2011

Characterization and prototyping of the rotating modulator hard x-ray/gamma-ray telescope

Brent Budden
Louisiana State University and Agricultural and Mechanical College, bbudden@phys.lsu.edu

CHARACTERIZATION AND PROTOTYPING OF THE ROTATING MODULATOR HARD X-RAY/GAMMA-RAY TELESCOPE

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in
The Department of Physics & Astronomy

by
Brent Budden
B.A., University of Pennsylvania, 2005
May 2011

To Kristin, Mom, and Dad.

Acknowledgements

First and foremost, I would like to thank my wife, Kristin Budden, for her tremendous support and enthusiasm for my accomplishements in this research project. The sacrifices she has made in order for me to continue and finish this work are beyond anything I could have expected.

I would like to thank my parents, John and Joelle Budden, whose pride in me has been a source of motivation in my studies since childhood. Thanks also to my brother, Mark Budden, who provided both encouragement and example in academic endeavors.

I would like to thank my fellow graduate students (of which there are too many to name) for joining me during countless hours of homework sessions and studying. And thank you to the faculty in the physics department at LSU for being both knowledgeable and engaging.

I owe a special debt of gratitude to my labmates. In particular, I thank Gary Case for his mentorship. His experience and selfless devotion in guiding the students in our group has contributed in countless ways to my own knowledge and research. I thank also Joachim Isbert, Jarrod Marsh, James Rodi, Yuan Zhang, Winston Capps, and the rest of the space science group. While we often worked on separate projects, I will miss the useful discussions on common challenges.

This work would not have been possible without the financial support of several organizations and people. Partial funding was made possible by the United States Department of Energy NNSA Cooperative Agreement DE-FC52-04-NA25683. I want to also thank NASA, the Louisiana Board of Regents, and the Louisiana Space Consortium – especially Greg

Guzik and John Wefel – for financial support under agreement NASA/LEQSF(2005-2010)-LaSpace, agreement NNX07AT62A, and grant NNG05GH22H. This support over the past five years enabled me to focus my energy on research instead of funding. I also wish to extend thanks to the Coates Foundation at LSU for additional research support.

I thank the staff of the electronics shop in the physics department – Brad Ellison, Randy Gould, Marcus Nauman, and Doug Smith – for frequent help and contributions to this work and for friendly matches on the tennis court. I also thank the staff of the machine shop; George Gascon, Brandon Amos, Vincent Vaugn, and Donnie Olano were knowledgeable and skillful in the design and building phases of this project.

I would like to thank the faculty and students of the medical physics program at LSU for allowing me to join them in classes and seminars, which helped in my research of imaging modalities and reconstruction. In particular, I thank Ken Hogstrom for his helpful advice in assembling a curriculum and Polad Shikhaliev and Kip Matthews for their always-open doors and useful discussions.

I thank Mark Devlin and his group at the University of Pennsylvania. It was as an undergraduate that I worked in his laboratory and realized my love for experimental physics. The experience I gained under his advisement led me to join the space science group here at LSU.

I would like to thank Gay Sutton, Michael Stewart, and Doug Granger. I appreciate their help as coworkers, but I am especially thankful for their friendship. I've enjoyed our time on the third floor of Nicholson and even more so our frequent lunches together.

Finally, I extend thanks to my Ph.D. committee for providing useful comments and suggestions. In particular, I thank my advisor, Mike Cherry, for providing me with the opportunity in his laboratory, steering my research focus, and challenging me to develop new techniques for x-ray/gamma-ray imaging.

Table of Contents

List of Tables

List of Figures

Abstract

A hard x-ray/gamma-ray telescope with high sensitivity and wide field of view would be capable of performing an all-sky census of black holes over a wide range of obscuration and accretion rates. As an example, NASA's Black Hole Finder Probe mission was designed to provide a 5σ flux sensitivity in a 1-year observation of $\sim$ 0.02 mCrab in the 10 - 150 keV energy range and 0.5 mCrab in the 150 - 600 keV energy range with 3 - 5 minutes of arc angular resolution. These are significantly higher sensitivity and resolution goals than those of current instruments. The design focus on sensitivity would make the instrument equally suitable for national security applications in the detection of weak shielded illicit radioactive materials at large distances (100 m - 1 km).

X-ray and gamma-ray imaging designs for astrophysics and security applications typically utilize a coded aperture imaging technique. The spatial resolution necessary, however, coupled with the specification of high sensitivity, requires a large number of readout channels (resulting in high cost and complexity) and limits the use of this technique to relatively low energies. As an alternative approach, an investigation is made here of the rotating modulator (RM), which uses primarily temporal modulation to record an object scene. The RM consists of a mask of opaque slats that rotates above an array of detectors. Time histories of counts recorded by each detector are used to reconstruct the object scene distribution.

Since a full study of RM characterization and capabilities has not been performed prior to this work, a comprehensive analytic system response is presented, which accounts for realistic modulation geometries. The RM imaging characteristics and sensitivity are detailed,

including a comparison to more common hard x-ray imaging modalities. A novel image reconstruction algorithm is developed to provide noise-compensation, super-resolution, and high fidelity.

A laboratory prototype RM and its measurement results are presented. As a pathfinder mission to an eventual astrophysics campaign, a one-day high-altitude balloon-borne RM is also described, including expected performance and imaging results. Finally, RM designs suitable for next-generation astrophysics and security applications are presented, and improvements to the RM technique are suggested.

Chapter 1

Introduction

1.1 Astrophysical Motivation

1.1.1 Science

A high-sensitivity hard x-ray/gamma-ray telescope with good energy and angular resolution would be capable of performing an all-sky census of black holes over a wide variety of obscuration and accretion rates: from stellar-mass black holes in our own galaxy, to intermediate mass (100 - 1000 solar mass) black holes produced by the very first stars, to supermassive black holes located in the nuclei of galaxies. Such a census would investigate how black holes form and evolve and their role in the development of the universe and was a prime goal of NASA's proposed Black Hole Finder Probe (BHFP) mission. BHFP was intended as a pathfinder mission to the Black Hole Imager, one of the "vision" missions of NASA's Beyond Einstein program (National Research Council, 2007) and a high priority in several NASA strategic plans (NASA, 2006).

The BHFP census (Grindlay et al., 2001; McConnell et al., 2004) required a 5σ flux sensitivity in a 1-year observation of 5×10^{-13} erg cm^{-2} s^{-1} ($\sim$ 0.02 mCrab) in the 10 - 150 keV energy range, and 0.5 mCrab in the 150 - 600 keV energy range, comparable to the sensitivity of the all-sky soft x-ray (0.5 - 2.5 keV) survey conducted by ROSAT (Truemper,

1982). The BHFP census would be approximately 1000× the sensitivity of the HEAO-1 A4 sky survey (Matteson, 1978), 1 - 20× the sensitivity of Swift (Gehrels et al., 2004), and 20× the sensitivity of CGRO/BATSE (Fishman et al., 1989) for gamma-ray bursts (GRBs). The field of view (FOV) of the instrument must allow for full-sky observation once per orbit and 20% continuous observation per source. Angular resolution of 3 - 5 arcmin is necessary to avoid source confusion from the more than 30,000 Active Galactic Nuclei (AGN) that will be detected (McConnell et al., 2004). Bright sources must be measured with a location accuracy of 10 arcsec to allow for follow-up observations by other satellite- or ground-based narrow-field instruments.

Sensitivity in the 10 - 600 keV energy range is especially important to a black hole survey mission, since most accreting black holes are hidden by interstellar dust and gas. While hard x-rays, infrared, and radio waves can all penetrate this dust with relatively little attenuation, x-rays have the advantage in that emission from accreting black holes is more easily distinguished from stellar emission. Additionally, previous missions have indicated the importance of this energy band to the discovery of black hole candidates. The Swift all-sky survey (Tueller et al., 2008) showed that the majority of astronomical sources in the energy range 15 - 200 keV are due to AGN, bright centers of galaxies believed to harbor supermassive black holes; the majority of these AGN were determined to be Seyfert galaxies. Observations in this same energy range by the IBIS telescope on INTEGRAL (Revnivtsev et al., 2004) showed that most sources located in the vicinity of our Galactic Center are due to X-ray binaries, many of which consist of a donor star accreting onto a black hole.

The positron annihilation line, at 511 keV, is the brightest emission line observed in the Galaxy (Purcell et al., 1997; Knödlseder et al., 2005), but its source is poorly understood. Black hole binaries, supernovae, pulsars, and cosmic ray interactions could play a role, while some of the more exotic recent suggestions include dark matter annihilation (Boehm et al., 2004; Cassé et al., 2004) and light super-conducting strings (Ferrer & Vachaspati, 2005). Gamma-ray observations of the Galactic bulge region, where the distribution of 511 keV is

brightest, would, with good angular resolution, better define the source distribution and help to uncover the origin of this emission.

The most sensitive gamma-ray imager currently flying – the SWIFT Burst Alert Telescope (BAT) coded aperture – observes up to ~ 150 keV (Gehrels et al., 2004). Sensitivity extending to energies up to 1 MeV would allow for the observation of an even wider range of astrophysical phenomena and would provide a means of achieving broader scientific objectives. As seen in data from CGRO/COMPTEL (McConnell et al., 2000), observation at energies above a few hundred keV could help to determine the underlying mechanism of non-thermal emission from Galactic black hole candidates (GBHCs) such as Cygnus X-1. Cygnus X-1 has been observed at energies up to several MeV in order to precisely define its spectrum (McConnell et al., 2002). The x-ray/gamma-ray sky is highly variable, and GBHCs are no exception to this rule. GBHCs are typically quiescent, with sudden bright outbursts in which they are observed in one of two states: the "breaking gamma-ray state," associated non-thermal state observed in the x-ray regime, and the "power-law gamma-ray state," associated with the x-ray thermal state (Grove et al., 1998; Remillard & McClintock, 2006). A new survey telescope should precisely define the spectra of GBHCs up to high energies during outburst to determine in which of these states the source currently resides and to observe candidates transitioning between the two states. Such observations would provide observational data to help constrain theoretical models.

A wide energy band would also enable the observation of the peak energy in GRBs, low energy cyclotron lines, and nuclear lines. In particular, supernovae and their remnants could be studied via the nuclear lines from radioisotopes emitted by these events. Type Ia supernovae, resulting from thermonuclear explosions of white dwarf binaries, emit their energy primarily in the form of radioactive ^{56}Ni and its daughter isotope, ^{56}Co. It is the energy released from these decays that is the primary cause of the optical emission. In gamma rays, the 158, 749, and 812 keV lines of ^{56}Ni and the 511 and 847 keV lines from ^{56}Co may be observed. Precise identification (requiring good energy resolution) combined

with flux measurements of these nuclear lines could help to determine the amount of ^{56}Ni in the supernova, and help to distinguish between Chandrasekhar and sub-Chandrasekhar mass explosions (Boggs, 2006).

Type II supernovae, explosions resulting from stellar core collapse, may be identified by the 68 and 78 keV nuclear lines of ejected ^{44}Ti. Additionally, it is believed a short flash of 511 keV due to ^{18}F (158 min half-life) should accompany Type II supernovae. It has also been suggested that other nuclear lines such as 478 keV from ^{7}Be could be used to indicate the supernova rate in the Galaxy (Grindlay et al., 2001). A broad energy range extending well beyond the 150 keV upper bound of the Swift survey (Tueller et al., 2010), good sensitivity, and high angular resolution are thus essential for a next generation hard x-ray/gamma-ray telescope mission.

1.1.2 Current and Proposed Missions

Since the late 1970's, numerous hard x-ray/gamma-ray instruments have sucessfully launched and observed in this energy regime. Two of the most recent missions employ coded aperture imaging techniques (see Sec. 2.2) and are described below.

The International Gamma-Ray Astrophysics Laboratory (INTEGRAL; Winkler et al. 2003), launched in 2002, is composed of four instruments independently focused on high-resolution spectroscopy, high-angular resolution imaging, and monitoring. Of these, the IBIS telescope (Ubertini et al., 2003) consists of a 1 m^2, 1.6 cm-thick tungsten coded mask with a uniformly redundant array pattern (Fenimore & Cannon, 1978). A detection plane of 0.26 m^2 of cadmium telluride (CdTe) sits beneath the mask, providing imaging sensitivity between 15 - 150 keV. A second detection plane of 0.29 m^2 of cesium iodide (CsI) scintillators sits below the CdTe, and provides imaging sensitivity up to 10 MeV using the Compton interaction of the incident gamma rays in the top detection plane. The full energy range is defined for a $9^\circ \times 9^\circ$ (fully-coded) FOV and $12'$ angular resolution. An active Bismuth Germanate

(BGO) veto system combined with passive collimation lead shields reduces the background contribution from cosmic diffuse flux and Earth albedo.

The Swift Gamma-Ray Burst Mission (Gehrels et al., 2004), launched in 2004, consists of three instruments for observation at hard x-ray, soft x-ray, and ultraviolet/optical energies to observe and study GRBs. The Burst Alert Telescope (BAT) is a random pattern coded aperture-based telescope, sensitive to an energy range of 15 - 150 keV. The mask consists of $5 \times 5 \times 1$ mm-thick lead pixels to provide $1'$ - $4'$ angular resolution. The detection plane, totaling 0.52 m^2 in area, is comprised of $4 \times 4 \times 2$ mm pixels of cadmium zinc telluride (CZT), a solid-state device with good energy resolution ($< 3\%$ at 662 keV). The geometry provides a $100^\circ \times 60^\circ$ (half-coded) FOV. A graded passive shield minimizes background.

Both INTEGRAL and Swift have been highly successful in their missions. The focus of INTEGRAL has been on sensitivity to a wide energy range, but with a narrow FOV. Swift has a wide FOV for the purpose of generating an all-sky catalog, but is only sensitive to energies up to 150 keV. Design for the next generation of x-ray/gamma-ray telescopes has focused on sensitivity, allowing for the observation of a wide range of sources, in particular for those that are dim, far away, or obscured by dust and gas. Two telescope designs that were proposed to the National Academy of Sciences' Beyond Einstein Program Assessment Committee (BEPAC) to satisfy the scientific requirements of the BHFP are examples: Like the majority of prior successful missions operating in this energy range, both of these instruments – EXIST and CASTER – used a coded aperture imaging modality, though the telescopes differed primarily in the detection plane design.

The Energetic X-ray Imaging Survey Telescope (EXIST; Grindlay et al. 2001, 2003) combined three coded aperture telescopes to achieve a $180^\circ \times 75^\circ$ FOV. Each sub-telescope featured a quasi-hemispherical tungsten coded aperture mask. To achieve the angular resolution necessary for the BHFP, the mask pixels were 2.5 mm wide, while a thickness of 7 mm provided sufficient attenuation for 600 keV gamma rays. The EXIST detection planes were composed of 5 - 10 mm thick CZT. Each detector plane had an array of 1.0 mm pixels

totaling 2.7 m^2 (8.4 m^2 total). Collimating shields located beneath and to the sides of the detector array attenuated Earth albedo and provided cosmic ray-induced background rejection. Additionally, the shield outputs were pulse-height analyzed for GRB detection up to 10 MeV. This large-area detector system required 4.6×10^6 electronics channels (Grindlay et al., 2003).

The Coded Aperture Survey Telescope for Energetic Radiation (CASTER; McConnell et al. 2004, 2005) addressed the costly nature of EXIST due to its use of CZT and large number of readout channels. CASTER was proposed as an array of coded aperture telescopes, each with a 5 mm-thick graded tungsten mask composed of 1.3 - 2.2 mm pixels. Inorganic scintillators, which have a long record of successful use in laboratory and space-based missions, were used as the primary detectors. The CASTER detection plane was an Anger array, a grouping of photomultiplier tubes (PMTs) optically coupled to the backside of a sheet of cesium iodide (CsI), sodium iodide (NaI), or lanthanum bromide ($LaBr_3$) scintillator. By measuring the relative outputs of the PMTs, the photon interaction location could be determined with a spatial resolution better than the spacing of the PMT anodes, significantly reducing the required number of channels when compared to EXIST. For a total 8 m^2 detector area, 3.2×10^5 channels were required.

Both CASTER and EXIST were considered and criticized by BEPAC for being too costly and complex (National Research Council, 2007). This result, arising primarily from the large number of readout channels required, was a consequence of the fine spatial resolution necessary for a coded aperture-based design. Combined with the scientific requirement of high sensitivity and good energy resolution, these telescopes could not satisfy the budget constraints of the Beyond Einstein probe missions. Although the Swift and INTEGRAL coded aperture telescopes have proven extremely successful, new imaging technology must be developed if a next generation follow-on mission (operating with high sensitivity and good efficiency above 150 keV) is to be feasible.

1.2 National Security Motivation

A hard x-ray/gamma-ray imager capable of the scientific performance described in Sec. 1.1.1 would also be suitable as a stand-off radiation detection system for security applications. In particular, the emphasis of next-generation designs on sensitivity makes the instrument capable of detecting weak, distant (100 m - 1 km) illicit nuclear materials. Such a capability would be useful for the surveillance of trucks at border crossing points or the monitoring of ships entering a port, both of which could potentially be carrying contraband radioactive cargoes (Cherry et al., 2006).

Fissionable nuclear materials shipped illicitly into the United States pose a serious threat to national security, as do lost and abandoned sources and waste from medical clinics (U.S. Congress, Office of Technology Assessment, 1995b,a; National Research Council, 2002). Though these sources are not useful for producing nuclear bombs, highly effective weapons of mass destruction and Radiological Dispersion Devices (RDDs) could be assembled from these readily available materials. An RDD, or "dirty bomb," employs the use of a non-nuclear explosive material (e.g. dynamite) to spread a radioactive material, presumably in a high population center, resulting in long term health hazards, psychological effects, and possibly immense decontamination and cleanup costs. While no terrorist or criminal organizations have ever exploded such a device to date, they have demonstrated the ability to obtain small amounts of ^{235}U and ^{239}Pu. Additional isotopes such as ^{137}Cs, ^{60}Co, and ^{241}Am are available as medical waste and in commercial irradiators and could also potentially be obtained and used in an RDD (Intl. Atomic Energy Agency, 2000; Ferguson, 2003).

To produce an effective device, gamma-emitting sources (e.g., ^{60}Co, ^{137}Cs, ^{192}Ir) would need to be procured in gram quantities for an RDD; enriched uranium or weapons-grade plutonium would need to be obtained in kilogram quantities for a bomb (U.S. Congress, Office of Technology Assessment, 1995b). Such large quantities would require shielding during preparation and transportion to protect the terrorists or criminals from hazardous radioactivity and partly to hinder detection before detonation. If ^{238}U or ^{60}Co is encased in

an 8.5 cm-thick lead shield, only 10^{-3} of the 1 MeV gamma rays escapes without interacting. In the case of 186 keV photons from ^{235}U decay, 5 mm of lead provides equivalent shielding. A stand-off detector must therefore be capable of detecting Curie quantities of radiological materials in the presence of a 10^3 factor of attenuation. Such severe attenuation of the radioactive flux broadens sharp nuclear lines. An effective detection instrument, therefore, need not have exceptionally fine energy resolution; instead, the focus should be on high sensitivity (Reilly et al., 1991; DOE, 2001). An additional premium is put on sensitivity due to the long distances over which the radiological materials should be detected. Sensitivity can be achieved by maximizing the detection plane area, minimizing background contributions, and optimizing angular resolution.

The United States Department of Energy's Domestic Nuclear Detection Office (DNDO) outlined the short term requirements and longer term goals of a radiation-monitoring device in the Stand-off Radiation Detection System Advanced Technology Demonstration program (Broad Agency Announcement BAA07-01). The constraints on the system design are independent of the detailed instrument, and provide a framework for the design of such a security device. Energy resolution must be 7.5% at 662 keV, with a goal of reaching 2%. Detection efficiency must be capable of detecting at least 15 photons per second in a 662 keV photopeak from a 1 mC ^{137}Cs source at a distance of 100 m. Finally, the FOV must span 115° with angular resolution less than 17°. If an instrument is to be deployed at a large number of border crossings, harbors, etc., costs must be minimized. Several coded aperture designs have been proposed for this application (e.g., see N06/N14 of Yu 2009), and there is a need for improved and more cost-effective approaches.

1.3 Plan of the Dissertation

The motivation of this dissertation is to address the shortcomings of the coded aperture technique to accomplish many of the scientific objectives outlined above for NASA's BHFP

by investigating an alternative approach to hard x-ray/gamma-ray imaging. This technique should satisfy the sensitivity and resolution requirements of a black hole census astrophysics mission at reduced cost and complexity. A telescope developed for this application would be equally suitable for national security, whereby trucks and ships could be monitored for shielded illicit nuclear materials. Coded apertures have been used extensively in both applications due to the fine position resolution achievable and minimal sidelobe contributions; the necessary spatial resolution however, increases the number of readout channels and raises costs. The subject of this dissertation is therefore an alternative imaging modality, rotating modulation, in which temporal and spatial modulation work in concert to encode transmissions from the FOV prior to interaction in an array of detectors. The combined use of these two modulations enables the RM to function without a heavy second grid as seen in some temporal modulators and to achieve good angular resolution comparable to the coded aperture with fewer detector elements and readout channels. Furthermore, as will be shown, the RM sensitivity is comparable to that of the coded aperture in a high-energy, wide FOV regime.

In Chap. 2, the rotating modulation technique is introduced through a brief history of its precursors, a comparison of alternate modalities, and an explanation of its basic mechanism. Chap. 3 examines the characteristic system response of an RM, a necessary component in solving the inverse problem of image reconstruction. An overview of several reconstruction algorithms is provided in Chap. 4, including a novel method developed specifically for the RM that resolves images with little-to-no spurious noise fluctuations and provides "super-resolution" beyond the geometric limit of the instrument. Sensitivity characteristics of the RM are compared against those of the coded aperture and a common temporal modulation technique in Chap. 5.

The development of a laboratory-scale RM prototype is described in Chap. 6. Measurements and characteristics are presented in support of the RM concept analysis results. In Chap. 7, a design for a high-altitude balloon-borne mission is proposed as a pathfinder

experiment to a future BHFP. The mission's expected performance and simulation results are presented, based on the extrapolation of measurements made on a high-altitude balloon flight of the detector material. The implications of these results, including the possibility of a balloon campaign to satisfy some of a black hole census' science requirements and a stand-off radiation detection system to detect illicit nuclear materials are discussed in Chap. 8.

Chapter 2

X-ray / Gamma-ray Imaging

2.1 Introduction

At optical wavelengths, photons are traditionally imaged using a refractive lens or a parabolic mirror angled at nearly normal incidence to the photon trajectory (Schwarzschild, 1905). In these cases, the observed data constitute a direct image of the object scene requiring little or no subsequent modification. As refractive indices of a material are a function of electromagnetic frequency, refraction and reflection play a smaller role for high-energy photons. For soft x-rays, compound grazing incidence mirrors (Wolter, 1952) have been successfully used (e.g. Weisskopf et al., 1995), but above ~ 50 keV, these techniques become less useful (in 't Zand, 2009).

One popular class of techniques to image the hard x-ray and soft gamma-ray spectrum (10 keV - 1 MeV) employs an encoding modulation of incident photons prior to detection. These "multiplexing" techniques are so named since each detection element responds to intensity characteristics of multiple regions of the observable field (in 't Zand, 2009). Consequently, subsequent analysis is required to "redistribute" these measurements to the object scene, thus perfoming image reconstruction. Here, we examine two of the most often-used multiplexing

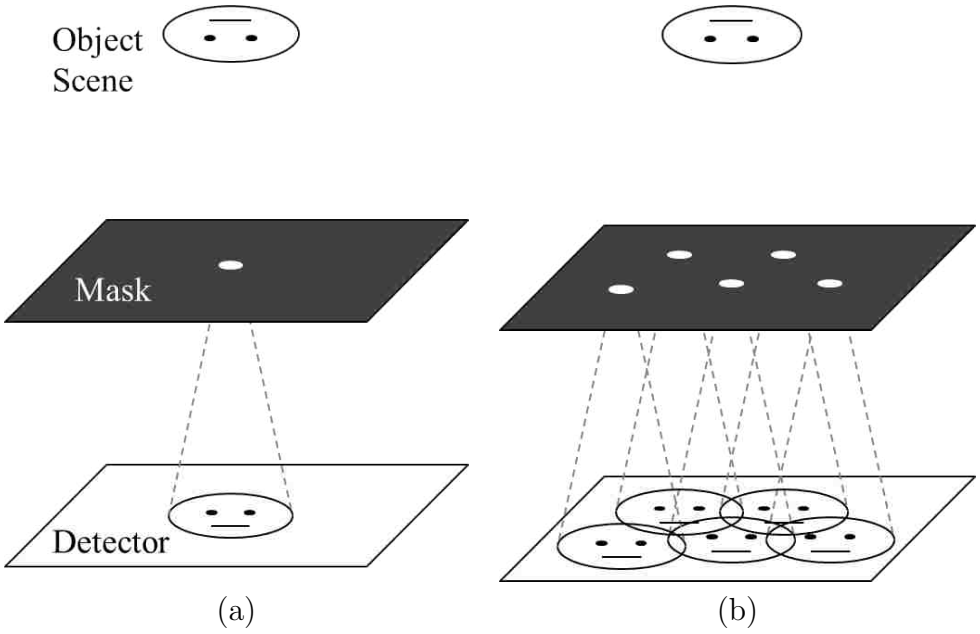


Figure 2.1: (a) The simplest imager design, the pinhole camera. (b) The multi-pinhole camera.

imagers, the coded aperture and the rotating modulation collimator, and we present the mechanism for the rotating modulator.

2.2 Coded Mask

The simplest imaging technique works for any wavelength of light: the pinhole camera (Fig. 2.1a). It employs a single "pinhole"-sized aperture above a position-sensitive detection plane. As light from the object scene is incident on the mask, all photons are blocked except those which pass through the aperture. The interaction location in the detector plane is uniquely correlated to the photon's incident angle, and so the sum of measurements constitutes a direct image distribution. Since all but a small percentage of the incident light is blocked, sensitivity is inherently low. At the cost of this sensitivity, angular resolution is improved by shrinking the aperture size, a. The ratio of the aperture width to its distance from the

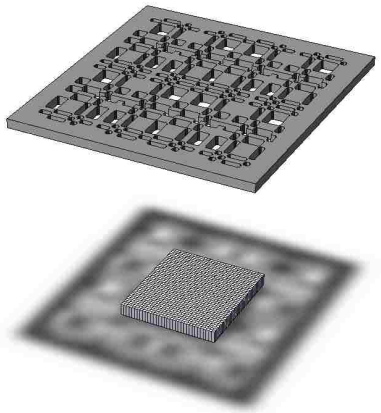


Figure 2.2: Coded mask system with mask suspended above position-sensitive detection plane.

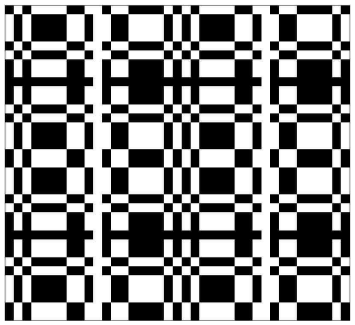


Figure 2.3: Basic pattern of a 17×19 Fenimore URA coded mask.

detection plane, L, provides the resolution of the system,

$$\delta\theta = \frac{a}{L}. \tag{2.1}$$

An effort to increase the sensitivity of the pinhole concept was introduced by Ables (1968) and Dicke (1968), whereby multiple pinholes are used to generate overlapping images on the detection plane (Fig. 2.1b). Angular resolution is unaffected, though the sensitivity is increased by a factor $\sqrt{N}$, where N is the number of pinholes present. Since the image offsets reflect the location of the apertures, the measured data can be deconvolved with the mask geometry to reconstruct the object scene.

Contemporarily, this spatial-encoding technique is referred to as a "coded mask" system (Caroli et al., 1987) and typically features 50% transmission with a half open/half closed configuration (Fig. 2.2). Pixel patterns are chosen such that sidelobes, artifacts of the mask pattern which arise in the image reconstruction process, are nullified. That is, patterns are sought whose autocorrelation yields a Kronecker delta function, such that any one shadow distribution corresponds uniquely to a single location in the field of view.

One common pattern, for example, is the uniformly redundant array (Fenimore & Cannon, 1978). When compared to a randomly distributed mask pattern, the URA provides a

greater signal-to-noise ratio (SNR) and fewer reconstruction atrifacts. The URA distribution, $A(i,j)$, is an $r \times s$ array of open/transparent (1) and closed/opaque (0) pixels where r and s are prime numbers, and $r - s = 2$:

$$A(i,j) = \begin{cases} 0 & \text{if } i = 0 \\ 1 & \text{if } j = 0, i \neq 0 \\ 1 & \text{if } C_r(i)C_s(j) = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{2.2}$$

where

$$C_\tau(i) = \begin{cases} 1 & \text{if there exists an integer } x, 1 \leq x < \tau \text{ such that } i = x^2 \bmod \tau \\ -1 & \text{otherwise.} \end{cases} \tag{2.3}$$

The mask is manufactured by creating a mosaic of the central URA, with dimensions twice that of the central URA (Fig. 2.3).

If the coded pattern is well chosen and the deconvolution technique sound, images may be produced with a coded mask camera that suffer little from the overlapping nature of the data. Mask pixels may be kept small to achieve good angular resolution without sacrificing sensitivity. The detection plane must, however, be capable of describing the shadow distribution accurately. (To satisfy the Nyquist criterion, the detector pixels must be at least twice as small as the mask pixel size.) For a high-sensitivity large-area design such as CASTER (McConnell et al., 2004) or EXIST (Grindlay et al., 2001), this requirement results in a large number of pixels and consequently readout channels.

2.3 Rotating Modulation Collimator

Oda (1965) first proposed the use of a modulation collimator (MC) to address the inability

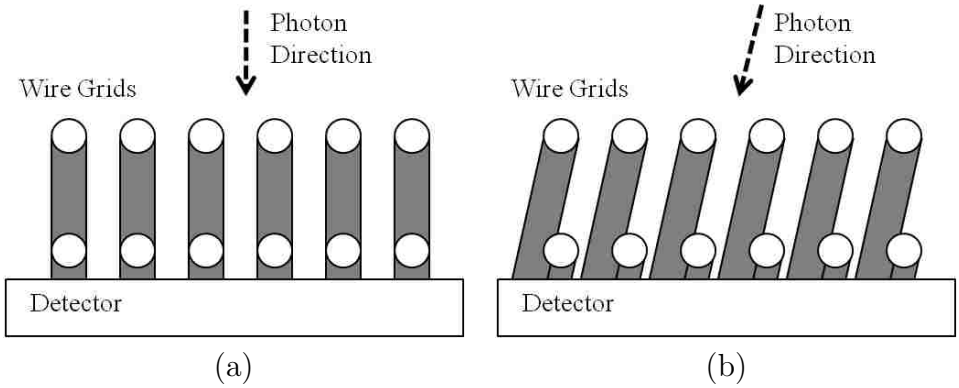


Figure 2.4: A modulation collimator measures (a) a high count rate when the wire grids are aligned with the angle of photon incidence, and (b) a low count rate otherwise. (Adapted from Oda 1965.)

of a single-pixel device to achieve good angular resolution. Two offset wire grids form the modulation device. One grid sits just above the detector while the other sits a distance L overhead (Fig. 2.4). For a single source, the top grid shadow falls exactly onto the bottom grid wires, exactly onto the spacings between them, or a combination thereof, resulting in a difference of intensity measurement by the detector depending on the angle of the source. Slewing of the instrument alters the angle of incidence, and so a time history of intensity measurements is recorded that can be correlated (with resolution according to a/L) to a source location in one dimension. Unlike narrow-field collimated devices, the MC allows for a wide FOV to be examined at once (in one dimension) while maintaining good angular resolution.

Mertz (1967) suggested an improvement to the MC design by rotating the wire grids (or slats at higher energies) in tandem about a central axis, sampling data in two dimensions (Fig. 2.5). Rotating modulation collimators (RMCs; Schnopper et al. 1968) have seen use in rocket (Schnopper et al., 1970) and balloon (Crannell et al., 1986; Gaither et al., 1996) experiments, as well as space-based missions with sensitivity to a range of x-ray and gamma-ray energies: e.g. Ariel-V (3 - 7 keV; Carpenter et al. 1976), SAS-3 (1.5 - 60 keV; Doxsey et al. 1976), and RHESSI (3 keV - 17 MeV; Hurford et al. 2002). RMCs are lauded for their

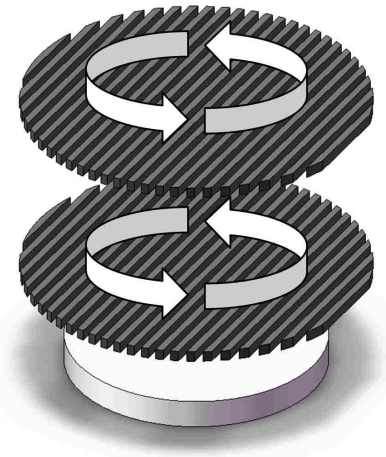


Figure 2.5: Rotating Modulation Collimator.

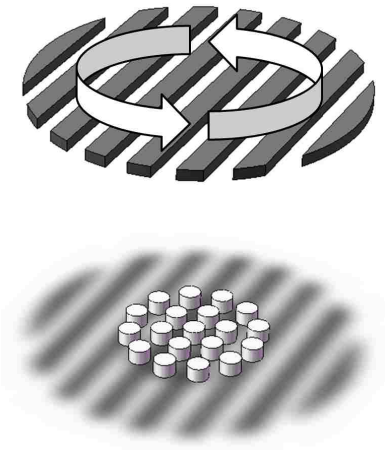


Figure 2.6: Rotating Modulator.

simplicity and ability to achieve good angular resolution, but have significant limitations. The use of two (typically 50% transparent masks) reduces the transmission throughput to half that of the coded mask, and the instrument response suffers from significant sidelobes which reduce image reconstruction fidelity (Cruise & Willmore, 1975).

2.4 Rotating Modulator

A Rotating Modulator (RM; Durouchoux et al. 1983) is another of the class of multiplexing imagers. It consists of a single grid of opaque slats spaced equidistance apart and suspended above an array of circular non-imaging detectors (Fig. 2.6). Typically, the detector diameters, slat widths, and slat spacings are equal. The grid rotates, periodically blocking the transmission of incident photons from the object scene onto the detection plane. A characteristic time history of counts is recorded by each detector for the entire length of the exposure, varying from 0 to 100% of incident exposure intensity. Like the coded aperture or RMC, these data are then correlated to a source or sources to reconstruct the object scene distribution.

The RM may be considered a trade-off between the coded aperture and RMC, featuring both temporal and spatial modulation. It features an arrayed detection system, ∼50% mask

transmission, and relatively low weight due to the use of a single modulating mask. In analogy to the coded aperture and RMC, the intrinsic angular resolution is defined by the ratio of slat spacing to mask-detector separation. Since this width is typically equal to the detector diameters in order to maximize sensitivity (as is discussed in Chap. 5), the resolution is effectively dictated by the mask-detector separation and detector diameter. Furthermore, RMs, like RMCs, typically exhibit extended sidelobes in their reconstructions (Durouchoux et al., 1983). In order for the RM to be a competitive imager with the coded aperture, it must be shown to approach or exceed the sensitivity of the coded aperture (particularly for a high-energy, wide FOV regime) and achieve resolving power beyond the geometric limit of the instrument geometry (i.e., "super-resolution") with a reconstruction technique that produces artifact-free images.

Compared to other multiplexing techniques, the RM has seen relatively little use. The WATCH experiment (Lund, 1981) is a self-described single-grid RMC, although its design is perhaps closer to that of an RM with some key differences (including slat detectors instead of circular detectors). Other than the prototype RM recently constructed in our laboratory (described in Chap. 6), we are aware of only one other RM in development (Shih et al., 2008) as of March 2011. The intention is to demonstrate the feasibility of the RM as a solution to hard x-ray/gamma-ray imaging for both a black hole census mission and for national security applications with a simple and cost effective design.

Chapter 3

RM Instrument Response[1]

The RM grid of slats modulates the transmission of photons from the object scene, S, describing a time history of counts recorded by each detector d. These data are folded modulo the mask rotation period to produce a single count profile for each detector, $O_d(t)$, which may be described by

$$O_d(t) = \sum_n P_d(t,n)S(n) + B + N(t), \tag{3.1}$$

where B is a constant background contribution and $N(t)$ is noise. $P_d(t,n)$ is the instrument response function, a collection of characteristic count rate profiles for point sources at all possible scene locations n. Chap. 4 examines deconvolution techniques for solving the inverse problem of Eq. 3.1 for the object scene S. The problem necessitates a pre-determined and well-defined modulation pattern, $P_d(t,n)$.

The technique for determining the expected modulation patterns should be computationally fast, allow for unconstrained instrument geometry, account for projection effects and non-uniform attenuation, and describe the cumulative shadowing by multiple slats simultaneously. Brute force Monte Carlo simulations accomplish these tasks, but the computation is time-consuming. A standard characteristic profile (Dadurkevicius & Ralys, 1985) that

[1]Reprinted by permission of Astrophysics and Space Science (Budden et al., 2011).

can be calculated analytically is described in Sec. 3.1. While suitable in many cases, this formula imposes tight constraints on instrument geometry and is too simplistic to account for non-uniform attenuation or shadow lengthening.

To achieve a wide FOV and high-energy sensitivity (which requires thick mask slats), a more robust analytical profile is necessary to accurately describe the instrument response. In Sec. 3.2, an advanced characteristic profile for the RM is presented that is capable of describing the complex modulation pattern and can be calculated analytically in a relatively short time. In Sec. 3.3, we show examples of count rate profiles generated with the standard and advanced formulae and with Monte Carlo simulations, and reconstructed images using modulation patterns resulting from each of these techniques.

3.1 Classical Response

The standard characteristic formula for a single-mask RM is presented by Durouchoux et al. (1983) and examined in greater detail by Dadurkevicius & Ralys (1985). An RM has slat width a, slat spacing b, and detector diameter c (Fig. 3.1). The mask is suspended a distance L from the detection plane. The standard formula imposes the constraint $a = b = c$ and assumes that the slats have infinitesimal thickness but attenuate 100% of the incident photons. An attenuation coefficient may be applied to correct for transmission through the slats, but clipping effects, spatially non-uniform attenuation, and shadow lengthening (all effects of finite mask thickness) are ignored. For the description of both the standard formula below and the more realistic advanced formula in Sec. 3.2, the mask is assumed to be centered midway between two slats and begins its period with the bars parallel to the lab frame's $\hat{y}$ direction; a simple transformation and offset parameter, however, easily provide an alternate case.

A point source in the object scene has intensity I_0 (as measured by a detector in the absence of a mask) and is located at azimuthal angle ϕ and zenith θ. For a detector centered

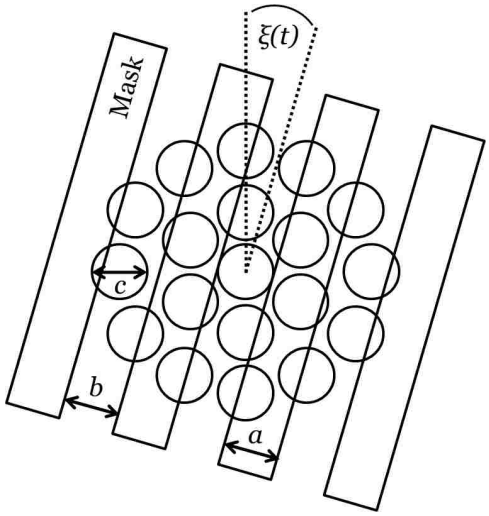


Figure 3.1: Top-down view of the RM geometry. A mask with slat width a and slat spacing b rotates above an array of circular detectors with diameter c, according to the function $\xi(t)$.

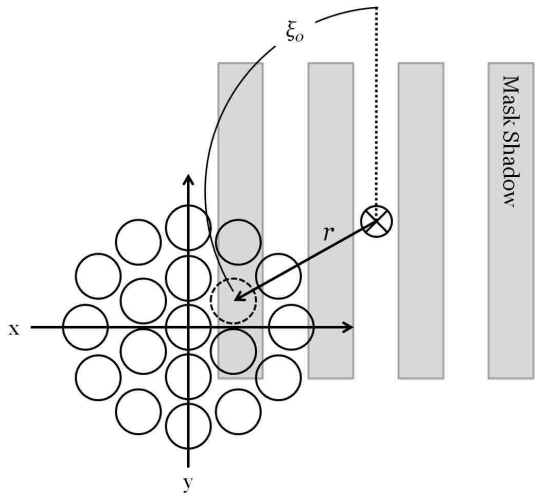


Figure 3.2: Top-down view of the RM, describing the polar location (r, ξ_0) of a particular detector (dashed outline) relative to the mask shadow origin ($\bigotimes$).

at (x_0, y_0) in the lab frame (relative to an origin coincident with the mask's rotational axis), a detector's polar coordinates (r, ξ_0) relative to the mask shadow's projected axis (Fig. 3.2) are given by

$$r = \sqrt{(x_0 + L\tan\theta\cos\phi)^2 + (y_0 + L\tan\theta\sin\phi)^2}, \tag{3.2}$$

and

$$\xi_0 = \tan^{-1}\left[\frac{y_0 + L\tan\theta\sin\phi}{x_0 + L\tan\theta\cos\phi}\right]. \tag{3.3}$$

The mask's angular orientation is $\xi(t) = \omega t$ for frequency ω. The x-component of the leading edge of the first slat in the $-\hat{x}$ direction from the origin is given by

$$x(t) = r\cos\left(\xi(t) + \xi_0\right). \tag{3.4}$$

To account for the periodic traversal by multiple slats, transmission through the spacings between slats, and symmetry of the system, a modified x coordinate is defined in units of

detector diameter a,

$$x^*(t) = 1 - \left| \left| \frac{x(t)}{a} \right| \bmod 2 - 1 \right|, \tag{3.5}$$

where the function $x \bmod y$ provides the remainder of x/y. The x^* component defines the traversal across the detector diameter of the leading edge of whichever bar shadows the detector at time t. The percentage of detector face shadowed is given by the integration over the area of a circle from zero to a fractional distance τ across its diameter,

$$F(\tau) = \frac{1}{\pi} \cos^{-1}(1 - 2\tau) - \frac{2}{\pi}(1 - 2\tau)\sqrt{\tau - \tau^2}. \tag{3.6}$$

The characteristic count rate profile measured by detector d is found by subtracting from 100% transmission the fractional shadowing of the detector described by Eq. 3.6 (with x^* as the input variable), scaled to the intensity of the source:

$$P_d(t) = I_0 \left(1 - F\left[x^*(t)\right]\right). \tag{3.7}$$

Examples of profiles generated with this formula are shown in Fig. 3.3.

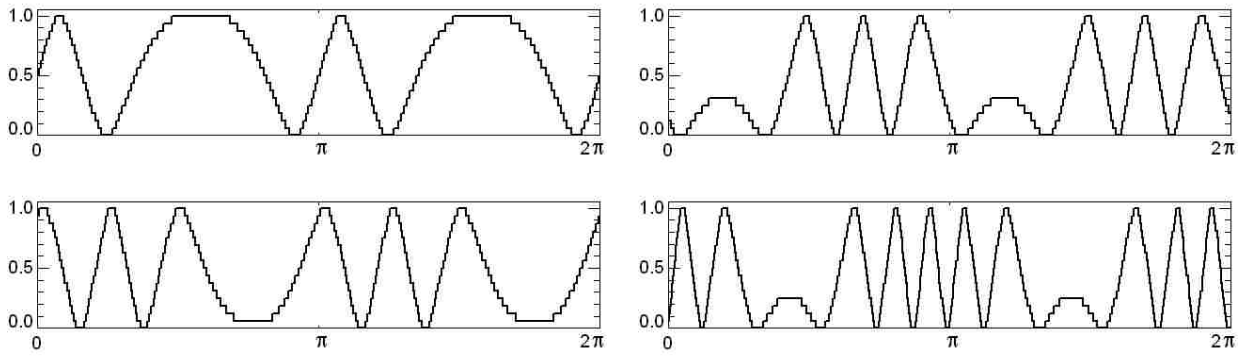


Figure 3.3: Four sample RM count profiles generated for different detectors for a single point source using the standard characteristic formula. The horizontal axis covers a full rotation.

3.2 Advanced Characteristic Formula

3.2.1 Introduction

Particularly at high energies where the mask may not be 100% opaque, a more realistic advanced characteristic profile must describe attenuation, clipping, and shadow-lengthening effects as a function of photon energy, and the cumulative shadowing of a detector by multiple slats simultaneously. The RM slat width a, slat spacing b, and detector diameter c should be defined independently to enable a comprehensive analysis of instrument sensitivity and resolution, and a finite slat thickness of h should be incorporated. An improved characteristic profile satisfying these requirements is described in the following section.

3.2.2 Description

A single slat shadow may be divided into three regions (Fig. 3.4). The middle region of the shadow is due to attenuation of photons that are incident on the "full thickness" portion of the slat, i.e. the photon trajectory penetrates both the top and bottom face of the slat; the resulting shadow is spatially uniform. The other two regions reside on the outside of the full thickness shadow and are a result of the attenuation of photons whose trajectory "clips" the slats, i.e. the trajectories pass through either the top or bottom face of the slats,

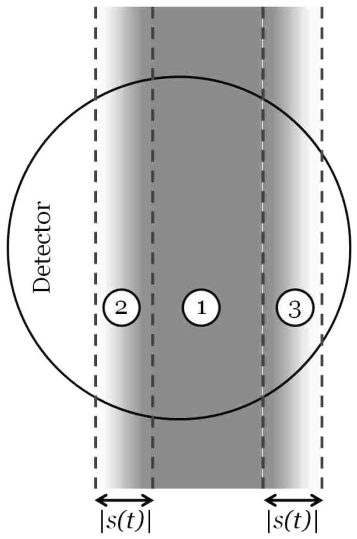


Figure 3.4: Top-down view of slat shadow on a detector face. A slat shadow in the advanced formula is divided into three regions: the "full thickness" region (1) where the shadow opacity is uniform, and the outer clipping sections (2,3) where the shadow opacity decreases exponentially away from region (1).

but not both. Since the distance a photon travels through this portion of the slat varies depending on how close it is to the slat edge, the opacity of this shadow will be non-uniform with exponential decrease away from the full thickness region.

