A Search for Low-Mass Dark Matter with the Cryogenic Dark Matter Search and the Development of Highly Multiplexed Phonon-Mediated Particle Detectors

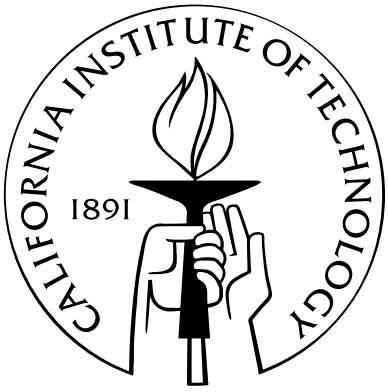
Thesis by

David Craig Moore

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

2012

(Defended May 8, 2012)

To Laurie

Acknowledgements

This thesis would not have been possible without the support and help of my family, friends, and collaborators. The CDMS experiment is the work of more than 80 individuals, who have contributed to the design, implementation, and day-to-day operation of the experiment, without which the data analyzed in this thesis would not exist. I am truly fortunate to have had the opportunity to work closely with, and learn from, a number of the excellent scientists and engineers who have made the CDMS experiment possible.

I am especially grateful to my adviser, Sunil Golwala, for giving me the opportunity to work with his group. Sunil's ability to identify and focus on the most important aspects of a problem has taught me how to approach technical problems in analysis or in the lab, and more importantly how to think about the bigger picture in order to work on interesting and productive topics. He has given me the freedom to work independently, while always being willing to provide advice when I became truly stuck. The original concept for a phonon-mediated detector using MKIDs was his, and without his support and encouragement even after the many wrong turns I took in the first few years of development, we would never have reached a working prototype.

Richard Schnee has provided insightful guidance on analysis techniques and understanding of the CDMS data, which has significantly improved the analyses presented in this thesis. Steven Yellin's advice on statistical methods has been a great help, in addition to his work to keep the SUF analysis cluster up and running. Thanks to Dan Bauer and Jim Beaty for providing constant guidance and advice during my shifts at Soudan. Both Bernard Sadoulet and Blas Cabrera have provided helpful insights on the low-threshold analysis and the design of the MKID-based phonon sensors presented here.

I was lucky to have the opportunity to learn directly from Walt Ogburn and Jeff Filippini after they came to Caltech as postdocs, who were both always willing to take time to share their extensive knowledge of CDMS with me despite the fact that they were busy

working to constrain CMB anisotropies. Jeter Hall and Walt deserve credit for suggesting a detailed low-threshold analysis of the Soudan data, as well as for providing insightful feedback and guidance throughout the analysis. The low-threshold analysis presented here relies significantly on Ray Bunker's work for the shallow-site data, which elucidated the most important challenges for understanding the CDMS detectors at low energy.

Zeesh Ahmed taught me how to use CAP, how to position correct, and how to take a complex analysis and reduce it to a set of manageable, well-defined tasks. I was incredibly lucky to have shared an office with Zeesh and to have worked closely with him on many projects. Some of the most fun (and productive) parts of grad school came during the c58 analysis when Jodi Cooley, Lauren Hsu, Matt Fritts, Oleg Kamaev, Scott Hertel, and Kevin McCarthy came to Caltech to work on analysis by day and croquet by night. Scott deserves special thanks for setting an inspiring example of just how many plots can be included in a single analysis note. I've greatly enjoyed and benefited from discussions over coffee about analysis, detector physics, and FPGA-based channelizers with Matt Pyle and Bruno Serfass. We've been lucky to have Bob Nelson join the CDMS group at Caltech, who has provided truly unique insights on a wide range of topics.

In addition to working with CDMS, I was extremely fortunate to be able to interact closely with Jonas Zmuidzinas and his group working on MKID development. Jonas' creativity and depth of knowledge are truly impressive, and many of the key breakthroughs needed to develop working phonon-mediated detectors stem from his ideas. The detectors discussed here build in large part on work done by Ben Mazin to develop optical/UV single photon detectors, and I'm especially grateful to Ben for teaching me how to design and test MKID devices as well as the unique joys of sub-Kelvin cryogenics. Bruce Bumble fabricated all the devices presented in this thesis, and such high quality devices would not have been possible without his careful attention to detail and willingness to work through many design iterations. Jiansong Gao's knowledge of resonator physics (and his resonator fitting code) were vital for acquiring high-quality MKID data and analyzing it correctly. I'm grateful to Tasos Vayonakis, Omid Noroozian, Ran Duan, Nicole Czakon, and Brett Cornell for making the MKID lab such a great place to work, and for teaching me about resonator design, device testing, and readouts.

Barbara Wertz and Kathy Deniston deserve special thanks for keeping the group running, without whom none of the work presented here would be possible. Thanks to both of

you for helping me countless times over the past several years and providing much support and encouragement along the way.

My parents started me on this path by instilling in me a deep love of science and learning when I was young and giving me the resources that I needed to pursue my interests. My brother, Rich, has always pointed me in the right direction, and continues to do so today. I can't thank them enough for their love and encouragement over the years.

Laurie, you were willing to come across the country with me so I could study at Caltech and have always supported me. Your love and support is the only reason any of this has been possible.

Abstract

A wide variety of astrophysical observations indicate that approximately 85% of the matter in the universe is nonbaryonic and nonluminous. Understanding the nature of this "dark matter" is one of the most important outstanding questions in cosmology. Weakly Interacting Massive Particles (WIMPs) are a leading candidate for dark matter since they would be thermally produced in the early universe in the correct abundance to account for the observed relic density of dark matter. If WIMPs account for the dark matter, then rare interactions from relic WIMPs should be observable in terrestrial detectors. Recently, unexplained excess events in the DAMA/LIBRA, CoGeNT, and CRESST-II experiments have been interpreted as evidence of scattering from WIMPs with masses $\sim$10 GeV and spin-independent scattering cross sections of 10^{-41}–10^{-40} cm^2.

The Cryogenic Dark Matter Search (CDMS II) attempts to identify WIMP interactions using an array of cryogenic germanium and silicon particle detectors located at the Soudan Underground Laboratory in northern Minnesota. In this dissertation, data taken by CDMS II are reanalyzed using a 2 keV recoil energy threshold to increase the sensitivity to WIMPs with masses $\sim$10 GeV. These data disfavor an explanation for the DAMA/LIBRA, CoGeNT, and CRESST-II results in terms of spin-independent elastic scattering of WIMPs with masses $\lesssim$12 GeV, under standard assumptions. At the time of publication, they provided the strongest constraints on spin-independent elastic scattering from 5–9 GeV, ruling out previously unexplored parameter space.

To detect WIMPs or exclude the remaining parameter space favored by the most popular models will ultimately require detectors with target masses $\gtrsim$1 ton, requiring an increase in mass by more than two orders of magnitude over CDMS II. For cryogenic detectors such as CDMS, scaling to such large target masses will require individual detector elements to be fabricated more quickly and cheaply, while maintaining the nearly background-free operation of the existing experiment. We describe the development of athermal phonon mediated

particle detectors using Microwave Kinetic Inductance Detectors (MKIDs), which could provide a simpler path to extending the CDMS detector technology to the ton scale. Results from prototype devices have demonstrated energy resolutions as good as $\sigma = 0.55$ keV at 30 keV, comparable to existing CDMS II detectors. Such designs can be scaled to kg-scale detector elements, while reducing the complexity of the detector fabrication and cryogenic readout electronics relative to existing designs. Since MKIDs are naturally multiplexed in the frequency domain, MKID-based designs also allow much finer pixelization of the phonon sensor, which is expected to enhance background rejection for large detectors while simultaneously reducing the number of wires needed to read out the detectors.

Contents

List of Figures

List of Tables

Chapter 1

Evidence for Dark Matter

Understanding the nature and evolution of the fundamental constituents of the universe is a central goal of cosmology, with significant implications for how we view the universe and our place in it. Over the past several decades, astrophysical observations have led to significant improvements in this understanding. Many complementary observational probes have led to the conclusion that the vast majority of the energy density in the universe is in the form of dark matter and dark energy, whose fundamental nature is poorly understood. "Ordinary" matter such as electrons, protons, and neutrons—of which we are made, as well as the earth, stars, and all other luminous objects that we observe—accounts for less than 5% of the total mass-energy. Understanding the nature of the "dark" components, which make up the remaining 95% of the energy density, is one of the most important open questions in physics, with significant implications for our fundamental understanding of cosmology and particle physics.

Figure 1.1 illustrates our current picture of the composition of the universe. In the present era, dark energy, which is consistent with a constant vacuum energy everywhere in space, makes up 72% of the total energy density. Although understanding the nature of dark energy is also a key open problem, in this dissertation we will focus on identifying and characterizing dark matter, which makes up the bulk of the remaining energy density (23% of the total energy density, or 85% of the matter density). Although known to interact gravitationally, dark matter is described as "dark" because it does not interact electromagnetically and emits no corresponding electromagnetic radiation which can sensed by telescopes. If dark matter interacts with ordinary matter through a force other than gravity, it must do so very weakly. A host of experiments aimed at directly detecting such interactions of dark matter in terrestrial detectors are currently in progress. Although no

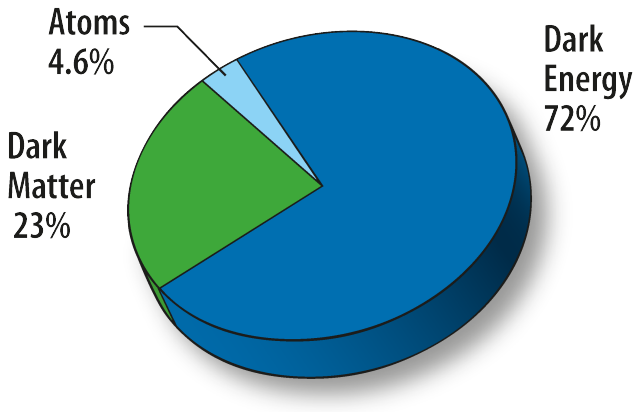


Figure 1.1: Current composition of the universe in the standard cosmological paradigm. The majority of the energy density consists of dark energy, while nonbaryonic dark matter provides ∼85% of the total matter composition. *Figure from NASA/WMAP Science Team [1]*

unambiguous detections of these interactions exist, several experiments currently claim to be seeing the first hints of such a signal.

The evidence supporting the cosmological paradigm described above will be reviewed in the following sections, focusing on observations pointing to dark matter as the dominant constituent of matter in the universe. Evidence for dark matter is seen in astrophysical observations at a wide variety of scales, from the scale of the smallest galaxies (Sec. 1.1.1), to clusters of galaxies (Sec. 1.1.3), to cosmological scales (Secs.1.2.1–1.2.3). Although theories can be found to explain some subset of these observations without the presence of dark matter, the dark matter paradigm is the most successful framework for simultaneously explaining all observations. Generic properties required for any dark matter candidate consistent with these observations will be discussed in Sec. 1.3.1, as well as specific candidates which meet these requirements in Secs. 1.3.2–1.3.5.

1.1 Astrophysical evidence

A variety of astrophysical observations point to the need for "missing mass," i.e., the mass of an astrophysical object or collection of objects inferred from its temperature, velocity distribution, or the lensing of light from background objects is significantly larger than would be expected from the visible matter alone. Although a complete summary of all such observations is beyond the scope of this work, the following sections highlight some of the

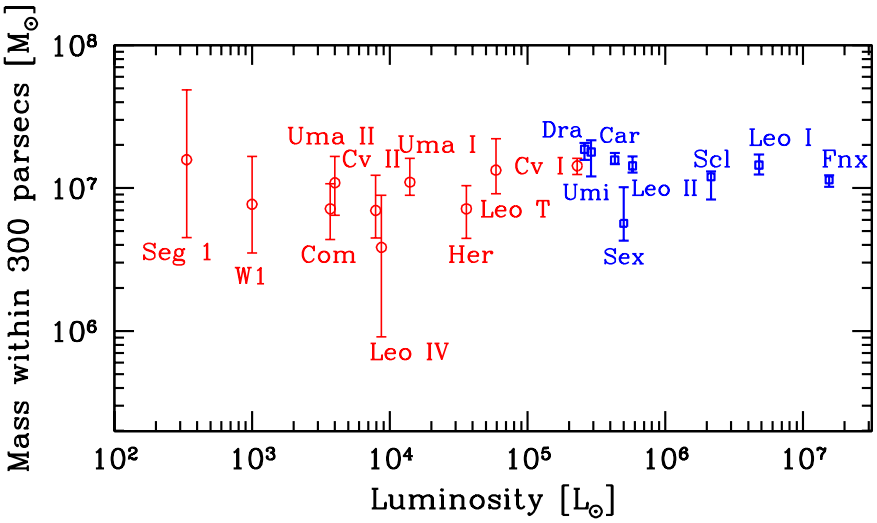


Figure 1.2: Mass inferred from the stellar velocity dispersions of the recently discovered ultra-faint dwarf galaxies orbiting the Milky Way (red). Large gravitational masses are inferred over several orders of magnitude in luminosity, indicating the need for significant nonluminous matter in the faintest objects. *Figure from Strigari et al. [4]*

most striking observations that suggest the need for a significant nonluminous component in a variety of astrophysical systems.

1.1.1 Dwarf galaxies

The smallest galaxies are some of the most dark-matter dominated objects in the universe. In recent years, a significant number of new Milky Way satellites with extremely low surface brightness ("ultra-faint" dwarf galaxies) have been discovered [2] using data from the Sloan Digital Sky Survey (SDSS) [3]. The mass of these objects is typically inferred from their stellar velocity dispersions, which can be used to determine the mass through the virial theorem by relating the stars' average kinetic energy to the gravitational potential. A summary of the inferred mass for the Milky Way satellites is shown in Fig. 1.2. For the faintest of the recently discovered ultra-faint dwarf galaxies, the mass-to-light ratio can approach nearly 1000 times the solar mass-to-light ratio. This is nearly two orders of magnitude larger than typical spiral and elliptical galaxies. This low luminosity is due to inefficient star formation in these systems, since it is difficult to accrete significant amounts of gas without the gas being stripped, e.g., by supernovae or tidal disruption from the host galaxy.

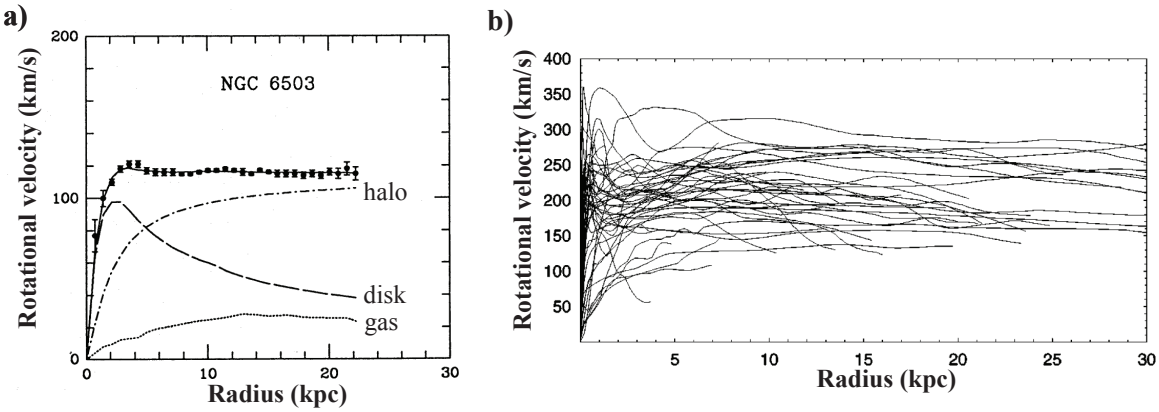


Figure 1.3: a) Rotational velocity versus radius for the spiral galaxy NGC 6503. The observed circular velocities (errorbars) are approximately constant at high radius, in contrast to the drop-off in velocity that would be expected if the bulk of the mass were contained in the luminous disk. A three-component fit to the rotational velocities is shown, including the inferred mass from the disk, gas extending past the disk, and the inferred halo of nonluminous matter extending to high radii. *Figure from Begeman et al. [5].* b) Rotational velocity versus radius for a variety of spiral galaxies, indicating that the constant dependence of the velocity at high radius is a generic property of these objects. *Figure from Sofue and Rubin [6]*

1.1.2 Spiral galaxies

On somewhat larger scales, spiral galaxies offer clear evidence for the need for a significant, nonluminous matter component surrounding the visible matter. The rotational dynamics of spiral galaxies offers a powerful probe of the mass of a galaxy since the circular velocity for a satellite in a given orbit depends only on the radius of the orbit and the total enclosed mass. This rotational velocity can be inferred from the Doppler shifts of various spectral lines (e.g., rotational CO lines or the 21 cm spin-flip transition of neutral hydrogen) as a function of radius from the center of the galaxy. Since neutral hydrogen is present at very high radii, these measurements can be extended well past the galactic disk, which contains the bulk of the luminous matter. Circular velocities can be inferred from the measured line-of-sight velocities after accounting for the proper motion and angle at which the disk is viewed.

These measurements have been performed for a large number of galaxies, with the typical dependence of the rotational velocity versus radius shown in Fig. 1.3a. If the mass of the galaxy were primarily concentrated in the luminous disk, simple Newtonian dynamics would predict that the rotational velocities would fall off like $1/\sqrt{r}$ at radii greater than several

kpc once past the edge of the disk. Instead, the rotational velocities are typically constant to very high radii, as shown in Fig. 1.3b for a large number of galaxies. This dependence is consistent with an additional nonluminous component, which is spherically distributed with a density $\rho \sim 1/r^2$ at high-radius, r, implying that the enclosed mass $M(r) \sim r$. In this picture, the luminous disk is surrounded by a much larger halo of dark matter extending well past the edge of the disk. This model, and its relevance to detecting the scattering of relic dark matter particles from the halo in terrestrial detectors, will be discussed in more detail in Sec. 2.1.

1.1.3 Galaxy clusters

Some of the strongest astrophysical evidence for dark matter is apparent on the scale of galaxy clusters, which are the largest objects in the universe to have undergone gravitational collapse. These objects typically contain hundreds of galaxies embedded in an intergalactic medium consisting of hot, x-ray emitting gas. The first evidence for the presence of significant amounts of nonluminous matter in galaxy clusters was found by Zwicky in 1933 [7] by measuring the velocity dispersion of galaxies in the Coma cluster and inferring the total mass from the virial theorem, analogous to the procedure described for dwarf galaxies in Sec. 1.1.1. In addition to improved measurements of the velocity dispersion, we now have additional independent measurements of the cluster mass from observations of gravitational lensing of background objects as well as the determination of the intracluster gas temperature from x-ray emission for a large number of clusters. New surveys which measure the distortion of the cosmic microwave background (CMB) due to inverse Compton scattering from hot electrons in clusters (the Sunyaev-Zel'dovich [SZ] effect) are also beginning to provide a new tool to detect clusters and constrain their mass.

Since clusters are large enough to provide a representative sampling of the cosmological matter density, these mass measurements provide an important cross-check on the dark matter and baryonic matter densities determined from the cosmological observations described in Secs. 1.2.1–1.2.3. The matter density, ρ_m, is typically specified in terms of the density parameter, $\Omega_m = \rho_m/\rho_c$, where $\rho_c = 3H^2/8\pi G$ is the critical density corresponding to a flat universe [8]. The total matter density consists primarily of cold dark matter and baryonic matter whose corresponding densities are denoted by Ω_c and Ω_b, respectively. As will be discussed in Sec. 1.2.1, a combination of cosmological measurements indicates that

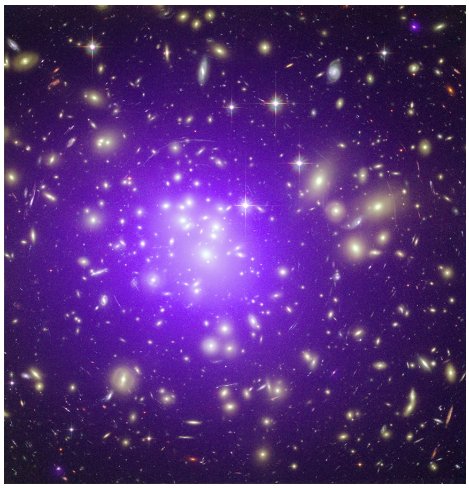


Figure 1.4: Composite image of Abell 1689 in the optical (yellow) and x-ray (purple). Abell 1689 is a strong gravitational lens, and multiple images and arcs characteristic of lensing are apparent. The bulk of the baryonic matter in the cluster lies in the hot intracluster gas, which is heated to $\sim 10^8$ K and emits thermally in the x-ray. *Image from Chandra X-ray Observatory, NASA/STScI [13]*

$\Omega_m = 0.267 \pm 0.025$, with $\Omega_c \approx 5\Omega_b$ [9].

Figure 1.4 shows a composite image of Abell 1689 at optical and x-ray wavelengths, demonstrating a few of the techniques which can be used to reconstruct the cluster mass. Spectroscopic measurements at optical and infrared wavelengths allow the determination of the line-of-sight velocities of the individual galaxies. Following Zwicky, such measurements have been used to determine the dynamical mass assuming virialized clusters for hundreds of objects (e.g., [10–12]). These measurements typically find $\Omega_m \approx 0.2$, with mass-to-light ratios of a few hundred times that of the sun, indicating that the mass of a cluster must be highly dominated by some nonluminous component.

The vast majority of the baryonic matter in the cluster (80–90%) lies not in the stars and galaxies, but within intracluster gas heated to nearly 10^8 K by the gravitational potential. This gas emits thermally at x-ray wavelengths, and galaxy clusters provide some of the brightest objects in the x-ray sky. For relaxed clusters that have not undergone a recent merger, the gas temperature determined from the x-ray emission can be used to infer the total cluster mass, under the assumption of hydrostatic equilibrium. This method has also been applied to a large number of clusters and gives good agreement with the inferred virial masses [14, 15].

In addition to their x-ray emission, clusters can be identified through the inverse Compton scattering of CMB photons from hot electrons in the intracluster gas. This effect produces a unique spectral distortion in the temperature of the CMB observed at earth in the direction of the cluster, with a decrement in intensity below 220 GHz and an increment above. It is especially effective for identifying high-redshift clusters since it measures a fractional change in the CMB brightness, for which the dimming at large distances is canceled by the higher intensity of the CMB at earlier times, leading to a signal that is independent of redshift. A recent survey of the southern sky using this effect has identified over 200 galaxy clusters, more than 100 of which were previously unknown [16]. Since the SZ effect depends on the total mass of a cluster [17], it provides another complementary measurement which can be used to determine the cluster mass once calibrated by x-ray and kinematic measurements.

Finally, the gravitational potential of a cluster distorts spacetime, leading to the lensing of background objects. For the strong gravitational lensing provided by Abell 1689, light from background objects arriving along multiple paths can be deflected to earth, leading to multiple images of the same background object at different locations around the lens. In addition, light from extended background objects can travel slightly different paths, leading to the arcs centered on the core of the cluster seen in the image. From a detailed reconstruction of the multiple images and distortion of background objects seen in the lensed image, the mass distribution in the lensing cluster can be reconstructed [18, 19]. At larger distances or for weaker lenses, the statistical distribution of a large number of only slightly distorted background images can be used to infer the mass [20, 21]. Lensing constraints are valuable since they provide an independent cross check on other methods of measuring cluster mass, which is valid regardless of the equilibrium assumptions needed for the dynamical methods described above [20].

The four independent cluster mass measurements described above give good agreement with the value of Ω_m derived from cosmological observations, with typical values of $\Omega_m \approx 0.2$–0.3 (e.g., [11, 14, 21]). These measurements further demonstrate that the luminous mass in these clusters accounts for only a small fraction of the total mass, leading to the conclusion that the dominant constituent of the mass in the clusters is due to dark matter.

This conclusion is reinforced by the recent discovery of several clusters undergoing mergers in which the distribution of mass and baryons are spatially separated. Two examples of

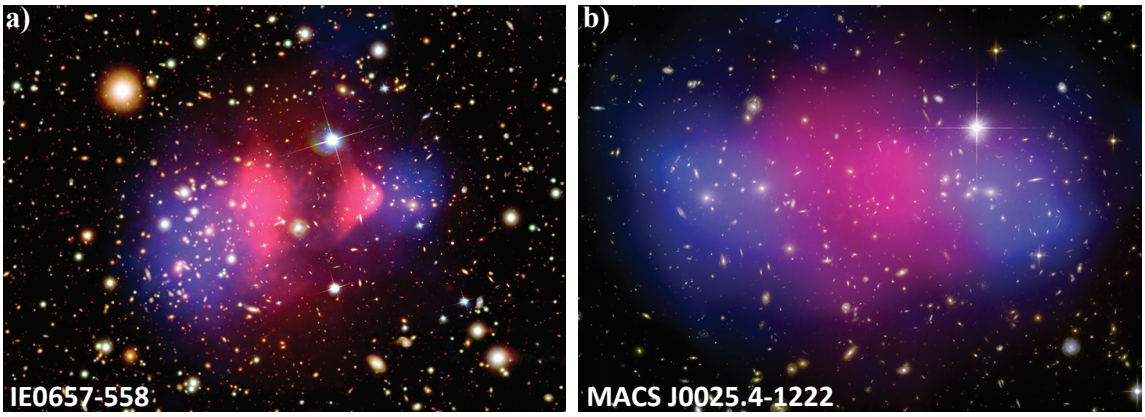


Figure 1.5: a) Composite optical and x-ray image of the Bullet cluster (IE0657-558). The optical image (yellow) from the Hubble Space Telescope is overlaid with the mass distribution reconstructed from lensing (blue) and the x-ray emission from the Chandra X-ray Observatory (pink), which traces the bulk of the baryons in the cluster. *Image from Clowe et al.; Chandra X-ray Observatory; NASA/STScI [13, 22].* b) Similar composite image for MACS J0025.4-1222. *Image from Bradač et al.; Chandra X-ray observatory; NASA/STScI [13, 23]*

these objects are shown in Fig. 1.5. In these composite images, the optical image from the Hubble Space Telescope is overlaid with both the total mass distribution inferred from lensing as well the x-ray emission, which primarily traces the intracluster gas. For both objects, a clear separation is seen between the mass that is reconstructed from lensing and the bulk of the ordinary matter in the gas. This separation is understood as the result of a collision of two clusters, in which the approximately collisionless dark matter passes through unimpeded, while the hot gas interacts and is stripped from the bulk of the matter distribution. The "Bullet cluster" (IE0657-558) provides a striking example of such a collision, with clear shock waves formed in the gas distribution. Since the gravitational potential is offset from the baryonic matter distribution, such observations cannot be explained without invoking a nonluminous, collisionless component that dominates the total mass of the cluster.

1.2 Cosmological evidence

While observations of astrophysical objects over a large variety of scales provide a clear need for a significant fraction of the matter in these objects to be nonluminous and collisionless, our best constraints on the density and properties of this dark matter come from cosmological information. In particular, detailed measurements of anisotropies in the cosmic microwave background (CMB), which provide the initial conditions for the growth of

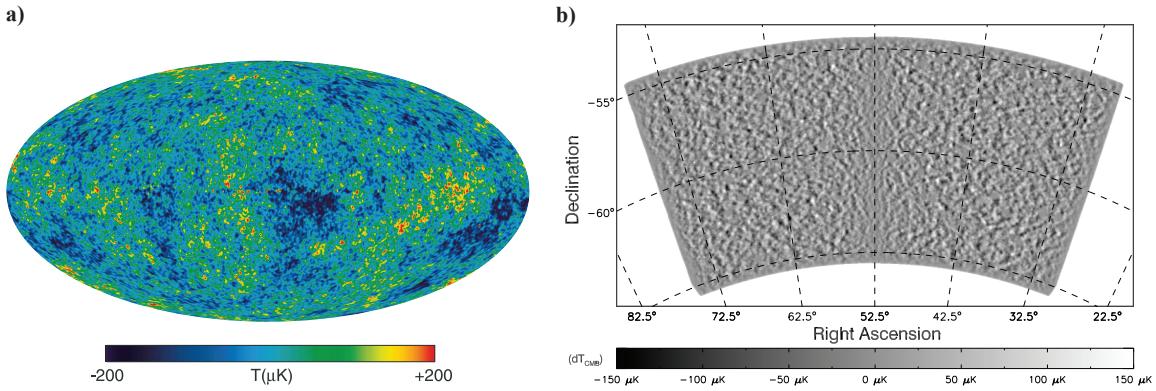


Figure 1.6: a) All-sky map of the CMB temperature anisotropies in galactic coordinates from WMAP. *Figure from Jarosik et al. [24].* b) Map of small scale anisotropies taken by SPT in a 236 square degree patch of the southern sky. The data have been high-pass filtered to remove modes with $l \lesssim 600$. *Figure from Keisler et al. [9]*

the large-scale structure (LSS) of the universe seen today, allow detailed constraints on the density and evolution of the matter in the universe to be determined. In addition, precise predictions of the amount of baryonic matter created in the early universe are in excellent agreement with CMB and astrophysical observations. The picture of the universe resulting from the combination of these complementary measurements is incredibly successful at describing the evolution of the universe over time, clearly establishing the need for dark matter and allowing precise measurements of its total density.

1.2.1 The cosmic microwave background

Approximately 370,000 years after the big bang ($z \sim 1100$), the universe had cooled sufficiently that neutral hydrogen could be formed. This epoch, typically referred to as "recombination," marked the point at which the universe became transparent to photons. The black body radiation released at this time is observed today as the cosmic microwave background (CMB), which is a 2.7 K black body spectrum observed to be highly uniform across the entire sky (with residual anisotropies at the level of 1 part in 10^5). Obviously, we no longer see such a highly uniform universe today, but instead see a rich variety of structures that were seeded by overdensities imprinted in the CMB anisotropies. Measurement of these anisotropies provides a rich data set from which the density of baryonic matter and dark matter during this period can be inferred.

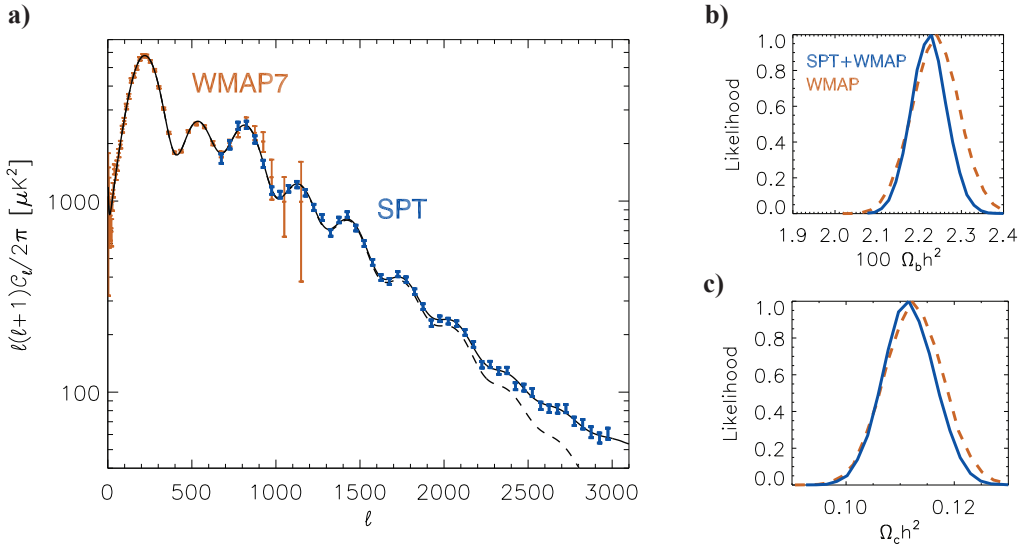


Figure 1.7: a) Power spectrum of the CMB temperature anisotropies measured by WMAP and SPT. The full sky-coverage of WMAP provides the tightest constraints on the power spectrum at large angular scales (small l), while the high angular resolution of SPT allows it to constrain the power spectrum at small angular scales (large l). The dashed line shows the contribution from the best-fit model of the primary CMB anisotropies, indicating that secondary anisotropies from distortion of the CMB by objects between us and the surface of last scattering are dominating at the smallest angular scales. b) Constraints on the baryon density, Ω_b, from CMB observations. The CMB constraints are degenerate with the dimensionless Hubble parameter, $h = 0.702 \pm 0.014$, which can be constrained from complementary cosmological measurements [25]. c) Constraints on the density of cold dark matter, Ω_c. *Figure adapted from Keisler et al. [9]*

Figure 1.6 shows maps of the CMB temperature anisotropies measured by the Wilkinson Microwave Anisotropy Probe (WMAP) satellite and South Pole Telescope (SPT). The power spectrum of the primary temperature anisotropies in the CMB determined from these measurements is shown in Fig. 1.7a. Prior to recombination, the photons and baryons were tightly coupled and underwent a series of acoustic oscillations as overdensities that began to gravitationally collapse felt a restoring force from the increased photon pressure. Non-baryonic dark matter did not feel the photon restoring force, and overdensities were able to collapse freely after the universe became matter dominated, providing the gravitational wells that drove the acoustic oscillations of the baryon-photon fluid. These oscillations manifest themselves as a series of "acoustic peaks" in the power spectrum shown in Fig. 1.7a.

The physics underlying these acoustic oscillations is well understood, and detailed modeling of the power spectrum can be used to constrain fundamental cosmological parameters,

such as Ω_b and Ω_c, as shown in Fig. 1.7b–c [9, 26]. In particular, the relative heights of the even and odd acoustic peaks constrain the ratio of baryonic matter to dark matter since the amplitude of the compressions and rarefactions are sensitive to the driving term from the dark matter gravitational wells as well as the inertia of the fluid set by the baryon density. When combined with other cosmological data, these observations provide the most precise determination of the matter composition of the universe and clearly demonstrate that the dominant matter constituent is nonbaryonic dark matter.

1.2.2 Big-bang nucleosynthesis

An independent handle on the density of baryonic matter can be determined from the primordial abundances of the light elements created during big-bang nucleosynthesis (BBN). Within a few minutes after the big bang, the universe had cooled sufficiently that light nuclei could be formed. The resulting primordial abundances of D, ^{3}He, ^{4}He, and ^{7}Li are a sensitive probe of conditions during this time. In particular, the time at which weak interactions become inefficient ("freeze out") due to the expansion of the universe determines the relic abundances of protons and neutrons, which can be used to synthesize these elements. The reaction rates can be determined from the known formation cross sections and the baryon number density, n_b. In addition, the freeze out time is determined by the universe's expansion in the regime where the energy density was radiation dominated, so the abundances are also sensitive to the photon density. The resulting light-element abundances are typically parametrized using the ratio of densities of the baryons and photons, $\eta = n_b/n_\gamma$. Since n_γ is well constrained by measurements of the CMB, the primordial abundance of the light elements directly constrains Ω_b.

The primary challenge in constraining Ω_b through BBN is to find an astrophysical system in which the abundance of a particular light element has not been altered from the primordial abundance by astrophysical processes. Deuterium typically provides the most precise constraints since it is not readily created in stellar nucleosynthesis. Recent measurements of D from Lyman-α absorption of light from background quasars in high-redshift, metal-poor systems indicate a D abundance of $\log(\mathrm{D/H}) = -4.56 \pm 0.40$, corresponding to a baryon density of $\Omega_b h^2 = 0.0213 \pm 0.0010$ [28]. This measurement dominates the best-fit BBN prediction, and gives good agreement with the values inferred from the CMB power spectrum, as shown in Fig. 1.8.

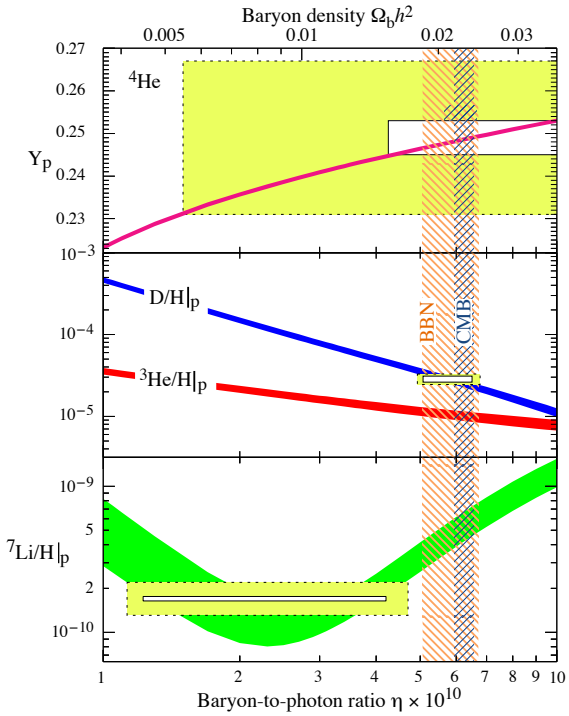


Figure 1.8: Predicted abundances of the light elements versus η from BBN. Colored bands denote the theoretical predictions, while the filled boxes summarize the current best measurements with corresponding statistical (white) and systematic (yellow) errors. The best-fit constraints from BBN are in good agreement with the value of Ω_b inferred from CMB observations, as shown by the vertical hatched bands. *Figure from Nakamura et al. [27]*

Although good agreement with CMB constraints is obtained for D and ^{4}He, there is significant discrepancy between measurements of the ^{7}Li abundance and the predicted abundance in the best-fit model, as shown in Fig. 1.8. Recent data has only served to increase this discrepancy [29]. Determining whether this mismatch is a result of unaccounted for systematics in the measurements of abundances or cross sections, or whether it could be evidence for new physical processes active during BBN is a key open question.

1.2.3 Structure formation

The final cosmological argument for the need for dark matter arises from the growth of structure from the nearly uniform distribution of baryonic matter imprinted on the CMB to the wide variety of structures observed today. Structure formation proceeds by the gravitational collapse of primordial overdensities in the matter distribution (seeded, e.g., by inflation). At late times the process is highly nonlinear and must typically be modeled by

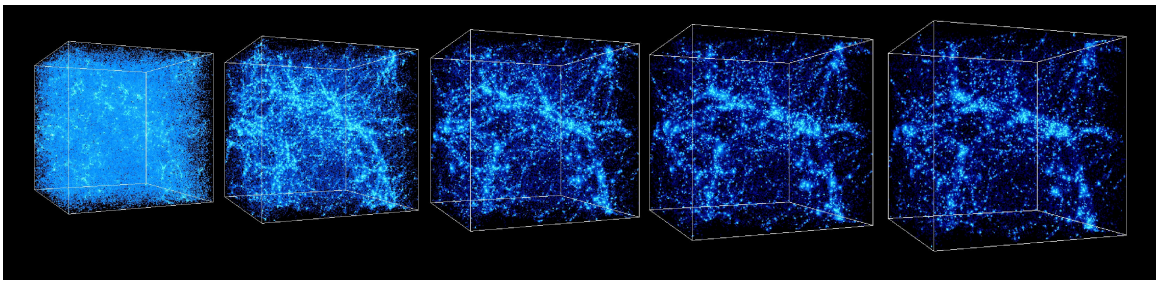


Figure 1.9: Numerical simulation of the formation of large-scale structure from $z = 10$ (leftmost) to $z = 0$ (rightmost). The much more uniform distribution of matter present at early times gradually collapses into overdensities forming the structures seen today. The box size is 43 Mpc, so only the largest structures and galaxy clusters are resolved in the image. *Figure from Kravtsov and Klypin [31]*

large numerical simulations [30], an example of which is shown in Fig. 1.9.

As discussed in Sec. 1.2.1, prior to recombination baryons were prevented from falling into gravitational wells formed by matter overdensities by the restoring force of the photon fluid. However, simulations indicate that observed baryonic density fluctuations fixed by the CMB anisotropies to be of order 10^{-5} at the time of recombination were too small to account for the amount of structure observed today. For these structures to form requires a significant amount of dark matter, which can fall into these overdensities prior to recombination, creating gravitational wells that the baryons can later collapse into after recombination.

As shown in Fig. 1.10, measurements of the matter power spectrum from large optical surveys such as SDSS [33] or the 2dF Galaxy Redshift Survey (2dFGRS) [32] can be used to constrain the total matter density, giving good agreement with the dark matter and baryon densities inferred from CMB observations. The amplitude and shape of the power spectrum provide clear evidence for a nonbaryonic dark matter density. Furthermore, the majority of this dark matter must be "cold," i.e., nonrelativistic throughout the formation of structure to prevent dark matter from escaping overdense regions, which would lead to a corresponding reduction in small scale power.

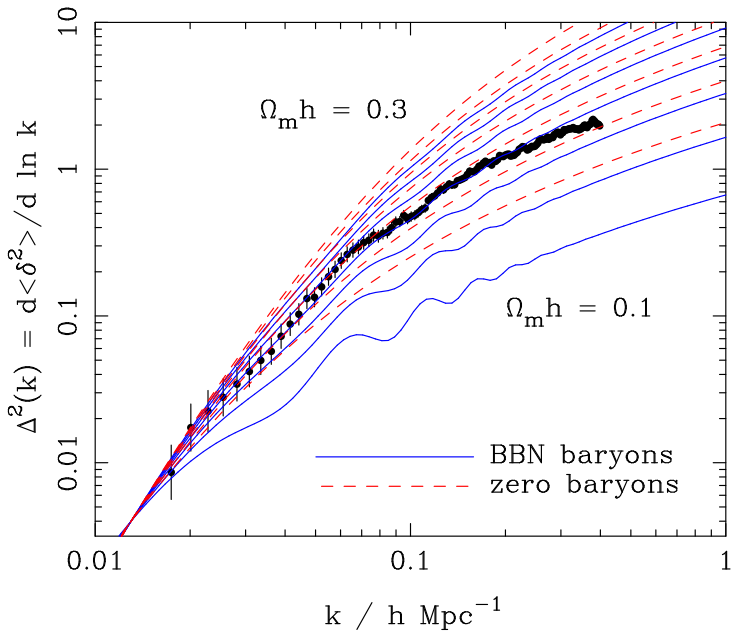


Figure 1.10: Comparison of the galaxy power spectrum from the 2dFGRS [32] to the expected power spectrum when the total matter density, Ω_m, is varied. The blue curves include baryons with $\Omega_b h^2 = 0.02$, which reduces the expected power relative to curves with the same Ω_m but no baryons (red, dashed) since the dark matter density is reduced. The amplitude of baryon acoustic oscillations in the matter power spectrum also increases as baryons become a larger fraction of the total mass due to their increased inertia. The data indicate $\Omega_m h \approx 0.2$ with a baryon fraction consistent with that produced by BBN, in good agreement with the values from CMB constraints. *Figure from Nakamura et al. [27]*

1.3 Candidates

1.3.1 Basic properties

Given the extensive astrophysical and cosmological evidence for dark matter, the next obvious question to address is what is its nature. From the observations described in Secs. 1.1–1.2 we can infer several generic properties that any dark matter candidate must have:

- **Dark matter must be abundant:** Astrophysical measurements indicate that the relic dark matter density accounts for ≈85% of the total matter density in the universe. Any dark matter candidate must be able to be produced with the correct abundance to account for this relic density.

- **Dark matter must be primarily "cold":** Comparisons of simulations of structure

formation to the observed matter power spectrum and the detailed pattern of acoustic peaks in the CMB indicate that the majority of dark matter must have been nonrelativistic ("cold") at the epoch where the universe moved from radiation domination to matter domination, and structure could begin to form. From detailed measurements of the small-scale power observed in the CMB and LSS, we can exclude neutrinos and other relativistic species from being the dominant component of dark matter (although such "hot" dark matter can contribute several percent of the total dark matter density).

- **Dark matter must be nearly dissipationless:** The size of dark matter halos indicates that they cannot efficiently cool by radiating photons and collapsing to form the dense structures seen in baryonic matter.

- **Dark matter must be nearly collisionless:** Constraints on the self-interactions of dark matter from objects such as the Bullet cluster and from the morphology of dark matter halos indicate that if dark matter possesses any interactions other than gravitational, they must be very weak. In addition, interactions via electromagnetism or the strong nuclear force are excluded by terrestrial searches for anomalously heavy nuclei [34, 35].

- **Dark matter must be primarily nonbaryonic:** The CMB and LSS structure require that the dark matter must have begun to fall into overdensities prior to recombination, while baryons were still tightly coupled to the photon fluid. In addition, BBN synthesis places tight constraints on the baryonic content of the universe, and any new dark matter candidate cannot dramatically alter the synthesis of light nuclei.

- **Dark matter must be stable:** Since a significant amount of dark matter remains today it must not readily decay on cosmological time scales.

These properties rule out baryons or the known neutrinos (and in fact any Standard Model particles) from constituting the dominant constituent of the dark matter. Thus, evidence of dark matter from astrophysics and cosmology is perhaps the strongest direct observational evidence for physics beyond the Standard Model. Many candidates have been proposed to account for the dark matter, with the most popular typically motivated by independent arguments from particle physics. In the following sections we briefly review

the most popular candidates, for which a large number of experimental efforts are now searching, as will be described in Chap. 2. However, as discussed in Sec. 1.3.5, a wide variety of candidates consistent with the constraints described above are possible, including those which are not easily accessible with current experimental techniques.

1.3.2 WIMPs

Weakly Interacting Massive Particles (WIMPs) are the candidate for which the largest number of experimental efforts are currently searching. They are a particularly well-motivated dark matter candidate since if there existed a new stable particle at the weak scale, it would be thermally produced in the early universe with the correct relic abundance to account for the observed dark matter density today. In addition, theories of physics beyond the Standard Model which address the "hierarchy problem"—i.e., why the Higgs mass (and by extension the mass of all Standard Model particles) is not pulled up to the cutoff scale by quadratic divergences in its radiative corrections—naturally produce new particles at the weak scale that cancel these divergences and provide excellent WIMP candidates. Finally, interactions from relic WIMPs are accessible to direct searches with existing detector technologies, leading to a large effort to detect WIMPs or constrain their properties.

Figure 1.11 reviews the canonical thermal production and freeze-out mechanisms for WIMPs [8]. In the early universe, when the temperature was much higher than the mass of the WIMP, $T \gg m_\chi$, creation and annihilation processes were in thermal equilibrium, and the comoving number density of WIMPs, n_χ, was constant. As the universe cooled and the temperature fell below m_χ, creation processes became inefficient, and the number density began to annihilate away exponentially following the Boltzmann factor, $n_\chi \sim e^{m_\chi/T}$. However, in an expanding universe, at some point the annihilation rate, $\Gamma_{ann} = \langle \sigma_{ann} v \rangle n_\chi$ falls below the Hubble expansion rate, H, and annihilation processes become inefficient. Here σ_{ann} is the WIMP annihilation cross section and v is the relative velocity of annihilating WIMPs. At this point, the relic density of WIMPs, n_{eq}, "freezes out" and provided the WIMP (or in theories with multiple species, the lightest WIMP) is stable on cosmological time scales, this density remains today.

The annihilation cross section, σ_{ann}, determines the relic abundance, n_{eq}. This abundance is exponentially sensitive to the cross section, with higher cross sections corresponding to lower relic abundances since the particles can efficiently annihilate for a longer pe-

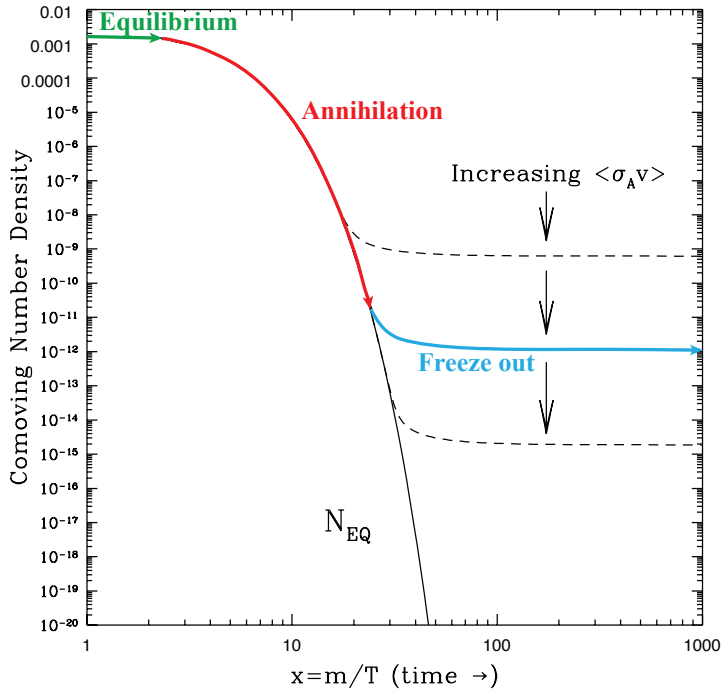


Figure 1.11: Evolution of the comoving WIMP number density versus time. In an expanding universe, the number density freezes out once the expansion rate overtakes the annihilation rate, leading to a relic WIMP density that can account for the observed dark matter density. *Figure adapted from Kolb and Turner [8]*

riod, delaying freeze out. An approximate calculation of this relic abundance indicates that [8, 36, 37]:

$$\Omega_c h^2 \approx \frac{3 \times 10^{-27} \text{ cm}^3 \text{ s}^{-1}}{\langle \sigma_{ann} v \rangle} \tag{1.1}$$

For the observed relic abundance from cosmological measurements, $\Omega_c h^2 \approx 0.1$ [9], this corresponds to annihilation cross sections $\sigma_{ann} \sim 10^{-37}$ cm^2 = 0.1 pb, where we have used that the freeze out temperature is $T_{fo} \approx m_\chi/20$ implying a typical velocity at freeze out $v_{fo} = \sqrt{3T_{fo}/2m_\chi} \approx 0.3c$ [36, 37]. Thus the observed relic density is reproduced for cross sections $\sigma_{ann} \sim 0.1$ pb and masses $m_\chi \sim 100$ GeV, typical of weak scale interactions. Given the need to stabilize the electroweak scale against quadratic divergences in the Standard Model, such a coincidence provides strong motivation for a connection between the dark matter particle and new physics at the weak scale.

1.3.2.1 Supersymmetry

Supersymmetry (SUSY) is one of the most-studied extensions to the Standard Model which addresses the hierarchy problem and provides natural WIMP candidates [27, 36, 38]. Supersymmetric theories extend the symmetry properties of spacetime in ordinary quantum field theory to provide a link between bosons and fermions. For each Standard Model particle, this leads to a corresponding "superpartner" with the opposite spin-statistics. Since fermions and bosons contribute radiative corrections to the Higgs mass of opposite signs, these superpartners naturally cancel the quadratic divergences and stabilize the electroweak scale. At the same time, the lightest of these new supersymmetric particles (LSP) provides a stable WIMP candidate, provided a new quantum number carried by the superpartners (typically denoted as "R-parity") is conserved. Since superpartners have not yet been observed, if supersymmetry is a true symmetry of nature it must be broken at some level. The stability of the electroweak scale can still be maintained provided the superpartner masses are less than a few TeV.

The Minimal Supersymmetric Standard Model (MSSM) denotes the extension to the Standard Model incorporating supersymmetry that adds the fewest possible number of new superpartners [27,38]. The standard naming convention for these new particles is that scalar superpartners of the Standard Model fermions acquire a preceding "s," while the fermionic superpartners of the scalar and vector bosons acquire a trailing "ino." Thus, the MSSM contains 2 new sfermions corresponding to both chiralities of the standard model fermions, a Majorana gaugino for each standard model gauge boson, a second Higgs doublet (required for consistency, in addition to the Standard Model Higgs) and the corresponding higgsinos.

As discussed above, although supersymmetry must be broken, this breaking must be "soft" to maintain the resolution of the hierarchy problem and prevent quadratic divergences [39]. Without detailed knowledge of the symmetry breaking mechanism, it must be parameterized phenomenologically by additional free parameters describing the sfermion masses, the gaugino masses, and their couplings, leading to more than 100 free parameters, even for the MSSM which contains the minimal number of new fields. Given this large parameter space, phenomenological studies are often carried out in a restricted parameter space known as the constrained MSSM (cMSSM or mSUGRA) (e.g., [40–45]). In this model, universal masses and couplings are assumed at the grand unification (GUT)

scale, limiting the new free parameters to a single scalar mass, M_0, a single gaugino mass, $M_{1/2}$, and a common trilinear coupling, A_0. Radiative electroweak symmetry breaking is assumed, giving two additional free parameters in the Higgs sector as the ratio of the Higgs vacuum expectation values, $\tan(\beta) = v_2/v_1$, and the Higgs mixing parameter, μ [27,38]. In the cMSSM, the LSP is typically the lightest neutralino, which is some linear combination of the bino, neutral wino, and two neutral higgsinos. Sneutrinos or gravitinos also provide neutral superpartners which could in principle be the LSP, but sneutrinos have been directly excluded by the earliest direct detection experiments [36], while gravitinos would have interaction cross sections far below those determined by thermal production, preventing their direct detection [46, 47] (such models are described as "superWIMPs" in Sec. 1.3.5).

The assumptions made in the cMSSM are highly idealized and in some cases not particularly well-motivated. Even if supersymmetry is realized in nature, the cMSSM is unlikely to be the correct description. Nonetheless, the detailed phenomenology of this model has been extensively studied due to its accessibility relative to the much larger parameter space of the unconstrained MSSM. The most natural portion of the cMSSM parameter space which would lead to the correct relic density involves neutralinos primarily annihilating to sleptons with masses $\lesssim 80$ GeV/c^2 [48]. However, such models were ruled out by early searches at LEP2 [49–51]. The remaining regions of parameter space typically achieve the correct relic density through enhanced neutralino annihilation due to resonances or coannihilation with other superpartners close in mass [27].

With recent data from the LHC [52, 53] and direct detection experiments such as XENON100 [54], the remaining cMSSM parameter space has been severely constrained. Recent global fits incorporating the LHC SUSY constraints from 1 fb^{-1} of data [52, 53], the Higgs constraints from 5 fb^{-1} of data [55], and the WMAP 7 yr data [26] are shown in Fig. 1.12. These fits indicate that only the portions of the stau coannihilation region (where the stau and LSP have similar masses, leading to resonant enhancement of the annihilation rate and allowing the relic density constraints to be met) at high mass remain favored at the 95% C.L., given all experimental data [40]. In addition, this region requires a relatively light Higgs mass, and if the recent hints that the Higgs mass is 125 GeV [55] are confirmed, the entire best-fit regions in the cMSSM shown in Fig. 1.12 will be disfavored [40–43].

While the cMSSM is widely studied, it should be stressed that constraints can be significantly relaxed in different regions of parameter space of the MSSM, or in simple extensions

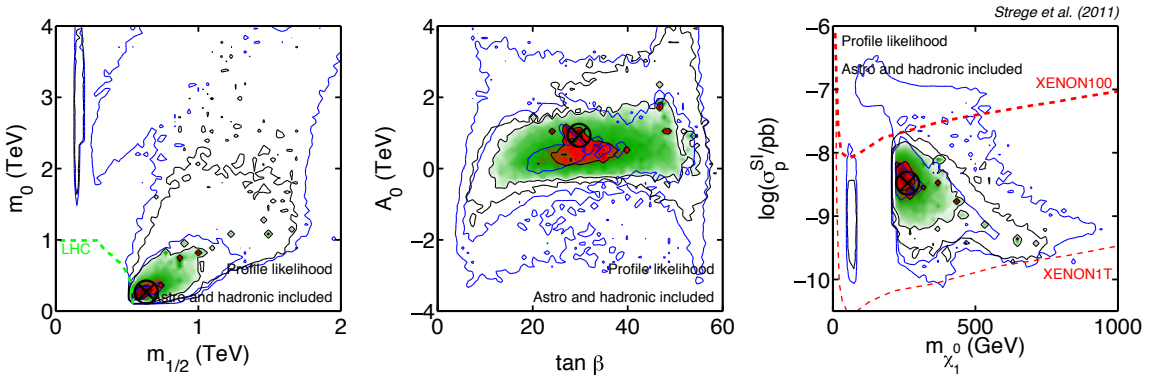


Figure 1.12: Best fit to the cMSSM parameter space including current accelerator, direct detection, and cosmological constraints using a profile likelihood method. The black contours indicate the 68%, 95%, and 99% allowed regions, while the blue contours show the result when direct detection data is not included. Overplotted are the approximate exclusions from the LHC and XENON100 data, and future projections for the reach of XENON1T. *Figure from Strege et al. [40]*

to it. In particular, this thesis will primarily focus on a region of parameter space at low mass ($m_\chi \sim 10$ GeV/c^2), which has been excluded in the cMSSM since LEP, but is of significant recent interest due to possible experimental signatures reported by several experiments (described in Sec. 2.1.4). Since nature is often more complicated than we initially envision, significant additional data will be required before supersymmetric WIMPs can be entirely excluded.

1.3.2.2 Non-supersymmetric candidates

Supersymmetry is not the only extension to the Standard Model which can address the hierarchy problem. A variety of models which typically incorporate a new symmetry at the TeV scale provide a natural dark matter candidate charged under this symmetry while allowing the freedom to cancel the Higgs mass divergence. Some of the most studied alternatives introduce extra spatial dimensions rather than supersymmetry to solve the hierarchy problem. In these models (first proposed by Kaluza [56] and Klein [57]), extra dimensions of size R are introduced and a tower of Kaluza-Klein states with $E \sim n/R$ for $n = 0,1,2,...$ corresponding to quantized momentum in the extra dimensions exist. In certain scenarios (e.g., [58–60]), the lightest Kaluza-Klein state (LKP) is stable and presents a viable WIMP candidate. Some extra-dimensional frameworks can also be used to address the hierarchy problem [59–62]. A large parameter space for WIMPs in extra-dimensional mod-

els exists, although colliders and direct detection experiments have begun to probe these models [63–65].

Additional proposed WIMP candidates include Little Higgs models [66], mirror dark matter [67], "minimal" extensions to the Standard Model adding scalar singlets [68], scalar doublets [69], and fermionic states [70], among others. There has also been significant recent interest in "asymmetric" dark matter models, in which the order of magnitude similarity between Ω_c and Ω_b is naturally accounted for since the relic dark matter density is determined by the baryon asymmetry of the universe [71]. Several of these models can accommodate WIMPs with masses $\sim$10 GeV/c^2 and will be discussed further in Sec. 1.3.3 below. A more detailed review of the non-supersymmetric WIMP candidates mentioned above as well as other proposed candidates can be found in [72].

1.3.3 Low-mass dark matter

Although some of the earliest models proposed WIMP masses as light as a few GeV [73–75], interest in light WIMPs was reduced as the simplest models were excluded by experimental data. In the cMSSM in particular, WIMPs with masses $\lesssim$40 GeV have been ruled out since LEP [49–51], with the recent LHC data and direct detection data described above increasing the lower limit to above 200 GeV. Motivated by the recent experimental results indicating possible evidence for a WIMP with $m_\chi \sim 10$ GeV described in Sec. 2.1.4, there has been significant theoretical interest in lower-mass supersymmetric WIMPs. If some of the assumptions of the cMSSM are relaxed, initial studies indicated that WIMPs as light as 6 GeV could be accommodated in the MSSM [76–78]. More recent studies incorporating Tevatron constraints indicated that such models would have difficulty accounting for cross sections $\gtrsim 5\times10^{-42}$ cm^2 [79–81], while those needed to account for the possible experimental signals under standard assumptions would be an order of magnitude larger [82]. In addition, recent updates to the scans of the MSSM parameter space from [76, 77] including LHC constraints raise the lower limit on the neutralino mass to 18 GeV [83]. Finally, light neutralinos in the MSSM typically produce a relic density which is too large to be compatible with cosmological measurements. In particular, for a $\sim$10 GeV neutralino in the MSSM, there is an approximate inverse relation between the relic density and scattering cross section [81], so pushing to the larger cross sections needed to explain experimental results would produce a relic density which cannot be reconciled with constraints from cosmology.

Although it is difficult to accommodate a neutralino WIMP with mass $m_\chi \sim 10$ GeV and spin-independent scattering cross section $\sigma_{SI} \gtrsim 10^{-41}$ cm^2 in the MSSM given current accelerator and cosmological constraints, numerous models beyond the MSSM exist in which a WIMP with such properties can arise. Fitzpatrick et al. have performed a relatively model-independent analysis of the relevant operators for connecting a low-mass WIMP to standard model particles [84]. They find that adding a ∼10 GeV scalar dark matter particle with scalar interactions and effective couplings to Standard Model equal to those needed to obtain a cross section compatible with experimental results naturally produces the correct relic abundance. In addition, a new scalar with vector interactions or a fermion with either scalar or vector interactions are also possible, although additional interactions are required to obtain the proper relic density in these scenarios [84]. Such interactions can be implemented in a variety of models. In the following sections, we briefly describe some of the most popular models which can provide WIMP candidates with mass ∼10 GeV.

1.3.3.1 Singlet scalars

The singlet scalar model provides the simplest possible extension to the Standard Model that includes a viable candidate for nonbaryonic dark matter [68, 85, 86]. In this model a single additional field is introduced which transforms as a singlet under the Standard Model gauge group, leading to 3 new free parameters: the mass of the new scalar, m_S, its self coupling, λ_S, and its coupling to the Higgs field, λ_L. In the small mass limit ($m_S \ll 100$ GeV) the annihilation cross section is $\sigma_{ann} v \propto \lambda_L^2 m_S^2 / m_h^4$ [68], and the coupling, λ_L, can be tuned to give the correct relic density for a wide range of WIMP masses, m_S.

The dependence of the spin-independent scattering cross section, σ_{SI}, at low mass is shown in Fig. 1.13. For WIMPs with masses from 5–10 GeV, the measured relic density of $\Omega_c h^2 \approx 0.1$ [9] can be obtained for cross sections from 10^{-41}–10^{-40} cm^2, as would be needed to explain the potential experimental signals described in Sec. 2.1.4. Although purely phenomenological, even this simple addition of a single scalar dark matter particle could naturally allow the correct scattering cross section and relic density needed for light-WIMP models consistent with experimental results [84, 87]. Extensions to this model that add additional fields in the dark sector (e.g., an additional Dirac fermion dark matter particle which couples to the Standard Model through the singlet scalar [88]) have also been proposed, allowing additional phenomenological freedom.

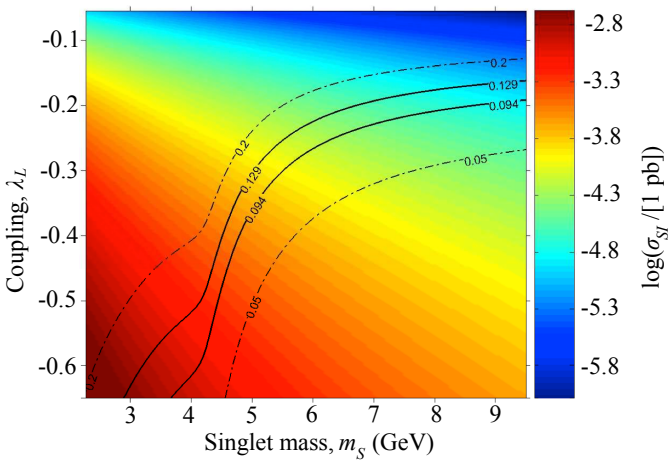


Figure 1.13: Spin-independent elastic scattering cross section, σ_{SI} (in pb = 10^{-36} cm^2), as a function of m_S and λ_L in the singlet scalar model assuming $m_h = 120$ GeV. Since $\sigma \propto 1/m_h^4$, the expected cross sections would be 20% lower if the results indicating that $m_h = 125$ GeV are confirmed. The black contours show the corresponding values of $\Omega_c h^2$. *Figure from Andreas et al. [87]*

1.3.3.2 Next-to-minimal supersymmetric standard model

As discussed in Sec. 1.3.3, it is difficult to accommodate a light neutralino in the MSSM while avoiding overproduction relative to the known relic density. This is primarily because most annihilation channels for a light neutralino are restricted since the masses of the sparticles, sleptons, and Higgs are constrained by colliders to be $\gtrsim$100 GeV [89]. However, in the next-to-minimal supersymmetry standard model (NMSSM) [90], or in other models which allow an additional scalar or pseudoscalar light Higgs consistent with collider constraints (e.g., [81, 91–93]), annihilation through this additional channel may allow relic density constraints to be met [81, 84, 89, 94].

The NMSSM extends the MSSM by adding an additional Higgs singlet. The lightest neutralino can then include a component from the superpartner of the singlet Higgs (the "singlino"). Scans of the NMSSM indicate that a light neutralino that is primarily bino, or a mixture of bino and singlino, can meet relevant collider constraints [84, 89]. However, in the strict NMSSM, the maximum cross sections that can be obtained given collider and cosmological constraints are generally at least an order of magnitude too small to explain the potential experimental signals [95], although sufficiently large cross sections can be obtained for limited regions of the parameter space [96]. More general extensions to the

NMSSM have also been proposed to allow a mostly singlino neutralino for which these cross sections can be obtained without significant fine-tuning [81, 97].

1.3.3.3 Asymmetric dark matter

Asymmetric dark matter describes a class of models in which the dark matter relic density is determined by the baryon asymmetry of the universe, rather than by thermal generation [98–100]. In the standard cosmological scenario, the relic abundance for baryons is fixed by a small baryon-antibaryon asymmetry of order 10^{-10} generated at early times [101]. This model is strongly supported by the excellent agreement of the observed baryon density with the predictions of BBN described in Sec. 1.2.2. In contrast, dark matter is typically assumed to be thermally produced (see Sec. 1.3.2), with a cross section and mass at the weak scale giving the proper relic abundance. Since the baryon relic density and dark matter density are generated by completely independent mechanisms, in this picture it is simply a coincidence that they agree within a factor of a few, $\Omega_c \approx 5\Omega_b$.

In contrast, in asymmetric dark matter models, the dark matter is assumed to be charged under quantum numbers related to baryon number, with effective interactions at early times which transfer the baryon asymmetry to the dark matter sector [71,102–107]. These effective interactions freeze out at high temperature, remaining fixed until the temperature falls below the mass of the dark matter particle and the symmetric dark matter component annihilates away, leaving only the remaining asymmetric component. Since the number densities of baryons and dark matter are approximately equal following this process, $n_b \sim n_c$, the relic dark matter density is determined by the ratio of masses:

$$\Omega_c = \frac{m_\chi}{m_p}\Omega_b \tag{1.2}$$

where m_χ is the dark matter mass and $m_p \approx 1$ GeV is the nucleon mass. For $m_\chi \sim 5$ GeV, the relic density constraints are naturally met, with the precise mass depending on the specific model considered. Although a mass in the 5–15 GeV range is generically predicted by these models, the scattering cross section relevant to direct detection is experiments is model dependent. Many such models have been proposed to obtain the cross sections $\sim 10^{-40}$ cm^2 needed to explain the possible experimental signatures (e.g., [71, 84, 106]).

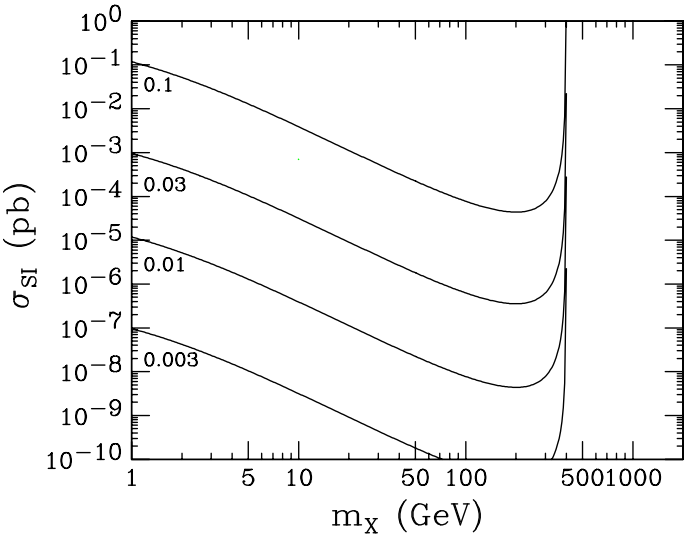


Figure 1.14: Spin-independent χ-proton elastic scattering cross section for an example WIMPless dark matter model with connector mass $m_Y = 400$ GeV. By varying the Yukawa coupling to the Standard Model, a large range of cross sections and masses can be obtained while maintaining consistency with relic density constraints. *Figure adapted from Feng and Kumar [108]*

1.3.3.4 WIMPless dark matter

WIMPless dark matter models [108–110] are models in which the correct thermal relic density is obtained for particles without weak interactions. The general motivation for such models is the observation that the relic density is given by [108]:

$$\Omega_\chi \propto \frac{1}{\langle \sigma_{ann} v \rangle} \sim \frac{m_\chi^2}{g_\chi^4} \tag{1.3}$$

where m_χ is the characteristic dark matter mass and g_χ is its characteristic coupling, which simply follows from dimensional analysis. As discussed in Sec. 1.3.2, for weak scale masses and couplings, the relic density is naturally reproduced. However, WIMPless models assume a dark matter particle without weak interactions, but for which the ratio m_χ^2/g_χ^4 is the same. Since electromagnetic and strong interactions are also ruled out for any dark matter candidate (as discussed in Sec. 1.3.1), this requires the dark matter to lie in a "hidden sector," which does not directly couple to the Standard Model.

Supersymmetric models with gauge-mediated supersymmetry breaking [111, 112] offer an example of such a model, where the structure of the hidden sector naturally enforces the

correct relic abundance set by m_χ^2/g_χ^4, regardless of the dark matter mass [108]. Thus, dark matter masses ∼5–10 GeV are easily accommodated in WIMPless models. Such models would be extremely hard to directly detect unless there is an additional connecting sector which contains fields charged under both the MSSM and the hidden sector. However, if such fields are added, a wide variety of direct detection signatures can be produced [108]. Figure 1.14 shows an example of one such model, where a connector sector containing a single left-handed doublet and right-handed singlet of fermions transforming under the Standard Model weak isospin SU(2) are introduced. For a connector mass, $m_Y = 400$ GeV, and a dark matter mass, $m_\chi \approx 5$ GeV, cross sections $\sim 10^{-40}$ cm^2 can be obtained for a Yukawa coupling, $\lambda_u \approx 0.03$.

1.3.3.5 Mirror dark matter

Mirror dark matter models [67, 113–115] propose that a second identical copy of the entire Standard Model exists and is coupled to the known particles Standard Model particles through a new Z_2 ("mirror") symmetry. In some scenarios, such models can address the hierarchy problem [113], with mirror nucleons or mirror electrons providing stable dark matter candidates [115]. Although dark matter candidates with masses ∼5 GeV can easily exist in these models and have been used to fit potential experimental signals at low mass [115–117], one must explain why mirror particles have different cosmological phenomenology than their Standard Model counterparts (i.e., why they are collisionless and dissipationless and can begin to form gravitational potentials prior to recombination). Such explanations typically assume different initial conditions for the mirror partners and Standard Model particles [115] potentially allowing different astrophysical phenomenology between the sectors.

1.3.4 Axions

While the previous discussion has focused on WIMPs, axions present a dark matter candidate which is well-motivated by particle physics but does not address the hierarchy problem. Instead, the axion is a natural solution to the "strong CP" problem, resulting from the stringent experimental constraints on CP violation in quantum chromodynamics (QCD)—most notably from limits on the neutron electric dipole moment [118]—in contrast to the expectation that the QCD Lagrangian naturally contains CP violating terms. The Peccei-Quinn mechanism solves the strong CP problem by introducing a new U(1) symmetry which is

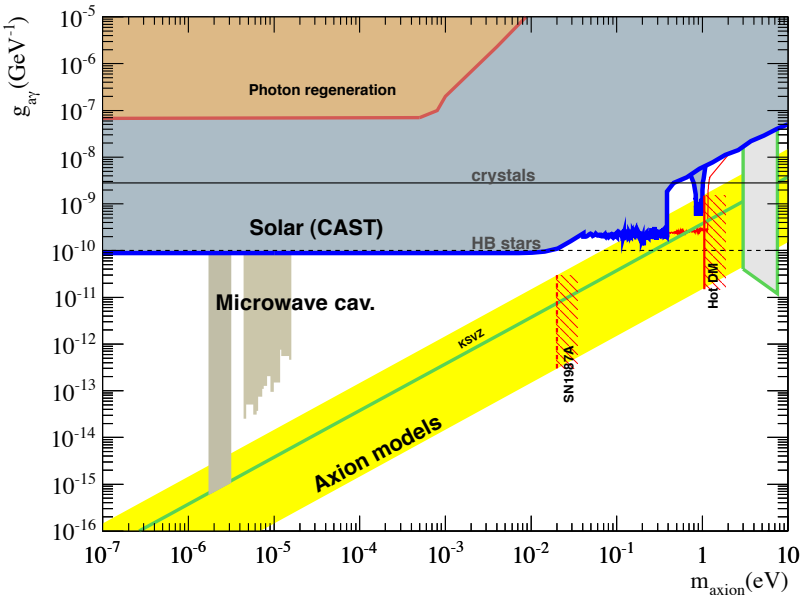


Figure 1.15: Summary of constraints on axion parameter space from various experimental and astrophysical bounds. The yellow band denotes the range theoretical predictions for the dependence of the axion mass on its coupling described by Eq. 1.4 (where $g_{a\gamma} \sim \alpha/2\pi f_a$). Microwave cavity searches (gray) are beginning to limit axion models with $m_a \sim 2$ μeV, while solar axion telescopes (blue) rule out parameter space near $m_a \sim 0.5$ eV. Astrophysical limits from hot dark matter and overcooling of supernovae and horizontal branch stars are also indicated. *Figure from Irastorza et al. [125]*

spontaneously broken, resulting in a minimum where the CP violating terms in the Lagrangian vanish [119, 120]. The axion is the resulting pseudo Nambu-Goldstone boson of the Peccei-Quinn U(1) symmetry. It is described by its mass, m_a, and the axion decay constant, f_a, for which [27]:

$$m_a \simeq (0.6 \text{ eV}) \frac{10^7 \text{ GeV}}{f_a} \tag{1.4}$$

Originally it was assumed that the axion was related to electroweak symmetry breaking, $f_a = O(250 \text{ GeV})$, but such axions have been excluded by accelerator searches [27]. Viable parameter space remains for "invisible" axions, in which $f_a \gg 250$ GeV. Since such axions would be relatively light, cosmological constraints on hot dark matter from the CMB and LSS limit their thermal production and require $m_A \lesssim 1$ eV, similar to neutrino constraints [9, 121, 122]. Nonetheless, a nonthermal population of axions could be produced during "vacuum realignment" after the Peccei-Quinn U(1) symmetry is broken [123, 124], resulting in a condensate of cold axions with a relic density $\Omega_A h^2 \approx 0.1$ for $m_A \approx 10$ μeV.

Experimental searches for axions typically rely on conversion of axions to photons in an applied magnetic field [126]. Microwave cavity searches, which attempt to resonantly convert cosmological axions into photons in a high-Q resonant cavity threaded by a large magnetic filed are beginning to rule out portions of the predicted parameter space if axions make up the majority of the cold dark matter [127]. Extensions to this technology to allow larger masses and lower couplings to be probed are currently in development [128]. Solar axion telescopes [129], which aim to convert axions produced in the sun to x-ray photons in the presence of a large magnetic field are also starting to probe predicted axion models for $m_a \approx 0.1$–0.5 eV. These experimental results and additional astrophysical constraints are summarized in Fig. 1.15. As shown in Fig. 1.15, axions remain a viable, well-motivated candidate for cold dark matter, with current experiments just beginning to probe the expected parameter space consistent with cosmological observations.

1.3.5 Other candidates

Finally, we briefly mention that while the candidates described above are some of the most popular, there are a wide variety of WIMP models meeting the constraints provided in 1.3.1, which would provide significantly different experimental signatures than the models described above. One such possibility, dubbed "superWIMPs," assumes WIMPs freeze out as normal in the early universe but then decay to a lighter, stable particle with significantly weaker interactions [46]. In these scenarios, e.g., SUSY models with a gravitino LSP, WIMPs could have the correct relic density but would be nearly impossible to directly detect.

Another substantially different possibility is that gravity itself should be modified, rather than introducing a new dark matter particle. One such widely studied phenomenological framework is Modified Newtonian Dynamics (MOND) [130], which introduces an acceleration scale, $a_o \approx 10^{-8}$ cm s^{-2}, below which the acceleration due to gravity is modified to reproduce the rotation curves of spiral galaxies. Although this model accounts for the dynamics of spiral galaxies as well as the correspondence between their total baryonic mass and circular rotation velocities at high radius (i.e., the baryonic "Tully-Fisher" relation [131]) with only a single new parameter, a_0, it does not by itself explain gravitational lensing or the detailed structure of the acoustic peaks in the CMB.

To attempt to account for these effects, fully relativistic theories which reduce to MOND in the classical limit have been developed (e.g., [132]). While it is still an open question of

whether such theories can fully satisfy cosmological constraints [133, 134], observations of galaxy clusters are also problematic. The total mass inferred from x-ray, lensing, and SZ observations is difficult to explain in a MOND-like framework [135]. In addition, nonequilibrium objects such as the Bullet cluster show a clear separation between the baryonic matter and gravitational potential. Explanation of such objects requires a significant fraction of the matter to be approximately collisionless, in contrast to the standard MOND framework, although explanations incorporating neutrinos have been suggested [136].

Although the vast majority of observations support the current cosmological paradigm in which nonbaryonic, particle dark matter accounts for roughly 85% of the matter in the universe, given current observational constraints this dark matter could have a wide range of properties. A survey of some of the most popular models has been given in the previous sections, although this summary is by no means comprehensive. While detection of such candidates would be definitive proof of the particle dark matter paradigm, the current lack of detection can only constrain the possible models for dark matter, since as discussed above it is possible to construct reasonable models in which the dark matter is extremely difficult to directly detect by conventional means. As will be described in Chap. 2, a wide variety of experimental techniques will likely be needed to detect dark matter and fully constrain its properties.

Chapter 2

Detection methods

Given the substantial evidence for particle dark matter presented in Chap. 1, it is of great interest to empirically probe the nature and composition of this dark matter. While a variety of possible experimental techniques have been developed, in this thesis we will focus on the direct detection of WIMPs through their elastic scattering in terrestrial detectors (Sec. 2.1). Other techniques, including indirect detection of the presence of dark matter through its annihilation products as well as direct production of dark matter or related particles in accelerators will be discussed briefly in Sec. 2.2.

2.1 Direct detection

Dark matter "direct detection" refers to a wide variety of experimental techniques developed to detect interactions from dark matter in particle detectors on earth. Given the standard assumptions about the distribution of dark matter in the Milky Way discussed in Sec. 2.1.1, for a $\sim$100 GeV/c^2 dark matter particle, we expect a flux of roughly 5×10^4 particles cm^{-2} s^{-1} at earth. Although the scattering cross section for WIMPs is small—by symmetry arguments, it should be roughly related to the annihilation cross section determined from the relic density—the expected flux is sufficient that the WIMP signal in certain models would be measurable in detectors with target masses from the kg to ton scale. The main challenge in such experiments is to develop a technique to identify rare WIMP interactions among a much higher rate of backgrounds from cosmic rays and natural radioactivity, which could mimic a WIMP signal. In Sec. 2.1.1, we discuss the rate and energy spectrum of the expected WIMP signal, while the primary backgrounds are discussed in Sec. 2.1.2. Uncertainties in the scattering rate due to our lack of knowledge of the detailed properties

of the dark matter halo and true WIMP model are discussed in Sec. 2.1.3. Possible positive detections of low-mass WIMPs reported by the DAMA/LIBRA, CoGeNT, and CRESST-II experiments are reviewed in Sec. 2.1.4.

