Lowering Backgrounds and Thresholds in the Search for Light Dark Matter with SuperCDMS

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

Jack Henry Nelson

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

Dr. Priscilla Cushman, Advisor

January, 2024

Acknowledgements

Having my graduate career almost perfectly span the SuperCDMS "interbellum" period between Soudan and SNOLAB makes the acknowledgement difficult for me to write. I have had the opportunity to learn from so many wonderful and brilliant people who made this collaboration what it is, in addition to having the chance to meet the next generation with the fresh ideas needed keep pushing forward. Needless to say, I'm excited to see what the next few years bring!

I am extremely grateful to my advisor, Prisca Cushman. You provided me with what most grad students covet – the freedom to pursue topics which interested me and make my own mistakes, while always providing guidance when I needed it and helping me to see the big picture. While working on SuperCDMS at UMN, I also absorbed a huge amount of information on the experiment, and how one does science in general, from Vuk Mandic, Matt Fritts and Anthony Villano. Though may have overlapped briefly, I also want to thank the SuperCDMS students who came before me: Mark Pepin, D'Ann Barker and Nick Mast, for showing me just how much grad students could accomplish. To those I leave behind in the local group: Zach Williams, Himangshu Neog and Shubham Pandey, I wish you all good luck. I will always remember the detour to the beach on our way back from the collaboration meeting. Even though I was terrified that we would miss our flight, I'm glad you dragged me along. Outside of SuperCDMS at UMN, I thank Liliya Williams for providing comments on this thesis.

The results from CPD could not have happened without the help of many people. In particular, I want to thank Xinran Li for spearheading the implementation of the noise boosting framework, and being patient with me as I tried to wrap my head around it. I am also grateful to Matt Pyle and Scott Hertel for guiding the analysis, and providing valuable feedback on my work. I thank Richard Germond for making the journey with me into SPICE/HeRALD world. It would not have been possible to indulge my habit of staring at the UMass data if it weren't for the help of Doug Pinckney and Pratyush Patel. Thank you for enabling me.

In the broader SuperCDMS collaboration, I first want to thank the reviewers of the CPD analysis: Yan Liu, Ray Bunker, Maddy Zurowski and Sukee Dharani. You all provided crucial comments which helped me realize my blind spots. I also thank Ziqing Hong for holding my hand at times through the approval process. For the background simulation work, I owe much to Rob Calkins, Mike Kelsey and Ben Loer for helping me understand SuperSIM.

Finally, thank you my family. My wife, Mia, my parents, Eric and Beth, and my sisters, Lydia and Laura. You supported me, even at times when I'm sure it looked like I would never finish. I couldn't have done it without you.

To my wife, Mia

You've always been there for me

Abstract

Cosmological observations have produced a wealth of evidence which demonstrates that the majority of the matter content in our Universe is "dark". The identity of this dark matter remains elusive, as none of the members of the standard model of particle physics accurately describe its properties. This has prompted the scientific community to launch a broad search, spanning decades in time, mass and sensitivity, with the goal of detecting and identifying this source of new physics.

The next generation of the Super Cryogenic Dark Matter Search (SuperCDMS) is currently under construction deep underground at SNOLAB. The experiment aims to expand the search for dark matter to lower masses ($\lesssim 10$ GeV/c^2) and greater sensitivities using silicon and germanium detectors by minimizing experimental backgrounds and operating detectors with superb energy resolution.

SuperCDMS will accomplish its low projected background in part by deploying a robust shield to protect its detectors from environmental radiation. This dissertation presents the results of simulations which demonstrate the success of the shield design at stopping radiogenic neutrons. The shield will be able to reduce these environmental sources to the point where coherent scattering from solar neutrinos are expected to dominate the nuclear recoil backgrounds.

In order to search for such light dark matter masses, SuperCDMS uses sensitive transition edge sensors to measure small energy depositions in the detectors. The ultimate energy resolution of these devices, expected to be < 1 eV, has not yet been realized. This dissertation describes the analysis of a dark matter search performed at the University of Massachusetts Amherst with a prototype detector which uses SuperCDMS style sensors to achieve a baseline energy resolution of 2.3 eV. The results of this search demonstrate sensitivity to dark matter candidates with masses as low as ~ 25 MeV/c^2.

Contents

List of Tables

List of Figures

Part I

Cosmology and the Case for Dark Matter

Chapter 1

Cosmological Preliminaries

1.1 A Crash Course in Cosmology

From a high level, we can think of the central problem of cosmology as taking a census of the Universe. Everything we know exists in the unending expanse of space and time. Given our innate curiosity, it's only natural for us to ask what the nature of this expanse is. What occupies it and what are the rules that govern it?

Before we can attempt to answer this question, we must choose a scale on which to perform this census. On the cosmological scale our understanding of the universe is best described via the Einstein field equations. On the microscopic scale we can describe the the contents of the universe through the Standard Model (SM) of particle physics. Einstein describes how Space-Time itself evolves and interacts with massive objects, while the SM describes the known particles occupying our universe and how they interact with one another. Einstein's theory is summarized through his eponymous field equations

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{1.1}$$

Here $g_{\mu\nu}$ is the metric tensor, which codifies how we locally measure distance. $R_{\mu\nu}$ is the Ricci curvature tensor and R the associated scalar, which describe the curvature of the metric. $T_{\mu\nu}$ is the stress-energy tensor of whatever occupies the space we're trying to describe, and G is Newton's gravitational constant. Note that this is not the most general statement of the Einstein Field Equations. In Eq. 1.1 we have neglected

a constant of integration, Λ. Λ is referred to as the "cosmological constant", and though it has a profound influence on the evolution of Eq. 1.1 we will suppress it now and reintroduce it later by folding it into $T_{\mu\nu}$. Where Greek letters index the four space-time dimensions. These equations give us a detailed description of how space-time warps in responds to the presence of mass and energy.

To apply 1.1 to the entire Universe, all we need to do is assume that the Universe is both homogeneous and isotropic. That is to say that there is no preferred location or direction in the Universe. This is argument is dubbed the Cosmological Principle. At first glance this seems to be a pretty poor assumption – our description of gravity will certainly look different in our solar system than it would if we were far from any star. But that's why it's called the "Cosmological Principle", we must assume that we're looking at the problem an scales large enough that any anisotropy or inhomogeneity averages out and is negligible.

In solving the field equations, we want to find $g_{\mu\nu}$ which satisfies Eq. 1.1 given a particular distribution of stuff in the universe, $T_{\mu\nu}$. If the Cosmological Principle is true, then this will be reflected in both $T_{\mu\nu}$ and $g_{\mu\nu}$. That is, they must both themselves be isotropic and homogeneous. For $T_{\mu\nu}$ we then interpret the time component as energy-density of the contents of space and the spatial components represent the corresponding pressure. This is also equivalent to assuming that the mass-energy content of the universe consists of a perfect fluid.

$$T_{00} = -\rho,\ T_{ii} = p \tag{1.2}$$

Where ρ is the energy density of the fluid and p is the pressure. For $g_{\mu\nu}$, we can write the general metric which contains the desired symmetries, called the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric

$$-c^2 d\tau^2 = -c^2 dt^2 + a(t)\big(\frac{dr^2}{1-kr^2} + r^2\big[d\theta^2 + \sin^2\theta d\phi^2\big]\big) \tag{1.3}$$

where $d\tau$ is the proper time interval, dt is the clock time interval and $\big(r, \theta, \phi\big)$ are the usual spatial spherical polar coordinates. k describes the intrinsic curvature of the Universe. Current measurements are consistent with a flat Universe which is described by $k = 0$ [1]. We will assume this throughout the rest of this section. $a(t)$ is the cosmological scale factor. If compare the Universe at some time t_1 and later at t_2, we will find $a(t_1) \neq a(t_2)$, as the Universe expanded. By convention we define $a(t_{now}) = 1$. Furthermore, a is closely related to the cosmological redshift z. If we observe a photon from the past, the wavelength of that photon

will have increased according to how much the scale factor has increased.

$$z \equiv \frac{\lambda_{now} - \lambda_{then}}{\lambda_{then}}, \quad \frac{\lambda_{now}}{\lambda_{then}} = \frac{1}{a_{then}} \ \ (a_{now} = 1) \tag{1.4}$$

$$\rightarrow a_{then} = (1+z)^{-1} \tag{1.5}$$

Once we've defined the scale factor $a(t)$, we can take a closer look at how it scales with the energy density. Looking at the role $a(t)$ plays in the FLRW metric, we can interpret da as a change in the length scale of the Universe, and hence $d(a^3)$ represents a change in volume. With this in mind, and recalling that we have assumed the content of the Universe is well described by a perfect fluid, it follows from mass-energy conservation

$$d(\rho a^3) = -p d(a^3) \tag{1.6}$$

That is, the change in the the energy must be opposite the pressure times the change in volume. From this it follows that

$$\frac{d\rho}{-3(p+\rho)} = \frac{da}{a} \tag{1.7}$$

We can simplify this further if we make an assumption about the equation of state of our perfect fluid. In particular, the pressure and energy density are related by

$$p = w\rho \tag{1.8}$$

Where w is a constant which defines a particular state. With this, we find that the energy density of a fluid scales with $a(t)$ as

$$\rho \propto a^{-3(1+w)} \tag{1.9}$$

To determine the time evolution of the scale factor, we plug the FLRW metric and our parametrization of the stress-energy tensor into Eq. 1.1, which gives a special case of the Einstein equations called the Friedmann equations (assuming a flat Universe)

$$H^2 \equiv \big(\frac{\dot{a}}{a}\big)^2 = \frac{8\pi G}{3}\rho \tag{1.10}$$

$$2\frac{\ddot{a}}{a} + \big(\frac{\dot{a}}{a}\big)^2 = -8\pi G p \tag{1.11}$$

Where we've introduced the Hubble parameter H. Of particular interest to us is Eq. 1.10, which is sometimes referred to as *the* Friedmann equation. Using Eq. 1.9 we find the solution to Eq. 1.10 is

$$a(t) \propto t^{\frac{2}{3(1+w)}} \tag{1.12}$$

Nte that this solution is valid for $w \neq -1$. For the case $w = -1$ we see from Eq. 1.9 that the energy density is independent of the scale factor

$$\rho(t) = \rho_0,\ w = -1 \tag{1.13}$$

Then the solution to Eq. 1.10 in this case is

$$a(t) \propto \exp\left(t\sqrt{8\pi G\rho_0/3}\right),\ w = -1 \tag{1.14}$$

We have implicitly assumed that everything in the Universe can be described as a perfect fluid with a single equation of state. The Universe is diverse, and we need to consider a mixture of several components for a detailed description. The components of interest are:

- Matter, $w = 0$
- Radiation, $w = 1/3$
- Cosmological Constant (Λ), $w = -1$

In the context of cosmology, matter and radiation are distinguished by whether the particle is relativistic or not (i.e a relativistic species satisfies $p \gg m$). The energy density of matter scales like $\rho \sim a^{-3}$, i.e. it is proportional to the particle number density. For radiation we have $\rho \sim a^{-4}$, where the extra factor of a^{-1} is introduced by redshift. The cosmological constant represents some "Dark Energy" whose energy density is independent of the scale factor. Hence as the Universe expands, more energy from Λ is added as well, further increasing the rate of the expansion. The physical origin of Λ remains one of the major unanswered questions in the field.

When comparing the prevalence of multiple components, it is convenient to express their abundance over time with respect to the present abundance and the "critical density" of the Universe.

$$\rho_c \equiv \frac{3H_0^2}{8\pi G} \tag{1.15}$$

$$\Omega_i \equiv \frac{\rho_i}{\rho_c} \tag{1.16}$$

Then, using the expressions we found relating the ρ to a for each component, we can rewrite Eq. 1.10 as

$$\frac{H^2}{H_0^2} = \Omega_{0,R} a^{-4} + \Omega_{0,M} a^{-3} + \Omega_{0,\Lambda} \tag{1.17}$$

Where the 0-subscript indicates the value at the present day (when $a = 1$), and the remaining subscripts (R, M, Λ) stand for radiation, matter and cosmological constant respectively. Under our assumption of a flat Universe, the density parameters must satisfy

$$\Omega_R + \Omega_M + \Omega_\Lambda = 1 \tag{1.18}$$

This says that the critical density ρ_c is equal to the average density of the Universe. Eq. 1.17 demonstrates our framing which compares the study of Cosmology to the conducting of a census. If we can poll the constituents of the Universe by measuring each Ω_i, then we arrive at a "complete"[1] description of the evolution of the scale factor for all times. One instructive feature of Eq. 1.17 is that it gives us a brief history of the Universe. At very early times most of the energy in the Universe was due to radiation, and during this period the the scale factor evolved as $a(t) \propto \sqrt{t}$. Extrapolating back to $t = 0$ implies that the scale factor vanished at this point. This picture of an ever expanding Universe originating from a singularity is generically referred to as Big Bang Cosmology. After the period of radiation came matter domination, and in the present day the energy density is mostly due to the cosmological constant. Since the radiation component has a well defined temperature $T \propto a^{-1}$, we conclude that the Universe was very hot in the past when $a \to 0$, and it has been cooling and expanding ever since.

1.2 ΛCDM and The Cosmic Microwave Background

We have motivated some of the central tenets of the standard model of cosmology. Mainly that the Universe started as a very hot singularity which has been expanding and cooling ever since, and that the rate of this expansion is determined by the contents of the Universe as described by the Friedmann equation.

To proceed we need to take a closer look at the matter piece. We know that there exists matter in the

[1] Modulo cosmic inflation

Universe from the constituents of standard model of *particle physics* (SM), because we exist. But this does not necessarily account for all the matter we see. The standard model of *cosmology* states that some fraction of the matter in the Universe must have the following qualities:

1. Has an extremely weak electromagnetic coupling
2. Does not decay to lighter particles on cosmological time scales
3. Was non-relativistic at the time of matter-radiation equality

The first condition leads us to call this matter "dark", and the third makes it "cold". With this piece we can call the standard model of cosmology by its proper name - "ΛCDM", with Λ representing dark energy and CDM for "Cold Dark Matter".

If we were to consider candidates from the SM which could satisfy these conditions, only the neutrino species pass the first two points. However from an investigation of neutrino cosmology, one finds that they were in fact relativistic at the time of matter-radiation equality and fail. Hence no particle identified to this day satisfies the necessary conditions to play the role of CDM.

The matter content of the Universe should then be partitioned between the baryonic matter and CDM.

$$\Omega_M = \Omega_B + \Omega_{DM} \tag{1.19}$$

With this piece we can call the standard model of cosmology by its proper name - "ΛCDM", with Λ representing dark energy.

We have now listed all the constituents we need to count in order to take a census of the Universe. To carry out this task, we must look for relics of the early Universe which we can compare to the present day. The Universe started out very hot and very dense, such that all of the constituents were relativistic, and all available degrees of freedom were excited an effective thermal bath. As the Universe cooled, the bath temperature eventually fell such that each of these degrees of freedom were no longer excited, and they effectively "froze-out" of their coupling to the bath. There is only one species we can directly observe to this day in the present day which has not scattered since freeze-out – the photon.

In the period between the formation of protons and hydrogen atoms, the mean free path of a photon was very short, limited by scattering off free electrons. The Universe in this era was hence an opaque plasma. At $z \sim 1100$, the Universe cooled sufficiently such that the free electrons could bind to protons. When this

transition occurred, the era of Recombination, the mean free path of the primordial photons increased to effectively infinity. This infinite streaming length means that we can still observe these photons. This relic is named the Cosmic Microwave Background (CMB), since the light we observe today is in the microwave region with a mean temperature of 2.73K [2]. The temperature measurement alone lets us check off a box for our census by measuring the giving us the radiation density of the Universe[2], since the internal energy of a black body is completely determined by its temperature.

$$\rho_R \approx \frac{\pi^2}{15} T_{CMB}^4 \rightarrow \Omega_{0,R} \approx 5.38 \times 10^{-5} \tag{1.20}$$

By observing the CMB we are essentially looking at a snapshot of the Universe right before it became transparent. When we measure the temperature of the CMB, we find it is remarkably isotropic, with temperature deviations $\sim 10^{-4}$ parts from the mean. This gives us a strong argument for the validity of the Cosmological Principal. It is however not perfectly isotropic, and any model of the Universe we propose must explain the observed lumpiness, as well as how those lumps propagated to form the structure of galaxies and galaxy clusters we see today.

To quantify the "lumpiness" of the CMB, we measure it's angular power spectrum. This is done by first decomposing the temperature map of the CMB into spherical harmonics

$$T(\theta, \phi) = \sum_{\ell,m} a_{\ell,m} Y_{\ell,m}(\theta, \phi) \tag{1.21}$$

Then the power of a particular ℓ-mode is given by

$$C_\ell = \langle |a_{\ell,m}|^2 \rangle \tag{1.22}$$

Looking at the power spectrum, we see a number of peaks, the structure of which is largely determined by the speed of sound in the plasma just before Recombination. We might expect a steady decrease in the peak amplitude, but rather observe that the odd number peaks appear to have a relative enhancement compared to the even ones. This feature can be explained by a massive species which was present at Recombination, but not coupled to the plasma. This species would have already began to cluster gravitationally, creating potential wells which drove oscillations in the plasma [3].

[2]Given that the the radiation content of the Universe is dominated by the CMB, which turns out to be a good assumption

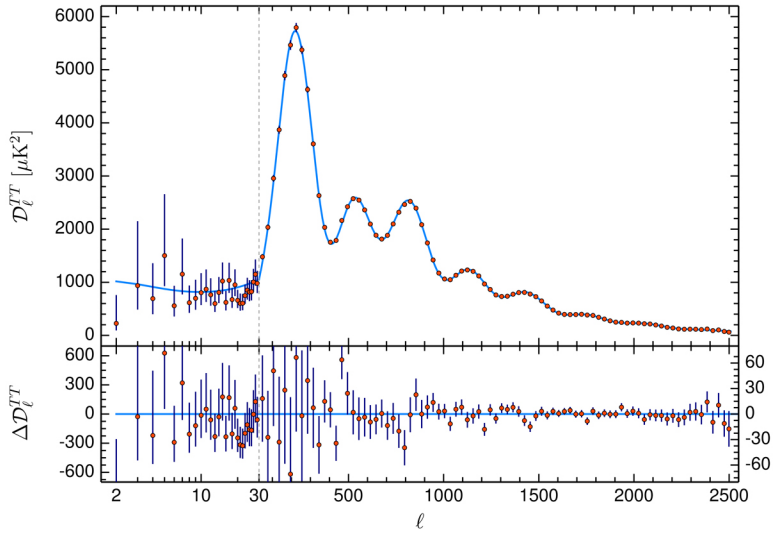


Figure 1.1: Measured CMB power spectrum from [1] (red points), along with the resulting fit to ΛCDM (blue curve).

The spectrum measured by the *Plank* satellite is shown in Fig. 1.1. Fitting the ΛCDM model to the CMB power spectrum requires at least six parameters, including Ω_{DM} and Ω_B. The results of this fitting shows remarkable agreement between the fit and the data. This gives us strong evidence that we have a model which well describes our snapshot of the early Universe, and this model requires the existence of some species of matter which is not contained in the SM.

Density Parameter	Value
Ω_Λ	0.685 ± 0.007
Ω_B	0.0493 ± 0.0006
Ω_{DM}	0.265 ± 0.007

Table 1.1: Density parameters measured by Planck [1]

1.3 Something Doesn't Add Up: The Case of the Missing Mass

In the previous section we argued that the CMB gives us evidence for the existence of DM not described by the SM. Of course, if this DM really exists and doesn't decay or interact readily we don't expect it to go anywhere between recombination and now. Surely enough, if we look at a generic galaxy we can see evidence of DM there too! Hence, in addition to the remarkable fit to the CMB we find in a model which includes

DM, its existence is further supported by observations on smaller scales.

In a gravitationally bound system, the relationship between the velocity v of an object orbiting at radius r and the mass contained within that radius $M(r)$ is clear from the Newtonian interpretation of Kepler's third law

$$v = \sqrt{GM(r)/r} \tag{1.23}$$

Hence we can determine the mass profile of a galaxy by measuring the velocity of many objects contained within that galaxy, and then plotting those velocities versus the radial distance from the galactic center. These velocities can in principle be measured by observing a known spectrum produced by the stars in the galaxy, and then determining the Doppler shift.

If we start from the hypothesis that all of the mass in a galaxy is baryonic, then it follows that most of that mass would be contained within stars. The number density profile of the stars in a galaxy can be estimated from the luminosity profile. When we go and perform this luminosity measurement on a number of different galaxies we see a clear trend – the vast majority of the stars are clustered near the center of the galaxy and there is a sharp drop-off in luminosity outside this core. This implies that at $r > r_{core}$, $M(r) \sim$ constant. Then from Eq. 1.23 we would expect $v \propto r^{-2}$ at large radii.

When we actually measure the velocity profile of many galaxies, we observe a very different pattern than expected from the luminosity. Rather than falling off, the velocity typically plateaus at some non-zero value. This behavior for several galaxies is shown in figure Eq. 1.2. This obvious contradiction from the initial prediction suggests that our initial assumption, that the mass of the galaxy is dominated by baryonic matter, is incorrect, and we typically resolve this contradiction by concluding that the galaxy contains a large portion of dark matter.

We can apply this same argument to the Milky Way by measuring its rotation curve. As a result of this measurement, we conclude that the local DM density near the Earth is $\rho_{DM} \approx 0.3$ GeV/cm^3 [5]. If the DM is a "halo" of particles which are approximately stationary with respect to the galactic center and the Earth is moving at velocity $v_0 \approx 230$ km/s with respect to this halo, then we expect an appreciable flux of DM particles to be constantly passing through the Earth. And if we set up a detector in our lab which sufficiently sensitive to the possible rare interactions between the DM and SM, we might even expect to see these interactions! Hence it is reasonable to expect that we might be able to detect and identify DM in our own backyard.

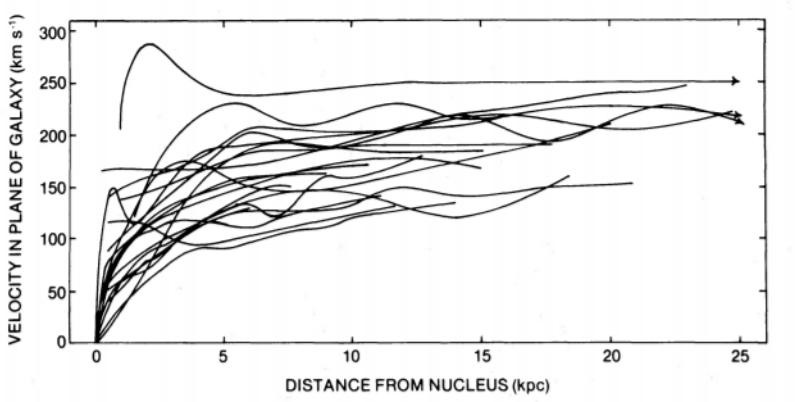


Figure 1.2: Rotation velocity curves from a sample of 21 spiral galaxies. Here we see that the velocity of a particular galaxy is roughly constant above ~ 5 kpc. Image from [4].

Chapter 2

Dark Matter Candidates

We have now argued that the Universe is populated by some DM, which is likely a particle not contained in the SM and yet to be identified. Now we must give some consideration to what this particle could reasonably be. Since we haven't directly observed this particle yet, our vetting process for potential DM candidates will naturally include some combination of theory motivation and speculation.

At the end of the day, we don't have that many requirements which our DM candidate must satisfy. The DM must be stable enough that it hasn't decayed already, and any coupling it has to the SM needs to be weaker than the EM interaction. It also must have been cold enough during the era of recombination to seed the observed structure of galaxies/clusters, and the amount of it needs to match the results of our Universal census. Our ability to generate models which satisfy these conditions is in some sense only constrained by the creativity of the theorists we know. We will discuss the construction of some viable theories at the end of this chapter, but our goal in this thesis is oriented around the fact that these theories exist rather than advocating for the likelihood of any particular one.

Most importantly, we don't know the mass m_χ or the cross section σ_0 of interactions between DM and SM. Our very first goal in searching for DM particles is to constrain these two parameters through a positive detection. Once this goal has been accomplished, one can use this information to discriminate between different theories of the DM identity. This will guide our theoretical discussion. We want to minimize our prejudice to particular theories at this point, but derive utility from the discussion by identifying reasonable places to search for DM in (m_χ, σ) space.

2.1 Recipe For the Present Density

Before we discuss any particular models, which may require us to make some assumption about the distribution of the DM in the early Universe, let us attempt to use what we know about the DM distribution today and extrapolate back in time. At some point in the history of the Universe, which we will characterize by temperature T_f, the total number of DM particles present became approximately fixed. Once this number is set, the number density of the species n_χ is determined solely by the evolution of the scale factor. Furthermore, we can use the fact that this expansion occurred such that entropy per comoving volume is conserved. With this we find

$$n_{\chi,0} = n_{\chi,f} \frac{g_{\star s,0}}{g_{\star s,f}} \left(\frac{T_0}{T_f}\right)^3 \tag{2.1}$$

where $g_{\star s,i}$ is the effective number of entropic degrees of freedom in the bath following the Big Bang at temperature T_i. The canonical story of our Cosmology assumes $g_{\star s,0}/g_{\star s,f} \sim 1\%$. Then once we relate the present average number density of the DM to the observed relic density, we arrive at the constraint

$$\Omega_{DM} = n_{\chi,f} \frac{m_\chi}{\rho_c} \frac{g_{\star s,0}}{g_{\star s,f}} \left(\frac{T_0}{T_f}\right)^3 \tag{2.2}$$

Regardless of the particular nature of the DM, this condition must be satisfied in order to see DM density we observe in the present day. We also note that this statement is also dependent on $g_{\star s,f}$.

What happened between the $T = 0$ and $T = T_f$ strongly depends on which particular set of assumptions we make. A lot has happened in the history of the Universe, and we can only glean evidence from a few particular moments in its evolution. Furthermore, we're trying to deduce what happened to a particle species we have yet to directly observe. Naturally, our story of how the DM got to T_f will include some combination of model building and speculation.

There are a number of reasonable questions we can ask about the DM besides the value of its mass and cross section, the answers to which will change its history:

1. Is the DM particle a boson or a fermion?

2. Is the DM population composed of a single particle species, or is it a collection of several?

3. If the DM couples to the SM, what particle mediates this interaction? How does the mass of the

mediator compare to the DM mass? Could this mediator be the DM itself?

4. Was the DM in equilibrium with the thermal bath following the Big Bang, or was its present abundance generated at later times?

Iterating over each of these questions with varying assumptions generates a huge number of pictures of DM, a handful of which might be viable from phenomenology point of view. For our purposes, rather than attempting the herculean task of exploring all of these iterations, we will focus for now single point. In particular, item 4 above has perhaps the most straightforward connection to the question of what happens before $T = T_f$. At the end of the chapter, we will take a broader look at the field of DM models and where we stand from an experimentalist's perspective.

2.1.1 Freeze-out Production

We will start by assuming that the DM is produced in abundance during the Big Bang, and is in equilibrium with the bath in the early Universe. This is reasonable assumption, since it puts the DM on somewhat of an equal footing with the rest of the SM. As we will discuss later, this scenario typically applies to models where the DM can be described as a Weakly Interacting Massive Particle (WIMP), which have played a unique roll in the history of the study of DM.

Some interaction is required in order for the initial equilibrium between the DM and SM to be maintained for a non-zero amount of time. For now we will restrict our picture to the case where these interactions look like $A\bar{A} \leftrightarrow \chi\bar{\chi}$, where A is some member of the SM and χ is our DM particle. That is, we only consider annihilations of SM particles to DM or vice versa. In the generic case, one should also account for the possibility of heavy SM particles to decay to DM ($A \rightarrow \chi\bar{\chi}$).

Generically, we expect the number density of the DM to evolve according to the Boltzmann equation

$$\frac{dn}{dt} + 3Hn = \left(n_{eq}^2 - n^2\right)\langle\sigma v\rangle \tag{2.3}$$

where n is the number density of the DM particle, n_{eq} is the number density at equilibrium, and $\langle\sigma v\rangle$ is the average annihilation cross section. We can view 2.3 as a kind of continuity equation applied to the DM number density. The H term on the left side accounts for the dilution of the number density due to the expansion of the Universe. On the right hand side, the positive term counts the DM particles generated

by annihilations of SM particles and the negative term represents DM annihilations which reduce the total number. The interactions which keep the DM in equilibrium occur at a timescale determined by the rate

$$\Gamma \approx \langle \sigma v \rangle n \tag{2.4}$$

Which follows from dimensional analysis. Notice that this is almost proportional to the right hand side of 2.3. On the other side of 2.3, we see that the expansion dilutes the number density. This subsequently decreases Γ, until eventually the rate is not sufficient to maintain equilibrium. A precise determination of when this occurs comes from solving 2.3, usually numerically. We can also apply a more intuitive estimate by assuming that this happens when the expansion rate matches the interaction rate.

$$\Gamma_{fo} \sim H_{fo} \tag{2.5}$$

This is the so-called freeze-out condition, determining the temperature below which the DM is no longer in equilibrium with the bath. We know that the total number of DM particles should be fixed before the period of Recombination, hence the Universe is radiation dominated during freeze-out and the expansion goes as

$$H_{fo} \sim \frac{T_{fo}^2}{m_P} \tag{2.6}$$

Where m_P is the Planck mass. Using this result in the freeze-out condition and combining it with 2.2 allows us to relate the present density to the cross section

$$\Omega_{DM} \sim \frac{1}{\rho_c} \frac{m_\chi}{T_f} \frac{g_{\star s,0}}{g_{\star s,f}} \frac{T_0^3}{m_P} \frac{1}{\langle \sigma v \rangle} \tag{2.7}$$

Notably, the present relic density is proportional to the *inverse* of the cross section. A weaker DM interaction results in a larger relic density. To evaluate this numerically, we must estimate m_χ / T_f. Given that the DM is non-relativistic ($m_\chi \gg T_f$) and in equilibrium before freeze-out, its number density can be written

$$n_f \approx \left(\frac{m_\chi T_f}{2\pi}\right)^{3/2} e^{-m_\chi / T_f} \tag{2.8}$$

A natural consequence of the non-relativistic assumption is that the behavior n_f is largely dominated by the exponential factor. Using this fact, along with some reasonable parametrization of $\langle \sigma v \rangle (m_\chi, T)$ [6] in 2.5,

one concludes

$$T_f \sim \frac{m_\chi}{20} \tag{2.9}$$

Due to the exponential behavior of 2.8, corrections to this relationship are only logarithmic in m_χ. Hence we generically expect the DM to freeze-out roughly a decade after $T = m_\chi$. With this, we have everything we need to evaluate 2.7 to arrive at an estimate of $\langle \sigma v \rangle$ which produces the present density Ω_{DM}. We find

$$\langle \sigma v \rangle \sim 3 \times 10^{-28} \text{ cm}^3/\text{s} \tag{2.10}$$

Remarkably, this is the same scale would expect for the Weak interaction, which motivates the name of the WIMP.

2.1.2 Freeze-in Production

While it is a natural assumption that the DM population was in equilibrium with the bath after the Big Bang, it is not one we are forced to make. We briefly attempted to motivate this choice by appealing to the notion that it places the DM and SM on equal footing. One could argue that, though this makes to the DM look more familiar to us, it is not necessarily well motivated. The SM and DM clearly play different roles in the ΛCDM model and they don't readily talk to each other, so why would we be so certain that they belong on an equal footing?

Let us consider the other extreme extreme of the spectrum. Assume that the DM has a negligible abundance in the early Universe, and the population we observe today was generated by a weak coupling to SM particles. We can study this scenario in a matter which is totally analogous to freeze-out case by simply neglecting the DM annihilation process. That is, the only interactions we consider are $A\bar{A} \to \chi\bar{\chi}$. Then the Boltzmann for this case looks like [7]

$$\frac{dn}{dt} + 3Hn \approx n_{eq}^2 \langle \sigma v \rangle \tag{2.11}$$

In order for this production to be insensitive to unknown physics in the very early Universe, we restrict ourselves to the case where the DM-SM coupling proceeds through renormalizable interactions. This is naturally accommodated if we assume that the mass of the mediator of the DM-SM interactions is very small compared to the DM mass, and then the production is dominated by the era $T \sim m_\chi$ [8, 9]. The form of 2.11 is only valid if the number density is never large enough for the DM annihilation process to become

significant, though in that case one could simply refer back to 2.3 to determine the evolution. DM models where 2.11 is applicable are thus sometimes refered to as Feebly Interacting Dark Matter (FIMP). These are distinguished from WIMPs by the fact that their coupling to the SM must be even weaker, such that they can never reach equilibrium with the bath [8].

We can immediately draw a conclusion from this which diverges from the freeze-out scenario. Namely, a larger coupling $\langle \sigma v \rangle$ results in a *larger* value of n_f for a given mass m_χ. This result is fairly intuitive. In the freeze-out case, a larger interaction means that the DM can remain in equilibrium for longer and hence the density has more opportunity to decline according to $n \propto e^{-m_\chi/T}$. For the freeze-out case, there is no equilibrium to maintain, and increasing the cross section simply increases the DM production rate.

Another comparison we can make between the freeze-in/out scenarios is how the temperature T_f, at which the total particle number is determined, changes. In the previous section, we discussed that DM typically freezes out about a decade after the temperature of the bath equals its mass. We can understand this by a similar argument to our last point. Since the remaining SM bath particles are all lighter than the DM, they no longer have sufficient energy to facilitate production of the DM via $A\bar{A} \to \chi\bar{\chi}$ as soon as $T \lesssim m_\chi$. But in the freeze-out case, the DM may still have sufficient energy to undergo the $\chi\bar{\chi} \to A\bar{A}$ process. Hence the DM particle number continues to change until the freeze-out condition is met about a decade later. When we consider the freeze-in case, the $\chi\bar{\chi} \to A\bar{A}$ process is always assumed to be negligible. From this, we conclude that the DM particle number is fixed by $T_f \sim m_\chi$ for the freeze-in scenario. In short, we generically expect freeze-in to occur earlier than freeze-out [7].

2.1.3 Everything in Between

We have just discussed two simple methods for the generating the observed abundance of DM in the universe under two very different sets of initial conditions. For the sake of simplicity, we imposed some fairly restrictive assumptions on this discussion. Mainly we only considered processes with a $2 \to 2$ topology, and we assumed that the DM population contains a single unique particle. The former assumption is incorrect for many scenarios. If the DM has any coupling to the SM and is lighter than half the Z-boson, we expect $Z \to \chi\bar{\chi}$ to have some contribution to the DM population [10].

We can also consider the case where other event topologies are important. For example, if the DM is bosonic then we might expect processes of the form $\chi\chi\chi \leftrightarrow \chi\chi$ to play an important role in the early Universe, along with its coupling to the SM. It is clear that the Boltzmann equation as we've written it is

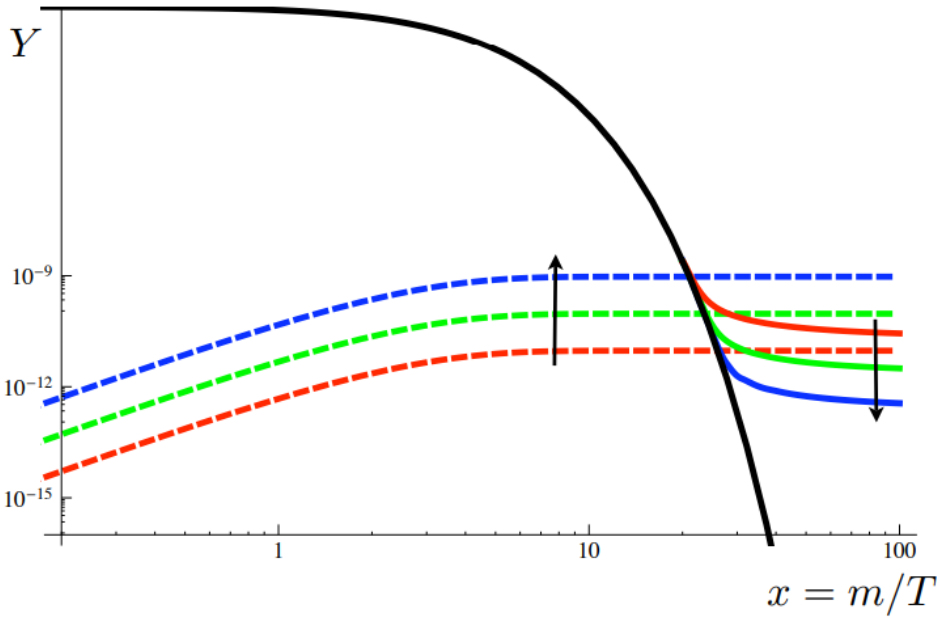


Figure 2.1: Evolution of the DM abundance $Y \propto n_\chi/T^3$ as a function of the age of the Universe, parametrized via m_χ/T. Solid lines indicate freeze-out scenarios, where the DM begins in equilibrium with the bath, and dashed lines represent the freeze-in case where the initial DM abundance in negligible. Arrows illustrate the effect of increasing the interaction strength in each case, which results in a smaller relic density for the freeze-out case and a larger one for the freeze-in case. Figure from [8].

not well equipped to handle this scenario, and one should consider a more detailed collision term along with entropy conservation in order to determine its evolution. This scenario applies to closely related DM models of the Strongly Interacting Massive Particle (SIMP) [11, 12] and the ELastically DEcoupling Relic (ELDER) [13]. A high level discussion of these models, including how the DM mass and coupling strength produce the relic abundance, is given in [14].