As the RM mask rotates, the width of a slat shadow varies due to the finite slat thickness by an amount $|s(t)|$, where

$$s(t) = h \tan\theta \sin(\xi(t) - \phi). \tag{3.8}$$

The absolute value of this parameter is the width at time t of either of the two clipping shadow regions. Because of this effect, the point of symmetry in the mask shadow shifts by an amount $s(t)/2$. The x-component of the leading slat shadow is modified from Eq. 3.4 to become

$$x'(t) = r\cos(\xi(t) + \xi_0) + \frac{s(t)}{2}, \tag{3.9}$$

where the polar coordinates (r, ξ_0) for the starting position of the detector in the frame of the projected mask shadow for a source at (θ, ϕ) are still given by Eqs. 3.2 and 3.3. Similarly, the width of the "full thickness" shadow region shrinks relative to a due to the increased percentage of incident photons that clip the slats (Fig. 3.5),

$$a'(t) = a - |s(t)|. \tag{3.10}$$

A coordinate x^* is defined (analogous to Eq. 3.5) to account for the mask's periodic nature, symmetry, and in this case, the ability for multiple bars to shadow the detector face simultaneously. Consequently, x^* is an array of time-dependent functions,

$$x^*_m(t) = \frac{1}{c}\left(\frac{1}{2}\left[a'(t) + c\right] - \left||x'(t)| \bmod (a+b) - (1+2m)\frac{a+b}{2}\right|\right), \quad -M \le m \le M, \tag{3.11}$$

where index m spans the integer values from $-M$ to M. The size of the array is equal to

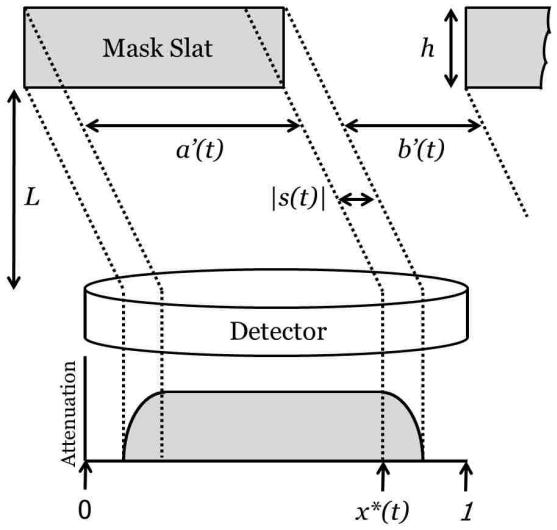


Figure 3.5: Side view diagram of the time-variable mask shadow parameters. The projected leading (bottom) edge of the bar position is given by $x^*(t)$. The full-thickness shadow region has width $a'(t)$, the clipping regions width $|s(t)|$, and $b'(t)$ describes the slat shadow spacing.

the total number of bars that may simultaneously shadow any one detector, $1 + 2M$, where M is the integer given by[2]

$$M = \text{floor}\left(\frac{a + s_{\max} + c}{a + b}\right). \tag{3.12}$$

The maximum $|s(t)|$ value, $s_{\max}$, occurs when the mask angle is 90° and 270° out of phase with the source azimuth:

$$s_{\max} = h \tan\theta. \tag{3.13}$$

The fraction of detector area shadowed (analogous to Eq. 3.6) becomes

$$F_0(\tau) = \frac{1}{\pi}\cos^{-1}(1 - 2\Lambda[\tau]) - \frac{2}{\pi}(1 - 2\Lambda[\tau])\sqrt{\Lambda[\tau] - \Lambda[\tau]^2}, \tag{3.14}$$

where we have introduced the constraint formula,[3]

$$\Lambda[\tau] = \min\{\max\{\tau, 0\}, 1\}. \tag{3.15}$$

In the standard formula from Sec. 3.2.2, this constraint is not necessary since only a single slat is evaluated, and its width equals the detector diameter. Since multiple bars

[2]The function floor(x) rounds x down to the nearest integer value.

[3]The function $\max\{x, y\}$ returns the maximum value of the two input values, and similarly, the minimum for $\min\{x, y\}$.

may now simultaneously contribute to the shadowing, the constraint limits the integration of shadow between the two edges of the detector area.

Eq. 3.14 and the Beer-Lambert law (which provides the transmission factor of photons through a material) determine the shadowing of the detector from all contributing slats,

$$G_0(t) = \left(1 - \exp\left[-\frac{h\sigma\rho}{\cos\theta}\right]\right)\left[F_0\left(x_m^*(t) + \frac{|s(t)|}{c}\right) - F_0\left(x_m^*(t) - \frac{a'(t)}{c} - \frac{|s(t)|}{c}\right)\right], \quad (3.16)$$

where σ is the mass attenuation coefficient (for a particular photon energy) and ρ is the density of the slat material. G_0 includes the shadow contributions from all three regions of the slat shadow with constant attenuation assumed. To account for the additional transmission through the clipping portion of the shadow, a function F_1 is defined,

$$F_1(\tau, t) = \frac{4}{\pi}\int_{\Lambda[\tau]}^{\Lambda[\tau+|s(t)|/c]} \exp\left[-Z(t)(\tau - x)\right]\sqrt{1-(2x-1)^2}\,\mathrm{d}x + F_0(\tau) - F_0\left(\tau + \frac{|s(t)|}{c}\right), \quad (3.17)$$

where

$$Z(t) = \frac{hc\sigma\rho}{|s(t)|\cos\theta}. \quad (3.18)$$

The first part of Eq. 3.17 is the integration of the transmission fraction of the exponentially-decreasing shadow opacity about the circular geometry of the detector; the second part removes the transmission (or lack of attenuation) already accounted for in Eq. 3.16. The transmission by clipped photons not accounted for in Eq. 3.16 is given by

$$G_1(t) = \exp\left[-\frac{h\sigma\rho}{\cos\theta}\right]\left[F_1(x_m^*(t), t) + F_1\left(1 - x_m^*(t) + \frac{a'(t)}{c}, t\right)\right]. \quad (3.19)$$

The advanced characteristic count rate profile combines these functions:

$$P_d(t) = I_0 \left(1 - \sum_m \left[G_0(t) - G_1(t) \right] \right). \tag{3.20}$$

3.2.3 Approximation for Practical Use

Equation 3.17 has no closed-form solution and is computationally expensive to evaluate numerically. Also, as the source azimuth ϕ and grid angle $\xi(t)$ align, $s(t) \to 0$ and $Z(t) \to \infty$; the span of the integral in F_1 thus approaches zero causing numerical solutions of F_1, and consequently G_1 (Eq. 3.19), to become unstable.

In most cases, the G_1 transmission contribution from the clipped photons will be a small contribution relative to G_0, and so may be simply ignored in Eq. 3.20. The shadow-lengthening effects are still described by G_0 with uniform attenuation assumed for the clipping shadow regions. A suitable approximation for G_1, however, is to use the ratio of the integral underneath the isolated exponential function in Eq. 3.17 to the full transmission, ignoring the circular shape of the detector:

$$\alpha = \frac{\cos\theta}{h\sigma\rho} \left(1 - \exp\left[-\frac{h\sigma\rho}{\cos\theta} \right] \right) - \exp\left[-\frac{h\sigma\rho}{\cos\theta} \right]. \tag{3.21}$$

A fortunate result of this approach is that α is constant over $s(t)$, and thus all grid angles. It must only be calculated once per source zenith, providing a computationally fast solution for approximating G_1,

$$\widetilde{G_1}(t) = \alpha \left[F_0 \left(x_m^*(t) + \frac{|s(t)|}{c} \right) - F_0 \left(x_m^*(t) \right) + F_0 \left(x_m^*(t) - \frac{a'(t)}{c} \right) \right.$$
$$\left. - F_0 \left(x_m^*(t) - \frac{a'(t)}{c} - \frac{|s(t)|}{c} \right) \right]. \tag{3.22}$$

3.3 Simulation and Results

We have performed custom-written Monte Carlo simulations (described in Sec. 6.4) with various instrument geometries and compared the results with the profiles derived from the advanced and, where suitable, the standard count rate formulae. For the results described below (unless otherwise indicated), a lead mask ($\rho = 11.34$ g/cm^3) is suspended $L = 1$ m above the detection plane (approximately mirroring the laboratory prototype geometry described in Chap. 6). Monoenergetic 662 keV photons have a total mass attenuation coefficient[4] of $\sigma = 0.103$ cm^2/g in the mask. Results are computed for various combinations of a, b, and c. Background is assumed to be zero, and only photopeak events are included in the analysis. Additional mask geometry and source parameters are selected for each scenario individually to demonstrate a particular advantage of the advanced formula over its standard counterpart.

3.3.1 Computational Speed

For a direct comparison of the computational expense for the advanced versus the standard formula, instrument response functions are calculated using both solutions for an RM with a 14° FOV divided into 12′ field bins (4900 elements) and count profiles broken up into 560 time bins. The calculations are performed using IDL 6.3 on a Windows machine with the software described in Sec. 6.4.

The instrument response based on the standard formula (one profile per sky bin) is computed in 0.8 s for each detector. The advanced formula, ignoring the G_1 term, takes 2.6 s, while inclusion of the approximated $\widetilde{G_1}$ term increases the time to 5.4 s. The processing time is still many orders of magnitude shorter than that required to determine the instrument response using a Monte Carlo simulation: If we require a 10σ result per time bin to derive a Monte Carlo profile that is suitable for the purposes of image reconstruction, the instrument response for a single detector takes ~ 1.3 days to compute.

[4]NIST XCOM, http://www.nist.gov/pml/data/xcom/

3.3.2 Count Rate Profiles

The count rate profiles as calculated by the two formulae are compared directly to the data recorded with Monte Carlo simulations. Two scenarios are first examined with $a = b = c = 4$ cm, so that the advanced formula can be compared directly to the standard formula (where $a = b = c$ is required).

An extreme case of shadow lengthening is demonstrated in Fig. 3.6 by increasing the thickness of the slats to 20 cm and placing a source at zenith angle $\theta = 6°$. The Monte Carlo result shows a profile that is asymmetric, with transmission peaks varying in height due to the broadened shadow. The standard formula is incapable of accurately describing this effect, since it assumes infinitesimal slat thickness; the advanced profile, however, is virtually identical to the Monte Carlo result. Though the thickness of the slats here is

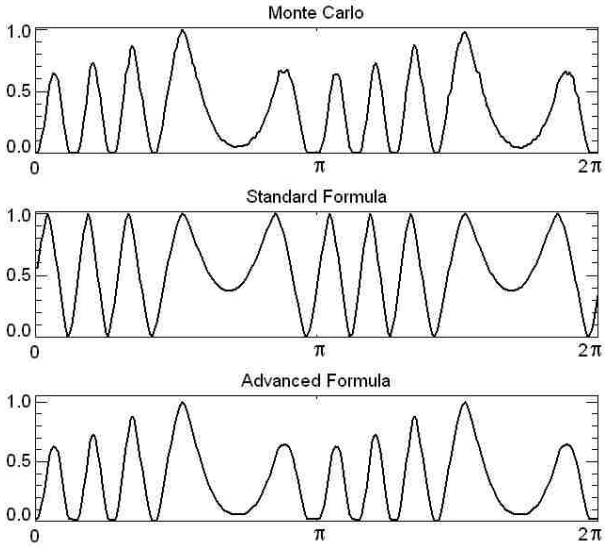


Figure 3.6: Count profiles for a mask with slat thickness 20 cm, source at $\theta = 6°$, and $a = b = c = 4$ cm. The horizontal axis is given in terms of mask rotation angle ξ, while the vertical axis is measured source intensity in arbitrary units. The result of shadow lengthening due to the large mask thickness is evidenced by the asymmetry in the profile.

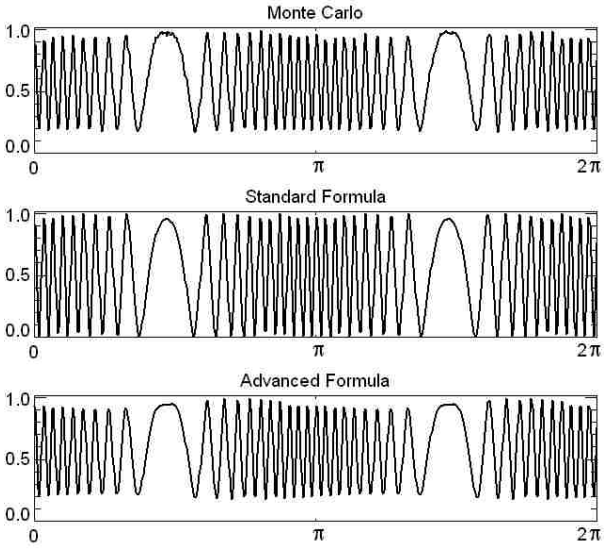


Figure 3.7: Count profiles for a mask with slat thickness 1 cm and source at $\theta = 50°$. Shadow lengthening due to the large incident angle is seen as an asymmetry in the envelope of the high frequency modulations, and the lack of attenuation due to the thin mask appears as a reduction of contrast in the profile.

greatly exaggerated for visual demonstration, the purpose is to describe a thick mask to image higher energy gamma rays, where even small inaccuracies of the system description could pose a large hindrance to reconstruction fidelity.

Shadow lengthening is again examined in Fig. 3.7 by instead increasing the source zenith angle to 50° with a slat thickness of only 1 cm. Two features of the Monte Carlo profile are observed here: (1) the asymmetry in the envelope of high frequency modulations between the two low-frequency peaks and (2) the decreased profile contrast (i.e., a minimum count rate which does not go all the way to zero) due to transmission through the mask. The standard formula is unable to describe either of these features, and so would be unsuitable for analysis with an instrument that has a large field of view.

We next examine the results of altering the mask geometry by removing the $a = b = c$ constraint. Since the standard formula utilizes only a single variable to account for these three mask parameters, a direct comparison is not possible. Instead, only the Monte Carlo and advanced profiles are presented. The profiles for an RM with mask geometry providing increased transmission is shown in Fig. 3.8. The slat spacing is increased to $b = 12$ cm, while

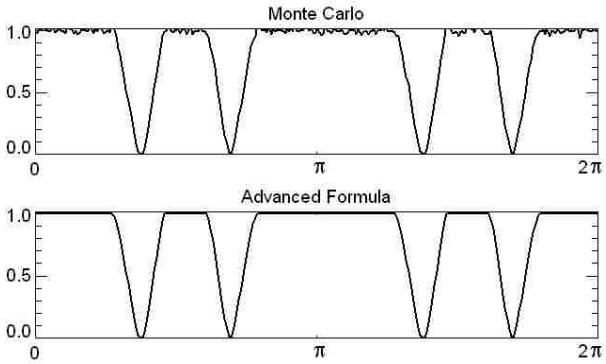


Figure 3.8: Spacing between mask slats = 12 cm, while slat and detector widths = 4 cm. The increased transmission is well represented by the advanced formula, as the profile intensity saturates several times during the rotational period.

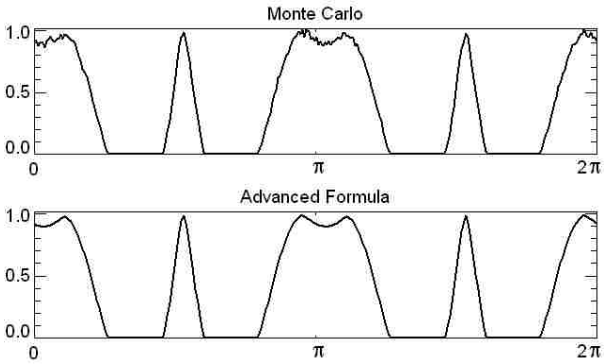


Figure 3.9: Slat widths = 12 cm, with slat spacing and detector diameters = 4 cm. The increased shadowing is seen as intervals of zero-intensity transmission.

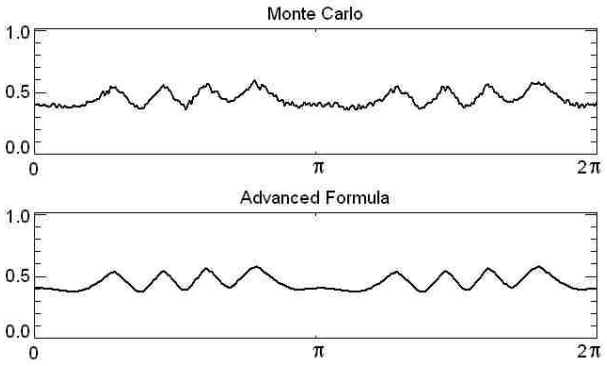


Figure 3.10: For slat width and spacing = 4 cm and detector diameter = 12 cm, up to two bars may simultaneously shadow the detector.

the slat and detector widths remain at 4 cm. The advanced profile describes the expected saturation of the transmission intensity several times during the rotational period.

Conversely, increased shadowing is shown in Fig. 3.9 by setting the slat widths to $a = 12$ cm, while the slat spacing and detector diameters are 4 cm. Again, the advanced formula describes the "dead time" of the profile, with several timespans of zero intensity transmission. Finally, the ability of the advanced formula to describe the simultaneous shadowing of multiple slats is shown in Fig. 3.10. Here, the slat width and spacing is 4 cm, and detector diameter is increased to 12 cm. While the ratio of these dimensions is again extreme for purposes of visual demonstration, the ability to individually define these variables is key to optimization of the instrument geometry, as described in Sec. 3.3.4, particularly when high energy sensitivity and large FOV are desired.

3.3.3 Image Reconstruction

The reconstructed images of Monte Carlo data based on both the standard and advanced characteristic count rate formulae are compared. For data deconvolution, the Noise-Compensating Algebraic Reconstruction (described in Sec. 4.6) is used.

First, if a 122 keV point source at small zenith angle is imaged by an RM with mask thickness 2 cm, providing 100% attenuation, both the standard and advanced formulae produce accurate image reconstructions (Fig. 3.11). The slightly extended reconstructions

reflect the locational uncertainty of the source due to noise in the data, as explained in Sec. 4.6.

Next, a 5 mm mask is used to image a 662 keV point source in the same location, allowing 55% transmission of photons incident on the slats (Fig. 3.12). The reduced contrast is properly accounted for in the advanced formula. In the image based on the standard formula, however, the increased transmission is reconstructed as spurious peaks in the image.

Finally, the 662 keV source is moved to a large zenith angle, $\theta \approx 48°$, and imaged by an RM with a 2 cm thick mask (Fig. 3.13). At such a large angle of incidence, a large fraction of photons clip the slats, and so this effect must be accurately represented in the instrument response formula. The effect of shadow lengthening is poorly described by the standard formula, and so the respective reconstruction is misrepresented and mislocated. The advanced formula, however, provides an accurate reconstruction of the source with high fidelity.

In addition to the subjective assessment of the reconstructions described, a quantitative result may also demonstrate the superiority of the advanced formula. An objective "figure of merit" should be defined based on the task of the imaging; in this case, we endeavor to demonstrate a higher fidelity reconstruction with the advanced formula over the standard

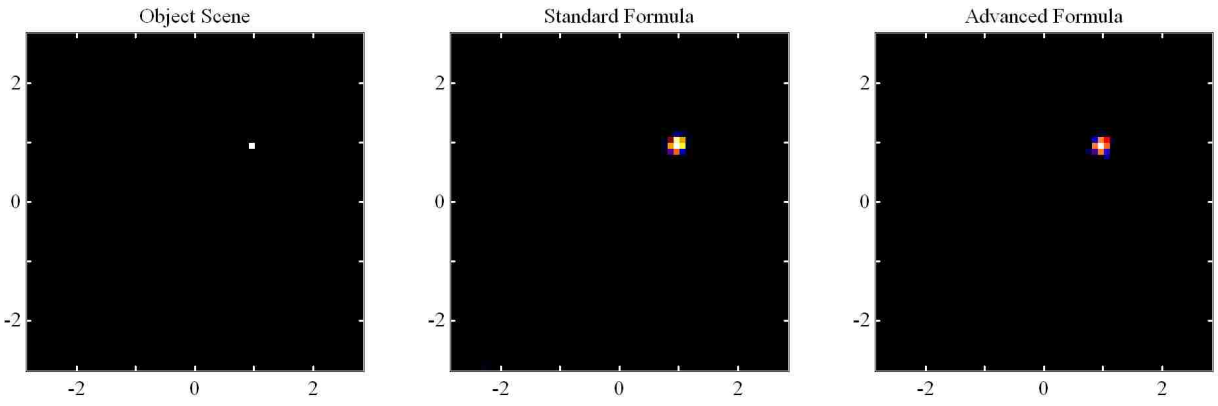


Figure 3.11: Monte Carlo simulation: A 122 keV point source is imaged by an RM with mask thickness 2 cm. The standard and advanced formula reconstructions both accurately depict the object scene. (Axes in degrees.)

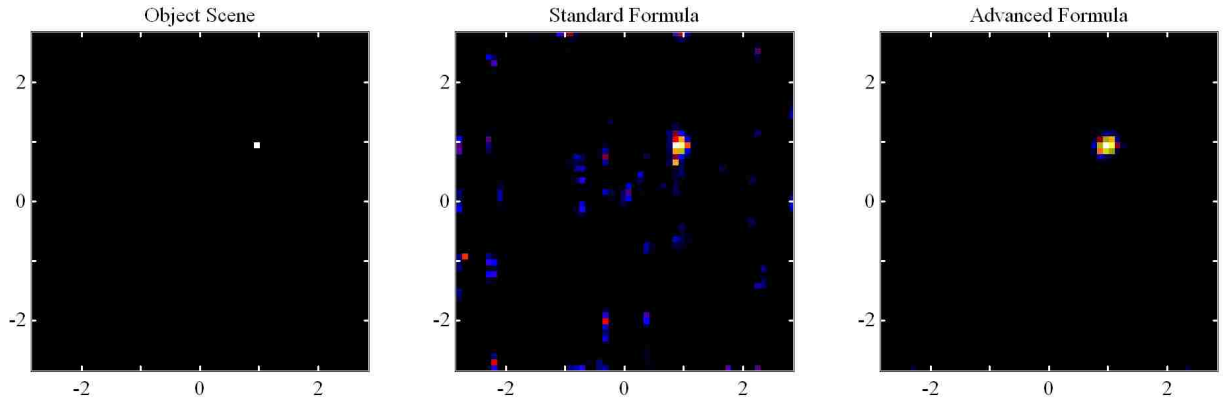


Figure 3.12: Monte Carlo simulation: A 5 mm-thick mask is used to reconstruct a 662 keV point source. Due to increased transmission through the mask, the standard formula reconstructs additional noise in the object field that is not present in the advanced formula reconstruction. (Axes in degrees.)

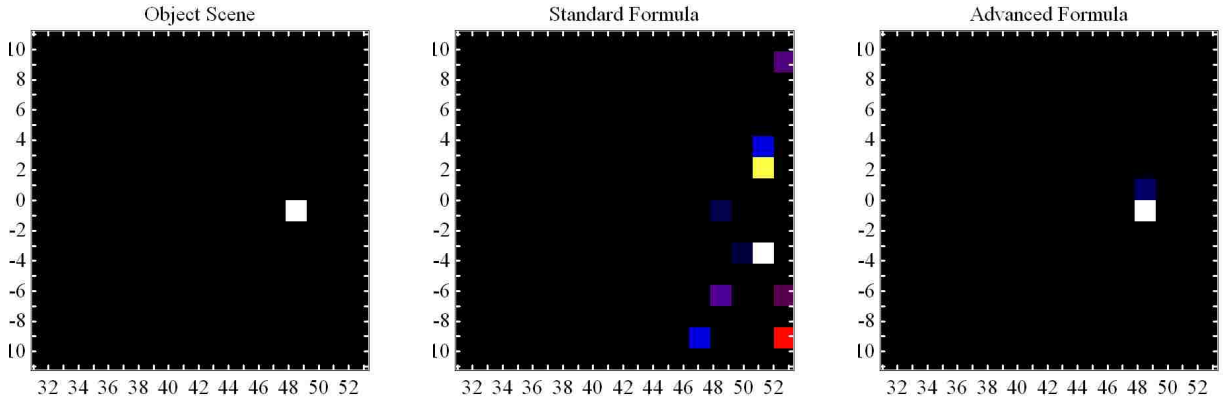


Figure 3.13: Monte Carlo simulation: A subsection (zoomed-in view) of a full reconstruction for an RM with mask 2 cm, and a point source located at $\theta \approx 48°$. At this large zenith angle, the standard formula is incapable of correctly resolving the source. Multiple spurious peaks are visible. Due to the wide FOV, large sky elements ($\sim 1.5°$) are used to reduce the size of the image matrix. (Axes in degrees.)

formula, and so the mean square error (MSE; Barrett & Myers 2003) is used. MSE is the average of the square of the deviations of the final reconstruction $F(n)$ from the original object scene $S(n)$:

$$\text{MSE} = \left\langle \, |F(n) - S(n)|^2 \, \right\rangle, \tag{3.23}$$

Table 3.1: Mean square error (Eq. 3.23) of standard and advanced reconstructions for Figs. 3.11 - 3.13. The smaller MSE for each scenario is in bold face, demonstrating the improved performance of the advanced formula.

Fig.	Mean Square Error	
	Standard Recon.	Advanced Recon.
3.11	225.4	**203.6**
3.12	279.6	**241.1**
3.13	200.7	**2.1**

where $\langle X \rangle$ denotes the average of the distribution X. A lower MSE value indicates a smaller deviation from the object scene, and thus a higher fidelity reconstruction.

Table 3.1 presents the MSE for the reconstructions in Figs. 3.11 - 3.13. In every scenario, the MSE is smaller for the advanced formula. In particular, for the source image with large zenith angle, the MSE of the advanced reconstruction is two orders of magnitude smaller than its standard counterpart, indicating the importance of this new characteristic profile for the wide-FOV regime.

3.3.4 Discussion

We have demonstrated that the advanced formula performs better than the standard instrument response by more accurately reconstructing the object scene. For a typical RM, where $a = b = c$, this advantage is most apparent for high photon energies (requiring a thick mask) and for large angles of incidence. In these cases, slat clipping and shadow lengthening have a significant effect on the profile shape, which must be accounted for in image reconstruction.

In the standard formula, the shadow opacity is uniform and its width assumed to remain unchanged despite the source zenith. Based on this assumption, the optimal values are given by a and b equal c. This equality results in the highest possible sensitivity of the instrument by ensuring that two conditions are met: (1) the profile has maximum contrast (varies from 0 to 100%), and (2) there is zero "dead time" (time intervals with no modulation). However, for a realistic finite slat thickness, $a = b = c$ results in dead time due to shadow lengthening,

and so to maximize sensitivity the spacing b should be increased. The standard formula does not provide a means of correcting for these effects, but the advanced formula allows for a detailed analysis of the sensitivity and angular resolution properties of any RM geometry. In Sec. 5.4, this opportunity is exploited to examine the optimal slat width and pitch for a typical RM design.

Chapter 4

Image Reconstruction

4.1 Imaging Problem

The measurement by any indirect linear imaging system may be described by an observational vector $O(m)$ in measurement space m. This vector does not constitute a direct representation of the object scene vector $S(n)$ (where n corresponds to a single element in the field of view) but rather the result of a correlation of the instrument response $P(m, n)$ with the object scene; i.e.,

$$O(m) = \sum_n P(m, n)S(n) + B(m) + N(m). \tag{4.1}$$

Included in this measurement is a background contribution $B(m)$ and noise $N(m)$ arising from statistical uncertainties in both the source and background rates and systematic effects (e.g., inaccurate description of the instrument response). To simplify the analysis, we make two assumptions: (1) the observation vector is digitized with sufficient precision and sampled above the Nyquist frequency of the region of interest in the object scene, and (2) the background contribution may be mostly removed by one of three techniques described in Sec. 4.2.

A decoding matrix $G(m, n')$ is defined to convert the observation vector back to image

space. Thus, a cross-correlation image is computed,

$$\begin{aligned} C(n') &= \sum_m G(m,n')O(m) && (4.2)\\ &\approx \sum_m G(m,n')\left(\sum_n P(m,n)S(n)\right) && (4.3)\\ &\approx \sum_n P'(n,n')S(n), && (4.4) \end{aligned}$$

where the approximation symbol ($\approx$) is used to denote that the background and noise terms have been dropped. A system matrix has been defined to provide a direct relationship between the object scene and cross-correlation image,

$$P'(n,n') = \sum_m G(n',m)P(m,n). \tag{4.5}$$

The challenge of image reconstruction is to solve the inverse problem of Eq. 4.4 for $S(n)$ from $C(n')$. A successful solution to this problem is dependent on the fundamental nature of the instrument response $P(m,n)$, an accurate definition of its structure and that of the decoding matrix $G(n',m)$, and the suppression of noise in the observation vector. In Sec. 4.5, we examine several possible solutions.

4.2 Background Removal

The background contribution in Eq. 4.1 must be removed prior to solving the system of equations defined in Eq. 4.4. Three methods may accomplish this task: (1) subtraction of an experimental measurement of its value, (2) use of an anti-mask, and (3) use of a proper decoding matrix for a constant B. The first of these solutions requires a "background run," or exposure time during which no sources are in the field of view. Such an exposure represents the contributions from external and internal backgrounds, and over sufficient time, should sufficiently approximate $B(m)$ for any subsequent source exposure. This method,

however, requires additional time to perform (eating into valuable observation time) and is more difficult in satellite and balloon-borne applications where sources can not be easily "removed" for such an exposure.

Anti-masks are typically used in coded aperture systems to remove artifacts from the reconstruction which arise from background nonuniformities in the detection plane, though the concept is applicable to any multiplexing instrument. If the background is not removed through a background exposure, then Eq. 4.3 is written as

$$C(n') = \sum_n \left(\sum_m G(m,n')P(m,n) \right) S(n) + \sum_m G(m,n')B(m), \tag{4.6}$$

The decoding matrix, $G(m,n')$ is chosen such that its average is zero, e.g.

$$G(m,n') = P(m,n) - \overline{P(n)}, \tag{4.7}$$

where $\overline{P(n)}$ is the instrument response at scene element n averaged over all detection elements m. For a typical multiplexing instrument for which $P(m,n)$ varies uniformly between zero and one, it may be rewritten as

$$G(m,n) = 2\left(P(m,n) - \frac{1}{2}\right). \tag{4.8}$$

An anti-mask is an exact opposite of the mask, with open elements switched with closed elements and vice versa. The instrument response for the anti-mask, $\widetilde{P}(m,n)$ is thus

$$\widetilde{P}(m,n) = 1 - P(m,n). \tag{4.9}$$

By combing Eqs. 4.8 and 4.9, the decoding matrix for the anti-mask is found to be the

negative of that of the mask:

$$\begin{aligned}\widetilde{G}(m,n) &= 2\left(\widetilde{P}(m,n)-\frac{1}{2}\right)\\ &= -2\left(P(m,n)-\frac{1}{2}\right)\\ &= -G(m,n). \end{aligned} \tag{4.10}$$

The combination of two exposures, one with the mask and one with the anti-mask, effectively remove the background term:

$$\begin{aligned}C(n')+\widetilde{C}(n') &= \sum_n\left(\sum_m G(m,n')P(m,n)+\widetilde{G}(m,n')\widetilde{P}(m,n)\right)S(n)\\ &\quad +\sum_m\left(G(m,n')+\widetilde{G}(m,n')\right)B(m)\\ &= \sum_n\left(\sum_m G(m,n')P(m,n)+\widetilde{G}(m,n')\widetilde{P}(m,n)\right)S(n). \end{aligned} \tag{4.11}$$

If the background term in Eq. 4.3 is constant across the spatial dimension m, then neither a background-subtraction nor anti-mask is required. Instead, the selection of a decoding matrix with average zero as defined in Eq. 4.7 is sufficient to accomplish this task. The cross-correlation image is given as

$$C(n')=\sum_n\left(\sum_m G(m,n')P(m,n)\right)S(n)+B\sum_m G(m,n'), \tag{4.12}$$

where it is easy to see that the second summation is equal to zero. In a multiplexing instrument such as the coded aperture, which uses spatial modulation to encode the object scene, background contributions across the detection plane are typically non-uniform, and so this technique may not be used. In a temporal-modulating imager which folds the exposure over short finite periods, such as the RM or RMC, the background may be approximately constant across the time elements of the profiles, and thus no background exposure (or anti-mask) is necessary.

4.3 Imaging System Analysis

The ability of an imaging device to achieve high fidelity is governed largely by the nature of its system response. The ability to well-represent and reconstruct the observed object is evaluated by consideration of the systems' impulse response, or point spread function (PSF). The PSF, $h(n';n_0)$ is the reaction of the instrument to a Kronecker delta function in the object scene located at n_0:

$$h(n';n_0) = \sum_n P'(n,n')\delta(n;n_0). \tag{4.13}$$

The PSF is characterized by several properties: support, linearity, and shift. The support of $h(n')$ has a large bearing on the subsequent analysis required, particularly on the ability to accurately reconstruct crowded fields and elongated structures. Compact support is desired, whereby nonzero values of $h(n')$ are confined to a finite number of elements. Infinite support implies that the PSF does not vanish even far away from the centroid of the point source. In this latter case, source confusion may present a significant challenge in accurate deconvolution. The ideal PSF is a delta function, in which case it is easy to see from Eq. 4.13 that the instrument response would be described by the identity matrix.

Linearity of the system ensures multiple observations obey superposition. As written in Eq. 4.1, this condition is always satisfied. Similarly, application of the decoding matrix also obeys linearity in the calculation of $C(n')$. Linearity of a final reconstructed image, however, is dependent on the deconvolution technique used for solving Eq. 4.4. Of the methods described in the sections below, this property is rarely satisfied.

Shift-invariance ensures that the PSF remains unchanged, only shifted, for a translation of a point source in the field of view. That is, the PSF must obey the relation that

$$h(n'-n_0;0) = h(n';n_0). \tag{4.14}$$

In this case, the matrix $P'(n, n')$ can be completely described by the vector $P'(n)$, and the image defined by

$$C(n) = P'(n) * S(n), \tag{4.15}$$

where $*$ is the convolution operator:

$$P'(n) * S(n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P'(n) S(n - n') dn'. \tag{4.16}$$

While the PSF describes the system's spatial response, the transfer function, $H(\eta) = \mathcal{F}\{P'(n)\}$, provides an analysis of the frequency response of the system. By taking the Fourier transform of both sides of Eq. 4.15, the convolution theorem (see Sec. 4.5.2) provides the relation

$$H(\eta) = \frac{\mathcal{F}\{C(n)\}}{\mathcal{F}\{S(n)\}}. \tag{4.17}$$

Thus, the transfer function is the ratio of the frequency components in the raw image to those in the actual object scene; i.e., it is a measure of how well the imaging system preserves spatial frequencies.

4.4 RM Imaging Properties

As described previously, the background-subtracted observation vector for a rotating modulator is a time history of counts for each detector d, described by

$$O_d(t) = \sum_n P_d(n, t) S(n), \tag{4.18}$$

where the sky element n may be given in terms of cartesian (x, y) or polar (θ, ϕ) coordinates. Each detector observation can thus be transformed to image space, producing a detector-

specific cross-correlation image:

$$\begin{aligned} C_d(n') &= \sum_t \left(P_d(t,n') - \overline{P_d}(n')\right) O_d(t) \\ &= \sum_n P'_d(n,n') S(n), \end{aligned} \tag{4.19}$$

where the decoding matrix has been selected as the mean-subtracted transpose of the instrument response for detector d, as described in Sec. 4.2. The instrument's cross-correlation image is found by summing these individual images over all detectors:

$$C(n') = \sum_d C_d(n') \tag{4.20}$$

$$\begin{aligned} &= \sum_d \left(\sum_n P'_d(n,n') S(n) \right) \\ &= \sum_n \left(\sum_d P'_d(n,n') \right) S(n) \\ &= \sum_n P'(n,n') S(n) \end{aligned} \tag{4.21}$$

Alternatively, a cross-correlation image could be generated for the entire instrument by first concatenating all detector profiles to produce a single observation vector. However, such a deconvolution problem requires a large array, making this technique slow. Instead, the observations are transformed to image space and summed there. Mathematically, the approaches are equivalent.

Eq. 4.21 implies the existence of a single system matrix for the RM given by

$$P'(n,n') = \sum_{d,t} \left(P_d^T(t,n') - \overline{P_d}(t,n')\right) P_d(n,t), \tag{4.22}$$

where each element of $P_d(n,t)$ is provided by the characteristic formulae presented in Chapter 3. Conceptually, an element of $P'(n',n)$ may be thought of as the relative brightness of pixel n' for a point source located at location n. The PSF appears as a central concentric peak

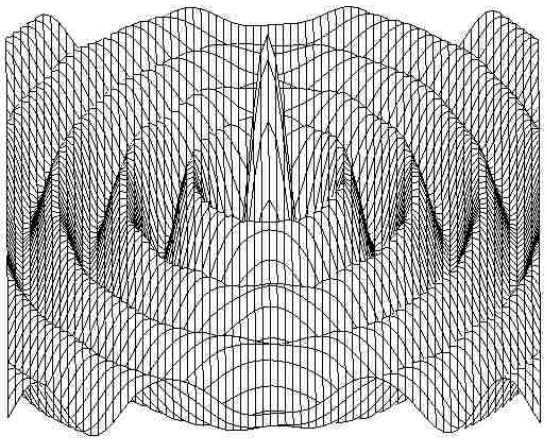


Figure 4.1: RM Cross-correlation image surface mesh of a single point source centered in the FOV. The concentric "ring" nature of the PSF is apparent.

with "ringed" sidelobes (Fig. 4.1). This distribution may be described analytically by a zeroth order Bessel function of the first kind (Mertz, 1976). Consequently, the PSF is of infinite support, which complicates image deconvolution.

4.5 Imaging Solutions

4.5.1 Direct Inversion

One solution to deconvolve the cross-correlation image is to multiply both sides of Eq. 4.4 by the inverse of $P'(n, n')$. Equivalently, this could be done with the observation vector, by selecting the decoding matrix such that $G(n', m) = P^{-1}(m, n)$. In this case,

$$C(n') \approx \sum_m P^{-1}(n', m)P(m, n)S(n) \tag{4.23}$$

$$\approx S(n) \tag{4.24}$$

This solution will only hold, however, if $P^{-1}(n', m)$ exists. Thus, $P(m, n)$ must be square and non-singular (its determinant must be non-zero), which in general is not true. While it is straightforward to fabricate $P(m, n)$ to be square by ensuring the number of measurement elements equals those of the object space, non-singularity of the system matrix is determined by the system geometry and method of encoding.

If $P(m,n)$ is non-singular, then a solution may still be difficult to find. Iterative techniques (e.g. Gaussian elimination) suffer from compounded rounding errors particularly for large matrices. As noted by Li & Wu (1994), the mathematical solutions to these linear systems will further deviate due to measurement errors and noise. Though infrequently so, if $P(m,n)$ is small and non-singular, then direct inversion offers a fast and simple method for deconvolving observed data.

4.5.2 Convolution Theorem

A similarly simple deconvolution technique may be employed for imaging systems which are linear and shift-invariant. The convolution theorem states that the convolution of two functions in the spatial domain is equivalent to the dot product of those functions in the Fourier domain (and vice versa):

$$\mathcal{F}\{C(n)\} = \mathcal{F}\{P'(n) * S(n)\} = \mathcal{F}\{P'(n)\} \cdot \mathcal{F}\{S(n)\}, \tag{4.25}$$

where $\mathcal{F}\{f(x)\}$ represents the Fourier transform of the function $f(x)$,

$$\mathcal{F}\{f(x)\}(\mu) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i x \mu} dx. \tag{4.26}$$

S may thus be solved for analytically,

$$S(n) = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{C(n)\}}{\mathcal{F}\{P'(n)\}}\right\} = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{C(n)\}}{H(\eta)}\right\}. \tag{4.27}$$

If the imaging system has small or zero response to a particular frequency component (i.e., there exists one or more η where $H(\eta) = 0$), then singularities arise in Eq. 4.27, making this technique impractical. If, however, the instrument responds well to all observable frequencies with a relatively flat transfer function, this technique, like direct inversion, provides a quick solution to the deconvolution problem.

4.5.3 Iterative Techniques

Eq. 4.1 represents a perfectly suitable system of equations to solve directly for the object scene. Yet, by employing a decoding matrix to generate a cross-correlation image as in Eq. 4.4, we are able to more easily examine the nature of the instrument response by placing both the object scene and measurements in the same domain. Additionally, this procedure allows some flexibility in the design of the matrix $P'(n, n')$: the decoding matrix $G(m, n')$ may be chosen to satisfy desirable conditions of $P'(n, n')$. For example, as shown in Sec. 4.2, a constant background may be removed without its measurement by careful choosing $G(m, n')$. For iterative techniques, it is useful to select the decoding matrix such that

$$G(m, n') = P(m, n'). \tag{4.28}$$

Thus, the system matrix is then

$$P'(n, n') = \sum_m P(m, n')P(m, n), \tag{4.29}$$

which, by definition, is positive and square-symmetric. While this definition can not take advantage of the auto-removal of constant background, these properties make Eq. 4.4 a more suitable system for a wide variety of iterative deconvolution techniques, which have been developed across a broad spectrum of applications including optical deblurring, radio interferometry, and medical tomography. Here, we note a few of the more common iterative algorithms which we have applied to the RM imaging technique.

CLEAN

Introduced by Högbom (1974), the CLEAN technique was originally developed for use in radio interferometry, and so named for its operation on the "dirty map," analogous to the cross-correlation image. This ad hoc technique iteratively seeks the highest value in the raw

image, and subtracts from it a fraction of the weighted PSF corresponding to that element's location and brightness. This "source" is added to a new "cleaned" map, which ultimately serves as the final reconstructed image. This procedure is iterated using the residual dirty map until some lower background threshold limit is reached.

CLEAN performs very well for sparse object scenes, and can even represent extended sources by breaking them down into a finite number of point-like sources. Additionally, the simplicity of the CLEAN operation makes the technique computationally fast. The fidelity of the reconstructed scene is, however, largely dependent on the support of the PSF. PSFs that are not compact present a significant problem in the form of source confusion.

Maximum Entropy Method

With its roots in information theory (Shannon, 1948), the Maximum Entropy Method (MEM; Jaynes 1957; Frieden 1972; Cornwell & Evans 1985) is a technique for solving the ill-posed deconvolution problem with missing data and noise. It asserts that the the best choice for the probability distribution of the reconstructed image is that which has maximum entropy while simultaneously achieving the best fit to the original data.

The "entropy" of an image is defined as the natural logarithm of its multiplicity, or total number of ways in which it could have been formed. For an image $F(n)$ with N pixels and M total counts, the entropy, S, is defined as

$$S = \ln\left(\frac{M!}{\prod_n^N F(n)!}\right). \tag{4.30}$$

The chi-squared goodness of fit of the newest iteration of the reconstruction, $F(n)$, is given by

$$\chi^2 = \sum_{n'}^{N} \frac{\left[\sum_n^N P'(n,n')F(n) - C(n')\right]^2}{C(n')}. \tag{4.31}$$

These constraints can be implemented as a Langrange multiplier problem, to iteratively find the best distribution which accomplishes these two goals simultaneously. The probability

distribution is then found iteratively by

$$F(n) = \frac{\exp\left(-\sum_{n'}^{N} \lambda(n') P(n,n')\right)}{\sum_{n}^{N} \exp\left(-\sum_{n'}^{N} \lambda(n') P(n,n')\right)}, \tag{4.32}$$

where the multipliers are determined for iteration $(k+1)$ by

$$\lambda^{(k+1)}(n') = \lambda^{(k)}(n') + \ln C(n') - \ln\left(\sum_{n}^{N} F(n) P(n,n')\right) \tag{4.33}$$

Maximum Likelihood Expectation Maximization

The likelihood $L(F)$ of a reconstructed image distribution F is equivalent to the probability of the data given the image, $P(O|F)$, i.e.

$$L(F) = P(O|F). \tag{4.34}$$

The Maximum Likelihood Expectation Maximization (MLEM; Dempster et al. 1977; Moon 1996) algorithm finds the solution to the deconvolution problem which maximizes this parameter.

Most commonly used in medical imaging modalities, MLEM assumes the observation vector $O(m)$ is Poisson-distributed, such that the uncertainty of each element is given by the square root of its value. By writing out the Poisson distribution function, an iterative solution for maximizing the likelihood can be found. The $(k+1)$ iteration is given by

$$F(n)^{(k+1)} = \frac{F(n)^{(k)}}{\sum_m P(m,n)} \sum_m \left[\frac{O(m)}{\sum_{n'} P(m,n') F(n')^{(k)}} P(m,n)\right]. \tag{4.35}$$

Algebraic Solutions

While statistical and ad-hoc techniques boast speed and the ability to perform well with incomplete and noisy data, exact agreement with the data can be achieved through the

use of algebraic techniques to solve the deconvolution problem. One technique devised by Kaczmarz (1937) is shown to converge for non-singular matrices (Barrett & Myers, 2003). This technique was reintroduced for medical imaging modalities by Gordon et al. (1970) as the Algebraic Reconstruction Technique (ART).

Other, more common solutions for solving the systems of linear equations include the Jacobi and Gauss-Seidel methods (Barrett et al., 1994). These techniques iteratively remove sidelobe contributions from each pixel caused by sources elsewhere in the field. The Jacobi iteration determines the contributions based on the last iteration of the reconstructed image:

$$F_i^{(k+1)} = \frac{1}{P'(n,n)} \left(C(n) - \sum_{n' \neq n} P'(n,n') F(n')^{(k)} \right). \tag{4.36}$$

Gauss-Seidel differs from Jacobi in that the updated elements in the current iteration are used in the calculation:

$$F_i^{(k+1)} = \frac{1}{P'(n,n)} \left(C(n) - \sum_{n'>n} P'(n,n') F(n)^{(k)} - \sum_{n'<n} P'(n,n') F(n)^{(k+1)} \right). \tag{4.37}$$

Algebraic techniques enforce exact agreement with the observation vector. Consequently, in the case of noisy or incomplete data, these solutions may produce spurious artifacts and reduced fidelity. Smoothing and regularization are typically employed to compensate for this shortcoming. The Direct Demodulation (DDM) technique (Li & Wu, 1994), for example, uses Jacobi or Gauss-Seidel iterations with a positivity constraint enforced as a physical parameter to dampen these oscillations. In Sec. 4.6, we present a novel algebraic solution based on DDM developed specifically for the RM.

Table 4.1: Characteristics of several reconstruction techniques as applied to the RM.

Technique	Feasible	Noise Suppression	Computational Speed	Super Resolution
Direct Inversion	No	—	Fast	—
Fourier Inversion	No	—	Fast	—
CLEAN	Yes	Yes	Fast	No
MEM	Yes	Yes	Fast	No
MLEM	Yes	Yes	Slow	Yes
Algebraic	Yes	No	Moderate	Yes
NCAR	Yes	Yes	Moderate	Yes

4.6 Noise-Compensating Algebraic Reconstruction

4.6.1 Introduction

The deconvolution techniques described in 4.5 were developed to solve the ill-posed reconstruction problem for a wide variety of system responses and noise conditions. For image reconstruction for an RM-based black-hole survey mission, a deconvolution technique must (1) be relatively fast, (2) remove extended sidelobes, (3) solve multiple point sources and extended distributions, (4) compensate for noise and incomplete data, and (5) achieve super-resolution.

Most of the techniques described are capable of satisfying some but not all of these requirements (see Table 4.1). Direct inversion is not feasible since the system matrix is typically large, contributing to errors in the calculation of its inverse. The non-ideal RM transfer function results in Fourier inversion being similarly not useful. CLEAN performs well for multiple point-like sources for PSFs with small finite support, but the extended RM sidelobes cause significant source confusion, and it is not able to resolve past the intrinsic resolution of the instrument. While we have previously shown imaging results using MEM (Budden et al., 2008a), it is unable to provide super-resolution or resolve complex source distributions. MLE performs very well for RM reconstruction, compensating for noise, achieving super-resolution, and providing good image fidelity. However, the assumption that the data

is Poisson-distributed means the detector data must be concatenated and not summed in image space. As mentioned in Sec. 4.4, this results in a computationally slow reconstruction.

4.6.2 Algorithm

We have developed a new technique derived from Direct Demodulation (DDM; Li & Wu 1994). In simulations, DDM has been previously shown to provide super-resolution with RMCs (Chen et al., 1998) and we have shown (Budden et al., 2009) the ability of DDM with an RM to achieve $\sim 8\times$ the geometric resolution and to resolve multiple and complex sources when little or no noise exists. For the reconstruction of actual measured data in the presence of background, we modify the DDM routine and implement a key step to compensate for noisy data. We refer to this technique as Noise-Compensating Algebraic Reconstruction (NCAR).

The Gauss-Seidel iterative method is used to solve algebraically for the object scene. The $(k+1)$ iteration of the reconstruction, $f(n)$, is determined by (Barrett et al., 1994)

$$f^{(k+1)}(n) = \frac{1}{P'(n,n)} \left(C^{(k+1)}(n) - \sum_{m<n} P'(m,n) f^{(k+1)}(m) - \sum_{m>n} P'(m,n) f^{k}(m) \right), \quad (4.38)$$

with $f^0(n) = 0$. Gauss-Seidel uses the results of calculations from the same iteration, and so to prevent pixel bias, it is prudent to randomize the order of m every iteration. A positivity condition is enforced as a physical constraint on each pixel of $f^{(k+1)}(n)$.

