# Cryogenic Detector Development for SuperCDMS

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Nicholas Glenn Mast

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Dr. Vuk Mandic, Advisor

November, 2020

# Acknowledgements

Completing this work has certainly been a journey. Along the way there were triumphs and setbacks, tremendous gains and devastating losses. Many things are quite different now than they were over six years ago when I first found myself a fresh grad student in old Tate hall. The research presented here has all taken place against a backdrop of changing personal, political, and societal conditions. I have only managed to navigate this all due to aid of many people, and I owe each of them my thanks for the influence they have had in my life. Because you are so numerous and I so forgetful, I am certain to neglect some of you, but here goes...

First I thank Prisca Cushman and Vuk Mandic who have together steered the course of the SuperCDMS group at UMN. Their scientific, professional, and personal guidance have helped shape me into a better physicist. In particular I thank my advisor Vuk for his calm and insightful manner and superb mentorship. Your steady hand has been a great help and comfort to me at many times, certainly more than I have let show. Thanks also to the post-docs I had the pleasure of working with at UMN, including Hassan Chagani, Matt Fritts, and Anthony Villano. Hassan and Matt, you helped me learn how to operate a dilution fridge with enough confidence to only occasionally wake bolt upright in the middle of the night in fear that I had left some important valve open and lost all the mix! From what I understand now, that is actually a sign of a reasonable level of competence, so thank you! Matt and Anthony, you helped me to learn many of the nuts and bolts of the particular blend of astrophysical, solid state, and experimental particle physics that is DM research. You have also shown through your diligent work and attention to detail how essential, but also how hard it is to produce good, solid, and valuable scientific results. While I am a better researcher for having worked with each of you, more importantly, I value you both as friends.

I have spent many hours of these years in the windowless basement labs of physics buildings, fretting over gauges and measurements as the ever-present whine of machinery slowly eroded my hearing and sanity. While that sounds like a particular brand of hell, the companionship of other inmates actually made these times (on average) fun! Allison Kennedy, you showed me how to balance the seriousness and frustration of research with whimsy and dark humor. David Strandberg taught me many practical lab things, but the most essential lessons were the importance of riddles and that D&D is cool! I also was fortunate to work in lab with many undergrad assistants including Gina Ristani, Max Buss, Jake Nelson, Gabe Spahn, and Dan Codoluto. Dan, I will always cherish our early morning wanderings through time and space on "Dan Day." In the first years of my time at UMN, I was also fortunate to work with Mark Pepin, D'Ann Barker, and Hannah Rogers. Though we spent less time together, I always enjoyed chatting with each of you in the grad student office. This is not only because you had windows and sunlight, but because you are each interesting and intelligent people that I am lucky to have met. In the last few years, I have had to succeed to the senior grad student throne as our group grew and changed, taking on our newer members; Jack Nelson, Derek Sincavage, Zachary Williams, and Michael Gardner. While my trips to FNAL and later the spread of COVID have prevented much direct interaction in the last year or two, I have enjoyed getting to know each of you and wish you the best in the coming years. I hope this group will find the camaraderie that I was fortunate to experience.

Speaking of FNAL, I owe immense gratitude to the SuperCDMS crew there; Noah Kurinsky, Dan Bauer, Lauren Hsu, Matt Hollister, and Pat Lukens. I was initially worried about being able to make an impact among such a group of skilled scientists, but all of you were immediately so kind and helpful, that I quickly felt welcome during my time there. Noah, in particular I thank you for your help and friendship during these years. Your skill and breadth of knowledge constantly impress me, but I am thankful also for the times you shared my woes and talked about life over a beer (or three). Those trips to Chicago would not have been nearly so fun without having the great fortune to work with the Northwestern crew as well. Thank you to Tali Figueroa for welcoming me during my too-brief visits to your lab and for your excellent example of how to lead a team in impactful physics research. Thanks also to Ziqing Hong and Tom Ren, you guys gave me the guidance and advice I needed to be of some help at NEXUS. I will

# Dedication

For my family.

## Abstract

The dark matter (DM) problem is currently one of the most pressing issues facing physics. This thesis presents some of my work and contributions to the detectors employed by the Super Cryogenic Dark Matter Search (SuperCDMS), one of the many experiments currently investigating DM. First, we present an overview of astrophysical and cosmological observations to give insight into our current understanding of DM and its role in the universe. Some potential particle DM candidates are also discussed. Then we describe the SuperCDMS experiment and relevant detector technologies. We go into particular detail in discussing the phonon sensors, deriving the relationships that define their fundamental working principles and describing their practical deployment. We also discuss the main noise contributions to the phonon sensors which can limit experimental sensitivity.

We then present some work related to reconstruction of event position in the detectors, including new analysis techniques and a novel cryogenic calibration source mover we have developed and tested. We also present tests of improving experimental scalability by increasing individual detector masses using the largest-diameter cryogenic Si detectors yet operated. We present first experimental results from the novel use of $(n, \gamma)$ processes to calibrate the low energy nuclear recoil energy scale. This work is essential to understand how DM-nuclear interactions will manifest in semiconductor detectors. We also take a short look at the newest, most sensitive detectors currently operated by SuperCDMS and discuss how they are allowing new insights into critical physics processes in all SuperCDMS detectors. Finally, we demonstrate how some of the various improvements discussed in this work can improve the experimental reach of the next generation of SuperCDMS at SNOLAB.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Dark Matter

This chapter provides an overview of the dark matter (DM) problem. The first section discusses the current evidence for a large amount of DM in the universe. The second section briefly outlines some popular candidates for particle DM. The last section then describes the many methods and experiments by which people are currently attempting to detect particle DM.

## 1.1 Evidence for Dark Matter

Diverse areas of physics supply compelling evidence that baryonic matter is only a small part of the total matter content of our universe. This section presents some of the main arguments from astrophysics and cosmology for the existence of non-baryonic DM. These arguments generally involve inferring the presence of DM through its gravitational interactions with normal baryonic matter. The treatment in this section is relatively brief. For an interesting overview of the historical evolution of our understanding of DM, the interested reader is directed to Ref. [1].

### 1.1.1 Galaxy and Cluster Dynamical Measurements

Lord Kelvin first applied the kinetic theory of gases to a galaxy of stars to make a dynamical estimate of the total mass [2]. He used estimates of the velocities of local stars to place a rough constraint on the amount of unseen matter in the galaxy. Early famous evidence for DM came from observations of galactic motion within the Coma

cluster [3]. In his 1933 paper, Zwicky used redshift measurements to calculate the dispersion of apparent velocities of 8 galaxies in the cluster. Using the Virial Theorem, a total mass two orders of magnitude larger than that typical of the luminous matter was found.

At the galactic scale, evidence for DM was first shown by measuring H$\alpha$ and NII lines in a set of 21 spiral galaxies to calculate their rotation curves, the stellar rotational velocities as a function of radius from the center of each galaxy [4]. It was found that, even across a wide range of galaxy size and luminosity, these rotation curves all exhibited approximately flat behaviors outside of their central nuclei and continuing out to $R_{opt}$, which encompasses 83% of total integrated light from the galaxy. A set of such measured velocity curves can be seen in Fig. 1.1. However, a simple application of Newton's Laws shows that, at radius, $R$, of a galaxy having enclosed mass, $M(R)$, the rotational velocity, $v(R)$, satisfies

$$v^2(R) = \frac{GM(R)}{R} \tag{1.1}$$

Thus, the observation that $v(R) = const$ implies that the mass distribution is going as $M(R) \propto R$. The total mass actually scales with radius even though the luminous mass distribution falls off. This implies the presence of significant DM halos on the galactic scale. More recently, similar results were found by using the 21 cm Hydrogen line to measure rotation curves of cold dust in spiral galaxies [5]. While arguments can be made that galactic and cluster dynamics are still not entirely understood, the existing evidence is sufficiently strong that any DM candidate must explain these dramatic effects. While the Milky Way's rotation curve is more difficult to determine than other galaxies, the present measurements and modeling indicate that even our host galaxy is dominated by DM [6].

### 1.1.2   Gravitational Lensing

Evidence for DM can also be found at the scale of galaxy clusters with the observation of gravitational lensing. As predicted by General Relativity, light coming from distant sources can be deflected by the mass of a foreground object, causing distortions in the image of the distant source. In particular, a lens of point mass $M$ will deflect a light
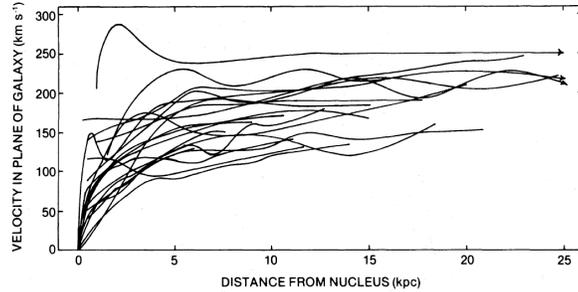
Figure 1.1: Measured rotation curves from a sample of 21 spiral galaxies. Despite differences in scale, all show flat rotation curves. Taken from [4].

ray by an angle $\alpha = 4GM/bc^2$, where $b$ is the ray's impact parameter [7]. Extreme examples of this effect are the arcs and rings predicted by first Chwolson in 1924 [8] and now commonly observed by the Sloan Digital Sky Survey and the Hubble Space Telescope. With detailed modelling, the amount and shape of the distortion can be used to estimate the foreground object's total mass and mass distribution. This approach does not rely on dynamical galactic models and is, in principle, a direct measure of the total mass of a galaxy or cluster. While this provides clear evidence for DM's existence, such strong lensing effects are relatively rare. A more fruitful approach has been to look at the much subtler, though more common, weak lensing regime.

Farther from the high-mass centers of lensing objects, the shapes of background galaxies appear only slightly deformed. This effect is known as weak lensing. Because the effect is so small, the distortion of images of multiple lensed objects must be analyzed statistically to constrain the lens mass. However, this method enables the measurement of detailed mass distributions out to larger radii than with strong lensing. In particular, halo radii extending out to 260 $h^{-1}$kpc have been measured [9] ($h$ is a factor commonly used to parameterize our ignorance of the value of the present Hubble constant with $H = 100\,h$ km s$^{-1}$ Mpc$^{-1}$). This is much farther than is found from either strong lensing or orbital tracers. In another study, a set of weak lensing galaxies was surveyed and the average mass distribution was fit to the commonly-used Navarro-Frenk-White density profile. By calculating likelihoods of various ellipticities of the DM halo, limits that effectively excluded simple spherical DM halos were found [10]. This further supports the existence of DM as opposed to alternative theories of gravity which cannot easily

reproduce such anisotropies at large radii.

In most galaxies and clusters, dark and luminous matter are found in equilibrium states and are thus spatially coincident, following their common gravitational potential. This makes measurements like rotation curves dependent upon assumptions about galactic content and evolution. However, in an event such as merger of two clusters, the intracluster plasma (which makes up the majority of the baryonic matter) will experience a fluid-like collision while the non-collisional galaxies and DM will not, allowing the mass components to be disentangled. Precisely this scenario has been examined in the Bullet Cluster [11]. When gravitational microlensing was used to map the mass distribution of the two colliding clusters, the mass was found to align with the spatial distribution of member galaxies rather than the distribution of the intergalactic plasma as measured by x-ray emission. Fig. 1.2 shows the calculated mass and plasma distributions. This rare event provides dramatic evidence that significant non-collisional DM halos exist, even at the cluster scale. This same structure also provides an opportunity to constrain the cross section for DM self interactions to $\sigma/m < 1$ cm$^2$/g [12].
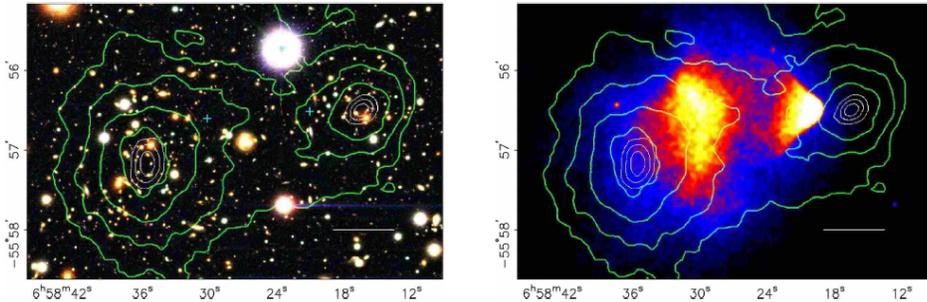


Figure 1.2: Distributions of matter in the Bullet cluster. Contours show mass distribution measured from weak lensing. The right panel includes x-ray imaging of intracluster plasma, demonstrating its separation from collisionless DM. Reproduced with permission from [11].

### 1.1.3 Cosmological Evidence

Some of the strongest evidence for the existence of DM comes from cosmological observations including large scale structure and the primordial elemental abundances resulting from the Big Bang. These arguments deal with the total mass and energy content of

the universe and show that, for the observed structure to be consistent with our current cosmological theories, DM is required to be cold, non-baryonic, and several times more abundant than ordinary matter.

### Friedman Equation and Matter Density

To connect the shape of our universe to the features of DM, a bit of background cosmology is required. The universe is observed to be homogeneous and isotropic on the largest scales. The Friedmann-Robertson-Walker (FRW) metric describes such a spacetime with a maximally symmetric subspace evolving in time, and is given by

$$ds^2 = dt^2 - a(t)^2 \left( r^2 d\Omega^2 + \frac{dr^2}{1 - kr^2} \right) \tag{1.2}$$

where $a(t)$ is the scale factor and $k \in \{1, -1, 0\}$ describes the curvature as closed, open, or flat [7]. Inserting the FRW metric into Einstein's Equations yields the Friedmann Equation,

$$H^2 = \frac{8\pi G \rho}{3} - \frac{k}{a^2} \tag{1.3}$$

where $\rho$ is the total energy density and the Hubble parameter, $H = \dot{a}/a$, has been introduced. This relates the contents, $\rho$, and curvature, $k$, of the universe to the scale factor expansion rate at any era. It is convenient to define the density parameter $\Omega_i = \frac{8\pi G}{3H^2} \rho_i$, where $\rho_i$ denotes contributions to the total energy density $\rho = \sum_i \rho_i$ from matter, radiation, and vacuum energy. This notation allows the Friedmann equation to be written as

$$\Omega = \Omega_M + \Omega_R + \Omega_\Lambda = 1 + \frac{k}{a^2 H^2} \tag{1.4}$$

where $\Omega_M$, $\Omega_R$, $\Omega_\Lambda$ are the matter, radiation, and vacuum energy contributions respectively. Note that $\Omega_M$ contains contributions from both baryonic ($\Omega_B$) and DM ($\Omega_\chi$). Thus the total energy density of the universe is directly related to its spatial curvature.

Constraints on $\Omega_M$ and $k$ can be found via a number of complimentary methods. The luminosity distance of SN1A are affected by the expansion history (i.e. $a(t)$) and can be fit with cosmological models that depend on parameters such as $\Omega_M$, $\Omega_\Lambda$, and $k$. Recent such studies have provided measurement of the present-day values of

$\Omega_{M,0} = 0.307 \pm 0.012$ and values of $k_0$ consistent with 0 [13]. The Cosmic Microwave Background power spectrum can also be used to constrain the matter density because a higher matter density tends to diminish smaller scale density fluctuations. The Planck collaboration has used this to determine a value of $\Omega_{M,0} = 0.3111 \pm 0.0056$ [14]. Baryon Acoustic Oscillations (BAO) can also constrain $\Omega_M$. Here, the largest scale of matter oscillations depends on the balance of matter and radiation density in the universe prior to recombination. The over-densities of these oscillations became the seeds for large-scale structure formation. By measuring the correlations scales of structure in the local universe BAO measurements alone can constrain $\Omega_{M,0} = 0.303 \pm 0.040$ [15].

As discussed in the remainder of this section, for the measured curvature, the baryon energy density is not large enough to account for the observed total matter density of $\Omega_{M,0} \approx 0.3$. This provides strong evidence of the influence of DM even at cosmological scales.

**Big Bang Nucleosynthesis**

Big Bang Nucleosythesis (BBN) describes the synthesis of the light elements in the beginning of the universe. Its predictions can be used to constrain the baryonic matter density, $\Omega_B$. In the first seconds after the Big Bang, the universe consisted of a hot plasma of elementary particles. The energy density was dominated by radiation, and weak interaction processes kept protons and neutrons in equilibrium [16]. As the scale factor grew, the radiation density fell, lowering the average temperature of the universe. At temperatures well below the nucleon mass, the equilibrium ratio of neutrons to protons became $n_n/n_p \approx \exp(-\Delta m/T)$ [17]. Eventually, the neutron-proton inter-conversion rate fell below the Hubble expansion rate and the weak interactions quickly froze out. This occurred at $T \approx 0.8$ MeV, giving $n_n/n_p \sim 1/6$. Due to free neutron Beta decay, this fell to a value of $\sim 1/7$ before production of the light elements began [18].

With the initial $n_n/n_p$ ratio set, the first bound element to form was deuterium via $p + n \leftrightarrow D + \gamma$. Even though the binding energy of deuterium is only 2.2 MeV, its formation was delayed until the photon temperature was much lower. Given the baryon to photon ratio, $\eta \equiv n_B/n_\gamma \sim 10^{-10}$, any deuterium formed was quickly photo-dissociated by high energy photons. So it was not until the temperature reached $\sim$

0.1 MeV that deuterium production really began to take hold.

Once this process began, most of the neutrons were quickly used to form deuterium and were thence combined into $^4$He, the most stable light element. Most of the deuterium formed went into $^4$He during this time. The precise amount was determined by the expansion history, which was governed by $\eta$. Then, given a good measurement of $n_\gamma$ (which we will see in Sec. 1.1.3 is provided by precision observation of the Cosmic Microwave Background) the amounts of primordial deuterium, $^4$He, and the other light elements, are good probes of $\Omega_B$.

$^4$He abundance can be measured via its emission lines in metal-poor extra-galactic HII regions in the local universe. Because it is also created by stars, $^4$He is correlated with metallicity of the environment. Extrapolation back to zero metallicity then gives an estimate of the primordial helium mass fraction, $Y_P$ [19]. Several recent analyses [20, 21, 22] of such systems report values around $Y_P \approx 0.245$, which corresponds to $\Omega_B h^2 \approx 0.02$.

While He emission has provided the most stringent estimates of $Y_P$, other methods such as measuring absorption of quasar light in primordial intergalactic gas clouds [23] or fitting for the effect of $Y_P$ on the damping tail of CMB anisotropies [24] yield independent confirmation of these values.

Deuterium abundance is also measured in high-redshift, low-metallicity quasar, which have nearly primordial abundances. It is identified by its isotope-shifted Lyman transitions absorption. Only a few such systems have been measured, but values of the D/H ratio yield $\eta = 5.6 \pm 0.5 \times 10^{-10}$ corresponding to $\Omega_B h^2 = 0.0205 \pm 0.0018$ [25]. Thus BBN provides multiple indications that less than 20% of the matter density of the universe can be accounted for by baryons, with the remainder assumed to be DM.

**Cosmic Microwave Background**

Observation of the Cosmic Microwave Background (CMB) also provides compelling evidence for the presence of substantial amounts of DM. As the scale factor continued to grow after BBN, the temperature of the universe fell until it reached $T \sim 0.3$ eV, at which point the first neutral atoms began to form in a period (confusingly) known as recombination. Once conditions were suitable for atoms to remain stable, the photon mean free path became long enough that they were essentially free streaming. The

energies of these free streaming photons still carry the imprint of the matter they last interacted with, providing valuable insight into conditions of the early universe.

The first remarkable feature of the CMB is that its spectrum is that of an incredibly uniform and nearly perfect blackbody of $T = 2.72548 \pm 0.00057$ K [26]. Atop that uniform signal, however, there are small temperature variations which give a map of particle density fluctuations in the surface of last scattering. The power spectrum of these temperature anisotropies is defined as

$$C_l = \frac{1}{2l+1} \sum_{m=-l}^{l} \left\langle |a_{lm}|^2 \right\rangle \tag{1.5}$$

where $a_{lm}$ are the coefficients of the expansion of $\delta T/T$ in spherical harmonics. The measurements of this angular temperature spectrum, shown in Fig. 1.3, are the most iconic results from CMB experiments such as Planck. Here the quantity $D_l = l(l+1)C_l/2\pi$ is directly related to the correlation of temperature between two points at a given angular separation.
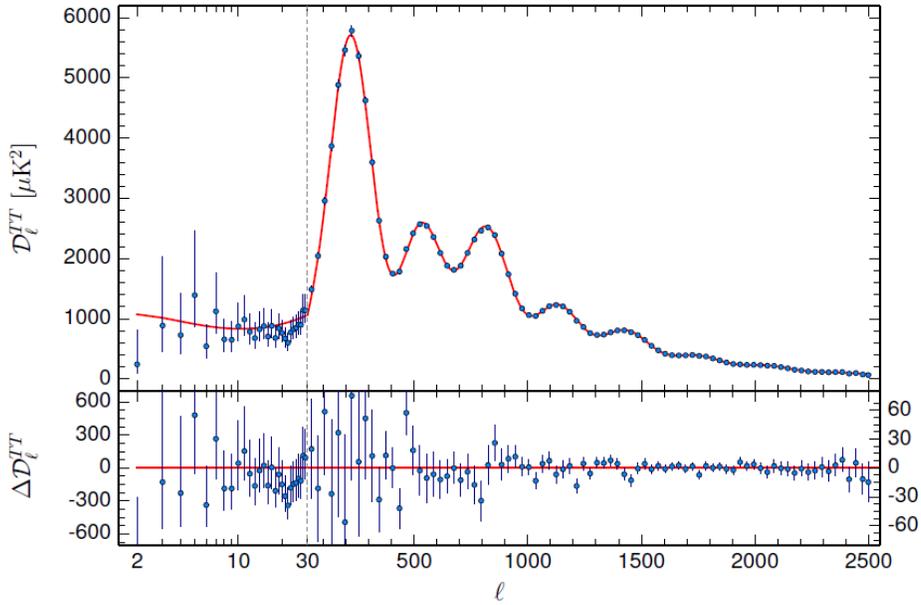


Figure 1.3: Cosmic Microwave background angular temperature spectrum as measured by the Planck satellite. Reproduced from [27] with permission.

The physics at the time of recombination involves only a few main constituents;

baryons, photons, and DM. The size and shape of the observed angular temperature spectrum is determined by the relative amounts of these ingredients and gives an independent measure of many cosmological parameters including $\Omega_B$ and $\Omega_M$. Before recombination, the content of the universe was a plasma, with electrons and photons tightly coupled through Thomson scattering. Any initial overdensities created potential wells into which baryons began to fall until photon pressure pushed them back out, setting up oscillations which traveled through the plasma at the speed of sound. The scale and magnitude of these oscillations are reflected in peaks of the CMB angular temperature spectrum [28].

By comparing the inferred physical scale of the lowest mode to the distance to the CMB, the BOOMERanG [29] and MAXIMA [30] experiments were able to conclude that the curvature constant, $k \approx 0$, indicating a flat, Euclidean universe. Clearly, the detailed characteristics of these oscillations depend on the content of the universe during the period of recombination. For example, higher baryon content would allow the fluid compressions to become denser, increasing the difference between adjacent peaks, which correspond to compressions and rarefactions. This is used to give a measure of $\Omega_B$. The total matter density, $\Omega_M$, can be measured through the Integrated Sachs Wolf Effect, the net redshifting of photons climbing out of potential wells as the universe expands. A variety of such effects contribute to the spectral shape. With a set of independently measured priors and a given cosmological model, the spectrum can generally be fit, giving values for the component densities. Some of the best such results are those obtained by the Planck Collaboration [14], which find a baryon density of $\Omega_B h^2 = 0.02233 \pm 0.00015$.

The model used assumes the presence of "cold" DM to allow the growth of structure. Numerical simulations of structure formation have shown that relativistic ("hot") DM, such as neutrinos, prevent structure formation up to the galactic scale ($\sim 10^{12} M_\odot$) [31]. This would imply a "top-down" formation of structure, which is refuted by observations of galaxies out to $z > 8$ [32], indicating that such particles are too hot to act as the DM. Therefore, the primary constituents must be cold (non-relativistic) relics and the Planck Collaboration finds a cold DM density of $\Omega_C h^2 = 0.1198 \pm 0.0012$.

The sections above have outlined several strong lines of evidence indicating that DM is now present in the universe and has been so since the very beginning. What's more,

its effects are observed from the largest cosmological scales to within our own galaxy.

## 1.2 Dark Matter Candidates

DM's gravitational and cosmological signatures give some insight into its possible identities. In this section, a few of the most popular solutions including MACHOs, Axions, and WIMPs are briefly discussed. It should be mentioned that there is no guarantee the DM is composed of only a single species, but may be a mixture of several. In addition to the candidates discussed in this section, there is a menagerie of other proposals, some rather exotic. Briefly mentioned are: sterile neutrinos, like SM neutrinos without weak interactions; gravitinos, the graviton superpartner; scalar DM and Wimpzillas, which may explain some galactic gamma ray excess; and Kaluza-Klein particles, which appear in theories of universal extra dimensions [33].

### 1.2.1 Baryonic Matter

A once popular theory for baryonic DM was Massive Astrophysical Compact Halo Objects (MACHOs). These are low mass stars or Jupiter-like objects in the galaxy's halo. Several searches were conducted to find such objects via their gravitational microlensing of nearby stars (e.g. in the Magelenic clouds) due to a Milky Way halo object passing through our line of sight. Though some objects were found, experiments such as EROS showed that MACHOs contribute only $\sim 8\%$ of the total halo mass [34]. While MACHOs may make up a piece of the galactic DM, they are not the primary constituent.

### 1.2.2 Axions

Axions were first suggested as a solution to the strong CP problem in particle physics. While the details of this argument are too complex to treat here, the upshot is that the existence of an axion with weak coupling to normal matter and radiation would help explain why CP violation is not seen in QCD interactions [35]. These same particles were later realized to be an excellent DM candidate. Unfortunately, due to its weak coupling and tiny mass, the axion is very difficult to detect directly.

Stellar physics is able to set stringent limits on the axion mass. Axions interact with photons via $\mathscr{L} = g_{a\gamma}\phi_a \mathbf{E} \cdot \mathbf{B}$ and can be created via Primakoff conversion, $a \leftrightarrow \gamma + \gamma$ if

external electric and magnetic fields are present [35]. This allows them to be produced in the plasma of stellar interiors. Axions, being weakly coupled to the stellar material, would have a much greater free-streaming length and provide another mechanism for cooling stellar interiors. Detailed statistical analyses of stellar populations have characterized the timescales of nuclear burning and led to an upper limit on the axion mass of $m_a \lesssim 0.4$ eV [36]. A still stronger limit of $m_a \lesssim 16$ meV can be set by considering the effect axions would have on the rate of energy loss in SN 1987A and the length of neutrino signal seen from that event [37]. A rough lower limit on the axion mass of $m_a \gtrsim 0.1$ meV can also be set if one allows the axions to make up all of the cold, non-baryonic matter [38].

### 1.2.3   WIMPs

Weakly Interacting Massive Particles (WIMPs) have been a well motivated DM candidate and were thus the primary target of most DM searches. These particles would have some mass, $m_\chi$ and self-annihilation cross-section $\sigma_\chi$. Soon after the Big Bang, when $T > m_\chi$, these particles remained in thermal equilibrium with photons and lighter species via pair creation/annihilation. As shown by simulations of large scale structure, the DM components must be cold [39], and thus must decouple at a lower temperature. Once $T < m_\chi$, the number density took the form $\sim \exp(-m_\chi/T)$. This caused the pair creation/annihilation to come to an abrupt halt, and the $\chi's$ to decouple from their surroundings. Thereafter, the comoving number density remained unchanged until today. If the $\chi's$ are to account for the cold DM density currently observed, they should have weak-scale annihilation cross sections as shown in detail in Appendix A. This leads to their being dubbed Weakly Interacting Massive Particles (WIMPs).

For the past several decades, the main thrust of DM search programs has been to look for the effects of interactions of WIMPs with baryonic matter. As discussed in Sec. 1.3, these efforts have excluded large swaths of natural WIMP parameter space, with no definitive discoveries yet.

### 1.2.4  Dark Sectors

Another possibility is that DM may not be charged under any standard model (SM) forces. Such DM might still have a rich life, and could be composed of a variety of undiscovered particles and forces in a so-called "Hidden" or "Dark" Sector [40]. The DM need not interact with SM particles (except, of course, gravitationally), but interactions connecting the Dark Sector to SM forces (sometimes given exciting names like "portals") may provide mechanisms to explain DM abundance and are naturally of general interest.

In general, Dark Sector (DS) models are extensions of the SM with weak (i.e. feeble, not Weak force) interaction with SM [41]. In fact, to still achieve DM relic abundance via freeze out, the DS must include non-SM, low mass states to mediate the DM-SM interactions [42]. DS-SM interaction could be direct via some mediator charged under both SM and DS forces, or indirect through kinetic mixing. Kinetic mixing occurs when virtual fermions charged under both the dark $U(1)_D$ and SM $U(1)_\gamma$ can form loops in the Feynman diagram, giving an effective interaction [43]. One popular such model is the "Dark Photon", a new vector particle A' which couples to SM photons in this way, with the strength of the mixing characterized by a mixing parameter, $\varepsilon$. Below the electroweak scale and for $m_{DM} \ll m_W$ the A' mixes directly with the SM photon and the interaction Lagrangian becomes

$$\mathcal{L} \supset -\frac{\varepsilon}{2} F_{\mu\nu} X^{\mu\nu} \tag{1.6}$$

where $F_{\mu\nu}$ and $X^{\mu\nu}$ are the EM and Dark photon field strength tensors respectively [44].

A field redefinition of $A^\mu \to A^\mu - \varepsilon X^\mu$ leads to a term coupling the DM to the EM current, $\varepsilon J_\mu X^\mu$. Recalling that $J_\mu = -e n_\mu$ this is like the EM current acquiring an extra dark charge $q_D = \varepsilon q$. This is useful because the interaction of such a particle with detector material will then be proportional to that of the photo-electric effect. In particular, the matrix element for absorption of the A' by electrons is related to that of the photon as $|\mathcal{M}|^2 = \varepsilon^2 |\mathcal{M}_\gamma|^2$ [45]. For masses larger than the band gap of a semiconductor detector, the dark photon absorption cross section, $\sigma_{abs,A'}$ is then related

to that of the photon, $\sigma_{abs,\gamma}$ as

$$\sigma_{abs,A'}(E_{A'} = m_{A'})v_{A'} \simeq \varepsilon^2 \sigma_{abs,\gamma}(\omega = m_{A'})c \tag{1.7}$$

where $v_{A'}$ is the Dark photon velocity and galactic DM is assumed to be non-relativistic with $E_{A'} \approx m_{A'}$ [46]. When considering this process in a material, there are additional in-medium corrections which are accounted for with the substitution of an effective mixing angle, $\varepsilon \to \varepsilon_{eff}$ [45]. This depends on the target's complex conductivity, $\tilde{\sigma} = \sigma_1 + i\sigma_2$. This quantity traditionally (and here confusingly) uses the same symbol as cross sections, but is something very different. It describes the material's response to electromagnetic oscillations and has units of $1/s$.

The dark photon absorption rate (per unit target mass per unit time) in a detector is given by

$$R = \frac{1}{\rho_{det}} \frac{\rho_{DM}}{m_{A'}} \langle n_e \sigma_{abs,A'} v_{A'} \rangle \tag{1.8}$$

where $\rho_{det}$ is the detector mass density, $\frac{\rho_{DM}}{m_{A'}}$ is the axion number density, and $n_e$ is the detector electron number density. Now, it can also be shown that the absorption rate is related to the real part of the complex conductivity, $\sigma_1 = \langle n_e \sigma_{abs,A'} v_{A'} \rangle$ [45]. Using this and Eq. 1.7, the absorption rate becomes

$$R = \frac{1}{\rho_{det}} \frac{\rho_{DM}}{m_{A'}} \varepsilon_{eff}^2(\omega = m_{A'}, \tilde{\sigma})\sigma_1(\omega = m_{A'}) \tag{1.9}$$

So the expected signal for absorption of a dark photon would be deposition of energy in a detector by absorption of a single energy of $m_{A'}c^2$ on electrons and with a rate given by Eq. 1.9. Because this simply scales with number of detector electrons, this allows even small detectors to search for such particles in any detector material. Such work has recently been performed by SuperCDMS [47, 48] and others.

## 1.2.5 Electron-Recoil DM

For low DM masses, interactions with detector electrons can become energetically favorable and progress can be made searching for such electron-recoil DM (ERDM) signals. Here, the interaction is a DM particle, $\chi$, scattering with a SM electron and the interaction is mediated by a dark photon, $A'$, which itself may be massive.

The ERDM relic abundance can be generated by low mass DM via a freeze-out (similar to WIMPs) or freeze-in mechanism [49]. With freeze-out scenarios, the DM starts with a large abundance in thermal equilibrium with the bath of SM particles. As the universe expands and $T < m_\chi$, DM formation from SM slows. Eventually interaction between the DM and SM all but ceases and the two populations evolve independently, with the DM abundance frozen-out at the value it had when last interacting. For freeze-in, there is assumed to initially be no substantial DM abundance, but feeble interactions between SM and DM allows some DM production. This slowly builds up the DM population as it heads towards equilibrium. The population grows as the universe cools until $T < m_\chi$, at which point it too becomes isolated. So in contrast with freeze-out where DM starts in equilibrium and moves away as the universe cools, freeze-in starts with almost no DM but the population moves towards equilibrium. The freeze-in scenario also allows for further evolution of the DM population, with higher-mass DM perhaps decaying into lighter, stable DM particles.

Light mediators predict a very low production/annihilation cross section, so they never reach thermal equilibrium and the population can only freeze in. However, they also lead to a kinematic enhancement of the the DM/SM scattering cross section that can bring the predicted DM-electron scattering rates within reach of current detector technologies [40]. Due to their low band gaps, semiconductor detectors are best suited to access the low-energy region where such signals may lie. However, the complex nature of the electron band structure in crystals make the calculation of ERDM scattering rates quite involved [50].

## 1.3  Dark Matter Searches

There are three paths that can lead to discovery of particle DM; indirect detection, production, and direct detection. Fig. 1.4 demonstrates how these complementary methods all work to constrain any unknown physical interactions between DM and SM. The indirect detection method looks for evidence of DM decay products both in and outside of our galaxy. By contrast, particle collider experiments offer opportunities to produce DM and search for the resulting missing momentum in otherwise well-understood particle collision events. Finally, the direct detection approach relies on the premise that

DM is present in our galaxy and that the Earth is continuously traveling through a cloud of it. This allows the possibility of occasional DM-SM interactions which could be measured with a suitable laboratory detector. These three avenues to discovery are each discussed in more detail in the sections below.
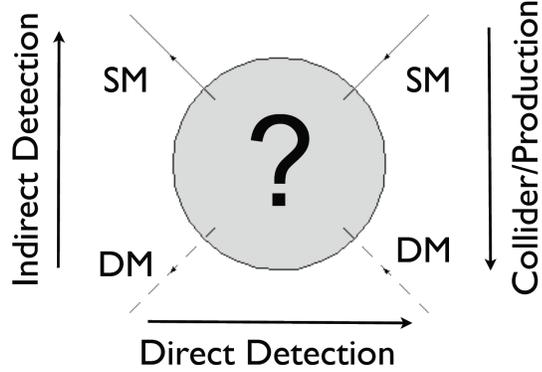


Figure 1.4: Paths to dark matter discovery. All of direct, indirect, and production approaches are currently being pursued by various research efforts.

### 1.3.1   Indirect Detection

Many experiments are looking for SM products (either primary or secondary) of DM self-annihilation or decay. The largest signals are naturally expected to come from areas of largest and nearest DM density such as the sun and the core of our galaxy. These experiments commonly look for unexplained excesses in the spectra of photons, cosmic rays, or neutrinos. However, such signals are expected to be small, so the searches require careful modelling and broad understanding of many background astrophysical processes.

For decay searches, the DM lifetime is constrained to be significantly longer than the age of the universe for there to still be a significant DM population today [51]. DM decay or annihilation in the early universe can affect the details of the CMB by causing excess ionization. This allows data from the Planck satellite to set limits on the DM self-annihilation cross section [24].

For annihilation to gamma rays, the measured flux from a DM halo should scale as the so-called "J-factor", which goes roughly as $J \sim \frac{V \rho_{DM}^2}{d^2}$ where $V$ is the volume of the

source and $d$ is the distance to it [52]. Thus the most promising sources are the galactic center and Milky Way dwarf satellites, which have dense, nearby globs of DM. The Fermi Gamma-Ray Space Telescope has measured the gamma ray spectrum from Milky Way satellite galaxies and found them to be consistent with expected backgrounds. This constrains the rate of DM annihilation to gamma rays from DM self-annihilation and provides limits on this cross section. The limits exclude the thermal relic cross section for DM masses from $\sim$5–100 GeV/c$^2$ and are competitive up to 10 TeV/c$^2$ [53]. The galactic center has a much higher J factor, but is co-located with many astrophysical processes which produce backgrounds that are not yet fully understood (and still hotly debated [54, 55]). Nevertheless, by combining current models of the DM distribution in the galaxy with gamma ray measurements (again from Fermi), limits on the DM annihilation cross section are found to be similar with those obtained from satellite galaxies [56], but depend strongly on the DM profile. Higher energy gamma signals are examined using arrays of atmospheric Cherenkov telescopes such as HESS [57], MAGIC [58], and VERITAS [59]. These experiments study both the galactic center and Milky Way satellites and are able to set limits on DM cross sections for masses from $\sim$0.1–100 TeV/c$^2$.

Most models of DM annihilation to cosmic rays will produce equal amounts of matter and antimatter. The cosmic ray spectrum from known astrophysical sources is generally dominated by matter, so many DM search programs seek to detect an excess of anti-matter cosmic rays [52]. These particles must travel diffusively through the galaxy and thus generally do not propagate as far or as cleanly as gamma rays. Antiproton data from the AMS-02 instrument aboard the International Space Station has been used to set strong limits including a possible indication of DM signal [60, 61]. However, there are significant systematic uncertainties in these measurements due to modelling of the complicated propagation of charged particles through the galaxy and ISM. AMS-02 [62] and PAMELA [63] also measure positron spectra and are able to set strong limits for DM masses from 10 GeV/c$^2$–1 TeV/c$^2$.

Neutrinos are another possible DM annihilation product. Because neutrinos below about 1 TeV are relatively unimpeded by matter, the strongest sources of these signals are actually expected to be from DM gravitationally captured by the Earth or Sun. Both

IceCube [64] and ANTARES [65] have had null results searching for the WIMP self-annihilation neutrino signal but have set some of the strongest limits for spin-dependent WIMP-proton cross section from 20 GeV/c$^2$–10 TeV/c$^2$. The Super-Kamiokande experiment also looks for this signal but is sensitive to lower neutrino energies and thus lower WIMP masses of 4–200 GeV/c$^2$ [66].

Several direct searches for the axion are also currently under way. The Axion Dark Matter Experiment (ADMX) uses a tunable microwave cavity threaded by a magnetic field to search for an axion signal. The experiment seeks to measure excess power output from the cavity due to the resonant conversion of axions to photons. Initial searches set limits for range of axion mass from 1.9–3.53 $\mu$eV/c$^2$, which excluded some axion models from making up all of the DM halo [67]. More recent efforts have produced limits in a narrower mass range of 2.66–3.31 $\mu$eV/c$^2$, but strong enough to exclude further benchmark axion-as-DM models [68]. The CERN Axion Solar Telescope (CAST) is another attempt to detect the solar axion flux by reconversion into a pair of photons in the magnetic field of the device. CAST has set set strong limits for axion masses below 20 meV/c$^2$ [69] and has reached the parameter space favored by some theoretical models at higher masses (0.64 eV/c$^2$ < $m_a$ < 1.17 eV/c$^2$) [70]. While an attractive DM candidate, a direct detection of the axion remains elusive.

### 1.3.2   Production

DM may be produced in a variety of ways in particle collision experiments. This type of experiment is not dependent on the details of the local DM density and low-energy interactions with SM matter because the DM production rate would be set by the accelerator's energy and luminosity.

The NA64 experiment at the CERN SPS has searched for missing energy in the scattering of high energy electron beams with fixed target nuclei [71]. In particular they considered the process of a dark photon created via kinetic mixing with bremsstrahlung photons from an electron scattering, followed by the decay of the dark photon into some DM particles which are not measured in the detector.

$$e^- Z \rightarrow e^- Z A'(A' \rightarrow invisible) \tag{1.10}$$

By analyzing such data, they have set limits on the kinetic mixing parameter, $\varepsilon$ for dark photon masses, $m_{A'}$, between 1 MeV/c$^2$ and 1 GeV/c$^2$.

The BaBar Collaboration has sought dark photon signatures in electron-positron collision data. They search for missing mass that would occur in the process

$$e^+e^- \to \gamma A'(A' \to invisible) \tag{1.11}$$

where the usual second photon is converted via kinetic mixing to a dark photon which subsequently decays to something in the dark sector. These experiments have yielded stronger limits than NA64 for $m_{A'}$ from 200 MeV/c$^2$–10 GeV/c$^2$ [72].

CMS and ATLAS have published results on dozens of searches for both WIMP and dark photon DM. An overview of these is beyond the scope of this work, but the searches generally involve studying evidence for DM models with missing transverse momentum in proton-proton collisions up to 13 TeV [73, 74]. An example exclusion limit on the spin-dependent WIMP interaction cross section is shown in Fig. 1.5. All such results have so far been generally consistent with the SM-only hypothesis.

### 1.3.3  Direct Detection

The direct detection path to DM discovery relies on a substantial DM presence in our galaxy. The local DM density has been measured to be in the range of 0.3–0.4 GeV/cm$^3$ using a host of measurements of the dynamics of bodies in our solar neighborhood [76]. As the sun orbits the galactic center, we pass through this DM halo at speed of $\sim$230 km/s [77]. This produces a substantial flux of DM through the Earth and allows the possibility of occasional DM-SM interactions which could be measured with a suitable terrestrial detector.

#### Direct Detection Signals

WIMPS have been the main target of direct detection experiments for the last few decades. WIMPs interact with normal matter via either elastic scattering directly off of nuclei, causing a recoil, or inelastic collisions by exciting an atom's electrons or nucleus to a higher energy state [33]. In these interactions, the WIMP could couple to the spin or mass of the target particle. Spin-independent WIMP couplings depend on the
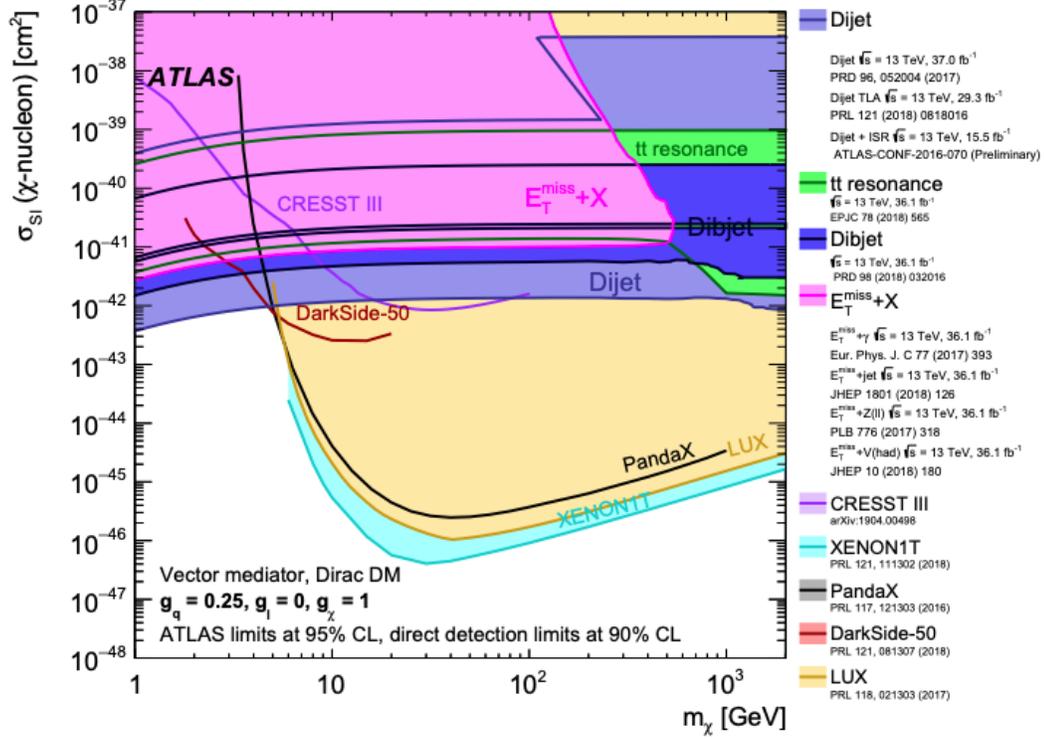
Figure 1.5: ATLAS limits on spin-independent DM-proton scattering assuming a vector mediator. Reproduced from [73] under CC License [75].

target's mass, making heavier target materials preferable, while target atom size isn't of paramount importance for a spin-dependent coupling.

While the spin-independent and spin-dependent interactions are the most discussed, there are also other interaction possibilities including some that depend on momentum transfer and/or angular momentum. An effective field theory (EFT) has been developed which considers such possibilities and no less than 11 different allowed interaction operators have been identified [78]. These operators can individually result in rather different signatures in detectors of different materials and and may even interfere with each other. While a detailed study of EFT is beyond the scope of this work, it is certainly a rich and active area of research which opens up new possibilities for how DM could manifest in our detectors [79, 80]. EFT analyses are also being considered by other direct DM search experiments such as LUX [81] and CRESST-II [82].

A first order calculation of detection rate and nuclear recoil spectrum for spin-independent coupling is given in [83]. There, it is shown that the rate of recoil events (in units of counts per kg of detector material per day) is

$$R_0 = \frac{361}{m_\chi m_N} \left( \frac{\sigma_0}{1 \text{ pb}} \right) \left( \frac{\rho_\chi}{0.3 \text{ GeV/cm}^3} \right) \left( \frac{v_0}{220 \text{ km/s}} \right) \text{kg}^{-1} \text{ day}^{-1} \qquad (1.12)$$

where $m_\chi$ and $m_N$ are the WIMP and target nucleus masses, respectively, in GeV/c$^2$, $\sigma_0$ is the WIMP-nucleon interaction cross section, $\rho_\chi$ is the local DM density, and $v_0$ is the average velocity of Earth through the WIMP cloud. Using a target material with $\mathcal{O}(10)$ nucleons, WIMP masses of $\lesssim 5$ GeV/c$^2$ will thus result in a few counts/kg/day. However, current limits ($\sigma_0 \lesssim 10^{-43}$ cm$^2$) reduce this rate to a few counts/kg/year.

As calculated in Sec. A.2 the shape of the measured nuclear recoil spectrum is driven by the Maxwellian DM velocity distribution. This leads to a detection challenge. The spectrum of these interactions is expected to be exponentially falling so many of the interactions occur at low energies, which are more difficult to measure. In particular the differential rate of nuclear recoil scattering is roughly

$$\frac{dR}{dE_{NR}} = \frac{R_0}{rE_0} e^{-E_{NR}/rE_0} \qquad (1.13)$$

where $E_{NR}$ is the nuclear recoil energy, $E_0 = \frac{m_\chi}{2} v_0^2$ is the average incoming WIMP energy, and $r = \frac{4m_\chi m_T}{(m_\chi + m_T)^2}$ is a classical elastic collision factor. Including further details like the target nuclear form factor and realistic models of the DM velocity distribution results in spin-independent recoil energy spectra like those shown in Fig. 1.6. There we can primarily see the effect of the form factor of different materials, which suppresses the recoil signal at certain energies.

Recent direct detection efforts have been shifting to searches for other types of DM such as those involving dark photons. These models involve interaction with electrons in the detector material and open up sensitivity to lower mass ($<$1 GeV/c$^2$) DM candidates. Through kinetic mixing, dark photons can convert to SM photons which are absorbed on detector electrons which then recoil at an energy of $m_{A'}$, the dark photon mass. The rate of this interaction in a detector was calculated in Eq. 1.9.

The physics of ERDM is much more complex than for DM that produces NRs. The target electrons are bound in orbitals with a range of momentum eigenstates which
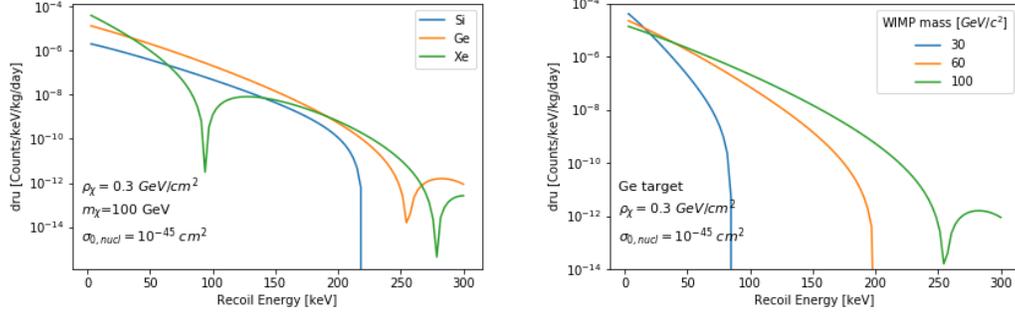
Figure 1.6: Left: Differential rate of spin-independent WIMP interactions with common detector materials. Right: Differential rate in Ge at several WIMP masses. $\sigma_{0,nucl}$ is the per-nucleon cross section scale.

may be much larger than the DM momentum, and so can not be treated as stationary targets. Additionally, the electron orbitals are not atomic because the target material is generally liquid or solid. In the case of crystalline detector materials, one even has to account for the band structure [84]. The details of this topic are beyond the scope of this thesis, but Fig. 1.7, reproduced from Ref. [50], provides some insight. There, we see the expected recoil spectrum from ERDM in Si and Ge as a function of number of electron/hole pairs produced, $Q$. It can be seen that the shape of these spectra are qualitatively similar to the NR spectra in Fig. 1.6, but at the energy scale of $\sim 10$ e/h pairs, which corresponds to $\sim 35$ eV.

### Experiments

Many different methods and technologies have been employed by a diverse array of research groups to detect DM signals directly. In the past few decades, this has grown into a field of its own right. In this section, a few of the leading experiments are highlighted.

Bubble chambers like PICO and superheated droplet detectors such as SIMPLE have been able to place competitive spin dependent limits on WIMP interactions. Both types of detectors are able to naturally discriminate electron and nuclear recoils as the target material temperature and pressure can be tuned such that electron recoils do not nucleate bubbles. The PICO Collaboration (formed from the merger of PICASSO[85]
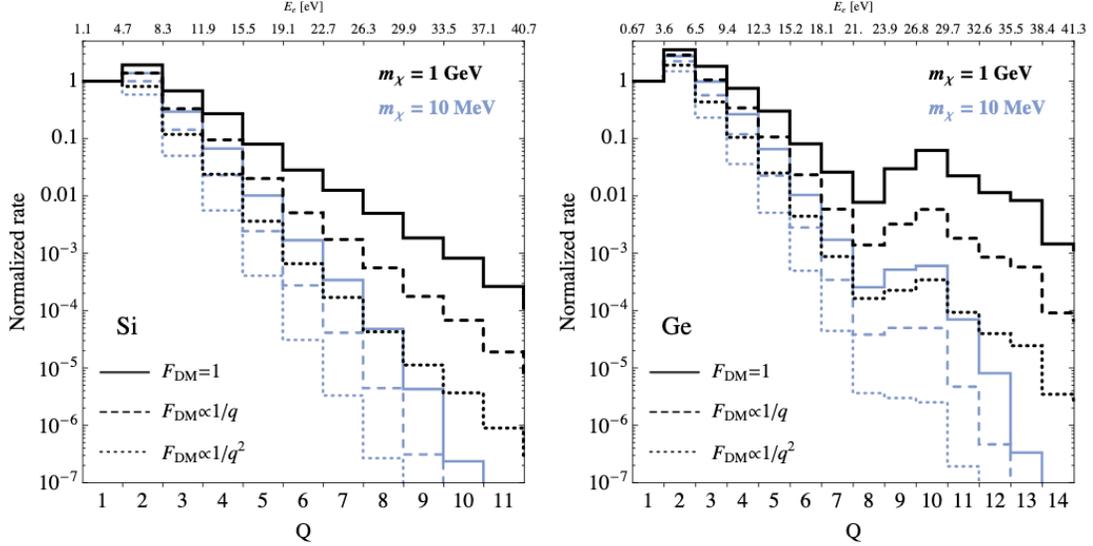
Figure 1.7: Normalized ionization signal spectrum of ERDM signals in silicon (left) or germanium (right). Here, $Q$ is the number of electron/hole pairs produced. Reproduced from [50] under CC License [75].

and COUPP[86]) used a superheated liquid $C_3F_8$ bubble chamber to search for WIMP signals in the SNOLAB facility. The device uses acoustic transducers to measure the acoustic power deposited during bubble nucleation events and arrays of cameras to monitor the timing and location of bubble creation in the containing vessel. During the most recent DM search, they were able to achieve thermodynamic threshold energies as low as $\approx 1.2$ keV. Using 1167-kg-days of search data, PICO found nuclear recoil signals consistent with neutron backgrounds, giving the strongest WIMP-proton spin-dependent limits for 6 GeV/c$^2$ < m$_\chi$ < 70 GeV/c$^2$ [87]. The SIMPLE experiment was similar to PICO, but used superheated $C_2ClF_5$ suspended in a gel matrix and microphones to measure acoustic signal amplitudes of nucleation events. SIMPLE had set strong spin-dependent limits [88] which were recently overtaken by PICO.

Another class of detectors uses a two-phase time projection chamber (TPC) partially filled with a noble liquid. The principle is that particle interactions in the liquid phase will produce scintillation light and liberate electrons. The electrons can then be drifted out of the noble liquid and into a region of gas phase where they will produce additional

scintillation light proportional to their number. The ratio and timing of these signals can be used not only to locate the event within the detector volume, but also to measure its energy and deduce whether the interaction was an electron or nuclear recoil. This technology is attractive because, being liquid, the detector mass can easily be scaled up to increase sensitivity and statistics. The largest backgrounds come from impurities in the noble liquids and much of the advancement in these devices has relied on improved methods for impurity removal. Because noble liquid detectors use relatively high mass target materials, they have been able to set the strongest high WIMP mass ($m_\chi > 6$ GeV/c$^2$) limits.

The Large Underground Xenon (LUX) experiment operated a TPC filled with a large, 250 kg fiducial target mass of liquid xenon (LXe). It was located at the Sanford Underground Research Facility near Lead, South Dakota. Like many sensitive direct search experiments, LUX was located in a deep underground lab to shield the detector from cosmic ray and muon backgrounds. In addition to the 4300 mwe (meters water equivalent) depth underground, the LUX TPC was shielded by a surrounding passive water tank [89]. The final LUX WIMP-search data taking ended in 2016 and the subsequent analysis found no evidence for a WIMP signal, excluding spin-independent WIMP-nucleon cross sections larger than $1.1 \times 10^{-46}$ cm$^2$ at a mass of 50 GeV/c$^2$. This is the strongest constraint on WIMPs above 10 GeV/c$^2$ to date. LUX has since merged with Zeplin-III, a smaller LXe TPC experiment, to form the LZ collaboration. Zeplin-III had only a 5.1 kg fiducial mass, but used a Gd-loaded plastic scintillator to moderate neutrons and act as an anti-coincidence veto system [90]. LZ combines the best of both experiments and will feature a 5.6 ton fiducial liquid mass surrounded by a gadolinium-loaded liquid scintillator veto [91]. The projected background rate of this device will allow a 1000 day exposure to reach a spin-independent cross-section sensitivity of $1.6 \times 10^{-48}$ cm$^2$ at 40 GeV/c$^2$ [92].

A few other notable TPC experiments which produce competitive limits are: Xenon1T, with a 2 ton LXe fiducial volume surrounded by a water Cherenkov active muon veto [93]; PandaX, a 580 kg LXe [94] at the China Jinping Underground Laboratory, the best-shielded laboratory in the world; and Darkside, a liquid argon TPC surrounded by both a layer of boron-loaded liquid scintillator and an active water Cerenkov detector to veto muons, neutrons, and other radioactive contaminants [95].

Solid state WIMP detectors are generally crystalline semiconductors and/or scintillators. Some solid state detectors use an electric field to drift liberated charges across a detector. The amount of charge is measured capacitively before recombining at the charged electrodes. In addition, the detectors can be operated as sensitive bolometers which measure the phonon energy deposited in the system by the interaction. Others measure a combination of scintillation light and phonons. As in TPCs, the simultaneous measurement of 2 signal channels allows event-by-event identification of recoil type (electron or nuclear). Generally, such devices cannot be fabricated above the kg-scale, so experiments using these technologies deploy a large number of identical modules to increase total exposure.

The EDELWEISS experiment uses cylindrical high purity germanium crystals with the surfaces fully instrumented with interdigitated charge measurement electrodes [96]. This arrangement allows for rejection of background events near the crystal surfaces. Each detector also has a pair of sensitive thermistors to measure the total deposited phonon energy. These devices are operated at temperatures of $< 20$ mK at the Modane Underground Laboratory in France. With an array 24 such detectors surrounded by layers of passive shielding and an active muon veto, they have set competitive limits for WIMP masses of 4-30 GeV/c$^2$ [97].

The CRESST experiment uses CaWO$_4$ crystals as the target mass and phonon detector. Phonons are measured using W-TES (transition edge sensors) read out with SQUID (superconducting quantum interference device) based amplifiers. A separate silicon-on-sapphire wafer is mounted in the same detector housing and used to convert scintillation light produced in the target crystal into phonons, which are also read out with W-TESs [98]. The CRESST devices are sensitive to lower WIMP masses than other technologies because the target material contains the relatively light oxygen atoms which have favorable recoil kinematics with WIMP masses down to $\sim 500$ MeV/c$^2$. A preliminary analysis of the spectrum from a new, more sensitive set of devices produced a world-leading limit below 1.7 GeV/c$^2$ [99].

The SuperCDMS (Cryogenic Dark Matter Search) experiment makes use of both germanium and silicon monocrystaline detectors. Like EDELWEISS, the SuperCDMS detectors are patterned with several charge biasing and measurement electrodes. Similar to CRESST, the SuperCDMS detectors make use of W-TESs read out by SQUIDs to

precisely measure phonon energies [100]. SuperCDMS has also developed a high voltage detector configuration which is optimized to search for low mass WIMPs and which set competitive limits in the $m_\chi < 10$ GeV/c$^2$ region [101]. An in-depth look at these detectors will be the subject of the next chapter.

Finally, of special note is the DAMA experiment, which operates an array of NaI crystals located in the underground Gran Sasso National Laboratory, and has long claimed detection of a WIMP signal [102]. Because it only measures scintillation information, the DAMA experiment is unable to distinguish between electron and nuclear recoils, but instead looks for the annual modulation of the total signal. DAMA has collected and impressive 2.46 ton-yr of exposure and observed an annual modulation signal with a $12.9\sigma$ confidence level [103]. However, the WIMP parameter space favored by DAMA has been searched and strongly excluded by many other experiments. While numerous non-DM explanations have been put forward, none has been able to sufficiently explain the signal. One recent suggestion involves the possible contamination of the nitrogen purge gas with radioactive argon, which can produce seasonal variations of 2.7 keV x-rays [104]. Another recent paper has concluded that the most recent DAMA/Libra data which extends to lower recoil energies than previous, is now incompatible with vanilla WIMP models [105].
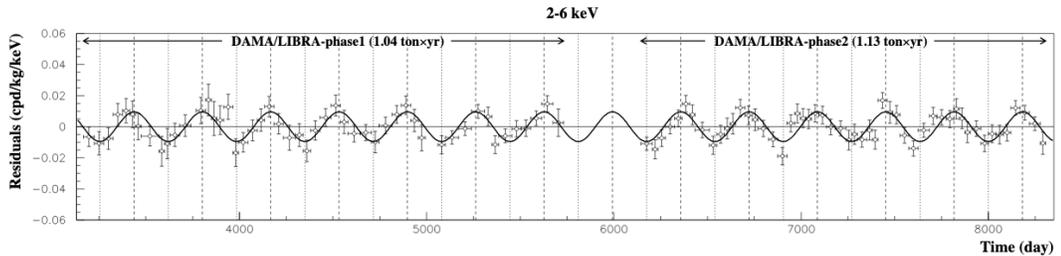


Figure 1.8: Annual modulation of the DAMA/Libra signal over the last 2 decades. The claim of a DM origin to this signal has long been refuted by competing experiments. Reproduced from Ref.[103] under CC License [75].

At any rate, several projects (COSINE-100[106], ANAIS[107], SABRE[108]) are repeating DAMA's measurement using similar NaI detectors to shed further light on this anomaly. Recent results from COSINE-100 experiment provide evidence that the nuclear recoil quenching factors (critical for calibrating the energy scale of NRs) that

DAMA/Libra has been using are incorrect at low energies [109]. Correcting for this effect, they are able to exclude the DAMA/Libra detection with 90% confidence. ANAIS has begun taking data and expects to be able to place similar constraints in the next few years. SABRE is still in a development phase, but will have the unique feature of placing twin detectors in the northern and southern hemispheres to distinguish between DM-like annual modulation and seasonal terrestrial variations [110].

We end this chapter with a plot summarizing the excluded spin-independent parameter space shown in Fig. 1.9. As discussed above, the liquid noble experiments have the best limits at high masses while solid state devices reach to lower masses. Projections for the next generations of SuperCDMS and LZ are also shown as dashed curves. The lower shaded region is where these experiments will begin to be sensitive to the background of solar and atmospheric neutrinos. Clearly the next few years will be an exciting time in the direct-DM search community as various groups push to carve out more of this parameter space!
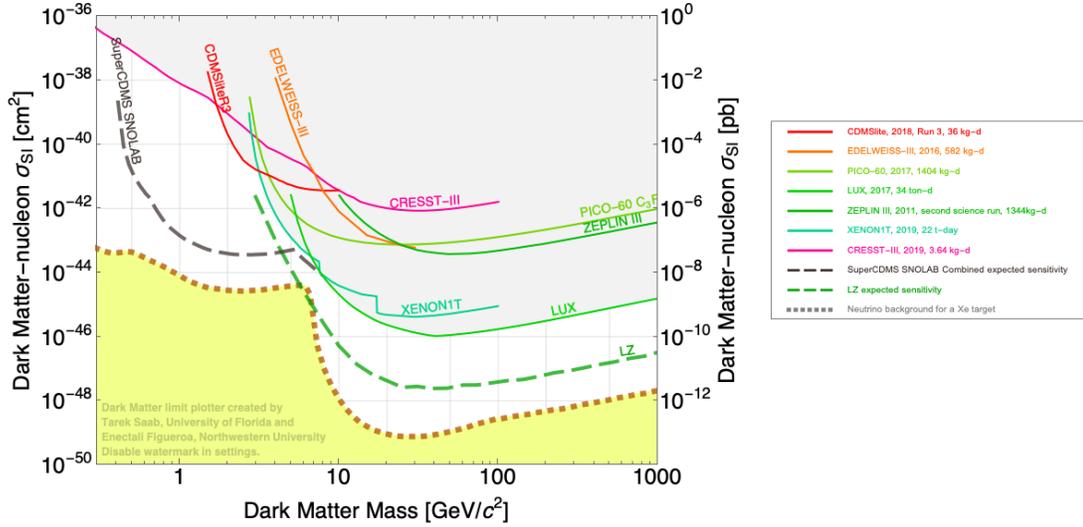


Figure 1.9: Current state of spin-independent parameter space. Produced using the Dark Matter Limit Plotter [111].

# Chapter 2

# SuperCDMS

## 2.1 Collaboration and Facilities

The Super Cryogenic Dark Matter Search (SuperCDMS) experiment operates arrays of monocrystaline germanium and silicon detectors at cryogenic temperatures ($<$50 mK). The detectors are instrumented to measure ionization and/or athermal phonon signals. The details of how they operate will be discussed in Sec. 2.2. This chapter will give a very brief overview of the history of SuperCDMS. The most recent phase of the experiment used these devices to search for DM and was conducted in the Soudan Underground Laboratory. The collaboration is currently preparing for a new campaign at the SNOLAB facility. There are also numerous test facilities at SuperCDMS member institutions which support the flagship experiment through testing and R&D efforts.

### 2.1.1 Early CDMS

The first CDMS DM searches were undertaken in 1996 at the Stanford Underground Facility (16 mwe) using early 1 cm thick detectors. Even though the site location only reduced the cosmic ray muon flux by a factor of 5 from that at the surface, the experiment was able to set an exclusion limit on the WIMP-nucleon cross section of $2.8 \times 10^{-42}$ cm$^2$ at 100 GeV WIMP mass. This result was important at the time as it excluded a region of parameter space in which the DAMA experiment had claimed detection with more than 84% confidence [112].

## 2.1.2 The Soudan Era

Beginning in 2001, the experiment was expanded and moved to the lower background environment of the Soudan Underground Laboratory at a depth of 2090 mwe. There, CDMS-II used a payload composed of 19 Ge and 11 Si detectors with upgraded sensitivities to exclude still lower cross sections ($1.8 \times 10^{-44}$ cm$^2$ at 60 GeV/cm$^2$) and search for annual modulation [113]. As shown in Fig. 2.1, the detectors were surrounded by a plastic scintillator to provide a veto measurement of any muons that did penetrate the rock overburden. Furthermore, the detectors were enclosed in ancient (radiopure) lead and polyethylene shielding. The lead was used to block gamma rays while the polyethylene thermalized neutrons, the dominant nuclear recoil background.

Analysis of the Si detector data revealed 3 candidate events, while known backgrounds estimated <1 event, hinting at an 8.6 GeV/cm$^2$ WIMP with $1.9 \times 10^{-41}$ cm$^2$ cross section [114]. Although a profile likelihood analysis favored the WIMP hypothesis, the data was not strong enough to indicate a discovery. However, this same parameter space was soon ruled out by the LUX experiment as well as by the next generation of CDMS, dubbed SuperCDMS.

SuperCDMS operated a new design of Ge detectors which greatly improved volume fiducialization. Analysis of 577 kg-days of data found that the measured recoil energy spectrum was consistent with expectations from background sources. This data set stronger limits for WIMP masses between 4 and 6 GeV/cm$^2$ [115].

An extension of this was the CDMSlite experiment. It used a single SuperCDMS iZIP with modified electronics to boost its voltage bias to 70 V, amplifying low energy phonon signals to probe lower WIMP masses. A 2016 analysis of this data excluded more of the parameter space between 1.6 and 5.5 GeV/cm$^2$ [116]. Fig. 2.2 shows spin-independent limits from leading experiments after the SuperCDMS Soudan era. Also shown are the target sensitivities of the next generation of detectors (dashed lines).

SuperCDMS ceased operations at the Soudan Underground Facility in the autumn of 2015. However, the wealth of high-quality data collected continues to yield interesting and relevant science. Since 2015, the SuperCDMS collaboration has used this data for publications studying WIMPS [116, 100, 118, 101], Effective Field Theory [79], Frenkel Defects [119], nuclear-recoil ionization yield [120], cosmogenic tritium production in Ge [121], and even dark photons and axion-like particles [122].
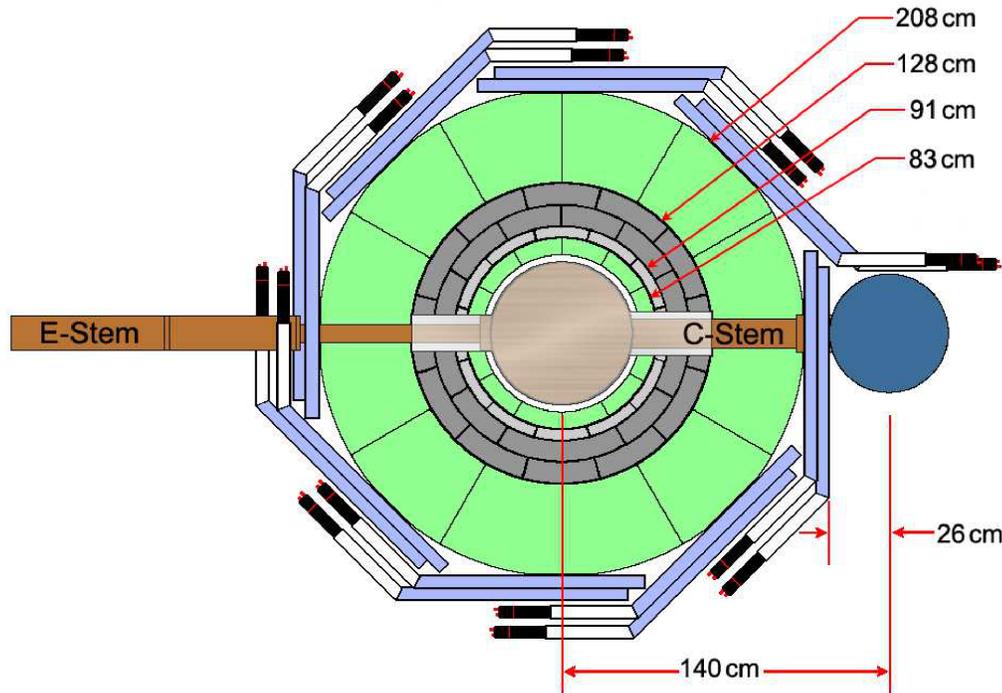
Figure 2.1: Top-down view of SuperCDMS-Soudan cryostat and shielding. Veto panels are shown in blue, polyethylene sheets in green, lead bricks in dark gray, and ancient lead in light gray. The E-Stem provided a path for the electronics and wiring to reach room temperature while the C-stem connected the various thermal stages of the cryostat to a dilution fridge.

### 2.1.3 SuperCDMS SNOLAB

As of this writing, the collaboration is preparing for the next phase of the experiment, SuperCDMS SNOLAB, which will incorporate more sensitive detectors, more stable cryogenics, and improved electronics and DAQ systems. The detectors and cryogenic infrastructure will be constructed of extremely radiopure materials. All of these materials are being assayed to ensure they have substantially lower background radiation than in the Soudan setup. A diagram of the core experimental infrastructure is shown in Fig. 2.3.
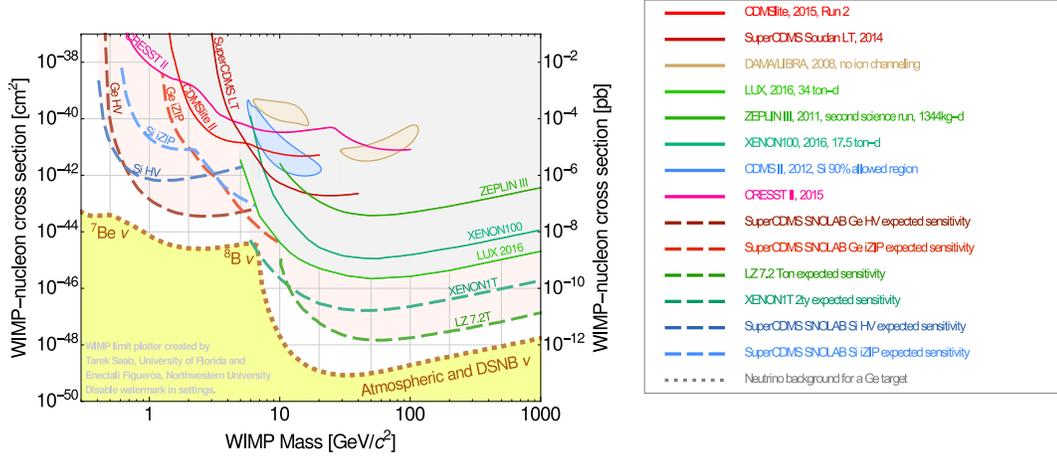
Figure 2.2: Measured and projected exclusion limits after the end of the SuperCDMS Soudan Era (circa 2016). Created using [117].

This experiment will be located at SNOLAB (6010 mwe), home of the famous Sudbury Neutrino Observatory (SNO), which helped solve the solar neutrino problem. This facility provides improved shielding from cosmic rays (nearly $3\times$ the effective depth of Soudan) and has strict clean room protocols to further reduce radiogenic background contamination.

SuperCDMS will deploy a variety of next-generation detectors to thoroughly characterize the experimental backgrounds and explore lower mass and cross section DM parameter spaces. These detectors are $2.3\times$ larger than those used at Soudan and feature more sensitive phonon sensors. They will also use a new type of high-electron-mobility transistor (HEMT) -based charge amplifier instead of the JFETs used in Soudan (and in the work presented in this thesis). A mixture of several detector types (iZIP and HV) and materials (Si and Ge) will be deployed to provide complementary of DM targets, measurement of the various backgrounds, and reach to sub-GeV DM. Accordingly, much of the CDMS detector development program of the last few years has been dedicated to optimizing the performance and understanding of these new detectors. The final sensitivity projections of the experiment are shown in Fig. 2.2

While many of the subsystems involved in this effort are in mature form, a late-stage cryostat redesign and COVID-related shutdowns have caused some delays in the project
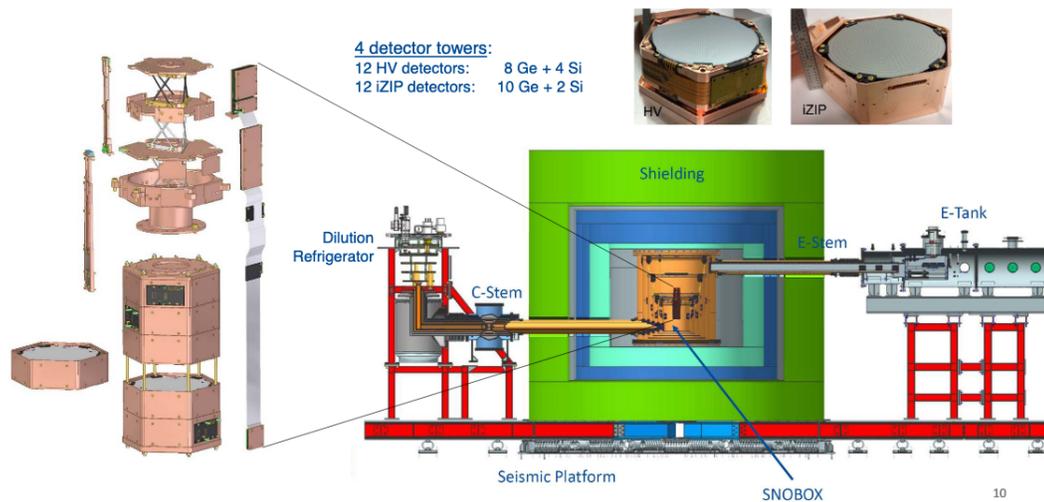
Figure 2.3: Side view of SuperCDMS SNOLAB experiment design featuring inner and outer neutron shielding (polyethylene) and gamma shielding (lead). The dilution refrigerator and the "E-tank" (where readout electronics will be mounted) are connected to either side of the "SNOBOX" cryostat. Also shown on the left is a diagram of a stack (or "tower") of detectors and photos of two detectors at the top right.

timeline. As of this writing, much of the electrical infrastructure and a radon filtration system have been installed and the seismic platform is currently being put in place at SNOLAB. Meanwhile, the dilution fridge has been delivered to FNAL and is being tested above-ground and the first detector towers have been assembled and are being tested at SLAC. The current arrival time of the cryostat cans is slated for late 2021 and our schedule indicates we can expect to do the first cool down with detectors in 2022. Following that, dedicated science data taking campaigns are expected to begin in 2023. While this is still a few years away, there is still plenty of momentum and enthusiasm within the collaboration to deliver world-leading science at SNOLAB.

### 2.1.4 SuperCDMS Test Facilities

While the main experiment at SNOLAB is designed and constructed, a huge amount of R&D and preparation has been taking place at other SuperCDMS facilities and member institutions. A few of them are described here.

**CUTE**

CUTE is the (slightly strained) acronym for the "Cryogenic Underground TEst facility". It is a cryogenic detector test facility, located underground at SNOLAB in the same experimental hall as the SuperCDMS experiment. It uses a cryogen-free dilution refrigerator (manufactured by Cryoconcept [123]) to reach base temperatures of $\sim 15$ mK. The combination of stringent background controls at SNOLAB and extensive water and lead shielding give CUTE a background of order of a few events/keV/kg/day in energies below a few keV. The main purpose of CUTE is to provide the ability to test detector properties that are difficult or impossible to study at surface facilities which are limited by background neutron and gamma radiation. CUTE has already begun some early operations and tests of SuperCDMS detector assemblies.

**NEXUS**

The "Northwestern EXperimental Underground Site" (NEXUS) is a shallow facility, located $\sim 100$ m underground in the MINOS near-detector experimental hall at the Fermi National Accelerator Laboratory (FNAL). It is built, maintained, and operated through a collaboration between Northwestern University and FNAL. The overburden lowers the muon flux to 0.5 muons/m$^2$/s. Similar to CUTE, it features a Cryoconcept [123] dry dilution refrigerator. A movable lead shield designed to surround the cryostat is expected to reduce the background to less than 100 events/keV/kg/day.

In addition to acting as another underground facility for detector testing, NEXUS will operate a Deuterium-Deuterium (D-D) neutron generator. This device can produce a large flux ($10^8$ neutrons/sec) of 2-3 MeV neutrons. In combination with an array of backing detectors, this facility will be used to characterize the response of SuperCDMS detectors to low-energy neutron scatters. Some early experimental runs at NEXUS were conducted beginning in the summer of 2019 and there are also plans to conduct low-mass ($< 1$ GeV/c$^2$) DM searches using a new style of small, but sensitive, detectors. This will be discussed further in Ch. 8. With the current setup, NEXUS can easily obtain exposures competitive with recent above-ground DM searches carried out with the same type of detector [47, 48].

NEXUS is not just meant to be used by SuperCDMS. It is already host to a TES

R&D test for the Ricochet Collaboration, which plans to measure Coherent Elastic Neutrino Nucleus Scattering (CEvNS) at a nuclear reactor [124]. There are also plans to study coherence times of superconducting microwave resonators [125] (a key component of certain types of quantum computers) that can be improved by moving to a low-background facility, as has recently been shown elsewhere [126]. Several other potential projects are also lined up including microwave kinetic inductance detectors (MKIDs) [127], testing exotic photodetector materials with meV band gaps, and state of the art application-specific integrated circuits (ASICs) at 4 K for charge measurement.

### UMN K100 and others

In addition to the facilities mentioned above, the SuperCDMS collaboration has always been heavily supported by a number of cryogenic test facilities located at member institutions. A few notable current contributors are UC Berkeley, Caltech, Northwestern, Queen's University, SLAC, Stanford, Texas A&M, and the University of Minnesota. These (and other) groups have all been integral in advancing the existing techniques and technologies and developing ever more sensitive detectors.

Of particular importance to the author is the University of Minnesota (UMN) detector test facility, as this is where much of the work in this thesis was undertaken. The UMN group operates a Kelvinox-100 $^3$He-$^4$He dilution refrigerator [128], mounted on vibration isolating columns, which is used to cool detectors to $\sim$30 mK. This facility tests and characterizes both production and R&D detectors by analyzing TES parameters, charge and phonon collection efficiency, detector neutralization hold time, and general detector performance. In addition, UMN is active in developing and testing new DAQ and analysis software as well as electronics and cold hardware (Sec. 5.2). The Super-CDMS test facility at UMN plays an integral role in analyzing concept detectors (Ch. 6) and characterizing production detectors (Sec. 5.1) for SuperCDMS-SNOLAB. Finally, UMN is even making new advances in the science and measurement of low energy NRs in Si and Ge (Ch. 7).

## 2.2 Detectors

This section will provide an introduction to how signals are generated and measured in SuperCDMS detectors. It will also describe several of the specific detector designs currently employed by SuperCDMS.

### 2.2.1 Electron- and Nuclear-Recoils

All versions of CDMS detectors measure some combination of ionization and/or phonon energy produced in a semiconductor (Si or Ge) crystal. We first discuss how incoming particle interactions leave their signatures in these quantities.

Incoming particles can interact with either the electrons or nuclei of the crystal. This interaction can be an elastic or inelastic scatter or absorption. Whatever the case, the result is generally that the affected electron or nucleus recoils in response. This leads to the classification of interactions as either an electron recoil (ER) or nuclear recoil (NR). ERs are common as they result from photon scattering/absorption from background or calibration sources. NRs occur from processes such as scattering of free neutrons, but are also how many models of DM are expected to interact. In either case, the recoil energy, $E_R$, goes into a combination of athermal phonons in the crystal lattice and promotion of electrons across the semiconductor band gap, creating electron/hole (e/h) pairs. The amount of energy that ends up in each of these systems depends on the recoil energy and type of particle interacting. Generally, NRs produce less charge than ERs at a given recoil energy. This difference is characterized by a quantity variously called the ionization yield, efficiency, or quenching factor. A similar behavior is observed in scintillators, with reduced light output occurring for NRs. In this work, we will use the name ionization yield and define it as

$$Y(E_R) = \frac{E_Q}{E_R} \tag{2.1}$$

where $E_Q$ is the amount of energy that goes into the production of e/h pairs and $E_R$ is the recoil energy. This is defined such that, for an ER, $E_Q = E_R$, giving $Y = 1$, while for NRs, the yield, and hence the number of e/h pairs produced, is lower. $E_Q$ is related to the number of electron/hole pairs that are produced, $N$ as

$$\langle N \rangle = E_Q/\varepsilon_\gamma \tag{2.2}$$

with $\varepsilon_\gamma \approx 3.0(3.8)$ eV in Ge(Si) [129, 130]. This is not to say that in an ER all the recoil energy is used up in producing e/h pairs. In fact the band gap minimum for Si (Ge) is only $E_{gap}$=1.12 (0.66) eV, so the remainder of the energy goes into the phonon system as well. $\varepsilon_\gamma$ and $Y$ are then just empirical quantities which describe the amount of free charge resulting from the interactions as the recoiling object (electron or nucleus) slows down in the crystal. The variance of the number of pairs produced is $\sigma_N^2 = F\langle N \rangle$, where $F$ is the Fano factor, which describes deviation from Poisson statistics resulting from correlations in the various microscopic interactions involved in the ionization process [131]. $F$ is around 0.12 (0.11) in Si (Ge) [132]. It should be noted that present measurements of $\varepsilon_\gamma$ and $F$ are generally for higher energies and material temperatures than the SuperCDMS detectors operate. However, there are current efforts to leverage current theory to extend our understanding to lower energies and temperatures [133]. Models and previous measurements of the yield in Si and Ge are shown in Fig. 2.4. Our work to address the tension in these measurements and lack of data at low energy will be discussed in Ch. 7.
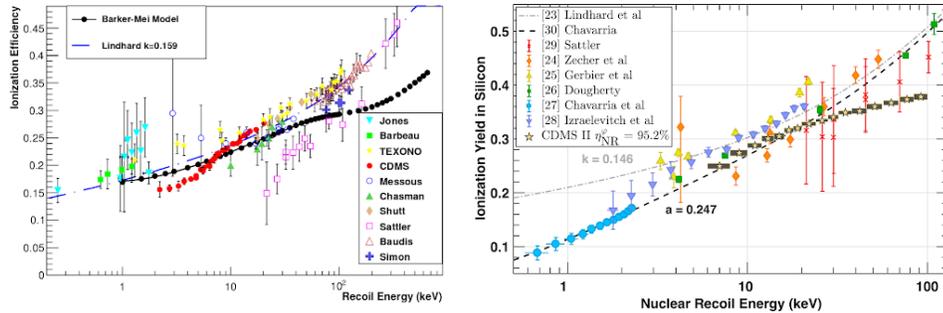


Figure 2.4: Left: Ionization yield measurements in Ge. Taken from [134]. Right: Ionization yield measurements in Si. Taken from [120]. There is significant tension in many of the results and very few measurements at recoil energies below 1 keV.

Because of the difference in ionization yield between ERs and NRs, a simultaneous measurement of the recoil energy and amount of charge produced can provide event-by-event discrimination between interaction types. The majority of backgrounds result in ERs so this discrimination is a powerful tool when searching for WIMP recoils on nuclei. The next two sections provides more details on how the charge and phonon measurements are actually performed.

### 2.2.2   Charge Signal

The detectors are operated with an axial electric bias field so liberated charges are drifted to either face across the detector bulk as shown in Fig. 2.5. As they travel, these carriers induce image charges on pickup electrodes patterned on the detector surfaces. The current flowing into the electrodes to produce the image charges is described by the Shockley-Ramo Theorem [135]. Although the charges moving through the detector bulk follow a path dictated by the bias electric field (and crystal properties), the induced signal in an electrode is determined by how the charge moves through that electrode's "weighting potential". The details of this will not be discussed here, but the result is that the signal measured on charge electrodes is induced both by the charges moving toward and away from it, and is (to first order) equal to the initial charge created in the detector. However, charge trapping sites in the crystal bulk can reduce and distort the measured signal. A more complete model of charge signal creation is presented in Appendix B.
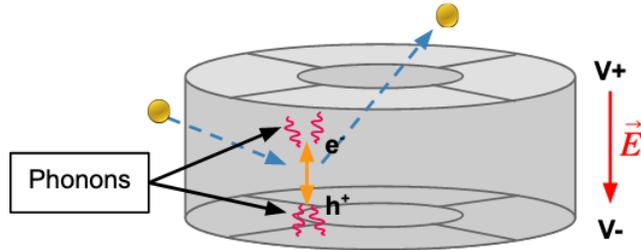


Figure 2.5: Anatomy of an event in a SuperCDMS-style detector. The incoming particle deposits energy which produces electron/hole pairs and phonons. An electric field across the detector drifts the liberated carriers to the two crystal faces. As they drift, more NTL phonons are produced.

Things are further complicated as the two types of charge carriers do not both propagate parallel to the electric field. Holes travel relatively straight through the crystal with minimal spreading. Electrons, however, have been found to propagate oblique to the electric field due to the shape of the conduction band minima (valleys). This is because, in a crystal, the carriers do not simply have a scalar mass, rather they have a mass tensor, so when a force is applied, they may accelerate in a different direction. (Imagine an electron in Si like a shopping cart with a bad wheel, it doesn't

quite want to go where you push it.) For example, in [100] Ge crystals (the orientation used in SuperCDMS detectors) they spread out in four directions, each $\sim 33°$ from the z-axis [136]. This effect is not observed in [100] Si due to the different geometry of the conduction bands. This effect is also suppressed at bias fields above $\sim 5$ V/cm where the carrier energies are large enough to allow inter-valley scattering. Fig. 2.6 shows the result of a simulation of these effects. This drifting causes charge signals on the side of a detector collecting electrons to have lower spatial resolution than signals on the hole side. This non-ideal behavior means that charge signals can depend on the species of carrier and original event position which must be accounted for in data analysis.
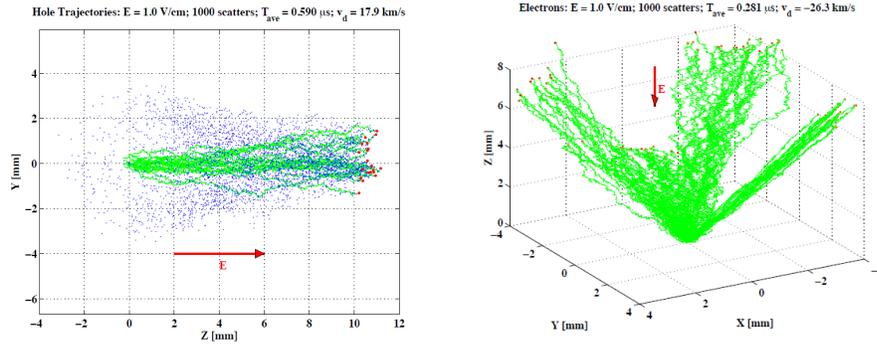


Figure 2.6: The results of a simulation of charge carrier propagation through a Ge crystal. Left: hole transport. Right: electron transport. The carrier trajectories are shown in green. In the left figure, the blue points show the locations where phonons were created in the simulation. Taken from [136].

The charge signals are picked up by metallic electrodes on the crystal and sensed via cryogenically cooled FET (field effect transistor) amplifiers. The amplifier outputs are connected to further levels of amplification and filtering outside of the cryostat before digitization. A simplified version of the first stage of the charge circuit is shown in Fig. 2.7. The shape of measured charge pulses is dictated by this circuit. SuperCDMS detectors are generally operated with bias fields of $\gtrsim 1$ V/cm. At standard operating temperatures, the charge carrier drift velocities are $\gtrsim 20$ km/s [136]. This means that, even in our largest devices, they fully traverse the 33 mm crystal thickness and induce their charge signal, $Q_{ind}$, on the measurement electrode in $\lesssim 1.7$ $\mu s$. Over these time scales the image charges that appear on the electrode are primarily sourced from the

coupling capacitor, $C_C$. The amplifier responds by changing $V_{out}$ to compensate via the feedback capacitor, $C_{FB}$, and so outputs

$$V_{out} = Q_{ind}/C_{FB} \tag{2.3}$$

The charge amplifier bandwidth limits the signal rise time to $\sim 1~\mu s$ [137, 138] so a charge pulse initially appears as a step function in voltage at the amplifier output. Thereafter, the signal decays exponentially with a characteristic fall time, $\tau_f = R_{FB}C_{FB}$, as the feedback branch discharges.
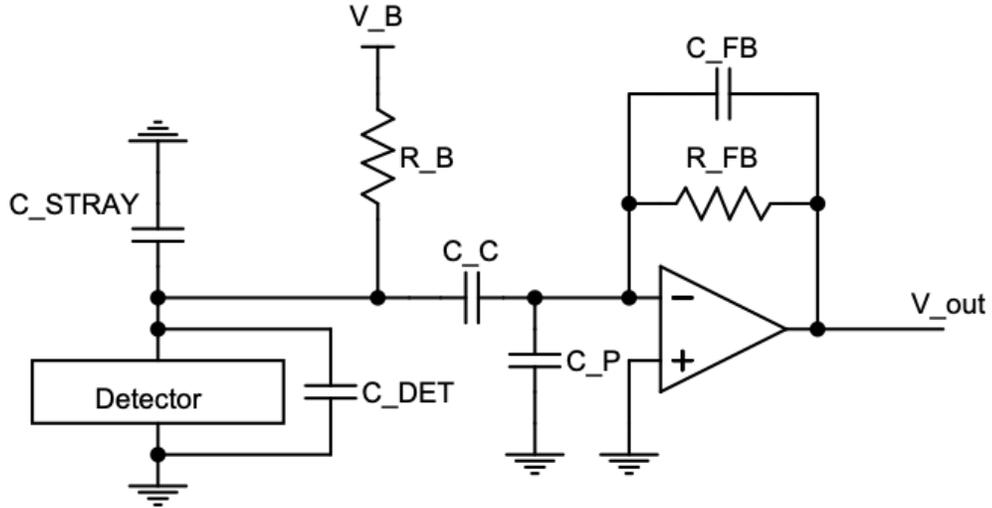


Figure 2.7: SuperCDMS charge readout circuit (pre SNOLAB).

The amount of charge induced on the measurement electrode is related to the amount of charge produced in the crystal and how it is drifted by the bias voltage. In a perfect crystal and with high bias voltage, the induced charge is simply equal in magnitude to the drifting charge, $Q_{ind} \approx Q = eN$, where $e$ is the electron charge and $N$ the number of e/h pairs produced. However, other effects of charge transport can change this relationship. First, the amount of drifting charge may not be the full amount initially ionized at the interaction site due to recombination of e/h pairs before they are drifted apart by the bias field. This will be discussed further in Ch. 6 where attempts are made to model and measure this effect. Next, once the carriers are drifting in the crystal,

some of them may be trapped on crystal impurities or defects. This stops them from drifting and contributing to the charge (and phonon signals). These drifting charges can also ionize additional carriers in a process known as impact ionization [139]. These will serve to increase the signal further. The contribution of these effects depend on a host of details such as bias field, detector geometry, and impurity and defect densities.

Now, over short time scales, the induced charge is primarily sourced from the feed-back capacitor, $C_{FB}$, via the coupling capacitor, $C_C$. However, there are also important effects due to non-negligible capacitances between the electrodes on opposite detector faces, $C_{det}$ as well as other stray capacitances between the electrode and ground, $C_{stray}$. The latter can either be directly between the electrode and housing or indirectly via other, nearby electrodes on the same face. These other capaciatances can also be sources for the image charge, and can reduce the signal seen at the amplifier by the gain factor

$$G_{cap} = \frac{C_C}{C_C + C_{stray}}.$$
(2.4)

At UMN, this has been measured in large devices to be as low as $G_{cap}$=0.7–0.8, but depends on the detector housing and electrode geometry. There may also be a significant ($\sim$ 100 pF) parasitic gate-to-ground capacitance, but this only affects things like the FET noise coupling and does not enter this capacitive gain ratio [137]. Now, over longer times than the amplifier response, the bias voltage power supply provides the amount of charge needed to restore the circuit to equilibrium. This time scale is set by $R_B C_{det} \sim 1$ ms, where $C_{det}$ is the detector capacitance. Putting these effects together, the measured signal amplitude is then

$$V_{out} \approx \frac{Q}{C_{FB}} \frac{C_C}{C_C + C_{stray}}$$
(2.5)

Knowledge of these various contributions (and later amplification stages) can allow an absolute calibration of the charge measurement which is very useful when studying charge creation and propagation effects. An example of a measured charge pulse can be seen in the left side of Fig. 2.8.

### 2.2.3 Phonon Signal

Here we give a very basic overview of phonon signals in the SuperCDMS detectors. This is no small topic and far more detailed resources (e.g. [140, 141, 142]) can provide
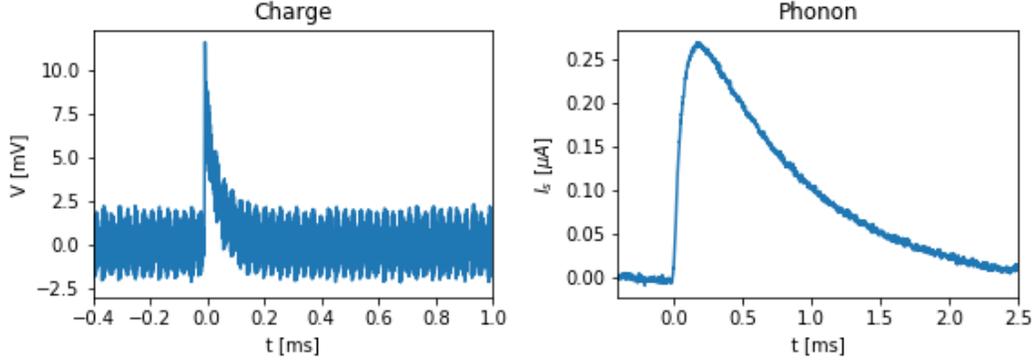
Figure 2.8: Example charge (left) and phonon (right) pulses from a 100 keV ER in an iZIP. The charge pulse has a delta function rise time and $\sim 100$ $\mu s$ fall time set by $R_{FB}C_{FB}$. The phonon rise and fall times are slower, being set by both the phonon transport and TES circuit response.

deeper insights. However, here we will cover the basics needed to understand their role in an operating detector.

Si and Ge, being Group 14 elements, have four electrons in the valence orbitals and four vacancies. Thus, when a crystal of one of these materials is formed, each atom can form covalent bonds with four nearest neighbors. In doing so, the crystals generally form a face centered cubic (FCC) lattice with eight atoms per unit cell as shown in Fig. 2.9. The lattice constant, $a$, is 5.4 (5.7) Å for Si (Ge). Other Group 14 elements, carbon and tin, can do the same. This is in fact, the same structure as diamond. Incidentally, lead also forms an FCC, but prefers to form metallic bonds.

Phonons are quantized vibrations which occur in the crystal lattice. They can be described as non-interacting bosons and so have a thermal number density described by the Planck distribution

$$\langle n \rangle = \frac{1}{\exp[\hbar\omega/k_B T] - 1} \tag{2.6}$$

where the phonon energy is $\hbar\omega$. At the mK temperatures at which the detectors are operated, this population is heavily suppressed and can be neglected for our purposes. However, significant populations of athermal phonons are produced when particles scatter with, or move through, the crystal. This is the signal we design the detectors to measure.
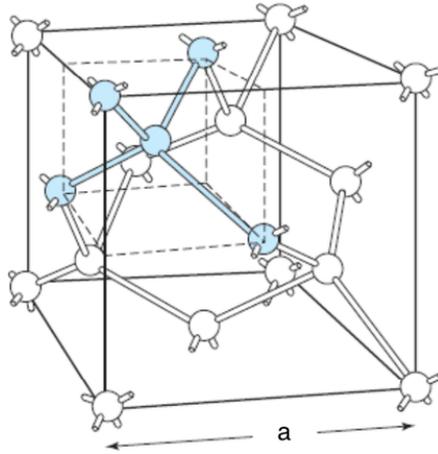
Figure 2.9: Unit cell of a diamond, FCC Bravais lattice. The unit cell contains eight atoms and each atom has four nearest neighbors. Taken from [143] with permission.

There are two main branches of phonons, acoustic and optical, which correspond to vibration of a pair of atoms in phase, or out of phase, with each other respectively. Optical phonons have higher energies and, as we will see below, quickly downconvert to acoustic phonons. Thus, for our purposes, we can assume that all phonons in the detector are acoustic. Phonons also come in several polarizations: fast transverse, slow transverse, and longitudinal [144]. The propagation speed, or sound speed, depends on the material and polarization, but ranges from 3–6 km/s in Ge [145] and from 4–9 km/s in Si [146]. In a pure crystal, the phonon energy is channeled preferentially along certain directions of the crystal lattice in an effect known as "phonon focussing". Beautiful measurements of this process have been obtained by exciting heat pulses in a small Ge cube and measuring the phonon flux on the opposite face with a bolometer as shown in Fig. 2.10.

Phonons can exist in a crystal at a range of energies. However, higher frequency ($\gtrsim 1$ THz) phonons of all types quickly downconvert to pairs of lower energy, acoustic phonons at a rate $\Gamma_{Downconvert} \propto \omega^5$ [148]. Even in a pure Si or Ge crystal, natural isotopic variation breaks the crystal symmetry. The lattice atoms are not all of the same mass and so do not comprise a uniform grid of perfect harmonic oscillators. Again, the high energy phonons are most affected, scattering off of isotopes at a rate $\Gamma_{Scatter} \propto \omega^4$ [149, 150].
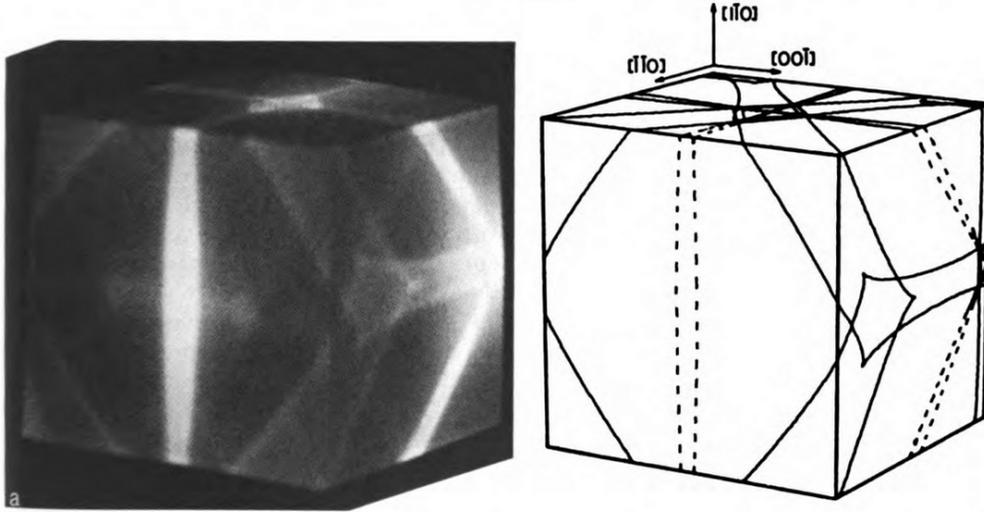
Figure 2.10: Left: Measurements of phonon signals from heat pulses in a 1 cm Ge cube. Right: Theoretical caustics of propagating ballistic phonons. Reproduced from [147] with permission.

The combination of these two effects results in high energy phonons propagating diffusively as they scatter and downcovert to lower energies. This continues until the population of phonons has low enough energies ($\sim$ 1 meV or $\sim$ 250 GHz) that they become essentially free streaming, or ballistic, with mean free paths similar to the detector scale ($\gtrsim$ cm). Eventually, all the phonons have mean free paths long enough to reach a surface of the crystal where they are either absorbed by implanted electronics or reflected from the bare surface. From this point on, the phonons move ballistically until they are either absorbed or thermalize (downconvert to thermal energies).

When a crystal is used as a particle detector, phonons are produced during both the initial scattering interaction and the subsequent charge drifting. The cascade of interactions that occur when an electron or nucleus recoils in the detector results in an initial population of "recoil phonons" which are high energy, near the Debye energy of $\sim$30 (50) meV in Ge(Si) [144]. At the low operational temperatures used, the charge carriers will immediately begin to lose energy by emitting phonons and settle down into the conduction band minimum. At this point, the total "primary" energy in the phonon

system will be

$$E_{prim} = E_R - NE_{gap} \qquad (2.7)$$

where $E_R$ is the recoil energy and $E_{gap}$ is the bandgap energy. If there is no bias field applied, the electrons and holes will immediately recombine, relinquishing the bandgap energy as "recombination" phonons, $E_{reco} = NE_{gap}$. If there is a bias applied, this recombination is deferred until the carriers are drifted and recombine with image charges at their respective electrodes. In either case, eventually the total recoil energy $E_R$ is deposited into the phonon system.

