events. By only including events that passed the known Cs-137 source direction, the SNR using multiple-interaction events improved considerably, but total SNR using all events was only 7% higher when imaging was included. Applying the Bonferroni correction when the source direction is unknown, the addition of imaging only improved detection performance at very low false-alarm rates on this system. Similar analysis can be applied to future imaging systems to quantify their imaging performance for detection and inform system design.

Still, imaging is important when searching for a source, by telling the user in which direction to look. Imaging is also the only way to detect a point source that may share emission lines with the background since in that case the presence of a peak is not sufficient evidence for the presence of the point source. Two algorithms have been applied to this situation. The GLRT and SIT use a model of a spatially uniform background with either unknown energy distribution, known shape of the energy distribution, or completely known background, and a set of point sources with some intensities, directions, and isotope identities. Performance for detection and identification has been shown for each case and, in almost all cases, exceeds that of detection based only on the spectrum using peak fitting. Known source location, source identity, or background shape can considerably improve detection performance. Finally, detection of

multiple sources using a modified GLRT performs as well as detection of a single source when the number of sources is known. Estimation of the number of sources present is also possible, and trials with two sources present estimated two sources a larger fraction of the time than trials with only one source.

Future Work

Future work may make incremental improvements in these methods to estimate better, handle more situations, and run faster. As in all science and engineering, there is also the possibility for fresh, new ideas that will perform better than these, which are not even dreamt of in our philosophy. In this section, I hope to briefly describe some of the incremental improvements that will make the imaging and detection systems more accurate, versatile, and rapid.

Imaging

Over the last decade, imaging algorithms for three-dimensional-position-sensitive systems have emulated those in the medical-imaging field, from simple back-projection, to filtered back-projection, to MLEM. It is hence useful to imitate some of the more recent advances in medical imaging, including calculation speed improvements, motion compensation, image regularization, quantitative imaging, and model improvements.

Model Improvements and Quantitative Imaging

One of the most pressing, yet incremental, issues in the current system is mismatch between the modeled response and the true response. This hurts detection, skews intensities of sources in different directions, reduces the usefulness of deconvolution into the isotope domain, and limits the use of single-interaction and coded-aperture imaging. Detector characteristics such as the finite detector voxel size, anode and edge dead layers, pixel jumping, and surrounding material currently are omitted from the model, as are physical processes such as pair production. Some of these factors need to be included in the system matrix calculation itself, while others, such as surrounding materials, need only be included in the simulations used to calculate the sensitivity. Together with dead time corrections, the system should be able to accurately estimate the

activity of radioactive sources at known distances. In addition, significant work is still needed to understand and predict the error in MLEM estimates.

Regularization

As mentioned on page 41, regularization can be used to improve image quality if we expect certain features in images. For instance, in all the situations described above, the energy-image is expected to be smooth, or even flat, in most regions, with a few sharp peaks in the image (point sources), and the spectrum (photopeaks). The extent of the flat regions may even be known by looking for object edges in an optical image of the same scene. Regularizers can be built to encourage energy-images that appear as expected and penalize energy-images that look significantly different.

Moving Objects and Detectors

Imaging the radiation produced from moving objects now occurs using the known object direction from optical cameras. Stronger sources could also be tracked by matching up the recorded data themselves, as is sometimes done in medical imaging to improve image quality.

With a moving detector, or multiple detectors, reconstruction will probably occur in three-dimensional space. However, it is not clear how to define the region over which to reconstruct. Reconstructing a high-resolution map everywhere can become quite memory and computationally intensive due to the possibly large regions and additional spatial dimension compared with 2D imaging. On the other hand, reducing the region size could lead to spectral and spatial artifacts if a hot source exists just outside the region.

Moving or multiple detectors also provide a rich field to investigate for detection algorithms. For instance, how should the detectors move or where should they be placed for optimal detection or search performance? Do the answers change when the detectors can image? How can you quantify the effect of detector or algorithm improvements in search scenarios? What methods can be used to reduce computation when there are many detector systems providing copious data? Would coincident events between detector systems ever be useful?

Calculation Speed

The limiting factor for many of these improvements is computation time. Very accurate system models are possible with numerical integration through all possible uncertainties, but doing so requires significant computation [148]. Therefore, by improving the system model, we may increase computation time. Many studies in medical imaging have shown order-of-magnitude improvement in image reconstruction speed using graphics processing units (GPUs). GPUs excel at the ray tracing required for calculating path lengths and intersections, so it is worth considering this approach for image reconstruction, especially in 3D reconstructions. It might also be worth considering alternative energy and spatial parameterizations, which could reduce calculation time by simplifying some calculations or reducing the number of bins in some regions. Other imaging and detection algorithms may also be faster.

Detection

There are a few limitations with the detection algorithms presented in this work. First, they assume a spatially uniform background. While this does not seem to affect performance for the backgrounds in laboratory environments, more non-uniform backgrounds could be problematic. The performance of these algorithms should be tested in other background environments, and, if necessary, other background spatial distributions should be included in the model by parameterizing background spatial distributions using an appropriate set of basis functions. Exploiting optical images of the scene may help to include background spatial distributions.