A normalized parameter β may be defined, which sums over all pixels in the residual to examine the agreement of the reconstruction with the cross-correlation,

$$\beta^{(k+1)} \equiv \frac{1}{\beta^{(0)}} \sum_{n'} \left| \sum_{n} P'(n',n) f^{(k+1)}(n) - C(n') \right|. \quad (4.39)$$

In Fig. 4.2a, this residual summation is plotted against the iteration number for a sample reconstruction. The successive reconstructions converge to the data (i.e. β shrinks) for an

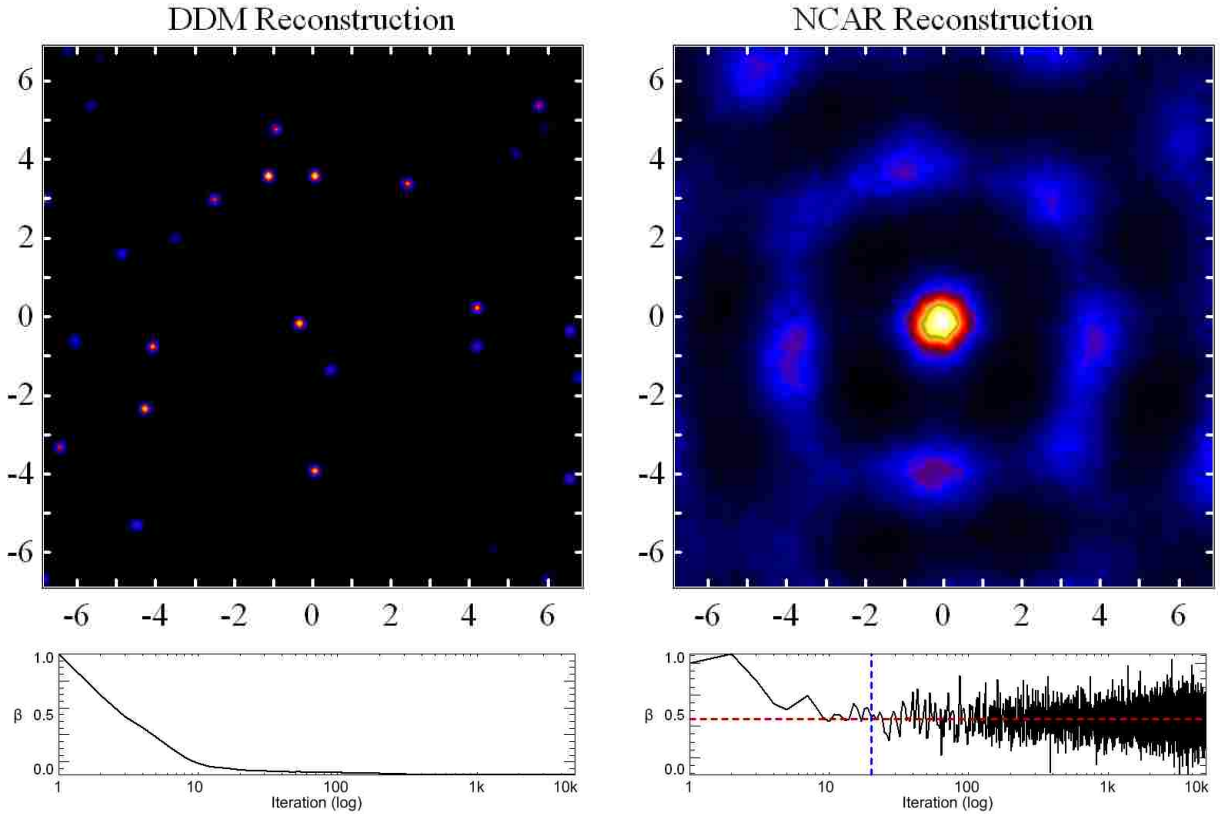


Figure 4.2: Low SNR simulation of a single point source centered in the FOV. The reconstructed image (top) and corresponding normalized residual summation, β, plotted against iteration number (bottom) are shown for (a) Direct Demodulation and (b) with the noise compensation additive, $R[\sigma(n)]$, given in Eq. 4.40. In (b), the vertical dashed line corresponds to κ, the iteration at which the maximum agreement has been reached, and the horizontal dashed line corresponds to the residual convergence value given the noisy data. Note the contrast to (a), which converges indefinitely to zero (i.e. perfect agreement).

indefinite number of iterations. This agreement, however, includes any fluctuations from noise within the data, and so this method causes spurious peaks and poor location accuracy of the true source, as seen in the reconstruction, when the SNR is low.

This "noise amplification" is the primary problem in image reconstruction using an algebraic technique. Regularization methods (e.g. smoothing, pixel-to-pixel variation constraints) may be employed, but this regularized image is still forced to agreement with the data, $C(n)$. The NCAR technique allows for deviation from the noisy data, by replacing $C(n)$ in Eq. 4.38 with $C^{(k+1)}(n)$, such that

$$C^{(k+1)}(n) \equiv C(n) + R^{(k+1)}[\sigma(n)]. \tag{4.40}$$

The function $R[\sigma]$ provides a random number from a normal distribution with average zero and standard deviation σ, where $\sigma(n)$ is the error in each pixel of the cross-correlation. If the error is assumed to be only from Poisson uncertainty, it can be shown for any multiplexing imager that

$$\sigma(n) = \sqrt{\sum_m O(m)P(m,n)^2}. \tag{4.41}$$

In the case of the RM, where multiple observational vectors are used to construct the cross-correlation image, the uncertainty is

$$\sigma(n) = \sqrt{\sum_d \sum_t O_d(t)P_d(t,n)^2}. \tag{4.42}$$

The reconstruction and residual summation, β, are plotted for NCAR in Fig. 4.2b. The agreement between reconstruction and data fluctuates about an envelope showing initial convergence followed by a leveling off beginning at some iteration κ. Once this minimum β is reached, the image has converged as well as possible to the data given the noise. Each successive iteration still contains spurious peaks arising from noise fluctuations, but because the component $R[\sigma(n)]$ has been added to randomize the noise, the amplitude and location

of these peaks are unique to each iteration, while the iterative images have the true sources in common. By now averaging over all the images after iteration κ, the spurious random peaks are suppressed while the true source distributions remain:

$$F^{(k+1)}(n) = \frac{f^{(k+1)}(n) + (k-\kappa)F^k(n)}{k+1-\kappa}, \quad k+1 > \kappa. \tag{4.43}$$

$F^{(k+1)}(n)$ is the final reconstruction shown in Fig. 4.2b. Using NCAR, statistical uncertainty in the data is manifest as an increase in the size of the reconstructed source in the form of fuzzy sidelobes. In cases where the actual size of the source is desired, NCAR may not be desirable; but with large noise in the data, an algebraic technique alone will produce spurious peaks and may not represent the true source size anyway. When it is acceptable that source size in the reconstructed image relates to uncertainty in the data, NCAR produces image reconstructions with a significant reduction in noise fluctuations compared to results using MEM, CLEAN, or algebraic solutions, as will be demonstrated in detail in Sec. 4.6.4.

4.6.3 Simulation

Custom-written Monte Carlo simulations (desribed in Sec. 6.4) are performed to test and verify the NCAR technique using software described in Sec. 6.4. The instrument response is calculated as described in Chapter 3. Since the instrument response assumes far-field imaging, sources are simulated at an "infinite" distance from the detector; i.e. photons from the same source arrive at equal angles of incidence. The geometry of the simulated RM reproduces that of the laboratory prototype, described in Chap. 6.

In Sec. 4.6.4, the NCAR technique is compared to its precursor, DDM, to demonstrate the effect of noise compensation. Additionally, we include a comparison to MEM and CLEAN, because of their wide use in astrophysical applications. For these simulations, the FOV is divided into 12×12 arcmin bins, totaling 4900 pixels for a single image. Each of the results

is that of a 30-minute simulated exposure. NCAR, DDM, and CLEAN are run for 10k iterations, and MEM for 50k – 100k to assure convergence.

4.6.4 Results

We first observe how a single point source with a measured rate of 10 cts/s (over all detectors) is imaged in the presence and absence of background. When no background exists (Fig. 4.3), all deconvolution techniques reconstruct the scene reasonably well, although spurious point sources are faintly visible in the MEM reconstruction.

When a background rate of 500 cts/s is included (Fig. 4.4), the MEM, CLEAN, and DDM reconstructions exhibit noise fluctuations that are manifest as spurious peaks in the image. The true source itself, however, remains relatively unaffected. The NCAR image shows the presence of diffuse background features, and the true source peak has broadened. Due to the nature of NCAR, the width of a point source reconstruction narrows as the SNR grows. This can be a desirable feature of an image, where a "blur" is typically considered an uncertainty on source location. As we continue to show, this blurring characteristic helps to remove spurious noise fluctuations.

In Figs. 4.5-4.6, object scenes with two sources, each with a measured rate of 50 cts/s, and a background contribution of 100 cts/s, are simulated. The background is kept relatively low here to demonstrate the intrinsic ability of NCAR to achieve super-resolution. In Fig. 4.5, the sources are separated by 3°, greater than the 1.9° geometric resolution of the instrument. NCAR produces the result that is the most free of spurious peaks and other degrading artifacts. Due to the broadened nature of the RM PSF, CLEAN in particular shows difficulty in resolving sources at this separation. In Fig. 4.6, the sources are moved to a separation of 1°, about half the intrinsic resolution of the instrument. MEM reconstructs an elongated structure, while CLEAN reconstructs only a single point source. Additional simulations confirm that MEM and CLEAN are both limited by the geometric resolution of the instrument. DDM and NCAR resolve both sources and provide the desired super-resolution.

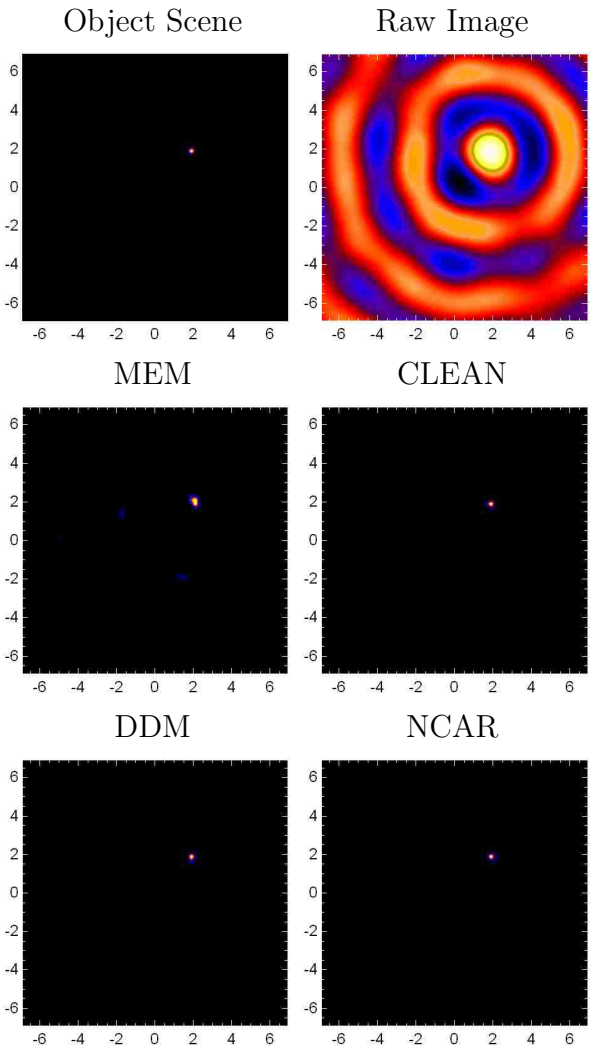


Figure 4.3: Monte Carlo Simulation results for a single point source (measured rate of 10 cts/s) and no background exposed for 30 minutes. Object scene is shown at upper left, raw cross-correlation image (Eq. 4.21) is at upper right, and results of MEM, CLEAN, DDM and NCAR are shown below. All reconstruction techniques perform reasonably well. (Axes in degrees)

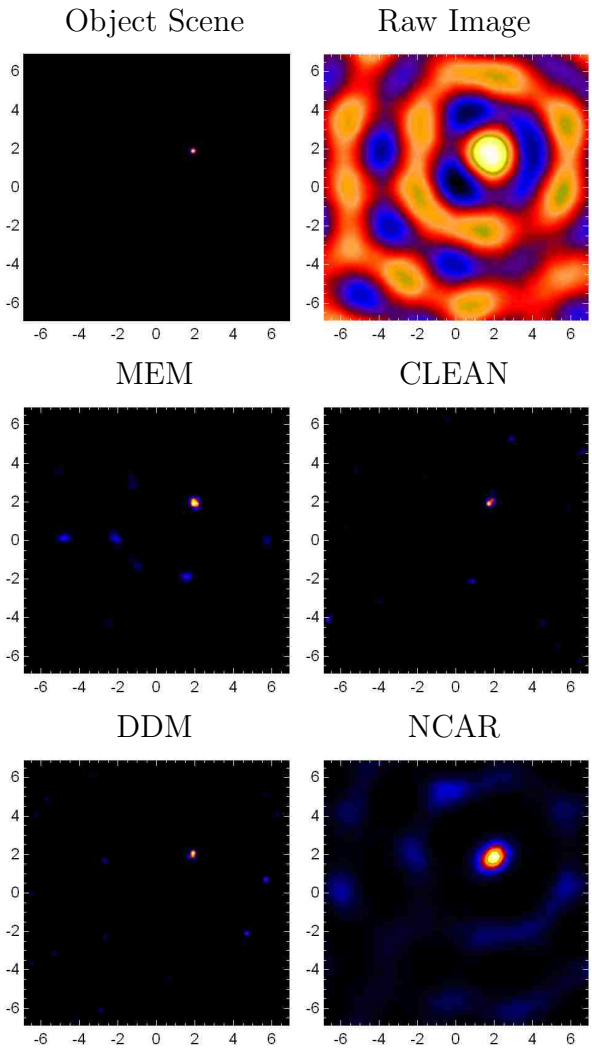


Figure 4.4: Monte Carlo Simulation results for a single point source (10 cts/s) with background (500 cts/s) exposed for 30 minutes. Note that, in the presence of background, the MEM, CLEAN, and DDM reconstructions contain spurious point-like sources and the image does not convey the uncertainty of the location of the true source. NCAR, however, smooths over the noise contributions, and places an uncertainty on the true source that is related to the SNR of the measurement. (Axes in degrees)

NCAR also provides image reconstructions for more complicated source distributions. Fig. 4.7 shows the results for 5 sources, each with a measured source rate of 50 cts/s and

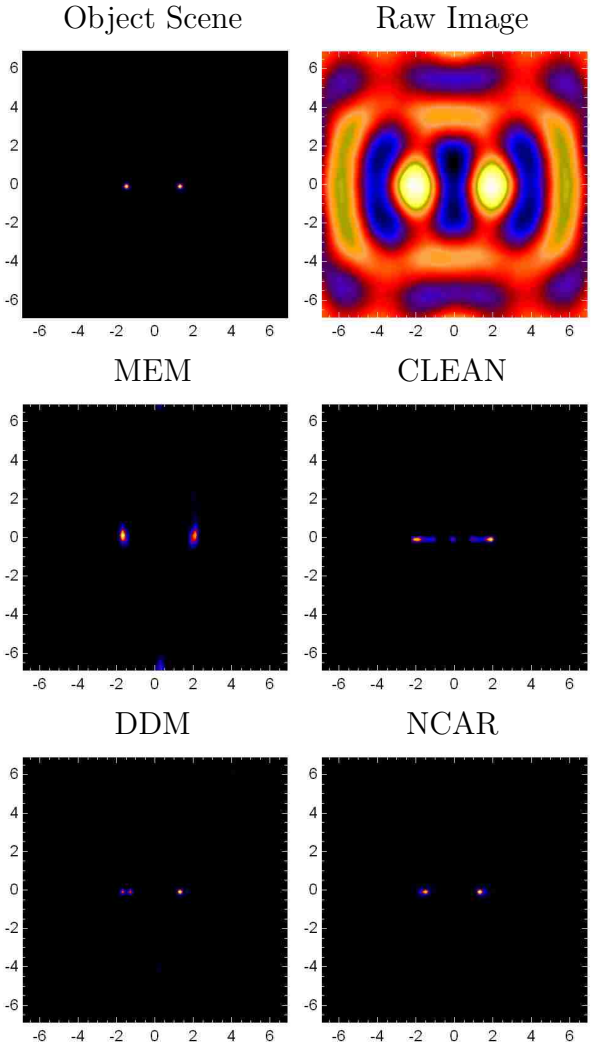


Figure 4.5: Monte Carlo Simulation results for two equal intensity sources (each 50 cts/s) separated by 3° exposed for 30 minutes in the presence of background (100 cts/s). Statistical and algebraic reconstruction techniques are both capable of resolving two sources that are at an angular separation greater than the geometric resolution defined by the instrument. (Axes in degrees)

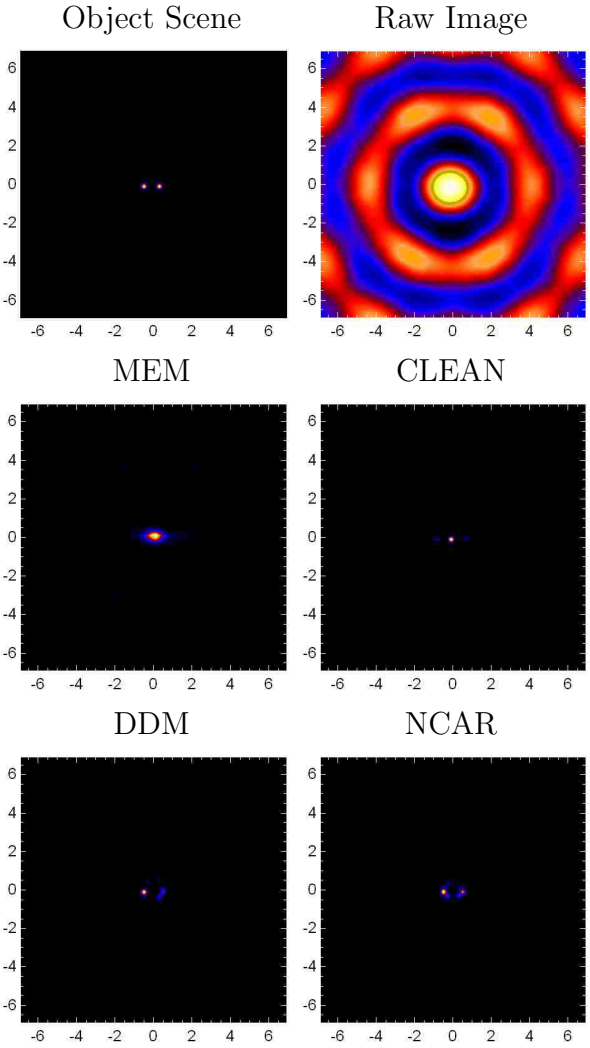


Figure 4.6: Monte Carlo Simulation results for two equal intensity sources (each 50 cts/s) separated by 1° exposed for 30 minutes in the presence of background (100 cts/s). At angular separations less than the geometric resolution of the instrument, MEM reconstructs a single elongated source and CLEAN sees only a single source, while DDM and NCAR are able to fully resolve the two sources. (Axes in degrees)

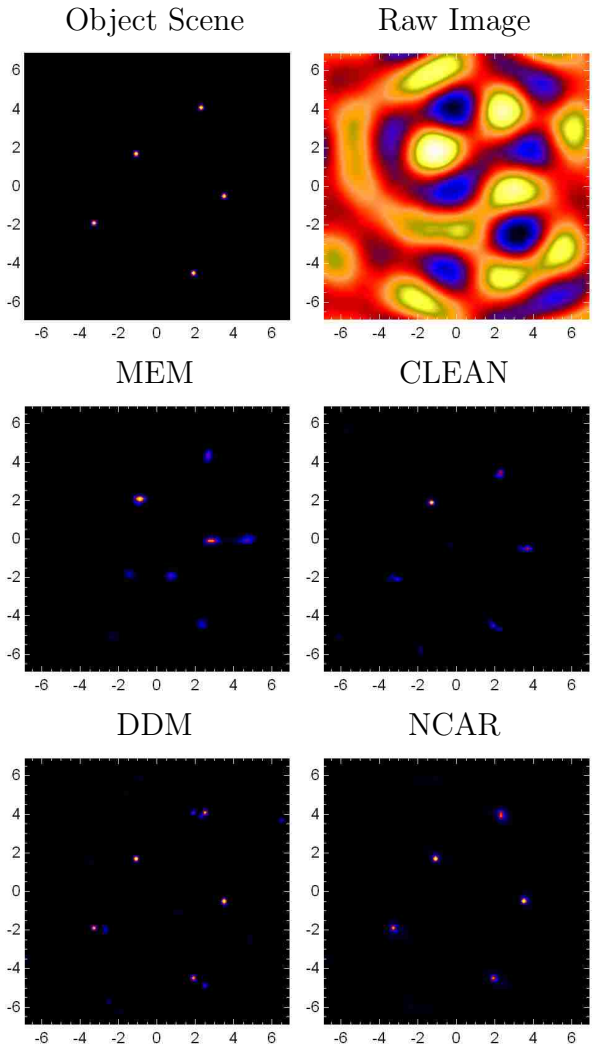


Figure 4.7: Monte Carlo Simulation results for five equal intensity sources (each 50 cts/s) in the presence of small background (50 cts/s) exposed for 30 minutes. Even with little background in the measurement, both MEM and DDM suffer from noise artifacts in the reconstruction due to the more complex object scene. NCAR, however, compensates for the noise and produces an image better representative of the object scene. (Axes in degrees)

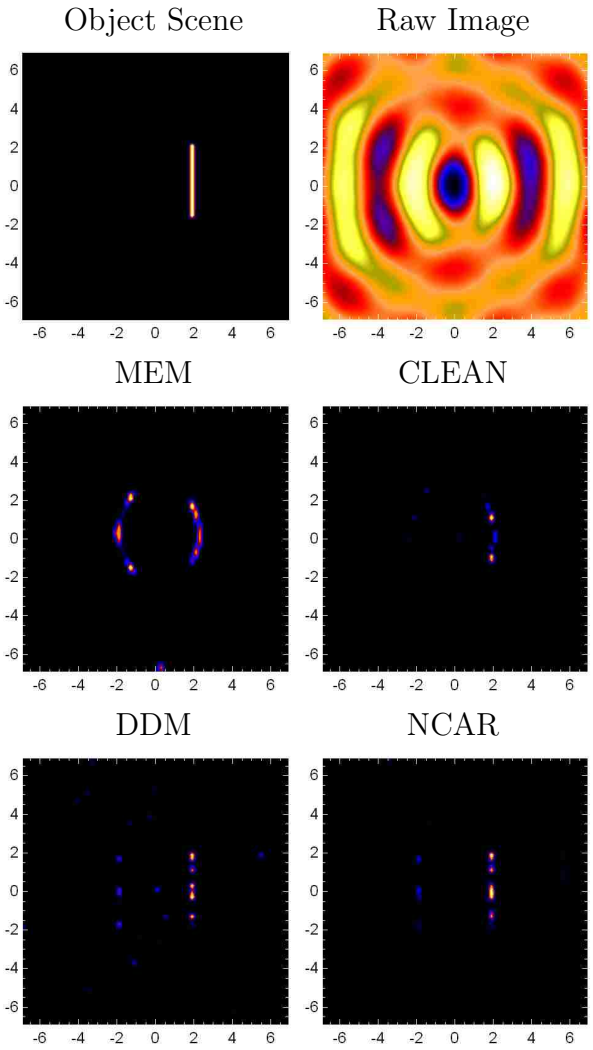


Figure 4.8: Monte Carlo Simulation results for a line source distribution (total measured rate 4.5×10^3 cts/s) exposed for 30 minutes in the absence of background. The PSF interference patterns cause MEM to perform poorly with a complex source, while DDM continues to suffer from from noise artifacts. NCAR removes most of the spurious sources. (Axes in degrees)

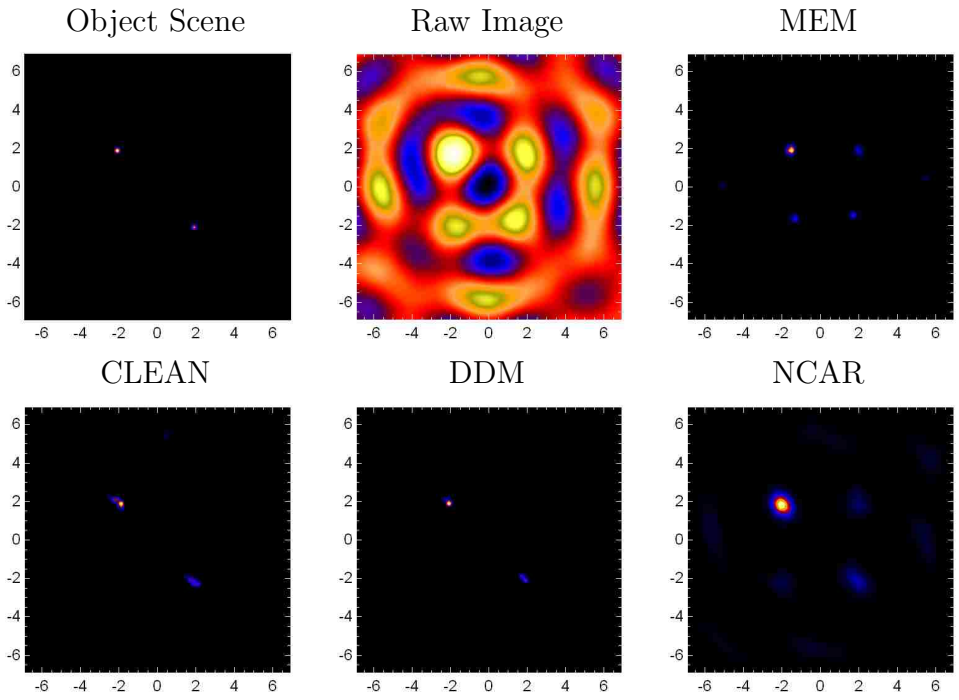


Figure 4.9: Monte Carlo Simulation results for weak source (lower right, 15 cts/s) in the presence of a stronger source (upper left, 25 cts/s) exposed for 30 minutes with significant background (500 cts/s). (Axes in degrees)

equivalent background rate. MEM is capable of resolving some, but not all of the sources entirely, while CLEAN performs slightly better. DDM suffers from spurious peaks near the true sources. NCAR, however, reconstructs each of the sources well, with no visible noise, but instead, a slight blurring. A line source distribution (Fig. 4.8) is also simulated. To examine the ability of the reconstruction techniques to resolve an extended source, the background rate is set to zero, and the line source has a total measured rate of 4.5×10^3 cts/s. MEM performs poorly with a significant spurious reconstruction to the left due to the interference of sidelobes. The CLEAN result has no significant spurious reconstructions, but the line is poorly resolved. DDM suffers from some spurious peaks arising from noise and interfering PSFs from the complex nature of the object scenes. NCAR provides a reconstruction of the line similar to DDM with most of the spurious peaks smoothed out.

As might be expected, the "broken" lines observed in the DDM and NCAR reconstruc-

tions in Fig. 4.8 are not due to sensitivity variations. As seen in Fig. 6.3, the sensitivity is relatively uniform, and the low-sensitivity sky regions are uncorrelated with the line breaks. Rather, this is a consequence of the statistical noise in the image causing the deconvolution to converge on an imperfect solution to the data. This effect is seen to increase for more complex object scenes (i.e. the number of non-trivial sky bin values increases), and is an inherent property of the RM response, not the deconvolution algorithm. The effect could be reduced by increasing the exposure time (thus improving SNR).

In order to observe two weak sources imaged simultaneously, a 25 and 15 cts/s source in the presence of 500 cts/s background are simulated in Fig. 4.9. MEM performs poorly, showing spurious reconstructions indistinguishable from the weaker of the two sources. CLEAN and DDM perform the best, with two sources clearly reconstructed with relative fluxes apparent. The "blurring" characteristic of NCAR, however, means that the weaker of the two sources is blurred to a greater degree, suppressing its apparent brightness further. For higher SNR, this discrepancy disappears.

An objective assessment of the NCAR technique is accomplished by first defining a new figure of merit for the image reconstructions. There is no singular objective value to describe image quality. Instead, the parameter defined must be based on the task of NCAR: to remove spurious fluctuations arising from noise, which are located randomly throughout the field. NCAR manifests these counts as a blurring of the true source location by populating adjacent image elements. The assertion is that this result is preferable in most cases since it allows for the identification of the true sources while removing false positives.

A standard MSE (Barrett & Myers, 2003) as described in Eq. 3.23 is not suitable to compare techniques here, since spurious fluctuations away from the true source carry the same weight as those near to the source. Instead, a weighted mean square error (wMSE) is defined:

$$\text{wMSE} = \left\langle w(n) \left| F(n) - S(n) \right|^2 \right\rangle. \tag{4.44}$$

In contrast to the standard MSE, the square of the deviation in each element n is multiplied

here by a weighting parameter $w(n)$. The weighting parameter is determined by the minimum distance between element n and a true source. If $\vec{N}$ describes the shortest vector between a true source in the field and element n, then

$$w(n) = 1 + |\vec{N}|. \tag{4.45}$$

Elements in which there is a true source have minimum weight (1), while those far from a true source have greater weight. A smaller wMSE thus indicates both higher fidelity and a consolidation of populated elements about the true source.

Table 4.2 presents the results of wMSE calculated for each of the image reconstructions by the various techniques in Figs. 4.3 - 4.9. For most of the object scene distributions, NCAR provides the best result (lowest wMSE) as expected. For the scenario in Fig. 4.4 (single point source with significant background), both MEM and DDM reconstructions provide superior results (lower wMSE) despite the spurious fluctuations apparent in the field. Here, the inherently low SNR causes significant diffuse background to be present in the NCAR reconstruction, which increases the wMSE. As shown in Fig. 4.2, for even lower SNR, noise manifestations in DDM dominate the true source representation. The diffuse background inherent in the NCAR reconstruction, however, may still cause a higher wMSE value despite

Table 4.2: Weighted mean square error (Eq. 4.44) for reconstructions shown in Figs. 4.3 - 4.9. The smallest wMSE values are in bold face. For most object scene distributions, NCAR provides the best result (lowest wMSE).

Fig.	Weighted Mean Square Error			
	CLEAN	MEM	DDM	NCAR
4.3	6.48×10^{-3}	4.04×10^{1}	1.64×10^{-2}	$\mathbf{2.00 \times 10^{-3}}$
4.4	2.20×10^{1}	$\mathbf{1.31 \times 10^{0}}$	2.91×10^{0}	1.42×10^{1}
4.5	5.33×10^{2}	2.87×10^{3}	1.62×10^{1}	$\mathbf{1.39 \times 10^{0}}$
4.6	1.09×10^{5}	4.16×10^{2}	$\mathbf{9.15 \times 10^{0}}$	1.36×10^{1}
4.7	2.88×10^{2}	9.11×10^{7}	1.65×10^{1}	$\mathbf{5.15 \times 10^{0}}$
4.8	5.09×10^{3}	3.29×10^{5}	9.22×10^{3}	$\mathbf{2.25 \times 10^{3}}$
4.9	6.24×10^{1}	4.65×10^{12}	$\mathbf{2.77 \times 10^{-1}}$	2.01×10^{1}

an apparent subjective assessment of its superiority. In this ultra-low SNR regime, therefore, wMSE is no longer a good indicator of image quality. It's definition and scope is therefore limited, but it is suitable for most of the scenarios examined.

While for two sources separated by 1° (Fig. 4.6), NCAR does not exhibit the lowest wMSE, it is almost equivalent to the lowest (that of DDM). Both DDM and NCAR demonstrate much lower wMSE than CLEAN and MEM due to their ability to reconstruct and separate the two sources. Finally, for the scenario in Fig. 4.9 (two sources with varying intensity in low SNR), DDM again demonstrates the lowest wMSE. This result is, however, expected, since this object distribution was presented specifically to demonstrate a regime in which NCAR is not preferable to other techniques.

4.7 Discussion

An algebraic technique such as NCAR is necessary to exploit the super-resolution capability of the RM while simultaneously suppressing noise fluctuations. NCAR will not, however, be suitable for all scenarios. As shown, its weakest performance is in resolving two weak sources with different measured rates. Because of the inherently low SNR, the peaks are "blurred" to different degrees, suppressing the weaker of the two and thus not accurately depicting relative source strength. In the other cases presented, however, NCAR performs quite well, reconstructing the object scene and suppressing noise fluctuations that typically plague algebraic solutions. In scenarios where the size of the source is of importance, NCAR will not produce good results with low SNR. If the ultimate desire is to survey and locate true sources, however, NCAR generates images free of spurious reconstructions with a visual representation of the locational uncertainty of a particular source measurement.

Chapter 5

Sensitivity

5.1 Signal-to-Noise Ratio

For a counting detector with only Poisson noise, the signal-to-noise ratio (SNR) of an unmodulated observation is given by

$$\mathrm{SNR} = \frac{C_S}{\sqrt{C_S + C_B}}, \tag{5.1}$$

where C_S is the total number of measured source events and C_B the total number of background events. These counts are related to the exposure time, T, and source and background rates, S and B, by

$$C_S = ST \quad \text{and} \quad C_B = BT, \tag{5.2}$$

and so Eq. 5.1 may be rewritten as

$$\mathrm{SNR} = S\left(\frac{T}{S+B}\right)^{1/2}. \tag{5.3}$$

A mutliplexing imaging device has an inherently lower SNR than an equivalent non-imaging detector for two reasons: (1) the existence of a modulation which blocks a fraction of incident photons and (2) some of the detected source counts may contribute to sidelobes,

decreasing the image value at the true source location. It can be shown (see Appendix B.3) that the SNR for a multiplexing imager is proportional to the standard deviation, σ, of the system matrix,

$$\mathrm{SNR}_{\mathrm{multiplex}} = \sigma S \left(\frac{T}{B + S(\mu + \gamma\sigma)} \right)^{1/2}. \tag{5.4}$$

The standard deviation parameter accounts for both the mask absorption and system response effect on the signal value, while the average, μ, and skewness[1], γ, of the system matrix are correlated with the noise contribution from source counts.

In most imaging modalities, the system matrices are evenly distributed about a mean, and so the skewness will be close to or equal to zero. Additionally, these systems typically detect 50% or less of incident photons (the other fraction is absorbed by the modulation device), and so $\mu \leq 1/2$. In astrophysical applications, background contributions often dominate observations. If $B \gg S/2$, then Eq. 5.4 may be simplified to

$$\mathrm{SNR}_{\mathrm{multiplex}} \approx \sigma S \left(\frac{T}{B} \right)^{1/2}, \tag{5.5}$$

which is equivalent to that of the background-dominated non-multiplexing observation multiplied by the standard deviation of the instrument response.

5.2 Comparison of Modalities

The coded aperture system matrix is represented by an array of ones (open pixels) and zeros (closed pixels), typically in equal numbers. The standard deviation of this distribution is $\sigma = 1/2$. The SNR of an observation by the coded mask imager is then

$$\mathrm{SNR}_{\mathrm{CA}} \approx \frac{1}{2} S \left(\frac{T}{B} \right)^{1/2}. \tag{5.6}$$

Unlike the coded aperture's discrete system response, RM and RMC system matrices

[1]Skewness is the third standardized moment of a distribution.

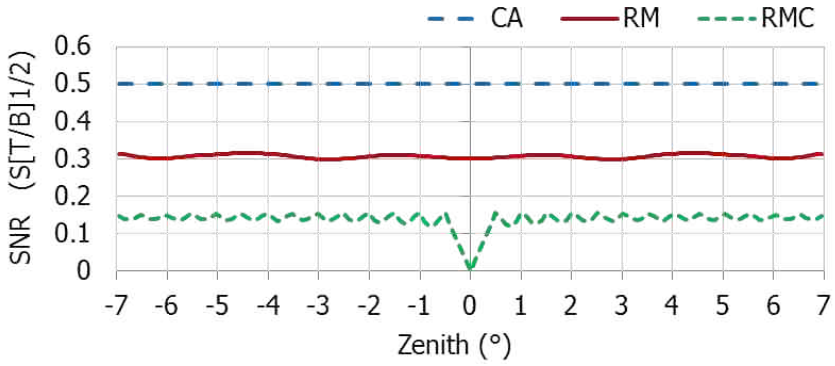


Figure 5.1: SNR of three multiplexing imagers in units of $S\sqrt{T/B}$.

are represented by a continuous range of values: $0 \rightarrow 1$ for the RM and $0 \rightarrow 1/2$ for the RMC. Furthermore, these profiles do not sample all parts of the sky equally, and so the standard deviation is dependent on the source location. For the RMC, the continuous nature of the profiles combined with the inherently low mask transmission reduces the sensitivity considerably compared to a coded aperture. On average, the standard deviation is $\sim$0.16, and so the SNR is

$$\mathrm{SNR}_{\mathrm{RMC}} \approx (0.16)S\left(\frac{T}{B}\right)^{1/2}. \tag{5.7}$$

The standard deviation for the RM is approximately twice as large:

$$\mathrm{SNR}_{\mathrm{RM}} \approx (0.31)S\left(\frac{T}{B}\right)^{1/2}. \tag{5.8}$$

In Fig. 5.1, the SNR of these three multiplexing imagers (with comparable geometries) are plotted as a function of source angle. The results are compared in units of $S\sqrt{T/B}$, and so an SNR of 1 corresponds to an observation by a non-modulating device. The coded aperture is found to be the ideal device in terms of sensitivity; the SNR is limited only by the number of open mask pixels, and the sensitivity is completely uniform across the sky. The RMC has the lowest SNR due to its intrinsically lower mask transmission. Furthermore, the

sensitivity drops to zero directly overhead, where the object scene is unmodulated; counts from this region are indistinguishable from background.

For an RM, the overall sensitivity is approximately twice as high as the average RMC SNR. The summation of the detector-dependent cross-correlation images smooths over regions of insensitivity, providing virtual uniformity across the sky. Additional detectors enhance this effect, though the mean sensitivity remains constant.

5.3 Verification of Sensitivity

Analytical simulations are performed using software described in Sec. 6.4 to verify the sensitivity calculation of the RM. A system geometry is defined with the same parameters as those of the laboratory prototype presented in Chap. 6. Resulting simulated count profiles are used to generate the raw cross-correlation image. A PSF is scaled to the intensity of the source and subtracted from this image. The standard deviation of the residual is taken as the noise component in calculation of the simulation SNR (Fig. 5.2).

Six pairs of source and background rates are selected for a fixed exposure time as a

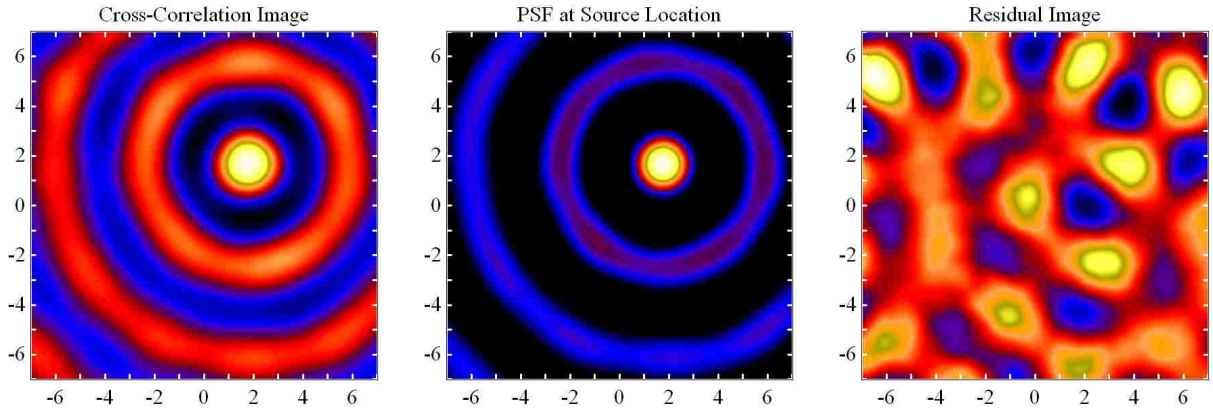


Figure 5.2: Simulation results verify the calculation of the RM SNR formula. (a) Cross-correlation image of a point source. (b) PSF centered at source location, (c) Residual image found by subtracting a scaled (b) from (a) and used to measure noise.

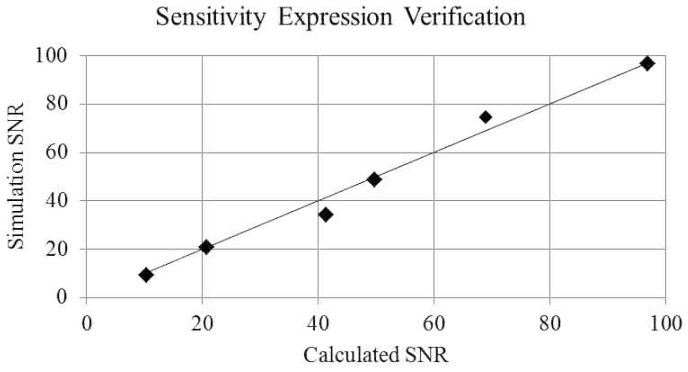


Figure 5.3: Simulation result for RM SNR vs. calculated values for a range of source and background rates.

selection of input SNRs. For each input SNR, four independent trials are run, and the simulation SNRs are averaged. These values are compared with those calculated using Eq. 5.8 in Fig. 5.3. A 1:1 fit to the analytically- and Monte Carlo-based sensitivity values has a coefficient of determination (R^2) value of 0.98, showing good agreement.

5.4 Sensitivity vs. Angular Resolution

In the simple case of zero bar thickness and perfect attenuation, the standard deviation of the RM response, and thus its sensitivity, is maximized for equal bar and spacing width. For finite bar thickness, shadow lengthening increases a', the bar shadow width, and so larger spacing is required to compensate to keep the sensitivity high. In Fig. 5.4, the SNR is plotted against the slat width percentage, $a/(a+b)$, the ratio of slat width to pitch. Here, the slat thickness is 1.5 cm, and so the maximum SNR occurs when the slat width is $0.37/0.63 = 59\%$ of the spacing width. For larger slat thickness, this maximum would occur at a smaller slat width fraction.

At slat width fractions larger than the ideal, detectors are shadowed during the profile for a finite period of time, reducing standard deviation and thus sensitivity. Conceptually, this is seen by a narrowing of the PSF, which, in general, aids sensitivity. However, the reduced transmission has a greater effect, causing the SNR to drop. This same effect occurs for reduced slat width fraction, where spacings are large, and detectors are fully-illuminated

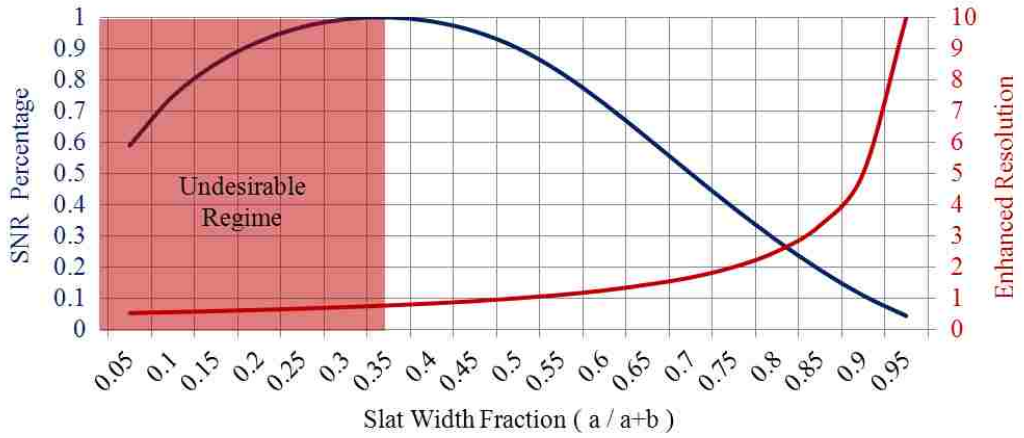


Figure 5.4: The SNR averaged over the FOV (blue) and the angular resolution enhancement (red) for varying slat width fractions, $a/(a+b)$, and grid thickness 1.5 cm.

for a finite time in the profile. Despite the increased transmission, the PSF is broadened, a similar lowering of the SNR occurs. This symmetric nature of the RM sensitivity curve is seen in Fig. 5.4, and verifies the description by Durouchoux et al. (1983) for zero grid thickness.

Also plotted in Fig. 5.4 is the enhancement of angular resolution, where a value of one corresponds to the default angular resolution with slat width and spacings equal. Since angular resolution is equal to the arcsine of the ratio of slat spacing and grid/detector displacement, angular resolution increases with increasing slat width fraction. This plot shows that for fixed detector size, enhanced resolution may be achieved by increasing the slat width fraction, but at the cost of lowered sensitivity. Furthermore, angular resolution is enhanced relatively little for large drops in sensitivity–e.g. for a resolution enhancement of 2, the sensitivity is 39% of its maximum value. At smaller than ideal slat width fractions, both sensitivity and resolution enhancement are reduced, and so this regime has undesirable properties for imaging.

5.5 Super-Resolution

By moving modulation of incident photons to the temporal domain, the RM and RMC gain a unique advantage over the coded aperture. Rather than the sampling of the instrument response being fixed by the instrument geometry as with a coded aperture, the time-tagged data may be rebinned during analysis. By sub-sampling the observation profiles (and the instrument response function), the object scene can likewise be further subdivided.

With the appropriate deconvolution algorithm, super-resolution, resolving power beyond the intrinsic geometric resolution of the instrument, may be achieved. There is, however, an inherent loss of sensitivity in achieving super-resolution. A proof outlined in Appendex B.4 suggests that the incorporation of a factor η of super-resolution implies the system matrix, P', and consequently its standard deviation, σ, should be modified with a coefficient $1/\eta^2$. By keeping P' fixed, this result may instead be included in the SNR expression by rewriting Eq. 5.5 as

$$\mathrm{SNR}_{\mathrm{multiplex}} \approx \frac{\sigma}{\eta^2} S \left(\frac{T}{B}\right)^{1/2}. \tag{5.9}$$

5.6 Detector Efficiency Consideration in SNR Comparison

As demonstrated in Sec. 5.2, due to the single mask design and ideal system matrix distribution, the coded mask provides a higher SNR than the RM for any observation. This result assumes perfect detector efficiency of all incident photons. In practice, the detector efficiency for a coded aperture is less than that for an RM, particularly for large zenith angles and high energies, due to an instrinsically smaller horizontal dimension.

