UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Gamma-Ray Observations of Solar Flares with RHESSI Imaging Spectroscopy and the GRIPS instrument

Permalink

https://escholarship.org/uc/item/8c98865f

Author

Duncan, Nicole

Publication Date

2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Gamma-Ray Observations of Solar Flares with RHESSI Imaging Spectroscopy and the GRIPS instrument

by

Nicole Duncan

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Steven Boggs, Chair
Professor Stuart Bale
Professor Gibor Basri
Doctor Albert Shih

Summer 2017

Gamma-Ray Observations of Solar Flares with RHESSI Imaging Spectroscopy and the GRIPS instrument

Copyright 2017
by
Nicole Duncan

Abstract

Gamma-Ray Observations of Solar Flares with RHESSI Imaging Spectroscopy and the GRIPS instrument

by

Nicole Duncan

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Steven Boggs, Chair

Solar flares can release $\sim 10^{33}$ ergs of power, accelerate particles to relativistic speeds, heat plasma to ∼15 million K and catastrophically reorganize 10^5 km long field structures in 100s – 1000s of seconds. Magnetic reconnection of large-scale field structures in the corona are thought to power flares, but the precise mechanisms that convert the stored magnetic energy into particle kinetic energy are poorly understood.

Flare spectra in the 20 keV – 10 MeV energy range are rich with information that provide a window into the underlying physics of flare particle acceleration. This hard X-ray (HXR)/gamma-ray emission can be used to understand electron and ion dynamics, particle abundances and the ambient plasma conditions in solar flares. Enhanced imaging, spectroscopy and polarimetry of flare emissions in this energy range are needed to address the current era of particle acceleration and transport questions, including: What causes the spatial separation between HXR emission generated by relativistic electrons and the gamma-ray line emission from energetic ions? How anisotropic are the relativistic electrons, and why can they dominate in the corona? How do the compositions of accelerated and ambient material vary with space and time, and why?

The *Reuven Ramaty High Energy Solar Spectroscopic Imager* (RHESSI) instrument [88], launched in 2002, provided the first combined imaging and high-resolution spectroscopy in the HXR/gamma-ray range. RHESSI's volumes of detailed study on electron-associated emission <1 MeV is in contrast to comparatively few ion-associated gamma-ray observations. Over the past two solar cycles RHESSI has imaged only five flares at the 2.2 MeV neutron-capture line and has been able to resolve ion lines in ∼30 events. My research aims to expand this small set gamma-ray flare observations by (1) using new techniques to study flares obscured by high-background counts in the existing RHESSI dataset and (2) providing new observations through the development and flight of the *Gamma-Ray Imager/Polarimeter for Solar Flares* (GRIPS) instrument.

To little girls who dream about big things, with eager eyes open to the full light of day. Though you will not read this manuscript, every time we meet I'll carry these words with me and will use them to show you how to dance your dreams into reality.

Contents

Contents **ii**

List of Figures **v**

List of Tables **xxiii**

1 Introduction **1**
1.1 Scope of Work 4

2 The Sun and Solar Flares **6**
2.1 The Sun 6
2.2 Solar Magnetic Activity 11
2.3 Solar Flares 14
2.4 Magnetic Reconnection 21
2.5 Particle Acceleration 24
2.6 The High-Energy Flare Spectrum 30
2.7 Connecting Ions and Relativistic Electrons 38
2.8 The Future of High-Energy Flare Studies 41

3 Gamma-Ray Detection and Measurement **42**
3.1 Photon Detection Basics 42
3.2 Photon Interactions with Matter 46
3.3 Semiconductor Detectors 52
3.4 Germanium Semiconductor Detectors 62
3.5 Outlook 64

4 HXR/Gamma-Ray Imaging **66**
4.1 Indirect Imaging 67
4.2 Compton Imaging 73
4.3 Laue Lens 76
4.4 Grazing-Incidence Optics 79
4.5 Fresnel Lens 81
4.6 Summary 83

5 Gamma-Ray Imaging-Spectroscopy of the 2003 October 29 X10 Flare with RHESSI **85**
5.1 The RHESSI Imager . . . 86
5.2 RHESSI Image Reconstruction . . . 90
5.3 2003 October 29 Flare . . . 91
5.4 Gamma-Ray Imaging-Spectroscopy . . . 93
5.5 Comparison to Background-Subtracted Spectroscopy . . . 97
5.6 2003 October 29 Imaging-Spectroscopy Result . . . 101

6 The Gamma-Ray Imager/Polarimeter for Solar Flares (GRIPS) Instrument **104**
6.1 Science Goals . . . 104
6.2 Instrument Overview . . . 106
6.3 Antarctic Flight Campaign . . . 111
6.4 Antarctic Flight Results . . . 114
6.5 Thermal Management . . . 120
6.6 Future Directions . . . 121

7 The GRIPS Spectrometer **123**
7.1 3D-GeD Detectors . . . 123
7.2 Spectrometer Mechanical Structure . . . 125
7.3 Flex Circuits . . . 128
7.4 Operating Voltages . . . 130
7.5 ASIC Electronics . . . 131
7.6 Status of the Spectrometer . . . 135

8 GRIPS Spectrometer Performance **137**
8.1 Efficiency . . . 137
8.2 Detector Spatial Resolution . . . 141
8.3 Energy Calibration . . . 143
8.4 Spectral Resolution . . . 150
8.5 Spectral Broadening . . . 153
8.6 Multiple-Trigger Events . . . 156
8.7 Summary . . . 160

9 GRIPS Imaging and Aspect Systems **161**
9.1 The Multi-Pitch Rotating Modulator (MPRM) Concept . . . 161
9.2 Mask Construction . . . 163
9.3 Aspect System . . . 164
9.4 Summary . . . 168

10 Conclusion **169**

A GRIPS Education and Public Outreach **171**

Bibliography **175**

List of Figures

0.1 The GRIPS team with the gondola at the Long-Duration Balloon facility outside McMurdo Station, Antarctica, during the 2015/2016 campaign. Pictured from left-to-right are: the author (Nicole Duncan), Dr. Hazel Bain, Brent Mochizuki, Dr. Pascal Saint-Hilaire (GRIPS PI), Dr. Albert Shih, Dr. Bennett Maruca, Niharkia Godbole and Dr. Amir Caspi. Not pictured are Jane Hoberman, Jerry Olsen, Dr. John Sample, Dr. Nicole Kelley, Dr. Gordon Hurford and Dr. Bob Lin.xxiv

2.1 *Left:* A composite image of the Sun on 2011 December 7 as observed by SDO [115]. The slices (from left) are images of the ∼6,000 K photosphere, the ∼1 million K transition region in EUV light, and two composite images showing coronal temperatures up to 2 million K. The final image is overlaid with a fit to the coronal magnetic field structures. *Right:* An artist's concept of the Milky Way galaxy overlaid with galactic longitude relative to the Sun, located in the Orion Spur. Image adapted from NASA/Adler/U.Chicago/Wesleyan/JPL [31]. . 6

2.2 Diagram of the Sun's layers drawn to scale. Image created by Kelvinsong [71]. 8

2.3 *Left:* Difference between the squared sound velocity predicted by the standard solar model and measured by helioseismology. The model accurately predicts velocity to .1%. *Right:* Heliosesimology measurements of the solar rotation rate as a function of radius at three latitudes. The central black line is the model value with errors represented by the shaded regions. Both images are from Kosovichev, et al. 1997 [77]. 8

2.4 *Left:* Image of the corona during a total solar eclipse on 1980 February 16 from NCAR's High Altitude Observatory (HAO). *Right:* Image of the chromosphere during a solar eclipse in France on 1999 August 11 by Luc Viatour. Both images were provided by UCAR [39]. 9

2.5 Temperature (solid line) and density (dashed line) of the solar atmosphere plotted along the height above the photosphere. The temperature in the solar corona climbs as the density falls off. Plot from Withbroe and Noyes 1977 [147]. 10

2.6 Artist's conception of the heliosphere and interplanetary space on a logarithmic distance scale with units of AU. One AU is the distance spanning the Sun to the Earth and is equivalent to 150 million km. At the time of this drawing, in 2013, the Voyager 1 probe was 125 AU and had crossed into interplanetary space. Image courtesy of NASA JPL/Caltech [70]. 11

2.7 Diagram illustrating the geometry of how a latitudinal flux tube rises and becomes twisted by the Coriolis force to produce a loop that is perpendicular to the original direction. [114]. 12
2.8 Yearly averaged sunspot number since 1610. The ∼11-year periodicity is variable between cycles, with a long period of suppressed activity in the 17th century called the Maunder Minimum when very few sunspots appeared. Plot from Schüssler, 2004 [114]. 13
2.9 False color image of coronal loops taken on 1999 Nov 6 in TRACE's 171 $\mathring{A}$ band. Image courtesy of TRACE/NASA [136]. 14
2.10 Illustration of the relative timing of flare emission energy bands during the pre-flare, impulsive and decay phases. The Hα flash occurs during the impulsive phase but peaks later. Image from Benz, 2002 [15]. 16
2.11 CSHKP model geometry for the standard model of solar flares. Image from Christe 2007 [32]. 18
2.12 Flare observations showing loop, footpoint, loop top and above the loop sources. The figure is a composite from different authors collected by Raymond et al. 2012 [110]. *(A)*: Common flare observation showing HXR footpoints from non-thermal emission (blue) and thermal SXR emission in the loop (red). From Krucker et al. 2008 [80]. *(B)*: Flare showing a non-thermal HXR source above the loop top. From Ishikawa et al. 2011 [69]. *(C)*: The same flare shown in panel (A) at a later time when a coronal loop-top source has developed but the footpoint emission has decayed. From Krucker et al. 2008 [80]. *(D)*: A rare observation showing non-thermal emission from the flare loop with no accompanying footpoint emission. From Veronig and Brown 2004 [138]. 19
2.13 *Left:* Standard model of solar eruptive events. Image from Christe 2017 [34] *Right:* Observation of a flare and CME that shows an apparent X-point and is suggestive of the standard model geometry. Image from LASCO/SOHO [82] . . 20
2.14 Diagram of interchange reconnection. This model proposes magnetic reconnection between a closed coronal loop and and open field line to explain the association of coronal jets and Type III radio bursts with flares. Image from Shibata & Magara 2011. [120] 21
2.15 Reconnection geometry for the Sweet-Parker (*left*) and Petschek (*right*) models. The field/current inflows come from the top & bottom, and the reconnected field lines outflow in the horizontal direction. The Sweet-Parker reconnection region (shown as a pink hued box of width 2L) is much smaller than the Petschek reconnection region (of width $2L^*$). Images from Zweibel & Yamada (2009) [152]. 22

2.16 Panels describing the WIND *in situ* measurement of Hall reconnection in Earth's magnetotail. *Left, top:* Diagram of the magnetotail showing the X-line and plasma sheet that the spacecraft flew through. *Left, bottom:* Diagram of the Hall geometry showing the different ion and electron diffusion regions and indicating the quadrupolar magnetic field structure that is a signature of Hall reconnection. To the bottom are indicated the magnetic field regions that WIND passed through. *Right:* Plots of density (a), velocity (b) and magnetic field strength (c-e) as measured by the WIND spacecraft as it passed through Earth's magnetotail X-line. The magnetic fields indicated in panel (d) show that the spacecraft measured the predicted fields in the Hall ion diffusion region and indicates when it detected a guide field. Plots from Oieroset et al. 2001 [98]. 25
2.17 Simulations for stochastic acceleration of particles inside magnetic islands inside the reconnection region. *Left:* Simulated particle trajectory (white curve) showing how the particle mirrors off either end of the magnetic island. *Right:* Particle energy (ϵ) as a function of its position, showing how energy increases with each mirroring event. Plots from Drake et al. 2006 [43]. 29
2.18 A HXR/gamma-ray spectrum from 3 keV to 10 MeV for the X4.8 flare on 2002 July 23, from RHESSI observations. Shown in color are the model fits to the thermal (red) & non-thermal (magneta) electron bremsstrahlung and gamma-ray line emission from accelerated ions (blue). The emission <30 keV is primarily due to thermal plasmas at ∼40 MK, and the spectrum >30 keV is a broken powerlaw dominated by accelerated electrons. Plot from Lin et al. 2011 [86]. 30
2.19 Gamma-ray model fits to the 2002 July 23 flare. Plot from Lin et al. 2003 [87]. . 31
2.20 *Left:* Gamma-ray polarimetry measurements from the 2002 July 23 and 2003 October 28 flares. The best-fit modulation curve is plotted as a solid line and the 100% modulation curve is plotted as a dashed line for comparison. Polarization was found in both flares at levels of 21%±9% and -11%±5%, respectively. *Right:* Diagram showing the detector-to-detector Compton scatter tracks that were used for analysis. Figures from Boggs et al. 2006 [20]. 33
2.21 The first spectrally resolved gamma-ray lines from a solar flare. These panels show RHESSI count spectra (crosses whose width denote binning and heights are 1 σ measurement error) for the 2002 July 23 X4.8 flare. A second peak is visible in the carbon and oxygen panels, its due to escape of a 511 keV annihilation photon (described in chapter 3). The lines are fit with a redshifted Gaussian plus bremsstrahlung and broad line contributions (thick line), the non-redshifted line is shown (thin line) for comparison. Plots from Smith et al. 2003 [127]. 34
2.22 Neon 1634 keV de-excitation line model for a downward-isotropic angular distribution of accelerated protons and α particles propagating in 30° tilt. The solid line shows the combined contributions of a narrow proton-induced (dashed) line and a broader and more heavily redshifted α-induced line (dash-dot) in an α/proton ratio of 0.5. The vertical line marks the de-excitation line for neon at rest. Plot from Smith et al. 2003 [127]. 36

2.23 *Top:* 511 keV spectra from the 2003 October 29 flare in two intervals when the line was broad (5 – 8 keV, 11:06 – 11:16 UT) and when it suddenly narrowed (∼1 keV, 11:18 – 11:30 UT), indicating that the temperature of the annihilation region dropped from $>10^5$K to $<10^4$ within 2 minutes. *Bottom:* Model fit to the broad 511 keV line shown above, also visible in this image are two peaks from the α-He^4 line complex. The model components are an orthopositronium continuum (solid), annihilation line Gaussian (dashed) and a component from the α-He^4 line complex (dotted). Plots from Share et al. 2004 [119]. 37
2.24 *Left:* Correlation study results from Shih et al. 2009 [121] showing that there is a tight relationship between the relativistic electron bremsstrahlung and the 2.2 MeV line fluence. This suggests that ions and relativistic electrons are accelerated by similar processes. The correlation is examined across two solar missions (RHESSI and SMM) and is found to extend over 3 orders of magnitude, from the largest flares to the limits of detectability by RHESSI. *Right:* The energy content of ions vs. electrons in units of 10^{30} ergs. The flares studied by Emslie et al. 2012 [48] show that the ion and electron energies are generally comparable, with the exception of two events in which the electron energy content was ∼2× higher. . 38
2.25 *Right:* RHESSI gamma-ray image in two different energy bands (200 – 300 keV & 2218 – 2228 keV) of the GOES class X17 flare on 2003 October 28, overlaid on a TRACE 195 Å image. The electron-associated bremsstrahlung emission is shown in red contours, and the imaged 2.2 MeV line is shown in blue contours. The electron and ion centroids are separated from one another by 14" ± 5" & 17"± 5" in each footpoint. Inset is a comparison of the GRIPS and RHESSI minimum beam widths illustrating how much finer resolution a GRIPS observation could have provided for this flare. Plot adapted from Hurford et al. 2006 [67]. 40

3.1 The ideal detector response for complete energy collection of a monoenergetic beam of photons. Image rendered by Cole Kelleher based on Knoll 2010 [74]. . . 44
3.2 Comparison of pulse height spectra illustrating resolution differences between a sodium iodide (NaI) scintillation detector and a lithium drifted germanium Ge(Li) detector for an Ag-110 source. Peaks are labeled in keV. Image from Philippot 1970 [103]. 45
3.3 The absorptive and scatter processes contributing to the total mass attenuation coefficient are plotted for Germanium in the range of 1 keV – 100 MeV. Shown are photoabsorption, Compton scattering (here called incoherent scattering), and pair production in the nuclear and electron fields. Plot provided by XCOM at NIST [**XCOM**]. 47
3.4 Diagram of the Compton scatter response for a beam of photons with energy E_γ. The Compton edge is the maximum energy transmitted to a recoil electron in a Compton scatter. The shape of the Compton continuum is given by the spin-averaged Klein-Nishina equation. Image rendered by Cole Kelleher based on Knoll 2010 [74]. 49

3.5 Simulated azimuthal scatter angles (χ) for a 100% polarized photon beam. The polarization angle, χ_0, can be determined from the modulation minima, P(min). Also marked are the modulation maximum P(max), zero modulation P_0, and the modulation amplitude A. Image from Zoglauer, 2006 [151]. 50
3.6 Example detector response for an intermediately sized detector wherein a monoenergetic (E_γ) beam of photons photoabsorb, Compton scatter and pair produce with incomplete energy collection. Compton scattering generates a continuum below the Compton edge, the counts between the Compton edge and photopeak result from multiple scatters through the detector without complete energy collection. Incomplete absorption of the pair-produced 511 keV photons from positron annihilation creates two additional escape peaks at 511 keV and 1.02 MeV below the photopeak. Image rendered by Cole Kelleher based on Knoll 2010 [74]. 51
3.7 Simplified illustration of how the energetically forbidden region (the bandgap) between valence and conduction bands determines material conductivity and classification. Insulators have large bandgaps typically >5 keV, semiconductor bandgaps are typically ∼1 keV and in metals the bands are effectively mixed because the valence band is not full of electrons. 53
3.8 Drift velocities for electrons (*left*) and holes (*right*) as a function of applied field strength in germanium, along the ⟨111⟩ crystal direction. The velocities at a range of temperatures are plotted and marked in Kelvin. A saturation velocity occurs when the curves plateau and the velocity no longer increases with the field. Plots from Ottaviani et al., 1975 [99]. 55
3.9 Diagram of a P/N junction with zero applied bias. The grey regions are charge neutral and the blue/red shaded regions are negatively/positively charged. Contact between the p- and n-sides naturally establishes a depletion region and creates an electric field. Reverse biasing the junction by placing a negative contact on the p-side and a positive contact on the n-side will cause the depletion region to grow. The magnitude of charge depletion on both sides is equal, so the depletion region extends farther into the less doped side (n-side in this example). Charge density (Q), electric field (E) and electric potential (V) across the P/N junction are plotted below the diagram. Image created by Adundovi and updated by Wereldburger758 on Wikimedia commons [1] 57
3.10 Simulated collected charge as a function of time for large and small pixels at 3 different interaction depths (.2L, .5L and .8L) within a detector active volume of depth L. The choice of electrode size determines the signal induced and charge collected. Figure from Barrett et al., 1995 [11] 60

3.11 *Left:* Sketch of a planar germanium cross-strip detector. A single gamma-ray enters the detector and Compton scatters through two sites within the crystal. The strip pairs (x1, y1) and (x2, y2) collect the charge pulses generated at the first and second sites. The intersection of the collecting strips locates the energy deposition sites in the 2D plane. *Right:* Real event data from the four electrodes on the detector sketched to the left. The data is fit to determine the charge arrival times. When combined with timing information from the opposite side, the depth of the interaction is determined. Figure from Vetter, 2007 [141] 64
3.12 Simulated signals on two electrodes (y_3 and x_3) in a germanium strip detector. The different panels shows how the photon interaction depth (0, 2, 4, 6, 8 and 10 mm) changes the signal timing on each electrode. The relative charge arrival times can be used to determine the interaction depth. Figure from Amman et al., 2000 [2] 65

4.1 The oldest known drawing of the camera obscura [59], published in Gemma Frisius' 1545 book De Radio Astronomica et Geometrica, describing how this single aperture imager was used to observe the solar eclipse of January 24, 1544. . . . 66
4.2 The IBIS telescope's coded mask on the INTEGRAL spacecraft. The coded apertures are distributed in a uniformly redundant array (URA) pattern. Image from Ubertini et al., 2003 [137]. 68
4.3 Schematic depicting coded-aperture imaging. In this implementation the detector (right) is smaller than the imaging mask (left), only allowing a partial shadow of the image mask to be recorded. The shadow pattern cast on the detector changes based on the source-to-mask direction. The original source distribution can be reconstructed by deconvolution of the shadow image cast onto the detectors. The mask shown is a URA, image from Busboom et al., 1998 [25]. 69
4.4 Photon flux modulation patterns generated by bi-grid collimators. *Top:* Spatially modulated flux generated by a grid-pair with slightly different grid pitches. The moire pattern phase shifts (the maximum transmission point moves) as the source location varies. *Bottom:* Temporally modulated flux generated by a grid-pair with identical pitch. As the source location varies the total transmitted flux varies. The angular resolution for both collimator types is set by the ratio of one-half the grid pitch to the separation. Image from Hurford 2010, [65]. 70
4.5 Moiré pattern generated by a single grid on the STIX instrument's X-ray imager, which is planned to fly aboard ESA's Solar Orbiter in 2017. STIX's collimator grids have slightly different pitches and orientations relative to one another. For the grid pair shown, the pitches are 666 μm/690 μm with orientations of 60°/64°. The black box denotes the size of the grids, and their resulting interference pattern is shown in red/yellow stripes denoting shadowing/transmission. The blue boxes show the CdTe sensor pixel sizes. Image from Krucker et al., 2013 [81]. 71

4.6 The raw data (top), its regularized correction (middle) and fourier component (bottom) for a single grid-pitch in an RMC. The raw data is not fully triangular because of apparent periodic motion of the source in the imaging FOV. Image from the first description of RMCs for astrophysical use by Schnopper et al., 1968 [113]. 73
4.7 *Top:* Schematic depicting a Compton scatter track that deposits energy (E1) to a recoil electron at position (x1, y1) in the first interaction and then photoabsorbs the remainder of its energy (E2) in a second detector at (x2, y2). In this example the recoil electron direction is not measured so the original photon can only be located to an annulus on the sky. *Bottom:* A superposition of 30 event circle reconstruction from a Co-60 source placed at α=20.6°, δ=-19.5°. The intersection of the circles is used to determine the source location. Both images are from vonBallmoos et al., 1989 [144]. 75
4.8 *Left:* Simulated Doppler broadening in the reconstructed origins for a Compton telescope. The colors denote angular deviation due to the momentum distribution of bound electrons in different shells of a silicon detector. *Right:* The minimum angular resolution (here denoted as "FWHM of ARM") as a function of energy for three detector materials in a Compton telescope. The minimum resolution becomes $\geq$1° for photons energies <500 keV. Images from Zoglauer, 2006 [151]. 76
4.9 Schematic of the two geometries for crystal diffraction. In the Bragg geometry (*left*) photons are primarily refracted from the crystal surface. In the Laue geometry the photons diffract along the volume of the crystal, increasing the efficiency. Image from Frontera & von Ballmoos, 2010 [54]. 77
4.10 Two Laue lens configurations wherein multiple crystal elements are combined with varying crystal materials and/or radii to achieve a wider energy bandpass than available with a single crystal. Different energies of light are diffracted through different angles onto a single detector plane. Lower-energy light is diffracted through narrower angles and E1 < E2. *Left:* This lens design is composed of concentric rings, generating discrete energy bands. *Right:* This design continually varies the radius in the so-called *Archimedes spiral* configuration, yielding a continuous energy band. Image from Frontera & von Ballmoos, 2010 [54]. 78
4.11 The reflectance as a function of grazing angle for X-rays at several energies for a single layer of nickel (solid lines) or iridium (dashed lines). Image from Glesener, 2012 [56]. 80
4.12 Diagram of the grazing incidence optical assembly from the InFocμS payload. A nested set of mirrors shallowly reflects light from two elements onto a detector plane. Image from Ogasaka et al., 2003 [97]. 81

4.13 Fresnel zone plates used as lenses to directly focus HXR/gamma-rays. *Top:* Collecting area of the zone plate, pointed toward the source. *(a):* A Fresnel Zone Plate (FZP) that uses opaque gratings to diffract light. The pitches become smaller at larger radii to deflect through larger angles. Shadowing from the opaque gratings lead to ∼10% efficiency *(b):* A Phase Zone Plate (PZP) which replaces the opaque grating with a material that induces a π phase shift to increase the throughput, increasing efficiency to ∼40%. *(c):* A Phase Fresnel Lens (PFL) which used a graded material thickness (or alternatively a graded index of refraction) to redirect all incoming light into an n=1 diffraction order, yielding a theoretical efficiency of 100%. Image from Skinner, 2001 [125]. 82

5.1 Schematic of the RHESSI telescope's imaging system. Nine subcollimators of matching grid pairs are set in front of nine monolithic germanium detectors. Also shown are the Solar Aspect System (SAS) components which provide precise pointing information used for image reconstruction. Image from Hurford et al., 2002 [68]. 86

5.2 Table of RHESSI grid parameters for the nine RMCs, from Hurford et al., 2002 [68]. 87

5.3 Simulated modulation profiles for a series of sources, inset to the right of each panel. Panel (1) shows an example source of unit flux. The successive images can be compared to the first panel with varying flux (2), position (3, 4) and source size (5, 6). Panel 7 illustrates the inverse problem at the core of RHESSI imaging: given a modulation pattern, what is the original distribution? Image from Hurford et al., 2002 [68]. 88

5.4 *Left:* Cross-section of RHESSI's co-axial detectors. The dashed line at the top indicates where the detectors are electrically segmented into front and rear sections. *Right:* The RHESSI cryostat before being integrated onto the spacecraft bus. Shown on top are the two sets of aluminum attenuators, one set engaged and sitting on top of the detectors, the other set out. Images from Smith et al., 2002 [128]. 89

5.5 Backprojection image from the 2003 October 29 flare using RHESSI's grids 7, 8 and 9 over the energy range of 5 keV – 25 keV. The characteristic sidelobes from incomplete sampling in the U,V plane are visible. The spacecraft rotation axis is marked by a white cross. Image provided by RHESSI Browser [123]. 91

5.6 *Top:* GOES X-ray bands with RHESSI observing times overlaid in blue and red bold traces, the gray traces are GOES observations without RHESSI observation times. RHESSI had good coverage of the impulsive phase (starting at ∼20:40 UT) and multiple observations during the decay. *Bottom:* RHESSI count rates, corrected for attenuator motions (attenuator states A0, A1 and A2 are labeled and shown in magenta bands). Counts are combined across multiple detectors and divided into coarse energy bands. Images provided by RHESSI Browser [123] 92

5.7 Image of the 2003 October 29 flare showing the RHESSI centroids for 200 keV – 300 keV HXR emission (red) and the 2.2 MeV line (blue) overlaid on a TRACE 195$\mathring{A}$ image. Image from Hurford et al., 2006 [67]. 93
5.8 Backprojection image from the 2003 October 29 flare using grids 6 & 9. This image indicates there is source structure on the 35.27" (grid 6) scale. To ensure that the peak of the backprojection dirty map is a direct measure of count flux, the source needs to be fully unresolved in the image. For this reason, only detector 9 (183.2") was used for gamma-ray imaging-spectroscopy in this work. 95
5.9 Spectra for the 2002 July 23 X4.8 flare (*top*) and 2003 November 02 X8 flare (*bottom*). Imaging-spectroscopy results are black crosses that denote spectral binning and measurement error. The background-subtracted spectra are shown in blue. The 2003 November 02 background-subtracted spectrum has been re-binned to match the imaging-spectroscopy binning. The 2003 November 02 event's summed model components are shown as green triangles. 98
5.10 Plots supporting the analysis of >2 MeV flux for imaging-spectroscopy of the 2002 July 23 flare. *Left:* Flare spectra integrated over two 3-minute intervals (00:20:00 – 00:32:00 UT & 00:33:00 – 00:36:00 UT) to examine if the >2 MeV count excess results from pile-up with lower-energy 300 keV – 500 keV photons. Since the flux remains constant in the >2 MeV bins but decreases for the lower-energy photons between intervals it is unlikely that pile-up is causing the excess counts. *Right*: The maximum pixel value for each backprojection map. Since the excess counts appear in the backprojected map itself they are not a result of how the map is interpreted. 100
5.11 2002 July 23 flare plots from Share et al. 2003 [118] showing the broad spectral feature that peaks at ∼450 keV, and can be attributed to α – He-4 fusion products Be-7 and Li-7. *Left:* Model fit to background-subtracted spectrum. *Right:* Expanded energy range showing the 511 keV line. The dashed line shows the Bremsstrahlung component and the dotted line is the Be-7 and Li-7 line complex model magnified by 5x. 101
5.12 *Top:* 2003 October 29 flare spectrum (black crosses) derived via imaging-spectroscopy over the time range of 20:40:00 – 21:07:00 UT. The spectral resolution is denoted by the width of the black crosses and the height is measurement error. The results of a spectral component fit are shown as green triangles. *Bottom:* Relativistic electron bremsstrahlung and 2.2 MeV line fluences correlation plot from Shih et al., 2009 [121]. The 2003 October 29 flare is added to the original plot as a green circle, showing that this flare fits into the tight linear correlation. 103

6.1 The fully integrated GRIPS instrument payload in December 2016 at the Long Duration Balloon (LDB) Camp on the Ross Ice Shelf outside McMurdo, Antarctica. The gondola is suspended from "The Boss" launch vehicle while the GRIPS and Columbia Scientific Ballooning Facility (CSBF) teams complete pre-flight checklists during a launch attempt. The imaging mask can be seen to the right in the photo, on the end of an 8-m boom. Below the payload hangs arrays of solar panels and a telemetry boom is mounted above the payload. 105
6.2 Schematic of the fully integrated GRIPS payload. The GRIPS and CSBF components are shown in red and the three piggyback instruments (Solar-T, TILDAE and SMASH) are shown in blue. The flight computer, aspect computer, power distribution unit and the majority of the signal processing electronics are housed in the electronics bay. The electronics bay and telescope boom are both covered in opaque foam, but are shown here as transparent. The solar panel arrays are suspended below the electronics bay floor. Not pictured (on the opposite side to the electronics bay) is a bay which housed the CSBF provided Science Instrument Package (SIP) that interfaced with the antenna boom, providing the communications pipeline between the GRIPS flight computer and ground support equipment. Image from Duncan et al. 2016 [45]. 107
6.3 The GRIPS spectrometer with BGO shields removed (their outlines marked on bottom plate). The image is taken from the side of the cryostat that will face Earth in flight, and look towards space. The cryostat housing is in the center of the image (gold plated housing) surrounded by the readout ASIC electronics boxes (smaller gold boxes) and their cable harnessing that is wrapped in wire mesh and teflon tape to provide electrical shielding and protection to the ribbon cables. An EMI shield is removed for the detector stack to the left, showing the density of signal traces that travel ∼25 feet in ribbon cables from the ASIC electronics to the FPGA Card Cages in the electronics bay. Protruding from the cryostat on the upper left are the feedthroughs which provide high voltage to the detectors. 108
6.4 Inside the GRIPS electronics bay prior to flight. Labeled are the primary system components, including: the flight, aspect and pointing control computers, the cryostat's temperature control board (cryoplate), the detector electronics, the power distribution unit and battery banks. 109
6.5 *Left:* The point response functions that compare the image fidelity resulting from GRIPS' 13 spatial frequencies to RHESSI's two frequencies (from grids 6 & 9) at gamma-ray energies. Note that GRIPS' response is virtually free of sidelobes. *Right:* The single MPRM imaging grid that contains 13 grid pitches for quasi-continuous sampling between 12.5" – 162". Mounted along the grid diameter are three optical lenses that project images of the Sun onto a fiducial plane as part of the aspect system that provides high-precision monitoring of the grid location and orientation (section 9.3). Images from Duncan et al. 2016 [45]. 110

6.6 *Top:* GRIPS flight path with time represented by the color gradient progressing from red, at launch, to blue on landing. Numerical annotation indicates the number of days at float. *Bottom:* Balloon altitude from launch with the color progression indicating time at float. Images from Duncan et al. 2016 [45]. 112
6.7 The author with the GRIPS payload in February 2016 after flight. Damage to the spectrometer's protective cage and bottom radiator plates suggest that the boom was not horizontal during descent and the spectrometer end impacted first before the payload flipped over to its final resting place. 113
6.8 GOES X-ray lightcurve for the duration of the GRIPS flight (marked by the horizontal black line). The vertical dashed line indicates the largest solar flare, GOES class C9.6, and the shaded areas indicate periods of precipitation from Earth's radiation belts. Image adapted from Duncan et al. 2016 [45] 115
6.9 Spectrograms for the RHESSI detectors during the C9.6 flare on 2016 Jan 28. Three of RHESSI's detectors (D1, D5 and D7) observed the flare. The spectra show that the majority of counts are below 10 keV. At ∼11:55 UT an attenuator comes in to reduce the low-energy flux. The maximum energy photons detected by RHESSI for this flare are ∼30 keV. Plots provided by RHESSI Browser [123]. 116
6.10 Distribution of low-energy thresholds in flight. To generate this plot the flight data was integrated over multiple days and the spectral cutoffs were fit to spectra across detectors 0 and 4. An example flight background spectrum is shown in figure 6.13. The majority of spectral cutoffs for detectors 0 and 4 range between 40 – 60 keV. 117
6.11 Detector 0 and detector 4 count rates integrated over 5 minute bins for the C9.6 flare that occurred on Jan 28, 2016 at ∼11:45 UT. The flare onset and ends of the impulsive and decay phases are marked with dashed lines. No obvious count rate excess occurs in either detector after the flare onset. D4's variability on this timescale comes from the duty cycling behavior discussed in section 8.1. Events with a maximum of two triggered strips per side were selected for this plot. . . . 117
6.12 Spectrograms spanning 11:21 UT - 13:06 UT for GRIPS detector 0 in 5-minute bins over the C9.6 flare that occurred on 2016 Jan 29 at ∼11:45 UT. The top plot shows that the majority of emission observed is in the 100 keV – 250 keV range (from background) and the bottom plot zooms in on the lower end of the energy portion of the range shown in the RHESSI plots (figure 6.9). These plots show that GRIPS has no obvious spectral changes during the flare impulsive phase in comparison to the pre-flare time period. GRIPS' spectral cutoff was ∼50 keV in flight and RHESSI primarily recorded flare emission <30 keV. 118
6.13 Antarctic background for a single strip on detector 0 integrated over multiple days in flight. The steep low-energy spectral cutoff is imposed by triggering thresholds in the electronics. To the far left is the noise pedestal at zero energy. A dashed line marks the predicted 511 keV line which is in good agreement with the peak seen in the background data (blue curve), after applying the ground-based calibration. 119

6.14 *Left:* The GRIPS gondola suspended from a custom-built truss (blue structure) inside NASA Plum Brook's B2 vacuum chamber for the 30 hr thermal vacuum test. *Right:* Banks of IR lamps were used to simulate the Earth's albedo from below and direct solar illumination from the front (not pictured). Pictures from Duncan et al. 2016 [45]. 121

7.1 The GRIPS spectrometer with top radiator plate removed. The main components of the spectrometer are diagrammed including the cryostat housing, cryocooler, cryocooler heatsink assembly (which includes thermoelectric coolers and a finned heatsink), ASIC electronics boxes, signal cables, BGO anti-coincidence shields, and side radiator plates. The spectrometer is on-axis pointed towards the sun along the telescope boom, in the direction marked "forward to imager". The picture is taken from the side of the cryostat that faces space, looking towards the Earth. 124

7.2 A single GRIPS 3D-GeD detector. Inset is a portion of the detector magnified to show the individual parallel electrodes. The detector is wire bonded to breakout boards, which connect to flex circuits that carry the signal through the cryostat shell, to exterior processing electronics. The HV side board is large because it also contains the circuits necessary to provide a steady voltage to that side's electrode strips. The LV boards are comparatively smaller because those strips are held at ground. Numbered are the electrode strips from 1 to 149 on each side, and a guard ring is labeled as 0. The detector edges are intentionally rounded and the guard ring is visible along the edges of the germanium face. Picture from Duncan et al. 2013 [44]. 125

7.3 When a photon interacts with the detector a cloud of electron/hole pairs are generated. The electrons and holes move towards the electrodes which supply a bias voltage across the crystal, their arrival locations and arrival times can be used to determine the site of photon interaction. *Left:* The 2D planar coordinates (X, Y) are determined by matching coincident signals from electrons on the HV strips and holes on the LV strips. *Right:* The interaction depth (Z) is inferred by the relative arrival times between holes (t_h) and electrons (t_e) at the detector strips. Since the drift times of both species are well-known for a given detector type and applied bias, the distance traveled by the charges through the crystal can be determined and the site of photon interaction can be inferred by taking the difference of their arrival times. 126

7.4 A 2D hitmap generated by matching coincident HV and LV electrode strips from a Cs-137 source in detector 0. The events shown triggered only one strip on each detector face and the image was flat-fielded. There are no counts in the corners (besides noise events) because the detector itself has rounded edges. Triggering efficiency and multiple-trigger noise led to the relatively dark electrodes that received few source triggers (40 LV strips and 10 HV strips). Visible in this image are two crystal defects in the upper left and right corners. 127

7.5 *Left:* The interior of the cryostat, integrated with two test detectors for preliminary testing. The detector stacks are housed inside an infrared shield (gold) then wrapped in thermal blankets (silver) before being inserted into the aluminum cryostat shell. The detectors are cooled by a copper cold finger. Shown in the foreground are two activated carbon molecular sieves used to enhance the vacuum. *Right:* The full flight detector stack integrated with electronics boards. Pictures from Duncan et al. 2016 [45]. 128
7.6 Diagrams showing the detector (green boxes) and BGO shield (blue shapes) placement for flight. *Top*: View of the detectors looking towards the rear of the boom, away from the Sun. The bottom of the image points toward the Earth and the BGO are arranged along the bottom to protect against interference from Earth's albedo and atmosphere. The detectors are labeled D0 – D5 and have individual reference names. *Bottom*: View looking along the telescope line-of-sight towards the Sun, from the rear of the boom. The rear BGO also protect against interference from Earth's albedo and atmosphere. Image courtesy of Albert Shih. . . . 129
7.7 The VATA453 ASIC chip on a GRIPS flight electronics board. The photosensitive ASICs are usually covered in a gold EMI housing, pictured to the left for an adjacent ASIC. A flex circuit is mechanically connected to the input traces that are wire bonded to the ASIC chip. The ASIC output passes through an LS64 level shifting chip (also designed by IDEAS) before exiting the board. 132
7.8 Electronics chain schematic showing how the detector is read out by the ASIC and FPGA. The ASIC handles triggering, signal shaping and the initial digitization step via a comparator. The FPGA completes the digitization by counting clock cycles while the comparator keeps the ASIC output low. Schematic provided by IDEAS. 133
7.9 Detailed schematic of the VATA453. For each detector strip input there are two parallel ASIC channels, a fast channel used for triggering (TA) and a slow channel (VA) used for charge shaping and analog-to-digital conversion. The VA-TA channels share a single output for each detector strip (total of 64 input and output channels), with triggering and conversion-mode switching handled by the FPGA. Schematic provided by IDEAS. 134
7.10 Example GRIPS barium-133 spectrum shown at three successive threshold settings of 5 (black), 9 (green) and 30 (magenta) out of a maximum setting of 31. The spectra have been normalized to emphasize the less active, higher-energy lines and illustrate the effects of thresholding. The known line energies are written on the plot in keV and the pedestal, which marks the zero-energy bin, is labeled. Plot adapted from Duncan et al. 2013 [44]. 135
7.11 Images of the spectrometer after GRIPS' flight in February 2016. Though spectrometer exterior structure and radiator plates sustained damage during landing, the cryostat, ASIC electronics and BGO were not visibly damaged. 136

8.1 Distribution of the time between events recorded by D0 and D4 for Am-241. The real source counts are Poisson-distributed in time, appearing as sloped lines on the semi-log plots. Data to the right of the dashed vertical lines were fit to exponentials whose decay constant represents the measured source rate for 59.5 keV photons. The fits used to determine the measured count rates are shown as solid blue lines. The fits have been extrapolated to the left of the fit region to show the expected source counts below the dashed vertical lines. In this region the remainder of the counts are due to noise and were excluded from the fitted exponential. These noise counts at short timescales are discussed in sec 8.6. D0's short timescale structure is zoomed in for detail in fig 8.17. 139

8.2 *Top*: Spectrograms for Cs-137 radiation at 5s intervals, showing duty cycling. Detector 0's energy dependent duty cycling occurs primarily below ∼250 keV. In contrast, D4's duty cycling is energy independent, affecting the entire spectrum uniformly. *Bottom:* Spectrograms for Am-241 radiation at 5s intervals. Detector 4's energy-independent duty cycling is consistent with data shown in the top plot for Cs-137 data. Detector 0, however, does not show the same energy-dependent state toggling as it did during the above Cs-137 data run. The linewidth difference between detector 0 & 4 is due to greater charge trapping on detector 0 (discussed in section 8.4). The lines at ~ 7 keV on D0 and ~ 2 keV on D4 are due to glitches (discussed in section 8.5). 140

8.3 Am-241 count rates for detector 4 and detector 5 normalized to emphasize their anti-correlated duty cycling. These two detectors are arranged in a stack with detector 4 closer to the source (top detector plane) than detector 5 (bottom detector plane). 141

8.4 Am-241 ground calibration data. The top plot shows the detector count rate at 1s intervals for source counts (selected for all triggers ≤ 2 per side) plotted against all triggers. The dashed horizontal line is a count rate threshold that separates the data into normal and suppressed states. *Bottom:* Correlation plots between noise triggers (all triggers - source triggers) to all triggers (*left*) and source triggers (*right*). These plots show that the normal and suppressed states are clearly two different populations that can be isolated from one another by a count rate threshold, marked as a dashed vertical line. Counts to the right of the dashed line are in the normal state and counts to the left are the suppressed state where the noise drops to zero. 142

8.5 Crossed tungsten rods were used to cast X-ray shadows along the detector diagonals. The rods can be seen as dark "X"s in figure 9.3. These plots show a the sum of counts along the detector diagonals for an Am-241 source, normalized to show two different detectors (D0 in red and D4 in blue) in single plot. Since the tungsten rods cast a shadow along a detector diagonal, the summed counts dip at the location of the rod. The width of the dip provides a measure of the detector spatial resolution. Since the diagonals do not have the same quantity of pixels, the summed counts have an overall decreasing trend. The smaller count fluctuations between diagonals are primarily due to variations in the energy thresholds for each pixel. 143
8.6 Example arrival time difference plots for Am-241 (*top*) and Cs-137 (*bottom*) sources. The difference of arrival times provides a measurement of photon interaction depth within the crystal. The ±500 ns edges are created by interactions that happen close the edge of the active volume. The lower-energy 59.5 keV photons have a shallower penetration depth than the 662 keV photons and are primarily collected near the side of the crystal that they're incident on. The higher-energy Cs-137 photons more uniformly penetrate the crystal. 144
8.7 Model fit to the 81 keV line and threshold for channel 148 on D4. The model fit is shown in green, the line and threshold fit locations are marked by dashed vertical lines. Above the line is a continuum which is approximated by a DC offset in the region surrounding the line. Only the photopeak is used for calibration and modeling the continuum accurately is not critical. 145
8.8 Am-241 spectra showing the 59.5 keV line on D0 (*left* panels) and D4 (*right* panels) for low-voltage (*top* panels) and high-voltage (*bottom* panels) sides. The spectral shape differences are due to increased trapping on detector sides facing away from the source relative to detector faces that are toward the source. In D0 the 59.5 keV radiation is incident on the HV side and on D4 it falls on the LV side. 146
8.9 Ba-133 strip spectrum for detector 0's low voltage channels 234 and 232 over the same integration time. The spectral counts are plotted in the common-mode subtracted (section 8.5) ADC bins and marked are the known line energies. The spectra have been normalized to the counts in the peak of the 81 keV line and channel 232's spectrum has been translated in ADC bins such that it aligns with channel 234. Channel 234 is electrically connected to a neighboring electrode on the detector and has ∼2× the effective area of channel 232. 147

8.10 Common-mode subtracted ADC bin values by ASIC channel on D0 and D4. The individual ADC bin values (solid dots) are nearly uniform across multiple channels within the same ASIC. Horizontal lines mark the median ADC bin value for each available line by ASIC. The line energies are 59.5 keV (purple), 81 keV (blue), 276 keV (yellow), 303 keV (magenta), 356 keV (green), 384 keV (orange) and 662 keV (red). There are only two ASICs instrumented on D4's LV side because the other two flex circuits were disabled (section 7.3). Selection cuts based on FWHM of the fitted lines reduced the set of channels that were used to determine the by-ASIC line values to the smaller set shown in these plots. 148
8.11 Plot showing how the slow shaper peaking time varies with input pulse energy for the GRIPS ASICs. The ADC bin values are given for different pulse energies as a function of sampling time on the slow shaper. The "neg" curves are pulses on the LV side, "pos" curves are pulses on the HV side and "pos on neg" is a positive charge signal on the LV side. A sampling time of 8000 us was chosen for the LV side and a time of 4000 us for the HV side. 149
8.12 *Left:* Fit residuals as percent of line energy by polynomial order for D0 channel 234. The magenta points show the polynomial deviation from linearity. *Right:* A 5th order polynomial interpolation for channel 234. 150
8.13 Fit residuals as percent of line energy by polynomial order for each ASIC on D0 and D4 LV sides. For each plot the last order plotted is the polynomial that was chosen for that ASIC's interpolation function. 151
8.14 Spectral resolution of low voltage detector strips on D0 (unsaturated colors) and D4 (saturated colors) at the six calibration line energies of 59.5 keV, 81 keV, 303 keV, 356 keV, 384 keV and 662 keV. The mode for each line is inset to each plot and given in keV with a $\pm.5$ keV error. 152
8.15 Ba-133 data for ASIC channel 152 on detector 0. This image shows how glitch rejection culls the extraneous peak at $\sim$bin 60. 154
8.16 59.5 keV spectra in raw ADC bins (labeled raw A/D bins) before common-mode subtraction is applied. The red curves are from data taken with the cryocooler on and the black curve is from data with the cryocooler off. The line is split into two lines by cryocooler effects for the channel in the upper plot and only minimally broadened for the channel in the lower plot. The two channels shown are are spatially close to one another, separated by one channel in the ASIC and by one strip on the detector. Despite the proximity of these channels to one another their responses to cryocooler noise are vastly different. 155

8.17 Distribution of the time between events recorded by D0 for Am-241, zoomed in to show the fast timescale detail from fig 8.1. The total events (dark purple points) are broken into components based on the triggering order. The legend abbreviates single-triggers as "st" and multiple-triggers as "mt". Any event that received more than 10 triggers was considered a multiple-trigger event since real source depositions can potentially trigger 10 electrodes. The components plotted are: the times between single-triggers followed by multiple-triggers (st-mt, red), multiple-triggers followed by multiple-triggers (mt-mt, green), and the combination of single-triggers followed by single-triggers and multiple-triggers followed by single-triggers (mt-st + st-st, violet). The first two components (st-mt & mt-mt) are due to multiple-trigger noise and the counts in the summed component (mt-st + st-st) are from the source. The data was fit with a decaying exponential (solid blue line) to the right of the dashed black line, to predict the source count rate. The fit has been extended into the region dominated by noise (dashed blue line) to show the expected source counts in this region. The fitted line matches the summed component (mt-st + st-st) well. 157

8.18 The total detector counts from an Am-241 source (same data as the purple curve of figure 8.1) as a function of total triggered events and the time between events. D0 labels the three different noise structures that are referenced in text. D0 was run in soft-veto multiple trigger mode (all events saved) and D4 was run in hard-veto multiple trigger mode (events with >25 triggers discarded). Though D4's hard veto caught most multiple trigger events above the threshold, occasionally some events with triggers greater than the threshold were saved. A key difference between noise on the two detectors is the persistence of events, with noise dominating the counts about 2x longer on D4 (1ms) than on D0 (.5ms). 158

8.19 The total detector counts from an Am-241 source (same data as the purple curve of figure 8.1) as a function of total triggered events and the time between events. These plots extend over a longer timescale than figure 8.18, shown to include some of the longer time between events where source counts dominate. At later times the source counts primarily trigger only a few strips, as expected for relatively the low-energy (59.5 keV) radiation in this dataset. 159

9.1 *Left:* The single modulation grid used for GRIPS' MPRM imaging system. The grid has thirteen slit/slat pitches made from 70/30 tungsten/copper slats of 2.5 cm line-of-sight thickness. The slit pitch varies from 1 – 13 mm, corresponding to thirteen imaging resolutions between 12.5 – 162 arcseconds. *Right:* Zoomed-in picture showing the slit/slat bundle arrangement on the imaging grid. The sum of all pitches in a bundle is 13mm. Transmission through the slits falls off as as line-of-sight angle increases. Pictures from Duncan et al. 2013 [44]. 162

9.2 Schematic illustration of GRIPS image reconstruction, for 1(a), 3(b), 10(c), 30(d), 100(e) and 1000(f) counts from a simulated point source. As the grid rotates photons are collected from a range of grid orientations and spatial frequencies. Image from Duncan et al. 2013 [44]. 163
9.3 *Left*: Photo of the fiducial plane (which is mounted to the top of the cryostat) during a pre-flight pointing test. In the photo are the three solar images, the fiducial plane, and the "×" patterns that are positioned above each stack of 3D-GeDs. The solar images in the lower left and in the upper right are at 90% and 80%, respectively, of the brightness of the central solar image. As the MPRM rotates, the outer solar images appear to orbit around the central solar image. *Right*: In-flight image from the pitch/yaw subsystem, with locations of the solar images (red × and circles) and illuminated fiducials (blue crosses) determined via a convolution filter. The right edge of the frame corresponds to the nadir direction, and shows faint illumination from incompletely blocked Antarctic albedo. Note also that the black tungsten rods, aligned with the "×" patterns, are illuminated by the central solar image. Pictures from Duncan et al. 2016 [45]. 165
9.4 In-flight image from the roll sensor showing the opposing horizon views and the Antarctic surface. The axes indicate pixel number. Picture from Duncan et al. 2016 [45]. 166
9.5 The roll sensor, mounted above the gondola in the orientation shown. The silvered, knife-edge, right-angle prism combines opposite horizons into a single image. 167

A.1 *Left:* Image illustrating the penetration depth of light into the atmosphere at different energies and why its possible to make ground based measurements of optical and IR light, but not HXR or gamma-rays. Balloons, rockets and satellites are needed to study high-energy light. Image courtesy of OpenStax Astronomy as a modification of a STScI/JHU/NASA image [7] *Right:* Simultaneous images of the Sun, showing how different energies of light reveal different phenomena, regions and temperatures. Image from the SDO archive [116]. 171
A.2 *Left, top:* GRIPS suspended from "The Boss" launch vehicle during a hang test. *Left, middle:* GRIPS team signatures on the cryostat, a tradition before flight. *Left, bottom:* Balloon filling with $2/3^{rd}$ of a ton of helium prior to launch. The helium tanks are visible on trucks to the right of a tractor. Image courtesy of Albert Shih. *Right:* GRIPS just after launch. The payload is suspended from a parachute (orange) attached to the balloon. After reaching the Stratosphere, the balloon will expand to 150 ft tall. Image courtesy of Albert Shih. 172

A.3 *Left:* Antarctic deed-field training where we learned about snow camping. The original recovery plan was to establish a 3 week field camp and groom an ice runway to accommodate a Basler aircraft at the landing site. Due to fuel restrictions we were instead given 3 out-and-back flights on a Twin Otter. Pictured in the background to left is Mt. Erebus, the southernmost active volcano, and Mt. Terror to right. *Right:* Since the landing site was ∼500 miles from the South Pole station, refueling the Twin Otter at fuel caches en route was necessary. Images courtesy of Hazel Bain. 173
A.4 Payload recovery in January 2017. The 800 lb. spectrometer was supported by fuel canisters and sawed off the frame onto a sled. Sliding across the ice was relatively easy but loading onto the plane via the wooden ramp required 6 people to push and manage safety straps. Hazel Bain is wearing an orange coat and the author is wearing a pink hat. 173
A.5 *Left:* Recovery team route from the GRIPS launch site on the Ross ice shelf outside McMurdo to the South Pole Station and impact site. We waited for weather at the South Pole for 4 weeks before conditions were clear for landing at the fuel cache and impact site. *Right:* The author at the South Pole station in December 2016. 174
A.6 NASA's LDB facility. Convective cooling from freezing fog during roll-outs resulted in some components being colder on the ground than in flight. 174
A.7 *Left:* The cryostat was powered during transport to the Plum Brook T-vac test and NASA systems integration. *Right:* Since no pressure vessels were used in flight, thermal straps and conductive heatsinks were used to mitigate heat on commercial electronics. 174

List of Tables

6.1 GRIPS instrument timeline showing the major project milestones and tests. . . 111

7.1 Detector numbers, crystal manufacturing numbers, depletion voltages (Depl V), operating voltages (Op V), and instrumentation on the LV and HV sides. Some detectors were not fully instrumented due to issues with the flex circuits which are discussed in section 7.3. The relative positioning of each detector is shown in figure 7.6. 130

Acknowledgments



Figure 0.1: The GRIPS team with the gondola at the Long-Duration Balloon facility outside McMurdo Station, Antarctica, during the 2015/2016 campaign. Pictured from left-to-right are: the author (Nicole Duncan), Dr. Hazel Bain, Brent Mochizuki, Dr. Pascal Saint-Hilaire (GRIPS PI), Dr. Albert Shih, Dr. Bennett Maruca, Niharkia Godbole and Dr. Amir Caspi. Not pictured are Jane Hoberman, Jerry Olsen, Dr. John Sample, Dr. Nicole Kelley, Dr. Gordon Hurford and Dr. Bob Lin.

A remarkable ensemble of people have given me the gift of their time; they have shared the knowledge, humor and kindness needed to persevere through GRIPS' and graduate school's many challenges. Thank you to the mentors who guided this metamorphosis from student-to-scientist, the team members whose dedication brought GRIPS to flight and the friends & family who provide a nourishing and supportive balance.

A few days after I was accepted to Berkeley Professor Bob Lin called me from London, where, I later realized, it was 2:00 in the morning. Anyone who knew Bob is probably laughing over that scenario like I am right now, with a small grin and some tears. I was lucky to have been his student for a few years before his passing. I look back on those early conversations and often think that he knew me better than I knew myself at that time; he devised the blueprint for the instrumentation-focused path that I am excited to continue beyond Berkeley.

Professor Steve Boggs, my graduate advisor, has provided the essential guidance I needed to stay on track in graduate school and a second home in the COSI group. Steve, you built an excellent group and I have learned a lot from you as both a scientist and a manager. Thank you for giving me just the right amount of freedom, pulling me back in when I wanted to take on too many projects, and having realistic expectations that were clearly defined.

The work in this thesis has been enriched by the mentorship of Dr. Albert Shih, my GSRP fellowship advisor at NASA Goddard SFC, the GRIPS project manager and the RHESSI Deputy Project Scientist. Albert, I had so much fun working with you through late nights, long phone conversations and strategizing solutions to the problems we were "lucky to have". Your guidance and your friendship have made the past 7 years productive and fun. Albert's detailed comments have greatly improved the readability of this manuscript and are immensely appreciated.

Dr. Pascal Saint-Hilaire, the PI of GRIPS, has been an important mentor and friend during my time at SSL. Before we began working together Pascal would go out of his way to teach me solar physics on the bus, or stop in the hallway for career conversations. These interactions went a long way towards making me feel at home in the lab.

Thank you to Professors Stuart Bale, Gibor Basri and Jonathan Wurtele for making my qualifying exam and dissertation experiences challenging and rewarding.

Teams make instruments possible. Thank you to the scientists and engineers who worked through all the challenges on GRIPS and brought it to flight. The GRIPS team (pictured and listed above) was a small but nimble group whose dedication to the project, and each other, made it possible. Thank you to Dr. Gordon Hurford for mentoring me in the imaging/aspect systems and many enlightening conversations about instrumentation. Thank you to Jane Hoberman and Brent Mochizuki for all the hours we spent together in the lab. Thank you to my van sister, Dr. Hazel Bain, for quick revisions to the GSE code, for talking with me 7 days straight on the way to Ohio, for kicking ice with me outside the high bay at LDB and for returning with me to Antarctica to recover GRIPS (and wait at the South Pole for 4 weeks). Thank you to Yvette Irwin for her attention to detail and patience with the GRIPS cabling. Thank you to Terry and Dean for being very accommodating in the machine shop.

Thank you to the COSI team for many useful discussions and sharing the highbay among other resources. A special thank you to Steve McBride who was explicitly not working on GRIPS, but who always talked through hardware issues with me anyway.

Friends inside and outside the lab brought me the support and balance I needed to be successful in graduate school and here I only name a few. Dr. Lindsay Glesner has been a dear friend and I have been so thankful to have her as a role model (though she would disagree to the role model part). Dr. Claire Thomas and Dr. Russell Avdek, thanks for moving to the neighborhood, hanging out at the lake and laughing at life with me.

When I was a little girl and told my parents that I wanted to be an astrophysicist, they took it seriously. Though not being scientists themselves, they actively supported this and many more dreams. Parents, you taught me how to dream, showed me dreams are possible, gave me the tools turn dreams into goals and filled me with the creativity needed to achieve them. Most importantly you taught me that there is more to life than relentless pursuit of achievement. Thank you.

Cole Kelleher nurtured the preparation of this thesis by providing Nora and myself with a peaceful home filled with plants, love and music. Cole, thank you for suggesting to drive my dog and I across the country so you could support us during this time, we've thrived in your loving care. I love you and I'm excited for our next adventure in Colorado!

Chapter 1

Introduction

The Sun is an astrophysical object whose proximity makes it a convenient natural laboratory to study high-energy plasma processes in stunning detail. The wealth of information learned about the Sun improves our understanding of the greater stellar population and provides a means to benchmark increasingly complex stellar modeling efforts. The exotic plasma conditions that comprise most of the universe are difficult to recreate on Earth, but have analogs in the dynamic solar atmosphere which can be observed and studied with precision. One of these features are solar eruptive events, whose *flares* and *coronal mass ejections* (CMEs) are capable of releasing $\sim 10^{33}$ ergs of power, catastrophically reorganizing 10^5 km long field structures, accelerating particles to relativistic speeds and heating plasma to $\sim$15 million K within 100s – 1000s of seconds.

Flares are the Solar System's most powerful explosions and remarkably efficient particle accelerators. They are capable of imparting up to 10s of percent of the energy released during an eruptive event into particle acceleration [48]. Ions and electrons are accelerated with roughly the same efficiency, each population receives about half of the energy that goes into particles. Magnetic reconnection has been shown to be capable of providing a mechanism to release energy from the stressed magnetic fields which power flares, but reconnection alone cannot accelerate particles to the 10s of GeV (ions) and 100s of MeV (electrons) observed during flares.

Flare spectra encode key information about energetic particle interactions, providing a window to study the properties and abundances of both accelerated and ambient populations. The most detailed gamma-ray imaging and spectroscopy to date has come from the *Reuven Ramaty High-Energy Solar Spectroscopic Imager* (RHESSI) instrument [88]. RHESSI images linked high-energy photon emission to observed spatial structures like magnetic field loops and footpoints, verifying the overall geometry of the standard flare scenario. In this picture reconnection of current-carrying magnetic fields releases energy when the stressed fields relax into lower-energy topologies. Inside and surrounding the reconnection region particles are efficiently accelerated. The particles exit the acceleration region bound to magnetic field lines, either escaping into interplanetary space or raining down into the solar atmosphere where increased density causes them to lose energy through collisions. Through these ener-

getic interactions particles emit photons in a variety of processes that impart the 20 keV – 10 MeV flare spectrum with a rich set of lines that hint at the underlying acceleration and transport mechanisms [94, 93]. A summary of the Sun, how its magnetic variability gives rise to solar flares, the process of reconnection, candidate particle acceleration mechanisms and features of the gamma-ray flare spectrum are provided in chapter 2.

RHESSI and its predecessor the *Solar Maximum Mission* (SMM) [140] provided evidence that flares accelerate ions and relativistic electrons proportionally; the ratio of their fluences has been shown to remain constant over three orders of magnitude, a correlation that extends to the limits of RHESSI's detectability and is independent of flare size [121]. This result suggests that these populations are accelerated together, and possibly by similar mechanisms.

Only five flares observed by RHESSI in the past two solar cycles were large enough to create images of the ion-associated emission. In four of these flares RHESSI found that the centroids of ion and relativistic electron-associated emission were significantly displaced from one another, by as much as 25" [66, 67]. This result is surprising; ions and electrons that are accelerated in the same region are thought to be transported out of the acceleration region along the same field lines, implying that they would enter the chromosphere together and have similar emission source locations. Unfortunately RHESSI's reduced imaging capabilities at gamma-ray energies limits these observations to the few largest flares of the past solar cycles.

To build a more complete understanding of particle acceleration in flares higher-resolution imaging with improved sensitivity >1 MeV is required. The original work in this manuscript is aimed towards expanding the small set of gamma-ray flare observations by (1) using new techniques to study flares obscured by high-background counts in the existing RHESSI dataset and (2) providing new observations through the development and flight of the *Gamma-Ray Imager/Polarimeter for Solar Flares* (GRIPS) instrument.

An X10 event that occurred on 2003 October 29 is the fifth flare in the RHESSI dataset that has been imaged at 2.2 MeV. This flare was obscured by large and irregular atmospheric background due to electron precipitation from Earth's radiation belts. This type of background is difficult to remove using RHESSI's usual background-subtraction methods, as a result the 2003 October 29 flare is not included in the relativistic electron/ion correlation study described above. However, since background counts are insignificant in RHESSI's style of indirect imaging, this flare can still be imaged despite the high background. In contrast to the other 4 gamma-ray flares that showed significant separations of the ion-associated 2.2 MeV and relativistic electron-associated HXR centroids, the 2003 October 29 event's centroid separation was found to be a statistically insignificant 13" ± 17" [67].

Though the 2003 October 29 flare is one of the five largest in RHESSI's observations, and the imaging results hint that it could be different from the other four in the set, this flare is not well-studied because of difficulty in deriving its spectrum. Chapter 5 explores using RHESSI imaging to separate background counts from flare counts in the gamma-ray energy range. This method is then used to generate a spectrum for the 2003 October 29 flare that puts it in context with the other large gamma-ray flares. Chapter 4 provides an introduction to indirect imaging and a review of competing imaging methods from 1 keV to 10 MeV.

The Gamma-Ray Imager/Polarimeter for Solar Flares (GRIPS) instrument is a balloon-borne solar telescope that had a successful flight from Antarctica in January 2016. GRIPS is designed to provide high-resolution spectroscopy ($\sim$50 keV – 10 MeV), imaging (12.5" – 162") and polarimetry ($\geq$150 keV) of flares. GRIPS' key technological improvements over the current solar state of the art at HXR/gamma-ray energies, RHESSI, include 3D position-sensitive germanium detectors (3D-GeDs) and a single-grid imaging system, the multi-pitch rotating modulator (MPRM). For photons that Compton scatter, which is the dominant process $\gtrsim$150 keV in germanium, the energy deposition sites can be tracked, providing polarization measurements as well as enhanced background reduction through Compton imaging. In GRIPS' energy range, indirect imaging methods provide higher resolution than focusing optics (>100 keV) or Compton imaging techniques. The MPRM grid-imaging system has a single-grid design which provides 2$\times$ the throughput of related bi-grid imaging systems like RHESSI. GRIPS' angular resolution is capable of imaging the separate relativistic electron/ion centroid emissions with precision over a variety of flare sizes. In chapter 6 the GRIPS instrument is introduced and its Antarctic flight discussed. Chapter 3 provides background on gamma-ray detection and discusses how position-sensitive germanium semiconductor detectors, such as those used on GRIPS, operate. Chapters 7 & 8 focus on GRIPS' spectrometer design and performance in detail. The MPRM imager and its aspect system, which provides the high-cadence pitch, yaw and relative roll solutions that are required for imaging, are discussed in chapter 9. In the past few years I have given numerous education and public outreach (EPO) lectures about the GRIPS project and our team's Antarctic adventure, a selection of these EPO materials can be found in the appendix.

Additional observations and enhanced imaging, spectroscopy and polarimetry of flare emissions in the HXR/gamma-ray energy range are needed to address the current era of particle acceleration and transport questions, including:

- What causes the observed spatial separation between the sites of relativistic electron and ion energy deposition in flare footpoints? Does this separation occur in all flares?
- How far does the correlation between ion and electron emission extend? Can we use electron emission as a proxy to understand and quantify ion emission?
- Do all flares accelerate ions?
- Are relativistic electrons isotropic or beamed in the corona?
- What is the composition/variance of accelerated material in space and time?

My research addresses these questions by expanding upon the set of RHESSI flares to include events obscured by high-background and providing new observations of gamma-ray flares with a next-generation imager/polarimeter.

1.1 Scope of Work

My doctoral work has three primary components: solar physics research focusing on gamma-ray emission from flares, space hardware development for the GRIPS balloon payload and missions operations for the RHESSI satellite. My advisors and I decided on these roles as a means to experience the whole space physics project lifecycle within the timeline of a graduate program. Through these roles I have worked alongside senior physicists and engineers to bring mission concept through design, integration and test, flight, operations and data analysis. I highly recommend a similar program of research, suborbital hardware development and spacecraft operations to other beginning graduate students.

The arc of my graduate research has been to expand on the limited set of gamma-ray flare observations with the motivation of understanding relativistic electron and ion dynamics. RHESSI's small set of gamma-ray flares have been relatively well-studied in this regard, with the exception of flares whose spectra have been obscured by irregular background. The method of imaging-spectroscopy leverages the fact that indirect imaging systems do not modulate, and hence do not image, background to separate out flare counts and generate spectra. The analysis contained in this manuscript examines extending this method into RHESSI's gamma-ray energy range as a means to constrain the relativistic electron and ion fluences in background-obscured flares. The primary goal was to put the mysterious 2003 October 29 X10 event in context with other large gamma-ray flares because its spatially resolved relativistic electron/ion centroids do not obviously share their results. For this project I collaborated with Dr. Bob Lin and Dr. Albert Shih, and used RHESSI's solarsoft IDL routines to support analysis.

GRIPS is a NASA Low Cost Access to Space (LCAS) suborbital payload primarily built at UC Berkeley (original PI Dr. Bob Lin and current PI Dr. Pascal Saint-Hilaire) that is designed to study solar flares by providing high-resolution imaging and spectroscopy. GRIPS had a successful Antarctic flight in January 2016. As the physicist responsible for implementing the detector, imaging and aspect subsystems my role was broadly defined – I filled a blend of systems engineering, integration, test, calibration, operations and data analysis roles. I developed systems requirements with senior scientists and worked with the engineering team for rapid systems delivery through aggressive prototype and test iterations. Working with small teams I was involved in clean room integration of flight detectors into the cryostat mechanical assembly, component and instrument wide noise reduction and thermal vacuum testing of both individual components and a full-scale thermal vacuum test at NASA Glenn Research Center's Plum Brook facility. I was the lead scientist responsible for tuning ASIC performance against an 18-D parameter space for the 1,800 detector strips across six flight detectors, providing detector calibrations, integrating the cryostat onto the gondola bus, delivering the imaging and aspect systems' mechanical components and maintaining systems-level knowledge of the spectrometer for operations. I also developed python tools to support GRIPS data analysis and analyzed the detector data for the largest flare that occurred during flight. I delivered the cryostat to Antarctica on-board a series of Air-National Guard flights and after our successful Antarctic flight campaign (2015/16) I led the following

season's Antarctic deep-field recovery of the flight hardware (2016/17).

RHESSI is a NASA Small Explorer mission that launched in 2002 as a PI-led mission (originally Dr. Bob Lin and currently Dr. Säm Krucker) that is operated from UC Berkeley's Space Sciences Lab. As a member of the Science Operations Team I help plan mission strategy and monitor payload systems. I take shifts supporting 24-hr flight operations as the on-call team member who monitors systems heath and coordinates with mission control specialists for payload operations. Frequently this involves adjusting payload parameters to maintain the health of systems now in their 15th year of operation, ensuring continuous high-quality data products from an aging satellite.

I want to take this opportunity to highlight two efforts that were outside of my technical research but valuable to my graduate experience. I believe that a vital duty of the scientific community is to disseminate knowledge so it can benefit society as a whole. With these projects I hope to have made contributions to the scientific community by reaching out to policy makers and to the greater public through education. I joined the Public Policy Committee of the Solar Physics Division (SPD) of the American Astronomical Society (AAS) as a student member during 2013 – 2015. As a part of the committee I helped organize congressional visits and generated materials to educate policy makers about space physics research and space weather hazards with the goal of increasing NASA Heliophysics division funding. My second effort was education and public outreach about solar flares and the GRIPS Antarctic campaign in the form of public lectures and twitter live updates. A selection of these EPO materials can be found in the appendix to this document.

Chapter 2

The Sun and Solar Flares

2.1 The Sun

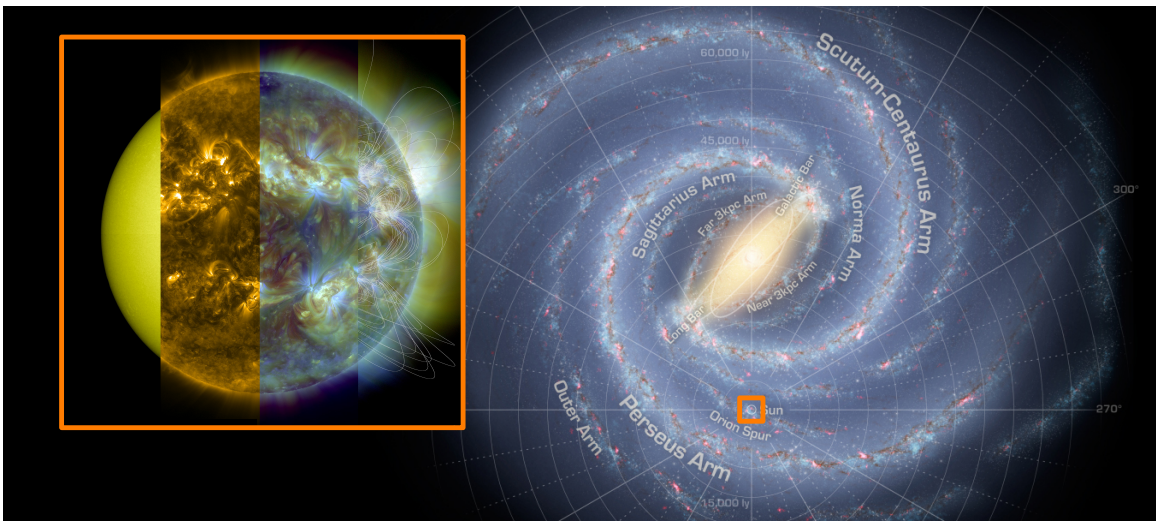

Figure 2.1: *Left:* A composite image of the Sun on 2011 December 7 as observed by SDO [115]. The slices (from left) are images of the ∼6,000 K photosphere, the ∼1 million K transition region in EUV light, and two composite images showing coronal temperatures up to 2 million K. The final image is overlaid with a fit to the coronal magnetic field structures. *Right:* An artist's concept of the Milky Way galaxy overlaid with galactic longitude relative to the Sun, located in the Orion Spur. Image adapted from NASA/Adler/U.Chicago/Wesleyan/JPL [31].

The Sun formed ∼4.5 billion years ago in the Orion Spur of the Milky Way galaxy (figure 2.1). Stars form through gravitational collapse of diffuse molecular clouds made from gas and dust. The collapse accretes material into a rotating sphere of superhot plasma whose mass determines its evolution, structure and lifetime. If the star gains enough mass the center

will reach temperatures >10 million K and hydrogen fusion begins. The energy produced by fusion establishes a hydrostatic equilibrium that keeps the star from collapsing further. The Sun is about halfway through its main hydrogen burning phase, in ∼5 billion years it will have exhausted its hydrogen fuel supply and will expand to become a red giant. Helium burning will set in, but since the Sun is not massive enough to continue fusing heavier elements its core will eventually collapse to become a white dwarf. Stars with ∼10× the mass of the Sun continue fusing heavier elements and explode as supernovae, whose iron cores collapse into neutron stars or black holes.

Solar Structure

The Sun does not have a physical surface so much as a boundary where the transparent outer plasma becomes opaque. Inside this ∼100 km thick layer, termed the *photosphere*, the density of H^- ions increases and the optical light generated by the plasma is quickly reabsorbed such that it does not escape. For this reason it is not possible to see deeper into the Sun.

Since we cannot directly sense light from the deeper regions of the Sun, models are used to infer its inner structure. The *standard solar model* uses the observable boundary conditions of luminosity ($L_\odot = 3.8\times10^9$ W), radius ($R_\odot = 7\times10^5$ km) and mass ($M_\odot = 1.989\times10^{30}$ kg) of the Sun to predict its structure. The final model input is the solar age, which has been roughly determined as 4.5 billion years via the age of meteorites in the Solar System. The solar model predicts that the Sun's interior is divided into layers that have differences in characteristics like temperature, pressure, density and energy-transfer mechanism. Figure 2.2 shows an illustration of the primary internal and external solar layers drawn to scale, and points to some features observed in the solar atmosphere.