2.1.1 Expected signal

The expected signal in direct detection experiments has been reviewed in multiple places (e.g., [36, 137, 138]). Here we provide a basic overview of the primary characteristics of the expected signal. Since WIMPs are gravitationally bound in the Milky Way halo, their characteristic velocity is set by the gravitational potential of the galaxy and is comparable to other galactic velocities, $v \approx 10^{-3}c$. Accordingly, WIMPs are nonrelativistic and the recoil energy for elastic scatters of WIMPs with a target is easily calculated as:

$$E_r = \frac{\mu^2 v^2}{m_T}(1 - \cos\theta_c) \tag{2.1}$$

where m_T is the target mass, v is the WIMP velocity in the nucleon rest frame, θ_c is the scattering angle in the center of mass frame, and $\mu = m_\chi m_T/(m_\chi + m_T)$ is the reduced mass of the WIMP-target system. Just from these simple kinematics, it is clear that WIMPs with masses from a few GeV to a few TeV cannot efficiently transfer energy to 0.5 MeV electrons, so electron recoils from WIMPs would typically have energies <1 eV and would be extremely difficult to detect in massive detectors. In contrast, the WIMP mass is well-matched to the nucleus mass, m_N, so nuclear recoils can provide significant transfer of the WIMP kinetic energy, $m_\chi v^2/2 \sim$ 1–100 keV. Thus, we expect only a nuclear recoil signature from WIMPs at keV energies, while most backgrounds at these energies are expected to produce electron recoils (see Sec. 2.1.2). This simple calculation also indicates that the minimum velocity that can produce a nuclear recoil of energy E_r is:

$$v_{min} = \sqrt{\frac{m_N E_r}{2\mu^2}} \tag{2.2}$$

The differential WIMP-nucleus scattering rate, typically given in units of counts keV^{-1} kg^{-1} day^{-1} (dru), can be written [36, 137]:

$$\frac{dR}{dE_r} = \frac{\rho}{m_N m_\chi}\int_{v_{min}}^{\infty} v f(v)\frac{d\sigma_{\chi N}}{dE_r}(v, E_r)dv \tag{2.3}$$

where ρ is the local dark matter density and $f(v)$ is its distribution as a function of velocity, v, in the detector rest frame. The differential scattering cross section, $d\sigma_{\chi N}/dE_r$, encodes the particle physics of the WIMP-nucleon interaction and is in general model dependent.

The rate given by Eq. 2.3 directly depends on the WIMP density at the Earth, which together with the WIMP mass and velocity distribution determines the expected flux. For uniformity, most experiments have adopted the canonical value of $\rho = 0.3$ GeV cm^{-3} [137]. This value provides roughly the central value of current measurements, although flattening of the halo or small-scale structure in the vicinity of the Earth could significantly alter this value [139–142]. In the "standard halo model" (SHM) conventionally used to calculate the expected recoil spectra for dark matter experiments, a Maxwellian velocity distribution in the galactic rest frame is assumed, truncated by the galactic escape velocity, v_{esc}, as:

$$f(v) = \begin{cases} Ae^{v^2/v_0^2} & v < v_{esc} \\ 0 & v \geq v_{esc} \end{cases} \tag{2.4}$$

where the normalization factor, A, is chosen to ensure that the integral of the velocity distribution is normalized to unity for a given v_{esc}. Here, v_0 is the characteristic (or most probable) speed, given by the local circular velocity at the radius of the sun. In the SHM, we assume $v_0 = 220$ km s^{-1}. In most recent direct detection searches (especially those sensitive to the high-velocity tail of the distribution, e.g., [143–145]), an escape velocity of $v_{esc} = 544$ km s^{-1} has been used [146]. The Earth is moving through this halo with velocity, $\mathbf{v_E}$, which is the vector sum of the sun's circular and peculiar velocity and the earth's orbital velocity around the sun. The relevant component of the peculiar velocity of the sun is taken to be 12 km s^{-1}. Taking the component of the earth's orbital motion along this direction gives a time dependent velocity relative to the rest frame of the halo:

$$v_E = 232 + 15\cos\left(2\pi\frac{t - t_0}{365.25\text{ d}}\right)\text{ km s}^{-1} \tag{2.5}$$

where $t_0 \approx$ June 2nd. This several percent annual modulation in the relative WIMP velocity seen at earth should produce a corresponding annual modulation in the counting rate, as discussed further in Sec. 2.1.4.

Given the halo model assumed above, we only need to determine the relevant cross

section in Eq. 2.3 to evaluate the expected WIMP rate. For nonrelativistic WIMPs, the differential scattering cross section with nucleons, $d\sigma_{\chi N}/dE_r$, takes on a fairly simple form [36, 147]:

$$\frac{d\sigma_{\chi N}}{dE_r} = \frac{m_N}{2\mu^2 v^2}\left(\sigma_0^{SI} F_{SI}^2(E_r) + \sigma_0^{SD} F_{SD}^2(E_r)\right) \tag{2.6}$$

The energy dependence is encoded by the form factor, $F^2(E_r)$, which describes the dependence of the coherent scattering cross section on the momentum transfer, $q = \sqrt{2m_N E_r}$. Equation 2.6 is separated into a spin-independent (SI) term due to the scalar coupling of χ to quarks, and a spin-dependent (SD) term due to axial-vector couplings, with [36, 148]:

$$\sigma_0^{SI} = \frac{4\mu^2}{\pi}\left[Z f_p + (A-Z) f_n\right]^2 \tag{2.7}$$

$$\sigma_0^{SD} = \frac{32\mu^2 G_F^2}{\pi}\frac{(J+1)}{J}\left[a_p\langle S_p\rangle + a_n\langle S_n\rangle\right]^2 \tag{2.8}$$

where Z, A are the atomic number and mass number of the target nucleus, f_p and f_n denote the model-dependent spin-independent coupling strengths to protons and neutrons, J, is the nuclear spin, and $\langle S_{p,n}\rangle$ are the expectation values of the proton and neutron spin for the nucleus. For the SUSY neutralinos described in Sec. 1.3.2.1, it is typically assumed that $f_p \approx f_n$. However, models for which this is not true have been recently proposed to explain the tension between possible experimental signals and null results from other target nuclei (see Sec. 2.1.4). Vector couplings could also lead to an additional SI component for a WIMP which is, e.g., a Dirac fermion, but they vanish for Majorana WIMPs (as expected for SUSY neutralinos) so we ignore such couplings here [138].

For the spin-independent case, where $f_n \approx f_p$, the nuclear form factor is given by the Fourier transform of the nucleon density, and typically parameterized in terms of the momentum transfer, q, as [137, 149, 150]:

$$F_{SI}^2(E_r) = \left(\frac{3j_1(qr_n)}{qr_n}\right)^2 e^{-q^2 s^2} \tag{2.9}$$

where j_1 is a spherical Bessel function and $s \approx 0.9$ fm is the nuclear skin thickness. The radius parameter is typically taken to be $r_n^2 = c^2 + (7/3)\pi^2 a^2 - 5s^2$ for $c = 1.23A^{1/3} - 0.6$ fm and $a \approx 0.5$ fm [137].

For the spin-dependent case, we can not generally assume that $a_p \approx a_n$ since their ratio is strongly model dependent [148], and the form factor cannot be simply factored

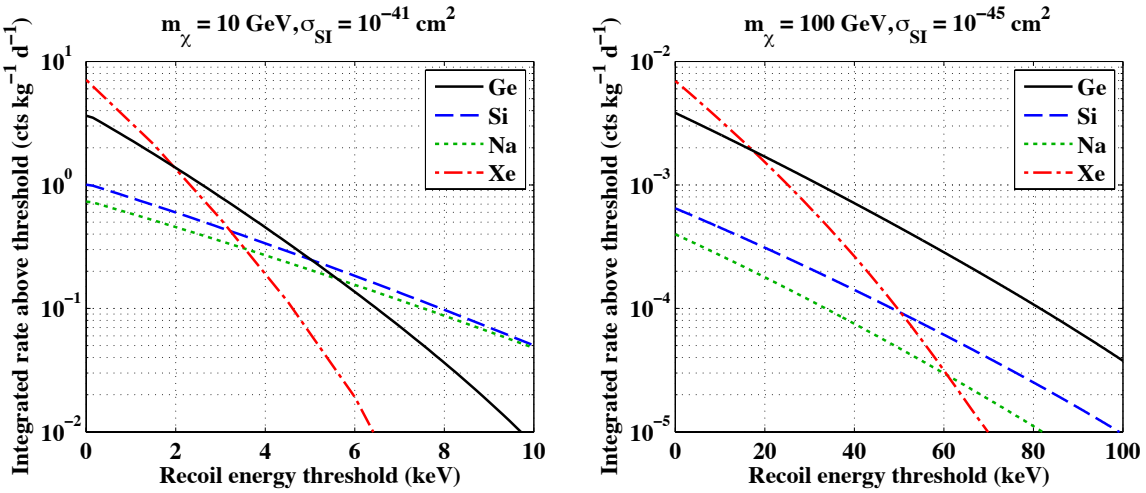


Figure 2.1: Comparison of the integrated WIMP rate (in counts kg^{-1} day^{-1}) as a function of the detector recoil energy threshold and target nucleus. (left) Rate assuming a 10 GeV WIMP, with a cross section close to that needed to explain the experimental results in Sec. 2.1.4. (right) Rate for a 100 GeV WIMP with a cross section just below current detection limits. Spin-independent elastic scattering and the SHM parameters described above are assumed.

out in a model-independent fashion. In this case, it is common to parameterize the form factor, $F_{SD}^2 \equiv S(q)/S(0)$ in terms of isoscalar, $a_0 = a_p + a_n$, and isovector, $a_1 = a_p - a_n$, components, with:

$$S(q) = a_0^2 S_{00}(q) + a_0 a_1 S_{01}(q) + a_1^2 S_{11}(q) \tag{2.10}$$

Here, the parameters S_{ij} depend on the nucleus and must be determined from nuclear structure calculations (e.g., [151, 152]).

From the parameterizations of the cross sections above, we can make several general conclusions. For the spin-independent scattering cross section with $f_p \approx f_n$, then the total scattering rate is proportional to A^2. At low momentum transfer, the WIMP-nucleon scattering does not probe the detailed nuclear structure, and the scattering amplitudes add coherently leading to the A^2 dependence. For spin-dependent scattering, the cross section scales as $(J+1)/J$ instead, leading to larger expected spin-independent sensitivity for most WIMP models and targets. However, models can be constructed in which the spin-independent scattering is significantly suppressed, leading to the need to also explore the spin-dependent parameter space to fully exclude SUSY WIMPs.

Figure 2.1 shows the expected total scattering rate for spin-independent elastic scattering

calculated from Eq. 2.3 for a variety of target nuclei. Two WIMP masses are shown in order to demonstrate the effects of the kinematic suppression of the recoil energy for heavy nuclei at low WIMP mass and the coherent enhancement of the spin-independent scattering rate. For low-mass WIMPs ($m_\chi \sim 10$ GeV), with recoil energy thresholds above ∼5 keV, lighter nuclei (e.g., Na and Si) give higher expected rates due to the better kinematic match between the WIMP and nucleus. If the detector threshold can be pushed below ∼5 keV, then the higher rate of coherent scattering from the heavier nuclei (e.g., Ge and Xe) can lead to improved sensitivity. Although this gives the total WIMP counting rate, the background spectrum, which also typically increases at low energy, must be taken into account to optimize the sensitivity at a given WIMP mass. For higher WIMP masses, the heavier targets can provide an order of magnitude higher rate than the lighter targets shown in Fig. 2.1 at typical recoil energy thresholds of ∼10 keV. Note that the rate for these heavy targets falls off more quickly with energy than the lighter targets due to form factor suppression (at high momentum transfer, the coherent scattering from the entire nucleus breaks down), so low thresholds are especially important for the heaviest nuclei.

2.1.2 Backgrounds

The WIMP rates shown in Fig. 2.1 indicate that even for a detector with a target mass of 100 kg, for the $m_\chi = 100$ GeV WIMP just below current detection thresholds, there would be a WIMP scatter in the detector only once every few days. At the same time, even if the detector is constructed from the most radiopure materials possible and shielded from external radiation, the typical residual background rate will be several orders of magnitude larger than the signal. Thus, direct detection experiments must develop techniques to reject these backgrounds to have any hope of seeing the much lower rate WIMP signal.

As discussed in Sec. 2.1.1, WIMPs are expected to produce only nuclear recoils above the detector threshold due to the kinematics of the scattering process. In contrast, most backgrounds in the few keV to few hundred keV range are due to γs and e^- from natural radioactivity or cosmic rays and will primarily produce electron recoils in the detector. In many detector materials, the energy deposition from an electron recoil or nuclear recoil will produce a different signature, allowing electron-recoil backgrounds to be identified and eliminated. A recoiling electron with $E_r \approx 10$ keV will have a much larger velocity than a nucleus with the same kinetic energy, leading to a sparser deposition of energy in the

detector for the electron recoil relative to the dense deposition from a nuclear recoil. If an experiment can tell the difference between these sparse and dense depositions of energy, most backgrounds can be eliminated on an event-by-event basis (e.g., [54, 153–155]). Neutron-induced nuclear recoils cannot be distinguished from WIMP-induced nuclear recoils in this fashion, so the detector must be shielded from incident neutrons. For this reason, dark matter direct detection experiments must typically be operated underground to reduce the neutron flux from cosmic ray induced particle showers.

Due to their low scattering cross section, WIMPs have negligible probability of interacting more than once in a direct detection experiment. In contrast, high energy γs or neutrons have a substantial probability of scattering multiple times. Detectors which have the ability to reconstruct the interaction location have an additional handle to reject these multiply scattering backgrounds. More significantly, if a large volume detector with negligible internal impurities can be produced, the finite penetration depth of external backgrounds will cause them to primarily interact near the outer regions of the detector volume, while WIMP-interactions will be distributed evenly throughout. By fiducializing to only the central region, excellent rejection of external backgrounds can be obtained. This technique has led to the development of detectors consisting of large volumes of highly pure liquid scintillators (typically liquefied noble gases), which currently provide the strongest constraints on the spin-independent scattering cross section, largely due to this "self shielding" of the inner detector volume [54, 156].

Finally, even if backgrounds cannot be completely eliminated, features of the expected WIMP signal may allow it to be identified in the presence of backgrounds given a sufficiently large number of signal events. As discussed in Sec. 2.1.1, the mean observed WIMP velocity varies throughout the course of the year due to the orbital motion of the earth around the sun. As shown in Fig. 2.2, this leads to an expected WIMP rate which varies by $\sim$5–10%, with a maximum near June 2nd and a minimum near Dec. 1st [157]. Even in the presence of significant unmodulated backgrounds, it may be possible to identify a WIMP signal through this annually modulated signal. As discussed in Sec. 2.1.4, two independent experiments are now claiming such a signal, possibly arising from WIMP interactions.

Detectors which measure the angular direction of a recoiling nucleus may also be able to identify a WIMP signal in the presence of significant backgrounds since the distribution of the recoil direction undergoes a diurnal modulation. Due to the solar system's velocity in

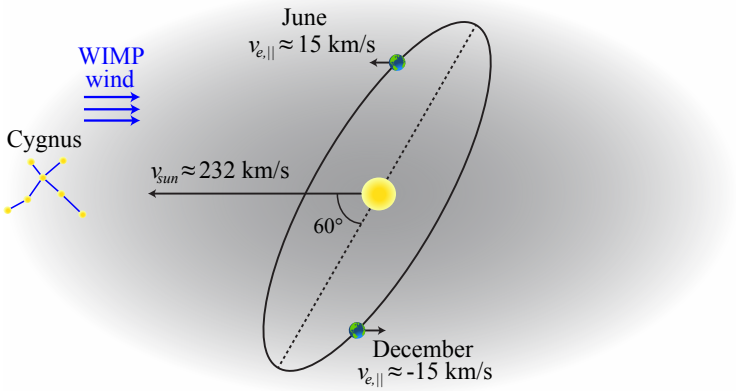


Figure 2.2: Schematic of annual modulation due to earth's rotation around the sun. In the rest frame of the halo, the sun is moving with $v_{sun} = 232$ km s^{-1} in the direction of the constellation Cygnus. The earth's orbital plane is inclined at 60° relative to this motion, while its circular velocity in the rest frame of the sun is $v_e = 30$ km s^{-1}. Thus, the component parallel to the sun's motion is $v_{e,||} = 15\left(2\pi\frac{t-t_0}{365.25\text{ d}}\right)$ km s^{-1}, with a maximum at $t_0 =$ June 2nd. The modulation of the relative velocity of the earth through the dark matter halo will give a corresponding modulation in the WIMP scattering rate in a direct detection experiment.

the rest frame of the halo, the WIMP flux (sometimes called the "WIMP wind") arrives primarily in the direction of our motion, which is towards the constellation Cygnus. A nuclear recoil signal with an angular distribution denoting a flux centered on Cygnus throughout the day would provide a robust signature of WIMPs that would be free from any known backgrounds. However, such detectors typically use gas targets to obtain sufficiently long recoil tracks that the direction can be measured, leading to target masses of only $\lesssim$10 g for the current generation of detectors [158–160]. Given current limits on the WIMP-nucleon scattering cross section, a large scale-up in target mass will be needed before directional detection experiments would be able to detect a WIMP signal.

2.1.3 Uncertainties on the expected signal

Uncertainties in either the WIMP velocity distribution or WIMP-nucleon scattering cross section for different target nuclei will lead to corresponding uncertainties in the expected scattering rate in direct detection experiments. In particular, the standard halo model (SHM) assumed above is a highly idealized version of the expected WIMP velocity distri-

bution. Although the parameters assumed are consistent with current experimental measurements, in some cases significant uncertainties remain. These assumptions typically scale the rate in roughly the same way for all experiments, so comparisons between experiments remain valid provided that all experiments assume the same halo model when calculating results. However, especially for WIMP masses or models where the rate is dominated by the high velocity tail of the distribution (e.g., the light WIMPs described in Sec. 2.1.4 below, or inelastic dark matter models [161]), uncertainties in the astrophysics model can lead to significant uncertainties in the comparison of results for detectors with different energy thresholds and target nuclei.

A variety of "nonstandard" halo models have been studied, including those which soften the unphysical, infinitely sharp escape velocity cutoff from Eq. 2.4 [140], elliptical or nonspherical models [162, 163], and departures from Maxwellian velocity distributions determined directly from N-body simulations [164, 165]. Given the complex nature of the hierarchical structure formation of galactic dark matter halos, detailed simulations (e.g., [166–168]) are required to understand the extent to which residual substructure in the halo and its velocity distribution remains. Such residual structure has been estimated to constitute several percent of the of the local density at earth and be peaked at velocities as much as $\sim$50% larger than v_0 [141]. For sufficiently low WIMP masses, this high-velocity component may allow a higher rate of WIMP scatters to be detected above the energy threshold of a direct detection experiment than would be expected from a standard Maxwellian distribution.

Even neglecting the detailed velocity structure, uncertainties in the canonical parameters assumed in the SHM remain. The local density is determined by modeling of the Milky Way's rotation curve and stellar velocity dispersions. A wide variety of such measurements has been performed (e.g., [140, 142, 169]), with typical values found in the range of $\rho \approx 0.2$–0.4 GeV/cm^3. Thus the average local density at the radius of the sun is only known to $\sim$30% at best, with the possibility of residual modeling systematics common to all methods. Independent methods which do not rely on global modeling of the mass distribution find a spread of values, ranging from those suggesting we may live in a dark matter "void" with negligible local density [170] to a density which is several times higher than the canonical value [171], again with possibly significant systematics. Since the rate in all experiments scales linearly with ρ, such variations should not affect the detailed compar-

ison of experiments, but if a detection is made, an accurate value is necessary for proper comparison to particle physics and cosmological constraints.

Accurate determination of the local circular speed, v_c, and the escape velocity, v_{esc}, is also necessary for detailed predictions of the scattering rate. A combination of recent measurements gives $v_c = 236 \pm 11$ km s^{-1} [172,173], slightly higher than the canonical value assumed in the SHM of 220 km s^{-1} [174]. The escape velocity for recent direct detection work is typically taken to be 544 km s^{-1}, which is the central value of the 90% CL interval, $498 < v_{esc} < 608$ km s^{-1} determined from a sample of high-velocity stars in the RAVE survey [146].

Given the substantial uncertainties in the properties of the halo and velocity distribution, techniques have been developed to make comparisons between direct detection experiments which are independent of the velocity distribution, $f(v)$ [175]. After accounting for the $\frac{d\sigma_{\chi N}}{dE_r} \propto 1/v^2$ dependence of the differential cross section, the scattering rate in Eq. 2.3 depends on $f(v)$ only through its integral:

$$g(v_{min}) = \int_{vmin}^{\infty} \frac{f(v)}{v} dv \tag{2.11}$$

where as above, for elastic scattering, $v_{min} = \sqrt{\frac{m_N E_r}{2\mu^2}}$. This factor $g(v_{min})$ is common to the scattering rate in all experiments, so when comparing rates the astrophysical dependence can be canceled out. For a given WIMP mass, m_χ, a unique mapping exists between the energy range and spectrum with which it will show up in each experiment [175]. If no candidate events are observed, then astrophysics independent limits can be determined in the $g(v)$ versus v plane, in analogy to the usual procedure of calculating limits in the σ versus m_χ plane. The most conservative form that such limits can take assumes a step function for $g(v)$, corresponding to a dark matter velocity distribution which is a δ function at some velocity v.

In addition to astrophysical uncertainties, a wide variety of particle physics models are consistent with current constraints on WIMP properties, as described in Sec. 1.3. This leads to a wide range of possible WIMP cross sections and masses, and possibly different interaction strengths with different target nuclei. An obvious example is models in which spin-independent scattering is suppressed, so that spin-dependent scattering dominates the expected rate. In this case, experiments using odd-proton targets such as ^{19}F or ^{23}Na would

have enhanced sensitivity if WIMP-proton scattering dominates [155, 176, 177]. Similarly, odd-neutron targets such as ^{29}Si, ^{73}Ge, ^{129}Xe, and ^{131}Xe would probe WIMP-neutron spin-dependent scattering [54, 178, 179]. Even in the case of spin-independent scattering, if the couplings to protons and neutrons are not equal (i.e., $f_p \neq f_n$ in Eq. 2.7), then different targets can have significantly different scattering rates. Recently, "isospin-violating dark matter" (IVDM) in which $f_n/f_p \approx -0.7$ [180–182] has been proposed to allow compatibility between the possible dark matter signals and null results described in Sec. 2.1.4 by suppressing the WIMP sensitivity of experiments with Xe targets [156]. Given the theoretical uncertainties in the precise form of the WIMP-nucleon interaction, results from multiple targets are necessary to fully constrain the allowed parameter space.

2.1.4 Light WIMPs

Figure 2.3 summarizes the current state of the constraints on the spin-independent elastic scattering cross section, σ_{SI}, from a number of different experiments. At high WIMP mass, the strongest constraints come from the XENON100 experiment [54], which currently places an upper limit on the spin-independent scattering cross section of $\sigma_{SI} < 10^{-44}$ cm^2 at $m_\chi \sim 100$ GeV, with a number of other experiments providing constraints within a factor of a few. These limits have begun to constrain cMSSM parameter space, with the favored parameter space now lying at $m_\chi \approx 250$ GeV and $\sigma_{SI} \approx 5 \times 10^{-45}$ after including collider constraints (which disfavor low WIMP masses) and direct detection constraints (which disfavor high cross sections).

Recently, there has been significant interest not in this high mass region, but in the region where $m_\chi \approx 10$ GeV and $\sigma_{SI} \approx 5 \times 10^{-41}$ cm^2. As will be discussed in the following sections, the DAMA/LIBRA [187,188], CoGeNT [189,193], and CRESST-II [154] experiments report results that could be compatible with spin-independent elastic scattering from a WIMP in this region of parameter space. However, as shown in Fig. 2.3, these regions only agree within a factor of 2–3, and naively appear to be incompatible with each other, as well as with constraints from XENON10 [156]. Nonetheless, the close proximity of the allowed regions suggests that all experiments may be detecting the same signal, and uncertainties in the astrophysics model [141, 164], detector response [82, 180, 194, 195], or low-energy backgrounds [190–192] may account for the apparent discrepancies. More data will likely be required, both from the experiments claiming signals, as well as those that appear to

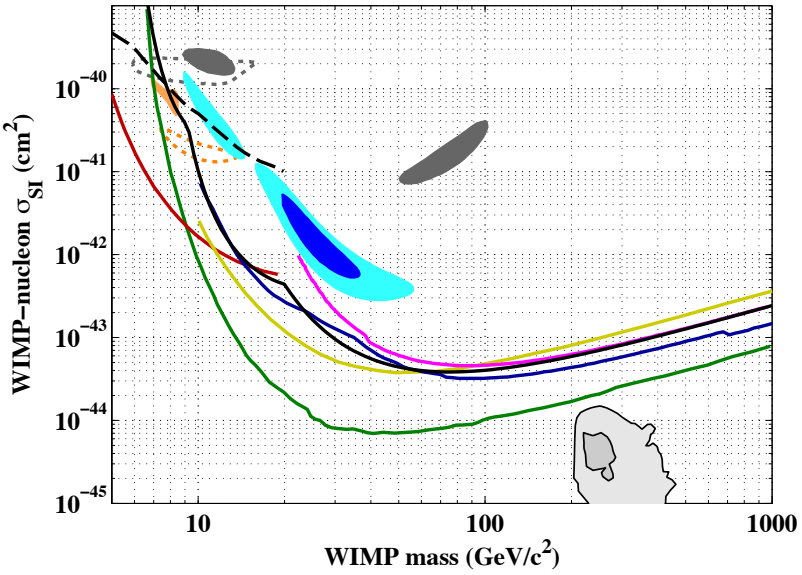


Figure 2.3: Summary of current direct detection constraints on the WIMP-nucleon spin-independent elastic scattering cross section, σ_{SI}, versus WIMP mass, m_χ. The lines show 90% CL upper limits on the cross section from XENON100 (green) [54], CDMS II (black) [153], EDELWEISS-II (magenta) [183], the combined analysis of CDMS II and EDELWEISS-II (dark blue) [184], ZEPLIN-III (yellow) [185], the XENON10 S2-only analysis (red) [156], and the CDMS shallow site (black, dashed) [186]. The light gray filled regions near $m_\chi \sim 250$ GeV show the 68% CL (inner) and 95% CL (outer) fits to the cMSSM including current direct detection and collider constraints from Fig. 1.12 [40]. The remaining filled regions show the 90% CL allowed regions consistent with possible signals in the DAMA/LIBRA (dark gray) [176, 187, 188] and CoGeNT (orange) [189] experiments, as well as the 68% CL (cyan) and 95% CL (blue) allowed regions from CRESST-II [154]. The orange dashed region shows an alternative calculation of the allowed parameter space for CoGeNT after subtracting a large (∼75%) surface event background [190–192], while the dotted gray region shows the allowed DAMA/LIBRA region after assuming a ∼50% uncertainty in the energy scale [82]. CDMS II constraints in the $m_\chi < 10$ GeV region will be presented in Chap. 5. All exclusion curves and allowed regions assume the SHM with $v_0 = 220$ km s^{-1} and $v_{esc} = 544$ km s^{-1} except the XENON10 S2-only exclusion ($[v_0, v_{esc}] = [230, 600]$ km s^{-1}), and the DAMA/LIBRA allowed regions with ($[v_0, v_{esc}] = [230, 600]$ km s^{-1}) and without ($[v_0, v_{esc}] = [220, 650]$ km s^{-1}) the energy scale uncertainties included. Using the more conservative values of $[v_0, v_{esc}] = [220, 544]$ km s^{-1} for these results would move the regions upward by ∼10–20% in the low-mass region, but would not qualitatively affect their agreement.

be in conflict with such a signal, before there is broad consensus in the community as to what these experiments are seeing. In the following sections, we review each of the potential signals and exclusions in the $m_\chi \lesssim 10$ GeV region. Constraints from CDMS II in this region will be discussed in Chap. 5.

2.1.4.1 DAMA/LIBRA

The longest standing potential signal in this region was first reported by the DAMA/NaI collaboration more than 10 years ago [196]. The current iteration of this experiment, DAMA/LIBRA, consists of 250 kg of extremely radiopure thallium-doped NaI scintillator crystals. DAMA/LIBRA has the largest exposure of any dark matter data, with nearly 1.2 ton yrs of data collected over 13 annual cycles [187]. DAMA measures only the scintillation light produced by each particle interaction, so it cannot distinguish between the expected nuclear-recoil signal from WIMPs and electron-recoil backgrounds on an event-by-event basis. Instead, it attempts to identify WIMP interactions through the presence of a residual annually modulating signal after removing the unmodulated rate. This task is made easier due to the high radiopurity of DAMA's detectors, leading to a low rate of unmodulated backgrounds in the low energy region ($\lesssim 1$ event keV$_{ee}^{-1}$ kg^{-1} day^{-1}). DAMA's energy threshold of 2 keV$_{ee}$ and target containing a light nucleus (Na) also make it sensitive to recoils from very low mass WIMPs, which produce recoil energies that may fall below the threshold of other experiments. DAMA's energy scale is given in keV$_{ee}$ units above, denoting "electron-equivalent" energy. Since nuclear recoils produce less scintillation light than electrons for the same total recoil energy, an electron-equivalent scale is typically used, where the recoil energy (in units of keV$_{nr}$ is given by E/q for E measured in keV$_{ee}$. Here, q is a target-dependent "scintillation yield" for nuclear recoils, for which measurements by the DAMA collaboration indicate $q_{Na} = 0.30 \pm 0.01$ and $q_I = 0.09 \pm 0.01$ [197].

As shown in Fig. 2.4, measurements of the residual counting rate in DAMA/LIBRA indicate the presence of an annual modulation at 8.9σ significance. This modulation has a phase consistent with that expected for a dark matter signal ($t_0 = 144 \pm 8$ days [187], while $t_0 = 152.5$ days in the SHM). As shown in Fig. 2.4b, the modulation peaks in the lowest energy bins and is absent at higher energies. It also occurs only for events interacting in a single detector module and is absent for multiple-detector hits. Although many suggestions have been proposed for backgrounds which could produce these features (e.g., [198–200]), to date the signal reported by DAMA/LIBRA remains unexplained by such a background [201] and has been interpreted as evidence for WIMP interactions. However, for spin-independent elastic scattering of WIMPs under standard assumptions, experiments such as XENON100 [54] and CDMS II [153] exclude the parameter space favored by the

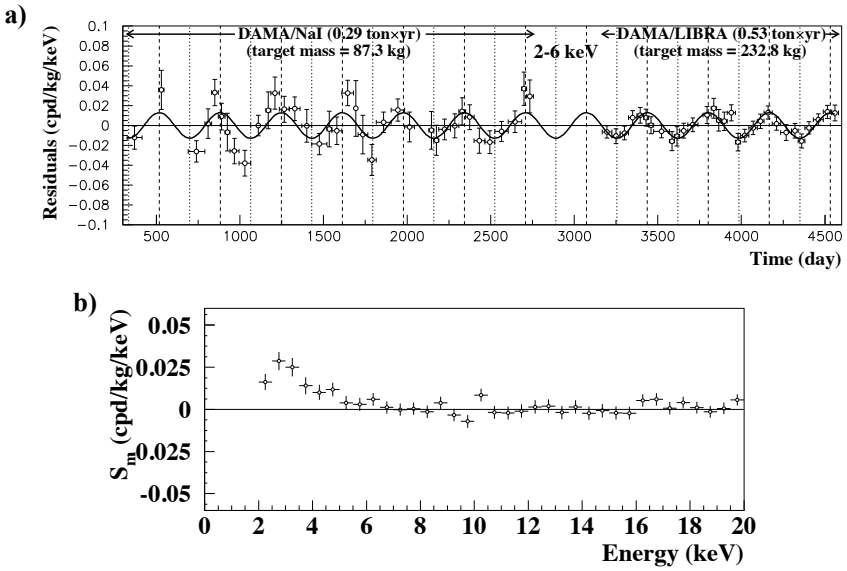


Figure 2.4: a) Residual counting rate for single-detector hits versus time for DAMA/NaI and DAMA/LIBRA from 2–6 keV$_{ee}$, after subtracting the mean total counting rate in each year. The horizontal axis gives the number of days since January 1st of the first year that DAMA/NaI operated. Clear evidence is seen for an annual modulation peaking in late May, over more than 10 annual cycles. b) Energy spectrum for the amplitude of the modulated rate. The modulation peaks in the lowest energy bins as expected for a WIMP signal, with no evidence for modulation above 8 keV$_{ee}$. *Figure from Bernabei et al. [188]*

DAMA/LIBRA modulation by several orders of magnitude if the measured recoil rate is primarily from I, as would be expected for WIMPs with masses $\gtrsim$20 GeV. Although the low-mass region corresponding to Na scattering (where the much higher rate of I recoils is missed since it falls below threshold) is also currently disfavored [144, 156], the exclusion in this region is less stringent, and uncertainties in the halo model or detector response at low energy may allow compatibility [82, 194, 195].

2.1.4.2 CoGeNT

The CoGeNT collaboration has developed a second detector technology with excellent sensitivity to WIMPs with masses $\lesssim$ 10 GeV, but with a target and measurement technique different than that employed by DAMA, in order to provide a cross check on their possible low-mass WIMP signal. CoGeNT operates a 440 g P-type point contact germanium ionization detector in the Soudan Underground Laboratory [189, 193]. Due to its point contact electrode, the capacitance of the detector is significantly reduced relative to standard

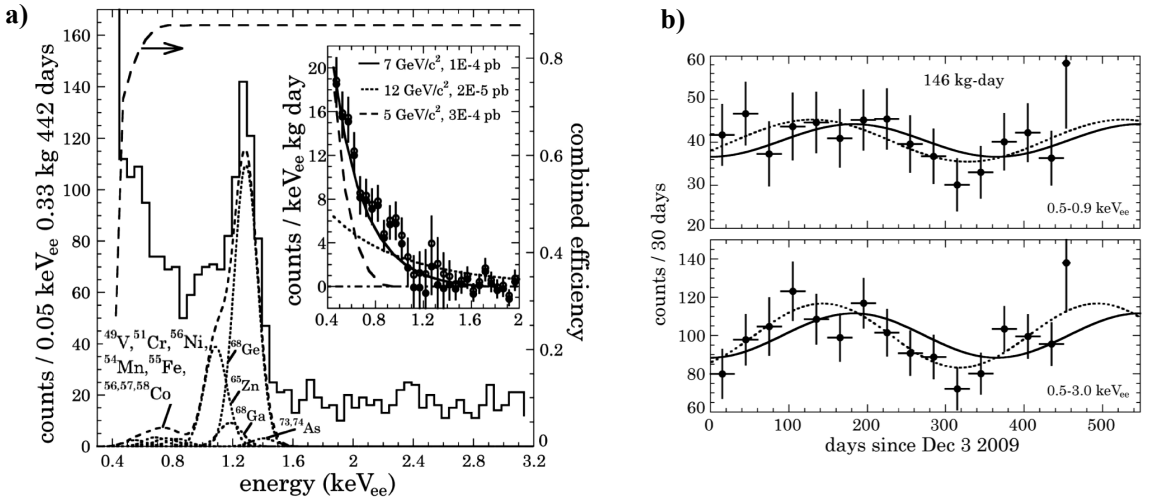


Figure 2.5: a) Total counting rate in CoGeNT at low energy after applying the surface event rejection criteria. A residual exponential excess remains after subtracting the constant Compton scatter background and known L-shell activation peaks near 1.3 keV$_{ee}$. This excess is similar to the spectrum expected for a WIMP with $m_\chi \sim$ 5–10 GeV and $\sigma_{SI} \sim 10^{-41}$–$10^{-40}$ cm^2, as shown in the inset. b) Total counting rate versus time in CoGeNT after removing surface events and subtracting the Compton and L-shell backgrounds. Evidence for annual modulation with a phase consistent with that expected for a dark matter signal is observed, with a significance of $\sim 2.8\sigma$ in the full energy range, limited by statistics. *Figure from Aalseth et al. [189]*

cylindrical geometries, allowing an extremely low ionization energy threshold of 0.4 keV$_{ee}$ in a nearly half-kg detector (for $q_{Ge} \sim 0.2$ at these energies, this corresponds to a $\sim$2 keV$_{nr}$ recoil-energy threshold). This geometry also allows the identification of interactions occurring within $\sim$1 mm of the detector surface to be identified due to the slower rising pulses for such events. This gives the ability to reject external low-energy backgrounds, which typically have a penetration depth $\ll$1 mm in the energy range of interest.

After applying their surface event rejection cut, CoGeNT initially reported a large exponential excess of events from 0.5–1 keV$_{ee}$ that could not be accounted for by known backgrounds [193], but which could be explained by a WIMP with a mass and cross section roughly compatible with that needed to explain the DAMA/LIBRA results described above [82]. Further data taking confirmed this excess and gave weak evidence ($\sim 2.8\sigma$, limited by statistics) for a $\sim$15% annual modulation in the counting rate in this excess [189], with a phase and spectrum consistent with that found by DAMA/LIBRA [190, 202]. This excess and evidence for a corresponding modulation is shown in Fig. 2.5. However, it has re-

cently been suggested that a significant fraction (∼75%) of the exponential excess observed by CoGeNT at low energy is due to leakage of surface electron recoils that cannot be completely distinguished from bulk events at low energy [190–192]. In this case, the parameter space consistent with a WIMP signal providing the remaining ∼25% of the excess is pushed to WIMP masses, $m_\chi \approx 10$ GeV and $\sigma_{SI} \approx 2 \times 10^{-41}$, improving agreement with the excess reported by CRESST-II (described in Sec. 2.1.4.3 below), but decreasing agreement with DAMA/LIBRA unless ion-channeling or local halo substructure is significant [176, 190].

If the modulation reported by CoGeNT is also due to WIMPs, then a modulation fraction of the remaining total WIMP rate of order 50–100% would be required, which is an order of magnitude larger than expected in the SHM. This modulation also appears to extend to higher energies than would be compatible with the measured total rate [202, 203], suggesting either a background origin or statistical fluctuation for the modulation detected above ∼1.2 keV$_{ee}$. This is supported by a search for a corresponding modulation from 1.2–3 keV$_{ee}$ in CDMS, which finds no significant evidence for modulation [203] at high energy, although the CDMS modulation analysis does not constrain the region below 1.2 keV$_{ee}$ where the WIMP signal would be expected. CoGeNT is continuing to acquire statistics with its existing detector module and plans an upgrade of its detector mass by roughly an order of magnitude. This larger data set will give an improved measurement of the modulation and possibly identify or exclude a background origin for the residual low-energy excess.

2.1.4.3 CRESST-II

CRESST-II is a third independent experiment that sees a low-energy excess of events that might be explained by spin-independent elastic scattering from a WIMP with $m_\chi \sim 10$ GeV and a cross section, $\sigma_{SI} \sim 5 \times 10^{-41}$ cm^2 [154]. The CRESST-II experiment consists of an array of $CaWO_4$ scintillating crystals, instrumented to readout out both the scintillation light and heat deposited by each particle interaction. Due to the reduced scintillation light yield for nuclear recoils relative to the total heat signal measured, most electron recoil backgrounds can be identified and removed on an event-by-event basis. CRESST-II's ∼10 keV$_{nr}$ threshold in the heat channel and the inclusion of a light nucleus (O) in its target allow sensitivity to WIMPs with masses ∼10 GeV.

A plot of the observed counting rate in CRESST-II for events with light yield consistent

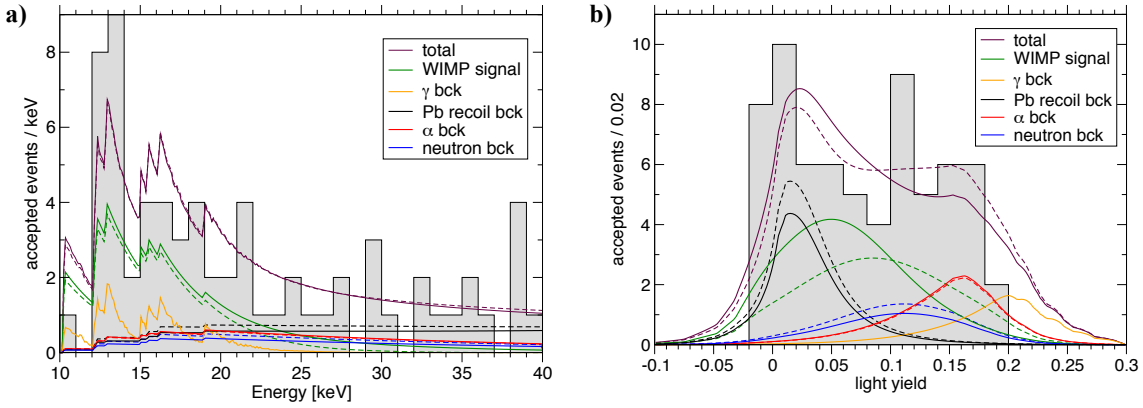


Figure 2.6: Observed spectra versus recoil energy (a) and light yield (b) for the nuclear recoil candidate events in CRESST-II. The colored curves denote the best-fit WIMP and background components from a maximum likelihood analysis in the light yield versus recoil energy plane. The dashed and solid lines correspond to the two local maxima found by the fit. A WIMP component (green) is preferred by the fit at $>4\sigma$. *Figure from Angloher et al. [154]*

with a nuclear recoil signal is shown in Fig. 2.6. A combined likelihood analysis of the recoil energy and light yield for nuclear recoil candidates indicates the preference for a nuclear recoil signal at $>4\sigma$ confidence, after including known backgrounds [154]. The corresponding allowed region extends to WIMP masses as low as 9 GeV and agrees with the cross section implied by DAMA/LIBRA and CoGeNT within a factor of a few in this region, as shown in Fig. 2.3.

Figure 2.6 demonstrates that most electron recoil backgrounds are expected to leak into the signal region from high light yield and do not appear to be capable of accounting for the exponential excess of events at light yields near zero. However, $^{210}\mathrm{Po} \rightarrow \alpha + {}^{206}\mathrm{Pb}$ decays are a known background at low light yield, if the α is absorbed in nonscintillating material (such as the detector clamps) and only the recoiling ^{206}Pb nucleus is detected. Simulations performed by CRESST-II indicate that the ^{206}Pb background should not increase at low energy, and thus the rate of this background measured in a reference region above 40 keV constrains the total rate at low energy in the likelihood fit, preventing the ^{206}Pb background from accounting for a large fraction of the low-energy exponential excess. However, independent simulations including sputtering and surface roughness effects indicate that an increasing spectrum for the ^{206}Pb might be expected at low energy [204]. To eliminate this possibility, CRESST-II is currently working to reduce the ^{206}Pb background by using

detector clamps with lower contamination [154].

2.1.4.4 Null results

Although the DAMA/LIBRA, CoGeNT, and CRESST-II results could be explained by spin-independent elastic scattering of low-mass WIMPs, other searches in this region appear to disfavor this possibility. The strongest constraints come from a dedicated search for low-mass WIMPs from the XENON10 experiment [156]. XENON10 typically measures both the scintillation light and ionization produced by each interaction in a liquid Xe target. The ratio of the ionization to scintillation signal gives the ability to discriminate between electron recoil backgrounds and nuclear recoils on an event-by-event basis. However, the scintillation yield limits the recoil energy threshold to ∼5 keV if both a scintillation and ionization signal are required [156,205]. Since Xe provides a heavy target nucleus ($A \approx 131$), a $m_\chi \lesssim 10$ GeV WIMP would not be expected to produce recoils above the threshold in the scintillation channel. A lower threshold can be obtained by requiring only an ionization signal, at the cost of increased acceptance of backgrounds.

The lack of a scintillation signal has two primary implications: the depth of the interaction can no longer be determined, and thus fiducialization to the center of the detector where backgrounds are lowest is more difficult, and it is no longer possible to discriminate between nuclear recoils and electron recoils. Nonetheless, using conservative assumptions for the calibration of the recoil energy scale from the ionization signal [156], an energy threshold of 1.4 keV_{nr} can be obtained, giving sensitivity to masses of $m_\chi \approx 10$ GeV and below.

Approximately 20 candidate events were observed in the energy range from 1.4–10 keV_{nr} in 15 kg days of exposure, leading the the limits shown as the red curve at low mass in Fig. 2.3, which conservatively assume that all candidate events could be from WIMPs. This result excludes the region consistent with CoGeNT, assuming the entire low-energy excess is due to WIMPs, by more than an order of magnitude for spin-independent elastic scattering of low-mass WIMPs under standard assumptions. Under the same assumptions, this result also excludes the possibility that only ∼25% of the low-energy excess in CoGeNT is actually due to WIMPs (and the remaining events are surface electron recoils) [190–192]. Astrophysical uncertainties cannot reconcile the result, since applying the astrophysics independent estimates described in Sec. 2.1.3 leads to the same order-of-magnitude exclusion. Com-

patibility may be possible in WIMP models in which the scattering from Xe is suppressed relative to Ge [180].

A low-energy analysis of data taken by CDMS at the shallow site using an energy threshold of $\sim$0.5 keV$_{nr}$ was also performed [186] to constrain possible low-mass WIMP interpretations of the DAMA/LIBRA, CoGeNT, and CRESST-II results. Like the XENON10 low-energy analysis, the lowered threshold in the CDMS analysis comes at the cost of increased backgrounds. Roughly 1000 candidate events were observed above threshold in an exposure of $\sim$71 kg days, corresponding to an efficiency corrected rate of $\sim$10 events keV$_{nr}^{-1}$ kg^{-1} day^{-1} at 2 keV$_{nr}$. Assuming all candidate events could be from WIMPs, but setting limits which take into account the candidate energy spectrum and detector distribution, gives the constraint shown as the black dashed line in Fig. 2.3. This result disfavors a low-mass WIMP interpretation for the DAMA/LIBRA modulation at the 90% CL, for spin-independent elastic scattering of low-mass WIMPs in the SHM, unless $q_{Na} > 0.4$ at low energy. It also disfavors the portion of the CRESST-II allowed region below 10 GeV, but cannot constrain the CoGeNT allowed region. Constraints from a similar low-threshold analysis of CDMS data taken at the deep site will be presented in Chap. 5.

2.2 Other techniques

Due to the uncertainties in the detailed properties of the WIMP-nucleon interaction and the galactic halo discussed in Sec. 2.1.3, other experimental probes will likely be required to fully constrain WIMPs even if there is broad consensus among direct detection experiments of such a signal. Although a detailed discussion is beyond the scope of this thesis, there are two primary additional techniques which can be used to constrain WIMP properties, independent of direct detection constraints. These include the "indirect detection" of WIMPs through observation of their Standard Model annihilation products (e.g., γ rays, neutrinos, or antimatter), as well as direct production of WIMPs or related particles in particle accelerators. At a qualitative level, these techniques correspond to the same Feynman diagram, but with the interaction proceeding along different directions as shown in Fig. 2.7. Thus, information determined from each experimental probe is highly complementary to the other techniques and can be used to break degeneracies and improve constraints on the properties of WIMPs.

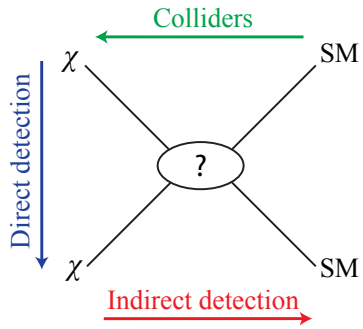


Figure 2.7: Qualitative schematic of complementarity between direct detection, indirect detection, and collider probes of dark matter. Each technique corresponds to the same basic set of Feynman diagrams, with χ denoting the dark matter particle, and SM denoting some set of Standard Model particles, but with time running in different directions. Due to these relations, constraints from multiple techniques can be used to obtain complementary information about the dark matter particle.

2.2.1 Indirect detection

Indirect detection of dark matter attempts to identify WIMP annihilations through the resulting Standard Model annihilation products (for a recent review, see e.g., [206]). Since the annihilation rate depends on the square of the WIMP density, the highest flux of such annihilation products should be seen from locations such as the galactic center or core of the sun, where the WIMP density is expected to be high. The challenge is to identify an annihilation signature which could not be mimicked by astrophysical backgrounds. For this reason, the ultra-faint dwarf galaxies discussed in Sec. 1.1.1 offer a promising target for such observations, since they are highly dominated by dark matter and astrophysical processes in these regions are much simpler than, e.g., in the galactic center. As shown in Fig. 2.8, a recent analysis of the γ-ray emission from these dwarf galaxies using observations from the Fermi Gamma-ray Space Telescope has begun to exclude the canonical thermal WIMP cross section ($\langle\sigma_{ann}v\rangle = 3 \times 10^{-26}$ cm^3 s^{-1}) for certain decay modes and $m_\chi \lesssim 30$ GeV [207]. Additional data taken with Fermi should improve these constraints.

There have been several anomalous indirect detection signals that have been interpreted as possible evidence for dark matter annihilations (e.g., [208–210]). One recent example is the excess of positrons observed above 10 GeV by PAMELA [211] (and later confirmed by Fermi [212]). Although such an excess was not expected to arise from astrophysical sources, interpretations in terms of dark matter [213–215] must account for a lack of a corresponding

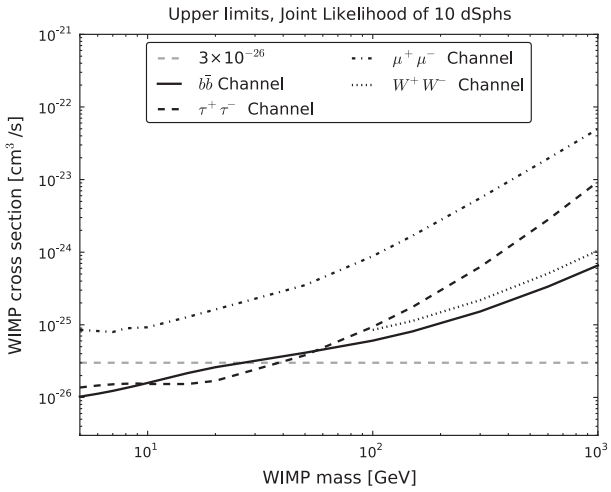


Figure 2.8: 95% CL upper limits on the velocity averaged annihilation cross section from a joint analysis of γ-ray emission from 10 dwarf spheroidal galaxies in the Fermi data. For certain annihilation channels, these results are beginning to constrain WIMPs with the canonical annihilation cross section for $m_\chi \lesssim 30$ GeV. *Figure from Ackermann et al. [207]*

antiproton signal in PAMELA [211]. More detailed study of astrophysical sources indicated that pulsars [216, 217] may also be able to explain the excess, and like other anomalies to date, this signal cannot be unambiguously interpreted as evidence for dark matter. Due to the complex astrophysical processes that can mimic a continuum WIMP signal, a narrow spectral feature in high-energy γ-rays originating from areas of high dark matter density has long been considered a "smoking gun" for dark matter annihilations, with negligible expected astrophysical backgrounds. A tentative detection of a just such a feature has recently been reported in the γ-ray spectrum at $E_\gamma = 130$ GeV [218], although more data will be required before such a signal can be confirmed.

It has also recently been suggested that excess γ-ray emission near the galactic center could be evidence for annihilations from low-mass WIMPs (similar to what would be needed to explain the DAMA/LIBRA, CoGeNT, and CRESST-II results) [219, 220]. This hypothesis is strengthened by synchrotron emission from the galactic center [221] as well as radio filaments [222], suggesting a corresponding population of high-energy electrons and positrons possibly created from dark matter annihilations. As shown in Fig. 2.9, a 10 GeV WIMP annihilating primarily to leptons with a thermal cross section may account for the excess in the spectrum of spatially extended γ-ray emission near the galactic center re-

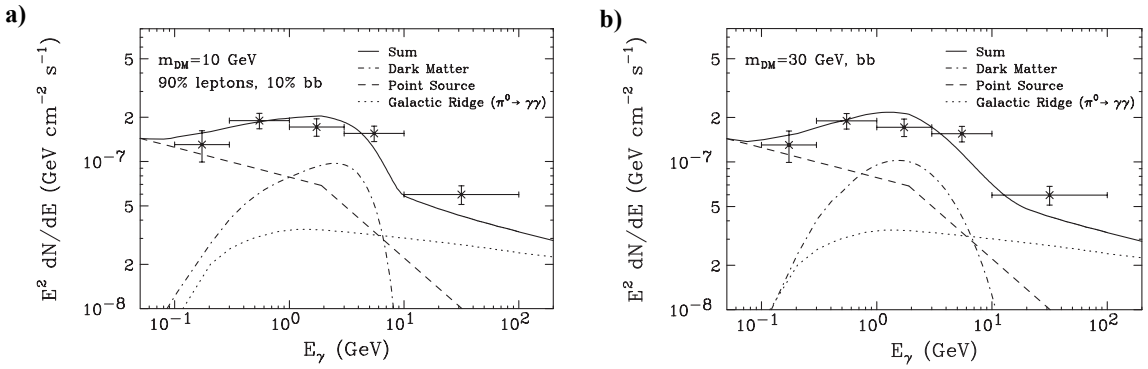


Figure 2.9: Fits to the spectrum of γ ray emission from the spatially extended excess near the galactic center reported in [220]. Either a 10 GeV WIMP annihilating primarily to leptons, or a 30 GeV WIMP annihilating to quarks could account for the excess emission above the assumed background model. *Figure from Hooper and Linden [220]*

ported by Hooper and Linden [220], although a heavier WIMP annihilating to quarks fits the spectrum equally well. These types of studies indicate the complementarity between direct and indirect detection methods [223], with the possibility of beginning to observe a positive signal for dark matter interactions with two independent techniques.

2.2.2 Accelerators

As discussed in Sec. 1.3, accelerators have the ability to constrain dark matter models through direct production of the WIMP or related particles in the laboratory. Collider probes are highly complementary [224] to indirect and direct searches since they probe the WIMP parameter space up to a given mass, roughly independent of the cross section. In contrast, direct and indirect searches are limited by cross section, but can probe WIMPs to a much higher range of masses than current accelerators. In addition, while colliders can produce and study a variety of strongly interacting particles predicted in many WIMP models, the WIMP itself will escape the detector unnoticed and only show up as some amount of missing energy in the event reconstruction (similar to the collider signature of a neutrino). Thus, determining the detailed properties of the WIMP and whether it can account for the observed relic density of dark matter through collider data alone may be difficult if the particles that can be measured by the collider do not fully constrain the theory [224, 225].

Since the LHC has not yet detected evidence for SUSY or other models of dark mat-

ter [40], it remains possible that the WIMP or related particles will be inaccessible to the LHC, in which case direct and indirect detection may offer the only near-term means of identifying particle dark matter [226]. Given the importance of understanding the properties of dark matter for particle physics and cosmology, and the remaining uncertainty in its particle nature, a wide variety of techniques including indirect, direct, and collider probes will be required to ultimately detect dark matter, or to rule out the most popular models discussed in Sec. 1.3.

Chapter 3

The Cryogenic Dark Matter Search

As discussed in the previous chapters, if the dark matter does consist of WIMPs, we expect interaction rates in terrestrial detectors of <1 event/(kg day), with typical deposited energies ∼10–100 keV. At the same time, backgrounds that can mimic a WIMP signal are orders of magnitude larger. To meet these challenges, the Cryogenic Dark Matter Search (CDMS) has developed Z-sensitive Ionization and Phonon (ZIP) detectors that measure both the ionization and phonon energy deposited by each interaction with an energy resolution of ∼0.25 keV. The design and operation of these detectors are discussed in Sec. 3.1. Section 3.2 describes the final CDMS-II experimental installation, which consisted of 30 ZIP detectors operated at the Soudan Underground Laboratory (SUL). In Sec. 3.3, we discuss the data acquisition and processing pipeline used to perform reconstruction and calibration of each particle interaction in the apparatus. Detailed descriptions of the CDMS-II detectors and experimental installation at the SUL can also be found in, e.g., [227–230].

3.1 ZIP detectors

CDMS-II ZIP detectors consist of 76 mm diameter, ∼10 mm thick germanium or silicon substrates lithographically patterned with both ionization and athermal phonon sensors, as shown in Fig. 3.1. Particle interactions in the substrate produce either a recoiling electron or nucleus with typical energies ∼10 keV in the range of interest for dark matter interactions. As described in the following sections, this recoil produces both ionization and nonequilibrium ("athermal") phonons. By measuring the detailed properties of the phonon signal, both the location and energy of the interaction can be determined. In addition, the ratio of the ionization signal to the phonon signal allows the discrimination of electron-recoil

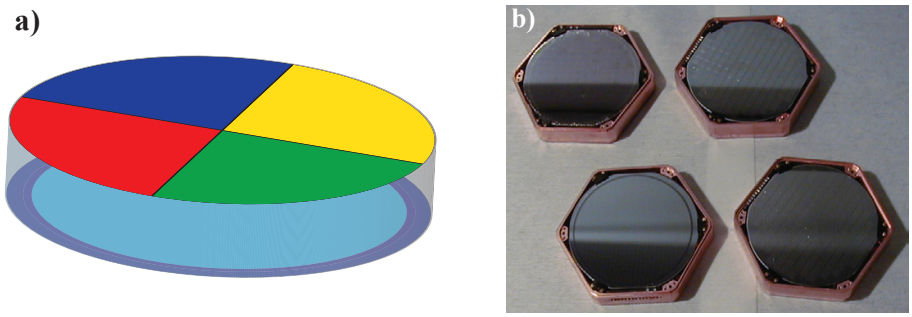


Figure 3.1: a) Schematic of ZIP sensor configuration. The top face of the detector is patterned with athermal phonon sensors, wired in quadrants. The bottom face consists of two concentric charge electrodes with the inner circular electrode defining the radial fiducial volume. b) Picture of 4 fabricated detectors in their housings. The charge face is shown for the bottom left detector, while the remaining detectors show their phonon faces.

backgrounds from the expected nuclear-recoil signal.

For spin-independent scattering, the A^2 enhancement of coherent scattering with nuclei gives a larger expected WIMP-interaction rate for Ge than Si. In addition, the higher density for Ge (5.3 g/cm^3) versus Si (2.3 g/cm^3) gives a $\sim$2 times larger mass for Ge than Si per detector element ($\sim$0.23 kg versus 0.11 kg, respectively). For an $m_\chi = 100$ GeV/c^2 WIMP, each Ge detector element has >10 times higher expected sensitivity to spin-independent elastic scattering of WIMPs than a corresponding Si detector element, provided background rates are similar. However, low-mass WIMPs ($1 \lesssim m_\chi \lesssim 10$ GeV/c^2) can produce recoils with sufficiently low energy that they fall below the detector energy threshold for Ge but can still be seen above threshold for Si due to its lower atomic mass. At these low WIMP masses, Si can be more sensitive to a WIMP signal than Ge. In addition, the inclusion of two nuclei in the target allows one to test the expected A^2 scaling if a positive signal is detected.

The full installation of detectors at the SUL consisted of 30 detectors (19 Ge and 11 Si) arranged in 5 stacks, or "towers", of 6 detectors each. Detectors are labeled by their tower number and location (from top to bottom) in the stack, so e.g., T1Z5 is the 5th detector from the top of the 1st stack.

3.1.1 Ionization signal

A particle interaction in the substrate produces a recoiling electron or nucleus, which can lose energy either by generating electron-hole pairs or by phonon emission. As discussed

in Sec. 3.1.3 below, nuclear recoils will produce less ionization than electron recoils due to the denser deposition of energy from the more slowly moving nucleus. For electron recoils, the average number of electron-hole pairs produced by the interaction can be written as $N_Q = E_R/\epsilon$ where E_R is the recoil energy and ϵ is the average energy required to create and electron-hole pair. Since most of the recoil energy is initially deposited in the phonon system, $\epsilon > E_g$, where $E_{g,Ge} = 0.74$ and $E_{g,Si} = 1.17$ eV are the band gap energies for Ge and Si, respectively, at 50 mK [231, 232]. Measurements of the ionization produced by electron recoils in calorimetric detectors indicate that $\epsilon_{Ge} = 3.0$ eV [233–235] and $\epsilon_{Si} = 3.82$ eV [232, 236] at 50 mK, in good agreement with theoretical models (see e.g., [232], Appendix C for a review). These numbers indicate that a 10 keV electron recoil would generate $\approx 3.3 \times 10^3$ electron-hole pairs, corresponding to only 30% of the deposited energy.

To collect this ionization signal, the charge electrodes patterned on the bottom surface of the detector were biased at $V_b = +3.0$ V for Ge and $V_b = +4.0$ V for Si for the data analyzed in this thesis, with the phonon sensors on the top surface grounded. The bias is supplied through a $R_b = 40$ MΩ resistor to decouple it from the detector on the time scale of the pulse. This small electric field is sufficient to drift the charge carriers produced by the interaction to the detector surfaces without significant recombination of the e-h pairs in the cloud of charges generated along the ionization track or in localized traps throughout the detector [233]. Ionized impurities can have large trapping cross sections and must be neutralized to prevent degradation of the charge collection. This is accomplished by exposing the detector to ionizing radiation from a strong radioactive source or an infrared LED, while grounding both detector surfaces. This produces a large population of e-h pairs, most of which will quickly recombine, but for which a small fraction can diffuse and become trapped in an ionized impurity. At 50 mK, the temperature is well below the typical activation energy of the impurity sites, allowing them to remain neutralized on the time scales required to collect WIMP search data. However, if the detector remains under bias for long periods, these sites can gradually become ionized by particle interactions, and a grounded LED flash is required every $\sim$10 hours to maintain neutralization in WIMP search running.

The charge electrodes are patterned into two concentric electrodes to define a fiducial volume, with the "inner" electrode covering 85% of the surface area of the detector, and the "outer" annular electrode covering 15%. Both the inner and outer electrodes are patterned

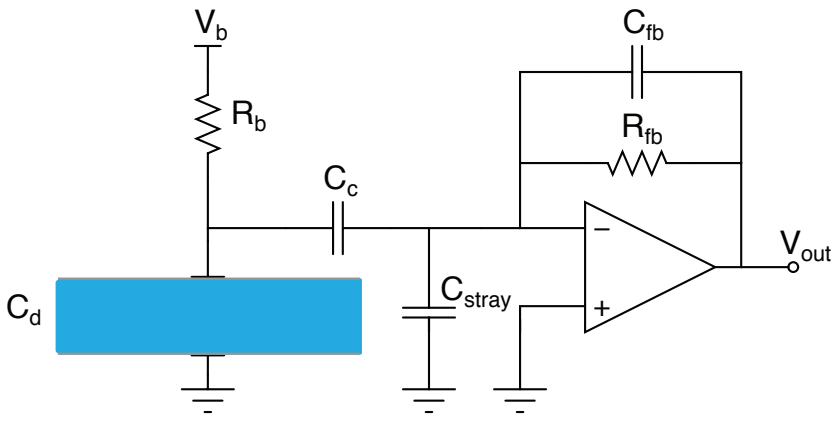


Figure 3.2: Schematic of charge readout circuit. The detector (blue) has capacitance $C_d = 93$ pF for the inner electrode (36 pF for the outer). The phonon face is held at ground, while the charge faced is biased through $R_b = 40$ MΩ, typically at $V_b = +3.0$ V for Ge. This bias is decoupled from the amplifier through $C_c = 300$ pF. Charge collected across the detector is measured by the image charges transferred to the feedback capacitor $C_{fb} = 1$ pF, which drains through the feedback resistor, $R_{fb} = 40$ MΩ, with characteristic time $\tau = 40$ μs. The stray capacitance to ground, $C_{stray} \approx 75$ pF is comparable to the detector capacitance, but still small relative to C_c. *Figure from J. Filippini [228]*

into a grid with a fill factor of ∼15% to limit collection of athermal phonons on the charge face of the detector where they cannot be sensed. Each electrode is read out by the charge amplifier circuit shown in Fig. 3.2. Since the charge carrier drift velocities at 50 mK are 20-40 mm/μs [237] and the amplifier response time is sufficiently fast, the rise time of the pulse is essentially instantaneous relative to the 0.8 μs sampling time of the ADC. The fall time of the pulse is set by the feedback circuit, $R_f = 40$ MΩ and $C_f = 1$ pF, so $\tau_f = R_f C_f = 40$ μs. The resulting charge pulses have a fixed shape given by a single exponential with decay time τ_f.

At sub-Kelvin temperatures and for $\lesssim$5 V/cm electric fields along the [100] crystalline axis in Ge, electrons do not propagate directly along the electric field direction, but rather at a ∼30° angle to the field [237]. In contrast, holes propagate along the field direction, leading to improved collection and decreased radial dispersion of the charge carriers when the charge face of the detector is biased to collect holes (i.e., $V_b < 0$). Operation of the detectors with a negative bias gives a larger effective fiducial volume and reduces the size of the high-radius region where charge carriers can be collected on the cylindrical walls of the detector. As discussed in Sec. 5.3.4, such poorly collected high-radius events are a

significant background at low energy. In this thesis, we consider data only with $V_b = +3$ V, giving higher backgrounds than would be expected for operation with a negative bias.

The charge readout noise provides a limit to the background rejection at low energy since the ionization signal for nuclear recoils and electron recoils can no longer be easily distinguished as the ionization signal becomes comparable to noise. The dominant theoretical noise contributions are the voltage noise of the JFET amplifier, measured to be typically 0.5 nV/$\sqrt{\text{Hz}}$ [227] referenced to the input, and Johnson noise from the bias and feedback resistors at 50 mK. As shown in Fig. 3.3, the measured charge noise at Soudan is limited by $1/f$-like pickup that dominates the noise below the 10 kHz signal frequencies of interest for pulse data. This excess noise likely results from a combination of electronic pickup (in particular 60 Hz and harmonics) and microphonic pickup from coupling to mechanical vibrations. These nonfundamental noise sources (in particular microphonic pickup) can vary significantly by detector, leading to nearly a factor of 2 variation in the reconstructed charge resolution for each detector from 0.25–0.50 keV$_{ee}$.

The ionization produced by the interaction also contributes to the total phonon signal due to the generation of Neganov-Luke phonons [238, 239]. After charge carriers are produced by an interaction, the electric field quickly accelerates these charges to a terminal drift velocity above the sound speed that is limited by the radiation of athermal ($\sim$1 K) phonons [237], in analogy to Cerenkov radiation. The total energy emitted as Neganov-Luke phonons is equal to the work done to drift the charges:

$$E_{NL} = eV_b N_Q \tag{3.1}$$

As discussed in Sec. 3.1.2, this additional contribution from Neganov-Luke phonons must be accounted for to correctly estimate the recoil energy of the event. These phonons also limit the maximum charge bias that can be applied while maintaining ionization-based discrimination between electron and nuclear recoils. Increasing the detector bias increases the Neganov-Luke phonon contribution, and eventually swamps the independent information available from the recoil phonons, leading to suppressed ionization-based discrimination between recoil types.

If one is willing to give up ionization-based discrimination, this technique can be used to significantly improve the resolution of the charge measurement. Since the noise in the

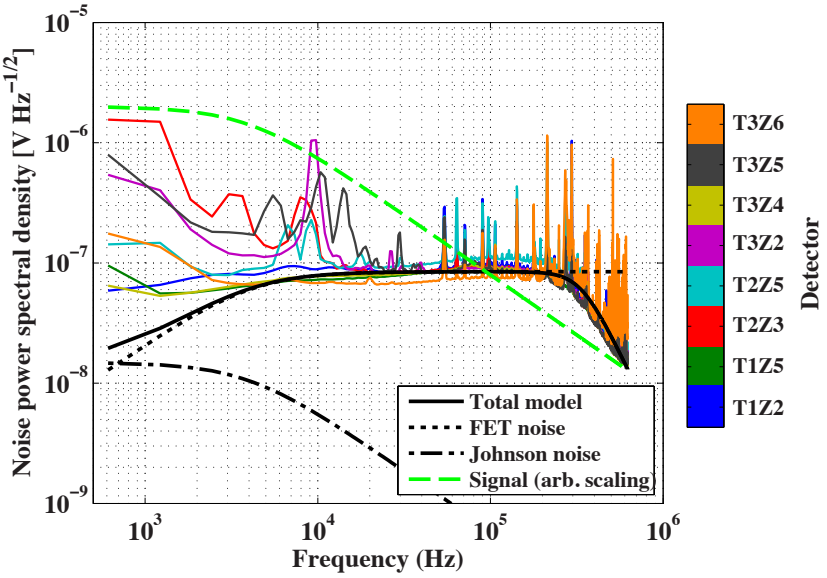


Figure 3.3: Measurement of the charge noise power spectral density referenced to the output of the electronics break out box for the 8 Ge detectors used in the low-threshold analysis discussed in Chap. 4. The noise model for the charge readout circuit was derived by T. Shutt in [234] and is dominated by the JFET noise. Using the measured 0.5 nV/$\sqrt{\text{Hz}}$ FET noise at 10 kHz [227] gives good agreement with the observed baseline above 20 kHz. All detectors show pickup at isolated spectral peaks above 50 kHz, which does not significantly impact the resolution since the signal frequencies of interest are $\lesssim$10 kHz (green, dashed). More problematic is the excess low-frequency pickup above FET noise, which dominates the signal-to-noise for the reconstructed signal. Several detectors (e.g., T2Z3, T3Z2 and T3Z5) show significant microphonic pickup from vibrations due to the electronics stem cryocooler at low frequencies and in peaks around 10 kHz. These detectors suffer from correspondingly poorer charge resolution.

phonon channels is independent of the voltage bias, V_b, but the Neganov-Luke phonon signal is proportional to V_b, amplification of the charge signal can be obtained by operating at high biases and using the phonon channels to read out the Neganov-Luke dominated signal [239]. J. Hall and R. Basu Thakur have successfully demonstrated such high voltage operation using CDMS II detectors, reaching analysis thresholds as low as 0.08 keV$_{ee}$. However, ionization-based background rejection is not possible in this operation mode, so future work will be needed to establish background levels. M. Pyle has pointed out that in this mode the constant Compton background will be suppressed by a factor proportional to the gain, relative to a nuclear recoil signal, since the large charge gain spreads the electron-recoil distribution over a larger energy range than the corresponding nuclear recoil distribution. In addition, due to the same effect, statistical background discrimination is possible by

comparing background spectra at multiple charge biases [240, 241]. In the limiting case, operation at high biases allows a measurement of the ionization signal alone, while operation at zero bias allows the measurement of the total recoil energy alone. Comparison of the spectra measured at high bias and at zero bias would then allow statistical ionization-based discrimination between electron recoils and nuclear recoils, even though the recoil type cannot be determined on an event-by-event basis.

3.1.2 Phonon signal

In addition to producing ionization, a recoiling electron or nucleus deposits the majority of its energy directly as athermal phonons due to interactions with the crystalline lattice [227, 242, 243]. The majority of these phonons, which we denote as "primary phonons," are emitted as high-energy optical phonons at the Debye frequency ($\sim$10 THz). We also include optical and acoustic phonons created as the initial population of hot charge carriers relax to the gap in the primary phonon signal, so that the total energy can be written as $E_{prim} = E_r - E_g N_Q$, where E_r is the recoil energy. These phonons are created at the interaction location and thus encode significant position-dependent information.

Following the creation of the primary phonons, an additional population of Neganov-Luke phonons are created during the $\lesssim$1 μs it takes to drift the carriers across the detector. The total energy emitted is given by the work done to drift the charge carriers, as described in Eq. 3.1. These phonons are emitted with frequencies well below the Debye frequency (0.1–0.6 THz) [244], but nearly all remain above the pair-breaking energy in Al, 2Δ = 0.09 THz.

Upon reaching the charge electrodes, the charge carriers relax to the Fermi level, releasing their remaining energy, $E_{relax} = E_g N_Q$ as "relaxation phonons." Since the $\sim$1 eV band gap energies are much larger than the <1 meV pair-breaking energy in the superconducting Al films, we expect that this process primarily creates pair-breaking phonons which contribute to the total measured athermal phonon signal. Combining all three populations gives the total phonon signal:

$$E_{tot} = (E_r - E_g N_Q) + (eV_b N_Q) + (E_g N_Q) = E_r + eV_b N_Q \tag{3.2}$$

These phonons must propagate to the sensors on the detector surfaces before they can

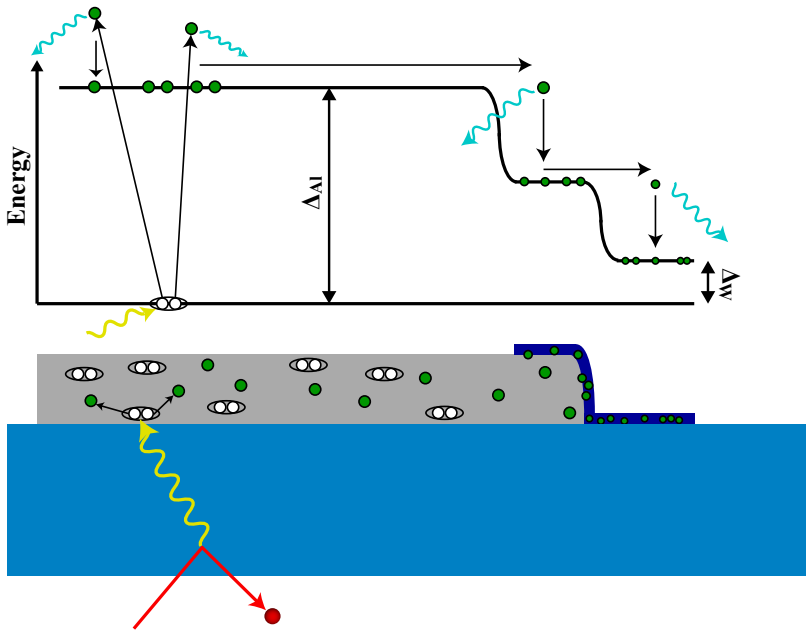


Figure 3.4: Schematic of athermal phonon collection and quasiparticle diffusion. An interaction in the substrate (light blue) creates ∼THz phonons ($\gg 2\Delta_{Al} = 90$ GHz) which can break Cooper pairs (white) in the Al absorbing films (gray), creating quasiparticles (green). These quasiparticles can diffuse into the W TES (dark blue) where they become trapped. The upper part of the figure shows the downconversion process, where energetic quasiparticles created by the pair-breaking phonon quickly downconvert to the gap edge by emitting phonons (creating additional quasiparticles if these phonons are sufficiently energetic). These quasiparticles can then diffuse into the Al/W interface region (where the gap is suppressed due to the proximity effect) or into the W TES itself, where they again rapidly emit phonons and fall below Δ_{Al}, becoming trapped.

be detected. Initially, the high-frequency phonons propagate quasidiffusively, with their mean free path limited to $l \ll 1$ mm by anharmonic decay into two lower energy phonons ($l \propto \nu^{-5}$) and elastic scattering from isotopic impurities ($l \propto \nu^{-4}$). These phonons rapidly undergo anharmonic decay until they become "ballistic," i.e. their mean free path becomes comparable to the dimensions of the substrate ($\nu \lesssim 1$ THz).

These ballistic phonons are collected in "Quasiparticle-trap-assisted electrothermal-feedback transition-edge-sensors" (QETs) patterned on the detector surface. A cross-section of a single QET element is shown in Fig. 3.4. Phonons with $\nu > 90$ GHz can break Cooper pairs in thick Al absorbing fins, creating quasiparticles well above the gap. These energetic quasiparticles emit pair-breaking phonons, creating a cascade in which ∼60% of the initial energy is converted into quasiparticles at the gap energy, while the remaining energy is lost

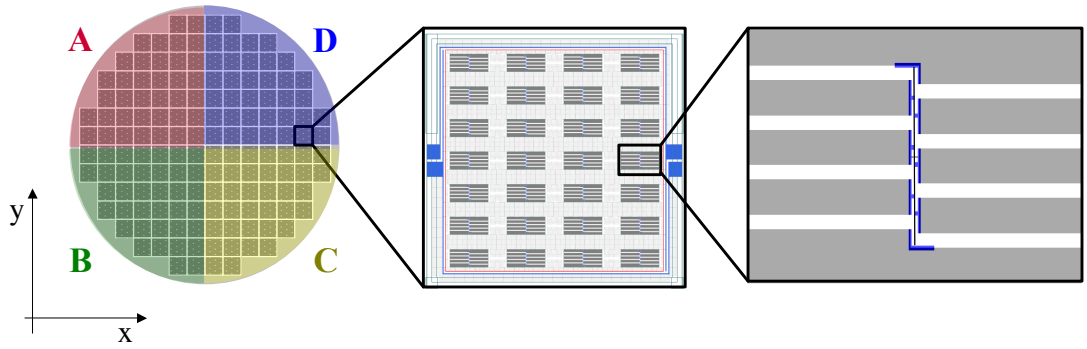


Figure 3.5: Phonon sensor layout. The top face of the detector is patterned with athermal phonon sensors divided into 4 quadrants labeled (A, B, C, D). Each quadrant consists of 37 identical tiles consisting of 28 TESs each, giving at total of 1036 TESs per quadrant, wired in parallel. The zoomed regions show the Al absorbing fins (gray), 8 of which are connected to each TES (blue). *Figure adapted from [247]*

as sub-gap phonons [245]. These quasiparticles can then diffuse into tungsten transition-edge sensors (TESs) connected to the end of the fin, where they rapidly emit a phonon and fall below the gap of the Al, becoming trapped [246]. This allows phonon energy incident on a large fraction of the detector surface area to be concentrated in small volume TESs. The Al collecting fins are 350 μm long, 50 μm wide and 300 nm thick, covering $\sim$15% of the top detector surface. As shown in Fig. 3.5, the phonon sensors are divided into four quadrants, each consisting of 1036 QET elements wired in parallel. Due to small variations in the TES transition temperatures across the device, the TESs must be wired in parallel rather than in series to maintain a stable voltage-biased configuration for each element, as discussed below. The TES geometry and choice of W are determined by the large number of TESs that must be wired in parallel. To maintain $\sim$1 Ω sensor resistance, each TES must have $\sim$1 kΩ resistance, requiring a superconductor with the desired T_c and high normal resistivity, as well as a long, narrow TES geometry.

The TESs themselves consist of 1 μm wide, 250 μm long, and 35 nm thick W films biased to operate in their superconducting transition ($T_c = 80$ mK). The W films are initially deposited with a $T_c \sim 120$ mK, after which they are ion-implanted with a position-dependent dose to lower their T_c to 80 mK and reduce any position-dependent nonuniformities in T_c [248]. In this transition, a small change in temperature will produce a large change in resistance, giving an extremely sensitive thermometer. Negative electrothermal feedback (ETF) can be applied by operating the TES in a voltage-biased configuration, as shown

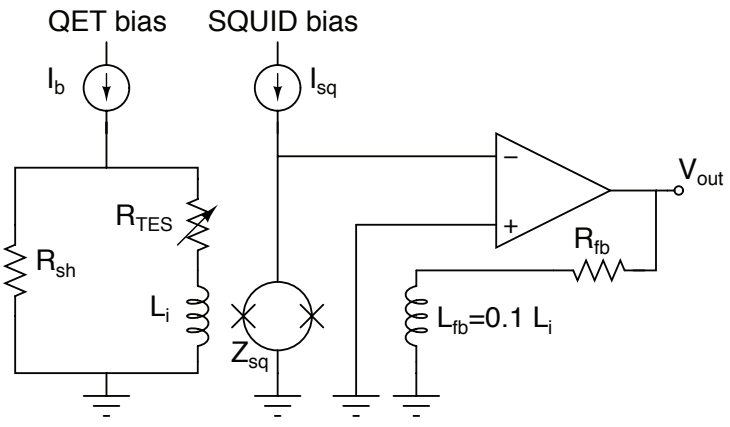


Figure 3.6: Schematic of phonon readout circuit. The TES (typically operated at $R_{TES} \approx$ 100–200 mΩ) is wired in series with the input coil, $L_i = 250$ nH and voltage biased by the shunt resistor, $R_{sh} = 25$ mΩ. The amplifier adjusts the current through the feedback coil, $L_{fb} = L_i/10$, to cancel the flux through the SQUID array (represented here as a single SQUID, Z_{sq}) from the input coil, giving a factor of 10 gain in the feedback current. This is translated to a voltage at the output, V_{out}, by the feedback resistor, $R_{fb} = 1.2$ kΩ. *Figure from J. Filippini [228]*

in Fig. 3.6. The voltage bias is maintained by a shunt resistor ($R_{sh} = 25$ m$\Omega \ll R_{TES}$) connected in parallel with the TES. In this configuration, the bias voltage is tuned so that the Joule heating ($P = I_{TES}V_b$) due to the bias current self-heats the electron system in the TES to the desired operating point along its transition. Typically, the TES array is biased only partway up the transition, with $R_{TES} \approx$ 100–200 mΩ. At low temperatures, the coupling of the electron and phonon systems is small (with thermal conductance $G \propto T^5$), providing a weak thermal link between the TES and the bath. This allows the TES to self-heat into its transition, while the substrate is maintained at 50 mK. When an event occurs, R_{TES} increases, decreasing the Joule heating and allowing the sensor to quickly return to its steady-state operating point. For the TESs in CDMS II, the ETF provides a characteristic recovery time of $\lesssim$40 μs, while phonon-mediated events have a longer fall time (~ 250 μs) due to the physical arrival time of the phonons at the absorbing fins.

