A Dark Matter Search Using the Final SuperCDMS Soudan Dataset and the Development of a Large-Format, Highly-Multiplexed, Athermal-Phonon-Mediated Particle Detector

Thesis by
Brett Durcan Cornell

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2018
Defended May 30th 2018

To my family

And to Huilin again besides

ACKNOWLEDGEMENTS

It is impossible for me to adequately thank everyone who helped make this dissertation possible. An experimental effort as large and complex as SuperCDMS is the culmination of decades of work by nearly a hundred individuals. I will try instead to acknowledge some small portion of them.

First I want to thank my advisor Sunil Golwala, I could not have asked for a more patient or knowledgeable mentor to help steer me through graduate school. When needed, you let your students direct themselves, struggle and grow. And at the same time you are always willing to take the time to meet with each of us, potentially for hours on end, any time we need it. I also want to thank the rest of our group here at Caltech. Many of the practical problems that I encountered in the lab as a fledgling experimentalist would have been for worse if not far Matt Hollister. The cryogenic systems here (and graduate students that work them) are the worse for your absence. Dave Moore, thanks for handing me such a high-quality hardware project. Your mentorship, notes, dissertation and code were all vital to the subsequent DM-MKID development here at Caltech. Yen-Young Chang, and Taylor Aralis, thank you both for making our group a more fun and lively place to work. Although we have probably overlapped a little too long at this point, your help with coding, fabrication, testing, and cryogenics has made my last few years much more enjoyable than they otherwise would have been. Finally I want to thank Bruce Bumble. You took a great deal of time and care to teach me micro-fabrication.

Many people from the greater CDMS collaboration also personally helped me in substantial ways. Bruno Serfass, you are indispensable. I have no idea how you took the time to help me navigate the Byzantine data production and analysis system that we as a collaboration have accumulated, but you always did. Dan Bauer, thank you for your depth of knowledge, and guidance. Without it serving on-site shifts would have been far less enjoyable. Steven Yellin, thank you for the statistics, the code and the SUF cluster. Ray Bunker and Lauren Hsu, without your early leadership the HT analysis would not have happened. I also want to thank Rob Calkins, Ben Loer, Noah Kurinsky, Peter Redl, Hang Qiu, and Elias Lopez Asamar for your contributions to this analysis. Most importantly, I want to thank Brad Welliver and Todd Doughty. You two did most of the hard play-making in this analysis, leaving me with the fun bit at the end.

I want to thank all the great friends I have made here in Los Angeles, without your distraction and support the last eight years would not have been possible.

Last I want to thank my family, Huilin and even Miso. You are the best.

ABSTRACT

Over the past eighty years, numerous complementary observations of our universe have indicated that our current description of physics is far from complete. The "ordinary matter", such as electrons, protons, photons and neutrons, that constitutes the bulk of all human physical experiences is actually only a minority (about 16%) of the total mass of the universe. The remaining 84% is very poorly understood, but has profound effects on the dynamics and evolution of our universe. Because it does not interact with light, and is not observable in telescopes on earth, this extra mass is usually referred to as "dark matter". Although the dark matter is poorly understood, Weakly Interacting Massive Particles (WIMPs) are a well-motivated candidate that can be directly detected via a non-gravitational interaction with normal matter, potentially allowing for direct terrestrial detection and characterization of this dark matter. This dissertation is focused on this direct WIMP detection and will be broken into two main parts.

The first part focuses on the blinded analysis of roughly three years of data collected from March 2012 to November 2015 by the SuperCDMS Soudan experiment. SuperCDMS Soudan consists of an array of 15, 0.6-kg, cryogenic, Ge iZIP particle detectors situated in a decommissioned iron mine in remote northern Minnesota. This analysis is optimized to be sensitive to theoretical WIMP masses above 10 GeV/c^2. This result set the strongest limits for WIMP–germanium-nucleus interactions for WIMP masses greater than 12 Gev/c^2 and can be found here [1].

The second part focuses on the development new kind of particle detector in the style of a SuperCDMS iZIP, designed to simplify fabrication and readout, improve phonon-based position reconstruction, and help to scale to larger target arrays. These detectors replace the TES-based phonon sensors of the iZIP with Microwave Kinetic Inductance Detectors (MKIDs).

PUBLISHED CONTENT AND CONTRIBUTIONS

R. Agnese et al. “Results from the Super Cryogenic Dark Matter Search Experiment at Soudan”. In: *Phys. Rev. Lett.* 120 (6 Feb. 2018), p. 061802. DOI: `10.1103/PhysRevLett.120.061802`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.120.061802`

I was the involved in all parts of this analysis and paper writing effort. I was the analysis team leader, chair of the paper writing committee, and the corresponding author for the paper.

B. Cornell et al. “Particle Detection Using MKID-Based Athermal-Phonon Mediated Detectors”. In: *Journal of Low Temperature Physics* 176.5 (Sept. 2014), pp. 891–897. ISSN: 1573-7357. DOI: `10.1007/s10909-013-1039-7`. URL: `https://doi.org/10.1007/s10909-013-1039-7`

I was responsible for all aspects of new detector development described in this paper as well as the primary writer.

B. Cornell et al. *High-resolution gamma-ray detection using phonon-mediated detectors*. 2012. DOI: `10.1117/12.926829`. URL: `https://doi.org/10.1117/12.926829`

I was responsible for all aspects of new detector development described in this paper as well as the primary writer.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION: DARK MATTER

Over the past eighty years, numerous complementary observations of our universe have indicated that our current description of physics is far from complete. The "ordinary matter", such as electrons, protons, photons and neutrons, that constitutes the bulk of all human physical experiences is actually only a minority ($\approx 5\%$) of the total mass-energy of the universe. The remaining 95% is very poorly understood, but is currently thought to be made of two main components as is shown in figure 1.1. The first acts as constant, non-zero, vacuum energy and is called "dark energy". Although dark energy constitutes most of the energy density (at $\approx 68\%$), and is an important current topic of research [2] this dissertation will instead focus on the direct characterization of the remaining $\approx 27\%$ called "dark matter". This substance is "dark" in the sense that it does not appear to interact electromagnetically, and because it neither emits nor reflects light, it is invisible to our telescopes here on earth. Although not directly visible, it has mass, or more correctly, it produces gravity that is in excess of what would be produced by the visible matter in our universe. This has profound and observable effects on the dynamics of many astrophysical and cosmological phenomena. Although they do interact gravitationally, if ordinary matter and dark matter interact via another force it would have to be very weakly. Over the better part of the last three decades, numerous experiments

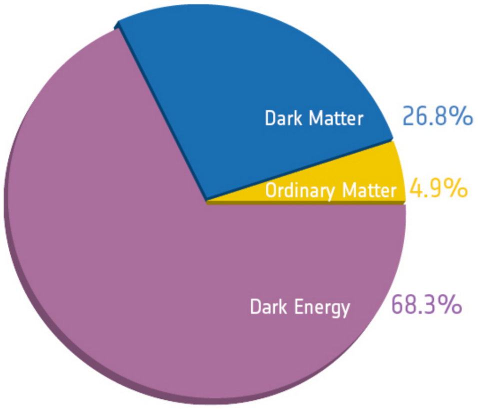


Figure 1.1: Composition of the universe as it is understood at this time. This result is from the first 15.5 months of the Planck mission [3]. Figure adapted from [4]

have been deployed to directly search for such an interaction, and as of this writing, no definitive signal has been observed. This dissertation will be focused on such a search.

1.1 Empirical Evidence

The body of evidence that lead to the development of the cosmological model described in the previous section is quite extensive, and a complete review of it would be well outside the scope of this dissertation. Many of these observed phenomena my be individually described by a verity of different theoretical mechanisms, only some of which include dark matter as a necessary component. When taken together, however, the dark matter and dark energy dominated cosmology of figure 1.1 is the only one consistent with all observation. I will take the remainder of this section to briefly outline some of this evidence, but for a more complete background please see [5].

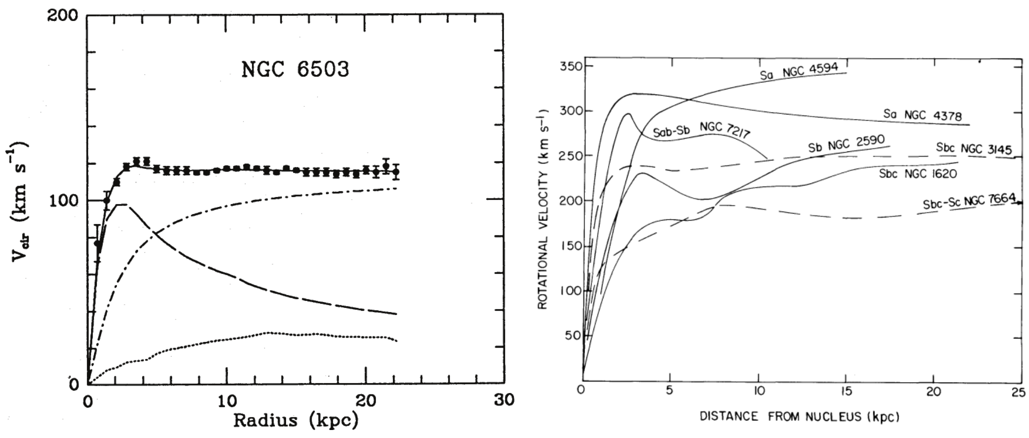
1.1.1 Astrophysical Evidence

Galaxies

If we want to gain insight into the mass distribution within a particular galaxy, there are two general ways to go about it. The distribution of luminous matter can be directly observed, or the dynamics of the luminous elements can be measured, and the mass inferred indirectly. Take spiral galaxies as an example. The velocity of a particular satellite will depend (assuming azimuthal symmetry) on the radius of its orbit and the total enclosed mass as

$$v(r) \propto \sqrt{\frac{m_{\text{enclosed}}(r)}{r}} \tag{1.1}$$

This $v(r)$ can be measured using the Doppler shift of various spectral lines of orbiting gases, particularly the 21 cm spin-flip transition of neutral hydrogen (HI). This hydrogen gas is present throughout the luminous disk, but also extends to orbits of very high radii (many times the size of the luminous disk). If the mass of the spiral galaxy were primarily within the luminous disk, we would expect a characteristic $\frac{1}{\sqrt{r}}$ falloff in the velocity of this HI gas once we were several kpc outside the disk. These velocity curves have been measured for many spiral galaxies (see



(a) (b)

Figure 1.2: Rotational velocity vs radius for a number of different galaxies. (a) The rotational velocity curve of the spiral galaxy NGC 6503 (black points with error bars), which is constant at high radius. A three component fit to this curve (solid) is shown and it includes the luminous matter of the disk (dashed), the gas which extends past the disk (dotted) and the remaining non-luminous dark matter halo (dot-dashed) [6]. (b) Similar rotation curves for 7 nearby galaxies depicting this flat radial dependence of rotational velocity at high radii [7].

figure 1.2) and they are typically constant out to many galactic radii, implying that $m_{\text{enclosed}}(r) \propto r$. This "halo" of non-luminous mass that surrounds the galaxy is though to be composed of dark matter, and it is much larger, both in mass and extent, than the visible disk. Rubin:1978 Although they lack the coherent orbits of a spiral galaxy, a similar approach can be used to compare the luminosity to the inferred mass for elliptical and dwarf galaxies [8]. These structures are believed to be gravitationally virialized allowing the total gravitational potential (and galactic mass) to be inferred from the velocity dispersion as in

$$M_{\text{tot}} \simeq 2\frac{R_{\text{tot}}\langle v^2\rangle}{G} \tag{1.2}$$

where R_{tot} is the total extent of the galaxy. Although nearby elliptical galaxies have been found to contain significant amounts of non-luminous matter [9], dwarf satellite galaxies have been found to have some of the highest mass-to-light ratios[1] of any observed objects [8].

---

[1] Approaching 1000 time the solar mass to light ratio.

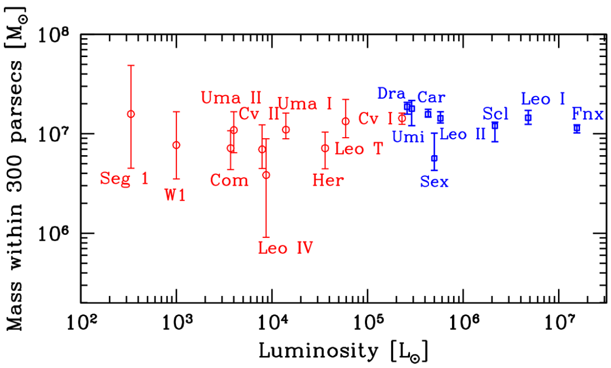


Figure 1.3: Inferred integrated mass measured via the velocity dispersion of a number of dwarf galaxies orbiting the Milky Way. Both those discovered by SDSS (red) as well as "classical" pre-SDSS dwarf satellites (blue). As can be seen, many of these recently discovered objects have significant amounts of non-luminous matter [10].

Galaxy Clusters

This inferred mass via velocity dispersion trick can be used to investigate objects much larger than individual galaxies. Galaxy clusters are groupings of hundreds to thousands of galaxies and are the largest known gravitationally bound[2] structures in the universe. The first observational evidence of dark matter was made in 1933 here at Caltech by Zwicky [11], by observing the velocity dispersion relation of galaxies in the Coma cluster. Equation 1.2 can then be used to infer the total mass of the galaxy cluster in a manner very similar to that in elliptical and dwarf galaxies. He inferred the total mass of the cluster it to be ~400 times as greater[3] than would be expected based on the luminous matter alone [12], leading him to postulate a non-luminous *dunkle materie* (dark matter) component.

The bulk of the baryonic mass of these large structures actually comes not from the galaxies themselves, but from from very hot, dust and gas called the intracluster medium (ICM) that exists between them. When fully virialized, this ICM reaches temperatures of nearly 10^8 K. This is hot enough to thermally emit x-rays [13], rendering galaxy clusters the brightest objects in the x-ray sky. Provided a cluster has not recently interacted, and the ICM is in hydrostatic equilibrium, the mass of the cluster can be inferred from the measured x-ray emission. This has been done for a large number of clusters and agrees well with the velocity dispersion

---

[2] And subsequently virialized.

[3] This large discrepancy was due in part to his assuming the Hubble parameter to be 558 $\frac{\text{km}}{\text{s Mpc}}$. A more modern accounting would put his over density at ~50.

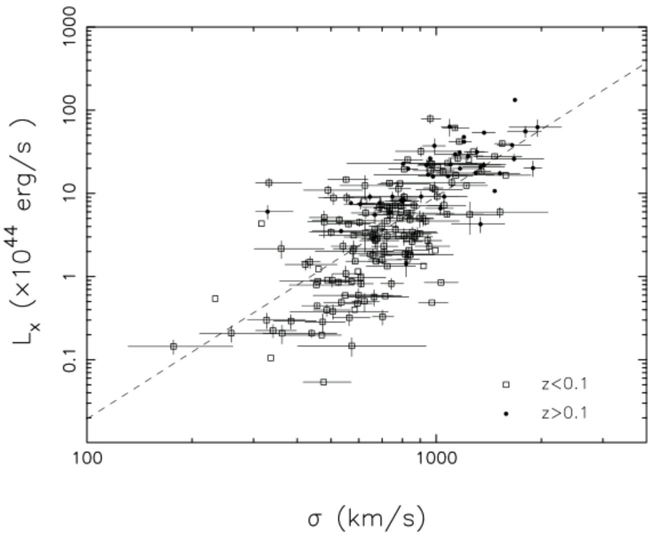


Figure 1.4: Measured x-ray luminosity vs velocity dispersion for 169 galaxy clusters. Comparing the best fit line (dashed) to theoretical models suggest that the mean baryonic mass fraction is about 17% [13]

measurements.

In addition to emitting x-rays, the ICM can also interact with and distort cosmic microwave background (CMB) light as it passes through the cluster. As these CMB photos travel through the ICM they can inverse Compton scatter off the hot gas of electrons, and when observed from earth this produces a characteristic distortion in the CMB spectrum known as the Sunyaev–Zel'dovich (or SZ) effect [14]. This effect depends on the total integrated pressure of the cluster and, assuming this can be related to the total thermal energy of the cluster, can provide an estimate of the total gravitational energy and mass of a cluster.

Finally, the mass of galaxy clusters can be probed using the general relativistic phenomena of gravitational lensing. This effect is due to geodesics being bent by large gravitational fields and is a wonderful empirical test of general relativity in its own right. Lensing is a particularly valuable probe when it comes to measuring cluster mass, as it only depends on gravity, and is free from the the kinematic assumptions[4] of the previously described methods. Observations of this lensing are usually broken into to rough categories based on the clarity of the effect when seen on earth. In cases of so-called "strong lensing" the lens strength, and the earth-lens-background-object alignment are such that the deflection is not in the linear

---

[4] Such as the hydrostatic equilibrium of the ICM and the full virialization of cluster objects.

regime, where it can be treated as a linear shape distortion and magnification effect. In this regime, multiple images of the background object can sometimes be seen. By studying examples of strong lenses[5] in detail, a mass distribution of the lens can be reconstructed from a careful accounting of all distorted or multiply-imaged background objects. These strong lenses are very rare, and each mass reconstruction must be done individually. As a result, inferring properties about the general non-luminous matter content of all galaxy clusters from them is not well motivated. "Weak lenses" are far more numerous as they only produce slight distortions in imaged background objects[6] and do not rely on the same kind of careful alignment that strong lenses do. Although each of these individual distortions are small, statistical correlations on large numbers of distorted objects can be used to infer the mass [17].

As previously mentioned, lensing can be used to measure a clusters's mass distribution even when that cluster is out of equilibrium, such as during a binary cluster merger event. These mergers are particularly interesting as a clear separation between the baryon distribution and mass distribution of the colliding clusters can sometimes be measured. This is understood to be caused by the (essentially) non-interacting nature of the dark matter. During such a collision, the baryonic ICM from each cluster will rapidly interact and begin to thermalize, effectively colliding inelastically. The majority of the mass of each cluster, however, is dark matter, which will pass through and continue on its course unimpeded. Two striking examples of such mergers are shown in figure 1.5. This phenomena in particular is very hard to explain without evoking some form of non-luminous non-baryonic, weakly interacting material.

1.1.2 Cosmological Evidence

Large Scale Structure Formation

In recent years, the Sloan Digital Sky Survey (SDSS) [18] and the Two-degree-Field Galaxy Redshift Survey (2dFGRS) [19] have made very detailed maps of large portions of the sky[7], cataloging hundreds of millions of individual objects photometrically as well as taking spectral measurements of ~1% of that total. Directly fitting

---

[5] Such as Abell [15] and 2218 [16].

[6] Examples of these distortions would be small changes in the apparent ellipticity or brightness of background galaxies.

[7] covering approximately 35% and 4% of the total sky respectively.

(a) IE0657-558 (b) MACS J0025.4-1222

Figure 1.5: Composite images of two binary galaxy cluster merger events. In both, an optical image taken by the Hubble Space Telescope is overlaid with the inferred mass distribution from gravitational lensing (blue) as well as the x-ray emission measured by the Chandra X-ray Observatory (red).

the measured amplitude and shape of the power spectrum of these objects provides compelling evidence of non-baryonic dark matter, but equally important is the physical distribution of these objects. Upon examination, this data indicates that even on length scales much larger than galaxy clusters[8] these astronomical bodies are not smoothly distributed, but rather form into "large scale structure" of filaments and sheets [20]. This is caused by quantum fluctuations in the very early universe that act as seeds for later gravitational accretion. Simulations (such as the one seen in figure 1.6) indicate that these would have to have been formed before recombination, at a time when baryonic matter would not have been able to collect in such a manner (due to the restoring force of the photon fluid). To achieve the structure seen today a significant amount of non-luminous cold (i.e. non-relativistic) matter is required to start collapsing at matter-radiation equality ($Z \sim 3000$) providing the necessary seed for the baryons at recombination ($Z \sim 1100$).

Big Bang Nucleosynthesis

Big Bang Nucleosynthesis (BBN) describes the processes that occurred in the first few minutes after the big bang, when free neutrons and protons combined to form a number of light stable isotopes (specifically ^{2}H, ^{3}He, ^{4}He, and ^{7}Li.). The probability of any of these nuclear reactions occurring scales as the square of the baryon density. As a result, cosmologies with a low baryon-to-photon ratio would be ex-

---

[8] i.e. not gravitationally bound.

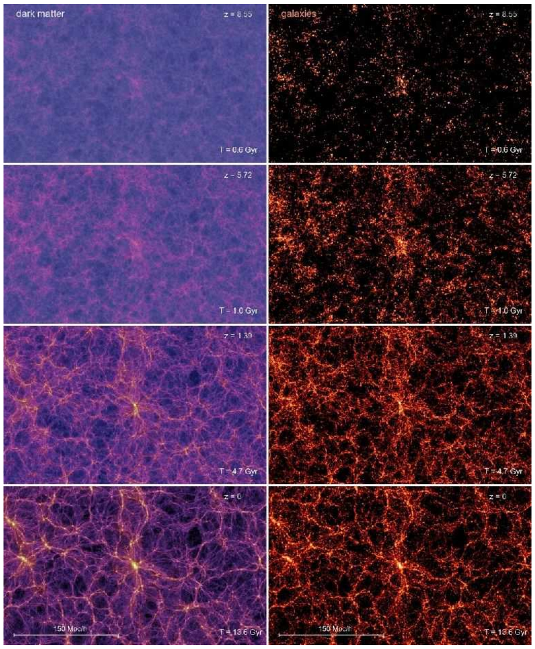


Figure 1.6: Plots depict the Millennium N-body simulation of large scale structure formation from $Z = 0$ (bottom) to $Z = 8.5$ (top). The panels on the right show the predicted positions of galaxies, while the left depicts the simulated dark matter distribution [20].

pected to have relatively more intermediary species (i.e. ^{2}H, ^{3}He) due to incomplete nucleosynthesis. This is shown in figure 1.7. Direct observational measurements of the relic abundance of these isotopes is not straightforward[9], but the measurement of the deuterium abundance alone is enough to indicate that the vast majority of mass in the universe must be non-baryonic [21]

Cosmic Microwave Background

Finally, the best measurement of cold dark matter (and the basis for figure 1.1) comes from detailed observation of the cosmic microwave background (CMB). Immediately after the Big Bang, the universe was opaque. Temperatures were above the atomic binding energy for hydrogen, and the the free electrons and protons strongly coupled to any photons via Thomson scattering, preventing their free propagation. After ~379,000 years things had cooled enough that neutral hydrogen could form, a process cosmologists refer to as "recombination"[10]. At this time pho-

---

[9] ^{3}He and ^{4}He are also produced by stellar nucleosynthesis, which complicates any measurement of their relic abundance.

[10] "Combination" would arguably make more sense, as by all evidence this was the first time in history that electrons and protons have ever combined.

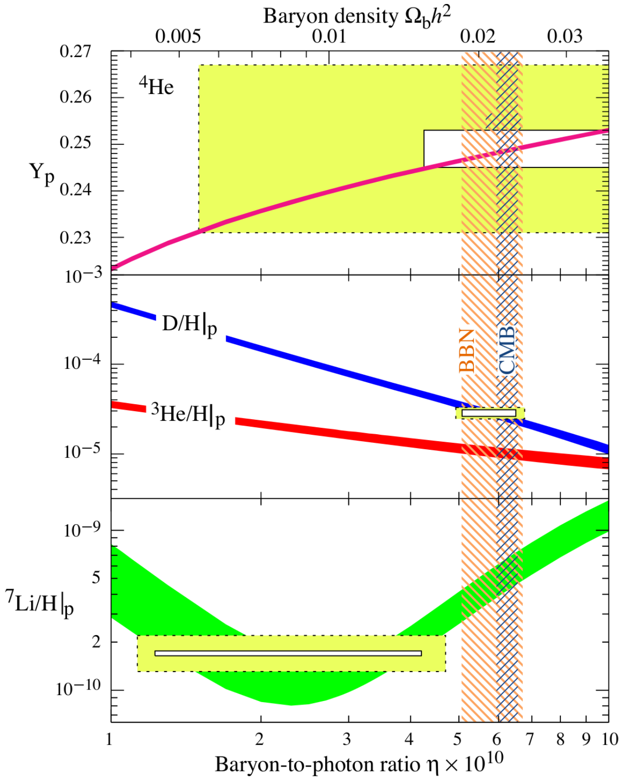


Figure 1.7: Predicted abundances of various light isotopes vs the baryon-to-photon ratio from BBN. The labeled horizontal colored bands are the theoretical predictions while the three rectangular boxes denote the current best observational measurements with statistical uncertainty (white box) as well as total uncertainty (yellow box) shown. The best fit constraints from BBN (orange hash) and CMB (blue hash) are very consistent, and are both in strong agreement with the observed deuterium abundance [21].

tons decouple from matter and are allowed to freely propagate through space. On earth, these recombination-epoch photons can be seen as a highly uniform 2.7 K blackbody spectrum. Although very close to being uniform in temperature, there are residual anisotropies in the CMB are at the level of 1 part in 10^5, and it is these anisotropies that later went on to form the LSS that we see in the modern universe. Detailed measurements of the nature of the anisotropies provides strong limits on the properties and composition of the universe's matter composition. From 2009 to 2013, the ESA's Planck mission made very high-detail maps of these CMB anisotropies, and since its decommissioning, is responsible for the highest precision measurements of several important cosmological parameters including the average relic density of ordinary and dark matter[22]. Prior to recombination, the photon-baryon fluid underwent a number a acoustic oscillations, again seeded by early

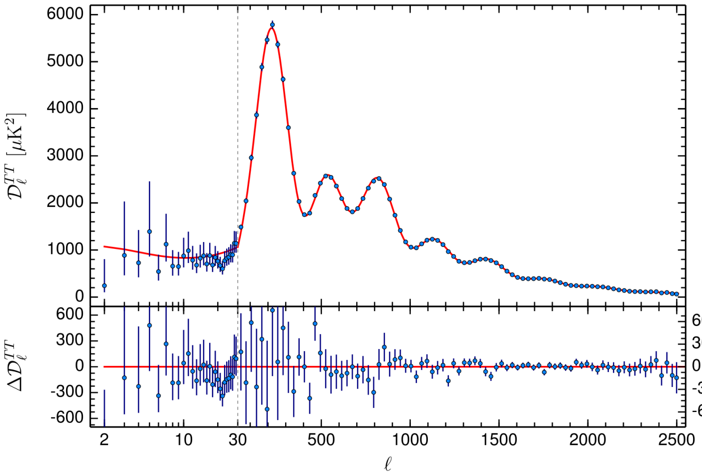


Figure 1.8: Angular power spectrum (as a function of multi-pole moment l) of the CMB temperature anisotropies (top) are shown fit to a theoretical model with the residuals also shown (bottom) [22].

quantum fluctuations in the local matter density. Both baryonic matter and dark matter were attracted to these local gravitational wells, but while baryons experiences a large restoring force from the photon fluid, dark matter was able to collapse freely. These oscillations are observable in the angular power spectrum as a series of "acoustic peaks" and can be seen in figure 1.8. The physics underlying this spectrum can be modeled in detail, allowing many cosmological parameters (such as the baryon density $\Omega_b h^2 \simeq 0.022$ and cold dark matter density $\Omega_c h^2 \simeq 0.120$) to be constrained[11]. A plot of this angular power spectrum is shown in figure 1.8

1.2 WIMP Dark Matter

After our brief discussion of the observed phenomena attributable to dark matter, we can outline a rough set of properties that any potential dark matter candidate must possess.

- **Non-luminous** From all of our astronomical observations, from dynamics of individual galaxies to merging galaxy clusters we know that whatever this extra mass is, it does not give off light.

---

[11] In cosmology, matter density ρ_m is usually denoted by the "density parameter" $\Omega_m = \frac{\rho_m}{\rho_c}$ where $\rho_c = \frac{3H^2}{8\pi G}$ is the critical density associated with a flat universe [23].

- **Dissipation-less** Given the large extent of astronomical DM-halos we know that the dark matter does not effectively dissipate energy or accrete into dense structures.

- **Non-baryonic** The observed CMB and LSS indicate that dark matter started to fall into gravitational over-densities before recombination when the photon-baryon coupling would have prevented ordinary matter from doing the same.

- **Cold** To form these early over-densities the dark matter would need to be non-relativistic or cold at the time of matter-radiation equality.

- **Collision-less** From galaxy-cluster mergers and halo dynamics we know that if dark matter self-interacts, it must do so very weakly.

- **Stable** At this time most matter is the universe is thought to be dark matter, so it can't decay on timescales shorter than the current age of the universe.

The Weakly Interacting Massive Particle (WIMP) is a class of theoretical particles which fulfill these requirements in compellingly natural way. Suppose there is a single undiscovered massive particle (which we will call χ) that is thermally produced in the very early universe. Wile the temperature is much larger than the mass of the WIMP ($T \gg M_\chi$), the creation and annihilation was in thermal equilibrium, leaving the co-moving WIMP density constant. At later times, when $T \leq M_\chi$ a relativity smaller fraction of the lighter particles (such as photons and baryons) will have the kinetic energy needed to produce a χ. Quantitatively this suppression of production will lead to an exponential suppression of the WIMP number density $n_\chi \propto e^{m_\chi/T}$. Due to the expansion of the universe, at some point WIMP-WIMP interactions will become rare enough that annihilation will be inefficient. At this point the WIMP density stops decreasing or "freezes out" and this relic density Ω_χ is the dark matter we observe today, a process that is shown in figure 1.9. This density can be approximately calculated from the annihilation cross section σ_{ann} as in

$$\Omega_\chi h^2 \simeq \frac{3 \times 10^{-27} \frac{\text{cm}}{\text{s}}}{\langle \sigma_{\text{ann}} v \rangle} \tag{1.3}$$

giving a rough annihilation cross section of $\sigma_{\text{ann}} = 0.5\text{pb}$. This is a cross-section consistent with weak-force interactions, giving a potential non-gravitational interaction mechanism, and a chance for terrestrial WIMP detection.

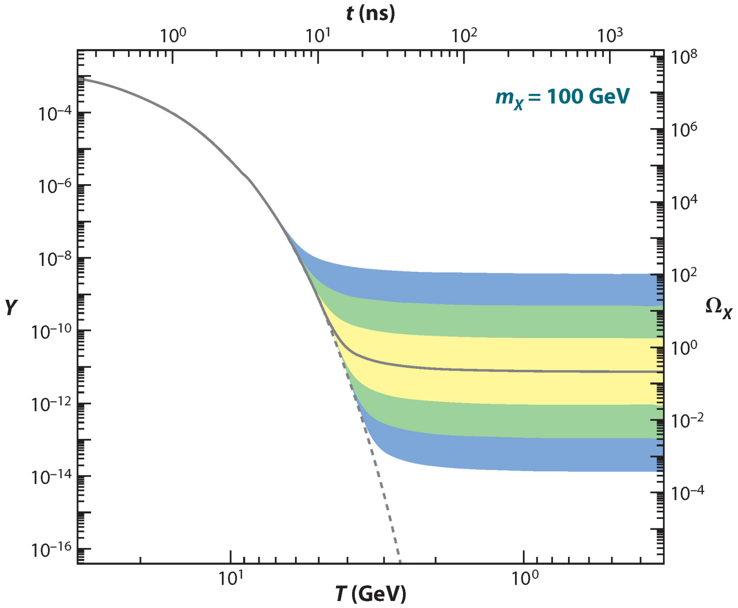


Figure 1.9: Curve depicting WIMP annihilation and freeze out. The co-moving WIMP number density (Y right) and the corresponding DM density (Ω_χ left) are shown vs the temperature of the universe. The Boltzmann-factor annihilation and freeze out is shown for the correct relic density (solid curve). The shaded regions are for cross sections that differ from this value by a factor of 10, 100 and 1000 from this value [24].

1.3 This Dissertation

This dissertation will focus on a terrestrial experimental effort to directly detect WIMP interactions with normal matter. It can be divided into two main parts. Chapters 2–5 will cover the analysis of roughly three years of data, collected in a subterranean array of massive cryogenic particle detectors called SuperCDMS Soudan. Chapters 6 and 7 will cover the development of new style of particle detector that could potentially be used to build a larger version of this type of experiment.

Chapter 2

SUPERCDMS SOUDAN OVERVIEW

2.1 Experiment Installation

SuperCDMS Soudan is a direct dark matter detector experiment consisting of an array of 15, 0.6-kg, Ge interleaved Z-sensitive Ionization and Phonon (iZIP) detectors. SuperCDMS Soudan is the successor to the Cryogenic Dark Matter Search (CDMS II) [25], and these efforts share much experimental infrastructure, including cryogenics, shielding, and electronics. Notably, both efforts took place in the Soudan Underground Laboratory (SUL or just Soudan). SUL was build inside of an old decommissioned iron mine in remote northern Minnesota, near the Boundary Waters Canoe Area Wilderness[1]. The iron-rich rock in the area is ideal for low background physics. It is dense which helps block cosmogenic radiation, as well as being structurally stable and safe for extended human occupation. The SUL is located on level 27 of the mine, which is approximately 713 m below the surface. CDMS-II started its first run in Soudan in 2002 and SuperCDMS ended operations nearly 13 years later at the end of 2015. For a more detailed overview of the experimental setup please see [26], but this chapter will provide an overview appropriate for understanding the work in this dissertation.

---

[1] In the summer, this area of the country is fantastically beautiful and I recommend visiting. In the winter there were weeks where the highest temperature I experienced outside was −20 °F.

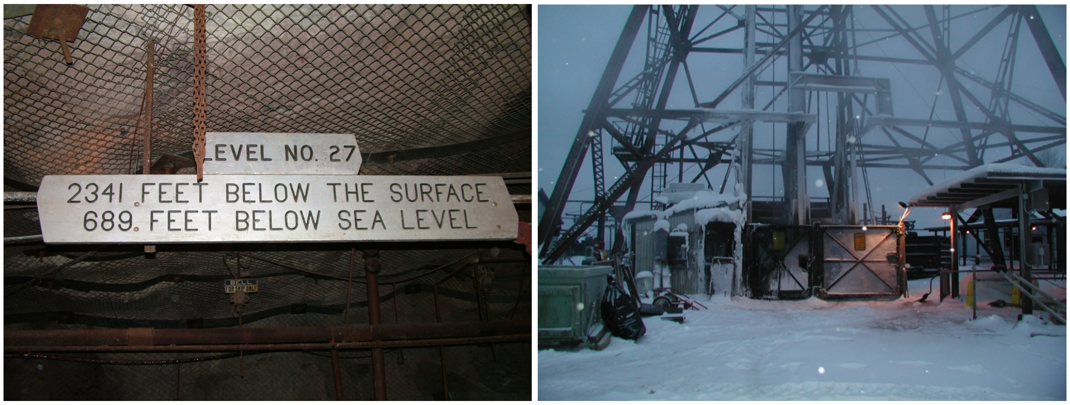


Figure 2.1: Left: The entrance to level 27 of the mine and the home of the SUL [27]. Right: the entrance to the mineshaft or *headframe* in the winter [28].

2.1.1 Shielding

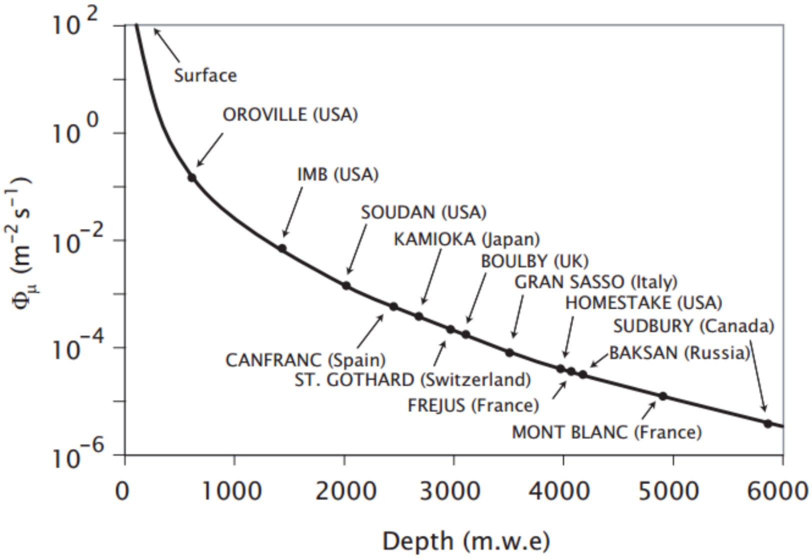


Figure 2.2: Cosmogenic muon flux vs. relative depth measured in meters of water equivalent (m.w.e.) for a selection of underground laboratories. Soudan offers a reduction in the cosmogenic muon rate of more than four orders of magnitude, but for exposures larger than the SuperCDMS Soudan effort, further reduction is required. The next generation of our experiment SuperCDMS SNOLAB will be installed in Sudburry [29].

For any low-statistics particle-physics experiment, reducing the number of unwanted interactions is of paramount importance. SuperCDMS Soudan is no different in this regard, and utilizes several levels of shielding in an attempt to reduce the rate of all non-DM iterations. The first, which was alluded to in the previous section, is the lab itself. The SUF is situated 713 m beneath the surface, an overburden of 2090 meters water equivalent (see figure 2.2). This reduces cosmic ray muon flux by a factor of approximately 50,000. For a target with the exposure of the SuperCDMS detector array, this reduces the cosmogenic neutron background to the order 0.1 event[2]. For larger exposures, a deeper lab is required for neutron-free operation. The second level of shielding protects our experiment from conditions inside of the lab itself, and consists of a radio-frequency-shielded cleanroom colloquially referred to as the RF room. This room protects the experiment from dust-born contaminates, such as radio-isotopes, as well as electronic interference from the rest of the lab. For a graphical layout of the RF room in relation to the other components of the lab please see figure 2.3. Inside of the RF room and immediately surrounding the detectors

---

[2] This is assuming the entire 9 kg detector array is used.

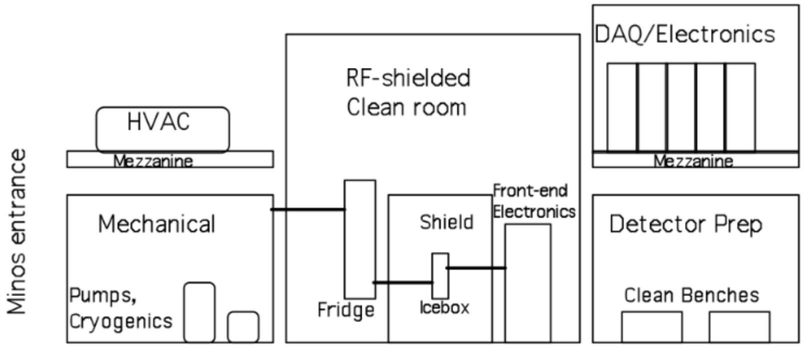


Figure 2.3: Cutaway view of the SuperCDMS components of the SUL [30].

themselves are about 6 layers[3] of additional shielding. The outermost of these is an active muon veto. This consists of an array of plastic scintillator panels arranged such that they cover all 4π sr around the detector. This allows active tagging of cosmogenic muons, which have an average flux of about 1 interaction per minute on the veto. Directly inside of this muon veto are 4 layers of lead and polyethylene. The outermost of these is 40 cm of polyethylene, and acts to moderate the energies of any neutrons. The next layer is 17.8 cm of lead which primarily acts as shielding against high energy photons. This lead is slightly radioactive in its own right, and is shielded by the next 4.4 cm layer of ancient lead. This ancient lead was recovered from a roman shipwreck off the coast of France and has much less ^{210}Pb than typical lead. Inside of the lead is a 10 cm thick layer of polyethylene which would moderate any neutrons that were not moderated by the outer layer (i.e. they were very high energy) or any neutrons produced by material inside the inner lead, as the lead acts sort of like a neutron "mirror". Inside of the ancient lead, and wrapped around the cryostat itself is a thin layer of mu metal. This helps to shield the very sensitive cold electronics from stray magnetic fields. Up to this point the shield is 99% hermetic, with two holes for the cold stem and the e-stem as can be seen in figure 2.4. Inside of this is the cryostat itself which, together with the cold hardware, provide an additional 3 cm of copper shielding.

2.1.2 Cryogenics

The iZIP detectors used in this experiment need to be at very low temperatures to operate correctly. This cooling is accomplished with a three part cryostat as is seen

---

[3] Depending on exactly what is counted as a "layer".

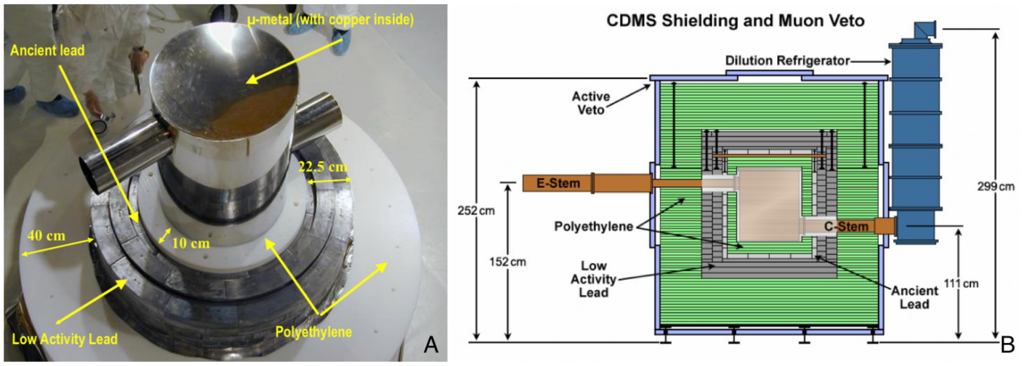


Figure 2.4: Depictions of the IceBox shielding. A: A picture of the passive shielding during construction. B: Cutaway schematic of our shielding, as well as the cryogenic (c-stem and DL-fridge) and electrical (e-stem) support infrastructure.

in figure 2.4B. The detector array itself is housed in a central, thermally-isolated vacuum chamber known as the “IceBox”. This is the volume around which we construct our shielding as described in the previous section. Most of the cooling power is provided by an Oxford Instruments 400S dilution refrigerator[4]. This fridge is not particularly radio pure, and as a result it is situated outside of the active muon shield. The refrigerator is connected to the IceBox via a shield pass-through called the “c-stem”. Inside of the c-stem is an assembly of concentric copper tubes connecting each thermal stage of the dilution fridge with its partner in the IceBox. The the inner-most vacuum chamber in this dilution refrigerator needs to be immersed in a bath of liquid He to maintain normal operation, which was accomplished both by a He reliquification system and, when that was not working, by external He transfers. These transfers, called “cryofills”, introduced enough external vibrational noise that data could not be taken while they were occurring. If the reliquification system was non-operational, a cryofill would be needed about once per day, and would take 90 min[5]. To prevent this unnecessary loss in live time, substantial effort was spent on maintaining the reliquification system.

Opposite the c-stem is another shield pass-through, containing all of the electrical readouts for our detectors called the “e-stem”. To help counter the increased heat-load produced by our JFET amplifiers, a Gifford-McMahon cryocooler was attached through the e-stem to the 4 and 77 K stages of the IceBox for the SuperCDMS upgrade. This cryocooler was a persistent source of noise during Super-

---

[4] Which provides 400 μW of cooling power at 100 mK.

[5] Most of this 90 min was re-cool the detector. This downtime was exploited to neutralize space-charge buildup as described in section 2.2.5, but that hardly made up for the loss in live time.

CDMS Soudan, especially during the end operations when the cooler was nearing the end of its engineered lifetime. This noise manifested during two portions of the cryocooler cycle[6] and were responsible for large low-energy trigger bursts. While this had a substantial impact on the Low Threshold [31] and CDMSlite [32] analysis efforts[7], which are focused on probing for WIMPs with mass $M_\chi > 10$ Gev/c^2, these cryocooler-induced events were largely removed by our increased analysis energy threshold.

2.1.3 Detector Payload

For a complete overview of the SuperCDMS Soudan detector payload please see [33]. In brief, the thermal, electrical, structural and radiological needs of our detectors require specialized hardware inside of the IceBox itself. This "cold hardware" was originally designed and built for the CDMS II experiment and subsequently modified for use with SuperCDMS' iZIPs, and can be broken down into five components.

- **Detector Housing** The iZIP disks themselves are mounted in hexagonal, high-purity copper housings via six Cirlex clamps. Electrical connections are made to each phonon and charge channel via Al wire-bonding to a small custom feed-through circuit mounded to the housing called a detector interface board (DIB). One side of the DIB contains bonding pads and a small LED[8], and the other a Mill-Max connector. There are two DIBs per iZIP, each carrying 12 channels: 8 for phonon bias and readout, 2 for ionization, and 2 for the LED.

- **Side Coax** This is a thin copper plate which, on one end, connects to the DIB via its Mill-Max connectors and carries these signals in vacuum coax wires to the detector tower base. It also contains a small feedback circuit on the DIB side to correctly bias and read out the ionization electrodes.

- **Tower** Our iZIPs are mounted in five groups of three detectors in what is usually referred to as a "stack". These stacks are mounted to the bottom of a

---

[6] Which were designated *chirp* and *thump* after their coincident sounds.

[7] To characterize cryocooler noise for low mass WIMP searches an accelerometer was installed on the e-stem to record the timing of these noise events.

[8] This LED is used for detector neutralization.

large hexagonal copper structure called a “tower”[9]. The tower carries signals from the stack, at base temperature, up to the SQUET card at 4 K. Moving upward it has 4 thermally isolated stages[10] to minimize thermal conduction to the detectors.

- **SQUET Card** At the top of the tower is a card containing the first stage amplifier for both the charge and phonon signals. The phonon signals are amplified using a very sensitive type of magnetometer called a superconducting quantum interference device (SQUID) which are thermally connected to ST temperature (about 800 mK). The primary charge amplification is done using JFETs which are at 4 K. These cards are named with a portmanteau of these two acronyms: “SQUET”[11].
- **Stripline** Finally, the electrical signals are carried from the SQUET carts, through the e-stem, to 50-pin bulkhead connectors where they are externally accessible. The vacuum chamber at he end of the e-stem that contains these vacuum feed-through connections is called the “e-box”.

When talking about a particular detector, there are two different naming schemes that are used interchangeably. This first is to use the tower number followed by the stack position, where the first (top) detector of the fourth tower would be IT4Z1. The detectors are also numbered consecutively starting at 1101, so IT1Z1 is 1101, IT1Z2 is 1102…, IT5Z2 is 1114 and IT5Z3 is 1115.

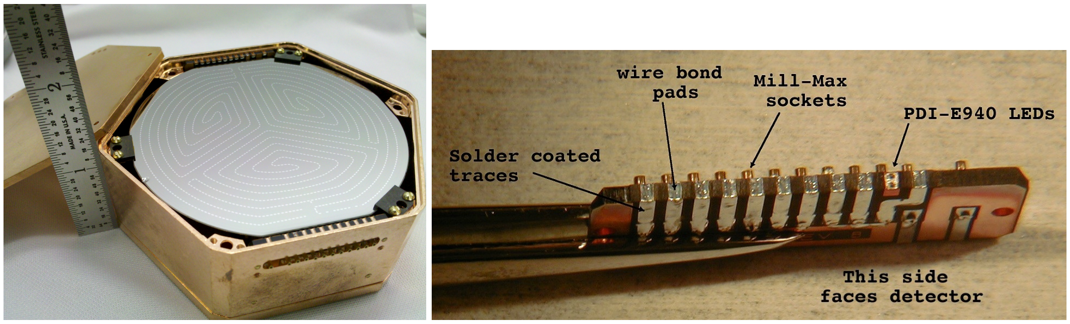
It was found in CDMS II that near-surface interactions, especially those from the ^{210}Pb decay chain, were an important source of background interactions. To better characterize this background two ^{210}Pb sources were installed *in situ*. One was placed on the top face of detector IT3Z1, and one on the bottom face of IT3Z3. These sources generated approximately 130 events per hour, and are of vital importance for the construction of our surface background model (see section 4.3.4).

---

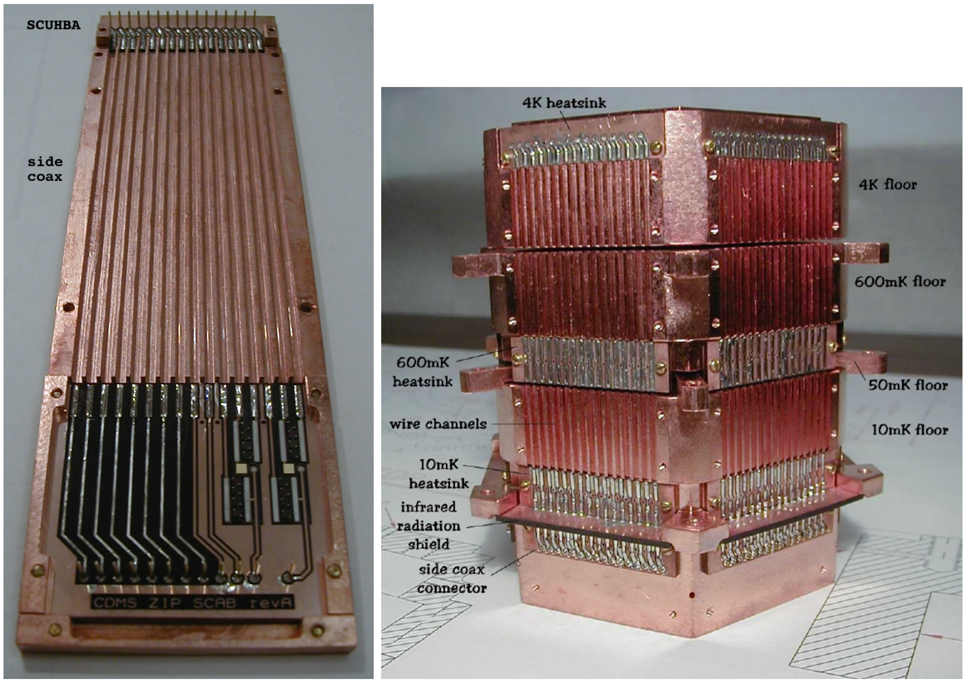
[9] The terms *tower* and *stack* are often used interchangeably in the CDMS collaboration to refer to the the stack, the tower, the combination of the two, or the entire assembly including the SQUET card. It is possible I will be similarly ambiguous in this dissertation.

[10] Starting from the stack side the stages are MC, CP, ST and 4K. They are connected thermally through the c-stem to the mixing chamber, cold plate, still and He bath of the dilution fridge respectively.

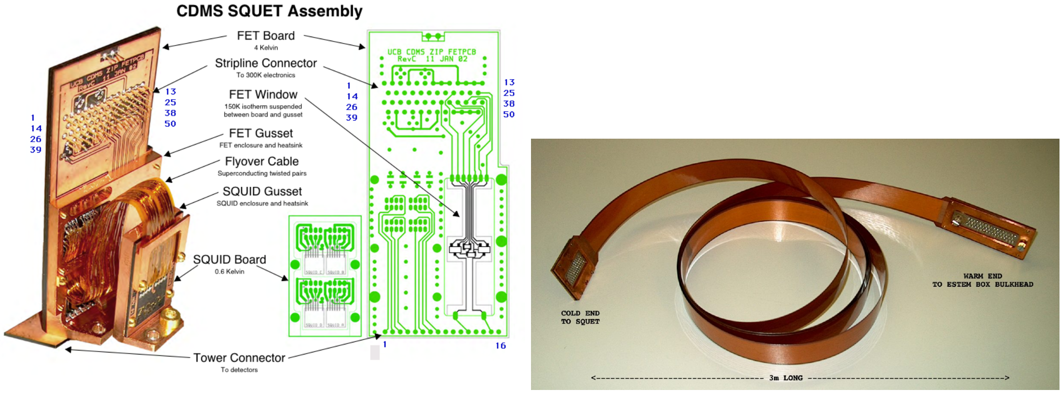
[11] As a people, experimentalists are nothing if not bad at naming things.



(a) Detector housing. (b) DIB connector



(c) Side Coax (d) Tower



(e) SQUET card (f) Stripline

Figure 2.5: Overview of cold electronics.

2.1.4 Warm Readout Electronics

After leaving the e-box the signals are input into a series of control and data-acquisition circuits collectively referred to as “warm electronics”. There are three main stages to the warm electronics system: the Front-End Boards, the Receiver-Trigger-Filter Boards, and the Trigger-Logic Board.

- **Front-End Board (FEB)** FEBs are custom 9U circuit boards that control the bias and readout of the phonon channels, charge channels, JFETs, SQUIDS, and LEDs. Each FEB is responsible for a single DIB, so each detector is read out with 2 FEBs. They are rack mounted and physically located near the e-box in the RF room itself.

- **Receiver-Trigger-Filter (RTF) Board** RTF boards are also custom 9U circuit boards, each one of which receives its input from a single FEB. As the name suggests these boards are used to set a trigger threshold so that only signals above a certain amplitude would be read out. This trigger is defined on the sum of all the phonon channels for a particular detector. In addition to issuing triggers, the RTF board also implements a couple of filters. A Butterworth band-pass filter with bounds of 300 Hz–3 kHz, is used on the signal that is sent to the trigger discriminator to prevent random noise from dominating our data acquisition. The digitized signal is low-pass filtered to prevent aliasing. These boards are rack mounted in the electronics room which is above the RF anteroom.

- **Trigger-Logic Board (TLB)** The TLB has a global overview of the entire experiment and is responsible for issuing global triggers. The TLB collects triggering information from the RTF boards as well as the muon veto, and the main DAQ control computer[12]. When a global trigger is issued, what happens depends on the operational mode (see section 2.3.3). In short, while collecting WIMP-search data, a global trigger will cause the entire detector array to be read out. This allows for multiple-interaction events to be tagged, and for the iZIP array to act as an active self-shield. During calibration data taking, the trigger rate is much higher[13] and as a result only detectors that issued a trigger are read out.

---

[12] Which is used to issue random triggers to better understand the noise environment.

[13] Around 20 Hz vs the 0.5 Hz for WIMP-search data.

2.2 iZIP Detectors

2.2.1 iZIP Overview

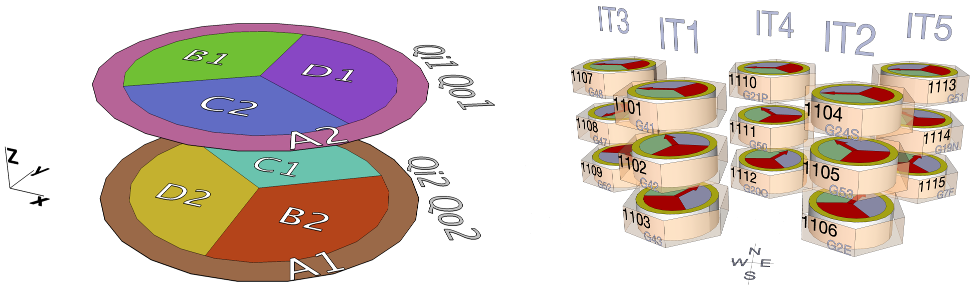
The majority of the experimental installation described above was put in place for the CDMS II experiment. The main upgrade offered by SuperCDMS soudan are the interleaved Z-sensitive Ionization and Phonon (iZIP) detectors. Each of the 15 iZIPs installed in the SuperCDMS soudan apparatus is a 0.6 kg crystalline high-purity[14] Ge disk that is 76 mm in diameter and 25 mm thick. These detectors were designed with two goals in mind: the reduction of backgrounds (via rigorous materials selection, handling and screening) and event-by-event discrimination between signal and background interaction events. For the sake of the latter requirement, these iZIPs are actually two detectors in one. Each detector face is photolithographically instrumented with interleaved ionization and phonon sensors. Between the two faces, the ionization channels are essentially identical, consisting of a disk-shaped inner electrode concentric with an annular guard electrode. The phonon sensor configuration is more complicated. Each face has an annular guard phonon collector that matches the guard ionization collector. The phonon sensors on the inner disk, however, are further subdivided into three equal wedges. The orientation of these wedges is rotated by 60 degrees between the two faces, to break any degeneracies that may arise during phonon position reconstruction. For an overview see figure 2.6. The following sections describe the phonon and ionization signal collection of an iZIP at an appropriately high level, for a much more in-depth treatment of this subject please see [34, 33].

2.2.2 Phonon Signal

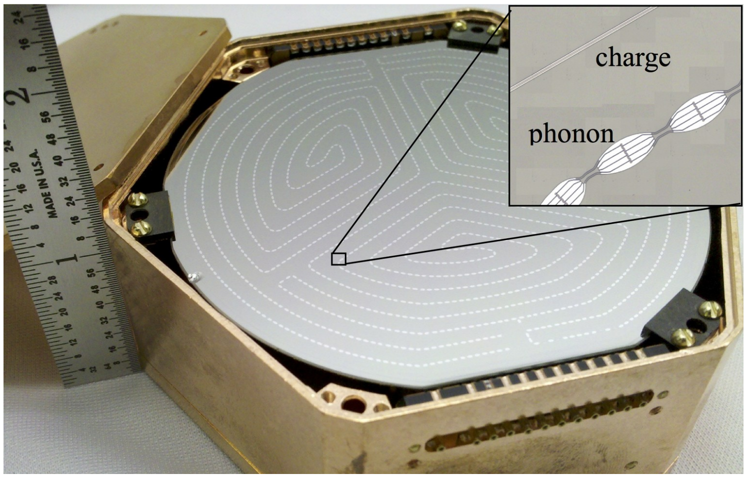
A particle scattering off of a Ge nucleus causes the nucleus to recoil in the crystal. This nucleus is large, highly charged, and quickly loses energy. This energy can be lost both to lattice site excitations (phonons) and if the recoil is energetic enough, to the liberation of valence electrons (and holes) into the conduction band of the Ge. Here, they can be drifted out for collection (for more on this see section 2.2.5). This process produces three different populations of phonons: primary recoil phonons, Luke-Neganov phonons, and electron-hole recombination phonons.

---

[14] A dislocation and impurity density of 10^{10} cm^{-3}



(a) Orientation of phonon collectors. (b) Tower array.



(c) Charge/phonon sensor detail.

Figure 2.6: iZIP detector overview.

- **Primary phonons** As mentioned above, a recoiling charged particle in a crystal will directly lose energy to lattice vibrations. These “primary” phonons are initially very high energy[15] and are localized around the interaction-induced recoil. At such high frequencies the phonons have very short mean free paths[16], due to two main energy-dependent scattering mechanisms: anharmonic decay, where the scattering rate is proportional to ν^5, dominates at frequencies above 1.6 THz, followed by isotope scattering, where the scattering rate is proportional to ν^4 [35]. Although isotope scattering is elastic, anharmonic decay is not, and as a result the phonons will rapidly down-convert in energy. This rapidly-scattering regime is referred to as quasi-diffuse propagation. As a result, this early portion (< 0.1 ms) of the phonon pulse contains information about the interaction location. As the frequency of the phonons decrease, both of these scattering rates will decrease, eventually reaching a

---

[15] They are initially at optical frequencies of around 10 THz.

[16] Much smaller than the physical size of the Ge crystal.

point where the mean free path is larger than the size of the detector. In this regime, the phonons are said to propagate "ballistically", and are uniformly distributed throughout the detector.

- **Neganov-Trofimov-Luke phonons** The second class of phonon comes from drifting the electron and hole pairs out of our detector and are called "Neganov-Trofimov-Luke" or just "Luke" phonons. As described in [36, 37], as a charge carrier is drifted across a crystal, it interacts with the lattice emitting phonons. This energy loss mechanism gives the drifting charges an effective terminal drift velocity, v_d. The kinetic energy of these charge carriers is low (~10 meV) and it will eventually be reclaimed as phonons during recombination (below), leaving a Luke-phonon energy, E_{Luke}, that is equal to the change in potential energy experienced by the charge carriers or

  $$\begin{aligned} E_{Luke} &= n_e e \Delta V_e + n_h e \Delta V_h && (2.1) \\ &= ne\left(\Delta V_e + \Delta V_h\right) && (2.2) \\ &= ne\Delta V_b && (2.3) \end{aligned}$$

  Where n_e and n_h are the number of electrons and holes respectively, which are produced in equal quantities. ΔV_e, and ΔV_h are the potentials that the electrons and holes drift through, but provided that each carrier is collected, the total Luke phonon energy only depends on the total bias voltage ΔV_b, and not the interaction location.

- **Recombination Phonons** During their production, each electron-hole pair is given potential energy equivalent to the band gap energy, E_g, in Ge. When the carrier reach the ionization channels on the faces of the detector they relax back to the Fermi Level, each giving back half of E_g.

2.2.3 Phonon Readout

As discussed above, during down-conversion, the phonon signal is localized and contains position information. At later times, the phonon distribution will thermalize and be uniformly distributed in the detector. Historically, most experiments of this type[17], were specifically designed to measure this thermal-equilibrium population of phonons. Aside from the obvious shortcoming of ignoring all position

---

[17] Such as Edelweiss [38] and CDMS II [39]

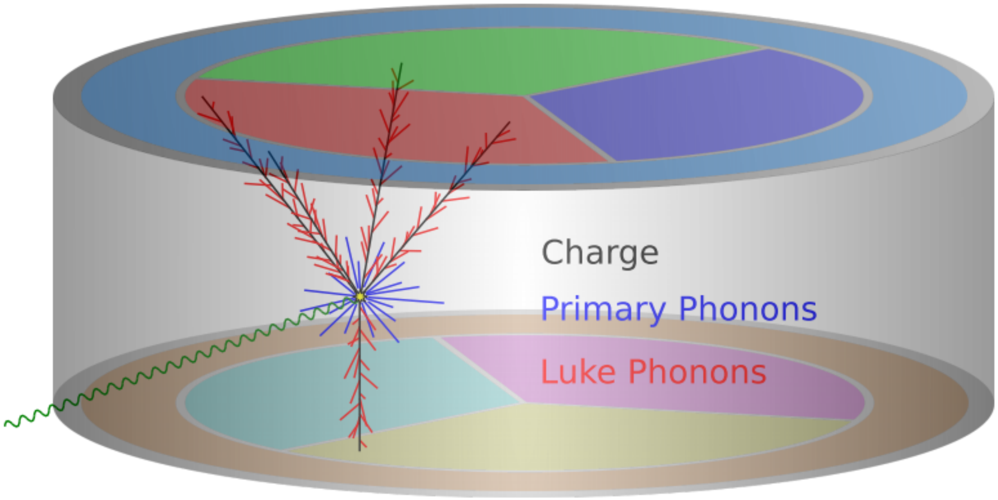


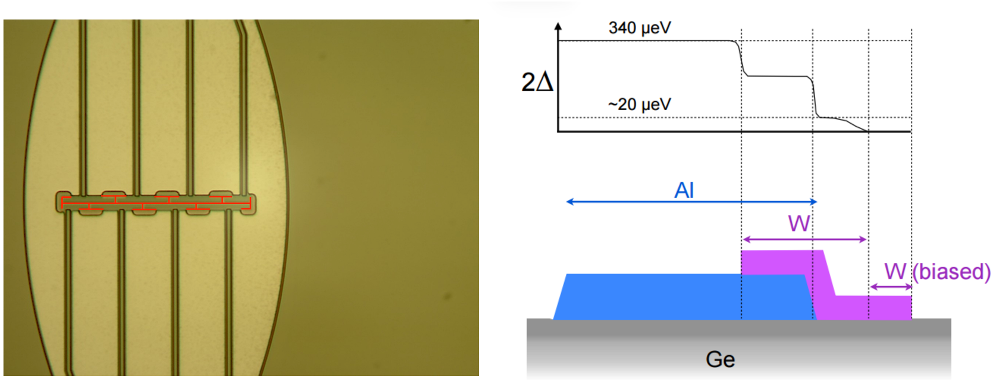
Figure 2.7: Qualitative particle interaction overview for an iZIP detector. An incident particle (green) interacts with an electron or a bound nucleus. This produces primary phonons (blue) as well as electron-hole pairs (black). By drifting the electrons and holes across the crystal Luke phonons are also produces (red). The oblique electron propagation is described in more detail in section 2.2.5.

information, this approach fundamentally limits the crystal size. Measuring the thermal phonon population is essentially measuring the change in temperature the entire crystal undergoes during a single particle interaction, which is inversely proportional to the total heat capacity of detector or $\Delta T \propto \frac{1}{C}$. At very low temperatures and substrate masses C is small enough that this temperature change can be effectively measured. However, if the detectors are made too large, this change in temperature will be vanishingly small. To overcome these limitations, iZIP detectors are designed to resolve a portion of the early athermal phonon signal.

The iZIP detector detects phonons using quasiparticle-trap-assisted electro-thermal-feedback transition-edge-sensors (QETs), which are two-component phonon sensors. These consist of an Al absorber fin that acts as a quasiparticle funnel, coupled to a W transition edge sensor (TES). Each of the eight phonon detectors described in section 2.2.1 consist of 458 QETs that are read out in parallel. These QETs are spread throughout the pixel area. When an incident phonon encounters one of the Al absorber fins it can split a Cooper pair in the superconducting Al ($T_c = 1.2$ K) provided the phonon energy is greater than the Cooper pair binding energy ($E_{\text{binding}} = 2\Delta(0)_{Al} = 340\mu$eV). The resulting pair of quasiparticles[18] can diffuse along the Al film until it encounters the W TES. Empirically, these quasi-

---

[18] These are similar to but distinct from electrons as we know them in the vacuum.