At equal geometrical resolutions, the coded aperture detector pixels are half the width of those on a comparable RM. Consider a coded aperture/RM with mask pixel/slat size a. The RM detector diameter, c_{RM}, is ideally given by $c_{\mathrm{RM}} = a$, since this condition maximizes

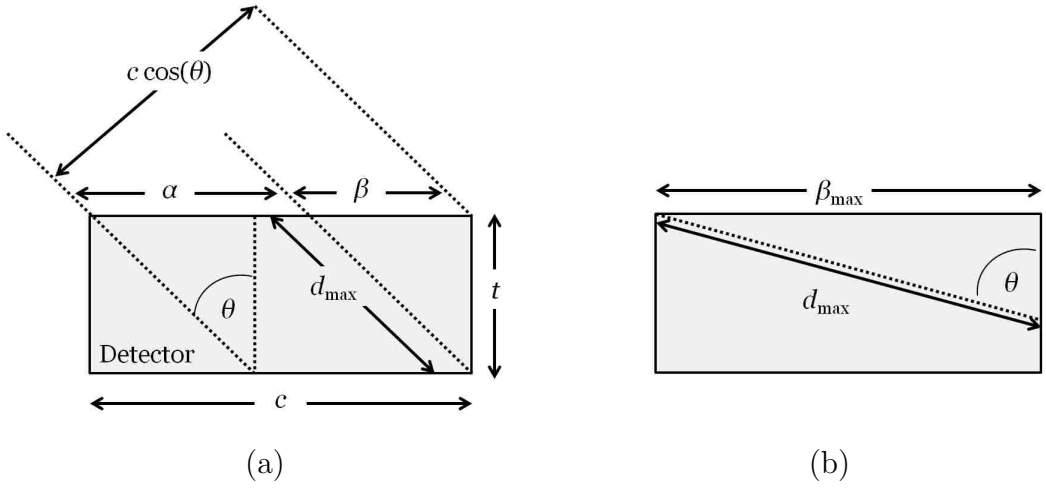


Figure 5.5: Diagrams describing the computation of the average distance, d_{ave} a photon traverses through a 1-dimensional detector for incidence angle θ in (a) the typical case, and (b) at extreme zenith angle.

the contrast and modulation of the profiles, thus maximizing the standard deviation and sensitivity. A coded aperture pixel, c_{CA}, however, is $c_{\text{CA}} = a/2$ or smaller due to the Nyquist criterion; the detection plane must sample the shadow pattern at twice the shadow's spatial frequency.

The inclusion of super-resolution capability requires further subdivision of these elements, such that the pixel size is $c_{\text{CA}} = a/(2\eta)$. For the RM, however, a subdivision of the observational data is a subdivision of temporal elements in the measured profiles. For data time-tagged to good precision, this procedure is trivial and the detector sizes remain unchanged. The inherently smaller pixel size of the coded aperture indicates that for obliquely incident photons (particularly at higher energies), the RM will have a greater detection efficiency, and possibly a comparable sensitivity to the coded aperture in this regime.

As a suitable approximation to this problem, a two-dimensional detector is shown in Fig. 5.5. A photon that enters through the top detector face may have a trajectory that passes through the bottom face or one of the side faces. The average mean path for detector

thickness t is derived in Appendix B.5, where it is shown that

$$d_{\text{ave}}(c) = \begin{cases} t\left(1 - \frac{t}{2c}\tan\theta\right), & \text{for } 0 \leq \theta \leq \tan^{-1}\left(\frac{c}{t}\right) \\ \frac{c}{2\tan\theta}, & \text{for } \theta > \tan^{-1}\left(\frac{c}{t}\right). \end{cases} \tag{5.10}$$

The detection efficiency is calculated based on the above equation,

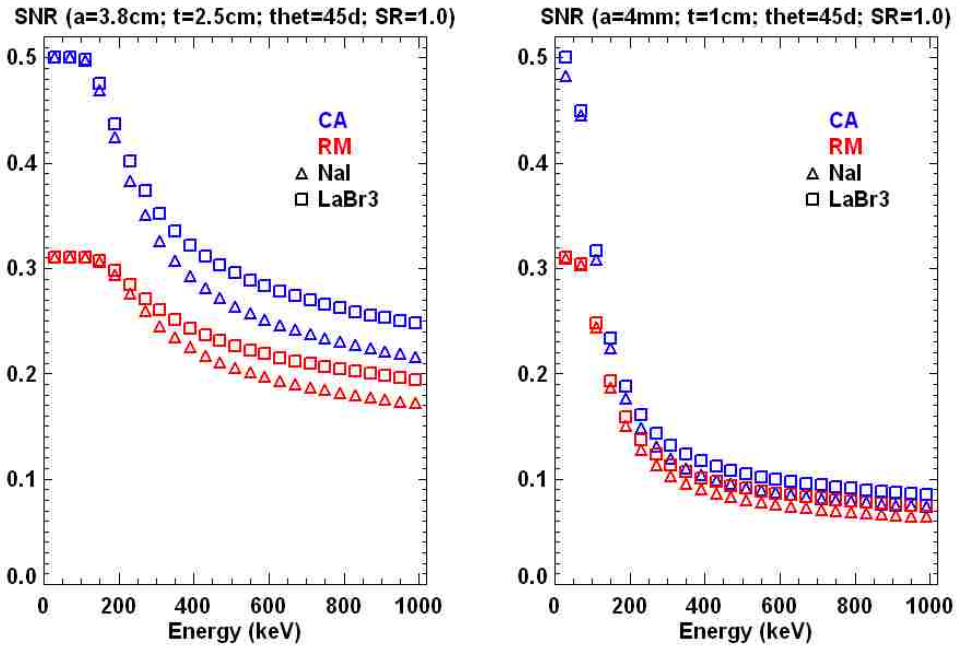
$$\epsilon_{\text{det}}(c) = 1 - \exp\left[-\sigma\rho d_{\text{ave}}(c)\right], \tag{5.11}$$

where σ is the mass attenuation coefficient and ρ the density of the detection material. If the background is due primarily to external contributions, then both the source rate S and background rate B are dependent on this detection efficiency; the detection efficiency should be included in Eq. 5.9 of the SNR,

$$\text{SNR} \approx \frac{\sigma}{\eta^2}\sqrt{\epsilon_{\text{det}}}S\left(\frac{T}{B}\right)^{1/2}. \tag{5.12}$$

If the background rate is instead due primarily to intrinsic contributions (as may be the case at low energies for scintillators such as $LaBr_3$ described in Sec. 7.1.1), then B is not primarily dependent on the detection efficiency, and Eq. 5.12 should be written with ϵ_{det} to the first power. The problem is therefore dependent on the detector material and environment to which the instrument is exposed. Eq. 5.12 therefore represents a conservative estimate in terms of comparing the RM detection efficiency to that of the coded aperture, and so in many cases the ratio of RM to coded aperture SNR will be better than is described here.

Eq. 5.12 is used to directly compare the sensitivity of the RM to the coded aperture. Several scenarios are examined, with parameters selected for mask pixel/slat width a, detector thickness t, photon energy E, photon zenith θ, and super-resolution factor η. In each case, one parameter is variable while the others are fixed. For each scenario, results are

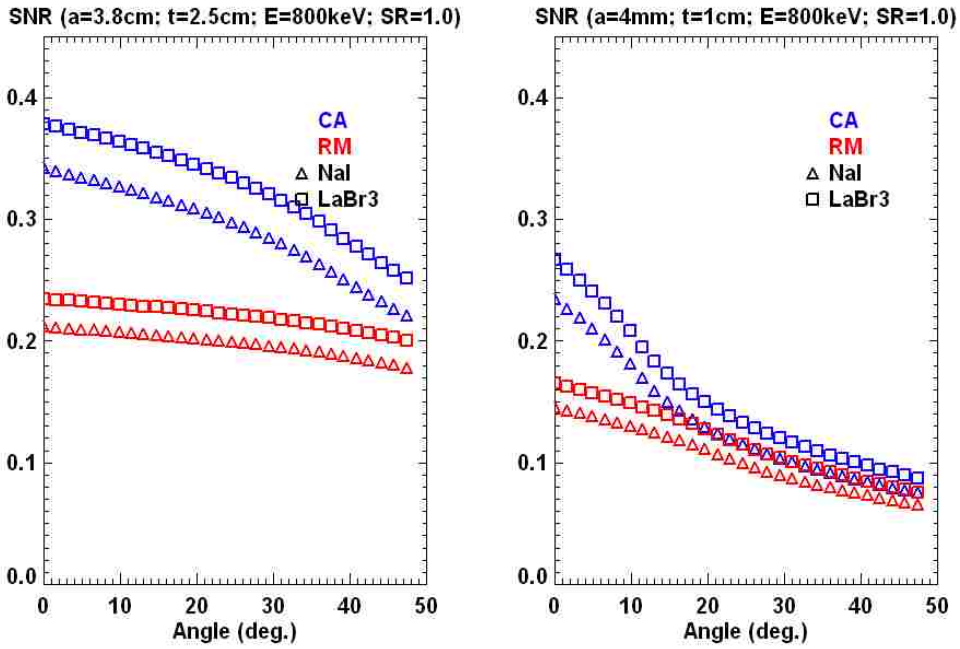


(a) (b)

Figure 5.6: SNR of the RM (lower, red curves) and coded aperture (upper, blue curves) are compared for photon energies ranging from 30 keV to 1 MeV with 45° zenith angle for (a) $a = 3.8$ cm, $t = 2.5$ cm, and (b) $a = 4$ mm, $t = 1$ cm.

calculated for two scintillator materials, sodium iodide and lanthanum bromide (described further in Sec. 7.1.1).

We first examine energy dependence at angle of incidence $\theta = 45°$. Small angles will show little change in the relative sensitivities of the two instruments and angles much larger than 45° are unreasonable in practice. Fig. 5.6(a) shows the results of using $a = 3.8$ cm and $t = 2.5$ cm (based on the RM prototype geometry discussed in Chap. 6). For low energies (< 150 keV), the detection efficiencies are high, with the RM providing SNR ≈ 0.31 and coded aperture providing SNR ≈ 0.50. At higher energies, the detector width advantage of the RM is noticeable, as the relative difference in SNRs between the two instruments becomes much smaller. Even at 1 MeV, however, the RM SNR does not exceed the coded aperture SNR. Furthermore, the shapes of the plots suggest the RM will not provide an advantage at even higher energies using this geometry. Fig. 5.6(b) shows a similar plot for

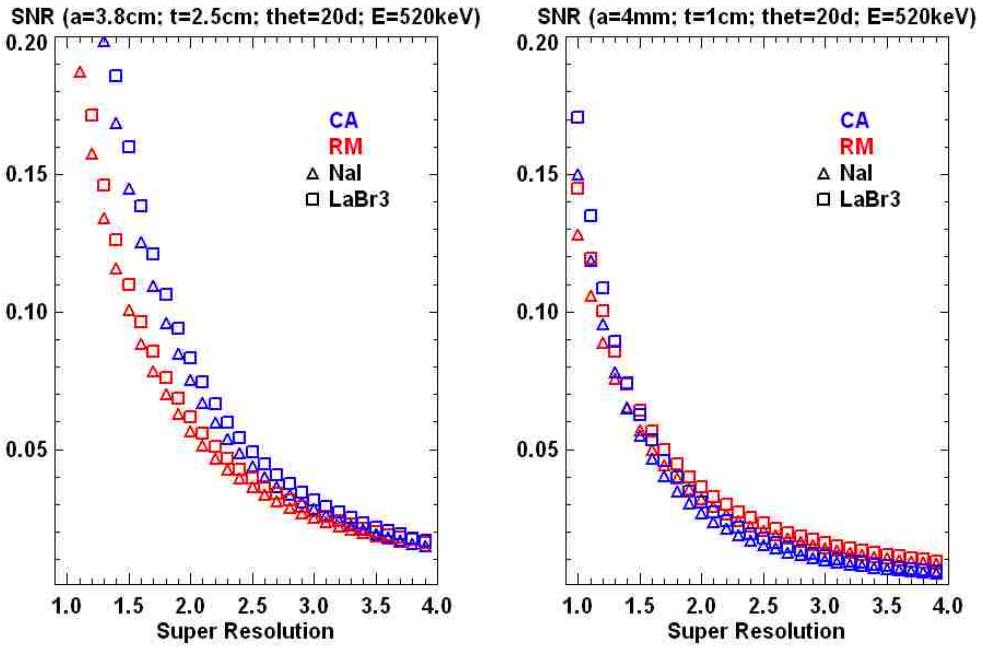


(a) (b)

Figure 5.7: SNR of the RM and coded aperture are compared for photon energy 800 keV and azimuth ranging from 0 to 50°, with (a) $a = 3.8$ cm, $t = 2.5$ cm, and (b) $a = 4$ mm, $t = 1$ cm.

$a = 4$ mm and $t = 1$ cm. Here, the RM advantage is more prominent, with the RM SNR nearly equaling the coded aperture at higher energies.

Next, SNR dependence on photon incidence (i.e., zenith) angle is examined. Photon energy is fixed at 800 keV. Fig. 5.7(a) examines the RM prototype geometry. At $\theta = 0$, SNRs are lower than the expected 0.31 and 0.50 due to imperfect absorption efficiency. The SNR of the coded aperture, however, drops off faster at higher angles than does the SNR of the RM, and the two become comparable at the highest of incident angles. While the RM SNR conceivably overcomes the coded aperture SNR at even higher angles, such shallow angles are not feasible in practice. Fig. 5.7(b) examines this dependency for the alternate geometry. At $\theta = 0$, SNRs are lower than in (a), due to thinner detectors. Convergence of the coded aperture and RM SNR is observed at higher angles, though the RM SNR still never matches that of the coded aperture.



(a) (b)

Figure 5.8: SNR of the RM and coded aperture are compared for photon energy 520 keV and zenith angle 20°, and super-resolutions ranging from 1 to 4, with (a) $a = 3.8$ cm, $t = 2.5$ cm, and (b) $a = 4$ mm, $t = 1$ cm.

Finally, Fig. 5.8 examines the SNR dependence on super-resolution of both instruments. In 5.8(a) (prototype geometry), the two SNRs begin convergence at around $\eta = 3 \rightarrow 4$, but due to the inherently low SNR obtained with $\eta > 1$ (0.025 at $\eta = 3$), the result may lack significance. In Fig. 5.8(b) SNR of the two instruments converges sooner, at $\eta \approx 1.5$, and the superior detection efficiency of the RM provides a better SNR than the coded aperture at higher super-resolutions.

5.7 Discussion

Modulation imaging instruments demonstrate a sensitivity proportional to the standard deviation of the system response distribution. From this effect alone, the coded aperture is the preferred choice over the RM since it can offer the maximum SNR for a given mask

transmission. Partially for this reason, coded apertures have seen widespread use in hard x-ray/gamma-ray imaging.

Besides this "coding power," however, detection efficiency must be considered in the analysis. Due to the mechanisms by which these two modalities modulate incident photons, the RM has larger detector elements, and thus better detection efficiency than a coded aperture. This is especially true in the case of large zenith angle, high photon energy, or a high factor of super-resolution. In these regimes, the high RM detector efficiency may result in an SNR comparable to (though rarely exceeding) that of a coded aperture.

The RM maintains two advantages over the coded aperture: (1) since the RM detectors are approximately twice the dimension as those of the comparable coded aperture, the number of readout electronics is smaller by a factor of four, reducing cost and complexity; (2) super-resolution may be achieved without altering the physical instrument and without sacrificing detection efficiency. In the case of high-energy and wide FOV, these results make the RM an attractive and viable alternative to a coded aperture design.

Chapter 6

Laboratory Prototype RM

6.1 Introduction

Empirical verifications of the feasibility of the RM technique (Sec. 2.4), system response formulae (Chap. 3), NCAR reconstruction algorithm (Sec. 4.6), and sensitivity characteristics (Chap. 5) are accomplished by the design, construction, and testing of a laboratory-scale prototype. The Lanthanum Bromide-based Rotating Aperture Telescope (LaBRAT) is a nineteen-channel RM, which features a custom data acquisition system (DAQ) and simulation/analysis software. In this chapter, these components and techniques are described, and measurement results are presented. This work serves as a precursor and proof-of-concept verification for the high-altitude balloon RM mission to be described in Chap. 7.

6.2 Mechanical

The LaBRAT prototype (Fig. 6.1) measures approximately 75 cm tall $\times$ 90 cm wide $\times$ 150 cm long. An aluminum tubing frame provides a rigid structure with low weight. The frame rests on four casters to provide mobility for aiming and transportation.

The mask is composed of eight lead slats, each 3.8 cm wide and 1.9 cm thick, providing $\sim$90% attenuation of 662 keV photons. The slats are spaced 3.8 cm apart and coupled to

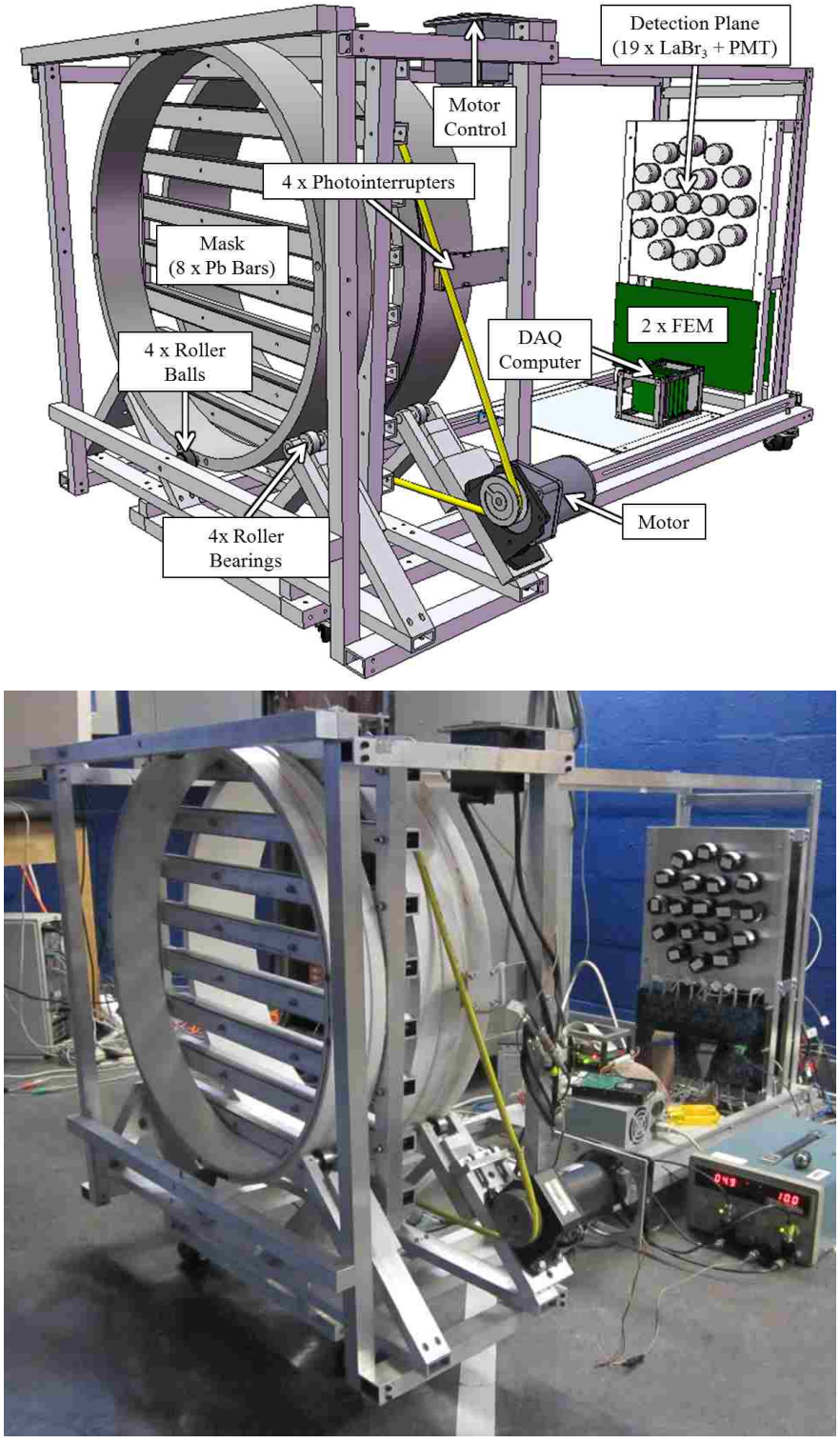
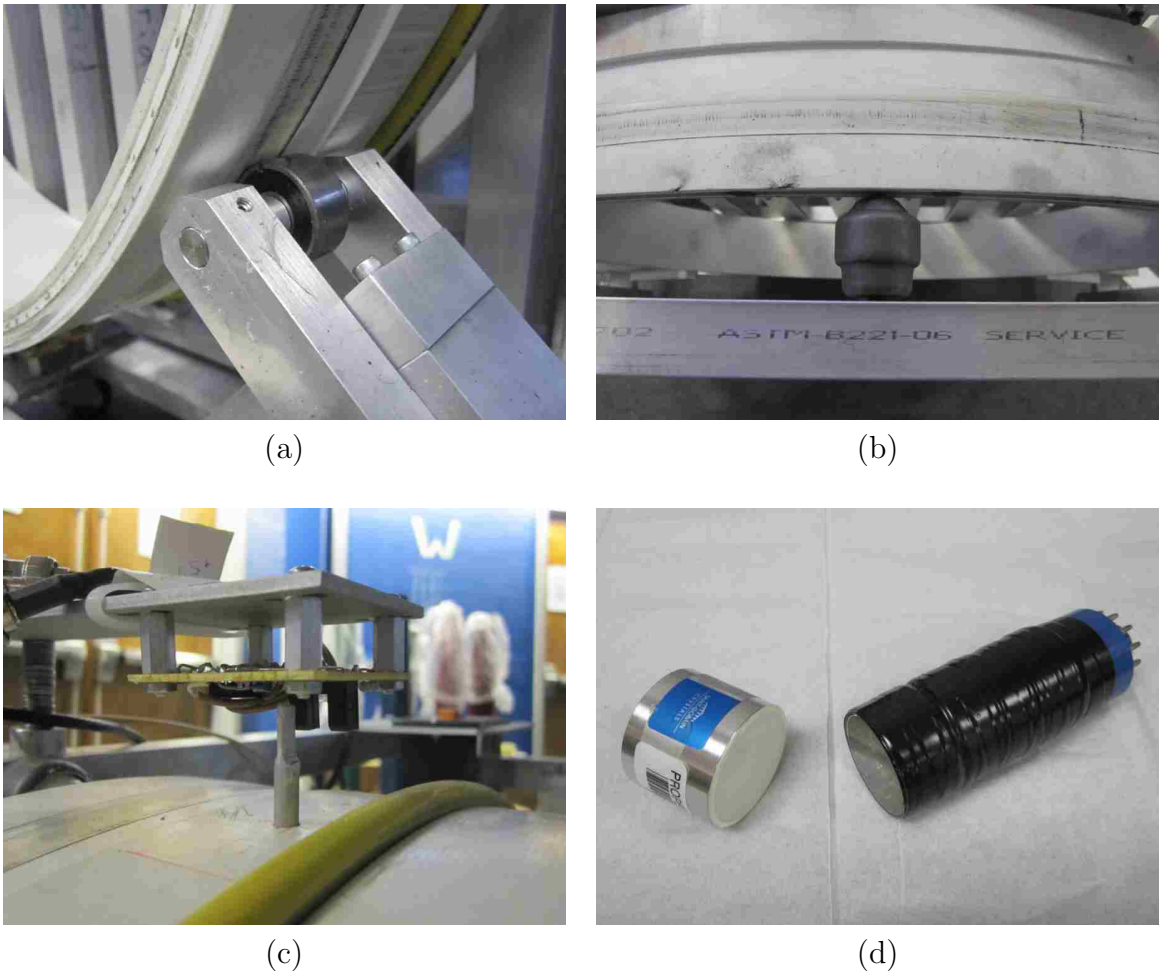


Figure 6.1: (Top) Annotated CAD drawing and (Bottom) Photograph of the LaBRAT prototype.



(a) (b)

(c) (d)

Figure 6.2: (a) Four roller bearings support the mask and allow it to rotate. (b) Four roller-balls keep the mask in place and prevent wobble. (c) Photointerrupter circuit/fork and mask pin for interruption. (d) 3.8 cm diameter, 2.5 cm thick hermetically-sealed $LaBr_3$ scintillator and 3.8 cm diameter Electron Tubes 9102B PMT.

2.5 cm square aluminum tubing on the inside face and 3.8 cm-wide aluminum plates on the outside face; this configuration provides the necessary support for the soft lead slats without impinging on the FOV.

To allow for rotation, the grid frame is sandwiched between two sections of 61 cm (outer diameter) PVC piping, which rest on four roller bearings mounted on base struts angled at 45° from horizontal (Fig. 6.2a). Four roller balls are positioned perpendicular to the mask face and dampen axial motion of the mask composite (Fig. 6.2b). A shallow groove in the rear PVC section allows for attachment of a 1 cm-thick rubber belt, coupled to a single-phase

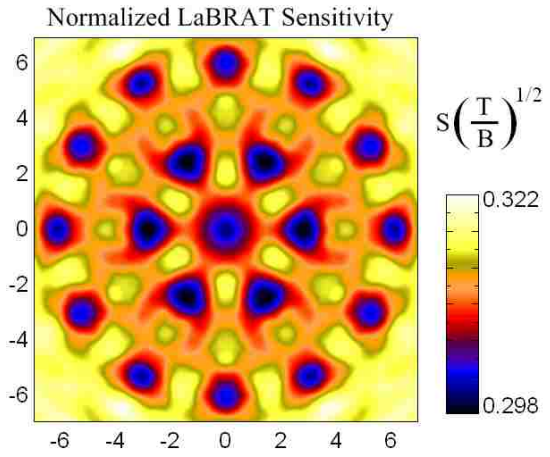


Figure 6.3: Contour plot of the relative sensitivity of LaBRAT's FOV. Relative uniformity is achieved by having multiple detector elements. Axes in degrees.

servo motor. The mask is driven at a frequency of ∼10 rpm; for stationary sources, the rate does not matter, though for transient sources, the rotation period must be much less than the characteristic time scale of the source.

Four infrared LED photointerrupters (PIs) are mounted about the circumference of the rear PVC section. A metal pin attached to the PVC interrupts the transmission between each PI fork once per revolution, providing quarter-period time-tagged markers to allow for correction of nonuniformities in the grid rotation speed during analysis (Fig. 6.2c). PI circuitry is described in Sec. 6.3.2.

The detector plane is comprised of nineteen cylindrical cerium-doped lanthanum bromide inorganic scintillating crystals (scintillation properties discussed in Sec. 7.1.1) arranged in a circular layout; one detector is aligned with the grid axis of rotation and is surrounded by outer concentric rings of six and twelve detectors. Each crystal measures 3.8 cm in diameter × 2.5 cm thick and is hermetically-sealed in an aluminum casing with a glass window on one end. This window is coupled to an Electron Tubes 9102B photomultiplier tube (PMT) with Saint-Gobain BC-630 optical grease (Fig. 6.2d). Each detector/PMT package is optically sealed with black tape to prevent light leakage.

The detector array is mounted on an aluminum frame 1.16 m behind the mask. This geometry provides a 13.8° fully-coded FOV and 1.9° full-width-at-half-maximum (FWHM) geometric angular resolution. As a result of using 19 detectors in LaBRAT, the sensitivity is

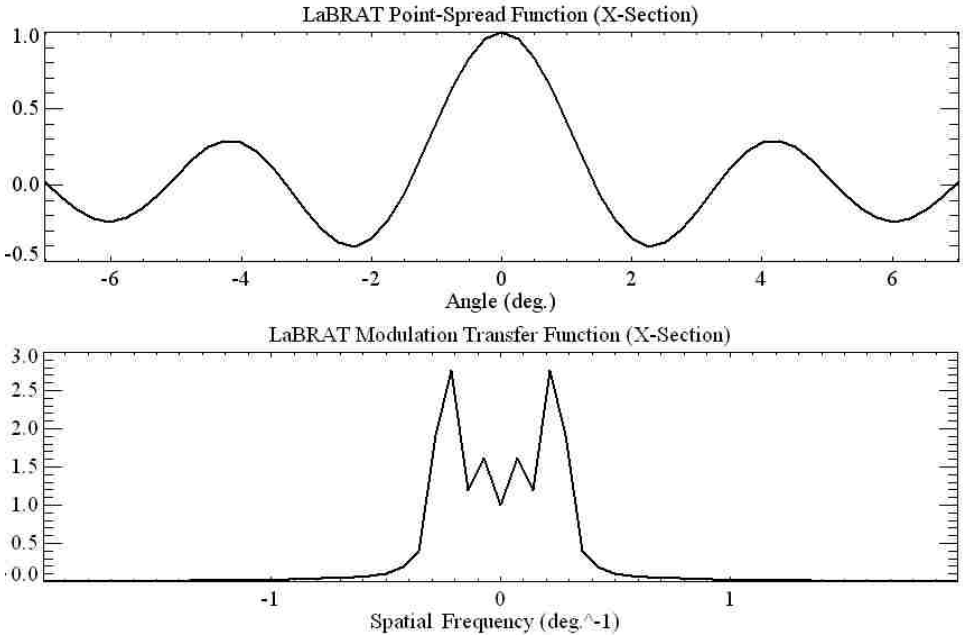


Figure 6.4: Cross-sections of the LaBRAT (Top) point-spread function and (Bottom) modulation transfer function.

fairly uniform. Fig. 6.3 is a contour plot of the sensitivity across the object scene, which, on average, is 0.31 $S(T/B)^{1/2}$, with less than $\pm 4\%$ maximum deviation from this value. Regions of low sensitivity are those that are unmodulated from the viewpoint of one of the detectors; unmodulated regions manifest as a profile with low standard deviation, shown in Chap. 5 to be directly proportional to sensitivity. Since each of the detectors sees this region in the direction of the grid axis, the pattern in the figure thus mimics the detector layout.

Given its FOV and geometric angular resolution, the LaBRAT PSF contains 2 concentric sidelobes about the central source location for a centered source. A cross-section of the response is shown in the top of Fig. 6.4. The modulation transfer function, MTF(η) is the absolute value of a normalized transfer function, $H(\eta)$, which was described in Sec. 4.3:

$$\mathrm{MTF}(\eta) = \frac{|H(\eta)|}{|H(0)|}. \tag{6.1}$$

The MTF is displayed in the bottom of Fig. 6.4. The highest frequency response is approximately ± 0.21 deg^{-1}, and then goes to zero for higher frequencies.

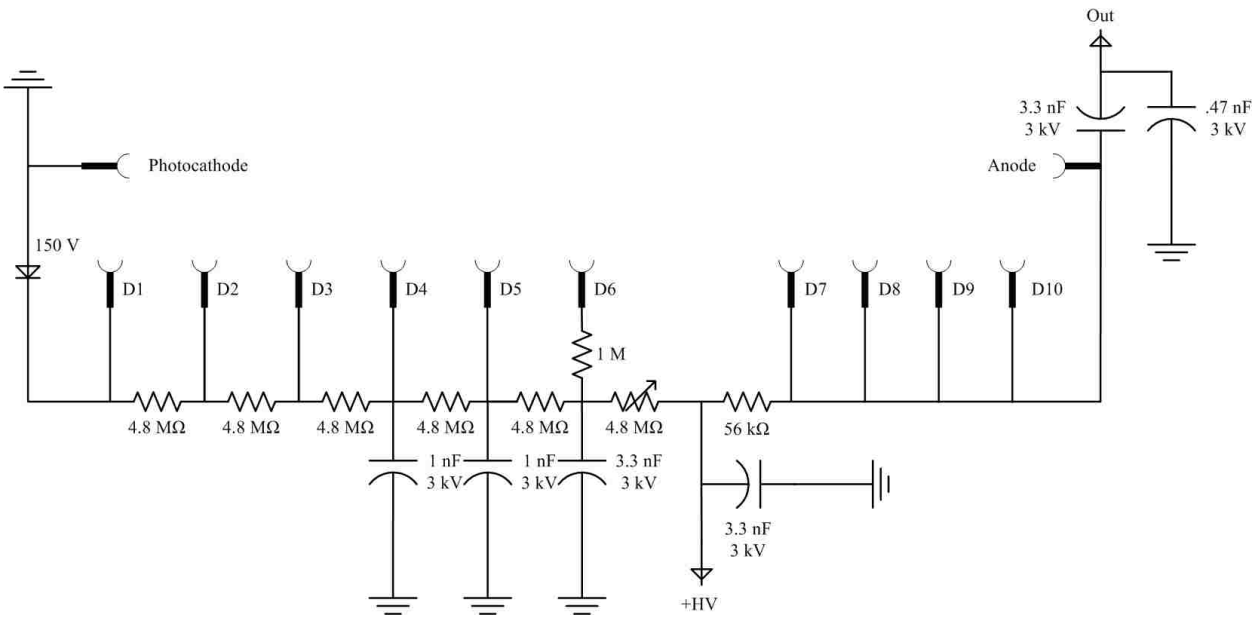


Figure 6.5: The 6-dynode voltage divider circuit used on each PMT base.

6.3 Electronics

6.3.1 PMT Voltage Divider

Due to the relatively high light output of $LaBr_3$, custom PMT interstage voltage dividers are constructed, in which the last four dynodes of a standard ten-dynode chain are shorted together, reducing gain from the standard value (Fig. 6.5). High voltages are selected independently (800 - 1200 V) for each PMT to ensure equalized output signals (i.e. photon energies are measured at approximately the same ADC values). A 150 V zener diode provides a constant voltage drop from the grounded photocathode to the first dynode. Sequential 4.8 MΩ resistors produce interstage voltage drops of 110 - 175 V up to the anode.

Capacitors, valued 1 - 3.3 nF, couple dynodes 4 - 6 and the high voltage input to ground; this placement provides the extra charge necessary to prevent a temporary drop in voltage at these later stages due to large photoelectron input. Another capacitor AC-couples the anode to the signal output, which is fed into the Front End Module (FEM) of the DAQ.

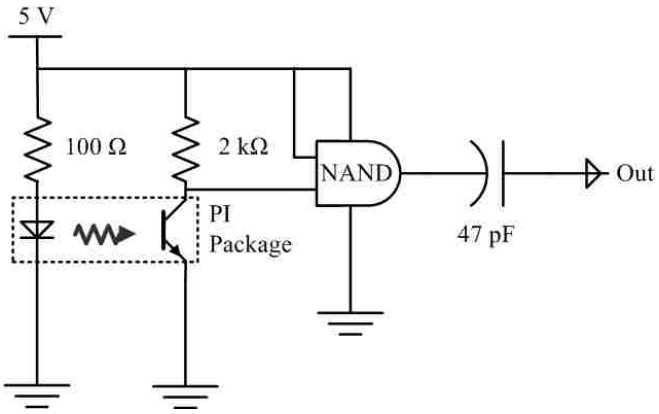


Figure 6.6: LaBRAT Photointerrupter Circuit Diagram.

6.3.2 Photointerrupter Circuit

The photointerrupter circuit (Fig. 6.6) makes it possible to adjust for nonuniformity in the grid rotation. It is designed to mimic the short negative pulse output of a PMT anode signal. A potential difference of 5 V is applied to the circuit. Infrared light from the LED allows the photodetector to provide a path to ground, keeping the collector voltage at 300 mV. The collector and constant 5 V are supplied as two inputs to a NAND Schmidt trigger ("one-shot"). Interruption of the beam causes the phototransistor to open, and the collector voltage rises to 3 V. The output of the NAND gate, typically held at 5 V, drops to zero. A 47 pF capacitor AC-couples this output to the signal.

The result of an interruption to the photointerrupter package is ultimately a 80 ms-wide negative pulse with an amplitude of 200 mV. The signal is plugged into one of the FEM channels where it is used as a trigger, digitized, and recorded in the same manner as any PMT output. The readout software designates this channel as a photointerrupter, and the time-tagged markers are used to fold the measured count profiles.

6.3.3 Data Acquisition

The 32-channel LaBRAT data acquisition system (DAQ) features two front-end modules (FEMs), a trigger logic module (TLM), an analog-to-digital converter (ADC) board, and a QNX-based computer system for readout control and archiving. The DAQ is custom

designed for LaBRAT, with conversions and operations timed specifically to match the rise and decay times of the $LaBr_3$ light output.

The negative anode signal from each of the 19 PMTs and four photointerrupters is fed into an FEM input channel where the signal is split. One branch is fed into a signal shaper. An integrator and differentiator first convert the signal into one with a constant 4 μs width. A subsequent second-order low-pass integrator inverts the polarity and produces a 10 μs-wide symmetric gaussian pulse. The pulse is then amplified to 0 - 5 V to match the sensitivity of the ADC. The second signal branch feeds into a discriminator shaper. An integrator and differentiator produce a signal with a 200 ns decay time to be sent to a 32-channel TLM. The signal is compared to upper and lower threshold voltages. A positive trigger condition sends a convert signal to each of the four eight-channel ADCs to begin conversion of the shaped anode pulses. A register is set containing the list of the channels triggered, which is read by the computer for use in the data record.

On the ADC board, a track-and-hold is timed to hold the maximum voltage from the signal shaper output pulse. The 12-bit ADC then digitizes each of the channel voltages to a value from 0 - 4095 and outputs the values to a register. The register is read by the computer and archived in the open data file.

6.4 Software

Object-oriented RM simulation and analysis software is written in IDL 6.3. The code features a graphical user interface (GUI), the ability to perform simulations, a DAQ file reader, and image reconstruction capability. The primary GUI window (Fig. 6.7) allows for all the parameters of the RM to be defined, including mask/detector geometry and data binning options. This information is utilized for both simulations and data analysis, which are selected by radio buttons. A click-sensitive source map allows for simulated object scenes to be quickly placed with specific source rates and energies, as well as both external and

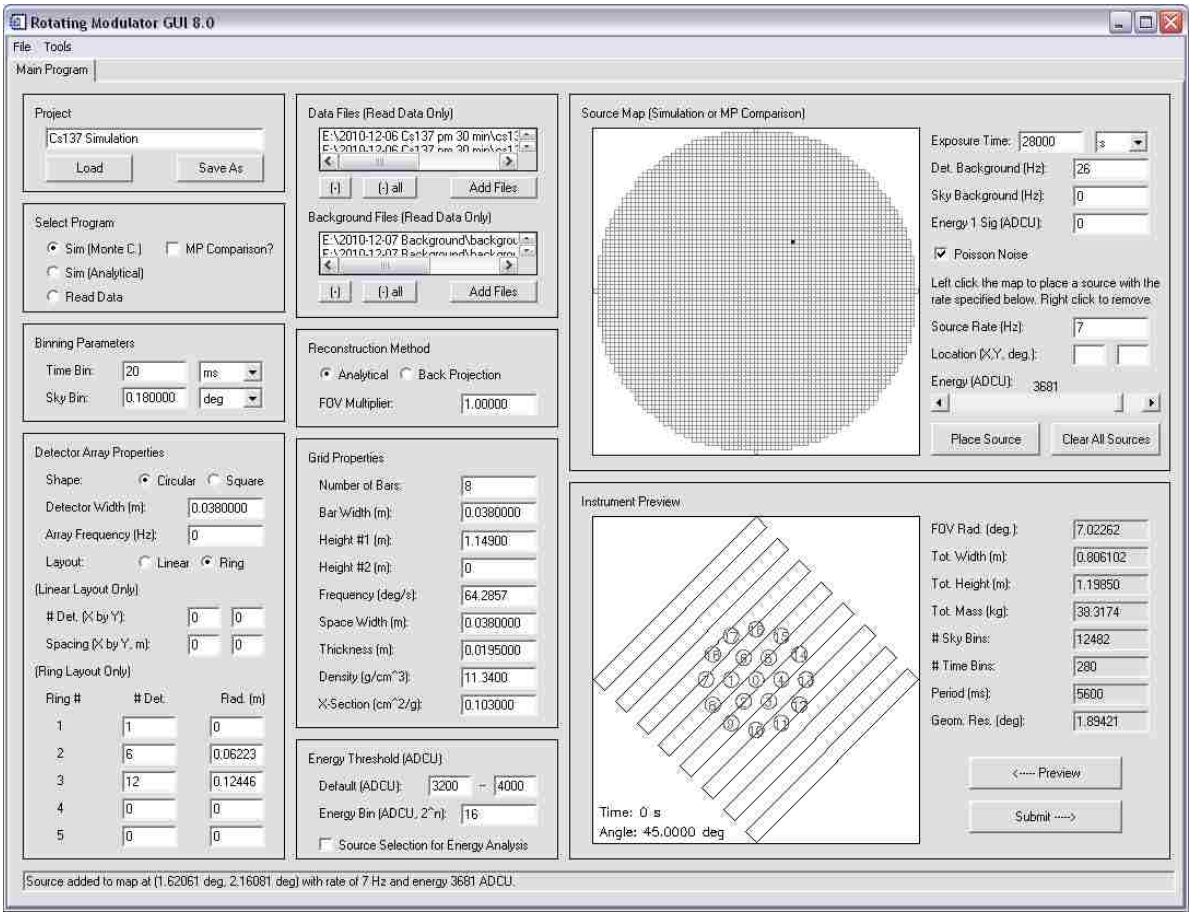


Figure 6.7: The primary window of the RM GUI software allows for easy designation of the RM geometrical parameters, source fields, binning options, data files, and tasks.

internal background contributions. For experimental analysis, data and background files may be selected. Project files containing all specified settings may be defined, saved, and loaded.

Simulations include both analytical and Monte Carlo. Analytical simulations query the pre-computed system matrix. Background and Poisson noise are subsequently added, and profiles are generated based on the source distribution. Monte Carlo simulations object-oriented and are custom written in IDL for this research. Objects are instantiated from classes defining the grid, detectors, and photons to randomly cast incident gamma rays on the instrument based on the object scene distribution. Attenuation in the grid only is taken into account based on the density and cross-section of the slats and energy of the photons.

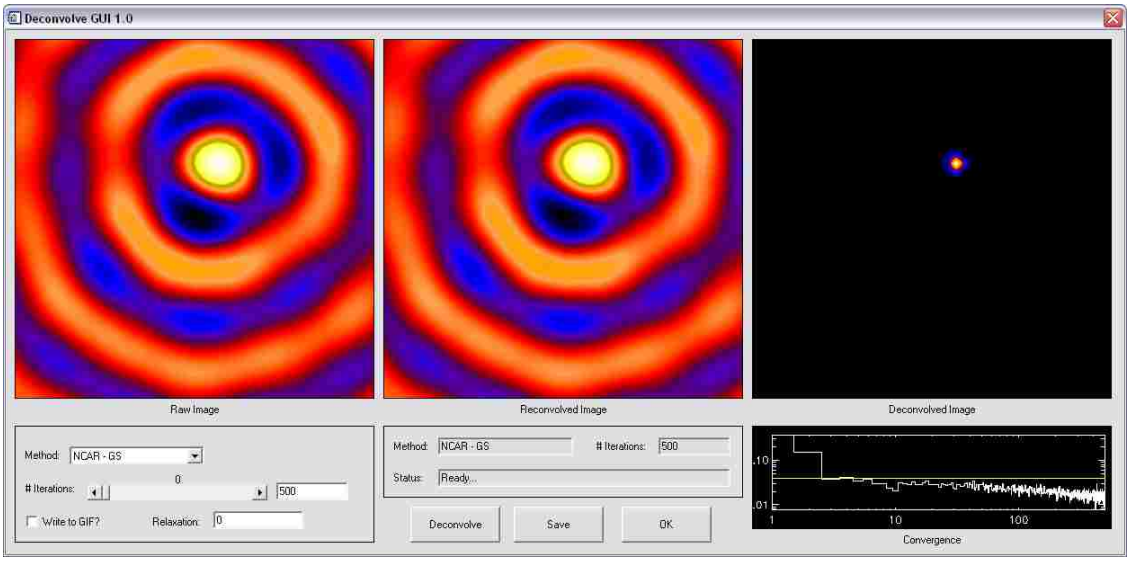


Figure 6.8: The deconvolution/reconstruction GUI allows one of several algorithms to be run for a specified number of iterations. The raw image (left), reconvolved image (middle), reconstruction (right) and convergence plot (bottom right) are shown.

Scattered photons are disregarded as these will contribute only to constant background and should be ignored during energy windowing. While analytical simulations produce the final folded profiles suitable for submission to the image reconstruction analysis, these custom Monte Carlo simulations produce raw data equivalent to that read in for experimental data.

For experimental data, a multi-dimensional array of ADC values and timestamps for each channel is produced after being read from each of a number of files associated with an exposure. The first nineteen channels contain the data from the $LaBr_3$ detectors. An energy window is selected over the range of interest to provide upper and lower energy thresholds for included events. The last four channels are the photointerrupters, and so threshold ADC values mark the timestamps where these interruptions occur. The raw data are subsequently folded over these markers to produce a single count profile for each of the nineteen detector channels. Background rates are determined from exposures taken in the absence of sources, which are read from a separate list of files; the same energy windows are used here, and the measured contribution from background is then subtracted from the profiles. The energy windows as selected on each channel may not be completely

consistent, and so to prevent detector bias, a normalization factor is calculated for each detector proportional to its background rate. The inverses of these normalization factors are multiplied by the respective instrument response in the subsequent analysis.

The system matrix is calculated as described in Chapter 3. A cross-correlation image is generated from the profiles resulting from simulation or experimental data analysis. This raw image and the system matrix are provided to an image deconvolution GUI, which allows several reconstruction algorithms to be run for a specified number of iterations. The GUI shows the raw image, the latest iterative reconstruction, this reconstruction reconvolved with the system matrix, and a plot of convergence (Fig. 6.8).

6.5 Results

6.5.1 Simulation

Analytical simulations have been performed with the software described in Sec. 6.4 using the LaBRAT geometry to examine the RM response (Sec. 4.4) and to verify the derived sensitivity formulas (Sec. 5.3). Monte Carlo simulations have similarly been used to validate the analytical RM system matrix (Secs. 3.3.3, 3.3.2) and to demonstrate the advantage of the NCAR reconstruction algorithm (Sec. 4.6.3) in reducing noise and achieving super-resolution.