Our discussion up to this point has also been restricted to the case where there is a single unique BSM particle involved in the generation of the DM. This may not be the case in reality, and DM could be a single component of some larger "dark sector". We will discuss what we might expect this dark sector to look like in more detail in the following section, but for now we note that the existence of the larger sector can easily alter the freeze-in/out picture if thermal equilibrium established separately within each sector. For a minimal example involving two BSM particles, Ref. [8] considers the case where a FIMP freezes-in before decaying to some lighter particle which makes up the observed DM.

For our final alternative to the freeze-in/out picture, we first note that from the previous sections we conclude that the relic abundance is determined by the DM mass and coupling strength to the SM - it is insensitive to the initial conditions set in the early Universe. This is a nice assumption since there is some predictive power in how the relevant parameters we aim to determine should relate to explain the Universe we observe. On the other hand, we aren't restricted to this assumption. We know that the abundance of baryons today day was determined by some matter-antimatter asymmetry set in the early Universe. This, along with the fact that the DM abundance is comparable to that of baryons ($\Omega_{DM} \sim 5\Omega_B$, see Table 1.1), motivates the hypothesis that their abundance may in fact be closely linked. In these models, the present DM abundance is thus inherited from some initial asymmetry rather than the particular strength of its coupling to the SM [15].

2.2 A Menu of DM Candidates

There are a number of different paths the DM could have taken to get where it is today and, and not too many requirements our DM model needs to satisfy. If we try to construct a model at the Lagrangian level, our only strict requirement is that we need to include a mass term for the DM, and then from considerations of the observed relic density we can motivate the inclusion of a term which couples the DM to the SM. Within these confines, we can potentially generate an infinite number of DM models, a more manageable

- but still large - subset of which will exhibit phenomenology compatible with present observations. Given this large potential model space, there are two philosophical approaches one can adopt

1. **Top-Down Approach**: We motivate a particular DM model by choosing one which arises naturally from a proposed solution to a distinct problem. Prominent examples include DM arising from Super Symmetry (SUSY), a theory closely related to resolving the "Hierarchy Problem" [16] - or axion DM, where the DM particle could also explain the vanishing neutron electric dipole moment [17].

2. **Bottom-Up Approach**: We allow our model to be more *ad hoc*, and consider it to be well motivated as long as it resolves the DM problem without making assumptions which we might deem unreasonable.

At a first glance, one might wonder why we'd ever prefer a bottom-up compared to a top-down approach. This view has certainly colored a large swath of history in the search for DM, where the fact that SUSY predicts new physics at mass and interaction scales compatible with WIMP DM has served as one of the leading hypotheses to explain the nature of DM [18]. Though certainly well motivated, we have yet to directly observe any DM candidate, and searches for SUSY particles at the LHC have only yielded null results. This is not to say that we should discard the hypothesis that the DM we see is made up of SUSY WIMPs. There still remains viable parameter space we have yet to constrain [19], and abandoning the theory is premature as long as this remains the case. But the current state of the field may perhaps serve as a warning against placing all of our "eggs" - in this case our theoretical and experimental effort - in a single basket corresponding to one theory. Despite how well motivated a particular explanation for DM might appear to us, we shouldn't get too far ahead of ourselves. All of our understanding of the physical world is ultimately derived from our measurements and observations of it, rather than a particular aesthetic.

Let us consider an example of a "bottom-up" model as an alternative. Suppose that the DM we see is actually part of some larger "hidden sector". One simple benchmark model for such a sector is one that looks just like QED [20]. We do this by introducing a new gauge group $U(1)_D$, under which the fermionic DM particle χ is charged. The breaking of $U(1)_D$ results in the massive vector boson A', dubbed the dark photon. This new sector can couple with the SM via the kinetic mixing portal. In particular we expect the Lagrangian to contain the term [21]

$$\mathcal{L} \supset -\frac{\epsilon}{2} F_{\mu\nu} F'^{\mu\nu} \tag{2.12}$$

Where F is the SM electromagnetic field, F' is the $U(1)_D$ gauge field and ϵ is the kinetic mixing parameter between the sectors. Through this kinetic mixing with the SM photon, A' is able to effectively couple with

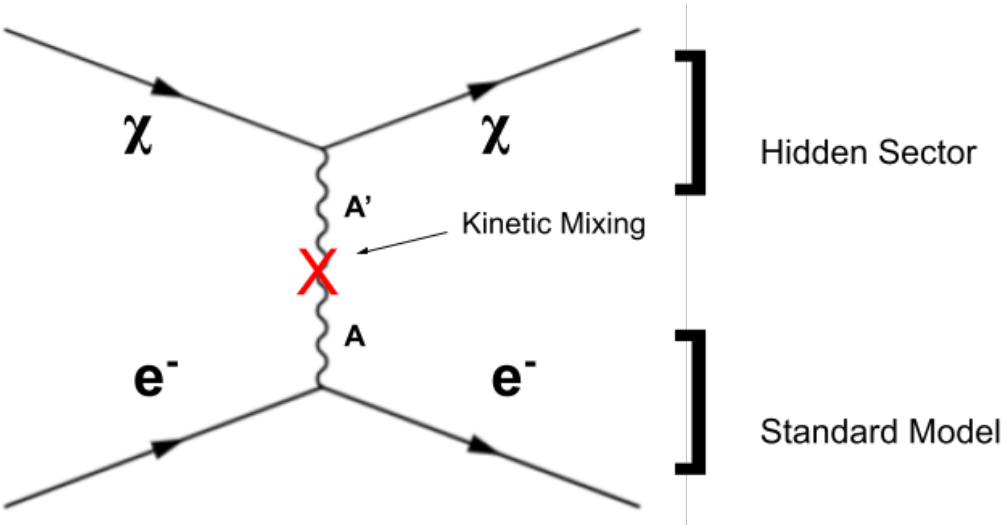


Figure 2.2: Toy diagram of a DM particle χ scattering off an electron via exchange of dark photon A' which kinetically mixes with the SM photon A.

the SM photon, and could we could detect χ interaction through this channel as illustrated in Fig. 2.2.

Given this model, there are a number of ways to generate the relic density depending on ϵ and the mass hierarchy $m_\chi/m_{A'}$, the details of which can be found in Ref. [20]. We can consider this a bottom-up model since we effectively wrote a Lagrangian with the intention of satisfying our list of requirements for the DM. Note however that the scenario we described arises from String Theory [22], so one may actually consider this to be a top-down approach if one finds String Theory to be well motivated. Even in the case where we admit that this model is *ad hoc*, it is not necessarily unreasonable. The SM is really the only model of particle physics which we know works. On the other hand, we know that whatever the nature of DM is, it must be described by some Lagrangian. Given this, one could consider it natural to assume that the Lagrangian describing DM would look a lot like some a of the SM Lagrangian.

We have now touched on a number of models which could explain the population of DM we observe. To put this all into context, we look at a comparison of these made in [23] as shown in Fig. 2.3. More details of the particular DM models we've discussed can be found in the references therein. Depicted here is a huge amount of space in the (m_χ, σ) plane. We highlight the region 10 MeV $\lesssim m_\chi \lesssim$ 10 GeV, as searches for DM in this mass range are the focus of this thesis. Notice that the so called "WIMP Miracle" appears in this plot as the overlap between the regions labeled "thermal" and "NMSSM" (Next-to-Minimal Supersymmetric Standard Model), but this is only a small area of the potential model space. Notably, the "hidden sectors"

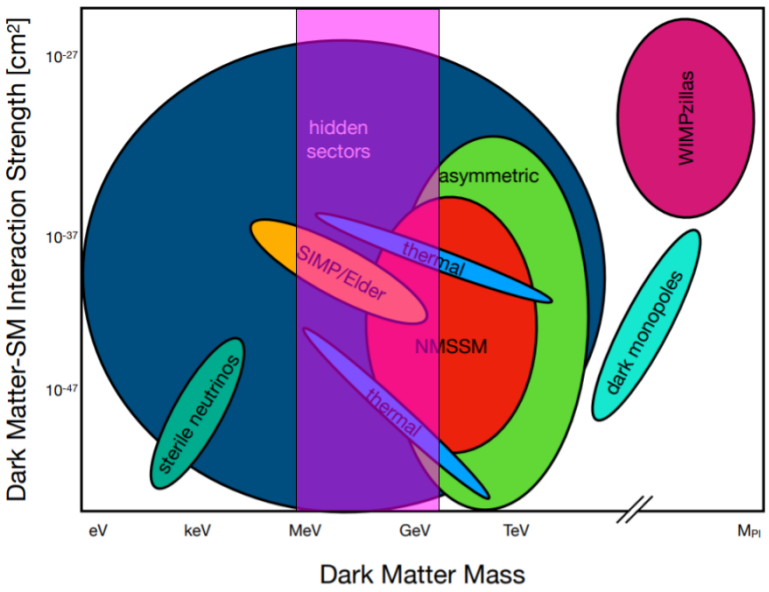


Figure 2.3: Cartoon of the landscape of models probed by direct detection experiments. Note that the interaction strength scale lumps together the possibility to scatter off of both nuclear and electron targets. We highlight the MeV-GeV mass range, where there are several models which current and near-future direct detection experiments can target. We also draw attention to the two regions in the plot labeled as "thermal". The top region refers to thermal freeze-out scenarios and the bottom to thermal freeze-in. i.e. the DM is not thermal itself, but is generated by interactions from the thermal bath. Figure adapted from Ref. [23].

region encompasses the entire mass range we've highlighted. We will refer to the region of parameter space < 1 GeV as "Light Dark Matter" (LDM), to distinguish it from the traditional WIMP parameter space where experiments such as SuperCDMS have historically focused their efforts.

When all the dust is settled, we don't have a unique model to describe the abundance of DM in the Universe, but rather a handful of viable ones. The point of this thesis is rather to discuss a few particular experiments which have sufficient sensitivity to be good candidates for detecting DM in the MeV-GeV mass range. With these considerations in mind, it is often prudent from an experimentalist's point of view to take a "shoot first, ask questions later" approach to the DM problem. If we can build a sensitive enough experiment, we can reasonably expect to observe *some* kind of signal since there's so much DM out there. Once we have identified a signal, we can start to worry more about particular models.

A similar, albeit less violent, version of this strategy is referred to in [23] as "delve deep, search wide". While it is worth investing more effort into exploring DM the models which we think are best motivated, we also want to cover enough ground that we don't paint ourselves into a corner. A crucial consequence of this paradigm is that we cannot reasonably expect a single experiment to do both. A number of claims of a detected DM signal have occurred in the history of DM searches [24–26], but reproducibility is the name of the game with the scientific method. Subsequent experiments with improved sensitivity have failed to observe a consistent signal [27, 28]. A single experiment cannot positively identify DM with confidence, so the search for DM is really a community effort which must include a diverse set of experiments exploring a

wide range of parameter space.

Chapter 3

Dark Matter Direct Detection

3.1 Considering All Avenues: Detection and Production of Dark Matter

We have argued that we have good reason to suspect plenty of DM in our Universe, and that it is reasonable to expect it to look like a particle which has a small but non-zero coupling to SM particles. Now all we have to do is find it and identify the particle! There are many potential avenues to accomplish this, which we can broadly classify as falling into one of three categories [29]:

1. Direct Detection: We look for DM in the local halo interacting with the SM by setting up a detector on the Earth. Since the solar system moves with a relatively large velocity with respect to the halo, we expect a considerable flux to pass through the detector which increases our chances of an observation.

2. Indirect Detection: We look for DM in some galaxy by looking for the SM byproducts of DM-DM annihilations. One possibility for this avenue is to view the center of a galaxy, where we expect a large fraction of the DM in the halo to cluster with appreciable kinetic energy, as a laboratory to facilitate these interactions [30].

3. Collider Production: We can look for DM in the byproducts of SM-SM annihilations in colliders. Since the produced DM will only weakly interact with the detectors which measure the collision products, the DM would appear to us as some missing momentum [31].

As discussed in the previous chapter, there are a large number of models containing a massive amount of parameter space where the DM could be hiding. None of these three pathways for DM identification should be considered "better" than the others, since a diverse toolbox is our best hope of covering this space. That being said, we should also note that it is difficult to compare the results from searches populating different branches. Each school of identification requires different assumptions about the DM distribution. For example, direct detection experiments must know the *local* DM density, indirect experiments may depend on the density at the center of some galaxy, and production experiments are naively independent of any particular DM halo. Our estimate of the DM density at a particular point in a halo is ultimately derived from a measurement of the rotation curve, and carries a sizeable systematic uncertainty. For one thing, our measurement of the local density in the Milky Way depends on our own peculiar velocity, while the measurement of a distant galaxy will be relatively insensitive to this. Since each branch of experiment carries its own distinct systematics, it is difficult to directly compare results from different branches. This really just brings us back to our "shoot first" model for discovering DM. First we need an experiment sensitive enough to see *something* consistent with DM. Next, other sensitive experiments in the same branch can validate whether they see something similar. Finally, we can validate if the supposed DM they observe DM occupies parameter-space accessible to other branches, where they could be independently confirmed. Only after all these steps have taken place can we confidently say that we've identified DM. This thesis will focus on direct detection experiments.

3.2 DM Direct Detection Rates

Before we can make any attempt discover DM or exclude any particular model using data from a DM search, we must know how many interactions we expect in our detector. We can immediately write an estimate of the interaction rate per target mass from dimensional analysis

$$R \sim \frac{N_A}{A}\sigma_{0,N}\langle v\rangle n_0 \tag{3.1}$$

Let us get a quantitative sense of this rate. We assume a Ge target with $A = 72.6$ amu. Then the average velocity of the take $\langle v\rangle \sim 220$ km/s from the galactic velocity of the sun, $\sigma_{0,N} \approx 10^{-36}$ cm^3 from the scale of the weak interaction. Assuming a WIMP mass $m_\chi = 10$ GeV/c^2, along with the estimated local halo density $\rho_0 = 0.3$ GeV/c^2/cm^3, we arrive at a rate $R \sim 10$ events/kg/year . The small size of this event rate demonstrates the great effort which direct detection experiments must put into background reduction.

This just gives us an estimate. We typically want to know the differential rate as a function of energy, so that we can properly account for the threshold of the detector and perform more powerful statistical tests when we analyze our spectrum. The rate can be described in terms of the differential number density dn via [5]

$$dR = \frac{N_A}{A}\sigma_{0,N}F^2(q)vdn \tag{3.2}$$

where A is the atomic mass of the target, N_A is Avogadro's number, and $\sigma_{0,N}$ is the cross section at zero momentum transfer for the DM to interact with a target nucleus. $F^2(q)$ is the momentum-dependent nuclear form factor dependant on the recoil energy.

The nuclear form factor roughly accounts for the finite size of the target nucleus. If the de Broglie wavelength of the DM is large compared to the size of the nuclear radius, then the target effectively appears coherently as a point particle and $F(q) \approx 1$. At smaller wavelengths, the structure of the nucleus becomes important and the interaction cross section no longer benefits from coherence such that $F(q) < 1$. Considering that the DM is non-relativistic, we can write its wavelength as

$$\lambda_\chi \approx \frac{h}{m_\chi v_0} \tag{3.3}$$

Since the relevant velocity $v_0 \approx 10^{-3}c$ is determined by the motion of the Earth for all terrestrial experiments, λ_χ is essentially defined by m_χ. For the primary searches of interest for this thesis are in the regime $m_\chi < 10$ GeV/c^2, we have $\lambda_\chi \gtrsim 125$ fm. Regardless of the particular target, we expect $r_N \lesssim 10$ fm. Hence $F(q) \approx 1$ is a good approximation for the results discussed in this work.

A consequence of the argument that we expect DM to scatter coherently off the nucleus in the low-mass limit is that we expect $\sigma_{0,N}$ to scale with with the size of the nucleus. In order to directly compare results from different targets, it is often preferable to describe the interaction in terms of the DM-*nucleon* cross section, $\sigma_{0,n}$. This can be related to the DM-nucleus cross section by [32]

$$\sigma_{0,N} = \left(\frac{\mu_N}{\mu_n}A\right)^2\sigma_{0,n} \tag{3.4}$$

where μ_N is the DM-nuclear reduced mass and μ_n is the DM-nucleon reduced mass. From this, we see that using a heavier target results in a significant enhancement in the cross section. Note that here we only

consider the case where the interaction is spin-independent (SI). For the spin-dependent (SD) case no such enhancement occurs, since a coherent interaction would factor in the total spin of the nucleus, which doesn't scale directly with A. In general we will restrict the discussion in this work to SD interactions.

The differential number density is in turn related to the phase-space distribution of the DM

$$dn = \frac{n_0}{k} f(\mathbf{v}, \mathbf{v_E}) d^3\mathbf{v} \tag{3.5}$$

where $f(\mathbf{v}, \mathbf{v_E})$ is the velocity distribution of the DM - with $\mathbf{v}$ the velocity of the DM particle with respect to the target, and $\mathbf{v_E}$ is the velocity of the Earth with respect to the DM halo, and k is a normalization constant.

Direct detection experiments adapt a Standard Halo Model (SHM) to describe $f(\mathbf{v}, \mathbf{v_E})$ for the purposes of evaluating dn. The SHM assumes that the DM velocities follow a a Maxwell-Boltzmann distribution in the rest frame of the halo with some velocity cutoff $v_{esc} \approx 544$ km/s [33], above which the DM is no longer gravitationally bound.

$$f(\mathbf{v}, \mathbf{v_E}; v_{esc}) = \exp\left(-\frac{(\mathbf{v} + \mathbf{v_E})^2}{v_0^2}\right)\Theta\big(v_{esc} - |\mathbf{v} + \mathbf{v_E}|\big) \tag{3.6}$$

Where $v_0 \approx 240$ km/s is the most probable DM velocity to observe, determined by the Sun's velocity with respect to the galactic center. With this definition, we can evaluate k.

$$k = \pi^{3/2} v_0^2 \big[v_0 \operatorname{erf}(v_{esc}/v_0) - \frac{2v_{esc}}{\sqrt{\pi}} \exp\left(-v_{esc}^2/v_0^2\right)\big] \approx (\pi v_0^2)^{3/2} \tag{3.7}$$

This simple model is quite useful. For one, recall that one can derive the Maxwell-Boltzmann velocity distribution by assuming that energy of the particles in the ensemble obey Boltzmann statistics. That is to say, if the DM particles in the halo really follow the Maxwell-Boltzmann velocity distribution (ignoring for now the escape velocity), then if we were to observe a random DM particle the probability P that it would have energy between E and $E + dE$ is given by

$$\frac{dP}{dE}(E) \approx \frac{1}{E_0} e^{-E/E_0} \tag{3.8}$$

Where E_0 is the average energy of the DM particles in the halo, as observed on Earth

$$E_0 = \frac{1}{2} m_\chi v_0^2 \tag{3.9}$$

Given this, all we need to estimate the differential rate in our detector is normalize it to the total interaction rate R_0, evaluate the average recoil energy $\langle E_R \rangle$ that we expect a DM particle with energy E to deposit in our detector, and reinstate an energy cutoff corresponding to v_{esc} such that we can't claim sensitivity to arbitrarily fast DM.

The SHM gives us something very nice to work with when determining the spectrum of DM in our detector, but we should also be a bit suspicious of its simplicity. For one, another consequence of assuming that the DM follows a Maxwell-Boltzmann velocity distribution is that it implies that the DM in the halo has established thermal equilibrium. Considering that our typical picture of DM strongly constrains its self-interaction cross section [34], this equilibrium cannot be established through DM-DM collisions like an ideal gas on a time-scale consistent with structure formation. Rather, equilibrium could be established through gravitational interactions producing a kind of collisionless pressure [35]. In general, N-body simulations indicate that the resulting velocity distribution of DM halos may be non-isotropic [36, 37]. Despite these shortcomings, it is still worthwhile to interpret our search results in terms of the SHM. At the very least it gives us a decent and tractable first-order approximation for the velocity distribution, and in general direct detection experiments cannot meaningfully compare results without adopting a standard model.

An interesting consequence of this is the dependence of the rate on $\mathbf{v_E}$. As the Earth orbits the Sun, $\mathbf{v_E}$ will modulate which in turn implies that we expect to generically observe an annual modulation in the rate. This variation in the rate has the potential to give experiments a great degree of sensitivity, since most typical sources of backgrounds will be constant throughout the year. It should be noted that experiments looking for this signal cannot automatically assume that there is no variation in the background, since seasonal variation in temperature and atmospheric pressure can effect concentrations of radon [38] and the production rate of cosmic rays [39]. One promising method for rigorously constraining the seasonal variation in the background would be to compare DM search results from both the northern and southern hemispheres, as proposed by the SABRE experiment [40].

3.2.1 Elastic Nuclear Recoils

If a DM particle with kinetic energy E scatters elastically off the target, it is a straightforward to show that the energy imparted to the detector is

$$E_R = \frac{rE}{2}(1 - \cos\theta) \tag{3.10}$$

where θ is the scattering angle and r is a factor of the DM and target mass

$$r = \frac{4m_\chi m_T}{(m_\chi + m_T)^2} \tag{3.11}$$

where m_T is the mass of the target atom. The factor r is quite important, because it characterizes the "kinetic matching" between the DM and the target. In order to maximize the fraction of its energy the DM can transfer to the target, we want $m_T \approx m_\chi$. Assuming that the scattering is isotropic in the center of mass frame, we have the average recoil energy, the recoil energy from a DM particle with energy E is uniformly distributed between $0 \leq E_R \leq rE$. This result, taken with 3.8 can be used to show that the expected differential rate with respect to recoil energy in our detector is

$$\frac{dR}{dE_R} \approx \frac{R_0}{rE_0} e^{-E_R/rE_0} \tag{3.12}$$

with the approximation becoming equality in the limit $v_E = 0$, $v_{esc} = \infty$. This limit gives us a good approximation, but we should take care to use realistic values when actually calculating the sensitivity of our experiment. The escape velocity v_{esc} is a particularly important parameter, since it sets a hard upper limit on the recoil energy we can observe. If one wants to perform an annual modulation search, it is clear that v_E is a critical parameter. For a more complete discussion, see [5, 32]. The total rate R_0 is determined by integrating 3.2 in this limit

$$R_0 = \frac{2}{\sqrt{\pi}} \frac{N_0}{A} \frac{\rho_\chi}{m_\chi} \sigma_{0,N} v_0 \tag{3.13}$$

where ρ_χ is the local DM mass-density. We adopt the recommendation of [33] in this work and take $\rho_\chi = 0.3$ GeV/c^2/cm^3.

As we search for DM at lower masses, it becomes clear from this discussion that our experimental reach for elastic recoils becomes kinematically suppressed. The scale of the recoil energy deposited in our detector

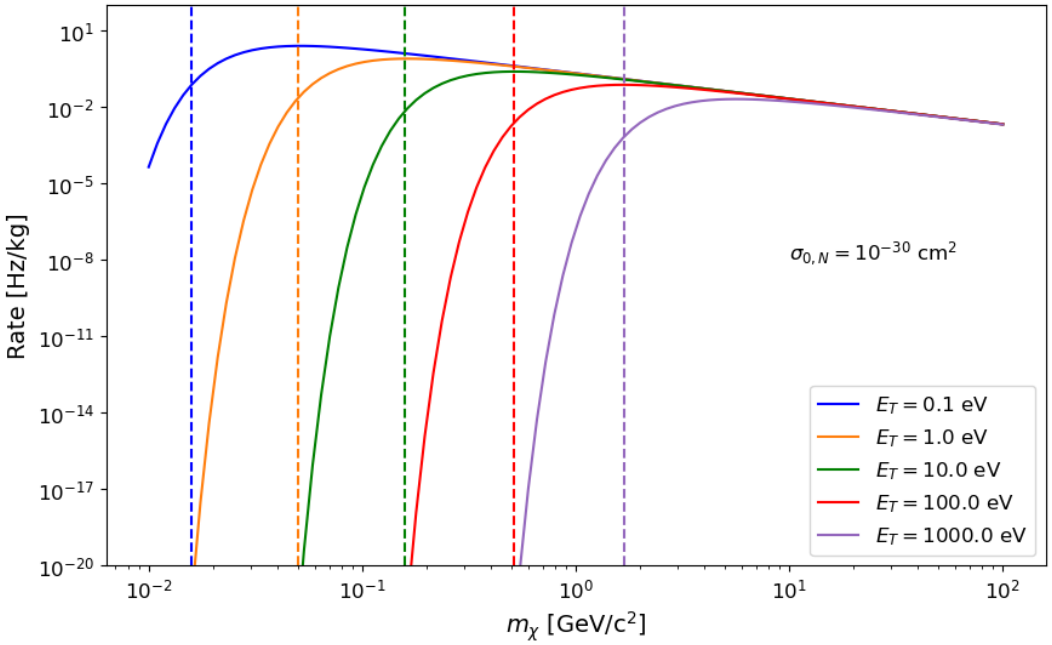


Figure 3.1: Expected rate per detector mass for a Si target as a function of DM mass m_χ. Solid curves represent the total integrated rate above a given threshold E_T in the limit $v_{esc} = \infty$. Dashed lines represent the cutoff on mass sensitivity imposed by a finite v_{esc}, below which the DM is not kinematically able to produce a recoil above threshold. Sensitivity to DM models where $m_\chi < 100$ MeV/c^2 requires $E_T < 10$ eV.

is defined by the ratio r

$$\langle E_R \rangle = rE_0 \approx \begin{cases} 2\frac{m_\chi^2}{m_T} v_0^2, & m_\chi \ll m_T \\ 4m_T v_0^2, & m_\chi \gg m_T \end{cases} \tag{3.14}$$

If we were sensitive to arbitrarily small E_R and are in the limit $v_{esc} = \infty$ this wouldn't be relevant and we'd always expect to observe rate R_0 in our detector. Of course this isn't realistic. Our detector always has some intrinsic non-zero trigger threshold E_T, below which we can't meaningfully distinguish between signal events and noise. Imposing this threshold and remaining in the $v_{esc} = \infty$ limit gives the expected event rate

$$R(E_T) = \int_{E_T}^{\infty} \frac{dR}{dE_R} dE_R = R_0 e^{-E_T/rE_0} \tag{3.15}$$

Even though there's a chance for an DM particle with arbitrarily large energy to interact with the detector, the finite threshold means our ability to observe this is suppressed when $E_T > rE_0 \propto m_\chi^2$. Imposing a finite v_{esc} punishes us even more, since we expect to see nothing in our detector when the largest allowed recoil energy falls below the detector threshold. This sets a lower bound on the DM we can search for with a given detector

$$r m_{\chi,e} \geq \frac{2E_T}{v_{esc}^2} \tag{3.16}$$

where the subscript e indicates that this is the lower mass bound for elastic recoils. In the limit $m_\chi \ll m_T$, this becomes

$$m_{\chi,e} \geq \sqrt{\frac{E_T m_T}{2v_{esc}^2}} \tag{3.17}$$

This behavior is illustrated in 3.1, which demonstrates that we require $E_T \sim \mathcal{O}(1 \text{ eV})$ in order to achieve sensitivity to DM with $m_\chi \sim \mathcal{O}(100 \text{ MeV/c}^2)$.

3.2.2 Inelastic Channels

Given the kinematic suppression which naturally arises from looking for LDM via elastic recoils on a massive target, there are two natural paths we can take to continue pursing DM detection at smaller masses:

1. We can choose a lighter target such that we improve the kinematic matching with LDM.
2. We can relax expand our searches to consider inelastic recoils, where the recoil energy is not strictly determined by the ratio of the DM and target mass.

In terms of choosing a less massive target, we are somewhat constrained. There are only a handful of elements lighter than Si, and fewer still which make suitable detector materials. There is active research and development into the use of diamond [41], sapphire [42] and superfluid He [43] as detector volumes. These would all in principle provide superior kinetic matching to LDM than Si. Clearly superfluid He would optimize the sensitivity at the lowest masses, though we note that the realizations of this technology currently under development rely on coupling the superfluid active volume to a more conventional solid-state detector to actually read out the excitations. The remainder of this chapter will focus on the case of inelastic recoils. When we consider these channels in general, the DM may interact with primarily with either the target electrons or nuclei, depending on the particular model. To distinguish between these classes of models, we refer to either Electron Recoil Dark Matter (ERDM) or Nuclear Recoil Dark Matter (NRDM).

When looking at a inelastic channels, it is in general possible for the incident DM to deposit all of its kinetic energy. In principle this means we can search for DM with a lower bound on the masses defined by

$$m_{\chi,ie} \geq \frac{2E_T}{v_{esc}^2} \tag{3.18}$$

where the subscript ie indicates that this is the lower mass bound for inelastic recoils Comparing this to 3.16, we see that all we have done is remove the factor r. Though this lifts the kinematic restriction on searching for lower masses, there is often a penalty on the rate associated with inelastic channels when we evaluate the matrix element between initial and final states.

The computation of the inelastic scattering rate is more involved than the elastic case, as the initial and final states are more complicated. We outline the general calculation which is discussed in detail in Ref. [44]. The transition rate induced by the DM scattering can always be calculated from first principles using Fermi's Golden Rule. First, let $\mathcal{M}(q)$ represent the momentum-dependent matrix element relating the outgoing/incoming DM particle state. We factorize the momentum dependence by evaluating at a reference q_0. $\mathcal{M}(q) = \mathcal{M}(q_0)\mathcal{F}_{med}(q)$, where $\mathcal{F}_{med}(q)$ is a form factor for the DM interaction which depends on the ratio of the DM and mediator mass m_χ/m_{med}. In particular, we have

$$\mathcal{F}_{med}(q) = \begin{cases} 1, & m_\chi \gg m_{med} \\ (\frac{q_0}{q})^2, & m_\chi \ll m_{med} \end{cases} \tag{3.19}$$

Notice that this implies that LDM with a light mediator is easier to detect, since it is associated with lower

momentum transfers. Next we define the reference cross section $\sigma_{ref} \equiv (\mu^2/\pi)\overline{|\mathcal{M}(q_0)|^2}$. With this, the interaction rate can be written as

$$\Gamma(\mathbf{v}) = \frac{\pi\sigma_{ref}}{\mu^2}\int \frac{d^3q}{(2\pi)^3}\mathcal{F}^2_{med}(q)S(\mathbf{q}, E_R) \tag{3.20}$$

Where $\mathbf{q}$ represents a particular momentum transfer to the target, and E_R is the corresponding energy deposition

$$E_R = \mathbf{q}\cdot\mathbf{v} - \frac{q^2}{2m_\chi} \tag{3.21}$$

We have also introduced the dynamic structure factor $S(\mathbf{q}, E_R)$, which is an extremely useful quantity as it encapsulates *all* the target-specific internal structure of the detector

$$S(\mathbf{q}, E_R) \equiv \frac{2\pi}{V}\sum_f \big|\langle f|\mathcal{F}_T(\mathbf{q})|i\rangle\big|^2\delta(E_f - E_i - E_R) \tag{3.22}$$

Here $|i\rangle, |f\rangle$ represent the initial/final states of target, and E_i, E_f the corresponding energy. We will discuss the particular evaluation of the structure factor in the following subsections.

Once we have this transition rate, we convert it to a detection rate by accounting for the density of targets in our detector, ρ_T and the phase-space distribution of the DM

$$\frac{dR}{d\omega} = \frac{n_0}{\rho_T}\int d^3v f(\mathbf{v}, \mathbf{v_E})\frac{d\Gamma}{d\omega} \tag{3.23}$$

One should briefly verify that this formalism reproduces the previous, somewhat heuristic, derivation of the elastic nuclear scattering rate. In particular we approximate the target nuclei as not interacting with their neighbors, the structure factor can be written as [44]

$$S(q,\omega) = 2A^2\pi\frac{\rho_T}{m_N}F_N^2(q)\delta\big(\omega - \frac{q^2}{2m_N}\big) \tag{3.24}$$

Plugging this into Eq. 3.23 and assuming a heavy mediator then gives us the familiar result. We're particularly interested in the case where the target is a crystal, which brings into question the assumption we made above that the target nuclei act independently. Clearly the existence of the crystal lattice is evidence that are interactions between lattice sites. However, as long as the recoil energy is sufficiently large compared to individual lattice excitations (i.e. phonons), the independent approximation turns out to be a fairly good

one [45]. Typically mean phonon energies are of the order $\omega_p \sim \mathcal{O}(100 \text{ meV})$.

In all the inelastic channels we consider, there is some electronic excitation. In this case, the dynamic structure function can be related to the Energy Loss Function (ELF) which is subsequently related to the dielectric function of the material $\epsilon(q, \omega)$

$$S(q, \omega) \approx \frac{q^2}{2\pi\alpha} \text{Im}\big[-1/\epsilon(q, \omega)\big] \tag{3.25}$$

Where the approximation holds in the low temperature limit. In principle one can calculate this empirically using measurements of $\epsilon(q, \omega)$, but current data does not probe momentum transfers at the scale relevant for DM scattering [45]. One must therefore resort to directly calculating $\epsilon(q, \omega)$, depending on the electronic structure of the material. For the case of an atomic target, where the initial electron state is bound and discrete and the final state is an approximately free electron, this calculation is straightforward given a relativistic Hartree-Fock description of the initial states [10]. For the case of a crystal target, the initial electron state may be tightly bound for the inner electrons or a continuum state for valence, and then the final state exists in the continuum conduction band. This calculation is more complicated, and usually better suited for Density Functional Theory (DFT) methods [46].

DM-Electron Scattering

The benefit of looking for LDM via scattering off electrons is two-fold. Not only is it an inelastic channel, which means that all the kinetic energy of the DM is in principle available to excite the electron, but the light mass of the electron also means that the energy transfer is more efficient on average. We do note that this kinematic matching advantage is not as trivial as we expressed in 3.11, since we can't definitively specify the initial momentum of the electron in the lab-frame. Typically we expect $v_e \sim \alpha c$, which implies a 10 MeV traveling at v_0 of would deposit ~ 10 eV when scattering off an electron [47].

The numerical package QEdark has been developed by the dark matter community to calculate the crystal form factors in germanium and silicon, allowing cryogenic experiments to effectively determine scattering rates of light dark matter off of electrons in their detectors [48]. It relies on the established electronic structure software package Quantum Espresso in order to calculate the electronic form factor using DFT techniques, and then convolve it with the DM velocity distribution to find the scattering rate. An example spectrum for a 100 MeV DM particle scattering on a Ge target via a heavy mediator is shown in Fig. 3.2. This tool is

well established, and has been used by several experiments to exclude DM-electron scattering for masses as low as $m_\chi \sim 1$ MeV [49–52]. It has been demonstrated however that the formalism implemented in QEdark does not properly account for screening between valence electrons [49], which makes the derived rates not conservative. Though QEdark remains a useful benchmark to compare results from different experiments, alternative packages such as DarkELF [46], EXCEED-DM [53] or QCDark [54] should be considered for calculating accurate sensitivities to DM-elctron scattering.

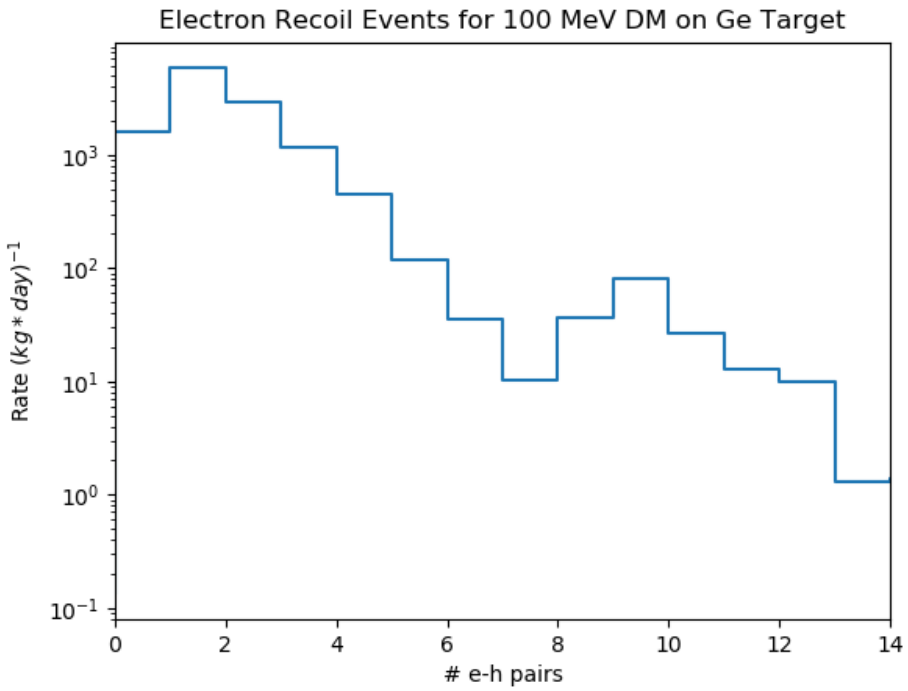


Figure 3.2: Spectrum for a 100 MeV dark matter particle scattering off electrons in Ge, assuming an average electron scattering cross section of 10^{-37} cm^2. Calculated using QEdark. Note the bump near 30 eV, which is due to the 3d electron shell in Ge.