Second, these detection algorithms require a known emission spectrum from the threat sources for best performance. Shielded sources, however, have an emission spectrum partially shifted to lower energies due to interactions in the shield. Certainly the direct radiation can still be detected as indicative of the source, but the additional gamma rays at lower energies may confuse identification and will not contribute to the source signal. One could build the type and amount of shielding into the model to handle situations with shielded sources.

Third, the algorithms do not make use of time-series data, which could be informative. For instance, some sort of background tracking could help to inform the

algorithm of the range of possible backgrounds, giving performance in between known and completely unknown background. The effect of poor background estimates should also be studied.

Fourth, the GLRT and SIT rely on a library of source identities, and detection will not occur unless a source matches one of these. Sources not in the library should be detected as well; however, expanding the library size leads to more false alarms and a greater computational burden. Other types of algorithms, such as anomaly detection, do not require an explicit library [149].

Finally, the GLRT and SIT are computationally slow unless tricks are used which somewhat degrade performance. Including shielded sources or non-uniform backgrounds in the model could worsen this. Also, the optimization algorithms for isotope identity and source location could certainly be faster. Beyond this, there are a multitude of other detection algorithms, such as pattern recognition, anomaly detection, or simple peak-fitting algorithms, which might have as good, or better, detection performance with less computation and fewer of the drawbacks described above.

System Improvements

The future work described above is independent of any hardware improvements. Possible changes to the hardware for better performance in different scenarios are extensive, but one stands out as practical for improving imaging and detection: digital waveform readout. Reading the digital waveforms from each electrode allows sub-pixel resolution of the interaction location and more precise understanding of the interactions, such as discriminating multiple interactions under a single pixel or identifying photoelectric absorptions in some situations. These new data can be used in the system model to improve imaging and detection.

Other Applications

Applications for 3D-position-sensitive semiconductor systems abound. In the field of detection and imaging for security applications, for instance, these systems could be used for imaging of thermal neutrons using the high thermal-neutron cross section of Cd-113 [150], muon tomography to find high-Z and dense objects in cargo [151,152],

active interrogation, or other missions. Certainly, this technology has come far, but there is still much to be done.

Bibliography

[1] Feng Zhang, "Events Reconstruction in 3-D Position Sensitive CdZnTe," University of Michigan, Ann Arbor, MI, PhD Thesis 2005.

[2] Willy R. Kaye, "Event Reconstruction in Pixelated CdZnTe Detectors," University of Michigan, Ann Arbor, MI, PhD Thesis 2011.

[3] Glenn F. Knoll, *Radiation Detection and Measurement*, 4th ed. USA: John Wiley & Sons, 2010.

[4] Feng Zhang and Zhong He, "New Readout Electronics for 3-D Position Sensitive CdZnTe/HgI2 Detector Arrays," *IEEE Transactions on Nuclear Science*, vol. 53, no. 5, pp. 3021-3027, October 2006.

[5] Feng Zhang, Zhong He, and Carolyn E. Seifert, "A Prototype Three-Dimensional Position Sensitive CdZnTe Detector Array," *IEEE Transactions on Nuclear Science*, vol. 54, no. 4, pp. 843-848, August 2007.

[6] Zhong He, "Review of the Shockley-Ramo theorem and its application in semiconductor gamma-ray detectors," *Nuclear Instruments and Methods in Physics Research A*, vol. 463, pp. 250-267, 2001.

[7] Z. He, G. F. Knoll, D. K. Wehe, and J. Miyamoto, "Position-Sensitive single carrier CdZnTe detectors," *Nuclear Instruments and Methods in Physics Research A*, vol. 388, pp. 180-185, 1997.

[8] Zhong He, "Three-Dimensional Position-Sensitive Wide Bandgap Semiconductor Gamma-Ray Imaging Detectors," in *Future Trends in Microelectronics*, Serge Luryi, Jimmy Xu, and Alex Zaslavsky, Eds. Hoboken, NJ, USA: John Wily & Sons, 2010, pp. 321-329.

[9] Yuefeng Zhu, Stephen 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*, vol. 58, no. 3, pp. 1400-1409, June 2011.

[10] Hao Yang, "Identification and Reconstruction of Multiple Interaction Under Single Anode Pixel Events," in *IEEE Room-Temperature Semiconductor Detectors Workshop Conference Record*, Valencia, Spain, 2011, In Press.

[11] Yuefeng Zhu, "Digital Signal Processing Methods for Pixelated 3-D Position Sensitive Room-Temperature Semiconductor Detectors," University of Michigan, Ann Arbor, PhD Thesis 2011.

[12] Weiyi Wang, William R. Kaye, Jaecheon Kim, Feng Zhang, and Zhong He, "Improvement of Compton Imaging Efficiency by Using Side-Neighbor Events," in *IEEE Nuclear Science Symposium Conference Record*, Knoxville, TN, 2010, pp. N41-144.

[13] S. J. Kaye, W. R. Kaye, and Zhong He, "Experimental demonstration of coded aperture imaging using thick 3D-position-sensitive CdZnTe detectors," in *IEEE Nuclear Science Symposium Conference Record*, Orlando, FL, 2009, pp. 1902-1906.

[14] C. E. Moss, C. L. Hollas, G. W. McKinney, and W. L. Myers, "Comparison of Active Interrogation Techniques," *IEEE Transactions on Nuclear Science*, vol. 53, no. 4, pp. 2242-2246, August 2006.