The drifting charges also produce phonons as they move across the crystal. The bias electric field accelerates the charges, but scattering with the lattice limits the effective carrier drift speed [151]. The change in charge carrier potential energy as it moves through the bias electric field is converted into phonons as the carrier scatters with lattice sites. Depending on the bias field applied, the spectrum of phonons produced may be initially diffusive (high energy) or ballistic [140]. As described by Neganov, Trofimov, and Luke [152, 153], each e/h pair will (absent trapping) contribute $eV_{bias}$ to the total phonon energy with $e$ the electron charge and $V_{bias}$ is the potential difference across the detector. These will be called "NTL phonons". Summing up the above phonon contributions gives the total energy deposited into phonons,

$$E_{PT} = E_{prim} + E_{reco} + E_{NTL} = E_R + NeV_{bias} \qquad (2.8)$$

On average, this can be written as

$$\langle E_{PT} \rangle = E_R + E_Q eV_{bias}/\varepsilon_\gamma \qquad (2.9)$$

One should be careful though, because the phonon quantity measured is not necessarily simply the total energy in the phonon system. The phonon sensors are on the two detector faces and are thus only privy to the phonon flux that reaches them. For a detailed understanding, then, one must also consider the location, spatial distributions, and spectra of the primary, NTL, and recombination phonons. A theoretical treatment of these different populations can be found in [151] and further useful discussion is in [140]. The more complete story is that the primary phonons can be thought of as emitted in an isotropic sphere at the recoil location and recombination phonons will be

emitted isotropically from the carrier recombination sites at the electrode/crystal interface. NTL phonons, however, are emitted preferentially in the carrier drift direction and along the entire drift path as shown in Fig. 2.11. So they roughly form a column of phonons propagating away from the initial recoil site towards the two detector faces. Furthermore, the spectrum of phonon energies is different in each of these cases and also depends on the material (Si or Ge) and carrier species (e- or h+). To fully model all of this requires detailed Monte Carlo simulations of all these processes, which SuperCDMS has been refining for many years.
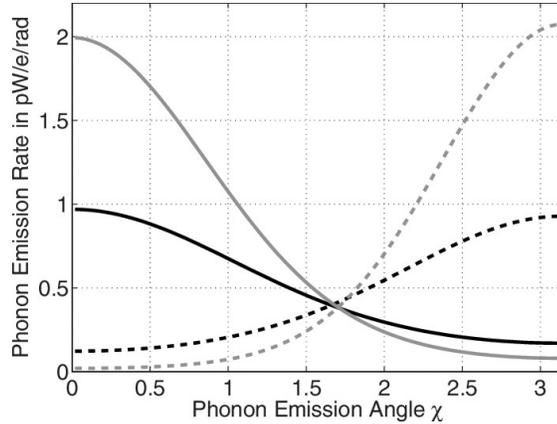


Figure 2.11: Theoretical angular distribution of NTL phonon power. Solid (dashed) lines are for electrons (holes). Black (gray) lines are for Ge at 300 V/m (Si at 400 V/m). The angle is measured from the net electron drift direction (i.e. the detector's z-axis). This shows that NTL phonons are emitted primarily in the forward direction for both types of carriers. Reproduced from [151] with permission.

Now, this is not to say that everything is hopelessly complicated. Because all of these phonon populations scatter and downconvert (albeit at different rates, locations, etc) to $0^{\text{th}}$ order they all end up as ballistic phonons rattling around inside the crystal, bouncing off of bare crystal faces or being absorbed by the phonon sensors. This is roughly the state of things after a few hundred $\mu s$, just a randomized phonon bath with all recoil position information lost but a total energy proportional to Eq. 2.9. So even without the detailed modelling, one can make decent measurements of recoil energies independent of event position. Of course taking into account these other effects is important to maximize detector performance and energy resolution.

The actual measurement of phonon energy is carried out with quasi-particle trap assisted electrothermal feedback transition edge sensors (QETs) [154]. These transition edge sensors (TESs) are operated in electro-thermal-feedback (ETF) mode and read out with SQUID based amplifiers. The detailed physics of QETs will be discussed more in Ch. 3, but we give a brief outline of their operation here. As seen in Fig. 2.12, each QET consists of a small tungsten film connected to several aluminum fins. The sensor operates at $\lesssim 50$ mK, well below the 1.2 K Al superconducting transition temperature. Phonons in the crystal which have sufficient energy ($\gtrsim 0.36$ meV) can enter the Al fins and break Cooper pairs into quasiparticles. The quasiparticles then diffuse through the fins, with a fraction of them making it into the tungsten. Those which make it can then transfer their energy by scattering with Cooper pairs and heating the tungsten. A small voltage bias maintains enough Joule heating in the tungsten film to hold it in stable equilibrium in the sharp transition between its superconducting and normal temperatures. When additional heat enters the tungsten, it causes the resistance to increase and the bias current to momentarily change. This changing current is monitored by a sensitive SQUID based amplifier and fed out of the fridge to be digitized.
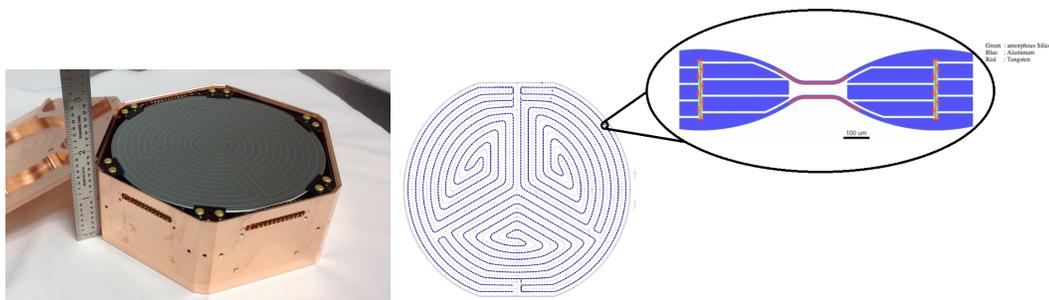


Figure 2.12: Left: iZIP detector in copper housing. Right: Charge and phonon sensor layout mask with TES details shown in the insert. The aluminum fins are shown in blue, amorphous silicon in green, and tungsten in red.

### 2.2.4  Detector Variations

SuperCDMS detectors have evolved several times since the earliest designs at Berkeley. For more info on previous designs one can consult one of many fine previous dissertations [155, 142, 156, 137]. Here we will give a brief overview of the current arsenal of detector

types commonly used by SuperCDMS.

The detectors are generally constructed out of either Ge or Si crystals to provide complementary sensitivities to different DM masses (Si recoil kinematic cutoff is lower than Ge) and backgrounds (Ge has fewer intrinsic radioactive backgrounds). There are two flagship device types that SuperCDMS will deploy at SNOLAB, the iZIP and HV detectors. Both are built from high purity [100]-oriented Si or Ge crystals shaped into cylinders 100 mm in diameter and 33.3 mm thick. The only difference is the phonon and charge sensors patterned on the crystals.

**Flagship SuperCDMS Detectors - iZIP and HV**

The interleaved Z-dependent Ionization and Phonon (iZIP) detectors feature a novel charge biasing scheme to help identify events near the crystal surfaces [157]. This is important as such surface events are known to exhibit incomplete charge collection, distorting the measured signal. One particular danger is that an ER with a low charge measurement may be reconstructed as an NR, so a poorly-measured gamma ray could be mistaken for a WIMP. To mitigate this effect, on each detector face there is both an inner charge bias/readout electrode covering most of the surface and an outer charge bias/detector ring at high radius. Comparison of the measured signal amplitudes in the inner and outer channels allows for radial fiducialization in order to remove events near the cylindrical surface.

To identify events near the flat detector faces, a clever biasing scheme has been borrowed from the Edelweiss Experiment [158]. Each bias electrode is also interleaved with grounded electrodes with a spacing of around 1 mm. This maintains an electric field which is axial in the bulk of the cylindrical detector, but scalloped within a distance of ~1 mm of each flat detector face. With this design, charge carriers from events occurring in the detector bulk traverse the crystal almost as if there is a purely axial field and induce the expected ionization signal. Charges from surface events, however, are likely to drift to the nearby biased or ground electrodes, inducing signal almost entirely on a single side. Fig. 2.13 shows this bias field near a detector face as well as an example of the resulting charge (a)symmetry in measured data. The bias voltages used here are typically ±4 V.

Another feature of the iZIP device is that the grounded electrodes are actually
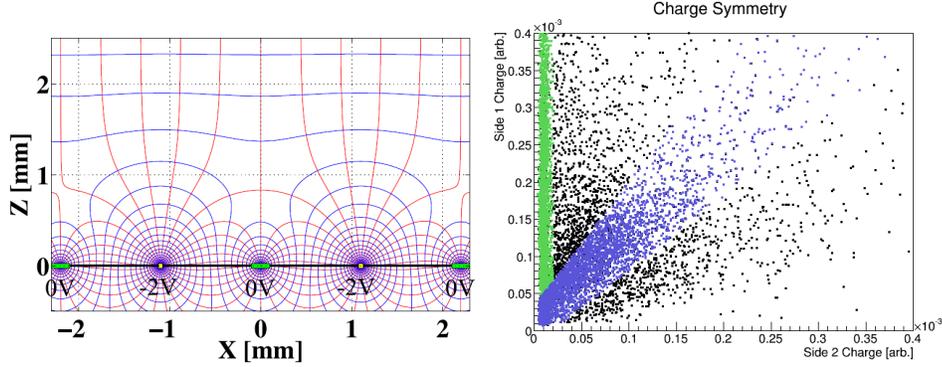
Figure 2.13: Left: Cross section of SuperCDMS iZIP detector showing electric field effects of interleaved electrodes. Taken from [159]. Right: Charge signal measured on opposite detector faces for a population of events. In green are near-surface events, which occur in the scalloped field region, inducing signal on only one detector face. Purple events are those occurring in the detector bulk which give a balanced signal on the two faces.

composed of QET phonon sensors as shown in Fig. 2.12. This allows a simultaneous measurement of the charge and phonon signals from individual particle events. Due to their different ionization yields, this can provide an event-by-event discrimination between ERs and NRs. An example of this is shown in Fig. 2.14. This difference can be exploited to select only NRs when searching for WIMP-nuclear recoils. It will also be useful at SNOLAB when studying the different types of background signals and recoil type discrimination. Unfortunately, below $\sim 10$ keV the charge amplifier noise limits the resolution of this device.

As shown in the left side of Fig. 2.15, the current generation of iZIP detectors feature 6 phonon and 2 charge channels on each face of the device. This type of detector was used for the position sensitivity studies which will be discussed in Sec. 5.2.

The HV detectors are composed of only phonon sensors and operated at higher voltage, typically $\sim 100$ V. By Eq. 2.9, this gives $\sim 30\times$ amplification of the NTL phonon energy. This dominates the primary phonons and effectively measures the number of charge pairs produced via the NTL energy. Because of this, the phonon and charge signals become degenerate and discrimination via yield is not possible (hence the lack of charge electrodes). However, the NTL amplification allows the detector energy threshold
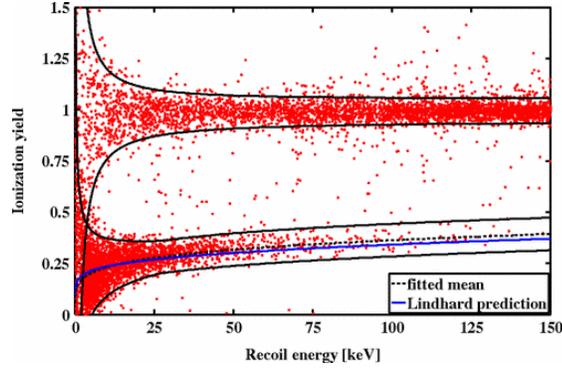
Figure 2.14: Ionization yield as a function of recoil energy from a calibration run of CDMS II. Taken from [160]. The upper band corresponds to ERs and the lower to NRs, which have a yield of $\sim 0.3$ in this energy range.
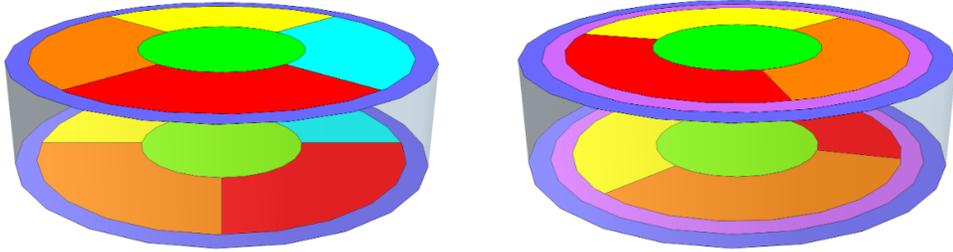


Figure 2.15: Left: iZIP phonon channel layout. The outer charge electrode is interleaved with the outer phonon channel while the inner charge electrode is interleaved with the remaining channels. Right: HV phonon channel layout. This device is composed of only phonon sensors.

to be greatly improved and the detector can become sensitive to lower energy interactions. This type of detector is the natural extension of the CDMSlite mode of detector operation developed at SuperCDMS Soudan [161]. The SNOLAB HV design features 12 phonon channels like the iZIP, but in a slightly different arrangement to optimize for measurement of radial position information. This is shown in the right side of Fig. 2.15.

Ideally, the detector resolution is limited by fundamental thermal noise effects in the QETs and readout electronics. In that case, one can simply increase the bias voltage to amplify the phonon signal and achieve arbitrary signal to noise ratio. Of course things are not that simple and additional noise sources begin to dominate after $\sim 20$ V. This is discussed in more detail in Ch. 4, but the net effect is that the optimal HV detector

performance has typically been achieved at $\sim 100$ V. This type of device was used for several of the studies highlighted in this thesis.

**Test Devices Small and Large**

Many different types of R&D detectors are tested in order to push forward various components of the technologies. Here we highlight two extremes, small phonon-only devices and large charge-only devices.

The HVeV line of detectors are a set of small, $\sim 1$ g, test devices which were designed to study improvements in QET sensitivity. Like the HV detectors, they have only phonon sensors and use the NTL effect to provide amplification. A photo of such a device in the "Animal" ADR (adiabatic demagnetization refrigerator) at Northwestern is shown in Fig. 2.16. This detector program has been very successful to the point of allowing single e/h pair sensitivity [162]. With this low energy threshold, these devices have been used to explore new parameter space for both dark photons and ERDM [47, 48] without the need for the strict backgrounds controls or large detector masses required of WIMP searches. With this sort of resolution, such devices have also become very valuable tools to study the nature of leakage currents observed at high voltages. Some ongoing studies of these devices at NEXUS will be discussed further in Ch. 8. As of this writing, there have only been a few generations of these devices produced in Si, but there are plans to fabricate similar designs on Ge as well.

At the other end of the spectrum, we have also demonstrated the largest ($> 1.3$ kg) cryogenic Si detectors yet operated [163]. These charge-only devices test the limits of current crystal growth technologies. They were also operated using a novel contact-free biasing scheme. This will be discussed further in Ch. 6.

### 2.2.5 DAQ and Data Processing

Signals from SuperCDMS detectors are controlled and read out using custom detector control and readout cards (DCRCs) [164]. The DCRCs interface with the cold electronics (SQUIDs, cryogenic FETs) and digitize the charge and phonon signals. Standard telnet protocols are used to communicate with DCRCs via ethernet in order to read trigger information and retrieve digitized pulse data. Communication and coordination of these boards is accomplished via a custom MIDAS-based DAQ system [165]. Because
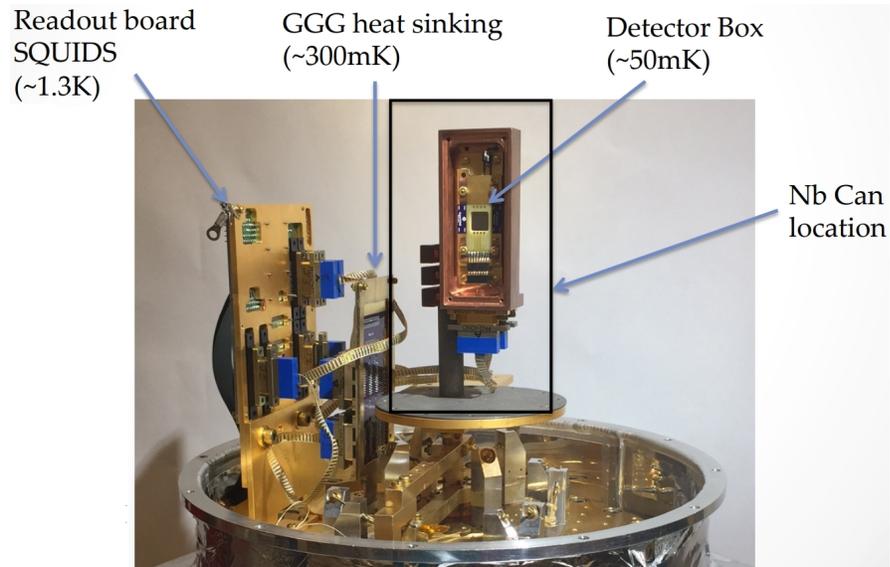
Figure 2.16: Photo of a 1 cm$^2$ HVeV detector mounted in the Northwestern ADR.

SuperCDMS detectors can produce huge amounts of raw data ($\sim$ Mb/s for an iZIP detector triggering at 10 Hz) a set of streamlined processing code names *cdmsbats*, has been developed to automatically extract and calculate meaningful quantities from each event. These quantities, such as integrals, amplitudes, and timing of charge and phonon pulses, can be used to determine the position and energy of recoil events.

Although they are not given much airtime in this document, a tremendous amount of work has been (and continues to be) devoted to development of the SuperCDMS DAQ and data processing algorithms. Each of these topics could expanded to a thesis in their own right. However, some will be discussed in more detail in later chapters when relevant.

# Chapter 3

# Phonon Sensors

## 3.1   TES Basics

The operation of a transition edge sensor (TES) relies on the steep variation of resistance as a function of temperature in the normal-superconducting transition region of a metal. This response allows the device to be sensitive to incoming heat. In SuperCDMS detectors, the heat signal is ultimately from phonons created by particle events in the detector (though the phonons themselves do not directly heat the TES as explained later). The superconducting critical temperature, $T_C$, of a metal also depends on the current going through it due to the magnetic field thus produced. Fig. 3.1 shows the resistance, $R(I,T)$, of a Mo/Au TES as a function of temperature and current near the critical phase transition. SuperCDMS generally uses W as the TES material, but the qualitative behaviors are the same. Below the critical temperature, $T_C$, the TES is superconducting and has 0 resistance while at higher temperatures it acts as a "normal" resistor and follows Ohm's law with resistance $R_N$. In between these two regions is a sharp transition where the resistance varies smoothly between 0 and $R_N$ with temperature. As can be seen, there is also a slight dependence on the current running through the device, but this is generally far less dramatic.

When operating a TES, it will be held near a quiescent point, $R_0 = R(I_0, T_0)$, generally midway through the transition. Then it is useful to describe the local topology
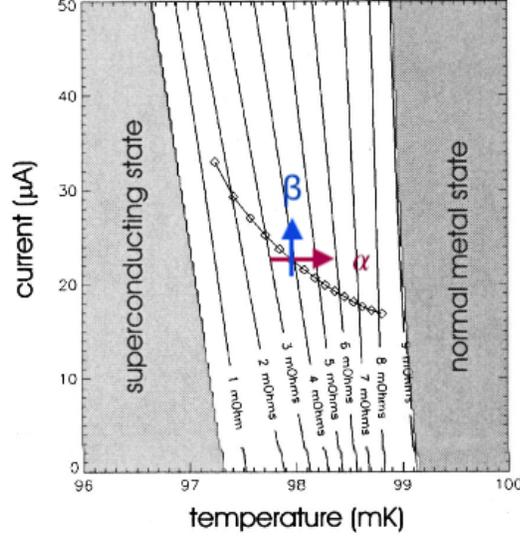
Figure 3.1: TES resistance as a function of current and temperature. The trajectory through this space when scanning the applied TES bias voltage is indicated. Also shown are the directions in which the $\alpha$ and $\beta$ parameters are defined. Image taken with permission from [166] with modification.

of this space using the logarithmic temperature and current sensitivities defined as

$$\alpha \equiv \left.\frac{\partial logR}{\partial logT}\right|_{I_0} = \left.\frac{T_0}{R_0}\frac{\partial R}{\partial T}\right|_{I_0} \tag{3.1}$$

and

$$\beta \equiv \left.\frac{\partial logR}{\partial logI}\right|_{T_0} = \left.\frac{I_0}{R_0}\frac{\partial R}{\partial I}\right|_{T_0} \tag{3.2}$$

Typical values for TES parameters are $\alpha \approx 100$ and $\beta \lesssim 3$ [167]. These allow us to expand the TES resistance near the quiescent point as,

$$R(I,T) \approx R_0 + \alpha\frac{R_0}{T_0}\delta T + \beta\frac{R_0}{I_0}\delta I \tag{3.3}$$

To operate such a device as a phonon detector, the TES temperature and current are not controlled directly. Rather the bias voltage across the TES is controlled and the resulting values of the temperature, current, and resistance depend on the rest of the electronic circuit, the TES's connection to a thermal bath, and any additional input power from phonon signals.

The TES is generally photo-lithographed onto a larger substrate which acts as a larger thermal bath. In SuperCDMS, the substrate is the large Si or Ge crystal. The device is operated by first cooling the bath lower than the TES $T_C$. Then, the TES can be warmed up to the transition region by passing a current through it and carefully adjusting the Joule heating. This is accomplished by connecting the TESs in parallel with a small shunt resistance, $R_{sh}$, as shown in the left side of Fig. 3.2. If $R_{sh} \ll R(I, T)$, the TES is then voltage biased with $V_{TES} \approx I_B R_{sh}$ and the power it dissipates will be $P_{Joule} = V_{TES}^2 / R(I, T)$. This causes the TES to remain stable to thermal fluctuations. Any increase(decrease) in temperature drives up(down) the resistance, which causes the Joule heating to decrease(increase). This effect is called Electro-thermal Feedback (ETF) and can be described by the power balancing equation

$$C\frac{dT}{dt} = P_{Joule} + P_{phonon} - \kappa(T^n - T_{bath}^n) \tag{3.4}$$

where $C$ is the TES heat capacity and $P_{phonon}$ any heating power from phonons entering the TES. The final term describes cooling via conduction to the crystal substrate. Here, the exponent $n$=4–6 is determined by the dominant thermal impedance between the crystal and TES [168]. Values of $n \approx 5$–6 are generally observed in SuperCDMS devices [142] and can be studied by observing the equilibrium value of $P_{Joule}$ while adjusting the crystal temperature.

With the temperature stabilized in ETF mode, the total power will be roughly constant. In its quiescent state, there are no incoming phonons and the TES draws enough Joule power to compensate heat loss. When the device is heated by phonons, it draws proportionally less current in order to stabilize. Thus, by monitoring the signal current flowing through the TES, $I_s$, the total energy deposited by phonons can be measured.

The right side of the TES circuit in Fig. 3.2 is solely dedicated to measuring $I_s$. This current is measured using an array of DC SQUIDs (Superconducting QUantum Interference Devices) [169]. The SQUID array acts as a sensitive magnetometer, responding with a voltage change proportional to the magnetic field induced by $I_s$ flowing through the input coil, $L_{In}$. An amplifier is operated in feedback mode to stabilize and linearize the SQUID amplifier response. Because the signal is flux coupled to the SQUID, a feedback resistor, $R_{FB}$ and coil, $L_{FB}$ are used. The amplifier will adjust $V_{out}$
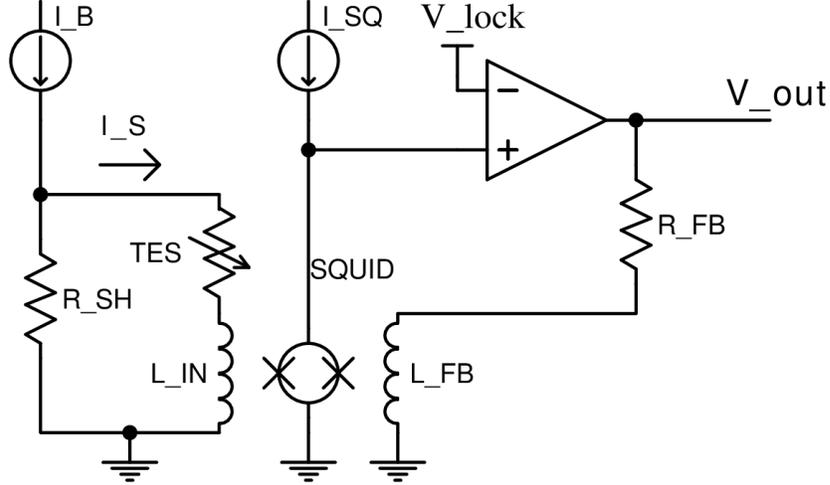
Figure 3.2: Circuit diagram of phonon measurement electronics. The phonon loop on the left is designed to hold the TES near its transition temperature. The right side of the circuit is a SQUID-based amplifier to monitor and amplify the signal current, $I_s$ passing through the input coil, $L_{In}$.

to compensate for changes in the magnetic flux through the SQUID. This results in the relation

$$V_{out} = I_s R_{FB} \frac{N_{IN}}{N_{FB}} \tag{3.5}$$

where $\frac{N_{IN}}{N_{FB}}$ is the ratio of turns in the input and feedback coils. In SuperCDMS electronics this ratio is typically 10–14 and $R_{FB} \sim 1 \ k\Omega$. Except for the current sources and amplifier, everything in this circuit diagram is maintained at cryogenic temperatures, between 10 mK and 1 K depending on the component. The output voltage from this first stage is further amplified and digitized by electronics outside the cryostat.

### 3.1.1 QETs

The TES sensitivity to input energy depends on its total heat capacity, $dQ/dT = C = cM$, where $M$ is the TES mass. A smaller device will respond with a larger temperature change for a given input power, so we want to make the TES mass as small as practical. However, in SuperCDMS we deploy these devices on the surface of a large Si or Ge crystal to measure the phonon energy present in the crystal and we want to absorb as

much of that energy as possible. One practical way to meet both demands is to add large ($\sim 10^4$ $\mu$m$^2$), flat absorbing fins to a small ($\sim 10^2$ $\mu$m$^2$) TES as shown in Fig. 3.3. In SuperCDMS devices, the TESs are made of W and the fins are Al. While Al is also a superconductor, it has $T_C \approx 1.2$ mK, and does not enter the transition region at the typical detector operating temperatures of $< 60$ mK.
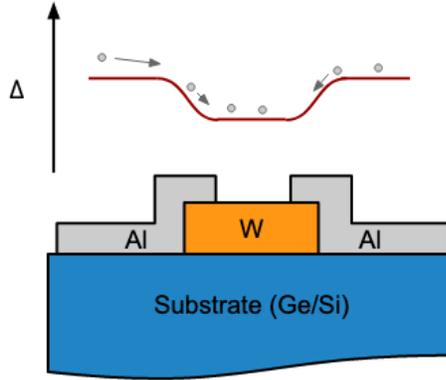


Figure 3.3: Diagram of QET consisting of a W TES with Al fins. Phonons in the substrate break Cooper pairs in the fins which then diffuse to the W and heat it.

This type of arrangement is called a Quasiparticle trapping assisted Electrothermal feedback Transition edge sensor (QET) [170]. Phonons from the detector substrate enter the superconducting Al and can break Cooper pairs if they deposit more than (twice) the superconducting energy gap $E \gtrsim 2\Delta(T = 0) \sim 3.5 k_B T_C$ [144]. In Al this is only 0.36 meV. Once a Cooper pair is broken, the quasiparticles will diffuse through the Al until they recombine or are trapped. Because the W has a much lower $T_C$ than Al, its superconducting band gap is lower. Because of this, the energy level decreases moving from the Al, through the overlap region, and into the W, forming a "quasiparticle trap". Once there, the quasiparticles will interact with the W electron system, raising its temperature. In this way, even though phonons are not absorbed in the TES directly, more of their energy is funneled to it. Now, the processes that transport this energy involve finite efficiencies. Only some of the phonons in the substrate will impinge upon the Al fins, some fraction of those will couple into the Al, and only some of those will have the energy to break a Cooper pair. Then the quasiparticles move diffusively

through the fin and only those that happen to reach the trap fall in. This last portion depends on details like the fin and overlap geometry and quasiparticle lifetime. There is very interesting physics happening in each of these steps, but detailed studies are beyond the scope of this thesis. Suffice it so say that much work has been done to tune the interface materials and QET geometry so as to maximize the total efficiency with which phonons in the crystal are converted to heat in the TESs [171].

## 3.2 $I_b - I_s$ curves

An important way to evaluate a TES circuit is to sweep the bias current, $I_b$, and evaluate the response of the signal current, $I_s$ (see Fig. 3.2). This is called taking an $I_b - I_s$ or an IV curve. The "IV curve" label comes from the fact that scanning $I_b$ is equivalent to scanning the voltage across the TES branch of the circuit. This procedure is done relatively slowly, so that at each set point, the circuit stabilizes. A diagram of such a curve is shown in Fig. 3.4. Its shape is explained below.
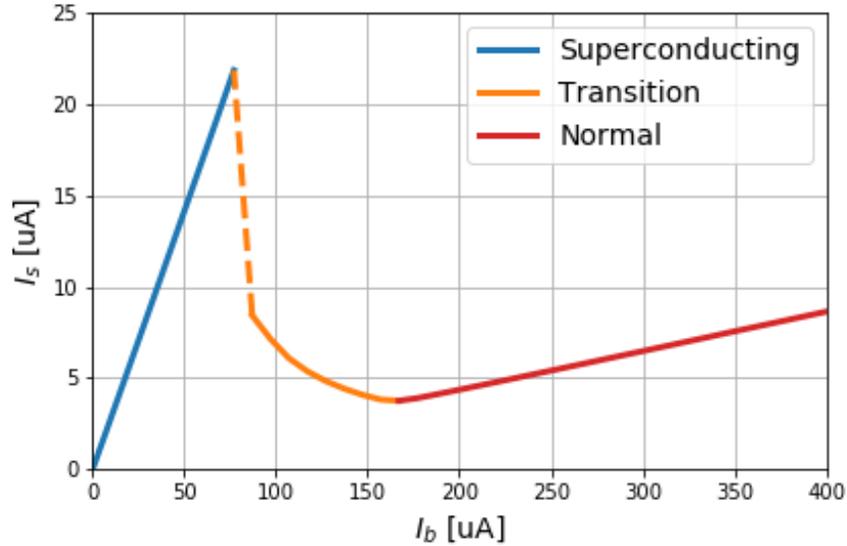


Figure 3.4: Measured $I_b-I_s$ curve for a TES. In the superconducting and normal regions, it behaves like a simple resistor, while the transition region is more complicated. The dashed line shows an unstable region of the curve where the TES will fall from transition into superconducting.

The TES circuit has the following DC relationship

$$I_b R_{sh} = I_s (R_L + R) \tag{3.6}$$

where $R = R(I_s, T)$ is the TES resistance and the load resistance, $R_L$, is the shunt resistance plus any parasitic resistances in the TES-shunt loop. At low values of $I_b$, the TES will remain superconducting (SC) and have 0 resistance. There, the circuit will follow

$$I_s(I_b)|_{SC} = I_b \frac{R_{sh}}{R_L}. \tag{3.7}$$

Likewise, at high $I_b$, the TES will generate enough Joule heat that it is driven normal (N). In that extreme, the circuit will also follow the simple relation

$$I_s(I_b)|_N = I_b \frac{R_{sh}}{R_L + R_N}. \tag{3.8}$$

In both of these cases, the $I_b - I_s$ curve will form a straight line with slope set by the resistance ratios. When the TES enters the transition region, as long as there are no phonon signals coming into the TES, it will adjust its resistance such that the Joule heating power will balance the bath cooling power.

$$I_s^2 R(I_s, T) = \kappa \left( T^n - T_{bath}^n \right) \tag{3.9}$$

We can use the DC circuit relation to rewrite this to make the dependence on $I_b$ more clear

$$(I_b R_{sh})^2 \frac{R(I_s, T)}{(R_L + R(I_s, T))^2} = \kappa \left( T^n - T_{bath}^n \right) \tag{3.10}$$

So for each value of $I_b$, the circuit will settle on some $T$ and $I_s$ that satisfies this relation. A simple simulation of such a scan is shown in Fig. 3.5. There, we have assumed the TES has a smooth transition region from superconducting to $R_N = 100$ m$\Omega$ and $\alpha = \infty$ so its resistance has no intrinsic dependence on $I_s$. The left panels show how the various parameters change throughout the scan. The right diagram indicates the trajectory through the resistance space of the TES. There are jumps in the temperature between the SC and transition region. This is also observed in real devices and is because the circuit is unstable when $R < R_L$.

When operating a TES in the lab, we generally only have access to the values of $I_b$ and $I_s$. From there we can determine the TES resistance and quiescent Joule power.
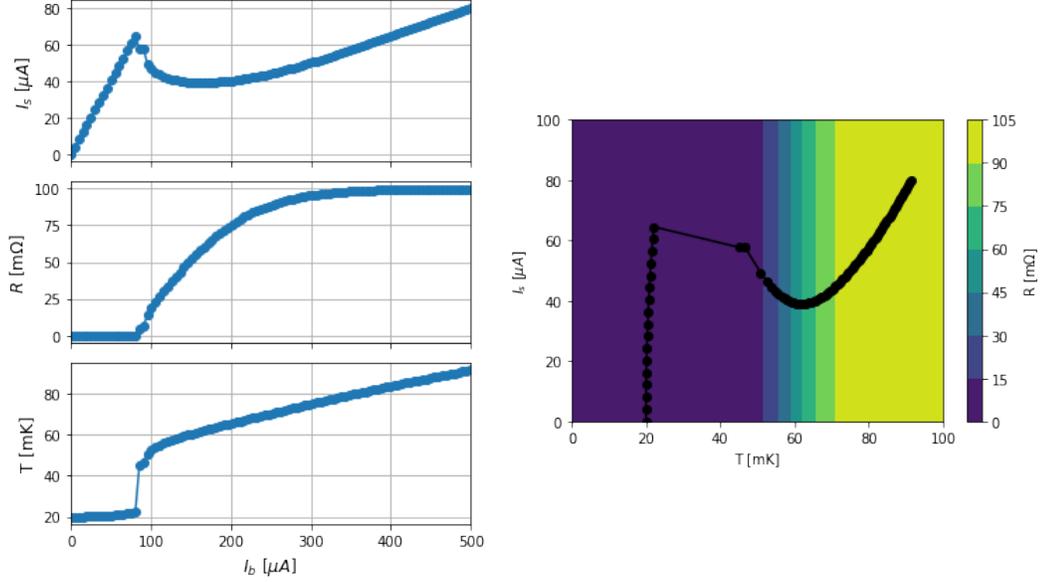
Figure 3.5: Simulated TES parameters during an $I_b - I_s$ scan for a TES with $\alpha = \infty$. Left: Corresponding $I_b - I_s$ curve and associated TES resistance. Also shown is the TES temperature, though this is not observed directly in practice. Right: Trajectory through the $R(I, T)$ space during the scan. There are points between the SC and transition regions where the circuit is unstable.

This information is used to select an operating point for the detector. Generally this is chosen to be near the bottom of the transition region, but not so low as to risk falling into SC. In practice, this is somewhere in the region of $R_0=0.2R_N$–$0.5R_N$.

## 3.3 Complex Impedance

We saw above that $I_b - I_s$ curves are useful in determining DC characteristics of the TES circuit, in particular for setting the TES operating resistance, $R_0$. But the frequency-dependent response is also important, as it contains information about parameters like $\alpha$ and $\beta$ as well as how the input coil inductance limits the sensor circuit bandwidth. The complex impedance of the TES circuit is used to characterize these quantities. This section shows a theoretical derivation for the complex impedance and (briefly) discusses methods used to measure it. Much of this section is an attempt to follow derivations in [154], and thus closely follows the treatment therein.

### 3.3.1 Derivation

The TES circuit can be represented by a Thevenin-equivalent circuit with bias voltage $V = I_{Bias}R_{sh}$ and load resistance $R_L = R_{sh} + R_{par}$ (which represents the total resistance of shunt and any parasitics in the TES-shunt loop), the input coil inductance $L$, and the TES, all in series. The TES dynamics are governed by two differential equations; one thermal and one electrical. The thermal equation (discussed earlier) is,

$$C\frac{dT}{dt} = P_{Phonon} + P_{Joule} - P_{Bath} \tag{3.11}$$

where $C$ is the heat capacity (to first order of TES and any absorber/fins, but this is really more complicated due to thermal impedances between them), $T$ is the TES temperature, $P_{Phonon}$ is any incoming phonon power, $P_{Joule}$ is heating power from TES bias current, and $P_{Bath}$ is cooling power to the thermal bath. The electrical equation is,

$$L\frac{dI_s}{dt} = V - I_s R_L - I_s R(I_s, T) \tag{3.12}$$

where $L$ is the input coil inductance, $I_s$ is the current in the TES, and $R$ is the TES resistance. In practice, $L$ includes feedback effects from the SQUID circuit which can effectively reduce the inductance [172]. Now, we can expand all of these parameters about their steady state values. The cooling power to the bath takes the form,

$$\begin{aligned} P_{Bath} &= K(T^n - T_{Bath}^n) \\ &\approx P_{Bath_0} + G\delta T \end{aligned} \tag{3.13}$$

where we have defined the differential thermal conductance parameter $G \equiv \frac{dP_{Bath}}{dT} = KnT^{n-1}$. Now we can also expand the joule heating power as,

$$\begin{aligned} P_{Joule} &= I^2 R_{TES} \\ &\approx P_{Joule_0} + 2I_0 R_0 \delta I + \alpha\frac{P_{Joule_0}}{T_0}\delta T + \beta\frac{P_{Joule_0}}{I_0}\delta I \end{aligned} \tag{3.14}$$

All of these expansions can then be substituted in to linearize the thermal and electrical differential equations. In the absence of electrothermal feedback (ETF), the thermal equation has a natural thermal time constant $\tau \equiv C/G$. This is the time scale of cooling from the bath. Under constant current conditions and with no input phonon

power, the thermal equation becomes

$$C\frac{d\delta T}{dt} = \alpha\frac{P_{Joule_0}}{T_0}\delta T - G\delta T \tag{3.15}$$

$$\Rightarrow \frac{d\delta T}{dt} = -\frac{(1-\mathscr{L})}{\tau}\delta T \tag{3.16}$$

where the low-frequency loop gain is defined as $\mathscr{L} \equiv \alpha P_{Joule_0}/GT_0$. This quantity relates to how sensitive the Joule power is to changes in the temperature at constant current. For a sensitive device, we desire a high value of $\mathscr{L}$. Integrating this equation gives the exponential decay of temperature excursions with the current-biased time constant $\tau_I = \tau/(1-\mathscr{L})$. Using these definitions and expansions, the differential equations become,

$$\frac{d\delta I}{dt} = \frac{\delta V}{L} - \frac{R_L + (\beta+1)R_0}{L}\delta I - \frac{\mathscr{L}G}{I_0 L}\delta T \tag{3.17}$$

$$\frac{d\delta T}{dt} = \frac{\delta P_{Phonon}}{C} + \frac{I_0 R_0(\beta+2)}{C}\delta I - \frac{\delta T}{\tau_I} \tag{3.18}$$

By considering the circuit response to a change in current while holding $\delta T = 0$ in Eq. 3.17, we can define a characteristic electrical time constant as

$$\tau_{el} = \frac{L}{R_L + (\beta+1)R_0}. \tag{3.19}$$

Now we can write things a little more compactly using these definitions and by putting the coupled differential equations together into a single matrix equation.

$$\frac{d}{dt}\begin{pmatrix}\delta I \\ \delta T\end{pmatrix} = -\begin{pmatrix}\frac{1}{\tau_{el}} & \frac{\mathscr{L}G}{I_0 L} \\ -\frac{I_0 R_0(\beta+2)}{C} & \frac{1}{\tau_I}\end{pmatrix}\begin{pmatrix}\delta I \\ \delta T\end{pmatrix} + \begin{pmatrix}\frac{\delta V}{L} \\ \frac{\delta P_{Phonon}}{C}\end{pmatrix} \tag{3.20}$$

To compute the complex impedance, $Z(\omega) = \frac{dV}{dI}$, we apply a signal to the bias current, which is in turn a Thevenin-equivalent voltage signal, $\delta V = \delta I_b R_{sh}$, and analyze the above equation for each Fourier component by using,

$$\delta X \rightarrow Re(\delta X e^{i\omega t}). \tag{3.21}$$

Then, we obtain the following relationship among the TES state variables,

$$M\begin{pmatrix}\delta I \\ \delta T\end{pmatrix} = \begin{pmatrix}\frac{\delta V}{L} \\ \frac{\delta P}{C}\end{pmatrix} \tag{3.22}$$

with

$$M = \begin{pmatrix} \frac{1}{\tau_{el}} + i\omega & \frac{\mathscr{L}G}{I_0 L} \\ -\frac{I_0 R_0 (\beta+2)}{C} & \frac{1}{\tau_I} + i\omega \end{pmatrix} \qquad (3.23)$$

These relationships will be useful both here and in later analysis of noise in the circuit. We will also find it useful to have the expression multiplied through by the inverse matrix, $M^{-1}$, whence

$$\begin{pmatrix} \delta I \\ \delta T \end{pmatrix} = M^{-1} \begin{pmatrix} \frac{\delta V}{L} \\ \frac{\delta P}{C} \end{pmatrix} \qquad (3.24)$$

with

$$M^{-1} = \frac{1}{(1/\tau_I + i\omega)(1/\tau_{el} + i\omega) + \frac{R_0}{L}\frac{\mathscr{L}(\beta+2)}{\tau}} \begin{pmatrix} \frac{1}{\tau_I} + i\omega & -\frac{\mathscr{L}G}{I_0 L} \\ \frac{I_0 R_0 (\beta+2)}{C} & \frac{1}{\tau_{el}} + i\omega \end{pmatrix} \qquad (3.25)$$

We can now use this to obtain an expression for the complex impedance. It is measured when there is no external signal power entering the TES from sources such as phonons, so $\delta P = 0$. We then simply read off the complex impedance as

$$Z = \frac{\delta V}{\delta I} = \frac{L}{M_{1,1}^{-1}} \qquad (3.26)$$

which, using the definitions of $\tau_{el}$ and $\tau_I$, becomes,

$$\boxed{Z(\omega) = R_L + i\omega L + R_0(\beta + 1) + \frac{R_0 \mathscr{L}}{1 - \mathscr{L}}\frac{\beta + 2}{1 + i\omega\tau_I}} \qquad (3.27)$$

We see that the first two terms of Eq. 3.27 are the impedance of the passive resistors and the inductor, so the TES impedance can be identified as the last two terms. This makes it clear that the overall circuit impedance depends on some important TES parameters including the operating resistance, $R_0$, the sensitivity of $R$ to changes in current, $\beta$, thermal properties $G$ and $C$ (which factor into $\tau_I$ and $\mathscr{L}$), and the sensitivity of $R$ to changes in temperature, $\alpha$ (a factor in $\mathscr{L}$).

We can begin to understand the shape of $Z$ by considering Fig. 3.6. There, we have plotted several curves of $Z(f)$ for different values of $\beta$ and $\mathscr{L}$. The first curve is for a near-ideal device with $\beta = 0$ and high $\mathscr{L}$. We observe that the low frequency impedance has a phase of $-\pi$, meaning the DC response of the circuit will be out of phase with a driving signal. This is an effective negative DC resistance and accounts for the change in slope sign in an $I_b - I_s$ curve. However, as the frequency increases, the amplitude
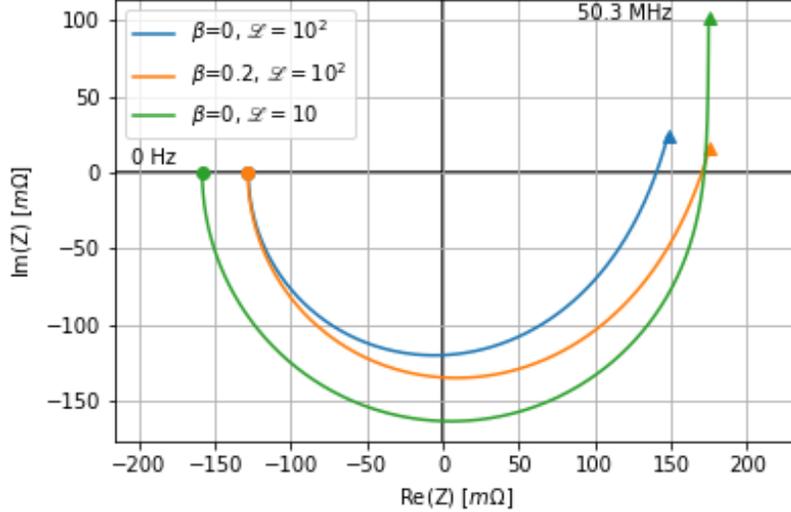
Figure 3.6: Complex impedance curves for several values of $\beta$ and $\mathscr{L}$.

does as well and the phase eventually comes back to 0. In the limit of infinite frequency, the phase approaches $\pi/2$ and the impedance becomes infinite. This behavior is driven by the presence of the inductor. This same general story holds when we relax the $\beta$ parameter, but the high frequency amplitudes are enhanced. By contrast, decreasing $\mathscr{L}$ (or, equivalently decreasing $\alpha$) enhances the low frequency impedance amplitude. The shape is, of course, dependent on other parameters as well, but this illustrates how intrinsic TES parameters can be extracted from measurement of complex impedance.

### 3.3.2 Measurement

As shown above, the complex impedance of the TES circuit contains information about various TES parameters. The goal of this section is to show how measurements can be used to extract those parameters. $Z$ can be be measured in various ways including by sweeping the frequency of a sine wave and recording amplitude and phase of response, injecting a square wave and fitting the time- or frequency-domain response, or injecting white noise and convolving the input and response [166]. In the sections below, we discuss only the sine and square wave methods which have been used often in SuperCDMS.

**Swept Sine Wave**

With the TES in transition with some DC bias current, $I_b$, a small sine wave perturbation is added, such that $I_b(t) = I_b + \delta I_b \sin(\omega t)$. This induces a a perturbation in the effective voltage of amplitude $\delta V = R_{sh} \delta I_b$. By measuring the phonon sensor output, one can then compute the complex admittance, $\delta I / \delta V(\omega) = 1/Z(\omega)$.
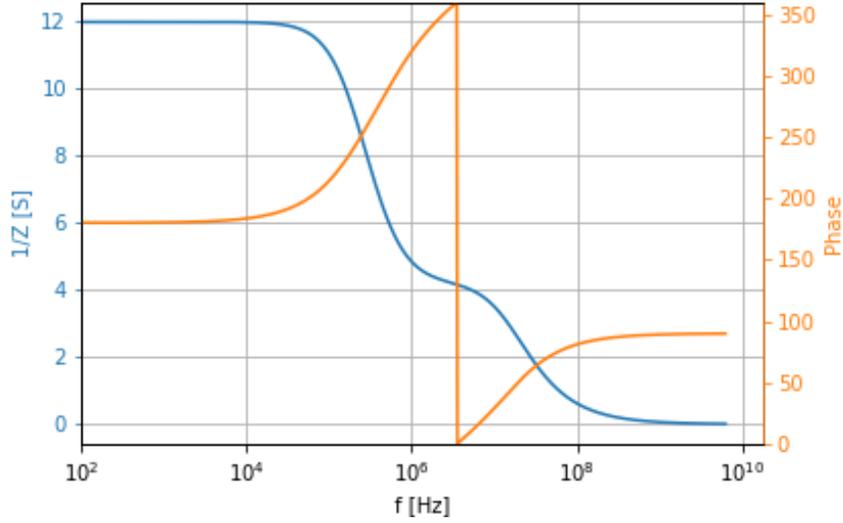


Figure 3.7: Shape of complex admittance magnitude and phase.

Fig. 3.7 shows an example plot of $1/Z(\omega)$ and demonstrates the key features. The position of the lower frequency shoulder is generally set by $\tau_I$ and the upper by $\tau_{el}$. At high frequencies, the complex impedance is dominated by the $i\omega L$ term, sending $1/Z \to 0$. Here, the inductor acts as an open circuit, shielding the TES branch from any current change. At intermediate frequencies, $1/\tau_I < \omega < 1/\tau_{el}$, $1/Z$ plateaus, giving

$$\frac{1}{Z}\bigg|_{1/\tau_I < \omega < 1/L} = \frac{1}{R_L + R_0(\beta + 1)} = \tau_{el}/L. \tag{3.28}$$

Here, we are below the inductor's cutoff, but the signal is too fast for the ETF to respond, so we just see the resistor-like terms. At frequencies below the current biased

time scale, $\tau_I$, the impedance tends to a constant value

$$\frac{1}{Z}\bigg|_{\omega=0} = \frac{1-\mathscr{L}}{R_L + R_0(\beta+1) + \mathscr{L}(R_0 - R_L)}. \tag{3.29}$$

In this regime, the ETF mechanism is operating, and a change in voltage results in a larger change in signal than it did without the feedback mechanism; i.e. the TES impedance is low here. What's more, for $\mathscr{L} > 1$, the impedance is negative. This is exactly what we saw in the $I_b - I_s$ curves earlier; when in transition, the IV curve of the circuit has a negative slope.

We can also understand how the phase changes with frequency. At low frequencies, the ETF is operating and the current is out of phase with the driving voltage. At frequencies higher than $1/\tau_I$, the TES branch basically operates like a resistor and the current is in phase with the driving voltage.

Early measurements of this type were undertaken at UMN using prototype detector control electronics which were limited to outputting triangle-wave test signals. Nevertheless, the measurements were somewhat successful despite the aliasing of multiple frequencies compared to a pure sine wave signal [173]. One such measurement is shown in Fig. 3.8. At frequencies above $\sim 20$ kHz, the output signal becomes small as the admittance falls and extraction of the phase becomes unreliable. However, the extracted fit parameters were all reasonable for the device under test. Higher precision measurements of this type are now being employed on devices at NEXUS.
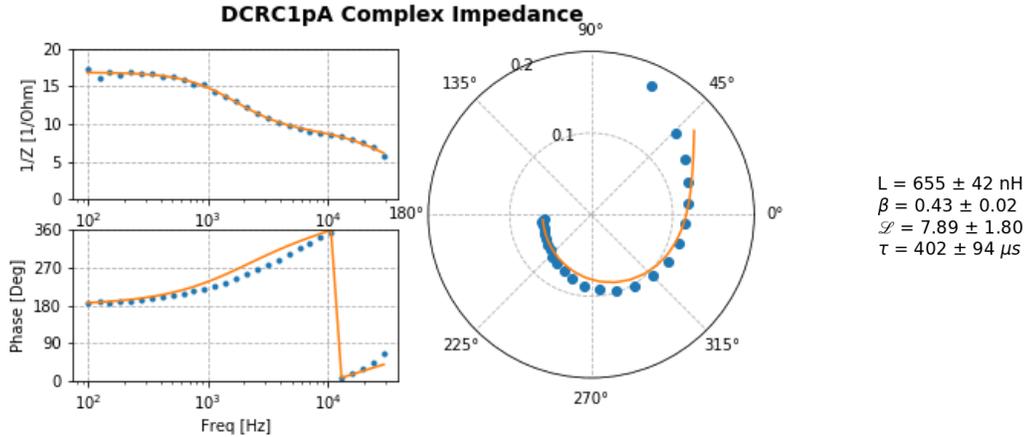


Figure 3.8: Complex impedance measurement at UMN using swept triangle wave.

**Square Wave**

The process described above can be tedious and time consuming since data must be taken separately at each frequency. Alternatively, one can probe all the harmonics at once by simply driving the circuit with a square wave signal. The Fourier coefficients of a driving square wave of frequency $\omega_D$ are the terms with all the odd multiples $\omega_D/2$,

$$\delta V(t) = \frac{2\delta V}{\pi} \sum_{n=1,3,5,...}^{\infty} \frac{1}{n} \sin\left(\frac{n\omega_D t}{2}\right) \tag{3.30}$$

Each mode will induce a response in the TES circuit as described by the shape of the complex impedance curve. In the time domain, we can expect that the early part of the response will be a small spike in amplitude in the same direction (in phase) as the driving square wave and having a time constant $\tau_{el}$. Then the signal will fall to a level prescribed by Eq. 3.29 and with time constant $\tau_I$. In practice it is easier to take the Fourier transform of this signal and calculate the impedance at each frequency as

$$Z(\omega = n\omega_D) = \frac{2}{n\pi} \frac{\delta V}{\delta I(\omega)}. \tag{3.31}$$

**Electrothermal Oscillation**

While we are looking at the details of complex impedance, we will briefly mention the issue of electrothermal oscillations. (For a more thorough treatment see [171] or [140]). This is an undesirable effect that can occur in TESs when the characteristic timescales, $\tau_I$ and $\tau_{el}$ are too close to each other. If these frequencies are within a factor of a few, the poles interact and cause the magnitude of the admittance to grow strongly. This effect can be seen in the left side of Fig. 3.9 where the admittance is plotted for a range of $\omega_{el}/\omega_I = \tau_I/\tau_{el}$ ratios. The plot on the right shows the impedance curves in the complex plane. There, the resonance occurs when the curve passes through the origin, where $|Z| = 0$.

The practical effect is that the circuit will respond with a huge change in signal current for small perturbations in voltage. When this effect is strong, the response is not quickly damped but sets up an oscillation. We will be able to see this more mathematically in the next section. One physical way to understand it is that the circuit is behaving like an RLC oscillator, with the TES heat capacity playing the role
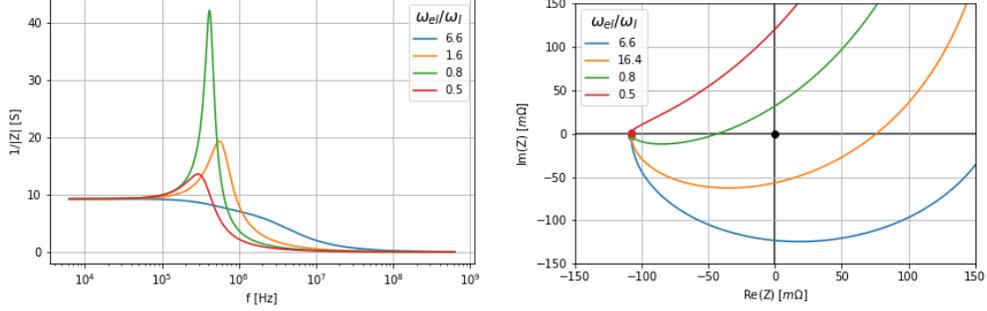
Figure 3.9: Resonance in complex impedance that results in electrothermal oscillation. Left: Complex admittance as a function of $\omega_{el}/\omega_I$ ratio. Right: Corresponding complex impedance curves.

of the capacitor. Basically, energy gets passed back and forth between the magnetic field in the input coil and the heat in the TES. The TES circuit and thermal parameters should ideally be designed to avoid such behavior as it can amplify noise sources as we will see later.

## 3.4 TES Energy Response

In this section, we will examine how the TES circuit responds to an input of phonon energy. In particular, we are interested in how that energy is related to the quantity $I_s$, which is our only measurement of the TES state when in operation.

### 3.4.1 Impulse Response

We start by considering the response of a TES to a delta function input of energy. As described in [154], the equations obtained in Eq. 3.20 can be solved by first calculating the eigenvectors, $\vec{v}_\pm$, and associated eigenvalues, $\lambda_\pm$. They take the following forms,

$$\vec{v}_\pm = \begin{pmatrix} \frac{1-\mathscr{L}-\lambda_\pm\tau}{2+\beta}\frac{G}{I_0 R_0} \\ 1 \end{pmatrix} \tag{3.32}$$

$$\lambda_\pm = \frac{1}{\tau_\pm} = \frac{1}{2\tau_{el}} + \frac{1}{2\tau_I} \pm \frac{1}{2}\sqrt{\left(\frac{1}{\tau_{el}} - \frac{1}{\tau_I}\right)^2 - 4\frac{R_0}{L}\frac{\mathscr{L}(\beta+2)}{\tau}}. \tag{3.33}$$

Now, for a delta function energy input, $E$, we set $\delta V = \delta P_{Phonon} = 0$ in Eq. 3.20 and then find a homogeneous solution of the form

$$\begin{pmatrix} \delta I_s \\ \delta T \end{pmatrix} = A_+ e^{-\lambda_+ t} \vec{v}_+ + A_- e^{-\lambda_- t} \vec{v}_-. \tag{3.34}$$

.

Then, using the initial conditions $\delta T(t = 0) = E/C$ and $\delta I_s(t = 0) = 0$, we can solve for the prefactors

$$A_\pm = \frac{E}{C} \frac{1 - \tau_\mp/\tau_I}{1 - \tau_\mp/\tau_\pm} \tag{3.35}$$

With this, we can write the measured response of the TES current to an impulse of energy as

$$\delta I_s(t) = \frac{E}{I_0 R_0 (\beta + 2)} \frac{(\lambda_- - 1/\tau_I)(\lambda_+ - 1/\tau_I)}{\lambda_+ - \lambda_-} \left( e^{-t/\tau_+} - e^{-t/\tau_-} \right) \tag{3.36}$$

This demonstrates the basic response of our phonon sensors; they should have a double exponential shape, with rise(fall) time set by $\tau_+(\tau_-)$. To understand this better, we can consider the case where the response controlled by the inductor, $\tau_{el}$ is much faster than the current-biased circuit response time, $\tau_I$. Then we see that

$$\tau_+ \rightarrow \tau_{el}$$
$$\tau_- \rightarrow \tau_{eff} = \tau \frac{1}{1 + \mathscr{L} \frac{R_0 - R_L}{(1+\beta) R_0 + R_L}}. \tag{3.37}$$

So the rise time of the pulse is set by the effective $L/R$ time constant of the system when in transition. This is sensible because we prescribed a delta function input and this is the fastest the circuit can respond. The fall time is scaled by the natural, thermal time $\tau$, but also gets faster for higher loop gain. We can also see the hint of a stability condition in the denominator of $\tau_-$. If the TES is biased too low in transition, such that $R_0 < R_L$, and if $\mathscr{L} > 1$, then the fall time changes sign, and the TES is unstable and will not return to equilibrium after an energy pulse.

Now, in practice, the phonon signals coming into the TESs are not quite a delta function, rather they bounce around in the detector crystal and can deposit power in

the TES over periods as long as a few ms. So we need to extract the total absorbed phonon energy for an arbitrary current pulse, not just the simple double exponential above. We derive expressions for this a few different ways in the coming sections.

The impulse response can also show us something about the electothermal oscillation case. The terms $\lambda_\pm$ are real-valued as long as the argument of the square root is positive. However, if the argument becomes negative, the falling exponential will obtain an oscillatory component as well. Depending on the strength of the resonance, this could manifest as simply a small oscillation on the falling edge of a pulse, or a strong, sustained oscillation. There are a variety of detector and circuit conditions that can cause the oscillation (which are evaluated in [171]), but a few general trends are that the oscillation can be made worse by moving the TES deeper into transition (since this generally increases $R_0 \mathscr{L}$) or better by using a higher base temperature (which decreases $\mathscr{L}$ at a given $R_0$). However, full understanding of this generally requires analysis of the particular circuit and TES parameters. Here we will simply follow the advice in [168] and say that one should ensure the electronics pole should be at least a factor of a few times higher than the characteristic TES response time.

### 3.4.2 Strong Feedback Case

Here, we treat the simple case when the TES is operating with strong feedback, by which we mean the circuit responds fast enough that the temperature does not vary appreciably. In this case, the power going to the bath will not change and the Joule heating power will exactly track any incoming phonon power. Likewise we will restrict ourselves to small perturbations so that the change in signal current is small.

When the TES absorbs some energy, $E$, it heats the TES, increasing its resistance which reduces $I_s$. The amount of absorbed energy is related to the change in TES current, $\delta I$. We will define $\delta I_s = I_0 - I_s > 0$ as a positive quantity since, in practice, we like to think of the $I_s$ signal as pulses instead of troughs. Now, in this case the amount of energy absorbed will be compensated for by a reduction in the Joule heating, thus

$$E_{abs} = \int V_{TES} \delta I dt \tag{3.38}$$

Using Kirchhoff's Voltage Law with the TES circuit gives

$$V_{TES} = I_b R_{sh} - (I_0 - \delta I_s) R_L - L\frac{d\delta I_s}{dt} \tag{3.39}$$

where $R_L = R_{sh} + R_p$ which allows us to write

$$
\begin{aligned}
E_{abs} &= (I_b R_{sh} - I_0 R_L) \int \delta I_s dt + R_L \int \delta I_s{}^2 dt \\
&= \left(1 - \frac{R_L}{R_L + R_0}\right) I_b R_{sh} \int \delta I_s dt + R_L \int \delta I_s{}^2 dt
\end{aligned}
\tag{3.40}
$$

where there is no net energy held in the inductor over time scales longer than $R_L/L$. This result indicates that the TES can be "self calibrating" in the sense that one can calculate the absorbed energy based solely on the circuit characteristics without need for detailed understanding of the TESs internal parameters (e.g. $\alpha$, $\beta$). In practice, however, higher order corrections can start to matter. These can either be measured and included in the model (as derived in the next subsection) or the device can be calibrated with known energy sources (e.g. lasers or radioactive sources).

### 3.4.3 Higher Order Corrections

Here follows a derivation of absorbed energy in a TES with higher order corrections. In particular we will account for parasitic resistances, the shape of the $R(I,T)$ surface near the bias point, and the variation in bath cooling power with temperature. Credit to M. Fritts and N. Kurinsky for previous work to include several of these effects [171].

For clarity, we will assume the shunt branch has a total resistance of $R_{sh} = R_{sh,0} + R_{sh,par}$ made up of the nominal shunt resistance and any parasitic resistance on that branch. The TES side has the TES resistance, $R_s = R(I,T)$, some parasitic resistance, $R_p$, and the inductor. The voltage across the two branches is always equal, so we have the DC relation

$$(I_B - I_s)R_{sh} = I_s(R_p + R_s) \tag{3.41}$$

which also holds at the quiescent point ($I_s = I_0, R_s = R_0$). Now, the bias power in the TES at any time is

$$P_s = I_s^2 R_s = I_s I_b R_{sh} - I_s^2(R_{sh} + R_p) \tag{3.42}$$

where in the final expression we have eliminated $R_s$ to make clear that we can measure the total bias power by monitoring $I_s$, since all other terms are constant. Now, we want

to measure the extra power that was absorbed when some phonon power comes into the TES. To do so, we want to compare the TES bias power during a pulse to the quiescent bias power. To 0th order, the quiescent bias power is just the TES bias power at the quiescent point, $P_0 = P_s(I_s = I_0, R_s = R_0) = I_s^2 R_0$. However, as the TES is heated, it moves in the transition, so we ought to also account for the fact that the quiescent bias power changes too. In other words, we need to account for the shape of $R(I, T)$. We will do this by Taylor expanding the bias power in transition about the quiescent point, $I_s = I_0$ as

$$
\begin{aligned}
P_s(I_s, T) = P_s|_{I_0} &+ \left.\frac{\partial P_s}{\partial I_s}\right|_{I_0, T_0} (I_s - I_0) + \left.\frac{\partial P_s}{\partial T}\right|_{I_0, T_0} (T - T_0) + \\
&+ \frac{1}{2}\left.\frac{\partial^2 P_s}{\partial I_s^2}\right|_{I_0, T_0} (I_s - I_0)^2 + \frac{1}{2}\left.\frac{\partial^2 P_s}{\partial T^2}\right|_{I_0, T_0} (T - T_0)^2 + \dots
\end{aligned}
\tag{3.43}
$$

We can evaluate the derivatives like

$$
\left.\frac{\partial P_s}{\partial I_s}\right|_{I_0, T_0} = \left.\frac{\partial (I_s^2 R_s)}{\partial I_s}\right|_{I_0, T_0} = 2I_0 R_0 + I_0^2 \left.\frac{\partial R_s}{\partial I_s}\right|_{I_0, T_0} = (2 + \beta) I_0 R_0
\tag{3.44}
$$

and similarly for the other terms. Doing so and using the definitions of $\delta I_s = I_0 - I_s$ and $\delta T = T - T_0$ allows to write the power in transition as

$$
P_s(I_s, T) \approx I_0^2 R_0 \left( 1 - (2 + \beta)\frac{\delta I_s}{I_0} + \alpha\frac{\delta T}{T_0} + \frac{1}{2}(2 + 4\beta + \beta_2)\left(\frac{\delta I_s}{I_0}\right)^2 + \frac{\alpha_2}{2}\left(\frac{\delta T}{T_0}\right)^2 \right)
\tag{3.45}
$$

where we have defined the unitless curvature terms as

$$
\begin{aligned}
\alpha_2 &= \frac{T_0^2}{R_0} \left.\frac{\partial^2 R_s}{\partial T^2}\right|_{I_0} \\
\beta_2 &= \frac{I_0^2}{R_0} \left.\frac{\partial^2 R_s}{\partial I^2}\right|_{T_0}
\end{aligned}
\tag{3.46}
$$

and have dropped any higher order contributions. This is the same result as obtained in Eq. 12 of [154], but now with the addition of the second order terms. Now, we are ready to connect this to the absorbed phonon energy. We know that the energy balance equation gives us

$$
\frac{dE}{dt} = P_s + P_{phonon} - P_{bath}
\tag{3.47}
$$

with $P_{bath} = \kappa \left(T^n - T_{bath}^n\right)$ and $n \approx 5$. At equilibrium there is no phonon signal power and the Joule heating is balanced by the bath cooling $P_0 = P_{bath,0}$. Now, when some phonon power comes in, we can integrate the above expression over the timescale of the pulse to obtain the total absorbed phonon energy

$$E_{abs} = \int P_{phonon} dt = \int \left(-P_s + P_{bath}\right) dt \qquad (3.48)$$

To get here, we use the fact that the sensor eventually returns to equilibrium, so the total change in energy is $\int dE = E(t = -\infty) - E(t = \infty) = 0$. Because we measure small changes in parameters from equilibrium, it is useful here to insert a factor of $P_0 - P_{bath,0} = 0$ to the above expression to obtain

$$E_{abs} = \int P_{phonon} dt = \int \left[\left(P_0 - P_s\right) - \left(P_{bath,0} - P_{bath}\right)\right] dt \qquad (3.49)$$

The first term can be obtained by inserting the expansion in Eq. 3.45. The second term can similarly be written in terms of an expansion in powers of $\delta T$ as follows.

$$\begin{aligned}
P_{bath,0} - P_{bath} &= \kappa(T_0^n - T_{bath}^n) - \kappa(T^n - T_{bath}^n) \\
&= \kappa(T_0^n - T^n) \\
&= -n\kappa T_0^{n-1} \delta T - \frac{n(n-1)}{2} \kappa T_0^{n-2} \delta T^2 + ... \\
&= -GT_0 \frac{\delta T}{T_0} - \frac{(n-1)GT_0}{2} \left(\frac{\delta T}{T_0}\right)^2
\end{aligned} \qquad (3.50)$$

Now, putting this all into the expression for absorbed energy and collecting like terms, we have

$$\begin{aligned}
E_{abs} = \int &\left[ I_0^2 R_0 (2+\beta) \frac{\delta I_s}{I_0} + (GT_0 - I_0^2 R_0 \alpha) \frac{\delta T}{T_0} \right. \\
&\left. - \frac{I_0^2 R_0}{2} (2 + 4\beta + \beta_2) \left(\frac{\delta I_s}{I_0}\right)^2 + \frac{1}{2} \left((n-1)GT_0 - \alpha_2 I_0^2 R_0\right) \left(\frac{\delta T}{T_0}\right)^2 \right] dt
\end{aligned} \qquad (3.51)$$

Now, when we operate the TES, we only have access to $\delta I_s$ but not $\delta T$. The best we can do here is defer to the local $R(I, T)$ topology to relate $\delta T$ to $\delta I_s$. We have

$$\delta T \approx \left. \frac{dT}{dR} \right|_{I_0} \delta R_s = \frac{T_0}{\alpha R_0} \delta R_s \qquad (3.52)$$

Then to see how this change in resistance will affect the signal current, we can expand the DC relation for small signals

$$R_s = \frac{I_b R_{sh}}{I_s} - R_L$$

$$\Rightarrow \delta R_s = \frac{I_b R_{sh}}{I_0^2} \delta I_s - \frac{2 I_b R_{sh}}{I_0^3} \delta I_s^2 + \dots \tag{3.53}$$

Putting these together we find

$$\delta T \approx \frac{T_0}{\alpha R_0} \left( \frac{I_b R_{sh}}{I_0^2} \delta I_s - \frac{I_b R_{sh}}{I_0^3} \delta I_s^2 \right)$$

$$\approx \frac{1}{G\mathscr{L}} \left( I_b R_{sh} \delta I_s - \frac{I_b R_{sh}}{I_0} \delta I_s^2 \right) \tag{3.54}$$

where the low-frequency loop gain is defined as

$$\mathscr{L} \equiv \frac{\alpha I_0^2 R_0}{G T_0}. \tag{3.55}$$

Now, inserting this relation into Eq. 3.51 and discarding higher order terms we obtain (after a page of algebra)

$$E_{abs} = \left[ 1 - 2\frac{R_L}{R_L + R_0} + \beta \frac{R_0}{R_L + R_0} + \frac{1}{\mathscr{L}} \right] I_B R_{sh} \int \delta I_s dt +$$

$$+ \left[ R_L \left( 1 - \frac{1}{\mathscr{L}} \right) - R_0 \left( \frac{1}{\mathscr{L}} + 2\beta + \frac{\beta_2}{2} \right) + \right. \tag{3.56}$$

$$\left. + \frac{(R_L + R_0)^2}{2 R_0} \left( \frac{n-1}{\alpha} - \frac{\alpha_2}{\alpha^2} \right) \right] \int \delta I_s^2 dt$$

Well, it took a bit of work, but we have now arrived at a result that gives a more detailed sense of how the absorbed energy depends on TES parameters. The first line of this expression is linear in the signal current. There we see some parts that are similar to the result in Eq. 3.40 and two new terms that involve $\beta$ and $\alpha$ (hiding in $\mathscr{L}$). The factor of two in the second term can be traced back to having now included the $\delta T$ term in the expansion of TES power in Eq. 3.43. Likewise the term involving $\beta$ comes from that same expansion in $\delta I_s$. The loop gain term, however arrived when we accounted for the change in bath power. For a good device, we will have large $\mathscr{L} \sim 100$ and small $\beta \lesssim 0.1$. But these terms may start to play a significant role otherwise. If we take the extreme limit of this, we obtain

$$\lim_{\alpha \to \infty, \beta \to 0} E_{abs} = \left( 1 - 2\frac{R_L}{R_L + R_0} \right) I_B R_{sh} \int \delta I_s dt + R_L \int \delta I_s^2 dt \tag{3.57}$$

This is in agreement with Eq. 3.40 except for the aforementioned factor of two which comes from including $\delta T$. This expression is what we commonly use to get a first estimate of the TES energy without doing a detailed analysis of all the parameters.

### 3.4.4 Empirical Calculation from $I_b - I_s$ Curve

The TES can be thought of as self calibrating, provided one knows the circuit parameters $R_L$, $R_0$, and $R_{sh}$. In practice, however, these values are $\sim$10–100 m$\Omega$ so the results can depend strongly on small parasitic resistances in the circuit wiring. Because most of the wiring in the circuit is only superconducting at cryogenic temperatures, it is actually somewhat difficult to measure them during detector operation, instead they must be inferred from the TES circuit response. However, it turns out that one can obtain a nice expression for $E_{abs}$ that incorporates this type of analysis. Credit goes to M. Fritts for first devising the results obtained here.

We start with the assumption that the absorbed phonon power is made up for by the reduction in TES power and any change in bath power as we argued earlier to get Eq. 3.49.

$$P_{abs} = P_0 - P_s - (P_{bath,0} - P_{bath}) \tag{3.58}$$

Now, instead of expanding $P_s$ in terms that are difficult to measure, we use the DC relation to write it in terms of $I_s$. This yields

$$P_s(I_s) = I_s^2 R_s = I_b R_{sh} I_s - R_L I_s^2 \tag{3.59}$$

and similarly $P_0 = P_s(I_0)$. Inserting these and doing a little re-arranging we find ourselves with some familiar terms

$$P_{abs} = \left( 1 - 2\frac{R_L}{R_L + R_0} \right) I_B R_{sh} \delta I_s + R_L \delta I_s - (P_{bath,0} - P_{bath}) \tag{3.60}$$

Now, we can treat the (local) variation in the bath power by realizing that this is exactly what an $I_b - I_s$ curve traces out. At each point in the scan we have $P_s(I_b) = P_{bath}$, and we record the resulting value of $I_s$. This gives us an empirical map of $P_{bath}(I_s) = P_s(I_s)$, which we can the use to calculate the change in bath power near the set point, $I_b = I_0$, as follows.

$$P_{bath} - P_{bath,0} = \delta P_{bath} \approx \delta P_s(I_s) \approx -\left. \frac{dP_s}{dI_s} \right|_{I_0} \delta I_s + \frac{1}{2} \left. \frac{d^2 P_s}{dI_s^2} \right|_{I_0} \delta I_s{}^2 \tag{3.61}$$

Note here that these are total derivatives, not the partials we used earlier, so the TES temperature is also varying. Now, if we evaluate the power slope term, we find

$$
\begin{aligned}
\frac{dP_s}{dI_s} &= 2I_s R_s + I_s^2 \frac{dR_s}{dI_s} \\
&= 2(I_b R_{sh} - I_s R_L) + I_s^2 \frac{dR_s}{dI_s} \\
&= I_b R_{sh} - 2I_s R_L + I_s R_{sh} \frac{dI_b}{dI_s}
\end{aligned}
\tag{3.62}
$$

Again, here these are total derivatives, so $\frac{dR_s}{dI_s}$ does not give terms of $\beta$, but rather is related to how the whole circuit responds. So in the third line we have evaluated the slope of $R_s$ by taking the derivative of Eq. 3.41 and solving for $\frac{dR_s}{dI_s}$. We also evaluate the 2nd derivative as

$$
\begin{aligned}
\frac{d^2 P_s}{dI_s^2} &= \frac{dI_b}{dI_s} R_{sh} - 2R_L + R_{sh} \frac{dI_b}{dI_s} + I_s R_{sh} \frac{d^2 I_b}{dI_s^2} \\
&= -2R_L + 2R_{sh} \frac{dI_b}{dI_s} + I_s R_{sh} \frac{d^2 I_b}{dI_s^2}
\end{aligned}
\tag{3.63}
$$

Evaluating these at $I_0$ and inserting back into the expression for $P_{abs}$ we find that many terms cancel and we are left with

$$
P_{abs}(\delta I_s) = -\left( \frac{I_s}{I_b} \frac{dI_b}{dI_s} \bigg|_{I_0} \right) I_b R_{sh} \delta I_s + \left[ \left( \frac{I_s}{I_b} \frac{dI_b}{dI_s} \bigg|_{I_0} \right) + \left( \frac{I_s^2}{2I_b} \frac{d^2 I_b}{dI_s^2} \bigg|_{I_0} \right) \right] \frac{I_b R_{sh} \delta I_s^2}{I_0}
\tag{3.64}
$$

Of course we are ultimately interested in the energy absorbed, so we just integrate Eq. 3.64 over the duration of the pulse to obtain

$$
\begin{aligned}
E_{abs} = &-\left( \frac{I_s}{I_b} \frac{dI_b}{dI_s} \bigg|_{I_0} \right) I_b R_{sh} \int \delta I_s dt + \\
&+ \left[ \left( \frac{I_s}{I_b} \frac{dI_b}{dI_s} \bigg|_{I_0} \right) + \left( \frac{I_s^2}{2I_b} \frac{d^2 I_b}{dI_s^2} \bigg|_{I_0} \right) \right] \frac{I_b R_{sh}}{I_0} \int \delta I_s^2 dt
\end{aligned}
\tag{3.65}
$$

This form is useful because the terms in brackets are unitless and invariant to the scale of $I_b$ or $I_s$. They are directly obtainable from just the *shape* of the TES $I_b - I_s$ curve. This means that any systematic uncertainties in the scales of various parameters only enter in the product of $I_b R_{sh} \delta I_s$. Of course none of this is immune to uncertainties in the offsets of these quantities. One more interesting feature is that this expression does

not contain $R_p$, the TES parasitic resistance. While $R_p$ will of course change the circuit behavior, the effects are incorporated in the shape of the measured $I_b - I_s$ curve, so the precise value need not be used here.

### 3.4.5 Bias in Integral of Squared Noisy Signal

While we often use specialized fitting algorithms to extract parameters from measured pulses (see OF section) we do sometimes wish to directly evaluate the integral of a measured trace. In that case, we must take the integrals of a noisy signal and its square. We will assume that the current signal has the form $\delta I(t) = \delta I_{true}(t) + n(t)$, where $\delta I_{true}(t)$ is the true current signal and $n(t)$ is a stationary noise term which we assume is uncorrelated with $I_{true}(t)$.

For the 1st order term, the integral of the noisy current signal will be an unbiased integral of the true current since

$$\langle \int \delta I dt \rangle = \int \delta I_{true}(t) dt + \int \langle n(t) \rangle dt = \int \delta I_{true}(t) dt. \tag{3.66}$$

However, the integral of the squared current acquires an offset bias from the $n^2(t)$ term.

$$\langle \int \delta I^2 dt \rangle = \int \delta I_{true}^2(t) dt + \int \langle 2 \delta I_{true}(t) n(t) \rangle dt + \int \langle n^2(t) \rangle dt = \int \delta I_{true}^2(t) dt + T \sigma^2$$
$$\tag{3.67}$$

where $T$ is the total trace length. This offset can be calculated and corrected for by estimating $\sigma$ from the pre-pulse baseline standard deviation. In practice, this correction is only of order 1% of the total absorbed energy.

# Chapter 4

# Phonon Signal Noise

There are several classes of noise present in the phonon signals; electronics noise, ultimately from resistor Johnson noise; thermal TES noise caused by thermal fluctuations between the TES and crystal; and power-coupled "noise" in the phonon system itself due to leakage currents in the crystal. This chapter describes how each of these sources contribute to the measured phonon signal and some studies carried out to understand and reduce their impact.

The phonon noise spectrum is determined by circuit and TES parameters. By fitting the noise distribution with the detector in various states, we can provide complementary information to e.g. complex impedance measurements about parameters which can be used to optimize the sensor signal to noise. In particular, the superconducting and normal states can tell us about the SQUID noise spectrum as well as values of $R_{sh}, R_p, R_N, L$ as well as the effective temperatures $T_{sh}, T_p$ of the shunt and parasitic resistances.

## 4.1 SQUID and Readout Noise

The readout noise includes everything from the SQUID to the warm electronics and ADC. In practice, this is dominated by the SQUID noise. The power spectral density (PSD) of the SQUID noise has been found to have the following form

$$S_{I,SQ}(\omega) = \frac{A_\phi^2}{\omega^\alpha} + I_{RO,\infty}^2 \tag{4.1}$$

76

where the first term is due to flux noise in the SQUID and the second term is related to Johnson Noise in shunt resistors within the SQUID chip [174]. The units of $A_\phi$ are such that $S_{SQ}$ will have the correct units of A$^2$/Hz.

This noise spectrum can be observed on phonon channels which do not have a TES installed (i.e. $R_{TES} = \infty$). It is not uncommon for this to be the case for some channels in a given experimental run, as the number of available SQUID channels often exceeds the number of actual readout channels on the device under test. Examples of such noise spectra from UMN and NEXUS, each fit with Eq. 4.1, are shown in Fig. 4.1. Note that these plots, and most of those shown in this chapter, are actually of $\sqrt{S_I(\omega)}$, with units of $A/\sqrt{Hz}$. The model was found to provide a reasonable fit to the noise below $\sim$50 kHz. At higher frequencies, the noise is generally dominated by a forest of discrete noise lines generated by the various components on the DCRC. In the UMN measurement shown, there is also a broad resonance related to the limited amplifier bandwidth. In any case, the phonon signal pulses come with characteristic rise times of $\sim$10 $\mu s$ and fall times of $\sim$1 ms, putting the frequency range of interest in 1-100 kHz, which is mostly free from these effects. (Methods such as Optimal Filtering which can further de-weight these noise contributions will be discussed later)
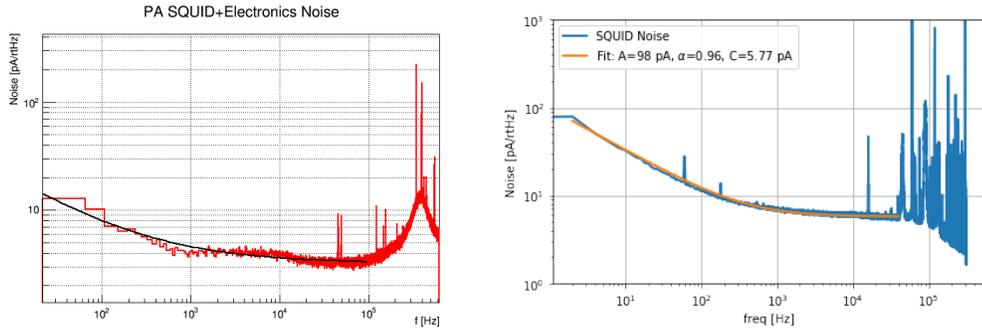


Figure 4.1: Example of phonon readout noise with fit to SQUID noise model measured at UMN (left) and NEXUS (right). The UMN measurement used CDMS II SQUIDs which meet the stringent requirement of <5 pA/rtHz above 600 Hz. The NEXUS measurement was done with a different SQUID design and resulted in somewhat higher noise.

## 4.2   Johnson Noise

Statistical fluctuations of the motion of electric charge in a conductor generates a voltage noise across its terminals with a voltage PSD given by

$$S_{V,JN} = 4k_B T R \tag{4.2}$$

where $R$ is the resistance and $T$ the temperature of the device. This was first observed by Johnson [175] and explained theoretically by Nyquist [176]. Note that this intrinsic voltage noise is "white", having no frequency dependence. Even though much of the TES circuit involves m$\Omega$ resistances at mK temperatures, this inescapable noise source can still be significant. The amount of this noise and how it couples to the measured signal current depends on the TES state as described in the following sections.

## 4.3   Superconducting and Normal Noise

When the TES is superconducting, the Thevenin-equivalent circuit is just the shunt and parasitic resistances, $R_{sh}$ and $R_p$, in series with the input coil, $L$. The current noise through in this circuit is just due to the Johnson noise on those "passsive" resistors (i.e. not the TES). The total passive Johnson voltage noise from the shunt and parasitic resistances is thus

$$S_{V,PJN} = 4k_B(R_{sh}T_{sh} + \sum R_p T_p). \tag{4.3}$$

Note that we can no longer simply lump $R_{sh}$ and $R_p$ together into $R_L$ as we did when dealing with the pure circuit response because they each come with their own temperature. The shunt resistor is a single component with single temperature, that of the fridge stage where it is mounted. The parasitic resistance is a little trickier since it is usually due to various wiring connections along the route from the SQUID (generally at the 1K stage) to the TES ($\sim$10 mK). Tracking down the locations and intermediate temperatures of all these contributions can be difficult, so often a single $R_p$ with an effective $T_p$ is used.

Now, this voltage noise is converted to current noise through the signal coil by the total circuit impedance. In SC mode, this is $Z_{SC} = R_{sh} + R_p + j\omega L$ and the current

noise is then given by

$$S_{PJN,SC}(\omega) = S_{V,PJN}/|Z_{SC}|^2 = \frac{4k_B(R_{sh}T_{sh} + \sum R_p T_p)}{(R_{sh} + R_p)^2 + \omega^2 L^2} \tag{4.4}$$

We can rearrange terms a little bit to write this as

$$S_{PJN,SC} = I^2_{PJN,SC,0}/(1 + (\omega/\omega_{RL,SC})^2) \tag{4.5}$$

This makes it clear that we get white noise at low frequencies with a single pole roll-off set by the total resistance and input coil inductance. Here we have defined $I_{PJN,SC,0}$ as the DC current noise in superconducting state and $\omega_{RL,SC}$ as the superconducting knee frequency given by

$$I_{PJN,SC,0} = \frac{\sqrt{4k_B(R_{sh}T_{sh} + \sum R_p T_p)}}{R_{sh} + R_p}$$
$$\omega_{RL,SC} = \frac{R_{sh} + R_p}{L} \tag{4.6}$$

Of course, the readout noise is always present independent of the TES state. The total observed noise is then the sum of these components

$$S_{SC}(\omega) = S_{RO}(\omega) + S_{PJN,SC}(\omega) \tag{4.7}$$

An example of fitting this to a SC channel in a Si HV detector in UMN R62 is shown on the left side of Fig. 4.2. Note that in this fit, there is an unexplained 1/f excess not described by the model. Though it has a similar shape to the SQUID noise, it is inconsistent with the SQUID noise magnitude measured on similar channels in the same experimental run.

The normal state noise is the same as the superconducting but with the addition of the TES normal resistance, $R_N$. This increases the total noise level and reduces the frequency of the impedance pole. These effects give a normal PJN noise of

$$S_{PJN,N} = (I^2_{PJN,N,0} + I^2_{TJN,N,0})/(1 + (\omega/\omega_{RL,N})^2) \tag{4.8}$$

where the passive Johnson noise DC level is now

$$I_{PJN,N,0} = \frac{\sqrt{4k_B(R_{sh}T_{sh} + \sum R_p T_p)}}{R_{sh} + R_p + R_N} \tag{4.9}$$
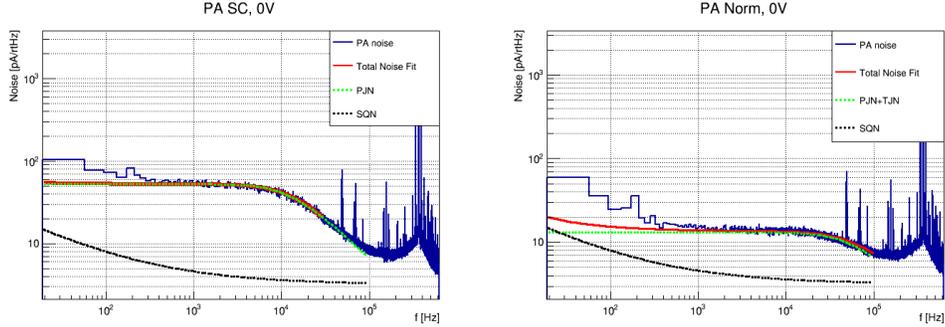
Figure 4.2: Examples of fitting the superconducting (normal) state noise for an HV detector. In the normal state (right) the PJN DC level is decreased and the pole frequency is increased compared to the SC noise.

the TES Johnson noise DC level is

$$I_{TJN,N,0} = \frac{\sqrt{4k_B R_N T}}{R_{sh} + R_p + R_N} \tag{4.10}$$

and the circuit impedance pole frequency is

$$\omega_{RL,N} = \frac{R_{sh} + R_p + R_N}{L} \tag{4.11}$$

We see that the normal state noise has the same functional form as the superconducting noise. However, the additional resistance tends to *decrease* the current noise and increase the pole frequency. This can be seen in the right side of Fig. 4.2.

## 4.4 Transition State Noise

When the TES is in transition, the noise analysis becomes a bit more intricate. The first reason is that the voltage fluctuations are no longer interacting with just a passive circuit, one must account for the feedback effects. There is also an additional noise component from thermal fluctuation noise between the TES and substrate which adds another frequency pole to the model. The analysis presented here follows some techniques used in [154, 140].

### 4.4.1 Passive Johnson Noise (PJN) in Transition

As found in Eq. 3.27, the circuit complex impedance is

$$Z(\omega) = R_L + i\omega L + R_0(\beta + 1) + \frac{R_0 \mathscr{L}}{1 - \mathscr{L}} \frac{\beta + 2}{1 + i\omega\tau_I}.$$

(4.12)

With this, the passive Johnson noise in transition becomes

$$S_{PJN} = \frac{S_{V,PJN}}{|Z|^2} = \frac{4k_B(R_{sh}T_{sh} + \sum R_p T_p)}{|Z|^2}$$

(4.13)

So this component has a shape set by the complex admittance and an overall scale set by the PJN voltage noise. This component is therefore subject to the same dangers of interacting poles ($\tau_I$ and $\tau_{el}$) that we saw in Sec. 3.3.2. Here, having the TES and electrical time constants too close will cause a large response to the small Johnson voltage noise. If the resonance is strong enough, this can become a dominant noise source.

### 4.4.2 TES Johnson Noise in Transition

The TES Johnson noise also features while in transition. We have to be a bit more careful here than with the passive resistances. The TES Johnson voltage doesn't just couple to the current via $Z$ because it's correlated with the TES power and the feedback mechanism. In other words, we can't simply read off the $\delta I/\delta V$ relationship from our earlier result

$$\begin{pmatrix} \delta I \\ \delta T \end{pmatrix} = M^{-1} \begin{pmatrix} \frac{\delta V}{L} \\ \frac{\delta P}{C} \end{pmatrix}$$

(4.14)

because now $\delta V$ across the TES is correlated with the $\delta P$ term.

Matt Pyle presented a nice argument to understand this in his thesis [140], and we use it here. Kirchoff's Voltage Law applied to the Thevenin equivalent TES circuit with a voltage noise $\delta V$ on the TES gives

$$V_B + \delta V = I_0 R_{sh} + L\frac{dI}{dt} + I_0 R_0$$

(4.15)

multiplying this expression through by another factor of $I_0$ yields an expression for the power in various parts of the circuit.

$$I_0 V_B = I_0^2 R_{sh} + \frac{d}{dt}\frac{LI^2}{2} + (I_0^2 R_0 - I_0\delta V)$$

(4.16)

On the left side we have the power supplied to the circuit, on the right there is the shunt Joule heating as well as the change in energy stored in the input coil. Finally, there is the steady-state TES Joule heating as well as the effect of TES voltage fluctuations. Here, we can see that a small TES voltage fluctuation is the same as a small (negative) power fluctuation. On average, $\delta V$ is 0, but any current that comes through the TES during a positive voltage fluctuation will come out the other side having higher than average potential energy. This is the amount of Joule heating power which was *not* lost in the TES, so this final term is the reduction in TES Joule heating power. In this way we can identify it with the power term, $\delta P = -I_0 \delta V$.

Inserting this result into Eq. 4.14, we obtain

$$\frac{\delta I}{\delta V} = \frac{M_{1,1}^{-1}}{L} - I_0 \frac{M_{1,2}^{-1}}{C} = \frac{1}{Z} + \frac{\mathscr{L}}{\tau L} \tag{4.17}$$

where the first term is just the complex admittance and the second is due to the power coupling between TES and inductance. This expression now tells us how an intrinsic fluctuation in the TES voltage couples to the signal current. In transition, the TES is at $T_0$, so the TES voltage Johnson noise is $S_{V,TJN} = 4k_B R_0 T_0$ and the observed current noise is

$$S_{TJN} = \left(\frac{\delta I}{\delta V}\right)^2 S_{V,TJN} \tag{4.18}$$

which works out to be

$$S_{TJN} = \frac{4k_B R_0 T_0}{(R_L + R_0(\beta+1))^2} \left(\frac{1 + i\omega\tau}{(1-\mathscr{L})(1+i\omega\tau_{el})(1+i\omega\tau_I) + \mathscr{L}(\beta+2)\frac{R_0\tau_{el}}{L}}\right)^2 \tag{4.19}$$

It should be noted that a more careful analysis using more advanced techniques [154] gives corrections of $\mathcal{O}(\beta)$, but that is beyond the scope of this work.

From Eq. 4.17, we see that this noise term is susceptible to the same resonances in $Z$ that we have seen elsewhere. The factor of $\mathscr{L}/\tau L$ increases the overall transfer function amplitude a bit, but does not introduce any new poles to the expression.

### 4.4.3    TES Thermal Fluctuation Noise

Now that the TES is in transition, it is sensitive to thermal fluctuations, which is exactly why we put it in there in the first place! However, doing so brings in another

noise source, that of thermal fluctuations across the thermal link between the TES and bath (crystal). In this section we again refer to previous works that have guided the following derivations [140, 177]. Generally, we will see how temperature fluctuations are related to power fluctuations in the TES, then calculate the magnitude of those thermal fluctuations in terms of intrinsic properties of the TES and use that to determine the noise spectrum observed in our output signal.

We start by considering the canonical ensemble of a body (TES) with heat capacity $C$ and temperature $T$ connected to a bath (crystal substrate) at temperature $T_{bath}$ by a thermal link having conductivity coefficient $G$. These quantities will be related to a random power fluctuation $\delta P(t)$ by the power balancing equation

$$C\frac{dT}{dt} = -G(T - T_{bath}) + \delta P(t) \tag{4.20}$$

We can solve this in the Fourier domain to obtain the linear frequency relation between power and thermal fluctuations in the TES

$$\delta T(\omega) = \frac{\delta P(\omega)}{G(1 + i\omega\tau)} \tag{4.21}$$

where $\tau = C/G$ is the usual natural thermal time constant for cooling of the TES. The power and temperature PSDs thus scale as

$$S_{\delta T}(\omega) = \frac{S_{\delta P}(\omega)}{G^2(1 + \omega^2\tau^2)} \tag{4.22}$$

At low frequencies, $\delta P \approx G\delta T$ as expected. But at frequencies above $\tau$, the temperature does not have time to respond to changes in power and $\delta T \approx 0$. Now, we expect the power fluctuations into the TES are from random thermal fluctuations and are uncorrelated at time scales of interest (i.e. $\mathcal{O}(\tau)$), or, explicitly

$$< \delta P(t)\delta P(t + t') > = \delta P_0^2 \; \delta(t') \tag{4.23}$$

We can use the Wiener-Khintchine relations to connect the auto-correlation function $C_x(t)$, of a quantity $x$ to the spectral density $S_x(\omega)$:

$$C_x(t) = < x(t')x(t' + t) > = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(\omega)cos(\omega t)d\omega \tag{4.24}$$

$$S_x(\omega) = \int_{-\infty}^{\infty} C_x(t)cos(\omega t)dt \tag{4.25}$$

This gives us $S_{\delta P} = \delta P_0^2$, i.e. a "white" spectrum of power fluctuations. We then relate the power fluctuations to intrinsic energy fluctuations in the system as

$$S_{\delta E}(\omega) = C^2 S_{\delta T}(\omega) = \tau^2 \frac{S_{\delta P}}{(1 + \omega^2 \tau^2)}. \tag{4.26}$$

Now, we can use a little thermodynamics to relate this to the fluctuations of energy in the system. Using the canonical partition function, we can calculate the variance of the energy in the TES as

$$
\begin{aligned}
\sigma_{\delta E}^2 &= \langle E^2 \rangle - \langle E \rangle^2 \\
&= -\frac{\partial}{\partial \beta} \left( \frac{\sum E_i e^{-\beta E_i}}{\sum e^{-\beta E_i}} \right) \\
&= -\frac{\partial}{\partial \beta} <E> \\
&= k_B T^2 \frac{\partial <E>}{\partial T} \\
\Rightarrow \sigma_{\delta E}^2 &= C k_B T^2
\end{aligned} \tag{4.27}
$$

The variance is just the integral over spectral density. So we can now integrate over Eq. 4.26 and solve for $S_{\delta P}$,

$$
\begin{aligned}
\sigma_{\delta E}^2 &= \int \frac{d\omega}{2\pi} S_{\delta E}(\omega) \\
C k_B T^2 &= \frac{\tau^2 S_{\delta P}}{2\pi} \oint \frac{d\omega}{(1 + \omega^2 \tau^2)} \\
\Rightarrow 2 G k_B T^2 &= S_{\delta P}
\end{aligned} \tag{4.28}
$$

where the integral can be solved via contour integration over either the upper or lower half-plane. There's one slight detail here that we have been able to gloss over before. When we are talking about spectral densities, we have always been implicitly dealing with one-sided PSDs. A full spectral density of some quantity, $S_X(\omega)$, is *really* defined over all positive and negative frequencies. But when dealing with fluctuations which are real-valued in the time domain (which is always the case for the physical quantities we are talking about), then $S_X(\omega) = S_X(-\omega)$ and people often work with the one-sided or folded spectral density, which is defined only for $\omega > 0$ and which is really

$$S_X(|\omega|) = S_X(\omega) + S_X(-\omega) = 2 S_X(\omega) \tag{4.29}$$

Confusingly, these are often represented by the same symbol, as we have done here. The upshot is that we have been secretly making sure this factor of 2 was present elsewhere, but now we need to explicitly account for it. Doing so gives us the result that the power noise spectrum due to thermal fluctuation noise is

$$S_{P,TFN} = 4Gk_BT^2F_{TFN} \qquad (4.30)$$

where we have (seemingly out of nowhere) added another factor, $F_{TFN}$. This unitless, $\mathcal{O}(1)$ factor is commonly added to correct for some hidden assumptions made in the above derivation. When exercising our thermodynamics skills, we were doing things that work under the assumption that the TES and bath were in thermal equilibrium. This is certainly not the case during normal TES operation in which the TES is kept at its transition temperature, above $T_{bath}$. $F_{TFN}$ is inserted to correct for this. Its theoretical value has been calculated in cases where the mean free path of energy carriers is both larger [178] and smaller than [179] the thermal link length. For our purposes, we will proceed in the ideal limit of $T_{bath} \to 0$, where $G \to nI_0^2R_0/T_0$ and $F_{TFN} \to 1/2$ [180]. For n=5 (electron-phonon coupling) this gives $S_{P,TFN} \to 10k_BT_0I_0^2R_0$. Finally, we want to couple this power noise to our measured value, the signal current. We again make use of Eq. 4.14 for the coupling term and obtain

$$
\begin{aligned}
S_{TFN} &= \left(\frac{dI}{dP}\right)^2 S_{P,TFN} \\
&= \left(\frac{M_{1,2}^{-1}}{C}\right)^2 10k_BT_0I_0^2R_0 \qquad (4.31) \\
&= \frac{10k_BT_0R_0}{(R_L + R_0(\beta+1))^2}\left(\frac{\mathscr{L}}{(1-\mathscr{L})(1+i\omega\tau_{el})(1+i\omega\tau_I) + \frac{R_0}{L\tau_{el}}\mathscr{L}(\beta+2)}\right)^2
\end{aligned}
$$

### 4.4.4 Total Transition State Noise

One may notice that the above sources of transition state noise all have similar functional forms. In particular, they share that same messy denominator that comes from the

factor of $|M|$ in all terms of $M^{-1}$. Here, we collect the compact forms of these terms.

$$S_{PJN} = \left(\frac{M_{1,1}^{-1}}{L}\right)^2 4k_B \sum T_i R_i = \frac{(1 + \omega^2\tau_I^2)4k_B \sum T_i R_i}{(|M|L\tau_I)^2} \qquad i \in \{shunt, parasitic\}$$

$$S_{TJN} = \left(\frac{M_{1,1}^{-1}}{L} - I_0\frac{M_{1,2}^{-1}}{C}\right)^2 4k_B T_0 R_0 = \left(\frac{1}{1 - \mathscr{L}}\right)^2 \frac{(1 + \omega^2\tau^2)4k_B T_0 R_0}{(|M|L\tau_I)^2}$$

$$S_{TFN} = \left(\frac{M_{1,2}^{-1}}{C}\right)^2 10k_B T_0 R_0 I_0^2 = \left(\frac{\mathscr{L}}{1 - \mathscr{L}}\right)^2 \frac{10k_B T_0 R_0}{(|M|L\tau_I)^2}$$

$$\tag{4.32}$$

and we can also explicitly state

$$|M|^2 = (\tau_I\tau_{el})^{-2} \left|(1 + i\omega\tau_I)(1 + i\omega\tau_{el}) + \frac{R_0\tau_{el}}{L}\left(\frac{\mathscr{L}}{1 - \mathscr{L}}\right)(\beta + 2)\right|^2 \tag{4.33}$$

The total noise in transition state is the sum of the spectral densities

$$S_{Tot} = S_{PJN} + S_{TJN} + S_{TFN} + S_{SQ} \tag{4.34}$$

where $S_{SQ}$ is the ever present SQUID readout noise. The shapes of these various noise contributions have been plotted for a model device in Fig. 4.3. The $|M|^2$ term in the denominator drives the high frequency roll-off of all terms and any resonance if the $\tau_{el}$ and $\tau_I$ poles interact. The TES Johnson noise also depends on the natural thermal time constant, $\tau = C/G$. This plot shows a model of a typical, well designed system where the thermal fluctuation noise is dominant in the frequency range of interest (generally 0.1–100 kHz). At higher or lower frequencies, the SQUID readout noise is the limiting factor.