Models of the solar internal structure have been verified by observing the propagation of acoustic (p waves) and surface gravity waves (f waves). The study of solar seismic waves, called *helioseismology*, has been extremely successful in developing a detailed view of the stratification and internal angular velocity of the Sun's inner structure, verifying the standard solar model predictions to within .1% (figure 2.3, left). The results of helioseismology were critically important in verifying the expected rate of solar neutrino production from fusion during the neutrino debate [35], wherein the measured flux of electron neutrinos at Earth was ∼33% of that predicted by the standard solar model. The strong agreement between the solar model and helioseismology indicated that the prediction of solar neutrino production was correct, suggesting that the neutrino problem solution would be found in modifications to the Standard Model of particle physics. In 2002 the solar neutrino problem was put to rest when neutrino observatories were able to confirm that neutrinos had mass, in disagreement with the Standard Model of particle physics which assumed they were massless. By assigning neutrinos mass, the electron neutrinos could change to previously undetected muon and tau neutrinos in the exact percentage that accounted for the "missing" solar electron neutrinos.

The core contains 1/2 the mass of the Sun and spans out to 15% of the solar radius with an average density of 20 g/cm^3. Inside the core proton-proton fusion generates ^{4}He from

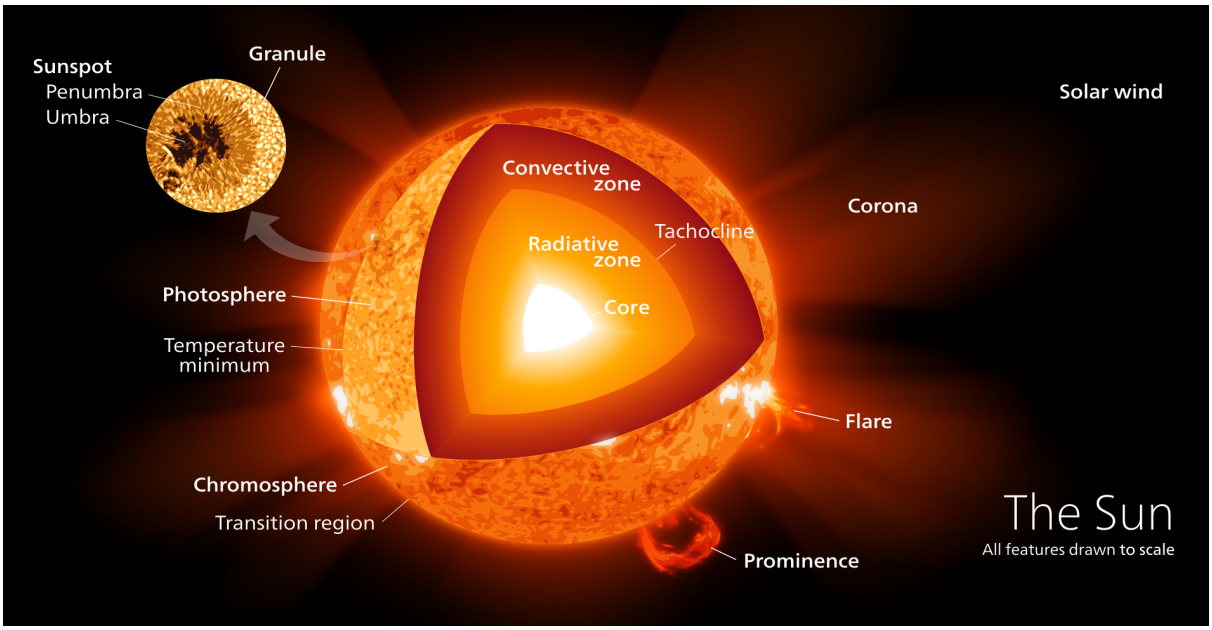

Figure 2.2: Diagram of the Sun's layers drawn to scale. Image created by Kelvinsong [71].

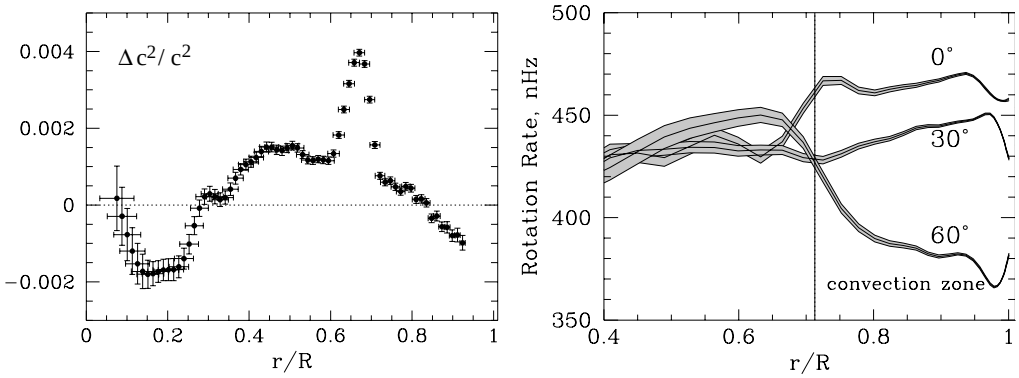

Figure 2.3: *Left:* Difference between the squared sound velocity predicted by the standard solar model and measured by helioseismology. The model accurately predicts velocity to .1%. *Right:* Heliosesimology measurements of the solar rotation rate as a function of radius at three latitudes. The central black line is the model value with errors represented by the shaded regions. Both images are from Kosovichev, et al. 1997 [77].

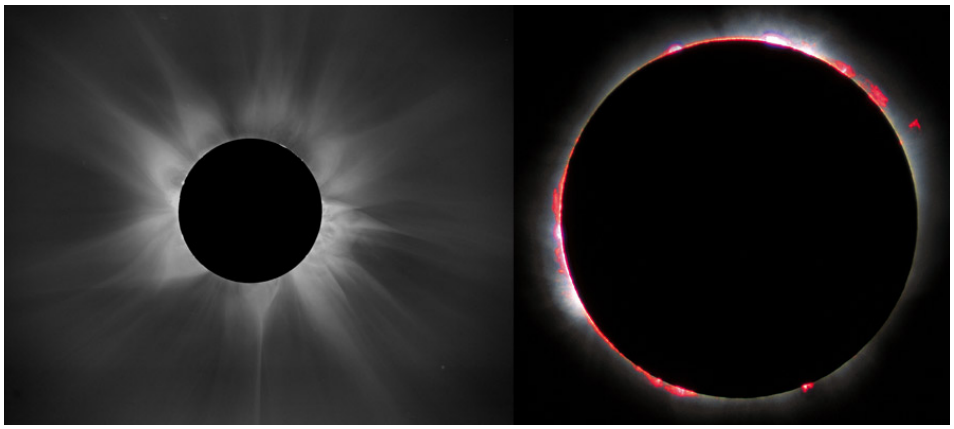

Figure 2.4: *Left:* Image of the corona during a total solar eclipse on 1980 February 16 from NCAR's High Altitude Observatory (HAO). *Right:* Image of the chromosphere during a solar eclipse in France on 1999 August 11 by Luc Viatour. Both images were provided by UCAR [39].

hydrogen nuclei, heating it to temperatures above 10^7 K.

Photons transport the energy generated in the core through the *radiative zone*, with a characteristic timescale of $\sim 10^7$ years. The radiative zone is the largest solar volume, spanning between .15 – .7 $R_\odot$. Density drops by three orders of magnitude by the radiative zone's outer edge, becoming .02 g/cm^3, and the temperature cools to $\sim 10^6$ K. The radiative zone and core uniformly rotate together.

The tachocline separates the uniformly rotating radiative zone from the differentially rotating convective zone, which imparts a strong shear to this layer. Figure 2.3 (right) plots the results of helioseismology measurements of the solar rotation rate as a function of depth at three different latitudes, and marks the top of the tachocline.

Bulk motions of hot plasma transfer energy from the tachocline up towards the visible solar surface in convective cells spanning from .7 – 1.0 $R_\odot$, inside the *convective zone*. Along this layer the denisty falls off by 5 orders of magnitude, reaching 10^{-7} g/cm^3. The rising plasma forms columns that are visible as bright spots, called *granules*, that form a mosaic pattern on the visible surface, shown inset to figure 2.2. The dark inner granule lanes are created by cool plasma downflows. Inside this zone the charged particle flows organize and drive local magnetic field behavior.

At the top of the convective zone the plasma becomes transparent to optical light, which can then escape into space and reach Earth. The photosphere, shown in the far left solar slice of figure 2.1 left, looks like a smooth sphere in optical light. The photospheric plasma absorbs and emits light like a blackbody. The optical light peaks in the yellow-green part of the visible band and measurement of its continuum suggests that the photosphere has a blackbody temperature of 5800 K.

The *chromosphere* is a $\sim$2500 km thick layer above the photosphere. Figure 2.4 (right) shows an image of the chromosphere during a solar eclipse. The rosy hue comes from H-α

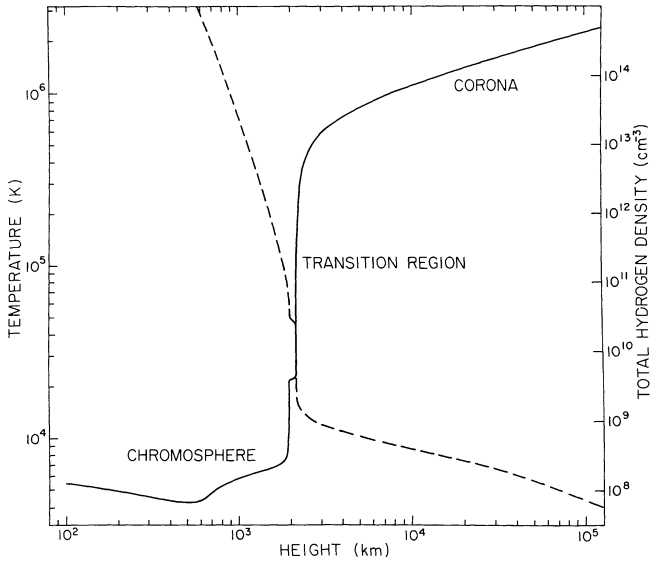

Figure 2.5: Temperature (solid line) and density (dashed line) of the solar atmosphere plotted along the height above the photosphere. The temperature in the solar corona climbs as the density falls off. Plot from Withbroe and Noyes 1977 [147].

emission; it is this coloring which gave the region its name. Imaging of the corona in H-α shows a forest of spicules that extend from the chromosphere into the upper solar atmosphere. Chromospheric temperature and density continue to decrease until the temperature reaches a minimum. At this point, inside the *transition region*, the temperature abruptly increases by two orders of magnitude while the density continues to fall off. Figure 2.5 plots the temperature and density across the chromosphere, through the transition region, and into the *corona* where the temperature climbs to $\sim$2 MK. Attempts to explain the heat input into the corona via small releases of stored magnetic energy in *nanoflares* or wave heating have not adequately explained this temperature inversion. The source(s) and mechanism(s) of energy input into the corona and how it can maintain this temperature, called the *coronal heating problem*, are as yet unidentified. Figure 2.1 (left) shows the 1 million degree transition region in EUV light as the second solar slice (from the left).

The hot and tenuous corona is the Sun's outer atmosphere. This region is highly structured, as shown in figure 2.4 (left) during a solar eclipse in 1980. Close to the disk are loop like structures of closed field lines along which particles travel and return to the Sun. Figure 2.1 shows the coronal loop structures in the final two solar slices. The third slice from the left is a combination of multiple SDO filters that show loops of hot plasma up to 2 million K. The final slice is another composite image overlaid with fits to the magnetic field lines in this region. The interactions of coronal loops impart this region with highly dynamic and explosive behavior in the form of flares and CMEs.

From the corona's radial structures blow the *solar wind*, a constant stream of electrons,

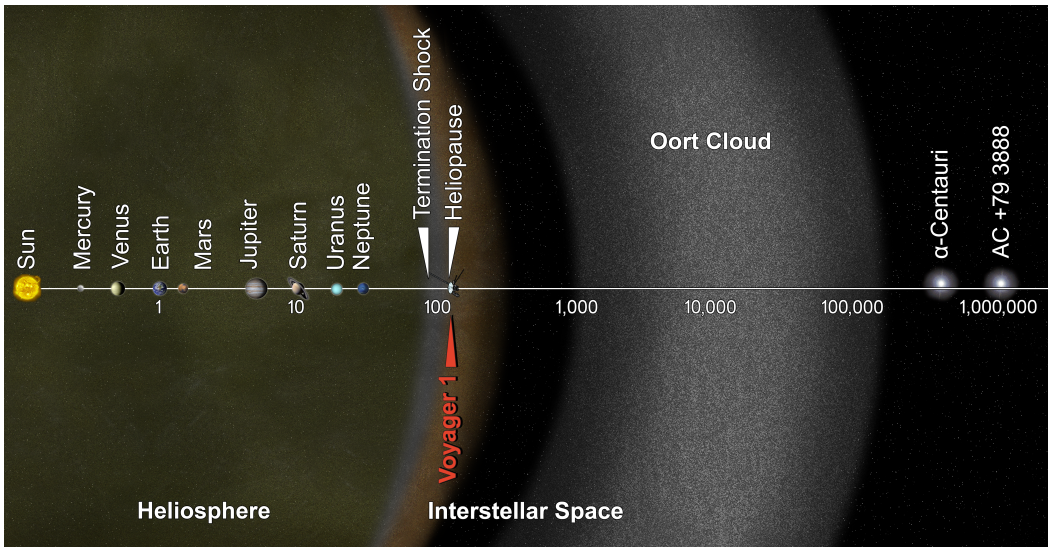

Figure 2.6: Artist's conception of the heliosphere and interplanetary space on a logarithmic distance scale with units of AU. One AU is the distance spanning the Sun to the Earth and is equivalent to 150 million km. At the time of this drawing, in 2013, the Voyager 1 probe was 125 AU and had crossed into interplanetary space. Image courtesy of NASA JPL/Caltech [70].

protons, alpha and heavier particles that flow from the Sun into interplanetary space. The equatorial part of the wind is comparatively slower than the polar wind, giving these two components the names *slow* and *fast solar wind*. The fast wind blows from *coronal holes*, which are regions characterized by magnetic field lines that extend so far into into interplanetary space that in the solar reference frame they're considered *open field lines*. The slower wind emanates from streamers above magnetically complex active regions.

The solar wind blows through interplanetary space, creating the constant stream of particles which populate the Earth's radiation belts and precipitate as auroras. Wind particles that are not caught by planets or other solar system bodies keep traveling until they reach a point where their pressure equals the pressure exerted by the interstellar medium at a point called the *heliopause*. This boundary marks the edge of the Sun's influence on local interstellar space. In 1977 NASA launched the Voyager 1 & 2 spacecrafts to study the outer solar system. In August 2012 the Voyager 1 probe reached the heliopause and crossed into interstellar space at 116 AU (17.3 billion km). These probes are the farthest man-made objects in space that are still in contact with Earth.

2.2 Solar Magnetic Activity

Solar activity is governed by magnetic fields over multiple temporal and spatial scales. The *solar dynamo* generates a roughly dipolar global magnetic field that switches polarity every

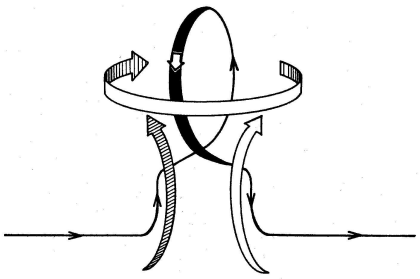

Figure 2.7: Diagram illustrating the geometry of how a latitudinal flux tube rises and becomes twisted by the Coriolis force to produce a loop that is perpendicular to the original direction. [114].

∼11 years. Smaller regions of concentrated magnetic flux, called *sunspots*, that appear as large dark spots in the mosaic of convection cells in the photosphere are typically ∼10,000 km across [53] (inset figure 2.2). Sunspots emerge on the order of days and can persist for weeks, with their overall population modulated on the same ∼11 year time scale as the dynamo. Tethered to the sunspots are flux tubes that rise into the chromosphere and corona, where they form loops that can twist and shear, building up energy as their topologies become more complex. Associated with these loops are CMEs and flares, the fastest manifestations of magnetic activity. CMEs are ejected from the Sun with velocities of 20 – 3000 km/s and flares can release energy and restructure field configuration within 100s – 1000s of seconds.

A solar dynamo governs the periodic reversal of the overall solar magnetic field as a result of differential rotation, bulk convective motions and meridinal flow of the plasma. With these mechanisms the poloidal and toroidal solar magnetic fields can regeneratively induce one another.

Differential rotation drags the plasma at the solar center faster than the plasma at the poles. High conductivity binds the particle and field motions together such that they are *frozen-in flux*. Since the ratio of plasma pressure to magnetic pressure (the plasma parameter β) is high in the convective zone, the plasma will drag along the magnetic field as it moves. In this way a poloidal magnetic field that extends longitudinally between the solar North and South poles will become stretched and winds around the Sun because of differential rotation. After many rotations the originally poloidal field is now toroidally wrapped around the sun many times.

The mechanism which regenerates of the poloidal component is an area of active research and the reader is referred to Charbonneau 2010 [30] for an excellent review of current solar dynamo models. The mechanism described by the Babcock-Leighton model [9, 85, 84] provides a good schematic description the process which leads to the dominant magnetic features and activity observed during the years of solar maximum, and how toroidal fields can generate poloidal fields. In this model the latitudinal flux tubes of the toroidal field rise to become loops inside the plasma convective flow, which then become twisted by the Coriolis force. The resulting loop is oriented perpendicular to the original flow, as illustrated

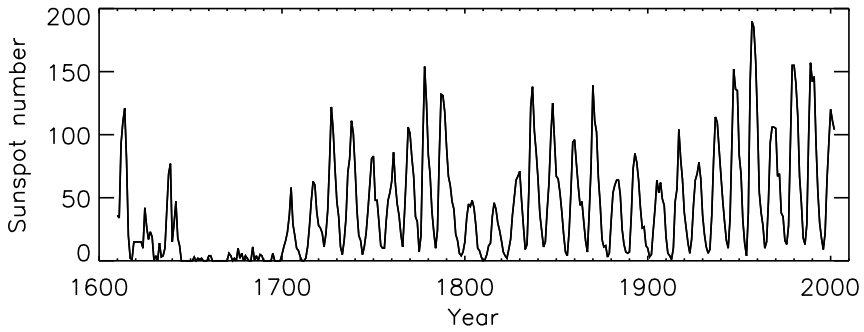

Figure 2.8: Yearly averaged sunspot number since 1610. The ∼11-year periodicity is variable between cycles, with a long period of suppressed activity in the 17th century called the Maunder Minimum when very few sunspots appeared. Plot from Schüssler, 2004 [114].

by figure 2.7. These rising flux tubes generate sunspots in the photosphere and rise into the solar atmosphere to fuel the eruptive activity of flares and coronal mass ejections. In the Babcock-Leighton model, the overall fields of the decaying sunspots then concentrate at the poles to form the poloidal field. As Charbonneau points out, extensive modeling of the fields confirms that this process does occur on the Sun and triggers the magnetic reversal of the poloidal component, but the key question is if this collection by decaying sunspots is the driving mechanism for the dynamo cycle or a side-effect of the active-region decay.

The quantity of sunspots on the solar surface and degree of solar activity in the form of flares and CMEs follow the same ∼11-year periodicity as the global magnetic field, and is termed a *solar cycle*. Figure 2.8 shows the daily average of sunspot numbers from 1610 to 2002. Though the ∼11 year solar cycle dominates, there is variability between each cycle duration (9 – 12 years) and magnitude, the most dramatic of which was an extended ∼60 year time period of extremely suppressed activity called the *Maunder Minimum* in the 17th century. Current dynamo models can roughly reproduce the periodic sunspot number along with other features salient to how the sunspots migrate longitudinally and the directionality of their polarity. The variability in cycles, however, is a major challenge in this field of research [30].

Sunspots form when magnetic flux tubes rise through the convective zone due to magnetic buoyancy and emerge into the photosphere to create regions of strong magnetic polarity (∼1000 G). At the photosphere, magnetic pressure from the strong fields reduce the gas pressure and suppress the upward convective flow of hot material, leaving the plasma comparatively cooler and darker than in the surrounding regions. The *Wilson effect* describes how the density difference in these regions lead to reduced opacity, resulting in a deeper apparent solar surface in a sunspot.

Magnetic flux ropes rise from the convective zone through sunspots, into the chromosphere and corona to form cascades of arches, giant *coronal loops*, *prominences* and *filaments*.

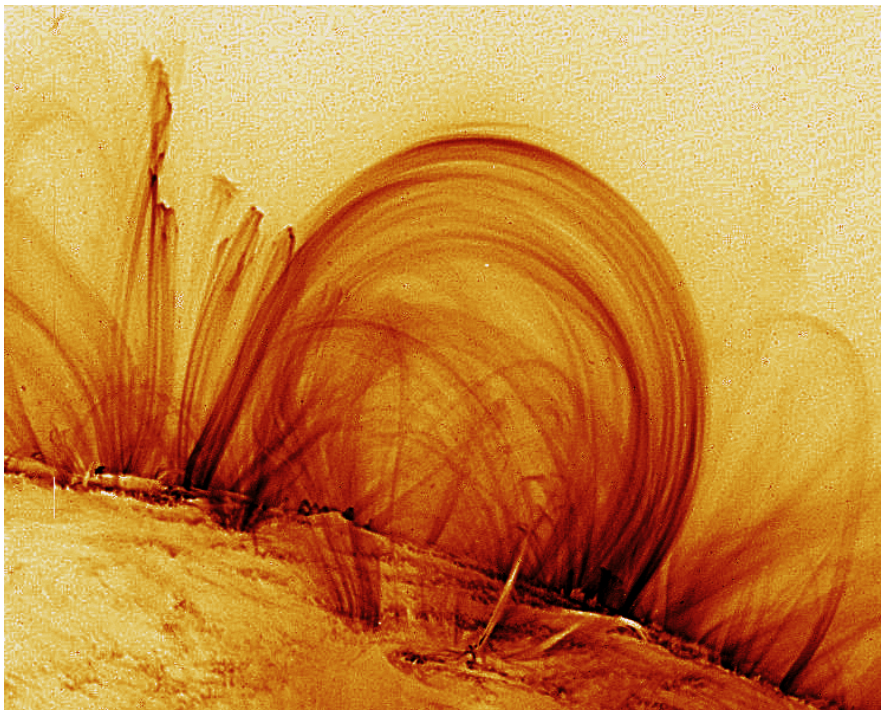

Figure 2.9: False color image of coronal loops taken on 1999 Nov 6 in TRACE's 171 $\mathring{A}$ band. Image courtesy of TRACE/NASA [136].

The most ubiquitous structure in the solar atmosphere is certainly the loop (figure 2.9), and the loops' footpoints are usually found in the magnetically complex areas of sunspot groups called *active regions*. In the solar atmosphere the plasma pressure has dropped significantly but the magnetic pressure and conductivity are high, leading to a low-β plasma where the motions of the magnetic field now determine the motions of particles. The particles are constrained to move along the field lines and as they move they emit light, tracing out the field structures.

Active regions of high magnetic complexity can experience sudden explosive energy releases that can accelerate particles to relativistic speeds, heat the local plasma to 15 – 20 MK and reconfigure the surrounding magnetic field in 100s – 1000s of seconds. These impressive displays are solar flares, the largest explosions in the solar system and the most tempestuous display of magnetic activity on the Sun. The original work in this manuscript is devoted to increasing observational capabilities for flares with the aim of understanding the processes which accelerate particles.

2.3 Solar Flares

Solar flares are characterized by a sudden energy release that is accompanied by broadband radiation and a reorganization of the local magnetic fields. The largest flares have been

observed to release an estimated 10^{32} – 10^{33} ergs over 100s – 1000s of seconds. Flare emission can span radio to gamma-ray energies whose timing, spatial location and duration can be used to understand the underlying physics of these events.

The magnitude of a flare is classified by the peak of its soft X-ray flux as measured by the *Geostationary Operational Environmental Satellite* (GOES) satellites using a logarithmic scale with letters that define the magnitude. The scale proceeds through A, B, C, M and X for fluxes of 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5} and 10^{-4} W/m^2, respectively. A flare of GOES class X2.5 would have a peak SXR flux of 2.5×10^{-4} W/m^2.

Flare Phases

The time history of a flare can be roughly divided into three phases: the *preflare*, *impulsive* and *decay phases*. An illustration of flare evolution plotted by energy is shown in figure 2.10. In the preflare phase the surrounding plasma gradually heats up and emits in the EUV and SXR. Active region prominences can brighten and show morphological changes minutes before the flare begins. These events signal that the region is starting to gradually leak energy and release stress before the rapid onslaught of the impulsive phase.

The flare onset is abrupt, the impulsive phase is marked by a sudden increase in broadband emission that lasts on the order of minutes. Most of the energy is released during the few minutes of this phase. The HXR/gamma-ray emission appears as two localized nodules in the chromosphere located at the site of the magnetic loop footpoints, and is occasionally accompanied by a third HXR source above the loop top in the corona. The sudden increase in Hα intensity and line width, called the *flash*, usually occurs during the impulsive phase though its peak in emission can occur later.

The thermal SXR emission peaks after the impulsive release of energy. The decay phase has a relatively long duration of one to multiple hours while the emission gradually drops off and approximately returns to the pre-flare baseline. During this time the arcades of loops that result from reconfiguration of the magnetic field during the impulsive phase are brightly illuminated by flare-heated plasma. High in the corona particles continue to be accelerated, generating radio emission, and possibly ejected into interplanetary space.

Standard Flare Scenario

The association between complex magnetic active regions and flares is well-established. Not only are flares accompanied by large-scale reconfigurations of the local magnetic fields, but they are also the only coronal power source which can provide the $\sim 10^{33}$ ergs released during the largest flare events. Tandberg-Hanssen & Emslie (1988) [135] showed that the other possible energy sources, gravitational and thermal potential energy, could only provide $\sim 10^{28}$ ergs for comparable flare sizes.

The stored magnetic energy is thought to release when the stressed field relaxes into a new configuration by the process of *magnetic reconnection*. A schematic depicting two reconnection models are shown in figure 2.15. The reconnection geometry consists of oppositely

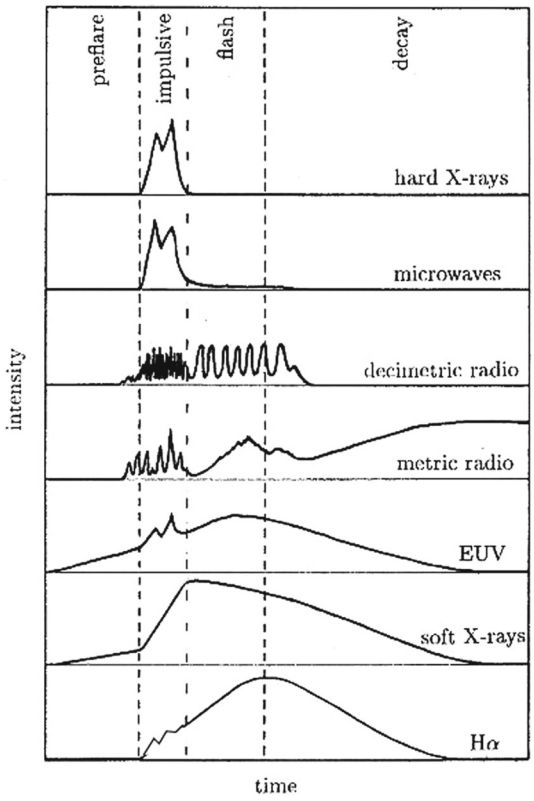

Figure 2.10: Illustration of the relative timing of flare emission energy bands during the preflare, impulsive and decay phases. The Hα flash occurs during the impulsive phase but peaks later. Image from Benz, 2002 [15].

oriented magnetic field lines that come together to form a neutral current sheet. Inside the current sheet the conditions are favorable for the originally anti-parallel field lines to change their topology and reconnect into outgoing loops. A description of the magnetic reconnection process is provided in section 2.4.

The *standard flare scenario* is a conceptual framework that describes the chain of events a flare progresses though. It does not address the underlying physics of how different phenomena occur so much as link the most common flare observations into a causal framework. In a recent flare review Benz (2017) [16] highlights the key observations which must be addressed and provides a description of this scenario, the *standard model*, summarized below.

- The SXR flux is correlated with the cumulative HXR flux, in the so-called *Neupert effect* [95]
- Compact HXR sources are frequently at the base of SXR loops [64, 112]

- A coronal HXR source can occasionally be observed coincident with, or above, the SXR loop top [90, 80]
- The energy contained in non-thermal accelerated electrons tends to exceed the energy in the thermal plasma
- There are spectral differences in the HXR loop top and footpoint sources
- The SXR emission measure increases during the impulsive phase, suggesting that more material is entering the flare loops

The standard model geometry is shown in figure 2.11, it evolved through the work of multiple scientists in its early incarnation (Carmichael 1964 [27], Sturrock 1966 [131], Hirayama 1974 [62], Kopp & Pneuman 1976 [76]) and for this reason it is also referred to by the abbreviation of their names as the *CSHKP model*. At that time the whole list of observations cited above were not available, leading to challenges and addenda to the model through the recent decades, but this general 2-D picture is still popular and relevant today.

The model begins when reconnection is initiated in the corona and particles are accelerated as a result. Particle acceleration occurs at the reconnection outflows, and the candidate acceleration mechanisms are discussed in section 2.5. The particles are either transported away from the Sun where they can escape into space, or, they stream in the opposite direction along the loop field lines back towards the chromosphere. The particles that head away from the Sun can continue to be accelerated in the high corona by shocks or other processes, and these particles are associated with the late-phase radio emission. If the accelerated particles escape into space, they can be detected by spacecraft in-situ and are termed *solar energetic particles (SEPs)*.

The particles that precipitate downwards travel relatively unobstructed through the corona. Magnetic mirroring, caused by the narrowing of flux tubes in the lower atmosphere, may trap the downward-flowing particles. Trapped particles will "bounce" between the mirroring points until they are eventually scattered into the mirror loss cone. Particles then continue their journey into the dense chromosphere where they lose non-thermal energy through bremsstrahlung, generating two compact HXR sources at the loop footpoints.

Sudden energy deposition in the chromosphere by flare-accelerated particles causes the local plasma to heat-up and expand to fill the coronal loops, in a process termed *chromospheric evaporation*. The hot plasma (10 – 20 MK) emits in the EUV/SXR, resulting in the gentle rise of this emission, and linking the cumulative HXR emission to the SXR.

Flare Geometry

The standard flare model outlined above reasonably explains the progression of flare observations, and it is well-accepted that stressed anti-parallel current-carrying (non-potential) magnetic flux tubes provide the energy input to the impulsive phase. The overall geometry of the magnetic fields in the corona that lead to conditions favorable for reconnection is not

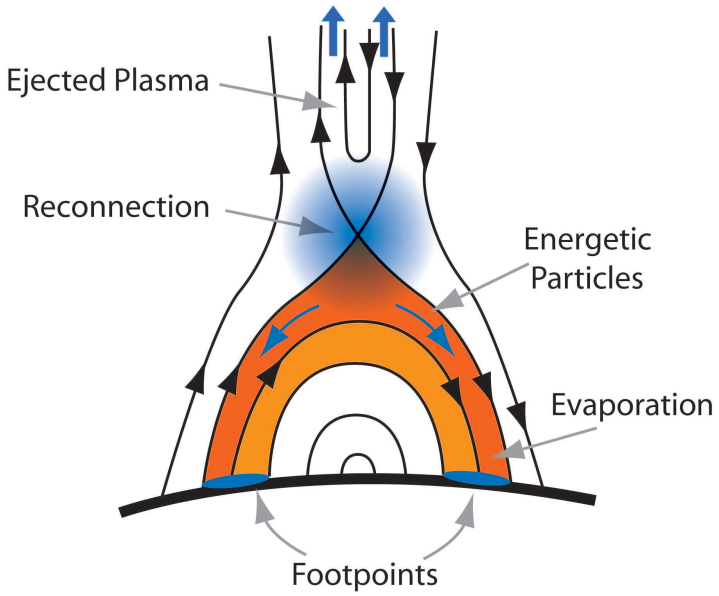

Figure 2.11: CSHKP model geometry for the standard model of solar flares. Image from Christe 2007 [32].

clear. Furthermore, any model that describes the geometry of the large-scale magnetic fields in the corona must also address the possible formation and launch of an accompanying CME and provide a mechanism that can impart it with a similar energy content as the flare.

Several models have been suggested for the greater coronal-field geometry that leads to reconnection and the standard scenario discussed above. The leading models suggest a single loop reconnecting with itself (CSHKP, one loop or dipolar), a single loop reconnecting with a open field line (interchange or tripolar) and two loops reconnecting with each other (two loop or quadrupolar). The diversity and serendipity of flare observations (e.g., figure 2.12) have garnered both supporting and conflicting observations for each of these models. Benz [16] points out that there is no reason to exclude these models from one another as there could be a combination of them occurring in the flare at the same time or at different stages leading to hybrid geometries and a vast array of observations. He therefore argues against a standard geometry for the preflare coronal magnetic structure.

The CSHKP framework is termed and considered the standard model for eruptive events today. It assumes a single loop whose sides are driven together to form the stressed anti-parallel flux tubes required for reconnection. The sides can be pushed together by other magnetic structures in the crowded corona, by motions of the chromospheric footpoints relative to the coronal loop top that can act to "wind-up" the loop, or other dynamic processes. The coronal geometry is shown in figure 2.13 (left). In the 3-D extension of this model an arcade of sheared loops reconnect with each other to generate a large twisted coronal flux rope through the center of the plasmoid, which can be launched as a CME.

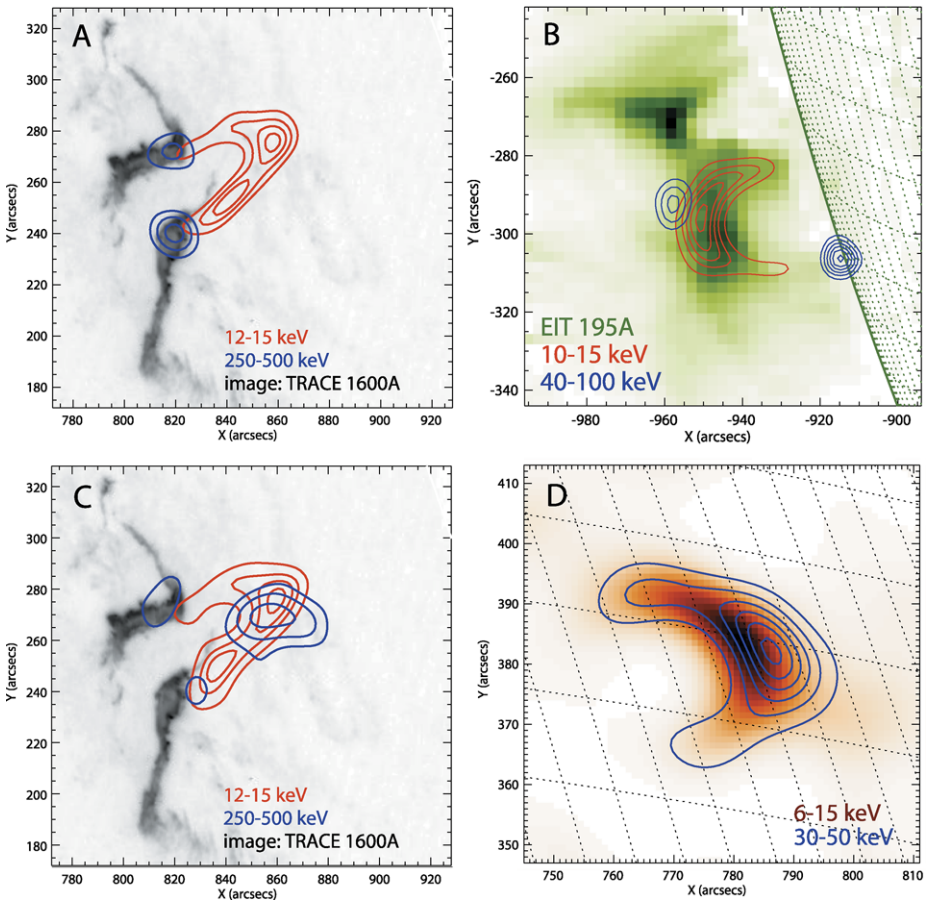

Figure 2.12: Flare observations showing loop, footpoint, loop top and above the loop sources. The figure is a composite from different authors collected by Raymond et al. 2012 [110]. *(A)*: Common flare observation showing HXR footpoints from non-thermal emission (blue) and thermal SXR emission in the loop (red). From Krucker et al. 2008 [80]. *(B)*: Flare showing a non-thermal HXR source above the loop top. From Ishikawa et al. 2011 [69]. *(C)*: The same flare shown in panel (A) at a later time when a coronal loop-top source has developed but the footpoint emission has decayed. From Krucker et al. 2008 [80]. *(D)*: A rare observation showing non-thermal emission from the flare loop with no accompanying footpoint emission. From Veronig and Brown 2004 [138].

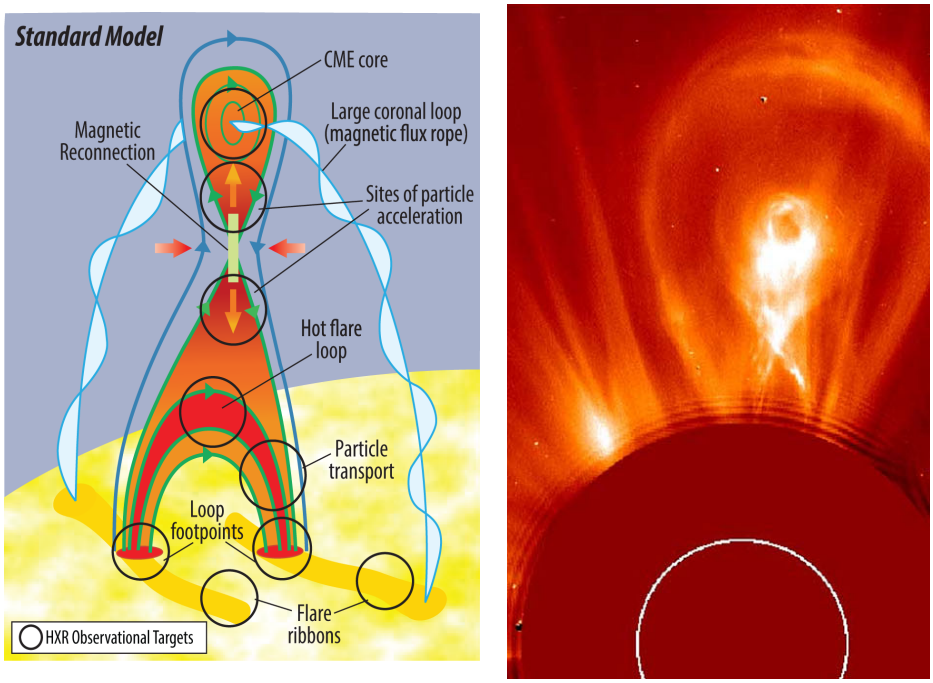

Figure 2.13: *Left:* Standard model of solar eruptive events. Image from Christe 2017 [34] *Right:* Observation of a flare and CME that shows an apparent X-point and is suggestive of the standard model geometry. Image from LASCO/SOHO [82]

There are decades of observations supporting the one-loop scenario. Figure 2.13 (right) shows a LASCO coronagraph image that displays a few key elements of the standard model for eruptive events, including the reconnection X-point and a hot plasmoid forming above. Evidence for this geometry includes vertical cusp-shaped structures seen in SXR after the flare, horizontal inflows of cold plasma and vertical outflows of hot plasma, synchrotron and HXR emission from above the plasmoid region indicating high-energy particle acceleratio and two-ribbon footpoints in Hα, EUV & X-ray (Benz 2017 [16] and references therein).

In the quadrapole geometry two loops with anti-parallel fields are pushed together and reconnect. Multiple scenarios could lead to this geometry, including two pre-existing loops meeting and collisions as a result of newly emerging flux with pre-existing loops. The defining characteristic of this model is that none of the contributing field lines are open. Though ejected flux is possible, it is not a requirement of the model, and this geometry is more popularly proposed for non-eruptive flares which do not have an associated CME.

Reconnection between an open field line and a closed loop has been suggested as a geometry to describe the coincident eruption of a jet with a flare as well as the appearance of Type III radio bursts that are consistent with electron beams escaping from the Sun along open field lines. Type III radio burst emission has been associated with 33% of RHESSI's HXR flares above C5 [17]. The model geometry is shown in figure 2.14. This process predicts three chromospheric HXR footpoints and two coronal HXR sources at the shocks. A recent

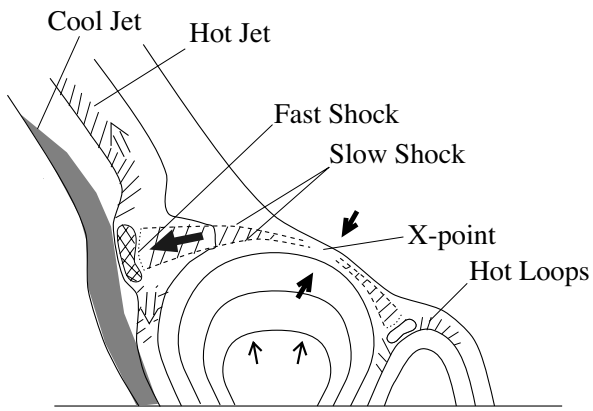

Figure 2.14: Diagram of interchange reconnection. This model proposes magnetic reconnection between a closed coronal loop and and open field line to explain the association of coronal jets and Type III radio bursts with flares. Image from Shibata & Magara 2011. [120]

observation provided the first observation of a double coronal HXR source, post-flare loop and Type III burst that is consistent with interchange reconnection [56].

2.4 Magnetic Reconnection

When a rising magnetic flux tube emerges into the corona it encounters other flux tubes and shears against them, increasing the magnetic complexity in the region. The field stress continues to build-up for days or weeks with continued flux emergence, and the relative motions between a flux tube's footpoints and loop top. Eventually the magnetic configuration becomes unstable and magnetic reconnection is initiated. Magnetic reconnection is the mechanism by which stressed field lines relax and release energy as a result. The released energy is in the form of accelerated particles and thermally heated plasma. Since reconnection is responsible for accelerating flare particles it must progress on a time scale which is faster than the impulsive phase of flares, 100s – 1000s of seconds.

This section introduces two different reconnection scenarios: the Sweet-Parker and Petschek models. Both of these models are equilibrium scenarios with constant inflows and outflows of plasma. Since impulsive energy release has been observed at the millisecond scale [73], the reconnection process must also account for bursty particle acceleration. Reviews of MHD and reconnection processes in the solar atmosphere, space plasmas & laboratory experiments can be found in Shibata & Magara 2011 [120], Yamada et al. 2010 [148], Zweibel 2009 [152], Aschwanden 2002 [6] and Tandberg-Hanssen & Emslie 1998 [135].

When two oppositely directed current-carrying magnetic fields come together a current sheet forms at their intersection. Within the current sheet there is a non-zero resistivity which breaks the frozen-in-flux condition that constrains the magnetic field and particle

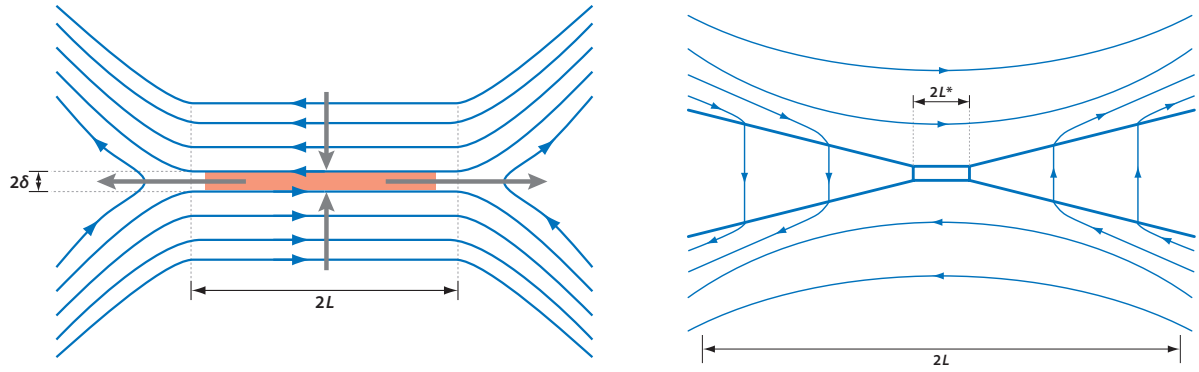

Figure 2.15: Reconnection geometry for the Sweet-Parker (*left*) and Petschek (*right*) models. The field/current inflows come from the top & bottom, and the reconnected field lines outflow in the horizontal direction. The Sweet-Parker reconnection region (shown as a pink hued box of width 2L) is much smaller than the Petschek reconnection region (of width 2L*). Images from Zweibel & Yamada (2009) [152].

motions throughout the greater corona. Energy is dissipated through resistive processes inside the diffusion region allowing the fields to relax and reconnect in the outflow.

Sweet Parker and Petschek reconnection

The simplest incarnation of magnetic reconnection occurs in the Sweet-Parker model (Sweet 1958 [133], Parker 1957 [100] & 1963 [101]) whose geometry is shown in figure 2.15 (left). The plasma inflows (vertical arrows) into a diffusion region that spans the length of the reconnecting field lines. The plasma outflows perpendicularly to the inflow (horizontal arrows).

Conservation of mass confines the outflow velocity (v_{out}) to be proportional to the inflow velocity (v_{in}) by the geometry of the diffusion region ($2\delta \times 2L$). This relationship consequently restricts the overall reconnection rate ($\mathrm{d}\Phi/\mathrm{dt}$).

$$v_{out}\delta \sim v_{in}L \tag{2.1}$$

$$\frac{d\Phi}{dt} = \frac{v_{in}}{v_{out}} \sim \frac{\delta}{L} \tag{2.2}$$

Since the width of the diffusion region is very narrow and the length can be a fraction of the coronal loop size (10^4 – 10^5 km), conservation of mass shows that the rate is severely limited when the plasma must pass through the reconnection region. An analysis of the reconnection time by Shibata & Magara 2011 [120] shows that the time scale for Sweet-Parker reconnection is $\sim 10^8$ – 10^9s, six orders of magnitude longer than the impulsive phase.

Reconnection alone cannot accelerate particles to the non-thermal velocities observed in flares. Equating the magnetic and gas pressures, then invoking conservation of mass yields

a plasma outflow velocity that is proportional to the inflowing magnetic field strength (B_{in}) and mass density (ρ).

$$v_{out} = \frac{B_{in}}{\sqrt{\mu_0 \rho}} \tag{2.3}$$

Using a typical coronal magnetic field strength of 10 gauss and density of 10^9 cm^{-3}, the above equation shows that particles could only be accelerated to a few keV in reconnection outflows. Though Sweet-Parker reconnection is not a good candidate to power the impulsive phase of flares because of long time scales and inadequate acceleration, it could be involved in preflare heating or as a flare-initiation process before a faster type of reconnection sets in.

Petschek's reconnection model [102], shown in figure 2.15 (right), proposes using slow-mode MHD shocks (solid blue lines) to redirect the plasma into the outflow region. The standing shocks form an X-region extending from a much smaller diffusion region ($L \sim \delta$). This configuration can redirect plasma flows well outside the diffusion region, bypassing the conservation of mass issues in the Sweet-Parker model. Petschek's model can achieve reconnection rates on the order of the flare impulsive energy release [120].

Collisionless Hall Reconnection

Steady-state magnetic reconnection in the Petschek model has been shown to provide a reasonably fast mechanism to remove energy from stressed magnetic fields. Petschek's model describes a steady-state mechanism: it does not address how reconnection is triggered and how the standing MHD shocks are generated. In a model suggested by Cassak et al. (2006) [28] collisionless Hall reconnection could provide a reasonable scenario to answer these questions.

The Hall model's pre-flare environment is undergoing a steady-state Sweet-Parker type reconnection. This allows the pre-flare environment to leak some energy as the overall fields become more stressed. The Sweet-Parker reconnection progresses until the diffusion region reaches a characteristic length scale at which point the region undergoes an abrupt transition to Hall reconnection, marking the beginning of the flare impulsive phase. The Hall reconnection rate has been shown to be up to six orders of magnitude faster than the rate predicted by the Sweet-Parker model in the corona [18]. The characteristic length scale is set by the thermal ion gyroradius in the presence of a strong out-of-plane (guide) magnetic field and otherwise set by the ion skin depth [29].

In the pre-flare environment, as Sweet-Parker reconnection progresses higher strength magnetic fields are drawn into the reconnection region. This causes the diffusion region's resistivity (η) to drop and its width (δ) to decrease [28].

$$\frac{\delta}{L} \sim \sqrt{\frac{\eta c^2}{4\pi v_A L}} \tag{2.4}$$

Where v_A is the Alfvén speed in the magnetic field upstream of the dissipation region.

Eventually the width of the dissipation region approaches the ion skin depth (d_i). At this length scale ions decouple from the magnetic fields, forming an ion diffusion region. The electrons remain coupled to the magnetic fields as the diffusion region width continues to decrease until the electron skin depth (d_e) is reached, at which point, they too decouple to form a smaller electron diffusion region. The ion and electron skin depths are a function of their plasma frequencies, w_{pi} and w_{pe}.

$$d_i = \frac{c}{\omega_{pi}} \qquad d_e = \frac{c}{\omega_{pe}} \tag{2.5}$$

$$w_{pi} = \sqrt{\frac{4\pi n e^2}{m_i}} \qquad w_{pe} = \sqrt{\frac{4\pi n e^2}{m_e}} \tag{2.6}$$

A current is generated within the diffusion regions by the difference in ion and electron motions. With this current the Hall term in Ohm's law can no longer be ignored.

$$\vec{E} + \vec{v} \times \vec{B} = \frac{1}{e n_e}(\vec{J} \times \vec{B}) \tag{2.7}$$

The current generated by the Hall term creates a characteristic quadrupolar magnetic field and introduces waves into the plasma. These waves are either whistler and/or kinetic Alfvén waves depending on β. Since both types of waves are dispersive, there will be an outflow jet created along the X-line. This sets-up the standing MHD shocks that are key to the Petschek reconnection model.

Evidence of Hall reconnection in space plasmas was first found in the Earth's magnetotail by the WIND spacecraft [98]. Figure 2.16 shows the details of the magnetotail geometry (left, top), the Hall reconnection geometry (left, bottom) and WIND's *in situ* measurements of the reconnection region (right). Earth's magnetosphere/magnetotail share similarities with flares; the outer field lines are approximately the same size as flaring loops, there is a current sheet between the North & South tail lobes and reconnection in the lobes leads to magnetospheric substorms. Since this observation, Hall reconnection has been measured *in situ* by other probes in Earth's magnetosphere [46] and at Mars [58], and has been demonstrated in laboratory plasmas [148]. A review of Hall reconnection, its role in space plasmas and possible application to flares can be found in Lin et al. 2011 [86].

2.5 Particle Acceleration

Magnetic reconnection alone cannot generate the large beams of highly energetic particles observed in flares. In Sweet-Parker type reconnection the particles can only be accelerated to a few keV. The diffusion region for Hall/Petschek type geometries is too small to energize a significant quantity of particles. So while magnetic reconnection is generally accepted as an integral part of the flare process, it is still unclear what processes are responsible for efficiently converting the magnetic potential energy into particle kinetic energy.

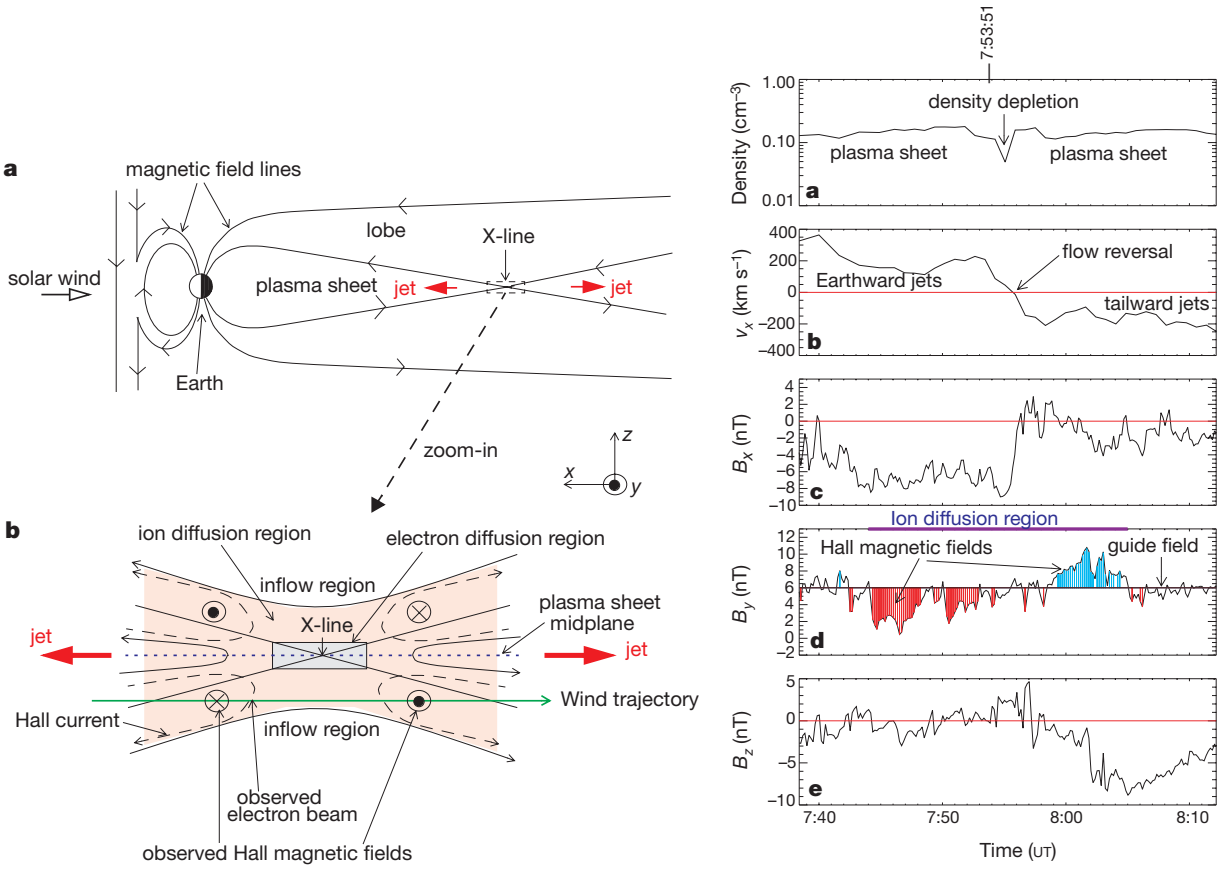

Figure 2.16: Panels describing the WIND *in situ* measurement of Hall reconnection in Earth's magnetotail. *Left, top:* Diagram of the magnetotail showing the X-line and plasma sheet that the spacecraft flew through. *Left, bottom:* Diagram of the Hall geometry showing the different ion and electron diffusion regions and indicating the quadrupolar magnetic field structure that is a signature of Hall reconnection. To the bottom are indicated the magnetic field regions that WIND passed through. *Right:* Plots of density (a), velocity (b) and magnetic field strength (c-e) as measured by the WIND spacecraft as it passed through Earth's magnetotail X-line. The magnetic fields indicated in panel (d) show that the spacecraft measured the predicted fields in the Hall ion diffusion region and indicates when it detected a guide field. Plots from Oieroset et al. 2001 [98].

Only a small fraction of the total energy released during a flare is liberated by the reconnection process itself. The primary role of reconnection is to change the magnetic topology of anti-parallel fields into turbulent regions composed of islands or loops, waves and shocks. The majority of energy release occurs when the turbulent fields relax into more potential configurations [86]. Particles are accelerated in the process. Good summaries of observations, candidate particle acceleration mechanisms and acceleration models are found in Benz 2017 [16], Raymond 2012 [110], Zharkova 2011 [150], Vilmer 2011 [142], Aschwanden 2002 [6] and Miller 2000 [91].

Acceleration Model Challenges

For an X-class flare the typical energy content in the HXR footpoints suggests that electrons need to be accelerated above 20 keV at a rate that can exceed 10^{37}/s [92]. Typical coronal densities ($\rho \sim 10^{11}/\text{cm}^3$) and loop volumes ($\text{V} \sim 10^{27}$) give a total available particle population of $\text{N} \sim 10^{38}$. For the particle acceleration rate predicted by observations, the entire acceleration region would be depleted of electrons in ∼10 s. This *number problem* poses a significant issue to possible acceleration models. The model needs to include a mechanism which replenishes the acceleration region with electrons on fast time scales so that acceleration can proceed through the duration of the impulsive phase.

Dennis et al. 2011 [40] consider the effects of such a large flow of particles through the flare loops. They estimate that a current of 10^{18} A would be generated as the particles flow from the coronal acceleration region into the chromospheric footpoints. For a flux tube of radius $\sim 10^9$ cm, the magnetic field associated with this current would be $B \sim 10^8$ G. Not only is the size of this field unrealistic, as it is five orders of magnitude larger than the strong fields above sunspots, but it would have an energy content of 10^{42} ergs, which is ten orders of magnitude greater than the energy released during the flare. Considering the self-inductance of such a current also presents problems, because as Dennis et al. point out, a current of this size cannot instantaneously appear. The current would generate a self-inductance of $\mathcal{L} \sim 10$ H, which would require a voltage of $V \sim 10^{18}$ V to initiate in 10 s. This is fourteen orders of magnitude higher than the typical accelerated electron energy.

A cospatial *return current* has been suggested as a method to overcome the challenges posed by such large acceleration currents. The return current could neutralize the beaming effects in the propagation region of the particle path, but is not viable in the acceleration region because the current would have to flow against the accelerating field. Detailed descriptions of the return current and the status of efforts to utilize it as a solution to coronal beaming issues is given by Zharkova et al. 2011 [150]. Though progress has been made in return-current theory and simulation, a complete solution has not been presented that can account for large-current issues in the acceleration region. The coronal-beaming problems have discounted models that rely on large-scale electric fields, making coronal acceleration models that feature stochastic processes more favorable.

A different solution to the number problem comes from Brown et al. 2009 [24] who propose that re-accelerating particles after they deposit their initial energy in the chromosphere

would reduce the total number of electrons required. They suggest that small-scale electric fields could provide a stochastic re-acceleration mechanism in the chromosphere. Siversky & Zharkova 2009 [124] suggest that particles travel back to the coronal acceleration region for re-acceleration after losing most of their energy in the chromospheric footpoints. Fletcher & Hudson 2008 [51] have suggested removing the site of particle acceleration from the corona altogether as a means to overcome beaming issues in the corona. In this scenario Alfvén waves propagate from the coronal reconnection site along the loop to the chromospheric footpoints where they locally accelerate particles.

Acceleration Mechanisms and Models

The possible acceleration mechanisms fall into three broad categories: acceleration by DC electric fields, shocks and wave resonance. The efficiency of each process varies depending upon the particle velocities such that it is possible for multiple processes occur at different stages of acceleration, generating seed populations that input into a second process.

DC Electric Field Acceleration

Particles accelerated by DC electric fields will feel a retarding force from interactions with other particles along its path, called the *Coulomb drag force*. The drag force is maximum when the particle reaches the thermal speed and decreases above this threshold. The Dreicer field ($\mathcal{E}_D$) is the value of the electric field ($\mathcal{E}$) which balances the electric Lorentz force and the Coulomb drag force. For typical flare parameters $\mathcal{E}_D \sim 10^{-4}$ V/cm [91]. For *sub-Dreicer fields*, where $\mathcal{E} < \mathcal{E}_D$, there is a critical velocity (v_c) below which the drag force dominates. Particles with speeds below v_c will be heated and particles with speeds greater than v_c will be accelerated to high energies. For *super-Dreicer fields* $\mathcal{E} > \mathcal{E}_D$, acceleration due to the electric field dominates for all particle velocities and the entire population can be accelerated to high energies.

A large scale super-Dreicer field could be generated in a reconnecting current sheet. It would be difficult to keep particles inside the current sheet due to a small magnetic field component that will act to dispel them through the sides. Miller notes [91] that this could also be beneficial, serving to replenish the acceleration region with new electrons. These large-scale fields, however, cannot provide the requisite particle energies as electrons are likely to escape after they reach ∼100 keV.

Shocks

Shocks are generated by a sudden, non-adiabatic transition, of particle velocities across the sound speed. In plasmas shocks are formed at the boundary between sub- and super-Alfvénic speeds. Particles can gain speed by interacting with the electric fields in the shock front, but must repeatedly interact with shocks to obtain significant acceleration. For randomly oriented shocks the particles will gain and lose energy stochastically and the net energy

will be gained as the square of the characteristic velocity, in the process known as *second-order Fermi* (or diffusive) acceleration. A converging set of shocks can accelerate particles systematically and energy will be gained linearly in what is known as *first-order Fermi acceleration*. If a particle is systematically redirected across the same shock front multiple times it will also gain energy proportionally to the characteristic velocity.

Shocks can be efficient accelerators, and have been shown to generate power-law spectra for simple particle distributions and situations [19]. They are a popularly invoked acceleration mechanism in astrophysics and are used to explain the acceleration and distribution of cosmic rays measured at Earth [50, 14, 8]. Shocks have been suggested for particle acceleration due to CMEs and have been shown to provide an acceleration mechanism for particles in the solar wind that generate Type II radio emission [105]. Shocks with orientations that are favorable for multiple particle crossings have been demonstrated in the Earth's magnetosphere and have been found in numerous areas of astrophysics. Zharkova et al. 2011 [150], however, point to (1) the need for a primary mechanism to generate an energetic seed population, (2) how simple model differences quickly deviate from power-law distributions, (3) the need for a guiding field to systematically redirect particles across a shock front and (4) the little direct evidence for shock fronts at loop tops as key issues for utilizing this shock scenario in the flare context.

Stochastic acceleration by moving magnetic islands was proposed by Drake et al. 2006 [43] as an efficient means to accelerate particles during reconnection. In this concept a tearing mode instability in the reconnecting current sheet generates multiple magnetic islands. As the islands relax they approximate moving magnetic mirrors, accelerating particles through numerous bounces. Interaction with many such islands accelerates particles to high speeds. Through numerical simulations this method showed that particles will mirror inside the magnetic island (figure 2.17, left) and this will cause them to systematically gain energy as a result (figure 2.17, right). Drake et al. 2009 showed that these systems can also pick-up thermal ions and accelerate them to Alfvénic speeds, at which point they too can enter the magnetic islands for higher energy acceleration. This 2D model becomes more complicated when extended into 3D.

Wave Resonance

Magnetic reconnection will generate an array of plasma waves which can be generated with a continuous distribution across spatial scales. Particles that travel with the wave's resonant velocity will be accelerated or decelerated depending on their phase difference with the wave. The resonance condition, given below, depends on the wave frequency (ω), the wavevector's component that is parallel to the magnetic field ($k_\parallel$), the particle's parallel component velocity ($v_\parallel$), the Lorentz factor (γ), the cyclotron frequency (Ω) and its harmonic (l).

$$\omega - k_\parallel v_\parallel = \frac{l\Omega}{\gamma} \tag{2.8}$$

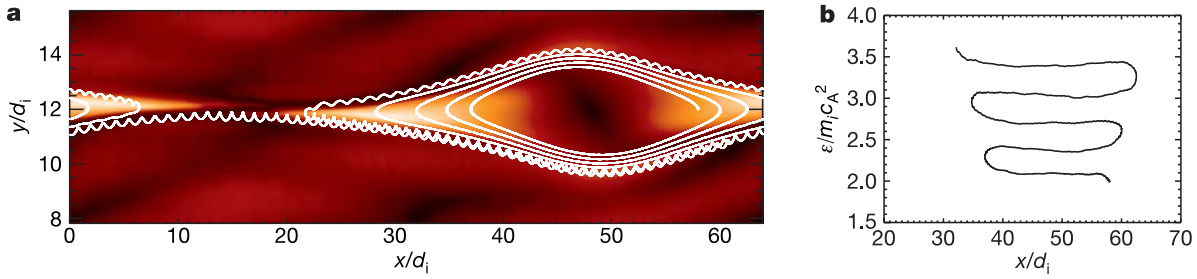

Figure 2.17: Simulations for stochastic acceleration of particles inside magnetic islands inside the reconnection region. *Left:* Simulated particle trajectory (white curve) showing how the particle mirrors off either end of the magnetic island. *Right:* Particle energy (ϵ) as a function of its position, showing how energy increases with each mirroring event. Plots from Drake et al. 2006 [43].

When a particle is in resonance with $l = \pm 1$ (*cyclotron resonance*), minute changes in the particle velocity will cause resonance with a different wave and the particle can jump to the higher (or lower) resonance, having gained (or lost) energy. The net effect of many such encounters through *resonance overlap* can lead to substantial energy gains as the particle climbs up the resonant modes. A turbulent cascade can generate many wave modes sequentially such that different particle populations are accelerated through cyclotron resonance at different times.

If a particle is in resonance with $l = 0$ (*Landau resonance*) it sees a constant electric field in its reference frame and can be continually accelerated. Particles in this resonance condition are said to *transit-time accelerate*, reaching high speeds quickly.

One limitation of the wave resonance model is that the seed population must have fairly high speeds, exceeding the wave resonance speed. Electron thermal energy in the corona is typically sufficient to meet this requirement and these species can be immediately accelerated by wave resonance. The greater mass of thermal ions, however, imparts them with lower thermal velocities which do not automatically meet this criterion. This leads authors to suggest multiple-mode acceleration schemes that pre-accelerate ions out of the thermal distribution. Miller [91] suggests using Alfvén waves to accelerate ions out of the thermal distribution and into high energies through cyclotron resonance. Fast-mode waves would then pick-up and accelerate the ions through Landau resonance to even higher speeds. The electrons could be immediately accelerated through the fast-mode process. Since the pre-acceleration of ions would take a non-negligible time to accomplish, Emslie et al. 2004 [47] pointed out that the time difference in the Miller model could possibly explain the spatial discrepancy between the electron and ion footpoint locations observed by Hurford et al. 2003 [66] & 2006 [67] (described in section 2.7).

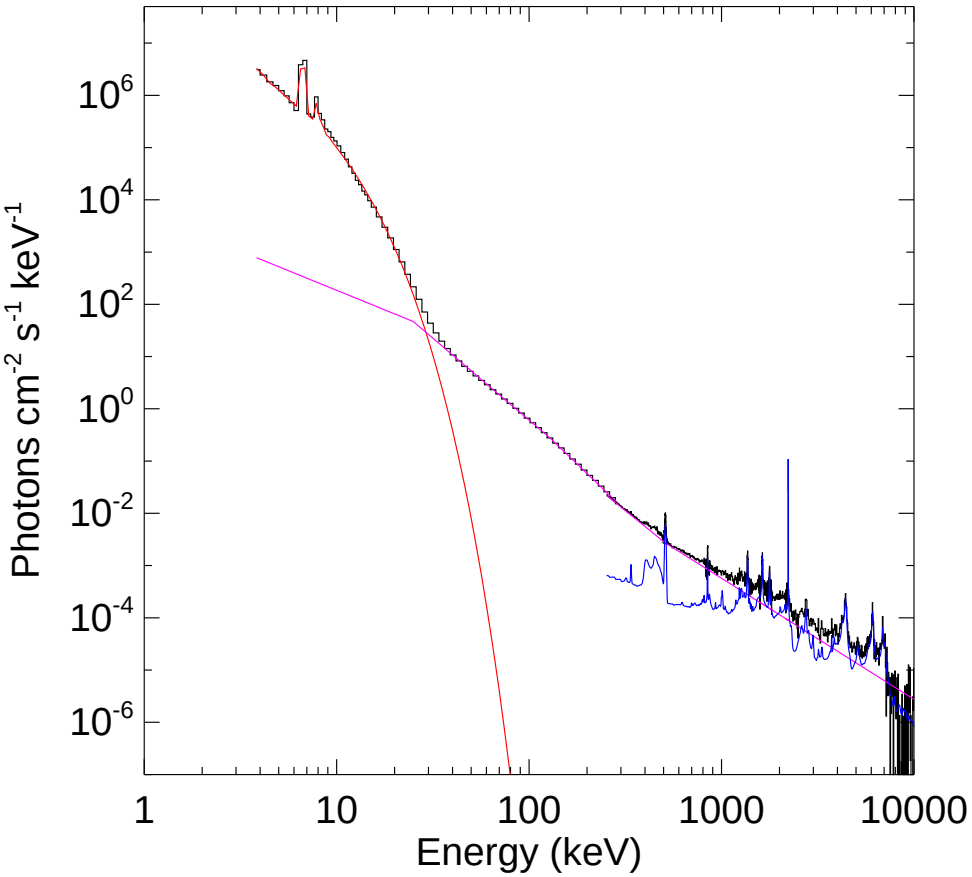

Figure 2.18: A HXR/gamma-ray spectrum from 3 keV to 10 MeV for the X4.8 flare on 2002 July 23, from RHESSI observations. Shown in color are the model fits to the thermal (red) & non-thermal (magneta) electron bremsstrahlung and gamma-ray line emission from accelerated ions (blue). The emission <30 keV is primarily due to thermal plasmas at ∼40 MK, and the spectrum >30 keV is a broken powerlaw dominated by accelerated electrons. Plot from Lin et al. 2011 [86].

2.6 The High-Energy Flare Spectrum

The HXR/gamma-ray flare spectrum is rich with information from a diversity of particle interaction processes that provide useful constraints on electron and ion populations. The spectral features typically observed in the HXR/gamma-ray energy range are shown in figure 2.18 (black) for an X4.8 flare that occurred on 2002 July 23. Plotted in color are model fits to the thermal Bremsstrahlung generated by ∼40 MK plasma (red), a broken-power law distribution created by non-thermal electrons (magenta) and a forest of lines (blue) that represent the summed contributions of narrow and broad lines from accelerated ions, an electron-positron annihilation line at 511 keV and the neutron capture line at 2.2 MeV.

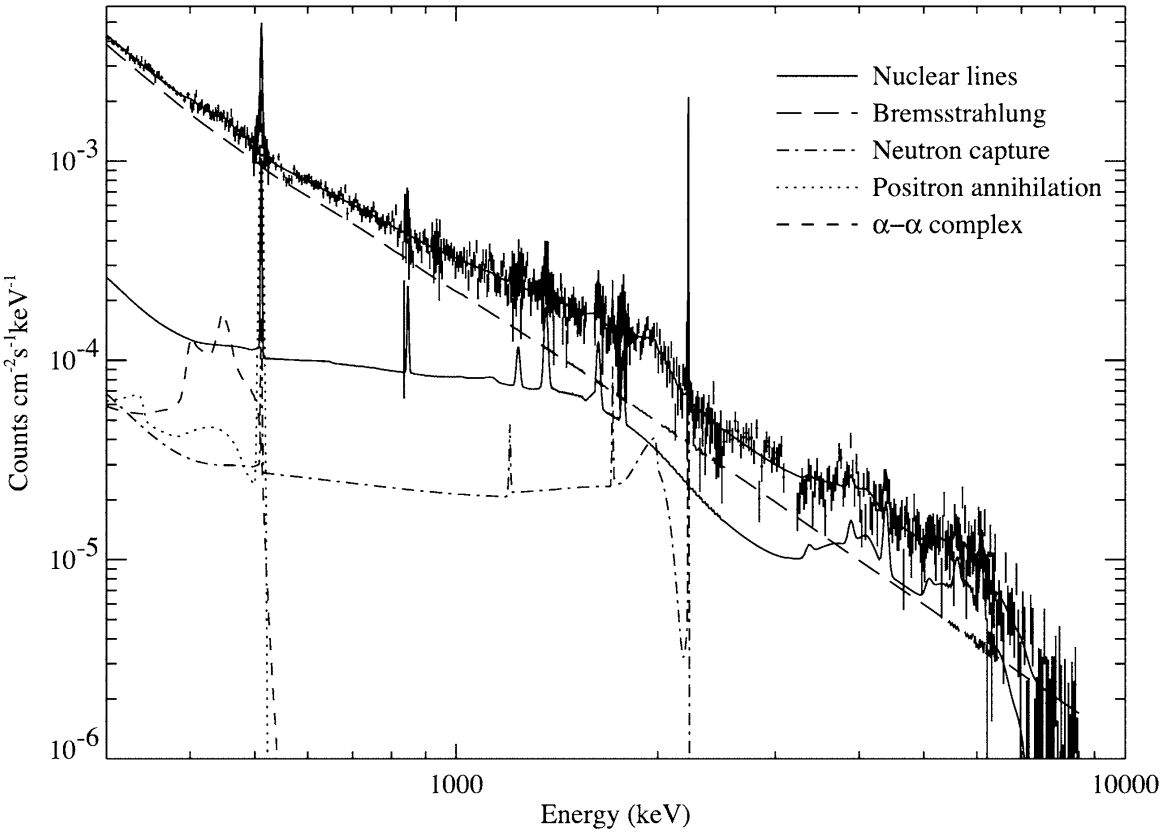

Figure 2.19: Gamma-ray model fits to the 2002 July 23 flare. Plot from Lin et al. 2003 [87].