[15] J. M. Hall et al., "The Nuclear Car Wash: Neutron interrogation of cargo containers to detect hidden SNM," *Nuclear Instruments and Methods in Physics Research B*, vol. 261, pp. 337-340, 2007.

[16] Steven W. Smith, "X-ray Backscatter Detection System," Patent 5,181,234, January 19, 1993.

[17] M. Lenti, "A 3-D imaging device using Compton scattering off the body," *Nuclear Instruments and Methods in Physics Research A*, vol. 588, pp. 457-462, 2008.

[18] Q. Looker et al., "Demonstration of imaging via backscattering of annihilation gamma rays," *Nuclear Instruments and Methods in Physics Research A*, vol. 615, pp. 295-300, 2010.

[19] Robert C. Runkle, Timothy A. White, Erin A. Miller, Joseph A. Caggiano, and Brian A. Collins, "Photon and neutron interrogation techniques for chemical explosives detection in air cargo: A critical review," *Nuclear Instruments and Methods in Physics Research A*, vol. 603, pp. 510-528, 2009.

[20] D. L. Chichester and S. M. Watson, "Multispectral UV-Visual Imaging as a Tool for Locating and Assessing Ionizing Radiation in Air," in *IEEE Nuclear Science Symposium Conference Record*, Knoxville, TN, 2010.

[21] Peter Marleau, James Brennan, Erik Brubaker, Nathan Hilton, and John Steele, "Active Coded Aperture Neutron Imaging," in *IEEE Nuclear Science Symposium Conference Record*, Orlando, FL, 2009, pp. 1974-1977.

[22] P. Lindén, J. K.-H. Karlsson, B. Dahl, I. Pázsit, and G. Por, "Localisation of a neutron source using measurements and calculation of the neutron flux and its gradient," *Nuclear Instruments and Methods in Physics Research A*, vol. 438, pp. 345-355, 1999.

[23] Nicholas Mascarenhas, James Brennan, Kevin Krenz, Peter Marleau, and Stanley Mrowka, "Results With the Neutron Scatter Camera," *IEEE Transactions on Nuclear Science*, vol. 56, no. 3, pp. 1269-1273, June 2009.

[24] Robley D. Evans, *The Atomic Nucleus*. New York, USA: McGraw-Hill, 1955.

[25] Paul Kirkpatrick and A. V. Baez, "Formation of Optical Images by X-Rays," *Journal of the Optical Society of America*, vol. 38, no. 9, pp. 766-774, September 1948.

[26] E. Caroli, J. B. Stephen, G. Cocco, L. Natalucci, and A. Spizzichino, "Coded aperture imaging in X- and gamma-ray astronomy," *Space Science Reviews*, vol. 45, no. 3-4, pp. 349-403, 1987.

[27] Andrzej Czarnecki and Savely G. Karshenboim, "Decays of Positronium," in *Proceedings of the 14th International Workshop on High Energy Physics and Quantum Field Theory (QFT)*, Moscow, Russia, 1999, pp. 538-544.

[28] Krysztof Kacperski, Nicholas M. Spyrou, and F. Alan Smith, "Three-Gamma Annihilation Imaging in Positron Emission Tomography," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 525-529, April 2004.

[29] D. Xu, "Gamma-ray imaging and polarization measurement using 3-D position-sensitive CdZnTe detectors," University of Michigan, Ann Arbor, MI, PhD Thesis 2006.

[30] Andreas Zoglauer and Gottfried Kanbach, "Doppler Broadening as a Lower Limit to the Angular Resolution of Next Generation Compton Telescopes," in *Proceedings of SPIE vol. 4851*, Waikoloa, Hawaii, 2003, pp. 1302-1309.

[31] Brian Plimley et al., "Reconstruction of electron trajectories in high-resolution Si devices for advanced Compton imaging," *Nuclear Instruments and Methods in Physics Research A*, In Press.

[32] Mark D. Hammig and Randy C. Stevenson, "The Degree of Enhancement in a Gamma-Ray Image Gleaned from Recoil-Electron Tracking," in *IEEE Nuclear Science Symposium Conference Record*, Dresden, Germany, 2008, pp. 2433-2439.

[33] D. Xu, Z. He, C. E. Lehner, and F. Zhang, "4π Compton imaging with single 3D position sensitive CdZnTe detector," *Proceedings of SPIE*, vol. 5540, pp. 144-155, 2004.

[34] Carolyn E. Lehner, Zhong He, and Feng Zhang, "4π Compton Imaging Using a 3-D Position-Sensitive CdZnTe Detector Via Weighted List-Mode Maximum Likelihood," *IEEE Transactions on Nuclear Science*, vol. 51, no. 4, pp. 1618-1624, August 2004.

[35] Crystal L. Thrall, Christopher G. Wahl, and Zhong He, "Performance of Five-or-More-Pixel Event Sequence Reconstruction for 3-D Semiconductor Gamma-Ray-Imaging Spectrometers," in *IEEE Nuclear Science Symposium Conference Record*, Dresden, Germany, 2008, pp. 1299-1301.

[36] S. E. Boggs and P. Jean, "Event reconstruction in high resolution Compton telescopes," *Astronomy and Astrophysics Supplement Series*, vol. 145, pp. 311-321, August 2000.