Now, despite the amount of effort and interest that has been paid to the problem of understanding noise levels in TESs, there are actually still significant unexplained discrepancies observed between theory and experimentally measured noise [154]. An example of this can be seen in Fig. 4.4 which is taken from a recent dedicated study of noise in a sample of TES geometries [181]. While such excesses are still unexplained, several sources have been excluded including; more complex thermal geometries involving extra heat capacities and conductances (extra Cs and Gs), environmental noise from electronics, and particle interactions with the TES. Pointing out this failure of
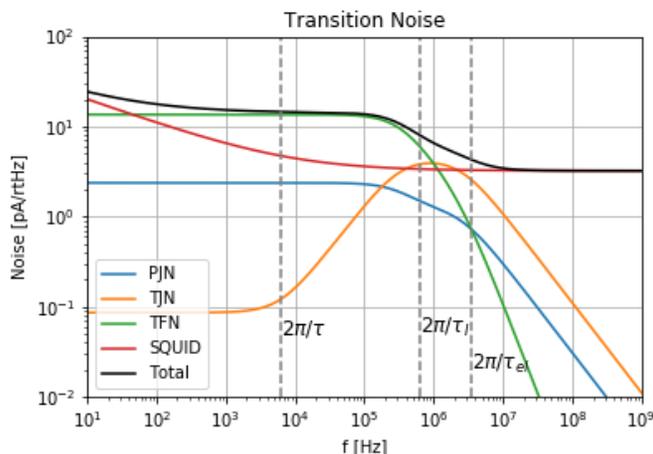
Figure 4.3: Model of total transition noise. In a well designed system, the thermal fluctuation noise is dominant in the signal frequency range, while readout electronics noise dominates at both lower and higher frequencies. Note that here we are plotting what's properly called the amplitude spectral density, $\sqrt{S_I}$, which has units of A/rtHz.

our understanding is not meant as a criticism of the models developed above, rather an indication that understanding fundamental TES noise is still an active and interesting area of research in the community!

## 4.5 PJN studies at UMN

Even though a complete understanding of TES noise sources is still a work in progress, we can still use the models of known sources to improve detector performance. In this section, we discuss the passive Johnson noise (PJN) in particular and show how it was improved at UMN.

### 4.5.1 Measuring PJN Contribution

Now, the models do provide good fits do data when the TES is SC or Normal, i.e. when there's no complicated feedback mechanism. As we saw above, then PJN is the only fundamental noise source. There is also, of course SQUID readout noise, but separate, dedicated measurements of this allow it to be accounted for. When the detector is in transition, the fundamental PJN voltage noise source is still present. The only difference
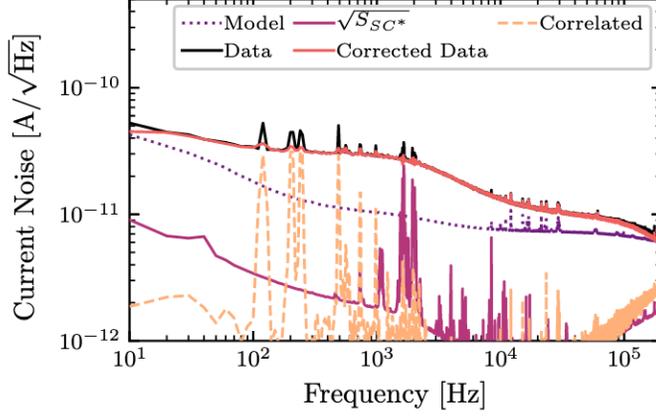
Figure 4.4: Comparison of model and measured TES noise. Measured noise data (black solid) is seen to exceed the noise model (purple dotted) even when correcting for correlated electronics noise (orange solid). Taken from [181].

is the circuit impedance through which it is converted to a current signal. This provides a way to scale the measured SC noise to estimate the PJN contribution when in transition. In particular, the DC noise levels are related as

$$S_{PJN,T}(\omega = 0) = S_{PJN,SC}(\omega = 0) \left( \frac{Z_{SC}(\omega = 0)}{Z_T(\omega = 0)} \right)^2. \tag{4.35}$$

where the $Z$ values can be obtained by analyzing complex impedance or $I_b - I_s$ characterization curves. It's informative to consider the ideal limit of large loop gain where we find

$$\left. \frac{Z_{SC}}{Z_T} \right|_{\omega=0, \mathscr{L}=\infty} = \frac{1}{1 - R_0/R_L} \tag{4.36}$$

So for the PJN to be suppressed in transition, we want $R_0 \gg R_L$. As described below, this turned out to be an issue in the setup at UMN.

In fridge run 62 at UMN (R62), we operated a SNOLAB HV Si prototype detector. The TESs on this device had $T_C$=65–80 mK and $R_N$=75–93 m$\Omega$. In addition, the SQUID chips that were used were of the same type used in SuperCDMS Soudan, which had 20 m$\Omega$ shunt resistors integrated onto the SQUID chips. With the benefit of hindsight, this is expected to give ratios of $|Z_{SC}/Z| \gtrsim 1$, causing a significant PJN contribution in transition.

To investigate this effect, noise data was acquired at several different points in transition. At each point, the DC impedance was obtained from the slope of the $I_b - I_s$ curve as indicated in Fig. 4.5. Each point was labeled by the quantity denoted $Ib \times Pre$, which is inversely proportional to the TES sensitivity to input energy. (It is just the first term of Eq. 3.57). $Z_{SC}$ can be measured from the slope of the superconducting portion of the $I_b - I_s$ curve.
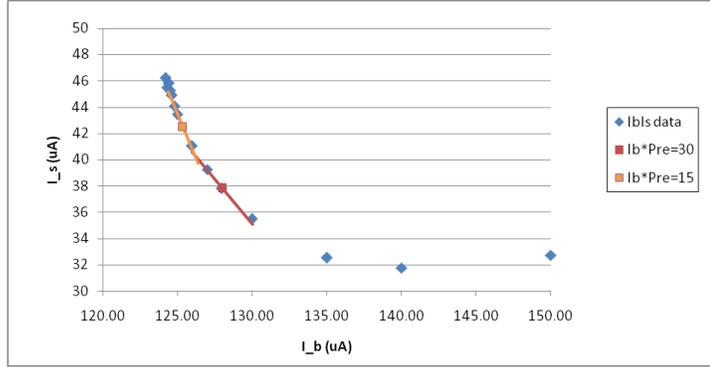


Figure 4.5: Example $I_b - I_s$ curve with several fitted slopes from UMN R62. A steep slope corresponds to a more sensitive region of the detector, but also couples PJN more strongly to the output signal.

The resulting impedance and ratios are shown in Table 4.1. Channel E was selected for this study, it had a measured SC impedance of $Z_{SC} = 31.1$ $m\Omega$. The transition impedance is effectively negative due to ETF. We see here that the steeper portion of the $I_b - I_s$ curve corresponds with a lower impedance. This means that, while the TES is more sensitive in these regions, the PJN will also be coupled more strongly to the output signal.

| Ib$\times$ Pre [$\mu A$] | Z [m$\Omega$] | $Z_{SC}/Z|_{DC}$ |
|---|---|---|
| 30 | -17.4 | $-1.79 \pm 0.27$ |
| 25 | -14.7 | $-2.12 \pm 0.41$ |
| 20 | -11.3 | $-2.75 \pm 0.39$ |
| 15 | -8.81 | $-3.54 \pm 0.61$ |

Table 4.1: Measured DC impedance ratios in R62.

The impedance ratios were then used to scale SC noise to the expected level in transition. These values were then inserted into a full noise model which was used to

fit the observed noise spectra. While there are other issues involved in decorrelating the relative weights of the other noise sources, this model could at least be used to estimate the contribution of PJN to the total noise. Fig. 4.6 shows these fits of the total transition state noise obtained by using the above scaling relation to fix the PJN contribution. The various other parameters were allowed to float within reasonable physical ranges in the fits. The total noise fit is shown in red, while the range of scaled PJN noise is shown in green. The width of the PJN range is dominated by uncertainties in $I_b - I_s$ slopes. The bias point, as indicated by the $Ib \times Pre$ value, is correlated with the inferred PJN contribution at different points in the transition region.
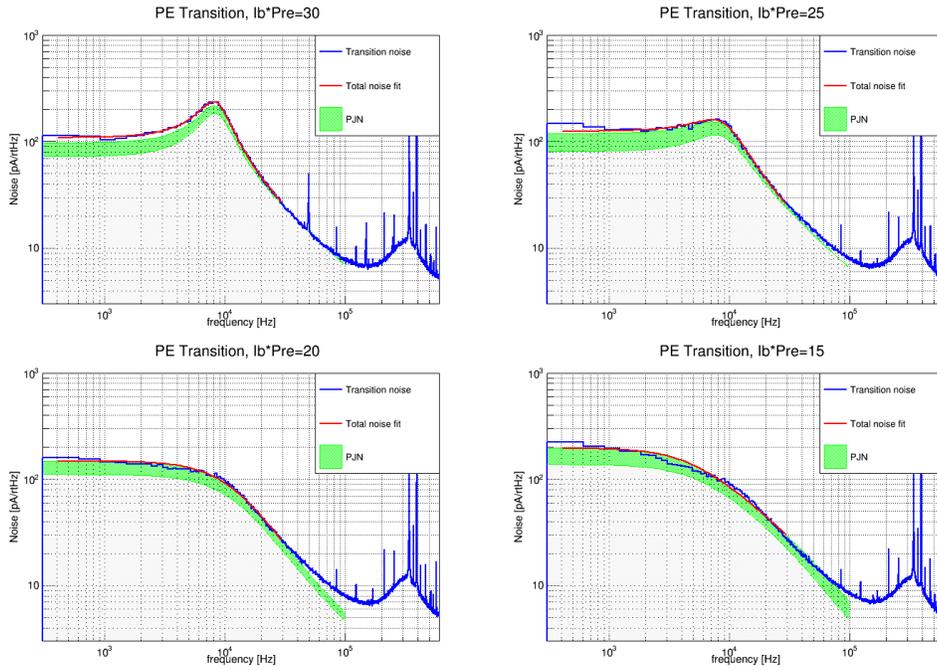


Figure 4.6: Fitted PJN contributions to R62 noise PSDs at OV.

In all cases, it is clear that the passive Johnson noise was the dominant noise source. Additionally, the DC noise level grew at points lower in the transition (lower $Ib \times Pre$) at a rate similar to that of the SC/Norm scaling ratios as would be expected for PJN. This showed that the electronics noise was dominated by the thermal noise in the shunt and/or parasitic resistances. Note that this was the dominant noise source when the detector is at low bias voltages, $\lesssim 10$ V. As described later, charge leakage effects in the

crystal eventually also contribute significantly to detector noise at high biases.

### 4.5.2 Improving PJN

As described above, the existing cold electronics chain were poorly matched to the Si HV detector's relatively low normal resistances, so the existing PJN voltage noise was not suppressed in transition. One way to address this was to reduce the Johnson noise by lowering $R_{sh}$. The nominal value used in SuperCDMS was 20 m$\Omega$, although production testing revealed these values could be as much as 40% higher. It was discovered that, although the shunt resistor was integrated onto the SQUID chips they actually had several available configurations of on-chip shunt resistances that could be changed with different wirebonding configurations. In particular, the 20 m$\Omega$ was achieved with a pair of 10 m$\Omega$ resistors in series. Adjusting the wirebond paths to put them in parallel allowed us to use a lower shunt resistance of 6.7 m$\Omega$.

The results of this on the SC noise can be seen in the noise PSD in Fig. 4.7 where the shunt resistance was decreased in R66. The data shown here is now from a Ge HV detector tested around the same time as the Si HV described above and which suffered similar PJN noise issues. As described in Eq. 4.6, the shunt resistance contributes to the Johnson voltage noise (numerator), so decreasing it will decrease the signal in transition. However, when observed in SC mode, the total circuit impedance is also decreased, so the SC noise increased, as expected. Additionally, the pole frequency was observed to decrease just as expected in R66 since $f_{RL} = (R_{sh} + R_p)/L$. The difference in noise above 100 kHz was due to warm readout electronics and is well above the signal range for phonon pulses.

The real target of this improvement, however, was the detector's baseline energy resolution, which was reduced by a factor of $\frac{275 \ eV_{ee}}{518 \ eV_{ee}} = 0.53$. This was actually slightly better than the theoretical improvement of $\sqrt{R_{sh,R66}/R_{sh,R61}} = 0.57$, which would be the case for $R_p = 0$. Variations from this are not unexpected due to factors like slight differences in fridge operating temperatures between the runs.
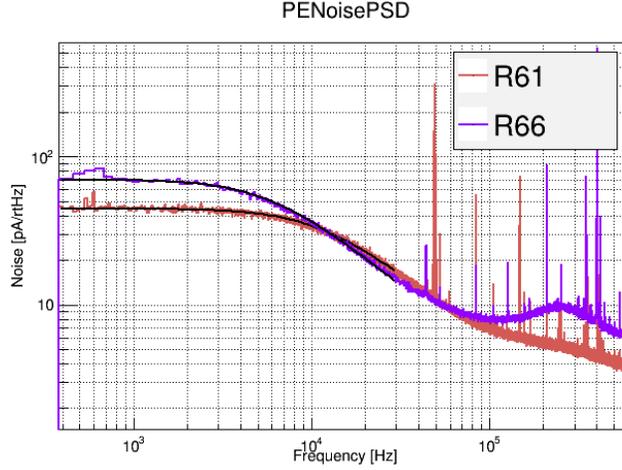
Figure 4.7: Effect of $R_{sh}$ reduction on SC noise. R61 used $R_{sh} = 20$ mΩ while R66 used $R_{sh} = 6.7$ mΩ. The SC noise level increases for lower $R_{sh}$ as expected but the noise level in transition will be decreased due to the decrease in PJN voltage noise.

### 4.5.3  PJN Noise Model Parameters

To fully understand this noise signal, we need a model for where the parasitic resistances are located and what their temperatures are. The most likely parasitic resistance culprits in this setup were the Al wirebonds and the Mill-Max pin-socket connections. The diagram in Fig. 4.8 shows these connections and the temperature stages at which they are located. The components that contribute to PJN noise are those located in the same loop as the TES and $R_{sh}$.

Note that the detector and SQUID temperatures are ideally below the Al super-conducting temperature of 1.2 K, so the wire bonds should have 0 resistance and not contribute to the Johnson noise, but this may not be the case if parts of the fridge are not as cold as designed. From this model, the dominant parasitic PJN contributor should be the Mill-Max pins connecting the SQUET card to the tower bottom, which is nominally at the LHe temperature of 4.2 K. Combining all of these gives expressions for the total parasitic resistance on each branch

$$R_{p,shunt} = R_{bond}$$
$$R_{p,TES} = 6 \times R_{pin} + 1 \times R_{bond}(@T_{SQUID}) + 2 \times R_{bond}(@T_{det})$$

$$(4.37)$$
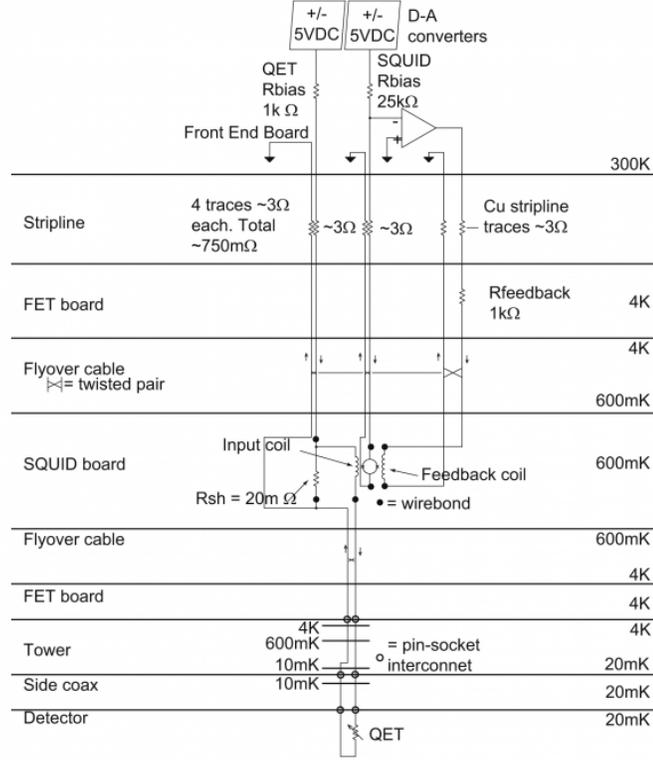
Figure 4.8: Diagram of the various components and thermal stages of the SuperCDMS phonon circuit. Taken from [182].

where we have assumed that all wirebonds have resistance, $R_{bond}$, unless they are cold enough to be SC, and each Mill-Max pin connection has resistance $R_{pin}$. Using these parameters and the assumed exact values of the shunt resistors, one can model the PJN using the above equation with the free parameters being $R_{bond}$, $R_{pin}$, and the fridge temperatures at the SQUET card and SQUID chip.

To fit the model, we used measured noise levels for a variety of $R_{sh}$ (including one channel with it purposefully shorted) and detector temperatures. This variety allowed some of the degeneracy in noise contributions to be broken. These data were fit to the above noise model with 4 free parameters: $R_{bond}$, $R_{pin}$, $T_{SQUID}$, and $T_{SQUET}$. The detector temperature was fixed at the measured values. The results of a best fit to the data points is shown in Table 4.3.

| $R_{sh}$ (mΩ) | $T_{det}$ (mK) | Channel | $I_{PJN,0,SC}$ (pA/rtHz) |
|---|---|---|---|
| 20 | 32 | S1PA,PB,PC,PE,PF | $45.0 \pm 0.9$ |
| 6.67 | 55 | S1PA-PF | $65.1 \pm 2.9$ |
| 6.67 | 22 | S1PE | $68.7 \pm 1.3$ |
| 0 | 22 | S2PB | $78.3 \pm 1.3$ |

Table 4.2: Measured average noise levels with various shunt resistances and tower temperatures.

| Parameter | Best Fit |
|---|---|
| $R_{bond}$ | 0.05 mΩ |
| $R_{pin}$ | 1.88 mΩ |
| $T_{SQUID}$ | 1.39 K |
| $T_{SQUET}$ | 3.97 K |

Table 4.3: Fitted parasitic resistance contributions and component temperatures.

This analysis indicated that the SQUID card actually runs hotter than the Al transition temperature, so the wire bonds there are not superconducting and do contribute slightly to the parasitic resistance. Without going into details, this was not entirely surprising based on how the SQUID card is heatsunk in the K100 fridge setup and is consistent with thermometer readings elsewhere in the fridge. These results indicate that the parasitic resistances an each branch were,

$$R_{p,shunt} = R_{bond} = 0.05 \ m\Omega$$
$$R_{p,TES} = 6 \times R_{pin} + 1 \times R_{bond} = 11.33 \ m\Omega$$

(4.38)

These values are also consistent with those obtained via analysis of $I_b - I_s$ curves, which supports the validity of the model. The takeaway here is that the dominant electronics noise source in the phonon circuit comes from the Passive Johnson noise of resistive components in the TES loop. The dominant culprit is the Mill-Max pin connection between the SQUET board and the bottom of the tower at 4 K. This design is not ideal from a noise perspective, part of the TES loop comes back to 4K between the SQUID and TES.

This wasn't an issue with earlier generations of detectors, but with the new, more

sensitive devices, it must be addressed. Unfortunately there are no clear easy fixes without significant redesign of the cold hardware and electronics. Such designs (currently underway) will incorporate this knowledge by removing such "warm" connections within the TES loop. Further, they will include heatsinking of the SQUID boards at colder temperature stages of the fridge to further minimize PJN contributions from the shunt resistance.

## 4.6   HV Noise

All of the noise contributions discussed above are due to the design of the phonon measurement and readout system. These sources are present in both iZIP and HV detectors and are the primary source of noise at low bias voltages. As the detector bias voltage is increased, these contributions remain unchanged, however a new source of noise begins to rear its head. This is especially a problem for HV detectors and is perhaps the largest unsolved technical detector issue moving into the SNOLAB era. This section describes efforts to understand and mitigate this effect at UMN.

### 4.6.1   Properties

The CDMSlite experiment [118] was the first to use high bias voltages to exploit the amplification of the NTL phonon signal. However, it was quickly discovered that phonon baseline noise increased dramatically when $\gtrsim 10$ V were applied to the detectors. The optimization between NTL gain and increased noise was found to be at 60–70 V, but varied depending on the particular detector and polarity. Since that time, the same type of noise has been observed in many SuperCDMS detectors to varying degrees.

An example of such measurements made with a Si HV device tested at UMN can be seen in Fig. 4.9. Here, the baseline phonon noise width has been characterized by applying the Optimal Filter algorithm to randomly triggered pulse-less traces. Also clear is the polarity asymmetry of this noise, which has been observed in other devices, both Si and Ge.

The manner in which this noise is distributed in frequency provides compelling evidence as to its origin. A clear example of this is shown in Fig. 4.10, which shows the phonon noise PSDs for a Si prototype detector operated at Berkeley. There we
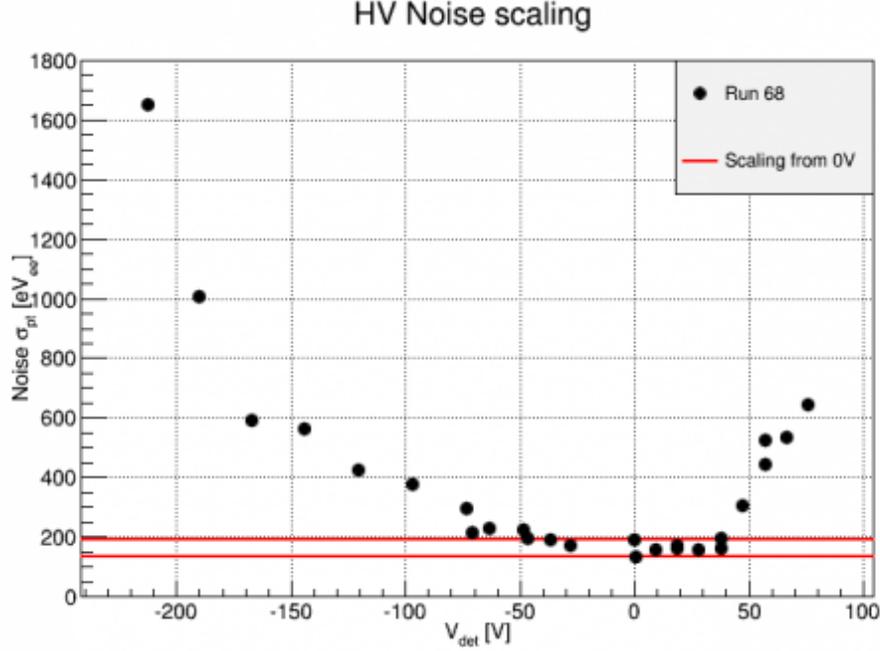
Figure 4.9: Phonon baseline noise as a function of applied bias voltage in a Si HV detector in UMN R68. The red lines show how electronics noise present at 0 V is expected to be constant. The two lines represent two different data sets and indicate the relative uncertainty observed in such measurements due to other stability factors.

see that the additional noise comes in at low frequencies and rolls off with a constant characteristic frequency.

This shape is well explained by a Poisson leakage of single charges across the crystal. They produce bursts of phonon energy which are so small and so frequently that the individual pulses are not observed but their aggregate spectrum is unmistakable. We can understand this shape theoretically as the result of "shot noise" which has been studied since the very early days of radiation counting experiments [183]. If we assume that single charges move across the crystal at uncorrelated times $t_1, t_2, ....$ An NTL phonon power is produced in the detector as

$$P(t) = eV(\delta(t_1) + \delta(t_2) + ...) \tag{4.39}$$

where $e$ is the electron charge and $V$ is the bias voltage. Calculating the PSD of this

signal over a trace length $T$ which contains $N$ such charge signals yields

$$\begin{aligned}
S_P &= |P(\omega)|^2 \\
&= \lim_{T\to\infty} \left| \frac{2}{\sqrt{T}} \int_0^T P(t)e^{-i\omega t} dt \right|^2 \\
&= \lim_{T\to\infty} \frac{2Ne^2V^2}{T} \\
&= 2eI_{leak}V^2
\end{aligned} \tag{4.40}$$

where in the second line we have used the truncated Fourier transform to consider a trace of finite length and the factor of 2 is because we want the 1-sided PSD. In the final line we have also defined an average leakage current of $I_{leak} = \lim_{T\to\infty} Ne/T$. So the phonon power noise from single charges drifting across the detector is expected to be white. But we need to propagate this signal into our TES and see how it manifests in the signal current. To do so, we approximate the TES impulse response to first order as an exponential with fall time $\tau_f$. In Sec. 3.4 we saw this is not a bad model with $\tau_f = \tau_-$. This can also be simply measured in larger energy pulses and is typically $\sim 1$ ms. In this case, the transfer function from phonon power to signal current is

$$\frac{dI}{dP} = f_{PCE} \left. \frac{dI}{dP} \right|_{\omega=0} \frac{1}{1 + i\omega\tau_f} \tag{4.41}$$

where $\left. \frac{dI}{dP} \right|_{\omega=0}$ is the DC TES response which depends on TES operating point. We have also included a necessary factor of phonon collection efficiency, $f_{PCE}$ which accounts for the fact that only a small fraction, $\sim 10\%$, of phonon energy in the crystal actually ends up in the TES system. With this, the leakage noise signal will manifest in the TES current as

$$S_I(\omega) = S_P \left| \frac{dI}{dP} \right|^2 = f_{PCE}^2 \left. \left| \frac{dI}{dP} \right|^2 \right|_{\omega=0} \frac{2eI_{leak}V^2}{1 + \omega^2\tau_f^2} \tag{4.42}$$

So we find that the HV noise is white with a high frequency roll off set by the pulse fall time. This is not surprising, as this noise produces the same signal as many tiny events that are constantly occurring. This also, unfortunately, means that the noise looks exactly like real events, so we can't hope to just filter it as we can with other noise sources.

A set of fits to this model are shown in Fig. 4.11 for a Ge HV detector operated in UMN R61. These show that the shot noise leakage model provides a good fit to the
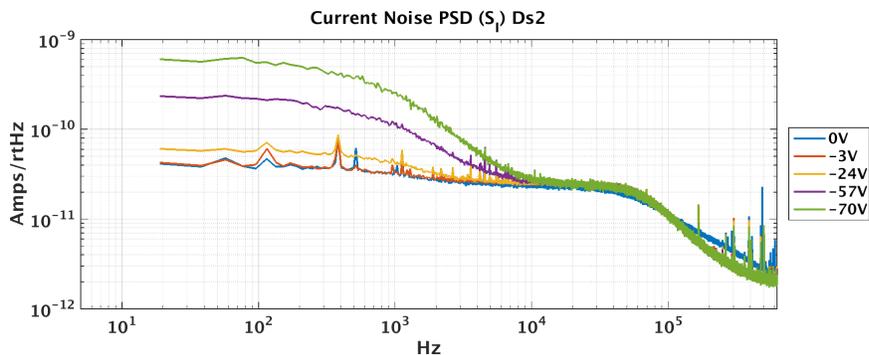
Figure 4.10: Phonon noise spectrum as function of bias voltage in detector S12C at Berkeley. The noise spectrum was observed to increase below 1 kHz at high bias voltages. Figure produced by Bill Page.

observed noise spectra, consistent with a flow of single carriers through the detector. There are still several open possibilities for the ultimate cause of this effect including leakage of carriers through the electrode-crystal interface or induced e/h pair creating in the crystal bulk.
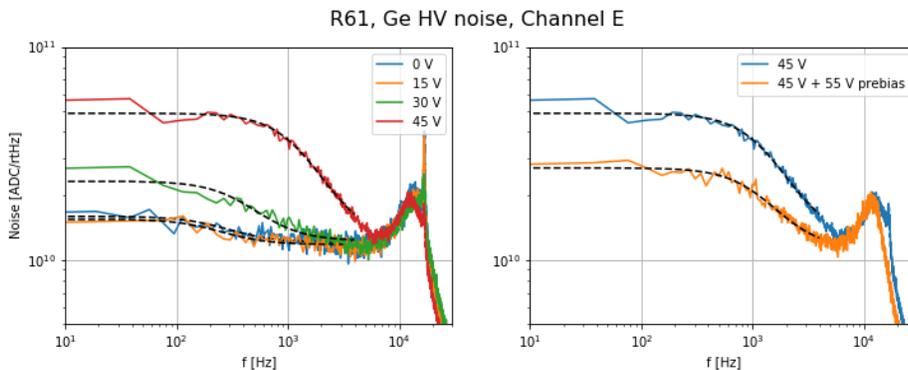


Figure 4.11: Phonon noise spectra as function of bias voltage in a Ge HV detector at UMN. Left: Noise is observed to increase with bias voltage. Black dashed lines indicate fits of excess noise to the shot noise model. Right: Prebiasing for several minutes has been shown to decrease the noise level.

The right side of Fig. 4.11 shows another interesting feature of this noise. That is the observation that the HV noise level can be reduced when "prebiasing" the detector by first applying a slightly higher voltage for around 5 min, then reducing the bias to the nominal level to take data. Some attempts have been made to describe this behavior as

the result of removing charges trapped in impurity sites in the crystal bulk [184]. The idea behind this is that such trapped charges would be only loosely bound at the nominal voltage and are occasionally coming loose to cause the noise. The argument then goes that prebiasing pulls these charges out, clearing shallow traps so that the remaining trapped charges are held tightly in deeper trapping potentials. Unfortunately models like this can not completely explain the remaining leakage levels or why the leakage usually remains relatively constant with time. Additionally, more recent observations carried out in HVeV devices have shown that much of this leakage is quantized, i.e. that the signals are from a single carrier moving through the full detector thickness, or an e/h pair produced in the bulk and both moving all the way to the electrodes. By contrast, the trap model would only result in a fractional signal since the freed electron would drift, but the trapping sites themselves are not mobile.

### 4.6.2 Contact-Free Electrodes

A common feature of the observed leakage has been an asymmetry with respect to bias voltage polarity. This hints at an asymmetry in leakage properties of different charge carriers. One theory is that carriers are being injected through the semiconductor/metal interfaces. If the leakage is occurring because carriers tunnel from the metallic electrode(s) and into the crystal, then removing contact between electrode and crystal surfaces would prevent such behavior. This has been tested with the use of biasing electrodes with a vacuum gap [185]. This study indicated that the dominant leakage is due to holes entering the crystal through an amorphous-Si blocking layer between the metallic electrode and semiconductor crystal. The use of such contact-free electrodes is a possible solution to the HV leakage problem, but it comes with several disadvantages.

The first is that carriers produced in the crystal in normal particle interaction events are usually drifted across the crystal and enter the electrodes where they are taken away by the power supply (really they meet an image charge provided by the supply). But, when such a vacuum electrode is used, the carriers moving to that side are not taken up, and so accumulate at the crystal face. As long as the detector remains biased, a surface charge builds up at this face and generates a counter-bias field, which lowers the effective electric field in the crystal bulk. This effect is seen and discussed in some of the work presented in Ch. 6. If the rate of accumulation is high enough, the counter

bias will cause the detector response to change with time as the NTL gain and charge collection efficiency are be reduced. This can either be accommodated by increasing the applied bias voltage over time to compensate and/or periodically grounding and resetting the detector. Another issue is that the contact-free face of the device can no longer be instrumented with phonon channels (though the biasing electrode can be used to measure charge signals). This reduces the total phonon collection efficiency of the device, hurting the low energy sensitivity.

### 4.6.3   Ongoing Studies

The precise origin of the HV leakage is still under investigation by CDMS and other groups [186]. Several avenues to understand and mitigate this effect are being worked on. One theory is that there are long-lived states in the crystal that are populated by high-energy events (such as cosmic rays) which then decay and release e/h pairs stochastically. This will be tested by comparing leakage rates in the same device operated above and below ground, such as at the NEXUS or CUTE locations. Similarly, there may be some long-lived fluorescing states in some material near the detectors (e.g. Cirlex clamps which hold the crystal or fiberglass PCB boards) which radiate low energy (sub-bandgap) photons at the crystal. These could cause a slow release of carriers from shallow impurity states. This possibility is discussed further in the context of HVeV studies at NEXUS in Ch. 8. The total amount of leakage is reduced in these small devices to the point that individual e/h pairs can be resolved. This allows further insight into the nature and origins of this mysterious noise source.

There are also ongoing efforts to counter this effect via; the use of contact-free electrodes with large HV detectors, the effect of initial HV over-biasing and other operational techniques on leakage noise, and alternative blocking layer recipes to replace the amorphous Si used between the metallic electrodes and crystal. With such concentrated work, we are optimistic that the HV leakage issues will soon be understood and overcome. In principle, this can allow single e/h resolution even in our large ($\sim$1 kg) detectors and unprecedented sensitivity to DM interactions.

# Chapter 5

# Event Position Reconstruction

It is important to properly reconstruct the location at which an event occurs in the detector. One reason is that, in general, the detector energy response has some dependence on event location. This is because of both surface effects which change the amounts of charge and phonons produced and electric field nonuniformities which affect the phonon energy produced. Another reason is that the event location can be used to veto undesirable background signals. In particular, low energy photons can only penetrate into the outer few mm of the crystal, so a well-defined fiducial volume can exclude a large fraction of the background signals.

It is thus essential to understand how to reconstruct not just the energy, but the position, $(r, \theta, z)$, of individual events based on the measured signals. This chapter describes several ways this has been studied. The first section discusses exploration of HV detectors where we used known event position distributions to infer the mapping from physical location to measured signals. The second section describes the development and testing of a novel cryogenic motorized system to accurately move a calibration source for position calibration.

## 5.1 HV Detectors

HV detectors will form a significant part of the SNOLAB detector payload. UMN was the first to operate prototypes of the SNOLAB HV detectors in the fall of 2016. At SNOLAB, the HV detectors will be symmetrically biased and all 12 channels will be read

out but this requires dedicated cold electronics that were not available at the time. This left us with 2 readout modes; 0 V with all 12 channels available for read out, or 1-sided HV with only the 6 channels on the unbiased face operating. The studies presented below were carried out with 1-sided HV biasing and thus only make use of phonon signals from one side of the device. This arrangement is the same as that employed in CDMSlite. Regardless, we can still make considerable progress in this way. All the insights gained will be applicable for a detector with full readout. We expect nominal accuracy gains for the radial and angular position reconstruction by averaging such measurements from both sides. As described below, accurate $z$-position measurements will certainly require full readout as well.

One practical difficulty was that all the detector signal and bias wiring was fed out through three 50-pin connectors, with lines from both sides of the detector present in each connector. A printed circuit board was designed and fabricated to act as an adapter between the 3 detector cables, DCRCs, and an HV power supply, breaking out the 150 lines to DCRCs and the HV bias network. More details of this work can be found in Appendix C.

The layout of the SNOLAB HV phonon channels is shown in Fig. 5.1. Each diagram represents the arrangement of channels on one face of the detector. The orange letter labels in parentheses indicate the channel names used in this work. To distinguish between sides, we will use a naming convention such that e.g. PAS1 indicates Phonon channel A on Side 1. This device only has phonon channels, so the "P" is not so informative, but the same system is used on iZIP devices where there are also charge, "Q", channels. Also shown in the figure are the positions of detector interface boards (DIBs). These are small circuit boards which connect the on-detector sensors to external electronics via wirebonds.

The particular design choices for the channel geometry are beyond the scope of this work, but the general goal was to maintain sensitivity to both radial and angular position information. Such detailed design drivers can be found in [171]. The rough idea was for the radial information to be captured by the relative phonon signals (both arrival time and energy) observed in PA and PB compared to PF. The positions of the DIBs and crystal flats in this diagram show that the inner ring channels (PC, PD, PE) on the two faces are offset from each other so as to provide complementary angular
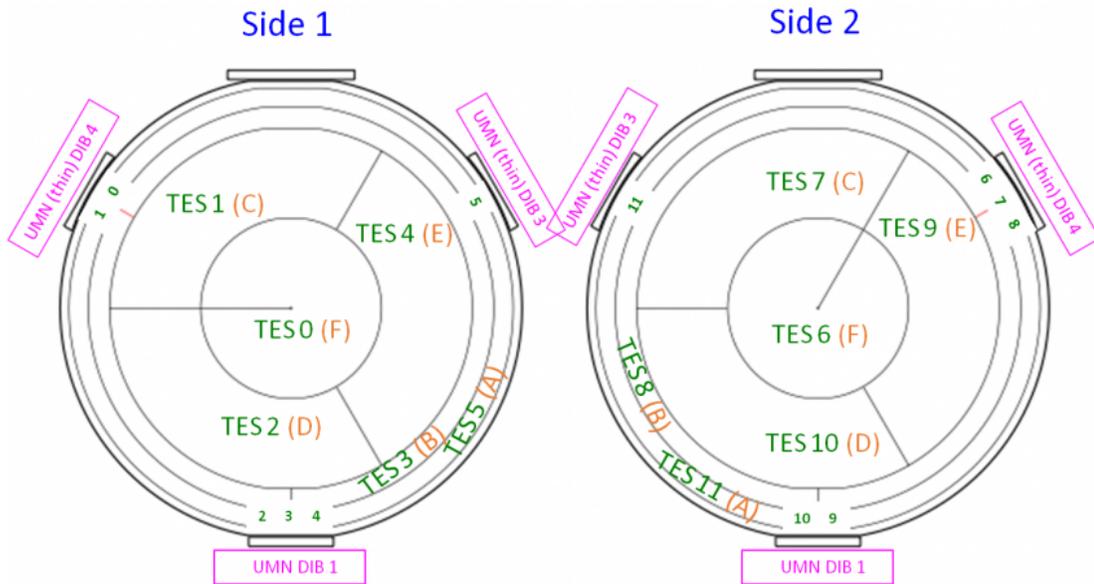
Figure 5.1: Diagram of SNOLAB HV channel layout. Green "TES" labels are only used for fabrication. Orange letter labels in parentheses are the channel names used in this work. DIB positions are also shown.

sensitivity. Finally, comparison of signals on the two sides also allow some sensitivity to vertical ($z$ axis) position.

### 5.1.1 (r, $\theta$) Estimation

**Angle**

Two different types of angular position parameters were constructed based on the detector channel map. The first type was constructed based on the relative amplitude of phonon pulses in the side 1 inner ring (PCS1, PDS1, PES1) channels. This was designed by assigning each channel a unit vector from the detector center and calculating the projection of that vector into $x$ and $y$ directions. Here, we have defined $x$ to be towards the center of PES1 and $y$ to be orthogonal and in the direction of PCS1. Then, for a given event, each channel will contribute a weight equal to its amplitude measured using the OF algorithm described in Appendix E. For compactness, we will denote this by e.g. $OF(PAS1)$. The resulting quantities will be called side 1 $x$ and $y$ partitions

and are defined as

$$
\begin{aligned}
pxpart1OF &= \frac{OF(PES1) - (OF(PCS1) + OF(PDS1))\cos(60°)}{OF(PCS1) + OF(PDS1) + OF(PES1)} \\
pypart1OF &= \frac{OF(PCS1) - OF(PDS1)}{OF(PCS1) + OF(PDS1) + OF(PES1)} \sin(60°)
\end{aligned}
\tag{5.1}
$$

Note that each quantity is normalized by the sum of individual amplitudes. This gives a natural range of the partitions from (-1,1), and makes them invariant to event energy for a detector with linear energy response.

For comparison, we constructed timing-delay based quantities using the relative start times of pulses as measured using the Rise Time Fall Time walk (RTFT) algorithm [142]. Briefly, the RTFT algorithm takes a pulse trace, smooths it with a low-pass Butterworth filter, and finds the trace maximum. It then "walks" down the rising and falling edges of the trace to locate the time at which various rise and fall percentages are reached. For example, in the parameter considered here, we use the 30% rise time of the pulse, i.e. the point at which the pulse first reaches 30% of its maximum amplitude. For simplicity, we will denote this calculation with the function $WKr30$, such that e.g. $WKr30(PAS1)$ will represent the 30% rise time of channel PAS1 in units of $\mu$s. We define the delay quantities somewhat like the OF partitions above, except we do not normalize and the signs are flipped because we expect large amplitudes and early rise times to correspond with events near the channel. Because they are not normalized, these quantities have units of time and typically lie in the (-100,100) $\mu$s range. They are weakly dependent on event energy, to the extent that pulse shapes saturate and change shape somewhat at higher energies.

$$
\begin{aligned}
pxdelWK &= (WKr30(PCS1) + WKr30(PDS1))\cos(60°) - WKr30(PES1) \\
pydelWK &= (WKr30(PDS1) - WKr30(PCS1))\sin(60°)
\end{aligned}
\tag{5.2}
$$

In the upper plots of Fig. 5.2 we show the distribution of $(x, y)$ quantities defined above using a set of data with 45 V bias. Note that the three channel geometry from which the quantities are derived is apparent in both plots. However, the timing-based quantities (left) generally give a smoother distribution. There have been only a few basic quality cuts placed on the data including a high baseline cut to remove muon tails, a saturation cut to remove any traces which exceeded the ADC range, and a threshold cut to ignore traces which triggered on noise. The lower plots have one additional

cut to select only events in the Ge K-shell activation line at 10 keV. This subset will be useful for reconstructing event positions because we know that it should be from events uniformly distributed in the detector volume. (This uniform assumption will be explored further in Sec. 5.1.3)
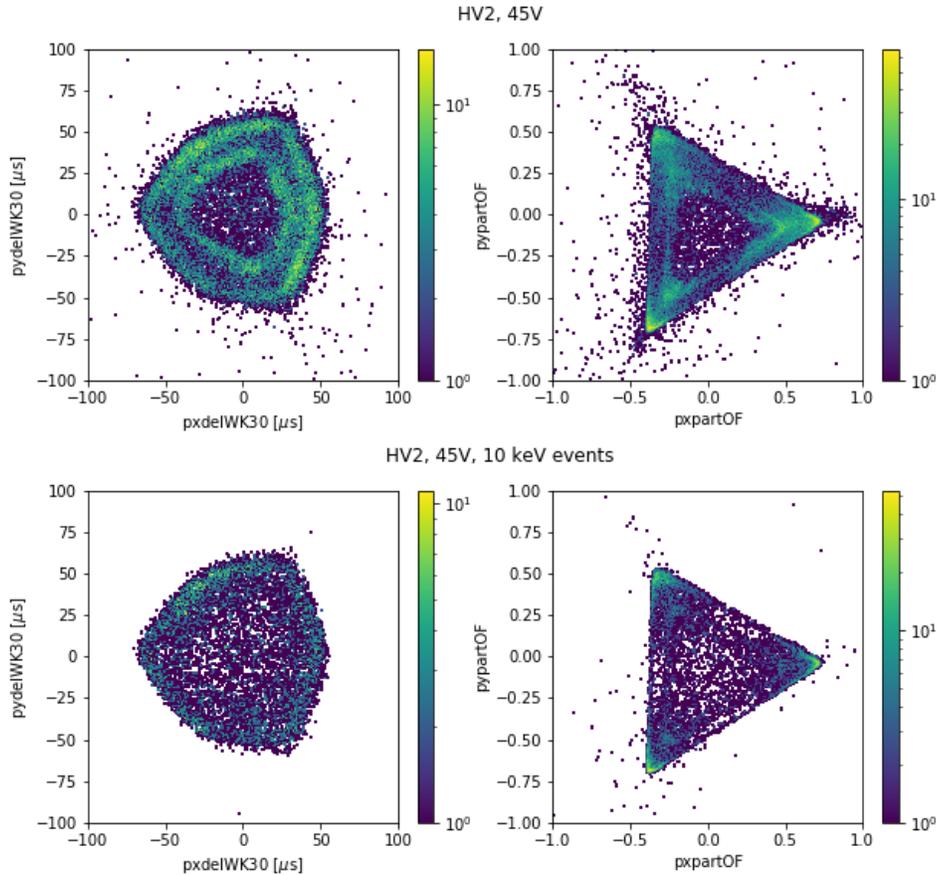


Figure 5.2: $(x, y)$ distributions from timing and amplitude of HV inner ring channels. Top: All events above threshold. Bottom: Only 10 keV events.

We can extract an angular position parameter, $\hat{\theta}$, of the 10 keV events from the $(x, y)$ parameters above. We define $\theta = 0$ as the center of PES1 with $\theta$ measured anti-clockwise. The normalized measured probability distribution functions (PDF) of these parameters are shown in the left side of Fig. 5.3. Also shown is the expected PDF of angle for a distribution of events uniformly distributed in the detector. We see that the timing-based angle is already somewhat uniform while the OF-based angle reconstructs

many events to be nearby the channel centers.
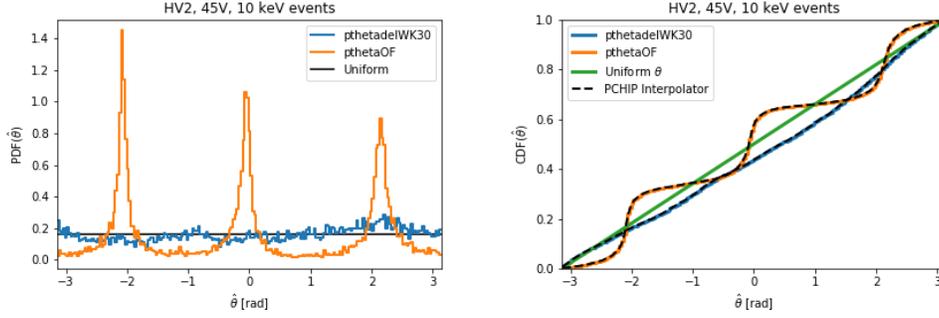


Figure 5.3: Left: Probability distribution of angular parameters. Right: Fitted CDFs of angular parameters.

The right side of Fig. 5.3 shows the measured cumulative distribution functions (CDFs) of the same parameters. They have each been approximated with Piecewise Cubic Hermite Interpolating Polynomials (PCHIPs) which use monotonic cubic splines to fit the data [187]. We will use this to make a transformation of the measured angular parameters so that their distributions are consistent with the uniform expectation. We make the assumption that the $\theta$ estimates, $\hat{\theta}$, are (on average) ordered monotonically. Then we can map to the true angle, $\theta$, by matching the value of the CDFs. It works out to be the case that the mapping function $f : \hat{\theta} \to \theta$ is

$$f(\hat{\theta}) = CDF_\theta^{-1}(CDF_{\hat{\theta}}(\hat{\theta})) \tag{5.3}$$

where $CDF_\theta$ and $CDF_{\hat{\theta}}$ are the true and measured CDFs of $\theta$ and $\hat{\theta}$ respectively. Now, we know that a uniform distribution gives

$$CDF_\theta = \frac{\theta + \pi}{2\pi}, \qquad \theta \in [-\pi, \pi) \tag{5.4}$$

which yields

$$f(\hat{\theta}) = \pi(2CDF_{\hat{\theta}}(\hat{\theta}) - 1) \tag{5.5}$$

We then use the spline fits as analytic estimates of $CDF_{\hat{\theta}}$ and can plot the resulting PDF of the transformed angular quantities as shown in Fig. 5.4. Both of the distributions are now more uniform than earlier, especially the OF-based quantity. This indicates that we have a measure of event angle that is reasonably consistent with expectations, with the delay-based quantity giving the most uniform results.
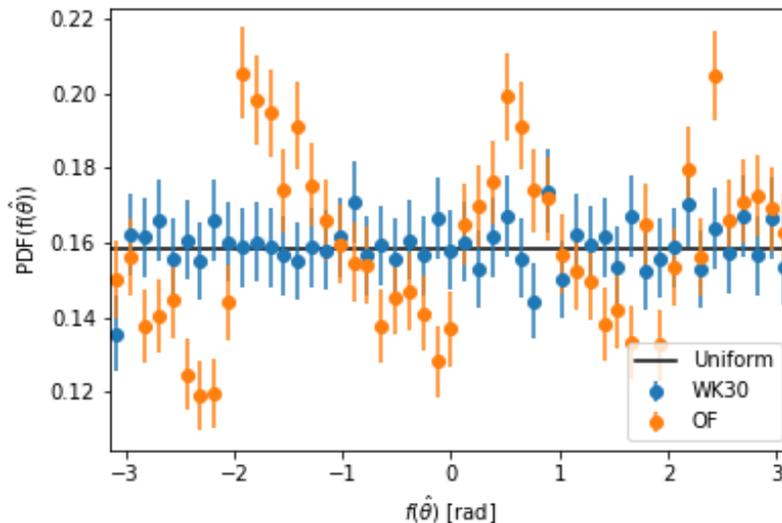
Figure 5.4: PDF of transformed parameters. Error bars show statistical uncertainties.

One simple improvement would be to repeat the above steps but with higher statistics in the 10 keV peak. There are also some edge effects apparent in the first few bins of the delay-based quantity. This would likely be alleviated by repeating this with the CDF measured from a different starting angle and averaging the two results. One could also repeat another iteration of the same CDF-matching procedure, now starting with the transformed angular values. Finally it would be interesting to try this with events from other neutron activation lines. Unfortunately, the next line at 1 keV was close to the threshold in this dataset and the measured distribution was thus affected by trigger efficiency issues.

**Radius**

Reconstruction of the event radius is important because events near the crystal surface at high radius will experience not only charge quenching surface effects but are the most likely to have their phonon signal distorted due to bias field non-uniformity.

We again explore the use of both timing and amplitude quantities to develop a good radial parameter. For such things, we also make use of the axially symmetric channels, PA, PB, and PF. Unfortunately, the noise in PA was prohibitively high in the UMN

R62 data used for these studies. However, PB has similar geometry and is expected to give reasonable results. Thus, using similar forms to those of the angular parameters, we define the following radial parameters

$$pRpartOF = \frac{OF(PBS1) - OF(PFS1)}{OF(PBS1) + OF(PFS1)}$$

$$pRdelWK30 = WKr30(PFS1) - WKr30(PBS1)$$

(5.6)

We also consider the radial information contained in the inner ring channels. We make use of the total magnitude of the quantities used before as

$$pRdelWK30xy^2 = pxdelWK30^2 + pydelWK30^2$$

$$pROFxy^2 = pxpartOF^2 + pypartOF^2$$

(5.7)

Fig. 5.5 shows each of these as a function of energy. The 10 keV activation line is prominent in each. In the upper right plot there is a hint that the parameter folds back upon itself at high radius because the activation line seems to curl back to low energies near pRpartOF=0.5.

This is explored further in Fig. 5.6, which shows the various radial parameters plotted against each other. Here we are again only considering the 10 keV events, which should have a uniform distribution in the crystal. This makes it clear that the individual quantities each fold back at high radius. However, they do so in different ways, so that the combination of two of them breaks the degeneracy. This is particularly clear in the second row. The third row, which uses pROFxy has some visible structure with three bands that are slightly different. This is an artifact of in imbalance in the amplitude calibration amongst the inner ring channels. It is the same reason the triangles in the right plots of Fig. 5.2 are slightly crooked.

For the next step, we select the combination of pRdelWK30xy and pRdelWK30 (middle left plot of Fig. 5.6) because they seem to separate the fold-back region well. This is in contrast to the combination of pRpartOF and pRdelWK30 (upper left plot) which has a simpler shape, but for which signals from multiple radii may reconstruct to the same point in the distribution. For example, at values of pRdelWK30 = 25 $\mu$s, there are two distinct populations, with some events at high radius around pRdelWK30xy = 30 $\mu$s and another group, presumably at lower radius, found near pRdelWK30xy = 50
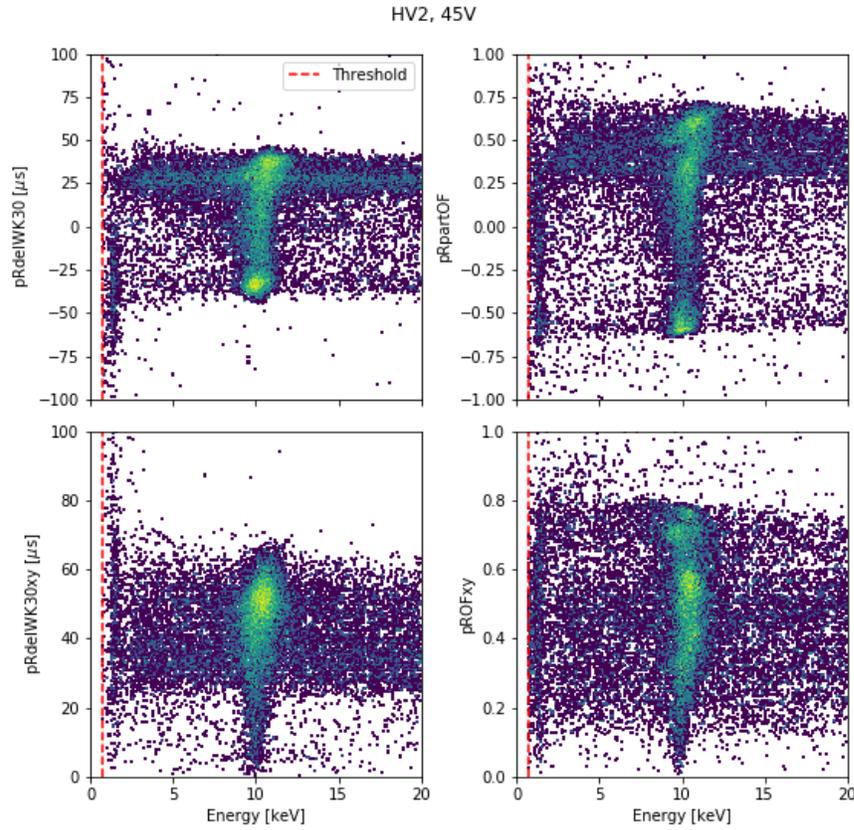
Figure 5.5: Various radial parameters as a function of energy.

$\mu$s. However, these groups would be degenerate if pRpartOF were used. Now, as with the angle, we assume that, in the pRdelWK30xy vs. pRdelWK30 plane, the events are on average sorted by their true radii. Then we can construct a new parameter, $\lambda$, which just measures the distance along the distribution. As shown in Fig. 5.7, we fit the shape of the distribution with a parametric curve. The $x$ and $y$ positions of the curve are described by $4^{\text{th}}$ order polynomials of $\lambda$, which runs from 0 to 1. Each measured data point was then mapped to the nearest $\lambda$ along the curve.

Now we simply repeat the steps used for the angular analysis to build a map from $\lambda$ to true radius, $r$. Fig. 5.7 shows the measured PDF and CDF of $\lambda$ as well as the PCHIP fit. For a uniform distribution of events in a cylinder, the theoretical distribution is $CDF_r(r) = (r/R)^2$, where $R = 50$ mm is the maximum detector radius. So we define
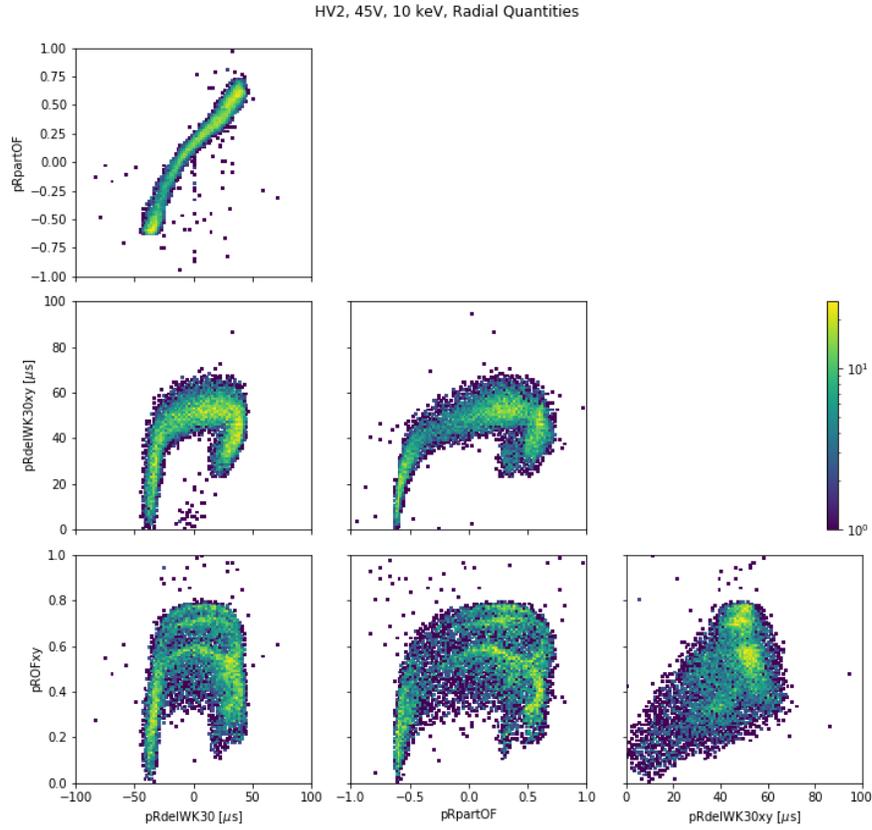
Figure 5.6: Comparison of the radial parameters.

the mapping function, $g : \lambda \to r$, as

$$g(\lambda) = CDF_r^{-1}(CDF_\lambda(\lambda)) = R\sqrt{CDF_\lambda(\lambda)} \tag{5.8}$$

Fig. 5.9 shows the PDF of the transformed radius. It is in reasonable agreement with the theoretical uniform distribution for which $PDF_r(r) \propto r/R$. The physical radii of the phonon channels are also shown for reference.

Now, we can combine our angular and radial estimates to build a 2D event map as shown in Fig. 5.10. This appears very close to uniform, but we can still see some structure at the PA/PB interface. This is likely an artifact of not having PA information in this dataset, but it is remarkable that it appears at the correct radius, even though we did not include any information about the individual channel radii in our modelling.
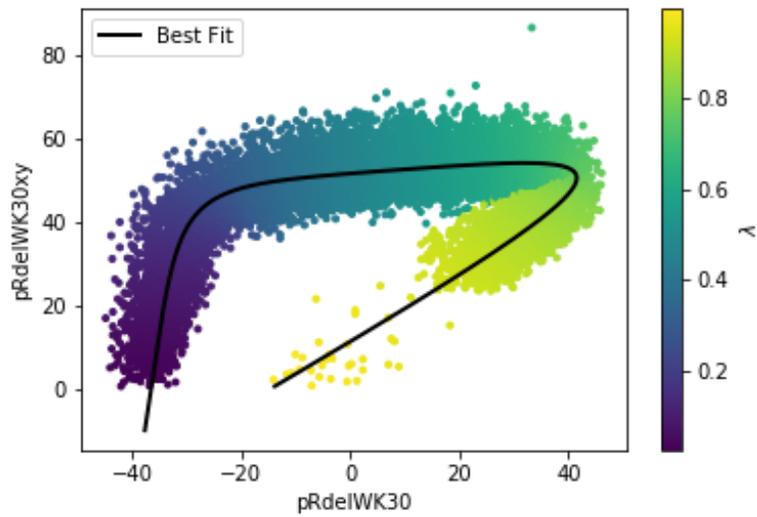
Figure 5.7: Parameterized $\lambda$ fit to radii distribution.

It would be good to repeat this with PA information as well, but this is the type of situation such redundancy (similar PA and PB) was designed in for.
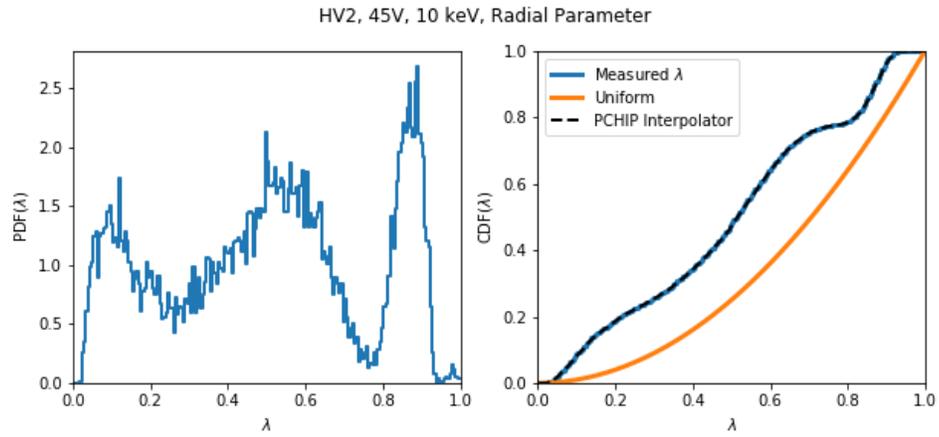
Figure 5.8: PDF and CDF of $\lambda$ radial parameter.
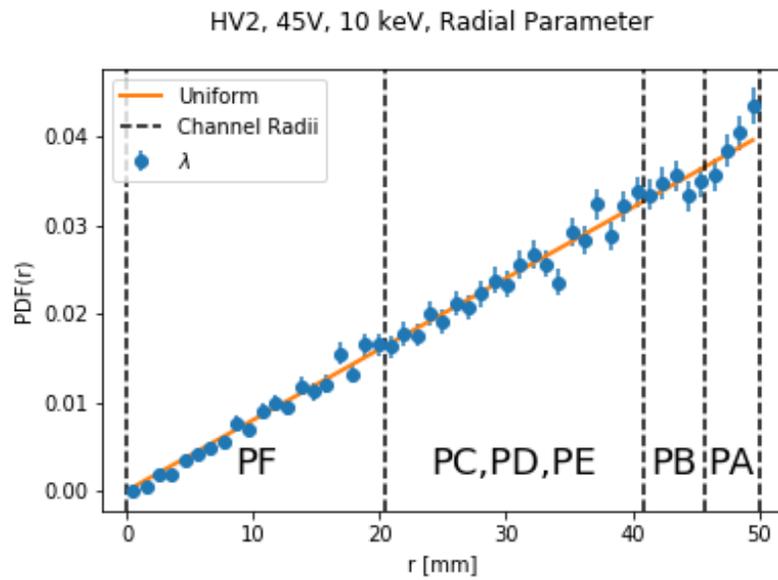


Figure 5.9: PDF of transformed radial parameter.
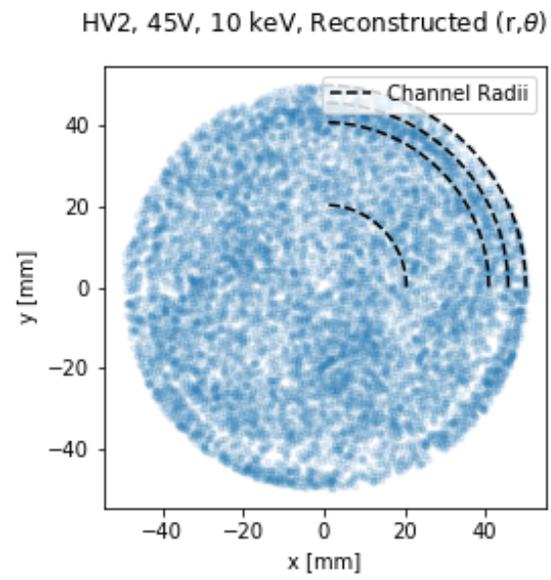
Figure 5.10: Reconstructed $(r, \theta)$ positions. Some residual structure is observed between the outer two channels, likely due to the lack of PA measurement in the analysis.

### 5.1.2 z Estimation

Unfortunately, it is much harder to obtain as satisfactory results for estimation of the event $z$ position. One major reason is the asymmetric bias/readout arrangement currently employed with HV detectors at UMN. All of the parameters we invented above used the idea of measuring event angle or radius by using information from channels that spanned a range of physical angles or radii. So too with $z$ estimation, the natural choice would be something like the following parameters which make use of channels on both faces:

$$
\begin{aligned}
pZpartOF &= \frac{\sum_{i=A}^{F} OF(PiS1) - OF(PiS2)}{\sum_{i=A}^{F} OF(PiS1) + OF(PiS2)} \\
pZdelWK &= \sum_{i=A}^{F} [WKr30(PiS2) - WKr30(PiS1)]
\end{aligned}
\tag{5.9}
$$

However, as described above, we only read out one side of the HV detector when operated with a bias applied. Another issue is that, while we can read out all 12 channels at 0 V bias, the 10 keV Ge activation line that we used above is near the trigger threshold and suffers from trigger efficiency effects which are correlated with event position.

So, we here instead study higher energy, 60 keV events from a $^{241}$Am source in a Si HV detector operated at 0 V. For this study, the source was collimated and located over the center of one face of the detector. The data was taken in UMN R62 during the winter of 2017. With a known distribution of these events, we will use similar steps to those above to develop a mapping to physical $z$ position. It is important to note that the work presented in this section is an extension of a study first carried out by D. Sincavage at UMN.

Now, another unfortunate issue was that channels PBS2, PCS2, and PDS2 were not operable in this data. To maintain angular symmetry in the definition of a $z$ parameter, we were thus restricted to using only the PA and PF channels on each face. We define

the following parameters:

$$pZpartA = \frac{OF(PAS1) - OF(PAS2)}{OF(PAS1) + OF(PAS2)}$$

$$pZpartF = \frac{OF(PFS1) - OF(PFS2)}{OF(PFS1) + OF(PFS2)}$$

$$pZdelayA = WKr30(PAS2) - WKr30(PAS1)$$

$$pZdelayF = WKr30(PFS2) - WKr30(PFS1)$$

(5.10)

A comparison of these quantities as a function of event energy can be seen in Fig. 5.11. In the left two figures, which use the inner, PF channels, there are clear vertical structures associated with the $^{241}$Am lines at 60 keV and from 10-20 keV. This is expected since the source was located above the center of the detector. Interestingly, there is also a region near the top and bottom where the $z$ parameters appear to curl back on themselves possibly due to some sort of surface effect. There is also a strong horizontal band of events which have PF-based partitions near 0. These are likely from events further out in radius, far from PF, which evidently do not impart as much $z$-position information in PF. The right two figures use the PA channels. There, the $^{241}$Am lines are observed to have a very narrow range of $z$ parameters, again because the signal is distant radially from the channels used.

As a compromise then, it is sensible to use both of the PA and PF quantities to maintain reasonable $z$ sensitivity at all radii. Essentially, we define this $z$ parameter to be the same as Eq. 5.9, but without PB–PE. Additionally, from here forward we will simply use the OF-based quantities, since the delay-based ones (lower figures in Fig. 5.11) appear to contain similar information with inferior resolution. Finally, we will also use a radial parameter, which we define using the inner and outer channels as we did in Sec. 5.1.1. Thus we have

$$pZpartAF = \frac{OF(PAS1) - OF(PAS2) + OF(PFS1) - OF(PFS2)}{OF(PAS1) + OF(PAS2) + OF(PFS1) + OF(PFS2)}$$

$$pRpartAF = \frac{OF(PAS1) + OF(PAS2) - OF(PFS1) - OF(PFS2)}{OF(PAS1) + OF(PAS2) + OF(PFS1) + OF(PFS2)}.$$

(5.11)

In Fig. 5.12, we show the $z$ vs. $r$ distribution using these parameters for a range of energies from 50–70 keV. The 60 keV events can be seen as a thin curve on the left side
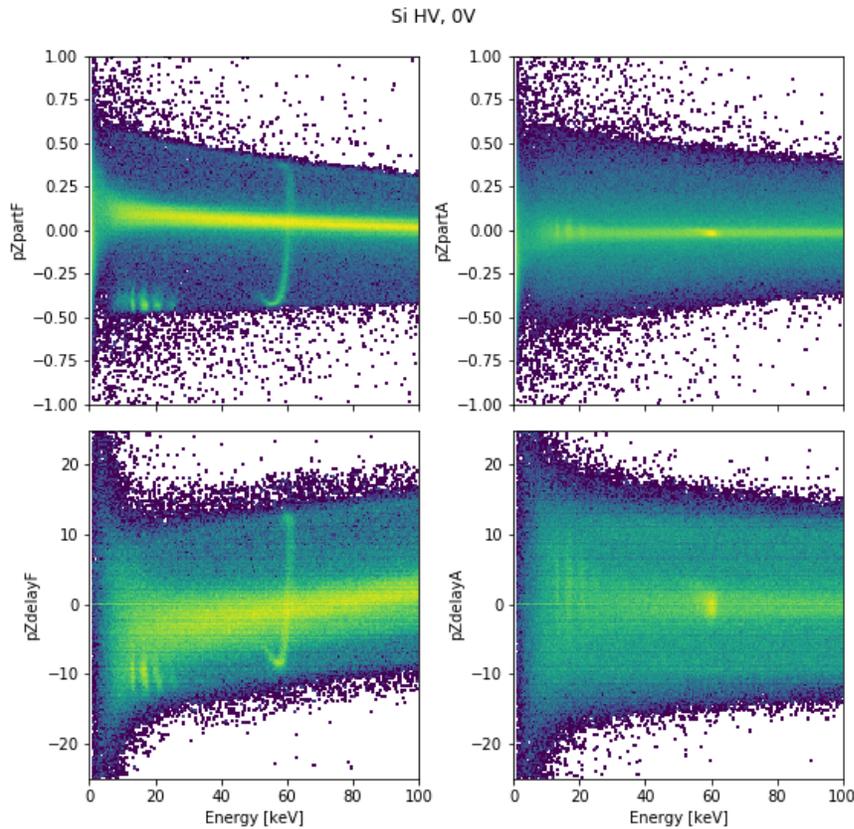
Figure 5.11: Comparison of several $z$ parameters as a function of event energy.

of the figure, at low radius. We can see here that both the radial and $z$ parameters are most sensitive near the channels which define them, i.e. at the extreme values of $z$ and $r$. In between, they are less sensitive and the distribution is squeezed. In the future, including the inner ring channels will help to normalize the shape of this distribution. Nevertheless, this is sufficient for our purposes here.

As in previous sections, we would like to obtain the distribution of $z$-parameter values for the known population, here the 60 keV events from $^{241}$Am. We first make a crude selection of the low-radius region as indicated in Fig. 5.12. Then, we consider these events as a function of energy as shown in the left side of Fig. 5.13. With the higher radius events removed, we can now see the 60 keV line much more clearly. However, there is still a background distribution of events that we do not want to pollute our
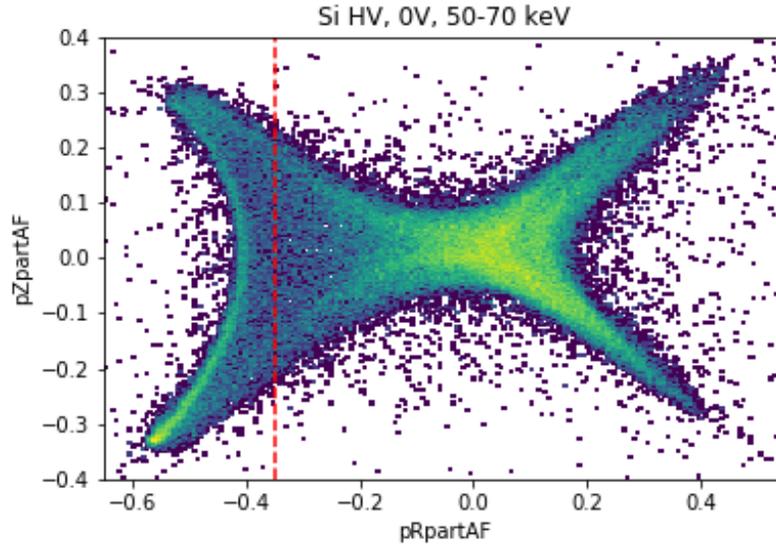
Figure 5.12: Distribution of events using radial and $z$ parameters based on PA and PF channels. Cut used to select low radius events is indicated with the dashed line.

sample. So we perform a background subtraction as follows. First, we normalize the energy dependence of pZpartAF by transforming the solid red lines which follow the distribution edge in the left side of Fig. 5.13, to each be flat as shown on the right figure. We call this transformed variable, pZpartAF_norm. Then, we select three regions that span equal ranges of energy as indicated with dashed lines. The lower and upper regions bracket the middle region, which contains our signal.

In the left side of Fig. 5.14, we show the distribution of pZpartAF_norm for each of the three regions. We then make the assumption that the background distribution is smoothly varying, so we can subtract the average of the lower and upper distributions and be left with the distribution due to 60 keV events from $^{241}$Am. This is shown as the red curve. Once this is obtained, we follow the now-familiar step of calculating and fitting its CDF as seen on the right side of Fig. 5.14.

We then need the expected CDF. The flux of a beam of photons directed into a material will fall off exponentially with a length scale, $\lambda$, which is 13.38 mm for 60 keV gammas in Si [188]. With a detector of thickness, $h = 33.3$ mm, we obtain

$$CDF_z(z) = \frac{1 - \exp[-z/\lambda]}{1 - \exp[-h/\lambda]} \tag{5.12}$$
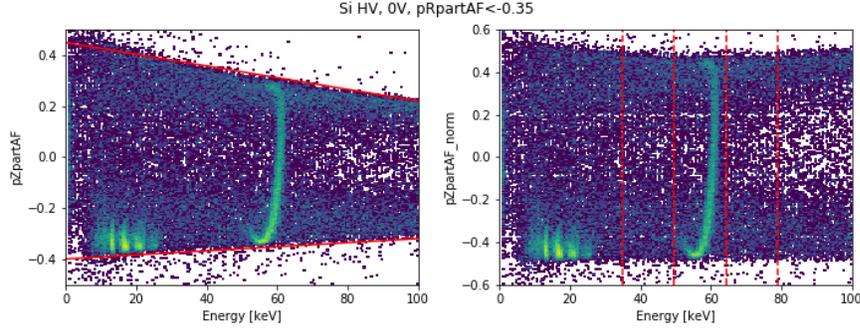
Figure 5.13: Distribution of $z$ parameter as a function of event energy for low radius events. Left: Using the original pZpart definition. Solid lines show the rough distribution range. Right: After stretching the distribution to normalize out energy dependence. Dashed lines indicate the three energy ranges used for signal and background subtraction.
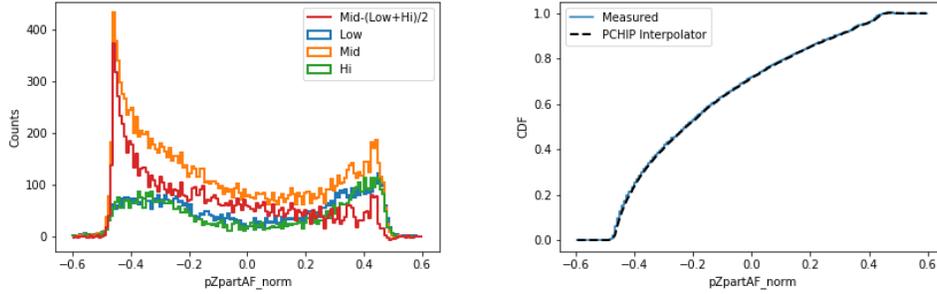


Figure 5.14: Left: Background subtraction of normalized $z$-parameter distribution for 60 keV events near the detector center. Right: Measured CDF of 60 keV radial parameter.

And thus we transform from any $z$ parameter, $zPar$, to an estimate, $\hat{z}$, of the physical position as

$$
\begin{aligned}
\hat{z} &= CDF_z^{-1}(CDF_{zPar}(zPar)) \\
&= -\lambda \ln\left[1 - (1 - \exp[-h/\lambda])CDF_{zPar}(zPar)\right].
\end{aligned}
\tag{5.13}
$$

Using this transformation, we produce the (again background-subtracted) distribution of $\hat{z}$ shown in Fig. 5.15. The transformed quantity provides a reasonable fit to the expected distribution, with the exception of the points very near $z = 0$ and $z = h$. This is not totally surprising because we did not "uncurl" the energy-dependence in the ends of pZpartAF, so the values there are not sorted properly. That means that $\hat{z}$ will

not properly sort the values in the curl-back region, but, because of the CDF-matching procedure we used, it should still be correct for events in the crystal bulk.
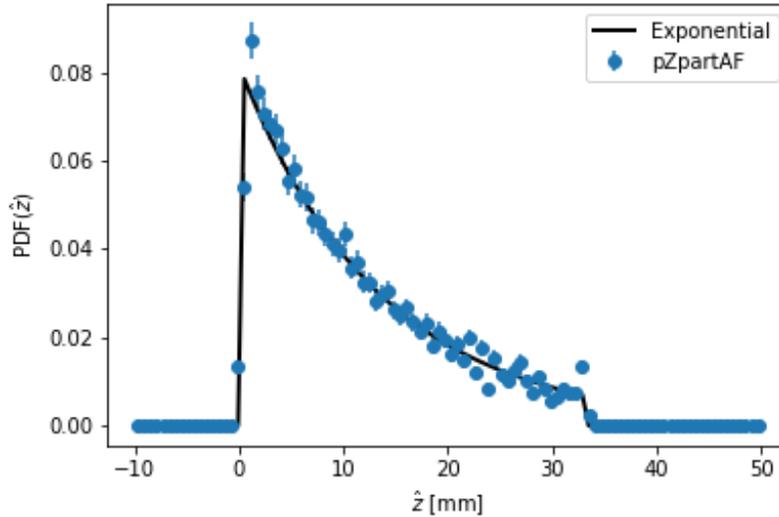


Figure 5.15: PDF of transformed $z$ parameter compared to exponential model.

There are two ways to deal with this. One, which is currently ongoing, is to study other properties of these near-surface events to understand the origin of this effect. Perhaps some aspect of the pulse shapes can break the $z$-parameter degeneracy and be used to uncurl the distribution. D. Sincavage has already seen some evidence that the rise times of channels on the opposite face provide some differentiation, but more work should be done to understand the ultimate cause of this behavior and how well it can be utilized for $z$ estimation. We do know of effects which reduce the amount of charge produced for events near the surface. However, this detector was operated with no bias voltage, so the phonon energy measured should simply be the total recoil energy, regardless of how many charges where produced. At face value, this makes the $z$ parameter curl somewhat surprising. It may be the case that this is simply a result of the phonon transport properties and channel arrangement. Or it could be that the initial, athermal diffusive phonon population is more efficiently down-scattered into the thermal continuum when events occur near crystal surfaces or electrodes, leaving a larger fraction of phonons below the QET sensitivity.

In the mean time, a more conservative approach is to simply remove those events from consideration. In the left side of Fig. 5.16 we show the distribution of $\hat{z}$ as a function of energy for events near the detector center. We also indicate two cut lines which remove all events within 5 mm of the two detector faces, which covers the region of the curl. The right figure shows the result on the measured energy distribution. While the low energy lines are mostly eliminated (20 keV penetration depth in Si is $< 1$ mm) the 60 keV distribution is cleaned up substantially and appears much more like the expected Gaussian.
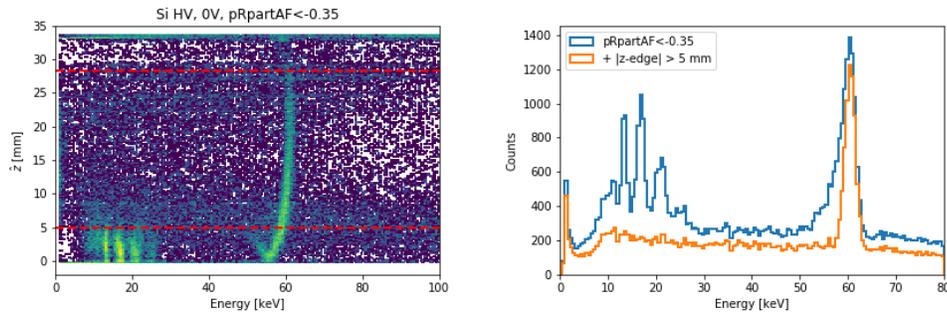


Figure 5.16: Left: Physical $z$ parameter as a function of energy. Red lines indicate cuts to remove events within 5 mm of the surface so as to exclude regions of reduced energy signal. Right: Effect of this $z$ fiducialization on measured energy spectrum.

Now we have constructed a $z$ parameter that appears to give a decent fiducialization. But it is important to remember of the several important caveats in this work. This analysis was done at 0 V, which allowed us to access all (working) channels, but which will also have a rather different phonon distribution than data taken with higher bias voltages. This is because the phonon population at 0 V is entirely recoil phonons, which are expected to be produced as an isotropic sphere emanating from the recoil site. At high voltage, the signal is instead dominated by NTL phonons emitted (roughly) in columns spanning the full detector thickness. So while the same parameters and techniques will likely be useful, the exact shapes of the distributions are expected to be different. In particular, more of the energy will be focused in the channels immediately above and below the event, so the position sensitivity we saw in Fig. 5.12 will likely be exacerbated and it will be of critical importance to have all phonon channels operating and read out.

We also only made this mapping for a single energy at low radius. This procedure should be repeated with a wider distribution of energies and positions to provide a full mapping. In Ge, this can be done using the neutron activation lines as before. However Si has no such lines, so this will likely need to be carried out with calibration sources placed, or moved, across a range of positions.

### 5.1.3 Uniform Neutron Activation Assumption

The above procedures for generating mappings from measurement parameters to real physical locations used the assumption that 10 keV events from neutron capture in Ge were uniformly distributed. We briefly examine this assumption by studying the results of a simulation generated for a neutron activation analysis.

In UMN R66, we used a strong PuBe source to expose a Ge HV detector to a large neutron flux and measured the resulting rate of neutron captures. Anthony Villano also developed a detailed Geant4 [189] model of the arrangement. In this simulation, neutrons were thrown from the PuBe source and propagated to the detector where they could interact by scattering or capturing.

In Fig. 5.17, we plot the simulated distribution of capture locations, $(r, \theta, z)$, in the Ge HV detector. These are compared to the expectations for a uniform volume distribution. We observed significant asymmetry, especially in the direction of the external PuBe source. One cause of this may be the occurrence of a significant amount of capture of non-thermal neutrons coming from the source.

The point here is that the actual spatial distribution of events may depend on details like the shielding and source geometry as well as the neutron energy spectrum used for activation. In these cases, one can still use all the same procedures to provide mappings from the measured signals to the true event location, but instead of calculating the simple analytic forms of the assumed CDFs, one should instead approximate the true CDFs by fitting simulated distributions like these.
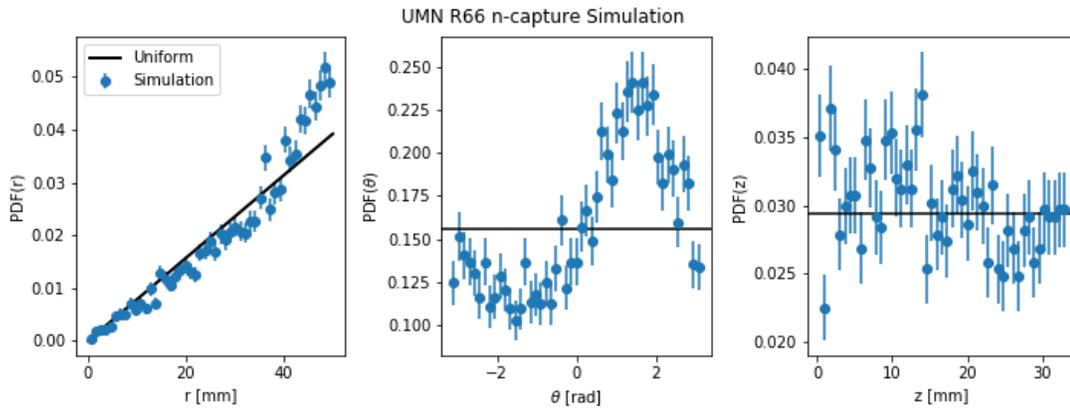
Figure 5.17: Simulated neutron capture distributions in UMN R66 displaying preferential capture in portions of the detector nearest to the neutron source. Error bars are statistical.

## 5.2 iZIP Detectors and an In-Situ Source Mover

In the previous section, we demonstrated how a suitable calibration source can be used to map the $z$ distribution at a single $(r, \theta)$ point in a detector. Building on this, it becomes desirable to develop a method to repeat such calibration by taking data with a source at a variety of locations around the detector. This is a non-trivial task for detectors which are operated inside a cryostat at mK temperatures and with low energy radioactive sources, which cannot efficiently penetrate to the detector from outside the cryostat. To address this issue, we have developed a prototype cryogenic device which is mounted in a detector housing and moves a source across a range of radial positions. This section will describe the design and operation of this source mover. We will also present some examples of how data collected with this device is being used to calibrate position quantities in a SNOLAB iZIP detector. Some of the work presented here has been published in [190].

### 5.2.1 Design and Construction

There are commercially available cryogenic positioning stages sold by companies like SmarAct, JPE, PI, etc. However, neither the cost ($\sim$\$10k) nor the available designs were suitable for the project at hand. In addition, many such cryogenic devices are piezoelectric-based actuators, which operate using $>100$ V AC signals. There was significant concern that, even if the existing SuperCDMS tower wiring could withstand such signals, there would be significant EMI disruption to the sensitive TES and SQUID electronics. So we instead opted to develop a homebrew alternative building on previous work which demonstrated simple modifications to a commercial stepper motor which allowed stable cryogenic operation [191].

Following a similar procedure, we obtained a partially assembled PF25-48P4 stepper motor from Nippon Pulse America Inc. Usually, the motor housing is sealed with a few tack welds in the final step of assembly. However, we were able to obtain samples which were not yet welded, but only for this particular model. The PF25-48P4 motor is a cylinder approximately 25 mm in diameter and 15 mm deep and is likely larger and more powerful than necessary for this purpose. It would be interesting to investigate more compact designs in the future. One could still use commercial motors, but may

have to first break the weld joint to open the motor case by carefully cutting the tack welds with a diamond saw or Dremel tool.

In our case, the motor used two axial drive coils, each consisting of ∼900 turns of fine copper wire with total resistances of 120 Ω. We unwound these coils and replaced them with Nb-Ti superconducting wire. Each coil was wound with 900 turns of 4.1 mil (3 mil core, formvar-clad) Nb-Ti wire from Supercon, Inc. Fig. 5.18 shows a photo of the crude rig used to wind the coils. A power drill was used to rotate the coil and wind Nb-Ti wire from a large spool in a controlled manner. The number of turns was counted by means of an electronic switch attached to a NIM scaler module. After winding, each coil had a room temperature resistance of 7.5 kΩ due to the high normal resistance of Nb-Ti.
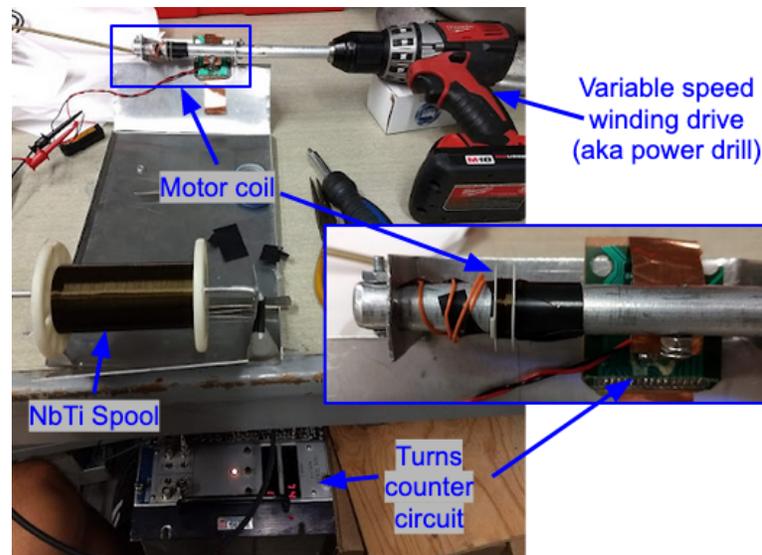


Figure 5.18: Photos of the motor coil winding jig. The small motor coil (inset) was wound with Nb-Ti from a large spool. The coil was mounted on a motorized axle and the number of turns was counted by a circuit consisting of an electronic switch and NIM scaler module.

One large issue with this design, which manifested in multiple ways, was the particular choice of wiring. Nb-Ti itself is an excellent material for this project as it superconducts below about 10 K. This can allow the currents required to power the motor to flow without generating Ohmic heating in the coil, an essential feature for

stable operation at mK temperatures. However, the normal resistance was so high that, with the pure Nb-Ti selected, we could not operate the motor at room temperature for testing. Additionally, bare Nb-Ti is notoriously difficult to make good electrical connections to. For this experiment, we tightly wound the stripped end of the Nb-Ti wire around an Al wire pigtail and coated the joint in Pb-Sn solder. Unfortunately, this may have resulted in joint resistances as high as a few Ohms, though this was difficult to quantify on the bench due to the large Nb-Ti resistivity. As will be discussed later, this likely lead to undesired heating when operating the motor. In retrospect, a more appropriate choice would have been some type of copper-clad Nb-Ti which is available commercially from Supercon. Such a material would allow room temperature operation, since the normal resistivity would be similar to that of the original copper motor wire [192], and would enable reliable electrical connections to other parts of the wiring with standard soldering techniques.

A brass #6-32 threaded rod, linked to the motor, was used to drive a source holder equipped with a small radioactive calibration source. As shown in Fig. 5.19 the source holder could be threaded onto the drive rod and fit over a guide rail which caused the motor rotation to linearly actuate the source position. With a count of 32 threads per inch and a motor step angle of 7.5°, a single step of the motor moved the source a distance of 16.54 $\mu$m. To prevent freezing or sticking of any lubricants or oils, all the mechanical and motor parts were cleaned with methanol in an ultrasonic cleaner bath. They were then only handled and assembled in a clean room environment. The calibration source used in this experiment was $^{241}$Am which has a prominent 60 keV gamma line. A 66 mil thick lead disk with an 18 mil hole collimated the source gammas and restricted the emission rate. A strip of Kapton tape placed over the collimator blocked alpha emission.

The motor was controlled by an EasyDriver [193] stepper motor controller and an Arduino Uno [194] microcontroller board. Stepper motors are actuated by sending square current signals of alternating polarity to the two coils. The motor was operated in full-step mode, giving 48 steps per rotation. Custom firmware was written for the Arduino so that the user could control the rate and duration at which these signals were sent. The amplitude of the motor drive signals was set by an external variable power supply.
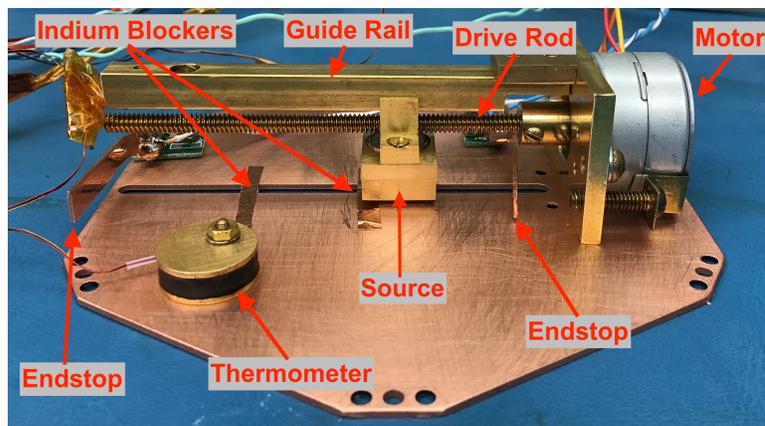
Figure 5.19: Stepper motor and source holder assembly.

## 5.2.2 Experimental Setup

The motor and source assembly was mounted on a copper plate with a slot through which the gammas could pass unimpeded. This plate was mounted in a spare detector housing 1.6 cm above a SuperCDMS detector, allowing the calibration source to illuminate one face of the detector.

One of the benefits of using a stepper motor was that the source position could be calculated to excellent accuracy by keeping careful track of all steps taken by the motor and using dead reckoning (i.e. assuming the distance moved relative to a known starting location). However, for these initial trials, it was not certain whether the motor would operate properly or might miss some steps. To hedge against this, two conductive endstops were used to validate the source position at the extreme ends of the source holder's range of motion. These endstops were simply rigid copper strips, insulated from the housing, and connected to wires leading out of the fridge. In this way, contact of the source holder with an endstop was indicated with a simple resistance measurement to ground.

We also installed two small indium blockers (1 mm thick, 2 mm wide) in the slot of the copper plate. These were thick enough to effectively shield the detector from the source when it was moved to those locations. The main purpose of these blockers was to provide a secondary means of locating the source by noting when the calibration signal was blocked. Again, this is not strictly needed in cases where the motor is operating

properly, but was a hedge against such issues. However, the blockers were also being tested for a longer-term purpose. It is often desirable to include a calibration source when collecting DM search data. However, the required high rate of the source may overwhelm possible low rate DM signals. The blocker and motor combination provide a means of effectively toggling the calibration source illumination on or off, allowing one to efficiently collect e.g. low rate DM search data sets interspersed with high quality calibration measurements.

As the first test of this design, the source mover assembly was mounted above a prototype SuperCDMS SNOLAB iZIP detector [195]. The assembly was oriented to move the source along the detector diameter as shown in Fig. 5.20. For this study, the detector was operated with biases of $\pm 4$ V on the two faces. The detector was read out using existing CDMS-II cold hardware [138] with minor modifications. We used prototype RevC DCRCs for control and readout of the charge and phonon channels.
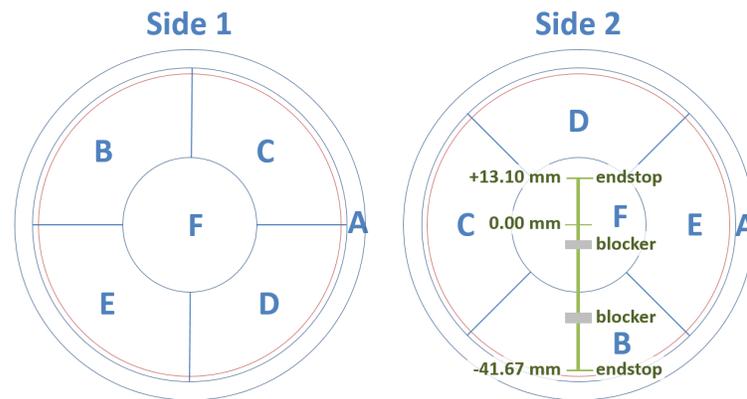


Figure 5.20: Detector layout and calibration source path. Divisions between phonon sensors (labeled A-F on each side) indicated in blue and between charge sensors (inner and outer on each side) in red. Source range of motion indicated in green. Diagram courtesy of M. Fritts.

### 5.2.3   Heating Issues

The detector and movable source were mounted in the UMN K100 dilution fridge and operated at $\sim 35$ mK during R71 in the fall of 2019. Unfortunately, we observed excess heating of the base temperature stage when operating the stepper motor. The likely

source was Joule heating in the aforementioned connection between the Nb-Ti coil and wiring harness. By sending low DC current ($< 5$ mA) down a single coil, we observed fridge heating consistent with a resistance of $\mathcal{O}(1\ \Omega)$ at the location of the Nb-Ti joint and wiring harness.

This was an issue when trying to operate the motor because it required drive currents of $\gtrsim 30$ mA through the coils to achieve reliable stepping motion. This lead to $\mathcal{O}(1\ \text{mW})$ heating which was strong enough to drive a portion of the adjoining Nb-Ti coil normal. Once this process began, the normal portion of the Nb-Ti would also heat, leading to a runaway increase in the coil resistance and heating. Evidence for this was found by monitoring the motor coil drive signals as shown in Fig. 5.21. We measured the motor driver board's applied voltage, $V_{in}$, and current drawn by a single coil, $I_{mon}$, in series with the $\sim 120\ \Omega$ wiring running from room temperature to 4 K.
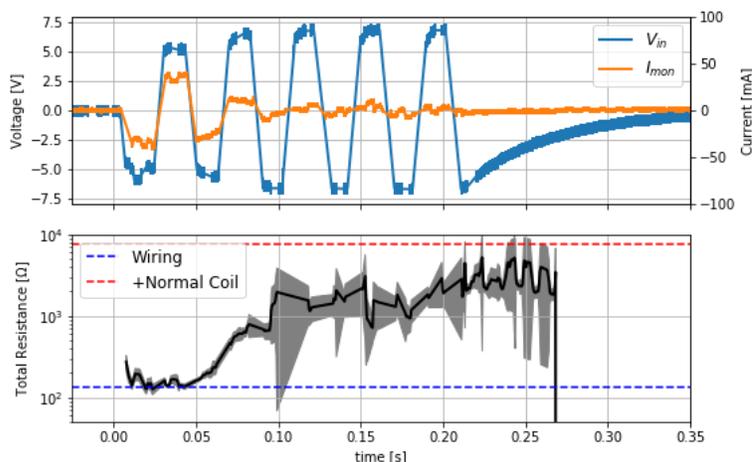


Figure 5.21: Top: Drive signals for a single motor coil while attempting to move 20 steps. Bottom: Inferred total resistance of coil and wiring. After the first few steps, the resistance was observed to increase. Drive voltage increased to compensate until it saturated at $\pm 7$ V. Such behavior indicated that the Nb-Ti coils were heated out of their superconducting states after just a few steps.

The signals observed when trying to move the motor by 20 steps are shown in the upper plot. The lower plot shows the inferred total resistance of the coil and wiring. Separate measurements of the fridge wiring (blue) and plus normally-conducting motor coil (red) are also shown for reference. The motor driver board was designed to output a

certain maximum drive current when stepping. To do so, it increased the drive voltage until either the desired current setting was reached or the voltage source railed. In this example, the resistance remained low and consistent with the fridge wiring for the first few steps until 50 ms. After that, the resistance was observed to increase to a few kΩ with further operation. When this occurred, the driver board tried increasing the voltage amplitude to compensate, but could not do so beyond the maximum power supply setting of 7 V.

As shown in the lower plot of Fig. 5.21, the resistance increased to a value consistent with a large fraction of the motor coil going normally-conducting. This measurement indicates that only part of the coil is normal, so it is probably not the Nb-Ti itself generating the initial heating, otherwise we would likely observe the full normal resistance. This is bolstered by the fact that the critical current of this Nb-Ti wire is expected to be $\gtrsim$ 10 A, far larger than anything we applied [196]. These observations lead to the conclusion that the heating is primarily driven by the resistance of the joint between the Nb-Ti coil ends and the short Al wiring pigtail which fed out from the detector housing to the tower wiring. Operating the device for much longer periods of time would overwhelm the fridge with the substantial Joule heating. This made it difficult to estimate whether friction or eddy currents caused significant additional heating. We believe that future designs with improved connections will alleviate this issue. This may be accomplished either by use of Cu-coated Nb-Ti or with other methods of producing low-resistance joints. One such technique which is currently being investigated is simple clamp connections utilizing another superconductor like Al. Another option is to use a relatively simple chemical procedure recently developed specifically for electroplating the ends of Nb-Ti wiring with Cu [197].

### 5.2.4  Stutter-Step Operation

Even with these issues, we were able to accomplish reasonable movement of the source. Heating was minimized by operating with a low drive current of 35 mA and only energizing the coils immediately before stepping the motor, then de-energizing them once the stepping was completed. Additionally the motor was only moved in bursts of 4–6 steps at a time, since after this the coils heated and the drive current could not be maintained. We also added 200 $\mu$F smoothing capacitors in parallel with each coil to reduce

possible eddy heating due to transient spikes in the drive current signal output by the controller board. The time scale of this smoothing required that the dwell between step signals be 40 ms. In this experiment, we did not work to optimize the balance between a shorter smoothing timescale, which allowed a shorter dwell for more steps before the coils heated, and the possible increased heating from transients.

When operated in this "stutter-step" mode, the base temperature would momentarily spike to 500 mK, but return to $< 50$ mK within 5 minutes, at which time we were free to move ahead another 4–6 steps. We automated this procedure which allowed the motor to operate at an average of $\approx 1$ step/min for arbitrary periods of time without overwhelming the refrigerator. This was certainly not the ideal mode of operation as it only gave a source translation speed of $\approx 1$ mm/hr. However, this was enough to allow a cursory study of the operation of this system with the source at a variety of positions across the 50 mm detector diameter during a cryogenic run lasting only a few weeks, which is typical of detector characterization studies. Of course improved heating and translation speeds can improve this somewhat, although the data taking itself also takes significant amounts of time.

While stepping the motor we ensured the source was moving by monitoring the coil drive signals as in Fig. 5.21. We also monitored the endstop resistances to confirm when the source holder reached the end of its motion. In this way we found that number of step commands sent to the motor was consistent with the number required to move this distance, implying 100% step efficiency. This allowed us to estimate the intermediate source locations to within 60 $\mu$m.

### 5.2.5  Phonon Collection Efficiency

It was discovered that, after each operation of the motor, the phonon collection efficiency (PCE) of the QETs was diminished. PCE is the efficiency with which phonons from the detector crystal are coupled to the transition edge sensor of the QET. We observed reduced signal amplitudes on all phonon channels after operating the motor, but unchanged charge signals as shown in Fig. 5.22. This indicates lower PCE as opposed to charge trapping or other crystal effects.