(a) QET detail. Al absorbers (silver) and W TES's (Orange)

(b) Trapping Overview

Figure 2.8: Left: Detail of eight QET's. Note the relative sizes of the silver Al absorber to the orange W TES. Right: A schematic depicting quasiparticle trapping. Quasiparticles created in the large blue Al absorber fins have a ground state energy of 340 μeV. The decrease in ground state energy in the W TES causes an increase in the number of low-energy quasiparticles that can't transition back into the AL absorber film and become trapped in the TES [41].

particles have a characteristic diffusion length in our Al of $l_D = 180\,\mu$m [40], which sets the design length of the absorbers. Once they reach the TES, the quasiparticles can enter the W film and break additional Cooper pairs in the process. The critical temperature for W (T_c = 50 mK) is much lower than it is for Al, and as a result the gap energy is much lower in the TES than it is in the absorber fin (340 vs 20 μeV). When these relatively high energy Al quasiparticles fall down into the relatively lower quasiparticle ground state energy of the W, the excess energy can go into breaking W Cooper pairs and creating more quasiparticles. Because these quasiparticles are all below the Al gap energy, they are effectively trapped in the W TES. In this way, the overall effect of this absorber fin / TES system is to collect quasiparticles over a large area of the detector surface ($\sim 5\%$ metal coverage) and funnel them into the TES where they are trapped and concentrated in a much smaller volume. A portion of the W TES is biased and heated into its superconducting transition and this process can repeat causing in the biased (zero gap) portion of the TES. A QET overview is given in figure 2.8.

Our W TESs are voltage biased via a shunt resistor (R_{sh} = 22 mΩ << $R_{\text{TES}} \approx$ 200 mΩ) and read out using the circuit show in figure 2.9. As it is biased, the Joule heating ($P = V_{\text{b}}^2/R_{\text{TES}}$) due to the bias current of the TES will cause it to self-heat into its[19] superconducting transition, a state referred to as negative electro-thermal feedback (ETF). When an interaction occurs, R_{TES} increases due to the concentration of quasiparticles trapped in the active (biased) portion of the TES. This decreases the Joule heating allowing the TES to cool back to its steady-state self-heating point in the superconducting transition. At the operational temperature of an iZIP, the phonon-electron coupling is very weak causing the thermal conductance between the TES and the substrate to be low as compared to the ETF conductance. As a result the TES re-cools from the decrease in Joule heating rather than from heat flow to the substrate. This allows the substrate to be maintained at base temperature while the W TES is held in the middle of its superconducting transition.

The change in current through each TES is read out using a dedicated, low-impedance, SQUID amplifier circuit as shown in figure 2.9. The TES is biased in series to an inductor called the "input coil" (L_{i}) which translates changes in current to changes in flux through the SQUID. Although the SQUID is a very sensitive magnetometer, it

---

[19] The exact point in the superconducting transition can be tuned by changing the bias voltage. Practically, this is controlled by the current sent down the bias network. Typically, you first have to drive enough current to make the TES normal, then ramp it down to get the desired bias voltage.

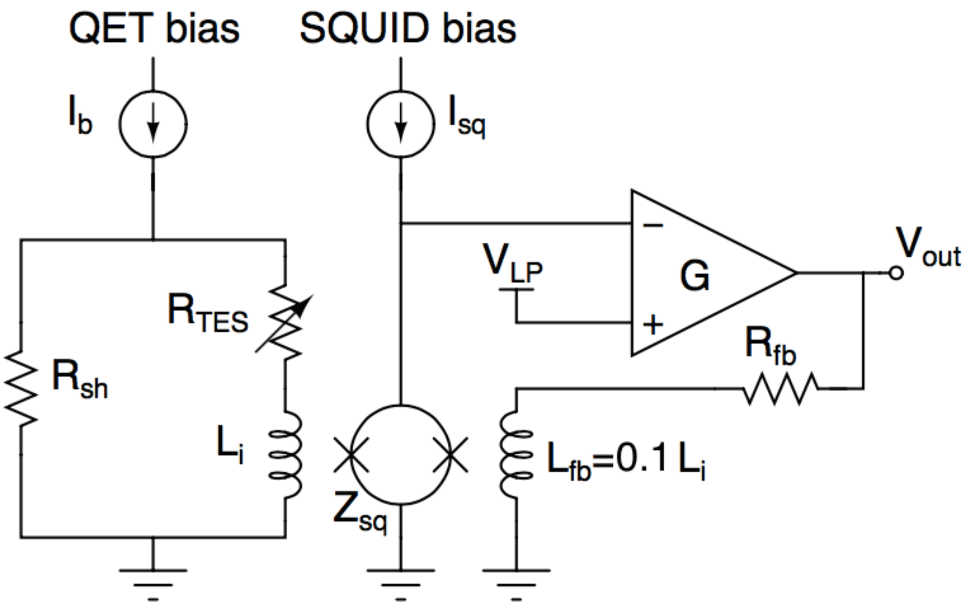


Figure 2.9: Overview of the QET readout circuit. Any change in quasiparticle-sourced dissipation (R_{TES}) will cause a change in the flux produced by the input coil (L_{i}). This flux will induce op-amp G to provide current to the feed-back inductor L_{FB} until the total flux through the SQUID is canceled.

operates most accurately in a closed-loop feedback circuit as shown in figure 2.9[20]. The amplifier in the feedback loop supplies the feedback inductor (L_{fb}) with current until the field it produces cancels field from the input coil. Because $L_{\text{fb}} = L_{\text{i}}/10$, the small voltage signal produced by the SQUID will be amplified by a factor approximately 10.

The noise from the phonon readout circuit has many different components, but is dominated by the Johnson noise from the shunt resister. This is due to the fact that the shunt resister is thermally tied to the still stage ($T \approx 600$ mK), which is substantially warmer than the ~50 mK TES. The resulting white noise should be ≈ 10 pA $\sqrt{\text{Hz}}$ with a high frequency roll off set by $L_{\text{i}}/R_{\text{TES}}$. The measured noise spectrum *in situ* had much higher than expected noise at low frequencies (see figure 2.10). This excess is believed to arise from 1/f noise from our electronics, together with vibrational coupling of our towers with external elements of the lab[21]. Although problematic for very low energy events, this analysis was relatively unaffected by the excess noise.

2.2.4 Phonon event reconstruction

At this point we have turned our collected phonon energy into an amplified voltage pulse. The next thing we need to do is reduce this pulse into useful physical quantities, such as the total collected phonon energy. To this end SuperCDMS uses several reconstruction algorithms, which I will detail below.

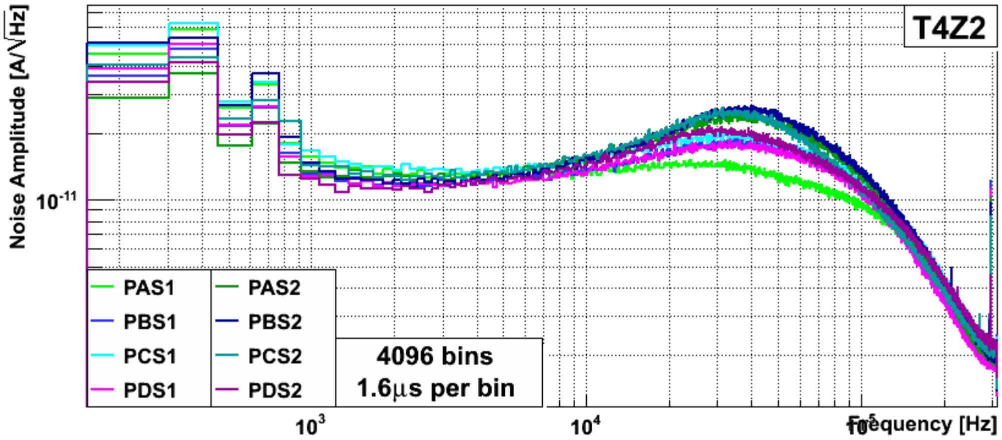


Figure 2.10: QET noise power spectral density (PSD) for IT4Z2.

---

[20] The SQUID’s voltage response is highly nonlinear.

[21] Such as the e-stem cryocooler.

Optimal Filter (OF)

The most important reconstruction algorithm is called the "Optimal Filter" (OF)[22]. The goal of any type of filter is the suppress frequencies that are overwhelmed by noise while keeping those that convey the desired signal. To do this the optimal filter compares a digitized time-domain voltage trace (V_t) to a signal template, S_t with the form

$$V_t = AS_t + n_t \tag{2.4}$$

Where n_t it is a stochastic noise signal with a PSD modeled after the noise environment in the detector and A is the pulse amplitude that estimates the interaction energy. This amplitude is found via χ^2 minimization. As the noise components are correlated in the time domain, this optimization is carried out in the frequency domain where they are independent. In this domain, χ^2 as a function of the amplitude, A, and pulse start time, t_0, takes the form

$$\chi^2(A, t_0) = \sum_{\nu} \frac{|\widehat{V}_\nu - A\widehat{S}_\nu e^{-\iota 2\pi\nu t_0}|^2}{\widehat{n}^2_\nu} \tag{2.5}$$

and has the effect of filtering hardest on frequencies of high signal to noise. For our phonon channels the signal template is determined empirically, and a single template is used for all channels on a given detector. This template is built from the average of many traces taken from high-quality bulk electron recoil events. The noise component n_t is similarly determined from randomly collected noise traces.

Non-Stationary Optimal Filter (NF)

The primary shortcoming of the OF is that it assumes a single template is a valid description of all phonon pulses in a given detector. While this is a good assumption for charge pulses (see section 2.2.7), phonon pulses exhibit significant position dependence. As is seen in figure 2.11 this position dependence occurs in the first $\sim$ 100 μs of the pulse. The goal of the NF is to de-weight this portion in the χ^2. The way we do it is to construct an average signal template using many traces and then subtract this template from the traces themselves obtaining the residuals. This

---

[22] For a very detailed overview of the OF formalism please see [42, 30, 43].

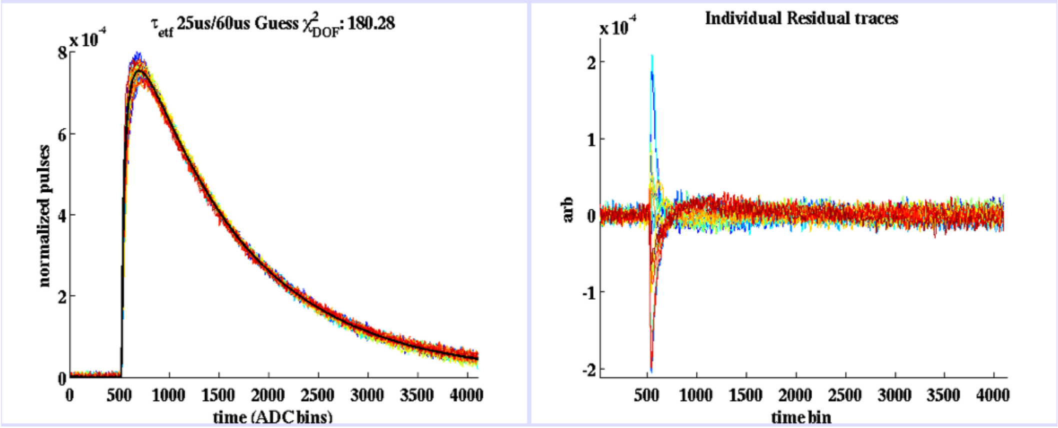


Figure 2.11: Non-stationary optimal filter motivation. Left: the raw phonon pulses for a bulk-recoil event (colored) vs the template used by the OF reconstruction algorithm (black). Right: the residuals between the template and the raw pulses. As can be seen the first few hundred ms of the pulse suffers from position dependence. This can be modeled as a kind of "noise" and de-weighted.

residual is then treated as a source of time dependent or "non-stationary" noise[23]. For more on this technique, please see [43]. Although used for estimating the total phonon energy collected during an interaction, the estimates for the energy collected for each individual phonon channel is reconstructed using the OF algorithm. This preserves the position information of the interaction which is of interest.

Phonon Pulse Integral (INT)

A much simpler way to reconstruct the energy from a given voltage pulse is to simply integrate it (or more correctly, sum the digitized pulse values). This method is susceptible to poor performance at low event energies due to the effects of low-frequency noise, but was useful in two contexts. First, our blinding criteria was set before the OF energy calibration was finalized, and as a result relied on INT quantities in its definition. Second, for pulses of very large amplitudes the TES is driven normal and the templates used in the OF are no longer accurate. The INT energy estimates, however, are still accurate. For this reason it was used in the high-energy α-particle study done to help characterize our surface background rate as described in section 4.3.4.

---

[23] Although in reality it is not noise but position dependence in the signal. If we understood how to correctly parametrize this position dependence we could add it to the template, rendering our noise stationary again.

Phonon Tail Fit (F5)

To preform the relative phonon energy calibration (section 2.3.4.2), the thermal portion[24] of the phonon pulse is fit to an exponential of the form $V(t) = A\exp(-t/750\,\mu\text{s})$. The fit value of A is then the channel weight used in the relative phonon calibration. This was also useful in the high energy α-particle study as an energy estimator when the TES response saturates.

2.2.5 Ionization Signal

For a more detailed overview of the iZIP ionization collection system please see [44, 45, 46]. The basic idea is the same as any traditional semiconductor particle detector; apply a bias voltage to a semiconductor crystal and measure the current arising from particle-interaction-produced charge carriers. iZIPs, however, operate at much lower temperatures and bias fields than most particle detectors. In Ge, as in all non-conductors, the Fermi level, E_f falls in the band gap and given that the probability that an electron energy state, E_e is occupied is given by the Fermi function:

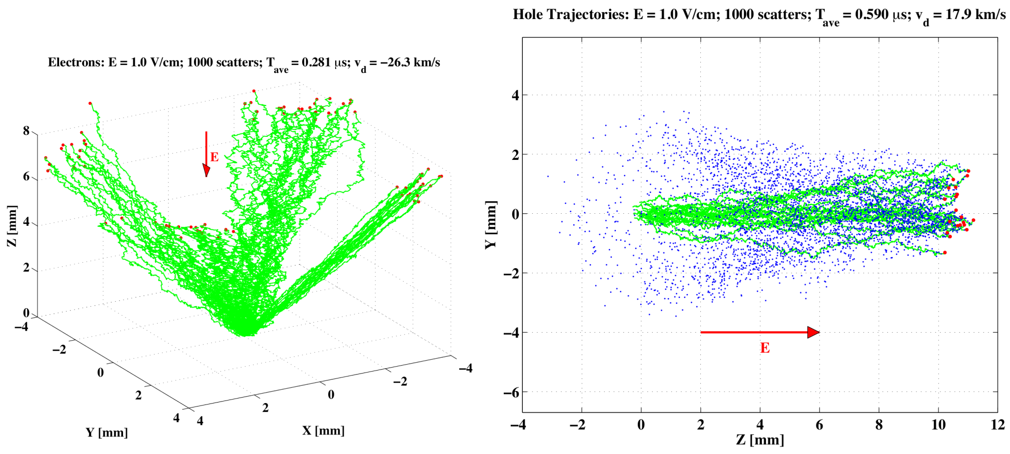
$$F(E_\text{e}) = \frac{1}{e^{(E_\text{e}-E_\text{f})/kT} + 1} \tag{2.6}$$

At low enough temperatures, even the lowest electron energy state in the conduction band will be too far from the Fermi Level to be thermally populated. In this regime the crystal is said to be “frozen out” and will act as an insulator. If an incident particle interacts with the crystal and imparts more than the band gap of energy, it is possible for an electron to be kicked out of the valence band and into the conduction band. This conduction electron may then be drifted to the positively biased electrode and collected[25]. The valence band vacancy, or hole, is similarly mobile and can be thought of as a quasiparticle with a charge of positive e. In Ge ($\Delta(0) = 0.74$ eV) the average energy required to produce an electron-hole pair is 3 eV. As a result of the shape of the Ge conduction band, conduction-band electrons behave as if they have a tensor mass [47]. This causes the electrons to propagate along four elliptical channels aligned along the [111], $[\bar{1}11]$, $[1\bar{1}1]$, and $[\bar{1}\bar{1}1]$ lattice directions. At the low drift voltages used in this analysis, electrons will largely re-

---

[24] After the ~0.1 ms down-conversion is complete

[25] “Collected” is slightly misleading. As the charge carrier approaches the ionization electrode it induces a mirror charge in the electrode. As it drifts closer, this mirror charge density increases, inducing a mirror current in the electrode. It is this mirror current that is detected. Upon reaching the boundary of the crystal the charge carrier itself relaxes back to the Fermi level.



(a) Electron Transport

(b) Hole Transport

Figure 2.12: Simulation of charge transport in an iZIP from [47]. Left (a): electron propagation simulation in Ge. Due to the indirect nature of the Ge band gap the electrons propagate along four channels aligned along the [111], $[\bar{1}11]$, $[1\bar{1}1]$, and $[\bar{1}\bar{1}1]$ lattice directions. Right (b): hole propagation simulation.

main in the channel they initially propagated into. From simulation, these channels were measured to be ≈33°from vertical as can be seen in figure 2.12. The holes are much simpler and propagate as if they have an effective mass that is a scalar, and any spreading in the XY direction is due to random scattering. This causes the effective spot size of the collected charge carriers to be quite different on the electron collecting face, and hole collecting face. This has important implications for position reconstruction (see section 2.4.4), and can also cause incomplete ionization signal collection due to surface trapping.

Surface Trapping

The simplified model of a our Ge semiconductor, with its valence and conduction band separated by a clean band gap, is fairly accurate in the bulk of the crystal, but begins to break down near the surfaces of the detector. The concern is that interactions near the surface of the detector could suffer from incomplete collection of the ionization signal for an interaction, which substantially confuses our ability to differentiate between signal and background on an event-by-event basis. Qualitatively, we approach this phenomena from two angles, one for interactions occurring near the flat top and bottom surface of the detector (called the detector “faces”) and the other for interactions near the vertical cylindrical portion of the detector (called the detector “sidewall”).

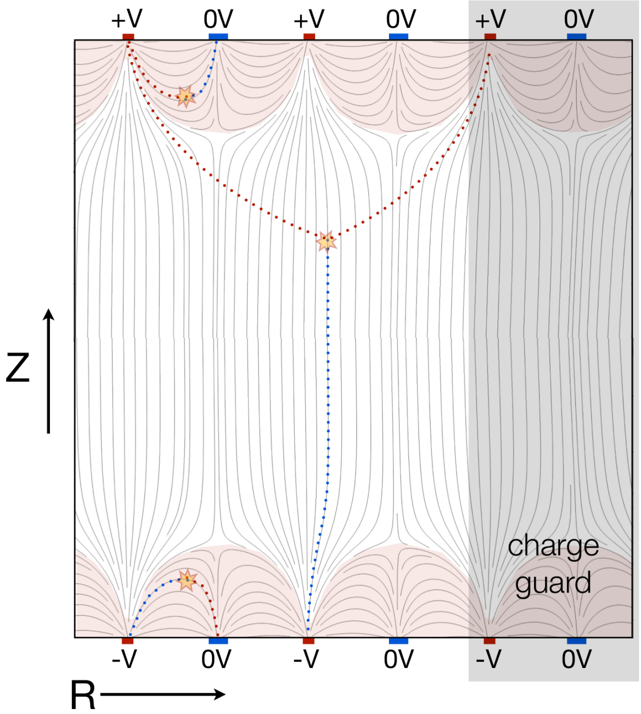


Figure 2.13: Ionization collection for three different classes of particle interactions. A bulk interaction is expected to undergo symmetric collection of electrons and holes. Any near-face interactions however will preferentially be collected on the same side, allowing for an ionization-based Z-position proxy to be constructed [44].

For interactions very near[26] the faces of our detectors a portion of the electron-hole cloud will diffuse directly into the ionization collecting electrode[27], and will suppress the ionization signal by up to 30%. This can be mitigated to some degree by electrically isolating the ionization electrodes from the crystal substrate. This was a large problem for the oZIP detectors used in CDMS II, and the impetus for the new interleaved design of the iZIP detector [34]. During normal operation, the ionization electrodes are biased to ±2 V, and the phonon lines are grounded. This creates a uniform drift field in the volume, but near the faces the field is tangential to the detector surface and much higher in strength. This can be seen in figure 2.13. Due to this field configuration only interactions occurring in the bulk of the crystal will induce ionization signals on both faces of the detector; those near the faces will see the electrons and holes collected on a single face. This allows analyzers to selectively exclude near-face events.

Interactions near the sidewall are also problematic. As this is a physical boundary,

---

[26] On the order of a few microns [48].

[27] A process referred to as “back diffusion”

the lattice is unable to repeat regularly. Even if the surface were perfectly smooth, regular, and pure, the Ge itself would have many dangling bonds that could potentially trap electrons and holes and prevent their collection. In reality the surface is rough and amorphous, allowing for many energy levels in what would otherwise be the band gap. These local potential energy wells can easily acts as traps at the low drift-fields used in our detectors. The oblique propagation exhibited by electrons makes the ionization signal collected on the top (electron-collecting) face especially susceptible to under-collection. This effect is dependent on the interaction location, and is addressed in analysis in a number of ways. First, when estimating the total ionization energy of the interaction, only the face that collects more ionization is used or

$$E^{\text{Ionization}}_{\text{total}} = \max\{E^{\text{Ionization}}_{\text{electron}}, E^{\text{Ionization}}_{\text{hole}}\} \tag{2.7}$$

Within the CDMS Analysis Package[28] (CAP) this total ionization energy estimate is spelled `qsummaxOF`. Second, the ionization collectors for each face are arranged into an inner disk collector surrounded by an outer annular electrode. This allow us to construct a proxy for the radial position of an interaction by examining the partitioning of signal between these two detectors. This technique is of central importance to this analysis and covered in much more detail in section 2.4.4.

Bulk trapping / neutralization

Although the very surface of the detector is always a problem from a charge collection standpoint the charge carriers may also run into propagation problems in the bulk of the crystal. Any impurity, or lattice defect may act as a scattering site, and at very low temperatures and bias voltages electrons and holes can be trapped at such sites as can be seen in figure 2.14. As holes are drifted downward toward the negatively biased face and electrons towards the positive bias, over time this trapping will produce a continuous space charge distribution that canceles the drift field, severely impacting the ionization signal. There are a few methods that are used to neutralize this space charge and, although the details depend on the operating mode of the detector in question, in general they involve grounding both faces of the detector and then producing many charge carriers in the detector. These will neutralize the space charge.

---

[28] A custom MATLAB-based analysis environment.

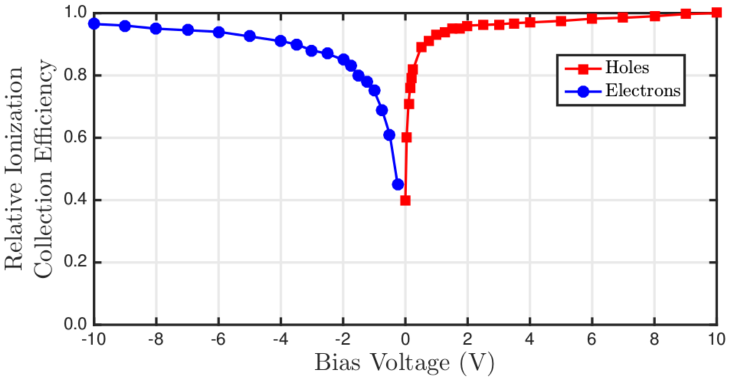


Figure 2.14: Collection efficiency vs bias voltage (V_b). At higher bias voltages most charge carriers can overcome any local traps, and the collection efficiency is quite high. This regime produces excessive Luke phonons however, and the optimal balance between the two effects was chosen at ±2 V [44].

During WIMP-search operation, event rates are low enough that the timescale of space-charge buildup is on the order of hours. As a result our WIMP-search data is taken in time-blocks or “series” of three hours, and between each series our detectors are neutralized via a process referred to as “flashing”. Each of our iZIP detectors is instrumented with 4 LEDs installed on the readout DIBs near its surface, which are briefly illuminated while the detectors are grounded between series. Although these photons only excite electron-hole pairs in the outer $\sim$1 mm of the crystal, the field from the space charge itself will, in the absence of a bias voltage, allow the detector to be neutralized. After neutralization, the detectors are re-cooled and the next series of data-taking may be started.

Event rates during calibration are much higher, and as a result it takes tens of minutes, rather than hours, for space-charge buildup to become a problem. This is the same timescale as a post-flash cool-down. In an effort to keep our data-taking duty factor above 50%, we can actually exploit the high event rate from calibration as our source of neutralizing charge-carriers. During calibration, neutralization simply requires briefly grounding voltage bias lines.

There is another source of space charge that requires special attention. During the initial cooling (or re-cooling) of our detector from high temperatures, many charge carriers from the conduction band will relax into meta-stable traps in the band gap (levels that are again caused by impurities, vacancies or lattice abnormalities). Unlike the space charge distribution that arises during normal detector operation, however, the trapped electrons and holes will be randomly positioned in the detector.

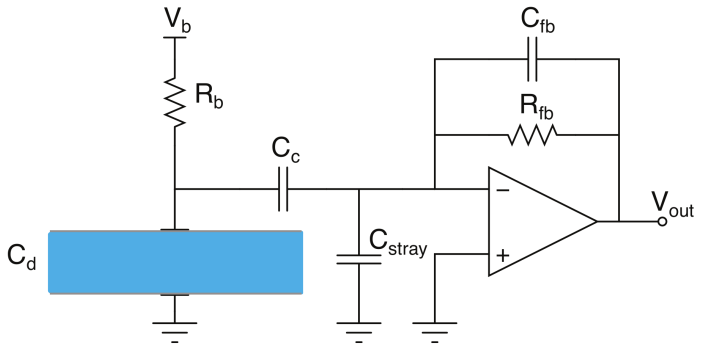


Figure 2.15: Transimpedance amplifier readout circuit for our iZIP's ionization collection lines [49].

As a result they will not induce the large scale drift field which typically aids neutralization, requiring an extended flashing procedure.

2.2.6 Ionization Readout

The small current signal produced by ionization collection on each electrode is converted to an amplified voltage signal via a dedicated negative-feedback transimpedance amplifier circuit as shown in figure 2.15. In Ge, the drift velocity $v_d \sim 30\ \frac{\text{mm}}{\mu\text{s}}$ [47], leading to charge drifting times $t_d \lesssim 0.8\ \mu$s. As it happens, our ADC has a digitization rate of 0.8 μs, making the rising edge of the ionization pulse unresolvable, and it appears completely vertical after digitization. As a result, the pulse shape is a single exponential with a fall time that is set by the feedback circuit itself. This is set by the feedback resister $R_{\text{fb}} = 40$ MΩ, and the parasitic feedback capacitance $C_{\text{fb}} \approx 1$ pF leading to a falltime of $\tau = R_{\text{fb}}C_{\text{fb}} \approx 40\ \mu$s. For a detailed study of the characteristic noise of this circuit please see [50]. From this, it is expected that given the 50 mK base temperature of SuperCDMS Soudan, the ionization readout circuit noise should be dominated by the JFET voltage noise ($\sim 0.5\ \frac{\text{nV}}{\sqrt{\text{Hz}}}$). For an example of the measured noise in the Soudan detectors see figure 2.16.

2.2.7 Ionization Event Reconstruction

As with the phonon signal, the next step is to reconstruct energy estimates from this ionization signal pulse. Unlike the phonon signal the charge signal is an instantaneous rise, followed by a single fixed exponential fall. This leaves only the

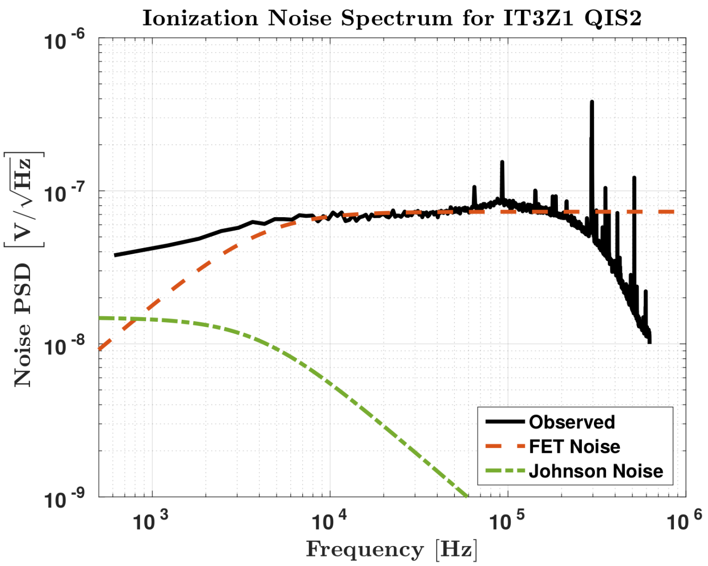


Figure 2.16: An example of an ionization noise power spectral density (PDS) for detector IT3Z1 during February 2013 with respect to the JFET gate. The FET noise (red) has been set at 0.5 $\frac{\mathrm{nV}}{\sqrt{\mathrm{Hz}}}$ to agree with the empirical spectrum (black) together with the known Johnson noise (green) [51].

amplitude to be fit. For this we turn again to Optimal Filtering (OF), with some slight complications.

Two-Template Optimal Filter (2x2 OF)

Unlike our phonon channels, the ionization channels are not independent; for a given detector face, the signal response of one channel can influence the signal response of the other. This arises from electronic cross-talk between the inner and outer electrodes. This can be modeled by a generalization of equation 2.4:

$$\begin{bmatrix} V_{\mathrm{t}}^{\mathrm{Inner}} \\ V_{\mathrm{t}}^{\mathrm{Outer}} \end{bmatrix} = \begin{bmatrix} S_{\mathrm{t}}^{\mathrm{Inner}} & S_{\mathrm{t}}^{\mathrm{C1}} \\ S_{\mathrm{t}}^{\mathrm{C2}} & S_{\mathrm{t}}^{\mathrm{Outer}} \end{bmatrix} \begin{bmatrix} A^{\mathrm{Inner}} \\ A^{\mathrm{Outer}} \end{bmatrix} + \begin{bmatrix} n_{\mathrm{t}}^{\mathrm{Inner}} \\ n_{\mathrm{t}}^{\mathrm{Outer}} \end{bmatrix} \tag{2.8}$$

Where the off-diagonal elements of the the signal template matrix S consist of two additional templates that model the cross talk. As with the one-template case, the A terms can be fit via χ^2 minimization in the frequency domain.

Ionization F5 Fitter (F5)

For completeness I should add that, as with the phonon channels, the charge signal can saturate for very high-energy interactions. This method simply fits any non-saturated portion of the pulse to an exponential with the known pulse-falltime.

2.3 Experimental Data

2.3.1 Data Overview

The exposure for this analysis (internally referred to as the "High Threshold" analysis or simply HT) can be divided into three large time periods. Run 133 (R133) which lasted from March 2012 until July 2013, R134 which was from July 2013 until July 2014, and R135 from September 2014 until November 2015. Several types of data were collected during the HT exposure period. WIMP-search data (WS), sometimes referred to as dark-matter search or DM-search data, is data that was collected in the absence of external calibration sources. ^{133}Ba calibration was performed frequently and provides a source of electron-recoil interactions. The peaks in the Ba spectrum allow for calibrating the detectors and the frequency of data collected allows for monitoring the stability of the detector response. A ^{252}Cf calibration source was deployed infrequently but at periodic intervals throughout the data period to provide a source of nuclear-recoil interactions. The majority of this data was collected during Runs 133 and 134[29] including the entirety of the WS exposure. R135 consisted of additional calibration data that was used in the HT analysis, but did not impact our overall WS-exposure. Since this WS set may contain actual WIMP events, a blinded version was used to develop the analysis (described in section 2.3.5 and 3.6.2). Approximately 70% of this time was used for DM-search data, while 10% was used for calibration and the remaining 20% was lost both to planned experimental downtime[30] and to periods of high detector noise. The total raw live time of these runs was 534 d. After removing data during which our detectors were not functioning normally (see section 3.3), we had a total exposure of 1690 kg days. This is summarized in figure 2.17

---

[29] With the exception of a 2 month period of very high cryocooler noise at the end of R134.

[30] Such as regularly occurring maintenance.

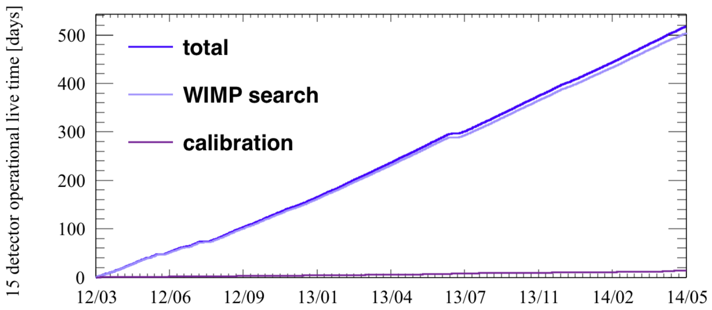


Figure 2.17: SuperCDMS Soudan operated from the end of 2011 to the start of 2015. During this period over 95% of the available time was used as WIMP-search exposure netting over 500 live-days.

2.3.2 Data Processing

The `cdmsbats` package is responsible for reconstructing the raw Soudan data into useful quantities to be used for analysis. The package is custom written in C++ and utilizes CERN's ROOT framework [52]. The inputs to `cdmsbats` consist of the raw data[31], monitoring information, templates for the phonon and charge pulses, and calibration constants. The latter two quantities are bootstrapped through analysis studies that utilize earlier versions of `cdmsbats`. For a much more detailed look at `cdmsbats` please see [53]. The output of `cdmsbats` are ROOT `nTuple`s containing blinded quantities that would be of interest to analyzers. These fall into two main categories. Reduced quantities (RQs) contain the uncalibrated quantities such as timestamps, estimated energies, and phonon pulse shape quantities. The calibrated quantities (RRQs[32]) contain the energies in units of keV as well as any estimate built from energy-based measurements such as an energy-partition-based position estimate. It is important to note that, due to the iterative nature of this process, the blinding criteria were set before the OF templates were finalized. As a result, the blinding only had access to INT-based RRQs. This has implications for the exact definition of our signal acceptance region as will be seen in section 3.6.

---

[31] *Raw* here means it is in a custom binary format used internally in the CDMS DAQ, and includes all digitized information including time-stream data.

[32] Although the source of this particular initialism is a topic of some debate among current SuperCDMS students, I have it on good authority that RRQ stands for "relational reduced quantities".

2.3.3 Data Acquisition

As seen in figure 2.17, the bulk of our data taking during the SuperCDMS Soudan runs was used to look for WIMPs and as a result most of the following data-acquisition discussion is geared toward WIMP search data. Calibration data required a slightly different operational mode for our detectors, and only these differences will be highlighted.

2.3.3.1 WIMP-search data

The bulk of our data was taken in this "WIMP-search" operational mode. We can break the DAQ into 4 steps:

- **Detector Neutralization** As detailed in section 2.2.5 space charge can build up from the initial cooling of our detectors as well as from normal operation. During WIMP-search data, this space-charge accumulation takes about three hours for our worst preforming detectors. To be neutralized, the ionization readout lines are grounded and our detectors are flashed with the surrounding LEDs. After flashing, we re-cool our detectors for 15 minutes.
- **Start-of-run randoms** Before taking any data the ionization readout lines (previously grounded) are biased to ±2 V. The QETs are also biased and self-heat into their superconducting transition. At this point 500 random event traces are taken in each of our detectors. These are used to study the noise in our detectors and are vital both for our optimal filtering algorithms as well as for analyzers to understand the noise environment of that particular run.
- **Data Taking** After the random events are acquired, the next three hours are spent taking DM-search data as well as a few intermittent random traces (at a rate about one tenth of particle interactions). After a global trigger is issued by the TLB, traces are read out of all detectors, with the triggering detector being specially tagged. This allows for our detectors to act as active shields for one another; coincident interactions are almost certainly due to backgrounds. Although it varies run to run, our WIMP-search data rate was typically less than 1 Hz. As this is the lowest event rate of any of our operational modes, and as a result has the lowest rate of background interactions,

we colloquially refer to this as "low background" data[33]. `cdmsbats` outputs two related datasets for this mode, one that is blinded and used through this analysis, and one that is complete and only used at the very end. These have the unfortunate names of `bg_permitted` and `bg_restricted` respectively.

- **End-of-Run randoms** After data acquisition has stopped we take 500 additional random traces. Originally, only the start-of-run randoms were acquired, but their close proximity in time to the biasing of our detectors made them susceptible to transient noise. As a result the end of run random traces were taken and are believed to be more representative of the noise behavior of the detector during the run. Typically another run is started immediately, so the detector would at this point be grounded and re-neutralized.

2.3.3.2 Calibration data

About 5% of experiment live time (spelled `LiveTime` in CAP) is taken using two ^{133}Ba sources. These are inserted partway through the veto-polyethylene-lead shielding surrounding the IceBox via small tubes near the e-stem and c-stem shield pass-throughs. This data is used throughout the analysis, from the energy calibration (section 2.3.4) to understanding and modeling γ-sourced backgrounds (section 4.3.3).

Nuclear recoils are calibrated using a ^{252}Cf source. This source is placed outside the IceBox shielding which is used to block most of emitted γ-rays. The main difference in data taking for calibration is that space charge buildup becomes a problem after only about 25 min. To keep our dead-time due to post-neutralization cooling to a minimum, the detectors are simply grounded for 5 minutes and calibration-induced events act to neutralize them. This grounding procedure is repeated 5 times during every run.

2.3.4 Energy calibration

The energy calibration of the ionization and phonon channels occurred in four stages. In the first stage, the working inner ionization channels were calibrated using γ lines from ^{133}Ba calibration data. Secondly, this inner-ionization calibration was then used to calibrate the outer channels. In the third stage, the relative

---

[33] This is somewhat confusing as almost all event in our "low background" data set will be background interactions. The term is used to emphasize the fact that no calibration source is present.

calibration of the phonon channels was determined by fitting the (ballistic) tails of the pulse for each channel. The last stage involved using the relative calibration to reprocess the data and compute the total phonon pulse quantities. These total phonon pulse quantities were calibrated by assuming that the ionization yield of bulk γ events has a mean of 1. The following subsections describe each of these stages but for even grater detail please see [54, 55, 56].

2.3.4.1 Ionization calibration

As they may have different voltage responsivity, all of the ionization channels should be calibrated independently. This was done for each side using hole collection, requiring that some ^{133}Ba calibration data be taken with the bias voltage reversed. This is done booth to take advantage of the smaller hole spot size, as well as to remove any systematic collection differences in the induced mirror current caused by the electron's oblique propagation behavior. First, the inner electrodes are calibrated by only examining events that induce no measurable response in the outer electrode, ensuring complete inner collection. There are a number of lines produced by ^{133}Ba but the 356 keV peak is particularly easy to measure and is the one primarily used in calibration. Because we can select the inner-collected events, calibration can be done by scaling the measured peaks to their known energy values. The outer electrode calibration is much harder. We cannot ensure a dataset that contains only well-collected outer events as these will typically experience some degree of sidewall trapping. Importantly, we are very interested in the class of events that experience both inner and outer collection (often called "shared" events), and these are what we utilize to preform the outer channel calibration. This is done by examining the distribution of ionization energy collected from the outer channel vs the inner channel. In this 2D plane all peaks are bands of over-density with a negative slope. This structure is exploited to map the known peaks of the inner channel to their analogues on the outer channel. For example, the 356 keV inner-collected peak that was previously calibrated can be found as an over-density on the x-axis. This over-density can be followed via the shared event population up to its counterpart on the y-axis as is seen in figure 2.18, calibrating the outer channel.

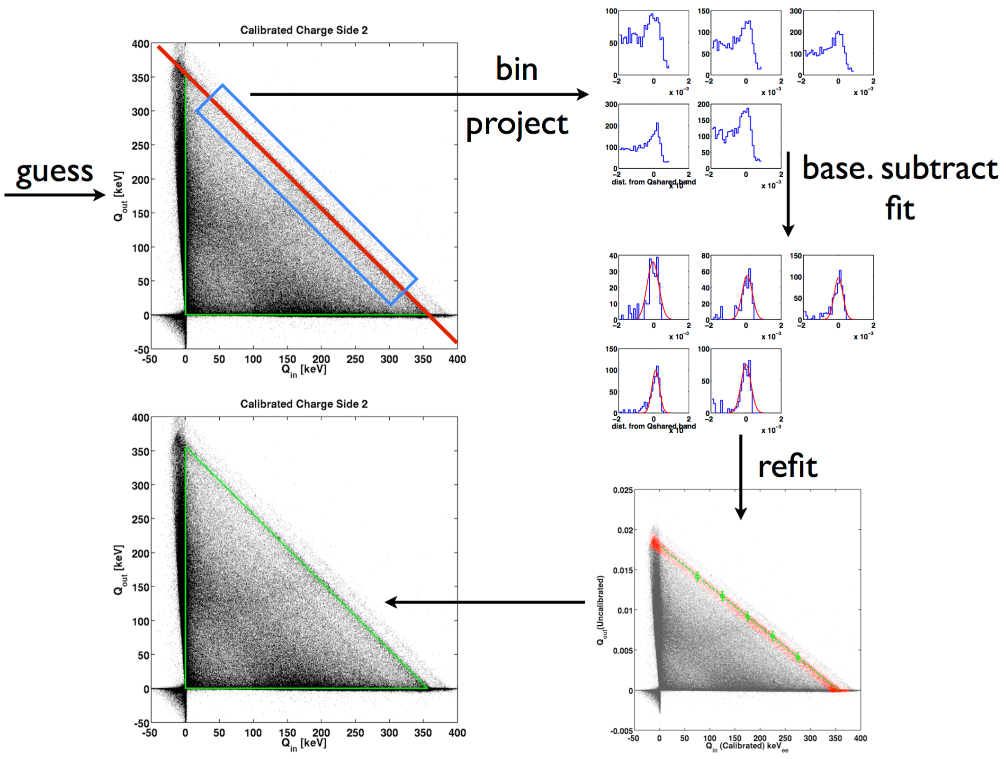


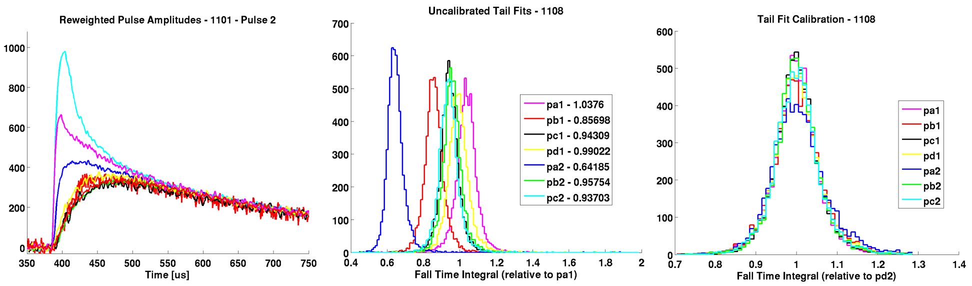
Figure 2.18: The outer ionization calibration utilized the previously calibrated inner-collected events. The over-density produced by the 356 KeV line (on the x-axis) can be fit to a linear regression via the shared events. In this way the outer-collected events (on the y-axis) can be fit and calibrated. Courtesy of A. Anderson.

2.3.4.2 Phonon calibration

The 8 phonon channels on a detector require two step calibration as well, but for very different reasons than the ionization channels. First, each phonon channel collects from an identical area of the crystal's surface. As a result, after 750 μs (i.e. uniformly distributed ballistic phonons) all phonon channels will have the same flux. Due to fabrication variance, however, not all phonon channels exhibit the same response amplitude even at this late time. To correct for this, these late-time tails are fit using the TFP method discussed in section 2.2.4. Each channel can then be assigned a correction factor to match its response to a reference[34] as seen in figure 2.19. After normalizing the relative channel response, the absolute energy calibration of the phonon collection can be done. First the total collected phonon energy is estimated from the summed trace. The ratio of the collected ionization

---

[34] The choice of which channel is the reference is arbitrary, but for R133 it was `PAS1` while for R134, due to shorts, it was changed to `PDS2`.



(a) Raw phonon pulses (rescaled).

(b) Relative integrated pulse amplitude.

(c) After relative phonon calibration.

Figure 2.19: Relative phonon calibration overview. (a) after a few hundred microseconds the relative responsiveness of our phonon detecting pixels should be receiving equal phonon flux. As can be seen in (b) this is not the case. The first stage in phonon calibration is to re-weight our phonon channels so the relative energy calibration is consist at late times (c) [57].

energy to this summed phonon energy (called "ionization yield"[35]) can then be constructed. In ionization yield, ^{133}Ba calibration events will form a Gaussian, and the absolute phonon calibration is defined by setting the mean of this distribution to unity.

2.3.5 Blinding

In order to avoid biasing the analysis, a loose signal region was defined and blinded. For an event to be blinded, it needs to pass several selection criteria that all suggest it has the potential to be signal-like, which I will briefly outline below. For the exact criteria, please see section 3.6.2

- **Single scatter** Since we expect WIMPs to interact through a nuclear recoil and scatter just once, an event must be a so-called "single-scatter" event to be blinded.

- **Nuclear-recoil band** As described in section 2.4.2, nuclear-recoil interactions have an ionization yield that is a an energy-dependent Gaussian. This is called the nuclear-recoil or NR-band mean and is distinct from the similar band caused by electron-recoil interactions. As we expect WIMPs to only interact via nuclear scattering, only events within 3σ of the NR band are considered for blinding.

---

[35] A quantity which is of great importance to us and is discussed in section 2.4.2

- **Reconstruction** All events that satisfy our hardware trigger conditions pass through the reconstruction algorithms presented in section 2.2.4, even if they are not actual particle interactions. Events of this type (such as low-frequency noise or electronics glitches) are often very poor matches for our event templates. Any events that do not reconstruct well will be removed from the analysis and are not blinded.
- **"Loose" Ionization Fiducial Volume** As discussed in section 2.4.4 it is possible to use the ionization signal to reconstruct the position of the interaction. As near-surface interactions are expected to be dominated by background events and thus removed in later analysis, events with highly asymmetric collection (near-face events), and events that are collected almost entirely in the outer electrode (near sidewall events) are not blinded.
- **Recoil Energy range** We also only consider blinding events with a particular range of energies, starting above the phonon noise threshold and extending to 150 keV recoil energy.

There are some periods of time when the blinding criterion was not applied to the WIMP search data set. These data sets are useful for background modeling since signal-like events are present in WIMP search data. Random triggered events are exempt from the blinding criterion since they are not eligible for inclusion in our final signal region. CDMSlite mode detectors are not included in the analysis and therefore also exempt from the blinding, along with detectors operating in reverse bias mode. The low-threshold analysis [31] has already unblinded data for several detectors for energies below 13.1 keV (defined against the total collected phonon energy `ptNF`) and we do not attempt to re-blind them. Since these open datasets may contain WIMP events, they are excluded from the HT analysis.

2.3.6 Naming and Identification

Although not pertinent to the science result, it is worth taking a brief moment to talk about how we have chosen to name various quantities of interest in CDMS, and specifically in the CDMS Analysis Package (CAP). In portions of this thesis these quantities will often be referred to using their CAP designation, so an overview is warranted.

- **Event Identification** To talk about a particular interaction event we need an unambiguous identification scheme. To do this, CAP uses the combination of two RQ's spelled `SeriesNumber` and `EventNumber`. As mentioned in section 2.3.3.1, events are acquired in three hour long runs called series. These series are identified using an 8 digit number and a 4 digit number separated by an underscore. The first two digits represent a location code and for all data taken at Soudan take the value of `01`. The next six digits represent the date the series started, and the final four represent the time as follows:

  $$\texttt{\{location code\}\{date in yymmdd\}\_\{hour\}\{minute\}} \tag{2.9}$$

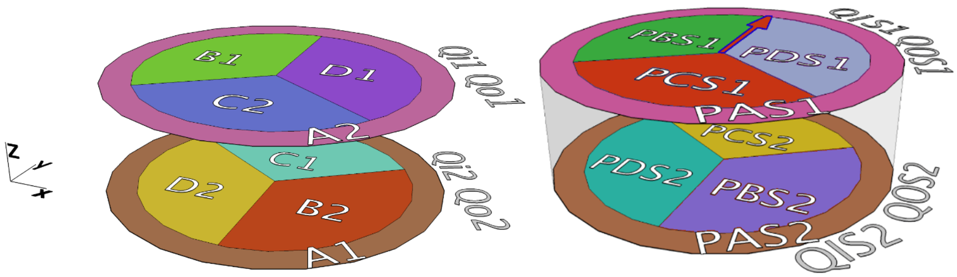
  During acquisition, data is written to files in 500 event chunks called "dumps". Within a given series, the i^{th} event in dump number d is referred to by the `EventNumber` where

  $$\texttt{EventNumber} = 10000d + i \tag{2.10}$$

- **Channel Names** As can be seen in figure 2.20, there are two distinct ways of referring to any particular channel. The "hardware name" uses the readout (or DIB) to identify the channel. These are written in capitals and take the form

  $$\{\texttt{channel type}\}\{\texttt{channel letter}\}\texttt{DIB}\{\texttt{DIB number}\}. \tag{2.11}$$

  The "analysis name" refers to the physical location of the sensor on the de-



(a) DIB channel names.

(b) Analysis channel names.

Figure 2.20: Analysis names (b), vs hardware or DIB names (a) for each phonon readout channel.

tector and takes the form

$$\texttt{\{channel type\}\{channel letter\}S\{side number\}}. \tag{2.12}$$

For example, the annular ring (channel letter `A`) phonon collector (channel type `P`) on side 1 is read out on DIB 2. The hardware name for this channel is `PADIB2` and the analysis name is `PAS1`.

- **RQ Names** For uncalibrated quantities (RQ's) the name scheme is fairly ad hoc, but they are identifiable by the fact that they start with capital letters. For easily described RQ's they are often just a camel case version of their description. An event's live time is `LiveTime` and the current base temperature is `BaseTemp`. Most quantities that are measured from a particular channel have the form {analysis channel name in caps}{under-case measured quantity}. The QET bias on the side one annular ring is `PAS1bias`.

- **RRQ Names** Calibrated quantities (RRQ's) are any quantity that requires reconstruction to an absolute energy measurement. These can be energy estimates themselves (which will have units of keV), as well as dimensionless quantities built from these estimates. They typically start with lower case letters and take the form (line-breaks inserted for ease of reading)

$$\begin{aligned}&\texttt{\{lowercase channel designation\}}\\&\texttt{\{quantity being estimated; default (nothing) is energy\}}\\&\texttt{\{reconstruction algorithm\}}\end{aligned} \tag{2.13}$$

  For example the phonon energy estimate made from summed traces and reconstructed using the NS filter is `ptNF`, while the partition-based radial position estimated from the ionization signal is `qrpartOF`.

- **Cut Names** The final kind of `CAP` name refer to a series of boolean masks that allow us to select for or against various subsets of our data. These are typically referred to as "cuts". These all start with a lowercase `c` and take the form

$$\texttt{c\{data selected for\}\_\{dataset qualifier\}} \tag{2.14}$$

  The dataset qualifier refers to the data that the cut is valid on. The most general designation refers to any cuts that are valid for all R133, R134 and R135 data and is `133`. An example is a cut that selects for electronic glitch-induced events `cGlitch_133`. There were two main versions of `cdmsbats` (`V52` and `V53`) used to produce datasets and certain cuts were specific to these production runs. The cut that selects events with poor optimal filter resolutions depends on `cdmsbats` version and, in the case of the analysis covered in this dissertation, is spelled `cBadOFRes_V53`. Additionally, there were cuts that were specific to the high threshold analysis effort. These had an additional `_HT` added after the normal dataset qualifier as a suffix.

To emphasize that these terms gain their semantics from their spelling rather than their phonetics they will be depicted in `monospace`. For a summarized table of various useful `CAP` quantities please see table 2.1.

2.4 iZIP Signal and Background Discrimination

As previously mentioned, iZIPs were designed with the goal of discriminating between signal and background, or more correctly between nuclear- and electron-recoil interactions on an event-by-event basis. Before we address exactly how this is accomplished, we need to discuss the fundamental quantities that we observe during particle interactions, and how exactly we build upon them. To describe a WIMP-like interaction the most important measurable quantity is the energy of the interaction. The interaction strength, potential WIMP mass, and astronomical distribution of WIMP dark matter all influence the recoil energy observed in a terrestrial WIMP scatter. As such, the energy domain of a particular direct DM detection experiment directly influences the mass range of potentially observable WIMPs for a given DM halo model. As discussed above, each event is read out using two distinct signal pathways, and it turns out these two signals are different for nuclear-recoil interactions than they are for electron-recoil interactions, allowing for the discrimination of signal from most backgrounds. In practice, we define a quantity called “ionization yield” which is the ratio between the two to convey this information. Ionization yield works very well at background discrimination provided that the total ionization and phonon signal can be accurately recorded. There are two cases in iZIP detectors where we run into difficulty with the ionization signal measurement. The first is at low recoil energies. Below recoil energies of ∼10 keV the ionization-yield distributions of nuclear- and electron-recoil interactions start to overlap. Second, if a particle interacts near the surface of a detector it is possible that some of the charge carriers may become trapped and fail to be collected. This effect reduces the ionization yield of interactions of these "surface events". To mitigate this effect we reconstruct a number of position estimators which convey the location of the particle interaction of our detector. I will go over all of these quantities in more detail below.

2.4.1 Recoil Energy

To calculate the recoil energy of a particle interaction, we start with something we actually measure: the total collected phonon energy, $E_{\text{total}}^{\text{Phonon}}$. These collected phonons come from three sources. First, there are the phonons directly produced in the initial recoil event, $E_{\text{recoil}}^{\text{Phonon}}$. These to rapidly scatter and down-convert in energy but the process is energy-conserving allowing most phonon energy to be collected and read out. As discussed earlier in section 2.2.3, by drifting electron-hole pairs across the detector we produce Luke phonons, $E_{\text{Luke}}^{\text{Phonon}}$, which also contribute to our total. Last, each charge carrier pair is initially imparted with potential (from the gap) as well as kinetic energy when they are initially created. This is eventually all turned into phonons as well. The kinetic energy is given up relatively quickly as the charge carriers relax to their drift velocities (near the conduction band ground state). when each charge carrier reaches a readout line it stops propagating and falls back to the Fermi level giving the reminder of its kinetic energy and all of its potential energy back as phonons. For bookkeeping purposes I will refer to all three of these phonon populations together as the relaxation energy, $E_{\text{relaxation}}^{\text{Phonon}}$. As an equation this looks like:

$$E_{\text{total}}^{\text{Phonon}} = E_{\text{recoil}}^{\text{Phonon}} + E_{\text{relaxation}}^{\text{Phonon}} + E_{\text{Luke}}^{\text{Phonon}} \tag{2.15}$$

This can be re-written as

$$E_{\text{total}}^{\text{Phonon}} = E_{\text{recoil}}^{\text{Phonon}} + N_{\text{eh}}\varepsilon + E_{\text{Luke}}^{\text{Phonon}} \tag{2.16}$$

where N_{eh} is the number of electron-hole pairs collected and ε is the average electron-hole creation energy. Assuming all charge carriers are collected this is the same as the the amount of initial recoil energy that produced electron-hole pairs in the first place, $E^{\text{Ionization}}_{\text{recoil}}$ which produces:

$$E^{\text{Phonon}}_{\text{total}} = (E^{\text{Phonon}}_{\text{recoil}} + E^{\text{Ionization}}_{\text{recoil}}) + E^{\text{Phonon}}_{\text{Luke}} = E^{\text{Recoil}} + E^{\text{Phonon}}_{\text{Luke}} \tag{2.17}$$

or

$$E^{\text{Recoil}} = E^{\text{Phonon}}_{\text{total}} - E^{\text{Phonon}}_{\text{Luke}} \tag{2.18}$$

where E^{Recoil} is the recoil energy we are interested in. We can now see that the recoil energy of a particle interaction in our detector is just the total collected phonon energy less the Luke phonon energy. For this to be useful, we need a way to measure the Luke phonon energy. As previously mentioned in section 2.2.3, Luke phonons are created as our charge carriers are drifted across the crystal, thus the total Luke phonon energy is set by the bias voltage, V_B, across which the electrons and holes drift or

$$E^{\text{Recoil}} = E^{\text{Phonon}}_{\text{total}} - N_{\text{eh}}|eV_B| \tag{2.19}$$

where e is the elementary charge. In CAP E^{Recoil} is spelled `precoiltNF` and subsequently will often be referred to this way.

2.4.2 Ionization Yield

We have touched on the concept of ionization yield a few times before, but as it is our most important way of differentiating signal from backgrounds we will review it in slightly more detail here. Ionization yield is defined as the ratio between the total ionization energy, $E^{\text{Ionization}}_{\text{total}}$, and the recoil energy of the interaction, E^{Recoil}, or

$$Y^{\text{Ionization}} = \frac{E^{\text{Ionization}}_{\text{total}}}{E^{\text{Recoil}}} \tag{2.20}$$

Remember from section 2.3.4.2 that the phonon signal calibration is defined by setting the ionization yield to of an electron-recoil interaction to unity. How does this compare to they ionization yield of a nuclear-recoil event? Qualitatively, electron recoils, which are caused by Compton scatters or charged-particle interactions, have long tracks through the detector, cross many lattice sites, and have a relatively low differential energy loss. As a result, a relatively large proportion of their energy is lost to electron-hole pair creation versus lattice vibrations. Nuclear-recoils on the other hand, which are caused by neutral particle interactions such as WIMP-scatter events, have a very short recoil length and a very high differential energy loss. As a result most of this energy is lost to lattice vibrations. Quantitatively, this can be measured by illuminating our detector with calibration sources designed to induce large populations of either electron- or nuclear-recoil interactions. For electron-recoil events we use a ^{133}Ba γ source, and for nuclear-recoil events a ^{252}Cf neutron source. As can be seen in figure 2.21, the ionization yield of nuclear-recoil type interactions is empirically about 25% of that of an electron-recoil type event, although there is a slight energy dependence. Lindhard [58] has developed the mostly widely used model for predicting the ionization yield of nuclear-recoils as a function of energy. The basic functional form is

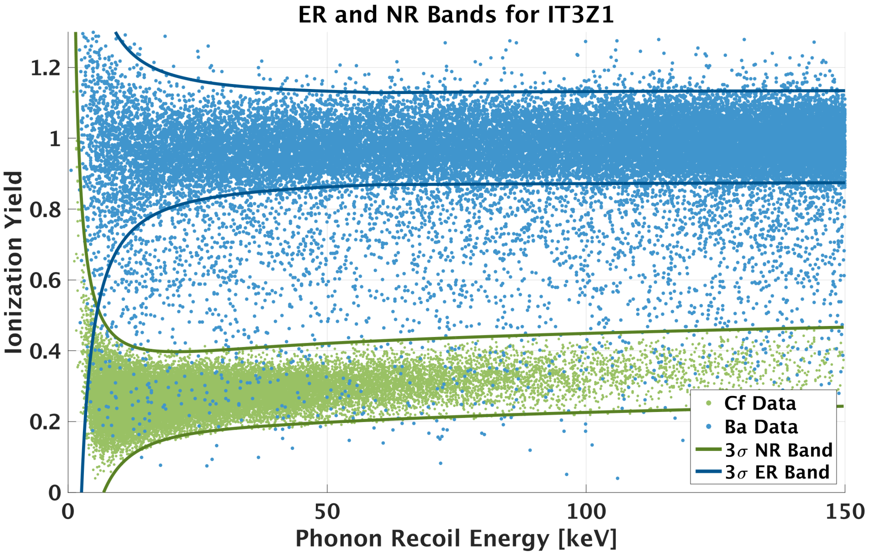


Figure 2.21: Plot of the ionization yield vs total recoil interaction energy for ^{252}Cf-sourced nuclear recoil events (green), and ^{133}Ba-sourced γ-ray interactions. This is the classic plane in which CDMS-style detectors can discriminate nuclear- from electron-recoil events. As can be seen, the electron-recoil events form a band (Generally Gaussian with a low-yield tail) centered around a yield of unity. The nuclear-recoil interactions form a distinct energy-dependent band that is ~25% of the ER band in ionization yield. The non-Gaussian tail of low-yield events arise from near-surface interactions that suffer from reduced ionization collection.

$$Y^{\text{Ionization}}(E) = \varepsilon \frac{\kappa g(\epsilon)}{1 + \kappa g(\epsilon)} \tag{2.21}$$

where

$$
\begin{aligned}
\kappa &= 0.113Z^{2/3}A^{1/2} \\
g(x) &= 3x^{0.15} + 0.7x^{0.6} + x \\
\epsilon &= 11.5Z^{-7/3}E
\end{aligned}
$$

and Z and A are the usual atomic number and atomic mass. For a Ge nucleus this model predicts ionization yield values in the 0.2 to 0.3 range for the 10-100 keV decade of interest for this analysis, and, at least in this energy range, is fairly consistent with our measured results. A comparison of this model to historical measurements can be seen in figure 2.22. Provided complete charge collection, ionization yield can differentiate between electron- and nuclear-recoil interactions. Using this technique, the probability of misidentifying an electron-recoil as a nuclear-recoil is 1 in 5.88×10^{-7} at recoil energies above 8 keV [59]. For our Soudan exposures and event rates, this is more than good enough to ensure a background free analysis, but obtaining this excellent discrimination sacrificed 70% of our iZIP detector volume to ensure that the complete-charge-collection condition was met. As an astute reader will notice, the electron-recoil events in figure 2.21 are not a simple tight Gaussian centered around 1. Instead there seems to be long tail of low-yield electron-recoil events have an increased probability of being misidentified as nuclear-recoil interactions. **These arise from near-surface interactions and removing them without sacrificing our exposure constitutes the bulk of the complexity of this analysis effort.**

2.4.3 Position Estimators

If we are to remove near-surface interactions we need some way to reconstruct an estimate of the interaction location in our detector. This is done by instrumenting our detector with multiple sensors in different locations and comparing the collected signals for a given interaction. In general, this can be done both by comparing signal amplitudes as well as signal timing. For our analysis, we utilize a number of different position estimates, but each boils down to comparing the partition of energy across various combinations of our different ionization and phonon channels, and each will be detailed in the following subsections.

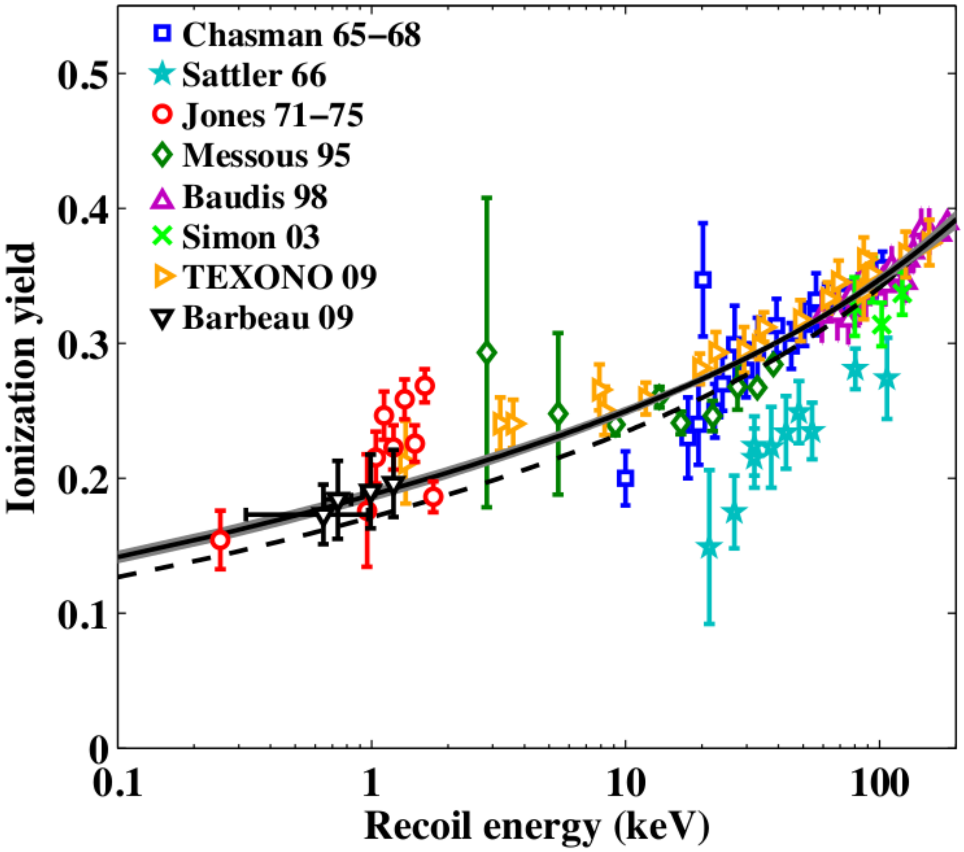


Figure 2.22: Previous empirical measurements of the ionization-yield energy dependence of a nuclear-recoil event (colored) plotted against a theoretical Lindhard model [41].

2.4.4 Ionization Based Position Estimation

As mentioned in section 2.2.5 our new iZIP detectors have a few improvements over the CDMS II oZIP-style detector. First, the ionization- and phonon-collecting lines are interdigitated, with the ionization lines biased at ±2 V and the phonon lines held at ground. This creates a uniform drift field in the bulk of the crystal and a highly tangential field near the face. The second major difference is that both faces are symmetrically instrumented.