Nuclear Recoils

The Migdal effect describes the inelastic process where an atomic target is ionized following a nuclear recoil. A cartoon of this process for a free atom is shown in 3.3. To understand why we expect prompt ionization to follow a NR with some non-zero probability, we consider the case of an isolated atom. Note that this problem is given as an exercise in [55], and we only motivate the solution here. If the atom begins at rest with the electron in the ground state and some momentum q is imparted to the nucleus, the electron must

"catch up" to the now moving nucleus. Then the excitation probability is given by the overlap of the ground state wave function in the frames before and after the momentum transfer. We find

$$P_{excite}(q) = 1 - \left(1 + \frac{1}{4}(qa)^2\right)^{-4} \tag{3.26}$$

where a is the Bohr radius. Hence the fact that a nuclear recoil can induce prompt ionization follows from elementary quantum mechanics. This simple result informs us of two intuitive results regarding the Migdal effect

1. More energy transferred to the nucleus is more likely to excite the atom.

2. It's easier to excite an electron further away from the nucleus, as it's less tightly bound.

For a real material the calculation is more complicated, as the valence electrons may interact with each other and are not simply described by atomic orbitals. However, this approximation should still hold if the analysis is restricted to tightly-bound, non-valence electrons, in which case the transition rate is calculated in [56]. This approach has been adopted by a number of experiments to calculate exclusion limits on NRDM exploiting the Migdal effect, with resulting sensitives to DM masses as low as $m_\chi \sim 10 - 100$ MeV. This has been done both with liquid noble [57–59] and crystal [60–62] targets.

Though the calculation must be altered in order to analyze the excitation of valence electrons via the Migdal effect, there is huge potential benefit to doing so. This follows from our conclusion that less tightly bound electrons should be easier to excite. Hence we expect an enhancement in the Migdal effect when looking at the valence band of a semiconductor compared to an atomic target. To see how the calculation can be generalized, note that the initial and final states of the system are almost identical to those in the case of DM-electron scattering [63]. This means that the Migdal rate should also be encoded in the ELF. This becomes more clear in the limit where the momentum transfer to the electron is small compared to the nucleus, in which case we have [64]

$$\frac{d\sigma_{ion}}{dE_R d\omega} \approx \frac{d\sigma}{dE_R}\frac{dP}{d\omega}(E_R) \tag{3.27}$$

where E_R is the recoil energy, ω the energy deposited into the electron system, $d\sigma/dE_R$ is the differential cross section for elastic nuclear scattering and $dP/d\omega$ is the differential ionization probability. This probability can be derived from the dynamic structure factor. Detailed calculations of the Migdal rate applied to the semiconductor valence band are given in [45, 64].

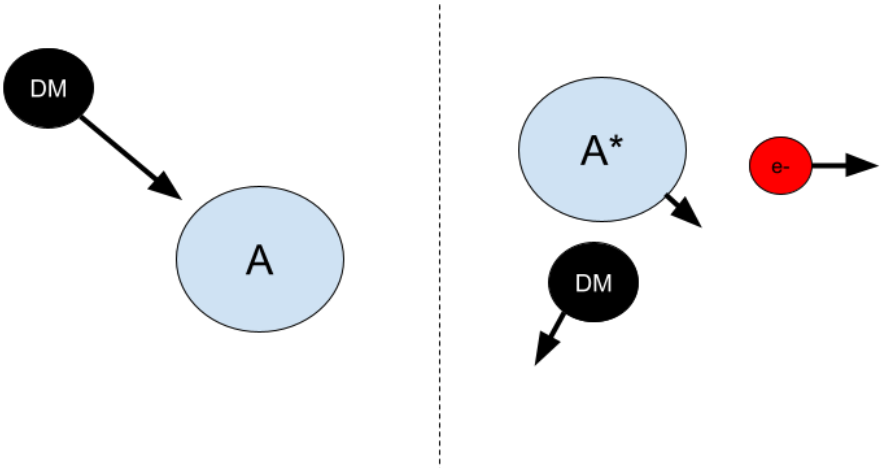


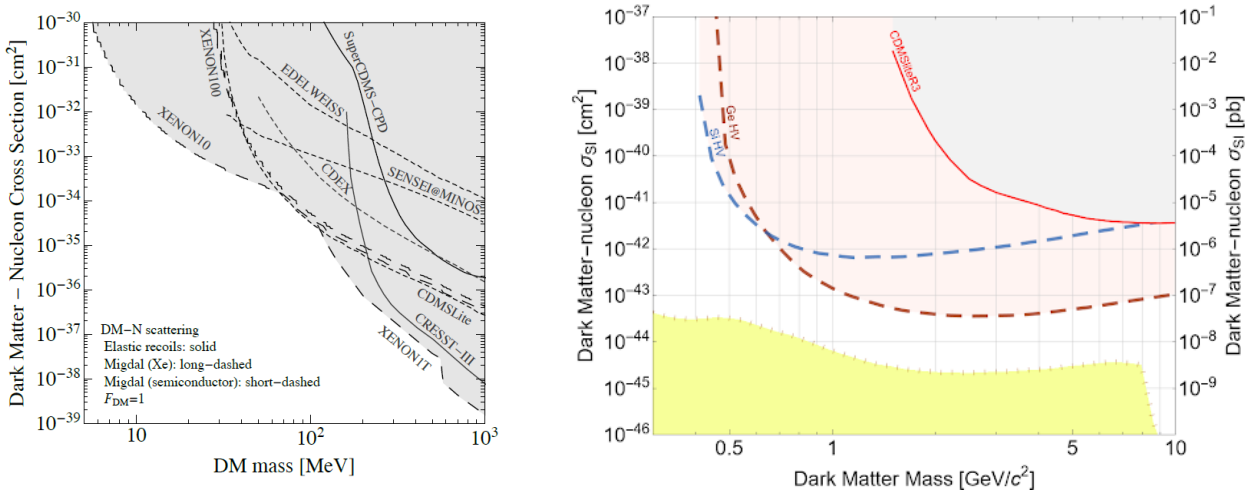
Figure 3.3: Sketch of the Migdal effect. An inelastic nuclear recoil with an atom excites the target and releases an electron.

At this point in time, some caution should be applied when applying the Migdal effect to any analysis. Though the Migdal effect is certainly theoretically well motivated, it has yet to be directly calibrated in a laboratory. Notably, a dedicated experiment to directly observe the effect in liquid xenon has yielded a null result [65]. This is not the nail in the coffin for the Migdal effect, as the result could be consistent with underestimating the in-medium recombination which would not occur at the energy scales relevant for a DM search. But the result demonstrates that our picture of the Migdal effect is incomplete. More measurements are required to calibrate the effect, both in liquid noble and semiconductor targets [66].

3.3 Where Are We and Where Are we Going?

The current landscape of the search for LDM is shown in Fig. 3.4a. We see that liquid noble detectors utilizing the Migdal effect currently claim to exclude $\sigma_{0,n} \gtrsim 10^{-32}$ cm^2 at $m_\chi \sim 10$ MeV. Though this is an impressive result, it is more optimistic than concrete until the effect can be directly calibrated and the non-observations in [65] are accounted for.

The present limits set by solid state detectors are confined to higher masses, but we don't necessarily expect this to remain the case. For one, we know that attaining sensitivity to LDM is closely tied to our ability to resolve increasingly small energy depositions. The analysis threshold of a liquid noble detector looking for electronic excitations is ultimately determined by the binding energy of the valence electrons (~ 10 eV). The threshold of a semiconductor detector will be limited by its band gap (~ 1 eV) if looking for



(a) (b)

Figure 3.4: (Left) Current exclusion limits on NRDM at low masses. The majority of the curves claiming sensitivity in this region take advantage of the Migdal effect. Figure from [67] (Right) Projected reach of the SCDMS HV detectors at SNOLAB. Note the different scales between the plots. The low background will let the detectors gain sensitivity to substantially lower cross sections, but the projected resolution limits the mass reach to $m_\chi \gtrsim 400$ MeV. Figure generated with the SCDMS Limit Plotter [68].

electronic excitaitons, and the mean phonon energy (~ 100 meV) if only considering nuclear excitations.

We compare the current state of LDM searches to the projected reach of the SuperCDMS-SNOLAB HV program in Fig. 3.4b, which predicts a silicon detector will be able to exclude $\sigma_{0,n} \gtrsim 10^{-41}$ cm^2 at $m_\chi \sim 500$ MeV. The projected sensitivity to smaller cross sections is a natural consequence of the low background design of the experiment. One should also note that these projections assume no background modeling, the inclusion of which will push the curve to even lower cross sections using a profile likelihood method (barring that DM is actually detected in this region). While this is a satisfying chunk of parameter space, we should consider what is required to make a similar push in the mass region below 400 MeV. In Chapter 8, we discuss results which attempt to accomplish this by leveraging existing SuperCDMS technology.

Part II

The SuperCDMS Experiment

Chapter 4

The SuperCDMS Experiment

4.1 Introduction

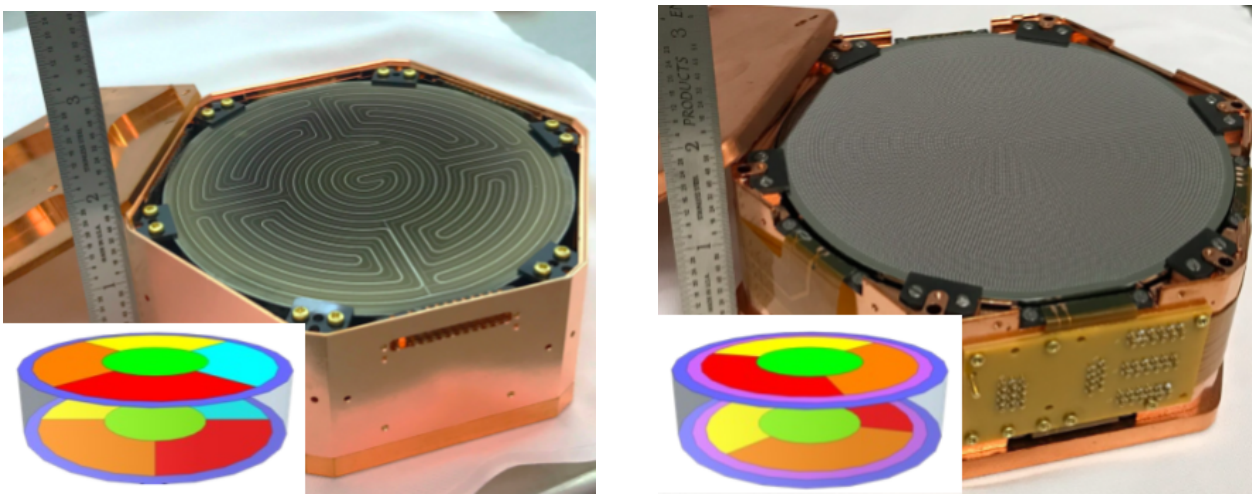
We have a number of requirements to perform an effective direct search for dark matter

1. Dark matter interactions with the detector are rare. In order to claim a detection, an experiment must be carried out in a low background environment so that the expected signal of a few events per kilogram-year can be distinguished from a statistical fluctuation in the background.

2. Any particular dark matter model we search for will interact with our detectors primarily either via nuclear or electron recoils. Hence our detectors need to be sensitive to energy depositions in at least one of these channels. If the detector is sensitive to both, the ability to distinguish between the two presents a powerful tool to reject backgrounds.

3. There is a wide range of viable masses for dark matter candidates. If our goal is to search for dark matter at the scale of $\sim$ 1-10 GeV/c^2, our detectors need baseline energy resolutions $\lesssim$ 100 eV.

In this section we discuss the details of the Super Cryogenic Dark Matter Search (SuperCDMS or SCDMS), and how it achieves each of these requirements. We will start with a description of the detectors employed by SuperCDMS, and the technology which allows them to achieve the required sensitivity in both the electron and nuclear recoil channels. Then we will discuss the experimental infrastructure, including the SNOLAB facility and shielding leveraged to reduce backgrounds within tolerance.



(a) (b)

Figure 4.1: Photos of SuperCDMS SNOLAB detectors and sensor layout. Left shows an iZip, and right an HV. We see that the phonon sensors are much sparser on the iZip. The respective phonon channel layouts are shown in the bottom left corner of each photo. Both detectors have the same number of phonon channels, but different layouts.

4.2 SuperCDMS SNOLAB Detectors

4.2.1 Detector Description

The detectors to be deployed by SuperCDMS at SNOLAB come in two designs and two materials, resulting in a matrix of four detector "flavors". The utilized materials are germanium and silicon, and the designs consist of the Interleaved Z-sensitive Ionization and Phonon detector (iZip), and the High Voltage detector (HV). Both designs share common dimensions with a 100 mm radius and 33.3 mm height, but operate with different readout channels and external voltage biases.

The inclusion of multiple target materials provides sensitivity to more signal channels, e.g. if the DM interaction is spin-dependent it will have different couplings to germanium and silicon. The two detector designs then aim to probe different regions of the DM parameter space. The iZip is capable of discriminating between electronic and and nuclear interactions in the bulk, making it well suited for characterizing backgrounds and searching for DM in the $m_\chi \approx 10$ GeV mass range. The HV detector trades this discrimination power for increased signal gain and lower analysis thresholds, and mainly probes the $m_\chi \lesssim 1$ GeV regime.

In a crystal semiconductor, there are two relevant energy systems which an incident particle can excite; the phonon system and the electron system. Whereas phonons are effectively nuclei excited in the quasi-

harmonic potential at the crystal lattice sites, the electron system is characterized by the effective continuum of states occupied by the valence electrons and the parallel continuum they occupy once excited, respectively termed the "valence" and "conduction" bands. When electron is excited from the valence to conduction band, it leaves an effective positive charge at the ionized lattice site and the entire system can be described as an electron-hole pair.

All SuperCDMS SNOLAB detectors share common dimensions with a 100 mm radius and 33.3 mm height. The iZip device is instrumented with two charge and six phonon channels on each side. The HV eschews charge channels as it doesn't attempt to discriminate between the charge and phonon signals, but increases the sensor coverage in its six phonon channels per side. The charge channels operate by coupling a sensitive capacitor to the detector surface. The external voltage bias drifts charge carriers across the detector bulk, and an image charge is induced on the coupling capacitor. The resulting signal is then amplified by a HEMT. Phonon channels are based on absorbing athermal phonons into a sensitive thermistor, and observing the change in resistance.

4.2.2 The Crystal System

Silicon and germanium targets are relatively light compared to their liquid noble counterparts. This difference in mass scale naturally defines a partition in the dark matter parameter space; semiconductor targets are well suited to search for lighter candidates ($\lesssim$ a few GeV), and liquid nobles for heavier ones ($\lesssim$ a few TeV). This division becomes even clearer when we consider what determines the analysis thresholds of the respective materials. For electronic interactions, the resolution of a liquid noble detector is ultimately determined by the target ionization energy ($\sim$ 10 eV), while the resolution of electronic excitations in a semiconductor is fundamentally limited by the bandgap ($\sim$ 1 eV); hence semiconductors are inherently more sensitive to electronic excitations. For a liquid noble, the detection of a nuclear recoil is mediated through the electron system via scintillation, hence the sensitivity of each channel is on a comparable scale.

The presence of the two systems mirrors the two kinds of interactions we expect to detect from an incident particle. An incident particle can either interact with a target electron, producing an Electron Recoil (ER), or a target nucleus and generate a Nuclear Recoil (NR). In both cases the deposited energy is promptly partitioned between the electron and phonon system, but the evolution of each system depends on the external voltage.

Each system exhibits a distinct energy scale, with fundamental phonon excitations at $\mathcal{O}(10)$ meV and the

scale of electronic excitations is determined by the semiconductor bandgap $\mathcal{O}(1)$ eV. Despite this separation of scales, the two systems are intimately related, with both arising from the structure of the crystal lattice[1]. NRs generate electron-hole pairs, whether at an individual lattice site via the Migdal effect or through interactions between neighboring lattice sites as described by Lindhard theory. Similarly, phonons always accommodate the excitation of electron-hole pairs, as evidenced by the fact that the energy required to produce the excitation is greater than the bandgap. We can exploit the coupling of electron-hole pairs to phonons by applying an external voltage across the crystal. In the presence of an E-field, the pair partners drift to their respective terminals and shed phonons proportional to the magnitude of the voltage via the Neganov-Trofimov-Luke (NTL) effect. Given that NRs and ERs excite both the electron and phonon system, our ability to distinguish between interaction types based on the detector response relies on characterizing the particular partition each interaction makes between the two channels.

4.2.3 Ionization Signal

The energy scale of electronic excitations in a semiconductor is ultimately determined by the bandgap. In operating our detectors we don't directly observe this energy since phonon production typically accompanies the excitation [69]. Rather we typically refer to the average energy required excite an electron-hole pair, ϵ_{eh}. This value is constant at energies large compared to the bandgap. At energies near the bandgap, corrections to ϵ_{eh} can be treated empircally as in [70]. The bandgap and ϵ_{eh} for Ge and Si are summarized in Table 4.1.

	Bandgap (eV) [71]		
Material	Indirect	Direct	ϵ_{eh} (eV) [72]
Ge	0.74	0.89	3.0
Si	1.17	–	3.8

Table 4.1: Characteristic electronic energy scales for Ge and Si at 0 K. At higher temperatures, all quantities will decrease.

If there is no external electric field applied to the crystal, any liberated electron will be drawn to the closest positive charge - likely the just ionized lattice site, but possibly another nearby impurity. Once the liberated charge recombines, all the excitation energy has been converted to phonons and there is nothing to indicate to the experimenter whether any ionization occurred in the first place. Hence if we want to observed the ionization directly we need to apply a voltage to the crystal, such that the liberated electron-hole pairs

[1] Both Si and Ge form a fcc diamond structure

drift across the crystal rather than promptly recombine.

As we will soon discuss, there is distinct phonon signal which accompanies the pair drift across the crystal. In principle one can infer the amount of ionization produced from measuring the phonons alone with sufficient resolution. However if one is interested in analyzing the ionization and phonon channels simultaneously, it is worthwhile to make independent measurements.

4.2.4 Phonon Signal

The phonon system is described by distinct phonon modes (optical and acoustic) with their own band structure and dispersion relation. Presently the technology we use to detect phonons has not yet reached the $\mathcal{O}(10\text{ meV})$ resolution needed to resolve individual quasiparticles. There are several paths of development to improve phonon readout systems, with the common goal of resolving individual phonons [73]. The realization of this goal would allow semiconductor based experiments to exploit the discrepancy between the fundamental nuclear and electronic energy scales in future dark matter searches.

At a high level, it is more convenient to describe the average evolution of the system as a whole rather than that of individual phonon modes. Any attempt to describe the detailed phonon dynamics typically relies on simulation techniques. Our zeroth order picture of the phonon evolution is that, promptly following the particle interaction, we have an isotropic soundwave of ballistic, acoustic phonons propagating from the interaction point.

Detecting a the small amount of heat expected from a DM interaction requires a very sensitive thermometer. In the case of SuperCDMS this is realized using the Transition Edge Sensor (TES). A TES is made of a superconducting material (in our case tungsten) instrumented on the face of the crystal. A voltage bias is applied across the TES, such that it is heated to just below its critical temperature $T_c \sim 50$ mK. Note that this requires us to operate the detector in a cryogenic environment sufficiently colder than T_c. When energy from the phonons reaches the TES, it is driven out of the superconducting transition into its normal phase, which is characterized by a large increase in the resistance of the device. Then we actually have to measure this change in current, which we do by placing an inductor in series with the TES and coupling it to a SQUID.

We can now see two aspects of the TES which will determine the resolution of the phonon signal. Since we're ultimately measuring the phonon energy via change in the temperature of the TES, we can increase our sensitivity either by reducing the heat capacity of the TES or by increasing the fraction of initial phonon

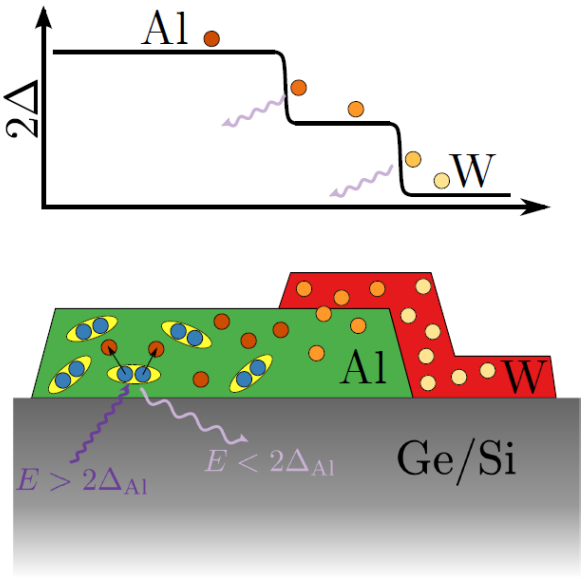


Figure 4.2: Cartoon of the QET. Phonons with sufficient energy break Cooper pairs in the aluminum, which then propagate to the tungsten TES. Figure from Ref. [74].

energy which reaches the TES. We lose sensitivity to the phonons when they thermalize by down-scattering off of impurities in the crystal bulk or surface. Hence we can increase the amount of phonon energy we collect by increasing the active area of the TES such that a given phonon has a greater chance to hit the TES before thermalizing. However increasing the size of the TES necessarily increases its heat capacity, so there is not an intrinsic gain in sensitivity from this scheme.

One can circumvent this trade-off by coupling the TES to a dedicated collection area, such that the phonon collection is improved without increasing the heat capacity. SuperCDMS uses aluminum fins as a collection target. Aluminum goes superconducting at ~ 4 K, much higher than tungsten. Hence we expect the large aluminum volume to remain superconducting when absorbing any reasonable amount of phonon energy. While the aluminum is superconducting, incident phonons with sufficient energy break Cooper pairs. These pairs propagate across the aluminum bulk, and some fraction reaches the TES, facilitated by an intermediate bandgap in a region where the tungsten and aluminum overlap. This device is termed the Quasiparticle-assisted Electrothermal-feedback TES (QET).

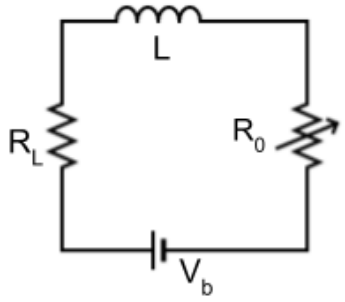


Figure 4.3: Thevenin diagram of the TES circuit. The resistance of R_0 will increase when some heat is deposited into the tungsten. The inductor L couples to the actual readout circuit.

TES Dynamics

Here we present a brief overview of how the energy deposited in the TES relates to the change in current, which is the basis of our phonon energy measurement. For a more complete discussion, one should always start by referring to [75], the so-called "TES Bible".

The crucial underlying principle of operating the TES is that its temperature and the current running through it are coupled by electrothermal feedback (ETF). In particular, the state variables (T, I) are governed by the system of ODEs

$$C\frac{dT}{dt} = I^2R(T,I) - \kappa[T^n - T_b^n] \tag{4.1}$$

$$L\frac{dI}{dt} = V_b - I[R_L + R(T,I)] \tag{4.2}$$

The first equation describes the heat power balancing of the device. The current through the TES produces Joule heating

$$P_J = I^2R(T,I) \tag{4.3}$$

Then the remaining temperature term represents the cooling of the TES due to its weak thermal coupling to the bath at temperature T_b. The second equation then describes the current response to the applied voltage bias V_b. The two equations are coupled through the resistance term, which in general depends on both the current and temperature, though the temperature sensitivity is much greater. This temperature dependence is characterized by a superconducting regime $R(T,I) \approx 0$ at low temperatures, a normal resistance regime $R(T,I) \approx R_N$ at high temperatures, and a sharp transition at some critical temperature T_c. Notably, in the case where the voltage across the TES is kept approximately constant, the temperature will return to its

equilibrium value following some deviation much faster than in the case where there is no coupling between the current and temperature.

We typically arrange our circuit with a small inductance, such that intrinsic electrical response time is small. In which case, we can approximate

$$V_b \approx I\big[R_L + R(T,I)\big] \tag{4.4}$$

$$\rightarrow R(T,I) \approx \frac{V_b}{I} - R_L \tag{4.5}$$

Then we can write the Joule power as

$$P_J \approx V_b I - R_L I^2 \tag{4.6}$$

Consider the steady state operation of the device, where some baseline current I_0 passes through the TES such that the Joule heating balances the bath power. Then suppose we have some energy deposited in the TES which causes a temperature deviation from the operating point and a subsequent current deviation $I = I_0 + \delta I$. Then we can track the change in the Joule heating.

$$\delta P_J = P_J - P_{J,0} \approx V_b(I_0 + \delta I - I_0) - R_L\big((I_0 + \delta I)^2 - I_0^2\big) \tag{4.7}$$

$$= (V_b - 2I_0R_L)\delta I - R_L(\delta I)^2 \tag{4.8}$$

This relationship defines the fundamental scale at which we expect the current to deviate in response to a given energy deposition.

4.2.5 The NTL Effect

In the previous section, we restricted the discussion to phonons which are produced from the initial recoil off the incident particle. We can also produce secondary phonons which can be used to amplify the signal. In the absence of an external electric field, any electron-hole pairs liberated by the particle interaction will immediately recombine with the closest ion and release their energy as phonons. These recombination phonons look much like the recoil phonons, and there is no way to determine how much of the recoil energy was expended as phonons versus ionization.

If we apply a voltage across the detector, the picture changes significantly. The induced electric field

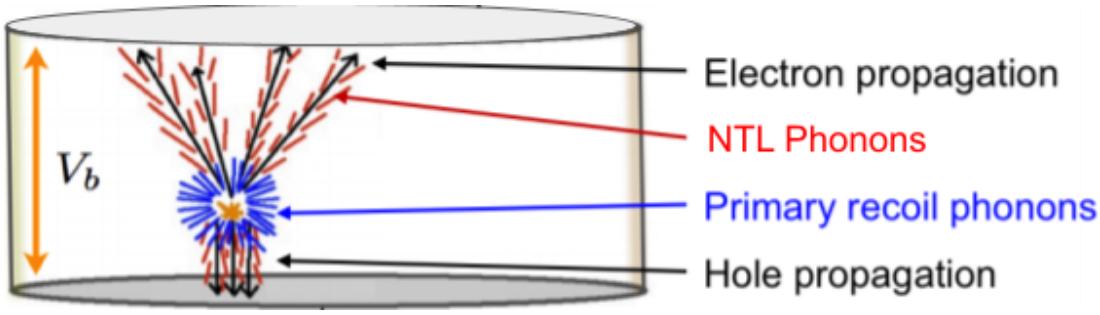


Figure 4.4: Cartoon of electron, hole and phonon propagation for a biased detector. Phonons are generated both by the prompt recoil, as well as by electorn-hole pairs being drifted across the detector according to the NTL effect. The amount of NTL phonons produced is proportional to the applied voltage V_b and the amount of liberated charge.

will do work on any free electron-hole pairs and drift the carriers across the crystal to their respective terminals. The rather than accelerate the entire time, the carriers quickly achieve a terminal velocity. The field continues to perform work on the carriers, and this excess energy is realized as secondary NTL phonons. These phonons behave very differently than those produced from recoil or recombination, in that they are emitted preferentially in the direction of the drifting carrier as opposed to quasi-isotropically. In order to quantify the signal amplification, we only need to recall the work-energy theorem

$$E_{NTL} = e \cdot n_{eh} V \tag{4.9}$$

where e is the elementary charge, n_{eh} is the number of pairs liberated by the particle interaction and V is the applied voltage. From this we see that the amplification is proportional to the applied voltage, which gives us a knob to increase our gain. We also notice in 4.9 that the only non-trivial component is determining how many pairs are created for a given interaction.

4.2.6 The Total Phonon Signal and Interaction Discrimination

Now let us simultaneously consider the phonon signal for a particle recoil in the presence of an external field. The total number of phonons is just the sum of those produced in the initial recoil and those generated via the NTL effect

$$E_{tot} = E_r + E_{NTL} = E_r + e \cdot n_{eh} V \tag{4.10}$$

Next we need to specify n_{eh}. Intuitively the number of liberated pairs is proportional to the recoil energy E_r, but the proportionality depends strongly on whether that recoil is off of an electron or a nucleus in the target. To quantify this dependence, we introduce the ionization yield, Y

$$Y(E_r) \equiv \frac{\epsilon n_{eh}}{E_r} \tag{4.11}$$

where ϵ is the average pair creation energy summarized in Table 4.1. If the recoil is off an electron, then $Y \sim 1$, regardless of the target or energy. This is intuitive, as we expect the initial recoil to impart all of its energy to the electron, which in turn readily gives any energy in excess of the pair creation threshold to the creation of additional electron-hole pairs. For a nuclear recoil, the picture is more complicated and is in general a function of the recoil energy. Any ionization must arise from the recoiling nucleus colliding with its neighbors and slowing down, which is an inefficient process. The benchmark model describing this process comes from Lindhard, which has been experimentally shown to be accurate in both Si and Ge at recoil energies $\gtrsim 1$ keV. At lower energies it is well established that the Lindhard model breaks down, and it is of great interest to the DM direct detection community to perform independent measurements of the ionization yield in this regime.

For a detector which measures charge and phonon energy simultaneously, such as the iZip, it is straight forward to measure the ionization yield. An example of this using ^{252}Cf calibration data from a Ge iZip used by SCDMS Soudan is shown in Fig. 4.5. The calibration source generates both electron and nuclear recoils through the emission of gammas and neutrons respectively, and the resulting bands from each interaction are clearly visible.

This event-by-event discrimination is useful for rejecting backgrounds from a dark matter search. A given dark matter model will only interact through a single channel, so any events from the other channel can be cut from the spectrum. In the past our dark matter searches were restricted to WIMPs, which interact via nuclear recoils and hence the electron recoils were the primary background of interest. If we instead want to search for a model of DM which interacts via an electron recoil, this relation is reversed and we would instead reject nuclear recoils as the background.

While the utility of the background discrimination capabilities presented by the iZip are obvious for a low-background experiment, there is an inherent limitation to the measurement of the ionization yield near the detector threshold [74]. Searching for LDM requires us to analyze our data as close to the ever-shrinking

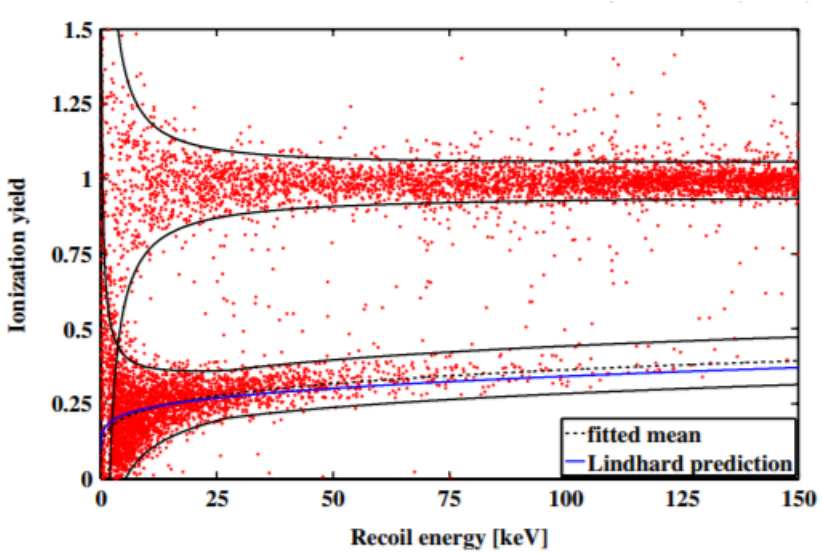


Figure 4.5: Ionization yield from ^{252}Cf calibration in a Germanium iZip. Upper band centered at $Y \approx 1$ is electron recoil events, and the lower band around $Y \approx 0.3$ are nuclear recoils. The mean of the NR band roughly follows the Lindhard theory down to ~ 1 keV, where the bands overlap and the yeild cannot reliably be estimated. Image from [76]

threshold as possible, hence the iZip isn't the optimal choice to search for LDM.

Consider the partition in energy between recoil and NTL phonons following a recoil

$$\frac{E_{NTL}}{E_r} = \frac{e \cdot n_{eh} V}{E_r} = \frac{e \cdot Y(E_r) V}{\epsilon} \tag{4.12}$$

For concreteness, consider the case of a germanium iZip operated at a bias of 10 V. For the yield, let's assume that we're looking at energies $E_r \lesssim 1$ keV, and conservatively take $Y(E_r) \sim 0.1$ in this regime. Then we have

$$\frac{E_{NTL}}{E_r} \sim 0.3, \text{ iZip Mode} \tag{4.13}$$

There are fewer NTL than recoil phonons, but they are on the same order of magnitude. This tells us that the applied voltage is really only serving the purpose of drifting the charge carriers across the crystal to facilitate the charge measurement, we're mostly measuring recoil phonons in the phonon channel. Since the discrimination is diminished at this energy, there is not a lot of utility in running in this mode. We do have a knob, in that we can increase the voltage bias. Say we increase the bias by a factor of 10, and call

this the "High Voltage" (HV) mode of operation. We then have

$$\frac{E_{NTL}}{E_r} \sim 3, \text{ HV Mode} \tag{4.14}$$

Now there are more NTL than recoil phonons, hence we've amplified our signal. This illustrates the advantages and trade-offs between the two modes of operation. In iZip mode we have good discrimination between signal and backgrounds, but are limited to higher analysis thresholds. In HV mode we lose this discrimination power, but we gain access to lower thresholds by amplifying the phonon signal.

4.2.7 Calibration Strategy

We want to push our detectors to thresholds of ~ 10 eV. In order to demonstrate that we have accomplished this goal, we need to actually calibrate the detector in this regime. It is easy enough to use a laser or x-ray source to generate excitations at this energy, but the problem becomes more complicated if we require the energy to be deposited in the detector bulk, as a photon $\lesssim 1$ keV will typically only penetrate $\lesssim 100\mu$m into the detector. Interactions near the surface are affected by the concentration of crystal defects and fringing effects in the applied E-field alter, which alter their reconstruction from events in the bulk, where our DM signal occurs.

A convenient calibration in germanium is provided by electron capture decay of ^{71}Ge to ^{70}Ga with a half-life of 11.3 days. This decay generates lines as low as 160 eV from capture of an M-shell electron [27]. The isotope can be produced in the detector bulk via excitation of naturally abundant ^{70}Ge with a neutron source, such as ^{252}Ca.

There are no such capture lines in silicon. One strategy to calibrate a Si detector in a position independent manner would be to observe the steps in the Compton background associated with scattering off the inner electron shells, though the feasibility of this depends strongly on the detector resolution. An alternative approach involves placing the silicon detectors in close proximity to germanium and exposing both to a high energy external source, such as ^{133}Ba. Since the germanium detector will already be calibrated from the internal activation, the observation of events where a gamma from a known line scatters in both detectors would allow one to "cross-calibrate" the silicon detector across a wide energy range. To facilitate this cross-calibration, the SuperCDMS detector towers are arranged such that each silicon detector is "sandwhiched" between two germanium ones.

4.3 Overview of SCDMS Backgrounds

DM interacts rarely with our detector, so we want our experiment to be as sensitive as possible. But with this sensitivity to DM also comes an increased sensitivity to backgrounds, so we need stringent control of our backgrounds.

Recall that we can generically classify all interactions as either NR or ER. Regardless of the ultimate source, ER backgrounds in the detector typically manifest as gammas and betas. NR backgrounds on the other hand usually arise from neutron radiation, or a heavy nucleus recoiling after some decay. Typical ER background rates tend to be higher than NR rates. This is not really an issue when operating an iZip, where we are typically looking for DM which interacts via NR and we are able to reject the ER background. For an HV detector we lose event-by-event discrimination power, and the background will be dominated by ERs regardless of our DM model. This suggests another synergy between the two detector modes. The iZips can be used to measure the ER background, which can then be subtracted from the HV DM search data.

4.3.1 Detector Contamination

Though SCDMS uses very high purity silicon and germanium in its detectors, they are not completely pure and some of the impurities are radiogenic. Of particular concern is ^{32}Si ($\tau_{1/2} \approx 150$ years), which occurs naturally in silicon, and ^{3}H ($\tau_{1/2} \approx 12$ years), which can be produced in both germanium and silicon through cosmic ray interactions in the detector bulk [77]. Both sources undergo β-decay, producing electron recoil backgrounds. The rate is considerable, and is currently expected to be the dominant background for the HV detectors [78]. The ^{3}H background is mitigated by limiting exposure of the detectors above ground, and great care must be taken to track this exposure.