We believe this was a result of the superconducting Al fins trapping magnetic flux from the motor drive coils. This flux in turn trapped diffusing quasiparticles in the
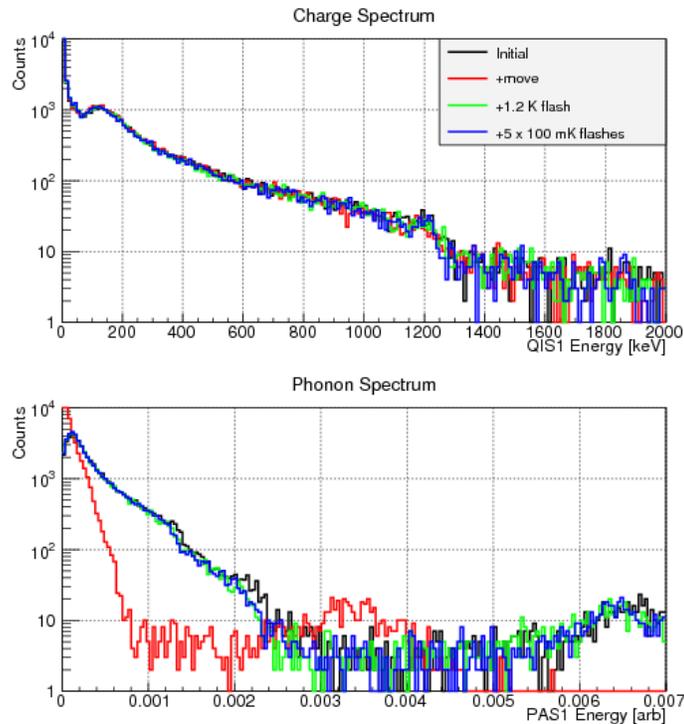
Figure 5.22: Charge and phonon spectra showing reduced PCE immediately after motor operation (+move) and restored PCE after 1.2 K heating (+1.2 K flash). Subsequent LED heating and detector operation (+5 x 100 mK flashes) did not affect PCE.

Al, reducing the fraction which heated the TES and thus lowering the observed signal [198]. Operating the motor heated everything at the base temperature stage including the detector and produced magnetic fields to rotate the motor. Measurements of the unmodified, copper-wound motor at room temperature indicated that these fields can be as as high as 400 $\mu$T just outside of the motor housing. We believe this heating could have driven the Al fins normal such that, when they cooled again, they trapped magnetic flux still present from the motor. Now, as mentioned above, the maximum temperature measured at the motor mounting plate during operation was $\sim$ 500 mK, less than the Al fin $T_C$=1.2 K. However, the temperature of the Al fins could be higher because they were exposed to direct IR radiation from the motor, and there is a delay of some few seconds until the detector temperature comes into equilibrium with the base temperature stage. This is because the detector crystal is purposefully thermally

isolated from the detector housing with the Cirlex clamps that hold the crystal in place.

Another piece of evidence supporting this theory was that we were able to restore the PCE to the nominal value by heating the detector above 1.2 K for a few seconds and letting it cool back down. This was accomplished using LEDs installed inside the detector housing. We also saw that a second round of LED heating did not further change the PCE, indicating that a single 1.2 K heating left it fully restored. Because of this, we adopted a procedure of moving the source from one desired location to the next using the stutter-step method, then briefly heating the detector above 1.2 K to restore PCE. After this heating, the detector would cool back to $\sim 35$ mK within 2 hours at which time the detector could be operated.

### 5.2.6   PCE Modelling

To further study the hypothesis that the PCE reduction was due to trapping magnetic field from the motor, we constructed a simple physical model to describe PCE as a function of ambient magnetic field strength. This subsection describes this model, previous PCE data we used to tune it, and the implications for motor operation.

**Al Quasiparticle Diffusion Model**

We model this effect by considering the diffusion of Quasiparticles (QPs) in the Al fins. To first order, we can just consider the 1-D diffusion of the QPs along the length of the fins [199]. The particles obey the diffusion equation

$$\dot{n} = D\nabla^2 n - \frac{n}{\tau} + s \tag{5.14}$$

where $n$ is the number density of QPs, $D$ is the diffusion coefficient, $\tau$ is the QP lifetime, and $s$ is a source term. For our purposes, the source term is given by

$$s(\vec{x}, t) = \delta(\vec{x}, t) \tag{5.15}$$

By combining these expression with appropriate boundary conditions for absorbers, Appendix D of [199] solves Eq. 5.14 for the geometry of QPs diffusing along a 1-D fin with a QP trap (TES-Al overlap region) at one end. In that case, the fraction of phonons that reach the TES, which we here call phonon collection efficiency (PCE), is

$$PCE = \frac{\lambda_d^2}{\lambda_a + \lambda_d \coth[1/\lambda_d]} \tag{5.16}$$

Here the (dimensionless) absorption length, $\lambda_a = D/vw$, where $v$ is an absorption coefficient for the QP trap and $w$ is the Al fin length and the diffusion length, $\lambda_d = \sqrt{D\tau}/w$.

Now, we need to include the additional effect of the magnetic field. We assume that the QP lifetime can be written as

$$\tau^{-1} = \tau_0^{-1} + \tau_{trap}^{-1} \tag{5.17}$$

where $\tau_{trap}^{-1}$ is the timescale over which a QP (with an otherwise infinite lifetime) would become trapped by pinned flux in the Al and $\tau_0$ is the QP lifetime without any additional magnetic flux effects. Further, we will assume that the trap lifetime is proportional to the mean free path for scattering on a trap, $l_B$. Then, the mean free path, which is inversely proportional to the trap density, will go like the inverse magnetic field flux through the Al fin. Thus we have

$$\tau_{trap} \propto l_B \propto \Phi^{-1} \propto B^{-1}. \tag{5.18}$$

where $\Phi$ is the magnetic flux through the fin and $B$ is the average magnetic field strength at the fin. We parameterize the scale of this effect with a constant, $c$ (having units of T×s), and obtain

$$\tau^{-1} = \tau_0^{-1} + B/c \tag{5.19}$$

which allows $B$ to enter into the PCE via $\lambda_d = \sqrt{D\tau}/w$. Now we have a simple PCE model with a few physical parameters.

**PCE vs. Magnetic Field Measurements**

In UMN R59 we measured how the presence of magnetic fields when cooling the detectors reduced phonon collection efficiency (PCE). This was accomplished by varying the position of an external mu-metal shield and measuring the phonon signal generated by the 60 keV line from [241]Am. The resulting measurements are shown in Fig. 5.23 where we see the expected reduction of PCE as the background magnetic field is increased.

Now, the measurements shown in Fig. 5.23 are a function of the magnetic field strength at the detector position but *without the dewar and fridge present.* Our model depends on the magnetic field present at the detector, so we need to account for additional magnetic shielding effects of the fridge components. We define the shielding
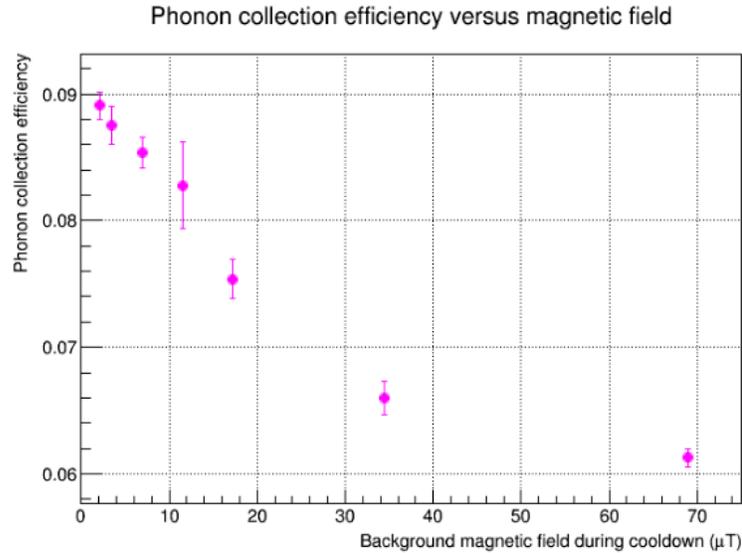
Figure 5.23: Measured reduction of PCE as a function of external magnetic field from UMN R59. Figure courtesy of M. Fritts.

factor, $s$, as the ratio of the $B$ field without the shield present to the the $B$ field with the shield present. For simplicity, we use the expression for a spherical magnetic shield [200].

$$s_{sphere} = \frac{1}{9\mu_r}\left((\mu_r + 2)(2\mu_r + 1) - \frac{2(\mu_r - 1)^2}{8r^3 + 12r^2 + 6r + 1}\right) \qquad (5.20)$$

where $\mu_r$ is the relative permeability and $r = \Delta/d$ is the ratio of the shield thickness to shield diameter. This expression gives values which are not that different than other simple geometries and is sufficient for this purpose. The form of this shielding function is shown in Fig. 5.24.

Now, the K100 fridge has a steel dewar jacket with 2 layers, each $\sim 0.25$" thick and diameter $\sim 24$" for which we use the relative permeability of carbon steel, $\mu_{r,steel} = 100$. It also has a stainless steel inner vacuum chamber (IVC) and some copper cans. However, copper has $\mu_r \sim 1$ and we assume the IVC is austentic stainless steel (most common) which also has $\mu_r \sim 1$. Apparently those structures do not add much magnetic shielding. So, using Eq. 5.20 and including 2 layers of carbon steel gives a shielding factor of $s \approx 5.3$.
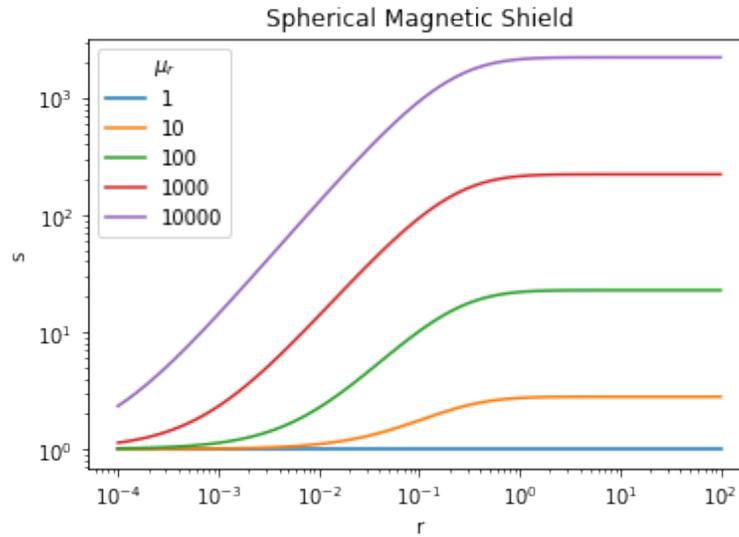
Figure 5.24: Shielding factor of spherical magnetic shield.

**Fitting and Extrapolation**

Now we can use the shielding factor and measured external fields during the mu-metal tests to calculate the field strength at the detector. Then we use the model from above to fit the data. We leave the following factors floating in the fit; $\tau_0$, $w$, $\lambda_a$, $D$, and $c$. The result of the fit is shown in Fig. 5.25. The best fit values are in Table 5.1 and are within plausible scales for these devices.

| Parameter | Value |
|-----------|-------|
| c | 186 $\mu$T $\mu$s |
| $\tau_0$ | 18 $\mu$s |
| w | 19 $\mu$m |
| $l_a$ | 3.0 pm |
| D | 0.15 $\mu$m$^2/\mu$s |

Table 5.1: Best fit parameters for the PCE model.

The fit has also been extrapolated to the region of PCE that matches observations of reduced PCE in R71 when operating the cryogenic stepper motor. The extrapolation
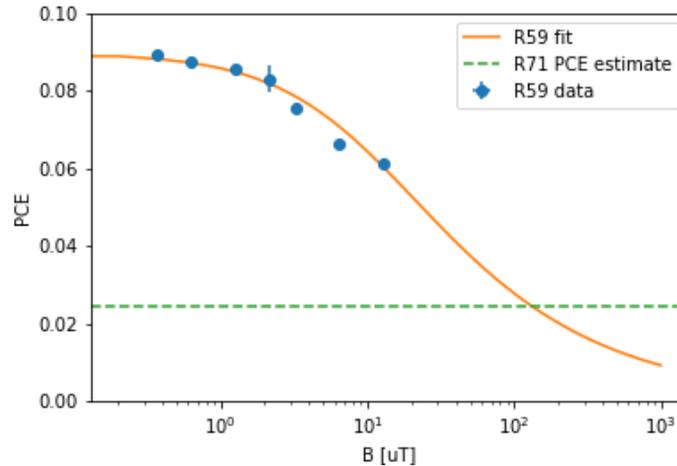
Figure 5.25: Fitted model to PCE measurements from mu-metal shield tests. Extrapolation shows that reduced PCE values observed in R71 were consistent with magnetic fields of the same scale produced by the stepper motor.

yields a value of $\sim 125$ $\mu$T for the magnetic field inside the detector housing during the stepper motor operation. This is plausible given the measured magnetic field directly adjacent to the motor housing was $\sim 400$ $\mu$T when operated. While this is a rather approximate model, it provides further support for the idea that the observed PCE loss in UMN R71 was consistent with magnetic flux from the stepper motor trapping in the Al fins after they were heated.

### 5.2.7 Position Mapping

Data was obtained with the calibration source at a range of different radial positions over the Ge iZIP detector. The interleaved charge and phonon sensors allowed the identification of events occurring within $\sim 1$ mm of the detector surface, where they suffer from reduced ionization yield [159]. These surface events were identified as having very asymmetric charge signals between side 1 and side 2 and were removed from analysis. The remaining bulk 60 keV events were used to examine the correlation of phonon signal parameters with source position.

Fig. 5.26 shows how several different types of position parameters vary with true source position. As in earlier analyses, we have employed parameters constructed using
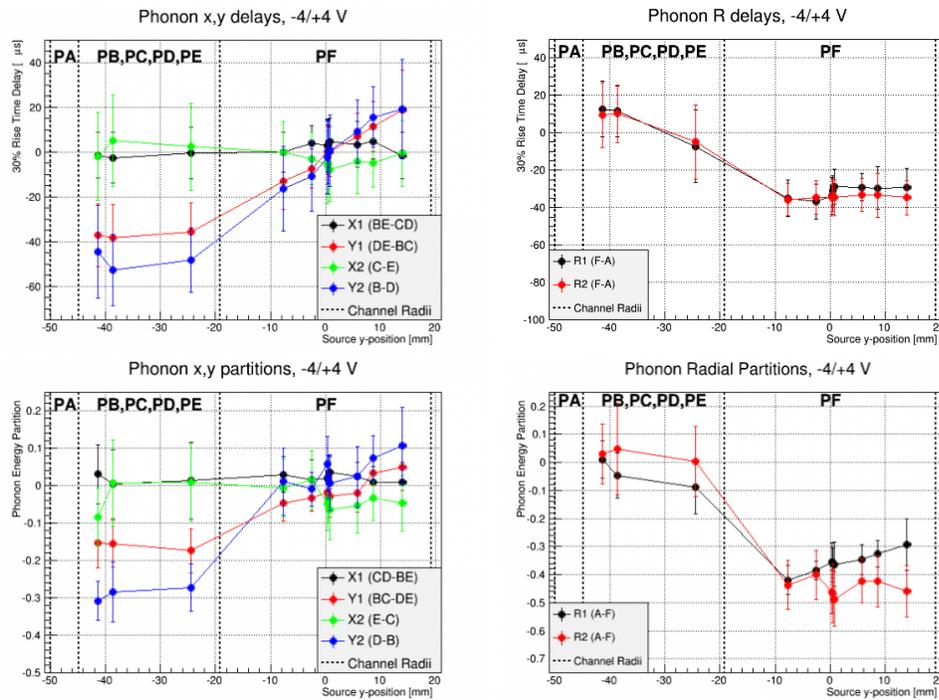
Figure 5.26: Various phonon position quantities for a range of calibration source positions. Points denote the average parameter values for bulk 60 keV events and error bars denote the width of each distribution. The radial and $y$ parameters all track the source movement while the $x$ parameters remain near 0.

the relative pulse amplitudes, combined into various partition parameters, and the relative timing between channel start times, combined into delay-based quantities. Recall, however that the iZIP channel layout shown in Fig. 5.20 is somewhat different than the HV layout used before, so the parameters are redefined accordingly as a set of delay-

and partition-based quantities in Eq. 5.21 and Eq. 5.22 respectively.

$$
\begin{aligned}
pxdel1WK &= \frac{WKr30(PBS1) + WKr30(PES1)}{2} - \\
&\qquad \frac{WKr30(PCS1) + WKr30(PDS1)}{2} \\
pydel1WK &= \frac{WKr30(PDS1) + WKr30(PES1)}{2} - \\
&\qquad \frac{WKr30(PCS1) + WKr30(PBS1)}{2} \\
prdel1WK &= -WKr30(PAS1) + WKr30(PFS1) \\
pxdel2WK &= -WKr30(PES2) + WKr30(PCS2) \\
pydel2WK &= -WKr30(PDS2) + WKr30(PBS2) \\
prdel2WK &= -WKr30(PAS2) + WKr30(PFS2)
\end{aligned}
\tag{5.21}
$$

$$
\begin{aligned}
pxpart1OF &= \frac{OF(PCS1) + OF(PDS1) - OF(PBS1) - OF(PES1)}{OF(PCS1) + OF(PDS1) + OF(PBS1) + OF(PES1)} \\
pypart1OF &= \frac{OF(PCS1) - OF(PDS1) + OF(PBS1) - OF(PES1)}{OF(PCS1) + OF(PDS1) + OF(PBS1) + OF(PES1)} \\
prpart1OF &= \frac{OF(PAS1) - OF(PFS1)}{OF(PAS1) + OF(PFS1)} \\
pxpart2OF &= \frac{OF(PES2) - OF(PCS2)}{OF(PES2) + OF(PCS2)} \\
pypart2OF &= \frac{OF(PDS2) - OF(PBS2)}{OF(PDS2) + OF(PBS2)} \\
prpart2OF &= \frac{OF(PAS2) - OF(PFS2)}{OF(PAS2) + OF(PFS2)}
\end{aligned}
\tag{5.22}
$$

The upper left plot shows the $x$- and $y$-position estimates using the relative 30% rise time delays of the four inner-ring channels, PB, PC, PD, and PE. The source was physically moved along the $y$ direction and that is generally reflected in these quantities. The $x$ parameters for both detector faces remained centered about 0, although there was some small amount of scatter in both the delay and partition parameters. The $y$ parameters both change with source position and are especially sensitive when the source is in the PF region. This is interesting because then the channels used to define

these quantities are bracketing the true source position. When the source is at the same radius as the inner ring channels, the sensitivity (slope) drops substantially and may even be flat. Interestingly, the $y$ parameter on side 2, which was the side the source was mounted on, generally exhibited larger delays between channels than did the side 1 $y$ parameter. The lower left plot of Fig. 5.26 depicts the distribution of $x$ and $y$ partitions constructed using relative OF energy estimates of the inner ring channels. They appear to show the same general trends as the corresponding delay-based parameters, although the sensitivity near the center of the detector seems to not be as high.

The right two plots show the radial parameters constructed using just the inner (PF) and outer (PA) channels on each side. The parameters in the upper right plot used the relative delay timing and the lower right plot used the OF energy partitions. In the delay parameters, we see a very similar trend between the two faces, with both relatively insensitive while the source was located over channel F, but with some reasonable slope once it moved over the inner ring channels. Again we are seeing that the parameters are most sensitive when the channels defining them bracket the true source position. The lower plot mostly shows these consistent trends with the possible exception of the side 1 quantity near the center, although this may be hard to interpret given the natural scatter of these points. In general here, we see that there is a rich set of parameter space to draw from in order to construct an optimal position mapping. One simple effort is described in Appendix D, where we develop a simple phonon model and fit it to the OF-based parameter distributions.

In the design tested, the maximum radial distance was limited by the width of the source and source holder. This prevented us from taking data with the source centered over the outer PA channel. However, it should be possible for future designs to be improved so as to allow source placement within a few mm of the crystal edge. Alternatively, while it will take more involved engineering work, the positioning of such a source scanner along the side of a detector, with its scanning path in the $z$ direction, could not only help map out the $z$-position dependence, but would also produce a population of calibration events at high radius. As discussed elsewhere, such events near the detector surfaces are the most important to study for fiducialization purposes.

Another way to study the signal propagation is by looking at the arrival time of phonon signals at the various channels. In this detector, we can use the charge signal to

determine when the event first occurred (to within $\sim 1$ $\mu$s), and study how long after that each channel first saw phonons. This was done for the same data studied above. The start time of each event was taken as the 20% rise time of the largest charge signal in the event. Then the average delay from this to the extrapolated start time of each phonon channel was calculated. An example of these distributions for a single source location is shown in the left side of Fig. 5.27. For this example, near the detector center, the PF channels first observed phonons within 5 $\mu$s, with S2, near the event, slightly sooner than S1. Then, the inner ring channels all started to receive phonons from 10–20 $\mu$s, and finally the outer PA channels at 25 $\mu$s. To study these delay distributions, we plot the mean of each distribution as a function of the radial distance from the source to the nearest edge of each phonon channel, i.e. the horizontal distance from the NTL phonon column to the channel edge. In a way, this assumes that the phonon signal starts to rise as soon as the first wave of phonons reaches the edge of the channel.
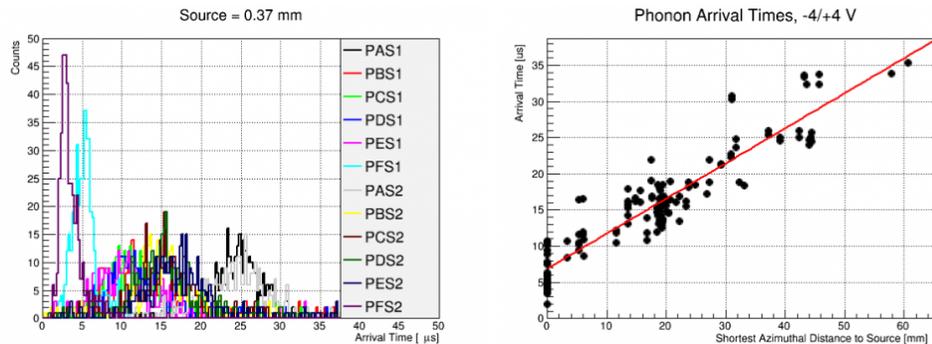


Figure 5.27: Left: Example of phonon arrival times relative to earliest charge signal where the PF channels are the physically nearest and have the earliest arrival times. Right: Arrival time is correlated with azimuthal distance from channel edge to source. The red fit line corresponds to a slope of $2.1 \pm 0.1$ km/s.

The slope of the distribution on the right side of Fig. 5.27 corresponds to $2.1 \pm 0.1$ km/s. Naively this is something like the phonon propagation speed, however the real story is more complicated. Ballistic phonons travel in straight lines at the speed of sound in Ge, which is 3–6 km/s [145]. But, we know that the phonons are generally first produced at higher energies and move diffusively before down-scattering to the ballistic regime [201]. There are then additional delays from the quasiparticle diffusion and TES response time. Thus the time delay observed here represents some effective combination

of these processes. In the iZIP detector design, this story is further complicated by the distribution of phonon production, with half of the NTL phonons being produced in the interleaved regions of the detector within a few mm of each face. Modelling all of these effects is beyond the scope of this work, but is mentioned to highlight how such datasets may be valuable to validation efforts of detector simulations and study models of charge and phonon transport.

### 5.2.8 Next Steps

For the particular design tested in these studies, solving the coil heating issues will enable faster source translation, though there will still be limits due to frictional heating. One can also trade longer pitch thread for increased speed at the loss of position accuracy. The inclusion of a small mu-metal shield around the motor could fix the problems associated with stray magnetic fields. Future modifications are planned in order to add an angular degree of freedom with a secondary motor to enable a full two-dimensional scan of the detector face. This will allow fuller mapping of both the radial and angular parameter space. While it would take more significant engineering efforts, including machining a vertical slot in an existing detector housing, it would be interesting to scan a source along a detector sidewall as well. A more minimal design of this type could be employed for in-run energy calibration. This could be as simple as a motor which moves a shutter to as to expose or block a source in a fixed position. Such a device could also be employed to actuate a cryogenic shutter for fiber optics currently used to calibrate HVeV detectors and which may suffer from IR leakage from higher temperature stages.

# Chapter 6

# Scaling Up - 150 mm Devices

## 6.1 Motivation

Perhaps the greatest weakness of using semiconductors as DM detectors is scalability. In practice it is difficult to grow pure monocrystaline semiconductors of the purity required above the kg-scale. An additional challenge is that semiconductor detectors must be operated at cryogenic temperatures, so large devices can also require significant cryostat design changes. This is in contrast to liquid noble experiments which are now reaching the ton-scale because, with less difficulty, one can always just "build a bigger bucket" to increase the total detector volume. Because of these limitations, experiments like SuperCDMS deploy arrays of many individual detectors, each of which must be instrumented separately, a further expense. It is thus desirable to understand how much larger individual detectors can be made. This chapter describes efforts to maximize the individual crystal volume while maintaining desirable detector operation characteristics.

As discussed in Ch. 4, leakage currents through the detector can reduce its sensitivity when operated at high bias voltages. If this leakage is occurring at the electrode-crystal interface, one possible solution is the use of contact-free, or vacuum-separated, electrodes. It has been demonstrated that such arrangements can reduce this leakage for small (56 cm$^3$) Ge devices [185]. The work in this chapter also demonstrates the feasibility of such contact-free bias schemes with larger (450–580 cm$^3$) Si crystals. Although these tests did not include phonon readout, they were the first demonstration of such bias schemes used with the largest cryogenic Si particle detectors yet operated

at temperatures below 100 mK.

## 6.2   Detector Designs

For these tests, two ionization detectors were fabricated out of n-type, [100] orientation, high purity, high resistivity (>15 kΩ-cm) Si crystals purchased from TopSil [202]. The two cylindrical crystals were both 150 mm in diameter. One was 25 mm thick, and the other was 33 mm thick. There are also currently 200 mm and 300 mm Si boules in use in the semiconductor industry. (There is even a proposed 450 mm Si standard) However, these tests used 150 mm boules due to limitations in the processing equipment used to polish and lithograph them. In addition, this size was pushing up against size restrictions in our cryogenic fridge which could not be overcome without significant engineering efforts.

The 33 mm thick crystal was used to construct the first, simple single-channel contact-free detector by mounting the bare crystal between two planar aluminum electrodes. This detector is denoted "S1501". As shown in the left side of Fig. 6.1, the ionization signal was measured on one electrode ("Readout") while the other was used to provide a voltage bias ("Bias"). The readout electrode could also be biased up to ±12 V, but most often the readout electrode was grounded and the bias electrode provided a 0–100 V bias. This device was tested in the fall of 2016.

The thinner, 25 mm thick crystal, denoted "S1502", had five concentric electrodes of equal areas patterned on one crystal surface as shown in Fig. 6.2. The electrodes were fabricated on one crystal face by depositing a layer of amorphous Si followed by a layer of aluminum. The amorphous Si layer has been part of the standard lithography program in SuperCDMS detectors (both Si and Ge) and so was included here. (Other, more recent studies are investigating how this layer may actually be a source of charge leakage.) The opposite crystal face of S1502 was left bare. The five channel geometries were then defined by wet-etching. The same contact-free electrode used with S1501 was mounted near the bare face to provide a bias voltage and the induced ionization signals were read out from the five deposited electrode channels. This device was tested in the fall of 2018.

The aluminum electrodes feature a 150 mm outer diameter top lip, used to secure
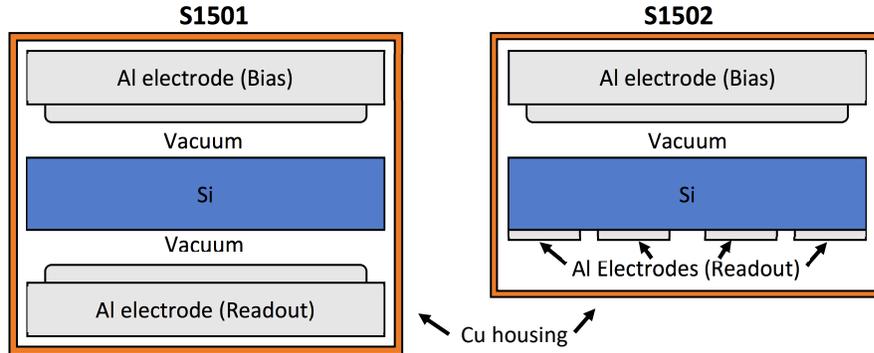
Figure 6.1: Left: S1501 detector design with two aluminum electrodes (gray) surrounding a 150 mm Si crystal (blue) with a vacuum gap between faces. All pieces are secured within a copper housing with Cirlex clamps (not pictured). Right: S1502 detector design with a single vacuum gap for biasing and a set of deposited readout electrodes on the crystal surface.

the electrode to its housing, and a solid 142 mm diameter cylinder which extended below the bottom edge of the lip. This cylindrical portion was designed to mate with mounting clamps in a dedicated housing. The detectors were mounted and enclosed in housings machined from OFHC copper. The housings were basically designed as scaled-up versions of the detector housings used in SuperCDMS Soudan. Both the crystals and the electrodes were held in place, electrically isolated from the housing using Cirlex clamps. Photos of the crystal and electrodes mounted in the housings can be seen in Fig. 6.3.

Each electrode was mounted with its flat face parallel to the crystal surface, with a gap of $\lesssim 1$ mm. As shown in Fig. 6.4, two shallow recesses along the radial periphery of the electrode face provided space for small circuit boards. Infrared (IR) and ultraviolet (UV) Light Emitting Diodes (LEDs) mounted on each board were used to periodically illuminate the detectors between data taking. The effect of these is discussed further in Sec. 6.4.4.

The bias voltage across the crystal, $V_{cr}$, was not the same as that applied to the detector as a whole, $V_{tot}$ due to the voltage drop across the vacuum gap(s). These quantities can be related by

$$V_{cr} = \frac{h}{h + \kappa d} V_{tot} \tag{6.1}$$
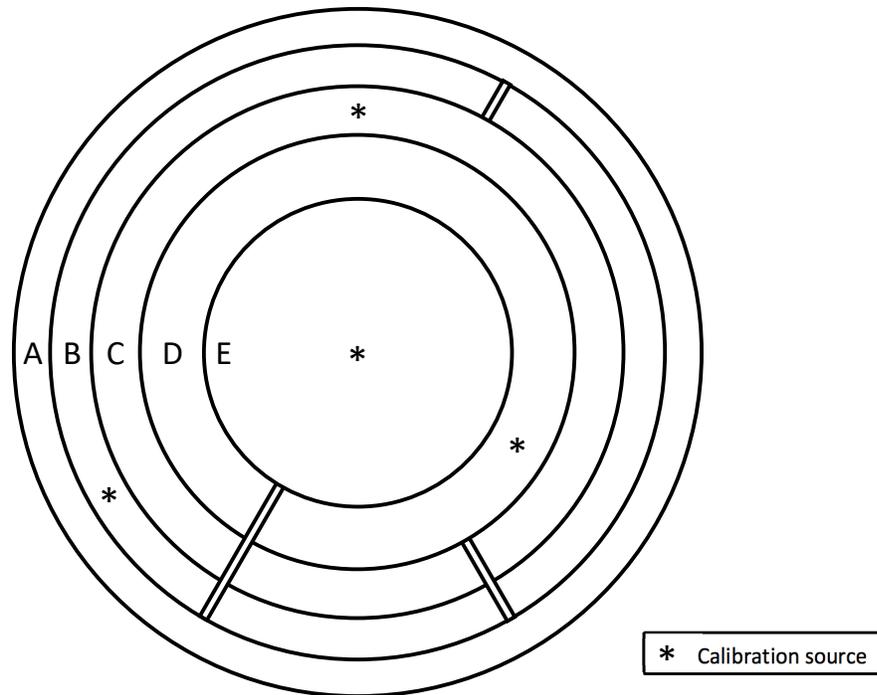
Figure 6.2: S1502 ionization electrode channel map. Locations of the $^{241}$Am calibration sources are marked. The segments of channels B and C were connected with wirebonds so that each forms a continuous annular electrode. Sections of the inner electrodes extend towards the detector radius to allow wirebond connections to readout electronics. The S1501 detector (not pictured) had a single, monolithic electrode with a single calibration source at the center.

where $h$ is the detector thickness, $d$ is the total vacuum gap width, and $\kappa = 11.47$ is the relative permittivity of Si near 0 K [203]. This expression does not include any fringing fields or effects of the grounded detector housing, which causes a non-uniform field at high radius.

The vacuum gaps of S1501 were measured to be $0.29 \pm 0.03$ mm and $0.40 \pm 0.07$ mm for the readout and bias sides respectively. These gaps resulted in $80\% \pm 2\%$ of the total bias field being applied across the crystal itself. They also set the total detector capacitance at $\sim$70 pF, similar to typical SuperCDMS detector capacitances of $\sim$100 pF [204]. S1502 had a single, larger gap of $0.9 \pm 0.1$ mm, giving $70\% \pm 3\%$ of applied bias across the crystal. This device used a thinner crystal and split the measurement electrode into 5 equal-area segments, so the capacitance of each of the five
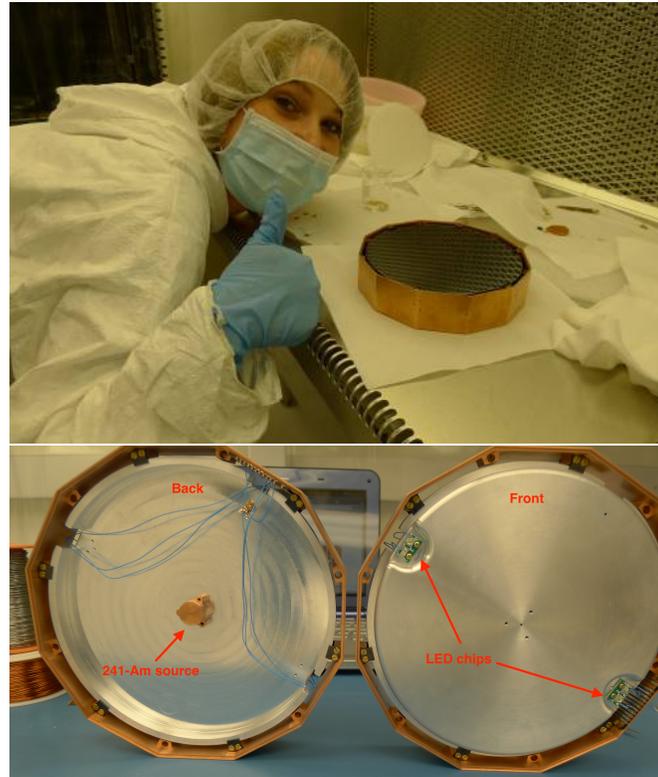
Figure 6.3: Top: A. Kennedy, proud after successfully mounting the S1501 crystal in a housing. Bottom: Contact-free electrodes in housings ready for assembly with S1501.

readout electrodes to the bias electrode was ~20 pF. Direct measurement of the vacuum gaps was difficult and is the main source of uncertainty in the magnitude of $V_{cr}$. As described below, it also lead to an uncertainty in the absolute amplitude of measured charge signals.

These devices were tested in the UMN K100 fridge with a base operating temperature ranging from 75–95 mK. The detector housings were designed to fasten to a conventional CDMS-Soudan "tower" [182], thus minimizing the amount of additional cold hardware to be fabricated for these tests. The K100 fridge has been able to achieve temperatures as low as 30 mK with smaller detector payloads. Unfortunately, due to the detector radius, we were unable to install a crucial component which normally heatsunk part of the tower to the Still ($\sim 1$ K) stage of the fridge. This resulted in the non-optimal operating temperature. However, these devices had only charge channels, which should
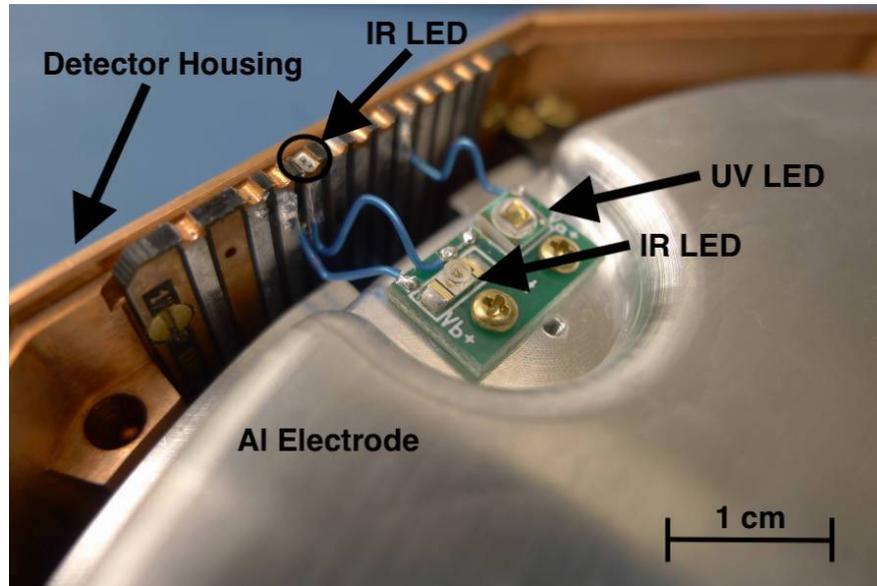
Figure 6.4: Photo of LED board mounted on electrode. Position of the IR LED mounted on the detector housing is also indicated.

operate without issue to temperatures above 1 K.

The detectors were connected to the tower wiring via different sets of modular coaxial wiring. For S1501, the single channel detector, simple flexible coaxial cables were used. These were included as a test to replace the standard rigid vacuum coaxial connectors used in SuperCDMS Soudan. The hope was that the flexible coax would allow us to operate different types of detectors with greater freedom of housing design and connector placement. Unfortunately, we found that it induced undesirable oscillations in the amplifier chain which distorted the pulse shapes. The design of the electronics used with the subsequent detector, S1502, was modified to as to accommodate the old vacuum side coax assemblies, and the pulse shapes were improved as discussed below.

The standard SuperCDMS cold electronics described in Sec. 2.2 were used to amplify the ionization signal. The LEDs and readout amplifiers were controlled and digitized with a set of prototype SuperCDMS Detector Control and Readout Cards (DCRCs) [164] connected to a computer running MIDAS-based data acquisition software [165]. The bias electrode was connected to an external power supply via a custom designed adapter board. With S1501, we used the HV board described in Appendix C, which

was meant for operating iZIP detectors in CDMSlite mode. With the S1502 detector, we needed the use of 3 DCRCs to read out the 5 charge channels. So we instead built custom D-Sub 50 boards to break out the appropriate lines for HV biasing.

## 6.3    Measured Charge Pulses

Analysis of these devices focussed mainly on the signals from the 60 keV gamma line from $^{214}$Am calibration sources. In cryogenic Si, it takes an average of 3.8 eV to produce an e/h pair [130], so these interactions each produced around $1.6 \times 10^4$ pairs where the gamma was absorbed. With a constant bias field applied by the electrodes, these charges drift across the crystal bulk, inducing a signal on the readout electrode [135], the magnitude of which can be predicted using the Shockley-Ramo Theorem [205, 206] including polarization effects of the crystal and gap dielectrics [207]. To first order, the expected signal at the readout electrode, $Q_{sig}$ is related to the total amount of charge liberated in the detector, $Q$ by

$$Q_{sig} = \frac{h}{h + \kappa d} Q \tag{6.2}$$

i.e. the signal is reduced by the same factor as the applied bias voltage in Eq. 6.1. This also leaves it subject to the same systematic uncertainties in $d$ that were discussed above.

The experimental flexible coaxial cabling used with S1501 introduced a high frequency pole which caused ringing in the pulse tail as can be seen in Fig. 6.5. This behavior was studied in detail and was reproduced on the bench at room temperature. The only clear difference we observed in the properties of the two designs was a $5\times$ higher parasitic capacitance between the signal wires and ground on the flexible coax. However, even including this in a full SPICE circuit simulation did not reproduce the observed results. We were ultimately unable to definitively determine the cause of the oscillation which prevented an absolute calibration of the S1501 signal from first principles. However, the relative event energy information was still contained in the maximum amplitude of the pulse. The whole pulse shape, including the ringing tail scaled linearly with input signal. This allowed each pulse to be fit with an average template constructed from characteristic events, to estimate the relative amplitude of

each event.

The S1502 pulses were all fit using the Optimal Filter (OF) algorithm. However, the S1501 data was also plagued with a peculiar form of readout glitch related to the DCRCs which occurred randomly and swapped adjacent sets of 8 ADC samples. At time scales above 3.2 $\mu$s this was not apparent, but it introduced many strange artifacts in the frequency domain which ruined the OF resolution. As an alternative, this data was simply fit in the time domain with a smoothed template, which was relatively insensitive to such issues.
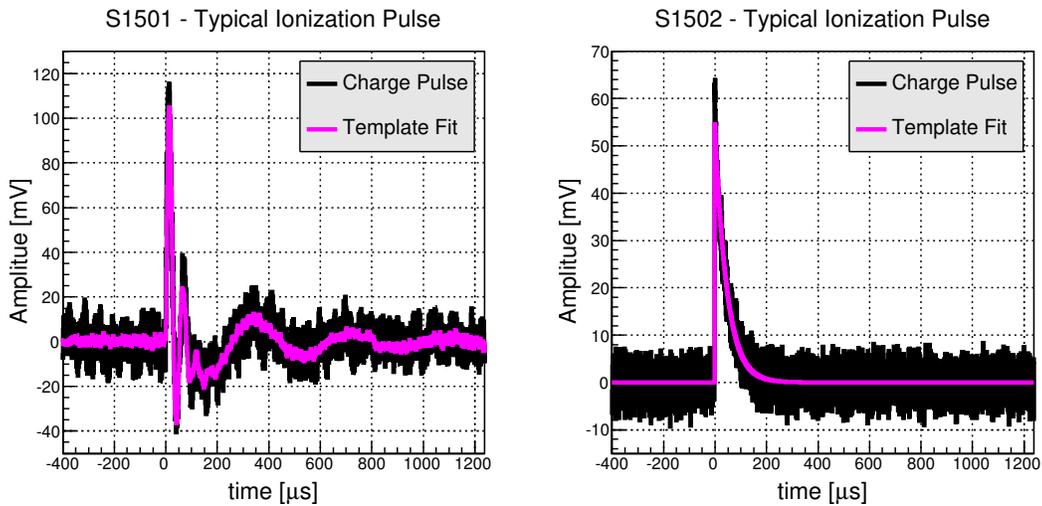


Figure 6.5: Typical ionization pulse signals shown in black, with fitted pulse templates in magenta. Left: S1501 detector pulse with distortion is still well-fit by average template. Right: S1502 channel E pulse shows effect of corrected readout circuit with simple exponential fall time.

### 6.3.1 Absolute Charge Calibration

With improved cold electronics, S1502 produced exponential pulses without ringing. This allowed an absolute calibration for all 5 channels based solely on the readout circuit design. As described in Sec. 2.2.2, the amplitude of the exponential charge signal voltage, $V_{out}$, is related to the induced charge signal as

$$V_{out} = \frac{Q_{sig}}{C_{FB}} \frac{C_C}{C_C + C_{stray}} \tag{6.3}$$

where $C_C = 300$ pF is the coupling capacitor and $C_{stray}$ is any other capacitance to ground. The feedback capacitance $C_{FB}$ is actually dominated by the capacitance of the feedback resistor itself, not because of any separate, discrete capacitor circuit element. The value of $C_{FB}$ can be inferred from the observed pulse fall time which is set by the first stage amplifier feedback branch, $\tau_F = R_{FB}C_{FB} = 46.2$ $\mu$s. Given a measured value of $R_{FB} = 40.2$ M$\Omega$, the average capacitance was found to be $C_{FB} = 1.15$ pF.

Obtaining $C_{stray}$ was a more nuanced task. Ideally, one would simply measure the capacitance between each readout electrode (A–E) and ground. However, the presence of the bias electrode complicates matters because it also provides an indirect capacitive coupling which we do not wish to include in Eq. 6.3. A model of the full network of elements and capacitances is shown in Fig. 6.6. This model includes the capacitance from each readout electrode to ground, the bias electrode, and to neighboring readout electrodes. There, $Cab$ represents the capacitance between elements $a$ and $b$ with $a, b \in \{L = Bias\ Electrode,\ G = Ground,\ A = Readout\ A,\ ...\ ,\ E = Readout\ E\}$.
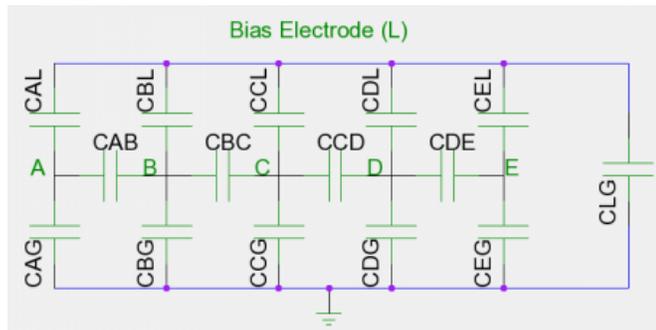


Figure 6.6: Capacitance network model for S1502 electrodes and housing.

To characterize all these capacitances, multiple bench measurements were performed on the detector and electrode stack with an HP4275A LRC meter. To isolate the different values, measurements were done with different combinations of elements shorted together, e.g. shorting channels $A$ and $B$ eliminates capacitance $CAB$. By combining many such measurement with the above capacitance matrix model, we were able to infer an average channel to ground capacitance of, $C_{Chan,G} = 25.5$ pF. This value represents the direct capacitance between a readout channel and ground. However, to first order each channel also sees ground via its neighboring channels, so these should

also be included. This indirect capacitance was measured to be, on average $70 \pm 20$ pF, where the uncertainty is set by the standard deviation of the measurements obtained for the individual channels. Note that this indirect capacitance also includes the effect of neighboring channels' coupling capacitor, since their amplifiers hold the other side of it at virtual ground. In this way, the indirect capacitance to ground can actually be larger than the direct capacitance to ground. Now, combining these effects gives an average effective capacitive signal gain factor of

$$G_{cap} = \frac{C_C}{C_C + C_{stray}} = 0.81 \pm 0.04 \tag{6.4}$$

A more careful analysis could be done which allows for variations in these factors due to the differing geometries of individual channels. However, that is beyond the scope of this work since the capacitance measurements and calculations involve significant systematic uncertainties that we are simply averaging over.

Similar analysis can be done to understand an observed $\sim 5\%$ cross-talk of signals between neighboring channels. In this process, a channel which observes an event induces small signals on the neighboring channels via capacitive coupling between the electrodes. While this coupling can explain a significant amount of the cross-talk observed, some of it will also be from overlapping Ramo potentials, which requires dedicated charge transport modelling to understand properly. However, this cross-talk is not expected to reduce the signal in the primary channel, and is thus not considered further in this analysis. Cuts were made in the datasets to ensure we only considered events which were not shared between two channels.

After the first stage of amplification, the DCRCs had a few stages of further amplification and filtering before the final voltage was digitized. These gains were different depending on the DCRC version. While these values were known theoretically from the amplifier circuit diagrams, we confirmed them by injecting fake signals on the bench. We showed that the absolute fixed change amplifier gains for the two revisions (Rev) of DCRC boards used were 4(5.5) for Rev C (Rev C.1). Putting all pieces together allowed us to calibrate from the fitted pulse heights back to charge induced at the electrode. This calibration allowed measurements of the absolute charge collection efficiency for detector S1502.

## 6.4 Experimental Results

Because this was such a novel detector, data was taken under a variety of conditions. This included a range of bias voltages, neutralization with IR and UV LEDs, and illumination with $^{241}$Am and $^{57}$Co gamma sources. The 60 keV gamma line from $^{241}$Am sources was used for energy calibration. In S1501, the source was mounted at the center of the readout electrode with an 0.008" diameter hole in a 0.375" thick lead disc collimating the source along the crystal axis. In S1502, sources were mounted on the bias electrode over the center of each channel except the outermost (channel A) as shown in Fig. 6.2. Each source was collimated with 0.375" thick lead discs with hole diameters of 0.018" for the center channel (E) and 0.006" for the other three (B–D). The center channel source had a different geometry and required a larger collimator hole to achieve a similar event rate to the others. The calibration events occurred more often on the side of the detector nearest the source, but because the mean free path of 60 keV gammas in Si is $\sim$3 cm [188], the calibration event rate only falls by about half over the depth of the crystal.

### 6.4.1 Gamma Simulations

We developed *Geant4* [189] simulations of these detectors to understand the distribution of energy depositions from the $^{241}$Am as well as from an external $^{57}$Co source. The penetration depth of characteristic gamma energies (60 keV for $^{241}$Am and 120 keV for $^{57}$Co) in both Si and Ge can be seen in Fig. 6.7. This simulation threw the gammas at a crystal with the same geometry as S1501. The gammas were thrown in an axial beam at the center of the detector ($z$ axis). The histograms show the energy-weighted positions of gammas which did not scatter out of the crystal, but deposited their full energy. The energy-weighted position (defined as $\vec{x}_{EW} = \frac{1}{E}\sum_{hit\ i} E_{dep,i}\vec{x}_i$) is used because the gamma may scatter several times in the crystal. The results showed that, in Ge detectors, the 60 keV gammas deposit most of their energy in the first few mm of detector material. This can exploited to compare the charge transport properties of electrons and holes. However, these events are deposited much more uniformly in Si, making it difficult to disentangle the charge carrier behaviors with these devices.

Because the sources were on opposite electrodes in the two 150 mm devices tested,
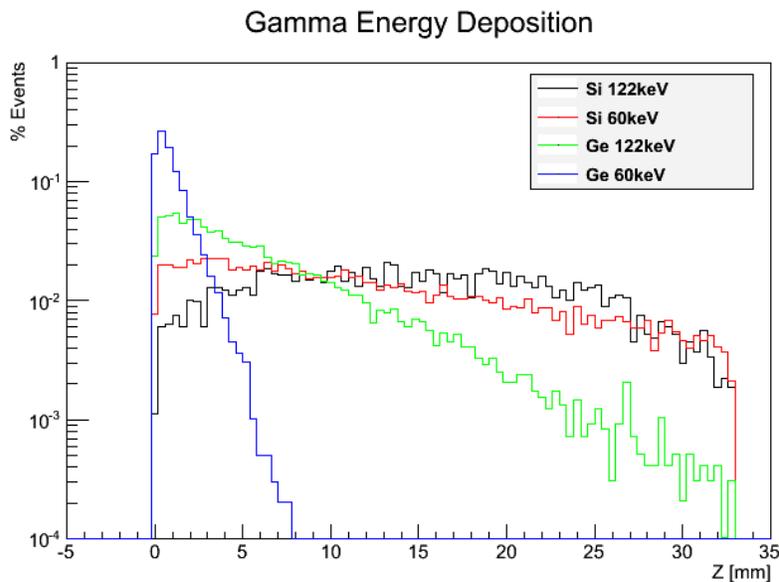
Figure 6.7: Energy deposition depth for Ge and Si. Almost all 60 keV events in Ge occur near the surface, while others penetrate farther.

the correspondence between bias field sign and dominant charge carrier is opposite. In S1501 (S1502), the calibration sources were on the readout (bias) electrode, so positive bias field configurations preferentially involved electrons (holes) traversing most of the crystal thickness and thus generating the bulk of the induced signal. However, as explained above, this is a weak effect due to the relative uniformity of calibration event depths.

At one point, a $^{57}$Co source was placed outside the fridge to illuminate detector S1501 from the radial direction. This was simulated using a full model of the K100 fridge geometry to understand where the 120 keV gammas would deposit energy. The results of this are shown in Fig. 6.8. It was found that there was a significant asymmetry in the distribution of event positions for gammas depositing over 100 keV. This helped us to understand other effects such as how the electric field evolved over the duration of detector operation.
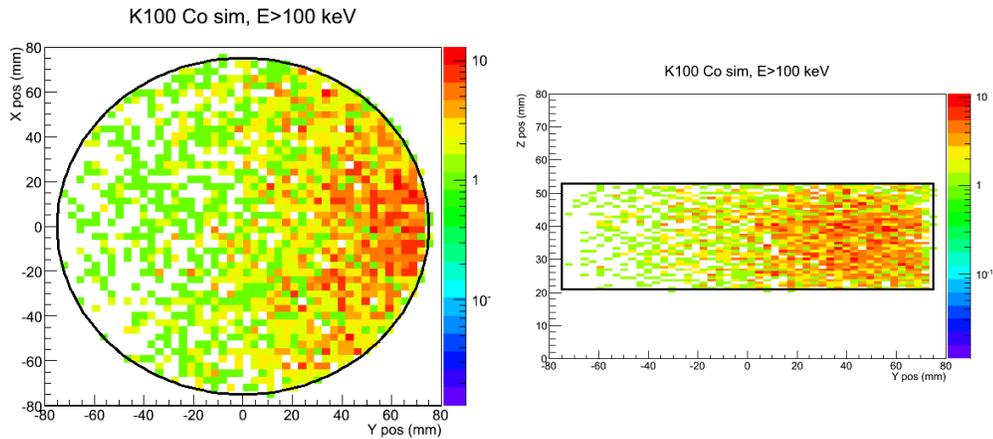
Figure 6.8: Simulated energy deposition locations of 120 keV gammas from external $^{57}$Co source in S1501. Simulation included the full K100 fridge geometry.

### 6.4.2 Charge Collection Efficiency

Both devices were used to study the amount of charge signal as a function of bias voltage, a critical aspect of detector operation. When the crystal bias is low, some of the e/h pairs produced when the gamma is absorbed may recombine immediately before the bias field drifts them apart. Additionally, as they are drifted to the electrode, they may become trapped on crystal impurities before traversing the entire crystal. Both processes reduce the amount of drifting charge and thus reduce the induced charge signal.

The charge collection efficiency (CCE) is defined as the ratio of measured charge induced on the readout electrode to that expected for a given event assuming no carriers are lost to these processes. As the bias voltage is increased, carriers are less likely to trap or recombine leading to increased CCE and thus larger signal amplitudes for a given interaction energy. This is demonstrated in Fig. 6.9 where the measured signal from a calibration peak is observed to shift with changing bias voltage.

For both detectors, the average amplitude of the 60 keV calibration line of $^{241}$Am was used to calculate the CCE as a function of bias field. Well-reconstructed events were selected from the data by making cuts on quantities like reconstructed energy, $\chi^2$ fit value to the templates shown in Fig. 6.5, and timing relative to the DAQ trigger. In the multichannel detector, S1502, we further required that there be a pulse in only a
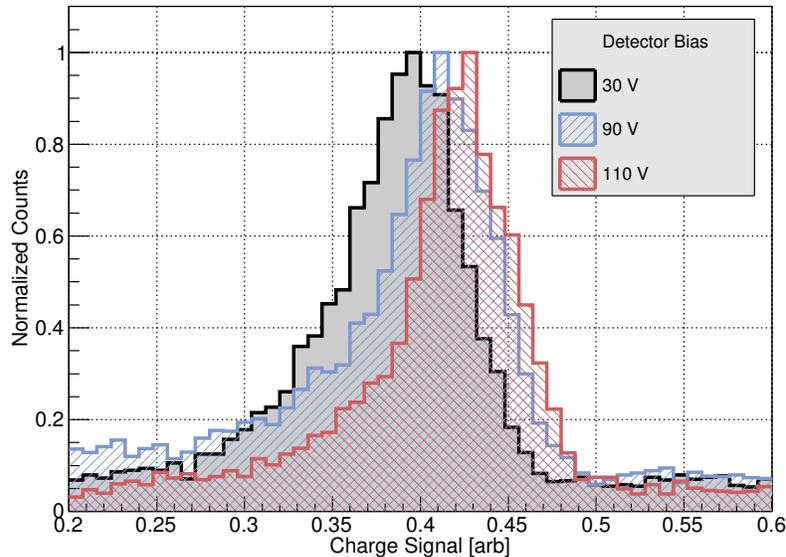
Figure 6.9: Shift of the 60 keV spectral line as measured by S1502 at various applied bias voltages. Fewer charges are collected at lower bias voltages because of trapping and recombination leading to lower CCE.

single channel and that the other channel pulse amplitudes were consistent with noise. This ensured that we only considered events in which all carriers were drifted to a single channel and helped reduce some of the effects of cross-talk. As discussed later (Sec. 6.4.4), the detector behaviour would change over time after several minutes of operation due to charge accumulating on the bare crystal surfaces. To account for this, the data used for CCE calculations was only taken from the first ∼2–10 minutes of data collected after applying the bias field.

The results of these measurements are shown in Fig. 6.10. The left plot of Fig. 6.10 shows the CCE curve for the single channel detector, S1501. As discussed above, electronics issues prevented an absolute calibration of this device, so the measured points were normalized by setting the highest measured CCE to unity. The plot on the right shows the results from the multichannel, S1502 device. Due to space restrictions, there was no calibration source on channel A, so it does not have a CCE curve. An important difference for this detector is that the CCE scale has been set using the absolute

calibration. The dominant uncertainties in that calculation were the relative size of the vacuum gap, $d$, and the capacitive gain factor. This systematic uncertainty scaled all channels by the same relative amount, so its magnitude was indicated with the orange band around CCE = 1 as opposed to with error bars on every point. For comparison, similar measurements from a much smaller (1 cm×1 cm×4.8 mm) device are shown on the same axes [208]. The $x$ axis on each plot is the bias electric field in the crystals, calculated using Eq. 6.1. Though not shown here, this quantity also suffers from the same systematic error caused by $d$.
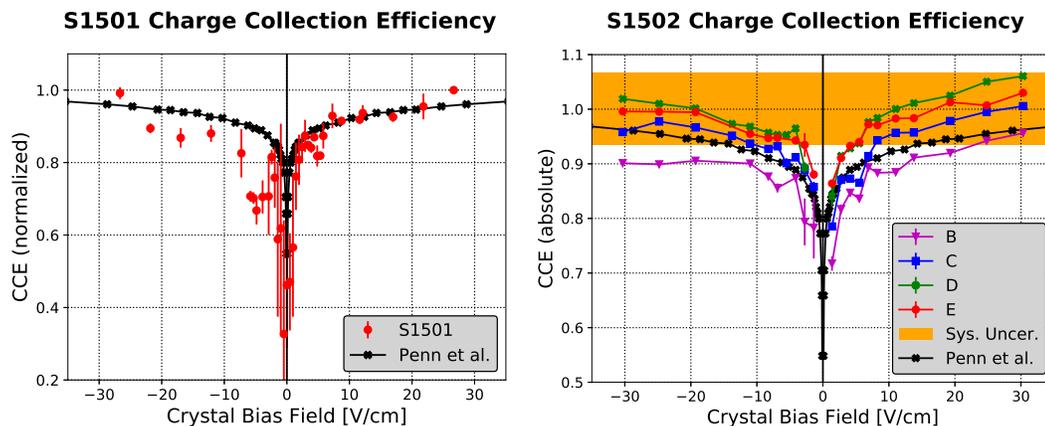


Figure 6.10: Charge collection efficiency as a function of applied bias. Left: S1501 relative charge collection efficiency, normalized to 1 at maximum CCE. Right: S1502 charge collection efficiency curves showing lower efficiency at higher radii. The systematic uncertainty of the absolute charge collection efficiency scale is shown as an orange band. Results from a smaller device shown for comparison [208].

These CCE curves all demonstrated the same qualitative behavior as a function of bias voltage. At high fields, the CCE is 100% within uncertainty. As the field decreases to $\sim$ 10 V/cm, the CCE decreases because carriers are lost to trapping and recombination. With no bias field, the carriers do not drift and no ionization signal is produced.

The S1502 measurements appear to show a modest decrease in CCE with radius. The inner channels, D and E, yield systematically higher CCE curves than the higher radius channels C and B. This is likely due to either an increase in charge recombination/trapping or a decrease in induced ionization signal due to a decrease in field

strength at high radius. Corrections of ∼5% near the crystal edge are not unexpected given the detector geometry [209]. To understand these effects in detail, more work needs to be done to both model the field and Ramo potentials throughout the whole detector, but to also examine the expected charge transport.

A simple model of carrier recombination and trapping to describe the shape of such CCE curves as a function of bias and crystal thickness was presented in [208]. The model describes diffusion of the electrons and holes in the initial charge cloud with a characteristic diffusion length scale, $L$. It assumes that all carriers which diffuse less than a field-dependent radius, parameterized as $bF^{-n}$, will recombine before the bias field, $F$, erodes the cloud. Those carriers will not contribute to the measured signal. The population of carriers which escape the initial charge cloud are then drifted across the crystal. These carriers have a chance to be trapped before reaching the electrode. This likelihood is characterized by trapping length parameterized as, $\lambda_t = cF^m$. This length scale describes the loss of carrier population, $N$, as a function of distance drifted, $z$ as , $dN/dz = -N/\lambda_t$. This assumes that the trap density is uniformly distributed in the crystal. Combining these effects with the signal reduction due to the vacuum gaps gives the following expression for charge collection efficiency as a function of field and detector thickness,

$$
CCE = \frac{2}{h + \kappa d} \left( 1 + \frac{b}{L} F^{-n} \right) \exp \left[ -\frac{b}{L} F^{-n} \right]
$$
$$
\times \left\{ bF^{-n} + cF^m \left( 1 - \exp \left[ -\frac{h - 2bF^{-n}}{2cF^m} \right] \right) \right\} \quad (6.5)
$$

The model was used to simultaneously fit the data presented here and that for smaller devices of similar resistivity presented in [208]. The resistivity is related to trap density, so these devices are expected to exhibit similar microscopic behaviours for carriers. For this portion of the analysis, the CCE data has been renormalized to match an extrapolated maximum CCE=1. The resulting fit curves for each thickness are shown in Fig. 6.11 and the corresponding best-fit parameters can be found in Table 6.1.

These fits indicate that, for devices less than a few cm thick, charge collection efficiency is dominated by trapping in the initial ∼0.7 mm charge cloud, before the carriers have drifted a significant distance from the initial event location. After the initial charge cloud, the length scale for trapping while drifting is ∼3 cm and increases
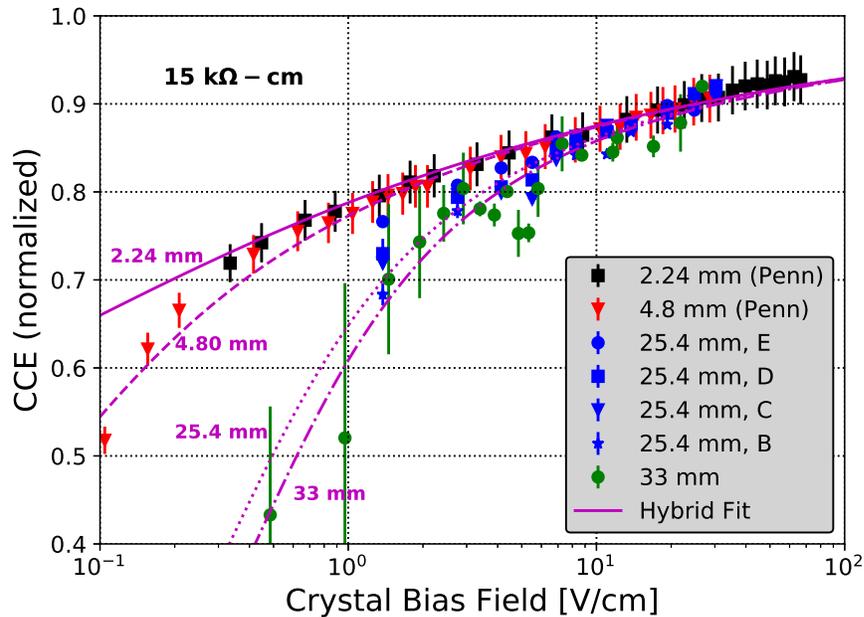
Figure 6.11: Fitting the trapping model of Ref. [208] to combined CCE data.

approximately linearly with field strength (m=1.13). This trapping length may become a limiting effect for crystals much thicker than ∼3 cm. Not only does the trapping exponentially reduce the measured signal (roughy as $\exp[-L/\lambda_t]$), but a thicker detector requires a higher bias voltage to achieve the same electric field ($F = V/L$). Of course the trapping length can also be increased by obtaining large crystals with lower impurity and lattice defect concentrations.

### 6.4.3 Accumulating Counter-Bias

As the detectors were operated, drifted charges accumulated at the bare detector faces. This accumulating charge caused a "counter-bias" field to develop over time, reducing the effective bias field in the crystal bulk and lowering the charge collection efficiency. When this occurred, the crystal was said to have lost "neutralization" and needed to be reset. An example of this behavior is shown in Fig. 6.12. After an initial period of stable operation, the 60 keV peak was observed to fall after several minutes as the electric field in the crystal was cancelled. As the counter-bias field grows, the effective bias field in

| Parameter | Description | Best Fit |
|-----------|-------------|----------|
| b | Max cloud radius at E=0 | $0.065 \pm 0.016$ cm |
| b/L | Ratio of b to diffusion length scale | $0.851 \pm 0.017$ |
| n | Max cloud radius inverse field exponent | $0.146 \pm 0.008$ |
| c | Average capture length at E=0 | $2.76 \pm 0.171$ cm |
| m | Capture length field exponent | $1.13 \pm 0.081$ |

Table 6.1: Best-fit parameters of hybrid trapping model to combined data set. Model and parameters are discussed in depth in Ref. [208].

the crystal falls and thus so does the charge collection efficiency, following the trends shown in Fig. 6.10. When the effective bias field fell to ∼5 V/cm, the measured signals began to rapidly decrease.

While investigating the distribution of accumulated surface charge in S1501, we compared the relative hold times of the spectral lines of internally mounted $^{241}$Am source and an external $^{57}$Co gamma source. We found that the external source line, which was preferentially illuminating the high-radius regions of the detector, was generally lost before the internal source, which was collimated down the detector's axis. This may simply be due to the lower electric field in the crystal at high radius due to fringing fields. This effect was expected to be more prominent for S1501 because the higher gap to crystal diameter ratio makes it less like the ideal parallel plate capacitor. However, it might also be an effect of the increased rate of charge production, and thus counter-bias accumulation, at high radius where the $^{57}$Co events were occurring. Because this device did not have multiple segmented electrodes, it was not possible to directly deduce the locations of individual events in the crystal.

With S1502, not only the 60 keV calibration peaks, but the entire measured energy spectrum was shown to fall at the same time and rate simultaneously in all channels. This indicated that the counter-bias was accumulating quasi-uniformly on the bare crystal face. One reason could be that the dominant source of charge was the spatially uniform background events, not the calibration sources (which were localized to one area of each readout channel). An alternative was that the accumulating charge at the bare crystal face naturally spread out evenly across the surface due to Coulomb repulsion.

In either case, the result is similar, a uniform surface charge density which increases as charges accumulate over time. This story was encapsulated in a simple model of

## S1502 Calibration Peak Decay
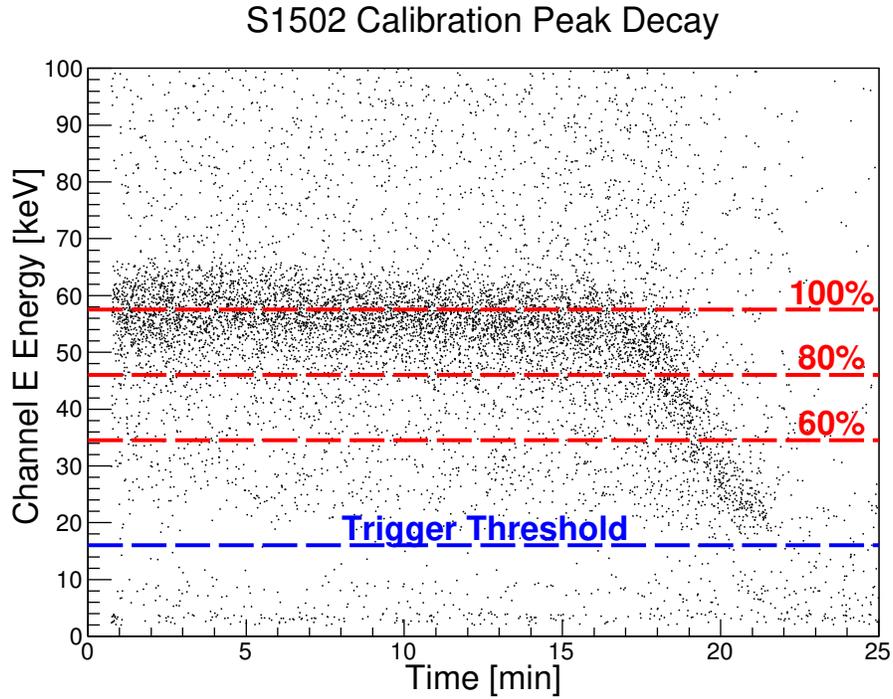


Figure 6.12: An example of the measured signal of a calibration peak in S1502 falling over time as liberated charge accumulated on the bare crystal face field.

counter-bias accumulation. In this model, the bias across the crystal should be

$$V_{cr} = \frac{h}{h + \kappa d}\left(V_{tot} - \frac{\sigma d}{\epsilon_0}\right) \tag{6.6}$$

where $\sigma$ is the accumulated surface charge density at the bare crystal face(s). This expression holds for both single and double vacuum interface devices (but with $d$ the sum of the two gaps), assumes uniform $\sigma$, and neglects fringing fields. The accumulated surface charge provides the counter-bias, $V_{CB}$, which is defined as the effective reduction in the total applied detector bias.

$$V_{CB} = \frac{\sigma d}{\epsilon_0} \tag{6.7}$$

We can combine Eq. 6.6 with the CCE curves to understand how the signals decrease as a function of time. We first assume there is a constant rate of energy deposition in the

detector from ionizing radiation. An estimate of this rate was calculated from the measured detector energy spectrum, calibrated linearly with the 60 keV peak and adjusted for livetime. This yielded a rate of 90.3 MeV/s deposited for a rate of e/h production of $2.4 \times 10^7$ pairs/s. This rate was found to be dominated by background radiation, not calibration sources, consistent with the uniform $\sigma$ assumption. Of course not all of the initially produced charges contribute to $\sigma$ as some recombine immediately and others are trapped while drifting. But this is exactly what CCE provides a measurement of! This simple model then assumes that charge is being produced at a constant rate, and accumulates (with efficiency set by field-dependent CCE) at the bare face to produce a counter-bias.
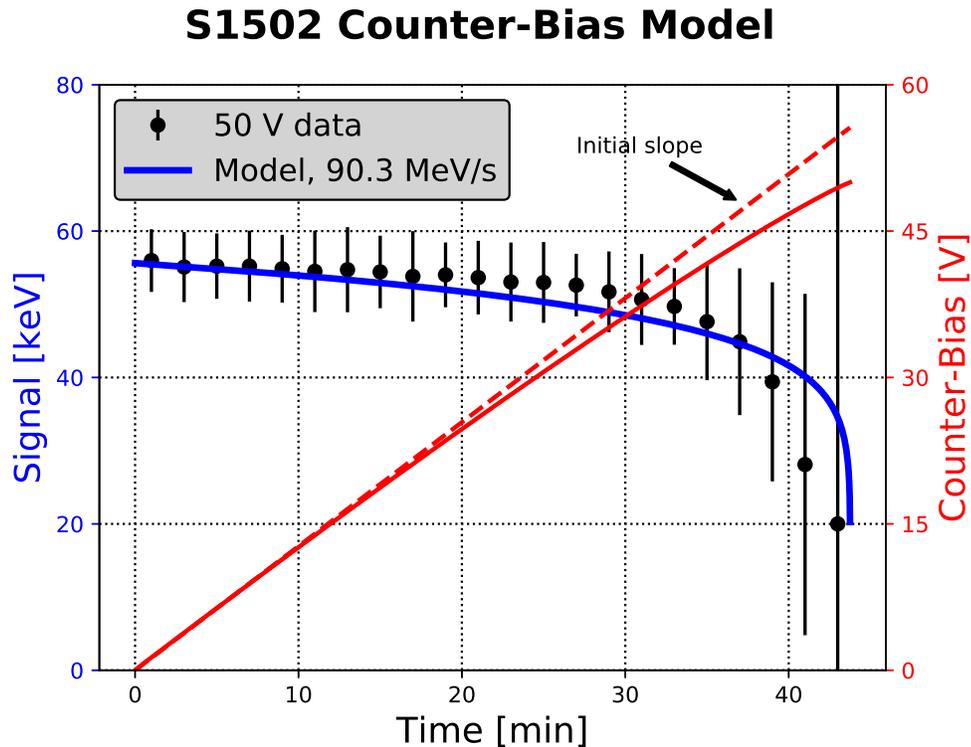


Figure 6.13: An example of the modeled counter-bias development and resulting calibration signal for the case of a total detector bias of 50 V and energy deposition rate in the detector of 90.3 MeV/s. The dashed line shows the initial counter-bias slope for illustration. The measured signal is due to 60 keV events with the effects of CCE applied. Shown for comparison are measured 60 keV peak signals.

An example of this model's output is shown in Fig. 6.13 where the counter-bias and 60 keV calibration line signal are plotted as a function of time. This is compared to a measured signal from a data series with 50 V total applied bias. The counter-bias initially grows at a constant rate (dashed line) set by the initial CCE and the energy deposition rate. As the counter-bias becomes substantial, CCE falls, lowering both the observed signal and the rate at which carriers accumulate at the surface which lowers the rate of counter-bias accumulation. This must occur because, if the counter-bias continued to accumulate at a constant rate, it would eventually become larger than the applied bias. This can't occur because, at some point, the total electric field in the crystal will be 0, at which point the produced charges will not drift, but simply recombine at the event site. Then the bias field will cease to evolve. The model relies on an empirical parameterization of the average CCE curve. Uncertainty in this parameterization may be the cause of the slight overestimation of the signal around the 40 min mark. Even with such slight deviations, be believe the model to be consistent with the observed data.

To quantify this signal loss in S1502, we measured the time it took for the 60 keV line to fall to 80% of its starting value (80% hold time) as a function of total detector bias voltage, $V_{tot}$. The measured data points are compared to the results of the above model as shown in Fig. 6.14. The orange band is the $1\sigma$ uncertainty associated with the spread in individual channel CCE measurements. The model agrees well with most of the data, especially for $|V_{tot}| > 25$.

Importantly, the model shown in Fig. 6.14 has no free parameters. The accumulation rate is set by the total energy spectrum and the shape by the CCE curves. It is thus compelling that it agrees well with the measured data and provides further evidence that the model of counter-bias due to accumulated surface charge is sufficient to describe the observed signal loss.

The largest discrepancies were observed for channel D in $-25 < V_{tot} < 0$. The source of this deviation is not yet fully understood, but is likely related to the fact that it was measured using DCRC RevC electronics, while the others were using RevC.1. In bench tests, we have observed both polarity and DCRC-version dependent gain variations, especially at lower signal amplitudes. Although efforts were made to correct for those effects here, they may still be present at some level in this data. In the future, if such

## S1502 80% Hold Time
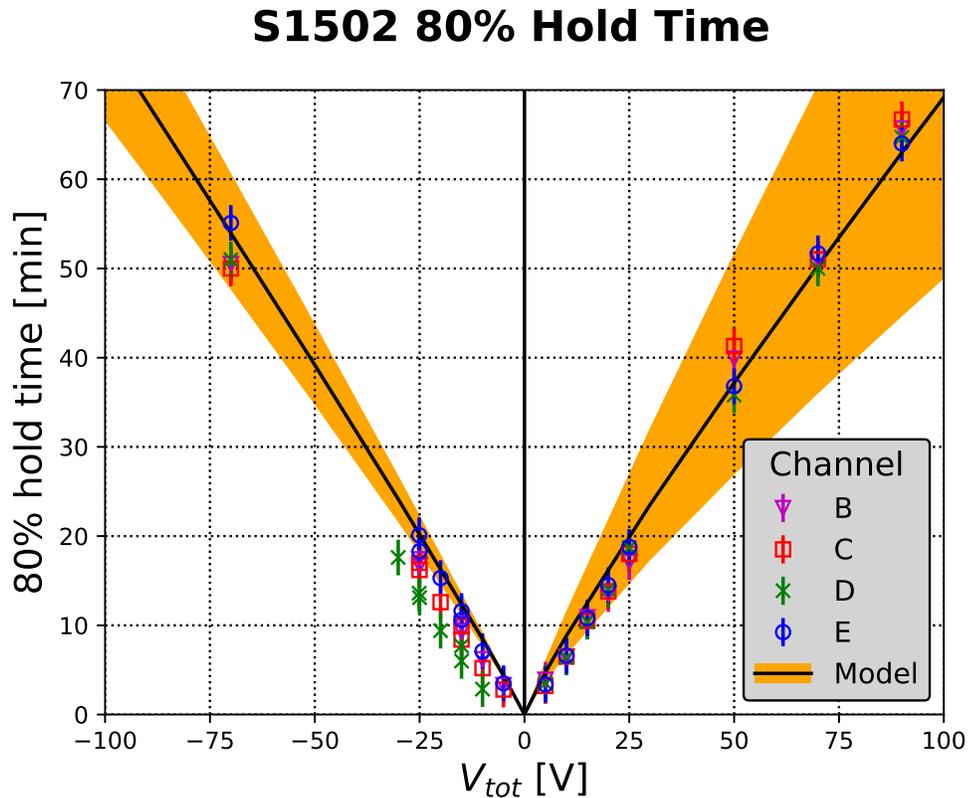


Figure 6.14: The measured 80% signal hold time as a function of total applied detector bias. The results of a simple model of accumulating surface charge are seen to fit the data well. The model uncertainty is higher at positive biases due to larger variation in measured CCE curves at positive biases. The discrepancy in channel D at negative biases is believed to be due to a readout electronics issue.

measurements are made again, it would be wise to take data with DCRC positions swapped to better understand and mitigate these effects.

We can also use this result to make a crude estimate of leakage current present in S1502. A conservative estimate is to require the leakage current be less than 10% of the charge observed from the background event rate (i.e. that we could change the slopes in Fig. 6.14 by 10% and still have an okay fit). This corresponds to a leakage of $\lesssim$380 fA. This is just a factor of a few larger than the leakage noise plaguing CDMS HV detectors, so it is not an unreasonable limit. More careful modelling and better charge signal reconstruction could further tighten these constraints. However, this indicates

that measurements of counter-bias accumulation over time may be a sensitive probe of low leakage current levels, provided the background event rate is low. This technique is powerful because it effectively integrates the current over long time scales which could allow the measurement of low, constant leakage currents. Additionally, it can also be achieved without sensitive phonon electronics, or even on a completely bare crystal! It just requires that at least one crystal face, patterned or bare, be left electrically floating so that charge can accumulate. This could allow more rapid testing of various crystal types and interfaces with only simple charge electronics.

In the UMN setup, the design of the charge readout circuit is such that it should operate stably as long as the NbTi detector tower wiring is superconducting (below 10 K). Even the charge signal noise is expected to be stable since it is dominated by the JFET which self-heats to 150 K [138]. This should allow leakage characterizations, such as described above, to be performed across a range of crystal temperatures, at least up to the LHe bath at 4 K. This would probe a regime where leakage should dramatically and predictably rise if due to culprits such as hopping conduction in the amorphous Si layer [210].

### 6.4.4 Detector Reset with LEDs

The loss of neutralization was mitigated by periodically removing the bias field, grounding the electrodes, and illuminating the detector with pulses of LED light for several seconds in a process called "flashing" [211]. This process is also commonly used for other SuperCDMS detector types to neutralize charge trapped on impurity sites. The standard flashing procedure consisted of powering the LED in 100 $\mu$s pulses at a rate of 200 Hz for 30 seconds (2% duty cycle). The LED photons produced e/h pairs in the crystal which could then diffuse to overcharged regions neutralizing any bound space charge. Flashing typically heated the fridge by $\sim$500 mK, after which the fridge generally needed 30 minutes to return to base temperature, at which point the bias field could be re-applied and data taking resumed. The detector behavior was reproducible, indicating that flashing returned the detector to the same neutralized state each time.

We made dedicated tests of neutralization in S1502 using varied amounts of flashing from three different types of LEDs. There were IR LEDs mounted on the inside of the detector housing, near the sidewalls (Photonic Detectors Inc., model PDI-E940SM). We

also installed another set of IR (Everlight, model IR25-21C/TR8) and UV (International Light Technologies, model E305SL) LEDs on small PCB chips mounted on the electrode face.

Both types of IR LEDs had similar spectra peaked at 940 nm (1.31 eV) which is above the Si indirect band gap (1.12 eV) but below the direct band gap (3.4 eV), but there was a difference in their packaging. The IR LEDs on the detector housing were specially ordered and totally bare diodes, not packaged into an SMD chip. This actually required them to be connected to using special conductive epoxy and wirebonds. By contrast the IR LEDs on the electrode were solder-mounted devices with a clear plastic lens. The UV LED spectrum was peaked at 310 nm (4.0 eV) which is higher than the Si direct band gap. These LEDs also came with a clear plastic lens. After each flashing, a 10 V total detector bias was applied for data collection. With a standard flash, this 10 V bias provided an 80% hold time of 6 minutes. We considered this condition as "full neutralization". We varied the following parameters in our studies: LED type, total flash time, LED pulse rate, and LED pulse width.

Fig. 6.15 summarizes the results of these flashing tests by comparing several parameters. The total amount of time the LED was on during flashing (the product of total flash time, pulse rate, and pulse width) should be proportional to the total amount of ionizing energy emitted. The temperature increase in the detector housing can be used as a proxy for the LED power usage. The LED's power is either converted directly to Joule heat in the chip or to light, which eventually gets absorbed by the crystal or detector housing. In either case, the total power is eventually converted into heating the detector tower. The lower plot shows the increase in tower temperature during flashing and demonstrates that the total "LED ON-time" was well correlated with LED power. Assuming that each LED had a relatively constant efficiency, it is then clear that this also correlated with LED light output. The upper plot of Fig. 6.15 shows that both types of IR LEDs provided full neutralization for total ON-times above $\sim 1$ ms. In contrast, the UV LEDs did not neutralize any better than simply grounding the detector without flashing even though they seemed to be emitting the same amount of power.

Another important observation was that grounding with no flashing at all also neutralized the detector. Grounding for 5 minutes produced partial neutralization (4 min hold time) and grounding for 25 minutes produced full neutralization ($> 6$ min hold

time). It is not unexpected that simply grounding provides some neutralization because there is a constant background event rate in the detector which liberates carriers that also help neutralize the detector. However, it seems that using the UV LEDs may have resulted in hold times that were somewhat *shorter* than without them. This is not yet fully understood, but one possible explanation may be that part of the UV LED spectrum extends up to $\sim 4.3$ eV and is able to produce photo-electrons from the Al electrode since the Al work function is 4.1–4.3 eV [212]. If this process occurs with reasonable efficiency, it could cause such electrons to accumulate on the bare crystal face during flashing, before the voltage bias is even applied. This could be tested by repeating the above tests as before, but instead applying a negative voltage bias for data taking. If there are indeed an initial distribution of electrons there, it will affect the effective bias and increase both the signal amplitude and hold time.

The difference in neutralization properties of the IR and UV LEDs may also be related to the penetration depths of the different photons into the Si crystal. If the photons are absorbed very near the crystal surface, the carriers produced may be less likely to neutralize the crystal by diffusing to trapping sites in the bulk material, but instead simply recombine near the surface. The absorption coefficients for UV and IR light (at room temperature) are $\sim 10^6$ cm$^{-1}$ and $\sim 10^3$ cm$^{-1}$ respectively [213]. The UV energy is greater than the Si direct bandgap so it is absorbed much more readily, with most of the UV energy being deposited within $\sim 10$ nm of the surface. The IR light, on the other hand, can only liberate carriers across the indirect bandgap of Si which also requires a lattice phonon, making it a less frequent interaction and increasing the IR penetration depth. At cryogenic temperatures, the decreased phonon population further lowers this likelihood, increasing the IR penetration depth to more than 10 $\mu$m. It is possible that this difference in photon absorption profile means carriers freed by UV light would (on average) diffuse further to neutralize the crystal bulk, making the UV process less efficient.

The flashing and cool down process for these devices generally took 30 minutes. Given hold times of 10 min – 1 hr, this was a significant component of the detector's operational duty cycle at the test facility. If operated in a low-background environment such as SNOLAB, the detector hold time would be longer by a factor proportional to the total energy deposition rate and could easily achieve duty cycles of >90% if operated

with $V_{tot} \gtrsim 50$ V.

## 6.5 Conclusions and Acknowledgements

The work here showed that large Si crystals can exhibit CCE comparable to smaller devices, though effects of trapping may become a problem for still larger devices if the crystal purity is not improved. Additionally, we showed that contact-free electrodes can be used alone, or in combination with traditional sensors on the crystal. This type of detector geometry has some limitations due to the accumulating counter-bias, it can also provide an avenue for rapid testing and prototyping of different crystal types without the need for more complicated electronics.

This projects described in this chapter began before my tenure with the UMN SuperCDMS group. Much of the design work for the detector, housing, fridge modifications, and electronics was carried out by my predecessors A. Kennedy, H. Chagani, and J. Zhang. That work was described in A. Kennedy [214]. I developed the HV circuitry for this detector as well as a set of miniature LED boards mounted in the detector housing to study the efficacy of both UV and IR light at neutralizing the detector. A. Kennedy, M. Fritts, and myself were the key contributors to the subsequent analysis and publication efforts. The main scientific results of testing these detectors were published in [163], where acknowledgement is given to additional, invaluable personnel at UMN and Texas A&M.
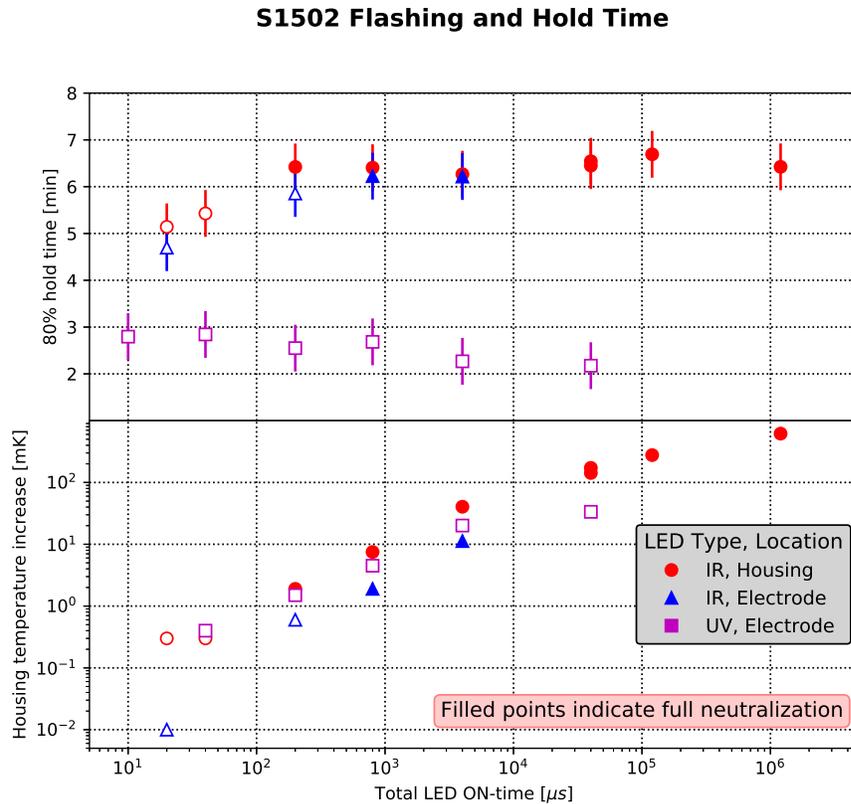
Figure 6.15: Upper: IR LEDs on both the electrode and the housing provided full neutralization over a wide range of total on-times. Housing LEDs performed a little better, providing full neutralization with shorter ON-times. UV LEDs did not neutralize any better than no flashing. Lower: Temperature increase of the housing was a good predictor of whether the detector neutralized completely for IR LEDs. Although they seemed to be providing similar power, the UV LEDs were not observed to help neutralize the detector.

# Chapter 7

# Low Energy NR Calibration

SuperCDMS is increasingly relying on the NTL gain of HV detectors to decrease the measured energy threshold and observe lower energy events. The total phonon energy generated in a detector for a recoil of energy $E_R$ is

$$E_{tot} = E_R \left( 1 + Y(E_R) \frac{eV}{\varepsilon_\gamma} \right) \tag{7.1}$$

where $Y(E_R)$ is the ionization yield (defined to be 1 for ERs and $< 1$ for NRs), $V$ is the detector bias voltage, and $\varepsilon_\gamma \approx 3.0(3.8)$ eV in Ge(Si) is the average energy to produce an e/h pair [129, 130]. Increasing $V$ amplifies the total phonon signal, but makes it increasingly degenerate with the charge signal energy, $E_Q = E_R Y(E_R)$. This method thus sacrifices the simultaneous charge and phonon measurement (which, in iZIP detectors provides ER/NR discrimination) for a highly amplified measurement of the charge signal via phonons.

Without event-by-event ER/NR discrimination, measured energy spectra are in general a combination of ER and NR events and it becomes essential to know the ionization yield down to low energies to correctly model the observations. As described in Sec. 2.2, there is generally significant tension in current measurements of $Y$ for both Ge and Si, some of which may be due to the variety of detector types and operating temperatures used. Additionally, there are very few measurements for $E_R < 1$ keV, and none made using the same bias field strength and mK temperatures employed by SuperCDMS.

Before we get too much further, there are several related energy scales we will discuss in this work. The first is $E_{tot}$, the total phonon energy produced for a given

event; this is the signal our phonon sensors measure (up to a constant factor of phonon collection efficiency). However, because $E_{tot}$ depends on detector bias, we generally calibrate the phonon signal to the equivalent recoil energy for ERs using e.g. known $^{241}$Am gamma lines. This calibration results in an "electron-equivalent" energy scale, $E_{ee} = E_{tot}/(1 + \frac{eV}{\varepsilon_\gamma})$, which we denote with units of eV$_{ee}$. For ER events, $E_{ee} = E_R$, so we recover the initial recoil energy. However, the total phonon energy of an NR event of the same recoil energy will be reduced because $Y(E_R) < 1$. Without ER/NR discrimination, we can not identify an individual event type and will reconstruct such an event to have an electron-equivalent energy of $E_{ee} = E_R(1 + Y(E_R)\frac{eV}{\varepsilon_\gamma})/(1 + \frac{eV}{\varepsilon_\gamma})$, which is less than $E_R$. So to understand how our detector will respond to NRs, we must map out $Y(E_R)$.

Now, one reason for a dearth of such measurements is that there are few suitable NR calibration sources. The current lowest energy yield measurements in Si were performed by fitting the quasi-exponential recoil energy distribution of neutron-nuclear elastic recoils caused by a photo-neutron source [215]. This type of measurement is relatively simple to carry out, but relies on detailed modelling of neutron transport to simulate the correct neutron energy flux at the detector. It also requires fitting of the resulting model to a featureless exponential energy spectrum, which results in significant correlations between derived yield values at different energies. Another method which is currently being undertaken by SuperCDMS is to use a neutron beam of fixed energy to scatter neutrons off the target detector and into an array of backing detectors. This way, coincidence between detectors can give individual recoil energies via simple scattering kinematics. In principle then, each scatter observed at a given angle becomes an independent probe of the yield at a given energy. This work is promising, but requires substantial infrastructural and technical investment because one must deliver the entire cryogenic setup to a suitable neutron beamline. SuperCDMS is also investigating a crude alternative to this whereby the neutrons are produced at the cryogenic setup using a D-D neutron generator and measured by custom backing detectors. The principle here is the same, but the neutron energy magnitude and spread are both higher than at a beamline.

An alternate method which was recently proposed is to use the nuclear capture of thermal neutrons and subsequent nuclear de-excitation, i.e. $(n, \gamma)$ reactions on the

detector nuclei. Some of these occur by emitting an $\mathcal{O}(10\text{ MeV})$ gamma [216], which results in the nucleus recoiling with energy $E_R \simeq E_\gamma^2/2mc^2$, where $m$ is the nucleus mass. In Si and Ge, this process gives $E_R < \mathcal{O}(1\text{ keV})$, right where measurements of NR yield are sought! The process may be difficult to identify, but coincident measurement of the escaping gamma could help isolate it from backgrounds. This chapter presents a proof-of concept of the use of such $(n, \gamma)$ reactions to produce and measure low energy NRs using a prototype SuperCDMS SNOLAB detector. The experiment highlighted here was performed during the summer of 2018 in UMN R68. A publication describing the main results is currently in preparation.

## 7.1  Experiment

For this experiment we operated a prototype SuperCDMS SNOLAB Si HV detector [217] using SuperCDMS Soudan SQUIDs and cold hardware [138] modified to reduce phonon readout noise. The neutron source was a strong PuBe source (1.4 Ci α, 62 µCi n) shielded by a paraffin barrel. The detector cryostat was also shielded on three sides and from below by 8" of polyethylene slabs and on four sides by 5/8" thick lead panelling. The paraffin and polyethylene served as moderators to thermalize the neutrons. Additionally, a 1' thick lead wall was placed near the PuBe barrel to block direct MeV gammas from the PuBe source. Two 8" diameter by 8" thick cylindrical NaI detectors were also mounted near the cryostat to search for coincident capture gammas. A top-down diagram of this arrangement is shown in Fig. 7.1.

A $^{241}$Am calibration source was mounted in the HV detector housing. This source has a primary gamma emission energy of 60 keV and multiple secondary lines between $10 - 30$ keV. A 66 mil thick lead disk with an 18 mil hole collimated the source gammas and restricted the emission rate. A strip of Kapton tape placed over the collimator blocked alpha emission and low energy ($< 3$ keV) gammas.

An FPGA on the DCRC was programmed to monitor the HV detector phonon signals in real time and issue triggers. A simple trigger quantity was calculated as the difference between the average signal in a 4 $\mu$s trigger window and the average baseline value in the preceding 12.8 $\mu$s. If this quantity exceeded a threshold value on one of three selected phonon channels, A(outer), C(mid), or F(inner), a trigger was issued and
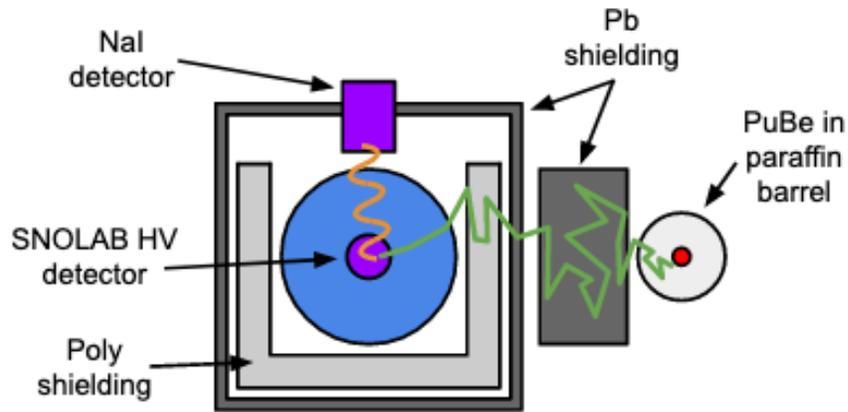
Figure 7.1: Top-down diagram of source and detector arrangement for neutron capture calibration study (not to scale). Cartoon of neutron thermalization and capture shown in green with resulting gamma emission in orange.

a 3.2768 ms trace around the trigger time was digitized for offline analysis. The trigger threshold values were set as low as possible given the sensor noise without completely overwhelming the DAQ system with a high rate of triggers. During some periods of data taking, the DCRC trigger rate could exceed the rate which the DAQ could handle. In this case, so called "stale" triggers which were not digitized in time would be dropped in an energy-independent manner. This lead to a finite DAQ efficiency which we accounted for in later analysis. At the beginning of each data series, a set of 1000 traces at random trigger times were selected and read out before the threshold triggering algorithm was enabled. These random triggers were primarily used to calculate noise spectra used in pulse fitting, but were also crucial in evaluating various analysis and DAQ efficiencies.

During the course of this experiment, we took many types of data with a range of bias voltages and external radiation sources to characterize and optimize the detector operation. In the end, the most useful datasets were all taken at a detector bias of -125 V. Note that this device was operated in so-called CDMS-lite mode, wherein one face of the detector was only used for biasing while the six phonon channels on the opposite face were held at ground and read out. We took three types of data in this mode; Background (Bkg) data which consisted of only the room background radiation and the $^{241}$Am source located in the detector housing, $^{22}$Na (Na) data where we also placed a strong $^{22}$Na source inside the lead shielding to search for Compton scattering signals,

and PuBe (PuBe) data in which we introduced the shielded PuBe source to produce thermal neutrons. In all, we operated the detector for 29.5 hours of Bkg, 81.7 hours of Na, and 97.8 hours of PuBe data, although the actual livetimes were only $60 - 80\%$ of these values due to the DAQ efficiency.

### 7.1.1 Noise

The detector energy threshold was limited by noise in the phonon signal. At 0 V, the Johnson voltage noise on passive components in the phonon circuit dominated the noise. This was reduced by lowering the TES shunt resistors as far as possible while maintaining stable operation. However, as described in Sec. 4.6, a source of bias-voltage-related noise quickly becomes dominant when a bias voltage of $\sim$10s of V is applied. Such noise is present to some extent in many devices and is associated with some sort of charge leakage through the detector. The fundamental mechanism, however, is still being investigated by SuperCDMS and other collaborations. The $E_{tot}$ noise levels measured in R68 are shown in the left side of Fig. 7.2 as a function of detector bias voltage. As commonly observed, there was a substantial difference in noise level with bias polarity. For the main data taking, we selected a bias voltage of -125 V as it optimized the noise width. That is to say, it was where the increase in signal due to NTL gain began to be overtaken by degradation from increasing leakage noise. This can be seen as the minimum point of the plot on the right side of Fig. 7.2 where the noise has been calibrated to eV$_{ee}$. This choice of bias voltage provided an NTL gain for ERs of $G_{NTL} = 1 + eV/\varepsilon_\gamma = 1 + 125/3.8 \approx 34$. In both plots, we also show the expected scaling of two measurements of the noise at 0 V, assuming that no additional noise sources entered.

### 7.1.2 OF Templates

Because the expected signal was very near threshold ($\sim$100 eV$_{ee}$), the Optimal Filter (OF) template shape was important. To construct the templates we selected 716 events from a range of times throughout the data run. The pulses were selected from events with energy between 0.1-4 keV$_{ee}$ which passed basic data quality cuts. These pulses were aligned to have the same start times and an average pulse shape was calculated
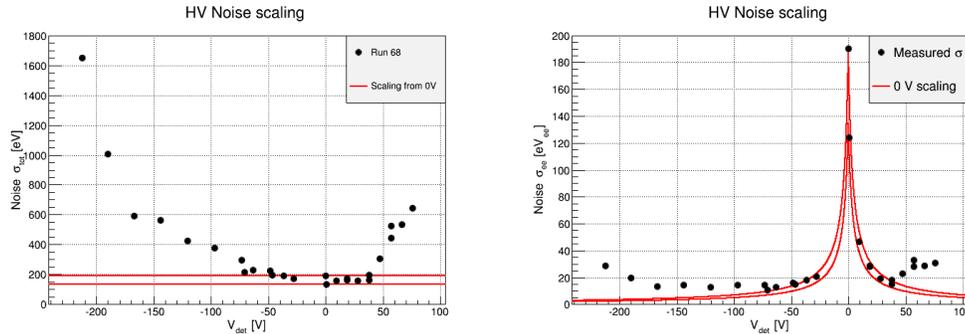
Figure 7.2: R68 noise width as a function of HV bias in units of total phonon energy (left) and electron-equivalent energy (right).

for each of the six phonon channels and used as the OF templates. A total phonon pulse, PT, was also calculated for each event by summing the six traces together with relative weights ranging from $0.65 - 1.3$, which were calculated to account for variations in intrinsic channel gains found from $I_b - I_s$ analyses as described in Sec. 3.2. An OF template was calculated from the average PT pulse as well and OF fits of the PT pulses were used as the main energy estimator.

The detailed pulse shapes of the final templates are shown in the left side of Fig. 7.3. The 0.8 $\mu$s width of each time bin is set by the DCRC sampling time. The pre- and post-pulse baselines are generally flat to within less than 1% of the peak height. This is much smaller than the average baseline-shift to pulse-height ratio of $\sim$50% of the weakest channel (PD) near threshold ($\sim$0.1 keV). This helps ensure that the OF algorithm fits the template to a pulse in the raw trace, as opposed to fitting any sloped baselines. In high-rate data, there can be a substantial baseline slope for some events when falling on the tail of a larger pulse. If the baseline change over the trace length is comparable to the pulse height and the template has a sloped baseline, the lowest fit $\chi^2$ can be found by matching the template to the baseline instead of the pulse. The flat template baseline prevents such shenanigans. There are small oscillations visible on the falling edges of some templates, particularly the inner-ring (PC – PE) channels. This is the result of electrothermal oscillations, the effect of which were minimized by selection of channel operating points, but which were still present to some degree. The right side of Fig. 7.3 shows an example of the OF fit to a 456 $\mathrm{eV_{ee}}$ PT pulse. To speed up computational

time, the start time was required to fall within a 100 $\mu$s "OFdelay" window around the trigger time.
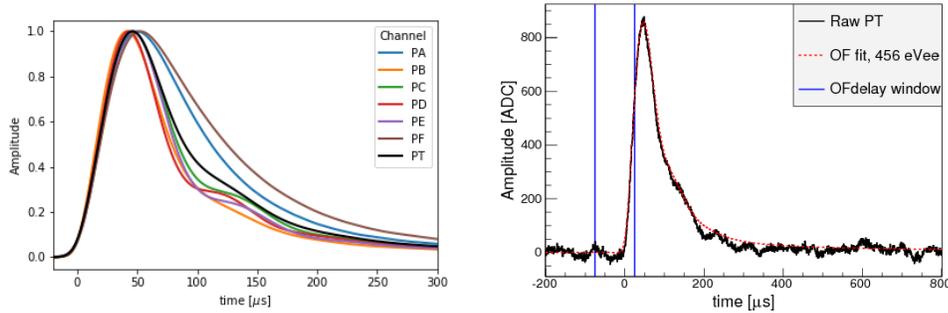


Figure 7.3: Left: R68 OF templates with electrothermal oscillations are visible on the falling edges of some channels. Right: Example fit to a 456 eV$_{ee}$ PT pulse.