Here, analytical simulations demonstrate the reconstruction capabilities in the limit of low noise (measured results are then presented in Sec. 6.5.2). In the high SNR regime, reconstruction fidelity is related primarily to the existence of sidelobes in the system response. This analysis empirically examines the impact of extended sidelobes on the RM technique. For this reason, complex object scene distributions are produced in Fig. 6.9: an object scene with numerous point sources, extended line sources, and several small extended sources with varying brightnesses. Cross-correlation images of the simulations show the difficulty of the deconvolution problem; the detailed object scenes must be reconstructed from raw images

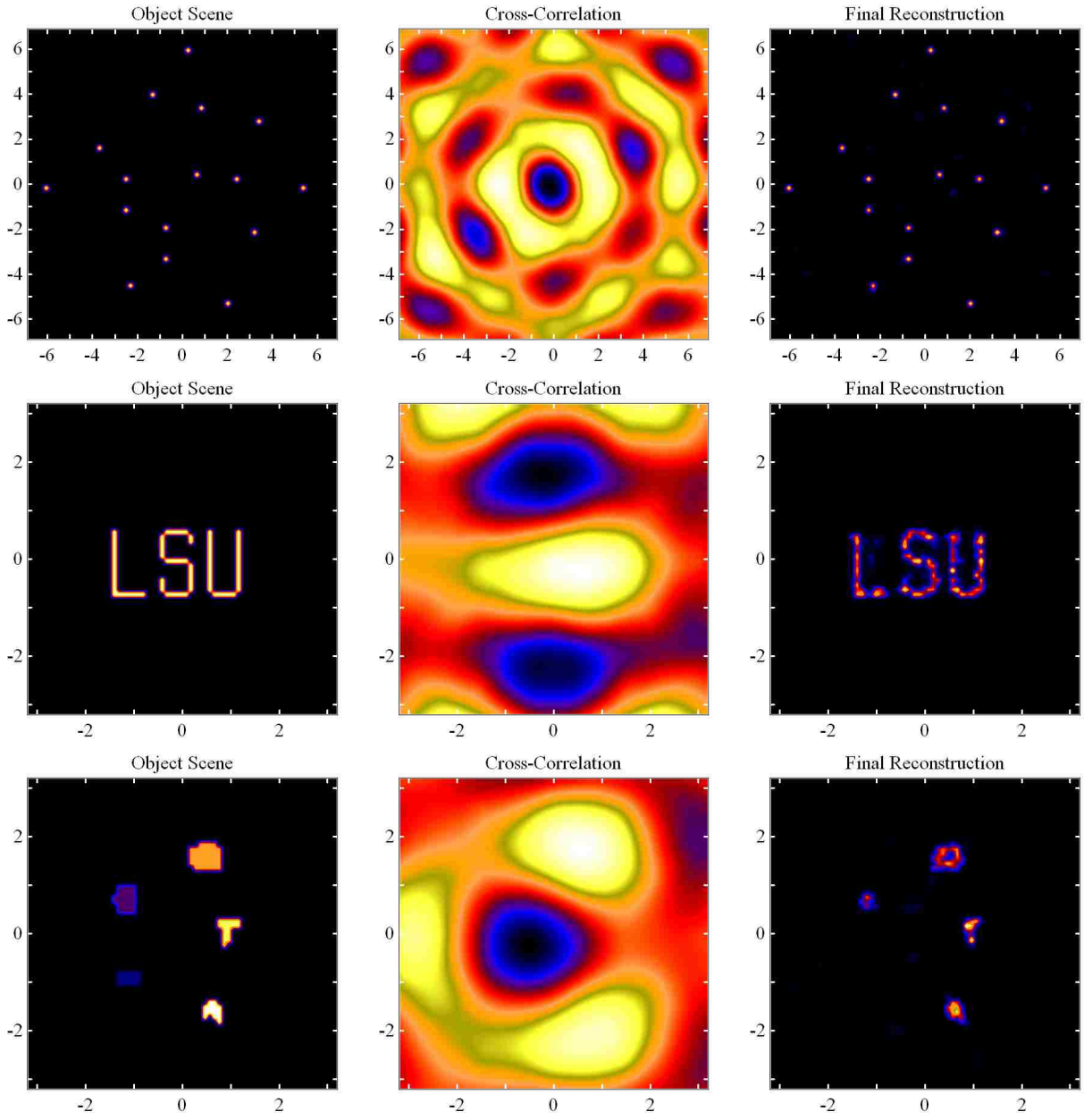


Figure 6.9: Object scenes, cross-correlation images, and final reconstructions demonstrate the ability of the RM technique to remove sidelobes in the high SNR regime for a variety of source distributions.

with very broad features. The final reconstructions are the result of the NCAR algorithm run for 200,000 iterations. Multiple sources are well reconstructed, despite the interference of sidelobes from each source. Similarly, extended line sources are reconstructed with only moderate noise arising from the convolution. Finally, sources with varying brightnesses maintain the ratios reasonably well; only the dimmest source is lost in the reconstruction.

6.5.2 Experimental

As described in Sec. 7.1.1, the use of $LaBr_3$ in the LaBRAT prototype provides better energy resolution than more commonly-used inorganic scintillators such as sodium idiode and cesium iodide. In Fig. 6.10, the gamma-ray lines from ^{133}Ba and ^{137}Cs isotopes are

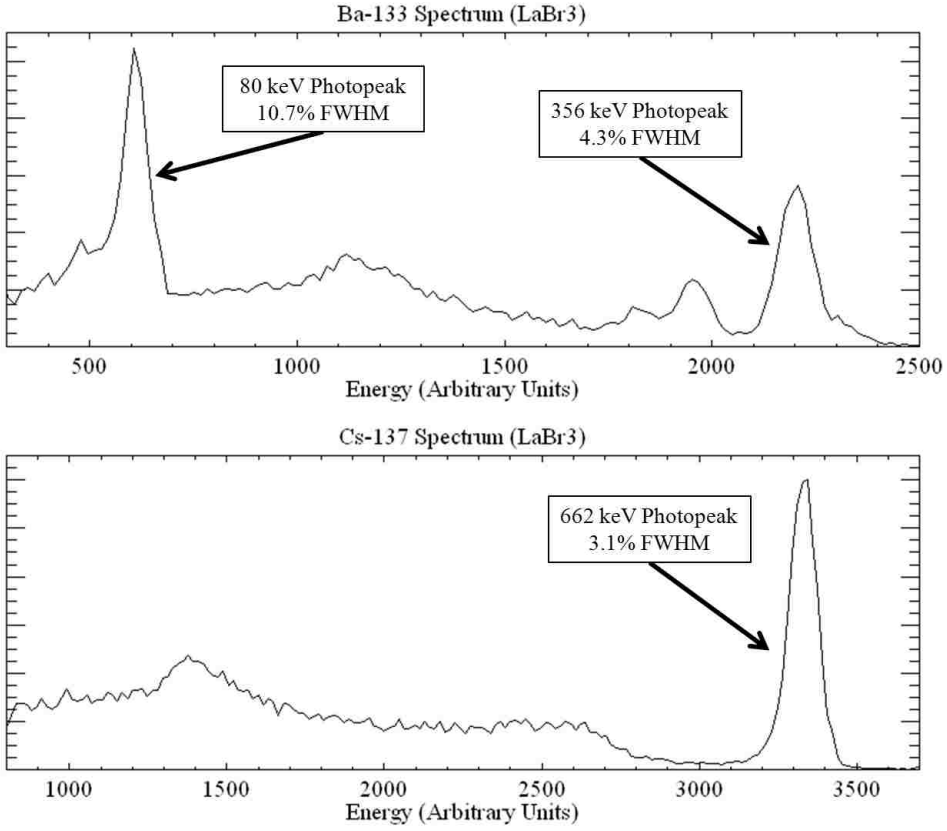


Figure 6.10: (Top) ^{133}Ba and (Bottom) ^{137}Cs spectrum show the resolutions obtained with the LaBRAT prototype at 80, 356, and 662 keV.

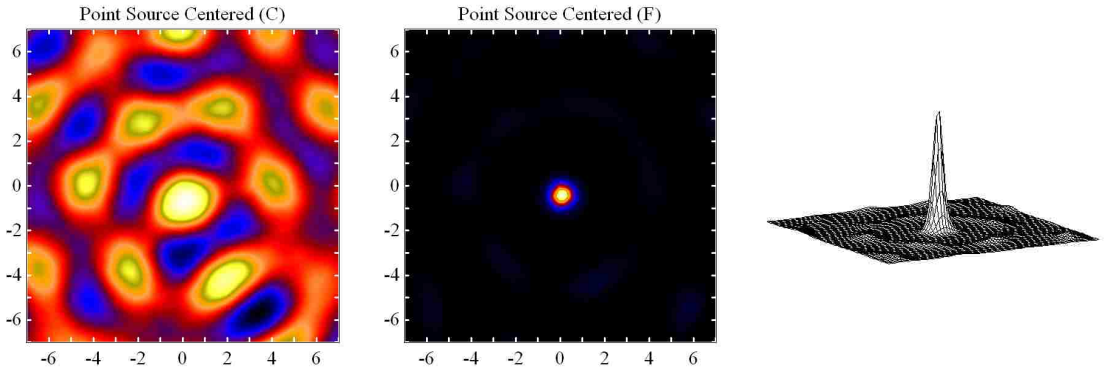


Figure 6.11: Contour plots of the cross-correlation (C), final reconstructed image (F), and surface mesh of final reconstruction for a single ^{137}Cs point source centered in the field of view.

measured to compare the energy resolutions at 80, 356, and 662 keV. At 662 keV, the resolution is measured to be 3.1%, close to the best reported value, 2.6%. (Short exposure time, low quantum efficiency of the PMT photocathode, and noise in the data acquisition all contribute to the worsening of this resolution.) At lower energies, the resolution worsens (10.7% at 80 keV) due to increased statistical uncertainties.

The first experimental result is for a single 100 μCi ^{137}Cs isotope. The source is placed $\sim$10 m from the detection plane. Since the RM was designed primarily as a far-field imager, the reconstruction algorithm assumes detected photons from a source are mutually parallel when incident on the instrument. Parallax can degrade the image, and so this distance provides a suitable tradeoff between minimal parallax and high measured source rate (for a brief discussion of a possible solution to this effect, see Sec. 8.3). The measured source rate is 6 cts/s and background rate is 30 cts/s (within the selected 662 keV photopeak window). For an exposure time of 45 minutes, the cross-correlation image and final reconstruction are shown in Fig. 6.11. Since the exposure time is relatively short given the inherently low source to background rates, the final reconstruction (using the NCAR deconvolution algorithm) exhibits some uncertainty on the source location indicated by a blurring. Longer exposure times would narrow this reconstruction.

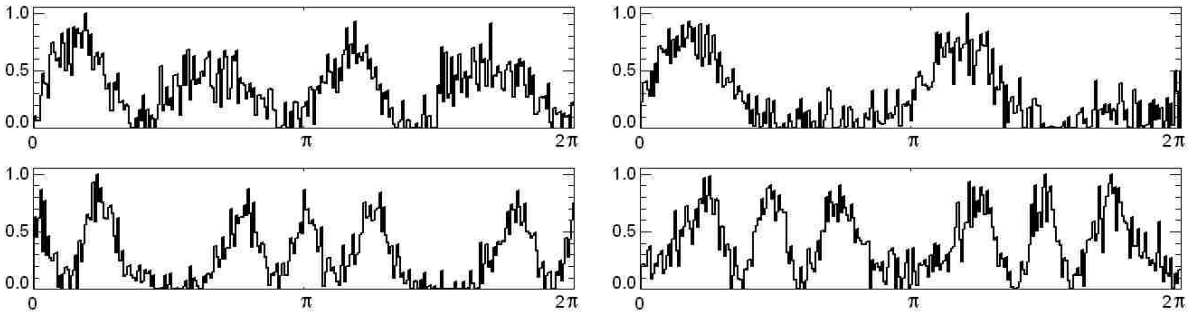


Figure 6.12: Normalized measured count profiles for middle and outer-ring detectors for one of the observations shown in Figure 6.13. The profiles differ due to the different x/y locations of the each detector.

Next, two 50 μCi ^{137}Cs sources are placed at the same distance at various separations to analyze the resolving power of the LaBRAT prototype. In the 662 keV photopeak, each source has a measured count rate of 3 cts/s amongst a background rate of 30 cts/s. Since multiple sources present source confusion and increased background, these exposures are longer than that seen in Fig. 6.11. Additionally, to demonstrate super-resolution, which is inherently less efficient, extended exposure lengths are required. Each of the following object scenes was imaged for 5 - 10 hours. Example count profiles for one of these exposures is shown in Fig. 6.12. The profiles are unique per detector location, and exhibit varying frequencies of modulation depending on the sources' apparent positions relative to the grid axis, as seen by each detector.

Fig. 6.13 shows the results with these two sources placed at varying angular separations, and probes the limit of the resolving power of the prototype. At 2° separation, the cross-correlation image clearly shows the two sources, and the NCAR reconstruction successfully removes the sidelobes. At 1° separation, roughly half the geometric resolution of the instrument, the cross-correlation image contains only a single central peak. An algorithm such as is NCAR is required to take advantage of the super-resolution capability of the RM's temporal modulation by deconvolving this raw image with the system response and resolving

two independent sources. At $35'$, over $3\times$ super-resolution, NCAR is again able to resolve each of these two sources.

When the sources are placed within $20'$, however, the reconstruction indicates the presence of two sources, but is unable to fully separate their peaks. A 2-dimensional gaussian function fit to the reconstruction displays a slightly wider x-dimension, $31'$, than its vertical dimension, $27'$; both widths, however, are larger than the $20'$ source separation. Thus, while the ability to achieve super-resolution has no theoretical limit, insufficient exposure times (to overcome the $1/\eta^2$ inefficiency described in Sec. 5.5) and systematic effects result in a practical limit.

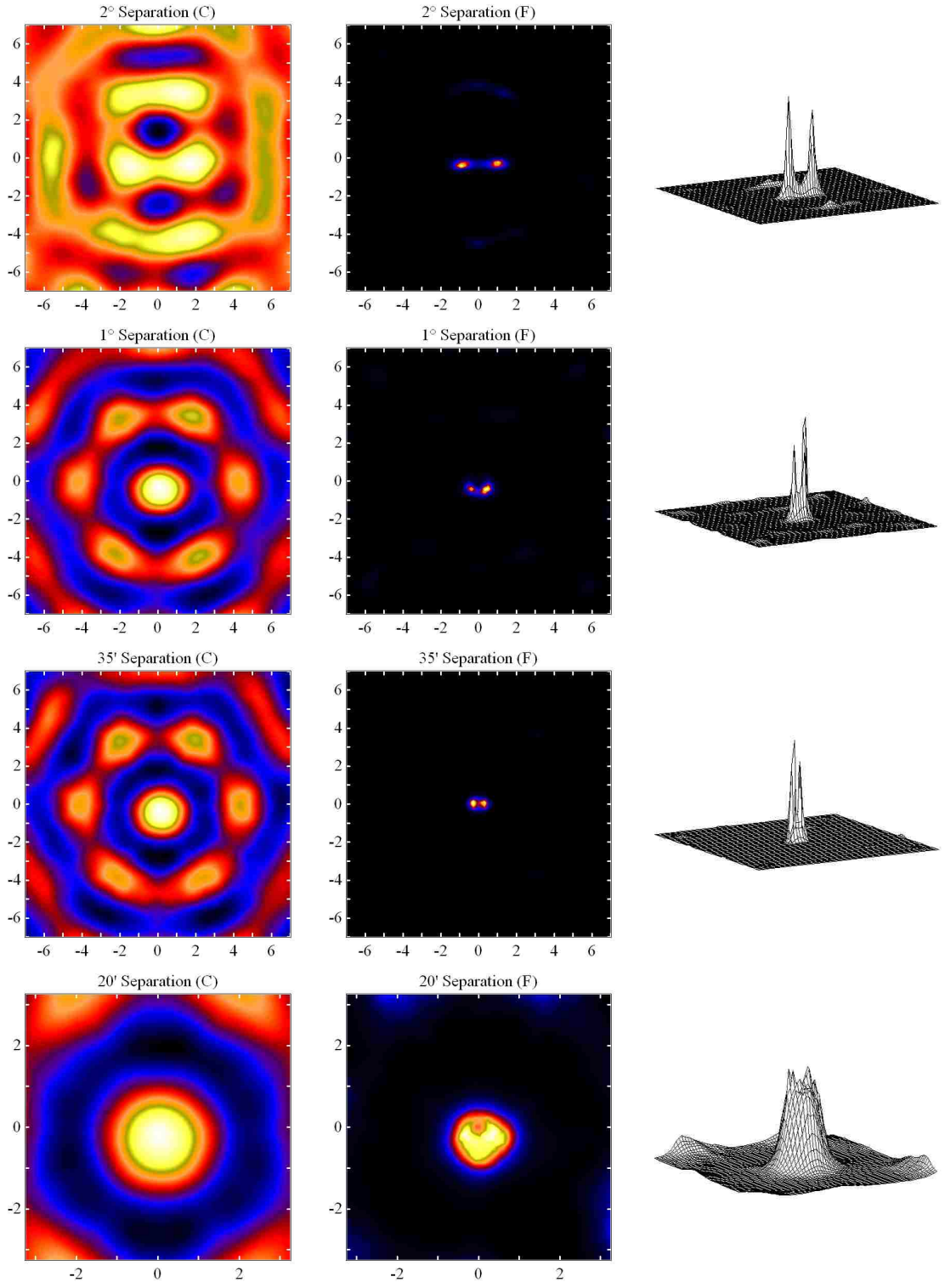


Figure 6.13: Contour plots of the cross-correlation (C), final reconstructed image (F), and surface mesh of final reconstruction for two ^{137}Cs point sources placed at varying angular separations. Top to bottom: 2°, 1°, 35′, and 20′ (zoomed in).

Chapter 7

High-Altitude Balloon-borne RM Mission Concept

7.1 Balloon Flight Measurement of $LaBr_3$ Background

7.1.1 Introduction

Satisfying the scientific requirements of a large-area RM astrophysics mission coupled with the demand for a low-cost instrument necessitate the use of a relatively inexpensive detector with good energy resolution capable of correctly identifying nuclear lines. A variety of scintillators and semiconductor devices are available with a wide range of costs and performance characteristics (see Table 7.1). Germanium semiconductors (Ge) provide the best resolution available (0.3% FWHM at 662 keV), but the operational temperature (< -150 °C) requires the use of liquid nitrogen for cooling. Alternatively, CZT semiconductor is often chosen for its excellent energy resolution ($< 3\%$), but at more than $5\times$ the cost of Ge, it is prohibitively expensive (e.g. as proposed for EXIST). Inorganic scintillators are typically used for their inherently low cost, but suffer from poorer resolution. Sodium iodide (NaI) offers resolution of 7% and cesium iodide (CsI) offers 10% (using a bialkalai PMT), and both cost a small fraction of the price of semiconductor detectors. These resolutions are sufficient to achieve the primary science goals of a black hole mission, but better resolution would enhance the secondary science goals (McConnell et al., 2004).

Table 7.1: Summary of Detector Characteristics (McConnell et al., 2004). Since 2004, CZT and $LaBr_3$ have both come down in cost, though they are still significantly more expensive than NaI and CsI.

	$LaBr_3$	NaI(Tl)	CsI(Tl)	CZT	Ge
Density (g/cm^3)	5.29	3.67	4.51	5.78	5.33
Light Output (ph/MeV)	63,000	39,000	52,000	N/A	N/A
$\Delta E/E$ (FWHM) @ 662 keV	$< 3\%$	7%	10%	$< 3\%$	0.3%
Peak λ (nm)	358-385	415	550	N/A	N/A
Fast Decay (ns)	25	230	1000	N/A	N/A
Hygroscopic	yes	yes	slightly	no	no
Cost ($ per cm^3)	30	2	4.50	3,000	500

Consequently, the use of $LaBr_3$ (manufactured by Saint-Gobain) has been suggested for CASTER, implemented in LaBRAT (Chap. 6), and proposed as the detection material for a high-altitude balloon-borne RM pathfinder mission for the BHFP (Sec 7.2). $LaBr_3$ offers a light yield of ∼60,000 photons/MeV (van Loef et al., 2001; Shah et al., 2003) and provides energy resolution of 2.6% FWHM, comparable to that of typical commercial CZT. The peak wavelength emission, 380 nm, is well-matched to the 25% peak quantum efficiency at 360 nm of borosilicate glass window Electron Tubes 9102B PMTs with bialkali photocathodes. Additionally, it features a decay time of < 25 ns, offering the potential to create a device with very good time resolution.

Compared to the other inorganic scintillators and semiconductor devices listed in Table 7.1, $LaBr_3$ is a relatively new detector technology and lacks the space-flight heritage of these other materials. A high-altitude mission to prove the spaceworthiness of $LaBr_3$ is especially important due to the presence of a relatively high internal background caused by two decay modes of naturally occuring ^{138}La (Rozsa et al., 2006). A 66% branching ratio by electron capture to ^{138}Ba produces a 1436 keV gamma ray and coincident 32 keV Ba X-ray. ^{138}La may also decay by beta emission to ^{138}Ce with a 34% yield, producing a 789 keV gamma and free electron.

In a self-counting background spectrum of $LaBr_3$ (Fig. 7.1), a beta-continuum is observed

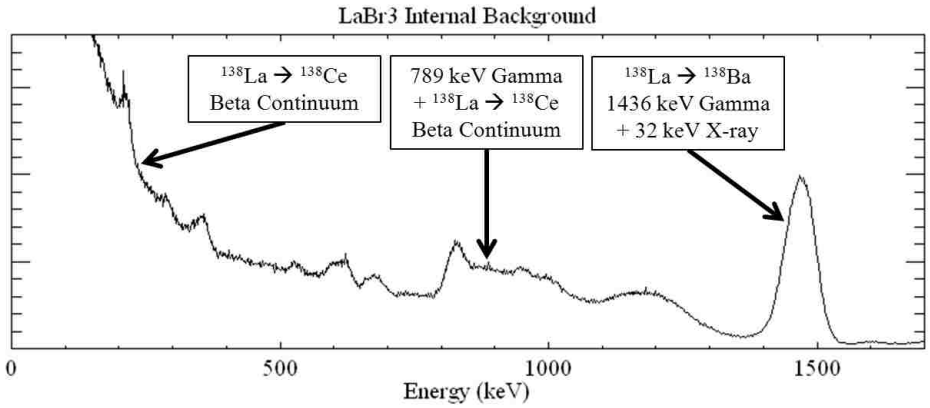


Figure 7.1: Self-counting background spectrum of $LaBr_3$ with features resulting from the two decay modes of naturally-occuring ^{138}La (Rozsa et al., 2006). Additional unlabeled spectral properties arise primarily from external background contributions

down to 255 keV from events in which the 789 keV gamma escapes detection. From 255 - 750 keV, a Compton continuum from both decay modes is visible. The 789 keV photopeak is present but has a wide high-energy shoulder extending to ∼1 MeV since the detection coincides with that of the beta. A prominent 1468 keV peak results from the electron capture decay mode, with its Compton edge and partial continuum visible down to ∼1 MeV.

The count rate in a 2.5 × 2.5 cm cylindrical $LaBr_3$ detector is measured to be 0.23 cts/sec in the 0 - 255 keV energy range, 0.07 cts/sec between 790 and 1000 keV, and 0.07 cts/sec in the 1468 keV gamma ray photopeak (Budden et al., 2008b). In order for the detector to be deemed spaceworthy, it must be demonstrated that background rates are sufficiently low, and that no induced activation of the crystal is observed in a near-space environment due to exposure to cosmic ray flux.

7.1.2 Mechanical

The Advanced Thin Ionization Calorimeter (ATIC; Guzik et al. 2004) is an LSU-developed balloon-borne observatory of cosmic ray nuclei and electrons spanning an energy range of 20 GeV – 200 TeV. The instrument consists of a silicon matrix for charge identification, a

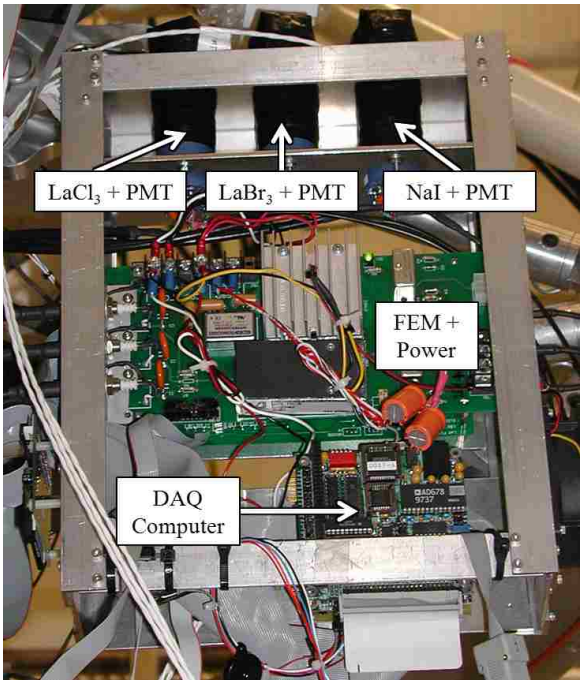


Figure 7.2: A photograph of the $LaBr_3$ experiment "piggy-backed" onto ATIC. Shown here is the experiment from a failed 2005 flight attempt where NaI and $LaCl_3$ were tested along with $LaBr_3$. (The succesful flight described in the text did not include the $LaCl_3$ detector.)

graphite target, scintillator hodoscopes, and a fully active calorimeter comprised of 400 2.5 $\times$ 2.5 $\times$ 25 cm Bismuth-Germanate (BGO) scintillating crystals viewed on-end by PMTs. ATIC has completed 3 successful flights in 2000, 2002, and 2007 from the NSF-NASA balloon facility at Williams Field in McMurdo Station, Antarctica. A 40 million ft^3 helium-filled balloon lifts the 1500 kg payload to approximately 120,000 ft, where circumpolar wind currents transport the instrument in a path around the continent.

The December 2007 ATIC flight provided an opportunity to test the feasibility of $LaBr_3$ use in a near-space environment by mounting a "piggy-back" payload onto the ATIC frame (Fig. 7.2). The experiment was comprised of one $LaBr_3$ and one NaI crystal, each a 2.5 cm diameter $\times$ 2.5 cm long cylinder viewed by a 3.8 cm Electron Tubes 9102B PMT. The anode output signals of each PMT were split; one end fed into an OR'd trigger, and the other end fed into the DAQ for digitization and recording to hard disk.

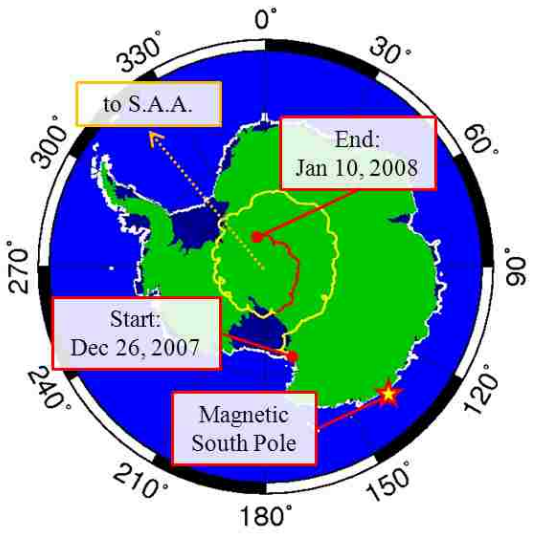


Figure 7.3: Trajectory of the 2007 – 2008 ATIC balloon flight. The star designates the location of the magnetic South Pole, and the arrow indicates the direction of the South Atlantic Anomaly (Budden et al., 2008b).

7.1.3 Results

ATIC launched on December 26, 2007 and followed a circumpolar trajectory (Fig. 7.3). On January 10, 2008 ATIC was brought down after an unexpected loss of pressure in the ATIC pressure vessel forced an end to the mission. ATIC remained aloft at an average of ∼120,000 ft, with variations corresponding to temperature swings from day/night cycles (Fig. 7.4). The $LaBr_3$ piggy-back mission successfully collected data for all 16 days of the flight.

Temperatures rose steadily during the flight from 25 °C to 35 °C, causing a 10% shift in gain. A 511 keV photopeak is present in both the NaI and $LaBr_3$ spectra, and is likely due to positron annihilation from pair production in the ATIC calorimeter. The 511 and 1468 keV photopeaks were utilized in the analysis to provide a gain correction of the $LaBr_3$ data. In the energy range of 700 - 1500 keV, the count rate on the ground was 2.5 cts/sec, while at float altitude, the rate rose to 7.0 cts/sec. While it rose slightly as the balloon neared the South Atlantic anomaly, the $LaBr_3$ count rates remained constant for most of the flight at about 3× the normal level. The total rates during the balloon flight for both $LaBr_3$ and NaI above 200 keV are plotted in Figure 7.5. $LaBr_3$ is found to have approximately 1.5× the background as NaI at float, with 6.2 cts/s/cm^3. A background-subtracted $LaBr_3$ spectrum is given in Fig. 7.6 for 100 – 700 keV; the 511 keV emission line is clearly visible.

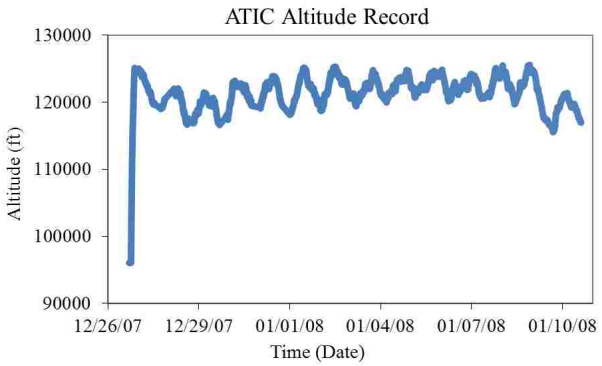


Figure 7.4: Plot of ATIC's altitude (in ft) during the 10-day flight (Budden et al., 2008b).

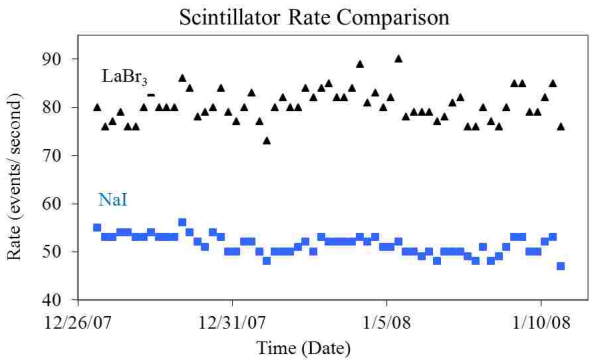


Figure 7.5: Count rates of a 2.5 cm diameter $\times$ 2.5 cm thick $LaBr_3$ and NaI crystals during the ATIC mission (Budden et al., 2008b).

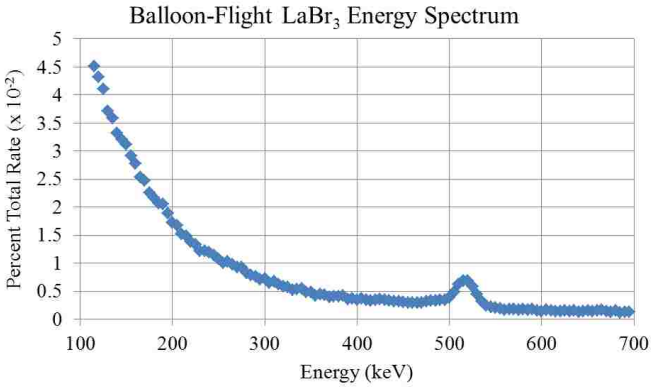


Figure 7.6: The $LaBr_3$ spectrum over the energy-range of interest from the ATIC flight shows the 511 keV bump and a large contribution from internal background due to naturally-occuring ^{138}La at lower energies.

While at least one previous balloon mission launched from New Mexico has seen the flight of $LaBr_3$ to a near-space altitude (McConnell, 2008), no experiment has flown both $LaBr_3$ and a comparable inorganic scintillator with proven spaceworthiness. In this experiment, $LaBr_3$ and NaI scintillators of equal size were flown to an altitude of 120,000 ft and were subjected to increased background due to "backsplash" from the ATIC calorimeter. $LaBr_3$ measured a higher background rate (by 50%) than NaI at float altitude, but did not suffer from induced activation by cosmic flux, verifying its ability to perform in a space environment.

7.2 High Altitude Rotating Modulator for Energetic Radiation Imaging

7.2.1 Introduction

The High Altitude Rotating Modulator for Energetic Radiation Imaging (HARMEnI) is the culmination of the prototypical work on the RM. HARMEnI will demonstrate the results and concepts described regarding the RM and its analysis including: the temporal modulation and imaging mechanism; the robustness of the advanced characteristic profiles, particularly for wide FOV and high energy photons; the validity of the NCAR imaging algorithm and its ability to perform well in the presence of strong background; reinforcement of $LaBr_3$'s ability to perform well in a near-space environment; the RM sensitivity characteristics; and the super-resolution capability of the RM by resolving nearby sources. The results of a successful HARMEnI mission might provide evidence of high-fidelity performance comparable to a coded aperture telescope, at reduced cost and complexity. This demonstration would make a convincing argument for the use of the RM in a full-scale astrophysics mission.

7.2.2 Mechanical and Electronics

The HARMEnI design (Fig. 7.7) is an extrapolation of the LaBRAT geometry. The fully-coded FOV is increased to 20° by inclusion of an additional two lead bars in the mask. The mask sits on a custom gear-toothed turn-table driven by a stepper motor for precise computer control over its motion and orientation. It is suspended 1.2 m above the array, offering the the same geometric resolution (and ultimately, resolving power) as LaBRAT.

The sensitivity is also increased by an additional concentric ring of scintillators; in total, HARMEnI consists of 37 $LaBr_3$ crystals packed into a tight configuration to minimize dead area; the minimum distance between detectors is 0.5 cm. The inclusion of additional detectors further smooths the sensitivity in the object scene by smearing the regions of un-

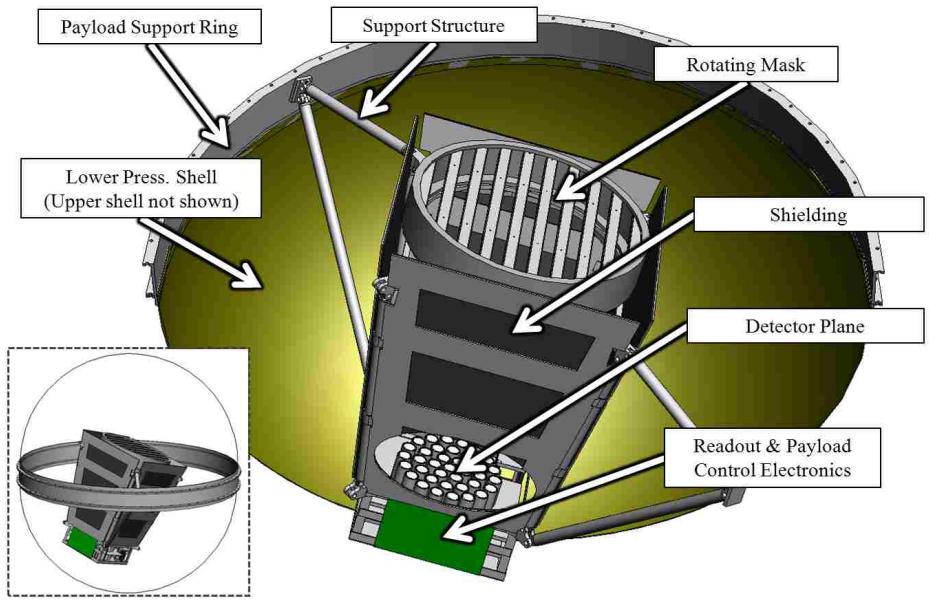


Figure 7.7: CAD figure of the HARMEnI concept. The balloon pressure shell and support ring have been cut away to show the instrument detail. Inset photo: The entire HARMEnI payload with the pressure shells made fully transparent.

modulated sky due to each detector. Dithering of the RM orientation about the vertical due to inevitable instrument swinging motions will result in additional uniformity across the FOV.

A major difference between LaBRAT and HARMEnI is the environment to which they are exposed. HARMEnI will encounter a significant amount of cosmic diffuse flux and Earth albedo, reducing the SNR of an observation. Most of this background is reduced by four passive graded 5 mm-thick lead-tin-copper shields, which extend from the detection plane to the grid. Aluminum plates support and connect the shields to the rest of the HARMEnI frame. The total instrument weight is $\sim$1000 kg.

HARMEnI utilizes the same electronic design as LaBRAT for the PMT base voltage dividers (Sec. 6.3.1) and DAQ (Sec. 6.3.3). Since the payload will be enclosed in the ATIC pressure shell, no further modification (e.g. potting) of the electronic components is necessary. The current 32-channel DAQ on LaBRAT is easily expanded to 64 channels for HARMEnI. The current LaBRAT computer is considered the Detector Control Unit

(DCU). Two additional PC/104 stacks contain the Flight Control Unit (FCU) for balloon craft communications, and the Data Archive Unit (DAU) for recording the DCU data output and FCU housekeeping records to a solid-state hard drive.

7.2.3 Balloon Payload

The HARMEnI telescope will launch in the same external balloon payload structure as that of ATIC (Guzik et al., 2004). This structure includes a Kevlar fabric pressure vessel (under ∼8 psi during flight) and pressure vessel ring. Four aluminum internal support rods attach HARMEnI to the ring and fix its elevation angle at 20° from zenith to provide optimal observation of its primary targets during the flight.

Azimuth is controlled from ground by a CSBF-supplied rotator (Solar Pointing System), which should allow the telescope to point to within 2 - 5° of the intended target. Given the wide FOV, this error is sufficiently low to allow for continuous observation of a particular source or nearby sources. Knowledge of the attitude, however, is critically important, especially to achieve and demonstrate the capability of super-resolution; a star tracker (Percival et al., 2007, 2008) provides attitude information to within 30′.

7.2.4 Flight Mission

The proposed HARMEnI mission is intended to image the Crab Nebula (RA 05h34m, Dec 22°00') and surrounding field during a one-day flight from Ft. Sumner, New Mexico. The Crab measures approximately 6′ x 4′ (at optical wavelengths) and so can be considered as a point source given HARMEnI's angular resolution. From the Ft. Sumner latitude of 34.5°, the Crab will rise to a maximum elevation angle of 12.5° from zenith. The trajectory of the Crab in the sky from rise to set will provide four hours of fully-coded observation. A secondary observational target is Cygnus X-1 (RA 19h58m, Dec 35°12'), which will be visible for a total of three hours.

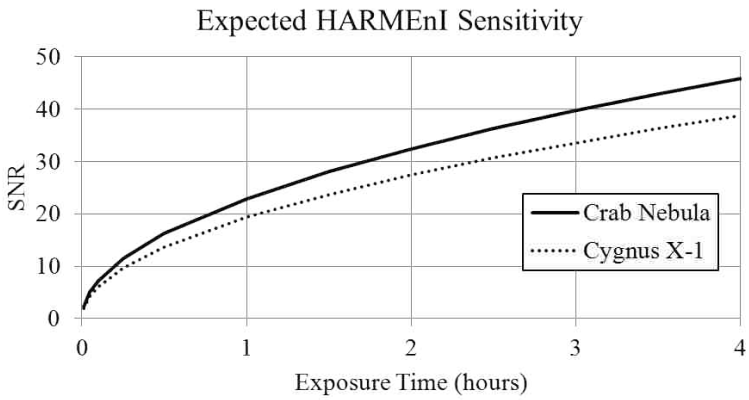


Figure 7.8: The significance of an observation by HARMEnI of the Crab Nebula will exceed 10σ in 11 minutes and Cygnus X-1 in 16 minutes. If observed for the maximum possible exposure time (4 hours for the Crab and 3 hours for Cygnus X-1), SNRs of 46 and 34 may be achieved.

7.2.5 Expected Performance

Laboratory experiments measure the rate per cubic element of internal background contribution from ^{138}La in a $LaBr_3$ crystal. In the energy range of 30 - 700 keV, this value is $\sim$0.6 cts/s/cm^3. An extrapolation to the detector volume of HARMEnI (1071 cm^3) determines an expected internal background rate of 679 cts/s.

These same measurements are normalized to the data from the $LaBr_3$ high-altitude flight discussed in Sec. 7.1 and subtracted to determine an energy-dependent cosmic diffuse flux rate. For the same energy range, the background contribution from external sources for HARMEnI is 5.6×10^3 cts/s. The inclusion of a 5 mm-thick lead shield reduces this contribution considerably, particularly at lower energies, to 364 cts/s.

Given the mask transmission ratio (50%) and sensitive detector area, count rates for the Crab are calculated to be 41 cts/s and Cygnus X-1, 35 cts/s. In the presence of the total 10^3 cts/s background contribution, a 10-sigma imaging result may be achieved in 11 minutes for the Crab and in 16 minutes for Cygnus X-1. If the observations are made for the full four or three hours exposure time, SNRs of 46 and 34 may be obtained (Fig. 7.8).

The software described in Sec. 6.4 simulates the HARMEnI data using these calculated rates. In Fig. 7.9, the NCAR algorithm deconvolves the data to reconstruct images of the

Crab from exposures lasting 3 minutes, 11 minutes and 4 hours. The confusion between the source location and the PSF sidelobes is apparent for the shorter exposures, but the Crab is still reconstructed well even for exposure times as short as 3 minutes.

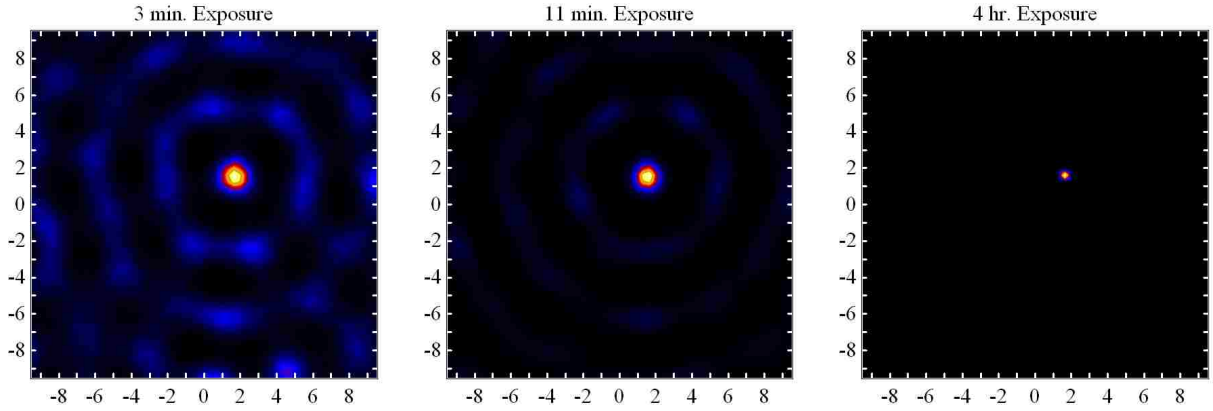


Figure 7.9: Simulation Results for the HARMEnI observation of the Crab Nebula over an exposure time of 3 minutes (5σ), 11 minutes (10σ), and 4 hours ($5\ \sigma$).

Chapter 8

Discussion

Motivated by the need for a high-sensitivity x-ray/gamma-ray imager for astrophysical and national security applications, the rotating modulator has been investigated as an alternative to the costly and complex coded aperture. The RM provides comparable sensitivity and angular resolution at high energies and for wide FOVs, with approximately 1/4 the number of readout channels. Additionally, the RM is able to provide super-resolution and may take advantage of insensitivity to background contributions with a proper choice of decoding matrix during reconstruction. While a laboratory prototype has been developed and a 1-day balloon flight proposed, the objectives for these experiments have been to demonstrate the RM performance capability and not to achieve specific scientific objectives.

Here, we instead propose the designs of two RM imagers: one as a full-scale balloon-borne astrophysics mission to survey black holes in the universe and the other as a stand-off radiation detection system to detect illicit nuclear materials. The geometry of the two instruments is based on the scientific objectives outlined in Chap. 1 and is tweaked using the advanced characteristic formula from Chap. 3 and the sensitivity calculations from Chap. 5. Additional enhancements to the RM imaging technique are then described that would make it more suitable for a wider range of applications and improve its results.

8.1 Stand-off Radiation Detection System

The United States Department of Energy's DNDO outlines the short term requirements for a stand-off radiation detection system, including the ability to detect 15 photons per second in a 662 keV photopeak from a 1 mC ^{137}Cs source at a distance of 100 m. ^{137}Cs has a 94.4% branching ratio and 90.1% branching fraction to produce a 662 keV photon, and so will produce

$$0.001\text{Ci} \times 3.7 \times 10^{10} \frac{\text{decays}}{\text{Ci}} \times 0.944 \times 0.901 = 3.1 \times 10^{7} \text{ ph/s}. \tag{8.1}$$

The density of air is 0.001225 g/cm^3 and its mass attenuation coefficient is 0.077 cm^2/g. Thus, at 100 m, the percentage of photons that are not attenuated is

$$\exp\left[-0.001225 \frac{\text{g}}{\text{cm}^3} \times 0.077 \frac{\text{cm}^2}{\text{g}} \times 10^4 \text{cm}\right] = 38.9\%. \tag{8.2}$$

The flux at this distance due to the source activity is

$$38.9\% \times \frac{3.1 \times 10^7 \frac{\text{ph}}{\text{s}}}{4\pi \times (10^4 \text{cm})^2} = 9.75 \times 10^{-3} \text{ ph/s/cm}^2. \tag{8.3}$$

$LaBr_3$ detectors with diameter 3.8 cm and thickness 5 cm are used, and so 85% of incident 662 keV photons will interact in the crystal (Rozsa et al., 2006). For an event to occur in the photopeak, however, the 662 keV photon may Compton scatter in the crystal so long as it ultimately interacts by the photoelectric effect and therefore does not escape the detector. Monte Carlo simulations are performed by Rozsa & Menge (2004) to determine the peak-to-total ratio of 662 keV photons in various configurations of $LaBr_3$ crystals. For a 3.8×5 cm detector, the peak-to-total ratio is found from Rozsa & Menge (2004) to be $\sim$0.45. Given this ratio and the attenuation efficiency, the overall photopeak detection efficiency is

$0.85 \times 0.45 = 0.41$, and so the flux of photopeak detections at 100 m is

$$0.41 \times 9.75 \times 10^{-3} \frac{\text{ph}}{\text{s} \cdot \text{cm}^2} = 3.73 \times 10^{-3} \text{ ph/s/cm}^2. \tag{8.4}$$

The area required to enable a photon detection rate of 15 cts/s is then

$$\frac{15 \text{s}^{-1}}{3.73 \times 10^{-3} \text{s}^{-1} \text{cm}^{-2}} = 4021 \text{ cm}^2. \tag{8.5}$$

If the mask is separated from the detection plane by 30 cm, then to achieve a 115° fully-coded FOV, the mask diameter must be 195 cm, requiring 26 lead slats. The mask should be at least 1.8 cm thick to attenuate 90% of the incident 662 keV photons, and so weighs a total of 213 kg. Sensitivities of the instrument geometry described here are derived for various slat width to pitch ratios. Due to the wide FOV and thick mask, sensitivity due to encoding power is maximum for a ratio of 0.25. Thus, slats should be 1.9 cm wide and spaced 5.7 cm apart. This maximization of sensitivity implies that on average, the mask has a 50% transmission despite it being only 25% opaque when viewed straight on. Thus, the area of the detection plane must be doubled to 8042 cm^2 to account for this attenuation, requiring 705 3.8 cm diameter detectors.

Due to the wide FOV, angular resolution will vary significantly across the scene. For sources directly ahead, the angular resolution will be the worst, at $\tan^{-1}(5.7/30) = 10.8°$. At the edge of the field of view, where apparent slit widths are smaller as seen by the detection plane, the resolution enhances to 5.6°. These resolutions are 1.5 - 3× better than those specified by for the DOE/DNDO stand-off radiation detection system.

The RM instrument described here has 705 readout channels, weighs less than 400 kg, and fits within a 2 × 2 × 1 m volume. It is capable of detecting photopeak events from 662 keV photons emitted by a 1 mC ^{137}Cs source at a distance of 100 m at a rate of 15 cts/s over a 115° field of view and with 5 - 10° angular resolution. A coded aperture with these qualities would require at least $\sim 4 \times 705 = 2820$ individual detectors and would suffer

from reconstruction artifacts arising from the finite thickness of the grid (especially at wide angles). While the coded aperture would have better overall sensitivity, the RM sensitivity is comparable for sources away from the center of the FOV. Furthermore, the RM has the advantage that better angular resolution (i.e., super-resolution) can be achieved over time with a prolonged exposure to the object scene.

8.2 High-Altitude Balloon-Borne Campaign

A high-altitude balloon campaign based on the HARMEnI design (Chap. 7) would be capable of accomplishing some of the scientific objectives of a black hole survey mission. If a mask is constrained to be 2.5 m across, then

$$\frac{2.5\text{m}}{2 \times 0.038\text{m}} \approx 32 \tag{8.6}$$

slats are needed to match the 3.8 cm-diameter detectors. At 1.8 cm thickness, the total mass is 500 kg. If the mask is separated from the detection plane by 218 cm, then the geometric resolution is defined at

$$\sin^{-1}\left(\frac{3.8\text{cm}}{218\text{cm}}\right) \sim 1^{\circ}. \tag{8.7}$$

Since super-resolution has been demonstrated with LaBRAT on the order of 3× the geometric limit, $20'$ resolving power should be achievable for bright sources. Unlike as was done for the stand-off imager for security applications, the slat width to pitch ratio is fixed here at 0.5. The reason for this is that reducing this value will worsen the angular resolution. Additionally, the FOV for this instrument will be much less than that of the security imager, and so the sensitivity will be maximized for a ratio closer to 0.5.

