9-2020

Low-Information Radiation Imaging using Rotating Scatter Mask Systems and Neural Network Algorithms

Robert J. Olesen

Low-Information Radiation Imaging using Rotating Scatter Mask Systems and Neural Network Algorithms

DISSERTATION

Robert J. Olesen, Capt, USAF

AFIT-ENP-DS-20-S-028

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENP-DS-20-S-028

LOW-INFORMATION RADIATION IMAGING USING ROTATING SCATTER MASK SYSTEMS AND NEURAL NETWORK ALGORITHMS

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy in Nuclear Engineering

Robert J. Olesen, B.S., M.S.

Capt, USAF

1 October 2020

AFIT-ENP-DS-20-S-028

LOW-INFORMATION RADIATION IMAGING USING ROTATING SCATTER MASK SYSTEMS AND NEURAL NETWORK ALGORITHMS

DISSERTATION

Robert J. Olesen, B.S., M.S.
Capt, USAF

Committee Membership:

Maj James E. Bevins, PhD
Chairman

Dr. John W. McClory
Member

Dr. Darren E. Holland
Member

Dr. Erik M. Brubaker
Member

Dr. James B. Cole
Member

AFIT-ENP-DS-20-S-028

Abstract

Developing fast, portable, and accurate radiation imagers remains an objective for many nuclear safety and security applications. While recent studies have demonstrated the directional capabilities of the single-detector rotating scatter mask (RSM) system for discrete, dual-particle environments, there has been little progress towards adapting it as a true imaging device. In this work, two algorithms were developed and tested using an RSM mask design previously optimized for directional detection and simulated ^{137}Cs signals from a variety of source distributions. The first, maximum-likelihood expectation-maximization (ML-EM), was shown to generate noisy images, with relatively low accuracy (145% average relative error) and signal-to-noise ratio (0.27) for most source distributions simulated. The second, a novel regenerative neural network (ReGeNN), performed exceptionally well, with significantly higher accuracy (33% average relative error) over all source types compared to ML-EM and drastically improved signal-to-noise ratio (0.85) in the reconstructed images.

This method was experimentally validated using an additively-manufactured mask. Measuring two point and one ring ^{22}Na source distributions, a modified ReGeNN was able to successfully train on simulated noisy signals and accurately predict the relative size and direction of the three measured sources. Training ReGeNN further revealed potential errors caused from overfitting, suggesting future improvement in ReGeNN architecture and training base is needed to obtain accurate activity profiles. To support future design optimizations, a ray tracing algorithm was also developed as an alternative to more rigorous Monte Carlo RSM simulations. This ray tracing code was shown to significantly improve computational efficiency, at a slight cost to the simulated signal accuracy for more complex mask designs.

Dedicated to my friends and family, near and far, new and old, for whom I will always be grateful

Acknowledgements

This work would not have been possible without the support and encouragement from numerous people throughout my time at AFIT. I would be remiss to not recognize my research advisor, Maj James Bevins, first and foremost for not only his advice and expertise within nuclear engineering, but also his continual willingness to support me as a person and officer in the Air Force. Likewise, I would also like to express my gratitude to Dr. Darren Holland, who has joined me on this project since the beginning, whose coding expertise laid the foundation for much of my computational work, and whose shared interest in board games have made for some fun discussions.

I would like to thank the other members of my committee, Dr. Erik Brubaker, Dr. James Cole, and Dr. John McClory for lending their knowledge to this endeavor. This research would not be where it is today without everyone involved, and I am grateful to have had such a supportive, and inquisitive, committee. Lastly, there are many fellow students and graduates who I would like to thank, as they encouraged me not just throughout my studies but entrusted me with valuable leadership experience despite my relatively new career as an officer. The lessons and connections I developed at AFIT will certainly outlast the few short years I stayed here.

Robert J. Olesen

Table of Contents

Page

List of Figures

List of Tables

LOW-INFORMATION RADIATION IMAGING USING ROTATING SCATTER MASK SYSTEMS AND NEURAL NETWORK ALGORITHMS

I. Introduction

In 1995, the International Atomic Energy Association (IAEA) launched the Incident and Trafficking Database (ITDB), allowing States to voluntarily report incidents regarding illicit trafficking and mishandling of special nuclear material (SNM) and other radioactive sources. By the end of 2019, 139 states had reported 3,896 incidents, some which contained fatal sources if exposed for just a few minutes [1]. However, only 290 cases were determined to likely be correlated with illicit trafficking or malicious use (Group I). Incidents for which there was insufficient data to determine intent (Group II) numbered 1,023. Remarkably, the largest number of reported incidents, 2,373, came from cases not related to trafficking or malicious intent (Group III). According to the report:

> The majority of incidents in Group III fall into one of three categories: the unauthorized disposal (e.g. radioactive sources entering the scrap metal industry); unauthorized shipment (e.g. scrap metals contaminated with radioactive material being shipped across international borders); or the discovery of radioactive material (e.g. uncontrolled radioactive sources). The occurrence of such incidents indicates deficiencies in the systems to control, secure, and properly dispose of radioactive material.

Figure 1.1, generated from data provided in the ITDB, highlights these deficiencies. The database includes incidents involving a host of SNM, including highly enriched uranium and plutonium among all three groups and a small number of illicit

interdictions involving kg worth of material. Scrap metal contaminated with enriched uranium has been reported as recently as 2014. In 2013, a 3,000 Curie source was discovered abandoned in a field in Mexico [2]. The vast majority of Group II and III sources originated from industrial and medical facilities. It should also be noted that most of the lowest-activity sources contained no positive confirmation of their recovery.

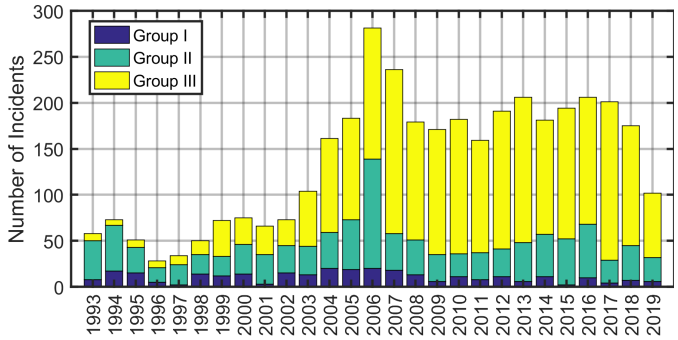


Figure 1.1. Number of incidents reported in the ITDB from 1993-2019 [1].

The data presented underscores the necessity to actively pursue and equip facilities and borders with suitable radiation detectors to either prevent the unauthorized transport of radioactive sources or to interdict the sources before they can cause irreparable harm. The rise in cases between 2003 and 2005 were a direct result of increasing the number of international radiation portal monitors at borders and scrap metal facilities. The United States is committed to support this endeavor, stating that they will "sustain and build upon the roughly 57,000 radiation detectors operating at U.S. seaports, border crossings, and within the American interior" in the 2018 Nuclear Posture Review [3].

This is just one such mission where the rotating scatter mask (RSM) system can be applied. Designed as a portable, directionally-sensitive detector at a fraction of the cost of current deployed systems, the RSM system has many use cases throughout the nuclear security paradigm. Converting the system to a true radiation imaging

system would have significant impacts in the community, narrowing the search space and minimizing health risks when searching for radioactive sources [4].

1.1 Background

Imaging radiation is not a novel concept nor is the prospect of portable imaging devices. In fact, one of the longest standing applications of radiation imaging originates from outside the nuclear security paradigm. Positron emission tomography (PET) and single-photon emission computerized tomography (SPECT) are well-founded techniques in the medical imaging community that utilize trace amounts of radioactive markers to image organ tissues [5]. However, medical imaging and nuclear security have an inverse relationship: in the former, the radiation source is predetermined, well-constrained, and surrounded by detectors; in the latter, the detection system is surrounded by an unknown source in potentially ill-defined environments.

As a result, radiation imagers designed for complex, unknown environments are often much more difficult to implement than their medical counterparts. The bulk of research regarding radiation imaging devices can be divided into two broad categories [6]. The first are coded-aperture devices that use two-dimensional arrays of detectors and a collimating mask to generate an encoded image that can be processed back into the original source distribution [7]. Some coded-apertures can be very large and expensive, containing many radiation detectors and lead masks, limiting their field-of-view (FOV) and portability [8, 9]. Modern advancements in miniaturized active detector arrays and novel system configurations have shown potential in remedying these limitations [10, 11, 12].

The second category of radiation imaging devices are scatter cameras, often known as Compton cameras when referring to gamma imaging [6]. These devices also use an array of detectors, but instead of generating encoded images, the images are back-

projected using well-defined kinematic physics of the interacting particles, which allows these devices to have nearly full 4π FOV and excellent background suppression [13, 14]. Compton cameras are particularly popular, with true portable, commercialized products readily available, albeit at a cost significantly larger than a standard detection system [15]. These devices may also be limited to certain particles and energy regimes, reducing their efficiency. Dual-particle scatter cameras are currently being designed and tested that should improve upon these current limitations if successful [16, 17].

However, a new design concept has arisen as a potential contender within the past decade: time-encoded imaging [18, 19, 20, 21]. In classical terms, these systems can be conceptualized as a standard detector output that is intentionally perturbed over time. If the perturbation is controlled and well-defined, and the source is assumed to be static, then theoretically the source image can be reconstructed. Remarkably, this concept also stems from an outside source: the field of astronomy. In fact, the rotating modulating collimator (RMC), used to image x-rays from distant stars, may very well be considered to be the predecessor to the RSM system [22, 23, 24].

Time-encoded imagers, such as the RSM, may be simple in design, typically containing one or a few number of standard radiation detectors and a mechanism to vary the signal over time (such as rotating an attenuating mask around a detector or rotating a set of detectors). The design minimizes the size and cost of the overall system, as most of the components can be acquired from commercially-available products. What these systems simplify in form factor, they make up for in complexity in data analysis. Whereas coded-apertures and scatter cameras generate two-dimensional signals, time-encoded imagers are one-dimensional. Reconstructing two-dimensional images from one-dimensional signals is nontrivial, requiring potentially complex algorithms and priors that constrain the solution space to realistic distributions, while still be-

ing robust enough to account for small perturbations that is expected in practical application.

This challenge was understood early on in the RSM development, and so most of the research thus far has focused on directional identification of point sources rather than true imaging [25, 26, 27]. The initial RSM demonstrations were promising, with alternative mask designs and analysis algorithms reducing both the system size and directional uncertainty while maintaining a nearly full 4π FOV [28, 29]. The concept has also been extended to neutron point sources [30]. Adapting the RSM system as a true radiation imaging system could have significant impacts in nuclear security and safety applications.

1.2 Problem Description

In regards to nuclear security and safety, a radiation imager must accurately locate and characterize potential sources in numerous, often unknown, environments. Doing so requires that the system assumes as little as possible about the environment and the source; if too much is assumed, then the system will be constrained to a narrow application space and may potentially misidentify, or even fail to detect, a source if the environment deviates too much from its testing environment. To pursue this endeavour, this study suggests that an optimal imager should have:

1. Large FOV to minimize "blind spots".
2. High detection efficiency to minimize time to identification.
3. Large signal-to-noise ratio (SNR) to minimize background and other environmental effects.
4. Particle discrimination and spectroscopy (i.e. high energy resolution) to best characterize potential sources.

5. Portability for rapid deployment and redeployment to meet dynamic mission requirements.

The RSM system was developed as a portable, directionally-sensitive detection system capable of discriminating and localizing neutron and gamma point sources in a nearly full 4π FOV. However, it may be possible to generate true source images using the same physical setup and data collection from the directional identification experiments. This task requires more advanced analysis techniques, as the RSM inherently generates under-determined systems of equations such that many, if not infinite, source images are plausible. A new system must also be developed and tested, as previous RSM prototypes were either not truly portable or not self-contained, requiring active manual assistance to run the equipment.

1.3 Research Objectives

The main objective of this research is to further the development of portable radiation imagers by integrating advanced imaging algorithms and data acquisition software with the RSM system. Previous research regarding the RSM, including system design and optimization, was conducted for the point source direction identification scenario. Therefore, the individual research objectives (highlighted in Table 1.1) constitute a branch in system development from the original RSM concept by developing a more computationally-efficient RSM simulation model, developing and analyzing advanced imaging algorithms, and designing and testing a new portable RSM system.

Table 1.1. Research Objectives

Objective	Chapters
I. Ray Tracing Simulations	3
II. Algorithm Development	4-5
III. Experimental Testing	6

The first objective supports future mask geometry optimization studies, developing surrogate radiation transport models using ray tracing to reduce the computation cost. The second objective explores algorithms suitable for the RSM system that enable imaging using the same system outputs as the directional identification platform. The two algorithms chosen are the maximum-likelihood expectation-maximization (ML-EM), due to its popularity among the radiation imaging community, and a novel neural network-based approach, due to recent advancements in machine learning techniques and general success in image-related problems. Finally, the last objective experimentally tests the RSM system using a new design developed for imaging and the algorithms developed in the second objective by measuring distributed gamma sources, demonstrating the RSM as a practical, portable gamma imager for the first time.

1.4 Assumptions & Limitations

While broadening the RSM application space towards distributed sources, this study is limited to environments available within the laboratory setting. As such, the effects from environmental background, shielding, and scattering may impose additional challenges throughout the analysis. The results presented assume that the number of detected events is large enough to maximize the SNR and reduce the impact from these environmental effects. The validity of this assumption will largely be dependent on expected source activities, active detection time, and surrounding environments for practical applications.

While previous studies have verified that the RSM is capable of localizing a neutron point source, this research focuses solely on gamma imaging. In order to generate a detector response library for algorithm development, MCNP (v6.1.4 and 6.2) was used to simulate a ^{137}Cs point source using a monoenergetic 662 keV photon beam.

The model would later be updated to simulate the experimental setup and ^{22}Na source with a monoenergetic 511 keV photon beam. The simulations used a simplified RSM model, excluding a majority of the mechanical and electrical components outside, and neglected any background effects. The detector responses were generated using the full-energy peak from the simulated detector spectra, with an energy deposited pulse height tally modeling the physical detector response. This study does not investigate any incident photon energy dependence, although the attenuation and detection probabilities are certainly gamma energy dependent.

Finally, distributed sources were modelled as a weighted collection of point sources to improve computational efficiency. It is expected that truly distributed sources will have slight variances in the detector responses that are not represented by this assumption; as a proof-of-concept study, it is assumed that these variations can be accounted for given a more specific application space. The distributions simulated and tested were also generally localized and monotonic. That is, these results do not test RSM imaging capabilities for radiation fields that have relatively large spatial extent and varying activity.

II. Theory

The fundamental component of the radiation imager is the simple radiation detector. Despite the prolific number of detectors available, detecting radiation is no trivial task. The probability and type of interaction is highly dependent on the specific form of radiation, the interacting particle's energy, and the detecting medium. For example, Figure 2.1 plots the gamma mass attenuation coefficient as a function of photon energy for NaI(Tl), a common inorganic photon scintillating detector material, for the three most common photon interactions: photoelectric absorption, Compton scattering, and pair production [31]. This type of detector emits light at a very specific wavelength when the photon interacts in the detector crystal; the amount of light produced is directly proportional to the deposited energy. This light is converted into an electrical pulse, also proportional to the deposited photon energy, which can then be recorded to create an energy spectrum.

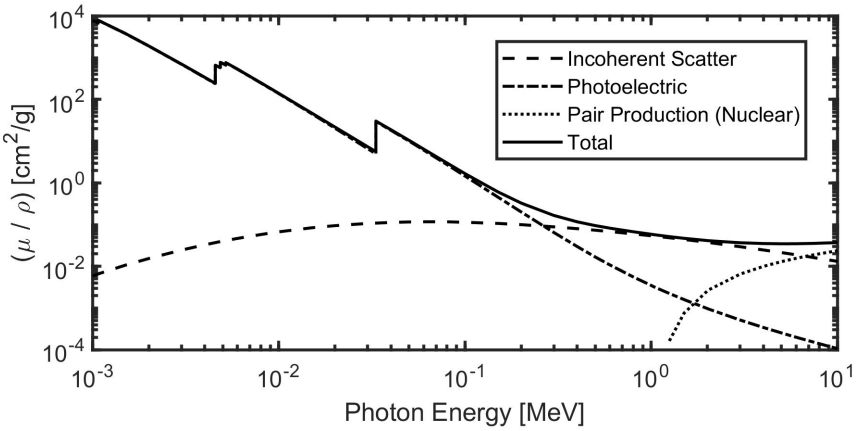


Figure 2.1. Mass attenuation coefficient vs photon energy for homogeneous NaI [31].

Figure 2.2 shows an energy spectrum for a ^{137}Cs source, a monoenergetic 662 keV gamma emitter. The large peak centered around 662 keV is known as the full-energy peak (FEP) or photopeak, created by the photon depositing all of its energy to the crystal through photoelectric absorption. The plateau-region in the lower energies is

caused by Compton scattering within the crystal, where the incident photon deposits only part of its energy in the crystal and scatters away at a lower energy. The energy deposited in Compton scattering follows a well-known distribution, with a maximum possible energy deposited forming the sharp drop-off known as the Compton edge. For a ^{137}Cs 662 keV photon, the maximum deposited energy is approximately 477 keV, which forms the Compton edge as shown in the energy spectrum.

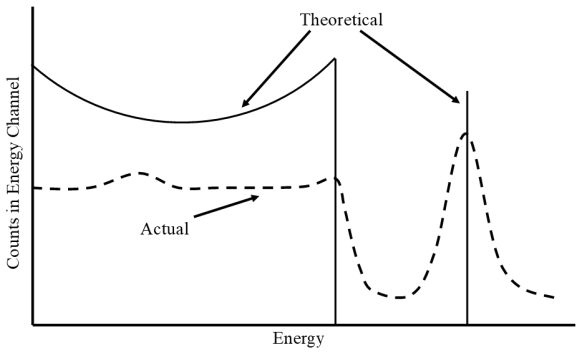


Figure 2.2. Example ^{137}Cs spectrum from a NaI(Tl) inorganic scintillating detector.

The 662 keV photon energy is below the 1022 keV pair production threshold (from energy-mass conservation). However, even for sources such as ^{60}Co, a common medical and industrial source that emits 1173 and 1333 keV gammas, most of the detected events are within the Compton regime and FEP. Other peaks may be present that are dependent on the specific source and surrounding environment, such as a backscatter peak caused from photons scattering back into the detector and a 32 keV barium x-ray peak, a daughter product of the initial ^{137}Cs beta decay.

Note that the peaks and Compton edge are not exact: the scintillation and electrical conversion are stochastic processes following a normal distribution. The spread of the distribution compared to the energy deposited is typically used to define the detector's energy resolution. While inorganic scintillators have reasonably good detection efficiencies (i.e. large mass attenuation coefficients), they are also subject to lower energy resolutions compared to semiconductor detectors.

2.1 Gamma Radiation Imaging

A brief history of gamma imaging devices and current technology is discussed in this section. It should be noted that this literature review is by no means comprehensive, as the number of systems that have been theorized, tested, and/or deployed would merit years of dedicating research and documentation beyond the scope of this work. Instead, this discussion is meant to serve as an introduction to the fundamental theories that most gamma imagers have been built upon. Within the context of this study, an imager was defined to be a system capable of reproducing a two-dimensional source image; although there certainly exist applications where one-dimensional, directionally-sensitive systems are applicable.

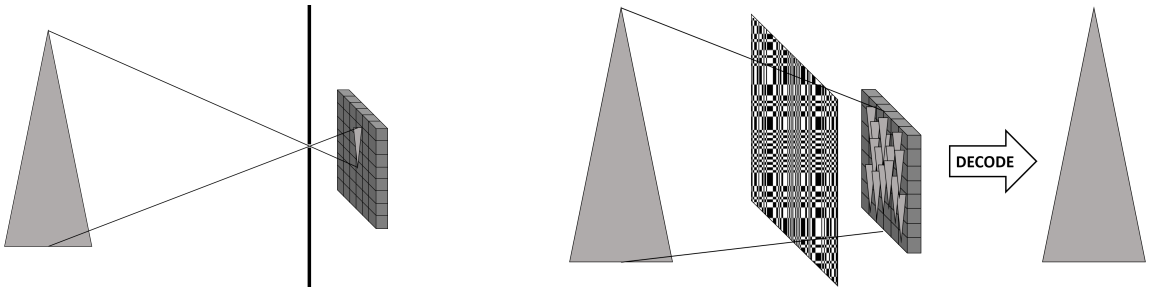
Throughout this discussion, it will be important to recall the five features of the ideal imager previously mentioned: 1) large FOV; 2) high efficiency; 3) large SNR; 4) particle discrimination and high energy resolution; and 5) portability. In reality, these parameters cannot be treated independently, as improvements in one are often obtained at the expense of others. For example, large detector volumes improve detection efficiency and FOV but at a significant cost to portability, SNR, and (potentially) spatial resolution. Most systems, therefore, excel in some areas while compromising in others, and the user must consider the specific application for which the device is to be used. For this reason, the benefits and limitations of each system are highlighted at the end of each section.

2.1.1 Coded Apertures

Coded apertures are perhaps the most similar to optical cameras in that they project the source image onto a 2D array of detectors, each detector acting as an individual "pixel". While optical cameras use lenses and mirrors to focus the light onto the array, no such devices readily exist for x-rays and gamma rays. Instead, an

aperture apparatus serves as a spatially modulating collimator to project the image onto the detector array.

Figure 2.3 illustrates the fundamental concepts of coded aperture imaging. The most basic design is the pinhole collimator, which functions almost exactly the same as a pinhole camera, generating a mirrored image of the source onto the detector array. This configuration requires that most of the particles be blocked by the collimator, significantly reducing the efficiency [7]. Coded aperture masks contain many pinholes which increase both the detection efficiency and SNR [32, 33]. However, the projection onto the detector array is no longer a single image but a combination of projections through each collimator. Complex algorithms must be used to reconstruct the original source image, such as filtered back-projections or maximum likelihood estimations.



(a) Pinhole Collimator (b) Coded Aperture

Figure 2.3. Schematic of the (a) basic pinhole collimator and (b) coded aperture concepts.

Coded apertures were first proposed in 1968 as a method for x-ray astronomy [33], but the principles have since been applied to gamma imaging [7, 34, 35]. A limited FOV is a consistent disadvantage among coded apertures; large arrays [8, 9] or multiple coded apertures are combined to improve the FOV [10, 36]. While the larger detector volumes also increase efficiency, systems of such scale are impractical for applications where portability is critical. The strong dependence of mask thickness and detector size on image reconstruction accuracy further limits scalability [11, 35]. Novel mask designs and reconstruction algorithms are being explored to help solve

these issues, while also enabling dual-particle imaging [12, 37].

2.1.2 Scatter Cameras

Unlike coded apertures, which uses arrays of detectors to generate an encoded image, scatter cameras, also known as Compton cameras when referring specifically to gamma imaging, takes advantage of the well-defined kinematics governing scattering events. Compton cameras consist of two more detector arrays working in coincidence. Figure 2.4 illustrates a basic two-array set-up and how it works. First, a photon with energy E_γ undergoes Compton scattering in one of the arrays, depositing some of its energy and scattering away at some angle, θ, with a lower energy, E'_γ. This scattered photon is then absorbed in the second array, depositing all of its energy and creating a second signal coinciding with the scattered event from the first detector.

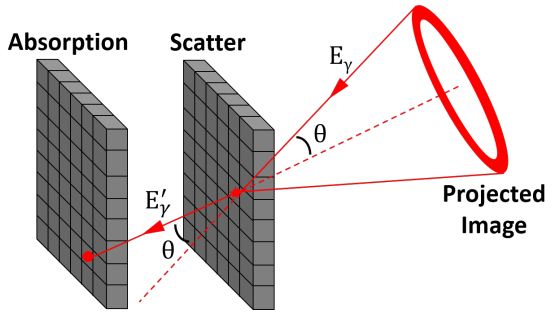


Figure 2.4. Schematic illustrating a Compton scatter camera.

The Compton scattering equation is given by

$$E'_\gamma = \frac{E_\gamma}{1 + \frac{E_\gamma}{m_e c^2}\left(1 - \cos(\theta)\right)}, \tag{2.1}$$

where m_e is the electron rest mass and c is the speed of light. The scattered photon energy is measured from the second detector, while the original photon energy can be recovered through the total energy deposited in the two detectors. Thus, the original scattering angle can be calculated. However, the reaction is isotropic about

the original direction. Therefore, calculating θ actually yields a cone of possible directions that the photon originated. Uncertainties in the energy deposition and interaction location spreads the cone out into a "ring" of potential directions.

These rings can be plotted in a 2D histogram which ideally converges to the original source image. The Compton telescope (COMPTEL) developed for the National Aeronautics and Space Administration (NASA) was one of the first successful applications of the Compton camera, capable of imaging gammas from 1-30 MeV [22]. As the detectors can accept gammas from all directions, modern gamma scatter cameras are capable of a nearly 4π FOV and have excellent SNR [13, 14]. Commercial Compton cameras have been produced [15], targeting the need for accurate imagers in monitoring and decommissioning of nuclear facilities.

Compton cameras are among the most portable imaging systems as they do not require any collimating or attenuating material to modulate the signal. Scatter cameras are also readily adapted towards neutron imaging, as the imaging itself is guided by particle kinematics. Dual-particle imaging becomes viable by using neutron sensitive detectors. This is significant as most special nuclear material are gamma and neutron emitters; imaging both particles would significantly increase the confidence of detection. Numerous studies are investigating various detectors and configurations of possible dual-particle scatter cameras [16, 17, 38].

It should be noted that such systems can be rather complex, due to the precise timing required to conduct coincidence measurements and the difficulty regarding neutron detection in general. Scatter cameras can also suffer from relatively low efficiencies, relying on multiple scatter events (and thus highly particle-energy dependant) and position-sensitive detector arrays. Such complexity makes these devices among the most expensive of the imaging systems.

2.1.3 Time-Encoders

This study classifies time-encoding imagers as any system that transforms and stores the source's spatial information in a temporal signal. Among the earliest adopters of time-encoding is the RMC, first conceptualized in the 1960s and later integrated onto the Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI) satellite, becoming the first instrument to image gammas from a solar flare [39, 40]. Figure 2.5 illustrates a basic RMC configuration.

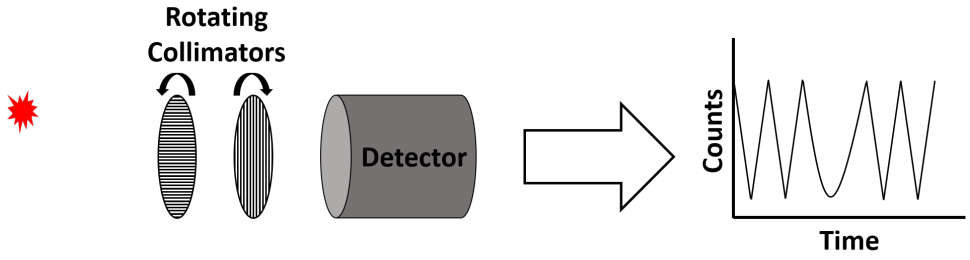


Figure 2.5. Schematic illustration of a standard RMC set-up.

An RMC is typically a set of two linearly-spaced collimating grids placed in front of a spatially-insensitive detector. Rotating the grids produces the temporal modulation in the signal. Reconstructing the source image is equivalent to similar Fourier inverse transforms in radio interferometry [6], with back-projection and iterative algorithms readily available.

A number of RMCs have been developed for more terrestrial applications regarding nuclear security [23, 24, 41]. Similar time-encoded imagers have been proposed and tested using alternative collimating mask configurations. While some have proposed devices that mechanically alter a coded aperture to impart a temporal modulation [21], most recent time-encoders have focused on enlarging the FOV by using a cylindrical, rotating collimator, as shown in Figure 2.6 [42]. Such devices have shown improvement in imaging point and extended sources over RMC systems, as well as being capable of dual-particle imaging [18, 20].

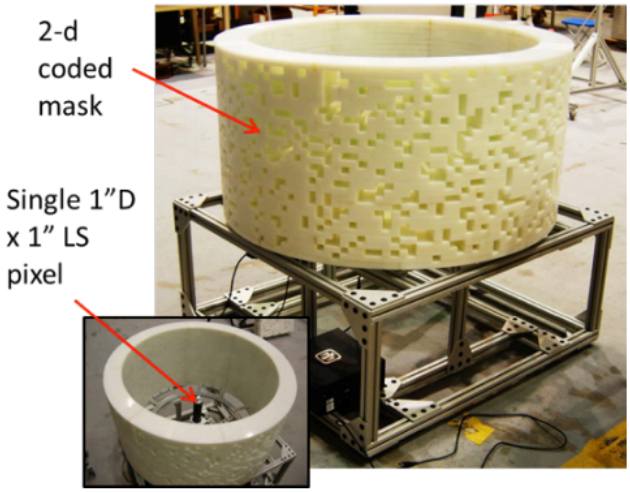


Figure 2.6. A cylindrical rotating collimator for time-encoded imaging [42].

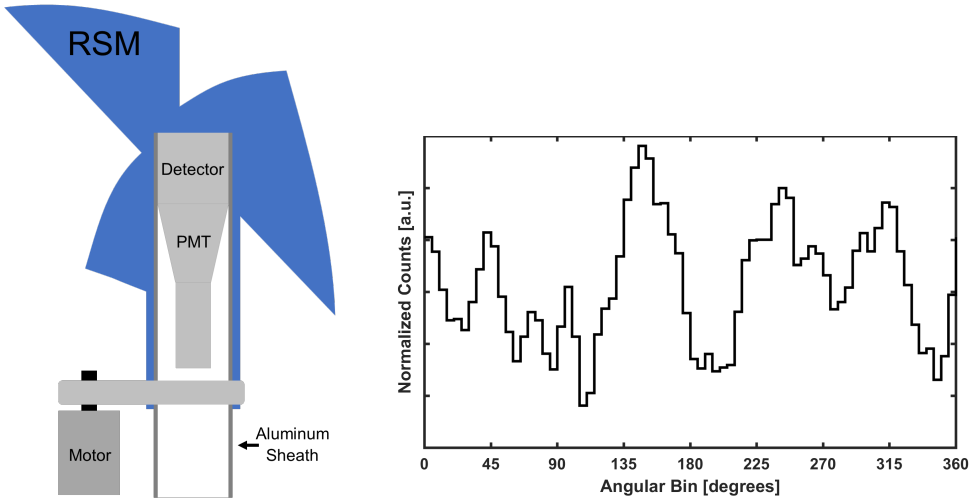
The largest advantage of these systems, compared to the imagers previously discussed, is their relative simplicity, foregoing the complex detector arrays and instead using simple, position insensitive detectors. The spatial reconstruction resolution is not as constrained to detector size, as opposed to coded apertures and scatter cameras where the angular resolution is closely related to the size and spacing of the detector array. Instead, the angular resolution depends on the collimator design and sampling rate, allowing separate optimization for both angular (collimator-driven) and energy (detector-driven) resolution. Most of the systems also require relatively few counts, improving the efficiency and decreasing the system uncertainty.

However, many of the time-encoded systems remain proof-of-concepts and recent studies indicate a lower limit on scalability for the rotating cylindrical collimators due to edge effects [19]. RMCs and time-modulated coded apertures also suffer from relatively poor FOV and efficiencies. The cylindrical collimators are improvements, although scatter cameras remain the best with near full-field imaging. More advanced reconstruction algorithms are being studied to help improve these systems, as most also contain artifacts in the reconstructions for the more extended source images.

2.2 Rotating Scatter Masks

The RSM replaces the collimator of other time-encoders with an *attenuating* mask of spatially-varying thickness. Similar to cylindrical rotating collimators, the mask surrounds a single, position-insensitive detector. By completely encasing the detector, the overall system can be reduced in size while theoretically increasing the FOV to nearly 4π steradians.

Figure 2.7 illustrates the basic components of an RSM system. Particles passing through the mask will be attenuated as they are scattered or absorbed before reaching the active detector volume. Rotating the mask generates a time-encoded detector response curve (DRC) similar to the signal from an RMC or other time-encoded imagers. Since the time-encoding is periodic, the DRC is typically expressed as a function of mask rotation angle over one full period.



(a) (b)

Figure 2.7. Schematic illustration of the primary components for (a) an RSM system and (b) an example DRC.

A detector response matrix (DRM) is formed from a collection of DRCs from point sources spanning a set of possible directions. The DRM can be calculated experimentally or, more easily, computationally through radiation transport simula-

tions. Assuming that the point source is stationary, the direction of a point source can be determined by comparing an experimentally measured DRC to the entire set within the DRM, using a metric such as the reduced chi-squared or modal assurance criterion to compare the DRCs [25, 28, 29].