## 7.2 Data Analysis

### 7.2.1 Baseline Correction

TES gain changes with position in the transition. If the TES is still cooling after a large pulse, the sensor baseline will be elevated and the gain will be lower than at the quiescent bias point. The left side of Fig. 7.4 shows this effect by plotting variations of the central channel, PF, baseline as a function of measured OF amplitude of the PT pulse, PTOFamps. The tilted calibration lines indicate that pulse amplitudes are reduced when the baseline is elevated. These lines allow a simple linear correction factor to be included for this effect, the result of which is shown in the figure on the right.

### 7.2.2 INTall Baseline Slope Correction

The current SuperCDMS pulse processing code calculates a trace integral of the baseline-subtracted PT pulse, a quantity named PTINTall. This is generally useful when identifying pileup events. However, if a pulse is on a sloped baseline, such as when falling on the tail of a larger pulse, the baseline will be over-subtracted, leaving much of the pulse
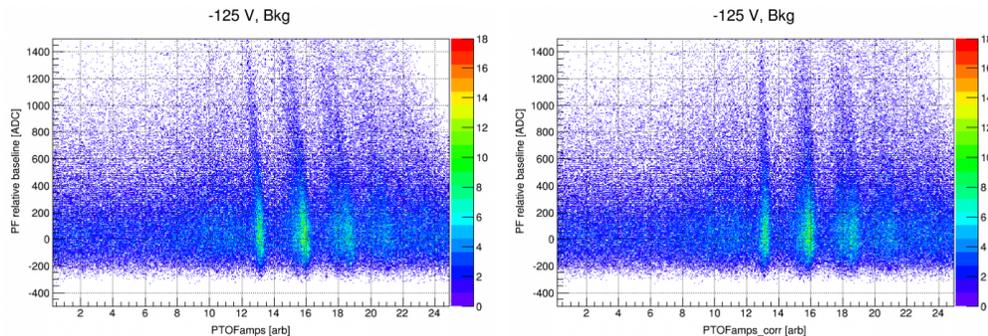
Figure 7.4: R68 baseline correction in background data. Left: Uncorrected pulse amplitudes vary with baseline value. Right: Simple linear correction removes slopes in calibration lines.

negative and reducing the calculated integral. To correct for this, the pre- and post-pulse baseline quantities, BSpre and BSpost respectively, were used to infer the slope upon which the pulse is sitting and a factor was added to the integral which adjusts for the "extra" area due to the sloped baseline.

Fig. 7.5 is a diagram of this idea. A pulse on an exponentially sloped baseline is shown in blue. It has been baseline subtracted so that the pre-pulse baseline is at 0. The orange shading shows the areas which contribute to the INTall quantity with those below the axis giving negative contribution. If the slope is large enough this can even result in a negative integral for a trace containing a positive-going pulse. The red line shows the correction factor suggested, it is simply the area of the triangle formed by the pre- and post-baseline averages and the trace length. The corrected quantity is named PTINTall_bscorr. While this factor is not perfect, it is a definite improvement on the standard INTall value. For highly sloped baselines, such as in the toy example shown above, the correction factor may be large. If the baseline is quite flat, it will be correspondingly small. As described, the correction is only made for pulses with a lower post- than pre-pulse baseline. For example, random noise may cause the end of the trace to be slightly higher than beginning even for a good pulse on a very flat baseline, in which case no correction is needed.
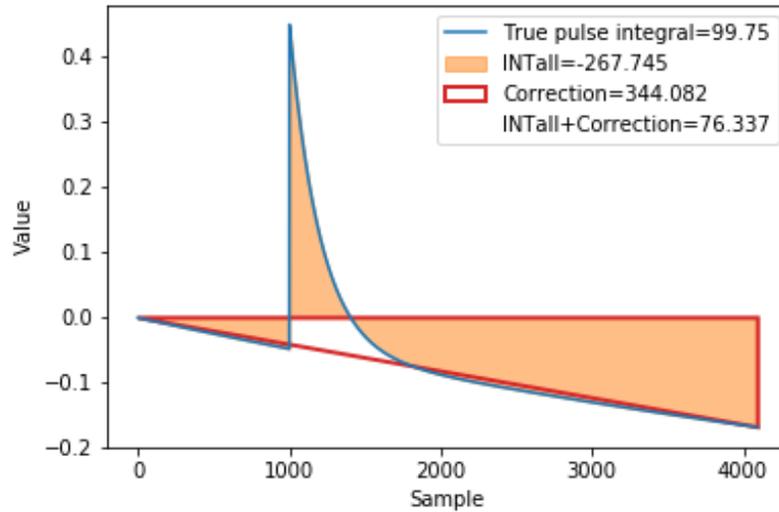
Figure 7.5: INTall sloped baseline correction. The sloped baseline would otherwise drastically reduce the calculated pulse integral.

### 7.2.3 Data Quality Cuts

A series of standard quality cuts were performed on both the PuBe and background data to remove undesirable events. These are such things as pileup events, electronics glitches, or poorly reconstructed pulses. Each cut is described in detail below.

**Baseline Cuts**

Events were cut if the pre-pulse baseline average or variance was excessively high. Both of these quantities were calculated using 320 $\mu$s of data in the region of the trace well before the DCRC trigger time. The variance cut was developed to remove a class of events which triggered twice. This occurred because, for sufficiently high energies, the falling edge of a pulse would oscillate and satisfy the trigger condition a second time, while the bulk of the pulse was in the pre-trigger region. Using a loose cut on the pre-pulse standard deviation quantity for channel F, named "PFstd", helped remove these events. However, it is not expected to have an energy-dependent efficiency on the true spectrum, it is only removing events that would have been double-counted otherwise.

Because the phonon signal is generally shared between all channels, it is sufficient to use a single channel to make this cut. The left side of Fig. 7.6 shows the distribution of PFstd as a function of the total phonon integral. Events above the cut line have anomalously high pre-pulse standard deviations and were removed.

Another simple cut was made on the average PF baseline quantity. This is the same quantity used in Sec. 7.2.1, here we simply placed an upper limit on baseline variations which were considered for analysis. This cut helped remove events which fell on the tail of an earlier pulse. The right side of Fig. 7.6 shows the variation in PF pre-pulse baseline quantity as a function of total pulse amplitude. The most prominent feature is the band of high energy ($\gtrsim$30 keV) events which reconstruct to much lower energies when on a high baseline. Because the baseline variation is uncorrelated with the true energy of the triggered event, a constant cut threshold was used. The location of the cut was selected to identify the upper band of high energy events for energies below $\sim$20 keV and remove them from analysis. The same cut value was used for all datasets.
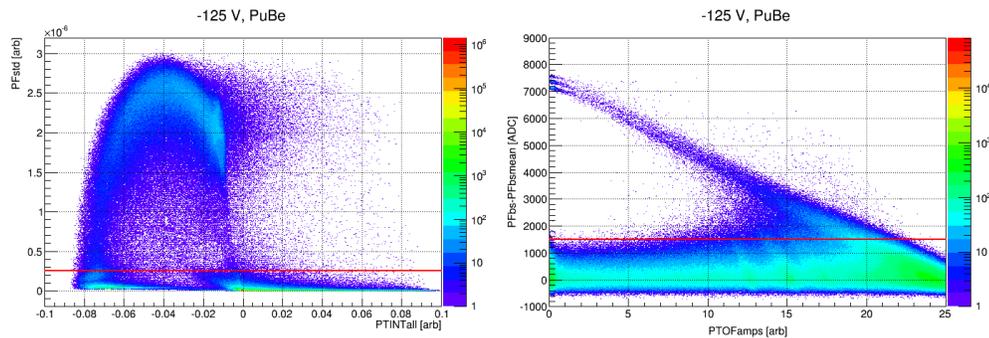


Figure 7.6: R68 baseline standard deviation cut (left) and baseline variation cut (right).

This cut also removed a class of retriggers which could leak past the baseline standard deviation cut above. Some retrigger pulses are very saturated and have a flat pulse top. If this flat top takes up most of the pre-pulse baseline region, it can have a low PFstd. However, it will also have a very high PFbs and thus be removed by this cut. So the combination of these two cuts serves to remove both retriggers and pulses on large tails. Because they are defined using data that occurs well before the triggered pulse, these cuts are independent of the triggered pulse true energy (though not the reconstructed energy) and thus should have a constant cut efficiency.

**OF/Integral consistency cut**

Pileup events can distort the measured pulse amplitude. Three cuts were used to remove such events. The first relied on the ratio of OF amplitude to pulse integral. Because the pulse shapes were roughly constant, the ratio of integral to OF amplitude proved a good measure of the presence of multiple pulses in a trace. This identifies pileup because the integral includes all pulses in the trace while OF just fits one in the OF search window. As shown in Fig. 7.7, this cut was defined in the OF amplitude vs. integral plane. To make the cut less subjective, vertical slices of the INT-OF plane were fit with the crystal ball function (gaussian with a power law tail defined in Appendix D of [218]. The slices were taken over 20 evenly spaced windows from PTOFamps = 0.1 to 16. Then, the set of points $3\sigma$ above each fit mean were all fit with a quadratic function to define the cut threshold. The y-intercept of the function was fixed at a corrected PT integral of $10^{-4}$ because the noise wall tended to make this fit parameter unstable. The resulting cuts were used to remove events with relatively high OF/INT ratios.

At energies above $\sim$20 keV, TES saturation became significant and the OF quantity also saturated. The 60 keV line is heavily saturated and can be seen at PTOFamps $\approx$ 24. Some effort was made to include the beginning of this saturated region in the OF/INT cut, but it eventually cuts off these saturated events and an upper analysis threshold is later set far below this region. The calibration lines all fell below PTOFamps = 20 on these plots. The resulting cuts remove events with relatively high OF/INT ratios. Events which fall above the band are generally pileup, the trace integral being too large given the triggered pulse amplitude. Events below the band, however, are generally good pulses with the remnant of a small pulse tail in the prepulse. This gives an elevated baseline, even lowering the corrected integral quantity. However, such events also come with elevated OF fit $\chi^2$ value, and will be removed by a later cut.

**OFdelay/WKr50 Timing Cuts**

A second type of pileup cut was made using the pulse timing quantities PTOFdelay and PTWKr50. PTOFdelay is the amount of time shift between the template and measured pulse as calculated by OF ($t_0$ in Eq. E.8). As explained in Appendix E, the OF algorithm finds the best fit within a preset window of $t_0$ values around the DCRC
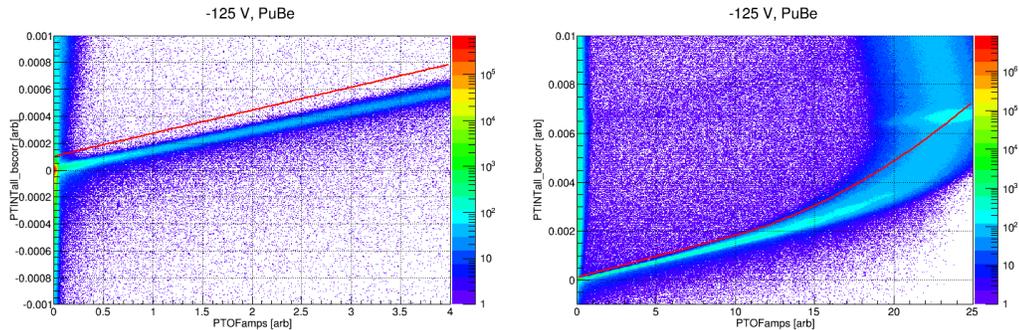
Figure 7.7: PT integral versus function of OF amplitude for PuBe data. Events above the red line are identified as pileup and removed. Left: Focused on the signal up to ∼4 keVee and Right: showing the full energy range, including saturated pulses. The 60 keV peak is visible at PTOFamps ≈ 24.

trigger time. We placed an analysis cut to exclude events which have OFdelay values at either edge of the allowed window. This is designed to catch situations where there is a larger pulse partially outside the window which the OF is trying to fit. This sort of pileup would also be removed by a cut on fit $\chi^2$ (see Sec. 7.2.3 below), but we included it here to better understand the nature of different event types being removed from consideration.

For the final pileup cut, we required a loose consistency between the timing values of PTOFdelay and PTWKr50, the time of the 50% rising pulse edge of PT, calculated with the RTFTWK algorithm [142]. This removes many noise events as well as otherwise good events with a pileup pulse in the final few bins of the trace which can leak through the OF-INT consistency cut. This cut was also needed to ensure the RTFTWK algorithm fit the same pulse as the OF algorithm and is required for later glitch removal cuts. In Fig. 7.8, only events inside the four lines were retained.

**Glitch and Spiky Cuts**

Two populations of odd events were found when setting standard quality cuts. One type was identified as DCRC readout glitches. For these, it appears that a portion of the trace is from some other, unrelated location in the DCRC memory buffer. Later bench tests have confirmed that such glitches occur rarely for some DCRCs, although the exact cause has not yet been determined. An example of this effect is shown in the
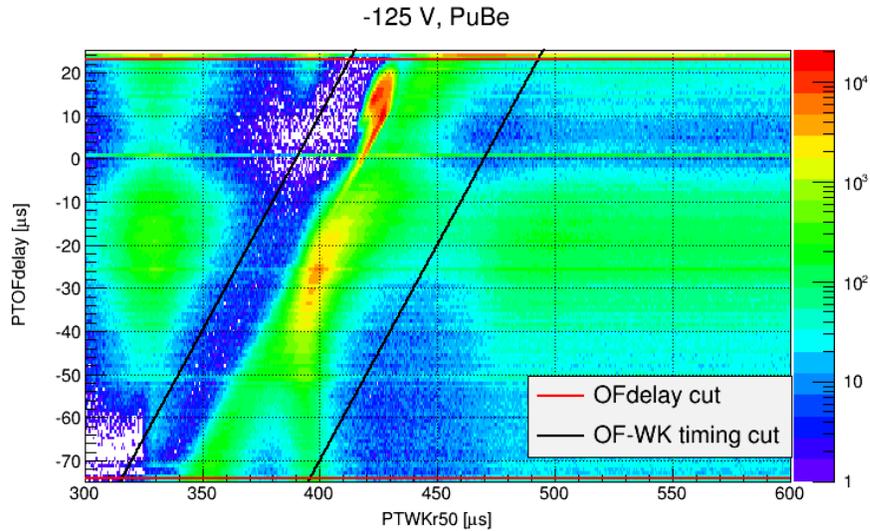
Figure 7.8: Relative pulse timing cuts on PuBe data. The two cuts remove events for which either the OF did not fit a pulse near the trigger time or for which the OF and RTFTWK timing quantities were inconsistent. The vast majority of good events lie in the small blob near PTWKr50 = 425 $\mu$s and PTOFdelay = 10 $\mu$s.

left side of Fig. 7.9 where two channels contain anomalous data points.

The other type is a train of "spiky" pulses which were observed only in the first part of the PuBe data taking campaign. Due to its transience, this was assumed to be an unidentified noise source coupling to the electronics. Examples of phonon signals from such events are shown in the right side of Fig. 7.9. There, we observe a train of pulses with very short ($< 50\ \mu$s) fall times across several channels. Both of these populations were identified and removed using their fall times which are so fast as to be unphysical for true phonon signals. The fall times were calculated as the time from the maximum of the PT pulse (PTWKr100) to the time it fell to 20% of this max (PTWKf20).

In Fig. 7.10 we show the distribution of these fall times as a function of OF amplitude. Good events form a band at fall times of about 110 $\mu$s. The "glitch" events appear as a band near 7 $\mu$s and the "spiky" events at around 25 $\mu$s. The spiky events were most prominent in PuBe datasets 1 and 2 (shown on the left), though there are a handful in the later data as well. Events falling below the black and red lines fail the glitch and spiky cuts respectively, and were discarded from further analysis.
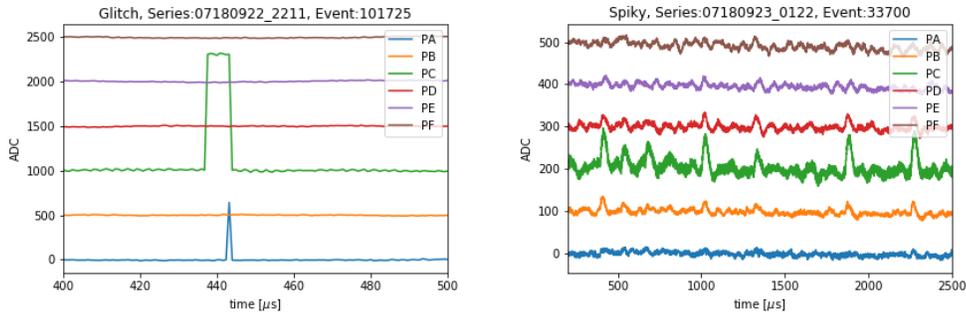
Figure 7.9: Phonon traces for glitch (left) and spiky (right) events.
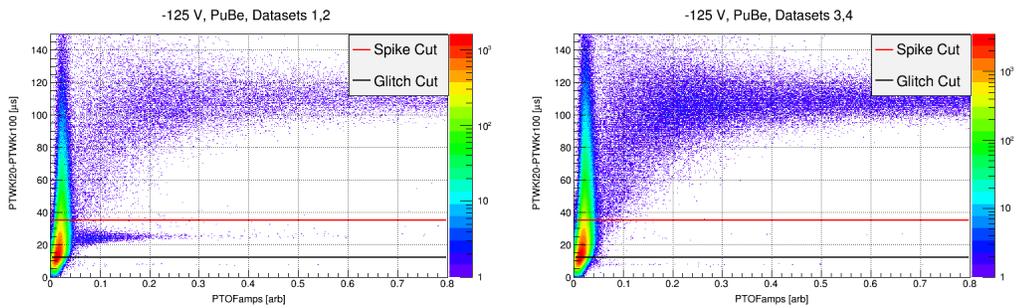


Figure 7.10: PT fall times as a function of PT amplitude for PuBe datasets 1 and 2 (left) and 3 and 4 (right).

### OFchisq Cut

The goodness of the OF fit is quantified with the PTOFchisq quantity which contains the frequency-domain fit $\chi^2$ value defined in Eq. E.8. As described in Appendix E, we expect $\chi^2 = N_{samples} = 4096$ for a good fit. Fig. 7.11 shows the distribution of $\chi^2$ values as a function of PT pulse amplitude. We can see that low energy pulses generally have $\chi^2$ values close to the ideal value, but they eventually worsen at higher amplitudes as TES saturation effects change the pulse shape to be different from the template. The cut was defined to include the main band of events in all three data types (PuBe, Na, and Bkg). It is shown as the red line, with all events above the line being removed from analysis.
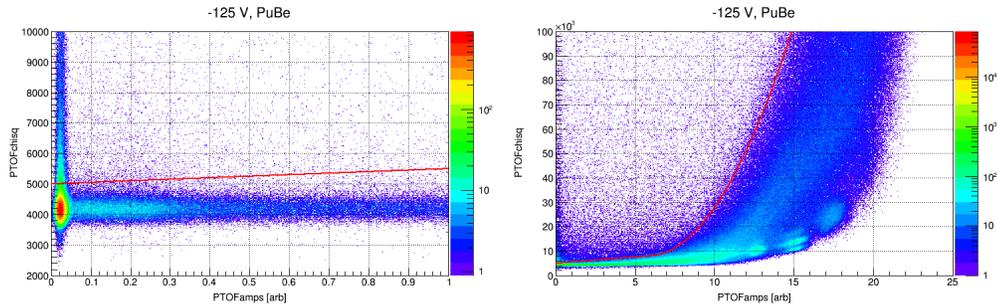
Figure 7.11: OFchisq for PuBe data. Left: Low amplitude ($\lesssim 1$ keV) pulses have near ideal $\chi^2$. Right: At higher energies the pulse shape starts to change from the template and $\chi^2$ correspondingly increases.

**Low Energy Trigger Burst Cut**

The final quality cut was made to remove a class of low-energy trigger bursts which were discovered in this data. These were first observed as an unexpected low-energy hump in Na calibration spectra but were also present in the PuBe data to a lesser extent. The events were found to come primarily in relatively short bursts during periods of $\sim 100$ ms. These sort of bursts can be seen as vertical stripes in Fig. 7.12 where the PT amplitudes below $\sim$500 eVee of a subset of triggered Na events are plotted as a function of event number (order in which events occurred). The scale of the y-axis is roughly equivalent to 500 eV$_{ee}$, so these events occur just above the 50 eV$_{ee}$noise threshold and appear in correlated clumps unlike what one would expect from an uncorrelated set of random processes. The physical source of these bursts is not yet understood, and they are a source of ongoing study in this and other devices (more on this in Ch. 8). However, for the purpose of this analysis, they were not included in any of our current physical models and thus we sought to exclude them.

Ideally we could calculate the true rate at which such triggers are occurring and remove time periods with elevated rates from analysis. However, the trigger rate was generally higher than the DAQ software could handle so some triggers were randomly discarded during data taking. Thus the actual rate of events digitized and saved to disk is lower than the trigger rate. To further complicate matters, the efficiency of this discarding process varies with triggering rate. When there are more triggers to handle, the DAQ becomes even less efficient. Additionally, our DAQ does not store a
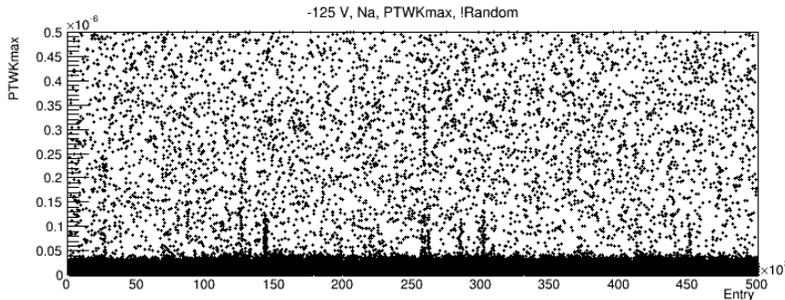
Figure 7.12: Low energy trigger bursts in Na data.

high fidelity global event time with each trigger, so we cannot use the relative timing of events to identify bursts. Instead we simply use their proximity in triggered event number, $dN$. As defined, $dN = 1$ for a pair of sequential events.

The DAQ discards triggers in a manner uncorrelated with event amplitudes (it in fact only has access to the trigger times and nothing else before deciding which events to save to disk). Thus the saved event amplitudes should be uncorrelated with order of occurrence. If we then divide the events into two categories (here events in/out of the $50 - 1000$ eV$_{ee}$ ROI), the distribution of said categories should be binomial if they are produced by random processes. In particular, the probability of two ROI events having $dN - 1$ non-ROI events between them should follow

$$P(dN|p) = p(1-p)^{dN-1} \tag{7.2}$$

where $p$ is the probability of any single event being in the ROI. Assuming the overall rate of bursts is relatively small, the presence of bursts will be seen as a deviation from this behaviour at low values of dN. This is precisely what is demonstrated in the left panel of Fig. 7.13 where we have histogrammed the distribution of dN values occurring in the PuBe data. The orange line is a fit of Eq. 7.2. As expected, the fit is quite good except for the excess at low dN (inset) due to the burst events.

It is then natural to cut events with low $dN$ values to remove contamination from burst events. The cut was defined to remove pairs of events associated with any $dN \leq dN_0$. The effect on the PuBe energy spectrum of using successively higher $dN_0$ values is shown in the right panel of Fig. 7.13. After a certain point, the spectral shape converges, indicating the burst events have been essentially eliminated.
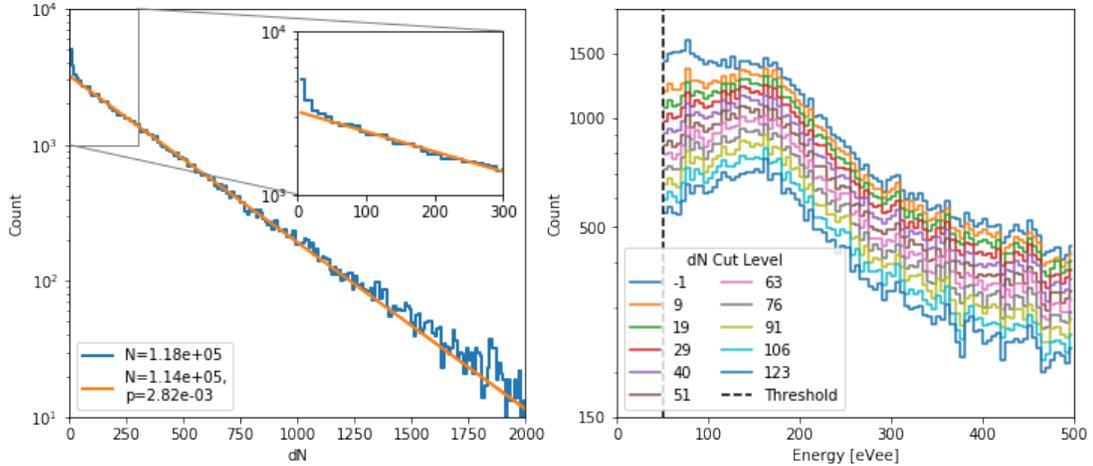
Figure 7.13: Left: dN distribution with fit. Excess due to bursts shown in inset. Right: Effect on PuBe spectrum of successively stronger burst cuts. Note y-axis is log scale to demonstrate that spectral shape converges for cut levels larger than $\approx 20$.

With this cut definition and the assumption that bursts are relatively rare, the cut efficiency can be calculated as

$$eff_{Burst} = P(dN > dN_0|p)^2 = (1-p)^{2dN} \tag{7.3}$$

A cut level of $dN_0 = 20$ was selected for the PuBe dataset because it was where the passage fraction of subsequent cuts was consistent with the theoretical efficiency for good events. The cut only removed events between the 50 $eV_{ee}$ analysis threshold and 1 $keV_{ee}$, where the resulting efficiency was $0.8933 \pm 0.0006$ and the estimated leakage fraction was $0.0007 \pm 0.0065$. Outside of $50 - 1000$ $eV_{ee}$, the efficiency of this cut was thus 1. The dN distribution of the Bkg dataset was consistent with having no burst events, so this cut was not applied to it.

### 7.2.4 Cut Efficiencies

The efficiency of each cut was estimated and the results are shown in Fig. 7.14 for both Bkg and PuBe datasets. The tail, pileup, and trigger burst cuts were constructed so as to be independent of the energy of the the triggered event. The tail cut efficiency was calculated from the passage fraction of randomly triggered traces and was found to

be $0.820 \pm 0.001$. For the pileup cut, $2.7\%$ of triggered events were estimated to have pileup based on the rate of pulses in randomly triggered traces. The pileup cut passage fraction was $0.965 \pm 0.001$, indicating that it was slightly stronger than necessary as it removed $3.5\%$ of the triggered events, so it was assumed this cut was working well and the efficiency for signal events was also assumed to be $0.965 \pm 0.001$. The trigger burst cut efficiency was estimated from the passage fraction of events in an energy sideband. It was found to be $0.989 \pm 0.001$. The spike cut efficiency was energy dependent. It was calculated as the passage fraction for PuBe datasets 3 and 4, which were relatively free of such pathological events. The $\chi^2$ cut selected the best-fit pulses, it did not remove any other known pathologies. As such, we assumed it operated on a set of real events, so the passage fraction of this cut is also the cut efficiency. Statistical uncertainties were included in the cut efficiency calculations. Each efficiency was calculated starting relative to the event population after applying all prior cuts, so the total cut efficiency is the product of the individual cut efficiencies. To average over statistical variations, a smooth function was fit to the total cut efficiency curves and was used for subsequent analysis. As can be seen in Fig. 7.14, an uncertainty of fixed-width was assigned so as to encompass the full range of variation observed. The uncertainty for background data was larger because the dataset had lower statistics and thus higher relative statistical uncertainties.
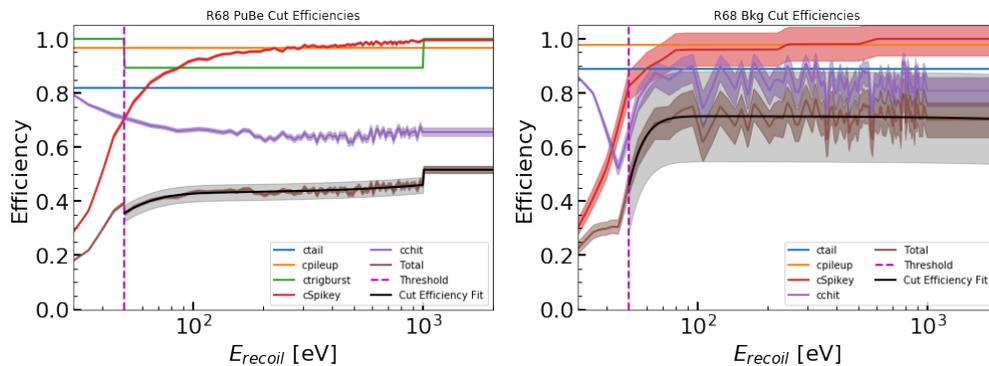


Figure 7.14: Estimated cut efficiencies for PuBe (left) and Bkg (right) datasets. Best-fit analytical function and uncertainty are shown as the black and gray band respectively.

### 7.2.5 DAQ Efficiencies

There were two known sources of inefficiency in the DAQ system. The first was the efficiency of the triggering system described in Sec. 7.1. We only triggered on some channels so events which had low energy or which were located far from a triggering channel may not have passed the trigger condition. The efficiency of this process was calculated by simulating the triggering on fake events. Fake events were constructed from scaled pulse templates added to real pulse-less noise traces selected from the set of randomly triggered events. The correct distribution of relative channel amplitudes was used by drawing from 325 prototype triggered events between 0.5 and 1 keV$_{ee}$, where the trigger efficiency was assumed to be 1. Fake events were constructed at 32 energies between 0 and 0.5 keV$_{ee}$ which were then passed through the triggering algorithm. The trigger efficiency was calculated as the fraction of fake events which did not trigger on the pure noise trace, but which did trigger when a pulse was injected. This procedure was repeated for a set of 1500 noise traces and prototype events selected from throughout the whole dataset. The uncertainty was dominated by systematic effects of using different sets of noise and prototype events. The resulting simulated trigger efficiency curve can be seen in Fig. 7.15.

The DAQ system also suffered from a "write efficiency". This is the efficiency with which triggered events are actually read out and saved to disk. The DCRC has a ~3 second buffer for which it maintains a list of triggers. The DAQ computer periodically polls the DCRC for this list and begins to request the portion of the buffer around each trigger point. The DCRC then transmits the relevant portions of its buffer to the DAQ computer, which then saves the data to disk for later analysis. Due to various latencies in this chain, data in the buffer may be overwritten before it can be transmitted. The DAQ system recognizes when this occurs and can drop remaining trigger points which the buffer has overwritten. When the rate of triggering is high, this occurs more frequently. This process effectively reduces the detector livetime from the nominal experimental exposure time.

The efficiency of this process can be estimated again using the set of random triggers in each data series. The true rate of events in the detector can be calculated from their frequency of occurrence in randomly triggered traces. Comparing this to the rate of triggered events saved to disk gives an estimate of the write efficiency. This was
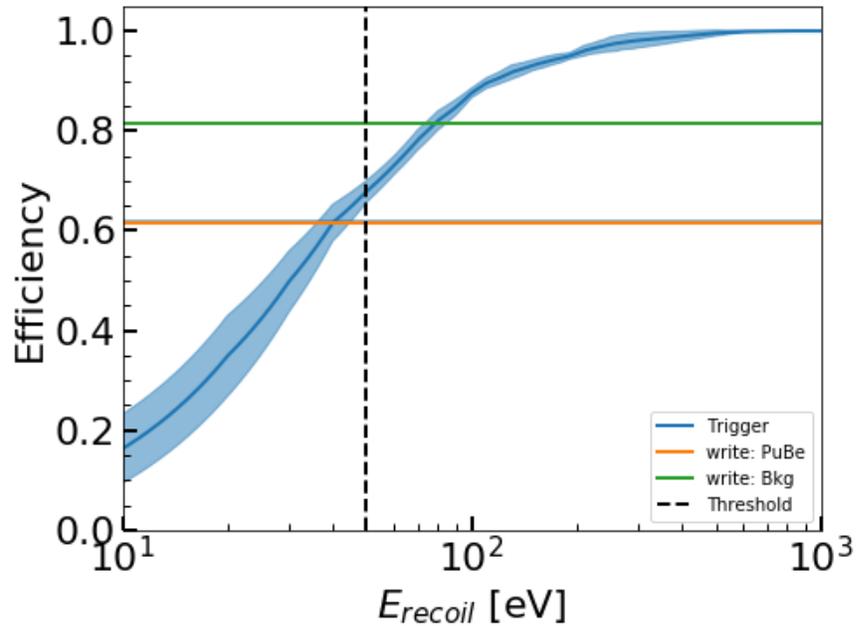
Figure 7.15: Trigger and write efficiencies.

calculated using events with energies well above the detector noise so there were no trigger efficiency effects involved, and there was practically no change of missing a pulse in a randomly triggered trace. The calculated efficiency was assumed to hold for all events due to the energy-independent nature of the trigger dropping process. This efficiency was calculated individually for each series and the average values were found to be $0.617 \pm 0.004$ ($0.815 \pm 0.004$) for the PuBe (Bkg) datasets used in this analysis. One could think of this as an effective livetime efficiency. However, we instead calculate the experimental livetime as the total detector exposure time and take into account effects like this DAQ efficiency and pileup effects only when considering the measured data.

### 7.2.6 Energy Calibration

Energy calibration of the total phonon signal to the ER scale was performed using several low gamma-ray and fluorescence lines associated with the $^{241}$Am calibration source. Fig. 7.16 shows the background spectrum used for calibration. A similar spectrum was

used to calibrate the PuBe data, but the calibration lines were more obscured by the increased PuBe-associated event rate.
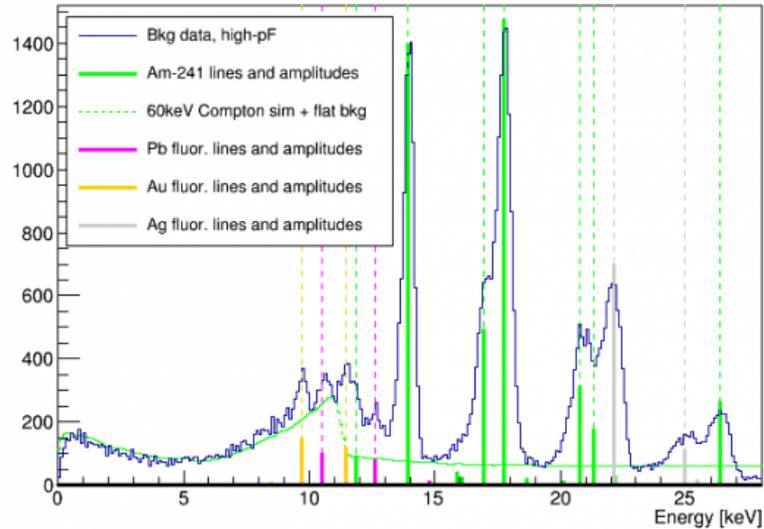


Figure 7.16: Calibration lines in the background energy spectrum. Figure produced by M. Fritts.

The rich set of calibration peaks was also used to develop the detector resolution model. This was defined as

$$\sigma_{ee}^2 = \sigma_0^2 + (B + \varepsilon_\gamma F)E_{ee} \tag{7.4}$$

where $\sigma_0$ is the baseline resolution, $F$ is the Fano factor [131], and $B$ incorporates any additional detector resolution effects. The widths of the identified ER calibration lines and baseline were measured and used to determine the best-fit values of $\sigma_0 = 10.27$ eVee and $B = 1.77$ eV$_{ee}$ as shown in Fig. 7.17. The lines shown as black points were not included in the fit because they were anomalously wide. The 12 keV gold fluorescence line was widened by a nearby $^{241}$Am line. The 17, 21, and 25 keV lines were all the left line of a double/triple set of lines, and all seemed to have extra low-energy tails. These tails were likely either associated with the same near-surface effects discussed in Sec. 5.1.2 or due to regions of the detector which had reduced NTL gain due to electric field variations.

$F$ describes the smearing of the signal due to the statistical variation in number of e/h pairs produced in each event which can, in general be different for ERs and NRs. For this fit, we used a value of F=0.1161 [132] since all peaks were from ERs. In later analysis, we describe the resolution using these fitted $\sigma_0$ and $B$ parameters along with an $F$ value appropriate for the event type.
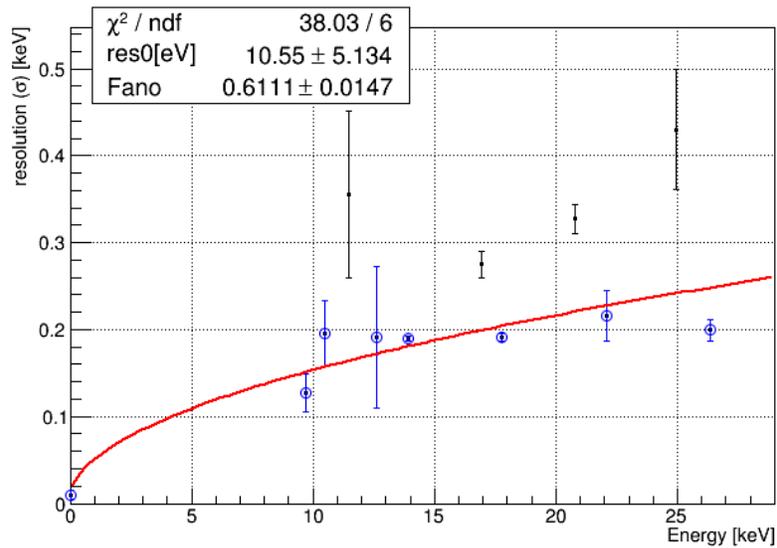


Figure 7.17: Detector resolution model fit to Bkg spectrum. In this parameterization, "res0"=$\sigma_0$ and "Fano"=$B/\varepsilon_\gamma + F$. Only blue points were included in the fit. Figure produced by M. Fritts.

### 7.2.7 NaI Detectors

The original conception of this $(n, \gamma)$ experiment also included the observation of $\sim$MeV escape gammas observed in external NaI detectors. The coincidence of signals in the NaI and HV detectors would bolster the identification of signals in the HV detector as nuclear recoils of particular energies. For UMN R68, we operated two 8" diameter, 8" thick cylindrical NaI detectors mounted near the cryostat and read out with photomultiplier tubes (PMTs). Unfortunately this arrangement only provided a low solid angle ($\sim$ 1%) coverage of the HV detector so it was not practical to use them for coincidence triggering and we did not have high enough statistics to observe a significant coincidence

population.

However, this run allowed an opportunity to develop and tune the NaI operation and coincident readout arrangement. In order to digitize the NaI signals using the MIDAS DAQ, they were fed into a spare DCRC to be digitized and synchronized with the HV detector signals. Because NaI signals are much faster ($\sim$ ns) than the 1.25 MHz DCRC digitization rate, the pulses were first passed through a pulse-widening amplifier before being sent to the digitizer. This shaping only introduced a delay on the order of a few $\mu$s, less than the phonon pulse rise time. Another parameter to optimize was the PMT bias voltage. The higher this was set, the greater gain the PMT obtained, however, at too high of values, the signals would saturate the pulse-shaping amplifier, with sometimes confusing effects for different amplifier models. In the end, there were at least 4 different configurations of NaI settings and amplifiers tested during the PuBe data taking campaign.

NaI spectra were taken with a variety of source configurations to ensure proper energy calibration. Fig. 7.18 shows the comparison of several spectra along with some identified line sources. It also shows the remarkable linearity of the calibration from pulse height to known gamma energy in the NaI detector. The two detectors used were observed to be quite linear up to the highest positively identified calibration line of 4.4 MeV. It is believed that this trend continues to at least $\sim$ 7 MeV. Future work should be done to simulate and identify the higher energy signals in these datasets and confirm the NaI detector calibrations up to the $\sim$10 MeV range expected from the $(n, \gamma)$ process.

Using the NaI and HV data taken, we searched for coincident $(n, \gamma)$ signals. Unfortunately, we did not observe a significant fraction of coincident events compared to that expected for random coincidence in the NaI and HV detectors. However, the work we did perform will be useful in planned (and NSF-funded!) extensions of this experiment which will feature large arrays of external NaI detector panels for more complete ($\gtrsim$25%) solid angle coverage.
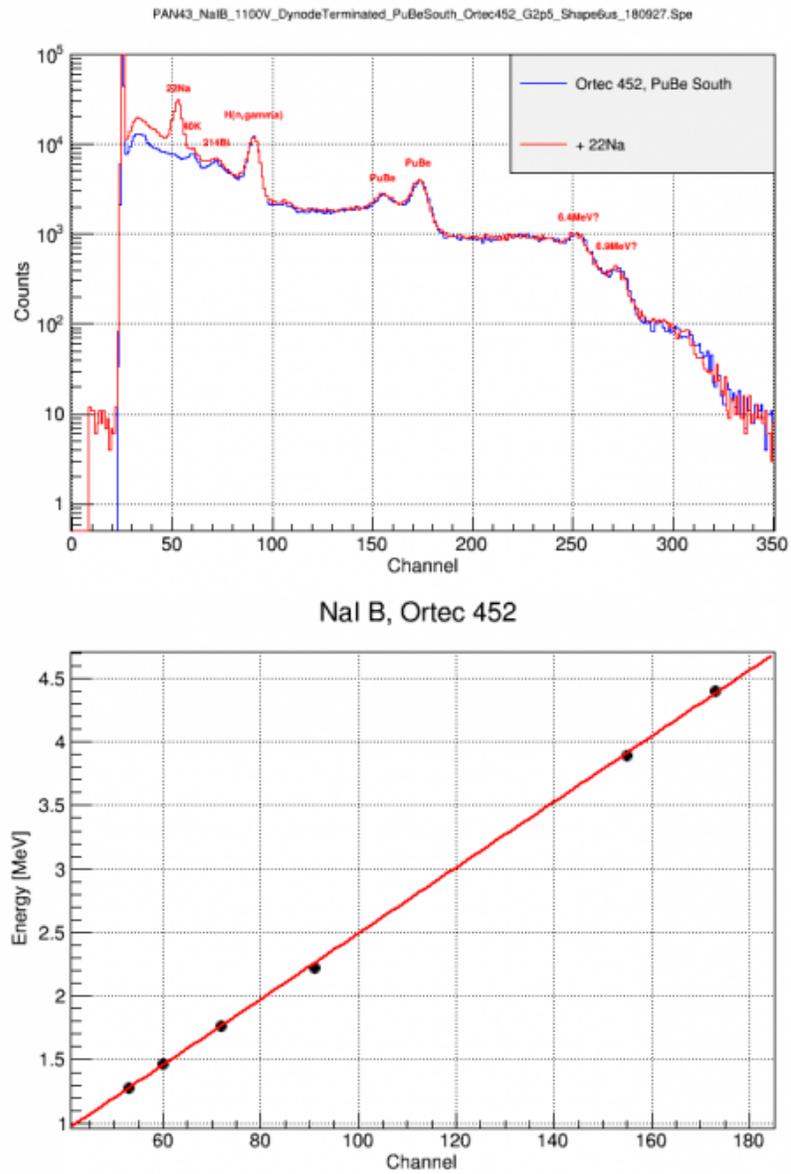
Figure 7.18: Example NaI spectra (top) and calibration curve (bottom).

## 7.3   Simulations

Two physics simulations were performed by A. Villano for this analysis and were used for comparison to measured data. The first was a *Geant4* [189] simulation of the PuBe source placed in a geometry describing UMN K100 fridge and lab. The PuBe source neutron spectrum was taken from [219]. For each simulated event, a neutron energy was drawn from the distribution and propagated through the geometry with scattering and/or capturing on the HV detector in the cryostat. Neutrons are produced by $^9\text{Be}(\alpha,\text{n})^{12}\text{C}^*$ reactions in the PuBe source. Each time a neutron was thrown from the source, a corresponding gamma was also thrown based on a simple model of the allowed nuclear decays of the excited $^{12}\text{C}^*$ daughter.

While some details of this simple gamma model could be improved (e.g. direct measurement of a similar PuBe source in [220] did not show the 3.2 MeV line at the level assumed in our model), we are not too concerned with it because the substantial amount of Pb shielding effectively blocked all direct gammas from the source. What is more important is that the neutron details are modelled correctly. This includes the initial spectrum, transport through shielding and fridge materials, and interactions with the detector. This also includes secondary processes such as gammas resulting from neutron capture on various shielding and cryostat materials. Previous studies of simulated neutron and gamma from AmBe sources measured by Ge detectors have found that material impurities in experiment, such as certain isotopes of Ti and Cr present in stainless steel of the cryostat, can be important due to their large neutron cross sections [221]. As we will see later in this analysis, there are some hints that the neutron propagation we have simulated is not in complete agreement with our results, possibly due to such effects.

A separate simulation was performed to model the details of the nuclear cascades and recoils which occurred after neutron capture on the Si detector nuclei. The most common Si isotopes ($^{28}\text{Si}$, $^{29}\text{Si}$, $^{30}\text{Si}$) were included with their natural abundances and were assumed to capture neutrons at relative rates set by their respective thermal neutron capture cross sections. Each has a set of many nuclear levels available for the excited nucleus to decay through following neutron capture. In general, they decay to the ground state via a cascade process, with the nucleus emitting a gamma and recoiling
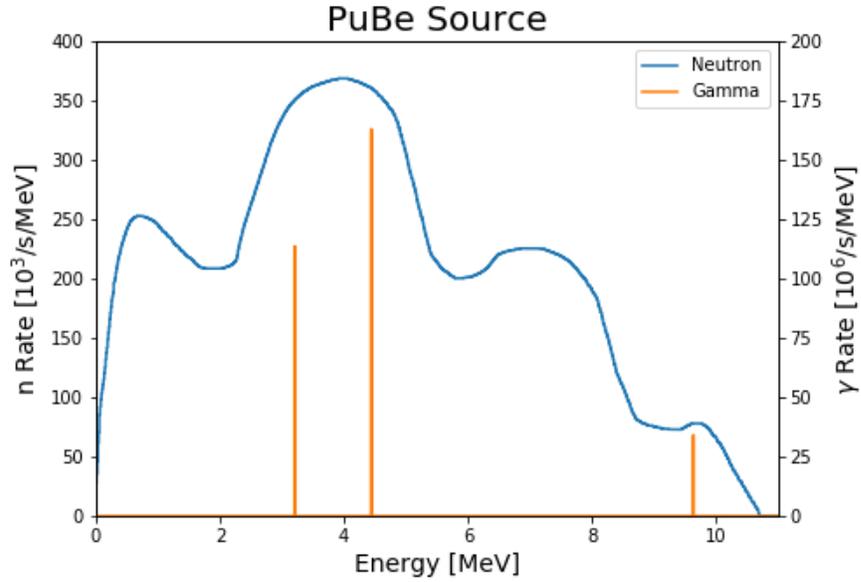
Figure 7.19: Spectra of PuBe source thrown in G4. Neutron spectra taken from [219]. Gamma energies were deduced from thrown neutron energy and $^{12}$C energy levels.

to conserve momentum at each step. There is an interplay between level lifetime and nuclear stopping time during the cascade process. If the nucleus recoils again before it has stopped from a previous recoil, the energy deposited in the crystal, $E_{dep}$, will be different. For a simple 2-step cascade, we can consider a few illustrative scenarios. If the decay time $t_d$ of the second decay (which is drawn from a distribution described by each level's half life) is much longer than the stopping time from the first recoil, $t_{stop}$ then

$$E_{dep}(t_d) = \frac{E_{\gamma,1}^2 + E_{\gamma,2}^2}{2mc^2}, \quad t_d > t_{stop} \tag{7.5}$$

where $E_{\gamma,i}$ is the energy of the i$^{\text{th}}$ gamma emitted and $m$ is the nucleus mass. However, if $t_d < t_{stop}$, then the deposited energy can range from

$$E_{dep,min} = \frac{(E_{\gamma,1} - E_{\gamma,2})^2}{2mc^2} \tag{7.6}$$

to

$$E_{dep,max} = \frac{(E_{\gamma,1} + E_{\gamma,2})^2}{2mc^2} \tag{7.7}$$

depending on the angle and timing of the second recoil with respect to the first.

This process can be calculated easily in a Monte Carlo simulation by requiring momentum and energy conservation at each step. This was done for the current study. However, a full analytic model of the process will allow for faster computation of recoil energy distributions. This calculation has been completed for a two-gamma cascade sequence in Appendix F. The general case of a multi-level cascade is a bit more difficult, but there are current fruitful efforts to complete this as well. The result of these simulations is a set of cascade events, each having multiple energy depositions from NRs and associated gammas. The simulation also describes whether the outgoing gammas scatter in the detector, in which case they would cause an ER as well. Such interactions will almost exclusively result in total signals higher than our 2 keV$_{ee}$ analysis range, and so are discarded later in analysis.

## 7.4   Fitting

### 7.4.1   Measured Spectrum and Background Subtraction

For this analysis we compare the simulated spectrum in the detector to the background-subtracted measured PuBe data. As demonstrated in Fig. 7.20, this was done by first calculating the reconstructed PuBe and Bkg spectra by dividing the measured count rates by the respective livetimes and estimated efficiencies after applying quality cuts. By including the efficiency functions in this calculation, we are in effect, reconstructing the true rate of events in the detector. Then, the reconstructed background rate was subtracted from the reconstructed PuBe rate, resulting in an estimate of the spectrum of events due to the presence of the PuBe source, which we call the reconstructed background-subtracted rate. In this calculation we include both statistical uncertainties from raw count rates and systematic uncertainties from cut, trigger, and other efficiency calculations.
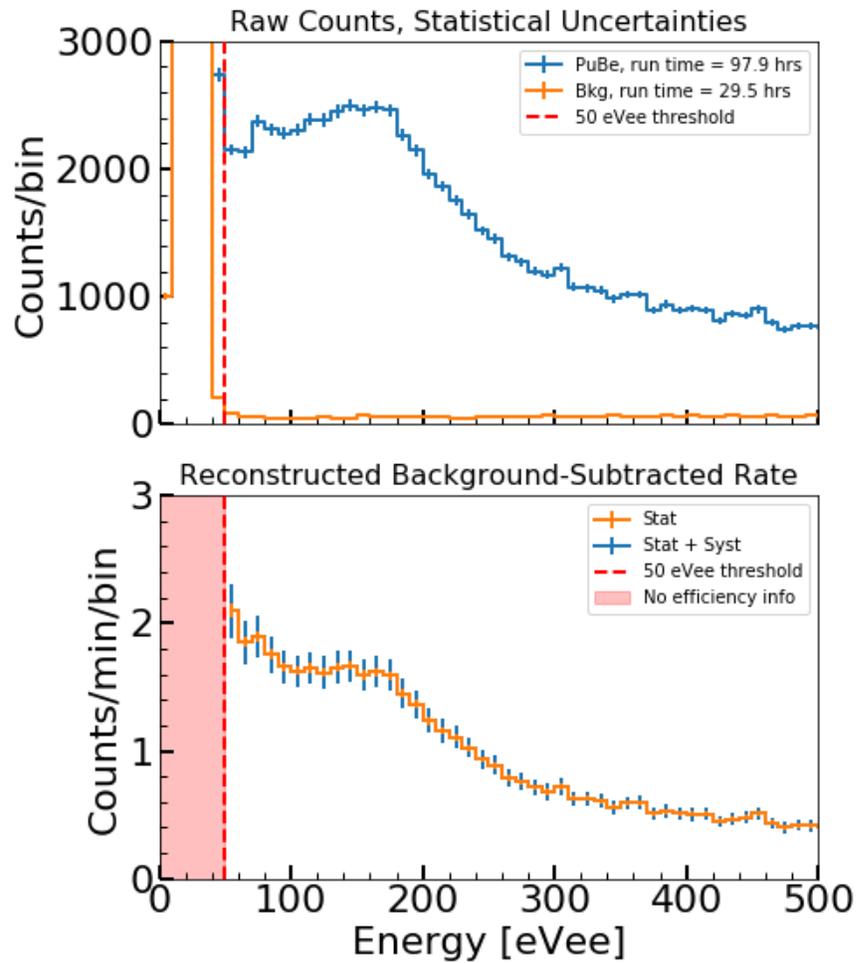
Figure 7.20: Top: Measured spectra for PuBe and Bkg data. Bottom: Reconstructed, background-subtracted PuBe rate.

### 7.4.2 Simulated Spectrum

The output of our simulations was a long set of particle interactions in the SuperCDMS HV detector. To compare these to our measurements, we first converted the hits into events at the detector and calculated the total phonon energy measured for each event. This required inclusion of material properties, such as the Fano factor and yield, as well as accounting for phonon NTL gain and some detector effects.

#### Phonon Energy

The signal measured in the SuperCDMS HV detectors is the total phonon energy produced in each event. Multiple hits from a single primary generally occur faster than the cryogenic detector time resolution, so they were grouped together to produce a single "event". The total measured event energy was the sum of total phonon energies due to individual hits. The total phonon energy of each hit was calculated as the sum of recoil, recombination, and NTL phonons produced.

To calculate this for simulated data, we treated each type of interaction separately. Because we calibrate our detector to the electron-equivalent energy scale ($eV_{ee}$), treating the simulated ERs is the simplest. For each event, we sum the deposited energy from the individual particle hits in the detector to get the total deposited energy in that event. Then, the electron-equivalent energy signal is

$$E_{ee}^{ER} = \sum_{hits} E_{hit} \tag{7.8}$$

For elastic NR events, we used a yield model (described below) to calculate the NTL phonon contribution and get a total phonon energy, $E_{tot}$, for each deposited hit energy, $E_{hit}$ as

$$
\begin{aligned}
E_{tot}(E_{hit}) &= E_{hit} + E_{Ion}eV/\varepsilon_\gamma \\
&= E_{hit}(1 + Y(E_{hit})eV/\varepsilon_\gamma)
\end{aligned}
\tag{7.9}
$$

where $V$ is the detector bias voltage and $\varepsilon_\gamma = 3.8$ eV in Si [130]. $E_{Ion} = E_{hit}Y(E_{hit})$ can be thought of as the amount of energy that goes into charge production. The $E_{tot}$ energies from all hits of an event are then summed to obtain the total event phonon energy. To convert this to the measured electron-equivalent value, we divide by the

assumed NTL gain for an ER, $G_{NTL} = 1 + eV/\varepsilon_\gamma$. This gives an electron-equivalent event energy of

$$E_{ee}^{NR} = \sum_{hits} E_{hit} \frac{1 + Y(E_{hit})eV/\varepsilon_\gamma}{1 + eV/\varepsilon_\gamma} \tag{7.10}$$

Treating the signals due to $(n, \gamma)$ is a little more subtle because they involve composite NR cascades. A. Villano has a nice explanation of this which I reproduce here. We assumed that such a recoiling nucleus in some cascade step begins at energy $E_0$ and deposits energy $\delta E$ before the next recoil occurs. We would then expect the ionization produced in that step to be that which would have been given from a single recoil of energy of $E_0$ less that which would have been given from a single recoil of $E_0 - \delta E$. This simply assumes that the ionization is independent of the slowing-down history of the particle, or, that any recoil of energy $E$ will give a reproducible amount of ionization regardless of the circumstances of its creation. This means that the amount of ionization energy in the $i^{th}$ cascade step is

$$E_{ion,i} = E_{0,i}Y(E_{0,i}) - (E_{0,i} - \delta E_i)Y(E_{0,i} - \delta E_i) \tag{7.11}$$

and the total electron-equivalent signal becomes

$$E_{ee}^{ng} = \sum_i \frac{\delta E_i + E_{ion,i}eV/\varepsilon_\gamma}{1 + eV/\varepsilon_\gamma} \tag{7.12}$$

where $i$ sums over recoils in the cascade. There is also the possibility of events with both ER and NR components. This was studied and it was found that $< 0.05\%$ of the simulated events above threshold are of this type, so they are not expected to affect the simulated spectrum.

**Yield Models**

We considered several physically-motivated functional yield forms in this analysis. The most commonly used model is that of Lindhard [222] (simplified form in [83]),

$$Y_L(E_R, k) = \frac{kg(\varepsilon)}{1 + kg(\varepsilon)} \tag{7.13}$$

with $k$ free and

$$g(\varepsilon) = 3\varepsilon^{0.15} + 0.7\varepsilon^{0.6} + \varepsilon$$
$$\varepsilon = E_R(keV)11.5Z^{-7/3} \tag{7.14}$$

where $Z = 14$ for Si. Note here that $\varepsilon$ is a rescaled recoil energy quantity not to be confused with $\varepsilon_\gamma$. This model is the standard in a range of materials at high energy, but is expected to break down in the low energies considered for this analysis. Several empirical and theoretical extensions have been made to this model which generally include some form of low-energy cutoff behavior. We define here each of the models used in this analysis. The Chavarria model, with $k$ and $a$ free, is described by

$$Y_{Chav}(E_R, k, a) = \left( \frac{1}{aE_R} + \frac{1}{Y_L(E_R, k)} \right)^{-1} \tag{7.15}$$

following the functional form published in [119]. The Sorensen model attempts to include the effects of a finite ionization threshold as

$$Y_{Sor}(E_R, k, q) = \max(Y_L(E_R, k) - q/\varepsilon, 0) \tag{7.16}$$

where the max function ensures the function is non-negative and we leave $k$ and $q$ as free parameters [223]. Finally, we also considered the Adiabatic Correction model, described as

$$Y_{AC}(E_R, k, \xi) = (1 - e^{-E_R/\xi})Y_L(E_R, k) \tag{7.17}$$

with $k$ and $\xi$ free as in [224]. A comparison of the various yield model shapes is shown in Fig. 7.21. It can be seen that all converge to the Lindhard model at high energies, but can cut off with different shapes.

**Fano Resolution**

After calculating the set of simulated electron-equivalent event energies, we generated the spectrum of measured energies including statistical smearing. An event produces some number, $N$, of e/h pairs with

$$\langle N \rangle = E_{ion}/\varepsilon_\gamma \tag{7.18}$$

The Fano factor, $F$, describes the statistical variance as

$$\sigma_N^2 = F\langle N \rangle \tag{7.19}$$

with $F$=0.12 – 0.15 for ERs below a few keV [132]. The Fano factor for NRs, $F_{NR}$, may be different and has not been well-characterized at low energies. It could in fact be
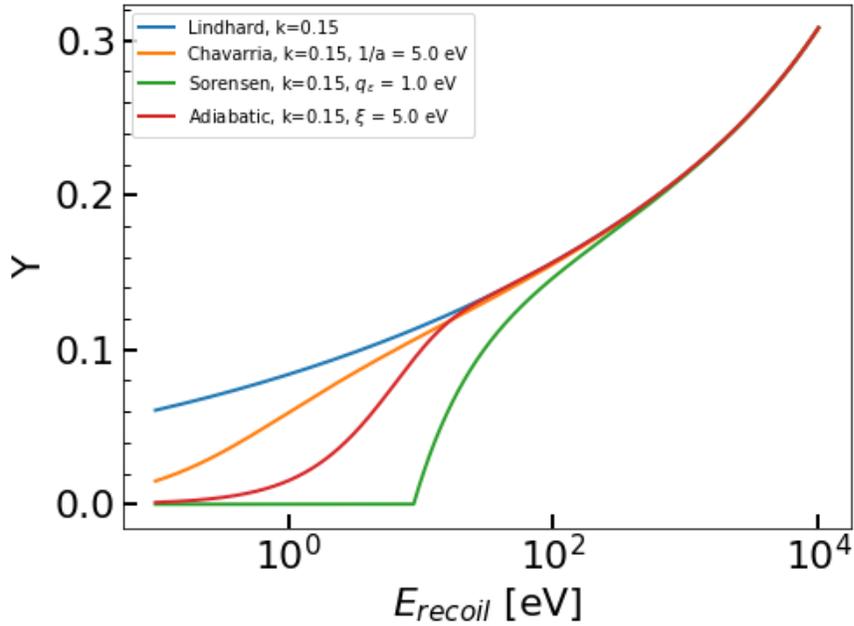
Figure 7.21: Example yield model shapes.

higher than the ER Fano factor by an order of magnitude [225], making the NR signals much more smeared out. In this analysis, we will generally leave it as a parameter to be fit. Here, as in many analyses, we assume that the distribution of $N$ is a Gaussian centered at $\langle N \rangle$ with width described by $\sigma_N$. There is evidence that the distribution is somewhat skewed [226] but we do not take such details into account here.

Now, including the Fano spreading could be done by randomly drawing the actual number of e/h pairs produced from the distribution described by $\langle N \rangle$ and $\sigma_N$ for each event, but this would just realize a single instance of the experiment. This process would need to be repeated many times to be sure any systematic effects are averaged out. However, this is computationally expensive so we used an alternative approach and simply filled the final simulated histogram using the weighted probability distribution of each event. In this way a single event will contribute a Gaussian distribution to the eV$_{ee}$ spectrum with a mean value, $\langle E_{ee} \rangle$, set by the event energy and eV$_{ee}$ width set by the Fano factor. This has the same effect as averaging many individual realizations of the spectrum.

To implement this, we first calculate the Fano contribution to the electron-equivalent peak width for each hit. As described in Appendix G, the expression for this Gaussian width is

$$
\begin{aligned}
\sigma_{E_{ee}}^2 &= (eV)^2 F \langle N \rangle / (1 + eV/\varepsilon_\gamma)^2 \\
&= (eV)^2 F \frac{\langle E_{Ion} \rangle}{\varepsilon_\gamma} / (1 + eV/\varepsilon_\gamma)^2
\end{aligned}
\tag{7.20}
$$

where the appropriate Fano factor is used for either ER or NR and the mean ionization energy is calculated as described above. Note that for $eV/\varepsilon_\gamma \gg 1$, we recover the familiar $\sigma_{E_{ee}}^2 = F \langle N \rangle \varepsilon_\gamma^2$. We also make the assumption that the individual hits or steps in an event are uncorrelated. While this may not be strictly true, it allows us to calculate the total statistical width of an event as the sum of all the widths of that event's hits or steps as

$$
\sigma_{eV\text{ee},tot}^2 = \sum_{hits} \sigma_{eV\text{ee},hit}^2.
\tag{7.21}
$$

At this point we have the full description of the electron-equivalent distribution for each event. This is what would be measured if the experiment were repeated a large number of times. Because these distributions take a nice analytical form, we can directly compute the contribution of each simulated event to the final, binned electron-equivalent spectrum by integrating the Gaussian over each bin width as

$$
\begin{aligned}
w_i &= \int_{E_i}^{E_{i+1}} \frac{1}{\sqrt{2\pi\sigma_{E_{ee}}^2}} \exp\left[-\frac{(E - \langle E_{ee} \rangle)^2}{2\sigma_{E_{ee}}^2}\right] dE \\
&= \frac{1}{2}\text{erf}\left[\frac{\langle E_{ee} \rangle - E_i}{\sqrt{2}\sigma_{E_{ee}}}\right] - \frac{1}{2}\text{erf}\left[\frac{\langle E_{ee} \rangle - E_{i+1}}{\sqrt{2}\sigma_{E_{ee}}}\right]
\end{aligned}
\tag{7.22}
$$

where $w_i$ is the weight in the $i^{th}$ bin which has bin edges $[E_i, E_{i+1})$. This expression is used to calculate the contribution of each event to each energy bin. Finally, the contributions from all events are summed to produce a single energy spectrum.

**Detector Resolution**

As described in Sec. 7.2.6, the detector energy resolution was parameterized with

$$
\sigma_{ee}^2 = \sigma_0^2 + (B + \varepsilon_\gamma F) E_{ee}
\tag{7.23}
$$

Now, this measurement includes the width due both detector effects and Fano broadening. We have described above how we account for the Fano effect for different event types, so we extract the intrinsic detector resolution as

$$\sigma^2_{E_{ee},Det} = \sigma^2_0 + BE_{ee} \tag{7.24}$$

The resulting detector resolution function is displayed in Fig. 7.22.
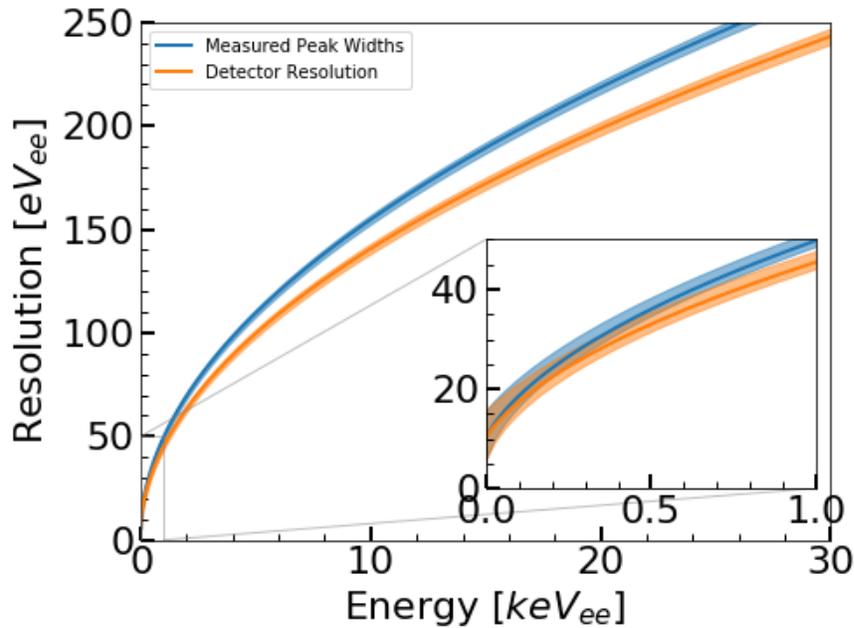


Figure 7.22: R68 detector resolution function. Shaded regions are due to fit parameter uncertainties.

We find here that, for energies less than 1 keV$_{ee}$, the intrinsic detector resolution is dominant. We apply this detector resolution by convolving the event energy spectrum with a Gaussian kernel which has an energy-dependent width set by the resolution function. This means that we take the contents of each bin of the event energy histogram and redistribute it according to the resolution function. The effect of this on the various components of the simulated spectra can be seen in Fig. 7.23. In this example, we used the Lindhard yield with $k = 0.15$ and $F_{NR} = 2$. Applying the detector resolution acts as a low pass filter on the energy spectrum, averaging over structures below the resolution scale. This results in a smoother ER spectrum. The elastic NR spectrum is not affected

much, because it has no real structure smaller than the resolution. The $(n, \gamma)$ signal peak height is somewhat reduced, but the total integral unchanged.
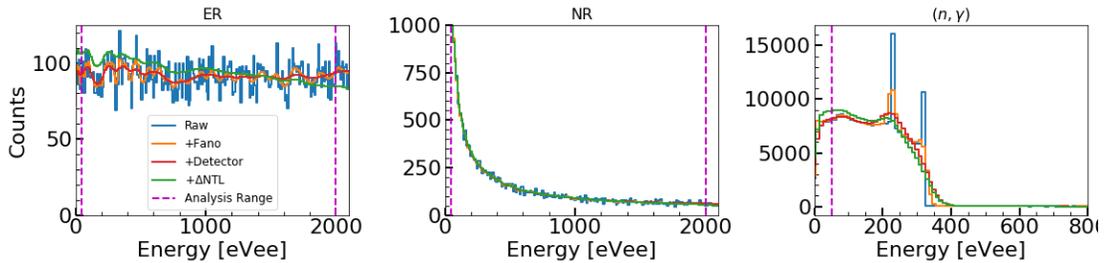


Figure 7.23: R68 detector resolution effects on the simulated energy spectra.

### $\Delta$NTL Model

In previous analyses of similar CDMS detectors, it was important to include the effect of the variation in NTL gain throughout the crystal [101]. The issue is that the bias field is not uniform in the entire detector volume, with some fringing effects near the edges. Events that occur in those regions will not experience the full NTL gain and have a reduced signal. In the present analysis, we have treated this as a systematic uncertainty and developed a simple model to estimate the size of the effect.

We designed a simple detector and housing model in COMSOL Multiphysics$^\circledR$ [227] and used it to calculate the electric field distribution for the 1-sided biasing configuration of this experiment. The model included a cylindrical Si detector with planar electrodes in a cylindrical copper housing. The real housing geometry is hexagonal with more complicated internal geometries, but for this purpose, it was sufficient to model this as a cylinder with inner diameter the same as the smallest inner diameter of the true housing. In this way we overestimated the influence of the grounded housing by placing more of it closer to the crystal. Fig. 7.24 shows the electric field computed in the detector with the upper electrode at -125 V and the bottom electrode and housing at ground. The important features are that the field is not uniform, but has fringes near the edges due to the grounded housing. There are some small meshing artifacts near the top of the detector which became smoother with finer meshing and have little affect on the results here.
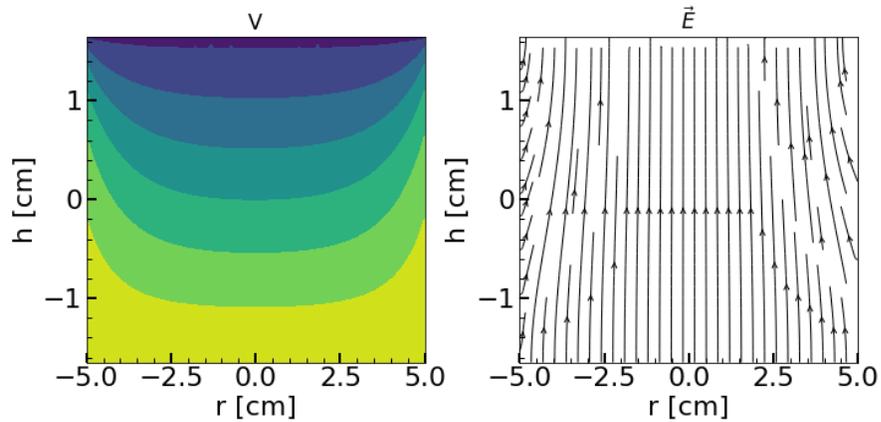
Figure 7.24: R68 detector bias field model. Left: Voltage as a function of radius and height. Right: Electric field lines.

We next created a simple model of charge transport in the detector by starting an event at a given location and following the electric field until a detector surface was reached. This was a relatively fast calculation and is probably a decent approximation given the constant drift speed experienced by carriers in Si. Examples of such tracks are shown in the left plot of Fig. 7.25. There it is clear that events too near the sidewalls may not traverse the full crystal and thus produce less NTL energy.
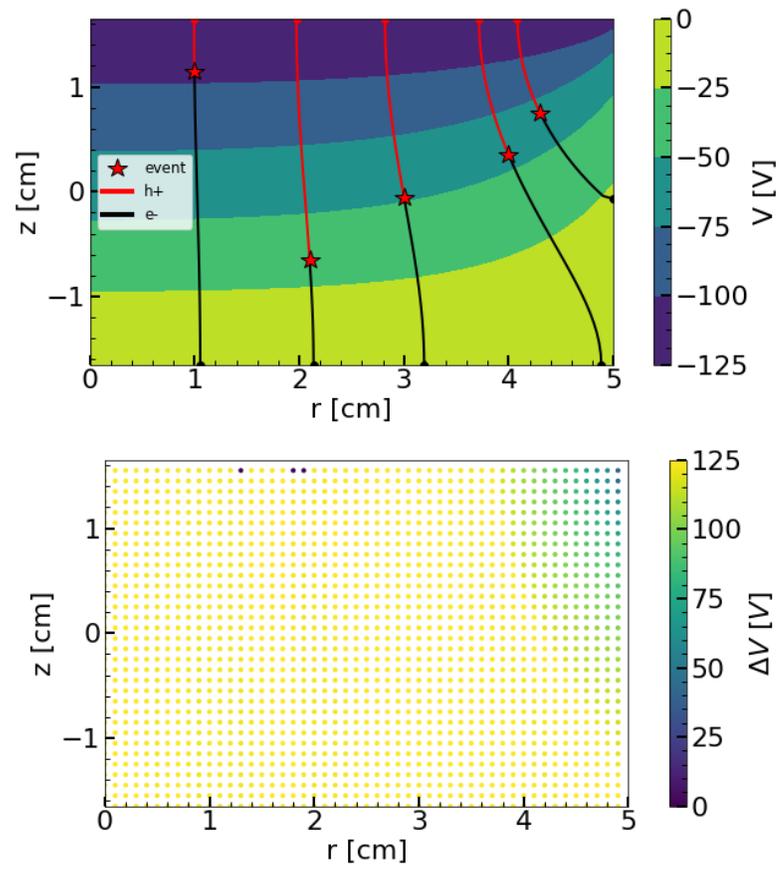
Figure 7.25: Top: R68 $\Delta$NTL simulation examples. Bottom: Map of $\Delta V$ calculations.

We generated a grid of starting positions in a single $r - z$ slice of the detector and simulated transport from each. We then calculated the total potential difference, $\Delta V$, experienced by the carriers in each event. This was simply the difference in potential at the endpoints of the electron and hole.

$$\Delta V = |V_{final}^{h+} - V_{final}^{e-}| \tag{7.25}$$

A map of those values is shown on the right side of Fig. 7.25. As expected there is a smooth trend with most events away from the upper corner experiencing the full voltage difference. There are also a few failed tracks due to the aforementioned meshing artifacts.

So how do ERs and NRs get reduced by this effect? They should each be reduced by the factor $f_R$, which is the ratio of the true NTL gain to the nominal value

$$f_R = \frac{G_{NTL}(\Delta V)}{G_{NTL}(V)} \tag{7.26}$$

This works a little differently for ERs and NRs due to the yield, but for the purposes of just characterizing the systematic uncertainty associated with this effect, we can make the following argument. If $Y \leq 1$ and $\Delta V \leq V$ then it follows that

$$f_R = \frac{1 + Y\Delta V/\varepsilon_\gamma}{1 + YV/\varepsilon_\gamma} \geq \frac{1 + \Delta V/\varepsilon_\gamma}{1 + V/\varepsilon_\gamma} \geq \Delta V/V \tag{7.27}$$

So reducing NRs by the same ratio as for ERs distorts them more than is needed. This makes it a conservative estimate of the size of the smearing effect. We then used the calculated $\Delta V/V$ values to construct a simple signal smearing model which could be applied to simulated energy spectra. This was done by first fitting the distribution of $\Delta V/V$ values as shown in Fig. 7.26. Because the grid of simulations was done on a single $r - z$ plane, we first weighted each point's contribution by the event radius, so as to obtain the correct weights for a uniform cylindrical volume.

Then we used a simple distribution model consisting of a delta function with an exponential tail. The shape of the tail was fit to the low $\Delta V/V$ distribution and the delta function was located at 1. By convolving this function with an energy histogram, we were able to estimate the amount of smearing due to variations in NTL gain. Examples of this effect for narrow energy lines are shown in the right side of Fig. 7.26. The effect
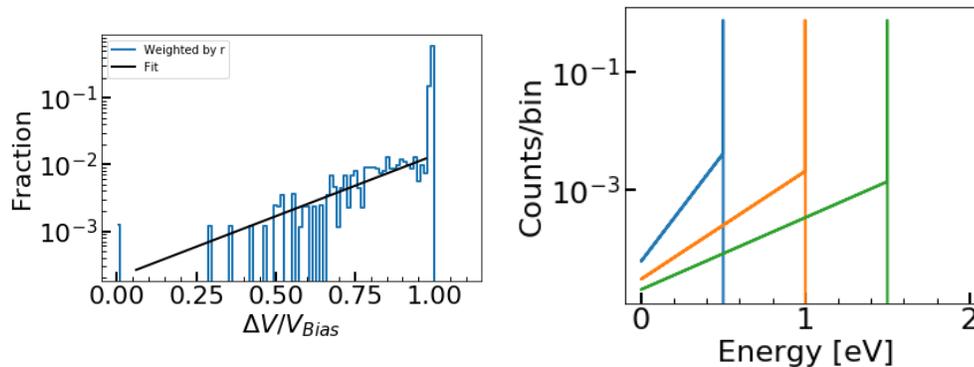
Figure 7.26: Left: Fitted histogram of weighted $\Delta V/V$ distribution. Right: Effect of $\Delta$NTL smearing model on single energy lines.

on the full simulated spectra is shown in Fig. 7.23. Again, this effect was only included for estimation of systematic uncertainties, and has been constructed to represent an overestimation of the effect.

### 7.4.3    MCMC Sampling Method

All of the above steps were used to build a simulated spectrum. In doing so, we had to include parameters such as the NR Fano factor and those that describe the NR yield function, which are actually the parameters of interest for this whole project. We sought the parameters that transformed our simulated hits into a measured spectrum consistent with the actual measurement we performed. We did this using a Markov chain Monte Carlo (MCMC) ensemble sampler called *emcee* [228]. The basic idea is that, given some model parameters, their prior probability densities, and a way to calculate a likelihood of our simulation agreeing with the measurement, the code will draw random values of the parameters, but in such a way that the distribution of those sampled values will converge to the parameters' posterior probability density functions (PDFs). A more in-depth explanation is beyond the scope of this work, but can be found in many places including Ref. [228].

**Priors and Likelihood**

Generally, we assumed no prior knowledge of the parameter values other than a broad, physically plausible range of values. We thus used a flat prior probability function within the given range. The log prior for a parameter $\theta \in [a, b]$ is then

$$\ln(\mathcal{P}_{flat}(\theta; a, b)) = \begin{cases} -\ln(b - a) & a \leq \theta \leq b \\ -\infty & \text{otherwise} \end{cases} \tag{7.28}$$

We also defined a likelihood to quantify how well a given simulated spectrum matches our measurement. Due to the various cuts and background subtraction procedure, the spectrum has significant systematic errors we wished to incorporate into the MCMC procedure. This prevented us from using a simple Poisson Likelihood as in Ref. [229]. Instead, to accommodate asymmetric uncertainties, we described the likelihood in each bin with a Split-Normal [230] function matched to the measured rate and uncertainties in each energy bin. The formula for the Split-Normal log likelihood function is then,

$$\ln(\mathcal{L}_{SNorm}(\vec{k}; \vec{\mu}, \vec{\sigma}_{low}, \vec{\sigma}_{hi})) = \sum_i^N \left[ \frac{1}{2} \ln(\frac{2}{\pi}) + \right. \\ \left. - \ln(\sigma_{low,i} + \sigma_{hi,i}) - \frac{1}{2} \left( \frac{k_i - \mu_i}{\sigma_i} \right)^2 \right] \tag{7.29}$$

where $\vec{k}$ is the set of simulated rates under test (calculated over $N$ energy bins), $\vec{\mu}$ is the set of measured rates for each bin, $\vec{\sigma}_{hi}$ ($\vec{\sigma}_{low}$) is the uncertainty for points above (below) $\vec{\mu}$, and

$$\sigma_i = \begin{cases} \sigma_{low,i} & k_i <= \mu_i \\ \sigma_{hi,i} & k_i > \mu_i \end{cases} \tag{7.30}$$

This reduces to a standard normal distribution for $\vec{\sigma}_{hi} = \vec{\sigma}_{low}$. We performed each MCMC sampling twice for each yield model with different uncertainty, $(\vec{\sigma}_{hi}, \vec{\sigma}_{low})$, values. One calculation included only the statistical uncertainties resulting from the finite measured and simulated count rates, while the other included all systematic effects from the various efficiencies and effects such as the $\Delta NTL$ model above.

**Free Parameters**

There were several free parameters that the MCMC was tasked with sampling over. A different yield model was selected for each calculation with either one or two free parameters. If the Lindhard model was used, this would only be the factor $k$, with a flat prior requiring $0.05 < k < 0.3$ . The other three models all had an additional parameter describing the cutoff scale, either $a$, $q$, or $\xi$ for the Chavarria, Sorensen, or Adiabatic Correction models respectively. These were generally allowed to range between values corresponding to cutoff energies of $0 - 2$ keV$_{\mathrm{nr}}$. We also let the NR Fano factor, $F_{NR}$ float with a generous range of $0 < F_{NR} < 30$.

In addition to these parameters describing the microscopic physics, we also included several scaling factors as nuisance parameters in the calculations. These were included because we found that our nominal simulated livetimes for the ER and NR components (from the *Geant4* simulation) resulted in simulated signal rates that were somewhat lower than the measured values. While we have not yet identified the source of this discrepancy, we proceeded under the assumption that the simulated spectra at least have the correct shapes, though the overall scales may not be correct. This is probably a safe bet for the ERs, since the spectrum is expected to be dominated by Compton scatters of higher energy gammas and thus quite flat and featureless. However, the elastic NR distribution is a quasi-exponential as expected, but the details of this shape are more important as we discuss later. The $(n, \gamma)$ distribution was simulated separately and is believed to be trustworthy, though we include a scaling factor for it as well for maximum freedom of the fit. The factors $f_{ER}$, $f_{NR}$, and $f_{NG}$ were thus included to scale the rate of the ER, elastic NR, and $(n, \gamma)$ signals respectively.

**Interpreting Results**

The MCMC algorithm generates a chain of parameter values that the algorithm visited, an example of which is shown in Fig. 7.27 below. In practice, the calculation is parallelized with multiple "walkers" traversing the parameters space. In the results here, we used 128 such walkers in each calculation. We then assessed the convergence of the algorithm to extract some meaningful statements about the parameters of interest. Ref. [228] follows the advice of Ref. [231] and recommends using the autocorrelation

time for this purpose. The autocorrelation time, $\tau_i$, of each parameter, $\theta_i$, is a measurement of how many iterations of the algorithm needed to get uncorrelated samples of the posterior PDF. Basically, each evaluation of the algorithm selects new parameter values that are somewhat correlated with the previous values and $\tau_i$ tells us how many samples away we must go to escape that effect. These values are shown as red lines for each parameter in Fig. 7.27.
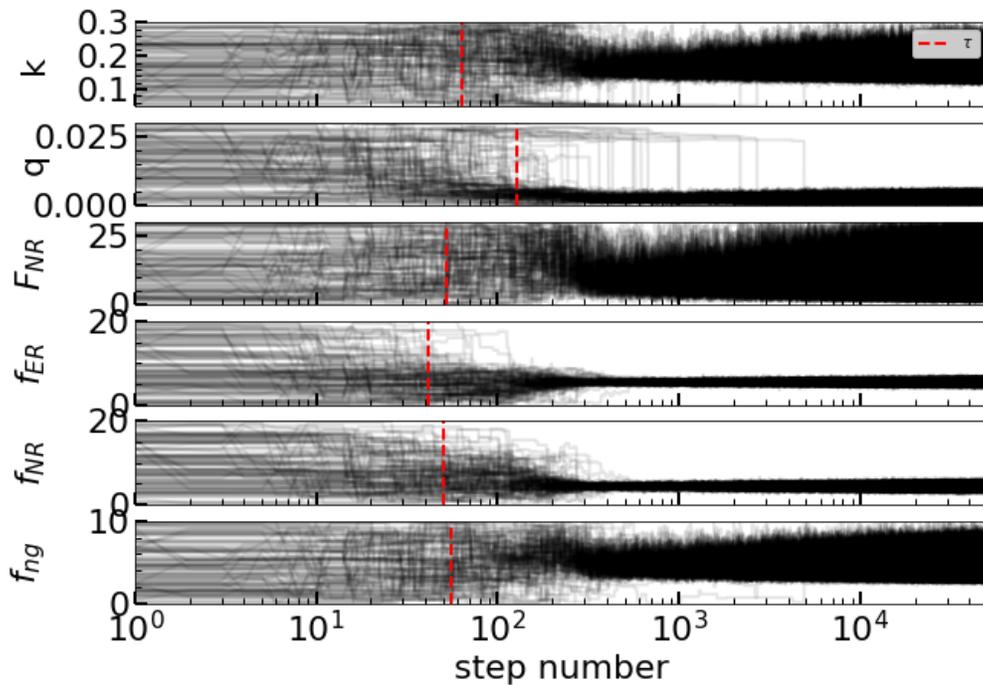


Figure 7.27: Example chain of MCMC sampler chains using the Sorensen yield model. This example used 128 walkers for 50k steps. Autocorrelation times are indicated for each parameter.

Once calculated, $\tau_i$ was used for several purposes. It can first be used to determine if the algorithm was run for long enough to accumulate statistically significant number of independent samples of the posterior PDF. In practice, the chain of values visited should be $\gtrsim 50 \times \tau_i$. If this is not the case, we simply ran the algorithm for a longer chain length. If the chain was long enough, then we took a sample of the chain values to

give an uncorrelated sampling of the parameter space. This is done by first discarding the first $2\bar{\tau}$ samples of the chain, where $\bar{\tau} = \sum_i^N \tau_i/N$ is the average length of all the parameters. This reduced initial correlation in the starting parameter values in a period sometimes called the "burn-in" phase. Then the chain was thinned by retaining only one value out of every $\bar{\tau}$ values.

At this point we evaluated the distribution of the remaining values which should give approximations to the posterior PDFs of the parameters. To examine these we generally produced a corner plot such as shown in Fig. 7.28. From there, one can examine numerical results such as the median and width of the parameter distributions. Such a figure demonstrates both the 1-dimensional parameter distributions as well as any correlations between parameters.
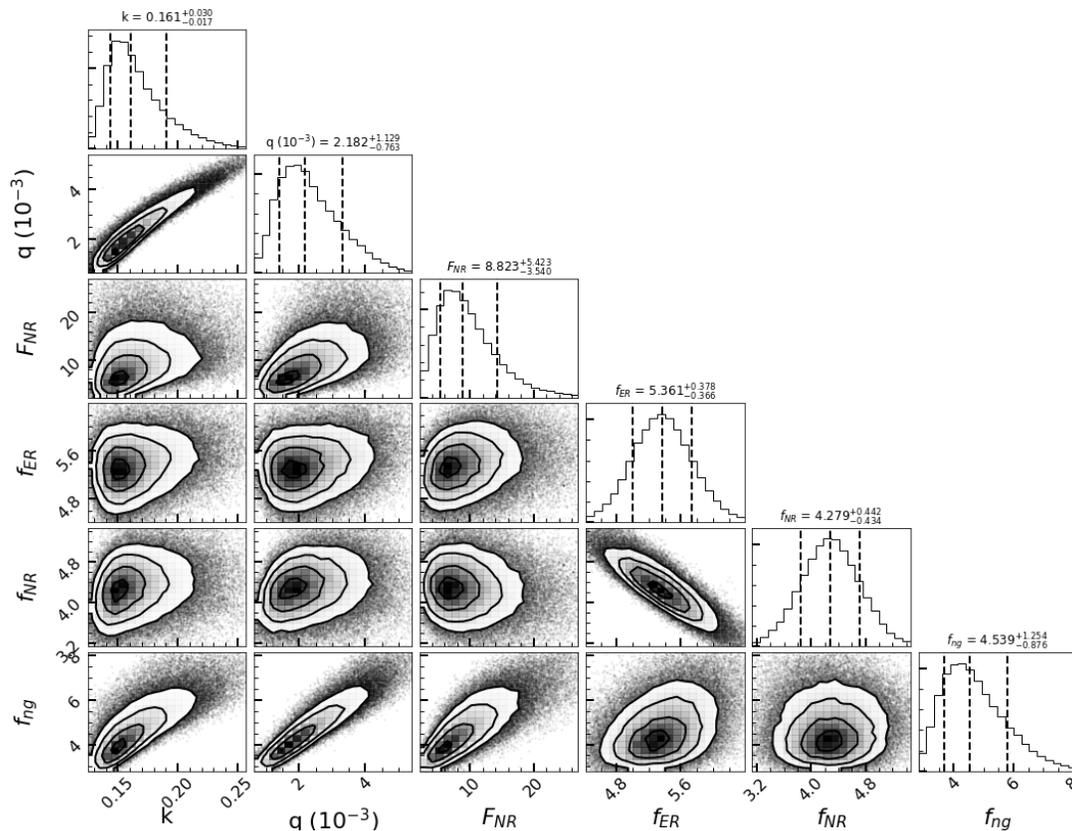


Figure 7.28: Example corner plot of MCMC sampler values. Vertical dashed lines on each 1-d histogram represent the median and 1-$\sigma$ equivalent percentiles.

The example shown in Fig. 7.28 was calculated using the Sorensen yield model which provided the best fit to the data. It demonstrates several common features observed in all models tested. First, there was a clear correlation between the Lindhard $k$ parameter and the yield cutoff parameter, here $q$. This is not surprising because increasing $k$ tends to increase the entire yield curve scale and moving the cutoff parameter to higher values will tend to pull the (low energy) yield back down. There was clear anti-correlation between $f_{ER}$ and $f_{NR}$. This is because these are the only two components at energies above $\approx 1\,\text{keV}_\text{ee}$, and both have relatively flat spectra there. So this is just indicating that their spectral sum is being kept at roughly constant value to match the measurement. The interplay between $f_{NG}$ and the parameters $k$, $q$, and $F_{NR}$ was more subtle. We will discuss this more when considering the resulting spectra below.

In general, we found that running the sampler for 5k – 50k steps was sufficient to obtain reasonable statistics for the tested yield models. We also found improved behavior when selecting sampler moves using an 80/20 mix of "DEMove" and "DESnookerMove" as suggested for multimodal problems in the documentation of [228]. With these settings, the step acceptance fraction, a measure of how efficiently the sampler selected new sample values, was generally 0.1 – 0.3, which is considered acceptable in such calculations.

With the multi-dimensional distributions of parameters displayed above, we can produce corresponding distributions of yield curve and simulated spectra. The left side of Fig. 7.29 shows yield bands generated from the MCMC sampler using the Sorensen yield model. To generate these bands we did not simply use the 1-d yield parameter distributions because this would ignore the valuable correlation information contained in the sampled distribution. Instead, we randomly selected 100 sets of sampled points from the distributions and calculated the 1-$\sigma$ equivalent width of the distribution of that set of yield curves. This was done for the calculations including just the statistical uncertainties in the likelihood as well as with all the previously discussed systematic effects.

These yield results are shown in comparison to several sets of previous measurements. One important feature is that this method generates a continuous yield curve which extends to low energy as opposed to a discrete set of measured points. This is possible because of two factors. First, we are assuming a functional form of the yield model,

so we are sensitive to its behaviour over the full range of our spectrum. Second, both the elastic NR and $(n, \gamma)$ spectra contain events which are composed of multiple, lower energy hits. In this way, well-measured events can be composed of several recoils, some or all of which may sample the yield curve below our *single-hit* analysis threshold. To demonstrate this, we have also included dashed lines denoting the yield parameters space (corresponding to the 50 $eV_{ee}$ to 2 $keV_{ee}$ analysis region) for which our experiment was sensitive to single-hit recoils.
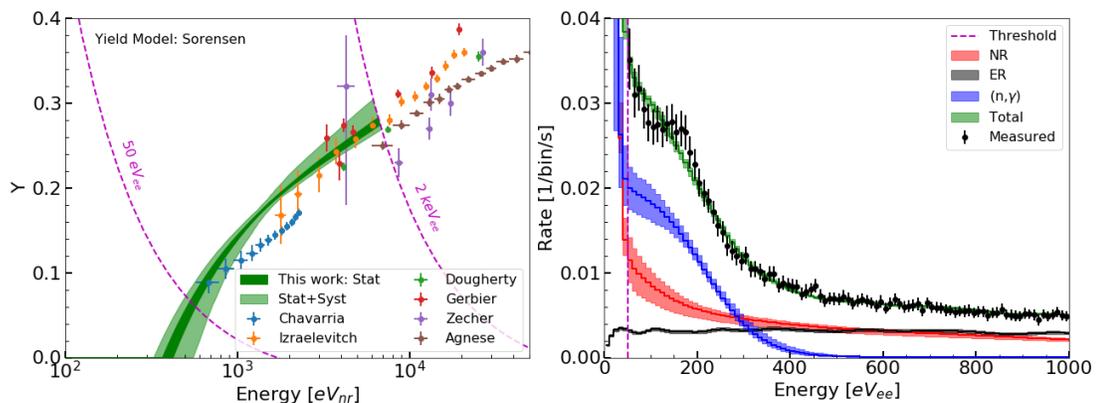


Figure 7.29: Left: Distribution of yield curves generated with MCMC using Sorensen model. The following previous measurements are shown for comparison: Chavarria [215], Izraelevitch [232], Dougherty [225], Gerbier [233], Zecher [234], Agnese [120]. Right: Corresponding range of best-fit background subtracted reconstructed spectra with systematic and statistical uncertainty. Shaded bands represent the 1-sigma equivalent range of rates in each energy bin.

Now, the yield curve shown here must be taken with a fair amount of context. While it was generated from the set of parameters which provided the best-fit spectrum, the absolute quality of that fit was not as good as we would like. The comparison of the measured and simulated spectra are shown on the right side of Fig. 7.29. Like the yield curves, the various spectral components were generated by sampling directly from the N-dimensional MCMC sample distribution with the band center and width denoting the mean and 1-$\sigma$ equivalent rates in each energy bin. We see here that the MCMC has done a fine job of matching the measured spectrum in most areas, except it has failed to reproduce the detailed shoulder near 200 $eV_{ee}$ which is associated with the $(n, \gamma)$ signal region.

At this point we are in a better position to understand some more of the correlations in the bottom row of Fig. 7.28. In all of the MCMC fits tried, the sampler tended to favor selecting a large value of $F_{NR}$ to smear out the otherwise sharp features of the $(n, \gamma)$ signal (cf. Fig. 7.23) and shift the resulting distribution (via $k$ and $q$). This was done to match the rising edge of the measured signal spectrum starting around 400 eV$_{ee}$ which the current elastic NR spectrum was unable to account for.

To further illustrate how much trust one should (or should not) place in the yield curve thus obtained, we also show the results using the worst fit yield model in Fig. 7.30. This result was produced using the basic Lindhard model. Again, the total simulated spectrum is a reasonable fit over much of the energy range except it does not capture the 200 eV$_{ee}$ shoulder. This again was accomplished by selecting a large $F_{NR} \approx 5$ and using the $(n, \gamma)$ signal to match the rising signal edge. But, while the spectra from this and the Sorensen model are rather similar, the yield curve is significantly different. In particular, the Lindard yield, with no cutoff parameter, remains high even down to very low energies.
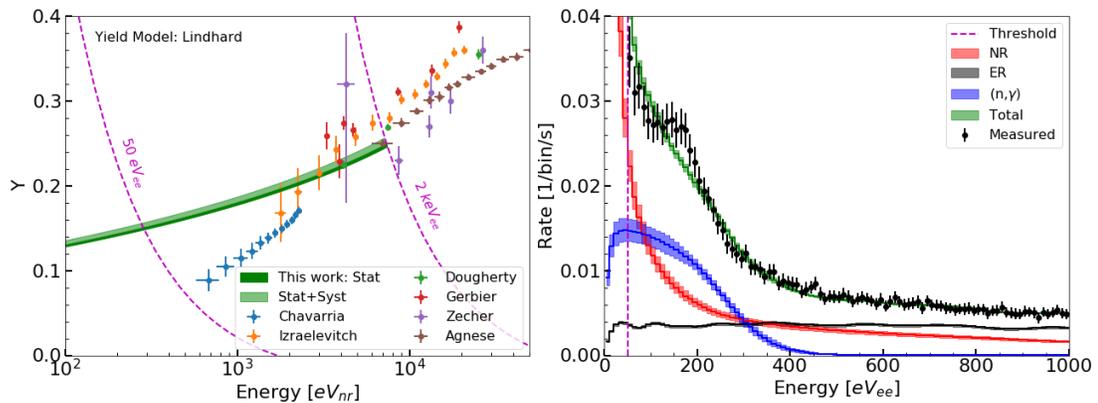


Figure 7.30: MCMC results for yield (left) and spectrum (right) using Lindhard model.

To be clear, these are not great fits but they are the best obtained using each model with the MCMC method and current simulation versions. Instead of showing all the spectra, we just present Table 7.1 of reduced $\chi^2$ values for the best fit spectra. These

were calculated over the $N$ spectral bins as

$$\chi^2/DOF = \frac{1}{N - M_{param}} \sum_{i=1}^{N} \left( \frac{R_{meas,i} - R_{sim,i}}{\sigma_{meas,i}} \right)^2 \qquad (7.31)$$

where the number of degrees of freedom, DOF, is the number of bins, $N$, minus the number of fitting parameters, $M_{param}$, and $\sigma_{meas,i}$ is the statistical uncertainty of the $i^{th}$ measured rate, $R_{meas,i}$.

| Model | $\chi^2$/DOF |
|---|---|
| Lindhard | 722.8/190=3.80 |
| Chavarria | 670.4/189=3.55 |
| Sorensen | 376.7/189=1.99 |
| AC | 525.0/189=2.78 |

Table 7.1: Table of best-fit Chisq/DOF values for several yield models. Calculated using only statistical uncertainties.

For completeness, we also report the best fit parameter values in Table 7.2. These were taken from MCMC fits performed with all statistical and systematic uncertainties included in the spectral likelihood distribution. The values were calculated from the 1-dimensional distribution of each parameter (e.g. the histograms in Fig. 7.28). The numbers reported are the median value and the uncertainties are 1-$\sigma$ equivalents, calculated as the distance from the median to the $16^{th}$ and $84^{th}$ percentiles of the distribution. Again, this does not capture any of the correlation between parameters, but gives a reasonable sense of their plausible values. We generally found that all models preferred $f_{ER}$ values around 5.4 – 6.1, somewhat larger than the $f_{NR}$ values of 3.0 – 4.3. Similarly, the $(n, \gamma)$ scaling factor ranged from $f_{NG} = 2.6$ – 4.5. The Sorensen model, which prodiced the best fit overall, did so with the highest NR scaling factors (both $f_{NR}$ and $f_{NG}$). The NR Fano factors were quite high for all models, but also had large relative uncertainties.

Unfortunately none of the tested yield models produced results fully consistent with the measured spectrum at all lowest energies. It is unclear if this is a deficiency of the yield models, the measurement, simulation, or the analysis methodology. As discussed above, one suspected cause is a deficiency in the elastic NR spectrum calculated from *Geant4*. We have found that the elastic NR spectrum alone does not provide a good

| Model | k | Cutoff Param. | $F_{NR}$ | $f_{ER}$ | $f_{NR}$ | $f_{NG}$ |
|---|---|---|---|---|---|---|
| Lindhard | $0.122^{+0.0043}_{-0.0044}$ | - | $5.45^{+2.5}_{-1.8}$ | $6.09^{+0.30}_{-0.29}$ | $3.01^{+0.28}_{-0.30}$ | $2.64^{+0.24}_{-0.21}$ |
| Chavarria | $0.138^{+0.016}_{-0.011}$ | $a^{-1} = 676^{+490}_{-360}$ eV | $5.55^{+2.9}_{-1.9}$ | $6.03^{+0.32}_{-0.31}$ | $3.22^{+0.32}_{-0.34}$ | $2.84^{+0.29}_{-0.25}$ |
| Sorensen | $0.161^{+0.030}_{-0.017}$ | $q/10^3 = 2.18^{+1.1}_{-0.76}$ | $8.82^{+5.4}_{-3.5}$ | $5.36^{+0.38}_{-0.37}$ | $4.28^{+0.44}_{-0.43}$ | $4.54^{+1.3}_{-0.88}$ |
| AC | $0.118^{+0.0043}_{-0.0043}$ | $\xi = 217^{+79.}_{-60.}$ eV | $4.36^{+2.2}_{-1.4}$ | $5.59^{+0.32}_{-0.31}$ | $3.50^{+0.31}_{-0.31}$ | $2.75^{+0.25}_{-0.23}$ |

Table 7.2: Table of best-fit parameters for several yield models calculated by including statistical and systematic uncertainties in MCMC sampling. Uncertainties are 1-$\sigma$ equivalents (i.e. calculated from the $16^{th}$ and $84^{th}$ percentiles).

fit to the rising signal edge at around $300 - 350$ eV$_{ee}$, slightly before the $(n, \gamma)$ signal is expected to contribute.

Furthermore, it turns out that the roughly exponential shape of the elastic NR spectrum is relatively insensitive to changes in yield. This means that changing the yield smoothly (within the constraints of our tested models) does not provide adequate distortion to match the measured signal. This is why the MCMC routine instead increased $F_{NR}$ (which the featureless NR spectrum shape is also insensitive to) and used the $(n, \gamma)$ signal to build the shape of the rising edge. This comes at the cost of smearing out the $(n, \gamma)$ structure which is likely the source of the measured shoulder near 200 eV$_{ee}$.

Another possibility is that the yield models tested are themselves inadequate. As we reach recoil energies nearer and nearer $\varepsilon_\gamma = 3.8$ eV and the threshold for Frenkel defect formation of $\sim 20$ eV [119], there may be crystal effects that manifest in non-smooth yield and Fano behaviors. More generally, it could be that the yield models we tried above simply do not have sufficient flexibility to capture the true yield behavior. To account for this, one can either develop more sophisticated models or try alternate fitting methods.

### 7.4.4 Integral Method

The MCMC method described above is powerful because it can provide full multi-dimensional estimates of the parameter PDFs. However, we found that it was not able to provide completely satisfactory fits to the measured spectrum. The fits agreed with the general spectral shape, but the yield models considered may not have had

sufficient flexibility to match the details near threshold. For these reasons we explored an alternative analysis method which can generate yield curves of arbitrary shape to nearly perfectly match the measured spectrum. This comes at the cost of some simplifying assumptions which will be discussed in detail.