Ionization Z position

In an oZIP style detector, interactions that happen within ~10 μm of a detector face run into a couple of problems. First, there is a layer of amorphous Ge near the face that has an entirely different band structure than the bulk of the crystalline Ge. It is possible to trap charge in this amorphous layer, leading to reduced collection. Second, the when they are initially produced, the charge carries are quite hot and until they relax down to near the conduction ground state their movement is dominated by thermal diffusion, rather than the drift field. As a result, these charge carriers can directly diffuse into the phonon collection circuits (which would produce no ionization signal) or even into the wrong ionization collector (potentially canceling actual signal). It is something akin to this "wrong side collection" behavior that iZIPs were specifically designed to exploit. In an iZIP, an interaction in the bulk will lead to symmetric collection of electrons and holes. An interaction within ~ 1 mm of either face, however, will encounter the transverse field as seen in figure 2.13. In this case charge carriers will not cross the crystal but instead be collected entirely on a single face[36]. As both faces of an iZIP are instrumented to collect charge, we can use this information to construct an ionization-based, Z-position estimate $Z^{\text{Ionization}}$ from ionization collection:

$$Z^{\text{Ionization}} = \frac{E^{\text{Ionization}}_{\text{electron}} - E^{\text{Ionization}}_{\text{hole}}}{E^{\text{Ionization}}_{\text{electron}} + E^{\text{Ionization}}_{\text{hole}}} \tag{2.22}$$

Where $E^{\text{Ionization}}_{\{\text{electron, hole}\}}$ is the energy estimate from the collection of electrons or holes respectively. This parameter varies from -1 (a hole-collection-side surface event interaction) to 1 (an electron-collection-side surface event interaction) with 0 being a bulk interaction. In practice, a uniformly illuminated detector will form three clustered distributions around these values of -1, 0, and 1 as seen in figure 2.23. In CAP, $Z^{\text{Ionization}}$ is spelled `pzpartOF`. As near-face surface event identification was a major design goal of our iZIP detectors, two ^{210}Pb sources were installed in situ to produce a controlled near-face surface event population for more detailed study. One was situated above the top face of detector iT3Z1 to produce electron-collection surface events while the other was situated below the bottom face or detector iT3Z3 for the creation of the equivalent hole-collection surface events. The first ~ 900 live-hours of our Soudan dataset were used to study our ability to identify these near-face surface events the results of which can be seen in figure 2.24. In an attempt to only study the effects of near-face surface events, as opposed to near-sidewall surface events, this study was restricted to interactions in which the ionization signal was completely collected by the inner electrode. This effectively reduced the detector mass to ~30% of its of its value. If we wish for the best possible sensitivity we need

---

[36] With the "wrong side" charge carriers being collected on the grounded electrodes and thus not sensed at all.

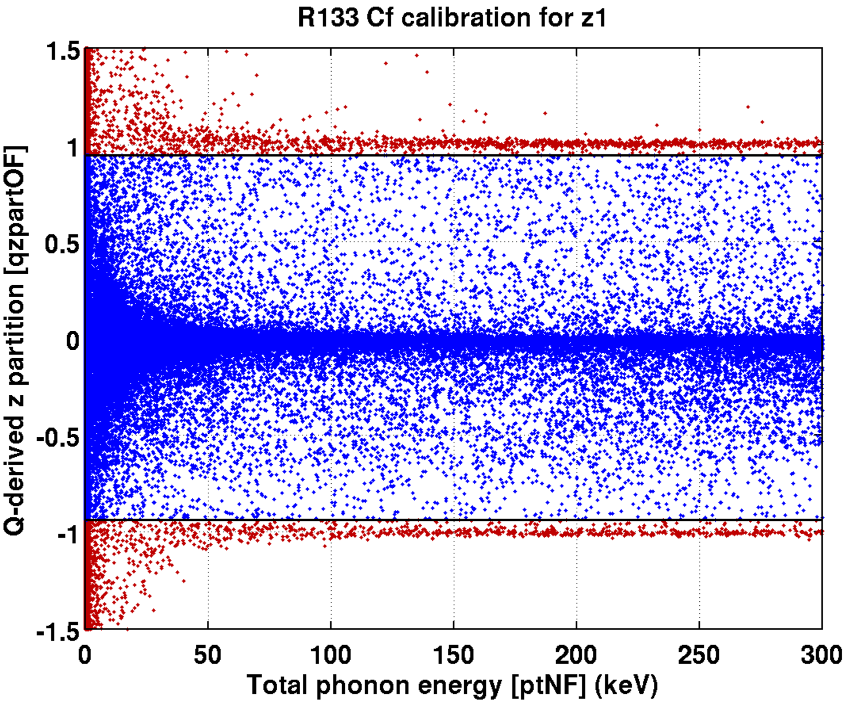


Figure 2.23: Plot of ionization z partition vs. recoil energy for a set of ^{252}Cf-sourced calibration events in detector IT1Z1. Example fiducializing cuts (which are very similar to those used in our "preselection region") are shown dividing surface events (red) from bulk events (blue).

to have a look at radial position in more detail as well.

Ionization radial position

As with near-face surface events, near-sidewall surface events may experience charge trapping if the drifted charge carriers encounter the sidewall. Unlike with near-face surface events, however, this phenomena may extend to interactions happening may millimeters away from any surface of the detector. This is especially pronounced when examining the ionization signal from electron collection. Due to the oblique propagation discussed in section 2.2.5, events occurring deep in the detector and far away from the electron-collecting face are epically susceptible. This effect may be mitigated by calculating the total number of charge-carriers created and the associated ionization energy estimate, $E^{\text{Ionization}}$, by simply calculating the ionization energy separately for the electron and hole collecting sensors and taking the maximum as was seen in equation 2.7. In CAP, this representation of $E^{\text{Ionization}}$ is spelled `qsummaxOF`. This estimate has the benefit of automatically discarding an ionization signal that is under-collected. This is helpful for interactions occurring near each of the faces, but runs into difficulty for high-radius events, where both electron and holes could be trapped. It is useful to then define an R-position estimate, $R^{\text{Ionization}}_{\{\text{electron,hole}\}}$

$$R^{\text{Ionization}}_{\{\text{electron,hole}\}} = \frac{E^{\text{Ionization Outer}}_{\{\text{electron,hole}\}}}{E^{\text{Ionization}}_{\{\text{electron,hole}\}}} \tag{2.23}$$

where $E^{\text{Ionization Outer}}_{\{\text{electron,hole}\}}$ is the ionization energy collected in the annular guard-ring electrode. The ionization R position estimate varies from 0 for an interaction in the bulk of the detector, to 1 for a high-radius event as can be seen in figure 2.25. Unlike with our ionization Z position estimate, there are two basic estimates to choose from, one from the electron signal and one from the holes. Each of these two estimates will be more or less reliable depending on the vertical interaction location. From a very abstract standpoint, this is not a problem. Any multidimensional machine-learning technique should be able learn which of these position estimates is more reliable based on all other available information. In practice, however, a single reliable radial position is vital for the exploratory phase of an analysis of this type. If human analyzers are going to want to remove the obviously very-high-radius events by hand, or better yet preform any sort of cross-check on the final analysis result, it is important that we have a good physical understanding of the variable we are cutting on.

A naive approach to construct a combined ionization radial position estimate would be to follow in the footsteps of our $E^{\text{Ionization}}$ definition and simply rely on the ionization radial position estimate from the face that collected more energy as in:

$$R^{\text{Ionization}} = \begin{cases} R^{\text{Ionization}}_{\text{electron}}, & Z^{\text{Ionization}} \geq 0 \\ R^{\text{Ionization}}_{\text{hole}}, & Z^{\text{Ionization}} < 0 \end{cases} \tag{2.24}$$

The problem with this definition is that for the majority of the interior volume of the detector, the charge collection on both faces will be essentially symmetric, but the value of $Z^{\text{Ionization}}$ will fluctuate slightly negative or positive due to noise. This will lead to potentially throwing away the more accurate radial estimator. An improvement would be to use pure electron or hole information only for events that are truly single-side collection, and a linear combination both electron and hole information for all others.

$$R^{\text{Ionization}} = \begin{cases} R^{\text{Ionization}}_{\text{electron}}, & Z^{\text{Ionization}} > 1 \\ R^{\text{Ionization}}_{\text{hole}}, & Z^{\text{Ionization}} < -1 \\ \frac{1+Z^{\text{Ionization}}}{2} R^{\text{Ionization}}_{\text{electron}} + \frac{1-Z^{\text{Ionization}}}{2} R^{\text{Ionization}}_{\text{hole}}, & \text{otherwise} \end{cases} \tag{2.25}$$

While this is an improvement, it makes the mistake of treating the electron and hole collection symmetrically. The definition that was settled on used the more tightly collimated hole collection side for every event where more holes are collected than electrons, and only utilizes this linear combination for events with an excess of electrons.

$$R^{\text{Ionization}} = \begin{cases} R^{\text{Ionization}}_{\text{electron}}, & Z^{\text{Ionization}} > 1 \\ R^{\text{Ionization}}_{\text{hole}}, & Z^{\text{Ionization}} < 0 \\ Z^{\text{Ionization}} R^{\text{Ionization}}_{\text{electron}} + (1 - Z^{\text{Ionization}}) R^{\text{Ionization}}_{\text{hole}}, & \text{otherwise} \end{cases} \tag{2.26}$$

As a final check to ensure a sensible, physically well motivated definition for radial position we turned to our on-going Detector Monte-Carlo (DMC) effort. Although not considered "production ready" at the time of this analysis, most of the known DMC discrepancies arise from difficulties simulating the phonon rather than ionization signal path, which is the more mature of the two. I have made a number of plots in figure 2.26 depicting how each of the above positions estimates maps to real physical interaction locations. In CAP $R^{\text{ionization}}$ is spelled `qrpartOF_zhalf`.

2.4.5 Phonon-based Position Estimates

Due to their relative positioning, as well as the physics of charge transport in our detectors, the ionization-based energy estimates from the previous section are fairly good at separating near-surface events from bulk events, but preform very poorly in three important contexts. First, estimating the relative positioning of bulk events is almost impossible using a measurement such as `qzpartOF`. All bulk events should have `qzpartOF` of identically zero. The width of the bulk-`qzpartOF`-distribution (the one centered around zero in figure 2.23) is almost entirely due to noise in the ionization collection measurement and does not accurately convey anything about the actual Z-position of the event. As a result, if we want to decide if a bulk event occurred in the top or bottom half of the detector we would need to turn to phonons.

Second, the ionization collectors are azimuthally symmetric; if we want to know the angular position of an interaction we need to use our phonon position estimates. Last, for very low-energy interactions the higher relative noise on the ionization channel will render phonon-based measures more accurate. The actual implementation of these phonon position estimates are based on simple partitioning of energy. They can be summarized as follows:

- **Phonon Z Partition**

$$Z^{\text{Phonon}} = \frac{E^{\text{Phonon}}_{\text{Side1}} - E^{\text{Phonon}}_{\text{Side2}}}{E^{\text{Phonon}}_{\text{Side1}} + E^{\text{Phonon}}_{\text{Side2}}} \tag{2.27}$$

  Where $E^{\text{Phonon}}_{\text{Side1}}$ is the OF-reconstructed energy estimate from all phonon channels on side 1.

- **Phonon R Partition**

$$R^{\text{Phonon}} = \frac{E^{\text{Phonon}}_{\text{Outer, Side1}} + E^{\text{Phonon}}_{\text{Outer,Side2}}}{E^{\text{Phonon}}_{\text{Total}}} \tag{2.28}$$

  Where $E^{\text{Phonon}}_{\text{Outer, Side\{1, 2\}}}$ and $E^{\text{Phonon}}_{\text{Total}}$ the OF-reconstructed energy estimates from the outer ring phonon channel and all 8 channels respectively.

- **Phonon X Partition**

$$X^{\text{Phonon}}_{\text{Side1}} = \frac{E^{\text{Phonon}}_{\text{D1}}\cos(\pi/6) + E^{\text{Phonon}}_{\text{B1}}\cos(5\pi/6) + E^{\text{Phonon}}_{\text{C1}}\cos(3\pi/2)}{E^{\text{Phonon}}_{\text{D1}} + E^{\text{Phonon}}_{\text{B1}} + E^{\text{Phonon}}_{\text{C1}}} \tag{2.29}$$

$$X^{\text{Phonon}}_{\text{Side2}} = \frac{E^{\text{Phonon}}_{\text{D1}}\cos(7\pi/6) + E^{\text{Phonon}}_{\text{B1}}\cos(11\pi/6) + E^{\text{Phonon}}_{\text{C1}}\cos(3\pi/2)}{E^{\text{Phonon}}_{\text{D1}} + E^{\text{Phonon}}_{\text{B1}} + E^{\text{Phonon}}_{\text{C1}}} \tag{2.30}$$

- **Phonon Y Partition**

$$Y^{\text{Phonon}}_{\text{Side1}} = \frac{E^{\text{Phonon}}_{\text{D1}}\sin(\pi/6) + E^{\text{Phonon}}_{\text{B1}}\sin(5\pi/6) + E^{\text{Phonon}}_{\text{C1}}\sin(3\pi/2)}{E^{\text{Phonon}}_{\text{D1}} + E^{\text{Phonon}}_{\text{B1}} + E^{\text{Phonon}}_{\text{C1}}} \tag{2.31}$$

$$Y^{\text{Phonon}}_{\text{Side2}} = \frac{E^{\text{Phonon}}_{\text{D1}}\sin(7\pi/6) + E^{\text{Phonon}}_{\text{B1}}\sin(11\pi/6) + E^{\text{Phonon}}_{\text{C1}}\sin(3\pi/2)}{E^{\text{Phonon}}_{\text{D1}} + E^{\text{Phonon}}_{\text{B1}} + E^{\text{Phonon}}_{\text{C1}}} \tag{2.32}$$

- **Phonon θ Partition** For each side we can define an angular partition as follows:

$$\theta^{\text{Phonon}} = \arctan 2(Y^{\text{Phonon}},\ X^{\text{Phonon}}) \tag{2.33}$$

where arctan 2 is the multi-valued inverse-tangent function.

2.4.6 Reconstructed Quantities of Interest Reference

Table 2.1: Table of reconstructed quantities of interest

Quantity	Description
pa1OF, pb1OF, pc1OF, pd1OF	side 1 OF phonon energies
pa2OF, pb2OF, pc2OF, pd2OF	side 2 OF phonon energies
psumOF	sum of all phonon quantities above
ptOF	OF phonon energy of summed traces
ptNF	non-stationary OF phonon energy of summer traces
prpartOF	(pa1OF + pa2OF)/ptOF
psum1OF	sum of all p{a,b,c,d}1OF quantities
psum2OF	sum of all p{a,b,c,d}2OF quantities
pzpartOF	(psum1OF - psum2OF)/ptOF
qi1OF, qo1OF	side 1 OF charge energies (inner, outer)
qi2OF, qo2OF	side 1 OF charge energies (inner, outer)
qsummaxOF	maxqi1OF + qo1OF, qi2OF + qo2OF

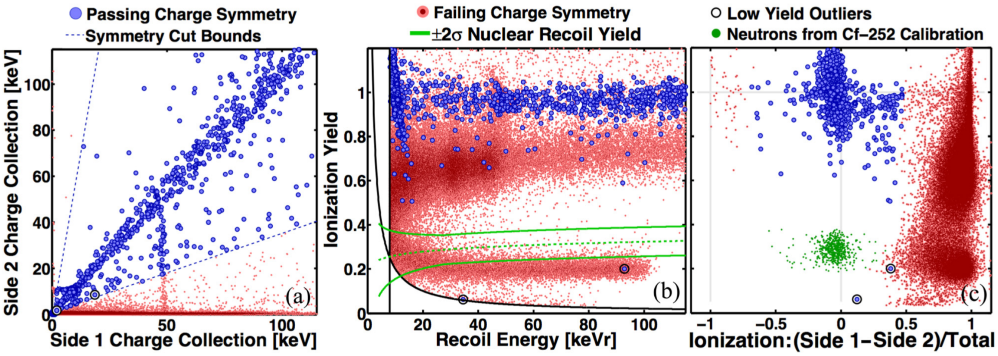


Figure 2.24: Plots showing the same unblinded ∼900 live hours from our detector with a ^{210}Pb-source implanted on the top face (IT3Z1). Events that pass (blue) and fail (red) the ionization symmetry cut are depicted, and refer to bulk and near-face interactions respectively. [surface rejection paper]. (a) The ionization symmetry cuts are shown (blue dotted lines) in the side 1 vs side 2 ionization collection space. The vertical over-density of points is from the 46.5 keV γ line from the ^{210}Pb source. (b) Classic ionization yield vs recoil energy discrimination plot. The 2σ NR-band is shown (green). Bulk electron recoil events (blue) are largely confined to the ER-band centered at a yield of 1 while the low-yield surface events (red) that would, without this symmetry cut, contaminate the NR-band are shown (red). (c) Plot of ionization yield vs ionization z-partition. In addition to the events in (a) and (b), there are ^{252}Cf-sourced nuclear-recoil events plotted (green). As can be seen these are characterized by being both in the bulk, and NR band and are a good proxy for our expected DM-sourced interactions.

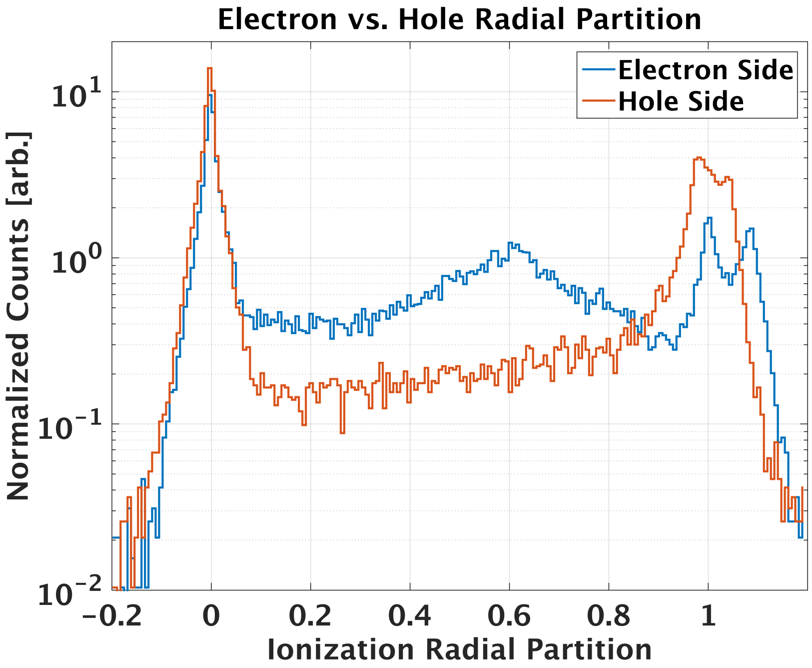


Figure 2.25: A histogram of the electron (blue) and hole (orange) ionization radial partitions from ^{133}Ba-sourced interactions. The peak centered on 0 consists of complete inner-collected events that can be safely assumed to occur in the bulk of the crystal. The peak at 1 consists of outer-collected events, that are generally considered to be of poor quality due to the potential of sidewall trapping. Due to oblique propagation of electrons these two populations are resolved more cleanly by the hole-collection data [51].

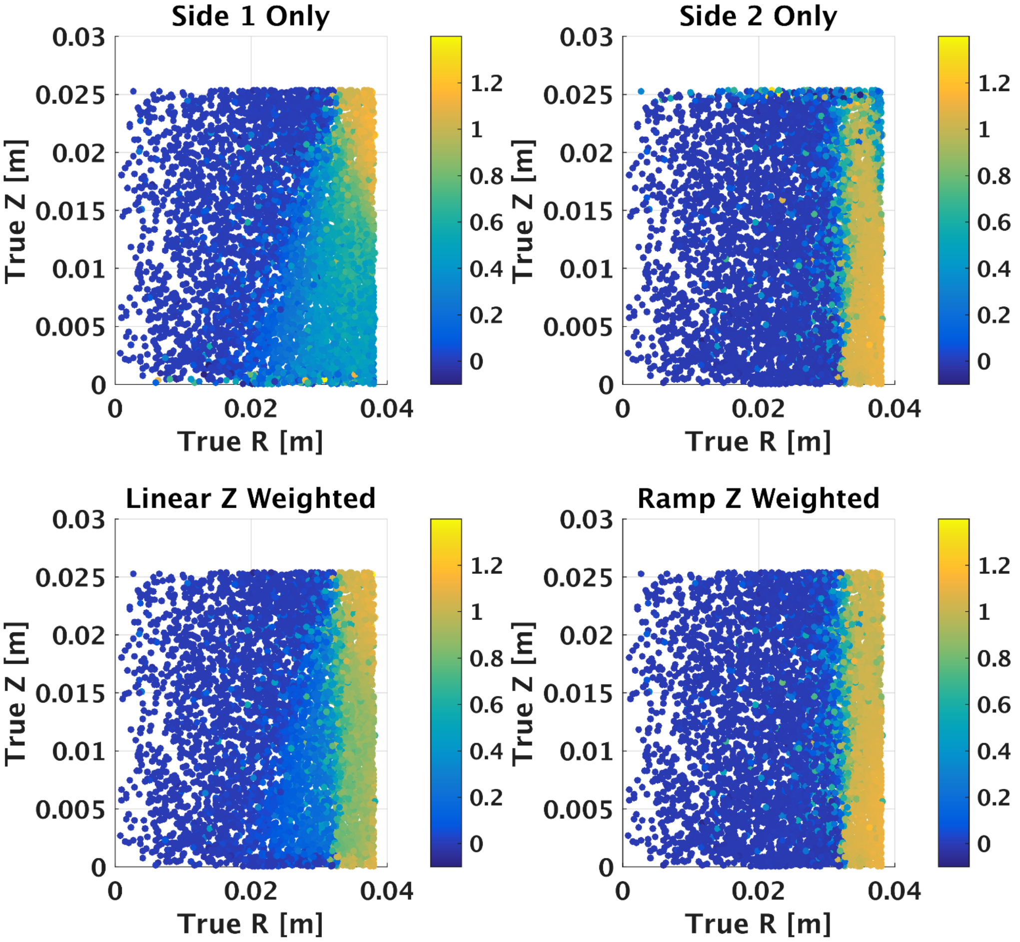


Figure 2.26: Comparison of the relative performance of various ionization-based radial position estimates. The x- and y-axis are the “true” particle interaction location (m) from the detector Monte Carlo (DMC), where (0, 0) is the center of bottom detector face. The estimate is depicted via the colormap. Clockwise from the top left:

(Side 1 Only) The estimate from the electron-collecting side (`qrpart1OF`).

(Side 2 Only) The corresponding estimate from the hole-collecting face (`qrpart2OF`).

(Ramp Z Weighted) The estimate that uses hole-collection for any event that collects more ionization energy on face 2, but a linear combination of electron and hole data (`qrpart_zhalf`).

(Linear Z Weighted) A simple linear combination of each side.

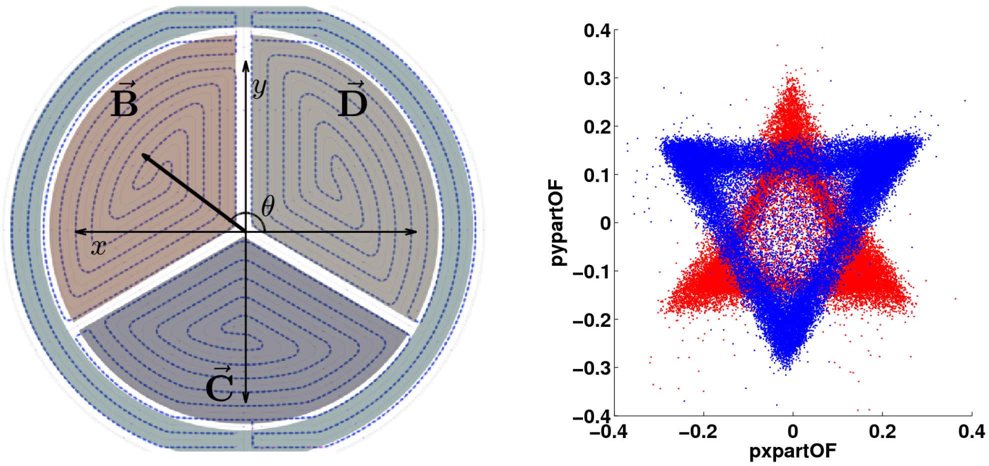


Figure 2.27: Due to the three wedge shaped phonon-collectors, the phonon signals can be used to reconstruct an X and Y position proxy based on energy partition. Left: the partitions are oriented assuming that the `D`, `B` and `C` collectors are aligned along the $\theta = 30°$, 150° and 270° directions respectively. Right: ^{133}Ba calibration data plotted in the phonon Y vs X partition space. Assuming the relative phonon calibration is correct, the population of events on each side should form an equilateral triangle.

Chapter 3

HIGH-THRESHOLD ANALYSIS: DATA SELECTION

3.1 Overview and science goals

The overarching goal of this analysis is to define a signal region that gives us the greatest sensitivity for detecting WIMPs of mass >10 GeV. In this regime the sensitivity scales with the exposure $Ex = MT$ where M is the target mass and T is the total experimental live time. This linear scaling is only accomplished in the absence of backgrounds in our signal region. If a background subtraction was preformed, the sensitivity scaling would reduce to $\sqrt{MT}$. Due to the blinded nature of this analysis, however, after the final signal-acceptance region is set, all observed events will be treated as signal and no background subtraction will be preformed. This is due to the uncertainty in modeling our backgrounds in our blinded data, and as a result as soon as we accept backgrounds into our signal region our sensitivity scaling will stop[1]. This informs an outline of the remainder of our analysis effort. We need to keep as much of our blinded data as we possibly can (exposure), but not at the expense of introducing *any* background events. For this reason, we often internally refer to this style of analysis as “zero-background”. To do this, we implemented a number of selection criteria referred to as “cuts”. To be considered as a WIMP candidate (i.e. inside of our signal region) an event must pass each of these cuts. They can be grouped into 5 categories:

- **Time-period Quality cuts** These cuts remove entire blocks of data and for the most part decrease the overall `LiveTime` of our experiment[2] of our experiment. The removal criteria for these cuts is time-based and typically removes series of events based on inferior operation conditions, such as high environmental noise.
- **Event-Based Quality cuts** During otherwise ideal operating conditions, events can still be of low-quality. Sporadic low-frequency noise or electronics glitches

---

[1] This is assuming that the rate of these misidentified background events scales with the exposure.

[2] The cut to remove randomly triggered events is an exception to this.

have the potential to satisfy our hardware trigger conditions, and will be (often very poorly) reconstructed as an event by `cdmsbats`. Seeing as all of the experimental live time surrounding this event is still valid, it would unnecessarily reduce our exposure if we removed it. Instead of simple `LiveTime` accounting this can be dealt with using a signal efficiency curve, which estimates that fraction of WIMP signal that would make it through the analysis as a function of energy.

- **Physics Cuts** Now that all of our events correspond to well-reconstructed particle interactions, we next remove any event that is physically incompatible with a WIMP-scatter interaction, or has a very high chance of being a background event. It is an important caveat that these events are not of any further interest to the analysis. As is discussed in section 2.4.2, an event with an ionization yield of 1 is almost certainly a background event, but these events are of remaining interest to the analysis and will be retained.
- **Preselection Cuts** All of the cuts above remove events from further consideration in this analysis. These preselection cuts serve to define a very loose signal acceptance region, but all the data surrounding this preselected region is retained. These surrounding regions are referred to as “sidebands”. As an example, to be in our preselection region an event must have an energy-dependent ionization yield consistent with nuclear-recoil interactions. Data that falls in this region is said to be inside of the nuclear-recoil band. Data that falls outside this region is said to be in the nuclear-recoil sideband. This is the region that use to build our background and signal models (see section 4.3).
- **Signal Region Definition** The remainder of this chapter will be devoted to describing the above cuts in more detail. The final cut that defines our signal region was quite complex to set and will be described in chapter 4.

The bulk of this chapter (sections 3.3-3.6) will be devoted to describing the first 4 types of cuts from above. Before we get into a discussion of our cuts, however, we first will examine which of the 15 iZIP detectors were included in the analysis.

3.2 Detector selection

Ideally we would use the entire detector array to maximize the exposure our our experiment. A number of our readout, and detector biasing channels developed shorts during the course of our 3 year run. Although many detectors experienced this with their phonon channels, ionization channel shorts proved much more problematic. Four our ionization channels, if a readout line was shorted, we would loose all ability to preform charge-based fiducialization, and as a result we removed these detectors. For detectors that experience shorting of the ionization bias lines the prognosis is actually worse; incorrectly biased ionization lines will drastically alter the drift field in the detectors, rendering most ionization-based measurements useless. Any detector with ionization bias shorts was also removed from the analysis. Although the causes of these shorts is still being looked into, all ionization shorts developed on the top surface[3]. This seems to point to conductive debris (such as small copper shavings from tapped screw holes) settling on the surfaces of the detector during detector mounting and installation. I have summarized the channel problems in table 3.1.

3.3 Time Period Based Quality Cuts

This first set of cuts are identifiable in that they remove entire series or time blocks, not just single events. These can affect experimental live time if applied to periods of active data-taking, and have an impact on our exposure. In general, they remove data where the experiment is in an ill-defined or unstable state, or where there is a known diagnostic indicating otherwise problematic operation. For an example of the effects of these cuts on the experimental live time on detector 1111 see figure 3.1.

Analysis Data Period (`cDataPeriod_V53_HT`)

During the three runs that make up the dataset for this analysis, there were periods where detectors were used in alternative operational modes (CDMSlite), or was focused on specific studies of our detector behavior (surface rejection paper). In addition there were periods where there were knows issues with the entire array. These were removed from further consideration. For a more detail see Table 5.4 in [33].

---

[3] The physical top of the detectors is the electron collecting side or side 1.

Table 3.1: Detector selection. Each detector that we have chosen to keep in this analysis is highlighted in green and the removed detectors highlighted in red.

Detector	Description
IT1Z1	PAS1 (and PDS2 to a lesser extent) were noisy for the entire SuperCDMS Soudan dataset. Use a 7 channel summed energy estimate for blinding.
IT1Z2	QOS1 bias line short for entire set. FET disabled on September 8, 2014. QIS1 feedback line short started right after R133 started. FET disabled June 27, 2012. PAS1 high-resistance short, PCS2 low-resistance short both on December 12, 2012. PCS2 grounded.
IT2Z1	Healthy!
IT2Z2	Healthy!
IT2Z3	PBS1 low-resistance, and PDS1 high-resistance short for entire dataset (disabled).
IT3Z1	Healthy! (although high LF noise)
IT3Z2	Healthy!
IT3Z3	PCS1 periodically noisy baseline (“square pulse”). 7 channel summed energy threshold for blinding.
IT4Z1	QIS1 feedback-line short March 28, 2012. FET grounded June 14, 2012 high noise causing temperature fluctuations.
IT4Z2	Healthy!
IT4Z3	Healthy!
IT5Z1	QIS1 and QOS1 feedback-line short from March 2, and April 25, 2012 respectively. Both FETs disabled on May 22, 2012. PAS1 low-resistance short May 22, 2012.
IT5Z2	CDMSlite detector. High-frequency spikes in QOS1 readout.
IT5Z3	QOS1 feedback-line short for entire experiment.

Flagged Series (`cBadSeries_133`)

This is the cut that removes series if flagged as bad by the on-site shift. It is about 99% efficient. It also removes series during times of high environmental noise.

CDMSlite Cut (`cCMDSlite_V53`)

During the course of our 3 runs, specific detectors were used in a slightly different operational mode for specialized low-threshold event detection called CDMSlite. Any detector that was flagged as being in this operational mode was explicitly re-

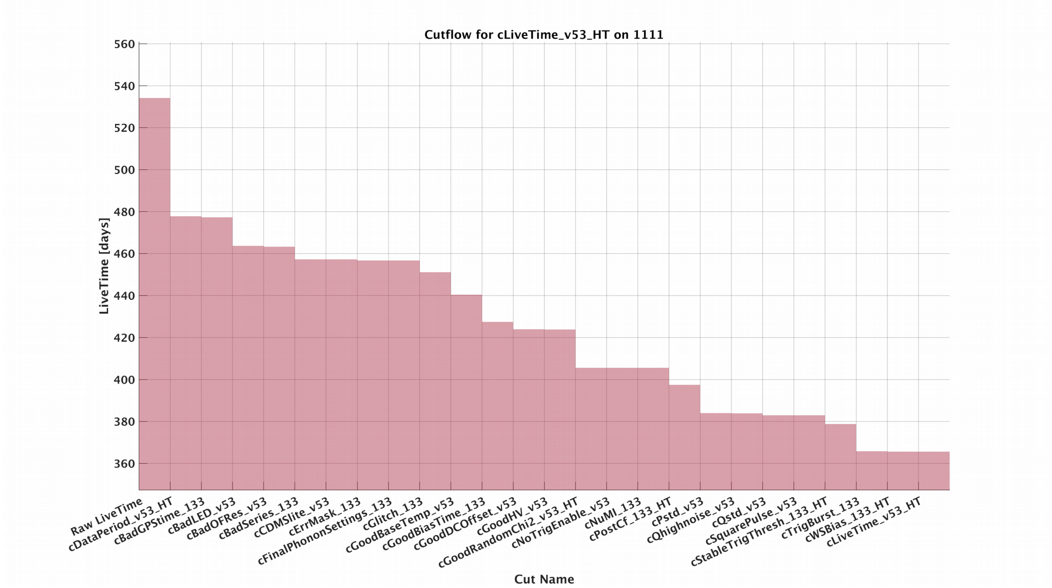


Figure 3.1: The loss in `LiveTime` from the application of our time period based quality cuts for detector IT4Z2 (1111). Aside from the Data Period cut there is no stand-out culprit for the loss of total exposure from these cuts. Rather, the sheer number of time-period based data quality issues that arose during the SuperCDMS Soudan dataset was responsible for the large cumulative effect [60].

moved.

Optimal Filter Resolution (~`cBadOFRes_V53`)

Aside from fitting our signal amplitude, the OF reconstruction procedure outputs a few diagnostics including resolution. This can be high if the noise template is large in amplitude for a given series. If the either the sum or the outer channel are 25% above the mean resolution for the given time block the series is cut.

Final Phonon Settings (`cFinalPhononSettings_133`)

During the first few weeks of operation the final bias voltages of some phonon channels had not been properly set. As this directly impacts the relative calibration, this period was removed.

Trigger Bursts (~`cTrigBurst_133`)

Transient noise can cause brief periods of very high trigger rates. The lifetime of these periods are very low and they are discarded. They are identified by removing events occurring during an `EventTime` that has a trigger rate that is either grater than 3 Hz or 8 times the median trigger rate for the series. Removes 5% of live time, and is shown in figure 3.2.

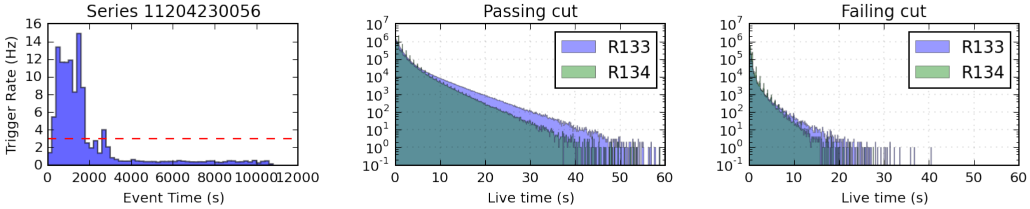


Figure 3.2: Trigger Burst cut. Left: events that occur at `EventTime`'s with high trigger activity (>3 Hz) are removed. As can be seen by the `LiveTime` distributions of the events passing (center) and failing (right) the cut this selection preferentially removes events with very short `LiveTime`. The peaks in this distribution are due to cryocooler-induced noise [61].

Crystal Neutralization (`cGoodBiasTime_133`)

Every 180 minutes we flash our crystals with an LED to get rid of trapped charges (see section 2.2.5 for more on this). Occasionally, there were times where we choose to turn this off, or the DAQ messed up and skipped a flash. We remove any such series.

Base Temperature (`cGoodBaseTemp_v53`)

Phonon gain is temperature dependent. If a detector is too much warmer than the temperature at which the phonon calibration was done, the series is removed. The maximum acceptable base temperature is different for each detector. It is determined to be the temperature at which the yield has increase by 5% relative to its value at 51 mK.

Trigger Timing (`cErrMask_133`)

The trigger for an event should be issued at time $t_0 = 0$ in the raw pulse. This is not always the case, and these events are removed.

External Timing (`cBadGPStime_133`)

Global experimental time for this analysis was recorded using GPS time. This was used to synchronize our experiment to external events (such as the NuMI beam). There were periods of the run where the GPS time was not properly recorded, and as a result if local time (EventTime) is different from GPS time by more than 17s the events are removed.

NuMI Beam (`cNuMI_133`)

SuperCDMS shares the SUL with the MINOS experiment [62]. The NuMI beam [63] produced at Fermilab and incident on MINOS, also passes through our detector array. Although it is unlikely we would detect an interaction from this, to be on the safe side we remove time periods in which the NuMI beam is active (within $\sim 60\,\mu$s).

Phonon sensitivity / DC Offset (`cGoodDCOffset_53`)

SQUIDS with trapped flux will result in a reduction in phonon response sensitivity. This condition is associated with an abnormal DC voltage offset in the phonon channel, and that is used as a proxy to cut these time periods.

Bad LED (`cBadLED_v53`)

In some of our detectors, using certain combinations of LEDs for detector neutralization can cause high ionization noise. During the initial period of run 133, before a good flashing scheme was developed, there are periods of time that fall victim to this increased noise. They are removed.

Stable Trigger threshold (`cStableTrigThresh_133_HT`)

To maintain high live time, the hardware trigger thresholds were changed many times during the course of the experiment. If the time period between such adjustments was too short, meaningful statistics about the detector's operation can't be gathered, and the time period is cut.

Random Trigger χ^2 (`cGoodRandomChi2_v53_HT`)

During the setting of the chi square cuts, it was noticed that there were sets of series with anomalous χ^2 distributions. There appeared to be an offset in the 0 energy distribution (see figure 3.3). It was eventually determined that this was caused by using contaminated noise templates in the processing. Originally these noise templates were built using start-of-run randoms, and it is believed that transient electronic noise from the initial biasing of the detector after neutralization was the source. Additionally, it was applied before fitting the charge χ^2 cut to improve it's accuracy. The cut was designed by studying the distribution of randoms for each given series. Any series with abnormally high χ^2 distributions were removed from the dataset. To reduce the live time lost to this effect, and additional 500 random events were taken as end-of-run randoms. These were not effected by this noise and were subsequently used to build our noise template.

Bias Mode (`cGoodHV_v53`)

The HT analysis operated with significant periods where one of the detectors was running in CDMSlite mode. While we do not include that exposure directly in the HT analysis, it does function as a multiple scatter veto. For a few series, however, the voltage generator required to operate the detector in high voltage mode was malfunctioning. This means that multiple scatters into the CDMSlite detector would be missed, compromising the background estimates. These periods were removed.

Trigger Enable (~`cNoTrigEnable_v53`)

Some events record zero live time due to the trigger enable bit not being recorded (for more on why this is see section 4.2.1). This removed such events.

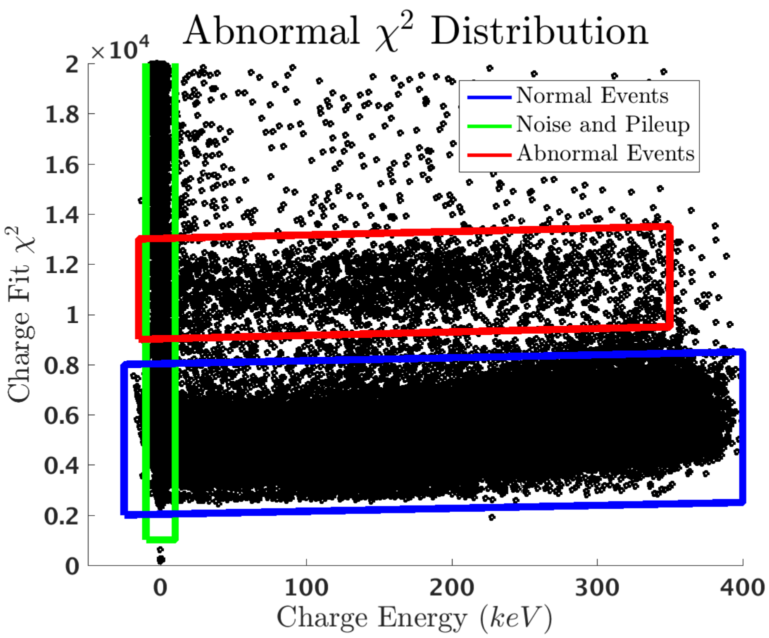


Figure 3.3: Ionization OF χ^2 residual vs total Ionization energy from ^{133}Ba calibration data. The well-reconstructed or “normal” events are shown in the blue box, while the large population of low-energy events in the green box are due to noise and pileup events. Both of these event populations were expected and seen in all series. In certain series there was a population similar in distribution to the normal events, but with systematically worse fits (offset by a constant amount). These are shown in the red box. It is believed that these events were due to transient noise from the start-of-run detector biasing that contaminated our start-of-run randomly collected traces [33].

Bias Voltage (`cWSBias_133_HT`)

For this analysis, every detector should be in the standard operating voltage modes with +2 V on Side 1 and -2 V on Side 2. This cut removes all periods where this is not the case.

Cf Activation (~`cPostCf_133_HT`)

To avoid the periods of high detector activation following the ^{252}Cf calibrations, the 48 hours immediately following these calibration periods is removed from the dataset. At this point, the neutron-activation event rate will have fallen below the cosmogenic background rate.

Electronics Glitch Cut (`cGlitch_133`)

Triggers can be caused from electronic glitches in the phonon readout channels. These appear as cascades of triggers in multiple detectors at once. These are removed by comparing the relative number of phonon to ionization triggers and looking for an excess.

High Ionization noise cut (`cQhighnoise_53`)

Periods of high ionization noise are not good candidates for a WIMP search since it can cause unexpected yield values and lead to mis-identification of events at low energy. Any given series will have some events where the ionization signal is indistinguishable from noise. This cut removes series where the fraction of such events is too high (5σ above a 20 event moving average).

Square Pulse (`cSquarePulse_v53`)

Two phonon channels (PDS1 for IT2Z1 and PCS1 for IT3Z3) can have events with large baseline offset noise, which to the eye look like square pulses (see figure 3.4). This cut examines the phonon pulse integral (less than 2σ below mean), fall time (if it is too fast), and baseline value to identify and remove these events.

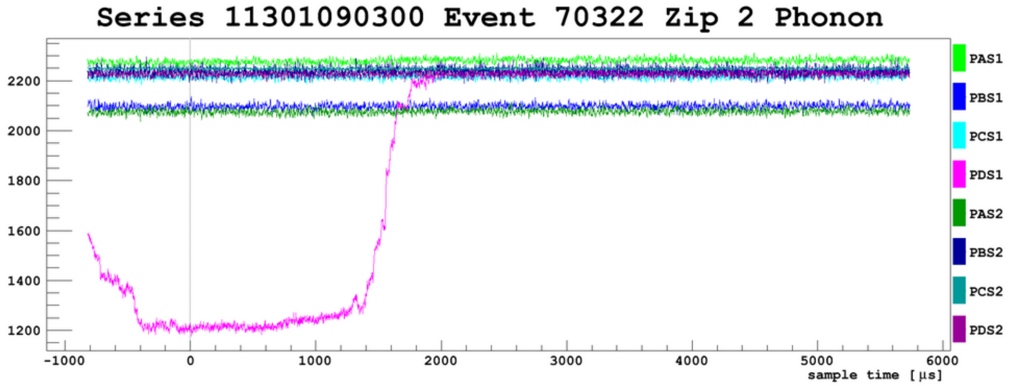


Figure 3.4: Digitized time-streams from each phonon detector during a so-called "square-pulse" event. This is actually noise in the baseline of `PDS1`, but due to the characteristically abrupt changes this noise appears similar to a square-pulse in shape.

Pre-Pulse cuts (`cPstd_v53` and `cQstd_v53`)

If an event has a pre-pulse baseline that is greater than 5σ from the series mean we get rid of it.

Randoms (~cRandom_V53)

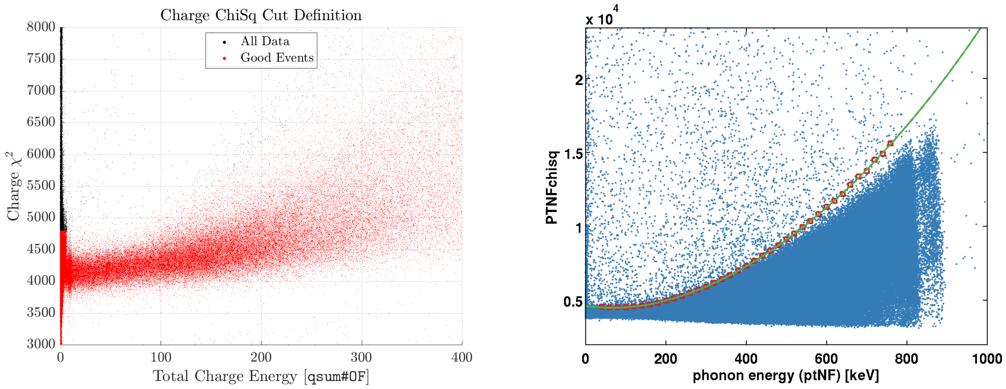
At the start and end of each run 500 random traces are taken and processed as normal events. There are also a smattering of such random traces taken during every run. These are used in many aspects of the analysis, from studying the noise behavior of our detectors, to constructing templates for the optimal filter algorithm. Given that they pass through our normal event reconstruction pipeline it is possible that a random event may reconstruct as an actual event or at least something event-like enough to be trouble. We explicitly remove the randoms to avoid this.

Reconstructed Quantity Outlier Rejection Cuts (`cGoodKSTest_V53_HT` and `cGoodQSTest_V53_HT`)

Finally, the last time-period cuts are the Kolgorov-Smirnov (KS) and Charge Stability specific Kolgorov-Smirnov (QS) tests which act as a safety check on the overall results, removing the outlier series. While off line analysis shifts and those responsible for assembling the list of series to process do a good job at marking and filtering out bad series from the analysis, there are still periods of subtle time variation in detector behavior or environmental effects that could affect the result. The standard CDMS approach to protect against this is the use of KS test to look for outlier series in the data by comparing distributions of key quantities. For more on this see [64].

3.4 Event Based Quality cuts

SuperCDMS is very sensitive to data quality. Many transient, spurious environmental noise sources can couple enough energy into our detectors to cause a trigger, even if there was no particle interaction. This next class of cuts removes this class of events by examining how well various portions of the event reconstruction process went.



(a) Ionization χ^2 [65]

(b) Phonon χ^2 [66]

Figure 3.5: Both the ionization (a) and phonon (b) χ^2 distributions suffer from flaring at high energy. To preserve these otherwise well-reconstructed events, each distribution is cut using a quadratic rather than a simple rectilinear selection.

Ionization χ^2 (`cQChiSq_V53`)

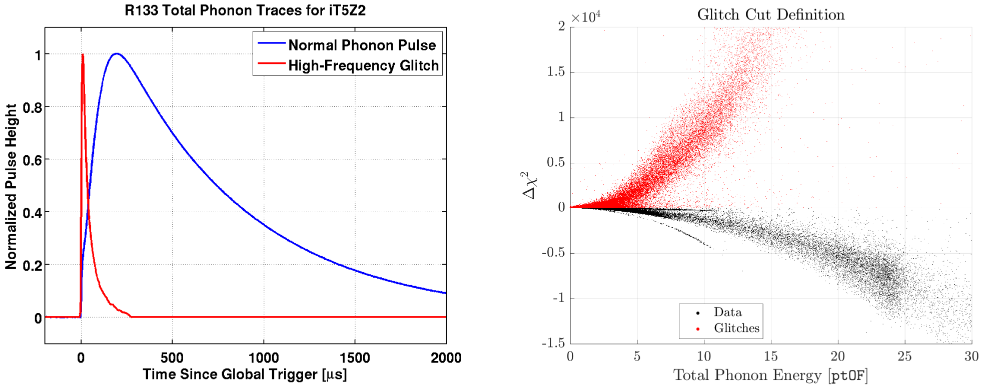
From section 2.2.7 the OF algorithm reconstructs events by fitting a template to the pulse and extracting an amplitude. Pulses are digitized with 4096 samples. If we were to construct Pearson's cumulative χ^2 test statistic between the fit template and our data, we would expect it to have follow a χ^2 distribution with 4095 d.o.f.[4]. As can be seen in figure 3.5a, this is only true for low energy interactions. For high energy events this is not true. As it turns out, our templates systematically have worse fits as energy increases. Our rule of thumb is any event with a Pearson's cumulative χ^2 test statistic that is greater than 3σ above the mean of the red band in figure 3.5a is removed. To do this the the $\mu + 3\sigma$ point is found for each energy bin, and the resulting points are fit to a quadratic. Every event above this line is removed. It should be noted that this is preformed on side-summed trace fits, so there are two of these goodness-of-fit tests per detector.

Phonon χ^2 (`cPChiSq_V53`)

Similar to the Ionization χ^2 cut, we also preform a goodness-of-fit cut on the phonon reconstruction. This is done in a very similar manner to the Ionization χ^2 cut, where for each energy bin the $\mu + 3\sigma$ point is fit to a quadratic everything above it is removed. This test is preformed on the summed phonon energy, and is only done once per detector. See figure 3.5b

---

[4] One degree of freedom is subtracted for the amplitude fit.



(a) (b)

Figure 3.6: The OF glitch cut removes a population of events characterized by very short fall-times. This is done by constructing templates for both a normal event and a so-called glitch event (a) preforming the OF fit and selecting all events that reconstruct as more signal-like than glitch-like (b) [67, 68].

Optima Filter Glitch Cut (`cGlitch1_V53`)

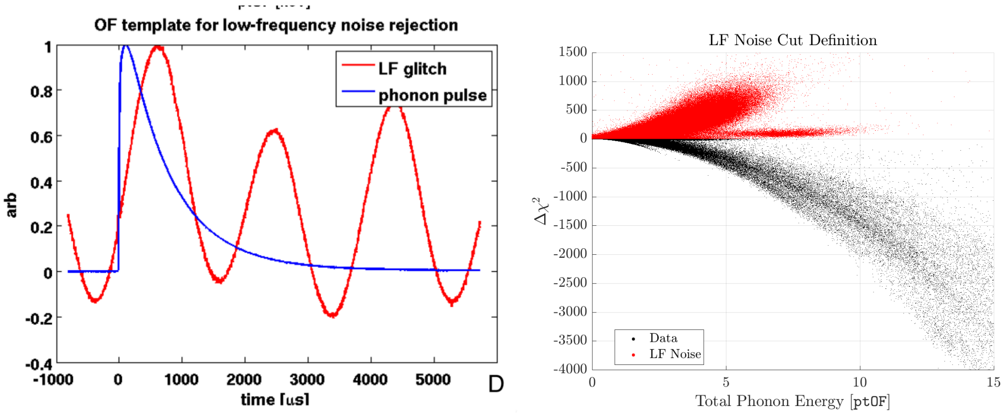
Occasionally, in the phonon readout channels, there are pulses with very short (~100 μs) fall-times. These are consistent with TES ETF relaxation time, and it is believed that glitches in the TES biasing voltage cause them. To remove these "glitch events" we turn to the OF. To remove these events we built a template of them, and then run the reconstruction algorithm using the glitch as the template. By comparing the χ^2 goodness-of-fit between the pulse and both templates (signal vs glitch) we can accept events that are more signal-like and reject events that have a better fit to the glitch template. This process can be seen in figure 3.6.

Low Frequency Noise Cut (`cLFnoise1_V53`)

Very low-frequency noise[5], can reconstruct in such a way that it looks like a nuclear-recoil interaction. It can also cause the energy of an actual interaction event to reconstruct incorrectly, if they occur simultaneously. This is dealt with identically to the OF glitch cut. We build a new LF-noise template and see if the OF fit to this noise template is better than the OF fit to the normal phonon template. This is shown in figure 3.7

---

[5] Which we believe is primarily due to the e-stem cryocooler.



(a) (b)

Figure 3.7: The LF noise cut is very similar to the glitch cut described in figure 3.6. An LF template is constructed (a) and events that reconstruct as more LF-noise than true event are rejected (b). [69, 51]

Good Start Time (`cGoodPStartTime_v53`)

During the OF fit, the start time of a pulse is allowed to float within a window around the trigger time. Ideally an event would be reconstructed near the center of the window. Some low energy pulses will have rise-times that are slow enough that the true start time of the pulse falls outside this window, causing the fit to “rail”. Similarly, due to the high event rate in some calibration datasets, “cross-detector pileup” events can occur that “rail” at the tail end of the OF search window. This occurs when the secondary trigger is issued just after the OF search window. These are excluded by ensuring that

$$(E_{\text{total}}^{\text{Phonon}} > E_{\text{min}}) \cap (-190\,\mu\text{s} < t_{\texttt{OF}} < t_{\text{threshold}}) \tag{3.1}$$

Where E_{min} and $t_{\text{threshold}}$ are detector-dependent thresholds. For more on this cut see [70].

3.5 Physics cuts

These are event-based cuts that remove events on physical grounds. These are either events that are inconsistent with the WIMP-scattering hypothesis, or have a very high probability of being a background interaction. In this sense, all remaining cuts in the analysis are “physics” cuts, however, the cuts listed below have an additional criteria: events that fail them are of no interest to our signal or background modeling efforts. As a result, they act to reduce the size of our dataset, and are generally applied uniformly with the above time-period and reconstruction cuts.

Muon Veto (`cVTStrict_133`)

Any event that is within 50 μs of a muon-veto trigger is removed.

Triggered event (`cTriggeredEvent_133_HT`)

The correct functional detector response includes an estimation of the trigger efficiency. This is calculated based on the number of events that actually trigger the detector, so for it to be valid, we can not include any pulses from a given detector that didn't actually trigger the experiment. This cut requires that any event under study for a given detector must have issued a trigger in that detector in the time window from -200 μs to $+100$ μs of the event.

Phonon Consistency Region cut (`cPCfSVM_2pct_Sandbox_v53_HT`)

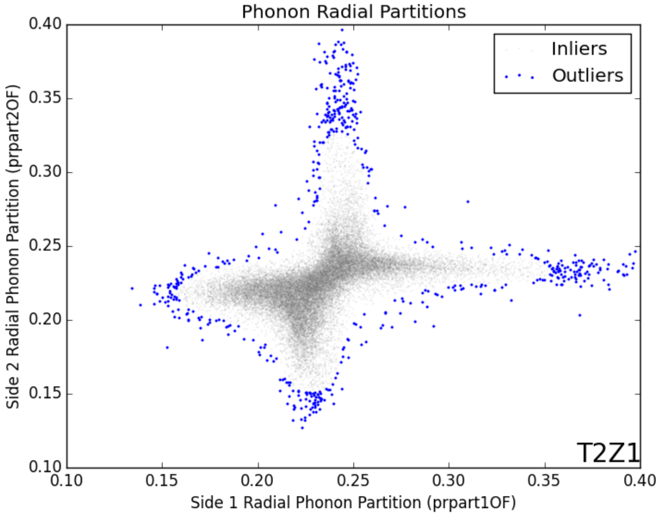


Figure 3.8: One-class support vector machines (SVM) can be used as an outlier rejection tool, which is the basis for the phonon consistency region cut. Plotted is the phonon radial partition for side 1 vs side 2 of ^{252}Cf calibration events for detector IT2Z1. The cut removes the outlying 2% of events (blue) from the phonon partition distribution [71].

As part of the BDT training for the previous low-threshold (LT) analysis [31], the team found that the results could be skewed significantly with the inclusion of certain outlier events. The majority of the LT preselection was done in terms of the charge partition and yield space, so any outliers in the phonon dimensions could lead to spurious results. In anticipation of similar problems to the LT efforts, we designed a cut to remove events that formed the extreme portions of the phonon partition distributions. The basis of this cut is the Support Vector Machine classification algorithm. The phonon partition quantities are combined in the SVM machinery and the outlying 2% of events were removed to make a clean sample.

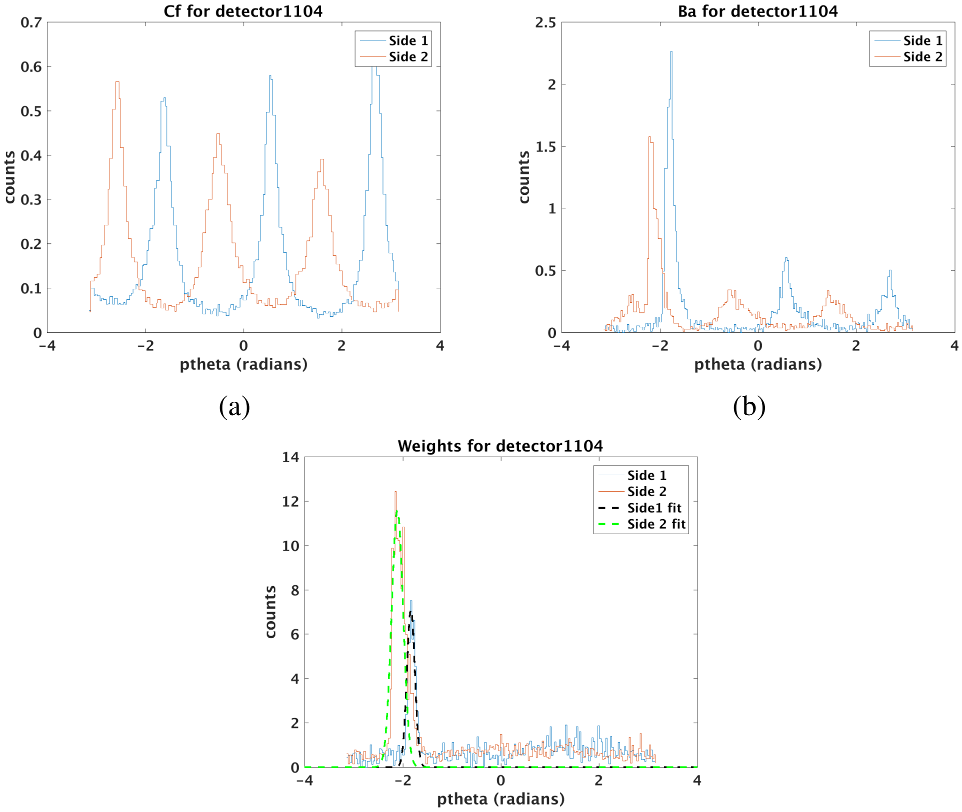
Localized Low-Yield Phonon cut (~`cSpot_v53_HT`)

During the investigation of the γ-sourced backgrounds, there was some surprise expressed at the large numbers of single-scatter γ's (from ^{133}Ba calibration) leaking into the nuclear-recoil band. When examining ionization quantity planes such as `qzpartOF` vs `qrpartOF` it was generally seen that these events had a preference for `qzpartOF` values larger than 0 and did typically occur at higher `qrpartOF` values. Unfortunately, they did not occur in a very tight cluster that would suggest easy removal. During some cross-checks, we noticed that these events were relatively tightly clustered in the phonon partition plane on both sides of the detector. These events occur along the flats of the wafer, near the DIB on each side of the detector, though the phenomena is more pronounced on Side 1. It is believed that these events are caused by the voltage on the DIB that is adjacent to the flat causing events to preferentially drift into the sidewall of the device. While this is a design flaw in the electronics, since the events are relatively localized, we can simply remove that portion of the detector. To accomplish this the "Spot cut"[6] was created to selectively remove these events. The variable in which the spot is most local is `pthetaOF`. Figure3.9 shows histograms of `pthetaOF` for both ^{252}Cf- as well as ^{133}Ba-calibration data. If we take the ^{252}Cf data to represent "normal" detector response we can use the ratio of ^{133}Ba to ^{252}Cf to gain an insight into where there is an excess of γ-sourced events.

`pxpartOF` and `pypartOF` can be dealt with in a similar fashion to `pthetaOF` except the 1d marginal distributions globally are hard to work with due to the triangular shape of the distribution in `pxpartOF`/`pypartOF` space. To simplify things we first

---

[6] So named because it appears that there is a "bad spot" in our detectors.



(a) (b)

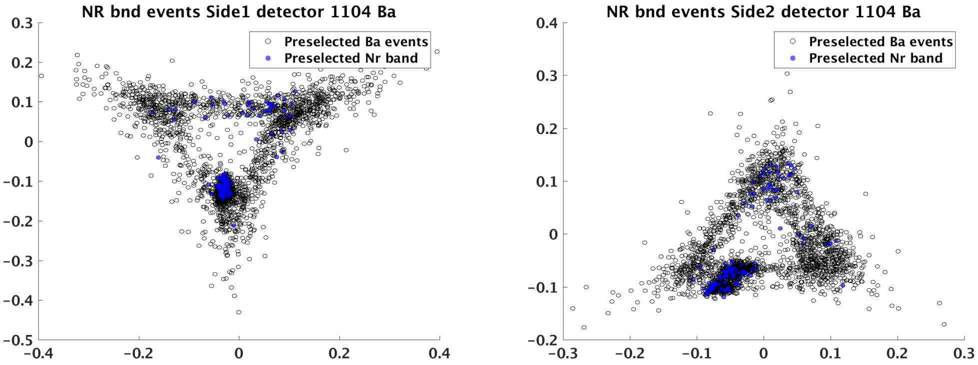
(c)

Figure 3.9: The “spot” of low-yield γ-sourced interactions is most localized in phonon θ-partition space. To find it the distribution of ^{252}Cf-sourced events (a) (assumed to represent uniform detector response) is subtracted from the same distribution constructed of low-yield ^{133}Ba-sourced events (b). The residual is fit to a Gaussian (c). This process is repeated in the X and Y phonon partitions and the intersection of these three Gaussians are removed.

constructed 2d histograms in `pxpartOF`/`pypartOF` space, found the peak and restricted ourselves to its local neighborhood. After restricting the domain to around the peak, we use the same re-weighting and fitting method as described above with the `pthetaOF` distribution. The intersection of the 2 sigma bands in all three variables calculated from the Gaussian fits to define the “spot”, and can be seen in figure 3.10.

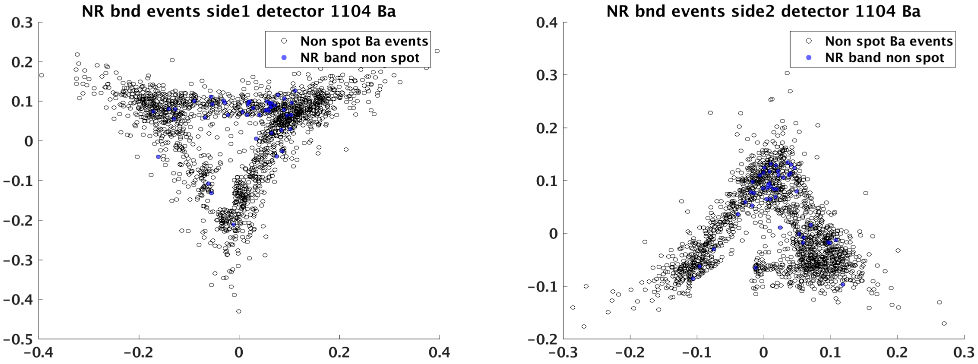
Ionization Threshold cut (`cQthresh_v53_HT`)

To remove the zero charge events, a charge threshold was included in the analysis. This cut was designed to remove any charge signals within 4σ of either the



(a) Side 1: before cut

(b) Side 2: before cut



(c) Side 1: after cut

(d) Side 2: after cut

Figure 3.10: Plots showing the action of the spot cut for detector IT2Z1. Each is the phonon Y vs X partition plane and depicts ^{133}Ba-sourced events with those that fall inside the 3σ NR-band highlighted in blue.

run averaged zero-energy noise distribution or the series-specific zero-energy noise distribution. The overall effect of this approach is to apply the average threshold in most cases, but to raise the threshold for more problematic series throughout the course of the run.

Energy Threshold (`cAnalysisThreshold_v53_HT`)

For each detector, the analysis threshold is the largest of the 95% trigger efficiency energy, the blinding lower energy limit, or a fixed value of 4 keV.

3.6 Modeling Preselection

All of the above cuts remove data that is either unreliable or is inconsistent with a WIMP scatter, but more than that, the data that is removed is largely uninteresting to the future pieces of the analysis. As a result, the preceding cuts are typically applied to all datasets, both WIMP-search as well as calibration, and only the events that pass all of them are considered for further analysis. This can be accomplished by simply using the reduced dataset resulting from the `cGoodEv`[7] cut application (referred to as a "skim") or by exploiting CAP's global mask functionality. Either way, the resulting dataset is smaller, and easier to work with. The remainder of the cuts discussed in this chapter are logistically different, however, in that both the passing and failing data is of continued interest. An example is the blinding cut. Our signal region is required to be a subset of our blinded region. Given that we do not have access to WIMP-search data inside this region, we must instead use calibration data to model our expected backgrounds[8]. If we want to validate our models, and compare them to actual WIMP-search data, however, we must do so *outside* of the blinded region. These regions of interest that reside outside of our potential signal region are referred to as "sidebands" of a particular cut, with this example being the "blinding sideband". Moving forward, this requires retaining the entire dataset, both WIMP-search and calibration, as well as the result of each successive cut's action on that dataset. In direct contrast with our previous cuts each addition to our preselection criteria significantly complicates our dataset and requires some new language to directly identify sub-components of interest, which I will outline below.