The change in current through the TES is read out with a single-stage SQUID array as shown in Fig. 3.6. The TES is connected in series with an input coil, L_i, which converts the change in I_{TES} into a change in flux through the SQUID array. Although the SQUID array provides an extremely sensitive magnetometer, its response is highly nonlinear and the array must be operated in closed-loop mode to increase its dynamic range. In this

mode, the voltage change across the SQUID is amplified and a negative feedback signal is applied through the feedback coil, L_{fb}, to cancel the change in flux through the SQUID. The feedback coil contains only 1/10th as many turns as the input coil, leading to a factor of 10 gain in the output current of the amplifier.

The phonon readout noise is determined by the Johnson noise of the shunt resistor, the noise current in the SQUID array, and the intrinsic noise of the TES, which is due to both fluctuations in the charge carriers (Johnson noise) and thermal fluctuations in the conductance between the electron system in the TES and the substrate. Detailed discussion of these theoretical noise sources for the CDMS TESs can be found in [227, 232, 249]. In practice, since the shunt resistor is heat sunk to the 600 mK stage rather than base temperature, Johnson noise from R_{sh} dominates, giving white noise at $\approx$15 pA/$\sqrt{\text{Hz}}$, rolled off by the $L_i/R_{TES} \approx$ 50–100 kHz time constant. Resonances in the SQUID arrays typically push the roll-off to higher frequencies, $f \sim$ 200–300 kHz. Given the phonon collection and observed signal sizes, this leads to baseline energy resolutions of σ = 0.1–0.2 keV$_{nr}$ for the Tower 1–3 (T1–T3) Ge detectors. The T4–T5 Ge detectors had smaller measured phonon amplitudes, possibly due to variations in the properties of the W or Al films, and gave σ = 0.3–0.5 keV$_{nr}$. The best phonon energy resolutions were obtained for the T2 Si detectors, which had baseline resolutions of σ = 0.05–0.1 keV$_{nr}$.

3.1.3 Yield discrimination

Since CDMS measures the ionization and phonon energy deposited by each interaction, it can discriminate between electron recoils and nuclear recoils on an event-by-event basis. This allows the identification of a nuclear-recoil signal amidst the much more prevalent electron-recoil backgrounds from residual radioactive decays in and around the detectors, which cannot be completely shielded. Typically, we define the "ionization yield" for each event as the ratio of the measured ionization energy, E_i, to the recoil energy E_r:

$$y = \frac{E_i}{E_r} \tag{3.3}$$

As discussed in Sec. 3.3.2, the ionization energy, E_i, is calibrated *in situ* using electron-recoil lines of known energy, and is reported in units of "keV$_{ee}$," which gives the recoil energy in keV for an electron recoil producing the same ionization signal as observed. The recoil

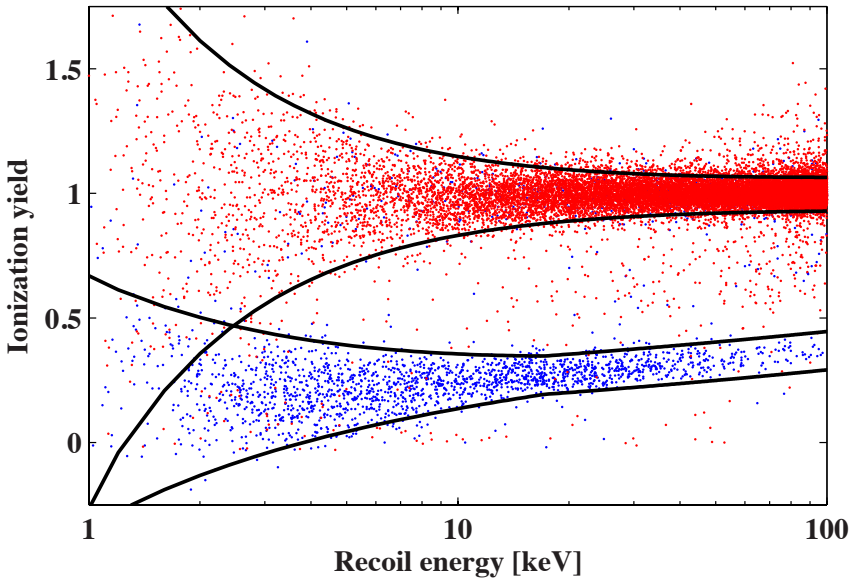


Figure 3.7: Ionization yield versus recoil energy for ^{133}Ba calibration data (red), which primarily consists of electron recoils from Compton scattering of γs, and ^{252}Cf calibration data (blue), which primarily produces neutron-induced nuclear recoils. The data shown were obtained with T1Z5, which has the lowest charge noise of the CDMS-II Ge detectors, leading to the best yield-based rejection of electron recoils at low energy. The black lines indicate the $\pm 2\sigma$ yield bands.

energy, E_r, is determined from the total phonon signal after accounting for the Neganov-Luke phonons following Eq. 3.2. Using these definitions, we have $y = 1$ for electron recoils.

In contrast, nuclear recoils produce a denser energy deposition in the detector, giving less ionization for the same recoil energy. An example of the ionization yield measured for nuclear recoils from the ^{252}Cf calibration source and electron recoils from the ^{133}Ba source are shown in Fig. 3.7. Clear separation between the bulk of the distributions is seen above 5 keV, although at low energies the populations begin to merge as the ionization signal becomes comparable to noise. For detectors with higher low-frequency noise, the merging of the electron-recoil and nuclear-recoil populations occurs at higher energy.

Figure 3.8 shows the fraction of electron-recoils from ^{133}Ba calibration data leaking into the 2σ nuclear recoil band as a function of recoil energy. At high energies, the yield-based discrimination is better than 10^3:1 for most detectors, and is dominated by a non-Gaussian tail of low-yield events below the bulk of the distribution. As discussed in the following section, these low-yield events are due to interactions very near the surface of the detector, where the ionization can be incompletely collected. Below 10 keV, the leakage fraction

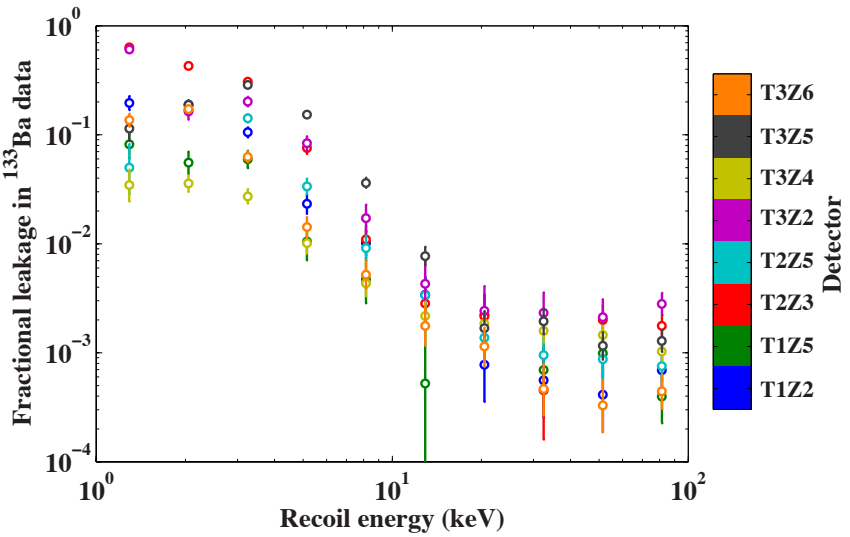


Figure 3.8: Fraction of ^{133}Ba calibration electron recoils lying within the $\pm 2\sigma$ nuclear recoil band for the 8 Ge detectors used for the low-threshold analysis described in Chap. 4.

begins to increase due to leakage from bulk electron recoils as the ionization signal becomes comparable to noise. Detectors such as T1Z5 and T3Z4 maintain >20:1 rejection down to 2 keV, while detectors such as T2Z3 and T3Z5, which have poorer ionization resolution due to microphonic pickup from the cryocooler, have significantly poorer rejection at low energy.

3.1.4 Surface events

Early investigations of the ionization collection in CDMS detectors indicated that a "dead layer" existed within 10–20 μm of the flat surfaces of the detectors, where ionization collection was found to be incomplete [234]. For events occurring in this dead layer, the local charge density immediately following the interaction can cause charges to diffuse and be collected in the "wrong" electrode before they can be drifted across the detector by the electric field due to the detector bias. In addition, defects present at the surface of the substrate could lead to a larger number of trapping centers in these regions. Although the addition of an amorphous Si layer between the substrate and electrodes was found to improve charge collection [234, 250], these poorly collected surface events remain the dominant background above 10 keV for CDMS-II.

The effect of the dead layer is clearly seen in a plot of the ionization yield for calibrations

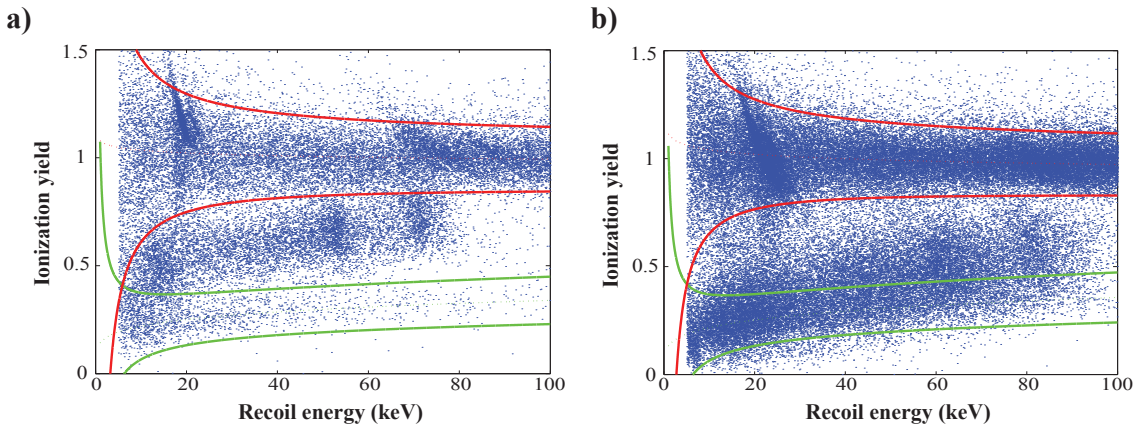


Figure 3.9: Calibration data for a collimated ^{109}Cd source positioned on the charge (a) and phonon (b) face of the detector, biased at $V_b = +3.0$ V. Fully collected bulk electron recoils appear in the 2σ electron-recoil band (red), including γ- and x-ray lines at 88 and 22 keV. Surface events, primarily due to the internally converted e^- from the source have reduced ionization collection and can leak into the 2σ nuclear-recoil band (green). These calibrations demonstrate the strong face-dependence of the yield, with phonon-face surface events typically appearing at lower yield than charge-face events. *Figure from V. Mandic [227]*

using a collimated ^{109}Cd source, as shown in Fig. 3.9. This source has x-ray and γ lines at 22 keV and 88 keV, as well as mono-energetic electrons at 63, 85 and 88 keV due to internal conversion of the γ. The relevant emissions are summarized in Table 7.2. The 88 keV *gamma* and 22 keV x-rays primarily interact in the bulk of the detector and lie around an ionization yield of 1. In contrast, the mono-energetic e^- interact within the dead layer and have suppressed ionization yield. In addition to the fully collected events, a broad energy distribution of low-yield events is seen due to back-scatter of e^- that do not deposit their full energy in the detector. Figure 3.9 also demonstrates that the majority of the surface events appearing at low yield occur on the phonon side of the detector. Comparisons of the yield distribution for the 22 keV x-rays to simulation indicate that the dead-layer is roughly twice as thick on the phonon side [227, 228] giving a larger fraction of phonon-side leakage into the nuclear recoil band than charge-side leakage.

Since CDMS measures the athermal phonon signal, the interaction location can be determined by the detailed partitioning of energy and relative delay of the signal in each sensor, as described in Sec. 3.1.5. In particular, for surface events, interactions with the metal films cause the high-energy recoil phonons to downconvert to ballistic phonons more quickly than the $\sim$5 μs time required for bulk phonons to undergo anharmonic decay [251,

252]. This leads to a phonon pulse which occurs more quickly after the charge pulse and with a sharper rising edge for surface events than for bulk recoils. In addition, nuclear recoils are found empirically to have slower rise times and longer delays on average than bulk electron recoils, likely due to their smaller Neganov-Luke phonon contribution. The smaller fraction of Neganov-Luke phonons has a significant impact on the rising edge of the pulse since these phonons are ballistic when they are emitted and do not need to undergo the downconversion process necessary for the primary phonons.

Using the detailed timing characteristics of the pulse, surface events leaking into the nuclear recoil band can be rejected with $\gtrsim$100:1 discrimination while maintaining $\sim$60% acceptance of nuclear recoils [153, 230]. Typically, a "timing quantity" is defined which combines the information from both the delay and rise time, and a discrimination cut is defined in the 2d yield versus timing plane, as shown in Fig. 3.10. As discussed above, most surface-events leaking into the nuclear recoil band occur on the phonon side of the detector. Luckily, since these phonon-side events occur very near the phonon sensors, they also tend to have faster timing, leading to improved timing-based rejection relative to charge-side surface events.

3.1.5 Position reconstruction

In addition to providing discrimination against interactions occurring near the detector surfaces, the athermal phonon signal allows the position of the interaction of be reconstructed using the relative partitioning of energy and timing delay between the phonon signal recorded in each sensor. We define the X and Y phonon partitions, x_p and y_p, and partition radius, r_p as:

$$x_p = \frac{\sum_i E_{p,i} x_i}{\sum_i E_{p,i}} = \frac{E_{p,c} + E_{p,d} - E_{p,a} - E_{p,b}}{E_{p,a} + E_{p,b} + E_{p,c} + E_{p,d}} \tag{3.4}$$

$$y_p = \frac{\sum_i E_{p,i} y_i}{\sum_i E_{p,i}} = \frac{E_{p,a} + E_{p,d} - E_{p,b} - E_{p,c}}{E_{p,a} + E_{p,b} + E_{p,c} + E_{p,d}} \tag{3.5}$$

$$r_p = \sqrt{x_p^2 + y_p^2} \tag{3.6}$$

where the index i runs over the phonon sensors $\{a, b, c, d\}$ located at the coordinates $\{(x_i, y_i)\} = \{(-1, 1), (-1, -1), (1, -1), (1, 1)\}$, respectively, and $E_{p,i}$ denotes the amplitude

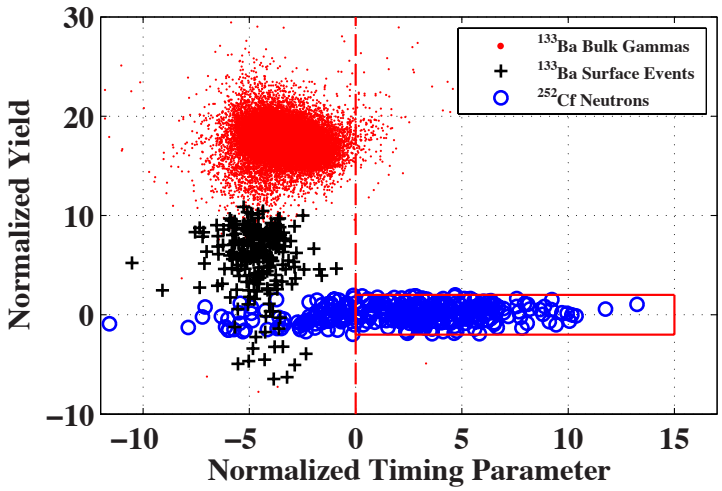


Figure 3.10: Combined surface event discrimination criteria using both ionization and timing for T1Z5. The vertical axis shows the ionization yield normalized to give the number of σ above or below the mean of the nuclear recoil distribution. Nuclear recoils from ^{252}Cf calibration data (blue, circles) lie at a normalized yield of 0, while bulk electron recoils (red, dots) are typically separated by $\sim$15 σ. Surface events with reduced ionization collection (black, pluses) can appear at low yield, but are separated from the bulk of the nuclear recoil distribution using a timing parameter describing the rise time and delay of the phonon signal relative to the charge signal. A signal acceptance region defined by the red box gives $\gtrsim$100:1 rejection of surface events in the 10–100 keV energy range, while maintaining $\sim$60% acceptance of nuclear recoils. *Figure from Ahmed et al. [153]*

of the phonon signal in the ith sensor. Similarly, the X and Y delays, x_d and y_d, and delay radius, r_d, are defined as:

$$x_d = \begin{cases} t_a - t_d & \text{for primary channel } a \text{ or } d \\ t_b - t_c & \text{for primary channel } b \text{ or } c \end{cases} \tag{3.7}$$

$$y_d = \begin{cases} t_b - t_a & \text{for primary channel } a \text{ or } b \\ t_c - t_d & \text{for primary channel } c \text{ or } d \end{cases} \tag{3.8}$$

$$r_d = \sqrt{x_d^2 + y_d^2} \tag{3.9}$$

where t_i is the time at which the phonon signal in the ith channel reaches 20% of its maximum height, and the primary channel denotes the channel with the largest value of $E_{p,i}$. Figure 3.11 shows an example of this reconstruction for collimated calibration data, demonstrating that the location of the collimator holes can be separately reconstructed

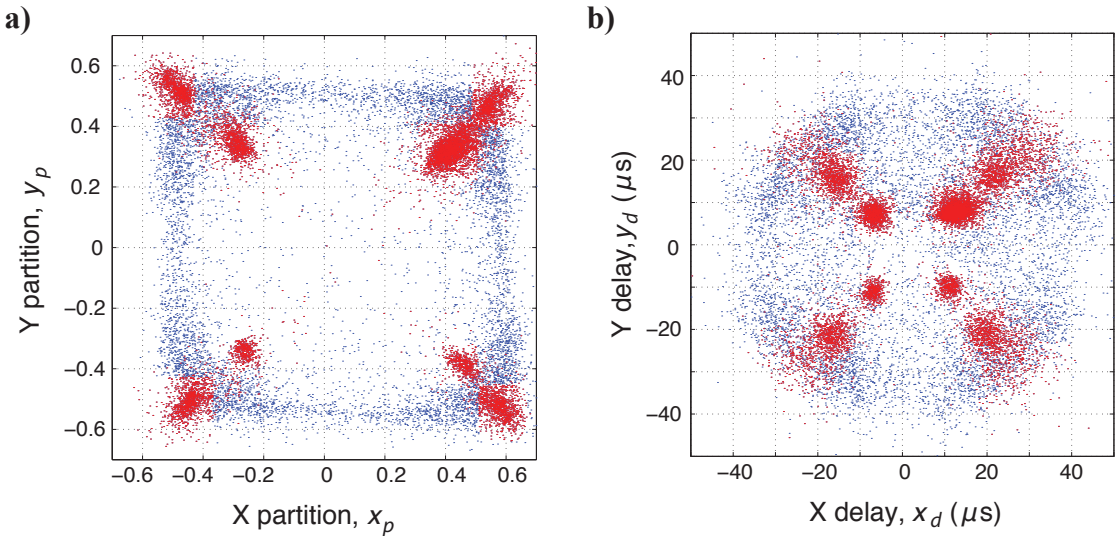


Figure 3.11: Position reconstruction for calibration data using the collimated ^{109}Cd source positioned on the phonon face of the detector. The position reconstructed from the partitioning of energy (a) and relative timing delays between the signal in each sensor (b) are shown. The red points highlight events within the spectral peaks of the ^{109}Cd source, demonstrating that the 8 collimated source positions along the 45° axes can be reconstructed, while the blue points lying away from the source emissions are background events uniformly distributed throughout the detector. *Figure from V. Mandic [227]*

using either the timing or partitioning information.

For events occurring near the outer cylindrical walls of the detector, phonons can be rapidly reflected of the wall and absorbed at lower radius, leading to a degeneracy between extremely high-radius and lower-radius events. As shown in Fig. 3.12, this degeneracy has different radial dependence for the partition and delay quantities, allowing the degeneracy to be broken and the physical position to be uniquely determined from a combination of these quantities. Figure 3.12 also shows that the primary timing quantity, which is used to discriminate between surface events and bulk events, has significant variation with position in the detector, even for calibrations with electron recoils alone. This position dependence is comparable to the variation in timing between surface electron recoils and nuclear recoils, leading to limited rejection of surface event backgrounds if the position-dependent effects are not taken into account. Reconstruction of the interaction location is vital for removing the position dependence of the timing discriminators and optimizing the background rejection of the ZIPs [230].

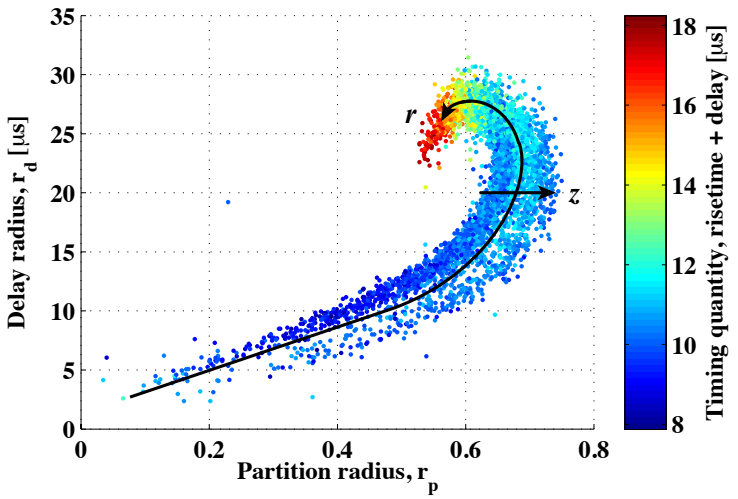


Figure 3.12: Plot of delay radius, r_d, versus partition radius, r_p, (known as a "shrimp plot") for electron-recoil calibration data from the ^{133}Ba source prior to position-dependent calibration. Although both r_d and r_p are seen to be double valued at high radius, they fold back differently, allowing the degeneracy to be broken by combining the radial information from both position estimators. Radial distance is measured along the curve of the "shrimp," while z-position can be inferred from the perpendicular direction, with phonon-face events at the same physical radius having larger values of r_p than for charge-face events. The points are colored by the combined timing parameter, demonstrating the significant radial and z-dependence of the pulse timing.

3.2 Experimental installation

3.2.1 Soudan Underground Laboratory

Since neutrons produce nuclear recoils that cannot be distinguished from WIMP-induced nuclear recoils on an event-by-event basis, a sensitive dark matter experiment requires extensive shielding to prevent neutrons from reaching the detectors. Neutrons can be of cosmogenic or radiogenic origin: cosmogenic neutrons result from showers containing neutrons induced by cosmic rays, while radiogenic neutrons result from radioactive decays in the materials surrounding the detectors. The cosmogenic component is especially troublesome because high-energy muons produced in atmospheric showers can be highly penetrating. These muons can interact in materials surrounding the detectors, producing neutrons that can mimic a WIMP signal. To sufficiently shield the experiment from particle showers produced by cosmogenic muons requires underground operation.

To shield from these cosmogenic neutrons, the CDMS-II experiment was operated in

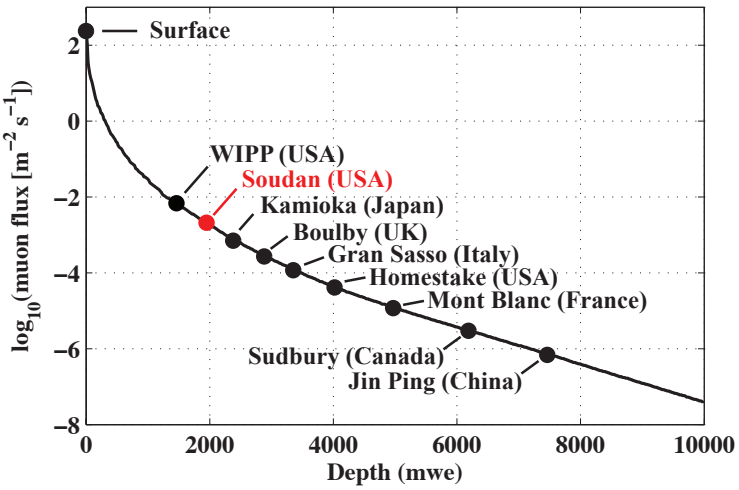


Figure 3.13: Muon flux versus depth for a selection of underground laboratories. At a depth of 2090 meters water equivalent (mwe), the muon flux at Soudan is suppressed by more than 4 orders of magnitude relative to the surface. Larger experiments will typically need to operate at deeper sites to reduce the muon-induced cosmogenic neutron background.

the Soudan Underground Laboratory (SUL) in northern Minnesota. The lab is on the 27th level of a decommissioned iron mine, at a depth of 714 m below the surface. The rock overburden at the SUL provides a cosmic ray flux that is equivalent to 2090 meters of water overburden, reducing the muon flux by a factor of 5×10^4 relative to the flux at the surface, as shown in Fig. 3.13. As discussed in Sec. 5.3.1, this is sufficient to reduce the expected cosmogenic neutron background to $\ll$1 event for the entire CDMS-II experiment. However, future experiments will require deeper sites such as SNOLAB in Sudbury, Ontario [253] to provide sufficient reduction of the muon background for more massive experiments.

3.2.2 Shielding

The remaining flux of cosmogenic neutrons at the SUL as well as radiogenic neutrons and γs are reduced to sufficiently low levels by several layers of passive and active shielding surrounding the experiment. The detectors themselves are housed in an inner cold volume known as the "icebox," which is connected to a commercial dilution refrigerator that sits outside the shielding through a horizontal "cold stem," as shown in Fig. 3.14. The outermost layer of the shield consists of an active scintillator veto made up of 40 overlapping panels that completely surround the detector volume and inner passive shielding. The scintillator thickness is 5 cm, allowing a minimum ionizing muon depositing 2 MeV g^{-1} cm^2 to deposit

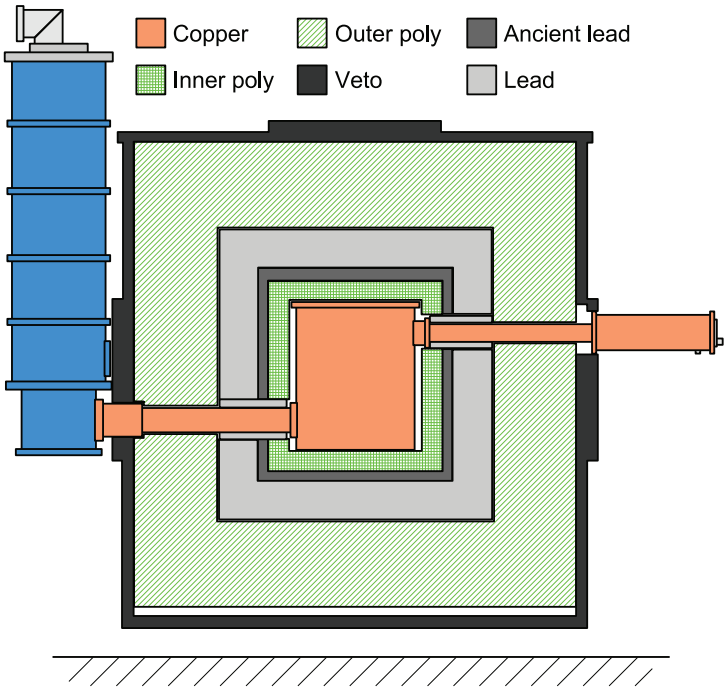


Figure 3.14: Cross-sectional view of active and passive shielding surrounding the icebox. The commercial dilution refrigerator sits outside the shield and is connected to the icebox through the cold stem. The icebox is surrounded (from inside to out) by an inner layer of polyethylene (poly), ancient lead, normal lead, a thicker outer layer of poly, and the active muon veto. *Figure from [232]; Original CAD from Susanne Kyre*

sufficient energy (∼10 MeV) that it can be distinguished from the more prevalent Compton scattering of radiogenic γ-rays that appear at lower energy ($\lesssim$ 2 MeV). This allows incoming muons to be tagged with ∼99.7% efficiency, while avoiding the large loss of livetime that would result if muons could not be distinguished from the 600 Hz rate of Compton scatters in the veto [254, 255].

Directly inside the active scintillator veto lie several layers of passive shielding: an outer layer of polyethylene, two layers of lead, and an inner layer of polyethylene, as shown in Fig. 3.14. The 40 cm thick outer polyethylene layer acts as a neutron moderator due to its composition of low-Z atoms, reducing the energy of external neutrons sufficiently that they do not produce nuclear recoils above the detector threshold. The lead serves to attenuate external γs and consists of an 18 cm thick outer layer in addition to an inner 4.5 cm thick layer of "ancient lead." Recently smelted Pb contains trace concentrations of ^{210}Pb, which has a half-life of 22.3 yrs and is a decay product of natural U in the

ores used to produce the Pb. In contrast, the ancient lead used in CDMS was recovered from a sunken ship near Nantes, France [256], and the ^{210}Pb has decayed sufficiently that the ^{210}Pb concentration is negligible. The innermost layer of polyethylene serves as an additional neutron moderator, primarily moderating neutrons produced from fission and (α,n) processes in the Pb shield. Without the inner moderator, neutrons that penetrated the Pb shield could scatter repeatedly without significant loss of energy giving them multiple chances to interact in the detectors.

To reduce the Compton scattering background from γs resulting from radon in the mine air, the volume within the Pb shield is continuously purged with N_2 boil-off from a dewar of liquid nitrogen. This purge was observed to reduce the Compton scatter rate from γs by a factor of >4 in the 10–100 keV region [228]. The icebox is additionally surrounded by a 2 mm thick mu-metal shield at room temperature that shields the inner volume from external magnetic fields which could affect the SQUID performance.

The icebox and inner concentric cans consist of oxygen-free high thermal conductivity (OFHC) copper, which provides low levels of radioactivity near the detectors. All materials within the icebox were screened for radioactivity with a high-purity germanium detector to measure or place limits on the level of contamination. Comparison of the observed γ spectrum with simulation indicates that ^{238}U and ^{232}Th in the inner concentric cans dominate the Compton rate at low-energy due to the large mass of the cans (260 kg) within the inner volume. The background measurement using the ZIP detectors themselves provides the most sensitive determination of the radioactive contamination of the experimental apparatus. Comparison with simulations indicates that the concentration of ^{238}U and ^{232}Th in the cans is 0.18±0.01 ppb and 0.56±0.06 ppb, respectively [229].

3.2.3 Cryogenics

The inner icebox volume is cooled by a commercial Oxford dilution refrigerator with a rated cooling power of 400 μW at 100 mK. Since the refrigerator itself is not constructed from radiopure materials, it is positioned outside the shielding and attached to the icebox through a series of concentric copper tubes that make up the cold stem. The fridge has several cold stages within the outer vacuum chamber (OVC), which provide isolation from room temperature to the ∼40 mK base temperature at the detectors. Each stage is connected to a corresponding Cu can within the icebox by the cold stem. An outer shield is maintained

at 77 K by a liquid N_2 bath, while an additional inner 4 K shield is maintained by the liquid He bath. The dilution refrigerator maintains the still can at ∼800 mK, the cold plate at ∼130 mK, and the mixing chamber at ∼40 mK.

The detectors are read out through a corresponding "electronics stem" which exits the icebox on the side opposite the cold stem. The electronics stem allows the detector wiring to exit the shield and is cooled by a Gifford-McMahon cryocooler at the 77 K and 4 K stages, without the need for an additional liquid cryogen bath. However, the vibrations introduced by the cryocooler were found to cause microphonic pickup on the charge channels for several detectors, as shown in Fig. 3.3.

The dilution fridge continuously cycles the ^{3}He-^{4}He mixture through the dilution unit, allowing long periods of stable operation at base temperature without warm-up. To prevent blockages of the impedances in the circulation loop, the mixture is cycled through external liquid N_2 and liquid He cold traps, as well as the internal liquid He cold trap supplied by Oxford, to remove any contamination that enters the mixture due to small, undetected leaks or outgassing of materials. The data analyzed in this thesis were taken during a series of 6 cryogenic runs between October 2006 and September 2008, separated by full or partial warm-ups of the dilution refrigerator to remove blockages or excess liquid He condensed in the icebox. Each of the 6 runs lasted from 2–6 months, and are denoted as cryogenic runs 123–128 (or R123–R128). Due to the time required to warm-up and recool the fridge and the possibility of detector malfunctions following thermal cycling, future work to prevent leaks an ensure stable operation over longer periods can have a significant impact on the livetime of the experiment.

3.2.4 Cold hardware

The detector support structure, wiring, and cryogenic amplifiers within in the icebox and electronics stem are known as the "cold hardware." The cold hardware components are assembled into "towers," as shown in Fig. 3.16a, which contain the detectors and cold readout electronics. The ZIPs are enclosed in hexagonal high-purity Cu housings, which are stacked in sets of 6 to form the lower portion of the towers. The housings within the tower do not have lids and are positioned so that the detectors face each other with ∼2.5 mm spacing. This close spacing between detector faces allows efficient tagging of multiple-detector scatters from surface events occurring on neighboring detector faces. The

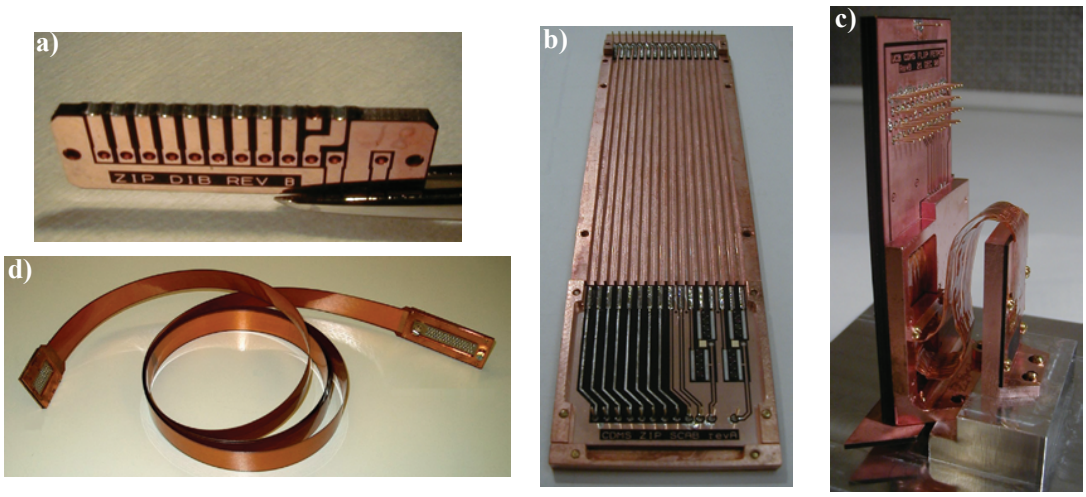


Figure 3.15: Photographs of various cold hardware components: a) DIB, b) Side coax, c) SQUET board, d) stripline. *Images from D. Seitz [257]*

top and bottom detectors in the stack, known as the "endcaps," do not have a neighboring detector on their outer face and have correspondingly poorer rejection of multiple-scatter surface events occurring on their exterior face.

Each detector is connected to a detector interface board (DIB), shown in 3.15a, which is attached to the inside of the detector housing along the substrate flat and provides wirebonding pads that are connected directly to the bias lines on the detector. Two infrared LEDs, which are used to neutralize the detectors as described in Sec. 3.1.1, are also soldered to the top of the DIB, illuminating the phonon face of its detector and the charge face of the detector above. Above the detector stack, the upper portion of the tower is sunk to each of the successive cold stages (mixing chamber, cold plate, still, 4 K bath) as shown in Fig. 3.16a, and thermally isolated from the remaining stages by a central graphite support structure. Radiation shields within the tower minimize infrared radiation emitted at the upper stages from reaching the detectors at base temperature.

A "side coax," shown in Fig. 3.15b, contains 16 vacuum coax lines that connect each of the DIBs at base temperature to the "SQUID and FET" (SQUET) cards mounted at 4 K. The side coax also contains the coupling capacitors and bias and feedback resistors for the ionization readout (see Fig. 3.2). Heat sinking the resistors at base temperature reduces their Johnson noise contribution to the ionization readout. The SQUET cards, shown in Fig. 3.15c, are positioned at the top of the tower and consist of two separate components: a

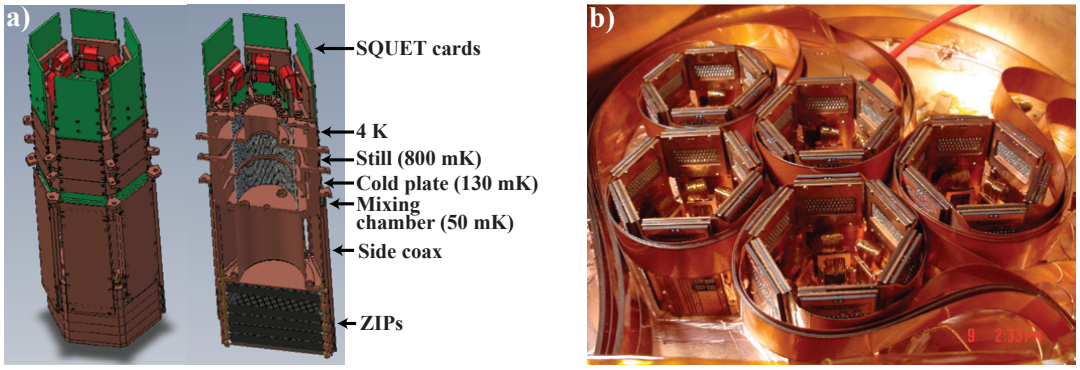


Figure 3.16: a) CAD drawing of assembled tower and cross-sectional view specifying the temperature stages and cold hardware components. *CAD drawings by P. Wikus.* b) Photograph showing the 5 towers installed in the Soudan icebox. The SQUET cards at the tops of the towers and striplines are visible above the 4 K radiation shield.

board housing the JFET charge amplifiers heat sunk to the 4 K stage and a smaller board housing the SQUIDs at the still stage. The FETs themselves sit on a Kapton membrane which allows them to self-heat to ∼140 K, where the optimal noise (∼0.5 nV/$\sqrt{\text{Hz}}$ at 10 kHz) is obtained [227]. The SQUIDs are heat sunk to the still stage at ∼800 mK to reduce the Johnson noise of the shunt resistors and the SQUID noise.

The upper portion of the SQUET card connects to the "striplines," shown in Fig. 3.15d, which provide a connection to the room temperature electronics through the electronics stem. Each stripline is a flexible, 3 m long ribbon cable consisting of copper traces sandwiched between dual ground planes with Kapton as the dielectric. The striplines are heat sunk at 4 K and 77 K and connected to 50-pin hermetic connectors on the electronics stem breakout box (e-box) at room temperature.

3.2.5 Warm electronics

The "warm electronics" take the signals exiting the e-box, provide second-stage amplification and signal conditioning, wait for an event trigger, and digitize and record the corresponding detector timestreams. The front end boards (FEBs) connect directly to the 50-pin connectors on the e-box and house the second-stage amplifier chain and biasing circuits for the cold electronics. The output of these FEBs are connected to receiver-trigger-filter (RTF) boards, which removes baseline offsets for the phonon channels and applies an analog 336-kHz Butterworth anti-aliasing filter. A summed charge pulse is generated from the inner

and outer charge channels, while a summed phonon pulse is generated from the 4 phonon inputs by the RTF board. These summed traces are used to issue triggers according to 5 thresholds set by software at run-time: P_{hi}, P_{lo}, Q_{hi}, Q_{lo}, and $P_{whisper}$, where P and Q denote trigger levels for the phonon and charge traces, respectively. For the 5-tower data, P_{lo} was typically set at 10–15σ above the baseline phonon noise to prevent spurious triggers from time-dependent noise and was the only trigger threshold used to determine whether to record an event. The *hi* triggers provide a somewhat higher threshold, while the *whisper* trigger provides a lower threshold. Although only the P_{lo} trigger is used for data acquisition, all trigger information is recorded and stored for each event.

The trigger logic board (TLB) accepts signals from the RTF boards and corresponding signals from the active scintillator veto and generates a "global trigger" if either the P_{lo} trigger is issued for any detector, the veto issues simultaneous triggers in multiple panels, or a random trigger is requested by the data acquisition software (DAQ) to monitor noise. When a global trigger is issued, the DAQ computers read out the digitized signal for each detector channel and veto channel. The phonon and charge channels are digitized by 14-bit Struck SIS 3301 analog-to-digital converters (ADCs), which provide a sampling rate of 1.25 MS/s and buffer a 2048 sample long time stream with the trigger at the 512th sample. Similarly, the veto signals are recorded by 12-bit Joerger VTR812 ADCs, which record 1024 samples at 5 MS/s.

The trigger information is separately held in a circular buffer so that, for each event recorded to disk, a record of the times and hit pattern for the 4 triggers immediately preceding the event as well the 5 triggers immediately after are stored. A slower DAQ system records phonon offsets and mean trigger rates as well as various experimental conditions and stores these data to disk separately throughout the data taking.

3.2.6 Data acquisition

The Soudan DAQ, described in detail in [254], is controlled by a "run control" server which manages the configuration and operation of the experiment, an "event builder," which acquires and records the data for each event, and a run control GUI which allows control and configuration of the experiment. Each component is implemented in Java or C++ and runs on dedicated servers within the electronics room at the SUL. The run control GUI is a Java-based network application which can be used to monitor the experiment remotely,

or directly control the experiment from the local network at the mine.

The largest overhead during data acquisition results from acquiring traces and recording them to disk. For the 5-tower operation, the maximum rate at which all the traces for each event could be recorded was 20 Hz, corresponding to a data rate of 12 MB/s. Although this is well above the ∼0.1 Hz trigger rate recorded in WIMP search data, it limits the rate of calibration events that can be recorded, increasing the time which must be spent acquiring calibration data. To improve the data rate during the high-rate ^{133}Ba calibrations, the DAQ was operated in "selective readout" mode, where traces were recorded only for detectors that issued a P_{lo} trigger. In this configuration, data rates of up to 70 Hz are possible, reducing the time needed for calibration by a factor of 3.5. For the WIMP search data as well as the lower-rate ^{252}Cf neutron calibrations, data is recorded in full readout mode, where traces are read out for each detector regardless of trigger information to ensure that particle-induced events falling under the trigger thresholds are recorded.

3.3 Analysis pipeline

The analysis pipeline takes the raw traces recorded by the DAQ for each event and produces reconstructed quantities describing the pulse characteristics that are ultimately used for identifying WIMP interactions. These include estimating the energy deposited as phonons and ionization, the location of the interaction inferred from the partitioning of phonon energy and relative timing delays between sensors, and the rise time of the phonon signal and delay relative to the charge pulse, which are used to distinguish surface from bulk events.

3.3.1 Processing pipeline

The data recorded to disk by the DAQ at Soudan are compressed, backed up to tape in the mine, and transferred to Fermilab for the primary data processing. Data were processed with one of two processing packages, `DarkPipe` for R123–124 or `BatRoot` for R125–128. These packages have equivalent functionality, although the latter was implemented using C++/ROOT rather than Matlab to take advantage of the efficient ROOT Ntuple format [258] and reduce processing time for the larger R125-128 data set. This initial processing reconstructs several basic characteristics of the charge and phonon pulses:

1. **Optimal filter reconstruction of charge amplitudes:** The amplitudes of the charge pulses for the inner and outer channels are reconstructed simultaneously using a 2-component template for each pulse [228, 259]. This procedure takes into account the cross-talk between the inner and outer charge channels and produces an amplitude estimate that maximizes the signal-to-noise given the fixed pulse shape set by the charge readout circuit and the measured noise for each data series. Since the start time is not known, a search within $[-100, +10]$ μs of the trigger for Ge ($[-50, +10]$ μs for Si) is performed, and the start time that maximizes the summed reconstructed inner and outer amplitudes is used.

2. **Time domain fit to charge amplitudes:** A second, time domain fit to the charge pulse is performed for highly saturating events, excluding samples which saturate the ADC. This reconstruction is primarily used to identify high-energy α interactions in order to constrain contamination of the detectors and surrounding materials.

3. **Optimal filter reconstruction of phonon amplitudes:** Phonon amplitudes are reconstructed using a single template optimal filter algorithm. Unlike the charge channels, the pulse shape is not fixed and varies significantly with position in the detector, leading to 10–20% systematic variations in the reconstructed pulse amplitude with position. As described in Sec. 3.3.3, these systematic variations are accounted for by applying a position-dependent calibration of the phonon energy [230].

4. **Nonparametric reconstruction of phonon pulse shape:** While the charge start time is determined directly from the charge optimal filter, the phonon pulse start time is determined with a time-domain algorithm due to variations in the pulse shape relative to the optimal filter template. This algorithm walks along the trace and determines the last-crossing time for a series of thresholds at 10, 20, 30,...,100% of the maximum pulse amplitude on the rising and falling edges of the pulse. The pulses are filtered with a 50 kHz low-pass Butterworth filter (90 kHz for Si) to improve the signal to noise on the few μs time scales of interest. A filter whose cutoff frequency varied with signal-to-noise was also implemented in `BatRoot`, although no significant improvement in the ability to reject surface events with the variable frequency filter was observed [230].

5. **Time domain fit to phonon pulses:** `BatRoot` also provides a reconstruction of the pulse amplitude, start time and rise time determined from a time-domain fit to each phonon pulse. This procedure accounts for the phonon pulse shape variation with position and provides significantly less position variation in the reconstructed pulse amplitudes than the optimal filter fit, prior to position-dependent calibration. It also provides improved signal-to-noise in the reconstructed pulse timing at the cost of potential systematic errors arising from misparameterization of the pulse shape.

3.3.2 Preliminary charge energy and position calibration

Following initial pulse reconstruction by `DarkPipe` and `BatRoot`, the charge and phonon energies are calibrated in physical units using calibration lines of known energy from an *in situ* ^{133}Ba source. An example of the charge energy and position calibration for a Ge detector is shown in Fig. 3.17. Figure 3.17a shows the uncalibrated charge energies for data from the ^{133}Ba source, as a function of Y-position in the detector reconstructed from the phonon delay, y_d. Variations of $\sim$5–10% with position, which are typically strongest in the Y-direction, are observed in the reconstructed energy of the 356 keV ^{133}Ba line. The position of this line as a function of x_d and y_d is fit to a polynomial form and the position dependence is normalized out. The overall amplitude of the charge signal is scaled to physical units using the ^{133}Ba spectral features at 356 keV and 384 keV.

At energies $\lesssim$10 keV, the X and Y delays cannot be reliably determined and the charge position correction is not effective. However, at these energies read out noise begins to dominate over the $\lesssim$10% systematic position variations, so application of the charge position correction would not improve the reconstructed resolution even with improved position estimators at low energy. As discussed in Sec. 4.2, the final energy calibration for the Ge detectors at low energy directly uses the position of the 1.3 keV$_{ee}$ and 10.4 keV$_{ee}$ lines present after the activation of the detectors by ^{252}Cf calibrations.

Calibration of the Si detectors is more challenging since no corresponding electron-recoil activation lines exist above threshold and because only a tiny fraction of 356 keV γ interactions are fully contained within the lower-mass Si detector modules. The initial charge calibration in Si is performed by summing the energy in a given Si detector with its precalibrated Ge nearest neighbors to find the location of the 356 keV peak for events that are fully contained in the set of neighboring detectors. However, direct calibration of

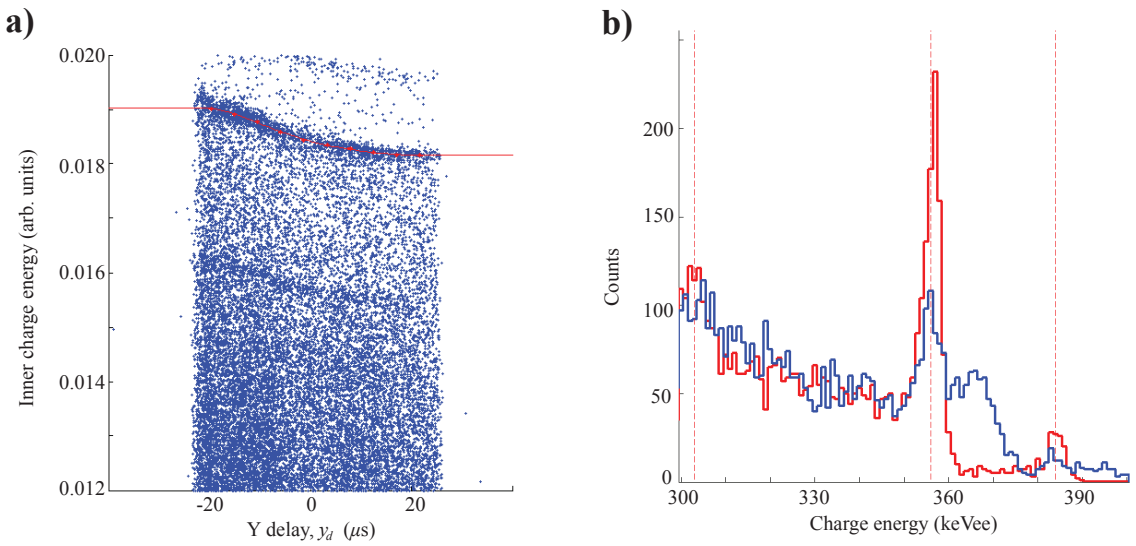


Figure 3.17: a) Position-based calibration of reconstructed charge energy for T1Z5. The 356 keV ^{133}Ba line appears as a concentration of events above which only the weaker 384 keV line is seen. The position dependence, which is strongest in the y-direction, is fit to a polynomial (red). b) Observed charge energy spectrum before (blue) and after (red) applying the position-dependent charge calibration and preliminary absolute calibration. The dashed lines show the location of the expected γ emissions from the source. After correction, a clear 356 keV peak is observed. *Figures from K. Sundqvist*

the nuclear-recoil energy scale in Si is also possible using a strong resonance in the ^{28}Si neutron scattering cross section, which appears as a corresponding feature in the ^{252}Cf calibration spectrum just above 20 keV [255, 260]. Constraints on the absolute calibration of the nuclear-recoil energy scale using this feature will be presented in [261].

3.3.3 Preliminary phonon energy and position calibration

The preliminary phonon energy calibration consists of two stages. First, a relative calibration between the 4 phonon sensors is performed using ^{133}Ba calibration data. For the relative calibration, the distributions of the phonon partition in each sensor are aligned to ensure that on average all sensors contribute equally to the summed phonon energy, as shown in Fig. 3.18a. The summed energy is then scaled using the precalibrated ionization energy to ensure that the mean ionization yield is equal to unity for electron recoils at ∼50 keV, as shown in Fig. 3.18b.

After this preliminary calibration, significant variations (10–20%) in the reconstructed energy with position and energy remain. In previous CDMS analyses that used ∼10 keV

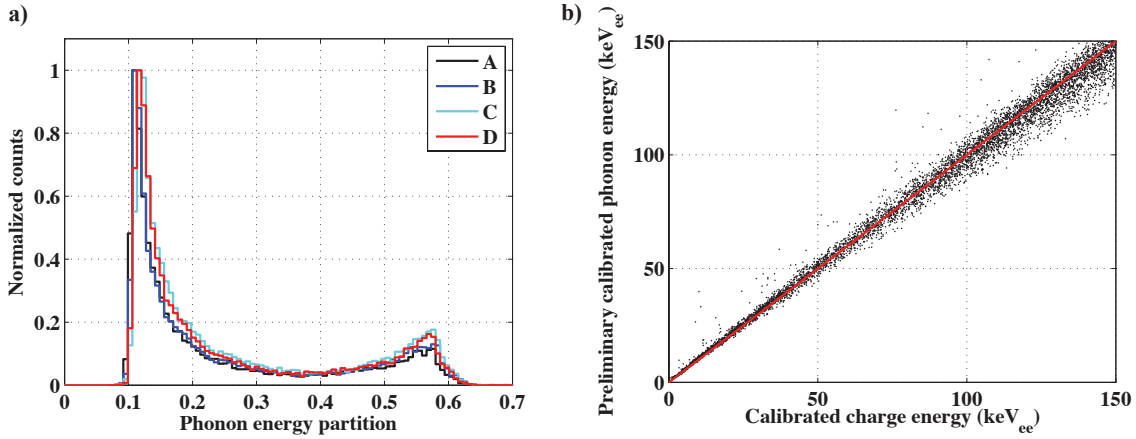


Figure 3.18: a) Relative phonon calibration. The distribution of the fractional energy in each sensor is normalized so that the peak for primary channel events at an energy partition of 0.55 is the same for all channels. b) Preliminary absolute phonon calibration. The phonon energy is matched to the precalibrated charge energy so that the ionization yield is unity for electron recoils. Data shown are for T1Z5.

thresholds (e.g., [153, 178, 256]), a position and energy-dependent correction [230] was applied to remove systematic variations in the phonon response and reconstruction. For energies $\lesssim$10 keV, the delay-based position estimators have poor signal to noise and the position-based correction no longer improves the resolution. For the low-energy analysis described in Chap. 4, only an energy-dependent calibration is applied following the preliminary calibration to remove nonlinearities in the phonon response with energy. This leads to an energy scale which is calibrated within 5% of the true energy in the 0-10 keV energy range. As for the charge energy, the final phonon energy calibration for Ge is performed using the position of the 1.3 keV and 10.4 keV activation lines, and is described in Sec. 4.2.

Chapter 4

Low-threshold analysis

Data taken with the full 5-tower installation at Soudan was previously analyzed using a 10 keV recoil energy threshold for the Ge detectors [153, 178, 228, 230]. This threshold was chosen to ensure that the expected background from surface event leakage was ≈0.5 events, increasing the discovery potential of the experiment since several events passing cuts would provide significant evidence for a WIMP signal without the need to develop a detailed background model that could suffer from significant systematics. For the $\gtrsim$50 GeV/c^2 WIMPs typically considered in the cMSSM, a 10 keV recoil energy threshold does not significantly reduce the sensitivity of the analysis since a large fraction of the velocity distribution can still produce recoils above threshold. However, for WIMPs with masses $\lesssim$20 GeV/c^2 this threshold is not optimal. Figure 4.1 shows the expected interaction rate for a 7 GeV/c^2 WIMP in Ge and Si as a function of recoil energy. With a 10 keV threshold, previous Ge analyses had no sensitivity to such models since the expected recoil energies fell below the analysis threshold.

The trigger thresholds (defined as the energy at which the triggering efficiency =50%) for the best Ge detectors were typically 1.5–2.5 keV. This allows sensitivity to particle interactions with recoil energies as low as 1–2 keV. An example of a 2 keV nuclear recoil phonon pulse is shown in Fig. 4.2, demonstrating that signal-to-noise of ≈10 is maintained in the phonon channels down to the trigger threshold. At energies below 10 keV, significant backgrounds are expected, as discussed in detail in Sec. 5.3. However, given the high expected rate for the model shown in Fig. 4.1 ($\sim$1 event per kg day), CDMS should still be able to provide interesting constraints on these models as long as backgrounds are lower than this rate.

In the following sections, we present the detector and data selection (Sec. 4.1), energy-

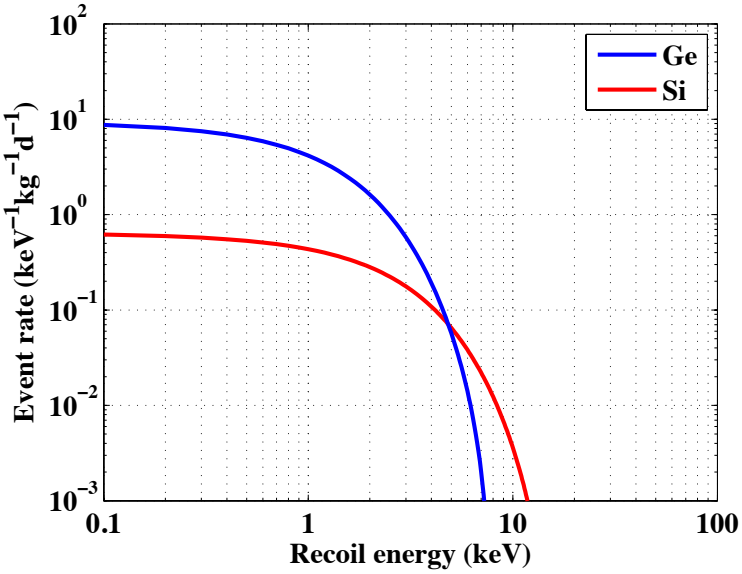


Figure 4.1: Event rate in Ge and Si for a $m_\chi = 7$ GeV/c^2 WIMP with cross section $\sigma_{SI} = 1.5 \times 10^{-40}$ cm^2. This WIMP model is consistent with the combined fit to the DAMA/LIBRA and CoGeNT signals from [82]. The standard halo model described in Sec. 2.1 is assumed, with an escape velocity of $v_{esc} = 544$ km/s and a characteristic velocity of $v_0 = 220$ km/s.

scale calibration (Sec. 4.2), reconstruction and data quality criteria (Secs. 4.3–4.4), and optimized ionization-based discrimination cuts (Sec. 4.5) used for a reanalysis of the Soudan 5-tower data with a 2 keV recoil energy threshold. The nuclear recoil candidate events resulting from this analysis and constraints on low-mass WIMPs will be presented in Chap. 5. This work builds on techniques developed for a previous low-threshold analysis of data taken at the shallow site [186, 255] developed by R. Bunker, by extending these techniques to the deep site and optimizing the sensitivity to WIMPs with masses from 5–10 GeV/c^2.

4.1 Detector and data selection

Since a reanalysis of CDMS data with a 2 keV recoil energy threshold will be background limited, the detectors with the lowest backgrounds will dominate the results and additional sensitivity is not gained by adding exposure from detectors with higher backgrounds. For this reason, only a subset of the detectors that were used in the analyses of the Soudan data with a 10 keV threshold are included in the low-threshold analysis presented here. As shown in Fig. 4.1, for an energy threshold of 2 keV, the Ge detectors have a higher expected

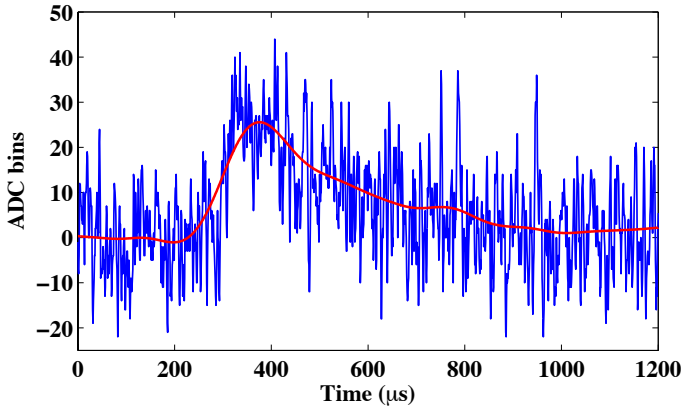


Figure 4.2: Total phonon signal for an example 2.0 keV nuclear recoil in T1Z5. The blue line shows the digitized pulse height versus time, while the red line shows the resulting trace after applying a 20 kHz low-pass filter.

rate per kg-day of exposure for WIMPs with masses from 5–10 GeV. In addition, the total background rates between the Ge and Si detector modules at low energy are expected to be similar, but the Si detectors are only ∼50% as massive, leading to higher relative backgrounds for the Si detectors per kg-day of exposure. Since we are mainly interested in constraining WIMP models with masses from 5–10 GeV/c^2 (the region of interest for the potential signals in the DAMA/LIBRA, CoGeNT, and CRESST-II experiments, as described in Sec. 2.1.4), only Ge detectors were considered for the analysis in this thesis. At WIMP masses below 5 GeV/c^2, Si can provide stronger constraints than Ge due to the higher recoil energies that can be transferred to the lighter Si nucleus. Although the Si energy scale is more difficult to calibrate at low energy (see Sec. 3.3.2), future analyses may be able to use the Si data to improve constraints for WIMPs with masses <5 GeV/c^2.

To determine the detectors with the best expected sensitivity to WIMPs with masses from 5–10 GeV/c^2, both the phonon energy resolution, which determines the trigger threshold, and the charge energy resolution, which determines the background rejection at low energies, were considered. The energy resolutions were measured by fitting the distribution of reconstructed energies for randomly triggered noise traces. To avoid bias due to the optimal filter search algorithm, the resolutions were calculated for fits with the pulse start time fixed at the trigger time. The reconstructed resolutions for the total phonon signal, p_t, and the charge energy for the inner electrode, q_i, are shown in Fig. 4.3.

As discussed in Sec. 3.1.2, the Tower 4–5 (T4–T5) Ge detectors have smaller phonon

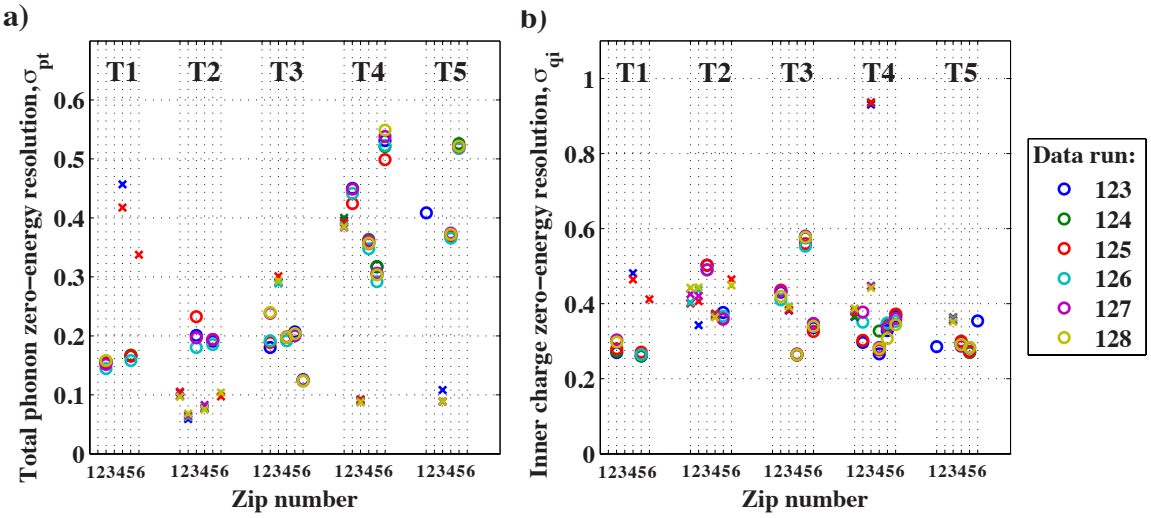


Figure 4.3: Measured baseline energy resolution, σ, for the total phonon (a) and inner charge (b) readout channels on each of the 30 ZIP detectors. Ge detectors are denoted by open circles while Si detectors are denoted by crosses. The resolutions are calculated from optimal filter fits with the template start time fixed at the trigger time to avoid the positive bias from the optimal filter search algorithm.

pulse amplitudes than the T1–T3 detectors, leading to poorer phonon energy resolution for these detectors shown in Fig. 4.3. For this reason, we use only the T1–T3 Ge detectors for this analysis. The ionization resolution for the T1–T3 detectors varies from 0.25–0.6 keV$_{ee}$, primarily due to variations in the low-frequency pickup of the charge readout channels, as shown in Fig. 3.3. Although the high-noise detectors (e.g., T2Z3 and T3Z5) have higher expected backgrounds, we do not exclude them from the analysis at this stage but rely instead on our limit-setting procedure to take into account backgrounds varying with detector (see Sec. 5.2).

For this analysis, we consider the data taken with the full 5-tower detector installation during cryogenic runs R123–128. These data span a $\sim$2 year time period from October 2006 through September 2008. As in [153, 178], T1Z1 and T1Z3 are excluded from the entire data-taking period due to detector malfunctions that prevented the readout of all charge and phonon channels. This leaves 8 Ge detectors in T1–T3 that were used to search for WIMP interactions. Some of these detectors were eliminated for certain cryogenic runs due to problems with calibration, neutralization, or failure of detector readout channels following cryogenic cycling. The runs in which each detector was operated are summarized in Table 4.1. All 30 detectors were used to identify particle interactions and reject events

Detector:	Mass (g):	Runs used:
T1Z2	228	R123–128
T1Z5	219	R123–126
T2Z3	219	R123, R125–127
T2Z5	239	R123, R125–127
T3Z2	231	R123, R125–128
T3Z4	239	R125–128
T3Z5	239	R123, R125–128
T3Z6	232	R123, R125–128

Table 4.1: List of detectors used to search for WIMP interactions for the low-threshold analysis. The mass of each detector is given [229, 230] as well as the data-taking runs in which each detector was operational.

with energy deposited in multiple detectors, as described in Sec. 4.3.

4.2 Energy scale calibration at low energy

Accurate calibration of the energy scale at low energies is crucial since the expected WIMP spectrum is exponential, and small errors in the energy scale can lead to larger errors in the estimated cross section. For this analysis, only the phonon energy was used to reconstruct the energy of the interaction to avoid incorporating the poorer signal-to-noise of the charge measurement at low energy [186, 255]. The energy is then estimated from the total phonon signal alone following Eq. 3.2. Since the charge signal is not used on an event-by-event basis, one must specify how much energy was deposited as Neganov-Luke phonons to correctly determine the recoil energy. This leads to two possible energy scales for the same total phonon signal, p_t, depending on recoil type. Electron recoils, which have a larger Neganov-Luke phonon contribution, will have a correspondingly larger total phonon signal than nuclear recoils of the same recoil energy. Assuming an event is an electron recoil, then the recoil energy, E_r, in units of keV$_{ee}$ is given by:

$$E_r = \frac{p_t}{2} \tag{4.1}$$

where we have used that:

$$E_r(p_t) = p_t - eV_bN_Q = p_t - \frac{eV_b}{\epsilon}E_Q \tag{4.2}$$

for $eV_b/\epsilon = 1$ for $V_b = 3.0$ V as discussed in Sec. 3.1.1. As in Sec. 3.3.2, the charge energy, E_Q, is calibrated to give the recoil energy, E_r, for an electron recoil that produces the same charge signal. Thus, $E_r = E_Q$ for electron recoils, and $p_t = 2E_r$ by definition of the calibration procedure.

If an event is instead assumed to be a nuclear recoil, then a smaller correction for the Neganov-Luke phonon contribution to the total phonon signal must be applied. The mean ionization energy for nuclear recoils, $\mu_{Q,NR}(p_t)$ is determined over the energy range from 2–100 keV using the distribution of nuclear recoils in ^{252}Cf calibration data, as described in Sec. 4.2.2. The recoil energy, assuming a nuclear recoil, in units of keV$_{nr}$ is then given by:

$$E_r(p_t) = p_t - \mu_{Q,NR}(p_t) \tag{4.3}$$

4.2.1 Calibration with Ge activation lines

An advantage of using Ge detectors for low-energy WIMP searches is that the phonon energy scale for electron recoils can be directly calibrated using activation lines at known energy that are present following calibrations with the ^{252}Cf source. ^{71}Ge is produced by thermal neutron capture during neutron calibrations and decays by electron capture to ^{71}Ga with a half-life of ∼11 days. Most of the time, this decay proceeds via capture of a K-shell electron, followed by emission of 10.36 keV$_{ee}$ of energy in x-rays and Auger electrons corresponding to the Ga binding energy. A smaller fraction of the time (∼10%), the decay proceeds via an L-shell capture, giving 1.29 keV$_{ee}$ in energy. Figure 4.4 shows the electron-recoil spectrum observed for each detector within 20 days following calibrations with the ^{252}Cf source. The Ga K-shell electron capture line is visible on all detectors, while the L-shell line is visible for the detectors with the lowest trigger thresholds.

The energy scale is calibrated for both ionization and phonons to ensure that the position of these lines, given by the mean of Gaussian fits to the observed counts, is not underestimated at the 90% confidence level. To perform this calibration, the energy scale is first roughly calibrated using the 356 keV$_{ee}$ line from the ^{133}Ba source, as described in Secs. 3.3.2–3.3.3. The position of the activation lines following this rough calibration are then fit to a Gaussian distribution to determine the mean energy, μ, and the one-sided 90% confidence interval for μ. A normalization is applied (typically 2–4%, depending on detector) to rescale the energies so that the lower edge of this 90% confidence interval is at 10.36

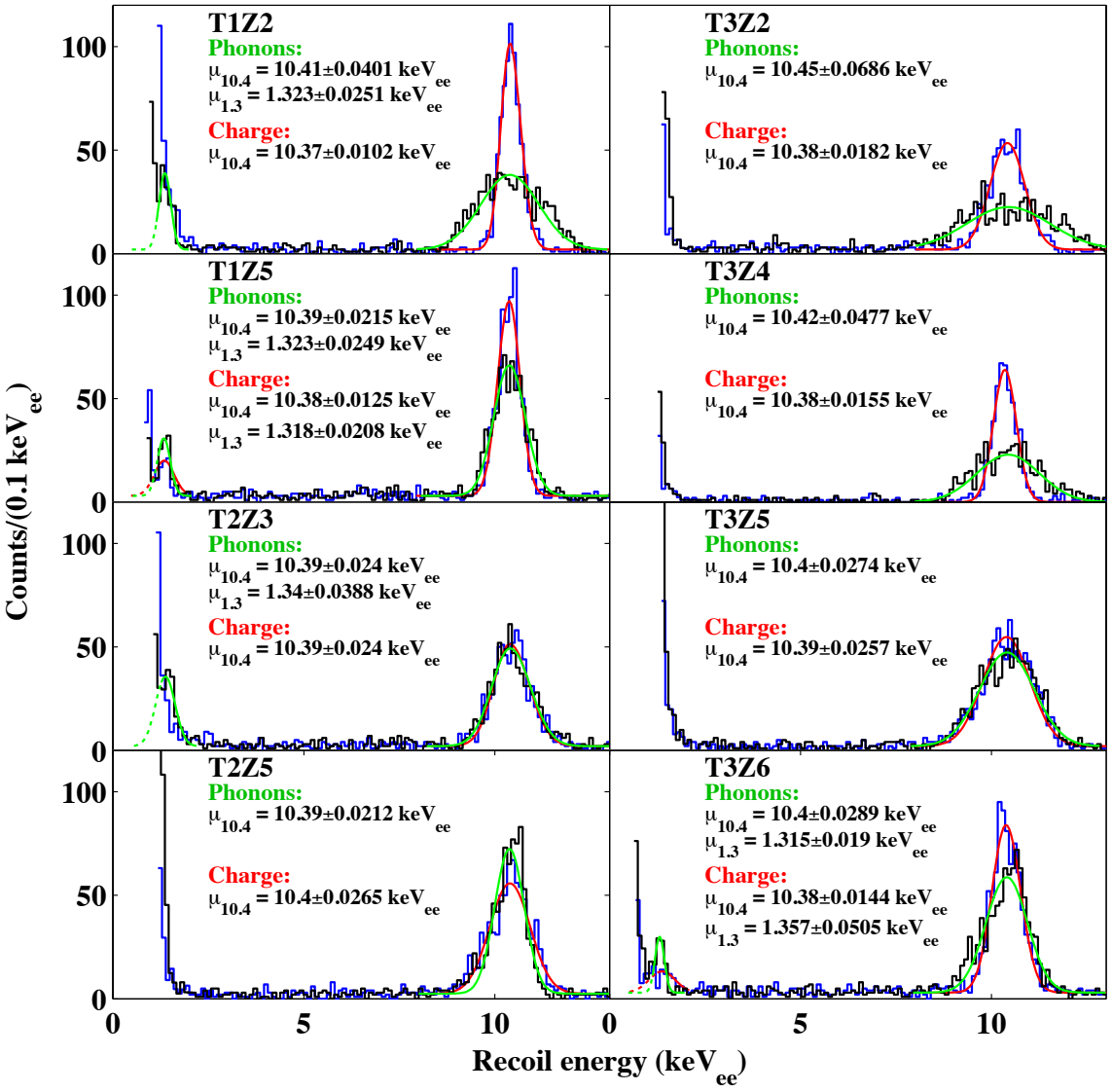


Figure 4.4: Electron-recoil spectrum for events within 20 days following a calibration with the ^{252}Cf source. The phonon-based recoil energy scale uses the phonon signal alone, assuming the Neganov-Luke phonon signal is consistent with an electron recoil. Both the charge-based and phonon-based energies are calibrated for each detector using the activation line at 10.36 keV$_{ee}$. For the detectors with the lowest trigger thresholds, the 1.29 keV$_{ee}$ activation line is visible, and is also used to calibrate the scale at low energies. The legend shows the measured position of the line and statistical uncertainty, after calibrating the energy scale to ensure the position of each line is not underestimated at the 90% CL.

and 1.29 keV$_{ee}$ for the K and L-shell lines, respectively. The spectrum after performing this calibration is shown in Fig. 4.4 for the 8 Ge detectors considered at low energy.

Figure 4.5 shows the time dependence of the counts observed for each line as a function

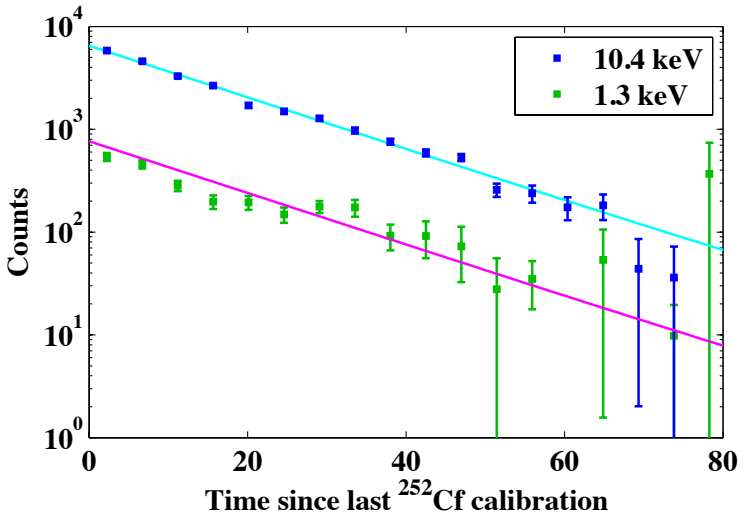


Figure 4.5: Measured counts for the 10.4 keV$_{ee}$ (blue) and 1.3 keV$_{ee}$ (green) lines versus time since the most recent ^{252}Cf calibration, after subtracting the constant Compton background rate measured from 3–8 keV$_{ee}$. A correction for the trigger efficiency has been applied to normalize the measured rate for the 1.3 keV$_{ee}$ line. A combined fit (cyan, magenta) to the data for both lines gives a measured half life of $t_{1/2} = 17.33 \pm 2.92$ days.

of time since the most recent ^{252}Cf calibration, summed over all detectors. For each time bin, the counts in a window around the line position are determined, and the corresponding rate from the constant Compton background is removed. A combined fit to the counts for the 10.4 and 1.3 keV lines versus time is performed to determine the half life with which the activation decays away. Only a single $t_{1/2}$ is fit, and the ratio of line amplitudes are fixed to the known K/L ratio $= 0.1175 \pm 0.002$ [262]. As shown in Fig. 4.5, the observed intensity ratio is in good agreement with this value. The best-fit half life was found to be $t_{1/2} = 17.33 \pm 2.92$ days. This value is $\approx 2\sigma$ higher than the known value of $t_{1/2} = 11.43 \pm 0.03$ days for ^{71}Ge [263], possibly indicating the presence of some residual cosmogenically activated ^{68}Ge with a longer half life of 271 days, which was not separately deconvolved.

4.2.2 Constraints on the nuclear-recoil energy scale

The nuclear-recoil energy scale is reconstructed from the phonon energy scale after applying the above calibration from the electron-recoil activation lines and accounting for the difference in the Neganov-Luke phonon signal following Eq. 4.3. This requires the determination of $\mu_{Q,NR}(E_r)$, which gives the mean ionization energy for nuclear recoils as a function of the recoil energy. We determine the ionization yield for nuclear recoils from fits to the

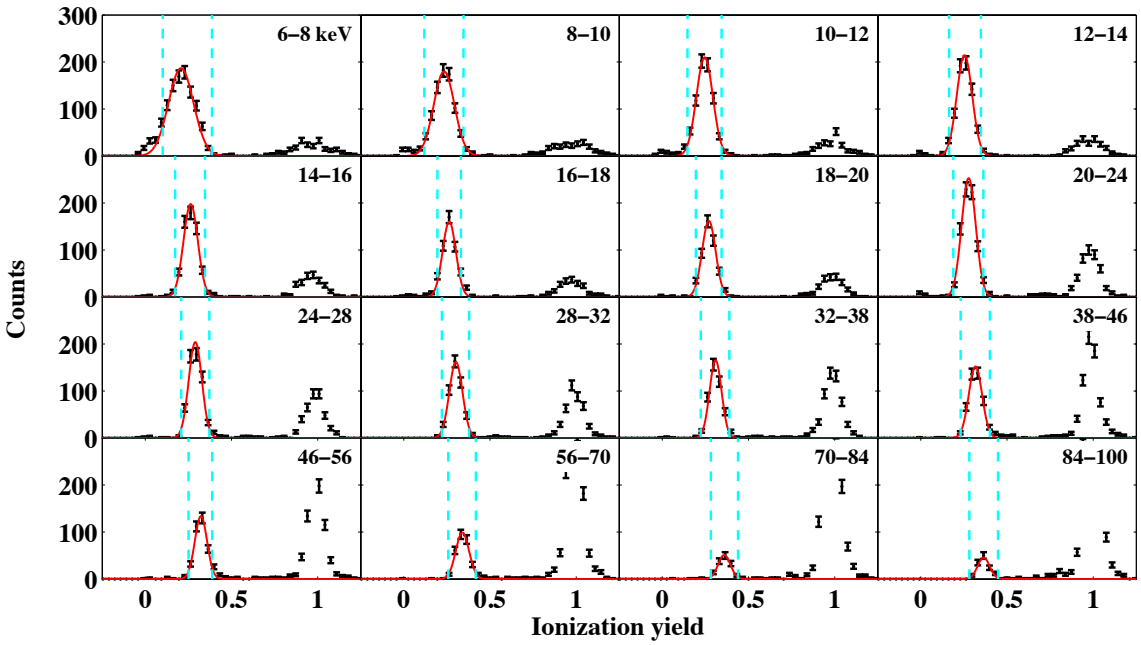


Figure 4.6: Gaussian fits to the ionization yield distribution in bins of recoil energy for nuclear recoils in ^{252}Cf calibration data for T1Z5. The cyan dashed lines show the fitting window, the lower edge of which is constrained to lie $>2\sigma$ above the mean of the ionization noise distribution at low energy. The high-yield peak that is excluded from the fits corresponds to residual electron recoils in the ^{252}Cf data.

distribution of neutron-induced nuclear recoils in ^{252}Cf calibration data. Above ∼4–8 keV, depending on detector, the bulk of the ionization distribution for nuclear recoils is well separated from noise, and fits to the ionization yield distribution in bins of phonon-based recoil energy directly give the mean ionization yield for nuclear recoils, as shown in Fig. 4.6 for T1Z5.