[37] R. A. Kroeger, W. N. Johnson, J. D. Kurfess, B. F. Philips, and E. A. Wulf, "Three-Compton Telescope: Theory, Simulations, and Performance," *IEEE Transactions on Nuclear Science*, vol. 49, no. 4, pp. 1887-1892, August 2002.

[38] J. van der Marel and B. Cederwall, "Backtracking as a way to reconstruct Compton scattered gamma-rays," *Nuclear Instruments and Methods in Physics Research A*, vol. 437, pp. 538-551, 1999.

[39] R. H. Dicke, "Scatter-Hole Cameras for X-Rays and Gamma Rays," *Astrophysical Journal*, vol. 153, pp. L101-L106, August 1968.

[40] K. P. Ziock, M. Cunningham, and L. Fabris, "Two-Sided Coded-Aperture Imaging Without a Detector Plane," in *IEEE Nuclear Science Symposium Conference Record*, Dresden, Germany, 2008, pp. 634-641.

[41] Ian D. Jupp, Kevin Byard, and Anthony J. Dean, "An improved sampling configuration for a coded aperture telescope," *Nuclear Instruments and Methods in Physics Research A*, vol. 345, pp. 576-584, 1994.

[42] A. Martineau, J. M. Rocchisani, and J. L. Moretti, "Coded aperture optimization using Monte Carlo simulations," *Nuclear Instruments and Methods in Physics Research A*, vol. 616, pp. 75-80, 2010.

[43] E. E. Fenimore and T. M. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Applied Optics*, vol. 17, pp. 337-347, 1978.

[44] Roummel F. Marcia, Zachary T. Harmany, and Rebecca M. Willett, "Compressive Coded Aperture Imaging," in *Proceedings of SPIE Conference on Computational Imaging VII* , 2009, p. 72460.

[45] Wonho Lee and David Wehe, "Hybrid gamma ray imaging - Model and results," *Nuclear Instruments and Methods in Physics Research A*, vol. 579, pp. 200-204, 2007.

[46] L. E. Smith, C. Chen, D. K. Wehe, and Z. He, "Hybrid collimation for industrial gamma-ray imaging: combining spatially coded and compton aperture data," *Nuclear Instruments and Methods in Physics Research A*, vol. 462, pp. 576-587, 2001.

[47] L. J. Schultz et al., "Hybrid coded aperture and Compton imaging using an active mask," *Nuclear Instruments and Methods in Physics Research A*, vol. 602, no. 2, pp. 267-274, September 2009.

[48] Willy R. Kaye, Nathan D. Bennett, Christopher G. Wahl, Zhong He, and Weiyi Wang, "Gamma-Ray Source Location by Attenuation Measurements," in *IEEE Nuclear Science Symposium*, Honolulu, Hawaii, 2007, pp. 1294-1298.

[49] R. C. Byrd, J. D. Drabanski, and B. L. Barraclough, "Monte Carlo calculations of the effective area and directional response of a polyethylene-moderated neutron counter," Los Alamos, Report LA-12850-MS, 1995.

[50] R. C. Byrd, "Directional fast-neutron detectors," Los Alamos, Report LA-12379-MS, 1992.

[51] D. Xu and Z. He, "Gamma-ray energy-imaging integrated spectral deconvolution," *Nuclear Instruments and Methods in Physics Research A*, vol. 574, pp. 98-109, April 2007.

[52] 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," In Preparation.

[53] L. Mihailescu, K. Vetter, and D. Chivers, "Standoff 3D Gamma-Ray Imaging," *IEEE Transactions on Nuclear Science*, vol. 56, no. 2, pp. 479-486, April 2009.

[54] S. J. Wilderman, J. A. Fessler, N. 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*, vol. 48, no. 1, pp. 111-116, February 2001.

[55] Howard Anton, *Elementary Linear Algebra*, 8th ed. New York: John Wiley & Sons, Inc., 2000.

[56] Scott J. Wilderman, W. Les Rogers, Glenn F. Knoll, and John C. Engdahl, "Monte Carlo Calculation of Point-Spread Functions of Compton Scatter Cameras," *IEEE Transactions on Nuclear Science*, vol. 44, no. 2, pp. 250-254, April 1997.

[57] R. C. Rohe, M. M. Sharfi, K. A. Kecevar, J. D. Valentine, and C. Bonnerave, "The Spatially-Variant Backprojection Point Kernel Function of an Energy-Subtraction Compton Scatter Camera for Medical Imaging," *IEEE Transactions on Nuclear Science*, vol. 44, no. 6, pp. 2477-2482, December 1997.

[58] Lucas C. Parra, "Reconstruction of cone-beam projections from Compton scattered data," *IEEE Transactions on Nuclear Science*, vol. 47, no. 4, pp. 1543-1550, August 2000.

[59] Michael J. Cree and Philip J. Bones, "Towards Direct Reconstruction from a Gamma Camera Based on Compton Scattering," *IEEE Transactions on Medical Imaging*, vol. 13, no. 2, pp. 398-407, June 1994.

[60] Dan Xu and Zhong He, "Filtered Back-Projection in 4pi Compton Imaging With a Single 3D Position Sensitive CdZnTe Detector," *IEEE Transactions on Nuclear Science*, vol. 53, no. 5, pp. 2787-2795, October 2006.