3.6.1 Nuclear-Recoil band cut

The most important aspect of our preselected region is the so-called "nuclear-recoil" or NR-band. The purpose of this cut is to only accept events are are very consistent with a nuclear-recoil interaction. To do this ^{252}Cf calibration data us used to select a large population of NR events, and all events with high ionization yield (`ytNF` $>$ 0.5) are removed to exclude stray ER contamination. The remaining population is binned in recoil energy and the populations in each have their ionization yield fit to a Gaussian. The mean of the resulting Gaussians μ is fit to the following power

---

[7] An "umbrella" cut that is simply all of the previously mentioned cuts applied together.

[8] A procedure that will be discussed at length in the next chapter.

law:

$$\mu = \alpha E_{\text{Recoil}}^{\beta} \tag{3.2}$$

and the standard deviation σ is fit to:

$$\sigma = \frac{\sqrt{aE_{\text{Recoil}}^{b} + c}}{E_{\text{Recoil}}} \tag{3.3}$$

The result of these fits can be seen in figure 3.11. Our preselection is defined as any data inside the 3σ NR-band. The remaining data (NR-sideband) is retained for use as an input to our background modeling.

3.6.2 Blinding cuts

Our actual treatment of the blinding criteria is slightly more complicated than I made it out to be when it was introduced in section 2.3.5. In a sense, blinding is a very rough draft of this entire analysis, after all, the goal is to identify a region of our parameter space that is consistent with a WIMP-scatter hypotheses. In this sense, there are components of the blinding cut that can be applied uniformly to all datasets and essentially behave as the data reduction cuts. For more on this please see [73].

Blinding Quality cuts

These are technically part of the blinding criteria but as they deal with data quality, events that fail them are of no interest and would only confuse further analysis. These are actually applied as a part of the series of cuts before the preselection and are only included here for completeness. These consist of a electronics glitch as well as a LF noise cut.

Blinding fiducialization

An event was a candidate to be blinded unless there was evidence that it was very close to the edge of our detector. This was done using ionization collection, but before the development of the position estimation parameters described in the previous chapter, and as a result the logic may be somewhat confusing. In essence, to select against hi-radius events, an event can be blinded if ionization is collected on the inner electrode ($> 2\sigma$ above the noise) or if there is clearly no outer collection ($< 8\sigma$ above the noise). In the Z, direction very similar logic is used except with side 1 and side 2 ionization collection. This data (both passing and failing) is included in the skim as it was of great interest in during the initial development of our cut-position optimizer. It allows for much higher statistics and greatly smooths things out, but there were notable problems that led to it being discarded for our final analysis pipeline. Namely it was possible that the optimizer would stop before it had cut away all unblinded points. The other problem was that it was very non-conservative.

NR Blinding Cut

In an analogue to the cut normal NR-band definition from section 3.6.1, the blinding criteria requires that events fall within the 3σ NR-blinding-band. The definition of this uses INT rather than OR or NF reconstructed quantities and as a result is slightly different than the normal NR-band. Events that fall in both of these bans are included in the preselection region. Events that are outside of both bands form our NR sideband. Events that are in one band but not in the other are referred to as the "NR-inconsistency-band" and are removed from further consideration.

3.6.3 Multiple Scatters

Due to their low interaction rate, the probability of a single WIMP scattering more than once in our detectors is vanishingly small. Naively, these multiple-interaction events could be identified by examining every event and tagging any where there is a simultaneous deposition of energy in another detector that exceeds some threshold. Low-frequency noise and glitches, however, make this far too permissive of a definition. To correct for this, the the LF noise and OF glitch cuts are applied to the secondary scatter events, and only events that pass both are tagged as multiple-scattering interactions. For more on this cut please see **??**

3.6.4 Saturation

High energy Compton or α particle interactions can cause a channel's ADC to saturate. As defining our background models uses a study of α interactions as an input, this data is tagged, but not completely removed.

3.6.5 Loose fiducialization

The purpose of these cuts is remove the bulk of the surface event interactions by hand, leaving only their tails. The blinding fiducialization cuts already do this, but they are defined using the INT reconstruction quantities and as a result they are somewhat inconsistent with the normal OF parameters that we will be using in this analysis. We also use the phonon z partition to remove the half of the two ^{210}Pb source detectors that are nearest the source. We save the surface event data to build our surface event background model.

3.6.6 Preselection Cut summarized

1. **blinding quality cut:** Applied as part of true preselection. Events removed by these cuts are not included in the modeling.
    - `~cGlitch_blind_v53`
    - `~cGlitch1_blind_v53`
    - `~cLFnoise1_blind_v53`
    - `~cVetoTime_blind_v53`
    - `~cDoNot_blind_v53_HT`
    - `cPrange_blind_v53_HT`
2. **blinding fiducial volume cut:** Events removed by these cuts are treated somewhat differently during model construction, but ultimately all events (passing and failing this cut) go into the final skim.
    - `cQin1_blind_v53`
    - `cQin2_blind_v53`
    - `cQsym_blind_v53`
3. **singles cut:** Used in model construction but events failing this cut do not go into the final skim.
    - `~cPmultTight_blind_v53_HT`

4. **NR cuts:** Used in model construction but events failing this cut do not go into the final skim.
   - `cNR_ytNF_v53_3sigma_HT`
   - `cNR_blind_HT`
5. **Saturation cuts:** Used in model construction but events failing this cut do not go into the final skim.
   - `cPsat_133`
   - `cQsat_133`
6. **Loose fiducialization cuts:** Used in model construction but events failing this cut do not go into the final skim.
   - 0.2 < `qrpartOF_zhalf` < 1.2
   - −0.8 < `qzpartOF` < 0.8
   - 0 < `pzpartOF` for iT3Z1
   - 0 > `pzpartOF` for iT3Z3

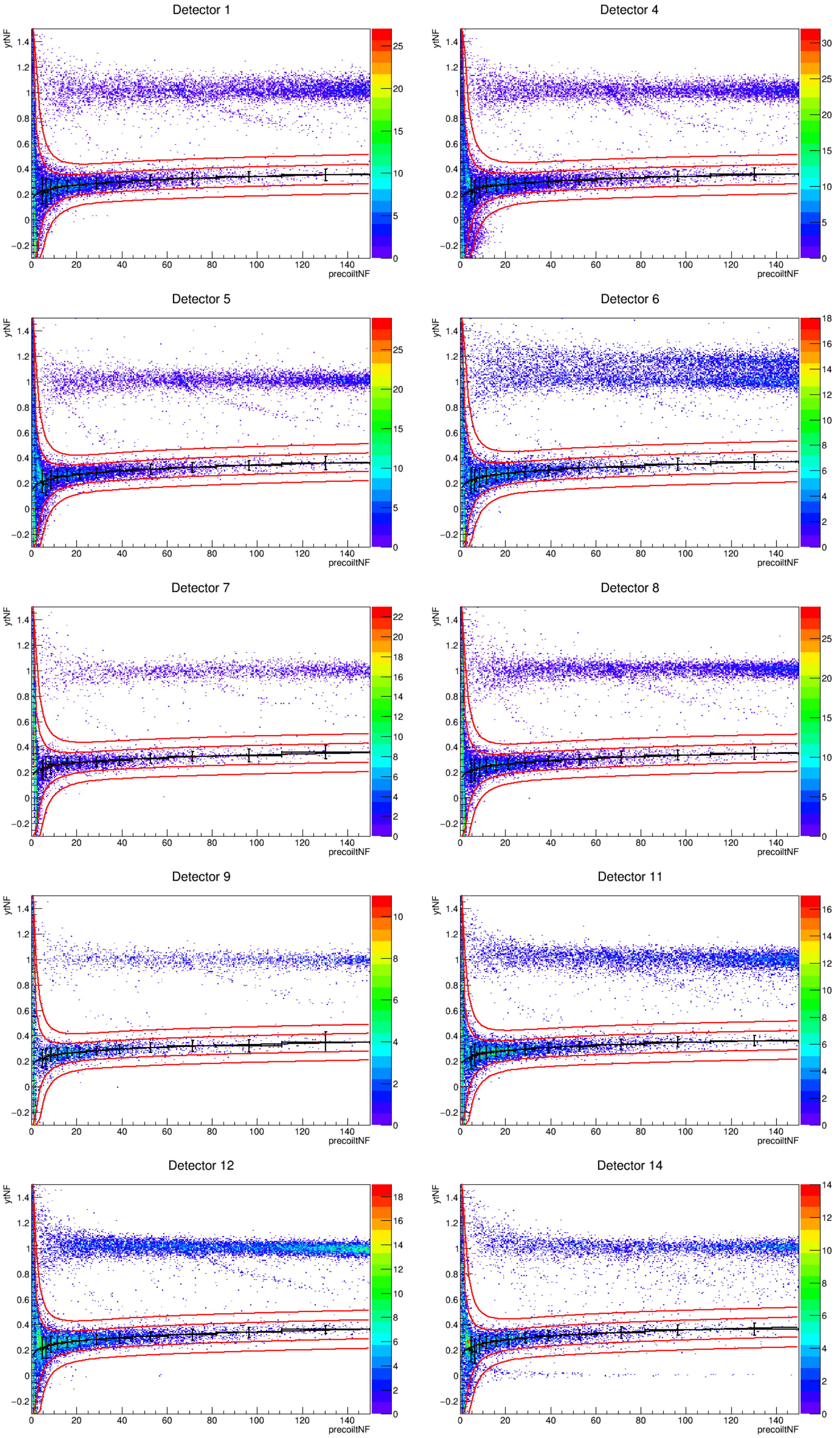


Figure 3.11: NR Band Fits
NR band fits for all detectors. The heatmap is a 2D histogram of ^{252}Cf-sourced events, and the 2 and 3σ NR-bands are shown in red [72].

Chapter 4

HIGH-THRESHOLD ANALYSIS: PARAMETRIC FIDUCIALIZATION

4.1 Signal Acceptance Region

4.1.1 Overview

Now that our data is well-formed we need to decide on a signal region we can unblind around. Ideally we would like to be able to quantitatively examine how different choices of signal region effect the expected sensitivity of our experiment. The two main components of this are the expected sensitivity to WIMP-Ge scattering events, and the expected number of background events misidentified as signal. In conducting a limit-setting analysis of this type, the general technique is to

1. Define a signal region that is expected to be free of background events.
2. After unblinding, treat any event inside the signal region as actual signal.
3. Using the Optimal Interval technique, set a limit, again assuming that all observed events are WIMP-scatter interactions.

As discussed in section 3.1, assuming we have no events after unblinding, our sensitivity to WIMP interactions above our energy threshold will scale like MT where M is the target mass and T is the experimental live time. As this technique assumes the signal region contains only signal, any background that is misidentified, or "leaks", into the signal region will immediately stop the MT scaling of sensitivity.[1]

---

[1] This is not exactly true. If the energy spectrum of these misidentified background events is substantially different from a WIMP spectrum, the Optimal interval method of limit setting will be able to exploit this fact, and continue to set stronger limits. For more information on the Optimal Interval method and out limit setting procedure please see section 4.5.4.

4.1.2 Parameters of Interest

To define a signal acceptance region that is free from background we need to first define some event parameters which can differentiate between signal and background. In SuperCDMS Soudan there are three basic kinds of discriminating parameters: energy estimates, interaction location or "position" estimates, and ionization yield. These have been previously described in some detail in section 2.4. For the exact CAP quantities used please see section 4.4.1.

4.1.3 Modeling

The next step is to understand how our expected sensitivity, which is to say how our signal acceptance and expected misidentified backgrounds, vary as a function of cut locations in each of these parameters. The complication here is that our final signal acceptance region is required to be a subset of the region of parameter space defined by our blinding procedure. As a result, we have to build models for what we expect our signal and backgrounds look like inside our blinded region. We can approximate our backgrounds using calibration-sourced data that has been modified to better match the unblinded regions of our WIMP-search data. Our signal model is slightly more complex. We use ^{252}Cf sourced neutron interactions to approximate a data-set of WIMP-induced nuclear-recoil events, which we modify to match what we would expect to see in our detectors from a given theoretical WIMP spectrum. This quantity would be proportional to our exposure, MT, but would need to be corrected to account for two complications. First, the event-based quality cuts so far described in our analysis have a probability of removing actual WIMP-interactions that varies as a function of recoil energy. This is usually referred to as the analysis efficiency or $\epsilon(E)$ where E is the recoil energy of the interaction. Second, we do not expect an equal number of WIMP interactions at all energies. This energy-dependent differential WIMP spectrum $\frac{\mathrm{d}R}{\mathrm{d}E}$ can be estimated theoretically. This exposure measurement that has been corrected by our expected WIMP-spectrum and analysis efficiency is called spectrum averaged exposure (SAE) and is formulated:

$$SAE = MT \frac{\int_{E_{min}}^{E_{max}} \mathrm{d}E \epsilon(E) \frac{\mathrm{d}R}{\mathrm{d}E}}{\int_{E_{min}}^{E_{max}} \mathrm{d}E \frac{\mathrm{d}R}{\mathrm{d}E}} \tag{4.1}$$

After we define models to understand what we expect our backgrounds and signal to look inside the blinded region, we can go about setting a final "fiducial volume" cut to define our signal acceptance region.

4.1.4 Fiducialization

Historically, many particle physics experiments define a signal acceptance region in part by excluding the physical periphery of the detector. This is done because detector response in the outer edges of the detector are often poorly understood, and because it is also the region where most background interactions happen. This reliable interior volume is typically referred to as a "fiducial volume". SuperCDMS is no different in these respects, but for this analysis we generalize the basic idea. As discussed in section 2.4 there are a number of different parameters, not just position, in our detector, that allow us to get a purer, more reliable signal. Cuts set on any combination of discriminating event parameters could be said to define a a "fiducial volume" which is a hypervolume in this parameter space. To maximize our sensitivity, we would ideally like to optimize the fiducial volume in some way, trading off between signal acceptance and expected misidentified background. To this end, we need a systematic way of ordering our potential fiducial volume cuts from most excepting of signal and background, to least, and could be characterized by a single fiducializing parameter. Our expected sensitivity could then be calculated as a function of this fiducializing parameter, and a set of cuts that maximizes it found. This procedure is termed "fiducialization" hereafter.

4.2 Spectrum Average Exposure Estimation

Before we can construct our signal and background models we need to estimate our SAE. As seen in equation 4.1, this requires estimating our exposure MT (or live time, given that the mass of our detectors is known), our analysis efficiency as a function of recoil energy $\epsilon(E)$, and choosing a differential WIMP-recoil spectrum to act as $\frac{\mathrm{d}R}{\mathrm{d}E}$. The following three subsections detail these three estimates.

4.2.1 Exposure Estimate

Estimating our *raw* exposure MT is the most straight forward of these steps. The mass of each detector is known, so this reduces to estimating the total live time re-

Table 4.1: `LiveTime Accounting`

Detector	LiveTime raw (day)	LiveTime corrected (day)
iT1Z1	285.57	285.65
iT2Z1	317.11	317.20
iT2Z2	322.19	322.29
iT2Z3	312.98	313.06
iT3Z1	234.33	234.40
iT3Z2	307.61	307.70
iT3Z3	242.50	242.57
iT4Z2	322.67	322.76
iT4Z3	347.33	347.43
iT5Z2	144.75	144.80

maining after the application of the time-period based cuts described in section 3.3. Ideally this could be done by simply adding up the `LiveTime` of all events that pass `cLiveTime_HT`. There is a subtle complication, that renders this less than ideal. `LiveTime` is recorded in units of milliseconds (ms), but the time stream is sampled by our ADC every 0.8 μs. This necessitates some degree of rounding in the recorded `LiveTime`. Normal WS-data has an event rate of about a Hz, so this should not introduce much of a systematic error in our accounting, and on average the rounding should be zero. However, during periods of higher trigger rates, there will be many events where the measured `LiveTime` is zero. These will not be averaged out and will cause us to under-estimate the total live time. To overcome this a second estimate of our experimental live time was derived from the history buffer information in our dataset, and used to correct this accounting. This is all summarized in table 4.1. For more detail on this estimate please see [74].

4.2.2 Efficiency Estimate

After accounting for the reduction in live time from our time period based cuts, we next turn to the event-based cuts and their effect on our exposure. For these cuts (such as one that selects well reconstructed events, or only allows events in a particular portion of the detector) the experimental live time surrounding the event is perfectly valid and throwing it away would underestimate our sensitivity to WIMP interactions. There is a chance, however, that a valid WIMP-scattering event would be removed by removed by these cuts. We account for this using an “efficiency curve”[2] or simply an efficiency. This is the probability, as a function of energy, of a true WIMP-interaction in our detector making it through all the cuts of our analysis and into the signal acceptance region and is denoted $\epsilon(E)$ in equation 4.1. To calculate this we use ^{252}Cf calibration data as a WIMP-scattering proxy and examine how our various cuts remove this population. Quantitatively, for the E^{th} energy bin we can calculate the efficiency of a cut (or series of cuts) as follows:

$$\epsilon_E = \frac{n_E^{\text{Passing}}}{n_E^{\text{Total}}} \tag{4.2}$$

where n_E^{Passing} are the number of NR-band calibration events that pass the cut and n_E^{Total} are the total number of NR-band ^{252}Cf events in the bin. There were a select number of cuts where, for various reasons, ^{252}Cf calibration data could not be used to calculate the cut efficiency. These are summarized below, but for a much more detailed description of their treatment please see [75].

Non-^{252}Cf Efficiency Cuts:

- **Trigger Efficiency** For a variety of reasons, there is some probability that an event that otherwise reconstructs perfectively well, does not induce a hardware trigger in our iZIP array. As discusses in section 3.5, we require that all events that make it into our signal region need to have issued a valid trigger. As a result, this population of events has, by construction, a trigger efficiency of unity. To get an unbiased population that would actually allow us to measure the efficiency of this cut, we turn to multiple-scattering events in WIMP-search data. During WIMP-search operation, if any iZIP issues a trigger the entire detector array is read out and reconstructed. By examining non-primary events that reconstruct well, but do not issue a trigger we can measure our energy dependent trigger efficiency.
- **Single-scatter Cut** ^{252}Cf data has a much higher probability of causing scattering events in multiple detectors than true WIMP interactions, and as a result it is not a good candidate to measure the efficiency of this cut. Naively, the efficiency of this could could be assumed to be 1. After all the probability of a real WIMP event being removed by this cut seems vanishingly small. In

---

[2] This would probably be better named “WIMP acceptance probability curve”, but the term *efficiency* is very strongly ingrained in our collaboration, and it would be difficult for me to avoid using that word.

reality, due to noise fluctuations, it is possible for a WIMP-like single scatter to fail this cut due to noise fluctuation in a secondary detector that reconstructs to an event. To measure this we use our randomly collected traces from WIMP-search operation. Just like WIMP interactions, randoms should only fail this cut due to secondary-detector noise fluctuations. This efficiency is independent of the energy of the primary interaction, and is fixed at about ~99%.

- **Ionization** χ^2 As describe in section3.3, early in the run it was discovered that certain series suffered from contamination of random events (used to build our OF noise template) resulting in problematically high or low χ^2 values. These were removed by the random trigger χ^2 cut (`cGoodRandomChi2_V53_HT`). Importantly, this cut was applied before the ionization χ^2 cut was developed. As a result of these removed series, the ionization χ^2 between ^{252}Cf calibration and WIMP search is systematically different. As the ionization pulse-shape does not depend on an event being an electron- or nuclear-recoil the ^{133}Ba calibration was used as a proxy instead.

- **Good Phonon Start Time** As described in section 3.4, this cut seeks to remove events that are poorly reconstructed due to the pulse start time being too close to the edges of the OF search window (or even outside of it). This effects depends on event rate, and as a result, a high percentage of calibration data is removed by it compared to WS data. Seeing as the criteria used to construct this cut are completely different than those used in the blinding definition, the blinded WIMP search data was thought to be a good proxy to the interactions that we would find in the signal region if it were unblinded. As a result this is what was used.

- **Muon Veto** In a sense, our muon veto cut is just a different kind of single-scatter cut, where the secondary detector is the active muon veto shield. As a result, we again turned to random traces from WIMP-search data to calculate its efficiency. By examining the fraction of random traces that were collected concurrently to a veto event (and thus removed by the cut) we can calculate the veto efficiency. This is again, not dependent on the interaction energy of the primary scatter, and a fixed value of ~98.6%.

The estimated efficiency for each detector after the time-period, quality, physics, and preselection cuts are shown in figure 4.1.

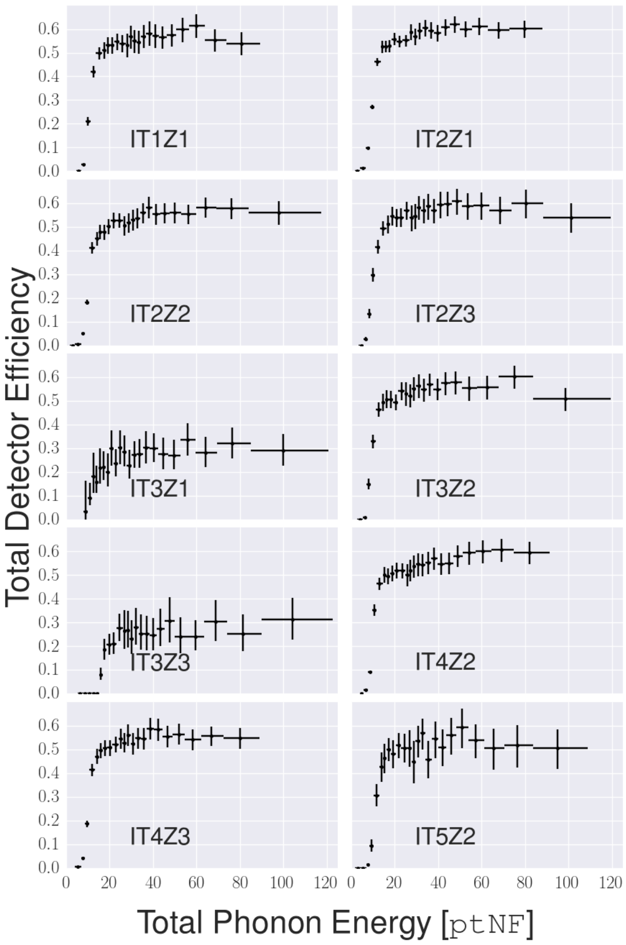


Figure 4.1: Efficiency curve estimates for each of the 10 HT detectors constructed to aid in our background and signal modeling. The errors include the systematic and statistical uncertainties. The notably reduced efficiency experienced by IT3Z1 and IT3Z3 is due to removing the half of each of those detectors closest to the installed ^{210}Pb sources. [33]

4.2.3 WIMP-spectrum estimate

Now that we have our exposure and total analysis efficiency estimates taken care of, the last piece we need before we get to the actual signal and background modeling is the differential WIMP-recoil spectrum, $\frac{\mathrm{d}R}{\mathrm{d}E}$. This analysis uses a slightly modified version of the theoretical WIMP halo model as described in [76]. In this model, the WIMPS are assumed to be smoothly and isotropically distributed in a sphere surrounding our galaxy. Although the average velocity of the dark matter in the halo is zero with respect to our galaxy, the WIMPs are thought to be fully virialized with a Maxwellian velocity distribution, f_v, which when taken with respect to the earth has the form

$$f_v(\vec{v}, \vec{v_E}) \propto \rho_0 e^{\frac{-|\vec{v}+\vec{v_E}|^2}{v_0^2}} \tag{4.3}$$

where $\vec{v}$, $\vec{v_E}$, and v_0 are the the velocity of the WIMPs, the velocity of the earth and the peak of the WIMP velocity distribution respectively. We take $|\vec{v_E}|$ to be 232 km s^{-1} which is the mean velocity of the earth as it travels around our galaxy, and v_o to be 220 km s^{-1}. The local WIMP mass density, ρ_0, is taken to be 0.3 GeV c^{-2} m^{-3}. This Maxwellian distribution is truncated at the high end astrophysically by the galactic escape velocity of $v_{\mathrm{esc}} = 544$ km s^{-1}. For a given recoil energy, E, in our detector there is also an associated minimum WIMP velocity that can induce such a recoil given by

$$v_{\min} = \sqrt{\frac{m_{\mathrm{Ge}} E}{2\mu^2_{[\chi,Ge]}}} \quad \text{where} \quad \mu_{[\chi,Ge]} = \frac{m_{\mathrm{Ge}} m_\chi}{m_{\mathrm{Ge}} + m_\chi} \tag{4.4}$$

The resulting differential WIMP-recoil spectrum has the form:[3]

$$\frac{\mathrm{d}R}{\mathrm{d}E} = k_r \frac{R_0 v_0 \sqrt{\pi}}{4 v_E r E_0} \begin{cases} \mathrm{erf}(\frac{v_{\min}+v_E}{v_0}) - \mathrm{erf}(\frac{v_{\min}-v_E}{v_0}) - \frac{4v_E}{\sqrt{\pi}v_0} e^{-(\frac{v_{\mathrm{esc}}}{v_0})^2} & 0 < v_{\min} < v_{\mathrm{esc}} - v_E \\ \mathrm{erf}(\frac{v_{\mathrm{esc}}}{v_0}) - \mathrm{erf}(\frac{v_{\min}-v_E}{v_0}) - \frac{2(v_E - v_{\min} + v_{\mathrm{esc}})}{\sqrt{\pi}v_0} e^{-(\frac{v_{\mathrm{esc}}}{v_0})^2} & v_{\mathrm{esc}} - v_E < v_{\min} < v_{\mathrm{esc}} + v_E \\ 0 & v_{\mathrm{esc}} + v_E < v_{\min} \end{cases} \tag{4.5}$$

Where the total event rate R_0, kinematic factor r, and normalization ratio k_r are defined as

$$k_r = \left[\mathrm{erf}\left(\frac{v_{\mathrm{esc}}}{v_0}\right) - \frac{2v_{\mathrm{esc}}}{\sqrt{\pi}v_0} e^{-(\frac{v_{\mathrm{esc}}}{v_0})^2} \right]^{-1} \tag{4.6}$$

$$r = \frac{4 m_\chi m_{\mathrm{Ge}}}{(m_\chi + m_{\mathrm{Ge}})^2} \tag{4.7}$$

$$R_0 = \frac{2\rho_0 N_{\mathrm{Avogadro}}}{\sqrt{\pi} m_\chi A} v_0 \sigma_0 F^2_{\mathrm{SI}}(E) \tag{4.8}$$

---

[3] For an outline of the full derivation of the differential WIMP-recoil spectrum please see [76], but it should be noted that the integration over the velocity spectrum that is outlined there can allow for negative scattering rates for some velocity combinations. For the full correct derivation please see [77] appendix A.

with σ_0 being the total zero-momentum cross section and $F_{\rm SI}(E)$ the spin-independent Helm nuclear form factor [78].

$$F_{\rm SI}^2(E) = \frac{9 j_1^2(r_0\sqrt{2m_{\rm Ge}E})}{2m_{\rm Ge}Er_0} e^{2m_{\rm Ge}Es^2} \tag{4.9}$$

where j_1 is a spherical Bessel function and $s \approx 0.9$ fm is the nuclear skin thickness.

4.3 Model dataset Construction

Now that we have a handle on how to calculate our SAE, our next step is to produce so called “model datasets”. These sets would ideally be high-statistics versions of a particular interaction type occurring inside our blinded region. There are four main interactions we try to model using this method. They are:

1. Low-yield γ-sourced background interactions.
2. ^{210}Pb sourced near-face background interactions.
3. Cosmogenic as well as radiogenic-sourced neutron background interactions.
4. Sensitivity to WIMP-Ge scattering interactions.

The production of these modeling datasets occurs in three steps. First, an appropriate subset of our calibration data is selected to act as a proxy for the interaction type being modeled. Second, systematic differences in the calibration dataset and the target data are corrected for. This correction occurs in different event parameters depending on the model. Third, the model is normalized so the absolute rate of interactions is consistent with our blinded WIMP-search dataset. These last two steps are complicated enough that I will outline them briefly below, before moving on to describe the actual model creation. It should be noted that this is probably an incomplete description of all potential sources of backgrounds that might find their way into our signal region, but these were the only sources that had calibration proxies with which we could construct models. Notably absent is a good model for ^{210}Pb sourced near-sidewall interactions, as there was no ^{210}Pb source installed around the periphery of any of our detectors. In the end, it turned out our models were fairly consistent with our blinded data, and that γ-sourced backgrounds, especially those at high radius, were the dominate source of background in our blinded region.

4.3.1 Systematic density correction

In many ways this is the most complicated step in our model creation. The models we are making are event-based: ^{133}Ba calibration events for our γ-sourced background model, ^{252}Cf calibration for our WIMP and neutron model, and surface source events for our ^{210}Pb model. If these calibration datasets were perfectly sampled from their corollary background distributions our work would be done—up to a normalization factor (more on this in the next section). Sadly, all of the calibration datasets differ systematically from the WIMP search set they are attempting to model. To offset this, we re-weight each event in our calibration dataset in such a way that if we were to construct a weighted pdf estimate (such as a histogram) from the calibration data, it would exactly match the corresponding (unweighted) pdf estimate in the WIMP search data. We colloquially refer to this new RRQ of re-weighing factors as the "weight vector". Construction of this new RRQ happens as follows (again leaving overall normalization for later).

1. Estimate the calibration and WIMP search distributions
2. Take the ratio of the two distributions. Call this the *weight distribution*.
3. Re-sample the weight distribution at points that correspond to our actual calibration events to recover the weight vector RRQ

Ideally this would be done simultaneously on every RRQ of interest to us (namely ionization and phonon radial and z partition, recoil and ionization energy, and ionization yield) using the full multi-parameter density estimates. Sadly, we quickly run into the curse of dimensionality. Even just using two parameters our data is too sparse to construct a meaningful density estimate; seven would be far too much to ask. As a result we simply construct one marginal weight vector for each RRQ we want to preform a correction on and then multiply them together on an event-by-event basis. This assumes that 1. there are no large inter-RRQ correlation of the systematics and 2. each RRQ's correction is just as important as any other.

4.3.2 Absolute normalization

How is it that we go from a weight vector (which describes a relative systematic correction) to a useful background or signal model? The answer is normalization.

Before normalization the weight vector describes our relative conversion from a calibration dataset to a dataset that has been sampled from our desired background or WIMP signal distribution. To turn that into a useful quantity we normalize the $\mathcal{L}^1$ norm of our weight vector to be the quantity we are interested in modeling (such as number of background events in our signal region for our background model). This is a convenient choice. It allows the units of our new normalized weight vector be $\frac{\text{model quantity}}{\text{calibration event}}$. It also makes the survival function of our model as a function of any other RRQ easy to calculate by simply preforming a cumulative sum.

4.3.3 Background Model: Gamma sourced events

4.3.3.1 Systematic gamma correction

Qualitatively, the first two steps of the systematic correction outlined above can be summed up by:

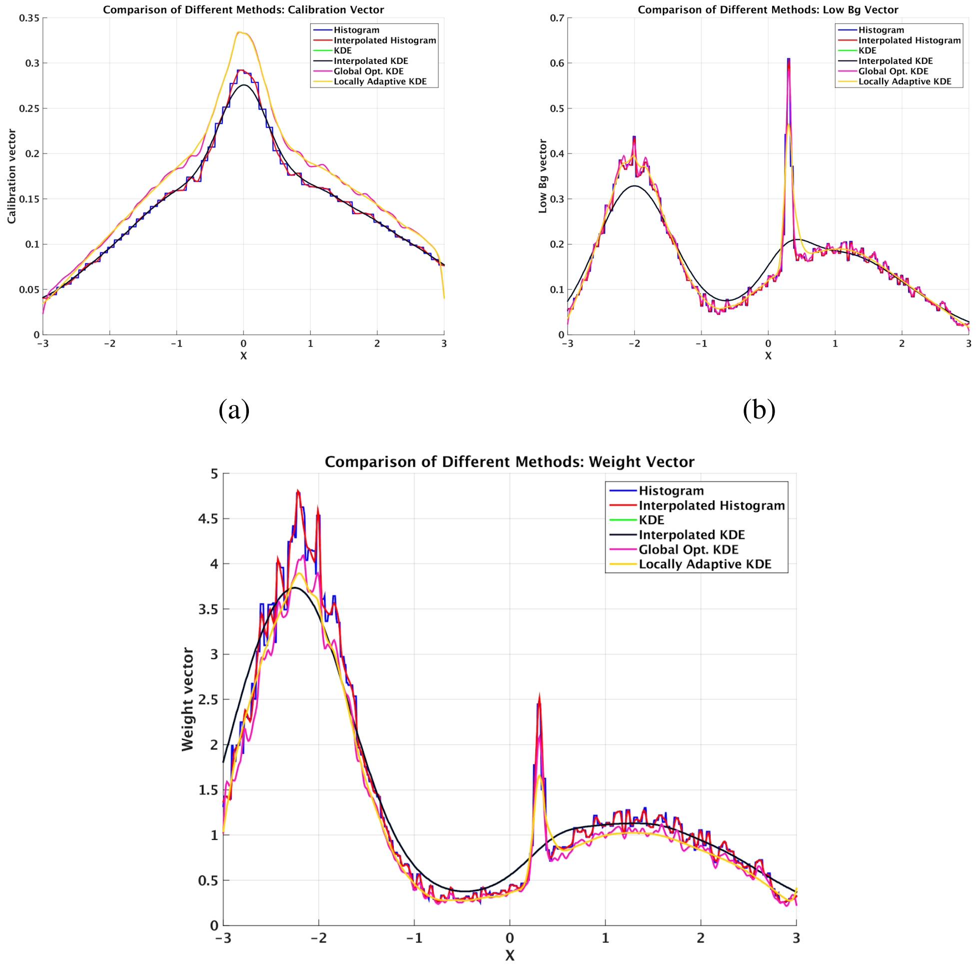
$$X_w(\vec{x}) = \frac{X_{WIMP}(\vec{x})}{X_{calib}(\vec{x})} \tag{4.10}$$

where X_{WIMP}, and X_{calib} are the marginal density estimate of RRQ $\vec{x}$ for the wimp search and calibration data sets respectively and X_w is the resulting weight distribution. For the gamma model, X_{calib} is constructed from the sideband, single-scatter, ^{133}Ba data that falls inside the fiducial blinding region. X_{Wimp} is constructed from similarly selected WIMP-search data. We restrict ourselves to the fiducial blinded region to ensure we do not contaminate our gamma model with large amounts of ^{210}Pb. Inside this fiducial blinded region we have no access to NR band single-scatter events in the WIMP-search data, so we utilize the NR-band sideband as a proxy. As a demonstration of how this works here is our re-weighing machinery acting on a contrived RRQ (that is just the sum of some Gaussian). Using the following fake RRQ definition:

$$\vec{x}_{calib} = \mathcal{N}(0.5, 2, 1e5) + \mathcal{N}(-1, 1, 1e4) + \mathcal{N}(0, 0.3, 1e5) \tag{4.11}$$

$$\vec{x}_{WIMP} = \mathcal{N}(1, 1, 1e4) + \mathcal{N}(-2, 0.5, 1e4) + \mathcal{N}(0.3, 0.04, 1e3) \tag{4.12}$$

Where $\mathcal{N}(\mu, \sigma, N)$ is a vector of N samples from a normal distribution with mean μ and standard deviation σ. The resulting RRQ and weight density estimation can be seen in figure 4.2. We have implemented many different estimation methods,



(a)

(b)

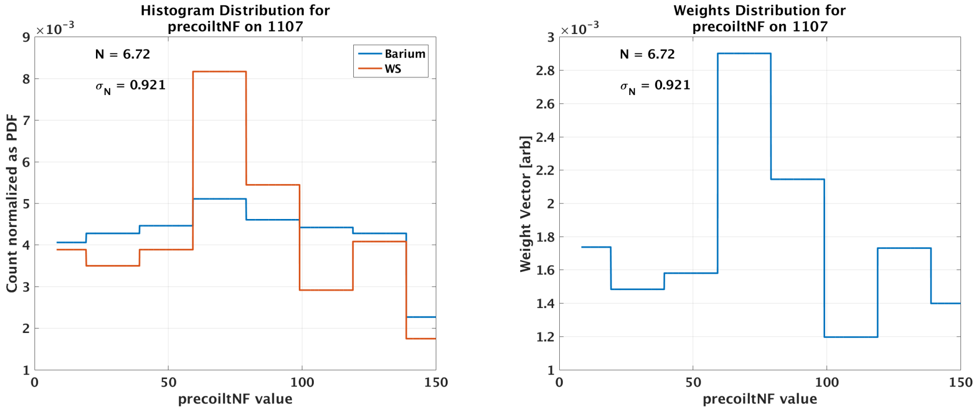
(c)

Figure 4.2: Overview of our systematic correction using the data from equation 4.11. The calibration data (a) is a high statistics dataset that extends into the blinded region. We want to use it to model a particular distribution in the WIMP-search data (b), but as we can see they are systematically different distributions. To correct this we divide the two and form the weight distribution (c). This can be used to sample weights for the points in (a). As can be seen here a number of density estimation methods were examined, but in the end a simple histogram (blue) was chosen.

but have settled on the histogram as the best preforming. At this point the actual per-event weights can be found for this RRQ, by sampling the weight distribution at each of the calibration dataset's points or:

$$\vec{x_w} = X_w(\vec{x_{calib}}) \tag{4.13}$$

This process are repeated for the discriminating parameters described in section 2.4, and the results can be multiplied event-by-event. The result is the un-normalized weight vector. Although it is lower in statistics and less clear than the fake example shown in figure 4.2, this process is plotted in figure 4.3.



(a) (b)

Figure 4.3: An example of γ-sourced weight vector construction for detector IT3Z1. This particular example corrects the recoil-energy distribution, and an identical process is carried out for all discriminating parameters. On the left are the `precoiltNF` distributions for ^{133}Ba-sourced (blue) as well as our WS sideband data (orange). On the right is the resulting weight distribution.

4.3.3.2 Gamma normalization

For our gamma model, we want to model the number of gamma events that are misidentified as signal by our detectors (obviously as a function of all other RRQs). To this end, we want:

$$||\vec{w^i}||_1 = \sum_k w_k^i = N_{NRSS}^i \tag{4.14}$$

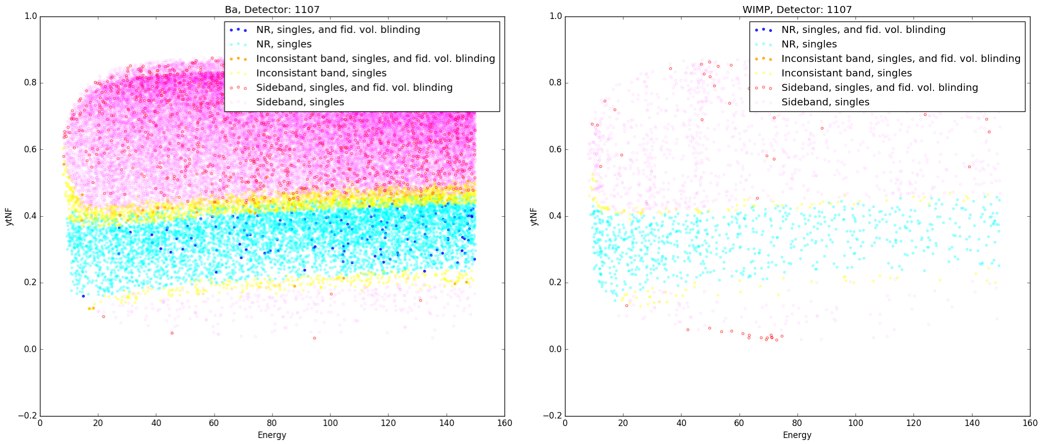
Where N_{NRSS}^i is the number of single scatter events that fall inside the NR band from our WIMP-search dataset in detector i. How do we measure N_{NRSS}^i? Outside the fiducial blinding region, this measurement is easy. We simply count the number of single scatter events in the NR band from our WIMP-search dataset. These events are mostly going to be at high Z and R positions (The fiducial blinding cut is defined in terms of charge symmetry (cQsym_blind) as well as radial partition (cQin1_blind and cQin2_blind) amongst other things). The number of NR single scatter events inside the fiducial blinding region is not accessible in the WIMP-search data, after all, these cuts together define the blinded data that has been removed. Instead of counting, it is estimated using the ^{133}Ba dataset and the sidebands as follows:

$$N_{NRSS} = N_{NRSS}^{unblind} + N_{NRSS}^{blind} \tag{4.15}$$

where

$$N_{NRSS}^{blind} = N_{NRSS}^{Ba} \frac{N_{SBSS}^{WIMP}}{N_{SBSS}^{Ba}} \tag{4.16}$$

where N_{NRSS}^{Ba} is the number of ^{133}Ba NR-band single scatters in the fiducial blinding region (shown as the dark blue points in figure 4.4), N_{SBSS}^{WIMP} is the number of WIMP-search sideband (out of NR band) single scatters in the blinding region, and N_{SBSS}^{Ba} is the number of ^{133}Ba sideband single scatters in the fiducial blinding region (dark red in figure 4.4). There was initially some interest in allowing our preselection region to include the fiducial unblinded region, which would set the total γ-sourced background model normalization to N_{NRSS}. This was to provide more statistics for our gamma model. This would rely either on our optimizer cutting out all fiducial unblinded events on its own (as they are not candidates for our signal



(a)

(b)

Figure 4.4: Ionization yield vs recoil energy depicting a portion of our preselected region. Both the ^{133}Ba-calibration (a) as well as the WIMP-search data (b) are shown. The events are colored based on various preselection criteria. Importantly to estimate the total number of single-scatter NR-band, gamma-sourced events that fall inside our blinding region for WIMP-search data (which would be colored dark blue but have been blinded) we start with the dark blue events in (a) and correct by the ratio of dark red events between (a) and (b).

region) or instituting a post-optimization cut to remove them. The former event did not come to pass (our optimizer was never so accommodating) and instituting a post-optimization further-fiducializing cut calls the entire purpose of optimization into question. As a result we restricted our preselection conditions to only include events inside of the fiducial blinding region, and the total γ-sourced background model normalization is N_{NRSS}^{blind}.

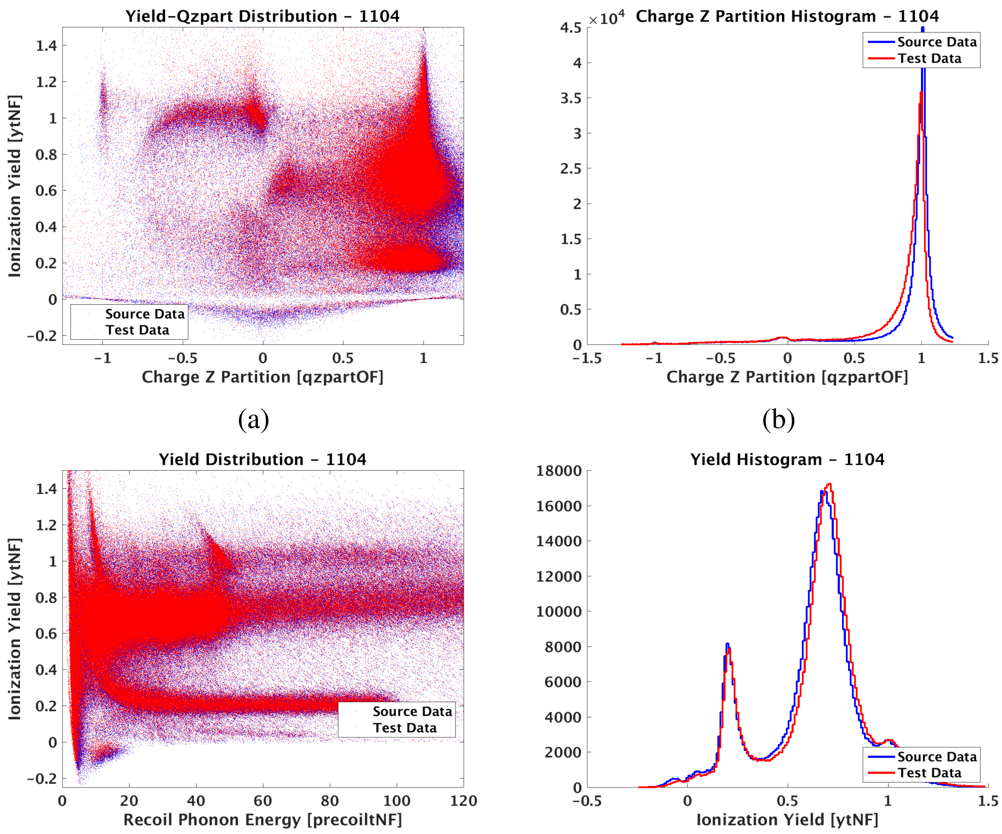
4.3.4 Background Model: ^{210}Pb sourced events

4.3.4.1 Systematic ^{210}Pb correction

The ^{210}Pb model creation deviates somewhat from the prescription we have used in the previous section. The ^{210}Pb calibration sources are only found on the face of two detectors (side 1 of IT3Z1, and side 2 of IT3Z3), and they are always on. They are part of the WIMP-search dataset, so while inter-dataset systematics are not that much of a concern, blinding is. In order for our model to extend into our signal region we have utilized the unblinded WIMP-search data (March - June 2012). The other main issue is inter-detector systematics. To deal with the different behavior between different detectors, the data from the source detector is shifted and smeared by the quadrature difference between it and the noise distribution of the target detector. Simply:

$$\vec{E}_{target} = \vec{E}_{source} + (\mu_{target} - \mu_{source}) + \mathcal{N}(0,\ \sqrt{\sigma^2_{target} - \sigma^2_{source}},\ length(\vec{E}_{source})) \tag{4.17}$$

Where E ranges over each of the base energy estimators from both the phonon channels (`p*#OF`) as well as the charge channels (`q*#OF`). As it turns out, the noise often turns out to be higher on the source detectors then the target detector. When this is the case, only the shift in mean is preformed. No extra randomly sampled Gaussian noise is added. This should be the conservative approach. Using these newly constructed base quantities, more useful discriminating quantities such as ionization yield charge partitions are constructed. Figure 4.5 shows how a few of these quantities are corrected by the above procedure for detector iT2Z1.



(a) (b)

(c) (d)

Figure 4.5: ^{210}Pb-sourced correction between detectors. The data from our side 1 source detector, IT3Z1 (blue) is plotted in various parameters of interest. The data in blue blue has been smeared and shifted to better match the noise characteristics of IT2Z1 [79].

4.3.4.2 ^{210}Pb normalization

Now that we have an idea of what the systematics look like, we need to figure out what the total rate of ^{210}Pb-sourced background events are. Once we do, we can normalize our weight vector just like in the γ-sourced model. Although the systematics for this dataset are small (as seen in the previous section), the source detector faces (iT3Z1 side 1 and iT3Z3 side 2) have substantially higher rates than the remaining detectors. How do we connect the ^{210}Pb-sourced events between detector faces? We use a specially unblinded dataset to measure the high-energy ^{210}Pb-sourced α rate [80]. The ^{210}Pb to ^{206}Pb decay produces a high energy α-decay, and as the sources are in secular equilibrium the rate of these α's should be proportional to ^{210}Pb-sourced surface events. These events are very high energy (~5 MeV) and expected to have very low relative ionization energy, and subsequently low yield.

As mentioned in sections 2.2.4, these high energy events can't be reconstructed using our OF filter quantities. Instead of being identified as quantities in `qsummaxOF` vs `precoiltNF` plane it does it in the `qsummaxF5` vs E^{Recoil} = `psumTFP−plukeqF5` as seen in figure 4.6. This process involves two steps in the rescaling. First, the total number of source α's (from iT3Z1 side 1 and iT3Z3 side 2) is estimated from the experimental live time:

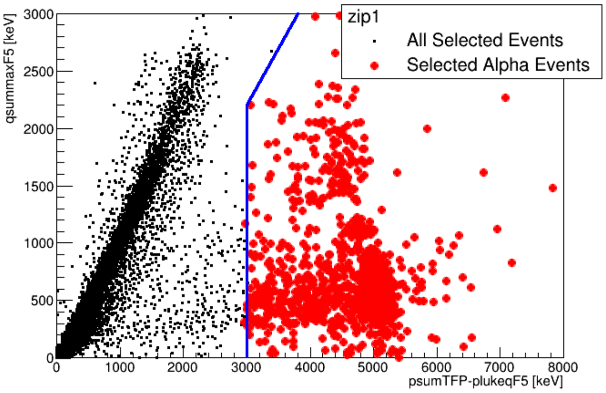


Figure 4.6: Selection of the high-energy ^{210}Pb-sourced α-particle interactions for detector iT1Z1. As these events tend to reconstruct very poorly using our traditional OF methods, the ionization energy is estimated using the F5 pulse shape extrapolation, and the phonon energy is reconstructed via a single exponential fit at late time (TFP). The alpha population (red) is characterized by being very high energy, but having a low ionization yield (as an alpha acts much like a recoiling nucleus in the crystal) [80].

$$N_s = N_t \frac{LT_s}{LT_t} \tag{4.18}$$

where N_s is the number of α's in the source sample dataset, N_t is the number of alphas in the study, LT_s is the live time of the source dataset, and LT_t is the live time in the unblinded α study. Then the weight normalization is determined by the ratio of the alphas in the source sample to the alphas in the full exposure for the non source detectors:

$$||\vec{w^i}||_1 = \frac{N_i}{N_s} \tag{4.19}$$

where $\vec{w_i}$ is the weight for a given detector/surface combination and N_i is the number of alphas found for that particular detector surface.

4.3.5 Background Model: Neutron events

Our models for neutron backgrounds are similar to the γ-sourced model detailed above in that calibration data (^{252}Cf in this case) was used as the base of the model dataset and as a result needed to be corrected to the WIMP-search data. The main difference is that instead of correcting the ^{252}Cf data to some unblind sideband, we use Geant4 Simulation [81] of our expected neutron backgrounds. This is because all large and distinct populations of ^{252}Cf-sourced neutron events will be inside our blinded region, rendering the NR-sideband-rescaling approach used in the γ model useless. Our neutron model actually consists of two separate models, one for cosmogenic and one for radiographic neutrons. The methodology for their construction is identical save for the value of the normalization factor.

4.3.5.1 Systematic Neutron correction

The output from the simulations gives an estimate for the uncorrected single scatter neutron spectrum, as well at the total rate of these interactions in our array. The basic approach for creating the background model is fairly simple. The underlying assumption is that the spatial distribution of the cosmogenic neutrons matches that for the ^{252}Cf-sourced neutrons (which is highly likely). Given this, the only important reweighing required is in terms of recoil energy. The weight vector, $\vec{w^i}$, is constructed as

$$\vec{w^i} = \epsilon^i \frac{N^i_{sim}}{N^i_{Cf}} \tag{4.20}$$

were for energy bin i, ϵ is the analysis efficiency, N_{sim} is the number of simulated neutron events, and N_{cf} is the number of calibration events. Due to the very low statistics of the cosmogenic simulation, constructing a good spectrum is difficult. To deal with this, we assume that the energy spectrum is identical to that of the radiographic neutrons and reuse the un-normalized weight vector from the radiogenic simulation for both models.

4.3.5.2 Neutron normalization

After formation of the uncorrected weight vector, we again normalize the sum of the weight vectors to equal the number of expected neutron events. For the radiogenic model that is 0.1094 ± 0.024, and for the cosmogenic it is 0.024 ± 0.024 events.

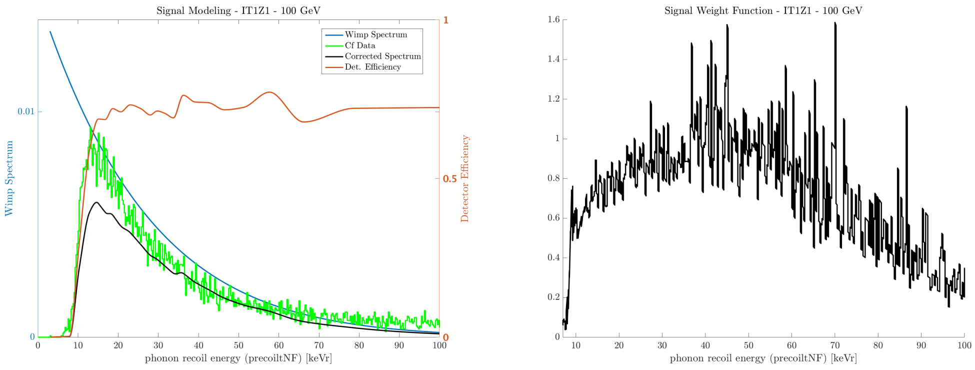
4.3.6 Signal Model Dataset

4.3.6.1 Systematic signal correction

Creating a weight vector RRQ for our signal model is very similar to the background models described above, with a few extra complications. The calibration dataset is treated in much the same was as in the γ model, just with NR-band ^{252}Cf instead of sideband ^{133}Ba data. With the construction of X_{calib} out of the way, what do we do about X_{WIMP}? With the gamma model we used sideband data to act as an empirical sample from our WIMP search gamma background. There is no corollary empirical dataset for a WIMP signal (as the existence of such a set would be a discovery of WIMP dark matter). Instead we use the theoretical WIMP energy spectrum together with our understanding of our detector's energy dependent efficiency to predict what our signal distribution would look like if we were to measure WIMPs in our detector. Clearly this approach only works for RRQs that we have some theoretical WIMP model for, which is currently limited to energy. So, in this case :

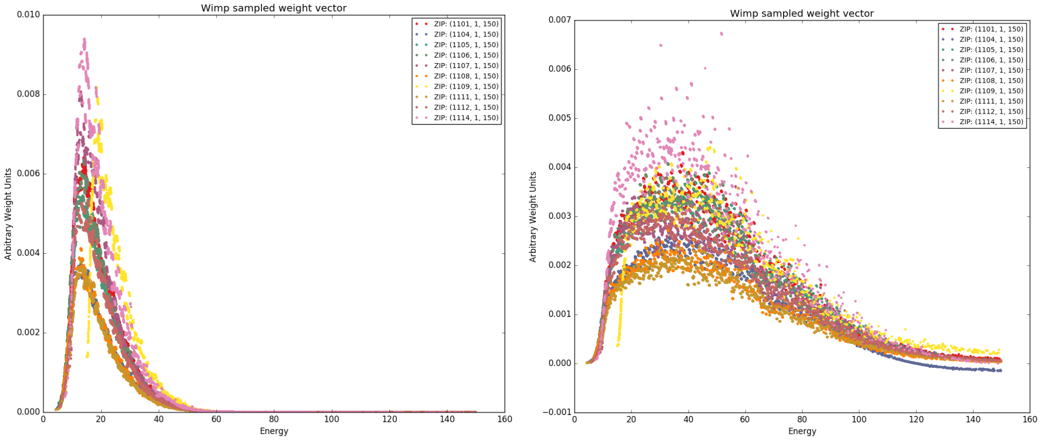
$$E_w(\vec{E}) = \frac{E_{WIMP}(\vec{E})}{E_{calib}(\vec{E})} = \frac{\epsilon(\vec{E})\frac{dR}{dE}}{E_{calib}(\vec{E})} \tag{4.21}$$

where $\epsilon(\vec{E})$ is the energy-dependent analysis efficiency and $\frac{dR}{dE}$ is the differential recoil spectrum for a WIMP. It should be noted that the shape of $\frac{dR}{dE}$ depends on the wimp mass, which is currently unknown. As a result, we will need to compute a different wimp model weight distribution for every potential WIMP mass of interest. For this analysis we calculated our signal model for masses of 10, 25, 50, 100, 250, and 500 GeV/c^2. This process can be see in figure 4.7. After constructing E_w, we can again re-sample the function with the ^{252}Cf recoil energy RRQ to recover the weight vector $\vec{E_w}$. Figure 4.8 shows scatter plots of the weight vector vs recoil energy for two different WIMP masses, and represent our expected sensitivity to WIMP-interactions as a function of energy.



(a) (b)

Figure 4.7: Construction of the weight distribution for our signal model. (a) illustrates the construction of the energy dependent WIMP sensitivity for detector iT1Z1. It is for a theoretical WIMP mass of 100 Gev. The blue line is a theoretical WIMP spectrum ($\frac{dR}{dE}$), the orange is the energy-dependent analysis efficiency ($\epsilon(\vec{E})$) for iT1Z1. The resulting corrected WIMP spectrum is shown in black. ^{252}Cf-sourced calibration is shown as the green histogram. The resulting weight distribution is shown in (b).



(a) (b)

Figure 4.8: Scatter plot of the WIMP model weight vector, $E_w(\vec{E})$ vs recoil energy for a theoretical WIMP-mass of 25 (a) and 250 (b) GeV/c^2.

4.3.6.2 Signal normalization

Unlike for our background models, we can't simply normalize our signal models to an expected number of events. Even for a given wimp mass, the event rate would depend on the interaction cross-section between the WIMP and the Ge in our detector. Instead of modeling the number of WIMP interactions, we instead model our sensitivity to WIMP interactions. This sensitivity depends on our total exposure, our detector's efficiency, a theoretical WIMP recoil spectrum and has historically been referred to as Spectrum Average Exposure or SAE. It is defined:

$$SAE = MT\frac{\int_{E_{min}}^{E_{max}} dE\epsilon(E)\frac{dR}{dE}}{\int_{E_{min}}^{E_{max}} dE\frac{dR}{dE}} \tag{4.22}$$

Where $\epsilon(E)$ is the energy-dependent analysis efficiency, $\frac{dR}{dE}$ is the differential recoil spectrum for a WIMP, and MT is the total post-live-time-cut exposure. Just like with our unnormalized weight vector the SAE depends on the WIMP mass, and we have used the same six masses as before to calculate the normalization factor. It should be noted that after normalization our signal weigh vector has units of $\frac{kg\ day}{event}$ and sums to the total SAE.

4.4 Parametric fiducialization Methods

4.4.1 Parameters of interest

As has been addressed in section 2.4, iZIPs can make a number of measurements during a particle interaction that allow it to discriminate between signal and background. Any and all of these are potentially useful as fiducialization parameters. There are seven quantities we use in order to define our final fiducial volume. Our goal, however is to be able describe each event in terms of a single parameter, that best describes how "signal-like" or "background-like" that particular interaction is. First, though I will explicitly delineate the parameters that are used to construct this "multivariate fiducilazing parameter".

`precoiltNF`

Given that our blinding region, and thus our final signal region, is restricted to the 3-σNR band, at this juncture, the recoil energy is one of our most powerful discriminating parameters. Our expected WIMP spectra are exponential at low energy, and taken together with our analysis efficiency, most WIMP interactions are expected in the lower end of our recoil energy range (see figure 4.8 from the last section). Our backgrounds are much more uniform in energy, causing the relative expected signal to background to vary quite a bit with energy.

`ytNF`

This is the classic CDMS discriminator as described in section 2.4.2. Although we are already restricting ourselves to the region that is both inside the 3-σNR band and below the bottom of the 3-σER band, it is possible that this will need further restriction to deal with our surface event population that has suppressed ionization yield.

`qsummaxOF`

This is our best estimate for the collected charge energy, and is a component of the ionization yield in that `qsummaxOF = ytNF` × `precoiltNF`. As can be seen this variable is not independent from the previous two, but it is important to note that our chosen machine learning method (see [decision trees]) can only inspect a single variable at a time and can only cut rectilinearly. This means that from the point of view of the decision tree the information provided by `qsummaxOF` is not redundant, but potentially useful for removing events with very low ionization collection.

`qzpartOF`

This is our main Z-direction fiducializing parameter.[4] It allows for the removal of near-face events such as ^{210}Pb-sourced β's and recoiling ^{206}Pb nuclei.

`qrpart_zhalf`

This is our main radial fiducializing parameter. The main source of potential backgrounds in our experiment comes from high-radius photon interactions, and is one of the most important parameters.

`pzpartOF`

This is included for similar reasons as `qzpartOF`. It does not preform as well as ionization-based parameters at near-surface position resolution, but there are two regimes where it adds additional useful information. The first is in the bulk of

---

[4] Fiducializing in the traditional sense that it helps to remove the poorly functioning surface layer of our target mass.

the crystal. Events that occur in the bulk will have essentially symmetric collection, leaving a very large proportion of events in the noise blob around a `qzpartOF` value of 0. For these events `pzpartOF` provides better Z-position information. Similarly at low recoil energies, the ionization noise is substantially worse than the phonon noise, rendering `pzpartOF` more useful.

`prpartOF`

In this same way that `pzpartOF` is the phonon analogue for `qzpartOF`, `prpartOF` is the phonon analogue of `qrpart_zhalf`. It similarly has improved radial-position information for bulk interactions, and those at lower recoil energies.

4.4.2 Dimensionality Reduction

As we have seen, our goal is to systematically change a single parameter and observe how that restricts or relaxes our final signal region in some way. This ability to order a proposed set of fiducializing cut positions is important for a few reasons. First it allows human analyzers to manually explore and enumerate these cuts by hand, both to build an intuition as to how various choices of cut behave as well as for understanding the results of our optimizer. As was previously mentioned, construction of a background model in more than a single dimension quickly runs into problems with the sparse nature of our data, especially the γ-model dataset. This is even more of a problem once we move to optimizing the cut locations. With a single fiducializing parameter per detector, our SAE, which we are trying to maximize, and our misidentified background, which is serving as a constraint on that maximization are both ten dimensional. In the most general case (assuming rectilinear cuts in all dimensions), using all seven discriminating parameters described above, we would have to maximize a seventy dimensional function subject to a seventy dimensional constraint. Even assuming we had the incredibly high statistics in our model datasets required to construct these functions, actually preforming the optimization would be practically intractable. To solve this we turn to a machine learning method called a “decision tree”.

4.4.3 Decision trees

An example of such a dimensionality reducer is a decision tree (DT). These are widely used in high energy particle physics in part because they are easy to construct and simple to inspect and understand. For a much more in-depth description of the theory of decision trees, please see Appendix B in [33]. In the language of the machine learning community decision trees are said to be a supervised classifier. Classifiers take a number of attributes or "input features" of a thing and attempt to assign it to one of a finite number of categories or classes. DT's are also supervised learners, which means that they need to be shown many example elements of each class in order to train themselves. In our case we have two classes: signal and background. Our DT uses the seven discriminating parameters of interest described in section 4.4.1 as input features. The output of our DT is a "score" or number in $[-1, 1]$. A very signal-like event would have a score of -1, a very background-like event would have a score of 1 and an event that has an equal probability of being signal and background would be scored at 0. In this sense, our DT would act as our dimensionality reducer: it takes the seven parameters that each have some ability to separate signal and background and produces a single score which is ideally better than any one of the input features at discriminating between signal and background. Although a complete description of Decision trees is outside the scope of this thesis a very brief outline of the method will be given below.

4.4.3.1 Single Tree overview

Decision Trees are simple enough that it is conceivable to construct one by hand provided enough time. This is because decision trees are simple collections of single variable, rectilinear splittings or decisions. As discussed above our tree is a classifier with two classes K one for signal K_S and one for background K_B. Given N samples of background N_B and signal N_S model events as training data, the tree trains itself as follows. For each of the seven inputs discusses in section 4.4.1 the BDT proposes an initial splitting cut to separate background and signal. At each proposed splitting it can calculate the probability that an event on each side of the split is signal or a background event by examining the ratio of events in the training like:

$$P_K = \frac{N_K}{N} \tag{4.23}$$

where the subscript K can represent signal S or background B. Using this we can define a metric of the impurity (called the *Gini impurity*) of each of these samples as:

$$f_{Gini} = \sum_{K}^{K\in\{S,\,B\}} P_k(1 - P_k) = 1 - \sum_{K}^{K\in\{S,\,B\}} P_K^2 \tag{4.24}$$

This impurity is calculated for the total sample f_{Gini}^{Total} before the proposed splitting as well as for the population on each side of the spiting (which we will call the Left and Right). Finally the information gain ΔG is calculated for each of these proposed splittings where

$$\Delta G = N f_{Gini}^{Total} - N^{Left} f_{Gini}^{Left} - N^{Right} f_{Gini}^{Right} \tag{4.25}$$

The splitting that produces the largest information gain is chosen as the first node in our decision tree. The procedure can then be recursively repeated on the Left and Right sub-populations.

The recursion will continue until the sub-populations are all pure signal or pure background. The problem with this is that the tree will be fitting individual events from our models, and will not preform well when scoring events that were not used in the training dataset (these reserved events are usually referred to as “testing events”). To prevent this over-training we can artificially limit the growth of the tree in a process called “pruning”. Two of these pruning methods were used in this analysis. The first is to limit the maximum recursion depth of the tree and is called the “maximum tree depth”. The second is to limit the minim number of events in a sample (as a percentage of the entire sample) that the DT will preform a split on and is called the “minimum leaf size”. To differentiate these DT-tuning parameters from the normal input parameters of the DT they are usually referred to as “metaparameters”. There are two complications which arise from this pruning. First, we need a way to choose good values for these metaparameters. Second, because the DT can only preform rectilinear cuts, a single pruned tree will typically produce decision boundaries which are very blocky and course. We will address both in the next section.