The photon spectrum is generated by particles as they lose energy. Since this physics is well-understood the photon spectrum can be fit with a variety of models that describe the deceleration processes, and its these model fits which give information about the original particle distributions. Assumptions, such as the level of collisionality of the interaction (target) region, enter into the models and can have significant effects on the predicted photon spectrum. A *thin-target* model interaction region has so little density that collisional energy loss can be ignored. In the *thick-target* case, however, the collisions are so frequent that an accelerated particle is assumed to lose all of its non-thermal energy through collisions. Another assumption that can significantly alter the predicted spectrum is the optical thickness, which is used to scale the quantity of photons observed to the quantity actually produced. Recent reviews that give the details of observations and how they can be interpreted by electron and ion emission are given by Raymond et al. 2012 [110], Fletcher et al. 2011 [52], Holman et al. 2011 [63] and Vilmer et al. 2011 [142].

Electron Spectrum

Bremsstrahlung emission from thermal and non-thermal electrons are the primary features of the HXR/gamma-ray spectrum, giving it the overall smooth shape. A photon is produced when a charged particle decelerates and is diverted by the Coulomb field of another charged particle. The biggest contribution comes from electron-ion interactions, though electron-electron emission could be important at relativistic energies [75].

The efficiency of the bremsstrahlung emission is heavily dependent upon the ambient medium. The flare loop is usually modeled as a thin-target where there is relatively little HXR/gamma-ray emission and as a thick-target in the denser chromosphere where the majority of emission in this energy range appears in brightly localized footpoints (figure 2.12, panel a). A third HXR component above the loop top is also observed in many flares when their comparatively brighter footpoints are behind the solar limb (figure 2.12, panel b) or appear later in the flare (figure 2.12, panel c) when the footpoint intensity has decreased. The spectral distribution of the loop-top source is flatter than the spectrum in the footpoint sources, indicating that higher-energy electrons are possibly staying trapped in the loop top while the lower-energy electrons precipitate towards the footpoints earlier.

The thermal bremsstrahlung component is modeled as a Maxwellian and the non-thermal component is frequently observed as a broken power law whose spectral index changes in the range of 600 – 900 keV. In theory the electron distribution could be determined in a model independent way, by inversion of the bremsstrahlung spectrum [23]. The method relies on the fact that a bremsstrahlung photon can only be produced by an electron of equal or greater energy. The spectrum is then determined by starting with the highest observed emission and working towards the lower-energy end of the spectrum. This method, however, is limited by counting statistics.

The >30 keV spectral shape motivates modeling the population of non-thermal electrons as a broken power-law, which rolls over to higher-energies at 10s of keV [110]. Holman et al. 2011 [63] suggest that the break could result from a diverse range of possibilities, such as non-uniform ionization in the target region, effects due to the return current or the acceleration process itself. Kontar et al. 2007 [75] showed that by considering the electron-electron bremsstrahlung component for relativistic electrons, it is possible to generate a broken-power-law photon spectrum from a single-power-law electron distribution.

If the population of non-thermal electrons is beamed, as it is thought to be in the standard flare scenario, then its emission will be polarized. Scattering of lower-energy electrons will reduce the measurements of bulk polarization, but the relativistic electrons which create gamma-ray bremsstrahlung should have a stronger polarization signature. Polarization could provide a powerful diagnostic for determining the beaming properties of electrons in the coronal loops, which would help to validate the coronal acceleration region assumption of the standard flare scenario and constrain the particle acceleration mechanisms discussed in section 2.5. Though RHESSI was not designed for polarization measurements, there has been some progress in using the different detectors to infer HXR polarization through Compton scattering. Boggs et al. 2006 [20] and Suarez-Garcia et al. 2006 [132] report marginal

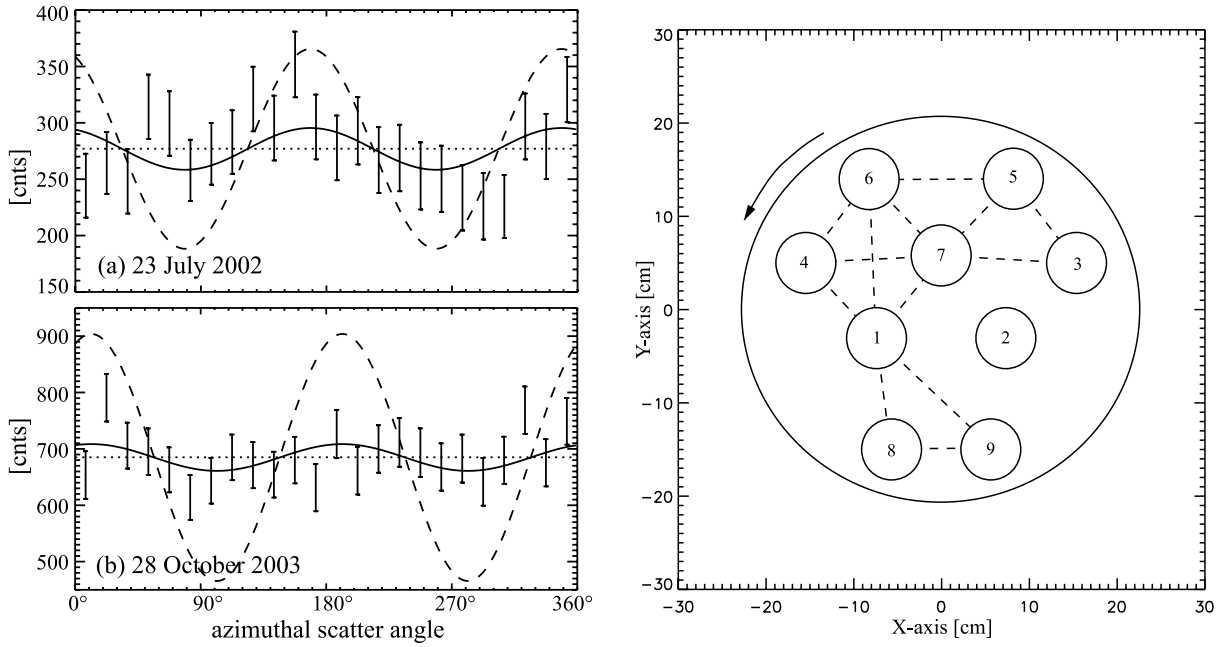

Figure 2.20: *Left:* Gamma-ray polarimetry measurements from the 2002 July 23 and 2003 October 28 flares. The best-fit modulation curve is plotted as a solid line and the 100% modulation curve is plotted as a dashed line for comparison. Polarization was found in both flares at levels of 21%±9% and -11%±5%, respectively. *Right:* Diagram showing the detector-to-detector Compton scatter tracks that were used for analysis. Figures from Boggs et al. 2006 [20].

polarization detections with RHESSI. Figure 2.20 (left) shows gamma-ray polarimetry results from the 2002 July 23 and 2003 October 28 flares from the Boggs study. The best-fit modulation curve is plotted as a solid line and the 100% modulation curve is plotted as a dashed line for comparison. The results from both studies indicate that an instrument with explicit polarization capabilities could have the potential of providing a well-constrained observation that can be used in field topology and acceleration models.

Ion Spectrum

The richness of the gamma-ray spectrum comes from the diverse interactions of accelerated ions as they generate lines through nuclear de-excitation, neutron capture and positron-annihilation. The overall contribution of ion-associated emission is shown as a blue trace in figure 2.18 and the individual components are plotted in figure 2.19. Information about the abundances of ambient and accelerated ions, and their energy spectrum can be inferred from the model fits. Detailed descriptions of the ion-interaction mechanisms can be found in Ramaty & Murphy 1987 [108] and Ramaty et al. 1979 [106], recent reviews of flare ion

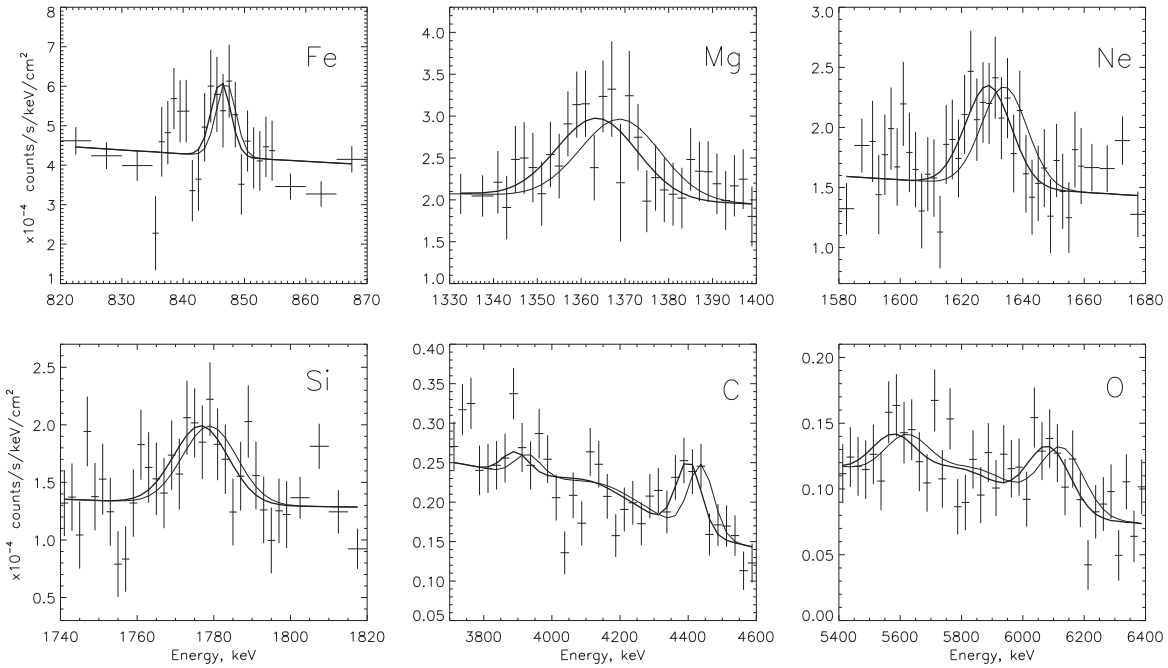

Figure 2.21: The first spectrally resolved gamma-ray lines from a solar flare. These panels show RHESSI count spectra (crosses whose width denote binning and heights are 1 σ measurement error) for the 2002 July 23 X4.8 flare. A second peak is visible in the carbon and oxygen panels, its due to escape of a 511 keV annihilation photon (described in chapter 3). The lines are fit with a redshifted Gaussian plus bremsstrahlung and broad line contributions (thick line), the non-redshifted line is shown (thin line) for comparison. Plots from Smith et al. 2003 [127].

interactions and their modeling efforts in the RHESSI era are given by Vilmer et al. 2011 [142] and Share & Murphy 2006 [117].

Flares can accelerate ions to GeV energies. *In situ* measurement of SEPs at earth show a power-law distribution with multiple segments with different spectral indexes. Motivated by these SEP observations and the well-observed power law distribution of energetic electrons, the accelerated-ion spectrum is also modeled as a power law. The ion spectral index is inferred by taking ratios of gamma-ray lines that have different threshold energies.

Nuclear De-Excitation Lines

Nuclei can be excited in a variety of processes that result in the emission of a characteristic gamma-ray upon de-excitation. Excitation processes include inelastic collisions, nuclear fragmentation and fusion reactions. The gamma-ray energy depends on the specific element and its level of excitation. Excitation lifetimes are short, typically $\sim 10^{-12}$ s, and these lines are appropriately termed *prompt emission lines*. Since the de-excitation time is short

compared to thermalization timescales the excited nucleus is likely to be still moving when it relaxes, resulting in Doppler shifts of the line energy.

The shapes and centroids of observed lines contain a wealth of information that can be used to understand the parameter space of accelerated and ambient abundances. The theoretical line width for de-excitation of a thermal nuclei is ∼100 eV for MeV de-excitation lines. The observed line width (in the absence of detector effects) will depend on the nuclei's recoil and thermal velocities; typical values are ∼100 keV. A randomly distributed ensemble of particles will have a Doppler-broadened line and the width of that line can be used to infer information about the particle interactions. The collision of energetic light nuclei on heavier particles will impart relatively little velocity to the heavy nuclei, resulting in a narrowly broadened line. This so-called *direct* reaction has typical line widths that are ∼1 – 2% of the de-excitation photon energy. Broad lines result from the inverse scenario, wherein heavy particles collide with lighter particles in an *indirect* reaction. In this case more energy is imparted to the excited nuclei, and their Doppler-broadened widths can be as much as ∼20% of the line energy.

The first spectrally resolved observation of prompt nuclear de-excitation lines came from RHESSI in the 2002 July 23 flare. The lines and their model fits are shown for iron, magnesium, neon, silicon, carbon and oxygen in figure 2.21. The fit redshifts suggest a population of accelerated particles that are not traveling radially inward, towards the solar surface. The redshifts decrease with increasing mass, indicating that they are traveling away from the observer and are consistent with loops that are tilted towards the Earth by ∼40° [127]. The model fit used for the neon de-excitation line at 1634 keV is shown in figure 2.22. The combined model profile is generated by considering the separate contributions of accelerated protons and accelerated α particles, providing a constraint on their accelerated ratio.

Neutron Capture Line

The neutron capture line is the thinnest and most prominent gamma-ray line. Particles with energies ∼20 MeV/nucleon create neutrons in nuclear-fragmentation interactions. The energetic neutrons travel deep into the solar atmosphere before thermalizing in the high-density photosphere. After losing sufficient energy the neutrons are captured by either photospheric helium-3 or hydrogen. Capture by hydrogen results in the formation of a deuteron and the release of a 2.223 MeV photon. Capture by helium-3, however, does not result in a measurable spectral signature. Since the neutrons have thermalized, the line is Doppler-broadened by thermal velocity only, resulting in a very narrow line. Thermalization delays this line emission by ∼100 s relative to the prompt nuclear de-excitation lines. The photosphere is slightly opaque to gamma-rays and this must be accounted for in 2.223 MeV photon-to-neutron model fits.

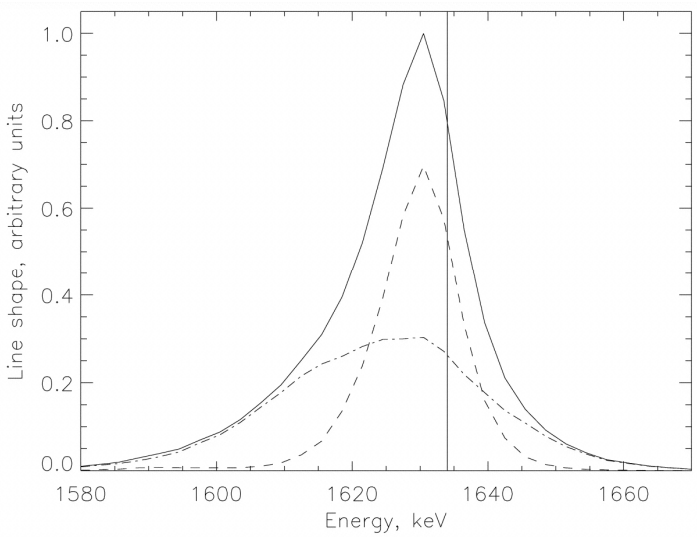

Figure 2.22: Neon 1634 keV de-excitation line model for a downward-isotropic angular distribution of accelerated protons and α particles propagating in 30° tilt. The solid line shows the combined contributions of a narrow proton-induced (dashed) line and a broader and more heavily redshifted α-induced line (dash-dot) in an α/proton ratio of 0.5. The vertical line marks the de-excitation line for neon at rest. Plot from Smith et al. 2003 [127].

Positron-Electron Annihilation Line

Positrons are created in particle collisions that result in radioactive nuclei and pions. Energetic positrons lose energy through bremsstrahlung and particle collisions before eventually annihilating with ambient electrons. Annihilation can be direct, where the positron and electron immediately annihilate into two 511 keV photons. Or, the electron and positron can become bound together to form the exotic atom positronium. Positronium is unstable and the two particles will eventually annihilate by creation of two 511 keV or three gamma-rays (in a continuum up to 511 keV) depending on the spin state.

Ambient plasma temperature, ionization state, composition and density can affect the positron-electron annihilation channel. Spectral fits to the 511 keV line can constrain these atmospheric parameters. Figure 2.23 shows the 511 keV line for the 2003 Oct 28 flare in two time intervals, illustrating how the interaction region changes over the course of the flare [119]. A continuum component from orthopositronium and from Compton scattering is visible in the first time interval, indicating that the ambient hydrogen density was greater than $\sim 10^{14}$ /cm^3. The line broadness during this interval ($\sim$5 – 8 keV) is consistent with annihilations in ionized medium at temperatures $>10^5$ K. In 2 minutes the line abruptly narrows to $\sim$1 keV, indicating a density of $>10^{15}$ /cm^3 with a temperature $<10^4$ K. The sudden narrowing of the 511 keV line suggests that the annihilation region can be heated to transition-region temperatures and then suddenly cool, while maintaining high ionization. Adding to the mystery is that no significant changes occurred in the bremsstrahlung, nuclear

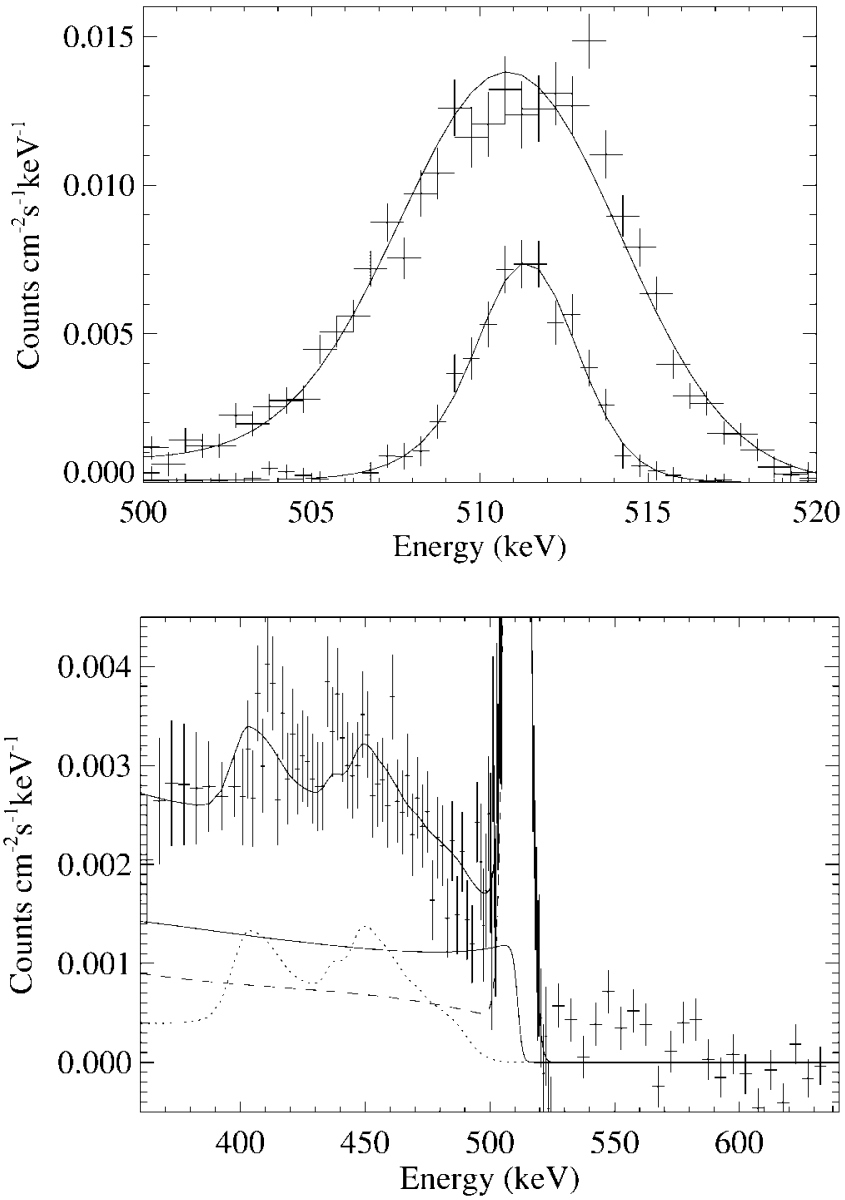

Figure 2.23: *Top:* 511 keV spectra from the 2003 October 29 flare in two intervals when the line was broad (5 – 8 keV, 11:06 – 11:16 UT) and when it suddenly narrowed (∼1 keV, 11:18 – 11:30 UT), indicating that the temperature of the annihilation region dropped from $>10^5$K to $<10^4$ within 2 minutes. *Bottom:* Model fit to the broad 511 keV line shown above, also visible in this image are two peaks from the α-He^4 line complex. The model components are an orthopositronium continuum (solid), annihilation line Gaussian (dashed) and a component from the α-He^4 line complex (dotted). Plots from Share et al. 2004 [119].

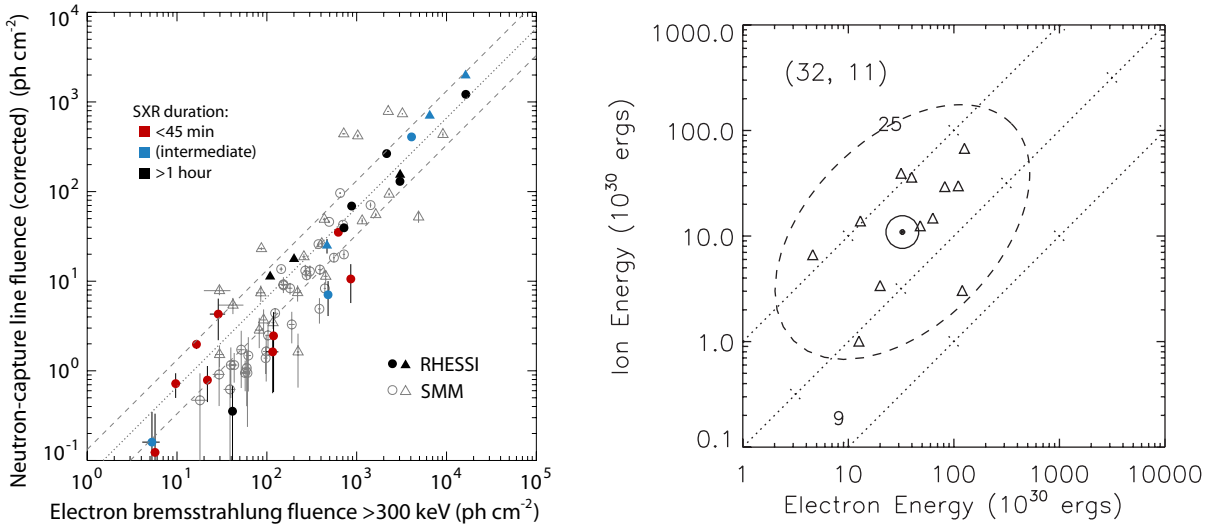

Figure 2.24: *Left:* Correlation study results from Shih et al. 2009 [121] showing that there is a tight relationship between the relativistic electron bremsstrahlung and the 2.2 MeV line fluence. This suggests that ions and relativistic electrons are accelerated by similar processes. The correlation is examined across two solar missions (RHESSI and SMM) and is found to extend over 3 orders of magnitude, from the largest flares to the limits of detectability by RHESSI. *Right:* The energy content of ions vs. electrons in units of 10^{30} ergs. The flares studied by Emslie et al. 2012 [48] show that the ion and electron energies are generally comparable, with the exception of two events in which the electron energy content was $\sim$2$\times$ higher.

or annihilation line fluxes when the line narrowed.

2.7 Connecting Ions and Relativistic Electrons

RHESSI and its predecessor SMM provided evidence that flares accelerate ions and relativistic electrons proportionally. Figure 2.24 (left) shows the results of a 2009 correlation study by Shih et al. that compares fluences between the 2.2 MeV neutron capture line and the >300 keV bremsstrahlung from relativistic electrons. The fluences in these two components show a linear proportionality that extends over three orders of magnitude, from the biggest flares through the limits of detection for RHESSI. This tight correlation suggests that a single mechanism is responsible for accelerating the $\geq$30 keV ions and >.3 MeV relativistic electrons.

Modeling of the spectral components indicate that both of the non-thermal electron and ion populations ($F(E)$) can be represented by power-law distributions in energy (E), whose

indexes (δ) and amplitudes (A) are determined by spectral fits. [86].

$$F(E) = AE^{-\delta} \tag{2.9}$$

In determining the total energy content of ions and electrons the *low-energy cutoff*, or *roll-off*, parameter is critical. From observations of neon-20 in a few large events the ion spectrum has been seen to extend down to ∼3 MeV [107]. There is virtually nothing known about ion energies below 1 MeV, and this is usually taken as the low-energy roll-off. Emslie et al. 2012 [48] describe the method they took to estimate the energy content of ions from measurement of the 2.2 MeV line. The basic procedure entails correcting the measured 2.2 MeV line fluence for attenuation through the solar atmosphere then using this value to determine the energy content in >30 MeV protons. To extrapolate the spectrum down to 1 MeV they chose an ion spectral index of 4, based on results of big flares whose indexes ranged between 3 – 5. The results are plotted in figure 2.24 (right) against the energy content of electrons.

The non-thermal electron spectrum is modeled as a broken-power-law that is forward-fit to the spectral data to get the indexes, break-point and energy roll-off. In producing figure 2.24 (right) Emslie et at. used the largest roll-off value that will fit the data, this produced lower-limit energies for electron energy content. Typical values for the non-thermal electron energy cut-offs are in the ∼20 keV range. Emslie et al. found that the total energy content in ions and electrons are roughly comparable and accounts for approximately 20% of the energy liberated during a solar eruptive event (with the remaining 80% going into the CME).

RHESSI provided the first gamma-ray flare image with the 2002 July 23 X4.8 flare. This image placed the ion-associated line emission in a compact source within the flaring region [66]. This image showed for the first time that the ions were accelerated by the flare itself and not by large-scale processes like CME-associated shocks, as was suggested during the SMM era [139]. The spatial extents of the flaring loops were too small to resolve the individual footpoints using RHESSI's finest resolution grid at gamma-ray energies (35") and a single 2.2 MeV centroid was imaged. This event showed a surprising result, the centroids of the ion and electron emission were displaced from one another by 25" ± 5" [67]. This spatial separation was unexpected because particles that are energized together, as the correlation results suggest, are thought to be transported out of the acceleration region along the same field lines and travel along the same loops into the chromosphere where they would have similar emission source locations.

The 2003 October 28 X17 event, shown in figure 2.25, is the only flare captured by RHESSI with a strong enough fluence and large enough spatial extent to image the separate 2.2 MeV flare foot points [67]. The 2.2 MeV and 300 – 500 keV footpoints are separated by 14" ± 5 & 17" ± 5", sharing the unexpected spatial result with the 2002 July 23 flare. Only three additional flares have been imaged at the 2.2 MeV line. The 2002 November 2 event shows a significant displacement of 11" ± 5" of the ion/electron centroids, and the single ion centroid may be preferentially associated with one of the two electron footpoints [67]. The 2005 January 20 X7 event had a similar result with only one 2.2 MeV centroid associated

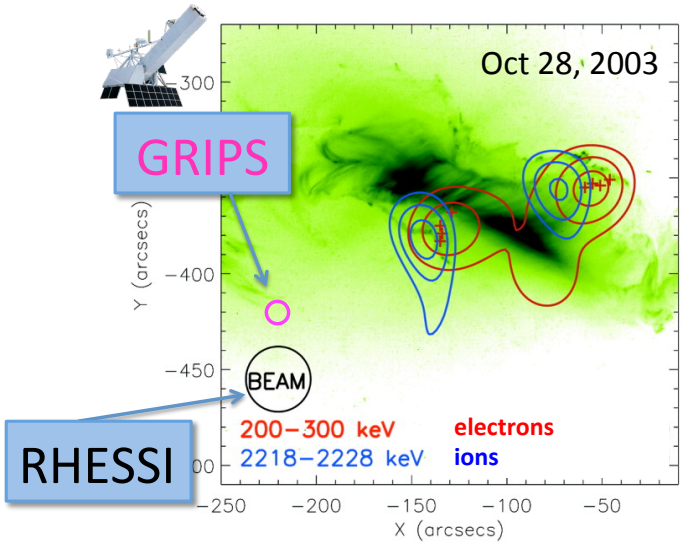

Figure 2.25: *Right:* RHESSI gamma-ray image in two different energy bands (200 – 300 keV & 2218 – 2228 keV) of the GOES class X17 flare on 2003 October 28, overlaid on a TRACE 195 Å image. The electron-associated bremsstrahlung emission is shown in red contours, and the imaged 2.2 MeV line is shown in blue contours. The electron and ion centroids are separated from one another by 14" ± 5" & 17"± 5" in each footpoint. Inset is a comparison of the GRIPS and RHESSI minimum beam widths illustrating how much finer resolution a GRIPS observation could have provided for this flare. Plot adapted from Hurford et al. 2006 [67].

with one of the footpoints [142]. The 2003 October 29 X10 event, shown in figure 5.7, could only be imaged using the coarser (183") of RHESSI's two gamma-ray imaging grids [67]. This event's 2.2 Mev centroid displacement was found to be a statistically insignificant 13" ± 17". The 2003 October 29 event is the only flare in the set of gamma-ray imaged flares with a inconclusive separation result and due to large background it has been relatively poorly studied and consequently was not included in Shih et al.'s or Emslie et al.'s correlation studies. Chapter 5 explores this flare in detail.

There has not been a satisfactory model explanation that predicts similar ion and relativistic electron energy content and also produces the observed footpoint separation, though attempts have been made. Emslie et al. 2004 [47] provided an explanation to the 2002 July 23 results using a stochastic acceleration model which predicted that ions and electrons would be accelerated in different loop sizes. The October 28 event, however, showed that the ion and electron loops are of similar sizes and orientations. Hurford et al. 2006 [67] find that gradient and curvature drifts would produce electron-ion separation in the correct direction, but for typical coronal parameters these drifts can only produce separations that are two orders of magnitude smaller than the observed separation. Five imaged flares is too small of a set to draw systematic conclusions about the spatial separation. Expanding

upon this small group of gamma-ray flare observations is necessary to explain the seemingly inconsistent proportional acceleration and spatial separation results.

2.8 The Future of High-Energy Flare Studies

Additional gamma-ray flare observations with fine spectroscopy, explicit polarimetry capabilities, enhanced sensitivity and improved imaging resolution that can separate the individual 2.2 MeV footpoints across multiple flare sizes are needed to address the current era of particle acceleration and transport questions. With these capabilities the ion/relativistic electron connection can be constrained, the beaming properties of particles investigated and the variations of accelerated and ambient material studied.

The Gamma-Ray Imager/Polarimeter for Solar flares (GRIPS) instrument is designed to meet the observational goals of a next generation solar telescope. GRIPS can provide $\sim$2$\times$ the photon throughput and $\sim$3$\times$ finer angular resolution than RHESSI at gamma-ray energies. These improvements will extend the limits of detectability for gamma-rays, and with 12.5" imaging resolution provide the ability to separate the footpoints in most gamma-ray flares. GRIPS' ability to measure polarization for photons $\geq$150 keV will allow for targeted studies of particle beaming and its germanium detectors have the fine resolution needed to investigate accelerated particles and ambient plasma conditions using the gamma-ray line spectrum. GRIPS had a successful first flight and technology demonstration aboard a long-duration balloon platform in January 2016, during the descending phase of solar cycle 24. The instrument, its key systems and Antarctic flight are described in chapters 6 – 9.

Chapter 3

Gamma-Ray Detection and Measurement

Gamma-ray interactions with matter are used to infer information about the original photons. There are many processes by which photons interact with matter, but only those in which the photon deposits some, or all, of its energy in the detector are relevant to photon detection. In the 1 keV – 10 MeV energy range there are three dominant processes that lead to detection: photoelectric absorption, Compton scattering and pair production. In each of these processes the energy of the original photon is transferred into scatter products (electron, positron and/or reduced-energy photon) whose combination depends upon the interaction type. Complete detection and summation of each absorption or scatter product energy would yield a measurement of original photon energy. The finite sizes of detectors and efficiency with which they collect and resolve the interaction products leads to complex spectra in this range of photon energies. Many of the topics covered in this chapter apply broadly to radiation detectors in the energy range of interest. Detailed discussion in this chapter is focused on semiconductor detectors because their superior energy resolution to scintillators and higher stopping power than gas proportional counters makes them the detector of choice for applications requiring high spectral resolution. Much of the discussion in this chapter is adapted from Knoll 2010 [74].

3.1 Photon Detection Basics

Since photons are not charged they can not directly interact with the detector medium; instead they must deposit some quantity of their energy in the detector to ionize atoms and transfer energy to electrons. Measuring the electron energy then provides a means to determine how much energy the original photon deposited in the detector. A good detector must therefore have sufficient photon stopping power and the ability to measure electron energy precisely. The $\sim$1000 times higher density of scintillation and semiconductor detectors puts them in favor over gas proportional counters because they can achieve much higher

stopping power in compact sizes.

Ideal Detector Response

In this section a model is built that describes the output of a single-photon-counting detector. The model assumes the ideal case where the entirety of the original photon energy is deposited in the detector and transferred to charge carriers that are collected without energy loss. The methods by which the photon energy is deposited, transferred and collected varies by detector type, material and configuration; these details will be described later in the chapter. The description provided here is the general case which is broadly applicable to single-photon-counting detectors which absorb and detect the full photon energy.

A detector converts the energy deposited by a photon into N charge carriers and those charge carriers are turned into an equivalent pulse of height H_0 that is digitized and recorded as a single count. Each count is histogrammed and displayed in what is termed a *differential pulse height spectrum*. For linear detectors the pulse height is simply a multiple (K) of the quantity of charge carriers. In the absence of statistical fluctuations all the photons collected would generate the same quantity of charge carriers and the output would be a delta function at H_0.

Real systems contain statistical fluctuations that will tend to vary the quantity of charge carriers produced. If the fluctuations are random and N is large then the quantity of charge carriers produced follow Poisson statistics and the distribution will have a standard deviation (σ) of $\sqrt{N}$. The output of the ideal detector would therefore be a Gaussian whose integral provides a measure of the total photon count. This peak at H_0 is termed the *photopeak* since it represents the full photon energy. An example photopeak differential pulse height spectra ($\frac{dN}{dH}$) is shown in figure 3.1.

$$\sigma = \sqrt{N} \tag{3.1}$$

$$H_0 = K\,N \tag{3.2}$$

Though many factors enter into the *Full-Width at Half Maximum* (FWHM) of a real detector system, the fundamental theoretical limit is provided by statistical fluctuations. For a linear detector response, the pulse height spectra's FWHM would simply be a product of the Gaussian FWHM and the proportionality constant, K.

$$FWHM = 2.35\,K\,\sigma = 2.35\,K\,\sqrt{N} \tag{3.3}$$

Energy Resolution

Two Gaussians are just distinguishable when they are separated by their FWHMs. The *energy resolution* (R) is defined as the ratio of the distribution's FWHM to the pulse height. Since the FWHM and pulse heights are related to the total charge carriers produced, the

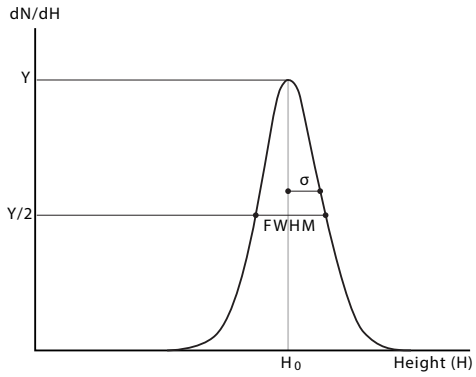

Figure 3.1: The ideal detector response for complete energy collection of a monoenergetic beam of photons. Image rendered by Cole Kelleher based on Knoll 2010 [74].

system energy resolution is fundamentally limited by how many charge carriers are produced for each photon interaction.

$$R\Big|_{Poisson} \equiv \frac{FWHM}{H} = \frac{2.35}{\sqrt{N}} \tag{3.4}$$

The quantities of charge carriers created by different detectors vary widely and are largely dependent on the steps required to generate an electrical signal. In scintillation counters, the incident photon energy is converted to energetic electrons, which then lose energy to the scintillation medium by exciting a proportional quantity of atoms. These atoms then relax to generate a proportional amount of secondary photons that are read out and turned into an equivalent electrical signal by detector electronics. This process involves many inefficient steps such that the energy required to generate a single charge carrier is ∼100s of eV. The relatively straightforward process of energy loss by kinematics and bremsstrahlung to generate a cascade of secondary electrons in a semiconductor is a comparatively efficient process, requiring only ∼3 eV to generate a charge carrier. Figure 3.2 illustrates the resolution differences between a sodium iodide (NaI) scintillation detector and a lithium-drifted germanium solid-state detector (Ge(Li)). The superior energy resolution achievable in solid-state detectors have made them the standard for applications requiring high spectral resolution.

In practice it is possible to achieve resolutions 3 – 4x better than the value predicted by Poisson statistics alone. Poisson statistics assume that each charge carrier is independently generated and random statistics govern the excitation process. Measured line resolutions indicate that this is not a good assumption for all detector types. The *Fano factor* (F) is used to correct for the discrepancy between the differences in observed and Poisson-calculated statistical fluctuations. Scintillators have Fano factors close to unity whereas semiconductor

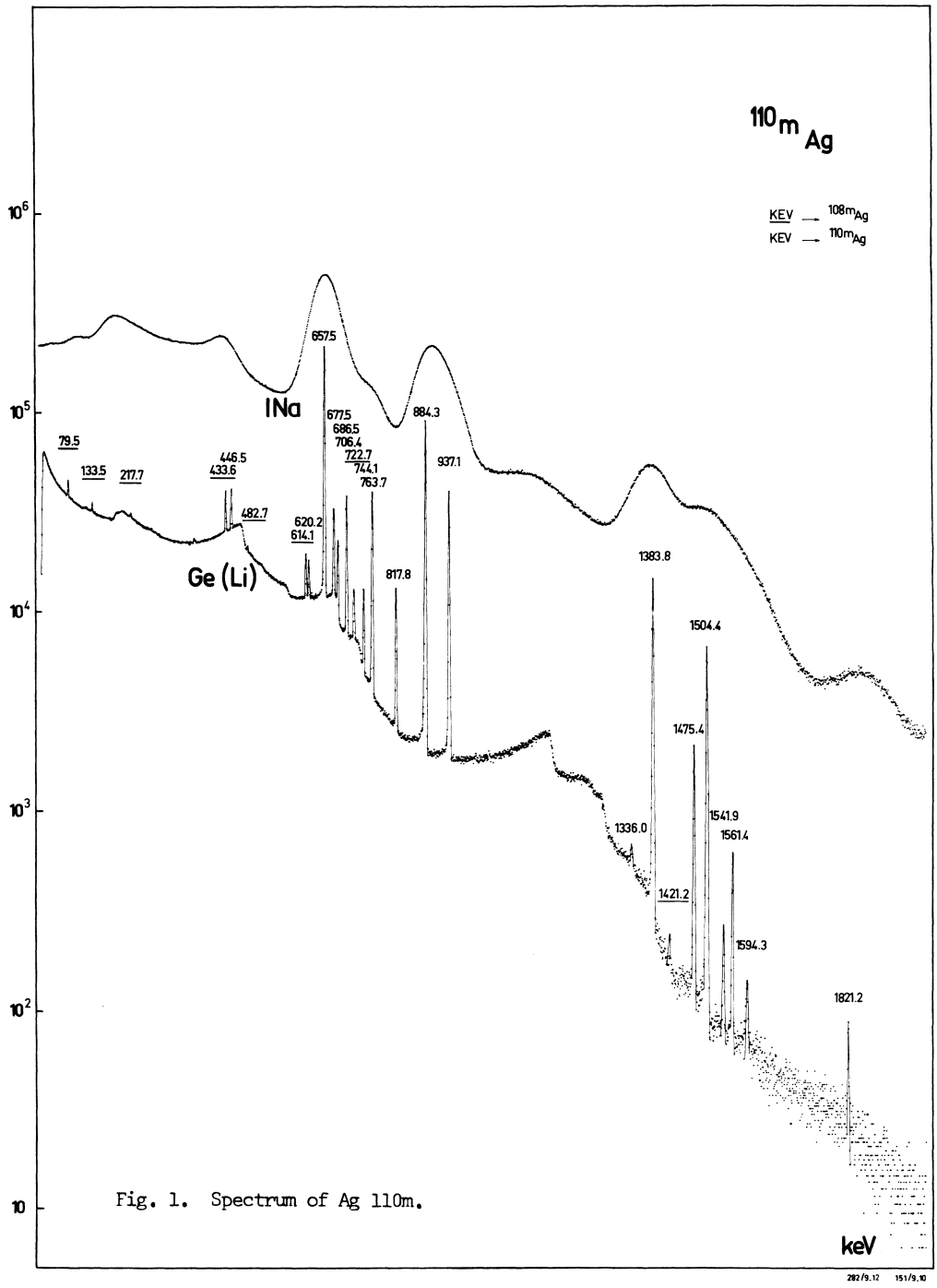

Figure 3.2: Comparison of pulse height spectra illustrating resolution differences between a sodium iodide (NaI) scintillation detector and a lithium drifted germanium Ge(Li) detector for an Ag-110 source. Peaks are labeled in keV. Image from Philippot 1970 [103].

detectors have values typically ∼.1.

$$FWHM_{Fano} = 2.35\,K\sqrt{F\,N} \tag{3.5}$$

$$R\Big|_{Fano} = \frac{FWHM_{Fano}}{H} = 2.35\sqrt{\frac{F}{N}} \tag{3.6}$$

A number of different factors will contribute to statistical fluctuations in a real detector system. If all the fluctuations are symmetric and independent of one another, then the resulting count distribution will tend towards a Gaussian centered at the nominal pulse height. This prediction holds even if the sources of fluctuations have a non-Gaussian distribution themselves. The FWHM of the resulting distribution is the sum of the individual FWHMs of each source added in quadrature.

$$FWHM_{tot} = \sqrt{FWHM_{Fano}^2 + FWHM_{noise}^2 + ...} \tag{3.7}$$

3.2 Photon Interactions with Matter

Three interactions are of interest to measurement for 1 keV – 10 MeV photons: photoelectric absorption, Compton scattering and pair production. Coherent scattering processes, which cause the incident photon to be diverted without energy transfer to the detector medium, are irrelevant to photon detection but their ability to redirect without energy loss is ideal for photon imaging (section 4.3).

A beam of photons passing through a medium will be attenuated as they interact with the material. This attenuation can be described by a decaying exponential that relates the exiting beam intensity (I) to the incident photon beam intensity (I_0) by a *mass attenuation coefficient* (μ) and the material thickness (t).

$$I = I_0 e^{-\mu t} \tag{3.8}$$

The mass attenuation coefficient is an energy-dependent quantity that accounts for absorption and scattering processes: photoelectric absorption (τ), Compton scattering (σ) and pair production (κ).

$$\mu = \tau + \sigma + \kappa \tag{3.9}$$

Figure 3.3 shows the combined mass attenuation coefficient as a function of photon energy, and plots the individual contributions of the absorption and scattering processes for Germanium. The sharp features in the photoelectric absorption are from the k-shell (11.1 keV) and L-shell (1.4 keV) edges. This plot illustrates that the different processes tend to dominate in certain energy bands.

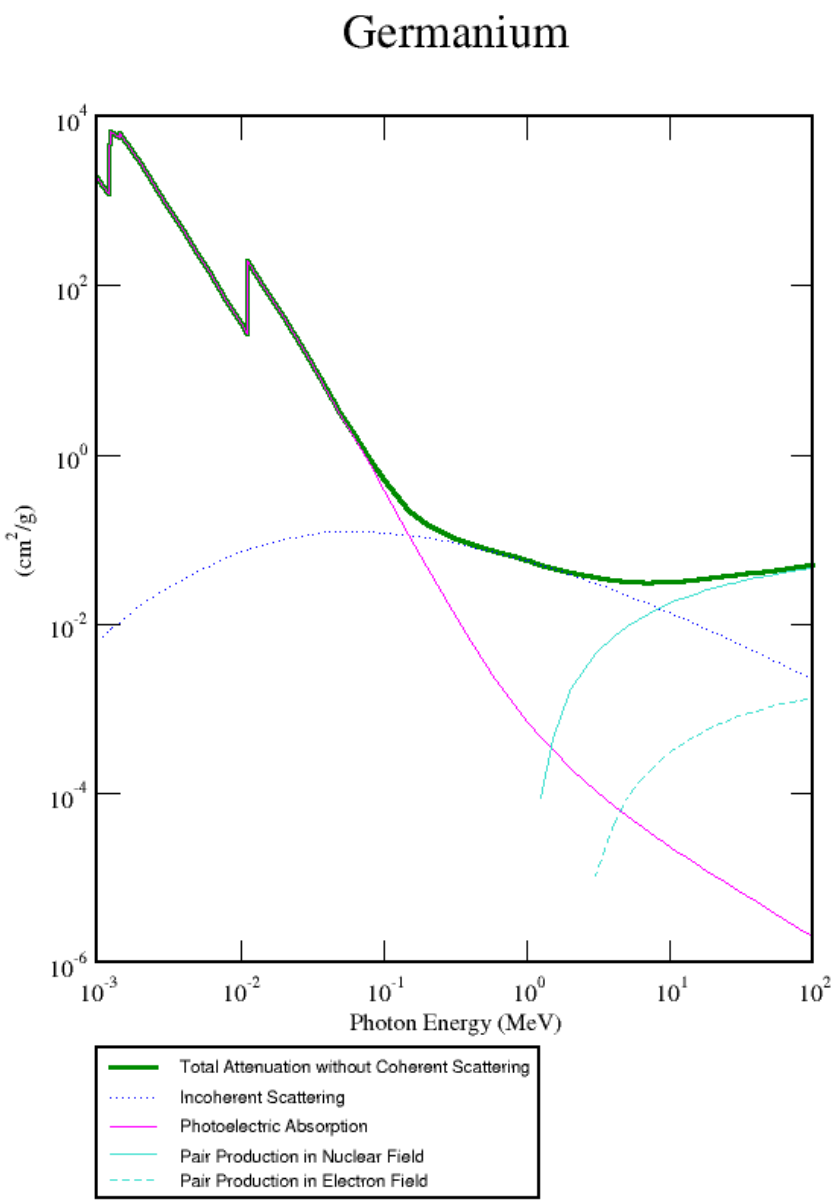

Figure 3.3: The absorptive and scatter processes contributing to the total mass attenuation coefficient are plotted for Germanium in the range of 1 keV – 100 MeV. Shown are photoabsorption, Compton scattering (here called incoherent scattering), and pair production in the nuclear and electron fields. Plot provided by XCOM at NIST [**XCOM**].

Photoelectric Absorption

Photoelectic absorption, or photoabsorption, occurs when a photon deposits its full energy into a material by interacting with a single atom. The original photon disappears and a photoelectron is ejected from the atom. The photoelectron kinetic energy (E_{e^-}) equals the photon energy (E_γ) minus the binding energy required to liberate the electron from its shell (E_B). The energetic photoelectron then loses energy through kinematic and radiative interactions with other particles, generating a cascade of secondary electrons.

$$E_{e^-} = E_\gamma - E_B \tag{3.10}$$

At gamma-ray energies the photoelectron is usually ejected from the tightly bound K-shell, leaving the atom in an excited state. The atom relaxes by emission of a characteristic X-ray or Auger electron of energy equivalent to the photoelectron binding energy. If the detector is of sufficient volume, the secondary X-ray or Auger electron then interacts with atoms to liberate more electrons in another cascade. If nothing escapes from the detector, the total energy of all the electrons equals the initial photon energy and the detector output will be a single photopeak.

In detectors whose volume-to-surface area ratio is small, or for interactions occurring close to the surface of the detector, the chance of characteristic X-ray escape increases. If the secondary X-ray escapes, the full photon energy is not deposited and one or more *X-ray escape peaks* can occur below the full photopeak energy.

Compton Scattering

A photon Compton scatters when a portion of its original energy is deposited in a single-site interaction, resulting in a recoil-electron and a scattered photon of reduced energy. Kinematics and conservation of energy lead to relationships between the energy of the initial photon (E_γ) and scatter products (E_{e^-}, $E_{\gamma'}$) as a function of the angle between them (θ) and the electron rest energy (m_0c^2).

$$E_{e^-} = E_\gamma - E_{\gamma'} \tag{3.11}$$

$$E_{\gamma'} = \frac{E_\gamma}{1 + (E_\gamma/m_0c^2)(1 - cos\theta)} \tag{3.12}$$

In the case of grazing incidence ($\theta = 0$) there is minimum energy transfer to the recoil electron and the detector will not measure any energy deposition. Maximum energy is transferred to the recoil electron in a direct collision ($\theta = \pi$) where the scattered photon reverses direction.

$$E_{\gamma'}(\theta = 0) \simeq E_\gamma \qquad E_{e^-}(\theta = 0) \simeq 0 \tag{3.13}$$

$$E_{\gamma'}(\theta = \pi) = \frac{E_\gamma}{1 + 2E_\gamma/m_0c^2} \qquad E_{e^-}(\theta = \pi) = E_\gamma \frac{2E_\gamma/m_0c^2}{1 + 2E_\gamma/m_0c^2} \tag{3.14}$$

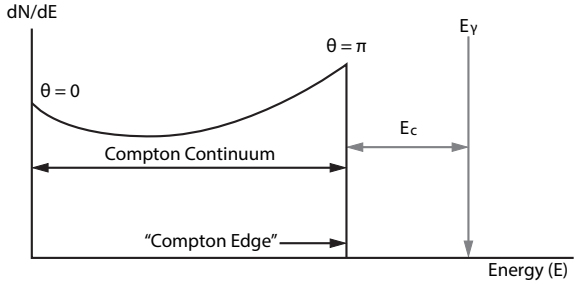

Figure 3.4: Diagram of the Compton scatter response for a beam of photons with energy E_γ. The Compton edge is the maximum energy transmitted to a recoil electron in a Compton scatter. The shape of the Compton continuum is given by the spin-averaged Klein-Nishina equation. Image rendered by Cole Kelleher based on Knoll 2010 [74].

The maximum recoil-electron energy is called the *Compton edge*, it represents the maximum detectable energy in a single Compton event. The gap between the Compton edge and the full photopeak energy (E_C) is a function of the incident photon energy. In the limit where the initial photon energy is large the scattered photon energy tends towards a constant value of .256 MeV.

$$E_C = E_\gamma - E_{e^-}|_{\theta=\pi} = \frac{E_\gamma}{1 + \frac{2E_\gamma}{m_0 c^2}} \tag{3.15}$$

$$\lim_{E_\gamma \mapsto \infty} E_C(\theta = \pi) \simeq \frac{m_0 c^2}{2} = .256\, MeV \tag{3.16}$$

Since all values of scattering angle are possible, a continuum of electron energies extends from zero to the Compton edge. This continuum shape is determined by the likelihood of a gamma-ray scattering into a certain angle. The angluar distribution of Compton scattered photons is given by the differential scattering cross sections, $\frac{d\sigma}{d\Omega}$, of the spin-averaged Klein-Nishina equations for polarized and unpolarized light. Polarization augments the scatter formula by introducing a preferential scattering angle in the azimuthal direction (χ), which is isotropic for unpolarized light.

$$(\frac{d\sigma}{d\Omega})_{unpol} = \frac{r_e^2}{2}(\frac{E_{\gamma'}}{E_\gamma})^2(\frac{E_{\gamma'}}{E\gamma} + \frac{E_\gamma}{E_{\gamma'}} - sin^2\theta) \tag{3.17}$$

$$(\frac{d\sigma}{d\omega})_{pol} = \frac{r_e^2}{2}(\frac{E_{\gamma'}}{E_\gamma})^2(\frac{E_{\gamma'}}{E\gamma} + \frac{E_\gamma}{E_{\gamma'}} - 2sin^2\theta cos^2\chi) \tag{3.18}$$

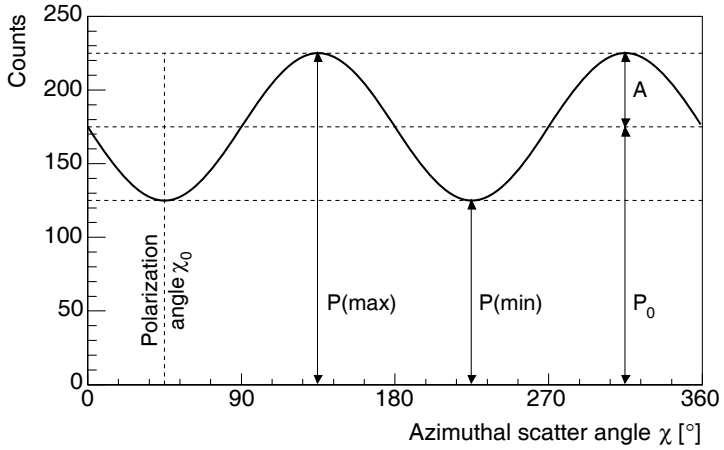

Figure 3.5: Simulated azimuthal scatter angles (χ) for a 100% polarized photon beam. The polarization angle, χ_0, can be determined from the modulation minima, P(min). Also marked are the modulation maximum P(max), zero modulation P_0, and the modulation amplitude A. Image from Zoglauer, 2006 [151].

For incident fluxes which have high degrees of polarization the scattered photons will be azimuthally modulated. This provides a convenient measure of polarization for detectors which can track the scattered photon. Figure 3.5 illustrates how a 100% polarized beam will be azimuthally modulated. The minimum corresponds to the plane of polarization for the incident flux. The amplitude of modulation decreases for higher-energy photons, limiting the usefulness of this technique to photons $\leq$10 MeV [83].

The discussion so far has concerned the effect of a single Compton scatter on an unbound electron. Multiple scatters, or photoabsorption of the scattered photon, within the same detector volume would occur so quickly that detector electronics would register them coincidentally. If the scattered photon is completely photoabsorped, the energy of the initial photon is conserved, adding to the spectral photopeak. Multiple Compton scatters without absorption of the final photon results in incomplete energy deposition, appearing as counts in the region between the Compton edge and photopeak. This effect is illustrated in figure 3.6. Detectors with adequate spatial resolution can distinguish between the individual sites along a Compton track and use this information to determine the scattered photon's energy and direction.

Pair Production

A photon interacts with the strong Coulomb field of a detector nucleus or electron to create an electron-positron pair in pair production. The minimum photon energy for this process is twice the rest mass of the electron, 1.02MeV. Photon energy in excess to the minimum

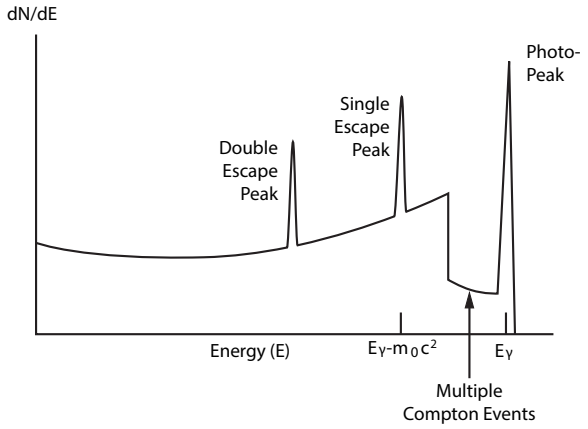

Figure 3.6: Example detector response for an intermediately sized detector wherein a monoenergetic (E_γ) beam of photons photoabsorb, Compton scatter and pair produce with incomplete energy collection. Compton scattering generates a continuum below the Compton edge, the counts between the Compton edge and photopeak result from multiple scatters through the detector without complete energy collection. Incomplete absorption of the pair-produced 511 keV photons from positron annihilation creates two additional escape peaks at 511 keV and 1.02 MeV below the photopeak. Image rendered by Cole Kelleher based on Knoll 2010 [74].

creation energy of the pair goes into particle kinetic energy.

$$E_{e^-} + E_{e^+} = E_\gamma - 2m_0c^2 \tag{3.19}$$

These charged particles will interact with the detector medium, loosing energy primarily through bremmstrahlung. The positron will travel until its kinetic energy is roughly equivalent to the thermal energy of electrons, typically a few millimeters, then interacts with an electron to directly annihilate into two 511 keV photons or via capture to positronium which can produce two 511 keV photons or three photons in a continuum.

The two annihilation photons can photoabsorb, Compton scatter or escape the detector volume without interaction. Since the positron thermalization and annihilation happens quickly, any energy deposition from the annihilation photons can be coincidentally detected with the pair-produced electron. If the detector is of sufficient size to collect the pair-produced electron and the annihilation photons, the total energy of the incoming photon is conserved. As detector volume decreases, the likelihood of photoabsorbing the annihilation photons decreases. In this case, the detector spectrum can have peaks at 511 keV and 1.02 MeV below the photopeak energy. These lines, called the *single* and *double escape peaks*, are illustrated on top of a Compton spectrum in figure 3.6.

3.3 Semiconductor Detectors

Semiconductor detectors offer high energy resolution in compact sizes with relatively fast timing characteristics, and a variable thickness that can be tailored to the application. Semiconductor detectors can be complicated to implement, needing a steady high voltage to be applied across the detector volume and in some cases requiring cooling systems to obtain operating temperatures in the ∼77 K range. They also are more susceptible to radiation damage and more expensive than comparable scintillators. Despite these difficulties, their far superior energy resolution (∼1 keV compared to 10s of keV for scintillators) make them the detector of choice for high resolution applications. The RHESSI satellite and GRIPS payload, discussed in chapters 5 & 6, both employed germanium semiconductor detectors for high-resolution solar spectroscopy.

Band Structure

The electrons in a single atom occupy atomic orbitals of discrete energy. When N identical atoms come together to form a solid, the degenerate atomic orbitals will split into N molecular orbitals such that no two electrons will occupy the same state, as required by the Pauli exclusion principle. Though the individual orbitals are still discrete, their energy spacing is so small that they effectively form a continuum of energy states, called an *energy band*. Lattice periodicities generate regular and predictable band structures of allowable and forbidden electron energies.

Inner-shell electrons are tightly bound to their parent nucleus and have relatively narrow band structures. The outer-shell electrons, which contribute to chemical bonding, have a highly banded structure because their orbitals overlap to a significant extent. These outer-shell electrons contribute to the material conductivity. Figure 3.7 shows a simple illustration of the two bands relevant to conduction, the *valence* and *conduction bands*. The valence band is the top-most filled energy level of a crystal, where all the electrons are in their ground state and still bound to their parent nucleus. There are just enough electrons in the solid insulators and semiconductors to fully occupy the valence band. Lying above the valence band is the conduction band, where excited electrons are free to move through the crystal.

The forbidden energy region between the valence band and conduction bands is termed the *bandgap*. The size of the bandgap determines the material conductivity and classifies the material as either an insulator or semiconductor. Energy input, by thermal or other means, is required to move electrons into the conduction band and generate a current. Insulators have relatively wide bandgaps typically of >5 eV, whereas semiconductor bandgaps are on the order of ∼1 eV. In metals the valence band is not completely full and the conduction band is effectively mixed in energy with the valence band. Currents flow easily in metals because the outermost electrons require very little energy gain to conduct through the material.

An electron can be thermally excited to the conduction band if the bandgap energy (E_g) to the material temperature (T) ratio is sufficiently small. The probability of thermal excitation ($P(T)$) is an exponential function of this ratio and the Boltzmann constant (k)

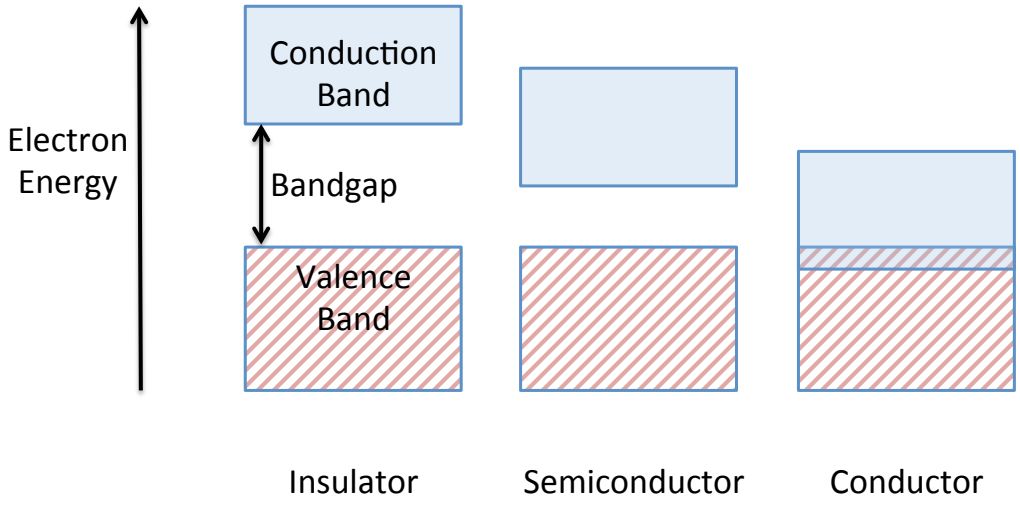

Figure 3.7: Simplified illustration of how the energetically forbidden region (the bandgap) between valence and conduction bands determines material conductivity and classification. Insulators have large bandgaps typically >5 keV, semiconductor bandgaps are typically ∼1 keV and in metals the bands are effectively mixed because the valence band is not full of electrons.

and scaled by a material-specific proportionality constant (C).

$$P(T) = CT^{3/2}e^{-\frac{E_g}{2kT}} \tag{3.20}$$

In materials with very small bandgap energies, such as semiconductors, the thermal excitation of electrons at room temperature can easily obscure the signals generated by photon energy deposition. It is necessary to cool germanium semiconductor detectors to <80 K to reduce thermal currents.

The excitation of an electron into a conduction band leaves a vacancy, termed a *hole*, in the valence band. When a second electron moves into the place of where the excited electron left a hole, there is now a vacancy at the position of the second electron and the hole is said to have moved to that position. Under an applied electric field the electrons flow towards positive potential and as a result the holes move towards negative potential, acting like a particles of positive charge. The observed conductivity of a material results from the combined motions of electrons in the conduction band and holes in the valence band.

Diffusion and Drifts

Thermally excited electrons and holes will randomly diffuse through a material until they recombine. The distance over which the particles move over an elapsed time (t) can be described by a thermal diffusion coefficient (D). In the absence of an applied field, a collection of electrons and holes generated at the same location would diffuse such that the cross-section of their spread would be Gaussian with a standard deviation (σ) given by the diffusion coefficient and elapsed time. The diffusion coefficient itself depends upon the material temperature and the ability of charges to move through the material, called the *mobility* (μ). The mobility encapsulates the ability of a species to move through the material given varying conditions like temperature and applied field.

$$\sigma = \sqrt{2Dt} \tag{3.21}$$

$$D = \mu \frac{kT}{e} \tag{3.22}$$

Under an applied electric field the electrons and holes will flow along field lines such that their total motion is a combination of random thermal velocity and a *drift velocity* due to the field. The drift velocity (v) is proportional to the applied electric field ($\mathcal{E}$) and mobility. Since the mobilities of electrons and holes are comparable to one another in silicon and germanium, the drift velocities of these species are roughly the same. Figure 3.8 shows the germanium drift velocities of electrons (left) and holes (right) as a function of electric field strength (here, E) along the $\langle 111 \rangle$ crystal direction for a range of temperatures (denoted in kelvin). The drift velocity increases more slowly with increasing electric field strength until the mobility plateaus and a *saturation velocity* is reached. Saturation velocities are on the order of $\sim 10^7$ cm/s for charge carriers in germanium, imparting these detectors with fast radiation response times.

$$v_e = \mu_e \mathcal{E} \qquad v_h = \mu_h \mathcal{E} \tag{3.23}$$

When an electric field is applied across a detector, mobilized charge carriers will flow towards the electrodes which generate the potential difference. The current induced by the charged particles as they approach these electrodes, termed *collecting electrodes*, provide the signal which is processed by the readout electronics. In the absence of thermal energy the charge carriers would follow the field lines to the electrodes and arrive in a time predicted by their drift velocities and the distance they traveled. Judicious choice of the electrode geometry, adequate timing resolution on the signal readout electronics and knowledge of the electric field structure internal to the detector can be used to reconstruct the position of where an electron-hole pair originated within the semiconductor.

Thermal energy adds a random velocity component to the drift velocity, acting to generate a spread in arrival locations and times at the collecting electrodes. Thermal diffusion creates a lower limit in the spatial and timing resolution achievable by position sensing detectors. The spatial spread generated by diffusion (σ_D) over a drift distance (x) can be estimated by

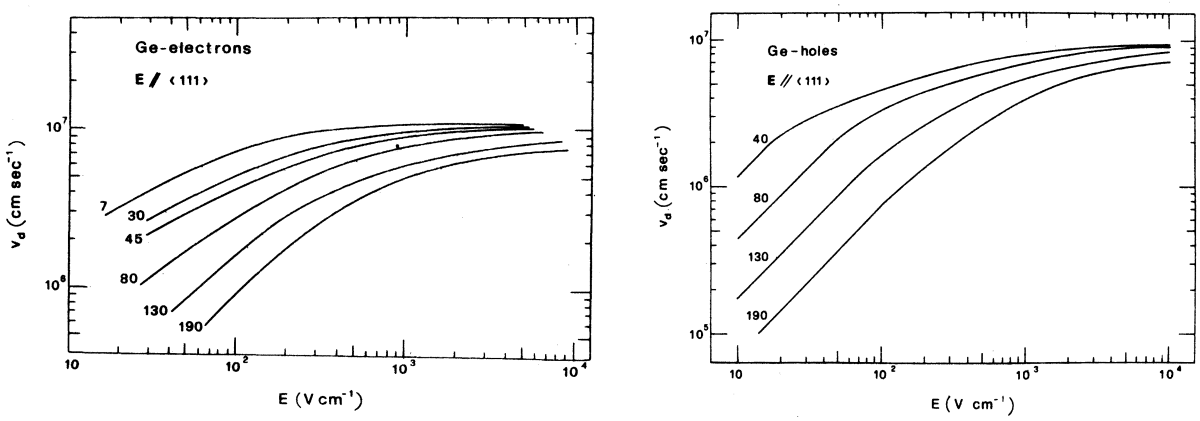

Figure 3.8: Drift velocities for electrons (*left*) and holes (*right*) as a function of applied field strength in germanium, along the ⟨111⟩ crystal direction. The velocities at a range of temperatures are plotted and marked in Kelvin. A saturation velocity occurs when the curves plateau and the velocity no longer increases with the field. Plots from Ottaviani et al., 1975 [99].

combining equations 3.21, 3.22 and 3.23. The spread in arrival times (σ_t) is then given by the spatial spread divided by the drift velocity. Minimizing the detector temperature will narrow these spreads.

$$\sigma_D = \sqrt{\frac{2kTx}{e\mathcal{E}}} \tag{3.24}$$

$$\sigma_t = \frac{\sigma_D}{v} \tag{3.25}$$

Impurities, Dopants and the Reverse Biased P/N Junction

A completely pure semiconductor, called an *intrinsic semiconductor*, has equal numbers of electrons and holes. In practice all detectors will contain impurities which create an electron/hole imbalance. The presence of additional electrons (called *donors* in *n-type* material) or holes (called *acceptors* in *p-type* material) create additional energy bands within the bandgap. These extra energy levels increase the material conductivity, exacerbating the effects of thermally generated currents. Under an applied electric field, the excess charge carriers will cause a continuous *leakage current* to flow. Any pulses generated by photon energy deposition will be lost to the comparatively large quiescent leakage current.

The extra energy bands introduced by impurities can act as trapping or recombination centers. *Trapping* and *recombination* of charge carriers lead to energy loss in spectral measurements by reducing the size of the detected pulse. When a particle is trapped, it is immobilized for extended periods of time, usually in an energy band in the middle of the bandgap. Though the trapped carrier might be ultimately released, it will be removed from

the primary charge pulse and its energy measurement lost. Recombination occurs when a migrating hole and electron come together and annihilate, indefinitely removing them from the pulse. Since trapping and recombination reduce the pulse size, these processes will preferentially move counts from the photopeak to the lower-energy side of the Gaussian. In detectors where trapping and recombination are significant the photopeak will develop a low-energy tail. Trapping and recombination can reduce carrier lifetimes by of $3 - 4$ orders of magnitude.

Intentionally introducing impurities to tailor material properties is called *doping* and the impurities are referred to as *dopants*. Heavily doped material, denoted as n^+- or p^+-type, have very high conductivity of their majority charge carrier and very low conductivity of their minority carriers. Thin layers of this heavily doped material are commonly used as electrical contacts on semiconductors, called *blocking contacts*. The very low minority carrier concentration makes it difficult to inject these carriers into the detector volume, effectively acting to block the minority charges from entering the detector volume.

No currently known detector fabrication process can completely remove impurities. Efforts to *compensate* semiconductor materials by attempting to balance donor and acceptor impurities is not perfect. Materials with even trace levels of impurities will act as n- or p-type with large leakage currents under applied fields.

A reverse-biased P/N junction is used to overcome quiescent leakage currents in semiconductor detectors. The junction is formed by bulk p-type and n-type material joined in a single crystal. Since it is difficult to bring together two different crystals, the junction is typically formed in a single contiguous crystal by causing a doping change from one side of the junction to the other.

Natural diffusion of p-side holes and n-side electrons across a P/N junction results in recombination at the interface. As more charges migrate across the junction a net negative charge accumulates on the p-side, and a net positive charge accumulates on the n-side. Inside this region the majority carriers have migrated away and the area is termed the *depletion region*. As the charge imbalance grows an electric field is established which acts to impede charge migration across the junction. Eventually charges stop flowing across the junction. This naturally occurring depletion region of an unbiased P/N junction is relatively narrow in comparison to the crystal volume.

The depletion region of a reverse-biased semiconductor has two key features that make it an ideal radiation detector: low leakage current and an internal electric field. Applying a bias voltage along the naturally established electric-field direction (*reverse biasing*) will cause the depletion region to grow. A negative contact on the p-side will attract the majority carrier holes, and a positive contact on the n-side similarly attracts that side's majority carrier electrons, increasing the depletion region on both sides of the junction. This depletion region is the detector's *active volume*.

The charge separation caused by a reverse biased P/N junction acts like a capacitor. As the electric field increases, the depletion depth (d) continues to grow and the capacitance (C) decreases. Solving for the field inside the depletion region gives a relationship between the depletion depth, externally applied voltage (V) and dopant concentration (N) that can

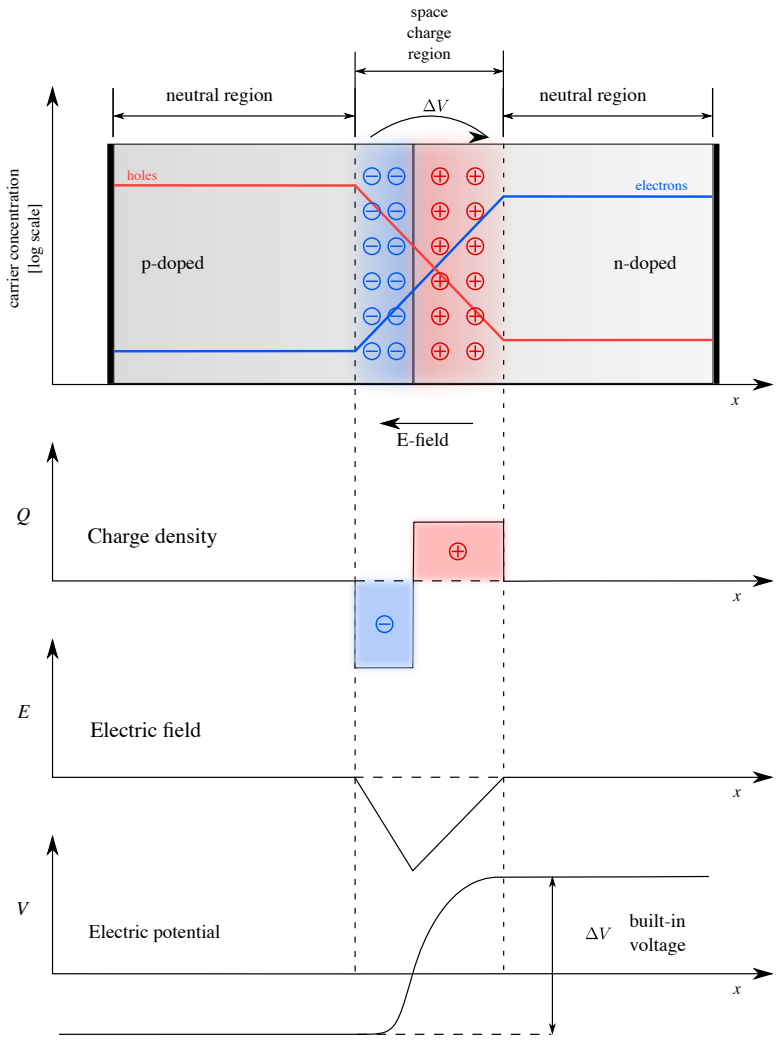

Figure 3.9: Diagram of a P/N junction with zero applied bias. The grey regions are charge neutral and the blue/red shaded regions are negatively/positively charged. Contact between the p- and n-sides naturally establishes a depletion region and creates an electric field. Reverse biasing the junction by placing a negative contact on the p-side and a positive contact on the n-side will cause the depletion region to grow. The magnitude of charge depletion on both sides is equal, so the depletion region extends farther into the less doped side (n-side in this example). Charge density (Q), electric field (E) and electric potential (V) across the P/N junction are plotted below the diagram. Image created by Adundovi and updated by Wereldburger758 on Wikimedia commons [1]

be used to relate the detector capacitance to the voltage. Below is the capacitance per unit area for a detector (with permittivity ϵ) in a planar configuration whose electrodes effectively form a parallel-plate capacitor [74].

$$C = \frac{\epsilon}{d} \simeq \sqrt{\frac{e\epsilon N}{2V}} \tag{3.26}$$

When the crystal is fully depleted, continuing to increase the electric field has no effect on the capacitance. Monitoring the capacitance as the field is increased provides a good measure of when the detector reaches full depletion, at a point termed the *depletion voltage*. Capacitance-voltage measurements are the standard method for determining the depletion voltage for semiconductor detectors. Since detector capacitance is a source of electronic noise, it is advantageous to minimize the capacitance by operating the detector fully depleted. A fully depleted detector also provides the most active volume for a detector of a given size.