Most of the RSM research thus far followed this approach, focusing exclusively on point source direction identification rather than true imaging. Multiple studies have been conducted that showcase the success using a single point source [25, 27], multiple point sources with different photon energies [26], and even extending the application to neutron point sources [30]. Imaging was explored during the initial development of the RSM, with iterative algorithms such as ML-EM recreating extended sources with reasonable accuracy. However, the large computational power required at the time for these algorithms greatly limited the system's practicality (along with the general lack of portability of the then-current designs) [25].

The RSM shares similar limitations and difficulties in reconstructing the source image as other time-encoded imagers. An ideal DRM would consist of a completely orthogonal set of DRCs over all potential source directions, so that each direction produces a completely "unique" DRC. This would minimize the uncertainty caused from statistical noise, reducing the chance of misidentifying the source direction. Now consider the actual system: the DRC is periodic in time, repeating after each rotation of the mask. Therefore, sources along the same polar angle have the same periodic DRC, just phase-shifted based on their "starting position" relative to the mask. This correlation among polar-equivalent source directions causes a dilemma.

As a hypothetical, consider two DRCs generated from two different source directions, the first forming a sine wave and the second forming a cosine wave, both with the same period and amplitude. Mathematically, the DRCs are orthogonal and would normally be considered independent and unique. However, sine and cosine are phase-

shifts of each other (by $\pi/2$ radians). Thus, the DRC of one also corresponds *exactly* to the DRC of the other, just phase-shifted by a different rotational starting position. Instead of being orthogonal, the DRCs are actually indistinguishable from each other, forming a degenerate DRC that corresponds to two different source directions. Direct inverse algorithms, such as inverse Fourier transforms and filtered back-projections, would prove incapable of properly reconstructing images.

Early studies concluded that the RSM's degenerate nature is physically and mathematically unavoidable [28]. Thus, research focused on reducing the degeneracy as much as possible, resulting in geometrically-complex mask designs through rigorous optimization procedures. Another study approached the problem by changing how the source direction is determined from the DRC by eliminating the need to produce a DRM altogether [29]. This method, instead, used features within the DRC to extract the source direction. By carefully controlling the mask design, these features could be fitted to a known function relating them back to the original source direction (similar in concept to how scatter cameras calculate the photon direction). The masks, termed Spartan RSMs, were designed such that any degeneracy resulted in uncertainty centered about the true source direction, rather than permitting multiple source directions to be possible.

Such a method resulted in significantly smaller mask designs, improving the system's detection efficiency and portability. However, the Spartan designs were only suitable for a single point source. The analysis method would become invalid for multiple point sources emitting the same photon energies, as the features could not be accurately extracted from the DRC. Thus, in order to conduct imaging, this study referred back to the original optimization study [28], using the RSM designs that minimized degeneracies to theoretically maximize the imaging performance. However, this required new imaging algorithms to be developed as creating a database of

all potential source distributions is impractical and limited.

2.3 Imaging Algorithms

In an RSM reconstruction, each pixel is defined as an independent source of particles. The total response, then, can be modeled as a linear combination of DRCs. Reconstructing the source image from a measured DRC is a matter of solving the $\mathbf{P}\boldsymbol{\lambda} = \mathbf{y}$ inverse problem, where $\mathbf{y}$ is the measured DRC, $\boldsymbol{\lambda}$ is the image in vector-form to be reconstructed, and $\mathbf{P}$ is the system response defined by the DRM. Specifically $p_{i,j}$ is the probability that a particle emitted from pixel j, with mean activity λ_j, is measured at mask rotational angle i.

The solution is not as trivial as setting $\boldsymbol{\lambda} = \mathbf{P}^{-1}\mathbf{y}$, as $\mathbf{P}$ can be ill-conditioned or even under-determined (although this can be *a* solution, it may not be the *only* solution). A solution is unique if and only if $\mathbf{P}$ has full column rank.

Now consider a DRC generated from n samples; the DRC has a total of n corresponding pixel directions, separated by a phase-shift in the rotational angle in $360°/n$ increments. For a single polar angle, there are n equations with n unknowns, allowing the inverse to be solved directly. If there are m different polar directions to be solved, each with n phase-shifted DRCs, then the system increases to $m * n$ equations for the same n-dimensional DRC. The system is under-determined, with $n\,(m-1)$ degrees of freedom. Since a solution *must* exist (the DRC was measured from some image), but $\mathbf{P}$ cannot have full column rank [28], the system is indeterminate with an infinite number of solutions.

Therefore, additional constraints are required to solve the system of equations and reconstruct the original source image. Using $\boldsymbol{\lambda} = \mathbf{P}^{+}\mathbf{y}$, where $\mathbf{P}^{+}$ is the Moore-Penrose inverse of $\mathbf{P}$, yields a least squares solution. However, this may not be the most "realistic" solution as such reconstructions tend to be noisy, whereas reality

dictates that most (artificial) sources are localized and relatively uniform.

2.3.1 Maximum-Likelihood Expectation-Maximization

First developed in 1976, ML-EM is an iterative algorithm that finds maximum likelihood estimates from observations that can be viewed as incomplete data of the overall system [43]. In other words, ML-EM finds the solution that maximizes $P(\mathbf{y}|\boldsymbol{\lambda})$: the probability of observing $\mathbf{y}$ given the estimated solution parameters (reconstruction) and a known probability distribution that the events are observed.

Deriving the ML-EM algorithm for the RSM system follows the original outline presented by Shepp and Vardi for radiation emissions [44]. The resulting equation (derived in Section 4.A) is a compact form that is remarkably similar to those used for positron emission tomography:

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\sum\limits_i p_{ij}} \sum_i \frac{y_i p_{ij}}{\sum\limits_k \lambda_k^m p_{ik}}, \tag{2.2}$$

where the m superscript denotes the iteration index.

Since this algorithm is effectively equivalent to the one derived by Shepp and Vardi [44], it maintains the same characteristics that:

- The log likelihood is concave, guaranteeing that the maximization step does indeed solve for a local maximum.
- The log likelihood strictly increases after each iteration and will, therefore, always converge to a solution.
- The total number of particles is conserved between each iteration. That is, $\sum\limits_{i,j} \lambda_j^{m+1} p_{ij} = \sum\limits_{i,j} \lambda_j^m p_{ij} = \sum\limits_i y_i$.

However, it is important to note that ML-EM does not guarantee a solution for the

global maximum and is only guaranteed to converge to one of potentially many local maxima.

Equation 2.2 assumes nothing about the actual shape of the reconstructed image. Many studies have been conducted on implementing Bayesian priors that constrain the algorithm to specific classes of images, often implementing smoothing functions that limit the solution to a localized distribution [45]. These methods introduce the prior to the likelihood function, $L(\boldsymbol{\lambda})$, usually taking the form

$$L(\boldsymbol{\lambda}) = \prod_{i,j} e^{-\lambda_{ij}} \frac{\lambda_{ij}^{n_{ij}}}{n_{ij}!} e^{-V(\lambda_j)}, \tag{2.3}$$

where $V(\lambda_j)$ describes some potential function that compares each pixel to its local surroundings.

The Median Root Prior (MRP) is a method that uses a Gaussian potential function

$$V(\lambda_j) = \beta \frac{(\lambda_j - M_j)^2}{2M_j}, \tag{2.4}$$

where M_j is the median of the pixel values in a neighborhood surrounding λ_j and β is a weighting term ranging from zero to one. The advantage of comparing a pixel to the median, rather than a value such as the average, is that this solution preserves sharp edges rather than oversmoothing the image. Previous studies have shown that using MRP in conjunction with ML-EM constrains the solutions to locally monotonic sources: ideal for realistic radioactive source distributions [46, 47].

The algorithm including the prior is remarkably similar to before:

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\sum_i p_{ij} + \beta \frac{\lambda_j^m - M_j}{M_j}} \sum_i \frac{p_{ij} y_i}{\sum_k \lambda_k^m p_{ik}}. \tag{2.5}$$

The amount of detail recoverable in the image is governed by the neighborhood size

used to calculate the median. Any feature size smaller than this neighborhood is treated as noise and consequently penalized.

2.3.2 Neural Networks

Other priors may be integrated into ML-EM to improve performance, but quantifying what is "realistic" or "man-made" is often difficult to put in mathematical form. Artificial neural networks have risen almost precisely for this scenario through the application of what is known as the universal approximation theorem.

This theorem states that multi-layer feed-forward networks, such as convolutional neural networks (CNNs), with locally-bounded piecewise continuous activation functions can approximate any continuous function to any degree of accuracy, as long as the activation function is not a polynomial [48]. This means that an artificial neural network can be trained to approximate the distribution pertaining to "realistic" sources without knowing the form of that distribution explicitly (the approximation is instead dictated by the given training data). A well-designed neural network integrates this into a single network that both learns the distribution and how to process the data.

Of course, one criticism of this method is that the distribution is hidden within the network, earning the oft-quoted "black box" label. Another limitation, which is important to note, is that while the universal theorem states that these networks can approximate any continuous function, developing the appropriate network that can do so is not guaranteed. As the task becomes more complex, requiring more complex architectures, the networks become extremely sensitive to hyper-parameters within the model and can become unstable or subject to over-fitting to the training data rather than approaching a generalized solution.

Despite these criticisms and limitations, many applications have already made use

of artificial neural networks. They have been applied to gamma and neutron spectroscopy [49, 50], and recent studies have successfully integrated them into coded-aperture systems [51, 52]. This work chose to study CNNs due to their rising popularity as the state-of-the-art for image deconvolution and generation [53, 54].

III. Monte-Carlo Radiation Transport Simulations

Designing, manufacturing, and experimentally testing each RSM design iteration with every source distribution is impractical. This method is prohibitively inefficient, particularly for neural network training, which would require thousands of experimental measurements with source distributions that are difficult to create. Instead, the system was modeled using the Monte-Carlo N-Particle (MCNP) radiation transport code developed by Los Alamos National Laboratory (LANL) [55]. Using the code significantly reduced the time to generate the full DRM and allowed multiple designs to be tested in quick succession. Simulating the transport also removed any systematic errors that may be present in the laboratory environment (e.g. scattering, background, and electronic noise) so that the DRM could be generated with high precision and low uncertainty. This method also neglects any source positioning uncertainties during the experimental measurements of the actual source distributions.

As the name implies, MCNP is a Monte-Carlo (i.e. stochastic) code that simulates the probabilistic nature of radiation interaction with material. The lifetime of a single source particle is recorded in a history by sampling various probability distribution functions (PDFs) that govern the initial source trajectory, interaction path length and type, and products following interaction. Some PDFs, such as the initial source distributions and trajectories, are defined by the user. Most interaction PDFs are defined from a cross-sectional library, using the material compositions provided by the user. Following the law of large numbers, the code tracks many source particle histories to form expected distributions of the system's behavior.

In the initial ML-EM imaging algorithm development, MCNP v6.1.4 (an unreleased beta version containing vital bug fixes) was used to model the radiation transport and detection [56]. However, LANL released an updated version, MCNP v6.2, during this time [55]. To maintain consistency and comparable data for the same

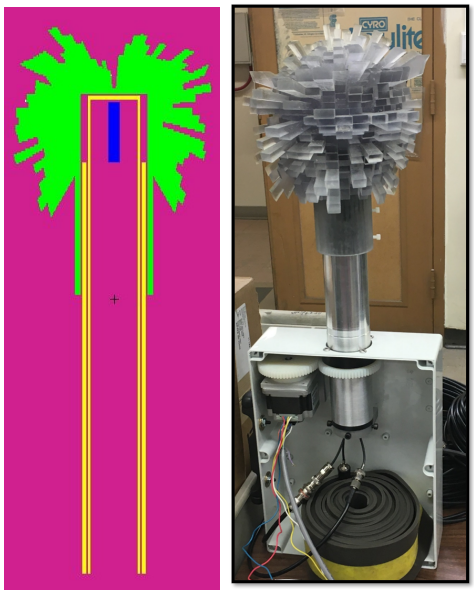
RSM design, the CNN study used the DRMs from MCNP v6.1.4 used in the previous ML-EM study. MCNP v6.2 was used in the later experimental work, as new simulations were already required due to a newer mask design (a scaled version of the "Mace" used in the ML-EM and neutral network development work) as well as a different detector crystal ($LaBr_3$ vs NaI) used for the experiments.

3.1 System Modeling

To improve computational efficiency, a simplified RSM model was used in the MCNP simulations (Figure 3.1a). A majority of the electrical and mechanical components required to rotate the mask were excluded in the simulation model, including the detector's integrated PMT. As shown in Figure 3.1b, the physical system was designed so that these components were within the RSM's "blind spot", a region where no mask material exists to modulate the signal. Excluding them from the model does not impact the results for FEP-based DRCs as they are outside the system's FOV. The masks were designed with a blind spot for this exact purpose, without significantly reducing the system's FOV.

The mask was modeled using homogeneous poly(methyl methacrylate) (PMMA), with 1.18 g/cc density. This composition was assumed to be approximately equal to the material used during the 3D manufacturing process (VeroClear^{TM}), as stated by the manufacturer, with the exact feed material composition being proprietary. This assumption would not impact the algorithm development; experimental measurements have demonstrated this approximation is acceptable [26, 27]. Due to the complex mask geometry, an unstructured mesh was used to create the mask geometry within MCNP (which necessitated using v6.1.4 and above).

For the algorithm development, the detector was modeled with a homogeneous mixture of NaI (the Tl in actual detectors is introduced as a low-level dopant and



(a) (b)

Figure 3.1. Comparison of (a) the RSM MCNP model and (b) the actual RSM system. For the MCNP model, only the PMMA mask (green), aluminum shells (yellow), and detector crystal (blue) were modeled.

does not measurably impact the crystal's cross-section), with 3.667 g/cc density. The actual shape of the detector crystal, a 3" × 3" right-circular cylinder, was used in the model. This was later changed to a 10 mm × 10 mm × 50.8 mm rectangular $LaBr_3$ detector (another inorganic scintillator) for the experimental measurements, and is the detector shown in Figure 3.1a.

Two aluminum shells were included to account for the rotating and stationary sleeves housing the detector. While these tubes would decrease the number of counts detected in the FEP, their rotational-symmetry means that they do not significantly affect the actual shape of the DRC and could have been excluded (a few mm of aluminum is relatively negligible at the photon energies being measured). However, they were included because they provided realism to the simulation without significantly increasing the computation time. For the same reason, a standard atmospheric composition was used as the surrounding medium.

3.2 Source Modeling

The DRM was generated using FEP events; only particles with a direct line-of-sight to the detector could contribute to the DRM. Otherwise, the particle would have to be scattered first before it can be absorbed by the detector. Every scattering event (assuming negligible coherent scattering) after the photon is emitted subsequently reduces the photon's energy and would therefore produce a signal in a lower energy channel than the FEP. This is another justification of excluding the mechanical and electrical components, as any directional dependence from these components would be counted outside the FEP region.

Additionally, constraining to the FEP-only events allows the computational efficiency to be further improved by using a geometric variance reduction technique restricting the initial source trajectories. Emissions from decay of radioactive materials are isotropic, with equal probability of being emitted in any direction. Using an isotropic emission in the simulation is inherently inefficient, as most of the particles would never interact with the detector at all, and only particles with a direct line-of-sight could contribute to the FEP counts. Therefore, for the MCNP simulations, the initial source directions were sampled uniformly in a cone that just covered the extent of the detector.

MCNP does not simulate the scintillation process and electrical conversion that occurs in physical systems. Instead, the detector response was modeled using a pulse-height tally. This tally records the total energy deposited in the active detector volume from an event and bins it into an energy histogram. Using this tally to simulate the detector response assumes that the signal is deposited locally (i.e. no secondary particles escape and all scintillation light is collected) and that the detector signal is proportional to the energy deposited by the photon. The former is a reasonable assumption for most commercial inorganic scintillators, and the latter was discussed

in Chapter 2. The tally does take into account multiple scatter events, whereby a photon can deposit its energy through multiple interactions in the crystal. As these events occur within picoseconds of each other, which is much faster than the timing resolution for most scintillators, a single pulse is formed equal to a pulse generated from a single event of equivalent gamma-ray total energy deposition.

Figure 3.2 shows an energy pulse-height tally of the RSM model, using a monoenergetic 662 keV photon source to simulate ^{137}Cs. The source was located 86.36 cm from the center of the detector and the sampling of the initial source direction was confined to the variance reduction cone previously discussed (the specific tally shown is for an arbitrary direction). The histogram is an approximation of the ideal energy spectrum that would be observed from the detector, with the counts normalized to the total number of simulated particles.

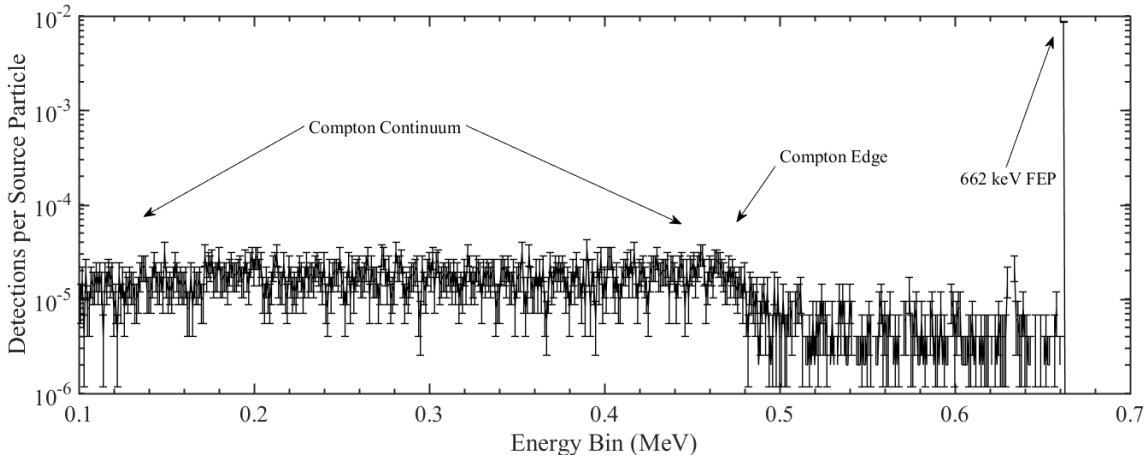


Figure 3.2. MCNP pulse-height tally of the simulated RSM system for a monoenergetic 662 keV source with a homogeneous 3" NaI detector.

The uncertainty from MCNP is given by the square root of the number of counts for each, as dictated by Poisson statistics. For the model, 500,000 source particles was found to reduce the relative uncertainty in the FEP bin to a maximum 2% over all simulated directions. It should be noted that the figure is not representative of the true spectrum one would observe in an actual RSM measurement, as not only

does this model exclude photons that scatter into the detector but it also neglects the stochastic nature of the scintillation and electrical conversion that result in broadening of the observed features. However, this model is sufficient for this study as long as the analysis is constrained to the FEP.

3.3 Ray Tracing Radiation Transport

The Mace RSM design was derived from an optimization study using full-scale, 20 cm masks with the intent that they would function as purely directional detection systems. As the algorithms for imaging are much different than those for directional detection, there is no guarantee that the results from those specific mask design studies would apply directly to imaging as well. The masks have also since been scaled down to be more portable, and, due to the non-linearity between pathlength and attenuation, the same design will likely have reduced SNR at smaller sizes, potentially further reducing the imaging capabilities.

Eventually, a more rigorous optimization study will be needed. A single design iteration will have to generate the mask model, simulate the measured DRCs to form the database, and feed those DRCs into the imaging algorithm to predict performance and update the parameters for the next iteration. Hundreds, if not thousands, of designs will likely be tested using a metaheuristic optimization algorithm, and generating the DRCs over the entire FOV in a full Monte Carlo simulation is computationally prohibitive. Instead, if the FEP assumption is kept (such that the DRCs are generated using FEP-only events), then a reduced-order model (ROM) can be developed to drastically simplify and accelerate the simulations.

Ray tracing can take advantage of this FEP assumption. In this method, the pathlength of a ray which passes through the mask and detector is calculated. If this ray represents the path of a photon, the probability, P_M, that the photon will not

interact with the mask is given by

$$P_M = e^{-\mu_M x_M}, \tag{3.1}$$

where μ_M is the mask's linear attenuation coefficient for the given photon energy (equal to the mass attenuation coefficient times density) and x_M is the pathlength through the mask. The probability, P_D, that the same photon will be absorbed by the detector is equally given by

$$P_D = C\ (1 - e^{-\mu_D x_D}), \tag{3.2}$$

where μ_D and x_D are again the detector's linear attenuation coefficient and pathlength and C is a scaling constant equal to the ratio of interactions resulting in a complete charge collections versus all possible photon interactions. It should be noted that C is not just the photoelectric-to-total cross section ratio as multiple interactions within the detector can still result in the photon energy being completely deposited. This ratio is highly dependent on detector geometry and is typically calculated using a Monte Carlo simulation. However, the simulation only needs to model the detector medium itself, and C only needs to be calculated once for all and can simply be included in the ROM afterwards.

Combining these two equations yields the probability, P, that a photon passing through the RSM will fully deposit its energy in the detection medium:

$$P = C\ e^{-\mu_M x_M}\ (1 - e^{-\mu_D x_D}). \tag{3.3}$$

Taking the average probability of many rays yields the estimated probability that a photon emitted from a source is detected in the FEP, which is the same value reported

by MCNP and can be used to generate the DRCs. This method has many advantages over MCNP as there are significantly fewer interactions, variables, and particles (i.e. rays) modeled, while still accounting for the varying mask and detector pathlength as a function of the source's polar angle.

This improvement in speed does come at the loss of information; specifically, this method neglects any photons that scatter into the detector and does not provide any insight on the expected spectrum below the FEP. However, any photons that do scatter into the detector will be at lower energy than the energy at which they were emitted (while some of them will be due to coherent scattering, this is usually negligible at the energy regimes being considered). Therefore, assuming that the background subtraction in the experimentally measured energy spectra can sufficiently isolate the FEP, ray tracing may be a viable alternative to MCNP that can significantly improve computational efficiency for an optimization study.

3.3.1 Initial Ray Tracing Comparisons

Using the techniques described, a ray tracing code was developed to support future optimization work. The code was initially evaluated with a Spartan mask design as it was easier to test due to the simpler geometry and corresponding DRCs. Figure 3.3 shows the model used, with a simplified geometry containing only the mask's surface planes, the cylindrical detector, and monodirectional rays to simulate a far-field source (the aluminum sleeves were excluded from both the ray tracing and MCNP models in this section). The Spartan model used a 7.5 cm maximum voxel length, with each voxel spanning 10° in both polar and azimuthal angles and ranging from 0° - 170° in ϕ. The model used a 1" right circular cylindrical NaI detector.

Similar to the MCNP geometrical variance reduction, rays that did not intercept the detector were not included as they would not contribute to the FEP. To further

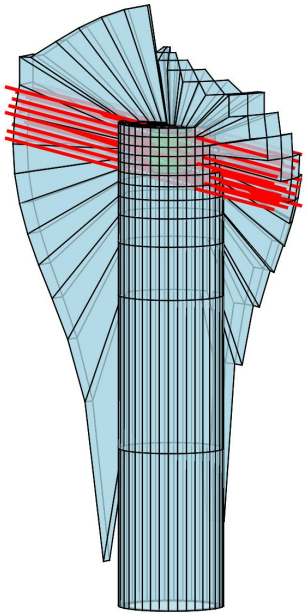


Figure 3.3. The ray tracing model for a Spartan mask. The model used a simplified geometry using only the mask surface (light blue) and a detector (light green) to calculate pathlength, with monodirectional rays (red) to simulate a far-field source.

improve computational efficiency, the intercepts were calculated for a subset of planes in an ROI immediately surrounding the rays (defined as two voxel widths) and for planes in between the ray's origin and the detector (therefore excluding planes that are known to not intercept or intercepts that occur after a photon would have already passed through the detector). For the model, u_M was set to 0.09975 cm^{-1}, while u_D was set to 0.2829 cm^{-1} for the NaI detector and 0.4496 cm^{-1} for a $LaBr_3$ detector tested later. These values were calculated from the total mass attenuation coefficient at 662 keV for the respective materials and densities [31].

Pathlengths were calculated by finding the intercepts of the ray as it passed through the mask and detector, which were then evaluated using Equation 3.3 to generate predicted DRCs. The ray tracing model accounted for multiple intercepts (i.e. a ray that enters and exits the mask multiple times before arriving at the de-

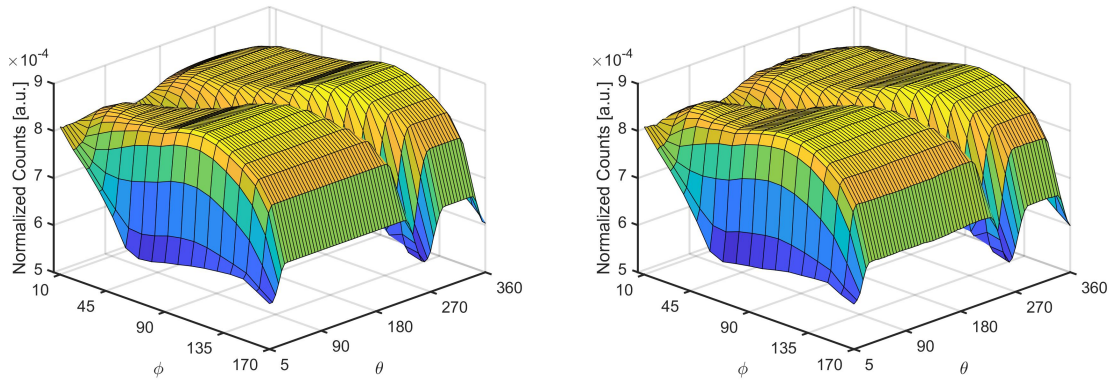
tector) by first ordering the intercepts as they occurred along the ray trace and then calculating the total pathlength as

$$x_M = \sum_{n=1,3,5,\ldots}^{N} ||\boldsymbol{x}_n - \boldsymbol{x}_{n+1}||, \tag{3.4}$$

where N is the total number of intercepts in between the ray origin and detector, $\boldsymbol{x}_n$ is the n^{th} intercept coordinate vector from the origin, and $||\ldots||$ is the euclidean distance operator. Duplicate coordinates were removed, which occurs when a ray intercepts the edge or corners of two or more adjacent planes. In the unlikely event that a ray was parallel to a plane and passed through it (i.e. it runs precisely along a voxel edge), the initial intercept, $\boldsymbol{x}_n$, was set to midpoint of the ray through the plane, as it represents the average distance a photon would intercept the mask given small perturbations. It should be noted that a small 1e-6° offset was included in all rays to prevent this type of intercept from occurring (without significantly altering the DRC), and increasing the number of rays simulated similarly minimized any risks from this approximation.

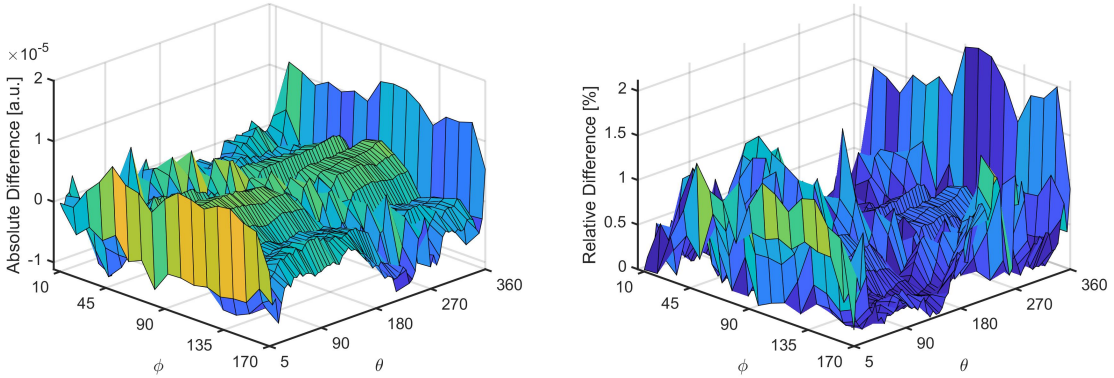
The same model was imported into MCNP v6.2, using a monoenergetic 662 keV photon beam and recording the events within the FEP to generate the DRCs, with enough particles simulated so that the uncertainty was less than 0.5%. Figure 3.4 compares the two methods by plotting their respective FEP DRMs. To account for the scaling constant C in Equation 3.3, the two DRMs have been normalized by their respective sums. The two show remarkable agreement; this is exemplified even more in Figure 3.5, which plots the absolute and relative differences between the two models. With a maximum relative difference of only 2.5% and an average relative difference of just 0.35%, the ray tracing was able to match the MCNP model within the variance expected based on the reported MCNP statistical uncertainty.

Figure 3.6 provides a closer look, plotting sample MCNP and corresponding ray



(a) Ray Trace (b) MCNP v6.2

Figure 3.4. (a) Ray Trace DRM versus (b) MCNP v6.2 DRM for a Spartan design. The DRMs have been normalized to the sum of counts and the uncertainty in the MCNP model is less than 0.5%. The good agreement between the two demonstrates the plausibility of using ray tracing as an alternative to more rigorous, computational expensive Monte Carlo simulations.



(a) Absolute Difference (b) Relative Difference

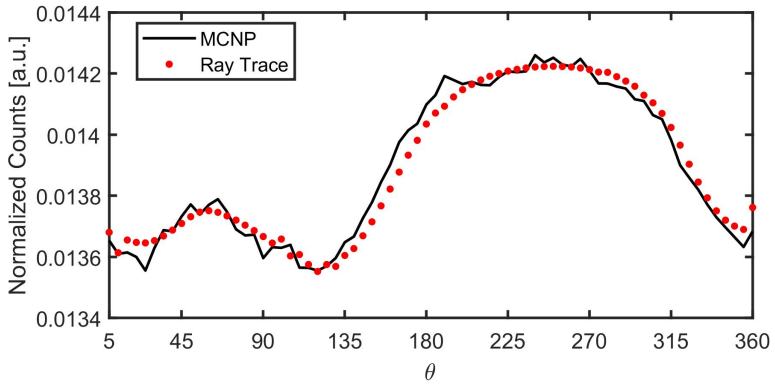
Figure 3.5. (a) Absolute difference and (b) relative difference (in reference to MCNP) between the ray tracing and MCNP DRMs. The largest error occurs near regions with the largest attenuation, with a maximum 2.1% relative difference. However, the average relative difference was only 0.35% between the models, validating ray tracing as a viable method to simulate the system response.

tracing DRCs at three different polar angles. In all cases, it is clear that the ray tracing was able to match the MCNP models, even capturing the non-ideal (and non-intuitive) DRC shapes at low polar angles caused by the detector geometry [29]. At higher polar angles, the DRCs are near-perfect matches (to within the 0.5% uncertainty in the

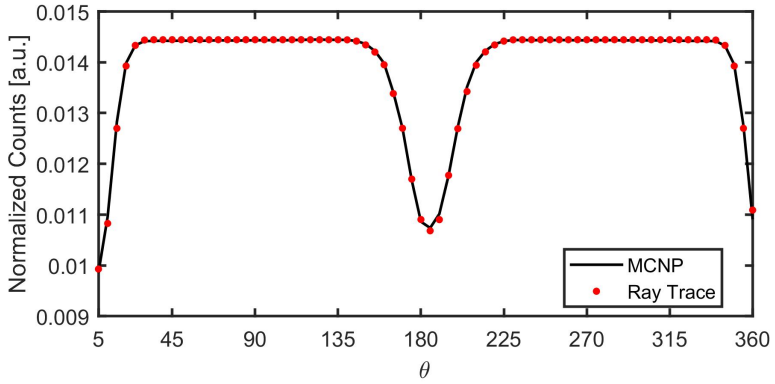
MCNP models), as the ray tracing was able to replicate the depth and width of the "valleys" to well within statistical uncertainty. Using the modal assurance criterion as a FOM (with mean subtraction) [28, 29], the DRCs in Figure 3.6 all have excellent 97.05%, 99.87%, and 99.18% similarity, with the $\phi = 10°$ being the worst-case and $\phi = 90°$ best case.

While the two models produced equivalent results, the two varied significantly in computation time. The ray tracing model, on a quad-core 4.0 GHz processor using MatLab, took less than 5 minutes to produce the DRCs, whereas the MCNP model, using a high-performance computing system with 40 nodes at 2.2 GHz, took approximately 1 hour to complete. While the comparison between the two is not exact given the different processing powers, it is still clear that the ray tracing method takes significantly less time compared to the MCNP simulations. It should also be noted that the ray tracing results shown were calculated using 1,223 rays covering the 1" crystal, but initial testing suggests that fewer rays may produce equivalent results, potentially further improving the computational efficiency.