At lower energies as the ionization signal becomes comparable to noise, the reconstructed energy is biased to higher values by the optimal filter search algorithm. Since the start time of the charge pulse will vary relative to the phonon trigger time depending on the signal-to-noise of the pulse, the algorithm performs a search for the start time by offsetting the template pulse within a 100 μs window prior to the phonon trigger time. It then finds the best estimate of the pulse start time from the time offset for which the amplitude estimator is maximized. As the ionization energy goes to zero, this leads to a positive bias in the reconstructed amplitude since the algorithm begins to choose the start time corresponding to the largest positive noise fluctuation within the search window.

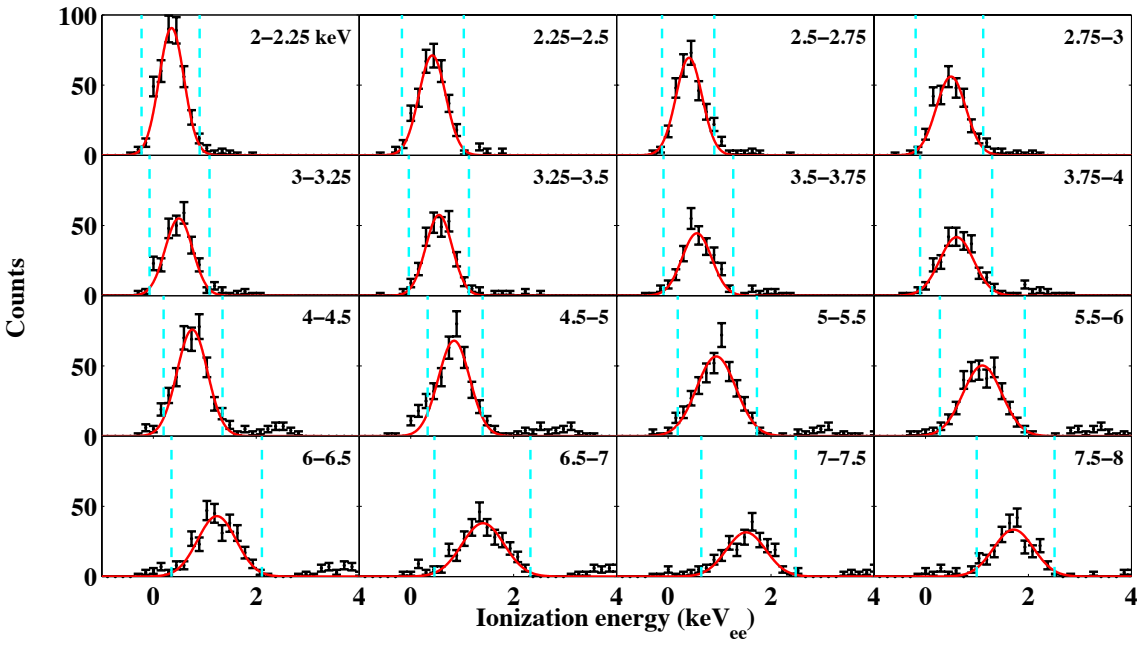


Figure 4.7: Gaussian fits to the ionization energy distribution in bins of recoil energy for low-energy nuclear recoils in ^{252}Cf calibration data for T1Z5. The cyan dashed lines show the fitting window, which has significant overlap with the noise distribution at the lowest energies. The energy bins above 6 keV give good agreement with the fits to the yield distribution in the same energy range shown in Fig. 4.6.

To measure the yields from 2–10 keV, we again begin by fitting the ionization energy distribution in bins of recoil energy as shown in Fig. 4.7 for T1Z5, recovering yield values biased high at low energies due to the optimal filter search. To account for this bias, for each detector we take the ∼2500 noise traces recorded by the DAQ prior to each ^{252}Cf calibration run for each detector. We then add a pulse with known energy to each noise trace and reconstruct the energy for each simulated pulse using the standard optimal filter reconstruction algorithm.

The mean reconstructed energy versus true energy is shown in Fig. 4.8a for T1Z5. At the lowest energies, the positive bias is reproduced by the simulation, giving a zero-energy bias identical to that measured from the randomly triggered noise pulses for each detector. Using the magnitude of the search bias as a function of reconstructed energy determined from the simulation, the measured mean of the ionization energy distribution is corrected to remove the effect of the search bias and recover the true mean, as shown in shown in Fig. 4.8b. As a cross-check on this correction, an alternative calculation was performed by

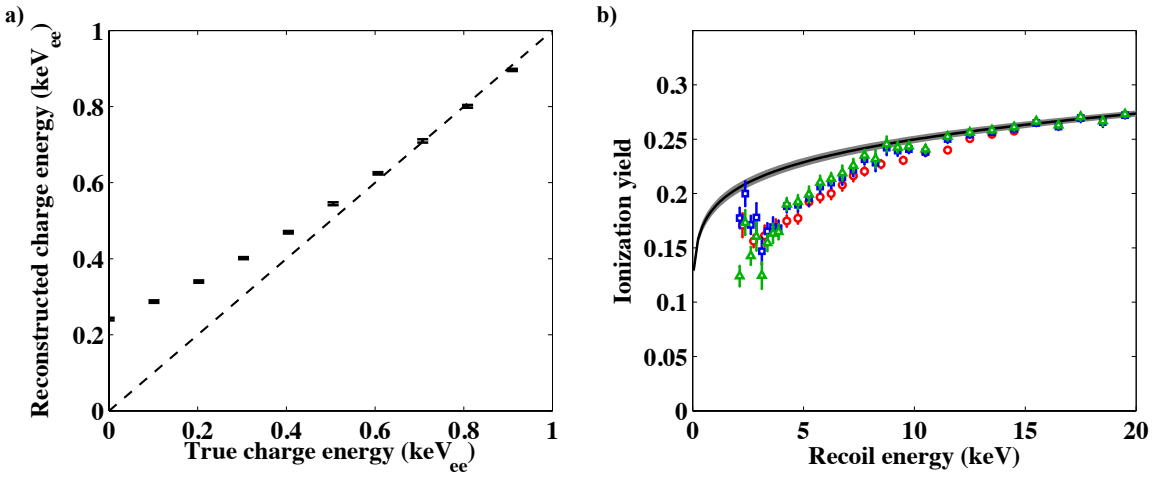


Figure 4.8: a) Reconstructed energy versus true pulse energy determined from simulated pulses for T1Z5. The dashed line shows the expected reconstructed energy for no bias, while the errorbars show the simulated optimal filter search bias. b) Comparison between the reconstructed yields for T1Z5 before (blue, squares) and after (green, triangles) applying the correction for the optimal filter search bias. The corrected values give good agreement with the yields determined by 1-parameter fits for the mean, constrained to fit only the portion of the ionization distribution $>2\sigma$ above noise (red, circles).

fitting only the portion of the ionization distribution $>2\sigma$ above noise, but constraining the amplitude and σ for the fit using the simulated ^{252}Cf spectrum and known energy resolution function for each detector, to allow only a single parameter fit for the mean. As shown in Fig. 4.8b, this alternative technique gives good agreement with the yields determined from fits to the full distribution after correcting for the optimal filter search bias.

The ionization yield for nuclear recoils in Ge has been previously measured in the 0.5–100 keV recoil energy range by several experiments [233, 264–275]. These measurements, which determine the ionization produced by a neutron-induced nuclear recoil while constraining the recoil energy using the kinematics of the scattering process, are shown in Fig. 4.9a. Also shown is the theoretical prediction for the ionization yield from Lindhard et al. [137, 276]:

$$y(E_r) = \frac{kg(\epsilon)}{1 + kg(\epsilon)} \tag{4.4}$$

where $k = 0.133Z^{2/3}A^{-1/2}$, $\epsilon = 11.5E_rZ^{-7/3}$, and $g(\epsilon) = 3\epsilon^{0.15} + 0.7\epsilon^{0.6} + \epsilon$. While the measurements agree with the Lindhard prediction within ∼10–20% over a wide energy range, the standard Lindhard model appears to underestimate the measured yields slightly

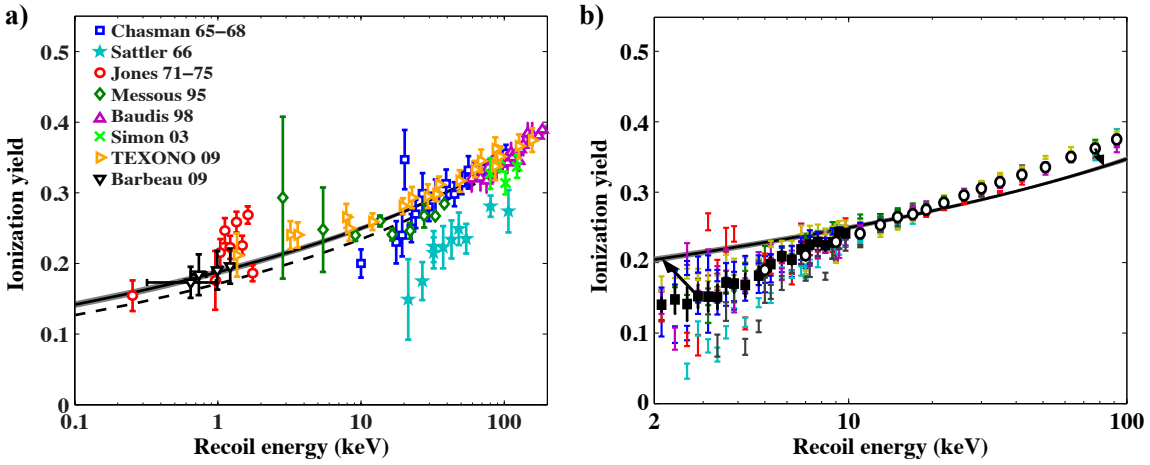


Figure 4.9: a) Previous measurements of the ionization yield for nuclear recoils in Ge. The dotted line shows the theoretical prediction from Lindhard et al. [276], while the solid line shows the best fit to the measurements assuming the form given in Eq. 4.5. b) Measurement of the ionization yield versus recoil energy in CDMS. The colored errorbars indicate the yields measured for each detector, while the black markers show the measured yields averaged over all detectors. The filled squares denote the low-energy yields measured after correcting for the optimal filter search bias, while the open circles show the yields determined from the fits to yield distributions at high energy. If the difference in the measured yields is ascribed entirely to an overestimate of the phonon-based recoil energy for nuclear recoils, the arrows indicate the effect of such a rescaling.

at low energy. Thus, we assume an empirical "Lindhard-like" dependence:

$$y(E_r) = C\frac{k'g(\epsilon)}{1+k'g(\epsilon)} \tag{4.5}$$

but where C and k' are free parameters. The best fit to previous measurements using Eq. 4.5 gives $C = 0.80$ and $k' = 0.24$ and is shown in Fig. 4.9a.

Figure 4.9b shows the measured yields in CDMS for each detector versus recoil energy, following the correction for the optimal filter search bias at low energy. At recoil energies $\gtrsim$20 keV, the yields measured for each detector agree within $\sim$5%, and are roughly 10% higher than the best-fit parameterization to previous measurements shown in Fig. 4.9a [228]. Below 10 keV, there is significant variation in the reconstructed yields versus energy, indicating that systematic variations between detectors dominate. Nonetheless, all detectors show yields which are lower than previous measurements in this energy range, with an averaged yield over all detectors that is $\sim$10–20% lower below 10 keV.

For each detector, the mean of the ionization energy distribution measured from ^{252}Cf

Detector:	**NR yield coefficients:**	
	A	B
T1Z2	0.1077	0.3154
T1Z5	0.1249	0.2697
T2Z3	0.1039	0.3295
T2Z5	0.0913	0.3602
T3Z2	0.0863	0.3972
T3Z4	0.1529	0.2060
T3Z5	0.0894	0.3803
T3Z6	0.1443	0.2230

Table 4.2: Parameterization of the measured mean ionization yield for nuclear recoils, $\mu_{Q,NR} = AE_r^B$, for each detector. These values are used to convert the measured total phonon signal to an equivalent recoil energy, assuming the Neganov-Luke phonon contribution is consistent with a nuclear recoil.

calibration data is parameterized by a power law of the form $\mu_{Q,NR} = AE_r^B$ over the energy range from 2–20 keV, where A and B are determined separately for each detector and listed in Table 4.2. The phonon-based recoil energy is then determined using these parameterizations to calculate the Neganov-Luke phonon contribution to the total phonon signal following Eq. 4.3. Due to the low ionization yield for low-energy nuclear recoils, only $\sim$15% of the total phonon signal arises from Neganov-Luke phonons, and any error due to uncertainties in the measurement of the ionization yield is reduced by the same factor, leading to a $<$3% systematic uncertainty on the recoil energy at 2 keV resulting from the Neganov-Luke correction [144].

Provided that the ionization collection efficiency for nuclear recoils at low energy does not differ from that for electron recoils, the yield measurements are inconsistent with an underestimate of the nuclear recoil energy scale. Under this assumption, the recoil energy at 2 keV is overestimated by 5%–20%, depending on detector. If instead the lower yields are due to enhanced recombination or trapping of charges for low-energy nuclear recoils at the relatively low drift fields used in CDMS, then directly using the measured ionization in CDMS gives the correct Neganov-Luke contribution. Thus, to determine the energy scale for the analysis presented in this thesis, we do not apply a corresponding correction based on the comparison of the ionization yield with previous measurements. This leads to a possibly conservative estimate of the recoil energies since an overestimated energy scale produces weaker limits on the scattering cross section.

4.3 Standard reconstruction and quality cuts

To ensure proper pulse reconstruction and detector operation for each nuclear recoil candidate event, several standard data quality and reconstruction criteria are applied (typically denoted as "cuts"). These basic cuts are identical to those previously developed for the 10 keV threshold analysis of these data [153, 178], and are described in detail in [228, 230]. For completeness, a brief description of each of the data quality and reconstruction cuts applied for this analysis are listed below. In general, these cuts remove time periods of abnormal detector operation (e.g., periods with high noise or malfunctioning readout channels), events for which the reconstruction algorithms fail (e.g., pileup of multiple interactions in a single trace), or events inconsistent with a dark matter signal (e.g., interactions in multiple detectors or the muon veto).

4.3.1 Data quality cuts

1. **Overall detector operation:** As summarized in Table 4.1, certain detectors were functional only during subsets of the entire data taking period due to failure of readout channels following cryogenic cycling, loss of detector neutralization, or insufficient calibration. In addition, as discussed in Sec. 4.1, only the 8 detectors with the lowest trigger thresholds were used to search for WIMP interactions. Detector selection for this analysis was implemented with the cuts `cGoodDet_123` and `cGoodDet_c58`.

2. **Series-to-series detector stability:** Within cryogenic data runs, each few hour data period was automatically tested to ensure consistency with a sample of 30 "known good" series using Kolmogorov-Smirnov tests. Cuts to remove short time periods with statistically significant deviations from the reference series were implemented as `cBadDet_bg_123`, `cBadDet_bg_124`, and `cBadDet_bg_c58`.

3. **Detector neutralization:** As discussed in Sec. 3.1.1 the detectors can gradually become deneutralized when operated under bias for long periods of time, leading to reduced ionization collection due to trapping of charges at impurity sites. To maintain stable collection, the detectors are regularly grounded and illuminated with pulses from IR LEDs. On rare occasions, this neutralization procedure may fail to occur, so each few hour data-taking period is monitored for an abnormally high frac-

tion of events with ionization yield below the bulk of the electron recoil distribution. Periods showing evidence of poor neutralization are removed with `cBadNeut_123`, `cBadNeut_124`, and `cBadNeut_bg_c58`.

4. **Event triggering:** Nuclear recoil candidate events are required to have trigger history information consistent with particle interactions in the detector rather than triggers randomly issued by the DAQ or issued by the veto multiplicity trigger. These conditions are implemented by removing events selected by the `cRTrig` or `cVTrig` cuts. In addition, a handful of events for which the trigger history bitmask is incorrectly recorded by the DAQ are removed by the cuts `cErrMask_123` and `cErrMask_c58`.

5. **Stable detector tuning:** Periods near the beginning of the data run prior to the final tuning of the QET bias settings and trigger thresholds are removed using the cuts `cStabTuning_123` and `cStabTuning_c58`. For R123–124, the cut `cBadRolloff_123` was applied to remove time periods with abnormal phonon sensor bandwidth, for consistency with previous analyses.

6. **Poor noise performance:** Several cuts were applied to remove isolated periods of time where the readout noise for the charge or phonon channels was abnormally high for a given detector. `cBadResTight_123` and `cBadResTight_c58` remove data series where the amplitude or timing resolution for the phonon signal is >25% higher than the median resolution throughout each data run. In addition, `cHighQNoise_123`, `cHighQNoise_124`, and `cHighQNoise_c58` remove time periods within a single data series where the energy resolution is $>4\sigma$ above the mean resolution, or where a significant rate of charge-only electronics glitches are observed.

7. **Bad detector regions:** Certain detectors had readout problems with a single phonon channel or portion of their ionization electrodes but still allowed events interacting far from the location of the malfunctioning sensors to be used. Of these partially functioning detectors, only T3Z2 was used in this analysis. For this detector, `cBadDetRegions_123` and `cBadDetRegions_c58` remove events whose primary phonon channel was C or D, which are close to a disconnected portion of the ionization guard ring on the right side of the detector.

8. **NuMI neutrino beam:** Although estimates indicate that neutrino-induced showers

coincident with the NuMI neutrino beam should provide a negligible background to CDMS, the cuts `cNuMI_123` and `cNuMI_c58` were applied to remove time periods within 60 μs of the NuMI beam. These cuts provide negligible loss of livetime since the beam is active only once every $\sim$2.5 s.

4.3.2 Reconstruction quality cuts

1. **Charge Optimal Filter goodness-of-fit:** The processing package computes the χ^2 statistic in the frequency domain for the optimal filter fit to the charge pulses for each event. Fits with abnormally high values of χ^2 are rejected by the cuts `cChiSq_123`, `cChiSq_124`, and `cChiSq_c58`, which primarily remove "pileup" events where additional interactions not centered at the trigger time are recorded within the $\sim$ 1.5 ms long trace. A corresponding cut is not applied using the phonon signals due to the broad distribution of the χ^2 statistic resulting from the systematic variations in pulse shape with position.

2. **Phonon prepulse baseline:** For each event, the standard deviation of the pulse baseline in the 500 samples prior to the trigger time is calculated and traces with a standard deviation that is $>4\sigma$ above the median are rejected by `cPstd_123` and `cPstd_c58`. These cuts also primarily remove pileup events as well as isolated periods of abnormally high noise.

4.3.3 Event topology

1. **Muon-veto coincidence:** Events with scintillator veto information consistent with a coincident muon interaction are rejected to avoid cosmogenic neutron-induced nuclear recoils. The cuts `cVTStrict_123` and `cVTStrict_c58` reject the vast majority (>99%, as discussed in Sec. 3.2.2) of these muon-coincident events while maintaining >97% livetime by differentiating the muons from the much higher rate of γ interactions in the scintillator panels.

2. **Multiple detector interactions:** Since WIMPs have negligible probability of interacting more than once in the detector apparatus, events with energy depositions $>4\sigma$ above noise in multiple detectors are rejected. Although only the 8 Ge detectors listed in Table 4.1 are used to identify WIMP interactions, all 30 detectors are used

to veto events interacting in multiple detector modules. This criterion is implemented as `cSingle_123` and `cSingle_c58`, which are denoted as the "singles" cut below.

Several cuts that were applied for the 10 keV threshold analyses presented in [153, 178] were not applied for this analysis due to poor acceptance of nuclear recoils at low energy. For this reason we do not require the ionization signal to be $>4\sigma$ above noise, as in previous analyses (implemented as `cQThresh_123` and `cQThresh_c58`) nor do we place a requirement on the start time of the phonon signal (implemented as `cGoodPStartTime`) since the timing delay can be poorly reconstructed at recoil energies $\lesssim$5 keV.

4.4 Additional low-energy quality and reconstruction cuts

For the low-threshold analysis presented here, several cuts required modification to improve background rejection and signal acceptance at energies below 10 keV. All cuts were developed using a subset of data consisting of $\sim$25% of all events evenly sampled throughout the run, as described in Sec. 4.4.1. In particular, cuts to remove electronics glitches, microphonic pickup from the cryocooler, periods with spurious events due to liquid He films condensed on the detectors, and activation of the detectors following neutron calibrations are described in Secs. 4.4.2–4.4.5. The fiducial volume cut was also optimized at low energy to improve the acceptance efficiency for nuclear recoils below 5 keV, as described in Sec. 4.4.6. Finally, the yield-based ionization selection of nuclear recoil candidate events was optimized to maximize signal acceptance while minimizing leakage from electron-recoil backgrounds above and below the signal region, as described in Sec. 4.5.

4.4.1 Data selection

Since backgrounds below 10 keV at Soudan had not been previously studied in detail, a subset of the exposure was reserved to study these backgrounds and optimize selection criteria at low energy. All selection criteria were determined with no knowledge of the candidate events in the remaining subset of data, which was used to calculate limits. The cut `cBlindLowE` was developed to select a subset of 25% of events spaced evenly throughout the data taking period, which we refer to in the following sections as the "open data." Each event recorded by the DAQ in a given data series is associated with an `EventNumber`, which provides a sequential ordering of the events recorded in each series. Events with

`EventNumber % 4 = 0`, where `%` denotes the modulus operation, were reserved to study low-energy backgrounds. All data selection cuts described in the following sections were determined using this subset of data prior to looking at the remaining 75% of the data, which was used to obtain the results presented in Chap. 5. This method of setting cuts using a statistically independent data set ensures that cuts could not be tuned to produce limits that were biased by statistical fluctuations. Since the analysis was expected to be background limited, the 25% decrease in exposure resulting from this procedure did not significantly decrease the expected WIMP sensitivity.

4.4.2 Glitch cut

Electronics "glitches" denote a broad class of events which are generated by the readout electronics rather than by particle interactions in the detectors. For low-energy nuclear recoils, the most dangerous background is from phonon-only glitches with no corresponding signal in the charge channels [232]. Using only ionization yield, these events can be difficult to distinguish from true nuclear recoils at energies $\lesssim$5 keV where the corresponding nuclear-recoil ionization signal becomes comparable to noise.

These glitches do have several characteristic features that allow them to be easily rejected with negligible loss of acceptance of true nuclear recoils. As shown in Fig. 4.10, the rise time and fall time of the pulses for glitch events is consistent with the ETF time constant of the TESs, rather than the slower rise and fall times characteristic of phonon mediated events. These pulse shape differences can be effectively identified by improvements to the optimal filter algorithm which fit a glitch template to the residual phonon traces obtained after subtracting off the best fit to the standard phonon pulse template [240].

In addition, phonon-only glitch events typically have high multiplicity, with coincident pulses appearing in multiple detectors. Thus, nearly all glitches are removed by the standard "singles" cut described in Sec. 4.3.3. However, very near threshold, the hardware trigger information recorded by the DAQ can be more sensitive to glitch events than the energy reconstructed by the optimal filter algorithm due to the mismatch between the glitch pulse shape and pulse template. Thus, an additional multiple scatter cut that has better rejection of low-energy glitch events was implemented by rejecting events which had $\geq$2 P_{lo} phonon triggers with at least two fewer Q_{lo} charge triggers. This condition was enforced by `cGlitch_le_c58`. This cut has >99% acceptance of nuclear recoil events in ^{252}Cf calibration

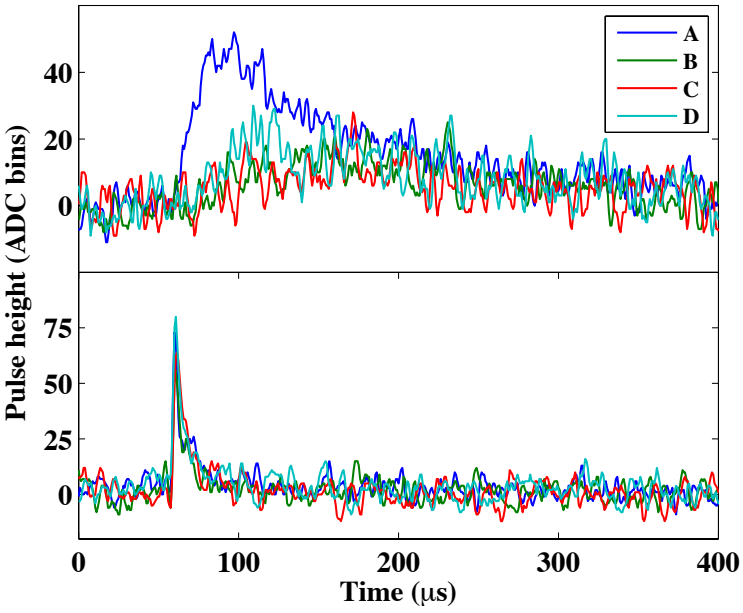


Figure 4.10: Comparison of typical 5 keV nuclear recoil phonon signal (top) to the signal from a phonon-only glitch event (bottom). The different traces denote the signal in each phonon channel (A–D). The glitch event is seen to have a fall time that is much shorter than the phonon lifetime, which sets the fall time of the nuclear-recoil event.

data that deposit energy above the software threshold in only a single detector. In addition, a manual inspection of low-energy traces in the open data indicated that all glitch events identified by eye were successfully removed by these trigger multiplicity criteria. Although multiplicity criteria were sufficient for this analysis, future analyses that use fewer detectors or extend to lower energies may require improved glitch removal techniques using optimal filter estimates of the glitch component, as described above.

4.4.3 Cryocooler cut

As discussed in Sec. 3.1.1, the electronics stem cryocooler introduces significant mechanical vibrations that contribute microphonic pickup, increasing the low-frequency charge noise for certain detectors (e.g., T2Z3 and T3Z5). However, the cryocooler period is 1.2 s while vibrations at the $\sim$1 kHz signal frequencies of interest typically decay away on $\sim$100 ms time scales. Thus, events occurring more than a few hundred ms after the the most recent cryocooler compression do not suffer significantly from microphonic pickup and can be recovered without a significant increase in the baseline noise. A comparison of the

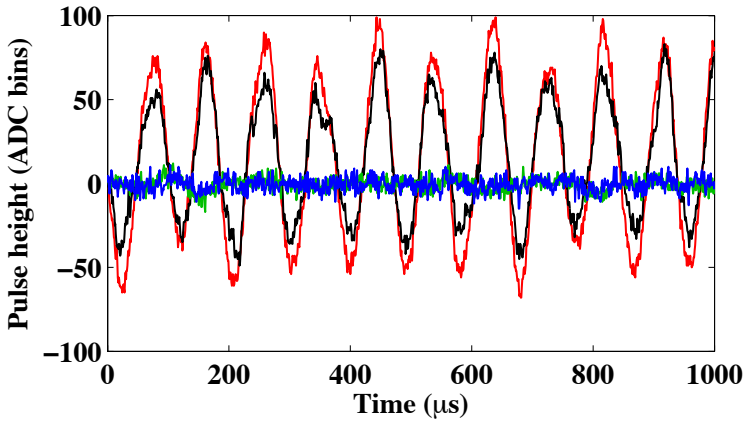


Figure 4.11: Example of randomly triggered charge noise traces for T3Z5 immediately after the cryocooler compression for the inner (black) and outer (red) electrodes. A large oscillation due to microphonic pickup with $f \approx 10$ kHz is obvious. These mechanical oscillations decay away on time scales ∼10–100 ms and the pickup for the inner (blue) and outer (green) channels is significantly reduced at times longer than a few hundred ms after the compression.

charge traces between an event immediately after the cryocooler compression, and one well separated is shown in Fig. 4.11 for T3Z5.

For this analysis, we follow the procedure from the 10 keV threshold analysis and use the standard deviation of the prepulse baseline in the charge traces to remove events with significant pickup from the cryocooler. For each detector, the distribution of the standard deviation for the prepulse baseline is determined. A cut is defined to remove events with a larger standard deviation than a given detector-dependent threshold, and the zero-energy resolution is calculated from randomly triggered noise traces as a function of the cut position. The cut condition for the low-threshold analysis was set to ensure that the reconstructed charge resolution was not increased by more than 20% above the baseline resolution with the cryocooler turned off. This is a more stringent criterion than that used in the higher threshold analysis for T2Z3, T2Z5, and T3Z5, since variations in the noise have a larger effect on background rejection for this analysis. These more restrictive cuts were implemented as `cQstd_le_123`, `cQstd_le_124`, and `cQstd_le_c58`.

4.4.4 He films cut

Due to a small vacuum leak in the electronics breakout box, periods during the end of R125 and R127 suffered from condensation of superfluid liquid He films on the detector surfaces.

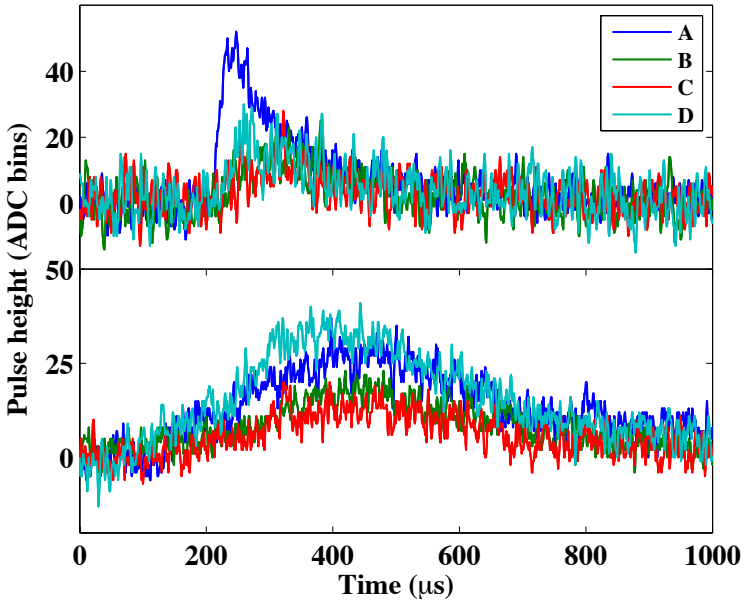


Figure 4.12: Comparison of a typical 5 keV nuclear recoil phonon signal (top) to a typical phonon-only signal from the high-rate phonon-only events present after the onset of He films (bottom). The different traces denote the signal in each phonon channel (A–D). The He film event is seen to have a significantly longer rise time and fall time than the particle induced event.

These films were found to produce a high rate of events with significant phonon signal but no corresponding charge signal, as shown in Fig. 4.12. Like the glitch events described above, such events can leak into the signal region at low energy where the ionization signal for true nuclear recoil events becomes comparable to noise.

Both the number of detectors showing spurious events and the rate of such events was found to increase over time near the end of R125 and R127 until the trigger rate became unacceptably high and the detectors were partially warmed up so that the He could be pumped from the icebox. For the low-threshold analysis, a more conservative exclusion of periods suffering from He films was used than in the 10 keV threshold analysis since such events could produce significant backgrounds at low energy even shortly after the onset of the films. The cut `cHeFilm_le_c58` removes data after the onset of the films, determined by monitoring the rate of events with zero yield over time and cutting periods after a sustained increase over several data series was observed. An additional buffer removing 5 data series before the onset of films was applied to ensure that any film-induced events were entirely excluded from the low-threshold analysis.

4.4.5 Neutron activation cut

As discussed in Sec. 4.2.1, neutron calibrations with the ^{252}Cf source produce ^{71}Ge which decays away via electron capture with a half-life of 11.4 days. For electron capture from the L-shell, the Ga binding energy of 1.3 keV$_{ee}$ is released as x-rays and Auger electrons. This leads to a higher than average rate of bulk electron recoils near 1.3 keV$_{ee}$ immediately after ^{252}Cf calibrations. Although bulk electron recoils are well separated from nuclear recoils in ionization yield at high energies, poorly collected 1.3 keV$_{ee}$ electron recoils, or those with a downward noise fluctuation in the reconstructed energy, can appear in the signal region. To reduce this background, data taken within 20 days following a ^{252}Cf calibration are excluded from this analysis. As discussed in Sec. 5.3.3, after removing these time periods, leakage from the 1.3 keV$_{ee}$ line is no longer the dominant expected background near threshold, and excluding longer periods following calibration would not provide significant further improvements in the sensitivity.

4.4.6 Fiducial volume cut

Due to the optimal filter search algorithm, the reconstructed charge energy for events where the ionization signal approaches zero is biased to values above zero, since the algorithm has the freedom to fit the largest positive noise fluctuation in the search window. This leads to reduced acceptance for the fiducial volume cut defined for the standard 10 keV threshold analysis at low energy, since it does not account for this effect. Figure 4.13 shows the definition of the fiducial volume cut for the low-threshold analysis. The cut is defined in the outer charge energy (`qo`) versus inner charge energy (`qi`) plane, by fitting the distribution of events along the the (`qo` = 0) axis for electron recoils from ^{133}Ba calibration data.

The majority of events interact in the volume under the inner electrode (∼85% of the total geometric volume), leading to an outer charge signal consistent with 0. We define the fiducial volume selection as the events with charge energy within 2σ of the mean of this distribution. However, at charge energies below ∼2 keV$_{ee}$, depending on detector, the optimal filter search bias causes the band of bulk events to shift to higher values of reconstructed outer energy, `qo`. Since the ionization yield for nuclear recoils is ∼0.2 below 10 keV, failure to account for this bias can lead to reduced acceptance of nuclear recoils when the search algorithm finds a large positive noise fluctuation. To account for this,

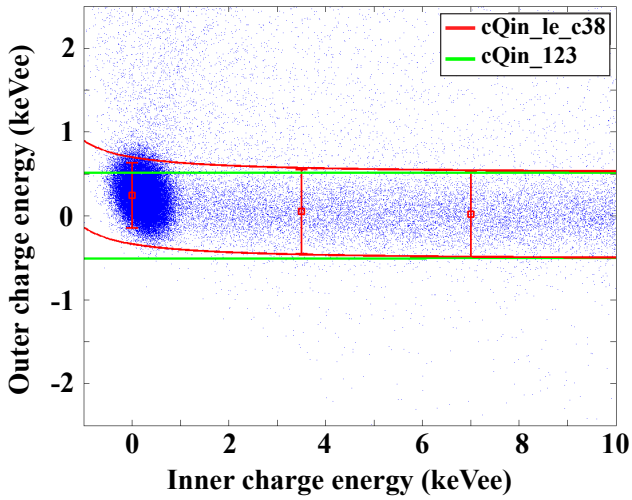


Figure 4.13: Definition of fiducial volume cut in the outer charge energy vs. inner charge energy plane for T1Z2. The distribution of events with outer ionization energy consistent with zero is fit to a Gaussian and the $\pm 2\sigma$ band is determined. For the low-threshold analysis, the bands shown in red are used, after accounting for the positive bias at low energy due to the optimal filter search algorithm. The acceptance of the standard cut (green) is reduced at low energy since it does not account for this bias.

the distribution of `qo` for bulk interactions was refit after allowing an increasing energy dependence as the ionization signal becomes comparable to noise. Although we cannot determine whether an event interacted in the fiducial volume at the lowest energies, this definition avoids an unnecessary reduction in the signal efficiency. This updated fiducial volume selection at low energies was implemented as `cQin_le_c38`.

4.5 Optimization of ionization-based background rejection

The ionization yield provides the primary discriminant between electron-recoil backgrounds and nuclear-recoil candidate events for the low-threshold analysis. The nuclear-recoil and electron-recoil distributions were determined from the ^{133}Ba and ^{252}Cf calibration data in the ionization energy versus total phonon energy plane, as shown in Fig. 4.14. Previous CDMS analyses (e.g., [153, 186]) have typically defined the corresponding distributions in the ionization yield versus recoil energy plane. However, for the low-threshold analysis it is preferable to work with ionization energy directly rather than ionization yield since the yield diverges as the recoil energy approaches zero. This makes it difficult to properly account for the width of the electron and nuclear recoil distributions in the ionization yield

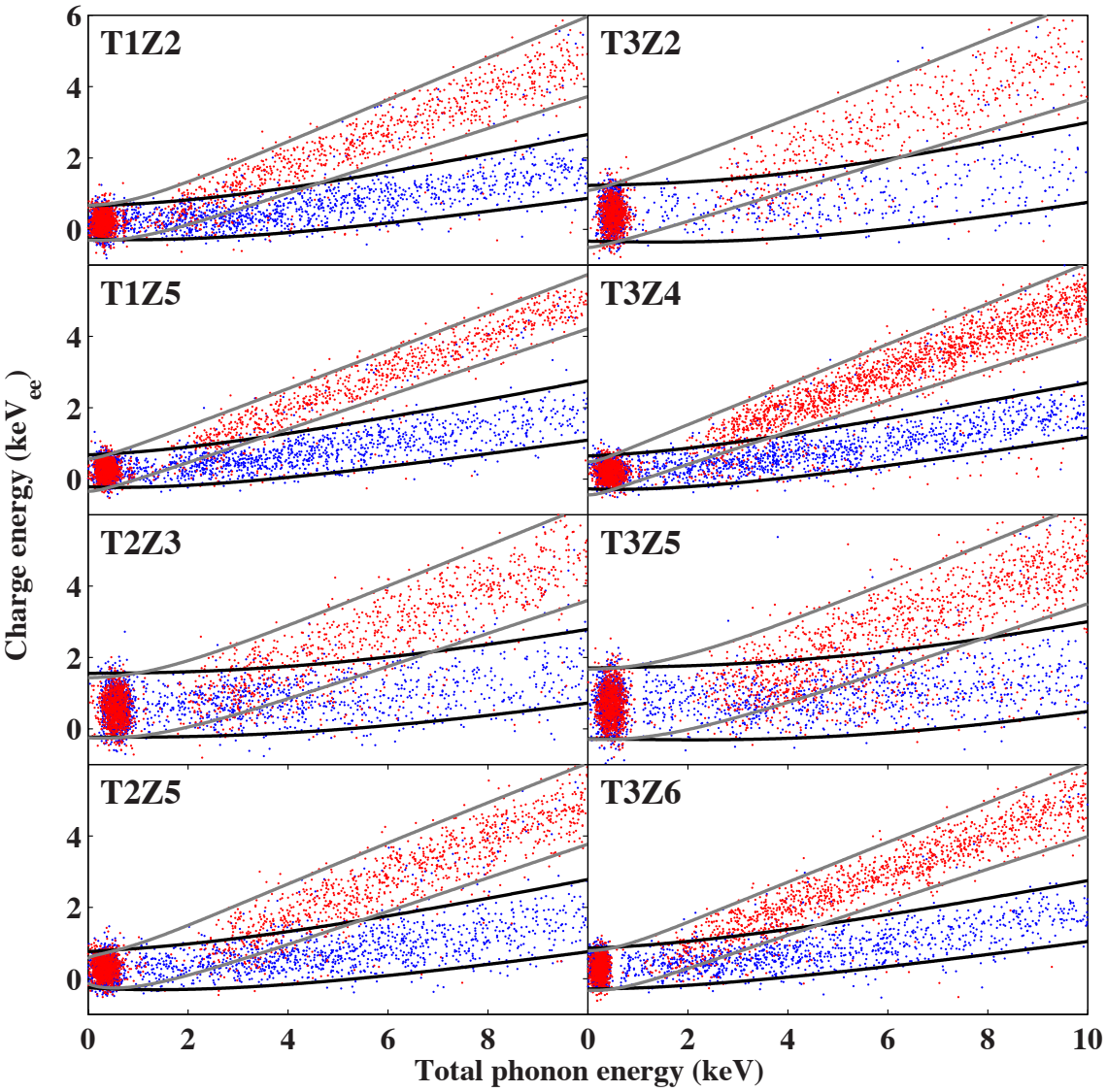


Figure 4.14: Calibration of ionization response for electron recoils from the ^{133}Ba source (red) and neutron-induced nuclear recoils from the ^{252}Cf source (blue). The gray lines denote the $\pm 2\sigma$ electron-recoil band, which is designed to contain $\sim$95% of bulk electron recoils. The corresponding $\pm 2\sigma$ nuclear-recoil band is shown by the black lines.

plane. In addition, we work directly with the total phonon energy (including the Neganov-Luke contribution) rather than the recoil energy in order to avoid folding in the poorer signal-to-noise of the charge measurement at low energy. As discussed in Sec. 4.2, there is a one-to-one mapping of the total phonon energy to the recoil energy only after a specific recoil type is assumed.

Fits to the electron-recoil and nuclear-recoil distributions in the ionization energy versus

total phonon energy plane are also shown in Fig. 4.14. The distribution of ionization energy is fit to a Gaussian for each total phonon energy bin, and the dependence of the mean and width of the ionization distribution versus energy are parametrized as:

$$\mu(p_t) = \frac{\alpha_1 p_t^{\alpha_2+1}}{1+\alpha_3 p_t^{\alpha_2}} + \alpha_4 \tag{4.6}$$

$$\sigma(p_t) = \beta_1 p_t^{\beta_2} + \beta_3 \tag{4.7}$$

where the functional form for $\mu(p_t)$ allows for additional curvature of the distribution at low energy due to the positive bias of the optimal filter search algorithm, while approaching a linear dependence at high energy. These fits define the electron and nuclear recoil "bands," which are used to select events that are consistent with each recoil type.

The distribution of nuclear recoils in the ^{252}Cf calibration data is compared to the observed distribution of events in the "open" WIMP search data in Fig. 4.15. Previous analyses (e.g., [153, 186]) have used a $\pm 2\sigma$ band to select nuclear recoil candidates in order to maximize signal acceptance. However, the distribution for nuclear recoils is more localized near the center of the nuclear recoil band than for the events in the open data at low energy, indicating that an ionization-based selection of only the central portion of the distribution should maximize signal acceptance while minimizing leakage from backgrounds above and below the bulk of the signal distribution.

To determine the nuclear-recoil band selection that maximizes sensitivity, limits on the scattering cross section were calculated using the open data, following the same procedure that will be used in Sec. 5.2 to set the final constraints from the full data set. We initially consider only an energy-independent selection and scan over each cut position defined by (n_{hi}, n_{lo}), where n_{lo} gives the location of the lower boundary of the nuclear-recoil band, normalized to the number of σ from the mean of the distribution, and n_{hi} gives the corresponding upper boundary. At each cut position, the nuclear-recoil acceptance is measured as a function of energy, and the statistical errors on how precisely this efficiency can be measured are calculated. The candidate nuclear-recoil events in the open data are then found, and the optimal interval limit is calculated assuming the most conservative efficiency values consistent with the measured efficiency at the 90% CL. The resulting sensitivity, as a function of cut position is shown in Fig. 4.16, assuming a WIMP mass, $m_\chi = 8$ GeV/c^2.

Both the limits computed after averaging the rates across detectors and those computed

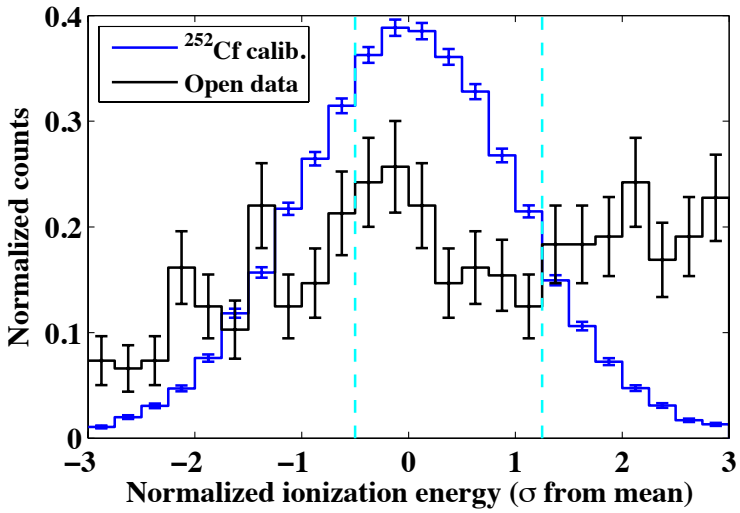


Figure 4.15: Comparison of the ionization energy distribution for neutron-induced nuclear recoils from the ^{252}Cf calibration data (blue) to the distribution of events in the open WIMP search data (black) for a recoil energy range of 2–10 keV, summed over detectors. The horizontal axis gives the normalized ionization energy, measured as the number of σ from the mean of the nuclear-recoil distribution. The cyan dashed lines show the location of the (+1.25,-0.5)σ nuclear-recoil band cut discussed below.

by concatenating the results from different detectors are shown in Fig. 4.16. These limit-setting procedures will be discussed in detail in Sec. 5.2. In both cases, a nuclear-recoil band selection within (n_{hi}, n_{lo}) = (+1.25, -0.5)σ gives approximately optimal sensitivity. For the concatenated limits the detector with the lowest rate dominates the sensitivity and there is a broad minimum with only ∼30% stronger limits than for a $\pm 2\sigma$ cut. For the averaged limits, higher rates on the detectors with poor charge resolution lead to a larger improvement (∼50%) for the tighter cut.

Enhanced sensitivity may be possible for certain WIMP masses by using an energy-dependent cut that increases the exposure at higher energies where backgrounds are expected to be less prominent. To check this possibility, the same optimization procedure was repeated with a cut designed to have equal leakage as a function of energy, as shown in Fig. 4.17. This cut is defined by the upper and lower boundaries at the 2 keV threshold (n_{hi}[2 keV], n_{lo}[2 keV]). At higher energies, the upper boundary is relaxed to remain at a fixed number of σ below the mean of the electron-recoil distribution, while the lower edge is fixed to a constant value in ionization energy. This procedure should give constant leakage from the bulk electron-recoil distribution, as well as from the population of events with

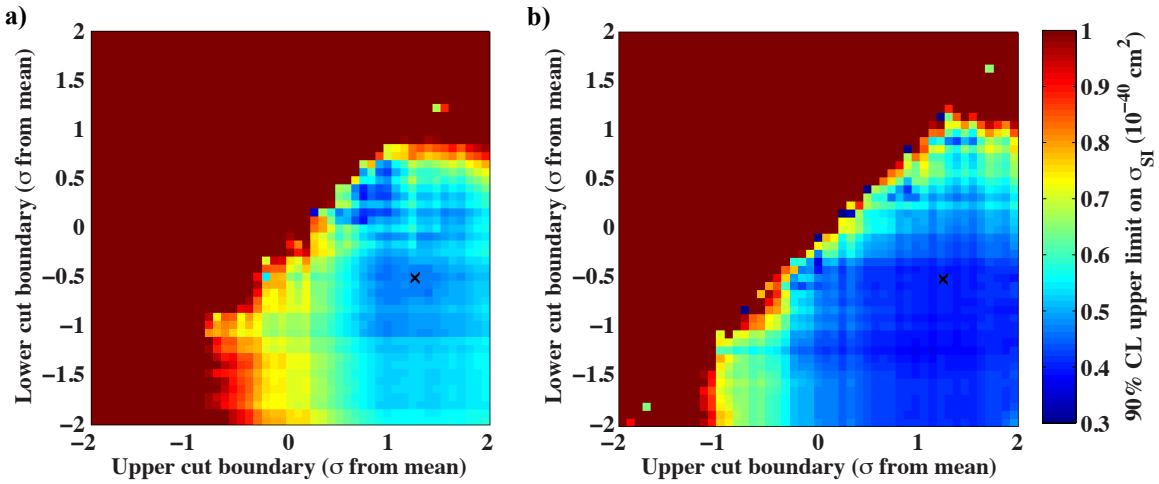


Figure 4.16: Limits on the spin-independent scattering cross section for $m_\chi = 8$ GeV as a function of cut position from the open data. The limits after averaging the rate over all detectors are shown in a), with limits where the energy intervals used for the optimal interval calculation are concatenated by detector shown in b). For b), the limits are dominated by the detector with the best ionization resolutions (T1Z5). The black x highlights the (+1.25,-0.5)σ cut considered below, which approximately maximizes the sensitivity of the analysis for either limit-setting procedure.

ionization energy consistent with 0. However, for WIMP masses $\lesssim$10 GeV/c^2, the expected WIMP spectrum falls off more quickly than the observed event rate in the open data, so an energy-dependent cut is not found to give significantly increased sensitivity in the region of interest for this analysis. Given the focus of this analysis on the 5–10 GeV/c^2 mass range, the simpler energy independent cut with (n_{hi}, n_{lo}) = (+1.25, -0.5)σ was used for the results presented in Chap. 5.

4.6 WIMP search exposure

The following sections calculate the total WIMP search exposure and signal acceptance efficiency for the low threshold analysis. The expected number of WIMP interactions for a given cross section depends only on the product of this exposure and efficiency, and in practice there is often ambiguity as to whether a given cut should be classified as removing exposure, or as removing acceptance efficiency. In the following sections, we follow the convention from previous analyses [228,230] where cuts that remove time periods, independent of the characteristics of an individual event, are applied before calculating the exposure in Sec. 4.6.1. The acceptance of the remaining cuts, which depend on characteristics of

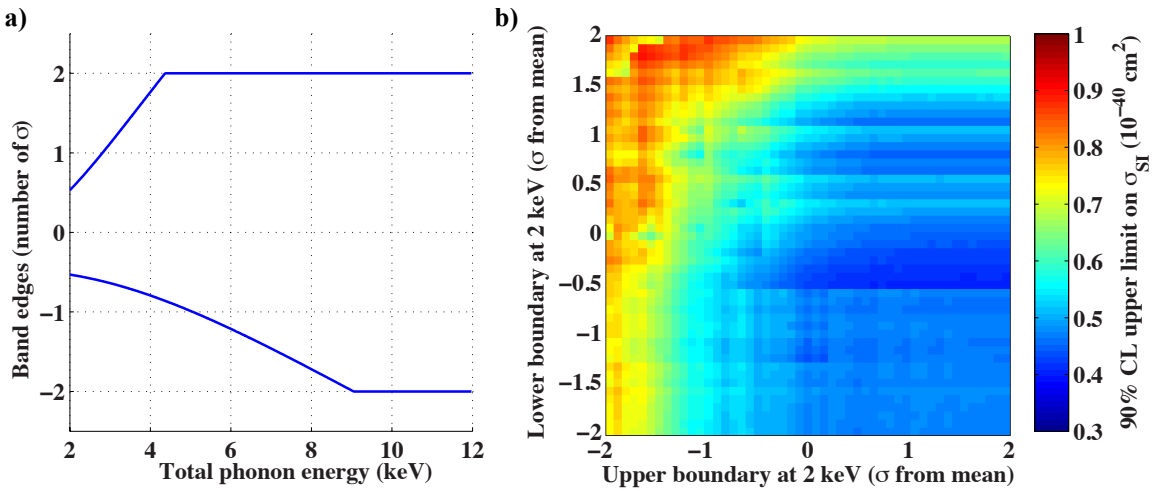


Figure 4.17: a) Example definition of an energy dependent nuclear-recoil band cut with approximately equal leakage as a function of energy for T3Z5. The cut is defined by its boundaries at the 2 keV total phonon energy threshold (here $\pm 0.5\sigma$). b) Limits on the spin-independent scattering cross section for $m_\chi = 8$ GeV/c^2 for a energy-dependent cut designed to give equal leakage as a function of cut position. The limits are calculated by concatenating the energy intervals from different detectors, and are dominated by the detector with the best ionization resolution (T1Z5).

the event and are often energy dependent, are included in the signal efficiencies given in Sec. 4.6.2.

4.6.1 Exposure

Prior to calculating the exposure, all cuts from Sec. 4.3.1 are applied. Of these cuts, the only which depends on the event characteristics is `cBadDetRegions`, which removes events occurring in the right half of T3Z2. For this cut, the passage fraction for neutron-induced nuclear recoils is separately determined from ^{252}Cf calibration data, and applied to the total exposure passing all other cuts. In addition, the cryocooler cut from Sec. 4.4.3 is applied since it primarily cuts time periods within a few hundred ms following each compression in the cryocooler cycle. The inverse of the selection cut for the open data described in Sec. 4.1 is applied, to remove the open data and leave only the remaining statistically independent subset of data used to calculate limits. Finally, the neutron activation and liquid He films cuts described in Secs. 4.4.5–4.4.6 are applied. For each event, the DAQ records the preceding time period for which it was live and waiting to trigger. This quantity is summed over the remaining events to find the total live time in days. The exposure, in

Detector:	**R123:**	**R124:**	**R125:**	**R126:**	**R127:**	**R128:**	**Total:**
T1Z2	10.25	6.59	10.78	7.80	4.58	3.45	43.44
T1Z5	9.44	7.44	10.16	7.99	0	0	35.02
T2Z3	9.44	0	9.19	5.80	3.60	0	28.03
T2Z5	11.83	0	11.08	7.40	4.42	0	34.71
T3Z2	2.80	0	2.14	1.47	0.82	0.62	7.84
T3Z4	0	0	11.75	8.40	5.51	3.92	29.58
T3Z5	7.57	0	7.98	4.90	3.14	2.16	25.75
T3Z6	8.60	0	10.62	8.14	5.02	3.77	36.14
Total:	59.92	14.03	73.70	51.88	27.08	13.92	**240.52**

Table 4.3: Total exposure for each detector and run after applying the selection cuts removing periods of abnormal detector operation. The total exposure summed over the 8 Ge detectors used to identify WIMP interactions is 241 kg days.

kg days, is then the product of this live time and the detector masses shown in Table 4.1. Table 4.3 shows this exposure for each detector and data run. The total exposure considered in this analysis, summed over all detectors, was 241 kg days.

4.6.2 Signal efficiency

The remaining cuts described in Secs. 4.3–4.5 are included in the signal acceptance efficiencies. For each cut, the acceptance is measured from calibration data as a function of energy and typically fit to a functional form describing its energy dependence. All cuts described in the above sections have high acceptance ($\gtrsim$98% for single-scatter bulk nuclear recoils), with the exception of the nuclear-recoil band and fiducial volume cuts, which have efficiencies of $\approx$60% and 80–95% at low energy, respectively. In addition, the inefficiency for the detector to trigger on particle interactions can be significant near threshold. In the following sections, the efficiency calculation for each of these dominant components of the signal acceptance is described. All efficiencies are calculated in terms of total phonon energy since this is the physical quantity measured by the detector on which they depend. To determine the signal acceptance, the efficiencies are then converted to recoil energy assuming a Neganov-Luke phonon signal consistent with a nuclear recoil.

4.6.2.1 Trigger efficiency

The efficiency with which the detectors trigger on low-energy single-detector particle interactions determines the fundamental lower limit to the energy threshold of this analysis. For each detector, the triggering efficiency is measured as a function of energy for the WIMP

search data combined over all data runs (and including the open data to maximize statistics). Events are binned in total phonon energy, and the efficiency is calculated as the ratio of the number of events in a given bin for which the detector triggers within the time interval defined by the optimal filter search window, relative to the total number of events in that bin.

Naively applying this procedure to all events would bias the measured trigger efficiency to higher values since true interactions where the trigger did not occur would be missed entirely, and the total number of interactions in the denominator would be too low. To avoid this triggering bias, only events for which at least one other detector triggers within the optimal filter search window were considered, to ensure that the event would be recorded regardless of whether a trigger was issued in the detector under test. After selecting this unbiased sample, the measured trigger efficiencies are shown in Fig. 4.18 for each detector. The efficiencies are measured in terms of the total phonon energy, p_t, and converted to recoil energy assuming the Neganov-Luke phonon contribution is consistent with a nuclear recoil. These efficiencies are well-described by an error function form:

$$\epsilon_{trig}(p_t) = \frac{A}{2}\left[1 + \mathrm{erf}\left(\frac{p_t - \mu}{\sqrt{2\sigma^2}}\right)\right] \tag{4.8}$$

with a mean, μ that varies between 1.5–3 keV for the detectors included in this analysis. The width, σ, is correlated with the measured detector resolutions shown in Fig. 4.3, but is typically a factor of $\sim$2 larger than the reconstructed resolution due to the poorer signal-to-noise of the causal hardware filter used for triggering, relative to the optimal filter.

4.6.2.2 Nuclear-recoil band efficiency

The acceptance for the $(+1.25, -0.5)\sigma$ nuclear-recoil band is measured directly using the ^{252}Cf calibration data. The data are first binned in phonon energy, and the efficiency is measured as the ratio of the number of events selected by the nuclear-recoil band cut to the total number of events in the calibration data that are $<3\sigma$ above the mean of the ionization distribution for nuclear recoils. The upper limit on the data selection is imposed to limit leakage of electron-recoil events present in the ^{252}Cf calibration data into the nuclear-recoil sample. We also restrict the event sample to only single-scatter events since this cut removes a significant fraction of the Compton scatter background in the electron-recoil band while

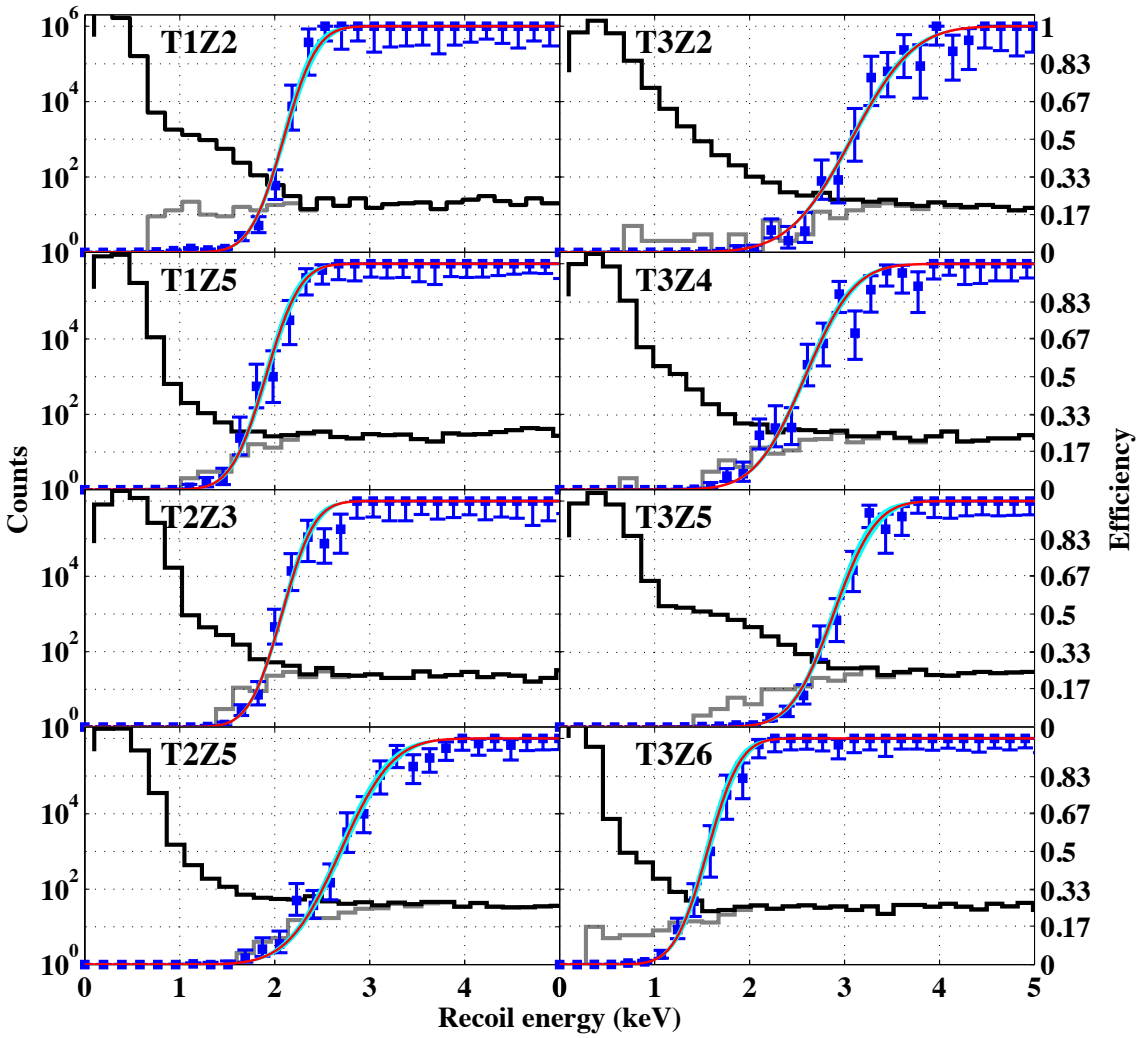


Figure 4.18: Measured trigger efficiencies for each detector as a function of recoil energy, assuming a Neganov-Luke phonon contribution consistent with a nuclear recoil. The total counts for each energy bin in the unbiased sample used to measure the efficiency are shown by the black histogram, with corresponding units on the left axis. The gray histogram shows the number of events in each energy bin for which the detector triggers. The blue squares show the maximum likelihood estimate of the efficiency with the error bars denoting the Clopper-Pearson 68% confidence intervals. The corresponding efficiency units are shown on the right axis. The red line shows the error function fit to the efficiencies, with the cyan band denoting the $\pm 1\sigma$ statistical errors determined from the covariance matrix of the fit.

maintaining the majority of the nuclear-recoil population. No lower bound on the ionization energy is included, so any neutron-induced nuclear recoils consistent with the zero-charge events discussed in Sec. 5.3.4 are included.

The measured efficiencies are approximately constant with energy, as was designed when

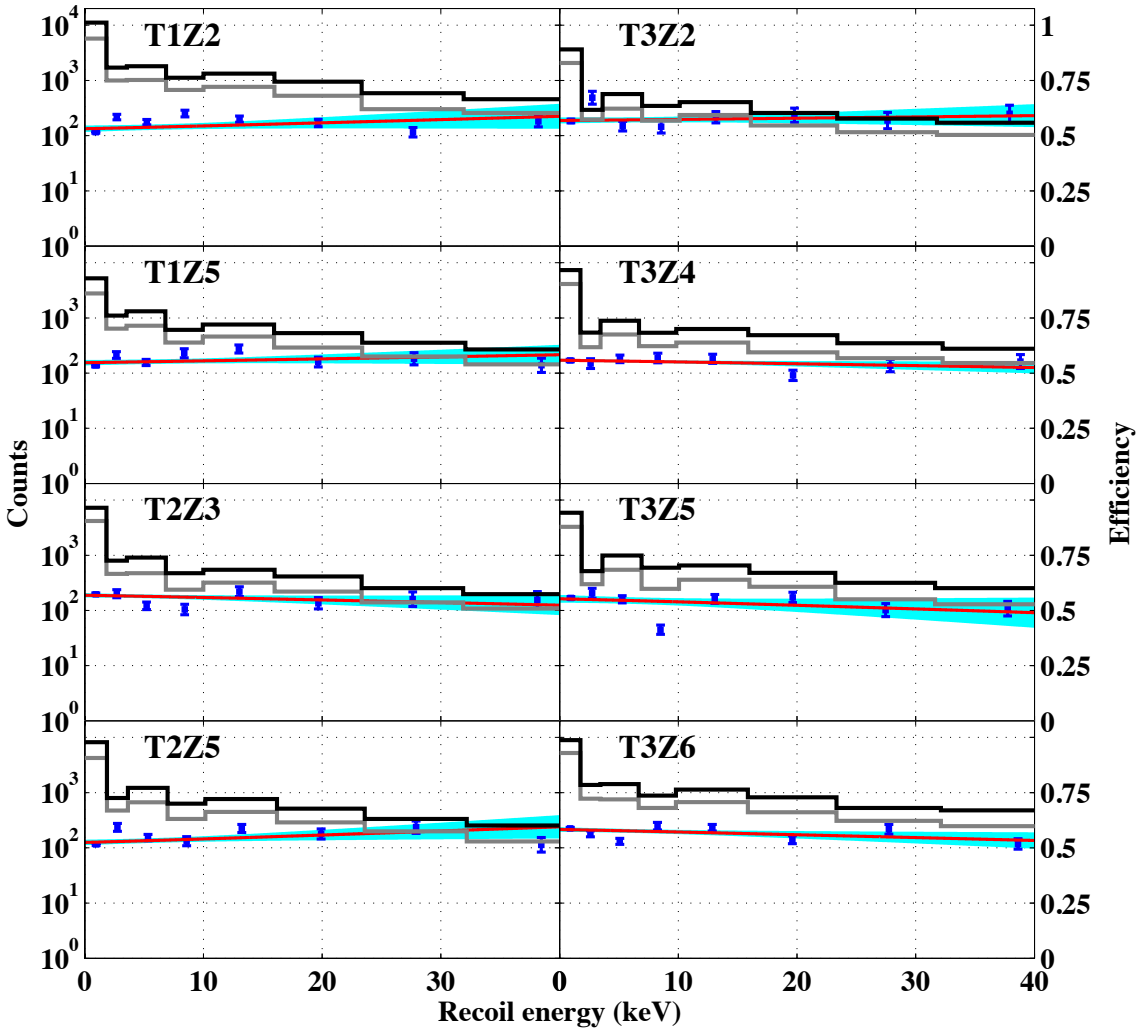


Figure 4.19: Measured nuclear-recoil band efficiencies for each detector as a function of recoil energy, assuming a Neganov-Luke phonon contribution consistent with a nuclear recoil. The total counts in the ^{252}Cf data that are $<3\sigma$ above the mean of the nuclear recoil distribution for each energy bin are shown by the black histogram, with corresponding units on the left axis. The gray histogram shows the number of events in each energy bin selected by the band cut. The blue squares show the maximum likelihood estimate of the efficiency, with the corresponding units on the right axis. The red line shows the linear fit to the efficiencies, with the cyan band denoting the $\pm 1\sigma$ statistical errors determined from the covariance matrix of the fit.

the energy independent nuclear-recoil band cut was defined. To allow for any residual energy dependence, they are fit to a linear model, as shown in Fig. 4.19. The measured efficiencies at low energy are roughly 60%, as would be expected for a Gaussian distribution in ionization yield.

4.6.2.3 Fiducial volume cut

The efficiency of the fiducial volume cut is also measured from ^{252}Cf calibration data, after selecting events within the $\pm 2\sigma$ nuclear-recoil band. The measured efficiencies are shown in Fig. 4.20. At low energies, as the ionization signal becomes consistent with noise, the cut efficiency approaches 95% as expected for the $\pm 2\sigma$ fiducial volume definition shown in Fig. 4.13. At higher energies, the efficiency approaches a roughly constant value near 60-70%. This acceptance is determined by the geometrical fiducial volume defined by the ionization electrodes and the typical diffusion of the charge carriers as they are propagated across the detector.

To parameterize these dependencies, the measured efficiencies are fit to:

$$\epsilon_{qin}(p_t) = \alpha_1 + \alpha_2 \mathrm{erf}\left(\frac{p_t - \alpha_3}{\sqrt{2\alpha_4^2}}\right) + \alpha_5 p_t \tag{4.9}$$

which transitions between the low-energy constant efficiency and the high-energy linear dependence following an error function form.

4.6.2.4 Total acceptance efficiency

The overall signal acceptance efficiency is then defined as the product of the measured efficiencies for the event reconstruction quality cuts, the fiducial volume cut, the nuclear recoil band cut, and the trigger efficiency, and is shown in Fig. 4.21.

The combined efficiency typically peaks between 2–4 keV near 50%, and falls to $\sim$40% at higher energies as the fiducial volume cut becomes more effective. The acceptance efficiency just above the 2 keV recoil energy threshold is dominated by the trigger efficiency, which varies significantly by detector. At recoil energies $\gtrsim$3 keV, the fiducial volume and nuclear-recoil band efficiencies dominate.

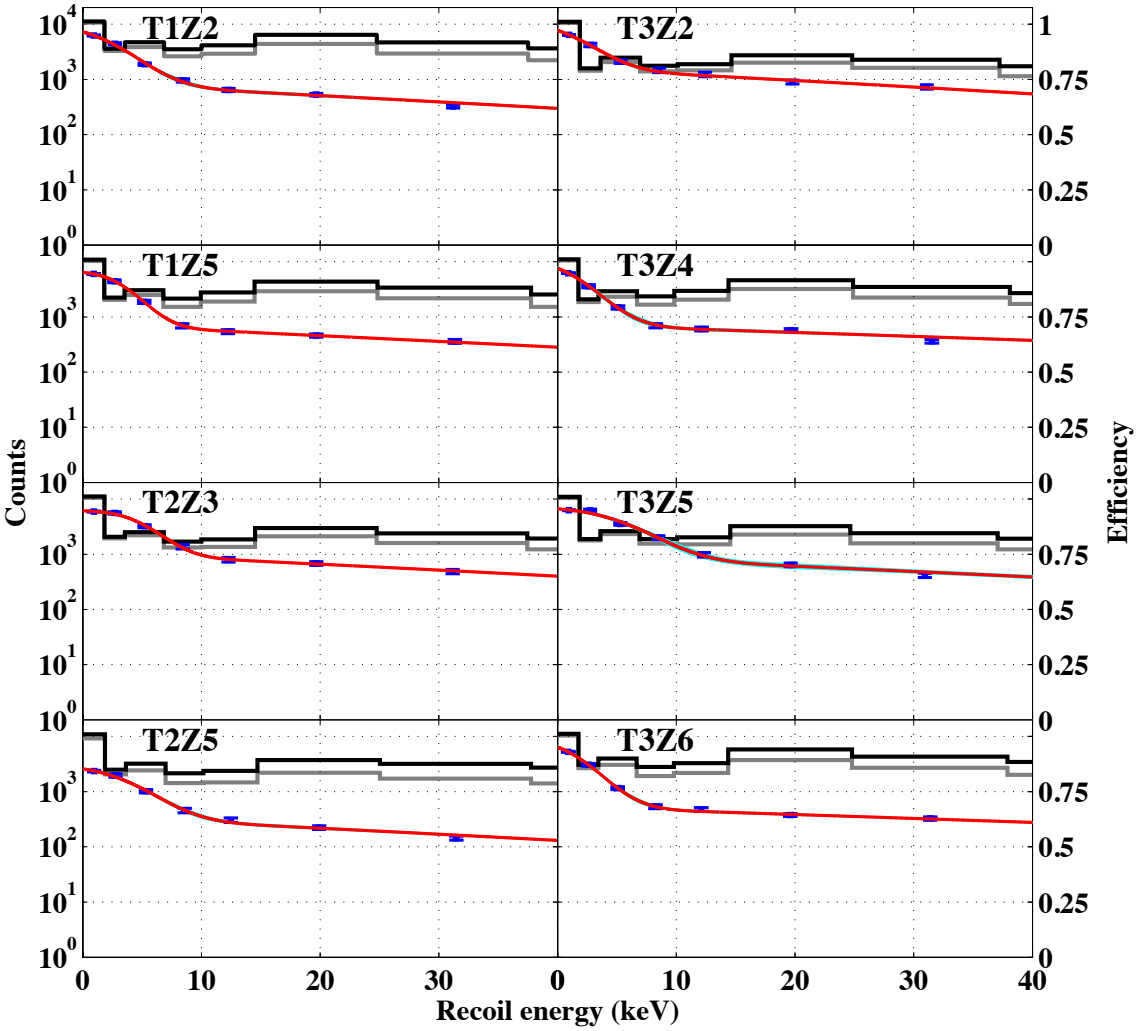


Figure 4.20: Measured fiducial volume efficiencies for each detector as a function of recoil energy, assuming a Neganov-Luke phonon contribution consistent with a nuclear recoil. All nuclear recoils in the ^{252}Cf data within 2σ of the mean of the nuclear-recoil distribution for each energy bin are shown by the black histogram, with corresponding units on the left axis. The gray histogram shows the number of events in each energy bin selected by the fiducial volume cut. The blue squares show the maximum likelihood estimate of the efficiency, with the corresponding units on the right axis. The red line shows the fit to the efficiencies, with the cyan band denoting the $\pm 1\sigma$ statistical errors determined from the covariance matrix of the fit. For nearly all detectors, the cyan band cannot be seen since its width is comparable to the width of the best-fit line.

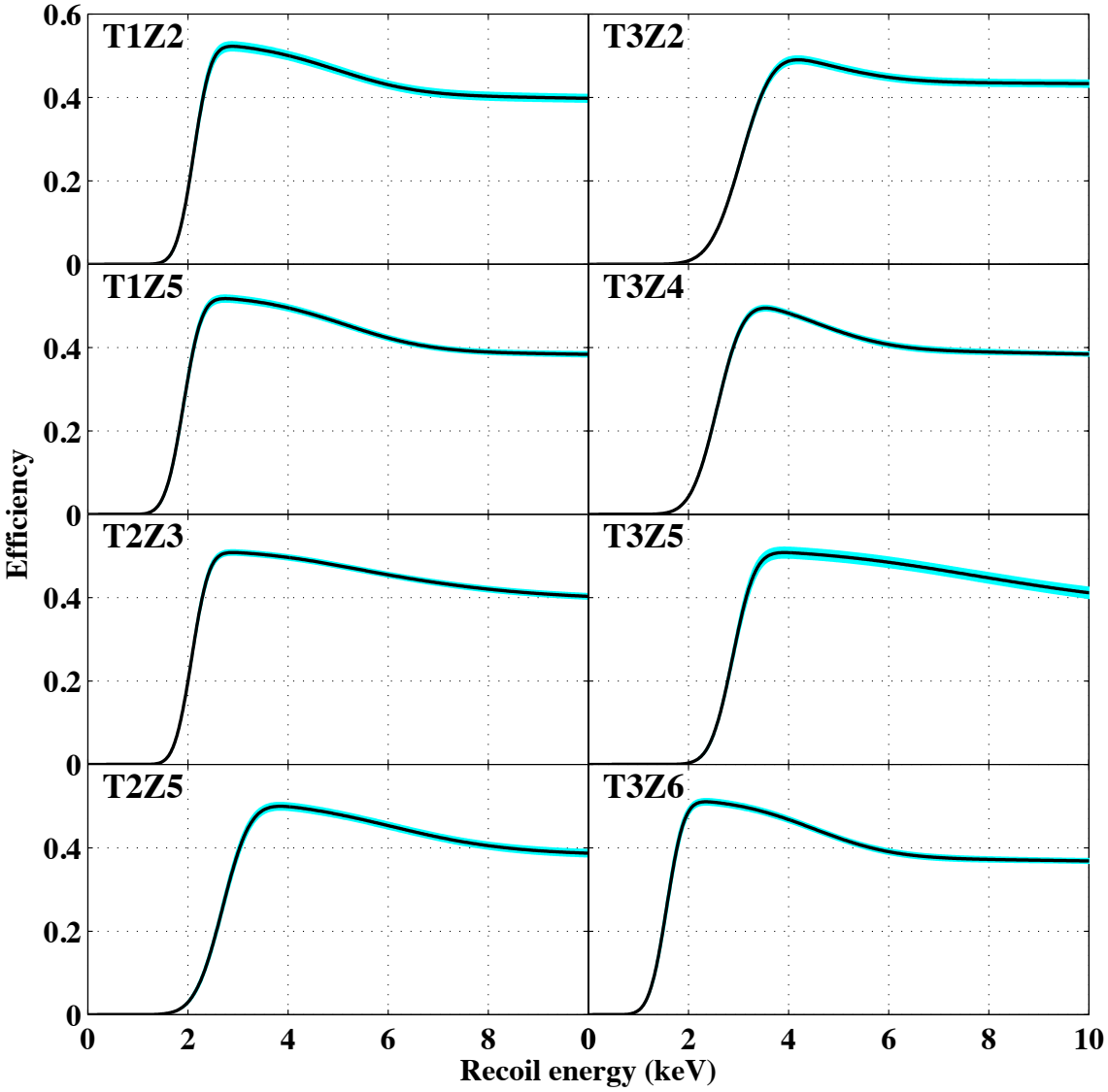


Figure 4.21: Total acceptance efficiencies (black) for each detector as a function of recoil energy, assuming a Neganov-Luke phonon contribution consistent with a nuclear recoil. The cyan band shows the $\pm 1\sigma$ statistical errors, which are typically $<5\%$ of the total efficiency over the energy range of interest.

Chapter 5

Constraints on light WIMPs

Chapter 4 defines optimized selection criteria at low-energy for identifying nuclear recoils consistent with a low-mass WIMP signal. In the following sections, we compute the nuclear-recoil candidate events at low energy (Sec. 5.1) and the corresponding limits on the WIMP-nucleon scattering cross section (Sec. 5.2). To be conservative, these limits assume that all candidate events could arise from WIMPs, even though significant backgrounds are expected at low energy, as discussed in Sec. 5.3. Systematic checks of the limits including variations of the event ordering, detector selection, and nuclear-recoil band cut are discussed in Sec. 5.4. Finally, these results are compared to the potential signals from DAMA/LIBRA, CoGeNT, and CRESST-II in Sec. 5.5.

5.1 Candidate events

The candidate events passing all selection criteria at low energy are shown in Fig. 5.1 in the ionization energy versus recoil energy plane. The distribution of neutron-induced nuclear recoils from ^{252}Cf calibration data is also shown, as well as the $(+1.25,-0.5)\sigma$ acceptance region for nuclear-recoil candidates. A total of 442 candidate events passed all selection cuts in the 2–100 keV energy range, with the vast majority appearing at low energy ($\lesssim$5 keV) as shown in Fig. 5.1. The number of candidates and total exposure by detector are shown in Table 5.1.

Several backgrounds outside of the signal region are also apparent in Fig. 5.1, including bulk electron recoils in a band above the nuclear-recoil signal region as well as events with ionization energy consistent with zero below the signal region. At low energy, these background distributions can have significant overlap with the nuclear-recoil band. Estimates

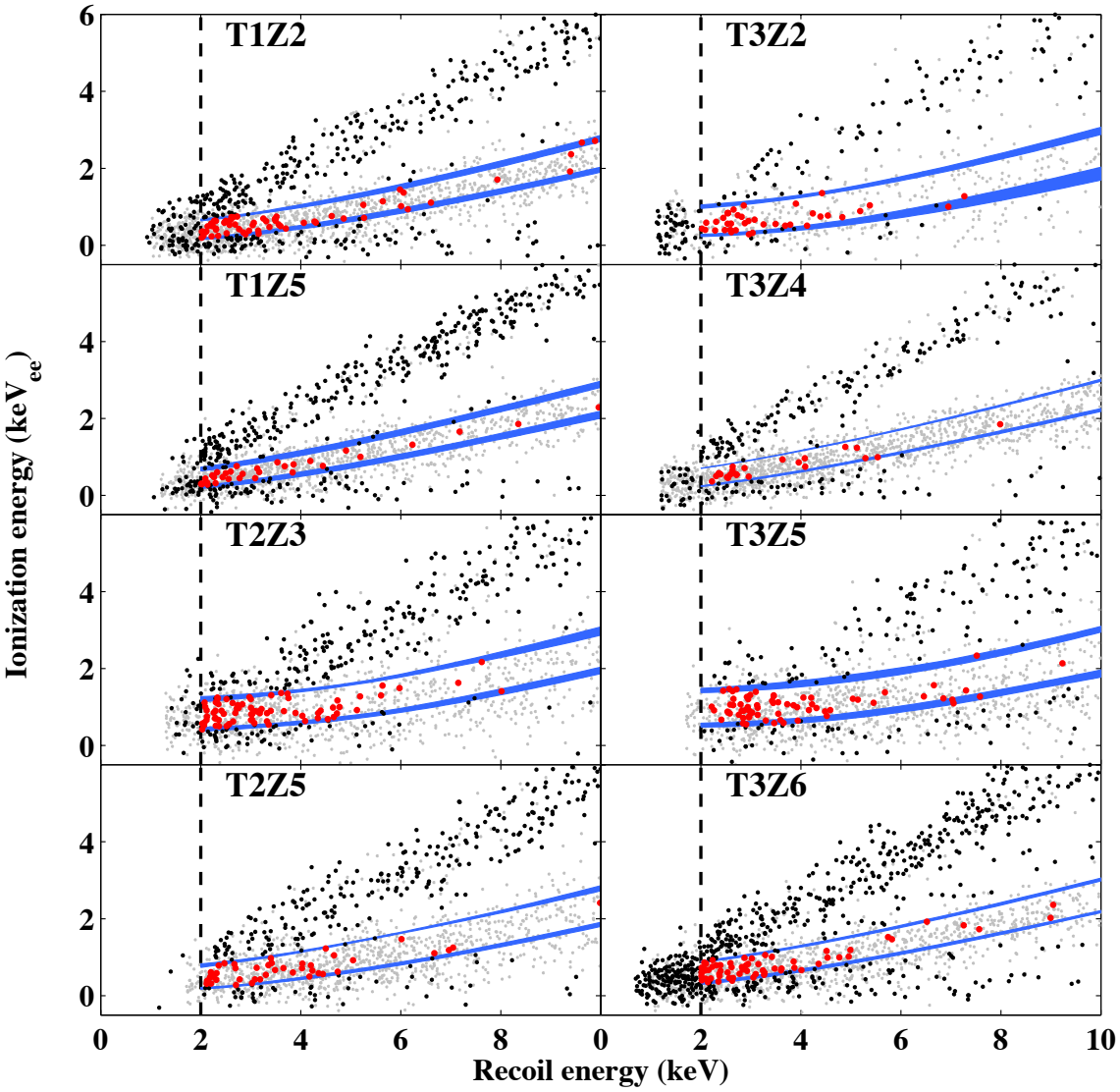


Figure 5.1: Candidate events in the ionization energy vs. recoil energy plane for each detector. Events in the WIMP search data passing all candidate selection cuts except the ionization-based nuclear-recoil band cut are shown in black. The $(+1.25, -0.5)\sigma$ nuclear-recoil selection region is defined by the blue bands, with the nuclear-recoil candidates lying within this band highlighted in red. The 2 keV recoil energy threshold for this analysis is denoted by the dashed line. The data are overlaid on the ^{252}Cf calibration data (gray dots), which shows the expected location of a nuclear-recoil signal. A phonon threshold cut requiring the reconstructed energy to be $>6\sigma$ above the mean of the readout noise distribution is applied.

of the leakage of these backgrounds into the signal region at low energy will be presented in Sec. 5.3.