[61] Alfred O. Hero, *Statistical Methods for Signal Processing*. Ann Arbor, MI: Course Notes (EECS 564), 2008.

[62] L. A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," *IEEE Transactions on Medical Imaging*, vol. MI-1, no. 2, pp. 113-122, October 1982.

[63] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1-38, 1977.

[64] William Hadley Richardson, "Bayesian-Based Iterative Method of Image Restoration," *Journal of the Optical Society of America*, vol. 62, no. 1, pp. 55-59, January 1972.

[65] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The Astronomical Journal*, vol. 79, no. 6, pp. 745-54, June 1974.

[66] H. H. Barrett, T. White, and L. C. Parra, "List-mode likelihood," *Journal of the Optical Society of America A*, vol. 14, pp. 2914-2923, November 1997.

[67] D. Xu, J. A. Fessler, Z. He, and Z. J. Meng, "Estimation of the sensitivity image in the maximum likelihood image reconstruction," Unpublished.

[68] Jeffrey A. Fessler, *Image Reconstruction: Algorithms and Analysis*. Ann Arbor, MI: Unpublished, 2009.

[69] Daniel J. Lingenfelter, Jeffrey A. Fessler, and Zhong He, "Sparsity regularization for image reconstruction with Poisson data," in *Proceedings of SPIE, Computational Imaging VII*, San Jose, CA, 2009, pp. 72460F/1-72460F/10.

[70] Scott J. Wilderman, N. H. Clinthorne, J. A. Fessler, and W. Les Rogers, "List-Mode Maximum Likelihood Reconstruction of Compton Scatter Camera Images in Nuclear Medicine," in *IEEE Nuclear Science Symposium Conference Record*, Toronto, Canada, 1998, pp. 1716-1720.

[71] Tom Hebert, Richard Leahy, and Manbir Singh, "Three-dimensional maximum-likelihood reconstruction for an electronically collimated single-photon-emission imaging system," *Journal of the Optical Society of America A*, vol. 7, no. 7, pp. 1305-1313, July 1990.

[72] Weiyi Wang, "Compton image reconstruction techniques with 3D Position sensitive room temperature CdZnTe detector arrays," University of Michigan, Ann Arbor, MI, PhD Thesis 2011.

[73] Jonathan S. Maltz et al., "Joint Spatio-spectral Compton Camera Reconstruction Algorithm for Arbitrary Detector Geometries," in *IEEE Nuclear Science Symposium - Oral Presentation*, Orlando, Florida, 2009.

[74] S. M. Robinson, W. R. Kaye, J. E. Schweppe, and E. R. Siciliano, "Optimal background attenuation for fielded radiation detection systems," *IEEE Transactions on Nuclear Science*, vol. 54, pp. 1279-1284, August 2007.

[75] S. Agostinelli et al., "GEANT4 – a simulation tooolkit," *Nucl. Instrum. Meth. A*, vol. 506, pp. 250-303, 2003.

[76] Jason M. Jaworski, Christopher G. Wahl, Weiyi Wang, Jeffrey A. Fessler, and Zhong He, "Model-Based Reconstruction of Spectral and Spatial Source Distribution for Objects with Known Motion," In Preparation.

[77] Jason M. Jaworski, Christopher G. Wahl, Weiyi Wang, Jeffrey A. Fessler, and Zhong He, "Model-Based Reconstruction of Spectral and Spatial Source Distribution from Objects with Known Motion," in *IEEE Nuclear Science Symposium Conference Record*, Knoxville, Tennessee, 2010, pp. 1518-1524.

[78] Weiyi Wang, Christopher G. Wahl, and Zhong He, "Maximum Likelihood Estimation Maximization Deconvolution in Spatial and Combined Spatial-Energy Domains for a Detector Array System," in *IEEE Nuclear Science Symposium Conference Record*, Honolulu, Hawaii, 2007, pp. 1965-1970.

[79] Brian D. Yanoff et al., "GE Intelligent Personal Radiation Locator System," in *Proceedings of SPIE Vol. 7306*, Orlando, Florida, 2009, pp. 730616/1-730616/11.

[80] Eric A. Wulf et al., "Compton imager for detection of special nuclear material," *Nuclear Instruments and Methods in Physics Research A*, vol. 579, pp. 371-374, 2007.

[81] K. Vetter et al., "High-sensitivity Compton imaging with position-sensitive Si and Ge detectors," *Nuclear Instruments and Methods in Physics Research A*, vol. 579, pp. 363-366, 2007.

[82] Carolyn E. Seifert and Mitchell J. Myjak, "Directionality in the Gamma Tracker Handheld Radioisotope Identifier," *IEEE Transactions on Nuclear Science*, vol. 56, no. 3, pp. 1351-1355, June 2009.

[83] Andreas Zoglauer et al., "MEGAlib - Simulation and Data Analysis for Low-to-medium-energy Gamma-ray Telescopes," in *Proceedings of SPIE 7011*, Marseille, France, 2008, p. 70113F.