The origin of ^{32}Si is ultimately cosmogenic as well, though it takes a much longer detour to reach our detectors. One hypothesis is that it is produced by cosmic ray interactions in the atmosphere, and then propagates to sources of bulk Si through the water cycle [79]. Notably, measurements of contamination in different Si samples appear to yield a wide range of activities. When estimating the total expected background, SuperCDMS adopts the conservative end of this range, but we have not yet directly measured the ^{32}Si activity in our detectors. Hence a dedicated measurement would be very interesting and has the potential for significant impact on our picture of the background.

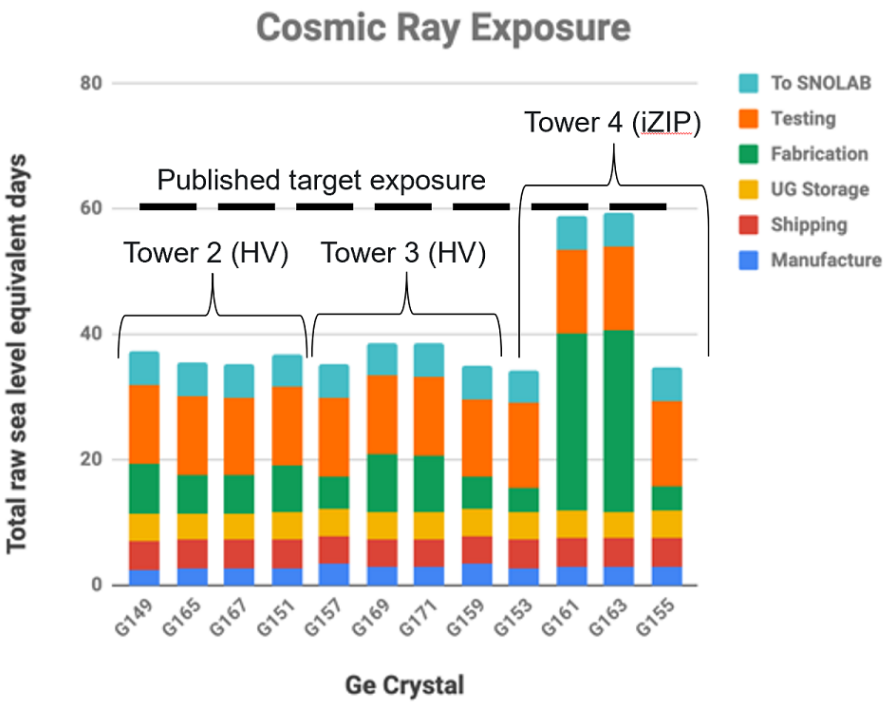


Figure 4.6: Above ground exposure of the SuperCDMS germanium crystals, which determines the tritium background. The plurality of the exposure for the HV detectors came from detector testing, while almost half of the iZip exposure was from fabrication. Detectors in all three towers are within the exposure goals presented in [78]. The remaining iZip tower (Tower 1) used previously fabricated crystals, and has a larger exposure.

4.3.2 Cosmic Rays

Cosmic ray protons can scatter inelastically off the atmosphere, creating secondary particles which can reach the surface of the earth and potentially be detected. To shield against this background, SCDMS is operated underground. SNOLAB is at a depth of 2 km underground, with the rock overburden providing about 6000 meters water equivalent (m.w.e.) of shielding. Of all the particles produced by cosmic rays which could potentially interact with the detector, only muons have a long enough mean free path to present a significant background at this depth. The muon flux is reduced by the overburden from $2 \times 10^{-2} cm^{-2} s^{-1}$ at sea level to $3 \times 10^{-10} cm^{-2} s^{-1}$ at SNOLAB [80].

Though the muon can hit a detector, this is a background of marginal concern. The expected $\sim$ GeV energy deposition by a muon is well above the region of interest when searching for DM, and the main drawback is that this produces a considerable amount of dead time as the detector cools back to baseline. Muons can still produce backgrounds capable of mimicking a DM signal via production secondary neutrons through muon capture or spallation with a nucleus in some nearby material [81]. These can occur from a muon interacting with the rock overburden or with the experimental shielding itself.

4.3.3 Material Contamination

All the materials used to construct the SCDMS experiment will have some amount of contamination from radioactive impurities. ^{238}U, ^{232}Th, ^{40}K and ^{60}Co are all long-lived isotopes which are common in many materials, with U and Th having long chains of decay daughters [78]. This collection of radioistopes can produce a variety of potential backgrounds, including gammas and βs inducing electron recoils, as well as neutrons from spontaneous fission or secondary products of α-decays causing nuclear recoils.

Since this contamination is so ubiquitous, SCDMS assays the radiopurity of all of the materials it uses. This ensures the background from material contamination will be low enough to not significantly affect the sensitivity of the experiment. The material assay program is carried out using a variety of methods, including gamma screening with high purity germanium detectors, and measuring the concentration of ^{238}U and ^{232}Th in materials via inductively coupled plasma mass spectrometry (ICP-MS). The assay also influences some of the shield design, with the high purity materials placed closest to the detectors in order to block potential backgrounds from less pure components.

4.3.4 Environmental Backgrounds

The rock in the cavern walls surrounding the experiment also contains radioisotopes. Typical concentrations in rock are 1.11 ppm ^{238}U and 5.56 ppm ^{232}Th [80]. U can produce neutrons either spontaneous fission, while both U and Th undergo α-decay which can result in neutrons through (α, n) reactions in the bulk of the rock. gamma-decays are also present in the decay chain of each isotope, potentially creating electron recoil backgrounds. Since the contamination of the rock cannot be lowered, the experiment must really on shielding to mitigate this background. We will discuss this background source in more detail in 5

4.3.5 Radon

Radon is a radioactive noble gas which occurs partway the decay chains of ^{238}U and ^{232}Th. Being an inert gas it has the ability to diffuse through certain materials before it decays to the next daughter in the chain. This can appear in our material radiopuritiy assays as “broken equilibrium”, where all the members of the chain above radon are measured as having a common activity, and those below have a distinct activity. The fact that radon is generated by decays of primordial radionuclides and can diffuse means that it is quite ubiquitous in the air.

Radon in the air can be mitigated in a straightforward by purging that air and replacing it with pure nitrogen. The SuperCDMS SNOLAB shield includes a radon barrier in order to perform this purge around the cryostat. This is effective enough to expect the background from prompt radon decays to be subdominant.

Unfortunately radon's ability to generate backgrounds is not limited to prompt decays. If a material is exposed to air contaminated with radon, some amount will diffuse into the material before decaying, leaving a concentration of radioactive daughters just under the surface of the material. Notably, this can occur with copper, which the SuperCDMS detector housings are constructed of. Hence we expect to see some contamination from radon daughters on the surface of the detector housings, which can create a considerable background. The most straightforward way to combat this is to limit the exposure of copper to dirty air by keeping it under a nitrogen purge.

4.3.6 Neutrinos

Our detectors are also sensitive to neutrinos interacting through coherent elastic scattering off the target nucleus, AKA CEνNS[2]. Reactions in the p-p chain make the Sun an effective neutrino factory. Once our detectors achieve a sufficient sensitivity these events will dominate the expected DM signal. At DM masses $m_\chi \sim 10$ GeV, this occurs at cross sections $\sigma \sim 10^{-45}$ cm^2, determined by the ~ 11 MeV endpoint of the ^{8}Be decay spectrum [82]. This background clearly cannot be shielded away. Our ability to search for DM past this "neutrino floor" will depend on our ability to subtract this background from our spectrum. Current sensitivity projections of the experiment identify CEνNS as the dominant source of NR backgrounds in our detectors [78].

4.3.7 Low Energy Backgrounds

As experimental searches for DM have pushed to unprecedented low thresholds, they have encountered new backgrounds $\lesssim 100$ eV which are not yet well understood. An excess event rate above the expected background due to Compton scattering and noise triggers has been observed in a number of different experiments. This excess has been observed both above and below ground and in different detector materials. Though one might naively hope that this excess is just the DM signal we've been searching for, the shape and energy scale do not appear compatible with this hypothesis [83]. Since the source of this background is not well understood, it is also difficult to confirm that each observation has a common cause.

[2]Coherent Elastic ν-Nucleus Scattering

The source of such an excess is an active area of research, and several production mechanisms have been proposed. One possibility is that it could be secondary ionization resulting from $\beta-$decays. Even if β doesn't hit the detector itself, it can generate low energy photons via Cerenkov or transition radiation [84, 85]. Another candidate is that these excess events are induced by stress on the detector, potentially due to how it is held. This hypothesis is supported by the observations that the background appears to be non-ionizing, and in cryogenic environments it appears to decrease with time since the last cool down [86]. Since the low energy excess is not well understood, it is very well possible that both of these sources can contribute, along with some yet to be identified. In any case, measuring, characterizing and understanding these backgrounds will be critical for pushing DM searches to lower masses.

4.4 SCDMS Shielding and SNOLAB Infrastructure

SNOLAB, located $\sim$ 2 km underground, sharing a shaft with an active nickel mine near Sudbury ON, was built as the site of the Sudbury Neutrino Observatory (SNO). Since its inception, it has expanded its facilities to accommodate other low background experiments and SuperCDMS-SNOLAB is currently presently (as of 2023) under construction therein. Locating the next generation of SuperCDMS here already gets us ahead on controlling our backgrounds. As already mentioned, the large rocky overburden substantially reduces the flux of cosmogenic muons from that at Soudan. Additionally, the SNOLAB facility is operates as a class-2000 cleanroom. This allows for close monitoring and control of external contamination in the form of dust and radon to the experiment, as well as location to perform radiopurity assays in a low-background environment.

To protect from the remaining external, such as though originating from radiogenics in the walls of the surrounding cavern, SCDMS relies on a shield made from water, high-density polyethylene (HDPE) and lead. Water and HDPE both contain a significant portion of hydrogen, which makes them effective at shielding from neutrons [78]. Lead, being a high Z material, does well at shielding gammas. However, it is also a potential target for muons to scatter and produce neutrons.

In order to take advantage of the complimentary shielding properties of each component, they are arranged in an alternating pattern surrounding the detectors. The outermost layer is a 60 cm thick water tank resting on a HDPE base, which acts as a first line of defense to reduce the environmental neutron flux by $\sim 10^6$. The next layer is then formed by 23 cm of lead, which attenuates the incident gamma flux by an estimated factor of $\sim 10^5$. blocks environmental gammas, and is estimated to attenuate the flux by $\sim$105. Finally

there is an additional neutron shield composed of 40 cm HDPE. This layer protects against any potential neutron backgrounds generated in the lead, which can be produced by cosmogenic muon interactions.

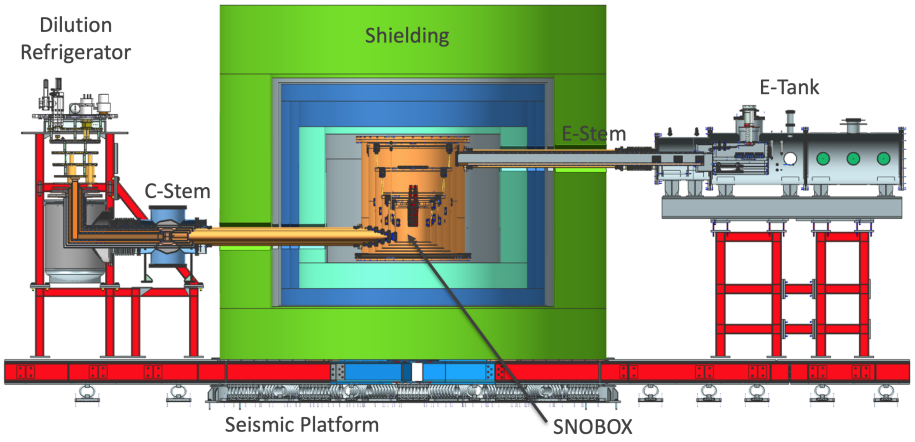


Figure 4.7: Diagram of SCDMS-SNOLAB shielding and nearby mechanical infrastructure. The green outer layer of the shield represents the water tank, followed by the dark blue lead shield and light blue inner poly. To the right sits the E-Tank which houses the warm electronics, and left is the fridge which cools the cryostat to base temperature. Both are connected to the cryostat by stem penetrations in the shield. The entire structure rests on a seismic platform to isolate from vibrations which may be produced by blasts from the neighboring mine.

Though this layered shield design is in principle very effective at stopping both gammas and neutrons, practical design considerations leave a few points of vulnerability. Not every component of the experiment can be enclosed in the shield, and it is necessary to leave two gaps, collectively referred to as the "stems", in the shielding which connect the internal and external components. One connects the cryogenic portion of the experiment to the dilution refrigerator outside the shield, and the other carries the signal readout cables from the detectors to the "E-Tank" which houses warm electronics. These are named the c-stem and e-stem, short for "cryo" and "electronics", respectively. The stems, along with the shield and external components can be seen in Fig. 4.7.

Particular attention must be paid to the stems when studying the environmental background, since they are the easiest point of penetration. It is also worth noting that neutrons an gammas will behave differently as they move through the stems. A gamma which goes through the stem at some large angle will hit the shield and typically stop, so the ones which penetrate the shield must travel down the line-of-sight of the stem.

If a neutron enters the stem at an angle, it has a good chance of bouncing and continuing to travel down the stem. Hence the neutrons which penetrate the shield are less likely to travel along the stem line-of-sight, so the offset position of the detectors does not protect from neutron backgrounds as effectively as it does for gammas. This makes it somewhat more difficult to shield from stem neutrons, and the bouncing behaviour complicates how these neutrons penetrate the shield.

Chapter 5

Simulation of Environmental Neutron Backgrounds

5.1 Motivation

Radiogenic neutrons are constantly being emitted from the walls of the cavern surrounding the experiment. This background is generated from decays of ^{238}U and ^{232}Th naturally abundant in the norite rock. Neutrons represent a particularly challenging background for dark matter which interacts via nuclear recoil, as a single isolated neutron scatter would be indistinguishable from a dark matter interaction.

Our shield was designed with blocking exactly these neutrons in mind. The outer water layer and the inner HDPE layer both contain large fractions of hydrogen, which serves as an excellent neutron moderator from the view of kinematic matching. One can view the water shield as our first line of defense against the cavern neutrons, while the inner poly is kind of a back up since the intermediate lead layer can be a target for the production of high energy neutrons from muon spallation. Hence we expect neutron events in the detector to be rare. But the shield isn't perfect at stopping neutrons. A very small but non-zero fraction of the neutrons should be able to traverse the entire shield. What's worse is the stems. We require two penetrations in the shield, one to connect the cryostat to the fridge and one to connect the warm to the cold electronics so that we can actually read the detector output. Though there is some material stuffed into the stems in the form of copper piping or wiring and plastic, they represent a mostly unobstructed path for

neutrons (or gammas for that matter) to pass through the shield and reach our detectors.

Since we're performing a rare event search, we need to have a good estimation of all our backgrounds including the cavern neutrons. The most robust and conceptually straightforward to carry out this estimation is via Monte Carlo simulation. Using some package which encodes all the relevant physics processes and cross-sections, in our case Geant4 [87], we implemented a model of our experiment and threw simulated neutrons at it to see what crosses the shield with sufficient energy to excite our detectors. The fact that we expect these backgrounds to be rare is great from the perspective of running the actual experiment, but is actually a significant challenge in terms of carrying out the simulation. That is, the shield is somewhat of a victim of its own success. The more efficient we are at stopping the neutrons from reaching our detectors, the more neutrons we need to simulate in order to arrive at a statistically significant estimate of the background.

5.2 Cavern Neutron Spectrum

The isotopes ^{238}U and ^{232}Th have half-lives comparable to the age of the Universe, at 4.5 and 14 billion years respectively. Hence we expect to find some primordial trace of these species to be present in the Earth's crust. Both isotopes are progenitors of complex decay chains which result in several α, β and γ decays, as well as spontaneous fission (SF). These spontaneous fission processes necessarily emit neutrons, and the majority of the neutrons are induced as secondaries following α-decays along the chains. An energetic α is able to scattering on nearby a targets in the surrounding rock, causing a spallation and releasing a free neutron - a process referred to as (α, n). A summary of these production mechanisms is given in [74]. The precise nature of the spectrum generated by the (α, n) process is highly dependent on the chemical composition of the medium in which the α-decays occur, but we gain some qualitative understanding of the shape by considering the energy released in the primary α-decay. Regardless of the details of the (α, n) process, the energy of the emitted neutron cannot exceed that of the inducing α. Hence we estimate the energy scale of (α, n) spectrum by considering the energy released by α-decays along the 238U and 232Th chains. We we look at these energies (see [74]), we see that α's are typically emitted at $\sim$4 MeV (e.g. the decays of each chain progenitor) and that they do not exceed 9 MeV with the highest energy being the 8.9 MeV decay of ^{212}Po.

The detailed neutron spectrum produced by the ^{238}U and ^{232}Th chains in the bulk of the rock surrounding the cavern, including both SF and (α, n) processes, is calculated using SOURCES4C [88]. This gives us the

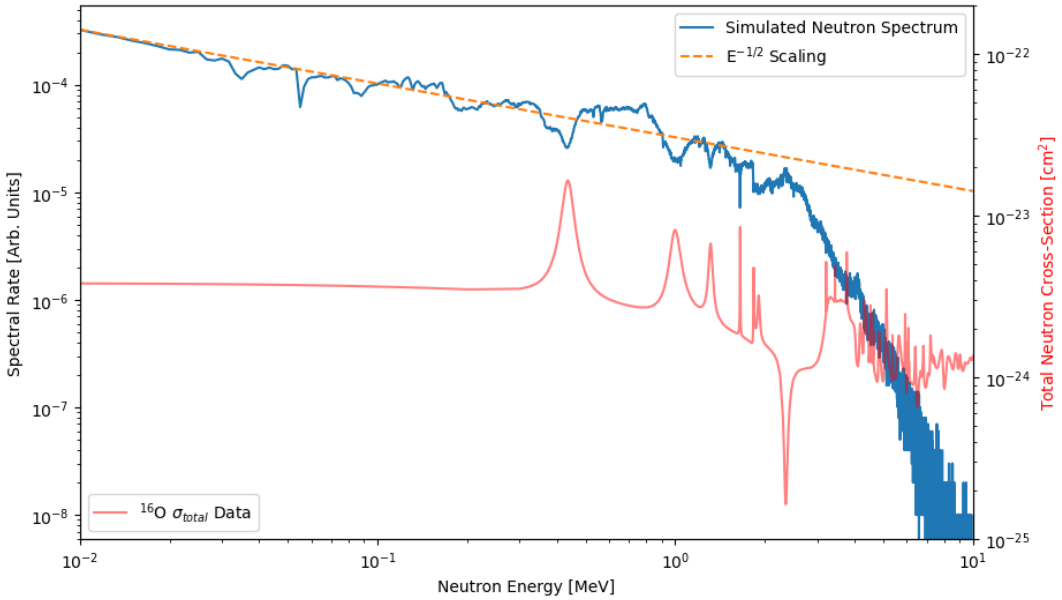


Figure 5.1: Simulated spectrum of neutrons produced by ^{238}U decays in the norite rock which enter the cavern. Neutrons are also generated by ^{232}Th decays in norite, as well as decays occurring in the thin shotcrete layer which coats the cavern. However, all these spectra have a similar enough shape that we can use the above spectrum as a reasonable proxy for all the neutrons. We also show the total neutron interaction cross section on ^{16}O for reference. There are clear dips in the simulated spectrum corresponding to resonance peaks in the cross section data.

energy of the primary neutrons, but these neutrons must travel through the rock itself before they can enter the cavern, which presents an opportunity for them to scatter and lose energy before reaching our detectors. In order to determine the spectrum which has been attenuated by the rock before entering the cavern, we use the spectrum derived from SOURCES4C to perform a Geant4 simulation of the neutron transport through norite. The generation of the spectrum is detailed in reference [89]. The spectral shape resulting from this simulation is shown in Fig. 5.1. We see that the shape generally matches our qualitative intuition - the rate scales like $\sim E^{-1/2}$ up to a few MeV, then sharply falls off until its ultimate endpoint around 9 MeV. There are numerous peaks and valleys on top of these generic features, corresponding to resonances where the free neutrons are absorbed or transmitted through the constituents of the surrounding rock. To confirm this, we show the total neutron interaction cross-section on ^{16}O [90], the most abundant element in the rock, for comparison with the rate entering the cavern. We see that all of the features in the cross-section data below the energy where the neutrons fall off are closely mirrored in the spectral shape.

The neutron environment underground spans a wide energy range, from thermal energies up to $\sim$ 10

Target	Detector Type	ROI Edge (eV)	Min. E_n (eV)
Ge	iZip	10^3	1.85×10^4
Si	iZip	10^3	7.5×10^3
Ge	HV	3	56
Si	HV	3	22

Table 5.1: Minimum kinetic energy of an incident neutron required to produce a recoil in the ROI for each detector flavor.

MeV. Since the qualitative neutron behavior can have a huge variation across this range, we should take a moment and attempt to define what energies we are most concerned about from a backgrounds perspective. For an elastic nuclear recoil, the largest possible energy ER which can be deposited in our detector by a particle with kinetic energy E_n is

$$E_R \leq \frac{4m_T}{m_n}\big(1 + \frac{M_T}{m_n}\big)^{-2} E_n \tag{5.1}$$

where m_T is the mass of the target and m_n is that of the incident particle. In this case we're interested in a neutron scattering off a Si or Ge atom, but the result also holds for a WIMP scattering off a generic target. Our projected analysis thresholds for the iZip and HV detector types are 1 keV and 3 eV respectively [78].

The minimum kinetic energy required for a neutron event to appear in our region of interest (ROI) for each relevant target material and detector type is summarized in Table 5.1. Recall that NR and ER backgrounds are distinguishable in an iZip detector but not an HV. For this reason our first concern when estimating the NR background is that in the iZips, since the NR background in the HV will be subdominant to ERs in most cases. We restricted our analysis the elastic NR in the iZips, and it is reasonable to kill neutron tracks at energies <1 keV. If we were more ambitious and interested in a complete background model for the HV, this cutoff is of course unreasonable. A choice for cutoff has already been made for us in some sense during the generation of the neutron flux spectrum in Fig. 1, which only describes the spectral shape down to 10 keV, below which we didn't throw any neutrons. Since this is the shape we used to simulate our backgrounds, we should keep this cutoff in mind in case we see any artifacts. We used this as the threshold to distinguish between "fast" and "thermal" neutrons, with the former category describing those which fall above the threshold. This is a somewhat crude classification, considering that in many context one might describe $\sim$ keV scale neutrons as "epithermal" and reserving the terms "thermal" for those with kinetic energies <1 eV. Though a bit unorthodox, this scheme allows us to directly compare those distributions without needing to renormalize.

The benefit of applying a energy cutoff stems from the fact that fast and thermal neutrons have qualitatively different behaviors. Not only are the total interaction cross-sections for thermal neutrons typically ∼1-2 orders of magnitude larger than those of fast neutrons, but very term "thermal" implies that the neutrons have scattered many times and achieved a Maxwellian velocity distribution. Fast neutrons on the other hand still carry the spectrum characteristic of their production mechanism. Hence thermal neutrons ideally behave like a kind of gas, while fast neutrons act as free-streaming radiation - that is, until they start to scatter and thermalize. In principle, we should care about all the neutrons. Even though a thermal neutron with kinetic energy 1/40 eV cannot generate an elastic recoil in our ROI, they have a considerable capture cross-section. This capture (on which isotopes?) is then followed by a nuclear deexcitation emitting an ∼MeV gamma, and imparting some ∼100 eV energy to the deexciting nucleus, resulting in a nuclear recoil within our HV ROI. For a simple two-body decay, the recoil energy is monoenergetic. However, the deexcitation does not always transition directly to the ground state, but typically occurs as a cascade over finite time. The details of this cascade are not modelled natively in Geant4, hence our simulation is not optimized for studying the background generated by thermal neutrons - and other specialized tools should be used ([91–93]).

5.3 Neutron Flux

In line with our discussion in the previous section, measurements of the neutron flux typically target either the thermal or fast populations. The standard reference for these values used by most experiments at SNOLAB is the "SNOLAB Handbook" [94], which gives flux values of 4140 ± 120 $1/\mathrm{m}^2/\mathrm{day}$ for thermal neutrons and ~ 4000 $1/\mathrm{m}^2/\mathrm{day}$ for fast neutrons. Two aspects which we should note about these numbers right off the bat:

1. The value for the thermal flux comes from an actual measurement, while the fast flux is an estimate.

2. Neither flux value is normalized to the solid angle.

I mention point 1 because it is often overlooked that our standard normalization for the fast neutrons is not actually based on a concrete measurement - so there is a considerable "systematic" uncertainty associated with all the rates we estimated. Point 2 is important to mention because it tends to be a perennial point of confusion. Without the details of the measurement performed to yield these values, the numbers themselves

are ambiguous. One can imagine two scenarios - one where the detector is placed in the middle of the cavern and detects neutrons in 4π, and one where the detector is up against the cavern wall and sees neutrons in $\sim 2\pi$. This point may in fact be moot for the fast neutrons since we don't know what measurements that flux estimate is based on, but this a concrete question for the thermal flux. Then, since the numbers are listed together in the SNOLAB handbook with the same units it is reasonable to assume that the same normalization applies to both values.

The details of the thermal flux measurement are discussed in Browne [95]. These measurements were performed using ^{3}He proportional counters to estimate backgrounds for the SNO experiment. From the context presented in the reference, we can conclude that the measurement covers all angles and hence the stated thermal flux is in 4π. Hence the proper normalization for the thermal flux should be 4140 ± 120 $1/\mathrm{m}^2/\mathrm{day}/(4\pi$ sr). Though the fast flux is not explicitly measured, some hints are provided which we can use to check the estimate presented in the SNOLAB handbook. First, Browne argues that the shape of the neutron spectrum at SNOLAB should resemble that at Modane. Then we look at measurements taken of the flux at Modane, where they found that $15 \pm 5\%$ of the total neutron flux populated the region 1-10 MeV [96]. Browne then claims that the total neutron flux over all energies observed in SNO is ~ 9000 $1/\mathrm{m}^2/\mathrm{day}/(4\pi$ sr), and hence argues that the flux in the 1-10 MeV range would be 1350 ± 450 $1/\mathrm{m}^2/\mathrm{day}/(4\pi$ sr). Once again the solid angle normalization is not made explicit in the text but, considering the normalization of the thermal flux and the fact that SNO was certainly sensitive to neutrons in 4π, I assume that this value is per 4π. Now we can use this number and compare it to our simulated spectrum. Integrating the curve in Fig. 5.1, we find that 36% of the rate occurs in the 1-10 MeV range. Hence the proper normalization for the total simulated spectrum would be 3750 ± 1250 $1/\mathrm{m}^2/\mathrm{day}/(4\pi$ sr), which is consistent with the value quoted for fast neutrons in the SNOLAB Handbook. We should keep in mind that this is still just an estimate, and that we relied on two big assumptions:

1. The shape of the neutron spectrum observed by SNO is sufficiently similar to that measured at Modane.

2. The differential neutron spectrum in the SNO cavity matches that in the Ladder Lab, where SuperCDMS is located.

The above assumptions have many opportunities to fail. The neutron spectrum depends first and foremost on the concentration of ^{238}U and ^{232}Th in the rock, and then the chemical composition of the rock itself. It is clear that we would expect these specifications to change when we go from Modane to SNOLAB, and even

within SNOLAB itself there should be some variation. For a realistic determination of the radiogenic neutron background, SuperCDMS (along with all the denizens of SNOLAB looking for dark matter) would benefit from a dedicated measurement of the fast neutron flux. But barring the existence of this measurement, we have argued that the value of 4000 $1/m^2/day/(4\pi$ sr) quoted in the SNOLAB Handbook is a reasonable number to use in our simulations.

5.4 Simulation Framework

The SuperCDMS collaboration uses the SuperSIM, a Geant4 application, for its simulation studies. The utility of SuperSIM over pure Geant4 code is that it gives us a modular framework to change between the many complex geometries and sources we're interested in.

There are a huge number of choices we made when running the simulation. Below I'll summarize these choices, dividing them between "floating" ones which vary as we iterated over different studies and "fixed" ones which we did not change.

Floating Parameters

1. **Geometry**: We need something for the particles to interact with. This includes all of the experimental infrastructure from the detectors themselves to the shield and the surrounding cavern. We must make choices regarding the fidelity of the details (we don't model individual screws or resistors), and there have been changes to the geometry as the experimental design was updated (e.g. reduction of the cryostat). Now the experiment is actually being built and we have a frozen geometry, but it's possible that certain components will be found to have insufficient detail when construction finishes.

2. **Particle Source**: We must model where the particles originate from. This involves either selecting a spectrum or delegating Geant4 to simulate the details of the decay process. We also must specify a volume or surface to throw the particles from.

3. **Active Volume**: We can't record every particle interaction in our little universe, and must choose where we actually perform detailed particle tracking. The obvious choice is the detectors themselves, since we're ultimately interested in how much radiation interacts with them. However, when simulating a rare event we typically expect nothing to actually reach the detector. In these cases it can be useful to record particles which cross a certain layer of the shield, which can either be used to make a qualitative

estimate of the expected background, or we can use the information to attempt to define a reasonable distribution which would allow us to rethrow particles from that particular shield layer.

Fixed Parameters

1. **Physics**: We can't simulate the entire standard model, we have to chose which processes are relevant to the scenario we're studying. For most of our background simulations, including the cavern neutrons, we use the Geant4 "shielding" physics list.

2. **Detector Response**: Geant4 describes the history of a particle in terms of individual microscopic steps, which occur on scales much too fine for our physical detector to resolve. We model our detector response according to the following definitions. Note that these serve as a kind of stopgap to use until a complete Detector Monte Carlo (DMC), which simulates all the processes of a particle interaction in our detector from phonon propagation to the electrothermal response of the TES, is ready for production.

   (a) A "detector hit" or "zip hit" refers to the sum of all G4 steps in a single detector within a 1 μs window of the first step in the detector. The energy deposited by the hit must also be above the analysis threshold, defined here to be 2 eV for HV and 350 eV for iZips, in order to be registered as a hit.

   (b) A detector hit is considered to be in the region of interest (ROI) if it deposits an energy in the range 3 eV-2 keV in an HV detector or 1-50 keV in an iZip.

   (c) If two detector hits occur within a 1 ms window of each other in separate detectors, they are considered to be "multiples". Otherwise the detector hit is classified as a "single". Since a dark matter particle is unlikely to interact in more than one detector, only singles represent backgrounds to a dark matter search.

   (d) We expect our detectors to respond differently to NRs vs ERs, but the simulation only records energy depositions. The identity of the interactions are tagged after the fact - if the particle depositing energy in the detector is a neutron or nucleus its deposition contributes to NRs and anything else contributes to ERs. An event is classified as a either an NR or an ER if $\geq 95\%$ of total the deposited energy in that event belong to that recoil type.

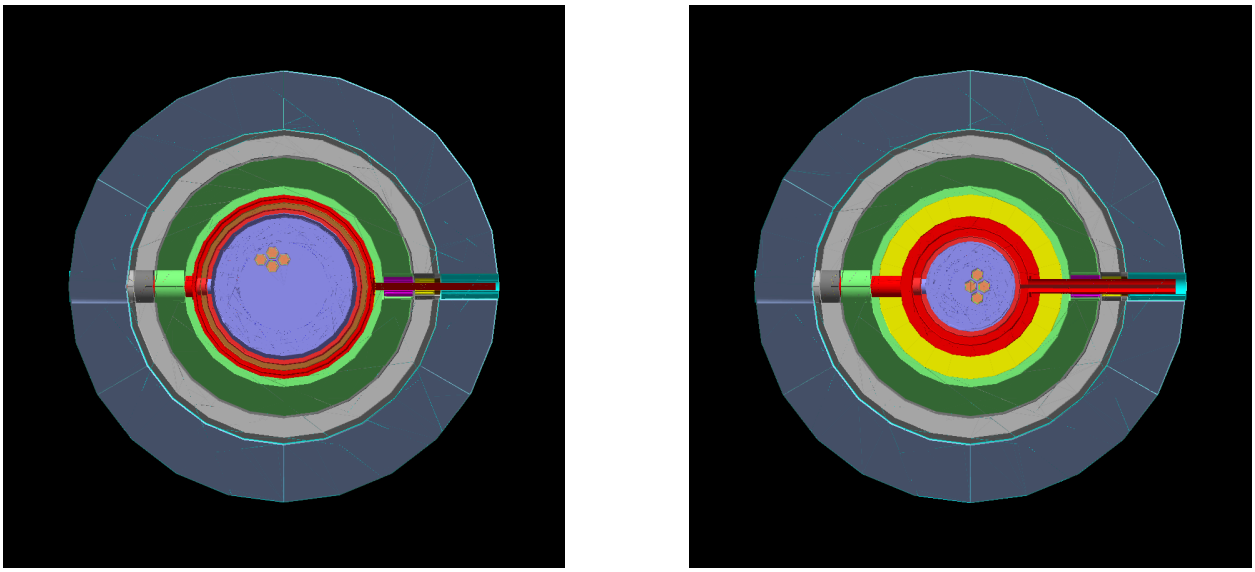
5.5 Simulation Geometry

In the following sections, we will discuss what we learned about the shield from our neutron simulations. We refer back to Fig. 4.7 for a cartoon of the shield and mechanical infrastructure closest to the detectors. In the follow discussion we will generally refer to the outer neutron shield as the "Water Tank" (WT) and the inner neutron shield as the "Inner Poly" (IP).

The design of the SuperCDMS shield has not substantially changed since we have started undertaking simulation campaigns to estimate our backgrounds. What has undergone significant changes are the copper cryostat cans contained within the shield. The original proposed cryostat had a an OVC with a radius of ∼90 cm and could accommodate 32 detector towers, but bids for the construction of a copper can that large while meeting the cleanliness requirements of the experiment would have pushed the schedule back several years. It was instead decided to proceed with a smaller cryostat, large enough to hold 7 towers with a ∼60 cm radius OVC, for the initial run of the experiment. It is this, small cryostat, geometry which is shown in Fig. 4.7. Note that reducing the cryostat did not affect the initial payload, which was always planned to be 4 towers. Expanding to a larger cryostat able to accommodate is still a potential option for a future upgrade.

Since we are thinking about the background from radiogenic cavern neutrons, we should consider whether we expect shrinking cryostat to affect this background. To first order, we don't expect a reduction in the mass of the copper between the detectors and the cavern wall to do much to make much difference on the incident neutron flux, considering that copper isn't a very efficient neutron moderator. On the other hand, in the large cryostat accommodated by the Sep20 geometry the towers had been strategically offset to avoid line-of-sight radiation coming down the stems. When the cryostat shrinks, it restricts our ability to offset the detectors as such - hence the detectors are located along the axis defined by the stems and potentially exposed to more line-of-sight neutrons.

The geometric models of the detectors, cryostat and shield implemented before and after the cryostat rescaling are shown in Fig. 5.2. We refer to the geometry before the rescaling as "Sep20" and that after as "Feb21", based on when these models were tagged in the SuperSim. Simulations to estimate the radiogenic neutron background were performed using both models and we will discuss the results of these in the following section. Given the time between the models and the design changes, there are a number of differences in the perspectives and analysis choices we brought to each analysis, despite the common initial goals. Besides the



(a) Sep20 (b) Feb21

Figure 5.2: Overhead cross section of the SuperCDMS shield and cryostat geometry in SuperSIM. The cryostat is shown in red, and clearly shrank after the "Sep20" geometry was updated to "Feb20". In both cases, the outer dark gray ring represents the water tank, followed by the light gray lead and then the dark green inner poly. The four copper tops of the detector towers can be seen near the center in both images. The towers are offset from the stems in the left, but along the stem axis in the right.

differences in the cryostat geometry, the important differences between the two simulation campaigns are as follows:

1. The campaign run for Sep20 neglected to include the striplines located in the e-stem.
2. Neutrons were thrown from the cavern wall in the Sep20 campaign, and from outside the water tank in Feb21.