As discussed in Sec. 4.2, the recoil energy for each event was determined from the

Detector:	**Number of candidates:**	**Total exposure: (kg days)**
T1Z2	65	43.4
T1Z5	38	35.0
T2Z3	77	28.0
T2Z5	45	34.7
T3Z2	42	7.8
T3Z4	25	29.6
T3Z5	70	25.8
T3Z6	80	36.1
Total:	**442**	**240.5**

Table 5.1: Number of observed nuclear-recoil candidate events from 2–100 keV and total exposure for each detector. The event rate varies significantly by detector, as expected for backgrounds which vary with each detector's ionization-based background rejection. The exposure denotes the total live time and does not include the signal acceptance efficiencies.

phonon signal alone, after subtracting the Neganov-Luke phonon contribution consistent with a nuclear recoil. This energy scale causes electron recoils to appear at higher energy than their true recoil energy, as shown in Fig. 5.2. An additional benefit of the larger Neganov-Luke phonon contribution for nuclear recoils is that the effective electron-recoil background rate, in units of events keV_{nr}^{-1} kg^{-1} day^{-1} is suppressed relative to the nuclear recoil rate. The measured constant Compton scatter rate in the electron-recoil band in CDMS is $\sim$1.8 events keV_{ee}^{-1} kg^{-1} day^{-1} in the 2–8 keV_{ee} energy region [277]. However, due to the larger total phonon signal for electron recoils relative to nuclear recoils of the same recoil energy, this background is stretched by a factor of $\sim$2, leading to a Compton rate of only $\sim$1 event keV_{nr}^{-1} kg^{-1} day^{-1} in the same energy range.

5.2 Optimum interval limits

From the candidate events shown in Fig. 5.1, we can compute upper limits on the strength of the WIMP-nucleon scattering cross section. We conservatively assume that all nuclear-recoil candidates could be from WIMPs, and determine limits using the optimum interval method [278–280]. Although this method cannot be used to identify a positive signal, it allows the determination of an upper limit on the cross section at a given WIMP mass while making no assumptions about whether an individual event is due to signal or background. As discussed in Sec. 5.3, significant backgrounds are expected for this analysis, but there are also potentially large systematic errors on the expected background distribution that

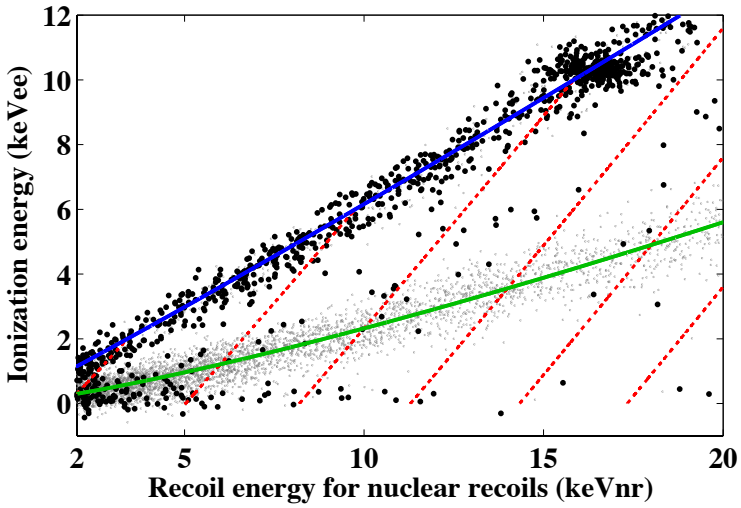


Figure 5.2: Ionization energy versus recoil energy for the ^{252}Cf calibration (gray) and WIMP search (black) data for T1Z5, assuming a recoil energy scale consistent with the Neganov-Luke phonon contribution for nuclear recoils. The means of the electron-recoil (blue) and nuclear-recoil (green) distributions determined from calibration data are also shown. The red dashed lines show contours of constant "true" recoil energy for a given ionization yield, demonstrating that the electron recoils are pushed to higher recoil energies using this scale (e.g., the 10.4 keV$_{ee}$ electron-recoil line appears at a nuclear recoil equivalent energy of 16 keV$_{nr}$). *Figure from Ahmed et al. [144]*

are difficult to quantify. Due to these uncertainties, we calculate conservative limits using the optimum interval method, which are free from any corresponding systematic errors on the background estimate. Even without detailed knowledge of the backgrounds, if the distribution of the backgrounds in some parameter is different than the expected WIMP signal, then the optimum interval method can provide stronger limits than would be possible if this difference in distributions were not taken into account.

To calculate limits using this method, the signal distribution and measured event distribution must be specified in terms of some parameter, ϵ, which is typically taken to be the recoil energy of the events. However, the best sensitivity is obtained by choosing ϵ to maximize the differences between the distribution of the signal and the expected backgrounds. For this analysis, we expect significant variations in the backgrounds by detector due to differences in the ionization-based discrimination of background events. Although the optimum interval method does not require a detailed understanding of these backgrounds, given only the knowledge that they should vary by detector we can improve the expected sensitivity of the method by specifying the measured event distribution in terms of a parameter

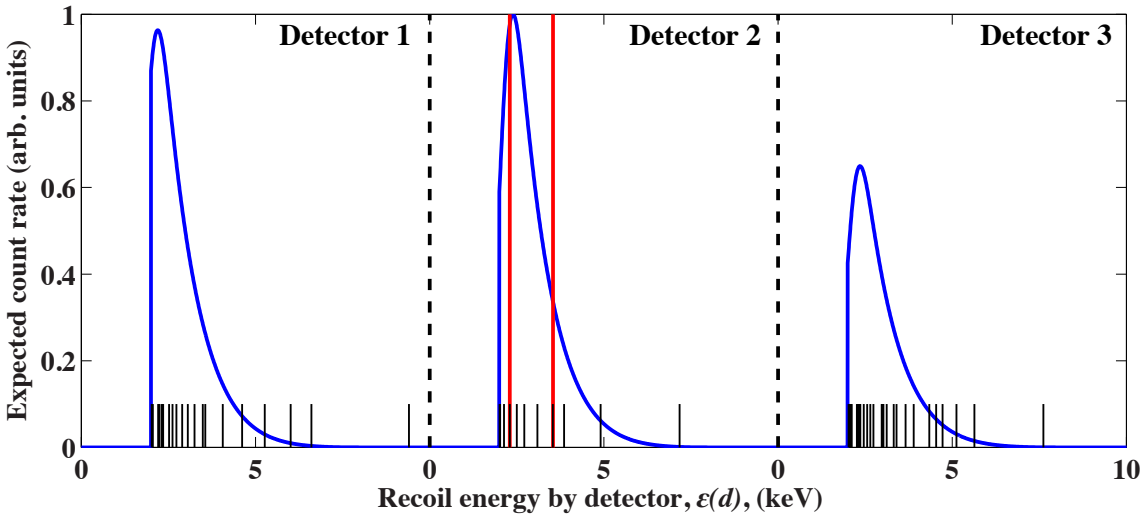


Figure 5.3: Schematic of the optimum interval method with energy intervals ordered by detector. The expected signal (blue) is calculated as a function of the recoil energy for each detector, $\epsilon(d)$, given the corresponding exposure and signal acceptance efficiency. This signal distribution is then compared to the distribution of the candidate events (black, lines) to find the "optimum" interval (denoted by the red lines) for which the strongest constraints on the signal size are obtained. The 90% CL upper limit on the signal is determined from this optimum interval using precomputed Monte Carlo simulations.

$\epsilon(d)$, which is determined both from the measured recoil energy of the event as well as the detector, d, in which it occurred.

Following a method proposed by S. Yellin, the energy distributions of the events measured by each detector are separately concatenated by detector [186,278], while maintaining the distinction between detectors as shown in Fig. 5.3. The cumulative signal probability (i.e., the probability of a signal event having a lower value of $\epsilon(d)$ than a given value) is then calculated for each observed event. For each interval containing n events and η cumulative signal probability, the probability of having $\leq n$ such events is determined by comparison to precomputed Monte Carlo simulations. For a given WIMP mass, this allows the determination of the 90% confidence level upper limit on the cross section consistent with the observed number of events. The "optimum" interval, which provides the tightest constraints on the cross section is then selected, and the appropriate statistical penalty for the freedom to choose this interval is determined from Monte Carlo.

For an event-rate limited analysis (i.e., a sufficiently large event rate that combining exposure from multiple detectors would not improve the limit) with backgrounds that vary

by detector, the optimum interval method with energy intervals concatenated by detector provides an automated method to choose the most constraining energy interval on the detector with the lowest background rate. In addition, the appropriate statistical penalty for the freedom to choose this lowest rate detector is automatically accounted for by the method. To avoid statistical biases arising from tuning the method used to calculate limits, the optimum interval method with energy intervals concatenated by detector was determined as the primary method to obtain limits prior to looking at the candidate events.

For each WIMP mass, we calculate the cumulative signal probability as a function of recoil energy and detector given the exposure and efficiencies measured for each detector in Sec. 4.6. We assume the standard halo model described in Sec. 2.1.1, with a local WIMP density at the earth of $\rho = 0.3$ GeV/cm^3, a characteristic WIMP velocity, $v_0 = 220$ km/s, and a galactic escape velocity of $v_{esc} = 544$ km/s [137, 146].

For spin-independent elastic scattering, the resulting 90% confidence upper limits on the WIMP-nucleon cross section are shown in Fig. 5.4a. The observed limits are also compared to the expected limits determined using the open data. To calculate these expected limits, 10^4 simulated candidate data sets were drawn from the distribution of events in the open data. The same limit-setting procedure was applied to these simulated data sets to determine the distribution of the expected limit as a function of WIMP mass. As shown in Fig. 5.4a, the observed limits for this analysis are in good agreement with those predicted from the open data.

Figure 5.4b shows the spin-dependent WIMP-neutron scattering limits from this analysis, which are significantly weaker in cross section due to the lack of coherent enhancement for spin-dependent scattering and small natural abundance ($\sim$8%) of the odd-neutron isotope ^{73}Ge in our target. The Ge form factor from [151] is used, and the same halo parameters as for the spin-independent calculation are assumed. The expected spin-dependent limits determined from simulated data sets drawn from the open data are also shown, following the same procedure as for the spin-independent case.

As discussed above, the limits shown in Fig. 5.4 conservatively assume all events could be from WIMPs and make no assumptions about the low-energy backgrounds. Since backgrounds are not subtracted, these limits are accurate regardless of whether the candidate events originate from backgrounds or WIMP-induced nuclear recoils. To be conservative, these limits give the primary result of this analysis as was specified before looking at the

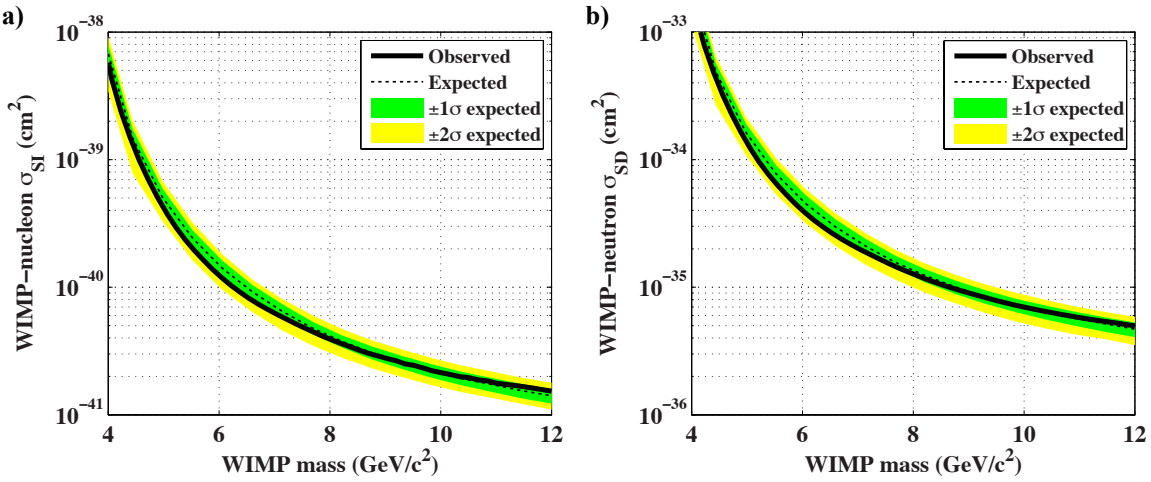


Figure 5.4: 90% CL upper limits on the WIMP-nucleon spin-independent scattering cross section (a) and the WIMP-neutron spin-dependent scattering cross section (b) versus WIMP mass. The observed limits from this analysis are shown as the thick black line. The expected limits are determined by applying the same limit-setting procedure to simulated data sets drawn from the open data.

candidate events at low energy. These results will be compared to other low-mass constraints and the potential signals reported by DAMA/LIBRA, CoGeNT, and CRESST-II in Sec. 5.5.

5.3 Expected backgrounds

Figure 5.5 reproduces the candidate events for T1Z5, and highlights several populations of events in sidebands of the signal region which are expected to leak into the signal region at low energy. These expected backgrounds include a population of events with ionization energy consistent with noise ("zero-charge" events), surface electron recoils with suppressed ionization collection, and bulk electron recoils at low energies where the ionization signal becomes consistent with noise. Each of these expected background populations is discussed in detail in the following sections, and the expected leakage into the signal region at low energy is determined.

5.3.1 Neutrons

Neutron-induced nuclear recoils cannot be distinguished from those produced from WIMPs on an event-by-event basis. However, the shielding of the experiment due to the rock

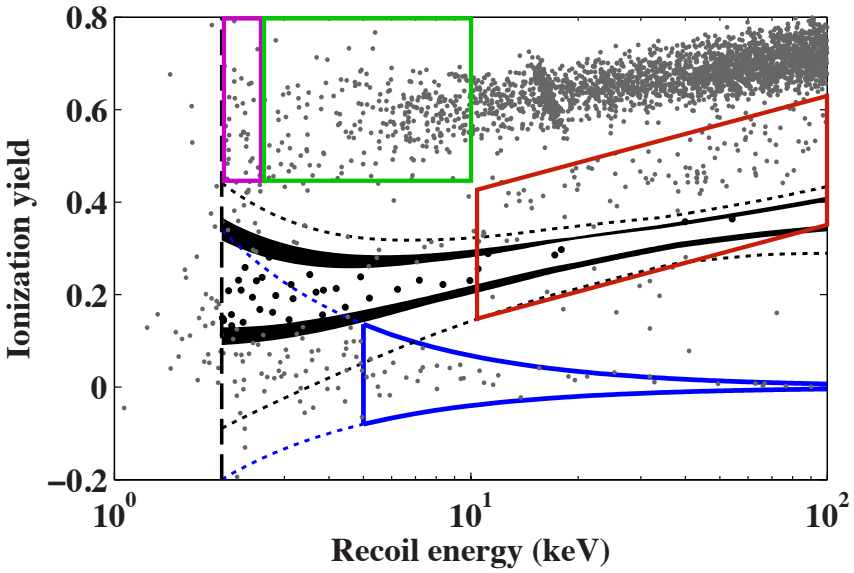


Figure 5.5: Backgrounds seen in sidebands of the data in the ionization yield versus recoil energy plane for T1Z5. The gray dots show the WIMP search events passing all cuts except the nuclear recoil band cut (black bands), while the black dots are the candidate events. Several populations of events can be identified which are expected to leak into the signal region at low energy, including zero-charge events (blue), surface electron-recoils (dark red), bulk Compton scatters (green), and bulk electron-recoils from the 1.3 keV activation line (magenta). The blue band (including the dotted lines extending below 5 keV) indicates the $\pm 2\sigma$ band consistent with the measured baseline ionization noise for this detector. The black dotted lines show the corresponding $\pm 2\sigma$ nuclear-recoil band. As can be seen in Fig. 5.1, T1Z5 has less leakage from bulk electron recoils than the average detector due to its excellent ionization resolution. Bulk electron recoil leakage is more significant for detectors such as T2Z5 and T3Z5.

overburden at the SUL, muon veto, and passive polyethylene surrounding the icebox reduce the expected background from both radiogenic and cosmogenic neutrons to <0.1 events for this exposure above 10 keV [153,178,228,230]. Extrapolating the approximately exponential radiogenic and cosmogenic spectra determined from GEANT4 simulations indicates that we expect $\ll 1$ neutron-induced nuclear recoil in the entire exposure considered above 2 keV, after summing over detectors. Thus, neutrons provide a negligible background for this analysis and are not considered further.

5.3.2 Surface events

As discussed in Sec. 3.1.4, electron recoils occurring within a $\sim$10 μm "dead layer" on the flat surfaces of the detector can have reduced charge collection relative to bulk electron recoils.

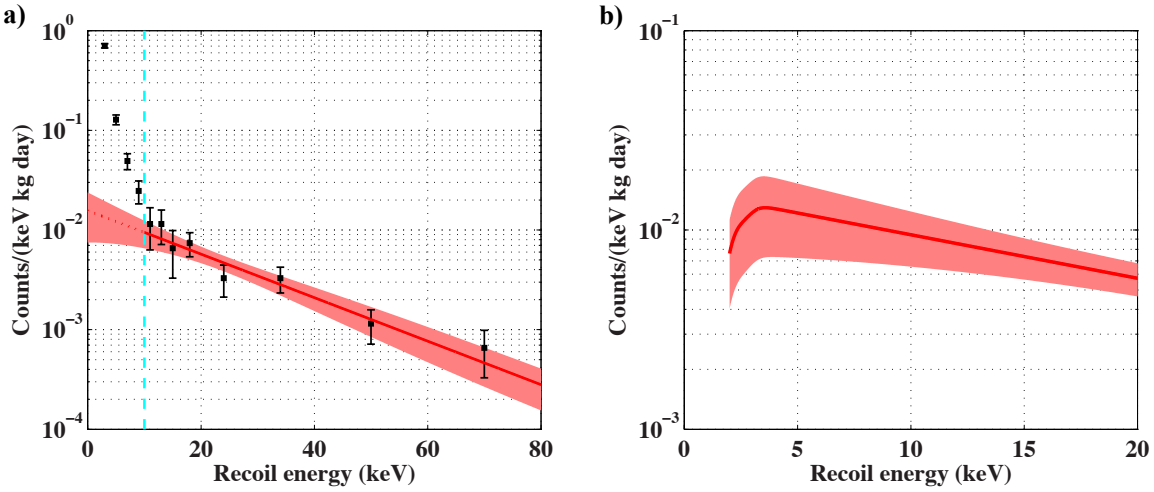


Figure 5.6: a) Rate of events in the (+1.25,-0.5)σ nuclear recoil band that are selected by the surface event timing cut (black, errorbars) after summing over detectors. The spectrum above a 10 keV threshold (cyan, dashed) is fit to an exponential (red), which is extrapolated to estimate the leakage at lower energy. Below 10 keV, an increased rate is observed corresponding to additional background components or a possible WIMP signal. b) Estimate of the background leakage due to the surface events assuming an extrapolation of the exponential spectrum measured in a) to low energy. The trigger efficiency has been applied for direct comparison with the observed rate.

These surface events are seen as a non-Gaussian tail of the electron-recoil distribution at low yield. An increasing fraction of these events can leak into the nuclear-recoil band at low energy as the signal-to-noise of the ionization measurement decreases. At energies above 10 keV, the typical penetration depth for external γs is much larger than the thickness of the dead layer, so the majority of surface events result from external e^-, which have typical penetration depths within the dead layer for energies $\lesssim$100 keV. Of these surface events, approximately half are due to e^- ejected from a neighboring surface by a γ via the photoelectric effect, and the remaining half result from βs emitted during the decay of ^{210}Pb, which is a long-lived decay product ($t_{1/2} = 22.3$ yrs) of ^{222}Rn that plates out on the detector surfaces through exposure to the air during fabrication [229, 230].

Above 10 keV, the energy spectrum of surface events leaking into the nuclear-recoil band is observed to be approximately exponential [230]. This dependence is shown in Fig. 5.6a, where we plot the rate of events in the (+1.25,-0.5)σ nuclear-recoil band that are selected by the surface event timing criteria developed for the 10 keV threshold analysis [153], after summing over all detectors. To estimate the surface event leakage at low energy, we assume

that the exponential spectrum observed above 10 keV can be extrapolated to the 2 keV threshold. For WIMP masses below 10 GeV, no WIMP-induced nuclear recoils are expected in the region above 10 keV used to determine the spectrum. The resulting expected leakage is shown in Fig. 5.6b, after extrapolating the observed exponential spectrum to low energy and accounting for the reduced trigger efficiency just above threshold.

5.3.3 Bulk electron recoils

At low energies, even well-collected bulk electron recoils can begin to leak into the nuclear-recoil signal region as their ionization signal becomes comparable to noise. Just above threshold, there is an additional contribution from the 1.3 keV activation line, which is pushed above the recoil-energy threshold for this analysis since the total phonon signal for a 1.3 keV electron recoil is equivalent to that for a 2.2 keV nuclear recoil, given the larger Neganov-Luke phonon contribution for the electron recoil.

To estimate the background from bulk electron recoils, the electron-recoil spectrum was measured from 2–10 keV after summing over all detectors, as shown in Fig. 5.7a. Only the portion of the electron-recoil band which was $>2\sigma$ above the mean of the nuclear-recoil distribution was considered, to avoid inclusion of any true nuclear recoil signal. Even if we conservatively assume all events within the $(+1.25,-0.5)\sigma$ nuclear recoil band are WIMPs, the expected number of WIMP-induced nuclear recoils above the 2σ nuclear-recoil band from 2–5 keV would be only $\sim$18 events. This would account for only $\sim$3% of the total of 593 events observed in the portion of the electron-recoil band in this region, indicating that leakage from a potential nuclear-recoil signal provides a negligible systematic for this estimate. The corresponding correction for the efficiency of selecting electron recoils only within the region above the 2σ nuclear-recoil band is included in the measured rate shown in Fig. 5.7a.

The measured rate is then fit to a Gaussian plus a constant background, after accounting for the trigger efficiency near threshold, as shown in Fig. 5.7a. To determine the expected leakage, the fraction of electron-recoil events expected to overlap with the $(+1.25,-0.5)\sigma$ nuclear recoil band as a function of energy, $\eta(E_r)$, was calculated, using the measured mean and width of the bands determined from calibration data and assuming a Gaussian

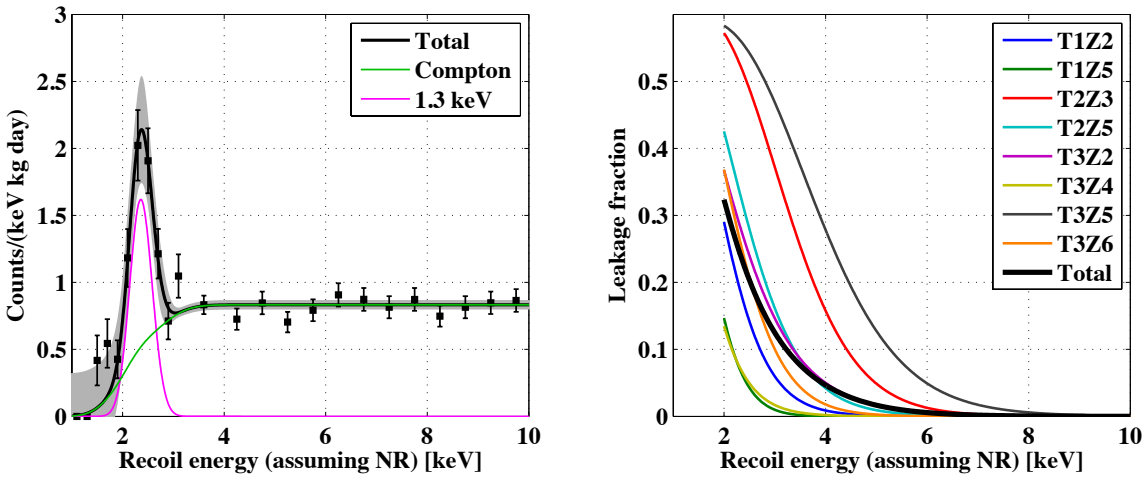


Figure 5.7: a) Measured electron-recoil spectrum versus energy after summing over all 8 Ge detectors, using the same phonon-based energy scale as for the nuclear-recoil candidate events. Only the portion of the electron-recoil distribution that is $>2\sigma$ from the mean of the nuclear-recoil distribution is considered, and the correction for the selection efficiency of electron recoils is applied. The rates are fit to a constant Compton background plus a Gaussian distribution for the 1.3 keV_{ee} activation line, after accounting for the trigger efficiency. b) Fraction of electron recoils expected in the (+1.25,-0.5)σ nuclear-recoil band for each detector. The solid black line shows the average over detectors, after weighting by the observed spectrum for each detector individually.

distribution in ionization energy:

$$\eta(E_r) = \frac{1}{2}\left[\text{erf}\left(\frac{q_{NR}^{(top)} - \mu_{ER}}{\sqrt{2\sigma_{ER}^2}}\right) - \text{erf}\left(\frac{q_{NR}^{(bot)} - \mu_{ER}}{\sqrt{2\sigma_{ER}^2}}\right)\right] \tag{5.1}$$

where $q_{NR}^{(top)}$ and $q_{NR}^{(bot)}$ denote the upper and lower limits of the (+1.25,-0.5)σ nuclear recoil band as a function of recoil energy, μ_{ER} denotes the mean of the electron recoil band and σ_{ER} denotes its width. The resulting values of $\eta(E_r)$ are shown in Fig. 5.7b, along with the spectrum-averaged value over all detectors.

The expected leakage into the nuclear-recoil signal region is then the product of the measured electron-recoil spectrum and the fraction of overlap between the populations. This leakage estimate for bulk electron recoils is shown in Fig. 5.8. Leakage from the 1.3 keV_{ee} activation line dominates at energies $\lesssim$2.5 keV, while leakage from the constant Compton background dominates the bulk electron recoil leakage at higher energies.

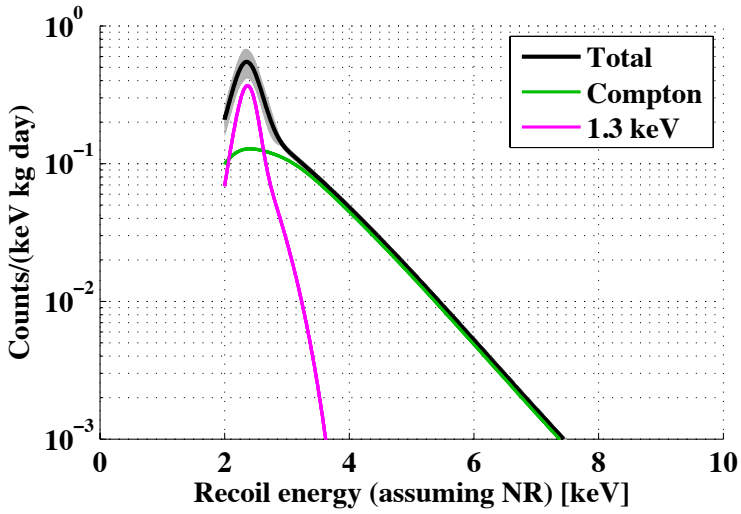


Figure 5.8: Expected bulk electron recoil leakage versus energy, using the same phonon-based energy scale as for the nuclear-recoil candidate events, after summing over detectors. The total expected leakage (black) and $\pm 1\sigma$ statistical errors (gray band) are shown, as well as the contributions from the 1.3 keV_{ee} activation line (magenta) and the constant Compton background (green).

5.3.4 Zero-charge events

As shown in Fig. 5.5, there is a significant population of events below the electron-recoil band with ionization energy consistent with zero, extending from threshold up to 100 keV recoil energy. At energies $\gtrsim$10 keV, the location of these zero-charge events can be inferred from the partitioning of energy and timing delays of the phonon signal recorded in each sensor. Figure 5.9 shows that these high-energy zero-charge events are dominated by events occurring at very high radius in the detector. These events are consistent with electron recoils where nearly all of the charge is trapped either on the cylindrical walls of the detector or by an increased density of ionized impurities at high radius, resulting in a signal in the outer electrode which is consistent with noise.

As discussed in Sec. 4.4.6, the fiducial volume cut requires that the reconstructed charge signal in the outer electrode be consistent with noise for the candidate events. Figure 5.10 shows the effect of this fiducial volume cut on the total rate of events observed at low energy, for both single or multiple-scatter events. Prior to fiducialization, the total rate of both single and multiple-scatter events is approximately constant with energy at $\sim$3.5 events keV^{-1} kg^{-1} day^{-1}, using the usual phonon-based energy scale which assumes the Neganov-Luke

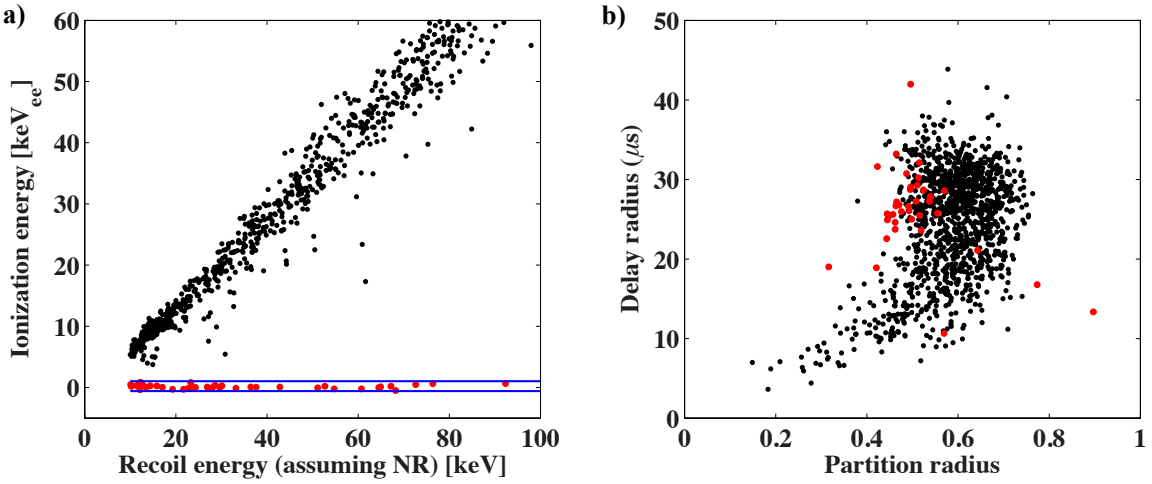


Figure 5.9: Selection of zero-charge events at high energy for T1Z2. The ionization energy distribution for events in the WIMP search data is shown in a), with the corresponding zero-charge population (red dots) selected as the events lying within $\pm 2\sigma$ of the baseline noise distribution. The corresponding reconstruction of the event location on the "shrimp" plot described in Sec. 3.1.5 is shown in b). The zero-charge events are seen to primarily lie at the "head" of the shrimp, which corresponds to events very near the cylindrical wall of the detector.

phonon contribution is consistent with a nuclear recoil. For the single-scatter events, additional contributions from the 10.4 keV$_{ee}$ and 1.3 keV$_{ee}$ activation lines are visible.

At energies $\lesssim$5 keV, the ionization-based fiducial volume cut becomes less effective and the number of events removed by the fiducial volume cut decreases, leading to an increase in the single-scatter counting rate at low energy. For a WIMP signal which is evenly distributed throughout the detector volume, this increase in the effective fiducial volume is accounted for by the acceptance efficiencies calculated in Sec. 4.6.2.3. The expected backgrounds, however, are not evenly distributed throughout the detector, as shown in Fig. 5.10. In particular, approximately 60% of the Compton single-scatter events from 5–10 keV are removed by the fiducial volume cut, since singles are more likely to occur in the outer region of the detector where the remaining energy can be deposited in the Cu detector housing rather than a neighboring detector. For the same reason, multiple scatter events are primarily concentrated in the inner fiducial volume since Compton scatters which escape from the inner fiducial volume are more likely to interact in a neighboring detector. In contrast, electron-recoils from the 10.4 keV$_{ee}$ activation line are evenly distributed throughout the detector volume and are significantly less likely to be removed by the fiducial volume cut.

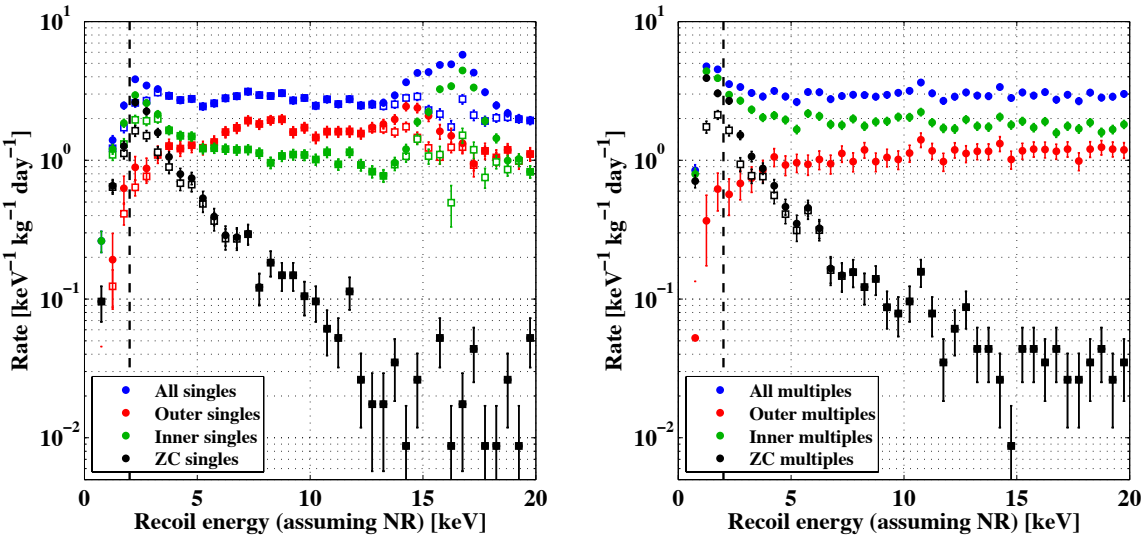


Figure 5.10: Raw counting rate for single-scatter (a) and multiple-scatter (b) interactions in the WIMP search data at low energy, summed over all 8 Ge detectors. No correction for the signal or background acceptance efficiencies has been applied. The total rate prior to applying the ionization-based fiducial volume selection is shown (blue, filled). Events with significant ionization energy in the outer electrode that are removed by the fiducial volume selection (red, filled) as well as events passing the fiducial volume selection (green, filled) are also shown. The black points show the rate of zero-charge events, which pass the fiducial volume selection and have reconstructed energy in the inner charge electrode consistent with noise. The open markers for the total, inner, and outer singles rates show the rate after subtracting the expected counts from the 10.4 and 1.3 keV$_{ee}$ activation peaks (using the rate in the K-shell peak to fix the subtraction of the L-shell peak in the region of interest). For the zero-charge events, the open markers show the rate after subtracting the expected rate of well-collected electron recoils lying within the $\pm 2\sigma$ zero-charge band.

Figure 5.10 also compares the rate of zero-charge events to the total counting rate in the detector. An approximately exponential increase in the rate of zero-charge events is seen in the energy range from 5–15 keV. The increase in rate of these events occurs as the "outer" rate of events removed by the fiducial volume cut decreases, giving an approximately constant total rate prior to fiducialization after accounting for the K-shell and L-shell peaks in the single-scatter spectrum. This increase in rate is expected for high-radius events for which the ionization signal in the outer electrode cannot be distinguished from noise. At energies above a few keV, the increasing spectrum can likely be accounted for by two effects. First, even for a high-radius event for which some fraction of the charge reaches the outer electrode, an increasing fraction have reconstructed energy which can fall below the threshold used by the fiducial volume cut as the total energy goes to zero. Second,

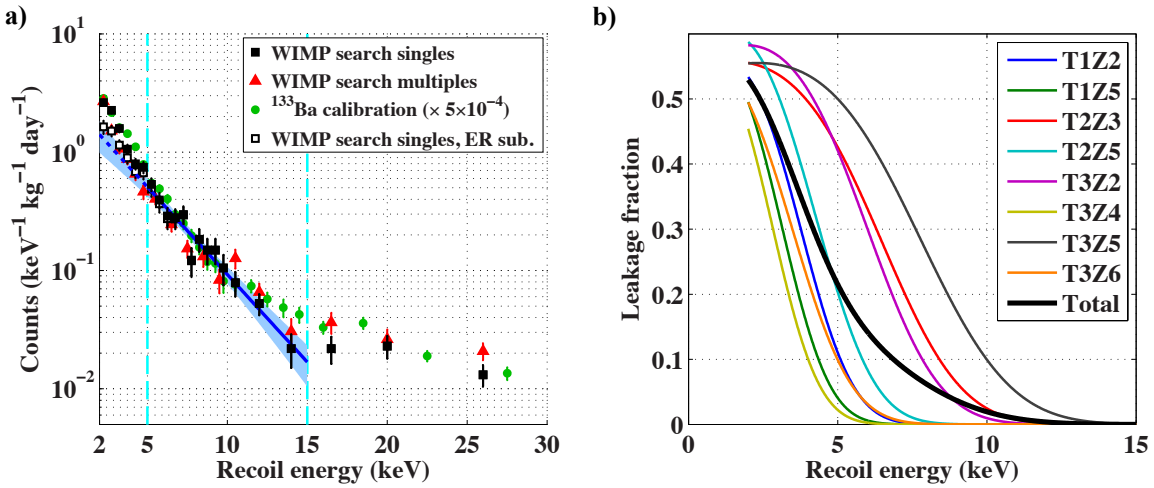


Figure 5.11: a) Fit to the observed single-scatter zero-charge event rate at low energy. The WIMP search single-scatter events within the 5–15 keV fitting window denoted by the cyan dashed lines are fit to an exponential (blue, solid), which is extrapolated to determine the expected zero-charge event rate a low energy. The blue band shows the $\pm 1\sigma$ errors determined from the covariance matrix of the fit, added in quadrature with an additional $\sim$20% systematic error due to the dependence of the extrapolation on the fitting window. b) Expected leakage fraction of zero-charge events into the $(+1.25,-0.5)\sigma$ nuclear-recoil band for each detector, assuming Gaussian distributions in ionization yield. The black curve shows the average over detectors, weighted by the observed zero-charge spectrum for each detector individually.

the decreasing penetration depth of low-energy external γs increases the probability that such events will interact in the high-radius dead layer where the ionization signal can be completely or partially trapped before reaching the electrodes. At high energies ($\gtrsim$20 keV), this rate of poorly collected events may include a significant contribution from surface e^- or recoiling ^{206}Pb nuclei resulting from $^{210}\text{Po} \rightarrow {}^{206}\text{Pb} + \alpha$, where the α is missed. However, the rate of zero-charge events at low energy in the multiple scatter data is too high for such an explanation and suggests that γ-induced electron recoils are the dominant expected background in the energy range of interest. At the lowest energies, leakage from well-collected electron recoils into the zero-charge band leads to a more sharply increasing spectrum since the ionization signal for the majority of electron recoils just above threshold is indistinguishable from noise on an event-by-event basis. The rate of zero-charge events after subtracting the expected contribution from these bulk electron recoils is also shown in Fig. 5.10.

To determine the expected leakage into the signal region at low energy for these zero-

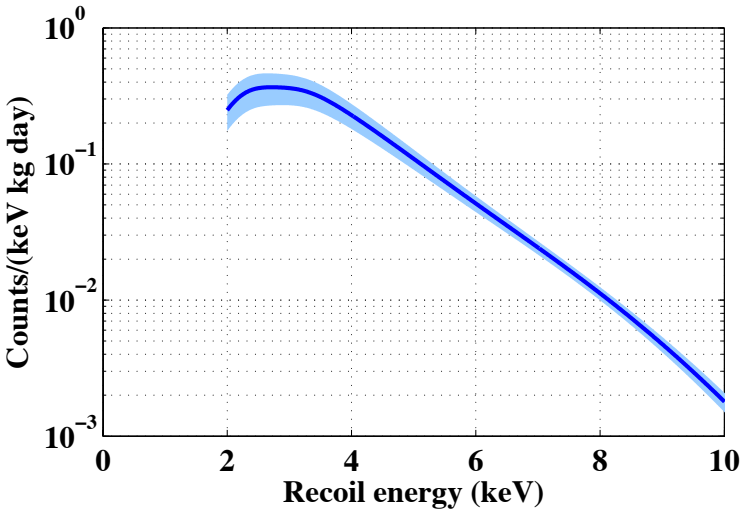


Figure 5.12: Expected zero-charge leakage versus energy after summing over detectors and extrapolating to threshold. The light blue band shows the $\pm 1\sigma$ errors from the fit.

charge events, the spectrum observed from 5–15 keV was fit to an exponential, as shown in Fig. 5.11a. At energies $\lesssim$5 keV, the rate of zero-charge events begins to have a non-negligible contribution from bulk electron recoils as the ionization signal becomes comparable to noise. To prevent double counting this background, the fit is restricted to energies above $\sim$5 keV. This lower limit also avoids contamination of the estimate of the zero-charge rate if WIMP-induced nuclear recoils are present in the data. Even under the highly conservative assumption that all events in the $(+1.25,-0.5)\sigma$ nuclear recoil band are due to WIMPs, we would expect only $\sim$15% of the single-scatter zero-charge events from 5–6 keV to be due to WIMPs. This fraction falls off rapidly with energy, so any contamination of the zero-charge population due to WIMP-induced nuclear recoils in this energy range should be negligible. Finally, we find that the expected rate above threshold varies by only $\sim$20% when the lower boundary for the fit is varied from 5–8 keV, which is included as an additional systematic error on the extrapolation to low energy shown in Fig. 5.11a.

As shown in Fig. 5.11a, the rate and spectrum of the multiple-scatter zero-charge events is similar to that for the single-scatter events. This agreement in rate is reasonable given that the total rates of single and multiple Compton scatters shown in Fig. 5.10 are similar. Zero-charge events are also seen in the ^{133}Ba calibration data with a similar spectrum to the WIMP search data. Their rate in the ^{133}Ba calibration data is significantly larger, with $\sim$4 times more zero-charge events observed from 5–10 keV in a total exposure that

was 500× smaller than the WIMP search data, leading to a rate per unit exposure that is $\sim 2\times 10^3$ higher. Although the rate of zero-charge events scales roughly with the γ rate, when normalizing by the total rate in the electron-recoil band only $\sim 1/6$ as many zero-charge events are observed in the ^{133}Ba calibration data relative to the WIMP search singles and multiples. While this comparison clearly indicates that electron recoils can produce zero-charge events with an increasing spectrum at low energy, the difference in relative rates between the WIMP search data and ^{133}Ba calibration data indicates that either a different geometrical distribution of incident γs or an additional population of zero-charge events that does not scale directly with the γ rate is present in WIMP search data.

Figure 5.11a also compares this extrapolation to the WIMP search singles rate after subtracting off the expected leakage of bulk electron recoils. The electron-recoil leakage into the zero-charge band is determined following the analogous procedure to that described in Sec. 5.3.3 for the nuclear-recoil signal region. This leakage accounts for the majority of the increase in rate above the exponential extrapolation shown in Fig. 5.11a, and gives reasonable agreement with the extrapolation below 5 keV once the additional electron-recoil leakage is removed.

Analogous to the estimation of the bulk electron recoil leakage in Sec. 5.3.3, the fraction of zero-charge events expected to overlap with the $(+1.25,-0.5)\sigma$ nuclear-recoil band was calculated, following Eq. 5.1. This leakage fraction as a function of energy is shown in Fig. 5.11b. The total expected zero-charge leakage, shown in Fig. 5.12, is then the product of the extrapolated rate of events below 5 keV and this leakage fraction.

5.3.5 Comparison to candidate events

Figure 5.13 compares the sum of the background estimates presented in Sec. 5.3.2–5.3.4 to the observed rate of candidate events at low energy after summing over detectors. The background estimates above agree well with the observed rate, and there is not significant evidence for an excess of candidate events above these backgrounds. The dominant expected background for WIMPs with masses from 5–10 GeV/c^2 are zero-charge events, which provide the largest expected leakage in the 2.5–8 keV range, after summing over detectors. For the detectors with the best ionization resolution, surface events can dominate the leakage from zero-charge events well below 8 keV. Just above threshold, leakage of bulk electron recoils from the 1.3 keV$_{ee}$ activation line contributes a comparable expected rate to the

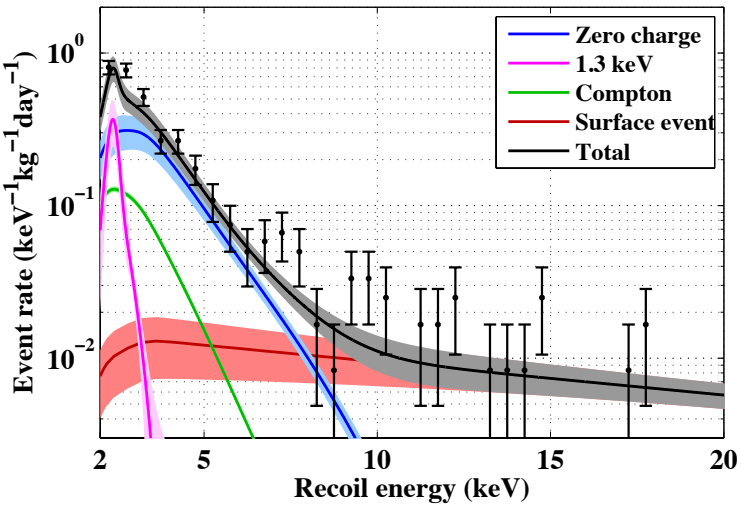


Figure 5.13: Comparison of the expected backgrounds to the measured rate at low energy, summed over all 8 Ge detectors. The expected background from each contribution is shown as a solid line, with the $\pm 1\sigma$ errors shown as the light band. Only the statistical errors from the fits to the distributions are shown, while systematic errors due to the extrapolation to low energy are not included. The total background estimate (black) agrees well with the total observed rate (errorbars). No correction for the nuclear recoil acceptance efficiencies has been applied.

zero-charge events. However, for $m_\chi \gtrsim 5$ GeV/c^2, this energy range is avoided by the optimum interval limit-setting procedure and further reduction of the 1.3 keV$_{ee}$ background by excluding longer periods after the ^{252}Cf calibrations would not have significantly improved the sensitivity of this analysis.

Despite the good agreement between the observed rate of candidate events and the background estimate, we cannot definitively exclude the possibility that the extrapolation of the zero-charge or surface-event single-scatter background is overestimated at low energy. If the exponential extrapolations do not accurately describe the background spectra, a WIMP signal or additional unknown background contribution could be present just above threshold. For this reason, the limits presented in Sec. 5.2 made no assumptions about the low-energy backgrounds and are valid regardless of the true nature of these events.

Such an overestimate of the zero-charge event background is unlikely since it would require significant differences between the spectrum of single-scatter and multiple-scatter zero-charge events below 5 keV, despite their good agreement above 5 keV. The presence of a corresponding multiple-scatter component from 2–5 keV strongly suggests a background

origin for these events. As shown in Fig. 5.11, the spectrum is also in good agreement with the low-energy rate of electron-recoil induced zero-charge events in ^{133}Ba calibration data.

5.4 Systematic checks of limits

In the following sections, we calculate the sensitivity of the limits discussed in Sec. 5.2 to various choices made in the limit and cut setting procedure. Although the limits presented in Sec. 5.2 are statistically unbiased since the limit-setting procedure was completely specified prior to looking at the candidate events at low energy, these checks give a sense of the dependence of the limits on the cut positions, detector selection, and combination of detectors used to determine the optimum interval method limits.

5.4.1 Detector selection and event ordering

As discussed in 5.2, the primary results from this analysis were calculated by concatenating the candidate event energies by detector. For a background limited analysis where backgrounds vary by detector, this procedure provides an automated way to choose the most constraining energy interval on the detector with the lowest backgrounds while including the appropriate statistical penalty for the freedom to choose this interval. Figure 5.14 compares the primary result to the limits that would be obtained by applying the optimum interval method to T1Z5 alone. The effect of the statistical penalty for the freedom to choose the most constraining detector is $\sim$30%, leading to slightly weaker limits than if T1Z5 had been identified as the detector expected to produce the strongest constraints prior to looking at the low-energy candidate events. In retrospect, T1Z5 could have been identified as the most constraining detector ahead of time since it also produces the strongest constraints for the open data, and it is the Ge detector with the best ionization resolution.

As was expected for the large number of observed candidate events, the most constraining energy interval for the primary result was completely contained within the candidate events for a single detector for the mass range shown in Fig. 5.4. Accordingly, the limits have no dependence on the ordering of detectors used when concatenating the energy intervals by detector. Nonetheless, the ordering of detectors used to calculate limits was specified to be in order of increasing detector number, starting at the top of T1 and ending at the bottom of T3, prior to looking at the low-energy candidate events.

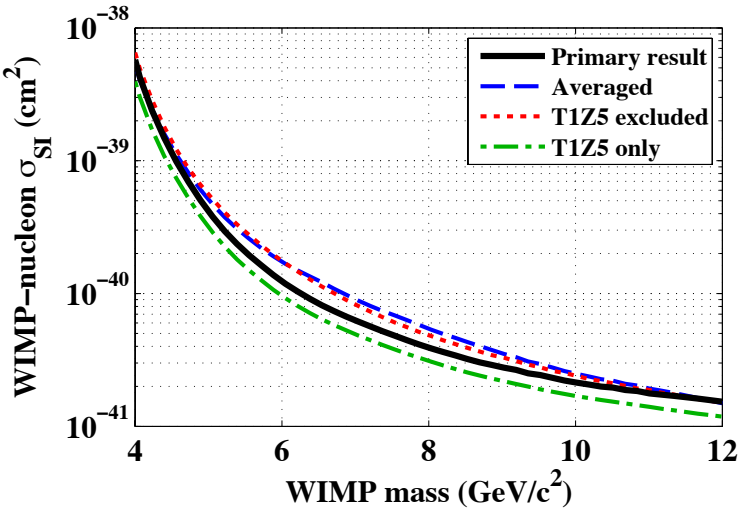


Figure 5.14: Comparison of limits from the primary result (black, solid) to limits computed by varying the detector selection or event ordering used for the optimum interval method. The limits calculated for T1Z5 alone (green, dot-dashed) demonstrate the statistical penalty imposed on the primary result for the freedom to choose the best detector in the limit-setting procedure. The "averaged" limits (blue, dashed) are computed by ordering events by recoil energy regardless of the detector in which they occurred. The optimum interval limits computed for events concatenated by detector, but excluding T1Z5, are also shown (red, dotted).

An alternative method of combining data from multiple detectors is to simply sum the candidate events over detectors without maintaining the distinction between individual detector elements. This is the standard practice used for previous CDMS analyses with a 10 keV recoil energy threshold [153,178], where backgrounds are expected to be comparable for each detector. The optimum interval limit-setting procedure is then performed with the candidate events ordered by energy. Figure 5.14 also shows the limits calculated with this ordering, which ignores the variation in expected backgrounds with detector. The limits for this "averaging" method are ∼40% weaker than the primary result at $m_\chi = 8$ GeV/c^2.

Finally, the optimum interval limits were again computed after concatenating energy intervals by detector, but with T1Z5 excluded. The resulting limits, shown in Fig. 5.14 are ∼30% weaker at $m_\chi = 8$ GeV/c^2 than the primary result where T1Z5 is included. In this case, the most constraining energy interval is entirely contained within the events from T3Z4 at WIMP masses $\gtrsim$6 GeV/c^2. It is not surprising that T3Z4 provides the strongest constraints in this mass region after T1Z5 is removed since it has the best ionization resolution of the remaining detectors.

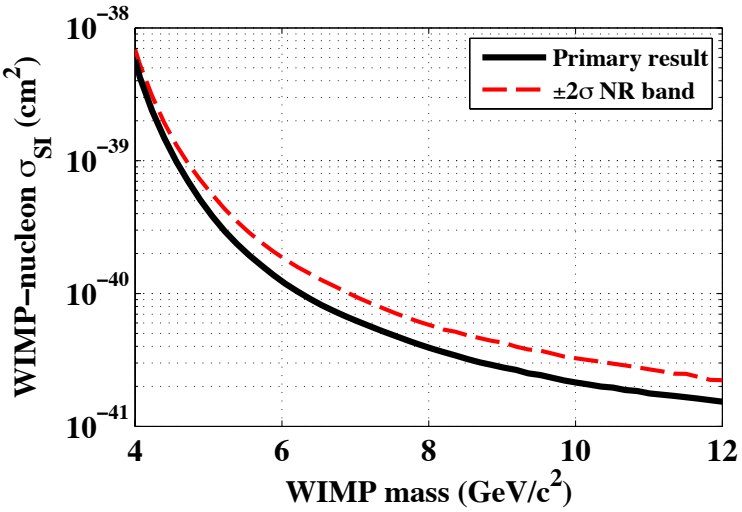


Figure 5.15: Comparison of optimum interval limits versus ionization-based nuclear recoil selection criteria. The primary result (black, solid), which used an optimized (+1.25,-0.5)σ nuclear recoil acceptance band gives limits which are a factor of $\sim$1.4 stronger at 8 GeV/c^2 than those obtained using a looser $\pm 2\sigma$ band (red, dashed).

5.4.2 Effect of ionization-based nuclear recoil selection

We can also calculate the effect of the (+1.25,-0.5)σ nuclear recoil band cut used for the primary result, relative to a looser $\pm 2\sigma$ cut which has better signal acceptance, but at the cost of higher expected backgrounds. A comparison of the limits obtained using the $\pm 2\sigma$ cut with the primary result is shown in Fig. 5.15, using the optimum interval method with energy intervals concatenated by detector. Events from T1Z5 again contain the most constraining energy interval, which is not surprising for the looser cut where backgrounds become more significant.

The optimized (+1.25,-0.5)σ nuclear-recoil band cut gives limits which are a factor of $\sim$1.4 stronger at $m_\chi = 8$ GeV/c^2, relative to the looser $\pm 2\sigma$ ionization-based selection. If the candidate events were distributed in ionization energy in the same way as the expected nuclear-recoil signal distribution determined from the ^{252}Cf calibration data, then we would not expect a significant change in the limits determined with either band cut after accounting for the relative change in signal acceptance. The improvement in the limits with the tighter band cut results from the presence of backgrounds in the WIMP search data that are preferentially distributed above and below the nuclear-recoil signal region.

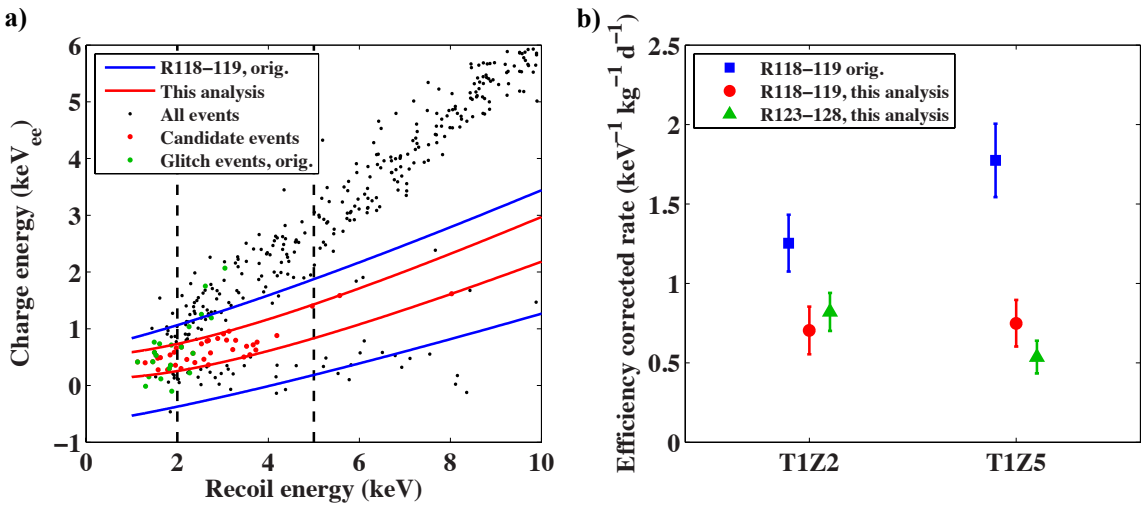


Figure 5.16: a) Comparison of nuclear-recoil candidate events in the R118–R119 data set for the analysis presented in [232] and using the cuts developed for the current analysis. The energy range considered in [232] was 2–5 keV, denoted by the black dashed lines. All single-scatter events in the WIMP search data passing the candidate selection criteria other than the ionization-based nuclear-recoil band cut are shown in black. The blue band denotes the $\pm 2\sigma$ nuclear-recoil selection region used in [232], while the red band denotes the (+1.25,-0.5)σ band from this analysis. Events removed by the glitch rejection cuts used in [232] are shown in green. b) Comparison of the rate of candidate events from 2–5 keV between the data sets and analyses. After applying the same cuts as used in this analysis, the rates observed in R118–R119 and R123–R124 are in good agreement.

5.4.3 Comparison to R118-R119

W. Ogburn also presented a rough low-energy analysis of data taken with T1Z2 and T1Z5 during R118–R119 at the SUL in Appendix B of his thesis [232]. Although presented only as a proof-of-principle analysis, the rate of low-energy events found in his analysis was larger than in the present analysis [281], most notably for T1Z5 which gave a factor of $\sim$2.5 times higher rate. However, this analysis used very different selection criteria at low energy than the current analysis of R123–R128 data. In particular, the ionization-based nuclear-recoil selection and electronics glitch removal cuts used in this analysis have been optimized for use at low energy relative to the cuts used in [232]. Figure 5.16 shows a comparison of the observed rate in R118–R119 and R123–R128 using the same cuts for both data sets.

As seen in Fig. 5.16, the primary difference between these analyses for T1Z5 is the use of the pulse-shape based electronics glitch removal cuts for the original analysis in [232], which have only $\sim$50% acceptance of nuclear recoils while removing $<$10% of the candidate events

from 2–5 keV. This cut alone leads to nearly a factor of 2 increase in the efficiency corrected rate relative to the current analysis, in which the glitches are effectively removed using their trigger multiplicity. In addition, the optimized $(+1.25,-0.5)\sigma$ nuclear-recoil selection region used in this analysis provides better rejection of events appearing near the edges of the wider $\pm 2\sigma$ nuclear-recoil selection region used in [232], leading to a further factor of $\sim$1.5 increase in the efficiency corrected rate for [232], relative to the current analysis. As discussed in [232], the $\pm 2\sigma$ bands used for the original analysis were extrapolated in the yield versus recoil energy plane from fits to higher energy calibration data, leading to a selection region which is too wide at low energy. Figure 5.16b shows that after comparing the candidate events using the same selection criteria, the rates observed in R118–R119 and R123–R124 are in good agreement.

5.5 Comparison to DAMA/LIBRA, CoGeNT, and CRESST

The primary motivation for this analysis was to test the potential signals reported by the DAMA/LIBRA, CoGeNT, and CRESST-II experiments, which are described in Sec. 2.1.4. Since both CDMS and CoGeNT use the same target material, the observed rate in each experiment can be directly compared, independent of the assumed astrophysics or particle physics model. As shown in Fig. 5.17, the observed event rate in CDMS is incompatible with the hypothesis that the majority of the low-energy excess events in CoGeNT are due to WIMP-induced nuclear recoils.

It has recently been proposed that only $\sim$25% of the low energy excess events in CoGeNT could be due to WIMPs, with the majority of the excess instead arising from electron recoils occurring on the detector surfaces, which can leak past surface-event rejection criteria at low energy [192]. In this case, the residual excess rate in CoGeNT is comparable to that observed for T1Z5 in CDMS. While the analysis presented here cannot test this residual CoGeNT excess, a maximum-likelihood analysis that accounts for backgrounds could be sensitive enough if the understanding of the backgrounds can be demonstrated to be accurate enough. However, as discussed in Sec. 2.1.4, if the annual modulation observed in CoGeNT is also due to WIMPs, then this would require a modulation fraction of nearly 50–100% in CoGeNT, which is an order of magnitude larger than would be expected in the standard halo model. A recent analysis of the same data presented in this thesis does not find a corre-

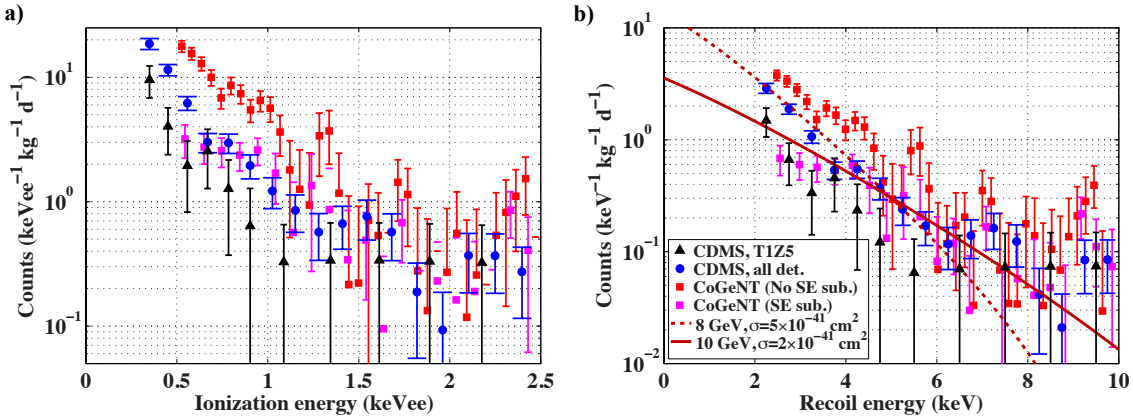


Figure 5.17: a) Comparison of observed candidate event rate at low energy for CDMS to the ionization energy spectrum measured by CoGeNT. The energy scale for CDMS uses the total phonon signal alone, and is converted to ionization energy in units of keV$_{ee}$ using the measured nuclear-recoil ionization yield for CDMS. The total low-energy excess in CoGeNT is shown after subtracting off the constant Compton background and L-shell activation lines (red, squares) as well as after subtracting off recent estimates of the surface event contamination at low energy (magenta, squares) following [192]. The CDMS spectra summed over all detectors (blue, circles) and for T1Z5 alone (black, triangles) are also shown. b) Same comparison as in a), but converting the CoGeNT data to recoil energy using the quenching factor assumed in [82]. This quenching factor is slightly higher than the ionization yields measured by CDMS, causing the spectra to appear less discrepant than when plotted versus ionization energy. The expected rate for two example WIMP models with masses ∼10 GeV are overlaid. Note that no background subtraction is employed for the CDMS data, even though backgrounds are expected to be significant, as described in Sec. 5.3.

sponding modulation in the low-energy CDMS event rate from 5–12 keV$_{nr}$ [203], disfavoring a low-mass WIMP explanation for the modulation in CoGeNT from 1.2–3.2 keV$_{ee}$ at >98% confidence and suggesting either an origin from backgrounds or a statistical fluctuation for the modulation found by CoGeNT in this energy range. However, the CDMS modulation analysis did not directly test the energy range below 1.2 keV$_{ee}$, where the majority of the WIMP signal would be expected.

Since DAMA/LIBRA and CRESST-II use different target nuclei than CDMS, comparison of the observed rate between experiments depends on the model of the WIMP velocity distribution and the details of the WIMP-nucleon interaction. To compare with our results, we assume spin-independent elastic scattering with equal coupling to the protons and neutrons, $f_p = f_n$. As in Sec. 5.2, the standard halo model is used with a local WIMP density at the earth of $\rho = 0.3$ GeV/cm^3, a characteristic WIMP velocity, $v_0 = 220$ km/s, and a

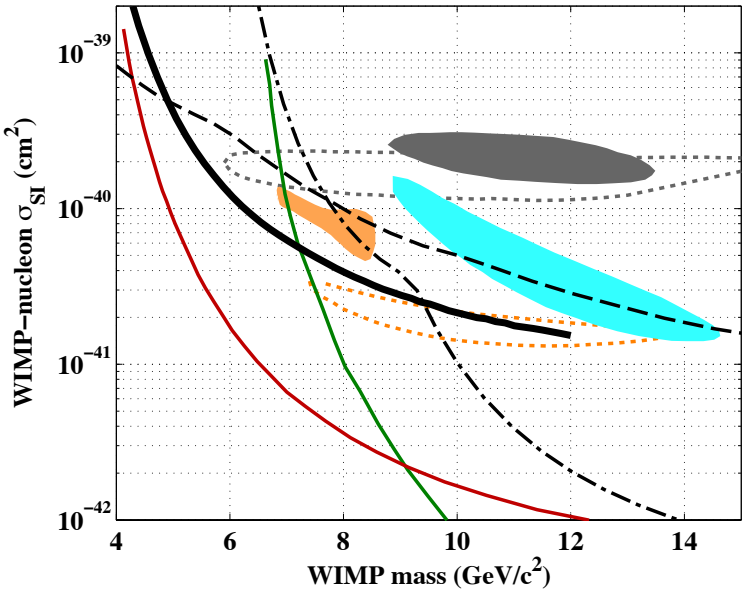


Figure 5.18: Comparison of limits from this analysis to previous results in the scattering cross section versus WIMP mass plane, assuming spin-independent elastic scattering. The 90% CL exclusion limits from this analysis (thick, black) disfavor parameter space consistent with CoGeNT [189] (orange, filled), DAMA/LIBRA [176] (gray, filled), and CRESST-II [154] (cyan, filled). An alternative calculation of the CoGeNT allowed region after subtracting the expected surface event background [190, 192] (orange, dotted) is also shown, as well as an alternative calculation of the DAMA/LIBRA allowed region which allows for larger quenching factors at low energy [82] (gray, dotted). We ignore the effect of ion channeling on the DAMA/LIBRA allowed regions since recent analyses indicate channeling should be negligible [194, 282]. Exclusion limits from the combined CDMS II data with a 10 keV threshold [153] (dash-dotted), the low-threshold analysis from the shallow site [186] (dashed), XENON100 [54] (green, solid), and a low-threshold analysis of the XENON10 data [156] (red, solid) are also shown.

galactic escape velocity of $v_{esc} = 544$ km/s [137, 146]. The resulting limits from Fig. 5.4a are compared to the parameter space consistent with other experiments in Fig. 5.18.

As shown in Fig. 5.18, these results disfavor an interpretation of the DAMA/LIBRA and CRESST-II experiments in terms of spin-independent scattering of WIMPs with $m_\chi <$ 10 GeV/c^2 at greater than 90% confidence, given the standard assumptions about the WIMP coupling and halo model discussed above. At the time of publication, these results [144] were the most constraining in the 5–9 GeV/c^2 mass range, although more recent results from a low-threshold analysis of the XENON10 data provide stronger constraints [156]. Given the uncertainties in the WIMP model, astrophysics models, and detector response at low energy, these results still provide useful constraints for models in which the XENON10

constraints can be evaded.

Chapter 6

Design of highly multiplexed athermal phonon sensors

6.1 Introduction and motivation

In the previous chapters we have discussed the constraints on WIMP models placed by the current generation of dark matter experiments, which have begun to probe the parameter space favored by supersymmetric WIMPs. Figure 6.1 shows the expected sensitivity of future extensions of the CDMS technology to the WIMP-nucleon spin-independent elastic scattering cross section. To detect WIMPs, or to exclude the remaining parameter space will ultimately require detectors with $\sim$1 ton or larger target mass. These projections assume that the background-free operation of existing experiments can be maintained, even while significantly increasing the size of the experiment.

To maintain background-free operation as the target mass increases requires either eliminating existing backgrounds with improved shielding or reduced contamination of the detector materials, or improving the ability to discriminate against backgrounds, by a factor comparable to the increase in mass. Since neutron backgrounds cannot be distinguished from WIMPs on an event-by-event basis, future experiments will need to operate at deeper sites such as DUSEL [283] or SNOLAB [253] and with larger passive neutron moderators (polyethylene or water shields) to eliminate cosmogenic neutron backgrounds and radiogenic neutrons from external sources. Careful screening of all materials inside the passive shielding is needed to reduce internal neutron backgrounds.

Unlike neutron backgrounds, electron-recoil backgrounds can also be reduced by improving the discrimination of the detector between nuclear and electron recoils. For the CDMS technology, large gains are most easily made by improving background discrimina-

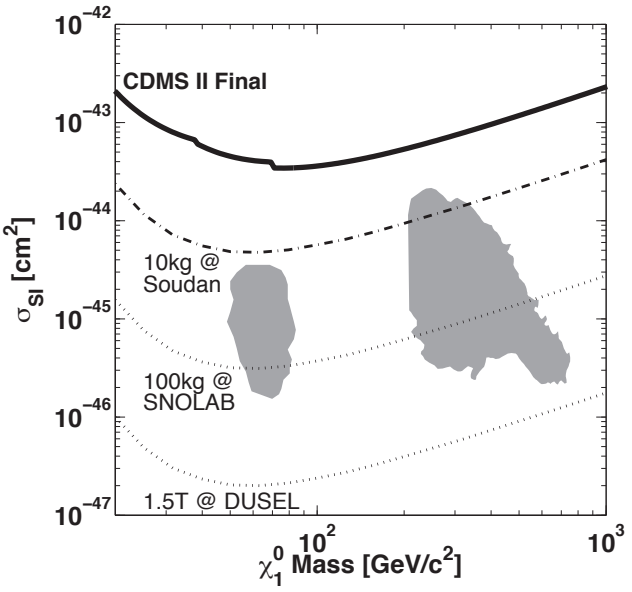


Figure 6.1: Favored parameter space in the cMSSM (gray) at the 99% CL in the cross section versus WIMP-mass plane. The favored regions are determined from profile likelihood fits including LHC and direct detection constraints [40], as discussed in Sec. 1.3.2.1. The solid and dashed curves indicate the current and future sensitivity for CDMS, assuming ≈0 background operation can be maintained.

tion, although improvements in screening of detector materials and shielding will also be implemented. New detector designs using interleaved charge electrodes (iZIPs) have shown conservative lower limits on the rejection of surface events that are $>3\times10^4$:1 from the charge signal alone [284]. Since these constraints are limited by neutron contamination at the surface test facilities, underground tests at Soudan to establish the ultimate surface rejection of these detectors are in progress.

As shown in Fig. 6.2, the interleaved charge electrodes provide a different field geometry for bulk and surface interactions, allowing these two populations of events to be differentiated. Near the surface, the electric field is primarily parallel to the surface and ionization will be collected only on a single detector face. In contrast, bulk events see a field perpendicular to the surface and will be collected symmetrically on both faces. When combined with current phonon-timing and yield-based discrimination, these detectors appear to have sufficient rejection that surface-event backgrounds will not be a limiting background for ton-scale experiments.

Although the path to extending the current TES-based iZIP technology to the ton scale

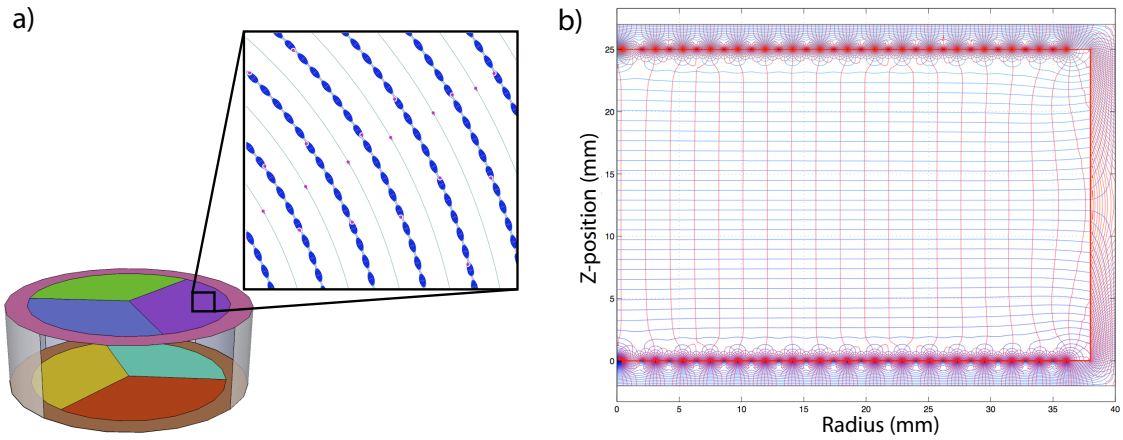


Figure 6.2: a) Schematic of the iZIP sensor layout. The TESs on each face are segmented into 3 inner phonon sensors, and an outer phonon guard ring. The zoom shows the arrangement of the interleaved charge (narrow, gray) and phonon (wide, blue) electrodes. The spacing between the charge and phonon lines is ∼1 mm. b) Cross section of the detector showing the field lines (red) and potentials (blue) resulting from the interleaved electrode geometry, biased at (+V,0) for the charge and phonon electrodes on the top face, and (-V,0) for the electrodes on the bottom face. Near the detector faces, the field is parallel to the surface, while throughout the bulk, the field is perpendicular to the surface. *Figures from S. Hertel and M. Pyle [284]*

is clear, it is not without significant technical challenges or costs. In Sec. 6.2, we briefly describe the major obstacles to extending TES-based designs to ton-scale detectors. We then describe a concept for multiplexed, highly granular phonon sensors based on microwave kinetic inductance detectors (MKIDs), and show how this technology mitigates many of the problems inherent in scaling TES-based detectors to a large experiment (Secs. 6.3–6.4). Finally, we describe the design and readout of prototype MKID-based athermal phonon sensors (Secs. 6.5–6.6).

6.2 TES-based design path

A ton-scale detector requires an increase in target mass of more than 2 orders of magnitude over the ∼5 kg CDMS II experiment described in Chap. 3. To make such a large experiment feasible requires a significant decrease in the cost of fabricating individual detector elements. For CDMS II detector production, the dominant cost per detector element (∼85%) was due to detector fabrication and testing, rather than raw materials. A simple scaling of this model is not feasible for a ton-scale experiment. Although many factors influence production times

and costs, a significant driver was the use of tungsten TESs, whose T_c was poorly controlled between depositions and across a single detector. Since the CDMS II design used a single-stage SQUID array and many TESs were wired in parallel to produce each $\sim$10 cm^2 sensor, a high-sheet resistance material is needed for the TES. Tungsten is one of the few elemental superconductors with the desired properties, but W films must be ion-implanted to control the transition temperature and uniformity [248]. This tuning requires additional testing and fabrication time to measure the T_c versus position for each detector, implant the detector, and retest to verify the TES properties.

Decreasing fabrication and testing time likely requires replacing the tungsten TESs with a material for which it is easier to reliably produce films with the desired properties. With the use of multistage SQUIDs, the need for a high-resistivity material can be avoided since the SQUID input coil inductance can be reduced, and the L/R time constant can still allow $\sim\mu$s resolution for significantly lower resistance. This reduced requirement on the resistivity allows more engineering freedom to find a robust TES material. A potential drawback is the proliferation in the number of SQUIDs, which provides an additional source of complexity and potential for loss of detector channels.

Even with more uniform materials that eliminate the need for ion-implantation, experience with CDMS II detectors indicates that significant position-based dependence will remain in the shape and magnitude of the phonon signal. In addition to the radial and depth dependence due to reflection of phonons from the substrate surfaces, any residual variation in TES properties over the device will lead to small scale variations in the response. Misreconstruction of the position and energy of the interaction due to these variations was a primary limitation to background rejection in CDMS II. Improving phonon-based position reconstruction and background rejection will require more finely segmented phonon sensors. Although the charge-based rejection in the iZIP appears to add sufficient discrimination for rejection of surface events in a ton-scale experiment, the fundamental limits to this discrimination are still unclear. Rare events such as multiple internal surfaces scatters (shown in Fig. 6.3), or interactions near low-field saddle points could ultimately limit the rejection of these detectors. Increasing the spatial resolution of the phonon sensors would allow an additional handle on these events, improving the ultimate rejection of the detectors.

Highly segmented sensors will require multiplexing multiple sensors on a single readout channel to reduce the cost and heat load that would be required to have a separate channel

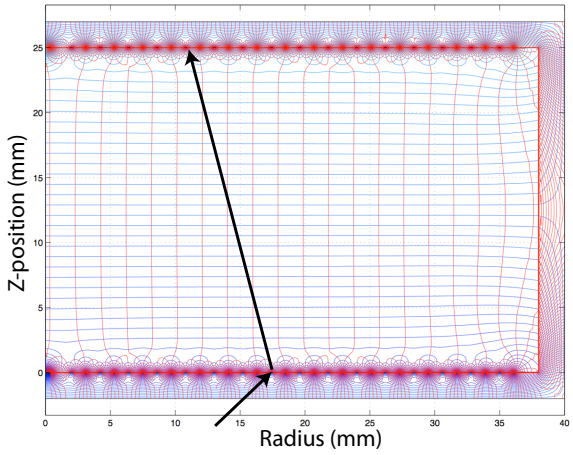


Figure 6.3: Example of a rare event topology which could evade the charge-based surface event rejection of the iZIP. A multiple internal scatter which deposits similar energy near both detector faces will appear to be a bulk event since energy is collected on both sides, but may have depressed ionization collection. Simulations indicate that the rate of such events should be $\lesssim$1 per 10^5 electron recoils, but such events could become a limiting background in large experiments [285]. Phonon-based information is needed to identify and reject such events.

for each sensor. Many SQUID multiplexing systems suitable for reading out TESs exist, in the time domain [286–288], code domain [289,290] or frequency domain [291–293]. Although these systems add significant complexity to the cryogenic readout electronics, rapid development of SQUID multiplexers for submm/mm and x-ray astronomy will continue to drive development and reliability. The extension of these multiplexing systems to CDMS technology is non-trivial due to the large bandwidths (100–200 kHz) needed to resolve the $\sim\mu$s rise times of the phonon signal. Continued development of code-domain or frequency-domain systems will likely be required to meet CDMS's bandwidth requirements.

The proposed extensions to TES-based CDMS detector technology described above increase the sensitivity of the experiment at the cost of increases in complexity. In the following sections, we describe a new sensor technology using microwave kinetic inductance detectors (MKIDs) [294]. As shown below, this technology naturally allows for highly segmented athermal phonon sensors with significantly reduced cryogenic readout complexity. In addition, MKIDs can be patterned from a single deposited film, with large (>10 μm) features, reducing fabrication complexity and increasing reliability.