[84] Nokia Corporation. (2011, March) Qt - A cross-platform application and UI framework. [Online]. http://qt.nokia.com/

[85] Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, and Francois Yergeau. (2008) Extensible Markup Language (XML) 1.0 (Fifth Edition). [Online]. http://www.w3.org/TR/2008/REC-xml-20081126/

[86] Leonardo Dagum and Ramesh Menon, "OpenMP: An Industry-Standard API for Shared-Memory Programming," *IEEE Computational Science & Engineering*, vol. 5, no. 1, pp. 46-55, January-March 1998.

[87] Dimitri van Heesch. (2011, March) Doxygen. [Online]. http://www.stack.nl/~dimitri/doxygen/index.html

[88] Alex Kingsbury, "History's Troubling Lessons," *U.S. News & World Report*, February 26, 2007.

[89] Charles A. Lo Presti, Dennis R. Weier, Richard T. Kouzes, and John E. Schweppe, "Baseline suppression of vehicle portal monitor gamma count profiles: A characterization study," *Nuclear Instruments and Methods in Physics Research A*, vol. 562, pp. 281-297, June 2006.

[90] C. G. Wahl, D. Alderson, and L. Pibida, "Studies for software optimization for gross counting portal monitors," *Nuclear Instruments and Methods in Physics Research A*, vol. 574, pp. 185-191, April 2007.

[91] Lloyd A. Currie, "Limits for Qualitative Detection and Quantitative Determination," *Analytical Chemistry*, vol. 40, no. 3, pp. 586-593, March 1968.

[92] Paul E. Fehlau, "Comparing a recursive digital filter with the moving-average and sequential probability-ratio detection methods for SNM portal monitors," *IEEE Transactions on Nuclear Science*, vol. 40, no. 2, pp. 143-146, April 1993.

[93] Tom Burr, James R. Gattiker, Kary Myers, and George Tompkins, "Alarm criteria in radiation portal monitoring," *Applied Radiation and Isotopes*, vol. 65, pp. 569-580, 2007.

[94] Kenneth D. Jarman, Robert C. Runkle, Kevin K. Anderson, and David M. Pfund, "A comparison of simple algorithms for gamma-ray spectrometers in radioactive source search applications," *Applied Radiation and Isotopes*, vol. 66, pp. 362-371, 2008.

[95] D. L. Stephens et al., "Induced temporal signatures for point-source detection," *IEEE Transactions on Nuclear Science*, vol. 52, pp. 1712-1715, October 2005.

[96] K. P. Ziock et al., "Source-search sensitivity of a large-area, coded-aperture, gamma-ray imager," *IEEE Transactions on Nuclear Science*, vol. 53, pp. 1614-1621, June 2006.

[97] ANSI, "American National Standard Evaluation and Performance of Radiation Detection Portal Monitors for Use in Homeland Security," Standard N42.35-2004, 2004.

[98] M. B. Greenfield, A. T. Domondon, N. Okamoto, and I. Watanabe, "Variation in γ-ray count rates as a monitor of precipitation rates, radon concentrations, and tectonic activity," *Journal of Applied Physics*, vol. 91, no. 3, pp. 1628-1633, February 2002.

[99] R. C. Runkle, T. M. Mercier, K. K. Anderson, and D. K. Carlson, "Point source detection and characterization for vehicle radiation portal monitors," *IEEE Transactions on Nuclear Science*, vol. 52, pp. 3020-3025, December 2005.

[100] D. Arnold, M. Blaauw, S. Fazinic, and V.P. Kolotov, "The 2002 IAEA intercomparison of software for low-level γ-ray spectrometry," *Nuclear Instruments and Methods in Physics Research A*, vol. 536, pp. 196-210, January 2005.

[101] Canberra Industries, Inc., Genie 2000 Spectroscopy Software, Version 3.1, 2007.

[102] J. Brutscher, R. Arlt, and K. H. Czock, "Isotope identification software for Gamma spectra taken with CdZnTe detectors," *Nuclear Instruments and Methods in Physics Research A*, vol. 458, pp. 189-195, 2001.

[103] R. T. Reiman, G. P. Lasche, R. L. Coldwell, and J. A. Nobel, "Performance Comparison of a New Approach to Nuclear Spectral Analysis with Commercial Codes," *IEEE Transactions on Nuclear Science*, vol. 47, no. 4, pp. 1522-1527, August 2000.

[104] R. Arlt et al., "Use of CdZnTe Detectors in Hand-Held and Portable Isotope Identifiers to Detect Illicit-Trafficking of Nuclear Material and Radioactive Sources," in *IEEE Nuclear Science Symposium Conference Record*, Lyon , France, 2000, pp. 4/18-4/23.

[105] Markku J. Koskelo, Pertti A. Aarnio, and Jorma T. Routti, "SAMPO80: An accurate gamma spectrum analysis method for microcomputers," *Nuclear Instruments and Methods*, vol. 190, pp. 89-99, 1981.

[106] W. K. Hensley, E. A. Lepel, M. E. Yuly, and K. H. Abel, "Adaption and implementation of the RAYGUN gamma-ray analysis code on the IBM PC," *Journal of Radioanalytical and Nuclear Chemistry*, vol. 124, no. 2, pp. 481-499, 1988.

[107] Josef Uher, Greg Roach, and James Tickner, "Peak fitting and identification software library for high resolution gamma-ray spectra," *Nuclear Instruments and Methods in Physics Research A*, vol. 619, pp. 457-459, 2010.