The striplines represent the readout cables connecting the detectors to the warm electronics, and are modeled as a thin strip made of Kapton and copper along with copper discs to hold them. Though this is a small amount of material, considering that neutrons tend to penetrate the shield through the stems we expect that even small changes in the stem region could result in large effects on the background. The effect of throwing from the cavern wall versus the water tank is perhaps more subtle. When we ran the Feb21 campaign we assumed that the exact same spectrum as shown in Fig. 5.1 and normalization of 4000 $1/\mathrm{m}^2/\mathrm{day}/(4\pi\ \mathrm{sr})$ which we used when throwing neutrons from the cavern wall apply uniformly to the surface of the water tank. However we also expected that some neutrons have scattered between the

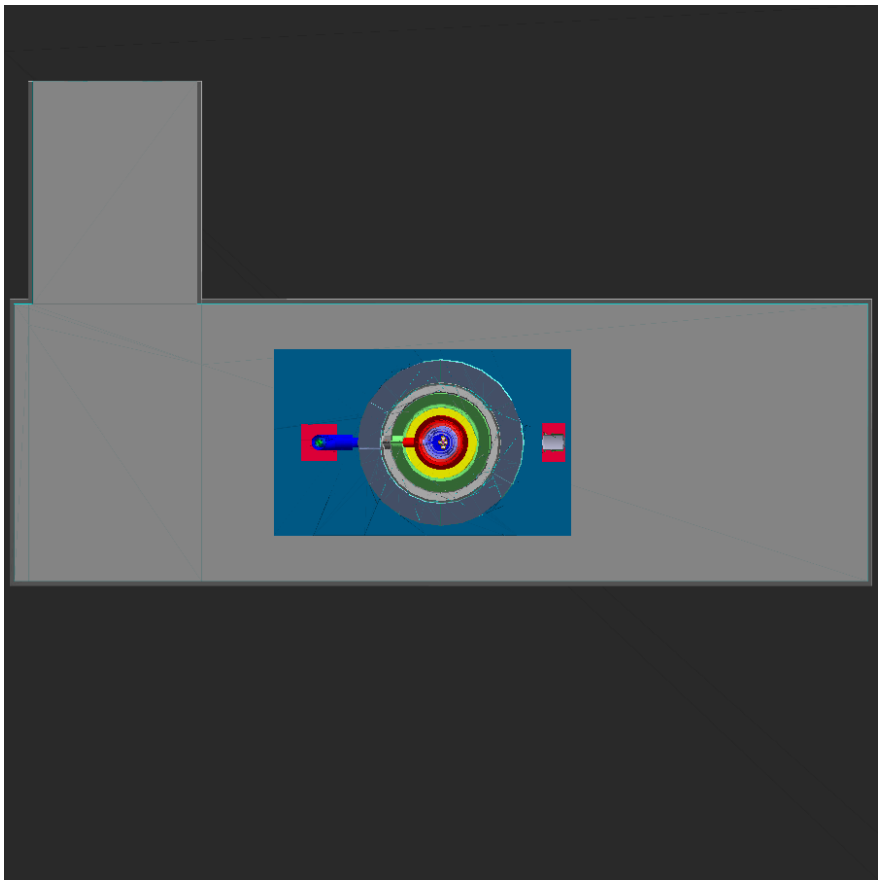


Figure 5.3: Feb21 detector geometry situated in the cavern at SNOLAB. Black represents the norite rock surrounding the gray cavern. Concentric rings near the center show the shield and cryostat, as seen in 5.2b. To the left of the shield is the dilution fridge, and the E-Tank is modeled to the right. These sit on the rectangular blue seismic platform, along with the shield.

time they exit the cavern wall and the time they reached the water tank, thus distorting the spectrum and normalization from our uniform assumption. The shape of the cavern with the Feb21 detector geometry situated inside is shown in Fig. 5.3.

With these two factors in mind, we don't claim that either campaign produced the most accurate estimate of the radiogenic cavern neutron background. Rather, we claim that each one contained a useful element of the background estimate, and that we could look at them simultaneously to come up with a reasonable estimate and help inform future simulation campaigns which will properly account for all of these factors. In particular, we saw from this discussion that the Feb21 campaign did provide us with robust information about how neutrons penetrated the shield, and the Sep20 campaign told us how neutrons transported between the cavern wall and the shield.

5.6 Radiogenic Neutron Simulations and Results

Over the course of my career in SuperCDMS, there have been a number of changes to both the planned experiment and how we perform our simulations. In this section I will discuss several simulations performed to estimate the radiogenic neutron background. Each simulation utilized different versions of G4 and SuperSim, as well as make different assumptions about the geometry of the experiment. This means that we can't necessarily make direct comparisons between the results of each simulation.

5.6.1 Neutrons Thrown From Water Tank

We threw neutrons from the outside of the water tank in order to estimate the background in the detectors. The neutrons were generated according to the original spectrum in Fig. 5.1, and distributed uniformly in space across the surface of the water tank. This served as an approximation for how we expect the flux originating from the wall to appear at the tank. i.e. we didn't account for the fact that they moderate in energy from bouncing around the cavern before reaching the shield. We determined the angular distribution according to Lambert's cosine law, which was meant to reflect the fact that the neutrons have scattered. We later performed a separate simulation throwing neutrons from the cavern to the water tank in order to estimate the sensitivity of our result to these assumptions. We only threw the neutrons oriented inward with respect to the shield. The surface of the water tank has an area of ~ 93 m^2, and the naive expected flux is 2000 1/m^2/day/(4π sr) entering the tank. We simulated $\sim 12 \times 10^9$ primaries, corresponding to ~ 180 years of effective livetime.

Rates in Detectors

First and foremost, we were interested in the background generated in our detectors. We evaluated this from our simulation using the integration time, ROIs and thresholds previously defined. The results are summarized Tables 2+3. For the neutron backgrounds, we're primarily interested in NRs in the iZips, but we included counts in the HV detectors as well as ERs for completeness. It is useful to compare these values to those in [78], where we confirmed that our result was of the correct order of magnitude as well as assess how these numbers factor into the entire background budget. A few things we noticed

1. There was a NR background present in the HV detectors, but it was orders of magnitude lower than the $\sim \mathcal{O}(100$ 1/kg/keV/year) rate in ER backgrounds we expected in these detectors (though we

	Rate [1/kg/keV/year $\times 10^{-5}$]			
Interaction	Ge HV	Si HV	Ge iZip	Si iZip
NR Event	800^{+200}_{-100}	1600^{+500}_{-400}	76^{+9}_{-8}	70^{+40}_{-30}
NR Single	400^{+100}_{-100}	700^{+400}_{-300}	33^{+6}_{-5}	30^{+30}_{-20}
ER Event	< 50	< 200	9^{+4}_{-3}	40^{+30}_{-20}
ER Single	< 50	< 200	2^{+2}_{-1}	< 20

Table 5.2: Simulated radiogenic cavern neutron background rates in ROI for each detector type. Reported uncertainties reflect a 68% Poisson confidence interval.

haven't taken any modeling of the yield into account, so it was not exactly a direct comparison), which is dominated by internal contamination of ^{3}H and ^{32}Si in the detectors, and then by material contamination of components housed within the shield.

2. There was a non-zero ER background induced by the neutrons. This was not due to elastic nuclear recoils, but gammas induced by neutron capture somewhere near the detectors. Once again, this background was negligible compared to the typical ER rates.

3. For all events generated in the ROI, $\sim 50\%$ passed the multiples cut.

Flux Through Inner Poly

Next we take made some assessments of how the neutrons looked after penetrating the shield. We could have technically looked at the neutron flux at the OVC given the simulation configuration, but we avoided this since the statistics at the OVC were too sparse to make robust claims about the distribution. Instead we evaluated the neutron flux exiting the IP toward the detectors.

The stems provide relatively unobstructed access from the water tank to the OVC, hence we expected most neutrons to come through the stems compared to the walls of the poly shield. To characterize the size of this discrepancy, We defined the stem region as that within a 20 cm radius of the central axis of each stem, slightly larger than the 17 cm physical radius of the stem. We used this definition to account for the fact that fast neutrons can penetrate a short distance in the WT to reach the stems and become effectively free-streaming. We also separated the neutrons entering through the top and bottom lids of the IP from those through the side wall.

The fluxes along with the corresponding rates are summarized in Table 6. We present these for "fast" neutrons, defined as those with a kinetic energy >10 keV, as well as over all energies. From this table we

	Flux $[1/\mathrm{m}^2/\mathrm{day}/(2\pi\ \mathrm{sr})\times 10^{-3}]$	
Region	All Energies	$E_n > 10$ keV
Top Lid	6.9 ± 0.2	0.56 ± 0.05
Bottom Lid	4.6 ± 0.1	0.34 ± 0.04
E-Stem	980 ± 10	722 ± 9
C-Stem	358 ± 7	265 ± 6
Side Wall (Ex. Stems)	8.3 ± 0.1	0.76 ± 0.03
Total	16.4 ± 0.1	7.37 ± 0.08

Table 5.3: Simulated neutron flux exiting the inner poly.

saw that if a neutron of any energy was observed through the IP, there was a ∼56% chance it is located in a stem. If we restricted ourselves to fast neutrons, this probability increases to ∼91%. Just as we expected, the stems played a huge role in determining the neutron flux despite their small size. This was especially true for the fast neutrons. We also noticed from this table that ∼2.7 times as many neutrons came through the e-stem compared to the c-stem, regardless of energy. This was just a consequence of the fact that the c-stem was occupied by a series of copper pipes connecting each stage of the cryostat to the fridge, which provided more shielding than the thin striplines which sparsely populated the e-stem. These results are also reflected in Fig. 5.4, which shows a heat map of the neutron flux through the stems and sidewall of the IP.

Once the flux through the IP was simulated, it is tempting to consider rethrowing from this distribution to estimate the rates in the detectors with better statistics. The problem is deceptively difficult however. Not only was the neutron flux not homogeneous over the surface of the IP, but there were correlations between the spectrum as well as the angular distribution to this spatial distribution. The neutrons which came through the stems were less likely to have scattered, and hence we expect them to have a spectrum similar to their generated spectrum, and we expect their momentum vector to be aligned with the central axis of the stem. We show the observed spectral dependence as on the spatial region in Fig. 5.4. As expected, the neutrons through the stems contained a large population of ∼MeV neutrons (along with a considerable thermal population), while the spectrum through the side wall fell off more sharply at this scale.

The angular distribution at each region, where the angle in question is defined as that between the neutron's momentum vector and the normal vector to the surface of the IP where the neutron exits, is shown for each region in Fig. 8. The distribution through the sidewall and top/bottom lids were well approximated by Lambert's cosine law, which is reasonable for particles which have gone through multiple random scatters. That through the stems on the other hand appeared to have a Lambertian component,

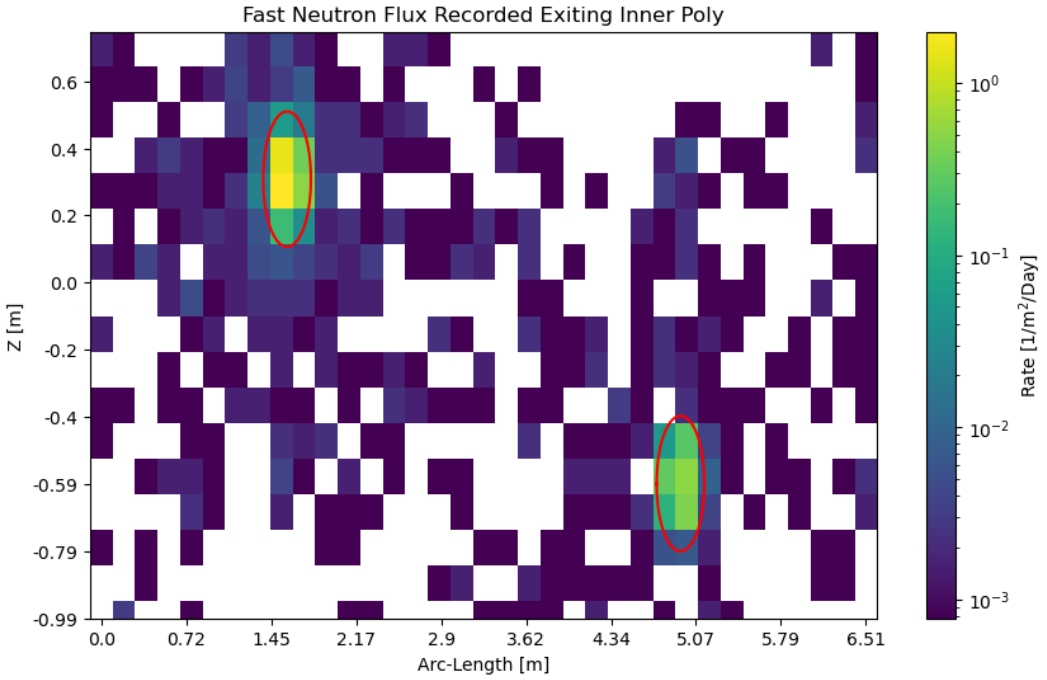


Figure 5.4: Heatmap of the neutron flux through the IP sidewall (including stems). The definition of each stem region is circled in red, where the upper circle represents the e-stem and the lower the c-stem, only considering neutrons with kinetic energy >10 keV. The importance of the stems is clear.

but was clearly heavily biased toward small angles - consistent with neutrons coming down the stem without scattering.

To proceed with rethrowing, one could consider finding the correlation between the angular and spectral distributions in each region, and then performing a series of dedicated simulations where you throw from each region preserving this correlation and then weighting these simulations to the appropriate flux to estimate a net spectrum in the detectors. We see the difficulty of this when we consider how coarse our definitions of the spatial regions are. Recall that we purposefully defined the stem region to be slightly larger than the radius of the physical stem since we wanted to account for the excess flux which we saw near the stems. This is just another way of saying that the boundary between the stem and side wall regions is not well defined. Hence if we really wanted to pursue this "region" based approach to rethrowing, we would ideally want to either define more regions or find some way to interpolate between them.

In any case, special care must be taken when undertaking any rethrowing strategy for simulating rare events. There are two primary reasons for this

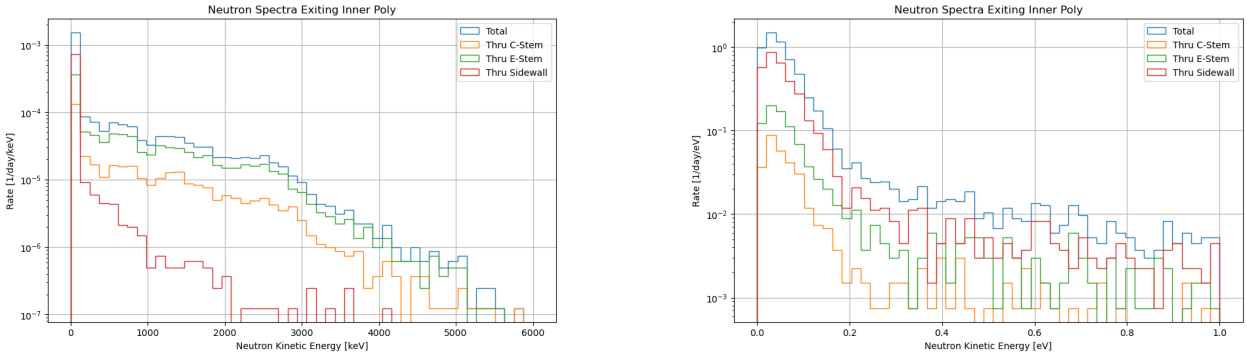
1. If our initial distributions are at all biased, this biased will be tend to be amplified by whatever the factor of efficiency we gain from rethrowing. On the other hand if we're careful from the outset we could "factor out" this bias, *a la* importance biasing.

2. Combining weighted simulations can give us the average spectrum, but in real data we must be acutely aware of the passage fraction and efficiency of the multiples cut. Since this is implemented on an event-by-event rather than an averaged basis, it is quite subtle to model this when rethrowing.

With these considerations, an earnest attempt at rethrowing/reweighting is beyond the scope of this work, but we have presented here an initial characterization of the how the flux, spectrum and angular distribution vary across the inner surface of the IP.

Transfer Function From Water Tank to Inner Poly

We also looked at where on the surface of the WT the neutron which penetrated the the shield originated. This information was useful because it allowed us to estimate what effect a more realistic flux at the WT (compared to our totally uniform assumption) would have on the flux we saw at the IP. The position distribution is shown in Fig. 5.7, where the left plot only considered the origin of neutrons which penetrated through the IP with an energy >10 keV and the right shows all neutrons which get through the IP. Unsurprisingly,



(a) Neutron spectrum over all energies.

(b) Neutron spectrum in the thermal regime.

Figure 5.5: Spectrum of neutrons exiting inner poly. We separate the spectra by region of inner poly, with red indicating those through the sidewall, orange through the C-Stem, green through the E-Stem and blue the total spectrum. At high energy, the spectrum is dominated by the E-Stem and those through the sidewall are clearly attenuated. At thermal energies, the spectrum is largely determined by the sidewall neutrons.

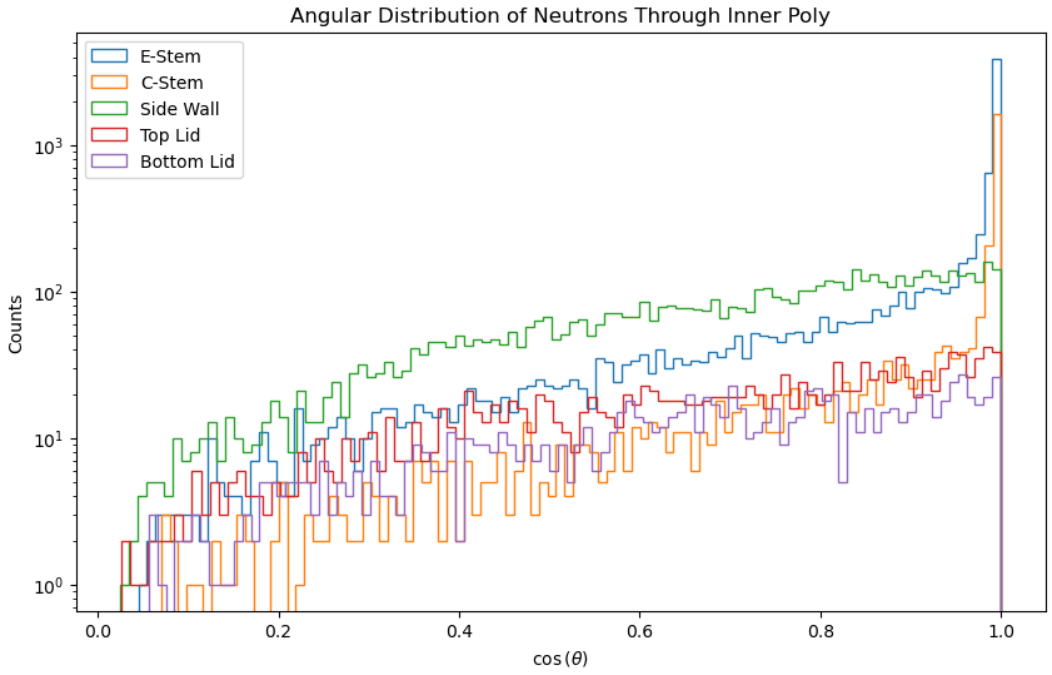
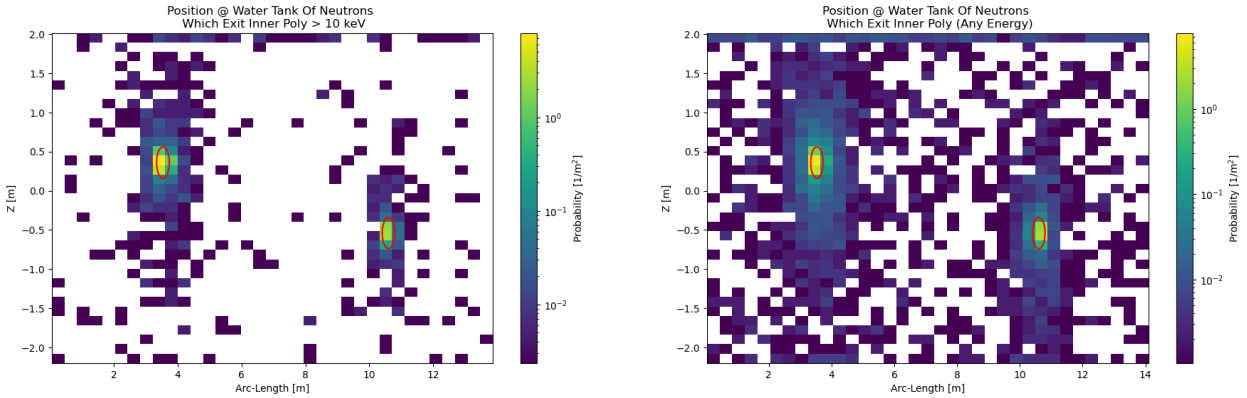


Figure 5.6: Exit angle of neutrons exiting the inner poly at all energies. We break up the angular dependence by the spatial region where the neutron exits. For neutrons which exit through the lids or sidewall (green, red, purple), the angular distribution is approximately Lambertian, proportional to $\cos\theta$. Events through the stems (blue, orange) appear to have a Lambertian component, but are clearly biased to small angles – indicating that they have likely haven't scattered.



(a) Origin of fast neutrons recorded through IP.

(b) Origin of all neutrons recorded through IP.

Figure 5.7: Starting point at WT for neutrons which reach IP. Overdensities (yellow) are centered at the stems (red circles), indicating that they are the preferred point of entry. We also see that these overdensities extend beyond the stem region, indicating that neutrons can leak into the stem without entering the stem directly. These overdensities extend even further if we don't require the incident neutron to have an energy > 10 keV.

most neutrons which made it through the IP originated near one of the stems. But we also saw that the dense region near the stems extended beyond our definition of the stem region up to $\sim$1 m for fast neutron events and slightly further when we included low energy neutrons. This demonstrates that the neutrons still "leaked" into the stem when they hit the shield nearby.

With the information of where the neutrons penetrating the shield originated from, we used the fact that the neutrons were generated uniformly over the WT to estimate the probability of an incident neutron penetrating the shield. We defined the "transfer function" T, or probability that a neutron incident on a given region of the WT will exit a given region of the IP, as follows

$$T_{ij} = \frac{a_{ij}}{N_j} \tag{5.2}$$

Where "i" indexes the exiting region, "j" the entering region, "a_{ij}" is the number of simulated neutrons observed at "i" originating from "j" and, "N_j" is the number of neutrons simulated in that region. Notice that this is a rather coarse definition of the transfer function as we only considered the spatial correlation. In general the probability transitioning between the two given regions in the shield will depend on its energy and angle of incidence, but here we suppressed this complication by integrating over a wide energy range. In Table 5.4 we present the transfer function derived from our simulation. From this we saw that a fast

		Region Entered					
		Top	Bottom	E-Stem	C-Stem	Side Wall	Total
Region Exited	Top	0	0	2.00×10^{-6}	6.67×10^{-7}	2.57×10^{-10}	2.67×10^{-6}
	Bottom	0	0	1.12×10^{-7}	1.21×10^{-7}	0	2.33×10^{-7}
	E-Stem	5.22×10^{-8}	1.15×10^{-8}	3.33×10^{-4}	1.82×10^{-7}	4.87×10^{-8}	3.33×10^{-4}
	C-Stem	2.39×10^{-9}	2.01×10^{-8}	0	1.21×10^{-4}	1.93×10^{-8}	1.21×10^{-4}
	Side Wall	5.26×10^{-9}	4.31×10^{-9}	2.55×10^{-5}	1.04×10^{-5}	4.62×10^{-9}	3.59×10^{-5}
	Total	5.99×10^{-8}	3.59×10^{-8}	3.61×10^{-4}	1.32×10^{-4}	7.29×10^{-8}	–

Table 5.4: Approximate transfer function for neutron transport between the WT and IP. Numbers reflect the estimated probability for an event with an incident fast neutron located at the "entering region" on the WT to result in a fast neutron located at the "exiting region" on the IP.

neutron which hits a stem on the WT has a $\sim 0.01\%$ chance of making it through the IP while remaining fast, while one incident on the sidewall of the WT only has a $\sim 10^{-5\%}$ chance.

5.6.2 Neutrons Thrown From Cavern Wall

We typically assume that throwing neutrons from outside the WT is a good proxy for the neutron flux incident from the cavern wall, but we don't expect them to be exactly the same. Neutrons will tend to scatter between the time they exit the wall of the cavern and hit the outside of the WT, so there will be some change in the two spectra. Additionally, the geometry of the cavern itself can influence the spatial distribution of the flux over the surface of the WT. A more accurate simulation of the background would take these effects into account.

A simulation to estimate the neutron background from the cavern was run with the Sep20 geometry. Neutrons were thrown according to the spectrum in Fig. 5.1 isotropically and uniformly over the surface of the cavern wall. An additional objective was to characterize the flux at intermediate layers of the shield in order to develop a rethrowing scheme which could allow us to quickly estimate how the background would change with a given change in the geometry. With the benefit of some hindsight, we know project is quite ambitious. In the previous section we demonstrated that there is a high degree of position dependence across the surface of a given shield layer, and this is correlated to the momentum associated with the neutron flux - the neutrons coming down the stems look very different than those which make it through the bulk of the shield. This correlation is subtle to characterize at a single given, and it is difficult to develop a robust picture which spans multiple layers. Furthermore, the goals of characterizing the background in the detectors and the flux at the outer layers of the shield are somewhat in tension. The probability of a neutron thrown

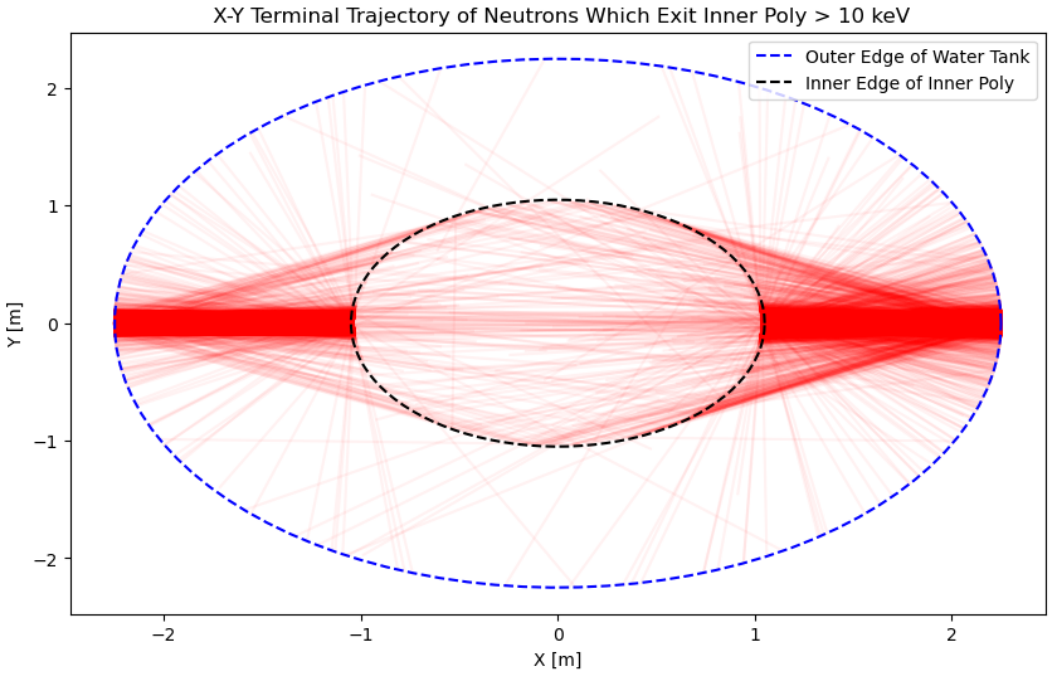


Figure 5.8: Map of terminal locations for neutrons which penetrate the IP with energy >10 keV. Note that this doesn't necessarily describe the actual trajectory of the neutron since we don't record the track between the endpoints. It is interesting that some events show the neutron crossing to the opposite side of the shield before reaching the IP. This serves as graphical representation of the matrix a_{ij} and we can see how many neutrons reach the IP through the stem as opposed to leaking into or out of it.

from the cavern hitting the WT is several orders of magnitude greater than hitting a detector. Given the huge number of primaries we needed to simulate to generate significant statistics in the detectors, we would quickly generate ∼TBs of data at the WT, which is much greater than what is needed to characterize the flux at the WT. Hence estimating the background and characterizing the flux are better suited for distinct studies rather than a single simulation campaign.

Given these difficulties, along with the fact that the Sep20 geometry was outdated, we didn't consider the result of this study to give us an accurate estimate of the background. However, all of the changes to the geometry which have occurred are contained within the shield, and we expect the simulation to give us a reasonable description of the flux at the WT. We evaluated this flux and compared it to the uniform assumption we made when throwing from the WT in order to estimate the effect this would have on our background studies.

Flux at Water Tank

Given our simulation set up, we recorded the position and momentum of neutrons thrown from the cavern wall which are incident on the outside of the WT. To keep our results consistent with our simulation when throwing from the WT, we only recorded neutrons with an kinetic energy >10 keV. It is also possible that a neutron bounced off the WT and scattered around the cavern sometime before hitting the WT again. This would have led us to double count these neutrons in our flux estimate. To mitigate this we only recorded the first neutron which touched the tank.

The observed flux at the WT is shown in Fig. 5.9. It is clear the distribution is not uniform, and some interesting features appear.

1. There is a clear shadowing effect around the e-stem and c-stem, consistent with shielding from the ETank and fridge respectively.

2. There are more neutrons at the top of the shield than the bottom, consistent with shielding from the seismic isolation platform.

3. There are more neutrons above each stem than at the same height on the rest of the shield. This could be explained if we consider the long axis of the cavern to have a moderate collimating effect.

We also looked at the spectrum of neutrons which were incident on the WT, and compared it to the spectrum originating from the cavern wall as shown in Fig. 12. We saw that flux $\gtrsim$ 1 MeV is moderated by a

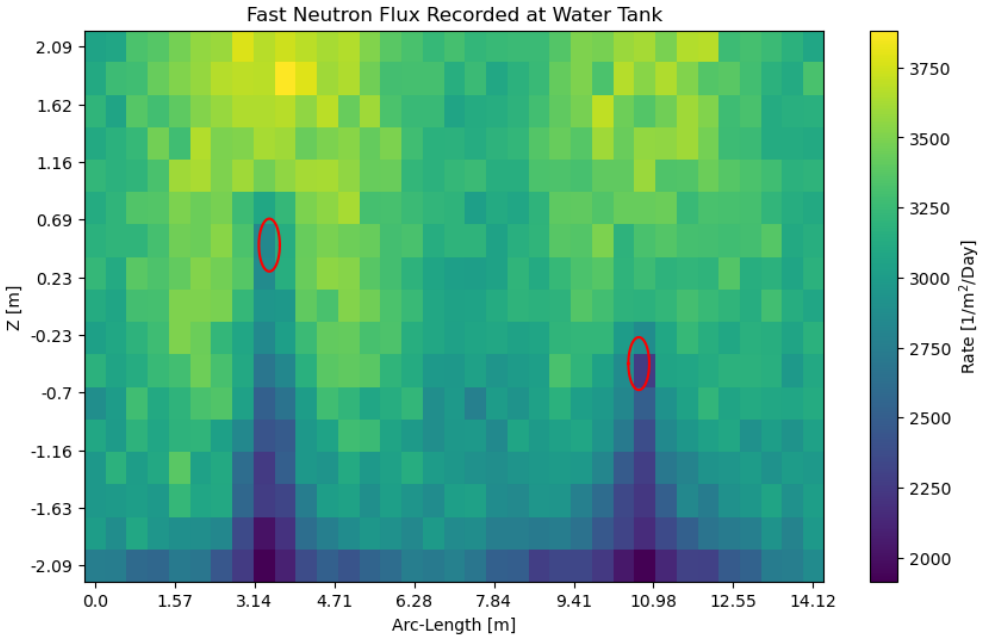


Figure 5.9: Neutrons thrown from cavern wall incident on the water tank. Location of the stems are red circles. We see a non-isotropic distribution, which is a result of shielding from nearby mechanical components, as well as the shape of the cavern.

factor of a few, and that $\lesssim$1 MeV was increased by a factor of a few. Hence when we threw from the WT and assume that the spectrum is identical to that originating from the cavern, we were slightly overestimating the contribution from neutrons $\gtrsim$1 MeV and underestimating those below. Note that we only considered the flux incident on the sidewall here, in order to avoid biasing ourselves by including the low flux on the bottom lid.

When we threw neutrons from the WT, we had assumed that the angular distribution of the neutron momenta followed Lambert's cosine law to account for scattering within the cavern. From our cavern wall

Region	Sim. Flux [1/m^2/day/(2π sr)]
Top Lid	3270 ± 6
Bottom Lid	247 ± 1
E-Stem	2750 ± 60
C-Stem	2320 ± 60
Side Wall	3133 ± 3
All	2651 ± 2

Table 5.5: Incident neutron flux on the water tank, averaged over each defined region.

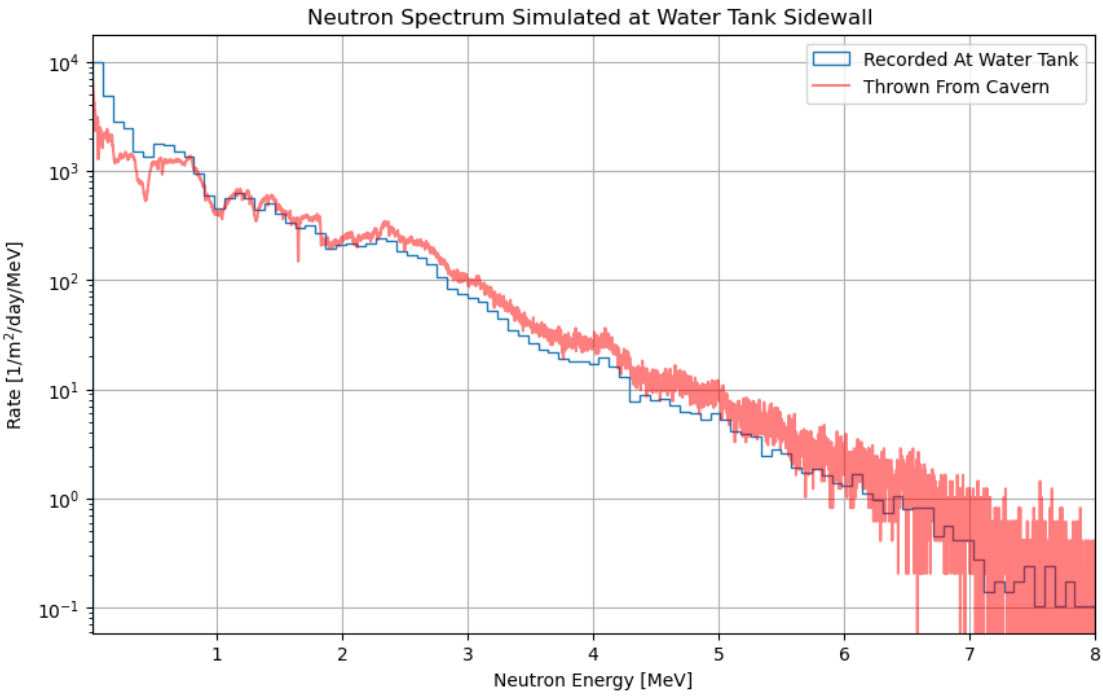
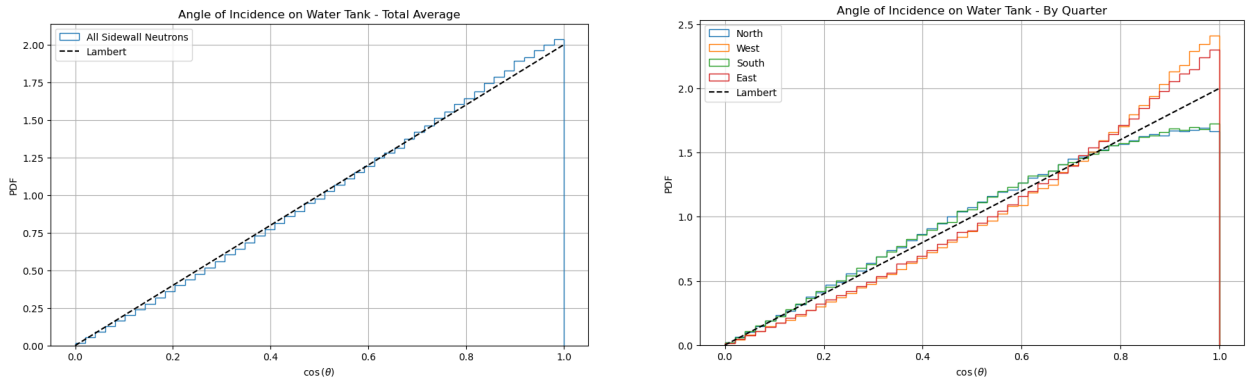


Figure 5.10: Comparison between the spectrum recorded at the WT (blue) to that thrown from the cavern wall (red). We see that the presence of the cavern moderates the spectrum above $\gtrsim 1$ MeV, and the observed spectrum exceeds the thrown one below this energy.



(a) (b)

Figure 5.11: Incident angle of neutrons hitting the sidewall of the WT when throwing from the cavern wall. Left plot shows the average over the entire sidewall, while the right plot breaks the wall into quarters. While the total average is approximately Lambertian, the sections of the shield aligned with the long axis of the cavern demonstrate a bias towards smaller incident angles.

simulation, we tested the validity of the assumption. Note we only considered neutrons incident on the sidewall here. The generated angular distribution is shown in Fig. 13, with the left plot averaging over the entire sidewall and the right breaking the sidewall into equal area quarters centered on the stems. The e-stem is positioned in the center of the "West" quarter. We see that, while the total average was decently well approximated by a Lambertian distribution, the East-West quarters were biased towards small angles of incidence and the North-South portions favor wider angles. This was consistent with our hypothesis that the shape of the cavern acts as a weak collimator in the East-West direction. The average distribution was in fact slightly biased towards small angles, indicating that more neutrons reached the tank along the East-West axis.