Under these mask constraints, a larger detector area narrows the fully-coded FOV, but increases sensitivity. By defining a 20° FOV, a total detection plane diameter of 173 cm is possible, with area 2.4 m^2, approximately 1/3 that of the proposed CASTER (McConnell

et al., 2004, 2005) and EXIST (Grindlay et al., 2001, 2003) designs. Given this diameter, a partially-coded FOV extends out to

$$2 \times \tan^{-1}\left(\frac{173\text{cm} + 250\text{cm}}{2 \times 218\text{cm}}\right) \approx 88^{\circ}. \tag{8.8}$$

Such a configuration requires 2062 3.8 cm-diameter detectors. At 5 cm thick, the total mass is 623 kg. Shields which extend from the detection plane to the grid and are 5 mm thick would add an additional 1063 kg to the weight, resulting in a combined weight (detectors, mask, and shields) of just over 2000 kg. By allowing another 1000 kg for structure and electronics, the mass is close to the maximum possible for a balloon payload.

To examine the sensitivity with this instrument design, we consider a 5-year campaign of 2 week high-altitude balloon flights, totalling 6.05×10^6 seconds. Given a 20° FOV, any one source may be visible for $\sim$ 4 hours per day, and so total exposure time of a single source would be 1.01×10^6 s. The results of the high-altitude flight of a $LaBr_3$ crystal (Sec. 7.1) are used to determine Crab and background count rates. A 5σ imaging sensitivity that can be achieved is 1.4 mCrab in the 30 - 150 keV energy range and 13.2 mCrab in the 150 - 600 keV energy range. For a coded aperture technique, the sensitivity would be slightly better, but 8.2k readout channels would be required. Additionally, resolving power would be fixed at 1°, instead of being able to resolve down to $20'$ or smaller for extended exposure times.

8.3 Additional Enhancements to RM Design

A benefit of a temporal modulation technique such as the RM is its ability to "automatically" remove background contributions with the proper choice of decoding matrix (described in Sec. 4.2). Since NCAR (and many of the techniques described in Sec. 4.5.3) requires a positive-valued PSF, the background term is not removed. Instead, a background exposure must be used to manually remove the constant offset from the observed count profiles. This process reduces source exposure time and is difficult in a balloon-borne or satellite

mission. A modification to the NCAR routine to handle negative-valued PSFs (e.g. through regularization) would enhance the RM technique and provide an additional advantage over the coded aperture, which due to its spatial nonuniformity in background contributions, must inherently perform a background exposure.

The removal of the constant background offset may have a secondary beneficial effect. Accorsi & Lanza (2001) suggests the use of an anti-mask with a coded aperture imager to remove near-field artifacts. Near-field artifacts will similarly arise with the RM technique, since the assumption of sources being at infinite distance causes nearby sources to reconstruct poorly; i.e., they are "out of focus." If the axis of an RM grid is aligned with the edge of a slat (as opposed to the center of a slat or slit), then the second half of the rotational period will be exactly the inverse of the first half, mimicking an anti-mask. If this property can be used to remove near-field artifacts in coded aperture imaging, it is conceivable that this technique may also be used with the RM. Such a property would make the RM even better suited to security applications, where the instrument may be within close proximity to the source.

Other than lower sensitivity, a disadvantage of the RM when compared to the coded aperture is its requirement of mechanical rotation of the grid. For a satellite-based RM telescope, mechanical rotation would not, however, be required as with a laboratory or balloon-borne experiment. Instead, the entire instrument would be rotated as is commonly done with satellite-based or rocket-borne RMCs (e.g. Hurford et al. 2002). Unlike the RMC, however, for the RM this whole-instrument rotation alters the system response since each detector location now moves with respect to the object scene. An investigation should therefore be made into the effect on coding power and sensitivity for this application.

For both astrophysical and security applications, a monolithic detector with an Anger array (i.e., gamma camera) might improve the sensitivity variations in the FOV as seen in Fig. 6.3. Additionally, this configuration would allow for any slat/slit widths (or combinations thereof) to sample a large range of the frequency spectrum. Such a design would, however, be

subject to spatial resolution limitations unlike the individual non-imaging detectors typically used in an RM. An investigation should be made to weigh the advantages and disadvantages of such a system.

Further improvements to the imaging technique could be made to maximize the signal-to-noise ratio of an observation. In the current system, all profiles are weighted equally, despite the fact that profiles with fewer modulations provide less coding power. Shikhaliev (2010) presents a generalized weighting method in a study on medical computed tomography which biases the reconstructed image towards specific "sub-images" of the data set in order to maximize the SNR of the final reconstruction. This same technique could be used with RM data, where each sub-image is either from a particular detector, or from a particular profile in the system response.

Chapter 9

Conclusion

The Black Hole Finder Probe, part of NASA's Beyond Einstein program, was proposed as a satellite mission to survey the universe for black holes over a wide range of accretion rates and sizes. To accomplish the primary scientific goals, sensitivity is required significantly greater than previous all-sky surveys in the energy range of 10 - 600 keV, with 3 - 5 minutes of arc angular resolution and a wide field of view capable of observing the entire sky once per orbit. The suggested instrument designs (EXIST and CASTER) relying on coded mask imaging techniques (McConnell et al., 2004, 2005; Grindlay et al., 2001, 2003) were deemed too costly and complex by the Beyond Einstein Program Assessment Committee (National Research Council, 2007). This judgement motivates the search for an alternate hard x-ray/gamma-ray imaging technique that can meet the sensitivity and resolution requirements with fewer electronic channels and lower cost.

Rotating modulation uses temporal and spatial encoding in concert to measure the object scene. A grid of opaque slats spaced equidistance apart is suspended above an array of non-imaging detectors. The grid rotates, modulating incident photons with a pattern specific to the RM geometry and source location. Each detector records a time history of counts, which is used to construct an image of the object scene convolved with the instrument response. The image contains sidelobes due to an extended point-spread-function. Deconvolution of

this raw image with the pre-computed system matrix may remove sidelobes and suppress noise to reconstruct the object scene.

A standard system matrix is described by the inventor of the RM concept (Durouchoux et al., 1983), but is simplistic in its assumptions about the grid geometry and photon attenuation. At high energies and wide zenith angle, a more realistic characteristic formula is necessary to describe shadow lengthening, non-uniform attenuation effects, and simultaneous slat shadowing. An advanced system response is derived and presented to satisfy this goal, and is shown to outperform the standard formula in these regimes.

The extended sidelobes inherent to the RM response necessitate a reconstruction algorithm that correctly deconvolves the raw image with little source confusion. Since the RM works primarily in the temporal domain, super-resolution is achievable, and so a reconstruction algorithm must also be able to demonstrate this capability. Algebraic techniques employed with positivity constraints as a form of non-linear regularization satisfy these requirements, but are subject to spurious fluctuations arising from noise. To compensate, the NCAR technique has been devised, which introduces random noise in the iteration based on Poisson uncertainty in the data. Once the reconstructed image has reached best convergence to the data, images are averaged to smooth over noise components. NCAR is shown to perform better than many common deconvolution techniques used in the field of astronomy.

One of the most significant advantages of the coded mask imager is its high sensitivity. Due to the discrete and symmetric mask pattern, the coded aperture has zero sidelobes. Measurements from a source location are attributed entirely to that location, maximizing the SNR of the observation. Temporal techniques such as the RM have continuous system formulas, which inherently have a lower standard deviation. The RM, in particular, is shown to have a sensitivity approximately 62% that of a comparable coded aperture, assuming perfect detection efficiency. At equal geometric resolutions, however, the coded aperture requires detector pixels approximately half that of the RM. At wide photon incidence angles (such as would be seen with a survey mission) and high photon energies, the detection

efficiency drops more rapidly for the coded aperture than for the RM. In these regimes, the RM is shown to be comparable in sensitivity to the coded aperture.

A laboratory prototype RM, LaBRAT, has been developed to demonstrate the RM mechanism and verify the derived characteristic formula, image reconstruction algorithm, and sensitivity calculations. The nineteen-channel $LaBr_3$-based prototype features a custom DAQ and readout software. Simulations based on the prototype geometry demonstrate the ability to deconvolve the raw data to resolve complex object scene distributions despite the extended sidelobes of the RM response. Experimental data verify the ability of the NCAR algorithm to achieve super-resolution and compensate for noise. LaBRAT serves as a proof-of-concept for a high-altitude balloon flight, which extends the number of detectors and field of view.

A proposed balloon mission, HARMEnI, is a 37-channel $LaBr_3$-based RM with 20° fully-coded FOV and 1.9° geometric angular resolution. A 2007 high-altitude flight of $LaBr_3$ – which suffers from internal background due to the decay of ^{138}La – and commonly-used NaI directly compared the latter, which has a large space heritage, to the former, which had only seen one previous high-altitude mission. In this high-energy regime, $LaBr_3$ demonstrated only $\sim$1.5$\times$ the background contribution as NaI and suffered from no induced activation. An additional challenge in the design of the mission is in the existence of cosmic diffuse flux and Earth albedo in this near-space environment. Shields extend from the detection plane to the mask to reduce this contribution by $\sim$90%. HARMEnI simulations show a 10σ result is achievable with an observation of the Crab Nebula in 11 minutes and 16 minutes for Cygnus X-1. The successful HARMEnI flight will complement the theoretical work and results of LaBRAT to serve as a proof-of-concept for the RM candidacy as a full-scale astrophysics mission for performing an all-sky black hole census.

Based on the advanced system response derived in Chap. 3 and sensitivity calculations described in Chap. 5, a full-scale balloon-borne astrophysics campaign is proposed to satisfy some of the scientific objectives of a black hole survey. For national security applications, a low-cost alternative to coded aperture designs is suggested using the RM technique to detect

shielded illicit nuclear materials at long distances. Improvements to both designs and the RM in general are suggested for further investigation to improve the imaging technique and provide additional advantages over commonly-used coded aperture designs.

Bibliography

2001, Technology R&D for Arms Control, Tech. rep., DOE/NNSA

Ables, J. G. 1968, Proceedings of the Astronomical Society of Australia, 1, 172

Accorsi, R., & Lanza, R. C. 2001, Applied Optics, 40, 4697

Barrett, H., & Myers, K. 2003, Foundations of Image Science, 1st edn. (Wiley-Interscience)

Barrett, R., et al. 1994, Templates for the Solutions of Linear Systems: Building Blocks for Iterative Methods, 2nd edn. (Philadelphia, PA: Society for Industrial Mathematics)

Boehm, C., Hooper, D., Silk, J., Casse, M., & Paul, J. 2004, Physical Review Letters, 92, 101301

Boggs, S. E. 2006, New Astronomy Reviews, 50, 604

Budden, B., Budden, M., Case, G. L., & Cherry, M. L. 2011, Astrophysics and Space Science, Accepted for Publication, DOI: 10.1007/s10509-011-0712-z

Budden, B., Case, G. L., & Cherry, M. L. 2008a, in Nuclear Science Symposium Conference Record, 2008. NSS '08. IEEE, 2976–2980

Budden, B., Case, G. L., & Cherry, M. L. 2009in (SPIE), 74490G

Budden, B., Case, G. L., Cherry, M. L., Hopson, R., Isbert, J., Sankaran, R., Smith, D., & Stewart, M. 2008b, in Nuclear Science Symposium Conference Record, 2008. NSS '08. IEEE, 3347 –3350

Caroli, E., Stephen, J. B., di Cocco, G., Natalucci, L., & Spizzichino, A. 1987, Space Science Reviews, 45, 349

Carpenter, G. F., Skinner, G. K., Wilson, A. M., & Willmore, A. P. 1976, Nature, 262, 473

Cassé, M., Fayet, P., Schanne, S., Cordier, B., & Paul, J. 2004, in ESA Special Publication, Vol. 552, 5th INTEGRAL Workshop on the INTEGRAL Universe, ed. V. Schoenfelder, G. Lichti, & C. Winkler, 65

Chen, Y., Li, T. P., & Wu, M. 1998, Astronomy and Astrophysics Supplement, 128, 363

Cherry, M. L., Case, G. L., & Welch, C. E. 2006, in Astroparticle, Particle and Space Physics, Detectors and Medical Physics Applications, ed. M. Barone, E. Borchi, A. Gaddi, C. Leroy, L. Price, P.-G. Rancoita, & R. Ruchti, 917–921

Cornwell, T. J., & Evans, K. F. 1985, Astronomy & Astrophysics, 143, 77

Crannell, C. J., Hurford, G. J., Orwig, L. E., & Prince, T. A. 1986, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 571, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 142–148

Cruise, A. M., & Willmore, A. P. 1975, Royal Astronomy Society, Monthly Notices, 170, 165

Dadurkevicius, V., & Ralys, D. A. 1985, Astrophysics and Space Science, 113, 233

Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, Journal of the Royal Statistical Society. Series B (Methodological), 39, 1

Dicke, R. H. 1968, Astrophysical Journal, 153, L101

Doxsey, R., et al. 1976, The Astrophysical Journal, 203, L9

Durouchoux, P., Hudson, H., Matteson, J., Hurford, G., Hurley, K., & Orsal, E. 1983, Astronomy and Astrophysics, 120, 150

Fenimore, E. E., & Cannon, T. M. 1978, Applied Optics, 17, 337

Ferguson, C. D., K. T. P. J. 2003, CNS Occasional Papers, 11

Ferrer, F., & Vachaspati, T. 2005, Physical Review Letters, 95, 261302

Fishman, G. J., et al. 1989, in Developments in Observations and Theory for Solar Cycle 22, ed. R. M. Winglee & B. R. Dennis, 96–106

Frieden, B. R. 1972, Journal of the Optical Society of America (1917-1983), 62, 511

Gaither, III, C. C., Schmahl, E. J., Crannell, C. J., Dennis, B. R., Lang, F. L., Orwig, L. E., Hartman, C. N., & Hurford, G. J. 1996, Applied Optics, 35, 6714

Gehrels, N., et al. 2004, Astrophysical Journal, 611, 1005

Gordon, R., Bender, R., & Herman, G. T. 1970, Journal of Theoretical Biology, 29, 471

Grindlay, J., et al. 2001, in American Institute of Physics Conference Series, Vol. 587, Gamma 2001: Gamma-Ray Astrophysics, ed. S. Ritz, N. Gehrels, & C. R. Shrader, 899–908

Grindlay, J. E., Craig, W. W., Gehrels, N. A., Harrison, F. A., & Hong, J. 2003, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 4851, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. J. E. Truemper & H. D. Tananbaum, 331–344

Grove, J. E., Johnson, W. N., Kroeger, R. A., McNaron-Brown, K., Skibo, J. G., & Phlips, B. F. 1998, Astrophysical Journal, 500, 899

Guzik, T. G., et al. 2004, Advances in Space Research, 33, 1763

Högbom, J. A. 1974, Astronomy and Astrophysics Supplement, 15, 417

Hurford, G. J., et al. 2002, Solar Physics, 210, 61

in 't Zand, J. 2009, http://astrophysics.gsfc.nasa.gov/cai/

Intl. Atomic Energy Agency. 2000, Categorization of Radiation Sources, Tech. Rep. GOV/2000/34-GC(44)/7

Jaynes, E. T. 1957, Physical Review, 106, 620

Kaczmarz, S. 1937, Bull. Acad. Polon. Sci. Lett., A35, 355

Knödlseder, J., et al. 2005, Astronomy and Astrophysics, 441, 513

Li, T.-P., & Wu, M. 1994, Astrophysics and Space Science, 215, 213

Lund, N. 1981, Astrophysics and Space Science, 75, 145

Matteson, J. L. 1978, in AIAA 16th Aerospace Sciences Meeting

McConnell, M. L. 2008, personal communication

McConnell, M. L., et al. 2000, Astrophysical Journal, 543, 928

—. 2002, Astrophysical Journal, 572, 984

McConnell, M. L., et al. 2004, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 5488, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. G. Hasinger & M. J. L. Turner, 944–955

McConnell, M. L., et al. 2005, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 5898, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. O. H. W. Siegmund, 1–12

Mertz, L. 1967, in Modern Optics, ed. J. Fox, 787

Mertz, L. 1976, Astronomy and Astrophysics Supplement, 45, 383

Moon, T. K. 1996, IEEE Signal Processing Magazine, 13, 47

NASA. 2006, Science Program for NASA's Astronomy and Physics Division, http://science.nasa.gov/media/medialibrary/2010/03/31/AstrophysicsScienceProgram.pdf

National Research Council. 2007, NASA's Beyond Einstein Program: An Architecture for Implementation (Washington, DC: National Academies Press)

National Research Council, C. 2002, Making the Nation Safer: The Role of Science and Technology in Countering Terrorism (Washington, DC: National Academy Press)

Oda, M. 1965, Applied Optics, 4:1, 143

Percival, J. W., Jaehnig, K. P., & Nordsieck, K. H. 2007, in Bulletin of the American Astronomical Society, Vol. 38, American Astronomical Society Meeting Abstracts, 135.22

Percival, J. W., Nordsieck, K. H., & Jaehnig, K. P. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7010, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Purcell, W. R., et al. 1997, Astrophysical Journal, 491, 725

Reilly, D., Smith, H., Ensslin, N., & Kreiner, S. 1991, Passive Nondestructive Assay of Nuclear Materials, Tech. Rep. LA-UR-90-732, Los Alamos National Lab.

Remillard, R. A., & McClintock, J. E. 2006, Annu. Rev. Astro. Astrophys., 44, 49

Revnivtsev, M. G., et al. 2004, Astronomy Letters, 30, 382

Rozsa, C. M., & Menge, P. R. 2004, Efficiency Calculations for Selected Scintillators, Tech. rep., Saint-Gobain Crystals

Rozsa, C. M., Menge, P. R., & Mayhugh, M. R. 2006, Performance Summary: BrilLanCe Scintillators $LaCl_3$:Ce and $LaBr_3$:Ce, Tech. rep., Saint-Gobain Crystals

Schnopper, H. W., Bradt, H. V., Rappaport, S., Boughan, E., Burnett, B., Doxsey, R., Mayer, W., & Watt, S. 1970, Astrophysical Journal, 161, L161+

Schnopper, H. W., Thompson, R. I., & Watt, S. 1968, Space Science Reviews, 8, 534

Schwarzschild, K. 1905, Astronomische Mitteilungen der Universitaets-Sternwarte zu Goettingen, 10, 1

Shah, K. S., Glodo, J., Klugerman, M., Moses, W. W., Derenzo, S. E., & Weber, M. J. 2003, IEEE Transactions on Nuclear Science, 50, 2410

Shannon, C. E. 1948, Bell Systems Technical Journal, 27, 379

Shih, A. Y., et al. 2008, American Geophysical Union

Shikhaliev, P. M. 2010, Physics in Medicine and Biology, 55, 5317

Truemper, J. 1982, Advances in Space Research, 2, 241

Tueller, J., Mushotzky, R. F., Barthelmy, S., Cannizzo, J. K., Gehrels, N., Markwardt, C. B., Skinner, G. K., & Winter, L. M. 2008, Astrophysical Journal, 681, 113

Tueller, J., et al. 2010, Astrophysical Journal Supplement, 186, 378

Ubertini, P., et al. 2003, Astronomy & Astrophysics, 411, L131

U.S. Congress, Office of Technology Assessment. 1995a, Environmental Monitoring for Nuclear Safeguards, Tech. Rep. OTA-BP-ISS-168, Washington, DC

—. 1995b, Nuclear Safeguards and the International Atomic Energy Agency, Tech. Rep. OTA-ISS-615, Washington, DC

van Loef, E. V. D., Dorenbos, P., van Eijk, C. W. E., Krämer, K., & Güdel, H. U. 2001, Applied Physics Letters, 79, 1573

Weisskopf, M. C., O'Dell, S. L., Elsner, R. F., & Van Speybroeck, L. P. 1995, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 2515, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. R. B. Hoover & A. B. Walker, 312–329

Winkler, C., et al. 2003, Astronomy & Astrophysics, 411, L1

Wolter, H. 1952, Annalen der Physik, 445, 94

Yu, B., ed. 2009, Nuclear Science Symposium Conference Record, 2009, IEEE

Appendix A

Derivation of RM Characteristic Formulae

In Sec. A.1, the Classical Characteristic Formula for an RM is derived. While practical in many cases, this formula suffers from assumptions not always suitable. Next, the imaging system is generalized to allow for more flexibility of design including unrestricted bar width, bar spacing, and detector diameter. This Generalized Characteristic Formula (Sec. A.2) also accounts for assymetries in the system whereas the classical formula does not. Finally, the Advanced Characteristic Formula is derived (Sec. A.3), which addresses attenuation properties of the bars and accounts for clipping/edge and shadow-lengthening effects.

This appendix encompasses a complete derivation of the formula. Verifications are herein performed including proper reduction characteristics and Monte Carlo simulation results.

A.1 Classical Formula

A.1.1 Fractional Area of Circular Detector

A unit circle centered at the origin is defined by

$$x^2 + y^2 = 1, \tag{A.1}$$

or, for radius r and central coordinates x_0 and y_0,

$$r^2 = (x - x_0)^2 + (y - y_0)^2. \tag{A.2}$$

This relation is written in the form of a function with $y \to C(x)$,

$$C(x) = \pm\sqrt{r^2 - (x - x_0)^2} \tag{A.3}$$

A function $F(x^*)$ is needed to describe the percentage of area of a circle integrated when a chord is swept a distance x^* (unit diameter) across its face. Since $0 \leq x^* \leq 1$, $r = x_0 = 1/2$ (i.e. the circle has diameter 1). Then,

$$\begin{aligned} C(x) &= \pm\sqrt{\left(\frac{1}{2}\right)^2 - \left(x - \frac{1}{2}\right)^2} \\ &= \pm\sqrt{\left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}(2x-1)\right)^2} \\ &= \pm\sqrt{\left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 (2x-1)^2} \\ &= \pm\frac{1}{2}\sqrt{1 - (2x-1)^2}. \end{aligned} \tag{A.4}$$

Integration finds the total area enclosed by the circle through the distance x^*:

$$\begin{aligned} A(x^*) &= \int_0^{x^*} C(x)dx \\ &= 2\int_0^{x^*} \frac{1}{2}\sqrt{1 - (2x-1)^2}dx \\ &= \int_0^{x^*} \sqrt{1 - (2x-1)^2}dx. \end{aligned} \tag{A.5}$$

A table of integrals provides the following indefinite integral solution:

$$\int \sqrt{a^2 - x^2}dx = \frac{1}{2}x\sqrt{a^2 - x^2} + \frac{1}{2}a^2 \tan^{-1}\left(\frac{x}{\sqrt{a^2 - x^2}}\right) + C. \tag{A.6}$$

In this case, $a = 1$, so

$$\int \sqrt{1-x^2}dx = \frac{1}{2}x\sqrt{1-x^2} + \frac{1}{2}\tan^{-1}\left(\frac{x}{\sqrt{1-x^2}}\right) + C. \tag{A.7}$$

The trigonometric identity, $\tan^{-1}(x/\sqrt{1-x^2}) = \sin^{-1}(x)$, is used:

$$\int \sqrt{1-x^2}dx = \frac{1}{2}x\sqrt{1-x^2} + \frac{1}{2}\sin^{-1}x + C. \tag{A.8}$$

The substitution $x' = 2x - 1$, $dx' = 2dx \rightarrow dx = dx'/2$ is made in Eq. A.5:

$$\begin{aligned}
A(x^*) &= \int_0^{x^*} \sqrt{1-(x')^2}\frac{1}{2}dx' \\
&= \left(\frac{1}{2}\right)^2 \left[x'\sqrt{1-x'^2} + \sin^{-1}(x')\right]_0^{x^*} \\
&= \left(\frac{1}{2}\right)^2 \left[(2x-1)\sqrt{1-(2x-1)^2} + \sin^{-1}(2x-1)\right]_0^{x^*} \\
&= \left(\frac{1}{2}\right)^2 \left[(2x-1)\sqrt{1-4x^2-1+4x} + \sin^{-1}(2x-1)\right]_0^{x^*} \\
&= \left[\frac{1}{4}(2x-1)\sqrt{4(x-x^2)} + \frac{1}{4}\sin^{-1}(2x-1)\right]_0^{x^*} \\
&= \left[\frac{1}{2}(2x-1)\sqrt{x-x^2} + \frac{1}{4}\sin^{-1}(2x-1)\right]_0^{x^*} \\
&= \frac{1}{2}(2x^*-1)\sqrt{x^*-x^{*2}} + \frac{1}{4}\left(\sin^{-1}(2x^*-1) - \sin^{-1}(-1)\right) \\
&= \frac{1}{2}(2x^*-1)\sqrt{x^*-x^{*2}} + \frac{1}{4}\left(\sin^{-1}(2x^*-1) + \frac{\pi}{2}\right).
\end{aligned} \tag{A.9}$$

The term in parentheses,

$$\begin{aligned}
\left(\sin^{-1}(2x^*-1) + \frac{\pi}{2}\right) &= -\sin^{-1}(1-2x^*) + \frac{\pi}{2} \\
&= \cos^{-1}(1-2x^*),
\end{aligned} \tag{A.10}$$

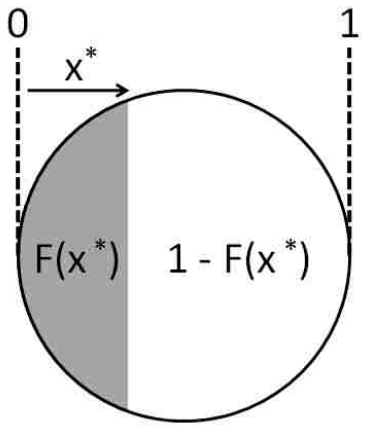


Figure A.1: Visual Representation of the percentage of circulation area integrated, $F(x^*)$.

provides the final $A(x^*)$,

$$A(x^*) = \frac{1}{4}\cos^{-1}(1-2x^*) - \frac{1}{2}(1-2x^*)\sqrt{x^* - x^{*2}}. \tag{A.11}$$

To find the fraction of circle area, Eq. A.11 is divided by the total area (see Figure A.1):

$$\begin{aligned} F(x^*) &= \frac{A(x^*)}{\pi\left(\frac{1}{2}\right)^2} \\ &= \frac{4A(x^*)}{\pi} \\ F(x^*) &= \frac{1}{\pi}\cos^{-1}(1-2x^*) - \frac{2}{\pi}(1-2x^*)\sqrt{x^* - x^{*2}}. \end{aligned} \tag{A.12}$$

This result provides the fractional area of a circle when integrated a unit distance x^* over its diameter. To relate this to fraction exposed or shadowed, grid specifics must be known. For a grid with its axis through the center of a bar, a particular detector may "start" completely shadowed, and so Equation A.12 will represent fraction exposed, while it will represent fraction shadowed for a grid with its axis centered on a spacing. The transformation between the two cases is seen to be

$$F(x^*) \rightarrow 1 - F(x^*), \tag{A.13}$$

so that the fractional area of a circle exposed for a given x^* fractional distance across the

detector is

$$F(x^*) = 1 - \frac{1}{\pi}\cos^{-1}(1-2x^*) + \frac{2}{\pi}(1-2x^*)\sqrt{x^* - x^{*2}}. \tag{A.14}$$

A.1.2 Detector Location in Shadow Frame

The grid rotation axis is defined as the origin. The grid shadow location must first be determined to describe detector shadowing. For a grid height L and source zenith θ, the shadow is offset from the grid in the detection plane by a magnitude

$$|\vec{R}| = L\tan\theta. \tag{A.15}$$

The x and y components of this position are determined by the source azimuth ϕ,

$$\begin{aligned} \vec{R}_x = |\vec{R}|\cos\phi &= L\tan\theta\cos\phi, \\ \vec{R}_y = |\vec{R}|\sin\phi &= L\tan\theta\sin\phi. \end{aligned} \tag{A.16}$$

A particular detector located at $\vec{r}' = (x_0', y_0')$ in the frame of the grid will be located at $\vec{r} = (x_0, y_0)$ in the frame of the grid shadow. Since

$$\begin{aligned} \vec{R}_x &= x_0 - x_0' \quad \text{and} \\ \vec{R}_y &= y_0 - y_0', \end{aligned} \tag{A.17}$$

it is found that

$$\begin{aligned} x_0 &= x_0' + L\tan\theta\cos\phi \quad \text{and} \\ y_0 &= y_0' + L\tan\theta\sin\phi. \end{aligned} \tag{A.18}$$

This result defines the components of a vector $\vec{r}$ to describe the location of a detector in the shadow frame for a source at (θ, ϕ):

$$\begin{aligned} \vec{r} &= x_0\hat{x} + y_0\hat{y} \\ &= (x_0' + L\tan\theta\cos\phi)\hat{x} + (y_0' + L\tan\theta\sin\phi)\hat{y} \\ r = |\vec{r}| &= \left[(x_0' + L\tan\theta\cos\phi)^2 + (y_0' + L\tan\theta\sin\phi)^2\right]^{\frac{1}{2}}. \end{aligned} \tag{A.19}$$

A.1.3 Determining Detector x^* Value

The grid shadow rotates, and thus the reference frame rotates. Modulation across $\vec{r}_y$ (defined to be parallel to the grid bars) is non-existent. Thus, only modulation across $\vec{r}_x$, which describes the binary nature of the bar pattern, is of concern. The time-dependent function of $x(t)$ is then

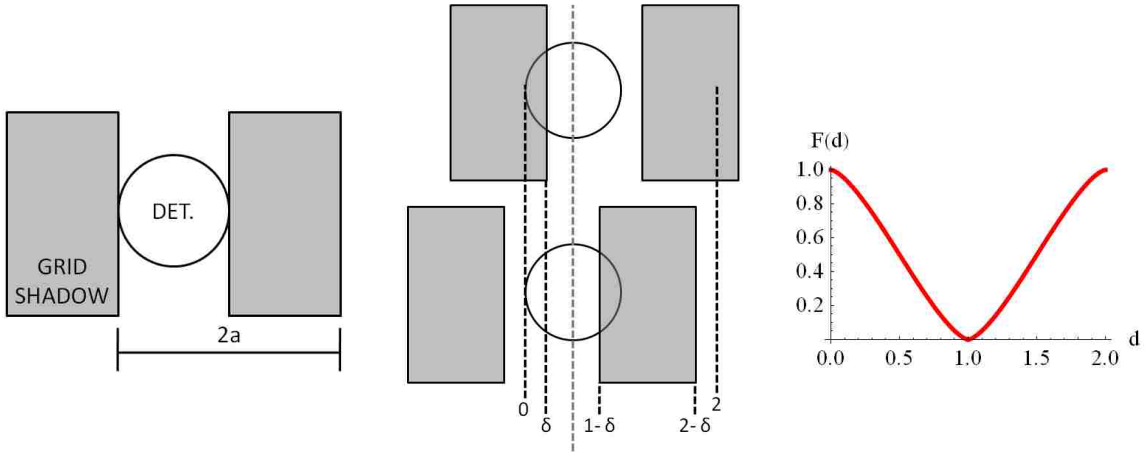
$$x(t) = r\cos(\xi(t) + \xi_0) \tag{A.20}$$

$$= \cos(\xi(t) + \xi_0)\sqrt{(x_0' + L\tan\theta\cos\phi)^2 + (y_0' + L\tan\theta\sin\phi)^2}, \tag{A.21}$$

where $\xi(t)$ is the orientation of the grid, and ξ_0 represents the starting angle of the detector in the (static) shadow frame. Note that $F(x^*)$ from the previous section is only valid for $0 \le x^* \le 1$. The final solution must be constructed as a piecewise function symmetric about $x = a$, with a period of $2a$ (see Fig. A.2a). The symmetry about 0 (i.e. $-x^*$ produces the same result as x^*) and the periodicity of $2a$ are taken into account by defining an index d in units of bar width a:

$$\begin{aligned} d &= \frac{1}{a}\left(\text{abs}\,(x) \mod 2a\right) \\ &= \frac{\text{abs}(x)}{a} \mod 2. \end{aligned} \tag{A.22}$$

The index d has range $0 \to 2$. $F(d)$, however, is symmetric about 1 (see Fig. A.2b,c),



(a) (b) (c)

Figure A.2: (a) Demonstration of $2a$ periodicity of grid pattern. (b) Demonstration of symmetry about $d = 1$ (c) Plot of $F(d)$.

and so a transformation $T(d)$ is required to satisfy a closed form,

$$x^* = T(d), \quad 0 \leq x^* \leq 1. \tag{A.23}$$

The relation of d and x^* that describes the same $F(x^*)$ over the entire range is considered:

$$\begin{aligned} 0 \leq d < 1 \qquad & x^* = d, \\ 1 \leq d \leq 2 \qquad & x^* = 2 - d. \end{aligned} \tag{A.24}$$

Zero is added to the right side of these relations to put them into forms that more closely resemble one another:

$$\begin{aligned} 0 \leq d < 1 \qquad & x^* = 1 + (d - 1) = 1 - (1 - d), \\ 1 \leq d \leq 2 \qquad & x^* = 1 + (2 - d - 1) = 1 + (1 - d) = 1 - (-1)(1 - d). \end{aligned} \tag{A.25}$$

It is apparent that the two relations are identical, except for the (-1) coefficient in one. For $0 \leq d < 1$, the term $(1 - d)$ is positive, while for $1 \leq d \leq 2$, $(-1)(1 - d)$ is positive. On each

of these ranges, the respective terms can be replaced with $|1-d|$ without altering the result:

$$0 \leq d \leq 2 \quad x^* = 1 - |1 - d|. \tag{A.26}$$

Plug in d from Eq. A.22 to arrive at

$$x^* = 1 - \left| \frac{|x|}{a} \bmod 2 - 1 \right|. \tag{A.27}$$

A.1.4 Summary

For an RM with grid co-axial on a spacing, the formula that describes the count profile for a unit-intensity source on a particular detector is

$$P_d(\theta, \phi, t) = F\left(1 - \left|\left| \frac{r(\theta,\phi)}{a} \cos(\xi(t) + \xi_0) \right| \bmod 2 - 1 \right|\right), \tag{A.28}$$

where a is the grid slat width/spacing, $\xi(t)$ is the angular orientation of the grid at time t, $r(\theta, \phi)$ is the distance between the detector position and the axis of the projected grid shadow from a grid height L,

$$r(\theta, \phi) = \left[(x_0 + L\tan\theta\cos\phi)^2 + (y_0 + L\tan\theta\sin\phi)^2\right]^{\frac{1}{2}}, \tag{A.29}$$

and $F(\tau)$ describes the fraction of exposed detector area,

$$F(\tau) = 1 - \frac{1}{\pi}\cos^{-1}(1-2\tau) + \frac{2}{\pi}(1-2\tau)(\tau-\tau^2)^{\frac{1}{2}}. \tag{A.30}$$

This formula is consistent with that described by Dadurkevicius & Ralys (1985).

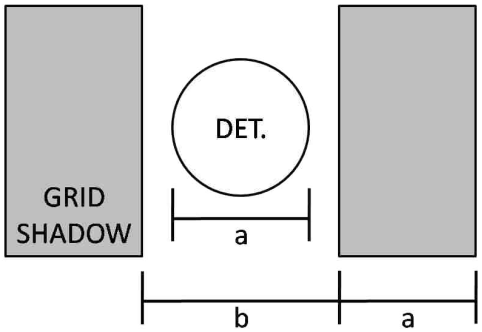


Figure A.3: Basic layout for $b > a$.

A.2 Generalized Formula

In the previous section, the characteristic RM formula is derived for equal slat, slit, and detector widths, ie. $a = b = c$. In this section, the formula is generalized to remove the dependency of these three paramenters. Such a derivation is necessary to complete that of the Advanced Formula in Sec. A.3. This derivation is performed by incrementally altering one of the three parameters and determining its effect on the system response.

A.2.1 Differing Bar (a) and Space (b) Width ($b > a$)

Fig. A.3 shows the layout for differing bar/space widths, with $b > a$. Detector width c is constrained to $c = a$. The nature of the profile is still periodic across one bar and one space width, $a + b$. Thus the detector x position may be folded across this range. The shadow pattern remains symmetric across the centered origin, $F(x) = F(-x)$. These two pieces of information yield a new position variable in units of detector diameter a:

$$d(x) = \frac{1}{a}\left(|x| \bmod (a+b)\right). \tag{A.31}$$

The shadow does not reach the detector until a distance $x = (b-a)/2$ has been traversed; a piecewise function will thus be required. A function with a closed form is desired, $x^* = T(d), 0 \leq x^* \leq 1$. For clarity, Fig. A.4 examines each stage of a complete slat shadow transversal.

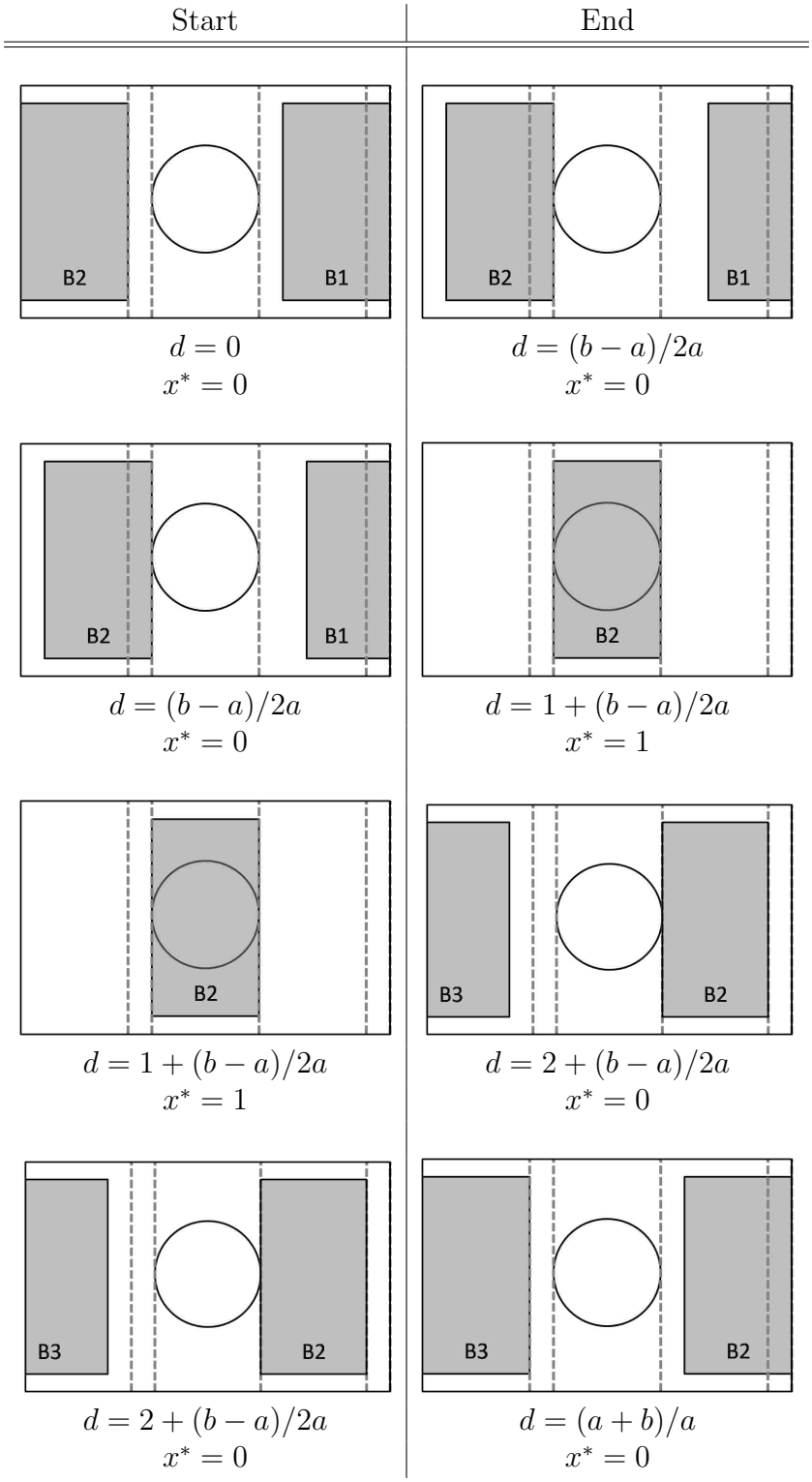


Figure A.4: A detailed examination of the piecewise nature for $b > a$.

Using the endpoints of each step as detailed in Fig. A.4, it is straightforward to construct the piecewise transformation function $x^*(d)$:

$$x^*(d) = \begin{cases} 0 & \text{if } 0 \le d \le (b-a)/2a \\ d - (b-a)/2a & \text{if } (b-a)/2a \le d \le 1 + (b-a)/2a \\ 2 + (b-a)/2a - d & \text{if } 1 + (b-a)/2a \le d \le 2 + (b-a)/2a \\ 0 & \text{if } 2 + (b-a)/2a \le d \le (a+b)/a. \end{cases} \tag{A.32}$$

The ranges can be simplified, since $1+(b-a)/2a = (a+b)/2a$ and $2+(b-a)/2a = (3a+b)/2a$:

$$x^*(d) = \begin{cases} 0 & \text{if } 0 \le d \le (b-a)/2a \\ d - (b-a)/2a & \text{if } (b-a)/2a \le d \le (a+b)/2a \\ 2 + (b-a)/2a - d & \text{if } (a+b)/2a \le d \le (3a+b)/2a \\ 0 & \text{if } (3a+b)/2a \le d \le (a+b)/a. \end{cases} \tag{A.33}$$

Consider only the cases where $x^* \neq 0$ (second and third parts of the piecewise Eq. A.33). Then,

$$\begin{aligned} x^*(d) &= \begin{cases} d - (b-a)/2a & \text{if } d < (a+b)/2a \\ 2 - (d - (b-a)/2a) & \text{if } d > (a+b)/2a \end{cases} \\ &= \begin{cases} d - (a+b)/2a + 1 & \text{if } d < (a+b)/2a \\ -d + (a+b)/2a + 1 & \text{if } d > (a+b)/2a \end{cases} \\ &= \begin{cases} 1 + (d - (a+b)/2a) & \text{if } d < (a+b)/2a \\ 1 - (d - (a+b)/2a) & \text{if } d > (a+b)/2a. \end{cases} \end{aligned} \tag{A.34}$$

Eq. A.34 is in a form that readily reveals its symmetry. For $d < (a+b)/2a$, the term $(d - (a+b)/2a)$ is negative, while for $d > (a+b)/2a$, the term $(d - (a+b)/2a)$ is positive.

Thus, the two pieces of Eq. A.34 can be rewritten as

$$x^*(d) = 1 - \left| d - \frac{a+b}{2a} \right|. \tag{A.35}$$

Eq. A.35 is only valid in the range $(b-a)/2a < d < (3a+b)/2a$, since outside of this range, $x^* = 0$. Eq. A.35, however, can never be greater than 1 due to the absolute value term on the right hand side. Under the condition that $|d - (a+b)/2a| > 1$, however, x^* is less than 0. This bound is defined by the ranges and easily verified:

$$\begin{aligned}
x^*(d = (b-a)/2a) &= 1 - \left| \frac{b-a}{2a} - \frac{a+b}{2a} \right| \\
&= 1 - \left| \frac{b-a-a-b}{2a} \right| \\
&= 1 - \left| \frac{-2a}{2a} \right| \\
&= 0 \\
, x^*(d = (3a+b)/2a) &= 1 - \left| \frac{3a+b}{2a} - \frac{a+b}{2a} \right| \\
&= 1 - \left| \frac{3a+b-a-b}{2a} \right| \\
&= 1 - \left| \frac{2a}{2a} \right| \\
&= 0.
\end{aligned}$$

Setting $x^* = 0$ outside this range is necessary to satisfy parts one and four of the piecewise Eq. A.33, and so a minimum constraint is imposed on x^* (plugging in Eq. A.31),

$$\begin{aligned}
x^* &= \max\left\{ 1 - \left| d - \frac{a+b}{2a} \right|, 0 \right\}, \quad b > a \\
&= \max\left\{ 1 - \left| \frac{|x| \bmod (a+b)}{a} - \frac{a+b}{2a} \right|, 0 \right\}, \quad b > a
\end{aligned} \tag{A.36}$$

where $\max\{x, y\}$ provides the largest of the two values x and y. The components of Eq.

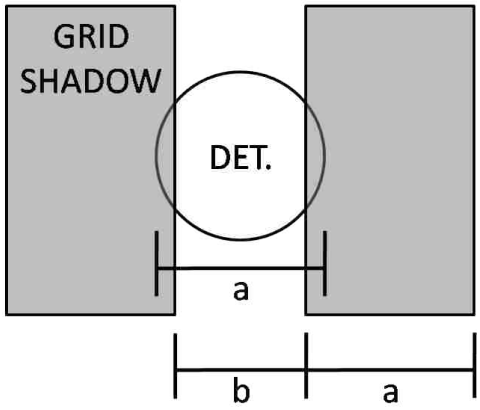


Figure A.5: Basic layout for $b < a$.

A.36 are easily interpreted to be rewritten as a conceptual 'pseudoformula':

$$x^* = \max\left\{[\text{Detector Width}] - \left|(|[\text{Location}]| \bmod [\text{Period}]) - [\text{Point of Symmetry}]\right|, 0\right\} \tag{A.37}$$

A.2.2 Differing Bar (a) and Space (b) Width ($b < a$)

For the case where $b < a$, the situation is similar to $b > a$, with the exception that multiple bars may simultaneously shadow the detector (see Fig. A.5). Fig. A.6 describes the piecewise nature of the system.