This analysis is essentially similar to the "Integral Method" described in [215]. The idea is really quite simple and related to the CDF-matching techniques we used way back in Ch. 5. We know that the measured spectrum is made up of contributions from ERs and NRs. Assuming that $Y$ is monotonic, the measured rate (in $eV_{ee}$) above some energy, $E_{ee,i}$, is equal to the rate from electron recoils above $E_{ee,i}$ plus that from nuclear recoils above some corresponding recoil energy, $E_{nr,i}$. We of course can not integrate the rate arbitrarily high in energy, so we fix the maximum value of $E_{ee,max} = 2$ keV and a corresponding $E_{nr,max}$ (which itself depends on yield, more on that below). Given a measured total spectrum and simulated spectra of all the electron and nuclear recoil components, we can then perform the integrals over a range of $E_{ee,i}$ and find the corresponding set of $E_{nr,i}$. This set of points then allows us to map out the yield curve. As described below, this assumes each event is composed of a single hit.

In our case, we have a measured spectrum which gives a rate, $R_{meas}(E_{ee})$ and simulated spectra for ERs of rate $R_{ER}(E_{ee})$, elastic NRs of rate $R_{NR}(E_{nr})$, and $(n, \gamma)$ (NG) events of rate $R_{NG}(E_{nr})$. Thus, the requirement described above can be written as

$$\sum_{E_{ee,i}}^{E_{ee,max}} R_{meas,i} = \sum_{E_{ee,i}}^{E_{ee,max}} R_{ER,i} + \sum_{E_{nr,j}}^{E_{nr,max}} (R_{NR,j} + R_{NG,j}) \qquad (7.32)$$

where e.g. $R_{meas,i} = R_{meas}(E_{ee,i})$ and $R_{NR,j} = R_{NR}(E_{nr,j})$. The procedure then is to select some $E_{ee,i}$ and decrease $E_{nr,j}$ until the condition is met, then record the pair $(E_{ee,i}, E_{nr,j})$. This is repeated for a range of $E_{ee,i}$ from $E_{ee,max}$ to the lower analysis threshold of $E_{ee,min} = 50$ eVee.

An example of how this works is shown in Fig. 7.31. The colored portions show a set of several integration regions for which the simulated spectra integrals (center and right plots) match that of the measured signal (left). The resulting set of $(E_{ee,i}, E_{nr,j})$ are shown in the lower plot, with the black point corresponding to $(E_{ee,max}, E_{nr,max})$ and the other colors corresponding to the above integrals. Finally, from those points
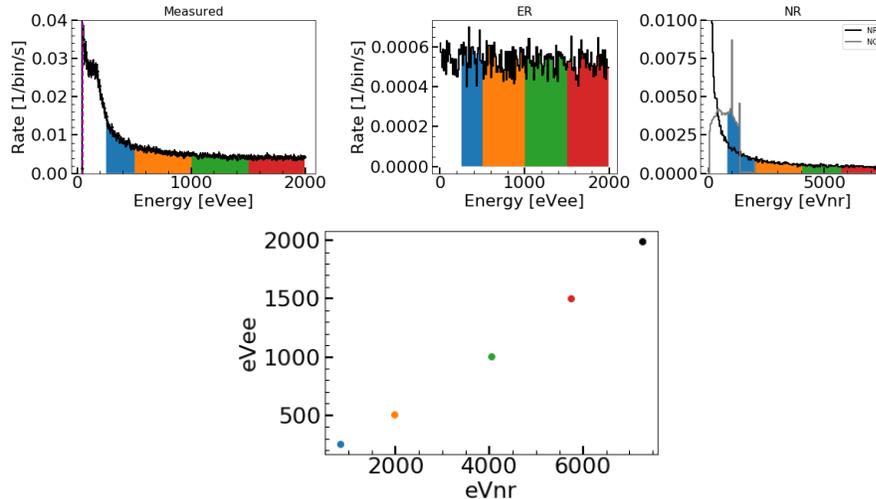
Figure 7.31: Cartoon of the Integral Method. Colored regions of integration on the measured spectrum(left) are made up of contributions from the various simulated components (right). Matching the integrated rates allows correspondence between the eVee and eVnr energy scales as shown at the bottom.

one can calculate the yield as

$$Y = \left( \frac{E_{ee}}{E_{nr}} (\varepsilon_\gamma/eV + 1) - \varepsilon_\gamma/eV \right). \tag{7.33}$$

Note that this relation holds for this class of detectors because we measure the total phonon energy from both the initial recoil and subsequent NTL phonons. This is in contrast with devices which are only sensitive to the charge produced, for which $Y = E_{ee}/E_{nr}$. However, this is a small effect given the operating bias used gives $\varepsilon_\gamma/eV = 0.03$.

**Single Hit Energy Assumption**

An important assumption required for the Integral Method is that each event is described with a single recoil energy which determines the yield. In a large detector this is not strictly the case because multiple scatters can occur. Furthermore, the $(n, \gamma)$ signal is in general a combination of several recoils within a single event. So, to perform this analysis we will make the simplifying assumption that the yield of nuclear recoil events is governed by the total recoil energy in the event, not the individual components.

We examine the validity of this in two ways. First, we know that a large portion of the simulated events are dominated by one energy deposition. We can show this be calculating the the ratios of maximum hit energy to the total event energy for the simulated data. I.e., for an event made up of $N$ hits of energies $E_i$, we calculate $\max(E_i)/\sum_i E_i$. We find that the fraction of events having $\max(E_i)/\sum_i E_i > 0.9$ is $0.89/0.80/0.93$ for ER/NR/NG events, so that the majority of events are dominated by a single hit. Now, this makes no difference for the ER events, because they will not have the yield applied to them. However, a minority of the NR and NG events might be affected.

We can also test the effect of making this assumption by comparing the distributions of events in eVee for an example yield function. As an extreme example, we choose a Lindhard yield with $k{=}0.146$ and a sharp cutoff of $Y(E_r < 100 \text{ eV}) = 0$. This is extreme because it will completely remove the influence of low energy hits. The effect of this is shown in Fig. 7.32. On the left are the NR spectra when either applying yield to the individual hits $(E_i)$ or to the sum of hits $(\sum_i E_i)$. For these events, there is effectively no change. On the right is the comparison for $(n, \gamma)$ events. The only meaningful change of this population is the shift of one prominent peak. However, this change is only about $10 \text{ eV}_{ee}$, comparable to a single bin width. So, while this may produce some difference if there is a hard yield cutoff, the change will likely be small. For now, we will thus assume that making the single hit assumption is reasonable. In the end, we can repeat this exercise for any derived yield function to evaluate the impact.
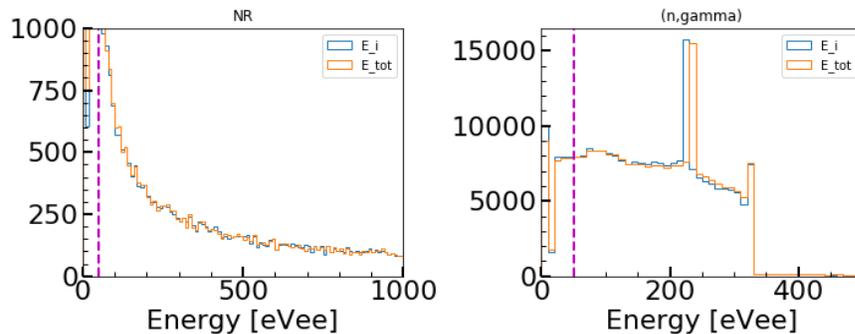


Figure 7.32: Comparison of effect when applying a yield function with cutoff to either the individual hits ("E_i") or total hit energy ("E_tot"). Left: The NR spectrum is negligibly changed. Right: Only one peak in the $(n, \gamma)$ distribution is affected significantly.

**Free Parameters**

There are several free parameters that must be selected before a yield curve can be calculated in this manner. They consist of an initial yield point and several rate-scaling factors for the simulated spectra. In this analysis, our entire energy range has direct NRs because the PuBe source emits neutrons at MeV energies and only some of them are thermalized. Thus we can start the integration procedure above the $(n, \gamma)$ signal, but we need to assume some starting value of $Y_{max} = Y(E_{nr,max})$.

As discussed earlier, our nominal simulated ER and NR rates do not agree with the measured rate for any reasonable range of yield values. At energies above 1 keV$_{ee}$, there will be no $(n, \gamma)$ events, just ER and elastic NR. To match the observed rate, the ER and NR rates would both need to be increased by a factor of $\sim 4$ relative to the expected rates. The cause of this discrepancy is not yet understood. The ER and elastic NR events are both generated by a *Geant4* simulation of the PuBe source in the K100 lab. The livetime of the simulation has been adjusted to match that of the measurement using our best estimate of the PuBe neutron production rate which is estimated to be known within 3% based on historical documentation of the source activity and adjusted for source composition changes over time [235]. There is some question as to the energy range over which the neutron production rate was initially measured, but it is reasonable to assume this included neutrons above 1 MeV, in which case the measured rate would only be low by 25% at most [236]. The PuBe source is still managed by the UMN group, so measurements of the current source rate could be made to confirm this. Perhaps the easiest (and safest) method would be to measure the 4.4 MeV gamma rate and infer the neutron production rate. Another possibility is that the measured rate in the HV detector is incorrect, however, no issues have been found so far after correcting for the effects of cut and trigger efficiencies as well as DAQ write efficiencies as described above. In contrast to the *Geant4* simulations, the livetime of the $(n, \gamma)$ simulation is calculated in an independent manner and is believed to be understood better.

In any case, we again introduce a set of empirical rate scaling factors, $f_{ER}$, $f_{NR}$, and $f_{NG}$ to scale to overall rate of the simulated spectra. These modify the rate balancing

condition as follows

$$\sum_{E_{ee,i}}^{E_{ee,max}} R_{meas,i} = \sum_{E_{ee,i}}^{E_{ee,max}} f_{ER}R_{ER,i} + \sum_{E_{nr,j}}^{E_{nr,max}} (f_{NR}R_{NR,j} + f_{NG}R_{NG,j}) \qquad (7.34)$$

While the values of $f_{ER}$ and $f_{NR}$ are somewhat unconstrained, we can make an independent estimate of $f_{NG}$. $f_{NG}$ scales the simulated rate of $(n,\gamma)$ events observed by the detector, which is proportional to the rate of neutron capture on Si atoms. For the MCMC analysis, we used an early estimate of the nominal total Si capture rate of 0.218 Hz. Since this has been used for most of the analysis thus far it corresponds to a value of $f_{NG} = 1$. Here, we re-evalute this calculation using updated measurements and cross-section references. This is a somewhat clumsy way to update the capture rate calculation, but we do it so as to allow direct comparison with the scaling factors used in the MCMC calculations.

The rate of thermal neutron activation with the same PuBe source and shielding setup was measured in UMN R66 using a Ge HV detector. The measurement was made by tracking the rate of increase in the 10 keV$_{ee}$ line associated with electron capture following activation of Ge-70. This measurement can be scaled to predict the total rate of neutron capture in the Si detector (of same volume, location, etc.) using

$$R_{Si} = R_{Ge-70} \frac{N_{Si}}{N_{Ge-70}} \frac{\sigma_{Si}}{\sigma_{Ge-70}} \qquad (7.35)$$

where $R_{Si}$ ($R_{Ge-70}$) is the activation rate of all Si isotopes (Ge-70), $N_{Si}$ ($N_{Ge-70}$) is the number of Si (Ge-70) nuclei in the detector, and $\sigma_{Si}$ ($\sigma_{Ge-70}$) is the thermal neutron capture cross section of Si (Ge-70). The Ge-70 activation rate was measured to be $R_{Ge-70} = 1.49 \pm 0.29$ Hz in R66. The thermal neutron capture cross section for Ge-70 is $\sigma_{Ge-70} = 3.05 \pm 0.13$ b [237]. The capture cross sections for the Si isotopes are given in [238] and their natural-abundance weighted average is $165 \pm 4$ mb. These values give an updated estimate of the Si activation rate of $R_{Si} = 0.444 \pm 0.089$ Hz which corresponds to scaling values of $f_{NG} = 2.04 \pm 0.41$.

To proceed with this analysis, we adopted a method of normalizing the simulated rate to the measured value near 2 keV$_{ee}$ by selecting a pair of $(f_{ER}, f_{NR})$ values. This was done by first selecting some trial value for $Y_{max}$ and using it to calculate the rate of ER and NR events between 1.99 keV$_{ee}$ and 2 keV$_{ee}$ (our highest energy bin). Then we

selected a value of $f_{ER}$ ($f_{NR}$) and calculated $f_{NR}$ ($f_{ER}$) such that the simulated rate agreed with the measured value. In this way, we always began the integration procedure with the highest energy point being self consistent. From there, the integration steps could proceed as described earlier.

## Resolution Effects

This procedure will find a yield curve that maps the simulated spectra to give a near-perfect match to the measurements, limited only by the simulation statistics. However, we have not yet taken into account the finite detector resolution which will smear simulated energies over a range of measured energy bins. To include this effect, we again followed the lead of the DAMIC analysis [215] and adopted the following procedure:

1. Generate $0^{th}$ yield curve, $Y_0(E_{nr})$.

2. Use $Y_0$ to transform $eV_{nr}$ event energies to $eV_{ee}$.

3. Smear $eV_{ee}$ energies by adding a random offset consistent with resolution function.

4. Transform smeared eVee energies back to $eV_{nr}$.

5. Calculate $Y_1(E_{nr})$ using smeared $eV_{nr}$ energies.

6. Iterate.

This was generally found to converge after as few as $2 - 3$ iterations and generated eV$_{ee}$ spectra smeared with the detector resolution as expected. However, because the energies are smeared by a random amount each time, it introduces a small random scatter in the final simulated spectra, though the agreement with measured rates is still satisfactory.

One interesting effect was that the inclusion of resolution effects provided some limited sensitivity to the yield curve immediately below our detector's energy threshold. This is because lower energy events can be smeared up and contribute to signal in analysis region. The size and shape of this contribution depends on the yield because it sets how far below threshold such values are and thus how likely they are to be smeared up. Now, this sensitivity is different from that obtained elsewhere in the analysis region. By definition, we do not have a measured rate below threshold and

thus cannot directly calculate a yield there. However due to such effects, one must assign *some* yield value below threshold. We have tried three such extrapolations. First we tried maintaining the same yield value from the lowest measured value down to 0 energy. However, this tended to systematically overestimate the simulated spectrum above threshold indicating the sub-threshold yield was too high. Conversely, setting the sub-threshold yield to 0 tended to underestimate the above-threshold spectrum. In the end we settled on a linear extrapolation of the yield below threshold. It was found that this allowed the algorithm to reliably produce simulated spectra consistent with the measurements and the expected resolution effects. With this procedure, we generally found it sufficient to make only one or two passes of the smearing procedure, with no significant changes in spectra and yield curves after further iteration.

**Calibrating to Previous Results**

We could not prescribe fixed values of $Y_{max}$, $f_{ER}$, or $f_{NR}$ a-priori so we were unable to just calculate a single yield curve from our data and simulations using this method. Instead, we chose to calibrate the high energy region (above the $(n, \gamma)$ signal) of our data to previous yield measurements in order to select a range of plausible $Y_{max}$, $f_{ER}$, and $f_{NR}$ values. In particular, we used the Izraelevitch measurements [232] to benchmark our calculations. These were the only measurements we are aware of which fully spanned our measured energy range from $1 - 2$ keV$_{ee}$.

To perform this calibration, we produced a large number of yield curves with a range of $Y_{max}$, $f_{ER}$, and $f_{NR}$ values and retained only those which were consistent with all of the Izraelevitch points within measurement uncertainties. We also calculated these full curves for the cases of $f_{NR} = 0$ i.e. no $(n, \gamma)$ and of $f_{NG} = 2.04 \pm 0.41$, our best estimate of the correct $(n, \gamma)$ flux. The curves which were consistent with the Izraelevitch measurements had scaling values in the range $f_{NR}$=5.6 – 9.0 and $f_{ER}$=1.3 – 4.2 with high $f_{NR}$ correlated with low $f_{ER}$.

We then plotted the envelope of all values attained in each set of calculations. This is shown in Fig. 7.33, with separate plots for the cases of with and without $(n, \gamma)$ signal. Here the edges of the envelopes do not represent any single yield curve, rather the set of extreme values achieved by any yield curve in the set. The solid line in each band represents the single yield curve which best matches the Izraelevitch measurements.

Similarly, we have plotted on the right the full range of rate values achieved by any calculation in the set. The solid lines in these histograms are the spectra corresponding to the same calculation as the solid best-fit yield line.
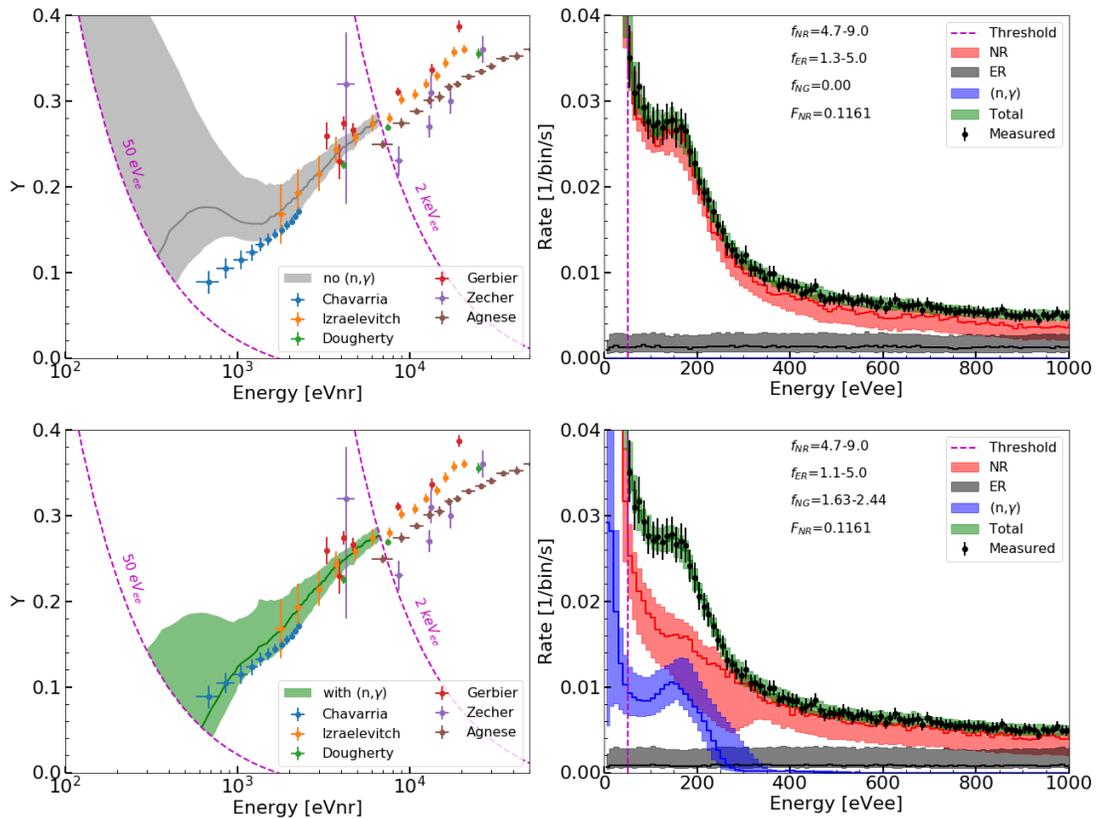


Figure 7.33: Left: Distribution of yield curves consistent with Izraelevitch data and the present analysis. The following previous measurements are shown for comparison: DAMIC [215], Izraelevitch [232], Dougherty [225], Gerbier [233], Zecher [234], Sattler [239], CDMS-II [120]. Right: Corresponding distribution of various spectral components. Top: Case with no $(n, \gamma)$ signal. Bottom: Case with expected $(n, \gamma)$ signal.

With these results, we find that the extrapolated yield curve is more consistent with the low energy DAMIC measurements when the $(n, \gamma)$ signal is included. We can also examine the spectra and see that, in the case without $(n, \gamma)$, the yield curve must rise at low energy to distort the exponential NR signal into matching the measured shoulder. In contrast, when the $(n, \gamma)$ signal is included, it describes the shoulder and the NR spectrum much more naturally, though there is sometimes still a small distortion in

yield and NR spectrum visible around 300 eV$_{\text{ee}}$(1.5 keV$_{\text{nr}}$). The results shown here were calculated with the assumption that the Fano factor for NRs, $F_{NR} = 0.1161$, the same as for ERs. Repeating this exercise with different values has shown almost no change in the resulting yield band for $F_{NR} < 2$. This is not surprising because the resolution is dominated by detector effects below 2 keV$_{\text{ee}}$ as described in Sec. 7.4.2.

After benchmarking our simulation livetimes at high energy, we can make an argument that the measured signal is only consistent with previous yield measurements if the $(n, \gamma)$ signal is present at the expected levels. In future studies, more work should be done to independently measure the PuBe neutron flux and better validate the *Geant4* simulations. This would eliminate the need for calibration to previous yield measurements. The yield curves obtained using this method are obviously tied directly to the previous yield points we chose to calibrate to. It is important to note that there is significant tension in existing measurements even at higher energies. These tensions may be due to the differences in the detector types, temperatures, electric field conditions, and crystal purities used in the various measurements. For example, the Izraelevitch measurements were obtained using a commercial Si drift detector which was 500 $\mu$m thick and operated at 220 K and 110 V bias while the CDMS-II (Agnese) data was taken with 10 mm thick ZIP detectors biased at $\pm 4$ V and operated at $\sim$50 mK. For this work, we have used the Izraelevitch measurements for calibration simply because they covered a complementary range of recoil energies. However, it could be possible that, for example, there is some temperature-dependent effect that reduces yield from the Izraelevitch to the CDMS-II (Agnese) values. In that case, lowering the calibration points used would naturally affect the rest of the yield curves calculated here.

## 7.5   Discussion

### 7.5.1   This Work

The two main results of this work are the MCMC fit using the Sorensen model as shown in Fig. 7.29 and the Integral Method fit with $(n, \gamma)$ included which was shown in the lower half of Fig. 7.33. Both of these roughly agree with previous yield values over the analysis range. Each method has its own strengths and weaknesses. The Integral Method does not project below threshold and requires some simplifying assumptions,

but it is flexible enough to generate arbitrary yield curves. In contrast, the MCMC method can generate multi-dimensional PDFs of the parameters of interest, but relies on functional yield models and may not produce sufficiently good spectral fits.

With MCMC, $F_{NR}$ is driven quite high because it smears the $(n, \gamma)$ signal to improve rising edge fit. The Integral Method does not need to do this, because it can adjust yield there. However, this results in a noticeable wiggle in the yield and NR spectrum around 300 $eV_{ee}$(1.5 $keV_{nr}$). These effects in both methods may be the result of an improper NR spectrum. This is an essential component to this analysis because it is somewhat degenerate with the targeted $(n, \gamma)$ spectrum. As mentioned above, this could be due to the effect of un-modelled neutron interactions with impurities in the cryostat material. Even if all materials were included, it as also important to have the correct physics implemented in the simulation. This is not trivial since we are simulating a large range of energies from meV to MeV. Recent studies have shown that even simulating the gamma spectrum correctly for a detector near an AmBe source requires careful selection of how certain low-energy neutron physics interactions are included in *Geant4* [240]. We should explore the systematic uncertainties introduced by the choice of such calculation methods, particularly those describing neutron transport.

### 7.5.2 Next Steps

While neutron scattering experiments will provide complementary NR yield measurements, the $(n, \gamma)$ technique has the advantage of being performed in-situ. The results shown here are the first to demonstrate that this is a viable method to generate such such low energy NRs to map out the ionization yield. To improve these results, it would be good to find an independent way to benchmark these simulations. One possibility is to use an iZIP detector either concurrently or in an otherwise duplicate run. While the iZIP can only reach down to $\approx 10$ $keV_{ee}$, the ER/NR discrimination would allow a nice measurement of the keV-scale neutron flux at the detector location and could be directly compared to simulations. An HV detector operated at $\sim 50$ V could then be used to provide an overlapping measurement from say $10 - 20$ $keV_{ee}$ and a subsequent low energy measurement with the lowest achievable threshold. In this way, one could better constrain the low energy neutron spectrum.

A more game-changing next step to make the most of the $(n, \gamma)$ events will be the

observation and tagging of coincident gammas and their incorporation into the analysis. An array of external detectors, like the NaI devices tried here but with large solid-angle coverage, would allow the selection of a very pure population of $(n, \gamma)$ events. This would essentially eliminate the dependence on simulations of the other spectral components (ERs and direct NRs). With sufficient energy resolution of the external detectors, this could also allow identification of recoil energies on an event-by-event basis, given the discrete nature of the various nuclear levels. Not only this, but the external detector location relative to the HV detector crystal axis may even allow a study of directional variation in NR response. This could be of immense benefit to any future efforts to validate DM signals via directionality.

# Chapter 8

# HVeV Detectors

We have already discussed improving DM detectors by increasing the mass, refining the position reconstruction, and by extending NR energy calibrations to low energies. This chapter discusses a new generation of small detectors originally designed to improve the detector energy threshold by sheer sensitivity and better characterization of HV leakage properties.

These devices are operated at high voltage and able to achieve eV-scale energy resolution, and have hence been named "HVeV". The original HVeV device, designed and tested at Stanford, was the first phonon-based detector capable of resolving single e/h pairs [162]. These devices are small square chips with dimensions of $1 \times 1$ cm$^2$ and 1 mm thick, about $1/2600^{th}$ the size of the current generation of SuperCDMS SNOLAB flagship detectors (iZIP and HV). In the small devices, the leakage current density is greatly reduced regardless of whether it is due to effects at the electrode surfaces (which scale with area) or in the crystal bulk (which scale with volume). So while HV detectors are currently limited to thresholds of 10s of eV$_{ee}$, the HVeVs now routinely achieve $\sim 0.1$ eV$_{ee}$. Such low thresholds have allowed SuperCDMS to probe new low mass DM parameter space even at above-ground facilities [47, 48].

Perhaps more importantly the HVeVs are an excellent testbed for improving the technologies used in all SuperCDMS detectors. Below we highlight two such aspects of this. First, we discuss the measurement of the increased energy efficiency of the latest TES designs on the HVeV. Then we examine a new class of correlated events that have been identified in these devices and which may improve our understanding of the HV

leakage currents. The HVeV data discussed in the energy efficiency section was taken with the Animal ADR at Northwestern University. Particular thanks to Z. Hong and R. Ren for that work. The pulse burst data was obtained by running a similar device in the NEXUS fridge at FNAL. For that work, thanks again are due to the Northwestern crew and N. Kurinsky and V. Novati who helped in various ways to bring the facility online and operate the fridge and detector. I was fortunate enough to play a role in many aspects of that work including designing the cold electronics boards (Appendix C), designing and machining some of the mechanical fridge components, as well as operating the run and performing data analysis.

## 8.1 Energy Efficiency

A natural way to improve detector sensitivity is to increase the efficiency with which phonon energy is coupled to the TESs. Recall that there are several stages in this process. After a recoil event in the crystal, the athermal phonons are initially propagating in the crystal substrate, basically bouncing around the crystal. Eventually, the phonons are either absorbed by an Al fin of a QET, absorbed by passive Al like bias rails, or they thermalize. The fraction of phonon energy which couples into QET fins is the first efficiency bottleneck. This is governed by a combination of phonon demographics (population and energies), crystal geometry, substrate/QET material interfaces, and electrode/QET geometry. Next, phonons in the fins only have a finite probability of breaking Cooper pairs to produce quasiparticles (QPs), again a function of phonon demographics. The QPs then must diffuse through the fins to the TES, become trapped there, and deposit energy to heat the TES. All of this depends on the detailed fin geometry, QP diffusion properties in the fin material, the effects of trapped flux in the fins, and the fin-TES overlap geometry. There are clearly many factors to consider when designing these devices, and such details are beyond the scope of this work. For more information, interested readers can start with the following references [140, 198, 171].

In this section we discuss the measurement of the total energy efficiency of an HVeV device which is set by the cumulative effects of all these processes. The results shown here along with some of the gory design details described above are presented in an upcoming device characterization paper [241]. Some innovations in the QET geometry

were included in this design and the theoretical energy efficiency was hoped to be near $\sim 30\%$, almost three times higher than that obtained in the current generation of large SuperCDMS HV detectors. However, it should be noted that the current state of such design work involves a number of semi-empirical models, estimations, and scaling rules, making it something between an art and black magic. The particular detector design studied in this work was named "NF-C" and was a re-optimization of the device operated in [242]. It was composed of two phonon channels as shown in Fig. 8.1 and was operated in the "Animal" ADR at Northwestern during Animal Run number 68 (AR68). The following sections will describe the device characterization and measurements performed to extract the detector's energy efficiency.
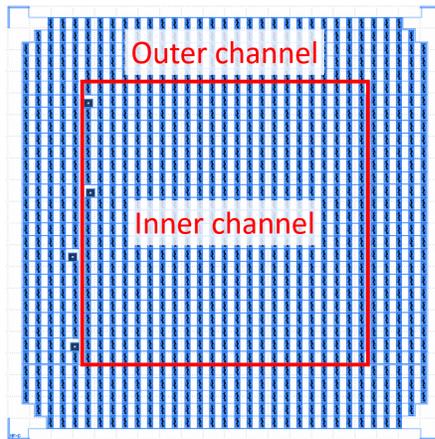


Figure 8.1: Phonon mask for the NF-C detector. Each channel is composed of approximately 500 QETs in parallel.

### 8.1.1  Estimating Circuit Parameters

As derived in Ch. 3, the phonon energy absorbed by a TES can be inferred from the change in signal current and circuit parameters as

$$
\begin{aligned}
E_{abs} \approx &\left(1 - 2\frac{R_L}{R_L + R_0}\right) I_b R_{sh} \int \delta I_s(t) dt \\
&+ R_L \int \delta I_s{}^2(t) dt
\end{aligned}
\tag{8.1}
$$

where $R_L = R_{sh} + R_{par}$ is the total resistance (shunt and parasitic) in the loop apart from the TES, $R_0$ is the TES operating resistance, and $I_b$ is the TES bias current.

Here we have defined $\delta I_s(t) = I_0 - I_s(t) > 0$ as the change in signal current during a phonon pulse relative to the quiescent value, $I_0$. This absorbed energy can be compared to the calculated total phonon energy, $E_{tot}$, to define the detector's energy efficiency, $\epsilon = E_{abs}/E_{tot}$. It is thus important to constrain the several circuit parameters in order to obtain a proper energy estimate.

These values are critical in determining the absolute energy calibration of the device under test but are difficult to measure or estimate accurately during a run, and even harder to get a handle on after the fact, so the values we used span our best estimate of plausible ranges. The shunt resistance in AR68 was fixed at the nominal value of $R_{sh} = 50$ m$\Omega$. Any deviation from this was absorbed into $R_p$ as described in the next section. The bias current, $I_b$, was realized with a voltage source and bias resistor. The bias resistance was estimated as $8100 \pm 50$ $\Omega$, with uncertainty mainly due to resistance variations with fridge temperature. The voltage source was more complicated as two different sources were used for the $I_b - I_s$ sweep and the fixed $I_b$ used for detector operation. As discussed below, there was evidence for an offset between the two sources, which was a major source of uncertainty. Finally, the signal current, $I_s$, is measured by a SQUID-based amplifier, with input-to-feedback coil turns ratio of $14.0 \pm 0.1$ and first-stage feedback resistance of $2040 \pm 50$ $\Omega$. There were also several second-stage amplifiers with gain factors estimated to be $1 \pm 0.01$. These are important because the digitized measurement of $I_s$ scales linearly with each of these three factors.

## 8.1.2 $I_b - I_s$ Analysis

As described in Sec. 3.2, the analysis of $I_b - I_s$ curves yields two important parameters used in the energy efficiency calculation; the parasitic resistances, $R_p$, and the operating resistance, $R_0$. The NF-C device has only two phonon channels, a square inner channel and an outer surrounding ring. Both TES channels have $T_C \approx 65$ mK. The upper row of Fig. 8.2 shows the observed $I_b - I_s$ curves for these two channels measured with the fridge cooled to 50 mK. They behaved very similarly to each other, with TES transitions observed below $I_b \approx 20$ $\mu$A and falling into superconducting (SC) below $I_b \approx 16$ $\mu$A. The curves shown there use the nominal circuit parameters.

Because the $R_p$ and the TES normal resistance, $R_N$ were not known a-priori, the only resistance scale was the (assumed value of) $R_{sh}$. Absent other information, one
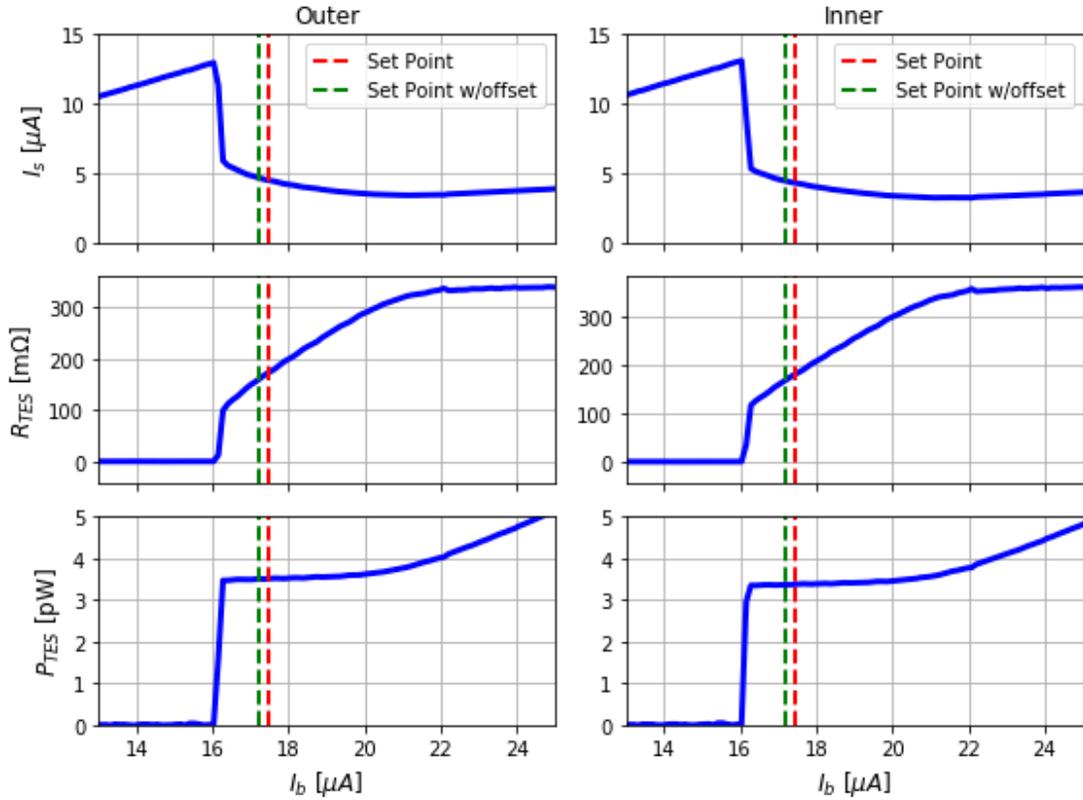
Figure 8.2: AR68 IV curve analysis. Blue curves show the values of $I_s$ (top), $R_{TES}$ (middle), and $P_{TES}$ (bottom) for both channels as calculated using the nominal circuit parameters. Vertical dashed lines represent the estimated range of operating point.

must make an assumption about the relative contributions to $R_p$ from the shunt and TES branches. For this work, we assumed they both contributed equally because there were roughly the same number of wires and connectors (the main sources of $R_p$) on each branch. From there, we used the slope of the SC portion of the curve to determine the value of $R_p$. With that, we were able to infer the TES resistance, $R_{TES}$, and finally the TES power, $P_{TES} = I^2_{TES}R_{TES}$, as shown in the middle and lower rows of Fig. 8.2.

One major systematic arose because the electronics which were used to control the TES bias during $I_b - I_s$ curves was different from that used when setting the fixed bias point during detector operation. It is possible that the 0 point of the two setups may have been offset from each other, but this was not discovered until well after the

experiment was completed. This is evidenced by the fact that the intersection of the SC and normal portions of the $I_b - I_s$ curves were offset from the origin. If the offset was somehow due to the fridge wiring, the fixed bias point setting would also undergo this shift. However, if the shift is inherent in the $I_b - I_s$ sweep electronics, the fixed bias point would not undergo the same shift. The magnitude of these shifts are only $\approx 0.2$ $\mu$A in $I_b$, but this can account for 10s of m$\Omega$ variation in the selected TES operating resistance, $R_0$. For these reasons, we included this possible systematic shift in the subsequent analysis.

To incorporate all of the known uncertainty sources, we repeated the $I_b - I_s$ analysis with every combination of values (min, nominal, max) for all of the parameters described above. We also repeated this both with and without the $I_b$ offset shift. Then we used the extreme ranges of the resulting values of $R_p$ and $R_0$ as our credible region. The results of this jackknife uncertainty analysis are shown in Table 8.1. Although it is not used in this analysis, we also list the calculated TES normal resistances, $R_n$.

| Parameter | Outer | Inner |
|:---:|:---:|:---:|
| $R_p$ [m$\Omega$] | $12.3 - 18.8$ | $11.2 - 17.5$ |
| $R_n$ [m$\Omega$] | $311 - 371$ | $332 - 396$ |
| $R_0$ [m$\Omega$] | $125 - 186$ | $136 - 196$ |

Table 8.1: Parameter ranges derived from HVeV $I_b - I_s$ curves in AR68.

As an aside, the assumption of equal $R_p$'s on the two branches is almost certainly not correct, but the values are difficult to nail down otherwise. One way would be to run the identical electronics and TES loop wiring but with a TES having known $R_n$. Then, the normal and SC slopes could be compared to determine both parasitic values independently. Another way to break the degeneracy is to operate identical runs but with different $R_{sh}$ values (e.g. as we did at UMN in Sec. 4.5). In principle, this was exactly the case in AR68 ($R_{sh}$=50 m$\Omega$) and several calibration runs of Animal operated at TUNL ($R_{sh}$=10 m$\Omega$). However, such analysis comparing those $I_b - I_s$ curves has not yielded sensible $R_p$ values. This could be an effect of small (m$\Omega$) changes in $R_p$ introduced when changing detector payloads or connecting to different shunt resistors. While this is not yet fully understood, it serves to highlight the difficulty of accurately nailing down the various tiny resistances which are so important in these systems.

### 8.1.3  Efficiency Measurement

The low-energy detector calibration was performed using short pulses from a 635-nm (1.95 eV) laser. The laser was operated at room temperature and its signal reached the detector through a plastic optical fiber, filtered against black-body and infrared radiation at 1.4 K (for more details see Ref. [242]). The laser diode was driven with a pulsed current source in pulses lasting 500 ns each. The laser produced Poisson-distributed photon bursts with averaged number of photons per pulse, $\lambda \approx 0.3$. The laser pulses were synchronized with a TTL signal which was also digitized and used for offline analysis. The detector operated at an HV bias voltage of $V_{bias} = 100$ V for high NTL gain, and the fridge temperature was stable at $50.00 \pm 0.01$ mK. Cuts were applied to select phonon pulses in the detector which were coincident with the laser signal, had energy above the noise threshold, and which had low pre-pulse baseline values.

The identified phonon pulses were fit with the Optimal Filter (OF) algorithm and the resulting energy spectrum is shown in Fig. 8.3. Here, the energy has been calibrated to the total phonon energy produced in an event. Given the applied bias voltage, the signal produced by the absorption of $n$ laser photons, each of energy $E_\gamma$=1.95 eV, is expected to be

$$E_{tot} = n(E_\gamma + e \times V_{bias}) = n \times 101.95 \text{ eV} \tag{8.2}$$

where $e$ is the electron charge and each absorbed photon produces 1 one e/h pair. Note that this looks a little different from earlier expressions for total phonon energy produced by higher energy ERs in e.g. SuperCDMS iZIP detectors. For such detectors, we can not resolve the integer number of e/h pairs, so instead we used the measured average value of $\varepsilon_\gamma = 3.8$ eV/pair. However, in the HVeV analysis, the signal is known to be groups of 1.95 eV photons and we could use the quantized e/h peaks for calibration.

The efficiency was calculated for individual phonon pulses by comparing the energy absorbed by the TES as calculated with Eq. 8.1 to the calibrated $E_{tot}$. The resulting distributions of efficiencies, for a given combination of systematically affected parameters, are shown in Fig. 8.4. The primary source of uncertainty in these values was the systematic uncertainty of circuit parameters such as the TES bias resistance, $R_{par}$, and the possible TES bias offset. For this figure, we have selected the most conservative set of assumptions to obtain a *lower estimate* of the efficiency. Other combinations of
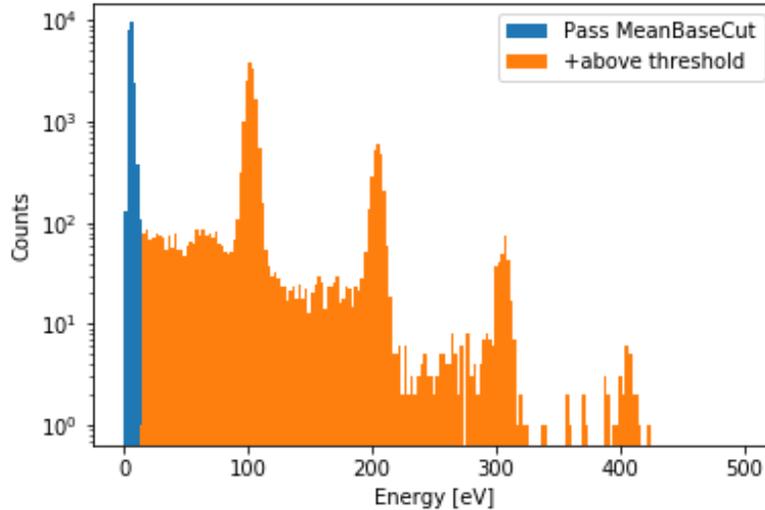
Figure 8.3: HVeV low-$\lambda$ laser calibration spectrum. The x-axis is calibrated to total phonon energy including NTL gain effects. Peaks correspond to signals generated by 0–4 e/h pairs liberated by laser photons.

systematic uncertainty values will shift the entire distribution to substantially higher values. The width of each distribution is set by a combination of the noise and energy variation of individual events.

As described above, we performed a jackknife uncertainty analysis to bound the systematic effects of the circuit parameters. The results of these calculations, extended to the efficiency calculation, are shown in Table 8.2 for different numbers of e/h pairs produced. The values quoted are the means of each distribution, along with the statistical uncertainty of the mean and the systematic uncertainty of the mean. The systematic effects are obviously dominant for all energies considered.

Recall now that the target efficiency of this device was in the neighborhood of 30%, comparable to the 27% obtained with its predecessor [242]. However, we are confident that this new design has achieved further improvement and may be reaching energy efficiencies as high as an order of magnitude better than that of the first HVeV device tested just a few years ago [162]. Given this possibility, it was unfortunate that uncertainties in the circuit parameters produced such large systematic uncertainties.
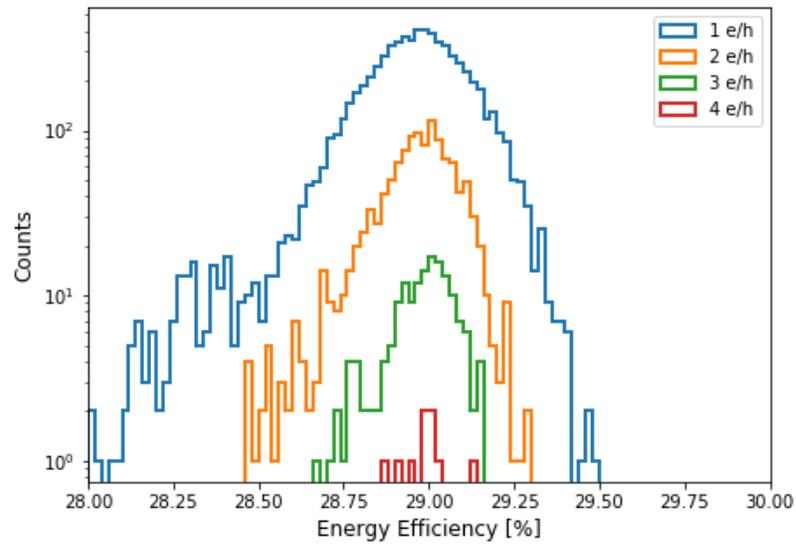
Figure 8.4: Distribution of energy efficiency in HVeV low-$\lambda$ data. Calculated using the most conservative combination of systematic uncertainties.

However, there are plans to repeat this measurement in the NEXUS fridge with a far more stable and better characterized circuit. Hopefully this work will be completed soon and confirm these results with improved precision.

| # e/h | Efficiency [%] | Stat. | Syst. |
|:---:|:---:|:---:|:---:|
| 1 | 43.7 | $\pm 0.0021$ | $^{+13.4}_{-14.8}$ |
| 2 | 43.5 | $\pm 0.0037$ | $^{+13.2}_{-14.5}$ |
| 3 | 43.3 | $\pm 0.0081$ | $^{+13.0}_{-14.3}$ |
| 4 | 43.0 | $\pm 0.032$ | $^{+12.8}_{-14.1}$ |

Table 8.2: Mean Energy Efficiency

## 8.2   Pulse Bursts

With the advent of the single-charge sensitive HVeVs, a number of interesting microscopic detector phenomena are able to be studied in new ways. For example, the distribution of events with energy between e/h peaks ("non-quantized" events) allows measurement of impact ionization and trapping probabilities [243]. The ability to differentiate quantized and non-quantized events has also allowed closer study of possible sources of leakage current when HV is applied. Recall that such leakage currents are the main limiting factor of the SuperCDMS HV detector resolutions.

In the course of studying such effects in an HVeV device at NEXUS, we have recently been examining an interesting and yet unexplained class of signals which have peculiar timing properties. In particular, when operating the detector at nominal bias voltages of ≈100 V, there are occasional, momentary increases in the measured pulse rate in excess of that otherwise expected from uncorrelated background sources. Anecdotally, such bursts have been observed everywhere HVeVs have been operated (Stanford, Northwestern, TUNL, and NEXUS) but have not yet been studied carefully. Most importantly, we do not yet know if these signals are generated by some phenomena external to the detector (e.g. correlated showers of photons) or internal to it (e.g. some gated leakage mechanism) at the electrodes. While such questions remain unanswered as of this writing, this section presents some initial efforts to describe and define this class of signal. While the work presented here is primarily focused on results obtained at NEXUS, much credit goes to R. Ren for ongoing studies of related phenomena at Northwestern including discussion of the implications for DM detection [244].

### 8.2.1   Anatomy of a Burst

The bursts generally contain between several and several dozen pulses all within a few ms. The individual pulses generally appear similar to have the same rise and fall times as other "regular" pulses, but their primary distinction is correlated arrival timing. An example of such a burst is shown in Fig. 8.5. In this example, the first two pulses are consistent with 2 e/h pair signals and the remaining pulses are each consistent with 1 e/h pair.

We developed crude criteria to identify bursts in the NEXUS datasets. A burst was
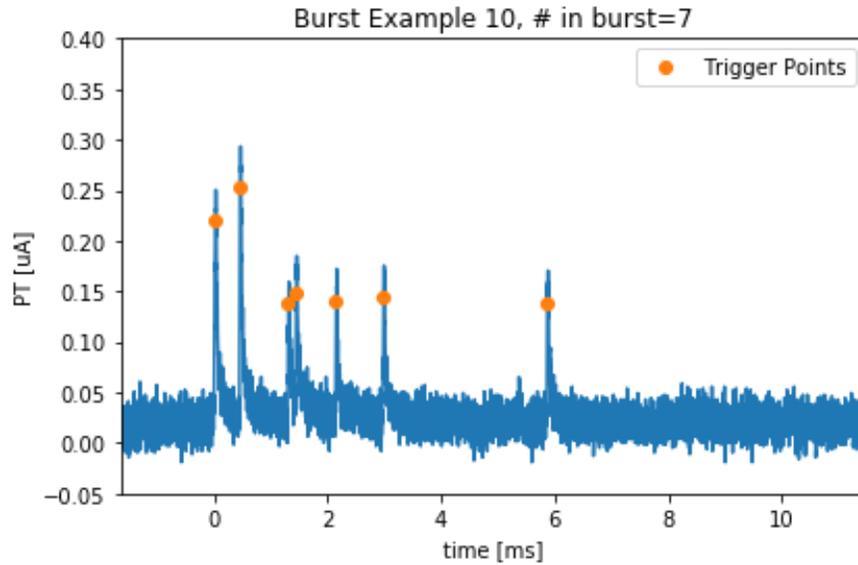
Figure 8.5: Example pulse burst in NEXUS data.

any chain of 3 or more pulses which were each separated from the next by less than 8 ms. This value was selected somewhat by-eye, but the results we show are relatively insensitive to it. To distinguish various types of pulses in the data, we developed a burst index (BI) system. Pulses which were not part of a burst were assigned BI=0, the first pulse in a burst, or burst "primary", was assigned BI=1, and the $k^{th}$ pulse in a burst was assigned BI=k. The set of pulses with BI>1 were also categorized as burst "secondaries". These labels will be useful in later sections.

### 8.2.2 Arrival Timing

One way to identify the presence of bursts in a large dataset is to examine the distribution of pulse interarrival times, $\Delta t$. If events are generated by a homogeneous Poisson process of rate $R$, the probability of observing $k$ events in time $t$ is

$$P(\text{k in interval t}) = \frac{(Rt)^k e^{-Rt}}{k!}. \tag{8.3}$$

Here, we specify the process is homogeneous, meaning $R$ is constant, at least over the length of the observation. If several such uncorrelated processes are occurring, each with different rates, $R_i$, the same expression holds with $R = \sum_i R_i$. This is the expected

behavior of the signals observed in our detectors. They are generally from processes like background radiation in the cryostat and environment which have half lives much longer than the few hours of data taking. Now, the probability density function (PDF) of interarrival times for such a process is

$$P(\Delta t) = Re^{-R\Delta t}. \tag{8.4}$$

So we expect to see a simple exponential distribution of $\Delta t$ values.

An example of the $\Delta t$ distributions typically observed in a detector at NEXUS is shown in Fig. 8.6. This shows the interarrival times of pulses from 24 hours of 100 V HVeV data. The left figure shows the distribution at long time scales relative to the $\approx 0.5$ ms width of a single pulse. There, the distribution is exponential as expected, with an implied rate of 7.3 Hz. However, there is an excess of events at lower $\Delta t$, which is shown on the right. This is not explained by the same homogeneous Poisson process and is an indication that the dataset contains correlated bursts of pulses. At time scales shorter than the average pulse width, pulses begin to pile up and the current generation of reconstruction algorithms cannot reliably fit all component pulses.
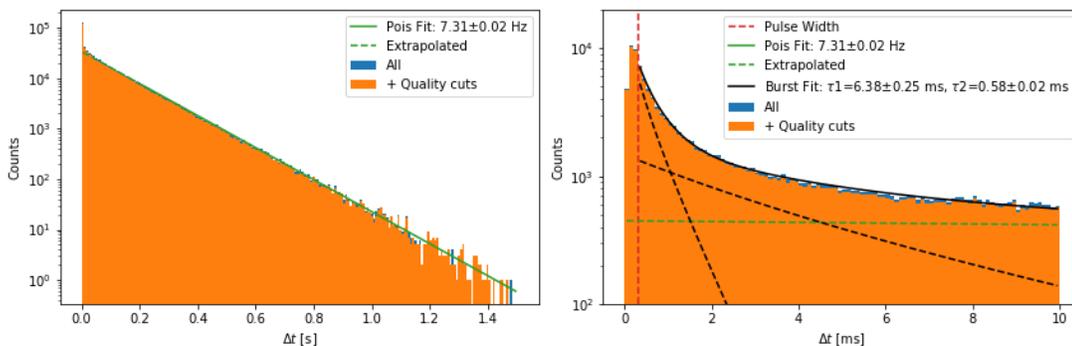


Figure 8.6: $\Delta t$ distribution of pulses in 24 hours of 100 V HVeV data at NEXUS. Left: At long time scales, the distribution is well-modelled by a homogeneous Poisson process. Right: There is an excess rate below $\approx 10$ ms which is indicative of pulse bursts. The excess has been fit as the sum of two characteristic burst timescales atop the extrapolated Homogeneous Poisson distribution.

The right side of Fig. 8.6 also shows an attempt to model the excess. To do so, we assumed that the pulses within a burst are exponentially distributed in time, with some characteristic lifetime, $\tau$. This is shown to be approximately true in the next section.

This is no longer a homogeneous Poisson process because the rate is changing with time. We can generate the PDF for an *inhomogeneous* Poisson process, by replacing

$$R\Delta t \rightarrow \int_0^{\Delta t} r(t')dt' \tag{8.5}$$

where $r(t)$ is the time-dependent rate. Now the interarrival PDF becomes

$$\begin{aligned} P(\Delta t) &= e^{-\int_0^{\Delta t} r(t')dt'} \times \frac{d}{dt}\int_0^{\Delta t} r(t')dt' \\ &= r(\Delta t)e^{-\int_0^{\Delta t} r(t')dt'} \end{aligned} \tag{8.6}$$

Using our assumption that the rate time dependence is exponential, $r(t) = \frac{1}{\tau}e^{-t/\tau}$, we obtain

$$P(\Delta t) = \frac{1}{\tau}e^{e^{-\Delta t/\tau} - \Delta t/\tau - 1}. \tag{8.7}$$

We can then add a distribution of this shape atop the already fitted homogeneous distribution and fit the low $\Delta t$ region. This neglects the timing between burst and non-burst pulses, but is a fine approximation when bursts are rare. However, we found that the excess is not well represented by a single $\tau$ value. Instead we needed to include an entire second such distribution as well. This is because even our exponential burst model does not yield an interarrival time distribution very different from that of the homogeneous Poisson. Basically this is because Eq. 8.7 is dominated by the $\exp(-\Delta t/\tau)$ term, and the double exponential does not contribute much to the total shape.

There is ongoing work to produce appropriate theoretical expressions for this timing distribution, but it is a nontrivial task given the physical process responsible is not yet known. One common related model is that of the gated or interrupted Poisson process (IPP) [245]. Here, possible events are generated by an underlying homogeneous Poisson process but they are also gated by another process which either inhibits or allows the events to occur. A nice derivation of the interarrival time distribution for such a scenario is presented in [246], which models the timing of Sferics (bursts of electromagnetic signals generated by lightning). That result produces a distribution described by two exponential values, as our excess appears to be. However, there are only burst events present in their analysis, and the longer fall time is thus set by the rate between the bursts themselves. In our case, there is also a significant constant background rate which would fill those timing gaps and which produces the long $\Delta t$ distribution. Of

course none of these address the case of non-constant, intra-burst timing distribution, which we will describe in the next section.

### 8.2.3 Other Properties

We present here some other interesting properties of pulse bursts. While this only provides a brief overview of their many facets, this selection may hold some clues as to their physical origin.

**Burst Multiplicity**

The total event rate in the detector was only $\approx 7.3$ Hz, so the occurrence of such bursts randomly from a homogeneous Poisson process was exceedingly rare, especially for bursts with high multiplicity (large number of pulses). Given our burst definition (all $\Delta t < \Delta t_{max}$), the probability of any set of $k$ sequential events being a burst is the probability of having $k - 1$ short $\Delta t$'s surrounded by two long $\Delta t$'s. This gives the following

$$
\begin{aligned}
P(\text{burst of k}) &= P(\Delta t \geq \Delta t_{max})P(\Delta t < \Delta t_{max})^{k-1}P(\Delta t \geq \Delta t_{max}) \\
&= e^{-2R\Delta t_{max}}(1 - e^{-R\Delta t_{max}})^{k-1}
\end{aligned}
\tag{8.8}
$$

The total number of sequential sets of $k$ events in a dataset of length $T$ is $RT - k$ (on average) so the total expected number of bursts of multiplicity $k$ in the dataset is

$$
N(k) = (RT - k)e^{-2R\Delta t_{max}}(1 - e^{-R\Delta t_{max}})^{k-1}
\tag{8.9}
$$

A comparison of this expected population to our observation is shown in Fig. 8.7. For a homogeneous Poisson rate of 7.3 Hz, there should have randomly been a few thousand bursts of three pulses, but a vanishingly small number of bursts containing more than six. However, we observed much higher values, including bursts containing as many as 40 pulses! Clearly something more interesting is happening here.

**Intra-Burst Timing**

In the upper plot of Fig. 8.8, we show histograms of the secondary pulse arrival times relative to the primary pulse. Recall that the category "secondaries" includes all pulses
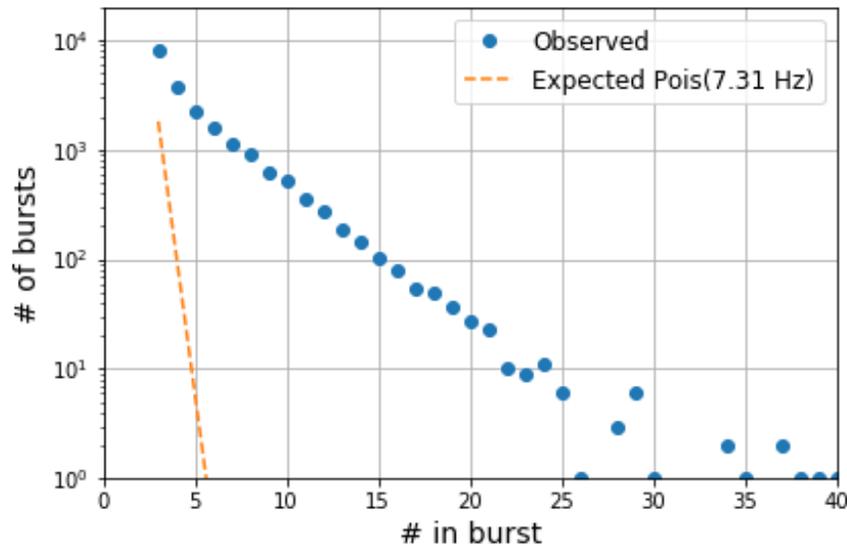
Figure 8.7: Observed number of bursts as a function of number of pulses in each burst. The expected number of bursts for a homogeneous Poisson process is shown for comparison.

in a burst other than the primary. This was done separately for bursts containing different numbers of pulses. All categories show a roughly exponential timing distribution, so secondaries are more likely to occur early in the early part of a burst and are rare beyond 30 ms. The 3-pulse distribution is somewhat distorted because it contains some background events and the burst definition cuts it off at 16 ms. Higher multiplicity distributions do not have such issues.

This figure is interesting because it demonstrates some correlation between intra-burst timing and multiplicity. If the bursts were produced by some process with statistical properties like radioactive decay, then the secondary arrival time distributions would be identical regardless of the multiplicity, i.e. the measured half life of a decay does not depend on the efficiency or solid angle coverage of one's detector. The lower plot of Fig. 8.8 shows the best-fit exponential fall times of each of the secondary timing distributions as a function of burst multiplicity. This may indicate the underlying process has a lifetime that grows with the total number of particles it will emit (and pulses it produces). If the process is occurring inside the detector, such as carrier injection from
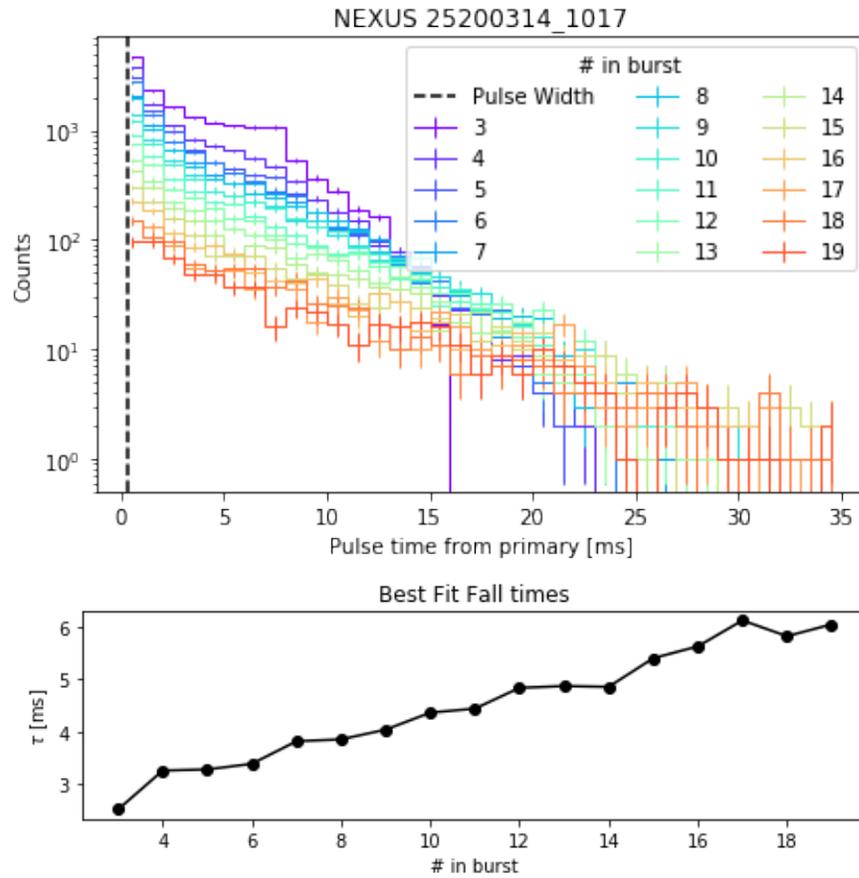
Figure 8.8: Top: Arrival time distribution of burst secondaries. Bottom: Fitted exponential fall times as a function of burst multiplicity.

an electrode, then this could be a direct look at the populations produced. If this is from some process external to the detector, doubtless there will be some additional geometric or energy-based efficiency with which it couples to the device that could distort these distributions to be different from that emitted.

**Energy Distributions**

We also studied the energy distribution of the various types of pulses in a burst to seek further hints as to their origin. A set of histograms of the individual pulse energies are shown in Fig. 8.9 where they are separated by BI. All types of pulses demonstrate

quantization, meaning they are either from a single carrier traversing the entire crystal or an e/h pair produced in the bulk, with both electron and hole drifted to the appropriate electrode. This rules out effects like the secondaries (BI>1) being trapped carriers that are released by the primary event, or some sort of stragglers from the primary event.
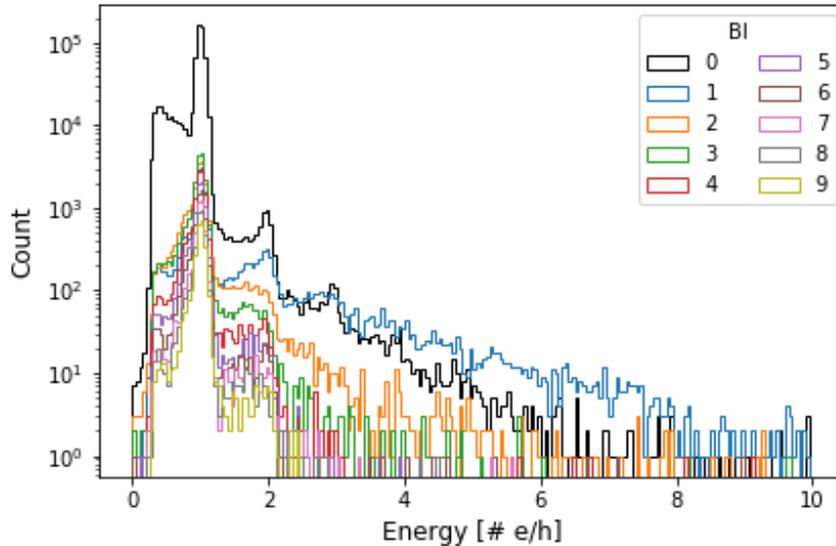


Figure 8.9: Comparison of energy spectra for various burst indices. Energy has been scaled to number of e/h pairs produced

We also observe that the spectrum of primaries (BI=1) is shifted to higher energies than the non-burst (BI=0) events, indicating fundamentally different processes at work. This probably means that the bursts are not phenomena that happen randomly to any primary, but rather that a class of events different from the normal background processes is occurring in bursts. The primaries are also generally reconstructed to higher energies than the secondaries, which are almost exclusively in the 1 and 2 e/h peaks. However, given the current pulse fitting algorithms, it is difficult to tell if the primary is actually a single higher energy event or simply many 1 e/h events occurring nearly simultaneously. One way to answer that may be to reconstruct the physical location of the events in the detector using the difference between the inner and outer phonon channels, but such studies have so far proven inconclusive.

Finally, we note that there is an interesting difference in the spectra below the 1

e/h peak. These events are generally assumed to be produced by 1 e/h events which experience charge trapping and thus do not produce the full NTL energy. Here, the non-burst events show a marked rise in this rate with lower energy. This could be explained by an effect such as carriers injected from the electrode trapping again in the region near the electrode instead of traversing the whole crystal. By contrast, the burst events, especially those of high burst index, show proportionally less sub-1 e/h signal. Again, this points to the bursts being produced by a fundamentally different process than the non-burst pulses.

### 8.2.4 Changing Burst Behavior

It is natural to ask what changes in environment or detector operation have some effect on these pulse bursts. We briefly mention a few such effects observed thus far. A. Mayer at Northwestern demonstrated early on that the rate of bursts appears to be independent of both detector bias voltage and laser rate. Note that this had not been tested for detector biases below 60 V, since the quantized signal is not observed there. This seems to leave open the possibility that the bursts are related to bias voltage, but there is some turn-on threshold below which they do not appear. However, some work by T. Ren has shown evidence that such bursts are still present when the detector is operated at 0 V [244]. If true, this would be very telling as leakage processes like carrier injection are expected to require at least some bias voltage.

One confusing finding was that the event rates increased after grounding and "baking" the detector with an IR LED (940 nm) overnight. The baking process has historically been used with larger SuperCDMS detectors because it was though to neutralize charge trapping sites [204]. Indeed, when operating SNOLAB HV detectors at UMN, we have found that without periodic LED illumination, full NTL amplification is not achieved and the resolution of calibration lines becomes severely degraded. However, the case seems to be different in the HVeV. Fig. 8.10 shows the rate of events during -100 V operation of a Si HVeV detector at NEXUS. The average rate of burst pulses (i.e. all pulses that are part of a burst) was seen to increase dramatically following the bake and took over a week to stabilize, albeit at a lower level than before the bake. The rate immediately after the bake was in fact so high that pileup effects made the detector nearly unusable. It seems that the non-burst rate was relatively stable through this,

contributing about 10 Hz to the total (marked "All Pulses" in Fig. 8.10).
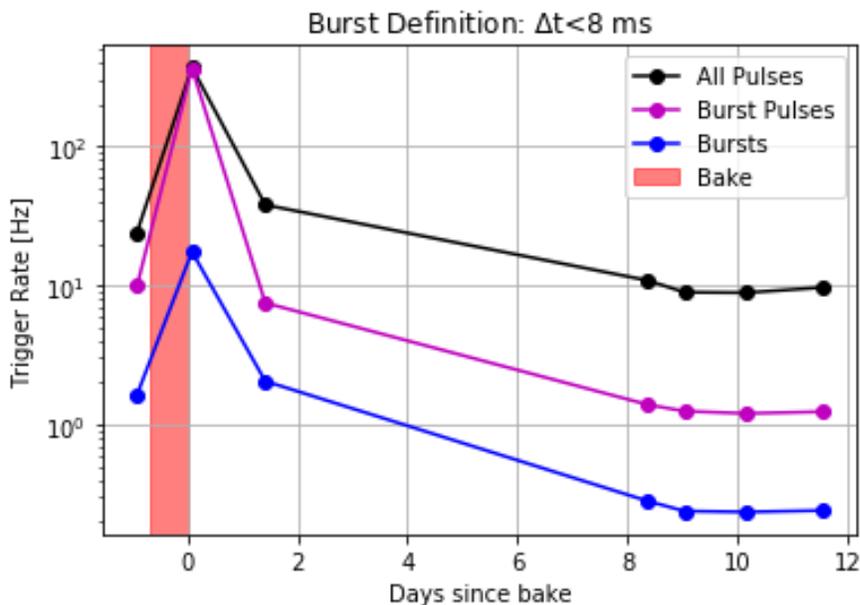


Figure 8.10: Trigger rate of events in HVeV detector at NEXUS. An overnight LED bake generally increased rates for several days afterwards. The average rate of burst pulses was observed to increase after the bake and took over a week to stabilize.

While it is confusing that the same behavior is not observed in larger devices, it is promising that LED baking appears to be one way to modulate the burst properties. It in fact appears to reduce the total rate, if one is patient enough. At very least, it will be important that any realistic model of bursts be able to explain the immediate rate enhancement after baking and the rather long time constant associated with the reduction thereafter. In upcoming experimental runs at NEXUS, there are plans to study baking further and to also test the effects of additional shielding external to the cryostat. If bursts are somehow related to low energy photons such as those from the LED, then perhaps shielding from the ambient background will also have some effect.

### 8.2.5 Speculations on Origins

At this point we have collected some pieces of circumstantial evidence related to pulse bursts and their properties. Here we very briefly speculate on two possible origins;

transition radiation and gated charge leakage. However, there are also other interesting possibilities such as processes where the crystal lattice itself is occasionally shifting and reorganizing near defects.

One source currently being studied by A. Robinson at the University of Montreal is transition radiation which is produced when relativistic charged particles pass a boundary between two different media. In the case of HVeV detectors, this could be initiated by the decay of radioactive contaminants on the copper housing surfaces releasing betas that pass into and out of copper and Si surfaces. It is possible that this can result in bursts of low energy photons produced in the detector cavity. The modelling of such processes is still being pursued by A. Robinson and his research group, but it is an intriguing possibility because it may be already mitigated by other efforts to produce the ultra-clean environment in the SNOLAB cryostat. In that case, the SNOLAB payload may not have to contend with these effects.

Another possible burst culprit is some sort of gated leakage from the electrodes. The idea here is that carriers from either the grounded phonon electrodes or the HV bias electrodes may be able to enter the crystal bulk via inter-band states due to lattice defects near the crystal faces. It is possible that a combination of such states will periodically trap and detrap carriers, which themselves can modulate the trapping on other nearby lattice sites. The periodic detrapping would then manifest as little bursts of single e/h signals. Bursts of events have been extensively studied by those interested in the origins of 1/f noise, especially in semiconductors. Such gating effects have been pointed to as the source of bursty events in MOSFETs [247]. By studying the noise properties of FETs, some have even been able to identify the individual trapping sites responsible for gating and characterize their potential barriers [248]. It is suggestive that such work describes processes in many ways similar to the pulse bursts observed in HVeVs. Interestingly, even the time scales are roughly correct, with gating processes lasting for periods of ms to a few seconds [249]. Unfortunately even our relatively small HVeV detectors are still much larger than the FETs studied elsewhere and would likely have so many leakage gates that only their aggregate effect could be observed. However, if such surface states can be shown to be a source of leakage, there is extensive experience in surface passivation techniques that have been developed in the semiconductor industry and could be applied here.

## 8.3 New Perspectives on Old Problems

The new HVeV detectors are proving an excellent testbed for understanding the main drivers of SuperCDMS detector technologies. We are using them to push the limits of TES sensitivity by increasing their energy efficiency. Even the current generation has reached the point where the factor limiting the sensitivity to small ER energies is not the TES, but rather the Si bandgap energy required to make a single e/h pair. Finally, the limiting factor in previous experiments like CDMSlite was the charge leakage in HV detectors. This is still the largest barrier to realizing the full sensitivity of the SuperCDMS SNOLAB experiment. With the HVeV, we are now able to resolve individual charge leakage events. This is helping us gain insight into the microphysical processes which govern their creation. In particular, correlated event bursts may point to processes such as gated carrier injection from electrodes or external processes like scintillation or transition radiation. Whatever the source turns out to be, it is nearly certain that it will be uncovered utilizing the unique properties of the HVeV detectors.

# Chapter 9

# Conclusions and Outlook

This thesis has touched on a range of topics related to improving SuperCDMS detectors via energy resolution, position reconstruction, NR energy calibration, and increasing detector sizes. After initially covering the experimental motivation, we delved into the world of TESs and discussed improvements in phonon circuit noise in Ch. 4. We then developed several methods to extract event position information given known physical distributions in HV detectors in Ch. 5. We also presented a new cryogenic source moving device which allows calibration of event positions to a new level of accuracy. In Ch. 6, we showed how detector masses may be scaled up by presenting tests of the largest-diameter cryogenic Si detectors yet operated. Next, we took a deep dive into an experiment undertaken to use $(n, \gamma)$ processes for low energy NR ionization yield measurements in Ch. 7. Finally, in Ch. 8, we presented some snapshots of ongoing work with HVeV devices that is both advancing TES sensitivity and probing the nature of leakage currents. Along the way we have explained how these various improvements and advancements serve the SuperCDMS experiment as a whole. In this final chapter we seek to make that somewhat more concrete by demonstrating the effect of such detector developments on the experimental sensitivity to WIMPs.

# 9.1 Effects on WIMP Sensitivity

We explored the effects of four different parameters on experimental sensitivity to WIMP DM; ionization yield, detector energy threshold, detector mass, and volume fiducialization. The most interesting and unique results in this thesis (low energy yield and 150 mm crystals) only pertain to Si and have not yet been studied in Ge. As such, we leave the discussion of such advancements in Ge to future authors pending such experiments. For each of the effects examined, we generated simulated experiments to determine the experimental sensitivity to WIMP parameters. Before presenting the results themselves, we describe the methods used to obtain them.

**Background Spectra**

Rare event searches such as SuperCDMS SNOLAB are limited by various backgrounds due to the detector material, cryostat, and laboratory environment. As such, Super-CDMS has devoted a tremendous amount of effort to assaying and tracking all materials and their contaminants as well as performing extensive modelling and simulations in efforts to understand and reduce all such backgrounds as far as possible [121, 217, 250, 251, 252, 253]. The fruit of this labor is now readily available within the collaboration via tools such as the "Background Explorer", which provides the current best estimates of experimental background rates in the SuperCDMS SNOLAB experiment.

Fig. 9.1 shows the expected background spectra of both ER (solid black) and NR (dashed black) single-scatter events in Si detectors as obtained from Background Explorer. In this analysis, we assumed that multiple-scatter events can be effectively vetoed and do not consider them further. The probability of a WIMP scattering multiple times in a detector is also negligible. Some of the important components of the backgrounds are also shown. The shape of a WIMP recoil spectrum is such that the low energy distribution (above the detector threshold) is generally the main driver of experimental sensitivity. We see here that the lowest few bins of the rate are dominated by NRs from coherent neutrinos (of solar and atmospheric origin) and ERs from detector contamination.

In Si detectors, the dominant contaminants are $^{32}$Si and $^{3}$He. Both are spallation
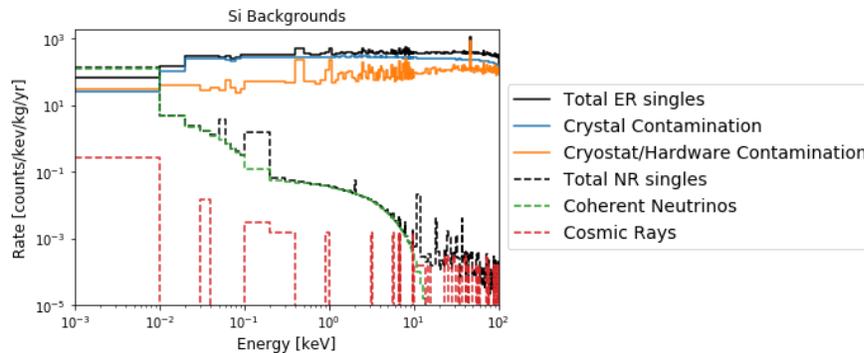
Figure 9.1: Estimated background rates for Si detectors at SNOLAB.

products of cosmic ray secondaries with other materials and result in $\beta$ decays. The $^{3}$He production mechanism is spallation of Si nuclei in the detector crystals [254], so efforts are made to minimize the time fabricated crystals are above ground to reduce this effect. This includes underground storage, avoiding air transport, and even optimizing ground transport routes to avoid time at high elevations. $^{32}$Si is a little trickier to mitigate since it is thought to be produced primarily by spallation of atmospheric argon and becomes incorporated into natural Si deposits via rain and seawater [255]. There have recently been successful efforts to manufacture isotopically pure $^{28}$Si samples using centrifugal separation [256] or mass spectrometry methods [257], but crystals of the size used in SuperCDMS are not yet commercially available or otherwise affordable in the quantities required.

**Generating Projected Sensitivities**

The background rates are used to generate simulated ER and NR events. For a given simulated experimental exposure, an appropriate number of energies are randomly sampled from the distributions shown in Fig. 9.1. The ER events are then converted to the NR scale (to units of $eV_{nr}$) using a yield model and detector bias voltage. An example of such a simulated spectrum for 9.6 kg-yr of Si exposure is shown in Fig. 9.2. The set of simulated event energies is then used to calculate a 90% C.L. exclusion value of the WIMP-nucleon scattering cross section using the Optimum Interval (OI) method [258, 259]. The OI method is designed to calculate an upper limit given measured data

in the presence of unknown backgrounds by considering especially sparse regions where few events are measured. It does not require binning of events and is a conservative approach as it does not assume any information about expected background rates. The recoil spectrum from WIMPs assumes a spin-independent interaction and the standard halo model as in Sec. A.2. In each scenario tested, the limit calculations are repeated five times and the median cross section at each WIMP mass is taken to form the final sensitivity estimate.
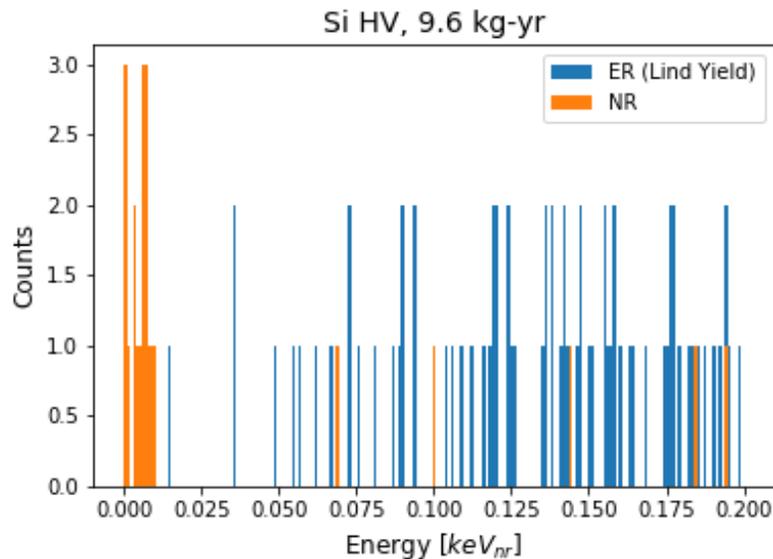


Figure 9.2: Simulated background spectrum for Si detectors at SNOLAB. In this example, the ER energies have been converted to $keV_{nr}$ using the default Lindhard yield with k=0.146 for Si.

For most calculations, we assume the initial SNOLAB payload, which will include two Si iZIPs and 4 Si HV detectors operated over 5 calendar years and expected to obtain 4.8 kg-yr and 9.6 kg-yr of exposure respectively [217]. For the HV detectors, we will assume a flat 85% fiducial volume efficiency for all events above detector threshold. This accounts for analysis cuts which will remove events at high detector radius which may be poorly reconstructed. This is somewhat conservative since the efficiency will be somewhat lower for ERs, which are preferentially near the detector surfaces, but this does not greatly affect the results presented here. The cuts used with iZIPs are slightly

more complex as explained in Sec. 9.1.3 section below.

When showing the resulting projected sensitivity plots, we included the neutrino floor and exclusion limits from several leading experiments for comparison. The neutrino floor is the region where signals from scattering of solar and atmospheric neutrinos will be competitive with DM interaction rates [260]. We only plotted the neutrino curve for Si, which is appropriate for the detectors simulated, but which is also generally more conservative (higher cross-section) than e.g. Ge or Xe. While there are ways to distinguish between neutrino and DM signals (e.g. directionality, annual modulation) this curve represents the regime where neutrinos will make DM discovery claims very difficult for the current generation of detectors. We also show a set of exclusion limits from the following experiments: CRESST-III [261], CDMSlite Run 3 [101], PandaX-II [262], and Xenon1T [263]. This combination forms the leading edge of explored WIMP parameter space in the region where SuperCDMS SNOLAB will be competitive. Finally, we include the official SNOLAB sensitivity projections calculated in [217]. The actual data points that form these neutrino and experimental curves were all taken from the database contained in the Dark Matter Limit Plotter [111].

### 9.1.1 Yield and Threshold

We start by considering the effect of the NR ionization yield function. As discussed in Ch. 7, this quantity is very important for SuperCDMS HV detectors because it is needed to convert measured phonon signals to the NR energy scale, $eV_{nr}$. For events that produce charge, the NTL amplification improves the sensitivity to low energy signals, including both ERs and NRs. Importantly though, the difference in yield between ERs and NRs leads to higher gains for ERs which effectively stretches their spectrum out relative to NRs. The degree of this stretching is determined by the yield function and detector bias. Further, if the yield has a cutoff below which NRs do not ionize carriers, then those events will experience no NTL gain. In this case, increasing the detector bias will amplify ER signals away from the low energy NRs.

We calculated projected sensitivities with the four different yield functions shown as solid curves in Fig. 9.3. The first is the basic Lindhard model with k = 0.146 as predicted for Si [222]. This model has been shown to be accurate at high energies, but appears to overestimate the yield below $\approx$ 5 keV. We also considered the Chavarria

parameterization with k = 0.146 and $a^{-1}$ = 4.0 keV. This empirical fit was selected to agree with Lindhard at high energies and the DAMIC measurements [215] at low energies [120]. In addition, we calculated sensitivity projections with two of the most promising yield functions obtained in the UMN R68 analysis discussed in Ch. 7. These were calculated using the Integral Method of Sec. 7.4.4 and the Sorensen MCMC analysis of Sec. 7.4.3. Note that these final two models both have a hard cutoff below which NRs do not produce ionization. For the Integral Method, we use the single yield curve which best agreed with the Izraelevitch data (red solid line in Fig. 9.3) as well as 100 randomly selected curves which fill the red shaded area. As described in Sec. 7.4.4, this covers a whole family of curves which were in plausible agreement with the Izraelevitch measurements and our data. Note, however, that the shaded region here is extrapolated to higher and lower energies than shown in Fig. 7.33.
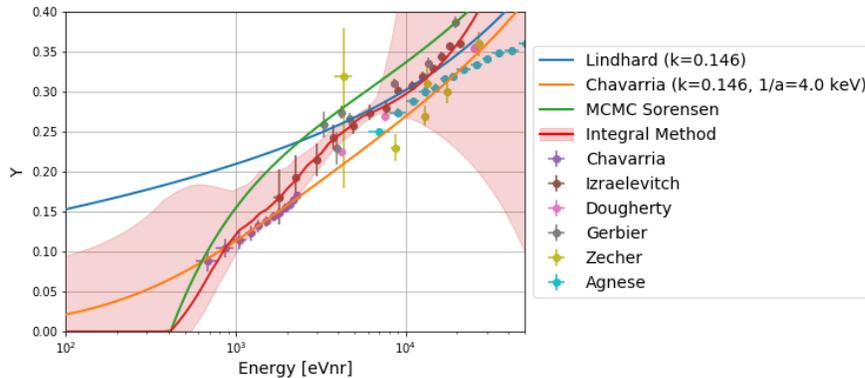


Figure 9.3: Comparison of yield models considered. Lindhard is valid at high energies while Chavarria is also consistent with DAMIC measurements. The Integral and MCMC Sorensen yield functions obtained in R68 both cut off near 400 $eV_{nr}$. The red shaded region is bounded by two extreme yield functions obtained with the Integral Method.

To perform the OI calculation we must also prescribe a detector energy threshold, in $eV_{nr}$, below which the signal efficiency is 0. Current SuperCDMS projections assume the HV detectors will be sensitive to single e/h signals when operated at 100 V bias [217]. This requirement is essentially that the threshold be 100 $eV_{tot}$, where $eV_{tot}$ denotes total phonon energy (recoil plus NTL). This translates to different $eV_{nr}$ values depending on the yield function. For Lindhard and Chavarria, the thresholds are 23 $eV_{nr}$ and 71 $eV_{nr}$ respectively. However, the Integral and MCMC Sorensen models cut off high

enough in energy that the only way for an NR to generate 100 $eV_{tot}$ is for there to be no ionization produced, with the full 100 eV coming from a 100 $eV_{nr}$ recoil. For this reason, the Integral and MCMC Sorensen models are assigned thresholds of 100 $eV_{nr}$ for this calculation. The resulting projected sensitivities under these assumptions are shown in Fig. 9.4.
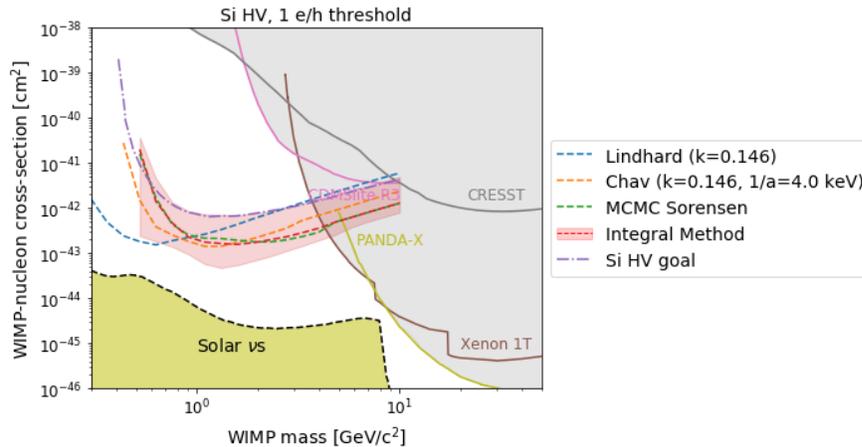


Figure 9.4: WIMP sensitivity curves for Si HV detectors using different yield models. Assuming nominal SNOLAB payload and exposure with 1 e/h phonon threshold.

We find here that the Lindhard model provides the greatest reach to low mass WIMPs due the lower energy threshold. The two cutoff yield models (Integral and MCMC Sorensen) give very similar results to each other, although the full range covered by the Integral Method is quite a bit larger. The Chavarria yield function naturally lies somewhere between these extremes, but has nowhere near the reach which would be obtained if Lindhard held at low energies. While the low-mass cutoff of a sensitivity projection is generally set by the energy threshold value, the reach in cross section at higher masses is also related to the background rate at higher energies, which is dominated by ERs. But, recall that the density of ER hits in $eV_{nr}$ units is set by the yield function, with lower yield values (such as with Integral and MCMC Sorensen) converting ERs to higher $eV_{nr}$ values, and thus thinning out the spectrum. In this way, the low-yield models can sometimes result in improved cross section reach at high WIMP masses. The sensitivities here are somewhat stronger than the projection from [217], labelled "Si HV goal". This is primarily due to the different yield functions

assumed, with agreement seen with Lindhard at high mass and Chavarria at low mass corresponding to agreements in the yield functions at high and low energies respectively.

Now, as discussed elsewhere in this thesis, the energy threshold of HV detectors is driven by leakage currents which are not fully understood. At this point, no HV detector has yet exhibited the 1 e/h resolution assumed above. We next consider a more conservative threshold case. In UMN R68, we worked very hard to obtain a 50 $eV_{ee}$ threshold with a Si HV detector. Fig. 9.5 shows the result of moving instead to that condition which corresponds to a range of thresholds from 246 $eV_{nr}$ for Lindhard to 603 $eV_{nr}$ for MCMC Sorensen. At these higher energies, even the cutoff yield models will generate some ionization.
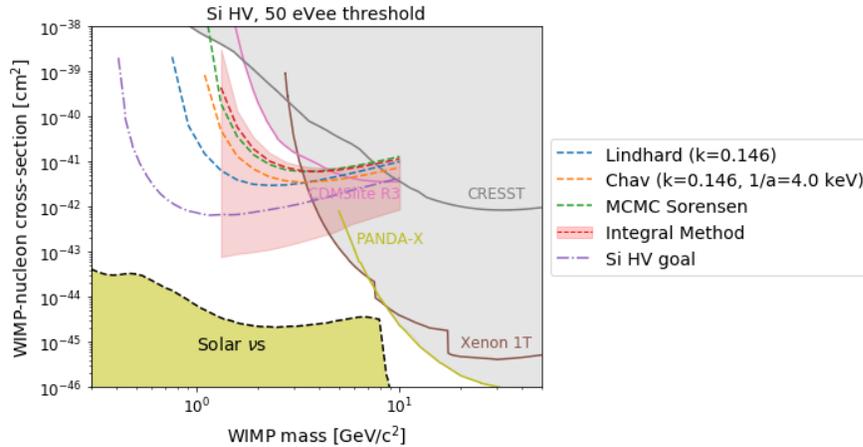


Figure 9.5: WIMP sensitivity curves for Si HV detectors using different yield models. Assuming nominal SNOLAB payload and exposure with 50 eVee phonon threshold.

This represents a rather dramatic scale back in experimental reach. We have raised the thresholds by factors of 5–10 and see the mass reach pulled back proportionally. The cross section also rises by a factor of 5 because the background rate near threshold rises as we increase the threshold energy. However, this should be taken as a very pessimistic scenario as the 50 $eV_{ee}$ analysis threshold of R68 was driven by an abundance of spurious triggers on phonon noise due to a simplistic triggering algorithm, non-optimal noise conditions, and the generally high-rate environment at a surface facility. At SNOLAB, the raw noise and rate will be improved and the simple threshold triggering will be replaced by a real-time Optimal Filter (OF) trigger, similar to that used in [264]. Given

these considerations, we assume a more realistic threshold of 10 eV$_{ee}$ can be achieved even without eliminating the leakage current. This value was selected as it corresponds to the OF baseline resolution observed in R68, which would be near that which would have been attainable with an OF trigger.
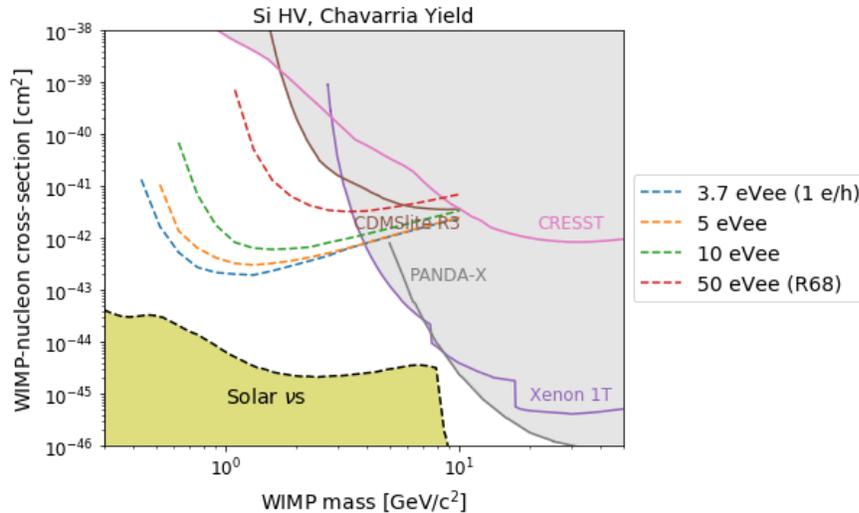


Figure 9.6: Effect of different analysis thresholds assuming the Chavarria yield function.

For comparison, Fig. 9.6 shows the sensitivity projections obtained with a range of thresholds. We see here how threshold primarily dictates the low-mass reach. However, once the threshold is high enough, it also begins to pull up the cross sections as well. We have selected the Chavarria yield function for this figure and will use it for the remaining studies. However, as we found above, if the yield function has a cutoff as seen in R68, the sensitivities will be similar, but slightly weaker.

### 9.1.2 Detector Size

We next consider the case of increasing exposure by changing the detector size. In Ch. 6, we demonstrated successful operation of 150 mm diameter Si crystals with charge readout. The next natural step will be to confirm that crystals of this size will also perform well as phonon detectors, similar to the current generation of 100 mm HV detectors. Though none have been fabricated yet, we considered a hypothetical upgrade of the SNOLAB payload where the four 100 mm Si HVs are replaced with 150 mm

versions. We also examined the case where the entire SNOLAB cryostat is used to house a payload of as many 150 mm Si detectors as will fit. The mixing chamber can is currently designed to house seven towers, each capable of holding a stack of six 100 mm detectors (however only four such towers will initially be deployed). Ignoring the many engineering details required for such an upgrade, we find that we could instead fit a maximum of four towers with housing wide enough for 150 mm detectors, for a total of 24 such detectors. Both of these scenarios were considered and compared to the nominal payload. The projected sensitivities calculated with the OI method are shown as dashed curves in Fig. 9.7.
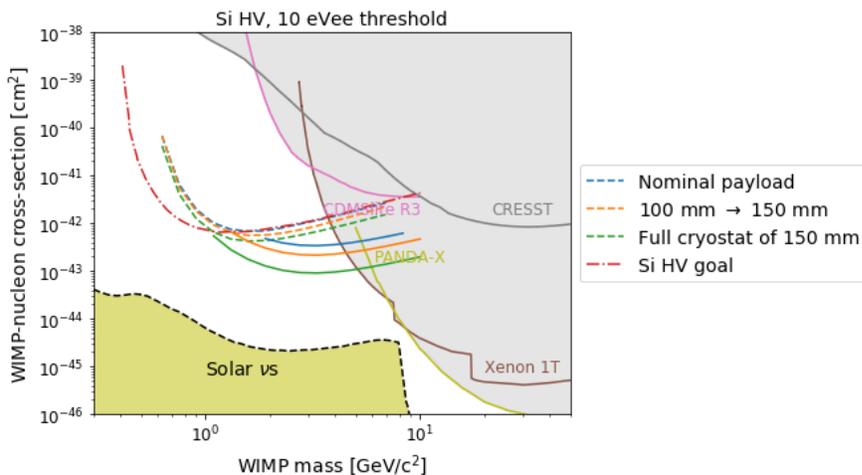


Figure 9.7: Effect of increasing Si HV detector mass. Orange curve shows effect of simply replacing the nominal Si HV payload of four 100 mm detectors with four 150 mm detectors. The green curve indicates the gains obtained by filling the entire SONLAB cryostat with only 150 mm Si HV detectors. For the methods tested here, dashed curves were obtained with the OI method while solid lines were calculated using a Poisson method including a background estimate [265].

Here we find a much less dramatic result than when considering yield and threshold above. When simply increasing the detector diameters to 150 mm, the cross section reach improved by a factor of 1.2. This is somewhat less than the raw ratio of masses (2.25), which is how the cross section would scale for a totally background-free experiment. Similarly, when moving from the nominal payload to a full cryostat of 150 mm

detectors, the cross-section improved by a modest factor of 1.7 even though the exposure increased by a factor of 13.5. This is because the experiment is primarily limited by background signals, not exposure.

The OI method does not use any information about the expected background rate when calculating a limit. For this reason, the projected sensitivity does not scale as strongly with exposure compared to other methods which do include such information. For example, if we instead treat this as a simple counting experiment with a single bin from the detector threshold to 1 $keV_{nr}$, we can calculate sensitivity projections using the Poisson method described in Sec. 28.6.4 of [265]. These are also shown in Fig. 9.7 using solid lines of corresponding color to the OI calculations. These sensitivities tend to scale with the Poisson uncertainty on the number of counts in the bin, so they improve proportionally to the square root of total exposure. However, as discussed in [217], we generally choose to use the conservative OI calculation for such projections to avoid biases induced from binning choices.

In all three cases calculated with OI, the low-mass reach did not change substantially because the energy thresholds were the same. So we find that even dramatically increasing the mass (or equivalently livetime) only provides modest scientific gains unless paired with threshold and/or background reductions. However, for a given realization of an experiment, stronger sensitivity projections may be set than with OI if one uses techniques (e.g. the Poisson method above or profile likelihood fits [101]) which take the expected background spectrum into account.

### 9.1.3 Fiducialization

Finally, we consider the case of volume fiducialization in the Si iZIP detectors. In Sec. 5.2, we demonstrated a device and procedures to improve position reconstruction in an iZIP. While this can improve fiducialization in HVs as well, the results are better in the iZIP, which is considered here.

We treated the iZIPs a bit differently from the HV detectors. First, we assumed the nominal payload and exposure of 2 Si iZIPs operated for 4.8 kg-yr at bias voltages of $\pm 5$ V. Recall that the interleaved phonon rails are grounded and the charge channels on opposite detector faces are biased to opposite signs, so this provides an NTL voltage of 10 V for events occurring in the crystal bulk. As in [217], we will assume a phonon

analysis threshold of 166 eVnr, which is driven by phonon circuit noise and is not affected by leakage currents at these low voltages. Importantly, the iZIP is able to discriminate between ER and NR events based on yield and has demonstrated rejection of ERs at the level of $1.7 \times 10^{-5}$ [159]. We will assume such a cut efficiency for ERs above the expected charge resolution of $\approx 1$ keV$_{nr}$, with ERs at lower energies passing the cut. The efficiency will not be a step function, but will roll off due to the Gaussian charge resolution. We incorporate this using an error function centered at 1 keV$_{nr}$ with width 0.3 keV$_{nr}$, consistent with the efficiency implied by Fig. 7 of [217].

We tested the position reconstruction effects in a simple manner. In iZIPs, cuts are generally made to remove events near the crystal surfaces by considering the charge signals. The interleaved channels allow events within about 1 mm of the faces to be identified and removed. The outer charge electrode, named "Qouter", covers radii from 43.2 – 50 mm and is used to identify events near the outer detector surface. Assuming this is able to identify all such events beneath the electrode, the remaining fiducial volume fraction from these two cuts is about 72%. For comparison, we considered a hypothetical improved radial cut based on studies of phonon signals using the source mover. By using detailed studies with such a device, one could develop a complete map of the phonon channel responses at all positions in the detector volume. This will be especially important at high radius where position-dependent corrections to energy can be found. With such a calibration, we believe one could also reliably reconstruct the radius to within 1 mm. This is approximately the resolution attained in the most recent studies with the source mover. Assuming the energy and radius are able to be properly reconstructed for high radius events, an improved radial cut could reliably remove all events with radii from 49–50 mm, improving the fiducial volume fraction to 93%. The results of simulations with these assumptions are shown in Fig. 9.8 where the projected SNOLAB sensitivity is also shown for comparison.

Here we observed sensitivity curves with two distinct lobes. The higher mass region corresponds to the higher recoil energy range where the yield cut is able to remove most backgrounds. In this region, the improved radial cut lowers the cross section linearly with added fiducial mass (i.e. by the a factor of $93/72 \approx 1.3$). At lower energies, there is still some ER background and the reach almost scales the square root of fiducial mass (though it is slightly weaker as we found in the 150 mm study above). The kink
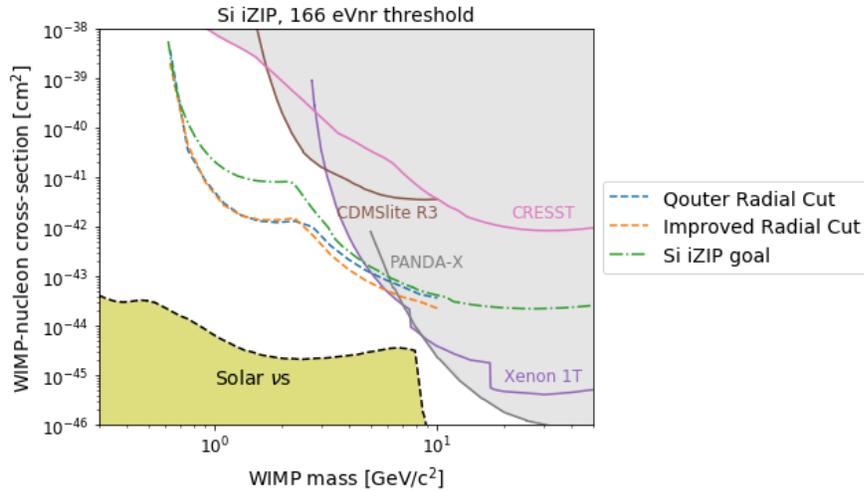
Figure 9.8: Experimental reach for nominal payload of Si iZIPs for different radial cuts. Using an improved phonon cut to exclude only the outer 1 mm of the detector volume slightly improves the cross-section reach.

in the curves is the transition between these two regimes and corresponds to the OI method selecting a different energy interval to set the limit. Both simulated curves are qualitatively similar to the official projection, but are somewhat lower between 1–2 GeV. This is due to the different yield function used in that analysis, with the Chavarria function used here smearing the ERs to higher energies, thus improving the reach between threshold and the onset of the yield cut at $\approx 1$ keV$_{\mathrm{nr}}$.

As modelled here, the fiducialization imposes an efficiency on both ERs and NRs which is equal to the fiducial volume retained. This is equivalent to assuming that events are uniformly distributed through the detector volume. While this is close to true for most NRs, and certainly correct for WIMPs, many ER backgrounds will be more densely located near the detector edges. This is, of course, one of the reasons we make fiducial cuts in the first place, because they preferentially remove ERs. While fully accounting for these effects would take more detailed modelling (including tracking of all event origins and interaction locations) doing so is only expected to shift the sensitivities in Fig. 9.8 by $\sim 10\%$.