4.4.3.2 Ensembles and metaparameter tuning

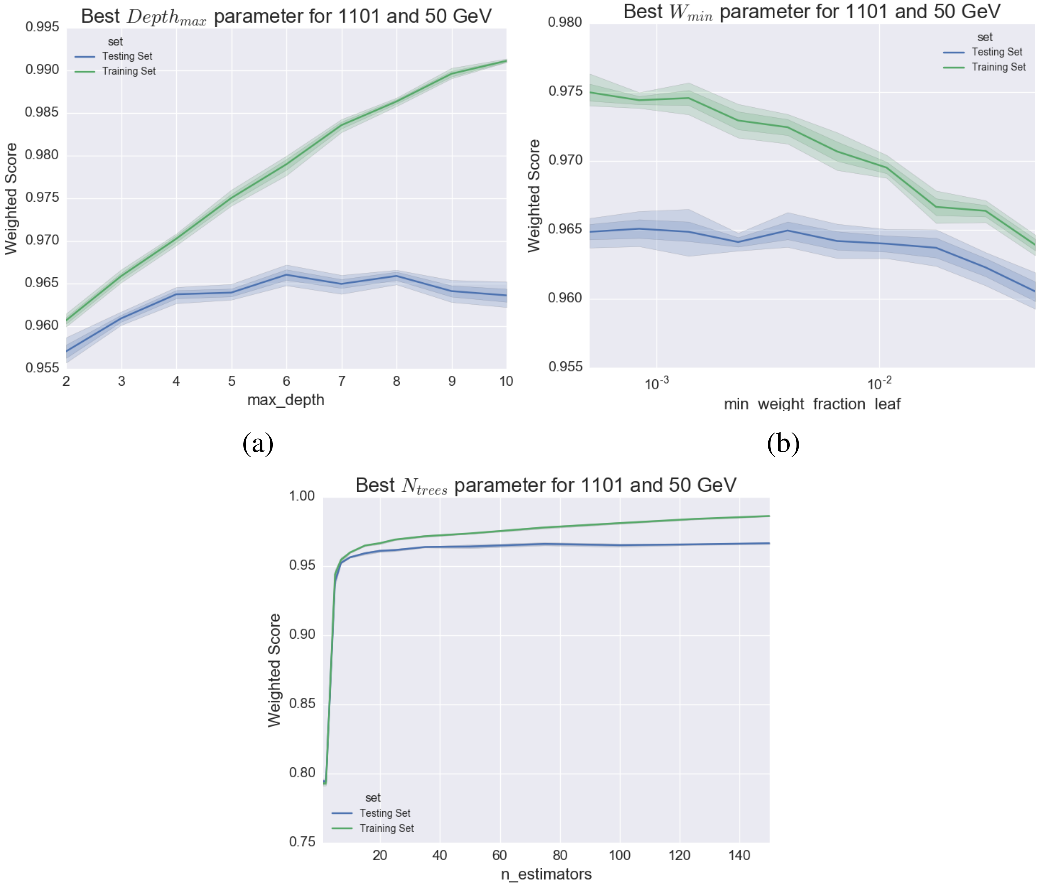
As mentioned in the previous section, our pruned tree will not preform very well on its own. Our signal and background populations will probably be most effectively separated using some kind of a smooth hyper-surface in the 7D input feature space, but the DT will only be able to produce a few coarse rectilinear blocks. To overcome this we construct an ensemble of trees that are grown under slightly different conditions from one another. Each of these trees can "vote" on the classification of an event, and that vote is averaged into a final score. There are many methods that are used to grow this ensemble, but the one used here is type of boosting called "gradient boosting". As a result, we typically refer to our machine learning method as a "boosted decision tree" or BDT. In essence, the first tree of our ensemble is grown as described in the previous section. The next tree is trained slightly differently. All training events that the first tree misclassified are re-weighted to be more important (essentially counting as more than a single event) as the next tree is grown. This allows the next tree to "try harder" to classify them correctly. The procedure is repeated (with the third three weighted to "try harder" than the second) to construct the entire ensemble.

This leaves us with tuning our metaparameters: the maximum tree depth, the minimum leaf size, and the number of trees in our ensemble. This was done by repeatedly spiting our data into training and testing sets, and measuring the classification performance of the resulting BDT at a variety of metaparameter combinations. The result of this is shown in figure 4.10. The values that were settled on were max depth: 5, min leaf weight: 1%, number of trees: 50.

4.4.4 Bootstrapping

4.4.4.1 Overview

At this point we have a specific machine learning method (defined as a gradient-boosted decision tree ensemble with its three metaparameters fixed at the values set in the previous section) that we believe will preform well at optimally using all of our discriminating parameters to separate our potential signal and misidentified background. What we need to do now is actually train a BDT, save it, and use it to score all events we could potentially be interested in. To be unambiguous, I will use the singular *a BDT* to refer to a 50-tree gradient-boosted ensemble described in



(a) (b)

(c)

Figure 4.9: Accuracy of our BDT at classifying an event at various metaparameter values. For our two pruning methods (a) and (b) we would like to be in a region where small changes in the metaparameter vale do not strongly effect the resulting accuracy. For the ensemble size (c) more should always be better, but eventually memory usage becomes a constraint [82].

section 4.4.3.2. In this context there arises a slight chicken-and-egg problem. We need to use some subset of our signal and background model datasets to train our BDT. However we would like to be able to produce scores for all events from all datasets. This includes the subsets of our calibration and WIMP search data used to train the BDT itself. This has the potential to introduce bias in the scoring of those particular events. After all, as previously discussed, a dataset independent of the training set is required to understand our BDT's behavior. There is a related problem, and that is one of low statistics. Our background models, especially the one describing our expected γ-sourced interactions, have a number of very rare, but

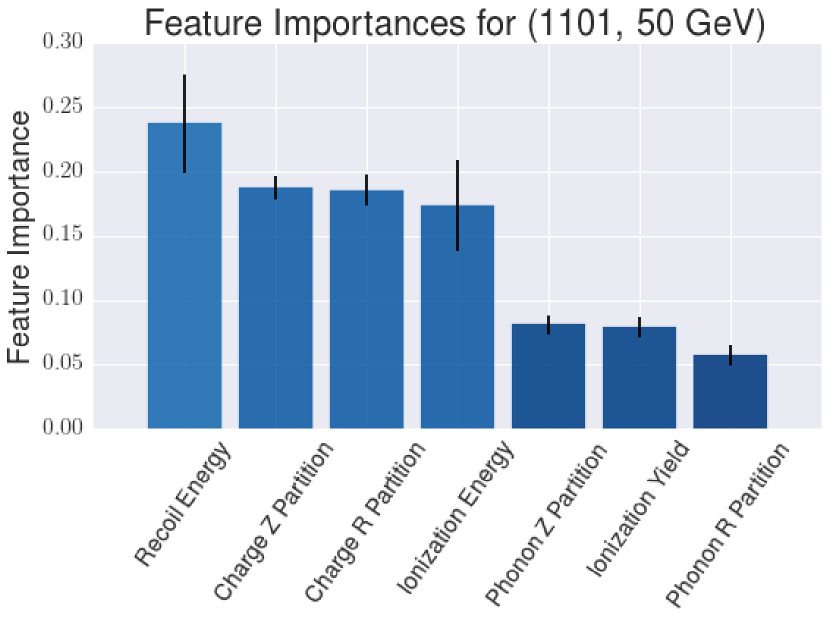


Figure 4.10: The relative number of times a particular input feature is chosen for a splitting, or the "feature importance" is plotted for a single BDT trained on the 50 GeV/C^2 model datasets for detector IT1Z1. This example has the recoil energy of the events as the most important discriminating parameter followed by the ionization position estimators. Given that this model is entirely inside of the NR-band, this is consistent with our expectations, and acts as a nice cross-check of our BDT behavior.

very important events[5]. As a result, any subset of data that we pick has the potential to miss a number of our very rare, but very important, low radius γ events. As with the solution to most chicken-and-egg problems the solution to this is referred to as bootstrapping[6]. Our requirements for this process are pretty simple: we want to utilize all of our data to train our method, but ideally when scoring a particular event we use a BDT that was not trained on this particular event. The process also needs to be deterministic. A single interaction[7] should always produce the same BDT score regardless of the context in which it is used—as a part of the testing set used to characterize the trees during their training, as a part of a signal or background modeling dataset, or event as just a normal calibration or WIMP-

---

[5] Important in the sense that they have a high weight, and are an important contribution to the total expected background

[6] To start a computer the CPU needs to load code from disk. But to know exactly where to go to get this code, and how to execute it once it it loaded requires some control logic or code. Which is, of course, located on disk. The process devised to solve this problem is also referred to as bootstrapping or simply booting.

[7] Defined by the unique combination of `EventNumber` and `SeriesNumber`

search event used by a third party, so while randomness is going to be an important component, it needs to be tracked. Finally, as a matter of bookkeeping it would be ideal if our final fiducializing parameter produced by this method ranged from [0, 1], with 0 being most signal-like and 1 most background-like. This is for consistency with our definition of `qrpart_zhalf`; ideally we can re-use the same optimization machinery for both parameters. This process is divided into two parts. *Training*, which uses our model datasets to bootstrap an ensemble of BDT's that are saved for later, and *scoring* which uses the saved BTD's to produce a BDT score for any event in the SuperCDMS dataset.

4.4.4.2 Training

The basic idea with bootstrapping is this: instead of training a single BDT on a random subset of our model datasets and using the remainder of data to test that BDT's performance, repeat this process many times, constructing a number of BDT's. Then average the results. The basic training procedure is as follows

1. For each BDT in our bootstrapping ensemble, construct a boolean mask where each entry is set as `true` or `false` with a probability of 0.5. This mask is the same size as our entire model dataset and is indexed by a unique event ID consisting of `EventNumber` and `SeriesNumber`. The mask is then saved to disk in `hdf5` format in such a way as to be unambiguous which BDT it is associated with.

2. Using this mask, our model datasets are split into testing (a mask vale of `false`) and training datasets (a mask value of `true`), which are used to train the BDT and test its performance.

3. For each individual BDT the minimum and maximum scores achieved from scoring the testing set are saved.

4. Finally, the BDT's themselves are serialized to disk. It should be noted the the native `Python` serialization methods found in the standard library's `pickle` and `cpickle` were as of this analysis incapable of serializing our decision tree objects. For this we utilized Michael McKerns' excellent scientific serialization library `dill` described here [83].

The above process is repeated 25 times. This number was selected as it ensures the probability that an event is not selected to be part of at least one BDT's training set is vanishingly small. In practice it was required that this be the case, and all events were used to train at least on of the 25 BDT's.

4.4.4.3 Scoring

The above training procedure may seem needlessly complex, but all components are required to construct a repeatable, unbiased BDT score. The scoring for a particular interaction, ε, is done as follows:

1. Load each of the 25 trees, masks, and bounds into memory.
2. For each tree, t, look up the mask value associated with `EventNumber`$_\varepsilon$ and `SeriesNumber`$_\varepsilon$. If the value is `true`, the score is set to zero and we skip to step 4.
3. If the mask value is `false` or the event is not found in the mask (indicating it was not a part of one of our model datasets) the event is given a raw score of s_{raw}. This score is than normalized using the max and min bounds found from step 3 in the training as

$$s_{norm} = \frac{s_{raw} - b^t_{min}}{b^t_{max} - b^t_{min}} \tag{4.26}$$

4. The resulting scores from events that are actually run through the BDT (and not manually set to zero) are then averaged to produce our final multivariate fiducializing parameter.

4.4.5 Multivariate results

To understand our new multivariate parameter's performance we can score our entire signal and background model datasets, the results of which can be seen in Fig.4.11. In closing the discussion of our BDT is it important to re-iterate the context in which it is being used. From Fig.4.11 it would appear we can read off the relative probability of an event with any particular BDT score as being a signal event or a background event. Assuming our background and signal model datasets

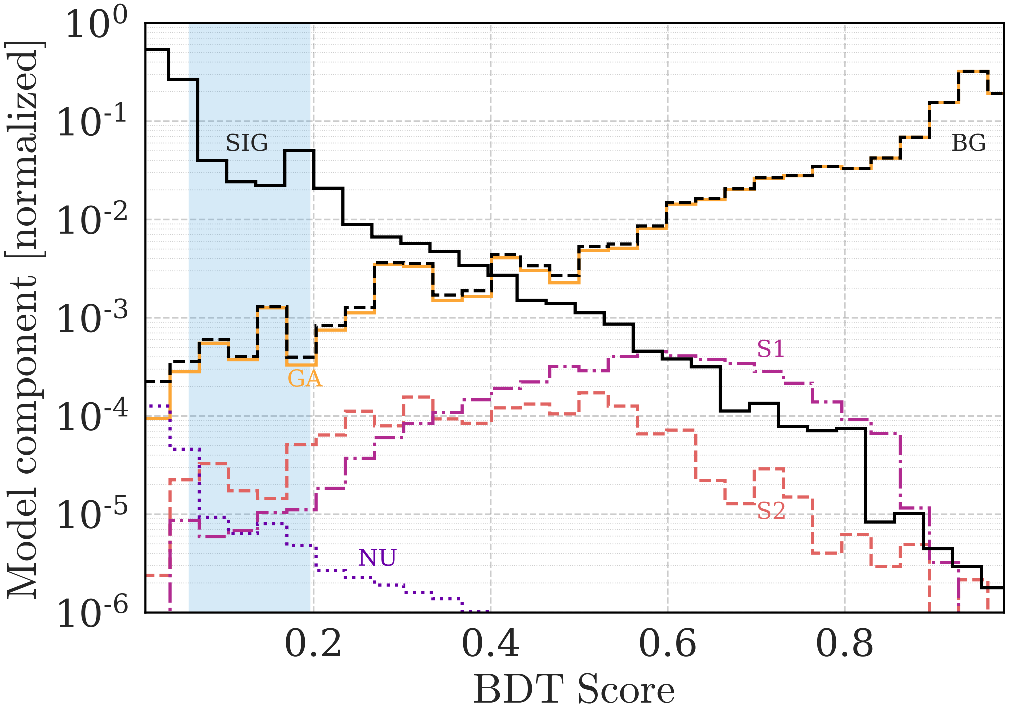


Figure 4.11: Histograms comparing background (BG: black dashed) and 50 GeV/c^2 signal models (SIG: black solid) in BDT score for the preselected events, summed over all detectors. The background model has been subdivided into its constituent components, which are, from darkest to lightest: radiogenic and cosmogenic neutrons (NU: purple dotted), upper surface ^{210}Pb chain (S1: fuchsia dot-dashed), lower surface ^{210}Pb chain (S2: coral dashed), and gammas (GA: orange). Both the signal and total-background model histograms have been normalized to unit integral for ease of comparison. The BDT cut for a given detector accepts events with BDT score between 0 and a detector-specific cut value. The range of these cut values over the ten detectors is shown (blue band).

are perfect, this would be the case. We will not, however, use the BDT ensemble to score the blinded WIMP-search events in the `bg_restricted` dataset. Instead, we are simply using the BDT as a dimensionality reducer. In this sense the only events that will be scored are unblinded WIMP-search and calibration events that we use as a part of our modeling datasets[8]. As a result, the bootstrapped-BDT score can be thought as just another fiducializing parameter that we will set our final cuts in while we are still blind. In this way any errors in the modeling or BDT will alter how well set our final signal acceptance region, but they will not bias, or change the

---

[8] After we unblind and the analysis is finished, we will score any events that appear in our signal region for a cross check, but no result will depend on accuracy of the BDT itself.

correctness, of our result.

4.5 Optimizing fiducialization

4.5.1 Cut setting and Optimization overview

Ideally we want to find fiducial cut positions that allow us to set the strongest limit. What we have direct access to, however, is total spectrum averaged exposure (SAE) and misidentified background ("leakage") as a function of cut position. We suspect that the optimal leakage will be less than one event, so we can simply pick a leakage (say 0.3 events) and maximize the SAE while constraining the leakage to be that value. We do this for a number of different leakages, starting at 0.02 events, ending at 0.98 events with steps of 0.02 events. For each chosen leakage constraint we: 1.) Define objective and constraint equations, 2.) preform a constrained optimization to find the cuts that have the highest SAE at the chosen leakage and 3.) set both a Poisson and an Optimal Interval 90% C.L. upper limit.

4.5.2 Objective and constraint

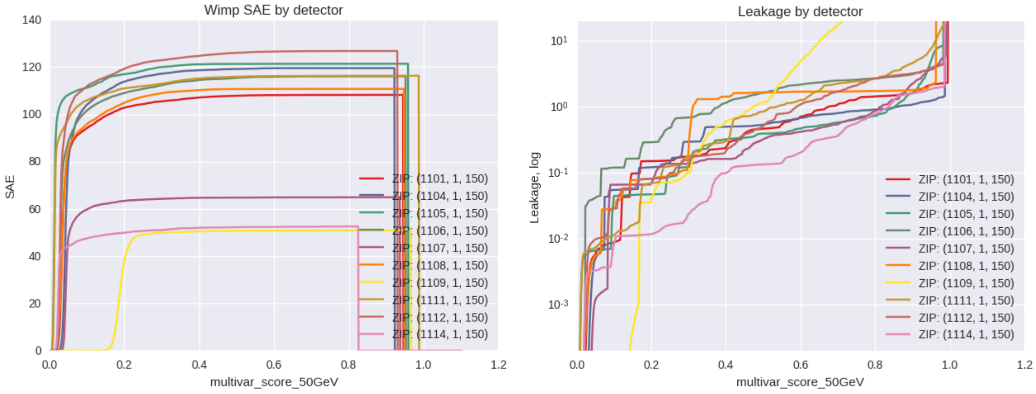
Our objective function $\mathrm{sae}(x_1, \ldots, x_n)$, is simply the total SAE as a function of the fiducializing cut position, x_i, in detector i. This SAE function is just the sum of the individual SAE's from each detector:

$$\mathrm{sae}(x_1, \ldots, x_n) = \sum_{i=1}^{n} \mathrm{sae}_i(x_i) \tag{4.27}$$

Where $\mathrm{sae}_i(x_i)$ is the detector dependent SAE. In its simplest form it can be found my taking a cumulative sum of the weight vector up to the suppled value of cut position in BDT-score space for detector i, x_i:

$$\mathrm{sae}(x_i) = \sum_{s}^{\forall s < x_i} w_s \tag{4.28}$$

Where the sum ranges over all model dataset events with BDT-score s that is less than x_i. This is simply an empirical CDF of the distribution that the weight vector is attempting to model. Our constraint function $leak(x_1, \ldots, x_n)$ would similarly be the sum of the the empirical CDFs of our background model and constructed in a completely analogous fashion. The detector-dependent components of these two functions can be seen in figure 4.12.



(a) (b)

Figure 4.12: Detector-dependent components of the objective and constraint function. (a) is a plot of the individual SAE functions from each of our detectors ($\mathrm{sae}_i(x_i)$) as a function of BDT score. (b) is the corresponding set of individual leakage functions ($\mathrm{leak}_i(x_i)$) plotted logarithmically vs BDT score.

A problem with the above naive construction is that it does not produce very smooth functions. The signal model has quite robust statistics even at very low BDT score, as does the neutron model. The problem is with the model for γ-sourced events. Although the absolute magnitude of our expected γ-sourced backgrounds is much higher than neutron sourced backgrounds, the number of events in the model dataset is substantially lower. This leads to trouble in our constraint function, where an otherwise smooth neutron distribution is punctuated by giant steps caused by these heavily-weighted γ-sourced events. A number of smoothing approaches were tested including weighted linear and cubic spline interpolations, as well as a fixed-windowed averaging scheme, but a 50 point nearest-neighbor averaging was what was finally settled on. While it has minor variation from the previously described naive cumulative sum, it provides a constraint function with gradients well-defined enough for our optimizer to make sense of, and is was is presented in figure 4.12.

4.5.3 Cut Optimization

The thing we are maximizing is SAE (at a given constrained leakage).

$$\underset{x}{\text{minimize}} \qquad \text{sae}(x_1, \ldots, x_n) \tag{4.29}$$

$$\text{subject to} \qquad \text{leak}(x_1, \ldots, x_n) = \textit{allowed leakage} \tag{4.30}$$

We do this for values of *allowed leakage* starting at 0.02 events and ending at 0.98 events with a step size of 0.02. This implies that for a single proposed example of sae and leak this optimization will need to be preformed about 50 times. As mentioned above there are a number of reasonable choices for construction of sae and leak. They could be simple functions of `qrpart_zhalf`, or some more complex multi parameter score. A particular ensemble of bootstrapped, gradient-boosted decision trees was settled on, but each proposed fiducialization technique needed to be run through this optimizer at least once to get a sense of its viability. This is further complicated by the fact that for a given choice of fiducializing parameter, five different versions of sae must be constructed and tested, one for each of the five proposed WIMP-masses used to construct our signal models. Two important requirements come of this. First, it is important that our optimization routine converge quickly, or barring that, that its convergence can be sensibly tracked and understood in some way. Second, the maximum found does not need to be exact. It is entirely possible that a number of cut positions will result in a total SAE that is very close to the global maximum achievable. Given the uncertainties inherent in our modeling, any of these choices would be an appropriate definition of our final signal-acceptance region. As a result, it is far more important to find a result close to the global maximum quickly than it is to find the "true" global maximum.

4.5.3.1 Brute Force

As this analysis is similar in many respects to the original CDMS II analysis, this was the first place we turned for optimization technique. Following the methods outlined in **??** a target function f was defined such that

$$f(x_1, \ldots, x_{10}) = \left(1 - \frac{\text{sae}(x_1, \ldots, x_{10})}{SAE_{tot}}\right)\left(1 - \frac{\text{leak}(x_1, \ldots, x_{10})}{\textit{allowed leakage}}\right) \tag{4.31}$$

where SAE_{tot} is total SAE before applying any final fiducialization cut. This target was then minimized by MATLAB's `patternsearch`, which essentially takes an initial point, p_0 and steps a small distance, ϵ, in each coordinate direction. This produces twenty proposed points, $p_0 \pm \epsilon \vec{x_1}, \ldots, p_0 \pm \epsilon \vec{x_{10}}$, at each one of which the function f is evaluated. If the value of f is smaller at one or more of the proposed points the smalled is chosen as the initial point for the next iteration and the value of ϵ is increased, typically by a factor of 2. If none of the proposed points produce an f that is smaller than that of the initial point, the value of ϵ is decreased and the proposed points re-calculated. This can be done to any depth of iterations desired until a minimum is found. The idea here is that by minimizing this simple target the first term in equation 4.31 will seek to maximize the SAE, while the second will satisfy the constraint on our allowed leakage. There are two related problems with this approach. This first is that it is easy to be trapped in a local minimum and not explore the entire parameter space. The second is that if this happens, it is very possible that the second term in equation 4.31 will not be zero, and our constraint condition will not be met. In practice this second problem, together with the very slow speed of this approach, rendered it intractable.

4.5.3.2 SLSQP

Rather than rely on finding a particular global optimum to satisfy our constraint we next turned to manually enforcing the constraint itself. To do this we turned to method of Lagrange multipliers. As a reminder we are trying to solve

$$\underset{x}{\text{minimize}} \qquad \text{sae}(\vec{x}) \tag{4.32}$$

$$\text{subject to} \qquad \text{leak}(\vec{x}) - \textit{allowed leakage} = l(\vec{x}) = 0 \tag{4.33}$$

The Lagrangian for this problem would be

$$\mathcal{L}(\vec{x}, \vec{\lambda}) = sae(\vec{x}) - \lambda l(\vec{x}) \tag{4.34}$$

Following the derivation in [84] stationary points of the above Lagrangian can be found numerically by iteratively solving the associated subproblem at iteration point x_i

$$\begin{aligned} \underset{d}{\text{minimize}} \quad & \nabla\mathcal{L}(\vec{x}_i, \vec{\lambda}) \cdot \vec{d}_i + \frac{1}{2}\vec{d}_i^T H\mathcal{L}(\vec{x}_i, \vec{\lambda}) d_i^T && (4.35) \\ \text{subject to} \quad & l(\vec{x}_i) + \nabla l(\vec{x}_i) \cdot \vec{d}_i = 0 && (4.36) \end{aligned}$$

where the search direction $\vec{d}_i(\vec{x}) = \vec{x} - \vec{x}_i$, and H is the Hessian operator. Solving this type of problem is much simpler than the previous one. Traditionally this type of optimization is referred to as quadratic programming, and is a well established sub-filed of numerical optimization. As an example, provided $H\mathcal{L}(\vec{x}_i, \vec{\lambda})$ is positive definite it reduces to solving.

$$\begin{bmatrix} H\mathcal{L}(\vec{x}_i, \vec{\lambda}) & (\nabla l(\vec{x}_i))^T \\ \nabla l(\vec{x}_i) & 0 \end{bmatrix} \begin{bmatrix} \vec{d}_i \\ \vec{\lambda} \end{bmatrix} = \begin{bmatrix} -\nabla\mathcal{L}(\vec{x}_i, \vec{\lambda}) \\ -l(\vec{x}_i) \end{bmatrix} \tag{4.37}$$

There are numerous high-quality implementations of this sequential quadratic programming optimization method and the one used in this analysis is from the SciPy project [85] in `scipy.optimize.minimize(method="SLSQP")`. This method has the benefits of reliably enforcing our constraint conditions and converging to a local optimum fairly quickly[9]. The downside here is that this method finds local maxima, and the global performance is very sensitive to the initial proposed point. We expect that the maximum SAE as a function of constrained leakage to be increasing and convex[10], and as can be seen in figure 4.16 this particular optimization method by itself is only accurate in a minority of constraint conditions. Although a useful tool, to find maxima that are close to the global value we need a more effective way of exploring our parameter space. For this we turned to Monte Carlo.

4.5.3.3 Simulated Annealing and Basin hopping

For a global method of exploring our entire parameter space we turned to the stalwart of Bayesian statistical inference, Markov-Chain Monte Carlo (MCMC). In place of maximizing a likelihood function in the presence of a prior, we are attempting to maximize (or at very least explore) our *sae* function. We enforce our constraint by using a square prior that is defined to be 0 unless the leakage value is

---

[9] On the order of a CPU minute.

[10] Think a curve of diminishing returns.

within 0.002 misidentified background events or

$$\text{prior}(x) = \begin{cases} 1 & \textit{allowed leakage} - 0.002 < x < \textit{allowed leakage} \\ 0 & \textit{else} \end{cases} \tag{4.38}$$

and our posterior function is our narrowly constrained objective. In its equilibrium state our MCMC will produce points x_m with a probability proportional to $\text{sae}(x_m)$. Although there are many variations of the exact algorithm the basic outline is as follows

1. Choose a proposal distribution, k, from which to draw new points. For simplicity we will restrict ourselves to k that are symmetric.
2. Choose an initial point in the parameter space x_0, and then for each iteration, i, at point x_i
3. Generate a proposed point x_{prop} by sampling from k
4. Calculate the acceptance ratio $\alpha = \text{prior}(x_{prop})\,\text{sae}(x_{prop})/(\text{sae}(x_i)\,\text{prior}(x_i))$
5. Generate a uniform random number r on [0,1].
6. If $r < \alpha$, or $\alpha > 1$, accept the proposed point and let $x_{i+1} = x_{prop}$. Otherwise do nothing and let $x_{i+1} = x_i$.

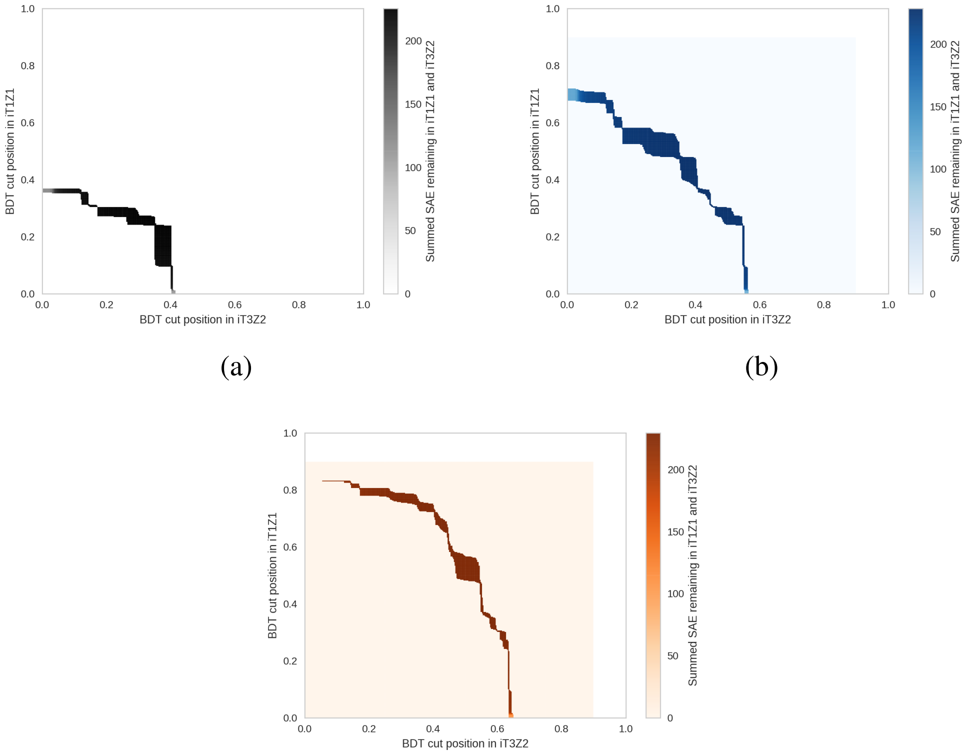
The problem, as we will see, is finding a good choice of k. If we let k be a 10-dimensional Gaussian centered at x_i we would have a kind of sampling referred to as Metropolis-Hastings which preforms a random walk around the parameter space. An example of how this can be used to preform a global maximization by slightly modifying the above as follows:

1. Assuming our proposal distribution is $\mathcal{N}(x_i, \sigma^2)$, follow the above algorithm until equilibrium has been reached.
2. At equilibrium, define a temperature parameter, T. Initialize it at some large value and decrement, or "cool" it by some fixed amount every iteration.
3. After drawing r, alter the acceptance criteria as follows: If $rT_i/T_0 < \alpha$, or $\alpha > 1$, accept the proposed point and let $x_{i+1} = x_{prop}$. Otherwise do nothing and let $x_{i+1} = x_i$.

As the iterations progresses the temperature cools and the probability of making a downhill transition goes to 0. This acts to trap the Markov Chain in a particular local maximum. It is stochastic, and will vary run-to-run, but as the density of samples is proportional to the size of a particular local peak, this method can produce very good (near global max) maxima. Due to its similarity to annealing in metallurgy, this process is referred to as "simulated annealing". To speed the process up, between each step a fast gradient-based optimization[11] may be preformed so that every MC step starts at a local maximum. Additionally, the distance between the randomly chosen MCMC step and the subsequent local maximum can be recorded. This can, especially after a few iterations that end at the same peak, inform about the size of the local peak (or basin in the case of a minimization). To find a new peak, the step size (given by σ in our above example) can be adjusted accordingly. This technique allows for many local maxima to be found, and is referred to as "basin hopping". There are a number of well-tested implementations of both approaches available to `MATLAB` (`GlobalSearch`, `simulannealbnd`) and `Python`(`scipy.optimize.basinhopping`) programming environments. None of these methods, or our hand-implemented versions of them, were able to reliably converge on a near-global minimum, or at least not on a timescale necessary for this analysis given our computing constraints. In each case, the difficulty arose from an inability to choose a proposal distribution k that would allow the MCMC to efficiently explore the space. The problem is that our constraint condition creates an objective that is highly anisotropic. To build an intuition, figure 4.13 plots a very simplified 2 dimensional example of our objective using the actual SAE and leakage from two of our detectors, at a variety of constraint leakages. As can be seen, step sizes on the order of the size of the space to be explored will often fall outside of the constrained area, leading to high rates of rejection and very slow movement of our sampler. Step sizes of a length scale necessary to stay in the constrained volume are, however, far smaller than the overall size of the volume, leading to very slow relaxation times. To solve this we turned to a recent approach for sampling such distributions proposed by Goodman and Weare [86].

---

[11] Such as the SLSQP method mentioned above.



(a) (b)

(c)

Figure 4.13: Three examples (at varying constraints) of a simplified 2D version of our target function, which consists of our SAE objective subject to a narrow prior that constrains it around a particular leakage value. As can be seen, the space we would like to sample is highly anisotropic, and efficiently generating new sampling points is difficult. The full 10D target space suffers from a much more extreme version of this problem.

4.5.3.4 An affine-invariant MCMC

A simple example of a distribution that suffers from this type of slow exploration is a highly-anisotropic 2-dimensional distribution of the form

$$p(x, y) \propto f(-\frac{(x-y)^2}{2\epsilon} - \frac{(x+y)^2}{2}) \tag{4.39}$$

where f is a univariate function. In the limit where ϵ is small this suffers from the same problem described above. Generating proposal samples efficiently is very difficult. Due to the angle of the anisotropy generating samples for each coordinate independently, such as in Gibbs sampling, will never efficiently explore the space. It could be possible in this example to find a general two-dimensional Gaussian with the right aspect ratio to explore the space effectively, but as the number of dimensions,N, increases the number of tuning parameters for the proposal distribution scales as N^2, making this very difficult in our case. We could, however, preform the following affine transformation[12]

$$x' = \frac{x - y}{\sqrt{\epsilon}} \quad y' = x + y \tag{4.40}$$

which transforms our problem to a much simpler isotropic one. Ideally, we would like to construct an MCMC sampler that is invariant under this class of transformations. In this way, even a poorly formed target function can be efficiency sampled from provided it can be made more isotropic by an affine transformation[13]. In practice no known single MCMC algorithm can achieve this, but it can be done with an *ensemble* of Markov chains [86]. The basic approach is to construct an ensemble, $\boldsymbol{X}$, consisting of L Markov chains, or *walkers*, $\boldsymbol{x}^k$. The intuition here is that the ensemble X can be exploited to generate proposed samples that are more consistent with the target function. The general procedure is as follows.

1. Choose a symmetric univariate distribution g
2. To propose a new position, $\boldsymbol{x}^k_{prop}$ for walker $\boldsymbol{x}^k$, first draw a random number z_k from g.
3. Randomly select a second walker $\boldsymbol{x}^j$ where $j \neq k$.
4. Construct the new proposed position via $\boldsymbol{x}^k_{prop} = \boldsymbol{x}^j - z_k[\boldsymbol{x}^k - \boldsymbol{x}^j]$
5. Accept or reject the new proposal in an identical fashion to what was described above.
6. Repeat the above for all L walkers. This constitutes a single step.

Goodman suggests a proposal distribution of the form

$$g(z) \propto \begin{cases} \frac{1}{\sqrt{z}} & \text{for } z \in [\frac{1}{a}, a] \\ 0 & \text{otherwise} \end{cases} \tag{4.41}$$

---

[12] An affine transformation is simply a liner transformation followed by a displacement and takes the form $\boldsymbol{x}' = A\boldsymbol{x} + \boldsymbol{b}$.

[13] It is important to note that this regularizing transformation does not ever need to be know, it just needs to exist. If a sampler is truly affine-invariant, it will produce identical samples regardless of how the underlying space is transformed.

where a is a tuning parameter.

For our optimization purposes we implemented a custom basin hopping technique as follows. For each value of allowed leakage

1. Initialize a starting point by allowing each detector to have a fraction of that leakage based on its total sensitivity to signal or for detector i,

   $$x_0^i = \text{arcleak}^i \left(allowed\ leakage \times \frac{SAE_i}{SAE_{tot}} \right) \tag{4.42}$$

   where arcleak^i is the inverse of the leakage function leak^i for detector i. This was implemented via brute-force interpolation.

2. Run `SLSQP` constrained maximization to find the nearest local peak. The results of this stage can be seen in figure 4.16.

3. Initialize 200 walkers by sampling their initial positions from an isotropic 10-dimensional Gaussian with $\sigma = 0.002$ centered at the local peak. Our number of walkers was set by memory constraints.

4. Allow the walkers to explore the parameter space for 2,000 steps. Our rule of thumb to set this number was that it need to be about ten relaxation times, which in some dimensions was roughly 200 steps as can be seen in figure 4.14.

5. At this point use the current position of each walker as the initial position for another `SLSQP` constrained maximization for each of them. This gives us a large set of local peaks.

6. Find the highest of these peaks and restart the entire procedure. In practice only a few iterations of this procedure [14] were required to produce vastly improved optimal points as can be seen in figure 4.16.

The exact implementation of the affine invariant sampling used in step 3 is described in [87] and provided as the `Python` package `emcee`. This was far more preformant than our hand-coded version.

---

[14] Tn our case our iteration depth was three.

4.5.3.5 Optimizer Results

To find the optimal cut positions at all constrained leakage values and five WIMP-signal models, in the end requires about a CPU year of computing time. This was carried out on the SUF cluster. A good deal of care was taken to exploit all available parallelism, and the actual running time of a series of such optimizations was a little over a week. The final results of the optimization can be seen in figure 4.15 and 4.16.

4.5.4 Limit Setting

At this point our optimizer has given the cut positions (for a given *allowed leakage*) that maximized the SAE of our experiment. We still do not know which of these *allowed leakage*, SAE combinations produce the best limit. To do this we turn to the Optimal Interval method **??**, which will be used during our final limit setting procedure. A full description of this method is beyond the scope of this dissertation, but in this context it is used to set a limit on the number of signal interactions with a known energy distribution. Essentially for each observed event with energy E_i, the cumulative signal probability, is calculated[15] which can be calculated for detector i from our signal model like so:

$$P^i_{cum}(E) = \frac{\sum_0^{E_r > E} w_{E_r}}{SAE^i} \tag{4.43}$$

Where w_{E_r} is the weight calculated for signal model event at a `precoiltNF` value of E_r. Following the procedure outlined by Yellin in [88] we can combine all ten of our detectors into a single large detector by preforming a weighted averaging[16] of our ten signal models into a single spectra. For each possible interval between pairs of observed events, or between an observed event and bounds of our energy domain, the number of contained events and the total cumulative signal probability are calculated. For each interval the 90% confidence level upper limit on the number of expected WIMP interactions is calculated by comparison a lookup table that was pre-computed from Monte Carlo. This *optimal* interval provides the lowest estimate on the number of WIMP interactions and, after taking the appropriate statistical penalty for the freedom to choose such an interval, allows the strongest limit to be

---

[15] By cumulative we mean total probability a WIMP interaction below the interaction energy E_i

[16] Weighted by the fraction of total SAE provided by a given detector or $SAE^i/SAEtot$.

set.

Seeing as the optimal interval method requires actual events with actual energies to produce a limit, we turn again to Monte Carlo. The procedure is:

1. For each constrained leakage we have the cut positions from the previous section. Apply them.
2. Calculate expected number of misidentified background events in each detector, μ_i, by summing the remaining background model weight vectors, as well as their expected energy spectrum by taking a simple weighted histogram.
3. Preform 5,000 MC experiments by sampling the correct number of events in each detector from a Poissonion with a mean of μ_i.
4. For each sampled event, sample an interaction energy from the appropriate background energy spectrum as calculated in step 2.
5. For each MC experiment calculate the cumulative signal probability for all interactions, and preform the Optimal Interval calculation for the 90% C.L. upper limit on the expected number of WIMP interactions.
6. Average the resulting upper limits and divide by the SAE to produce a WIMP rate per unit exposure. Because we know the theoretical flux of a proposed wimp with this mass through our detectors, we can calculate the 90% C.L. upper limit on the WIMP-nucleon cross section. The results of this process can be seen in figure 4.17.
7. Repeat for all ~50 constrained leakage values and all 5 signal model datasets.

Although conceptually fairly straightforward, some care was taken to make this computationally tractable. In particular the standard optimal interval `Fortran` code required some refactoring to allow it to be compiled as a shared library. In this way the `UpperLim` function found in `UpperLimNew2.f` can be called directly with no context-switching overhead. The build process required to achieve this is extremely fragile, and we were unsuccessful at automating it on all systems. For this reason we have not released this publicly as an improvement to the code.

4.5.4.1 Cut selection

We will be using the Optimal Interval method to set our final limit, but before we do so we need to choose a final set of fiducializing cuts. This means we need to choose a mass to optimize around as well as a leakage that will give us the best limit. From Figure 4.17 we can see that the cross section largely stops improving for allowed leakages above 0.14 events. There is very little incentive to optimize for higher leakage: although the optimal interval will still potentially allow us a good limit, we would expect more leaked background events. To be conservative, we decided to choose the tightest cut that did not sacrifice sensitivity. To this end the Poisson minimum was selected.

Having chosen a constrained leakage value of 0.14 events, we still have five sets of optimized cuts to choose from, one for each of the five signal model masses. Ideally, we want a final cut position that provides strong limits at a variety of potential WIMP masses, not just the masses that they were optimized for. For this end, we took each of our five optimized cut positions and calculated the 90% C.L. upper limit on the expected WIMP rate for each of our five signal models. The results of this cross check can be seen in figure 4.18. As can be seen each set of cuts preforms the best for the WIMP mass at which it was optimized. This is expected, but is a nice consistency check. Although not the strongest limit at any mass other than 50 GeV c^{-2} the set of cuts optimized on that particular signal model dataset preforms very well at all WIMP masses. As a result the final cut set defining our signal acceptance region is that which was optimized for a constrained leakage of 0.14 misidentified background events assuming a WIMP mass of 50 GeV c^{-2}.

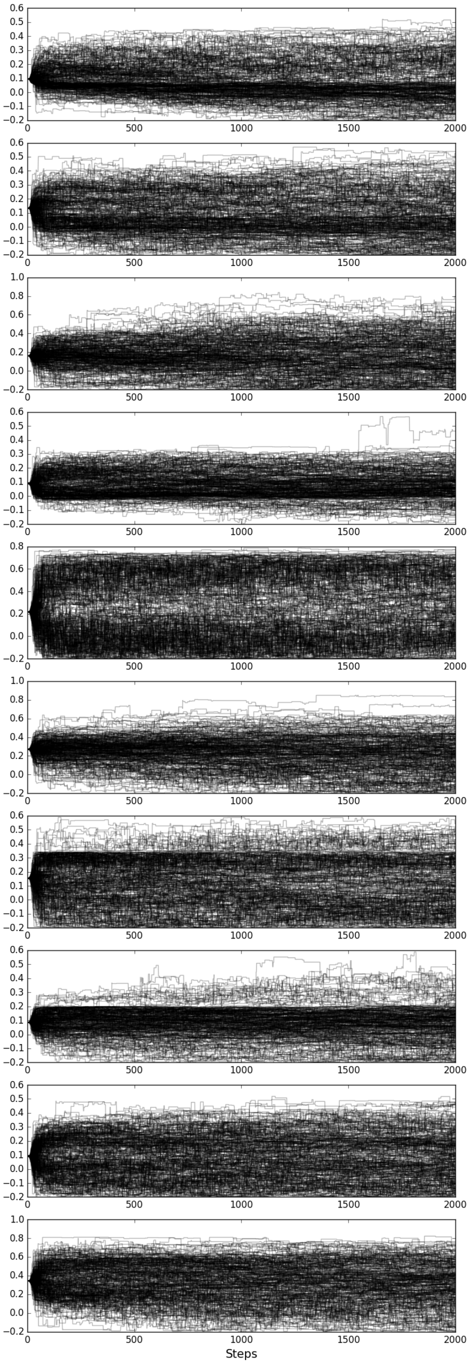
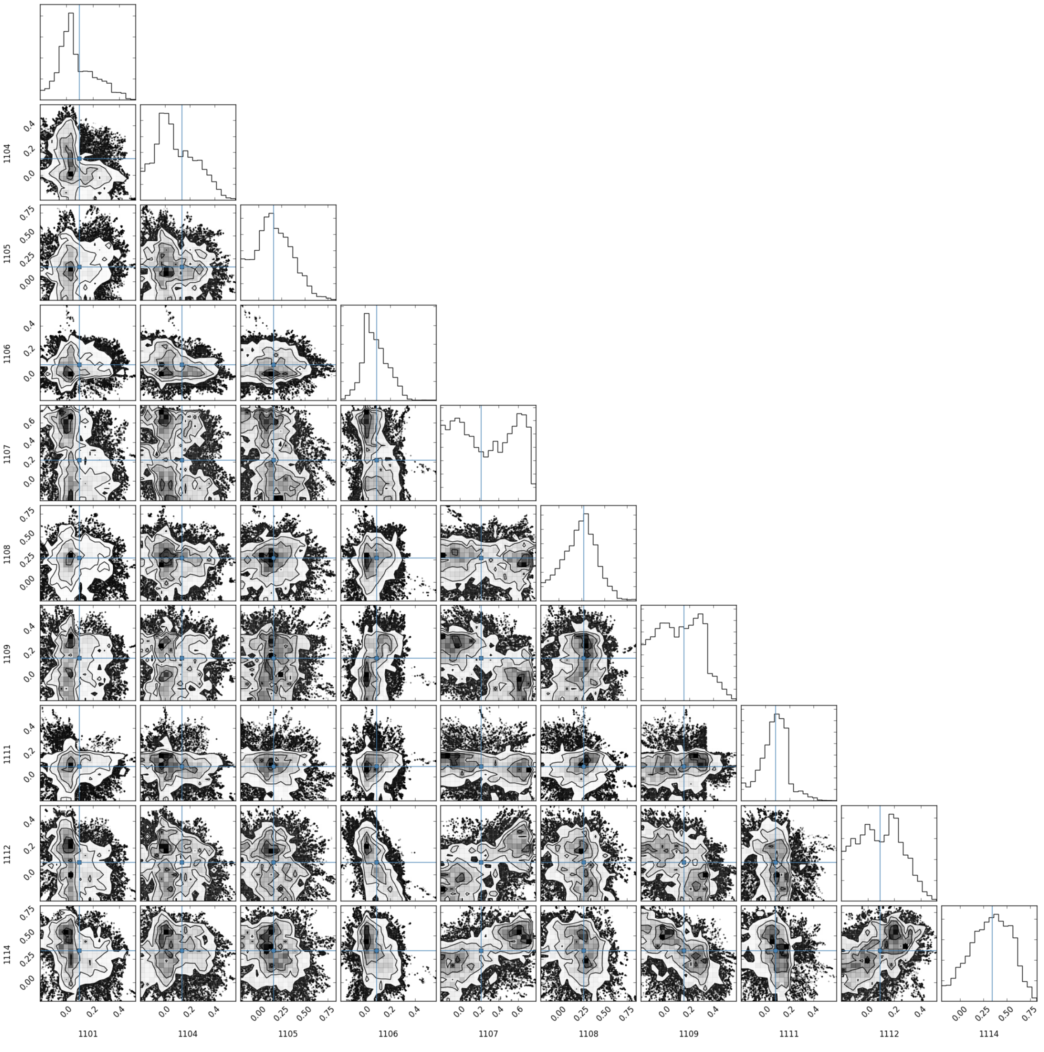
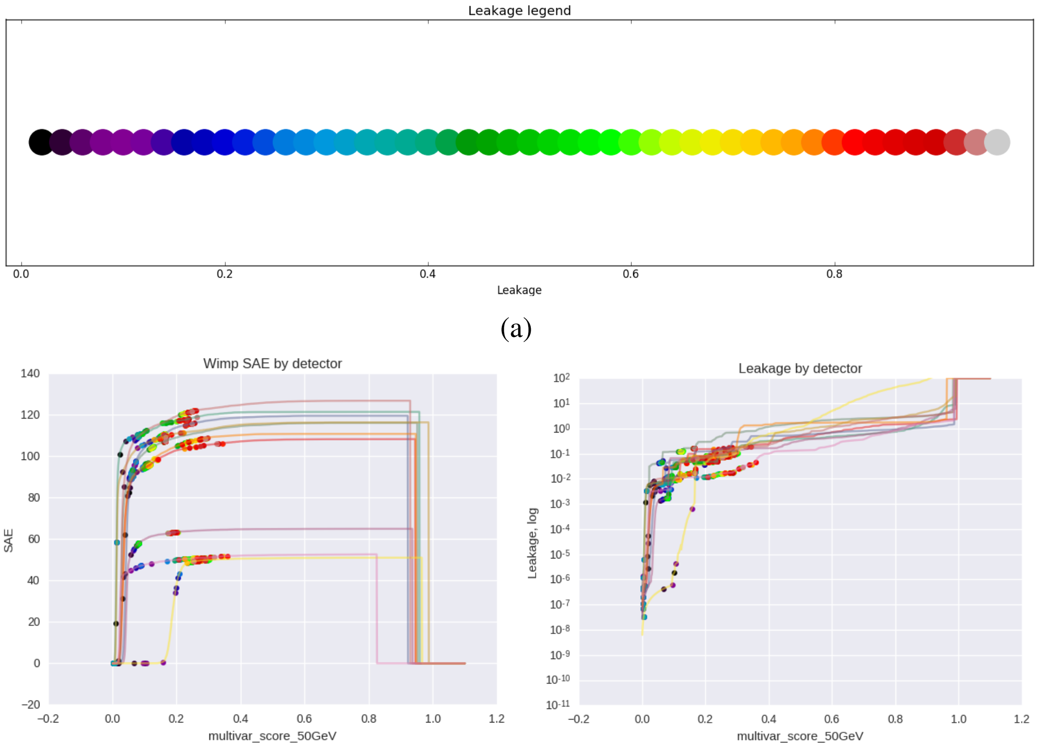


Figure 4.14: Plots of the position in BDT score of our walkers vs step number for each of our detectors. As can be seen it takes ~200 steps for the BDT score to be fully explored in all detectors (Mass 50 GeV, leakage is 0.6 events).



(a)

Figure 4.14: Marginal distributions of our sampled posterior for a WIMP with a mass of 50 GeV/c^2, and a constrained leakage of 0.6 events. Due to the angle of the anisotropy it is not visible in any marginal distribution rendering Gibbs sampling inefficient. The initial point found by the SLSQP optimizer is shown (blue).



(a)

(b)

(c)

Figure 4.15: Optimal cut positions are shown for a 50 GeV/c^2 WIMP mass. The constrained leakage value is depicted via the color (a) and the resulting optimal detector-dependent BDT-cut position is shown against the SAE and leakage functions used to preform the optimization.

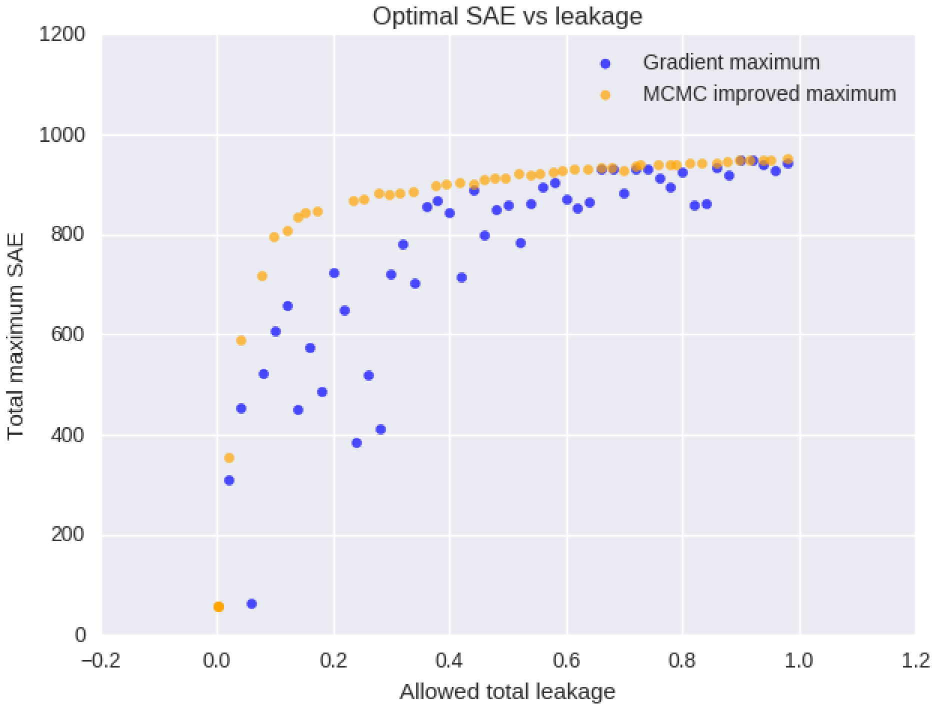


Figure 4.16: Total maximized SAE vs allowed constrained leakage for a 50 GeV/c^2 WIMP. We expect this to be a curve of diminishing returns, and as can be seen the SLSQP-based optimizer (blue) preforms very poorly at a number of constraint conditions. A few iterations of our affine-invariant basinhopping method (orange) vastly improve this result.

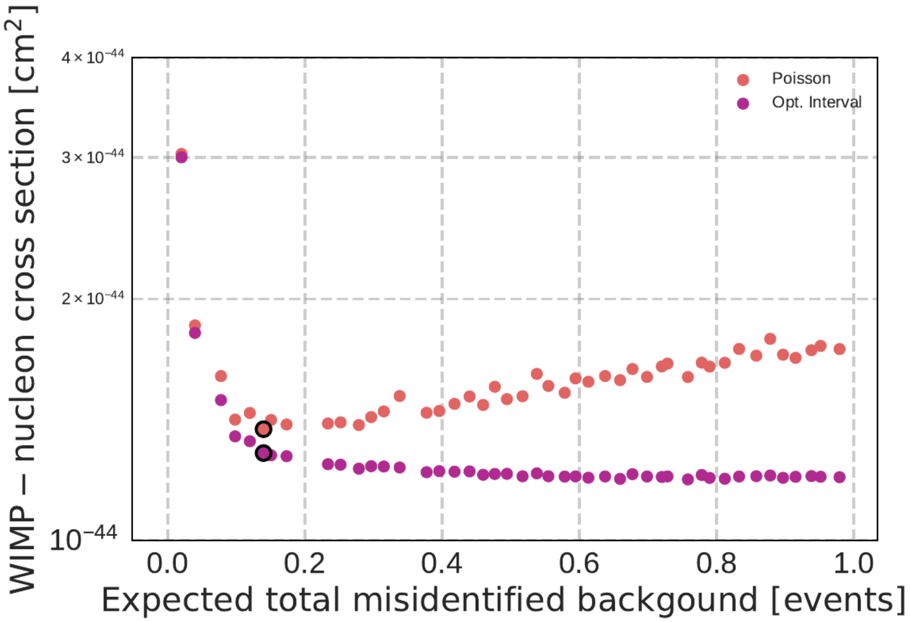


Figure 4.17: The 90% C.L. upper limit on the spin-independent WIMP-nucleon cross section vs allowed leakage calculated using the optimal interval method (fuchsia) as well as using a simple Poisson upper limit (salmon). We used the Poisson minimum to select the cuts to unblind around (black highlighted) which correspond to an allowed leakage of 0.14 events.

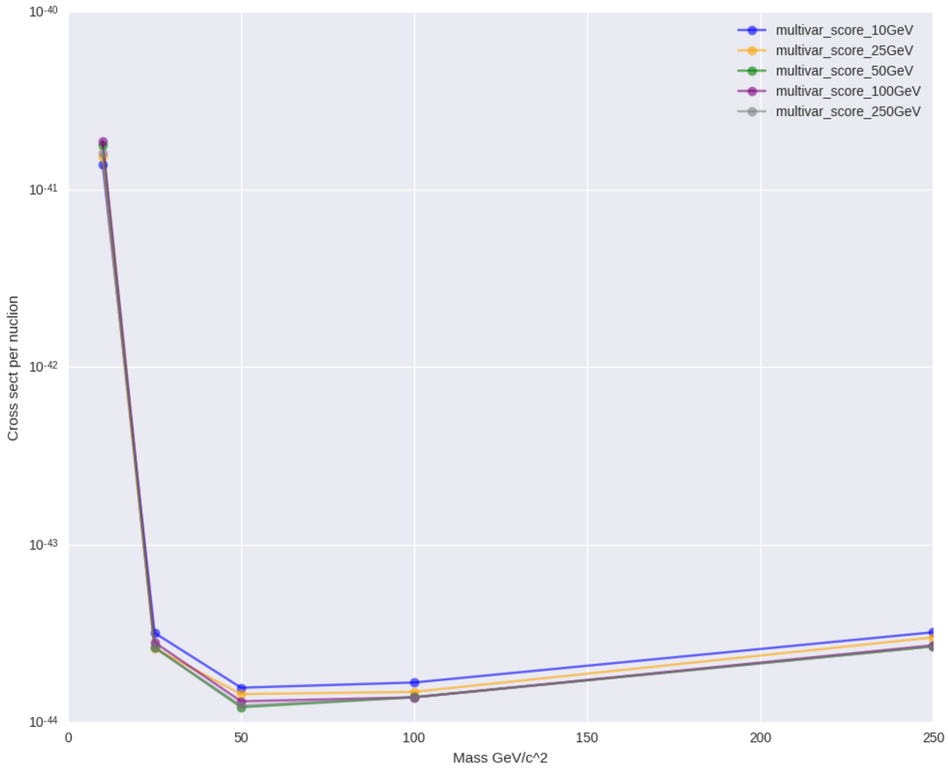


Figure 4.18: Plotted is the expected 90% C.L. upper limit of the spin-independent cross section predicted by Monte Carlo from our signal and background models. The limit was set using cut definitions optimized for a particular WIMP mass (shown by line color). As can be seen each set of cuts cuts preforms the best for the WIMP mass at which it was optimized, but the 50 GeV c^{-2}
model produced a set of cuts that is well-preforming on a large energy domain.

Chapter 5

HIGH THRESHOLD RESULTS

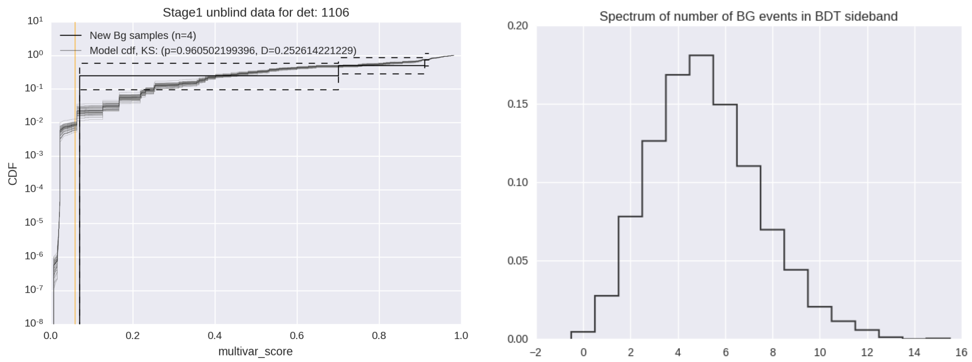
5.1 Stage One (Surface) Unblinding

Now that we have set a signal region with an optimal set of fiducializing cuts, we are theoretically ready to unblind. To allow for further validation of our background models, however, we preformed unblinding in two stages, the first of which only unblinds data outside the BDT cut. This was called “stage one” unblinding. After the stage one unblinding we had access to a new “BDT sideband” consisting of previously blinded, preselected singles-scatter interactions falling inside the 3σ NR-band but outside the newly defined BDT cut. There are two general ways that we can validate our background model[1] against this new data. First, we can check to see if the magnitude of the model (which is normalized to the number of expected background events) is consistent with the number of newly unblinded events observed in the BDT sideband. To do this, we did another Monte Carlo, realizing 5,000 experiments, and drawing background events from our model (propagating all known statistical uncertainties that arose during our model creation). We can then compare the results of this MC to the actual measured number of events in our new BDT sideband. The second way we can validate our model is to compare the CDF of our model in BDT space to the ECDF of the actual unblinded sideband. We also preformed a KS test of the sideband against the model. We have plotted these tests for detector iT3Z4 in figure 5.1.

There are two ways that a detector can be a cause for concern. First, the number of events in the sideband can be higher than expected from the model (say by a few sigma or more). Second, if we form an empirical cumulative distribution function (ECDF) of the actual sideband data we can compare it to the CDFs used in the leak function. If the ECDF of the actual events is systematically higher than the CDF from the model, this would indicate that we have more events at a lower (more signal-like) BDT score than anticipated by the model. There were two detectors in particular that exhibited both of these traits: iT2Z2 and iT4Z2. iT2Z2 has 26 observed events; a few sigma higher than our model predicted. Its ECDF also

---

[1] We can validate the sum of the background models.



(a) The BG model CDF vs BDT Score for iT2Z3

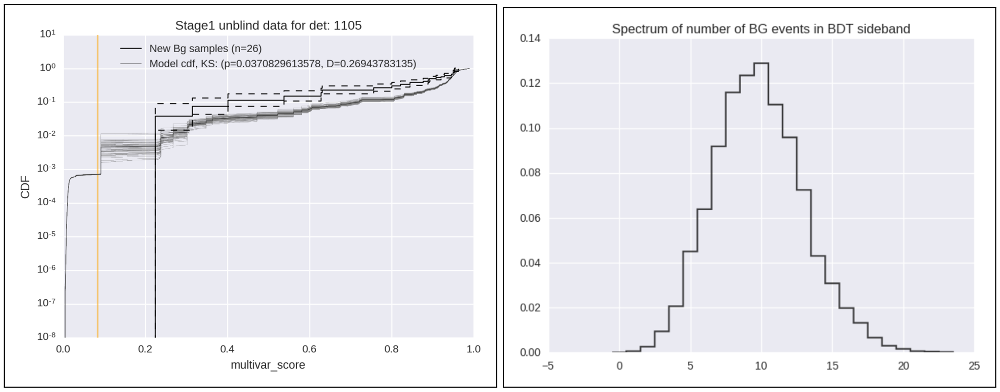
(b) A histogram of expected number of events in the newly unblinded sideband

Figure 5.1: Model validation for IT1Z3. The new BDT sideband for this detector contains $N = 4$ events. (a) plots the ECDF of these 4 events (dark black) against the background models drown from our MC (light gray), which are used to illustrate the uncertainties in our modeling. The BDT cut position is depicted with a vertical orange line. For this detector, the distribution of events in BTD space is highly consistent with our model, at least in the new BDT sideband. (b) is a histogram from the same MC depicting the number of events in the sideband predicted by our model. Four events is again highly consistent.

consistently high at low BDT score as shown in figure. iT4Z2 has 32 observed events; 2 sigma higher than our model predicted. Its ECDF also very high at low BDT score. Both of these effects can be seen for both detectors in figure 5.2 These detectors were of some concern, but ultimately it was decided that in all probability, we would have more sensitivity by keeping them, than by removing them (see the next section for more on this).

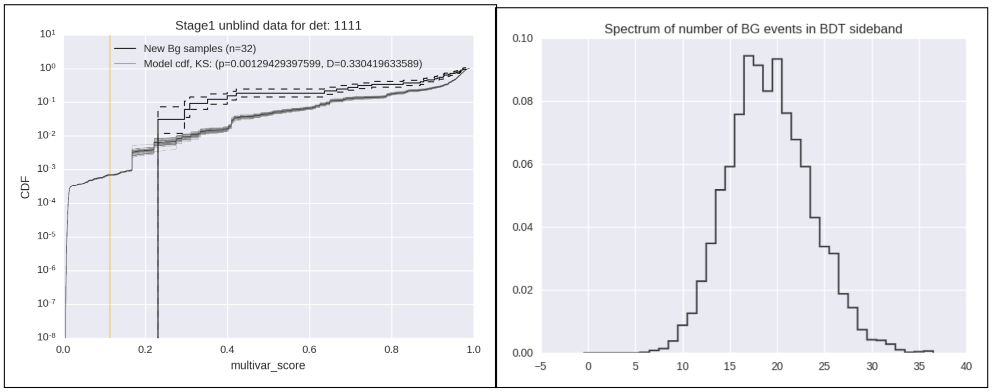
5.2 Sensitivity and Expected Background Re-estimation

Although at this point there was still some disagreement in the shape of the models in BDT space vs the unblinded sideband data, we can gain insight into the systematic discrepancy between the total magnitude of our background model, and what is happening in the sideband via a naive scaling. That is to say, if our model predicted that we would see 4 events in this new sideband, but after unblinding the sideband there were actually 8, we can scale every component of our background model (neutrons, surface, and γ's) by a factor of 2 so that, at least in magnitude, our model



(a) The BG model CDF vs BDT Score for IT2Z2

(b) A histogram of expected number of events in the newly unblinded sideband



(c) The BG model CDF vs BDT Score for iT2Z3

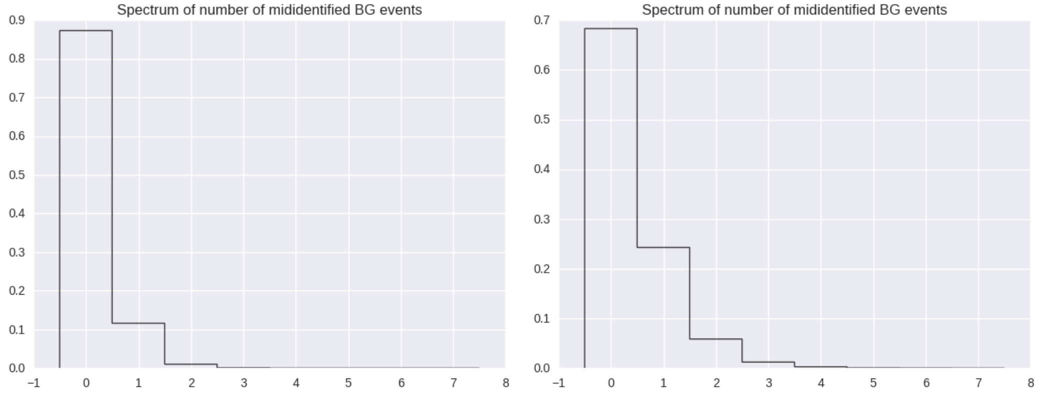
(d) A histogram of expected number of events in the newly unblinded sideband

Figure 5.2: Validation plots for IT2Z2 as well as IT4Z2, in the same style as figure 5.1. (a) and (b): For detector IT2Z2, the ECDF is consistently high (at the 1 to 2σ level), but the larger problem is with the magnitude. The new sideband has $N = 26$ events in it, a value high enough that is was not seen in any of the 5,000 realizations of our experiment, which expected ∼10 events. (c) and (d):

would be consistent with our newly observed sideband. After this simple re-scaling, we can then re-estimate the expected number of events that will appear in our signal region from MC, and compare that to the expected number that we had previously estimated from before the stage 1 unblinding. This is shown in figure 5.3.