The function $x^*(d)$ is constructed as follows:

$$x^*(d) = \begin{cases} d + (a-b)/2a & \text{if } 0 \le d \le (a-b)/2a \\ d + (a-b)/2a & \text{if } (a-b)/2a \le d \le 1 - (a-b)/2a \\ 2 - d - (a-b)/2a & \text{if } 1 - (a-b)/2a \le d \le (a+b)/a - (a-b)/2a \\ 2 - d - (a-b)/2a & \text{if } (a+b)/a - (a-b)/2a \le d \le (a+b)/a. \end{cases} \tag{A.38}$$

Eq. A.38 only needs to have two pieces, however the first and fourth pieces will include a contribution from a second bar not noted here. The function is written in a form easier to

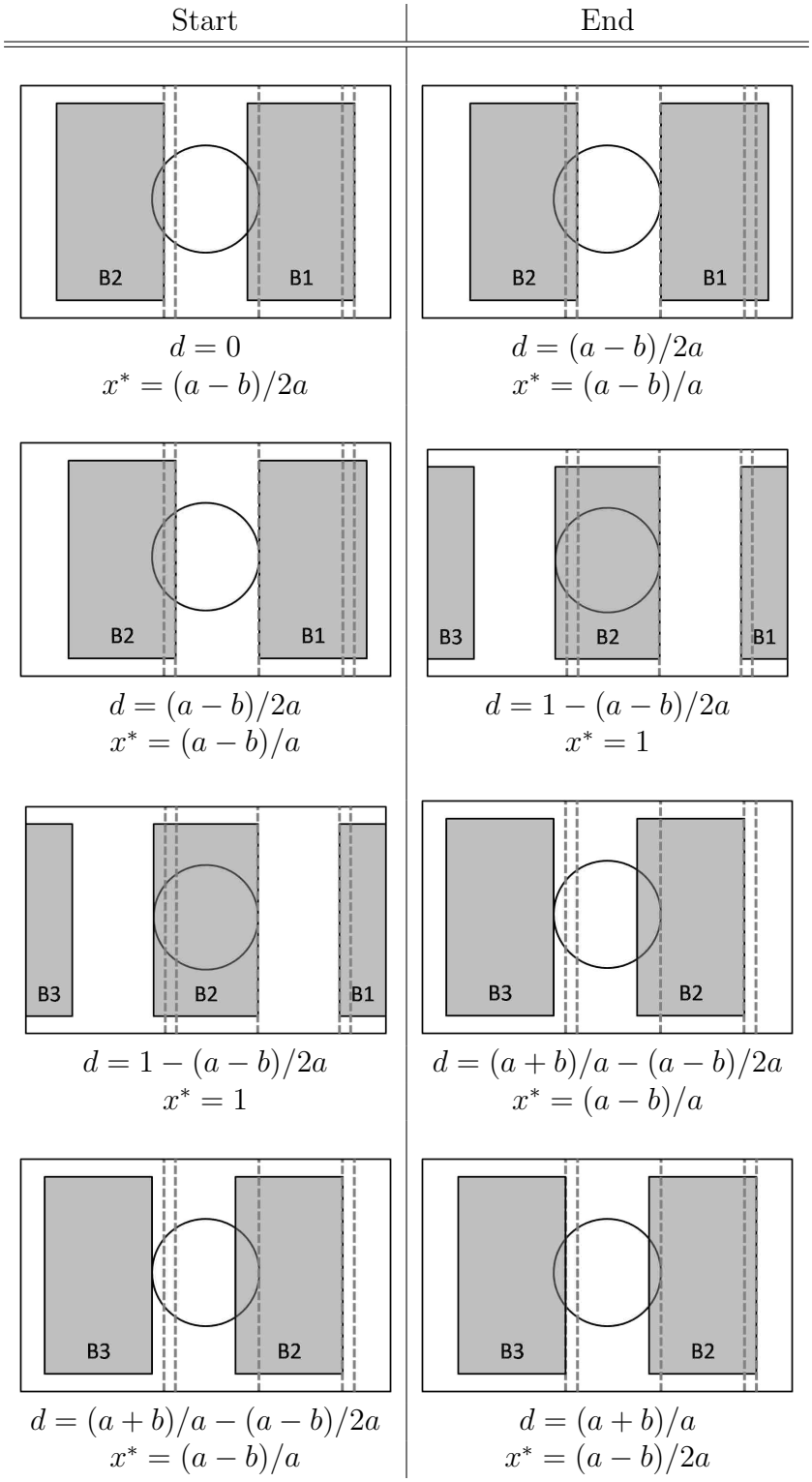


Figure A.6: A detailed examination of the piecewise nature for $b < a$.

read:

$$x^*(d) = \begin{cases} d + (a-b)/2a & \text{if } 0 \le d \le (a-b)/2a \\ d + (a-b)/2a & \text{if } (a-b)/2a \le d \le (a+b)/2a \\ 2 - (d + (a-b)/2a) & \text{if } (a+b)/2a \le d \le b/a + (a+b)/2a \\ 2 - (d + (a-b)/2a) & \text{if } b/a + (a+b)/2a \le d \le (a+b)/a. \end{cases} \tag{A.39}$$

Keeping in mind that two more pieces (for $d < (a-b)/2a$ and $d > b/a + (a+b)/2a$) must incorporate a second bar, Eq. A.38 is rewritten as only two pieces:

$$\begin{aligned} x^*(d) &= \begin{cases} d + (a-b)/2a & \text{if } d < (a+b)/2a \\ 2 - (d + (a-b)/2a) & \text{if } d > (a+b)/2a \end{cases} \\ &= \begin{cases} d + a/a - (a+b)/2a & \text{if } d < (a+b)/2a \\ 2 - (d + a/a - (a+b)/2a) & \text{if } d > (a+b)/2a \end{cases} \\ &= \begin{cases} 1 + (d - (a+b)/2a) & \text{if } d < (a+b)/2a \\ 1 - (d - (a+b)/2a) & \text{if } d > (a+b)/2a. \end{cases} \end{aligned} \tag{A.40}$$

In Eq. A.40, if $d < (a+b)$ then $(d - (a+b)/2a)$ is negative, while if $d > (a+b)/2a$, then $(d - (a+b)/2a)$ is positive. The two pieces are thus combined:

$$x^* = 1 - \left| d - \frac{a+b}{2a} \right|. \tag{A.41}$$

Eq. A.41 is identical to Eq. A.35. That is, the same equation may be used to describe the traversal of a bar, any width a and spacing b across the detector; it is only the lack of shadow for $b > a$ and addition of shadow for $b < a$ that distinguishes between the two cases.

In Fig. A.6, the second bar, whether it be B1 or B3, is always offset from the location of B2 by $\pm(a+b)/a$ (in units of detector width). This suggests that for a second bar,

$d \rightarrow d \pm (a+b)/a$; thus

$$x^*[] = \left\{ 1 - \left| d - \frac{a+b}{2a} \right|, 1 - \left| d \pm \frac{a+b}{a} - \frac{a+b}{2a} \right| \right\}, \tag{A.42}$$

where [] denotes an array. Equation A.42 incorporates all three bars which may appear on the detector during a single bar's period, however, it is apparent that the latter two terms are negative ($x^* < 0$) in extreme cases. Thus, a positivity constraint is implemented (on the latter terms only) as was done for $b > a$. The final equation is

$$x^*[] = \left\{ 1 - \left| d - \frac{a+b}{2a} \right|, \max \left\{ 1 - \left| d \pm \frac{a+b}{a} - \frac{a+b}{2a} \right|, 0 \right\} \right\}, \quad b < a. \tag{A.43}$$

A.2.3 Differing Bar (a) and Space (b) Width (any b and a)

The two previous sections examined the formulation of a fractional distance x^* across a circle's diameter for two cases: with bar spacing greater than bar width ($b > a$), and with bar spacing less than bar width ($b < a$). The two equations (A.36 and A.43) are extremely similar, with only two significant differences: (1) in Eq. A.36, a positivity constraint is implemented on the bar of interest, since part of its period consists of no detector shadowing; (2) in Eq. A.43, the left and right bars must be taken into account, and a positivity constraint included. There is no reason that distinctions (1) and (2) can not be incorporated into a single formula. The left and right bars can be added to Eq. A.36, since they make no contribution. Similarly, a positivity constraint can be added to the specified bar in Eq. A.43 without alteration of the result.

Thus, an array of values is obtained for the general case $b \neq a$:

$$x^*[] = \left\{ \max \left\{ 1 - \left| d - \frac{a+b}{2a} \right|, 0 \right\}, \max \left\{ 1 - \left| d \pm \frac{a+b}{a} - \frac{a+b}{2a} \right|, 0 \right\} \right\}. \tag{A.44}$$

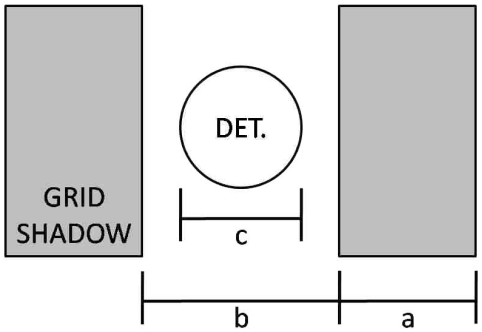


Figure A.7: Basic layout for $b > a > c$.

For completion, Eq. A.31 is plugged in and the result is simplified:

$$\begin{aligned} x^*[] &= \left\{ \max\left\{ 1 - \left| \frac{|x| \bmod (a+b)}{a} - \frac{a+b}{2a} \right|, 0 \right\}, \right. \\ &\quad \left. \max\left\{ 1 - \left| \frac{|x| \bmod (a+b)}{a} \pm \frac{a+b}{a} - \frac{a+b}{2a} \right|, 0 \right\} \right\} \\ &= \left\{ \max\left\{ 1 - \frac{1}{a} \left| |x| \bmod (a+b) - (a+b)/2 \right|, 0 \right\}, \right. \\ &\quad \max\left\{ 1 - \frac{1}{a} \left| |x| \bmod (a+b) + (a+b)/2 \right|, 0 \right\}, \\ &\quad \left. \max\left\{ 1 - \frac{1}{a} \left| |x| \bmod (a+b) - 3(a+b)/2 \right|, 0 \right\} \right\}. \end{aligned}$$

The final result for $b \neq a$ is then

$$x^*[] = \max\left\{ 1 - \frac{1}{a} \left| |x| \bmod (a+b) + n\frac{(a+b)}{2} \right|, 0 \right\}, \quad n = -1, 1, -3. \tag{A.45}$$

A.2.4 Differing Bar (a), Space (b), and Detector (c) Widths ($b > a > c$)

By not constraining the detector diameter, c, to the same size as the bar width, a, a new level of complexity is added. Every possible case will not be examined for the three variables. Rather, it is sufficient to examine a couple of the cases and infer a general solution from these

results. The derivation begins with the simplest of cases, $b > a > c$. Eq. A.31 requires only a small change, since it already incorporates the periodicity $(a + b)$ and symmetry across $(a + b)/2$. The alteration comes in the factor, $1/a$. Since this is no longer the width of the detector, d is now

$$d(x) = \frac{1}{c}\left(|x| \bmod (a + b)\right). \tag{A.46}$$

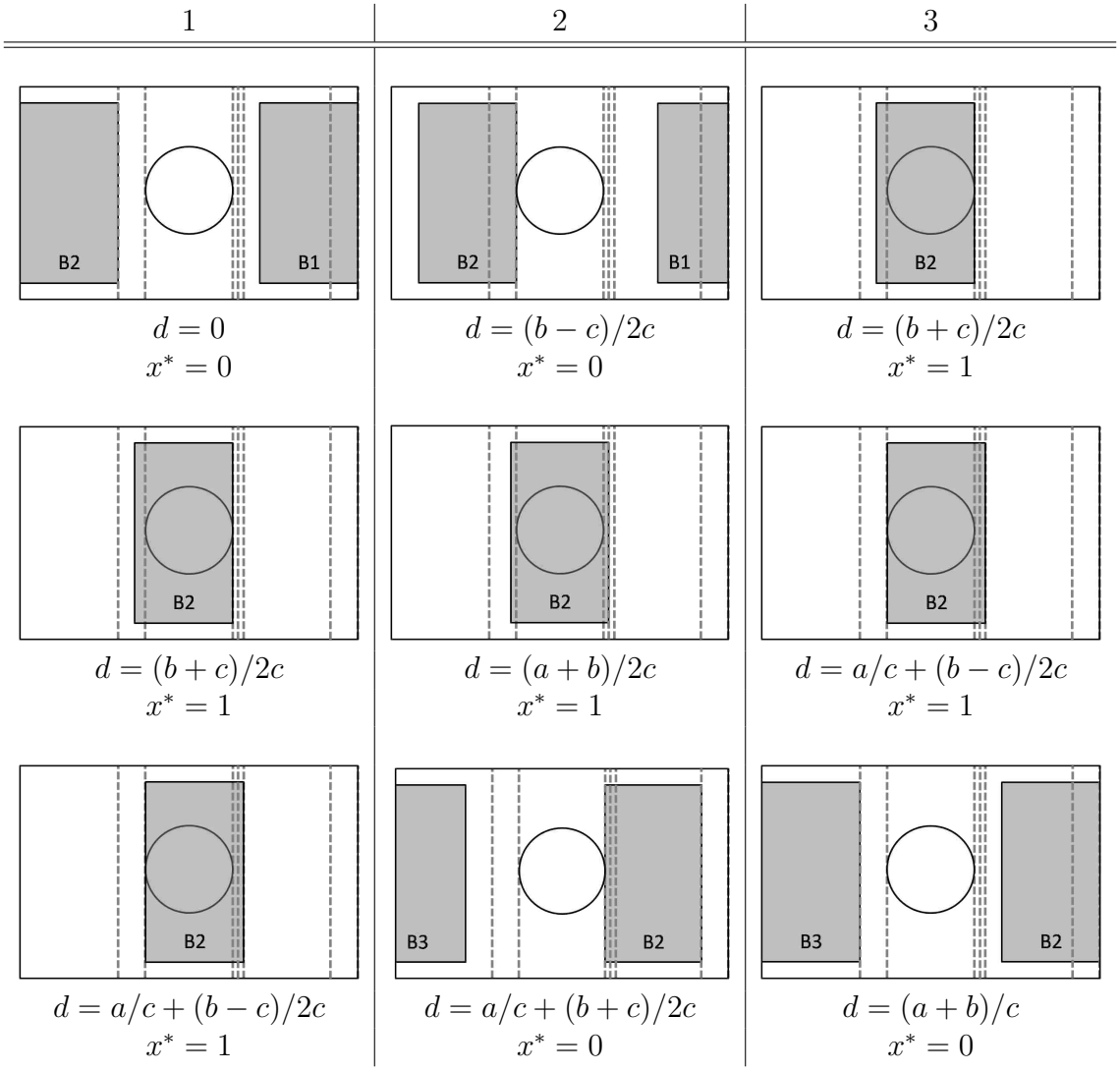


Figure A.8: A detailed examination of the piecewise nature for $b > a > c$.

Fig. A.8 is constructed to examine each stage of the bar shadow traversal and provides

the following piecewise function:

$$x^*(d) = \begin{cases} 0 & \text{if } 0 < d < (b-c)/2c \\ d - (b-c)/2c & \text{if } (b-c)/2c < d < (b+c)/2c \\ 1 & \text{if } (b+c)/2c < d(a+b)/2c \\ 1 & \text{if } (a+b)/2c < d < a/c + (b-c)/2c \\ (1+a/c) + (b-c)/2c - d & \text{if } a/c + (b-c)/2c < d < a + (b+c)/2c \\ 0 & \text{if } a + (b+c)/2c < d < (a+b)/c. \end{cases} \tag{A.47}$$

The second and fifth pieces of Eq. A.47 are considered and rewritten in a more suitable form (point of symmetry recognized according to Eq. A.37):

$$\begin{aligned} x^*(d) &= \begin{cases} d - \frac{b-c}{2c} & \text{if } d < \frac{a+b}{2c} \\ \left(1 + \frac{a}{c}\right) + \frac{b-c}{2c} - d & \text{if } d > \frac{a+b}{2c} \end{cases} \\ &= \begin{cases} d - \frac{a+b}{2c} + \left(\frac{1}{2} + \frac{a}{2c}\right) & \text{if } d < \frac{a+b}{2c} \\ -d + \frac{1}{2} + \frac{a}{2c} + \frac{a+b}{2c} & \text{if } d > \frac{a+b}{2c} \end{cases} \\ &= \begin{cases} \frac{1}{2} + \frac{a}{2c} + \left(d - \frac{a+b}{2c}\right) & \text{if } d < \frac{a+b}{2c} \\ \frac{1}{2} + \frac{a}{2c} - \left(d - \frac{a+b}{2c}\right) & \text{if } d > \frac{a+b}{2c} \end{cases} \\ &= \begin{cases} \frac{1}{2}\left(1 + \frac{a}{c}\right) + \left(d - \frac{a+b}{2c}\right) & \text{if } d < \frac{a+b}{2c} \\ \frac{1}{2}\left(1 + \frac{a}{c}\right) - \left(d - \frac{a+b}{2c}\right) & \text{if } d > \frac{a+b}{2c} \end{cases} \\ &= \frac{1}{2}\left(1 + \frac{a}{c}\right) - \left|d - \frac{a+b}{2c}\right|. \end{aligned}$$

The other pieces of Eq. A.47 are maximum and minimum constraints on x^*, so (by plugging

in Eq. A.46) the result becomes

$$\begin{aligned} x^* &= \min\left\{\max\left\{\frac{1}{2c}(a+c) - \left|d - \frac{a+b}{2c}\right|, 0\right\}, 1\right\} \\ &= \min\left\{\max\left\{\frac{1}{2c}(a+c) - \left|\frac{|x| \bmod (a+b)}{c} - \frac{a+b}{2c}\right|, 0\right\}, 1\right\} \\ x^* &= \frac{1}{c}\min\left\{\max\left\{\frac{1}{2}(a+c) - \left||x| \bmod (a+b) - \frac{a+b}{2}\right|, 0\right\}, 1\right\}, \\ & \quad b > a > c. \end{aligned} \tag{A.48}$$

The comprehensive pseudoformula further generalizes Eq. A.37:

$$\begin{aligned} &[\text{Bar/Detector Width Average}] - |[|\text{Location}]| \bmod [\text{Period}] - [\text{Point of Symmetry}]| \\ &\quad (\text{constrained between 0 and 1}) \end{aligned} \tag{A.49}$$

A.2.5 Differing Bar (a), Space (b), and Detector (c) Widths ($a > c$, any b)

A detailed analysis here is not required. By not constraining b to be greater than a and c, the only difference from the last case is that multiple bars (up to two) may shadow the bar simultaneously. Thus, a jump from Eq. A.48 to Eq. A.51 is made as was done from Eqs. A.36 and A.43 to A.45. The solution is, again, an array of three values for x^*:

$$x^*[] = \frac{1}{c}\min\left\{\max\left\{\frac{1}{2}(a+c) - \left||x| \bmod (a+b) + n\frac{a+b}{2}\right|, 0\right\}, 1\right\}, \tag{A.50}$$

$$(n = -1, 1, -3), \quad (a > c). \tag{A.51}$$

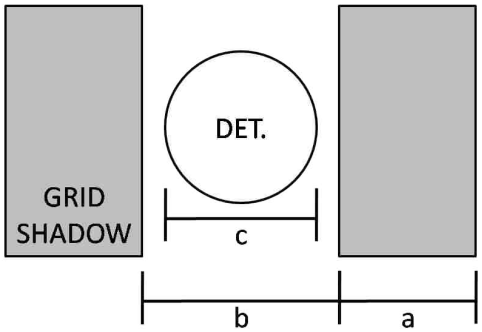


Figure A.9: Basic layout for $c > a$.

A.2.6 Differing Bar (a), Space (b), and Detector (c) Widths ($c > a$, any b)

Eq. A.51 contains two implications: (1) no more than two bars may ever shadow a portion of a single detector simultaneously, for to do so would require $2(a+b) < c$, and (2) a single bar always reaches the point of shadowing the entire detector since $a > c$. These assumptions are not valid, however, for the general case, where bar shadows may not cover the entire detector. Also, there exists no limit to the number of bars that may cast a shadow simultaneously onto a detector provided the detector diameter is large enough. Implication (2) is first considered. A piecewise function for x^* for $c > a$ may be constructed as in previous cases, but Fig. A.10 reveals a problem: the solution for x^* is ill-defined for $d = (a+b)/2c$. While this location may be found, plugging it into the formula for $F(x^*)$ yields an incorrect result since the left-most side of the detector is no longer shadowed.

The constraint requirements should be dropped from Eq. A.51:

$$x^*[] = \frac{1}{c}\left(\frac{1}{2}(a+c) - \left||x| \bmod (a+b) - n\frac{a+b}{2}\right|\right), \quad (n=-1,1,3), \quad (a>c), \qquad \text{(A.52)}$$

and, instead, Eq. A.12 is modified to include the constraints,

$$F(x^*) = \frac{1}{\pi}\cos^{-1}(1-2M[x^*]) - \frac{2}{\pi}(1-2M[x^*])\sqrt{M[x^*]-M[x^*]^2}, \qquad \text{(A.53)}$$

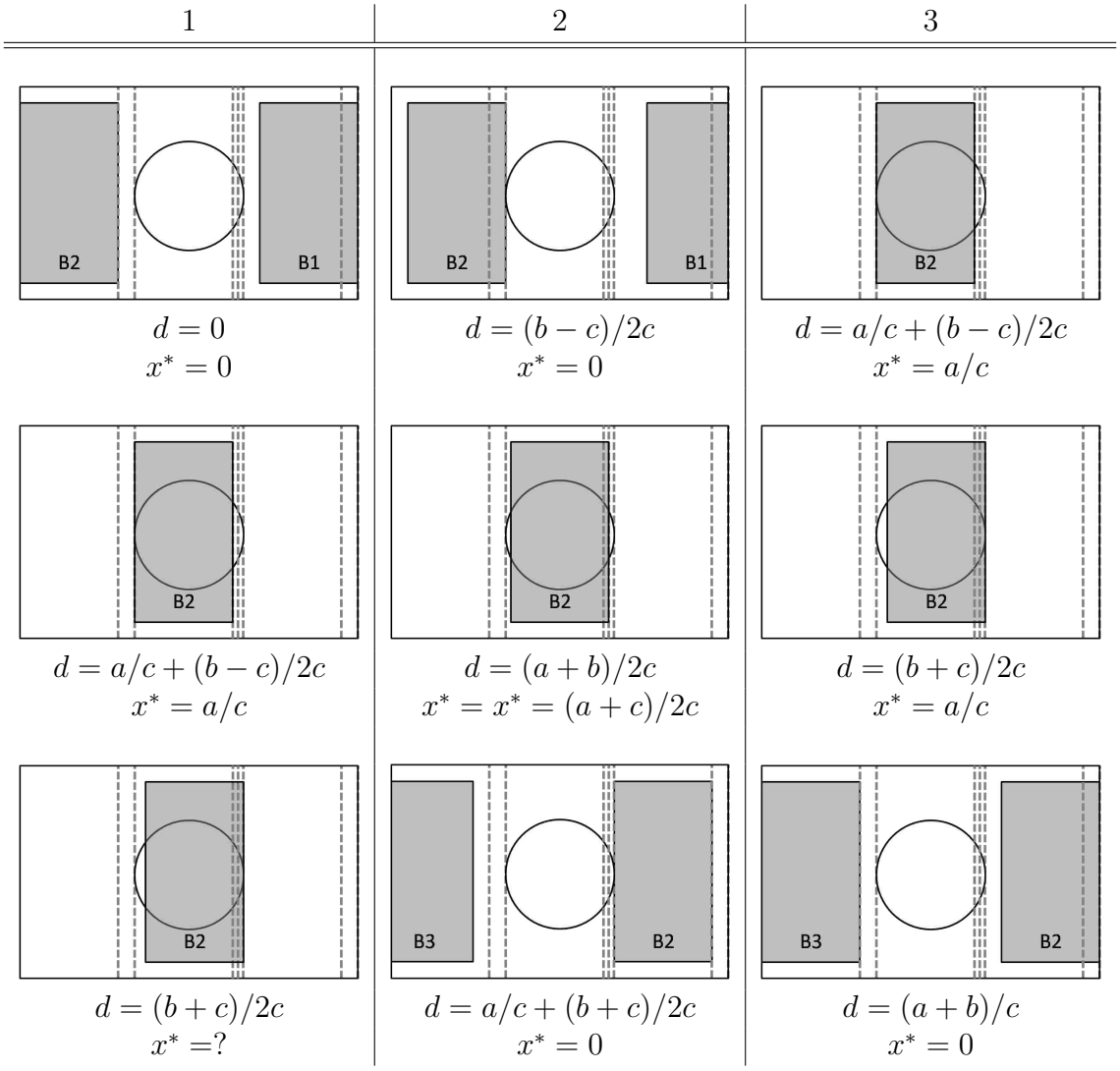


Figure A.10: A detailed examination of the piecewise nature for $c > a$.

where

$$M[\tau] = \min\{\max\{\tau, 0\}, 1\}. \tag{A.54}$$

The difference in this reorganization is nonexistant for previously examined cases, but will come of importance in the following step. A transformation of $F(x^*)$ is to remove the portion of unshadowed detector to the left of the shadow in question:

$$F(x^*) \to G(x^*, a, c) = F(x^*) - F(x^* - a/c). \tag{A.55}$$

The total number of bars shadowing the detector simultaneously must be determined, ensuring that $x^*[]$ is an array containing the leading edge locations of all bars. For $(a+b) > (a+c)$, only one bar will shadow the detector at a time. The relation $a+b = a+c$ (i.e. $b = c$) represents the smallest value for b that may occur without multiple shadows appearing simultaneously. For slightly smaller values of b, two more bars will appear (one before and one after the bar of interest). Thus, the number of bars that must be considered simultaneously per one bar's period is

$$N = 1 + 2 \text{ floor} \left(\frac{a+c}{a+b} \right). \tag{A.56}$$

In Eq. A.52, $N = 3$ and these bars are denoted by the coefficients

$$n = -1, 1, 3.$$

Extrapolating this result, the coefficients for $N = 5$ are

$$n = -3, -1, 1, 3, 5,$$

and so on. Using these examples, a more general solution for n is constructed:

$$\begin{aligned} n &= \{-3, -1, 1, 3, 5\} \\ &= \{-4, -2, 0, 2, 4\} + 1 \\ &= 2\{-2, -1, 0, 1, 2\} + 1 \\ &= 1 + 2\{-(N-1)/2 ... (N-1)/2\}. \end{aligned}$$

Since this result is messy, N is redefined,

$$N = \text{ floor} \left(\frac{a+c}{a+b} \right), \tag{A.57}$$

and then n is simply

$$n = 1 + 2\{-N...N\}. \tag{A.58}$$

Eq. A.52 is generalized to

$$x^* = \frac{1}{c}\left(\frac{1}{2}(a+c) - \left||x| \bmod (a+b) - n\frac{a+b}{2}\right|\right), \quad n = 1 + 2\{-N...N\}. \tag{A.59}$$

A.2.7 Summary

A general solution for x^* has been constructed which will work under most circumstances for varying bar and spacing widths, a and b. Once c is allowed to become smaller than these two parameters, however, a modification to $F(x^*)$ must take place to account for multiple bars. Additionally, constraints must be enforced to keep x^* within the range $[0, 1]$. The results of this section are summarized below.

For an RM with bar width a, spaced an equidistance apart, b, and detectors, diameter c, if the grid is co-axial on a spacing, the count profile of a unit intensity source located at (θ, ϕ) is

$$P_d(\theta, \phi, t) = 1 - G(x^*, a, c), \tag{A.60}$$

where $G(x^*, a, c)$ represents the fraction of the detector shadowed,

$$G(x^*, a, c) = \sum_n \left(F(x_n^*) - F(x_n^* - a/c)\right). \tag{A.61}$$

The function $F(x^*)$ is defined as

$$F(\tau) = \frac{1}{\pi}\cos^{-1}(1 - 2M[\tau]) - \frac{2}{\pi}(1 - 2M[\tau])\sqrt{M[\tau] - M[\tau]^2}, \tag{A.62}$$

where $M[\tau]$ represents the constraint formula,

$$M[\tau] = \min\{\max\{\tau, 0\}, 1\}, \tag{A.63}$$

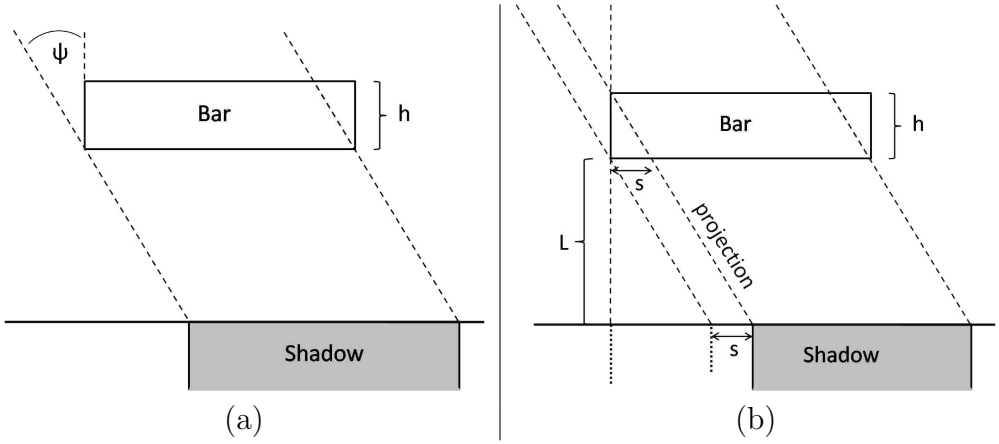


Figure A.11: Side views of (a) the "thickness problem" and (b) the definition of s.

and x_n^* is an array of fractional distances across the detector diameter,

$$x_n^* = \frac{1}{c}\left(\frac{1}{2}(a+c) - \left||x| \bmod (a+b) - n\frac{a+b}{2}\right|\right), \quad n = 1+2\{-N...N\} \tag{A.64}$$

with

$$N = \text{ floor}\left(\frac{a+c}{a+b}\right). \tag{A.65}$$

The position of the detector in the shadow frame, x, is

$$x = r(\theta,\phi)\cos(\xi(t)+\xi_0), \tag{A.66}$$

where r is given by

$$r(\theta,\phi) = \left[(x_0 + L\tan\theta\cos\phi)^2 + (y_0 + L\tan\theta\sin\phi)^2\right]^{\frac{1}{2}}. \tag{A.67}$$

A.3 Advanced Formula

The formulation of generalized modulation profiles describing unequal bar/spacing/detector width parameters was devised specifically to set up the ground work necessary for the in-

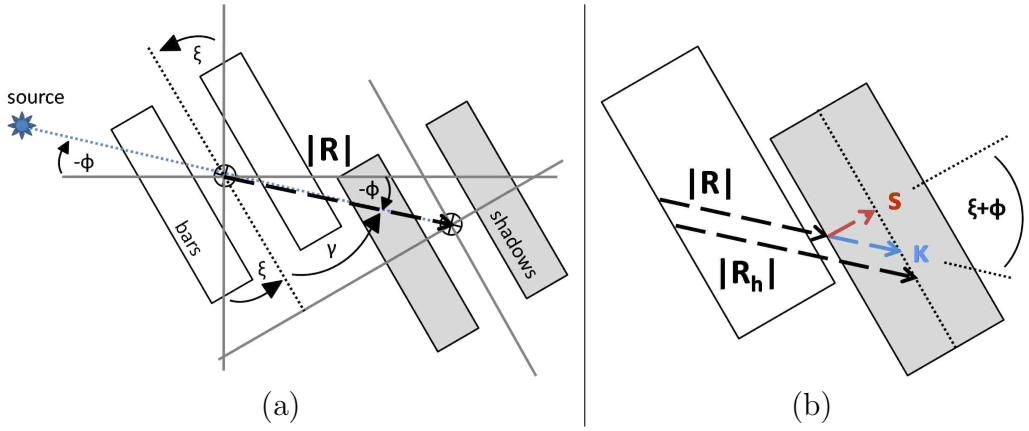


Figure A.12: Top-down views of (a) the relation of $|\vec{R}|$, ξ, and ϕ, and (b) the means of finding s.

corporation of thickness–and thus attenuation and clipping effects–into the analytical RM solution. As with the formulation of the advanced modulation profiles, the Advanced Characteristic Formula is derived by stepping through sequential generalizations of the geometry.

A.3.1 Consideration of Constant Thickness Portion

The modulation profiles derived in Eqs. A.1 and A.2 assume zero bar thickness, and will thus cast a shadow as described in Fig. A.11(a). The derivation begins by including only the portion of the light which has traversed the maximum amount of material possible. That is, any photons that may have "clipped" the bar are ignored. As seen in Fig. A.11(b), the "right" side of the projection remains in place, while the "left" side is moved in by an amount s. Similarly, for a projection in the negative direction, the "right" side would move in by an amount s, and the "left" side would remain fixed. In either case, the width of the bar's shadow, a', is smaller than the bar width itself, a, by an amount $|s|$:

$$a' = a - |s|. \tag{A.68}$$

The magnitude (and sign) of s modulates as a function of the grid rotation, ξ, and the azimuthal location of the source, ϕ. Figure A.12(a) offers a pictorial view of these relations.

The positive ξ and ϕ are defined in opposite directions, and so increasing either one will have an identical effect on s. Figure A.12(a) also describes the vector magnitude $|\vec{R}|$, which may be recalled from Eq. A.15,

$$|\vec{R}| = L \tan\theta.$$

This magnitude is the distance from the grid origin to the shadow origin projected onto the shadow plane, i.e., it is the projected distance from *any* point on the grid bottom and its counterpart in the shadow plane. Thus, in Fig. A.12(b), a new vector magnitude, $|\vec{R}_h|$, is defined, which is the projected difference between the *top* of the grid with thickness h and the corresponding location on the shadow,

$$|\vec{R}_h| = (L + h) \tan\theta. \tag{A.69}$$

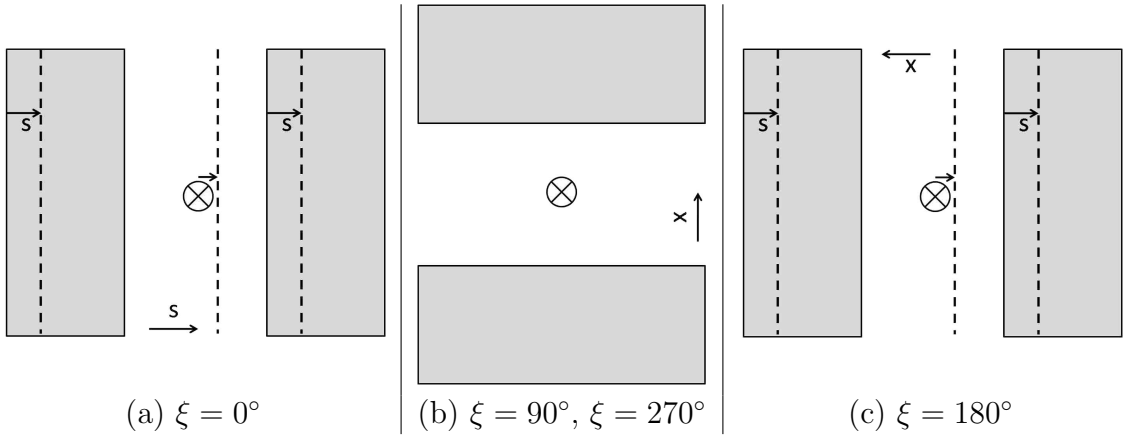
The difference in length between this vector and $|\vec{R}|$ is

$$\kappa = |\vec{R}_h| - |\vec{R}|. \tag{A.70}$$

This distance, however, is in the direction of projection. The width s is desired, which is in the direction of the shadow's x-component. The length s relates to κ by

$$\begin{aligned} s &= \kappa \cos(\xi + \phi) \\ &= (|\vec{R}_h| - |\vec{R}|) \cos(\xi + \phi) \\ &= [(L + h) \tan\theta - L \tan\theta] \cos(\xi + \phi) \\ &= h \tan\theta \cos(\xi + \phi). \end{aligned} \tag{A.71}$$

The projected width of the bar shadows, without bar clipping, is determined from Eqs. A.68



(a) $\xi = 0°$ (b) $\xi = 90°$, $\xi = 270°$ (c) $\xi = 180°$

Figure A.13: Demonstration of the now-offset point of symmetry, and how it relates to grid rotation. The axis of rotation is designated by $\otimes$, while the point of symmetry is designated by the dashed bar between the shadows. Source azimuth, $\phi = 0$ (into the page and to the right) for sake of simplicity.

and A.71:

$$\begin{aligned} a' &= a - |s| \\ &= a - |h \tan\theta \cos(\xi + \phi)| \\ &= a - h \tan\theta |\cos(\xi + \phi)|. \end{aligned} \tag{A.72}$$

Similarly, the width of the shadow spacings increases by an amount $|s|$:

$$b' = b + h \tan\theta |\cos(\xi + \phi)|. \tag{A.73}$$

The parameter requiring modification in the modulation formula is x. In the generalized formula, x is defined with the origin centered on the projected rotation axis, which is also a point of symmetry in the pattern, hence the $|x|$ in Eq. A.64. The rotation axis, however, no longer lies on the point of symmetry in the shadow, except for when the projection direction is parallel to the grid (i.e. has no x-component, or $\xi(t) = -\phi$). Fig. A.13 demonstrates this concept. Since b expands to $b' = b + s$, the point of symmetry will shift by $s/2$. Thus, a new

point of symmetry is defined, x', to be

$$\begin{aligned} x' &= x + \frac{s}{2} \\ &= x + \frac{h}{2}\tan\theta\cos(\xi+\phi). \end{aligned} \tag{A.74}$$

Eq. A.64 is rewritten with the newly derived variables:

$$x_n^* = \frac{1}{c}\left(\frac{1}{2}(a'+c) - \left||x'| \bmod (a'+b') - n\frac{a'+b'}{2}\right|\right), \quad n = 1+2\{-N...N\}. \tag{A.75}$$

However, since $a'+b' = a+b$, it may also be written as

$$x_n^* = \frac{1}{c}\left(\frac{1}{2}(a'+c) - \left||x'| \bmod (a+b) - n\frac{a+b}{2}\right|\right), \quad n = 1+2\{-N...N\}, \tag{A.76}$$

where

$$\begin{aligned} a' &= a - h\tan\theta|\cos(\xi+\phi)| \quad \text{and} \\ x' &= x + \frac{h}{2}\tan\theta\cos(\xi+\phi). \end{aligned}$$

Finally, the parameter N is rewritten with these new variables:

$$N = \text{floor}\left(\frac{a'+c}{a+b}\right). \tag{A.77}$$

Note that Eq. A.77 is a function no longer of only constants, but now of a' also, which is time dependent. Thus the array, n, changes temporally. While possible to implement, this consequence requires additional computational time. Instead, it is prudent to find the maximum *possible* value of N and let constraint conditions, $M[\tau]$, satisfy boundary requirements. The maximum N occurs simultaneously with maximum a', which is just a. Thus Eq. A.65 is preferred over Eq. A.77.

A.3.2 Inclusion of Attenuation

Inclusion of proper attenuation in the characteristic formulae will lower the profile contrast, producing a noticeable effect for small thickness, density, or cross sections. The Beer-Lambert law gives the fraction of light absorbed by a thickness δ,

$$f(\delta) = 1 - e^{-\sigma\delta\rho}, \tag{A.78}$$

where ρ is the density of the material, and σ is the mass attenuation coefficient for a particular energy. For the constant thickness portion, the distance, δ, through which a photon will travel is a function only of the source zenith angle and the bar thickness, h:

$$\delta = \frac{h}{\cos\theta}. \tag{A.79}$$

Eq. A.79 is plugged into Eq. A.78,

$$f(\theta) = 1 - \exp\left[-\frac{h\sigma\rho}{\cos\theta}\right]. \tag{A.80}$$

For implementation into the profile, rewrite Eq. A.61 with $a \rightarrow a'$,

$$G(x^*, a, c) = \sum_n (F(x_n^*) - F(x_n^* - a'/c)).$$

For clarity, this equation may be rewritten as

$$G(\theta, \phi, t) = \sum_n \left(F[x_n^*(\theta, \phi, t)] - F[x^*(\theta, \phi, t) - a'(\theta, \phi, t)/c]\right). \tag{A.81}$$

This coefficient is included in Eq. A.61 (with $a \rightarrow a'$),

$$G_1(\theta, \phi, t) = \sum_n f(\theta)\left(F[x_n^*(\theta, \phi, t)] - F[x^*(\theta, \phi, t) - a'(\theta, \phi, t)/c]\right), \tag{A.82}$$

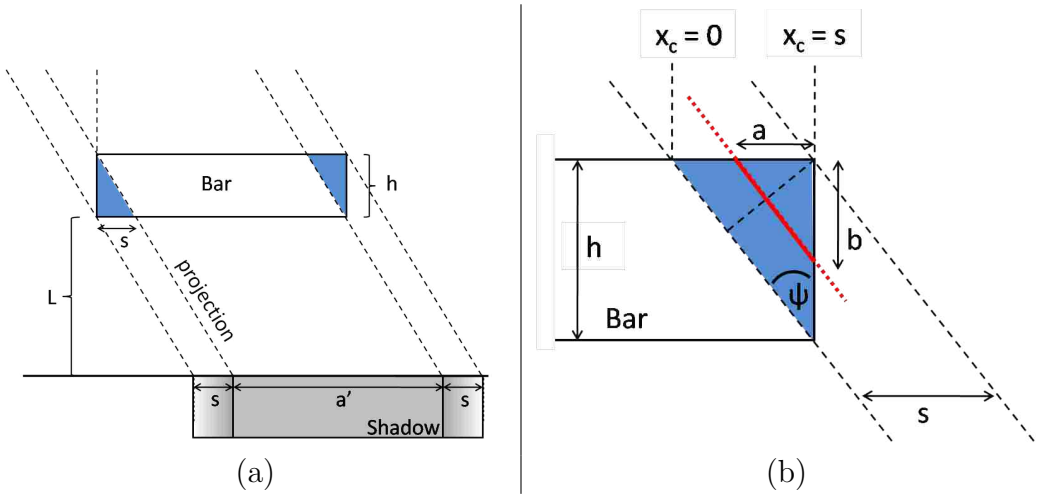


Figure A.14: (a) Inclusion of "clipped" portion of the bar. (b) Determining the amount of material traversed in the clipping portion.

where

$$x_n^*(\theta,\phi,t) = \frac{1}{c}\left(\frac{1}{2}\left(a'(\theta,\phi,t)+c\right) - \left|\,|x'(\theta,\phi,t)| \bmod (a+b) - n\frac{a+b}{2}\right|\right), \quad n = 1+2\{-N...N\}. \tag{A.83}$$

A.3.3 Determining Distance Traversal through Clipped Portion

Fig. A.14(a) shows the clipping portion of the bar still unaccounted for in the formula. Fig. A.14(b) examines the clipping portion in detail. A function for $\delta(x)$, the distance traversed through the bar as x_c goes from $0 \rightarrow 1$ is desired. A piecewise function may be written:

$$\delta(x_c) = \begin{cases} \frac{h}{\cos\theta} & \text{if } x_c = 0 \\ 0 & \text{if } x_c = s. \end{cases} \tag{A.84}$$

From this figure, it is apparent that

$$s = h\tan\psi, \tag{A.85}$$

though angle ψ remains unknown. Eq. A.71 provides another defintion of s:

$$s = h \tan\theta \cos(\xi + \phi).$$

These two equations are combined to solve for the unknown angle ψ,

$$\tan\psi = \tan\theta \cos(\xi + \phi). \tag{A.86}$$

A component of the traversal distance (the component perpendicular to the x-direction of the grid frame, or in the plane of view of Fig. A.14b) is related to lengths a and b by

$$\delta_\perp = \sqrt{a^2 + b^2}. \tag{A.87}$$

These distances are easily found geometrically:

$$\begin{aligned} a &= s - x_c, \quad \text{and} && \text{(A.88)} \\ b &= \frac{a}{\tan\psi} \\ &= \frac{s - x_c}{\tan\psi} \\ &= \frac{s - x_c}{\tan\theta \cos(\xi + \phi)}, && \text{(A.89)} \end{aligned}$$

and plugged into Eq. A.87:

$$\begin{aligned} \delta_\perp &= \sqrt{a^2 + b^2} \\ &= \sqrt{(s - x_c)^2 + \frac{(s - x_c)^2}{\tan^2\theta \cos^2(\xi + \phi)}} \\ &= (s - x_c)\sqrt{1 + \frac{1}{\tan^2\theta \cos^2(\xi + \phi)}}. && \text{(A.90)} \end{aligned}$$

At $x_c = 0$,

$$\begin{aligned}
\delta &= \frac{h}{\cos\theta}, \quad \text{and} \\
\delta_\perp &= s\sqrt{1 + \frac{1}{\tan^2\theta\cos^2(\xi+\phi)}} \\
&= h\tan\theta\cos(\xi+\phi)\sqrt{1 + \frac{1}{\tan^2\theta\cos^2(\xi+\phi)}} \\
&= h\sqrt{\tan^2\theta\cos^2(\xi+\phi)+1}.
\end{aligned}$$

Apparently,

$$\begin{aligned}
\delta &= \frac{\delta_\perp}{\cos\theta\sqrt{\tan^2\theta\cos^2(\xi+\phi)+1}} \\
&= (s-x_c)\sqrt{1 + \frac{1}{\tan^2\theta\cos^2(\xi+\phi)}}\,\frac{1}{\cos\theta\sqrt{\tan^2\theta\cos^2(\xi+\phi)+1}} \\
&= \frac{s-x_c}{\cos\theta}\,\frac{1}{\sqrt{\tan^2\theta\cos^2(\xi+\phi)+1}}\,\frac{\sqrt{\tan^2\theta\cos^2(\xi+\phi)+1}}{\tan\theta\cos(\xi+\phi)} \\
&= \frac{s-x_c}{\cos\theta\tan\theta\cos(\xi+\phi)}, \\
|\delta(x_c)| &= \frac{s-x_c}{\sin\theta|\cos(\xi+\phi)|}.
\end{aligned} \tag{A.91}$$

Using Eq. A.80, the fraction of attenuation in the clipped portion is determined by

$$f(\theta, x_c) = 1 - \exp\left[-\frac{(s-x_c)\sigma\rho}{\sin\theta|\cos(\xi+\phi)|}\right]. \tag{A.92}$$

A.3.4 Weighting the Integrated Area for Clipping Portion

Unlike the constant thickness portion of the shadow, the clipping shadow region is not spatially uniform – i.e, the shadow is darkest at $x_c = 0$, and zero at $x_c = s$. The problem is illustrated in Fig. A.15. To include this bar shadow portion into the modulation formula, the "right" clipped portion as appears in Fig. A.15b is first examined. A portion of the

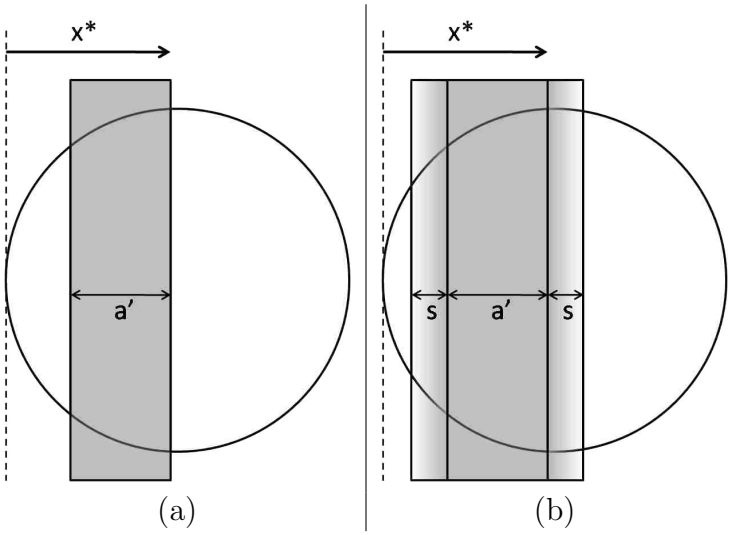


Figure A.15: (a) The integration over the portion of circle area when the clipping was ignored – constant weighting. (b) The integration over the circle area with clipping effects – weighting is a function of x^*.

circle is integrated with a specific weighting factor, where the weight is zero at the line that designates the upper integration bound, and increases to the left according to Eq. A.92. The limits of integration of Eq. A.8 are changed to x^* to $x^* + |s|/c$; this factor of c is necessary since the limits of integration must be in unit diameter. The abbreviation $s_c = |s|/c$ is used, and a weighting factor, f, is added:

$$A(x^*) = \int_{x^*}^{x^*+s_c} f\sqrt{1-(2x-1)^2}dx. \tag{A.93}$$

But f must be defined in terms of both the part of the shadow in the integration, x, and the upper bound of the integration, x^*:

$$f(x) = 1 - \exp\left[-\frac{(|s| + cx^* - cx)\sigma\rho}{\sin\theta|\cos(\xi+\phi)|}\right]. \tag{A.94}$$