Though the detector is fully operable with maximal active volume at the depletion voltage, in practice it is desirable to operate the detector at even higher voltages, such that the crystal is said to be *over-depleted*. Over-depleting increases the field uniformity inside the crystal and as the field strength increases, so too do the carrier drift velocities. Faster drift velocities reduce the likelihood of trapping/recombination and improve the spatial/temporal pulse spreading caused by thermal diffusion. The excellent $\sim$1 keV energy resolutions reported by semiconductor detectors are achieved by using over-depleted crystals that have reached their saturation velocity.

A reverse-biased detector cannot endure arbitrarily high electric fields. Eventually the fields become strong enough to accelerate electrons to speeds which can ionize atoms, generating large currents that can cause permanent damage to the detector. Typically the currents across the detector surface and bulk active volume are monitored as the field is turned up on a detector for the first time. The current should be continually monitored for signs of breakdown as the crystal passes the depletion voltage, is over-depleted and reaches its final operating voltage. Transient current spikes (>1 μA) can indicate that the detector is on the edge of the breakdown voltage and should be immediately turned down to avoid permanent damage.

The depletion depth reached by a given voltage is a function of the doped semiconductor's resistivity (ρ) [74]. Since the resistivity depends on the purity of the material before the doping process, it is highly desirable to start with the purest material possible before doping.

$$d \simeq \sqrt{2\epsilon V \mu \rho} \tag{3.27}$$

A semiconductor's maximum depletion depth is determined by a combination of the material purity and crystal breakdown voltage. The ability to generate larger depletion regions motivates the choice of germanium for gamma-ray detection over silicon. Gamma-ray detectors need to be on the scale of a few centimeters to ensure adequate absorption, and high purity germanium detectors can achieve these depletion depths. Silicon semiconductors cannot reach similar levels of purity with current manufacturing techniques, and consequently their maximum depletion depths are on the order of a few millimeters.

Pulse Generation, Signal Formation and Arrival Times

When a photon interacts within the active volume of a semiconductor, the deposited energy is turned into electron-hole pairs. The quantity of pairs generated is proportional to the original photon energy, typically ∼3 eV per pair for germanium. The electrons and holes then drift towards opposite sides of the detector, accelerated by the internal electric field. The charges follow field lines to the collecting electrodes, but thermal diffusion imparts a spread in the arrival positions and times. If the photon deposits energy outside the active volume, or where there isn't an electric field, then the electron-hole pairs migrate due to diffusion alone and are not collected by the electrodes en masse.

As the charges drift through the detector volume they induce currents on the electrodes. These induced currents are the signals which are measured and analyzed by the detector readout electronics. When the charge carriers themselves arrive at the electrodes the signal ends. The Shockey-Ramo theorem for induced charge describes how charge motion is converted to induced charge for a variety of different electrode shapes and configurations via the concept of a *weighting field*. A full description of the process can be found in Knoll 2010 (Appendix D) [74].

The choice of electrode geometry has a significant impact on the shape and timing of the induced signals. Figure 3.10 shows simulated signals induced on large and small pixels, illustrating how pixel choice changes the signal shape and timing. The curves represent three different charge interaction depths (.2L, .5L, .8L) within an active region of depth L [11]. For a simple planar electrode detector with large pixels, signals are induced on the electrodes as soon as the charges start moving. The signal is proportional to the distance between the drifting electrons and holes, rising linearly with time. When one charge is collected the slope of the signal rise drops and continues to linearly rise with a new slope until the other charge is collected. In detectors where electrons and holes have different drift velocities the slopes of the two collection curves will be different for each species, and it can be determined which species arrived at the electrodes first, as shown in the example. For materials like germanium, where holes and electrons drift with approximately the same velocity, the slope magnitude drops by 50% when one species reaches the electrode and it is impossible to determine if the electron or hole arrive first.

Signal formation is vastly different when the electrodes are partitioned into pixels or strips (figure 3.10, bottom), with the so-called *small electrode effect*. In this configuration the signal on any one electrode is small until the charge is very close to the collecting electrode, at which point it rapidly rises until the charge is collected. This rapid rise marks the arrival of the charge at the electrode. The other electrodes effectively shield the collecting electrode from receiving a large signal until the charges are close.

The shape of the induced signal and its timing can be used to infer spatial information about the charges. The time it takes for charges to be collected by an electrode is set by their drift velocity through the crystal. Since the drift velocity is well known for a given detector voltage setting, it is possible to use the arrival time to infer the distance over which the charges traveled. In a material where electrons and holes drift at the same speed, an

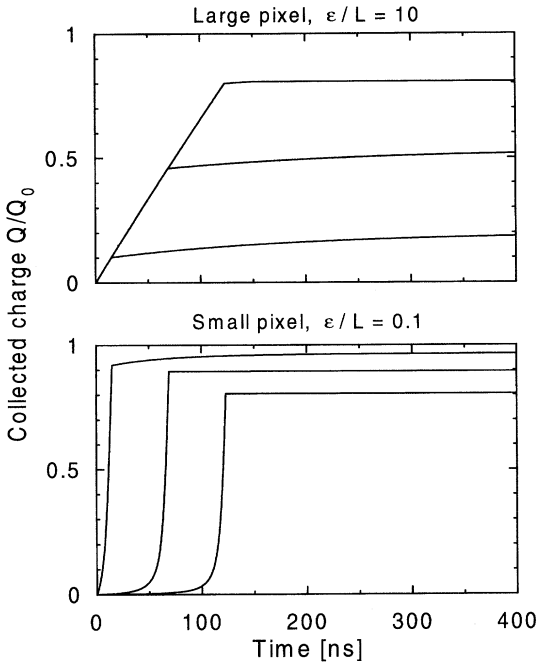

Figure 3.10: Simulated collected charge as a function of time for large and small pixels at 3 different interaction depths (.2L, .5L and .8L) within a detector active volume of depth L. The choice of electrode size determines the signal induced and charge collected. Figure from Barrett et al., 1995 [11]

electron-hole pair generated at the center of an active region would arrive simultaneously at the opposite detector electrodes. The difference in their arrival times can be used to determine the depth at which they were created.

The large-pixel configuration can be used to determine the arrival time information necessary for the depth measurement, but there is ambiguity for materials where the electrons and holes drift at similar speeds. In these detectors it is impossible to determine if the electron or hole was collected first, and only the absolute distance from the center of the active region can be determined. In materials with significantly different drift velocities (like CZT) it is possible to determine which species arrived first and there is not ambiguity in determining which side of the crystal an interaction occurred.

Small-pixel detectors resolve the ambiguity issue for materials where the drift velocities are the same because these pixels collect signals from electrons and holes independently. The independent measurements provide a means to determine the actual interaction depth unambiguously. Another advantage of small-pixel detectors is that their sharp signal rise conveniently marks the charge arrival time with precision. Determining arrival times for large-pixel detectors is less precise because of their linear collection curves.

Leakage Current

A detector's leakage current is the baseline parameter which determines its minimum detectable signal; all electron-hole pairs must be measured on top of this signal. Although a reverse-biased semiconductor's resistivity is very high, the 100s to 1000s of volt potential drop across the crystal will cause leakage current to flow. In practice, given good detector configuration and operating conditions, the steady-state leakage current can be held below the microamp range and this is sufficient for radiation detection. The leakage current can be divided into two major components: bulk and surface currents.

The issue of large leakage currents due to crystal impurities in semiconductor materials is primarily solved by using a reverse-biased P/N junction, as described above. In this configuration the reverse-biased field resists the flow of majority carriers across the junction. The field direction is, however, favorable for the flow of minority charge carriers through the active region. These minority carriers generate a small quiescent leakage current across the crystal. The magnitude of this current is often negligible in comparison to thermally generated leakage current.

Bulk leakage currents through the detector active volume are mostly due to thermal excitations of valence electrons. The leakage currentv (I_L) generated by a reverse biased detector of bandgap E_b is proportional to temperature [129].

$$I_L \propto T^2 e^{-\frac{E_b}{2kT}}$$

Silicon and cadmium-telluride detectors, with bandgap energies of 1.12 eV and 1.52 eV, can operate at relatively warm temperatures of ∼300 K. Germanium's small bandgap of .67 eV necessitates cooling to much colder temperatures of <80 K.

Electrodes can be a significant source of leakage current. Simple ohmic contacts allow charges of either sign to pass freely into and out of the detector. These contacts will inject a carrier for each carrier sunk at the opposite electrode, thus maintaining charge balance within the detector and generating a large leakage current. High purity detectors of typical size (1 cm^2) and high resistivity (50,000 Ωcm) can see leakage currents as high as .1 A as a result. Blocking electrodes made of heavily doped n^+- or p^+-type material are frequently used to generate very high-resistivity for minority charge carriers to pass between the electrodes and detector volume. In this case, when a reverse bias is applied across the detector the blocking contacts don't allow the flow of carriers from the opposite side back through the active volume.

Surface leakage currents can vary widely in magnitude. Large electric potential gradients at the edges of the crystal can generate currents across the detector surface. Surface contamination from debris, vacuum pump oil, fingerprints or other condensed vapors can provide low-resistance paths for currents to flow. Techniques for reducing the field edge effects and assembly in clean rooms can greatly reduce surface leakage from these factors. Separating the surface from the detector active volume electrically, or physically, can minimize the impact of surface leakage current on signal collection. This can be done physically by etching the crystal or electrically with a guard ring around the electrodes.

3.4 Germanium Semiconductor Detectors

Gamma-ray detection requires large detector active volumes to ensure adequate energy deposition. Even within solid materials of high density, detector volumes of a few centimeters are required for efficient gamma-ray detection. The purity of the semiconductor material before doping determines how large of an active volume can be created before crystal breakdown voltages occur, as discussed above. With currently available manufacturing techniques, germanium provides the largest active volumes and is consequently the material of choice for gamma-ray detection. Germanium's relatively small bandgap energy ($B_{g,Ge,300K} = .67$ eV) adds complication to its operation in comparison to silicon ($B_{g,Si,300K} = 1.1$ eV) and other semiconductors. Thermally generated leakage currents require germanium to be cooled to <80 K whereas the other semiconductors can operate close to, or at, room temperature.

Fabrication

Germanium crystals approach intrinsic carrier quality through manufacturing techniques that either compensate for impurities through doping or refinement that reduces the absolute impurity concentration. In the 1960s a process emerged wherein insterstitial litium ions were used to compensate for impurity concentrations. This process, called lithium ion drifting, could compensate over a maximum depth of 2 cm. Germanium detectors fabricated by this method were given the designation of Ge(Li). These detetors fell out of favor when advanced refinement techniques emerged in the 1980s. These new ultra-high-purity germaium detectors, HPGe, achieved impurity concentrations of $\sim 10^{10}$ atoms/cm^3 and depletion depths of multiple centimeters. Advanced refinement techniques do not exist for silicon, and at this time modern silicon detectors rely on the lithium ion drift method to compensate for impurity concentrations.

Though both lithium-drifted and HPGe detectors have properties close to intrinsic germanium and comparable penetration depths, the HPGe detectors have a key advantage which sets them apart from an operational standpoint. The HPGe detectors can withstand theramal cycling whereas the Ge(Li) detectors suffer degredation and must be continually cooled. For this reason HPGe detectors are more widely used today.

Fabrication of germanium detectors starts with a relatively pure crystal, whose bulk volume is either slightly n- or p-type. Heavily doped material is deposited onto the surface by a variety of methods to form electrode contacts [141]. The P/N junction is formed at the electrode whose type is opposite to the bulk detector material. This is the rectifying contact and the depletion region will grow from this electrode. If the depletion voltage is reached then the entire volume is active. Typically a voltage is applied to the rectifying contact while the other is held constant; for an n-type bulk material the p-type electrode receives a negative voltage, and for a p-type bulk the n-type electrode receives a positive voltage. A dead-layer will exist in the region close to the electrode and judicious choice of the doped electrode material will affect the dead-layer depth.

Detector Geometry

The most basic germanium detector is a a bulk wafer with two simple planar electrodes. This type of detector provides no spatial resolution of where the photon interaction occurred within detector. The total volume of a wafer is usually limited by the diameter of the crystal to $\sim$10s of cm^3.

Coaxial detectors are made by removing the core of a cylindrical crystal and placing an electrode in the bore. The critical depletion depth runs from the exterior surface electrode to the inner bore electrode and the crystal can be made long in the axial direction, greatly increasing the detector active volume over planar geometries. The coaxial detectors can be open on both ends or have a single closed-end. An example of a closed-end type coaxial detector is shown in figure 5.4 for the RHESSI satellite. Detectors in coaxial configurations have reached active volumes of $\sim$100s of cm^3. Since the inner electrode can be made small, these large-volume detectors can made with lower capacitance than would be possible with planar detectors.

Detector spatial resolution can be achieved by using multiple electrodes to segment the detector into different collection regions and provide 1 – 3D position sensitivity. Electrode segmentation on both planar and coaxial geometries is possible and has been demonstrated for a variety of electrode geometries, materials and deposition processes. A good review of the breadth of germanium detector and electrode configurations can be found in Vetter, 2007 [141].

Planar germanium cross-strip detectors can be used to determine the position of photon energy deposition in 3 dimensions. These types of detectors, termed *3D-GeDs* or *Double Sided Strip Detectors* (DSSDs), are used to provide high spatial resolution to the the GRIPS spectrometer which is described in chapters 6, 7 and 8. Figure 3.11 shows a sketch of the strip geometry to the left, and example signals for the collecting electrodes and their neighbors to the right. This image demonstrates that one key advantage of 3D spatially resolved detectors is the ability to detect the independent locations in a Compton scatter track, even within the same crystal.

The 3D-GeD uses orthogonal strip electrodes to collect holes in one dimension and electrons in the other. The intersection of the two electrode strips which collect charges locate the photon interaction site in the 2D plane parallel to the electrodes. Figure 3.12 shows a simulation involving one site of photon energy deposition, and combines the signals collected on opposite strip electrodes (y_3 and x_3) into a single panel. The small electrode effect effectively shields the collection electrode from an induced signal until the charge pulse is close, resulting in a sharp rise (or fall) preceding full collection and allows the electron and hole signals to be collected independently. The signal edge is used to determine the charge arrival time for each species and their relative timing provides a measure of the depth. Each successive panel illustrates how the signal timing is effected by photon interactions occurring at different depths.

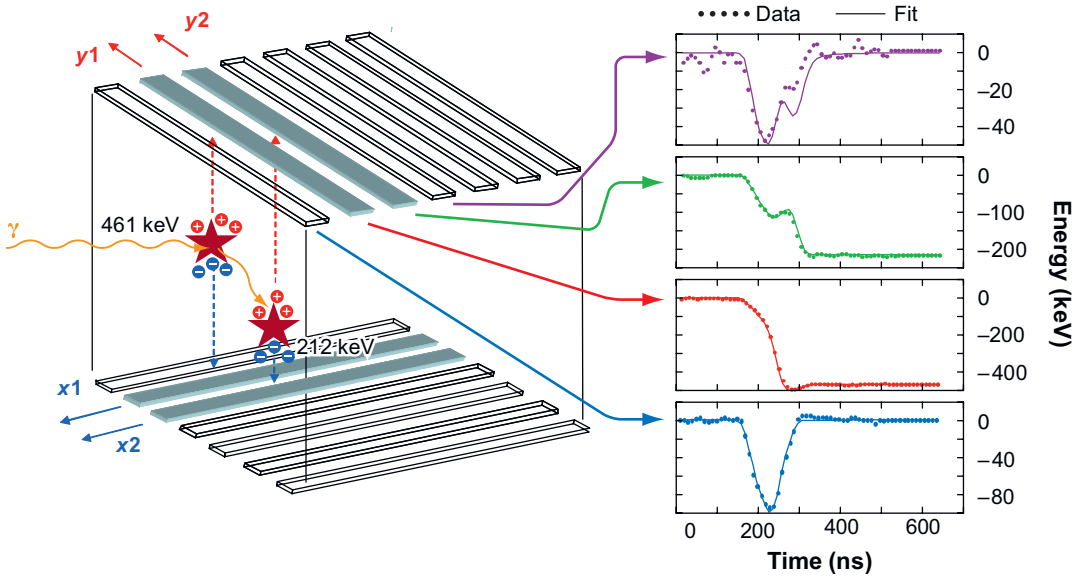

Figure 3.11: *Left:* Sketch of a planar germanium cross-strip detector. A single gamma-ray enters the detector and Compton scatters through two sites within the crystal. The strip pairs (x1, y1) and (x2, y2) collect the charge pulses generated at the first and second sites. The intersection of the collecting strips locates the energy deposition sites in the 2D plane. *Right:* Real event data from the four electrodes on the detector sketched to the left. The data is fit to determine the charge arrival times. When combined with timing information from the opposite side, the depth of the interaction is determined. Figure from Vetter, 2007 [141]

3.5 Outlook

The availability of high spatial resolution gamma-ray detectors capable of determining the photon interaction site in 3 dimensions was demonstrated as recently as 2000 [2] and first flown for astrophysics in 2005 on the NCT balloon as a Compton imager [36]. NCT showed that a 3D-GeD with a 2 mm electrode spacing could provide high-resolution spectroscopy ($\sim$1 keV) and independently track Compton scatters to provide imaging (chapter 4). The advent of 3D-GeD detectors with a finer electrode spacing of .5 mm was the key enabling technology for the GRIPS telescope's indirect imaging system (chapter 9). Indirect imagers can achieve high imaging resolution, but at the expense of sensitivity because absorbing grids are used to modulate signals. The 3D-GeD spatial sensitivity allows GRIPS to remove one of the modulation grids, improving sensitivity over similar bi-grid designs. The .5 mm detector resolution made it possible to design a 12.5" resolution system on an 8 m boom, an instrument size accessible to balloon platforms.

Germanium semiconductor detectors provide the highest spectral resolution, greatest

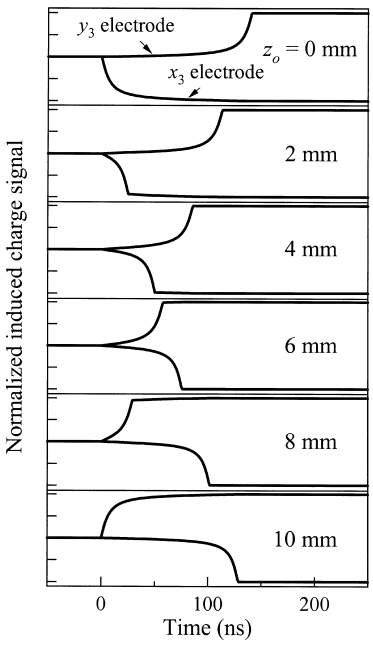

Figure 3.12: Simulated signals on two electrodes (y_3 and x_3) in a germanium strip detector. The different panels shows how the photon interaction depth (0, 2, 4, 6, 8 and 10 mm) changes the signal timing on each electrode. The relative charge arrival times can be used to determine the interaction depth. Figure from Amman et al., 2000 [2]

photon stopping power and fastest signal response available over the 1 keV – 10 MeV spectrum, making them the detector of choice for high-resolution spectroscopy in this energy range. These detectors can be implemented with or without spatial resolution in a range of detector volumes and shapes. Recent advances in high spatial resolution detectors have enabled new and advanced imaging for astrophysics in this energy range.

Chapter 4

HXR/Gamma-Ray Imaging

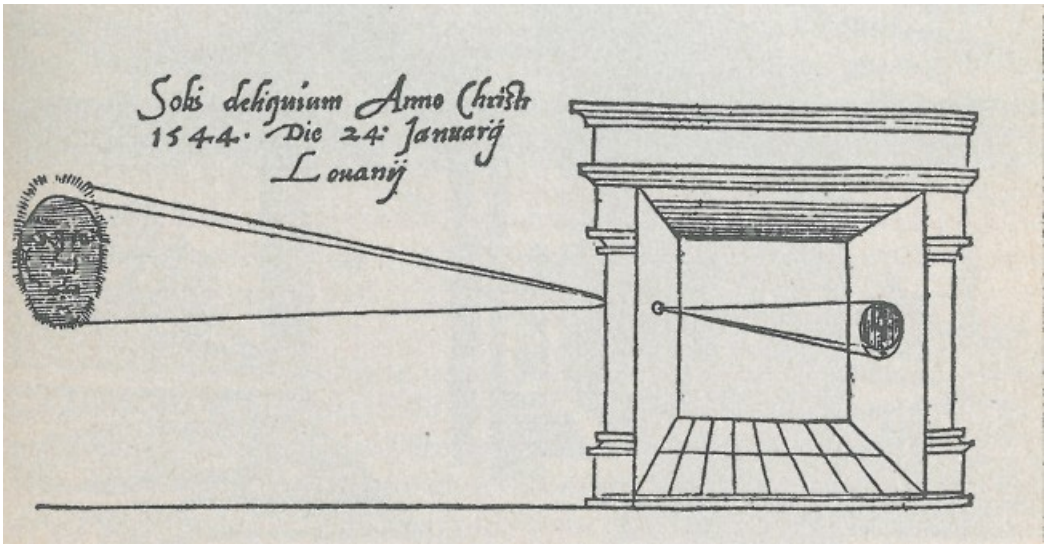

Figure 4.1: The oldest known drawing of the camera obscura [59], published in Gemma Frisius' 1545 book De Radio Astronomica et Geometrica, describing how this single aperture imager was used to observe the solar eclipse of January 24, 1544.

Indirect imagers made from masks, grids & collimators (section 4.1) and Compton techniques (section 4.2) provided the first images of the Sun and greater universe in the 1 keV to ∼10 MeV energy range. Indirect imaging has been able to achieve excellent ∼2" angular resolution but faces a tradoff in sensitivity from the use of absorbing layers in the telescope line-of-sight and limits the field of view (FOV) to photons that pass through those layers. Compton imaging can in principal have a full-sky FOV without any photon attenuation, but it is fundamentally limited to a ∼1° angular resolution by the momentum distribution of bound electrons in the detector material. Despite these tradeoffs in sensitivity, angular resolution and FOV, both methods have been extremely successful and are the standard imaging techniques in the full 1 keV to ∼10 MeV energy band today.

Though indirect imaging and Compton imaging have been successful, both of these methods are signal-to-noise limited because their photon collection areas are the same as the photon detection area. Since background scales with detector area, increases in sensitivity come at the cost of increased noise. Decoupling the photon collection area from the detector area is possible by using focusing optics to collect photons and divert them onto a smaller detector. In the 1 keV to ∼10 MeV energy range, however, the index of refraction is slightly less than unity, resulting in long focal lengths and shallow reflection angles that make direct photon imaging technically challenging. Techniques using crystal diffraction (Laue lenses, section 4.3), grazing incidence optics coated in multilayers (section 4.4) and diffraction zone plates (Fresnel lenses, section 4.5) are being explored as direct focusing options. Manufacturing advancements have enabled the production of multilayered mirrors that have made grazing incidence optics the state-of-the art for astrophysics observations <100 keV, but direct photon imaging has yet to be realized through the whole band. These systems, though attractive because of signal-to-noise and fine angular resolution, have the drawbacks of heavily constricted FOVs and very narrow energy bands in comparison to the indirect and Compton methods.

4.1 Indirect Imaging

Absorbing layers (masks, grids and collimators) can be used to selectively transmit or block incoming photons, modulating the incoming flux. The resulting spatial (figure 4.3 & figure 4.4, top) or temporal (fig 4.4, bottom) variation encodes information about the source's angular distribution. Deconvolution of the modulated flux, using properties of the well-known imaging target metrology, reconstructs the original photon distribution. The resulting images are probability distributions generated by the deconvolution technique. Design choice in the coded-aperture style imaging implementation is a tradeoff between the relative importance of angular resolution, sensitivity and FOV. The RHESSI (chapter 5) and GRIPS imaging systems (chapter 9) employ strategies based upon these methods.

Coded-Aperture Masks

The camera obscura, depicted in figure 4.1 for observations of a 1544 solar eclipse, uses a small *pinhole* aperture in an opaque mask to project an (inverted) image onto a distant screen. The pinhole size must be small in comparison to the distant object and the angular resolution is given by a ratio of the pinhole diameter to its separation from the imaging screen. In the early 19th century this type of single aperture imager was developed into the first photographic camera when photosensitive materials were used as the imaging screen. Since the only requirements of this imager are that (1) the imaging aperture is small in comparison to the source and (2) the aperture diameter is large enough to avoid diffraction, this device can be used to generate images of high-energy light distributions. In a HXR/gamma-ray implementation the detector spatial sensitivity combined with the aperture-to-detector

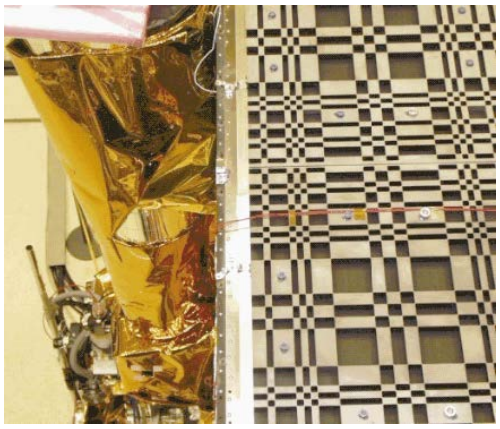

Figure 4.2: The IBIS telescope's coded mask on the INTEGRAL spacecraft. The coded apertures are distributed in a uniformly redundant array (URA) pattern. Image from Ubertini et al., 2003 [137].

distance would set the angular resolution. The sensitivity of such a device is severely limited, as the only radiation reaching a detector would be the flux over the single pinhole, making it unsuitable for high-energy solar and astrophysical observations, which are generally low-flux.

The pinhole flux problem can be resolved by using an imaging mask with a set of multiple pinholes that generate many projected images, increasing the effective area in proportion to the total area of the apertures [42]. These *coded-aperture* masks, shown in figure 4.2 for the IBIS telescope, can significantly increase the effective area, reaching up to ∼50% of the total mask area. Deconvolution of the overlapping image patterns can reconstruct the original source image for simple photon distributions, like those of astrophysical point-sources.

The basic geometry of a coded-aperture imager is shown in figure 4.3. In this design the detector is smaller than the imaging mask, only allowing a partial shadow of the image mask to be recorded. The shadow pattern cast on the detector changes based on the source-to-mask direction. For this implementation the detector must have a spatial resolution that at least matches the aperture size. This requirement usually sets the angular resolution for these imagers based upon available detector technologies. The FOV is set by the ratio of the detector and/or mask size to their separation distance. The FOV can approach ∼1 sr making this option good for monitoring multiple sources, and in survey or burst monitoring applications. Two disadvantages of this method are (1) a limited dynamic range resulting in difficulty resolving multiple sources of different brightnesses inside a single FOV and (2) a degraded angular response for extended sources.

$$resolution = \frac{1}{2}\frac{pitch}{separation}$$

There are numerous ways to choose the distribution of apertures in the coded mask. One class of patterns, called the uniformly redundant array (URA), has the convenient property of a point response function that has flat sidelobes and an angular response that is defined by the geometry of the individual apertures [49], shown in figure 4.2. This type of array has been

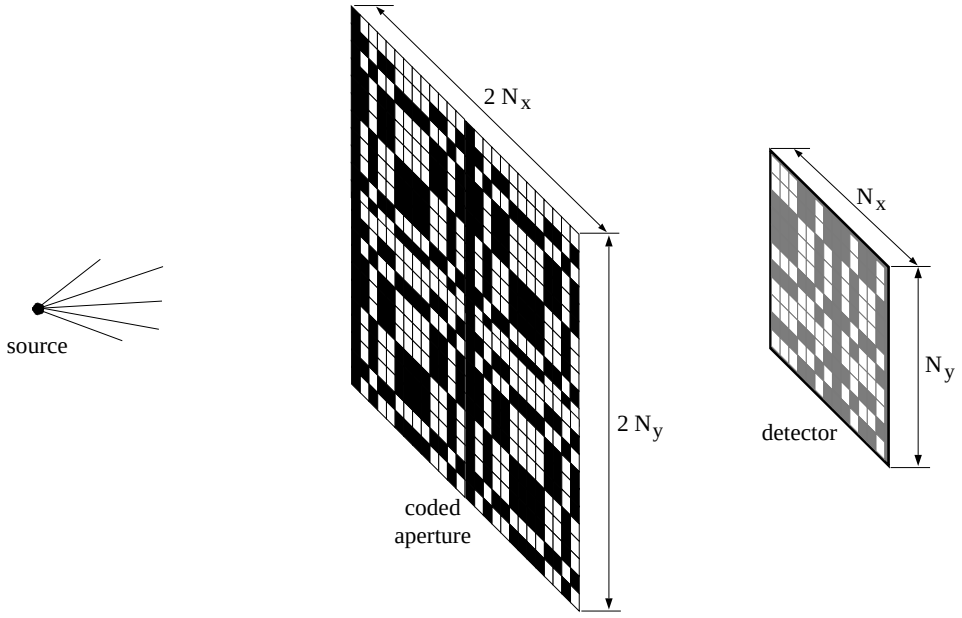

Figure 4.3: Schematic depicting coded-aperture imaging. In this implementation the detector (right) is smaller than the imaging mask (left), only allowing a partial shadow of the image mask to be recorded. The shadow pattern cast on the detector changes based on the source-to-mask direction. The original source distribution can be reconstructed by deconvolution of the shadow image cast onto the detectors. The mask shown is a URA, image from Busboom et al., 1998 [25].

widely employed in astronomical observatories, including BAT on the SWIFT mission [13] and IBIS on INTEGRAL [137]. URAs, or more generally binary arrays with flat sidelobes, are widely regarded as the best choice for coded aperture imaging.

Collimators

A set of co-aligned grids are used to modulate incoming flux in modulating collimator designs [96]. Collimators have loose detector spatial resolution requirements, or remove this requirement completely, making them favorable over coded-aperture designs in applications where detector spatial resolution is coarse, or high-precision detectors are difficult to implement, and high angular resolution is required.

The grid-pairs in temporally modulated collimators are co-aligned with identical grid pitches and orientations. As the source-to-grid orientation changes the total flux transmitted through the grids fluctuates. Figure 4.4 (bottom) illustrates how flux changes for two source locations. In this configuration the detector requires no spatial resolution and the angular resolution is given by the ratio of one-half the grid pitch to separation.

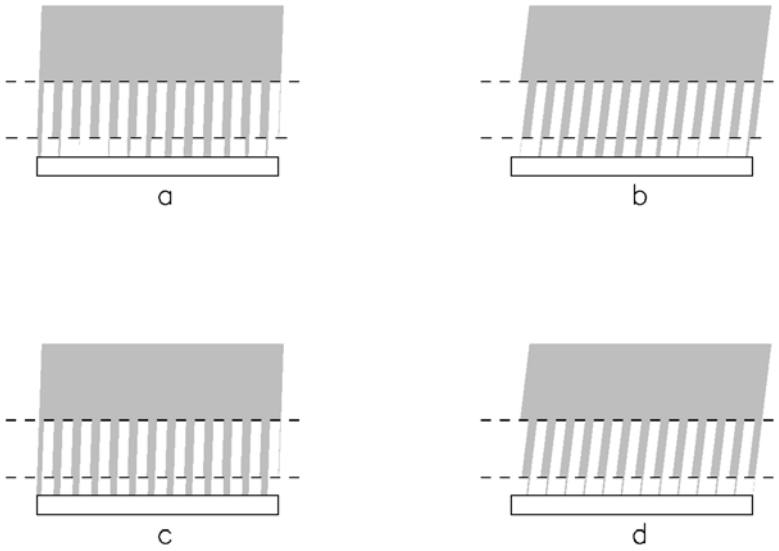

Figure 4.4: Photon flux modulation patterns generated by bi-grid collimators. *Top:* Spatially modulated flux generated by a grid-pair with slightly different grid pitches. The moire pattern phase shifts (the maximum transmission point moves) as the source location varies. *Bottom:* Temporally modulated flux generated by a grid-pair with identical pitch. As the source location varies the total transmitted flux varies. The angular resolution for both collimator types is set by the ratio of one-half the grid pitch to the separation. Image from Hurford 2010, [65].

$$resolution = \frac{1}{2}\frac{pitch}{separation}$$

A requirement of time-modulated collimator imaging is that there is constant motion between the source and grid-pair, in the plane perpendicular to the line of sight. This is frequently accomplished by rotating the grids relative to the source to create quasi-periodic transmission profiles (figure 4.6, top), in an implementation termed *rotating modulating collimators* (RMCs) [113]. This can be implemented through fixed grid-to-detector orientations aboard rotating spacecraft and a rotating grid relative to fixed detectors on-board non-rotating platforms such as balloons.

Spatially modulated bi-grid designs have two sets of grids whose pitches and/or orientations are slightly offset from one another. The image formed by the source through the two grids is a moiré interference pattern whose phase depends on the source-to-grid orientation. An illustration of how the moiré pattern phase shifts as a function of source location is shown in figure 4.4, top. The moiré pattern generated by a single grid-pair on the STIX instrument's X-ray imager [81] is shown in figure 4.5. The moiré pattern is shown in red/yellow to denote shadowing/transmission. As the source location changes, the phase of this pattern moves across the detectors (shown in blue). The detectors need to be big enough and have

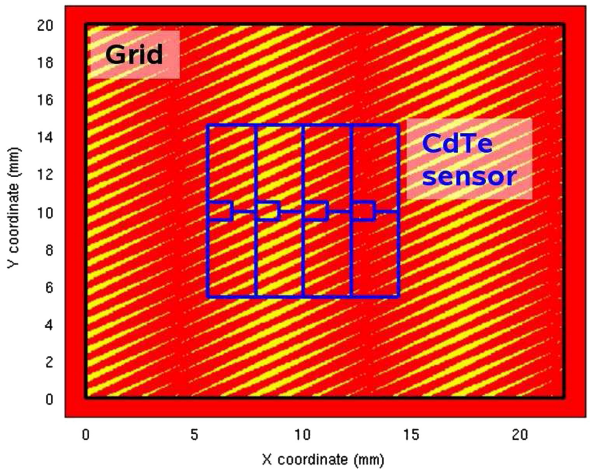

Figure 4.5: Moiré pattern generated by a single grid on the STIX instrument's X-ray imager, which is planned to fly aboard ESA's Solar Orbiter in 2017. STIX's collimator grids have slightly different pitches and orientations relative to one another. For the grid pair shown, the pitches are 666 μm/690 μm with orientations of 60°/64°. The black box denotes the size of the grids, and their resulting interference pattern is shown in red/yellow stripes denoting shadowing/transmission. The blue boxes show the CdTe sensor pixel sizes. Image from Krucker et al., 2013 [81].

enough spatial resolution to detect the overall pattern's phase change but do not need to resolve the fine structure of the pattern itself. The detector spatial resolution requirement is set by the imaging grid diameter, but the image's spatial resolution is set by the ratio of one-half the grid pitch to their separation.

The maximum imaging energy for either implementation is set by the transmission properties of the grid, with the requirement that the grid is opaque for photon modulation to occur. For high-energy light, above ∼1 MeV, this can lead to grids with thicknesses on the order of ∼2.5 cm with high-Z materials such as tungsten. These thicknesses lead to shadowing within the grid itself, limiting the FOV to the ratio of one-half the grid pitch to thickness. The lower energy limit is set by the grid pitch, since the front grid can act as a diffraction grating for sufficiently long wavelengths. These performance trade-offs between energy range, angular resolution and FOV can be relevant in design. For RHESSI, the resolution above ∼1 MeV is limited to a finest resolution of 35" by a 1° FOV requirement, and its finest grid pitch's (2.3") energy range is limited to imaging >4 keV due to diffraction. The high angular resolution and loosely constrained detector requirements of bi-grid collimators come at the expense of sensitivity. The two absorbing grid layers reduce the photon throughput to ∼25% of the incoming photon flux, whereas single-grid designs still suffer from a ∼50% reduction.

$$FOV = \frac{1}{2}\frac{pitch}{thickness}$$

The periodic modulation of an RMC's flux in time (total transmitted flux) or space (phase shifted maxima) will have a quasi-triangular form. The raw flux for a time-modulating RMC is plotted in figure 4.6 (top), showing how the total transmitted flux varies in time for a bi-grid RMC. The primary sinusoidal component of the regularized data (figure 4.6, middle) represents a single Fourier component (figure 4.6, bottom) of the source distribution. Measuring the amplitude and phase of the transmitted flux provides a measure of this single Fourier component [104]. For a point source the transmission amplitude is related to the source intensity and transmission phase is related to its location. Considering the transmission profiles as discrete Fourier components can simplify analysis of the angular response and it is possible to reconstruct images by combining the Fourier components.

Bi-grid modulating collimators have been greatly successful in providing high angular resolution imaging for solar and astrophysics observations. The RHESSI satellite [88], launched in 2002 and still in active service as of 2017, employs nine time-modulating RMCs for solar imaging on-board a rotating spacecraft. RHESSI's nine RMCs are able to achieve a finest resolution as high as 2.3 arcseconds. RHESSI's imaging concept and hardware are described in detail in section 5.1. The forthcoming STIX instrument is a spatially-modulated bi-grid design scheduled to fly on ESA's Solar Orbiter in 2017 [81].

Outlook

Until recently, the fine spectral resolution of germanium detectors came at the cost of coarse spatial resolution. This limited applications requiring high-angular resolution and fine spectral resolution to dual-grid RMC designs, compromising sensitivity. In the early 2000s, 3-dimensional position sensitive germanium cross-strip detectors (3D-GeDs) were developed [2] that had excellent spatial resolution without compromising the fine spectral resolution of germanium. The NCT payload demonstrated 3D-GeDs with 2 mm spatial resolution and $\sim$1 keV spectral resolution in a Compton imager [37]. The GRIPS payload recently demonstrated that a finer-pitch version of the NCT detectors was able to achieve $\sim$.5 mm spatial resolution and $\sim$3.5 keV spectral resolution [45] (chapter 8).

GRIPS' imaging system, the Multi-Pitch Rotating Modulator (MPRM) (section 9.1), uses the 3D-GeD's fine spatial sensitivity to implement a single-grid modulating imager that combines the benefits of RMC and coded-aperture methods. This single-grid design combines the high image fidelity generated by the RMC's straightforward fourier components that can simultaneously image over a wide range of spatial resolutions, with the $\sim$2$\times$ greater sensitivity of the coded-aperture method. The GRIPS MPRM is designed for 12.5" – 162" resolution with a .5 mm minimum grid pitch over an 8 m imaging boom. With the maturation of deployable boom technology, at 1 MeV a satellite-borne GRIPS-like MPRM imager using 3D-GeD detectors could achieve better than 5" resolution over a 20 m boom.

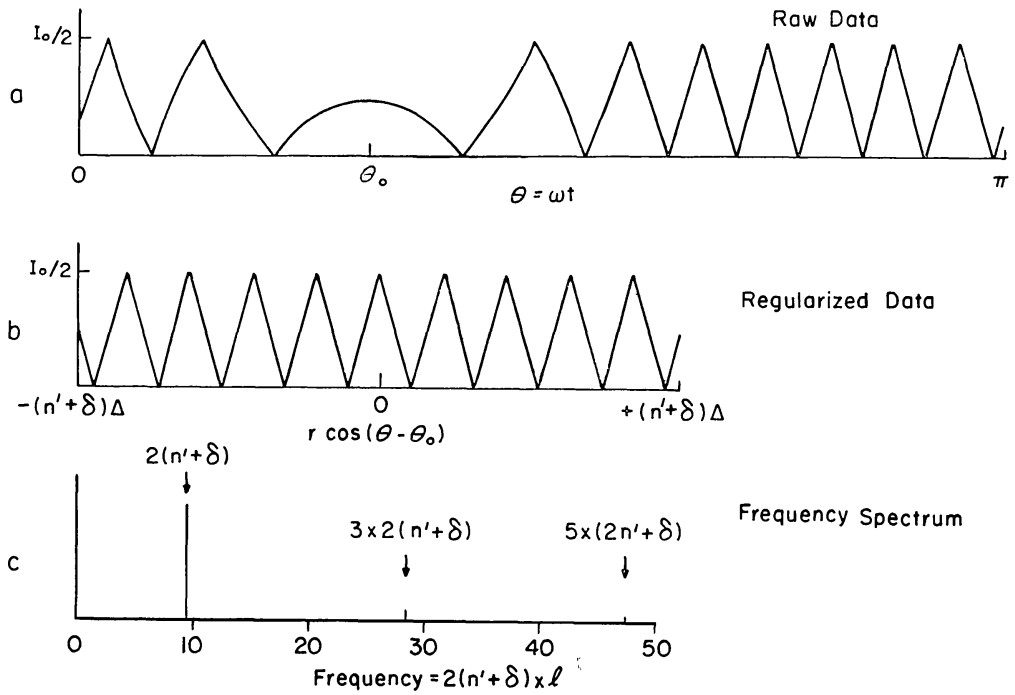

Figure 4.6: The raw data (top), its regularized correction (middle) and fourier component (bottom) for a single grid-pitch in an RMC. The raw data is not fully triangular because of apparent periodic motion of the source in the imaging FOV. Image from the first description of RMCs for astrophysical use by Schnopper et al., 1968 [113].

4.2 Compton Imaging

Compton scattering is the dominant absorption process between ∼150 keV and ∼8 MeV in germanium, with the exact cross-sections varying by the scattering material. In a Compton event, the incident photon deposits only a portion of its energy into the crystal, liberating a recoil electron, and scattering a lower-energy photon. The incident photon's origin can be determined if the energy and direction of both the recoil electron and scattered photon are measured. A basic kinematic sketch, cross-sections and spectral effects of Compton scattering within materials are discussed in section 3.2. The current section discusses the method of Compton imaging, how it compares to other imaging choices, and its merit in the solar context.

Compton imaging does not rely upon lenses, mirrors, masks or grids. Full measurement of the scatter products' energy and direction can locate the true source location for a single photon [38]. A partial measurement of the scatter products locate the single photon to within an annulus (recoil electron direction unknown) or great circle (energy of both scatter products unknown) on the sky, and the intersection of these reconstructions locate the source within

a handful of photons [151]. Figure 4.7 (top) illustrates the basic geometry of the Compton interaction and shows the reconstructed event circle when the recoil electron direction is not measured. The bottom plot shows a superposition of 30 event circles, whose intersection is used to determine the source location.

Since there are no requirements to pass photons through a physical lens or mask the Compton imager can have a potentially full-sky FOV. In practice the FOV will depend on the spatial geometry of the detectors, how the Compton scatter tracks are measured and the presence of ancillary equipment/passive blocking material.

Compton telescopes have typical angular resolutions of $>1^\circ$ and rely upon detector configuration and/or spatial sensitivity to determine scatter directions. The recent COSI flight reported achieving $\sim 5.6^\circ$ angular resolution at 662 keV with germanium strip detectors of $\sim$2 mm spatial resolution [89]. The Compton telescope's coarse angular resolution is without the compromise on sensitivity seen by the higher angular resolution indirect imagers, which sacrifice 50% – 75% of incoming flux. The low fluxes of astrophysical sources make Compton imagers a popular choice for this community in studies of medium gamma-ray energies ($\sim$.2 – 20 MeV) with compact sources. The low angular resolution makes these imagers impractical for studying the Sun however, since the gamma-ray structures of interest need to be resolved on the arcsecond scale.

The Compton technique does not suffer from the ambiguity of indirect imagers, which can only locate the source to within a projection of the grid slits for a single photon, and require hundreds of photons to make an image. Although relatively coarse angular resolution renders the Compton technique impractical for solar imaging, its ability to place a single photon to within a few degrees on the sky can be used for background rejection. This has the potential of drastically reducing background counts for indirect Solar imagers, whose limited sensitivity is usually compensated for with increased detector area.

Outlook

A complete kinematic treatment of the Compton event must account for the original momentum of the scattered electron. The Compton cross section for bound electrons in a material with atomic number Z modifies the simpler unbound cross section with a term called the *incoherent scattering function*, S_i^I [111]. Proper consideration of the momentum distribution changes the calculated electron scatter angle. Since this original bound electron momentum cannot be measured there will be fundamental uncertainties in the reconstructed photon origins, an effect termed *Doppler broadening*. Figure 4.8 (left) shows the angular deviation of the calculated photon origin due to Doppler broadening in silicon, assuming all scattered measurements are perfect. The effect is stronger for more tightly bound electrons and weaker for outer shell electrons whose initial momentum more closely approximates a unbound electron. This effect introduces imaging errors that fundamentally limit the angular resolution of a Compton telescope.

$$(\tfrac{d\sigma}{d\Omega})_{bound,i} = (\tfrac{d\sigma}{d\Omega})_{unbound} S_i^I(E_i, \theta, Z)$$

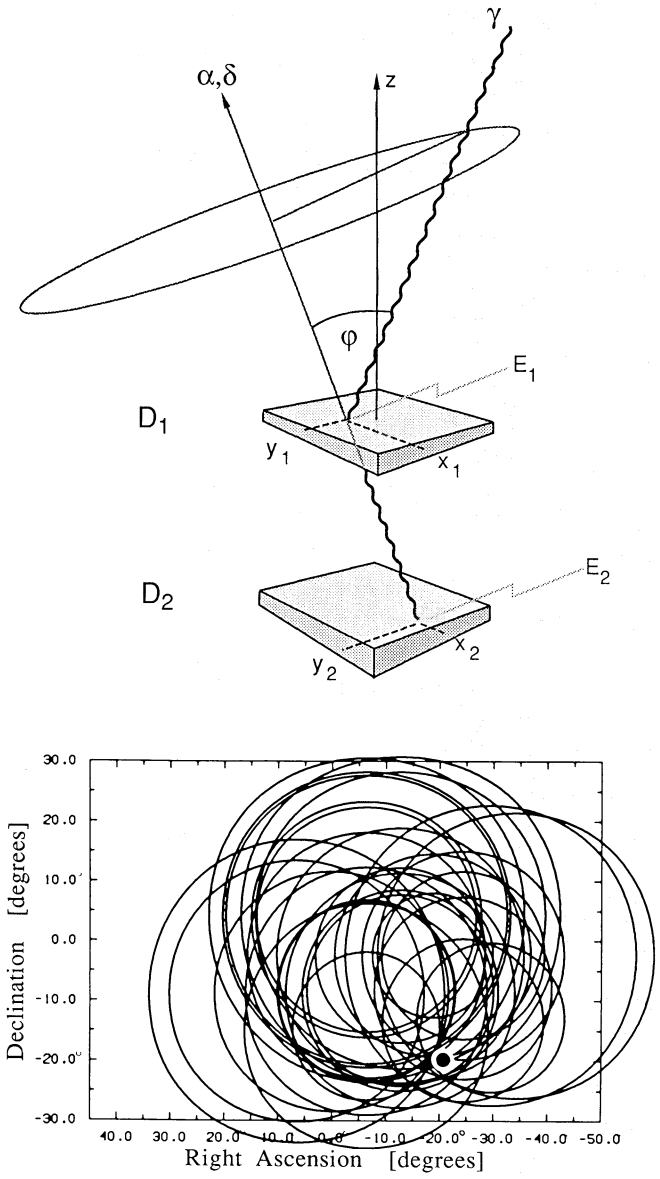

Figure 4.7: *Top:* Schematic depicting a Compton scatter track that deposits energy (E1) to a recoil electron at position (x1, y1) in the first interaction and then photoabsorbs the remainder of its energy (E2) in a second detector at (x2, y2). In this example the recoil electron direction is not measured so the original photon can only be located to an annulus on the sky. *Bottom:* A superposition of 30 event circle reconstruction from a Co-60 source placed at α=20.6°, δ=-19.5°. The intersection of the circles is used to determine the source location. Both images are from vonBallmoos et al., 1989 [144].

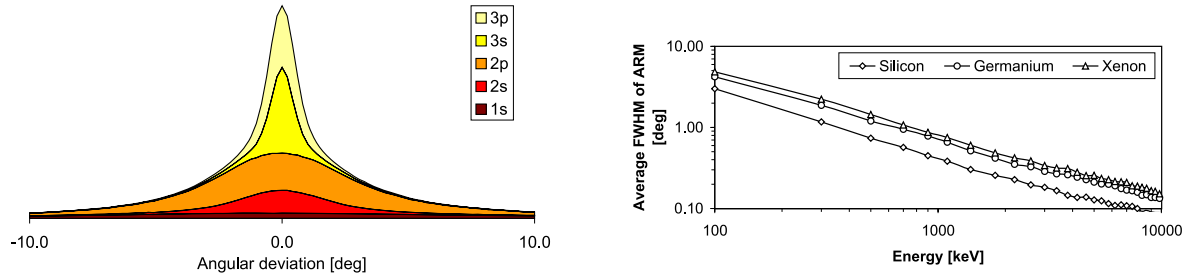

Figure 4.8: *Left:* Simulated Doppler broadening in the reconstructed origins for a Compton telescope. The colors denote angular deviation due to the momentum distribution of bound electrons in different shells of a silicon detector. *Right:* The minimum angular resolution (here denoted as "FWHM of ARM") as a function of energy for three detector materials in a Compton telescope. The minimum resolution becomes $\geq 1^\circ$ for photons energies $<$500 keV. Images from Zoglauer, 2006 [151].

Though a judicious choice of detector material can slightly improve the Compton telescope's angular resolution, for the most popular detector choices in this energy range there are only slight improvements to the response (figure 4.8, right). The angular resolution of the Compton imager is generally limited to $\geq 1^\circ$ for photon energies $<$500 keV. For higher-energy photons the angular resolution limits approach $\sim .1^\circ$ for 10 MeV photons. Though this resolution becomes comparable to observatories like IBIS [146], a coded aperture system with a reported resolution of $.2^\circ$, it is much coarser than the recently flown GRIPS instrument, whose single-grid imager is capable of 12.5 arcsecond resolutions at comparable energies [45]. These two indirect imager systems of course come at the expense of a $\sim$50% reduction in photon throughput and the added complexity of a mechanical grid with aspect and pointing requirements that the Compton telescope can be implemented without. Angular resolution science requirements rule out Compton imaging for solar HXR/gamma-ray research, but its arguably simpler implementation and trade off of (possibly unnecessary) angular resolution for improved sensitivity and greater FOV makes it an attractive tool for the astrophysics community.

4.3 Laue Lens

HXR/gamma-rays undergo crystal diffraction when the incoming light's wavelength and angle (λ, θ) are related to the crystalline interplanar spacing (d_{hkl}) by the Bragg condition:

$$2d_{hkl}sin\theta = n\lambda$$

Crystal diffraction can occur close to the surface of materials as a reflection in the *Bragg geometry*, or as the light is transmitted through the material in the *Laue geometry*. For the

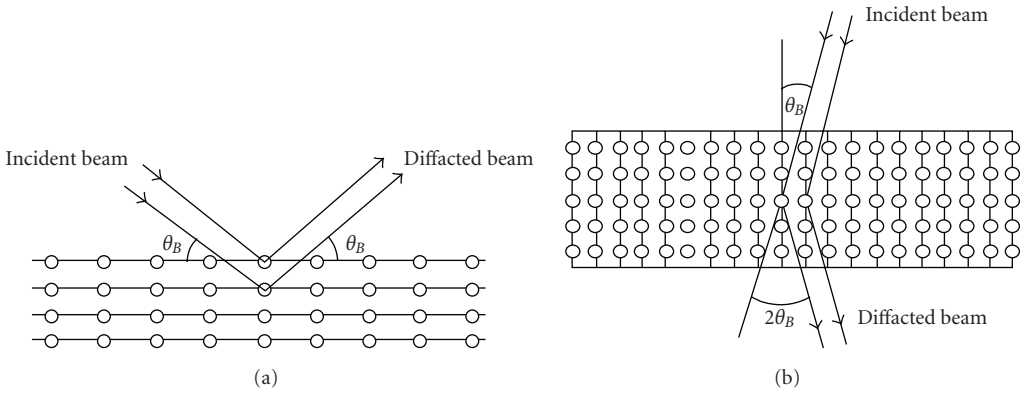

Figure 4.9: Schematic of the two geometries for crystal diffraction. In the Bragg geometry (*left*) photons are primarily refracted from the crystal surface. In the Laue geometry the photons diffract along the volume of the crystal, increasing the efficiency. Image from Frontera & von Ballmoos, 2010 [54].

Laue geometry light has the entire depth of the crystal to interact coherently, increasing the diffraction efficiency. For this reason the Laue configuration has been pursued by the HXR/gamma-ray energy (>100 keV) astrophysics community as a method to directly focus gamma-rays. The Bragg geometry is widely employed in multi-layer coatings for grazing incidence optics, as discussed in the section below.

The energy band of a single crystal is very narrow because of the Bragg condition's fixed dependence on the crystal periodicity. A number of schemes have been proposed for extending a single crystal's passband energy and increasing the diffraction efficiency [54, 26] including creating mosaics wherein the crystal planes are oriented slightly differently, curved crystal planes and bent quasi-mosaic crystals. Even with these schemes the single crystal pass band is limited to a few keV.

The diffracted photon energy (E) as a function of the radial distance (r) from the lens axis and focal length (f) can be determined from the Bragg equation. Assuming first order diffraction (n =1) and small angles, the relevant lens equations are given by:

$$E \simeq \frac{hc\,f}{d_{hkl}r}$$
$$r \simeq \frac{hc\,f}{d_{hkl}E}$$

To extend the energy passband of the lens multiple crystal elements can be arranged at varying radii. The median energy is set by a ratio of the focal length to the radius. Focal lengths on the order of 20 – 100 m have been proposed for instruments with energy ranges of 80 – 600 keV [143].

A Laue lens as envisioned for astrophysics would be an ensemble of crystals of differing materials (varying the crystal spacing d_{hkl}) and/or orientations (varying reflected angle θ)

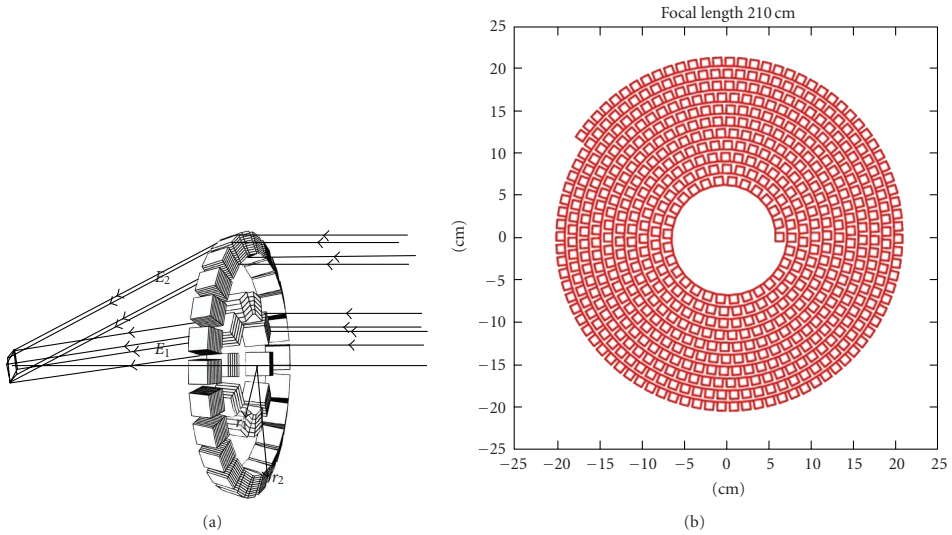

Figure 4.10: Two Laue lens configurations wherein multiple crystal elements are combined with varying crystal materials and/or radii to achieve a wider energy bandpass than available with a single crystal. Different energies of light are diffracted through different angles onto a single detector plane. Lower-energy light is diffracted through narrower angles and E1 $<$ E2. *Left:* This lens design is composed of concentric rings, generating discrete energy bands. *Right:* This design continually varies the radius in the so-called *Archimedes spiral* configuration, yielding a continuous energy band. Image from Frontera & von Ballmoos, 2010 [54].

assembled such that a wide band of energies can be collected and projected onto a single small detector. The general concept of a Laue lens is shown in figure 4.10. Crystals can be arranged in concentric circles yielding discrete energy bands, or in a spiral pattern to generate a continuous energy band.

Outlook

In practice the Laue lens design is limited by a small FOV ($\sim$1 arcminute), narrow passband (likely $\sim$100s of keV) and complex mechanical arrangements to hold and focus the individual crystals. The 2003 stratospheric balloon flight of the CLAIRE payload provided the first, and to date only, astrophysical test of a Laue lens [145]. CLAIRE used a 45 cm diameter diffraction lens of made from 556 germanium-silicon crystals to focus 170 keV light in a 3 keV bandpass. The payload was a success, detecting 33 photons from the Crab within a narrow observing window of 72 minutes.

The Laue lens field is continuing to make significant progress in crystal research [26], demonstrating long focal length lenses [143], narrowing of projected spot sizes [54] and finding solutions to configuration and focusing challenges [12, 143] for the multi-element

arrays. Certainly the NuSTAR [61] mission has demonstrated that current deployable boom technology that can make lens arrays feasible at focal lengths on the order of ∼10m today. Complimenting the Laue lens with a second imager, whose FOV and pass-band is broad, in combined payload is an interesting way to leverage the Laue lens strengths for targeted science goals. Efforts of this kind, combining a Laue lens (.8 – .9 MeV) with a Compton imager (.1 – 10 MeV) were suggested by the DUAL mission concept in a white paper to the 2010 U.S. Astrophysics Decadal Survey [22].

4.4 Grazing-Incidence Optics

In astronomy direct focusing of soft X-rays (<10 keV) through grazing-incidence optics has a history going back to 1965 aboard suborbital rocket flights [55], and space-based observatories since the 1990s (Chandra and XMM-Newton [5]). With the advent of new manufacturing processes to generate smooth mirror surfaces and multilayer optics in precision geometries and with the advent of high-resolution position-sensitive detectors for the focal plane, the grazing-incidence method has extended into the HXR range (≤100 keV) aboard several balloon (HEFT [60], HERO [109] and InFOCμS [97]) and rocket (FOXSI [57]) campaigns. Combined with advances in deployable boom technology, these mirrors have been successfully operating onboard the NuSTAR observatory since 2012 [61]. NuSTAR reports an 18 arcsecond FWHM angular resolution and a 12 arcminute x 12 arcminute FOV, making it comparable to similar rotating grid resolutions without sacrificing sensitivity and only a moderate trade in FOV.

Since X-rays have an index of refraction (n) that is less than 1, total external reflection occurs for incident angles (α_I) less than the critical angle (α_c), as described by Snell's law:

$$cos\alpha_I = ncos\alpha_T \tag{4.1}$$

$$n = 1 - \delta + i\beta \tag{4.2}$$

$$\delta = \frac{n_e r_0 \lambda^2}{2\pi} \quad ; \quad \beta = \frac{\mu\lambda}{4\pi} \tag{4.3}$$

The critical angle is a function of the wavelength and drops off sharply with increased energy. Figure 4.11 shows the X-ray reflectance from 1 keV – 80 keV, illustrating how shallow the grazing angle becomes for HXR/gamma-ray energy light.

Total external reflection of X-rays by a metal is sufficiently efficient to focus SXR. The efficiency drops off with energy and a single reflection does not provide the efficiency needed for imaging HXR. Depth-graded multilayers were developed to improve the efficiency of grazing incidence optics and provide wide-band responses for HXR photons[60]. These multilayers alternate high and low refractive indicies to generate reflections at the boundary of each layer, which then constructively interfere to increase the total reflection efficiency. The condition for constructive interference is effectively the same as the Bragg condition for crystal diffraction, but in this case the relevant distance is the layer depth (D) and not a

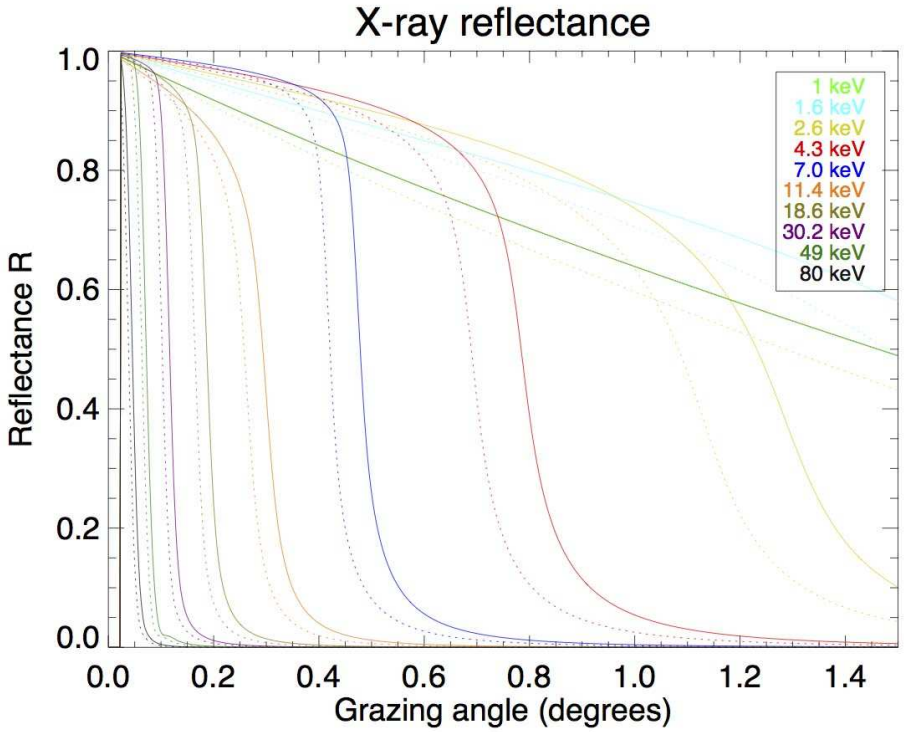

Figure 4.11: The reflectance as a function of grazing angle for X-rays at several energies for a single layer of nickel (solid lines) or iridium (dashed lines). Image from Glesener, 2012 [56].

geometric crystal spacing (d_{hkl}).

$$2Dsin\theta = n\lambda \tag{4.4}$$

The general geometry of the grazing incidence optical array for HXR imaging is shown in fig 4.12. A number of different mirror shapes, mostly based around approximations to the Wolter type geometries [4], have been implemented to focus the light onto a small detector. The photon collection area is drastically increased by nesting many sets of mirrors.

Outlook

Grazing incidence telescopes have become the standard for astrophysics in the <100 keV range with NuSTAR and have been demonstrated for solar physics with the FOXSI rocket flights [57] and NuSTAR solar observing campaigns. Development of depth-graded multilayers and evolving manufacturing techniques enabled grazing incidence technology to overcome efficiency and surface roughness issues previously limiting this method to <10 keV. Expanding the grazing incidence telescope beyond the <100 keV limit of current imagers will be

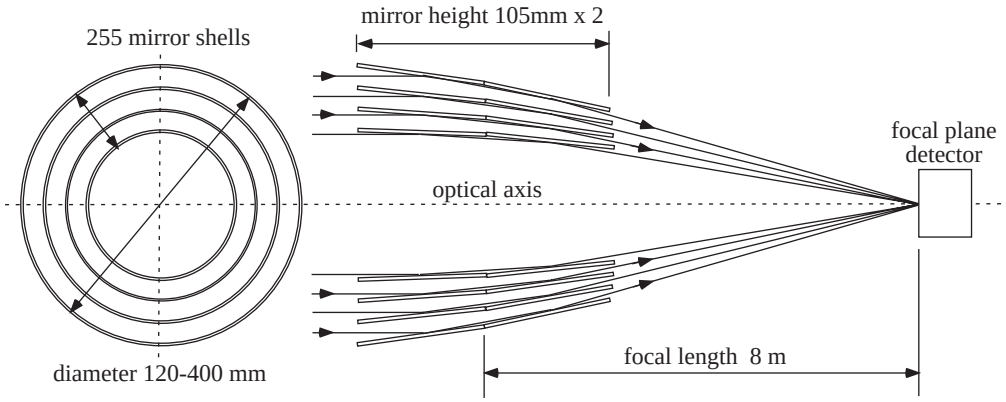

Figure 4.12: Diagram of the grazing incidence optical assembly from the InFocμS payload. A nested set of mirrors shallowly reflects light from two elements onto a detector plane. Image from Ogasaka et al., 2003 [97].

technically challenging. Current manufacturing techniques generate surfaces smooth enough for the grazing incidences required for <100 keV photons, but surface roughness becomes a limiting factor for much higher-energy light, requiring the development of new manufacturing techniques. Furthermore, though multilayers can be used in this energy range (and are in gamma-ray microscopy), the quantity of layers required for adequate efficiencies would drastically increase making these optics heavier and thicker than current versions.

4.5 Fresnel Lens

Fresnel lenses use circular gratings (zone plates) to diffract and focus light. An illustration of several types zone plates are shown in figure 4.13. Modern lens designs use a graded thickness (or graded refractive index) to focus light in the n=1 diffraction order, with a theoretical 100% collection efficiency [125]. Fresnel lens imaging has been widely employed to focus HXR/gamma-rays for microscopy, but an astrophysical implementation has yet to be realized due to the very long, $\sim 10^6$ km, focal lengths needed.

The theoretical lowest spatial resolution (s_{min}) achievable for two diffracted sources in the same focal plane (and consequently the minimum angular resolution, θ_{min}) is given by the Rayleigh criterion: when the maximum of one Airy pattern coincides with the first minimum of the other, the two patterns are just resolvable from one another. The minimum resolution is a tradeoff between the telescope focal length (f), lens diameter (d_{lens}) and the finest pitch

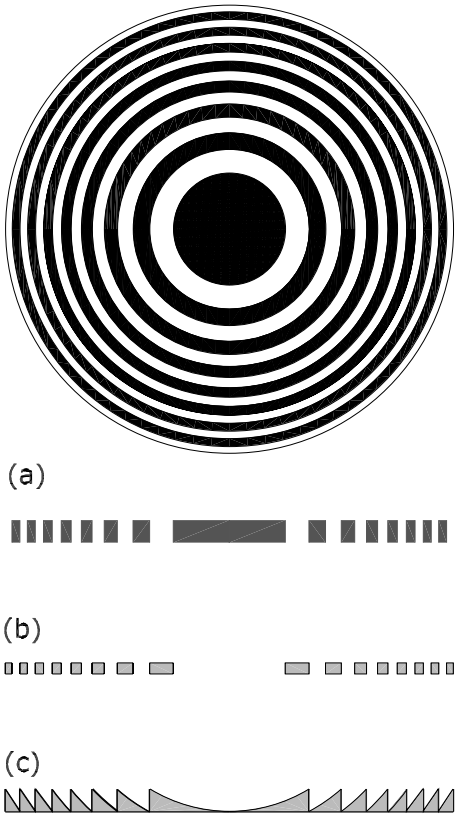

Figure 4.13: Fresnel zone plates used as lenses to directly focus HXR/gamma-rays. *Top:* Collecting area of the zone plate, pointed toward the source. *(a):* A Fresnel Zone Plate (FZP) that uses opaque gratings to diffract light. The pitches become smaller at larger radii to deflect through larger angles. Shadowing from the opaque gratings lead to ∼10% efficiency *(b):* A Phase Zone Plate (PZP) which replaces the opaque grating with a material that induces a π phase shift to increase the throughput, increasing efficiency to ∼40%. *(c):* A Phase Fresnel Lens (PFL) which used a graded material thickness (or alternatively a graded index of refraction) to redirect all incoming light into an n=1 diffraction order, yielding a theoretical efficiency of 100%. Image from Skinner, 2001 [125].

of the diffraction grating (p):

$$s_{min} = 1.22\frac{\lambda f}{d} \tag{4.5}$$

$$= .61p \tag{4.6}$$

$$\theta_{min} = \frac{s_{min}}{f} \tag{4.7}$$

Using these equations as a guide, a telescope sensitive to 511 keV would have a focal length of 10^6 km, a mask diameter of 5 m, an angular resolution of .1 μarcseconds and a very small FOV of $\sim$100s of μarcseconds [125].

Outlook

A 2005 study by NASA Goddard Space Flight Center's Integrated Mission Design Center (IMD) [78] provided a detailed mission study for two theoretical Fresnel telescope concepts. The study examined the feasibility of formation flying by two spacecraft, one carrying the lens and the other carrying the detector, at a distance of 10^5 km for a "Pathfinder" mission (10 μarcsecond resolution) and at a distance of 10^6 km (1 μarcsecond resolution) in a full mission concept. The study found that the formation flying mission is viable inside the science requirements of pointing, control and aspect knowledge with current, or expected, technology.

Though difficult to ground qualify a lens with a 10^6 km focal length, efforts have been made to build and qualify scaled versions [79]. A mission with a focal length of 1000 km to achieve milli-arcsecond imaging at 5 – 11 keV with a modified Phase Fresnel Lens was proposed in 2008 [126], but not selected for continued study. A 2012 study demonstrated a Fresnel lens concept at 5.4 keV in NASA Goddard's 100 m beamline with the goal of providing .1 arcsecond imaging resolution for solar HXR studies [41]. Fresnel lenses are an interesting concept to study the medium-energy gamma-ray universe in dramatic detail, trade studies and scaled concept designs show that a mission design is feasible. In the coming decades one can imagine that a small FOV instrument with μarcsecond angular resolution might become very attractive to the planetary/stellar physics community. The full Fresnel diffraction limited concept could provide a window into extra-solar planetary systems, capable of resolving the corona and winds of distant stars.

4.6 Summary

Indirect imagers provide the highest resolution imaging solution for the complete 1 keV to $\sim$10 MeV energy range. The availability of detectors that can provide high resolution position sensitivity has made it possible to develop the GRIPS single-grid indirect imager that has a high image fidelity from sampling many spatial frequencies simultaneously and a high photon throughput compared to bi-grid designs. The position-sensitive detectors can be

used as imagers in their own right, through careful tracking of the Compton scatter products single-photon origins can be determined to annuli on the sky. This Compton imaging is fundamentally limited to $\sim 1^{\circ}$ resolution, making it a poor choice for solar studies but a powerful tool for background rejection and polarimetery measurements. Both Compton and indirect imaging rely on a combined photon collection and detection area which can lead to high background counts. Direct focusing optics like Laue lenses, Fresnel lenses and grazing-incidence optics decouple the photon collection and detection areas, dramatically increasing signal detectability by reducing noise. Direct focusing methods have yet to be realized for >100 keV energies in astrophysics, though these communities are making steady progress towards demonstrating their viability in space.

Chapter 5

Gamma-Ray Imaging-Spectroscopy of the 2003 October 29 X10 Flare with RHESSI

One of the largest flares to occur in the past two solar cycles was an X10 event on 2003 October 29. Though this flare had excellent observational coverage by RHESSI it is relatively unstudied. Deriving a spectrum for this flare is complicated by a large and unpredictable atmospheric background that occurred as a result of energetic particle precipitation from Earth's radiation belts.

The dataset of gamma-ray flares is extremely limited. As discussed in chapter 2 there are only a handful flares in the RHESSI and SMM datasets which have detectable 2.2 MeV neutron-capture lines. Spectral fits for these flares show that the fluence of bremsstrahlung produced by relativistic electrons is tightly correlated to the ion-generated 2.2 MeV line fluence, suggesting a relationship between how these two populations are accelerated and transported [121]. The set of flares capable of being imaged at the 2.2 MeV line is much smaller. In the past 15 years of observation RHESSI has only observed five flares large enough to image the 2.2 MeV line. Four of these flares (2002 July 23, 2003 October 28, 2003 November 2 and 2005 January 20) showed a significant displacement of the 2.2 MeV image centroid relative to the HXR centroids [66, 67], seemingly inconsistent with the correlation result (fully discussed in section 2.7). The fifth flare imaged at gamma-ray energies is the 2003 October 29 event. Imaging of this event at 2.2 MeV was limited to grid 9, with a relatively coarse 183" resolution. The observed ∼13" centroid separation for the HXR and 2.2 MeV emission is within statistical uncertainty (17") and the spatial result is inconclusive. Since this flare's spectrum was obscured by background it is not included in the set of flares in the correlation study. Generating a reliable spectra for the 2003 October 29 flare that puts it into context with the other four flares with imaged 2.2 MeV lines, and the ensemble of flares on the correlation plot, can significantly improve these limited datasets.