6.3 Microwave kinetic inductance detectors

MKIDs are a new class of superconducting detector that allow for a naturally frequency-domain multiplexed readout and simple sensor design. Each sensor consists of a superconducting LC resonator operating at GHz frequencies. Thousands of sensors can be read out using a single coaxial line. MKIDs eliminate the need for complicated cryogenic readout systems like the multistage SQUID multiplexers used in large TES arrays, moving the readout complexity to room-temperature electronics. The development of this room-temperature signal processing hardware is driven by the communications industry, leading to rapid scaling of the capabilities of readout systems with time. MKIDs are currently being developed for submm/mm [295], optical/UV [296], and x-ray astronomy [297]. In the following sections, we describe the adaptation of MKIDs for use as athermal-phonon sensors in a CDMS-style detector.

6.3.1 Operating principle

The operating principle of MKIDs is shown in Fig. 6.4. A superconducting film provides the inductor in an electrical resonator. Although the superconductor has no DC resistance, the inertia of the Cooper pairs introduces a nonzero AC impedance. Since the Cooper pairs do not scatter, when an electric field is applied to the superconductor, energy can be stored in the kinetic energy of the Cooper pairs. When the field is reversed there is a phase lag between the field and the current as the Cooper pairs are decelerated. This appears as an additional inductance in the superconducting state, known as the "kinetic inductance." Since the kinetic inductance arises from the kinetic energy of the Cooper pairs, if an excitation in the superconductor breaks Cooper pairs, then the average momentum per Cooper pair increases for the same current. This leads to an increase in the kinetic inductance, pushing the resonant frequency of the circuit lower. In addition, the resulting quasiparticles increase the dissipation of the resonator, reducing the quality factor. By monitoring the phase and amplitude of the signal transmitted past the resonator, the frequency and dissipation in the resonant circuit can be measured, allowing the determination of the quasiparticle number.

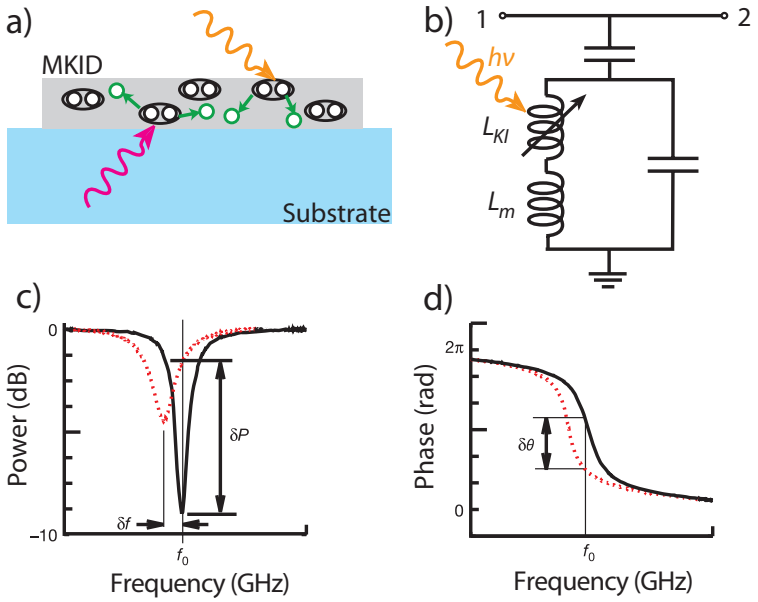


Figure 6.4: MKID detection principle. a) Schematic cross-sectional view of an MKID. Phonons from the substrate (light blue) or photons incident on the superconducting film (light gray) with energies $>2\Delta$ can break Cooper pairs (black circles), creating quasiparticles (green circles). b) Circuit diagram of MKID. The kinetic inductance L_{KI} acts as a variable inductor in an LC circuit weakly coupled to a microwave feedline. Incident energy that breaks Cooper pairs causes an increase in the kinetic inductance and dissipation in the circuit. c) Microwave transmission versus frequency before (black solid) and after (red dotted) incident energy breaks Cooper pairs in the film. The increased inductance shifts the resonance to lower frequency, while the increased dissipation decreases the quality factor. d) Phase response versus frequency before (black solid) and after (red dotted) incident energy breaks Cooper pairs. A phase shift, $\delta\theta$, is seen when monitoring the transmission at a given frequency. *Figure adapted from Day et al. [294]*

6.3.1.1 Surface impedance response

The surface impedance of a superconductor can be written as:

$$Z_s = R_s + jX_s = R_s + j\omega L_s \tag{6.1}$$

where at non-zero temperature, the surface resistance R_s arises from dissipation due to quasiparticles and the surface inductance, L_s, depends on the kinetic inductance, L_{KI}. We denote the "kinetic inductance fraction," $\alpha = L_{KI}/L$ as the fraction of the total inductance ($L = L_{KI} + L_m$) due to kinetic inductance. The magnetic inductance, L_m, depends only on geometry and is independent of the Cooper pair density. For $T \ll T_c$, typically the surface

reactance, $X_s \ll R_s$, since the thermal quasiparticle density vanishes.

The superconducting films considered in this thesis lie in the local, thin film limit, where the mean free path, l, is limited by the film thickness, t, and is much smaller than the coherence length and London penetration depth, $l \ll \xi_0, \lambda_L$. In this limit, the surface impedance can be related to the complex conductivity, $\sigma = \sigma_1 - j\sigma_2$, as [298]:

$$Z_s = \frac{1}{(\sigma_1 - j\sigma_2)t} \tag{6.2}$$

The complex conductivity in the superconducting state was given by Mattis and Bardeen in [299] as:

$$\sigma_1 = \frac{\sigma_n}{\hbar\omega}\left[2\int_{\Delta}^{\infty}\left(f(E) - f(E+\hbar\omega)\right)g(E)dE + \int_{\Delta-\hbar\omega}^{-\Delta}\left(1-2f(E+\hbar\omega)\right)g(E)dE\right] \tag{6.3}$$

$$\sigma_2 = \frac{\sigma_n}{\hbar\omega}\int_{\Delta-\hbar\omega}^{-\Delta}\left(1-2f(E+\hbar\omega)\right)g(E)dE \tag{6.4}$$

where $f = 1/(\exp(E/k_bT)+1)$ is the Fermi function, and:

$$g(E) = \frac{E^2 + \Delta^2 + \hbar\omega E}{\sqrt{\Delta^2 - E^2}\sqrt{(E+\hbar\omega)^2 - \Delta^2}} \tag{6.5}$$

In the general case, these integrals must be evaluated numerically. However, in the low frequency ($\hbar\omega \ll \Delta$) and low temperature ($k_bT \ll \Delta$) limit, Gao et al. [300] have derived approximate analytic formulas for the response of the complex conductivity due to pair breaking:

$$\frac{\sigma_1}{\sigma_n} = \frac{2\Delta_0}{\hbar\omega}\frac{\delta n_{qp}}{N_0\sqrt{2\pi k_bT\Delta_0}}\sinh\left(\frac{\hbar\omega}{2k_bT}\right)K_0\left(\frac{\hbar\omega}{2k_bT}\right) \tag{6.6}$$

$$\frac{\sigma_2}{\sigma_n} = \frac{\pi\Delta_0}{\hbar\omega}\left(1 - \frac{\delta n_{qp}}{2N_0\Delta_0}\left[1 + \sqrt{\frac{2\Delta_0}{\pi k_bT}}e^{-\frac{\hbar\omega}{2k_bT}}I_0\left(\frac{\hbar\omega}{2k_bT}\right)\right]\right) \tag{6.7}$$

where $\Delta_0 = \Delta(T=0)$ and δn_{qp} is the change in quasiparticle density. In the approximations below, $k_bT \ll \Delta$, so $\Delta \approx \Delta_0$, and we drop the subscript on Δ_0 in the following sections. In

the thin film limit, Eqs. 6.1 and 6.2 indicate that:

$$\frac{\delta Z_s}{Z_s} = \frac{\delta\sigma}{\sigma} \quad \Rightarrow \quad \frac{R_s(n_{qp})}{X_s(0)} = \frac{\sigma_1}{\sigma_2(0)} \tag{6.8}$$

$$\frac{\delta X_s(n_{qp})}{X_s(0)} = \frac{\delta\sigma_2}{\sigma_2(0)} \tag{6.9}$$

where we have used $R_s(0) = \sigma_1(0) = 0$. We can then write the change in the surface resistance and reactance due to a change in the quasiparticle density using Eqs. 6.6 and 6.7 as:

$$\frac{R_s}{X_s(0)} = S_1(\omega, T)\frac{\delta n_{qp}}{2N_0\Delta} \tag{6.10}$$

$$\frac{\delta X_s}{X_s(0)} = S_2(\omega, T)\frac{\delta n_{qp}}{2N_0\Delta} \tag{6.11}$$

where we have defined:

$$S_1(\omega, T) = \frac{2}{\pi}\sqrt{\frac{2\Delta}{\pi k_b T}}\sinh\left(\frac{\hbar\omega}{2k_bT}\right)K_0\left(\frac{\hbar\omega}{2k_bT}\right) \tag{6.12}$$

$$S_2(\omega, T) = 1 + \sqrt{\frac{2\Delta}{\pi k_b T}}e^{-\frac{\hbar\omega}{2k_bT}}I_0\left(\frac{\hbar\omega}{2k_bT}\right) \tag{6.13}$$

6.3.1.2 Resonator response

MKIDs are typically operated as $\lambda/4$ distributed resonators or lumped element LC circuits which are weakly coupled to a microwave feedline, as shown in Fig. 6.4b. Following Gao [298], we can write the transmission past the device as a function of frequency, f:

$$S_{21} = 1 - \frac{Q/Q_c}{1 + 2jQ\left(\frac{f-f_0}{f_0}\right)} \tag{6.14}$$

where Q is the total quality factor of the resonator, $Q^{-1} = Q_c^{-1} + Q_i^{-1}$. The coupling quality factor, Q_c, denotes energy lost from the resonator to the feedline through the coupling capacitor, while the internal quality factor, Q_i^{-1}, includes all other energy losses in the resonator, e.g., due to dissipation from quasiparticles, grain boundaries, or radiation. The resonant frequency is given by $\omega = 1/\sqrt{LC}$. Fig. 6.5 plots the resonator response versus frequency from Eq. 6.14 in the IQ plane, where I and Q are defined as $S_{21} = I + jQ$.

By measuring the complex transmission data for a frequency sweep around the res-

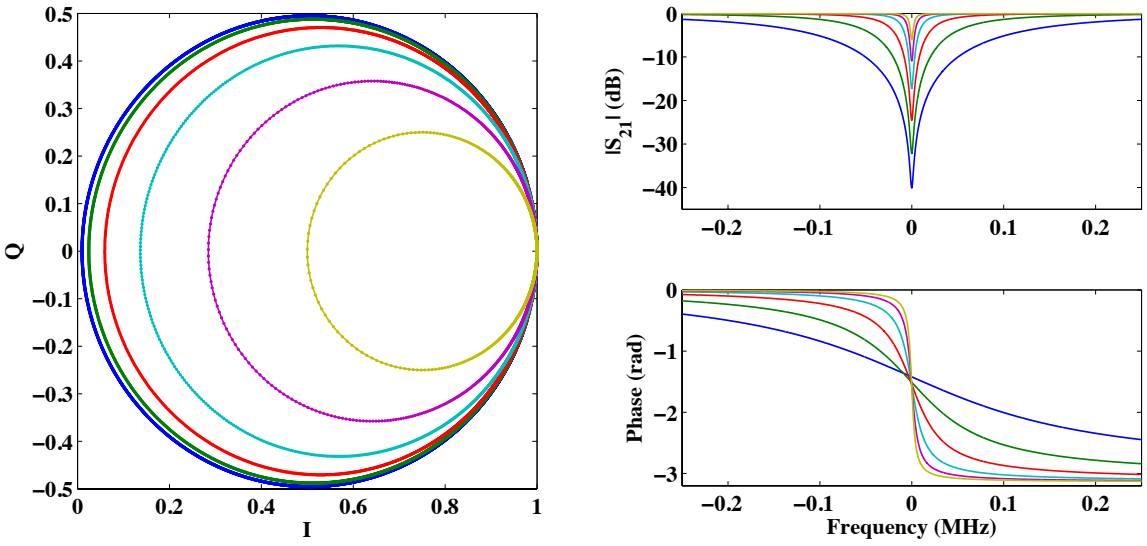


Figure 6.5: (left) Transmission past the resonator in the IQ plane as a frequency tone is swept past the resonance for $Q_i = 10^6$ and Q_c=1×10^4 (blue), 2.5×10^4 (green), 6.3×10^4 (red), 1.6×10^5 (cyan), 4.0×10^5 (magenta), and 1×10^6 (yellow). An upward frequency sweep traces out the resonance loop in the counterclockwise direction. (right) Amplitude (top) and phase (bottom) of the transmission past the resonator versus frequency.

onance, f_0, Q_c and Q_i can be determined by fitting the resonance loop to Eq. 6.14, as described in [298]. In particular, several of these parameters can be read off directly from the resonance loop using Eq. 6.14. The width of the resonance feature is determined by the resonator Q. In the limit where $Q_c \ll Q_i$, so $Q \approx Q_c$, then:

$$Q = \frac{f_0}{\Delta f} \tag{6.15}$$

where Δf is the full width of the resonance feature at half the maximum transmission (-3 dB). The diameter of the resonance circle gives the ratio between the resonator Q and coupling Q, which can be inverted to solve for Q_i:

$$d = \frac{Q}{Q_c} = \frac{Q_i}{Q_i + Q_c} \tag{6.16}$$

Finally, the resonant frequency, f_0, can be determined by the point at which the imaginary portion of the transmission vanishes (i.e., the transmission is 180° out of phase). In practice, impedance mismatches, cable delays, and parasitic inductances can rotate the resonance loop in the IQ plane. In this case, f_0 can still be easily determined from the resonance loop

as the frequency where $|dS_{12}/df|$ is maximized.

The response of the resonator to a change in f_0 or Q_i can be calculated from Eq. 6.14:

$$\delta\mathrm{Im}[S_{21}] = \frac{2Q^2}{Q_c}\frac{\delta f_0}{f_0} \quad \Rightarrow \quad \delta\theta = 4Q\frac{\delta f}{f} \tag{6.17}$$

for θ measured relative to the center of the resonance loop, where the second relation uses the small angle approximation, $\delta\theta \approx \delta\mathrm{Im}[S_{21}]/(d/2)$. In addition, we have:

$$\delta\mathrm{Re}[S_{21}] = \frac{Q^2}{Q_c}\,\delta\left(\frac{1}{Q_i}\right) \tag{6.18}$$

In practice, the internal Q saturates at some low-temperature value, $Q_{i,max}$ due to loss mechanisms other than thermal quasiparticles. In this case, we can write $Q_i^{-1} = Q_{i,qp}^{-1} + Q_{i,max}^{-1}$, where $Q_{i,qp}$ is the internal quality factor due to thermal (or phonon-generated) quasiparticles alone.

We can combine Eqs. 6.10 and 6.11 with the resonator response from Eq. 6.17 and Eq. 6.18 to determine the change in the transmission past the resonator due to a change in the quasiparticle density. Since $\omega = 1/\sqrt{LC}$ and $Q_{i,qp} = R/\omega L$, we have:

$$\frac{\delta f_0}{f_0} = -\frac{\alpha}{2}\frac{\delta X_s}{X_s(0)} \tag{6.19}$$

$$\delta\left(\frac{1}{Q_i}\right) = \alpha\frac{R_s}{X_s(0)} \tag{6.20}$$

where we have used that L_m is constant, and only a fraction α of the total inductance, L, is due to kinetic inductance. Finally, this gives the resonator response to a change in quasiparticle density, δn_{qp}:

$$\delta S_{21} = \frac{\alpha Q^2}{Q_c}\frac{\delta n_{qp}}{2N_0\Delta}\left(S_1(\omega,T) - jS_2(\omega,T)\right) \tag{6.21}$$

6.3.1.3 Amplifier limited sensitivity

This section derives the amplifier limited sensitivity of an MKID following the calculation by J. Zmuidzinas in [301, 302]. The noise contributed by an amplifier with noise temperature, T_N, is given by:

$$\delta S_{21} = \sqrt{\frac{k_b T_N}{4P_f\tau_{qp}\chi_{BW}}} \tag{6.22}$$

where P_f is the feedline power reaching the HEMT and the quasiparticle lifetime τ_{qp} determines the noise bandwidth. The factor $\chi_{BW} \leq 1$ accounts for the decrease in resolution if the phonon energy arrives on an extended timescale relative to τ_{qp}. In the limit that the rise time of the phonon signal is much shorter than its fall time, $\tau_{rise,ph} \ll \tau_{ph}$, then [303]:

$$\chi_{BW} \approx \frac{\tau_{qp}}{\tau_{ph} + \tau_{qp}} \tag{6.23}$$

χ_{BW} approaches unity in the limit that the phonon signal arrives as a δ-function.

Using the resonator response from Eq. 6.21, we can write the minimum resolvable change in the number of quasiparticles as:

$$\delta N_{qp} = \frac{Q_c}{\alpha Q^2} \frac{2N_0 \Delta V}{S_1(\omega, T)} \sqrt{\frac{k_b T_N}{4 P_f \tau_{qp} \chi_{BW}}} \tag{6.24}$$

where V is the resonator volume, and we have assumed that dissipation readout is used (i.e., the real part of δS_{21} is measured). For a given feedline power, P_f, the power absorbed by the quasiparticles on resonance is [298]:

$$P_r = \frac{2Q^2}{Q_{i,qp} Q_c} P_f \tag{6.25}$$

For sufficiently low temperatures such that the thermal quasiparticle density is negligible, quasiparticles generated by the readout power dominate, and the quasiparticle number is:

$$N_{qp} = \eta_{read} \frac{P_r \tau_{qp}}{\Delta} \tag{6.26}$$

where $\eta_{read} \leq 1$ is the efficiency for the readout power to create quasiparticles. These quasiparticles cause a corresponding change in the internal Q:

$$Q_{i,qp} = \frac{X_s(0)}{\alpha R_s} = \frac{2N_0 \Delta V}{\alpha S_1(\omega, T) N_{qp}} \tag{6.27}$$

Optimum sensitivity is achieved when $Q_{i,qp} \approx Q_{i,max}$, since further increases in readout power will generate sufficient quasiparticles to reduce Q_i. In this limit, the feedline power is:

$$P_f = \frac{Q_c Q_{i,qp}}{2Q_r^2} P_r = \frac{Q_c}{Q_r^2} \frac{1}{\eta_{read} \tau_{qp}} \frac{N_0 \Delta^2 V}{\alpha S_1(\omega, T)} \tag{6.28}$$

Combining Eq. 6.24 and 6.28 gives:

$$\delta N_{qp} = 2\sqrt{\frac{\eta_{read} k_b T_N N_0 V}{\chi_c \chi_{qp} S_1(\omega, T) \alpha Q_{i,qp} \chi_{BW}}} \tag{6.29}$$

where we have defined $\chi_c = 4Q^2/Q_i Q_c \leq 1$ and $\chi_{qp} = Q_i/Q_{i,qp} \leq 1$. Sensitivity is maximized when $Q_i = Q_c \Rightarrow \chi_c = 1$.

For frequency readout, dielectric noise from two-level systems (TLS) [298,304] dominates the amplifier noise for many applications. However, for the large geometry resonators described in Sec. 6.5.1, amplifier noise is dominant at the signal frequencies of interest. Frequency readout then benefits from the additional response in the frequency direction:

$$\delta N_{qp}^{freq} = \frac{\delta N_{qp}^{diss}}{\beta(\omega, T)} \tag{6.30}$$

where:

$$\beta(\omega, T) = \frac{S_2(\omega, T)}{S_1(\omega, T)} \tag{6.31}$$

For the materials and operating conditions of interest in this thesis, $\beta \approx 2$–3.

6.3.2 Multiplexing

A primary advantage of MKIDs is their ability to be easily multiplexed in the frequency domain. This multiplexing is a natural result of the narrow-band MKID response shown in Fig. 6.5. Fig. 6.6 shows a typical multiplexing scheme. Each MKID is tuned to resonate at a slightly different frequency than other resonators on the device and coupled via a weak coupling capacitor to a single microwave feedline. Since the transmission past each resonator is unity away from its resonant frequency, the signal from each resonator can be monitored without interference from the other MKIDs.

The resonator bandwidth limits the multiplexing factor that can be achieved. From Eq. 6.15, $\Delta f \propto Q^{-1}$, so higher-Q resonators can be packed more tightly in frequency. However, even for resonators separated by several line widths in frequency, significant capacitive or inductive coupling can produce coupled modes. For large arrays, resonator designs that reduce the capacitive coupling between resonators and frequency codings that widely separate nearest neighbor resonators in frequency space are needed to achieve high multiplexing factors with spatially localized resonances [305].

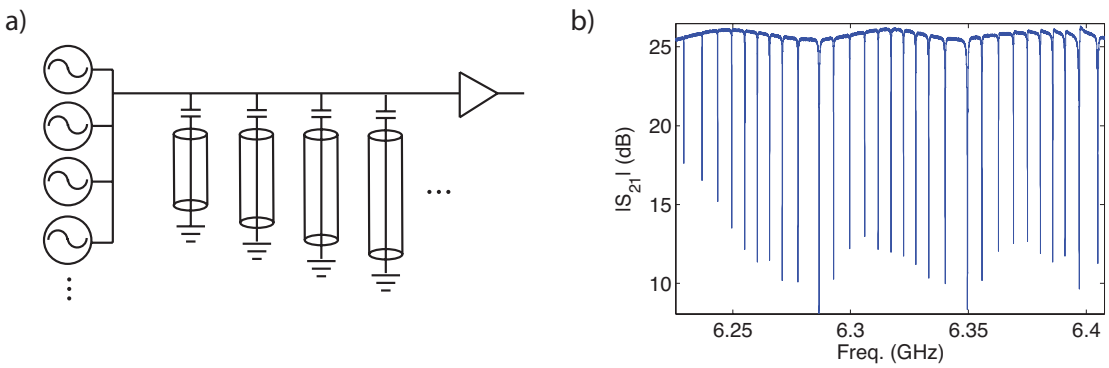


Figure 6.6: a) Schematic of multiplexing scheme for transmission line resonators. Each MKID is tuned to a slightly different resonant frequency and coupled to a single microwave feedline. A comb of frequencies corresponding to each MKID is input into the device and the transmitted signal is amplified by a single low-noise cryogenic amplifier. b) Measured transmission, $|S_{21}|$, versus frequency for an array of MKIDs. Each MKID appears as a narrow dip in the transmission near its resonant frequency.

6.4 MKID-based ZIP detector

Figure 6.7 shows the concept for an MKID-based ZIP detector. This design is based on the iZIP design discussed in Sec. 6.1, and consists of interleaved ionization and phonon sensors, with the microwave feedline acting as the ground electrode for the ionization measurement. In the MKID-based detector, the TES islands are replaced with ∼250 lumped-element microwave resonators, each with active inductor area ∼2 mm^2, giving comparable active metal coverage to existing TES-based designs (∼5% of both sides).

The following sections describe two different designs for the MKID-based athermal phonon sensors. The first design, described in Sec. 6.4.1, was based on previous experience with coplanar waveguide (CPW) resonator geometries from x-ray [297] and optical detectors [306], as well as the quasiparticle-trapping absorbers used in CDMS TES-based detectors. Due to difficulties in successfully transmitting quasiparticles from the absorber to MKID with this design, a simplified design that eliminated the absorbers was pursued. This design, in which the phonon energy is directly absorbed into lumped element MKIDs, is described in in Sec. 6.4.2.

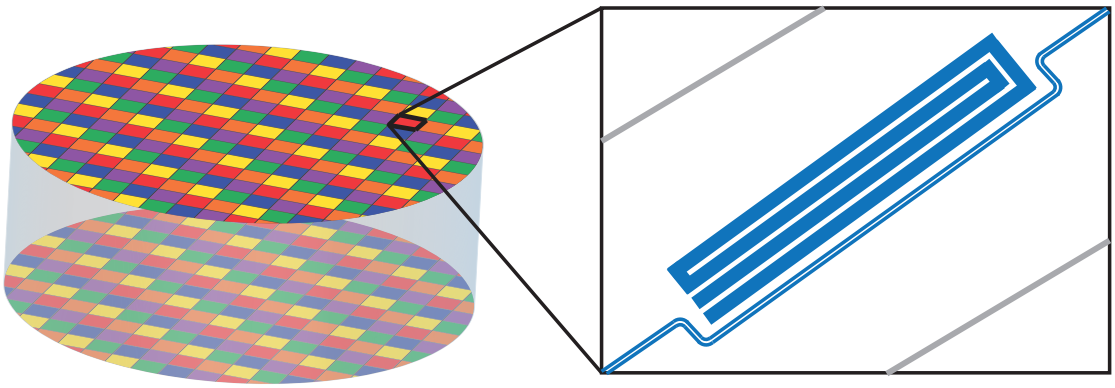


Figure 6.7: Concept for an MKID-based iZIP detector. Each face of the substrate is patterned with several hundred mm^2 MKIDs, giving a significantly more granular sensor than in existing TES-based designs. The TES islands in the standard iZIP are replaced with a 0.3 mm x 4 mm meandered MKID (blue). The microwave feedline (blue) provides the ground for the interleaved charge electrodes (gray).

6.4.1 Strip detectors

The initial concept for large-area MKID-based sensors suitable for use as athermal phonon collectors was proposed by S. Golwala [307] and is shown in Fig. 6.8. This design is similar to the existing CDMS TES-based design, in which large-area Al collectors are used to absorb the incident phonon energy, which is then concentrated in a small-volume resonator through quasiparticle trapping [246]. The absorbers are arranged along the ground planes of the resonator, which is made of a thinner, lower gap material such as titanium. This design allows coverage of the detector surface by the phonon absorber, with the energy then concentrated into small, sensitive detectors. Calculations [307, 308] indicated that energy resolutions of < 100 eV would be attainable with nearly full coverage of the detector surface.

The dimensions of the absorbers are determined by the typical length of a $\lambda/4$ CPW resonator ($\approx$5 mm) and the quasiparticle diffusion length, $l = \sqrt{D\tau}$, where D is the diffusion constant and τ is the quasiparticle lifetime in the absorber. To measure this diffusion length, a series of strip detectors [309,310] were designed and fabricated. The strip detector geometry is shown in Fig. 6.9a. Interactions occurring in the central strip break Cooper pairs, creating quasiparticles which can diffuse and become trapped in the lower-gap MKIDs on either end. From the relative energy collected in each resonator, the location of the interaction can be determined. The total energy collected as a function of position, $E_{coll}(x)$,

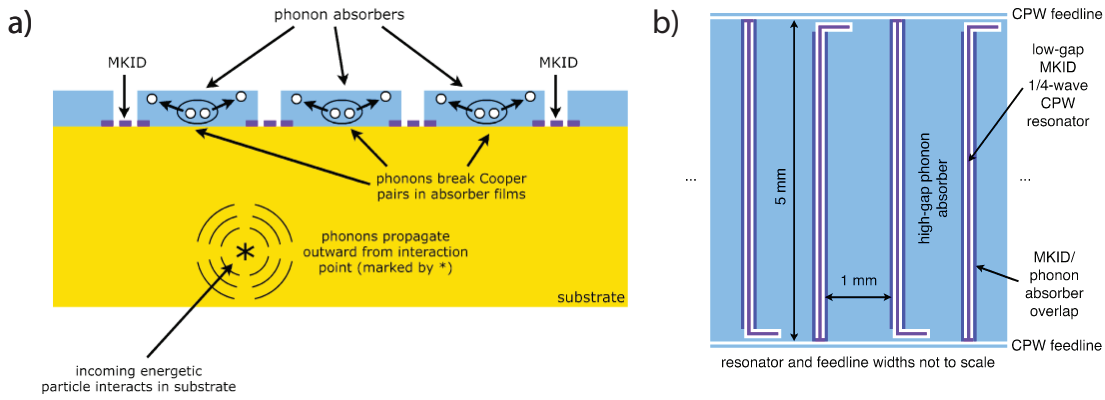


Figure 6.8: a) Cross section of absorber-based detector concept. Thick phonon absorbers (light blue) are coupled to small volume CPW resonators (purple). Particle interactions in the substrate create phonons which propagate to the absorbers and break Cooper pairs, creating quasiparticles. These quasiparticles can diffuse and be trapped in the lower gap MKID, concentrating the incident energy into a small volume. b) View of absorbers and resonators from above. The typical absorber size is 1 mm x 5 mm. *Figure adapted from Golwala et al. [307]*

determines the diffusion length, l [309]:

$$E_{coll}(x) = E\frac{\sinh(\beta x/L)}{\sinh(\beta)} \tag{6.32}$$

where E is the initial deposited energy, x is the position of the interaction, L is the length of the strip and $\beta = L/l$.

The first strip detector devices that were tested used aluminum MKIDs with tantalum absorbers since these materials were previously used to fabricate x-ray strip detectors [297]. For an MKID-based phonon detector consisting of hundreds of MKIDs, each resonator would collect only tens of eV of energy, requiring an increase in the responsivity of the devices by 2–3 orders of magnitude relative to the x-ray strip detectors. This increased sensitivity was achieved by reducing the MKID thickness from 200 nm to 20 nm and reducing the CPW center strip width from 3 μm to 1 μm. Decreasing the resonator size decreased the volume by a factor of 30 and increased the kinetic inductance fraction, α, from 0.03 to 0.75. Since the phase response scales as $\alpha Q/V$, these changes provided a factor of $\sim$750 increase in responsivity.

Fig. 6.9b shows the relative pulse height for coincident pulses in two 25 nm thick Al resonators attached to a $L = 100$ μm long, 100 nm thick Ta absorber. The strip was

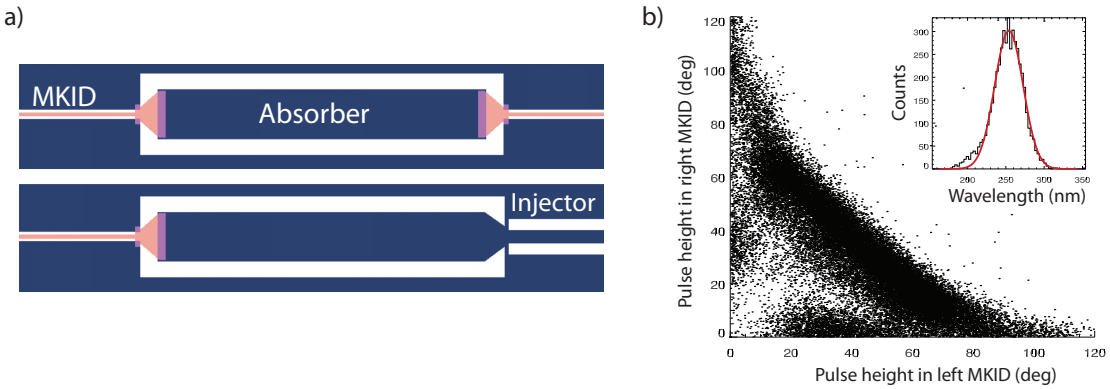


Figure 6.9: a) Schematic of the strip detector geometry. The top panel shows the layout for a standard strip detector, where an absorber (dark blue) is connected to lower-gap MKIDs (light red) on either end. In the bottom panel, a quasiparticle injector is formed by replacing the lower-gap MKID on the right with an MKID made of the same material as the absorber. b) Pulse heights for 20 nm thick Al MKIDs attached to a 100 μm long Ta strip, illuminated by 254 nm photons. The inset shows the energy spectrum for events with greater than 20 degrees of phase shift in both MKIDs. *Figure from Moore et al., [306]*

illuminated with 254 nm photons. Fitting the observed pulses as a function of position along the strip to a 2-dimensional diffusion model with recombination time, τ, gave a Ta diffusion constant, $D = 14 \pm 2$ cm^2/s, and quasiparticle lifetime, $\tau = 35 \pm 6$ μs, implying $l = 216 \pm 30$ μm. These results are consistent with the diffusion length inferred from fits of the data to Eq. 6.32.

Although Fig. 6.9b shows results from a working strip detector with sub-eV resolution, these results were not reproducible. The devices made afterwards showed high signal-to-noise pulses when illuminated by 254 nm photons, but the pulses in resonators connected to the same strip were uncorrelated in time, indicating that they arose from direct hits in the MKIDs rather than quasiparticle diffusion and trapping from the strip. Nearly 30 subsequent devices were tested, both using the same design and fabrication procedures as the working device, as well as others with a variety of changes to the trapping geometry, cleaning of the MKID-absorber interface, and order of deposition of the MKID and absorber. A similar lack of trapping was found for devices with Al absorbers and lower-gap Ti or AlMn MKIDs. Although x-ray detectors with MKID thicknesses $\gtrsim$200 nm were reliably fabricated using the same resonator geometry and fabrication procedures, no corresponding method was found for reliably producing the thinner detectors needed for an athermal-phonon absorber.

To study the transition between the working x-ray detectors with MKID thicknesses $\gtrsim$200 nm and the 20 nm thick UV photon detectors that did not work reliably, a new geometry was developed to allow direct injection of quasiparticles into the absorber using strong microwave pulses. This eliminated the need for an external photon source with the desired energy. These "quasiparticle injectors" are shown in Fig. 6.9a, where the lower gap MKID on the right side of the strip is replaced with an MKID made of the same material as the absorber strip. Applying a strong microwave pulse at the resonant frequency of the injector creates quasiparticles at the shorted end, which can then diffuse across the absorber and be detected by the MKID on the left side of the strip. While the detailed mechanism for generating quasiparticles using a pump tone well below the gap is likely complex, it is not necessary to understand the quasiparticle generation process in detail for the tests reported below.

Using these injector devices, previous results from Ta/Al strip detectors illuminated by x-ray and UV photons were confirmed [306]. A quasiparticle response was seen in a device with 320 nm thick Al MKIDs when driving the Ta injector with a strong microwave pulse. In contrast, no quasiparticle response was observed when driving thinner injector devices with 25 nm MKIDs.

Since the MKID/absorber interface appeared to prevent quasiparticles from reaching the MKID, single-layer devices that eliminated this interface were designed. These devices consisted of single 60 nm thick Al film patterned into the standard strip detector geometry. To allow quasiparticle trapping in the MKID, the triangular trapping region was selectively implanted with 4.8×10^{14} Mn ions/cm^2 at 20 keV using a photoresist implant mask. On-chip test devices implanted with the same Mn dose indicated that $T_c \approx 500$ mK in the implanted regions. Although the Mn implantation was found to reduce the resonator internal Qs to $\sim2\times10^4$, annealing the devices in vacuum at 300 °C for 60 minutes increased the the Q_i to $\gtrsim4\times10^5$. After annealing, the T_c was also increased to 720 mK, which gave good agreement with cosputtered AlMn devices with the same Mn concentration [311].

Quasiparticle transmission was seen in these AlMn/Al device using the same geometry as Al/Ta devices, consistent with the hypothesis that removing the MKID/absorber interface would improve transmission. Fig. 6.10 shows the response in the left MKID when driving the injector with a 5 μs microwave pulse. The pulse response was fit to a diffusion model to determine the aluminum diffusion constant, $D_{Al} = 20 \pm 5$ cm^2/s. The quasiparticle

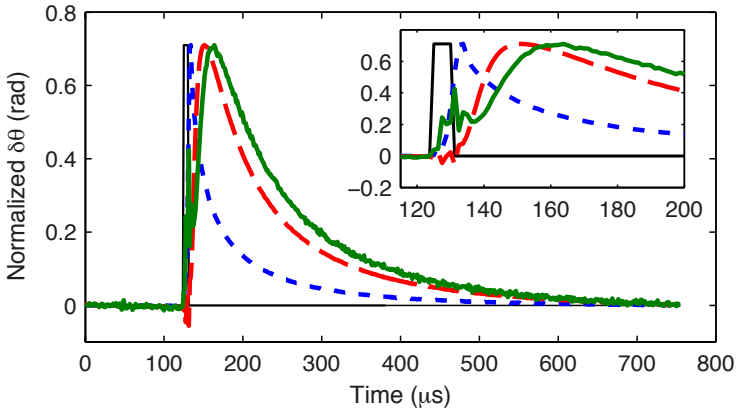


Figure 6.10: Normalized MKID response when a 5 μs long square wave pulse (black) is applied to the corresponding quasiparticle injector for a 200 μm strip (red, dashed) and 400 μm strip (green, solid). The blue, dotted curve shows the response when the MKID is driven directly. An enlarged version of the rising edges of the pulses is shown in the inset. The delayed MKID response allows the quasiparticle diffusion constant in the absorber to be measured. All pulses are normalized to the pulse height for the 200 μm strip. The transient oscillations visible immediately after the microwave power is switched on are due to the readout electronics and not the response of the resonators. *Figure from Moore et al., [306]*

lifetime in AlMn was measured from the pulse fall times to be $\tau_{AlMn} = 130 \pm 4$ μs. This lifetime is consistent with measurements of the quasiparticle lifetime in unimplanted Al films, so we take $\tau_{Al} \approx \tau_{AlMn}$ and calculate a diffusion length $l_{Al} = \sqrt{D\tau} = 508 \pm 69$ μm for these devices. This diffusion length is sufficiently large to allow 0.5 mm wide Al absorbing fins without significant quasiparticle recombination in the absorber. However, the trapping probability for this device was low due to the relatively small difference in gap between the Al and implanted AlMn. Optimized devices with high trapping probability would likely require higher implanted Mn doses to sufficiently suppress the gap after annealing.

Although the implanted AlMn/Al devices appeared to solve the quasiparticle trapping problems which plagued previous designs, the added implantation and annealing steps increased the fabrication complexity of the design. Since reducing fabrication time and cost is a key driver for use of MKIDs in next-generation experiments, these implanted designs were not ideal. In the following section we present a design where the phonon energy is directly absorbed in the MKID. This design eliminates the need for quasiparticle trapping and can be patterned from a single deposited film, allowing for faster and more reliable detector fabrication.

6.4.2 Direct absorption in large-area MKIDs

As an alternative to the strip-detector architecture considered in Sec. 6.4.1, designs allowing for direct absorption of the quasiparticle energy in the MKID were considered. A similar design developed independently by Swenson et al. [312] has been used to demonstrate time-resolved phonon-mediated detection of high-energy interactions from cosmic rays and natural radioactivity using MKIDs. Although these designs do not benefit from the concentration of quasiparticles into a small volume resonator, in this section we show that the ∼100 eV energy resolutions needed for a CDMS-style iZIP detector can still be obtained.

We calculate the sensitivity for the direct-absorption design following the original calculation by S. Golwala [303]. The sensitivity of an individual resonator was given in Eqs. 6.29 and 6.30. As will be demonstrated Chap. 7, the frequency noise is amplifier limited for the resonator geometries and signal frequencies of interest in this thesis, so we assume amplifier-limited frequency readout for this calculation.

For an interaction in the detector that deposits energy E, the number of quasiparticles created in each resonator is:

$$N_{qp} = \eta_{ph}\frac{E}{N_r\Delta} \quad \Rightarrow \quad \delta E = \frac{N_r\Delta}{\eta_{ph}}\delta N_{qp} \tag{6.33}$$

where we assume that the phonons uniformly illuminate the N_r resonators on the detector surface, and η_{ph} is the overall efficiency for conversion of the incident energy into quasiparticles. Over the full detector, we have N_r measurements of the energy so the overall detector resolution is:

$$\sigma_E = \frac{1}{\sqrt{N_r}}\delta E = \frac{2\Delta}{\eta_{ph}\beta(\omega,T)}\sqrt{\frac{\eta_{read}k_bT_NN_0VN_r}{\chi_c\chi_{qp}S_1(\omega,T)\alpha Q_{i,qp}\chi_{BW}}} \tag{6.34}$$

The number of resonators, N_r, can be expressed in terms of the area of a given resonator, A_r, the surface area of the substrate, A_{sub}, and the total fraction of the surface area which is covered by resonators, η_{fill}, as $N_r = \eta_{fill}A_{sub}/A_r = \eta_{fill}A_{sub}t_r/V$, for a resonator of thickness, t_r.

Resonator bandwidth requirements determine χ_c. The time constant with which the

resonator can respond, τ_r, depends on the strength of the coupling to the feedline:

$$\tau_r = \frac{Q}{\pi f_0} \tag{6.35}$$

Ideally, the MKID would provide enough bandwidth to resolve the rising edge of the phonon signal. For sufficiently large substrates that the time to propagate from the interaction site to the detector surfaces is much longer than the phonon-downconversion time, the pulse rise time is given by $\tau_{rise,ph} = t_{sub}/c_s$, where t_{sub} is the substrate thickness and c_s is the speed of sound in the crystal. For Ge, $c_s \approx 5$ mm/μs, so characteristic rise times are $\tau_{rise,ph} \approx 5$ μs for a 2.5 cm thick detector. To resolve this rise time requires:

$$Q \leq \frac{\pi f_0 t_{sub}}{c_s} \tag{6.36}$$

which gives $Q \leq 5{\times}10^4$ for $f_0 = 3$ GHz and $t_s = 2.5$ cm. For the devices considered in Chap. 7, we typically have $Q_i > 10^6$, so we are in the overcoupled limit $Q_c \ll Q_i$ and $\chi_c \approx 4Q/Q_i = 4\pi f_0 t_{sub}/c_s Q_i$. Using the above relations, we can rewrite Eq. 6.34 as:

$$\sigma_E = \frac{\Delta}{\eta_{ph}\beta(\omega,T)}\sqrt{\frac{\eta_{read}\eta_{fill}}{\pi\alpha\chi_{BW}S_1(\omega,T)}\frac{N_0 k_b T_N}{f_0}\frac{A_{sub}t_r c_s}{t_{sub}}} \tag{6.37}$$

While the rise time is set by the characteristic propagation time before the first bounce, the decay of the phonon signal additionally depends on the filling fraction and phonon absorption probability per bounce, p_{abs}, as $\tau_{ph} = (t_{sub}/c_s)(1/\eta_{fill}p_{abs})$. This assumes that there are no sources of phonon energy loss other than creation of quasiparticles in the resonators. For iZIP detectors, $\tau_{ph} \approx 900$ μs, while quasiparticle lifetimes for the films considered in this work are typically $\tau_{qp} \approx 50$–100 μs. Given these time constants, the limit $\tau_r \approx \tau_{rise,ph} \ll \tau_{qp} \ll \tau_{ph}$ applies, and Eq. 6.23 can be simplified to give $\chi_{BW} \approx \tau_{qp}/\tau_{ph}$.

The absorption probability per bounce depends on the probability of transmission from the substrate into the MKID, p_t, and the characteristic pair-breaking length in the MKID, λ_{pb}. Since the average distance travelled in the film per bounce is $>2t_r$, then the absorption probability must be at least $p_{abs} = 2p_t t_r/\lambda_{pb}$. Using this parametrization, we can write

$\chi_{BW} = \tau_{qp}/\tau_{ph} = 2\eta_{fill}\, p_t\, t_r\, c_s\, \tau_{qp}/t_{sub}\, \lambda_{pb}$. Inserting into Eq. 6.37 gives:

$$\sigma_E = \frac{\Delta}{\eta_{ph}\beta(\omega,T)}\sqrt{\frac{\eta_{read}}{\alpha p_t}\frac{A_{sub}k_bT_N}{2\pi f_0}\frac{N_0\lambda_{pb}}{\tau_{qp}S_1(\omega,T)}} \tag{6.38}$$

It is somewhat surprising that this final result depends on only a limited range of parameters, but this results from the simplifying assumptions made in the above calculation. In particular, this result holds only when:

$$\begin{aligned} &\tau_r \approx \tau_{rise} \ll \tau_{qp} \ll \tau_{ph} \\ &Q_c \ll Q_i \\ &t_r \ll \lambda_{pb} \\ &n_{qp}^{thermal} \ll n_{qp}^{read} \end{aligned} \tag{6.39}$$

We can control α, Δ, τ_{qp} and f_0 through our choice of materials and resonator geometry. The expected sensitivity would also be improved if lower noise amplifiers become available. We have less control over the remaining parameters, which have less significant variations with material.

As an example, we can numerically evaluate Eq. 6.38 for two possible resonator materials, Al or TiN. The choice of resonator materials will be discussed further in Sec. 6.5.3. For Al, the single spin density of states is $N_0 = 1.72{\times}10^{10}$ μm^{-3} eV^{-1} [313]. For TiN, we use $N_0 = 8.7{\times}10^{9}$ μm^{-3} eV^{-1} [301]. The sensitivity for Al is then:

$$\sigma_E = (245\text{ eV})\left(\frac{0.3}{\eta_{ph}}\frac{1.9}{\beta(\omega,T)}\frac{\Delta}{200\ \mu\text{eV}}\right)\times \sqrt{\frac{\eta_{read}}{1}\frac{0.1}{\alpha}\frac{0.5}{p_t}\frac{A_{sub}}{100\text{ cm}^2}\frac{T_N}{5\text{ K}}\frac{3\text{ GHz}}{f_0}\frac{\lambda_{pb}}{1\ \mu\text{m}}\frac{100\ \mu\text{s}}{\tau_{qp}}\frac{1.6}{S_1(\omega,T)}} \tag{6.40}$$

while TiN (tuned for $T_c \approx 0.75$ K) gives:

$$\sigma_E = (38 \text{ eV}) \left(\frac{0.3}{\eta_{ph}} \frac{2.1}{\beta(\omega,T)} \frac{\Delta}{115\ \mu\text{eV}} \right) \times \sqrt{\frac{\eta_{read}}{1} \frac{0.75}{\alpha} \frac{0.5}{p_t} \frac{A_{sub}}{100\ \text{cm}^2} \frac{T_N}{5\ \text{K}} \frac{3\ \text{GHz}}{f_0} \frac{\lambda_{pb}}{1\ \ \mu\text{m}} \frac{100\ \mu\text{s}}{\tau_{qp}} \frac{1.2}{S_1(\omega,T)}} \tag{6.41}$$

The assumed values for η_{ph}, α, and τ_{qp} are based on measurements of the prototype devices in Chap. 7. Even with the conservative values assumed, the expected baseline resolution for a massive ($A_{sub} = 100$ cm^2) detector is comparable to the average CDMS II baseline resolution of ∼250 eV. Using $T_c \approx 0.75$ K TiN could improve this resolution to $\lesssim 40$ eV due to its higher kinetic inductance and lower gap. Lower noise amplifiers, such as the broadband NbTiN parametric amplifier currently being developed by Eom and collaborators at JPL [314] could improve the resolution by as much as a factor of 6, if they are able to reach the quantum limit of $T_N = 0.14$ K at 3 GHz and no additional phase noise above amplifier noise is present. Finally, we note that this calculation assumes $Q_c \ll Q_i$ to be able to resolve the μs rising edge of the pulse. If sufficient position resolution can be obtained from the amplitude of the signal in each MKID without need for the detailed timing information, then the resolution could be improved by raising Q_c as $\sigma_E \propto 1/\sqrt{Q_c}$.

6.5 Design of 20-channel prototype devices

As a first step towards a full-sized MKID-based CDMS-style detector, we designed smaller prototype devices which could be easily fabricated at the Micro Devices Laboratory at JPL. The aim of these devices was to provide the first demonstration of energy- and position-resolved phonon-mediated detection of particle interactions using MKIDs. Fabrication constraints limited the substrate area to 22 mm x 20 mm, the full area of a single stepper field. Similarly, 1 mm thick substrates were the largest that could easily be aligned using the stepper. For the substrate material, high-resistivity (> 5kΩ cm) Si was used due to the experience fabricating MKIDs on this material at JPL. The resonator design, materials, and layout are described in the following sections, with experimental results presented in Chap. 7.

6.5.1 Large-area resonators

The area coverage required for a single resonator is set by the desired total active metal coverage and the number of channels that can be read out on a single feedline. We used the surface coverage fraction of the current iZIP detectors as a baseline for our prototype devices. For the iZIPs, approximately 5% of the surface area on both faces is covered with Al absorbing fins. For the 22 mm x 20 mm prototype devices, $\sim$5% area coverage was attainable using 20 MKIDs, each with an inductor area of $\sim$1 mm^2.

Our initial resonator design employed the simple LEKID geometry first proposed by S. Doyle et al. [315]. This design has no ground plane, which is a requirement for incorporating MKIDs into an iZIP style design where DC electric fields are maintained along the surface of the crystal. An example of this resonator geometry is shown in Fig. 6.11a. The resonator consists of a meandered inductor and an interdigitated capacitor, coupled to a coplanar strip (CPS) feedline. Simulations using SONNET [316] were performed to determine an inductor and capacitor geometry with the desired area and a resonant frequency $\gtrsim$3 GHz, limited by the lower frequency cutoff of our HEMT amplifier. The resonant frequency depends both on the resonator geometry and the kinetic inductance. For design purposes, we assumed a surface inductance of $L_s = 20$ pH/sq, based on measurements of TiN films with thicknesses $\sim$100 nm. For future detectors, moving to a lower frequency cryogenic amplifier will enable each resonator to cover a larger active area.

The resulting resonator design consisted of a 70 μm wide meandered inductor and a 4-fingered interdigitated capacitor with 20 μm gaps between fingers. The total inductor area at 3 GHz is $A_{ind} = 1.07$ mm^2, with a capacitor area of $A_{cap} = 0.26$ mm^2. The simulated current distribution from SONNET for this resonator design is also shown in Fig. 6.11. The effective inductor area is slightly reduced from the nominal value due to current variations across the inductor. Since the current peaks at the center of the inductive meander, energy deposited near the center will produce a somewhat larger resonator response than if the equivalent energy were deposited at the edge of the inductor. From the simulated current distribution, one can calculate the effective active area, $A_{eff} = \eta A_{geom}$, i.e., the area of a resonator with completely uniform current density across the inductor which would have

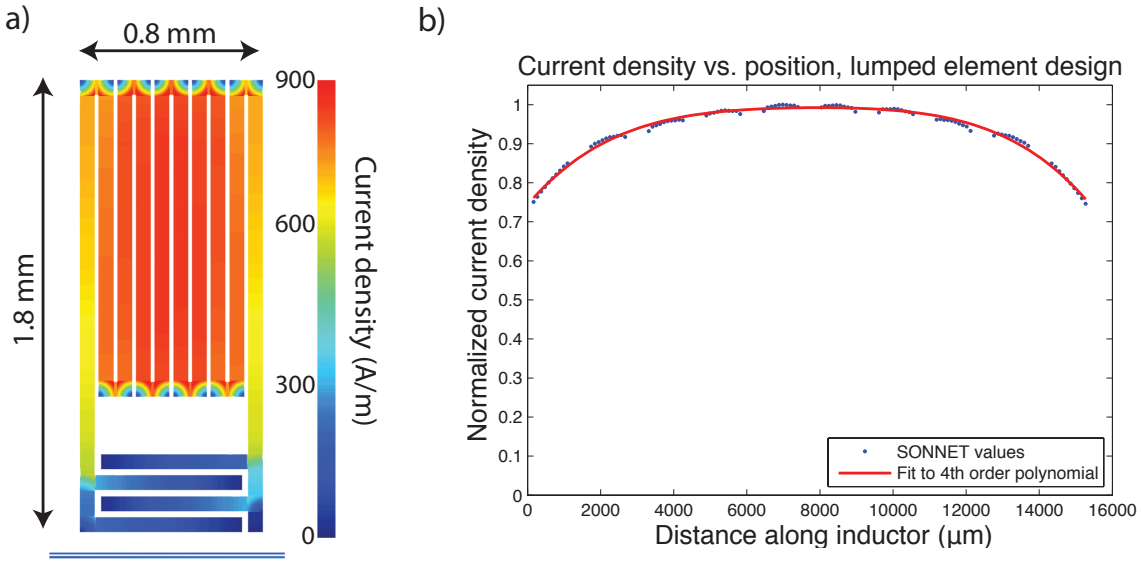


Figure 6.11: a) Large-area MKID geometry used for the prototype arrays. The resonator consists of a meandered inductor connected to an interdigitated capacitor in a standard LEKID design. The inductor width is 70 μm, with total area $\sim$1 mm^2. The color scale indicates a SONNET simulation of the current density distribution near the resonant frequency. b) Normalized current as a function of position along the inductive meander, starting from the bottom left corner of the inductive section of the resonator. The red curve shows a polynomial fit indicating that the effective inductor area due to decreased current density at the edges is $\sim$90% of the total geometrical area.

the same phase response when uniformly illuminated with phonons. Here, η is defined as:

$$\eta = \int \left[\frac{j(z)}{j_{max}}\right]^2 dz \tag{6.42}$$

where $j(z)$ is the current density along the resonator, and j_{max} is the peak value. SONNET simulations indicate that $\eta = 0.89$ for this geometry, indicating that we have fairly good uniformity across the devices. A tapered resonator design that reduces current density nonuniformities will be discussed in Sec. 6.5.2.

The resonator quality factor, and accordingly its bandwidth, is controlled by its coupling to the microwave feedline. This coupling is determined by the feedline geometry and the distance of the resonator from the feedline. As discussed in Sec. 6.4.2, we require $Q \lesssim 5{\times}10^4$ to give the $\sim\mu$s time resolution needed to resolve the rising edge of the phonon pulses. SONNET simulations were performed to determine the expected coupling quality factor, Q_c, for the resonator geometry from Fig. 6.11 as a function of distance from the feedline.

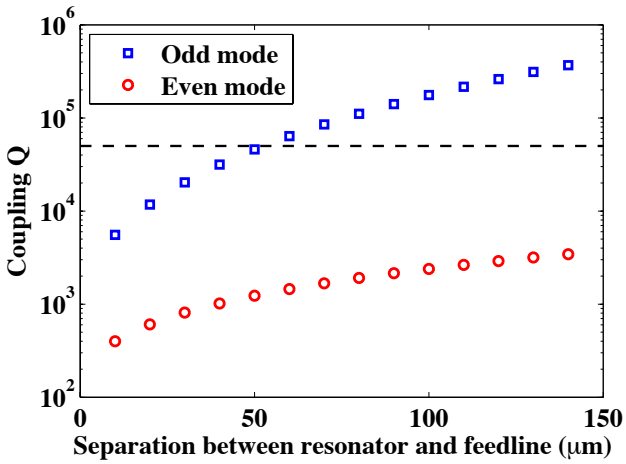


Figure 6.12: Simulated coupling Q versus separation between resonator and feedline from SONNET. The even CPS mode shows significantly higher coupling (and 1–2 orders of magnitude lower coupling Q) than the odd CPS mode for the same separation. The dotted line shows the design Q, corresponding to a separation of 55 μm from the feedline.

For our prototype devices, we chose $Q_c = 5 \times 10^4$, corresponding to a separation of 55 μm between the resonator and feedline for coupling to the odd CPS mode, as shown in Fig. 6.12.

6.5.2 Array layout

Resonators of the geometry described above were arranged in a 20-element array, as shown in Fig. 6.13a. Each resonator was tuned to a different resonant frequency by varying the length of the inductive section of the resonator, while leaving the capacitive section and coupler unchanged. For nominal values of the surface inductance (L_s = 20 pH/sq), the design frequencies were evenly spaced by 10 MHz from 3–3.2 GHz. To minimize coupling between resonators, the frequency layout was optimized so that resonators which were physically close were far apart in frequency space. Using the coding shown in Fig. 6.13a, nearest neighbor resonators are at least 30 MHz apart in frequency.

As described in the previous section, a CPS feedline with 12 μm wide conductors and a 6 μm gap (i.e., a 12–6 geometry) was used to excite the resonators. The characteristic impedance of this feedline was 210 Ω at L_s = 20 pH/sq, which required an impedance transformer to give a good match with the 50 Ω characteristic impedance of the input and output coaxial cables. A single-pole on-chip LC transformer was designed, as shown in Fig. 6.13b. SONNET simulations indicated that this transformer gave negligible return loss

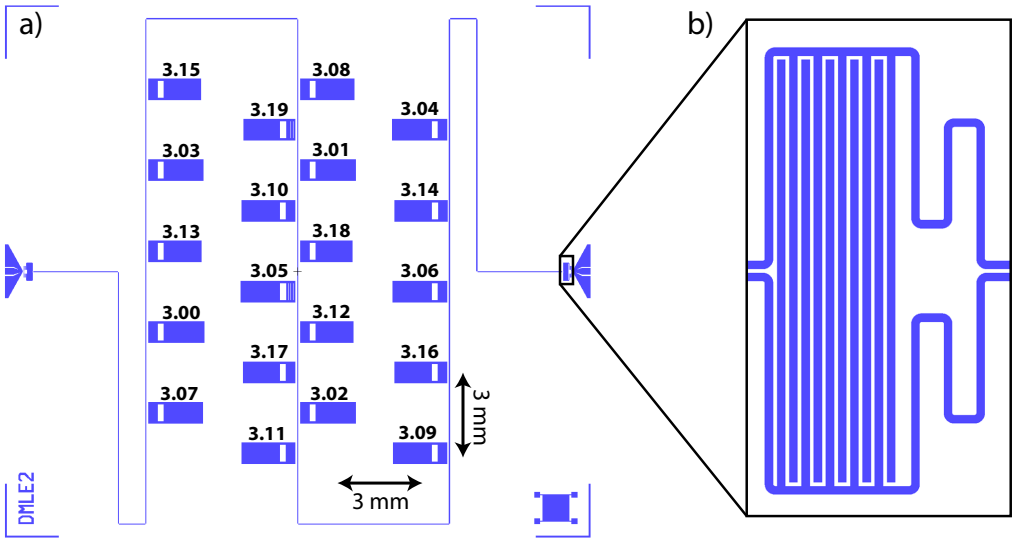


Figure 6.13: a) Schematic of array layout. Twenty resonators are arranged in a grid and excited by a meandered CPS feedline. The center-to-center row and column spacing between resonators is 3 mm. The numbers just above each resonator indicate the design frequency in GHz. b) Zoom showing the on-chip LC transformer used to match the high-impedance CPS line to the 50 Ω CPW transition board. The transformer consists of a meandered inductive section in each line followed by a shunt capacitor.

in a ∼500 MHz bandwidth around 3.1 GHz.

Tests of this array design found that the coupling Qs varied from $\sim 10^3$–10^5, despite the design of $Q_c = 5 \times 10^4$ for all resonators. Two possible sources for this variation were identified:

1. As shown in Fig. 6.12, the coupling to the even CPS mode is significantly higher than the odd mode, for which the resonators were designed. If the resonators are able to couple some fraction of their energy into this mode, the measured coupling Q will be lower than the design value.

2. Despite the frequency coding to separate physically close resonators in frequency space, SONNET simulations indicated that residual coupling existed between certain diagonal pairs of resonators, since there are no canceling currents along the short side of the resonator. These coupled modes will have different Q_c than for isolated resonators.

To mitigate these problems, several changes to the array design were tested in SONNET

and are described below.

Reducing the coupling of the resonator to the even mode is most easily accomplished by lowering the impedance of the odd mode relative to the even mode. For the even mode, the feedline conductors and the box lid form what is essentially a vacuum microstrip, with the lid serving as the ground plane. The characteristic impedance of this microstrip mode, $Z_0 \propto d/w$, where d is the gap between the feedline and box lid, and w is the width of the feedline. SONNET simulations indicated that for the 12-6 CPS feedline with $L_s = 20$ pH/sq and device box geometry described above, the impedance of the even mode was 194 Ω, comparable to the impedance for the odd mode.

For high-resistivity superconductors, where $L_s \gg 1$ pH/sq, it is difficult to reduce the impedance of the CPS line below 100 Ω unless sub-micron gaps are used between the conductors. Given fabrication tolerances, such small gaps are problematic, and a more viable solution is to use a low surface inductance material for the feedline. For example, for 25 nm Al, with $L_s = 0.8$ pH/sq, a 30–2 CPS line has a characteristic impedance of 50 Ω. By adding processing layers, more complicated feedlines or different materials are also possible, e.g., finite ground plane CPW lines (FGCPW). However, FGCPWs are also mode rich, and in addition to the even (microstrip) mode, they contain a coupled slotline mode which must be suppressed by tying together the ground conductors with air-bridges. We expect future devices with feedlines that have characteristic impedance of 50 Ω for the odd mode, while maintaining > 200 Ω impedance for the even mode, to better match the coupling Q expectations shown in Fig. 6.12.

New resonator geometries that reduce inter-resonator coupling were also tested in SONNET. Fig. 6.14 shows a mm^2 resonator design where the inductive section is meandered in a spiral pattern, following similar designs from Noroozian et al. [305]. In this design, currents in the inductive section are canceled by opposite currents flowing in a closely spaced, parallel meander. Simulations verified that this spiral design reduced coupling between resonators to a negligible level, assuming the same frequency spacing for nearest neighbor resonators as in the 20-element array.

In addition, these resonator designs taper the width of the inductive section to maintain a constant current density along the inductor. Simulations indicate that for the tapered design, the current is uniform to within 5% across the entire inductive meander, which maximizes the effective active area of the design. Although the improvements to the feedline

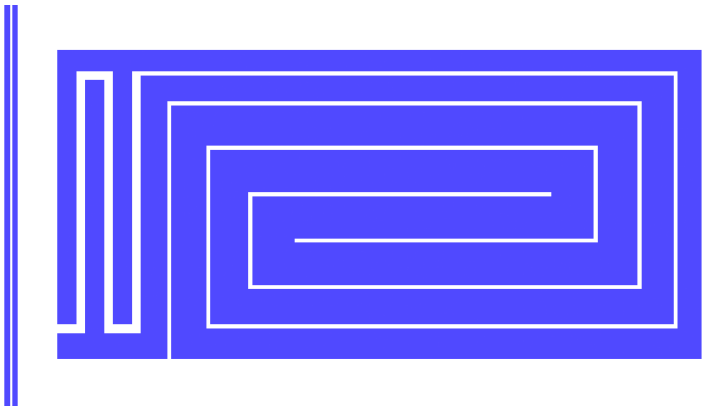


Figure 6.14: Spiral resonator design with reduced coupling to other resonators. At large distances, the current in each inductive meander is canceled by the corresponding opposite current in the adjacent meander. The line width is also tapered to maintain approximately uniform current density throughout the inductive section.

and resonator geometry described above were not implemented for the results shown in Chap. 7, future devices will incorporate these changes.

6.5.3 Resonator materials

To be able to efficiently collect athermal phonons, the arrays must be fabricated from a superconductor with a gap that is small enough that ballistic phonons can break Cooper pairs (i.e., $2\Delta \ll 1$ THz). Fig. 6.15 shows a simulation of the phonon downconversion process in Si, in the case that anharmonic decay of phonons dominates the downconversion process. To fully collect the phonon energy requires a superconductor with $T_c \lesssim 2$ K from which high Q_i resonators can be fabricated. In addition, $L_s = \hbar R_n/(\pi\Delta)$ [299], so maximizing the kinetic inductance fraction requires low-gap materials with high normal-state surface resistance, R_n.

Nitride materials, such as TiN_x or $NbTiN_x$, provide good candidates for a sensitive resonator material since they have high normal-state resistivity, $\rho_n \approx 100 - 200$ $\mu\Omega$ cm, and have been shown to make extremely high quality resonators, with $Q_i > 10^7$ in some cases [302]. In addition, their T_c is tunable from ~0.5 K to > 4 K by varying the nitrogen content of the films. However, for the substoichiometric films with T_c<2 K, needed to collect ballistic phonons, the film parameters are highly sensitive to deposition conditions, and large variations in gap and resonator properties between depositions and across a single

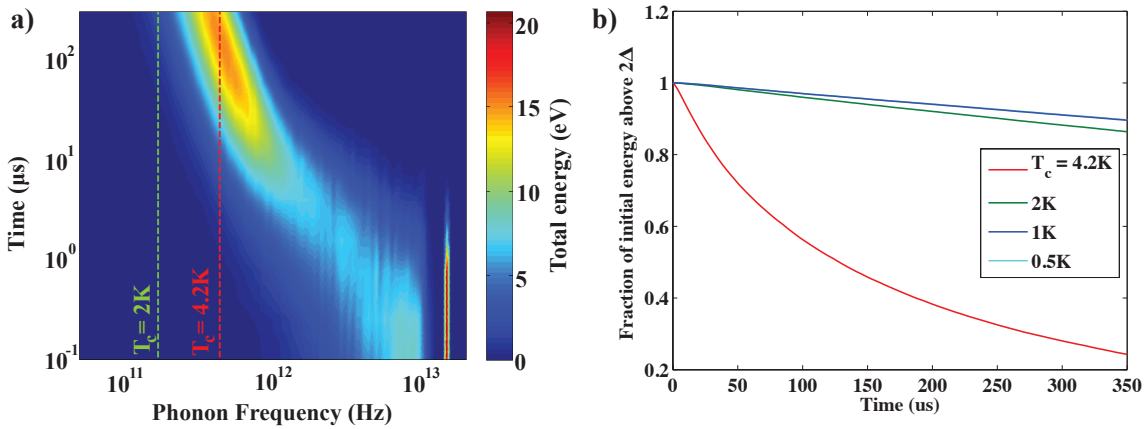


Figure 6.15: a) Simulation of the distribution of phonon energy versus time in a Si substrate following a particle interaction, using the CDMS detector Monte Carlo [317]. Phonons are initially created at the Debye frequency (∼10 THz) and rapidly undergo anharmonic decay until their mean free path becomes comparable to the size of the substrate. The simulation assumes no energy loss mechanisms other than anharmonic decay. b) Fraction of phonons which are energetic enough to break Cooper pairs in a superconductor of the given T_c. For $T_c < 2$ K, nearly all phonons remain above 2Δ on the time scales needed to collect the energy in absorbing films.

film are observed.

Although Al has lower normal-state resistivity than the nitrides, thin Al films have a $T_c \sim 1.2$–1.3 K, and it is straightforward to make resonators with $Q_i > 10^6$ (e.g., [318]). No tuning of film composition is required, so Al is less sensitive to deposition conditions and shows higher uniformity than the nitrides. Al is also known to be a good athermal phonon collector from experience with Al absorbers in TES-based CDMS detectors.

Other materials with T_c from 0.5–1 K, such as Ti and AlMn were also investigated as possible MKID materials. Although pure Ti was found to produce extremely low quality resonators ($Q_i < 10^4$), resonators fabricated from AlMn films with T_c as low as 500 mK showed internal quality factors and quasiparticle lifetimes comparable to Al. Devices patterned from cosputtered AlMn films [311] with $T_c < 1$ K should have increased sensitivity relative to Al devices since $\sigma_E \propto \Delta$, as shown in Eq. 6.38. Future work will determine if AlMn films of the desired uniformity and reproducibility can be produced.

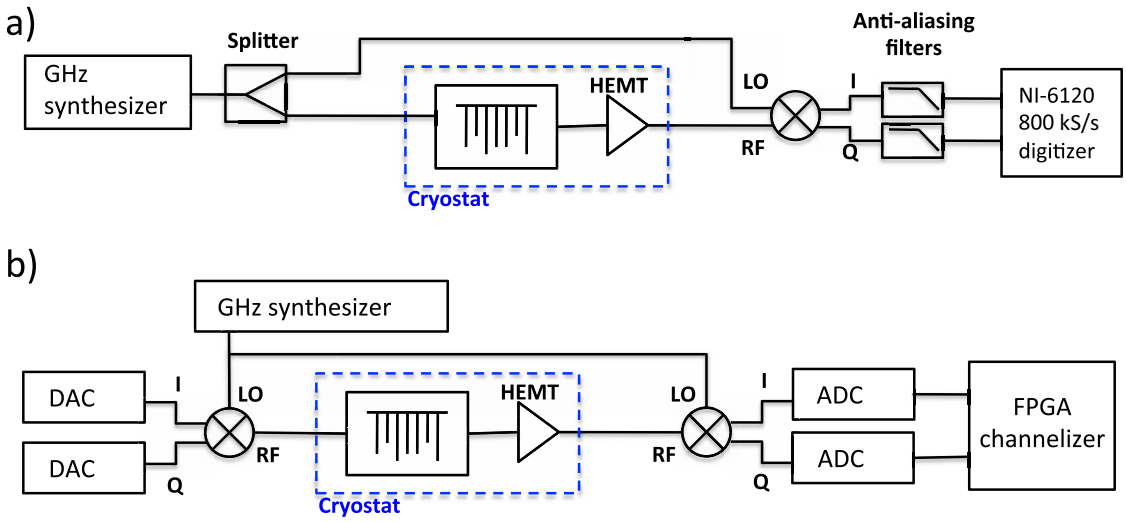


Figure 6.16: a) Schematic of a single tone analog readout. The GHz synthesizer and mixers are typically implemented using standalone analog components, making the cost of scaling this scheme to a large number of resonators prohibitive. b) Schematic of a multitone readout. With this setup, a single GHz synthesizer and pair of mixers can be used to read out hundreds of tones within ∼500 MHz of bandwidth.

6.6 Multichannel readout

The standard homodyne mixing technique for reading out a single resonator is shown in Fig. 6.16a [318]. A tone at the resonance frequency is generated by a microwave synthesizer, split into two copies, one of which is sent through the cryostat. This tone excites the MKID with the corresponding resonant frequency, and the transmitted signal is amplified by a low-noise HEMT amplifier. After exiting the cryostat, the signal transmitted past the resonator is mixed with the original copy generated by the synthesizer using an IQ mixer. The phase and amplitude of the signal transmitted past the resonator can then be determined by proper combination of the I and Q outputs of the mixer. Any change in the transmission past the resonator due to, e.g., quasiparticle creation, will cause a shift in the phase and amplitude of this output signal.

For small numbers of resonators, the readout is typically implemented using analog synthesizers and mixers, with a separate set of components needed for each resonator. Although conceptually simple, this system is not easily scaled to large numbers of resonators due to the high cost per channel required by standalone analog components.

A more scalable solution is to take advantage of the recent advances in digital electronics

to generate the frequency comb needed to excite an array of resonators and perform the channelization and mixing digitally. In this scheme, shown in Fig. 6.16b, fast digital-to-analog converters (DACs) generate a comb of frequencies at baseband, usually by playing back a predetermined lookup table (LUT). For currently available DACs, the baseband frequencies are limited by the sampling rate to $f_{DAC} \lesssim 1000$ MHz. This frequency comb is then mixed up to the RF band corresponding to the resonant frequencies (typically 1–10 GHz) using a single upconverting IQ mixer. By using two DACs to generate both the I and Q components of the frequency comb, the ambiguity between positive and negative frequencies is removed and the full bandwidth of the DAC can be utilized, rather than the standard Nyquist limit of $f_{DAC}/2$.

After the signal is upconverted, it is sent through the cryostat and HEMT, and mixed back to baseband using a downconverting IQ mixer. This baseband signal is digitized by dual analog-to-digital converters (ADCs), after which Field Programmable Gate Array (FPGA)-based signal processing hardware is used to channelize the signal in real time, producing the phase and amplitude of the signal transmitted past each resonator in the array. Further real-time processing can then be performed directly on the FPGA, e.g., time-averaging for slow signals or triggering algorithms for pulse data.

6.6.1 ROACH-based hardware

For this work, the multichannel digital readout described above was implemented using the Reconfigurable Open Architecture Computing Hardware (ROACH) developed by the Collaboration for Astronomy Signal Processing and Electronics Research (CASPER) [319] at UC Berkeley. The ROACH provides a flexible digital signal processing (DSP) platform consisting of a Xilinx Virtex-5 FPGA with interfaces to four 10 Gbps ethernet connections, onboard DRAM memory, and 2 Z-DOK connectors that provide an interface for connection of custom data acquisition hardware. An integrated PowerPC processor running Linux interfaces with the FPGA and allows control of the system.

For MKID-based readouts, Prof. Ben Mazin at UCSB and collaborators have developed DAC and ADC boards to interface with the ROACH [320]. The DAC board consists of dual 16-bit, 1 GS/s Texas Instruments DAC5681 chips allowing generation of a frequency comb with up to $\sim$1 GHz of bandwidth. However, the ADC boards provide the limitation on system bandwidth with dual 12-bit Texas Instruments 550 MS/s ADS5463 chips. As

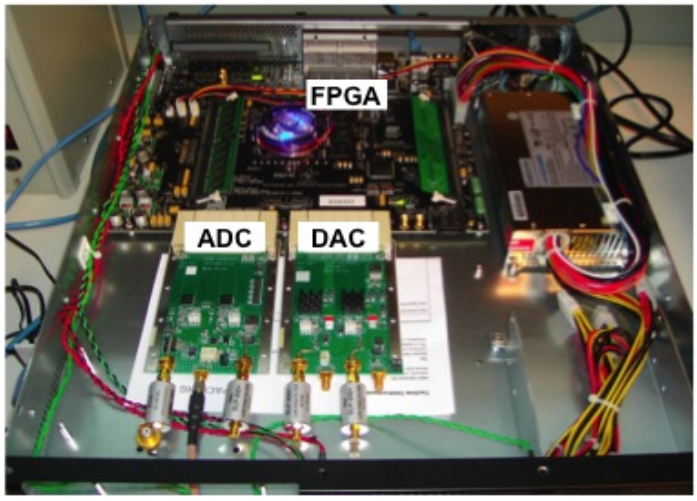


Figure 6.17: Photograph of ROACH hardware with custom DAC and ADC boards.

described below, even the ∼500 MHz bandwidth available with the current ADC boards is sufficient to readout an entire MKID-based iZIP style detector, although ADC capabilities are constantly improving. The number of channels which can be multiplexed with a single ROACH board will correspondingly grow with time. A photograph of the ROACH board and custom DAC/ADC interface boards used for this work is shown in Fig. 6.17.

6.6.2 DDC channelizer

The channelization step is performed by custom firmware running on the ROACH, which takes the baseband signal digitized by the ADC and produces the phase and amplitude of the signal transmitted past each resonator. For the device testing described in this work, a simple 20-channel digital downconverter (DDC) was developed to perform the channelization. An overview of the DDC firmware is shown in Fig. 6.18 and described below. This system is the digital analog of the homodyne mixing scheme shown in Fig. 6.16a, where each analog synthesizer and IQ mixer is now replaced with their digital equivalent. It is implemented using a combination of the CASPER DSP blocks and the standard Xilinx DSP library. To more easily fit the design within the resources of the FPGA, the design described below was implemented using a clock speed of 340 MHz rather than the full 550 MHz of which the system is capable.

The first step is to digitally mix the input signal from the ADC with the corresponding input tone. For each readout channel, a sine and cosine waveform is generated using using a

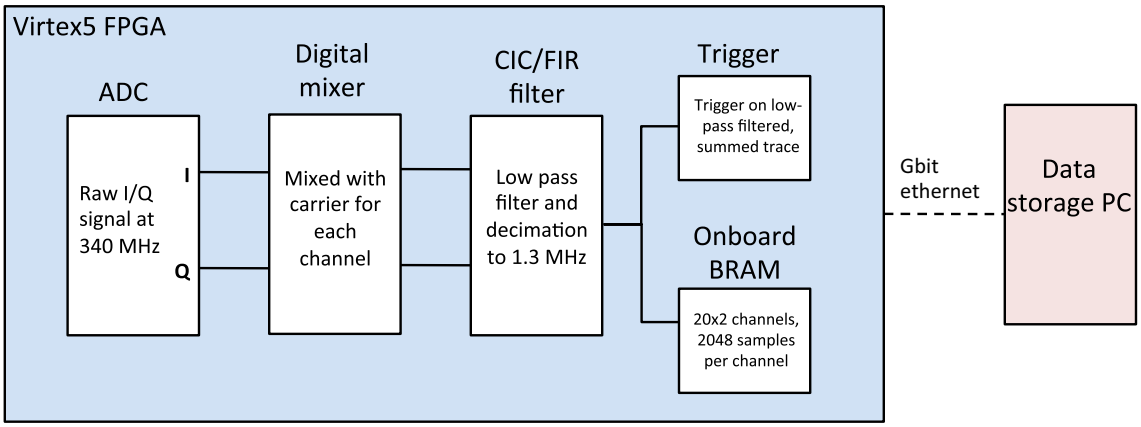


Figure 6.18: Schematic of DDC channelizer firmware. The signal from the ADC is digitally mixed with the input signal for each channel, and then low-pass filtered and decimated to an output data rate of 1.3 MHz. A 2048-sample time stream for each resonator is stored in onboard memory and read out to a data acquisition computer each time a trigger is issued by the firmware.

Xilinx Direct Digital Synthesizer (DDS) block. This block provides a look-up table (LUT) of the waveform which is played back and multiplied against a copy of the digitized input signal from the ADC. The resulting signal is then low-pass filtered and decimated by a factor of 256 using a 2-stage filter bank, leaving a bandwidth of 1.3 MHz for each channel.

The decimation is implemented in two stages to reduce the FPGA resources needed to perform the filtering. The first stage consists of a Cascaded Integrator-Comb (CIC) filter, which provides an efficient decimating filter using only adders and subtractors, with no multipliers. This stage reduces the data rate by a factor of 32 from the input sample rate of the ADC. However, the CIC filter does not provide a sufficiently sharp roll-off for isolation of the channels, so a second-stage Finite Impulse Response (FIR) filter is required to tune the filter-bank frequency response. Since the CIC has reduced the sampling rate by a factor of 32, the resulting data rate is low enough that all 20 channels can be filtered by a single time-domain multiplexed FIR. Following the FIR filter, a second decimation stage reduces the rate by a further factor of 8. The combined response function of the 2-stage filter is shown in Fig. 6.19. The two-stage filter has a flat passband with the 3 dB point of the roll off at 1.3 MHz.

After low-pass filtering, the data rate has been reduced by a factor of 256 to 1.3 MHz, but there are now 40 independent data streams (I and Q for 20 channels). These data

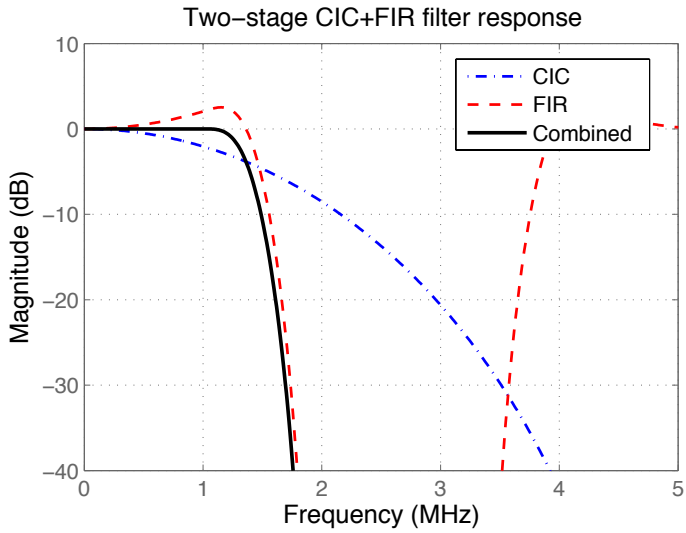


Figure 6.19: Frequency response of the two-stage decimating filter. The first-stage CIC filter (blue, dash-dotted) produces a slow roll off which is compensated by the second-stage FIR filter (red, dashed). The total response (black) has a flat passband (<0.1 dB ripple) with 3 dB point at 1.3 MHz and >20 dB isolation for frequencies above 1.6 MHz.

streams are buffered in a circular buffer in the on-chip block RAM (BRAM) holding 2048 samples for each channel (∼1.5 ms long timestreams), and a copy is sent to a trigger module. The trigger logic calculates the phase signal from I and Q for each channel and sums the phase timestream over all resonators. If desired, a limited subset of resonators to be used for triggering can be selected at run time. The summed phase is then low-pass filtered using a single-stage FIR with a cutoff frequency of 20 kHz to remove high-frequency noise and inter-tone mixing products. When the trigger logic detects an excursion above threshold in the filtered trace, the data buffer is frozen and the FPGA waits for the data acquisition computer to read out the current data before resuming.

While data taking is in progress, a Python server running on a separate data acquisition (DAQ) computer monitors the ROACH and waits for a trigger to be issued by the FPGA. When a trigger is recorded, the DAQ reads out the data buffer using the KATCP protocol [321] through the PowerPC, and stores the traces to disk. It then notifies the ROACH that the read is finished and waits for a new trigger. The data rate of this system is currently limited to $\lesssim 5$ Hz by the overhead required to acknowledge triggers and transfer data between the ROACH and the DAQ computer via KATCP. Future systems using the 10 Gb ethernet ports on the ROACH can increase the achievable data rate by several orders

of magnitude.