Revisiting the WT to IP Simulation

Combining the simulated flux we saw at the wall of the WT when throwing from the cavern wall, we utilized our transfer function to estimate how our recorded flux exiting the IP might change if we had thrown from the cavern wall. Keep in mind that this is a rough estimate, since we haven't weighted any of the energy dependence in the transfer function and we know there is some modulation in the shape of the spectrum which occurs between the cavern and the WT. We can write the estimate as

$$\vec{R}_{poly} = T\vec{R}_{tank} \tag{5.3}$$

	Flux $[1/m^2/day/(2\pi\ sr)\times10^{-3}]$	
Region	Simulated	Scaled
Top Lid	6.9	9.2
Bottom Lid	4.6	5.9
E-Stem	980	1350
C-Stem	358	419
Side Wall (Ex. Stems)	8.3	11
Total	16.4	21.7

Table 5.6: Rate of fast neutrons through the IP from our simulation throwing from the WT compared to scaling the flux observed outside the WT when throwing from cavern wall using our transfer function.

where $\vec{R}$ is the vector of rates[1] at each region of the shield layer, and T is our transfer function. The result of this scaling is summarized in Table 5.6. The total estimated fast flux increased by $\sim 30\%$ when scaling to the flux from the cavern. To first order, this was consistent with the increase in the flux observed around the stems when throwing from the cavern wall.

5.6.3 Neutrons Thrown From E-Stem

A common theme through the results we have presented thus far has been the importance of the stems. This fact was very apparent when attempting to estimate the background from the cavern wall simulation. The striplines had been omitted from the e-stem the geometry, and this resulted in a clear increase in the NR background compared to previous estimates. Conversely, this suggests that providing the stems with additional shielding could result in a substantial reduction of the NR background. To test this hypothesis, we considered placing a HDPE "Pipe Shield" (PS) surrounding the e-stem located between the WT and the ETank (49 cm thick, 50 cm radius) and simulated the effect this has on the flux near the e-stem.

In this study, we were only interested in the neutrons near the e-stem, and it was chosen to generate the neutrons in a hemisphere with radius of 2m centered about the middle of the e-stem on the outside of the WT. This was large enough to enclose the ETank, along with the PS. Neutrons incident on the outside of the WT were recorded. We also attempted to model the effects of the transfer function between the cavern wall and WT when throwing the neutrons, so that we could make a direct comparison between the flux observed in this study to that in the cavern wall simulation. We accounted for this using two steps which differ than what we did when throwing from the WT

[1] $flux \times area$

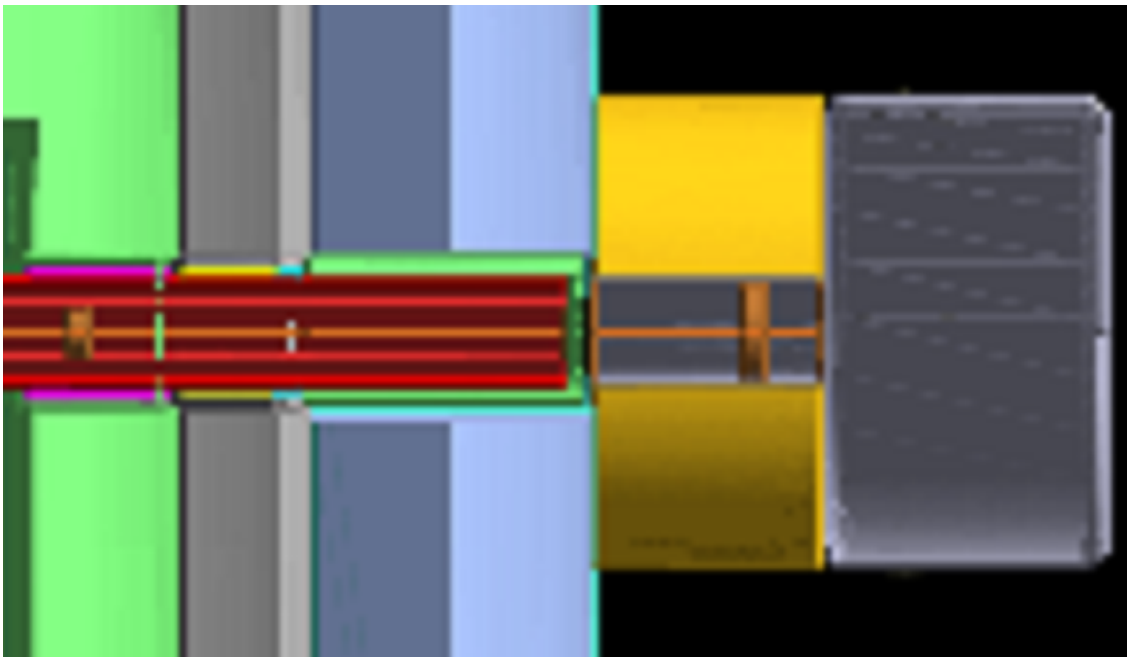


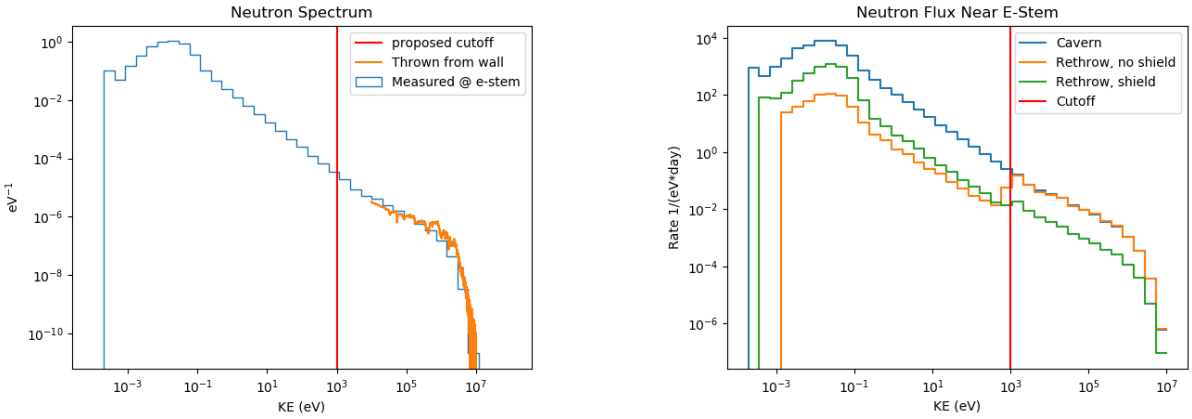
Figure 5.12: Cross section of the shield near the E-tank (gray), including the hypothetical addition of the PS (yellow). The PS rests around the E-Stem between the WT and the E-Tank. Copper and kapton striplines occupy the center of the E-Stem.

1. We threw neutrons according to the observed spectrum near the e-stem when thrown from the cavern wall, rather than the original cavern wall spectrum. We described this spectrum down to 1 keV, as opposed to the 10 keV cutoff on the original spectrum.

2. We defined our normalization when throwing from the hemisphere such that we observed the same number of neutrons when throwing from the hemisphere without the PS as we did when throwing from the cavern wall. This is opposed to assuming the same normalization when throwing from the cavern wall to the hemisphere.

In Fig. 15, we compare the spectra observed near the e-stem when throwing from the cavern to that throwing from the hemisphere with and without the PS. We see that the spectrum from the hemisphere without the PS gives good agreement with the cavern both in rate and shape above our energy cutoff. Then, when we added the PS and threw from the hemisphere we see that the rate above our cutoff was reduced by an order of magnitude. Given our simple scaling argument in the previous section, we naively expected this to reduce the flux at the IP by a similar amount.

It is also interesting that below our cutoff, we saw an increase in the rate with the PS present compared



(a) (b)

Figure 5.13: Spectral shape observed near the e-stem when thrown from the cavern wall, as compared to the spectrum thrown from the cavern wall. We throw neutrons originating from the hemisphere sampling the blue observed spectrum above the red line. Spectral rate when throwing from the hemisphere with (green) and without (orange) the addition of the PS. The orange spectrum agrees well with the blue above the cutoff, and the addition of the PS reduces the rate in this regime by an order of magnitude.

to with it absent. This isn't too surprising, considering that additional shielding provides more opportunities for the neutrons to thermalize. In any case, the observed increase in the low energy neutron rate was small compared to that generated from the cavern, and even there we didn't directly generated low energy neutrons.

5.7 Discussion

Through our simulation of the radiogenic cavern neutrons, we arrived at a reasonable estimate of the background. Notably, these backgrounds are expected to contribute $\lesssim 20\%$ of the NR background in iZips expected from CEνNS. These simulations were performed by throwing the neutrons from the outside of the water tank, which requires us to make some simplifying assumptions about the neutron transport between the cavern wall and the shield. A rough, energy independent estimate of the transfer function from the cavern to the water tank and from the water tank to the inner poly indicate that these assumptions can raise the rate on the order of $\sim 10\%$.

This study highlighted the difficulty of estimating the transfer function by stitching together spectra from different regions on a given volume. Since the distribution of particle fluxes through the shield are neither homogeneous or isotropic, it requires a prohibitive amount of statistics to properly characterize the

correlations in the high-dimensional parameter space which arises. Given that the desired purpose of this stitching approach is to reduce computation time, there is a diminishing benefit to accumulating the statistics to robustly carry out such an analysis. Alternatives such as importance biasing represent a promising route, but it has yet to be clearly demonstrated that such methods can be implemented in a way which preserves the timing resolution of our detectors which is critical for vetoing backgrounds. In any case, it is difficult to validate any "shortcut method" without having a full statistics end-to-end simulation for comparison. At the end of the day, it is hard to beat the brute force approach of simulating many primaries in a trusted geometry. The geometry of the SNOLAB cavern surrounding SuperCDMS is not quite at the level of detail required to do this for our experiment. The drift ceiling does not have a uniform height as currently modeled, and other nearby mechanical infrastructure not included in the geometry could produce similar shadowing effects as observed 5.9. Active work is underway to update our geometric model such that a large statistics cavern to detector simulation can be performed.

Overall, the cavern neutron background is not negligible, but compared to CEνNS as well as other sources of material contamination in the shield, we expect it to be sub-dominant in the overall background budget. In future upgrades to SuperCDMS, it is conceivable that the dominant sources material contamination can be mitigated and the CEνNS background may be modeled sufficiently to subtract the background, in which case the cavern neutron background will become more important. It is therefore worth considering the addition of an external neutron shield surrounding the stems to reduce the external background in this scenario.

Part III

Searching for Light Dark Matter

Chapter 6

SuperCDMS Small Detector Program

We want to search for DM masses below 1 GeV. This is motivated by the many viable models at this mass scale. We (SuperCDMS), along with the community of DM hunters at large, are looking to adopt a philosophy of "Delve Deep, Search Wide". SuperCDMS already has a plan to delve deep in the $1 - 10$ GeV mass range at SNOLAB by performing a large exposure, low background search using low (~ 100 eV) threshold detectors. While we work to implement this goal, we also want to think about searching wide. We can do this by developing new detectors which leverage existing SuperCDMS technology to achieve lower thresholds, further expanding our reach to lighter masses. In addition to providing this thrust, low threshold prototype detectors further support the program by providing tools with a unique capability to characterize backgrounds in the energy range critical for LDM searches.

6.1 LDM and Small Detectors

Consider that we want to use a TES to measure a phonon signal induced by DM interactions, and that we want to make our resolution as fine as possible in order to attain sensitivity to LDM. Under ideal operating conditions, where we've mitigated any external noise input to the sensor, the resolution of a TES will be limited by thermal fluctuation noise (TFN). If this is the case, and the temperature of the thermal bath is sufficiently small compared to the critical temperature T_c of the TES, then the energy resolution of the detector scales as [97, 98]

$$\sigma_E \sim \frac{T_c^3}{\epsilon_{net}} \sqrt{k_B \Sigma V_{TES} \tau_{BW}} \tag{6.1}$$

where ϵ_{net} is the net energy efficiency, Σ is a material dependent characterization of the thermal conductance between the TES and the thermal bath, V_{TES} is the volume of the TES and τ_{BW} is a time representative of the signal bandwidth. This scaling tells us that, under our idealizing conditions, we only have a handful of ways to improve the resolution:

- Improve the net efficiency
- Make T_c lower
- Make the TES smaller
- Make the detector faster

Though it's clear from the scaling relation Eq. 6.1 that the strongest dependence is on T_c, reducing the critical temperature presents a technical challenge. For one, we must keep T_c reasonably higher than the bath temperature, or we risk not being able to operate the detector in transition. Tungsten occurs in two phases, an $\alpha-$phase with $T_c \approx 12$ mK and a $\beta-$phase with $T_c \approx 5$ K. Real thin-films of tungsten exist in some combination of these phases, with intermediate T_c's. Typical values observed in SuperCDMS detectors are $T_c \sim 50$ mK. The particular T_c value realized have been shown to change with the concentration of magnetic impurities [99] and composition of the sputtering gas [100, 101]. Putting aside the difficulty tuning T_c, let us turn our attention to parameters which we have better handles on, namely the TES volume and net efficiency.

The dependence of V_{TES} of the resolution is clear. The TES is really just a very sensitive thermometer. When some heat is deposited in the TES, the temperature rises and we can measure a corresponding change in the resistance. If we reduce the heat capacity, in this case just by reducing the total amount of material in the TES, we get a larger change in temperature for a given heat input. What's more, reducing the TES volume is easy! Just put fewer sensors on our detector, and we're done.

The optimization of the net efficiency is much more subtle, since there are lots of ways to lose energy. It is useful to be explicit regarding what we mean by the efficiency. In its simplest conception, the efficiency is

just the ratio of the total recoil energy deposited in the detector and the energy absorbed by the TES.

$$\epsilon_{net} \equiv E_{abs}/E_{dep} \tag{6.2}$$

Given that the electrothermal response of the TES is sufficiently fast, the absorbed energy by the TES is well approximated by the energy removed by electrothermal feedback (see appendix A), which can be determined by directly integrating the current trace.

Taking a look at our previous strategy of minimizing the TES volume, this seems in direct opposition to the goal of increasing the energy efficiency. Naively, we expect the energy efficiency to be proportional to the fraction of detector surface covered by a sensor. Considering that the resolution scales stronger with the efficiency than the volume, for a fixed TES thickness we should expect the resolution to degrade when we reduce the TES volume! Our ability to subvert this relies on the fact that our sensor is not merely a TES, but rather a QET. This means that there are actually several steps which occur between the production of phonons in the detector bulk and the absorption of this energy by the TES.

Approximate Life Cycle of Energy Which Appears in our Signal:

1. The phonon hits an aluminum fin.
2. The incident phonon creates a quasiparticle by breaking a Cooper pair in the aluminum.
3. The quasiparticle diffuses across the fin.
4. The quasiparticle is trapped by the tungsten TES.

Each step in this list has a sub-unity probability to succeed, and the product of all steps is the net efficiency [98]

$$\epsilon_{net} = \epsilon_1 \times \epsilon_2 \times \epsilon_3 \times \epsilon_4 \tag{6.3}$$

Our argument that reducing the number of sensors will reduce the net efficiency is only true in so far as it reduces ϵ_1. But from this we see that we still maintain or even improve the net efficiency if we compensate by increasing $\epsilon_{2,3,4}$. The optimization of $\epsilon_{3,4}$ depends on the details of the QET geometry, and can be quite subtle. We leave the discussion of these parameters beyond the scope of this thesis, but refer the reader to Refs. [97, 102, 103] for details.

For now, we focus on a simple method for optimizing ϵ_2. In order for a phonon to break a Cooper

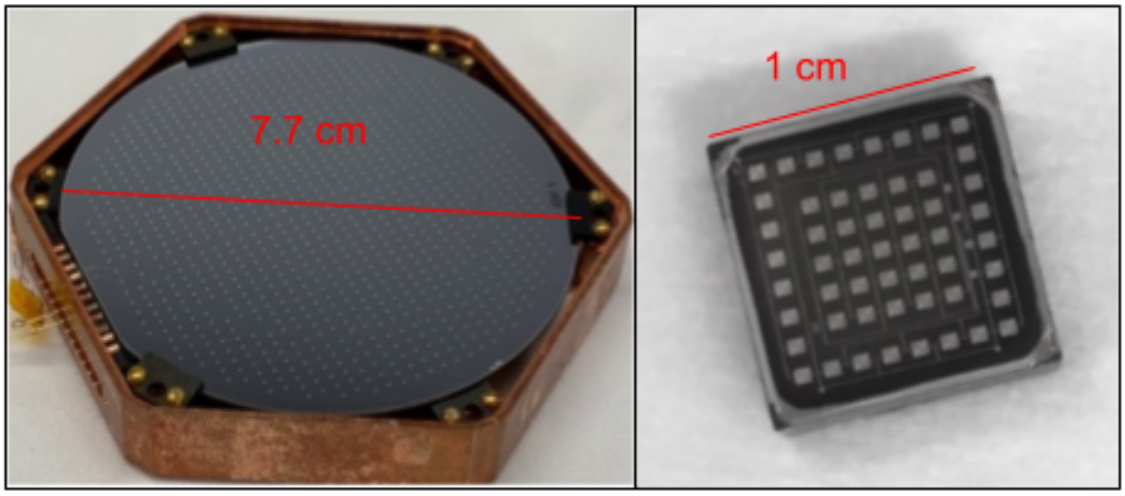


Figure 6.1: Photos of the CPD detector (left) and HVeV (right). Note the difference in surface area and density of sensor pads. CPD is much larger in area, has a less dense sensor layout and is designed to operate under no applied external voltage.

pair in the aluminum fin, its energy must be at least twice the superconducting band gap in aluminum $\omega_{ph} > 2\Delta_{Al} \approx 0.7$ meV. Shortly after the particle interaction in the bulk, a large fraction of the generated phonon population satisfies this condition. However, as the phonons propagate they have the opportunity to down-convert and fall below the energy threshold required to propagate through the aluminum fin. It has been demonstrated that the down-conversion process is dominated by scattering of phonons off the detector surface [104]. This suggests that we can minimize the population which falls below the relevant threshold, and hence maximize ϵ_2, by minimizing the uninstrumented surface area of the detector. This can be achieved without changing the sensor layout by thinning the detector.

6.2 The CPD and HVeV Programs

SuperCDMS has thus far deployed two detector technologies beyond its flagship iZip and HV designs in its quest to search for DM at sub-GeV masses. These detector programs are dubbed the Cryogenic PhotoDetector (CPD) and the High Voltage eV-resolution detector (HVeV). Here we discuss how these detectors inherit from and iterate upon existing SuperCDMS technology to accomplish sensitivity to lower mass DM.

Both detectors are built on a thin silicon substrate and make use of the same QET phonon readout inhered from SuperCDMS as their signal channel. Furthermore, both use a fairly small TES volume in order to maximize their sensitivity. They do this through different geometries however - with HVeV using a high

sensor coverage fraction over a small surface area and CPD a low coverage over a large area (see Fig. 6.1). Another stark difference is their operating conditions. The HVeV is designed to be operated under a ~ 100 V voltage bias, while the CPD does not incorporate any bias rails and operates at 0 V.

Both detectors have demonstrated baseline resolutions $\sigma_E \sim 1$ eV, and are hence both excellent candidates for LDM searches [105, 106]. Their particular differences do make them better suited for different models of LDM. At 0 V, CPD measures pure recoil energy and is best for targeting NRDM. The HVeV on the other hand makes use of the NTL effect to amplify the charge signal, which makes it best suited to search for ERDM (though its sensitivity at 0 V should not be discounted [98]). This suggests a synergy between the detectors which make them complementary to one another - they can target LDM in a similar mass range but through different channels. This synergy also helps each detector cover the other's blind spots. Being a large area detector makes CPD somewhat sensitive to position dependent effects which can degrade its resolution [107], but also gives it a large coverage ideal for operating as a low energy background veto. The HVeV on the other hand demonstrates less position dependence [108], but the large NTL amplification can introduce charge trapping/leakage effects which contribute to its background. It is also interesting to note that the low energy excess introduced in Chapter 4.3.7 has been observed in both detectors. Leveraging the synergy between the two has the gives us a diverse set of tools to characterize this excess. For example, simultaneous operation of sensitive devices at high voltage and zero voltage would immediately allow us to characterize what portion of the observed excess is non-ionizing, and the time dependence of the ionizing and non-ionizing populations could be compared.

The remainder of this thesis will discuss results obtained from the CPD program. Though we will not consider the implications of these results for the HVeV in detail, it is useful for the reader to keep in the back of their mind that these detectors are "siblings" in some sense, and there is a huge potential for the programs to support each other. Furthermore, they can also support the broader SuperCDMS by exploring the requirements to gain sensitivity to LDM as well as characterize backgrounds in the relevant energy range. We also note that the CPD program has expanded beyond SuperCDMS and is a central pillar to the SPICE/HeRALD experiments currently being developed to search for LDM using a wide array of diverse targets [109]. The two collaborations are presently developing expertise in conjunction.

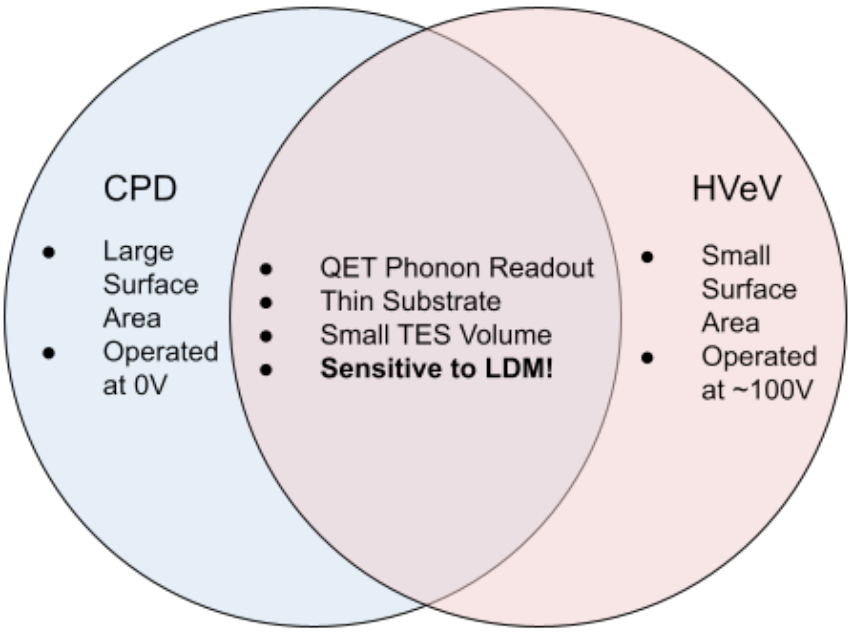


Figure 6.2: Venn diagram of a qualitative comparison between the CPD and HVeV detectors.

Chapter 7

CPD Calibration and Simulation Studies

This chapter describes simulations performed in to understand calibration data taken with a ^{55}Fe source to support a DM search using CPD at the Cryogenic Underground TEst facility (CUTE). The motivation for this DM search was to expand on the success of that performed with CPD at SLAC [105], but now repeat it in a low background environment. It was observed that the calibration introduced an unexpectedly large background, which the studies presented here were performed in order to characterize. The high background prevented use of the calibration source *in situ*, and additional sources of noise resulted in a baseline resolution worse than what was observed at SLAC. With these considerations, a full-fledged DM search was not performed, though a partial analysis is documented in Ref. [110][1].

7.1 The CUTE Facility

Though an independent test facility, CUTE has close ties to SuperCDMS. Its original design requirements were specified such that it can accommodate the testing of a SuperCDMS detector tower, and both facilities are housed in the same cavern drift at SNOLAB. As such, CUTE features a cryostat capable of reaching ~ 12 mK base temperature, which is surrounded by ~ 10 cm of low activity to act as a gamma shield and

[1]Note that the name applied to the detector CPD has changed over time. The name used in the reference "PD2" refers to the same detector as "CPD" here.

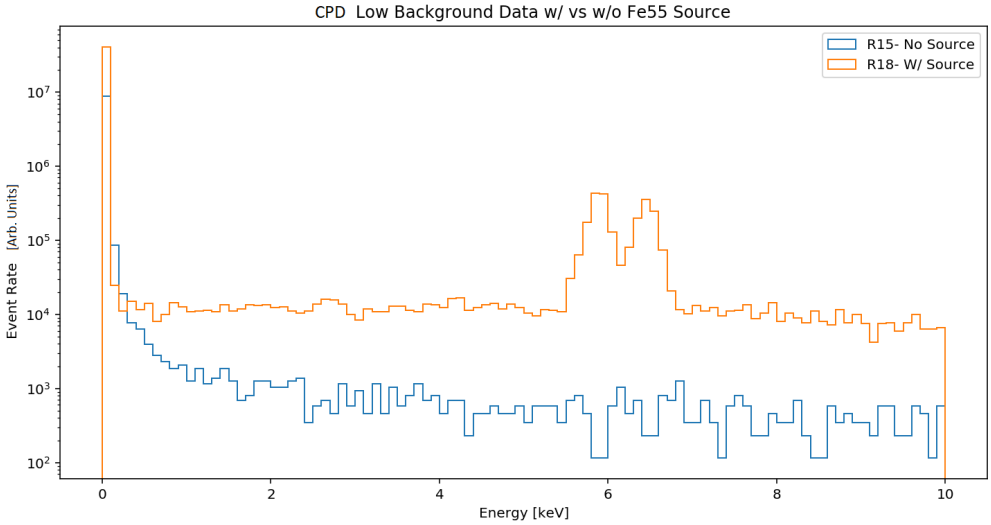


Figure 7.1: Comparison of low background data on CPD at CUTE between runs which include (orange) and exclude (blue) the ^{55}Fe source from the stack. In addition to the expected peaks at 5.9 and 6.5 keV, we was the background increases by an order of magnitude when the source was present.

then a $\sim$ 1.5 m thick water tank to block the cavern neutrons [111].

7.2 Simulation of Low Energy Backgrounds from ^{55}Fe

During the attempted DM search, a ^{55}Fe source was deployed in the detector payload to act as a low energy calibration for CPD. The source produces 5.9 keV (K_α) and 6.5 keV (K_β) characteristic x-rays [112]. Aluminum foil is placed between the source and detector in order to produce an additional line at 1.5 keV from aluminum fluorescence. The source is held in a copper detector housing which can be included in the stack, so it is deployed on a run-by-run basis. In runs which include the source we can easily see the expected x-ray peaks, but they are accompanied by a dramatic increase in the background, as shown in Fig. 7.1. This background makes the observation of the aluminum fluorescence peak very difficult, and precludes using the source for in-situ calibration during a dark matter search. This study aims to understand the observed increase in background by simulating the photon spectrum from ^{55}Fe decay, as well as radiogenic contaminants in the source holder.

When comparing the background between runs, we'll restrict ourselves to the region 2-3.7 keV. This is within the low energy range we're interested in calibrating, and avoids spectral features such as aluminum fluorescence peak. Comparing the rate in this region shows that introducing the source increases the back-

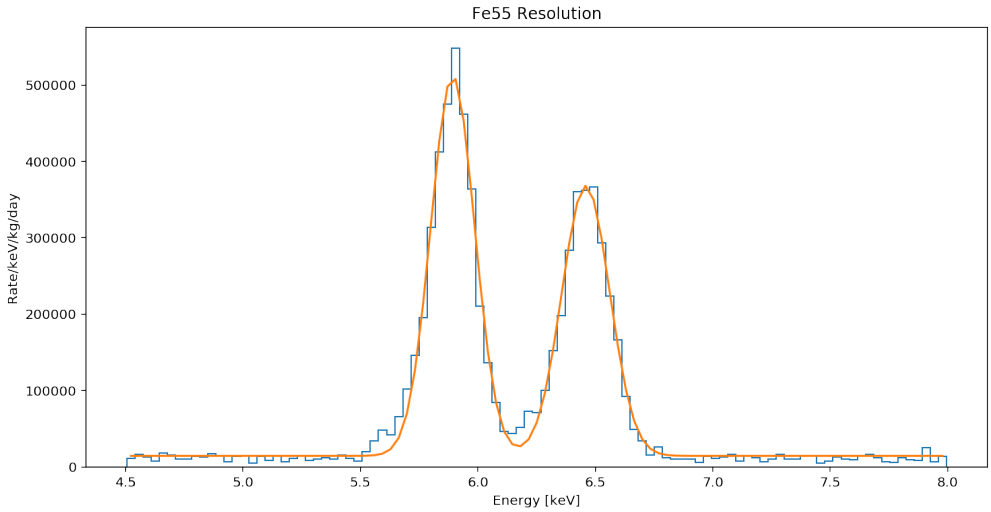


Figure 7.2: Observed calibration peaks in CPD at CUTE (blue). We fit two Gaussians with a flat background (orange) to this data. The peak locations agree well with the expected energies, and the shape is well approximated by the Gaussians.

ground by a factor of $\approx \times 15$. The original hypothesis raised when presented with this observation was that the increase in background was simply due to Compton scatters of the calibration x-rays, but this argument can be easily dismissed by noting that the scattering cross section of a 6 keV x-ray in silicon is $< 1\%$ than that for absorption [113]. The next consideration was then that high levels of radioactive contamination in the source holder, such as ^{60}Co in the copper, could explain the background. A simulation was then undertaken to resolve this question.

The Geant4 simulation we used to model the background does not include the macroscopic physics which gives our detectors finite resolution. In order to make a direct comparison between the simulation and data, we included a model of the energy dependent resolution, based on the observed baseline resolution and that of the ^{55}Fe x-ray peaks. The baseline resolution was estimated from the width of a Gaussian fitted to the zero-delay OF amplitude of in run randoms. For the x-rays, we fit two Gaussians with a flat background to the calibrated data.

The source itself is made of a thin layer of ^{55}Fe electroplated onto a small stainless steel disk. The disk sits on a copper bed which is mounted to the bottom lid of the housing. The bed is covered with a layer of aluminum foil, and then sealed with a copper lid. Finally another copper lid is placed over the entire housing. Small pinholes in each lid allow a column of x-rays to escape the encapsulation. Additional foil is placed over the outer pinhole in order to produce more fluorescence. The total thickness of foil used in R18

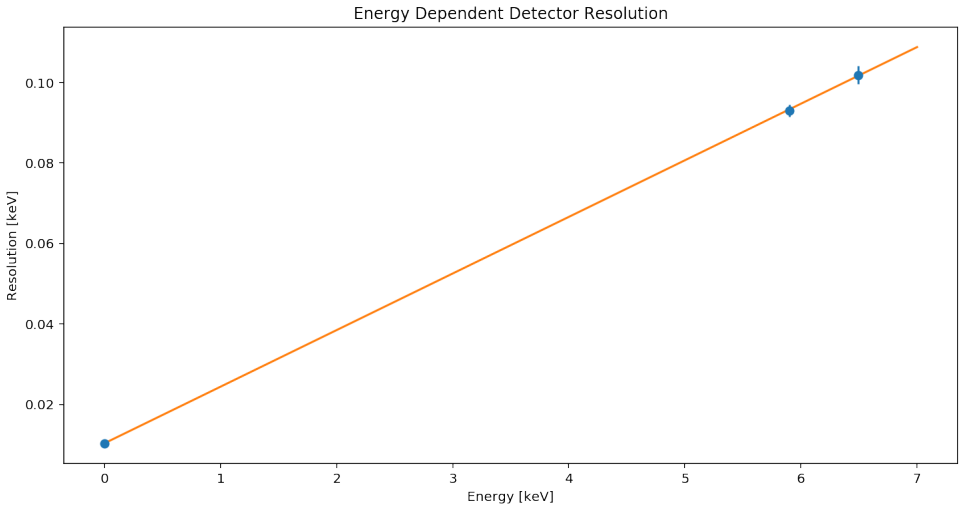


Figure 7.3: Observed resolution for CPD at CUTE, as a function of energy. The baseline resolution is estimated from the amplitude of random triggers, and the higher energy x-rays from our fit in 7.2. The energy dependence is well approximated by a linear function.

Line Energy [keV]	Photons/100 Disintegrations
0.556	0.175
0.567	0.175
0.685	0.175
5.888	8.45
5.899	16.57
6.490	1.7
6.535	1.7

Table 7.1: Characteristic ^{55}Fe x-rays

is nominally ~ 224 μm.

^{55}Fe decays to ^{55}Mn via the electron capture process $p + e^- \rightarrow n + \nu_e$ with a half-life 2.74 years. When the electron is captured it leaves a gap in its original shell which is subsequently filled by an electron from a higher level. This transition results in the emission of either an Auger electron ($\sim$ 5 keV), x-ray or both (Radiative Auger Effect). A 5 keV electron is attenuated by the aluminum foil, so we didn't expect any to escape the encapsulation and omit Auger electrons from our model of the source in this study. Relative intensities of the emitted characteristic x-rays are given in Table 7.1 [112].

In addition to the characteristic x-rays, the decay of ^{55}Fe can also emit bremsstrahlung photons as the electron undergoing capture interacts with the nuclear electric field. This process is called inner bremsstrahlung (IB) to distinguish it from "outer bremsstrahlung" which is produced by free electrons stopping in some

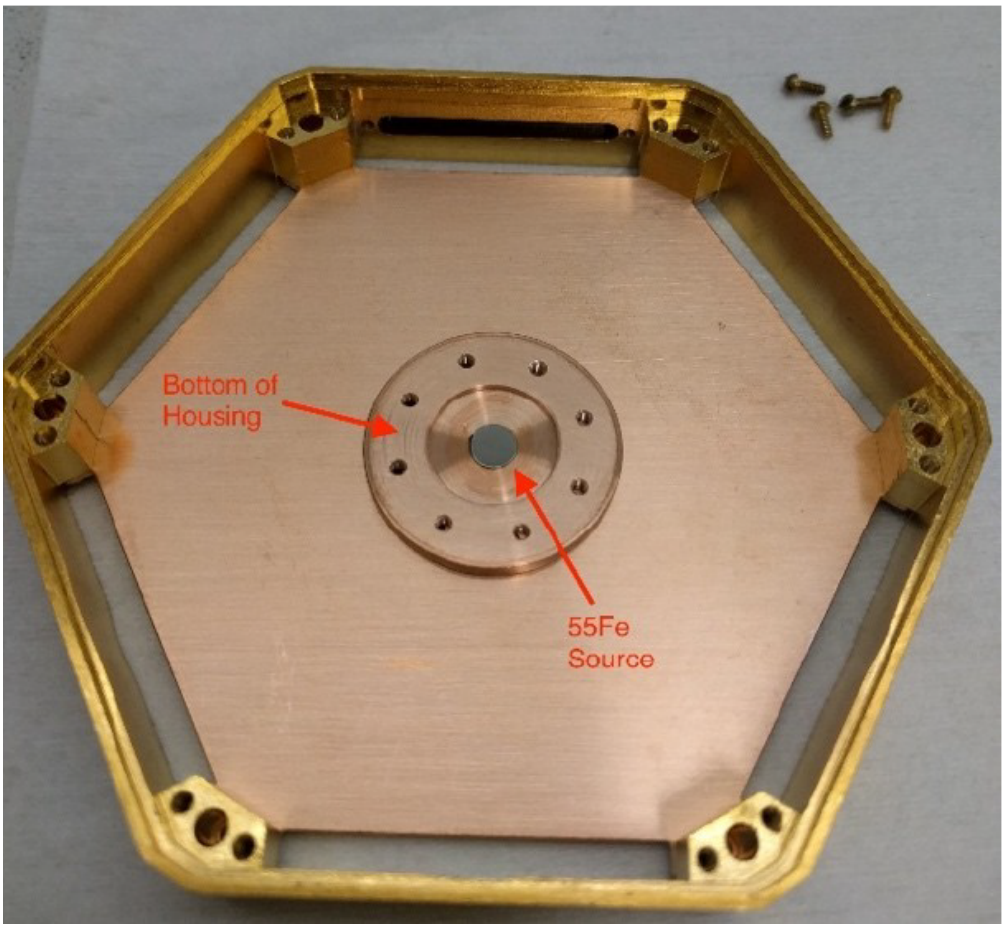


Figure 7.4: Photo of source holder used. The gray active source button is visible in the center of the holder. When closing the encapsulation, a layer of aluminum foil is added over the source button, and then a small copper lid with a collimating pinhole is fixed over copper surrounding the button. Then a larger hexagonal with another collimating pinhole is placed over the entire housing.

Line	Energy (keV)	ϵ	λ (μm)
K_α	5.90	0.25	30.5
K_β	6.51	0.034	40.6

Table 7.2: Characteristic parameters of calibration lines

foreign material. The resulting spectrum is a continuum up to 231 keV (the Q-value for the decay). The intensity between 35-231 keV is 3.24×10^{-5} per K-shell capture decay [114]. Though IB is suppressed compared to the characteristic x-rays, any photons produced above ∼20 keV can penetrate the thin (2mm) layer of copper [113] between the source and detector. Hence we included the IB spectrum in our simulation, since we expected a rate comparable to the characteristic x-rays.