Imaging-spectroscopy is commonly used as a tool to separate background from source counts in lower-energy studies (<300 keV) because there are sufficient source photons to

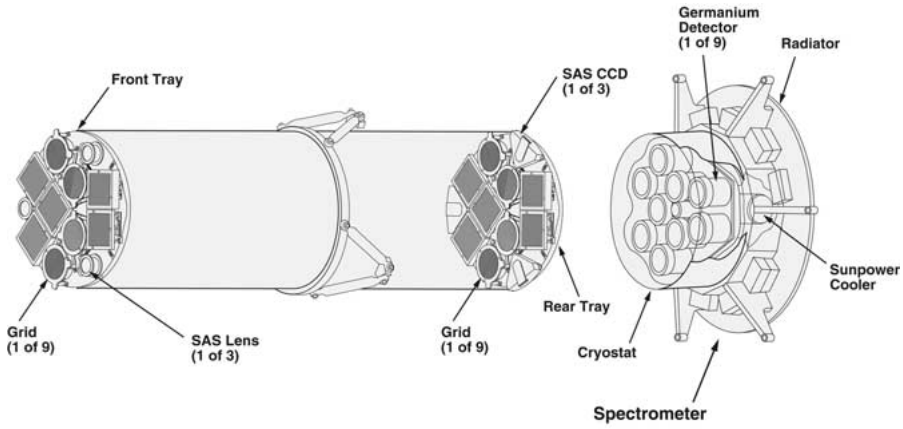

Figure 5.1: Schematic of the RHESSI telescope's imaging system. Nine subcollimators of matching grid pairs are set in front of nine monolithic germanium detectors. Also shown are the Solar Aspect System (SAS) components which provide precise pointing information used for image reconstruction. Image from Hurford et al., 2002 [68].

generate images with moderate energy bands (∼15 keV) and the full imaging capability of RHESSI's nine RMCs are available. At gamma-ray energies the imaging-spectroscopy method has been mostly unexplored. The gamma-ray range faces challenges resulting from limited imaging capabilities and low fluxes, even for the largest flares in the RHESSI catalog.

This chapter explores using imaging-spectroscopy in the gamma-ray energy range by comparing the derived spectrum against the results from background-subtracted spectra in well-studied flares, and applies this technique to the 2003 October 29 flare. Parameter fits to the image-derived spectra are used to determine the relativistic electron and 2.2 MeV line fluences, and this flare is compared against other flares in the correlation study. It is found that the 2003 October 29 flare shows a tight correlation between relativistic electrons and ions, similar to other flares of its size.

5.1 The RHESSI Imager

The RHESSI instrument uses a set of nine Rotating Modulating Collimators (RMCs) to indirectly image photons from 3 keV to ∼10 MeV. Detailed descriptions of the imaging concept, hardware and mission overview can be found in literature [68, 149, 88]. This section provides a basic introduction to RHESSI imaging as context for using its backprojected images to derive spectra.

Each RMC consists of two grids with matching pitches, co-aligned in front of a monolithic germanium detector. The grids selectively allow photons to be transmitted/blocked by a series of alternating transparent slits and opaque slats. As the spacecraft rotates the source photons are time modulated with an amplitude proportional to the incident flux, a phase

TABLE I
RHESSI – nominal grid parameters.

Subcollimator number	1	2	3	4	5	6	7	8	9
Pitch (mm)	0.034	0.059	0.102	0.177	0.306	0.530	0.918	1.590	2.754
slit width (mm)	0.020	0.035	0.061	0.106	0.184	0.318	0.477	0.811	1.487
FWHM resolution (arc sec)	2.26	3.92	6.79	11.76	20.36	35.27	61.08	105.8	183.2
Max. transmission	0.60	0.60	0.60	0.60	0.60	0.60	0.52	0.51	0.54
Grid thickness (mm)	1.2	2.1	3.6	6.2	10.7	18.6	6.2	6.2	30.0
Slat material	Mo	W	W	W	W	W	W	W	W
Field of view (deg)	1.0	1.0	1.0	1.0	1.0	1.0	4.4	7.5	2.8

Figure 5.2: Table of RHESSI grid parameters for the nine RMCs, from Hurford et al., 2002 [68].

which gives source location information and a period set by the design parameters. This type of time-modulating RMC design is described in more detail in section 4.1. The parameters of each grid are given in table 5.2. Since the grid thicknesses and materials vary, each grid has slightly different transmission properties. Only grids 6 & 9 have sufficient thicknesses to be opaque to >1 MeV photons, and consequently are the only two which can be used for imaging in this energy range.

The RMCs time modulate incoming flux, generating a *modulation profile* that represents a convolution of the source properties with the grid. In the grid/detector reference frame the source moves in a circle about the spacecraft rotation axis. Since the grids only modulate flux which is perpendicular to the direction of the slits, the modulation profile will not be a constant sinusoid; the period of time when the source moves parallel to the grid (i.e. when it changes direction) will not be modulated and transmission will be at the maximum value. This can be seen in the panels of figure 5.3. Panels 1 – 6 show the modulation profiles for the sources shown to the right. Panel (1) shows a generic source of unit flux and then shows sources with varying flux (2), position (3, 4) and source size (5, 6). Panel 7 illustrates the inverse problem at the core of RHESSI imaging: given a modulation profile, what is the original distribution? Further complicating the inversion is the fact that there could be multiple sources inside the same FOV.

Inside RHESSI's spectrometer [128] are nine high-purity germanium crystals with coaxial geometries. The detectors are electrically segmented into front and rear sections which are independently read out, time tagged and digitized. A cross-section of a single detector (left) and a rendering of the cryostat housing showing all 9 detectors (right) is shown in figure 5.4. The front segment's ∼1 cm thickness primarily detects photons from 3 keV to ∼200 keV. The rear segment's additional collection volume extends the sensitivity up to the full ∼17 MeV energy range. Since the general shape of flare spectra in RHESSI's energy range are powerlaws, there can be ∼12 orders of magnitude fewer counts in the higher-energy gamma-

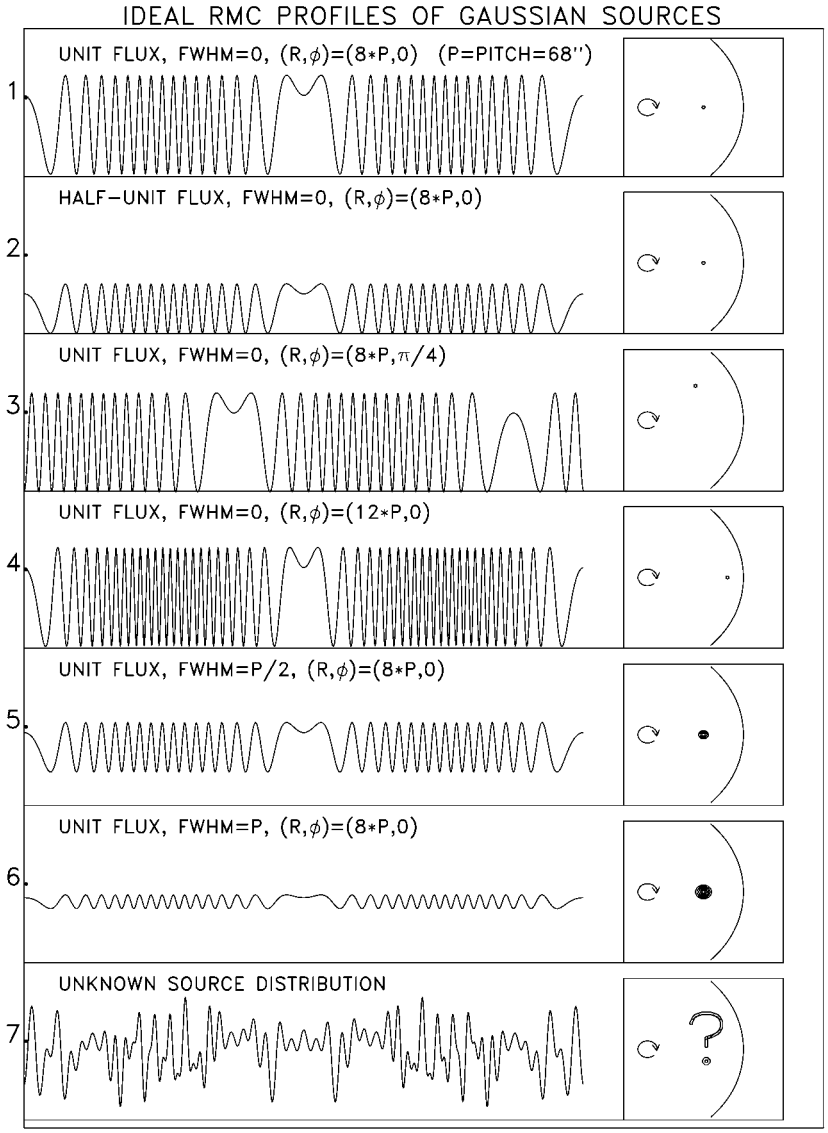

Figure 5.3: Simulated modulation profiles for a series of sources, inset to the right of each panel. Panel (1) shows an example source of unit flux. The successive images can be compared to the first panel with varying flux (2), position (3, 4) and source size (5, 6). Panel 7 illustrates the inverse problem at the core of RHESSI imaging: given a modulation pattern, what is the original distribution? Image from Hurford et al., 2002 [68].

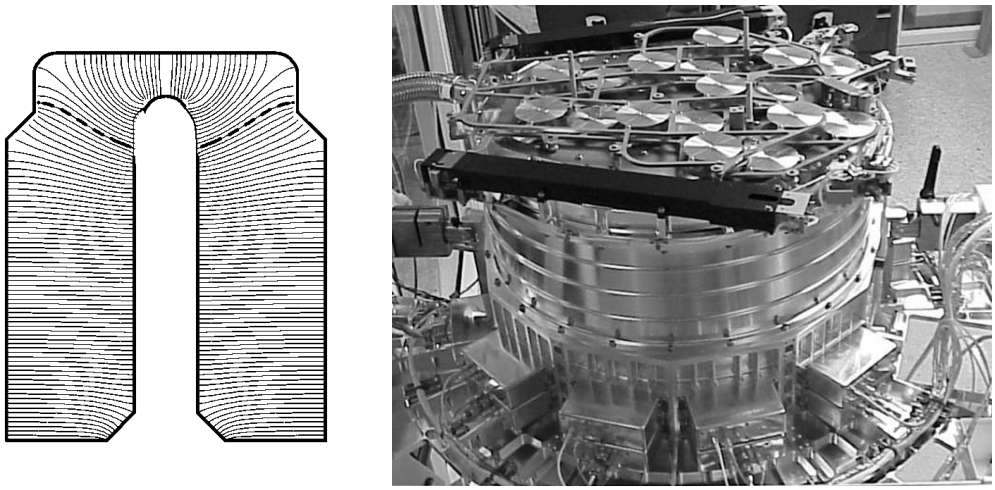

Figure 5.4: *Left:* Cross-section of RHESSI's co-axial detectors. The dashed line at the top indicates where the detectors are electrically segmented into front and rear sections. *Right:* The RHESSI cryostat before being integrated onto the spacecraft bus. Shown on top are the two sets of aluminum attenuators, one set engaged and sitting on top of the detectors, the other set out. Images from Smith et al., 2002 [128].

rays than those of the lower-energy Soft X-rays (SXR). In RHESSI's detectors the front segments "shield" the rear segments from the low-energy photons, effectively extending the dynamic range during high flux events. The detectors themselves have no spatial resolution beyond the front and rear segment designation.

The finite timing of the readout system becomes a consideration for high count rates [128]. Since a new event cannot be triggered while a previous event is being readout there will be counts lost during times of high flux. The RHESSI electronics use a parameter called the *deadtime* to account and correct for events lost during previous event processing. As the count rate exceeds a software threshold the spacecraft automatically deploys a set of attenuators to reduce the low-energy photon flux and maintain high sensitivity to photons >100 keV. There are two sets of attenuators onboard which deploy either independently or together, providing three staged levels of physical decimation. Each of these methods provide a means to correct the measured count rate and recover the actual count rate.

During a single trigger window multiple events can deposit energy into the detector simultaneously. Since the RHESSI detectors have no spatial resolution these two photons cannot be discerned from one another in detector data alone and their combined energy will be read out as a single photon energy. This pulse *pile-up* can manifest as excess counts above a high-energy line. If the higher-energy and lower-energy photon sources are in spatially different locations then imaging can discern between the two different sources to confirm that pile-up is occurring, but the spectral information of the exact photon energies will be lost.

5.2 RHESSI Image Reconstruction

Backprojection is the most straightforward method of generating images from RHESSI's time-tagged counts [68]. Counts are time binned and a 2-D probability distribution of where the photons could have originated from, called a *modulation pattern*, is generated. This modulation pattern essentially projects an image of the grid into the sky, assigning the same probability to each slit location and a zero probability to each opaque grid location. For a single time bin the modulation pattern looks like a series of stripes on the Sun. As the spacecraft spins, the modulation patterns are collected over a range of grid orientations. The modulation patterns from different times and orientations are linearly added to generate the backprojection map. The source regions will continue to build up as more patterns are added together and the other regions will become fainter in comparison. The regions outside the source will also be built up, but at a much slower rate than the source location because no photons are collected when the source is covered by an opaque slat. Collecting photons for a range of modulation patterns adds to the image fidelity. Figure 9.2 illustrates how backprojection maps are created and linearly added for the GRIPS instrument. This figure differs from a single RHESSI RMC because multiple angular resolutions are added in the same backprojection image and this generates an image of higher fidelity, but the method is conceptually the same.

Backprojection is equivalent to adding Fourier components together in an inverse 2-D Fourier transform. A single grid pitch at a single orientation generates a modulation profile that is an analog to a point on the U,V plane. As the grids rotate through 360° a circle is drawn in the UV plane, describing the Fourier components which can be used to make an image. With RHESSI's 9 RMCs of varying pitch, its response has equivalently 9 circles in the UV plane. Incomplete sampling of this plane leads to characteristic lobes in a point-response function for a Fourier transform. These lobes are visible in the RHESSI backprojection images. Figure 6.5 shows the point-response function of RHESSI's gamma-ray response (two spatial frequencies from grids 6 & 9) in comparison to the GRIPS instrument's point-response. GRIPS has 13 spatial frequencies that are sampled through 360°. RHESSI's incomplete sampling generates large side-lobes whereas GRIPS' additional sampling generates a response that is virtually free of sidelobes.

After RHESSI's initial backprojection map is generated, an algorithm can be applied to further resolve the image. These algorithms are nonlinear, meaning that once they are applied separate maps cannot be added together in a straightforward manner. Each of the techniques make assumptions about the source that are used to generate predicted modulation profiles. These predicted profiles are then iteratively compared to the observed profile until they reach an acceptable agreement with observations. A description of these methods can be found in literature [68].

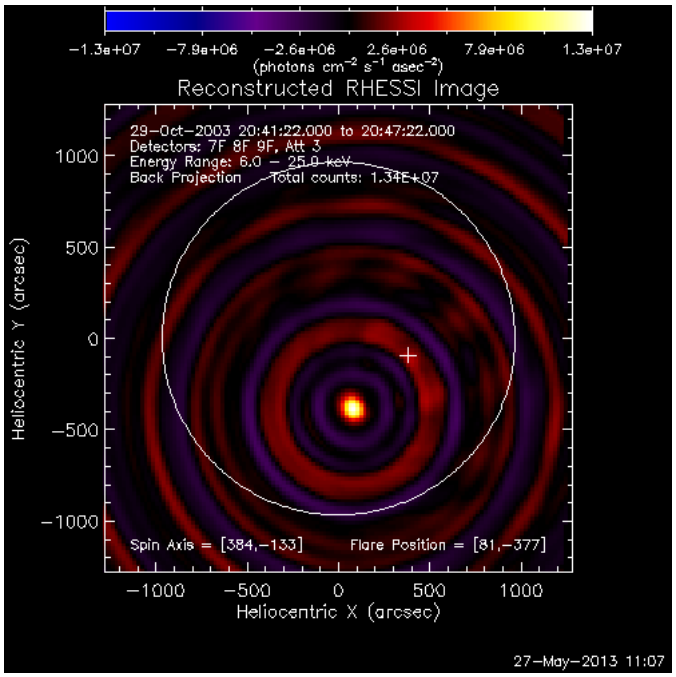

Figure 5.5: Backprojection image from the 2003 October 29 flare using RHESSI's grids 7, 8 and 9 over the energy range of 5 keV – 25 keV. The characteristic sidelobes from incomplete sampling in the U,V plane are visible. The spacecraft rotation axis is marked by a white cross. Image provided by RHESSI Browser [123].

5.3 2003 October 29 Flare

The X10 flare that occurred on 2003 October 29 at ∼20:40 UT is shown in figure 5.6 (top) for the GOES X-ray flux bands in 3-second intervals, with the RHESSI observation times overlaid in red and blue bold traces. RHESSI has good coverage of this flare through the impulsive phase and catches multiple observations of the decay. Continuous solar coverage is not possible due to RHESSI's orbit; night periods occur when the spacecraft is behind the Earth (shaded in purple bands) and observations are obscured by high radiation background when the spacecraft passes through the South Atlantic Anomaly (SAA, shaded in green bands).

Particles from an X17 event that occurred on 2003 October 28 at ∼11:00 UT reached the Earth by the following day where they were trapped in the radiation belts. Precipitation of these energetic particles from the radiation belts generated large and unpredictable atmospheric background counts preceding and during the 2003 October 29 flare, complicating simple background-estimation techniques. Figure 5.6 (bottom) is the RHESSI count rate over the interval of ∼19:50 UT to ∼21:30 UT, combined across six detectors and divided into coarse energy bands. This plot shows a large background signal extending over the 12 keV – 800 keV energy range that directly precedes the flare (before 20:40 UT).

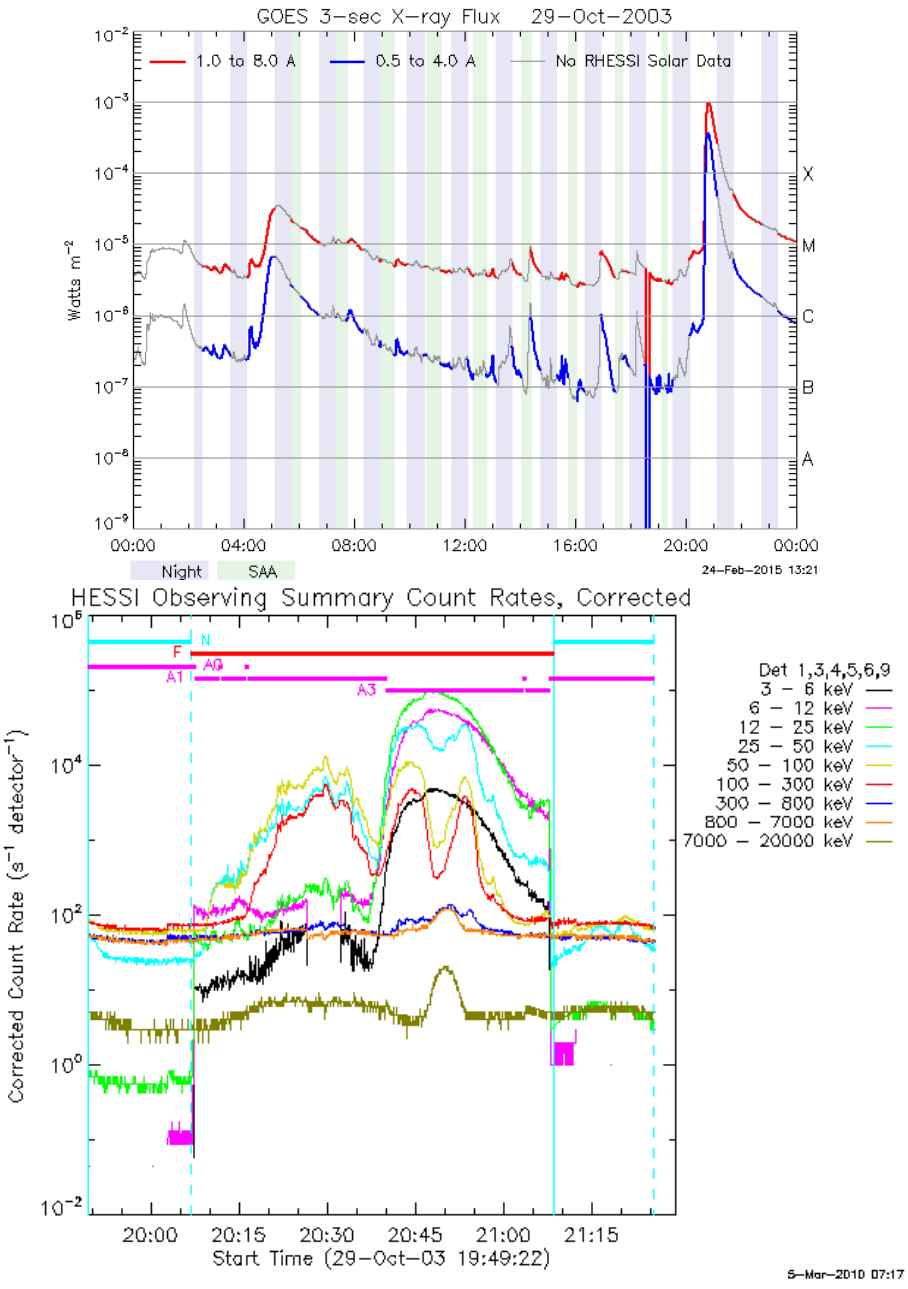

Figure 5.6: *Top:* GOES X-ray bands with RHESSI observing times overlaid in blue and red bold traces, the gray traces are GOES observations without RHESSI observation times. RHESSI had good coverage of the impulsive phase (starting at ∼20:40 UT) and multiple observations during the decay. *Bottom:* RHESSI count rates, corrected for attenuator motions (attenuator states A0, A1 and A2 are labeled and shown in magenta bands). Counts are combined across multiple detectors and divided into coarse energy bands. Images provided by RHESSI Browser [123]

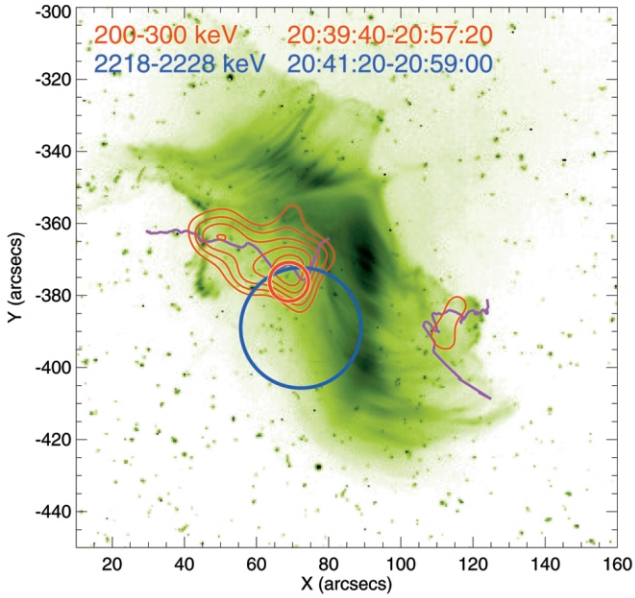

Figure 5.7: Image of the 2003 October 29 flare showing the RHESSI centroids for 200 keV – 300 keV HXR emission (red) and the 2.2 MeV line (blue) overlaid on a TRACE 195$\mathring{A}$ image. Image from Hurford et al., 2006 [67].

Since the atmospheric background generated by the precipitating electrons is large and diffuse, it does not affect imaging. RHESSI centroids for 200 keV – 300 keV HXR emission (red) and the 2.2 MeV line (blue) overlaid on a TRACE 195$\mathring{A}$ image is shown in figure 5.7 for the 2003 October 29 event [67]. The fine red lines show the 50 – 90% high-resolution contours of the HXR footpoints and the purple lines denote their motion. The 2.2 MeV line could only be imaged by grid 9 with 183" resolution, its 17" 1 σ error radius is shown as the thick blue line. Since HXR/2.2 MeV centroids are only displaced by 13" the displacement for this flare is within the 1 σ error and not statistically significant. The centroid displacements for the other 3 flares with imaged 2.2 MeV lines are significantly displaced from the HXR emission, by as much as 25". A review of Hurford et. al's findings for the complete set of imaged 2.2 MeV flares is found in section 2.7.

5.4 Gamma-Ray Imaging-Spectroscopy

For typical RHESSI spectroscopy the ambient background can be easily quantified by averaging the quiet solar data from the 15 orbits preceding and following the flare in the same geomagnetic location. The ambient background is then subtracted from the total detected spectra to obtain the predicted source spectrum. For most flares this method yields an acceptable energy-dependent background model and constrains the source counts well.

However, in times of high and irregular background, this method cannot be effectively used to generate reliable source spectra.

RHESSI's imaging properties can be leveraged to separate compact source counts from extended source counts, like background. RHESSI's RMCs do not modulate flux from sources whose angular extent is much greater than the angular resolution of the grids. Extended sources effectively become an overall normalization to the modulation profile, acting like a DC offset. This can be seen in figure 5.3's panels 5 and 6. Since these sources are much larger than the grid angular resolution, there is always some portion of the source visible through the grid, generating the DC offset and reducing the modulation amplitude. For an extremely big source the modulation would be negligible in comparison to the overall DC offset.

Since the background is effectively separated out by imaging, the imaged counts can then be used as a measure of source counts in a method called *imaging-spectroscopy*. Though useful this method has several drawbacks that make spectroscopy with background models directly computed and subtracted more attractive for general use.

In RMC-style imaging a single photon cannot be backprojected to the source, its origin can only be guessed to within ∼50% of the instrument field of view. Many photons over multiple different grid orientations must be binned and added together to generate an image whose peak location is consistent with the source and whose peak maximum is proportional to the source.

Images with sufficient signal-to-noise ratios (SNR) are needed to resolve the source, and usually this requires energy ranges that are much broader than the FHWM energy resolution of RHESSI's high-purity germanium detectors. The germanium detectors have excellent energy resolutions of ∼1 keV at low-energies, reaching ∼10 keV at higher-energies [128]. Even the best observed flares in the RHESSI catalog require energy bins that are 5 – 20x broader than can be achieved with spectroscopy alone.

Each backprojected image represents a proportional flux over its imaged energy range. Photon distributions that are structured on the angular resolution scale of the imaging grid will have real source counts distributed in the backprojection sidelobes. Post-backprojection techniques seek to improve image quality by redistributing this flux into the image peaks. In the high-flux, low-energy, portions of the spectrum there are enough counts and available Fourier components to effectively apply these post-processing techniques.

The imaging-spectroscopy technique is well used in RHESSI's lower-energy range (<300 keV) where there are ample counts to generate moderate-energy-band images, and the full complement of imaging RMCs available for use. This method is relatively unexplored in the gamma-ray energy range, however, where there are comparatively few counts in even the largest and best-observed flares. Further complicating the use of imaging-spectroscopy in gamma-rays are limited imaging capabilities; only grids 6 & 9 are thick enough to modulate (and image) photons >1 MeV.

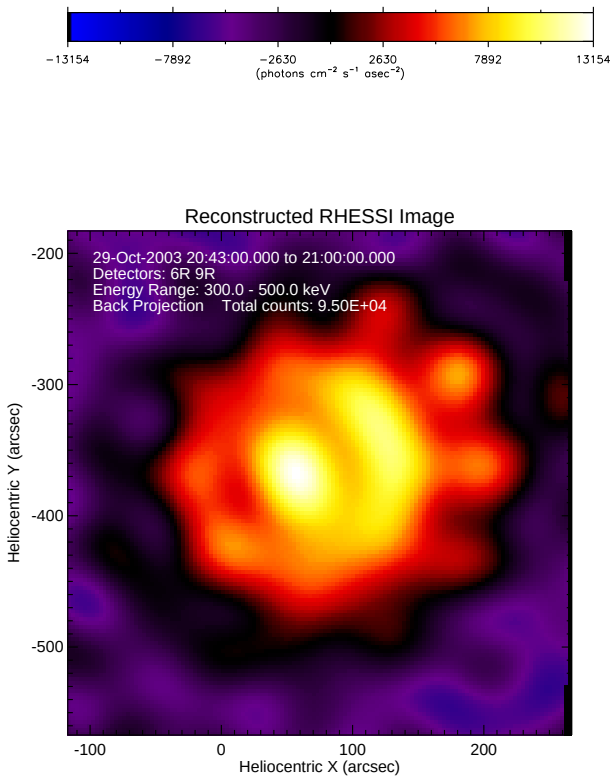

Figure 5.8: Backprojection image from the 2003 October 29 flare using grids 6 & 9. This image indicates there is source structure on the 35.27" (grid 6) scale. To ensure that the peak of the backprojection dirty map is a direct measure of count flux, the source needs to be fully unresolved in the image. For this reason, only detector 9 (183.2") was used for gamma-ray imaging-spectroscopy in this work.

Method

The methods used to increase image fidelity and redistribute mapped counts into the back-projected image peaks require high statistics that aren't available in the gamma-ray range. Instead, the imaged-flux must be derived from the peak of the backprojection map directly. The source must be fully unresolved on the spatial scale of the modulation profile for the peak in the backprojection map to contain all the flux. Figure 5.8 indicates that there is structure on the 35.27" (grid 6) scale for the 2003 October 29 flare. To ensure that the response most closely simulates a fully unresolved point source only grid 9 (183.2") was used for imaging-spectroscopy in this manuscript.

Hurford et al. 2002 [68] describe the method of generating a RHESSI backprojection image, which is summarized here. A backprojection map region of interest is chosen that adequately covers the flaring region but is small enough to exclude the rotation axis, to avoid

ghosting issues caused by incomplete spatial sampling. RHESSI's Solar Software (SSW) IDL routines then calculate the modulation patterns (P_{im}) for each pixel (m) and time bin (i) and combine them into the backprojection map. Since the sensitivity is proportional to the modulation profile variance a flat-fielding correction is applied to the individual maps before they are combined. The mean over all roll angles is subtracted out for each pixel and then each value is divided by its variance over all roll angles.

The intensity of each pixel (I_m) is a linear combination of the count rates (C_i) and the flat-fielded modulation patterns ($\widehat{P}_{im}$).

$$I_m = \frac{1}{A_{eff}} \sum_{i=1}^{N} \frac{C_i}{\Delta t_i} \widehat{P}_{im} \tag{5.1}$$

The image is normalized by the effective area of the detector (A_{eff}) and the time interval (Δt_i) of the image. This normalization ensures that the peak of the backprojection map equals the strength of the dominant source. The RHESSI software returns a backprojected image over the selected FOV, energy and time range. The image itself is a set of intensities for each pixel.

To determine the photon flux from the backprojection map the Solar Software (SSW) *hsi_map_evaluator* routine was used. This routine determines the photon flux (Γ) by finding the maxium pixel value in the backprojection map (I_{max}) and then passes the counts through the detector response matrix (drm). The total integration time (t_{map}), livetime (τ) and the detector's cross-sectional area (A) are also accounted for.

$$\Gamma = \frac{I_{max}}{A\ drm\ t_{map}\ \tau} \tag{5.2}$$

The image's statistical signal-to-noise ratio (SNR) was used as a control parameter to choose the narrowest energy bands possible for imaging while retaining adequate signal fidelity. Certain bins around key features (511 keV, 2.2 MeV and the escape peaks) were chosen over a fixed energy range, usually with very high SNR. The binning across the remainder of the spectrum was iterated until the SNR for the image was in the range of 4 – 5. The SNR for a backprojected map is given by a ratio of the mean modulation amplitude ($\bar{A}_{mod}$) and source counts (C_{src}) to the root of the total map counts (C_{map}), as returned by the *hsi_map_evaluator* routine.

$$SNR = \frac{\bar{A}_{mod}\ C_{src}}{\sqrt{2\ C_{map}}} \tag{5.3}$$

The imaging-spectroscopy results were compared with well-studied flares. Since the responses of the imaging system and spectrometer are different, and converting counts to photons involves additional layers of assumptions, it was necessary to choose how to make comparisons. To minimize the number of times a model-dependent forward fit entered into the analysis, the comparison was done in the *count flux* space of the imager.

Comparison spectra derived through background-subtracted spectroscopy were forward fit to generate a photon model spectrum. Then, this photon model was passed through the detector response matrix of the imager. The resulting spectrum represented the expected count flux in an image, given the photon model derived through traditional spectroscopy. The count flux for imaging-spectroscopy is straightforward to obtain. The photon flux derived from *hsi_map_evaluator* (Γ) is multiplied by the detector efficiency.

5.5 Comparison to Background-Subtracted Spectroscopy

Gamma-ray imaging-spectroscopy was compared against background-subtracted spectra for the 2002 July 23 X4.8 and 2003 Nov 02 X8 flares. The results from both methods are plotted in figure 5.9 as fluences in the count space of the imager. The imaging-spectroscopy results are plotted as black crosses whose widths denote the spectral binning and heights represent uncertainty. The background-subtracted spectrum for 2002 July 23 is plotted as a blue line in the top plot. The 2003 November 02 flare's background-subtracted spectrum has been rebinned to match the imaging-spectroscopy binning and then plotted as blue triangles in the bottom plot. The green triangles represent the parameter fit to the 2003 November 02 flare's imaging-spectroscopy results. The overall spectral shapes, key features and magnitude of fluxes are in good agreement between the two methods for both flares.

The next sections address several differences between the spectra derived by imaging-spectroscopy and background-subtraction methods. These features include discrepancies between the >2 MeV continuum, the count fluence of the 511 keV line and the appearance of a broad feature around ∼430 keV in the imaging-spectroscopy result. While these differences are interesting they do not significantly affect the parameter fits for the >300 keV relativistic electron bremsstrahlung and 2.2 MeV line fluences which are the primary motivators for applying this method to the 2003 October 29 flare. The purpose of the next sections are to highlight these anomalous features, explain our efforts to characterize them to date and point to some interesting research avenues.

Spectrum above 2 MeV

The >2 MeV count fluences differ between the imaging and background subtraction methods for both flares. The 2003 November 02 flare's the count flux begins to diverge around the 511 keV line and smoothly extends into the MeV range. This discrepancy could be explained by an energy-dependent background model that overpredicts the background at higher energies, resulting in a steeper spectrum. The 2002 July 23 flare's discrepancy, however, is more difficult to interpret. The sudden rise in flux for the imaged spectrum over the 2.23 MeV – 2.95 MeV range could not easily result from a slight overestimation of a smooth background. For the 2002 July 23 flare we must determine if this count excess is the result of real source

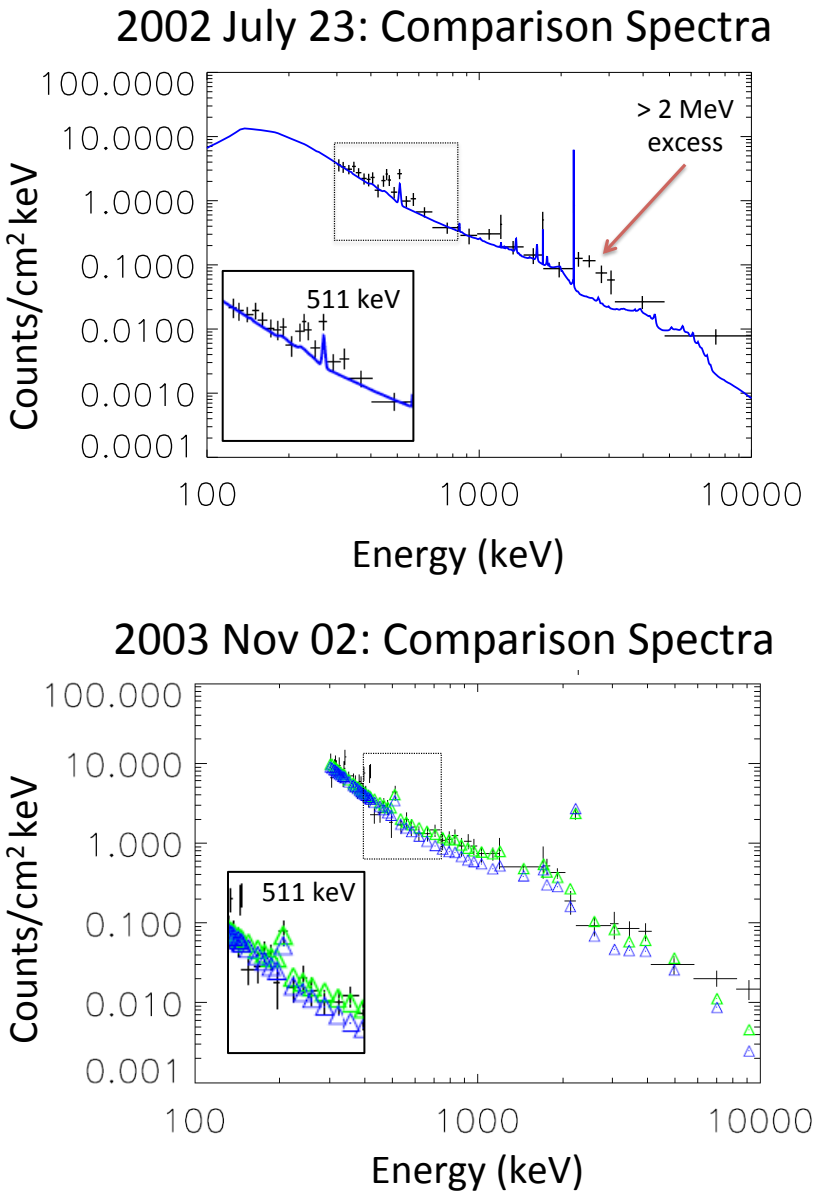

Figure 5.9: Spectra for the 2002 July 23 X4.8 flare (*top*) and 2003 November 02 X8 flare (*bottom*). Imaging-spectroscopy results are black crosses that denote spectral binning and measurement error. The background-subtracted spectra are shown in blue. The 2003 November 02 background-subtracted spectrum has been re-binned to match the imaging-spectroscopy binning. The 2003 November 02 event's summed model components are shown as green triangles.

photon accumulation or if it is an artifact of the method that manifests only for certain flare/grid parameters.

The count flux generated by imaging-spectroscopy is derived from the maximum pixel in a backprojection map. Figure 5.10 (right) shows the maximum pixel value for each energy bin in the 2002 July 23 imaged spectra. The >2 MeV count excess is present in the backprojected map itself, verifying that the excess is not a result of how the map is analyzed. This result suggests that these photons were modulated by the grids, implying that they originated from a compact source in the instrument FOV. The locations of the centroid in each image are consistent with the flare location, further suggesting that the excess counts come from a compact source at the flare site.

The most plausible source of real photons causing an apparent excess above a strong line in times of high-flux is pulse pile-up. When two photons deposit their energy in a RHESSI detector at the same time there is no way to determine that two different photon events occurred, and their combined energy is read-out as a single count event. This phenomena is a well-known consequence of high-flux X-ray events. To determine if the excess counts result from pile-up of 300 keV – 500 keV photons with 2.2 MeV photons the flare counts were compared in two 3-minute intervals, shown in figure 5.10 (left). The two intervals are shown in black and green with the same energy binning. The HXR flux decreases between the first and second intervals, but the >2 MeV flux remains constant. These results are not consistent with pile-up because the count excess was still present after the lower-energy flux dropped off.

To check that these excess counts were not a result of detector 9's response, an imaged spectrum was generated using detector 6. The same count excess appears in detector 6's imaged spectrum above the 2.2 MeV line. This suggests that the excess counts are not a simple artifact of the modulation patterns or modulation amplitudes used for grid 6 or grid 9 during image reconstruction.

The >2 MeV count excess is seen in two different grids and spanning two time periods. It is not observed in both comparison flares and its properties are consistent with real source counts but not pile-up. This excess flux inconsistency is still unresolved. Since the relativistic electron bremsstrahlung fit to the continuum is mostly anchored by the high-flux low-energy spectral components, and these fluxes are in agreement between the two methods, the count excess is inconsequential in deriving the relativistic electron bremsstrahlung model component.

Broad Feature at ∼450 keV

The 2002 July 23 flare's imaging-spectroscopy result shows a broad line at ∼450 keV which does not have a component of the same magnitude in the background-subtracted spectrum. This line could possibly be the gamma-ray line complex of Be-7 and Li-7, which are produced by flare accelerated α particles when they fuse with He-4 in the chromosphere. Share et. al [118] reported finding evidence of these lines during the 2002 July 23 event in the background-subtracted spectral data. They show that a downward oriented beam of α particles would

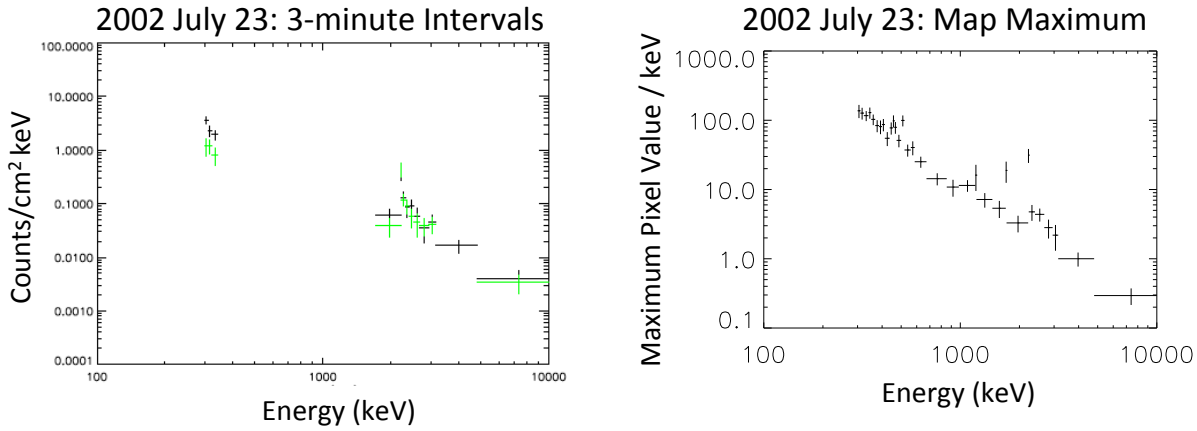

Figure 5.10: Plots supporting the analysis of >2 MeV flux for imaging-spectroscopy of the 2002 July 23 flare. *Left:* Flare spectra integrated over two 3-minute intervals (00:20:00 – 00:32:00 UT & 00:33:00 – 00:36:00 UT) to examine if the >2 MeV count excess results from pile-up with lower-energy 300 keV – 500 keV photons. Since the flux remains constant in the >2 MeV bins but decreases for the lower-energy photons between intervals it is unlikely that pile-up is causing the excess counts. *Right*: The maximum pixel value for each backprojection map. Since the excess counts appear in the backprojected map itself they are not a result of how the map is interpreted.

result in a broad complex which peaks at 450 keV. Their model fit to the background-subtracted data is shown in figure 5.11 (left).

Figure 5.11 (right) shows that the 450 keV line complex is much weaker than the 511 keV line in background-subtracted spectra. In the imaged-spectra, however, the 450 keV feature is much stronger and of roughly comparable magnitude to the 511 keV flux. Though the component reported by Share et al. is well aligned to the feature in the imaged spectrum, its unclear why there would exist such a large discrepancy between the line strengths generated by imaging and background-subtracted spectroscopy.

511 keV

A close-up showing the 511 keV line is inset to each of the comparison plots (figure 5.9). The 511 keV line flux for the 2003 November 02 event is well matched between the two methods. There is a discrepancy, however, between the two methods for the 2002 July 23 event's 511 keV flux. This result could indicate that RHESSI's response matrix is overestimating the amount of locally generated 511 keV for this event, but more study is needed to fully quantify the discrepancy. A next step of this work is to use imaging-spectroscopy to study multiple 511 keV lines to experimentally investigate the background model for this line.

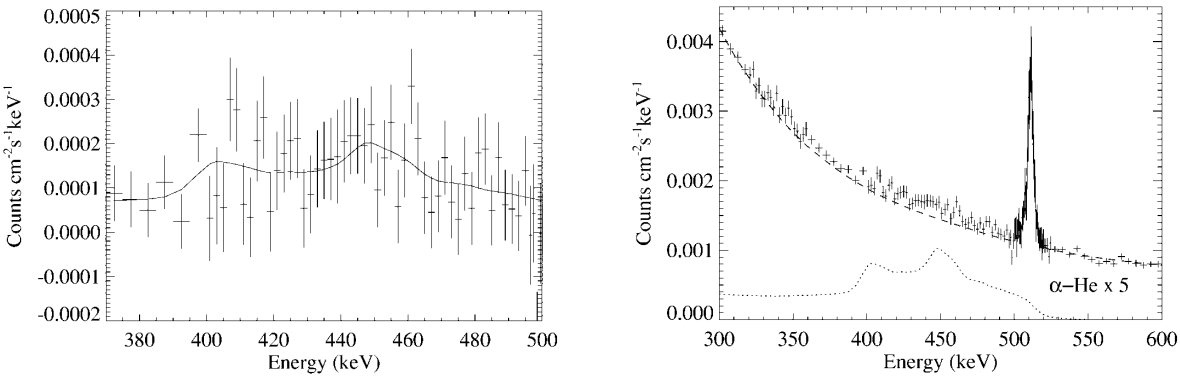

Figure 5.11: 2002 July 23 flare plots from Share et al. 2003 [118] showing the broad spectral feature that peaks at ∼450 keV, and can be attributed to α – He-4 fusion products Be-7 and Li-7. *Left:* Model fit to background-subtracted spectrum. *Right:* Expanded energy range showing the 511 keV line. The dashed line shows the Bremsstrahlung component and the dotted line is the Be-7 and Li-7 line complex model magnified by 5x.

Summary

The imaging-spectroscopy method reproduces the general shape and key features of the gamma-ray flare spectrum well. The reported discrepancies between the >2 MeV flux, the broad feature at ∼450 keV and the 511 keV line flux warrant further study and provide an interesting avenue to constrain RHESSI's response matrix. Since the photon spectrum generated by relativistic electrons is primarily anchored by the high flux 300 keV – 500 kev energy range, and the 2.2 MeV line is well defined with high statistics, the discrepancies outlined above were not found to significantly affect these two parameter fits.

5.6 2003 October 29 Imaging-Spectroscopy Result

The method of imaging-spectroscopy outlined in section 5.4 was applied to derive a spectrum for the 2003 October 29 flare. In figure 5.12 (top) the spectrum is shown as black crosses whose width indicates the spectral binning and whose height denotes error. The spectrum generated by imaging-spectroscopy clearly shows the prominent features of a typical gamma-ray flare: positron annihilation line at 511 keV, neutron capture line at 2.2 MeV and an overall broken-powerlaw spectrum as well as the two escape peaks resulting from secondary photon escape from the detector.

The spectrum was fit with templates in OSPEX and the resulting summed model components are plotted as green triangles in figure 5.12 (top). There is good agreement between the spectrum and model fit, especially at the lower-energies which anchor the electron relativistic bremsstrahlung component. At higher-energies the fitted spectrum predicts slightly lower fluences than derived from imaging. This could possibly be due to pion emission, which

was observed in the 203 October 28 flare, but was not included in the fit shown here. The key features of the model fit are given below:

>300 keV relativistic electron bremsstrahlung fluence: 5788 ± 322 photons/cm^2
2.2 MeV line fluence: 926 ± 180 photons/cm^2
Broken powerlaw at: 661 keV

These fit values were used to compare the 2003 October 29 event with other large flares. The derived relativistic electron and neutron-capture fluences were used to place the 2003 October 29 flare on Shih et al.'s correlation plot [121]. This X10 flare is added as a green circle in figure 5.12 (bottom). The imaging-spectroscopy results indicate that the 2003 October 29 event is one of the strongest 2.2 MeV flares observed in the past 2 solar cycles and that it fits into population of other large flares well.

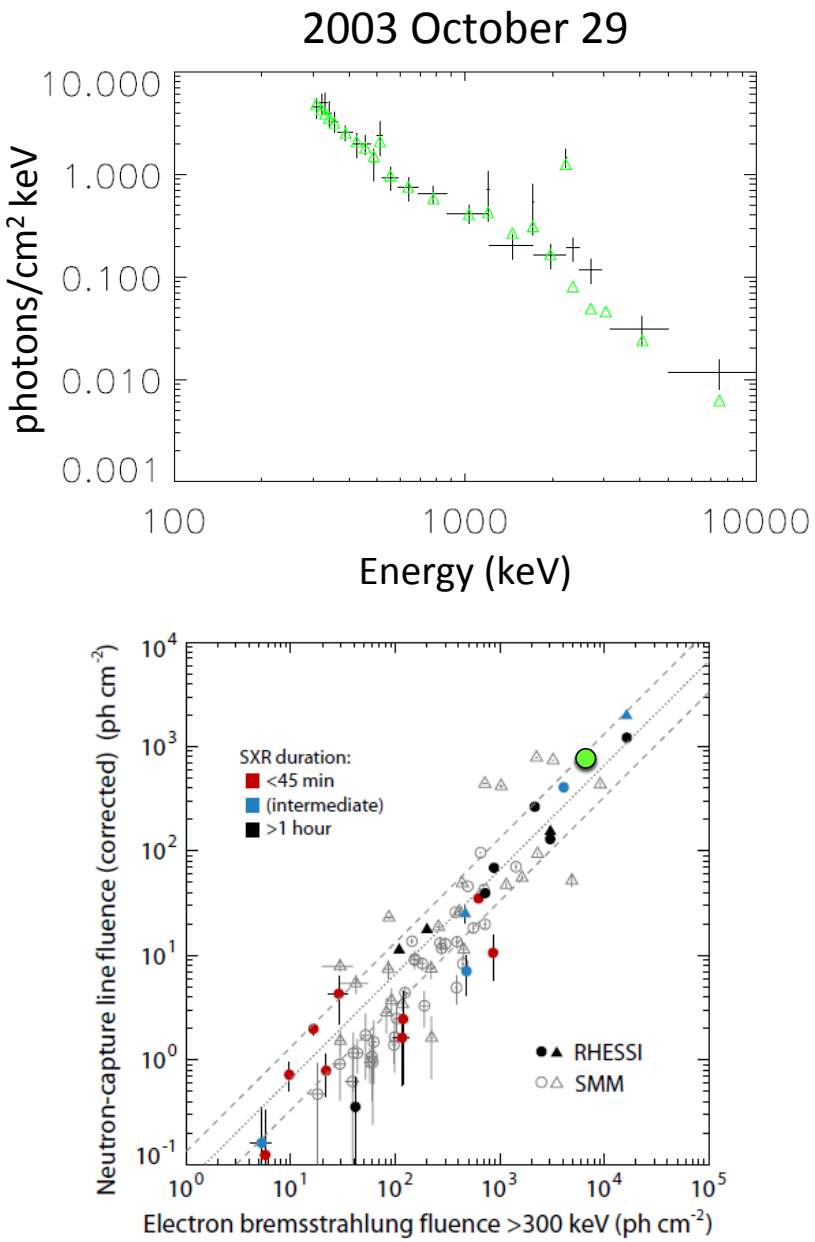

Figure 5.12: *Top:* 2003 October 29 flare spectrum (black crosses) derived via imaging-spectroscopy over the time range of 20:40:00 – 21:07:00 UT. The spectral resolution is denoted by the width of the black crosses and the height is measurement error. The results of a spectral component fit are shown as green triangles. *Bottom:* Relativistic electron bremsstrahlung and 2.2 MeV line fluences correlation plot from Shih et al., 2009 [121]. The 2003 October 29 flare is added to the original plot as a green circle, showing that this flare fits into the tight linear correlation.

Chapter 6

The Gamma-Ray Imager/Polarimeter for Solar Flares (GRIPS) Instrument

The Gamma-Ray Imager/Polarimeter for Solar flares (GRIPS) instrument is designed to study particle acceleration and transport by providing high-resolution imaging (12.5 – 162 arcseconds), spectroscopy and polarimetry of high-energy emission ($\sim$50 kev to $\gtrsim$10 MeV) during solar flares. GRIPS is a NASA Heliophysics Low Cost Access to Space (LCAS) mission led by Pascal Saint-Hilaire, formerly led by Bob Lin, at UC Berkeley's Space Science Laboratory (SSL). Most of the telescope and its support systems were designed, built and tested at SSL with contributions from Lawrence Berkeley National Laboratory (LBNL), NASA's Goddard Space Flight Center (GSFC), and UC Santa Cruz.

The GRIPS payload (Figure 6.1) flew on a long-duration balloon mission from Antarctica in January of 2016. Over GRIPS' $\sim$12-day flight, 21 C-class flares occurred and all systems performed as designed. Due to the remoteness of the landing site ($\sim$500 miles from the Amundsen-Scott South Pole Station) and the lateness within the austral summer season, recovery resources were limited. All flight data for the primary instrument and its three piggyback instruments were recovered in February 2016, but the instrument itself remained in the Antarctic deep field over the 2016 winter. In the following 2016/2017 austral summer season all of the GRIPS flight hardware was recovered.

6.1 Science Goals

Solar flares can release up to 10^{33} ergs within 100s–1000s of seconds [48]. These events impart up to tens of percent of the released energy into particles, accelerating electrons to 100s of MeV and ions to 10s of GeV. Large-scale field reconfigurations resulting from magnetic reconnection in the corona are thought to power flares, but the precise mechanisms that convert the stored magnetic energy into particle kinetic energy are poorly understood. High-resolution imaging and spectroscopy of the hard X-ray/gamma-ray energy range during flares is required to build a complete picture of these processes and understand the energetics

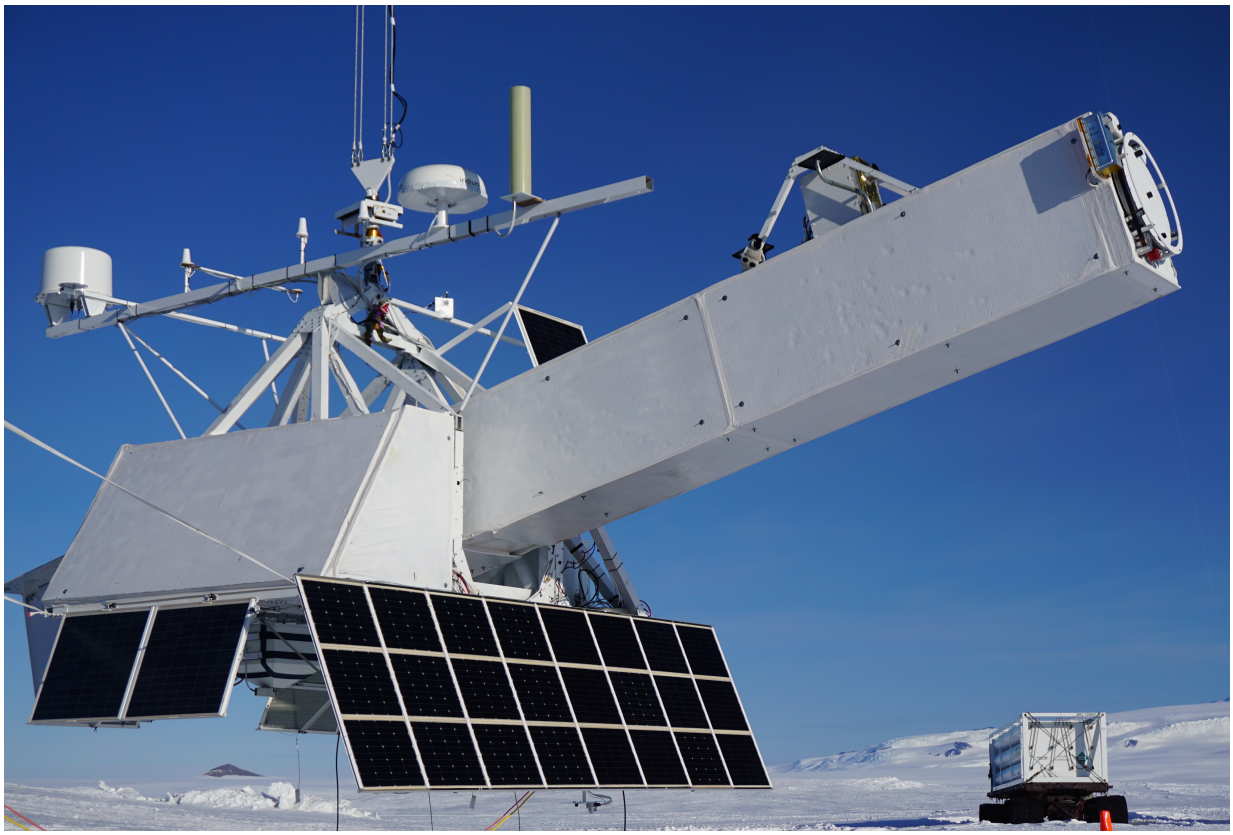

Figure 6.1: The fully integrated GRIPS instrument payload in December 2016 at the Long Duration Balloon (LDB) Camp on the Ross Ice Shelf outside McMurdo, Antarctica. The gondola is suspended from "The Boss" launch vehicle while the GRIPS and Columbia Scientific Ballooning Facility (CSBF) teams complete pre-flight checklists during a launch attempt. The imaging mask can be seen to the right in the photo, on the end of an 8-m boom. Below the payload hangs arrays of solar panels and a telemetry boom is mounted above the payload.

of solar eruptive events.

Observations from RHESSI linked high-energy photon emission to observed spatial structures like magnetic fields, verifying the overall geometry of the standard flare scenario. In the RHESSI and Solar Maximum Mission (SMM) [140] observations, ion and relativistic electron fluences were shown to be correlated, showing that flares accelerate ions and relativistic electrons proportionally and suggesting that these populations are accelerated together, possibly by similar mechanisms (figure 2.24, left) [121]. Since gamma-ray imaging has become available with the RHESSI instrument, there have been only five flares which had enough 2.2 MeV emission to image the flare footpoints. These flares showed that the centroids of ion and relativistic electron emission can be displaced from one another by $\geq$10,000 km (figure 2.25) [66, 67]. This result is surprising; in the standard flare scenario ions and electrons

are accelerated together and transported out of the acceleration region along the same field lines where they would deposit their energy in the same place within the chromosphere. The correlation results support this scenario, showing that the ion and relativistic electron acceleration is linked. To address these seemingly inconsistent results new observations from instruments with finer imaging and increased sensitivity at gamma-ray energies are required. GRIPS was designed to provide these observations and address the new set of high-energy flare questions generated by RHESSI and SMM, including: What causes the spatial separation between HXR emission generated by relativistic electrons and the gamma-ray line emission from energetic ions? How anisotropic are the relativistic electrons, and why can they dominate in the corona? How do the compositions of accelerated and ambient material vary with space and time, and why?

6.2 Instrument Overview

The GRIPS telescope consists of a germanium spectrometer/polarimeter and a single-grid imager separated by an 8 m boom. GRIPS' key technological advances over the current HXR/gamma-ray solar state of the art (RHESSI) are 3D position-sensitive germanium detectors (3D-GeDs) and a single-grid indirect imaging system, the multi-pitch rotating modulator (MPRM). Figure 6.2 shows a schematic of the complete payload and its systems. Separate chapters provide detailed descriptions of the MPRPM imaging system (chapter 9), the spectrometer (chapter 7) and its performance (chapter 8). GRIPS' first flight details and instrument descriptions can be found in literature [122, 44, 45].

The GRIPS spectrometer (pictured in figures 6.3 and 7.1) is able to resolve the site of individual photon interactions in three dimensions. When photons Compton scatter in the detector crystals, which is the dominant process in the range of ∼150 keV to ∼8 MeV, GRIPS can resolve each energy deposition site and reconstruct the scatter path. Determining the scattering track enables polarization measurements and coarse Compton imaging (∼1° resolution) that can be used to image extended sources or eliminate background.

Detector position sensitivity is the enabling technology for GRIPS' MPRM single-grid imaging system. The site of the photon interaction is used to determine the modulation pattern and project it onto the Sun. A single grid can provide ∼2× greater photon throughput than related bi-grid designs, such as RHESSI, which had detectors with no position sensitivity.

GRIPS' MPRM mask design, shown in figure 6.5 (right), is designed with significantly improved image fidelity and nearly 3× higher imaging resolution than RHESSI at gamma-ray energies. The MRPM grid rotates through 360° orientation at ∼10 rpm, providing quasi-continuous imaging resolution at 13 spatial resolutions between 12.5" – 162". At gamma-ray energies RHESSI is limited to imaging with two of its nine grids (numbers 6 & 9) because the remaining grids do not have adequate stopping power to modulate gamma-rays. Limited to two grids RHESSI can only provide imaging at the comparatively coarse angular resolutions of 35" and 182". The significant gains to image fidelity with GRIPS' 11 additional spatial

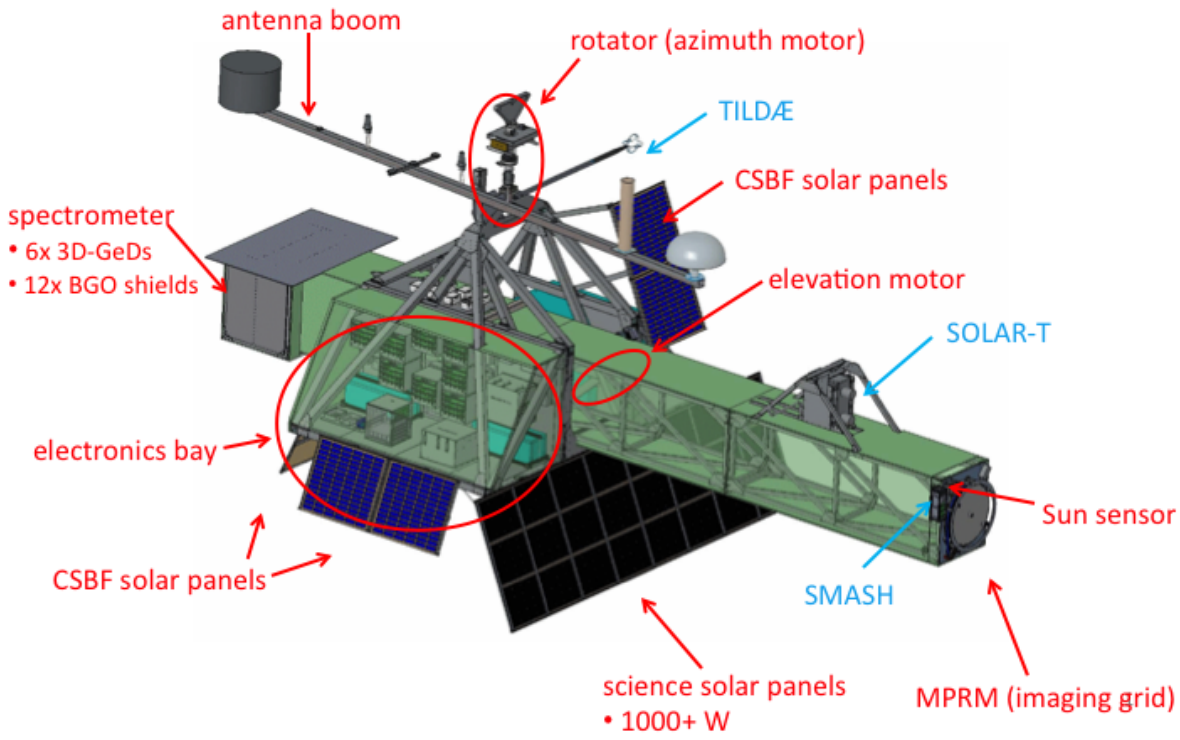

Figure 6.2: Schematic of the fully integrated GRIPS payload. The GRIPS and CSBF components are shown in red and the three piggyback instruments (Solar-T, TILDAE and SMASH) are shown in blue. The flight computer, aspect computer, power distribution unit and the majority of the signal processing electronics are housed in the electronics bay. The electronics bay and telescope boom are both covered in opaque foam, but are shown here as transparent. The solar panel arrays are suspended below the electronics bay floor. Not pictured (on the opposite side to the electronics bay) is a bay which housed the CSBF provided Science Instrument Package (SIP) that interfaced with the antenna boom, providing the communications pipeline between the GRIPS flight computer and ground support equipment. Image from Duncan et al. 2016 [45].



Figure 6.3: The GRIPS spectrometer with BGO shields removed (their outlines marked on bottom plate). The image is taken from the side of the cryostat that will face Earth in flight, and look towards space. The cryostat housing is in the center of the image (gold plated housing) surrounded by the readout ASIC electronics boxes (smaller gold boxes) and their cable harnessing that is wrapped in wire mesh and teflon tape to provide electrical shielding and protection to the ribbon cables. An EMI shield is removed for the detector stack to the left, showing the density of signal traces that travel ∼25 feet in ribbon cables from the ASIC electronics to the FPGA Card Cages in the electronics bay. Protruding from the cryostat on the upper left are the feedthroughs which provide high voltage to the detectors.

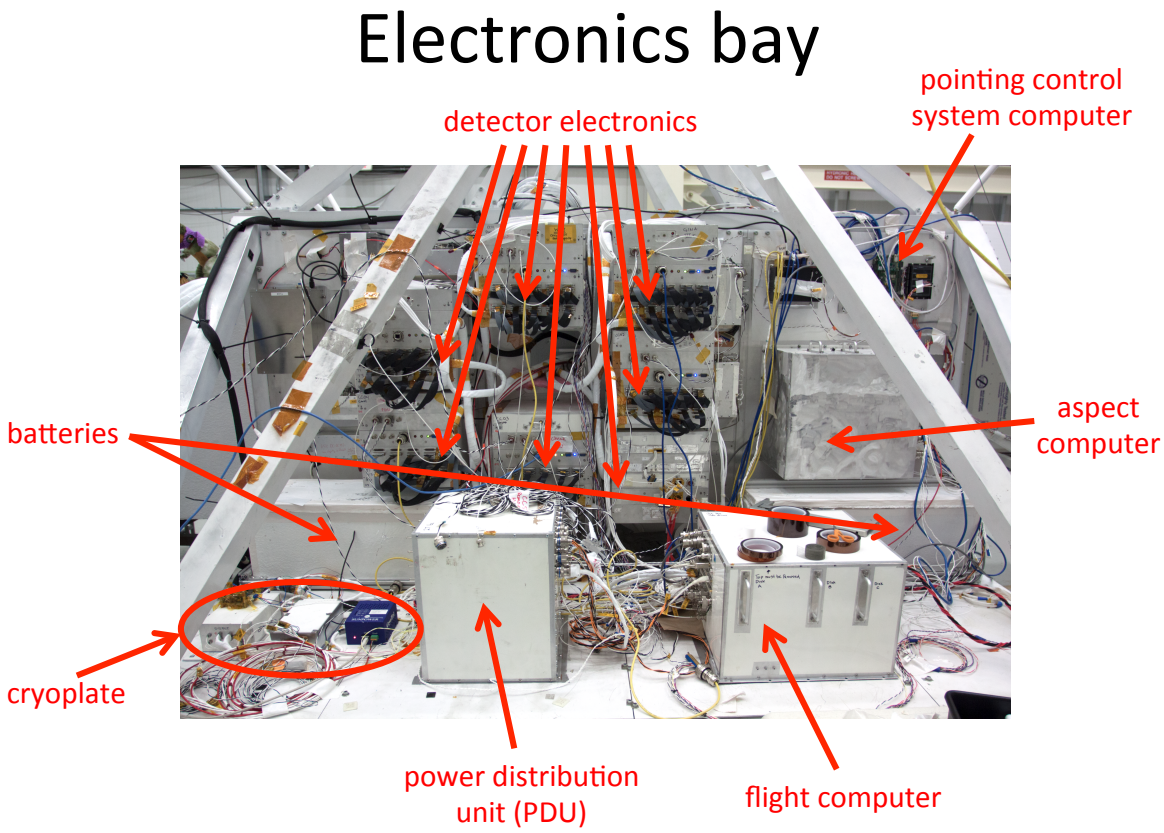

Figure 6.4: Inside the GRIPS electronics bay prior to flight. Labeled are the primary system components, including: the flight, aspect and pointing control computers, the cryostat's temperature control board (cryoplate), the detector electronics, the power distribution unit and battery banks.

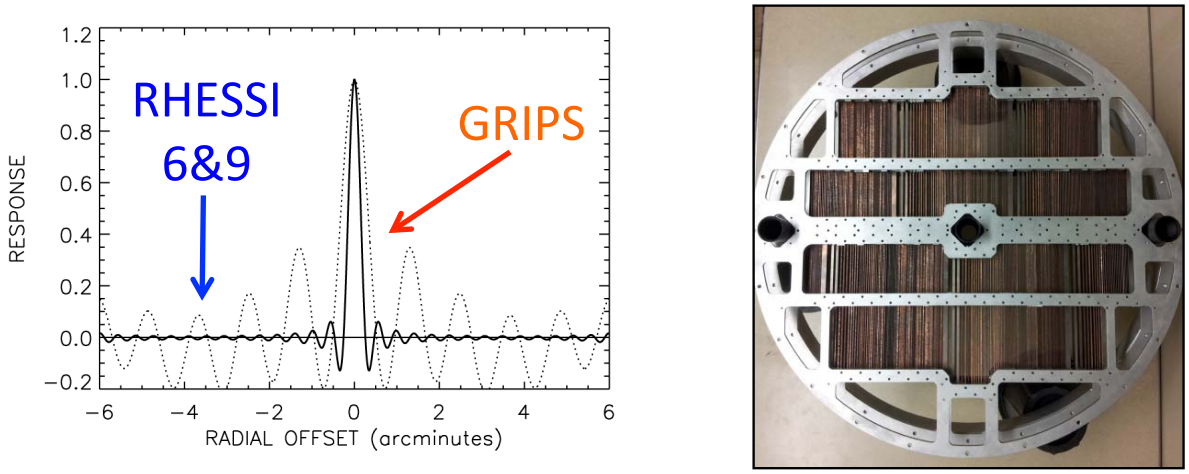

Figure 6.5: *Left:* The point response functions that compare the image fidelity resulting from GRIPS' 13 spatial frequencies to RHESSI's two frequencies (from grids 6 & 9) at gamma-ray energies. Note that GRIPS' response is virtually free of sidelobes. *Right:* The single MPRM imaging grid that contains 13 grid pitches for quasi-continuous sampling between 12.5" – 162". Mounted along the grid diameter are three optical lenses that project images of the Sun onto a fiducial plane as part of the aspect system that provides high-precision monitoring of the grid location and orientation (section 9.3). Images from Duncan et al. 2016 [45].

frequencies is shown in figure 6.5 (left). This plot compares the gamma-ray point response functions of RHESSI and GRIPS. GRIPS' response (solid curve) is compact and virtually free of the large sidelobes in RHESSI's response (dashed curve).

The single grid MPRM design avoids the tight engineering and manufacturing constraints typical of bi-grid designs. These RMC-type imagers rely on precision grid-to-grid alignment in order to use a detector with little or no spatial sensitivity. For GRIPS the grid is only required to cast a shadow on the detectors, the 8 m boom is free to twist and flex during flight. The exact location of the mask relative to the detectors is closely monitored by the imager aspect system (section 9.3). This position and the well-known grid metrology is used in image reconstruction. The aspect system provides high-precision monitoring of telescope pitch, yaw and roll. This enables arcsecond-precision aspect knowledge with a boom pointed to only 0.5° accuracy.

Project Timeline

GRIPS was funded through two NASA Low Cost Access to Space (LCAS) program proposals in 2008 and 2012 under NASA's Heliophysics division. The instrument was primarily designed and built at UC Berkeley's Space Sciences Laboratory (SSL) in collaboration with NASA's Goddard Spaceflight Center (GFSC), Lawrence Berkeley National Laboratory

(LBNL) who provided the detectors and UC Santa Cruz who worked on the initial ASIC boards. Due to decreasing solar cycle activity and the scientific ballooning flight schedule, the typical continental test flight required for off-continent long-duration balloon flights was replaced with a full-scale thermal vacuum chamber test at NASA Glenn Research Center's Plum Brook facility (section 6.5). Table 6.1 shows a timeline of GRIPS' major tests and milestones.

Dates	Milestones
2008	1st proposal selected
2012	2nd follow-on proposal funded
2015 Mar	Plum Brook T-vac Test
2015 Jul – Aug	CSBF/NASA Systems Integration
2015 Oct 31 – Dec 25	Pre-flight Integration
2015 Dec 26 – 2016 Jan 19	7 aborted launch attempts
2016 Jan 19 – 30	Antarctic LDB Flight
2016 Feb 2	Data Vault Recovery
2016 Nov 24 - 2017 Jan 16	Full Instrument Recovery

Table 6.1: GRIPS instrument timeline showing the major project milestones and tests.

6.3 Antarctic Flight Campaign

The GRIPS payload underwent a period of integration, pointing control, aspect system testing, and instrument calibrations beginning on 2015 October 31. The instrument final integration, test and launch occurred at NASA's Columbia Scientific Balloon Facility (CSBF) Long Duration Balloon (LDB) camp on the Ross Ice Shelf near McMurdo Station, Antarctica. The majority of GRIPS' hardware was transported via shipping container to Antarctica, but the more delicate spectrometer was flown by the United States Air National Guard (ANG) to McMurdo, stopping at several locations en route. Due to logistical complications with requiring access to power to keep the cryostat cold, the cryostat was instead transported warm, at the expense of extending the period of integration in Antarctica by $\sim$2 weeks to cool the cryostat back down to $<$80 K.