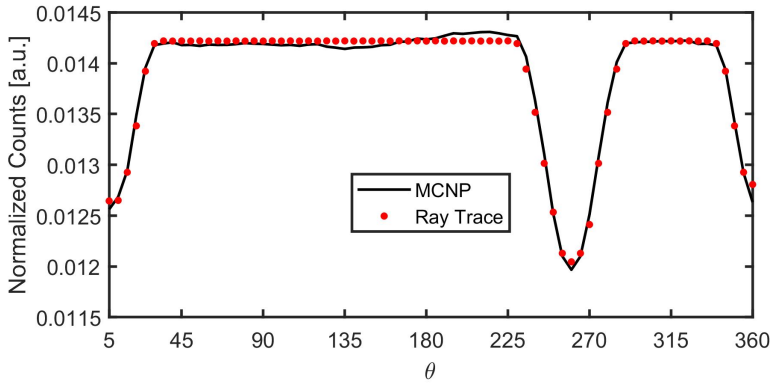
Increasing the model complexity noticeably reduced the ray tracing accuracy. Figure 3.7 compares the DRCs using ray tracing and MCNP for the experimental configuration shown in Figure 3.1a, with the MiniMace mask design and 10 mm $\times$ 10 mm $\times$ 50.8 mm $LaBr_3$ detector. Although the ray tracing simulation was able to capture the prominent trends shown in the MCNP DRC (which has a slightly increased 2% uncertainty compared to the previous Spartan DRCs), it did not capture the sharp features and tended to "smooth" the DRC. However, the agreement is still remarkably good given the increased computational efficiency, with an overall 91.98% similarity as measured by the modal assurance criterion. One possible explanation for the discrepancy is the uncertainty within the linear attenuation coefficient for $LaBr_3$ used in the ray tracing model, and future work can explore fine-tuning this parameter.



(a) $\phi = 10^\circ$ (97.05%)



(b) $\phi = 90^\circ$ (99.87%)



(c) $\phi = 170^\circ$ (99.18%)

Figure 3.6. Ray trace versus MCNP DRCs for three different ϕ directions, with similarity based on the modal assurance criterion given in parentheses. The ray trace was able to match the MCNP model extremely well, even capturing the non-ideal DRC shapes at low polar angles. The depths and widths of the "valleys" are almost exactly the same between the two (note that the MCNP DRCs have a 0.5% uncertainty which cannot be seen in the plots).

Similarly, more rays could be modeled which may further improve the accuracy.

It should be noted that the MiniMace RSM is expected to be the most complex

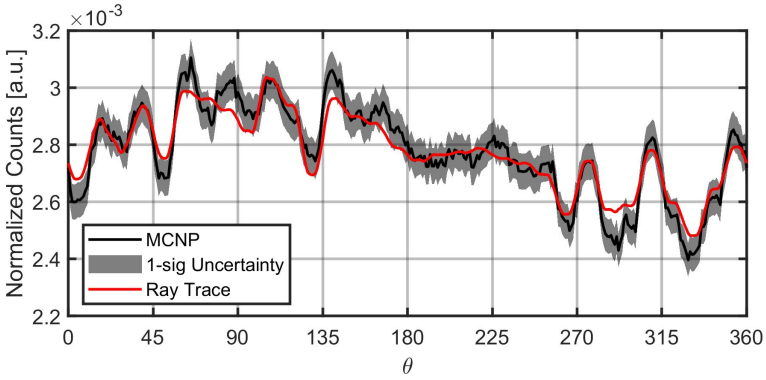


Figure 3.7. Example DRC comparing ray tracing to MCNP simulation (with shaded 2% uncertainty) for the MiniMace experimental mask. The ray-tracing DRCs are noticeably worse representations of the response as determined by MCNP than the Spartan comparisons, but the ray tracing was still able to capture the prominent features. Using the modal assurance criterion, the two DRCs are 91.98% similar.

design to be tested, with simpler designs potentially having similar, if not better, capabilities. Thus, the results in Figure 3.7 represent the most extreme scenario, which is still relatively accurate. These results demonstrate that ray tracing is likely a viable alternative to more costly algorithms. This could significantly reduce the computation time required during a robust design optimization and analysis for RSM imaging. While the design optimization is not further explored in this work, the ray tracing code developed provides a strong foundation for future studies to implement.

IV. Maximum-Likelihood Expectation-Maximization Reconstructions

This chapter is derived from a paper accepted by *Radiation Measurements*, under the title "Maximum Likelihood Reconstructions for Rotating Scatter Mask Imaging." The collaborating authors are Robert J. Olesen, Darren E. Holland, Erik M. Brubaker, and James E. Bevins. While the format of the journal article was modified to match the dissertation style, the contents of this chapter remain unaltered.

The article focuses on the ML-EM algorithm development for processing DRCs into a radiation image. As the chapter presents the paper in its entirety, much of the background has already been discussed in the previous chapters. For readers who are already familiar with the RSM and ML-EM concepts, Section 4.4.3 presents the criteria under which ML-EM will be tested while Section 4.5 discusses the results. Additionally, a detailed discussion on the ML-EM derivation and numerical stability of the algorithm is provided in Appendices 4.A and 4.B.

4.1 Abstract

The rotating scatter mask system is capable of determining the direction of a single point gamma or neutron source. However, the point source approximation does not hold for many realistic applications, requiring more detailed reconstructions with this system. This study is the first to characterize the rotating scatter mask as a gamma imager, or camera, through a detailed analysis of the relative error, precision, noise, and convergence time as a measure of reconstruction performance. Simulated ^{137}Cs sources and detector responses were generated in MCNP v6.1.4 and v6.2 with high fidelity and low statistical uncertainty. An analysis of variance was applied to maximum-likelihood expectation-maximization algorithms in order to determine the most significant factors in reconstructing the image, focusing on the source's shape,

size, and direction relative to the detector. Parameters for a Median Root Prior smoothing function were optimized to balance performance over a variety of source distributions. The algorithm performed reasonably well for point sources and for sources that spanned less than 30° in size and were located near the equatorial region of the rotating system. In most other scenarios, the algorithm either oversmoothed the image, resulting in blurry images, or completely failed to reconstruct the image. Increasing the resolution improved the reconstruction quality, while increasing the neighborhood size for the Median Root Prior reduced the image's noise, but at a significant cost to the computational efficiency. These results demonstrate that rotating scatter mask imaging is possible and introduces a potential alternative to other imagers. However, they also demonstrate that new techniques must be developed to increase the system's performance and robustness for gamma imaging applications.

4.2 Introduction

There is a continually growing interest in portable, real-time radiation imaging systems within the nuclear safety and security fields. Radioactive sources can be found in many peaceful applications, ranging from medical to power generation. Radioactive material localization is particularly important for nuclear power facilities due to the storage of radioactive waste and increased number of plants requiring characterization in preparation for decommissioning [57]. Accurately monitoring these facilities maintains positive control of potential contamination and minimizes the associated health risks for inspectors and personnel. Localization also significantly improves nuclear security by enabling the interdiction of illicit nuclear material and providing more precise details on the homogeneity and location of the material [58]. Sites may also be monitored using these systems for treaty verification purposes to detect any unwarranted changes in storage and production [59].

Coded-aperture imaging and Compton cameras have become popular devices among modern gamma-ray imagers [60]. However, the former is limited in its field-of-view (FOV) and is often composed of an array of attenuating masks and scintillating detectors, inhibiting the portability of the device [7]. Compton cameras benefit from a full 4π FOV and being highly portable, but they have reduced efficiencies due to their sensitivity in incident energy and cannot be used for imaging neutron sources [61, 62]. Advanced neutron scatter cameras are currently being investigated that address many of these problems [16, 17]. However, all these methods require high channel counts for accurate imaging and are often complex, costly systems.

The rotating scatter mask (RSM) was developed as an inexpensive single-detector system, capable of determining the direction for radioactive sources with a near-4π FOV [25]. By using a single standard scintillating detector, the RSM offers improved efficiency over Compton cameras and reduced size over coded-apertures, while significantly reducing the complexity and cost. Initial studies suggest that the RSM may also be used to monitor gamma and dual-particle environments [30, 26].

Thus far, the research surrounding the RSM has focused on determining the direction of point sources. This work seeks to establish an image reconstruction algorithm that evolves the RSM into an imaging device. For this study, the maximum-likelihood expectation-maximization (ML-EM) algorithm was chosen due to its wide deployment in the medical imaging community and coded-aperture systems [63]. The remainder of the chapter is organized as follows. Section 4.3 describes a brief overview of the RSM system. Then, the ML-EM algorithm is explained and derived for the RSM in Section 4.4, along with the figures of merit (FOM) used to analyze algorithm performance. Reconstructed images using the ML-EM algorithm and RSM response simulations are shown and discussed in Section 4.5, while the key findings and future recommendations are summarized in Section 4.6.

4.3 Rotating Scatter Masks

Figure 4.1 shows a schematic indicating the primary components of an RSM system. A mass of acrylic-like material, referred to in this work as the mask, surrounds a single, directionally-insensitive scintillating detector. Unlike the masks used in coded apertures, which project an image onto an array of detectors, this mask is more closely related to Bonner spheres, attenuating the particles before they reach the detector. However, unlike a Bonner sphere, the mask is geometrically complex so that the attenuation varies based on the source direction. This mask is attached to a motor so that it may freely rotate about the detector's longitudinal axis; in the presence of a static radioactive source, this rotation creates a periodic, time-modulated detector response that can be used to characterize the source's direction. This methodology follows the same process as one- and two-dimensional time encoded imaging systems [42].

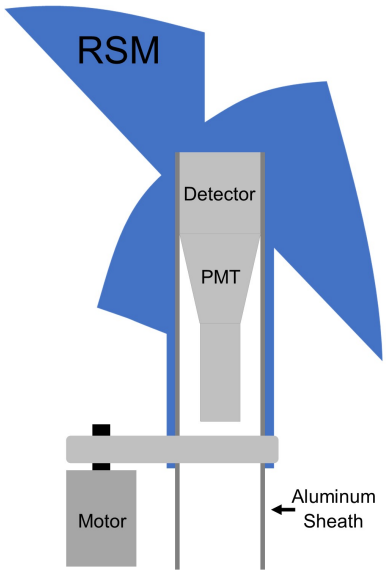


Figure 4.1. Schematic showing the RSM system's primary components [29]. An acrylic-like mask (labelled RSM) slides over an aluminum sheath surrounding a scintillating detector and is attached to a motor so that it may freely rotate about the detector.

The RSM's initial computational (using both MCNP [55] and GEANT4 [64]) and

experimental development demonstrated the ability to identify the direction of both neutron and gamma-emitting point sources to within 10° or better [25, 29, 30, 26, 27]. These studies noted that the largest limiting factor for the direction identification accuracy was the degenerate nature of the detector response curves (DRCs), thereby limiting the design's application for imaging. An optimization study on the mask geometry was conducted to improve and minimize these degeneracies, resulting in the "Mace" RSM design shown in Figure 4.2 [28, 29]. While optimized for directional measurements, the Mace design should have the best imaging characteristics of the RSMs designed so far.

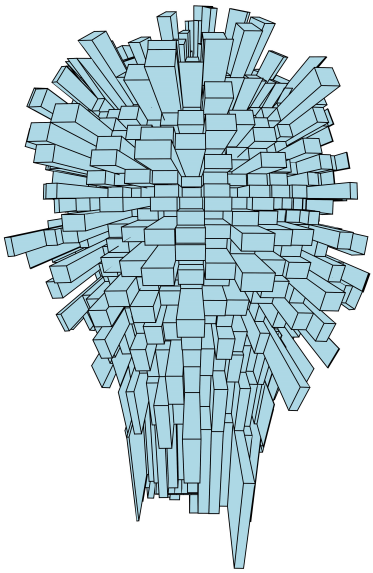


Figure 4.2. Design for the Mace RSM. The complex geometry decreases the response curves' degeneracies, theoretically improving the imaging capability.

The Mace DRCs form the basis set for this work. These DRCs were generated using the full-energy peak from a ^{137}Cs-simulated 662 keV point source located 86.36 cm from the center of a 3" x 3" NaI(Tl) detector, modeled in MCNP v6.1.4. The DRCs span from 5° - 170° in the polar angle ϕ (referenced from the axis of rotation) and 0° - 360° in the azimuthal (rotational) angle θ, both in 5° increments for a total of 2,448 source directions and corresponding DRCs. Distributed sources were modeled as a

collection of point sources so that the total detector response could be modeled as a weighted sum of the individual DRCs.

4.4 Methodology

In imaging, the goal is to solve for the initial source direction(s) given a measured detector response and a known detector response matrix (DRM). However, degeneracies in the DRCs result in an underdetermined problem such that an infinite number of solutions (images) exist that produce the same observed detector response [28]. For these types of problems, the goal is to produce a solution that meets a predefined set of criteria based on expectations of the true solution. This work adopts the ML-EM algorithm to reconstruct the image due to its prevalence in similar radiation imaging applications.

4.4.1 Maximum-Likelihood Expectation-Maximization

First developed in 1976, ML-EM is an iterative algorithm that seeks to find maximum likelihood estimates from observations that can be viewed as incomplete data of the overall system [43]. In common terms, suppose there are many solutions, λ, to a problem that could account for an observation, $\mathbf{y}$. ML-EM finds the solution that maximizes $P(\mathbf{y}|\lambda)$: the probability of observing $\mathbf{y}$ given the estimated solution parameters (reconstruction) and a known probability distribution that the events are observed.

Deriving the ML-EM algorithm for the RSM system follows the original outline presented by Shepp and Vardi for radiation emissions [44]. Suppose that a source is located in pixel j (corresponding to a specific direction) with an unknown mean emission of particles, λ_j. The number of detected particles measured at mask angle θ_i is given by y_i. Following the derivation in 4.A, the resulting equation is a compact

form that is remarkably similar to those used for positron emission tomography:

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\sum_i p_{ij}} \sum_i \frac{y_i p_{ij}}{\sum_k \lambda_k^m p_{ik}}, \tag{4.1}$$

where the m superscript denotes the iteration index and p_{ij} is the conditional probability that a particle emitted from pixel j is detected at the mask rotational angle θ_i (and are thus the normalized probabilities calculated from MCNP). The conditional probability requires that $\sum_i p_{ij} = 1$, allowing the equation to be further simplified as well as improving the numerical stability. However, it should be noted that there is no strict restriction in requiring the conditional probabilities for the general ML-EM algorithm.

For this study, the initial activity was set to one for each pixel ($\lambda_j^0 = 1$ for all j). Since this algorithm is effectively equivalent to the one derived by Shepp and Vardi [44], it maintains the same characteristics that:

- The log likelihood is concave, guaranteeing that the maximization step does indeed solve for a local maximum.
- The log likelihood strictly increases after each iteration and will, therefore, always converge to a solution.
- The total number of particles is conserved between each iteration. That is, $\sum_{i,j} \lambda_j^{m+1} p_{ij} = \sum_{i,j} \lambda_j^m p_{ij} = \sum_i y_i$.

However, it is important to note that ML-EM does not guarantee a solution for the global maximum and is only guaranteed to converge to one of potentially many local maxima.

4.4.2 Median Root Prior

Equation 4.1 assumes nothing about the actual shape of the reconstructed image. Many studies have been conducted on implementing Bayesian priors that constrain the algorithm to specific classes of images, often implementing smoothing functions that limit the solution to a localized distribution [45]. These methods introduce the prior to the likelihood function, usually taking the form

$$L(\boldsymbol{\lambda}) = \prod_{i,j} e^{-\lambda_{ij}} \frac{\lambda_{ij}^{n_{ij}}}{n_{ij}!} e^{-V(\lambda_j)}, \tag{4.2}$$

where $V(\lambda_j)$ describes some potential function that compares each pixel to its local surroundings.

The Median Root Prior (MRP) is a method that uses a Gaussian potential function

$$V(\lambda_j) = \beta \frac{(\lambda_j - M_j)^2}{2M_j}, \tag{4.3}$$

where M_j is the median of the pixel values in a neighborhood surrounding λ_j and β is a weighting term ranging from zero to one (this term and its relationship to the conditional probabilities, numerical stability, and uncertainty are discussed in 4.B). The advantage of comparing a pixel to the median, rather than a value such as the average, is that this solution preserves sharp edges rather than oversmoothing the image. Previous studies have shown that using MRP in conjunction with ML-EM constrains the solutions to locally monotonic sources: ideal for realistic radioactive source distributions [46, 47].

Following the same steps in 4.A, the algorithm including the prior is remarkably similar to before:

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\sum_i p_{ij} + \beta \frac{\lambda_j^m - M_j}{M_j}} \sum_i \frac{p_{ij} y_i}{\sum_k \lambda_k^m p_{ik}}. \tag{4.4}$$

Note that when $\beta = 0$, the algorithm reduces to the standard ML-EM form; thus, Equation 4.4 will be referred to as the generalized ML-EM algorithm.

The amount of detail recoverable in the image is governed by the neighborhood size used to calculate the median. Any feature size smaller than this neighborhood is treated as noise and subsequently penalized. For this study, the smallest possible MRP neighborhood, an initial 3x3 area surrounding each pixel, was used so as to extract the most amount of detail from the images. Note, the neighborhood cannot be 2x2, or any other even-sided square, as there is no center pixel and thus would introduce some form of directional bias to the final solution. The algorithm was determined to have converged when the sum of the all the relative differences between pixels of two successive iterations was less than 1%.

4.4.3 Performance Criteria

A factor screening analysis was performed using the principles from analysis of variance (ANOVA) to determine the most significant factors associated with the performance of the image reconstruction. ANOVA is a common statistical technique developed by Ronald Fisher in the 1920's that provides a statistical test for comparing two or more population means. This process is conducted by analyzing the ratio of variances from predetermined factors and using the F-test to estimate the probability of the measured variances, denoted by the p-value. ANOVA can also be used to produce a model for predicting factor effects. In a linear model, where the factors x_1 and x_2 are chosen to predict the measured response, y, the linear model takes the form

$$y = c_1 x_1 + c_2 x_2 + c_{12} x_1 x_2, \tag{4.5}$$

where c_1 and c_2 are the constants of proportionality for the primary factor effects and c_{12} is the constant of proportionality for the first-order interaction term. The

complete, quadratic model extends the relationship to the form

$$y = c_1 x_1 + c_2 x_2 + c_{12} x_1 x_2 + c_{11} x_1^2 + c_{22} x_2^2. \tag{4.6}$$

Under the null hypothesis, all constants of proportionality are equal to zero, and a significant p-value indicates a non-zero factor effect, which can then be estimated by comparing the factor variance to the overall group variance.

ANOVA is traditionally used to form these models and accurately predict a measured response based on known, predetermined factor values. However, in radiation imaging (particularly for source interdiction) the details of the source are unknown prior to the measurement. Thus, the objective of this analysis was to quantify the general impact of factors in a manner that allows for direct comparison between the factors and that can be applied to different image-reconstruction algorithms, rather than generating a predictive model. The three primary factors chosen were the source's: 1) shape, 2) size, and 3) center polar angle. For simplicity, the azimuthal angle was excluded as preliminary results indicated it had no impact on the image reconstruction for a single source. Four categories of shapes, shown in Figure 4.3, were chosen that mimic realistic source distributions:

1. Disc: Corresponding to a cross-section of a cylinder or sphere. This mimics nuclear waste containers used in storage as well as spherical forms of special nuclear material.

2. Ring: Corresponding to a hollow cylinder. This mimics contaminated storage drums as well as the annular shape commonly used to store special nuclear material.

3. Junctions ("X" shape): Used to simulate contaminated pipes or multiple fuel rods.

4. Square: Used to capture sharp corners that are not available in the other images. Extensions of this shape include rectangles that mimic the cross-section of cylinders including straight pipes, fuel rods, and waste storage drums.

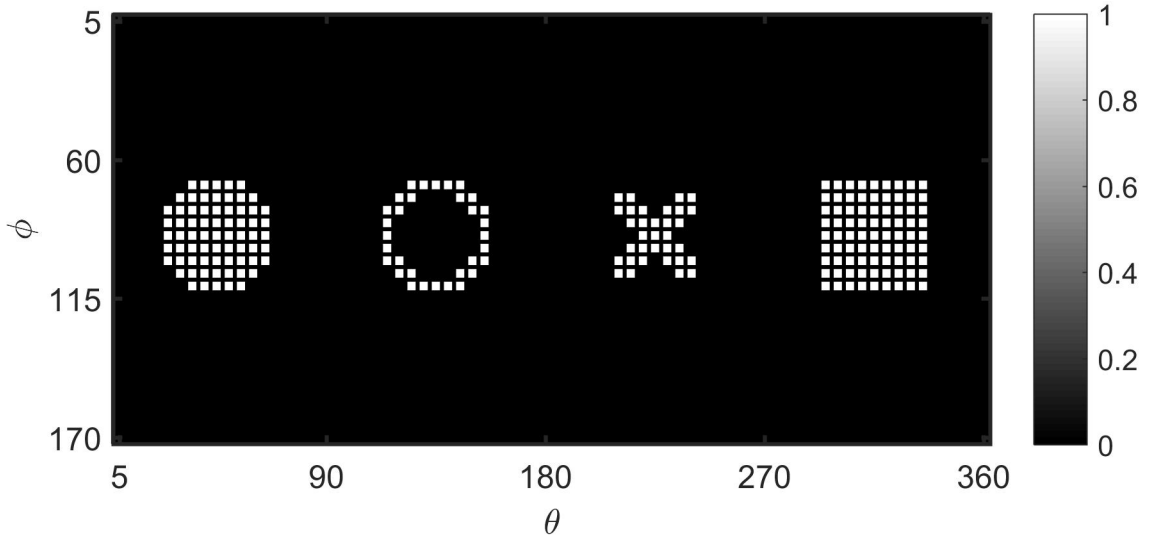


Figure 4.3. From left to right the four source shapes used: disc, ring, junction, and square. In this study, the mathematical convention for spherical coordinates was used where ϕ is the polar angle and θ is the azimuthal angle. The color scale indicates the pixel's relative activity.

The size parameter corresponds to the maximum angular extent of the source. Three values of 0°, 30°, and 50° were chosen, whereby the maximum corresponds roughly to a 1:1 ratio of source size to stand-off distance (i.e., a source 1 meter in length with a center located 1 meter away would span approximately 50°). Finally, three polar angles of $\phi = 30°$, 90°, and 150° were chosen for the source's center based on previous studies indicating that the RSM is less sensitive for sources near the poles [29]. The combination of these three factors was assumed to be robust enough to simulate most realistic sources for the cases considered in this study.

Under ANOVA, the null hypothesis assumes that all factors affect the reconstruction equally regardless of their value. For the statistical analysis, any factor with a p-value less than 0.10 was deemed to be statistically significant. This threshold was

chosen to ensure that reasonably-sized effects could be accounted for without overgeneralizing the analysis. A full factorial model was used for the factor screening, giving a total of 36 source images. However, the 0° size distributions included 9 replicated point sources, so a total of 27 *unique* source distributions were tested. In order to conduct the full factorial analysis, only the 24 distributed sources were included in the model as the three point sources would bias the results (by definition, a point may only be one size and so there is not a logical way to categorize and quantify it within the set of distributed sources). These three point sources were analyzed separately and compared to the general distributed source performances.

The relative error, E, between the true image, $\boldsymbol{\lambda}$, and the reconstructed image, $\boldsymbol{\lambda}^m$, was used to quantify the image reconstruction accuracy:

$$E = \frac{\sum_j |\lambda_j - \lambda_j^m|}{\sum_j \lambda_j}. \tag{4.7}$$

The convergence time for each of the images was also tested using ANOVA by recording the total number of iterations the algorithm took before reaching the 1% convergence criterion. Finally, the ML-EM's reconstruction precision and standard deviation was measured in order to quantify the image noise.

The algorithm's precision was defined as the ratio of the total activity contained within the true source distribution of the reconstruction to the overall total reconstructed activity. Thus, the precision value quantifies how much of the original source the ML-EM algorithm was able to reconstruct compared to "false sources", or phantoms, which the relative error alone does not indicate. Equation 4.8 provides the mathematical formula, assuming that the source distribution is monotonic (uniform

activity) and has been normalized to the maximum activity:

$$\text{Precision} = \frac{\sum_j \bar{\lambda}_j \bar{\lambda}_j^m}{\sum_j \bar{\lambda}_j^m}, \tag{4.8}$$

where $\bar{\lambda}$ denotes the normalized, relative activity for each pixel. A perfect reconstruction would yield 100% precision.

With this definition, it is still possible that a reconstruction could have a high precision but contain a noisy image within the shape's distribution. Therefore, the standard deviation of the reconstruction *within* the true distribution was also recorded. This FOM is defined as:

$$\sigma = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left(\bar{\lambda}_j \bar{\lambda}_j^m - 1\right)^2}, \tag{4.9}$$

where N is the number of pixels and the mean is assumed to be one based on an assumption of uniform activity. An ideal reconstruction would have a low relative error, high precision, a low standard deviation, and low convergence time. The combination of these four FOM provides a sufficient set of quantifiable representations for the reconstruction's quality, which can then be used for comparing the performance of future imaging algorithms.

4.5 Results

4.5.1 Prior Weighting

In order to determine the optimal prior weighting, 20 β values logarithmically-spaced between 10^{-5} and 1 were tested for each image. The reconstructions' relative errors as a function of β are plotted in Figure 4.4 for the three point sources and in Figure 4.5 for the 24 other distributed sources (intermediate values are interpolated for easy visualization). As expected, the algorithm's performance varied inversely

between the point and distributed sources. With the three points, the increasing β caused the image to become oversmoothed due to the neighborhood size discussed in Section 4.4.2. Conversely, increasing β generally had a positive effect for the distributed sources by increasing the locality and smoothness of the reconstructed image. This inverse relationship between point and distributed sources and β can be readily seen in Figure 4.6a, which plots the average relative error for the point sources and the average for the distributed sources on two separate lines.

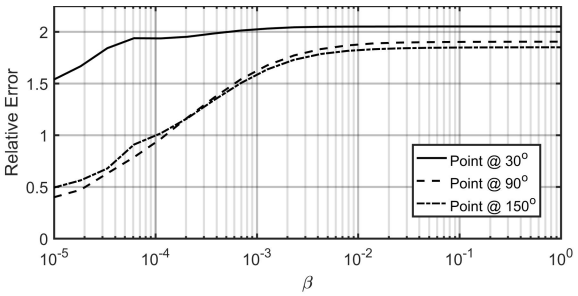
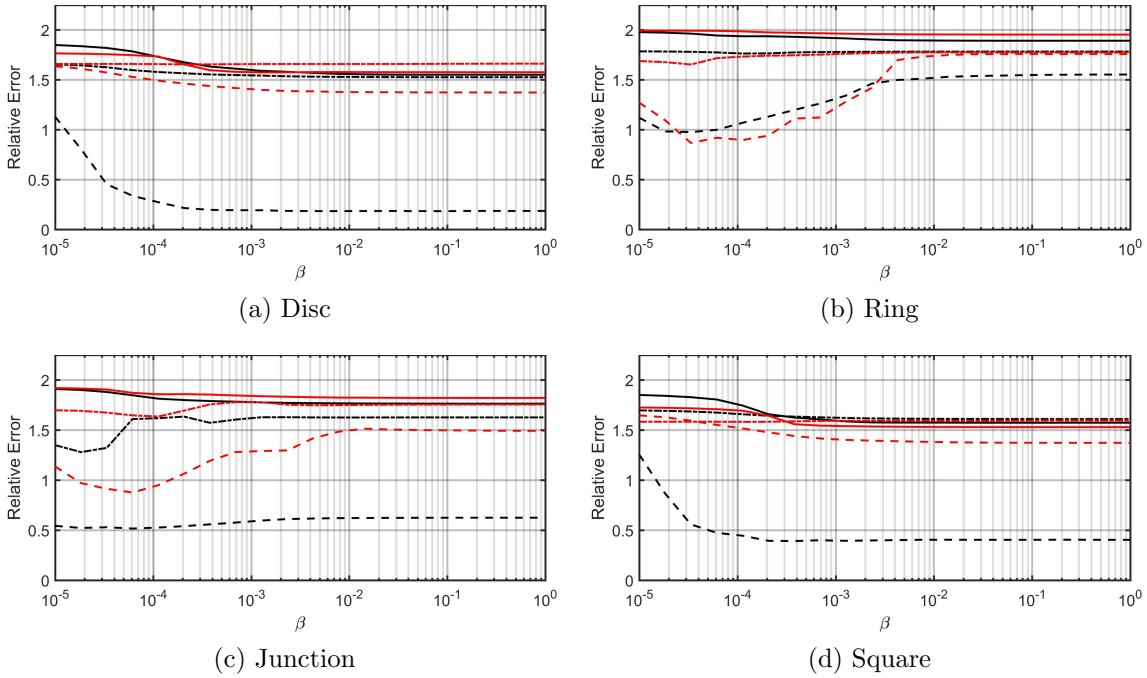


Figure 4.4. Point source relative errors versus the prior weight for the three point source images (continuous lines are interpolated between the points for easy visualization). As expected, the relative error increases with increasing prior weight due to oversmoothing.

There also appeared to be diminishing returns as β was increased. As shown in Figure 4.6a, both the point and distributed reconstructions converged, with most of the distributed sources reaching a steady-state relative error after $\beta \approx 10^{-3}$ and the point sources reaching a maximum around $\beta \approx 10^{-2}$. However, Figure 4.5 shows that whereas the relative error for the disc and square sources generally monotonically decreased with increasing β and reached a minimum around $\beta \approx 10^{-3}$, the relative error for the ring and junction sources reached a minimum about $\beta \approx 10^{-4}$ before increasing again and reaching a steady-state after $\beta \approx 10^{-2}$.

This relationship can be easily explained by the shape of the images. For both the disc and square sources, the images were generally smooth with no "sharp" features within a neighborhood. Thus, they matched the Gaussian prior assumption better so that increasing the prior weight improved the reconstruction. The ring and junction



(a) Disc (b) Ring

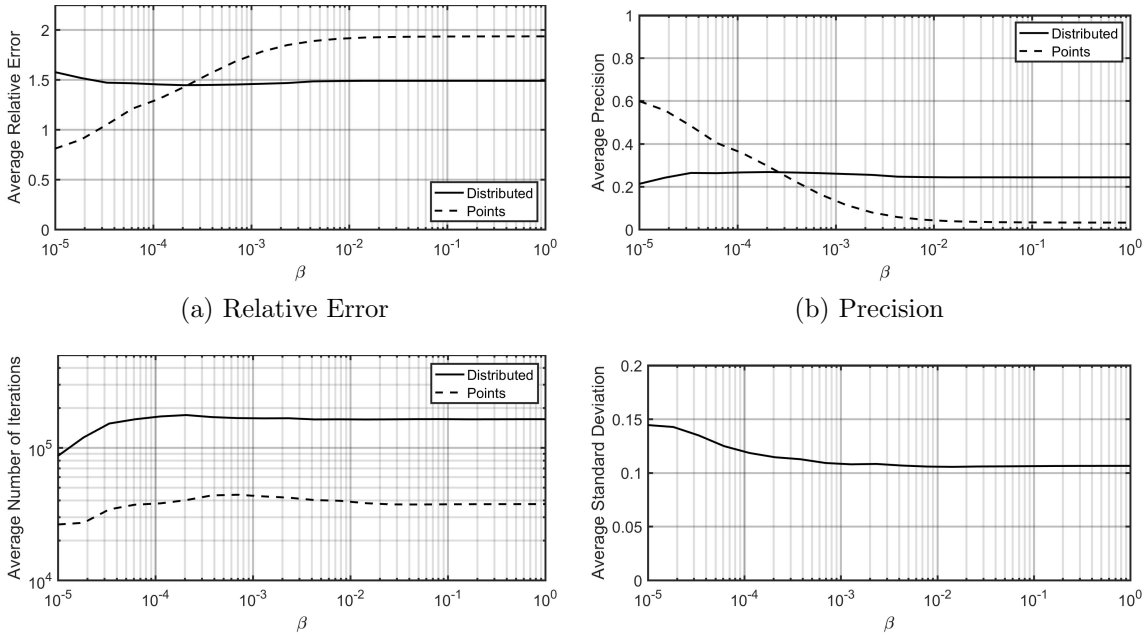
(c) Junction (d) Square

Figure 4.5. Distributed source relative errors versus the prior weight for sources located at 30° (solid), 90° (dashed), and 150° (dash-dot) polar angles. Black lines correspond to sources with a 30° angular extent while red lines correspond to the larger 50° angular extent.

sources contained many more sharp corners or thin edges, features that were consistently smaller than the 3x3 neighborhood size used to calculate the median. Therefore, initially increasing the prior weight improved performance by helping maintain the locality, but increasing the weight further caused oversmoothing and decreased the accuracy of the reconstruction.