[108] V. S. Kondrashov, Z. D. Moroz, A. A. Kolyshkin, and Remi Vaillancourt, "A computer program for the identification of nuclides by using median estimates of peak areas in gamma-ray spectra," *Nuclear Instruments and Methods in Physics Research A*, vol. 328, pp. 542-546, 1993.

[109] Dirk Arnold, Menno Blaauw, Stjepko Fazinic, and Vladimir P. Kolotov, "The 2002 IAEA intercomparison of software for low-level γ-ray spectrometry," *Nuclear Instruments and Methods in Physics Research A*, vol. 536, pp. 196-210, 2005.

[110] Sven P. Nielsen and Sigurour E. Palsson, "An intercomparison of software for processing Ge γ-ray spectra," *Nuclear Instruments and Methods in Physics Research A*, vol. 416, pp. 415-424, 1998.

[111] I. De Lotto and A. Ghirardi, "Automatic peak-locating techniques for γ-ray spectra," *Nuclear Instruments and Methods*, vol. 143, pp. 617-620, 1977.

[112] T. Vidmar, A. Likar, M. Korun, B. Vodenik, and K. Kossert, "A novel approach to the analysis of HPGe spectra - Part I," *Nuclear Instruments and Methods in Physics Research A*, vol. 587, pp. 68-75, 2008.

[113] W. W. Black, "Application of correlation techniques to isolate structure in experimental data," *Nuclear Instruments and Methods*, vol. 71, pp. 317-327, 1969.

[114] William R. Russ, "Library correlation nuclide identification algorithm," *Nuclear Instruments and Methods in Physics Research A*, vol. 579, pp. 288-291, 2007.

[115] A. L. Connelly and W. W. Black, "Automatic location and area determination of photopeaks," *Nuclear Instruments and Methods*, vol. 82, pp. 141-148, 1970.

[116] A. Robertson, W. V. Prestwich, and T. J. Kennett, "An automatic peak-extraction technique," *Nuclear Instruments and Methods*, vol. 100, pp. 317-324, 1972.

[117] Mustafa Karakaplan, "Fitting Lorentzian peaks with evolutionary genetic algorithm based on stochastic search procedure," *Analytica Chimica Acta*, vol. 587, pp. 235-239, 2007.

[118] W. Westmeier, "Background subtraction in Ge(Li) gamma-ray spectra," *Nuclear Instruments and Methods*, vol. 180, pp. 205-210, 1981.

[119] M. H. Zhu et al., "Iterative estimation of the background in noisy spectroscopic data," *Nuclear Instruments and Methods in Physics Research A*, vol. 602, pp. 597-599, 2009.

[120] Miroslav Morhac, "An algorithm for determination of peak regions and baseline elimination in spectroscopic data," *Nuclear Instrument and Methods in Physics Research A*, vol. 600, pp. 478-487, 2009.

[121] M. A. Mariscotti, "A method for automatic identification of peaks in the presence of background and its application to spectrum analysis," *Nuclear Instruments and Methods*, vol. 50, pp. 309-320, 1967.

[122] Clair J. Sullivan, Scott E. Garner, Krastan B. Blagoev, and Doug L. Weiss, "Generation of customized wavelets for the analysis of γ-ray spectra," *Nuclear Instruments and Methods in Physics Research A*, vol. 579, pp. 275-278, 2007.

[123] W. S. Murray, K. B. Butterfield, and W. Baird, "Automated Radioisotope Identification Using Fuzzy Logic and Portable CZT Detectors," in *IEEE Nuclear Science Symposium Conference Record*, vol. 2, Lyon, France, 2000, pp. 9/129-9/134.

[124] William S. Murray, Kenneth B. Butterfield, and William Baird, "Handheld CZT Radiation Detector," Patent 6,781,134, August 24, 2004.

[125] Robert J. Estep, Benjamin A. Sapp, and William Russ, "Improved Isotopic Identification for NaI Spectroscopic Portal Monitors Using the Material Basis Set (MBS) Method," in *IEEE Conference on Technologies for Homeland Security*, May 2007, pp. 168-173.

[126] P. Olmos et al., "A New Approach to Automatic Radiation Spectrum Analysis," *IEEE Transactions on Nuclear Science*, vol. 38, no. 4, pp. 971-975, August 1991.

[127] Paul E. Keller, Lars J. Kangas, Gary J. Troyer, Sherif Hashem, and Richard T. Kouzes, "Nuclear Spectral Analysis via Artificial Neural Networks for Waste Handling," *IEEE Transactions on Nuclear Science*, vol. 42, no. 4, pp. 709-715, August 1995.

[128] Lars J. Kangas, Paul E. Keller, Edward R. Siciliano, Richard T. Kouzes, and James H. Ely, "The use of artificial neural networks in PVT-based radiation portal monitors," *Nuclear Instruments and Methods in Physics Research A*, vol. 587, pp. 398-412, 2008.

[129] L. Bouchet, "A comparative study of deconvolution method for gamma-ray spectra," *Astronomy & Astrophysics Supplement Series*, vol. 113, pp. 167-183, 1995.

[130] L. J. Meng and D. Ramsden, "An Inter-comparison of Three Spectral-Deconvolution Algorithms for Gamma-ray Spectroscopy," *IEEE Transactions on Nuclear Science*, vol. 47, no. 4, pp. 1329-1336, August 2000.