Although this system is sufficient to read out the 20-channel prototype devices described above, a simple 1-stage, DDC-based channelizer is not extendable to the several hundred readout channels needed for a full detector. The design described above already uses nearly $\sim$80% of the available resources on the Virtex-5 FPGA for only 20 channels. A more scalable design is being developed by B. Serfass at UCB and S. McHugh at UCSB, which implements a two-stage DDC allowing 256 channels at $\sim$1 MHz of bandwidth per channel. This design uses a polyphase filter bank (PFB) to provide a coarse channelization prior to the final digital mixing stage. This reduces the input data rate to the mixers and filters, allowing a single set of mixers and filters to time-multiplex a correspondingly larger number of channels.

Chapter 7

Results from prototype phonon-mediated devices

In Chap. 6, we discussed the design of prototype phonon-mediated MKID arrays. In the following sections, we describe results from testing arrays fabricated from both Al and high-resistivity TiN and NbTiN films. Device fabrication is described in Sec. 7.1, measurements of basic array parameters are described in Sec. 7.2, phonon-mediated detection and pulse reconstruction are described in Sec. 7.3, and the measured energy resolution from calibration data is described in Sec. 7.5.

7.1 Device fabrication and mounting

The devices described in this thesis were fabricated by B. Bumble at the Microdevices Laboratory at JPL. The arrays were patterned from a single film (either Al, TiN, or NbTiN) deposited by dc magnetron sputtering at ambient temperature onto high-resistivity ($\rho > 5$ kΩ cm), 100 mm diameter Si wafers. Typically, 1 mm thick substrates were used, as these were the thickest that could be easily patterned without modification to the Canon EX3 stepper. Both (100) and (111) orientation wafers were used, although no significant differences in the resonator properties were observed between the two crystal orientations. To remove native oxide, the wafers were deglazed in hydrofluoric acid (HF) prior to deposition of the films.

The transition temperature of the TiN and NbTiN films can be controlled by varying the nitrogen content of the films [302]. In the JPL system, the films are deposited by sputtering from either Ti or NbTi targets in an argon atmosphere, while injecting a controlled amount of N_2 during deposition. Figure 7.1 shows the measured T_c for TiN and NbTiN films

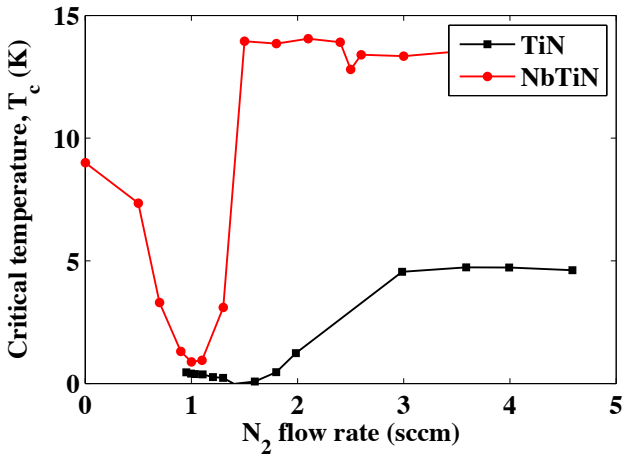


Figure 7.1: Measured T_c versus N_2 flow rate during deposition for TiN (black, squares) and NbTiN (red, circles). For NbTiN, the T_c can be tuned to ≈1 K for flow rates near 1 sccm of N_2 during deposition. For TiN, flow rates near 2 sccm give a $T_c \approx 1$ K. *Figure adapted from LeDuc et al. [302]; NbTiN data provided by B. Bumble*

deposited at JPL as a function of the flow rate of N_2 during deposition.

Both the nitride and Al films were patterned by UV projection photolithography followed by inductively coupled plasma etching with a chlorine-based chemistry (BCl_3/Cl_2). For the Al films, larger values of Q_i were obtained for Al films by depositing a thin (∼5–10 nm) layer of niobium immediately following the Al deposition, patterning the bilayer and then removing the Nb protection layer with a fluorine-based etch (SF_6), which removes only the Nb and stops on the Al.

Following patterning, the devices were diced into 22 mm x 20 mm chips and mounted in the testing enclosure. Figure 7.2b shows a chip mounted in its housing, which is a custom gold-plated copper enclosure that provides a transition between female Anritsu K-connectors on the input and output sides to 50 Ω microstrip-to-CPW transition boards that can be wirebonded directly to the CPW bond pads on the device. The chip is supported by 1 mm x 1.5 mm ledges in each corner, as well as 2 mm x 1.5 mm ledges directly underneath the bond pad, which support the center of the chip during wirebonding. Chips were typically affixed to the box using a thin layer of GE varnish applied to the supports at the corners and under the bond pads, with a total contact area of 12 mm^2.

After mounting the chips in the testing enclosure, a calibration source could be positioned to illuminate the side of the substrate opposite the face on which the resonators were

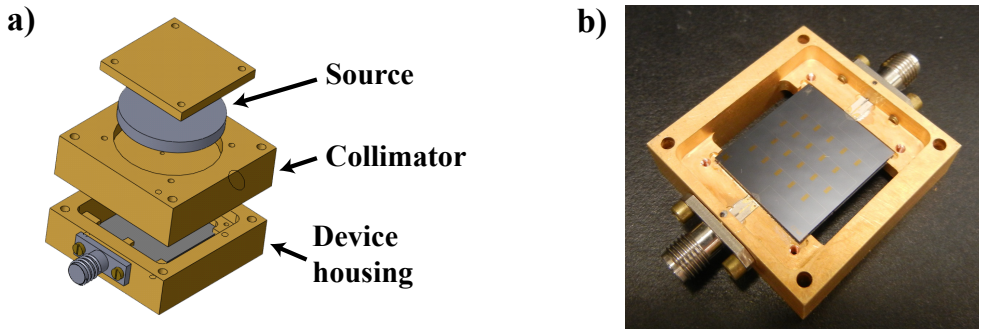


Figure 7.2: a) Schematic of device mounting hardware, including source holder. b) Photo of device mounted in testing enclosure.

patterned, as shown in Fig. 7.2a. The devices were then cooled to 50 mK in an Oxford Kelvinox 25 dilution refrigerator. A detailed description of the dilution refrigerator test bed can be found in [318] and [322].

Nearly 15 devices were fabricated and tested over the course of 24 months, summarized in Table 7.1. A detailed discussion of the testing results is presented in the following sections.

7.2 Resonator parameters

Figure 7.3a shows the measured coupling Q for a variety of devices as a function of frequency relative to the center of the array. As discussed in Sec. 6.5.1, the coupling Q for all resonators on the array was designed to be $5{\times}10^4$ based on SONNET simulations, while the measured Q_c varies from 10^3 to $5{\times}10^5$, with a median $Q_c \approx 10^4$. The variation in Q_c is likely due to the presence of position-dependent coupling of the resonators to the even CPS mode, in addition to the odd mode for which the Q_c simulations were performed. Future devices with lower-impedance feedlines are expected to improve the Q_c uniformity [305], as described in Sec. 6.5.2.

The internal Q for the same devices is shown in Fig. 7.3b. The Q_i is not strongly dependent on material, with most measured values from $Q_i = 10^5$–10^6. Therefore, for the typical resonator with $Q_c \approx 10^4$, the limit $Q \approx Q_c \ll Q_i$ applies, although a small fraction of resonators have $Q_c \approx Q_i$. All values of Q_i in Fig. 7.3b were measured with no source illuminating the substrate, and at temperatures, $T \ll \Delta$, so that thermal quasiparticles are negligible. It is not known what currently limits the maximum value of Q_i for these devices to $<10^6$. Possible sources of residual low-temperature dissipation include the presence of

Table 7.1: Summary of device parameters and testing.

Wafer ID:	**Resonator material:**	**T_c:**	**Resonator thickness:**	**Substrate thickness:**	**Testing summary:**
B100201	TiN	≈1.3 K	100 nm	1 mm	R143: Illuminated with ^{55}Fe, small pulses on all resonators, exponential spectrum
B100430.1	TiN	≈1 K	100 nm	1 mm	R146: Small pulses from ^{55}Fe source, exponential spectrum
B100430.2	TiN	≈0.2 K	100 nm	1 mm	R144: ∼20° pulses from ^{55}Fe source, spectral peak seen on single resonator
B100430.3	TiN	≈0.5 K	100 nm	1 mm	R149: Spectral peaks on all resonators from ^{129}I source, baseline $\sigma_E \approx 0.3$ keV, Q_i∼5×10^5; R150: Collimated ^{55}Fe source, Q_i∼1×10^5; R151: Collimated ^{109}Cd source, Q_i∼5×10^4
B100708	TiN	≈1.5 K	50 nm	1 mm	R148: Peak from ^{55}Fe source, baseline $\sigma_E \approx 0.6$ keV
B101027.1	NbTiN	≈0.65 K	50 nm	1 mm	R154: ^{129}I data, baseline $\sigma_E \approx 0.5$ keV; R150: Collimated ^{109}Cd data
B110506.3	TiN	4.2 K	75 nm	1 mm	R152: Stoichiometric TiN, uniform α and Δ across array, but no pulses from source
B110606.2	TiN	2.6 K	75 nm	1 mm	R153: fairly uniform α and Δ across array, small pulses from ^{109}Cd source, Q_i∼5×10^4
B110609	NbTiN	5.7 K	75 nm	1 mm	R153: T_c higher than desired ≈2 K, no pulses seen from ^{129}I source
B110830	Al	1.25 K	75 nm	1 mm	R155-R159: Varied sources and mounting to determine dependence of Q_i and phonon lifetime. No differences between securing chips with Ge varnish or nylon clamps. ^{109}Cd source lowered Q_i to ∼5×10^4
B111007	Al	1.35 K	25 nm	1 mm	R160-R161: ^{129}I source, baseline $\sigma_E \approx 0.4$ keV
B111111.1	Al	1.2 K	75 nm	0.3 mm	R162: Thinner substrate device to test effect on phonon collection and lifetimes, no source
B111111.2	Al	1.35 K	25 nm	0.3 mm	R162: ^{129}I source, phonon collection consistent with thicker substrates, ∼50% shorter pulse fall times
B111219	Al	1.4 K	15 nm	1 mm	R163: ^{129}I source, baseline $\sigma_E \approx 0.4$ keV

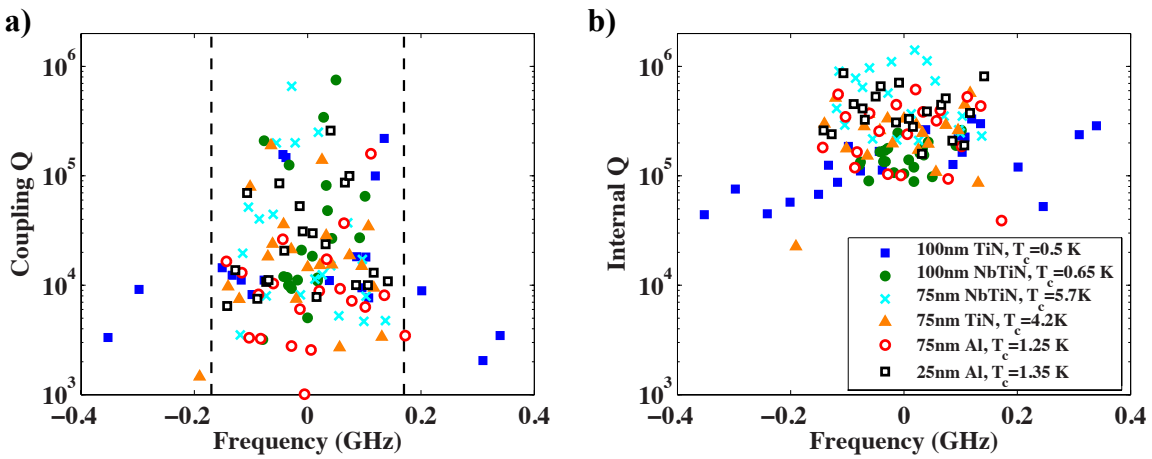


Figure 7.3: a) Measured Q_c versus frequency for representative arrays fabricated from TiN, NbTiN, and Al. The measured Q_c varies by more than 2 orders of magnitude from 10^3–5×10^5, with most resonators having $Q_c \approx 10^4$. The frequency scale denotes the resonators' position relative to the center frequency of the array. The design frequencies were spaced evenly by 10 MHz within 0.2 GHz of bandwidth, although the observed frequencies cover a wider range. The vertical dashed lines indicate the maximum bandwidth of the readout system described in Sec. 6.6. b) Measured Q_i versus frequency. The measured internal Qs vary from 10^5–10^6, so $Q_i \gg Q_c$ for the typical resonator. The variation of Q_c and Q_i is not strongly dependent on the resonator material, although the $T_c = 0.5$ TiN device has the largest spread in frequency.

a nonthermal quasiparticle population (e.g., due to stray pair-breaking radiation from the 4 K or 800 mK stages of the dilution refrigerator) or film defects at grain boundaries or the film surface.

It was found that illuminating the device with the ^{109}Cd source lowered the measured Q_i to $\lesssim 5\times10^4$ relative to $Q_i \approx 5 \times 10^5$ measured for the same device, in the same housing, when the source was removed. This provides an example where the low-temperature Q_i was limited through a non-thermal quasiparticle density (in this case originating from the source). Calibrating this response to data taken at elevated temperature indicated that to reproduce the measured Q_i required a steady-state quasiparticle population of $N_{qp} = 6\times10^{10}$ summed over all 20 resonators. The power required to maintain this density is given by: $P = N_{qp}\Delta/\eta_{ph}\tau$, where η_{ph} is the overall efficiency for converting incident energy into quasiparticles in the resonators and τ is the characteristic time with which the energy is lost. Conservatively taking $\eta_{ph} = 1$, and using the measured value of $\Delta = 200$ μeV for 75 nm Al, gives $P \approx 20$ nW, for a characteristic decay time given by the quasiparticle lifetime, $\tau = 100$ μs. However, the total power emitted by the 10 μCi source is only 5 nW.

If excess generation of quasiparticles by the source was the cause of the decrease in Q_i, a significant fraction of the power from the source must have reached the substrate even though there was a 8 mm thick Cu collimator exposing only 5% of the active area of the source to the chip. In addition, the time constant with which the energy decays must have been longer than the quasiparticle lifetime.

Figure 7.3a also shows the spread in the measured frequencies relative to the center frequency of the array for each of the devices. Although the array was designed to fit within 200 MHz of bandwidth, the typical measured frequency spacing was 50% larger, leading to an array bandwidth ≈300 MHz. For the Al and NbTiN arrays, all 20 resonators still fell within the total system bandwidth of 340 MHz for the readout system described in Sec. 6.6. However, the TiN device with $T_c = 0.5$ K showed variations in the resistivity and gap of order 20% across the device, leading to an array bandwidth >600 MHz. For this device, only 14 resonators could be read out simultaneously given the readout system bandwidth.

In addition to the relative variations in the array spacing described above, the central frequency of the array varies with the kinetic inductance fraction. The kinetic inductance fraction is given by $\alpha = L_{KI}/L_{tot}$, where the total inductance, $L_{tot} = L_{KI} + L_m$ consists of the kinetic inductance plus the magnetic inductance, L_m, as discussed in Sec. 6.3.1.1. If we define, $f_m = 1/2\pi\sqrt{L_m C}$, then we can determine α from the shift in the measured resonant frequency, f_0, as $T \to 0$ relative to f_m [298]:

$$\alpha = 1 - \left(\frac{f_0}{f_m}\right)^2 \tag{7.1}$$

As will be discussed in Sec. 7.5.4, α can also be determined from fits of the resonator response versus temperature to Mattis-Bardeen theory. For the Al devices, we use these fits to determine $f_m = 5.21$ GHz, for the central resonator, which agrees with the value of f_m from SONNET simulations within 5%. The resulting value of α for each device is then calculated from the observed array frequencies and is shown in Fig. 7.4. From Eqs. 6.2 and 6.7, the surface inductance is given by, $L_s = \hbar R_{s,n}/\pi\Delta$, where $R_{s,n}$ is the sheet resistance in the normal state. The 75 nm thick TiN and NbTiN films typically have $R_{s,n} = 20$–$30\ \Omega$/sq, leading to $\alpha \approx 0.25$ even for the films with $T_c \sim 5$ K. The lower T_c, high-resistivity films show values of α as high as 0.8. In contrast, due their lower surface resistance, even the

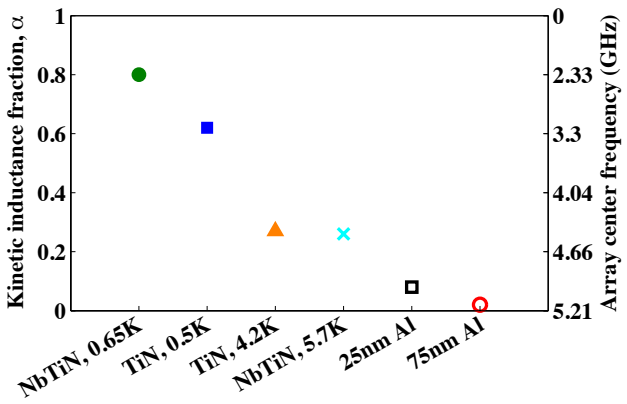


Figure 7.4: Kinetic inductance fraction, α, versus material. The high-resistivity films with $T_c < 1$ K show α as high as 0.8. In contrast, the lower-resistivity Al film gives much smaller values of α. The right axis shows the corresponding frequency of the central resonator in the array, from which α can be calculated using Eq. 7.1.

25 nm Al devices have α of only $\sim$0.1.

Although the low-T_c, high-resistivity films show the largest values of α, they show the poorest resonator-to-resonator uniformity. As shown in Fig. 7.1, small changes in composition can lead to large changes in the film properties. Gradients in the local N_2 density over the surface of the chip can lead to corresponding gradients in α and Δ across the device. These gradients vary the spacing of the resonators in frequency, as shown in Fig. 7.3, and also lead to variations in the relative resonator response, which must be removed to reduce variations of the detector response with interaction location. Figure 7.5 compares the frequency response of the resonators versus temperature for all 20 resonators on several devices. While the $T_c = 0.5$ K TiN device has poor uniformity evidenced by a wide range of responses across the array, both the $T_c = 4.2$ K TiN device and the Al device are less sensitive to deposition conditions and produce much more uniform resonators. For larger devices with $A_{sub} = 100$ cm^2, nonuniformities with position across the substrate will be a significant concern.

7.3 Phonon-mediated particle detection

Figure 7.6 shows the phase response in each MKID following an example 200 keV interaction in the substrate due to an interaction from a cosmic ray or natural radioactivity. Large, prompt pulses are seen for resonators 9 and 20, while smaller, delayed pulses are seen

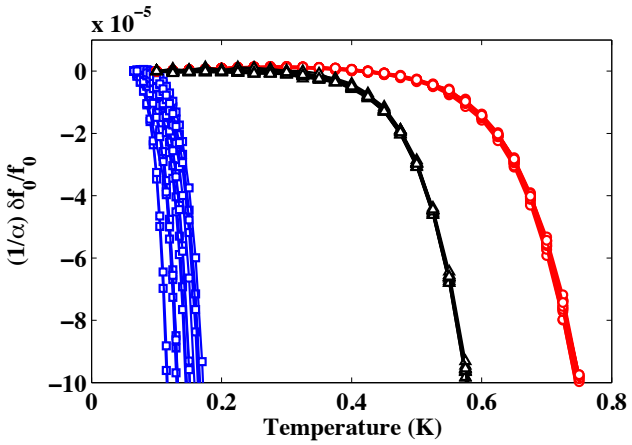


Figure 7.5: Fractional frequency response, $\frac{\delta f_0}{f_0}$, scaled by $1/\alpha$ versus temperature for the $T_c = 0.5\ K$ TiN device (blue, squares), the $T_c = 4.2\ K$ TiN device (red, circles), and the 25 nm Al device (black, triangles). For each device the response of all 20 resonators is plotted. While the $T_c = 4.2\ K$ TiN and Al devices show good uniformity, with the response of all resonators lying along the same curve, the lower-T_c TiN device shows a wide range of responses at each temperature.

for the resonators further from the interaction site. By summing the energy collected by each resonator, the total energy of the interaction can be determined. In addition, the relative partitioning of energy between the channels and the delay between the phonon signal detected in each channel can be used to determine the location of the interaction. Comparing with the resonator locations from Fig. 6.13a indicates that this interaction occurred near the top of the central column of resonators.

In the following section, we describe in detail how the position and energy of the interaction is determined from the response seen in each resonator. All data shown in this section were taken using the 25 nm Al device from B111007, illuminated by the ^{129}I source.

7.3.1 Energy reconstruction

For each interaction in the substrate, 40 2048-sample long timestreams are recorded (1 trace for I and Q for each resonator). From these traces, we would like to determine the energy and location of the interaction in the substrate. The first step is to calculate the phase and amplitude of the signal transmitted past the resonator from I and Q. Prior to taking data, a sweep in frequency around the resonance is performed to determine the shape of the resonator response. The resonances loops are then fit to the expected response from

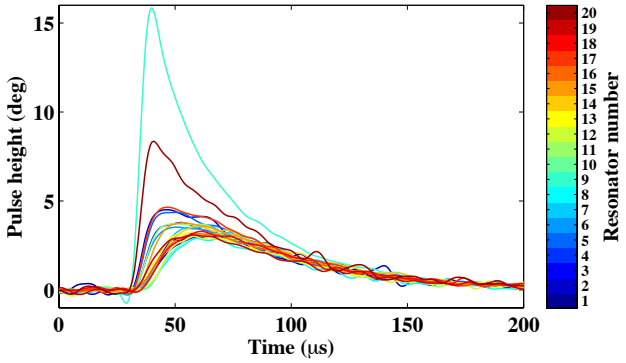


Figure 7.6: Coincident phase pulses in each resonator for a 200 keV phonon-mediated interaction in the substrate. The coloring indicates the resonator ordering in frequency space, with resonator 1 corresponding to the lowest frequency and resonator 20 denoting the highest frequency. A 200 kHz low-pass filter has been applied, and the pulse heights have been rescaled by the relative resonator responsivity as described in Sec. 7.3.1 below.

Eq. 6.14 [298] to determine the loop center, z_c, resonant frequency and quality factor. For the timestream for resonator i, we define $z_i(t) = I_i(t) + jQ_i(t)$, and calculate the change in the phase and amplitude of the response, relative to the center of the loop:

$$\delta\phi(t) = \phi_0 - \arctan\left(\frac{\mathrm{Im}[z_i(t) - z_c]}{\mathrm{Re}[z_i(t) - z_c]}\right) \tag{7.2}$$

$$\frac{\delta a(t)}{a} = \frac{|z_i(t) - z_c|}{r_0} \tag{7.3}$$

where ϕ_0 is the phase at the probe frequency in the steady state, $r_0 = |z_i(0) - z_c|$ is the radius of the loop in the steady state, and the signs have been defined to make the quasiparticle response move in the positive phase and amplitude direction.

As discussed in Sec. 7.4 below, the noise in the phase and amplitude direction is dominated by HEMT noise at the signal frequencies of interest. Since the phase response is larger by a factor of $\beta = S_2/S_1 \approx 2$–3, the signal-to-noise is dominated by the phase information and we can ignore the amplitude information with negligible loss of sensitivity. The energy deposited in each resonator is initially determined by three different methods:

1. Each trace is filtered by a 20 kHz low-pass filter, preserving the pulse fall time but removing high-frequency noise and position-dependent variations in the rising edge of the pulse. The energy is then estimated by integrating the filtered pulse.

2. An optimal filter fit is performed in the frequency domain using a single template to

fit the pulses.

3. A two-template optimal filter fit is performed, allowing for both prompt and delayed components in the pulse with different fall times.

7.3.1.1 Single template optimal filter

While the integral estimator is less sensitive to systematic variations in the pulse shape with position, the optimal filter estimate gives improved noise resolution at low energies for a known pulse shape. If we can write the pulse timestream as $p(t) = As(t) + n(t)$, where A is the amplitude of the pulse, $s(t)$, is the normalized pulse shape and $n(t)$ is a noise realization, then the optimal filter estimate of the pulse amplitude, $\hat{A}$, is (for a review see, e.g. [259] or [228]):

$$\hat{A}(t_0) = \frac{\sum_n \frac{\tilde{s}_n^* p_n}{J_n} e^{2\pi j t_0 f_n}}{\sum_n \frac{|\tilde{s_n}|^2}{J_n}} \tag{7.4}$$

where J_n is the noise power spectral density, t_0 is the start time offset of the pulse relative to the template, $f_n = f_s(n/N)$ for for sampling frequency, f_s, and trace length, N, and the index n runs over the length of the trace. It can be shown that this estimate is "optimal" in the sense that it maximizes the signal-to-noise of the amplitude estimator in the presence of Gaussian noise with a known power spectral density [259].

In general, the start time offset, t_0, is unknown so a search must be performed to find the value of t_0 that provides the best fit to the data. For a known pulse shape and stationary noise, the best estimator of the time offset, $\hat{t}_0$, is given by the value of t_0 that maximizes the amplitude estimator, $\max[\hat{A}(t_0)]$ [259]. Thus, to determine $\hat{A}$ we calculate $\hat{A}(t_0)$ for a variety of t_0 and take the maximum. This search adds negligible computation time since the time offset can be applied by the phase factor shown in Eq. 7.4 and only a single Fourier transform must be computed for each pulse. However, any search for the start time will be biased by noise very near threshold where a positive noise fluctuation in a neighboring bin may cause the start time to be misestimated, as discussed in Sec. 4.2.2. This bias causes the mean of $\hat{A}$ to be positive for noise traces where no signal is present. This bias vanishes as the width of the search window goes to zero or the energy of the pulse becomes much larger than the average single-bin noise fluctuation.

For a given event, these estimates give a single pulse amplitude, ϕ_i for each of the i resonators. To construct the total energy of the event, we must sum the energy over resonators:

$$E = \left(\frac{E_0}{\phi_0}\right) \sum_i \frac{\hat{A}_i}{r_i} \tag{7.5}$$

where r_i is a dimensionless factor accounting for the relative responsivity of resonator i, and (E_0/ϕ_0) converts the phase response to energy in units of keV, determined by calibration lines with known energy.

7.3.1.2 Relative responsivity

To determine the relative responsivity of the resonators, r_i, high-energy interactions depositing $\geq$100 keV of energy in the substrate from cosmic rays or natural radioactivity are used. At long times after the interaction, phonons have diffused sufficiently that the initial position-dependent response has decayed away and the resonators are uniformly illuminated by phonons. The relative responsivity can then be found by scaling the exponential tails of the pulses in each resonator at long times to give a common amplitude, following the relative responsivity calibration used for the iZIP detectors. In practice, the energy estimate for the pulse is dominated by the long time scale information, so we find that it is sufficient to compare the relative amplitude estimates between channels, as shown in Fig. 7.7a, rather than computing the relative responsivity directly from the traces.

As shown in Fig. 7.7b, scaling by the relative responsivity improves the resolution by reducing position-dependent smearing that arises as the energy partitioning across different sensors is varied. Following this initial rescaling, we attempt to remove any small remaining variations in responsivity by solving for the optimized values of r_i that maximize the signal-to-noise of calibration lines from the source. The optimization is performed using Matlab's `fminsearch`, which provides a gradient-free direct search algorithm. Figure 7.7b shows that this optimization provides a small ($\approx$ 10% improvement in resolution relative to the scalings determined in Fig. 7.7a.

After rescaling by the relative responsivities, we denote the "primary channel" as the channel showing the largest phase response. Similarly, we denote the "opposite channels" as the set of channels that show the smallest, delayed response. For the example event shown in Fig. 7.6, channel 9 is the primary channel, while the group of 11 resonators showing a

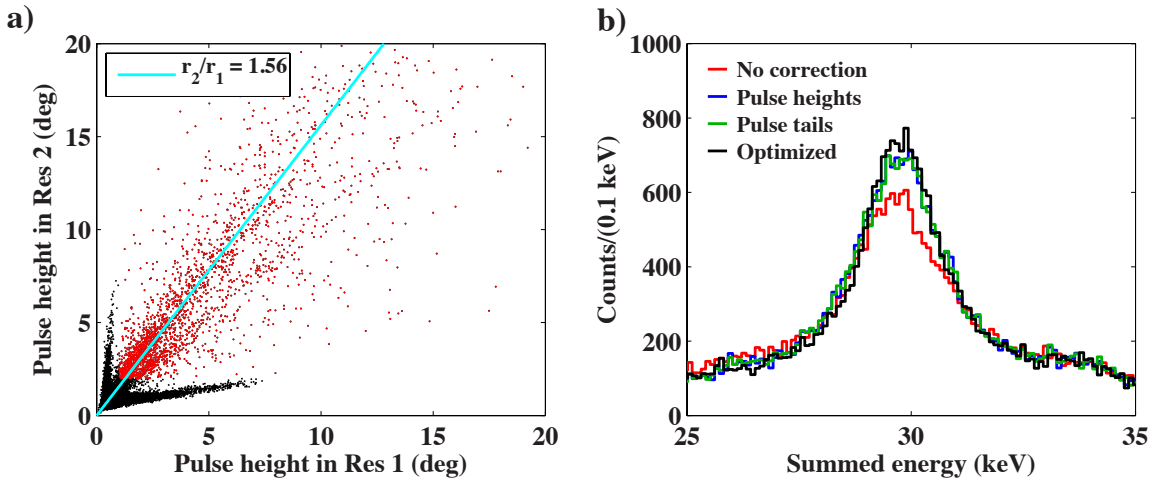


Figure 7.7: a) Pulse height in resonator 2 versus pulse height in resonator 1. The responsivity of resonator 2 relative to resonator 1 is determined by fitting the slope of the response for points above ≈100 keV (red). For the 30 keV events from the source (black), a larger range of slopes is possible due to the relative variation in energy partitioning with the interaction location. b) Effect of rescaling the resonators by their responsivity on the energy resolution of the 30 keV ^{129}I line. Both the rescaling factors determined from the slope of the relative amplitudes shown in a) as well as the rescalings inferred from matching the pulse tails at long times give similar improvement over the resolution with no rescaling. An additional ∼10% improvement is obtained by optimizing the scalings using a multidimensional minimization.

nearly identical delayed response would be labeled as opposite channels.

7.3.1.3 Two-template optimal filter

As shown in Fig. 7.6, the primary and opposite channels have significantly different pulse shapes, and a single pulse template cannot be used to properly determine the energy, as was assumed in Sec. 7.3.1.1. Instead, we find that our pulses are generally described by two components: $p(t) = A_p s_p(t) + A_o s_o(t) + n(t)$, where A_p is the amplitude of the prompt component, which decays away with the quasiparticle lifetime, and A_o is the amplitude of the delayed component, which decays with the phonon lifetime. The pulse shape for the prompt component and delayed component is given by s_p and s_o, respectively.

These templates are determined by first averaging many opposite channel pulses and fitting to a double-exponential functional form to determine s_o:

$$s_o(t) = a_o \left(1 - e^{-(t-t_0)/\tau_o}\right) e^{-(t-t_0)/\kappa_o} \tag{7.6}$$

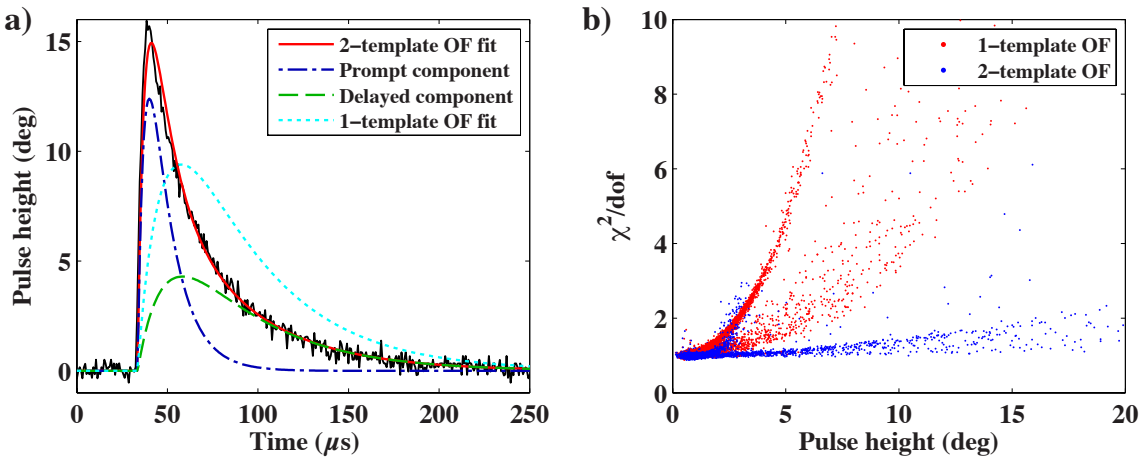


Figure 7.8: a) Comparison of single-template and two-template optimal filter fits to the primary channel pulse for the 200 keV event shown in Fig. 7.6. While the pulse is poorly described by a single template with the phonon fall time, it is well described by a sum of a prompt component, which decays with the quasiparticle lifetime, and a delayed component, which decays with the phonon lifetime. b) Improvement in the reduced χ^2 for the 2-template optimal filter fit relative to the single-template fit. The χ^2 for the population of primary channel events that was misfit with the single-template fit is significantly improved.

where a_o is the overall amplitude, typically normalized such that $\max(\{s_o\}) = 1$, τ_o is the pulse rise time and κ_o is the fall time. The prompt template, s_p, is then determined by fitting an averaged primary channel pulse to a model with a fixed component equal to s_o plus an additional double-exponential form:

$$s_p(t) = a_p \left(1 - e^{-(t-t_0)/\tau_p}\right) e^{-(t-t_0)/\kappa_p} + s_o(t) \tag{7.7}$$

where the pulse start time, t_0, is constrained to be the same for both the prompt and delayed components. We can repeat the minimization in the frequency domain to obtain the optimal filter amplitude estimates for each component:

$$\begin{pmatrix} \hat{A}_p \\ \hat{A}_o \end{pmatrix} = \begin{pmatrix} \sum_n \frac{|\tilde{s}_{p,n}|^2}{J_n} & \sum_n \frac{\mathrm{Re}(\tilde{s}^*_{p,n}\tilde{s}_{o,n})}{J_n} \\ \sum_n \frac{\mathrm{Re}(\tilde{s}^*_{p,n}\tilde{s}_{o,n})}{J_n} & \sum_n \frac{|\tilde{s}_{o,n}|^2}{J_n} \end{pmatrix}^{-1} \begin{pmatrix} \sum_n \mathrm{Re}\left(\frac{\tilde{s}^*_{p,n}}{J_n}\tilde{p}_n\right) \\ \sum_n \mathrm{Re}\left(\frac{\tilde{s}^*_{o,n}}{J_n}\tilde{p}_n\right) \end{pmatrix} \tag{7.8}$$

where the index n runs over the length of the trace and J is the power spectral density of the noise. Given the energy estimates for each component, the total energy of the pulse is

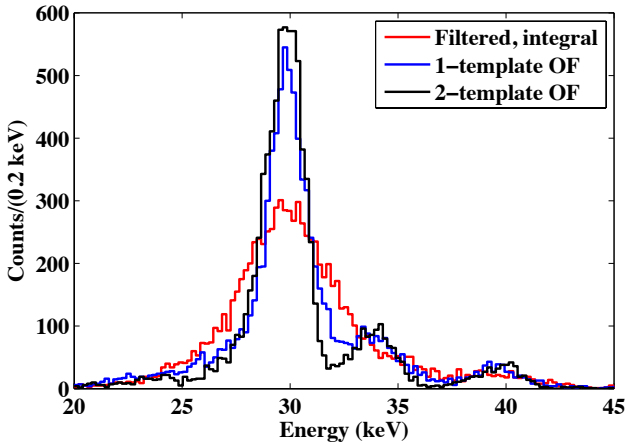


Figure 7.9: Comparison of reconstructed energy spectrum for the integral, single-template optimal filter and two-template optimal filter energy estimates. The integral estimate gives the poorest resolution ($\sigma_E = 2.47$ keV), with the two-template optimal filter ($\sigma_E = 0.85$ keV) giving ≈10% better resolution than for the single template ($\sigma_E = 0.95$ keV).

reconstructed as:

$$E = \left(\frac{E_0}{\phi_0}\right) \sum_i \frac{\hat{A}_{p,i} + w\hat{A}_{o,i}}{r_i} \tag{7.9}$$

where $w = \left(\sum_n s_o\right) / \left(\sum_n s_p\right)$ weights the primary and opposite channels by their integrals. Since the two pulse components measure the quasiparticle creation rate in the prompt and delayed components convolved with the quasiparticle decay time, w gives the ratio of the total number of quasiparticles created in these components, noting that the linear convolution does not affect the ratio. This weighting was varied around this value to empirically verify that it gives the optimal signal-to-noise in the reconstructed pulse amplitude.

An example two-template fit to the primary channel pulse for the 200 keV event discussed above is shown in Fig. 7.8a. The single-template optimal filter uses a pulse shape similar to the opposite channel pulse and cannot be used to accurately estimate the energy of the primary channel pulse. In contrast, the pulse is well described by the two-template fit. Figure 7.8b shows the typical improvement in the reduced χ^2 of the fit for the two-template fit relative to the single-template fit. The quadratic increase in the χ^2 with energy for the single-template fit is significantly reduced for the two-template fit, with $\chi^2/\text{dof} \approx 1$ for nearly all events from the 30 keV source. At higher energies, a weak quadratic dependence with energy is seen due to residual mismatch of the templates to the true pulse shape.

The effect of this improved reconstruction of the primary channel pulses on the energy

resolution is shown in Fig. 7.9. For fully collected events interacting in the center of the detector substrate, the two-template fit improves the reconstructed resolution by ≈10% relative to the single-template fit. It also reduces the non-Gaussian tails of events occurring between the expected spectral lines from the ^{129}I source. However, the single-template fit does remarkably well considering the significant differences in the template from the true pulse shape. This occurs because the primary channel pulse makes up only a small fraction (typically $\leq 10\%$) of the total energy of the pulse, while the remaining channels are better fit by the single-template pulse shape. Although the integral estimate does not require the assumption of a particular pulse shape, it shows the poorest resolution since it does not optimally take into account the noise spectral density.

7.3.2 Position reconstruction

In addition to reconstructing the energy of the interaction, the location of the interaction can be determined both by the relative partitioning of energy between the sensors and by the arrival time of the phonon signal in each sensor. Figure 7.10 shows an example of the position-dependent information available for a 200 keV interaction and demonstrates that the interaction location can be reconstructed. The sensors nearest the interaction site collect a larger fraction of the total deposited energy than those far from the interaction site. Figure 7.10a shows the energy collected in each sensor for the 200 keV event discussed above. The distribution of pulse heights is approximated by a Gaussian, with the peak of the Gaussian corresponding to the interaction location. This energy partitioning indicates that the interaction occurred near the top of the center column of resonators as previously discussed.

Figure 7.10b shows an independent reconstruction of the event location using the relative arrival time of the phonon signal in each sensor. We define the arrival time as the time at which the pulse in each channel has reached 20% of its maximum pulse height. The relative arrival time is then calculated as the delay between the arrival time in a given sensor and the arrival time in the primary channel. For plotting purposes, in Fig. 7.10b we show the time by which the pulse in each channel precedes the arrival time in the slowest channel to give its maximum value in the primary channel. These arrival times are well described by a cone with a linear dependence on the distance from the primary channel. The peak of this distribution can be used to estimate the interaction location, and again indicates that the

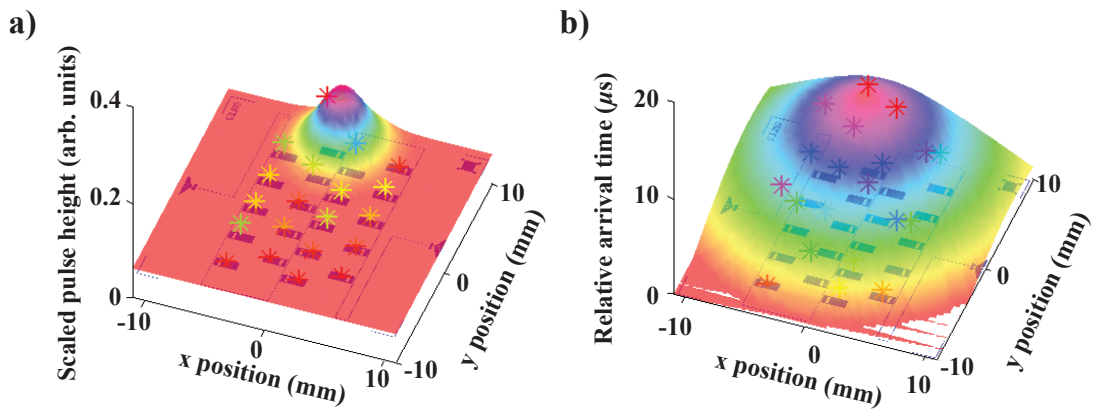


Figure 7.10: a) Reconstruction of the interaction location from the partitioning of energy between channels for the 200 keV event discussed above. The channels closest to the interaction location show the largest amplitude signal, while channels farther away from the interaction site show a smaller response. b) Reconstruction of the interaction location from the arrival time of the phonon pulses in each channel for the same event. The arrival times are calculated relative to the slowest channel so that the maximum value occurs for the primary channel. The phonon timing agrees with the energy partitioning and indicates that the interaction occurred near resonators 9 and 20.

event occurred near the top of the central column of the array. The pulse in the opposite channels arrives ≈ 15 μs after the pulse arrival in the primary channel.

In practice, rather than fitting a Gaussian to the distribution of the energy in each resonator, it is often simpler to define the x and y energy partitions as:

$$P_x = \frac{\sum_i x_i E_i}{\sum_i E_i}, \quad P_y = \frac{\sum_i y_i E_i}{\sum_i E_i} \tag{7.10}$$

where (x_i, y_i) are the coordinates of the ith resonator, and E_i is its energy. An example of the interaction locations inferred from this energy partition for 30 keV events from the ^{129}I source are shown in Fig. 7.11a, where the coloring indicates the primary channel for each event. The partition estimator does not give the location of the event directly in physical units, but instead is some function, $P_{x,y} \equiv P_{x,y}(x, y, z)$, where (x, y, z) is the true location of the interaction in the substrate. If there is sufficient information in the partition quantities that this mapping is one-to-one, it can be inverted to uniquely determine the event location.

In particular, these partition quantities depend on both the x and y position of the event, as well as on the z position. Figure 7.11b shows the partition plot above colored by

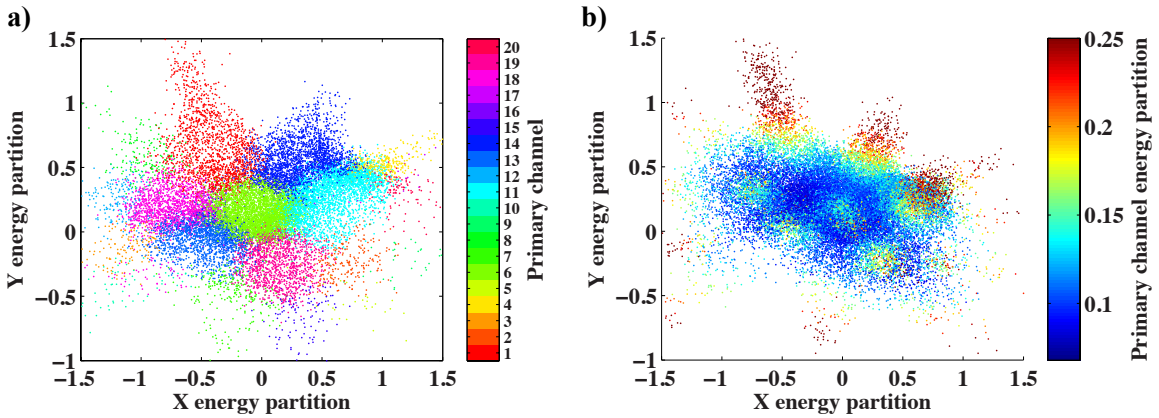


Figure 7.11: a) x and y energy partitions for 30 keV x-rays uniformly illuminating the substrate. The coloring indicates the primary channel. b) Same as a), but with the coloring indicating the primary energy partition, denoting the fraction of the energy collected in the primary channel. Surface events, which have relatively larger values of the primary energy partition, appear as outward going points for the outer sensors, and peaks in the center of the inner sensors.

the primary energy partition, defined as the ratio of the energy collected in the primary channel to the total energy of the event, $P_{prim} = E_{prim}/E$. Events occurring near the edge of the array typically have a large primary energy partition and lead to the "spiky" features along the edge of the partition plot. In addition, events interacting close to the top surface of the substrate have a larger primary energy partition, leading to the dependence of $P_{x,y}$ on z. Future devices with phonon sensors on both substrate faces would allow an additional measurement of z through the partitioning of energy and delay between the phonon signals collected on each face.

Although the timing delays provide an independent estimate of the event location, the signal-to-noise of this estimator is lower than for the partition estimator at low energies. To attempt to improve the signal-to-noise of the delay-based reconstruction, the device was divided into 4 quadrants containing 5 resonators each, and the traces were summed over all resonators in each quadrant. An example of the summed traces for each quadrant for a 30 keV event is shown in Fig. 7.12a. The summed pulses are fit to a double exponential form, and the start time is defined as the time at which each best-fit pulse crosses a threshold set at 50% of the amplitude of the smallest quadrant pulse. We then define the x and y delays

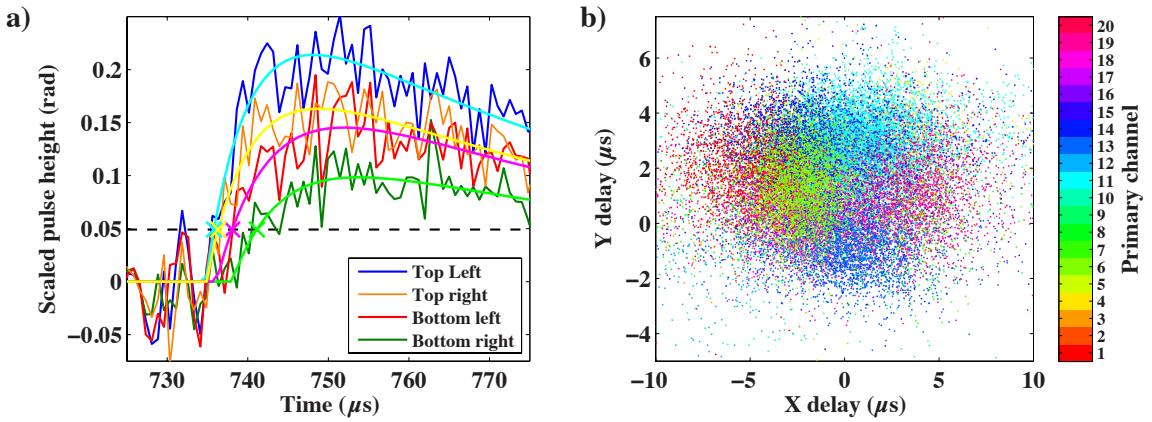


Figure 7.12: a) Determination of the start time for a 30 keV event using the summed pulse in each detector quadrant. The pulses from each resonator are scaled by their relative responsivity prior to computing the summed pulse. The start time for each pulse is determined as the time it crosses the threshold (black, dashed) given by 50% of the smallest amplitude pulse. Typical time delays between the primary and opposite channel pulses are $\lesssim 10$ μs. b) Reconstruction of the interaction location using the relative start times of the summed pulses in each quadrant. The coloring indicates the primary channel for each event.

as:

$$x_{del} = \begin{cases} t_{TL} - t_{TR} & \text{for primary channel in } TR \text{ or } TL \\ t_{BL} - t_{BR} & \text{for primary channel in } BR \text{ or } BL \end{cases} \tag{7.11}$$

$$y_{del} = \begin{cases} t_{BL} - t_{TL} & \text{for primary channel in } TL \text{ or } BL \\ t_{BR} - t_{TR} & \text{for primary channel in } TR \text{ or } BR \end{cases} \tag{7.12}$$

where t_{TL}, t_{TR}, t_{BL} and t_{BR} denote the start times in the top left, top right, bottom left and bottom right quadrants, respectively. The resulting delay-based position reconstruction is shown in Fig. 7.12b.

Comparison of the delay-based reconstruction in Fig. 7.12 with the partition-based reconstruction from Fig. 7.12a, shows that while the delay-based reconstruction is significantly less precise, some position-dependent information can be extracted from the pulse timing. Since the timing resolution depends on the signal-to-noise of the pulses in each channel, improving this delay-based reconstruction requires using higher-energy pulses or summing over many channels to improve the signal-to-noise at low energies.

7.3.2.1 Position correction

Using the partition and delay-based reconstruction of the event location, we can attempt to remove position-dependent variations in the reconstructed pulse amplitude, analogous to the CDMS position correction discussed in Chap. 3. For each event with partition location (x_p, y_p) and delay location (x_d, y_d), we define the distance to the jth event:

$$d(j) = \sqrt{(x_p(j) - x_p)^2 + (y_p(j) - y_p)^2 + (x_d(j) - x_d)^2/w_d + (y_d(j) - y_d)^2/w_d} \qquad (7.13)$$

where w_d is a relative weighting between the delay and partition information. Using this metric, the set of n events with the smallest values of $d(j)$ are chosen, and their reconstructed energy spectrum is fit to determine the location of lines of known energy from the calibration source, as shown in Fig. 7.13a.

Figure 7.13b shows the reconstructed energy of the primary ^{129}I peak for each event's nearest neighbors. Events near the edge of the device are seen to have systematically lower energy collection than events interacting in the center of the device, likely due to an increased loss of phonons to the detector housing near the edges. Even in the center of the device, ≈1 keV variations in the reconstructed energy with position lead to smearing of the energy resolution prior to applying the position-based correction.

To remove this position dependence, the reconstructed energy for each event can be corrected by the location of the peak for the event's nearest neighbors:

$$E_{corr} = \left(\frac{29.8 \text{ keV}}{E^{nn}_{29.8}}\right) E \qquad (7.14)$$

where E is the reconstructed energy of the event prior to position correction, $E^{nn}_{29.8}$ is the location of the primary ^{129}I peak for the event's nearest neighbors, and E_{corr} is the position-corrected energy. To perform this correction, the size of the neighborhood and relative weighting between the partition and delay quantities must be determined. Corrections were performed for a variety of parameters, with $n = 50$, 100, 200, 400, 600, 1000 and $w_d = 0.1$, 1, 10, 100, 1000. Since the delay-based position estimators are significantly noisier than the partition-based quantities at these energies, including the delay information did not improve the position correction and the best resolutions were found for $w_d \to \infty$. The number of nearest neighbors, n, must be small enough that the neighborhoods are smaller than the

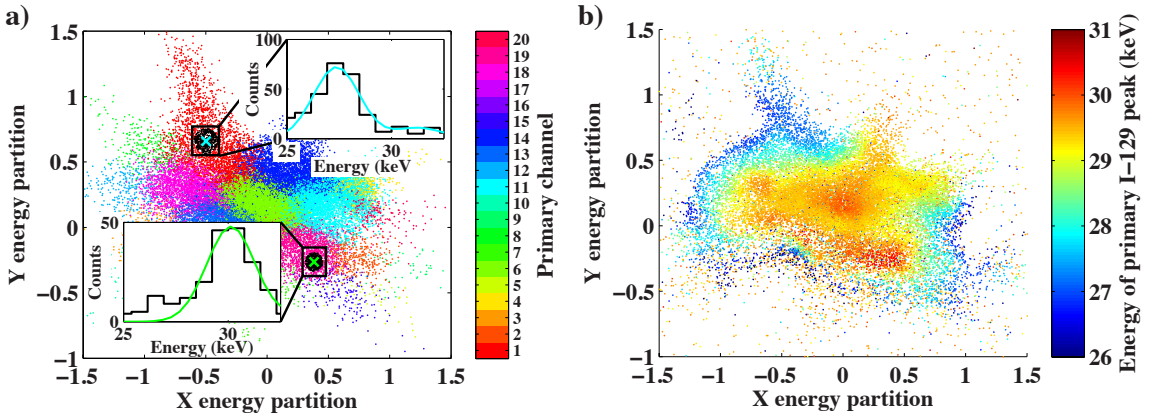


Figure 7.13: a) Example of position-dependent energy reconstruction for two different locations on the partition plot, prior to position correction. The black regions show the 400 nearest neighbor events for the example events indicated by the cyan and green crosses. The spectra for the events in the black regions are shown in the insets. For the neighborhood of the cyan event, the spectral peak from the ^{129}I source lies at a reconstructed energy of 27 keV, while for the green event, the peak lies above 30 keV. This variation produces a position-dependent smearing of the resolution when the spectrum is summed over the entire device. b) Variation in the reconstructed energy of the ^{129}I peak versus the interaction location determined from the partition-based position estimator.

length scale of variations in the reconstructed energy, but large enough that statistical errors do not dominate. It was found that $n = 400$ neighbors per 10^5 total calibration events gave the best reconstructed resolution. This minimum is rather broad, and $n = 200$–600 gave a resolution within 10% of the optimum resolution.

The results of this correction for $n = 400$ and $w_d = \infty$ are shown in Fig. 7.14. After correction, the non-Gaussian tail of poorly collected events below the 29.8 keV peak is significantly reduced. This tail primarily consists of events interacting near the edge of the detector, which have poorer phonon collection than for events interacting near the center. The position correction accounts for this by increasing the corrected energy for edge events, reducing the size of this low-energy tail. In addition, the large-scale variations in reconstructed energy for events interacting in the bulk are corrected, leading to a 30% improvement in the corrected resolution relative to the uncorrected resolution. Figure 7.14b, shows the location of the 29.8 keV peak for each event's nearest neighborhood. Comparison with 7.13b indicates that the large-scale position dependencies are being effectively removed by the position correction. However, some areas of miscorrection remain, especially near the edges of the device where the nearest neighborhoods grow physically large.

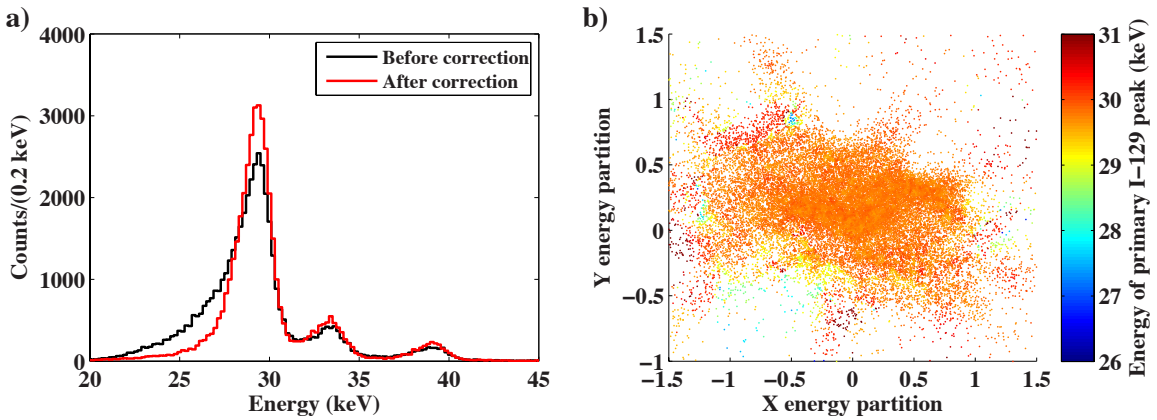


Figure 7.14: a) Comparison of the reconstructed energy spectrum for events from the ^{129}I source before and after applying the position-based correction. After correction, the low-energy non-Gaussian tail of poorly collected events is significantly reduced. In addition, the resolution for events occurring in the bulk of the detector improves by 30%, from $\sigma_E = 1.01$ keV prior to correction to $\sigma_E = 0.73$ keV after correction. b) Variation in the position-corrected energy of the ^{129}I peak versus the interaction location determined from the partition-based position estimator. Comparison with Fig. 7.13b indicates that the position correction successfully removes the majority of the large-scale systematic variations in the reconstructed energy with position.

7.4 Noise

For many applications of MKIDs, two-level systems (TLS) in dielectric layers on the resonator or substrate surfaces provide excess frequency noise, which can be orders of magnitude larger than the HEMT noise at frequencies $O(1\text{ Hz})$ [298, 304]. However, this noise typically has a $f^{-1/2}$ frequency dependence, so at the $O(1\text{ kHz})$ frequencies of interest for our pulse data, TLS system noise is less significant. In addition, the resonators considered in this work have larger geometries (i.e., 20 μm spacing between fingers in the capacitor) than resonators used for other applications, which further reduces the TLS contribution to the noise due to the smaller electric fields in the dielectric regions [298].

The noise in the phase and amplitude direction as a function of readout power for a typical resonator is shown in Fig. 7.15, both at the resonance frequency and away from the resonance. At frequencies above 10 kHz, the measured noise was found to be in good agreement with the white noise floor set by our HEMT amplifier:

$$S_{HEMT} = \frac{k_b T_N}{P_f} \tag{7.15}$$

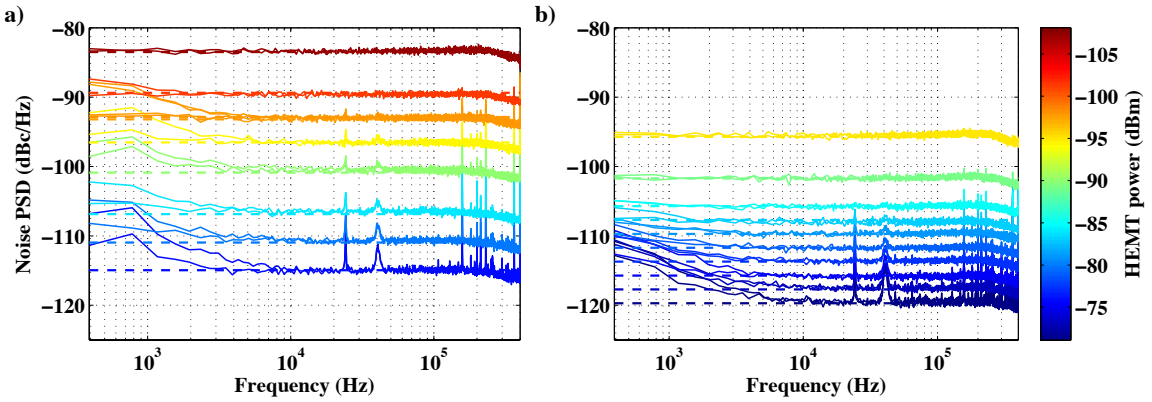


Figure 7.15: Measured phase and amplitude noise versus power at the HEMT. The noise at the resonant frequency is shown in a), while the noise several MHz outside the resonance feature is shown in b). For the on-resonance data, we account for the suppression of power reaching the HEMT relative to the off-resonance point due to the reduction in power transmitted past the device. The noise at $f \geq 10$ kHz is well described by white HEMT noise for $T_N = 5$ K. At lower frequencies, $1/f$ noise is seen both on- and off-resonance, indicating that the HEMT is contributing some additional low-frequency noise at the highest powers.

where $T_N = 5$ K is the noise temperature of the HEMT at $f \approx 3$ GHz, P_f is the feedline power reaching the HEMT, and S_{HEMT} is the single-sided noise power spectrum referenced to the carrier power, given in units of dBc/Hz.

At lower frequencies, excess $1/f$ noise is seen above the white noise floor, both for the on-resonance data and the off-resonance data. Since a portion of this noise is also seen off-resonance, it must be at least partly due to $1/f$ noise from the HEMT. Such noise, which can result, e.g., from HEMT gain fluctuations, is common to all resonators and can be removed by a more sophisticated analysis that uses off-resonance tones to monitor and remove common-mode fluctuations.

Figure 7.16 shows a detailed comparison of the on- and off-resonance noise for a typical TiN resonator from B100430.3 at a drive power just below the point at which the resonance shape becomes distorted. Empirically, this choice of power gives the best signal-to-noise in the reconstructed energy. The deviations from white noise at $f \lesssim 10$ kHz are seen to be identical in the off-resonance noise and the on-resonance amplitude noise, indicating that this excess noise is due to the HEMT. On-resonance, an additional few dB of noise are seen in the phase direction, possibly due to a small TLS contribution. However, the $\sim$2 dB additional noise in the phase direction is small compared to the increased responsivity in

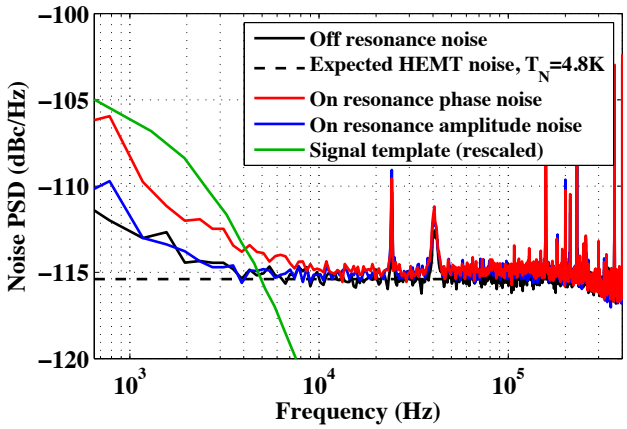


Figure 7.16: On- and off-resonance phase and amplitude noise for an example TiN resonator at the readout power used to take pulse data. Both the off-resonance (black) and on-resonance amplitude (blue) noise show a few dB excess above white noise (dashed) at frequencies ≤ 1 kHz, due to $1/f$ noise from the HEMT. The on-resonance phase noise (red) shows an additional few dB excess over the HEMT noise, possibly due to a small TLS contribution. The frequency dependence of the signal shape is overlaid (green).

the phase direction ($\beta \approx 3$), leading to improved signal-to-noise for phase readout relative to amplitude readout.

We can also compare this measured noise with the white HEMT noise that we assumed in the sensitivity calculation in Eq. 6.22. The power at the HEMT is given by Eq. 6.28, which we can numerically evaluate for the measured parameters for this device ($\tau_{qp} = 50$ μs, $\Delta = 75$ μeV, $\alpha = 0.6$, $Q_c = 10^4$) to give $S_{HEMT} = -105$ dBc/Hz, where we have conservatively assumed $\eta_{read} = 1$. In practice, $\eta_{read} < 1$, so the HEMT can be driven at somewhat higher powers and the noise decreased. Even though the calculation of the expected sensitivities in Eq. 6.22 assumes that the noise is white at all frequencies of interest, it is still conservative since this white noise floor for $\eta_{read} = 1$ is still higher than the measured excess noise at low frequencies.

7.5 Energy resolution

In the following sections we show the measured energy resolutions for several different devices using the reconstruction techniques described in Sec. 7.3.1. These data were taken while illuminating the substrate face opposite the resonators with one of 3 different sources:

- 100 nCi ^{129}I source, with x-rays from ≈30–40 keV and no collimator, leading to full

illumination of substrate.

- 10 μCi ^{109}Cd source, with ≈20 keV x-rays, and 63 and 85 keV mono-energetic internal conversion electrons collimated to a single 0.7 mm diameter hole through 8 mm of Cu. For 5 mm spacing between the collimator and substrate the spot size at the chip was 1.6 mm.

- 10 μCi ^{55}Fe source, with ≈6 keV x-rays collimated to single 0.5 mm diameter hole through 4 mm of Cu, giving a spot size of 1.8 mm at the chip.

Table 7.2 lists the expected emitted and absorbed line intensities for each source. The sources cover the ∼10–100 keV range of recoil energies of interest for WIMP scattering. However, at energies $\gtrsim$50 keV the interaction length for photons in Si is $\gg$1 mm and Compton scattering begins to dominate, so the photon energy is unlikely to be contained in the substrate. Due to the much shorter penetration depth of electrons with similar energies, the mono-energetic electrons from the ^{109}Cd source can still be used to calibrate the detectors for electron recoils from 50–100 keV.

7.5.1 $T_c = 0.5$ K TiN, B100430.3

Data were taken with the $T_c = 0.5$ K TiN device from B100430.3 with the ^{129}I, ^{109}Cd, and ^{55}Fe sources. This device had a large kinetic inductance fraction, $\alpha \approx 0.6$, as well as a T_c well below 1 K leading to high responsivity. However, as shown in Fig. 7.3, film nonuniformities across the device increased the array bandwidth, and only 14 resonators could be read out simultaneously. In addition, the resonators did not retain the same ordering in frequency as the design, leading to ambiguity in the physical location of the resonator corresponding to a given resonant frequency. Due to these complications, a position-dependent correction was not applied for the data below.

The reconstructed energy spectrum for data taken with the ^{129}I source is shown in Fig. 7.17. A rough cut is applied to remove interactions near the surface and edges by restricting the primary energy partition, $P_{prim} = E_{prim}/E_{tot} < 0.25$, where E_{prim} is the energy collected in the primary channel and E_{tot} is the total energy. Although this cut does not significantly improve the resolution of the primary peak, it does remove some of the misreconstructed events in the non-Gaussian tails of the distribution. The best-fit resolution at 30 keV is $\sigma_E = 0.95$ keV, which is about a factor of 3 larger than the measured baseline

Energy (keV):	**Emitted intensity (%):**	**Type:**	μ/ρ **(cm^2/g):**	**Absorbed intensity (%):**
^{129}I:				
29.461	20.5	x-ray, Xe $K_{\alpha 2}$	1.229	29.7
29.782	37.7	x-ray, Xe $K_{\alpha 1}$	1.187	53.1
33.562	3.5	x-ray, Xe $K_{\beta 3}$	0.816	3.5
33.624	6.7	x-ray, Xe $K_{\beta 1}$	0.811	6.7
34.419	2.0	x-ray, Xe $K_{\beta 2}$	0.754	1.9
34.496	0.4	x-ray, Xe $K_{\beta 4}$	0.748	0.4
39.58	7.5	γ	0.485	4.7
^{109}Cd:				
21.99	29.5	x-ray, Ag $K_{\alpha 2}$	3.052	30.6
22.163	55.7	x-ray, Ag $K_{\alpha 1}$	2.979	56.8
24.912	4.76	x-ray, Ag $K_{\beta 3}$	2.073	3.7
24.943	9.2	x-ray, Ag $K_{\beta 1}$	2.065	7.2
25.455	2.3	x-ray, Ag $K_{\beta 2}$	1.939	1.7
88.04	3.61	γ	0.038	0.1
62.5	40.8	e^-, IC_K		20.4
84.7	44.8	e^-, IC_L		22.4
88.0	9.3	e^-, IC_M		4.7
^{55}Fe:				
5.89	16.9	x-ray, Mn K_{α}	153.0	89.7
6.49	2.91	x-ray, Mn K_{β}	116.6	10.3

Table 7.2: Primary x-ray, γ and e^- emissions for the ^{129}I, ^{109}Cd and ^{55}Fe sources. The emission types, energies and emitted intensities are from [263]. To calculate the absorbed intensities for the x-ray and γ lines, the mass-attenuation coefficients, μ/ρ, for photoelectric absorption in Si from [323] are used. The absorbed intensity is then given by $I = I_0(1 - \exp[-(\mu/\rho)\rho l])$, where I_0 is the incident intensity, $\rho = 2.33$ g/cm^3 is the density of Si and $l = 1$ mm is the substrate thickness. The emitted intensities give the fraction of photons or electrons per source decay, while the absorbed intensities are normalized to give the expected fraction of interactions in the substrate from each emission type.

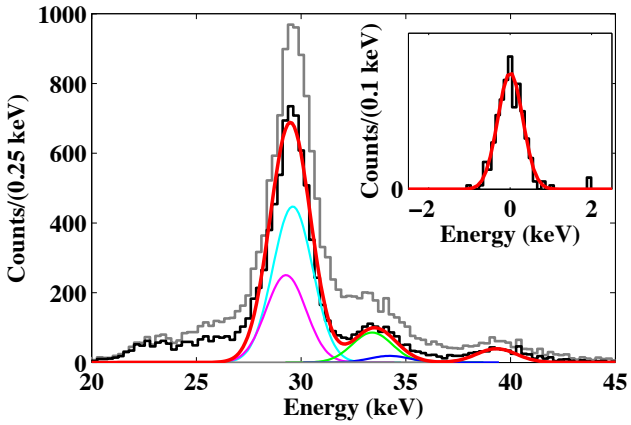


Figure 7.17: Reconstructed energy spectrum for the $T_c = 0.5$ K TiN device from B100430.3 illuminated by the ^{129}I source. The light histogram shows all data, while the dark histogram removes surface events by requiring the primary channel energy partition, $P_{prim} < 0.25$. Due to the ambiguity in the position reconstruction for this device, no cut to remove events near the substrate edges is applied. The data are fit to a model consisting of the expected lines at 29.5 (magenta), 29.8 (cyan), 33.6 (green), 34.4 (blue), and 39.6 (orange) keV, where only a single mean, amplitude and resolution are fit, and the relative line positions and amplitudes are fixed to the known values from Table 7.2. The best-fit resolution is $\sigma_E = 0.95$ keV at 30 keV. The inset shows the reconstructed energy after applying the same optimal filter pulse estimation to randomly triggered noise traces. The distribution is fit to a Gaussian to determine the baseline resolution of $\sigma_E = 0.30$ keV.

resolution. This indicates that the resolution for this device was dominated by position-dependent variations in the energy collection or reconstruction.

Since position reconstruction and correction was problematic for this device, we retested the same device with collimated ^{55}Fe and ^{109}Cd sources to attempt to remove any position variation in the reconstructed energy. The results of these calibrations are shown in Fig. 7.18. The 6 keV x-rays from the ^{55}Fe source were collimated to a <2 mm diameter spot on the back of the chip. In addition, the characteristic absorption length for 6 keV x-rays in Si is 30 μm, so most events interact very close to the substrate face opposite the resonators. As shown in Fig. 7.18a, the measured energy resolution from the 6 keV peak is $\sigma = 0.27$ keV, consistent with the baseline resolution of 0.3 keV for this device. Although not separately resolved, the K_β peak is visible as an excess of events in the high energy tail. These results indicate that position dependence was indeed dominating the resolution for the ^{129}I calibration, and for sufficiently well-collimated, low-energy events interacting near the center of the back surface of the substrate, the baseline resolution can be recovered.

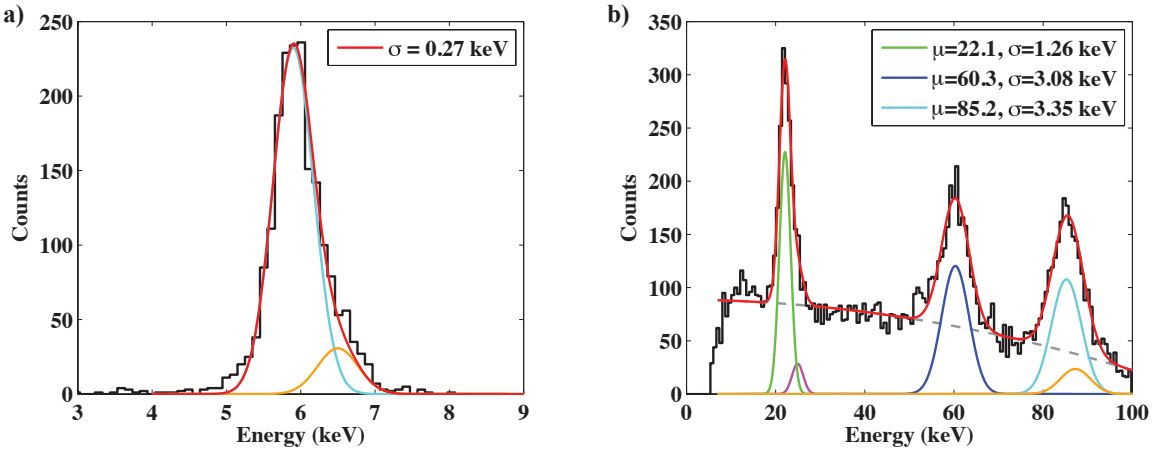


Figure 7.18: a) Reconstructed energy spectrum for the $T_c = 0.5$ K TiN device from B100430.3 illuminated by the collimated ^{55}Fe source. No position-dependent correction or cuts are applied. The data are fit to the expected K_α (cyan) and K_β (orange) lines with only a single free amplitude, mean and resolution. The relative position and amplitude of the lines are constrained using the values from Table 7.2. The measured resolution of $\sigma = 0.27$ keV is consistent with the baseline resolution for this device. b) Reconstructed energy spectrum for the $T_c = 0.5$ K TiN device from B100430.3 illuminated by the collimated ^{109}Cd source. The K_α (green) and K_β (magenta) lines, as well as the mono-energetic e^- lines at 63 (blue), 85 (cyan), and 88 keV (orange) are observed. A free quadratic background is included to account for backscattered electrons that do not deposit their full energy in the substrate. Significant position dependence is seen, with the resolution degrading from $\sigma = 1.3$ keV at 22 keV to $\sigma = 3.4$ keV at 85 keV. The baseline resolution was also degraded to $\sigma = 0.55$ keV due to an excess steady-state quasiparticle population generated by the source.

To gauge the response of the device at higher energies, where e^- must be used due to the substrate thickness, the same device was calibrated with x-rays and mono-energetic electrons from the ^{109}Cd source, with the measured spectrum shown in Figure 7.18b. As discussed in Sec. 7.2, illuminating the devices with the ^{109}Cd source was found to introduce an excess steady-state quasiparticle population that degraded the internal Q relative to illumination with the weaker ^{129}I or ^{55}Fe sources. This increased quasiparticle population degraded the resonators sufficiently that only 8 of the original 14 resonators could be read out simultaneously, leading to increased position dependence of the response. Even though the source was collimated to a < 2 mm diameter spot, some position dependence remained, and poorer resolution was observed for the higher-energy e^- lines relative to the x-ray line at 20 keV. In addition, the baseline resolution for these data was degraded to $\sigma = 0.55$ keV due to the reduced Q_i. After accounting for the smaller number of resonators used, this is

a factor of 2.5 worse than the baseline resolution without the ^{109}Cd source in place.

7.5.2 $T_c = 0.65$ K NbTiN, B101027.1

Following on the success of the sub-keV baseline energy resolution measured for the $T_c = 0.5$ K TiN device, additional devices made of high-resistivity materials were fabricated, from both TiN and NbTiN films with a range of T_c. The primary aim was to improve uniformity, allowing all 20 resonators to fit within the readout bandwidth and maintaining the frequency ordering for most resonators, so that their physical position could be determined. While both TiN and NbTiN films with $T_c > 2$ K showed the desired uniformity, due to the larger gap, lower kinetic inductance and shorter quasiparticle lifetime, these films did not have sufficient sensitivity to resolve 30 keV x-rays.

The $T_c = 0.65$ K NbTiN device from B1011027.1 did show improved frequency spacing, with all 20 resonators fitting within the readout bandwidth, in addition to high responsivity. Two resonators had anomalously low Q_i and did not show evidence of pulses, so only 18 resonators were read out simultaneously. Even though nearly all resonators could be read out, there still was significant variation in the ordering of the frequencies relative to the design. As shown in Fig. 7.19, this leads to poor reconstruction of the interaction location.

To determine the physical location of the resonators, the correlation matrix, R, between the pulses for each resonator was calculated using bulk events from the 30 keV peak, where:

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}, \quad C_{ij} = E\left[(\phi_i - \mu_i)(\phi_j - \mu_j)\right] \tag{7.16}$$

where E denotes the expectation value (calculated as the average over all events), ϕ_i is the energy in the ith resonator after rescaling by the relative responsivity and μ_i is the mean of ϕ_i over all events. From this correlation matrix, we choose each resonator's n nearest neighbors as the resonators corresponding to the n largest values in each column of R_{ij}. We then minimize the function:

$$f(\{i\}) = \sum_{n,i} \sqrt{(x_i(\{i\}) - x_n(\{i\}))^2 + (y_i(\{i\}) - y_n(\{i\}))^2} \tag{7.17}$$

where for a given permutation of the resonators on the device, $\{i\}$, that maps the resonators from their physical ordering to their ordering in frequency, (x_i,y_i) is the physical location

of the ith resonator and (x_n,y_n) is the physical location of its nth nearest neighbor from the correlation matrix.

This function is minimized over the set of all permutations $\{i\}$. For the current device where 18 out of 20 resonators were read out, there are $20!/2! \approx 10^{18}$ different possibilities for $\{i\}$, indicating that we cannot directly test all permutations. Instead, we use the genetic optimization algorithm from Matlab's global optimization toolbox [324] to minimize the function above.

The genetic algorithm generates an initial population consisting of sets of permutations $\{i\}$ based on an initial guess of the correct ordering. It then calculates the "fitness" of each member of the population using the function $f(\{i\})$ above. The algorithm then generates a new population of permutations by applying small modifications to the most "fit" members of the population and eliminating members with poor fitness. Two types of modifications are allowed, chosen at random with equal probability for each member of the parent population:

1. "Mutations" where two elements of the parent permutation are chosen at random, and their locations are interchanged.

2. "Insertions" where unused elements that are not contained in the parent permutation are inserted at random locations, shifting the remaining blocks but leaving their ordering intact. Any extra elements are then removed from a random location.

This procedure of generating new populations of permutations and selecting the most fit members to generate the next generation is repeated until the algorithm converges to a solution, after which repeated generations do not further improve the maximum fitness. This algorithm was tested on randomized data with a known ordering and found to reliably converge to the correct permutation of resonators, provided the initial guess specified $\geq 75\%$ of the resonators in their correct position. For initial conditions far from the correct solution, the algorithm does not generally converge to the proper ordering.

Using this optimization, the physical location of each resonator for this device was determined from correlations between resonators in the ^{129}I data. Figure 7.19 shows the reconstructed partition locations before and after optimizing the ordering. Although mapping the resonators from frequency to physical location assuming the design resonator ordering produces a poor reconstruction of the event location, after reordering based on the correlation data, significant improvement in the uniformity of the partition plot is observed.

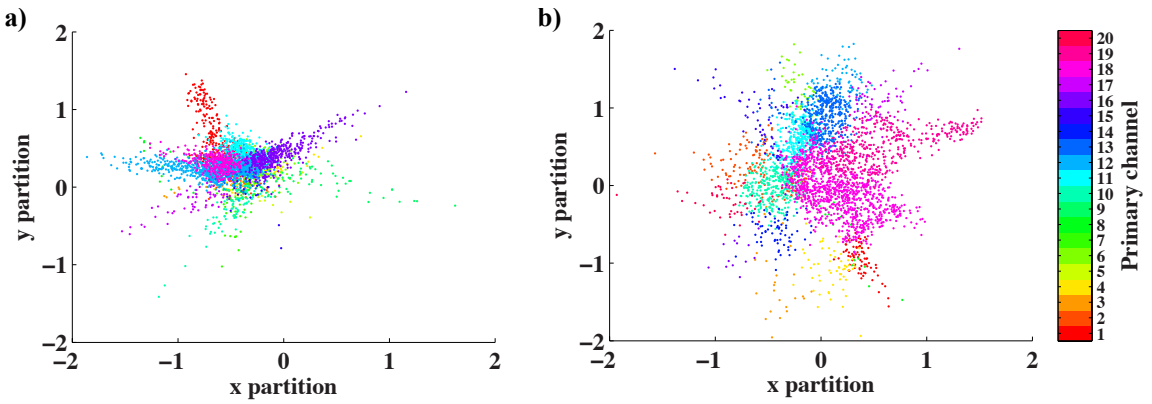


Figure 7.19: Comparison of the interaction location reconstructed from the energy partition before (a) and after (b) reordering the resonator locations based on correlations in the pulse height between resonator pairs. This optimization procedure allows the frequency ordering to be untangled, significantly reducing the degeneracies in the reconstructed partition location.