Figure 4.6 also plots the average values for the other FOM. These plots show that the point sources perform significantly better at low β values, as expected since the inclusion of the MRP distorts the single pixel images to the point where the precision drops to nearly zero at large prior weights. Remarkably, the average precision for the distributed source reconstructions, Figure 4.6b, was relatively invariant against β, with a consistently poor 25% average precision.

Figure 4.6c shows that increasing β generally increases the number of iterations



(a) Relative Error

(b) Precision

(c) Iterations Required

(d) Standard Deviation

Figure 4.6. Average FOM for the distributed and point sources (note that the standard deviation could only be calculated for the distributed sources). The point sources generally perform better at low β values while the distributed sources have an optimal performance around $\beta = 10^{-4}$.

until convergence for both types of sources, although the number of iterations for both sources reached a relatively steady-state around $\beta \approx 10^{-4}$. Finally, the overall trend in Figure 4.6d shows that increasing β generally decreases the standard deviation within the true source region of the reconstruction. These results suggest that $\beta \approx 10^{-4}$ would be a favorable value to balance performance between point and distributed sources. In the next section, the ANOVA was conducted with reconstructions using $\beta = 10^{-4}$ to determine the most significant factors affecting the FOM. Section 4.5.3 summarizes the specific dependencies based on the ANOVA results.

4.5.2 Image Reconstruction ANOVA

4.5.2.1 Accuracy

The left hand columns in Table 4.1 show the ANOVA results for the linear, full factorial model of the relative errors for the 24 distributed source images. Under this model, each factor was considered as well as their first-order interactions (denoted by *). Using the predefined $p < 0.10$ significance level, none of the factors were shown to have a statistically significant impact on the relative error and the null hypothesis that all factors affect the reconstruction equally could not be rejected.

Table 4.1. Relative Error ANOVA for ML-EM with MRP

Factor	p-value (Separated by Model)		
	Linear	Quadratic	Simplified Quadratic
Shape	0.97	0.80	0.71
Size	0.44	0.13	0.067
Direction	0.64	0.34	0.26
Shape*Size	0.92	0.60	
Shape*Direction	1.00	0.99	
Size*Direction	0.98	0.97	
Direction^2		0.0001	< 0.0001

However, previous experiments suggest that a linear model may not fully capture the system as the RSM was shown to perform better near the 90° polar directions and worsen as the source moved towards either pole [29, 26]. Therefore, a quadratic model was tested by including a second-degree polynomial term for the direction. These results are given in the middle two columns of Table 4.1. With a 0.0001 p-value for the squared direction term, indicating a statistically significant factor, the null hypothesis may be rejected and the intuition that the relative error depended non-linearly on the direction was validated. Note that neither the shape nor size can be tested for a quadratic component as the former is categorical and the latter only has two data points.

The quadratic model ANOVA also shows that the interaction terms have relatively large p-values, meaning they likely do not affect the relative error. These terms may be excluded from the model to create better statistics on the remaining terms and more precisely quantify the statistically significant factors (this can also cause previously insignificant factors to become significant due to the increased statistics). The resulting simplified quadratic model is given by the right hand columns of Table 4.1. Excluding the interaction terms decreased the size first-order effect below the 0.10 significance threshold, whereas the shape and direction's first-order effect remained above the threshold. Ultimately, the ANOVA results show that the relative error depends largely on the square of the polar direction of the true source's center, with a minor dependence on size.

4.5.2.2 Convergence Time

The process from the previous section was repeated for the ANOVA with the number of iterations required until convergence. The resulting simplified quadratic model is shown in Table 4.2; again, most first-order interactions and the linear term for direction were determined to not be statistically significant. Unlike the relative error, it was the shape, rather than the size, that was determined to be a prominent first-order effect with a p-value of 0.06 and every category of shape having a significant p-value less than 0.10. Additionally, while the overall shape and size interaction term was not deemed significant, the individual ring and size interaction had a p-value of 0.036, indicating that this particular configuration impacted the results (the other categorical interactions did not meet the significance threshold). This subdivision is possible as the shape is a categorical factor, allowing for each specific category to be analyzed separately.

The ANOVA results show that the source's size is not statistically significant with

Table 4.2. Convergence Iterations ANOVA for ML-EM with MRP

Simplified Quadratic Model		
Factor		p-value
Shape		0.06
Size		0.22
Direction		0.17
Shape*Size	All	0.18
	Ring*Size	0.036
Direction^2		0.0004

respect to the number of iterations. This result was initially surprising, as intuition would suspect that larger images would take longer to converge since more pixels should be changing through each iteration. Figure 4.7 plots the number of iterations until convergence versus the angular extent of the 24 distributed sources. The plot largely indicates no correlation between the two, thereby supporting the ANOVA results. It is likely this is caused by the prior effectively blocking the reconstructions into minimum 3x3 area chunks and suppressing the effect caused by increasing the area of the actual source shape. However, it should be noted that the point sources required approximately an order of magnitude fewer iterations (Figure 4.6c), suggesting a certain threshold exists once the source reaches the size of the neighborhood used.

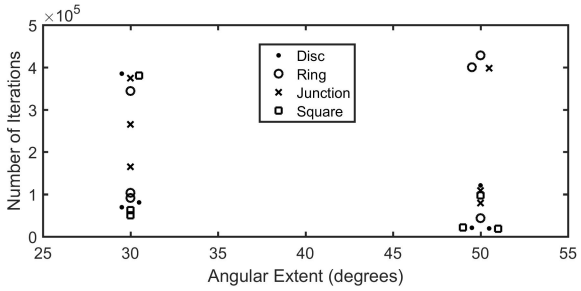


Figure 4.7. Number of iterations until convergence versus the angular extent of the distributed sources. There appears to be no correlation as the range is effectively equivalent; the number is only dependent on the source shape and direction. Note that similar data points have been artificially spread out laterally for better visuals.

4.5.2.3 Image Noise

Tables 4.3 and 4.4 show the reduced quadratic models for the reconstructions' precision and standard deviation. Similar to the other two FOM, the noise was determined to be highly dependent on the second-order directional component. The size component showed a reasonably small 0.092 p-value for the size factor, just surpassing the threshold limit and indicating some small dependence. The ANOVA for standard deviation also indicated that shape had a significant effect, with the ring and junction having p-values of 0.010 and 0.0053, respectively.

Table 4.3. Precision ANOVA for ML-EM with MRP

Simplified Quadratic Model	
Factor	p-value
Shape	0.72
Size	0.092
Direction	0.99
Direction^2	<0.0001

4.5.3 Factor Effects

Using the significant factors determined by ANOVA, each model was fitted using a least squares regression to understand the general effect for each factor. As an example, the model for relative error is given by

$$E = c_1 \left(\frac{\text{Size} - 40^\circ}{10^\circ} \right) + c_2 \left(\frac{\text{Direction} - 90^\circ}{60^\circ} \right)^2 + c_3, \tag{4.10}$$

where c_1 and c_2 are the proportionality constants to be fitted and c_3 is the bias. The terms in parentheses were automatically scaled with a range [-1,1] for each factor so that the proportionality constant's magnitude also represents the relative importance of the factor. For factors where shape was determined to be significant, each shape term was fitted to its own proportionality constant (if the equation were to be used

Table 4.4. Standard Deviation ANOVA for ML-EM with MRP

Simplified Quadratic Model		
Factor		p-value
Shape	All	0.014
	Ring	0.010
	Junction	0.0053
Size		0.72
Direction		0.15
Direction^2		<0.0001

to predict the relative error, the constants corresponding to the actual shape would be used while the others would be discarded). Table 4.5 summarizes the estimated proportionality constants for each FOM.

Table 4.5. Predictive Modeling from ANOVA

Factor	Proportionality Constant			
	Relative Error	Precision	Standard Deviation	Iterations ($\times 10^5$)
Disc	–	–	–	–0.56
Ring	–	–	–0.04	+0.63
Junction	–	–	+0.04	+0.60
Square	–	–	–	–0.67
Size	+0.11	–0.05	–	–
Direction^2	+0.84	–0.43	–0.12	–1.95
Ring*Size	–	–	–	+0.81
Bias	+0.89	+0.55	+0.20	+3.02

All FOM followed the same trend for the direction: the algorithm performed best for source directions near $\phi = 90^\circ$ and performance degraded as the source moved towards either pole. The magnitudes of the constants also indicate that the direction was the most important factor for all FOM (corresponding to the ANOVA results). For both relative error and precision, size was shown to have a small impact, with the performance decreasing as the source size increased. The ring sources surprisingly decreased the standard deviation (likely because the true image contained relatively few pixels), whereas the junction increased the standard deviation (which contained more pixels than the ring, but still not as localized as the disc or square).

The number of iterations was shown to have the most dependency, and variance, based on the shape of the source. Both the disc and square decreased the number of iterations until convergence while the ring and junction increased the number of iterations. The size of the ring also impacted the number of iterations, as smaller rings converged faster than the larger ones. These dependencies are relatively intuitive, as the disc and square sources match the Gaussian prior assumption better; it is more surprising that the shape did not have as much of an impact in the other FOM that are more related to the quality of the reconstruction.

One final discussion must be made about the relatively large bias present in all the models in Table 4.5. The consistently large bias indicates that the algorithm performed relatively poorly on average (for the precision, the bias indicates a maximum of roughly 55% for most sources). This result is demonstrated for the relative error in Figure 4.8, which plots the cumulative probability (x-axis) of the 24 reconstruction relative errors (y-axis) on the right and a box plot of the same data on the left. The probability axis is scaled so that any normal distribution about the median would appear linear. Superimposed on the probability plot is a red line connecting the first and third quartile and a dashed line extrapolating to the extremes. Under the ANOVA null hypothesis, most of the data points should fall along the normal line; therefore, this plot is a quick factor screening visualization to see if there are outliers from the normal or any non-normal distribution to the data in general.

Figure 4.8 shows that most of the data fall close to the superimposed normal line. While six fall outside the normal, only three may be classified as outliers (defined as more than 1.5 times the interquartile range from the first or third quartile). These three outliers correspond to the smaller disc, square, and junction sources located at the 90° polar direction, which is precisely in accordance with the results from Table 4.5. The other three sources outside the normal correspond to both rings and

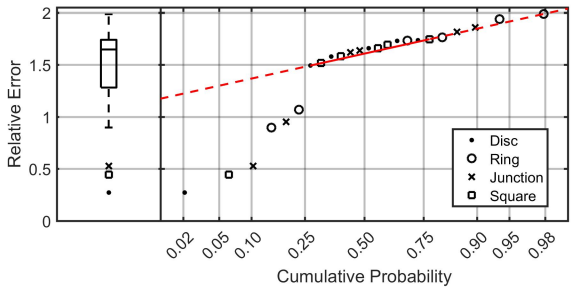


Figure 4.8. Box plot (left) and normal probability plot (right) for the 24 distributed source reconstruction errors. An interpolated solid red line connects the first and third quartile while the dashed red line extrapolates this line. The six data points that fall outside the normal indicate that there are statistically-significant factors affecting the reconstructions' performance.

the larger junction source located at 90°, again indicating that the direction was the largest driving factor. The remaining reconstructions had a relative error ranging from 1.5-2. Along with the other FOM showing the same pattern, these results indicate that ML-EM with MRP readily works for the specific range of sources that are relatively small and located along the equator, but on average performs poorly for most reconstructions as is apparent in the plots from Figures 4.5 and 4.6.

The performance of the ML-EM reconstruction as a function of these different parameters is highlighted in Figures 4.9 and 4.10, which show sample visual comparisons of the true and reconstructed source images for some of the best and worst performances, respectively. The smaller disc and point sources were reconstructed well with excellent agreement visually to the true source distribution. The image of the ring demonstrates the oversmoothing caused by the MRP, for while the algorithm did not fully reconstruct the image, the original ring shape was preserved and there were very few phantoms within the image. This preservation is promising for applications where detection and localization are more important over exact source shapes.

However, most of the reconstructions that followed the normal curve in Figure 4.8 were most similar to the images in Figure 4.10. These images portray the larger disc

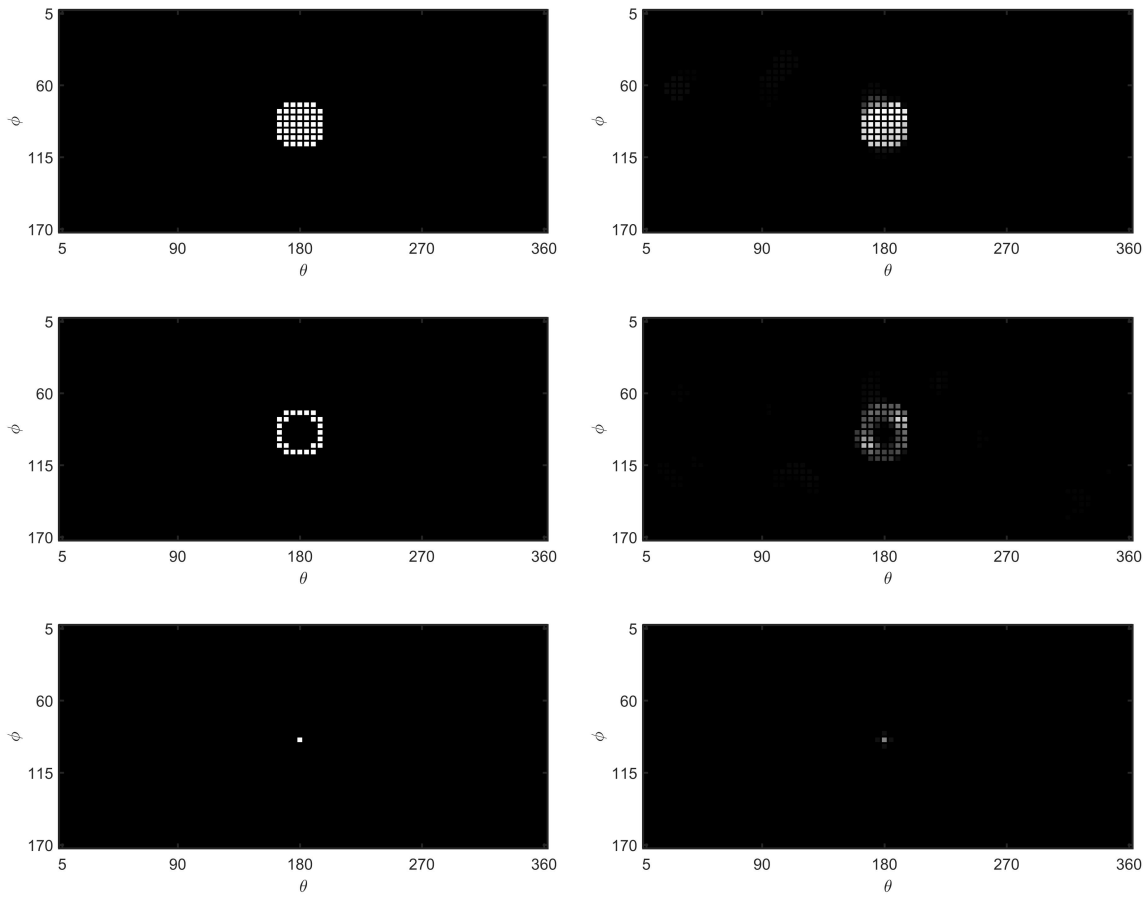


Figure 4.9. True source images (left) and their reconstruction (right) using ML-EM with MRP, $\beta = 10^{-4}$. The algorithm performed reasonably well for mid-sized sources but often oversmoothed the image, which is most noticeable in the ring reconstruction.

located at 30° and 90°, along with their reconstructions, which failed to capture any form of the original shape and contain many phantoms. The distortion in the top true image in Figure 4.10 is caused from projecting spherical coordinates onto a two-dimensional grid. The poor reconstruction occurred for all distributed sources near the poles and for the larger distributed sources at the equator.

4.5.4 Resolution Effects

As discussed in Section 4.4.2, the MRP method penalizes any features smaller than the neighborhood used to calculate the median. Increasing the neighborhood size may

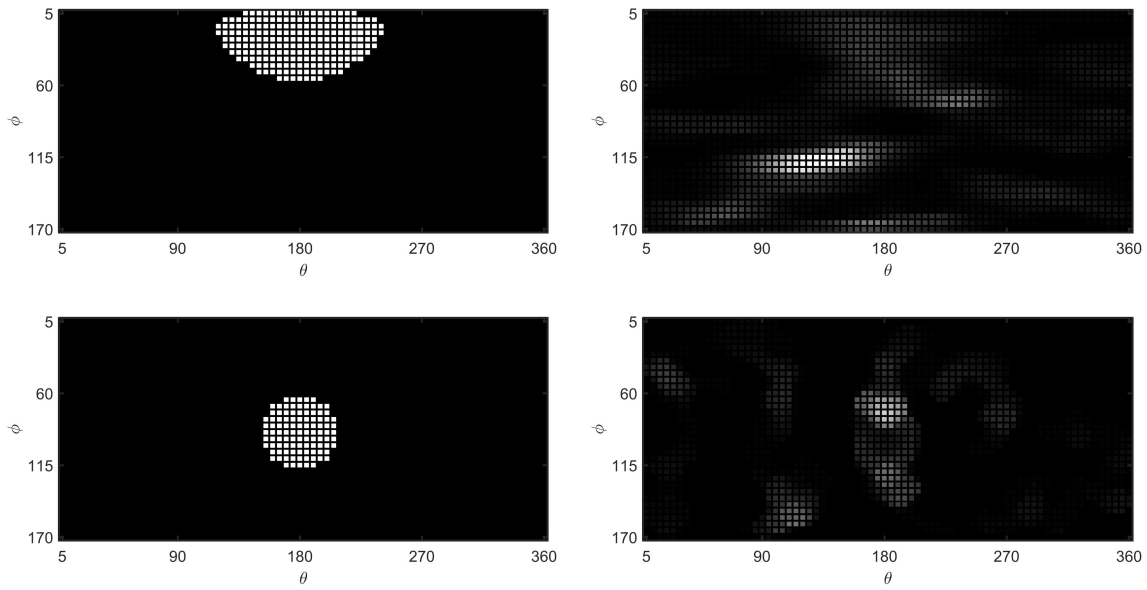


Figure 4.10. True source image (left) and its reconstruction (right) using ML-EM with MRP, $\beta = 10^{-4}$, for the larger disc located near the pole (top) and equator (bottom). Both reconstructions contained many phantoms and failed to capture the original shape's distribution.

improve performance for the more distributed sources, but a 5x5 pixel neighborhood with 5-degree resolution is rather unrealistic for most sources and would only further penalize distributions such as the ring, junction, and localized point sources. Therefore, responses were created in MCNP v6.2 [55] at 1-degree resolution so that larger neighborhoods may be used (thus allowing a simultaneous study on neighborhood size and angular resolution).

The four distributed sources in Figures 4.9 and 4.10 (sans point sources) were reconstructed at this higher resolution, using two different neighborhood sizes. The small disc and ring results are shown in Figure 4.11. On the left, a 3x3 neighborhood was used to calculate the median while the right used a 5x5 neighborhood; all other parameters, including convergence criteria, remained the same as the previous 5-degree resolution reconstructions.

Visually, it is clear that using higher resolution DRCs did not significantly improve

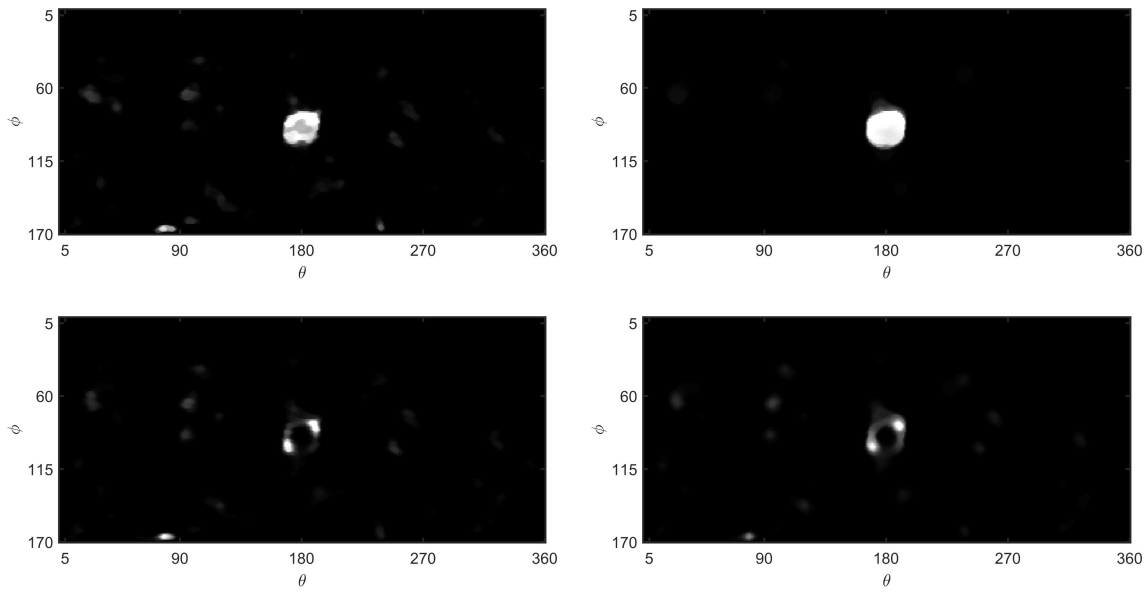
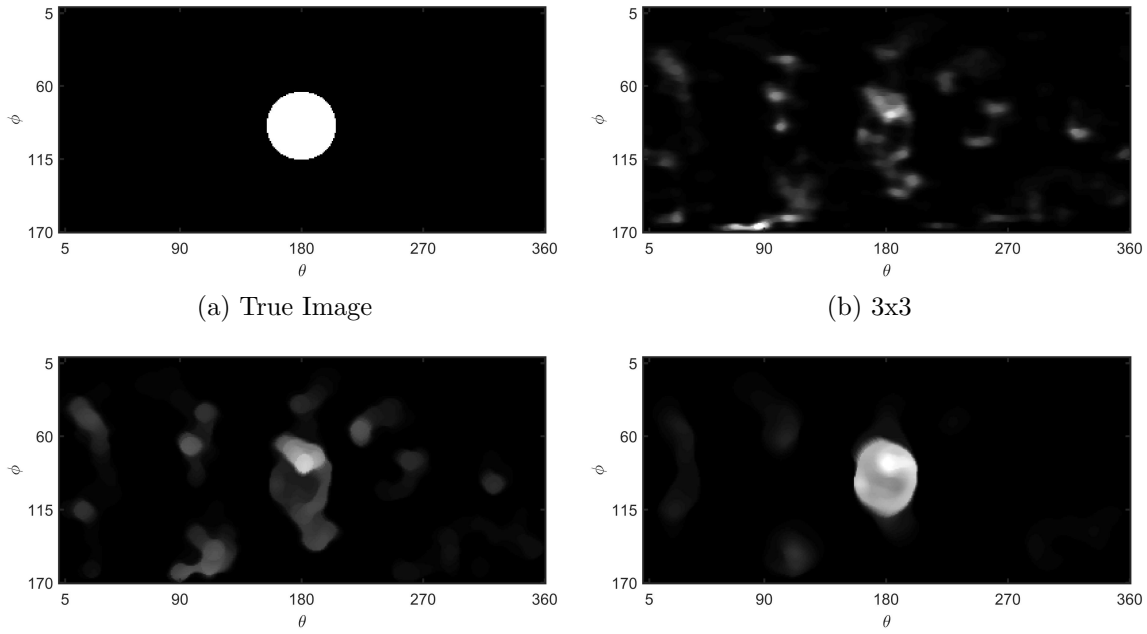


Figure 4.11. High-resolution, small disc and ring source reconstructions using a 3x3 (left) and 5x5 (right) neighborhood. The same $\beta = 10^{-4}$ and 1% convergence criterion were used for the ML-EM MRP algorithm. The higher resolution did not significantly improve the performance compared to the 5- degree resolution reconstructions, but the larger neighborhood did improve the reconstruction for each image. However, both the increased resolution and larger neighborhood significantly increased the computation time.

performance as the reconstructed images contained the same artifacts as the images generated from the lower resolution DRCs, including the phantom sources that appear in the reconstruction. However, the 5x5 neighborhood performed better in every scenario and reduced the noise within each image by penalizing small, noisy phantoms that appeared with the 3x3 neighborhood. Figure 4.12 shows reconstructions for the large disc at the equator for neighborhood sizes up to 7x7 pixels. The largest neighborhood size significantly improved the reconstruction without distorting the overall shape and size that occurred from oversmoothing at lower resolutions. It should be noted that although the 7x7 neighborhood, corresponding to 7-degree spans, is actually spatially smaller than the 3x3 neighborhood at 5-degree resolution, it performed better in localizing the reconstruction.

Finally, Figure 4.13 shows the reconstructions for the large disc near the pole



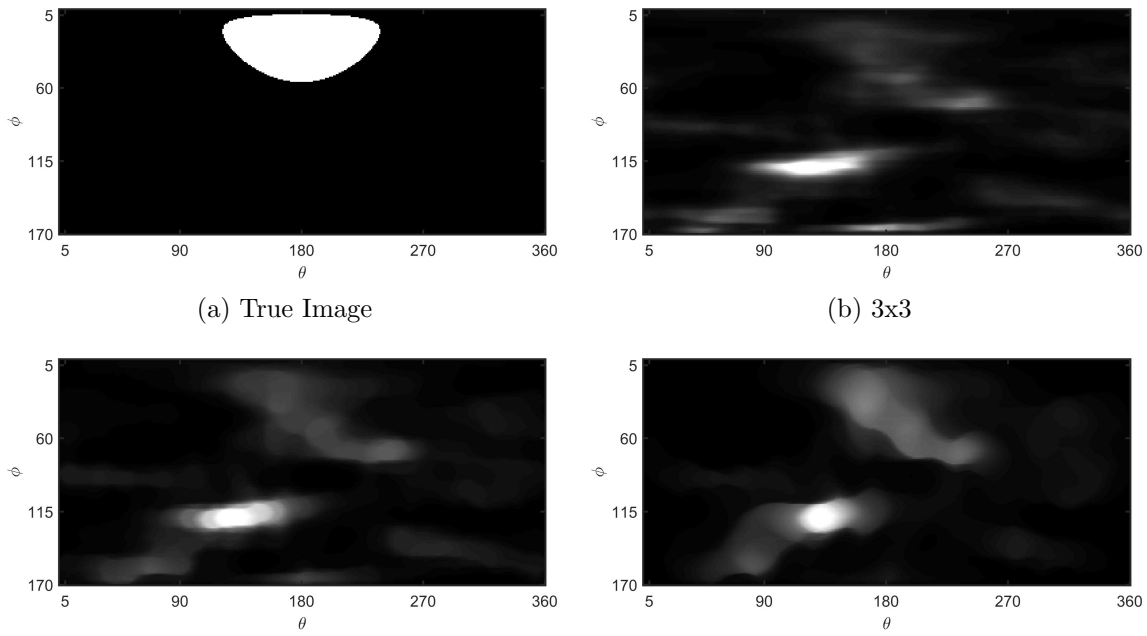
(a) True Image (b) 3x3

(c) 5x5 (d) 7x7

Figure 4.12. High-resolution reconstructions for the large disc source at the equator for increasing neighborhood sizes. The largest neighborhood size significantly improved the reconstruction quality, smoothing the image while preserving the overall size and shape of the disc.

for various neighborhood sizes. As with the other reconstructions, the image quality improved with larger neighborhood sizes. However, the reconstructions contained a very prominent aberration around $\theta = 135°$, $\phi = 110°$. The algorithm converged towards this aberration regardless of the specific neighborhood used. In application, there would be no method to discern which part of the image reflects phantoms and which are true sources, and this figure demonstrates perhaps the largest limitations with RSM imaging using the ML-EM MRP algorithm.

The improvements gained by the higher resolution and larger neighborhood sizes came at the cost of computational efficiency. Increasing the number of pixels by an order of magnitude increased the number of calculations required by roughly two orders of magnitude. The number of iterations required until convergence also increased, by almost two orders of magnitude for some of the images. These effects



(a) True Image (b) 3x3

(c) 5x5 (d) 7x7

Figure 4.13. High-resolution reconstructions for the large disc source near the pole for increasing neighborhood sizes. Visually, the reconstruction accuracy improved slightly with larger neighborhoods, but the aberration around $\theta = 135°$, $\phi = 110°$ remained.

caused the algorithm to take 100 to 1,000 times longer to reconstruct on the same computing system, requiring computation times from minutes at 5-degree resolution to many hours at the 1-degree resolution.

For applications requiring real-time analysis, these time scales are not practical. The increased resolution also requires longer detection times and more counts, further limiting the practicality for these high-resolution reconstructions. This is not to suggest that there is no scenario where such reconstructions are feasible: waste storage facilities, nuclear power plants, and isotope production facilities provide the optimal scenario for monitoring and characterizing the radioactivity at these higher resolutions. These results show that high-resolution reconstructions are possible and do not introduce any more artifacts than are present at lower resolutions. For applications limited in detection time or counts, the lower resolution reconstructions would provide the best real-time analysis without substantial information loss.

4.6 Conclusion

This is the first study to characterize the RSM as a gamma imager. Imaging with the RSM system would provide a low cost, efficient, and compact alternative to Compton cameras and coded-aperture devices. An ML-EM algorithm was derived for the RSM system based on principles for similar radiation detection scenarios. The MRP was implemented to improve the algorithm's performance by biasing the reconstructions towards more realistic source distributions.

A favorable prior weight of $\beta = 10^{-4}$ was chosen as larger values did not show any significant improvement in the distributed source reconstructions and only degraded the point source reconstructions. An analysis of variance was then conducted to determine if and how the source's shape, size, and direction relative to the RSM affected the reconstruction's relative error, precision, noise, and convergence time. It was determined that the most significant factor was the source's direction, with optimal performance for all FOM for sources near the equatorial region of the rotating mask. There was also shown to be small dependencies on the source's size for relative error and precision as smaller sources generally improved performance, although a strong correlation existed between the source size and neighborhood size used for the MRP.

The source's shape only significantly impacted the convergence time; distributions that matched the MRP assumption better generally decreased the number of iterations required until convergence. In all cases, there was an inherently large bias within the FOM, indicating that ML-EM performed poorly on average. Of the 24 distributed sources simulated and reconstructed, only three had relative errors of about 50% or less and most were centered about 175%. Increasing the resolution and neighborhood size generally improved the reconstruction quality, but at a significant cost to computational efficiency.

The overall poor performance reveals that further developments, either through alternative priors within ML-EM or alternative reconstruction algorithms altogether, are required to accurately and consistently reconstruct radioactive source images, particularly for widely distributed sources such as contaminated areas where the precise direction of the source and its extent are unknown. However, these results, particularly at high resolution, demonstrate that RSM imaging is possible and may serve as a practical alternative to other imagers if the analysis techniques can be improved. Alternative methods and algorithms, such as advanced machine learning techniques, are being investigated that impose priors and constraints that bias the reconstruction towards realistic source distributions. These methods may provide benefits over ML-EM as they do not require explicit derivations of the complex priors that govern "realistic" images, while still maintaining high predictive capabilities.

Acknowledgments

This research was supported by the Defense Threat Reduction Agency under grant HDTRA-18-27434 and 19-29109. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the United States Air Force, the Department of Defense, or the United States Government.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Appendix

4.A ML-EM Derivation

The detailed derivation of the ML-EM algorithm for the RSM system discussed in Section 4.4.1 is presented. This derivation follows the outline presented by Shepp and Vardi [44].

Suppose that a source is located in pixel j (corresponding to a specific direction relative to the detector) with an unknown mean emission of particles, λ_j. The probability of detecting n_j particles from that pixel is governed by Poisson statistics:

$$p\left(n_j\right) = e^{-\lambda_j} \frac{\lambda_j^{n_j}}{n_j!}. \tag{4.11}$$

In order to account for the mask's rotation and attenuation through the material, the probability p_{ij} is introduced, which indicates the probability that a particle from pixel j is detected when the mask is at rotation angle θ_i. Thus, the mean *effective* emission for a pixel at any specific mask angle is given by

$$\lambda_{ij} = \lambda_j p_{ij}, \tag{4.12}$$

and the probability of detecting n_{ij} particles is

$$p\left(n_{ij}\right) = e^{-\lambda_{ij}} \frac{\lambda_{ij}^{n_{ij}}}{n_{ij}!}. \tag{4.13}$$

For this work, p_{ij} was further defined to be the conditional probability of the particles *detected*, such that response curves described the probability density function for a detected particle from any given pixel. That is, the conditional probabilities require $\sum_i p_{ij} = 1$ for any pixel j (if a particle was detected, it must have been detected

by the RSM at some point along its rotation) and are the normalized probabilities calculated from the MCNP simulations. However, the normalization constant is different for each polar angle as more particles may be detected at some angles than others due to the varying attenuation. This varying normalization effectively resizes the detector response curves during the ML-EM iterations such that the final solution must be "un-normalized" to get the true number of *emitted*, rather than just detected, counts from the source.