The result is plugged into the area integral, with limits constrained between 0 and 1 (see Fig. A.16):

$$\begin{aligned} A(x^*) &= \int_{M[x^*]}^{M[x^*+s_c]} \left(1 - e^{-\frac{(|s|+cx^*-cx)\sigma\rho}{\sin\theta|\cos(\xi+\phi)|}}\right) \sqrt{1-(2x-1)^2}dx \qquad (A.95) \\ &= \int_{M[x^*]}^{M[x^*+s_c]} \sqrt{1-(2x-1)^2}dx \\ &\quad -e^{-\frac{(|s|+cx^*)\sigma\rho}{\sin\theta|\cos(\xi+\phi)|}} \int_{M[x^*]}^{M[x_s^*]} e^{\frac{cx\sigma\rho}{\sin\theta|\cos(\xi+\phi)|}} \sqrt{1-(2x-1)^2}dx \\ &= \int_{M[x^*]}^{M[x^*+s_c]} \sqrt{1-(2x-1)^2}dx - e^{-\frac{(|s|+cx^*)\sigma\rho}{\sin\theta|\cos(\xi+\phi)|}}\Gamma(x^*), \qquad (A.96) \end{aligned}$$

where

$$\Gamma(x^*) = \int_{M[x^*]}^{M[x^*+s_c]} e^{\frac{cx\sigma\rho}{\sin\theta|\cos(\xi+\phi)|}} \sqrt{1-(2x-1)^2}dx, \qquad (A.97)$$

or, defining a constant $Z = c\sigma\rho/(\sin\theta|\cos(\xi+\phi)|)$,

$$\Gamma(x^*) = \int_{M[x^*]}^{M[x^*+s_c]} e^{Zx} \sqrt{1-(2x-1)^2}dx. \qquad (A.98)$$

The first integral of Eq. A.96 has already been solved (Eq. A.62). It may be written in terms of the function $F(\tau)$,

$$\int_{M[x^*]}^{M[x^*+s_c]} \sqrt{1-(2x-1)^2}dx = \frac{\pi}{4}\left[F(x^*+s_c) - F(x^*)\right], \qquad (A.99)$$

where the integration constraints are already included in $F(\tau)$. The fraction of the detector shadowed by a single clipping portion is then

$$\begin{aligned} A(x^*) &= \frac{\pi}{4}\left[F(x^*+s_c) - F(x^*)\right] - \int_{M[x^*]}^{M[x^*+s_c]} e^{Z(x-x^*-s_c)} \sqrt{1-(2x-1)^2}dx \\ &= \frac{\pi}{4}\left[F(x^*+s_c) - F(x^*)\right] - \frac{\pi}{4}[\widetilde{F_1}(x^*+s_c) - \widetilde{F_2}(x^*)], \qquad (A.100) \end{aligned}$$

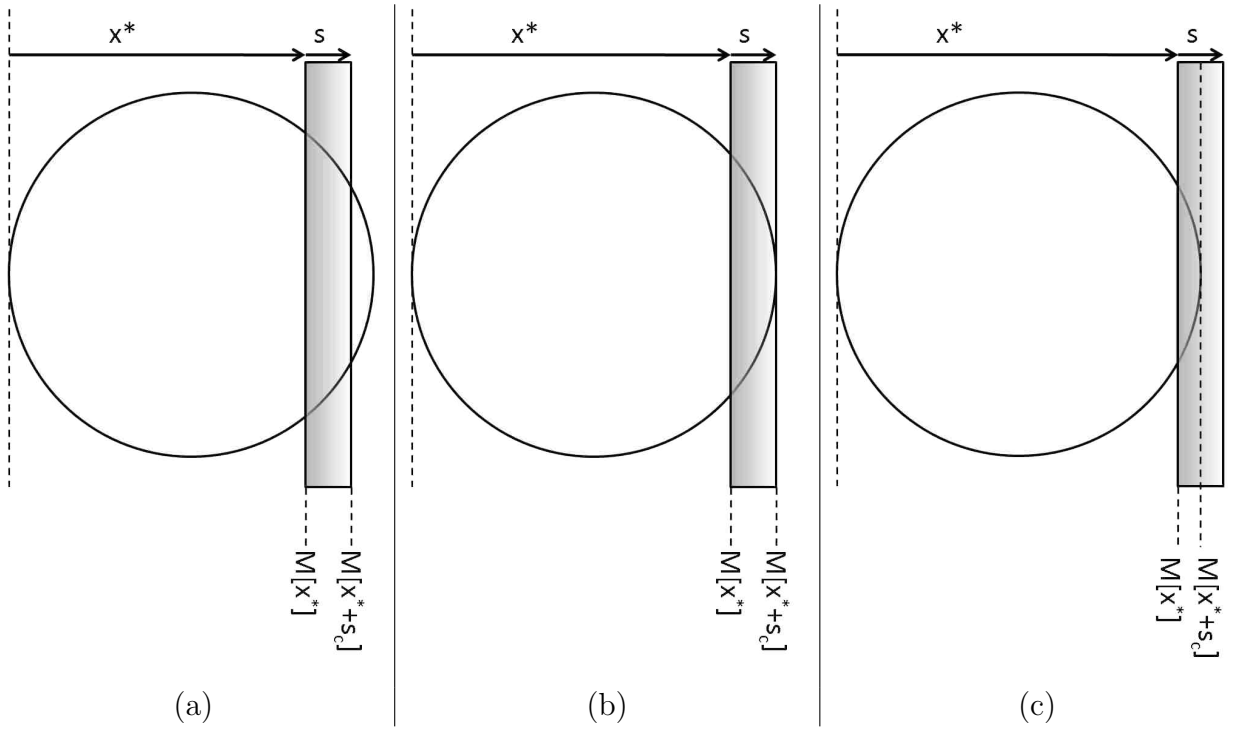


Figure A.16: The integration limits for the clipping shadow (a) totally inside the circular detector, (b) just at the edge, and (c) falling outside the detector area. Notice the integration range is limited by the detector edges (including the left side not shown).

where

$$\widetilde{F_1}(\tau) = \frac{4}{\pi}\int_0^{M[\tau]} e^{Z(x-\tau)}\sqrt{1-(2x-1)^2}dx, \quad \text{and} \tag{A.101}$$

$$\widetilde{F_2}(\tau) = \frac{4}{\pi}\int_0^{M[\tau]} e^{Z(x-\tau-s_c)}\sqrt{1-(2x-1)^2}dx. \tag{A.102}$$

This result only accounts for the portion of photons clipped by the "right" clipping portion, and the fractional area of the detector shadowed is then given by

$$\begin{aligned} F_{R,C} &= \frac{A(x^*)}{\pi\left(\frac{1}{2}\right)^2} \\ &= \frac{4}{\pi}A(x^*) \\ &= [F(x^*+s_c)-F(x^*)]-[\widetilde{F_1}(x^*+s_c)-\widetilde{F_2}(x^*)]. \end{aligned} \tag{A.103}$$

Similarly, the left clipped portion is determined by

$$\begin{aligned} F_{L,C}(x^*) &= F_{R,C}(1-(x^*-a'/c)) \\ &= F_{R,C}(1-x^*+a'/c) \\ &= F(1-x^*+a'/c+s_c)-F(1-x^*+a'/c) \\ &\quad -\widetilde{F_1}(1-x^*+a'/c+s_c)+\widetilde{F_2}(1-x^*+a'/c). \end{aligned} \tag{A.104}$$

Thus, the total fraction of attenuation due to the clipping portions of the bar is

$$\begin{aligned} F_C &= [F(x^*+s_c)-F(x^*)]+[F(1-x^*+a'/c+s_c)-F(1-x^*+a'/c)] \\ &\quad -[\widetilde{F_1}(x^*+s_c)-\widetilde{F_2}(x^*)] \\ &\quad -[\widetilde{F_1}(1-x^*+a'/c+s_c)-\widetilde{F_2}(1-x^*+a'/c)]. \end{aligned} \tag{A.105}$$

For multiple bars, the entire clipping contribution is

$$\begin{aligned} G_2(t) &= \sum_n F_C(t) \\ &= \sum_n \left(F\left(x_n^*(t)+\frac{|s(t)|}{c}\right)-F(x_n^*(t)) \right. \\ &\quad +F\left(1-x_n^*(t)+\frac{a'(t)}{c}+\frac{|s(t)|}{c}\right)-F\left(1-x_n^*(t)+\frac{a'(t)}{c}\right) \\ &\quad -\left[\widetilde{F_1}\left(x_n^*(t)+\frac{|s(t)|}{c}\right)-\widetilde{F_2}(x_n^*(t))\right] \\ &\quad \left. -\left[\widetilde{F_1}\left(1-x_n^*(t)+\frac{a'(t)}{c}+\frac{|s(t)|}{c}\right)-\widetilde{F_2}\left(1-x_n^*(t)+\frac{a'(t)}{c}\right)\right]\right). \end{aligned} \tag{A.106}$$

Since the effective width of bar shadows increases with inclusion of the clipping portion, it is necessary to modify N (where $1+2N$ equals the total number of bars considered at one time) by changing the numerator to $a'+2s_{\max}+c$. Since

$$a = a' + s_{\max}, \tag{A.107}$$

where $s_{\max} = h \tan\theta$, then

$$N = \text{floor}\left(\frac{a + s_{\max} + c}{a + b}\right). \tag{A.108}$$

A.4 Reduction

In an effort to reduce and reorganize the characteristic formula into a more readily intuitive form, Eq. A.82 and Eq. A.106 may be split into four separate shadowing parameters:

$$P_d(t) = I_0\left(1 - \sum_n \left[G_0(t) - G_1(t) - G_2(t)\right]\right), \tag{A.109}$$

where $G_0(t)$ refers to the shadow by full attenuation across the entire bar shadow width (full and clipping portions):

$$G_0(t) = F_0\left[x_n^*(t) + \frac{|s(t)|}{c}\right] - F_0\left[x_n^*(t) - \frac{a'(t)}{c} - \frac{|s(t)|}{c}\right], \tag{A.110}$$

and the function $F_0[\tau]$ is a renamed version of $F[\tau]$ for clarity. $G_1(t)$ and $G_2(t)$ constitute the remaining parameters from Eq. A.82 and Eq. A.106, which describe the transmission of photons through the constant thickness and clipping regions of the bar shadows, respectively,

$$G_1(t) = \exp\left[-\frac{h\sigma\rho}{\cos\theta}\right]\left(F_0[x_n^*(t)] - F_0\left[x_n^*(t) - \frac{a'(t)}{c}\right]\right), \tag{A.111}$$

$$G_2(t) = F_1\left[x_n^*(t), t\right] + F_1\left[1 - x_n^*(t) + \frac{a'(t)}{c}, t\right], \tag{A.112}$$

and the following symmetrical relation has been used:

$$F\left[1 - x_n^*(t) + \frac{a'(t)}{c} + \frac{|s(t)|}{c}\right] - F\left[1 - x_n^*(t) + \frac{a'(t)}{c}\right] = F\left[x_n^*(t) - \frac{a'(t)}{c}\right] - F\left[x_n^*(t) - \frac{a'(t)}{c} - \frac{|s(t)|}{c}\right]. \tag{A.113}$$

The function $F_1(\tau)$,

$$F_1[\tau,t] = \frac{4}{\pi}\int_{M[\tau]}^{M[\tau+\frac{|s(t)|}{c}]} e^{-Z(t)\left(\tau-x+\frac{|s(t)|}{c}\right)}\sqrt{1-(2x-1)^2}dx, \tag{A.114}$$

defines the integration of an exponential (shadow) across the surface of the circular detector.

A.5 Approximation

Eq. A.114 has no closed form solution, and is difficult to solve numerically in software. Instead, a parameter α is defined to describe the ratio of the integrated transmission of the exponential curve and the integrated transmission from τ to $\tau + s_c$, ignoring the circular shape of the detector:

$$\begin{aligned}\alpha &= \frac{\int_{\tau}^{\tau+s_c} e^{-Z(t)[\tau+s_c-x]}dx}{(\tau-s_c-\tau)\times(1-0)} \\ &= \frac{1}{s_c}\int_{\tau}^{\tau+s_c} e^{-Z(t)[\tau+s_c-x]}dx.\end{aligned} \tag{A.115}$$

$Z(t)$ may be described in terms of $|s(t)|$:

$$\begin{aligned}Z(t) &= \frac{c\sigma\rho}{\sin\theta|\cos(\xi(t)+\phi)|}\times\frac{h\tan\theta|\cos(\xi(t)+\phi)|}{|s(t)|} \\ &= \frac{h\sigma\rho}{(|s(t)|/c)\cos\theta} \\ &= \frac{h\sigma\rho}{s_c\cos\theta}.\end{aligned} \tag{A.116}$$

Then α is computed:

$$\begin{aligned}
\alpha &= \frac{1}{s_c}\int_{\tau}^{\tau+s_c} e^{-\frac{h\sigma\rho}{s_c\cos\theta}(\tau+s_c-x)}dx && \text{(A.117)}\\
&= \frac{1}{s_c}e^{-\frac{h\sigma\rho}{s_c\cos\theta}(\tau+s_c)}\int_{\tau}^{\tau+s_c} e^{\frac{h\sigma\rho}{s_c\cos\theta}x} && \text{(A.118)}\\
&= \frac{1}{s_c}e^{-\frac{h\sigma\rho}{s_c\cos\theta}(\tau+s_c)}\left[\frac{s_c\cos\theta}{h\sigma\rho}\right]\left(e^{\frac{h\sigma\rho}{s_c\cos\theta}(\tau+s_c)} - e^{\frac{h\sigma\rho}{s_c\cos\theta}(\tau)}\right) && \text{(A.119)}\\
&= \frac{1}{s_c}\left[\frac{s_c\cos\theta}{h\sigma\rho}\right]\left(e^{\frac{h\sigma\rho}{s_c\cos\theta}(\tau+s_c-\tau-s_c)} - e^{\frac{h\sigma\rho}{s_c\cos\theta}(\tau-\tau-s_c)}\right) && \text{(A.120)}\\
&= \frac{1}{s_c}\left[\frac{s_c\cos\theta}{h\sigma\rho}\right]\left(1 - e^{-\frac{h\sigma\rho}{\cos\theta}}\right) && \text{(A.121)}\\
&= \frac{\cos\theta}{h\sigma\rho}\left(1 - e^{-\frac{h\sigma\rho}{\cos\theta}}\right). && \text{(A.122)}
\end{aligned}$$

The percentage is constant for all values of s_c. This approximation is implemented in Eq. A.112:

$$\begin{aligned}
G_2(t) &= \alpha\left(F_0\left[x_n^*(t) + \frac{|s(t)|}{c}\right]\right. && \text{(A.123)}\\
&\quad \left. -F_0\left[x_n^*(t)\right] + F_0\left[x_n^*(t) - \frac{a'(t)}{c}\right] - F_0\left[x_n^*(t) - \frac{a'(t)}{c} - \frac{|s(t)|}{c}\right]\right). && \text{(A.124)}
\end{aligned}$$

Eq. A.114 may be rewritten as

$$F_1[\tau,t] = e^{-\frac{h\sigma\rho}{\cos\theta}}\frac{4}{\pi}\int_{M[\tau]}^{M[\tau+s_c]} e^{-Z(t)(\tau-x)}\sqrt{1-(2x-1)^2}dx. \qquad \text{(A.125)}$$

Since this constant exponential term is outside the integral, F_1 may be redefined as

$$F_1[\tau,t] = \frac{4}{\pi}\int_{M[\tau]}^{M[\tau+s_c]} e^{-Z(t)(\tau-x)}\sqrt{1-(2x-1)^2}dx. \qquad \text{(A.126)}$$

The profile formula may be expressed in terms of F_0 and F_1:

$$P(t) = I_0 \left[1 - \sum_n \left(\left(1 - e^{\frac{-h\sigma\rho}{\cos\theta}} \right) \left(F_0[x_n^*(t) + s_c] - F_0[x_n^*(t) - a_c - s_c] \right) \right.\right.$$
$$- e^{-\frac{h\sigma\rho}{\cos\theta}} \left(F_1[x_n^*(t)] - (F_0[x_n^*(t) + s_c] - F_0[x_n^*(t)] \right)$$
$$\left.\left. + F_1[1 - x^*(t) + a_c] - (F_0[1 - x^* + a_c + s_c] - F_0[1 - x^* + a_c]) \right) \right]. \quad \text{(A.127)}$$

To better organize this result, F_1 is redefined again:

$$F_1[\tau, t] = \frac{4}{\pi} \int_{M\tau}^{M[\tau + s_c]} e^{-Z(t)(\tau - x)} \sqrt{1 - (2x - 1)^2} dx + F_0[\tau] - F_0[\tau + s_c]. \quad \text{(A.128)}$$

Then,

$$P(t) = I_0 \left[1 - \sum_n \left(\left(1 - e^{\frac{-h\sigma\rho}{\cos\theta}} \right) \left(F_0[x_n^*(t) + s_c] - F_0[x_n^*(t) - a_c - s_c] \right) - \right.\right.$$
$$\left.\left. e^{-\frac{h\sigma\rho}{\cos\theta}} (F_1[x_n^*(t)] + F_1[1 - x_n^*(t) + a_c]) \right) \right]. \quad \text{(A.129)}$$

Based on Eq. A.129, G functions are also redefined:

$$G_0(t) = \left(1 - e^{-\frac{h\sigma\rho}{\cos\theta}} \right) \left(F_0[x_n^*(t) + s_c] - F_0[x_n^*(t) - a_c - s_c] \right), \quad \text{(A.130)}$$
$$G_1(t) = e^{-\frac{h\sigma\rho}{\cos\theta}} (F_1[x_n^*(t)] + F_1[1 - x_n^*(t) + a_c]), \quad \text{(A.131)}$$

to satisfy

$$P(t) = I_0 \left[1 - \sum_n \left(G_0(t) - G_1(t) \right) \right], \quad \text{(A.132)}$$

where G_0 describes the shadowing by the entire bar with attenuation accounted for, and G_1 desribes the transmission through the clipping portion.

In consideration of the approximation α (Eq. A.122) the transmission coefficient is now

Table A.1: Definition of parameters in the advanced characteristic formula.

Var	Description	Var	Description
a	RM Bar Width	σ	Grid Cross Section for Specified Energy
b	RM Bar Spacing	x_0	Detector X Location Relative to Grid Axis
c	Circular Detector Diameter	y_0	Detector Y Location Relative to Grid Axis
h	RM Bar Thickness	ω	Grid Rotation Frequency
L	Grid Height	θ	Source Zenith
ρ	Grid Density	ϕ	Source Azimuth
		I_0	Source Intensity

accounted for in G_0, and so must be subtracted so that α becomes

$$\alpha = \frac{cos\theta}{h\sigma\rho}\left(1 - e^{-\frac{h\sigma\rho}{\cos\theta}}\right) - e^{-\frac{h\sigma\rho}{\cos\theta}}. \tag{A.133}$$

A.5.1 Summary

An RM and source of interest are defined by the parameters described in Table A.1. The count profile on detector d is given by

$$P_d(t) = I_0\left(1 - \sum_n \left[G_0(t) - G_1(t)\right]\right), \tag{A.134}$$

where $G_0(t)$ represents the fraction of the detector shadowed, weighted by the amount of attenuation by the grid bar(s) at time t. This parameter accounts for the apparent expansion and contraction of the bar shadow widths depending on source zenith and grid angle,

$$G_0(t) = \left(1 - e^{-\frac{h\sigma\rho}{\cos\theta}}\right)\left(F_0[x_n^*(t) + s_c] - F_0[x_n^*(t) - a_c - s_c]\right), \tag{A.135}$$

where $F_0[\tau]$ describes the fraction of circular area integrated for a chord swept a unit diameter distance τ across the face:

$$F_0(\tau) = \frac{1}{\pi}\cos^{-1}(1 - 2M[\tau]) - \frac{2}{\pi}(1 - 2M[\tau])\sqrt{M[\tau] - M[\tau]^2}, \tag{A.136}$$

and $M[\tau]$ represents the constraint formula

$$M[\tau] = \min\{\max\{\tau, 0\}, 1\}. \tag{A.137}$$

The parameter $G_1(t)$ accounts for the additional transmission of photons through the clipping portions of the bars, and is given by

$$G_1(t) = e^{-\frac{h\sigma\rho}{\cos\theta}} \left(F_1[x_n^*(t), t] + F_1 \left[1 - x_n^*(t) + \frac{a'(t)}{c}, t \right] \right), \tag{A.138}$$

where

$$F_1[\tau, t] = \frac{4}{\pi} \int_{M\tau}^{M[\tau+s_c]} e^{-Z(t)(\tau-x)} \sqrt{1-(2x-1)^2} dx + F_0[\tau] - F_0[\tau + s_c], \tag{A.139}$$

and

$$Z(t) = \frac{c\sigma\rho}{\sin\theta |\cos(\xi(t) + \phi)|}. \tag{A.140}$$

Since Eq. A.138 has no closed form solution, a suitable approximation may be made by

$$\widetilde{G_1}(t) = \alpha \left(F_0 \left[x_n^*(t) + \frac{|s(t)|}{c} \right] - F_0 \left[x_n^*(t) \right] + F_0 \left[x_n^*(t) - \frac{a'(t)}{c} \right] \right.$$
$$\left. - F_0 \left[x_n^*(t) - \frac{a'(t)}{c} - \frac{|s(t)|}{c} \right] \right), \tag{A.141}$$

where α describes the approximate fractional transmission difference due to lack of attenuation through the clipping portion,

$$\alpha = \frac{cos\theta}{h\sigma\rho} \left(1 - e^{-\frac{h\sigma\rho}{\cos\theta}} \right) - e^{-\frac{h\sigma\rho}{\cos\theta}}. \tag{A.142}$$

Variable $x_n^*(t)$ is an array of fractional distances across the detector diameter,

$$x_n^*(t) = \frac{1}{c} \left(\frac{1}{2} \left(a'(t) + c \right) - \left| |x'(t)| \bmod (a+b) - n\frac{a+b}{2} \right| \right), \quad n = 1 + 2\{-N...N\}, \tag{A.143}$$

where

$$N = \text{ floor}\left(\frac{a + s_{\max} + c}{a + b}\right), \tag{A.144}$$

and $2N + 1$ is the total number of bars that must be considered at a time. The width of the constant thickness region of the bar shadow is

$$a'(t) = a - |s(t)|, \tag{A.145}$$

where s is the change in the shadow width due to clipping,

$$s(t) = h \tan\theta \cos(\xi(t) + \phi), \quad \text{and} \quad s_{\max} = h \tan\theta. \tag{A.146}$$

The x-component of the shadow location, corrected for shadow lengthening, is

$$x'(t) = r \cos(\xi(t) + \xi_0) + \frac{s}{2}. \tag{A.147}$$

The distance from the detector center to the projected shadow axis is

$$r = \left[(x_0 + L \tan\theta \cos\phi)^2 + (y_0 + L \tan\theta \sin\phi)^2\right]^{\frac{1}{2}}, \tag{A.148}$$

the angular orientation of the grid is given by

$$\xi(t) = \omega t, \tag{A.149}$$

and ξ_0 is

$$\xi_0 = \tan^{-1}\left[\frac{y_0 + L \tan\theta \sin\phi}{x_0 + L \tan\theta \cos\phi}\right]. \tag{A.150}$$

Appendix B

Derivation of Sensitivity

B.1 Non-Imaging

For a counting detector, the SNR of an observation is the ratio of the total number of source counts and the uncertainty in the data. For a source with measured strength S and background rate B, the total number of measured counts, O, is given by

$$O = (S + B)T, \tag{B.1}$$

where T is the length of the exposure. If the uncertainty in the data is assumed to be due only to Poisson statistics, then the uncertainty in O is

$$\mathrm{U}[O] = \sqrt{(S + B)T}. \tag{B.2}$$

The total measured counts due to the source only will be referred to as o, so that

$$o = ST. \tag{B.3}$$

Since the uncertainty in the measurement must include the noise due to background,

$$\mathrm{U}[o] = \mathrm{U}[O]. \tag{B.4}$$

The SNR of an observation is then

$$\begin{aligned} \mathrm{SNR}^{(\mathrm{NI})} &= \frac{o}{\mathrm{U}[o]} \\ &= \frac{o}{\mathrm{U}[O]} \\ &= \frac{ST}{\sqrt{(S+B)T}} \\ &= S\left(\frac{T}{S+B}\right)^{1/2}. \end{aligned} \tag{B.5}$$

B.2 Analytical Derivation

An indirect multiplexing instrument does not directly detect the source S, but instead, the instrument response performs some operation on the incident photons before or during measurement so that the sum of measurements i can be correlated to sky location j. In general, the observation by a multiplexing instrument is given by

$$O_i = \left(\sum_j s_j P_i^j + b\right) T, \tag{B.6}$$

where P_i^j is the instrument response, which describes this correlation of image and measurement spaces. The vector s_j is the measured source rate per detection element or bin for a source at each location j if all incident photons were detected (no mask or other modulation); i.e., $S_j = N s_j$ is the object scene for N detection elements. Similarly, b is the background rate per detection element.

When determining the SNR for a multiplexing instrument, it must be calculated with

respect to a particular observation at object scene j. Therefore, for this calculation,

$$s_j = \begin{cases} s_{j_0} & \text{for } j = j_0 \\ 0 & \text{otherwise,} \end{cases} \tag{B.7}$$

where j_0 is the particular element of the object scene under analysis. This assertion is equivalent to the statement that the SNR for a point is defined when it is the only source in the image. The observation for element j_0 becomes

$$O_i(j_0) = \left(s_{j_0} P_i^{j_0} + b\right) T. \tag{B.8}$$

The background-subtracted source observation is just

$$o_i(j_0) = s_{j_0} P_i^{j_0} T. \tag{B.9}$$

To calculate the SNR of a measurement, the measurement must be converted to image space. A cross-correlation is performed on the source observation with a pre-determined conversion matrix/vector $G_i^{j_0}$:

$$\begin{aligned} C_{j_0} &= \sum_i o_i G_i^{j_0} \\ &= \sum_i s_{j_0} P_i^{j_0} T G_i^{j_0} \\ &= s_{j_0} T \sum_i P_i^{j_0} G_i^{j_0}. \end{aligned} \tag{B.10}$$

The cross-correlation is now in units of sky elements, j, allowing for the significance of a

particular pixel of C to be determined. The uncertainty in C_{j_0} is

$$\begin{aligned}
\mathrm{U}[C_{j_0}] &= \left(\sum_i \mathrm{U}\left[o_i G_i^{j_0}\right]^2\right)^{1/2} \\
&= \left(\sum_i \mathrm{U}\left[O_i G_i^{j_0}\right]^2\right)^{1/2} \\
&= \left(\sum_i O_i {G_i^{j_0}}^2\right)^{1/2} \\
&= \left(\sum_i bT {G_i^{j_0}}^2 + \sum_i s_{j_0} P_i^{j_0} T {G_i^{j_0}}^2\right)^{1/2} \\
&= \left(bT \sum_i {G_i^{j_0}}^2 + s_{j_0} T \sum_i P_i^{j_0} {G_i^{j_0}}^2\right)^{1/2}. \qquad \text{(B.11)}
\end{aligned}$$

The SNR of an observation is thus

$$\begin{aligned}
\mathrm{SNR}[C_{j_0}] &= \frac{C_{j_0}}{\mathrm{U}[C_{j_0}]} \\
&= \frac{s_{j_0} T \sum_i P_i^{j_0} G_i^{j_0}}{\left(bT \sum_i {G_i^{j_0}}^2 + s_{j_0} T \sum_i P_i^{j_0} {G_i^{j_0}}^2\right)^{1/2}} \\
&= s_{j_0} T^{1/2} \frac{\sum_i P_i^{j_0} G_i^{j_0}}{\left(b \sum_i {G_i^{j_0}}^2 + s_{j_0} \sum_i P_i^{j_0} {G_i^{j_0}}^2\right)^{1/2}}. \qquad \text{(B.12)}
\end{aligned}$$

The goal now is to determine the function $G_i^{j_0}$. Ultimately, the goal is to solve the inverse multiplexing problem to return from measurement space to image space. It is therefore obvious that some form of the instrument response is needed. However, one can not simply multiply the observation by a matrix with varying values, for this will alter the caculation of the statistical significance. One must, therefore, demand that the average of the function $G_i^{j_0}$ be zero:

$$\overline{G_i^{j_0}} = 0. \qquad \text{(B.13)}$$

The solution to this problem is to use the instrument response, and subtract its mean across

elements i:

$$G_i^{j_0} = P_i^{j_0} - \overline{P^{j_0}}. \tag{B.14}$$

Eq. B.14 is plugged into the calculation of SNR:

$$\begin{aligned}
\mathrm{SNR}[C_{j_0}] &= s_{j_0} T^{1/2} \frac{\sum_i P_i^{j_0}(P_i^{j_0} - \overline{P^{j_0}})}{\sqrt{b\sum_i (P_i^{j_0} - \overline{P^{j_0}})^2 + s_{j_0}\sum_i P_i^{j_0}(P_i^{j_0} - \overline{P^{j_0}})^2}} \\
&= \frac{s_{j_0} T^{1/2}\left(\sum_i {P_i^{j_0}}^2 - \sum_i P_i^{j_0}\overline{P^{j_0}}\right)}{\sqrt{b\sum_i \left({P_i^{j_0}}^2 + \overline{P^{j_0}}^2 - 2P_i^{j_0}\overline{P^{j_0}}\right) + s_{j_0}\sum_i P_i^{j_0}\left({P_i^{j_0}}^2 + \overline{P^{j_0}}^2 - 2P_i^{j_0}\overline{P^{j_0}}\right)}} \\
&= \frac{s_{j_0} T^{1/2}\left(N\overline{{P^{j_0}}^2} - N\overline{P^{j_0}}^2\right)}{\sqrt{b\left(N\overline{{P^{j_0}}^2} + N\overline{P^{j_0}}^2 - 2N\overline{P^{j_0}P^{j_0}}\right) + s_{j_0}\sum_i \left({P_i^{j_0}}^3 + P_i^{j_0}\overline{P_i^{j_0}}^2 - 2{P_i^{j_0}}^2\overline{P^{j_0}}\right)}} \\
&= \frac{s_{j_0} T^{1/2} N\left(\overline{{P^{j_0}}^2} - \overline{P^{j_0}}^2\right)}{\sqrt{bN\left(\overline{{P^{j_0}}^2} - \overline{P^{j_0}}^2\right) + s_{j_0}\left(N\overline{{P_i^{j_0}}^3} + N\overline{P^{j_0}}^3 - 2N\overline{{P^{j_0}}^2}\,\overline{P^{j_0}}\right)}} \\
&= s_{j_0} T^{1/2} N^{1/2} \frac{\overline{{P^{j_0}}^2} - \overline{P^{j_0}}^2}{\sqrt{b\left(\overline{{P^{j_0}}^2} - \overline{P^{j_0}}^2\right) + s_{j_0}\left(\overline{{P_i^{j_0}}^3} + \overline{P^{j_0}}^3 - 2\overline{{P^{j_0}}^2}\,\overline{P^{j_0}}\right)}}.
\end{aligned} \tag{B.15}$$

Recall that $s_{j_0} = S_{j_0}/N$ and $b = B/N$:

$$\begin{aligned}
\mathrm{SNR}[C_{j_0}] &= \frac{S_{j_0}}{N} T^{1/2} N^{1/2} \frac{\mathrm{var}(P_i^{j_0})}{\sqrt{\frac{B}{N}\,\mathrm{var}(P_i^{j_0}) + \frac{S_{j_0}}{N}\left(\overline{{P_i^{j_0}}^3} + \overline{P^{j_0}}^3 - 2\overline{{P^{j_0}}^2}\,\overline{P^{j_0}}\right)}} \\
&= S_{j_0} T^{1/2} \frac{\mathrm{var}(P_i^{j_0})}{\sqrt{B\,\mathrm{var}(P_i^{j_0}) + S_{j_0}\left(\overline{{P_i^{j_0}}^3} + \overline{P^{j_0}}^3 - 2\overline{{P^{j_0}}^2}\,\overline{P^{j_0}}\right)}}.
\end{aligned} \tag{B.16}$$

The number of detection elements, N, has dropped out of the expression as expected. The term $\mathrm{var}(P_i^{j_0})$ is the variance of $P_i^{j_0}$. The term in parentheses in the denominator should be simplified. Skewness of a probability distribution X, γ, with mean μ and standard deviation

σ is the third standardized moment of the distribution, given by

$$\begin{aligned}
\gamma = \text{skew}(X) &= \text{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \\
&= \frac{\text{E}[(X-\mu)^3]}{\sigma^3} \\
&= \frac{\text{E}[(X-\mu)(X^2+\mu^2-2X\mu)]}{\sigma^3} \\
&= \frac{\text{E}[X^3+X\mu^2-2X^2\mu-X^2\mu-\mu^3+2X\mu^2]}{\sigma^3} \\
&= \frac{\text{E}[X^3+3X\mu^2-3X^2\mu-\mu^3]}{\sigma^3} \\
&= \frac{\text{E}[X^3]+3\text{E}[X]\mu^2-3\text{E}[X^2]\mu-\mu^3}{\sigma^3} \\
&= \frac{\text{E}[X^3]+3\mu^3-\mu^3-3\mu\text{E}[X^2]}{\sigma^3} \\
&= \frac{\text{E}[X^3]-3\text{E}[X]\text{E}[X^2]+2\text{E}[X]^3}{\sigma^3} \\
&= \frac{\text{E}[X^3]-3\text{E}[X]\text{E}[X^2]+2\text{E}[X]^3}{\text{var}(X)^{3/2}},
\end{aligned} \tag{B.17}$$

where $\text{var}(X) = \sigma^2$. Thus,

$$\text{skew}(P_i^{j_0}) = \frac{\overline{P^{j_0}{}^3} - 3\overline{P^{j_0}}\,\overline{P^{j_0}{}^2} + 2\overline{P^{j_0}}^3}{\text{var}(P_i^{j_0})^{3/2}}. \tag{B.18}$$

Using this result, the term in the denominator of the SNR expression can be simplified:

$$\begin{aligned}
\left(\overline{P_i^{j_0}{}^3} + \overline{P^{j_0}}^3 - 2\overline{P^{j_0}{}^2}\,\overline{P^{j_0}}\right) &= \overline{P_i^{j_0}{}^3} - 3\overline{P^{j_0}}\,\overline{P^{j_0}{}^2} + 2\overline{P^{j_0}}^3 + \overline{P^{j_0}}\,\overline{P^{j_0}{}^2} - \overline{P^{j_0}}^3 \\
&= \text{var}(P_i^{j_0})^{3/2}\ \text{skew}(P_i^{j_0}) + \overline{P^{j_0}}\left(\overline{P^{j_0}{}^2} - \overline{P^{j_0}}^2\right) \\
&= \text{var}(P_i^{j_0})^{3/2}\ \text{skew}(P_i^{j_0}) + \overline{P^{j_0}}\ \text{var}(P_i^{j_0}) \\
&= \text{var}(P_i^{j_0})\left(\overline{P^{j_0}} + \text{skew}(P_i^{j_0})\text{var}(P_i^{j_0})^{1/2}\right).
\end{aligned} \tag{B.19}$$

Eq. B.19 is plugged back into the SNR expression:

$$\begin{aligned}\mathrm{SNR}[C_{j_0}] &= S_{j_0}\frac{\mathrm{var}(P_i^{j_0})T^{1/2}}{\sqrt{B\ \mathrm{var}(P_i^{j_0}) + S_{j_0}\mathrm{var}(P_i^{j_0})\left(\overline{P^{j_0}} + \mathrm{skew}(P_i^{j_0})\mathrm{var}(P_i^{j_0})^{1/2}\right)}} \\ &= S_{j_0}T^{1/2}\sqrt{\frac{\mathrm{var}(P_i^{j_0})}{B + S_{j_0}\left(\overline{P^{j_0}} + \mathrm{skew}(P_i^{j_0})\ \mathrm{stddev}(P_i^{j_0})\right)}}, \end{aligned} \tag{B.20}$$

$$\mathrm{SNR}[C_{j_0}] = S_{j_0}\ \mathrm{stddev}(P_i^{j_0})\left(\frac{T}{B + S_{j_0}\left(\overline{P^{j_0}} + \mathrm{skew}(P_i^{j_0})\ \mathrm{stddev}(P_i^{j_0})\right)}\right)^{1/2}, \tag{B.21}$$

or, more simply,

$$\mathrm{SNR}^{(\mathrm{MP})}[C_{j_0}] = S_{j_0}\sigma\left(\frac{T}{B + S_{j_0}(\mu + \gamma\sigma)}\right)^{1/2}. \tag{B.22}$$

If the instrument response P has a uniform distribution, then the skewness is zero. In this case, the SNR reduces to

$$\mathrm{SNR}^{(\mathrm{MP})}_{(\mathrm{uniformP})}[C_{j_0}] = S_{j_0}\left(\frac{T}{B + \overline{P^{j_0}}S_{j_0}}\right)^{1/2}\ \mathrm{stddev}(P_i^{j_0}), \tag{B.23}$$

where $\overline{P^{j_0}}$ is the average P, or transmission coefficient for the instrument.

If the background rate is much higher than the source rate, $B \gg S_{j_0}$, then the expression may be even further simplified to that which is presented in Durouchoux (1983).

$$\mathrm{SNR}^{(\mathrm{MP})}_{(\mathrm{highBG})}[C_{j_0}] = S_{j_0}\left(\frac{T}{B}\right)^{1/2}\ \mathrm{stddev}(P_i^{j_0}). \tag{B.24}$$

B.3 RM Sensitivity

The statistical significance for a single detector in an RM is already fully evaluated by the equation for multiplexing SNR:

$$\mathrm{SNR}^{(\mathrm{RM},1)}[C_{j_0}] = S_{j_0}\ \mathrm{stddev}(P_i^{j_0}) \left(\frac{T}{B + S_{j_0}\left(\overline{P^{j_0}} + \mathrm{skew}(P_i^{j_0})\ \mathrm{stddev}(P_i^{j_0})\right)} \right)^{1/2}. \quad \text{(B.25)}$$

However, an RM typically consists of multiple detectors, and so it is useful to calculate the combined significance of element j_0 of the image derived from all detectors.

The cross-correlation image for an RM is composed of the individual cross-correlations determined from each of the detectors:

$$\begin{aligned} C_{j_0}^{\mathrm{tot}} &= \sum_d C_{j_0}^d \\ &= \sum_d \left(\frac{s_{j_0}}{D}\right) T \left(\sum_i P_{i,d}^{j_0} (P_{i,d}^{j_0} - \overline{P_d^{j_0}}) \right), \end{aligned} \quad \text{(B.26)}$$

where d refers to the detector element, and D the total number of detectors. The source strength s_{j_0} has been divided by D to distribute the source strength amongst the entire detection area:

$$\begin{aligned} C_{j_0}^{\mathrm{tot}} &= \frac{s_{j_0} T}{D} \sum_d \left(N \left(\overline{{P_d^{j_0}}^2} - \overline{P_d^{j_0}}^2 \right) \right) \\ &= \frac{s_{j_0} T N}{D} \sum_d \left(\overline{{P_d^{j_0}}^2} - \overline{P_d^{j_0}}^2 \right) \\ &= S_{j_0} T \sum_d \frac{1}{D} \left(\overline{{P_d^{j_0}}^2} - \overline{P_d^{j_0}}^2 \right). \end{aligned} \quad \text{(B.27)}$$

Clearly, the instrument response that directly relates the observed data to the sky is

given by the summation in the above expression,

$$P'_{j_0} = \frac{1}{D} \sum_d \left(\overline{P_d^{j_0}{}^2} - \overline{P_d^{j_0}}^2 \right). \tag{B.28}$$

Thus, the sensitivity for an RM with D detectors is given by

$$\mathrm{SNR}^{(\mathrm{RM})}[C_{j_0}] = S_{j_0}\ \mathrm{stddev}(P'_{j_0}) \left(\frac{T}{B + S_{j_0} \left(\overline{P'_{j_0}} + \mathrm{skew}(P'_{j_0})\ \mathrm{stddev}(P'_{j_0}) \right)} \right)^{1/2}. \tag{B.29}$$

The properties of the distribution of P'_{j_0} are given by

$$\overline{P'_{j_0}} = \frac{1}{2}, \tag{B.30}$$

$$\mathrm{skew}(P'_{j_0}) = 0, \tag{B.31}$$

since, like the coded aperture, the the mask has 50% transmission and the distribution of the instrument response is uniform. The standard deviation, however, can be not be further reduced analytically. The final analytic expression for an RM with D detectors is given by

$$\mathrm{SNR}^{(\mathrm{RM})}[C_{j_0}] = S_{j_0}\ \mathrm{stddev} \left(\frac{1}{D} \sum_d \left(\overline{P_d^{j_0}{}^2} - \overline{P_d^{j_0}}^2 \right) \right) \left(\frac{T}{B + \frac{1}{2} S_{j_0}} \right)^{1/2}. \tag{B.32}$$

Numerically, the standard deviation of the RM instrument response is calculated to be $\mathrm{stddev}(P') \approx 0.31$, and so

$$\mathrm{SNR}^{(\mathrm{RM})}[C_{j_0}] = S_{j_0}(0.31) \left(\frac{T}{B + \frac{1}{2} S_{j_0}} \right)^{1/2}. \tag{B.33}$$

B.4 Super-Resolution

The total number of counts in the cross-correlation image, C_j, should be normalized to equal the total number of counts measured by the instrument, α:

$$\sum_{j}^{M} C_j = \alpha, \tag{B.34}$$

where M is the total number of object scene elements. If the object scene is sub-sampled at a factor of η, then all $M\eta^2$ elements must be summed:

$$\sum_{j}^{M\eta^2} C_j = \alpha. \tag{B.35}$$

The calculation for C_j is then plugged in:

$$\begin{aligned} \alpha &= \sum_{j}^{M\eta^2} \sum_{i} O_i P_i^j \\ &= \sum_{i} O_i \left(\sum_{j}^{M\eta^2} P_i^j \right) \\ &= \sum_{i} O_i \left(M\eta^2 \overline{P^j} \right) \\ &= M\eta^2 \left(\sum_{i} O_i \right) \overline{P^j}. \end{aligned} \tag{B.36}$$

The total number of measured counts α is obviously constant, yet the right side of Eq. B.36 claims this value will scale with the square of the factor of super-resolution, η^2. Clearly, one of the terms on the right must be scaled to account for this consequence. M is not an option, since it is defined as the total number of elements in the object scene at geometric resolution. Similarly, $\sum_i O_i$ is fixed by the measured data. Thus, the instrument response

function must be scaled:

$$P_i^{j_0} \rightarrow \frac{P_i^{j_0}}{\eta^2}, \tag{B.37}$$

$$\text{stddev}(P_i^{j_0}) \rightarrow \frac{\text{stddev}(P_i^{j_0})}{\eta^2}, \tag{B.38}$$

$$\overline{P_i^{j_0}} \rightarrow \frac{\overline{P_i^{j_0}}}{\eta^2}, \tag{B.39}$$

$$\text{skew}(P_i^{j_0}) \rightarrow \text{skew}(P_i^{j_0}). \tag{B.40}$$

Plugging this change into the multiplexing SNR formula reveals the effect of sub-sampling:

$$\begin{aligned} \text{SNR}^{(\text{MP})}[C_{j_0}] &= S_{j_0}\frac{\text{stddev}(P_i^{j_0})}{\eta^2}\left(\frac{T}{B + S_{j_0}\left(\frac{\overline{P^{j_0}}}{\eta^2} + \text{skew}(P_i^{j_0})\frac{\text{stddev}(P_i^{j_0})}{\eta^2}\right)}\right)^{1/2} \\ &= \frac{S_{j_0}}{\eta^2}\text{stddev}(P_i^{j_0})\left(\frac{T}{B + \frac{S_{j_0}}{\eta^2}\left(\overline{P^{j_0}} + \text{skew}(P_i^{j_0})\ \text{stddev}(P_i^{j_0})\right)}\right)^{1/2} \end{aligned} \tag{B.41}$$

B.5 Calculation of Average Mean Free Path

The average length of photon trajectory at zenith angle θ across the detector described by Fig. 5.5 is desired. Only photons that are incident on the top face of the detector are considered, else they will manifest as background signal. Considering the span α over which the trajectories achieve their maximum distance, $d_{\max}$, and the span β over which these photons clip the detector, one can obtain an average trajectory distance in terms of the individual average trajectories:

$$d_{\text{ave}} = \left(\frac{d_\alpha\alpha + d_\beta\beta}{\alpha + \beta}\right)\cos\theta, \tag{B.42}$$

where $\cos\theta$ normalizes the distance by accounting for the effective detector area for zenith angle θ. From the geometry shown in Fig. 5.5(a),

$$\begin{aligned} \alpha &= c - t\tan\theta, && \text{(B.43)} \\ \beta &= t\tan\theta, && \text{(B.44)} \\ d_\alpha &= \frac{t}{\cos\theta}, \quad \text{and} && \text{(B.45)} \\ d_\beta &= \frac{d_\alpha}{2} = \frac{t}{2\cos\theta}. && \text{(B.46)} \end{aligned}$$

Plugging in,

$$\begin{aligned} d_{ave} &= \left[\frac{(t/\cos\theta)(c - t\tan\theta) + (t/2\cos\theta)(t\tan\theta)}{c}\right]\cos\theta \\ &= \left[\frac{t}{c\cos\theta}\left((c - t\tan\theta) + \frac{1}{2}t\tan\theta\right)\right]\cos\theta \\ &= \frac{t}{c}\left(c - \frac{1}{2}t\tan\theta\right) \\ &= t\left(1 - \frac{t}{2c}\tan\theta\right). && \text{(B.47)} \end{aligned}$$

The case when zenith angle is so large that all photons clip the detector (Fig. 5.5b) must also be included, which occurs when

$$\begin{aligned} \beta &> c \\ t\tan\theta &> c \\ \theta &> \tan^{-1}\left(\frac{c}{t}\right). && \text{(B.48)} \end{aligned}$$

In this case, the maximum trajectory is

$$d_{\max} = \frac{c}{\sin\theta}. \qquad \text{(B.49)}$$

The normalized average is thus

$$\begin{aligned} d_{\text{ave}} &= \frac{d_{\max} + 0}{2} \cos\theta \\ &= \frac{c}{2\sin\theta} \cos\theta \\ &= \frac{c}{2\tan\theta}. \end{aligned} \tag{B.50}$$

The results yield a piecewise function for the average trajectory distance across the detector for any zenith angle:

$$d_{ave}(c) = \begin{cases} t\left(1 - \frac{t}{2c}\tan\theta\right), & \text{for } 0 \leq \theta \leq \tan^{-1}\left(\frac{c}{t}\right) \\ \frac{c}{2\tan\theta}, & \text{for } \theta > \tan^{-1}\left(\frac{c}{t}\right). \end{cases} \tag{B.51}$$

Appendix C

Reprint Permission for Chap. 3

Request - 04/08/2011

Dear Roman Khanna,

On 3/25/2011, my manuscript, ASTR2628, was accepted for publication in Astrophysics and Space Science. I have not yet transferred copyright to Springer, but I would like to preemptively obtain permission to reprint this manuscript in my Ph.D. dissertation. This work may be viewable on the web.

Can you provide me with permission to do so?

Thank you,
Brent Budden

Permission - 04/11/2011

Dear Brent,

you are free to use your accepted pre-proof-version of the paper in your thesis (also in the online-version) giving the full reference (incl. DOI) to the Ap&SS paper. If you need to reprint the published Springer version, please obtain permission via the permission (rightslink) button on article website.

With best regards,
Ramon Khanna

–
Dr. Ramon Khanna
Senior Editor Physics & Astronomy
Springer

Vita

Brent Scott Budden was born in New Orleans, Louisiana, in July, 1983. He was raised in the rural outskirts of Abbeville, Louisiana. In 2001, he began his undergraduate studies at the University of Pennsylvania in Philadelphia, where he earned a bachelor of arts in physics with a minor in mechanical engineering in 2005. Brent returned to Louisiana in 2005 with his wife-to-be to begin graduate research at Louisiana State University in Baton Rouge. He plans to receive the degree of Doctor of Philosophy in physics in 2011 under the supervision of Dr. Michael L. Cherry.