The change in expected backgrounds is of less interest than the change in our expected sensitivity. As a result, during the MC in which we rescaled the magnitude of the background models to match the magnitude of the new BDT-sideband we also set an Optimal Interval limit. We then repeated the MC with detector 1111 removed, detector 1105 removed, both 1111 and 1105 removed, and finally with both 1111 and 1105 included but an entire extra background even added to each.

For more detail on this see [89]. Keeping the two troubled detectors, even in the presence of two extra background events added on top of our background model, provided better sensitivity than removing them from the analysis in most MC experiments. For this reason we decided to go ahead with unblinding our entire detector array keeping 1111 and 1105.



(a) Original model.

(b) Re-scaled model.

Figure 5.3: Normalized histograms of the number of misidentified background events in our final signal region predicted via MC using our (a) original background model, and (b) using the same model, but with the magnitude in each detector re-scaled to match the newly unblinded BDT sideband.

5.3 Stage Two (Bulk) Unblinding

On April 11th 2017 we unblinded the data. There was a single event in detector 1105 with a recoil energy of 42.79 keV. For details of the events see table 5.1. The raw pulse data is shown in figure 5.4.

The RQ and RRQ values for this event are sensible and there is no obvious reason to reject this it. The χ^2 value of PBS1 is somewhat high, presumably due to the peakiness of the pulse, however we defined our data quality cut on `ptNF`, and this event passed that cut. To see a visualization of the relative positions of all of our blinded preselected events see figure 5.5.

5.4 Constraints on WIMP parameter space

We only have an estimate of our live time since we only approximately knew the amount of live time removed by the blinding cut. It is not large but must be ac-

Table 5.1: Parameters from the single event in our signal region.

RRQ	Value	RQ	Vaue
PA1OF	7.8506	PAS1OFchisq	4270.3
PB1OF	11.2705	PBS1OFchisq	15044.0
PC1OF	6.4124	PCS1OFchisq	4778.7
PD1OF	5.9619	PDS1OFchisq	4625.4
PA2OF	6.2650	PAS2OFchisq	4247.5
PB2OF	6.2286	PBS2OFchisq	4423.3
PC2OF	7.0141	PCS2OFchisq	4235.1
PD2OF	7.8345	PDS2OFchisq	4327.2
psum1OF	31.455	PTNFchisq	4195.4
psum2OF	27.3422	PTOFchisq	4573.4
pzpartOF	0.0556		
ps1OF	31.142		
ps2OF	27.8603		
precoiltNF	42.79		

counted for. After unblinding we directly calculated the live time of our signal region and found a total raw exposure of 1689.6 kg day. As a reminder this raw exposure is taken after all of the time-period based cuts are applied, but not the event-based cuts. The breakdown can be seen in table 5.2.

Table 5.2: Final exposure accounting

Detector	Mass [g]	LiveTime [d]
IT1Z1	609.5	291.3
IT2Z1	597.4	323.8
IT2Z2	591.3	328.9
IT2Z3	579.1	318.9
IT3Z1	603.4	209.8
IT3Z2	591.3	313.9
IT3Z3	609.5	214.0
IT4Z2	597.4	328.9
IT4Z3	594.3	354.4
IT5Z2	606.5	148.0
Total	5979.7	2831.9

Up to this point we have calculated the analysis efficiencies (mostly) on a set of ^{252}Cf data. Because neutrons are likely to multiple scatter, they have a higher probability of being classified as a shared event relative to WIMPs. This means that generally, the fiducial volume efficiency is different for WIMPs and ^{252}Cf calibration data and we need to derive a correction factor if we are to accurately construct a limit. To do this we produced two datasets using our Detector Monte Carlo (DMC), one of WIMP scatters and one of ^{252}Cf scatters. We then apply the entire set of analysis cuts (including the final BDT based cut), and measure the energy dependent efficiency of each of these two datasets. The true (WIMP) efficiency is then just the efficiency calculated from ^{252}Cf calibration data multiplied by the ratio of efficiencies of DMC WIMPs to DMC ^{252}Cf neutrons. The magnitude of this correction is between 10% and 20%, dominated mainly by multiple-scatters and is shown in figure 5.6. The final corrected efficiency curve can be seen in figure 5.7.

Using this corrected efficiency we can calculate our 90% C.L. upper limit for our WIMP-nucleon cross section as described in section 4.5.4. The resulting limit excludes new parameter space for DM–germanium-nucleus interactions in the mass range 13–127 GeV/c^2. Using standard scalings [76] between nuclei for spin-independent DM-nucleon interactions, limits obtained for other experiments with targets made of other, non-Ge nuclei can be compared and are overlaid in figure 5.8. This result was published in PRL and can be found at [1].

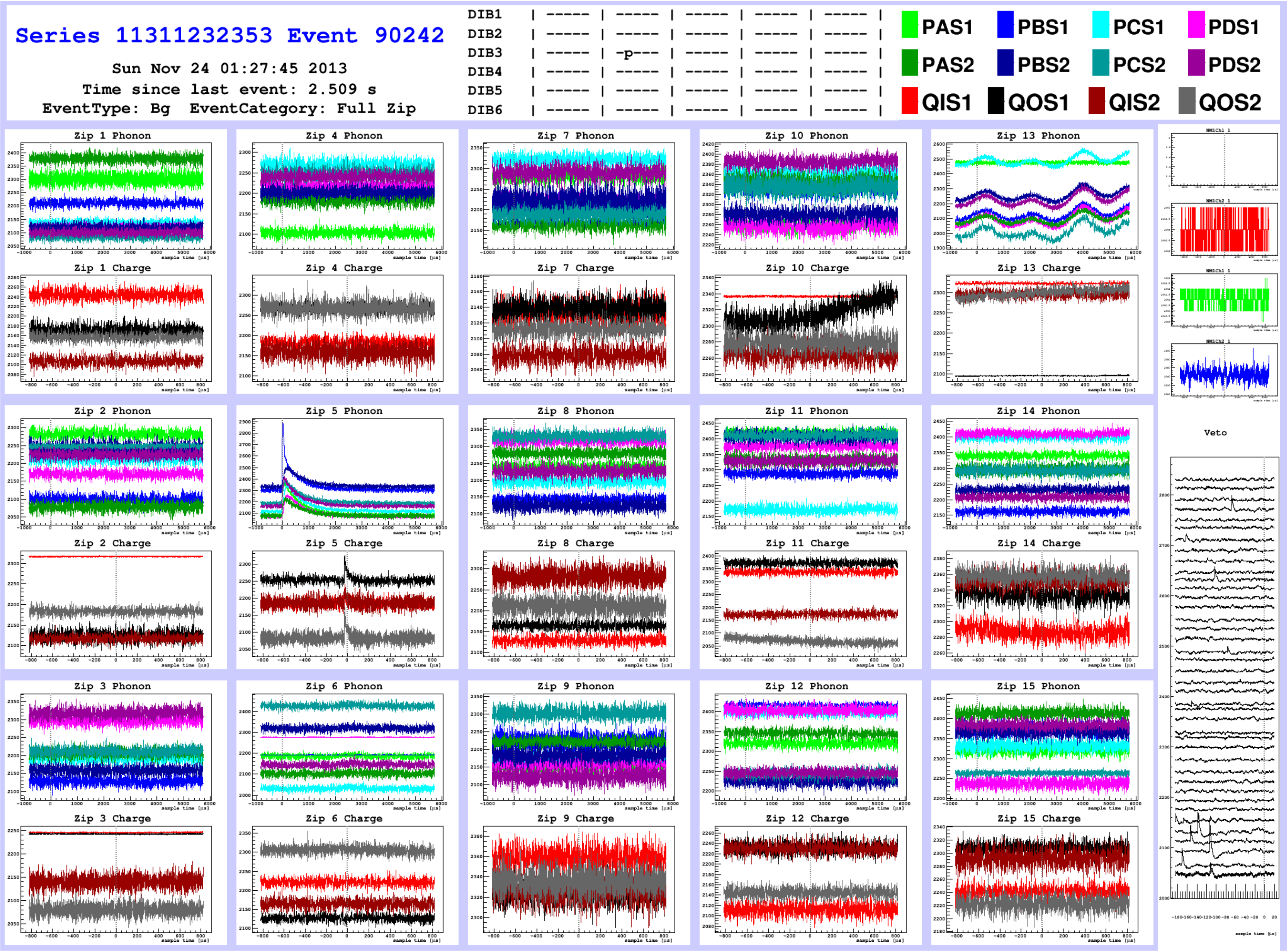


Figure 5.4: Raw event diagnostic data. The only slight abnormality is the peakiness of the phonon pulse in PBS1.

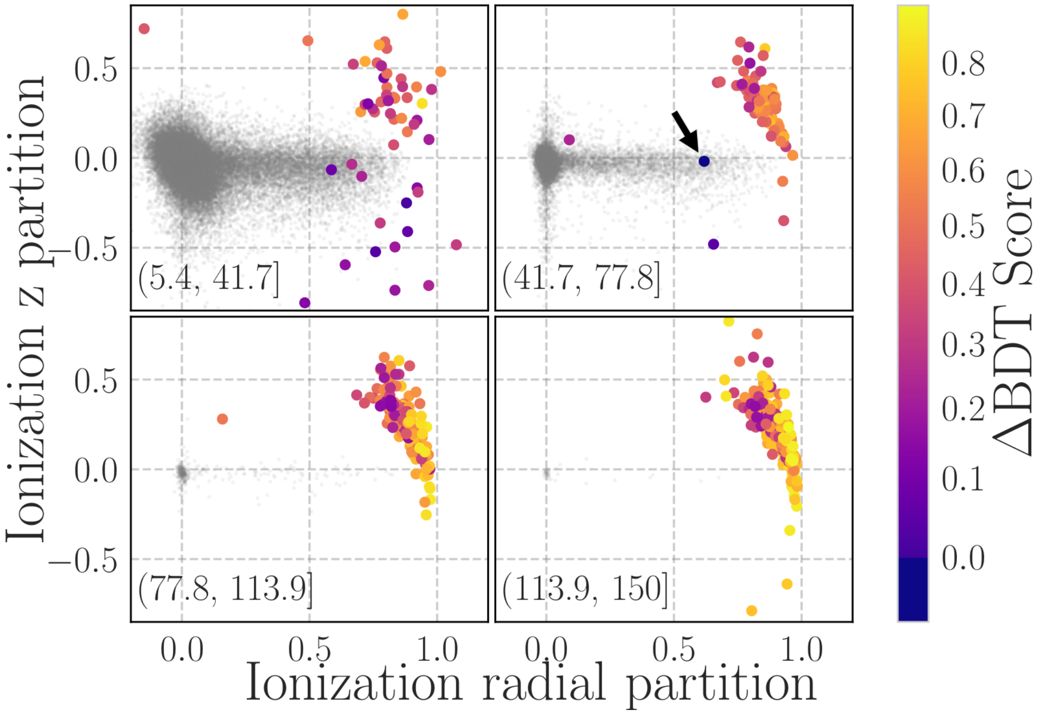
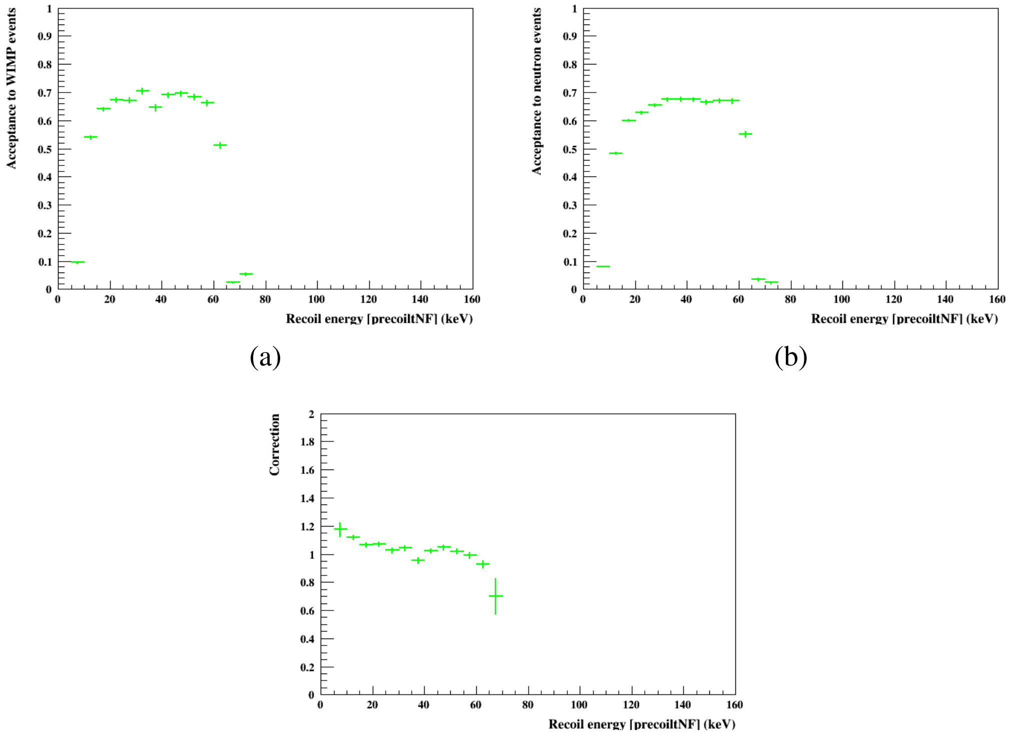


Figure 5.5: Scatter plots of ionization *z* vs. radial partitions for all DM-search events passing preselection cuts (large, colored) and signal model events passing the preselection and BDT cuts (small, gray). The events are divided into four even energy bins, labeled in keV. The events for all ten detectors are present, and each DM-search event has been colored by the distance from the BDT cut position in the detector that registered the event to the BDT score of the event itself. This sets the BDT cut position at ΔBDT = 0 and allows BDT scores to be compared between detectors. The single event accepted by the BDT cut is indicated with an arrow (and has ΔBDT < 0).



(a)

(b)

(c)

Figure 5.6: Efficiency correction from DMC. A large number of WIMP and ^{252}Cf interaction were produced using the DMC. We then apply the entire set of analysis cuts to the WIMP data (a) and the ^{252}Cf data (b) so that a relative correction factor (c) can be calculated [90].

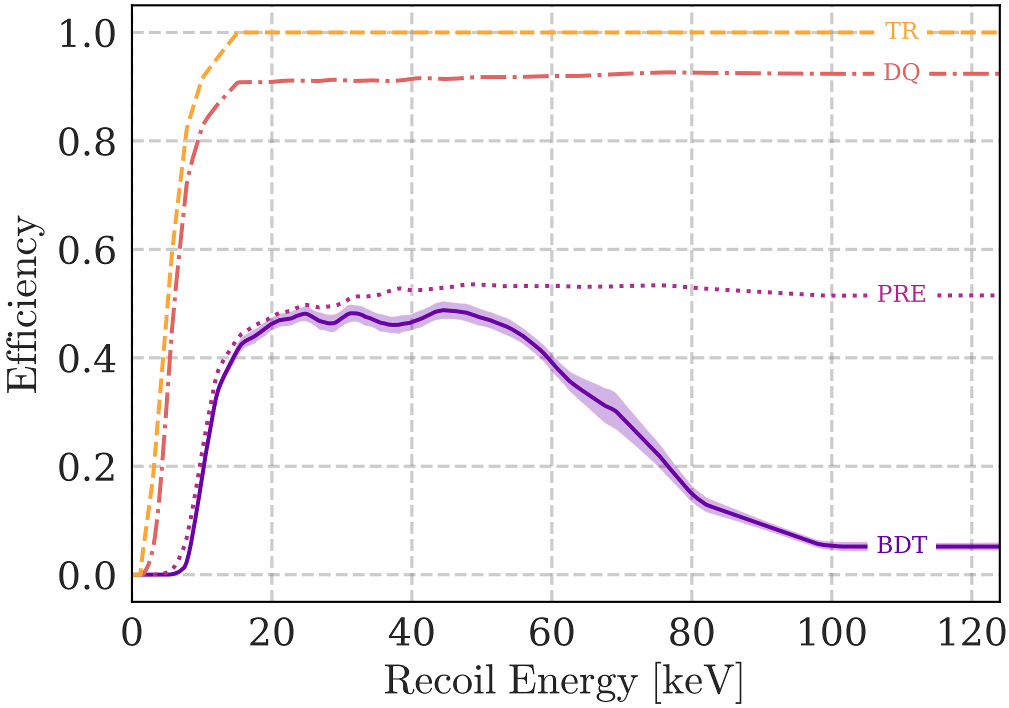


Figure 5.7: The total exposure-weighted efficiency is shown after sequential application of event selection criteria, averaged over all detectors. From top to bottom: hardware phonon trigger (TR), data quality (DQ), event preselection (PRE), and BDT discrimination (BDT). A 68% CL uncertainty band on the overall efficiency is shown.

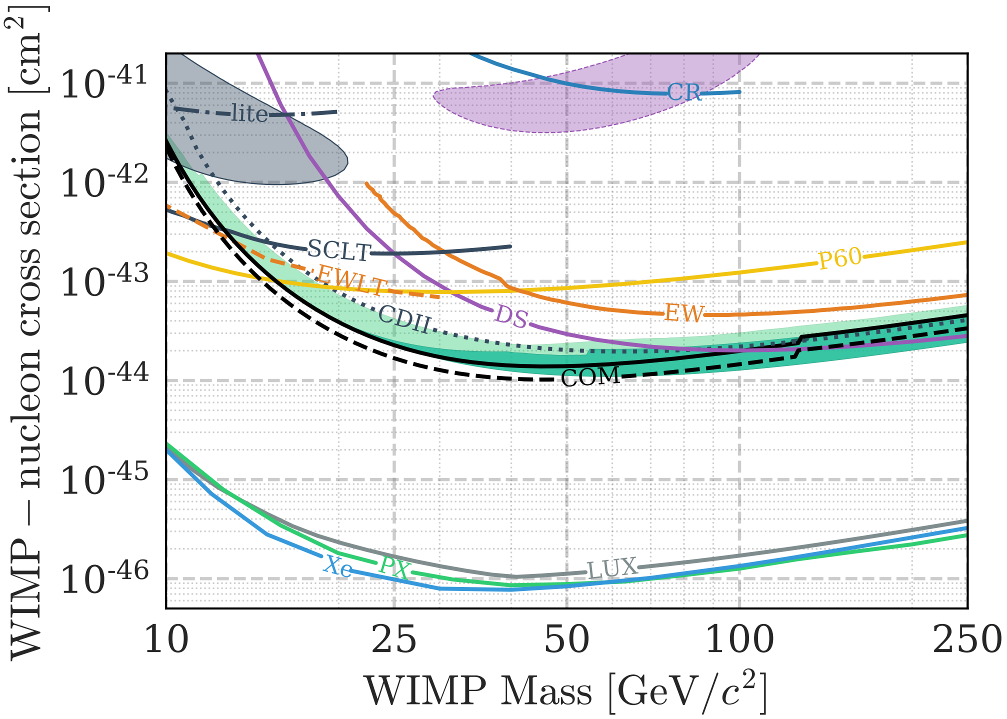


Figure 5.8: The 90% confidence upper limit on the DM-nucleon cross section (solid black) based on our single observed event. The range of the pre-unblinding 68% (95%) most likely expected upper limits are shown as dark green (light green) bands. Closed contours shown are CDMS II Si [91] (solid gray, 90% C.L.) and DAMA/LIBRA [92] (dotted purple, 90% C.L.). The remaining 90% C.L. exclusion limits shown are, in order of increasing sensitivity at 25 GeV/c^2, CRESST (CR) [93], CDMSlite Run 2 (lite) [94], EDELWEISS (EW) [95], SuperCDMS Soudan low threshold (SCLT) [96], DarkSide (DS) [97], PICO-60 (P60) [98], EDELWEISS low mass (EWLT) [99], CDMS II Ge alone (CDII) [100] as well as a combined limit with this result (COM), PandaX-II (PX) [101], LUX (LUX) [102], and XENON1T (Xe) [103].

Chapter 6

HIGHLY-MULTIPLEXED, ATHERMAL-PHONON-MEDIATED PARTICLE DETECTORS

6.1 Motivation

Future upgrades to the SuperCDMS experiment, such as the SuperCDMS SNOLAB effort [104] have a number of competing concerns to deal with in their attempts to increase sensitivity. Larger exposures requires our experiment to posses both larger individual detectors as well as more detectors in our array. As seen in the previous chapter, to maintain good background exclusion a significant fraction of our detector needed to be sacrificed in our fiducialization process. This was particularly evident at high radius, where oblique electron transport both complicated position reconstruction and suppressed ionization yield. Increasing the size of our detectors requires keeping the aspect ratio low[1], necessitating the scaling of the instrumented surface of the detector. This has the problem of scaling detector costs. Defect rates scale with instrumented surface area, and detector fabrication itself has historically been a large budgetary pressure in an experiment of this type. Improving our phonon-based position reconstruction is also of significant interest for future experiments. Phonons do not suffer from the kind of trapping and under-collection, or strangely asymmetric propagation of charge-carriers, making them attractive as detector size increases. As will be discussed in section 7.4, there is also significant interest in a high-voltage operational mode that forgoes ionization readout in the interest of decreased energy thresholds. In this style of operation phonon-based fiducialization is all that is available. In this sense, good position reconstruction and detector costs are in direct tension with our current iZIP instrumentation scheme. A single pixel of QETs[2] requires a dedicated readout pathway and all of the support infrastructure of wiring and cold electronics detailed in section 2.1.3. This complexity is also a large cost center, and is driven both by pixelization, as well as by increases in array size. In the following two chapters I will give an overview of our ongoing effort to develop a new type of SuperCDMS-style particle detector that

---

[1] The SuperCDMS SNOLAB detectors keep the ~1:3 aspect ratio of the SuperCDMS Soudan iZIP.

[2] Each of which contains many sensors wired together in series.

has the potential to significantly simplify the readout and fabrication of our phonon sensing elements, allowing for increased pixelization, better event reconstruction, much larger detector arrays and above all lower costs.

6.2 Microwave Kinetic Inductance Detectors

Microwave Kinetic Inductance Detectors (MKID's) are a locally developed, thin-film, superconducting, pair-breaking detector that may be used in many of the same applications as TES's. Unlike TES's MKID's do not require complex cryogenic electronics (such as the SQUID-based readout discussed in section 2.1.3), allowing virtually all of DAQ complexity to be handled at room temperature. Additionally, it is very easy to read out many individual MKID's on a single electrical channel[3]. As mentioned in section 2.2.3 each of an iZIP's 8 phonon pixels consist of hundreds of TES-based phonon sensors that are read out in parallel by a SQUID-based amplifier. This topology requires a dedicated readout channel for each pixel, limiting the potential number of pixels for this style of cryogenic detector. This can be overcome by dedicating a subset of the readout domain within each channel to an individual TES sensor, allowing the readout response to discriminate which sensor generated a particular response event. This practice is widespread in all telecommunications and is usually referred to as “multiplexing”. TES arrays have been successfully multiplexed in code-devision domain [105] the time domain [106, 107] as well as frequency domain [108], however this typically requires large additional complexity in the cryogenic electronics (usually in the form of multi-staged SQUID circuits).

MKID's are typically very high quality-factor (Q) resonating circuits with resonant frequencies typically in the GHz range. When coupled to a microwave transmission like they act as a “notch” filter[4] for the signal in the line. In this way the signal response of the MKID will be isolated in frequency space, leading to natural frequency-domain multiplexing as is seen in figure 6.1. MKID's are an active area of research and are currently being developed for use in sub-mm, optical, and x-ray astronomy as well as for rare-event searches in particle physics.

---

[3] By *channel* I am referring to a single physical readout pathway.

[4] Which is very much like a narrow band-stop filter.

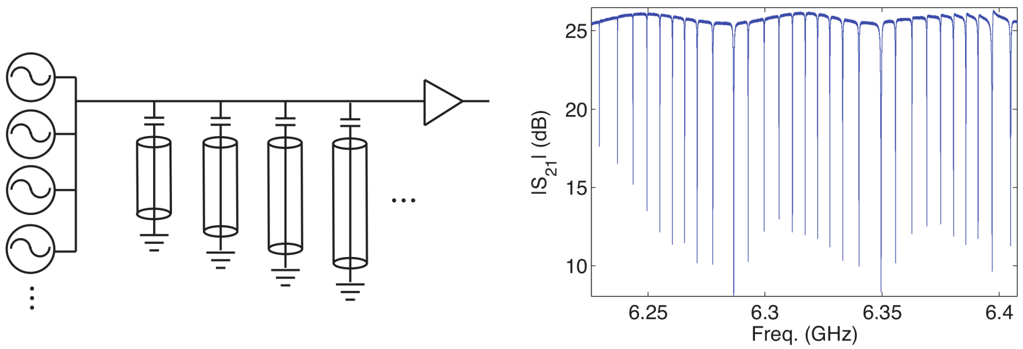


Figure 6.1: Multiplexing overview: (a) Schematic of numerous capacitively coupled resonators on a single feedline. Each is shown as a half-wave resonator with a different length and correspondingly different f_0. The resulting magnitude of the complex transmission through the feedline is shown in (b). Each resonator acts as a very narrow band-stop filter, limiting it's response to small, distinguishable portion of the readout domain. This is know as frequency-domain multiplexing [49].

6.2.1 Physics Overview

The Cooper pairs in a superconductor don't scatter or dissipate energy as they travel, causing the resistance (at $T = 0$ K) of a superconductor to be zero. These Cooper-pairs do, however, have mass and inertia. As a result, if an AC electric field is applied to a superconductor, the induced super-current will lag in phase behind the field, as it takes time for the field to accelerate the pairs. This implies that superconductors have a non-zero (and purely imaginary) impedance even at $T = 0$ K. As opposed to a normal inductor, which accomplishes a phase lag by conservatively storing energy in a magnetic field, this phase lag is accomplished by conservatively storing kinetic energy in the Cooper-pair condensate. This phenomena is aptly called "kinetic inductance". If an incident particle (such as an IR photon or a phonon) has an energy greater than the Cooper-pair binding energy, it can interact and create two quasiparticles. Unlike the condensate, the quasiparticles do scatter, which has two effects. First, it increases the resistance, or energy dissipation, of the superconductor. Because they scatter and lose kinetic energy, quasiparticle also do not contribute to the overall kinetic inductance of the metal. This decreases the mass of of the super-current condensate, which increases the kinetic energy required to maintain a given super-current and subsequently (and somewhat counter-intuitively) increases the kinetic inductance. An *LC* resonant circuit that is built out of such a superconductor will then experience changes in resonant frequency, phase, and quality factor from the splitting of Cooper-pairs inside of its inductive

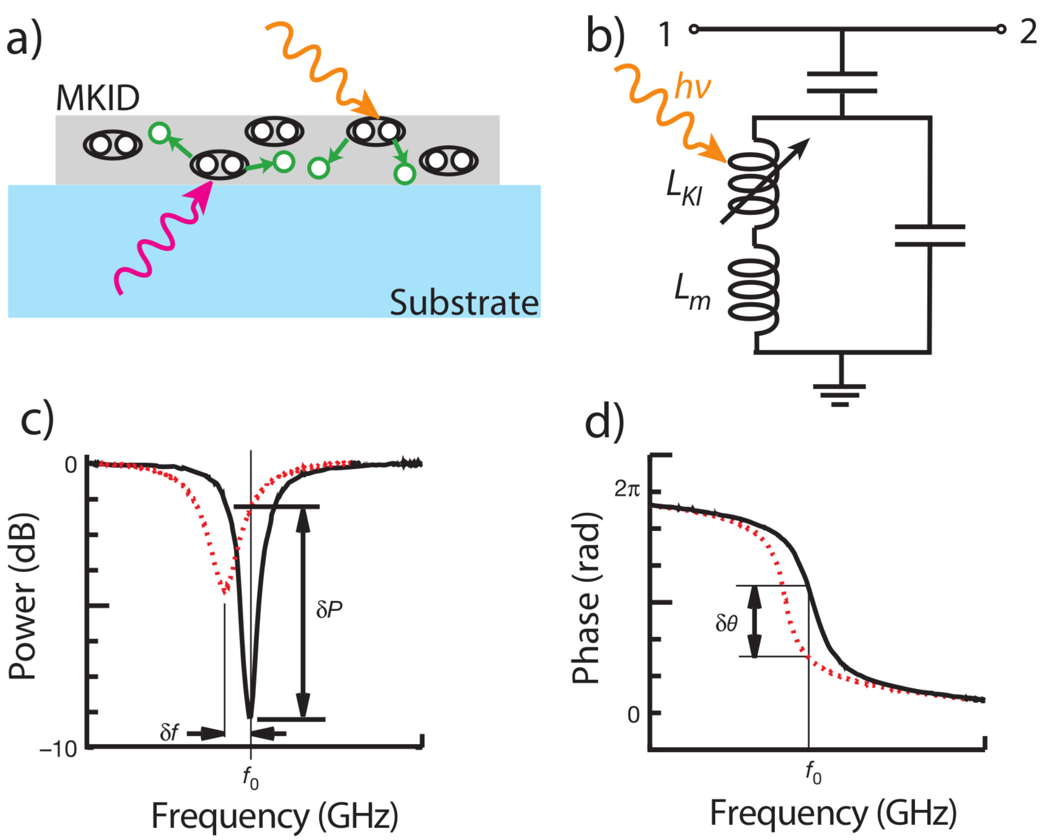


Figure 6.2: Physics of MKID response: a) Incident particles such as photons (orange) or phonons (pink) with energies $> 2\Delta$ can interact with a cooper pair (black) and create two quasiparticles (green). b) a schematic of the MKID circuit. L_{Ki} is the variable kinetic inductance while L_m is the normal magnetic inductance. Cooper pair creation causes in increase both in dissipation as well as inductance in the circuit. c) Magnitude of the complex transmission through the coupled feedline vs frequency. Response of an MKID is shown both before (black solid) and after (red dotted) the creation of a population of quasiparticles. d) Same as in c) but depicting the phase response of the MKID [109] via [49]

element. For a graphical depiction please see 6.2.

Different superconducting materials exhibit this effect to varying degrees. We can characterize this by splitting the inductance L of a superconducting inductor into two pieces

$$L = L_{\text{Kinetic}} + L_{\text{Magnetic}} \tag{6.1}$$

L_{Magnetic} is the component of the inductance that stores its energy in magnetic fields and is only dependent on the physical geometry of the inductive element. L_{Kinetic} is the kinetic inductance and is defined as the component that can be changed by the creation of quasiparticles in the superconductor. A given material is then characterized by the “kinetic inductance fraction”:

$$\alpha = \frac{L_{\text{Kinetic}}}{L} \tag{6.2}$$

6.2.2 MKID Signal Response

Quantitatively, how can we connect this change in quasiparticle density in our MKID to a change we can measure? As mentioned in the previous section, MKID’s act as notch filters, and can be probed by measuring the complex transmission as a function of frequency $S_{21}(f)$ through a feedline that is coupled to the resonator. Following [110, 49] we can write the this as

$$S_{21}^{\text{MKID}}(f) = 1 - \frac{Q/Q_C}{1 + 2\iota Q\left(\frac{f-f_0}{f_0}\right)} \tag{6.3}$$

with ι as the imaginary unit[5]. The resonator’s response is described in terms of its quality factor Q and resonant frequency $f_0 = 1/\sqrt{LC}$. As a concept, quality factor describes the energy loss from our resonant circuit and here we have broken it into two components:

$$\frac{1}{Q} = \frac{1}{Q_I} + \frac{1}{Q_C} \tag{6.4}$$

where the coupling quality factor Q_C is due energy lost from coupling to the feedline (which is the basis of our measurement), and the internal quality factor Q_I describes all other energy loss mechanisms. Figure 6.3 shows a few representations of equation 6.3, in the complex plane, as well as its polar parameters (magnitude and phase) vs frequency. Our MKIDs operate in the coupling-quality-factor dominated regime where $\frac{1}{Q_C} \gg \frac{1}{Q_I}$ and $Q_c \approx Q$.

It should be noted that equation 6.3, describes the complex transmission *at the resonator*. In general, if we were to measure the transmission at a later point along

---

[5] Although i is typically used in most math and physics texts to denote the imaginary unit, electrical engineers typically favor the use of the letter j to prevent a collision with the letter that is used to denote current I. Seeing as the current density is represented in all fields with the letter J I find this very unsatisfactory. As a result I will use the Greek letter ι.

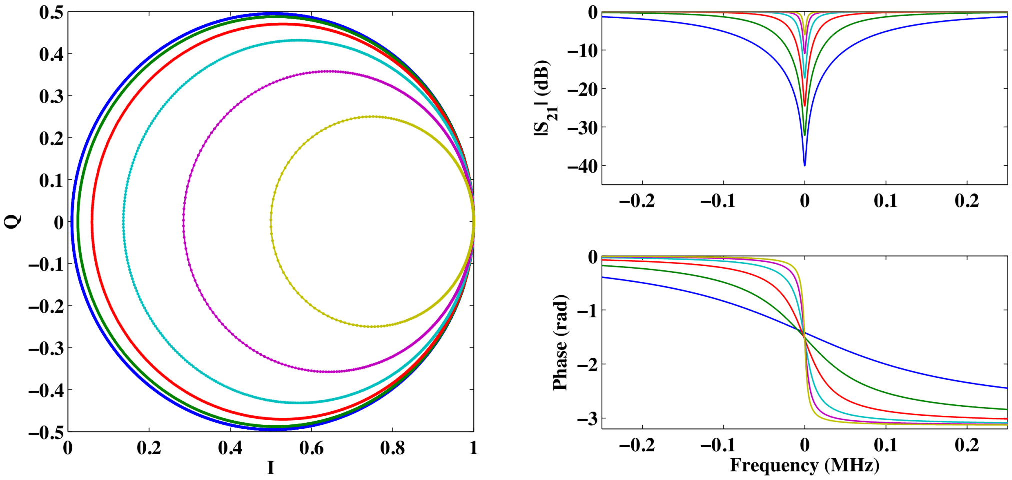


Figure 6.3: Visualization of equation 6.3 in various parameters. Left: The IQ plane generally represents the real and complex components of S_{21}^{Measured}, but in this case we are depicting the real and imaginary components of S_{21}^{MKID}. Shown is the response for a resonator with $Q_I = 10^6$ and $Q_C = 1 \times 10^4$ (blue), 2.5×10^4 (green), 6.3×10^4 (red), 1.5×10^5 (cyan), 4×10^5 (magenta), and 10^6 (yellow). Right: Magnitude vs frequency (top) and phase vs frequency (bottom) for the same set of resonators.

our transmission line[6] time τ after the resonator it would be

$$S_{21}^{\text{Measured}}(l) = AS_{21}^{\text{MKID}} e^{-2\pi\iota f\tau} \tag{6.5}$$

where A describes the gain, or attenuation, the signal experiences during the time period. As a result, if we were to plot $S_{21}^{\text{Measured}}(f)$ as in figure 6.3 it would be rotated by an angle of $2\pi f\tau$. To differentiate it from the complex plane of equation 6.3 the real and imaginary components of S_{21}^{Measured} are usually depicted as I and Q respectively[7] and this rotated plane is called the IQ plane.

We can write the surface impedance (Z_s) of a superconductor in terms of the surface resistance R_s and the surface inductance L_s as

$$L_s = R_s + \iota X_s = R_s + \iota\omega L_s = R_s + \iota\omega\mu_0\lambda_{\text{eff}} \tag{6.6}$$

Where X_s, is the surface reactance, and ω the angular frequency. Following [Gao]

---

[6] Assuming this is a nicely behaved feedline with no strange impedance mis-matches.

[7] As a result of this extremely unfortunate choice of notation the letters I and Q are both quite semantically oversubscribed.

because $f_0 = \frac{1}{2\pi\sqrt{LC}}$ and $Q_i = \omega L/R$ we have

$$\frac{\delta f_0}{f_0} = \frac{-1}{2}\frac{\delta L}{L} = \frac{-\alpha}{2}\frac{\delta L_{\text{Kinetic}}}{L_{\text{Kinetic}}} = \frac{-\alpha}{2}\frac{\delta X_s}{X_s} \tag{6.7}$$

$$\delta\frac{1}{Q_i} = \frac{\delta R}{\omega L} = \alpha\frac{\delta R_s}{X_s} \tag{6.8}$$

At very low temperatures ($T \ll T_C$) the resistance term will vanish and the overall impedance will be dominated by the reactance term. In this regime ($R_s \ll X_s$ and $T \ll T_C$) we can use equation 6.3 and find

$$\delta S_{21}^{\text{MKID}}|_{f=f_0} = \frac{Q^2}{Q_C^2}\left(\delta\frac{1}{Q_I} - 2\iota\frac{\delta f_0}{f_0}\right) \approx \alpha\frac{Q^2}{Q_C^2}\frac{\delta Z_S}{|Z_S|} \tag{6.9}$$

In the local, thin film limit, the thickness d limits the mean free path l, and both are much smaller than the coherence length ξ_0 and London penetration depth λ_L. In this limit we relate the surface impedance to the complex conductivity σ like

$$Z_s = \frac{1}{\sigma d} = \frac{1}{(\sigma_1 + \iota\sigma_2)\,d} \tag{6.10}$$

allowing equation 6.9 to be written

$$\delta S_{21}^{\text{MKID}} = \alpha\frac{Q^2}{Q_C^2}\frac{\delta\sigma}{|\sigma|} \tag{6.11}$$

From [111] we can express the components of this complex conductivity ($\sigma = \sigma_1 + \iota\sigma_2$) in terms of three integrals

$$\begin{aligned}\frac{\sigma_1}{\sigma_n} &= \frac{2}{\hbar\omega}\int_{\Delta}^{\infty}\frac{(f(E) - f(E+\hbar\omega))(E^2 + \Delta^2\hbar\omega E^2)}{\sqrt{(E^2+\Delta^2)((E+\hbar\omega)^2 - \Delta^2)}}\,\mathrm{d}E \\ &+ \frac{1}{\hbar\omega}\int_{\Delta-\hbar\omega}^{-\Delta}\frac{(1 - 2f(E+\hbar\omega))(E^2 + \Delta^2\hbar\omega E^2)}{\sqrt{(E^2+\Delta^2)((E+\hbar\omega)^2 - \Delta^2)}}\,\mathrm{d}E \end{aligned}\tag{6.12}$$

$$\frac{\sigma_2}{\sigma_n} = \frac{1}{\hbar\omega}\int_{\max[\Delta-\hbar\omega,\ -\Delta]}^{\Delta}\frac{(1 - 2f(E+\hbar\omega))(E^2 + \Delta^2\hbar\omega E^2)}{\sqrt{(E^2+\Delta^2)((E+\hbar\omega)^2 - \Delta^2)}}\,\mathrm{d}E \tag{6.13}$$

where the $f(E)$ is the standard Fermi function. In the low frequency ($\hbar\omega \ll \Delta$) and low temperature limit ($k_bT \ll \Delta$) [Gao et al LTD] derived an approximately analytic form in terms of the change in quasiparticle density δn_{qp}.

$$\frac{\sigma_1}{\sigma_n} = \frac{2\Delta\delta n_{qp}}{\hbar\omega N_0\sqrt{2\pi k_b T\Delta}}\sinh\left(\frac{\hbar\omega}{2k_bT}\right)\mathrm{K}_0\left(\frac{\hbar\omega}{2k_bT}\right) \tag{6.14}$$

$$\frac{\sigma_2}{\sigma_n} = \frac{\pi\Delta}{\hbar\omega}\left\{1 - \frac{\delta n_{qp}}{2N_0\Delta}\left[1 + \sqrt{\frac{2\Delta}{\pi k_b T}}e^{\frac{-\hbar\omega}{2k_bT}}\mathrm{I}_0\left(\frac{\hbar\omega}{2k_bT}\right)\right]\right\} \tag{6.15}$$

Where I_0 and K_0 are respectively the modified Bessel functions of the first and second kind, and N_0 is the single-spin density of states at the Fermi level. This allows us to rewrite equation 6.11 as

$$\delta S_{21} = \alpha\kappa\frac{Q^2}{Q_C}\delta n_{qp} \tag{6.16}$$

where

$$\begin{aligned}\kappa = & \frac{1}{\pi N_0}\sqrt{\frac{2}{\pi k_b T\Delta}}\sinh\left(\frac{\hbar\omega}{2k_bT}\right)\mathrm{K}_0\left(\frac{\hbar\omega}{2k_bT}\right) \\ & + \iota\frac{1}{\pi N_0\Delta_0}\left[1 + \sqrt{\frac{2\Delta}{\pi k_b T}}e^{-\frac{\hbar\omega}{2k_bT}}\mathrm{I}_0\left(\frac{\hbar\omega}{2k_bT}\right)\right]\end{aligned} \tag{6.17}$$

which finally relates the change in measured complex transmission to the change in the our MKID's quasiparticle population.

6.2.3 Potential Sensitivity

6.2.3.1 Single MKID Sensitivity

Our particular MKIDs are limited by the noise from our first stage HEMT [49]. In this regime we can calculate the sensitivity of one of our MKIDs following [112]. Assuming a δ-function deposition of energy in our MKID, and a HEMT amplifier noise temperature of T_{noise} the noise component of the complex transmission is

$$\delta S_{21} = \sqrt{\frac{k_b T_{\text{noise}}}{4P_{\text{read}}\tau_{\text{qp}}}} \tag{6.18}$$

where P_{read} is the readout power at the HEMT and τ_{qp} is the quasiparticle recombination lifetime, which determines the noise bandwidth. From [113] we can show that if the energy deposition is not δ-function-like, but extended over a period of time, then the $\frac{1}{\sqrt{\tau_{qp}}}$ in equation 6.18 needs to be replaced by $\frac{1}{\sqrt{\tau_{\mathrm{qp}}\chi_{\mathrm{BW}}}}$, where $\chi_{\mathrm{BW}} \leq 1$ describes the degradation in resolution due to the extended deposition time of the energy in our MKID. The timescales of importance here are the quasiparticle recombination lifetime τ_{qp}, the risetime of the energy deposition τ_{rise} and the subsequent falltime set by the total phonon lifetime in our device τ_{phonon}. In the limit where $\tau_{\mathrm{rise}} \ll \tau_{\mathrm{phonon}}$ we can show [113]

$$\chi_{\mathrm{BW}} \approx \frac{\tau_{\mathrm{qp}}}{\tau_{\mathrm{qp}} + \tau_{\mathrm{phonon}}} \tag{6.19}$$

Using equation 6.16 the minimum resolvable (1 σ) change in quasiparticle number that can be resolved from a dissipation readout scheme is

$$\delta N_{\mathrm{qp}}^{\mathrm{diss}} = \frac{2Q_C N_0 V\Delta}{\alpha Q^2 \mathrm{S}_1(\omega,\ T)} \sqrt{\frac{k_b T_{\mathrm{noise}}}{4P_{\mathrm{read}}\tau_{\mathrm{qp}}}} \tag{6.20}$$

where V is the volume of the MKID and $\mathrm{S}_1(\omega,\ T)$ is the real part of equation 6.17 or

$$\mathrm{S}_1(\omega,\ T) = \Re\kappa = \frac{1}{\pi N_0}\sqrt{\frac{2}{\pi k_b T\Delta}} \sinh\left(\frac{\hbar\omega}{2k_bT}\right) \mathrm{K}_0\left(\frac{\hbar\omega}{2k_bT}\right) \tag{6.21}$$

From [110], in the low temperature limit, where the majority of quasiparticles are supplied via readout power, the power absorbed by an MKID on resonance is

$$P_{\mathrm{absorbed}} = \frac{2Q^2 P_{\mathrm{readout}}}{Q_I^{\mathrm{qp}} Q_C} \approx \frac{N_{\mathrm{qp}}\Delta}{\eta_{\mathrm{read}}\tau_{\mathrm{qp}}} \tag{6.22}$$

where η_{read} is the efficiency for quasiparticle creation by the readout power, and

$$Q_I^{\mathrm{qp}} = \frac{2N_0 V\Delta}{\alpha \mathrm{S}_1(\omega,\ T) N_{\mathrm{qp}}} \tag{6.23}$$

is the component of the internal quality factor due to quasiparticle dissipation alone. Our sensitivity is maximized when P_{readout} (and subsequently Q_I^{qp}) is maximized until $Q_I^{\mathrm{qp}} \approx Q_C$, at which point the dissipation from thermally produced quasiparticles will start to affect the total Q. In this high readout power limit

$$P_{\mathrm{readout}} = \frac{Q_c N_0 V \Delta^2}{\alpha \mathrm{S}_1(\omega,\ T) Q^2 \eta_{\mathrm{read}} \tau_{\mathrm{qp}} \chi_{\mathrm{BW}}} \tag{6.24}$$

Combining equations 6.20 and 6.24, we get

$$\delta N_{\mathrm{qp}}^{\mathrm{diss}} = 2\sqrt{\frac{\eta_{\mathrm{read}} N_0 V k_b T_{\mathrm{noise}}}{\alpha \mathrm{S}_1(\omega,\ T) Q_I^{\mathrm{qp}} \chi_{\mathrm{BW}} \chi_{\mathrm{c}} \chi_{\mathrm{qp}}}} \tag{6.25}$$

where

$$\chi_{\mathrm{c}} = \frac{4Q^2}{Q_C Q_I} \tag{6.26}$$

$$\chi_{\mathrm{qp}} = \frac{Q_I}{Q_I^{\mathrm{qp}}} \tag{6.27}$$

In reading out the frequency response of MKID devices, the noise $\delta N_{\mathrm{qp}}^{\mathrm{freq}}$ is often dominated by so-called “two-level system” noise that arises from unwanted dielectric material (such as $\mathrm{SiO_2}$) on our substrate. For the devices in this thesis, this is not true, and the sensitivity to frequency response is improved over dissipation response by the ratio of the real and imaginary components of equation 6.17 or

$$\delta N_{\mathrm{qp}}^{\mathrm{freq}} = \frac{\delta N_{\mathrm{qp}}^{\mathrm{diss}}}{\beta(\omega,\ T)} \tag{6.28}$$

where

$$\frac{\Re\kappa}{\Im\kappa} = \frac{\mathrm{S}_1(\omega,\ T)}{\mathrm{S}_2(\omega,\ T)} = \frac{1}{\beta(\omega,\ T)} \tag{6.29}$$

For devices in this thesis β ~2-3.

6.2.3.2 Array Sensitivity

Now that we have an understanding of the sensitivity of a single MKID, we want to derive the expected energy resolution of a particular device. As mentioned in the previous section, our devices expect more sensitivity from frequency rather than dissipation readout and as a result, the following discussion will assume a frequency readout scheme. Assuming we have an array with N_{r} resonators, a particle interaction that produces $\delta E^{\mathrm{phonon}}$ of phonon energy in the substrate will induce a quasiparticle population in each MKID following

$$\delta N_{\mathrm{qp}} = \frac{\eta_{\mathrm{ph}}}{\Delta N_{\mathrm{r}}} \delta E^{\mathrm{phonon}} = \frac{\eta_{\mathrm{ph}} V}{\Delta \eta_{\mathrm{fill}} A_{\mathrm{sub}} d} \delta E^{\mathrm{phonon}} \tag{6.30}$$

where η_{ph} is the total efficiency for converting phonon energy in our substrate into quasiparticles (for more on this see section [device mount section]), and $N_{\mathrm{r}} = \frac{\eta_{\mathrm{fill}} A_{\mathrm{sub}} d}{V}$ where η_{fill} is the fraction of the substrate surface covered in MKIDs, A_{sub} is the surface area of the substrate, d is the metal film thickness, and V is the resonator volume. Given that we are splitting the total energy measurement into the N_{r} MKID's the total energy resolution will be

$$\sigma_E = \frac{\delta E^{\mathrm{phonon}}}{\sqrt{N_{\mathrm{r}}}} \tag{6.31}$$

Combining equations 6.25, 6.28, 6.30, and 6.31 we have

$$\sigma_E = \frac{2\Delta}{\eta_{\mathrm{ph}}\beta(\omega,\ T)} \sqrt{\frac{\eta_{\mathrm{fill}} A_{\mathrm{sub}} d \eta_{\mathrm{read}} T_{\mathrm{noise}} k_b N_0}{\alpha \mathrm{S}_1(\omega,\ T) Q_I^{\mathrm{qp}} \chi_{\mathrm{BW}} \chi_{\mathrm{c}} \chi_{\mathrm{qp}}}} \tag{6.32}$$

We defined χ_{BW}, χ_{c} and χ_{qp} in the previous section, and although $\chi_{\mathrm{qp}} = \frac{Q_I}{Q_I^{\mathrm{qp}}}$ will remain unchanged, our specific device requirements will allow for some simplifications to both χ_{BW} and χ_{c}. These are mainly due to pulse shape considerations.

Our desire to resolve the rising edge of the phonon pulse in our detectors will set a limit on the MKID response time τ_{responce}. This is set by the bandwidth of our resonator via the quality factor Q or

$$\tau_{\mathrm{rise}} \approx \tau_{\mathrm{responce}} = \frac{Q}{\pi f_0} \Rightarrow Q \leq \frac{\pi f_0 d_{\mathrm{sub}}}{c_{\mathrm{Ge}}} \tag{6.33}$$

where the phonon pulse rise time is given by the ratio of the substrate thickness d_{sub} to the speed of sound in Ge c_{Ge}[8]. For our devices, where $c_{\mathrm{Ge}} \approx 0.5\ \frac{\mathrm{cm}}{\mu\mathrm{s}}$, $f_0 \approx 3$ GHz, and $d_{\mathrm{sub}} = 2.5$ cm, we calculate that $Q \lesssim 5 \times 10^4$. The resonators in this dissertation have much higher internal quality factors ($Q_I > 10^6$) leaving us in the over-coupled regime where $\frac{1}{Q_C} \gg \frac{1}{Q_I}$ and $Q_c \approx Q$. In this regime

$$\chi_{\mathrm{c}} \approx \frac{4Q}{Q_I} = \frac{4\pi f_0 d_{\mathrm{sub}}}{c_{\mathrm{Ge}} Q_I} \tag{6.34}$$

As discussed in the last section the falltime is set by the phonon pulse fall time ($\tau_{\mathrm{phonon}} \approx 900\ \mu$s for an iZIP substrate) and the quasiparticle recombination time

---

[8] This is only true for substrate that are thick enough that the time is takes for a phonon to propagate from the interaction site is longer than the phonon down-conversion time, which is the case for massive inch thick detectors.

($\tau_{\mathrm{qp}} \approx$ 50–100 μs for our films). In this $\tau_{\mathrm{rise}} \approx \tau_{\mathrm{responce}} \ll \tau_{\mathrm{qp}} \ll \tau_{\mathrm{phonon}}$ limit equation 6.19 simplifies to

$$\chi_{\mathrm{BW}} \approx \frac{\tau_{\mathrm{qp}}}{\tau_{\mathrm{phonon}}} \tag{6.35}$$

Following [113], assuming the only energy loss mechanism available to our phonons is the creation of quasiparticles in our detector we can show that

$$\tau_{\mathrm{phonon}} = \frac{d_{\mathrm{sub}}\lambda_{\mathrm{pb}}}{2c_{\mathrm{Ge}}\eta_{\mathrm{fill}}p_{\mathrm{t}}d} \tag{6.36}$$

where p_{t} is the probability that a incident phonon will transmitted from the Ge crystal into the metal of the MKID and λ_{pb} is the phonon pair-breaking length in the metal film. Using the above, we can re-write equation 6.32 as

$$\sigma_E = \frac{\Delta}{\eta_{\mathrm{ph}}\beta(\omega,\ T)}\sqrt{\frac{A_{\mathrm{sub}}d\eta_{\mathrm{read}}T_{\mathrm{noise}}k_bN_0\lambda_{\mathrm{pb}}}{2\pi f_0\alpha p_{\mathrm{t}}\mathrm{S}_1(\omega,\ T)\tau_{\mathrm{qp}}}} \tag{6.37}$$

For the 25 nm Al film iZIP style devices in this thesis we have calculated [49] this to be $\sigma_E \approx$ 245 eV which is comparable to the ~100 eV baseline resolution of an iZIP. This could potentially be further improved by choosing a film material with better properties (such as a higher kinetic inductance fraction and lower gap), but it indicates that devices of this style have the potential to be a promising TES alternative.

6.3 Cm^2 Proof-of-Concept Device

With the above criteria in mind, we fabricated a proof-of-concept device at the Jet Propulsion Lab's Micro-devices Laboratory (MDL). This prototype consisted of 20 MKIDs on a 20 mm by 22 mm by 1 mm chip. The complete description of the development and testing of this family of devices is outside the scope of this thesis and can be found here [49]. The testing of these devices was some of the earliest work I was involved in at Caltech, and much of their design methodology, and considerations informed our subsequent development effort. As a result, I will present a short summary in the following section.

6.3.1 Device Design

As mentioned, this device consisted of 20 MKIDs on a 22 mm by 22 mm by 1 mm chip patterned on a ⌀ 100 mm high-resistivity (>5 KΩ) Si substrate. These substrates were chosen due to their compatibility with the processing equipment at MDL[9]. Size and number of MKIDs was designed around keeping the proportion of the detector surface that is covered with phonons collectors consistent with an iZIP ($\eta_{\text{Fill}} \approx 0.05$). The MKID's themselves were modeled after the Lumped Element Kinetic Inductance Detectors (LEKIDS) proposed by S. Doyle in [114]. The complex transmission response of these MKIDs (S_{21}^{MKID}) was then simulated using SONNET [115] and fit to equation 6.9. This allowed for the frequency and quality factor of each MKID to be engineered to specific design values. Due to our HEMT amplifier's low-frequency cutoff of ~3 GHz, these resonators were designed to work in the 3-3.2 GHz band with 10 MHz spacing. To be able to resolve the athermal phonon pulse (rise-time $\sim$ 1 μs) we required that $Q_c \lesssim 5 \times 10^4$, which is set by tuning the MKID-feedline gap. As we saw in section 6.2.3.2 materials with higher kinetic inductance fraction α have improved sensitivity, and concomitantly higher surface inductance L_s. As $L_s = \frac{\hbar R_s}{\Delta \pi}$, these materials will typically be small-gap and have a high normal state resistivity. These devices were designed around a family of high-resistivity ($\rho_n \sim 150\ \mu\Omega$cm) superconducting ceramic materials called sub-stoichiometric titanium nitride (TiN$_x$) with $L_s \approx 20$ pH/sq used in simulation. The resulting resonator layout and geometry can be seen in figure 6.4.

6.3.2 Results

After fabrication, the devices were tested in an Oxford Instruments Kelvinox 25 dilution refrigerator ($T_{\text{Base}} \approx 70$ mK). The readout system utilized the FPGA-based Reconfigurable Open Architecture Computing Hardware (ROACH) developed by the Collaboration for Astronomy Signal Processing and Electronics Research (CASPER) [116]. This system utilizes dual digital-to-analog converters (DAC's) which generate a comb of tones at baseband ($f_{\text{DAC}} < 500 MHz$) to excite the resonators. These signals are then up-converted to the 3 GHz band at which the MKID's operate using a custom IQ mixer. After being run through our device, this signal is then down-converted using the same model IQ mixer and digitized

---

[9] Specifically, the Cannon EX3 stepper system. The thickest wafer this system can easily align to is 1 mm. Similarly, a single stepper-field was 22 mm by 20 mm when projected on the substrate.

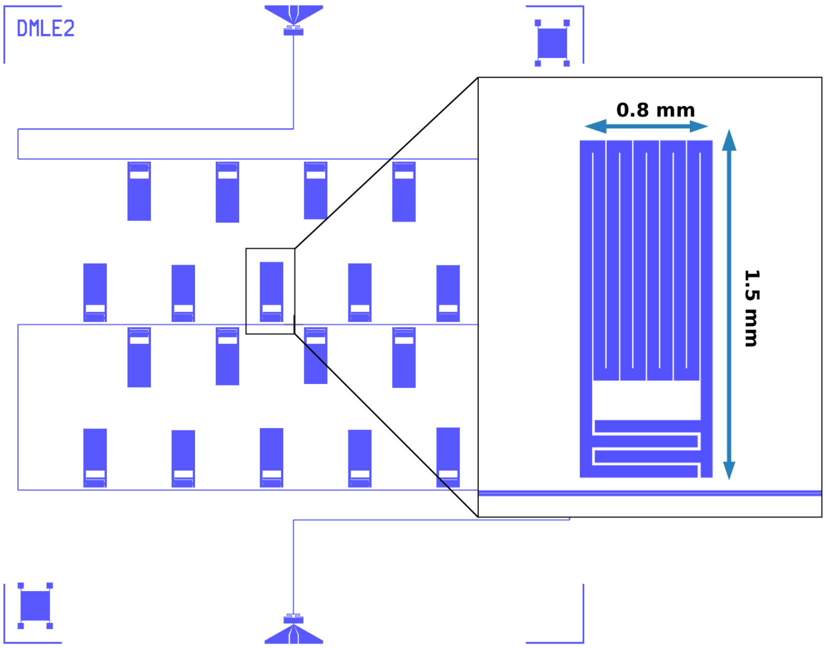


Figure 6.4: Schematic of the mask used to pattern our proof-of-principle device. The layout consists of a simple rectilinear grid of 20 resonators. Inset detail: each MKID is roughly 0.8 mm by 1.5 mm although the exact dimensions are slightly different from one another to ensure different L_m, and achieve frequency-domain multiplexing.

using dual analog-to-digital converters (ADC's). Custom firmware developed for this ROACH system then implements a 20-channel digital downconverter (DDC) to isolate each frequency channel. The system provides the amplitude and phase of the signal transmitted past each resonator at 1.3 MHz per channel, with on-board buffering and triggering. The total system bandwidth is 340 MHz, easily encompassing the 200 MHz bandwidth of the array. For a schematic of this please see figure 6.5. TiN_x films proved problematic for these devices. Changing the amount of nitrogen in the TiN_x films changes the gap, and allows T_c to be tuned from 4 k for stoichiometric TiN to ~0.5 K for highly substoichiometric films. Collecting ballistic phonons required films with $T_c < 2$K, which is well into the substoichiometric regime. Each MKID's responsivity is highly sensitive to this exact chemistry, and it proved impossible to deposit films with a high enough uniformity to be useful.

An example of the phase response for each of the 20 resonators following a 200 keV interaction in the substrate from a cosmic-ray can be seen in Fig. 6.6. The location of the interaction can be reconstructed from the relative fraction of energy partitioned into each of the 20 MKIDs (just as in section 2.4.5). This allows for a local

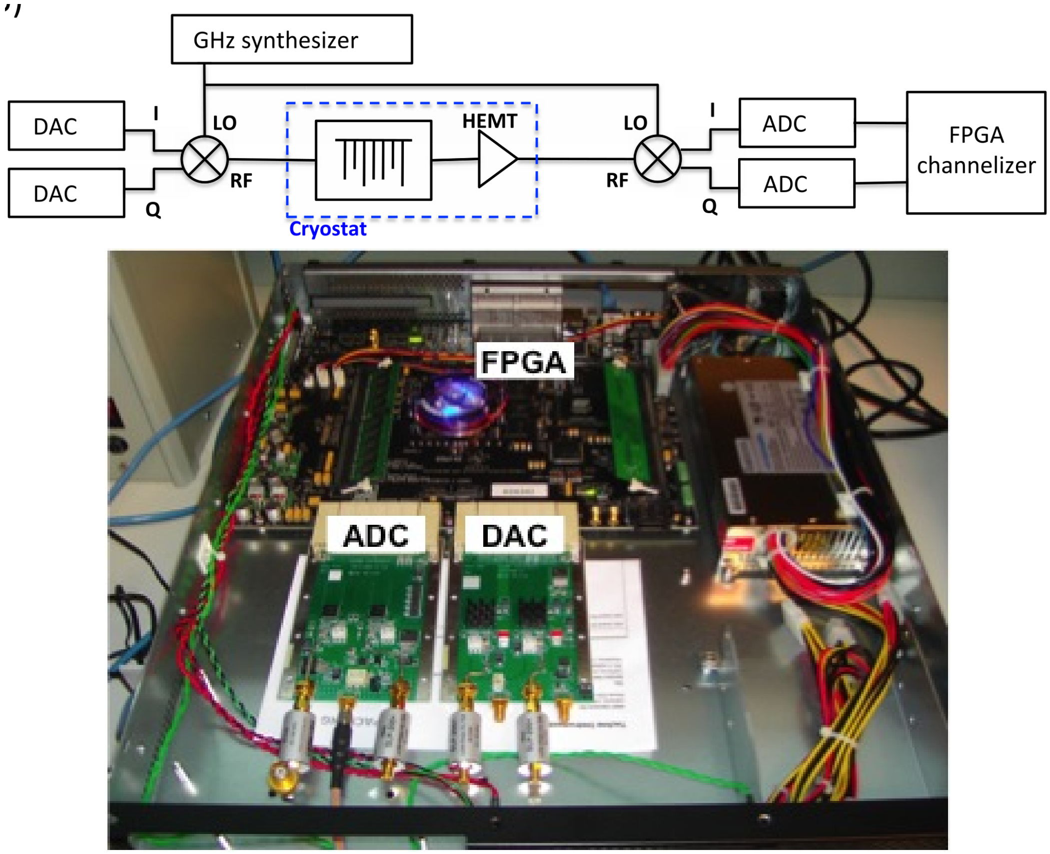


Figure 6.5: (a) Schematic of our ROACH-based readout system. Many hundreds of MKIDs can be read out simultaneously using these relative simple and cheap room temperature electronics, which are shown in (b) [49].

energy recalibration to be preformed as seen in Fig. 6.7. The best-fit resolution after the position dependent correction is preformed and the periphery of the device is excluded is σ = 0.55 keV at 30 keV (from an ^{129}I line). This is almost a factor of 2 better than the uncorrected resolution and brings us within 40% of the baseline resolution of σ = 0.38 keV.

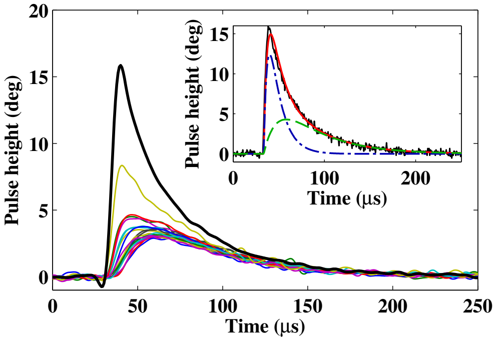


Figure 6.6: Example phase response of a 200 keV cosmic ray event. The pulses have all been normalized against their respective MKID's relative responsively, and an anti-aliasing filter has been applied. The resonator closest to the interaction site (thick black pulse) has both the greatest energy partition as well as the most prompt response. (inset) This is used to construct a two template optimal filter fit to the data (solid fit). This contains a prompt component (dotted-dashed) that reflects the quasiparticle lifetime in the MKID as well as a slower component (dashed) representing the phonon lifetime in the substrate.

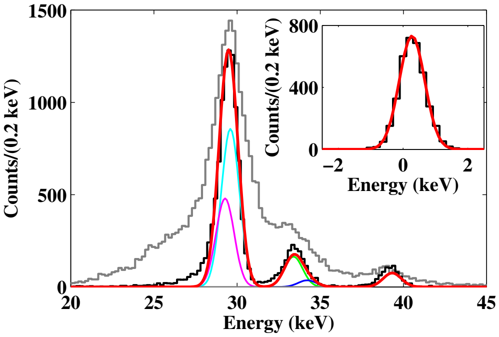


Figure 6.7: Observed spectrum when the substrate face opposite the resonators is illuminated with x-rays from an ^{129}I source. The light histogram shows the reconstructed energy for all events prior to correcting for position-dependent smearing of the response; the dark histogram is after. The best fit resolution at 30 keV is σ_E=0.55 keV. (inset) Fit to the reconstructed energy spectrum for randomly triggered noise traces. The inferred baseline resolution is σ_E=0.38 keV.

Chapter 7

LARGE FORMAT DETECTORS

7.1 MKID-Based iZIP detector

After the success or our proof-of-concept devices discussed in the last chapter, we next turned to prototyping an iZIP that would utilize MKIDs as the phonon-sensing element. The most obvious change here would be one of scale. The proof-of-concept was a 2 cm x 2 cm x 1 mm chip with twenty resonators, while iZIPS are single-wafer devices that each use an entire ⌀76 mm x 25 mm thick substrate. Aside from a simple scaling-up of our device, as we have seen in chapter 2, iZIPs contain a few important design features not present in our proof-of-concept. In what proved to be a somewhat ambitious decision, we choose to tackle all of these considerations simultaneously. As a result, the initial series of full-wafer devices we produced very much resemble an iZIP with it's QET's replaced by MKIDs and will be referred to as "iZIP-MKID" or "iZID" devices.