The probabilities for each pixel are independent from one another. Therefore, the probability of detecting n total particles over all pixels is

$$P\left(n|\boldsymbol{\lambda}\right) = L(\boldsymbol{\lambda}) = \prod_{i,j} e^{-\lambda_{ij}} \frac{\lambda_{ij}^{n_{ij}}}{n_{ij}!}, \tag{4.14}$$

where the notation, $L(\boldsymbol{\lambda})$, has been used to define this probability as the likelihood of the image $\boldsymbol{\lambda}$ and $n = \sum_{i,j} n_{ij}$. However, it is much more convenient to work with the log likelihood, as it reduces the equation to a sum of independent components:

$$l(\boldsymbol{\lambda}) = \ln(L(\boldsymbol{\lambda})) = \sum_{i,j} \left(n_{ij} \ln(\lambda_{ij}) - \lambda_{ij} - \ln(n_{ij}!)\right). \tag{4.15}$$

From this equation, the ML-EM algorithm proceeds in two steps:

1. Expectation step: Using an initial guess $\boldsymbol{\lambda}^m$, where m is the iteration index, calculate the expected number of detected counts, n', given the actual number of counts detected, $\boldsymbol{y}$.

2. Maximization step: Maximize $l(\boldsymbol{\lambda})$ using the expected number of counts from the previous step in order to find $\boldsymbol{\lambda}^{m+1}$.

The maximized $\boldsymbol{\lambda}^{m+1}$ becomes the initial guess for the next expectation step; this is repeated until $\boldsymbol{\lambda}$ has converged.

As the pixels are independent, the expected number of counts from a Poisson distribution given a mean effective emission λ_{ij} and a *detected* number of counts, y_i, for the mask at angle θ_i reduces accordingly from Bayesian statistics to [63]

$$n'_{ij} = E[n_{ij}|y_i, \lambda_{ij}^m] = \frac{y_i \lambda_{ij}^m}{\sum\limits_k \lambda_{ik}^m} = \frac{y_i \lambda_j^m p_{ij}}{\sum\limits_k \lambda_k^m p_{ik}}. \tag{4.16}$$

Note that the dummy index k in the denominator is equivalent to the pixel index j. The derivative of $l(\boldsymbol{\lambda})$ with respect to λ_j is

$$\frac{\partial l}{\partial \lambda_j} = \sum_i (\frac{n_{ij}}{\lambda_j} - p_{ij}), \tag{4.17}$$

where Equation 4.12 was substituted into Equation 4.15. Setting the derivative equal to zero and solving for λ_j yields

$$\lambda_j = \frac{\sum\limits_i n_{ij}}{\sum\limits_i p_{ij}}. \tag{4.18}$$

Substituting in the expected value n'_{ij} from Equation 4.16 for n_{ij} yields the ML-EM algorithm for the RSM system:

$$\lambda_j^{m+1} = \frac{\lambda_j^m}{\sum\limits_i p_{ij}} \sum_i \frac{y_i p_{ij}}{\sum\limits_k \lambda_k^m p_{ik}}. \tag{4.19}$$

4.B Prior Weighting and Numerical Stability

In the Median Root Prior, a Gaussian prior was implemented of the form

$$P = e^{-\beta \frac{\left(\lambda_j - M_j\right)^2}{2M_j}}, \tag{4.20}$$

where λ_j is the activity of the jth pixel, M_j is the median of a neighborhood surrounding that pixel, and β is the prior weighting. The prior weight effectively changes the width of the Gaussian, which is centered around M_j and has a variance equal to M_j/β. For very large β, the variance approaches zero and forces the pixel to take on the median value. Conversely, as β approaches zero, the variance approaches infinity and the pixel is allowed to take on any value (the prior has a value of one for any pixel and median).

By definition, there is no limit to the range for β other than requiring it to be non-negative. However, most applications often limit the prior weight between zero and one, a seemingly arbitrary yet reasonable decision. There are two subtleties that justify this choice that are often not addressed; this section seeks to address any concerns about the range of the prior.

The two justifications are grounded in physical uncertainty limitations and numerical stability. In the former, the MRP is implemented in a radiation detection scenario, an application that is governed by Poisson statistics. In Poisson processes, the standard deviation is the square root of the counts; larger values may be permitted due to uncontrollable noise but smaller uncertainties are not justifiably representative of the actual uncertainty. Perhaps not coincidentally, limiting β from zero to one also limits the standard deviation of the prior to a lower limit of the square root of the median number of counts, which is precisely what would be expected for a Poisson distribution.

The second justification is perhaps even more subtle but also more significant when applied to successfully implementing ML-EM. The algorithm, including the Median Root Prior, may be written as the condensed form

$$\lambda_j^{m+1} = \frac{\mathrm{A}\ \lambda_j^m}{\sum_i p_{ij} + \beta \frac{\lambda_j^m - M_j}{M_j}}, \tag{4.21}$$

where many of the terms have been condensed into the variable A. It is not unreasonable to imagine a scenario where a small, but nonzero λ_j exists such that $\lambda_j \ll M_j$ so that the algorithm for one iteration becomes

$$\lambda_j^{m+1} \approx \frac{\mathrm{A}\ \lambda_j^m}{\sum_i p_{ij} - \beta}. \tag{4.22}$$

The positivity constraint requires that $\beta \leq \sum_i p_{ij}$, but if $\beta \approx \sum_i p_{ij}$, the denominator becomes extremely small and the next iteration's pixel value immediately increases to a significantly large value. Because this new value would likely be greater than the median, the next iteration would have an extremely large denominator, significantly shrinking the pixel's magnitude in the next iteration and repeating the process continuously such that convergence is not reached.

This process demonstrates the *computational* limitations for numerical stability if either the prior weight or the probabilities are not conditioned properly. Using the conditional probabilities so that $\sum_i p_{ij} = 1$ simplifies the analysis by creating an upper limit of $\beta = 1$ regardless of the specific j index being used. In conjunction with the physical justifications, limiting β from zero to one (or near one) is not just an arbitrary decision but rather a logical result of the physical limitations and numerical mathematics.

V. ReGenerative Neural Network Reconstructions

This chapter is directly derived from a paper submitted to *Radiation Measurements*, under the title "Regenerative Neural Network for Rotating Scatter Mask Radiation Imaging." The collaborating authors are Robert J. Olesen, James B. Cole, Darren E. Holland, Erik M. Brubaker, and James E. Bevins. While the format of the journal article was modified to match the dissertation style, the contents of this chapter remain unaltered.

The article focuses on designing and testing the neural network model for converting simulated DRCs into a radiation image. These results are also compared to the ML-EM results as a baseline for performance. As the chapter presents the paper in its entirety, much of the background has already been discussed in the previous chapters. For readers who are already familiar with the RSM and neural network concepts, Section 5.4 provides a more detailed discussion on the specific neural network architecture developed, called ReGeNN, while Section 5.5 discusses the results and comparisons to ML-EM.

5.1 Abstract

This paper describes the design and testing of a new form of convolutional neural network, a regenerative neural network (ReGeNN), for application to rotating scatter mask gamma imaging. The network was trained using detector responses for realistic source distributions simulated in MCNP v6.1.4. ReGeNN was shown to reconstruct the source images with excellent quality when trained under ideal conditions. When comparing to standard maximum-likelihood expectation-maximization algorithms, ReGeNN reduced the relative error from 145% to 33% and increased the precision from 27% to 85% averaged over the 24 distributed sources tested. The

network also demonstrated robust learning capabilities after successfully training on noisy input data, with only relatively minor degradation to the source reconstruction quality. An analysis of variance study determined that the most significant factor affecting the reconstruction quality was the source's shape, with ring-type source distributions having the worst performance. The interaction between the source's size and direction was also discovered to have a small effect as larger sources located near the bottom of the system's field-of-view contained more phantoms within the reconstruction. Reconstruction quality was lower for responses exceeding the training noise level and for source distributions not included in the training set, indicating the importance of robust training data. The results show a significant improvement over more conventional algorithms, suggesting that real-time gamma imaging with the rotating scatter mask may be not only plausible, but practical for the first time. ReGeNN may readily be adapted for similar time-encoded radiation imaging systems, but the neural network methods described also have significant application potential towards other imaging systems.

5.2 Introduction

Monitoring, localizing, and characterizing radioactive sources remains a priority in applications for nuclear security and safety [4, 57, 65, 66]. Radiation imaging has become a staple in the biomedical community, but such advances resulted from the source's well-defined and constrained nature in such applications [67]. For nuclear security and safety, the composition and location of the source is not necessarily constrained, and interference from environmental shielding, scattering, and background radiation further complicates the ability to monitor, localize, and/or characterize radioactive sources. Overcoming these challenges requires a device with large field of view (FOV), high detection efficiency, and large signal-to-noise ratio (SNR), but these

parameters are often in direct conflict with each other. Traditional coded-aperture systems are often limited in FOV and efficiency [7], although recent developments have shown some improvements through alternative detector arrangements and reconstruction algorithms [10, 51]. Compton cameras continue to have excellent SNR and FOV, but at a higher cost to efficiency and being unable to perform dual-particle imaging [7, 13, 14, 15]. Building on the principles from Compton cameras, novel dual-particle scatter cameras are currently being investigated that may remedy some of these limitations [16, 17].

This study evaluates an alternative, novel imaging approach through the rotating scatter mask (RSM) system. The RSM system is one of many recently-developed time-encoded imaging systems, which converts a source's spatial information into a lower-dimensional temporal signal. Such systems benefit from increased efficiencies and reduced complexities, often containing only a single, standard detector [18, 19, 20, 21, 42]. Previous development and evaluation has demonstrated the RSM as a directional, dual-particle detection system with relatively high detection efficiencies and nearly full 4π FOV [25, 27, 28, 29, 30]. A previous study indicated that gamma imaging may be possible through appropriate reconstruction algorithms, without requiring any changes to the physical setup or data collection [68]. However, that study discovered that the maximum-likelihood expectation-maximization (ML-EM) algorithm, a technique often used throughout a variety of radiation imaging systems, failed to accurately reconstruct most source images, indicating that alternative priors or more advanced algorithms are required to enable RSM imaging capabilities. These results are discussed further in Section 5.4.

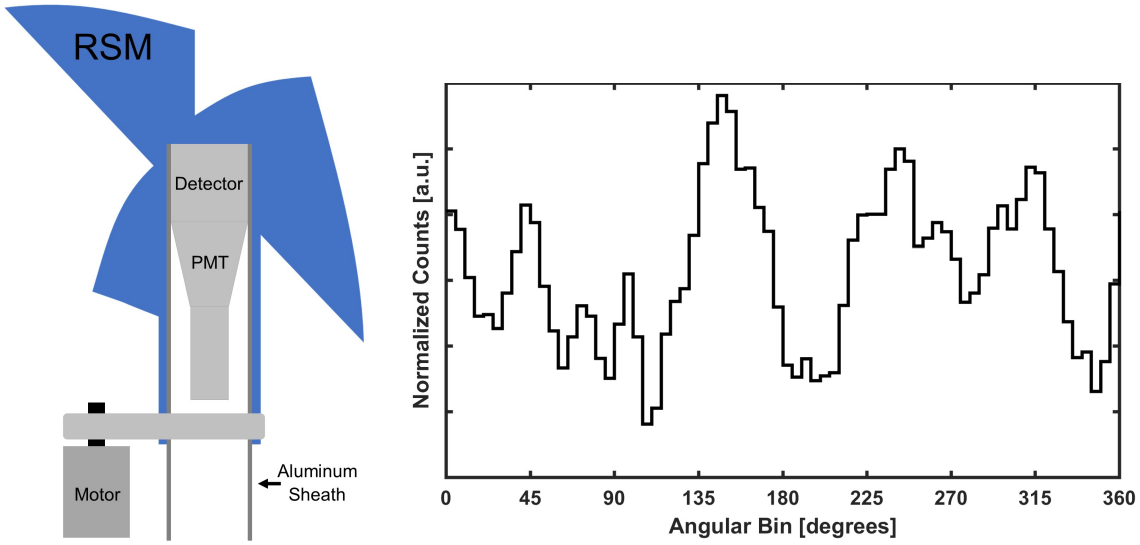
Artificial neural networks are prime candidates for this task, particularly convolutional neural networks (CNNs) designed specifically for image reconstruction and deconvolution [54]. While a CNN architecture has been proposed and successfully in-

tegrated for a coded-aperture system [51], there is relatively little literature regarding the appropriate design and training of a CNN for radiation imaging applications as compared to the comprehensive list of studies for more conventional algorithms such as ML-EM. In this paper, a novel CNN encoder-decoder architecture is proposed as an alternative imaging algorithm for the RSM system, which can be easily modified to fit other radiation imaging applications. The CNN architecture is described in detail, with all hyper-parameters defined and outlined for readers to successfully reproduce and alter the design for specific radiation imaging applications. The imaging performance with the proposed CNN architecture is compared to the previous study using ML-EM as a benchmark to highlight the increased image reconstruction performance.

5.3 Rotating Scatter Mask Imaging

The RSM system uses a spatially-varying attenuator, referred to as the mask, surrounding a standard scintillating detector, as shown in Figure 5.1a. The mask is connected to a motor so that it rotates along the detector's longitudinal axis; in the presence of a stationary radioactive source, this rotation generates a periodic, time-varying detector response. A detector response curve (DRC) is generated by recording the detected events versus rotation angle (Figure 5.1b), which ideally can be used to reproduce the source image. This process is similar to other time-encoded imaging systems [19, 20, 21, 42].

For this study, the mask design known as the "Mace" was chosen, in accordance with the ML-EM study previously mentioned [68]. The Mace geometry is shown in Figure 5.2. The DRCs were generated using the 662 keV full-energy peak from a simulated ^{137}Cs point source located 86.36 cm from the center of a 3" x 3" NaI(Tl) detector, modeled in MCNP v6.1.4, with a maximum 2% relative uncertainty for any source direction. The DRCs span from 5° – 170° in the polar angle ϕ (referenced from



(a) (b)

Figure 5.1. Schematic representing the RSM system's primary components (left). An acrylic-like mask (labelled RSM) slides over an aluminum sheath surrounding a scintillating detector and is attached to a motor so that it may freely rotate about the detector, generating a time-varying DRC (right).

the axis of rotation) and the full 0° – 360° in the azimuthal (rotational) angle θ, both in 5° increments, for a total of 2,448 point source directions and corresponding DRCs. Distributed sources were modeled as a collection of point sources; the total detector response was modeled as a weighted sum of the individual point source DRCs [68].

5.4 Methodology

Collapsing the source's spatial information into a temporal DRC forms an underdetermined system such that a single DRC can correspond to many, if not infinite, source distributions [28]. It is required, then, to define an algorithm that narrows the search space to those solutions that correspond to "realistic" or "expected" distributions. For ML-EM, the algorithm is designed to find the solution that maximizes the probability of observing the measured DRC [43, 44]. However, the most probable detected sources through ML-EM are often dispersed or noisy, and the mask design itself can play a critical role in biasing the results (a source is more likely to

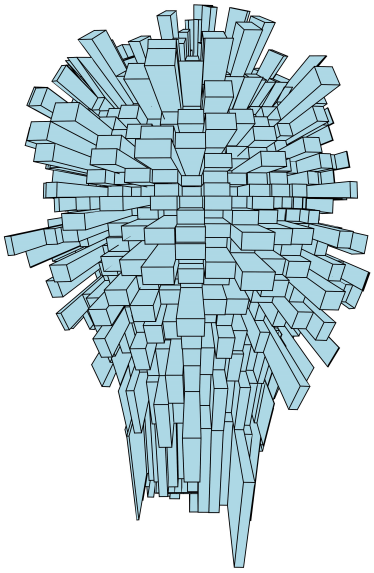


Figure 5.2. Design for the Mace RSM. The complex geometry theoretically improves the imaging capability by increasing the orthogonality between DRCs.

be detected if it goes through a thinner portion). Many man-made sources are often compact with roughly uniform activity; to account for this, many ML-EM algorithms employ a prior, such as the median root prior (MRP), that biases the algorithm to converge towards smoother, localized distributions [47].

For systems like coded apertures or medical devices, which have 2D inputs, this can greatly improve reconstruction quality. For the RSM, which has a 1D input, ML-EM with MRP only improves the reconstructions for a specific set of source distributions, making it impractical for generalized gamma imaging [68]. Implementing a more rigorous prior may improve performance, but quantifying what is "realistic" or "man-made" is often difficult to put in mathematical form. Artificial neural networks have risen almost precisely for this scenario through the application of what is known as the universal approximation theorem.

A version of the theorem states that multi-layer feed-forwards networks, such as CNNs, with locally-bounded piecewise continuous activation functions can approximate any continuous function to any degree of accuracy, as long as the activation

function is not a polynomial [48]. This means that an artificial neural network can be trained to approximate the distribution pertaining to "realistic" sources without knowing the form of that distribution explicitly (the approximation is instead dictated by the given training data). A well-designed neural network integrates this into a single network that both learns the distribution and how to process the data. Of course, one criticism of this method is that the distribution is hidden within the network, giving these approaches the "black box" interpretation. Another limitation, which is important to note, is that while the universal theorem states that these networks can approximate any continuous function, developing the appropriate network that can do so is not guaranteed. As the task becomes more complex, requiring more complex architectures, the networks become extremely sensitive to hyper-parameters within the model and can become unstable or over-fit to the training data.

Despite these criticisms and limitations, many applications have already made use of artificial neural networks. They have successfully been integrated into gamma and neutron spectroscopy [49, 50], and a recent study has demonstrated their potential in radiation imaging for coded-aperture systems [51]. Of interest to this study are CNNs due to their rising popularity as the state-of-the-art approach to image deconvolution and generation [53, 54]. The following section describes a novel CNN architecture that adapts the approach designed for computer vision image generation into a radiation image reconstruction algorithm.

5.4.1 Neural Network Architecture

The CNN architecture used for RSM imaging is shown in Figure 5.3, which borrows heavily from generative CNNs [53]. The generative CNN is designed to create generic classes of images (such as chairs, tables, and cars) using a predefined one-dimensional array that encodes information about the image. For radiation imaging, the source's

two-dimensional spatial information is only compressed into a one-dimensional DRC; the individual values in the DRC do not serve as any specific encoding. This difference is subtle; both applications use supervised learning, but the difference between encoding and compression plays a crucial role when creating appropriate training data and designing the network's architecture. To distinguish this neural network for radiation imaging from other CNN applications, this proposed architecture will be referred to as the regenerative neural network (ReGeNN) throughout this study.

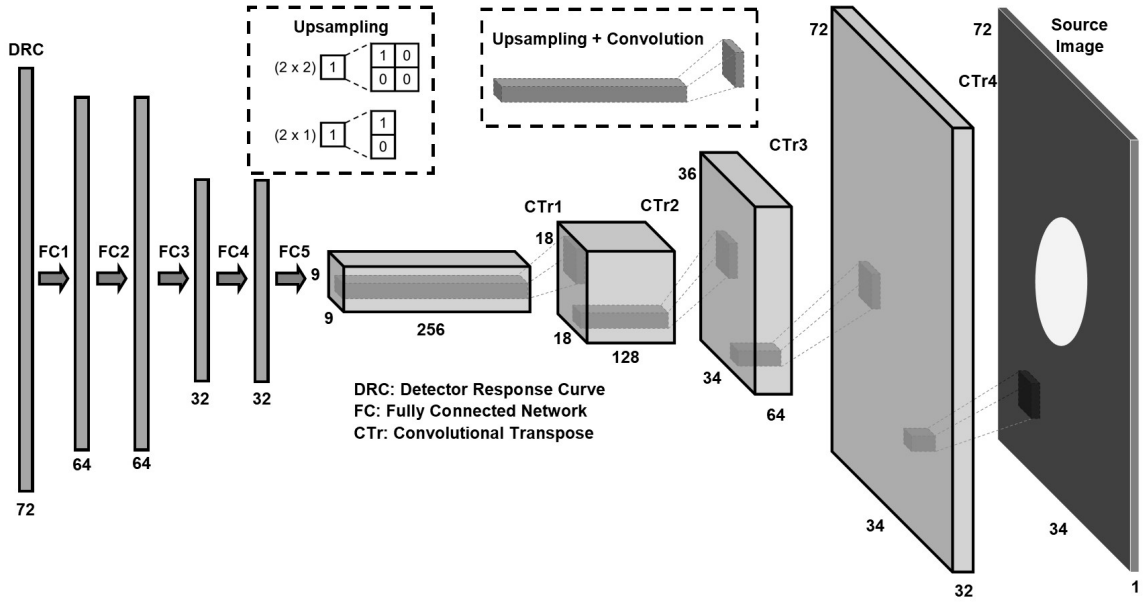


Figure 5.3. Architecture for the ReGeNN used to convert the DRC back into the source image. ReGeNN utilizes fully connected layers as an encoder followed by "convolutional transpose" layers as a decoder that upsamples into the reconstructed image.

ReGeNN is composed of nine total layers, split into an "encoder-decoder" format with the one-dimensional DRC as the input and the two-dimensional source reconstruction as the output. The number of layers was chosen based on a previously successful generative neural network [53], formatting the layers to match the 72-valued DRC input and desired 72 x 34 (θ x ϕ) output image resolution. The numbers shown in Figure 5.3 indicate the number of elements for each dimension in the layers. The first five layers form the "encoder" portion, using fully-connected layers (labelled

FC), while the other four layers use a combination of upsampling and convolution to form the convolutional transpose layers (labelled CTr) that comprise the "decoder" portion.

To convert from the one-dimensional encoder to the two-dimensional image, the fifth fully connected layer transforms the previous 32 x 1 array into a 20,736 x 1 array, which is then reshaped into a 9 x 9 x 256 array. During convolution the width refers to number of filters used for each layer; for example, CTr1 uses 256 different filters that produce their own 9 x 9 "image". For upsampling, a single pixel is converted into a 2 x 2 or 2 x 1 array, with one pixel equal to the original pixel value and zero elsewhere. CTr1 and CTr2 use a 2 x 2 upsampling, followed by a 2 x 1 upsampling for the remaining layers.

In order to increase the number of layers while preserving the original image's polar angular resolution (34 bins), the first and last columns were removed after the CTr2 upsampling to form a 36 x 34 multichannel image, rather than the 36 x 36 multichannel image that would result from 2 x 2 upsampling. Note that cropping occurs after the convolution and should not affect network performance (ideally, the network would learn that those columns provide no additional information and would weight those columns less). Future RSM designs could match resolution to layer upscaling to provide better network designs that would eliminate the need to crop the data. CTr3 conducts the final upsampling to form the original 72 x 34 image resolution, while the last CTr4 layer collapses the remaining 32 filters into the final reconstructed image. Each convolution uses a 3 x 3 filter, with the number of filters subsequently halved after each layer (until the final output layer). A stride of one with zero padding on the edges was used to preserve the size after each convolution.

Each hidden layer used the ReLU activation function, defined as $f(z) = \max(0, z)$, while the final output layer used the sigmoid activation function, defined as $f(z) =$

$1/\left(1+e^{z}\right)$. The decision to use the sigmoid activation function as the final layer was driven by defining each pixel to represent the relative source activity so that the output range was constrained to [0,1]. Training was conducted in mini-batches of 128 samples each, with batch normalization conducted after each activation function to improve convergence [69]. It was heuristically determined that setting the momentum for the batch normalization to 0.8 was ideal for this application. Training was conducted using the Adam optimizer (an efficient variant of stochastic gradient descent) with the learning rate set to 0.0001, $\beta_1 = 0.5$, and $\beta_2 = 0.999$ [70]. Readers knowledgeable on the Adam optimizer may note that $\beta_1 = 0.5$ is lower than what is typically recommended; the default value of 0.9 often resulted in training instability and was lowered based on another CNN architecture design [71].

The cross-entropy was chosen for the loss function as it is sensitive to small differences between the ideal and reconstructed images, which improves the overall training rate. For an image with arbitrary pixel values, the pixels are binned into various classes (i.e. intensities), and the cross-entropy loss is defined as

$$\text{Loss} = -\sum_{j}\sum_{c} \hat{\lambda}_{j,c}\, \log(\lambda_{j,c}), \tag{5.1}$$

where $\hat{\lambda}_{j,c}$ and $\lambda_{j,c}$ are the true and reconstructed pixel values, respectively, summing over every pixel, j, for each class, c. By design, the pixels in the true source images for this study were either "off" or "on", with pixel intensities of 0 or 1, respectively. The output from the neural network represents the probability that the pixel is on (i.e. $\lambda_{j,1} \equiv \lambda_j$) and the probabilities must sum to one such that $\lambda_{j,0} + \lambda_{j,1} = 1$ or $\lambda_{j,0} = 1 - \lambda_j$, where the class index has been dropped. Substituting these values into Equation 5.1 yields what is more generally known as the binary cross-entropy loss

function:

$$\text{Loss} = -\sum_{j} \left(\hat{\lambda}_j \, \log(\lambda_j) + (1 - \hat{\lambda}_j) \, \log(1 - \lambda_j) \right), \tag{5.2}$$

which was used as the loss function for ReGeNN.

The binary cross-entropy loss function proved more efficient than a more traditional mean-square error loss function. When using the mean-square error as the loss function, the network would often converge to a monotonic image with each pixel set to 1/2,448 (the total number of pixels), effectively averaging out the solution space and converging to a local minimum. As the binary cross-entropy is logarithmic, small deviations from the correct pixel value created larger loss values compared to the mean-square error, which in turn sped up the training and improved performance.

5.4.2 Training Data

The specific source distributions are discussed in detail in the next section. In accordance with the previous ML-EM study, five unique shapes (four distributed and one point source) were tested. The distributed sources spanned five discrete sizes ranging from $10^\circ - 50^\circ$ angular extent (in 10° increments) as measured from the source's center. At a 72 x 34 image resolution, corresponding to 5° resolution in both the azimuthal and polar angle respectively, a total of 2,448 source directions were possible with 51,408 unique source variations in shape, size, and direction. Of these variations, 45,000 were included as part of the training data while the remaining 6,408 were used as a validation set to ensure that the network was not over-fitting to the training data. The images and DRCs used to evaluate ReGeNN were not included in the training data to prevent biasing, providing a better representation for the generalizing capability of the network.

Training was conducted in mini-batches of 128 training images throughout each epoch (an epoch is defined as one complete run through all the training data). The

training images used for each mini-batch were randomly sampled from the complete training set in each epoch to prevent biasing caused by any local clustering of similar source distributions in the mini-batch of a particular epoch. All training was conducted using the Keras platform, using Tensorflow as the backend [72]. It is recommended for networks of this size to use the GPU as the processing unit, with this study finding a 10-fold decrease in run time when using a 1,280 CUDA core GPU over a quad core 4 GHz CPU. A repository with the code used to construct, train, and evaluate ReGeNN is available, as the network may readily be applied to similar time-encoded techniques with minor adjustments to the input/output size and may be used as a foundation for other imaging techniques [73].

5.4.3 Performance Criteria

An analysis of variance (ANOVA) was conducted to determine the most significant factors associated with ReGeNN performance and to provide a quantitative comparison to the ML-EM study, which goes into further detail on ANOVA principles [68]. For the purpose of this work, it is important to note that ANOVA is a statistical test that compares two or more population means using the probability of the measured variances, denoted by the p-value. ANOVA can be used to generate a predictive model of the mean; for example, the quadratic model can be defined as

$$y = c_1 x_1 + c_2 x_2 + c_{12} x_1 x_2 + c_{11} x_1^2 + c_{22} x_2^2 + \text{BIAS}, \tag{5.3}$$

where x_1 and x_2 are chosen factors to predict the mean value, y. The constants of proportionality c_1 and c_2 represent the first-order effects, c_{12} represents the first-order interaction effect, and c_{11} and c_{22} represent the second-order effects. Under the null hypothesis, the mean is invariant to the factor effects, with the constants of proportionality equal to zero. For this study, any factor with a p-value less than

0.10 was deemed statistically significant. As with the ML-EM ANOVA study, this threshold was chosen to account for reasonably-sized effects without over-generalizing the results.

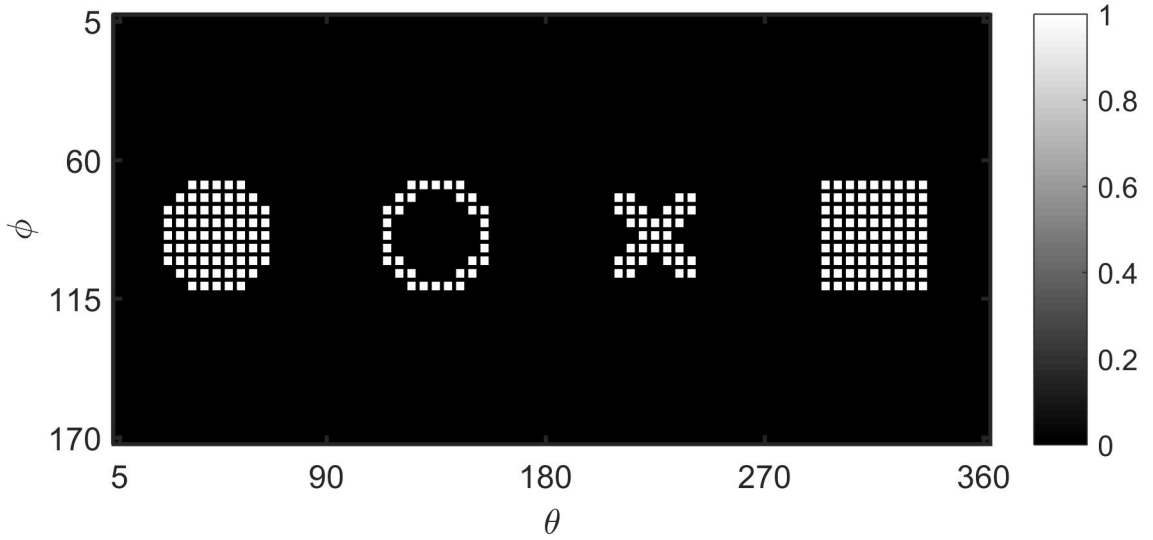


Figure 5.4. From left to right the four source shapes used: disc, ring, junction, and square (examples shown use a 40° angular extent, with the center polar angle at 90°). In this study, the mathematical convention for spherical coordinates was used where ϕ is the polar angle and θ is the azimuthal angle. The color scale indicates the pixel's relative activity.

Three factors were chosen to test under ANOVA: the source's 1) shape, 2) size, and 3) center polar angle. The four shapes tested, plotted in Figure 5.4, were chosen to mimic realistic source distributions:

1. Disc: Corresponding to a cross-section of a cylinder or sphere. This mimics nuclear waste containers used in storage as well as spherical forms of special nuclear material.

2. Ring: Corresponding to a hollow cylinder. This mimics contaminated storage drums as well as the annular shape commonly used to store special nuclear material.

3. Junctions ("X" shape): Used to simulate contaminated pipes or multiple fuel

rods.

4. Square: Used to capture sharp corners that are not available in the other images. Extensions of this shape include rectangles that mimic the cross-section of cylinders including straight pipes, fuel rods, and waste storage drums.

The source size was defined as the maximum angular extent referenced from the source's center direction, with two discrete values of 30° and 50° chosen to be included in the ANOVA test. Finally, the source direction was defined as the source's center polar angle direction, with three discrete values of $\phi = 30°$, 90°, and 150° included in the ANOVA test. These factor levels correspond to the previous ML-EM study, allowing for a one-to-one comparison between the two algorithms. A full factorial model was conducted, corresponding to 24 unique source distributions. Three point sources were included at the designated polar angles, but were not included in the ANOVA to prevent biasing (as point sources, by definition, do not have a size).

The performance criteria were defined with the following figures of merit (FOM):

$$E = \frac{\sum_j |\hat{\lambda}_j - \lambda_j|}{\sum_j \hat{\lambda}_j}, \tag{5.4}$$

$$\text{Precision} = \frac{\sum_j \hat{\lambda}_j \lambda_j}{\sum_j \lambda_j}, \tag{5.5}$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left(\hat{\lambda}_j \lambda_j - \mu \right)^2}, \tag{5.6}$$

where E is the relative error between the true and reconstructed images, precision is defined as the ratio between the sum of pixels (corresponding to activity) contained within the region of interest (ROI) and the sum of pixels in the entire reconstructed image, and σ is defined as the standard deviation within the reconstructed image ROI.

For this study, the ROI was defined as the pixels corresponding to the original source distribution, and μ was defined as the average value of $\hat{\lambda}_j \lambda_j$ within the ROI. The equations assume that the images have been normalized to the maximum activity, and the true source distribution is monotonic (uniform activity). An ideal image would have low relative error, high precision, and low standard deviation.