## 9.2  Outlook

The work presented in this thesis has covered several advancements related to cryogenic detector performance that are crucial to the SuperCDMS experiment. Of all these aspects, I have come to believe that mitigation of the HV leakage is currently the most important technical barrier to overcome. Some argue that the leakage source is inherent to surface operation of the detectors because it is somehow related to cosmic rays or other backgrounds. In such a case, the leakage will be solved and single e/h thresholds achieved by simply moving to the well-shielded SNOLAB environment. However, if this not the case, it is likely that further studies of the leakage using HVeVs, as discussed in Ch. 8, will be essential. To be sure, the experiment can still be successful in carving out substantial new parameter space even if the leakage persists at current levels (see Sec. VII of [217]). However, we should continue to aggressively research these effects in order to realize the full sensitivity of the SuperCDMS experiment at SNOLAB.

Even if the leakage is improved underground, efforts should be made to understand and mitigate it at the surface. This will help advance the use and utility of the fantastic QET sensor technology developed by SuperCDMS members. These leakage effects are also interesting more broadly because they are likely related to dark currents observed in a range of other types of semiconductor detectors and devices as well. Beyond that, advancements can be made in the current detector architecture through features such as QET efficiency, crystal size, and channel layout optimization for position sensitivity. Additionally, to extend the useful range of all such devices, it is important to complete the work of characterizing NR ionization yield down to the lowest possible energies. We have shown that novel techniques involving $(n, \gamma)$ processes can play a useful role in that effort. These improvements are not just important for academic reasons, rather SuperCDMS must continue its history of innovative detector R&D or risk being outpaced by our many (and formidable) competitors.

My grad student tenure has spanned the sometimes awkward period between two eras of SuperCDMS; Soudan and SNOLAB. Ending one experiment and gearing up for another has had its share of fits and starts. The collaboration has had to transform from a finely tuned DM data-taking and analysis machine to scattered bands of speculative R&D projects paired with an organizing force planning a massive scientific and

engineering undertaking. Now, with construction at SNOLAB underway, the shift has again begun to forming a cohesive force capable of staffing and operating this ambitious project. Through all such changes, there has necessarily been tension between our desire to push the technologies ever further and the practical need to build a reliable, working experiment. While much of my work has been in the former category, there will soon need to be a large shift of laboratory and analysis effort to the main DM search at SNOLAB. While this will likely mean fewer collaboration resources devoted to R&D it is important that such work continue as much as possible to create future generations of SuperCDMS DM searches. However, this transition also comes with new opportunities for analysis and simulation R&D. While I have not mentioned it at all in this thesis, considerable work is being done by younger members to take new techniques from data science and machine learning and bring them to bear on thorny problems like detailed event reconstruction. While there will be formidable challenges ahead, I am confident that the dedicated and skilled members of the SuperCDMS collaboration are up to the task!

As presented at the beginning of this work, much has been learned about DM since the beginning of the last century. We have come to understand its role in the evolution of the universe, and its effect on structure formation and evolution. However, the ultimate nature of DM remains one of the largest open questions in physics. This situation is certainly not due to lack of effort, since scores of experimental searches (built of uncountable millions of person-hours) have pursued various and creative approaches to gleaning any further evidence about the properties of DM. While much has been learned about what DM is not, there are certainly still many possibilities open which we have either not considered theoretically or are not yet able to achieve technologically. To describe this situation, I borrow the unofficial slogan of the Madison Symmetric Torus (another perhaps-Quixotic project where I began my physics career) which reminds us that, "NATURE BATS LAST". Nevertheless, the physics community is undaunted and as energetic as ever in this quest! Besides, if this stuff was easy, it wouldn't be so fun!

# References

[1] Gianfranco Bertone and Dan Hooper. History of dark matter. *Rev. Mod. Phys.*, 90:045002, Oct 2018.

[2] Baron William Thomson Kelvin. *Baltimore lectures on molecular dynamics and the wave theory of light*. 1904.

[3] F. Zwicky. Republication of: The redshift of extragalactic nebulae. *General Relativity and Gravitation*, 41(1):207–224, Jan 2009.

[4] V. C. Rubin, W. K. Ford, Jr., and N. Thonnard. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R = 4kpc) to UGC 2885 (R = 122 kpc). *Astrophysical Journal*, 238:471–487, June 1980.

[5] K. G. Begeman, A. H. Broeils, and R. H. Sanders. Extended rotation curves of spiral galaxies: dark haloes and modified dynamics. *Monthly Notices of the Royal Astronomical Society*, 249(3):523–537, 04 1991, http://oup.prod.sis.lan/mnras/article-pdf/249/3/523/18160929/mnras249-0523.pdf.

[6] Fabio Iocco, Miguel Pato, and Gianfranco Bertone. Evidence for dark matter in the inner Milky Way. *Nature Physics*, 11(3):245–248, 2015.

[7] Sean M. Carroll. *Spacetime and geometry: An introduction to general relativity*. Pearson, 2004.

[8] O. Chwolson. Über eine mögliche form fiktiver doppelsterne. *Astronomische Nachrichten*, 221(20):329–330, 1924.

[9] Philippe Fischer, Timothy A. McKay, Erin Sheldon, Andrew Connolly, Albert Stebbins, Joshua A. Frieman, Bhuvnesh Jain, Michael Joffre, David Johnston, Gary Bernstein, James Annis, Neta A. Bahcall, J. Brinkmann, Michael A. Carr, István Csabai, James E. Gunn, G. S. Hennessy, Robert B. Hindsley, Charles Hull, Željko Ivezić, G. R. Knapp, Siriluk Limmongkol, Robert H. Lupton, Jeffrey A. Munn, Thomas Nash, Heidi Jo Newberg, Russell Owen, Jeffrey R. Pier, Constance M. Rockosi, Donald P. Schneider, J. Allyn Smith, Chris Stoughton, Alexander S. Szalay, Gyula P. Szokoly, Aniruddha R. Thakar, Michael S. Vogeley, Patrick Waddell, David H. Weinberg, Donald G. York, and (The SDSS Collaboration). Weak lensing with sloan digital sky survey commissioning data: The galaxy-mass correlation function to 1 $h^{-1}$ mpc. *The Astronomical Journal*, 120(3):1198–1208, sep 2000.

[10] Henk Hoekstra, H. K. C. Yee, and Michael D. Gladders. Properties of galaxy dark matter halos from weak lensing. *The Astrophysical Journal*, 606(1):67–77, may 2004.

[11] Douglas Clowe, Maruša Bradač, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *The Astrophysical Journal*, 648(2):L109–L113, aug 2006.

[12] Maxim Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, L. David, W. Forman, C. Jones, S. Murray, and W. Tucker. Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56. *Astrophys. J.*, 606:819–824, 2004, astro-ph/0309303.

[13] D.M. Scolnic et al. The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample. *Astrophys. J.*, 859(2):101, 2018, 1710.00845.

[14] N. Aghanim et al. Planck 2018 results. VI. Cosmological parameters. 7 2018, 1807.06209.

[15] G. E. Addison, G. Hinshaw, and M. Halpern. Cosmological constraints from baryon acoustic oscillations and clustering of large-scale structure.

*Monthly Notices of the Royal Astronomical Society*, 436(2):1674–1683, 09 2013, https://academic.oup.com/mnras/article-pdf/436/2/1674/3949663/stt1687.pdf.

[16] Keith A. Olive. TASI lectures on dark matter. In *Particle physics and cosmology: The quest for physics beyond the standard model(s). Proceedings, Theoretical Advanced Study Institute, TASI 2002, Boulder, USA, June 3-28, 2002*, pages 797–851, 2003, astro-ph/0301505.

[17] Edward W. Kolb and Michael S. Turner. The Early Universe. *Front. Phys.*, 69:1–547, 1990.

[18] Brian Fields and Subir Sarkar. Big-Bang nucleosynthesis (2006 Particle Data Group mini-review). 2006, astro-ph/0601514.

[19] M. Peimbert and S. Torres-Peimbert. Chemical composition of H II regions in the Large Magellanic Cloud and its cosmological implications. *The Astrophysical Journal*, 193:327–333, October 1974.

[20] Erik Aver, Keith A. Olive, and Evan D. Skillman. The effects of He I $\lambda 10830$ on helium abundance determinations. *JCAP*, 07:011, 2015, 1503.08146.

[21] Antonio Peimbert, Manuel Peimbert, and Valentina Luridiana. The primordial helium abundance and the number of neutrino families. *Rev. Mex. Astron. Astrofis.*, 52(2):419–424, 2016, 1608.02062.

[22] Vital Fernández, Elena Terlevich, Angeles I Díaz, Roberto Terlevich, and F F Rosales-Ortega. Primordial helium abundance determination using sulphur as metallicity tracer. *Monthly Notices of the Royal Astronomical Society*, 478(4):5301–5319, 05 2018.

[23] Ryan J. Cooke and Michele Fumagalli. "measurement of the primordial helium abundance from the intergalactic medium". *Nature Astronomy*, 2(12):957 – 961, 2018.

[24] P. A. R. Ade et al. Planck 2015 results - xiii. cosmological parameters. *A&A*, 594:A13, 2016.

[25] John M. O'Meara, David Tytler, David Kirkman, Nao Suzuki, Jason X. Prochaska, Dan Lubin, and Arthur M. Wolfe. The deuterium to hydrogen abundance ratio toward a fourth QSO: HS 0105+1619. *The Astrophysical Journal*, 552(2):718–730, may 2001.

[26] D. J. Fixsen. The temperature of the cosmic microwave background. *The Astrophysical Journal*, 707(2):916–920, nov 2009.

[27] R. Adam et al. Planck 2015 results. I. Overview of products and scientific results. *Astron. Astrophys.*, 594:A1, 2016, 1502.01582.

[28] Wayne Hu, Naoshi Sugiyama, and Joseph Silk. he physics of microwave background anisotropies. *Nature*, 386:37–43, march 1997.

[29] S. Masi, P. De Bernardis, G. De Troia, M. Giacometti, A. Iacoangeli, F. Piacentini, G. Polenta, P.A.R. Ade, P.D. Mauskopf, J.J. Bock, J.R. Bond, C.R. Contaldi, D. Pogosyan, S. Prunet, J. Borrill, A. Boscaleri, E. Pascale, K. Coble, P. Farese, T. Montroy, J.E. Ruhl, B.P. Crill, V.V. Hristov, W.C. Jones, A.E. Lange, P. Mason, G. De Gasperis, P. Natoli, K. Ganga, E. Hivon, N. Vittorio, A.H. Jaffe, L. Martinis, F. Scaramuzzi, A. Melchiorri, C.B. Netterfield, F. Pongetti, and G. Romeo. The boomerang experiment and the curvature of the universe. *Progress in Particle and Nuclear Physics*, 48(1):243 – 261, 2002.

[30] R. Stompor, M. Abroe, P. Ade, A. Balbi, D. Barbosa, J. Bock, J. Borrill, A. Boscaleri, P. de Bernardis, P. G. Ferreira, S. Hanany, V. Hristov, A. H. Jaffe, A. T. Lee, E. Pascale, B. Rabii, P. L. Richards, G. F. Smoot, C. D. Winant, and J. H. P. Wu. Cosmological implications of the MAXIMA-1 high-resolution cosmic microwave background anisotropy measurement. *The Astrophysical Journal*, 561(1):L7–L10, nov 2001.

[31] J. R. Bond, G. Efstathiou, and J. Silk. Massive neutrinos and the large-scale structure of the universe. *Phys. Rev. Lett.*, 45:1980–1984, Dec 1980.

[32] M. D. Lehnert, N. P. H. Nesvadba, J.-G. Cuby, A. M. Swinbank, S. Morris, B. Clément, C. J. Evans, M. N. Bremer, and S. Basa. Spectroscopic confirmation of a galaxy at redshift z = 8.6. *Nature*, 467:940, Oct 2010.

[33] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: Evidence, candidates and constraints. *Phys. Rept.*, 405:279–390, 2005, hep-ph/0404175.

[34] Tisserand, P., Le Guillou, L., Afonso, C., Albert, J. N., Andersen, J., Ansari, R., Aubourg, É., Bareyre, P., Beaulieu, J. P., Charlot, X., Coutures, C., Ferlet, R., Fouqué, P., Glicenstein, J. F., Goldman, B., Gould, A., Graff, D., Gros, M., Haissinski, J., Hamadache, C., de Kat, J., Lasserre, T., Lesquoy, É., Loup, C., Magneville, C., Marquette, J. B., Maurice, É., Maury, A., Milsztajn, A., Moniez, M., Palanque-Delabrouille, N., Perdereau, O., Rahal, Y. R., Rich, J., Spiro, M., Vidal-Madjar, A., Vigroux, L., and S. Zylberajch (The EROS-2 collaboration). Limits on the macho content of the galactic halo from the eros-2 survey of the magellanic clouds. *A&A*, 469(2):387–404, 2007.

[35] Leslie J Rosenberg and Karl A. van Bibber. Searches for invisible axions. *Physics Reports*, 325(1):1 – 39, 2000.

[36] Georg G. Raffelt. Axions in astrophysics and cosmology. In *Dark matter in cosmology, clocks and test of fundamental laws. Proceedings, 30th Rencontres de Moriond, 15th Moriond Workshop, Villars sur Ollon, Switzerland, January 22-29, 1995*, pages 159–168, 1995, hep-ph/9502358.

[37] Georg G. Raffelt. *Astrophysical Axion Bounds*, pages 51–71. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[38] K.A. Olive. Review of particle physics. *Chinese Physics C*, 38(9):090001, aug 2014.

[39] George R. Blumenthal, S. M. Faber, Joel R. Primack, and Martin J. Rees. Formation of galaxies and large-scale structure with cold dark matter. *Nature*, 311:517–525, 1984.

[40] Marco Battaglieri et al. US Cosmic Visions: New Ideas in Dark Matter 2017: Community Report. In *U.S. Cosmic Visions: New Ideas in Dark Matter College Park, MD, USA, March 23-25, 2017*, 2017, 1707.04591.

[41] Mauro Raggi and Venelin Kozhuharov. Results and perspectives in dark photon physics. *Riv. Nuovo Cim.*, 38(10):449–505, 2015.

[42] Tongyan Lin. Dark matter models and direct detection. *PoS*, 333:009, 2019, 1904.07915.

[43] Bob Holdom. Two U(1)'s and Epsilon Charge Shifts. *Phys. Lett.*, 166B:196–198, 1986.

[44] Joerg Jaeckel. A force beyond the Standard Model - Status of the quest for hidden photons. *Frascati Phys. Ser.*, 56:172–192, 2012, 1303.1821.

[45] Yonit Hochberg, Tongyan Lin, and Kathryn M. Zurek. Absorption of light dark matter in semiconductors. *Phys. Rev.*, D95(2):023013, 2017, 1608.01994.

[46] Haipeng An, Maxim Pospelov, Josef Pradler, and Adam Ritz. Direct Detection Constraints on Dark Photon Dark Matter. *Phys. Lett.*, B747:331–338, 2015, 1412.8378.

[47] R. Agnese et al. First Dark Matter Constraints from a SuperCDMS Single-Charge Sensitive Detector. *Phys. Rev. Lett.*, 121(5):051301, 2018, 1804.10697. [erratum: Phys. Rev. Lett.122,no.6,069901(2019)].

[48] D.W. Amaral et al. Constraints on low-mass, relic dark matter candidates from a surface-operated SuperCDMS single-charge sensitive detector. 2005.14067.

[49] Lawrence J. Hall, Karsten Jedamzik, John March-Russell, and Stephen M. West. Freeze-In Production of FIMP Dark Matter. *JHEP*, 03:080, 2010, 0911.1120.

[50] Rouven Essig, Marivi Fernández-Serra, Jeremy Mardon, Adrián Soto, Tomer Volansky, and Tien-Tien Yu. Direct detection of sub-gev dark matter with semiconductor targets. *Journal of High Energy Physics*, 2016(5):46, May 2016.

[51] Tracy R. Slatyer. Indirect Detection of Dark Matter. In *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics : Anticipating the Next Discoveries in Particle Physics (TASI 2016): Boulder, CO, USA, June 6-July 1, 2016*, pages 297–353, 2018, 1710.05137.

[52] Dan Hooper. TASI Lectures on Indirect Searches For Dark Matter. *PoS*, TASI2018:010, 2019, 1812.02029.

[53] A. Albert et al. Searching for Dark Matter Annihilation in Recently Discovered Milky Way Satellites with Fermi-LAT. *Astrophys. J.*, 834(2):110, 2017, 1611.03184.

[54] Tansu Daylan, Douglas P. Finkbeiner, Dan Hooper, Tim Linden, Stephen K. N. Portillo, Nicholas L. Rodd, and Tracy R. Slatyer. The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter. *Phys. Dark Univ.*, 12:1–23, 2016, 1402.6703.

[55] Richard Bartels, Suraj Krishnamurthy, and Christoph Weniger. Strong support for the millisecond pulsar origin of the Galactic center GeV excess. *Phys. Rev. Lett.*, 116(5):051102, 2016, 1506.05104.

[56] Dan Hooper, Chris Kelso, and Farinaldo S. Queiroz. Stringent and Robust Constraints on the Dark Matter Annihilation Cross Section From the Region of the Galactic Center. *Astropart. Phys.*, 46:55–70, 2013, 1209.3015.

[57] Lucia Rinchiuso. Latest results on dark matter searches with H.E.S.S. *EPJ Web Conf.*, 209:01023, 2019, 1901.05299.

[58] M.L. Ahnen et al. Indirect dark matter searches in the dwarf satellite galaxy Ursa Major II with the MAGIC Telescopes. *JCAP*, 03:009, 2018, 1712.03095.

[59] Benjamin Zitzer. The VERITAS Dark Matter Program. *PoS*, ICRC2017:904, 2018, 1708.07447.

[60] Alessandro Cuoco, Michael Krämer, and Michael Korsmeier. Novel Dark Matter Constraints from Antiprotons in Light of AMS-02. *Phys. Rev. Lett.*, 118(19):191102, 2017, 1610.03071.

[61] Ming-Yang Cui, Qiang Yuan, Yue-Lin Sming Tsai, and Yi-Zhong Fan. Possible dark matter annihilation signal in the AMS-02 antiproton data. *Phys. Rev. Lett.*, 118(19):191101, 2017, 1610.03840.

[62] Alejandro Ibarra, Anna S. Lamperstorfer, and Joseph Silk. Dark matter annihilations and decays after the AMS-02 positron measurements. *Phys. Rev. D*, 89(6):063539, 2014, 1309.2570.

[63] O. Adriani et al. Cosmic-Ray Positron Energy Spectrum Measured by PAMELA. *Phys. Rev. Lett.*, 111:081102, 2013, 1308.0133.

[64] M.G. Aartsen et al. Search for annihilating dark matter in the Sun with 3 years of IceCube data. *Eur. Phys. J. C*, 77(3):146, 2017, 1612.05949. [Erratum: Eur.Phys.J.C 79, 214 (2019)].

[65] S. Adrián-Martínez et al. Limits on dark matter annihilation in the sun using the antares neutrino telescope. *Physics Letters B*, 759:69–74, 2016.

[66] K. Choi et al. Search for neutrinos from annihilation of captured low-mass dark matter particles in the Sun by Super-Kamiokande. *Phys. Rev. Lett.*, 114(14):141301, 2015, 1503.04858.

[67] S. J. Asztalos, G. Carosi, C. Hagmann, D. Kinion, K. van Bibber, M. Hotz, L. J Rosenberg, G. Rybka, J. Hoskins, J. Hwang, P. Sikivie, D. B. Tanner, R. Bradley, and J. Clarke. Squid-based microwave cavity search for dark-matter axions. *Phys. Rev. Lett.*, 104:041301, Jan 2010.

[68] T. Braine, R. Cervantes, N. Crisosto, N. Du, S. Kimes, L. J. Rosenberg, G. Rybka, J. Yang, D. Bowring, A. S. Chou, R. Khatiwada, A. Sonnenschein, W. Wester, G. Carosi, N. Woollett, L. D. Duffy, R. Bradley, C. Boutan, M. Jones, B. H. LaRoque, N. S. Oblath, M. S. Taubman, J. Clarke, A. Dove, A. Eddins, S. R. O'Kelley, S. Nawaz, I. Siddiqi, N. Stevenson, A. Agrawal, A. V. Dixit, J. R. Gleason, S. Jois, P. Sikivie, J. A. Solomon, N. S. Sullivan, D. B. Tanner, E. Lentz, E. J. Daw, J. H. Buckley, P. M. Harrington, E. A. Henriksen, and K. W. Murch. Extended search for the invisible axion with the axion dark matter experiment. *Phys. Rev. Lett.*, 124:101303, Mar 2020.

[69] V. Anastassopoulos et al. New cast limit on the axion–photon interaction. *Nature Physics*, 13:584–590, 2017.

[70] M. Arik et al. Search for Solar Axions by the CERN Axion Solar Telescope with $^3$He Buffer Gas: Closing the Hot Dark Matter Gap. *Phys. Rev. Lett.*, 112(9):091302, 2014, 1307.1985.

[71] D. Banerjee et al. Dark matter search in missing energy events with NA64. *Phys. Rev. Lett.*, 123(12):121801, 2019, 1906.00176.

[72] J. P. Lees et al. Search for Invisible Decays of a Dark Photon Produced in $e^+e^-$ Collisions at BaBar. *Phys. Rev. Lett.*, 119(13):131804, 2017, 1702.03327.

[73] Morad Aaboud et al. Constraints on mediator-based dark matter and scalar dark energy models using $\sqrt{s} = 13$ TeV $pp$ collision data collected by the ATLAS detector. *JHEP*, 05:142, 2019, 1903.01400.

[74] Albert M Sirunyan et al. Search for dark matter produced in association with a single top quark or a top quark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 03:141, 2019, 1901.01553.

[75] Creative commons attribution 4.0 international license. `https://creativecommons.org/licenses/by/4.0/`.

[76] J. I. Read. The Local Dark Matter Density. *J. Phys.*, G41:063101, 2014, 1404.1938.

[77] Ronald Drimmel and Eloisa Poggio. On the solar velocity. *Research Notes of the AAS*, 2(4):210, nov 2018.

[78] A.Liam Fitzpatrick, Wick Haxton, Emanuel Katz, Nicholas Lubbers, and Yiming Xu. The Effective Field Theory of Dark Matter Direct Detection. *JCAP*, 02:004, 2013, 1203.3542.

[79] K. Schneck et al. Dark matter effective field theory scattering in direct detection experiments. *Phys. Rev.*, D91(9):092004, 2015, 1503.03379.

[80] H. Rogers, D.G. Cerdeno, P. Cushman, F. Livet, and V. Mandic. Multidimensional effective field theory analysis for direct detection of dark matter. *Phys. Rev. D*, 95(8):082003, 2017, 1612.09038.

[81] D.S. Akerib et al. An Effective Field Theory Analysis of the First LUX Dark Matter Search. 3 2020, 2003.11141.

[82] G. Angloher et al. Limits on Dark Matter Effective Field Theory Parameters with CRESST-II. *Eur. Phys. J. C*, 79(1):43, 2019, 1809.03753.

[83] J.D. Lewin and P.F. Smith. Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil. *Astroparticle Physics*, 6(1):87 – 112, 1996.

[84] Rouven Essig, Jeremy Mardon, and Tomer Volansky. Direct Detection of Sub-GeV Dark Matter. *Phys. Rev.*, D85:076007, 2012, 1108.5383.

[85] Constraints on low-mass wimp interactions on 19f from picasso. *Physics Letters B*, 711(2):153 – 161, 2012.

[86] E. Behnke, T. Benjamin, S. J. Brice, D. Broemmelsiek, J. I. Collar, P. S. Cooper, M. Crisler, C. E. Dahl, D. Fustin, J. Hall, C. Harnish, I. Levine, W. H. Lippincott, T. Moan, T. Nania, R. Neilson, E. Ramberg, A. E. Robinson, M. Ruschman, A. Sonnenschein, E. Vázquez-Jáuregui, R. A. Rivera, and L. Uplegger. Direct measurement of the bubble-nucleation energy threshold in a cf$_3$I bubble chamber. *Phys. Rev. D*, 88:021101, Jul 2013.

[87] C. Amole et al. Dark Matter Search Results from the Complete Exposure of the PICO-60 C$_3$F$_8$ Bubble Chamber. *Phys. Rev. D*, 100(2):022001, 2019, 1902.04031.

[88] M. Felizardo, T. A. Girard, T. Morlat, A. C. Fernandes, A. R. Ramos, J. G. Marques, A. Kling, J. Puibasset, M. Auguste, D. Boyer, A. Cavaillou, J. Poupeney, C. Sudre, H. S. Miley, R. F. Payne, F. P. Carvalho, M. I. Prudêncio, A. Gouveia, and R. Marques. Final analysis and results of the phase ii simple dark matter search. *Phys. Rev. Lett.*, 108:201302, May 2012.

[89] Matthew Szydagis et al. A Detailed Look at the First Results from the Large Underground Xenon (LUX) Dark Matter Experiment. In *Proceedings, 10th International Symposium on Cosmology and Particle Astrophysics (CosPA 2013): Honolulu, Hawaii, USA, November 12-15, 2013*, 2014, 1402.3731.

[90] D. Yu. Akimov, H. M. Araújo, E. J. Barnes, V. A. Belov, A. Bewick, A. A. Burenkov, V. Chepel, A. Currie, L. DeViveiros, B. Edwards, C. Ghag, A. Hollingsworth, M. Horn, W. G. Jones, G. E. Kalmus, A. S. Kobyakin, A. G. Kovalenko, V. N. Lebedenko, A. Lindote, M. I. Lopes, R. Lüscher, P. Majewski, A. St. J. Murphy, F. Neves, S. M. Paling, J. Pinto da Cunha, R. Preece, J. J. Quenby, L. Reichhart, P. R. Scovell, C. Silva, V. N. Solovov, N. J. T. Smith, V. N. Stekhanov, T. J. Sumner, C. Thorne, and R. J. Walker. Wimp-nucleon cross-section results from the second science run of zeplin-iii. *Physics Letters B*, 709(1):14–20, 2012.

[91] D.S. Akerib et al. The LUX-ZEPLIN (LZ) Experiment. *Nucl. Instrum. Meth. A*, 953:163047, 2020, 1910.09124.

[92] D.S. Akerib et al. Projected WIMP sensitivity of the LUX-ZEPLIN dark matter experiment. *Phys. Rev. D*, 101(5):052002, 2020, 1802.06039.

[93] E. Aprile et al. The XENON1T Dark Matter Experiment. *Eur. Phys. J. C*, 77(12):881, 2017, 1708.07051.

[94] Xiangyi Cui et al. Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment. *Phys. Rev. Lett.*, 119(18):181302, 2017, 1708.06917.

[95] P. Agnes et al. DarkSide-50 532-day Dark Matter Search with Low-Radioactivity Argon. *Phys. Rev. D*, 98(10):102006, 2018, 1802.07198.

[96] E. Armengaud et al. Performance of the EDELWEISS-III experiment for direct dark matter searches. *JINST*, 12(08):P08010, 2017, 1706.01070.

[97] E. Armengaud et al. Constraints on low-mass WIMPs from the EDELWEISS-III dark matter search. *Journal of Cosmology and Astroparticle Physics*, 2016(05):019–019, may 2016.

[98] G. Angloher, M. Bauer, I. Bavykina, A. Bento, A. Brown, C. Bucci, C. Ciemniak, C. Coppi, G. Deuter, F. von Feilitzsch, D. Hauff, S. Henry, P. Huff, J. Imber, S. Ingleby, C. Isaila, J. Jochum, M. Kiefer, M. Kimmerle, H. Kraus, J. C. Lanfranchi, R. F. Lang, B. Majorovits, M. Malek, R. McGowan, V. B. Mikhailik, E. Pantic, F. Petricca, S. Pfister, W. Potzel, F. Pröbst, W. Rau, S. Roth, K. Rottler,

C. Sailer, K. Schäffner, J. Schmaler, S. Scholl, W. Seidel, L. Stodolsky, A. J. B. Tolhurst, I. Usherov, and W. Westphal. Commissioning run of the cresst-ii dark matter search. *Astroparticle Physics*, 31(4):270–276, 2009.

[99] F. Petricca et al. First results on low-mass dark matter from the CRESST-III experiment. *J. Phys. Conf. Ser.*, 1342(1):012076, 2020, 1711.07692.

[100] R. Agnese et al. Results from the Super Cryogenic Dark Matter Search Experiment at Soudan. *Phys. Rev. Lett.*, 120(6):061802, 2018, 1708.08869.

[101] R. Agnese et al. Search for Low-Mass Dark Matter with CDMSlite Using a Profile Likelihood Fit. *Phys. Rev.*, D99(6):062001, 2019, 1808.09098.

[102] R. Bernabei, P. Belli, F. Cappella, R. Cerulli, F. Montechia, F. Nozzoli, A. Incicchitti, D. Prosperi, C. J. Dai, H. H. Kuang, J. M. Ma, and Z. P. Ye. Dark Matter search. *Nuovo Cimento Rivista Serie*, 26(1):1–74, January 2003, astro-ph/0307403.

[103] R. Bernabei et al. First model independent results from DAMA/LIBRA-phase2. *Nucl. Phys. Atom. Energy*, 19(4):307–325, 2018, 1805.10486.

[104] D.N. McKinsey. Is DAMA Bathing in a Sea of Radioactive Argon? 3 2018, 1803.10110.

[105] Sebastian Baum, Katherine Freese, and Chris Kelso. Dark Matter implications of DAMA/LIBRA-phase2 results. *Phys. Lett. B*, 789:262–269, 2019, 1804.01231.

[106] Govinda Adhikari, others, and The COSINE-100 Collaboration. An experiment to search for dark-matter interactions using sodium iodide detectors. *Nature*, 564(7734):83–86, 2018.

[107] J. Amaré et al. First Results on Dark Matter Annual Modulation from the ANAIS-112 Experiment. *Phys. Rev. Lett.*, 123(3):031301, 2019, 1903.03973.

[108] Emily Shields, Jingke Xu, and Frank Calaprice. Sabre: A new nai(t1) dark matter direct detection experiment. *Physics Procedia*, 61:169 – 178, 2015. 13th International Conference on Topics in Astroparticle and Underground Physics, TAUP 2013.

[109] Y.J. Ko et al. Comparison between DAMA/LIBRA and COSINE-100 in the light of Quenching Factors. *JCAP*, 11:008, 2019, 1907.04963.

[110] M. Antonello et al. The SABRE project and the SABRE Proof-of-Principle. *Eur. Phys. J. C*, 79(4):363, 2019, 1806.09340.

[111] T. Saab and E. Figueroa-Feliciano. Dark matter limit plotter. v5.14.

[112] R. Abusaidi et al. Exclusion limits on the wimp-nucleon cross section from the cryogenic dark matter search. *Phys. Rev. Lett.*, 84:5699–5703, Jun 2000.

[113] Z. Ahmed et al. Search for annual modulation in low-energy CDMS-II data. 2012, 1203.1309.

[114] R. Agnese et al. Silicon detector dark matter results from the final exposure of cdms ii. *Phys. Rev. Lett.*, 111:251301, Dec 2013.

[115] R. Agnese et al. Search for low-mass weakly interacting massive particles with supercdms. *Phys. Rev. Lett.*, 112:241302, Jun 2014.

[116] R. Agnese et al. New results from the search for low-mass weakly interacting massive particles with the cdms low ionization threshold experiment. *Phys. Rev. Lett.*, 116:071301, Feb 2016.

[117] T. Saab and E. Figueroa-Feliciano. Dark matter limit plotter. v2.16.

[118] R. Agnese et al. Low-mass dark matter search with CDMSlite. *Phys. Rev.*, D97(2):022002, 2018, 1707.01632.

[119] R. Agnese et al. Energy Loss Due to Defect Formation from $^{206}$Pb Recoils in Super-CDMS Germanium Detectors. *Appl. Phys. Lett.*, 113(9):092101, 2018, 1805.09942.

[120] R. Agnese et al. Nuclear-Recoil Energy Scale in CDMS II Silicon Dark-Matter Detectors. *Nucl. Instrum. Meth.*, A905:71–81, 2018, 1803.02903.

[121] R. Agnese et al. Production Rate Measurement of Tritium and Other Cosmogenic Isotopes in Germanium with CDMSlite. *Astropart. Phys.*, 104:1–12, 2019, 1806.07043.

[122] T. Aralis et al. Constraints on dark photons and axion-like particles from Super-CDMS Soudan. 2019, 1911.11905.

[123] Cryoconcept. 4, avenue des Andes, 91 952 COURTABOEUF, France.

[124] J. Billard et al. Coherent Neutrino Scattering with Low Temperature Bolometers at Chooz Reactor Complex. *J. Phys. G*, 44(10):105101, 2017, 1612.09035.

[125] U. Patel, Ivan V. Pechenezhskiy, B. L. T. Plourde, M. G. Vavilov, and R. Mc-Dermott. Phonon-mediated quasiparticle poisoning of superconducting microwave resonators. *Phys. Rev. B*, 96:220501, Dec 2017.

[126] L. Cardani et al. Reducing the impact of radioactivity on quantum circuits in a deep-underground facility. 2020, 2005.02286.

[127] B. Cornell et al. Particle Detection Using MKID-Based Athermal-Phonon Mediated Detectors. *Journal of Low Temperature Physics*, 176:891–897, 2014.

[128] Oxford Instruments. Woods Tubney, Abingdon, Oxfordshire OX13 5QX, United Kingdom.

[129] S.O.W. Antman, D.A. Landis, and R.H. Pehl. Measurements of the fano factor and the energy per hole-electron pair in germanium. *Nuclear Instruments and Methods*, 40(2):272 – 276, 1966.

[130] R.H. Pehl, F.S. Goulding, D.A. Landis, and M. Lenzlinger. Accurate determination of the ionization energy in semiconductor detectors. *Nuclear Instruments and Methods*, 59(1):45 – 55, 1968.

[131] U. Fano. Ionization yield of radiations. ii. the fluctuations of the number of ions. *Phys. Rev.*, 72:26–29, Jul 1947.

[132] B.G. Lowe. Measurements of fano factors in silicon and germanium in the low-energy x-ray region. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 399(2):354 – 364, 1997.

[133] Karthik Ramanathan and Noah Kurinsky. Ionization Yield in Silicon for eV-Scale Electron-Recoil Processes. 4 2020, 2004.10709.

[134] D. Barker, W.-Z. Wei, D.-M. Mei, and C. Zhang. Ionization efficiency study for low energy nuclear recoils in germanium. *Astroparticle Physics*, 48:8 – 15, 2013.

[135] Zhong He. Review of the Shockley–Ramo theorem and its application in semiconductor gamma-ray detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 463(1):250 – 267, 2001.

[136] B. Cabrera, M. Pyle, R. Moffatt, K. Sundqvist, and B. Sadoulet. Oblique propagation of electrons in crystals of germanium and silicon at sub-Kelvin temperature in low electric fields. *arXiv e-prints*, page arXiv:1004.1233, Apr 2010, 1004.1233.

[137] S. Golwala. *Exclusion Limits on WIMP-Nucleon Elastic Scattering Cross-Section from the Cryogenic Dark Matter Search.* PhD thesis, University of California, Berkeley, 2000.

[138] D.S. Akerib et al. Design and performance of a modular low-radioactivity readout system for cryogenic detectors in the CDMS experiment. *Nuclear Instruments and Methods A*, 591:476 – 489, 2008.

[139] K. Sundqvist. *Carrier Transport and Related Effects in Detectors of the Cryogenic Dark Matter Search.* PhD thesis, University of California, Berkeley, 2012.

[140] M. Pyle. *Optimizing the Design and Analysis of Cryogenic Semiconductor Dark Matter Detectors for Maximum Sensitivity.* PhD thesis, Stanford University, 2012.

[141] Kevin McCarthy. *Detector Simulation and WIMP Search Analysisfor the Cryogenic Dark Matter Search Experiment.* PhD thesis, Massachusetts Institute of Technology, 2013.

[142] R. Clarke. *An Athermal Phonon Mediated Dark Matter Detector With Surface Event Discrimination.* PhD thesis, Stanford University, 2003.

[143] Chenming Hu. *Modern semiconductor devices for integrated circuits.* Prentice Hall, Upper Saddle River, N.J., 2010.

[144] Neil W. Ashcroft and N. David Mermin. *Solid state physics*. Holt, Rinehart and Winston, New York, 1976.

[145] G. A. Northrop and J. P. Wolfe. Ballistic phonon imaging in germanium. *Phys. Rev. B*, 22:6196–6212, Dec 1980.

[146] Silicon (si), sound velocities: Datasheet from landolt-börnstein - group iii condensed matter. volume 41a1$\beta$: "group iv elements, iv-iv and iii-v compounds. part b - electronic, transport, optical and other properties".

[147] G. A. Northrop and J. P. Wolfe. *Phonon Imaging: Theory and Applications*, pages 165–242. Springer US, Boston, MA, 1985.

[148] S. Tamura and H. J. Maris. Spontaneous decay of ta phonons. *Phys. Rev. B*, 31:2595–2598, Feb 1985.

[149] Shin-ichiro Tamura. Isotope scattering of dispersive phonons in ge. *Phys. Rev. B*, 27:858–866, Jan 1983.

[150] S. Tamura, J. A. Shields, and J. P. Wolfe. Lattice dynamics and elastic phonon scattering in silicon. *Phys. Rev. B*, 44:3001–3011, Aug 1991.

[151] G. Wang. Phonon emission in germanium and silicon by electrons and holes in applied electric field at low temperature. *Journal of Applied Physics*, 107(9):094504, 2010, https://doi.org/10.1063/1.3354095.

[152] B. S. Neganov and V. N. Trofimov. Colorimetric method measuring ionizing radiation. *Otkryt. Izobret.*, 146:215, 1985.

[153] P. N. Luke. Voltage-assisted calorimetric ionization detector. *Journal of Applied Physics*, 64(12):6858–6860, 1988, https://doi.org/10.1063/1.341976.

[154] K.D. Irwin and G.C. Hilton. *Transition-Edge Sensors*, pages 63–150. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[155] T. A. Shutt. *A Dark Matter Detector Based on the Simultaneous Measurement of Phonons and Ionization at 20 mK*. PhD thesis, University of California, Berkeley, 1993.

[156] V. Mandic. *First Results from the Cryogenic Dark Matter Search Experiment at the Deep Site.* PhD thesis, University of California, Berkeley, 2004.

[157] P.L. Brink, B. Cabrera, J.P. Castle, J. Cooley, L. Novak, R.W. Ogburn, M. Pyle, J. Ruderman, A. Tomada, B.A. Young, J. Filippini, P. Meunier, N. Mirabolfathi, B. Sadoulet, D.N. Seitz, B. Serfass, K.M. Sundqvist, D.S. Akerib, C.N. Bailey, M.R. Dragowsky, D.R. Grant, R. Hennings-Yeomans, and R.W. Schnee. First test runs of a dark-matter detector with interleaved ionization electrodes and phonon sensors for surface-event rejection. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 559(2):414 – 416, 2006. Proceedings of the 11th International Workshop on Low Temperature Detectors.

[158] A. Broniatowski et al. A new high-background-rejection dark matter ge cryogenic detector. *Physics Letters B*, 681(4):305 – 309, 2009.

[159] R. Agnese et al. Demonstration of surface electron rejection with interleaved germanium detectors for dark matter searches. *Applied Physics Letters*, 103(16):164105, 2013, https://doi.org/10.1063/1.4826093.

[160] Z. Ahmed et al. Search for inelastic dark matter with the cdms ii experiment. *Phys. Rev. D*, 83:112002, Jun 2011.

[161] R. Agnese et al. Search for Low-Mass Weakly Interacting Massive Particles Using Voltage-Assisted Calorimetric Ionization Detection in the SuperCDMS Experiment. *Phys. Rev. Lett.*, 112(4):041302, 2014, 1309.3259.

[162] R. K. Romani et al. Thermal detection of single e-h pairs in a biased silicon crystal detector. *Appl. Phys. Lett.*, 112(4):043501, 2018, 1710.09335.

[163] N. Mast et al. Performance of the First 150 mm Diameter Cryogenic Silicon Ionization Detectors with Contact-Free Electrodes. *Nucl. Instrum. Meth.*, A904:15–22, 2018, 1804.10634.

[164] S. Hansen, F. DeJongh, J. Hall, B. A. Hines, M. E. Huber, T. Kiper, V. Mandic, W. Rau, T. Saab, D. Seitz, and K. Sundqvist. The cryogenic dark matter search

test stand warm electronics card. In *IEEE Nuclear Science Symposuim Medical Imaging Conference*, pages 1392–1395, Oct 2010.

[165] Maximum integrated data acquisition system [computer software]. `https://midas.triumf.ca/MidasWiki/index.php`, 2016-2017.

[166] Mark A. Lindeman, Simon Bandler, Regis P. Brekosky, James A. Chervenak, Enectali Figueroa-Feliciano, Fred M. Finkbeiner, Mary J. Li, and Caroline A. Kilbourne. Impedance measurements and modeling of a transition-edge-sensor calorimeter. *Review of Scientific Instruments*, 75(5):1283–1289, 2004, https://doi.org/10.1063/1.1711144.

[167] Mark Anton Lindeman. *Microcalorimetry and the transition-edge sensor*. PhD thesis, University of California, Davis, October 2000.

[168] K. D. Irwin, S. W. Nam, B. Cabrera, B. Chugg, G. S. Park, R. P. Welty, and J. M. Martinis. A self-biasing cryogenic particle detector utilizing electrothermal feedback and a squid readout. *IEEE Transactions on Applied Superconductivity*, 5(2):2690–2693, June 1995.

[169] Robin Cantor. *Dc Squids: Design, Optimization and Practical Applications*, pages 179–233. Springer Netherlands, Dordrecht, 1996.

[170] S.W Nam, B Cabrera, B Chugg, R.M Clarke, C Fertig, K.D Irwin, and B.A Young. Squid based w-al quasiparticle trapping assisted transition edge sensor. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 370(1):187 – 189, 1996. Proceedings of the Sixth International Workshop on Low Temperature Detectors.

[171] N. Kurinsky. *The Low-Mass Limit: Dark Matter Detectors with eV-Scale Energy Resolution*. PhD thesis, Stanford University, 2018.

[172] B. A. Hines, K. M. Sundqvist, D. N. Seitz, and M. E. Huber. Flux-coupled direct feedback in a squid amplifier. *IEEE Transactions on Applied Superconductivity*, 21(3):262–266, 2011.

[173] D. N. Green and S. C. Bass. Signal representation with triangular basis functions. *IEE Journal on Electronic Circuits and Systems*, 3(2):58–68, 1979.

[174] S. M. Anton et al. Magnetic flux noise in dc squids: Temperature and geometry dependence. *Phys. Rev. Lett.*, 110:147002, Apr 2013.

[175] J. B. Johnson. Thermal agitation of electricity in conductors. *Phys. Rev.*, 32:97–109, Jul 1928.

[176] H. Nyquist. Thermal agitation of electric charge in conductors. *Phys. Rev.*, 32:110–113, Jul 1928.

[177] Richard F. Voss and John Clarke. Flicker ($\frac{1}{f}$) noise: Equilibrium temperature and resistance fluctuations. *Phys. Rev. B*, 13:556–573, Jan 1976.

[178] W. S. Boyle and K. F. Rodgers. Performance characteristics of a new low-temperature bolometer. *J. Opt. Soc. Am.*, 49(1):66–69, Jan 1959.

[179] John C. Mather. Bolometer noise: nonequilibrium theory. *Appl. Opt.*, 21(6):1125–1129, Mar 1982.

[180] Dan McCammon. Thermal equilibrium calorimeters: An Introduction. pages 1–34, 3 2005, physics/0503045.

[181] C.W. Fink et al. Characterizing TES Power Noise for Future Single Optical-Phonon and Infrared-Photon Detectors. 4 2020, 2004.10257.

[182] D. S. Akerib et al. Design and performance of a modular low-radioactivity readout system for cryogenic detectors in the CDMS experiment. *Nucl. Instrum. Meth.*, A591:476–489, 2008.

[183] Walter Schottky. On spontaneous current fluctuations in various electrical conductors. *Journal of Micro/Nanolithography, MEMS, and MOEMS*, 17(4):1 – 11, 2018.

[184] R. B. Thakur. *The Cryogenic Dark Matter Search Low Ionization-Threshold Experiment.* PhD thesis, University of Illinois at Urbana-Champaign, 2015.

[185] Nader Mirabolfathi, H. Rusty Harris, Rupak Mahapatra, Kyle Sundqvist, Andrew Jastram, Bruno Serfass, Dana Faiez, and Bernard Sadoulet. Toward Single Electron Resolution Phonon Mediated Ionization Detectors. *Nucl. Instrum. Meth.*, A855:88–91, 2017, 1510.00999.

[186] S. Marnieros et al. Controlling the leakage-current of low temperature germanium detectors using xef$_2$dry etching. *Journal of Low Temperature Physics*, 176(3):182–187, 2014.

[187] F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980.

[188] J.H. Hubbell and S.M. Seltzer. Tables of x-ray mass attenuation coefficients and mass energy-absorption coefficients (version 1.4). `http://physics.nist.gov/xaamdi`, 2004.

[189] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003.

[190] N. Mast, M. Fritts, D. Sincavage, P. Cushman, and V. Mandic. An In-Situ Movable Calibration Source for Cryogenic Particle Detectors. *Nucl. Instrum. Meth. A*, 971:164070, 2020, 2002.00893.

[191] F.S. Porter et al. A stepper motor for use at temperatures down to 20 mK. *Physica B: Condensed Matter*, 194-196:151 – 152, 1994.

[192] J.P. Bugeat, R. Petit, and D. Valentian. Electric motor with superconducting winding. *Cryogenics*, 27(2):59 – 62, 1987. Space Cryogenics Workshop.

[193] B. Schmalz. EasyDriver Stepper Motor Driver, v4.5.

[194] Arduino AG.

[195] H. Chagani. First measurements of supercdms snolab 100 mm diameter germanium dark matter detectors with interleaved charge and phonon channels. *Proceedings of Science*, 161, 2014.

[196] Supercon Representative D. Frost. private communication.

[197] Kyle DeBry and Gregory Lafyatis. Note: Electroplating process for connectorizing superconducting nbti cables. *Review of Scientific Instruments*, 89(7):076108, 2018.

[198] P. Brink. *Non-Equilibrium Superconductivity induced by X-ray Photons*. PhD thesis, Magdalen College, Oxford, 1995.

[199] R. A. Moffatt. *Two-Dimensional Spatial Imaging of Charge Transpor in Germanium Crystals at Cryogenic Temperatures*. PhD thesis, Stanford University, 2016.

[200] J. F. Hoburg. Principles of quasistatic magnetic shielding with cylindrical and spherical shields. *IEEE Transactions on Electromagnetic Compatibility*, 37(4):574–579, Nov 1995.

[201] S. A. Hertel and M. Pyle. Phonon Pulse Shape Discrimination in SuperCDMS Soudan. *J. Low. Temp. Phys.*, 167:1173–1178, 2012.

[202] Topsil GlobalWafers A/S. Siliciumvej 1, DK-3600 Frederikssund, Denmark.

[203] J. Bethin, T.G. Castner, and N.K. Lee. Polarizabilities of shallow donors in silicon. *Solid State Communications*, 14(12):1321 – 1324, 1974.

[204] A. Phipps. *Ionization Collection in Detectors of the Cryogenic Dark Matter Search*. PhD thesis, U.C. Berkeley, 2016.

[205] W. Shockley. Currents to conductors induced by a moving point charge. *Journal of Applied Physics*, 9(10):635–636, 1938, https://doi.org/10.1063/1.1710367.

[206] S. Ramo. Currents induced by electron motion. *Proceedings of the IRE*, 27(9):584–585, Sept 1939.

[207] Louis-André Hamel and Manuel Julien. Generalized demonstration of ramo's theorem with space charge and polarization effects. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 597(2):207 – 211, 2008.

[208] M. J. Penn, B. L. Dougherty, B. Cabrera, R. M. Clarke, and B. A. Young. Charge collection and trapping in low-temperature silicon detectors. *Journal of Applied Physics*, 79(11):8179–8186, 1996, https://doi.org/10.1063/1.362552.

[209] Hitoshi Nishiyama and Mitsunobu Nakamura. Capacitance of disk capacitors. CHMT-16:360 – 366, 06 1993.

[210] Nguyen Van Lien and B.I. Shklovskii. Hopping conduction in strong electric fields and directed percolation. *Solid State Communications*, 38(2):99 – 102, 1981.

[211] H. Chagani et al. Ionization Measurements of SuperCDMS SNOLAB 100 mm Diameter Germanium Crystals. *J. Low. Temp. Phys.*, 167(5-6):1125–1130, 2012.

[212] Crc handbook of chemistry and physics., 1978.

[213] S. M. Sze. *Physics of semiconductor devices.* John Wiley & Sons, New York, 2nd ed.. edition, 1981.

[214] A. Kennedy. *SuperCDMS Prototype Detector Design and Testing.* PhD thesis, University of Minnesota, 2017.

[215] A. E. Chavarria et al. Measurement of the ionization produced by sub-kev silicon nuclear recoils in a ccd dark matter detector. *Phys. Rev. D*, 94:082007, Oct 2016.

[216] M. A. Islam, T. J. Kennett, and W. V. Prestwich. Thermal neutron capture in silicon. *Phys. Rev. C*, 41:1272–1275, Mar 1990.

[217] R. Agnese et al. Projected sensitivity of the supercdms snolab experiment. *Phys. Rev. D*, 95:082002, Apr 2017.

[218] M. J. Oreglia. *A Study of the Reactions psi prime –¿ gamma gamma psi.* PhD thesis, Stanford University, 1980.

[219] Leona Stewart. Neutron spectrum and absolute yield of a plutonium-beryllium source. *Phys. Rev.*, 98:740–743, May 1955.

[220] J. Scherzinger, R. Al Jebali, J.R.M. Annand, K.G. Fissum, R. Hall-Wilton, S. Koufigar, N. Mauritzson, F. Messi, H. Perrey, and E. Rofors. A comparison

of untagged gamma-ray and tagged-neutron yields from 241ambe and 238pube sources. *Applied Radiation and Isotopes*, 127:98 – 102, 2017.

[221] Miloslava Baginova, Pavol Vojtyla, and Pavel P. Povinec. Investigation of neutron interactions with ge detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 897:22 – 31, 2018.

[222] J. Lindhard, V. Nielsen, M. Scharff, and P.V. Thomsen. Integral equations governing radiation effects. (notes on atomic collisions, iii). *Kgl. Danske Videnskab., Selskab. Mat. Fys. Medd.*, 33(10), 1 1963.

[223] Peter Sorensen. Atomic limits in the search for galactic dark matter. *Phys. Rev. D*, 91:083509, Apr 2015.

[224] D. J. Ficenec, S. P. Ahlen, A. A. Marin, J. A. Musser, and G. Tarlé. Observation of electronic excitation by extremely slow protons with applications to the detection of supermassive charged particles. *Phys. Rev. D*, 36:311–314, Jul 1987.

[225] Brian L. Dougherty. Measurements of ionization produced in silicon crystals by low-energy silicon atoms. *Phys. Rev. A*, 45:2104–2107, Feb 1992.

[226] Alan Owens, G.W. Fraser, and Kieran J. McCarthy. On the experimental determination of the fano factor in si at soft x-ray wavelengths. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 491(3):437 – 443, 2002.

[227] COMSOL Multiphysics®. v. 5.2. www.comsol.com. COMSOL AB, Stockholm, Sweden.

[228] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The MCMC hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306–312, mar 2013.

[229] B. J. Scholz, A. E. Chavarria, J. I. Collar, P. Privitera, and A. E. Robinson. Measurement of the low-energy quenching factor in germanium using an $^{88}$Y/Be photoneutron source. *Phys. Rev. D*, 94:122003, Dec 2016.

[230] Kenneth F. Wallis. The two-piece normal, binormal, or double gaussian distribution: Its origin and rediscoveries. *Statist. Sci.*, 29(1):106–112, 02 2014.

[231] Jonathan Goodman and Jonathan Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computer Science*, 5(1):65–80, 2010.

[232] F. Izraelevitch et al. A Measurement of the Ionization Efficiency of Nuclear Recoils in Silicon. *JINST*, 12(06):P06014, 2017, 1702.00873.

[233] G. Gerbier, E. Lesquoy, J. Rich, M. Spiro, C. Tao, D. Yvon, S. Zylberajch, P. Delbourgo, G. Haouat, C. Humeau, F. Goulding, D. Landis, N. Madden, A. Smith, J. Walton, D. O. Caldwell, B. Magnusson, M. Witherell, B. Sadoulet, and A. Da Silva. Measurement of the ionization of slow silicon nuclei in silicon for the calibration of a silicon dark-matter detector. *Phys. Rev. D*, 42:3211–3214, Nov 1990.

[234] P. Zecher, D. Wang, J. Rapaport, C. J. Martoff, and B. A. Young. Energy deposition of energetic silicon atoms within a silicon lattice. *Phys. Rev. A*, 41:4058–4061, Apr 1990.

[235] R E Tate and A S Coffinberry. Plutonium-beryllium neutron sources their fabrication and their yield. 10 1958.

[236] I. Heertje, N.P. Van Westen, and A.H.W. Aten. Measurement of pu-be neutrons with activation detectors. *Physica*, 30(10):1846 – 1848, 1964.

[237] O. Iwamoto, M. Herman, S. F. Mughabghab, P. Obložinský, and A. Trkov. Neutron cross-section evaluations for 70,72,73,74,76ge. *AIP Conference Proceedings*, 769(1):434–437, 2005.

[238] S. Raman, E. T. Jurney, J. W. Starner, and J. E. Lynn. Thermal-neutron capture by silicon isotopes. *Phys. Rev. C*, 46:972–983, Sep 1992.

[239] A. R. Sattler. Ionization produced by energetic silicon atoms within a silicon lattice. *Phys. Rev.*, 138:A1815–A1821, Jun 1965.

[240] M. Kasztelan, K. Jedrzejczak, and J. Szabelski. Comparison of mc geant4 simulation with the measurements of gamma photons produced by neutrons. *Modern Physics Letters A*, 34(06):1950046, 2019.

[241] R. Ren et al. Design and characterization of gram-scale phonon-mediated cryogenic particle detectors for ev thresholds. *In Prep.*

[242] Z. Hong, R. Ren, N. Kurinsky, E. Figueroa-Feliciano, L. Wills, S. Ganjam, R. Mahapatra, N. Mirabolfathi, B. Nebolsky, H. D. Pinckney, et al. Single electron–hole pair sensitive silicon detector with surface event discrimination. *Nucl. Instrum. Meth. A*, 963:163757, 2020.

[243] F. Ponce et al. Measuring the Impact Ionization and Charge Trapping Probabilities in SuperCDMS HVeV Phonon Sensing Detectors. *Phys. Rev. D*, 101(3):031101, 2020, 1910.02162.

[244] R. Ren et al. *In Prep.*

[245] Ferdinand Grüneis. Extremal properties of an intermittent poisson process generating 1/f noise. *Fluctuation and Noise Letters*, 15(04):1650028, 2016, https://doi.org/10.1142/S0219477516500280.

[246] D. A. Chrissan and A. C. Fraser-Smith. A clustering poisson model for characterizing the interarrival times of sferics. *Radio Science*, 38(4), 2003, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2002RS002693.

[247] Ferdinand Grüneis. 1/f noise, intermittency and clustering poison process. *Fluctuation and Noise Letters*, 01(02):R119–R130, 2001, https://doi.org/10.1142/S0219477501000238.

[248] A. Karmann and M. Schulz. Characterization of individual defects in mosfet's. *Applied Surface Science*, 39(1):500 – 507, 1989.

[249] M.J. Kirton and M.J. Uren. Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/f) noise. *Advances in Physics*, 38(4):367–468, 1989, https://doi.org/10.1080/00018738900101122.

[250] D. Barker. *SuperCDMS Background Models for Low-Mass Dark Matter Searches.* PhD thesis, University of Minnesota, 2018.

[251] Mark David Pepin. *Low-Mass Dark Matter Search Results and Radiogenic Backgrounds for the Cryogenic Dark Matter Search.* PhD thesis, University of Minnesota, 2016.

[252] Y. Chen. *High-Energy Neutron Backgrounds For Underground Dark Matter Experiments.* PhD thesis, Syracuse University, 2016.

[253] M. Penalver Martinez. *Neutron background estimation for direct WIMP searches at SuperCDMS Soudan.* PhD thesis, University of Durham, 2016.

[254] F.T. Avignone, R.L. Brodzinski, J.I. Collar, H.S. Miley, E. Garcia, A. Morales, J. Morales, R. Nuñez-Lagos, J.H. Reeves, C. Saenz, and J.A. Villar. Theoretical and experimental investigation of cosmogenic radioisotope production in germanium. *Nuclear Physics B - Proceedings Supplements*, 28(1):280 – 285, 1992.

[255] Devendra Lal, Edward D. Goldberg, and Minoru Koide. Cosmic-ray-produced silicon-32 in nature. *Science*, 131(3397):332–337, 1960, https://science.sciencemag.org/content/131/3397/332.full.pdf.

[256] N. V. Abrosimov et al. A new generation of 99.999% enriched 28si single crystals for the determination of avogadro's constant. *Metrologia*, 54(4):599–609, jul 2017.

[257] K. J. Dwyer, J. M. Pomeroy, D. S. Simons, K. L. Steffens, and J. W. Lau. Enriching 28si beyond 99.9998 % for semiconductor quantum computing. *Journal of Physics D: Applied Physics*, 47(34):345105, aug 2014.

[258] S. Yellin. Finding an upper limit in the presence of an unknown background. *Phys. Rev. D*, 66:032005, Aug 2002.

[259] S. Yellin. Extending the optimum interval method. 2007, 0709.2701.

[260] F. Ruppin, J. Billard, E. Figueroa-Feliciano, and L. Strigari. Complementarity of dark matter detectors in light of the neutrino background. *Phys. Rev. D*, 90:083510, Oct 2014.

[261] A. H. Abdelhameed et al. First results from the cresst-iii low-mass dark matter program. *Phys. Rev. D*, 100:102002, Nov 2019.

[262] Xiangyi Cui et al. Dark matter results from 54-ton-day exposure of pandax-ii experiment. *Phys. Rev. Lett.*, 119:181302, Oct 2017.

[263] E. Aprile et al. Light Dark Matter Search with Ionization Signals in XENON1T. *Phys. Rev. Lett.*, 123(25):251801, 2019, 1907.11485.

[264] I. Alkhatib et al. Light Dark Matter Search with a High-Resolution Athermal Phonon Detector Operated Above Ground. 7 2020, 2007.14289.

[265] R.Michael Barnett et al. Review of particle physics. Particle Data Group. *Phys. Rev. D*, 54:1–720, 1996.

[266] V. Mukhanov. *Physical Foundations of Cosmology*. Cambridge University Press, Oxford, 2005.

[267] Gerard Jungman, Marc Kamionkowski, and Kim Griest. Supersymmetric dark matter. *Physics Reports*, 267(5):195 – 373, 1996.

[268] Riccardo Catena and Piero Ullio. The local dark matter phase-space density and impact on WIMP direct detection. *JCAP*, 1205:005, 2012, 1111.3556.

[269] Andrzej K. Drukier, Katherine Freese, and David N. Spergel. Detecting cold dark-matter candidates. *Phys. Rev. D*, 33:3495–3508, Jun 1986.

[270] J. Engel. Nuclear form factors for the scattering of weakly interacting massive particles. *Physics Letters B*, 264(1):114 – 119, 1991.

[271] E. Burstein, J. J. Oberly, and J. W. Davisson. Infrared photoconductivity due to neutral impurities in silicon. *Phys. Rev.*, 89:331–332, Jan 1953.

[272] W. Kaiser and H. Y. Fan. Infrared absorption, photoconductivity, and impurity states in germanium. *Phys. Rev.*, 93:977–980, Mar 1954.

[273] Joseph A. Formaggio, E. Figueroa-Feliciano, and A.J. Anderson. Sterile Neutrinos, Coherent Scattering and Oscillometry Measurements with Low-temperature Bolometers. *Phys. Rev. D*, 85:013009, 2012, 1107.3512.

[274] E. Armengaud et al. Searching for low-mass dark matter particles with a massive Ge bolometer operated above-ground. *Phys. Rev.*, D99(8):082003, 2019, 1901.03588.

[275] William Alexander Page. Data acquisition for supercdms snolab. Master's thesis, The University of British Colombia, Vancouver, 8 2015.

[276] E. Figueroa-Feliciano, B. Cabrera, A.J. Miller, S.F. Powell, T. Saab, and A.B.C. Walker. Optimal filter analysis of energy-dependent pulse shapes and its application to tes detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 444(1):453 – 456, 2000.

[277] T. Alharbi. Principal component analysis for pulse-shape discrimination of scintillation radiation detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 806:240 – 243, 2016.

[278] UCB Pyle Group. QETpy. https://github.com/ucbpylegroup/QETpy.

[279] H. PAUL and A. SCHINNER. program MSTAR, version 3 (Nov., 2002; http://www.exphys.jku.at/stopping/).

# Appendix A

# WIMP Parameter Calculations

## A.1 Thermal Freezeout of DM

We wish to calculate the expected relic abundance of a DM candidate. In the early universe, the energy density was so high that all SM particles were in thermal equilibrium. We will assume that the DM was also initially in equilibrium with the SM bath, but the population became frozen-in as the universe cooled and thermal production of the DM became rare. If that DM was stable, it then remained until the present. In what follows we follow the derivations of [266, 42, 17] to calculate the required $m_\chi$ and annihilation cross section for $\chi's$ to be the dark matter. Plank units ($G = c = \hbar = k_B = 1$) are used throughout.

The rate of DM annihilation in the early universe is given by

$$\Gamma = n_\chi \langle \sigma v \rangle \tag{A.1}$$

where $n_\chi$ is the DM number density and $\langle \sigma v \rangle$ is the thermally averaged product of the annihilation cross section and relative velocity. This process will keep the DM in thermal equilibrium with the SM bath as long as it occurs often enough.

Decoupling occurs when the reaction rate is approximately equal to the Hubble expansion rate, $\Gamma \approx H$. One way to understand this is that it is when the mean time between interactions, $1/\Gamma$, has slowed to be about the age of the universe, $1/H$. We require the DM relics to be non-relativistic, as required by large scale structure. In that case, they must decouple when the temperature of the universe is relatively low,

$T < m_\chi$.

When a species is in equilibrium, the phase space density is described by

$$f(p) = \frac{g}{\exp[(E(p) - \mu)/T] \pm 1} \tag{A.2}$$

where the +(-) sign is for fermions (bosons), $g$ is the degeneracy of a single particle state, $E$ is the particle's energy, and $\mu$ is the chemical potential. If we assume that the DM is non-relativistic ($E \approx m + p^2/2m$) and $\mu = 0$, then we can relate the number density of the DM to the temperature as

$$\begin{aligned}
n_\chi &= \int \frac{d^3p}{(2\pi)^3} f(p) \\
&\approx g_\chi \frac{4\pi}{(2\pi)^3} \int_0^\infty \exp\left[-\frac{m_\chi}{T} - \frac{p^2}{2m_\chi T}\right] p^2 dp \tag{A.3} \\
&\approx g_\chi \left(\frac{m_\chi T}{2\pi}\right)^{3/2} e^{-m_\chi/T}
\end{aligned}$$

This will let us track the DM population as the universe cools. Importantly, we see that the population is exponentially sensitive to the freeze-out temperature. So what's the rest of the universe doing at this same time? At these early times the universe is radiation dominated ($\Omega \approx \Omega_R$) and

$$H^2 \approx \frac{8\pi}{3}\rho_R \tag{A.4}$$

where $\rho_R$ is the energy density of when the universe is radiation dominated. Now, we can also relate this energy density to the temperature and the number of particle species which were still relativistic. The energy density of a species in equilibrium is given by

$$\rho = \int \frac{d^3p}{(2\pi)^3} E(p) f(p). \tag{A.5}$$

We can solve this by inserting the expressions from above and substituting the dimensionless parameters, $x = E/T$, $\alpha = m/T$. This allows us to write (again assuming $\mu = 0$),

$$\rho = \frac{gT^4}{2\pi^2} \int_\alpha^\infty \frac{(x^2 - \alpha^2)^{1/2}}{e^x \pm 1} x^2 dx. \tag{A.6}$$

This dimensionless integral is solved in [266] and the results, accounting for particles and antiparticles, give

$$\rho = \begin{cases} \frac{\pi^2}{30} gT^4 & \text{Bosons} \\ \frac{7}{8}\frac{\pi^2}{30} gT^4 & \text{Fermions} \end{cases} \tag{A.7}$$

Now we can figure out the total radiation-dominated energy density by summing the contributions from the various particle species which were relativistic at a given time (temperature). This is commonly written as

$$\rho_R = \frac{\pi^2}{30} g_*(T) T^4 \tag{A.8}$$

with $g_*(T) = N_B + \frac{7}{8} N_f + \frac{7}{8} \left(\frac{4}{11}\right)^{4/3} N_\nu$ being the number of "effective relativistic degrees of freedom" at temperature $T$, i.e. those particle species which have $m \ll T$. Here $N_B$ is the number of baryons, $N_f$ is the number of fermions, and $N_\nu$ is the number of neutrino species which were relativistic. The extra factor of $4/11$ for the neutrinos is a correction for their effective temperature being higher than the rest of the radiation bath. The values for $g_*$ range from 106.75 when all SM particles were in equilibrium to 3.36 today using just the photon (2 polarizations) and three neutrinos (and their antiparticles).

Now we are ready to combine all of the above results. Doing so, we find the conditions at freeze-out to be described by

$$\Gamma \approx H$$
$$\langle \sigma v \rangle g_\chi \left(\frac{m_\chi T_F}{2\pi}\right)^{3/2} e^{-m_\chi/T_F} \approx \left(\frac{8\pi^3}{90} g_{*F} T_F^4\right)^{1/2} \tag{A.9}$$

where the F subscript denotes the value of a quantity at the time of freeze-out. Using the approximation $mv^2 \sim T \Rightarrow \langle \sigma v \rangle \sim \sigma_F (T/m)^{1/2}$ (where $\sigma_F$ is the effective cross-section at freeze-out), one finds

$$\sigma_F m_\chi \approx \frac{8\pi^3}{3\sqrt{10}} \frac{\sqrt{g_{*F}}}{g_\chi} e^{m_\chi/T_F}$$
$$\approx 10^3 \frac{\sqrt{g_{*F}}}{g_\chi} \tag{A.10}$$

where we have used $m_\chi/T_F = 4$ as the non-relativistic requirement in the last line. Now, the density of $\chi's$ today can be calculated as

$$\Omega_\chi = \frac{8\pi}{3H_0^2} m_\chi n_{\chi,0} \tag{A.11}$$

where $n_{\chi,0}$ is their present number density. Because both the number density, $n_\chi$, and entropy density, $s$, of $\chi's$ after decoupling scale as $a^{-3}$, the quantity $n/s$ is a constant.

Using this and the fact that, for a relativistic fluid, $s \propto gT^3$, we can describe the number density evolution as

$$n_{\chi,0} = n_{\chi,F}\frac{s_0}{s_F} = n_{\chi,F}\frac{g_{*S,0}}{g_{*S,F}}\left(\frac{T_0}{T_F}\right)^3 \tag{A.12}$$

where $g_{*S}(T) = N_B + \frac{7}{8}N_f + \frac{7}{8}\frac{4}{11}N_\nu$ is the effective degrees of freedom, now with a slightly different temperature scaling to match the entropy density. This is nearly the same as $g_*$, except at low energies where $g_{*S} = 3.9$.

$$\Omega_\chi = \frac{8\pi}{3H_0^2}m_\chi n_{\chi,F}\frac{g_{*S,0}}{g_{*S,F}}\left(\frac{T_0}{T_F}\right)^3 \tag{A.13}$$

Finally, substituting Eq. A.3 and Eq. A.10 into Eq. A.13 and using $T_0 = T_{CMB} = 2.73$ K as well as the same non-relativistic condition as above, we obtain

$$\Omega_\chi h_{70}^2 \sim \frac{g_{*F}^{1/2}}{g_{*S,F}}\left(\frac{3\times 10^{-37}cm^2}{\sigma_F}\right) \tag{A.14}$$

where $H_0 = h_{70}$ 70 km/s/Mpc. For $T \gtrsim 1$ GeV, $g_{*F} \approx g_{*S,F} \sim 100$ and $\sqrt{g_{*F}}/g_{*S,F} \approx 0.1$. This shows that a DM candidate particle can account for the observed cold DM abundance ($\Omega_C h_{70}^2 = 0.24$ [14]) via the freeze-out mechanism described above. Remarkably, the required annihilation cross section is similar to the electroweak scale ($\sigma_{EW} \sim 10^{-38}$ cm$^{-2}$) [266]. It is important to note that the interaction itself need not involve the weak force, it would just occur at the same scale. This is, however what gives WIMPs their name.

## A.2 Theoretical WIMP Detection Rate

In this calculation, we follow the standard derivations of [83, 267] to calculate the interaction rate of the galactic WIMP halo with a terrestrial detector as well as their expected nuclear recoil spectrum.

The interaction rate per unit mass of target material for a WIMP is naively

$$R = \frac{\rho_\chi}{m_\chi m_N}\sigma\langle v\rangle \tag{A.15}$$

where $\rho_\chi$ is the local DM mass density, $m_\chi$ is the WIMP particle mass, $m_N$ is the target nucleus mass, $\sigma$ is the cross section for DM-nucleus elastic scattering, and $\langle v\rangle$

is the average DM velocity. However, the details of the DM velocity distribution and the energy-dependence of the scattering cross section are also important. The latter is commonly described as

$$\frac{d\sigma}{dq^2} = \frac{\sigma_0}{2\mu^2 v^2} F^2(\vec{q}) \tag{A.16}$$

where $\vec{q}$ is the momentum transfer in a single scatter, $\mu = \frac{m_\chi m_N}{m_\chi + m_N}$ is the DM-nucleus reduced mass, and $F(\vec{q})$ is the nuclear form factor which encapsulates the momentum dependence of the interaction. $\sigma_0$ is the effective cross-section scale, it is the same as the total cross section for the case of $F = 1$. To include these effects, we write the differential cross section as

$$
\begin{aligned}
dR &= \frac{\rho_\chi}{m_\chi m_N} v f(\vec{v}) dv \frac{d\sigma}{dq^2} dq^2 \\
&= \frac{\rho_\chi \sigma_0}{4\mu^2 m_\chi m_N} \frac{f(\vec{v})}{v} F^2(\vec{q}) dv dq^2
\end{aligned}
\tag{A.17}
$$

where $f(\vec{v})$ is the local DM velocity distribution. In a real detector, we generally measure the energy of the recoiling nucleus, $E_R$, after an interaction. We connect momentum transfer to this quantity by noting that, for the simple 2-body scatter, the momentum transfer is

$$q^2 = 2\mu^2 v^2 (1 - \cos\theta) \tag{A.18}$$

where $\theta$ is the scattering angle in the center of momentum frame. Then the energy imparted to the nucleus is

$$E_R = q^2/2m_N = \frac{\mu^2 v^2}{m_N}(1 - \cos\theta). \tag{A.19}$$

Integrating over the velocity distribution, we can then write the rate as a function of recoil energy

$$\frac{dR}{dE_R} = \frac{\rho_\chi \sigma_0}{2m_\chi \mu^2} F^2(E_R) \int_{v_{min}}^{\infty} \frac{f(v)}{v} dv \tag{A.20}$$

where $v_{min} = \sqrt{E_R m_N/2\mu^2}$ is the minimum DM velocity that can impart an energy $E_R$ (i.e. when $\theta = \pi$). The integral is formally out to $v = \infty$, but the distribution generally cuts off at some characteristic galactic escape velocity, $v_{esc}$. The local DM velocity distribution can be somewhat complex, but it is standard to assume a Maxwellian to first order. That distribution is written as

$$f(v) = \frac{4v^2}{v_0^3 \sqrt{\pi}} \exp[-v^2/v_0^2] dv \tag{A.21}$$

where $v_0$ is a characteristic DM velocity and the distribution has been normalized such that the integral over all angles and velocities is unity, $\int f(\vec{v})d^3v = 1$. Inserting this result and performing the integral we obtain,

$$\begin{aligned}
\frac{dR}{dE_R} &= \frac{\rho_\chi \sigma_0}{\sqrt{\pi}v_0 m_\chi \mu^2}F^2(E_R)\exp[-v_{min}^2/v_0^2] \\
&= \frac{\rho_\chi \sigma_0 v_0}{m_\chi m_N}\frac{2}{\sqrt{\pi}}F^2(E_R)\frac{\exp[-E_R/rE_0]}{rE_0}
\end{aligned} \tag{A.22}$$

where $r = 4\mu^2/m_\chi m_N$ is a classical elastic collision factor and $E_0 = m\chi v_0^2/2$ is the characteristic DM kinetic energy. This expression describes the spectrum of recoil energies that would be measured by a terrestrial detector. One important result is that the spectrum will be a featureless exponential in recoil energy. The general shape can be changed by the form factor as well, but the fact remains that the interactions are preferentially at low energy. The total rate that can be seen by a real detector thus depends on the energy threshold, $E_T$, below which depositions of $E_R$ are indistinguishable from noise. For the case of F=1, the total rate integrated above this threshold is given by

$$R(E_T) = R_0 \exp[-E_T/rE_0] \tag{A.23}$$

where $R_0 = \frac{\rho_\chi \sigma_0 v_0}{m_\chi m_N}\frac{2}{\sqrt{\pi}}$ is the total interaction rate. We can rewrite $R_0$ in units of counts/kg/day as

$$R_0 = \frac{361}{m_\chi m_N}\left(\frac{\sigma_0}{1 \text{ pb}}\right)\left(\frac{\rho_\chi}{0.3 \text{ GeV/cm}^3}\right)\left(\frac{v_0}{220 \text{ km/s}}\right) \text{ kg}^{-1} \text{ day}^{-1} \tag{A.24}$$

where $m_\chi$ and $m_N$ are the WIMP and target nucleus masses, respectively, in GeV/c$^2$. We have normalized the equation using some "standard" [268] values for $\rho_\chi$ and $v_0$. This shows that, using a target material with $\mathcal{O}(10)$ nucleons, WIMP masses of $\sim$5 GeV would result in just a few counts/kg/day.

Using a more realistic $v_{esc}$ results in a much more complicated solution, which only differs from this result by 5%. In addition, here we have neglected the Earth's motion around the sun which causes a slight ($\sim 10\%$) annual modulation of this spectrum as the Earth moves either with or against the galactic WIMP wind [269]. To see this small effect, one would first have to have a statistically significant measurement of the total spectrum. However, it would also provide discrimination against many categories of background signals.

Finally, we haven't yet shown the effects of the form factor, $F$. This takes into account things like the mass distribution of the target nucleus for spin independent, or the distribution of nucleon wave functions in the case of (the more complicated) spin dependent coupling. The calculation of form factors is too complex to discuss here, but [270] is a good treatment of nuclear form factors specifically for WIMP interactions.

Including the current best models used by SuperCDMS for the target nuclear form factor and DM velocity distribution results in recoil energy spectra for spin-independent interactions like those shown in Fig. A.1. There we can primarily see the effect of the form factor of different materials, which suppresses the recoil signal at certain energies. The figure on the right also demonstrates how the spectrum changes in Ge for different WIMP masses. These are the types of spectra we are hoping to one day observe in our detectors.
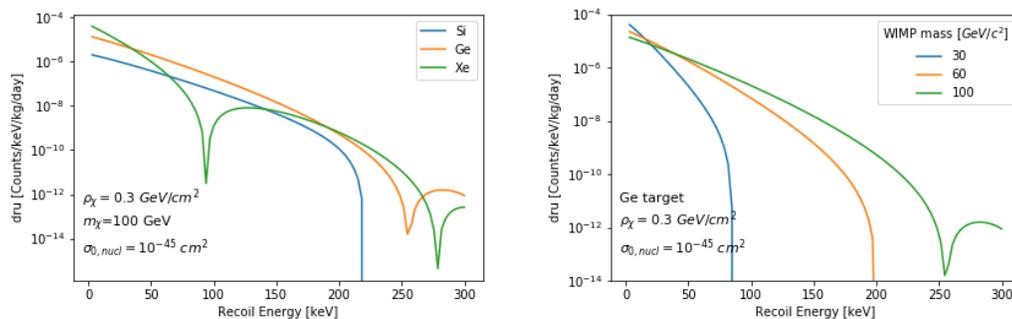


Figure A.1: Left: Differential rate of spin-independent WIMP interactions with common detector materials. Right: Differential rate in Ge at several WIMP masses. $\sigma_{0,nucl}$ is the per-nucleon cross section scale.

# Appendix B

# Charge Signal Model

The charge signal observed at the detector electrodes are governed by the Shockley-Ramo theorem [135]. For a uniform electric field of bias voltage, $V$, across a planar detector of thickness, $L$, the theorem states that the amount of charge measured on the electrode, $dQ$, just depends on the amount of drifting charge in the detector, $q$ and the distance it has moved through the weighting potential, $dx$ as

$$dQ = q\frac{dx}{L}. \tag{B.1}$$

A cartoon of this is shown in Fig. B.1. For a more complex detector and electrode geometry, the weighting potential will be more complicated, but SuperCDMS devices are not so different from this and the simple case will give some good intuition.

## B.1   Basic Model

In our devices both the electrons and holes are drifted, but in opposite directions. We'll consider a planar detector with bias $V(x = 0) = 0$ and $V(x = L) = V$ and an event that produces $N_0$ e/h pairs at $x_0$. If all charges are drifted to the detector edges, then an electrode at the $V = 0$ face will see an induced charge signal of

$$dQ = -eN_e\frac{dx}{L} + eN_h\frac{dx}{L}. \tag{B.2}$$
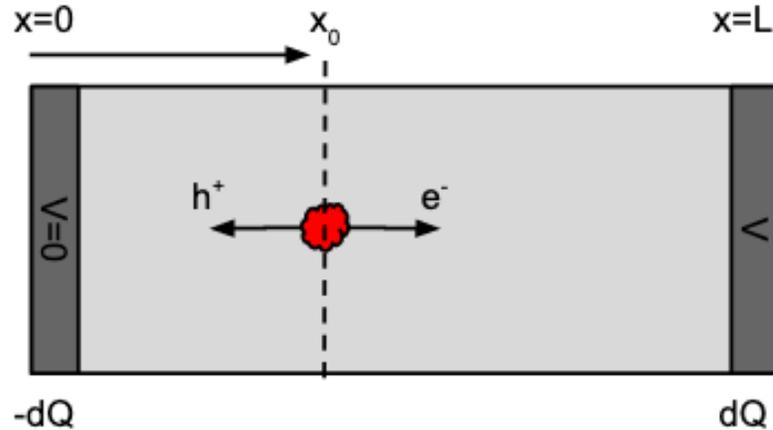
Figure B.1: Cartoon of charge motion in a planar detector.

As the charges drift, the total induced signal will be

$$Q = \int_{x_0}^{L} -eN_0 \frac{dx}{L} + \int_{x_0}^{0} eN_0 \frac{dx}{L} = -eN_0. \tag{B.3}$$

This is nice because, even though the signal may be produced by drifting holes, electrons, or a combination, the measured signal is independent of this and the location in the detector.

## B.2   Trapping

Now, if some carriers only traverse part of the distance to the electrode, they will only contribute that same fraction of their charge to the measured signal. We'll start with a population of $N_0$ holes (electrons) travelling towards $x = 0$ ($x = L$). These will be trapped with a characteristic length scale, $\lambda_{t,h}$ ($\lambda_{t,e}$). These are defined as

$$dN_h = -N_h \frac{dx}{\lambda_{t,h}}$$
$$dN_e = -N_e \frac{dx}{\lambda_{t,e}} \tag{B.4}$$

where $N_h(x)$ and $N_e(x)$ is the population of holes (electrons) as it moves through the detector. Here, we are assuming that the trap density, and thus $\lambda$, is constant throughout the detector volume which should be the case for a well made device. It is also

important to note here that these $\lambda$'s are the effective length the carrier drifts in the $x$ direction before being trapped. The carriers also have significant diffusive motion that needs to be considered to connect this to microscopic physics such as a mean free path for interactions. With this definition, the populations of carriers will fall as they traverse the crystal as

$$N_h(x) = N_0 \exp\left[-\frac{x_0 - x}{\lambda_{t,h}}\right] \tag{B.5}$$

$$N_e(x) = N_0 \exp\left[-\frac{x - x_0}{\lambda_{t,e}}\right] \tag{B.6}$$

The measured charge from this population is then

$$\begin{aligned} Q &= -e \int_{x_0}^{L} N_e(x)\frac{dx}{L} + e \int_{x_0}^{0} N_h(x)\frac{dx}{L} \\ &= eN_0 \frac{\lambda_{t,h}}{L}\left(1 - \exp\left[-\frac{x_0}{\lambda_{t,h}}\right]\right) + eN_0 \frac{\lambda_{t,e}}{L}\left(1 - \exp\left[-\frac{L - x_0}{\lambda_{t,e}}\right]\right) \end{aligned} \tag{B.7}$$

We see now that there will be position dependence in the measured charge values if the trapping effects are significant. In general, this will tend to smear the measured signal down. Note that this reverts to the simple trapping-free case for $\lambda_{t,h} \gg L$ and $\lambda_{t,e} \gg L$.

## B.3 Same Species Impact Ionization

We can easily extend this model to include the effects of same species impact ionization. That is the case where a drifting electron (hole) causes an additional electron (hole) to be released from the crystal. This can occur when carriers are held in shallow traps or the bias voltage is large enough that drifting charges have significant kinetic energy to ionize additional carriers. In principle this can occur when carriers are held in shallow traps or if drifting charges have significant kinetic energy to ionize additional carriers across the band gap. The energy scale of shallow impurities is $\sim 40$ meV in Si [271] and $\lesssim 10$ meV in Ge [272]. This is around the maximum kinetic energy obtained by individual drifting carriers in Si and Ge for the $\lesssim 100$ V/cm electric fields used in SuperCDMS. This energy scale is an order of magnitude lower than the gap minimum for Si (Ge) of 1.12 (0.66) eV. Thus, we will only concern ourselves here with impact ionization from shallow traps which result in a single additional freed carrier as opposed to across the band gap ionization which would produce an e/h pair.

Now, when a carrier is freed from a trap, it begins to drift as well and contributes to the signal starting from the location where it was freed. The trapping center itself is not mobile, and so does not drift or change the signal. We define an impact ionization length scale similar to the way we did the trapping scale but with a different sign,

$$
\begin{aligned}
dN_h &= N_h \frac{dx}{\lambda_{II,hh}} \\
dN_e &= N_e \frac{dx}{\lambda_{II,ee}}
\end{aligned}
\tag{B.8}
$$

where $\lambda_{II,hh}$ corresponds to impact ionization of holes by other holes and $\lambda_{II,ee}$ is for impact ionization of electrons by other electrons.

Because the expressions are similar, the measured signal when both trapping and same species impact ionization are included will take the same form as Eq. B.7 but with the replacement of $\lambda_{t,h}$ with a single characteristic hole length scale

$$
\lambda_h^{-1} = \lambda_{t,h}^{-1} - \lambda_{II,hh}^{-1}
\tag{B.9}
$$

and similarly for electrons

$$
\lambda_e^{-1} = \lambda_{t,e}^{-1} - \lambda_{II,ee}^{-1}.
\tag{B.10}
$$

So we see that the trapping and impact ionization will compete and whichever dominates will determine the signal size.

Fig. B.2 shows some examples of how these length scales affect the measured charge signal. On the left we see that these introduce dependence on the event position, $x_0$. Trapping ($\lambda_x > 0$) generally reduces the measured signal while impact ionization ($\lambda_x < 0$) increases it. The right figure shows the effect of these on the measured energy spectrum for a set of events distributed uniformly in $x_0$. The scales selected for the length scales are generally much larger than are observed in a good detector, but they illustrate that reducing such effects is essential if one is going to make any use of the device's energy spectrum.

## B.4  Opposite Species Impact Ionization

Now we will add in impact ionization of one species by the other. The physics is similar to the above case, with a drifting carrier delivering enough energy to free a trapped
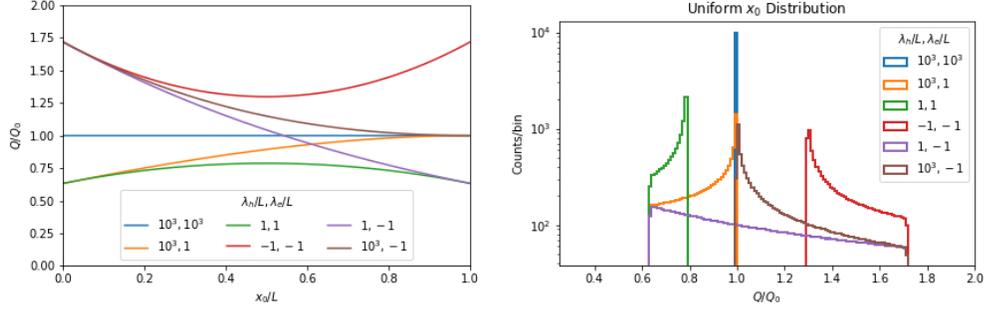
Figure B.2: Left: Measured charge signal as a function of event position for a variety of trapping/impact ionization length scales. Right: Corresponding distributions of measured energies assuming a uniform distribution of event positions.

carrier along the way, but now we are dealing with electrons freeing holes or vice versa. The resulting signal contribution is generated is somewhat different though, because the newly freed carrier will now move in the opposite direction to the type that freed it.

We first consider electrons ionizing holes with a characteristic length scale of $\lambda_{eh}$. Like the other $\lambda's$ above, this is not directly a mean free path, just the drift length scale over which this effect is important. At this point there are multiple generations of carriers to deal with, so they will be identified with roman numerals. $N_e^{\mathrm{I}}$ will denote the first generation of electrons which includes all the effects described above and $N_h^{\mathrm{II}}$ will denote the second generation of holes which are produced by the impact ionization. With this, the number of holes *produced* at location, $x$, is

$$dN_h^{\mathrm{II}}(x) = N_e^{\mathrm{I}}(x)\frac{dx}{\lambda_{he}}. \tag{B.11}$$

These holes will then drift back towards $x = 0$ from where they are produced and may experience trapping and same-carrier impact ionization on the way. From the results of the earlier sections we can propagate each back to a position $x$ with a factor of $\exp[-(x_h^{\mathrm{II}} - x)/\lambda_h]$, where $x_h^{\mathrm{II}}$ is the starting position of the carrier. To calculate the contribution of these charges to the measured signal, we first integrate over their starting

positions to obtain the number present at a given position as

$$
\begin{aligned}
N_h^{\mathrm{II}}(x) &= \int_{\max\{x,x_0\}}^{L} dx_h^{\mathrm{II}} \frac{N_e^{\mathrm{I}}(x_h^{\mathrm{II}})}{\lambda_{he}} \exp\left[-\frac{x_h^{\mathrm{II}} - x}{\lambda_h}\right] \\
&= \frac{N}{\lambda_{he}} \int_{\max\{x,x_0\}}^{L} dx_h^{\mathrm{II}} \exp\left[-\frac{x_h^{\mathrm{II}} - x_0}{\lambda_e}\right] \exp\left[-\frac{x_h^{\mathrm{II}} - x}{\lambda_h}\right] \qquad \text{(B.12)} \\
&= \frac{N\lambda}{\lambda_{he}} \left(\exp\left[\frac{\max\{x,x_0\}}{\lambda}\right] - \exp\left[-\frac{L}{\lambda}\right]\right) \exp\left[\frac{x_0}{\lambda_e} + \frac{x}{\lambda_h}\right]
\end{aligned}
$$

Where the $\max\{x, x_0\}$ term accounts for the fact that the first generation of electrons only traverse from $x_0$ to $L$ and $\lambda^{-1} = \lambda_h^{-1} + \lambda_e^{-1}$ Now, to calculate the resulting signal, we simply use the Ramo theorem as before to find

$$
\begin{aligned}
Q_h^{\mathrm{II}} &= \int_0^L e N_h^{\mathrm{II}}(x) \frac{dx}{L} \\
&= eN \frac{\lambda}{\lambda_{eh} L} \left\{ (\lambda_h + \lambda_e)\left(1 - \exp\left[\frac{x_0 - L}{\lambda_e}\right]\right) - \lambda_h \exp\left[-\frac{x_0}{\lambda_h}\right]\left(1 - \exp\left[\frac{x_0 - L}{\lambda}\right]\right)\right\}
\end{aligned}
$$

$$\text{(B.13)}$$

A plot of this expression for several length scales is shown in Fig. B.3. If all length scales are much larger than the detector width, then this effect does not contribute to the signal. However, once $\lambda_{eh}$ gets closer to $L$, an enhancement of the signal is observed. This will be greater for events at low $x_0$, since, for the scenario under consideration, the electrons will drift further and produce more holes. Also shown is how this effect will contribute to the measured charge signal distribution. Note that this additional energy will be added on top of that measured from the other effects described above, so a value of $Q_h^{\mathrm{II}}/Q_0$ indicates no contribution *from this effect*.

We can find a similar expression for the case where there is $h \to he$ impact ionization. This can be obtained from Eq. B.13 with the substitutions $e \leftrightarrow h$ and $x_0 \leftrightarrow L - x_0$.

This effect is much trickier to work out if we simultaneously allow electrons to ionize holes and vice versa. This is difficult to handle because the ionized charges will drift in the opposite direction from those which caused the ionization and if this is happening from both directions, the ionized carriers will drift through each other and recombination effects may become important. To really understand this then, one needs to perform simulations that take into account carrier drift paths and relative timing.
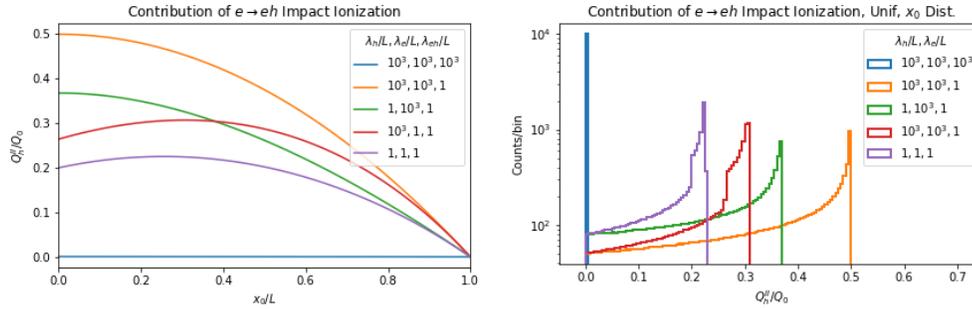
Figure B.3: Left: Contribution of signal due to $e \to eh$ impact ionization as a function of event location for a variety of relevant length scales. Right: Resulting distributions of contribution to measured signal.

## B.5    Relation to NTL Phonon Signal

One final note is that many of the expressions derived above also translate directly to the measured energy from NTL phonons. This is because, for a planar detector, both the measured charge signal and the NTL phonon energy depend linearly on the charge's motion through the bias field,

$$dQ = q\frac{V(x)}{V_0}dx \tag{B.14}$$

and

$$dE_{NTL} = qV(x)dx \tag{B.15}$$

In this way, all of the above expressions can also be used to calculate the NTL phonon energy with the substitution $E_{NTL} = QV_0$. Again, of course, this is only the case for planar devices having a uniform electric field.

# Appendix C

# Electronics Designs

## C.1 HV board

We desired to bias SNOLAB iZIP detectors to voltages higher than what the DCRC was capable of ($\pm 13$ V). To do so, a circuit was designed to interrupt the relevant signal lines between the fridge and DCRCs and re-rout them to an external power supply while passing through normal signal lines to the DCRCs. The design was modeled by Jeter Hall's CDMSlite board and Hassan Changani's HV board for a charge only detector.

I completed the final design and board layout and had the boards fabricated by Custom Circuit Boards. They were populated with components by hand in the lab. A separate board was designed to bias each side of the detector, but only one would be used at a time. For example, the board to bias all of Side 2 was designed to pass the 6 phonon channels and 2 charge channels from Side 1 through to the DCRCs while connecting the remaining lines to HV. To help shield the traces from RF interference, the board has an internal ground plane and is housed in a metal box. Because one entire side of the detector is biased, only 2 RevC DCRCs (4 phonon channels each) were required to read out the 6 channels as opposed to the 3 DCRCs needed when reading out all 12 phonon channels. An image of the baord pattern is shown in Fig. C.1. There are places at the bottom to mount 3 D-Sub 50 connectors which mate with the wiring that goes to the detector. At the top are 2 positions where DCRCs can connect. An HV connector and pads for the bias and test resistors are on the left side. Note that one channel crosses over the three orange ground planes from the bottom center to the

top left. This channel proved to be more susceptible to external noise due to the gap in shielding.
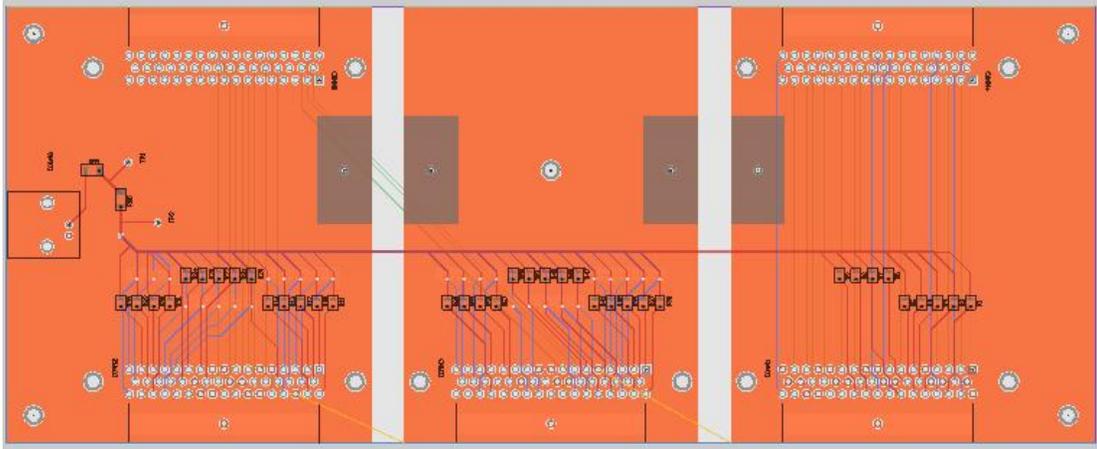


Figure C.1: HV board design.

The HV board is set up so that a 400 MΩ bias resistor, $R_b$, and a 1 MΩ test resistor, $R_t$, are in series with the HV input. We measure voltage, $V_t$, across $R_t$ to calculate the leakage current, and from that calculate the voltage at the crystal, $V_c$, by considering the voltage drop across the resistors. The formula is expected to be

$$V_{det} = V_0 - \frac{R_b}{R_t} V_t = \left( \frac{V_0}{1 \text{ V}} \right) - 400 \left( \frac{V_t}{1 \text{ } \mu\text{V}} \right) \tag{C.1}$$

This assumes $R_b \gg R_t$. Deviation from this approximation is only 0.25%. There are also 40 MΩ resistors on each biased line, but because there were > 50 such lines all in parallel, their effective resistance was even less than $R_t$.

Empirical evidence in Run 54 however suggested the formula was something closer to $V_c = V_0 - 430 \times V_t$ (with $V_0$ in [V] and $V_t$ in [$\mu$V]). One thing we neglected was the finite input impedance, $R_i$, of the multimeter used to measure $V_t$. The Fluke 87V multimeter we used has a 10 MΩ input resistance in voltage mode, which should appear in parallel to $R_t$. Taking $R_i$ into account we get

$$V_{det} = V_0 - \frac{R_i + R_t}{R_i} \frac{R_b}{R_t} V_t = \left( \frac{V_0}{1 \text{ V}} \right) - 440 \left( \frac{V_t}{1 \text{ } \mu\text{V}} \right) \tag{C.2}$$

which is much closer to the empirical formula used in Run 54.

After fabrication, it was discovered that the PCB design contained an error and the FET heater and FET Temperature sense lines were not connected for the charge readout channels. This is was mended by soldering small wires connecting the appropriate D-Sub pins on the underside of the board. Charge readout seemed to work properly thereafter and these fixes did not adversely affect the charge channel noise.

It was also discovered later that, due to a miscommunication, the channel map used in constructing this device was only correct for particular detectors, but was different from the standard channel mapping. The result is that one HV and one readout channel were swapped. A temporary rerouting board was constructed to swap these two channels, effectively undoing the mapping error built into the HV board. This was successfully used in one run, but we eventually just changed the detector wirebonding scheme to match the HV board. This resulted in the following mapping which was used for biasing side 2 of the detector.

| DCRC Connection | Detector Channel |
| --- | --- |
| DCRC1 PA | PE |
| DCRC1 PB | PB |
| DCRC1 PC | PD |
| DCRC1 PD | - |
| DCRC2 PA | - |
| DCRC2 PB | PA |
| DCRC2 PC | PC |
| DCRC2 PD | PF |

Table C.1: Channel map for Side1-readout HV mode

Additional modifications were made over the following years as the board was used to operate a variety of detector types. A few of those are mentioned below.

- While testing USD1, it was found that, with only DIB1 and DCRC1 connected, the HV board-box combo picked up a 60 Hz noise contribution, even when disconnected from the HV source. This was found to be alleviated by connecting the box, which was initially floating, to the HV board ground plane.

- The side 2 HV board was modified to remove biasing from the unused side 2 charge lines. This was done because shorts to chassis in the cold end of these lines were thought to be causing high bias leakage currents. This was accomplished by

removing the QI/QO related HV bias resistors. The side 1 HV board was also modified in this way.

- Modified the side 2 HV board to bridge the ground planes spanned by the crossover channel (PA). Previous bench tests indicated that this may improve the crossover channel noise, which was observed in practice. Note that the ground planes were designed to be separate at the HV board to prevent ground loops, since the star ground of the system was elsewhere.

- This board was also used with monolithic contact-free bias electrodes. So instead of biasing all the phonon channel lines for the unused side, we only used a single QET bias/return line pair to transmit this HV to the electrode via modified cold wiring. This helped reduce leakage occurring in the SQUET cards due to the large voltage difference which could occur between SQUIDs on the same chip when only some were biased to HV.

## C.2  Filter Switchboard

It was found that high frequency noise, likely from electronics on the DCRC itself, can couple to control wires and affect detector electronics performance. In particular, any AC component in the TES bias will heat the TESs more than intended. To investigate this, I developed a pi-filter board which allowed filters to be switched in or out of individual lines.

The board used SMD pi-filters (TUSCONIX model 4700-005) which have insertion loss over 5 dB at 10 MHz and achieves more than 60 dB above 1 GHz. An array of 50 such filters was arranged on a single board, each with a switch to include or remove the filter element as shown in Fig. C.2. The filter elements were located on each of the 50 lines between two D-Sub 50 connectors, which allowed the board to be inserted in series with standard detector wiring circuits.

A set of these boards were used at UMN in R56 to test the effect of filtering various lines for an iZIP detector. A matching set was sent to Queen's University as well, for similar testing. It was found that one effect of using the filters was to shift the $I_b - I_s$ curve to higher $I_b$ values. This is interpreted as follows. The low pass filters remove
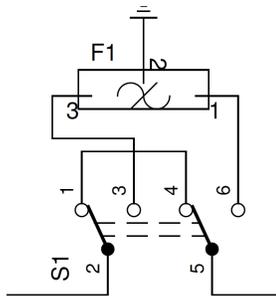
Figure C.2: Single switchable filter circuit element with pi-filter, F1, and DPDT switch, S1.

AC noise components which travel down the bias lines to the TES. Like the applied DC bias, this noise heats the TES, delivering more power to it than calculated via $I_b$. It was found that the amount of AC power depended on the DCRC hardware version as well as the channel number, indicating that the noise pickup may depend on the physical electronics layout. Additional tests showed that some of the charge signal readout lines were surprisingly the most important for reducing this noise component. Results from these tests informed the designs of later DCRC versions which now include on-board filtering for most wiring that enters the fridge.

## C.3    NEXUS Cold Electronics

I also helped produce the final PCB designs for all the SuperCDMS cold electronics at the NEXUS facility. Initial design work was done by collaborators at Northwestern including Z. Hong and B. Neblosky. N. Kurinsky also provided valuable input.