The expected rate from a given line is described as follows

$$\Gamma_i = Ak\epsilon_i e^{-L/\lambda_i} \tag{7.1}$$

where A is the total source activity, k is the "geometric efficiency" - or probability for an x-ray to pass through the collimator in the absence of any foil, ϵ_i is the branching fraction of the line, L is the foil thickness and λi is the attenuation length of the line of interest in Al.

7.2.1 Geometric Efficiency

We estimated the efficiency of the collimator by considering the radius of each pinhole and the distance between the source and the the second hole. The collimator design is sketched in Fig. 7.5, and the relevant dimensions are given in Table 7.3. The first collimator was essentially on top of the source, and we approximated its effect as simply reducing the size of the active source. Then we determined the efficiency of the second collimator by considering the solid angle subtended between it and the center of the source.

$$k = \frac{1}{2}\left(\frac{r_1}{r_s}\right)^2\left(1 - \cos\left(\theta\right)\right) \tag{7.2}$$

where

$$\theta \equiv \arctan\left(r_2/d\right) \tag{7.3}$$

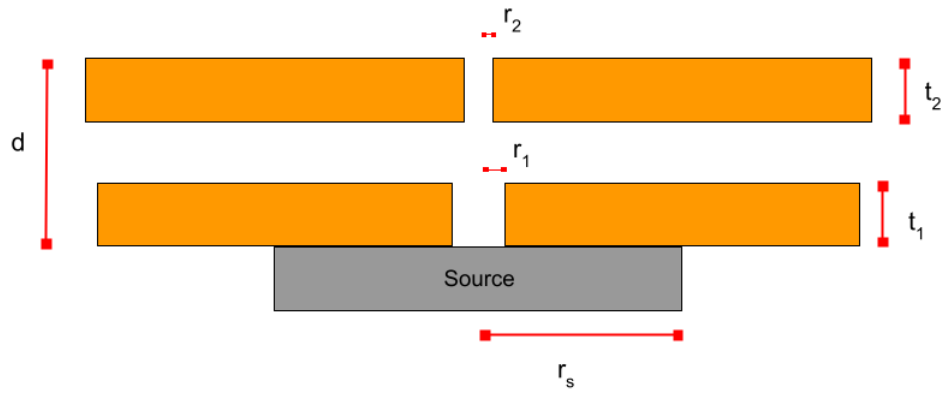


Figure 7.5: Cartoon showing the collimator configuration. Between the radioactive source (radius r_s) and the detector, we have two collimating lids with pinhole radius r_i and thickness t_i. The distance between the source and the end of the second collimator is described by d. Our goal is to estimate the fraction of x-rays we expect to escape this encapsulation.

For a thin collimator, we applied the small angle approximation, and find

$$k \approx \frac{1}{4}\left(\frac{r_1}{r_s}\right)^2\left(\frac{r_2}{d}\right)^2 \tag{7.4}$$

Hence the resulting efficiency follows from a consideration of the solid angle subtended by the collimator. For the nominal dimensions of the collimator, we found $k = 1.85 \times 10^{-4}$. However these were only the dimensions specified from the design, and the actual lengths were not measured. Hence there was a relatively large systematic uncertainty on the geometric efficiency.

This approximation did not account for the finite thickness of the collimator lids. In order to confirm the validity of this, we performed a toy Monte Carlo simulation. This simulation assumed the x-rays are generated uniformly and emitted isotropically across the surface of the source, and that they then travel in straight lines, stopping when they touch the copper. We then simply counted how many of the lines we drew which didn't intersect with any copper surface boundary. From 10^8 trials of the toy MC, we calculated a geometric efficiency of $(1.78 \pm 0.01) \times 10^{-4}$, which agreed with the above calculation within 5%. Hence including the finite thickness of the lids slightly decreased our geometric efficiency, which is a reasonable conclusion.

Parameter	Length [mm]
r_s	2.35
r_1	0.75
r_2	0.5
t_1	1
t_2	1
d	5.86

Table 7.3: Dimensions of source holder. Parameters are defined in Fig. 7.5. Note that these are the values specified in the design rather than the result of an actual measurement.

7.2.2 Foil Thickness

The expected rate depends strongly on the thickness of the aluminum foil, which is nominally 224 μm. This is a parameter we needed to specify in the simulation, and we determined the value from the data. Looking at the ratio of the expected K_α and K_β rates factors out the dependence on the source activity and geometric efficiency, giving us a function of the foil thickness

$$\frac{\Gamma_\alpha}{\Gamma_\beta} = \frac{\epsilon_\alpha}{\epsilon_\beta} \exp\left(-L(\frac{1}{\lambda_\alpha} - \frac{1}{\lambda_\beta})\right) \tag{7.5}$$

From our fit to the observed peaks shown in 7.2, we found the ratio of rates

$$\left.\frac{\Gamma_\alpha}{\Gamma_\beta}\right|_{obs} = 1.28 \pm 0.03 \tag{7.6}$$

from which we infered a foil thickness

$$L_{obs} = 215 \pm 3 \ \mu\text{m} \tag{7.7}$$

Recall that the purpose of placing the aluminum foil over the collimator was to stimulate the production of 1.5 keV fluorescence x-rays. The ~ 200 μm thickness of the foil was rather large compared to the ~ 10 μm attenuation length of x-rays in the foil. Considering this, it is not too surprising that the we couldn't resolve the desired calibration line in our data. There are two competing processes which determined the amount of observable fluorescence produced in the foil as we changed the thickness. Making the foil thicker gave more opportunities for the fluorescence to be generated, but at the same time made it more likely that both the incident and emitted radiation were attenuated before they escape. Naively, if the attenuation length of

both the incident and emitted x-rays were of a similar scale, we might guess that making the foil thickness equal to the average of the two lengths would balance the competing processes and maximize the fluorescence output.

$$L_{max} \sim \frac{\lambda_0 + \lambda_f}{2} \tag{7.8}$$

Where L is the foil thickness, and λ_0, λ_f are the attenuation lengths of the incident and emitted x-rays respectively.

Let us evaluate how the fluorescence which exits the foil varies as we changed the thickness in order to estimate how we could maximize the output. We assumed that both the incident and fluorescence radiation is monochromatic. The target is an infinite plane of thickness L. In the x-ray regime, photons interact with the material primarily via photoabsorption, and we have a simple correspondence between the photoabsorption cross section σ and attenuation length of light in the material λ

$$\lambda_i = \frac{1}{n\sigma_i} \tag{7.9}$$

where n is the target density and i indexes a particular energy. We considered a thin beam of photons incident normal to the target surface with energy E_0 and attenuation length λ_0. The intensity of the exciting light throughout the target then follows Beer's law

$$I_e(x) = I_0 e^{-x/\lambda_0} \tag{7.10}$$

Then we expect to produce a fluorescence intensity between x and $x + dx$ of

$$dI_f = I_e(x)\frac{W}{\lambda_0}dx = \frac{W}{\lambda_0} I_0 e^{-x/\lambda_0} dx \tag{7.11}$$

where W is the probability for the fluorescence line of interest to be emitted following the excitation. This gives us the amount of fluorescence produced, but obviously it's only possible to observe a fraction of this as the radiation is attenuated by the target. We assumed that the radiation is emitted in the same direction as incident x-ray. This is clearly not physically motivated and will overestimate the number of x-rays which penetrate the foil, but it is an assumption which greatly simplifies the problem and produces a correct order

of magnitude estimate of the optimal thickness.

If we assumed that the fluorescence photons are emitted parallel to the exciting photons, then a photon produced at depth x from the incident surface must travel a distance $L - x$ to escape. The probability for the photon to survive this distance also follows Beer's law. Then we then describe the expected intensity we expected to escape

$$dI_{out} = dI_f e^{-(L-x)/\lambda_f} \tag{7.12}$$

$$= I_0 \frac{W}{\lambda_0} e^{-x/\lambda_0} e^{-(L-x)/\lambda_f} \tag{7.13}$$

and we get the total intensity by integrating over the thickness of the target

$$I_{out}(L) = \int_0^L \frac{dI_{out}}{dx} dx \tag{7.14}$$

$$= I_0 W \frac{1}{1 - \frac{\lambda_0}{\lambda_f}} \left(e^{-L/\lambda_f} - e^{-L/\lambda_0} \right) \tag{7.15}$$

And finally we take the derivative with respect to L and solve for the optimal thickness. We find

$$L_{max} = \ln \left(\frac{\lambda_0}{\lambda_f} \right) \frac{\lambda_0 \lambda_f}{\lambda_0 - \lambda_f} \tag{7.16}$$

If we applied the more realistic assumption of isotropic emission of the fluorescence, we would find this slightly overestimates the optimal thickness.

Let us approximate the incident ^{55}Fe x-rays as having a 6 keV energy, and the emitted aluminum fluorescence line is 1.5 keV. We then have $\lambda_0 \approx 32$ μm and $\lambda_f \approx 9$ μm - which gives $L_{max} \approx 16$ μm. Compare this to our naive estimate where we took the optimal thickness as the average of the attenuation lengths and we would find $L_{max} \sim 20$ μm.

Clearly, the foil thickness used to calibrate CPD at CUTE is about an order of magnitude thicker than the optimal thickness. Though this may seem to be an oversight, there is another practical consideration we haven't accounted for which can make this a reasonable choice. If our goal is to use the ^{55}Fe source to perform an *in situ* calibration for a DM search, care should be taken such that the rate of calibration events doesn't eat into a significant amount of experimental exposure. With the actual source configuration used, CPD sees a ^{55}Fe x-rays at a rate of ~ 0.03 Hz. By decreasing the foil thickness to our optimal estimate, we would expect this rate to increase to ~ 25 Hz. The DM search utilized a 50 ms trace length, which means that we

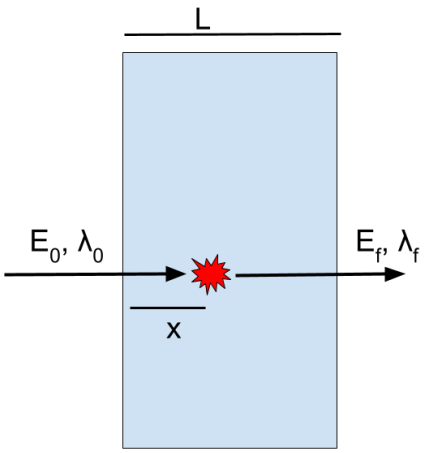


Figure 7.6: Cartoon of fluorescence produced by x-rays incident on a thin film. An incident x-ray with attenuation length λ_0 penetrates depth x into a film of thickness L. Some fluorescence is produced with a new attenuation length λ_f, which has some probability to escape the foil. We model both the incident and emitted radiation as normal to surface of the foil. This overestimates the total amount of fluorescence which escapes, but provides a reasonable estimate for the optimal thickness.

expected pileup from the source to occur in $\sim 0.15\%$ of recorded traces in the actual configuration and in $\sim 70\%$ for the "optimal" configuration. Certainly not ideal if the goal is an exposure limited experiment.

7.3 Source Simulation

7.3.1 Simulation Geometry

The geometric model used in our simulation is shown in Fig. 7.7, with the source and CPD detector explicitly labeled. There were a number of details included in the model which are not strictly relevant for our goal of understanding the background introduced by the source, but are left in the model for completeness. Namely, several other detectors, shown in red, were present in the payload for this run but we don't analyze them in detail here. We also note that our model was not perfect. For example, the source in our geometry was much closer to CPD than in reality. This would have affected the size of the spread of events across the surface of CPD. While this would have influenced the reconstructed energy if we had performed a complete detector Monte Carlo which simulated the propagation of phonons and hence introduced position dependence. we did't expect it to have any effect in this study where we just summed all the recorded energy depositions simulated in the detector.

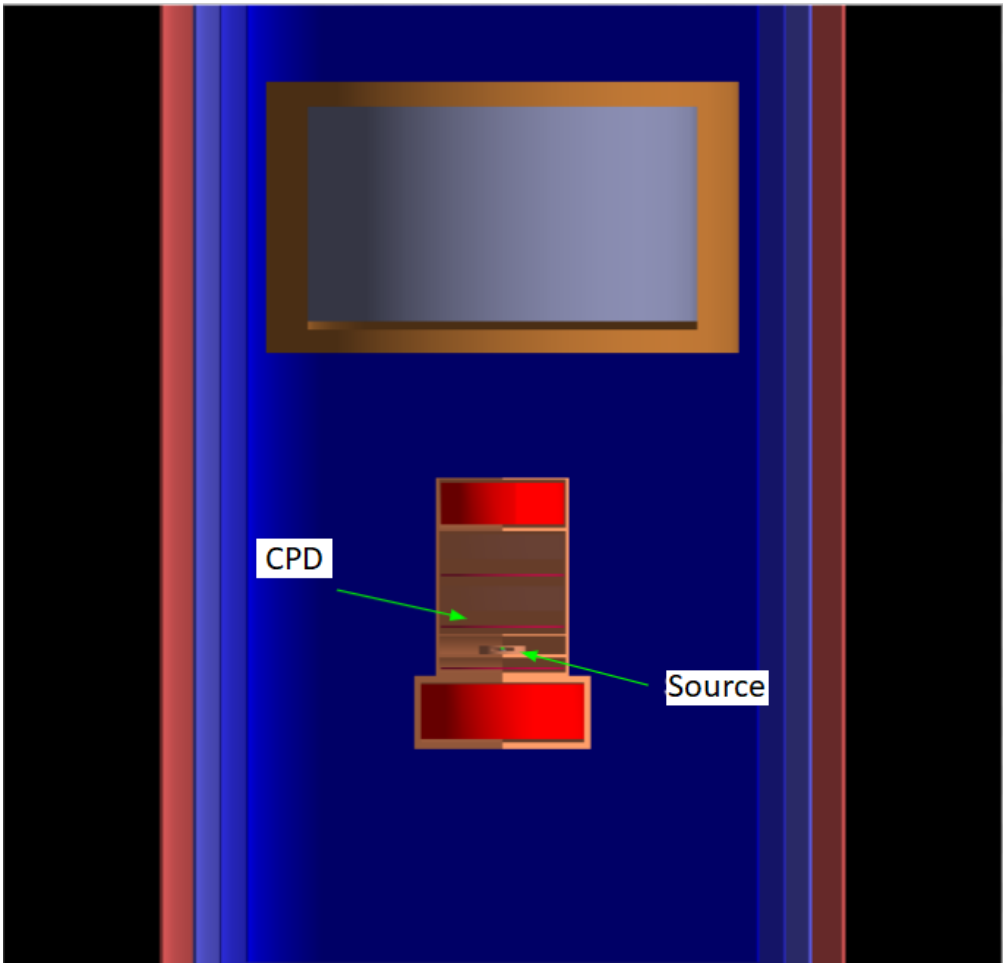


Figure 7.7: SuperSim rendering of detector stack in CUTE fridge. In addition to the source and CPD, we also see several other detectors included in the stack which were not investigated for this analysis. The gray object at the top of the cyrostat is a lead shield which is important for modeling environmental backgrounds but does not affect the background from the source.

7.3.2 Background Components

^{55}Fe X-rays

To model the characteristic x-rays from ^{55}Fe decay, we threw photons from the source according to the spectrum defined in Table . Photons at this low energy cannot penetrate the copper lid of the source holder, and can only exit through the collimator hole. To reduce the required number of simulated primaries, we biased the distribution by only throwing x-rays in a cone w/ half-angle of 5°within a radius of 1 mm (compared to the source radius of 2.35mm). We need to account for this bias when normalizing the simulation, and note that the geometric efficiency was changed by this bias as well. Regardless of how we biased the distribution, we want our simulation to always describe the same rate of photons escaping the collimator. From Eq. 7.1, we then have

$$A_0 k_0 = A_b k_b \tag{7.17}$$

Here the subscript "0" indicates the physical (unbiased) activity and geometric efficiency, and "b" indicates the biased values in our simulation. We determined the biased geometric efficiency by modifying our toy MC to throw the x-rays in a cone rather than 4π. From 10^8 trials of the toy MC, we calculated a geometric efficiency of 0.219 with a relative uncertainty of 5×10^{-3}%. Then the biased activity is

$$A_b = \frac{k_0}{k_b} A_0 = 150 \text{ Bq} \tag{7.18}$$

Then we determined the normalization for our simulation by defining an effective livetime per primary thrown, τ. For a process which happens with probability p per decay of ^{55}Fe, we have $\tau = 1/(pA)$. The probability for x-ray emission, $p_x = 0.289$ is given by summing the intensities in Table 7.1. Then the x-ray livetime per primary was $\tau_x = 0.02$ seconds/primary. We generated 10^{10} primaries in our simulation, corresponding to 6.34 years of livetime.

In order to validate the behavior of the x-rays, we repeated the simulation while varying thicknesses of aluminum foil and compared the detection rate of each line with Eq. 7.1. Fitting a simple exponential to the rate of each line as a function of the foil thickness, the decay constant gave the attenuation length and the prefactor was proportional to the geometric efficiency.

We saw that these values agreed with the ones we previously determined within 5%, so our simulation performed as expected. Next we determined the total detected x-ray rate from the simulation, and compared

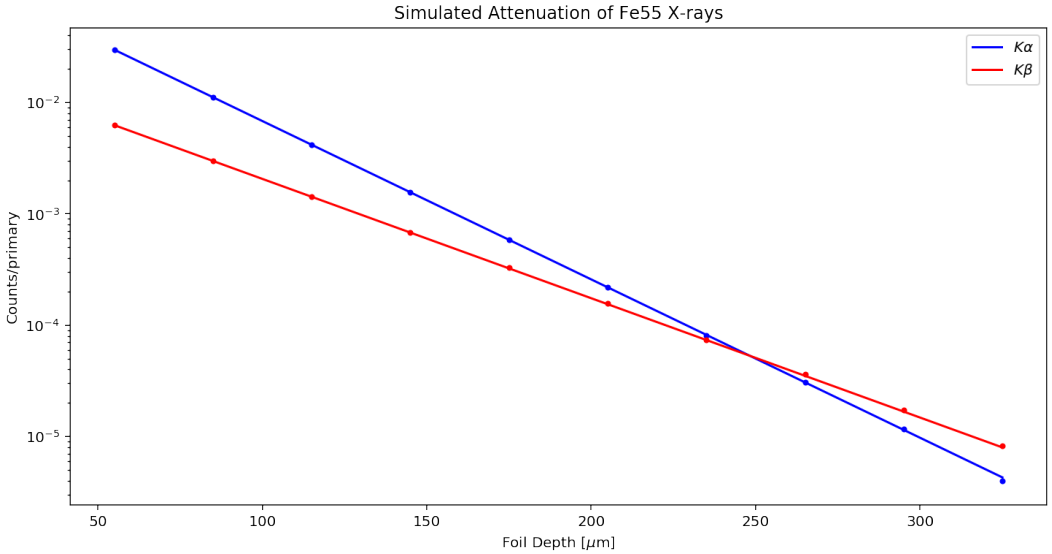


Figure 7.8: Effect of varying foil thickness on detection of x-ray peaks in the simulation. The blue line shows the intensity of the K_α line simulated in the detector, and red the K_β. The slope of each agrees well with the attenuation lengths in Table 7.2. The K_α is ~ 10 times as intense in the limit of no foil, and the lines achieve equal intensity at a foil thickness of ≈ 250 μm.

Method	Integrated Rate [mHz]
Calculated	$12.9^{+3.9}_{-1.3}$
Simulated	$12.7^{+3.9}_{-1.3}$
Observed	26.5 ± 0.3

Table 7.4: Total rate of ^{55}Fe x-rays.

it with the expected values from Table 7.3 and the observed rate in the data. We estimated the observed rate by integrating the spectrum between 5.4-6.9 keV and subtracting a flat background inferred from the side bands of this region. The results are given in Table 7.4. Note that the uncertainty on the expected and simulated rates were taken from the uncertainty of the source activity, and that on the observed rate was determined from counting statistics.

Though the simulation agrred well with our calculated rate, both underestimated what we observe in the data by $\sim 50\%$. Since we already accounted for the systematic uncertainty on the activity this suggests that we have likely underestimated the geometric efficiency of the collimator, which is reasonable since we have not actually measured the dimensions.

^{55}Fe Inner Bremsstrahlung

We introduced the notion that the electron capture decay can be accommodated by IB if the captured electron effectively interacts with the Coulomb potential of the nucleus as it is captured. In this case, the decay looks like $p + e^- \rightarrow n + \nu_e + \gamma$. Notice that this is essentially a 3-body decay mediated by the Weak interaction, and heuristically we expect the spectrum of the emitted γ to resemble that of an electron following the nuclear β-decay $n \rightarrow p + e^- + \bar{\nu_e}$. For the case of ^{55}Fe, the endpoint occurs at 231 keV. Photons at the higher end of this spectrum are not attenuated by the thin copper lid of the housing, and we expected to detect these with CPD. The IB process is suppressed by the additional vertex which appears with the creation of the γ, and we expect $\sim 10^{-5}$ IB photons per K-shell capture [114]. Though this is a relatively rare process, the suppression was comparable to that which is imposed on the K-shell x-rays by the presence of the collimator. Hence we actually expected the observed rate of IB photons to be comparable to that of the K-shell x-rays.

There is currently no model for the IB process in Geant4. For this study, we accounted for IB by taking the spectral shape from [115] and throwing photons according to this spectrum from the surface of the source. The spectrum is shown in Fig. 7.9. Note the rise at low energy. This occurs since the spectrum must reproduce characteristic x-ray lines, which made modeling the shape difficult in this region. Excluding photons below this energy from the simulation shouldn't bias the result too much, since we didn't expect a large fraction to be able to penetrate the housing lid.

The probability for an IB photon to be produced in the range 17-231 keV (the range in Fig. 8) is given by

$$p_{IB} = (\frac{I_2}{I_1}) I_1 p_K \tag{7.19}$$

$$I_1 \equiv \int_{35\text{ keV}}^{231\text{ keV}} dE \frac{dW}{dE}(E) \tag{7.20}$$

$$I_2 \equiv \int_{17\text{ keV}}^{231\text{ keV}} dE \frac{dW}{dE}(E) \tag{7.21}$$

where $p_K = 0.885$ is the probability for decay via K-shell capture [112], and $\frac{dW(E)}{dE}$ is the differential spectral shape in Fig. 8. From [114], we have $I_1 = 3.24 \times 10^{-5}$. We then determined the effective livetime $\tau_{IB} = 1/(p_{IB} A) = 0.146$ seconds/primary. Note that we used the nominal activity rather than the biased activity, since we threw the primaries in 4π. We simulated 10^7 primaries, corresponding to 19 days of

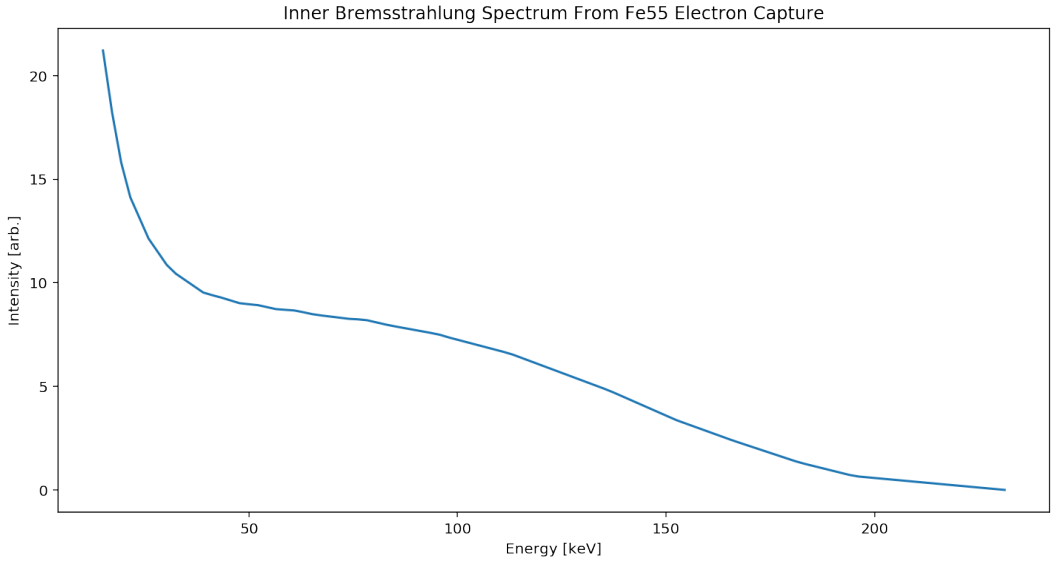


Figure 7.9: Spectral shape of IB photons produced via electron capture in ^{55}Fe [115]. Above 50 keV, the curve resembles a β-spectrum with a 231 keV endpoint. The sharp rise at lower energies can be interpreted as the tail of the ~ 6 keV x-ray peak, which is several orders of magnitude greater in intensity.

livetime.

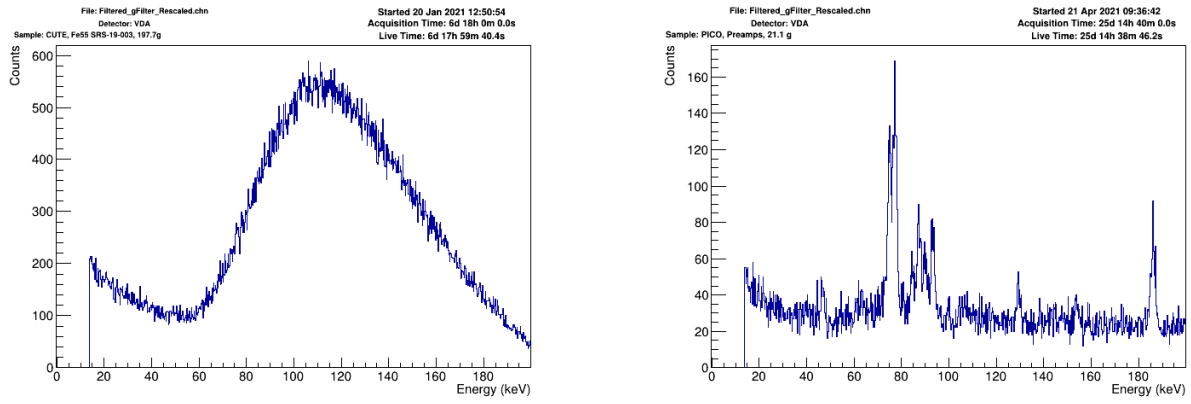
Radiogenics in Source Holder

CPD was also sensitive to radioactive contamination in the copper source holder (e.g. cosmogenic ^{60}Co and primordial ^{238}U). This contamination could have contributed to the excess background we observed, so we to included it in our model. The radiopurity of the source holder was assayed at Snolab using the VDA HPGe detector [116].

We simulated each component measured in the assay as being distributed uniformly throughout the

Species	Measured Activity [mBq/kg]
^{238}U - Upper	< 1.4
^{238}U - Lower	< 248
^{57}Co	< 54
^{60}Co	175 ± 8
^{210}Pb	$< 6.6 \times 10^5$
^{235}U	6 ± 4
^{232}Th	7 ± 2
^{40}K	7 ± 20

Table 7.5: Results of source holder radiopurity assay



(a) (b)

Figure 7.10: (left) Low energy spectrum from assay of the source holder. The wide peak at 120 keV is likely due to IB in the source, and is the cause of the high limit on ^{210}Pb measurement. (Right) Assay result from another sample in the same detector over the same energy range. Several are resolved, which demonstrates that the broad peak in the left plot is not due to insufficient resolution of the detector.

bulk source holder. The decay physics were handled via G4RadioactiveDecay (i.e. we use the AddIsotope command in SuperSim). Note that we assumed that the equilibrium of the ^{238}U chain is broken, and hence distinguished between the upper chain which started from ^{234}Th and the lower from ^{226}Ra.

The results of the assay are given in 7.5. Note the large rate suggested due to ^{210}Pb. Using the upper limit value of the activity would result in a rate more than twice the observed background. The limit on the activity of ^{210}Pb was inferred from the measurement of the 46.5 keV line. Fig. 7.10 shows the assay spectrum near this region in the VDA detector for the source holder (left) and a different sample for reference (right). Comparing these spectra, we saw that several lines we expected to resolve when assaying the source holder were washed out by a broad peak around 120 keV. This peak was roughly consistent with what we expect from the IB spectrum. The turn on point ∼50 keV was consistent with IB photons below this energy being attenuated by a few millimeters of copper, and the turn off point roughly matched the 231 keV endpoint of the spectrum. Hence it is likely that IB photons obscured the 46.5 keV line, resulting in the high limit on ^{210}Pb. For this reason we omitted ^{210}Pb from our simulation of the radiogenic background.

7.3.3 Simulation Results

Our simulation suggested that the background produced when deploying the ^{55}Fe source was dominated by IB photons at low energy. The rate of this background determined by the simulation overestimated the

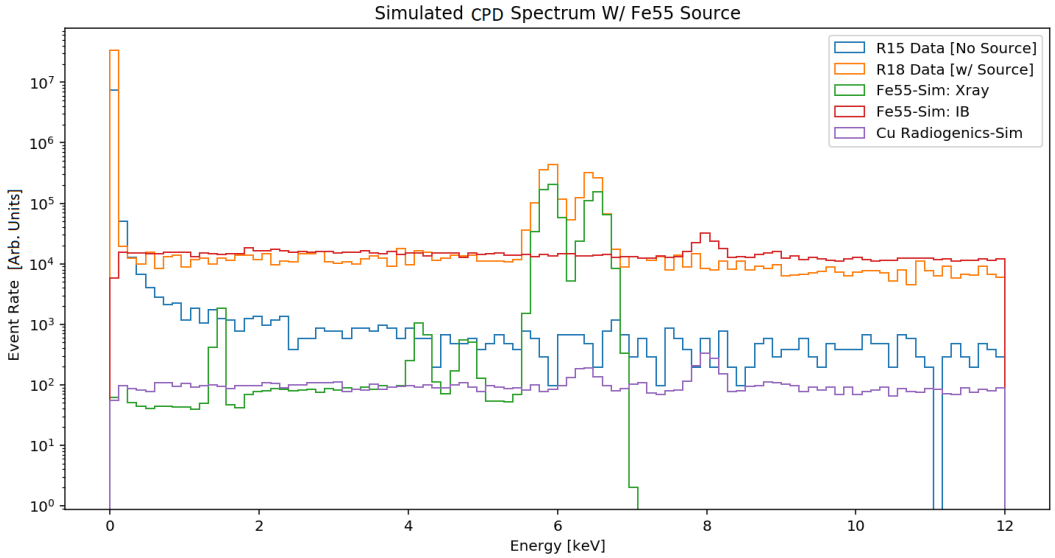


Figure 7.11: Simulation of source background, broken down by component. The orange and blue histograms are the actual data collected at CUTE, with and without the source respectively. Green histogram is our simulation of the characteristic x-rays, purple the radiogenic background from the copper source holder, and red the IB background. Our model underestimates the x-ray peak rate and overestimates the total background, but both are of the correct order of magnitude. This gives us a strong argument that the observed increase in the background is dominated by IB events.

observed increase by 40%. Since we also underestimated the x-ray rate by 50%, these discrepancies were not resolved by scaling the source activity. Though the disagreement was substantial, saw that they agreed within an order of magnitude. We concluded that the IB background was large compared to that produced by contamination of the source holder, barring exceptionally large ^{210}Pb contamination.

The model we used for IB was crude. The shape of the spectrum was obtained by digitizing Fig. 6 in [115]. Another data point was added to force the endpoint at 231 keV. There was also some uncertainty associated with the low energy cutoff at 17 keV. The attenuation in copper at this energy is comparable to the thickness of the lids, so it's possible that we excluded photons which could in fact penetrate the source holder. Our ultimate goal would be a full implementation of the IB process in Geant4, which is beyond the scope of this work. This study did demonstrate the that the IB process dominated the background, and more careful modeling is needed if one wants to use ^{55} Fe for in situ calibration of a DM search.

One somewhat surprising implication of the the simulation was that it suggested that a peak at 8 keV could be resolved. This peak was not observed the actual data. The energy is consistent to copper fluorescence emission from the source holder induced by IB photons. Even though it is at a higher energy than

the aluminum line, the general difficulty of calibrating our detectors in this energy range makes it worth considering whether the copper line could actually be resolved. Note that this feature did not necessarily depend on the hypothesis that the observed background is dominated by IB. Any ionizing radiation that originated from the source holder at high enough energy to produce the observed background would also generate copper florescence. Though we appeared to overestimate the IB background, the simulated peak was prominent enough that it should have be visible when scaling the simulated background. We consider here a few possibilities to explain the discrepancy in observing the peak:

- The rate of fluorescence depends on the distribution of copper in the simulation geometry, which we hadn't described with sufficient fidelity. Though we know not all of the source holder dimensions were modelled correctly, note that the attenuation length for an 8 keV photon on copper is ~ 22 μm, much less than the 1 mm thick lid closest to CPD. Hence we expected the observed fluorescence photons to predominately originate from near the surface of the lid, and the rate should have been mostly insensitive to the overall distribution of copper.

- The peak is potentially observable, but we din't see it because our calibration above the x-ray peaks wasn't sufficient. The calibration is defined without any information above 6.5 keV, and we know we started to see non-linearity effects in the estimator in this region. Though the calibration included a saturation correction, we didn't have an opportunity to validate it at higher energies. One simple but extreme example is to look at the endpoint. With the present calibration we saw an endpoint at $\sim$60 keV, whereas we know the IB background extends up to 231 keV. On one hand this isn't surprising, CPD has a finite dynamic range, above which we won't expect to estimate the energy, and the end of this range may very well be $<$100 keV. But it does raise the question of what exactly the dynamic range of CPD is, and how we would estimate it.

- The simulation didn't account for differences in detector response between surface and bulk events. Clear clustering was observed in the Amplitude vs. Integral plane when investigating the ^{55}Fe x-ray peaks [110]. A 6 keV photon only penetrates ~ 50 μm in silicon, so one hypothesis is that this clustering was due to the fact that the x-rays create surface events. The background due to Compton scattering from higher energy photons would instead be distributed throughout the bulk, potentially producing a different detector response.

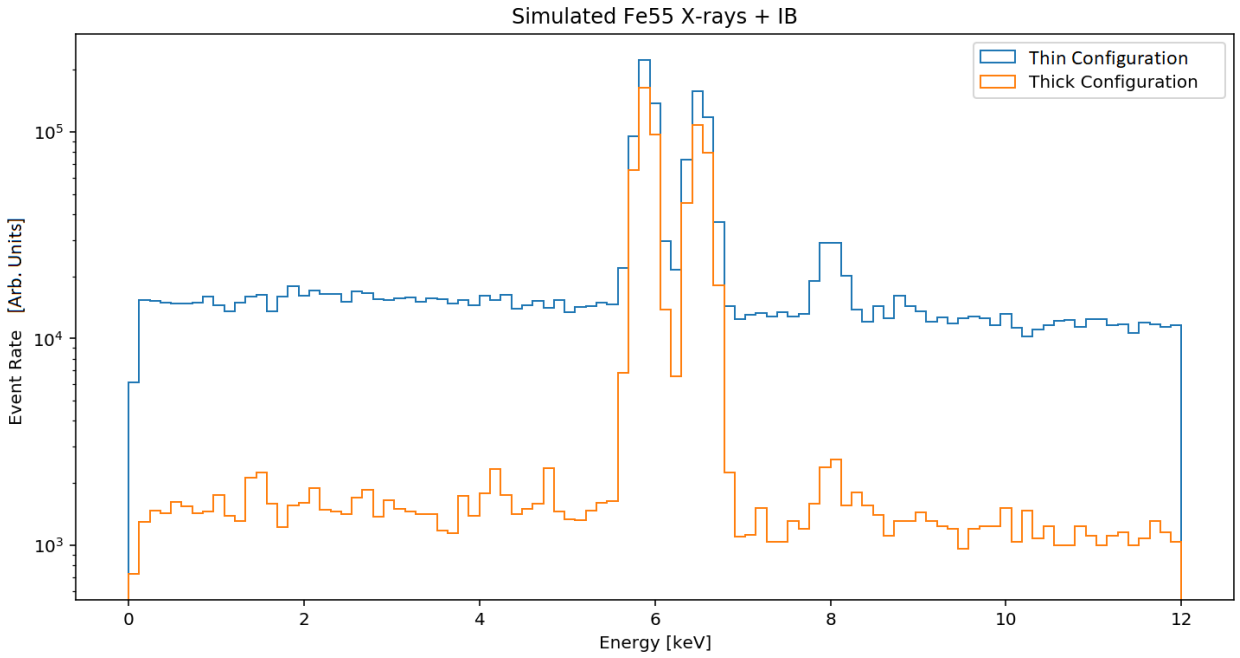


Figure 7.12: Comparison of source background with increased lid thickness. The blue histogram is with nominal lid thickness of 1 mm, while the orange considers increasing this thickness to 1 cm. The, which is dominated by IB, is significantly reduced by increasing the lid thickness with, only a moderate penalty on the calibration x-ray rate.

7.3.4 Shielding IB x-rays

If IB photons were the dominant background, we could consider shielding them. One possibility would be to increase the thickness of the copper lids in the source holder. The attenuation length for a 200 keV photon in copper is ∼7 mm. We considered a model of the source holder with the dimensions altered to accommodate a 1 cm thick lid, and then repeated the simulation described above for the ^{55}Fe x-rays and IB photons. This test suggested that using a lid of this thickness provided roughly an order of magnitude reduction of the IB background.