[131] Cs. Sukosd, W. Galster, I. Licot, and M. P. Simonart, "Spectrum unfolding in high energy gamma-ray detection with scintillation detectors," *Nuclear Instruments and Methods in Physics Research A*, vol. 355, pp. 552-558, 1995.

[132] M. Moralles, D. A. B. Bonifacio, M. Bottaro, and M. A. G. Pereira, "Monte Carlo and least-squares methods applied in unfolding of X-ray spectra measured with cadmium telluride detectors," *Nuclear Instruments and Methods in Physics Research A*, vol. 580, pp. 270-273, 2007.

[133] Si-guang Wang, Ya-jun Mao, and Pei-jia Tang, "Method of unfolding multiplet regions of X- and g-ray spectra with a detection efficiency constraint avoiding inflecting the peak shapes for correction results," *Nuclear Instruments and Methods in Physics Research A*, vol. 600, pp. 445-452, 2009.

[134] John Mattingly and Dean J. Mitchell, "A Framework for the Solution of Inverse Radiation Transport Problems," in *Nuclear Science Symposium Conference Record*, Dresden, Germany, 2008, pp. 1265-1270.

[135] Dean J. Mitchell, "Variance Estimation for Analysis of Radiation Measurements," Sandia National Laboratories, Albuquerque, NM, Sandia Report SAND2008-2302, 2008.

[136] Klaus P. Ziock et al., "Large area imaging detector for long-range, passive detection of fissile material," *IEEE Transactions on Nuclear Science*, vol. 51, pp. 2238-2244, October 2004.

[137] Ron Wurtz, Klaus-Peter Ziock, Lorenzo Fabris, and Rion Graham, "Comparing Imaging and Non-Imaging Techniques for Reducing Background Clutter and Resolving Distant Point Sources," in *IEEE Nuclear Science Symposium Conference Record*, Fajardo, Puerto Rico, 2005, pp. 338-342.

[138] Dan J. Lingenfelter, Jeff A. Fessler, Clayton D. Scott, and Zhong He, "Benefits of Position-Sensitive Detectors for Radioactive Source Detection," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4473-4483, September 2010.

[139] C. G. Wahl and Z. He, "Sensitivity of gamma-ray source detection using 3D-position-sensitive semiconductor detectors," in *IEEE Nuclear Science Symposium Conference Record*, Dresden, Germany, 2008, pp. 3334-3338.

[140] Carolyn E. Lehner, Zhong He, and Glenn F. Knoll, "Intelligent Gamma-Ray Spectroscopy Using 3-D Position-Sensitive Detectors," *IEEE Transactions on Nuclear Science*, vol. 50, no. 4, pp. 1090-1097, August 2003.

[141] A. J. Sankoh, M. F. Huque, and S. D. Dubey, "Some comments on frequently used multiple endpoint adjustment methods in clinical trials," *Statistics in Medicine*, vol. 16, pp. 2529-2542, 1997.

[142] Christopher G. Wahl and Zhong He, "Point-Source Detection using Energy and Imaging Information from 3D-Position-Sensitive Semiconductor Detectors," *IEEE Transactions on Nuclear Science*, vol. 58, no. 3, pp. 605-613, June 2011.

[143] Jason Jaworski. (2011, March) Private Communication.

[144] S. M. Kay, *Fundamentals of Statistical Signal Processing*. New York: Prentice Hall, 1998.

[145] R. E. Bethel and K. L. Bell, "Maximum likelihood approach to joint array detection/estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, pp. 1060-1072, July 2004.

[146] Steven M. Kay, "The Multifamily Likelihood Ratio Test for Multiple Signal Model Detection," *IEEE Signal Processing Letters*, vol. 12, no. 5, pp. 369-371, May 2005.

[147] Christopher G. Wahl and Zhong He, "Point-Source Detection Using 3D-Position-Sensitive Semiconductor Detectors with Estimated Background," in *IEEE Nuclear Science Symposium Conference Record*, Knoxville, Tennessee, 2010, pp. 1097-1100.

[148] Sebastian Schone et al., "A Common Approach to Image Reconstruction for Different Applications of Compton Cameras," in *IEEE Medical Imaging Conference, Conference Record*, Knoxville, Tennessee, 2010.

[149] David Michael Pfund, Kenneth D. Jarman, Brian D. Milbrath, Scott D. Kiff, and Daniel D. Sidor, "Low Count Anomaly Detection at Large Standoff Distances," *IEEE Transactions on Nuclear Science*, vol. 57, no. 1, pp. 309-316, February 2010.

[150] Carolyn E. Seifert, Debra S. Barnett, and Mitchell J. Myjak, "Prospects for Thermal Neutron Detection & Imaging with the GammaTracker Handheld Radioisotope Identifier," in *Proceedings of SPIE*, San Diego, CA, USA, 2008, pp. 7079/1-7079/7.

[151] Dan Xu, "Muon Detection with 3D Position Sensitive CdZnTe Radiation Detector," University of Michigan, Ann Arbor, MI, Unpublished.

[152] L. J. Schultz et al., "Image reconstruction and material Z discrimination via cosmic ray muon tomography," *Nuclear Instruments and Methods in Physics Research A*, vol. 519, pp. 687-694, 2004.