Astute observers may note that the convergence time has been excluded when comparing to the ML-EM analysis. Unlike iterative algorithms such as ML-EM, where the convergence time directly corresponds to the post-processing wait time for predicting an image after obtaining the measurement, the training of ReGeNN is completely pre-processed. After the network has been trained, the image reconstruction is near-instantaneous, providing *real-time* analysis. This is an automatic benefit gained from using a neural network over other iterative techniques.

5.5 Results

5.5.1 Neural Network Training

ReGeNN was trained over 1,000 epochs, equating to 352,000 total updates to the network when using the mini-batch mode. The training history is plotted in Figure 5.5, along with the validation history after each epoch. These plots are used to verify that the network is training appropriately and converging to a solution (training loss trends towards zero) without over-fitting (validation loss trends towards zero near the same rate as the training loss). The plot shows that both loss values are converging towards a solution, although there appears to be instabilities within the validation set, as indicated by the sharp spikes present in the validation loss curve.

These instabilities are likely caused by the stochastic training method, with each spike corresponding to a small perturbation towards a local minimum. This is supported by the inset within Figure 5.5, which shows the training and validation loss

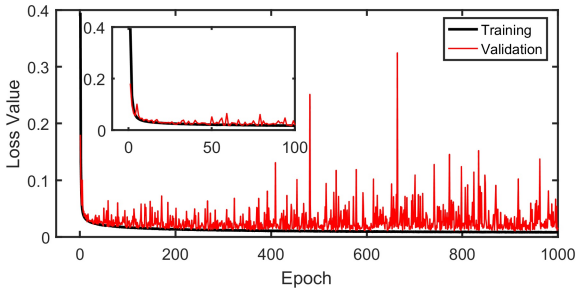


Figure 5.5. ReGeNN training history using the 45,000 ideal DRCs as the training set (values shown are the average loss value after each epoch). The spikes in the validation set are a result of the stochastic method of the Adam optimizer and using the logarithmic binary cross-entropy loss function. The first 100 epochs (inset) show generally good agreement between the two loss curves.

for the first 100 epochs. While small jumps exist within the validation curve, the overall trend is relatively smooth and the validation loss is approximately equal to the training loss. The extent of the "instabilities" are also exaggerated by the plot scaling; Figure 5.6 shows the loss history from 600 to 700 epochs, centered about the largest spike in the validation curve. This figure shows that the spike only occurs for one epoch before dropping back down to near-training loss values. This is indicative of the stochastic nature of using Adam as the optimizer whereby the network momentarily trains towards a local minimum based on the training data, moving away from the global minimum that would include the validation data. Using the binary cross-entropy also further accentuates this small perturbation as the logarithmic nature can cause relatively small differences to have large loss values (although the loss values shown are still relatively small such that any epoch would likely produce meaningful results). Continuing the training moves the network back towards the global solution, and the presence of local minima perturbations in this study does not significantly impact the results presented.

While not used for updating the network, the accuracy throughout the training was tracked. Here, accuracy is defined as the mean percentage of pixels that were "classified" correctly using a 0.5 threshold (0.5 probability that the pixel is "on"). The

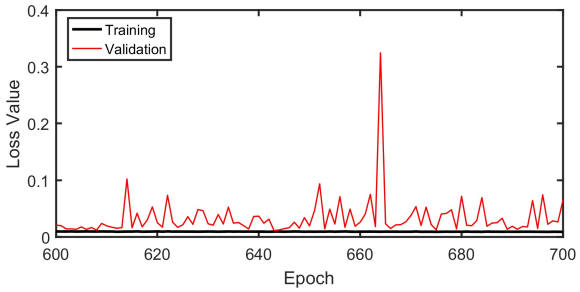


Figure 5.6. Zoomed in view of the ReGeNN training history between 600 and 700 epochs, showing the largest spike in the validation loss. However, the spike occurs for only one epoch, demonstrating the stochastic nature and verifying that the network is not over-fitting to the training data.

accuracy history is plotted in Figure 5.7. The perturbations from the binary cross-entropy loss history remain present in the accuracy curves, although at significantly smaller magnitudes. As the accuracy is more representative of a qualitative visual performance, the plot verifies that the network is training appropriately and the large magnitude of the spikes in the loss function are primarily caused by the sensitivity of the logarithmic binary cross-entropy function and do not indicate that the network is significantly over-fitting.

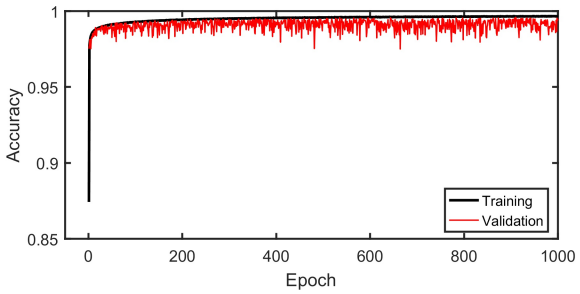


Figure 5.7. Average accuracy values for the training and validation datasets over each epoch. While small perturbations remain in the validation history, the magnitude is much smaller than the equivalent binary cross-entropy loss value.

The training can also be evaluated qualitatively. Visually, the original source distributions (Figure 5.8a) and the ReGeNN reconstructions after 1,000 epochs (Figure 5.8b) show remarkable agreement. Only two images deviate significantly (visually) from the original: the point source and large ring located at $\phi = 150°$, with the former

identifying a completely different point source direction and the latter generating a secondary "phantom" ring source near the top of the image. However, the overall excellent agreement between the two figures shows that ReGeNN converged properly as none of these images were included in the training, demonstrating the network's generalizing capabilities.

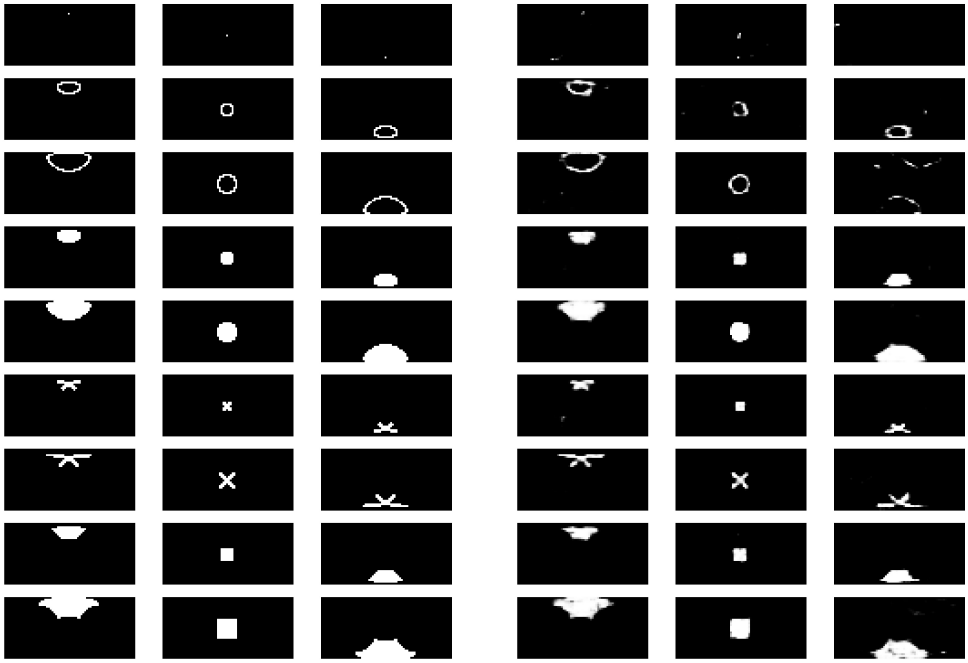
5.5.2 Image Reconstruction ANOVA

The ANOVA results for the 24 distributed sources of Figure 5.8 are presented in Table 5.1. These results show that most factors had p-values significantly larger than the 0.10 threshold chosen, so that the null hypothesis that they have no effect on the mean could not be rejected. However, the shape had p-values less than 0.10 for all three FOM, while the shape and direction interaction term (denoted by *) also fell below the significance threshold for the relative error and precision. Thus, the null hypothesis may be rejected, and there is statistical support that these factors have an effect in their respective FOM. It should be noted that some p-values, such as the size term for precision and direction-squared term for σ, are close to the threshold and may warrant further investigation in future studies regarding RSM design.

Table 5.1. FOM ANOVA for ReGeNN Reconstructions

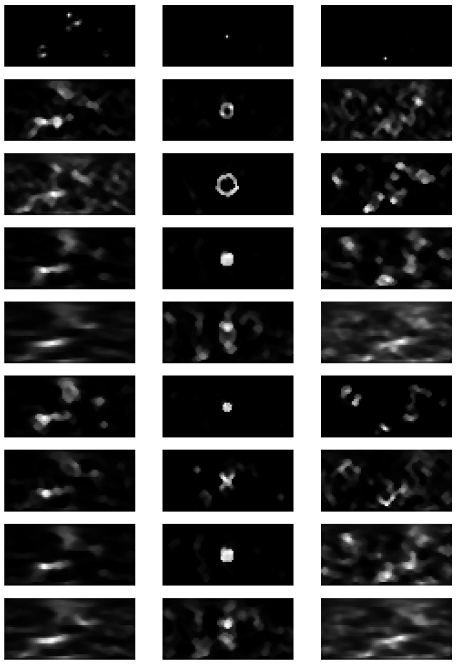
Factor	p-value		
	E	Precision	σ
Shape	$<$ **0.0001**	$<$ **0.0001**	**0.08**
Size	0.30	0.11	0.91
Direction	0.74	0.90	0.77
Shape*Size	0.93	0.87	0.78
Shape*Direction	0.92	0.59	0.69
Size*Direction	**0.02**	**0.02**	0.23
Direction^2	0.39	0.38	0.14

To understand how these factors affected the FOM, each model was fit using a least squares regression of just the significant factors. For example, the model for the

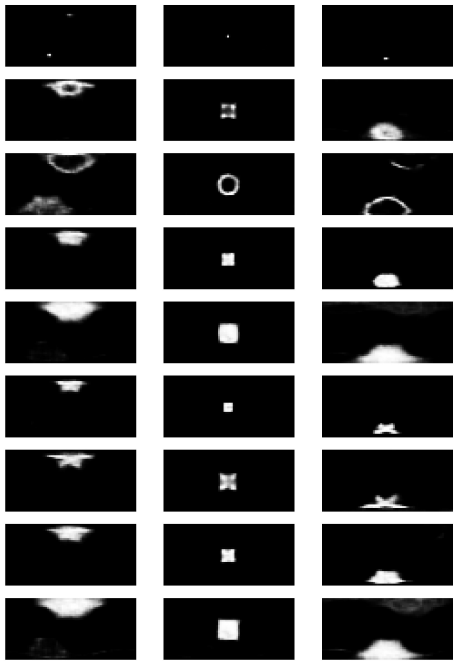


(a) True Source Images

(b) ReGeNN



(c) ML-EM with MRP



(d) ReGeNN with Noisy Training

Figure 5.8. Comparing the (a) true source distributions with (b) ReGeNN reconstructions shows significant improvement over (c) ML-EM with MRP, with ReGeNN recreating the source distribution at much higher detail over all source types. (d) Adding noise into the training data does not significantly impact the reconstructions, with the primary effect being blurring of sharp edges. In each sub-figure, the first column corresponds to $\phi = 30°$, the second column to $\phi = 90°$, and the third column to $\phi = 150°$source direction.

relative error would be given by

$$E = c_{\text{Shape}} + c_1 \left(\frac{\text{Size} - 40^\circ}{10^\circ} \right) \left(\frac{\text{Direction} - 90^\circ}{60^\circ} \right) + \text{BIAS}, \tag{5.7}$$

where c_{Shape} is the shape constant of proportionality for the actual source distribution and c_1 is the size and direction interaction constant. The terms in parentheses have been mapped to the range [-1,1] for the factor levels provided. In this way, all constants of proportionality are equally scaled and reflect the relative importance for each factor. Under this model, the bias roughly approximates the average expected FOM. These models are summarized in Table 5.2; for simplicity, only the significant factors are included with "–" indicating that the factor does not impact the FOM.

Table 5.2. Predictive Modeling using Significant Factors from ReGeNN ANOVA

Factor	Proportionality Constant		
	E	Precision	σ
Disc	–0.18	+0.09	–
Ring	+0.43	-0.23	+0.08
Junction	–	–	–
Square	-0.20	+0.12	–
Size	–	–	–
Direction	–	–	–
Shape*Size	–	–	–
Size*Direction	+0.10	–0.05	–
Direction^2	–	–	–
Bias	+0.29	+0.86	+0.20

When the categorical shapes were broken into their constituent components, the junction shape was determined to not significantly impact any FOM. The disc and square sources generally improved performance, lowering the relative error to as small as 9% and increasing the precision to as much as 98%. However, the ring sources were shown to have an inverse effect as reconstructions from those shapes had larger relative error, lower precision, and higher standard deviation (i.e. noise) within the ROI. It is also interesting to note that the ring reconstructions were the only ones that

affected the standard deviation. Finally, the size/direction interaction term indicates ReGeNN underperformed for sources with a larger size and near the southern pole of the RSM ($\phi = 150°$).

These results support the visual comparisons between Figures 5.8a and 5.8b. The disc, junction, and square source distributions are generally reconstructed in high detail and accuracy regardless of their size or direction, while the ring reconstructions are not as uniform as the original source shape. The right column reconstructions, corresponding to $\phi = 150°$, also contain more phantoms (false sources or background noise) than the other ϕ directions, particularly for the large disc and, to a lesser extent, the larger square source. This is quantified in Tables 5.1 and 5.2 through the size/direction interaction term. Through both quantitative and qualitative analysis, it is therefore possible to conclude that only the shape and combined size/direction of the source distribution significantly affects the reconstruction quality using ReGeNN.

5.5.3 Training for Noise

Following the success training ReGeNN under ideal conditions, the network was retrained using noisy DRCs to test its capabilities for more realistic applications. With this test, 10 "noisy" DRCs were generated for every possible source distribution by sampling each point along the DRC with a Gaussian distribution (the Gaussian was determined to be sufficient to approximate Poisson counting statistics given the number of expected counts in realistic scenarios). Each Gaussian was generated with a 2% relative standard deviation, using the ideal DRC as the mean value; the corresponding source image was kept to be the same monotonic activity from before. This generated a total of 514,080 DRCs to retrain ReGeNN, while another 51,408 noisy DRCs were used as the validation set. All hyper-parameters previously described were kept the same, except the number of epochs was lowered to 100 due to the larger training data

set. This corresponded to an equivalent number of updates to the network during training as compared to the 1000 epoch run using the smaller, ideal training dataset.

The training history using the noisy data set is plotted in Figure 5.9. Both the training and validation sets trended towards a minimum, indicating that the network was appropriately training towards a global solution. The validation loss curve is relatively smooth compared to the previous training using the ideal dataset (note that the scale in Figure 5.9 is 4x smaller than in Figure 5.5). This was in part due to using a larger dataset for both the training and validation data, as the larger training dataset helped the network to avoid local minima while the larger validation dataset averaged over more images to suppress the large loss values caused by the binary cross-entropy from just a few bad reconstructions. In fact, the first 30 epochs demonstrate the ideal training scenario, with little variation between the training and validation loss, indicating that the optimizer is training towards a global solution. After 30 epochs, however, the network has become fine-tuned and small perturbations caused by the stochastic nature of the optimizer cause it to momentarily reach a local minimum, although the perturbations are much smaller in magnitude than previously observed. Reconstructions with this noisy training and using the ideal DRCs as the input are presented in Figure 5.8d.

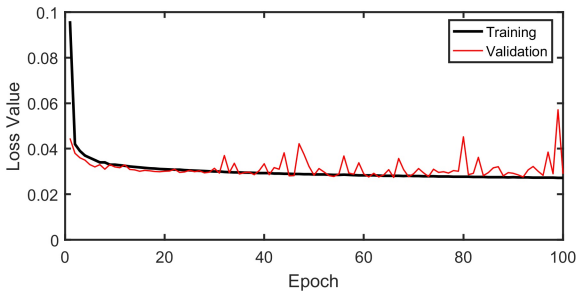


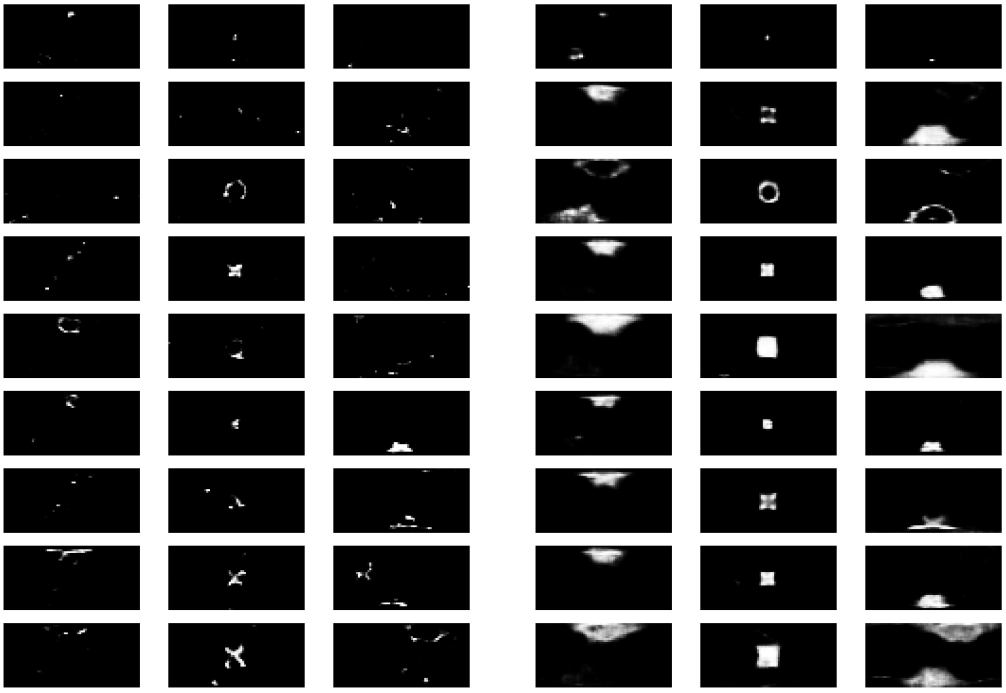
Figure 5.9. ReGeNN training history using 514,080 noisy DRCs as the training set. Due to the larger data set, the number of epochs was reduced to preserve the same number of updates as the ideal training. As before, the training and validation loss values converged towards a minimum, but the larger dataset helped reduce the magnitude of the spikes in the validation set.

The reconstructions again show excellent overall agreement, at a slight cost to blurring edges and loss of finer details. This loss is most notable for the junction and ring sources, although some features remain to distinguish them from the disc and square source distributions, which are harder to discern from each other. This is a relatively intuitive result as the difference between the square and disc distributions is only a few pixels, and even a minor amount of noise is enough to complicate the reconstruction process. Phantoms are also present, but not noticeably more than before under the ideal training (the only major phantom source appears in the first column, third row for one of the larger ring sources). It is interesting to note that training ReGeNN with noisy data helped clean the reconstructions for point sources, even if the reconstructions are not perfect "points". In general, these results show that ReGeNN is capable of handling noisy training data for practical applications with minimal degradation to the reconstruction.

However, obtaining an ideal DRC is not realistic, as noise and counting statistics dictate that there will be some fluctuations in the measured versus simulated DRCs. To test the practical applications using ReGeNN, the network was fed *noisy* DRCs using the same method to generate the noisy training data. The reconstructions are shown in Figure 5.10 for both networks trained on ideal and noisy DRCs, respectively.

As expected, the network performed poorly when trained under ideal conditions but given noisy inputs. When trained with noisy DRCs, the network was able to reconstruct the source distributions reasonably well even when the input data contained noise. The reconstructions in Figure 5.10b are not qualitatively different than those in Figure 5.8d, although quantitatively, the reconstruction accuracy is degraded slightly overall when given noisy input data.

Finally, Figure 5.11 plots the reconstructions when the input DRCs have a higher noise level (5%) than the training data (2%). There is a significant decrease in



(a) Ideal Training/Noisy Inputs (b) Noisy Training/Noisy Inputs

Figure 5.10. Comparing reconstructions using noisy input data to simulate typical experimental noise. (a) ReGeNN trained with ideal DRCs performed noticeably worse than (b) ReGeNN trained with noisy DRCs. The input data contained the same noise variance as the noisy training data.

performance, although the point source reconstructions remain relatively unaffected. While the general direction of the distributed sources is reconstructed properly, aside from a few outliers, the network failed to reconstruct the size and shape of most sources. In conjunction with the previous figures, these results suggest that while ReGeNN can be used for practical applications where measurements will contain some level of variance, careful consideration must be taken in designing the training dataset. The level at which the network trains will dictate the acceptable level of variance in the observed DRCs; future experimental work could further test this limitation. However, these results are nonetheless promising as they are substantially more robust than algorithms such as ML-EM, as will be discussed in the next section.

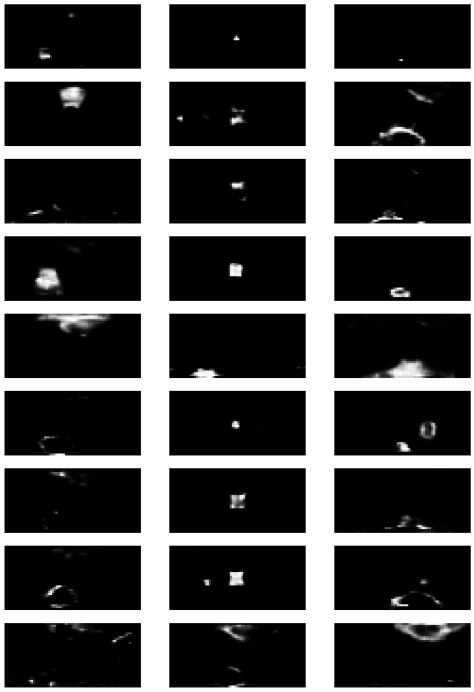
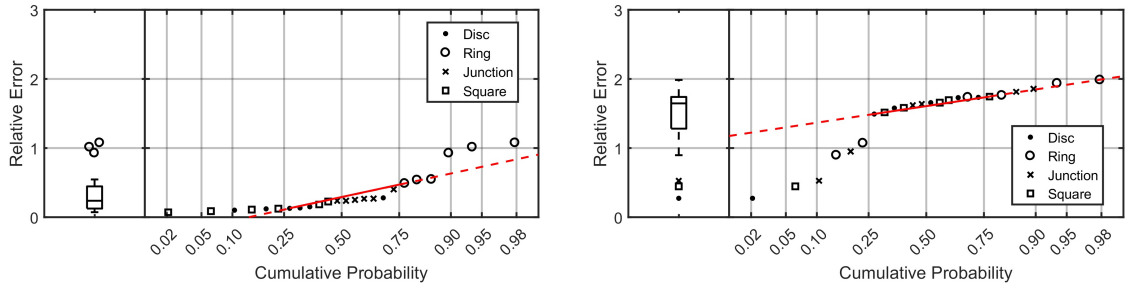


Figure 5.11. Reconstructions using ReGeNN when the input noise level (5%) is higher than the training data (2%). There is a noticeably decrease in accuracy, indicating that the measured data must have a variance equivalent to or smaller than the training data.

5.5.4 Comparison to ML-EM

The performance comparison between ReGeNN and ML-EM is readily summarized in Figures 5.12 and 5.13, which plots the relative error for the distributed sources as a box plot and a normal probability plot for the three reconstruction methods tested. It should be noted that the ML-EM study did not include Poisson noise and so is more directly comparable to ReGeNN trained with and using ideal DRCs as inputs. The probability axis is scaled such that any normal distribution would appear linear; superimposed on the plots are solid lines connecting the first and third quartiles, with dashed lines extrapolating it to the extremes. Under the null hypothesis, the FOM would be equal with some normally-distributed noise accounting for any variance so that most of the data would fall along the superimposed line. Thus, these plots are used for quick factor-screening (and model-assurance) by looking at the

shape of the data and seeing if any data points fall outside of the normal distribution.



(a) ReGeNN (b) ML-EM with MRP

Figure 5.12. Comparing the relative error for the distributed sources using ReGeNN and ML-EM with MRP. The plots show that ReGeNN has an overall higher accuracy and consistency than ML-EM, decreasing the average relative error from 145% **to only** 33%**.**

The ANOVA conducted already accounts for the outliers shown in the plots. What is most significant about these figures are the mean and spread of the data; ML-EM had an average relative error of 145% , while ReGeNN under ideal training had only 33%, lower than even the best reconstructions from ML-EM. The box plots also show that ML-EM had a much larger spread than the reconstructions using ReGeNN. This is a significant improvement across the full phase space tested that was not fully anticipated. Previous measurements indicated performance near the poles may be degraded, resulting in a more limited FOV than desired. However, ReGeNN had only a minor dependence on the source direction (as a secondary interaction effect with source size). Unlike ML-EM, which depended heavily on the direction-squared term, the biggest factor in ReGeNN was the source shape.

Using noisy training data did degrade performance, as shown in Figure 5.13, but primarily for the ring reconstructions; all other source distributions were reconstructed better than the majority of reconstructions from ML-EM, with the main source of error being a slight blurring of edges. Even for the ring sources, which had the largest relative errors, the ring source's general shape are still readily apparent in the reconstructions. This did result in a larger mean relative error, at 65%, although it

maintained reasonable visual quality and accurate source location and spatial extent information.

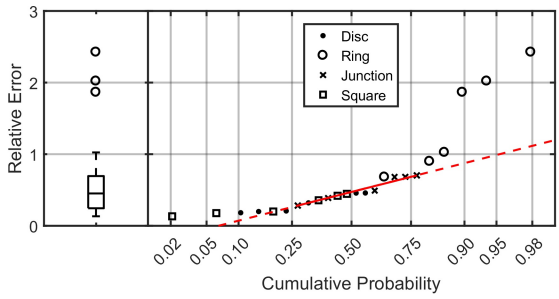


Figure 5.13. Normal probability plot for ReGeNN trained with noisy DRCs. While the relative error has slightly increased overall compared to ideal training datasets, it still shows a significant improvement compared to ML-EM.

Table 5.3 summarizes the various methods over the distributed sources, giving the average value for each FOM among the reconstructions. ReGeNN trained under ideal conditions performed the best in all categories except for σ, although the value for ML-EM is artificially low as most reconstructions failed in general, smearing the reconstruction and thus lowering $\hat{\lambda}_j \lambda_j$. The standard deviation within the ROI also decreased when ReGeNN was trained using noisy data, a result of the network blurring the images and removing some finer details. It should be noted that these results suggest that σ may not be a viable FOM on its own, as ML-EM appears to have a better FOM even though the ReGeNN reconstructions are more visually similar. However, it is included as the term is used in various other image quality measurements such as contrast-to-noise ratios, and provides a more complete quantitative representation for the reconstructions.

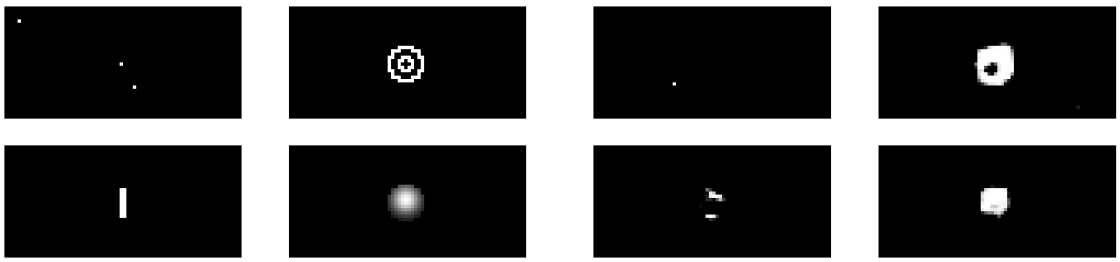
Both ReGeNN models outperformed ML-EM significantly, with the ideal scenario having better FOM and more consistency between the various source distributions. Visually comparing the point source reconstructions shows that ReGeNN trained with noisy DRCs performed the best at localizing the source for all three directions.

Table 5.3. Comparing Performance between Reconstruction Algorithms

Method	Average FOM		
	E	Precision	σ
Desired Value	Lo	Hi	Lo
ML-EM with MRP	145%	27%	12%
ReGeNN (Ideal Training)	33%	85%	24%
ReGeNN (Noisy Training)	65%	70%	14%

5.5.5 Network Robustness

While the results thus far demonstrate that ReGeNN is capable of accurately reproducing source images from the RSM DRCs, the network was trained using similar source shapes and sizes. To test the network robustness, four images were chosen that were similar to the original training source distributions but deviate significantly enough that they are clearly visually different. The distributions chosen for this set were multiple sources, concentric rings, a line source, and a localized Gaussian source. These images are plotted in Figure 5.14a.



(a) True Source Images (b) ReGeNN

Figure 5.14. (a) Images used to test network robustness and (b) their corresponding reconstructions using ReGeNN. The failure to accurately capture the original image details demonstrates the limitations of machine learning-based algorithms and the need to generate robust training data.

The corresponding reconstructions from ReGeNN, trained using the ideal DRCs of the original dataset, are shown in Figure 5.14b. ReGeNN generally underperformed for these four images, although the concentric rings and localized Gaussian maintained their approximate size, direction, and general shape. However, the multiple points

and line source failed completely to reconstruct an accurate image.

These results demonstrate the limitation of ReGeNN, and machine-learning based approaches in general: the network is only as robust as the training set it is given. It is not reasonable to expect that the algorithm should perform well for source distributions that it was not trained on. However, this limitation is not unique to ReGeNN; after all, the under-determined nature of the problem means that any algorithm (including ML-EM) must make an assumption about the expected source distribution. These images demonstrate the necessity to generate training data robust enough for the application space that it is expected to perform in.

5.6 Conclusion

This study presented an alternative, CNN-based reconstruction algorithm, known as ReGeNN, for RSM gamma imaging and other similar time-encoded detection systems. An RSM system using a Mace mask design was simulated in MCNP v6.1.4 to form a basis set of radiation source distributions and corresponding DRCs. The simulated results show that the trained ReGeNN could be used for RSM imaging, reducing the average relative error of the reconstructions to 33% and increasing the average precision to 85% when trained under ideal scenarios. For more realistic applications, the network was also shown to successfully reconstruct source distributions when trained with noisy data and given noisy inputs, with some minor degradation in the images. Both models outperformed ML-EM and suggests, for the first time, that imaging with the RSM system is not only plausible, but practical for this application.

An ANOVA study concluded that the shape and size/direction interaction terms were the biggest factors affecting the reconstruction relative error and precision. Ring sources generally had the worst (but still reasonable) FOM, while larger sources directed near the southern pole of the RSM also slightly degraded performance. This is

in direct contrast to the ML-EM ANOVA, which discovered a large dependence based on the source's direction and size, with the algorithm generally failing to reconstruct most source distributions. Another significant advantage was that the reconstruction was near instantaneous using ReGeNN, as nearly all the computation time was due to training the algorithm, a pre-processing step performed before taking data. This method could have great potential applications in nuclear security and safety where responsive feedback is necessary. However, the results also show the importance of developing robust, comprehensive datasets for those applications, as the network failed to accurately reconstruct DRCs that exceeded the trained noise level or distributions that were not included in the training set.

ReGeNN is not limited to just the RSM system, as the network could readily be adjusted for similar time-encoded systems with minimal changes to the network architecture. The techniques and methods may also be applied to other imaging applications, although careful consideration must be given when formatting the inputs and changing the network layers.

The results presented are only simulations, and a more rigorous experimental study is required to fully vet and demonstrate the RSM imaging capabilities. Future work can explore experimental validation and expand upon methods to improve the performance of ReGeNN in statistics-limited scenarios. Additionally, future work can also investigate a thorough mask design optimization for imaging applications, with ReGeNN integrated into the model as the reconstruction algorithm, to improve the overall coupled system performance.

Acknowledgments

This research was supported by the Defense Threat Reduction Agency under grant HDTRA-19-29109. The views expressed in this article are those of the authors and do

not necessarily reflect the official policy or position of the United States Air Force, the Department of Defense, the Department of Energy, or the United States Government.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administrations under contract DE-NA0003525.

VI. Experimental Imaging

Thus far, the experimental studies conducted with an RSM system were confined to determining a point source's direction [27, 30, 26], or finding multiple point sources with unique gamma energies [26]. Aside from the initial RSM development [25], which concluded that imaging could not be conducted at the time, no distributed source had been measured, much less reconstructed into a true radiation image. In this chapter, the ReGeNN reconstruction algorithm is tested with measured DRCs for two point sources and a mock ring source.