GRIPS was declared flight ready on 2015 December 26 after a pre-flight compatibility test. Over the following weeks, surface weather conditions needed for launch were scarce; in particular, even on clear days, strong surface winds exceeded the required 5 knot upper threshold for launch. During this time the payload had seven launch attempts (*roll-outs*) that were aborted before successfully launching on 2016 January 19 at 01:41 UT. Each roll-out consisted of a gondola pickup by CSBF's launch vehicle *The Boss*, followed by a full communications and instrument checkout. During several roll-outs, the payload was

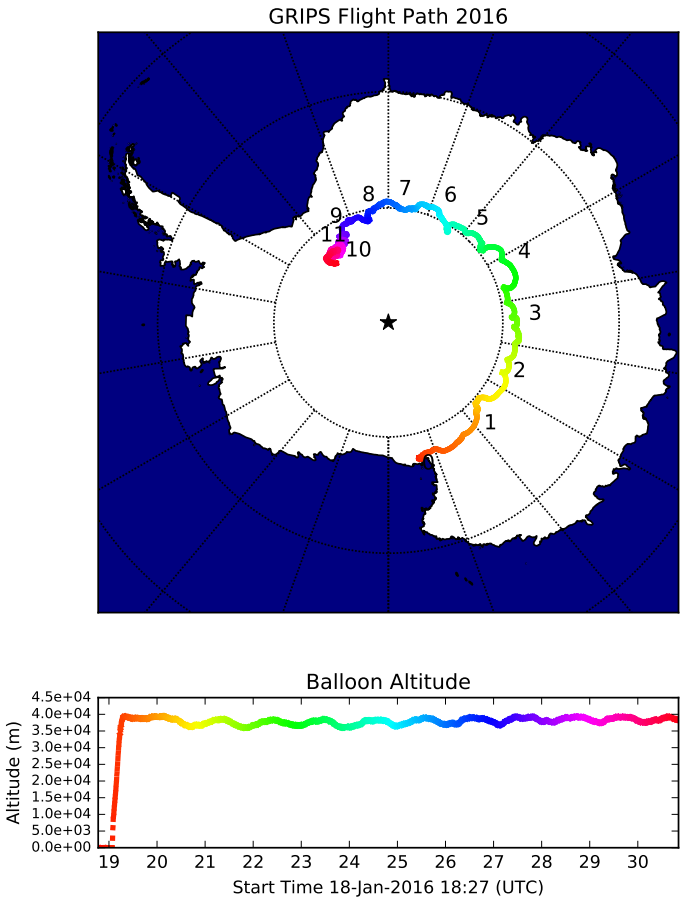

Figure 6.6: *Top:* GRIPS flight path with time represented by the color gradient progressing from red, at launch, to blue on landing. Numerical annotation indicates the number of days at float. *Bottom:* Balloon altitude from launch with the color progression indicating time at float. Images from Duncan et al. 2016 [45].

subjected to colder temperatures than it would encounter during flight, including freezing fog and light snow.

After ∼5 hours of ascent, the balloon reached a float altitude of ∼39,000 m (bottom of figure 6.6). The balloon altitude varied diurnally in flight, reaching a maximum altitude of 39,459 m at 03:06 UT on January 20 and a minimum altitude of 35,873 m at 20:08 UT on January 21. Figure 6.6 (top) shows the balloon flight path around the Antarctic

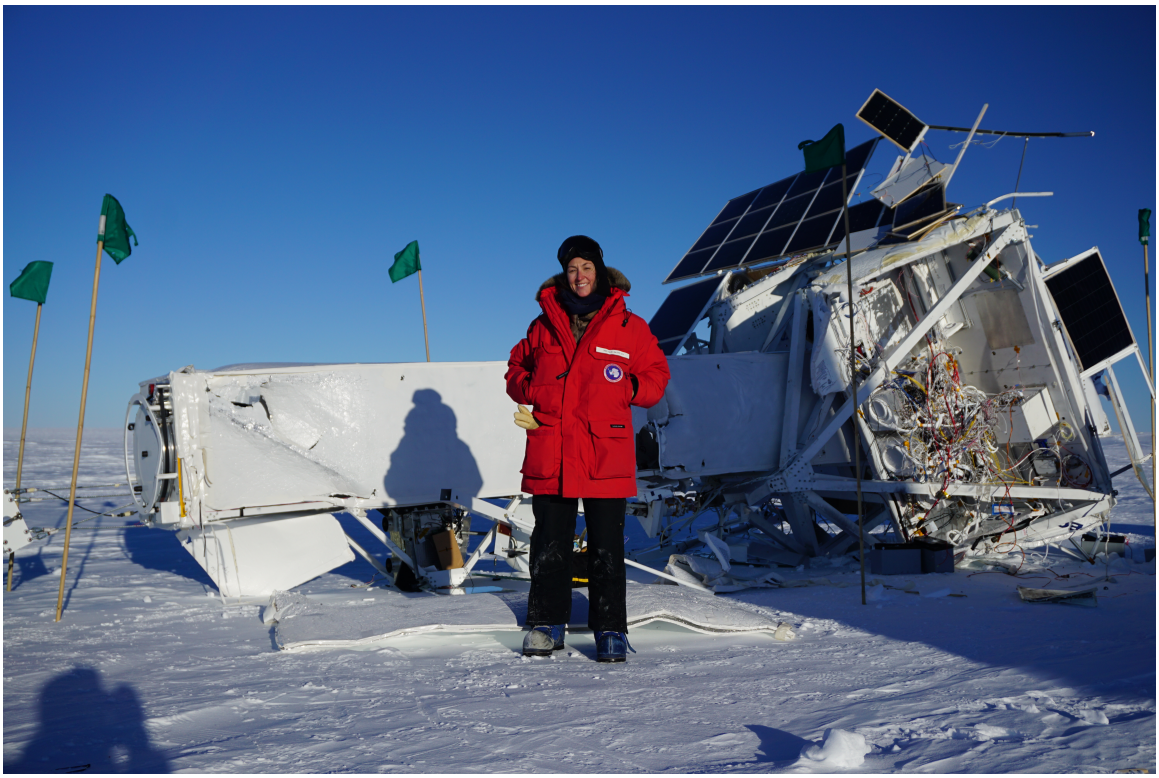

Figure 6.7: The author with the GRIPS payload in February 2016 after flight. Damage to the spectrometer's protective cage and bottom radiator plates suggest that the boom was not horizontal during descent and the spectrometer end impacted first before the payload flipped over to its final resting place.

continent, with days at float annotated and represented by the color gradient (red at launch, blue/purple on landing).

During flight all of GRIPS' instrument and support systems performed nominally. After nearly 12 days at float, preparations were made for landing. Detector data acquisition ended at 19:51 UT on 2016 January 30, with flight termination occurring shortly after, at 20:51 UT. The balloon landed at 21:30 UT in the Queen Maud Land region of the Antarctic continent, at 83°6.18 S, 40°54.08 W. Structural damage to the spectrometer's bottom radiator plates suggest that the boom was not horizontal upon landing. The elevation motor, used to keep the boom horizontal during descent, likely broke during the ∼10 G chute shock, causing the much heavier spectrometer end to fall below the electronics and SIP bay floors. After initial impact the gondola flipped over into its final resting position, shown in figure 6.7. The aluminum gondola was designed for the possibility of rolling and it sustained the most damage, protecting the GRIPS and piggyback instrument systems as intended.

Due to landing occurring late in the 2015/2016 Antarctic summer season and the remoteness of the landing site (∼500 miles from the Amundsen-Scott South Pole Station),

only the data vaults were initially recovered. The remainder of the instrument remained in the field over the Antarctic winter and preparations were made for a full instrument recovery in the following austral 2016/2017 summer. A 3-week field camp, including grooming an ice runway to accommodate a Basler aircraft, was planned at the GRIPS landing site to fully remove all ∼4,000 lbs of flight hardware. After early-season damage to the Basler and fuel shortages from reduced Air National Guard flights, the recovery effort was scaled back to 3 out-and-back day flights on-board a Twin Otter plane. The flight distance between the South Pole station, where recovery efforts were staged, and the landing site exceeded the Twin Otter's range requiring refueling a a fuel cache enroute. After five weeks of weather delays, the GRIPS recovery team was able to make three trips to the gondola in early January 2017 and recover all the flight hardware, leaving only the gondola's aluminum frame in the field.

6.4 Antarctic Flight Results

Solar activity was relatively low during GRIPS' ∼12 day flight. The Sun produced 21 GOES C-class flares, the largest of which was a C9.6 flare on 2016 January 28 at ∼11:45 UT. Figure 6.8 shows the GOES X-ray lightcurves at 0.5–4.0 Å(blue) and 1.0–8.0 Å(red) over the GRIPS flight. The corresponding GOES flare classifications are noted on the right-hand axis. Also marked are two extended periods of radiation belt electron precipitation events observed in the GRIPS data.

C9.6 Event on 2016 January 28

The largest solar flare that occurred during GRIPS' flight was a C9.6 at ∼11:45 UT on 2016 Jan 28. This flare was simultaneously observed by RHESSI, which had good coverage of the entire flare from onset through impulsive and decay phases. Figure 6.9 shows the counts recorded by three of RHESSI's detectors during the event. The majority of counts come in below 10 keV. The sudden shift to slightly higher-energy photons in the image is artificially created by RHESSI attenuators coming in to reduce the low-energy flux. The maximum energy photons detected by RHESSI for this flare are ∼30 keV.

The majority of GRIPS' detector channels had a low-energy spectral cutoff of ∼50 keV (figure 6.10). This threshold has a finite width and channel-to-channel variability such that some lower-energy counts are present in data, but the effective area drops off very quickly below ∼50 keV. Figure 6.12 shows detector 0 output for the 11:21 UT – 13:06 UT time range, marked as dashed lines are the flare onset at ∼11:45 UT and the end of the impulsive phase at ∼12:31 UT. The bottom plot zooms into the lower energy range, showing that D0 spectral cutoff was too high to observe the <30keV emission reported by RHESSI from this C9.6 flare. Figure 6.11 shows GRIPS' D0 and D4 energy independent count rates over an expanded time period (10:39 UT – 16:24 UT) around the C9.6 flare. The uniformity of detector count rates across the flaring period indicate that GRIPS mostly observed background during this time.

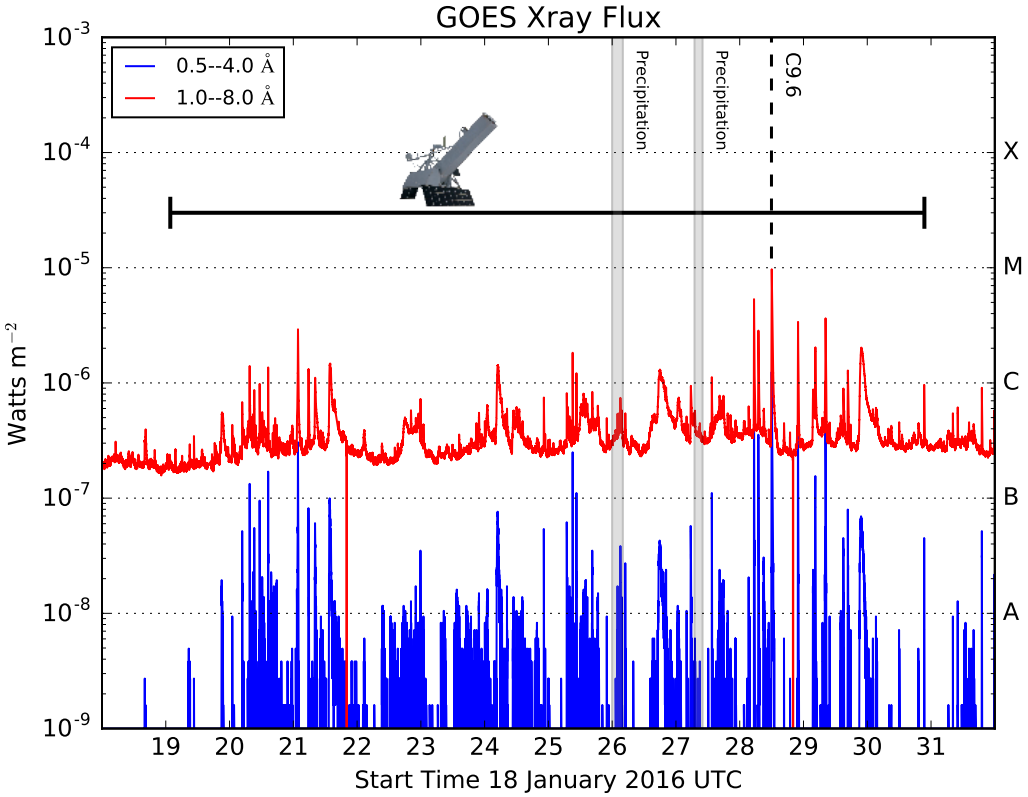

Figure 6.8: GOES X-ray lightcurve for the duration of the GRIPS flight (marked by the horizontal black line). The vertical dashed line indicates the largest solar flare, GOES class C9.6, and the shaded areas indicate periods of precipitation from Earth's radiation belts. Image adapted from Duncan et al. 2016 [45]

The apparent variability in D4's count rate is due to the well-documented duty cycling behavior described in section 8.1. At this time only two of GRIPS' six detectors have the full energy calibration which is required to generate the detector integrated spectra shown in this section. Summing the spectra from the additional four detectors would increase the photon collection area and possibly recover flare counts.

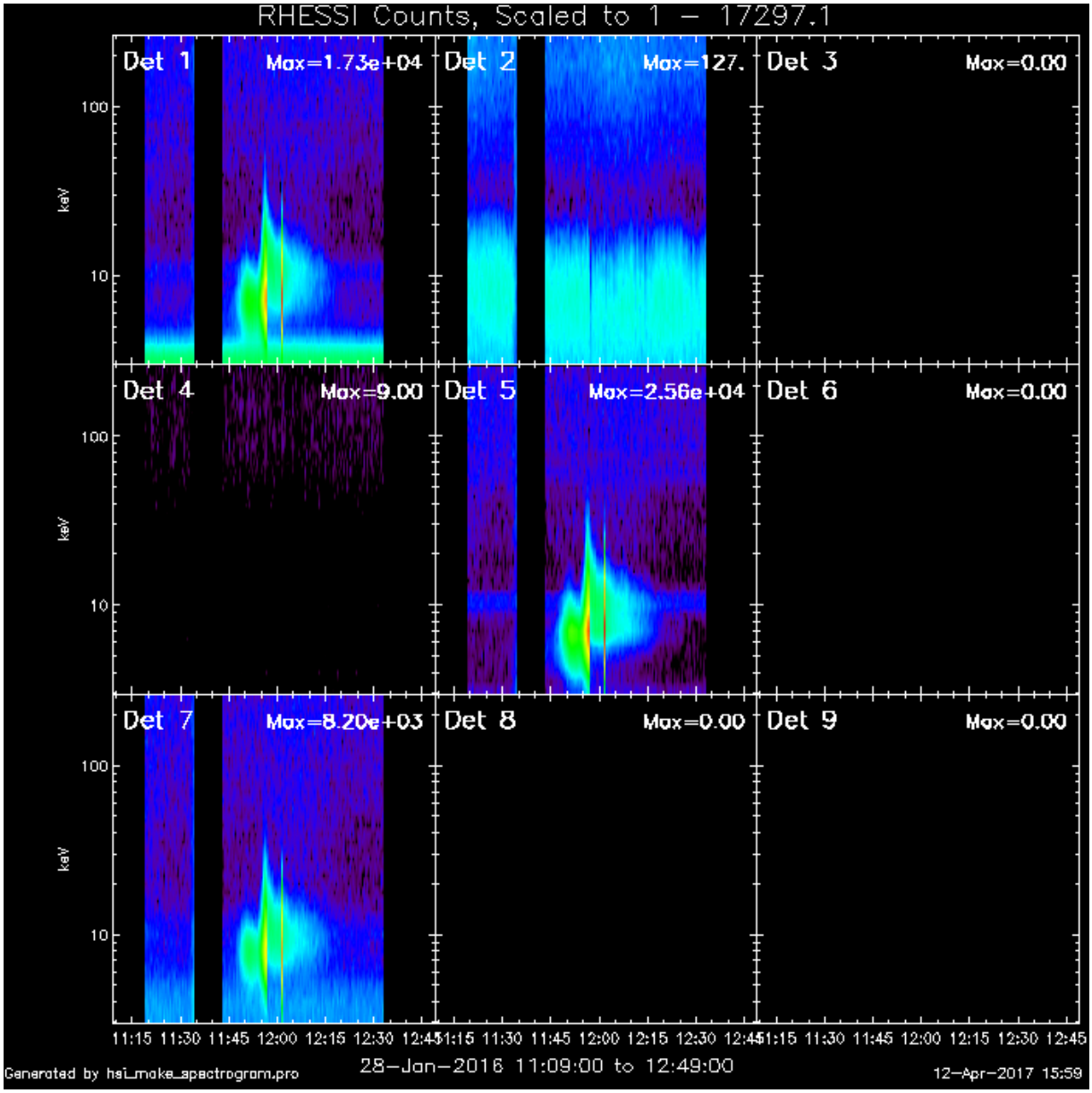

Figure 6.9: Spectrograms for the RHESSI detectors during the C9.6 flare on 2016 Jan 28. Three of RHESSI's detectors (D1, D5 and D7) observed the flare. The spectra show that the majority of counts are below 10 keV. At ∼11:55 UT an attenuator comes in to reduce the low-energy flux. The maximum energy photons detected by RHESSI for this flare are ∼30 keV. Plots provided by RHESSI Browser [123].

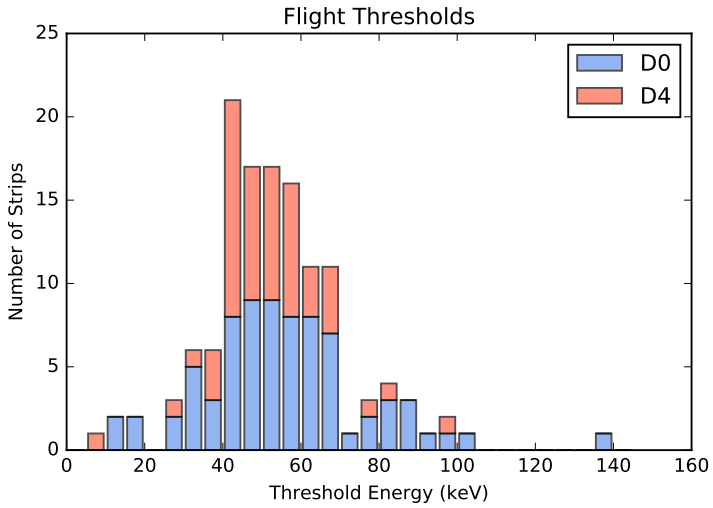

Figure 6.10: Distribution of low-energy thresholds in flight. To generate this plot the flight data was integrated over multiple days and the spectral cutoffs were fit to spectra across detectors 0 and 4. An example flight background spectrum is shown in figure 6.13. The majority of spectral cutoffs for detectors 0 and 4 range between 40 – 60 keV.

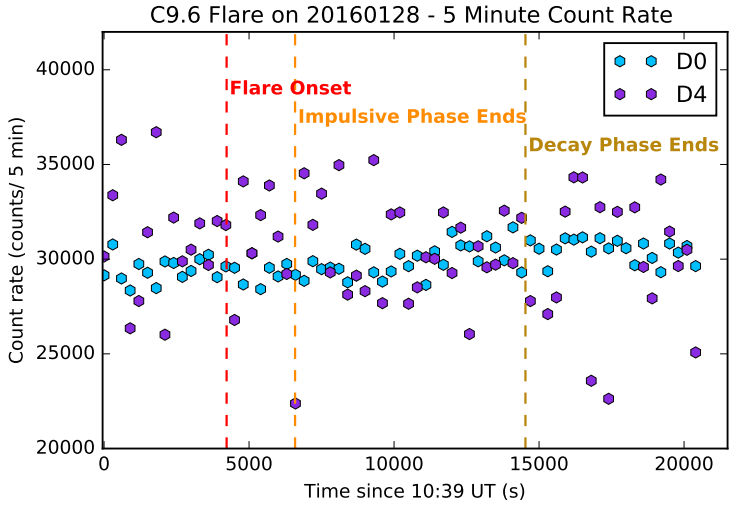

Figure 6.11: Detector 0 and detector 4 count rates integrated over 5 minute bins for the C9.6 flare that occurred on Jan 28, 2016 at ∼11:45 UT. The flare onset and ends of the impulsive and decay phases are marked with dashed lines. No obvious count rate excess occurs in either detector after the flare onset. D4's variability on this timescale comes from the duty cycling behavior discussed in section 8.1. Events with a maximum of two triggered strips per side were selected for this plot.

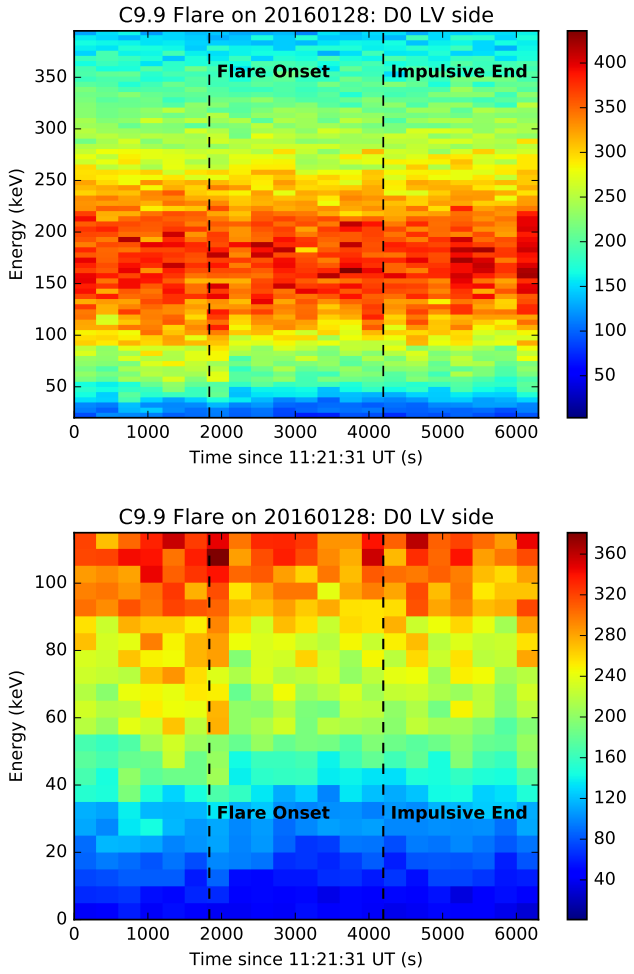

Figure 6.12: Spectrograms spanning 11:21 UT - 13:06 UT for GRIPS detector 0 in 5-minute bins over the C9.6 flare that occurred on 2016 Jan 29 at ∼11:45 UT. The top plot shows that the majority of emission observed is in the 100 keV – 250 keV range (from background) and the bottom plot zooms in on the lower end of the energy portion of the range shown in the RHESSI plots (figure 6.9). These plots show that GRIPS has no obvious spectral changes during the flare impulsive phase in comparison to the pre-flare time period. GRIPS' spectral cutoff was ∼50 keV in flight and RHESSI primarily recorded flare emission <30 keV.

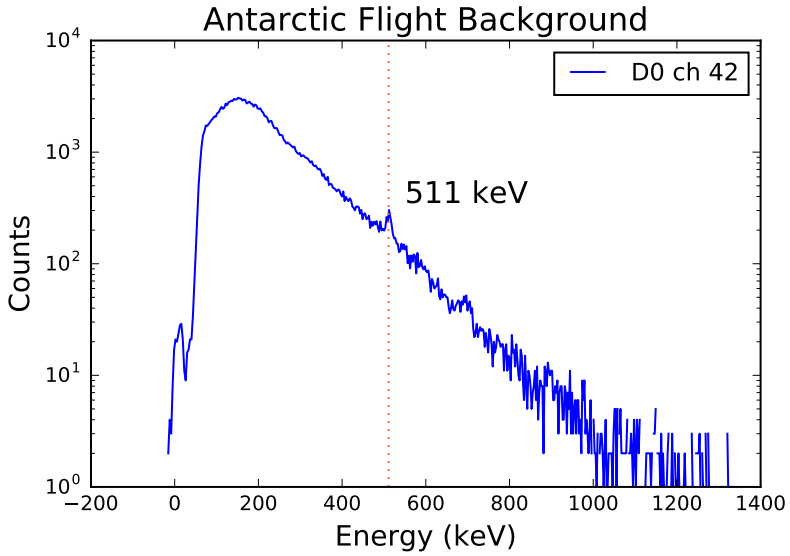

Figure 6.13: Antarctic background for a single strip on detector 0 integrated over multiple days in flight. The steep low-energy spectral cutoff is imposed by triggering thresholds in the electronics. To the far left is the noise pedestal at zero energy. A dashed line marks the predicted 511 keV line which is in good agreement with the peak seen in the background data (blue curve), after applying the ground-based calibration.

Flight Background

Figure 6.13 shows background counts on a single strip integrated over multiple days in flight. The Antarctic background gradually decreases with energy and the sharp cutoff is imposed by the detector electronics at the threshold location. The peak below the spectral cutoff, called the *pedestal*, is due to noise. The pedestal is located at zero energy because no real charge was collected with the noise trigger. Marked with a dashed line is an instrumental 511 keV peak caused by photon interaction with the detector, or, with passive material surrounding the detectors. The good agreement between the known line energy and spectral value gives confidence that our ground calibrations remained valid in flight.

Summary

The largest event that occurred during GRIPS' flight was below the limits of detectability for detectors 0 and 4, using the event selection described chapters 7 & 8. Summing the flare data over all detectors, selecting for channels with lower than average thresholds and improved data selection could yield the ∼100 photons needed to generate a reliable image. The remaining flare events are smaller in magnitude, and are not expected to have flare detections. The two extended periods of relativistic electron precipitation from Earth's radiation belts have positive detections in GRIPS' BGO shields and will be studied with the

spectrometer.

6.5 Thermal Management

Thermal management is a significant challenge to stratospheric ballooning. Direct solar illumination, Antarctic albedo and electronics provide thermal inputs which must be properly balanced with radiative losses to space in order to prevent components from overheating, or, freezing in space. A thermal model was developed for GRIPS to effectively balance the heat loading and losses to keep the system components within their operational ranges. The thermal management plan included conducting heat from hot components then radiating to space via the gondola frame or radiation panels, both of which are painted white for good emissivity. The entire electronics bay and cryostat were covered in foam to provide insulation. Components exterior to the foam insulation, like cameras and motors, were heated with commandable resistive heaters whose thermostats were controlled via software in the flight computer. In March 2015 the GRIPS gondola underwent a 30-hour full instrument thermal vacuum (T-vac) test at NASA Glenn Research Center's Plum Brook facility to simulate the thermal and near-vacuum ($\sim$3 mbar) environment at stratospheric altitudes. The Plum Brook test provided a highly-targeted and low-risk avenue to test the GRIPS thermal model and several other payload systems that are usually validated with comparatively higher-risk continental test flights.

GRIPS' T-vac test was carried out in the Spacecraft Propulsion Research Facility (B-2), which consists of a 33'$\times$120' chamber. The test lasted 30 hours at $\sim$2 torr and 77 K. The gondola was suspended from a specially built truss which allowed for $\pm5^\circ$ pointing maneuvers while in the chamber (Figure 6.14).

To simulate float conditions the B-2 chamber's cold wall and lid were cooled with liquid nitrogen, and the chamber was evacuated to $\sim$2 torr. The Sun's intense point-source radiation striking the front of the payload and the diffuse Earth albedo beneath it were simulated using banks of calibrated IR lamps. Direct solar illumination was simulated with lamps placed (1) in front of the electronics bay and (2) in front of the boom. Earth albedo lamp banks were placed (3) under the electronics bay and (4) under the spectrometer. Heaters were applied to some elements that were impractical to illuminate with IR lamps.

The GRIPS thermal model predictions reasonably matched the readings from the $\sim$80 temperature sensors used during the test. A critical parameter in making accurate predictions was properly accounting for the radiation environment of the B2 chamber's interior, namely that it cannot be modeled as a perfect black body due to degraded paint on the walls and the grated metal floor. Component temperatures in flight were in reasonable agreement with the overall GRIPS thermal model.

Harsh Antarctic ground conditions introduced factors that are not traditionally accounted for in a float model, adding complexity to pre-flight tests and instrument checkouts. In particular, convective cooling is much stronger on the ground than in the rarefied upper stratosphere. On a few roll-outs, the payload experienced freezing fog conditions which

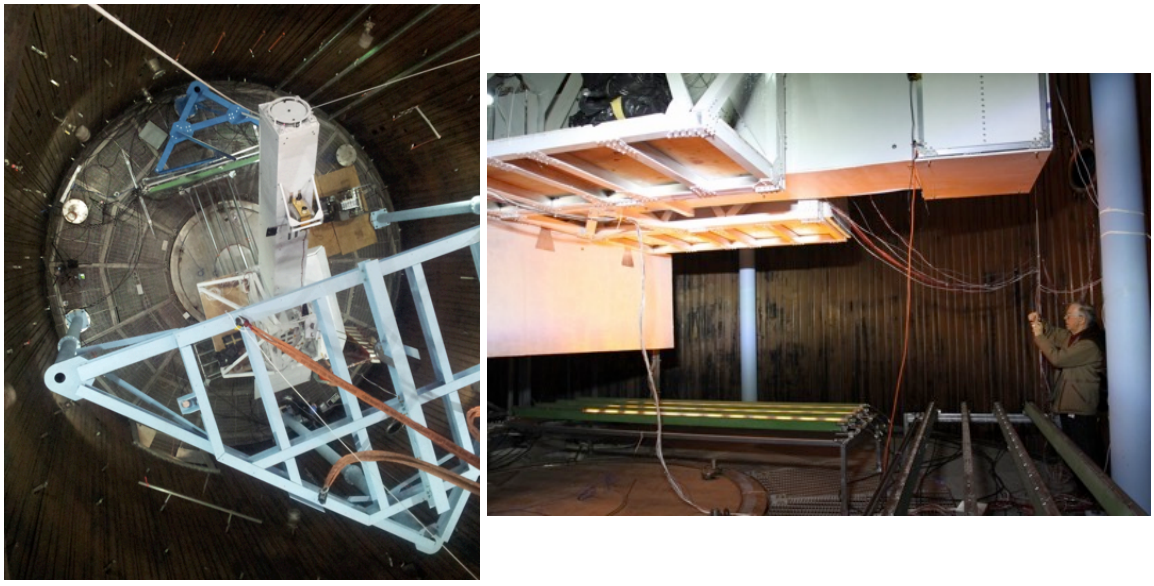

Figure 6.14: *Left:* The GRIPS gondola suspended from a custom-built truss (blue structure) inside NASA Plum Brook's B2 vacuum chamber for the 30 hr thermal vacuum test. *Right:* Banks of IR lamps were used to simulate the Earth's albedo from below and direct solar illumination from the front (not pictured). Pictures from Duncan et al. 2016 [45].

caused components to reach lower temperatures than were observed in flight. To negate these ground-only effects, a space heater was added in the electronics bay and occasionally the electronics for the BGO anti-coincidence shields needed to be shut off. On very sunny days with long hang times it was necessary to rotate the gondola as the sun moved to prevent differential heating on the spectrometer. Since the gondola is on-axis pointed at the Sun the spectrometer was not designed to accommodate prolonged periods of direct solar illumination.

6.6 Future Directions

The GRIPS instrument had a successful Antarctic flight in 2016 and the flight hardware was returned in spring of 2017. The immediate next steps for the GRIPS program is to checkout the flight hardware and assess its viability for re-flight. The state of the hardware will largely determine how the program moves forward. As an alternative to a future re-flight of GRIPS as a long-duration balloon (LDB) payload, GRIPS can be re-designed to be an ultra-long-duration balloon (ULDB) payload, making use of NASA's super-pressure balloon. The super-pressure balloon allows for flights of several months in duration, with the near-certain chance that the payload and its data vaults would not be recovered. Detailed analysis of data from GRIPS's recent flight will guide the development of data-reduction techniques so that all of the science information can be retrieved through telemetry links

alone. Recurring flights of LDB or ULDB GRIPS payloads during the active years of the solar cycle would ensure that many large gamma-ray flares are observed.

A future space-based solar instrument can also make use of GRIPS' key technologies. Given the ongoing progress in deployable booms, one can envision a spacecraft with scaled-up versions of the MPRM and the 3D-GeD spectrometer, separated by 20 meters or more. Such an instrument would have unprecedented sensitivity to solar gamma rays at a few MeV and an angular resolution of better than 5 arcsec, which would resolve the smallest structures expected for accelerated ions in flares. Observations at this scale could provide a complete picture of flare particle energetics.

Chapter 7

The GRIPS Spectrometer

The GRIPS spectrometer consists of six high-purity 3D position-sensitive germanium detectors (3D-GeDs) housed in a single cryostat, located opposite to the imaging grid at the end of an 8 m telescope boom. The spectrometer is surrounded by anti-coincidence shields made from bismuth germanate (BGO) scintillators that provide background rejection. The 3D-GeD detector output is triggered, shaped, sampled and begins digitization at the cryostat by *Application Specific Integrated Circuits* (ASICs) before being routed to the electronics bay to complete digitization and storage. The cryostat keeps the detectors under vacuum and a mechanical cryocooler maintains an interior temperature of <80 K. Heat loading from the cryocooler and electronics is radiated to space by plates surrounding the cryostat shell to the top and sides. Figure 7.1 shows these main spectrometer components with the top radiator plate removed.

7.1 3D-GeD Detectors

GRIPS' 3D-GeDs are high-purity germanium cross-strip detectors that can be used to locate and measure each photon energy deposition in three dimensions. These detectors extend the virtually pixelated detector concept that was successfully demonstrated on the NCT missions in 2005 [36] and 2009 [10] and on the COSI mission in 2016 [72]. The GRIPS detectors are of the same size (7.5 cm×7.5 cm×1.5 cm), but have a 4× finer (0.5 mm) electrode pitch. As with the earlier generations, these detectors were developed and fabricated at Lawrence Berkeley National Laboratory.

Each 3D-GeD is fabricated from a single germanium crystal. The planar crystal surfaces are segmented into orthogonal electrode strips which are used to apply a bias voltage across the crystal, virtually pixelating the detector active volume (figure 7.2). When a photon interacts with the detector medium a cloud of electron/hole pairs are created. These charges drift to opposite detector faces under the applied bias where they are collected, forming the signal which provides a measurement of the energy deposited by the photon interaction. Chapter 3 contains a detailed discussion of photon interaction processes in semi-conductors.

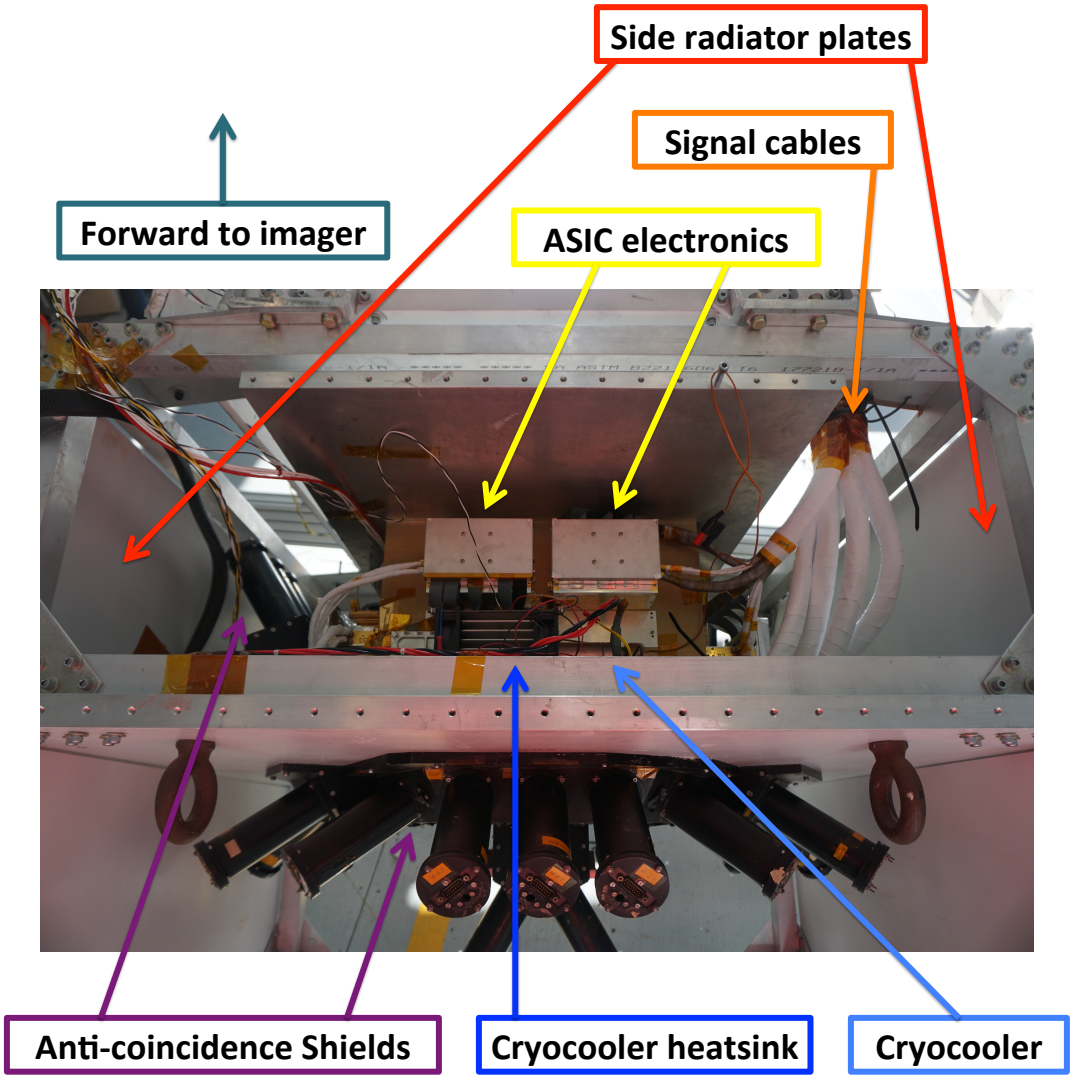

Figure 7.1: The GRIPS spectrometer with top radiator plate removed. The main components of the spectrometer are diagrammed including the cryostat housing, cryocooler, cryocooler heatsink assembly (which includes thermoelectric coolers and a finned heatsink), ASIC electronics boxes, signal cables, BGO anti-coincidence shields, and side radiator plates. The spectrometer is on-axis pointed towards the sun along the telescope boom, in the direction marked "forward to imager". The picture is taken from the side of the cryostat that faces space, looking towards the Earth.

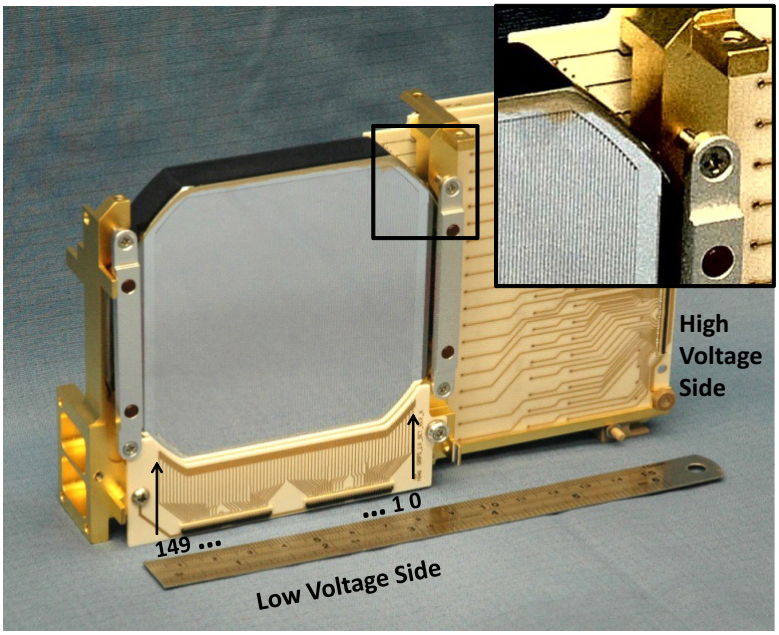

Figure 7.2: A single GRIPS 3D-GeD detector. Inset is a portion of the detector magnified to show the individual parallel electrodes. The detector is wire bonded to breakout boards, which connect to flex circuits that carry the signal through the cryostat shell, to exterior processing electronics. The HV side board is large because it also contains the circuits necessary to provide a steady voltage to that side's electrode strips. The LV boards are comparatively smaller because those strips are held at ground. Numbered are the electrode strips from 1 to 149 on each side, and a guard ring is labeled as 0. The detector edges are intentionally rounded and the guard ring is visible along the edges of the germanium face. Picture from Duncan et al. 2013 [44].

Figure 7.3 depicts how the segmented electrodes and arrival times are used to determine the photon interaction site. Coincident triggers on the orthogonal high voltage (HV) and low voltage (LV) strips locates the interaction site in a 2D plane. Figure 7.4 shows a 2D hitmap that results from collecting many such interactions. Given good knowledge of the electron and hole drift speeds, the difference in their arrival times on the opposite detector faces can be used to infer the photon interaction depth [2, 3].

7.2 Spectrometer Mechanical Structure

The nominal operating temperature for germanium semi-conductor devices is <80 K. Detectors are placed under vacuum then cooled via conduction by mechanical heat-pumps or a passive heat-sink (typically liquid nitrogen). Both methods present challenges such as noise, providing adequate power for a heat-pump or constrained operational lifetime due to a consumable like liquid nitrogen. Over GRIPS' planned 30-day long-duration balloon mission,

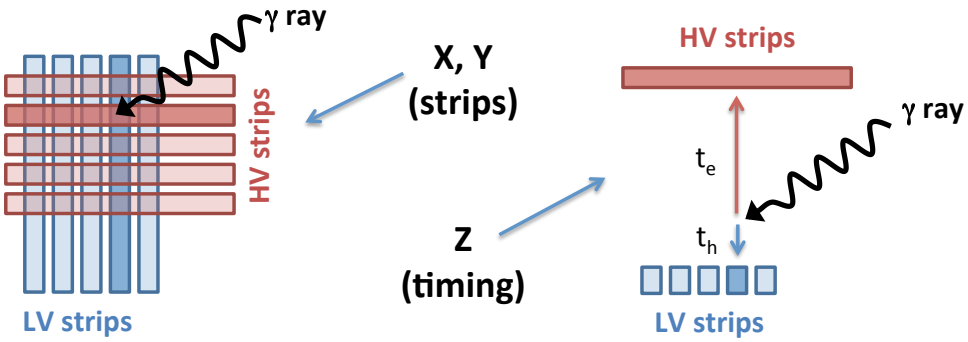

Figure 7.3: When a photon interacts with the detector a cloud of electron/hole pairs are generated. The electrons and holes move towards the electrodes which supply a bias voltage across the crystal, their arrival locations and arrival times can be used to determine the site of photon interaction. *Left:* The 2D planar coordinates (X, Y) are determined by matching coincident signals from electrons on the HV strips and holes on the LV strips. *Right:* The interaction depth (Z) is inferred by the relative arrival times between holes (t_h) and electrons (t_e) at the detector strips. Since the drift times of both species are well-known for a given detector type and applied bias, the distance traveled by the charges through the crystal can be determined and the site of photon interaction can be inferred by taking the difference of their arrival times.

the spectrometer heat load required the use of a mechanical cryocooler. The weight of liquid nitrogen to provide equivalent cooling was prohibitive in utilizing this option.

The flight detectors are arranged in two planes stacked on top of one another along the telescope line-of-sight (figure 7.6). The planar faces are oriented such that the maximum amount of detector surface area points towards the sun. The detectors are operated in coincidence mode so that they can identify and catch photons that scatter between detectors in Compton scatter or pair production events. The detectors are stacked to provide additional active volume to catch these scatter events. Of the eight detectors mounted in the cryostat only six were instrumented due to performance issues of the remaining two detectors. The six instrumented detectors populate the entire top plane and two on the bottom plane such that there two 2-detector stacks and two single-detector stacks.

The detector stacks are attached to a copper coldfinger within the aluminum cryostat (figure 7.5). The coldfinger is mechanically cooled by a Sunpower CryotelCT Stirling cycle cryocooler which operated at ∼85 W to maintain an ∼80 K detector temperature during flight. Heat is passively rejected from the cryocooler exterior to the cryostat through conduction to three radiator plates facing towards space. To accommodate the wide range of ground operating conditions, including integration and final hang test in Texas where exterior temperatures can exceed 35°C, an array of four thermoelectric coolers provided additional cooling by pumping heat from the cryocooler onto a finned heatsink with forced airflow.

The analog signals from GRIPS ∼1,800 instrumented electrode strips pass through the

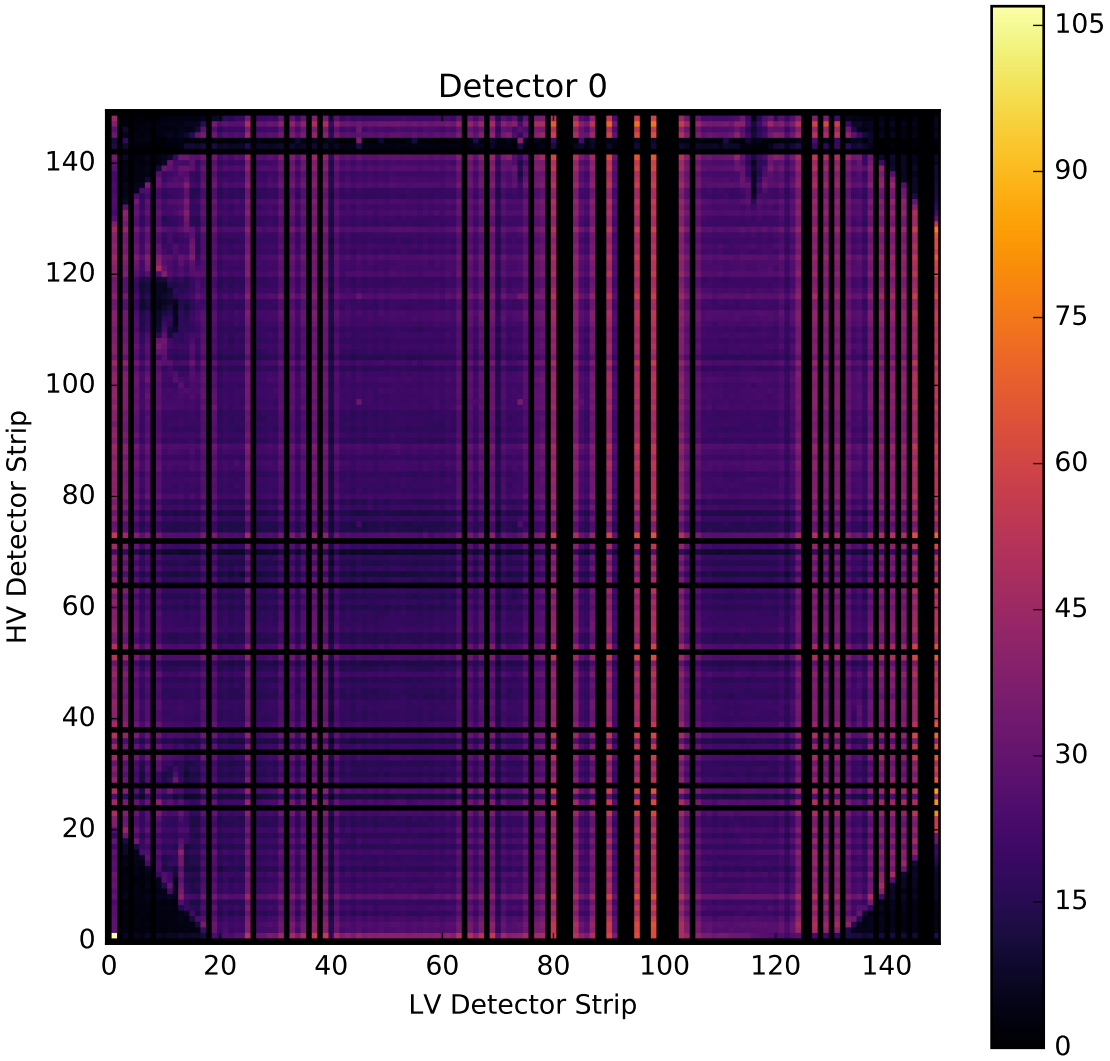

Figure 7.4: A 2D hitmap generated by matching coincident HV and LV electrode strips from a Cs-137 source in detector 0. The events shown triggered only one strip on each detector face and the image was flat-fielded. There are no counts in the corners (besides noise events) because the detector itself has rounded edges. Triggering efficiency and multiple-trigger noise led to the relatively dark electrodes that received few source triggers (40 LV strips and 10 HV strips). Visible in this image are two crystal defects in the upper left and right corners.

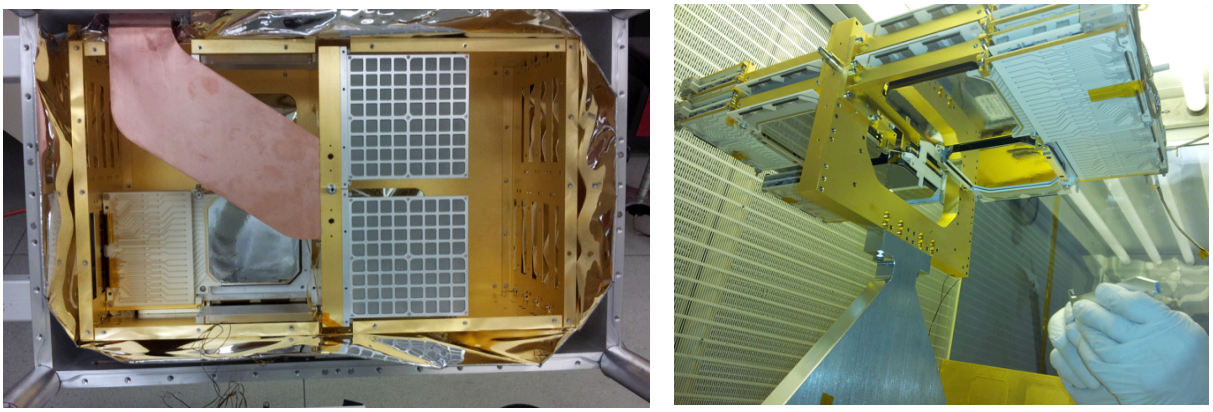

Figure 7.5: *Left:* The interior of the cryostat, integrated with two test detectors for preliminary testing. The detector stacks are housed inside an infrared shield (gold) then wrapped in thermal blankets (silver) before being inserted into the aluminum cryostat shell. The detectors are cooled by a copper cold finger. Shown in the foreground are two activated carbon molecular sieves used to enhance the vacuum. *Right:* The full flight detector stack integrated with electronics boards. Pictures from Duncan et al. 2016 [45].

cryostat wall via 48 polyamide flexible film circuits (*flex circuits*) and custom designed vacuum feedthroughs using Viton seals. Flex circuits were chosen the accommodate the requisite signal density and minimize the signal path length from the detector to readout electronics. Immediately outside the cryostat are 48 ASICs (8 ASICs per detector) which handle triggering, shaping and the first step of analog-to-digital conversion via a Wilkinson ramp comparator. The GRIPS ASIC was designed by IDEAS as an extension of the VATA450 ASIC family used on the ASTRO-H spacecraft [134] and the FOXSI sounding rockets [81]. Section 7.5 and chapter 8 discuss the GRIPS ASIC implementation and performance.

The cryostat is surrounded by 5 cm thick bismuth germanate (BGO) scintillators (figure 7.6) that provide active anti-coincidence shielding. The BGOs surround the cryostat in a hexagonal well around the sides and bottom, to protect against interference from Earth's albedo and atmosphere. The BGO output is read out by 36 individual PMTs. This shield was originally developed for the Max 91 HIgh REsolution Gamma-ray/hard X-ray Spectrometer (HIREGS) [21], a balloon-borne predecessor to the RHESSI spacecraft, and has been successfully re-flown for NCT, and now GRIPS. With a threshold of ∼100 keV, the BGO shields significantly reduce the instrumental background.

7.3 Flex Circuits

A total of 48 polyamide flex circuits (.012 inch thickness) carry signals from the detector breakout boards through the IR shield, thermal blanketing and cryostat wall. After exiting the cryostat the flex circuits immediately enter an EMI shielded box where the signals are

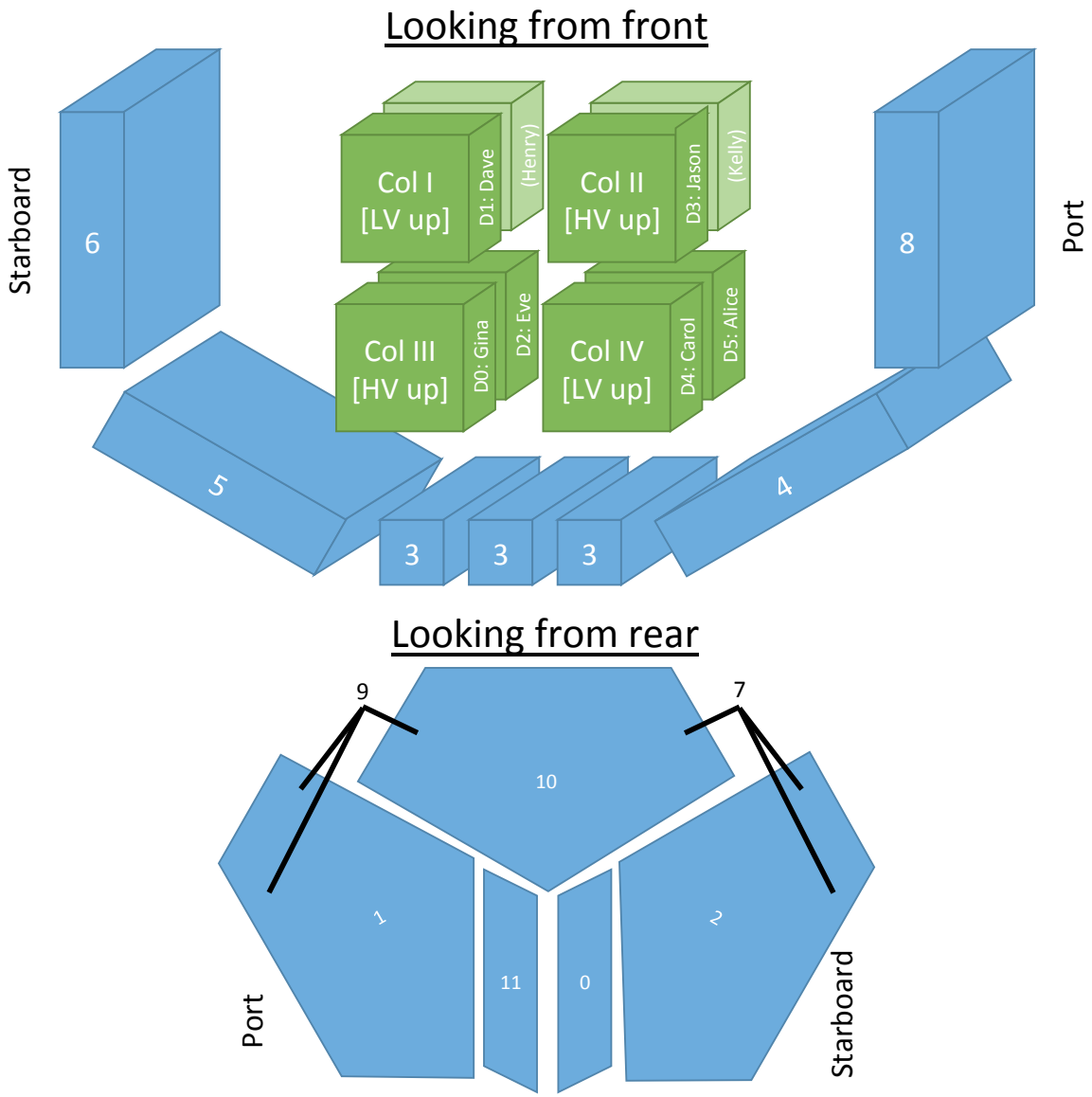

Figure 7.6: Diagrams showing the detector (green boxes) and BGO shield (blue shapes) placement for flight. *Top*: View of the detectors looking towards the rear of the boom, away from the Sun. The bottom of the image points toward the Earth and the BGO are arranged along the bottom to protect against interference from Earth's albedo and atmosphere. The detectors are labeled D0 – D5 and have individual reference names. *Bottom*: View looking along the telescope line-of-sight towards the Sun, from the rear of the boom. The rear BGO also protect against interference from Earth's albedo and atmosphere. Image courtesy of Albert Shih.

Detector	Crystal	Depl V	Op V	LV Side	HV side
D0	HP50814-3	200 V	404 V	fully instrumented	fully instrumented
D1	HP50814-2	250 V	412 V	fully instrumented	fully instrumented
D2	HP50765-4	600 V	306 V	fully instrumented	1/2 disabled
D3	HP42808-1	400 V	500 V	fully instrumented	1/4 disabled
D4	HP50847-1	300 V	610 V	1/2 disabled	fully instrumented
D5	HP50614-1	300 V	610 V	fully instrumented	fully instrumented

Table 7.1: Detector numbers, crystal manufacturing numbers, depletion voltages (Depl V), operating voltages (Op V), and instrumentation on the LV and HV sides. Some detectors were not fully instrumented due to issues with the flex circuits which are discussed in section 7.3. The relative positioning of each detector is shown in figure 7.6.

processed by ASIC electronics.

The cryostat was initially loaded with two test detectors that had a total set of 16 flex circuits. When the cryostat was rebuilt with the full set of 8 detectors, necessitating 48 flex circuits, additional strain relief was added as the circuits exited the IR shield. These modifications resulted in an increased path length that rendered two of the flex circuits too short to mate with exiting readout hardware. The signal traces on these two flex circuits, every other strip along detector 4's low voltage side, were disabled by connecting them to ground.

During the final stages of detector integration three flex circuits were found to have delaminated, causing vacuum leaks. Replacing these three flex circuits at the time of discovery would have risked the other 43 flex circuits that had already been integrated through the cryostat wall. To fix the vacuum leaks all 3 flex circuits were pushed back through the cryostat shell and their electrode strips were disabled by connecting them to ground. The channels disabled were on every other strip of detector 2's high voltage side, and every other strip across one half of detector 3's high voltage side. Table 7.1 summarizes which detector strips were instrumented and which were disabled by connecting their flex circuit traces to ground.

7.4 Operating Voltages

The cryostat was thermally cycled and warmed to room temperature for transport from the laboratory facility in Berkeley, California, to the campaign site in McMurdo, Antarctica. After arrival at the Long Duration Balloon facility the detectors were cooled and leakage current tested to determine the operational voltages for flight. For each detector the voltage was turned-up slowly while monitoring the leakage currents across (1) the guard ring and (2) the summed electrode strips on the low voltage side. Small transients during turn-on

are acceptable, but any current spike greater than 1 μA is considered a failing condition and the voltage is immediately turned down to prevent permanent damage to the detector. After thermal cycling detector 2 failed leakage current testing and was operated below its full depletion voltage. Though not fully depleted, the detector had enough active volume to successfully operate as a monolithic detector that provided spectral information but no spatial resolution.

Following leakage current testing the detector strips were individually instrumented and brought past the depletion voltage. A Ba-133 source was used to evaluate the spectra at multiple voltage settings past the depletion voltage. There are two broad classes of noise in the GRIPS spectrometer: (1) noise that leads to spectral broadening and (2) noise that generates false triggers. In general a high operating voltage is desired so that the charge drift velocities are saturated and the effects of charge sharing, trapping and recombination that lead to spectral broadening are minimized (discussed in chapter 3). In the GRIPS spectrometer it was found that increasing the operating voltages also increased the 2nd class of noise, generating false triggers that would contribute to detector deadtime and obscure the spectral signal. Though methods were adopted to mitigate the spectral effects of multiple-trigger noise (discussed in section 8.6) their presence affects detector livetimes and complicates event selection. In determining operating voltages, the benefits of higher voltages in improving spectral broadening were balanced by the negative effects of multiple trigger noise and the final voltages were set experimentally by finding the highest voltage setting with an acceptable level of noise. Table 7.1 summarizes the depletion and operating voltages used in calibration and flight.

7.5 ASIC Electronics

A single 3D-GeD has 298 electrodes, each with a dedicated signal processing chain from triggering through analog-to-digital conversion and single event storage. ASICs were chosen to handle the processing load of ∼1,800 separate channels while keeping reasonable constraints on the size, weight, power and thermal loading for the instrument. The GRIPS electronics chain consumes 6x less power than comparable discrete electronics: 25mW vs NCT's 150mW per channel. GRIPS' ASICs (figure 7.7) are a revision of IDEAS' VATA453, a chip originally designed for KIPAS/ISAS on Japan's NeXT mission and used with revisions on the Focusing Optics X-ray Spectroscopic Imager (FOXSI) rockets [57], also built by UC Berkeley's Space Sciences Lab.

Each ASIC has 64 independent signal channels (figure 7.8). The parallel channels consist of a charge-sensitive preamplifier that feeds into parallel fast-shaping and slow-shaping paths for triggering (TA) and energy measurements (VA), respectively. The initial charge-sensitive preamplifier stage converts charge induced on an electrode into an equivalent voltage. The fast channel discriminator provides a trigger signal if the incoming pulse exceeds a user-set threshold level. The slow channel shapes the voltage signal from the preamplifier and provides the input to a Wilkinson ramp comparator used for analog-to-digital conversion.

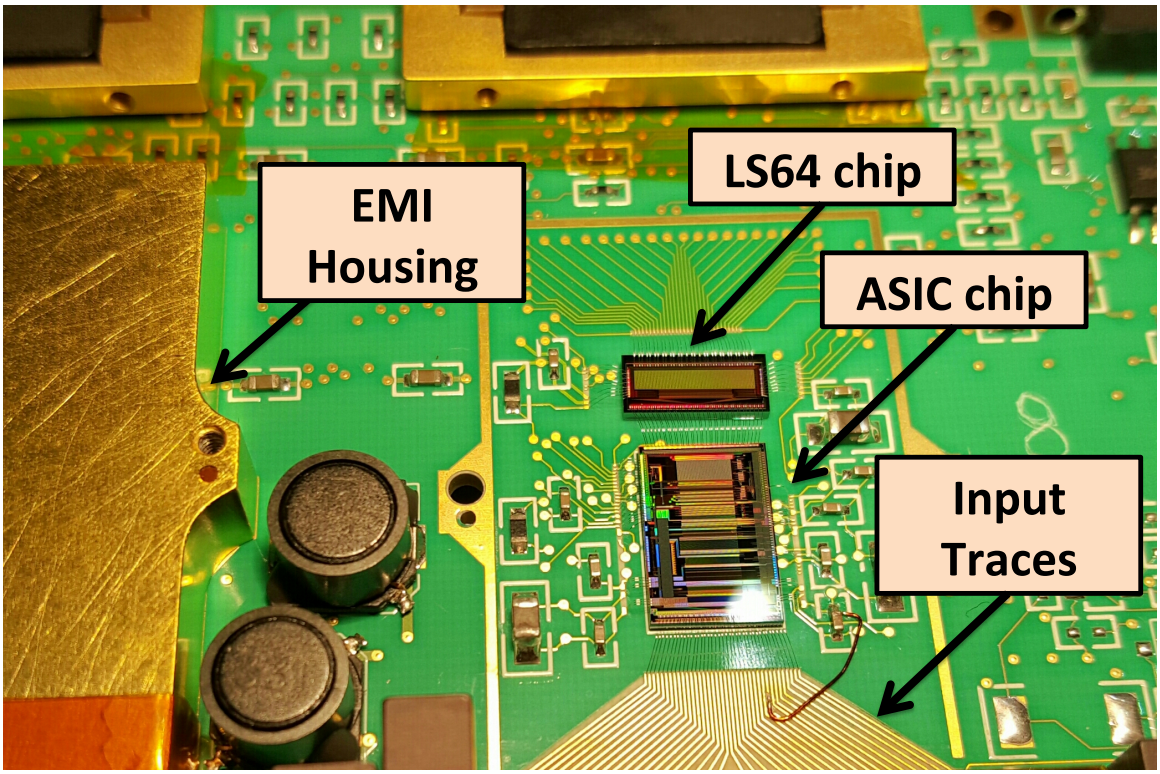

Figure 7.7: The VATA453 ASIC chip on a GRIPS flight electronics board. The photosensitive ASICs are usually covered in a gold EMI housing, pictured to the left for an adjacent ASIC. A flex circuit is mechanically connected to the input traces that are wire bonded to the ASIC chip. The ASIC output passes through an LS64 level shifting chip (also designed by IDEAS) before exiting the board.

The outputs of all ASIC channels are transmitted over ∼25 ft of ribbon cables (pictured in figure 6.3) from the main spectrometer to FPGAs in the electronics bay. The FPGAs complete the digitization step by counting clock cycles as the ASIC holds their output low for a time equivalent to the integrated charge. The quantity of elapsed clock cycles is the digital conversion value of the input signal, referred to as the *ADC bin* value in this manuscript. Effects associated with this two-step digitization are discussed in chapter 8. ASIC control settings such as gain, thresholds, shaping constants and analog-to-digital comparator ramp speed are adjustable through on-chip DACs.

A discriminator in the ASIC fast channel implements each channel's trigger threshold independently. The global trigger threshold is set for an entire ASIC via a single DAC setting. Each channel's trigger threshold can then be independently adjusted from the global setting. In practice the single global setting was found to be adequate. Since each channel has independent trigger circuits there is a spread in threshold energies and widths across the detector strips.

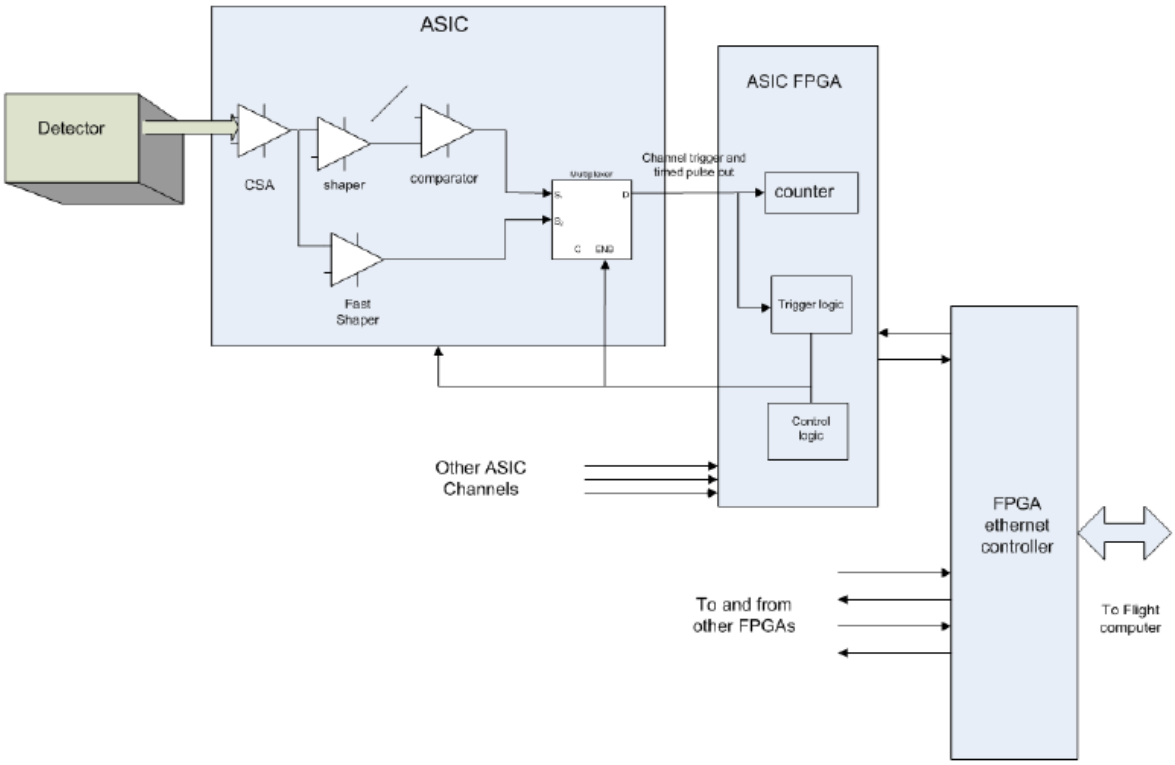

Figure 7.8: Electronics chain schematic showing how the detector is read out by the ASIC and FPGA. The ASIC handles triggering, signal shaping and the initial digitization step via a comparator. The FPGA completes the digitization by counting clock cycles while the comparator keeps the ASIC output low. Schematic provided by IDEAS.

After the signal has been digitized to an ADC bin value the pulses are histogrammed into the differential pulse height spectra (described in section 3.1). Calibration sources of known line energies are used to determine a gain function to convert ADC bins to energy. Figure 7.10 illustrates several key features of a detector spectrum from a single detector strip. The traces are a series of barium-133 spectra taken at successive threshold settings (range of $1 - 31$ in arbitrary units). In this plot the counts have been normalized to arbitrary units for comparison of the spectral features and each line is marked with its energy in keV. The pedestal is composed of conversions that were triggered due to noise but had no real charge to shape and it marks the zero-energy bin of the analog-to-digital conversion. As the threshold is raised triggering due to noise ceases and the pedestal disappears, the prominent 81 keV line is progressively cut out and the less active, higher-energy lines become visible over the high-energy continuum.

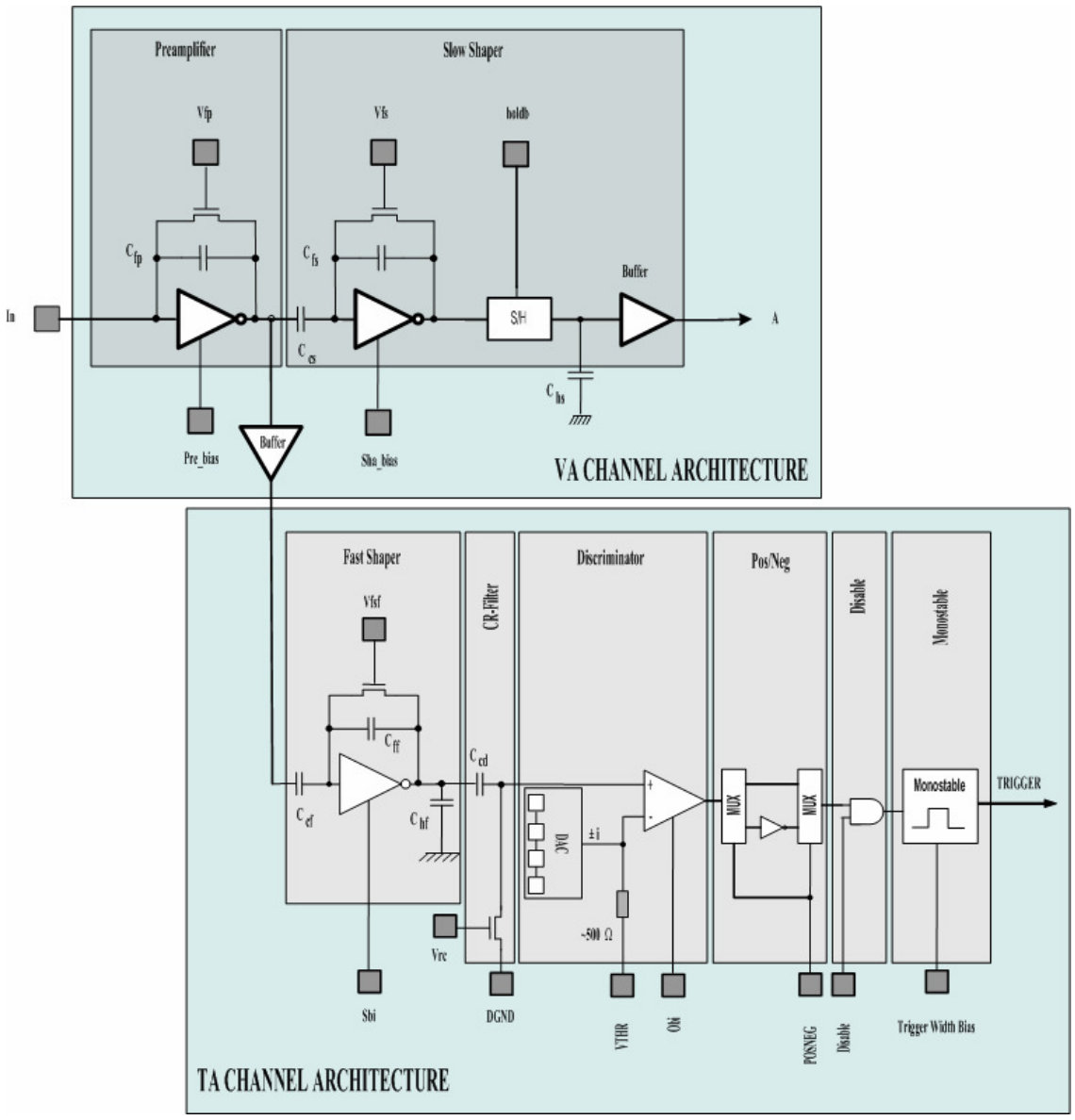

Figure 7.9: Detailed schematic of the VATA453. For each detector strip input there are two parallel ASIC channels, a fast channel used for triggering (TA) and a slow channel (VA) used for charge shaping and analog-to-digital conversion. The VA-TA channels share a single output for each detector strip (total of 64 input and output channels), with triggering and conversion-mode switching handled by the FPGA. Schematic provided by IDEAS.