As shown in the wiring diagram in Fig. C.3, the design included boards mounted at the 4 K, 1 K, and MC stages of the fridge. Each board contained some electronics and also acted as a feed-through to pass wiring into the relevant can/stage. All boards, therefore feature ground planes exposed at the edges or corners to mate with the can surface and provide a relatively continuous conductive plane to shield from electromagnetic interference (EMI). One feature shown here is the option to mount the TES shunt resistors at either 1 K, which gives a smaller wiring loop area for reduced vibrational noise coupling but at the cost of higher passive Johnson noise, or at the MC stage, which reduces PJN, but increases the wiring loop area.
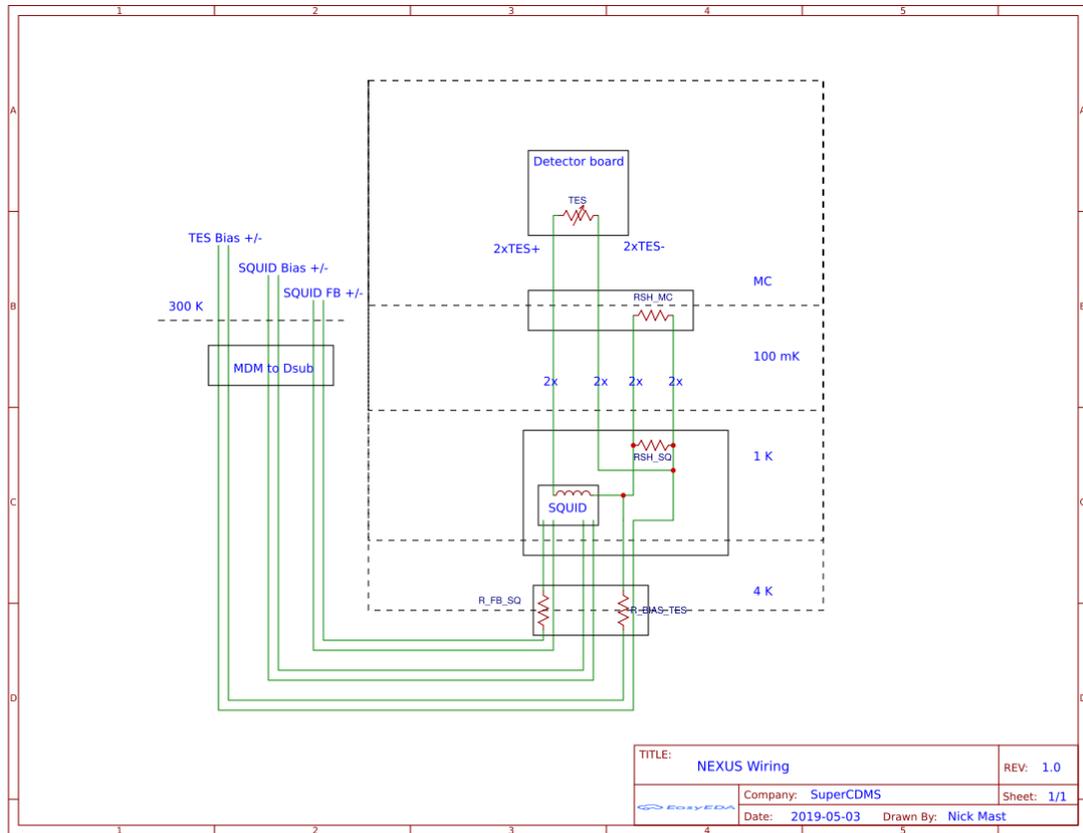
Figure C.3: NEXUS wiring cartoon

## C.3.1    4 K board

After entering the OVC (outer vacuum chamber) at room temperature, the wiring wraps around the 4 K can and interfaces with the 4 K board. As shown in Fig. C.4, this board was designed with pads to mount TES bias resistors, SQUID feedback resistors, and breakout lines for thermometry. These resistors were placed here to reduce Johnson noise, but may be relocated in future designs to counteract voltage-coupled cross-talk occurring between channels in the 300 K to 4 K wiring.

## C.3.2    1 K board

The 1 K board, shown in Fig. C.5, was designed to mate with SQUID boards assembled inside a NIST-designed, shielded Nb box. It also features optional pads for mounting
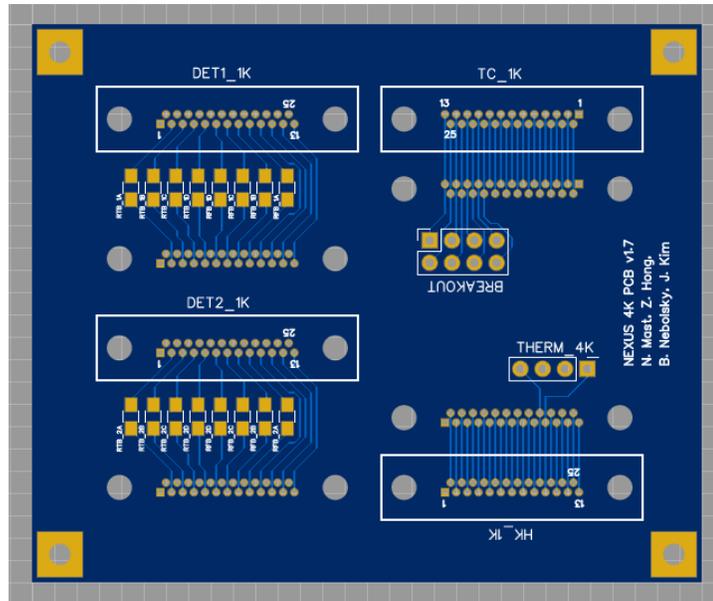
Figure C.4: NEXUS 4K board.

shunt resistors close to the SQUID. Like the 4 K board, there are additional breakout points for thermometry or accessories.

One important design consideration was the minimization of parasitic noise within the TES-SQUID loop. In service of this, the relevant traces on the 1 K board were designed to be extra wide (24 mil vs. 8 mil) and left bare so they could be coated with solder to make superconducting paths. Standard Sn-Pb solders will become superconducting at ~7 K. Additionally, the wiring from 1 K to the MC stage is all doubled (i.e. each connection is made by 2 wires in parallel) to reduce parasitic resistance.

### C.3.3   MC boards

There are two type of boards that connect to the mixing chamber (MC) stage. One type simply feeds through wiring going to the detectors and thermometry. The other provides an alternative mounting location for the TES shunt resistors. This location will have much lower passive Johnson noise than if mounted at 1 K. These boards also provide a breakout point for dedicated 4-wire measurement of $T_C$ test samples. Both are depicted in Fig. C.6.
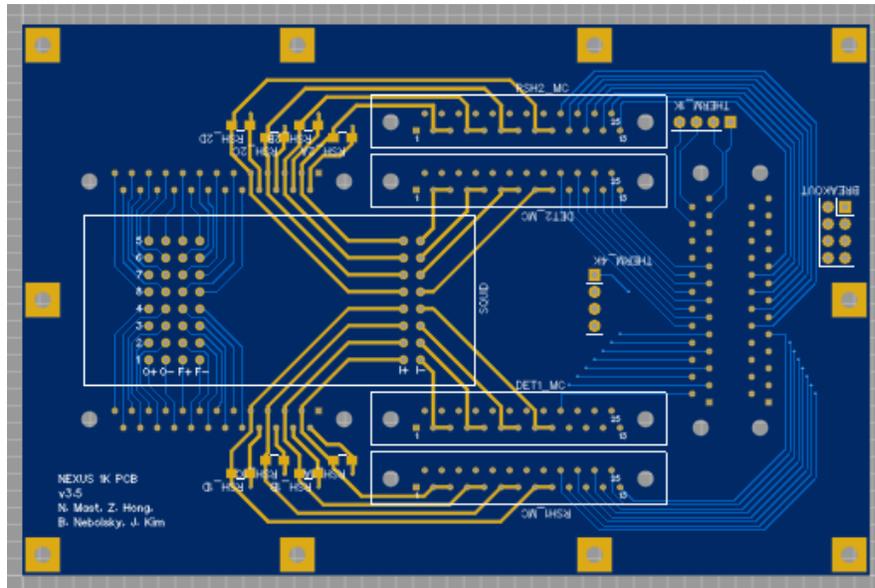
Figure C.5: NEXUS 1K board. Thick, superconducting traces are shown as gold lines.

All of these electronics were tested and installed in NEXUS in the spring of 2019. As of this writing, they have performed quite well in the first six runs of the NEXUS cryostat. They have been used to operate two different generations of HVeV detectors and a prototype TES chip for the Ricochet collaboration [273].
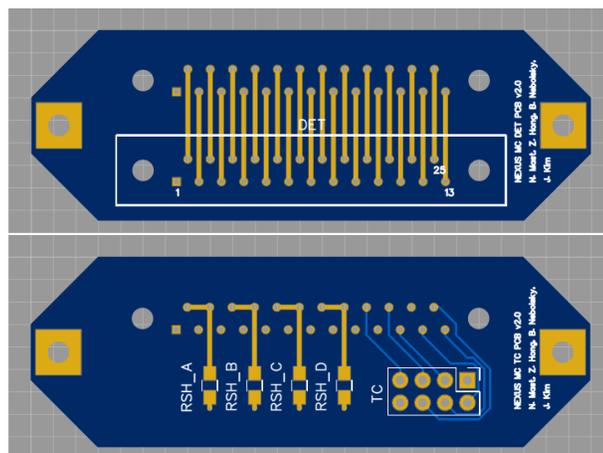
Figure C.6: NEXUS MC boards. Top: simple feed-through board. Bottom: Shunt resistor mounting pads and $T_C$ test breakouts.

# Appendix D

# Toy Phonon Model

This section describes a toy model of phonon signals in the CDMS detectors. It also describes how well the model can reproduce some observed signals from the UMN R71 test of the source mover.

The purpose of this model is not in any way to replace the full *Geant4*-based detector simulation being developed by members of SuperCDMS. That ambitious project will include all of the relevant micro-physics of charge transport, phonon production and propagation, QET and TES behavior, and even the readout circuitry. Instead, the simple model here is only to explore how well a few basic properties and a cartoon-level understanding of phonons in the detector can be used to match observed signals.

The goal of this model is to produce numerical estimates of the total phonon energy absorbed by each channel during an event in an iZIP or HV detector. This can be used to calculate the distribution of e.g. position parameters and phonon collection efficiency (PCE) for events at different locations. The model starts with the following assumptions:

1. Primary phonons are produced in an isotropic sphere centered at the event location and move outward from there.

2. NTL phonons are produced roughly in forward-facing cones along carrier drift path.

3. Recombination phonons are in a sphere where a drifting carrier reaches an electrode.

4. Phonons propagate as straight rays until they encounter a crystal surface.

5. When a phonon strikes an instrumented surface a fraction, $f_1$, are absorbed by the relevant phonon channel.

6. All energy not absorbed on the first pass is assumed to be redistributed through the rest of the crystal volume with a fraction $f_2$ being absorbed by phonon channels.

We will model these several complicated distributions numerically, with a simple Monte-Carlo type simulation. This is necessary because, as shown in Appendix A of [171], analytic expressions of even the simple case of a uniform sphere of phonons require messy expressions involving elliptic integrals and do not lend themselves easily to arbitrary phonon channel geometries.
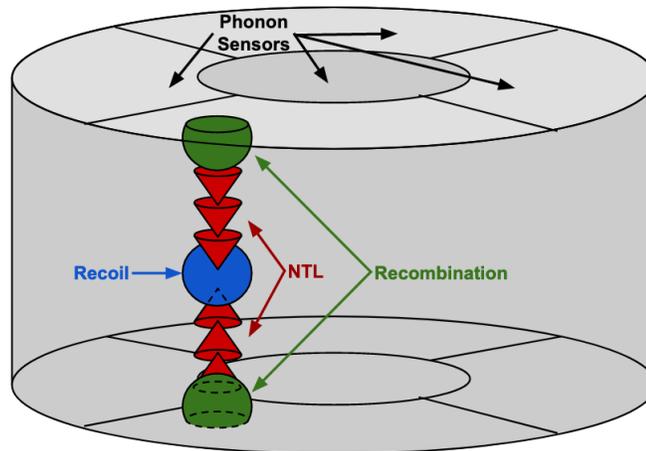


Figure D.1: Cartoon of different phonon populations in the top model. Recoil and re-combination phonons are emitted in spheres from the event and recombination locations. The NTL phonons are thrown along the $z$ positions as the carriers drift.

The primary phonons are generated by producing a distribution of rays emanating from the event location. The direction of each ray is selected randomly from a uniform distribution. The recombination phonons are produced similarly, but near both detector faces at the same $x$ - $y$ position as the event. The $z$ position is an adjustable parameter, but is set to be 0.1 mm from each detector face. This is a clumsy way of accounting for the fact that diffusive phonons produced near the surface should deposit energy in

the nearby channel. If the sphere was centered at the detector surface, none of the rays would intersect the nearby channels, but would only propagate as a half-sphere back into the detector volume. A cartoon of this is shown in Fig. D.1.

The NTL phonons are produced at a number of points spaced along the vertical ($z$ direction) from the event location to each face of the detector. In an HV detector, these points are evenly spaced as the voltage changes linearly and the NTL phonons are produced uniformly along the drift path of the carriers. In an iZIP, the total bias voltage change is different in the bulk and near the surfaces due to the interleaved ground rails. For example, with biases of +/- 4 V on side 1/2, the total bias change from a biased rail on one face to the other is 8 V. However, effect of the ground rails is to effectively compact the first and last 2 V into the near-surface regions as shown in Fig. D.2. Because of this, more of the NTL phonons are produced near the two detector faces. This is effected by producing phonons at points evenly spaced in *voltage*, not position.
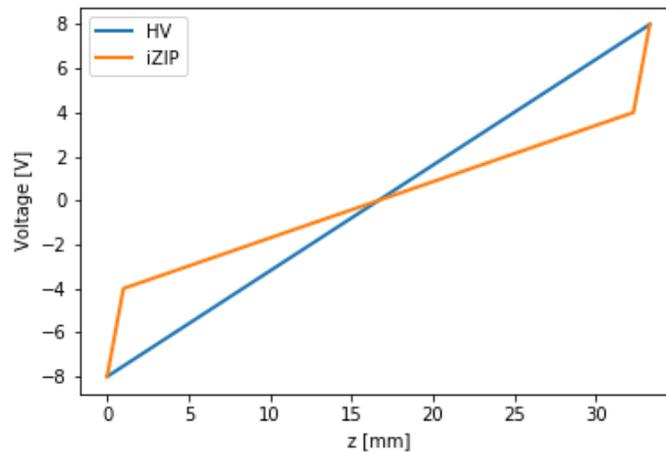


Figure D.2: Voltage as a function of $z$ position in HV and iZIP detectors. The interleaved iZIP fields tend to increase the electric field within $\sim$ 1 mm of the detector faces. This leads to different NTL phonon emission densities.

NTL phonons are produced primarily in the forward-facing direction along the carrier drift path [151]. The detailed shape of this distribution depends on the material and bias field. We model the distribution as Gaussian in the polar angle and uniform

in azimuthal. The polar angle distribution is centered in the forward direction and described by a width parameter, $\sigma_\phi$. It is also only defined from 0 to $\pi$ and normalized over that range. Like the primary and recombination phonons, these are modelled by throwing rays from the phonon production location which are drawn from the relevant angular distribution probability function. Once emitted, the NTL phonons follow the same absorption and scattering rules.

The rays from all phonons either hit a flat crystal face or the outside surface. Those which hit a face are absorbed in the relevant phonon channels with a fixed likelihood of $f_1$. This factor should be related to the QET surface coverage and PCE. In a real detector, after the first few reflections of ballistic phonons, they have pretty much been randomized and lost any discernible position information. We model this by collecting all remaining phonons which were not absorbed on the first pass and redistributing them amongst all the phonon channels. The fraction which are redistributed is $f_2$, which means the PCE is limited by $PCE < f1 + f2(1 - f1)$.

Now, we usually assume that, on average, the phonon energy is distributed evenly among the phonon channels provided they each have the same number of TESs. This is used for such things as relative phonon channel calibration. This is reasonable if one assumes that, late in the phonon event, there is a quasi-uniform bath of ballistic phonons banging around in the detector. Then it is natural to assume the total phonon power each channel observes is directly proportional to the surface area coverage of the crystal. In this case, for a 12 channel device like the SNOLAB detectors, each channel will receive $1/12 = 0.083$ of the remaining phonon energy. However, an alternative calculation would be to produce a uniform distribution of phonons in the crystal, each with a random direction, and find the distribution of energy absorbed in the first hit of those phonons. This leads to a radial weighting factor, with e.g. the inner (outer) channel receiving a fraction of 0.10(0.06) of those phonons. While there is more consistent evidence pointing to the former, uniform model being more correct, we include a parameter in this model to adjust the weighting factor, $w_i$, of energy redistribution to each channel $i$. The parameter, $f_{FA} = 1$, is defined as the ratio of inner (PF) to outer (PA) channel weighting factors, with the other channels' factor linearly interpolated and the sum normalized to 1. For example, a value of $f_{FA} = 1$ corresponds to an even weighting

across all 12 channels, defined by the weighting matrix

$$w = \begin{pmatrix} w_{A1} & w_{B1} & w_{C1} & w_{D1} & w_{E1} & w_{F1} \\ w_{A2} & w_{B2} & w_{C2} & w_{D2} & w_{E2} & w_{F2} \end{pmatrix} = \frac{1}{12} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{D.1}$$

where e.g. $w_{A1}$ is the weighting factor applied to channel PAS1. By contrast, a value of $f_{FA} \approx 1.6$ corresponds to

$$w = \begin{pmatrix} 0.0625 & 0.075 & 0.0875 & 0.0875 & 0.0875 & 0.1 \\ 0.0625 & 0.075 & 0.0875 & 0.0875 & 0.0875 & 0.1 \end{pmatrix} \tag{D.2}$$

which approximates the uniform phonon density scenario. For this example, we have assumed the HV mask, so the channels PC, PD, and PE receive the same weighting factor because they are at the same radius.

The model also included inputs for such things as event energy, ionization yield, detector material (Si or Ge), and the depth of the interleaved iZIP field. There are a great many details that this model ignores or parameterizes over. One potentially important omission is the oblique propagation of electrons in Ge at low bias fields. This may lead to larger discrepancies for such events. All of the steps described above were easy enough to implement that they were simply written in a single Python notebook from which dozens of simulated events could be generated per second on a single computer core.

As a test of this model, we attempted to fit some of the positions dependence observed in UMN R71 as described in Sec. 5.2. The model only calculates total energy absorbed by each channel and says nothing about relative pulse timing. So we used the partition quantities shown in Fig. 5.26. In addition, to constrain the overall scale of the signal, we simultaneously fit to a total PCE of 12%, which was previously measured in this detector. The fit was done as a simple $\chi^2$ minimization. The resulting fit distributions are shown in Fig. D.3. In the left and center plots, the points and error bars are taken from the measured data and the dashed lines are the result of the best fit model. While the agreement is not perfect, the model is undoubtedly able to reproduce the rough features observed in the experiment. The right plot shows the distribution of phonon collection efficiencies for the positions calculated. The dashed black line shows the 12% target value.
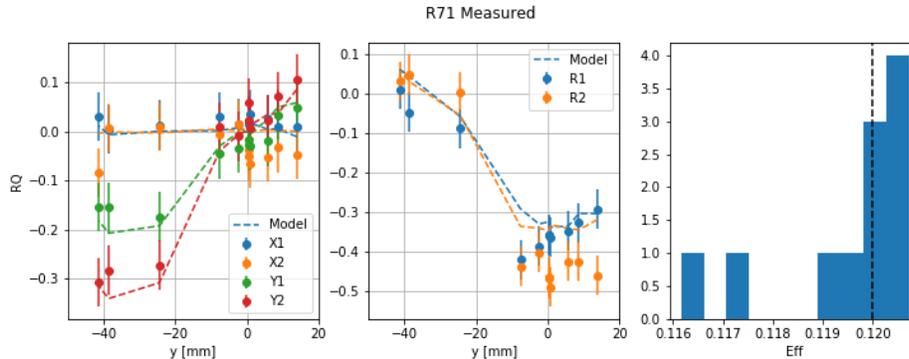
Figure D.3: Model fit to observed position parameters in R71 source scan data.

The best values obtained for the fitting parameters are shown in Table D.1. This analysis used constant values for several other parameters including the iZIP interleaved field depth of 1 mm and recombination depth of 0.1 mm, but the fits were observed to be relatively insensitive to these parameters within reasonable physical bounds. The best fit parameters are also quite reasonable. Given the best fit value of, $\sigma_\phi$, the NTL angular distribution was somewhat flat, with about 68% of NTL phonons being emitted in the forward-going direction. This is reasonable for a detector operated at such relatively low bias field ($\pm 4$ V/33 mm = 2.42 V/cm). Cf. the example calculation in [151] which gives roughly $\sigma_\phi \approx 0.4\pi$ at 3 V/cm in Ge. The absorption fractions $f_1$ and $f_2$ are harder to diagnose directly, but they appear reasonable given the fraction of the detector surfaces covered by QET sensors was $\approx 35\%$ for this device [171] and that we have observed $PCE \approx 12\%$. Finally the redistribution factor $f_{FA}$ is quite near to 1, indicating a relatively uniform sharing of energy after the initial phonon bounce. This is in agreement with measurements of absorbed energy after performing absolute calibrations of individual channels using information from $I_b - I_s$ curves as described in Ch. 3.

This model is a vast oversimplification of the real phonon physics that occurs in a detector. However, it has proven to yield energy values reasonably consistent with the measurement of R71 without too much pain. These results are encouraging and merit further study since they hint that the simple mechanisms described can capture a large amount of the relevant detector behavior. It should be investigated whether this same model can fit other data sets in iZIP and HV detectors. If so, it may be

| Parameter | Value |
|:---------:|:-----:|
| $\sigma_\phi$ | $0.54\pi$ |
| $f_1$ | 0.03 |
| $f_2$ | 0.10 |
| $f_{FA}$ | 0.96 |

Table D.1: Best fit toy phonon model parameters for the R71 source scan data.

used to relatively quickly generate predicted estimates of relevant measured quantities. With more validation, this could be a useful and efficient tool in developing position reconstruction algorithms.

# Appendix E

# Pulse Fitting

The raw data collected by SuperCDMS detectors comes in the form of digitized pulse traces from the various charge and phonon sensors. Because of the high number of events studied, the pulses are analyzed en masse with a set of software developed and maintained by the SuperCDMS collaboration. The software, called *cdmsbats*, is used to run various algorithms and fits on the signals in order to extract quantities relevant for reconstructing event energies, positions, and particle types.

However, as discussed earlier, all of these measurements come with some form or another of irreducible noise, which eventually limits measurements of low energy particle events. These measurement limits translate directly into limits on the range of DM parameter space that the detectors are sensitive to, and thus the scientific impact of the whole experiment. Thus it is of critical importance not only to limit this noise as much as possible when designing the systems, but also to develop techniques to measure the underlying signals as well as possible in the presence of whatever noise remains. This chapter describes some of the algorithms used in SuperCDMS to reconstruct true pulses from measured, noisy data. It also explores their known biases and resolutions, in particular in regimes where the signal to noise level becomes low.

# E.1 Optimal Filter Fitting

## E.1.1 OF and OF0

The most heavily used pulse fitting algorithm in SuperCDMS is known as Optimal Filter (OF) fitting. Broadly, this method uses information about the expected signal and noise to extract an estimate of the pulse amplitude. The details have been covered exhaustively in many SuperCDMS theses over the past 25 years [155, 137, 171]. As such I will only present some of the concepts relevant for this thesis and leave the interested reader in the capable hands of previous authors.

### Standard Formalism

The problem can be stated as follows. A measured signal, $v(t)$, is assumed to consist of the desired signal and noise as

$$v(t) = As(t) + n(t) \tag{E.1}$$

where $s(t)$ is a signal shape, or "template", normalized to an amplitude of 1, $A$ is the amplitude we wish to extract and $n(t)$ is the noise. We are assuming here that the noise is stationary, i.e. its spectrum does not change when a signal is present. In principle the template can be calculated theoretically from TES circuit parameters, but in practice it is constructed as something like an average of many measured pulses. The best fit can be obtained via $\chi^2$ minimization in the frequency domain where there is assumed to be no correlation across frequencies. This is not the case in the time domain, where the signal is strongly correlated at time scales less than the pulse fall time. The $\chi^2$ expression is then

$$\chi^2 = \int \frac{|v(f) - \hat{A}s(f)|^2}{J(f)} df \tag{E.2}$$

where $v(f)$ and $s(f)$ are the Fourier transforms of the measured signal and template respectively while $J(f)$ is the Power Spectral Density (PSD) of the noise. Here, $\hat{A}$ is the estimate of the true signal amplitude, $A$. $J(f)$ is generally calculated from an average of many measured, pulse-less traces. The minimization condition, $d\chi^2/d\hat{A} = 0$ leads to the following expression for the best fit amplitude,

$$\hat{A} = \frac{\int s^*(f)v(f)/J(f)df}{\int |s(f)|^2/J(f)df} \tag{E.3}$$

We can define a filter function as

$$\phi'(f) = \frac{s^*(f)/J(f)}{\int |s(f)|^2/J(f)df}. \tag{E.4}$$

This just depends on the noise and expected pulse shape, so it can be calculated once and reused for fitting many events on the same channel. Then we can reduce this fitting procedure to a product of the measured signal and our filter in frequency space, hence the name "Optimal Filtering". The expression for the best fit amplitude then becomes

$$\hat{A} = \int v(f)\phi'(f)df. \tag{E.5}$$

So there we have a nice, clean expression for the optimal pulse amplitude in the presence of noise, problem solved! However, without stating it explicitly, we assumed above that the signal and template are aligned in time, which is to say that $\hat{A}$ is a real number. This is not always the case so we should also allow for another degree of freedom to time shift the template. The above procedure is still quite useful, but we give it the name "OF0", to denote the fixed offset nature. In general, the template and signal may not be aligned properly due to factors such as variations in triggering time. The OF0 procedure is easily extended to accommodate a time offset, $t_0$ by including a phase with the amplitude term as

$$\chi^2 = \int \frac{|v(f) - \hat{A}e^{-i\omega t_0}s(f)|^2}{J(f)}df \tag{E.6}$$

With a corresponding redefinition of the filter,

$$\phi'(f) = \frac{s^*(f)e^{-i\omega t_0}/J(f)}{\int |s(f)|^2/J(f)df} \tag{E.7}$$

we again obtain Eq. E.5 for the amplitude estimate. Unfortunately, because Eq. E.6 is nonlinear in $t_0$, there is no analytic expression to estimate the time offset. In practice, we instead calculate $\hat{A}$ over a range of $t_0$ and pick that with the lowest $\chi^2$. Since Eq. E.6 can be recast [171] as

$$\chi^2 = \int \frac{|v(f)|^2}{J(f)}df - \hat{A}(t_0)^2 \int \frac{|s(f)|^2}{J(f)}df \tag{E.8}$$

we see that the best time offset is the one which maximizes $\hat{A}^2$. Strictly speaking, to minimize $\chi^2$, one should choose $\max(|\hat{A}|)$. However, if the polarity of the expected

signal is known to be positive, one can include that information by selecting $\max(\hat{A})$. Some of the consequences of this choice are discussed in the following section.

To ensure accurate fits, the template pulses must faithfully represent the average pulse shape. One of my first analysis tasks was to develop improved templates for UMN test facility data. The templates are constructed by selecting a set of characteristic data and calculating the average scaled pulse shape for each channel. Various data quality and pulse selection algorithms were implemented to build the templates. Fig. E.1 demonstrates the improvement in goodness of fit for a set of charge pulses measured on a prototype Ge iZIP in UMN R56. These are compared to fits of the same data using the previous template version, which had been calculated using data from a different device. Not only was the $\chi^2$ shown to be generally lower with the new template, but the improvement was especially good at higher energies. In the months following that work, we undertook a large scale re-processing of UMN detector data. This set of updated templates was an integral part of the processing, giving increased event by event energy resolution.
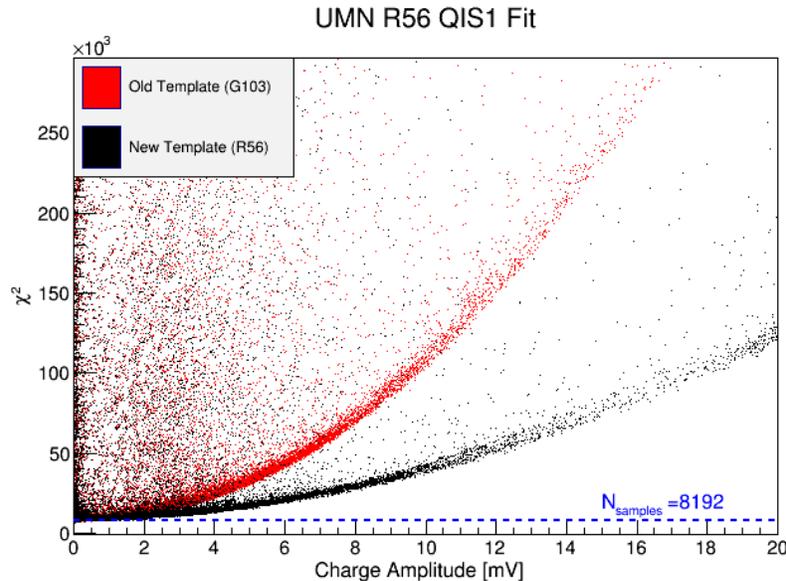


Figure E.1: Improvement in OF fit $\chi^2$ with new charge signal templates.

We can also understand the quadratic shape of these plots as follows. First, we

introduce the discrete version of the $\chi^2$ expression as described in [137].

$$\chi^2 = \sum_{n=-N/2}^{N/2-1} \frac{1}{\Delta f} \frac{\langle |v(f_n) - \hat{A}s(f_n)|^2 \rangle}{J(f_n)} \tag{E.9}$$

with $\Delta f = 1/T$ with total trace length of $T$, $f_n = n\Delta f$, and $N = N_{samples}$ is the number of digitized samples per trace and is generally 4096 (8192) for phonon (charge) traces measured at UMN. Note that, besides superficial differences in frequency-space notation, expressions in [137] use a total trace length of 2T while here we use T. Recall that we are assuming the measurement is composed as a signal shape plus noise as $v(f_n) = As(f_n) + n(f_n)$. Thus, if we have the correct template shape, OF fits should yield a constant $\chi^2$ value of

$$\begin{aligned}
\langle \chi^2 \rangle &= \sum_{n=-N/2}^{N/2-1} \frac{1}{\Delta f} \frac{\langle |n(f_n)|^2 \rangle}{J(f_n)} \\
&= \sum_{n=-N/2}^{N/2-1} \frac{1}{\Delta f T} = N
\end{aligned} \tag{E.10}$$

where we have used the discrete PSD definition $J(f_n) = T\langle |n(f_n)|^2 \rangle$. This is indeed observed to be case at low amplitudes in Fig. E.1. However, there is inevitably some level of mismatch between the true pulse shape and the template used. This will generally lead to quadratic increase in $\chi^2$ as follows. Suppose that the OF template, $\hat{s}(f_n)$ is slightly different from the true pulse shape, $s(f_n)$ as $\hat{s}(f_n) = s(f_n) + \xi(f_n)$. Then we obtain the following

$$\begin{aligned}
\langle \chi^2 \rangle &= \sum_{n=-N/2}^{N/2-1} \frac{1}{\Delta f} \frac{\langle |As(f_n) + n(f_n) - \hat{A}s(f_n) - \hat{A}\xi(f_n)|^2 \rangle}{J(f_n)} \\
&\approx \sum_{n=-N/2}^{N/2-1} \frac{1}{\Delta f} \frac{\langle |n(f_n)|^2 \rangle + \langle |\hat{A}\xi(f_n)|^2 \rangle}{J(f_n)} \\
&\approx N + \hat{A}^2 \sum_{n=-N/2}^{N/2-1} \frac{1}{\Delta f} \frac{\langle |\xi(f_n)|^2 \rangle}{J(f_n)} \\
&\approx N + \hat{A}^2 \times \text{const.}
\end{aligned} \tag{E.11}$$

where we have assumed in the second step that the template is close enough to the true pulse that $\hat{A} \approx A$ and used the fact that noise is uncorrelated with the signal. So we find that there will be a quadratic term associated with the template-signal mismatch. As seen in Fig. E.1, this can be reduced with improved templates. If the signal shape also varies with energy, the $\chi^2$ will grow quadratically at least, if not stronger. We have also glossed over it here, but in a real detector, there *can* be correlation between the signal and noise because the noise spectrum changes as TES moves through its transition. If not accounted for (see Fig. E.1), $\chi^2$ will also have a linear component from terms like $\langle n(f_n) \hat{A} s(f_n) \rangle \neq 0$.

### E.1.2 OF Resolution and Bias

The expected resolution and bias of the Optimal Filter can be calculated analytically. This section extends such calculation found in [137] and Appendix A of [274]. If the OF template and pulse are perfectly aligned and $t_0$ is fixed, one can show that the OF0 amplitude is unbiased, $\langle \hat{A} \rangle = A$, and has a width given by

$$\sigma_{\hat{A}}^2 = \langle \hat{A}^2 \rangle - \langle \hat{A} \rangle^2 = \left[ \int |s(f)|^2 / J(f) df \right]^{-1}. \tag{E.12}$$

The resolution of this algorithm then is set by the degree to which the signal and noise frequency spectra differ. Note, also that this is independent of signal amplitude, $A$. The probability density of measuring $\hat{A}$ given a true amplitude of $A$ is then just

$$P(\hat{A}|A) = \frac{1}{\sqrt{2\pi\sigma_{\hat{A}}^2}} exp\left( -\frac{(\hat{A} - A)^2}{2\sigma_{\hat{A}}^2} \right) \tag{E.13}$$

This can be seen in Fig. E.2 for the case where $\sigma_{\hat{A}} = 0.1$. The density plot shows the probability of measuring $\hat{A}$ given a true value of $A$. The dashed black line has a slope of 1 for comparison. There we see that the estimator, $\hat{A}$, is unbiased, so it follows a straight line, even when the signal is far below the noise level. Additionally, the variance of the estimator is constant everywhere.

Generally, the pulse and template are not perfectly aligned, so we use OF, not OF0. As descried above, this is accomplished by calculating the OF0 amplitude for a range of $t_0$ and then selecting either the largest $|\hat{A}|$ (equivalent to lowest $\chi^2$) or the highest $\hat{A}$
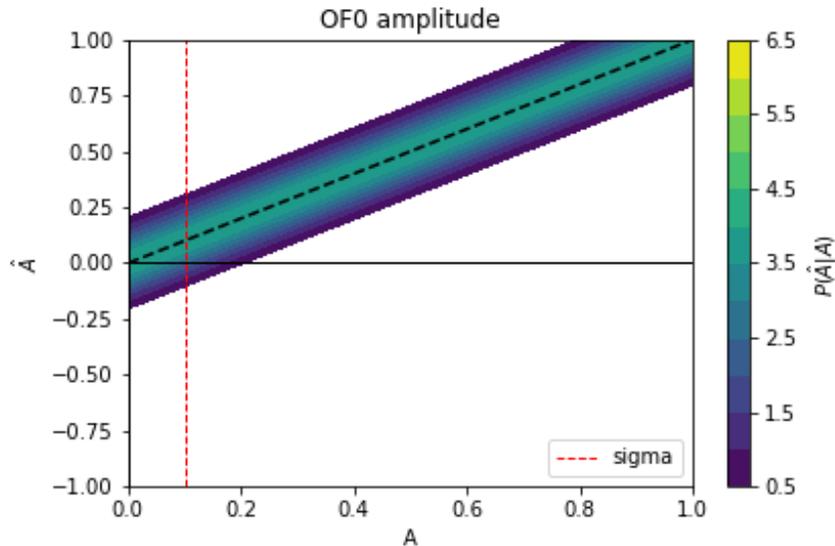
Figure E.2: Amplitude reconstruction probability density of the OF0 algorithm.

(lowest $\chi^2$ for positive-going signal). In either case, this selection biases $\hat{A}$ away from $A$.

We start by considering that we calculate $\hat{A}(t_0)$ for a number, $N$, of $t_0$ values. This is set by defining a time window around the nominal trigger time over which $t_0$ is allowed to vary, with $N$ then being the number of discrete time bins within that window. Assuming the $\hat{A}(t_0)$ are independent, they are each drawn separately from Eq. E.13. In practice, there are correlated over scales characteristic of the signal and noise, $N_{corr}$, and $N \rightarrow N_{eff} \approx N/N_{corr}$, but the same results follow.

Now, assume that we have a signal with amplitude $A$ in just one of the $N$ bins. To find the distribution of $\hat{A}$, we look at the probability distribution for the largest fluctuation from $N$ independent random variables $X_i$. With $P_i(X_i)$ the probability distribution of each $X_i$, and $X^*$ the largest fluctuation, we can write the distribution for $X^*$ when selecting $\max\{|X_i|\}$

$$P(X^*) = \sum_i^N P_i(X^*) \prod_{k \neq i} P(|X_k| < X^*) \tag{E.14}$$

or when selecting $\max\{X_i\}$

$$P(X^*) = \sum_i^N P_i(X^*) \prod_{k \neq i} P(X_k < X^*) \tag{E.15}$$

For to the problem at hand, we see that $P(X)$ is just the Gaussian, Eq. E.13. This means that $P(|X_k| < X^*)$ and $P(X_k < X^*)$ are just integrals of that Gaussian (i.e. erf's) with appropriate limits.

For the case of $\max\{|X_i|\}$, we then have

$$P(\hat{A}_{|max|}|A) = P(\hat{A}|A) \left[ \int_{-|\hat{A}|}^{|\hat{A}|} P(\hat{A}|0)d\hat{A} \right]^{N-1}$$

$$+ (N-1)P(\hat{A}|0) \left[ \int_{-|\hat{A}|}^{|\hat{A}|} P(\hat{A}|0)d\hat{A} \right]^{N-2} \left[ \int_{-|\hat{A}|}^{|\hat{A}|} P(\hat{A}|A)d\hat{A} \right] \tag{E.16}$$

The first term is the probability that the bin which actually has the signal has amplitude $\hat{A}$ while the $(N-1)$ others are smaller. The second term is the probability that one of the $(N-1)$ bins has has amplitude $\hat{A}$ while the others, including the bin with the signal fluctuation, are smaller.

We get a similar expression if we select $\max\{X_i\}$,

$$P(\hat{A}_{max}|A) = P(\hat{A}|A) \left[ \int_{-\infty}^{\hat{A}} P(\hat{A}|0)d\hat{A} \right]^{N-1}$$

$$+ (N-1)P(\hat{A}|0) \left[ \int_{-\infty}^{\hat{A}} P(\hat{A}|0)d\hat{A} \right]^{N-2} \left[ \int_{-\infty}^{\hat{A}} P(\hat{A}|A)d\hat{A} \right] \tag{E.17}$$

Performing the integrals yields

$$P(\hat{A}_{|max|}|A) = \frac{1}{\sqrt{2\pi\sigma_A^2}} exp\left(-\frac{(\hat{A}-A)^2}{2\sigma_A^2}\right) \left[ erf\left(\frac{|\hat{A}|}{\sqrt{2}\sigma}\right) \right]^{N-1} +$$

$$\frac{N-1}{\sqrt{2\pi\sigma_A^2}} exp\left(-\frac{\hat{A}^2}{2\sigma_A^2}\right) \left[ erf\left(\frac{|\hat{A}|}{\sqrt{2}\sigma}\right) \right]^{N-2} \frac{1}{2} \left[ erf\left(\frac{|\hat{A}|-A}{\sqrt{2}\sigma}\right) - erf\left(\frac{-|\hat{A}|-A}{\sqrt{2}\sigma}\right) \right]$$

$$\tag{E.18}$$

for max $\{|X_i|\}$. If instead, we are selecting max $\{X_i\}$, the distribution is

$$P(\hat{A}_{max}|A) = \frac{1}{\sqrt{2\pi\sigma_A^2}}exp\left(-\frac{(\hat{A} - A)^2}{2\sigma_A^2}\right)\left[\frac{1}{2} + \frac{1}{2}erf\left(\frac{\hat{A}}{\sqrt{2}\sigma}\right)\right]^{N-1} +$$

$$\frac{N-1}{\sqrt{2\pi\sigma_A^2}}exp\left(-\frac{\hat{A}^2}{2\sigma_A^2}\right)\left[\frac{1}{2} + \frac{1}{2}erf\left(\frac{\hat{A}}{\sqrt{2}\sigma}\right)\right]^{N-2}\left[\frac{1}{2} + \frac{1}{2}erf\left(\frac{\hat{A} - A}{\sqrt{2}\sigma}\right)\right] \quad \text{(E.19)}$$

In both cases, the following identities are useful when dealing with Gaussian integrals. Let $G(x)$ be the normalized Gaussian, then one can show

$$\int_{-\infty}^{0} G(x)dx = \int_{0}^{\infty} G(x)dx = \frac{1}{2}$$

$$\int_{-y}^{0} G(x)dx = \int_{0}^{y} G(x)dx \quad \text{(E.20)}$$

$$\int_{-\infty}^{y} G(x)dx = \frac{1}{2} + \int_{0}^{y} G(x)dx = \frac{1}{2}\left(1 + \int_{-y}^{y} G(x)dx\right)$$

Fig. E.3 shows how these OF probability distributions appear. Again, we are using an arbitrary values of $\sigma_{\hat{A}} = 0.1$ and $N = 5$ for illustrative purposes. For large amplitudes, the estimator is still unbiased, but below $\sim 2\sigma_A$ it becomes biased away from $\hat{A} = 0$. This is because it has the freedom to find the largest noise excursion from 0 within the available window. The difference between selecting either the max $\{|X_i|\}$ or max $\{X_i\}$ close to $A = 0$ is also apparent here. The figure on the left is the case of max $\{|X_i|\}$, where at $A = 0$ the algorithm selects the largest deviation of either polarity and the distribution becomes bimodal. On the right is the case of max $\{X_i\}$, so the largest positive going excursion is selected. Both algorithms suffer from a bias away from the true pulse amplitude as the signal falls below the noise level.

A larger window width, $N$, will allow the algorithm to find true pulses which are away from the expected trigger time, but will also increase the amount of bias at low amplitudes. This is demonstrated in Fig. E.4. Here, the probability density is shown for $A = 0$ and various values of $N$. At $N = 1$, we are in the OF0 situation, the algorithm has no freedom to choose and we obtain a Gaussian with width $\sigma_A$. However, as $N$ is increased, the additional freedom generally allows for higher amplitudes to be found somewhere within the window, and the fit is biased away from $\hat{A} = 0$.
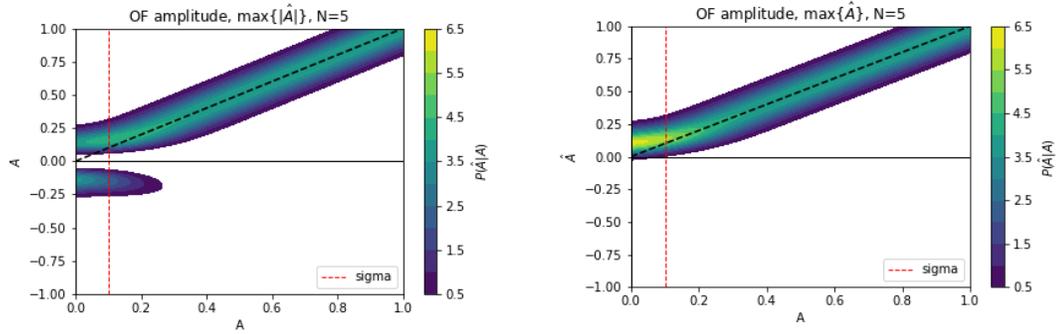
Figure E.3: Amplitude probability density of the OF algorithm when selecting $\max\{|X_i|\}$ ($\max\{X_i\}$) shown on the left(right). Both with N=5.
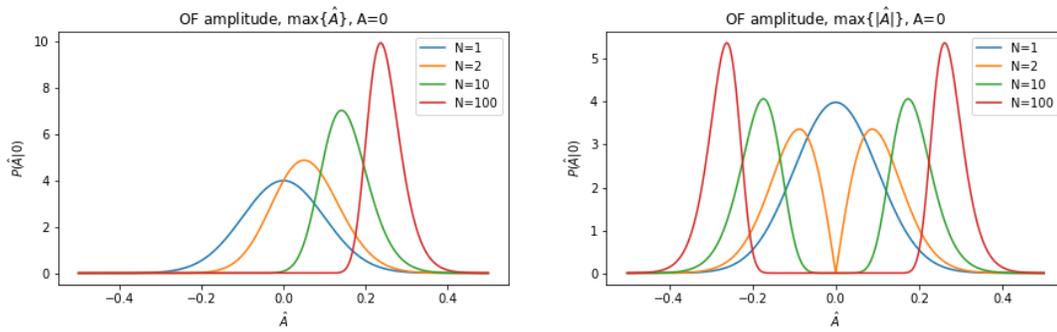


Figure E.4: OF probability density for $A = 0$, $\sigma_A = 0.1$ for different values of $N$. Case of $\max\{|X_i|\}$ ($\max\{X_i\}$) shown on the left(right).

With the above understanding, we see that, when applying the OF algorithm to triggered data, we should select a width for the OFdelay window (proportional to $N$) which is as low as possible, in order to reduce the low amplitude bias. The limit to this, of course depends on variations in pulse start time from event to event introduced by the DCRC triggering algorithm itself.

Now, we can use the above expressions to model the shape of noise distributions in real data. The left side of Fig. E.5 shows fits of Eq. E.13 and Eq. E.19 to the OF0 and OF distributions of randomly triggered PuBe data from UMN R68. This is an important dataset for the NR calibration experiment discussed in Ch. 7, but for now we just mention that it was taken with six channels of a Si HV detector operated at -125 V. The best-fit OF parameters can then be used to describe the expected position

of the noise wall in triggered science data. That is shown on the right side, where the OF(0) models have been used to predict the distortion introduced when measuring a uniform energy spectrum. The blue histogram, labelled $E$, is a uniform spectrum of $10^4$ energy values. The orange OF0 curve shows how the lowest energy events will sometimes be reconstructed with negative amplitudes, and thus negative energies, due to noise fluctuations. This tends to roll off the flat spectrum at low energies. The OF algorithm (here we use $\max\{X_i\}$) instead preferentially reconstructs low energy events to have slightly higher energies, creating an excess in the spectrum that starts at around $3\sigma_{\hat{A}}$. This has been used to inform the sensible placement of lower analysis thresholds and, more ambitiously, can be used to fit components of measured spectrum onto the leading edge of the noise wall.
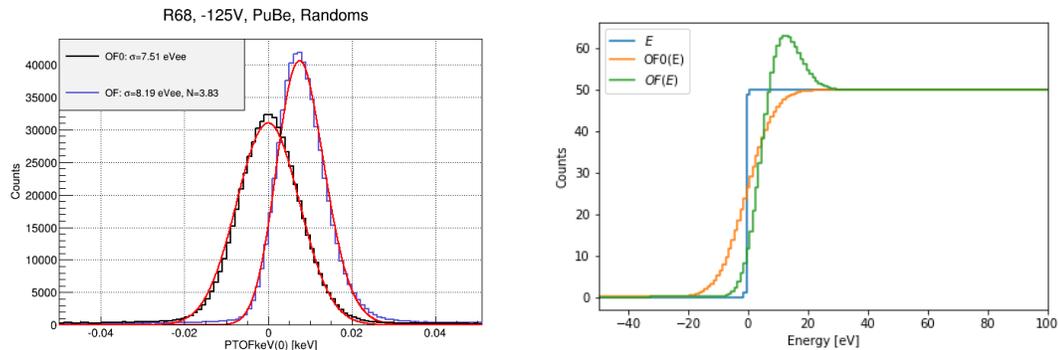


Figure E.5: Left: Fits to noise distributions in R68 PuBe data. Right: Simulation of how these effects will distort a uniform energy spectrum.

### E.1.3 Digitization Concerns

In much of the OF theory above, we assumed continuous quantities. In practice, measured signals are digitized at discrete time intervals. The amplitude digitization is so fine grained that it is not really an issue here. However, the expressions have to be modified slightly when dealing with the time digitization. The expressions derived above can be written in discrete form by replacing integrals with sums over the number of sampled points and the insertion of some factors of the sampling rate. The prescription for this procedure is given in [137]. One consequence of this discretization is that there are a finite number of $t_0$ values at which Eq. E.8 can be calculated and the true $t_0$ value is

likely between two of these. To accommodate this, instead of using the best calculated $t_0$ point, we actually fit the set of $(t_0, \hat{A})$ points near the maximum $\hat{A}$ with a quadratic and interpolate the best values.

Another significant complication has also arisen for application of OF algorithms to SNOLAB data. The problem is that the SNOLAB will use a new digitization scheme called "hybrid readout", where different portions of the saved traces will be saved with different digitization rates. As shown in Fig. E.6, the area near the triggered pulses will be fully sampled and a longer time portion will be saved with $16\times$ down-sampled values. The logic behind this method is to achieve sensitivity to low frequency baseline noise (of the type which plagued SuperCDMS Soudan) without having a prohibitively large data rate. Unfortunately, this introduces complications due to frequency correlations which occur when calculating Fourier transforms for the OF. These effects have been thoroughly investigated by B. Page [275]. He has shown that there is still a solution to this situation, but it involves computationally expensive (and not always stable) inversion of large matrices. At the time of this writing, members of the SuperCDMS Reconstruction and Machine Learning working group are studying alternative methods to handle this. They include machine learning approaches and methods such as empirically combining the results of separate OF fits to the finely sampled region and a down-sampled version of the full trace.
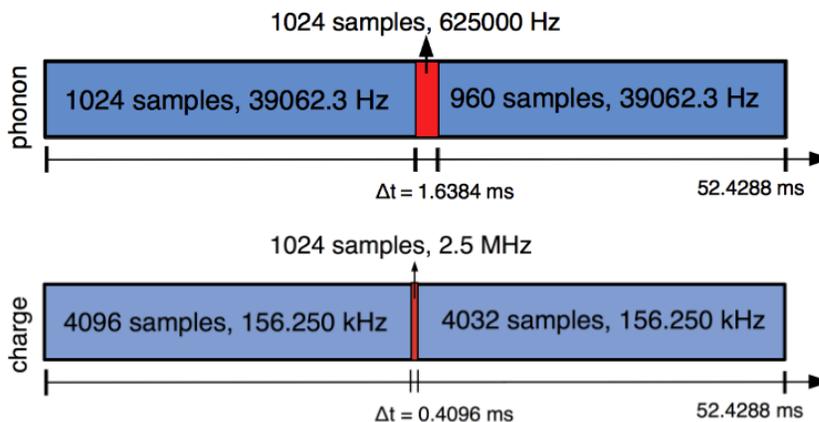


Figure E.6: Diagram of SNOLAB hybrid readout with variable sampling rate. Taken from [275].

## E.2  OF Variants

The standard OF and OF0 algorithms discussed above rely on the assumption that the pulse shape is independent of the pulse amplitude. For real detectors, this is generally true for the charge channels and for the phonon channels when the energy absorbed by TESs is small. However, for events with large energies, or when using large bias voltages to amplify phonon signals, significant local TES saturation can occur causing non-linear TES response. This results in phonon pulse shapes that change both amplitude and shape with absorbed energy [276]. Likewise, there can be event-position dependent effects because channels at different distances from an event will experience different phonon flux timing patterns. This can lead to pulse shape variations among channels in a given event. This section describes several variations on the vanilla OF algorithms with additional flexibility to account for second order effects such as these.

### E.2.1  Two-Template OF

One simple extension to the OF described above is include a second template and corresponding amplitude in the fit. This is commonly called the Two-template Optimal Filter (2TOF). While the primary template is generally constructed as the average pulse shape as before, the secondary template can be some next-order correction such as the average residual of single template fits. As before, the $\chi^2$ is constructed,

$$\chi^2 = \int \frac{|v(f) - (\hat{A}_1 s_1(f) + \hat{A}_2 s_2(f))e^{-i\omega t_0}|^2}{J(f)} df \qquad \text{(E.21)}$$

but is now minimized with respect to both amplitudes, resulting in

$$\frac{d\chi^2}{dA_i} = 0 = \int \frac{-2s_i^*(f)v(f)e^{-i\omega t_0} + 2\hat{A}_i|s_i(f)|^2 + 2\hat{A}_j s_i^*(f)s_j(f)}{J(f)} df \qquad \text{(E.22)}$$

from which we obtain the set of equations

$$\int \frac{s_i^*(f)v(f)e^{-i\omega t_0}}{J(f)} df = \hat{A}_i \int \frac{|s_i(f)|^2}{J(f)} df + \hat{A}_j \int \frac{s_i^*(f)s_j(f)}{J(f)} df \qquad \text{(E.23)}$$

The integrals on the right depend only on the templates and noise, so can be pre-calculated. The left integral is simply the application of the optimal filter for each

template to the measured signal. This system of equations can then be solved by substitution or matrix operations to yield the set of template component amplitudes, $\hat{A}_i$. Due to the additional degree of freedom, this algorithm generally yields better fits to real pulses, but comes with higher computational cost. This procedure can be extended to an arbitrary number of templates, with the same solution method and the size of the corresponding matrices growing with the number of templates.

**2TOF Template Testing**

The 2TOF algorithm was first used in SuperCDMS to analyze the CDMSlite pulses [118]. There, the two templates were constructed as the average pulse shape and the average residual, after fitting with the first template. These were called the slow and fast templates respectively, because the first average captured the general pulse shape, while the residual captured shorter timescale variations on the rising edge of the pulses. This procedure provided improved pulse fits and the resulting fast template amplitude obtained in 2TOF fits was used to characterize the event position in the detector.

To further explore the use of the 2TOF fitting method, I tried an alternative approach in template generation and tested it using UMN data. Using the mean pulse shape and residuals as the two templates is something akin to an expansion of the pulses using a certain basis. With those selections, it should provide a good fit for the average pulse and also for pulses with shapes only slightly different from the average. In a way, this is like starting with the average shape and working outwards. (One could imagine constructing further, higher order templates, by taking the residuals of those 2TOF fits, and so on.) One concern was that this way of calculating the two templates may not provide a set of templates that span the set of all pulse shapes fully. This is because the calculation of the average pulse will be skewed toward the most common pulse shapes, and not be well suited to fit rare events. This may be an issue because these devices are used for rare event searches but the majority of measured pulses are are from background data. As an alternative, I tried an approach where the two templates were selected to be at extreme ends of the range of pulse shapes, such that the 2TOF fits would be something like an interpolation between them.

Fig. E.7 indicates the sets of measured pulse shapes from which the templates were generated. The left figure shows a distribution of rise and fall times measured from a

set of real pulses in a Ge HV detector in UMN R61. The examples shown here were produced using measurements from all 12 phonon channels with the detector operated at 0 V bias. (Similar results were obtained with the detector operated at higher biases, but are not shown here.) Pulses from the extreme ends of this were used to construct the fast and slow templates. The highest density area was used to construct a single template OF for comparison. The resulting PT template shapes are shown on the right where the variation in both rise and fall times is clearly demonstrated. The ripple on the falling edge of PTfast is a result of electrothermal oscillations prominent in this setup. These are more likely to occur when the sensors receive a more sudden influx of phonon power, which is also when the rise times are faster.
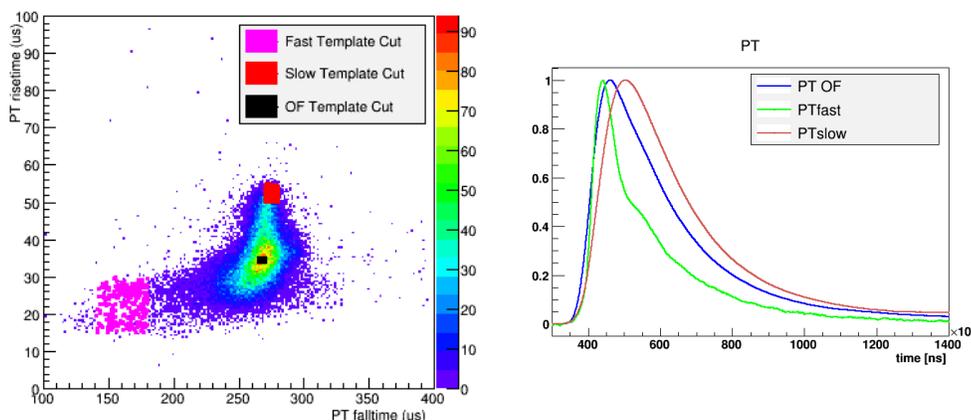


Figure E.7: Left: Selection regions for 2TOF and OF templates. Right: Resulting template shapes.

Fig. E.8 shows several examples of the fits in the time domain. As expected, the two template combination yields better fits to fast and slow events. However, the single OF template (perhaps not surprisingly) gives a better fit to pulses with average shape. It seems that, for intermediate pulse shapes, the 2TOF tends to give a peakier fit than needed. One notable feature is that both sets of templates have a more gradual initial rise than the raw pulses. This is likely due to the templates being built from traces with an assortment of start times which were not properly aligned.

After fitting a set of data with these new templates, an event energy quantity was constructed from a linear combination of the two PTOF amplitudes. The calibration
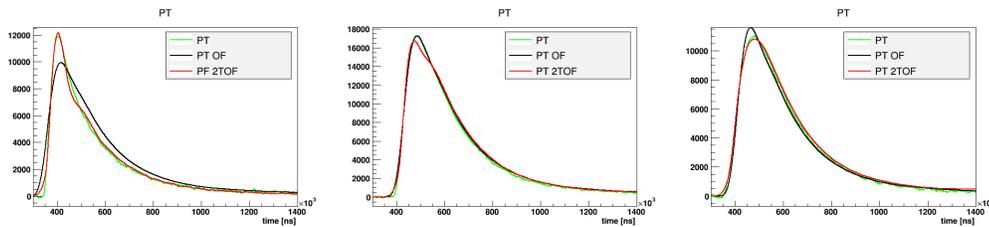
Figure E.8: Example fits to fast, medium, and slow pulses. 2TOF fits the extreme pulses best, but OF does a bit better for the medium-speed pulses.

factors were set using the 10.4 keV activation line in Ge (Appendix B of [251]). Fig. E.9 shows a scatter plot of the fast and slow template amplitudes with a linear fit to the calibration line. Note that the 10.4 keV pulses exhibited a range of mixing between the template types from almost entirely "fast", to completely "slow". This indicates that the method allowed the pulses to be fit over the full span of pulse shapes. There is a small population of events at either end of the calibration line with one negative amplitude. These are not yet fully understood, but many exhibit worse OF fit $\chi^2$ values than events with both amplitudes positive. There is also some evidence that the events with negative fast amplitudes have a rising pulse edge which is slow, but of a different shape than that of the slow template. Including a small negative fast component just happens to allow an improved fit to that shape.

The real test of the algorithm is shown in Fig. E.10. There is shown a comparison between the event energy as calculated using 2TOF and the single template OF. Both quantities have similar overall spectra, but the 2TOF gives a far better resolution. It is believed that this improvement is indeed due to the additional degrees of freedom to fit the range of pulse shapes allowed by the 2TOF.

While this alternative template selection process has shown improvements against the single template OF, it has not yet been directly compared to the CDMSlite style of a mean and residual - generated template set. It would be interesting to compare the relative coverage of the full space of pulse shapes between these two methods. Yet another interesting alternative is to use principle component analysis (PCA) to generate the templates. This method was specifically designed to select a minimal set of basis vectors (what we are calling here templates) to describe a data set [277].
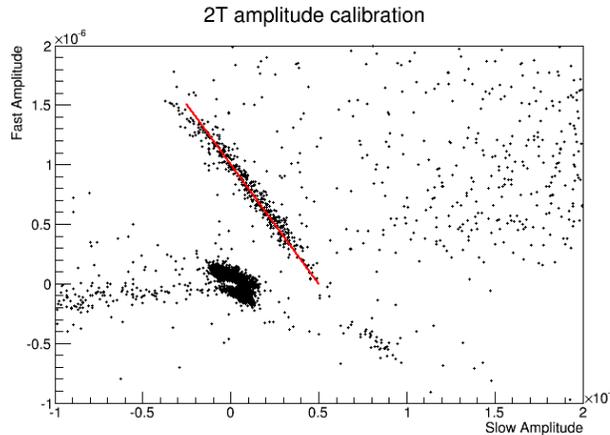
Figure E.9: Calibration of energy quantity from the 2TOF amplitudes. The solid red line indicates the relative scaling between the amplitudes which corresponds to a constant 10.4 keV energy.

### E.2.2  NSOF, MxN OF

There are still more advanced versions of OF which have been developed. One example is the MxN OF, which performs a simultaneous fit over N channels, each as a sum of M templates. There is also the Non-Stationary OF (NSOF) which is described in detail in Appendix E of [184]. When we developed the basic OF formalism above, we made the assumption that the noise was stationary, with its spectrum constant in time. However, this is not strictly realized in a real device. For example, in the phonon channels, the TES moves in the transition region as the phonons heat it. This changes the circuit impedance and thus the noise spectrum during the pulse. The NSOF takes such time correlated noise into account. In doing so, the mathematics becomes much more complex, with the noise PSD, $J(f)$, being promoted to a full frequency-space auto-covariance matrix, $V(f, f')$ and the $\chi^2$ definition becomes

$$\chi^2 = \int df \int df' \left(v(f) - \hat{A}s(f)e^{-i\omega t_0}\right) \frac{1}{V(f, f')} \left(v(f') - \hat{A}s(f')e^{-i\omega t_0}\right) \qquad \text{(E.24)}$$

This can also be solved analytically. For this method, one must generate both the pulse template and the auto-covariance of the noise instead of the simple PSD. The latter is calculated using the average residual noise after fitting each pulse with the template.
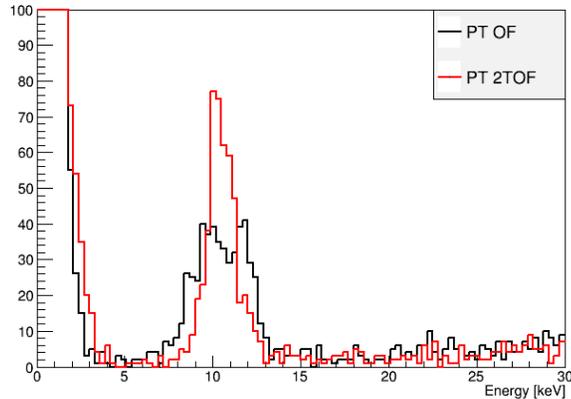
Figure E.10: Improvement in 10 keV calibration line with 2TOF.

This method was used as the energy estimator in CDMSlite. These algorithms have been described exhaustively elsewhere. Instead of repeating that work, we instead devote the next section to describe some tests of a new twist on OF which has not yet received as much coverage.

### E.2.3 Nonlinear Optimal Filter

Another way to account for changing pulse shapes is to allow the template shape to vary as well. In this method, we minimize the same $\chi^2$ expression as in the normal OF, but use a template whose shape is controlled by a set of parameters, $\vec{p}$.

$$s(t) \to s(\vec{p}, t) \tag{E.25}$$

Minimizing $\chi^2$ with respect to $\vec{p}$ will, in general, result in a set of non-linear equations, so we can not write a closed-form expression for the best fit $\vec{p}$ values. We thus call this the "Non-linear Optimal Filter", or NLOF. To find the best fit then, we must use the $\chi^2$ goodness of fit with a standard minimization algorithm. This makes NLOF much more computationally intensive than the above methods. (Really, even the vanilla OF is nonlinear in $t_0$, but since there is just the single parameter to minimize across, so we did not require a very sophisticated architecture.) In principle, NLOF can be more flexible to adapt to changing pulse shapes, but it relies on the user to develop a template

function which can faithfully represent the full range of pulse shapes but which also has a well-behaved $\chi^2$ in the $\vec{p}$ parameter space.

## E.3 NLOF Bias Study

This section summarizes a study on the effects of performing NLOF fits with a set of different template functions, some having more degrees of freedom than needed. In particular, we were interested in understanding how NLOF bias and resolution behaves for pulses near the noise level. It was done using the NLOF fitting and TES noise simulation functions found in the QETpy package [278] which was developed by the Berkeley SuperCDMS group.

A set of 2000 fake pulses, described by a simple 2 pole exponential function, were generated at 20 different amplitudes. At each amplitude, 100 pulses were generated with different instances of noise drawn from the same noise PSD. They were then fit using the NLOF using a template having either 1, 2, 3, or 4 poles which were compared to standard OF(0) fits. The form of the pulse shapes are given in Eq. E.26. There, $A$, $B$, and $C$ are all amplitudes, $t_0$ is the pulse start time, $\tau_r$ is the rise time, and the $\tau_f$ values are all fall times. The so-called 1 pole NLOF fit also uses the 2 pole function, but with the rise time a fixed parameter.

$$
\begin{aligned}
f_{2pole}(t) &= \Theta(t') \left( A e^{-t'/\tau_f} - A e^{-t'/\tau_r} \right) \\
f_{3pole}(t) &= \Theta(t') \left( A e^{-t'/\tau_{f1}} + B e^{-t'/\tau_{f2}} - (A+B) e^{-t'/\tau_r} \right) \\
f_{4pole}(t) &= \Theta(t') \left( A e^{-t'/\tau_{f1}} + B e^{-t'/\tau_{f2}} + C e^{-t'/\tau_{f3}} - (A+B+C) e^{-t'/\tau_r} \right) \\
t' &= t - t_0
\end{aligned}
\tag{E.26}
$$

An example of an NLOF fit is shown in Fig. E.11. The top left plot shows a comparison of the signal and pulse in frequency domain. The other figures show the fit in the time domain. This example used the 2 pole function, so the free parameters were amplitude, start time, and rise and fall times.

For the standard OF and 1,2 pole fits, the pulse amplitude is a natural parameter to compare. However, with 3 and 4 pole fits, there is no single pulse amplitude parameter that completely describes the pulse. For this reason, the integral of the best fit pulses
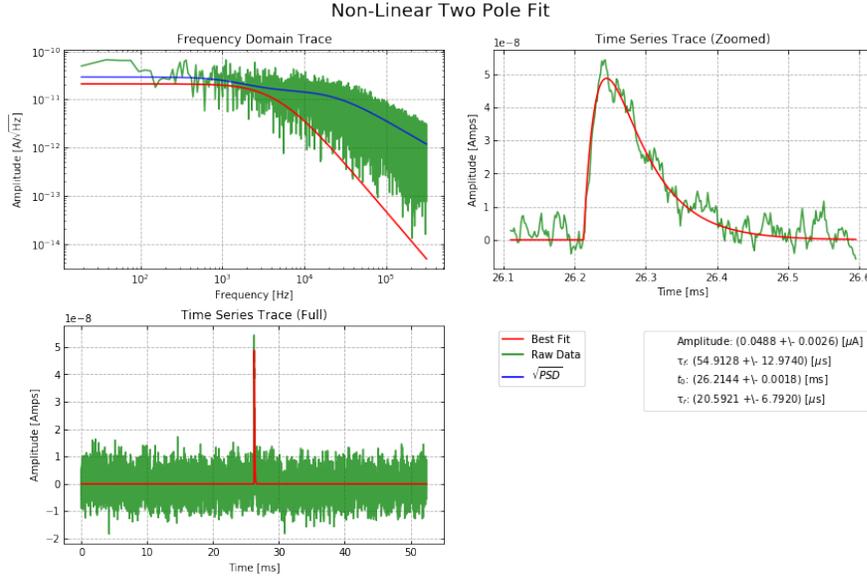
Figure E.11: Example NLOF fit with a 2 pole template function.

were compared. This not only is a well defined quantity for all fit procedures, it is the quantity which is actually proportional (to 1st order) to absorbed energy even for changing pulse shapes Eq. 3.56. Because the NLOF pulse templates are known analytically, the best fit integral can easily be calculated from the best fit parameters as follows:

$$
\begin{aligned}
I &\equiv \int_{t_0}^{\infty} f(t)dt \\
I_2 &= A(\tau_f - \tau_r) \\
I_3 &= A\tau_{f1} + B\tau_{f2} - (A+B)\tau_r \\
I_3 &= A\tau_{f1} + B\tau_{f2} + C\tau_{f3} - (A+B+C)\tau_r
\end{aligned}
\tag{E.27}
$$

Both the OF and NLOF algorithms also provide estimates of the uncertainty of returned quantities. For the OF, the expected integral resolution is calculated as

$$
\sigma_{I,OF} = \left( \int df \frac{s^* s}{J} \right)^{-1/2}
\tag{E.28}
$$

where $s$ is the template, normalized to integral of 1 and $J$ is the noise PSD. The NLOF algorithm returns the fit covariance matrix $\Sigma$, the entries of which are the covariance

of the model fit parameters near the best fit values. Given the Jacobian, $J'$ of the appropriate integral expression above (not to be confused with the PSD, $J$, which is a very different creature), the estimated variance of the pulse integral can be calculated as

$$\sigma_{I,NLOF} = \left( J'\Sigma J'^{T} \right)^{1/2} \tag{E.29}$$

where $J'$ is evaluated using the best-fit parameter values. The covariance matrix, $\Sigma$ involves only the first derivatives of the pulse fit function, so this variance estimate of the integral is only good to first order in the parameters. By comparing the sample variance from fitting several example pulses to this estimated variance, we can get a sense of when the model effectively becomes linear in the fit parameters.

Unlike the other OF variations discussed, the NLOF algorithm does not have an analytic solution, so the fit parameters obtained are dependent on the $\chi^2$ minimization settings, including initial parameter values and bounds. We describe here some tests of these settings, again in the context of fitting a noisy 2 pole true pulse.

### E.3.1 Default Parameter Bounds

In one test, the NLOF algorithm was given fixed initial $\tau$ values of; 20 $\mu$s rise time and falltimes of 100, 300, and 500 $\mu$s respectively. Initial values of $t_0$ and $A$ were estimated from the maximum point in the raw trace while $B$ and $C$ were both initialized at $A/3$. From there, the parameter bound requirements were to be within a factor of 100 of the initial amplitudes, a factor of 10 for the fall times, and within 30 bins for $t_0$. The fits were compared between seven different OF algorithms, described in Table E.1.

| Algorithm | Description | Fit Parameters, $\vec{p}$ |
|---|---|---|
| OF0 | OF with fixed $t_0$ | $A$ |
| OF | OF which maximizes $|\hat{A}|$ | $A$, $t_0$ |
| OFpos | OF which maximizes $\hat{A}$ | $A$, $t_0$ |
| 1-pole | NLOF with 2-pole template and $\tau_r$ fixed | $A$, $t_0$, $\tau_f$ |
| 2-pole | NLOF with 2-pole template | $A$, $t_0$, $\tau_r$, $\tau_f$ |
| 3-pole | NLOF with 3-pole template | $A$, $B$, $t_0$, $\tau_r$, $\tau_{f1}$, $\tau_{f2}$ |
| 4-pole | NLOF with 4-pole template | $A$, $B$, $C$, $t_0$, $\tau_r$, $\tau_{f1}$, $\tau_{f2}$, $\tau_{f3}$ |

Table E.1: OF fitting algorithms compared in this study.

Fig. E.12 shows comparisons of the fitted and true integrals (in units of $\mu A \times \mu s=$ pC) for each algorithm. The mean value at each true integral value, $I$, is shown as the solid blue line and the standard deviation is represented as the shaded region. The dashed red line indicates how perfect agreement would appear. Immediately one can see that OF0 provides the best fit, as expected. OF and OFpos are comparable down to low amplitudes, but not below the noise level. One important note here is that the distribution of OF near $I = 0$ is bimodal (as shown in Fig. E.4) so the mean and standard deviation do not capture the full story. Now, the NLOF fits performed decently at high energy, but showed a plateau similar to OFpos near $I = 0$. They also had noticeably higher variance.
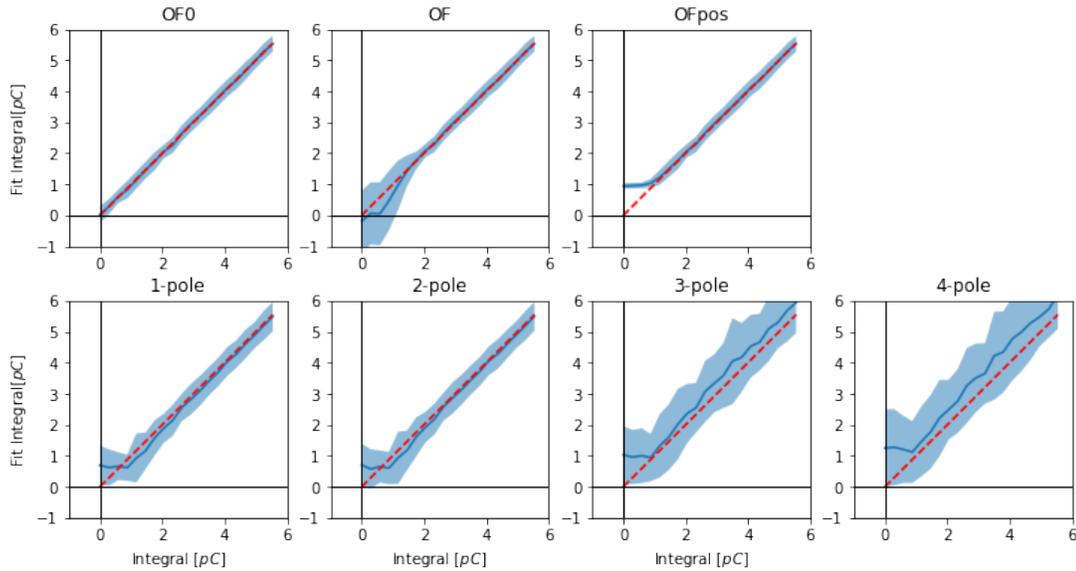


Figure E.12: Maps of fitted integrals for various OF fits. Solid blue lines indicate the average values while the shaded region denotes the standard deviation. The red dashed lines are of slope 1 and are shown for comparison.

Fig. E.13 provides another perspective on this same dataset. The figure on the left presents the deviation of the mean fit integral from the true integral (i.e discrepancy between the dashed red and solid blue lines in Fig. E.12). The error bars represent the standard error of the mean, $\sigma/\sqrt{N}$, i.e. the statistical uncertainty of these values due to the limited number of fitted pulses. Here we see that the OF0 is unbiased

everywhere and OF and OFpos are unbiased above the noise, all consistent with our theoretical treatments above. Near $I = 0$, OFpos becomes biased positive as expected. However, OF, free to chose any amplitude and start time, appears to be biased low in the noise. However, this may be an artifact of low statistics and the bimodal distribution. Theoretically OF should have an equal likelihood of selecting a positive or negative amplitude for a pure noise trace. (In fact, dedicated testing with this algorithm has borne this out, but it can take hundreds of samples for the mean to converge at low amplitudes.) Now, the 1 and 2 pole fits slightly underestimate the integral until they reach the noise, at which point they start to overestimate in a similar manner to OFpos, though interestingly with less bias near $I = 0$. The 3 and 4 pole fits were found to always overestimate the integral by as much as 50%. Interestingly, they are slightly less biased near the edge of the noise wall, but become more biased again further into the noise. The presence of positive bias in 2 and 4 pole fits is expected at some level because the amplitudes $B$ and $C$ were constrained to be positive and non-zero and the fall times $\tau_{f2}$ and $\tau_{f3}$ were required to be finite. However, this was not enough to account for such large biases. In addition, those parameters were generally not near their boundary values.
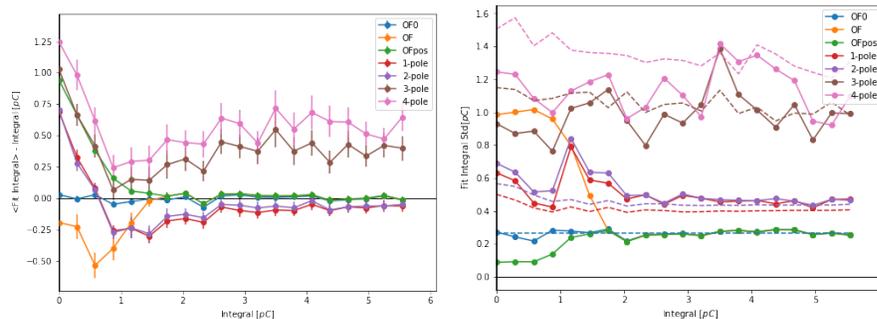


Figure E.13: Left: Deviations of mean fit integral from true pulse integral. Error bars represent standard error of the mean. Right: Standard deviation of fit integrals. Solid lines give the measured stdev of fit integrals while dashed lines show the theoretical expected stdev.

The right side of Fig. E.13 shows the standard deviation of fit values as a function of true integral. The solid lines represent the measured standard deviation of the fitted integral of the sample population at each true integral (i.e. 1/2 the shaded width in

Fig. E.12). The dashed lines are the average of the estimated integral variances based on the fit covariance as described in Eq. E.28 and Eq. E.29. Those are the theoretical estimates of variance from the individual fits. As expected, OF0 had consistently low standard deviation, consistent with the theoretical resolution. The standard deviation of OF becomes less meaningful at low amplitudes due to the distribution splitting as described above. OFpos actually becomes more precise below the noise, though this is paid for with decreased accuracy. The 1 and 2 pole fit standard deviations both come in with about $2\times$ the OF0 value, while the 3 and 4 pole values are roughly $4\times$ that of OF0.

Both the measured and estimated variances indicated that adding unnecessary degrees of freedom degrades the resolution in the scenario tested. All methods appeared consistent with their theoretical variance estimates well above the noise. However, below $I \sim 3.5$ pC, the 4 pole fit performed somewhat better than the estimate. The 1 and 2 pole fits generally yielded variances $10 - 20\%$ higher than the estimate except in a region near 1 pC where they deviate substantially. It is not yet understood what is driving this sudden deviation. As a reminder, the variance estimates for the NLOF fits are only strictly correct when $\chi^2$ is effectively linear in the parameters. We of course know this is generally not the case, so in some ways it is remarkable how close the agreement is.

### E.3.2 Guess NLOF $t_0$ from OF

Much of the loss of resolution of the NLOF fits near threshold in the above trials may simply be due to poor initial guesses for the start time, $t_0$. By default, $t_0$ was naively selected to be at the point of the raw trace with the largest positive amplitude. To improve this, it was suggested to use the start time found by the fixed template OF algorithms as the guess $t_0$ given to NLOF. This was done in 2 different manners. One trial made no assumptions about the pulse polarity, so utilized the OF start time as a guess, and imposed no amplitude bounds in the NLOF fits. The other trial assumed positive-going pulses and correspondingly used the OFpos start time and restricted the NLOF amplitudes to positive values.

Fig. E.14 shows the set of fits obtained when $t_0$ is selected using the OF start time. It shows that the NLOF fits generally are similar to the OF algorithm, with a flaring

standard deviation near $I = 0$. This supports the idea the the fit behavior near the noise is very sensitive to the start time selected. In this case, they are likely all trying to fit a similar noise excursion, instead of the true pulse. There is also a huge anomaly in one point of the 4 pole fit. This was due to a single fit which failed to converge. As of this analysis, we have not implemented any automatic procedures to check for, and handle such occasions.
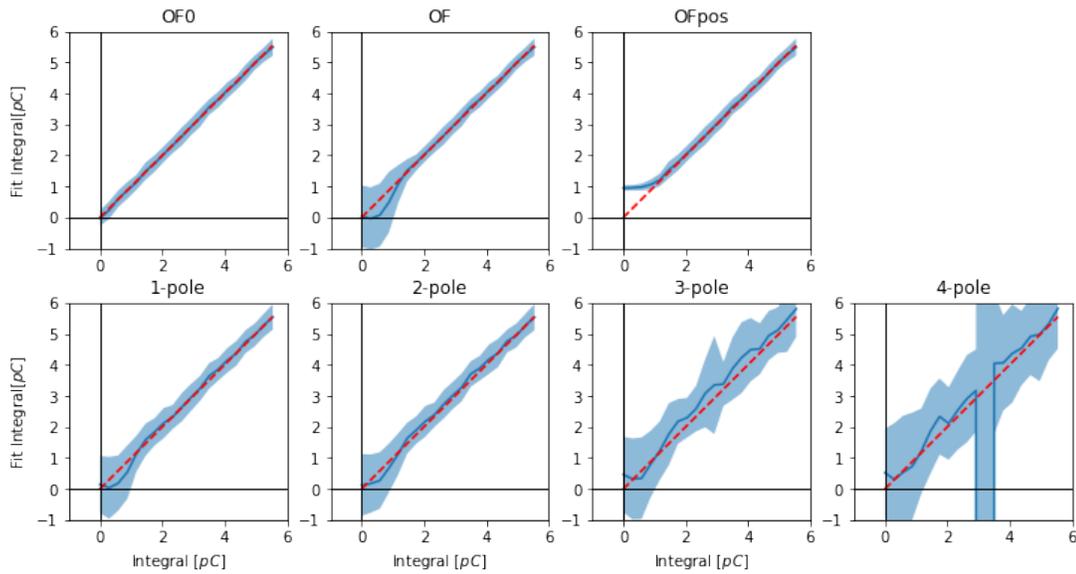


Figure E.14: NLOF fits with $t_0$ initialized by OF.

A similar result is found when OFpos is sued to supply the $t_0$ guess. This can be seen in Fig. E.15, where all of the NLOF fits follow the same general trend, with a slight positive bias near 0 (again, ignoring a few outliers due to failed fits). By comparing to the first trials above, this procedure has been shown to improve the 1 and 2 pole fits biases to match that of OFpos. In addition, it has lowered their variances near the noise considerably.

### E.3.3  Conclusions

The first steps have been made to systematically study biases in NLOF fits. A simple 2-pole pulse model was used to generate fake events which were fit with a variety of OF and NLOF algorithms. Known biases in OF integrals were observed and similar
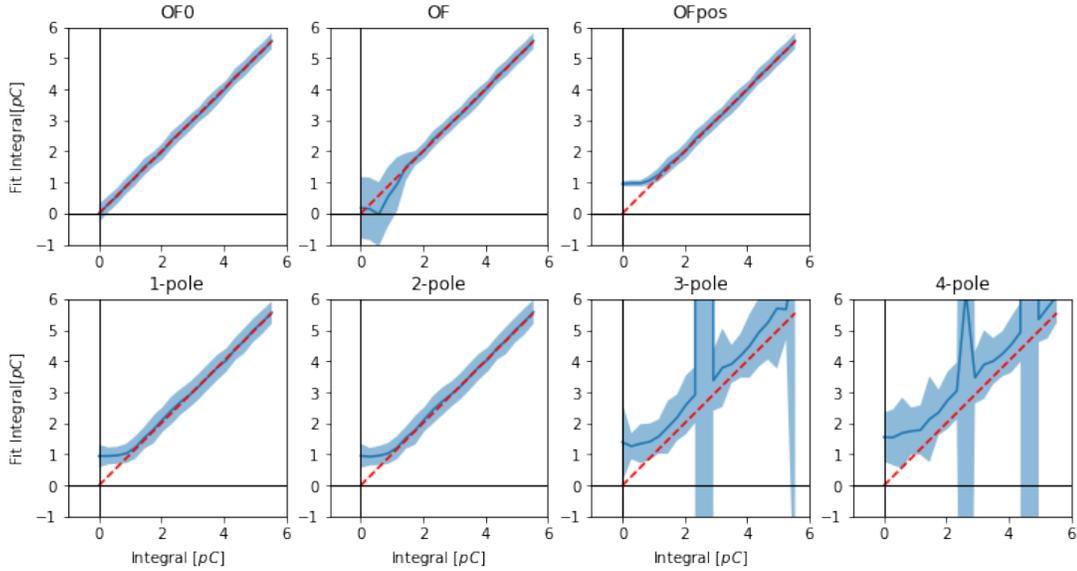
Figure E.15: NLOF fits with $t_0$ initialized by OFpos.

biases in NLOF integrals were seen as well. Additional bias and variance was found when the pulse template had unneeded degrees of freedom. Finally, it was shown that the output of standard OF algorithms can be useful to set starting parameter values in the more flexible NLOF, and can help improve the fit results thus obtained. The NLOF algorithm is a powerful tool, but there is still much to understand about how it may behave near the noise threshold. Though not discussed here, we have also found it fruitful to examine the distribution of the individual fit parameters (instead of just the implied integral). Future studies may also examine more complex situations where the underlying true pulse shape varies as a function of energy. It is expected that this will be a scenario in which the full flexibility of NLOF will be utilized.

# Appendix F

# Analytical (n,$2\gamma$) Decay model

This appendix presents a simple analytic model of the recoil energy distribution expected for a composite recoil cascade. We assume a recoiling nucleus has only 2 transitions to get to the ground state and the nucleus recoils in a random direction at each transition.

## F.1 Kinematics

We will call the energies of the first and second emitted gamma $E_{\gamma,1}$ and $E_{\gamma,2}$ respectively. With each recoil, the nucleus receives momentum $p = E_\gamma/c$ in some direction. These are low enough that the recoiling Si nucleus is non-relativistic, $v/c \sim 0.001$. For this reason, we will also assume that the linear stopping power, $S = -\frac{\partial E}{\partial x}$ is constant. Note that this greatly simplifies the complicated slowing-down process of the recoiling nucleus. Some calculations indicate the stopping power should actually decrease by a factor of a few as the nucleus slows down following the recoil [279]. However, this is a stochastic process that depends on the manner in which the nucleus initially sheds electrons and then later picks new ones up as it neutralizes and how it moves through the crystal lattice, etc. These details are not well understood at this time, so we make the constant $S$ assumption. Though it requires significantly more study, the $(n, \gamma)$ recoil techniques discussed elsewhere in this work may some day even be used as a probe of such microphysics!

In this simple model, the nucleus experiences the first decay at $t = 0$ while at rest. This imparts momentum $p = E_{\gamma,1}/c$ and the nucleus recoils with kinetic energy

$E(t = 0) = \frac{E_{\gamma,1}^2}{2mc^2}$ and velocity $v(t = 0) = E_{\gamma,1}/mc$. It then starts to slow through interactions with the crystal with constant acceleration is $a = S/m$. We can combine these to calculate the total energy deposited as a function of time

$$E_{dep,1}(t) = -x(t)S = -\left(\frac{E_{\gamma,1}}{mc}t + \frac{1}{2m}St^2\right)S \qquad (F.1)$$

The second gamma decay occurs at some later time, $t_d$. If the first recoil comes to a stop before the second recoil occurs, then the total deposited energy is easy to calculate as

$$E_{dep}(t_d) = \frac{E_{\gamma,1}^2 + E_{\gamma,2}^2}{2mc^2}, \quad t_d > t_{stop} \qquad (F.2)$$

where $t_{stop} = E_{\gamma,1}/|S|c$ is the time it would take the nucleus to come to a stop from the initial recoil. Of course we are more concerned with the case where the second decay occurs in-flight. Then, the remaining momentum from the first decay will be $p_1(t_d) = E_{\gamma,1}/c + St_d$. The second decay imparts additional momentum $p_2 = E_{\gamma,2}/c$ in a direction which differs from $p_1$ by some angle $\phi$. In this case, the resulting total momentum of the nucleus immediately after the second decay is

$$p^2(t_d) = p_1^2(t_d) + p_2^2 + 2p_1(t_d)p_2 \cos(\phi) \qquad (F.3)$$

Now, all of the energy from this second decay will get deposited. So to find the total deposited energy during the cascade sequence we take

$$E_{dep}(t_d, \phi) = E_{dep,1}(t_d) + E_{dep,2}(t_d, \phi) = \left(\frac{E_{\gamma,1}^2}{mc^2} - \frac{p_1^2(t_d)}{2m}\right) + \frac{p^2(t_d)}{2m}, \quad t_d < t_{stop} \quad (F.4)$$

which simplifies to

$$E_{dep}(t_d, \phi) = \frac{E_{\gamma,1}^2 + E_{\gamma,2}^2}{2mc^2} + \frac{E_{\gamma,1}E_{\gamma,2}}{mc^2}\left(1 - \frac{t_d}{t_{stop}}\right)\cos(\phi), \quad t_d < t_{stop} \qquad (F.5)$$

A plot of this is shown in Fig. F.1. We see that for long decay times, the angle does not matter and we get the quadrature sum of the energies. For short decay times, a boost along the same direction as the initial decay enhances the deposited energy. If the decays are perpendicular, we also recover the quadrature sum of energies as expected.
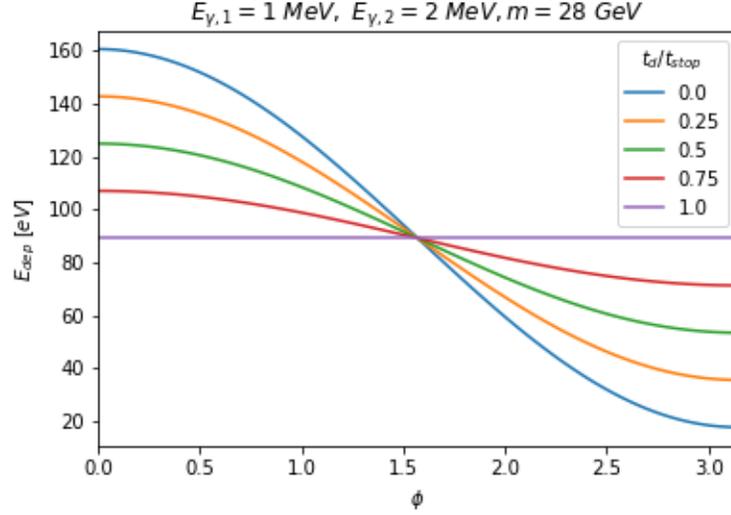
Figure F.1: Deposited energy versus angle for a range of decay lifetimes.

## F.2   Probabilities

Now we want to see how likely each recoil energy is. That is, we want the probability density function (PDF), $f_{E_{dep}}(E_{dep})$ of deposited energy $E_{dep}$. We know the PDFs for $t_d$ and $\cos(\phi)$ are

$$f_{t_d} = \frac{1}{\tau} \exp(-t_d/\tau)\Theta(t_d) \tag{F.6}$$

and

$$f_{\cos(\phi)} = \frac{1}{2}\Theta(1 - \cos(\phi))\Theta(1 + \cos(\phi)) \tag{F.7}$$

where the Heaviside step function, $\Theta(x)$, imposes the proper range of $t_d$ and $\cos(\phi)$. We just need to see how these PDFs change as we send them through each operations used to calculate $E_{recoil}$. For simplicity, we start by rewriting

$$E_{dep} = A + (B - C \times t_d)\cos(\phi) \tag{F.8}$$

where $A$, $B$, and $C$ are constants determined by the gamma energies, etc. We start with the operation of $-C \times t_d$. In general, when multiplying a random variable, $x$, having PDF $f_X(x)$ by a constant, $C$, the PDF of the resulting $Z = CX$ is $f_{CX}(z) = \frac{1}{|C|}f_X(\frac{z}{C})$.

Applying this rule, we obtain

$$f_{-Ct_d}(z) = \frac{1}{|C|} f_{t_d}\left(\frac{-z}{C}\right) = \frac{1}{C\tau} \exp(z/C\tau)\Theta(-z/C) \tag{F.9}$$

This is the PDF of the product of $-C$ and the random variable $t_d$. It retains the exponential shape, but is flipped and stretched out. Now we want to add an offset $B$ to obtain $B - Ct_d$.

When summing two random variables, $Z = X + Y$ with PDFs $f_X(x)$ and $f_Y(y)$, the result is the convolution $f_{X+Y}(z) = f_X(x) * f_Y(y) = \int f_X(z-y)f_Y(y)dy$. A special case is if one of those values, $Y = C$ is a constant. Then $f_Y(y) = \delta(y - C)$ and $f_{x+C} = f_X(z - C)$. So, applying this rule, we can obtain the next stage of the calculation

$$f_{B-Ct_d}(z) = f_{-Ct_d}(z - B) = \frac{1}{C\tau} \exp(\frac{z-B}{C\tau})\Theta(-\frac{z-B}{C}) \tag{F.10}$$

Next we need to take the product with $\cos(\phi)$. The rule for the product of two random variables $Z = XY$ is

$$f_{XY}(z) = \int f_X(x)f_Y\left(\frac{z}{x}\right)\frac{1}{|x|}dx \tag{F.11}$$

Using this, with the expressions above, we obtain

$$\begin{aligned}
f_{(B-Ct_d)\cos(\phi)}(z) = \\
= \frac{\exp(-B/C\tau)}{2C\tau} \int_{-\infty}^{\infty} \frac{\exp(x/C\tau)}{|x|}\Theta\left(\frac{B-x}{C}\right)\Theta(1 - z/x)\Theta(1 + z/x)dx
\end{aligned} \tag{F.12}$$

The constraints work out to mean that the integral is non-zero in two regions where $-\infty < x < -|z|$ or $|z| < x < B$. So we get two integrals and the whole expression is only valid for $|z| < B$.

$$\begin{aligned}
f_{(B-Ct_d)\cos(\phi)}(z) = \\
= \frac{\exp(-B/C\tau)}{2C\tau} \left(\int_{-\infty}^{-|z|} \frac{\exp(x/C\tau)}{|x|} + \int_{|z|}^{B} \frac{\exp(x/C\tau)}{|x|}\right)\Theta(B - z)\Theta(B + z)dx
\end{aligned}$$
$$\tag{F.13}$$

Now, these integrals do not have a nice closed form, but we can clean things up using the special function known as the "Exponential Integral", $E_i(x) = -\int_{-x}^{\infty} \frac{\exp(-t)}{t}dt$. A

couple changes of variables lets us write the above expression as

$$
f_{(B-Ct_d)\cos(\phi)}(z) =
$$
$$
= \frac{\exp(-B/C\tau)}{2C\tau} \left[ E_i\left(\frac{B}{C\tau}\right) - E_i\left(\frac{|z|}{C\tau}\right) - E_i\left(-\frac{|z|}{C\tau}\right) \right] \Theta(B-z)\Theta(B+z) \tag{F.14}
$$

The last step is to add in the constant, $A$ using the sum rule from before. We then obtain

$$
f_{E_{dep}} = f_{A+(B-Ct_d)\cos(\phi)}(z) =
$$
$$
= \frac{\exp(-B/C\tau)}{2C\tau} \left[ E_i\left(\frac{B}{C\tau}\right) - E_i\left(\frac{|z-A|}{C\tau}\right) - E_i\left(-\frac{|z-A|}{C\tau}\right) \right] \times \tag{F.15}
$$
$$
\Theta(B-z+A)\Theta(B+z-A)
$$

We also cheated a little bit and neglected the case when the second decay happens after the first has completely stopped. That is when $t_d > B/C$. This case occurs with probability $\exp(-B/C\tau)$, and we always get $E_{dep} = A$. So there is an additional piece to the total PDF that looks like $\exp(-B/C\tau)\delta(z-A)$

We can also go back and identify the constants $A$, $B$, and $C$ as follows:

$$
A = \frac{E_{\gamma,1}^2 + E_{\gamma,2}^2}{2mc^2}
$$
$$
B = \frac{E_{\gamma,1}E_{\gamma,2}}{mc} \tag{F.16}
$$
$$
C = -\frac{E_{\gamma,2}}{m}S
$$

A plot of the PDF of $E_{dep}$ for several values of $\tau$ is shown in Fig. F.2. The height of the central peaks are a little awkward to display because both the delta function and smooth PDF are formally infinite at $A$.

We see that, for short decay times, we get a uniform distribution because the total energy just depends on emission angle. At longer decay times, a peak appears around $A$, as it becomes less likely for the second decay to occur in-flight. In Fig. F.3, we show the relative fraction of events that appear in the delta-function peak. When the intermediate state lifetime is longer than the stopping time, there is basically no smearing, and all events deposit energy $A$. But as the lifetime becomes comparable to the stopping time, this quickly falls off.
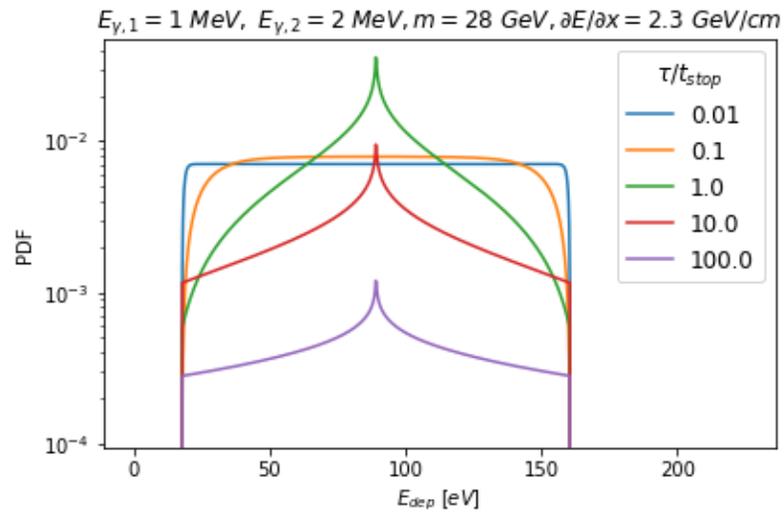
Figure F.2: Example PDF of deposited energy for a range of lifetimes.
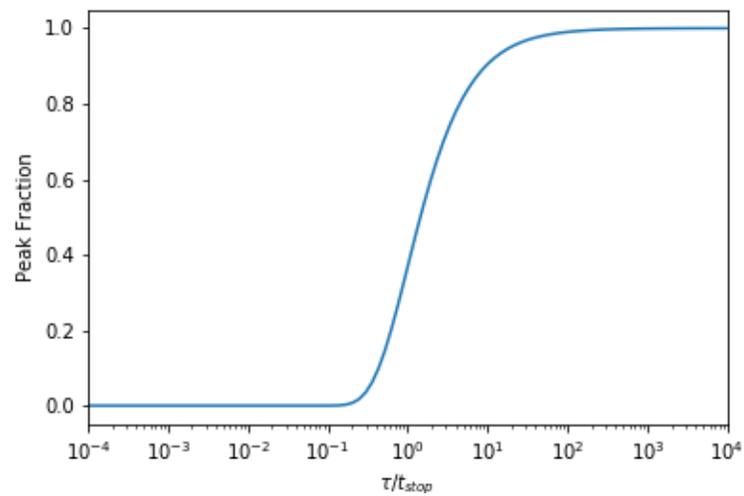


Figure F.3: Fraction of events in the delta-function peak as a function of lifetime.

# Appendix G

# Fano Width

## G.1 Contributions to Peak Width

In general, we parameterize the resolution of some detector at a measured energy, $E_m$ as

$$\sigma^2 = \sigma_0^2 + BE_m + (AE_m)^2. \tag{G.1}$$

Here, $\sigma_0$ is the baseline resolution, which is generally set by electronics or charge leakage noise and can be measured as the width of the randoms distribution. The $B$ term is related to the statistics of charge creation. Other detector or reconstruction effects can contribute to both $A$ and $B$.

## G.2 Fano Factor

When an interaction occurs in a semiconductor detector, a number of e/h pairs are produced. The average number is

$$\langle N \rangle = \frac{\langle E_{Ion} \rangle}{\varepsilon_\gamma} \tag{G.2}$$

where $\varepsilon_\gamma$ is $\sim 3.0(3.8)$ for Ge(Si) and $E_{Ion}$ is the amount of energy which goes into the production of e/h pairs. For a single recoil, we have $\langle E_{Ion} \rangle = Y(E_{Recoil})E_{Recoil}$, where $Y$ is defined to be 1 for electron-recoils and is a function of energy for nuclear recoils.

For $E_{Ion} \gg \varepsilon_\gamma$, the distribution of the number of charges produced is Gaussian

$$P(N) \propto \exp\left[-\frac{(N - \langle N \rangle)^2}{2F\langle N \rangle}\right] \tag{G.3}$$

where $F$ is the Fano factor [131] which describes the departure from Poisson statistics due to the correlated nature of charge production. This distribution has a mean of $\langle N \rangle$ and variance $\sigma_N^2 = F\langle N \rangle$.

Now, we generally do not measure $N$ directly. In a simple case, we instead measure some voltage, $V$ from our equipment which is linearly related to $N$, $V = aN$. In that case, the distribution of this measured voltage should look like

$$P(N) \propto \exp\left[-\frac{(\frac{V}{a} - \langle \frac{V}{a} \rangle)^2}{2F\langle \frac{V}{a} \rangle}\right] = \exp\left[-\frac{(V - \langle V \rangle)^2}{2Fa\langle V \rangle}\right] \tag{G.4}$$

i.e. the mean is now $\langle V \rangle$ and the variance in the same units as $V$ is $\sigma_V^2 = Fa\langle V \rangle$. Note that the *relative* width of the peak has not change just because we changed units. We just need two factors of $a$ to rescale, i.e. $\sigma_V^2 = a^2 \sigma_N^2 = aF\langle aN \rangle = Fa\langle V \rangle$.

The case of measuring phonon signals in a CDMS detector is a little different. We measure the total phonon energy, $E_T$, and divide by the NTL gain to get a measurement in electron-equivalent units, $E_{ee} = E_T/(1 + eV/\varepsilon_\gamma)$. Now, the total phonon energy is the sum of the initial recoil energy and the NTL phonon energy, $E_T = E_R + NeV$. But, across many trials with the same conditions, $E_R$ is constant while $N$ varies as described above. Thus $E_T$ is not *directly* proportional to the number of e/h pairs produced in a particular event. But the variance of $N$ directly gives us the variance of $E_T$ as

$$\sigma_{E_T}^2 = (eV)^2 \sigma_N^2 = (eV)^2 F\langle N \rangle \tag{G.5}$$

So the constant $E_R$ changes the offset of the $E_T$ distribution, but the width is set by variation in $N$. This leads to the following expression for variance of the calibrated electron-equivalent energy

$$\sigma_{E_{ee}}^2 = (eV)^2 F\langle N \rangle/(1 + eV/\varepsilon_\gamma)^2 = (eV)^2 F\frac{\langle E_{Ion} \rangle}{\varepsilon_\gamma}/(1 + eV/\varepsilon_\gamma)^2. \tag{G.6}$$