6.1 Experimental Setup

As part of the experimental validation, a new RSM system was designed and built to accommodate newer mask designs and improve the system's portability and modularity. The primary components, shown previously in Figure 3.1b, remain unchanged, with a high-density plastic mask affixed to a rotating sleeve connected to a motor. Inside the rotating aluminum sleeve is another stationary aluminum sleeve that houses a scintillation detector. The detector, with its integrated PMT, is powered by a standalone high-voltage power supply (HVPS), and its output is fed into a CAEN DT5790N digital acquisition (DAQ) system. Both the DAQ and the driver that controls the motor are connected to a laptop which acts as the primary user interface.

For this system, custom software was developed that integrated the DAQ signal generating the measured gamma spectrum with the motor's angular encoder. For each event, an angular tag associated with the mask rotation, a timing tag, and the short charge and long charge integration (for pulse shape discrimination when necessary) were recorded. This data could then be post-processed at any angular

or energy resolution defined by the user to generate the corresponding DRCs. The remainder of this section describes each component and the sources used in more detail.

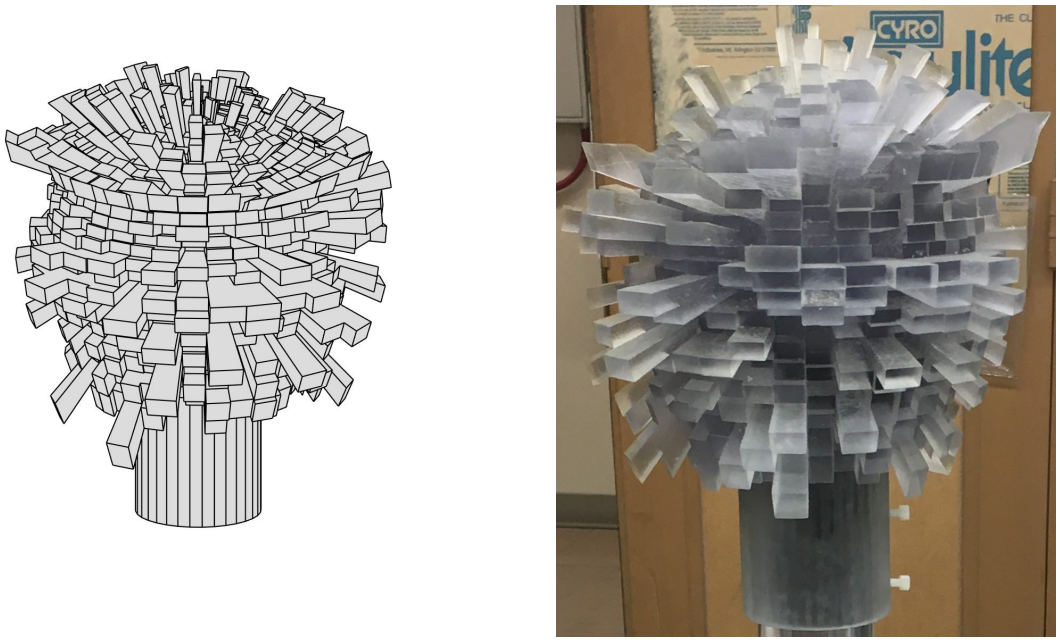
6.1.1 Mask Design

For the experimental measurements, a miniature version of the Mace was printed using the same PMMA-approximate material (VeroClear^{TM}) modeled previously. The maximum voxel length for this "MiniMace" was set to 7 cm to maximize attenuation while minimizing the torque on the system's motor. The inner diameter measured 5.65 cm to accommodate the detector, PMT, stationary aluminum detector sleeve, and rotating aluminum sleeve that couples the mask to the motor. A 0.5 cm thick cylindrical shell, extending along the bottom of the mask, was included to provide mechanical stability and structural support when mounted. The support shell also allowed for easy assembly and disassembly, as the mask was affixed through friction and two set screws rather than permanently affixing it to the rotating sleeve, improving the overall modular capability.

The mask design was constrained to span from 0° - 150° polar angles, rather than the previous 170° maximum, to fully accommodate the mechanical and electrical components of the actual system. The MiniMace contained 30 x 32 voxels spanning the polar and azimuthal angles, respectively, resulting in equivalent 5°- and 11.25°-angular extents. Figure 6.1 shows a CAD model highlighting these features, as well as the physical mask after printing.

6.1.2 $LaBr_3$ Detector

Due to detector availability, a $LaBr_3$(Ce) inorganic scintillator, which was small enough to fit in the stationary aluminum sleeve, was chosen to conduct the exper-



(a) CAD Model (b) Printed Mask

Figure 6.1. (a) CAD model for the MiniMace design and (b) the resulting 3D printed mask. Note the two set screws on the right of the image.

iments. Table 6.1 summarizes and compares the $LaBr_3$(Ce) crystal properties to NaI(Tl). While the mass attenuation coefficient for $LaBr_3$(Ce), shown in Figure 6.2, is not substantially higher than NaI(Tl), the increased density of the $LaBr_3$(Ce) crystal means a larger macroscopic cross section and thus higher intrinsic efficiencies. In general, $LaBr_3$(Ce) outperforms NaI(Tl), providing better energy resolution, faster timing, and higher efficiencies and count rates.

Table 6.1. $LaBr_3$(Ce) versus NaI(Tl) Properties

Properties	$LaBr_3$(Ce)	NaI(Tl)
Energy Resolution @ 662 keV	2.6%	∼6%
Photoelectron Yield [Relative to NaI(Tl)]	165%	100%
Primary Decay Time [ns]	16	250
Density [g/cc]	5.08	3.67

Originally, the $LaBr_3$(Ce) crystal was believed to have been a 10 mm × 10 mm right circular cylinder. Other than the small size reducing the overall efficiency, this would have been ideal as it provided a relatively symmetric shape and would minimize

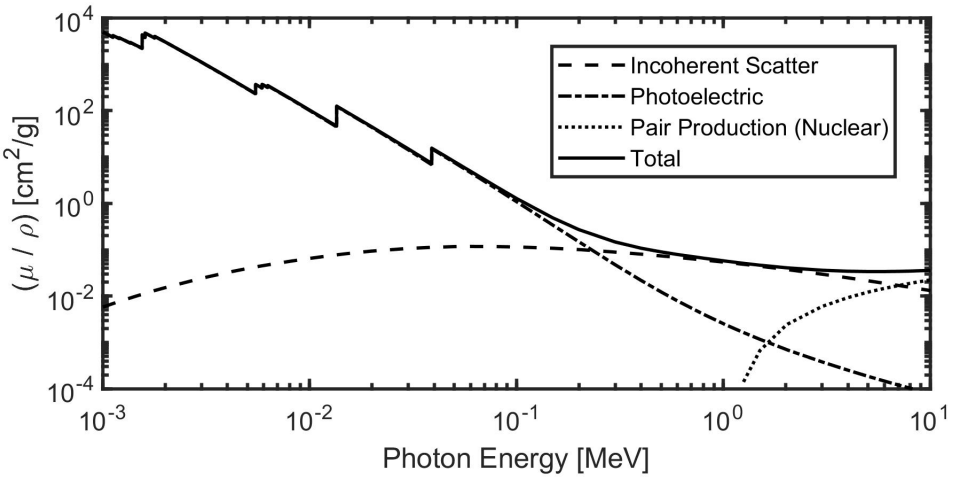


Figure 6.2. Mass attenuation coefficient versus photon energy for homogeneous $LaBr_3$. Combined with a larger crystal density, this detector has a larger intrinsic efficiency than standard NaI detectors.

any directional bias. The small size would also reduce the amount of interference from photons passing through and being detected through multiple voxels (in fact, if efficiency is of no concern, a point detector would be ideal), improving the SNR and uniqueness of the DRCs.

However, after consulting with the manufacturer, it was revealed that the detectors were a prototype design using a 10 mm $\times$ 10 mm $\times$ 50.8 mm rectangular prism crystal. This would potentially cause directional biasing as the effective pathlength would be dependent on the azimuthal angle of the source and the specific orientation of the crystal. An MCNP simulation calculating the count rate versus the source's azimuthal angle, plotted in Figure 6.3, shows that this biasing is minimal and within the expected noise level of the measured DRCs. The experimental results later demonstrate that this biasing is predominately negligible given the general uncertainty in the measured DRCs and limitations within ReGeNN.

The extended length further complicates the DRCs, as multiple voxels along the polar direction are now convolved into the final signal. This is somewhat alleviated as long as the detector length is properly accounted for in the model. It was believed

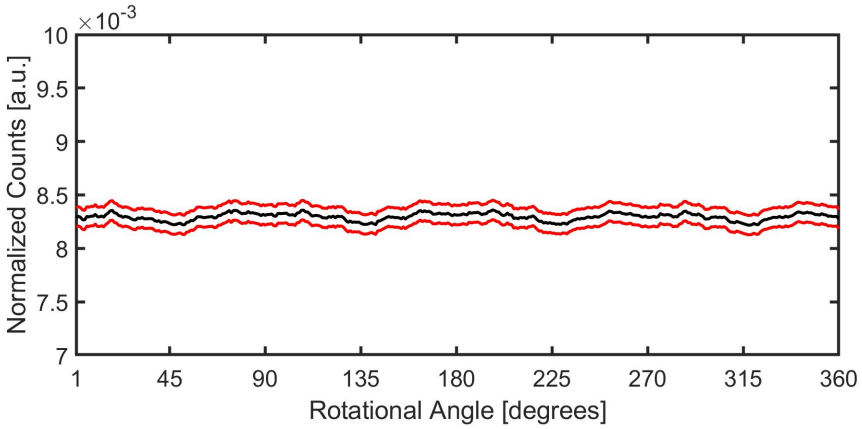


Figure 6.3. Simulated DRC (black) for the bare $LaBr_3$(Ce) rectangular detector and its corresponding one standard deviation uncertainty (red). While observable, the directional bias is well within the limits of expected counting statistics and is a relatively negligible effect.

that the overall effect of this non-ideal crystal shape would reduce the imaging efficacy by decreasing the uniqueness among DRCs over the FOV (the original design development assumed the majority of the detected particles would pass through a single voxel) but not remove it entirely. Future work is recommended to use different detector crystals better suited for this purpose, but is not explored in these experiments.

The $LaBr_3$(Ce) detector, with PMT included, was mounted to the stationary aluminum sleeve by attaching four foam blocks to the detector housing (Figure 6.4). These foam blocks provided sufficient friction to hold the detector in place while also producing a relatively uniform pressure to keep the detector centered within the sleeve. The low density and composition of the foam do not significantly attenuate the gammas and were excluded from the MCNP model.

6.1.3 Gamma Source

To maximize count rate and minimize collection time, a high-activity ^{22}Na source was used. ^{22}Na emits 511 keV gammas as a result of positron emission followed by

Figure 6.4. The $LaBr_3$(Ce) detector with integrated PMT and foam supports. The active part of the crystal is contained in the smaller aluminum housing in the lower right hand corner of the photo.

annihilation, along with 1274 keV gammas through another decay process. Three images were tested for this experiment: two point sources at varying azimuthal and polar angles and one ring source with a 20° angular extent centered at $\phi = 90°$. The first point source was chosen to be along $\phi = 90°$, $\theta = 147°$, while the second source was directed at $\phi = 75°$, $\theta = 327°$, ensuring that both azimuthal and polar angle dependencies were captured.

In order to simulate a distributed ring source, a ^{22}Na button source was placed on a 3-axis motor system. A computer program was generated that moved the button source in a circular pattern following the ring distribution (similar programs could be created to create the junction sources as well as rasters to mimic the filled-in disc and square sources). The RSM rotational period and the source's movement were de-synced to avoid correlation between the two movements and remove any unintentional directional biasing (i.e. avoiding the scenario where the source was always located at the same position along its arc for a given mask rotation angle).

For both the point and "distributed source", the ^{22}Na button was placed at least 2.22 m from the center of the detector. This distance ensured that the point sources could be approximated as a far-field planar source, simplifying the required MCNP

simulations and removing any distance dependencies.

6.2 ReGeNN Architecture and Training

A ReGeNN architecture similar to the one examined and tested in Chapter 5 was chosen, shown in Figure 6.5. The same 5° image resolution as all previous studies was chosen, resulting in 72 pixels along the azimuthal angle. However, to avoid arbitrary cropping within the network, the polar angle image size was reduced to 24 pixels, ranging from 30° - 145°. Removing polar angles below 30° is consistent with previous studies concluding a reduction in DRC quality at these angles [28, 29]. It should be noted that is possible to include these angles by changing the image resolution and changing the subsequent ReGeNN layers. However, as the goal of this research was to develop a functional, and not necessarily optimal, architecture, these modifications were deemed acceptable within the scope of the project.

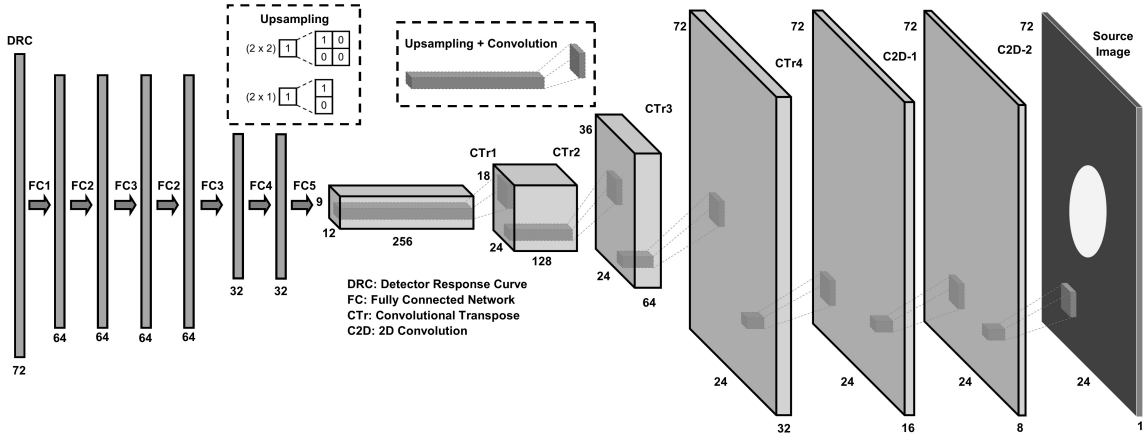


Figure 6.5. Architecture for the ReGeNN used to convert the DRC back into the source image for DRCs from the MiniMace design. Note that the previous design was modified to include additional layers to compensate for the expected added complexity of using a smaller mask and non-ideal crystal shape. The layer sizes were also changed to match the smaller FOV.

Due to the expected complications from scaling the mask down and the non-ideal crystal shape, the depth of the network was increased. In the encoder portion, two additional 64-node layers were added, theoretically improving the network's capability

to identify and distinguish key features within the DRC. On the decoder side, two additional convolutional layers (labelled C2D-1 and C2D-2) were included to similarly improve the image reconstruction quality during deconvolution. Additionally, the filter sizes in the decoder were modified as a variable size based on the upscaling of the current layer. That is, when the image resolution was upscaled, the filter size was equivalently upscaled so that the filter was consistently scanning along the same effective solid angle within the image throughout the network. Therefore, CTr1 used a 3x3 filter, CTr2 used a 6x6 filter, CTr3 used a 6x12 filter, and all remaining layers used a 6x24 filter. It was believed that this modification would help with the distributed source reconstructions by allowing the network to "see" more of the image in the final layers, where most of the image detail would be located, providing more information to each subsequent layer as a basis for image reconstruction.

As before, the training data was generated using MCNP v6.2 simulated DRCs (model shown in Figure 3.1a). A monodirectional beam of 511 keV gammas was modeled to simulate the far-field ^{22}Na sources, and the distributed sources were generated as a linear combination of point sources. However, to improve the simulated signal accuracy to the experimental measurements, the images and corresponding DRCs were generated at 1° resolution and later downsampled to 5°, better approximating the distributed sources. To compensate for signal noise within the measurements, either due to systematic uncertainties and/or counting statistics, the network was trained by adding an additional 2% relative uncertainty to the MCNP simulations for each of the DRC points. This addition was conducted using randomly sampled Gaussian noise with the MCNP results as the mean. As the MCNP DRCs already have an approximately 2% uncertainty, the final training data contained a simulated $\approx$3% "noise" level.

The training database consisted of the same points, rings, discs, and squares from

those used in Chapter 5. Due to mechanical constraints of what could be simulated in the lab, the junction sources were changed from an "X" to a "+". The source's angular extents ranged from 10° - 50° (in 10° increments), and the direction of the source's center was constrained such that the entirety of the source was captured in the image. In total, 27,648 unique source shapes and directions were generated.

Ten noisy DRC sets and their corresponding images were used as the training dataset, with another set of noisy DRCs generated for the validation set, for a total of 276,480 training sets and 27,648 validation sets. The image order was randomized to improve stability and prevent ReGeNN from biasing towards any specific distribution. The network was optimized using the binary cross-entropy loss function and the Adam algorithm, a more computationally-efficient version of stochastic gradient descent, with the learning rate set to 0.0001, $\beta_1 = 0.5$, and $\beta_2 = 0.999$ in accordance with Chapter 5. Training was conducted using mini-batches of 256 training sets, with batch normalization included after every layer (momentum set to 0.8) to stabilize the network during training and improve performance.

6.3 Results

6.3.1 ReGeNN Training History

Figure 6.6 plots the loss function for both the training data and validation data over 100 epochs (each epoch referring to one complete run through all the training data). Both the training and validation loss function decrease at the same rate initially, indicating a proper training of the network. After about 15 epochs, the training loss function continues to decrease, but the validation curve remains relatively constant and even begins to increase as training progressed.

This trend is the fundamental sign for over-training, whereby the network begins to simply "memorize" the training data, rather than learning the underlying connection

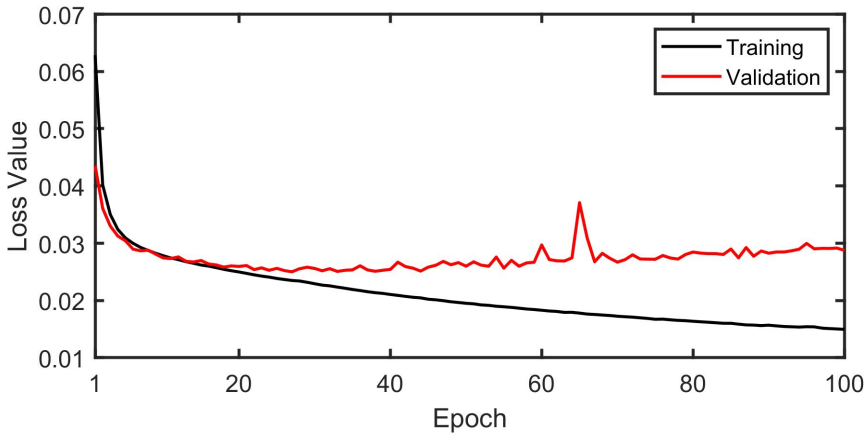
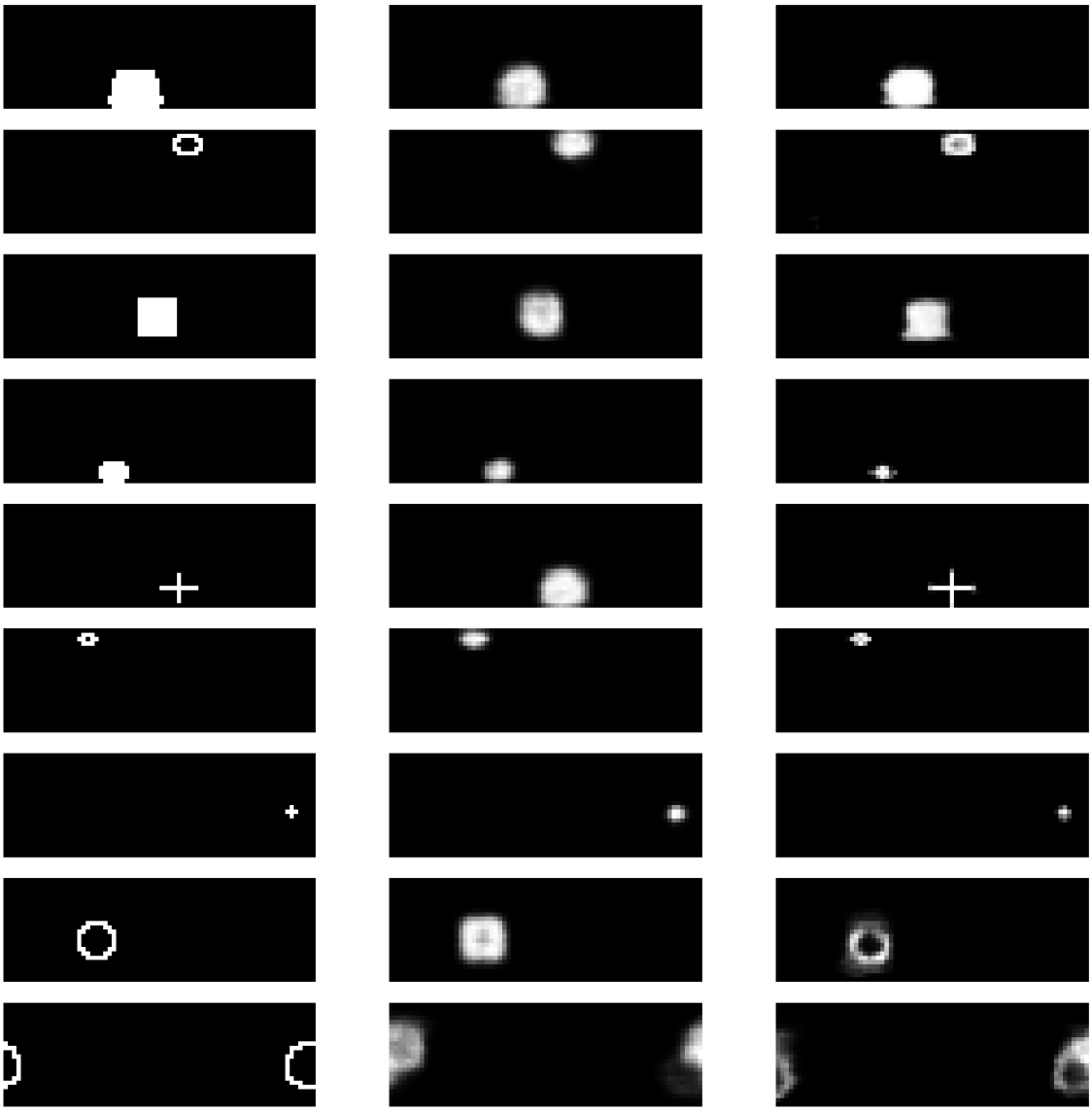


Figure 6.6. ReGeNN training history using 276,480 noisy DRCs as the training set. While the average traing loss value continually decreases, the validation started deviating significantly after 20 epochs, increasing as training progresses and indicating that the network has likely overfitted to the training data.

between input and output (i.e. local versus global solutions). However, plotting a sample of the validation reconstructed images after 100 epochs seems to show no such trend. These reconstructions, shown in Figure 6.7c, are remarkably accurate with a reasonable degree of precision and resolution (visually-speaking), and are actually *better* than the reconstructions after 10 epochs (Figure 6.7b) despite having a larger average loss function.

Investigating this discrepancy further reveals that the network may have overtraining, but not necessarily to the training dataset as a whole. Rather, this version of ReGeNN trained to prefer distributed sources over point sources. As shown in Figure 6.8, the reconstructions of the point sources in the validation dataset failed more often to reconstruct the point sources accurately compared to the distributed sources. This was likely caused by a majority of the training data containing distributed sources (especially when one recognizes that discs and squares are very similar in appearance at this resolution), with only 17,280 of the 276,480 DRCs (6.25%) corresponding to point sources. It is still interesting to note, however, that the majority of the reconstructions after 100 epochs still appear "better" than those at 10

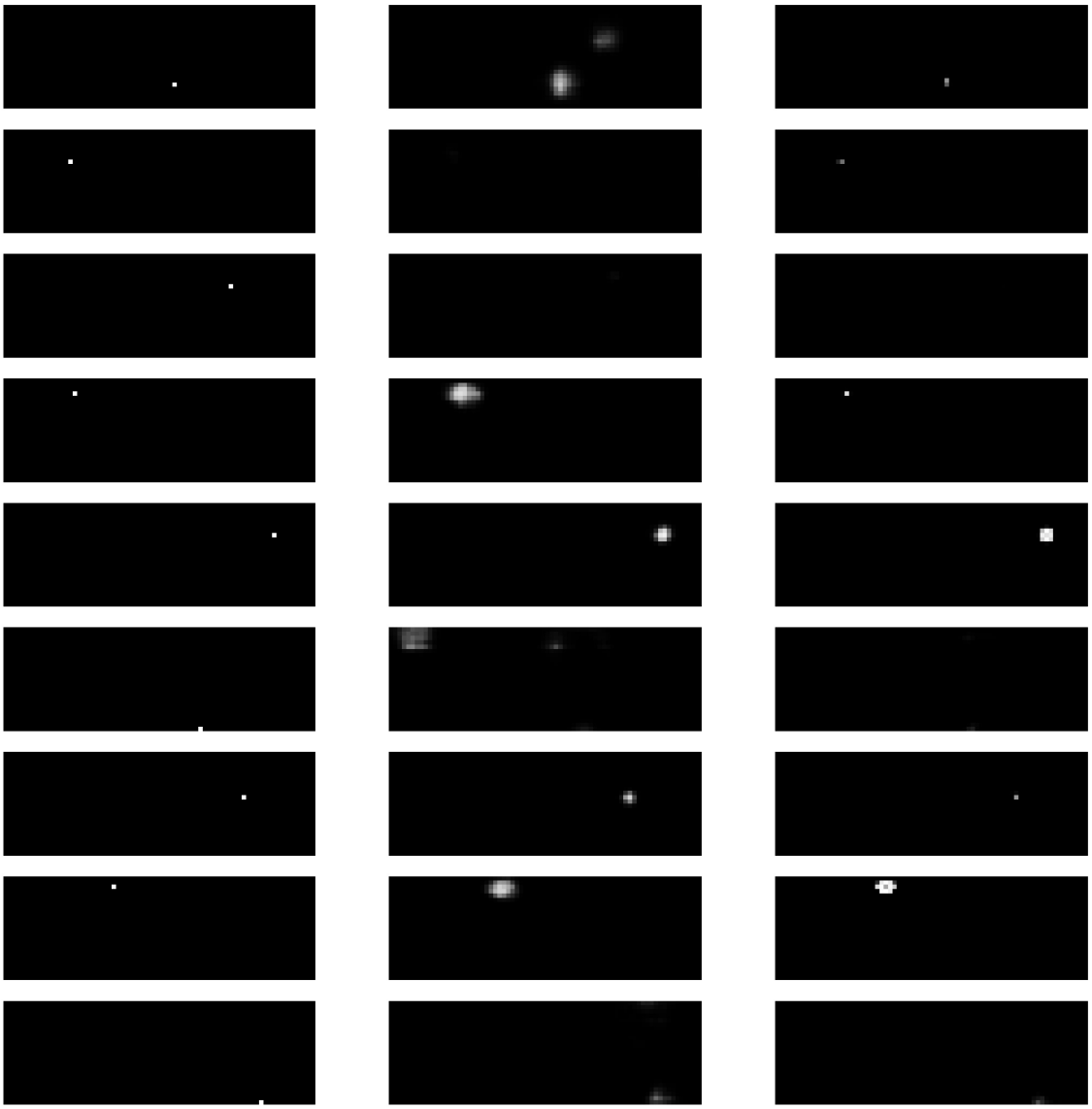


(a) True Source Images (b) 10 Epochs (c) 100 Epochs

Figure 6.7. (a) Sample source distributions and their validation reconstructions after (b) 10 and (c) 100 epochs. Counter to the training history, the later epochs seemed to have significantly improved reconstructions, suggesting no overfitting occurred.

epochs, and so there remains some uncertainty as to what is driving the increase in loss value within the validation set. While not explored further in this study, this suggests that ReGeNN could be further improved through better balancing of the training data, although the performance was already notably higher than expected given the non-ideal circumstances surrounding the printed mask and $LaBr_3$ detector

shape.



(a) True Source Images (b) 10 Epochs (c) 100 Epochs

Figure 6.8. (a) Sample point sources from the validation set and their reconstructions after (b) 10 and (c) 100 epochs. The network would often fail at reconstructing an accurate image for points throughout training, although it still appears that the later epochs generally had better performance, counter to what the training history would suggest.

6.3.2 Experimental Measurements and Images

To generate the measured FEP DRCs, background subtraction was required to remove the Compton background from the 511 and 1274 keV gammas from the ^{22}Na source as well as events from environmental background radiation sources. This was done by fitting the measured energy deposition spectrum to a linear combination of a Gaussian FEP and a smoothed step function (a step function convoluted with a Gaussian as defined by the FEP). As the mask rotated, an energy deposition spectrum was formed for each angular bin; once fitted, the resulting number of counts in the fitted Gaussian FEP was used to create the DRC.

Ideally, the Gaussian FEP would only change in magnitude between each angular energy deposition spectra, as the mean and variance are factors determined by the detector characteristics. The background, however, would change throughout rotation, as certain RSM orientations may have more Compton scatters into the detector over others due to the mask's asymmetric geometry. Therefore, the Gaussian mean and variance were estimated using the entire spectrum, over all rotational angles, in order to provide the most counts and best statistical certainty. These values were then set as the mean and variance in the individual angular spectra, with only the magnitude allowed to change. An example of a measured angular spectrum and the corresponding fits is shown in Figure 6.9, with a reduced chi-squared statistics of 2.26 between the fitted and measured spectrum. Similar agreement was seen throughout all experimental measurements with reduced chi-squared values ranging from 0.70 to 3.75.

The measured DRCs for the two point sources and 20° ring are plotted in Figure 6.10, along with their expected, simulated DRCs from MCNP. The plots show reasonably good agreement visually, indicating that the assumptions used in the MCNP models were adequate.

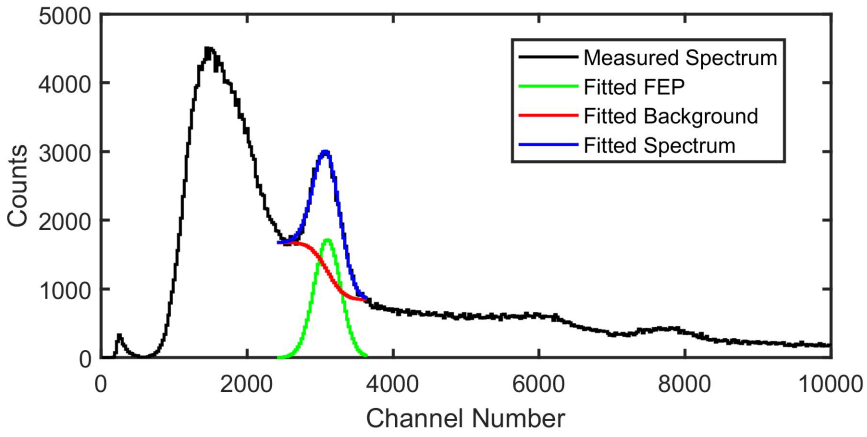
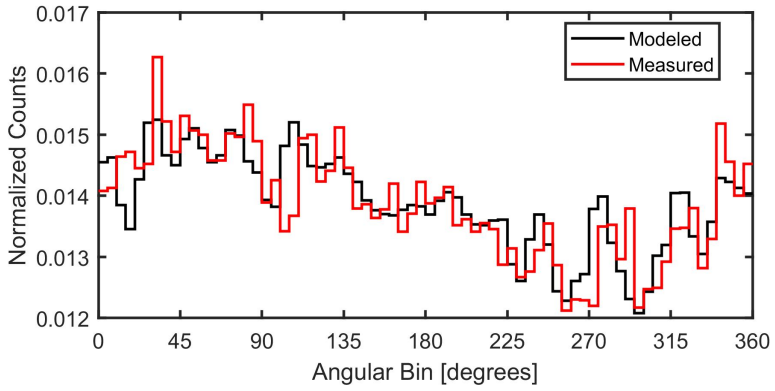


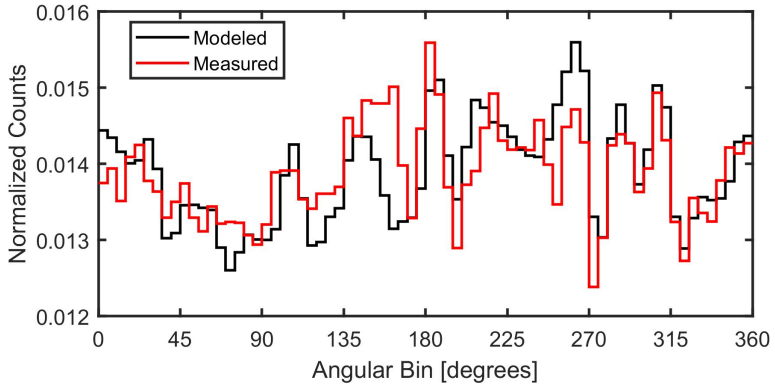
Figure 6.9. Example gamma energy deposition spectrum of a ^{22}Na point source at $\phi = 90°$, $\theta = 327°$ for counts collected from $\theta = 0°$-$5°$ and the corresponding FEP and background fits. The peak at around channel 3000 is the 511 keV peak, with the much smaller 1274 keV peak at around channel 8000. This specific example had a reduced chi-squared of 2.26, while the entire dataset ranged from 0.70 to 3.75.