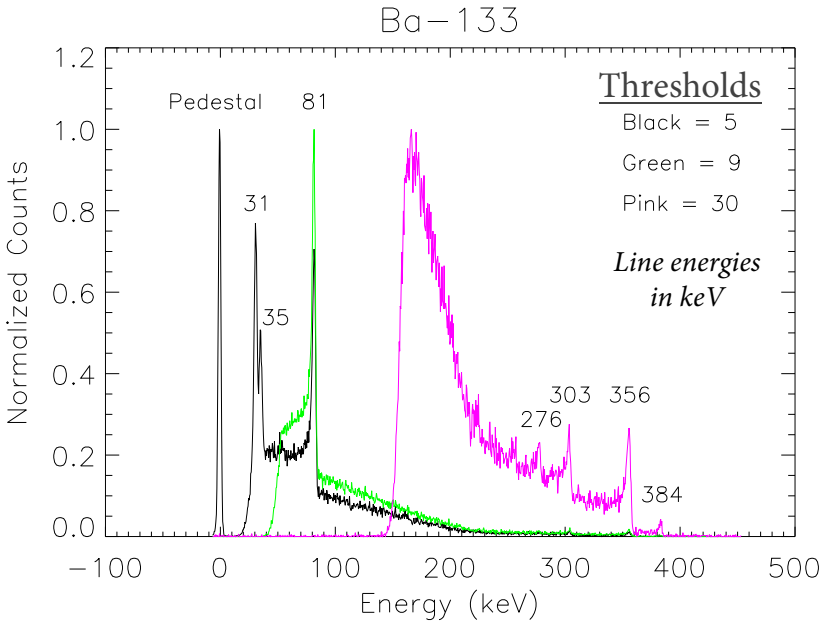

Figure 7.10: Example GRIPS barium-133 spectrum shown at three successive threshold settings of 5 (black), 9 (green) and 30 (magenta) out of a maximum setting of 31. The spectra have been normalized to emphasize the less active, higher-energy lines and illustrate the effects of thresholding. The known line energies are written on the plot in keV and the pedestal, which marks the zero-energy bin, is labeled. Plot adapted from Duncan et al. 2013 [44].

7.6 Status of the Spectrometer

After GRIPS' January 2016 flight the instrument spent the 2016 Antarctic winter in the field and the hardware was recovered in January 2017. The spectrometer stayed above the snow-pack, minimal snow was found surrounding the cryostat vessel and the electronics remained dry. Though the external frame and radiator plates sustained damage upon landing (figure 7.11), the cryostat, ASIC electronics boxes and BGO were not visibly damaged. The instrument hardware returned to Berkeley in May 2017. Since the spectrometer sustained a >10 G impact, the cryostat will be opened to assess the state of the detectors and their components. The ASIC and FPGA electronics can be tested independently and attached to the cryostat once it has been inspected.

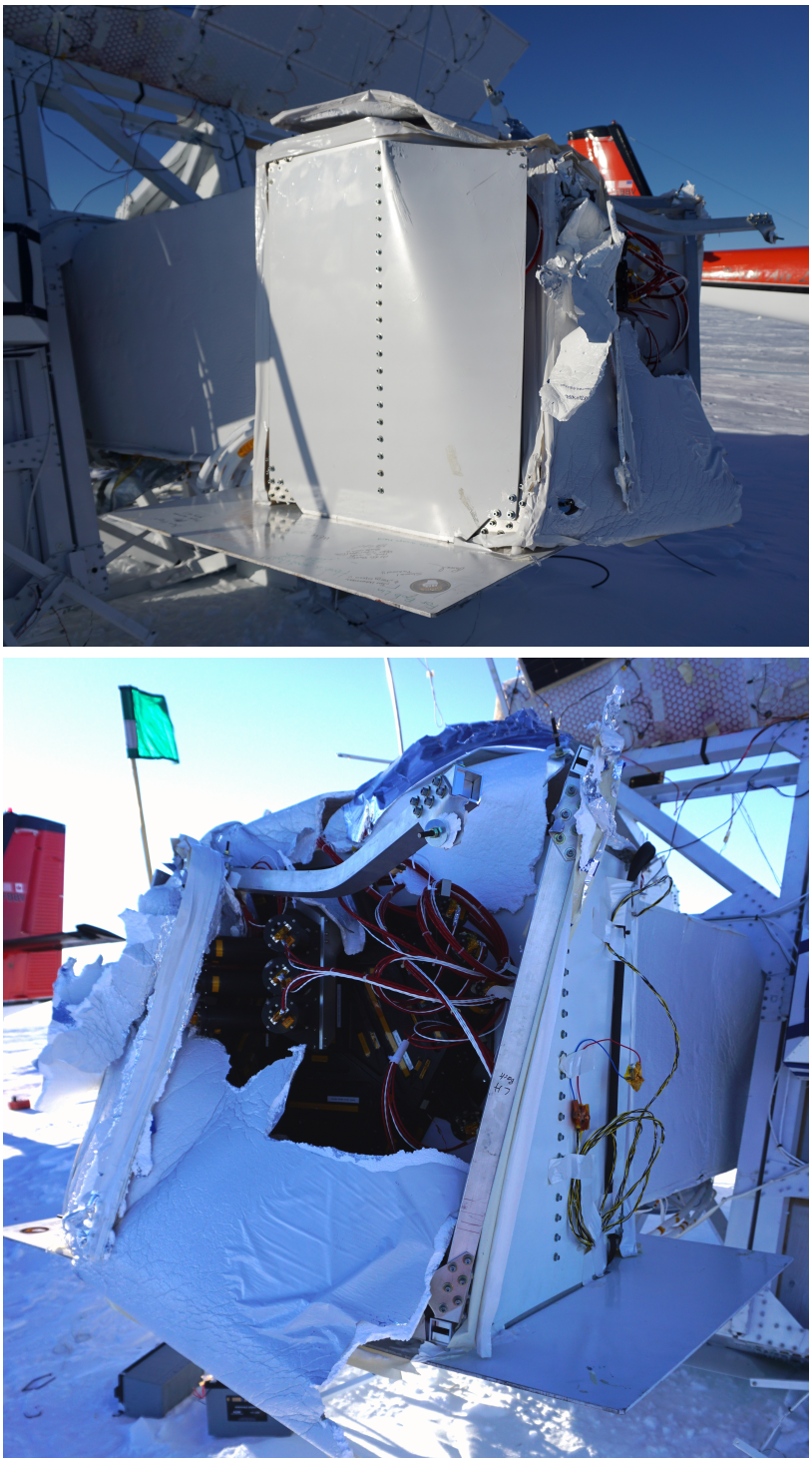

Figure 7.11: Images of the spectrometer after GRIPS' flight in February 2016. Though spectrometer exterior structure and radiator plates sustained damage during landing, the cryostat, ASIC electronics and BGO were not visibly damaged.

Chapter 8

GRIPS Spectrometer Performance

The GRIPS flight detector performance is evaluated in detail for the 300 strips across the low voltage (LV) sides of detector 0 (D0) and detector 4 (D4). The studies in this chapter will be expanded to the full 1,800 strips based upon initial investigation of this representative set.

8.1 Efficiency

Detector efficiency is the ratio of the measured count rate to the true rate of photons incident on the detector. Efficiencies less than unity can result from numerous factors in the detector and readout electronics implementation.

The GRIPS detector efficiency (ϵ) was determined by taking the ratio of the photopeak count rate (N) to the predicted source rate at the detector (S_D)

$$\epsilon = \frac{N}{S_D} \tag{8.1}$$

Detector geometry, source-to-detector distance and line-of-sight attenuation through passive materials determine the quantity of source photons that arrive at the detector. The source emission was assumed to be isotropic such that the projected counts are a simple ratio of the detector solid angle (Ω) to the full sky (4π). Attenuation through the cryostat lid was modeled as a decaying exponential of the aluminum mass-attenuation coefficient (μ) [130], the total aluminum thickness (x) and its density (ρ).

$$\begin{aligned} S_D &= S\,\frac{\Omega}{4\pi}\,e^{-\mu\rho x} \\ \mu &= .2778\ cm^2/g \\ x &= 0.332\ cm \end{aligned} \tag{8.2}$$

The measured count rate (N) can be determined by fitting a decaying exponential to the distribution of time between events at the detector. This method of determining the

count rate is more robust than simply counting the events per time because it circumvents ambiguity in the detector livetime (from duty cycling, discussed below) and isolates the true source counts from noise (section 8.6). Figure 8.1 shows the time between event plots for detectors 0 & 4. Noise in the GRIPS detectors is highly correlated, coming in very quick succession after an event. These noise events are largely confined to the region to the left of the dashed lines and the real source counts are easily isolated to the right of the line. Comparison of the fitted count rates from the time between events plots are in good agreement with the average count rates when appropriate event selections (to exclude noise and duty cycling) are made.

An Am-241 source was placed 1.33 m along the telescope line-of-sight to evaluate GRIPS' on-axis efficiency at 59.5 keV. The total aluminum thickness along the telescope line-of-sight is .332 cm. Though thermal blanketing was also present in the GRIPS line-of-sight, attenuation due to these thin layers of mylar and absorption in the air were not considered in this calculation. Using the fitted count rates and eqn. 8.1 the on-axis efficiencies for these two detectors at 59.5 keV were measured as:

$$\epsilon_{D0} = 12.3\ \%$$
$$\epsilon_{D4} = 13.8\ \%$$

Duty Cycling

During normal operation the detectors were observed to toggle between states of nominal operation and a suppressed state characterized by fewer triggered counts. This duty cycling was slightly different in each detector, did not follow a deterministic pattern and occurred both in flight and laboratory testing. Described below are observations of this phenomena and the steps taken to mitigate its effects.

Duty cycling manifested differently in each detector. Detector 4 toggled between states that affected the energy spectrum uniformly: either the detector was operating nominally or its entire spectrum was suppressed. In this case the total amount of time spent in the suppressed state is easily corrected for by scalar multiplier. In detector 0 the situation is more complicated: the duty cycling is energy dependent and predominately affects the <250 keV portion of the spectrum. Figure 8.2 (top) shows duty cycling in D0 and D4 spectrograms during a Cs-137 test in the laboratory.

Figure 8.2 (bottom) shows spectrograms of Am-241 for detector 0 & 4. Detector 4's duty cycling is consistent with the behavior shown in Cs-137 data, nominal operational states interspersed with suppressed periods. Detector 0, however, does not show the same state toggling behavior as in the Cs-137 data. Here, there is no evidence of state toggling, though the radiation (59.5 keV) is in the energy range affected by toggling in the Cs-137 data. The only detector-setting difference between two datasets is the trigger threshold, which does not affect the comparison since it is set higher for Cs-137 than for Am-231.

Detectors in the GRIPS cryostat are arranged into two planes: four detectors are in the "top" position pointing towards the target and two detectors are below in the "bottom"

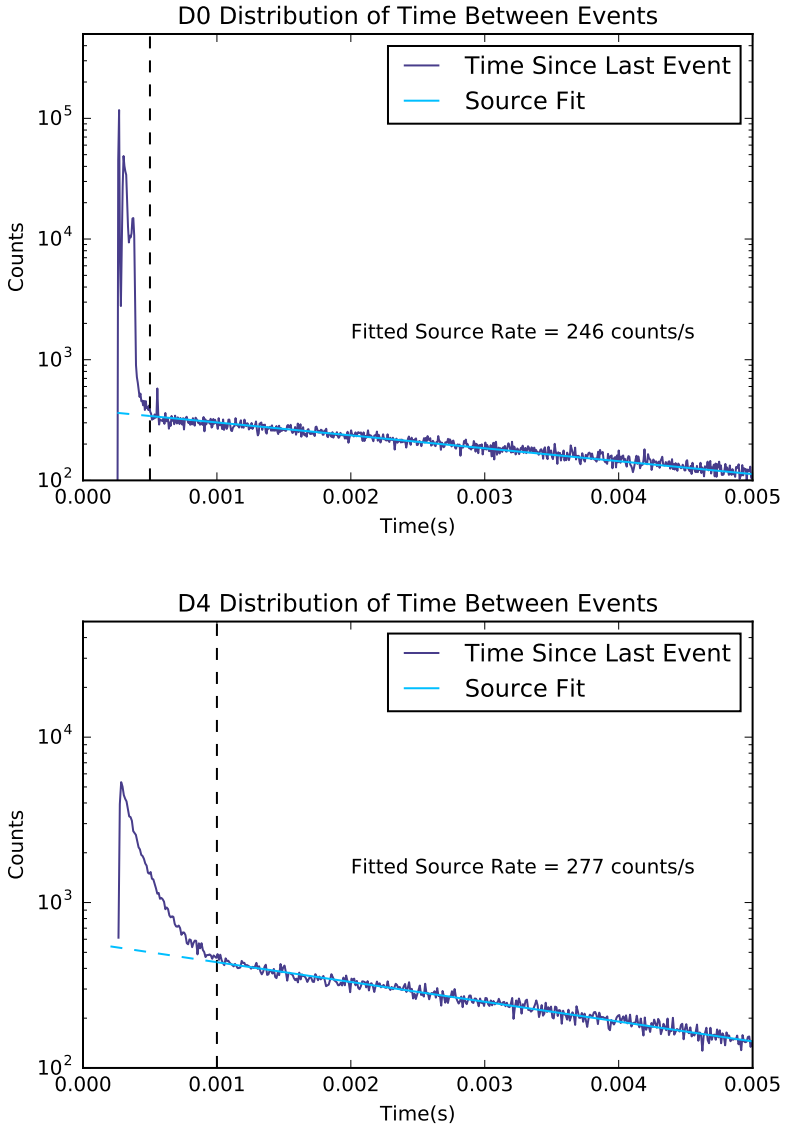

Figure 8.1: Distribution of the time between events recorded by D0 and D4 for Am-241. The real source counts are Poisson-distributed in time, appearing as sloped lines on the semi-log plots. Data to the right of the dashed vertical lines were fit to exponentials whose decay constant represents the measured source rate for 59.5 keV photons. The fits used to determine the measured count rates are shown as solid blue lines. The fits have been extrapolated to the left of the fit region to show the expected source counts below the dashed vertical lines. In this region the remainder of the counts are due to noise and were excluded from the fitted exponential. These noise counts at short timescales are discussed in sec 8.6. D0's short timescale structure is zoomed in for detail in fig 8.17.

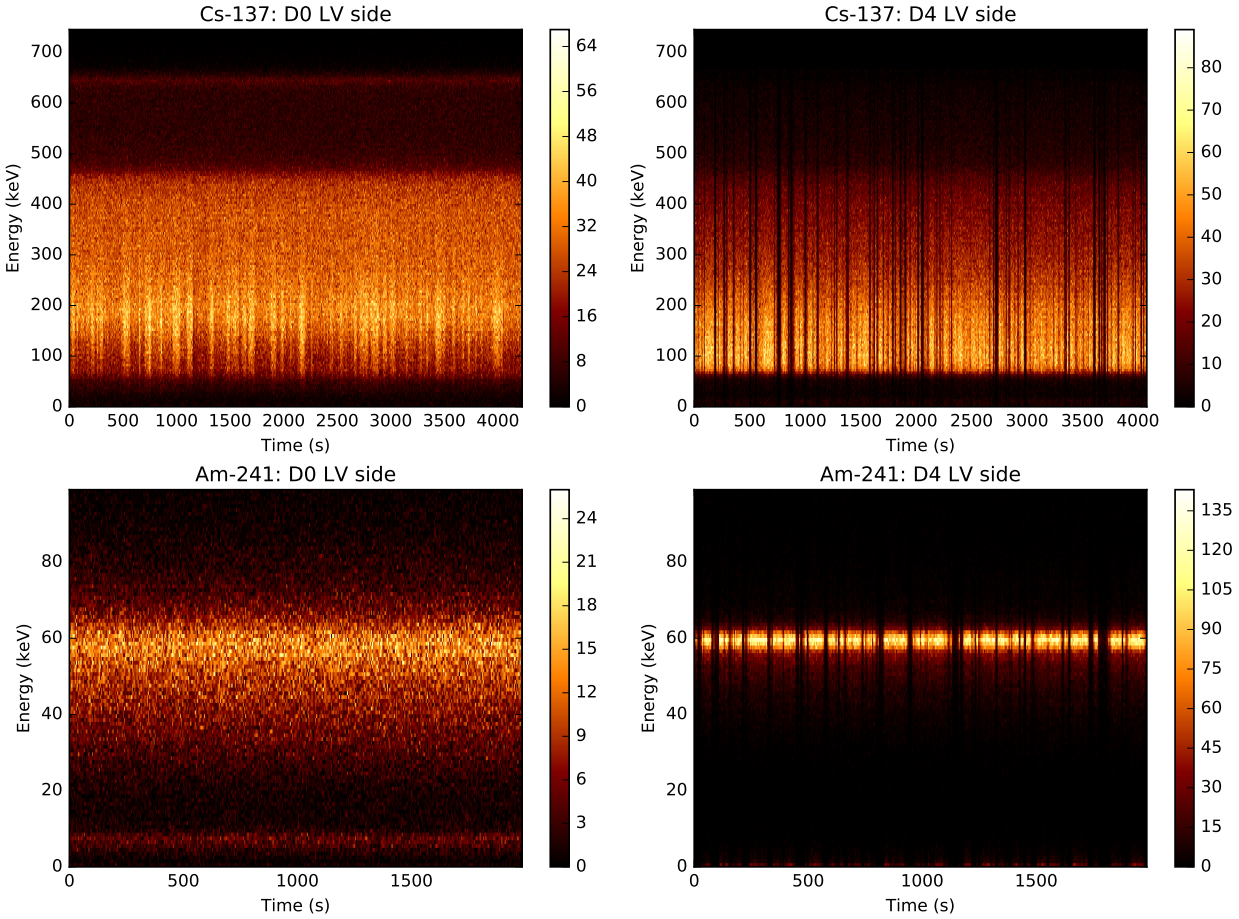

Figure 8.2: *Top*: Spectrograms for Cs-137 radiation at 5s intervals, showing duty cycling. Detector 0's energy dependent duty cycling occurs primarily below ∼250 keV. In contrast, D4's duty cycling is energy independent, affecting the entire spectrum uniformly. *Bottom:* Spectrograms for Am-241 radiation at 5s intervals. Detector 4's energy-independent duty cycling is consistent with data shown in the top plot for Cs-137 data. Detector 0, however, does not show the same energy-dependent state toggling as it did during the above Cs-137 data run. The linewidth difference between detector 0 & 4 is due to greater charge trapping on detector 0 (discussed in section 8.4). The lines at ∼ 7 keV on D0 and ∼ 2 keV on D4 are due to glitches (discussed in section 8.5).

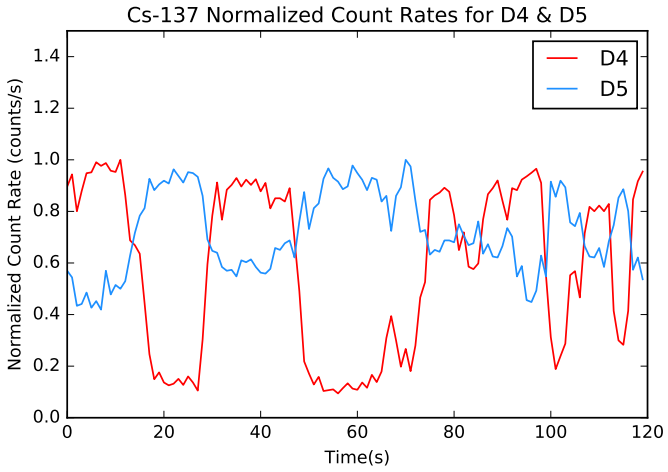

Figure 8.3: Am-241 count rates for detector 4 and detector 5 normalized to emphasize their anti-correlated duty cycling. These two detectors are arranged in a stack with detector 4 closer to the source (top detector plane) than detector 5 (bottom detector plane).

plane. The state toggling was anti-correlated in both D0(top)/D2(bottom) and D4(top)/D5(bottom) stacks. Normalized D4 and D5 count rates are shown in figure 8.3 for Cs-137, showing how these two detectors toggle between normal and suppressed states together.

The total count rate and the source count rate are shown for detector 4 in figure 8.4 (top). In nominal operation the total count rate (red) and source count rate (teal) difference (effectively the noise) is consistently non-zero. When the detector duty cycles to the suppressed state the noise suddenly drops, and the difference of total count rate to source count rate goes suddenly to zero. Plotting correlations between the noise and the total count rate (figure 8.4 bottom, left) or source count rate (figure 8.4 bottom, right) show that the two states are different populations that can be isolated from one another by a threshold, marked by a vertical line. This threshold was used as a filter to identify and correct for detector 4 duty cycling in post-processing software. Implementing a software correction for detector 0 is not as straightforward as with detector 4. More work needs to be done in characterizing detector 0's energy dependent state toggling to build a reliable filter and correction.

8.2 Detector Spatial Resolution

Crossed tungsten rods were embedded into the cryostat top; for low-energy radiation these rods casted X-ray shadows onto the detectors. These shadows were used to determine the 2D detector spatial resolution by comparing the width of the shadow to the electrode pitch and rod diameter (0.46 mm). Figure 8.5 shows counts summed along the detector-diagonal direction for two flight detectors with 59.5 keV incident radiation. The dips at pixels 105 and 125 indicate the location and width of the tungsten rod shadow, showing that its projection

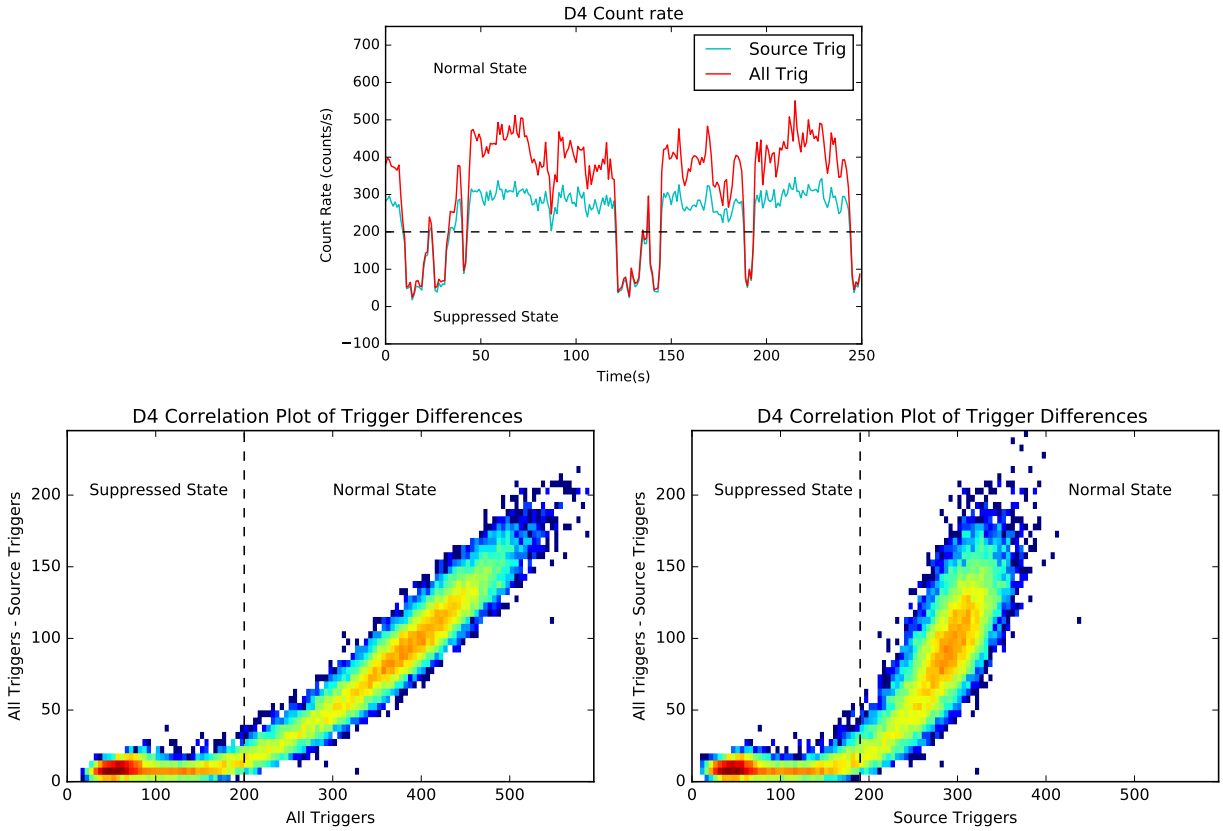

Figure 8.4: Am-241 ground calibration data. The top plot shows the detector count rate at 1s intervals for source counts (selected for all triggers ≤ 2 per side) plotted against all triggers. The dashed horizontal line is a count rate threshold that separates the data into normal and suppressed states. *Bottom:* Correlation plots between noise triggers (all triggers - source triggers) to all triggers (*left*) and source triggers (*right*). These plots show that the normal and suppressed states are clearly two different populations that can be isolated from one another by a count rate threshold, marked as a dashed vertical line. Counts to the right of the dashed line are in the normal state and counts to the left are the suppressed state where the noise drops to zero.

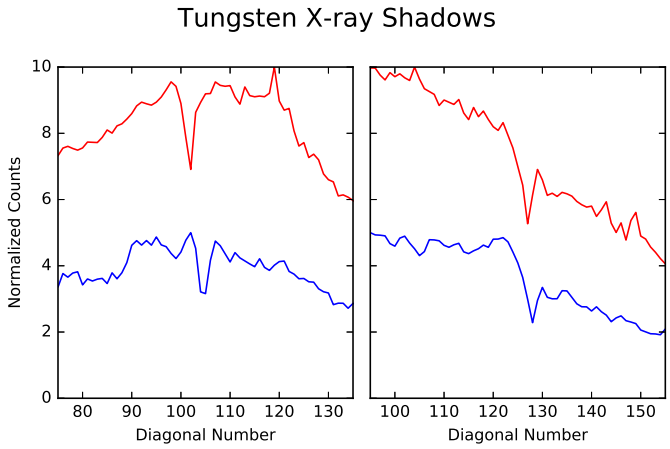

Figure 8.5: Crossed tungsten rods were used to cast X-ray shadows along the detector diagonals. The rods can be seen as dark "X"s in figure 9.3. These plots show a the sum of counts along the detector diagonals for an Am-241 source, normalized to show two different detectors (D0 in red and D4 in blue) in single plot. Since the tungsten rods cast a shadow along a detector diagonal, the summed counts dip at the location of the rod. The width of the dip provides a measure of the detector spatial resolution. Since the diagonals do not have the same quantity of pixels, the summed counts have an overall decreasing trend. The smaller count fluctuations between diagonals are primarily due to variations in the energy thresholds for each pixel.

is approximately the width of one strip pitch: .5 mm. Sub-strip level position accuracy can be inferred by comparing the extent of charge sharing between two neighboring strips.

Collection-time differences between electrons and holes on opposite detector faces measure how deep within the crystal a photon deposited energy, as discussed in section 3.4. Figure 8.6 shows a histogram of collection-time differences for two sources in GRIPS detector 0. Both traces go to zero at ±500 ns because this marks the physical edge of the detector active volume. Assuming that the entire detector width (1.5 cm) is active volume, these plots indicate that the charges are moving at $\sim 1.5 \times 10^6$ cm/s inside the semiconductor. This is in good agreement with the value predicted by figure 3.8 for a germanium detector operating at ∼80 K and a 600 V bias. Since GRIPS has a timing resolution of ∼50 ns FWHM, the depth can be inferred to within ∼1 mm. The 3D-GeD cross-strip segmentation combined with depth resolution through timing yields a 0.5×0.5×1 mm virtual pixelation within the detector active volume.

8.3 Energy Calibration

Radioactive sources of known line energies were used to determine the conversion from ADC bin values output by the detector readout chain to real photon energy absorbed by the

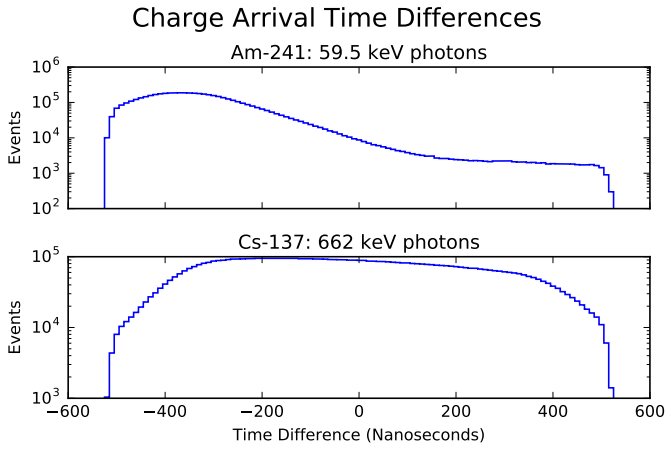

Figure 8.6: Example arrival time difference plots for Am-241 (*top*) and Cs-137 (*bottom*) sources. The difference of arrival times provides a measurement of photon interaction depth within the crystal. The ±500 ns edges are created by interactions that happen close the edge of the active volume. The lower-energy 59.5 keV photons have a shallower penetration depth than the 662 keV photons and are primarily collected near the side of the crystal that they're incident on. The higher-energy Cs-137 photons more uniformly penetrate the crystal.

detector. The sources used for energy calibration were Am-241 (59.5 keV), Ba-133 (81 keV, 276 keV, 303 keV, 356 keV, 384 keV) and Cs-137 (662 keV). Due to the size of ionization clouds $\geq$1 MeV and their associated effects, for Co-60 (1173 keV and 1333 keV) there were few single-strip energy depositions. The full 10 MeV energy range is achieved by summing over multiple strips from charge sharing in single-site energy depositions, multiple energy depositions in a Compton scatter track within & between detectors or from pair production.

The higher-energy lines of Ba-133 (276 keV, 303 keV, 356 keV and 384 keV) as well as Cs-137's 662 keV line were fit with a delta function convolved with a Gaussian to represent the shaping circuitry's spectral response. The continuum was modeled as a DC offset split at the line center. Fitting the 81 keV line required multiplying the same line + continuum used for higher-energy lines with a threshold function. The threshold was represented by a Heaviside function convolved with a second Gaussian to represent the triggering circuitry's response. Figure 8.7 shows the 81 keV fit features including the line position and threshold location.

The low-energy 59.5 keV line from Am-241 showed strong spectral effects due to charge loss by trapping (discussed in section 8.4). Figure 8.8 shows example Am-241 spectra for individual strips on D0 and D4 LV (top) and HV (bottom). The effects of charge trapping are minimal for D4 LV and D0 HV, whose line shapes are Gaussian with low-energy tails. D0 LV and D4 HV lines have a completely different shape than the previous set, showing large spectral broadening that affects the entire low-energy side of the lines. The 59.5 keV

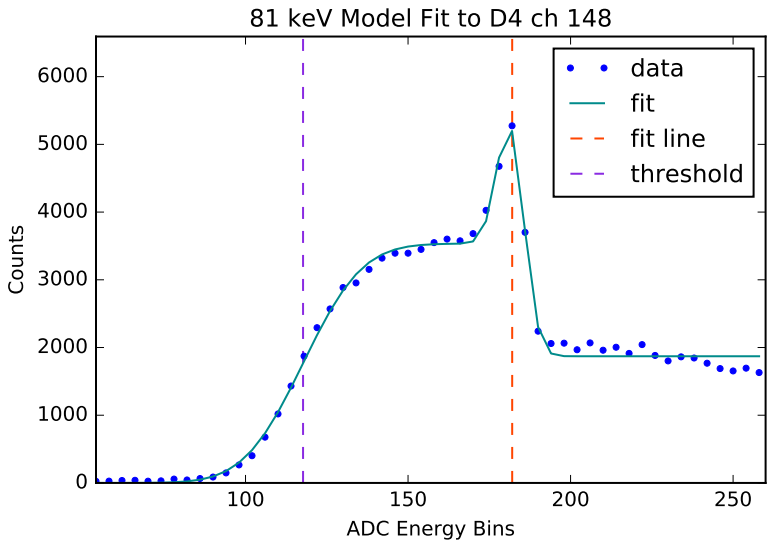

Figure 8.7: Model fit to the 81 keV line and threshold for channel 148 on D4. The model fit is shown in green, the line and threshold fit locations are marked by dashed vertical lines. Above the line is a continuum which is approximated by a DC offset in the region surrounding the line. Only the photopeak is used for calibration and modeling the continuum accurately is not critical.

photons predominately deposit their energy close to the surface of the crystal, as can be seen in the depth distribution plot in figure 8.6. D0 is oriented with its HV side towards the source and D4 is oriented such that its LV side is towards the source. The increased spectral broadening on their opposite faces is likely due to increased trapping since on average those charges would be traveling much farther before collection (discussed in section 8.4).

Both detector 0 and detector 4 Am-241 spectra were fit with split Gaussians: a higher-energy Gaussian that represents a convolution of the line with the instrument response and a lower-energy Gaussian that represents the tailing effect. The location where the two Gaussians joined was a model parameter that was constrained differently depending on which side of the detector the strip is on. For the detector side that faces the source (D0 HV & D4 LV) the line and split locations were not constrained to one another. For the detector side that faces away from the source (D0 LV & D4 HV) the two Gaussians were constrained to join at the line location. In this case the FWHM resolutions quoted for 59.5 keV are derived from the sigmas of the higher-energy Gaussian fits.

Charge sharing for energy depositions above ∼200 keV was a significant challenge to collecting full photopeak spectra on a single strip. Figure 8.9 compares spectra from a strip that is electrically connected to its neighbor (channel 234) and the spectrum from a single strip. Though having similar 81 keV magnitudes relative to the continua, channel 232 is missing the higher-energy lines observed in channel 234. This spectral difference is due to

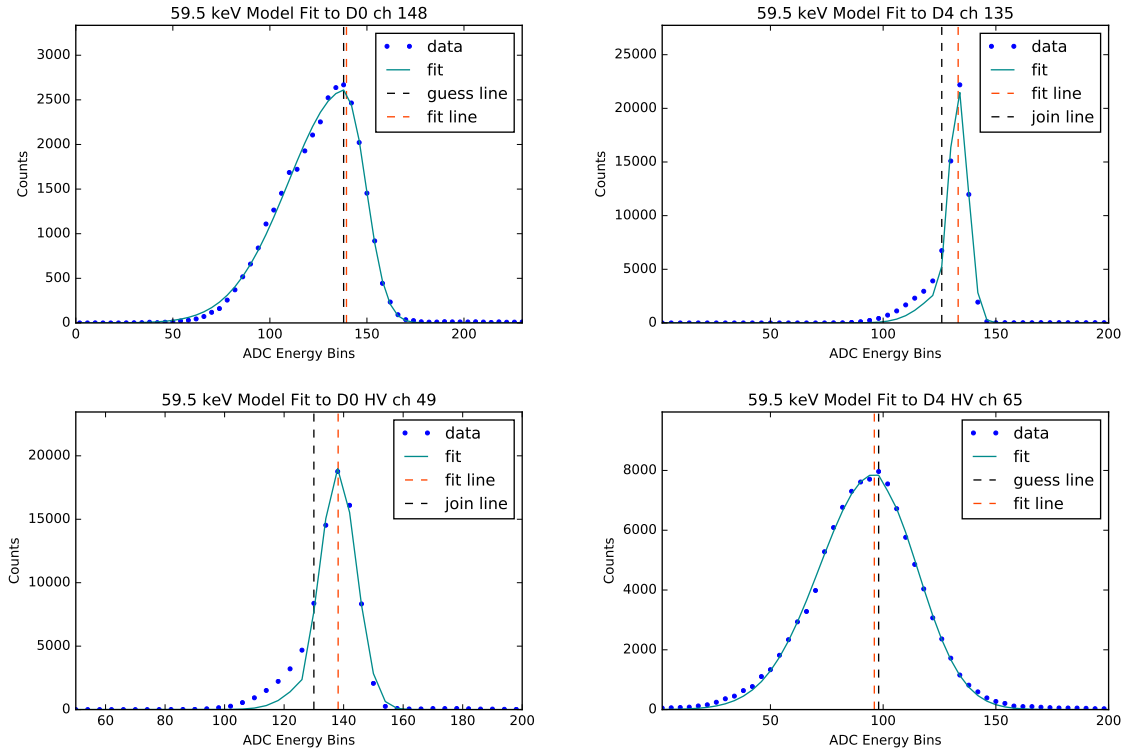

Figure 8.8: Am-241 spectra showing the 59.5 keV line on D0 (*left* panels) and D4 (*right* panels) for low-voltage (*top* panels) and high-voltage (*bottom* panels) sides. The spectral shape differences are due to increased trapping on detector sides facing away from the source relative to detector faces that are toward the source. In D0 the 59.5 keV radiation is incident on the HV side and on D4 it falls on the LV side.

charge sharing between neighboring electrode strips. Channel 234 has ∼2× the effective area of channel 232 and is able to collect the majority of the charge cloud across its two electrode strips, whereas channel 232 can only collect a portion of that charge cloud. Most of the strip spectra resemble channel 232 over this integration time. With very long integration times the higher-energy lines do become resolvable, but their magnitudes are small and the set is sparse such that not every channel has a complete set of calibration lines.

GRIPS's data analysis is particularly challenging due to its ∼1,800 detector channels that must be individually calibrated. A dense electrode spacing (.5 mm) imparts a narrow geometric area per strip, increasing the integration times required for sufficient calibration statistics and increasing the effects of charge sharing between strips. In many cases duty cycling (section 8.1) toggled the detectors to a suppressed state where source counts were not triggered, effectively reducing the count rate by half over an integration period. Low operating voltages (section 7.4) likely led to increased trapping and spread of the charge cloud within the detector active volume, further complicating energy calibration by reducing

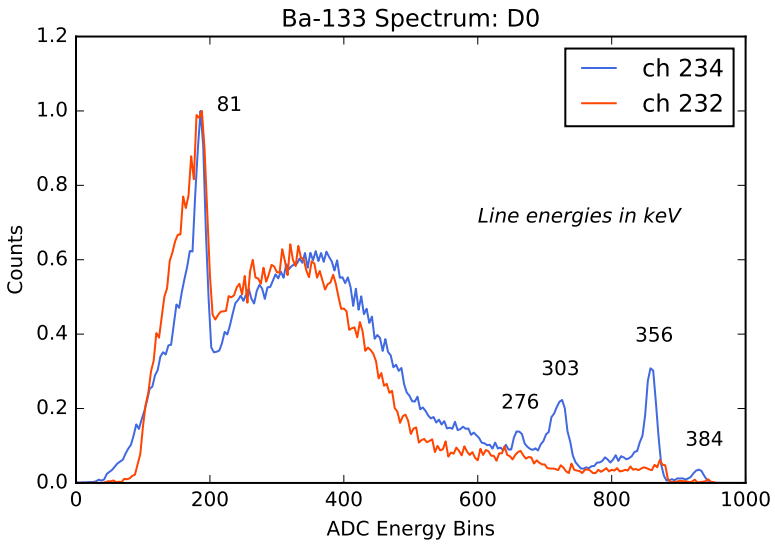

Figure 8.9: Ba-133 strip spectrum for detector 0's low voltage channels 234 and 232 over the same integration time. The spectral counts are plotted in the common-mode subtracted (section 8.5) ADC bins and marked are the known line energies. The spectra have been normalized to the counts in the peak of the 81 keV line and channel 232's spectrum has been translated in ADC bins such that it aligns with channel 234. Channel 234 is electrically connected to a neighboring electrode on the detector and has $\sim 2\times$ the effective area of channel 232.

the overall quantity of single-strip events for higher-energy lines. For some channels glitches and cryocooler noise (section 8.5) blurred the lower-energy lines and removed good source counts from the spectra. With these difficulties the calibration dataset was significantly reduced by livetime and event selection.

For a given line energy the ADC bin values had little variability for individual channels within a single ASIC. This property was used to provide calibration points to channels with incomplete sets of lines. Figure 8.10 shows the ADC bin values for channels on the LV side of D0 and D4. Horizontal lines mark the median ADC value for each calibration line for the channels on a single ASIC. These median line values were used to generate by-ASIC gains for channels which did not have a full set of calibration lines. These by-ASIC gains were checked against ADC values for lines that were intentionally left out of the set of lines used to determine the median values. This check yielded good agreement with the known line energies.

After charges are collected on an electrode the signal is shaped by analog circuitry, and the voltage is sampled at a fixed time that should occur at the peak of the shaped signal. This single voltage measurement is then digitized and represents the magnitude of the signal. The shape of the signal and the timing of the peak depends on the analog R-C time

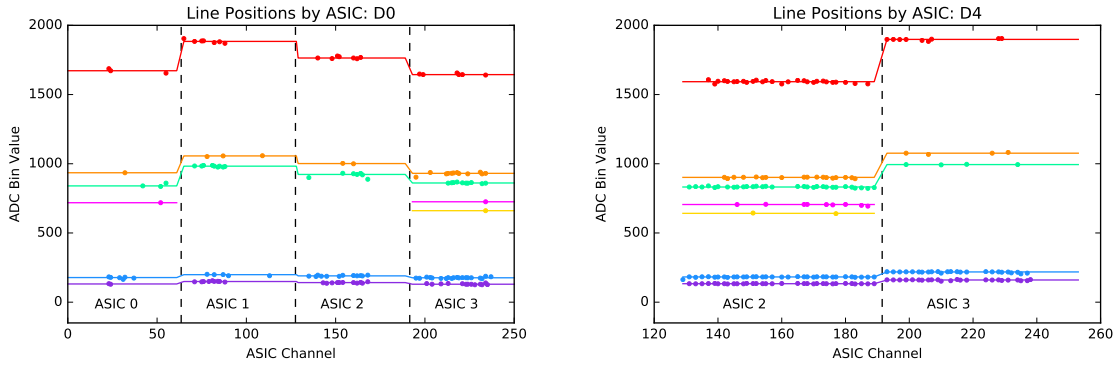

Figure 8.10: Common-mode subtracted ADC bin values by ASIC channel on D0 and D4. The individual ADC bin values (solid dots) are nearly uniform across multiple channels within the same ASIC. Horizontal lines mark the median ADC bin value for each available line by ASIC. The line energies are 59.5 keV (purple), 81 keV (blue), 276 keV (yellow), 303 keV (magenta), 356 keV (green), 384 keV (orange) and 662 keV (red). There are only two ASICs instrumented on D4's LV side because the other two flex circuits were disabled (section 7.3). Selection cuts based on FWHM of the fitted lines reduced the set of channels that were used to determine the by-ASIC line values to the smaller set shown in these plots.

constants. Sampling the signal at the peaking time is critical to accurately measuring the charges collected at the electrode. Incorrect sampling times, either before or after the peak, will result in partial measurements of the full energy deposition. Determining the correct peaking time for a circuit can be done experimentally by varying the peaking times while holding the input constant. The peak of these curves represent the correct sampling time. Figure 8.11 shows these curves for the GRIPS ASICs on an LV and HV strip for multiple energy inputs. Ideally all the curves for multiple energies would peak at the same time, and frequently they do with discrete analog components. In the GRIPS ASICs, however, the slow-shaper peaking times vary with energy.

A single sampling time must be specified for each ASIC over its entire operational energy range. Since the GRIPS ASICs have energy dependent peaking times, there will be sampling errors in the charge measurement. There will be smaller errors if the signals are sampled on the decay because the signal rise time (before the true peaking time) is faster (has a steeper slope) than the decay time (after the peak). The direction of the shaping curves and preference to sample on the falling edge led to choosing a shaping time consistent with the highest expected single-site energy deposition on the LV side and minimum energy deposition on the HV side. Sampling on the decaying edge means that any lower-energy (higher-energy) measurement on the LV (HV) side will result in a lower charge measurement than what is actually collected by the strip. The net effect on the LV side is an apparent line shift to lower ADC bins for low-energy lines. The apparent line shift depends on the line energy, introducing nonlinearity into the single-strip gain calibration.

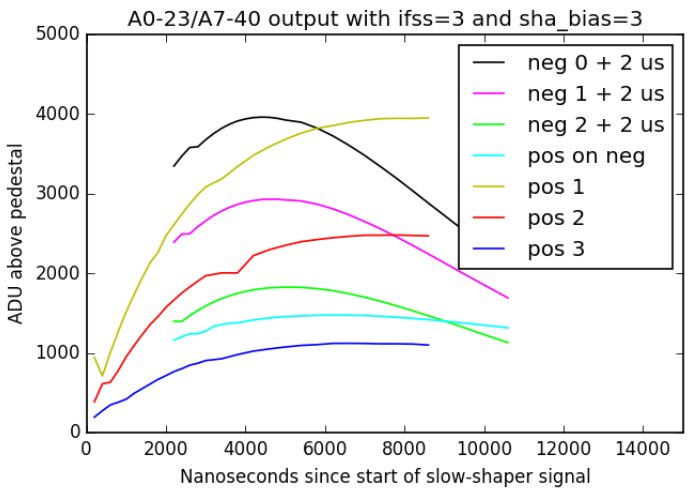

Figure 8.11: Plot showing how the slow shaper peaking time varies with input pulse energy for the GRIPS ASICs. The ADC bin values are given for different pulse energies as a function of sampling time on the slow shaper. The "neg" curves are pulses on the LV side, "pos" curves are pulses on the HV side and "pos on neg" is a positive charge signal on the LV side. A sampling time of 8000 us was chosen for the LV side and a time of 4000 us for the HV side.

The uniform sampling rate introduces non-linearity into the GRIPS gain calibration. This non-linearity primarily affects the lower-energy portion of the LV spectrum since the sampling time was explicitly chosen to match higher-energy peaking times. Above ∼200 keV the spectrum is linear. A good fit function would therefore be mostly linear with a non-linear tail extending into lower energies. For the purposes of this report a polynomial interpolation is provided that fully constrains the calibration data, but does not provide a general energy calibration function representing the true electronics response. The interpolation provides an adequate means to determine the bin-to-energy conversions within the calibration range and serves as a pathfinder for future efforts to fully characterize the true electronics response over the entire energy range.

Channel gains were fit to polynomials over the range of 59.5 keV to 662 keV. Figure 8.12 shows an example of the residuals as a percent of line energy by polynomial order (left) and the 5th order polynomial interpolation for detector 0 channel 234 (right). Similar interpolations were done for the by-ASIC line values using the maximum number of lines available for each ASIC. The residuals for the by-ASIC values are shown in figure 8.13. For each ASIC the lowest-order polynomial interpolation which fully constrained the calibration data was chosen to provide bin-to-energy conversions in this manuscript.

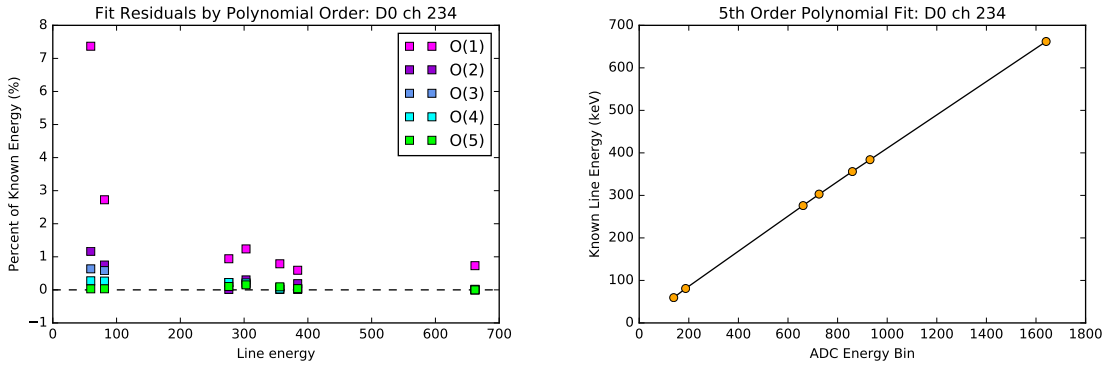

Figure 8.12: *Left:* Fit residuals as percent of line energy by polynomial order for D0 channel 234. The magenta points show the polynomial deviation from linearity. *Right:* A 5th order polynomial interpolation for channel 234.

8.4 Spectral Resolution

Individual-strip FWHM spectral resolutions are plotted in figure 8.14 for the low-voltage sides of detector 0 (unsaturated colors) and detector 4 (saturated colors). The FWHMs were derived from calibration data using radioactive sources of known line energies placed along the telescope line-of-sight and a polynomial interpolation was used to convert from ADC bins to energy. The modes are inset to each plot and given in keV with $\pm.5$ keV uncertainty. The spectral resolution is $\sim 3.5 \pm .5$ keV across the GRIPS energy range. The minimum FWHM line width is set by the ASIC, which has a manufacturer-reported equivalent noise at the input of $\sim$2.5 keV. The ASIC minimum resolution is greater than the $\sim$1.9 keV resolution (at 662 keV) reported by other groups with similar 3D-GeD detectors implemented using discrete analog components and a $4\times$ coarser electrode spacing (2 mm) [72]. Factors which led to the additional $\sim$1 keV spectral broadening in the GRIPS implementation above the 2.5 keV minimum resolution are discussed below and in the next section.

Tailing on the low-energy side of the 59.5 keV line was observed on both D0 and D4 (section 8.3), indicating that charge loss is an important factor for these detectors. Relatively low operating voltages ($\sim$2x their depletion voltages, see section 7.4) in the GRIPS detectors contributed to charge loss. Operating detectors at low voltages results in slower drift velocities, whose longer charge-collection times increase the likelihood of trapping/recombination. Trapping and recombination result in charge loss, generating tails on the low-energy side of lines [74].

For half of the detectors in the cryostat the LV side was oriented "up", pointing along the boom to where the radiation source was located. The other half of the detectors were oriented with the LV side pointed "down" away from the source. Since low-energy radiation tends to be absorbed by the crystal close to the incident surface, for low-energy photons the half the charge carriers must travel almost the full thickness of the crystal before being collected on the opposite side. The increased path length results in increased charge loss

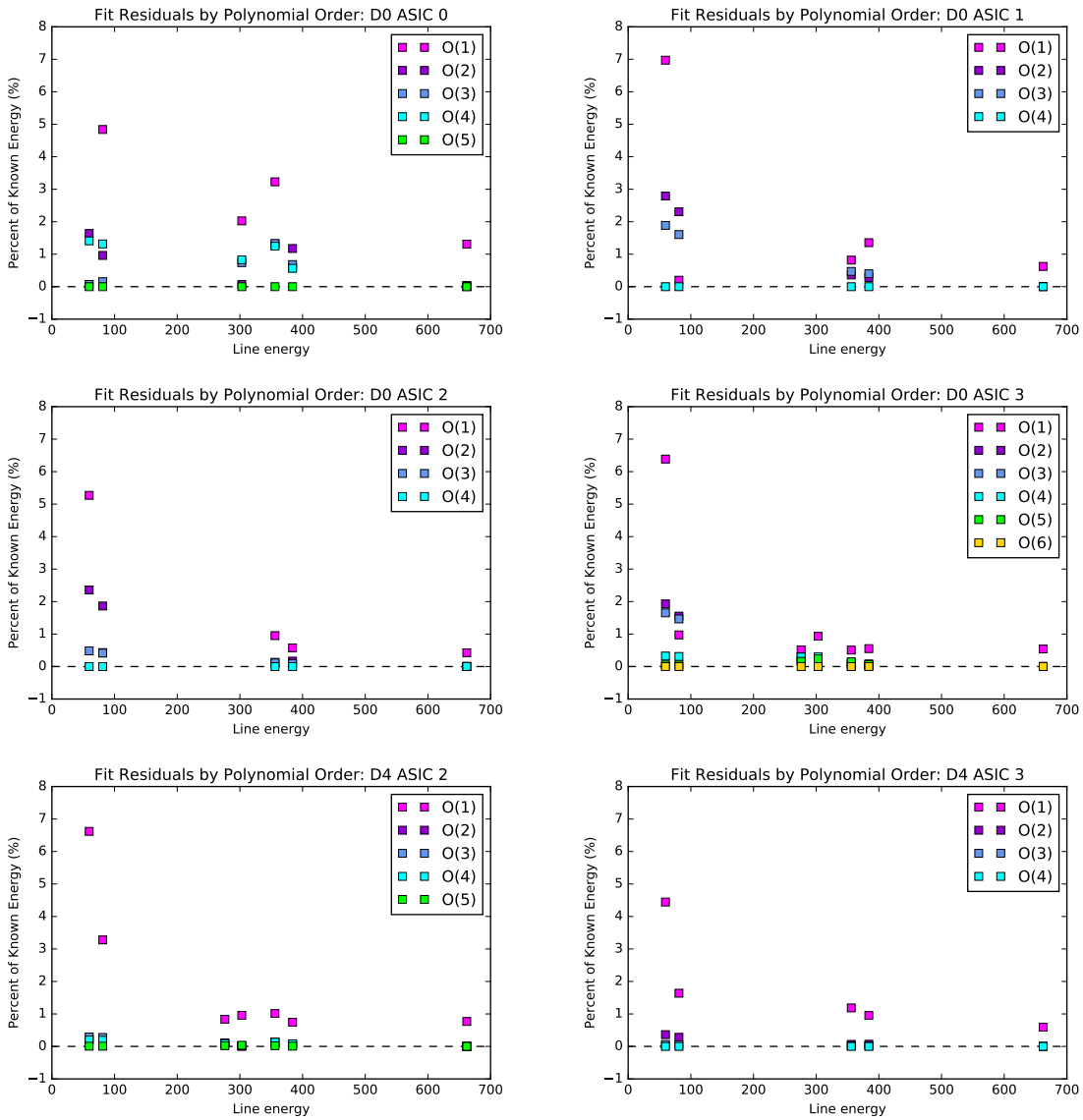

Figure 8.13: Fit residuals as percent of line energy by polynomial order for each ASIC on D0 and D4 LV sides. For each plot the last order plotted is the polynomial that was chosen for that ASIC's interpolation function.

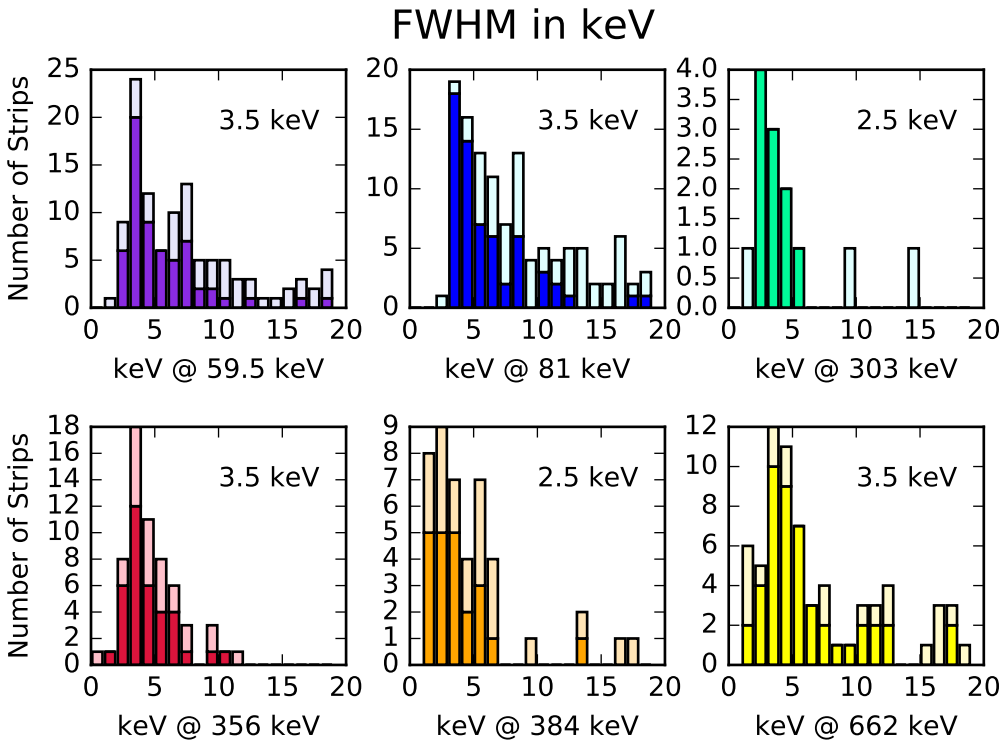

Figure 8.14: Spectral resolution of low voltage detector strips on D0 (unsaturated colors) and D4 (saturated colors) at the six calibration line energies of 59.5 keV, 81 keV, 303 keV, 356 keV, 384 keV and 662 keV. The mode for each line is inset to each plot and given in keV with a ±.5 keV error.

(due to trapping/recombination) and fewer single-strip energy depositions (due to spreading of the charge cloud) [74]. These effects will tend to broaden lines and increase tailing. The relative orientation of the source to readout side of D0 (LV "down") and D4 (LV "up") combined with the effects of low-operating voltages are the likely cause of D0's broader and fewer single-trigger spectral lines. The cryostat construction (with a 1" thick aluminum baseplate situated on the LV "down" side) prohibited taking low-energy calibration data from the anti-sunward orientation.

The method of common-mode subtraction (discussed in section 8.5) corrects for large-scale baseline shifts in the ASIC circuitry, reducing measured line widths by 10s of percent. With refinement of the common-mode method and channel grouping, the measured FWHM resolutions have the potential to narrow even further towards the ∼2.5 keV minimum set by the ASIC electronics.

8.5 Spectral Broadening

This section identifies several sources of spectral broadening, describes how they are rectified and identifies ways to further improve the GRIPS spectral performance with post-processing software.

Common-Mode Subtraction

In a process referred to as common-mode subtraction, large-scale baseline shifts are quantified and removed from the raw ADC bin values for groups of highly-correlated channels. Common-mode subtraction has been very successful, reducing line widths by 10s of percent. Refinements to the groups and technique are expected to further narrow our reported line resolutions.

The baseline voltages supplied by the chip to its internal stages suffered baseline shifts that were highly correlated between channels. These baseline fluctuations contributed to spectral broadening through the VA channel's ADC comparator. They are also a candidate for generating the false triggers on the TA channel's triggering comparator that occur during multiple-trigger noise (section 8.6). Other groups operating this family of ASIC chip also reported common-mode issues and the need for mitigation techniques (by private communication). Other sources of noise that systematically affect multiple electronics channels, such as the low-frequency noise caused by the mechanical cryocooler described later in section 8.5, are also reduced by the common-mode subtraction technique.

The common-mode groups were determined empirically. A 64 x 64 matrix of Spearman's rank order coefficients are generated as a correlation measure for all the channels on each ASIC. For each channel the 5 most correlated channels were chosen as its common-mode group. The groups are symmetry breaking: the 5 most correlated channels for a single channel are not necessarily most correlated with each other. Though the groups are not associated with a simple spatial layout in the ASIC, or on the detector, they persist over time and do not change significantly with detector or ASIC power cycling.

For each member of the group a scale factor is applied to the ADC bin value since the relative extent of noise experienced by each channel differs. After applying the scale factor, the median of the CMS group values is then subtracted from the triggered channel's ADC value. This algorithm could be improved by using a more advanced method of selecting the common-mode value instead of the simple weighted median. Discarding channels in the common-mode group which record a glitch or trigger coincidently could further refine this technique.

When a triggering event occurs all the ASIC channels switch to conversion mode and the entire detector's 298 channels are read out. The triggered channel, its six closest neighbor strips and the common-mode groups for all 7 channels are saved to disk for each channel that is triggered. For a single-strip trigger this scheme saves 12 strips when the common mode groups are highly symmetric and 42 strips in the worst case when there are diverse groups. The remainder of the conversions are not saved. The raw ADC values are recorded and

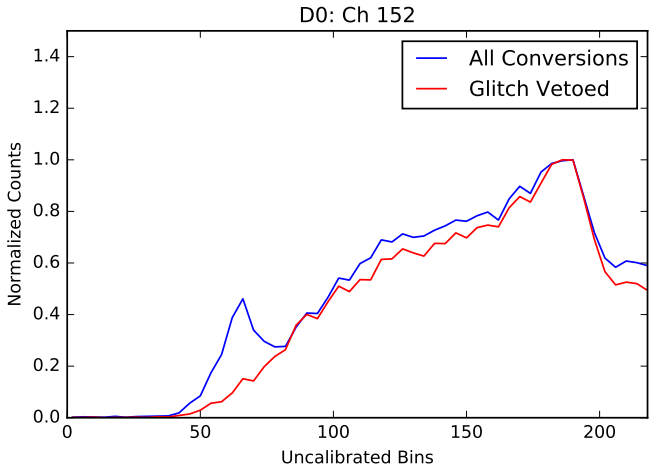

Figure 8.15: Ba-133 data for ASIC channel 152 on detector 0. This image shows how glitch rejection culls the extraneous peak at ∼bin 60.

common-mode subtraction is performed in post-processing analysis. This selective saving scheme was implemented to reduce the data load for flight.

Glitches

Each ASIC channel has a single output which switches between triggering and charge-conversion mode. After exiting the ASIC the signal travels over ∼25 ft of tightly packed ribbon cables before being read out by an FPGA. Upon receiving a trigger (a logic high) the FPGA sends a signal to the ASIC to change into charge conversion mode. During this time the ASIC outputs are kept at logic low and the FPGA counts clock cycles until the output returns a logic high, signaling the end of the analog-to-digital conversion. The quantity of clock cycles elapsed is the raw binning of the analog-to-digital conversion.

In the GRIPS nomenclature, glitches occur when there are voltage changes at the input of the FPGA that were not initiated by regular triggering and ADC signal processing. When the FPGA receives a logic high from the signal line during charge-conversion mode it waits to see if the signal change was a transient by checking if the signal goes low again. If it does return to logic low within a specified time interval the event is marked as a glitch.

Two types of glitches were observed. The first type of glitch is due to cross-talk between the signal lines when they switch from logic low to logic high with a fast edge. This glitch generates a localized peak in the strip spectra, as can be seen in figure 8.15 as a blue trace. The red trace shows how rejection for glitch-flagged events culls the extraneous peak at ∼bin 60, leaving a clean threshold edge and an untouched 81 keV line.

The second type of glitch occurs around the real conversion edge. This glitch is thought to result from ground bounce on the ASIC board that causes jitter in the comparator. In this case there are multiple transients around the real ADC edge, causing an apparent spread

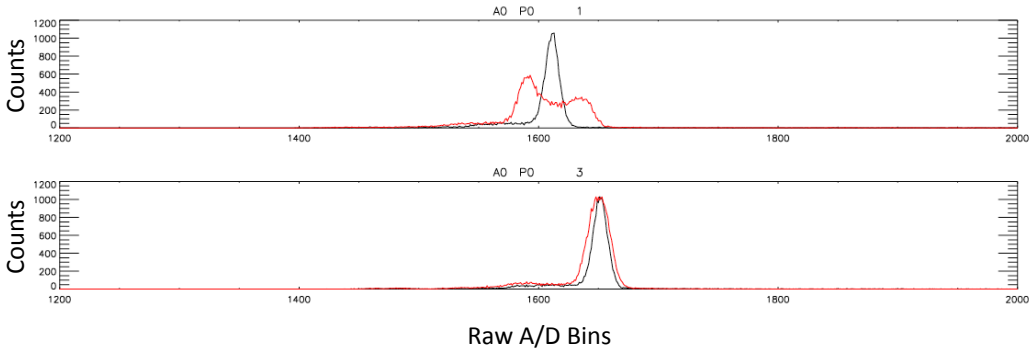

Figure 8.16: 59.5 keV spectra in raw ADC bins (labeled raw A/D bins) before common-mode subtraction is applied. The red curves are from data taken with the cryocooler on and the black curve is from data with the cryocooler off. The line is split into two lines by cryocooler effects for the channel in the upper plot and only minimally broadened for the channel in the lower plot. The two channels shown are are spatially close to one another, separated by one channel in the ASIC and by one strip on the detector. Despite the proximity of these channels to one another their responses to cryocooler noise are vastly different.

in the energy conversion. Glitch rejection of these events can narrow the line FWHM at the expense of rejecting real source data.

Each ASIC channel had a different glitch behavior determined by its susceptibility to cross-talk or ground bounce, resulting in glitch peaks and spectral blurring of varying magnitude across ASIC channels and detectors. In most cases glitch-rejected data was used for low-energy line studies of the threshold, Am-241 and Ba-133. All-conversion data was used for higher-energy lines of Ba-133, Co-60 and Cs-137. Development of a more selective glitch filter, to replace the current "ignore"/"don't ignore" all glitches, could retain more "good" counts in the low-energy calibration spectra providing better statistics for line fitting.

Cryocooler Noise

A mechanical cryocooler (Sunpower Cryotel CT) was chosen to cool the GRIPS detectors to the ∼80 K required for operation. A DC voltage of 28 V was input into the cryocontroller which created a 60-Hz rectified sine-wave output with a maxmium 15 A peak-to-peak current for input to the cryocooler. After integration into the cryostat 60 Hz interference was observed to varying extent across numerous signal traces. Figure 8.16 shows the response of two channels on D4 to the cryocooler. The red curve is for Am-241 59.5 keV radiation with the cryocooler on and the black curve shows the spectra with the cryocooler momentarily switched off. These channels are located physically close to one another on the detector (one strip between them) and in the ASIC (one channel between them) but have different responses to the cryocooler interference. For the top channel the true line is split

into two peaks which alternate being filled with counts at a rate of ∼60 Hz. The second channel suffers slight line broadening due to the cryocooler. The ASICs, which handle signal processing from the front-end pre-amp through the analog-to-digital conversion, were not designed to provide significant low-frequency filtering and additional stages could not be added to attenuate the 60-Hz noise. Appropriate choice of the common-mode groups and shielding effectively reduced the bulk effects of cryocooler interference (e.g., line splitting), but the finer effects of cryocooler noise contributed to spectral broadening.

8.6 Multiple-Trigger Events

Each GRIPS detector is run on coincidence mode between its high-voltage and low-voltage sides. If multiple channels trigger within a single coincidence window (600 μs) each channel is marked as a trigger to that single event. Multiple triggers due to physical effects such as charge sharing or Compton scattering through the same crystal can generate a handful of coincident triggers. Coincident triggers can also be generated by noise events that cause the FPGA to read a logic high value on the signal line during the coincidence window. These noise triggers could be generated at any point in the triggering electronics chain, ranging from microphonics on the input of the pre-amp, noise on the triggering comparator or glitches along the signal line feeding into the FPGA.

An on-board firmware veto was implemented in the FPGAs to catch and process multiple-trigger events caused by noise. Since multiple-triggers due to real energy deposition would only cause a few channels to trigger, the software veto was based on the number of channels triggered per event. In the case of a single-crystal, two-site energy deposition with charge sharing across 5 electrode strips, for each deposition there would be a total of 10 strips that would record a real trigger. Using this case as a guideline the software veto caught events with greater than 10 – 25 triggers. Based upon system performance and the severity of multiple-trigger noise the software veto was configured as either a soft veto which simply flagged the event (D0, D1, D2 & D3) or a hard veto which did not save data from the event (D4 & D5). Most of the analysis presented in this document discards multiple-triggered events from the data.

Figures 8.1 and 8.17 (zoomed in to show the D0 fast timescale structure in detail) show the time between events for noise counts in comparison to real source counts. In figure 8.17 the total events (dark purple points) are broken into components based on the triggering order. Any event that received more than 10 triggers was considered a multiple-trigger event ("mt") and events with fewer triggers as source-trigger ("st") events since real source depositions can potentially trigger 10 electrodes. The components plotted are: the times between source-triggers followed by multiple-triggers (st-mt, red), multiple-triggers followed by multiple-triggers (mt-mt, green), and the combination of source-triggers followed by source-triggers and multiple-triggers followed by source-triggers (mt-st + st-st, violet). The first two components (st-mt & mt-mt) are due to multiple-trigger noise and the counts in the summed component (mt-st + st-st) are from the source. This plot indicates that multiple-

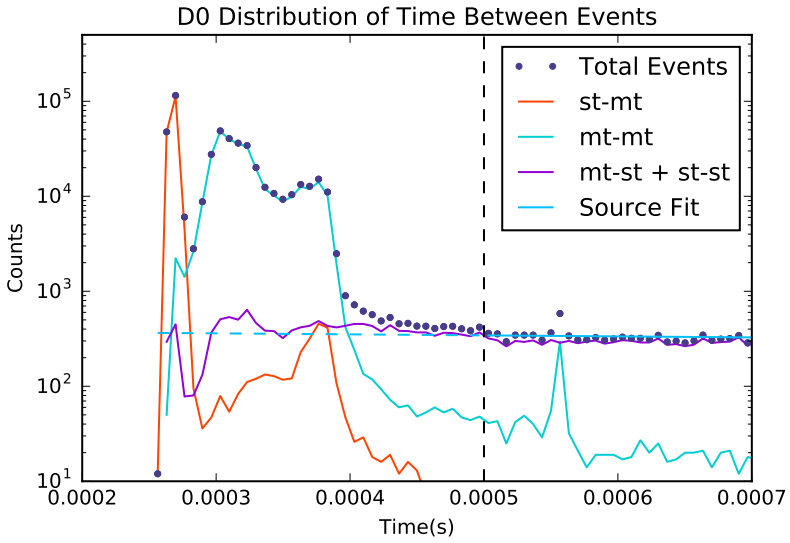

Figure 8.17: Distribution of the time between events recorded by D0 for Am-241, zoomed in to show the fast timescale detail from fig 8.1. The total events (dark purple points) are broken into components based on the triggering order. The legend abbreviates single-triggers as "st" and multiple-triggers as "mt". Any event that received more than 10 triggers was considered a multiple-trigger event since real source depositions can potentially trigger 10 electrodes. The components plotted are: the times between single-triggers followed by multiple-triggers (st-mt, red), multiple-triggers followed by multiple-triggers (mt-mt, green), and the combination of single-triggers followed by single-triggers and multiple-triggers followed by single-triggers (mt-st + st-st, violet). The first two components (st-mt & mt-mt) are due to multiple-trigger noise and the counts in the summed component (mt-st + st-st) are from the source. The data was fit with a decaying exponential (solid blue line) to the right of the dashed black line, to predict the source count rate. The fit has been extended into the region dominated by noise (dashed blue line) to show the expected source counts in this region. The fitted line matches the summed component (mt-st + st-st) well.

trigger events occur quickly after source-triggers (∼275 us) and then the system can receive successive multiple-trigger events through ∼400 us before relaxing. It is not immediately obvious what causes the regular timing of the three main noise peaks, nor why the st-mt and mt-mt have later peaks as well. This regular timing could point to where in the spectrometer system multiple triggers originate.

Figure 8.18 shows the histogrammed time between event counts (purple points in of 8.17) as a function of total triggers per event. Figure 8.19 shows the same histogram of time between event counts as a function of total triggers per event over the extended time range shown in figure 8.1. The noise peaks from figure 8.17 are clearly visible in these plots. These plots indicate that the mechanisms responsible for multiple-trigger noise could be different on the LV and HV side. The spread of multiple-trigger events on D0's LV side suggests that

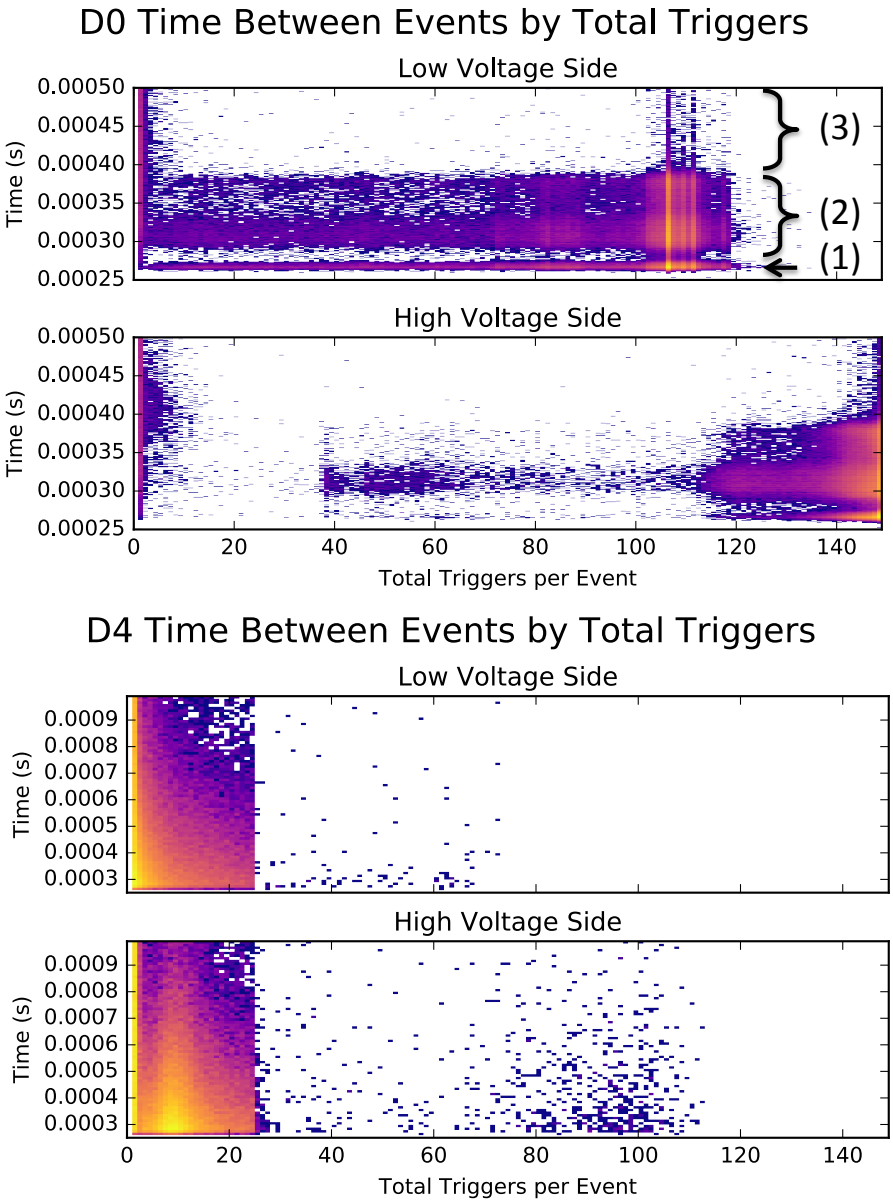

Figure 8.18: The total detector counts from an Am-241 source (same data as the purple curve of figure 8.1) as a function of total triggered events and the time between events. D0 labels the three different noise structures that are referenced in text. D0 was run in soft-veto multiple trigger mode (all events saved) and D4 was run in hard-veto multiple trigger mode (events with >25 triggers discarded). Though D4's hard veto caught most multiple trigger events above the threshold, occasionally some events with triggers greater than the threshold were saved. A key difference between noise on the two detectors is the persistence of events, with noise dominating the counts about 2x longer on D4 (1ms) than on D0 (.5ms).