7.1.1 Design Considerations

Similar to the proof-of-concept described in section 6.3, the size of the new iZID's MKIDs was set by our readout electronics and our desire to keep our active metal coverage consistent with an iZIP's at the few percent level. Our current HEMT amplifier has a lower operating frequency limit of 3 GHz, which sets an effective upper limit on the size of our resonators. We are also currently limited to multiplexing 250 MKIDs per FPGA board, which together with our coverage requirement, sets a lower limit. Luckily, resonators that are $\sim$1 mm^2 allow for 250 MKIDs to be designed for the 3 to 3.5 GHz band while maintaining a $\eta_{\text{fill}} \approx 0.05$ metal coverage. This means that the overall size of the resonator can stay roughly the same as it was in our proof of concept.

In addition to the phonon based readout, ionization collectors or "charge lines" have been added which may be used to bias the substrate and read out an ionization signal. This would allow us to make the complementary phonon and ionization measurements that are the hallmark of SuperCDMS' iZIP detectors. The addition of these charge lines, which are interleaved with the MKID feedline, sets a limit on the distance the MKID can protrude from the feedline to ∼ 500 microns. This charge line / MKID pitch is set by the overall DC capacitance requirements of the iZIP style design. As discussed in section 2.4 the interleaved nature of the iZIP electrodes only help to identify near-face surface-event interactions. Near-sidewall interactions are detected using ionization lines that are broken into two “pixels”: an inner disk surrounded by an annular guard electrode. This forced us away from the rectilinear layout of the proof-of-concept devices, to something more azimuthally symmetric. The final consideration we made when designing our new resonators was to ensure that no edge formed a dipole antenna in the far field, to minimize cross resonator coupling.

7.1.2 Feedline Design, Simulation, and Layout

This large formate device is designed to use the same 25 nm Al film from our prototype device. For simulation purposes we assume that while superconducting, the film has a surface impedance of $L_s = 0.8$ pH/sq. This surface impedance, together with our desire to have a well-matched input impedance of 50 Ω, sets the CPS center-gap to conductor-width ratio to be 1:15 ($Z_{\text{in}} \propto \frac{\text{center-gap}}{\text{conductor-width}}$). Given the very limited space available it is desirable to make the CPS feedline as narrow as possible. This is both because a wider feedline would, in and of itself, take up more space, but also because a wider CPS would decrease the coupling quality factor, necessitating a greater separation between the feedline and the MKID itself. For these reasons the basic design constraint of the feedline was the center gap. Our photo-mask supplier has a line width tolerance of 2 microns, which initially set our feedline gap width to 2 μm and the conductor width to 30 μm.

The layout of the feedline[1] for our prototype cm^2 device is a simple, rectilinear co-plainer stripline or CPS. Most all of the feature edges, in fact, are strictly aligned along the x or y axis. There is quite a bit to recommend this particular style of layout for most integrated circuit designs, mostly for reasons of accurate digitization. Simulation, as well as the fabrication of photolithographic masks, is conducted on a grid. Provided all edges of a physical feature are aligned with the grid directions, local grid spacing can be of order the size of the feature. As soon as an edge deviates from this prescribed geometry, however, this can start to break down. Our desire to have an azimuthally-symmetric iZID, drove us away from this simple geometry

---

[1] As well as the resonators themselves if we are being pedantic about it.

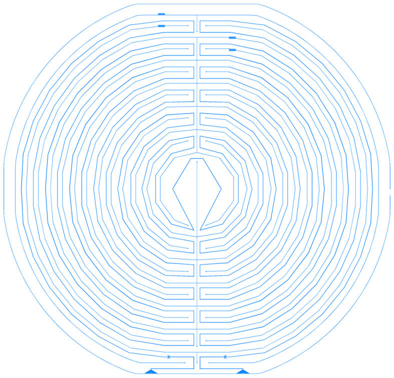


Figure 7.1: Layout for the feedline and charge-lines on our iZID device. The spacing is set to keep the high tangential surface field from the iZIP, and the circular layout is necessary due to the separate inner and outer charge collection requirements. The bonding pads for each of these charge collection circuit can be seen in the upper left (annular guard ring) and upper right (inner disk) part of the plot.

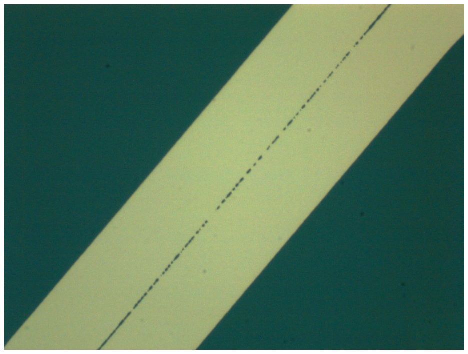


Figure 7.2: An example of shorts due to digitization errors in the initial iZID mask production.

and had three implications. Simulation of the device had to be done in sections that could be aligned to a grid, making the full behavior of our device somewhat hard to predict. Second, the initial fabrication of our devices suffered from "digitization shorts" (see figure 7.2) which were due to our masks fabricator's smallest feature size of 2 microns only being true in the $\hat{x}$ and $\hat{y}$ directions. In the $\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)$ direction the smallest allowed feature size is actually $2\sqrt{2} \approx 2.8$ microns. This lead to an increase in the CPS center gap to 3 microns. For an overview of the layout please see figure 7.1.

7.1.3 Resonator Design and Simulation

The difficulty in designing an MKID for our new large-format detector is that the proof-of-concept's near-square aspect ratio is fundamentally incompatible with the iZIP-style interdigitated ionization line / phonon line scheme. This is illustrated by figure 7.3. With this constraint in mind, we settled on a simple "fold over" form factor for our meandered inductor as seen in figure 7.4.

The MKID is 200×4000 microns. The inductive section is 45 microns with a 7 micron gap; the interdigitated capacitor fingers are 40 microns with a 20 micron gap.

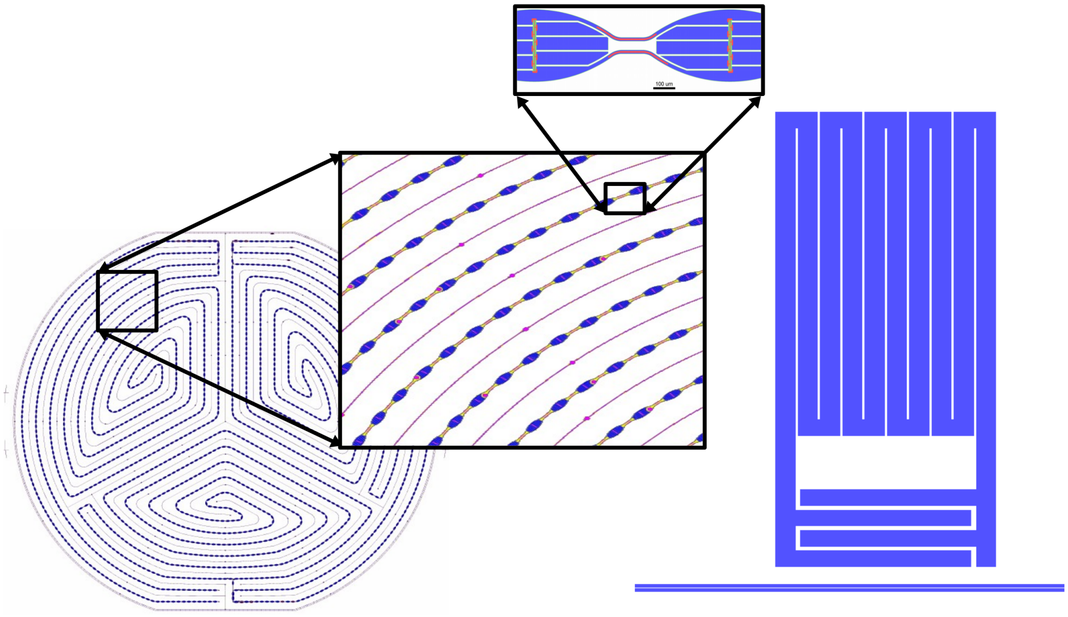


Figure 7.3: This is a comparison of our prototype MKID design (right) with a zoom of our phonon-collecting QET on a SuperCDMS iZIP. The MKID and the zoomed graphic of the QET are shown to scale. Part of the difficulty of the new large device layout is that there is not way to fit our prototype MKIDs in the superCDMS soudan style layout.

These resonators may be tuned to different resonant frequencies by changing the length of the inductive section. Using SONNET's high frequency electromagnetic simulation software we were able to simulate potential cross-talk between nearest neighbor pairs of our design as seen in figure 7.5. SONNET was also used to simulate the the resonator's frequency response, allowing us to engineer our desired resonant frequencies. For a lumped element MKID we expect that the resonant frequency $f_\mathrm{r} \propto \frac{1}{\sqrt{l}}$ where l is the length of the meandered inductor. For a half-λ resonator we expect the relationship to be $f_\mathrm{r} \propto \frac{1}{l}$. As we will see in section 7.1.3.2, the high-aspect ratio of this particular MKID geometry leaves these resonators somewhere between these two regimes. For the 3-3.5 GHz band of interest, however, a linear approximation works very well as can be seen in Figure 7.7. We then used this fit to design 250 resonators with resonant frequencies evenly spaced in the 500 MHz between 3 and 3.5 GHz. The following subsections will give a more detailed accounting of each of these design steps.

Figure 7.4: This MKID is 200 × 4000 microns, and utilizes a new “fold-over” inductor topology. Each portion of the inductive meander carries a current opposite to all of its nearest neighbors. In this way, any potential dipole coupling should cancel at length scales larger than the meander size.

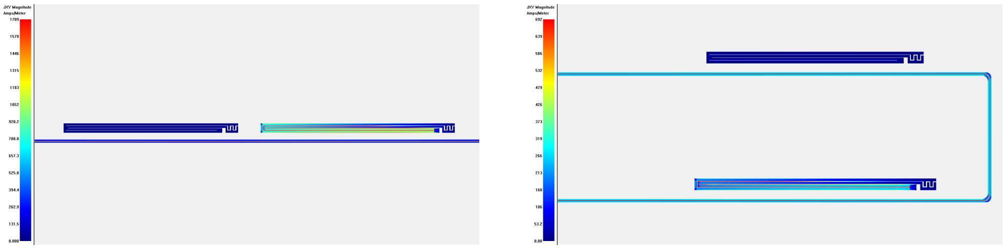


Figure 7.5: Current response during our azimuthal (left) and radial (right) coupling simulations. In both cases, for typical MKID separations, there was no observable induced current in the non-primary resonator. The resonator in the bottom in the radial tests did suffer from a deviation in its designed resonant frequency, which was later corrected for.

7.1.3.1 Frequency Engineering

After laying out a resonator that will fit in the space alloted, we need to be sure we can reliably engineer both the frequency as well as the quality factor of each MKID in our array. To do this we turn again to SONNET for simulation. The goal is to stimulate one end of a small feedline segment that passes near an MKID and measure the complex transmission, S_{21}, from the other terminus as seen in figure 7.6. The MKID should produce a characteristic response in $|S_{21}|$, which is a dip or “notch” around its resonant frequency, f as seen in section 6.2.2. The quality factor can also be directly read off of this plot by examining the full with half max, or the width of S_{21} at a value of -3 dB. We need to do this manually for a number of different inductor lengths, until we can accurately model and predict the relationship. The simplest way of doing this is as follows:

1. Assuming we have some coarse knowledge of the resonant frequency of our MKID, say that is should be between 3.2 and 3.4 GHz, we can pick a domain over which to do a frequent sweep.
2. SONNET is fairly clever and will adaptively sample more points around the resonant frequency, but given the design quality factor of 50,000 the width of a resonator will be ~0.00006 GHz, far smaller than the 0.2 GHz sweep range. As a result, the peak will be very under-sampled, and accurately measuring Q or even f_0 would be impossible.
3. This is overcome by refining and repeating the simulation on a much narrower bandwidth, say 0.0001 GHz around the value of f found in step 2. At this point the sampling density can be great enough to accurately measure both Q and f_0.

The crux of the problem is that near the points of interest for our measurement, $|S_{21}|$ is very rapidly changing. To overcome this we utilize a method devised by Gao [117]. The key insight here is that by inserting a third port in the center of the inductive meander as shown in figure 7.6 the input impedance, Z_{in}^3 of will, near resonance, take the form

$$Z_{in}^3(f) = R\left[1 + 2jQ\left(\frac{f - f_0}{f_0}\right)\right] \tag{7.1}$$

where R is an effective series resistance due to the loads at ports 1 and 2. Following [117], the resonance frequency can be found by examining the point where $\Im Z_{in}^3 \equiv 0$. At this point it would also be true that $\Re Z_{in}^3 \equiv R$. Similarly at this point

$$Q = \frac{kf_0}{2R} \quad \text{where } k = \frac{\mathrm{d}\Im Z_{in}^3}{\mathrm{d}f} \tag{7.2}$$

$Z_{in}^3(f)$ slowly varies as can be seen in figure7.6, allowing us to conduct a wide frequency sweep. This allows for fast extraction of the resonant frequency and quality factor of the resonators. Using this method we simulated eight resonators of

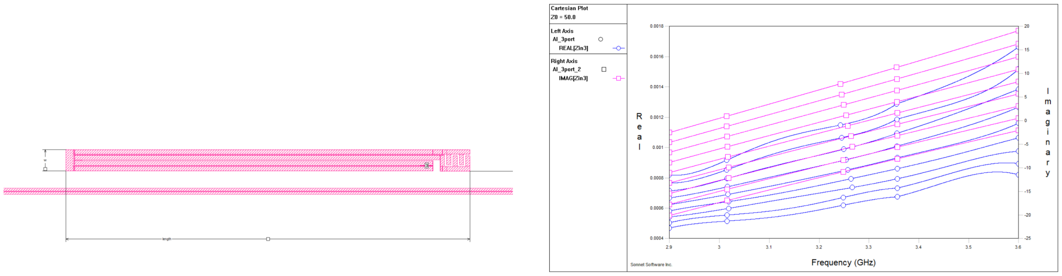


Figure 7.6: Modeling the resonant frequency using the three port method. On the left is a particular MKID geometry in SONNET with the third port shown. On the right is the real (blue) and imaginary (pink) components of the input impedance to the third port for a series of MKIDs with different lengths. Finding the resonant frequencies and quality factors of these MKIDs is as simple as finding the zeros of these lines.

varying inductor length, l, and fit the result to a line of the form $f_o = ml + b$ as can be seen in figure7.7, giving values of $m = -s0.0006274$ GHz/μm and $b = 5.85$ GHz.

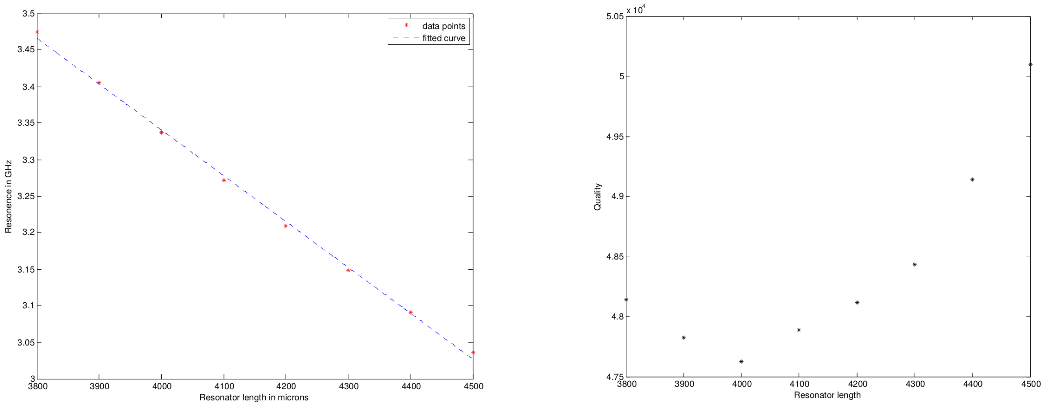


Figure 7.7: The resonant frequency (left) and quality factor (right) vs meandered inductor length for a series different MKIDs simulated in SONNET. As our MKIDs are not entirely lumped element ($f_0 \propto \frac{1}{\sqrt{l}}$), or half-wave (($f_0 \propto \frac{1}{l}$)) we simply fit the relation on the left to a line, which approximates it very well in this range.

The coupling quality factor could also be subtly effected by the length of the meandered inductor designed length. If it is, the separation between our MKID and the CPS feedline would need to be engineered such that regardless of the inductor length Q_C remained constant. As discussed in section 7.1.1, we require that $Q \lessapprox 5 \times 10^4$ in order to achieve the microsecond resolution needed to resolve the rising edge of the phonon pulses. Although there was some variation in the measured Q as a function of the inductor length it was on the $\sim 5\%$ level, far lower than the variation in quality factor from our prototype device. There was also no clear or monotonic frequency dependence, and as a result the MKID-feedline separation was kept constant for all resonators..

7.1.3.2 Current Uniformity

Although designed to have the same $\sim 5\%$ metal coverage as an iZIP, current non-uniformity within the meandered inductor may effectively act to lower this number. The kinetic inductance effect depends on the AC super-current in the inductor. This current will peak near the center of the inductive portion and tend towards zero at the capacitor. Any particle interaction near this current peak will have a much greater phase response than one near the capacitor. We can calculate this effectively reduced area, A_{eff}, like

$$A_{eff} = \eta A_{actual} \tag{7.3}$$

with

$$\eta = \int_{s=0}^{l} \left[\frac{j(s)}{j_{\max} l} \right]^2 \mathrm{d}s \tag{7.4}$$

where $j(s)$ is the current density parameterized by arc length, s, and j_{max} is its peak value. For a lumped element MKID, we expect $j(s) \approx j_{\max} \Rightarrow \eta \approx 1$. For a $\frac{\lambda}{2}$-wave resonator , $j(s) \approx \sin(2\pi \frac{s}{l}) \Rightarrow \eta \approx \frac{1}{2}$. This process is illustrated in figure7.8. For this MKID geometry, $\eta = 0.67$, leading to an effective metal coverage of 3%. Although this is a decrease from the performance of our prototype MKID geometry where $\eta = 0.89$, the combination of required frequency, aspect ratio, and metal type made improvements quite difficult. This reduced effective area was deemed acceptable for an initial large format device, but for further improved MKID geometries please see section 7.4.

7.1.3.3 Coupling Correction and Frequency Encoding

To ensure that our resonators respond at their design frequency, we conducted a series of coupling simulations to ensure there was no cross-talk between nearest neighbor pairs of resonators. The physics of this phenomena for a specific MKID

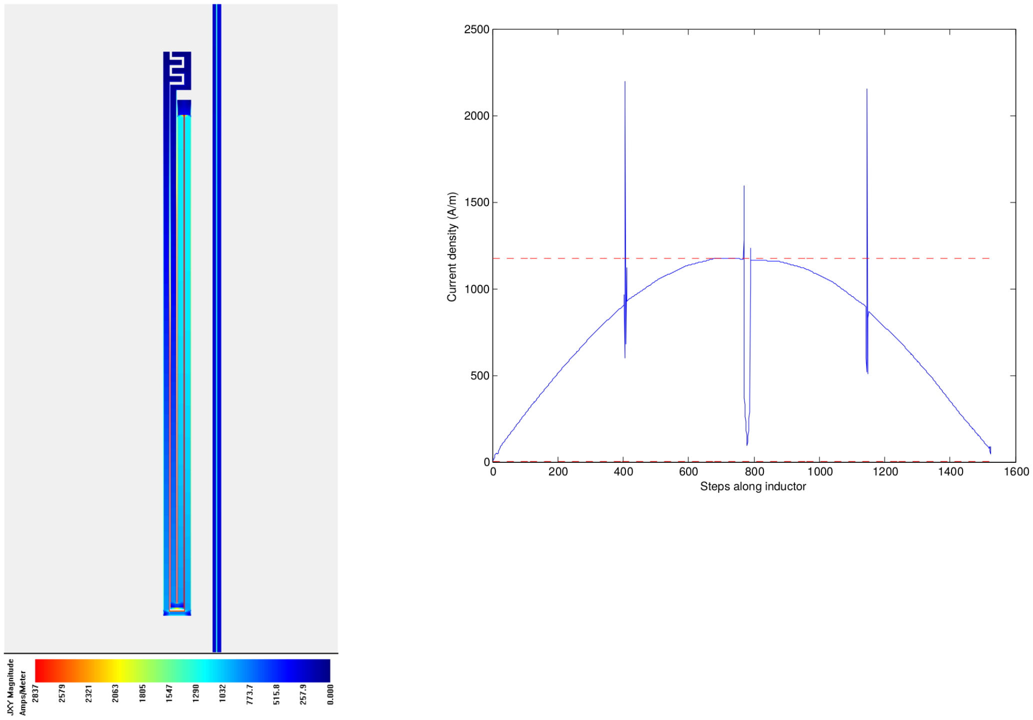


Figure 7.8: Left: Current uniformity simulation in SONNET. Right: current density vs distance along the feedline for our simulated device (blue). A perfectly lumped-element MKID would have uniform current (red), maximizing the sensitivity for a given metal coverage.

geometry is discussed in more detail here [118]. Due to computing resource constraints, we approximated our inter-MKID coupling with two basic geometries of nearest neighbor pairs, radial and azimuthal (shown in figure 7.5). These approximations allowed us to ensure that we could design non-interacting resonators wile keeping an efficient Manhattan geometry for simulation purposes. As can be seen in figure 7.8, for MKIDs reasonably separated in frequency space, there is no cross-coupling for either the radial or azimuthal pair. However, during the radial coupling simulations, we did discover unwanted coupling from the lower MKID to the vertical feedline segment. The difference between the actual length of the MKID and the expected length of the MKID based on the initial free resonator simulations for this geometry is quadratic as a function of resonator length. We applied this correction to the affected resonators as seen if figure 7.9.

The final frequency encoding of the resonators needs to ensure that nearby MKIDs are sufficiently far in frequency space as to not interfere with one another. Our previous device used a random frequency placement of the resonators, however

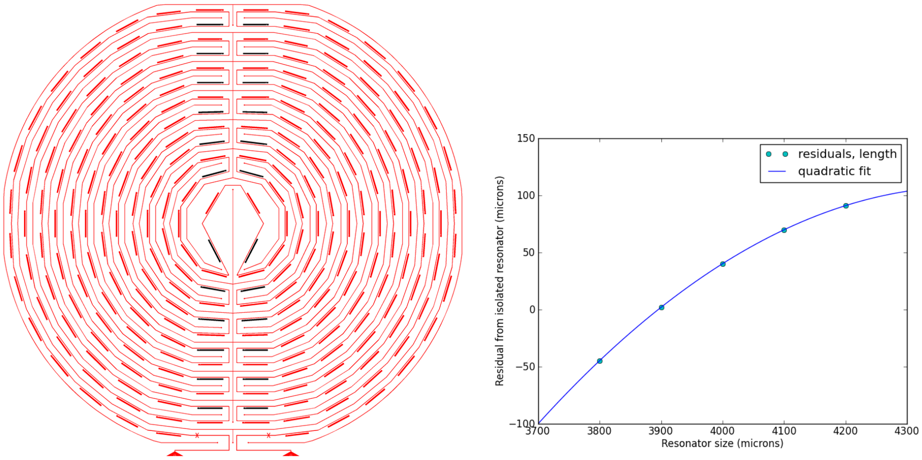


Figure 7.9: Left: resonators where the simulated resonant frequency deviated from the free resonator simulations from section 7.1.3.1 via coupling to the nearby vertical feedline feature. Right: The residual $r = f_{\text{design}} - f_{\text{simulated}}$ between the isolated resonator used to the initial design and the coupling simulation can be modeled by a simple quadratic.

this proved to add unnecessary compilation to analysis, especially if two or more MKIDs were change the order of their resonant frequencies. For the iZID mask we divided our resonators into six equal sub-domains, and alternate between domains as the MKIDs are placed along the feedline, azimuthally.

7.1.4 Mounting Design

Although necessary for scaling our technology to the 10^2–10^3 kg target-mass regime useful for rare-event searches, increased detector size does pose some problems from an energy resolution standpoint. To maintain a fixed metal coverage, and resonator size (set by our readout frequencies), it is necessary to instrument the surface of our detectors with many more individual detectors resonators. While this is beneficial from the standpoint of interaction-position reconstruction, the signal of a given interaction will be divided among more individual MKID sensors, worsening the overall energy resolution as $\sigma_E \propto \sqrt{N_{\text{resonators}}}$. While this may seem irreducible, there mitigating benefits to moving to larger substrates. Recall from equation 6.37 that $\sigma_E \propto \frac{1}{\eta_{ph}}$. To get a handle on our quasiparticle production efficiency, we can compare the phase and dissipation response to a known population of thermally produced quasiparticles and compare to the responses during an interaction event of know energy [110].

Following [110] in the low-frequency ($\hbar\omega \ll \Delta$) and low-temperature ($k_bT \ll \Delta$) limit the temperature dependence of the dissipation and frequency in a resonator have the form

$$\frac{f(T)-f_0}{f_0} = \frac{-\alpha}{4\Delta N_0}\left[1+\sqrt{\frac{2\Delta}{\pi k_b T}}e^{-\frac{\hbar\omega}{2k_bT}}\right] n_{qp} I_0\left(\frac{\hbar\omega}{2k_bT}\right) \tag{7.5}$$

$$\frac{1}{Q_i(T)} - \frac{1}{Qi} = \frac{\alpha}{\pi N_0}\sqrt{\frac{2}{\pi k_b T\Delta}\sinh\left(\frac{\hbar\omega}{2k_bT}\right)K_0\left(\frac{\hbar\omega}{2k_bT}\right)} \tag{7.6}$$

where

$$n_{qp} = 2N_0\sqrt{2\pi k_b T\Delta e^{-\Delta/k_bT}} \tag{7.7}$$

by fitting equations 7.5-7.6 (seen in figure 7.10) to the Q and f of our our MKID's we can determine α and Δ[2]. Equation 7.7 can then be used to convert the frequency response as a function of n_{qp} (insert in figure 7.10). This allows the total number of quasiparticles N_{qp} created for a particular response event to be calculated by observing the frequency shift of all the detectors. For a calibration-sourced event with know energy (E_{calib}) we can calculate the phonon collection efficiency $\eta_{ph} = \frac{N_{qp}\Delta}{E_{\text{calib}}}$. For our proof-of-concept device $\eta_{ph} = 0.07$ for 25 nm films and the relationship to film thickness is shown in figure 7.10. If only 7% of the interaction energy of an event causes quasiparticle production, what are the other energy loss mechanisms? Aside from losing energy into the films, phonons can be lost by thermal contact to the device holder itself. The proof-of-concept devices were directly mounted to a series of ledges machined into the gold-plaited copper devices holder for a contact area of ∼12 mm^2, and at 20 mm^2 the active metal covers only ∼ 1.7 times as much area. To remedy this our iZIDs are mounded using six Cirlex clamps that also have a contact area of ∼12 mm^2, while increasing the active area to 250 mm^2. We expect this to reduce relative losses to the device box to improve by an order of magnitude.

---

[2] For our proof-of-concept $\alpha = 0.074 \pm 0.005$ and $\Delta = 204.4 \pm 2.1$ μeV.

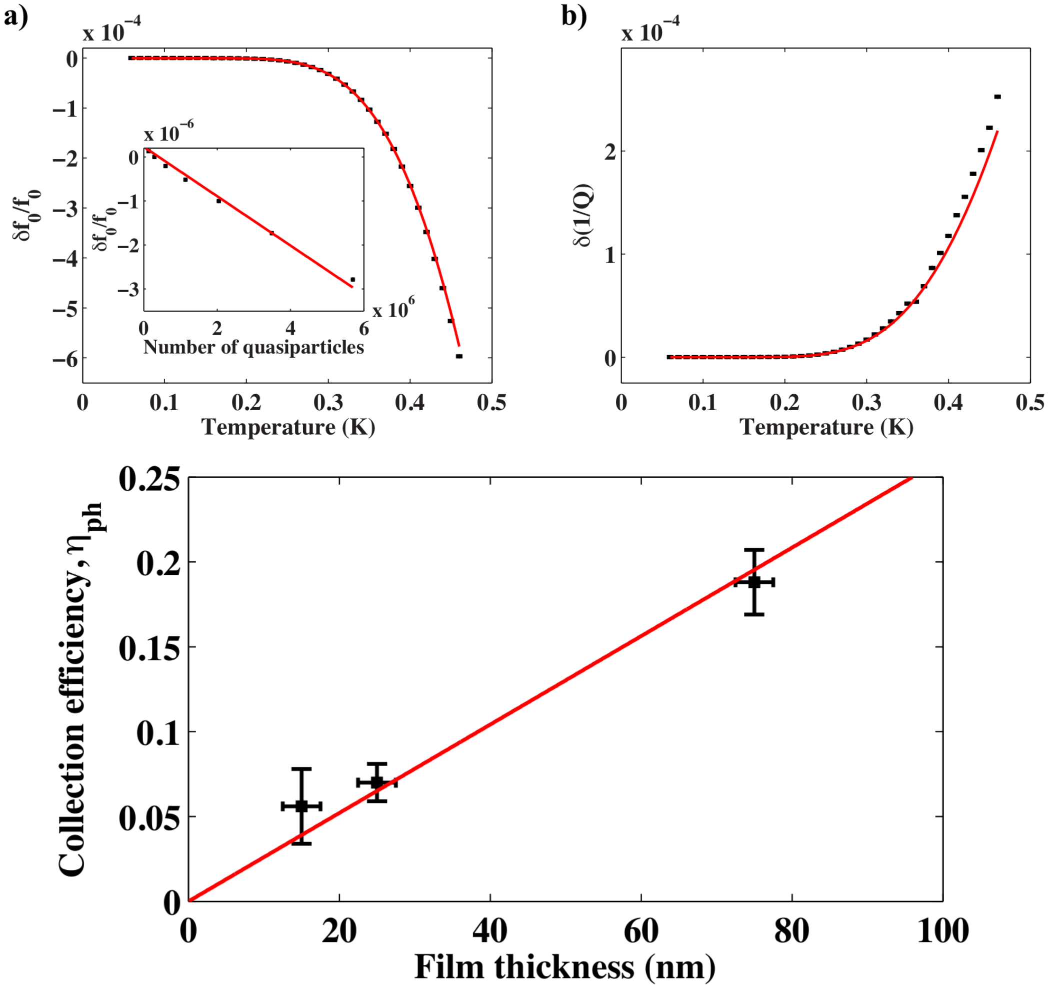


Figure 7.10: Top: The relative frequency (a) and dissipation (b) response from a thermally produced population of quasiparticles in our proof-of-concept devices. These are fit to equations 7.5 and 7.6 (red) respectively. This allows for the number of quasiparticles to be directly calculate from a particular response event (inset of a). Bottom: The measured collection efficiency vs film thickness for Al proof-of-concept devices. We believe this can be much improved in our larger iZID devices.

7.2 Initial Aluminum Devices

7.2.1 Fabrication

All of our devices were fabricated photolithographically[3], using the facilities at the Micro Devices Lab (MDL) at JPL. We utilize this technique to build the thin-film

---

[3] Photolithography is a wonderful, morphologically honest word. *Photo* (light) *litho* (stone) and *graphy* (writing) are all etymologically Greek and seem to indicate a process that writes or prints in stone using light. This is actually a fairly accurate description of the term.

superconducting circuits described earlier in this chapter. A simple outline of the process we use can be described as follows:

- **Substrate Preparation** If we are printing metal circuits we need to print them onto something. Circular wafers cut from semiconductor crystals, such as Si and Ge are typically used, but dielectrics such as SiO_2 are often used as well.

- **Metal Deposition** Next the substrate is coated with a very thin layer of the metal out of which the circuit will be printed. This is commonly done via vacuum vapor deposition. This entails placing our substrate in a vacuum chamber, and filling the chamber with a gaseous metal that is not in thermodynamic equilibrium (a metal vapor). This vapor will then condense on all of the surfaces it encounters inside the vacuum chamber, including our substrate. By regulating details of the vapor production, vacuum strength, and deposition rate, very pure metal films may be deposited on the substrate[4].

- **Photoresist Application** After the metal is deposited a very thin layer of light-reactive polymer called “photoresist” is applied to the surface of the substrate. To achieve a uniform application, the photoresist is usually prepared in a solvent such as methoxy-propyl acetate (PGMEA) so it may be applied to the substrate surface as a liquid. The substrate is then rotated or “spun” at a few thousand rpm, leaving a thin uniform coating of photoresist. Afterward, the solvent is baked off leaving a solid, light-reactive film.

- **Patterning** Next, the pattern of the circuit itself is transfered to the photoresist. This is done by the selective illumination of the photoresist using a specific wavelength of light (typically in the UV range). For a “positive” photoresist, light is blocked, or masked, were the metal is to remain. With a “negative” photoresist, it is the opposite.

- **Development** After patterning, the device is then placed in a bath of a corrosive solution called a developer. The chemistry of a specific developer is selected to match details of the photoresist and circuit, but in short with a positive resist, portions of the photoresist that were exposed to light will become more soluble in the developer. With a negative resist they will become less

---

[4] If the deposition rate is very slow, certain materials may even be deposited epitaxially, that is to say grown as single crystals.

so. After all of the exposed portions of photoresist are dissolved, the device is cleaned with water and dried using compressed air.

- **Etching** At this point the substrate will have a thin metal layer covering its entire surface. On top of this is a pattern of the desired circuit made out of the remaining photoresist. All that remains is to remove all metal that is not protected by photoresist, a process called “etching”. The two basic approaches to this are “wet etching”, where the metal is dissolved in a corrosive bath, and “dry etching”, where the metal is removed by physical ion bombardment, or chemically reacting with the ions. Although wet etching does have its merits[5], dry etching is a more controllable and repeatable process, and it is what was utilized for all devices in this dissertation. After etching the remaining photoresist may be removed. This can be done using a bath of organic solvent such as acetone, or via an additional dry etch step called oxygen ashing.
- **Post-production** At this point, we have printed a circuit on our substrate. For more complicated designs, many such circuits may be printed on top of one another by repeating the above steps. The work of fabrication, however, is not quite over. To test our devices they need to be mounted in a 3in wafer holder as discussed in section 7.1.4, and all circuit connections made via AL wire bonding. If the substrate dimensions are incompatible with the final device dimensions this can involve cutting the device to the required dimensions, commonly referred to as “dicing”.

The devices described in this section were all single-layer 25 nm Al realizations of the design described in beginning of this chapter and shared very similar fabrication technique as described below.

7.2.1.1 Wafer selection and preparation

This series of AL iZID devices was fabricated on the same substrates used for our cm^2 proof-of-concept devices. They are high-resistivity[6] Si wafers that are 1 mm thick and 100 mm, or ~4 inches, in diameter. The crystals used all had a crystal orientation of [100], which is to say that the face of the substrate was cut parallel to a face of the silicon crystal’s cubic unit cell. Ideally, this would minimize dangling

---

[5] Notably the fixed capital requirement is many orders of magnitude below that of dry etching.

[6] $\rho > 5\ \mathrm{k\Omega\ cm}$

surface bonds and hopefully slow oxide accumulation, but in practice both [100] and [111] orientation wafers were used to fabricate our proof-of-concept devices, without much measured difference in MKID response. The wafer dimensions were set to be compatible with the requirements of the Canon EX3 stepper, which was used to pattern the original devices. This set the maximum wafer thickness at 1 mm and diameter at 100 mm. This geometry proved to be suboptimal for our large-format devices and was later changed as is detailed in section 7.4.

To minimize the potential two level system noise in our resonators, surface oxides are removed from the wafer before metal is deposited on it. This process is usually called deglazing, and is accomplished with the application of hydrofluoric acid (HF) to the substrate surface. Two techniques were utilized, although no appreciable difference was noticed in the subsequent fabrication or behavior of the devices. The first deglazing technique, is to immerse the wafer in an a pre-mixed 1:10 dilution[7] HF bath for 30 s. This is done manually and to ensure even surface etching the wafer is agitated by hand for the duration of the bath. Although preformed on an acid bench with many layers of PPE, this potentially exposes the fabricator to HF, which is extremely hazardous[8]. As a slightly safer alternative, an Idonus HF vapor system was also used, which typically takes 3 min (1 each for fill, etch and empty). Testing for oxides is straightforward. Pure silicon is hydrophobic, and a properly deglazed wafer is not wettable.

7.2.1.2 Metal deposition

After deglazing, 25 nm of Al is vacuum deposited on one of the wafer's face using a Kurt J. Lesker load-locked UHV sputtering system. This system utilizes a physical vapor deposition (PVD)[9] technique called DC magnetron sputtering. There is a bit to unpack from that name, but in essence each component of it is referring to the details of the vaporization mechanism of the aluminum. In DC sputtering[10] the Al vapor source, called the "target", and the wafer are placed on opposing sides

---

[7] A "pure" HF bath is usually delivered as 30% gas dissolved in water at STP.

[8] Although it is corrosive, HF is also a contact poison. Rather than suffer immediate burns as in HCl exposure, HF is rapidly absorbed into the body and exposure is not immediately evident. Once absorbed, it binds to free Ca^{2+} and Mg^{2+} ions, and can damage the nervous, cardiovascular, muscular and skeletal systems.

[9] Physical vapor deposition is *physical* in the sense that the vapor is sourced from a pure solid or liquid that is physically vaporized in some way. This is in contrast to a vapor produced by a chemical reaction, which would be called *chemical* vapor deposition.

[10] From the Latin spūtāre meaning *to spit*.

(about 8 cm apart) of a vacuum chamber that is then filled with rarefied Ar gas ($P \approx 5$ mTorr). The wafer and the Al target are then biased with a high DC voltage. Electrons emitted from the highly biased Al target cathode, collide with the nearby Ar gas, ionizing it onto a plasma. The positively charged Ar ions are then driven into the target cathode by the bias field, setting off collision cascades within the Al. If any branch of this cascade interacts with a surface Al atom, and has more energy than the surface binding energy of Al, that atom may be ejected (or "sputtered") into vapor. To concentrate the near-target plasma formation, and to prevent damage to the wafer from high-energy electron bombardment, two concentric rare-earth magnets are placed on the back side of the Al target, one circular magnet in the center of the target, and an annular ring of reversed polarity at the edge. These confine most of the electrons to a toroid near the target surface, and are the "magnetron" portion of the name. This technique allows for a film growth of 43 nm/min, allowing for our 25 nm films to be deposited in $\approx$ 35 s.

7.2.1.3 Photoresist

After deposition, we applied a photoresist called AZ 5214 using a Solar-semi integrated system. This system can handle 100 mm wafers and automates the application, spinning, and baking of photoresists. AZ 5214 is a positive photoresist[11] consisting of a photoactive compound called naphthoquinone diazide and a novolak resin dissolved in PGMEA. The wafter is then spun at 4000 rpm for 40 s, resulting in a characteristic thickness of $\approx 1.4\ \mu$m. Afterward, to fix the resist and bake off all the solvent, the wafer is heated to 90–110 C for 90 s and then cooled.

7.2.1.4 Patterning and development

To pattern the photoresist we utilized a contact mask. Our contact mask is a 1:1 scale pattern of the desired circuit laid out in chrome on a substrate of quartz, which is commercially fabricated for us by Photo Sciences, Inc. This mask is then pressed, chrome side down, against the photoresist on top of the wafer. The backside of the

---

[11] AZ 5214 is actually what is referred to as an "image reversal" photoresist. These are able to switch behaviors from positive resist to negative resit depending on process details. The decision to use this particular class of resists was actually somewhat forward looking and will be discussed in some detail in section 7.3.5

quartz mask is then illuminated for 20 s with a Hg arc lamp[12]. This was conducted on a Karl Suss MA-6 mask aligner which assists in mask and substrate alignment, inspection, and illumination, but is entirely manually controlled. For these single-layer Al devices, alignment was fairly straightforward. For compatibility with our current iZIP infrastructure, our circuit is designed to fit on a 3 in diameter wafer, so on these larger wafers we just attempted to roughly center it, and align the flats. This can be seen in figure 7.11.

After patterning, the device is developed in a aqueous solution of 2.28% tetramethylammonium hydroxide called AZ MIF[13] which is a strong base. The wafer is submerged in developer and agitated for 60 s before being rinsed with deionized water and forcibly dried with compressed air.

7.2.1.5 Etching

Etching was done via a process called reactive ion etching or RIE. To do this we utilize an Unaxis Shuttleline[14] inductively coupled plasma (ICP) etcher system which consists of an oblate vacuum chamber inside of which we place our sample. The flattened top of this chamber contains a large, plainer electromagnet, concentrically coiled around a central viewport window. Gas, typically a reactive halide or oxidizer or some kind, is introduced into the system to a pressure of $P \approx 10$ mTorr and is lit into a plasma by a strong RF ($f_{\mathrm{ICP}} = 2$ MHz) electromagnetic field supplied by the coil. The chamber is electrically grounded, but the wafer is electrically isolated via a ceramic He-gas-cooled coverplate. Under this coverplate is a second, lower-power RF generator which is used to induce a bias field ($f_{\mathrm{bias}} = 13.56$ MHz). As they are lighter and much more mobile, the oscillating bias field causes the electrons in the plasma to collide with the electrically isolated wafer at a much higher rate than the ions, self-biasing the wafer with a negative charge. This bias attracts the positively charged reactive ions in the chamber, allowing for a higher rate, more directional etch. It is important to note that metal removal via physical sputtering is

---

[12] This particular process uses what is commonly referred to as a 400 or “clear” lens. This is transparent for wavelengths $\lambda > 300$ nm.

[13] Developers are generally hydroxides of some kind. The simplest of these utilize alkaline metal cations such as KOH or NaOH and are actively used in photolithographic development. These metallic cations can interfere with certain types of circuit production, such as the growth of an SiO_2 insulator. To emphasize that a non-metallic quaternary ammonium cation, $N(CH_3)_4^+$ is being used this developer is called “metal ion free” or MIF.

[14] As of this writing, I believe that Unaxis Shuttleline is now defunct and owned by Plasma-Therm.

a sub-dominant phenomena, allowing for much tuning of etch behavior and rate by changing the chemistry[15]. Inductively lighting the plasma allows for the formation of very high plasma densities. This large ion flux on the wafer provides fast non-isotropic etches, while minimizing sputtering and damage from ion bombardment. The progress of the etch can be monitored interferometrically via the viewport on the top of the vacuum chamber.

Our process consists of two etches. The first introduces 10 mT of O_2 at a flow rate of 20 sccm, which is lit into a plasma. This oxygen plasma etching, or "ashing"[16], is very selective towards organic matter, and can be used to remove photoresist if desired. In this context, we want to clean, or de-scum, while leaving our photoresist intact. To do this we ash for a short time ($\sim$ 10 s) at a low RF power (100 W ICP and 60 W bias), which leaves the photoresist intact while removing surface contaminates that may interfere with the main etch. After ashing, the oxygen is pumped out and a gas mixture consisting of 20 sccm BCl_3 and 15 sccm Cl_2 is introduced. This is lit using with an ICP and bias power of 350 W and 30 W respectively. This is allowed to etch until the metal is visibly removed for these devices.

7.2.1.6 Dicing

After etching, the photoresist is cleaned using a combination of an ultrasonic acetone bath and a dedicated oxygen ashing system[17]. For resist removal, we use 125 W of RF, and 20 sccm (10 mT) of O_2 for about 2 min. After cleaning, the wafer is coated with a thick protective layer of photoresist which is deposited via low speed spinning.

At this point we have fabricated a 3 inch device onto a 4 in Si wafer. To fit it in our holder we cut the wafer in a process known as dicing. Traditionally this is used to cut out a regular series of chips that were all fabricated on a single wafer[18]

---

[15] The ICP and bias field powers can be independently tuned to achieve both highly chemically reactive etches, as well as highly physical etches. The processes described in this thesis all operate in the former regime.

[16] Although organic material will be oxidized and pumped out of the system as carbon dioxide and water vapor, any stray, non-organic subcomponents will be left behind as a residue of ash.

[17] The make and model of which are lost to the dust bin of history. It does have something of a classic 70's paint job, which is nice. It also has a serious problem with its load/store instructions, and writing values to it's memory is...frustrating.

[18] In the case of our prototype detector, [12?] usable devices would be diced out of a single 100 mm wafer.

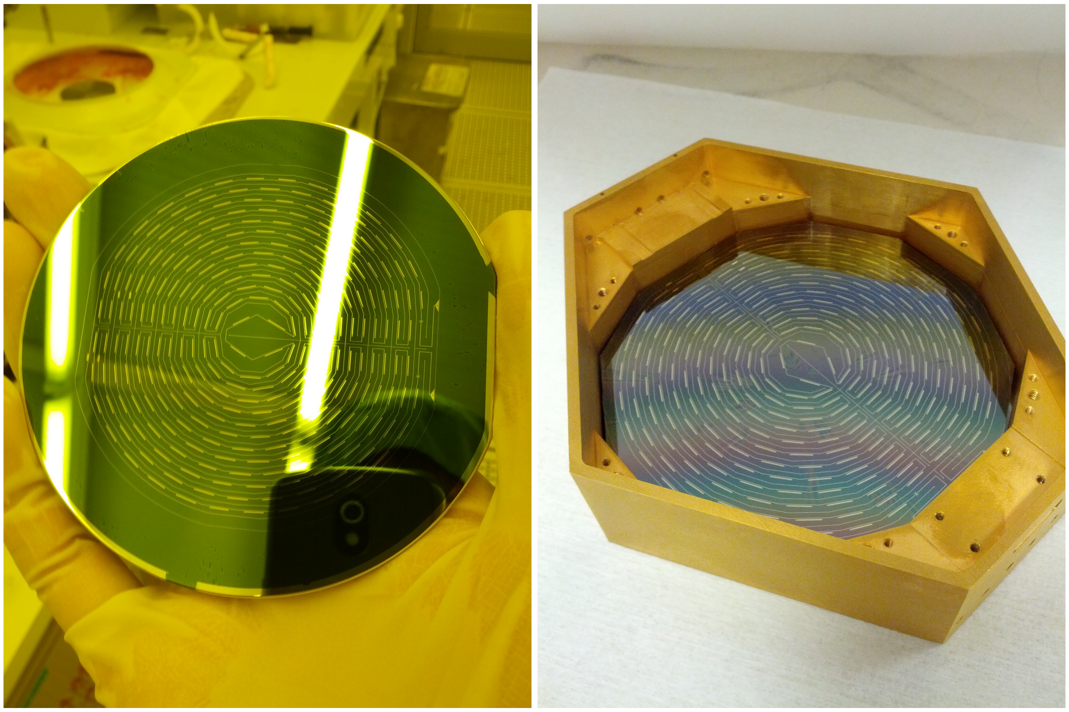


Figure 7.11: Left: The very first iZID-style device that was successfully fabricated. The pattern was developed to be compatible with our current 76 mm iZIPs, but was fabricated on a 100 mm wafer. Right: after dicing, the wafer can be mounted in our low-contact-area box.

and as such the dicer is typically programmed to cut a series of parallel lines in the substrate, rotate it by 90°, and cut another series of parallel lines. Although designed to hold a round 3 inch wafer, the interior of our device holder is actually an irregular dodecigon, and programming the dicer to correctly cut it out requires a great deal of care and manual intervention. Of all steps previously described in our device production process, this was the most error prone, and ultimately unnecessary.

7.2.2 Device Testing

These devices were were cooled and tested in the same Oxford Instruments Kelvinox 25 dilution refrigerator as our proof-of-principle device. The actual testing procedure deviated slightly from the one used for those devices, due to problems with the RF transmission of the devices. To better explain how this occurs I will briefly outline the exact ideal testing procedure for a device of this type.

To use the ROACH-based readout scheme described in section 6.3.2 we need to program our ROACH to generate a comb of frequency peaks, each one of which corresponds to the resonant frequency of a particular MKID. Although we do attempt to engineer these frequencies, there are inevitably modeling and fabrication errors which can lead to drift. As a result, the first order of business with each new device is to measure the resonant frequency and quality factor of each resonator. This is done using a piece of equipment called a two-port Vector Network Analyzer (VNA). A VNA is a combination of a broad spectrum signal generator together with an RF receiver at the import port and a second RF receiver in the output port. These are used in concert to probe the complex transmission (S_{21}) and complex reflection (S_{11}) of a device under test. This is accomplished using a variable frequency carrier wave source that is swept through the bandwidth under investigation. The dual input and output receiver have been very precisely calibrated to this source both in phase and amplitude. The result of this measurement can be fit to equation 6.3 and f_0, Q_C, Q_i, and Q can be extracted[19]. As this is a simple frequency sweep, the bandwidth is quite low and only static parameters may be probed. As a result, this is not a useful technique for observing dynamic phonon-induced quasiparticle populations, but it can be used to measure resonator response due to thermally induced quasiparticle and it is the basis for the $\delta f/f$ measurement described in section 7.1.4.

Our initial Al devices suffered from very poor RF transmission as can be seen in figure 7.12. As a result, individual resonators were very difficult to find and fit in a reproducible and predictable way. Although testing of these devices never progressed to probing individual MKIDs with our ROACH-based system, we were able to tease out certain properties of our MKIDs using analysis of their thermal response.

7.2.3 Average device properties

As mentioned in section 7.1.4, to calculate the kinetic inductance fraction and superconducting gap of our MKIDs, the frequency (and quality factor) of a resonator must be determined at a variety of different temperatures. Given the large RF structure mentioned in the previous section, finding, fitting, and following a particular resonance peak as it responded to the thermally produced quasiparticles, without getting it mixed up with an RF peak of another resonator proved to be intractable. Although we were not able to measure the properties of our individual MKIDs directly, we were able to estimate the average temperature response of the entire array. To do this we turned to cross correlation.

---

[19] For a more exact and detailed description of this fitting please see Appendix A in [110].

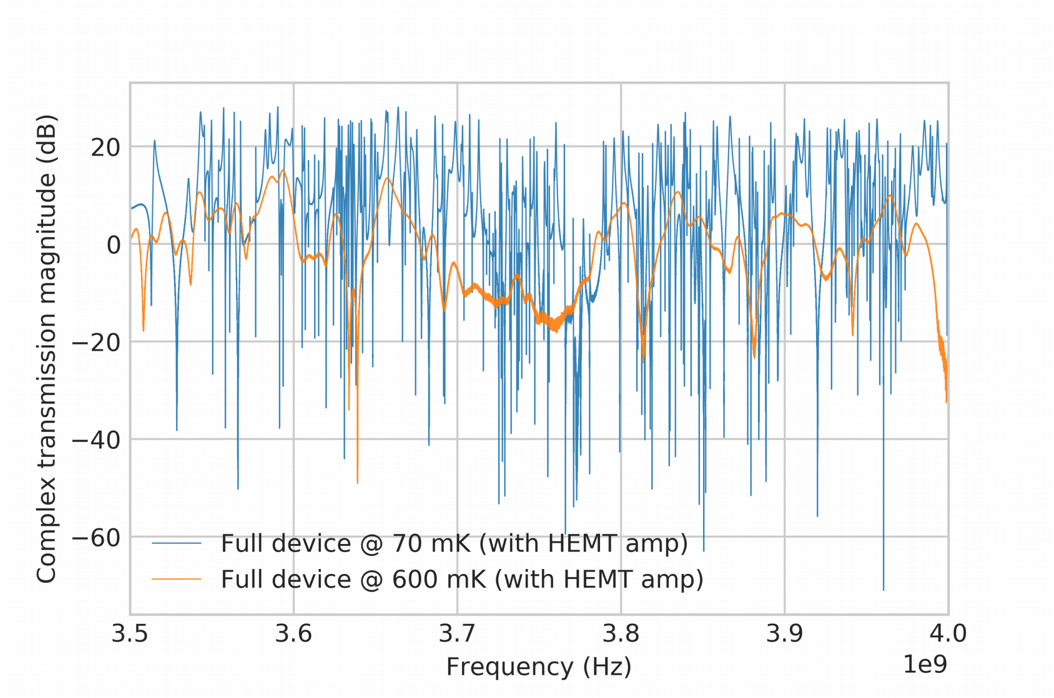


Figure 7.12: The results of our initial testing for our 25 nm Al iZID device. The magnitude of the complex transmission through our feedline is shown at our base temperature of 70 mK (blue) as well as at 600 mK (orange), where dissipation has suppressed our resonator's response. As can be seen, there is significant remaining RF structure. This makes finding and fitting individual resonators very difficult, and eventually lead to re-thinking the entire iZID design.

A cross correlation between two discretely sampled functions, h and g is defined as

$$(h \star g)_n \overset{\text{def}}{=} \sum_{m=-\infty}^{\infty} \overline{h}_m g_{m+n} \tag{7.8}$$

and it informs about the similarity of the two functions as a function of the number of sample offsets. If we let h be the measured magnitude of the complex transmission, or $|S_{21}^{\text{Measured}}|$, through our array and take g to be the identical measurement taken at a higher temperature, the peak of the resulting cross-correlation, $h \star g$, should occur at the offset, n which cancels the frequency shift experienced by the resonant frequencies of the MKIDs caused by thermal quasiparticles. This frequency offset, n, vs. temperature relationship can be fit to equation 7.5. The results of this fit can be seen in figure 7.13 and give $\alpha = 0.023$ and $\Delta = 0.178$ meV, both of which are reasonable for 25 nm Al films.

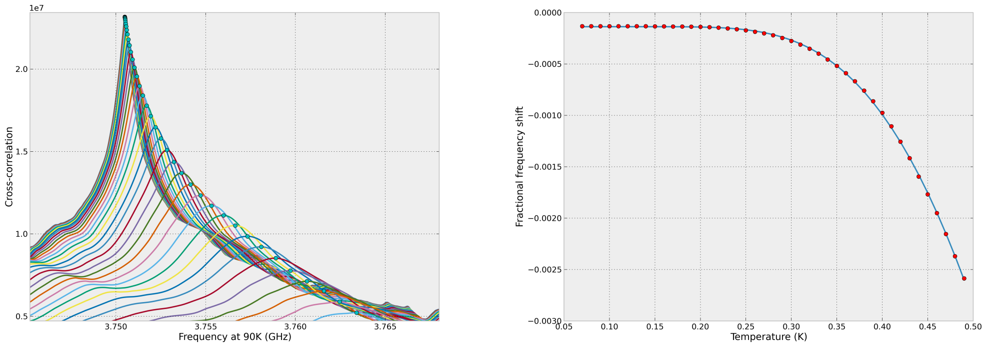


Figure 7.13: Left: The cross-correlations of frequency sweeps at various temperatures against a frequency sweep at 90 mK. They are plotted against 90 mK frequency domain, but the choice is arbitrary. The maximum of each cross-correlation was found and plotted (blue dots). Right: The result of fitting a Mattis-Bardeen curve to the "average" frequency shift as given by the maximized points from the plot on the left.

7.3 Nb Feedline Devices

To isolate the problems with the RF transmission through our device we decided to switch to using Nb instead of Al for our charge-lines and feedline. The higher Nb superconducting transition temperature ($T_c = 9.3$ K) allows for rapid testing in our newly constructed 4 K fridge. This is much faster, and cheaper to cool than our dilution fridge and is a perfect testbed for RF testing of the feedline. We also decided to transition to a two-layer process for our finished devices, so we could keep using Nb feedlines once we got them working.

7.3.1 Unit Test Overview

To test the RF properties of our device and develop an effective two-layer fabrication process we designed four new devices that we colloquially referred to as "unit test" devices. These consisted of

- **Balun test device** This is a simple short feedline segment designed to test the the impedance matching of our baluns as well as our wire bonds and device box.
- **Feedline-only device** Is simply the CPS feedline from an iZID.

- **Feedline and charge-line device** This is the feedline as well as the interdigitated charge-collecting lines.
- **Bi-layer feedline and MKID device** This device is made in two layers. The iZID's MKIDs are on a single contact mask, and the feedline, and charge-lines are on a separate mask.

The designs of these devices can be seen in figure 7.14

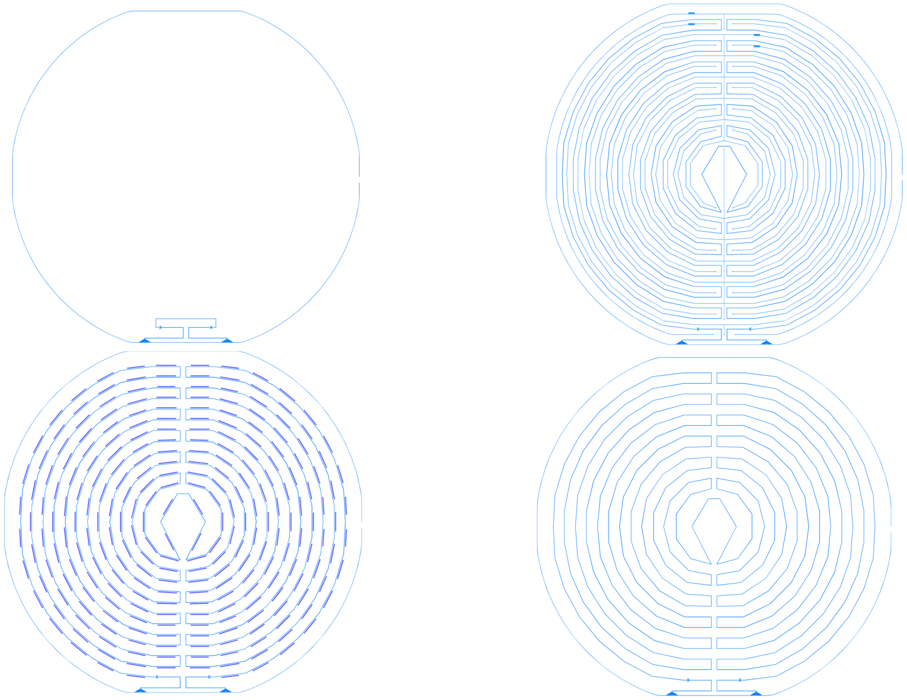


Figure 7.14: Our unit test device masks, clockwise from upper-left: Balun test device, Feedline and charge-line device, Feedline-only device, and Bi-layer feedline and MKID device. These were each used to test a component of our iZID, to better find the source of the large RF structure that was seen in initial single-layer 25-nm-Al devices.

7.3.2 Testing Setup

As it is new I will take a moment to describe our 4k testing setup. Our 4K cryostat is a machined, rectangular "tank"-style system with interior dimensions of 16.5" in width X 16.5" in depth X 8" high. It is cooled by a Sumitomo Heavy Industries RDK-101D Gifford-McMahon mechanical cooler, rated for 200 mW cooling capacity at 4 K. The fridge (and cooler) has two cold stages. The first is an aluminum radiation shield enclosing the cold volume, while the second is a 12" square copper cold plate. The working height above the cold plate is 4". The entire cryostat is mounted to a Lista bench, with the cold head installed inverted from below. The base of the cold head is mounted to a rectilinear, four posted frame assembly that reaches the floor. The base of this frame is weighted to dampen vibrations from the cooler motor. This entire assembly is connected to the cryostat using a steel bellows, and all thermal connections are made with braided copper cable, again to act as vibration dampening. Vacuum connections are provided through the baseplate, with a primary connected to a turbomolecular pump, as well as a secondary for pressure measurement and backfilling. As discussed in section 7.2.2 RF properties of the device were probed using an HP VNA. In this particular setup, as it is only for RF testing, there is no in-line HEMT amplifier.

7.3.3 Balun test device

7.3.3.1 Fabrication

As a sanity check, the first set of these unit-test devices that we fabricated were the “balun test devices”, the design for which can be seen in the upper left of figure 7.14. These devices are designed to test the RF properties of all parts of our readout circuit *except* for the long CPS feedline itself. The components being tested by this device include the SMA connectors and Duroid bond-pads on the box itself, our wire-bonding technique, and most importantly the balun, which transitions our CPW feedline into a CPS feedline and up to this point had only been simulated. Two devices were fabricated on the same day, and followed a very similar procedure to that used in the production of our full Al devices and is summarized here.

- The same ⌀ 100 mm wafers were selected, but were not deglazed with HF. Reducing surface oxides is less important for a simple impedance matching test of this kind, and the wafers themselves were found to be reasonably oxide free. As a result, deglazing was deemed not worth the potential hazard.
- 25 nm of Nb were deposited under very similar sputtering conditions to the 25 nm Al films.
- After sputtering a 1.4 micron layer of AZ 5214 was deposited and baked using the Solarsemi system. The device was then patterned and developed.

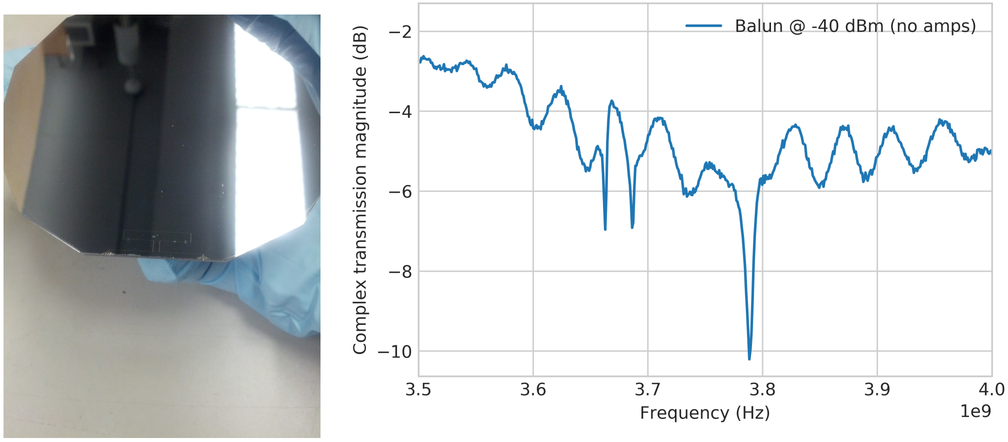


Figure 7.15: Left: Photo of a balun test device after testing. The dangling Al wire-bonds can be seen. Right: Magnitude of the complex transmission of the device on the right. The baseline transmission of this fridge is ≈ -2 dB, and the transmission profile through this device is quite good.

- While the ICP etching system is identical to the one used for etching our first Al devices, the halide chemistry was different. All of our Nb etches utilized a fluorine-based chemistry[20], this one specifically using a mixture of 23 sccm dichlorodifluoromethane, 10 sccm carbon tetrafloride, and 6 sccm oxygen (for 10 mT total).
- After cleaning, the wafer was coated and diced.

A picture of a balun test device can be seen in figure 7.15

7.3.3.2 Testing and Results

After being cleaned ultrasonically in a bath of acetone, and forcefully air dried, the devices were mounted and cooled to 4 K. Only two of these devices were made and they both behaved very similarly. The results of one of the tests can be seen in figure 7.15. In the 3.5-4 GHz band of interest the transmission is nicely flat with an average attenuation of ~ -5 dB, which is not terribly below the ≈ -2 dB attenuation of our fridge.

---

[20] For more on this please see section 7.3.5

7.3.4 Feedline-only and feedline-charge-line devices

After the sanity check provided by the balun device seemed to indicate that the problem was with the feedline (or perhaps the feedline interacting with the charge-lines in some way) we next turned to to our "feedline-only" and "feedline-charge-line" unit test devices. As it turns our RF structure problems lie with the feedline itself, and no real difference of behavior was observed with or without the addition of charge-lines. As a result, this section will incorporate the results and technique from both as they are so similar.

7.3.4.1 Fabrication

The fabrication of these feedline devices was shaped by a persistent problem with low device yields. By this I mean that the majority of devices were either obviously faulty upon inspection, or when tested provided very poor transmission. The crux of the problem is that we fabricate only a single device per wafer, the feedline is over a meter in length, and the central gap is only 2 microns wide, which is roughly the same size as most dust particles encountered at MDL. This creates a situation where any fabrication issue on any part of the wafer can easily cause a short in our feedline. If we were fabricating an array of smaller devices, we could simply choose to select the subset that were functional, but as our devices covers the entire wafer any variation across the wafer was unacceptable. We did eventually develop a procedure that was capable of occasionally producing working feedlines which I will outline below.

- **Wafer selection** One of the biggest changes was our transition to ⌀ 76 mm (or 3 inch) wafers. These were commercial high-resistivity Si and were also 1 mm thick[21]. It was decided that dicing, aside from being error-prone and time consuming, was a penitential sources of shorts. Metal and Si debris can become lodged in the protective wafer coating, and our facilities to clean and inspect wafers on campus (where they are mounted) is very limited compared to MDL. For repeatability, we developed a process where the final cleaning and inspection of our devices can be done at MDL in a cleanroom environment. As far as wafer preparation, we found that HF deglazing prior to film

---

[21] Ideally, we would like to use the thickest wafers we can, as this will best match our actual DM detectors. Previously, our ⌀ 100 mm wafer thickness was was set by our need to use the Canon EX3 stepper, but limit for the Karl Suss MA-6 contact mask aligner is the same.

deposition had little effect on subsequent fabrication or testing of these simple feedline circuits.

- **Film Deposition** Deposition was done as described in in section 7.2.1.2, but as with the balun test devices, a Nb target was substituted for the Al one. After deposition, the entire surface of the wafer was microscopically inspected. With such a thin layer of Nb, it was occasionally the case that areas of incomplete deposition or "pits" could occur. These are small micron sized areas of exposed Si, where there is no metal. If they were found the film could be re-deposited, or the location of the pit could be noted and avoided during patterning.

- **Photoresist Deposition** The smallest wafer accepted by the automated Solarsemi spinning system without modification is 100 mm. Although we could bond our 76 mm wafer to a 100 mm carrier with wax, we had better success spinning photoresist by hand using the Solitec manual spinner systems. During this transition is was also noticed that AZ 5214 has a tendency to coagulate as it ages. These clots are typically 1-10 microns in size and can interfere with patterning in two ways. First they are less soluble in developer than normal resist, and can cause unwanted metal to remain after etching. Second, they can interfere with the spin itself, causing azimuthal inconsistencies in the thickness of resist. To decrease the number of these clots we first run our AZ 5214 through a 0.2 micron Teflon filter using a syringe. After baking, we then do a through microscopic inspection of the resist. If any clots are found that could potentially interfere with our feedline pattern, we remove the resist and spin a new coat.