Figure 6.11 shows the reconstructed images using ReGeNN after both 10 and 100 epochs. After 10 epochs, ReGeNN demonstrated remarkable performance in localizing the source within the correct region. While the point sources are not entirely localized, the reconstruction size is similar to previous results from Chapter 5, suggesting an overall resolution limit (although this perhaps could be improved using more balanced training data). However, the highest pixel activity in the point reconstructions did not correspond with the true point direction, with both reconstructions being 5° (one pixel) off. This resulted in rather large relative errors, 606% and 660% despite the reconstruction being visually close and only a single pixel away from the true direction.

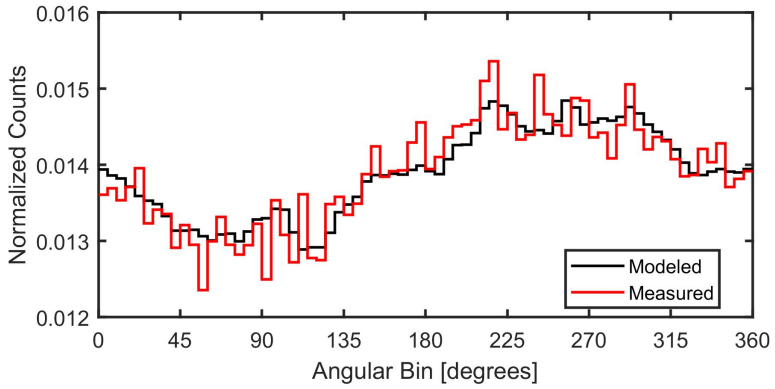
The ring reconstruction is also rather smoothed, and resembles a disc more so than a ring. However, the center corresponds precisely to the actual ring center, and the overall size is remarkably accurate (the bulk of the source is contained within the 20° angular extent) and is prominently more distributed than the point sources. The relative error for the ring reconstruction is much better, at just 109%. Even with just a cursory glance, it is clear that all three images represent three different sources,



(a) Point at $(90°, 327°)$



(b) Point at $(75°, 147°)$

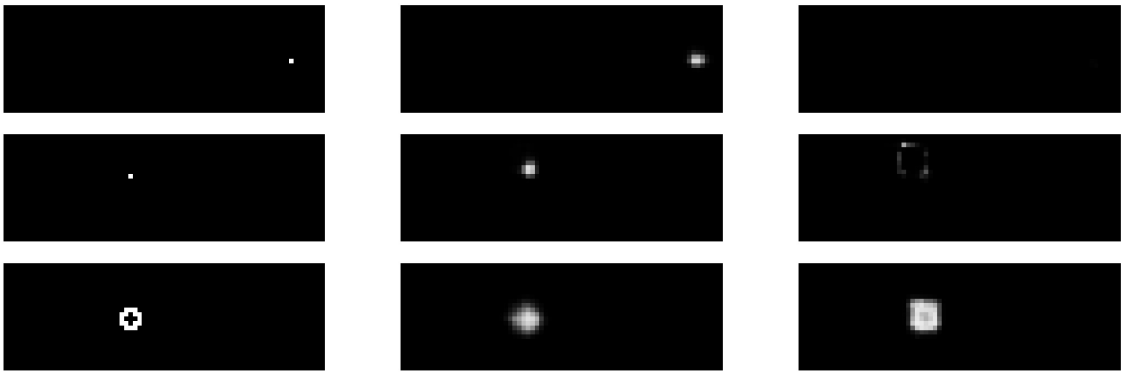


(c) Ring centered at $(90°, 147°)$

Figure 6.10. MCNP predicted DRCs versus measured DRCs for (a-b) the two point sources and (c) 20° ring. Coordinates for the source's center are given as (ϕ, θ). In all cases, the measured DRCs showed generally good agreement with the predicted MCNP models.

and the results are not all dissimilar from images of commercially-available imagers [74, 75].

The reconstructions after 100 epochs demonstrate the over-fitting issue previously



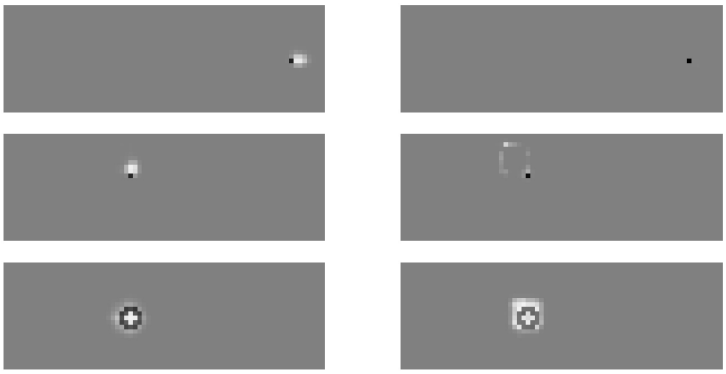
(a) True Source Images (b) 10 Epochs (c) 100 Epochs

Figure 6.11. (a) Experimental source distributions and their reconstructions after (b) 10 and (c) 100 epochs. At 10 epochs, the reconstructions are relatively accurate. The point source reconstructions were significantly worse after 100 epochs, but the ring reconstruction showed some improvement. In all cases, the reconstructions showed remarkable clarity, with practically zero noise in the image background.

discussed. Both point source reconstructions failed to even remotely reproduce the image. This highlights the largest risks regarding ReGeNN, and any neural network-based algorithm in general – when the network fails, it does not just fail to accurately reproduce the image, but it may fail to reproduce an image *entirely*. For example, the reconstruction of the first point source would seem to indicate the no source is present at all, as the image appears blank. This discrepancy is partially mitigated by the fact that having measured counts is already an indication that a source is present, so it is unlikely that this would cause a source to go undetected by most monitoring agencies. Rather, this showcases where such a design, or perhaps more specifically the ReGeNN algorithm, does not perform as expected. This would require the user would have to resort to more traditional methods to locate the source.

What is striking about these results is that the relative errors are actually *smaller* at 106% and 454% compared to the values after 10 epochs, despite the fact that Figure 6.11b agrees visually more than Figure 6.11c. This discrepancy seems to be an issue with quantitatively comparing images when the majority of an image is empty. Using a point source as an example, an empty reconstruction and a reconstruction

with a single pixel *anywhere* that is not the true direction have the same relative error (and the same loss value during training). However, a reconstruction that is just one pixel off (as in Figure 6.11b) is undeniably better than a reconstruction with a random pixel turned on, but this is not captured through the relative error or binary cross-entropy loss value during training. This suggests that more practical loss functions may need to be developed that are not as rigid as just the pure difference between images, and this may explain the upward trend in the validation loss as they have worse numerical reconstructions but significantly better visual quality.



(a) 10 Epochs (b) 100 Epochs

Figure 6.12. (a) Absolute difference between the true and reconstructed images after (a) 10 and (b) 100 epochs. From top to bottom, the total relative difference is 606%, 660%, **and** 109% **at 10 epochs and** 106%, 454%, **and** 146% **at 100 epochs.**

The ring reconstruction after 100 epochs does not appear to have this concern, as not only does the reconstruction similarly match the size and direction of the actual source as it did after 10 epochs, but it actually improved in reconstructing the specific distribution. Rather than resembling a normally distributed source (perhaps akin to what a spherical source may look like), the image is much more similar to that of a ring, although the relative error again *increased* from 109% to 146% compared to the reconstruction at 10 epochs. These results provide further evidence that the increase in validation loss during training was likely caused by a mixture of an imbalanced training dataset (distributed reconstruction performed better than points) and a non-

ideal loss function (higher loss values despite better visual quality).

Of note in all the reconstructions, regardless of the specific epoch, is the notable background suppression within the reconstructed images. Despite failing to reconstruct the image in some instances, the sources are entirely localized with no secondary phantoms and extremely high contrast-to-noise ratios. These images are not just a vast improvement over the simulated results from ML-EM [68], but are competitive with other imaging systems [51, 52]. These results mark a significant step in RSM development. This is the *first* instance where a distributed source has been measured and accurately reconstructed using the RSM system.

6.4 Conclusion

A new RSM system was designed and tested, using a miniature Mace design for the mask and $LaBr_3(Ce)$ as the detector to measure and image three different ^{22}Na source distributions. For this experiment, two point sources at varying polar and azimuthal angle were chosen as well as a 20° angular extent ring, centered on the RSM's equatorial plane. The ring distribution was mimicked experimentally by moving the source in a circular pattern and collecting data over a sufficient period to uniformly sample the ring.

With training data from MCNP simulations, a ReGeNN architecture was developed to reconstruct the measured DRCs. To account for the non-ideal detector shape and reduction in attenuation, and thus similarly signal strength, from the smaller mask, additional layers were included to expand the network from previous designs and theoretically improve performance. Ten DRC sets for images comprised of points, squares, rings, discs, and junctions, were generated for the training data, with 2% relative uncertainty added in addition to the 2% uncertainty in the MCNP model to account for statistical fluctuations in the experimental measurements. While the

training history seemed to indicate that the network began overfitting after about 15 epochs, analysis of the validation data revealed that the majority of the reconstructions were still improving with further training. It is believed that the increasing loss value in the validation data beyond 15 epochs is likely due to the network weights being biased to distributed sources over the more localized point sources. This may be due, in part, to an imbalanced training data set and warrants additional investigation.

After 10 epochs, ReGeNN was able to accurately reconstruct the direction for each of the three sources as well as reasonably estimate their relative size. These reconstructions showed remarkable clarity and high contrast-to-noise, with virtually zero noise in the image beyond a general blurring of the image to within 5° - 10°. After 100 epochs, the ring distribution reconstruction improved, with finer details unique to the ring (as compared to a disc or square of equal size) becoming visually discernible. However, the network failed to reconstruct images entirely for the two point sources, supporting the claim that the training data is likely imbalanced.

Regardless, these results show significant improvement from previous experiments and mark the first time the RSM system has been used to experimentally image a source successfully with such fidelity. Future work can explore fine-tuning the ReGeNN network to improve performance. The mask created for this experiment and the detector used in the system were chosen based on availability from current designs and equipment; as such, it is believed that the results could be further improved through a more robust system design optimization. Lastly, more distributions should be measured and tested to explore both algorithm and design limitations.

VII. Conclusions

7.1 Summary of Findings

This research demonstrates that gamma imaging is possible with the RSM system, providing a potential alternative to other gamma imagers that is portable and cost-effective. As part of the study, two imaging algorithms were investigated for comparison. The first, ML-EM, has strong foundations in various radiation imaging methods and reconstructs the image that is most likely to be detected given a measured response. A prior was applied to further improve performance by attempting to constrain the images to localized, monotonic distributions. While this method produced reasonably good images for a small selection of source distributions, primarily points and smaller disc sources, ML-EM ultimately failed to capture fine details and would generate noisy images for larger sources and sources near the edge of the RSM FOV. The algorithm was also limited by convergence rate, limiting the practicality for real-time analysis.

The second algorithm, termed ReGeNN, is a novel convolutional neural network-based method taking inspiration from recent successes in image processing using machine learning. This method proved far more successful, accurately reconstructing a majority of the source distributions simulated. Of the source distributions simulated, ReGeNN reduced the average relative error in the reconstructions from 145% in ML-EM to 33%, with relatively low background noise in the image. ReGeNN was also able to be trained with noisy datasets to account for variance within experimental measurements, with only minor degradation to the image quality.

To validate this method, a new RSM system was developed to measure a small set of ^{22}Na distributions. This system used a miniaturized Mace design developed in previous studies and the algorithm development, improving the overall portability.

During the experimental validation, a custom code was developed, integrating the various components and providing a more self-contained, modular detection system. Another ReGeNN architecture was developed and trained using MCNP-simulated responses (at 2% uncertainty) with an added 2% noise level added to account for experimental uncertainty. The image reconstructions from the experimental measurements showed that ReGeNN was able to capture the source's relative size and direction with reasonably high fidelity. The reconstructions featured excellent background suppression within the image, with high contrast-to-noise.

7.2 Benefits & Limitations

These findings mark the first successful application of the RSM as a gamma imager, with images comparable to already available devices at potentially a fraction of the cost. The system benefits from its rather simple design, with even the most technical aspect of the RSM, the single scintillating detector, being widely commercially available. Coupled with the additively manufactured mask, the RSM is relatively cost-efficient and can be manufactured on-demand to meet dynamic mission requirements. The new system is also much more portable than previous design iterations, although there is still extensive work to be done before a fieldable prototype can be deployed. The ReGeNN imaging algorithm allows for real-time analysis, with the majority of the computation time being consumed in the training prior to the measurement.

While ReGeNN was developed for the RSM, it is by no means applicable only to this research. The methodology can be adapted to other systems, particularly for similar time-encoded imagers, but it also has application to the general signal processing field. However, the experimental results demonstrates the pitfalls of neural network-based algorithms resulting from over-fitting to the training data. Careful

consideration must be taken to appropriately train the network for expected environments, while still being robust enough to handle a wide variety of datasets. Understanding the limitations of such networks, and thus RSM imaging capabilities, is critical to understanding the best scenarios and applications.

While previous research has concluded that the same RSM system could be used for directional dual-particle detection, this study limited itself to gamma events. There is nothing to suggest that the same techniques could not be applied for neutron imaging, although care must be taken given the more complex nature of neutron transport and detection. Similarly, this research only analyzed localized sources with discrete gamma energy emissions. The extent to which the system can handle radiation fields is unknown, but it is expected that the ML-EM approach may work better over ReGeNN due to the increased complexity and uncertainty associated with radiation fields.

7.3 Recommendations for Future Work

The sources measured in this study were by no means exhaustive, and future testing can explore more complex source types and distributions. Particularly, the attenuation through the mask and detection probability, and thus the resulting measured DRC, is known to be energy dependent, but no extensive research has been conducted on how best to account for such dependence in the fielded environment where the source is unknown. It may also be beneficial to extend the DRC measurements beyond the FEP regime, incorporating Compton events and significantly increasing the detection efficiency. In conjunction with the directional RSM progression, future work can also expand to neutron and dual-particle imaging.

A more rigorous mask design optimization, incorporating the ray-tracing radiation transport methodology developed from this work, is suggested to improve per-

formance. While the basis for the mask used in the experiment was the result of an optimization study, scaling down the mask ultimately altered the signal due to the non-linearity in attenuation and detection, and the previous optimization study was performed without consideration of the image reconstruction methodology. This optimization is also intricately tied to detector type and geometry and should be investigated further.

The conflicting evidence from the validation loss throughout training suggesting overfitting despite having generally improved validation reconstructions is still not well understood. In fact, the inherent flexible and evolving nature of neural networks, with many architectures and hyperparameters continuously being discovered, means that there is still substantial research needed with regards to network optimization. In addition to testing various architectures, future work can investigate various datasets and their impact on training performance.

Bibliography

1. IAEA, "IAEA incident trafficking database (ITDB): Incidents of nuclear and other radioactive material out of regulatory control 2020 fact sheet," https://www.iaea.org/resources/databases/itdb.

2. ——, "Mexico says stolen radioactive source found in field," https://www.iaea.org/newscenter/news/mexico-says-stolen-radioactive-source-found-field.

3. United Sates Department of Defense, "Nuclear posture review report," February 2018.

4. K. Vetter, R. Barnowksi, A. Haefner, T. H. Joshi, R. Pavlovsky, and B. J. Quiter, "Gamma-ray imaging for nuclear security and safety: Towards 3-d gamma-ray vision," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 878, pp. 159–168, 2018.

5. A. Rahmim and H. Zaidi, "Pet versus spect: strengths, limitations and challenges," *Nuclear Medicine Communications*, vol. 29, no. 3, pp. 193–207, 2008.

6. K. Vetter, L. Mihailescu, K. Nelson, J. Valentine, and D. Wright, "Gamma-ray imaging methods," Lawrence Livermore National Laboratory (LLNL), Livermore, CA, Tech. Rep., 2006.

7. M. J. Cieślak, K. A. Gamage, and R. Glover, "Coded-aperture imaging systems: Past, present and future development–a review," *Radiation Measurements*, vol. 92, pp. 59–71, 2016.

8. K. Ziock, J. Collins, L. Fabris, S. Gallagher, B. Horn, R. Lanza, and N. Madden, "Source-search sensitivity of a large-area, coded-aperture, gamma-ray imager," *IEEE Transactions on Nuclear Science*, vol. 53, no. 3, pp. 1614–1621, 2006.

9. A. L. Hutcheson, B. F. Phlips, E. A. Wulf, L. J. Mitchell, W. N. Johnson, and B. E. Leas, "Maritime detection of radiological/nuclear threats with hybrid imaging system," *2013 IEEE International Conference on Technologies for Homeland Security (HST)*, pp. 360–363, 2013.

10. D. Hellfeld, P. Barton, D. Gunter, L. Mihailescu, and K. Vetter, "A spherical active coded aperture for 4π gamma-ray imaging," *IEEE Transactions on Nuclear Science*, vol. 64, no. 11, pp. 2837–2842, 2017.

11. K. Ziock and M. Blackston, "Real time depth-of-interaction correction for coded-aperture, gamma-ray images," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 916, pp. 56–65, 2019.

12. S. Sun, Y. Liu, and X. Ouyang, "Near-field high-resolution coded aperture gamma-ray imaging with separable masks," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 951, 2020.

13. Y. Kim, J. H. Kim, H. S. Lee, H. R. Lee, J. H. Park, J. H. Park, H. Seo, C. Lee, S. H. Park, and C. H. Kim, "Development of compton imaging system for nuclear material monitoring at pyroprocessing test-bed facility," *Journal of Nuclear Science and Technology*, vol. 53, no. 12, pp. 2040–2048, 2016.

14. Y. Kim, J. H. Kim, J. Lee, and C. H. Kim, "Large-area compton camera for high-speed and 3-d imaging," *IEEE Transactions on Nuclear Science*, vol. 65, no. 11, pp. 2817–2822, 2018.

15. C. G. Wahl, W. R. Kaye, W. Wang, F. Zhang, J. M. Jaworski, A. King, Y. A. Boucher, and Z. He, "The polaris-h imaging spectrometer," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 784, pp. 377–381, 2015.

16. J. E. Goldsmith, M. D. Gerling, and J. S. Brennan, "A compact neutron scatter camera for field deployment," *Review of Scientific Instruments*, vol. 87, no. 8, 2016.

17. K. Weinfurther, J. Mattingly, E. Brubaker, and J. Steele, "Model-based design evaluation of a compact, high-efficiency neutron scatter camera," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 883, pp. 115–135, 2018.

18. J. Brennan, E. Brubaker, M. Gerling, P. Marleau, M. Monterial, A. Nowack, P. Schuster, B. Sturm, and M. Sweany, "Source detection at 100 meter stand-off with a time-encoded imaging system," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 877, pp. 375–383, 2018.

19. N. P. Shah, J. VanderZanden, and D. K. Wehe, "Design and construction of a 1-d, cylindrical, dual-particle, time-encoded imaging system," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 954, 2020.

20. X. Liang, X. Pang, D. Cao, D. Li, Z. Zhang, S. Liu, T. Hu, Y. Zhang, X. Wang, F. Meng *et al.*, "Self-supporting design of a time-encoded aperture, gamma-neutron imaging system," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 951, 2020.

21. D. Wang, I. N. Ruskov, H. Hu, Y. N. Kopatch, D. N. Grozdanov, N. A. Fedorov, and F. A. Aliyev, "Gamma-ray imaging with a time-modulated random coded aperture," *Review of Scientific Instruments*, vol. 90, no. 1, 2019.

22. V. Schonfelder, R. Diehl, G. Lichti, H. Steinle, B. Swanenburg, A. Deerenberg, H. Aarts, J. Lockwood, W. Webber, J. Macri *et al.*, "The imaging compton telescope comptel on the gamma ray observatory," *IEEE Transactions on Nuclear Science*, vol. 31, no. 1, pp. 766–770, 1984.

23. D. Smith, G. Hurford, and S. Boggs, "Rotating modulation collimator imagers," *New Astronomy Reviews*, vol. 48, no. 1, pp. 209–213, 2004.

24. B. Kowash, D. Wehe, and J. Fessler, "A rotating modulation imager for locating mid-range point sources," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 602, no. 2, pp. 477–483, 2009.

25. J. G. FitzGerald, "A rotating scatter mask for inexpensive gamma-ray imaging in orphan source search: Simulation results," *IEEE Transactions on Nuclear Science*, vol. 62, no. 1, pp. 340–348, 2015.

26. Z. T. Condon, "Multisource direction identification using a rotating scatter mask," Master's thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH, 2018.

27. J. V. Logan, D. E. Holland, L. W. Burggraf, J. A. Clinton, and B. E. O'Day III, "Monte carlo and experimental analysis of a novel directional rotating scatter mask gamma detection system," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 947, 2019.

28. D. E. Holland, J. E. Bevins, L. W. Burggraf, and B. E. O'Day, "Rotating scatter mask optimization for gamma source direction identification," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 901, pp. 104–111, 2018.

29. R. J. Olesen, B. E. O'Day, D. E. Holland, L. W. Burggraf, and J. E. Bevins, "Characterization of novel rotating scatter mask designs for gamma direction identification," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 954, 2020.

30. B. V. Egner, "Development of a mixed-radiation directional rotating scatter mask detection system," Master's thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH, 2019.

31. M. J. Berger, J. H. Hubbell, S. M. Seltzer, J. Chang, J. S. Coursey, R. Sukumar, D. S. Zucker, and K. Olsen, "XCOM: Photon Cross Section Database (version 1.5)," National Institute of Standards and Technology, Gaithersburg, MD., 2010. [Online]. Available: https://www.nist.gov/pml/xcom-photon-cross-sections-database

32. E. E. Fenimore and T. M. Cannon, "Coded aperture imaging with uniformly redundant arrays," *Applied Optics*, vol. 17, no. 3, pp. 337–347, 1978.

33. R. Dicke, "Scatter-hole cameras for x-rays and gamma rays," *The Astrophysical Journal*, vol. 153, pp. L101–L106, 1968.

34. K. P. Ziock, M. Burks, W. Craig, L. Fabris, E. Hull, and N. Madden, "Real-time generation of images with pixel-by-pixel spectra for a coded aperture imager with high spectral resolution," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 505, no. 1-2, pp. 420–424, 2003.

35. M. Gmar, M. Agelou, F. Carrel, and V. Schoepff, "Gampix: A new generation of gamma camera," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 652, no. 1, pp. 638–640, 2011.

36. S. Sun, Z. Zhang, L. Shuai, D. Li, Y. Wang, Y. Liu, X. Huang, H. Tang, T. Li, P. Chai *et al.*, "Development of a panorama coded-aperture gamma camera for radiation detection," *Radiation Measurements*, vol. 77, pp. 34–40, 2015.

37. T. Lee, S.-W. Kwak, and W. Lee, "Investigation of nuclear material using a compact modified uniformly redundant array gamma camera," *Nuclear Engineering and Technology*, vol. 50, no. 6, pp. 923–928, 2018.

38. M. D. Gerling, J. E. Goldsmith, and J. S. Brennan, "Miner-a mobile imager of neutrons for emergency responders," *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1–4, 2014.

39. G. J. Hurford, E. J. Schmahl, R. A. Schwartz, A. J. Conway, M. J. Aschwanden, A. Csillaghy, B. R. Dennis, C. Johns-Krull, S. Krucker, R. P. Lin *et al.*, "The rhessi imaging concept," *The Reuven Ramaty High-Energy Solar Spectroscopic Imager (RHESSI)*, pp. 61–86, 2003.

40. G. Hurford, R. Schwartz, S. Krucker, R. Lin, D. Smith, and N. Vilmer, "First gamma-ray images of a solar flare," *The Astrophysical Journal Letters*, vol. 595, no. 2, p. L77, 2003.

41. B. Kowash and D. Wehe, "A unified near-and far-field imaging model for rotating modulation collimators," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 637, no. 1, pp. 178–184, 2011.

42. J. Brennan, E. Brubaker, M. Gerling, P. Marleau, K. McMillan, A. Nowack, N. Renard-Le Galloudec, and M. Sweany, "Demonstration of two-dimensional time-encoded imaging of fast neutrons," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 802, pp. 76–81, 2015.

43. 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. 1–22, 1977.

44. L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 113–122, 1982.

45. S. Alenius and U. Ruotsalainen, "Bayesian image reconstruction for emission tomography based on median root prior," *European Journal of Nuclear Medicine*, vol. 24, no. 3, pp. 258–265, 1997.

46. S. Alenius, U. Ruotsalainen, and J. Astola, "Using local median as the location of the prior distribution in iterative emission tomography image reconstruction," *IEEE Transactions on Nuclear Science*, vol. 45, no. 6, pp. 3097–3104, 1998.

47. S. Alenius and U. Ruotsalainen, "Generalization of median root prior reconstruction," *IEEE Transactions on Medical Imaging*, vol. 21, no. 11, pp. 1413–1420, 2002.

48. M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural networks*, vol. 6, no. 6, pp. 861–867, 1993.

49. T. Kin, Y. Sanzen, M. Kamida, K. Aoki, N. Araki, and Y. Watanabe, "Artificial neural network for unfolding accelerator-based neutron spectrum by means of multiple-foil activation method," *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1–2, 2017.

50. E. K. Elmaghraby, M. Tohamy, and M. Comsan, "Determination of isotopes activity ratio using gamma ray spectroscopy based on neural network model," *Applied Radiation and Isotopes*, vol. 148, pp. 19–26, 2019.

51. R. Zhang, P. Gong, X. Tang, P. Wang, C. Zhou, X. Zhu, L. Gao, D. Liang, and Z. Wang, "Reconstruction method for gamma-ray coded-aperture imaging based on convolutional neural network," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 934, pp. 41–51, 2019.

52. R. Zhang, X. Tang, P. Gong, P. Wang, C. Zhou, X. Zhu, D. Liang, and Z. Wang, "Low-noise reconstruction method for coded-aperture gamma camera based on multi-layer perceptron (in-press)," *Nuclear Engineering and Technology*, 2020.

53. A. Dosovitskiy, J. T. Springenberg, M. Tatarchenko, and T. Brox, "Learning to generate chairs, tables and cars with convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 4, pp. 692–705, 2016.

54. K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

55. C. J. Werner *et al.*, *MCNP6.2 Release Notes*, Los Alamos National Laboratory, report LA-UR-18-20808, 2018.

56. J. T. Goorley *et al.*, *Initial MCNP6 Release Overview - MCNP6 version 1.0*, Los Alamos National Laboratory, report LA-UR-13-22934, 2013.

57. M. Laraia, *Nuclear Decommissioning: Planning, Execution and Decommissioning.* Woodhead Publishing, 2012.

58. R. S. Woolf, B. F. Phlips, A. L. Hutcheson, L. J. Mitchell, and E. A. Wulf, "An active interrogation detection system (actinides) based on a dual fast neutron/gamma-ray coded aperture imager," *2012 IEEE Conference on Technologies for Homeland Security (HST)*, pp. 30–35, 2012.

59. P. Hausladen, J. Newby, F. Liang, and M. Blackston, "The deployable fast-neutron coded-aperture imager: Demonstration of locating one or more sources in three dimensions," 2013. [Online]. Available: http://info.ornl.gov/sites/publications/files/Pub46191.pdf

60. K. Vetter, L. Mihailescu, K. Nelson, J. Valentine, and D. Wright, "Gamma-ray imaging methods," Lawrence Livermore National Lab (LLNL), Livermore, CA (United States), Tech. Rep., 2006.

61. S. Peterson, D. Robertson, and J. Polf, "Optimizing a three-stage compton camera for measuring prompt gamma rays emitted during proton radiotherapy," *Physics in Medicine & Biology*, vol. 55, no. 22, 2010.

62. M. McCleskey, W. Kaye, D. S. Mackin, S. Beddar, Z. He, and J. Polf, "Evaluation of a multistage cdznte compton camera for prompt γ imaging for proton therapy," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 785, pp. 163–169, 2015.

63. T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

64. J. Allison, K. Amako, J. Apostolakis, P. Arce, M. Asai, T. Aso, E. Bagli, A. Bagulya, S. Banerjee, G. Barrand *et al.*, "Recent developments in geant4," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 835, pp. 186–225, 2016.

65. P. Martin, N. Tomkinson, and T. Scott, "The future of nuclear security: Commitments and actions–power generation and stewardship in the 21st century," *Energy Policy*, vol. 110, pp. 325–330, 2017.

66. I. Jovanovic and A. S. Erickson, *Active Interrogation in Nuclear Security: Science, Technology and Systems*. Springer, 2018.

67. Z. Mu and Y.-H. Liu, "Aperture collimation correction and maximum-likelihood image reconstruction for near-field coded aperture imaging of single photon emission computerized tomography," *IEEE Transactions on Medical Imaging*, vol. 25, no. 6, pp. 701–711, 2006.

68. R. J. Olesen, D. E. Holland, E. M. Brubaker, and J. E. Bevins, "Maximum likelihood reconstructions for rotating scatter mask imaging," *Radiation Measurements*, vol. 137, 2020.

69. S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?" *Advances in Neural Information Processing Systems*, pp. 2483–2493, 2018.

70. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

71. A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

72. F. Chollet *et al.*, "Keras," https://keras.io, 2015.

73. R. J. Olesen *et al.*, "Regenn," Mar 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3725845

74. C. G. Wahl, W. Kaye, W. Wang, F. Zhang, J. Jaworski, Y. A. Boucher, A. King, and Z. He, "Polaris-h measurements and performance," *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1–4, 2014.

75. O. Ivanov, V. Potapov, A. Safronov, and A. Stepanov, "The improving of the angular resolution of gamma-ray images of portable compton camera with special iterational method," *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1–5, 2019.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
01-10-2020	PhD Dissertation	Mar 2018 – Oct 2020

4. TITLE AND SUBTITLE

LOW-INFORMATION RADIATION IMAGING USING ROTATING SCATTER MASK SYSTEMS AND NEURAL NETWORK ALGORITHMS

5a. CONTRACT NUMBER

5b. GRANT NUMBER

HDTRA-19-29109

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Olesen, Robert J, Capt

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENP-DS-20-S-028

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

ATTN: LTC Nickolas Duncan (nickolas.duncan@westpoint.edu)
Defense Threat Reduction Agency
8725 John J. Kingman Rd. Stop 6201
Fort Belvoir, VA 22060-6201

10. SPONSOR/MONITOR'S ACRONYM(S)

DTRA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

This material is declared work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

Developing fast, portable, and accurate radiation imagers remains an objective for many nuclear safety and security applications. While recent studies have demonstrated the directional capabilities of the single-detector rotating scatter mask (RSM) system for discrete, dual-particle environments, there has been little progress towards adapting it as a true imaging device. In this study, two algorithms were developed and tested using an RSM mask design previously optimized for directional detection and simulated ^{137}Cs signals from a variety of source distributions. The first, maximum-likelihood expectation-maximization (ML-EM), was shown to generate noisy images, with relatively low accuracy (145% average relative error) and signal-to-noise ratio (0.27) for most source distributions simulated. The second, a novel regenerative neural network (ReGeNN), performed exceptionally well, with significantly higher accuracy (33% average relative error) over all source types compared to ML-EM and drastically improved signal-to-noise ratio (0.85) in the reconstructed images. This method was experimentally validated using an additively-manufactured mask. Measuring two point and one ring ^{22}Na source distributions, a modified ReGeNN was able to successfully train on simulated noisy signals and accurately predict the relative size and direction of the three sources. Training ReGeNN further revealed potential errors caused from overfitting, suggesting future improvement in ReGeNN architecture and training base is needed to obtain accurate activity profiles. To support future design optimizations, a ray tracing algorithm was also developed as an alternative to more rigorous Monte Carlo RSM simulations. This ray tracing code was shown to significantly improve computational efficiency, at a slight cost to the simulated signal accuracy for more complex designs.

15. SUBJECT TERMS

Radiation Imaging, Nuclear Security, Rotating Scatter Mask, Single Detector, Time Encoding, Gamma Detection, Neural Networks, Ray Tracing, Maximum Likelihood

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Maj James E. Bevins, AFIT/ENP
U	U	U	U	152	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636, x4767; james.bevins@afit.edu