Figure 7.20 shows the reconstructed energy spectrum for this device when illuminated by the ^{129}I source. After applying a position-dependent cut to remove events interacting near the edge or surfaces of the substrate, the reconstructed resolution is $\sigma = 0.99$ keV at 30 keV. The baseline energy resolution of $\sigma = 0.51$ keV was roughly a factor of 2 better, indicating that this resolution is still dominated by position-dependent effects. In particular, one of the resonators that could not be read out was located near the center of the device, and events interacting near this resonator are not removed by the position-dependent cut described above. Although the overall resolution is comparable to that measured for the TiN device in Sec. 7.5.1, the non-Gaussian tail of poorly collected events at low energy seen in Fig. 7.17 is significantly reduced due to the increased number of channels that were read out.

7.5.3 25 nm Al, B111007

In parallel with the devices using nitride films, devices using thin Al films were also fabricated to check the dependence of the pulse fall times and phonon collection using a material with which devices with uniform properties could be reliably fabricated. Although the kinetic inductance fraction of these films was only $\alpha \approx 0.1$ even for 25 nm films, the uniformity of these devices was significantly better than for the nitrides. In addition, due to the improved energy collection described in Sec. 7.5.4, energy resolutions comparable to the nitride

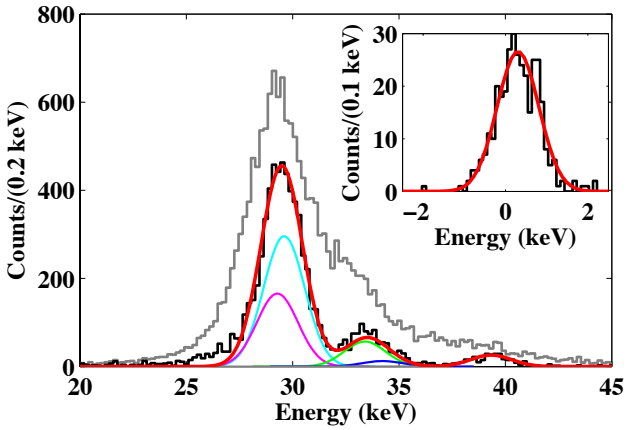


Figure 7.20: Reconstructed energy spectrum for the $T_c = 0.65$ K NbTiN device from B101027.1 illuminated by the ^{129}I source. The light histogram shows all data, while the dark histogram removes surface and edge events with a position-dependent cut. As before, the data are fit to the expected spectral lines from Table 7.2 with a best-fit resolution is $\sigma_E = 0.99$ keV at 30 keV. The baseline resolution for this device was $\sigma_E = 0.51$ keV, as shown in the inset.

devices were obtained for sufficiently thin films.

A large calibration data set consisting of 10^5 events from the ^{129}I source was obtained for the 25 nm thick Al device from B111007. All 20 resonators for this device had $Q_i > 2 \times 10^5$, showed phonon-mediated pulses from the source and could be read out simultaneously within the readout bandwidth. Mapping the resonator frequencies to physical locations was straightforward since nearly all resonators retained the design ordering in frequency. Applying the procedure described in Sec. 7.5.2 to map the physical location of the resonators based on correlations in the pulse response indicated that only resonators 16 and 17 had flipped their frequency ordering. The partition and delay plots for this device were shown in Fig. 7.11 and 7.12, respectively.

After applying the position correction procedure described in Sec. 7.3.2.1, the resulting reconstructed energy spectrum for this device is shown in Fig. 7.21. The reconstructed energy resolution for events occurring in the bulk of the device was $\sigma = 0.54$ keV at 30 keV, within 40% of the baseline resolution of $\sigma = 0.38$ keV. The improved position correction and reconstruction for this device relative to the nitride devices discussed above lead to nearly a factor of 2 improvement in the reconstructed energy resolution for bulk events. This resolution is now approaching the baseline resolution of the device, so to significantly improve the resolution further will require improving the responsivity or phonon collection

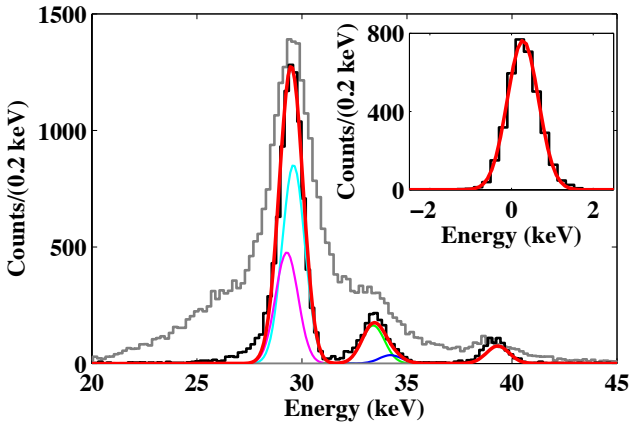


Figure 7.21: Reconstructed energy spectrum for the 25 nm Al device from B111007 illuminated by the ^{129}I source. The light histogram shows all data, while the dark histogram removes surface and edge events with a position-dependent cut and applies a position-dependent correction. As before, the data are fit to the expected spectral lines from Table 7.2 with a best-fit resolution $\sigma_E = 0.54$ keV at 30 keV. The baseline resolution for this device was $\sigma_E = 0.38$ keV, as shown in the inset.

of the device.

7.5.4 Energy collection

By measuring the frequency and dissipation response of the resonators versus temperature, and comparing to the observed response from phonon-mediated particle interactions, we can determine the overall efficiency for converting phonon energy created by the interaction into quasiparticles in the sensitive sections of the resonators. In the low-frequency ($\hbar\omega \ll \Delta$) and low-temperature ($k_bT \ll \Delta$) limit, the quasiparticle density as a function of temperature is [298]:

$$n_{qp} = 2N_0\sqrt{2\pi k_b T\Delta}\; e^{-\frac{\Delta}{k_bT}} \tag{7.18}$$

Figure 7.22 shows the frequency and dissipation response as a function of temperature for the 25 nm Al device. These data were first fit to the Mattis-Bardeen temperature dependence from Eqs. 6.6–6.9 and 6.19–6.20 to determine $\alpha = 0.074 \pm 0.005$ and $\Delta = 204.4 \pm 2.1$ μeV. Using Eq. 7.18, the frequency response versus temperature can be converted to a frequency response versus n_{qp} as shown in the inset to Fig. 7.22a.

We can convert the measured frequency response to a phase response versus number of quasiparticles using Eq. 6.17. Given the measured phase shift versus number of quasipar-

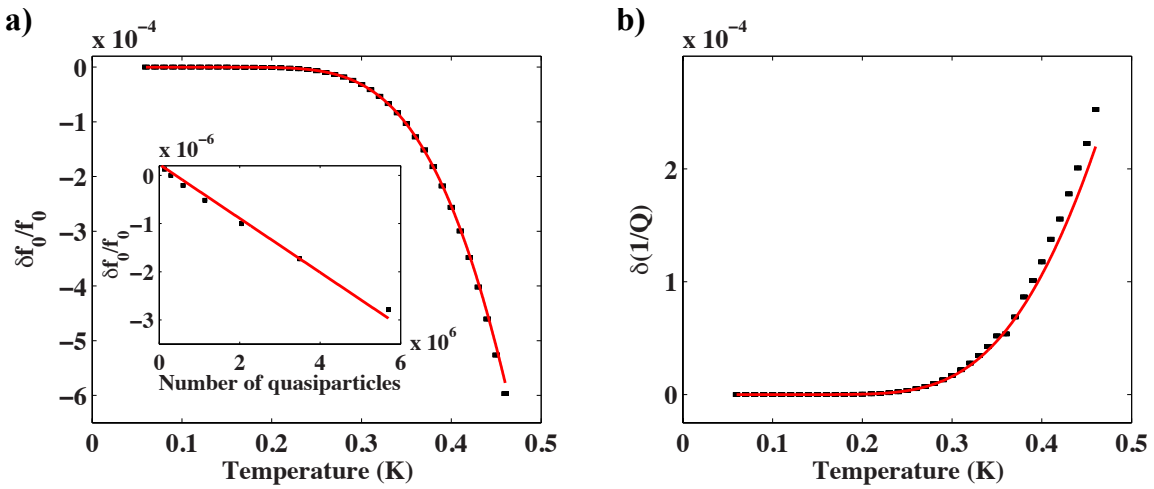


Figure 7.22: a) Fractional frequency shift, $\delta f_0/f_0$, versus temperature for a typical resonator from the 25 nm Al device. The frequency and dissipation data are simultaneously fit to Mattis-Bardeen theory (red) to determine $\alpha = 0.074 \pm 0.005$ and $\Delta = 204.4 \pm 2.1$ μeV. The inset shows $\delta f_0/f_0$ after converting the horizontal axis to the number of thermal quasiparticles at each temperature. For small phase shifts, the expected linear dependence is observed. b) Change in dissipation versus temperature for the same device.

ticles for each resonator, $\delta\theta/\delta n_{qp}$, we can divide the phase shift in each resonator for each 30 keV event from the calibration source by $\delta\theta/\delta n_{qp}$ to determine the number of quasiparticles required in each resonator to produce the observed response. Summing over resonators gives the total number of quasiparticles created for each event. This distribution is fit to a Gaussian to determine the mean number of quasiparticles, N_{qp}, and the phonon collection efficiency is calculated as $\eta_{ph} = N_{qp}\Delta/(29.8 \text{ keV})$. For the 25 nm Al device, we find $\eta_{ph} = 0.070 \pm 0.011$. The same procedure was repeated for 75 nm and 15 nm thick Al devices, and a comparison of the measured collection efficiencies shown in Fig. 7.23.

As shown in Fig. 7.23, there is an approximately linear increase in the collection efficiency with film thickness, as expected in the limit that the film thickness $t \ll \lambda_{pb}$, where $\lambda_{pb} \approx 1$ μm is the characteristic pair-breaking length in Al. As the films are made thinner, the volume is also reduced linearly, so the quasiparticle density remains approximately unchanged. In addition, α increases with decreasing thickness, giving higher responsivity for the same quasiparticle density and improving the signal-to-noise of the device. However, thinner films also saturate at lower powers, decreasing the maximum readout power that can be used and increasing the amplifier noise contribution. The quasiparticle lifetime was also found to decrease from $\tau_{qp} = 73 \pm 7$ μs for the 75 nm thick device to $\tau_{qp} = 15 \pm 4$ μs

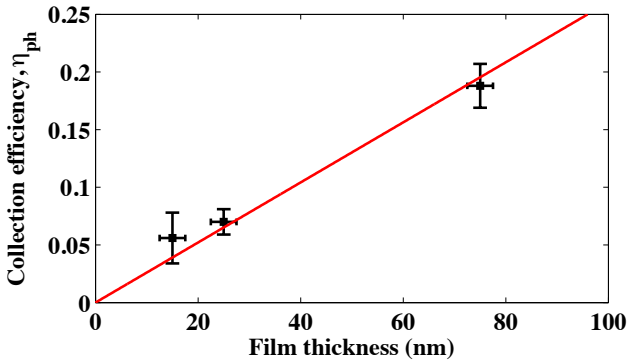


Figure 7.23: Measured phonon collection efficiency, η_{ph}, as a function of film thickness for the Al devices. An approximately linear increase in collection is observed as the film thickness is increased.

for the 15 nm thick device. Taken together, these tradeoffs led to a factor of 3 improvement in the baseline resolution of the 25 nm device relative to the 75 nm device, but no further improvement in the resolution when decreasing the film thickness to 15 nm. From these tests, the Al film thickness that maximizes signal-to-noise was found to be ≈15–25 nm.

The frequency and dissipation response versus temperature for the $T_c = 0.65$ K device is shown in Fig. 7.24. Unlike the Al data, the shape of the frequency response versus temperature is not well described by Mattis-Bardeen theory. Instead, an approximately linear dependence for $\delta f_0/f_0$ is observed at temperatures $T \ll \Delta$. Such a dependence appears to be a generic property of resonators fabricated from TiN and NbTiN films. A similar response was observed for the $T_c = 0.5$ K TiN device as well as by Barends et al. for $T_c = 16$ K NbTiN films [325]. Improved fits to the data can be obtained after introducing a broadening of the gap, Γ, in the density of states following Barends et al., $\Delta \rightarrow \Delta - j\Gamma$.

This gap broadening smears the divergence at the gap in the density of states and introduces a distribution of quasiparticle states below the gap edge. The presence of these sub-gap states increases the number of thermally excited quasiparticles at low temperature, which could account for the observed linear frequency dependence at $T \ll \Delta$. The number of quasiparticles can be calculated from [298]:

$$n_{qp} = \int f(E)DOS(E)dE = 4N_0 \int_0^\infty \frac{1}{1+e^{\frac{E}{k_b T}}} \mathrm{Re}\left[\frac{E}{\sqrt{E^2-(\Delta-j\Gamma)^2}}\right] dE \qquad (7.19)$$

where $DOS(E)$ is the broadened density of states, $f(E)$ is the Fermi function and the

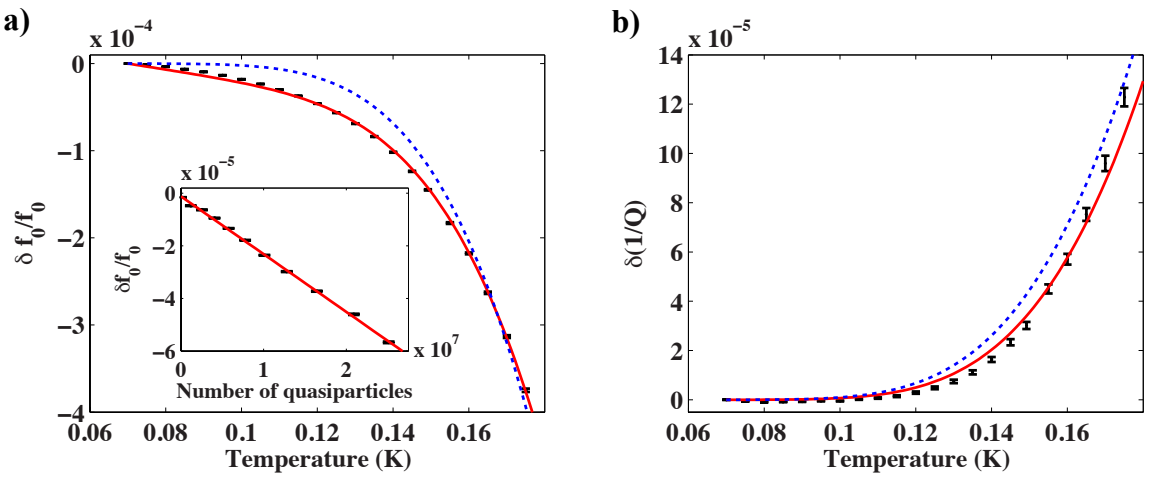


Figure 7.24: a) Fractional frequency shift, $\delta f_0/f_0$, versus temperature for an example resonator from the $T_c = 0.65$ K NbTiN device. Unlike the Al data, these data are not well described by standard Mattis-Bardeen theory (blue, dotted), but improved fits (red, solid) can be obtained by including as smearing $\Gamma \approx 10^{-3}\Delta$ in the density of states, following [325]. This smearing is also included in the conversion of the temperature to quasiparticle number and the expected linear dependence of $\delta f_0/f_0$ vs. N_{qp} is observed, as shown in the inset. b) Change in dissipation versus temperature for the same device.

integral must be computed numerically. To determine the broadening parameter, the change in frequency and dissipation were fit to Eqs. 6.3 and 6.4, which were numerically integrated after including the broadening in the density of states as above.

For the NbTiN device, there is significant resonator-to-resonator variation, so the fits are performed separately for each resonance. The best-fit broadening for this device varies from $\Gamma/\Delta \approx (0.5\text{–}2)\times10^{-3}$, while $\alpha \approx 0.7\text{–}0.75$ and $\Delta \approx 90\text{–}105$ μeV. Using Eq. 7.19, the phase response versus number of quasiparticles for each resonator was calculated, and compared with the response of the device to phonon-mediated 30 keV x-rays. After summing over resonators, the overall collection efficiency for the NbTiN device was calculated to be $\eta_{ph} = (9.2 \pm 3.0)\times10^{-3}$, which is more than an order of magnitude lower than for the Al devices of the same thickness. The same calculation was performed for the $T_c = 0.5$ K TiN device, again incorporating the gap broadening in the density of states. The TiN device gave a comparable collection efficiency as for the NbTiN device, with $\eta_{ph} = (6.5 \pm 3.1)\times10^{-3}$. In both cases, a correction was applied for the number of resonators that were simultaneously read out for the 30 keV calibration data, assuming approximately uniform illumination of all resonators for bulk events.

	T_c=0.5K TiN:	T_c=0.65K NbTiN:	25 nm Al:
α	0.6–0.7	0.7–0.75	0.074 ± 0.005
Δ (μeV)	70–85	90–105	204.4 ± 2.1
η_{ph}	$(6.5 \pm 3.1)\times10^{-3}$	$(9.2 \pm 3.0)\times10^{-3}$	0.070 ± 0.011
f_0 (GHz)	2.8–3.6	2.25–2.45	4.85–5.1
τ_{qp} (μs)	80–100	70–80	18 ± 4
$S_1(\omega, T)$	1.0	1.2	0.9
$\beta(\omega, T)$	2.3	2.3	3.1
σ_{exp} (keV)	0.25	0.26	0.39
σ_{meas} (keV)	0.30	0.51	0.38

Table 7.3: Measured device parameters for the TiN, NbTiN, and Al devices. For the TiN and NbTiN devices, a range of values is given indicating the resonator-to-resonator variation. The last two rows compare the expected resolution, σ_{exp}, from Eq. 6.38 to the measured baseline resolution, σ_{meas}, from calibration of each device with the ^{129}I source.

7.5.5 Comparison to expected resolution

Using the measured values of the resonator parameters, noise and phonon collection from above, we can compare the measured baseline resolution of 0.3–0.6 keV with the expected resolution given by Eq. 6.38. Table 7.3 lists the relevant parameters for each device. As in Sec. 6.4.2 we use $N_0 = 1.72\times10^{10}$ μm^{-3} eV^{-1} for Al and $N_0 = 8.7\times10^{9}$ μm^{-3} eV^{-1} for the nitrides. All devices have substrate area $A_{sub} = 8$ cm^2 and were read out with the same HEMT with $T_N = 5$ K, as shown in Sec. 7.4. It is difficult to directly measure η_{read}, p_t, or λ_{pb}, but as shown in Table 7.3, for $\eta_{read}/p_t \approx 1$ and $\lambda_{pb} \approx 1$ μm, the expected resolution from Eq. 6.38 agrees well with the measured baseline resolution for the 3 devices.

The agreement between the sensitivity calculation and the measured resolutions for the TiN, NbTiN and Al devices further supports the conclusion in Sec. 7.5.4 that <1% of the incident phonon energy is being collected in the inductive portions of the resonators for the TiN and NbTiN devices. Even though the nitride devices have significantly larger responsivity for the same energy deposition, this increased responsivity is canceled by lower energy collection and the final baseline resolution is comparable to the resolution for the 25 nm Al device.

This calculation also indicates how the sensitivity can be improved. Increasing the phonon collection by limiting losses of phonon energy to the detector housing should give a linear improvement in the energy resolution. As shown in Fig. 7.13, the poorer energy collection for events interacting near the edge of the substrate indicates that the loss of phonons

to the box can be significant. Future devices will be patterned on larger substrates with clamping mechanisms similar to those used in CDMS. With improved mounting, $\eta_{ph} \approx 0.3$, as observed in the iZIP devices, should be attainable. Lower gap, higher-resistivity materials can also improve the resolution provided that good phonon collection can be maintained. Although the TiN and NbTiN films tested to date appear to have poor collection, investigation of other high-resistivity materials may provide a more sensitive alternative to Al, provided high-quality, uniform resonators can be produced.

Chapter 8

Conclusions and future work

8.1 Low-mass WIMPs

Over the past several years, DAMA/LIBRA, CoGeNT, and CRESST-II have reported what may be the first direct evidence for WIMP interactions in terrestrial detectors [154, 187, 189]. In addition, there is evidence for excess γ-ray and synchrotron emission from the galactic center, which could be explained by annihilations from a WIMP with the same properties [219]. Given these results, there is significant excitement that we could be on the verge of discovering the particle nature of dark matter after more than 20 years of searching for such interactions, providing the first direct detection of the dominant constituent of the matter in the universe. If true, the next generation of detectors would be able to precisely identify the WIMP mass and scattering cross section, and begin to study the distribution of WIMPs in the vicinity of earth in detail, opening up a new field of "WIMP astronomy."

Nonetheless, given the significance of such a discovery, we must be certain that the signals being reported are due to WIMPs and not due to more mundane backgrounds that could mimic WIMP interactions. The reported excess events occur near the energy threshold of each of the experiments, where it is most difficult to understand the detector response and backgrounds. This difficulty is evidenced, for example, by the uncertainty in the surface event background for CoGeNT, for which initial fits to the observed spectrum constrained to be <30% of the low energy excess [193], while more recent estimates indicate the surface event contamination is closer to 75% [191, 192]. Given the low recoil energies of the candidate events, the experiments are also only sensitive to WIMPs in the high-velocity tail of the distribution, where there could be significant deviations from the simplified halo model typically assumed, complicating comparisons between experiments.

At the same time, there is not detailed quantitative agreement between the potential signals in each experiment, with the best-fit cross section varying by roughly an order of magnitude between the DAMA signal and the CoGeNT signal after subtracting the recent estimates of the surface event background (with CRESST-II suggesting a region directly in between). If the modulation reported by CoGeNT is confirmed with higher statistics data, and if the current surface event background estimate is correct, it would imply a modulation fraction of 50–100%, which is nearly an order of magnitude larger than would be expected given the earth's rotational velocity around the sun.

In addition, the results presented in this thesis disfavor a low-mass WIMP explanation for DAMA/LIBRA, CRESST-II, and the entire CoGeNT excess, for spin-independent elastic scattering of low-mass WIMPs, under standard assumptions. Compatibility between the CDMS results presented here and CoGeNT is possible if only $\sim$25% of the low-energy excess in CoGeNT is due to WIMPs [191,192]. Although the analysis presented here cannot test such a signal, a likelihood analysis accounting for the backgrounds could improve the sensitivity of CDMS II to such a signal if the backgrounds can be modeled with sufficient precision. However, such a scenario is disfavored by the low-energy rate in the XENON10 experiment [156], unless there are significant uncertainties in the detector response [195] or the WIMP scattering cross section is suppressed in Xe relative to Ge [180].

In general, given the uncertainties in the properties of the dark matter halo and the scattering cross section, there are models in which compatibility between the positive signals and null results can be obtained for experiments using different target nuclei if the "standard assumptions" are relaxed. The results in this thesis provide a more model independent check of the CoGeNT excess than liquid Xe experiments, since both CDMS and CoGeNT use Ge as the target nucleus. The key challenge to increasing the sensitivity of CDMS II to low-mass WIMPs is to understand the low-energy backgrounds in detail. CDMS has several additional handles on low-energy backgrounds which CoGeNT does not have, including the distribution of backgrounds in ionization yield and the distribution of multiple scatter events, for which any WIMP signal would be absent. In contrast, CoGeNT has the ability to remove surface event backgrounds, but as in CDMS, this timing-based rejection is less effective in the energy range of interest for low-mass WIMPs. Even with the additional information provided by the CDMS detector technology, detailed modeling of the backgrounds at low-energy is challenging.

There are several potential techniques which could be used by future analyses to improve the understanding of the low-energy backgrounds in CDMS. The dominant low-yield electron-recoil backgrounds are zero-charge events and surface events, both of which show up in ^{133}Ba calibration data. Thus, it may be possible to characterize these backgrounds in calibration data where no WIMP signal can be present and to use this knowledge to parameterize the energy and ionization-yield dependence of backgrounds in the WIMP search data. However, there are known systematic differences in the energy distribution and total rate (relative to the bulk electron-recoil rate) of the surface-event and zero-charge backgrounds between calibration and WIMP search data [153, 230]. These differences must be understood in detail to successfully use calibration data to constrain backgrounds in WIMP search data.

A technique with potentially fewer systematics (although at the cost of lower statistics) is to use the distribution of multiple-scatter events in the WIMP search data to constrain the corresponding distribution of single-scatter electron recoils, many of which arise from the same sources. For both the zero-charge and surface-event distributions, we expect that the singles and multiples have the same origin, but for singles the additional scatters are lost to the detector housing or other uninstrumented material. Using the multiples, it may be possible to construct a background model (or at least parameterize the recoil and ionization energy dependence) that can then be applied directly to the single-scatter data. A likelihood analysis incorporating these backgrounds could then be performed to improve the sensitivity to a small WIMP component that may be present in addition to backgrounds, as would be predicted by the revised estimation of the surface event background in CoGeNT. However, the identification of a WIMP component in such an analysis would require a large effective background subtraction, requiring the correspondence between the energy spectrum and background rate in the singles and multiples to be carefully understood.

Given the difficulty of understanding the low-energy backgrounds in detail, it is preferable to develop detector technologies with improved background rejection at low energy to avoid the need for detailed background modeling. A number of techniques are being pursued to improve the sensitivity of CDMS detectors at low energy, largely through the work of J. Hall, M. Pyle, and R. Basu Thakur:

- **Lower backgrounds with iZIP detectors and reduced contamination:**

Figure 6.2a shows the phonon sensor design for the iZIP detectors currently installed at the SUL. These detectors include a phonon guard ring which allows fiducialization from the partitioning of energy between the inner and outer phonon sensors, in addition to the charge based fiducialization used in CDMS II. Thus, phonon fiducialization should allow the elimination of the majority of "zero-charge" events, which were consistent with high-radius events leaking past the charge-based fiducial volume cut due to suppressed ionization collection. Using iZIP data from the UC Berkeley detector test facility, measurements of the phonon-based rejection indicate that a leakage fraction of $\lesssim 10^{-2}$ for high-radius events should be possible from 2–10 keV, while maintaining >50% nuclear recoil acceptance [326]. If zero-charge events can be eliminated at this level, then leakage from bulk electron recoils and low-yield surface events are expected to dominate the backgrounds in the nuclear-recoil signal region. CDMS is currently working to reduce contamination of the detectors, housings, and icebox for the next-generation experiment at SNOLAB, to minimize radiogenic neutrons (which scale with the U/Th contamination that also provides the dominant γ rate) as well as surface events due to Rn plate out during detector and hardware fabrication. Reduced contamination is necessary to maintain sufficiently low backgrounds in WIMP searches at high mass and will also help to reduce low-energy backgrounds where rejection is less effective.

- **Achieving lower thresholds through Neganov-Luke amplification:**
  The best CDMS II Ge detectors had baseline energy resolutions, $\sigma_{pt} \approx 125$ eV for the total phonon signal, corresponding to trigger thresholds ∼1.5 keV (due to time dependent noise, the trigger thresholds in CDMS II were set at ∼10–15σ, although in an optimized low-threshold experiment they could likely be lowered to 5–6σ). As discussed in Sec. 3.1.1, a portion of the phonon signal arises from Neganov-Luke phonons whose total energy is equal to the work done by the electric field to drift the charges across the detector [238, 239]. By operating the detectors at high voltage biases, a large gain on the ionization signal can be obtained by measuring the ionization through the corresponding Neganov-Luke phonons, allowing much lower energy thresholds. Energy thresholds as low as 0.08 keV$_{ee}$ (corresponding to 0.4–0.8 keV$_{nr}$ for $q_{Ge} = 0.1$–0.2) have been demonstrated using existing CDMS II detectors with

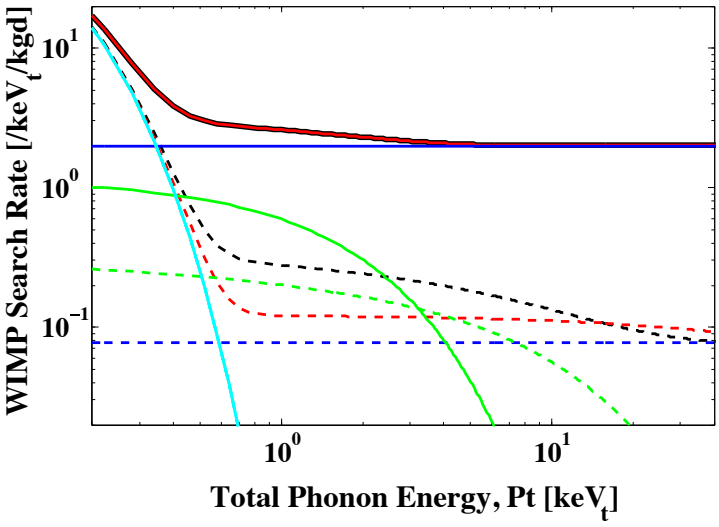


Figure 8.1: Demonstration of statistical background rejection of electron recoils possible from operating at multiple biases. For $V_b = 0$ V (solid lines), a total spectrum (black, solid) consisting of a nuclear-recoil signal (green, solid) plus a larger constant electron-recoil background (blue, solid) cannot be distinguished from an electron-recoil background that increases exponentially at low energy (red). However, at $V_b = 75$ V (dashed lines), the same total spectrum (black, dashed) containing nuclear recoils (green dashed) in addition to a constant electron-recoil background (blue, dashed) is easily distinguishable from the spectrum containing only electron recoils (red, dashed) due to the different Neganov-Luke gains for the two recoil types. For all the spectra, the contribution from noise events is included (cyan). *Figure from Pyle et al., [241]*

this technique.

- **Statistical discrimination of backgrounds through multiple bias operation:** The drawback to high-voltage operation is that only the ionization signal is measured since the initial recoil phonon signal is dominated by the secondary Neganov-Luke phonons. Thus, rejection of electron-recoil backgrounds on an event-by-event basis is no longer possible. Nonetheless, by operating at multiple biases, statistical rejection of backgrounds may still be possible [241]. In the limit of $V_b = 0$ V, only the phonon signal is measured, so the spectrum would be a sum of electron recoils and nuclear recoils given the total recoil energy for each. At high bias, only the ionization signal is measured since the Neganov-Luke phonon signal dominates the recoil phonons. Due to the lower ionization yield for nuclear recoils relative to electron recoils, the electron recoils experience a larger gain than for nuclear recoils, and the shape of the summed

spectrum changes (i.e., if electron recoils dominate, the spectrum at high bias will be flatter than a nuclear-recoil dominated spectrum in units of total phonon energy). As shown in Fig. 8.1, by comparing the observed spectral shape at multiple biases, the ionization yield can be determined on average, allowing the statistical subtraction of electron-recoil backgrounds.

- **Elimination of electron-recoil backgrounds below the single e-h pair ionization threshold:** The dependence of the spectral shape on recoil type described above allows a nuclear-recoil component to be identified even in the presence of an electron-recoil background due to the larger "stretching" of the spectral component due to electron recoils. In the extreme limit where a detector with a sufficiently low phonon energy threshold is developed that single electron-hole pairs can be identified, complete suppression of electron-recoil backgrounds would be possible. Such a detector would allow identification of nuclear recoils below the peak corresponding to a single e-h pair, where a phonon signal from a nuclear recoil could still be observed. In this region, it is not possible to observe electron recoils since the Neganov-Luke gain would push the energy out of the relevant range. However, additional phonon-only backgrounds, such as crystal relaxation or cracking events (e.g., similar to those observed in early CRESST detectors [327]) may appear at low energies and could present a significant background that does not originate from electron recoils. Such a technique would require massive detectors with extremely low phonon energy thresholds ($\sim$10 eV). Extensions to the CDMS technology using lower T_c TESs and high voltage biases may allow such thresholds to be obtained.

Given the apparent tension between the results from DAMA/LIBRA, CoGeNT, and CRESST-II and null results at low mass, more data will likely be required before there is broad consensus in the dark matter community as to whether these experiments are detecting WIMPs, or whether low-energy backgrounds could be mimicking a WIMP signal in each experiment. Planned upgrades to DAMA/LIBRA to reduce the detector energy threshold will give important information on the spectral shape of the observed modulation at low energy. In addition, CoGeNT is continuing to take data with its existing detector module and plans to upgrade to 4 larger detectors with comparable energy resolution to increase its target mass by roughly an order of magnitude. CRESST-II is currently working

to reduce contamination in its detector clamps to reduce the ^{206}Pb background.

In addition, SuperCDMS has installed new detectors at the SUL with lower expected low-energy backgrounds. The variety of techniques described above may allow CDMS detector technology to further improve its sensitivity to WIMPs with masses $\lesssim$10 GeV. At the same time, large experiments such as XENON100 [54], LUX [328], COUPP [155], EDELWEISS [183], and SuperCDMS will continue to improve sensitivity to high-mass WIMPs while potentially improving low-mass constraints if thresholds are sufficiently low. Thus, future data from experiments reporting a signal, as well as additional detector technologies which can provide a cross check on such signals, will be required to ultimately test whether the low-mass WIMP hypothesis can explain these experimental results.

8.2 Athermal-phonon mediated detectors using MKIDs

If the potential low-mass WIMP signals are not ultimately confirmed, ton-scale detectors will be needed to fully probe the parameter space consistent with supersymmetric WIMPs. Microwave kinetic inductance detectors (MKIDs) provide a natural technology for scaling the CDMS detector design to large target masses. Since MKIDs are easily multiplexed in the frequency domain, they allow highly pixelized phonon sensors, while significantly reducing the number of wires needed to read out a large experiment and eliminating the need for the complex cryogenic multiplexing electronics required for TES-based designs. At the same time, the MKID-based athermal phonon sensors presented in this thesis are easily fabricated from a single Al film, allowing simple detector production.

The prototype devices presented here provide the first demonstration of position and energy-resolved phonon-mediated particle detection using MKIDs. For the devices presented in this thesis, the best energy resolution obtained in an array of resonators patterned on $22 \times 20 \times 1$ mm Si substrates was $\sigma_E = 0.54$ keV at 30 keV. For comparison, the median CDMS II energy resolution was 1.0 keV at 20 keV, after applying the position dependent calibration, indicating that these prototypes already offer comparable performance to existing detectors.

Although we expect larger MKID-based devices, which require more MKIDs to maintain the same metal coverage, to have somewhat poorer energy resolution ($\sigma_E \propto \sqrt{N_r}$, where N_r is the number of MKIDs), the prototype devices also indicate several avenues to improve the signal-to-noise. The calculation of the expected resolution of these devices agrees well

with the measured resolution, after taking into account the phonon collection efficiency, kinetic inductance fraction, gap energy, and amplifier noise temperature. However, the measured phonon collection efficiency in the devices with the best energy resolution was only ∼7%, which is nearly a factor of 4 lower than the collection measured in iZIP devices with similar total metal coverage (but 300 nm thick Al films). Although this collection efficiency was found to scale with the MKID thickness, with ∼20% collection obtained in 75 nm thick films, the use of thicker films quickly decreases the kinetic inductance fraction for Al, and optimum signal-to-noise was found for 25 nm thick films. Using a material with higher normal-state resistivity would increase the kinetic inductance fraction, allowing thicker films and improved collection. In addition, improved mounting schemes that reduce transmission of phonons to the device housing may allow the collection to be improved even with the 25 nm films currently in use.

The measured kinetic inductance fraction, α, for the same devices was also only ∼7%, leaving significant room to increase the responsivity by increasing α. Lower gap, higher normal-state resistivity materials can have $\alpha \sim 1$ even for 100 nm thick films. Although TiN and NbTiN are high-resistivity materials with tunable T_c that produce high-Q resonators, the increased responsivity in these devices was found to be canceled by poorer phonon collection. However, a wide variety of lower-gap, higher resistivity materials are available, such as AlMn or other nitrides and silicides. If high-Q, uniform resonators with good phonon collection can be fabricated from one of these materials, the kinetic inductance can be increased significantly.

Finally, the dominant noise contribution is from the HEMT amplifier, which has a white noise spectrum consistent with a noise temperature $T_N = 4.8$ K above 5 kHz, with ∼3 dB excess noise at 1 kHz where the bulk of the signal is concentrated. More advanced analyses that remove correlated noise from the amplifier (which is common to all resonators) may be able to retain the white noise spectrum to below 1 kHz, possibly improving the resolution by $\sqrt{2}$. Lower noise amplifiers, such as the wideband parametric amplifier currently being developed at Caltech and JPL [314], may further reduce the noise. If such an amplifier were able to reach the quantum limit, it could provide up to an additional factor of 6 improvement in σ_E, although any additional sources of excess noise such as two-level systems (TLS) would limit the possible improvement. Further work will be required to fully characterize TLS noise in these devices.

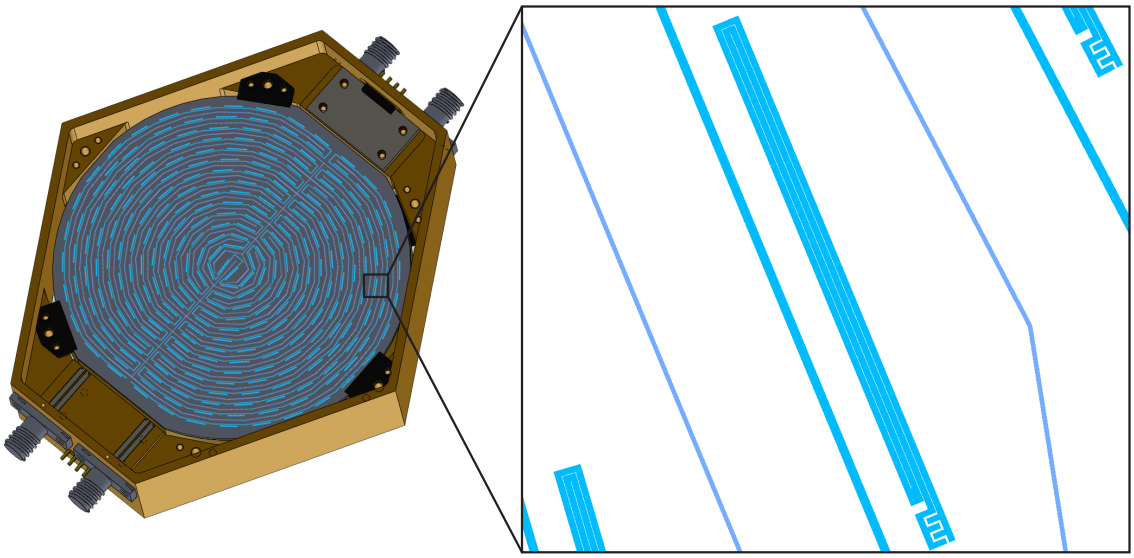


Figure 8.2: Schematic of the large substrate array design and testing enclosure. The substrate is mounted to the detector housing with 6 Cirlex clamps (black), identical to the mounting used for CDMS detectors. Both substrate faces would be patterned with an interleaved array of charge and phonon sensors, as in the iZIP. In the zoom, a single 4 mm long by 0.3 mm wide resonator is shown, as well as the interleaved charge and phonon electrodes. The MKID feedline serves as the ground for the charge measurement. The phonon sensors on each face consist of ∼256 MKIDs, read out by one coaxial cable for each face. *Resonator design and layout from B. Cornell*

Despite the significant potential for improvement, these designs already provide sufficient energy resolution for dark matter direct detection experiments. B. Cornell is currently leading the effort to scale the prototype design to a full 75 mm diameter, 10 mm thick Ge substrate. Such a detector would allow the demonstration of simultaneous readout of the charge and phonon signal in a massive detector, as well as the determination of the position resolution and background rejection possible with a highly pixelized phonon sensor. Improvements to the detector mounting scheme may also allow increased phonon collection. A preliminary design for the detector layout and resonator geometry is shown in Fig. 8.2.

With a full detector prototype, the energy resolution and background rejection can be optimized in a realistic device. In addition to the possible improvements through materials choice and substrate mounting discussed above, it may be possible to improve device performance by optimizing the resonator geometry. In particular, if background rejection from the energy partition information alone is found to be sufficient, much higher Q resonators could be used, which could increase the responsivity by more than an order of magnitude

at the cost of reduced sensor bandwidth. Even if the timing information is still needed, some subset of resonators could be weakly coupled to give a high-resolution energy measurement, while the remaining resonators could be more strongly coupled to resolve the timing of the prompt athermal signal. Other possible improvements to the resonator design include tapered resonators with reduced capacitive area to eliminate dead metal (similar to the design in Fig. 6.14), optimization of metal coverage and resonator area, or use of a higher gap superconductor (e.g. Nb) for the charge lines, feedline, and resonator capacitor to reduce uninstrumented metal capable of absorbing ballistic phonons.

Although significant engineering challenges remain, the results from prototype phonon-mediated MKID devices have demonstrated a path to extending similar designs to ∼kg scale substrates. These devices may provide a simpler and cheaper way to extend the CDMS detector technology to next-generation rare event searches such as the direct detection of dark matter. Such experiments may provide the first unambiguous detection of particle dark matter, potentially confirming the WIMP hypothesis and allowing detailed studies of the nature of dark matter and its distribution in the galactic halo.

Bibliography

[1] Wilkinson Microwave Anisotropy Probe (WMAP) Team, http://map.gsfc.nasa.gov.

[2] J. D. Simon and M. Geha, Astrophys. J. **670**, 313 (2007), arXiv:0706.0516 [astro-ph].

[3] D. G. York et al., Astrophys. J. **120**, 1579 (2000), arXiv:astro-ph/0006396.

[4] L. E. Strigari et al., Nature **454**, 1096 (2008), arXiv:0808.3772 [astro-ph].

[5] K. G. Begeman, A. H. Broeils, and R. H. Sanders, Mon. Not. Roy. Astron. Soc. **249**, 523 (1991).

[6] Y. Sofue and V. Rubin, Ann. Rev. Astron. Astrophys. **39**, 137 (2001), arXiv:astro-ph/0010594.

[7] F. Zwicky, Helv. Phys. Acta. **6**, 110 (1933).

[8] E. W. Kolb and M. Turner, *The Early Universe*, Westview Press, 1990.

[9] R. Keisler et al., Astrophys. J. **743**, 28 (2011), arXiv:1105.3182 [astro-ph.CO].

[10] A. Muzzin, H. K. C. Yee, P. B. Hall, and H. Lin, Astrophys. J. **663**, 150 (2007), arXiv:astro-ph/0703369.

[11] N. A. Bahcall et al., Astrophys. J. **585**, 182 (2003), arXiv:astro-ph/0205490.

[12] M. Girardi, S. Borgani, G. Giuricin, F. Mardirossian, and M. Mezzetti, Astrophys. J. **530**, 62 (2000), arXiv:astro-ph/9907266.

[13] CHANDRA X-ray Observatory/NASA, http://chandra.harvard.edu.

[14] A. Vikhlinin et al., Astrophys. J. **628**, 655 (2005), arXiv:astro-ph/0412306.

[15] A. Vikhlinin et al., Astrophys. J. **692**, 1033 (2009), arXiv:0805.2207 [astro-ph].

[16] C. L. Reichardt et al. (2012), arXiv:1203.5775 [astro-ph.CO].

[17] N. Aghanim et al. (2012), arXiv:1204.2743 [astro-ph.CO].

[18] J.-P. Kneib, R. S. Ellis, I. Smail, W. J. Couch, and R. M. Sharples, Astrophys. J. **471**, 643 (1996), arXiv:astro-ph/9511015.

[19] P. Natarajan, J.-P. Kneib, I. Smail, and R. S. Ellis, Astrophys. J. **499**, 600 (1998), arXiv:astro-ph/9706129.

[20] H. Hoekstra and B. Jain, Ann. Rev. Nucl. Part. Sci. **58**, 99 (2008), arXiv:0805.0139 [astro-ph].

[21] E. S. Sheldon et al., Astrophys. J. **703**, 2232 (2009), arXiv:0709.1162 [astro-ph].

[22] D. Clowe et al., Astrophys. J. Lett. **648**, L109 (2006), arXiv:astro-ph/0608407.

[23] M. Bradač et al., Astrophys. J. **687**, 959 (2008), arXiv:0806.2320 [astro-ph].

[24] N. Jarosik et al., Astrophys. J. Suppl. **192**, 14 (2011), arXiv:1001.4744 [astro-ph.CO].

[25] N. Suzuki et al. (2011), arXiv:1105.3470v1 [astro-ph.CO].

[26] E. Komatsu et al., Astrophys. J. Suppl. **192**, 18 (2011), arXiv:1001.4538 [astro-ph.CO].

[27] K. Nakamura and Particle Data Group, J. Phys. G Nucl. Phys. **37**, 075021 (2010).

[28] M. Pettini, B. J. Zych, M. T. Murphy, A. Lewis, and C. C. Steidel, Mon. Not. Roy. Astron. Soc. **391**, 1499 (2008), arXiv:0805.0594 [astro-ph].

[29] R. H. Cyburt, B. D. Fields, and K. A. Olive, JCAP **11**, 12 (2008), arXiv:0808.2818.

[30] V. Springel et al., Nature **435**, 629 (2005), arXiv:astro-ph/0504097.

[31] Simulations were performed at the National Center for Supercomputer Applications by Andrey Kravtsov (The University of Chicago) and Anatoly Klypin (New Mexico State University). Visualizations by Andrey Kravtsov.
http://cosmicweb.uchicago.edu/

[32] S. Cole et al., Mon. Not. Roy. Astron. Soc. **362**, 505 (2005), arXiv:astro-ph/0501174.

[33] W. J. Percival et al., Astrophys. J. **657**, 645 (2007), arXiv:astro-ph/0608636.

[34] T. K. Hemmick et al., Phys. Rev. D **41**, 2074 (1990).

[35] P. F. Smith, Contemp. Phys. **29**, 159 (1988).

[36] G. Jungman, M. Kamionkowski, and K. Griest, Phys. Rep. **267**, 195 (1996), arXiv:hep-ph/9506380.

[37] G. Gelmini and P. Gondolo, DM production mechanisms, in *Particle Dark Matter: Observations, Models and Searches*, edited by G. Bertone, pp. 121–141, Cambridge University Press, 2010.

[38] J. Ellis and K. A. Olive, Supersymmetric dark matter candidates, in *Particle Dark Matter: Observations, Models and Searches*, edited by G. Bertone, pp. 142–163, Cambridge University Press, 2010.

[39] L. Girardello and M. T. Grisaru, Nucl. Phys. B **194**, 65 (1982).

[40] C. Strege et al., JCAP **3**, 30 (2012), arXiv:1112.4192 [hep-ph].

[41] G. Bertone et al., JCAP **1**, 15 (2012), arXiv:1107.1715 [hep-ph].

[42] J. Ellis and K. A. Olive (2012), arXiv:1202.3262v1 [hep-ph].

[43] H. Baer, V. Barger, and A. Mustafayev, (2012), arXiv:1202.4038v1 [hep-ph].

[44] O. Buchmueller et al. (2011), arXiv:1110.3568v1 [hep-ph].

[45] O. Buchmueller et al. (2011), arXiv:1106.2529v1 [hep-ph].

[46] J. L. Feng, A. Rajaraman, and F. Takayama, Phys. Rev. Lett. **91**, 011302 (2003).

[47] J. L. Feng, Non-WIMP candidates, in *Particle Dark Matter: Observations, Models and Searches*, edited by G. Bertone, pp. 190–203, Cambridge University Press, 2010.

[48] H. Baer and M. Brhlik, Phys. Rev. D **53**, 597 (1996).

[49] R. Alemany et al., Combined LEP GMSB Stau/Smuon/Selectron Results, 189-208 GeV, LEPSUSYWG/02-09.2.

[50] A. Heister et al., Phys. Lett. B **583**, 247 (2004).

[51] ALEPH Collaboration et al., Phys. Rep. **427**, 257 (2006).

[52] ATLAS Collaboration, G. Aad et al., Phys. Lett. B **710**, 67 (2012).

[53] CMS Collaboration, S. Chatrchyan et al., Phys. Rev. Lett. **107**, 221804 (2011).

[54] XENON100 Collaboration, E. Aprile et al., Phys. Rev. Lett. **107**, 131302 (2011).

[55] CMS and ATLAS collaborations, http://cdsweb.cern.ch/record/1406786.

[56] T. Kaluza, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **1921**, 966 (1921).

[57] O. Klein, Z. Phys. **37**, 895 (1926).

[58] T. Appelquist, H.-C. Cheng, and B. A. Dobrescu, Phys. Rev. D **64**, 035002 (2001).

[59] L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999), arXiv:hep-ph/9905221.

[60] L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 4690 (1999), arXiv:hep-th/9906064.

[61] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, Phys. Lett. B **436**, 257 (1998), arXiv:hep-ph/9804398.

[62] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, Phys. Lett. B **429**, 263 (1998), arXiv:hep-ph/9803315.

[63] S. Arrenberg, L. Baudis, K. Kong, K. T. Matchev, and J. Yoo, Phys. Rev. D **78**, 056002 (2008).

[64] CMS Collaboration, V. Khachatryan et al., Phys. Lett. B **697**, 434 (2011), arXiv:1012.3375 [hep-ex].

[65] S. Chang, K. Y. Lee, and J. Song, Phys. Rev. D **85**, 055006 (2012), arXiv:1112.1483 [hep-ph].

[66] M. Schmaltz and D. Tucker-Smith, Ann. Rev. Nucl. Part. Sci. **55**, 229 (2005), arXiv:hep-ph/0502182.

[67] R. Foot, Int. J. Mod. Phys. A **19**, 3807 (2004), arXiv:astro-ph/0309330.

[68] C. P. Burgess, M. Pospelov, and T. ter Veldhuis, Nucl. Phys. B **619**, 709 (2001), arXiv:hep-ph/0011335.

[69] L. Lopez Honorez, E. Nezri, J. F. Oliver, and M. H. G. Tytgat, JCAP **2**, 28 (2007), arXiv:hep-ph/0612275.

[70] R. Enberg, P. J. Fox, L. J. Hall, A. Y. Papaioannou, and M. Papucci, JHEP **11**, 14 (2007), arXiv:0706.0918 [hep-ph].

[71] D. E. Kaplan, M. A. Luty, and K. M. Zurek, Phys. Rev. D **79**, 115016 (2009), arXiv:0901.4117 [hep-ph].

[72] G. Servant, Dark matter at the electroweak scale: non-supersymmetric candidates, in *Particle Dark Matter: Observations, Models and Searches*, edited by G. Bertone, pp. 164–189, Cambridge University Press, 2010.

[73] B. W. Lee and S. Weinberg, Phys. Rev. Lett. **39**, 165 (1977).

[74] H. Goldberg, Phys. Rev. Lett. **50**, 1419 (1983).

[75] K. Griest, Phys. Rev. D **38**, 2357 (1988).

[76] A. Bottino, N. Fornengo, and S. Scopel, Phys. Rev. D **67**, 063519 (2003).

[77] A. Bottino, F. Donato, N. Fornengo, and S. Scopel, Phys. Rev. D **78**, 083520 (2008).

[78] D. Hooper and T. Plehn, Physics Letters B **562**, 18 (2003), arXiv:hep-ph/0212226.

[79] D. Feldman, Z. Liu, and P. Nath, Phys. Rev. D **81**, 117701 (2010).

[80] E. Kuflik, A. Pierce, and K. M. Zurek, Phys. Rev. D **81**, 111701 (2010).

[81] A. V. Belikov, J. F. Gunion, D. Hooper, and T. M. P. Tait, Physics Letters B **705**, 82 (2011), arXiv:1009.0549 [hep-ph].

[82] D. Hooper et al., Phys. Rev. D **82**, 123509 (2010).

[83] A. Bottino, N. Fornengo, and S. Scopel (2011), arXiv:1112.5666 [hep-ph].

[84] A. L. Fitzpatrick, D. Hooper, and K. M. Zurek, Phys. Rev. D **81**, 115005 (2010).

[85] V. Silveira and A. Zee, Phys. Lett. B **161**, 136 (1985).

[86] J. McDonald, Phys. Rev. D **50**, 3637 (1994).

[87] S. Andreas, T. Hambye, and M. H. G. Tytgat, JCAP **10**, 34 (2008), arXiv:0808.0255 [hep-ph].

[88] Y. G. Kim and S. Shin, JHEP **5**, 36 (2009), arXiv:0901.2609 [hep-ph].

[89] J. F. Gunion, D. Hooper, and B. McElrath, Phys. Rev. D **73**, 015011 (2006).

[90] J. Ellis, J. F. Gunion, H. E. Haber, L. Roszkowski, and F. Zwirner, Phys. Rev. D **39**, 844 (1989).

[91] C. Panagiotakopoulos and A. Pilaftsis, Phys. Rev. D **63**, 055003 (2001).

[92] A. Menon, D. E. Morrissey, and C. E. M. Wagner, Phys. Rev. D **70**, 035005 (2004).

[93] T. Han, P. Langacker, and B. McElrath, Phys. Rev. D **70**, 115006 (2004).

[94] F. Ferrer, L. M. Krauss, and S. Profumo, Phys. Rev. D **74**, 115007 (2006).

[95] J. F. Gunion, A. V. Belikov, and D. Hooper, (2010), arXiv:1009.2555v1 [hep-ph].

[96] P. Draper, T. Liu, C. E. M. Wagner, L.-T. Wang, and H. Zhang, Phys. Rev. Lett. **106**, 121805 (2011).

[97] D. G. Cerdeño and O. Seto, JCAP **8**, 32 (2009), arXiv:0903.4677 [hep-ph].

[98] D. B. Kaplan, Phys. Rev. Lett. **68**, 741 (1992).

[99] S. M. Barr, R. Sekhar Chivukula, and E. Farhi, Phys. Lett. B **241**, 387 (1990).

[100] S. Nussinov, Phys. Lett. B **165**, 55 (1985).

[101] G. Steigman, Ann. Rev. Nucl. Part. Sci. **57**, 463 (2007), arXiv:0712.1100 [astro-ph].

[102] S. Dodelson, B. R. Greene, and L. M. Widrow, Nucl. Phys. B **372**, 467 (1992).

[103] G. R. Farrar and G. Zaharijas, Phys. Rev. Lett. **96**, 041302 (2006).

[104] H. An, S.-L. Chen, R. N. Mohapatra, and Y. Zhang, JHEP **3**, 124 (2010), arXiv:0911.4463 [hep-ph].

[105] T. Cohen, D. J. Phalen, A. Pierce, and K. M. Zurek, Phys. Rev. D **82**, 056001 (2010).

[106] T. Cohen and K. M. Zurek, Phys. Rev. Lett. **104**, 101301 (2010).

[107] M. L. Graesser, I. M. Shoemaker, and L. Vecchi, JHEP **10**, 110 (2011), arXiv:1103.2771 [hep-ph].

[108] J. L. Feng and J. Kumar, Phys. Rev. Lett. **101**, 231301 (2008).

[109] J. L. Feng, V. Rentala, and Z. Surujon, Phys. Rev. D **84**, 095033 (2011).

[110] J. L. Feng, V. Rentala, and Z. Surujon, Phys. Rev. D **85**, 055003 (2012), arXiv:1111.4479 [hep-ph].

[111] L. Alvarez-Gaumé, M. Claudson, and M. B. Wise, Nucl. Phys. B **207**, 96 (1982).

[112] M. Dine, A. E. Nelson, and Y. Shirman, Phys. Rev. D **51**, 1362 (1995).

[113] R. Barbieri, T. Gregoire, and L. J. Hall, (2005), arXiv:hep-ph/0509242.

[114] Z. Berezhiani, Int. J. Mod. Phys. A **19**, 3775 (2004), arXiv:hep-ph/0312335.

[115] R. Foot, Phys. Lett. B **632**, 467 (2006), arXiv:hep-ph/0507294.

[116] R. Foot (2012), arXiv:1203.2387v1 [hep-ph].

[117] R. Foot, Phys. Lett. B **703**, 7 (2011), arXiv:1106.2688 [hep-ph].

[118] C. A. Baker et al., Phys. Rev. Lett. **97**, 131801 (2006).

[119] R. D. Peccei and H. R. Quinn, Phys. Rev. D **16**, 1791 (1977).

[120] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).

[121] S. Hannestad, A. Mirizzi, G. G. Raffelt, and Y. Y. Y. Wong, JCAP **8**, 1 (2010), arXiv:1004.0695 [astro-ph.CO].

[122] A. Melchiorri, O. Mena, and A. c. v. Slosar, Phys. Rev. D **76**, 041303 (2007).

[123] J. Preskill, M. B. Wise, and F. Wilczek, Phys. Lett. B **120**, 127 (1983).

[124] L. F. Abbott and P. Sikivie, Phys. Lett. B **120**, 133 (1983).

[125] I. G. Irastorza et al., JCAP **6**, 13 (2011), arXiv:1103.5334 [hep-ex].

[126] D. A. Dicus, E. W. Kolb, V. L. Teplitz, and R. V. Wagoner, Phys. Rev. D **18**, 1829 (1978).

[127] S. J. Asztalos et al., Phys. Rev. Lett. **104**, 041301 (2010).

[128] D. Lyapustin (2011), arXiv:1112.1167v1 [astro-ph.CO].

[129] CAST Collaboration, M. Arik et al., Phys. Rev. Lett. **107**, 261302 (2011).

[130] M. Milgrom, Astrophys. J. **270**, 365 (1983).

[131] R. H. Sanders and S. S. McGaugh, Ann. Rev. Astron. Astrophys. **40**, 263 (2002), arXiv:astro-ph/0204521.

[132] J. D. Bekenstein, Phys. Rev. D **70**, 083509 (2004).

[133] R. Wojtak, S. H. Hansen, and J. Hjorth, Nature **477**, 567 (2011), arXiv:1109.6571 [astro-ph.CO].

[134] J. D. Bekenstein and R. H. Sanders, Mon. Not. Roy. Astron. Soc. **421**, L59 (2012), arXiv:1110.5048 [astro-ph.CO].

[135] E. Pointecouteau and J. Silk, Mon. Not. Roy. Astron. Soc. **364**, 654 (2005), arXiv:astro-ph/0505017.

[136] G. W. Angus, H. Y. Shan, H. S. Zhao, and B. Famaey, Astrophys. J. Lett. **654**, L13 (2007), arXiv:astro-ph/0609125.

[137] J. D. Lewin and P. F. Smith, Astropart. Phys. **6**, 87 (1996).

[138] D. G. Cerdeño and A. M. Green, Direct detection of wimps, in *Particle Dark Matter: Observations, Models and Searches*, edited by G. Bertone, pp. 347–369, Cambridge University Press, 2010.

[139] M. C. Smith, S. H. Whiteoak, and N. W. Evans, Astrophys. J. **746**, 181 (2012).

[140] A. M. Green, Mod. Phys. Lett. A **27**, 30004 (2012), arXiv:1112.0524 [astro-ph.CO].

[141] M. Lisanti and D. N. Spergel (2011), arXiv:1105.4166v1 [astro-ph.CO].

[142] E. I. Gates, G. Gyuk, and M. S. Turner, Astrophys. J. Lett. **449**, L123 (1995), arXiv:astro-ph/9505039.

[143] XENON100 Collaboration, E. Aprile et al., Phys. Rev. D **84**, 061101 (2011).

[144] CDMS Collaboration, Z. Ahmed et al., Phys. Rev. Lett. **106**, 131302 (2011), arXiv:1011.2482v3 [astro-ph.CO].

[145] CDMS Collaboration, Z. Ahmed et al., Phys. Rev. D **83**, 112002 (2011).

[146] M. C. Smith et al., Mon. Not. Roy. Astron. Soc. **379**, 755 (2007).

[147] A. Kurylov and M. Kamionkowski, Phys. Rev. D **69**, 063503 (2004).

[148] D. R. Tovey, R. J. Gaitskell, P. Gondolo, Y. Ramachers, and L. Roszkowski, Phys. Lett. B **488**, 17 (2000), arXiv:hep-ph/0005041.

[149] R. H. Helm, Phys. Rev. **104**, 1466 (1956).

[150] J. Engel, Phys. Lett. B **264**, 114 (1991).

[151] V. I. Dimitrov, J. Engel, and S. Pittel, Phys. Rev. D **51**, R291 (1995).

[152] M. T. Ressell et al., Phys. Rev. D **48**, 5519 (1993).

[153] CDMS Collaboration, Z. Ahmed et al., Science **327**, 1619 (2010).

[154] G. Angloher et al. (2011), arXiv:1109.0702v1 [astro-ph.CO].

[155] E. Behnke et al. (2012), arXiv:1204.3094v1 [astro-ph.CO].

[156] XENON10 Collaboration, J. Angle et al., Phys. Rev. Lett. **107**, 051301 (2011).

[157] A. K. Drukier, K. Freese, and D. N. Spergel, Phys. Rev. D **33**, 3495 (1986).

[158] DMTPC Collaboration, S. Ahlen et al., Phys. Lett. B **695**, 124 (2011), arXiv:1006.2928 [hep-ex].

[159] E. Daw et al., Astropart. Phys. **35**, 397 (2012).

[160] K. Miuchi et al., Phys. Lett. B **686**, 11 (2010), arXiv:1002.1794 [astro-ph.CO].

[161] D. Smith and N. Weiner, Phys. Rev. D **64**, 043502 (2001).

[162] N. W. Evans, C. M. Carollo, and P. T. de Zeeuw, Mon. Not. Roy. Astron. Soc. **318**, 1131 (2000), arXiv:astro-ph/0008156.

[163] A. M. Green, Phys. Rev. D **66**, 083003 (2002).

[164] M. Kuhlen et al., JCAP **2**, 30 (2010), arXiv:0912.2358 [astro-ph.GA].

[165] C. McCabe, Phys. Rev. D **82**, 023530 (2010).

[166] J. Diemand, M. Kuhlen, and P. Madau, Astrophys. J. **657**, 262 (2007), arXiv:astro-ph/0611370.

[167] J. Stadel et al., Mon. Not. Roy. Astron. Soc. **398**, L21 (2009), arXiv:0808.2981 [astro-ph].

[168] V. Springel et al., Mon. Not. Roy. Astron. Soc. **391**, 1685 (2008), arXiv:0809.0898 [astro-ph].

[169] R. Catena and P. Ullio, JCAP **8**, 4 (2010), arXiv:0907.0018 [astro-ph.CO].

[170] C. Moni Bidin, G. Carraro, R. A. Mendez, and R. Smith (2012), arXiv:1204.3924 [astro-ph.GA].

[171] S. Garbari, J. I. Read, and G. Lake, Mon. Not. Roy. Astron. Soc. **416**, 2318 (2011), arXiv:1105.6339 [astro-ph.GA].

[172] J. Bovy, D. W. Hogg, and H.-W. Rix, Astrophys. J. **704**, 1704 (2009), arXiv:0907.5423 [astro-ph.GA].

[173] S. E. Koposov, H.-W. Rix, and D. W. Hogg, Astrophys. J. **712**, 260 (2010), arXiv:0907.1085 [astro-ph.GA].

[174] F. J. Kerr and D. Lynden-Bell, Mon. Not. Roy. Astron. Soc. **221**, 1023 (1986).

[175] P. J. Fox, J. Liu, and N. Weiner, Phys. Rev. D **83**, 103514 (2011).

[176] C. Savage, G. Gelmini, P. Gondolo, and K. Freese, JCAP **4**, 10 (2009), arXiv:0808.3607 [astro-ph].

[177] C. Savage, P. Gondolo, and K. Freese, Phys. Rev. D **70**, 123513 (2004).

[178] CDMS Collaboration, Z. Ahmed et al., Phys. Rev. Lett. **102**, 011301 (2009).

[179] CDMS Collaboration, D. S. Akerib et al., Phys. Rev. D **73**, 011102 (2006).

[180] J. L. Feng, J. Kumar, D. Marfatia, and D. Sanford, Phys. Lett. B **703**, 124 (2011), arXiv:1102.4331 [hep-ph].

[181] S. Chang, J. Liu, A. Pierce, N. Weiner, and I. Yavin, JCAP **8**, 18 (2010), arXiv:1004.0697 [hep-ph].

[182] Z. Kang, T. Li, T. Liu, C. Tong, and J. M. Yang, JCAP **1**, 28 (2011), arXiv:1008.5243 [hep-ph].

[183] EDELWEISS Collaboration, E. Armengaud et al., Phys. Lett. B **702**, 329 (2011), arXiv:1103.4070 [astro-ph.CO].

[184] CDMS and EDELWEISS Collaborations, Z. Ahmed et al., Phys. Rev. D **84**, 011102 (2011).

[185] D. Y. Akimov et al., Phys. Lett. B **709**, 14 (2012), arXiv:1110.4769 [astro-ph.CO].

[186] CDMS, D. S. Akerib et al., Phys. Rev. D **82**, 122004 (2010).

[187] DAMA/LIBRA Collaboration, R. Bernabei et al., Eur. Phys. J. C **67**, 39 (2010), arXiv:1002.1028 [astro-ph.GA].

[188] DAMA/LIBRA Collaboration, R. Bernabei et al., Eur. Phys. J. C **56**, 333 (2008), arXiv:0804.2741 [astro-ph].

[189] CoGeNT Collaboration, C. E. Aalseth et al., Phys. Rev. Lett. **107**, 141301 (2011).

[190] C. Kelso, D. Hooper, and M. R. Buckley, (2011), arXiv:1110.5338 [astro-ph.CO].

[191] J. I. Collar and N. E. Fields (2012), arXiv:1204.3559 [astro-ph.CO].

[192] J. I. Collar, Talk at TAUP 2011 Workshop, Munich, Germany (Sep. 5–9, 2011).

[193] CoGeNT Collaboration, C. E. Aalseth et al., Phys. Rev. Lett. **106**, 131301 (2011).

[194] C. Savage et al., Phys. Rev. D **83**, 055002 (2011).

[195] J. I. Collar (2011), arXiv:1106.0653 [astro-ph.CO].

[196] DAMA/NaI Collaboration, R. Bernabei et al., Phys. Lett. B **450**, 448 (1999).

[197] DAMA/NaI Collaboration, R. Bernabei et al., Nuovo Cimento Rivista Serie **26**, 010000 (2003), arXiv:astro-ph/0307403.

[198] K. Blum (2011), arXiv:1110.0857 [astro-ph.HE].

[199] J. P. Ralston (2010), arXiv:1006.5255 [hep-ph].

[200] D. Nygren (2011), arXiv:1102.0815 [astro-ph.IM].

[201] DAMA/LIBRA Collaboration, R. Bernabei et al. (2012), arXiv:1202.4179 [astro-ph.GA].

[202] P. J. Fox, J. Kopp, M. Lisanti, and N. Weiner, Phys. Rev. D **85**, 036008 (2012).

[203] CDMS Collaboration, Z. Ahmed et al. (2012), arXiv:1203.1309v1 [astro-ph.CO].

[204] M. Kuźniak, M. G. Boulay, and T. Pollmann (2012), arXiv:1203.1576 [astro-ph.IM].

[205] XENON10 Collaboration, J. Angle et al., Phys. Rev. Lett. **100**, 021303 (2008).

[206] T. A. Porter, R. P. Johnson, and P. W. Graham, Ann. Rev. Astron. Astrophys. **49**, 155 (2011), arXiv:1104.2836 [astro-ph.HE].

[207] Fermi-LAT Collaboration, M. Ackermann et al., Phys. Rev. Lett. **107**, 241302 (2011).

[208] D. Hooper, D. P. Finkbeiner, and G. Dobler, Phys. Rev. D **76**, 083012 (2007).

[209] HESS Collaboration, F. Aharonian et al., Phys. Rev. Lett. **97**, 221102 (2006).

[210] W. de Boer, C. Sander, V. Zhukov, A. V. Gladyshev, and D. I. Kazakov, Phys. Rev. Lett. **95**, 209001 (2005).

[211] O. Adriani et al., Nature **458**, 607 (2009), arXiv:0810.4995 [astro-ph].

[212] Fermi-LAT Collaboration, M. Ackermann et al., Phys. Rev. Lett. **108**, 011103 (2012).

[213] P. Meade, M. Papucci, A. Strumia, and T. Volansky, Nucl. Phys. B **831**, 178 (2010), arXiv:0905.0480 [hep-ph].

[214] M. Cirelli, M. Kadastik, M. Raidal, and A. Strumia, Nucl. Phys. B **813**, 1 (2009), arXiv:0809.2409 [hep-ph].

[215] L. Bergström, T. Bringmann, and J. Edsjö, Phys. Rev. D **78**, 103520 (2008).

[216] D. Hooper, P. Blasi, and P. Dario Serpico, JCAP **1**, 25 (2009), arXiv:0810.1527 [astro-ph].

[217] P. D. Serpico (2011), arXiv:1108.4827v1 [astro-ph.HE].

[218] C. Weniger (2012), arXiv:1204.2797v1 [hep-ph].

[219] D. Hooper (2012), arXiv:1201.1303v1 [astro-ph.CO].

[220] D. Hooper and T. Linden, Phys. Rev. D **84**, 123005 (2011).

[221] D. Hooper and T. Linden, Phys. Rev. D **83**, 083517 (2011).

[222] T. Linden, D. Hooper, and F. Yusef-Zadeh, Astrophys. J. **741**, 95 (2011), arXiv:1106.5493 [astro-ph.HE].

[223] L. Bergström, T. Bringmann, and J. Edsjö, Phys. Rev. D **83**, 045024 (2011).

[224] G. Bertone et al., Phys. Rev. D **85**, 055014 (2012), arXiv:1111.2607 [astro-ph.HE].

[225] E. A. Baltz, M. Battaglia, M. E. Peskin, and T. Wizansky, Phys. Rev. D **74**, 103521 (2006).

[226] G. Bertone, D. Cumberbatch, R. Ruiz de Austri, and R. Trotta, JCAP **1**, 4 (2012), arXiv:1107.5813 [astro-ph.HE].

[227] V. Mandic, Ph.D. thesis, University of California, Berkeley (2004).

[228] J. Filippini, Ph.D. thesis, University of California, Berkeley (2008).

[229] T. Bruch, Ph.D. thesis, Universitat Zurich (2010).

[230] Z. Ahmed, Ph.D. thesis, California Institute of Technology (2012).

[231] C. D. Thurmond, J. Electrochem. Soc. **122**, 1133 (1975).

[232] W. Ogburn, Ph.D. thesis, Stanford University (2008).

[233] T. Shutt et al., Phys. Rev. Lett. **69**, 3531 (1992).

[234] T. Shutt, Ph.D. thesis, University of California, Berkeley (1993).

[235] X.-F. Navick, M. Chapellier, F. Déliot, S. Hervé, and L. Miramonti, Nucl. Instr. Meth. Phys. Res. A **444**, 361 (2000).

[236] B. A. Young and K. M. Yu, Rev. Sci. Instr. **66**, 2625 (1995).

[237] B. Cabrera, M. Pyle, R. Moffatt, K. Sundqvist, and B. Sadoulet, (2010), arXiv:1004.1233v1 [astro-ph.IM].

[238] B. Neganov and V. Trofimov, Otkryt. Izobret. **146**, 215 (1985).

[239] P. N. Luke, J. Appl. Phys. **64**, 6858 (1988).

[240] M. Pyle, Ph.D. thesis, Stanford University (2012).

[241] M. Pyle et al. (2012), arXiv:1201.3685v1 [astro-ph.IM].

[242] R. M. Clarke, Ph.D. thesis, Stanford University (1999).

[243] T. Saab, Ph.D. thesis, Stanford University (2002).

[244] G. Wang, J. Appl. Phys. **107**, 094504 (2010).

[245] A. G. Kozorezov et al., Phys. Rev. B **61**, 11807 (2000).

[246] N. E. Booth, Appl. Phys. Lett. **50**, 293 (1987).

[247] C. L. Chang, Ph.D. thesis, Stanford University (2004).

[248] B. A. Young et al., J. Appl. Phys. **86**, 6975 (1999).

[249] K. D. Irwin, Appl. Phys. Lett. **66**, 1998 (1995).

[250] T. Shutt et al., Nucl. Instr. Meth. Phys. Res. A **444**, 340 (2000).

[251] R. M. Clarke et al., Appl. Phys. Lett. **76**, 2958 (2000).

[252] V. Mandic et al., Nucl. Instr. Meth. Phys. Res. A **520**, 171 (2004).

[253] SNOLAB, http://www.snolab.ca/.

[254] J. Sander, Ph.D. thesis, University of California, Santa Barbara (2007).

[255] R. Bunker, Ph.D. thesis, University of California, Santa Barbara (2011).

[256] CDMS, D. S. Akerib et al., Phys. Rev. D **72**, 052009 (2005).

[257] D. Seitz, Cold Hardware Website, CDMS Internal Documentation.

[258] ROOT: A data analysis framework, http://root.cern.ch/.

[259] S. R. Golwala, Ph.D. thesis, University of California, Berkeley (2000).

[260] B. L. Dougherty, Phys. Rev. A **45**, 2104 (1992).

[261] CDMS Collaboration, Z. Ahmed et al., in prep. (2012).

[262] J. N. Bahcall, Phys. Rev. **132**, 362 (1963).

[263] L. Ekström and R. Firestone, Table of Radioactive Isotopes, http://ie.lbl.gov/toi/.

[264] C. Chasman, K. W. Jones, and R. A. Ristinen, Phys. Rev. Lett. **15**, 245 (1965).

[265] C. Chasman, K. W. Jones, R. A. Ristinen, and J. T. Sample, Phys. Rev. **154**, 239 (1967).

[266] C. Chasman, K. W. Jones, H. W. Kraner, and W. Brandt, Phys. Rev. Lett. **21**, 1430 (1968).

[267] K. W. Jones and H. W. Kraner, Phys. Rev. C **4**, 125 (1971).

[268] K. W. Jones and H. W. Kraner, Phys. Rev. A **11**, 1347 (1975).

[269] Y. Messous et al., Astropart. Phys. **3**, 361 (1995).

[270] A. R. Sattler, F. L. Vook, and J. M. Palms, Phys. Rev. **143**, 588 (1966).

[271] L. Baudis et al., Nucl. Instr. Meth. Phys. Res. A **418**, 348 (1998).

[272] P. Barbeau, J. Collar, and P. Whaley, Nucl. Instr. Meth. Phys. Res. A **574**, 385 (2007).

[273] P. Barbeau, Ph.D. thesis, U. Chicago (2009).

[274] E. Simon et al., Nucl. Instr. Meth. Phys. Res. A **507**, 643 (2003).

[275] A. Benoit et al., Nucl. Instr. Meth. Phys. Res. A **577**, 558 (2007).

[276] J. Lindhard et al., K. Dan. Vidensk. Selsk., Mat.-Fys. Medd. **33**, 10 (1963).

[277] CDMS Collaboration, Z. Ahmed et al., Phys. Rev. D **81**, 042002 (2010).

[278] S. Yellin (2011), arXiv:1105.2928v2 [physics.data-an].

[279] S. Yellin (2007), arXiv:0709.2701v1 [physics.data-an].

[280] S. Yellin, Phys. Rev. D **66**, 032005 (2002).

[281] J. Collar (2011), arXiv:1103.3481v1 [astro-ph.CO].

[282] N. Bozorgnia, G. B. Gelmini, and P. Gondolo, JCAP **11**, 019 (2010).

[283] Deep Underground Science and Engineering Laboratory, http://www.dusel.org/.

[284] M. Pyle et al., AIP Conf. Proc. **1185**, 223 (2009).

[285] K. A. McCarthy, Studying MISS event backgrounds using the DMC, CDMS internal note (2012).

[286] J. A. Chervenak et al., Appl. Phys. Lett. **74**, 4043 (1999).

[287] C. D. Reintsema et al., Rev. Sci. Instr. **74**, 4500 (2003).

[288] P. A. J. de Korte et al., Rev. Sci. Instr. **74**, 3807 (2003).

[289] K. D. Irwin et al. (2011), arXiv:1110.1608v1 [astro-ph.IM].

[290] M. D. Niemack et al., Appl. Phys. Lett. **96**, 163509 (2010).

[291] C. D. Reintsema et al., J. Low Temp. Phys. **151**, 927 (2008).

[292] M. A. Dobbs et al. (2011), arXiv:1112.4215v1 [astro-ph.IM].

[293] J. A. B. Mates, G. C. Hilton, K. D. Irwin, L. R. Vale, and K. W. Lehnert, Appl. Phys. Lett. **92**, 023514 (2008).

[294] P. K. Day, H. G. LeDuc, B. A. Mazin, A. Vayonakis, and J. Zmuidzinas, Nature **425**, 817 (2003).

[295] P. R. Maloney et al., SPIE Conf. Proc. **7741** (2010).

[296] B. A. Mazin et al. (2011), arXiv:1112.0004v1 [astro-ph.IM].

[297] B. A. Mazin et al., Appl. Phys. Lett. **89**, 222507 (2006), arXiv:cond-mat/0610130.

[298] J. Gao, Ph.D. thesis, California Institute of Technology (2008).

[299] D. C. Mattis and J. Bardeen, Phys. Rev. **111**, 412 (1958).

[300] J. Gao et al., J. Low Temp. Phys. **151**, 557 (2008).

[301] J. Zmuidzinas, Amplifier-limited mkid energy resolution: Prospects for ultrasensitive photon-counting detectors, Caltech KIDs internal report (2009).

[302] H. G. Leduc et al., Appl. Phys. Lett. **97**, 102509 (2010), arXiv:1003.5584 [cond-mat.supr-con].

[303] S. R. Golwala, Expected Energy Resolution of a Phonon-Mediated Detector using TiN MKIDs, Caltech KIDs internal report (2010).

[304] J. Gao, J. Zmuidzinas, B. A. Mazin, H. G. Leduc, and P. K. Day, Appl. Phys. Lett. **90**, 102507 (2007), arXiv:cond-mat/0609614.

[305] O. Noroozian, P. K. Day, H. G. LeDuc, B. H. Eom, and J. Zmuidzinas, J. Low Temp. Phys., in publication (2011).

[306] D. C. Moore et al., AIP Conf. Proc. **1185**, 168 (2009).

[307] S. Golwala et al., J. Low Temp. Phys. **151**, 550 (2008).

[308] J. Gao, Estimation of energy resolution of dark matter KIDs, Caltech KIDs internal report (2004).

[309] H. Kraus et al., Phys. Lett. B **231**, 195 (1989).

[310] J. Jochum et al., Ann. Phys. **505**, 611 (1993).

[311] S. W. Deiker et al., Appl. Phys. Lett. **85**, 2137 (2004).

[312] L. J. Swenson et al., Appl. Phys. Lett. **96**, 263511 (2010).

[313] W. L. McMillan, Phys. Rev. **167**, 331 (1968).

[314] B. H. Eom, P. K. Day, H. G. Leduc, and J. Zmuidzinas (2012), arXiv:1201.2392v1 [cond-mat.supr-con].

[315] S. Doyle, P. Mauskopf, J. Naylon, A. Porch, and C. Duncombe, J. Low Temp. Phys. **151**, 530 (2008).

[316] SONNET High Frequency Electromagnetic Software http://www.sonnetsoftware.com/.

[317] K. A. McCarthy et al. (2011), arXiv:1109.0551v1 [astro-ph.IM].

[318] B. A. Mazin, Ph.D. thesis, California Institute of Technology (2005).

[319] Collaboration for Astronomy Signal Processing and Electronics Research http://casper.berkeley.edu.

[320] R. Duan et al., SPIE Conf. Proc. **7741** (2010).

[321] Karoo Array Telescope Communication Protocol, http://pypi.python.org/pypi/katcp.

[322] M. Eckart, Ph.D. thesis, California Institute of Technology (2007).

[323] M. Berger et al., XCOM: Photon Cross Sections Database, http://www.nist.gov/pml/data/xcom/.

[324] Matlab Global Optimization Toolbox, http://www.mathworks.com/help/toolbox/gads/.

[325] R. Barends et al., Appl. Phys. Lett. **92**, 223502 (2008), arXiv:0804.3499 [cond-mat.mes-hall].

[326] M. Pyle, Low Energy Phonon Only Fiducial Volume, CDMS Internal Note (2012).

[327] J. Åström et al., Nucl. Instr. Meth. Phys. Res. A **559**, 754 (2006).

[328] D. N. McKinsey et al., J. Phys. Conf. Ser. **203**, 012026 (2010).