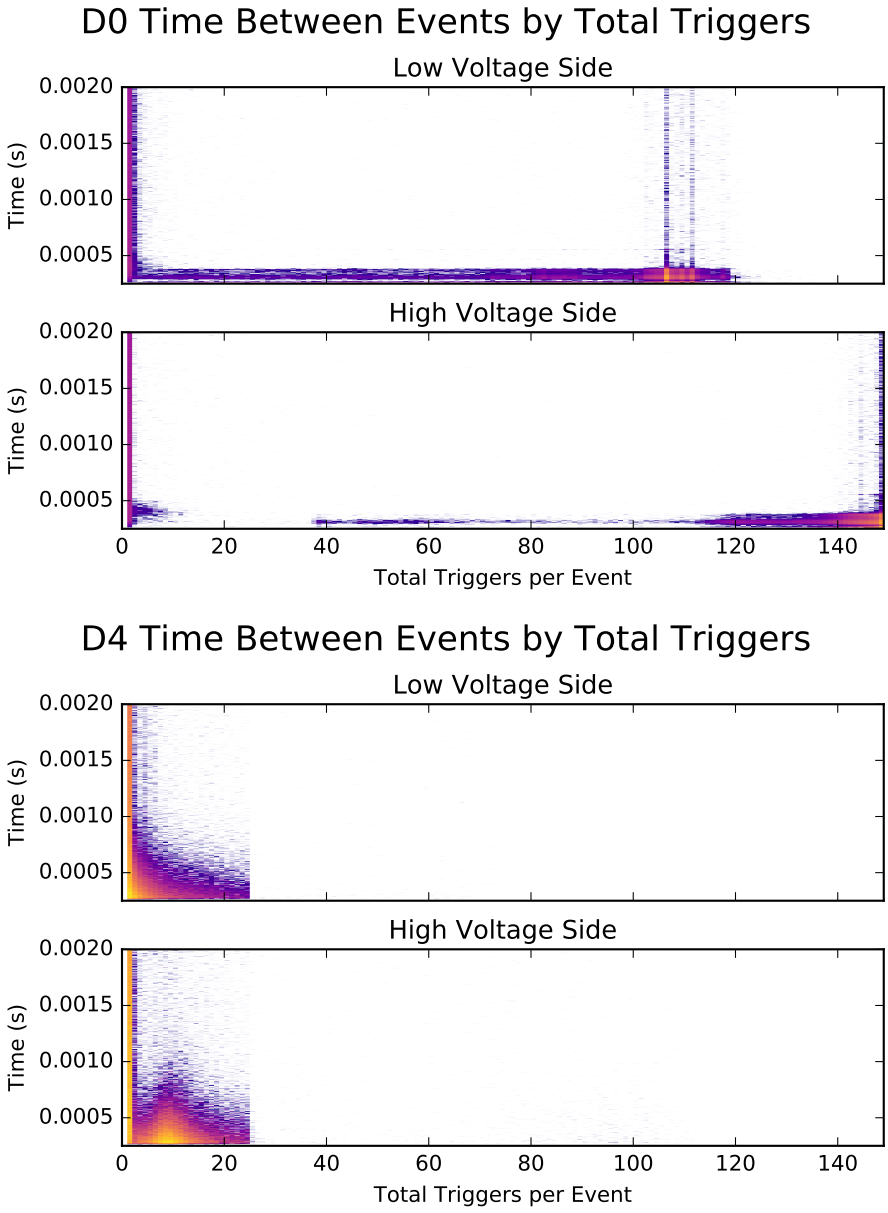

Figure 8.19: The total detector counts from an Am-241 source (same data as the purple curve of figure 8.1) as a function of total triggered events and the time between events. These plots extend over a longer timescale than figure 8.18, shown to include some of the longer time between events where source counts dominate. At later times the source counts primarily trigger only a few strips, as expected for relatively the low-energy (59.5 keV) radiation in this dataset.

the multiple-trigger noise can occur for any number of channels within a single ASIC and coincidently between ASICs, extending up to the entire side (which is ∼120 channels since there are non-triggering channels spread across this side, see figure 7.4). The situation on D0's HV side is markedly different. Here, there is a gap between the few triggers expected for source counts and the multiple-trigger events that begin at ∼38 triggers. The total triggers are sparse until ∼120 channels with most of the multiple-trigger events including the entire side of the detector. This could indicate that the multiple-trigger noise is not locally generated inside a single ASIC on the HV side, but is caused by something that affects an entire ASIC uniformly.

Detector 0's multiple-triggers have three primary components: (1) a fast multiple-trigger component (st-mt curve in figure 8.17 and first horizontal structure in figure 8.18) which comes in very quickly after a source-trigger that is characterized by any number of channels triggering together and possibly a result of the baseline fluctuations, (2) a component which comes in slightly later (mt-mt curve in figure 8.17 and second two horizontal structures in figure 8.18) in which almost all the channels on the detector trigger together, also attributable to noise and (3) a component which persists to long time-scales and causes the entire detector to record a trigger (mt-st component, vertical structures on the right of figure 8.19, top). Detector 4 shows a similar fast component to D0 (1) though it persists nearly 2x longer. On detector 4 any possible whole detector effects are removed by its hard multiple-trigger veto.

8.7 Summary

A survey of ∼300 strips across detectors 0 & 4 low voltage sides provided an initial detailed study of the spectrometer performance. Detectors 0 & 4 were found to have 12.3% and 13.8% efficiency with ∼3.5 keV spectral energy resolution and a rough .5×.5×1 mm spatial resolution. Duty cycling and multiple-trigger events reduced livetimes and complicated event selection, but post-processing software and firmware vetoes were successful in mitigating their effects. The baseline spectral resolution was set to ∼2.5 keV from the ASIC electronics. A combination of large-scale baseline shifts in the ASIC, cryocooler interference, glitches and low operating voltages contributed to an additional 1 keV spectral broadening. Improving the common-mode subtraction technique is expected to narrow lines further. Deviations from non-linearity in energy calibrations were expected due to energy-dependent peaking times in the ASIC electronics. ADC bin-to-energy conversion functions, which fully constrained the 59.5 – 662 keV energy range, were provided. These ground calibrations were shown to remain valid in flight by predicting the energy of the 511 keV instrumental background line (chapter 6). Energy calibrations were complicated by low single-site energy depositions due to charge sharing, line tailing from low operating voltages and the need for long integration times from duty cycling. Coarse resolution at strip-pitch granularity (.5 mm) was confirmed by tungsten shadow images and finer inter-strip resolution is anticipated in future work that will consider charge sharing between strips.

Chapter 9

GRIPS Imaging and Aspect Systems

In GRIPS' HXR/gamma-ray energy range indirect imaging provides higher spatial resolution than focusing optics (>100 keV) or Compton reconstruction techniques. GRIPS' imaging system, the Multi-Pitch Rotating Modulator (MPRM), is a single-grid imager that builds upon heritage from RHESSI's bi-grid Rotating Modulating Collimators (RMCs) [68]. High-precision detector spatial resolution enables the single-grid design, increasing photon throughput by $\sim$2$\times$ over related bi-grid designs. A review of HXR/gamma-ray imaging techniques is provided in chapter 4 and RHESSI RMC imaging is discussed in chapter 5.

The imaging aspect system tracks the positioning and orientation of the mask relative to the detectors for image reconstruction, and correcting for the twist & flexure of the 8 m telescope boom. The aspect system also provides high-cadence monitoring of the instrument pitch, yaw and relative roll, enabling 12.5" imaging with a telescope pointed to .5° rms.

9.1 The Multi-Pitch Rotating Modulator (MPRM) Concept

GRIPS' rotating grid partially obscures the Sun, selectively blocking/transmitting photons with alternating opaque slats and transparent slits (figure 9.1, left). The shadow of this pattern falls on the detector, whose spatial resolution is 1/2 the smallest grid pitch. The imaging resolution (R) is set by the ratio of the grid pitch (p) and the mask-to-detector separation (L). An 8 m telescope boom separates the mask from the detectors and the pitches vary between 1 – 13 mm, setting an imaging resolution of 12.5 – 162". The availability of detectors with .5 mm spatial resolution enabled this this fine imaging resolution to be designed with parameters achievable on a balloon platform.

$$R = p/2L = 12.5 - 162" \tag{9.1}$$

The source flux is spatially modulated at a regular rate by rotating the grid at $\simeq$10 rpm. The resulting flux represents a convolution of the grid/rotation parameters with the spatial properties of the source photon distribution. This modulated flux is backprojected

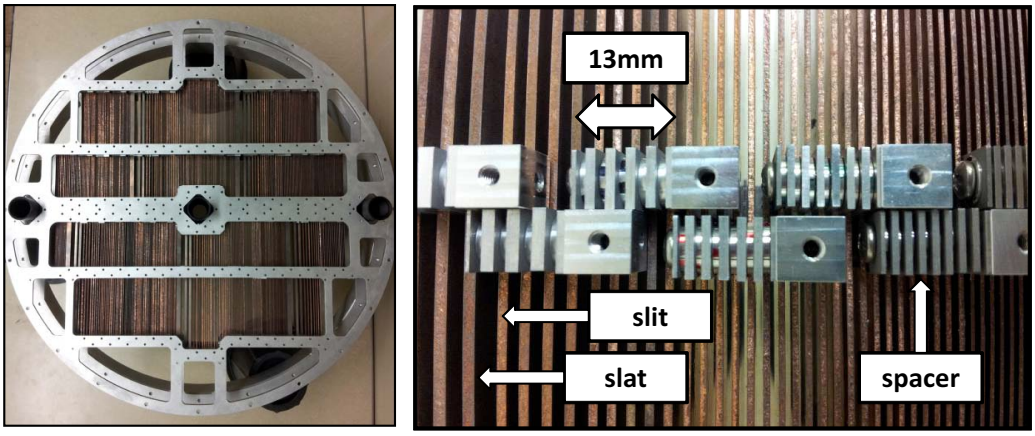

Figure 9.1: *Left:* The single modulation grid used for GRIPS' MPRM imaging system. The grid has thirteen slit/slat pitches made from 70/30 tungsten/copper slats of 2.5 cm line-of-sight thickness. The slit pitch varies from 1 – 13 mm, corresponding to thirteen imaging resolutions between 12.5 – 162 arcseconds. *Right:* Zoomed-in picture showing the slit/slat bundle arrangement on the imaging grid. The sum of all pitches in a bundle is 13mm. Transmission through the slits falls off as as line-of-sight angle increases. Pictures from Duncan et al. 2013 [44].

onto the Sun to recover the original photon distribution. GRIPS' point response function is virtually free of the side-lobes (figure 6.5, left) that are characteristic of RHESSI at gamma-ray energies. By sampling 13 frequencies simultaneously GRIPS can generate high-fidelity backprojection images over a range of spatial scales within a few hundred photons.

A schematic illustration of backprojection for a point source using the MRPM grid is shown in figure 9.2. In this method the image is iteratively built-up by adding together backprojection maps for individual photons at different grid pitches and orientations. Panel (a) shows a single photon map backprojected onto the Sun. The red strips indicate the transparent slit locations mapped to the solar disk, indicating where the photon could have originated. Since the opaque slats block photons, the black strips mark where photons could not have originated from. As the grid rotates and more photons are collected in various grid orientations these slit/slat maps are added together. The successive panels show the result created by 3 (b), 10 (c), 30 (d), 100 (e) and 1000 (f) photons as the grid rotates. This summed backprojection map can be further refined by using deconvolution techniques such as CLEAN. This backprojection + deconvolution prescription is the basic method used for RHESSI imaging analysis (chapter 5) and the GRIPS effort builds upon its heritage.

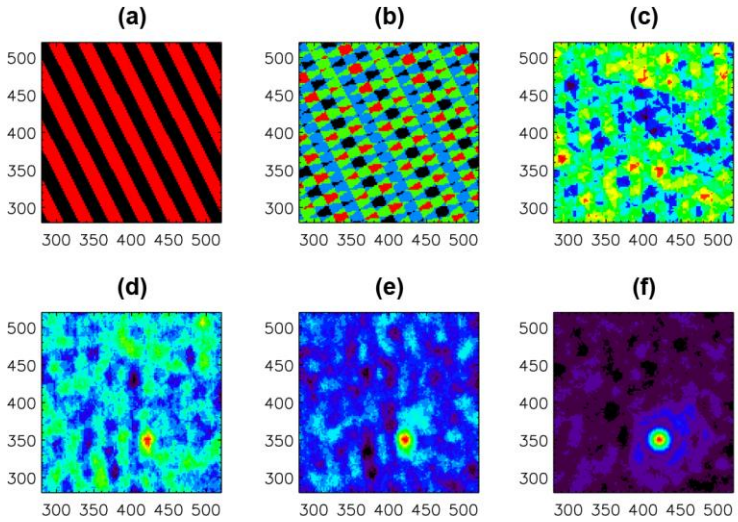

Figure 9.2: Schematic illustration of GRIPS image reconstruction, for 1(a), 3(b), 10(c), 30(d), 100(e) and 1000(f) counts from a simulated point source. As the grid rotates photons are collected from a range of grid orientations and spatial frequencies. Image from Duncan et al. 2013 [44].

9.2 Mask Construction

The opaque slats on the MPRM imaging mask are arranged into bundles of a single pitch (figure 9.1, right). The sum of all slats + slits in a bundle is 13mm, and each bundle type contains a different total number of slats & slits. A 1 mm pitch has 13 slats & 13 slits arranged into a bundle, whereas a 13 mm pitch has 1 slat & 1 slit per bundle. A series of all 13 pitches constitute a period, and there are 2.5 periods on the mask. The periods are offset from one another provide maximum spatial frequency coverage across a detector plane.

Blades are interwoven with spacers at three points along the blade length, creating a well-defined slit metrology. These three clamping points are also used to affix the bundle to a mask structure. A belted motor system rotates the grid via an outer ring on the mask circumference.

The grid is constructed from 491 off-the-shelf blades composed of a 70/30 tungsten/copper alloy, assembled into 70 total bundles that are installed onto the mask structure. The blades are nominally 20.3 cm long and 2.54 mm deep, making them opaque to HXR/gamma-rays along the telescope line-of-sight. The single-grid design does not require the tight manufacturing constraints of bi-grid designs, allowing GRIPS to adopt a design perspective that minimized production costs by using loosely toleranced, off-the-shelf tungsten blades. 18 measurements per blade carefully mapped their physical characteristics and the remaining mask hardware was customized to fit the blades' aggregate properties. The grid metrology

was photo-characterized on both sides at multiple orientations. Slat & slit locations derived from the images will be used to create a model of the grid to be used with backprojection.

9.3 Aspect System

The GRIPS aspect system provides the high-cadence, high-resolution pointing knowledge needed for 12.5" imaging resolution and ∼arcsecond image placement with a telescope pointed to only .5° rms. Precision monitoring of instrument pointing allows for relaxed pointing-control and mechanical requirements of the telescope system (e.g., boom twist and flexure). The aspect system makes three primary measurements: the pitch/yaw offset of the optical bench to Sun center, the orientation of the MPRM, and the absolute roll of the optical bench. The aspect system implementation has undergone significant design changes since it was described in early literature [122, 44], in part due to lessons learned from the Solar Aspect System (SAS) of the HEROES balloon instrument [33], whose design was adapted from the original GRIPS design. The HEROES project also contributed components of the roll sensor to the GRIPS system. A complete description of the final aspect implementation is also available in a later manuscript [45].

A single subsystem simultaneously measures pitch/yaw offset and MPRM orientation. This subsystem uses three 3-cm diameter plano-convex BK-7 lenses (focal length of 7780 mm) mounted across the MPRM mask diameter (figure 9.1, left) to project three images of the Sun onto a *fiducial screen*, or *fiducial plane*, mounted to the top of the cryostat (figure 9.3). The fiducial plane is monitored by a 1280 × 960 pixel CCD camera (Prosilica GT1290) that monitors the motion of the projected Suns. Since the solar images and fiducials contain all the necessary information to determine pitch/yaw and orientation, there are no stringent alignment requirements on this camera itself beyond including the aspect target within the camera FOV.

The pitch/yaw offset of the optical bench to Sun center is determined by measuring the position of the central solar image relative to the center of the cryostat via the fiducial plane. The fiducial plane coordinates are well-constrained to the detectors mechanically by the cryostat itself. Crossed tungsten rods embedded in the cryostat top (and visible in the pitch/yaw subsystem images) cast X-ray shadows onto the detectors, linking the pitch/yaw coordinate system to the detector spatial coordinate system. The boom width limits the pitch/yaw field of view to be within a few degrees of the Sun. With the effective plate scale of 16" per pixel, the pitch/yaw knowledge accuracy is better than a few arcseconds.

The MPRM grid orientation is determined by tracking the three solar images as they rotate on the fiducial screen. The two outrigger lenses have neutral-density filters that reduce the intensity of their images to 90% and 80% of the central image. The intensity is a key to distinguish between orientations that are 180° apart, making the total set of images nondegenerate. During nominal pointing, as the MPRM rotates at ∼10 rpm, one of the outrigger images may fall beyond the fiducial plane and outside the camera FOV. The MPRM orientation can still be determined by the remaining two images of the Sun.

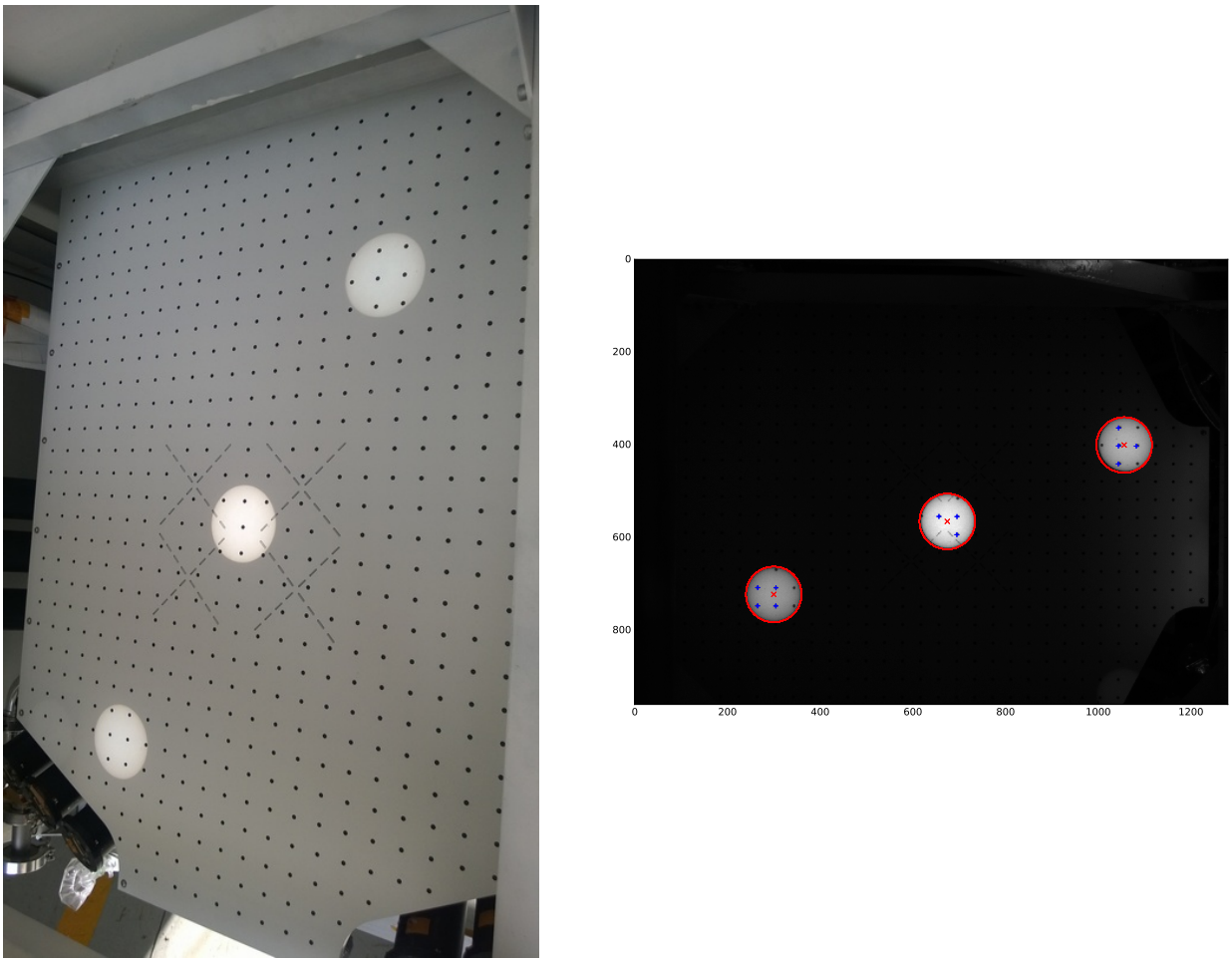

Figure 9.3: *Left*: Photo of the fiducial plane (which is mounted to the top of the cryostat) during a pre-flight pointing test. In the photo are the three solar images, the fiducial plane, and the "×" patterns that are positioned above each stack of 3D-GeDs. The solar images in the lower left and in the upper right are at 90% and 80%, respectively, of the brightness of the central solar image. As the MPRM rotates, the outer solar images appear to orbit around the central solar image. *Right*: In-flight image from the pitch/yaw subsystem, with locations of the solar images (red × and circles) and illuminated fiducials (blue crosses) determined via a convolution filter. The right edge of the frame corresponds to the nadir direction, and shows faint illumination from incompletely blocked Antarctic albedo. Note also that the black tungsten rods, aligned with the "×" patterns, are illuminated by the central solar image. Pictures from Duncan et al. 2016 [45].

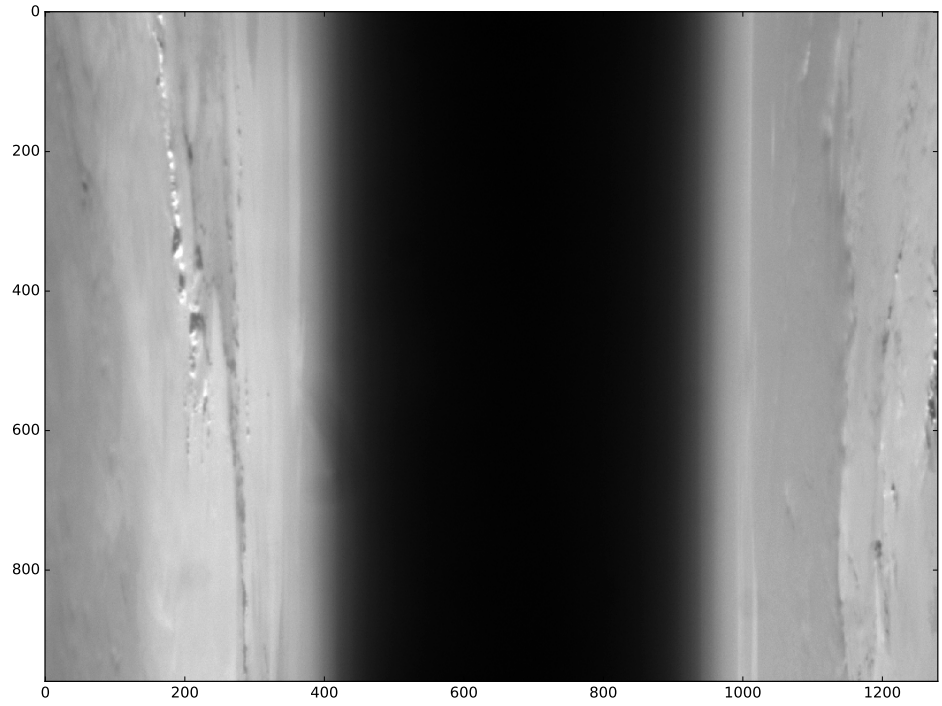

Figure 9.4: In-flight image from the roll sensor showing the opposing horizon views and the Antarctic surface. The axes indicate pixel number. Picture from Duncan et al. 2016 [45].

The fiducial plane is a thin aluminum sheet painted flat white that is bolted to the top of the cryostat, which is painted flat black. The high-contrast fiducials are created by precision machining holes into the white fiducial plane, revealing the black cryostat top below. The crossed tungsten rod images were similarly generated, by cutting channels in the white aspect screen to reveal the black painted rods. This implementation not only takes advantage of machining tolerances vs screen printing tolerances, but it also avoids the speckling observed in the HEROES SAS images that resulted from the partial glossiness of screen-printing inks. The HEROES SAS experience also showed that the originally planned expanding grid of fiducials added an unnecessary level of complication as part of a mostly stable system with the entire fiducial screen within the camera FOV, so GRIPS used a simpler grid of constant spacing between its fiducials.

The absolute roll of the optical bench is determined by a separate subsystem mounted above the gondola. The roll sensor combines opposing views of Earth's horizon into a single image (figure 9.4). Locating the relative positions of the horizons in the image provides a measure of instrument roll. A silvered, knife-edge, right-angle prism mounted below a 1280×960 pixel CCD camera (also a ProSilica GT1290) combines the two horizons into a



Figure 9.5: The roll sensor, mounted above the gondola in the orientation shown. The silvered, knife-edge, right-angle prism combines opposite horizons into a single image.

single image (figure 9.5). This system can monitor absolute roll continuously, without the measurement drift that can occur in gyroscopes. With the plate scale of 1.1 arcminutes per pixel, the roll knowledge accuracy is better than 15 arcseconds. The HEROES project contributed the roll sensor's physical assembly and the mirrored prism, with the primary change being the exchange of the CCD camera.

Pre-flight calibration of the pitch/yaw subsystem occurred in tandem with the pointing control system. The wedge angles of each of the three lenses were calibrated by individually rotating each lens through four 90-degree increments while keeping the other two lenses fixed. Pre-flight calibration of the roll subsystem consisted of using a theodolite as a visual feature within the camera's FOV and measuring the precise optical angular position of the theodolite using the theodolite itself. The association of the pixel coordinates of the theodolite and the corresponding precise optical angles over a dozen positions of the theodolite allows determination of the plate scale and distortion of the roll-subsystem camera.

Camera images were saved at 4 Hz to a dedicated aspect computer running Ubuntu Linux. The framerate was chosen to meet science requirements for pointing knowledge. The aspect computer provided the command interface for controlling the cameras, produced real-time telemetry, and recorded the images for post-flight analysis on three $\sim$1 TB solid state drives. To provide enough storage for a potential month long flight, images were saved using

a combination of Rice compression and row decimation to reduce the storage needs with minimal impact to science.

9.4 Summary

The imaging and aspect systems performed nominally in flight. The pitch/yaw subsystem verified the performance of the pointing control system and provided a fine aspect solution. The roll subsystem measured the expected variety of roll motions, including a 24-second harmonic motion, or pendulation, of the entire balloon-payload system with a peak-to-peak amplitude of $\sim$1 arcminute.

The largest flare that occurred during GRIPS' flight was below the limits of sensitivity for the spectrometer using the event selection strategies described in chapters 7 and 8. Improved event selection and background reduction via software tools or Compton techniques has the potential to provide the $\sim$100 photons needed to generate an image from flight data. If there are not enough photons to generate an image from a flare, the imager could be tested using a well-collimated beam line. In this case, a single 3D-GeD detector could be used with the mask to validate the concept in a scaled test.

Chapter 10

Conclusion

The HXR/gamma-ray flare spectrum is rich with features that can be used to understand the dynamics of accelerated particles and the properties of ambient plasma. The most detailed gamma-ray imaging and spectroscopy to date has come from RHESSI, which provided two key advances for high-energy flare observations: high-resolution spectroscopy capable of resolving the individual ion lines and gamma-ray imaging. With individual line fluences and shapes the ratios of accelerated particles could be determined and the total energy content of ions estimated in ∼30 of the largest RHESSI flares. Relativistic electrons and ions were found to be proportionally accelerated, and their energy content shown to be tens of percent of the total energy liberated during a solar eruptive event. In one of the largest flares, fine resolution of the 511 keV line showed that the annihilation region could reach an unexpected combination of transition region like temperatures while at chromospheric densities.

RHESSI provided the first gamma-ray images of ion emission, showing compact sources associated with the flaring region which are not consistent with the SMM-era suggestion that ions are accelerated by large-scale structures. These images also revealed the unexplained displacement between ion-associated gamma-ray emission and electron-associated HXR emission. RHESSI's limited imaging resolution at gamma-ray energies made these imaging observations possible for five flares, and only one of those events was large enough to separately resolve the individual 2.2 MeV footpoints. RHESSI provided a wealth of observations which indicate there is more to be learned about particle acceleration and transport in flares. Expanding upon RHESSI's small gamma-ray flare dataset with additional observations and enhanced capabilities is required to build a complete picture of ion and relativistic electron dynamics.

Flares that are obscured by high and irregular background are poorly studied in RHESSI literature. Chapter 5 explored extending the method of imaging-spectroscopy into the gamma-ray energy range to extract spectra from background in highly obscured flares. This method is regularly utilized in the lower-energy HXR spectrum where there are many source counts to generate images with narrow energy binning and high signal-to-noise with the full dynamic range provided by RHESSI's nine imaging collimators. In the photon-limited gamma-ray range, however, imaging is complicated by poor statistics and a reduced set of

two imaging grids. Imaging-spectroscopy was shown to reproduce the key features and fluences of a gamma-ray spectrum for two well-studied flares. Applying this method to the 2003 October 29 flare yielded a clean spectrum with the expected gamma-ray features. The spectrum was fit with relativistic electron and ion components and the fluences were used to add this flare to Shih et al.'s 2009 correlation study results. This flare's spectroscopy results fit into the ensemble of other larges flares well; its 2.2 Mev and relativistic electron fluences fall into the tight linear proportionality observed in other gamma-ray flares.

The imaging-spectroscopy results indicate that the amount of locally generated 511 keV on RHESSI might be overestimated. Since the 511 keV line provides a rich diagnostic for the ambient plasma conditions, investigating this feature between the two methods as a means to constrain RHESSI's response matrix would be beneficial. This could be done by systematically comparing the 511 keV line fluxes in the two methods over many flares.

GRIPS is a next-generation flare observatory whose technology was successfully demonstrated on a long-duration balloon platform in a 2016 Antarctic flight. GRIPS' indirect imager and 3D-position sensitive germanium detectors provide a gamma-ray dedicated system capable of increasing the throughput of gamma-rays by $\sim 2\times$, improving imaging resolution by $\sim 3\times$ and providing a means to detect polarization while maintaining similar high-resolution spectroscopy to RHESSI.

During GRIPS' $\sim$12 day flight 21 C-class flares occurred. The largest event was a C9.6 which was below the limits of detectability for the spectrometer using the currently implemented event selection techniques. Improved event selection and noise reduction through software filters or Compton techniques could provide a means to recover the $\sim$100 photons needed to generate an image from flare data. With the current event selection and software tools, the spectrometer had an efficiency of $\sim$13%, a FWHM line energy of $\sim$3.5 keV and a $.5\times.5\times1$ mm detector spatial resolution. Chapters 6 – 9 described the GRIPS instrument, its systems, their performance, and Antarctic flight in detail.

The future of GRIPS is uncertain; much of the path forward will depend upon evaluation of the recently recovered spectrometer and electronics. Since the heavy instrument frame was not recovered, it could make sense to redesign the GRIPS suborbital mission concept to fly aboard an ultra long-duration balloon (ULDB) flight. Aboard a ULDB, whose flight times could be in the range of months, a GRIPS balloon flight would have a better likelihood of catching gamma-ray flares during the next solar maximum. The GRIPS technology has the most utility in space where it can monitor the Sun quasi-continuously. A GRIPS satellite observatory with a 20 m boom could provide gamma-ray imaging of 5" or better and can be realized with technology available today.

The Sun is a dynamic laboratory that provides unparalleled opportunities to study high-energy astrophysical processes with precision and detail. The tumultuous solar atmosphere is a window that can offer high-resolution images of acceleration processes that drive particles to relativistic speeds, heat plasma to millions of kelvin and catastrophically reorganize 10^4 – 10^5 km long field structures in the span of minutes. No other astrophysical object has been studied to a similar level of detail as the Sun and still we find that there is more to learn about our nearest star.

Appendix A

GRIPS Education and Public Outreach

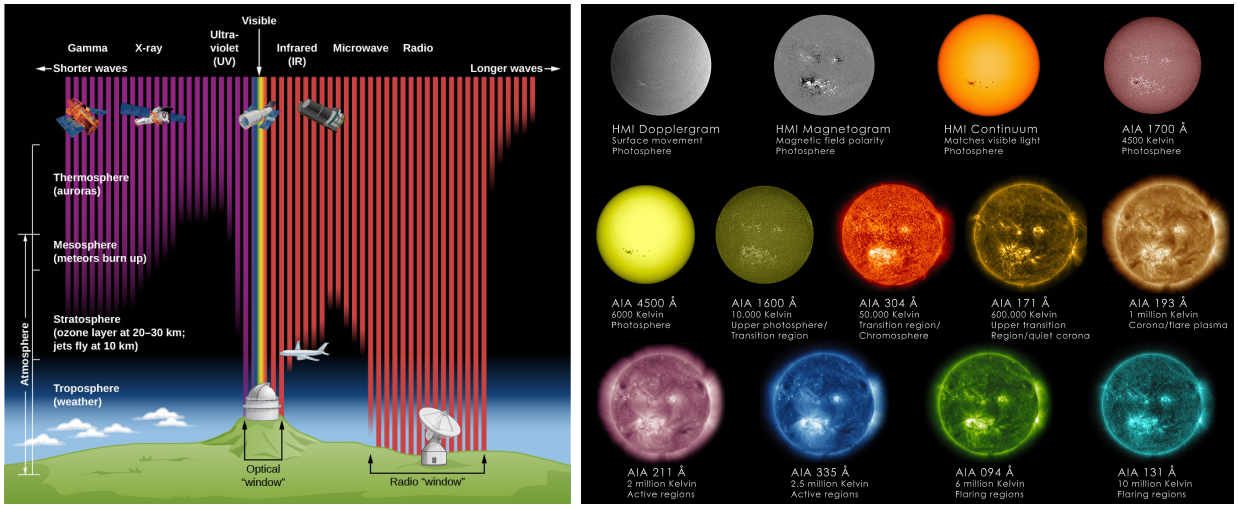

Figure A.1: *Left:* Image illustrating the penetration depth of light into the atmosphere at different energies and why its possible to make ground based measurements of optical and IR light, but not HXR or gamma-rays. Balloons, rockets and satellites are needed to study high-energy light. Image courtesy of OpenStax Astronomy as a modification of a STScI/JHU/NASA image [7] *Right:* Simultaneous images of the Sun, showing how different energies of light reveal different phenomena, regions and temperatures. Image from the SDO archive [116].



Figure A.2: *Left, top:* GRIPS suspended from "The Boss" launch vehicle during a hang test. *Left, middle:* GRIPS team signatures on the cryostat, a tradition before flight. *Left, bottom:* Balloon filling with $2/3^{rd}$ of a ton of helium prior to launch. The helium tanks are visible on trucks to the right of a tractor. Image courtesy of Albert Shih. *Right:* GRIPS just after launch. The payload is suspended from a parachute (orange) attached to the balloon. After reaching the Stratosphere, the balloon will expand to 150 ft tall. Image courtesy of Albert Shih.



Figure A.3: *Left:* Antarctic deed-field training where we learned about snow camping. The original recovery plan was to establish a 3 week field camp and groom an ice runway to accommodate a Basler aircraft at the landing site. Due to fuel restrictions we were instead given 3 out-and-back flights on a Twin Otter. Pictured in the background to left is Mt. Erebus, the southernmost active volcano, and Mt. Terror to right. *Right:* Since the landing site was ∼500 miles from the South Pole station, refueling the Twin Otter at fuel caches en route was necessary. Images courtesy of Hazel Bain.



Figure A.4: Payload recovery in January 2017. The 800 lb. spectrometer was supported by fuel canisters and sawed off the frame onto a sled. Sliding across the ice was relatively easy but loading onto the plane via the wooden ramp required 6 people to push and manage safety straps. Hazel Bain is wearing an orange coat and the author is wearing a pink hat.

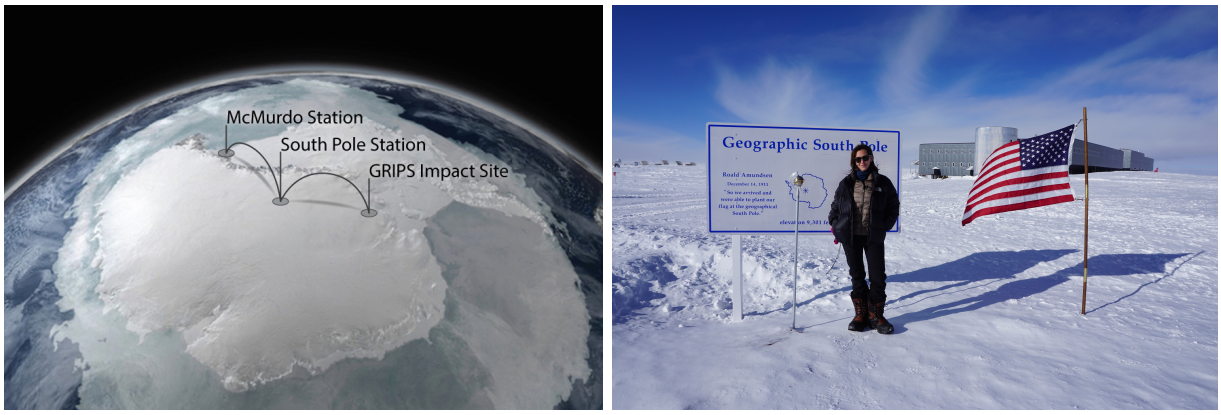

Figure A.5: *Left:* Recovery team route from the GRIPS launch site on the Ross ice shelf outside McMurdo to the South Pole Station and impact site. We waited for weather at the South Pole for 4 weeks before conditions were clear for landing at the fuel cache and impact site. *Right:* The author at the South Pole station in December 2016.



Figure A.6: NASA's LDB facility. Convective cooling from freezing fog during roll-outs resulted in some components being colder on the ground than in flight.



Figure A.7: *Left:* The cryostat was powered during transport to the Plum Brook T-vac test and NASA systems integration. *Right:* Since no pressure vessels were used in flight, thermal straps and conductive heatsinks were used to mitigate heat on commercial electronics.

Bibliography

[1] Wereldburger758 Adundovi. *Pn-junction-equilibrium-graph. [CC BY-SA 3.0] via Wikimedia Commons*. URL: https://commons.wikimedia.org/wiki/File:Pn-junction-equilibrium-graph.svg.

[2] M. Amman and P. N. Luke. "Three-dimensional position sensing and field shaping in orthogonal-strip germanium gamma-ray detectors". In: *Nuclear Instruments and Methods in Physics Research A* 452 (Sept. 2000), pp. 155–166. DOI: 10.1016/S0168-9002(00)00351-X.

[3] S. Amrose et al. "Calibration of 3D positioning in a Ge cross-strip detector". In: *Nuclear Instruments and Methods in Physics Research A* 505 (June 2003), pp. 170–173. DOI: 10.1016/S0168-9002(03)01044-1.

[4] B. Aschenbach. "X-ray telescopes". In: *Reports on Progress in Physics* 48 (May 1985), pp. 579–629. DOI: 10.1088/0034-4885/48/5/001.

[5] B. Aschenbach et al. "Imaging performance of the XMM-Newton x-ray telescopes". In: *X-Ray Optics, Instruments, and Missions III*. Ed. by J. E. Truemper and B. Aschenbach. Vol. 4012. procspie. July 2000, pp. 731–739. DOI: 10.1117/12.391615. eprint: astro-ph/0007256.

[6] M. J. Aschwanden. "Particle acceleration and kinematics in solar flares - A Synthesis of Recent Observations and Theoretical Concepts (Invited Review)". In: *ssr* 101 (Jan. 2002), pp. 1–227. DOI: 10.1023/A:1019712124366.

[7] OpenStax Astronomy. *Electromagnetic Spectrum*. URL: http://cnx.org/contents/LnN76Opl@9.1:Fe7BqV48@4/The-Electromagnetic-Spectrum#OSC_Astro_05_02_Radiation2.

[8] W. I. Axford, E. Leer, and G. Skadron. "The acceleration of cosmic rays by shock waves". In: *International Cosmic Ray Conference* 11 (1977), pp. 132–137.

[9] H. W. Babcock. "The Topology of the Sun's Magnetic Field and the 22-YEAR Cycle." In: *apj* 133 (Mar. 1961), p. 572. DOI: 10.1086/147060.

[10] M. S. Bandstra et al. "Detection and Imaging of the Crab Nebula with the Nuclear Compton Telescope". In: *apj* 738, 8 (Sept. 2011), p. 8. DOI: 10.1088/0004-637X/738/1/8. arXiv: 1106.0323 [astro-ph.IM].

[11] H. H. Barrett, J. D. Eskin, and H. B. Barber. "Charge Transport in Arrays of Semiconductor Gamma-Ray Detectors". In: *Physical Review Letters* 75 (July 1995), pp. 156–159. DOI: 10.1103/PhysRevLett.75.156.

[12] N. M. Barrière et al. "Developing a second generation Laue lens prototype: high-reflectivity crystals and accurate assembly". In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Vol. 8147. procspie. Sept. 2011, p. 81471D. DOI: 10.1117/12.899893. arXiv: 1111.6700 [astro-ph.IM].

[13] S. D. Barthelmy et al. "The Burst Alert Telescope (BAT) on the SWIFT Midex Mission". In: *ssr* 120 (Oct. 2005), pp. 143–164. DOI: 10.1007/s11214-005-5096-3. eprint: astro-ph/0507410.

[14] A. R. Bell. "The acceleration of cosmic rays in shock fronts. I". In: *mnras* 182 (Jan. 1978), pp. 147–156. DOI: 10.1093/mnras/182.2.147.

[15] A. Benz, ed. *Plasma Astrophysics, second edition*. Vol. 279. Astrophysics and Space Science Library. June 2002. DOI: 10.1007/978-0-306-47719-5.

[16] A. O. Benz. "Flare Observations". In: *Living Reviews in Solar Physics* 14, 2 (Dec. 2017), p. 2. DOI: 10.1007/s41116-016-0004-3.

[17] A. O. Benz et al. "Survey on Solar X-ray Flares and Associated Coherent Radio Emissions". In: *solphys* 226 (Jan. 2005), pp. 121–142. DOI: 10.1007/s11207-005-5254-5. eprint: astro-ph/0410436.

[18] J. Birn and M. Hesse. "Geospace Environment Modeling (GEM) magnetic reconnection challenge: Resistive tearing, anisotropic pressure and hall effects". In: *jgr* 106 (Mar. 2001), pp. 3737–3750. DOI: 10.1029/1999JA001001.

[19] R. D. Blandford and J. P. Ostriker. "Particle acceleration by astrophysical shocks". In: *apjl* 221 (Apr. 1978), pp. L29–L32. DOI: 10.1086/182658.

[20] S. E. Boggs, W. Coburn, and E. Kalemci. "Gamma-Ray Polarimetry of Two X-Class Solar Flares". In: *apj* 638 (Feb. 2006), pp. 1129–1139. DOI: 10.1086/498930. eprint: astro-ph/0510588.

[21] S. E. Boggs et al. "A high resolution gamma-ray and hard x-ray spectrometer (HIREGS) for long duration balloon flights". In: *Advances in Space Research* 21 (1998), pp. 1015–1018. DOI: 10.1016/S0273-1177(97)01090-9.

[22] S. Boggs et al. "DUAL Gamma-Ray Mission". In: *ArXiv e-prints* (June 2010). arXiv: 1006.2102 [astro-ph.IM].

[23] J. C. Brown et al. "Evaluation of Algorithms for Reconstructing Electron Spectra from Their Bremsstrahlung Hard X-Ray Spectra". In: *apj* 643 (May 2006), pp. 523–531. DOI: 10.1086/501497.

[24] J. C. Brown et al. "Local re-acceleration and a modified thick target model of solar flare electrons". In: *aap* 508 (Dec. 2009), pp. 993–1000. DOI: 10.1051/0004-6361/200913145. arXiv: 0909.4243 [astro-ph.SR].

[25] A. Busboom, H. Elders-Boll, and H. D. Schotten. "Uniformly Redundant Arrays". In: *Experimental Astronomy* 8 (1998), pp. 97–123.

[26] R. Camattari. "Laue lens for astrophysics: Extensive comparison between mosaic, curved, and quasi-mosaic crystals". In: *aap* 587, A21 (Mar. 2016), A21. DOI: 10.1051/0004-6361/201526745.

[27] H. Carmichael. "A Process for Flares". In: *NASA Special Publication* 50 (1964), p. 451.

[28] P. A. Cassak, J. F. Drake, and M. A. Shay. "A Model for Spontaneous Onset of Fast Magnetic Reconnection". In: *apjl* 644 (June 2006), pp. L145–L148. DOI: 10.1086/505690. eprint: physics/0604001.

[29] P. A. Cassak, J. F. Drake, and M. A. Shay. "Catastrophic onset of fast magnetic reconnection with a guide field". In: *Physics of Plasmas* 14.5, 054502 (May 2007), p. 054502. DOI: 10.1063/1.2734948.

[30] P. Charbonneau. "Dynamo Models of the Solar Cycle". In: *Living Reviews in Solar Physics* 7, 3 (Sept. 2010), p. 3. DOI: 10.12942/lrsp-2010-3.

[31] NASA/Adler/U. Chicago/Wesleyan/JPL-Caltech. *Milky Way and Our Location*.

[32] S. D. Christe. "Solar hard X-ray microbursts". PhD thesis. University of California, Berkeley, 2007.

[33] S Christe et al. "A Solar Aspect System for the HEROES mission." In: *IEEE* (2014), pp. 1–9.

[34] S. Christe et al. "Exploring impulsive solar magnetic energy release and particle acceleration with focused hard X-ray imaging spectroscopy". In: *ArXiv e-prints* (Jan. 2017). arXiv: 1701.00792 [astro-ph.SR].

[35] J. Christensen-Dalsgaard et al. "The Current State of Solar Modeling". In: *Science* 272 (May 1996), pp. 1286–1292. DOI: 10.1126/science.272.5266.1286.

[36] W. Coburn et al. "First results from the balloon flight of the NCT prototype". In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Ed. by O. H. W. Siegmund. Vol. 5898. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Aug. 2005, pp. 13–21. DOI: 10.1117/12.615876.

[37] W. Coburn et al. "Preliminary laboratory performance of the NCT prototype flight electronics". In: *X-Ray and Gamma-Ray Instrumentation for Astronomy XIII*. Ed. by K. A. Flanagan and O. H. W. Siegmund. Vol. 5165. Feb. 2004, pp. 131–138. DOI: 10.1117/12.506037.

[38] A. H. Compton. "A Quantum Theory of the Scattering of X-rays by Light Elements". In: *Physical Review* 21 (May 1923), pp. 483–502. DOI: 10.1103/PhysRev.21.483.

[39] NCAR's High Altitude Observatory (corona) and Luc Viatour (chromosphere). *Sun's Atmosphere - Corona and Chromosphere*. URL: https://scied.ucar.edu/imagecontent/suns-atmosphere-corona-and-chromosphere.

[40] B. R. Dennis, A. G. Emslie, and H. S. Hudson. "Overview of the Volume". In: *ssr* 159 (Sept. 2011), pp. 3–17. DOI: 10.1007/s11214-011-9802-z. arXiv: 1109.5831 [astro-ph.SR].

[41] B. R. Dennis et al. "Very High-Resolution Solar X-Ray Imaging Using Diffractive Optics". In: *solphys* 279 (Aug. 2012), pp. 573–588. DOI: 10.1007/s11207-012-0016-7. arXiv: 1205.4762 [astro-ph.IM].

[42] R. H. Dicke. "Scatter-Hole Cameras for X-Rays and Gamma Rays". In: *apjl* 153 (Aug. 1968), p. L101. DOI: 10.1086/180230.

[43] J. F. Drake et al. "Electron acceleration from contracting magnetic islands during reconnection". In: *nat* 443 (Oct. 2006), pp. 553–556. DOI: 10.1038/nature05116.

[44] N. Duncan et al. "Detector and imaging systems for the gamma-ray imager/polarimeter for solar flares (GRIPS) instrument". In: *Solar Physics and Space Weather Instrumentation V*. Vol. 8862. Proceedings of the SPIE. Sept. 2013, 88620W. DOI: 10.1117/12.2024667.

[45] N. Duncan et al. "First flight of the Gamma-Ray Imager/Polarimeter for Solar flares (GRIPS) instrument". In: *Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray*. Vol. 9905. July 2016, 99052Q. DOI: 10.1117/12.2233859. arXiv: 1609.08558 [astro-ph.IM].

[46] J. P. Eastwood et al. "Average properties of the magnetic reconnection ion diffusion region in the Earth's magnetotail: The 2001-2005 Cluster observations and comparison with simulations". In: *Journal of Geophysical Research (Space Physics)* 115, A08215 (Aug. 2010), A08215. DOI: 10.1029/2009JA014962.

[47] A. G. Emslie, J. A. Miller, and J. C. Brown. "An Explanation for the Different Locations of Electron and Ion Acceleration in Solar Flares". In: *apjl* 602 (Feb. 2004), pp. L69–L72. DOI: 10.1086/382350.

[48] A. G. Emslie et al. "Global Energetics of Thirty-eight Large Solar Eruptive Events". In: *ApJ* 759, 71 (Nov. 2012), p. 71. DOI: 10.1088/0004-637X/759/1/71. arXiv: 1209.2654 [astro-ph.SR].

[49] E. E. Fenimore. "Coded aperture imaging: the modulation transfer function for uniformly redundant arrays". In: *ao* 19 (July 1980), pp. 2465–2471. DOI: 10.1364/AO.19.002465.

[50] E. Fermi. "On the Origin of the Cosmic Radiation". In: *Physical Review* 75 (Apr. 1949), pp. 1169–1174. DOI: 10.1103/PhysRev.75.1169.

[51] L. Fletcher and H. S. Hudson. "Impulsive Phase Flare Energy Transport by Large-Scale Alfvén Waves and the Electron Acceleration Problem". In: *apj* 675, 1645-1655 (Mar. 2008), pp. 1645–1655. DOI: 10.1086/527044. arXiv: 0712.3452.

[52] L. Fletcher et al. "An Observational Overview of Solar Flares". In: *ssr* 159 (Sept. 2011), pp. 19–106. DOI: 10.1007/s11214-010-9701-8. arXiv: 1109.5932.

[53] Peter Foukal. *Solar Astrophysics*. 2004.

[54] F. Frontera and P. von Ballmoos. "Laue Gamma-Ray Lenses for Space Astrophysics: Status and Prospects". In: *X-Ray Optics and Instrumentation, 2010. Special Issue on X-Ray Focusing: Techniques and Applications, id.215375* 2010, 215375 (2010), p. 215375. DOI: 10.1155/2010/215375. arXiv: 1007.4308 [astro-ph.IM].

[55] R. Giacconi et al. "Solar X-Ray Image Obtained Using Grazing-Incidence Optics." In: *apj* 142 (Oct. 1965), pp. 1274–1278. DOI: 10.1086/148404.

[56] Lindsay Glesener. "Faint Coronal Hard X-rays From Accelerated Electrons in Solar Flares". PhD thesis. University of California at Berkeley, 2012.

[57] L. Glesener et al. "The FOXSI solar sounding rocket campaigns". In: *Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray*. Vol. 9905. procspie. July 2016, 99050E. DOI: 10.1117/12.2232262.

[58] J. S. Halekas et al. "In situ observations of reconnection Hall magnetic fields at Mars: Evidence for ion diffusion region encounters". In: *Journal of Geophysical Research (Space Physics)* 114, A11204 (Nov. 2009), A11204. DOI: 10.1029/2009JA014544.

[59] J. H. Hammond. *The camera obscura: a chronicle.* 1981.

[60] F. A. Harrison et al. "Development of the High-Energy Focusing Telescope (HEFT) balloon experiment". In: *X-Ray Optics, Instruments, and Missions III*. Ed. by J. E. Truemper and B. Aschenbach. Vol. 4012. procspie. July 2000, pp. 693–699. DOI: 10.1117/12.391608.

[61] F. A. Harrison et al. "The Nuclear Spectroscopic Telescope Array (NuSTAR) High-energy X-Ray Mission". In: *apj* 770, 103 (June 2013), p. 103. DOI: 10.1088/0004-637X/770/2/103. arXiv: 1301.7307 [astro-ph.IM].

[62] T. Hirayama. "Theoretical Model of Flares and Prominences. I: Evaporating Flare Model". In: *solphys* 34 (Feb. 1974), pp. 323–338. DOI: 10.1007/BF00153671.

[63] G. D. Holman et al. "Implications of X-ray Observations for Electron Acceleration and Propagation in Solar Flares". In: *ssr* 159 (Sept. 2011), pp. 107–166. DOI: 10.1007/s11214-010-9680-9. arXiv: 1109.6496 [astro-ph.SR].

[64] P. Hoyng et al. "Origin and Location of the Hard X-Ray Emission in a Two-Ribbon Flare". In: *apjl* 246 (June 1981), p. L155. DOI: 10.1086/183574.

[65] G. J. Hurford. "X-ray imaging with collimators, masks and grids". In: *ISSI Scientific Reports Series* 9 (2010), pp. 223–234.

[66] G. J. Hurford et al. "First Gamma-Ray Images of a Solar Flare". In: *apjl* 595 (Oct. 2003), pp. L77–L80. DOI: 10.1086/378179.

[67] G. J. Hurford et al. "Gamma-Ray Imaging of the 2003 October/November Solar Flares". In: *apjl* 644 (June 2006), pp. L93–L96. DOI: 10.1086/505329.

[68] G. J. Hurford et al. "The RHESSI Imaging Concept". In: *solphys* 210 (Nov. 2002), pp. 61–86. DOI: 10.1023/A:1022436213688.

[69] S. Ishikawa et al. "On the Relation of Above-the-loop and Footpoint Hard X-Ray Sources in Solar Flares". In: *apj* 737, 48 (Aug. 2011), p. 48. DOI: 10.1088/0004-637X/737/2/48.

[70] NASA JPL/Caltech. *Solar System, in Perspective*. URL: https://www.nasa.gov/mission_pages/voyager/multimedia/pia17046.html.

[71] Kelvinsong. *[CC BY-SA 3.0], via Wikimedia Commons*. URL: https://commons.wikimedia.org/wiki/File:Sun_poster.svg.

[72] C. A. Kierans et al. "The 2016 Super Pressure Balloon flight of the Compton Spectrometer and Imager". In: *ArXiv e-prints* (Jan. 2017). arXiv: 1701.05558.

[73] A. L. Kiplinger et al. "Millisecond time variations in hard X-ray solar flares". In: *apjl* 265 (Feb. 1983), pp. L99–L104. DOI: 10.1086/183966.

[74] Glenn Knoll. *Radiation Detection and Measurement*. Michigan: Wiley, 2010.

[75] E. P. Kontar et al. "Electron-Electron Bremsstrahlung Emission and the Inference of Electron Flux Spectra in Solar Flares". In: *apj* 670 (Nov. 2007), pp. 857–861. DOI: 10.1086/521977. arXiv: 0707.4225.

[76] R. A. Kopp and G. W. Pneuman. "Magnetic reconnection in the corona and the loop prominence phenomenon". In: *solphys* 50 (Oct. 1976), pp. 85–98. DOI: 10.1007/BF00206193.

[77] A. G. Kosovichev et al. "Structure and Rotation of the Solar Interior: Initial Results from the MDI Medium-L Program". In: *solphys* 170 (1997), pp. 43–61. DOI: 10.1023/A:1004949311268.

[78] J. Krizmanic, G. Skinner, and N. Gehrels. "Formation Flying for a Fresnel Lens Observatory Mission". In: *Experimental Astronomy* 20 (Dec. 2005), pp. 497–503. DOI: 10.1007/s10686-006-9038-1. eprint: astro-ph/0601047.

[79] J. Krizmanic et al. "Development of ground-testable phase fresnel lenses in silicon". In: *Experimental Astronomy* 20 (Dec. 2005), pp. 299–306. DOI: 10.1007/s10686-006-9030-9. eprint: physics/0601012.

[80] S. Krucker et al. "Hard X-ray emission from the solar corona". In: *aapr* 16 (Oct. 2008), pp. 155–208. DOI: 10.1007/s00159-008-0014-9.

[81] S. Krucker et al. "The focusing optics x-ray solar imager (FOXSI): instrument and first flight". In: *Solar Physics and Space Weather Instrumentation V*. Vol. 8862. procspie. Sept. 2013, 88620R. DOI: 10.1117/12.2024277.

[82] LASCO/SOHO. *SOHO Science Archives*. URL: https://sohowww.nascom.nasa.gov/data/archive/.

[83] F. Lei, A. J. Dean, and G. L. Hills. "Compton Polarimetry in Gamma-Ray Astronomy". In: *ssr* 82 (Nov. 1997), pp. 309–388. DOI: 10.1023/A:1005027107614.

[84] R. B. Leighton. "A Magneto-Kinematic Model of the Solar Cycle". In: *apj* 156 (Apr. 1969), p. 1. DOI: 10.1086/149943.

[85] R. B. Leighton. "Transport of Magnetic Fields on the Sun." In: *apj* 140 (Nov. 1964), p. 1547. DOI: 10.1086/148058.

[86] R. P. Lin. "Energy Release and Particle Acceleration in Flares: Summary and Future Prospects". In: *ssr* 159 (Sept. 2011), pp. 421–445. DOI: 10.1007/s11214-011-9801-0. arXiv: 1110.1805 [astro-ph.SR].

[87] R. P. Lin et al. "RHESSI Observations of Particle Acceleration and Energy Release in an Intense Solar Gamma-Ray Line Flare". In: *apjl* 595 (Oct. 2003), pp. L69–L76. DOI: 10.1086/378932.

[88] R. P. Lin et al. "The Reuven Ramaty High-Energy Solar Spectroscopic Imager (RHESSI)". In: *solphys* 210 (Nov. 2002), pp. 3–32. DOI: 10.1023/A:1022428818870.

[89] A. Lowell et al. "Positional calibrations of the germanium double sided strip detectors for the Compton spectrometer and imager". In: *High Energy, Optical, and Infrared Detectors for Astronomy VII*. Vol. 9915. procspie. Aug. 2016, 99152H. DOI: 10.1117/12.2233145.

[90] S. Masuda et al. "A loop-top hard X-ray source in a compact solar flare as evidence for magnetic reconnection". In: *nat* 371 (Oct. 1994), pp. 495–497. DOI: 10.1038/371495a0.

[91] J. A. Miller. "Particle Acceleration in Solar Flares". In: *Highly Energetic Physical Processes and Mechanisms for Emission from Astrophysical Plasmas*. Ed. by P. C. H. Martens, S. Tsuruta, and M. A. Weber. Vol. 195. IAU Symposium. May 2000, p. 277.

[92] J. A. Miller et al. "Critical issues for understanding particle acceleration in impulsive solar flares". In: *jgr* 102 (July 1997), pp. 14631–14660. DOI: 10.1029/97JA00976.

[93] R. J. Murphy. "Solar Gamma-Ray Spectroscopy". In: *ssr* 130 (June 2007), pp. 127–138. DOI: 10.1007/s11214-007-9197-z.

[94] R. J. Murphy and G. H. Share. "What gamma-ray deexcitation lines reveal about solar-flares". In: *Advances in Space Research* 35 (2005), pp. 1825–1832. DOI: 10.1016/j.asr.2005.03.004.

[95] W. M. Neupert. "Comparison of Solar X-Ray Line Emission with Microwave Emission during Flares". In: *apjl* 153 (July 1968), p. L59. DOI: 10.1086/180220.

[96] M. Oda. "High-Resolution X-Ray Collimator with Broad Field of View for Astronomical Use". In: *ao* 4 (Jan. 1965), pp. 143–143. DOI: 10.1364/AO.4.000143.

[97] Y. Ogasaka et al. "Development of supermirror hard x-ray telescope and the results of first observation flight of InFOCuS flight observation". In: *X-Ray and Gamma-Ray Telescopes and Instruments for Astronomy.* Ed. by J. E. Truemper and H. D. Tananbaum. Vol. 4851. procspie. Mar. 2003, pp. 619–630. DOI: 10.1117/12.461531.

[98] M. Oieroset et al. "In situ detection of collisionless reconnection in the Earth's magnetotail". In: *nat* 412 (July 2001), pp. 414–417. DOI: 10.1038/35086520.

[99] G. Ottaviani, C. Canali, and A. A. Quaranta. "Charge Carrier Transport Properties of Semiconductor Materials Suitable for Nuclear Radiation Detectors". In: *IEEE Transactions on Nuclear Science* 22 (Feb. 1975), pp. 192–204. DOI: 10.1109/TNS.1975.4327640.

[100] E. N. Parker. "Sweet's Mechanism for Merging Magnetic Fields in Conducting Fluids". In: *jgr* 62 (Dec. 1957), pp. 509–520. DOI: 10.1029/JZ062i004p00509.

[101] E. N. Parker. "The Solar-Flare Phenomenon and the Theory of Reconnection and Annihiliation of Magnetic Fields." In: *apjs* 8 (July 1963), p. 177. DOI: 10.1086/190087.

[102] H. E. Petschek. "Magnetic Field Annihilation". In: *NASA Special Publication* 50 (1964), p. 425.

[103] J. C. Philippot. "Automatic Processing of Diode Spectrometry Results". In: *IEEE Transactions on Nuclear Science* 17 (1970), pp. 446–488. DOI: 10.1109/TNS.1970.4325723.

[104] T. A. Prince et al. "Gamma-ray and hard X-ray imaging of solar flares". In: *solphys* 118 (Mar. 1988), pp. 269–290. DOI: 10.1007/BF00148596.

[105] M. P. Pulupa, S. D. Bale, and J. C. Kasper. "Langmuir waves upstream of interplanetary shocks: Dependence on shock and plasma parameters". In: *Journal of Geophysical Research (Space Physics)* 115, A04106 (Apr. 2010), A04106. DOI: 10.1029/2009JA014680.

[106] R. Ramaty, B. Kozlovsky, and R. E. Lingenfelter. "Nuclear gamma-rays from energetic particle interactions". In: *apjs* 40 (July 1979), pp. 487–526. DOI: 10.1086/190596.

[107] R. Ramaty and N. Mandzhavidze. "Gamma-rays from Solar Flares". In: *Highly Energetic Physical Processes and Mechanisms for Emission from Astrophysical Plasmas*. Ed. by P. C. H. Martens, S. Tsuruta, and M. A. Weber. Vol. 195. IAU Symposium. May 2000, p. 123.

[108] R. Ramaty and R. J. Murphy. "Nuclear processes and accelerated particles in solar flares". In: *ssr* 45 (1987), pp. 213–268. DOI: 10.1007/BF00171995.

[109] B. D. Ramsey et al. "First Images from HERO, a Hard X-Ray Focusing Telescope". In: *apj* 568 (Mar. 2002), pp. 432–435. DOI: 10.1086/338801.

[110] J. C. Raymond et al. "Observational Aspects of Particle Acceleration in Large Solar Flares". In: *ssr* 173 (Nov. 2012), pp. 197–221. DOI: 10.1007/s11214-012-9897-x.

[111] R. Ribberfors. "Relationship of the relativistic Compton cross section to the momentum distribution of bound electron states". In: *prb* 12 (Sept. 1975), pp. 2067–2074. DOI: 10.1103/PhysRevB.12.2067.

[112] P. Saint-Hilaire, S. Krucker, and R. P. Lin. "A Statistical Survey of Hard X-ray Spectral Characteristics of Solar Flares with Two Footpoints". In: *solphys* 250 (July 2008), pp. 53–73. DOI: 10.1007/s11207-008-9193-9. arXiv: 1111.4247 [astro-ph.SR].

[113] H. W. Schnopper, R. I. Thompson, and S. Watt. "Predicted Performance of a Rotating Modulation Collimator for Locating Celestial X-Ray Sources". In: *ssr* 8 (Sept. 1968), pp. 534–542. DOI: 10.1007/BF00175005.

[114] M. Schüssler. "The Sun and Its Restless Magnetic Field". In: *Lecture Notes in Physics, Berlin Springer Verlag*. Ed. by K. Scherer, H. Fichter, and B. Herber. Vol. 656. Lecture Notes in Physics, Berlin Springer Verlag. 2004, p. 23. DOI: 10.1007/b100037.

[115] SDO. *Like Painting on the Sun's Layers*.

[116] SDO. *SDO image gallery*.

[117] G. H. Share and R. J. Murphy. "Gamma Radiation From Flare-Accelerated Particles Impacting the Sun". In: *Washington DC American Geophysical Union Geophysical Monograph Series* 165 (Oct. 2006), p. 177. DOI: 10.1029/165GM17.

[118] G. H. Share et al. "Directionality of Flare-accelerated α-Particles at the Sun". In: *apjl* 595 (Oct. 2003), pp. L89–L92. DOI: 10.1086/378176.

[119] G. H. Share et al. "RHESSI e^+-e^- Annihilation Radiation Observations: Implications for Conditions in the Flaring Solar Chromosphere". In: *apjl* 615 (Nov. 2004), pp. L169–L172. DOI: 10.1086/426478.

[120] K. Shibata and T. Magara. "Solar Flares: Magnetohydrodynamic Processes". In: *Living Reviews in Solar Physics* 8, 6 (Dec. 2011), p. 6. DOI: 10.12942/lrsp-2011-6.

[121] A. Y. Shih, R. P. Lin, and D. M. Smith. "RHESSI Observations of the Proportional Acceleration of Relativistic 0.3 MeV Electrons and 30 MeV Protons in Solar Flares". In: *apjl* 698 (June 2009), pp. L152–L157. DOI: 10.1088/0004-637X/698/2/L152.

[122] A. Y. Shih et al. "The Gamma-Ray Imager/Polarimeter for Solar flares (GRIPS)". In: *Space Telescopes and Instrumentation 2012: Ultraviolet to Gamma Ray*. Vol. 8443. Proceedings of the SPIE. Sept. 2012, 84434H. DOI: 10.1117/12.926450.

[123] Albert Shih. *RHESSI Browser*.

[124] T. V. Siversky and V. V. Zharkova. "Stationary and impulsive injection of electron beams in converging magnetic field". In: *aap* 504 (Sept. 2009), pp. 1057–1070. DOI: 10.1051/0004-6361/200912341. arXiv: 0907.1911 [astro-ph.SR].

[125] G. K. Skinner. "Diffractive/refractive optics for high energy astronomy. I. Gamma-ray phase Fresnel lenses". In: *aap* 375 (Aug. 2001), pp. 691–700. DOI: 10.1051/0004-6361:20010745.

[126] G. K. Skinner et al. "The milli-arc-second structure imager (MASSIM): a new concept for a high angular resolution x-ray telescope". In: *Space Telescopes and Instrumentation 2008: Ultraviolet to Gamma Ray*. Vol. 7011. procspie. July 2008, 70110T. DOI: 10.1117/12.789568.

[127] D. M. Smith et al. "High-Resolution Spectroscopy of Gamma-Ray Lines from the X-Class Solar Flare of 2002 July 23". In: *apjl* 595 (Oct. 2003), pp. L81–L84. DOI: 10.1086/378173. eprint: astro-ph/0306292.

[128] D. M. Smith et al. "The RHESSI Spectrometer". In: *solphys* 210 (Nov. 2002), pp. 33–60. DOI: 10.1023/A:1022400716414.

[129] H. Spieler. *Semiconductor Detector Systems*. 2005.

[130] National Institute of Standards and Technology. *Mass Attenuation Coefficients*.

[131] P. A. Sturrock. "Model of the High-Energy Phase of Solar Flares". In: *nat* 211 (Aug. 1966), pp. 695–697. DOI: 10.1038/211695a0.

[132] E. Suarez-Garcia et al. "X-Ray Polarization of Solar Flares Measured with Rhessi". In: *solphys* 239 (Dec. 2006), pp. 149–172. DOI: 10.1007/s11207-006-0268-1. eprint: astro-ph/0609778.

[133] P. A. Sweet. "The Neutral Point Theory of Solar Flares". In: *Electromagnetic Phenomena in Cosmical Physics*. Ed. by B. Lehnert. Vol. 6. IAU Symposium. 1958, p. 123.

[134] et.al Takahashi T. "The ASTRO-H X-ray Observatory". In: *Space Telescopes and Instrumentation 2012: Ultraviolet to Gamma Ray*. Vol. 8443. procspie. Sept. 2012, 84431Z. DOI: 10.1117/12.926190. arXiv: 1210.4378 [astro-ph.IM].

[135] E. Tandberg-Hanssen and A. G. Emslie. *The physics of solar flares*. 1988.

[136] TRACE. *TRACE Image Gallery*. URL: http://www.lmsal.com/TRACE/POD/TRACEpod.html.

[137] P. Ubertini et al. "IBIS: The Imager on-board INTEGRAL". In: *aap* 411 (Nov. 2003), pp. L131–L139. DOI: 10.1051/0004-6361:20031224.

[138] A. M. Veronig and J. C. Brown. "A Coronal Thick-Target Interpretation of Two Hard X-Ray Loop Events". In: *apjl* 603 (Mar. 2004), pp. L117–L120. DOI: 10.1086/383199.

[139] W. T. Vestrand and D. J. Forrest. "Evidence for a spatially extended component of gamma rays from solar flares". In: *apjl* 409 (June 1993), pp. L69–L72. DOI: 10.1086/186862.

[140] W. T. Vestrand et al. "The Solar Maximum Mission Atlas of Gamma-Ray Flares". In: *apjs* 120 (Feb. 1999), pp. 409–467. DOI: 10.1086/313180.

[141] K. Vetter. "Recent Developments in the Fabrication and Operation of Germanium Detectors". In: *Annual Review of Nuclear and Particle Science* 57 (Nov. 2007), pp. 363–404. DOI: 10.1146/annurev.nucl.56.080805.140525.

[142] N. Vilmer, A. L. MacKinnon, and G. J. Hurford. "Properties of Energetic Ions in the Solar Atmosphere from γ-Ray and Neutron Observations". In: *ssr* 159 (Sept. 2011), pp. 167–224. DOI: 10.1007/s11214-010-9728-x. arXiv: 1110.2432 [astro-ph.SR].

[143] E. Virgilli et al. "The LAUE project and its main results". In: *Optics for EUV, X-Ray, and Gamma-Ray Astronomy VI*. Vol. 8861. procspie. Sept. 2013, p. 886107. DOI: 10.1117/12.2023593.

[144] P. von Ballmoos, R. Diehl, and V. Schoenfelder. "Imaging the gamma-sky with Compton telescopes". In: *aap* 221 (Sept. 1989), pp. 396–406.

[145] P. von Ballmoos et al. "CLAIRE: First light for a gamma-ray lens". In: *Experimental Astronomy* 20 (Dec. 2005), pp. 253–267. DOI: 10.1007/s10686-006-9071-0.

[146] C. Winkler et al. "The INTEGRAL mission". In: *aap* 411 (Nov. 2003), pp. L1–L6. DOI: 10.1051/0004-6361:20031288.

[147] G. L. Withbroe and R. W. Noyes. "Mass and energy flow in the solar chromosphere and corona". In: *araa* 15 (1977), pp. 363–387. DOI: 10.1146/annurev.aa.15.090177.002051.

[148] M. Yamada, R. Kulsrud, and H. Ji. "Magnetic reconnection". In: *Reviews of Modern Physics* 82 (Jan. 2010), pp. 603–664. DOI: 10.1103/RevModPhys.82.603.

[149] A. Zehnder et al. "RHESSI imager and aspect systems". In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Ed. by S. L. Keil and S. V. Avakyan. Vol. 4853. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Feb. 2003, pp. 41–59. DOI: 10.1117/12.460290.

[150] V. V. Zharkova et al. "Recent Advances in Understanding Particle Acceleration Processes in Solar Flares". In: *ssr* 159 (Sept. 2011), pp. 357–420. DOI: 10.1007/s11214-011-9803-y. arXiv: 1110.2359 [astro-ph.SR].

[151] Andreas Zoglauer. "First Light for the Next Generation of Compton and Pair Telescopes". PhD thesis. Max-Planck-Institut für Extraterrestrische Physik, 2006.

[152] E. G. Zweibel and M. Yamada. "Magnetic Reconnection in Astrophysical and Laboratory Plasmas". In: *araa* 47 (Sept. 2009), pp. 291–332. DOI: 10.1146/annurev-astro-082708-101726.