- **Patterning and Development** Patterning was done using the same Karl Suss MA-6 mask aligner system that was described in section 7.2.1.4. Before patterning a series of devices, the quartz/chrome contact masks were cleaned in a bath of acetone followed by Micro-90[22] and deionized water. Although this would ensure a very clean and defect-free patterning, occasionally the contact mask would become fused to photoresist and would have to be removed using a solvent. To prevent this a thin layer of isopropanol was applied to the contact surface of the mask and allowed to air dry. This thin solvent film was

---

[22] This is relatively mild multipurpose alkaline cleaning agent made by the International Products Corp. I am not exactly sure what it is made out of, but it cleans very well.

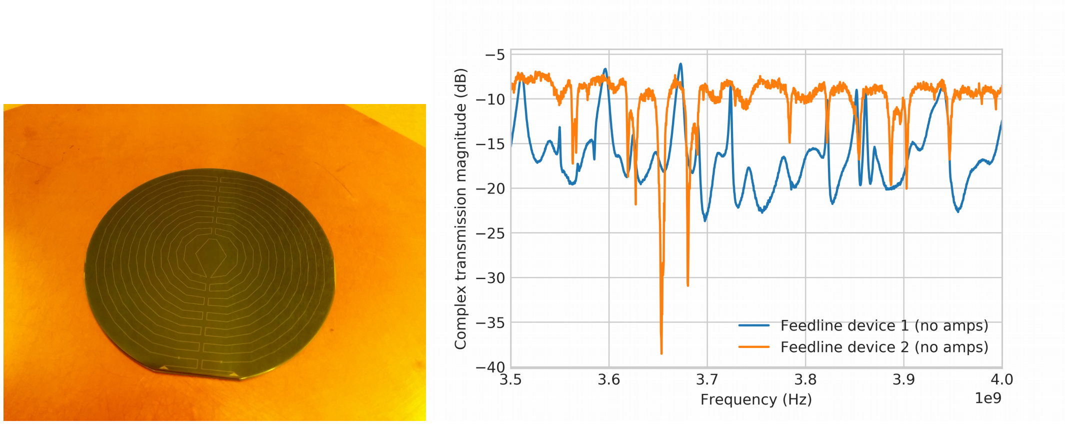


Figure 7.16: Left: an example of a 25-nm Nb feedline-only device. A number of these inspected very well, but upon testing 50% had very poor transmission. Right: a plot of the magnitude of the complex transmission of two devices that inspected identically well. Device 1 (blue) had very poor transmission and would be unusable in a device with MKIDs. Device 2 (Orange) does have a few residual RF resonances, but there is enough flat frequency domain that the majority of MKIDs on such a device could be found and fit.

completely transparent and did not interfere with patterning, but did prevent the mask and device from fusing.

- **Etching** The main etch was conducted in the Unaxis Shuttleline ICP etcher using the same Freon-12 / carbon tetrafloride chemistry discussed in section 7.3.3.1. The pre-etch ashing[23] was conducted in the dedicated ashing system, rather than as an initial step in the ICP etch recipe. This was to allow for a final cleaning of any residual ash using a compressed air gun, as well as a final microscopic inspection of our device. At this point the entire process is still reversible; the wafer could be ultrasonically cleaned and returned to its post metal deposition state. After etching, however, the surface of the wafer will no longer be flat, and even if the Nb feedline were removed, re-fabrication of the feedline would not be possible.

- **Post-fabrication** After etching, the device would be cleaned and throughly inspected. It would then be packed in a wafer caddy for transportation to our on-campus testing facility for mounting, wire bonding, and testing.

---

[23] Or more properly the pre-etch oxygen de-scumming.

7.3.4.2 Testing and Results

The majority of the feedline tests devices were faulty upon inspection, but those that appeared to be defect free were brought to campus and tested using the same dry 4K fridge and HP VNA described in section 7.2.2. The results of these simple transmission tests was plagued by low yields. Although some devices with usable feedlines were produces, most devices that inspected well had persistent problems with transmission. This is shown in figure 7.16. For this style of detector to work simple fabrication is a critical design criteria, and as a result we decided to re-think our approach to a full-wafer detector, which is discussed in section 7.4.

7.3.5 Bi-metal Device

Concurrently with our fabrication and tests of our 25 nm Nb feedline-only devices we began the development of a process that would allow our feedlines to be fabricated in Nb, but our resonators to be made out of Al. In general there are two ways to go about this. Either the Nb feedline can be fabricated first, and then the Al MKIDs added as a 2nd layer, or it can be done the other way around. Each approach has strengths and weaknesses, but both processes were successfully used to produce devices that, at least by eye, inspected well. Sadly both suffered from the same problem of low feedline yields that was discussed in the previous section. Because of these low yields, we decided to change course with our large-format device design before we produced a working example of the iZIP-style layout discussed in this section. As a result the remainder of the section will focus on the fabrication methods developed to produce these devices.

7.3.5.1 Standard Fabrication

This procedure is very similar to the standard Al film procedure described in section 7.2.1 followed by the Nb procedure from 7.3.4.1. An abbreviated procedure is as follows.

- **Al Deposition** 25 nm of Al is deposited using the same procedure as in section 7.2.1. During the same vacuum cycle, a thin, 3 nm, layer of TiN[24] is

---

[24] This is done using a pure Ti target, which is sputtered using a mixture of Nitrogen and Argon. The Ti vapor reacts with the Nitrogen gas to produce TiN vapor which condenses on the substrate.

deposited directly on top of it. TiN is a hard-wearing ceramic, and serves as a protective layer to the relatively soft, very thin, and easily damaged Al.

- **Spinning, Patterning, and Developing** Applying, patterning, and developing the photoresist is exactly as described in section 7.3.4.1.

- **Al Etching** The etch consists of the pre-etch ash in the dedicated ashing system followed by the same borontetracolride / chlorine etch from section 7.2.1 in the ICP etching system. The device is then cleaned in the asher followed by a water bath. Ideally this will react with any residual Cl from the etch, protecting the remaining Al from further degradation.

- **Nb Deposition** After cleaning, 25 nm of Nb is deposited on the wafer.

- **Spinning, Patterning, and Developing** Although identical to the first layer, care must be taken to be sure that the alignment of the second layer is consistent with the positioning of the first. This is done on the Karl Suss MA-6 microscopically by hand.

- **Nb Etching** Similarly to the Al etch from section 7.2.1 this etch utilizes a fluorene based chemistry. Fluorene etches Nb and TiN at a much faster rate than it etches Al. To exploit this specificity and to protect the Al MKIDs from Cl etching the Freon 12 is removed from the etch leaving a carbon tetrafloride (40 sccm) and oxygen (2 sccm) chemistry at 10 mTorr. The ICP and bias powers were 400 and 40 W respectively. The device can now be cleaned in either the asher or an acetone bath.

Results can be seen in figure 7.17. This production method is the easiest, and some variant of it is probably what we will want to use for production of two-layer devices in the future. However, it does expose fragile Al to many fabrication steps, and it also prevents us testing the feedline in any way before we go through the trouble of making the MKIDs (remember that getting a good feedline can require lots of re-fabrication and wafer handling). This is mitigated to some degree by the TiN coating, but alternatively, we can fabricate the layers in reverse order.

7.3.5.2 Liftoff Fabrication

In a perfect world, we would like to do the same process as above but be able to fabricate the Nb feedline layer first. As we could see from our normal Nb feedline process, it often needs to be re-done in order to produce working device. Ideally this could be done without the risk of damaging the Al MKIDs. Sadly we do not have access to any etches that chemically selects for Al, and not Nb, so an etch through Al cannot be made to predictably stop[25] when it reaches the Nb feedline. To get around this we utilize a process knows as *lift-off*.

- **Nb Feedline Fabrication** First, the Nb feedline is fabricated on a clean substrate using the same procedure as in 7.3.5.1. The device is then cleaned.
- **PMMA Application** Before any Al is deposited a $\sim$ 100 nm layer of poly(methyl methacrylate)(PMMA) is applied via spinning at 4000 rpm for 45 s. It is then set by baking at 180°C.
- **Resist Application** After the PMMA is set, a 1.4 micron layer of AZ 5214 is spun on the surface. And fixed by baking at 110°C for 50 s.
- **Patterning** The device is then patterned with the Al MKID pattern. At this point it should be noted that there is no Al currently present on our device, just our Nb feedline below a $\sim$ 100 nm layer of PMMA below a 1.4 micron layer of AZ 5214.
- **Image Reversal and Development** After patterning our resist is “reversed”. This reversal consists of a bake at 120°C for 2 min followed by a re-exposure of the entire substrate to UV light, a process typically referred to as a “flood” exposure. AZ 5214 is an interesting resist in that it contains a cross-linking agent that is activated by being heated above 110 °C but only in areas of resist that are exposed to UV light. This renders the previously-soluble UV-exposed portions of the design insoluble in developer. The subsequent flood exposure acts on the previously-unexposed areas and renders them soluble. These steps together act to reverse AZ 5214 from a positive resist to a negative resist. This “image reversal”(IR) is actually specifically designed to facilitate lift-off. Typically with a positive resist, as the UV light passes through the resist is interacts with and is attenuated by the photoactive compound[26] (PAC). This so-called “bulk effect” causes the top layer of photoresist to be more soluble in developer than the bottom layer, leading to a positively sloped edge profile of 75 to 85°after development. The effect is reversed when the resist is used

---

[25] In reality just drastically reduces its etch rate.

[26] Naphthoquinone diazide in the case of AZ 5214

in IR mode; areas of the resist initially exposed to more light undergo a higher degree of cross-linking and become less soluble in developer. This leads to an overhung, or negatively sloped edge profile, which facilitates the lift-off process as described below. After reversal, the device is developed normally, with a 60 s bath in AZ 300 MIF.

- **PMMA Etch** After development the entire surface of the device will be covered in resist save for the pattern made by the MKIDs. The next step is to etch away the PMMA under-layer where it is exposed, leaving bare Si exposed in a pattern that matches the desired location of the Al MKIDs. To do this we use the dedicated oxygen asher at 125 W with a 150 V bias, a flow rate of 20 sccm 0_2 at 15 mTorr for about 120 s. Although this process will etch the resist layer it does so at a very slow rate allowing for the preferential removal of the exposed PMMA. A picture of a device after PMMA has been removed in this way can be seen in figure 7.17.

- **Al Deposition** At this point the device is loaded into the sputtering system and a 25 nm film of Al is deposited on the surface. As is the case with all sputtering systems, Al should condense on all portions of the wafer in a direct line of sight to the target. This includes on top of the photoresist, and at the exposed Si at the bottom of the MKID-shaped trenches. Because we we used an IR resist, however, the sidewalls of the PMMA trenches will be in the shadow of the negatively-sloped resist overhang, leaving them exposed, and a discontinuity in the Al film.

- **Al Lift-off** After deposition, the wafer is placed in a warm acetone whirlpool bath for a few hours. The acetone attacks the exposed PMMA sidewall and undercuts the photoresist and Al film that are deposited above it. This allows the unwanted Al to be literally lifted off of our device by the solvent, leaving Al only where it was deposited on bare Si.

The benefits of liftoff are that it allows for feedline testing and protects the delicate Al from subsequent development steps, but the exact edge geometry is less controlled than with a traditional etch.

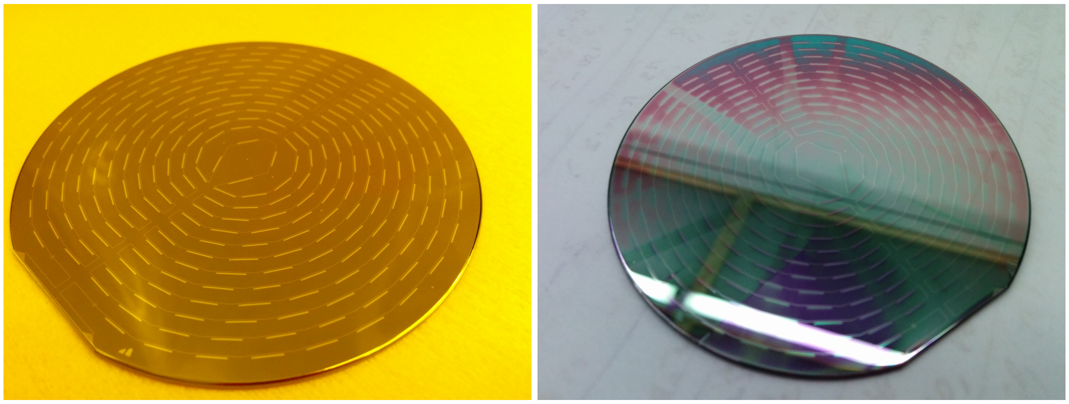


Figure 7.17: Left: Bi-layer device after "standard fabrication" is complete. Although the most straightforward, this exposes our delicate 25 nm Al MKIDs to may subsequent processing steps. Right: Liftoff device midway through fabrication. Here the feedline has already been fabricated and MKIDs are deposited into the etched pits that can be seen in the photo.

7.4 Simplified Phonon-Only Device

7.4.1 Forgoing Ionization Readout

For the ∼30 year history of the field of direct dark matter detection, it has focused on WIMP masses in the 10 GeV to 10 TeV millennium, and for interaction cross sections weaker than the weak interaction [119]. In part, this focus was driven theoretically, where the lightest neutral supersymmetric particles would make a natural WIMP candidate [32]. However, lack of evidence from accelerator experiments has cast some doubt on this mechanism [120]. More recently, theorists have noted that because the DM and baryonic matter have similar energy densities the relic density may be generated by an asymmetry related to CP-violation in baryons [121]. This would relate the baryonic and DM number densities leading to model of DM of a few GeV. Probing for low-mass WIMPs requires sensitivity to very low energy interactions. Recall from equation 2.15 that the total collected phonon energy comes from three sources, with the Luke phonon energy $E_{\mathrm{Luke}} = n_{eh} e V_b$ being due to the drifting of electron-hole pairs across our detector. During the traditional operation of an iZIP, $V_b = 4$ V which allow E_{Luke} to me estimated from the collected ionization signal. Below recoil energies of a few keV the ionization energy measurement starts to become lost in the ionization readout noise. In order to get around this energy threshold problem, in 2012 we developed a new operational mode for our iZIP detectors *in situ*. This consists of highly biasing our detector near its breakdown

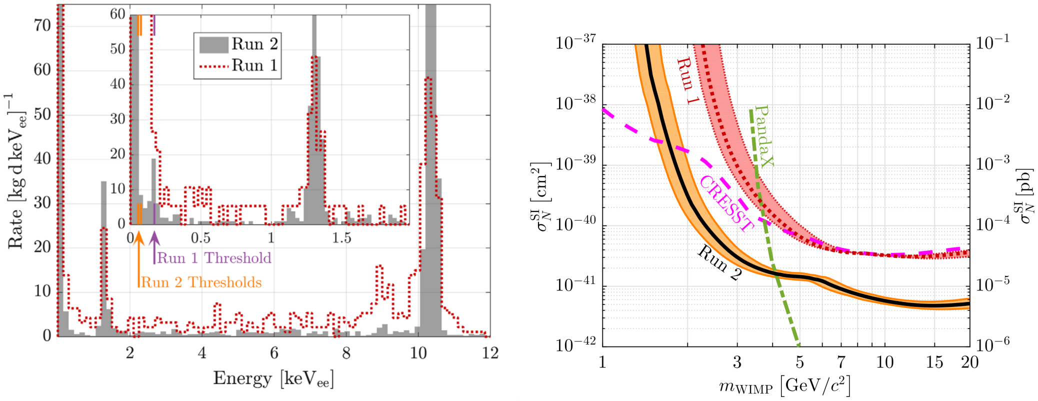


Figure 7.18: CDMSlite results. Left: the efficiency-corrected spectrum of events inside the signal region for Run 1 (red dotted) and Run 2 (gray shaded. The 1.03 and 10.37 keV ^{71}Ge-activation peaks are visible. Right: Spin-independent WIMP-nucleon cross section 90% C.L. upper limit for Run 1 (red dotted with red uncertainty band) and Run 2 (black with orange uncertainty band). Results from the CRESST-II [93] and Panda-X [122] experiments are also shown.)

potential ($V_b \approx 70$ V) and using the resulting amplified E_{Luke} signal to estimate the number of electron-hole pairs produced during the initial particle interaction n_{eh}. Although a direct ionization energy measurement is below the detection threshold, the amplified Luke phonon measurement allow us to substantially reduce our threshold for resolving low energy interactions. The first proof-of-principle data-collection period called *Run 1* collected 6.25 kg d of exposure (9.56 kg d raw) and a larger dataset from 2014 called *Run 2* collected 70 kg d (80.26 kg d raw). Because there is only one energy measurement, that from our phonons, we lose both our ability to discriminate between signal and background based on ionization yield, as well as our ability to preform the clean ionization-based fiducialization discussed in the first half of this dissertation. As a result hundreds of events were present in the signal region after unblinding. Although this reduces the the strength of the limit that can be set, the low thresholds allowed the Run 2 result to set the world-leading spin-independent WIMP-nucleon cross section from WIMP masses in the 1.5-4 GeV/c^2 range. This result can be seen in figure 7.18.

This operational mode was so successful that a new style of detector was developed that forgoes ionization collection completely. These use biased QET's and are referred to as High Voltage (HV) detectors. In the future upgraded SuperCDMS experiment that is planned for deployment in SNOLAB, fully half of the 24 detectors will be HV. With this new interest in phonon-only collection, as well as a desire to simplify our iZID we decided to pivot away from a full iZIP style detector and instead focus on an MKID-only design.

7.4.2 Design Considerations

After the persistent problems with our iZID devices, our main re-design consideration was to increase the yields and RF performance of our feedlines. There multiple areas where we felt this could be improved.

- First, the local dipole field created by a CPS feedline extends further than that of a co-plainer waveguide (CPW) feedline. We suspected that part of the attenuation and RF structure experienced by our working iZID feedlines was due to this dipole field coupling to our metal device holder. As a result we decided to move to CPW feedlines.
- Second, we wanted to keep the minimum feature size $\geq$ 10 microns. This will both reduce rates of incidental shorting, as well as make such potential shorts easier to detect using automated tools.
- As mentioned in section 7.1.2 our iZID has a feedline layout that is designed around a pair of interdigitated, azimuthally-symmetric ionization collectors. Forgoing this ionization detection allows us to return to the simple rectilinear geometry of the proof-of-concept. This allows for simpler fabrication and testing, and also decreases the length of the feedline.
- As seen in figure 7.10, thinner Al films make a trade off between an increase in kinetic inductance fraction (α), and a decrease in the phonon collection efficiency (η_{ph}). It was found that 25 nm Al was optima for our proof-of-concept, and as a result these were the films we used in our iZID devices. In our movement to 2-layer devices that utilize Nb feedlines, however, this film thickness was not fully reconsidered. Thicker feedlines are much less likely to suffer from pitting or other fabrication defects, and as a result we decided to increase our Nb film thickness to 300 nm.

Although our main concern was with the feedline, relaxing the charge collecting condition allowed us much more room for any potential MKID design. As a result we decided to transition back to a lower aspect ratio resonator that should be more uniform in current density, more lumped element, and more similar to the original prototype.

7.4.3 Feedline Test Devices

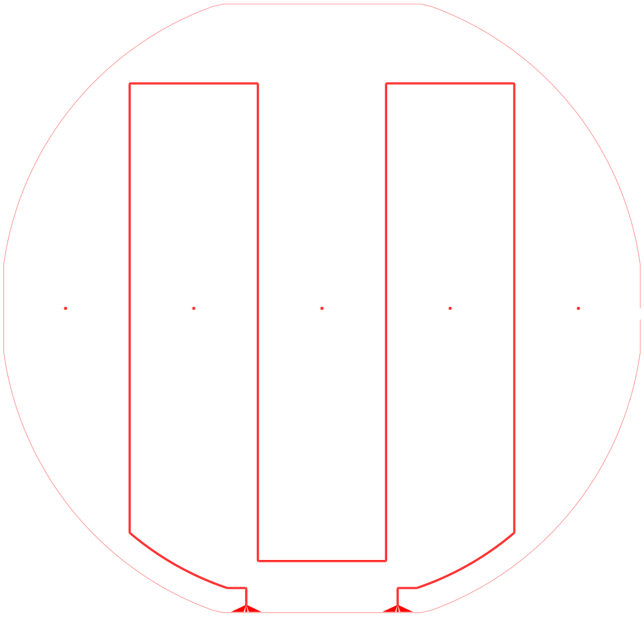


Figure 7.19: An example of a simple CPW feedline test device. The overall consideration was to cover as much of the wafer surface as possible using a short grid layout. This example with 4 vertical sections was usually referred to as a “fork”.

In a departure from the development of our iZID devices, the first series of HV-MKID style devices we produced consisted of simple CPW feedlines, an example of which is shown in figure 7.19. The design of this CPW feedline was driven by a desire to have a characteristic impedance of 50 Ω, and a gap of at least 10 microns. As mentioned in the previous section we had also determined to fabricated these feedlines out of 300 nm Nb films. Assuming L_s = 0.13 pH/sq [123], we again used SONNET to determine a center strip:gap ratio. There was some concern that these CPW feednlines would require a set of complex grounding ties to prevent a “slot mode” were both of the finite ground lines and center conductor are oscillate in lock-step at the same voltage. In a sense the entire conductor-finite-ground-plane assembly would act as a microwave stripline against the ground-plane of our device holder. These devices were fabricated using the same technique as described in section 7.3.4.1, with the metal deposition being ten times longer. Aside from being generally easier to inspect and fabricate reliably, it was found that the RF structure of these feedlines could be drastically reduced by reducing the spurious coupling to the device holder. This was done both by cooling and testing our device with cover plate to the device box removed[27], as well as with a layer of Ecosorb material installed just inside the lid. The results of these transmission tests can be seen in figure 7.20.

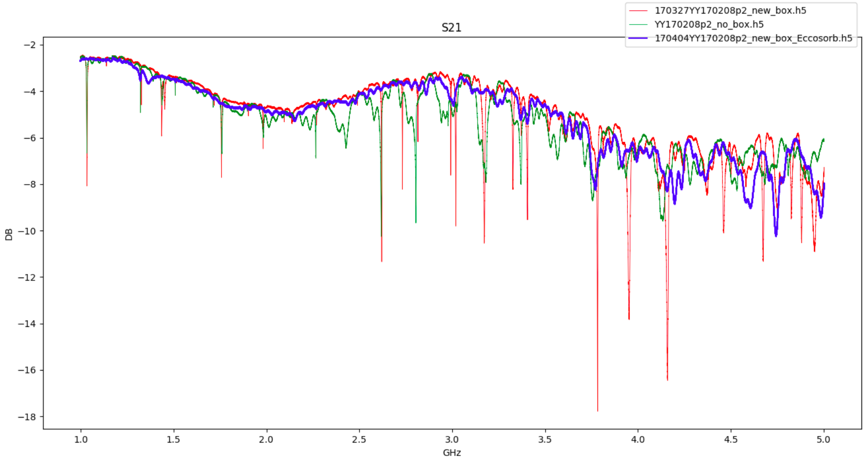


Figure 7.20: The magnitude of the complex transmission through the “fork” device depicted in figure 7.19. This device was tested with the device box entirely closed (red), with the lid of the box removed (green) and with a layer of Eccosorb installed in the lid (blue). Courtesy of Yen-Young Chang

7.4.4 Single-layer 300 nm Nb HV-MKID

After the relative success of our new CPS feedline, we decided to design a series of large format phonon-only devices. Part of the problem with our iZID devices was the early introduction of incidental complexity in the design, and as a result we wanted to keep these HV-MKID devices as simple as possible. Although our ROACH-based readout hardware would allow for up to 250 MKIDs on a device, we judges it to be prudent to start simpler. To this end wrote a program that lets us define the number of snakes of our feedline, and the spacing of our MKIDs and it will layout a mask with the positions of the MKIDs and feedlines optimized. In this way, we can automate increasing the complexity of our device. An example of such a device with 10 feedline meanders and 80 MKIDs is shown in figure 7.21. It is important to note that this style of tunable complexity is much simpler due to the rectilinear nature of these devices.

---

[27] Although this does reduce the feedline coupling to the box, it also exposes the device to potential IR radiation and is not a viable longterm solution.

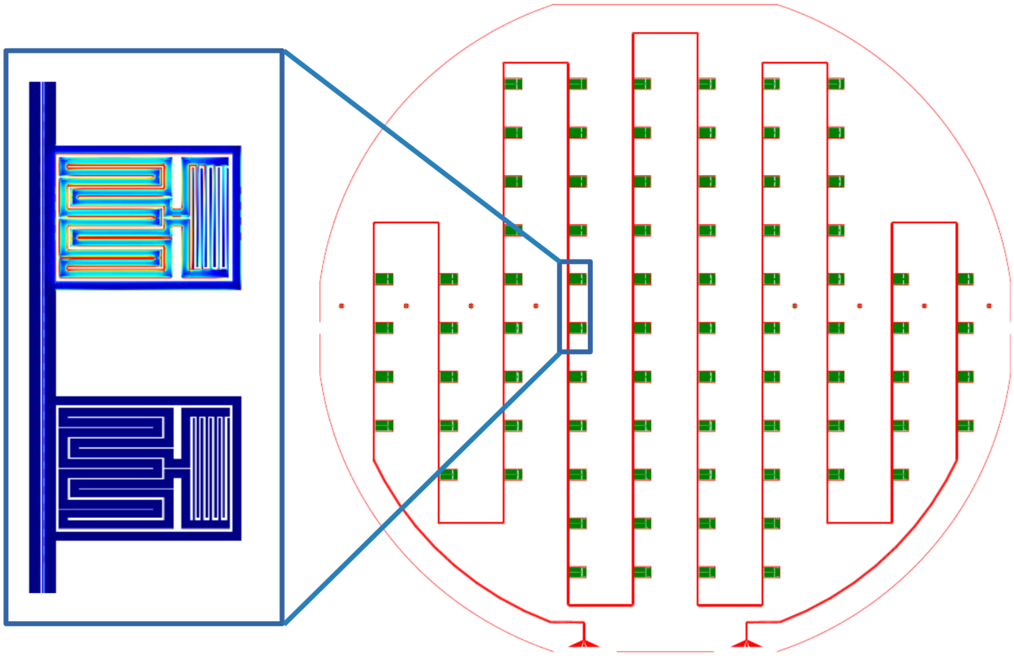


Figure 7.21: Layout of our 80-resonator HV-MKID device. Detail is a SONNET simulation testing for resonator cross-coupling. The new-ground shield together with the “fold-over” design of the resonators proved very successful as preventing any cross-talk.

The resonator design, device fabrication, and testing of these HV-MKID devices was lead by Yen-Yung Chang, and more detail can be found here [124]. In an attempt to end this dissertation on a positive note I will summarize the results.

The MKID’s used in these devices were designed incorporate components from both the iZID as well as the proof-of-concept’s resonator design. They are low-aspect ratio, lump-element resonators with highly uniform currents ($>95\%$ current uniformity), but also utilize the same “fold-over” topology as an iZID MKID so that all nearest neighbor currents tend to cancel. To further prevent inter-resonator coupling, the finite ground-planes of our CPW are extended around each MKID in a design inspired by [125]. For the geometry of these resonators see figure 7.21. For this device, 80 MKIDs were designed to be evenly spaced in the 3.05–3.45 GHz band.

These devices are designed to be two-layer, with the MKID’s fabricated out of 25 nm Al and the feedline out of 300 nm Nb. The first examples of these HV-MKID devices, however, were fabricated out of a single Nb layer. We do not expect any responsivity out of these very high T_C MKIDs, but the RF properties, such as relative resonance location, and identification should inform our future effort. This single thick-layer fabrication was very similar to that described in section 7.3.4.1 but film deposition was ten times longer (300 nm film) and uses Fuji GKR-6760 photoresist, which is a 300 nm deep UV resist. We undercoat with PMMA and then spin at 1400 rmp for 60 s. Apply at 600 rpm to prevent mixing with PMMA. UV exposure was determined to work well above about 90 s.

The complex transmission through this device is shown in figure 7.22. All 80 resonances could be identified and fit. A and two-layer examples of this device (with a 300 nm Nb feedline and 25 nm Al MKIDs) are currently being fabricated and tested, and I am hopeful that a working device will be forthcoming soon.

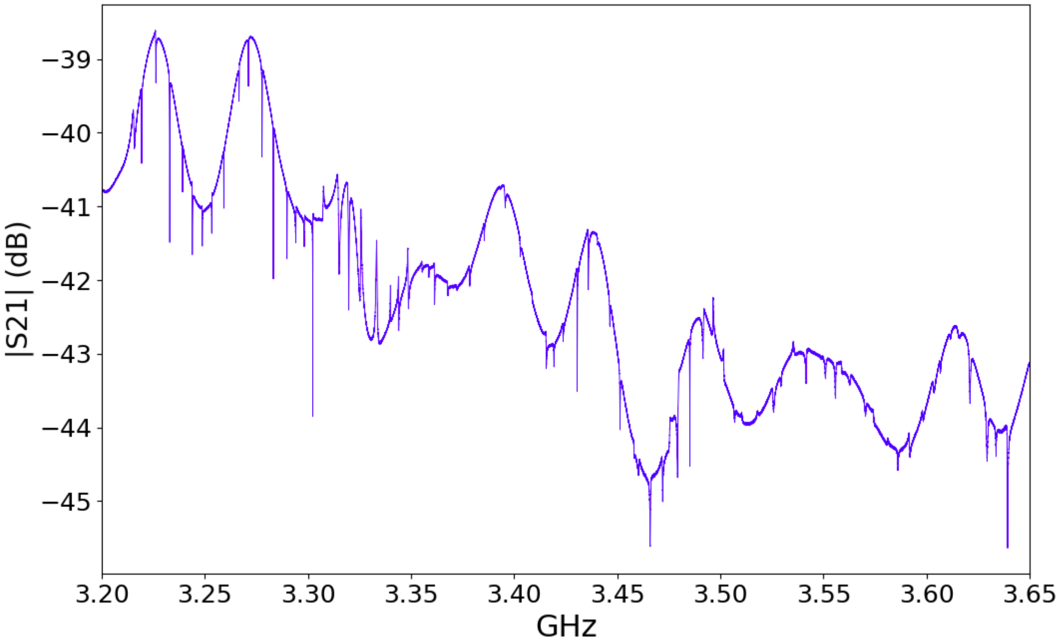


Figure 7.22: The magnitude of the complex transmission through our single-layer 300 nm Nb HV-MKID device. Although the Nb MKIDs have substantially depressed quality factors when compared to the 25 nm Al design material, this provides a wonderful test of the RF properties of this style of device. Each of the 80 resonators can be identified and fit, and two-layer examples of this device (with a 300 nm Nb feedline and 25 nm Al MKIDs) are currently being fabricated and tested.

BIBLIOGRAPHY

[1] R. Agnese et al. “Results from the Super Cryogenic Dark Matter Search Experiment at Soudan”. In: *Phys. Rev. Lett.* 120 (6 Feb. 2018), p. 061802. DOI: `10.1103/PhysRevLett.120.061802`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.120.061802`.

[2] Joshua A. Frieman, Michael S. Turner, and Dragan Huterer. “Dark Energy and the Accelerating Universe”. In: *Annual Review of Astronomy and Astrophysics* 46.1 (2008), pp. 385–432. DOI: `10.1146/annurev.astro.46.060407.145243`. eprint: `https://doi.org/10.1146/annurev.astro.46.060407.145243`. URL: `https://doi.org/10.1146/annurev.astro.46.060407.145243`.

[3] P. A. R. Ade et al. “Planck 2013 results. XV. CMB power spectra and likelihood”. In: *Astron. Astrophys.* 571 (2014), A15. DOI: `10.1051/0004-6361/201321573`. arXiv: `1303.5075 [astro-ph.CO]`.

[4] *ESA Science and Technology: Plank’s new cosmic recipe*. `http://sci.esa.int/planck/51557-planck-new-cosmic-recipe/`. Accessed: 2018-05-01.

[5] M. Roos. “Dark Matter: The evidence from astronomy, astrophysics and cosmology”. In: *ArXiv e-prints* (Jan. 2010). arXiv: `1001.0316 [astro-ph.CO]`.

[6] K. G. Begeman, A. H. Broeils, and R. H. Sanders. “Extended rotation curves of spiral galaxies - Dark haloes and modified dynamics”. In: *mnras* 249 (Apr. 1991), pp. 523–537. DOI: `10.1093/mnras/249.3.523`.

[7] Vera C. Rubin, W. Kent Ford Jr., and Norbert Thonnard. “Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa through Sc”. In: *Astrophys. J.* 225 (1978), pp. L107–L111. DOI: `10.1086/182804`.

[8] Joshua D. Simon and Marla Geha. “The Kinematics of the Ultra-Faint Milky Way Satellites: Solving the Missing Satellite Problem”. In: *Astrophys. J.* 670 (2007), pp. 313–331. DOI: `10.1086/521816`. arXiv: `0706.0516 [astro-ph]`.

[9] M. Cappellari et al. “The SAURON project - IV. The mass-to-light ratio, the virial mass estimator and the Fundamental Plane of elliptical and lenticular galaxies”. In: *mnras* 366 (Mar. 2006), pp. 1126–1150. DOI: `10.1111/j.1365-2966.2005.09981.x`. eprint: `astro-ph/0505042`.

[10] Louis E. Strigari et al. “A common mass scale for satellite galaxies of the Milky Way”. In: *Nature* 454 (2008), pp. 1096–1097. DOI: `10.1038/nature07222`. arXiv: `0808.3772 [astro-ph]`.

[11] F. Zwicky. "Die Rotverschiebung von extragalaktischen Nebeln". In: *Helvetica Physica Acta* 6 (1933), pp. 110–127.

[12] Sidney van den Bergh. "The Early history of dark matter". In: *Publ. Astron. Soc. Pac.* 111 (1999), p. 657. DOI: `10.1086/316369`. arXiv: `astro-ph/9904251 [astro-ph]`.

[13] Craig L. Sarazin. "X-ray emission from clusters of galaxies". In: *Rev. Mod. Phys.* 58 (1 Jan. 1986), pp. 1–115. DOI: `10.1103/RevModPhys.58.1`. URL: `https://link.aps.org/doi/10.1103/RevModPhys.58.1`.

[14] John E. Carlstrom, Gilbert P. Holder, and Erik D. Reese. "Cosmology with the Sunyaev-Zel'dovich Effect". In: *Annual Review of Astronomy and Astrophysics* 40.1 (2002), pp. 643–680. DOI: `10.1146/annurev.astro.40.060401.093803`. eprint: `https://doi.org/10.1146/annurev.astro.40.060401.093803`. URL: `https://doi.org/10.1146/annurev.astro.40.060401.093803`.

[15] Jose M. Diego et al. "A free-form lensing grid solution for A1689 with new multiple images". In: *Mon. Not. Roy. Astron. Soc.* 446 (2015), pp. 683–704. DOI: `10.1093/mnras/stu2064`. arXiv: `1402.4170 [astro-ph.CO]`.

[16] J. P. Kneib et al. "Hubble space telescope observations of the lensing cluster abell 2218". In: *Astrophys. J.* 471 (1996), p. 643. DOI: `10.1086/177995`. arXiv: `astro-ph/9511015 [astro-ph]`.

[17] Erin S. Sheldon et al. "Cross-correlation Weak Lensing of SDSS Galaxy Clusters III: Mass-to-light Ratios". In: *Astrophys. J.* 703 (2009), pp. 2232–2248. DOI: `10.1088/0004-637X/703/2/2232`. arXiv: `0709.1162 [astro-ph]`.

[18] Will J. Percival et al. "The shape of the SDSS DR5 galaxy power spectrum". In: *Astrophys. J.* 657 (2007), pp. 645–663. DOI: `10.1086/510615`. arXiv: `astro-ph/0608636 [astro-ph]`.

[19] Shaun Cole et al. "The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications". In: *Mon. Not. Roy. Astron. Soc.* 362 (2005), pp. 505–534. DOI: `10.1111/j.1365-2966.2005.09318.x`. arXiv: `astro-ph/0501174 [astro-ph]`.

[20] V. Springel, C. S. Frenk, and S. D. M. White. "The large-scale structure of the Universe". In: *Nature* 440 (Apr. 2006), pp. 1137–1144. DOI: `10.1038/nature04805`. eprint: `astro-ph/0604561`.

[21] K Nakamura and Particle Data Group. "Review of Particle Physics". In: *Journal of Physics G: Nuclear and Particle Physics* 37.7A (2010), p. 075021. URL: `http://stacks.iop.org/0954-3899/37/i=7A/a=075021`.

[22] P. A. R. Ade et al. "Planck 2015 results. XIII. Cosmological parameters". In: *Astron. Astrophys.* 594 (2016), A13. DOI: `10.1051/0004-6361/201525830`. arXiv: `1502.01589 [astro-ph.CO]`.

[23] Edward W. Kolb and Michael S. Turner. “The Early Universe”. In: *Front. Phys.* 69 (1990), pp. 1–547.

[24] Jonathan L. Feng. “Dark Matter Candidates from Particle Physics and Methods of Detection”. In: *Annual Review of Astronomy and Astrophysics* 48.1 (2010), pp. 495–545. DOI: `10.1146/annurev-astro-082708-101659`. eprint: `https://doi.org/10.1146/annurev-astro-082708-101659`. URL: `https://doi.org/10.1146/annurev-astro-082708-101659`.

[25] Z Ahmed et al. “Dark matter search results from the CDMS II experiment.” In: *Science* 327.5973 (Mar. 2010), pp. 1619–21. ISSN: 1095-9203. DOI: `10.1126/science.1186112`. URL: `http://www.sciencemag.org/content/327/5973/1619`.

[26] D. S. Akerib et al. “Exclusion limits on the WIMP-nucleon cross section from the first run of the Cryogenic Dark Matter Search in the Soudan Underground Laboratory”. In: *Phys. Rev. D* 72.5 (Sept. 2005), p. 052009. ISSN: 1550-7998. DOI: `10.1103/PhysRevD.72.052009`. URL: `http://link.aps.org/doi/10.1103/PhysRevD.72.052009`.

[27] *UC Berkeley CDMS Homepage*. `http://cdms.berkeley.edu/UCBlabs/Main/DeepUndergroundScience`. Accessed: 2018.

[28] *Soudan Underground Mine State Park Wikipidia Page*. `https://en.wikipedia.org/wiki/Soudan_Underground_Mine_State_Park`. Accessed: 2018.

[29] J. M. Carmona et al. “Neutron background at the Canfranc Underground Laboratory and its contribution to the IGEX-DM dark matter experiment”. In: *Astropart. Phys.* 21 (2004), pp. 523–533. DOI: `10.1016/j.astropartphys.2004.04.002`. arXiv: `hep-ex/0403009 [hep-ex]`.

[30] Jeffrey Peter Filippini. “A Search for WIMP Dark Matter Using the First Five-Tower Run of the Cryogenic Dark Matter Search”. PhD thesis. UC, Berkeley, 2008. DOI: `10.2172/1415812`. URL: `http://cdms.berkeley.edu/Dissertations/filippini.pdf`.

[31] R. Agnese et al. “Search for Low-Mass Weakly Interacting Massive Particles with SuperCDMS”. In: *Phys. Rev. Lett.* 112.24 (2014), p. 241302. DOI: `10.1103/PhysRevLett.112.241302`. arXiv: `1402.7137 [hep-ex]`.

[32] R. Agnese et al. “Low-mass dark matter search with CDMSlite”. In: *Phys. Rev.* D97.2 (2018), p. 022002. DOI: `10.1103/PhysRevD.97.022002`. arXiv: `1707.01632 [astro-ph.CO]`.

[33] William Todd Doughty. “SuperCDMS Soudan High Threshold WIMP Search”. PhD thesis. UC, Berkeley, 2016.

[34] Matt Christopher Pyle. “Optimizing the design and analysis of cryogenic semiconductor dark matter detectors for maximum sensitivity”. PhD thesis. Stanford U., 2012. DOI: `10.2172/1127926`. URL: `http://purl.stanford.edu/bj167zz6539`.

[35] Steven W. Leman. “Invited Review Article: Physics and Monte Carlo techniques as relevant to cryogenic, phonon, and ionization readout of Cryogenic Dark Matter Search radiation detectors”. In: *Review of Scientific Instruments* 83.9 (2012), p. 091101. DOI: `10.1063/1.4747490`. eprint: `https://doi.org/10.1063/1.4747490`. URL: `https://doi.org/10.1063/1.4747490`.

[36] P. N. Luke. “Voltage-assisted calorimetric ionization detector”. In: *Journal of Applied Physics* 64.12 (1988), pp. 6858–6860. DOI: `10.1063/1.341976`. eprint: `https://doi.org/10.1063/1.341976`. URL: `https://doi.org/10.1063/1.341976`.

[37] B. S. Neganov and V. N. Trofimov. “Colorimetric method measuring ionizing radiation”. In: *Otkryt. Izobret.* 146 (1985), p. 215.

[38] A. Juillard. “Status and Prospects of the EDELWEISS-III Direct WIMP Search Experiment”. In: *J. Low. Temp. Phys.* 184.3-4 (2016), pp. 897–903. DOI: `10.1007/s10909-016-1493-0`.

[39] Z. Ahmed et al. “Dark Matter Search Results from the CDMS II Experiment”. In: *Science* 327 (2010), pp. 1619–1621. DOI: `10.1126/science.1186112`. arXiv: `0912.3592 [astro-ph.CO]`.

[40] Jeffrey Yen. “Phonon Sensor Dynamics for Cryogenic Dark Matter Search Experiment”. PhD thesis. Stanford U., 2015. DOI: `10.2172/1352050`. URL: `http://cdms.berkeley.edu/Dissertations/yen.pdf`.

[41] Scott A. Hertel. “Advancing the Search for Dark Matter: from CDMS II to SuperCDMS”. PhD thesis. MIT, 2012. DOI: `10.2172/1248359`. URL: `http://lss.fnal.gov/cgi-bin/find_paper.pl?thesis-2012-19`.

[42] Sunil Ramanlal Golwala. “Exclusion limits on the WIMP nucleon elastic scattering cross-section from the Cryogenic Dark Matter Search”. PhD thesis. UC, Berkeley, 2000. DOI: `10.2172/1421437`. URL: `http://lss.fnal.gov/cgi-bin/find_paper.pl?thesis-2000-59`.

[43] Ritoban Basu Thakur. “The Cryogenic Dark Matter Search low ionization-threshold experiment”. PhD thesis. Illinois U., Urbana, 2014. DOI: `10.2172/1182551`. URL: `http://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2014-33.shtml`.

[44] Arran T. J. Phipps. “Ionization Collection in Detectors of the Cryogenic Dark Matter Search”. PhD thesis. UC, Berkeley, 2016. DOI: `10.2172/1350522`. URL: `http://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2016-36.pdf`.

[45] Kyle Michael Sundqvist. “Carrier Transport and Related Effects in Detectors of the Cryogenic Dark Matter Search”. PhD thesis. UC, Berkeley, 2012. DOI: `10.2172/1128107`. URL: `http://lss.fnal.gov/archive/thesis/2000/fermilab-thesis-2012-54.pdf`.

[46] Robert Moffatt. “Two-Dimensional Spatial Imaging of Charge Transport in Germanium Crystals at Cryogenic Temperatures”. PhD thesis. Stanford U., 2016. DOI: `10.2172/1350526`. URL: `http://cdms.berkeley.edu/Dissertations/moffatt.pdf`.

[47] B. Cabrera et al. “Oblique propagation of electrons in crystals of germanium and silicon at sub-Kelvin temperature in low electric fields”. In: *ArXiv e-prints* (Apr. 2010). arXiv: `1004.1233 [astro-ph.IM]`.

[48] T. Shutt. “Studies of the dead layer in BLIP dark matter detectors”. In: *7th International Workshop on Low Temperature Detectors (LTD 7) Munich, Germany, July 28-August 1, 1997*. 1997.

[49] David Craig Moore. “A search for low-mass dark matter with the cryogenic dark matter search and the development of highly multiplexed phonon-mediated particle detectors”. PhD thesis. Caltech, 2012. DOI: `10.2172/1247705`. URL: `http://thesis.library.caltech.edu/7043/`.

[50] Thomas Alan Shutt. “A dark matter detector based on the simultaneous measurement of phonons and ionization at 20 mK”. PhD thesis. UC, Berkeley, 1993. DOI: `10.2172/1425600`. URL: `http://lss.fnal.gov/cgi-bin/find_paper.pl?thesis-1993-28`.

[51] B. C. Welliver Jr. “Dedicated searches for low and high mass wimps with the SuperCDMS Soudan iZIP detectors”. PhD thesis. University of Florida, 2016.

[52] R. Brun and F. Rademakers. “ROOT: An object oriented data analysis framework”. In: *Nucl. Instrum. Meth.* A389 (1997), pp. 81–86. DOI: `10.1016/S0168-9002(97)00048-X`.

[53] *CDMSBATS documentation.* `http://titus.stanford.edu/dokuwiki/doku.php?id=software:cdmsbats`. Accessed: 2018.

[54] A. Anderson. *Initial R133 Charge Calibration.* `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/120410_3/`. SuperCDMS internal note. Accessed: 2018.

[55] A. Anderson. *R133 Relative Phonon Calibration.* `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/120410_3/`. SuperCDMS internal note. Accessed: 2018.

[56] A. Anderson. *R133 Absolute Phonon Calibration.* `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/120517/index.html`. SuperCDMS internal note. Accessed: 2018.

[57] Ray Bunker. *Summary of R134 Production Preparations*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/140409_rb/`. SuperCDMS internal note. Accessed: 2018.

[58] J Lindhard et al. "INTEGRAL EQUATIONS GOVERNING RADIATION EFFECTS. (NOTES ON ATOMIC COLLISIONS, III)". In: Vol: 33: No. 10 (Jan. 1963).

[59] R. Agnese et al. "Demonstration of surface electron rejection with interleaved germanium detectors for dark matter searches". In: *Appl. Phys. Lett.* 103.16 (Oct. 2013), p. 164105. ISSN: 00036951. DOI: `10.1063/1.4826093`. URL: `http://scitation.aip.org/content/aip/journal/apl/103/16/10.1063/1.4826093`.

[60] B. Welliver. *Where has all the livetime gone?* `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/20150721_livetime/20150721_livetime.html`. SuperCDMS internal note. Accessed: 2018.

[61] Jon Wilson. *Trigger Burst Cut, R133 and R134*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/2015_02_03/TriggerBurst.html`. SuperCDMS internal note. Accessed: 2018.

[62] P. Adamson et al. "The MINOS scintillator calorimeter system". English (US). In: *IEEE Transactions on Nuclear Science* 49 II.3 (June 2002), pp. 861–863. ISSN: 0018-9499. DOI: `10.1109/TNS.2002.1039579`.

[63] P. Adamson et al. "The NuMI Neutrino Beam". In: *Nucl. Instrum. Meth.* A806 (2016), pp. 279–306. DOI: `10.1016/j.nima.2015.08.063`. arXiv: `1507.06690 [physics.acc-ph]`.

[64] Brad Welliver. *Barium c335 HT KS and QS Tests Summary*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/20151010_ba_ksSummary/20151010_ba_ksSummary.html`. SuperCDMS internal note. Accessed: 2018.

[65] Todd Doughty. *Charge Chi Square Cut (v5-3)*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/130531/`. SuperCDMS internal note. Accessed: 2018.

[66] Adam Anderson. *Update to Phonon Chi-square Cut (cPChiSq*$_v$53). `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/130528/index.html`. SuperCDMS internal note. Accessed: 2018.

[67] Ray Bunke. *Initial Study of R133 Glitch Events*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/120530/`. SuperCDMS internal note. Accessed: 2018.

[68] Ray Bunke. *Defining a New R133 Glitch Cut*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/120903/`. SuperCDMS internal note. Accessed: 2018.

[69] Yu Chen. *Glitch Cut Update for R133 Low Threshold Analysis*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/130904_chen/`. SuperCDMS internal note. Accessed: 2018.

[70] Adam Anderson. *Redefining cGoodPStartTime Redux*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/130705/`. SuperCDMS internal note. Accessed: 2018.

[71] Noah Kurinsky. *High Threshold Sandbox Cut*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/160223/`. SuperCDMS internal note. Accessed: 2018.

[72] Rob Calkins. *Nuclear and Electron Recoil Band Fits for the ytNF-precoiltNF Yield Plane*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/150416_rc/temp.html`. SuperCDMS internal note. Accessed: 2018.

[73] Brad Welliver. *CAP Blinding Cut*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/2014_07_blind/140717_blind.html`. SuperCDMS internal note. Accessed: 2018.

[74] Hang Qiu. *LiveTime Estimation and Uncertainty for HT Analysis*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/150812_hq/livetime_ht_hq.html`. SuperCDMS internal note. Accessed: 2018.

[75] Todd Doughty. *Estimated Efficiency*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/160524_TD/`. SuperCDMS internal note. Accessed: 2018.

[76] J.D. Lewin and P.F. Smith. “Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil”. In: *Astropart. Phys.* 6.1 (Dec. 1996), pp. 87–112. ISSN: 09276505. DOI: `10.1016/S0927-6505(96)00047-3`. URL: `http://www.sciencedirect.com/science/article/pii/S0927650596000473`.

[77] F. Donato, N. Fornengo, and S. Scopel. “Effects of galactic dark halo rotation on WIMP direct detection”. In: *Astropart. Phys.* 9 (1998), pp. 247–260. DOI: `10.1016/S0927-6505(98)00025-5`. arXiv: `hep-ph/9803295 [hep-ph]`.

[78] Richard H. Helm. “Inelastic and Elastic Scattering of 187-Mev Electrons from Selected Even-Even Nuclei”. In: *Phys. Rev.* 104 (5 Dec. 1956), pp. 1466–1475. DOI: `10.1103/PhysRev.104.1466`. URL: `https://link.aps.org/doi/10.1103/PhysRev.104.1466`.

[79] Todd Doughty. *Lead Source Surface Background Model*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/151124/`. SuperCDMS internal note. Accessed: 2018.

[80] Andy Kubik. *Surface Alpha Rate Study*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/151216_ak/SurfaceAlphaStudy.html`. SuperCDMS internal note. Accessed: 2018.

[81] Elias Lopez Asamar. *Estimation of the cosmogenic neutron background in the SuperCDMS Soudan high-threshold WIMP search*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/161102_ELA/index.html`. SuperCDMS internal note. Accessed: 2018.

[82] Todd Doughty. *Defining a BDT Criteria*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/160627/`. SuperCDMS internal note. Accessed: 2018.

[83] M. M. McKerns et al. "Building a Framework for Predictive Science". In: *ArXiv e-prints* (Feb. 2012). arXiv: `1202.1056 [cs.MS]`.

[84] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization*. second. New York, NY, USA: Springer, 2006.

[85] Eric Jones, Travis Oliphant, Pearu Peterson, et al. *SciPy: Open source scientific tools for Python*. [Online; accessed <today>]. 2001–. URL: `http://www.scipy.org/`.

[86] J. Goodman and J. Weare. "Ensemble samplers with affine invariance". In: *Communications in Applied Mathematics and Computational Science, Vol. 5, No. 1, p. 65-80, 2010* 5 (2010), pp. 65–80. DOI: `10.2140/camcos.2010.5.65`.

[87] D. Foreman-Mackey et al. "emcee: The MCMC Hammer". In: *pasp* 125 (Mar. 2013), p. 306. DOI: `10.1086/670067`. arXiv: `1202.3665 [astro-ph.IM]`.

[88] S. Yellin. "Some ways of combining optimum interval upper limits". In: (2011). arXiv: `1105.2928 [physics.data-an]`.

[89] Brett Cornell. *HT Unblinding – Stage 1*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/20160707_opt_bdc/fid_opt_final/unblinding.html`. SuperCDMS internal note. Accessed: 2018.

[90] Elias Lopez Asamar. *Determination of the acceptance of the fiducial volume cuts in the high-threshold WIMP search at SuperCDMS Soudan*. `http://titus.stanford.edu/cdms_restricted/Soudan/R133/ebook/170720_ELA/index.html`. SuperCDMS internal note. Accessed: 2018.

[91] R. Agnese et al. "Silicon Detector Dark Matter Results from the Final Exposure of CDMS II". In: *Phys. Rev. Lett.* 111.25 (Dec. 2013), p. 251301. ISSN: 0031-9007. DOI: `10.1103/PhysRevLett.111.251301`. URL: `http://link.aps.org/doi/10.1103/PhysRevLett.111.251301`.

[92] C Savage et al. “Compatibility of DAMA/LIBRA dark matter detection with other searches”. In: *J. Cosmol. Astropart. Phys.* 04 (Apr. 2009), pp. 010–010. ISSN: 1475-7516. DOI: `10.1088/1475-7516/2009/04/010`. URL: `http://stacks.iop.org/1475-7516/2009/i=04/a=010`.

[93] G. Angloher et al. “Results on light dark matter particles with a low-threshold CRESST-II detector”. In: *The European Physical Journal C* 76.1 (2016), p. 25. ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-016-3877-3`. URL: `http://dx.doi.org/10.1140/epjc/s10052-016-3877-3`.

[94] R. Agnese et al. “New Results from the Search for Low-Mass Weakly Interacting Massive Particles with the CDMS Low Ionization Threshold Experiment”. In: *Phys. Rev. Lett.* 116 (7 Feb. 2016), p. 071301. DOI: `10.1103/PhysRevLett.116.071301`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.116.071301`.

[95] E. Armengaud et al. “Final results of the EDELWEISS-II WIMP search using a 4-kg array of cryogenic germanium detectors with interleaved electrodes”. In: *Physics Letters B* 702.5 (2011), pp. 329–335. ISSN: 0370-2693. DOI: `https://doi.org/10.1016/j.physletb.2011.07.034`. URL: `http://www.sciencedirect.com/science/article/pii/S0370269311008240`.

[96] R. Agnese et al. “Search for Low-Mass Weakly Interacting Massive Particles with SuperCDMS”. In: *Phys. Rev. Lett.* 112 (24 June 2014), p. 241302. DOI: `10.1103/PhysRevLett.112.241302`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.112.241302`.

[97] P. Agnes and others. “Results from the first use of low radioactivity argon in a dark matter search”. In: *Phys. Rev. D* 93 (8 Apr. 2016), p. 081101. DOI: `10.1103/PhysRevD.93.081101`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.93.081101`.

[98] C. Amole et al. “Dark Matter Search Results from the PICO−60 C_3F_8 Bubble Chamber”. In: *Phys. Rev. Lett.* 118 (25 June 2017), p. 251301. DOI: `10.1103/PhysRevLett.118.251301`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.118.251301`.

[99] E. Armengaud et al. “Search for low-mass WIMPs with EDELWEISS-II heat-and-ionization detectors”. In: *Phys. Rev. D* 86.5 (Sept. 2012), p. 051701. ISSN: 1550-7998. DOI: `10.1103/PhysRevD.86.051701`. URL: `http://link.aps.org/doi/10.1103/PhysRevD.86.051701`.

[100] R. Agnese et al. “Improved WIMP-search reach of the CDMS II germanium data”. In: *Phys. Rev. D* 92 (7 Oct. 2015), p. 072003. DOI: `10.1103/PhysRevD.92.072003`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.92.072003`.

[101] Xiangyi Cui et al. “Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment”. In: *Phys. Rev. Lett.* 119.18 (2017), p. 181302. DOI: `10.1103/PhysRevLett.119.181302`. arXiv: `1708.06917 [astro-ph.CO]`.

[102] D. S. Akerib et al. “Results from a Search for Dark Matter in the Complete LUX Exposure”. In: *Phys. Rev. Lett.* 118 (2 Jan. 2017), p. 021303. DOI: `10.1103/PhysRevLett.118.021303`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.118.021303`.

[103] E. Aprile et al. “First Dark Matter Search Results from the XENON1T Experiment”. In: *Phys. Rev. Lett.* 119 (18 Oct. 2017), p. 181301. DOI: `10.1103/PhysRevLett.119.181301`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.119.181301`.

[104] R. Agnese et al. “Projected sensitivity of the SuperCDMS SNOLAB experiment”. In: *Phys. Rev. D* 95 (8 Apr. 2017), p. 082002. DOI: `10.1103/PhysRevD.95.082002`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.95.082002`.

[105] G. M. Stiehl et al. “Code-division multiplexing for x-ray microcalorimeters”. In: *Applied Physics Letters* 100.7 (2012), p. 072601. DOI: `10.1063/1.3684807`. eprint: `https://doi.org/10.1063/1.3684807`. URL: `https://doi.org/10.1063/1.3684807`.

[106] J. A. B. Mates et al. “Demonstration of a multiplexer of dissipationless superconducting quantum interference devices”. In: *Applied Physics Letters* 92.2 (2008), p. 023514. DOI: `10.1063/1.2803852`. eprint: `https://doi.org/10.1063/1.2803852`. URL: `https://doi.org/10.1063/1.2803852`.

[107] Jörn Beyer, Dietmar Drung, and Kent D. Irwin. “Bias conditions of dc-SQUID for a time-domain SQUID multiplexer”. In: *Review of Scientific Instruments* 75.2 (2004), pp. 502–506. DOI: `10.1063/1.1638877`. eprint: `https://doi.org/10.1063/1.1638877`. URL: `https://doi.org/10.1063/1.1638877`.

[108] M. A. Dobbs et al. “Frequency multiplexed superconducting quantum interference device readout of large bolometer arrays for cosmic microwave background measurements”. In: *Review of Scientific Instruments* 83.7, 073113-073113-24 (July 2012), pp. 073113-073113-24. DOI: `10.1063/1.4737629`. arXiv: `1112.4215 [astro-ph.IM]`.

[109] Peter Day et al. “A broadband superconducting detector suitable for use in large arrays”. In: 425 (Oct. 2003), pp. 817–21.

[110] Jiansong Gao. “The physics of superconducting microwave resonators”. PhD thesis. Caltech, 2008. URL: `http://resolver.caltech.edu/CaltechETD:etd-06092008-235549`.

[111] D. C. Mattis and J. Bardeen. “Theory of the Anomalous Skin Effect in Normal and Superconducting Metals”. In: *Phys. Rev.* 111 (2 July 1958), pp. 412–417. DOI: `10.1103/PhysRev.111.412`. URL: `https://link.aps.org/doi/10.1103/PhysRev.111.412`.

[112] Jonas Zmuidzinas. *Amplifier-limited MKID Energy Resolution: Prospects for Ultransensitive Photon-Counting Detectors*. Tech. rep. KIDs Internal Report. California Institute of Technology, Apr. 2009.

[113] Sunil Golwala. *Expected Energy Resolution of a Phonon-Mediated Detector using TiN MKIDs*. Tech. rep. KIDs Internal Report. California Institute of Technology, Dec. 2010.

[114] Duncombe C. “Lumped Element Kinetic Inductance Detectors”. In: *J. Low Temp. Phys.* 151 (2008), p. 530.

[115] *Sonnet Software*. `http://www.sonnetsoftware.com/`. Accessed: 2018.

[116] J. Hickish and D. Werthimer. “The Collaboration for Astronomy Signal Processing and Electronics Research (CASPER): Lowering barriers to entry in radio astronomy research with open-source, community-driven digital signal processing hardware and software”. In: *2016 Progress in Electromagnetic Research Symposium (PIERS)*. Aug. 2016, pp. 1670–1670. DOI: `10.1109/PIERS.2016.7734754`.

[117] Jiansong Gao. *A fast way to determine f_r and Q from Sonnet simulation*. Tech. rep. KISS Internal Report. NIST, Oct. 2011.

[118] Omid Noroozian. “Superconducting microwave resonator arrays for submillimeter/far-infrared imaging ”. PhD thesis. Caltech, 2012. URL: `http://resolver.caltech.edu/CaltechTHESIS:06122012-091920562`.

[119] Mark W. Goodman and Edward Witten. “Detectability of certain dark-matter candidates”. In: *Phys. Rev. D* 31 (12 June 1985), pp. 3059–3063. DOI: `10.1103/PhysRevD.31.3059`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.31.3059`.

[120] G. et al. Aad. “Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In: *The European Physical Journal C* 75.7 (July 2015), p. 299. ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-015-3517-3`. URL: `https://doi.org/10.1140/epjc/s10052-015-3517-3`.

[121] David E. Kaplan, Markus A. Luty, and Kathryn M. Zurek. “Asymmetric dark matter”. In: *Phys. Rev. D* 79 (11 June 2009), p. 115016. DOI: `10.1103/PhysRevD.79.115016`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.79.115016`.

[122] Andi Tan et al. “Dark Matter Results from First 98.7 Days of Data from the PandaX-II Experiment”. In: *Phys. Rev. Lett.* 117 (12 Sept. 2016), p. 121303. DOI: `10.1103/PhysRevLett.117.121303`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.117.121303`.

[123] A. I. Gubin et al. “Dependence of magnetic penetration depth on the thickness of superconducting Nb thin films”. In: *Phys. Rev. B* 72 (6 Aug. 2005), p. 064503. DOI: `10.1103/PhysRevB.72.064503`. URL: `https://link.aps.org/doi/10.1103/PhysRevB.72.064503`.

[124] Y.-Y Chang et al. “Development of a Massive, Highly Multiplexible, Phonon-Mediated Particle Detector Using Kinetic Inductance Detectors”. In: (Apr. 2018).

[125] E S. Battistelli et al. “CALDER - Neutrinoless double-beta decay identification in TeO_2 bolometers with kinetic inductance detectors”. In: 75 (May 2015).

Appendix A

BDT UNBLINDING ISSUES

Before unblinding I was conducting some validation checks at the command prompt. This consisted of plotting scatter plots in various spaces (z vs r partition, yield vs energy etc.) for Cf data and comparing to what the optimizer had done to the signal model (which is just Cf data). I did not expect exact matches (the Signal model was just c34 data for example) but they should look very similar (with similar pass fractions). Initially, this was not the case the pass fraction for the signal model was (very roughly) 60% but the BDT cut in matlab was removing 90% of the data. Clearly something was wrong. It turns out, that two of the input RRQ's (prpartOF and pzpartOF) had their order reversed in apply_BDT when compared to when the BDT's were initially trained. After fixing this problem, things looked much better. At this point I wrote a script to produce all of these scatter plots for inclusion in a note. In the same script I also produced the same plots for the new stage 1 unblinded wimp search data (that is data the falls outside the BDT cut but inside the blinded region). After these plots were made I was having a closer look that the Cf plots an noticed that there were some slightly more subtle problems that I had not appreciated while I had been making the plots interactively. So, I reloaded the saved scores from the signal model, found the matching Cf events (by Series and Event number) and re-scored them from matlab using the apply_BDT script. The scores did not match. At this point I realized that there was a problem with "round-tripping" the BDT: it preformed differently after being saved and re-loaded. I stopped (deleted the stage 1 wimp search scatter plots I had made) and set about fixing the BDT. As it turns out there were a few problems.

First I noticed that the scores coming out of apply_BDT (after saving) were off by something like a factor of 2. The problem here was that in apply_BDT every event was scored by every tree, while originally about half of the events were removed to train the trees and had their score set to 0. Averaged out over all the trees this suppressed the score by about a factor of 2. I fixed this by tracking the number of times an event was actually scored and normalizing by that instead of the number of trees in the bootstrap.

After fixing the problem described above out optimizer started going crazy. This was due to the fact that some events would never be scored in a single tree (50% chance of being scored in each tree and ten trees. If you have enough events there will be a few that are left out.). These events that were never included were normalized (or divided by) 0. I had actually already coded against this possibility by checking for inf values (which is was floating point $x/0.0$ produces), but did not realize that the ieee spec also allows for 0.0/0.0 to be a special NaN value that is distinct from inf. This was not a fun bug to find. Eventually the fix here was to simply increase the number of trees in the bootstrap until all events were included at least once all the time. I settled on 25.

At this point the scores were looking fine from a sytematic standpoint (the distribution was right), but they were still not exactly the same on an event-by-event basis. The problem here was tracking all the randomness in the bootstrap. To fix this all of the decisions as to which event was used as training and testing for every tree was saved along with he Event and Series number of that event. Then, in the apply_BDT script each event can be deterministically scored in exactly the same was it was in the signal and background model. In the case of the new event that was not used in the model (such as a WIMP search event), each of the 25 trees are used to produce a score.

At this point I could sometimes deterministically reproduce the BDT score using the apply_BDT script. If I was checking things interactively, I could often get it to work, but as soon as I tried to run an automated check on everything the scores suddenly would not quite match up (but they would be very very close). To explain what was going wrong here we need to talk a little about exactly how the normalization to [0,1] works. To get the most out of the dynamic range the scores in each tree are linearly mapped to [0, 1] in such a way that that minimum score is exactly 0 and the maximum score is exactly 1. To do this ween need to know that maximum and minimum score produced by our bg and signal models for each tree. These values were cached along with the trees themselves, but it was done in a way that allowed the same tree from different detectors to over-write each other. This meant that when I took a particular model (from say 1101) scored and saved it, and then immediately turned around and used apply_BDT to make the corresponding scores in MATLAB everything seemed to work exactly. But, if I made all ten models and then tried to reproduce in MATLAB things would be off by a small amount. The fix was easy (give each detector its own cache) but this was also not a fun bug to find.