We considered this solution because it was trivial to implement in the simulation. In reality, it would be much more efficient to shield the IB x-rays with a lead annulus. In any case, we have demonstrated that the IB background can in fact be moderated to the point where it is comparable to the environmental background - while at the same time preserving the efficiency of the calibration x-rays.

Chapter 8

Dark Matter Search With CPD

8.1 Introduction

A upgraded version of the CPD device, was operated at the University of Massachusetts Amherst (UMass) in a cryoconcept fridge with the purpose of performing a search for LDM. This analysis represents a joint effort between the SuperCDMS and SPICE/HeRALD collaborations, both of which sharing expertise on CPD style detectors. The device was operated during UMass Run 26 (R26) and Run 28 (R28), and read out with Magnicon SQUIDs by a National Instruments DAQ. Details on the UMass infrastructure can be found in Ref. [43], which describes a different analysis performed in the same fridge. The data stream was continually read out, such that no triggering was done during data collection and all processing was performed offline.

The two runs demonstrated very different noise conditions. The noise environment observed in R28 was favorable compared to R26, and it was decided to perform the LDM search using the R28 data. This chapter will discuss the development of cuts, signal model and evaluation of systematic uncertainties for the DM search carried out using this data set.

Several studies were performed using the data from R26 to characterize the behavior of the experimental baseline and glitches observed in data. These studies do not directly affect the results of the R28 LDM search, but still are of some relevance since the cuts used in the R28 analysis were largely inherited from those developed for R26. These studies may also be of general interest for other contexts where it is necessary to characterize and remove low frequency noise. With these considerations, we will do not discuss the R26

analysis in this chapter, but relegate it to the relevant appendix.

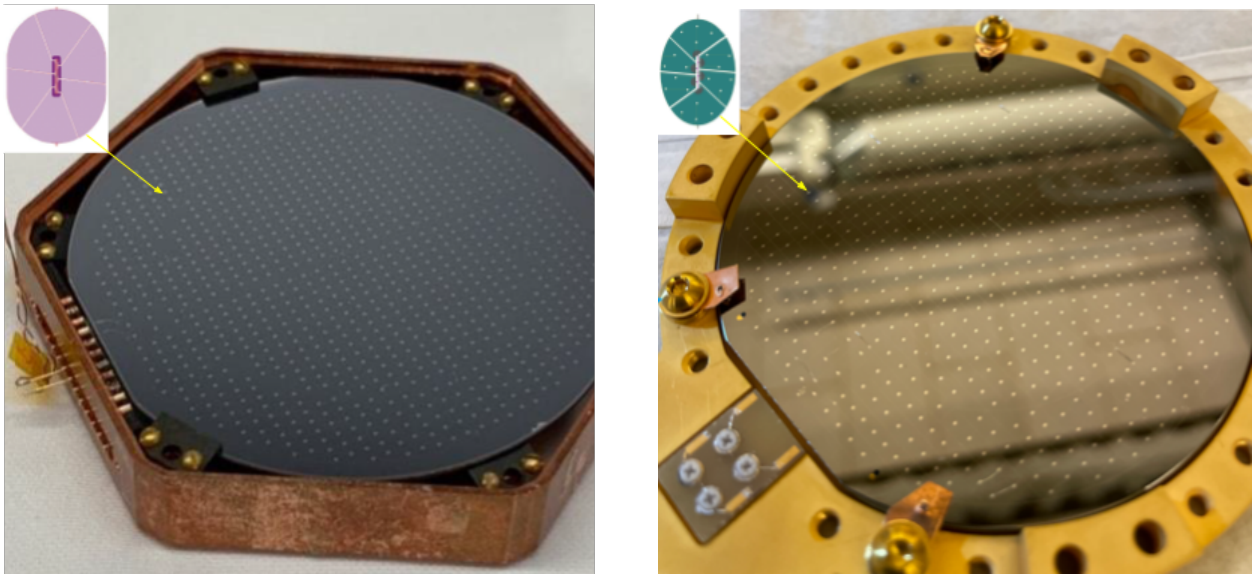
8.1.1 Comparison of CPD Generations

Both the original CPD and upgraded CPD are silicon detectors with a ≈ 3.3 cm radius and 1 mm thickness. The differences between the two are introduced in the sensor layout. The dimensions of the QET were adjusted such that the region where the aluminum and tungsten overlap features a thicker tungsten layer. This has been shown to increase the efficiency of quasiparticles in the aluminum fin to deposit their energy into the tungsten, thus increasing the net efficiency [103]. Additionally, there are fewer total sensors on the upgraded CPD, which decreases the TES heat capacity and increases the detector sensitivity. A comparison of the relevant parameters for each version is given in Table 8.1, and a visual comparison of the detectors in Fig. 8.1.

In the following sections, we will discuss the measurement of the net efficiency and baseline energy resolution of CPD, but we take a moment now to compare the results between CPD generations. These comparisons give us a good empirical demonstration of the resolution scaling introduced in Eq. 6.1. For one, we see that the net efficiency is improved in upgraded CPD despite the fact that there are fewer sensors. This gives us evidence that we can increase the efficiency even if we expect to collect fewer phonons. The improvement must come from the updated QET design increasing the fraction of quasiparticles which are collected by the TES. Perhaps even more striking is the fact that the baseline resolution improves in the second generation even though T_c is higher. From the scaling law we saw that the strongest dependence on the resolution follows T_c. This tells us that we have effectively improved the efficiency and reduced the heat capacity enough in the second generation to overcome this unfavorable scaling. The upgraded CPD demonstrates a better baseline resolution than the original. Hence we expect to improve our sensitivity LDM at smaller masses with this detector, even if we were to simply repeat an updated version of the original experiment.

Quantity	Original CPD [117]	Upgraded CPD
T_c (mK)	42.5	51
Net Efficiency (%)	≈ 17	≈ 25
# QETs	1031	673
Active Coverage Fraction (%)	1.9	0.68
Baseline Resolution (eV)	3.86	2.26

Table 8.1: Comparison of detector parameters across CPD generations.



(a) (b)

Figure 8.1: Photos of the two CPD generations. (Left) shows the original CPD, and (right) is the upgraded version. Note the difference in QET geometry, and density of sensor pads on each surface. The upgraded version uses a smaller QET pad, and the sensors are distributed more sparsely across the detector surface.

8.2 Calibration

8.2.1 Measurement of Net Efficiency

We defined the net efficiency ϵ of our device such that if there is some energy deposition E_{dep} in the bulk, then the energy absorbed by the TES E_{abs} is given by

$$E_{abs} = \epsilon E_{dep} \tag{8.1}$$

Under the assumption that the electrothermal response time of the TES is sufficiently fast, the energy deposited in the TES is well approximated by E_{ETF}, the energy removed by electrothermal feedback. This can be estimated directly by integrating the observed current trace following the deposition

$$E_{abs} \approx E_{ETF} = \int dt \big[(V_b - 2I_0R_L)\delta I(t) - R_L(\delta I(t))^2\big] \tag{8.2}$$

This property has led some to describe the TES as "self-calibrating" [75], though this is perhaps too strong of a statement given the time and effort people put into calibrating these devices. Really what this property does is give us a robust way to measure the net efficiency. All we need to do is make a histogram of E_{ETF} from some calibration data and identify a peak of known energy. Then the ratio of the known energy of that peak and the reconstructed energy E_{ETF} is our measurement of ϵ.

Our implementation of this measurement for the upgraded CPD closely followed that carried out for the original analysis [117]. For our peak of known energy, we used the 1.49 keV aluminum fluorescence line induced on a thin foil by a ^{55}Fe source emitting $\approx$ 6 keV x-rays. The resulting spectrum is shown if Fig. 8.2. From this, we inferred a net efficiency $\epsilon = 25.5\%$ [118].

8.2.2 Calibration of Pulse Amplitude

The net efficiency is an important number because it is a physical characteristic of the detector and sensor layout, and it gives us a robust energy estimator. On the other hand, this estimator does not give us the best possible resolution. Under the assumption that we have a good description of the signal template and the underlying noise environment, we can improve the resolution by applying an Optimal Filter (OF) to the observed pulse an use the filtered amplitude as our estimator.

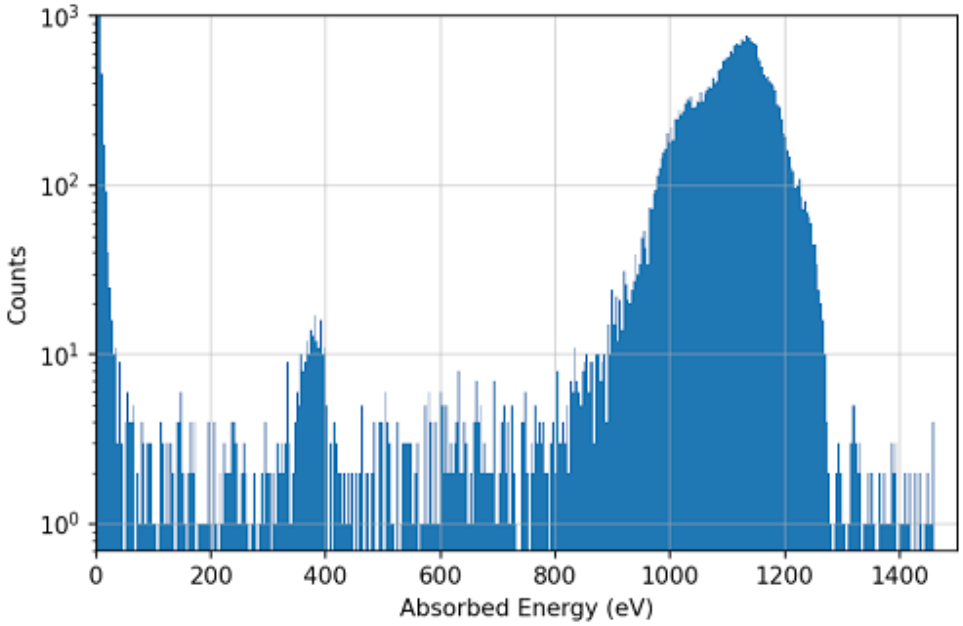


Figure 8.2: Histogram of absorbed energy estimator used to determine the net efficiency. The 1.5 keV appears just below 0.4 keV in the "Absorbed Energy" parameter, which suggest a net efficiency of $\approx 25\%$. The structure around 1.1 keV in absorbed energy is due to the ~ 6 keV ^{55}Fe x-rays. The two peaks are not well resolved and have a reduced net efficiency due to the onset of saturation at this energy.

Say that our noise PSD is described by $J(f)$, and we have a template for our signal $\tilde{s}(f)$. Notice that we're describing everything in terms of frequency, since we typically assume that the noise is correlated in the time-domain but not the frequency-domain. Implicit in this is that we actually record the data in the time-domain and then relate it to frequency via the Fourier transform. Next, consider that we observe some actual triggered data $\tilde{v}(f)$ and want to estimate the amplitude A of the underlying signal. We can do this by evaluating the goodness-of-fit via the χ^2, and then choosing the amplitude which minimizes it

$$\chi^2(A) = \frac{|\tilde{v}(f) - A\tilde{s}(f)|^2}{J(f)} df \tag{8.3}$$

Note that this is the simplest realization of the OF. One can consider the case where one allows for the template to shift in time, or there to be multiple signals present in the data. For a more complete discussion of the general case, including how to go about deriving the optimal estimators, see [102].

Let's return to the time-domain. Suppose that we have recorded an event $\delta I(t)$ and applied the OF to determine the amplitude A such that $\delta I(t) \approx As(t)$. If we evaluate our integral estimator for this event, we have

$$E_{dep} \approx \frac{1}{\epsilon} \int dt \big[A(V_b - 2I_0 R_L) s(t) - A^2 R_L s(t)^2 \big] \tag{8.4}$$

When searching for LDM, we're most interested in the limit of small amplitude. Hence we apply the approximation

$$A \ll \frac{V_b}{R_L} - 2I_0 \tag{8.5}$$

Comparing this to our data, this limit holds until we start to see saturation effects. Then we approximate the energy integral as

$$E_{dep} \approx \frac{1}{\epsilon} A(V_b - 2I_0 R_L) \int s(t) dt \tag{8.6}$$

What we really want is to take the observed amplitude A and calibrate it to the deposited energy via some constant η

$$E_{dep} = \eta A \tag{8.7}$$

Comparing this to 8.6, we immediately read off the relevant coefficient

$$\eta = \frac{1}{\epsilon}(V_b - 2I_0 R_L) \int s(t) dt \tag{8.8}$$

We have already discussed that we can estimate ϵ by identifying a peak of known energy in the E_{ETF} histogram. Then the determination of parameters (V_b, I_0, R_L) are standard measurements taken at the beginning of a run with at TES device. The definition of the template can be somewhat subtle, but the principle is straight forward. We just need to identify a collection of "good" pulses and average them, then apply some filter or empirical fit to smooth it out. Once we have all this, our detector is effectively calibrated in the saturation-free limit. With the calibration determined, we can estimate our baseline resolution by looking at the distribution of OF amplitudes reconstructed for random triggers. For this analysis we find $\sigma_E = 2.3$ eV

8.3 Cut Development

8.3.1 Good Time Cut

Our dark matter search data is a subset of the data collected during Run 28 of the cryoconcept fridge located at UMass Amherst. A roughly 40 hour period of DM search data was collected at the beginning of the run, which is referred to as "Trial 1". The run was then suspended approximately two weeks, during which other studies of the detector were performed. Another 50 hour period of DM search data was then taken at the end of the run, which we call "Trial 2". The reason for collecting search data at the beginning and end of the comparatively long fridge run was a cross check of the hypothesis that the event rate would decrease over time as the crystal relaxed from the stress induced by the clamps. It was found that the integrated rate from our trigger threshold up to 40σ did decrease from ≈ 0.8 to ≈ 0.6 Hz. While it was concluded that we didn't have enough information to make a more detailed statement about the evolution of the trigger rate from these two data points, we decided that the dark matter search should be performed on the Trial 2 data since the lower background rate benefits our sensitivity. Throughout the rest of this note we will try to make the distinction between the Trial 1 data and Trial 2 data where relevant, but any unspecified references to "Run 28" are directed at Trial 2 since it contains the relevant DM search data.

After selecting the period of data taking for the DM search, we define two more livetime cuts to ensure the data quality. We call the first the "broad time cut", which was defined qualitatively to restrict our analysis to an interval where the baseline was observed to be stable. The second is called the "fine time cut", which involves defining a quantitative threshold to remove periods where a statistical increase in the trigger rate is observed. Taking these two definitions together form a single livetime cut, which we refer to

as cGoodTime.

Broad Time Cut

To motivate the broad time cut, it is useful to look at the baseline (defined as the average of the first 9.5ms of the 20 ms trace) estimated from random triggers through Trial 2 of Run 28, which is shown in Fig. 1. We see that there is a sharp rise in the mean baseline around 10 hours into the data taking. Since the baseline is a measure of the signal gain, it is clear that any attempt to include the entire period in a single analysis would likely require a time-dependent calibration. We take a step further into comparing the data before and after this shift in the baseline and compare the PSDs estimated after applying the **QETpy.autocuts** algorithm, which removes outliers in the distribution of prepulse baseline, slope and amplitude of randoms to curate the PSD [119]. It is apparent that the decrease in the magnitude of the baseline is associated with a large increase in the noise below 1 kHz.

With these considerations in mind, it was determined that only data from the beginning of the data collection period should be included in the dark matter search. We decided that only performing the search on the first 8.3 hours would provide us with sufficient statistics while providing enough of a buffer period between the shift in the baseline to leave the data unaffected by the introduction of the excess low frequency noise. This choice defines the "Broad Time Cut", any data taken after 8.3 hours into Trial 2 is removed from our livetime.

Fine Time Cut

When defining the broad time cut, we observed that the introduction of low frequency noise after the baseline shift also caused a large increase in the trigger rate (see Run 28: Cuts Summary, a similar relationship is shown for Trial 1). In order to ensure the consistency of the noise conditions for our dark matter search data, we decided to remove periods where the trigger rate appears to increase beyond what we would expect from statistical fluctuations. We define the trigger rate to be the integrated rate in the region where we expect noise triggers to appear, 4.5-40 σ or 10-90 eV. Also note that we have not applied any quality cuts besides the Broad Time Cut at this point. The estimated rate as a function of time over the course of the run is shown in Fig. 2.

In order to estimate the distribution of the rate, we must bin our data in time. We start by considering five minute intervals, which is the length of a saved data file. Keep in mind that all of our cut development

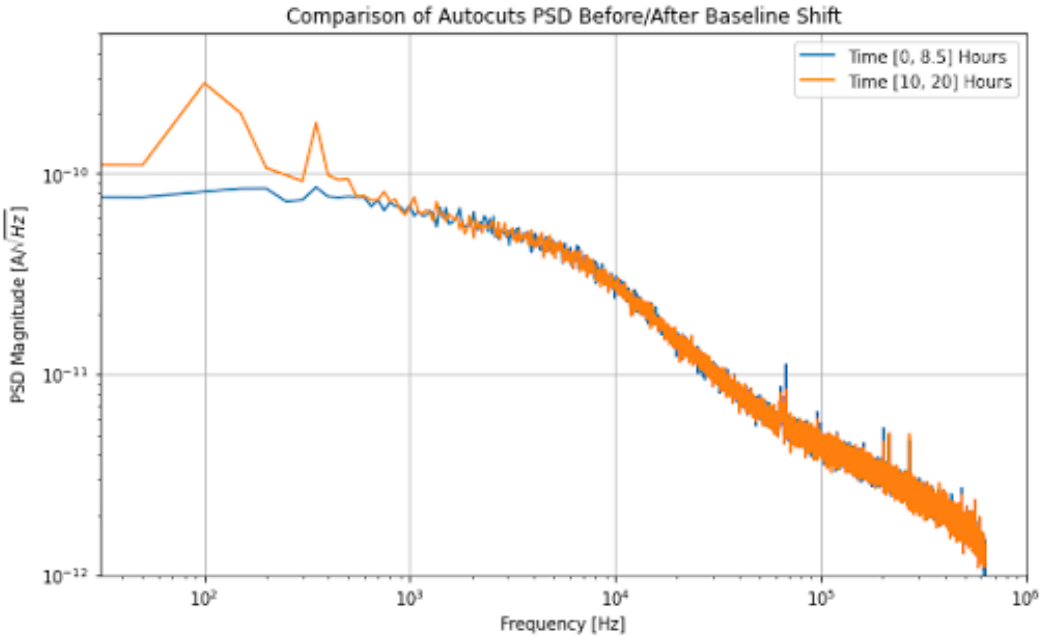


Figure 8.3: Change in noise between periods separated by the broad time cut. The blue curve shows the noise in the period preserved by the cut, while the orange is from a subset of the period removed. There are clear, large peaks in the noise PSD in the removed period.

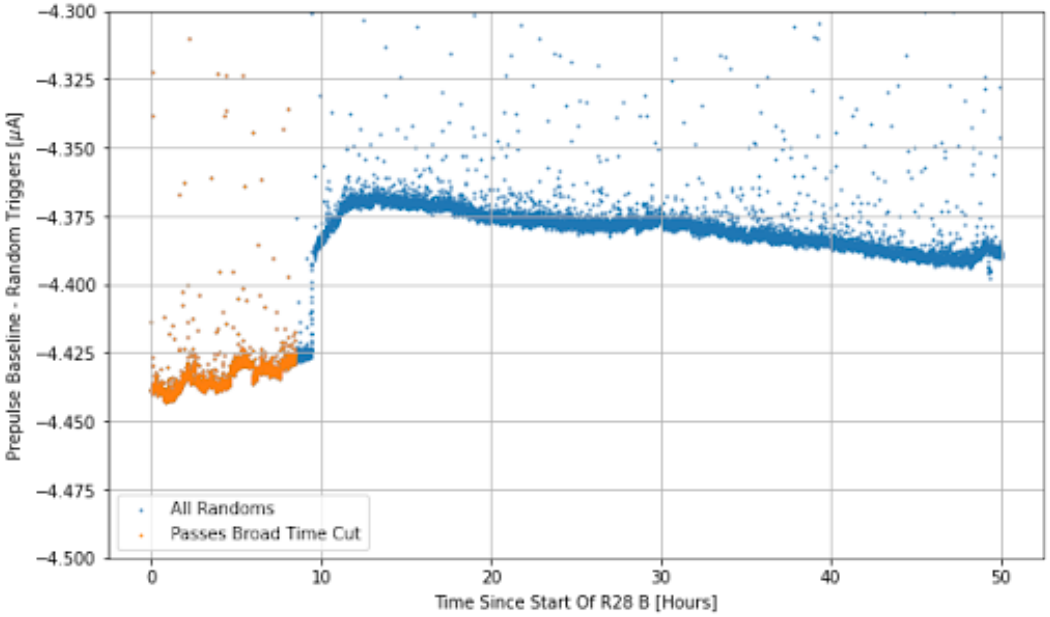


Figure 8.4: Application of the broad time cut across the entire data series. Blue points represent all data in the series, while those in orange are preserved by the cut. We see that there is a distinct jump in the baseline which is accompanied by the increase in noise shown in 8.3. Define the broad time cut to only include data before this jump.

studies were performed on the 50% of data which wasn't blinded, and that alternating files were subjected to blinding (i.e. when indexing the files sequentially, every file with an even index is blinded). Hence five minutes is the longest interval we can consider without having to correct for the blinded periods. From this we found a mean rate of 0.69+/-0.01 Hz. The proposed implementation of the cut is to fit a Gaussian to the distribution and remove periods with a rate 2σ above this mean. The resulting fit is shown in Fig. 3 (blue), where the fit σ is 62+/-6 mHz. This puts the proposed cut threshold at 0.813 Hz.

To confirm that it is reasonable to describe the rate distribution as Gaussian, we repeated this study with finer time bins - reducing the chosen interval from 5 to 1 minute. Note that reducing the time interval also reduces the statistics of any particular estimate of the rate, hence we expect to observe larger statistical fluctuations in the distribution. The derived fit parameters are summarized in Table 1, and the fit is drawn in Fig. 3 (red). We see that both distributions are reasonably well approximated by a Gaussian, the means agree regardless of the bin size, and the width increases as expected.

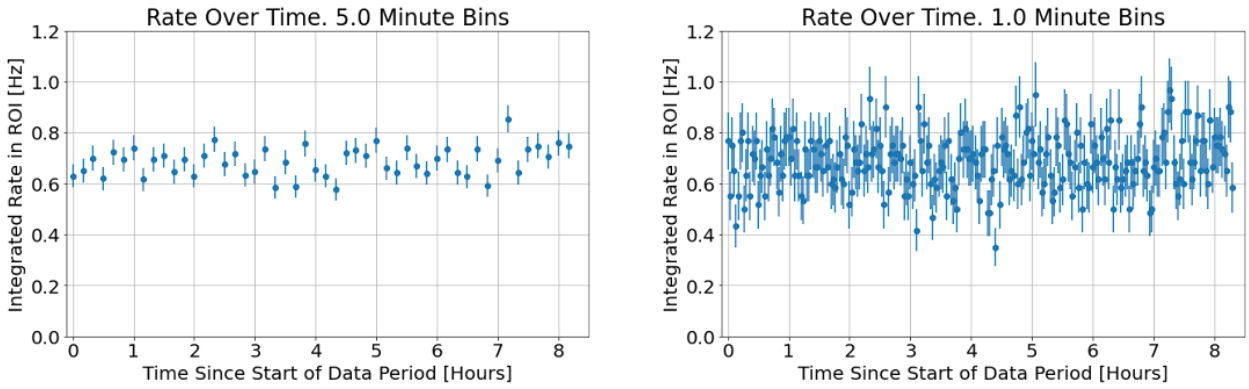
With our proposed cut definition, the choice of 5 minute bins would remove 5 minutes from our livetime, while choosing 1 minute bins would remove a total of 9 minutes. Hence the choice we make will make a negligible difference in our dark matter search result, especially when we consider that our sensitivity is limited by backgrounds rather than exposure. Taking this into consideration, we decided to implement the cut with 5 minute bins, since it is technically simpler to implement the cut at the level of individual data files rather than define smaller time intervals. After unblinding, to apply this cut we will simply count the rate in each unblinded file and compare it to the threshold defined here, rather than re-evaluate the rate distribution. The rate over time is shown once again in Fig. 4, highlighting the single 5 minute period of non-blinded data removed by the cut.

8.3.2 Pre-Pulse Baseline Cut

When the TES is in a steady state, we can interpret the baseline current as a measure of where we are in the transition. If we look at a trace and see that the prepulse baseline deviates from our expectation, we lose confidence in our energy reconstruction since the event could potentially have a different gain.

Since the detector was run above ground, we expect a large rate of muons to heat up the detector. These muon events will have long thermal tails and any low energy events which get reconstructed on these tails will have an elevated prepulse baseline.

In Figure 5 we see the prepulse baseline over the course of the R28 after cGoodTime is applied. The plot



(a) (b)

Figure 8.5: Rate over time averaged over two different time bins. Using the shorter time bins (right) introduces larger fluctuations on account of there being fewer events per time bin. Despite our particular choice of bin size, we see that the rate consistently averages ≈ 0.7 Hz.

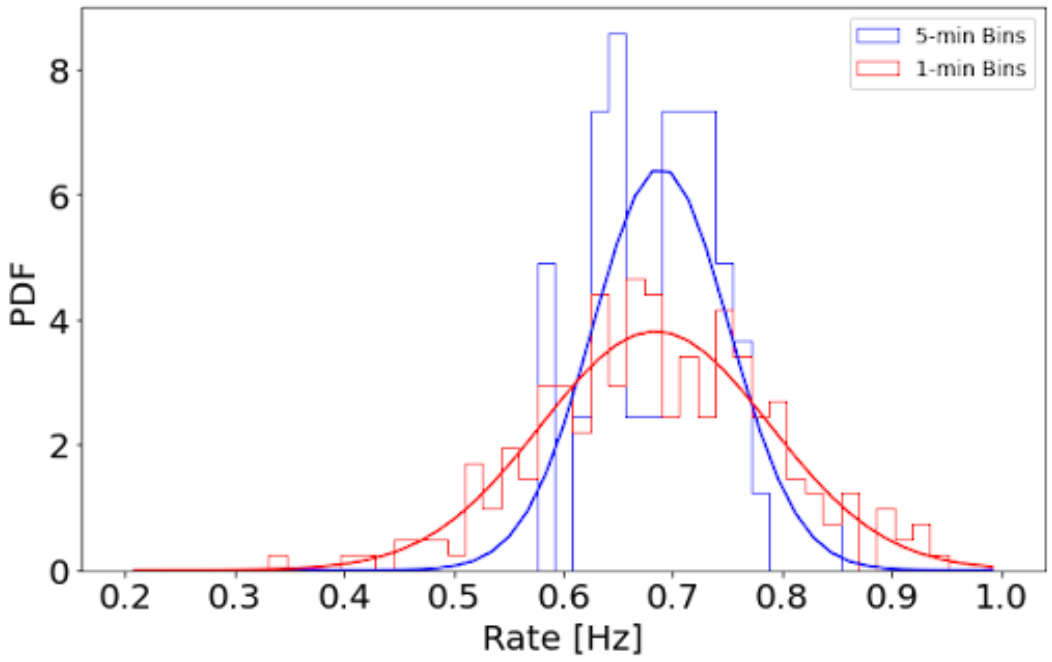


Figure 8.6: Comparison of rate distribution for each bin choice. We fit a Gaussian to each histogram, and see that the shape agrees reasonably well for each choice of time bin. The increased width of the 1-minute bin distribution is consistent with the expected increase in statistical fluctuations.

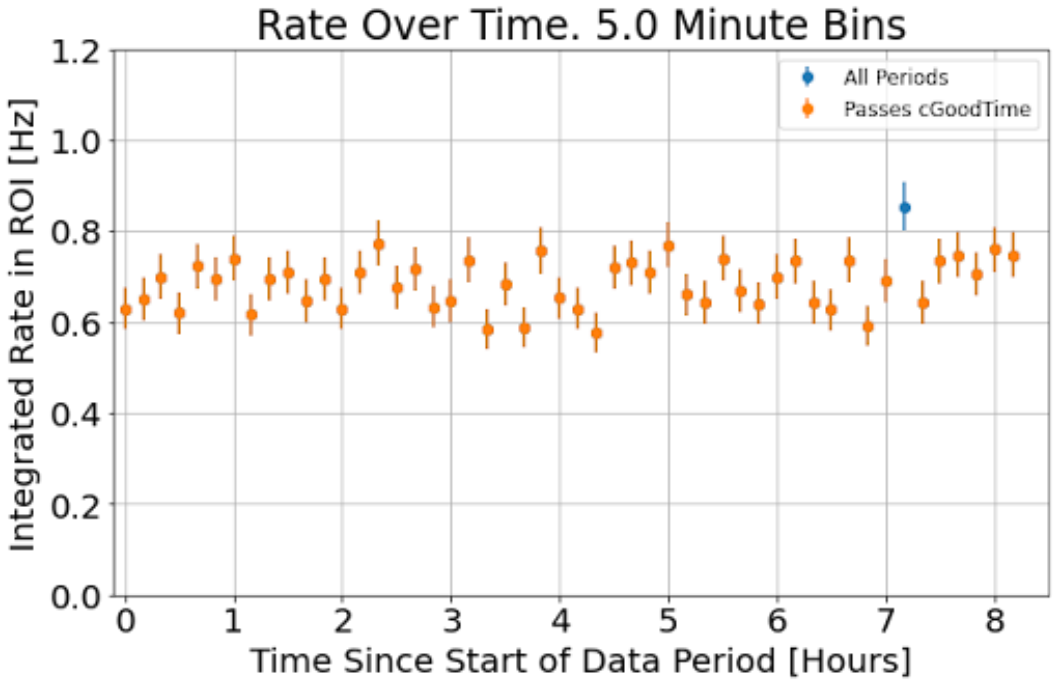


Figure 8.7: Application of the fine time cut to data preserved by the broad time cut. Periods which pass the cut are colored orange. Only a single period is removed by the cut, where we observe an anomolous upward fluctuation in the rate.

shows the "recorded" TES current converted from the ADC, which differs from the true current through the TES by an arbitrary DC offset from the SQUID. Here we use 20ms traces, and the prepulse baseline is defined as the average of the first 9.5ms of the trace.

We define the prepulse baseline cut by dividing the data series into time intervals and fitting a gaussian to the most prominent peak in the distribution of baseline for triggered events. The chosen time interval is 30 minutes, and events within 3σ of the fit mean are considered to pass the cut. The choice of the 30 minute interval takes two considerations into account. We want to choose a time which is short enough such that we're not averaging over large fluctuations in the baseline, but long enough that we have robust statistics when fitting the distribution of the baseline. Looking at the baseline over time, shown in Fig. 5 we see that the baseline is relatively stable on the scale of an hour, justifying our choice in terms of the former concern.

Considering the question of sufficient statistics, given our raw trigger rate of $\sim$0.7 Hz we expect $\sim$1300 samples of the baseline in each interval, which is surely sufficient to perform our fit. One could consider even finer time bins with these statistics, but our choice not to do so was somewhat historically motivated. The original plan with this analysis was to perform a dark matter search on the data collected during UMass fridge R26. This data set demonstrated a number of peculiar behaviors in the baseline which we had to correct for. One not-well-understood feature is that frequency, stochastic, sudden shifts in the baseline were observed over the course of the run (see Prepulse Baseline Cut). Hence this relatively long period was chosen such that these shifts were well resolved when looking at the baseline distribution. This phenomenon is not observed in R28, but we decided not to modify the cut definition since it already satisfies the criteria we just outlined.

8.3.3 Slope Cut

To estimate the slope we simply select a region before and after the trigger and average the time and current in each region. The slope is then the change in mean current over the change in mean time.

$$\hat{m} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0} \tag{8.9}$$

The information that this estimator gives us will depend on our choice of pre-trigger and post-trigger region. Here we define the region using the first 9.92 ms of the pulse and last 5 ms. This definition was chosen so that there was very little change in the slope with energy in the unsaturated region. Notice, under

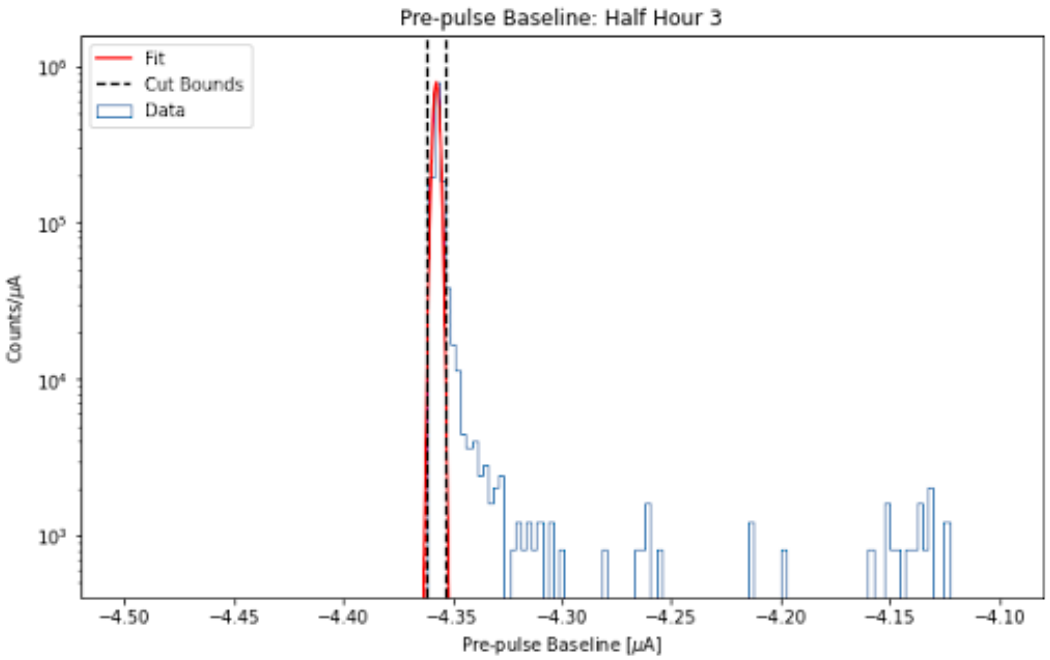


Figure 8.8: Example fit to a half-hour period to define outliers in the baseline distribution. A Gaussian (red) is fit to the most prominent peak in the distribution. Events within 3σ of this fit for this half hour are preserved by the cut. We repeat this definition for each half hour of collected data to generate the cut for the entire data series.

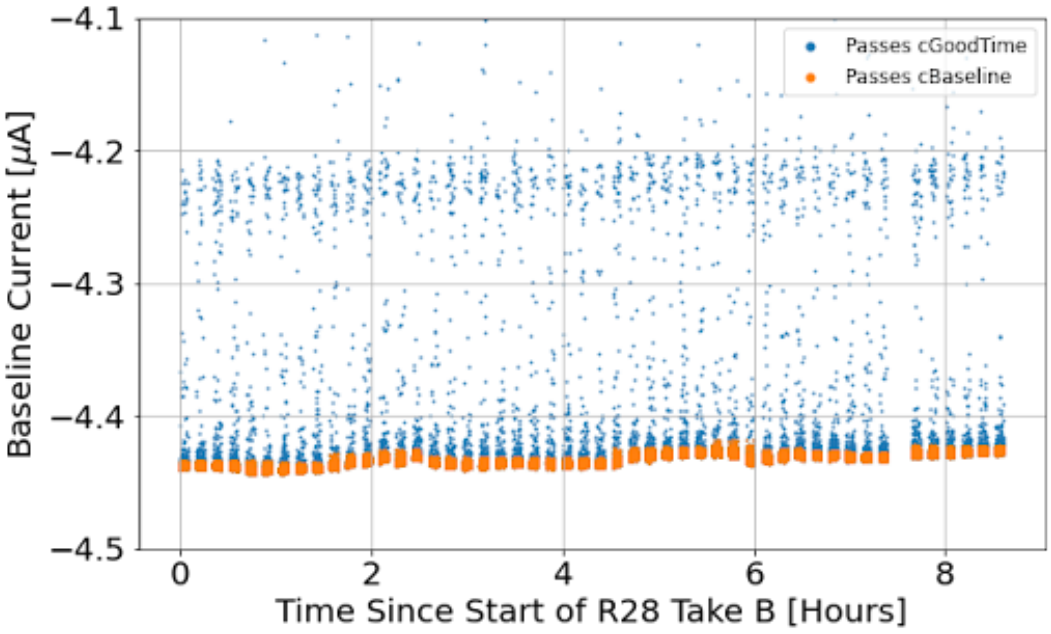


Figure 8.9: Application of the baseline cut to the data series. Events in orange pass the cut. The cut preserves a consistent baseline within $\sim 5\%$. A broad distribution in elevated baseline can be observed above the region preserved by the cut, which are simply pilep events from the calibration source which happen to fall on the prepulse baseline.

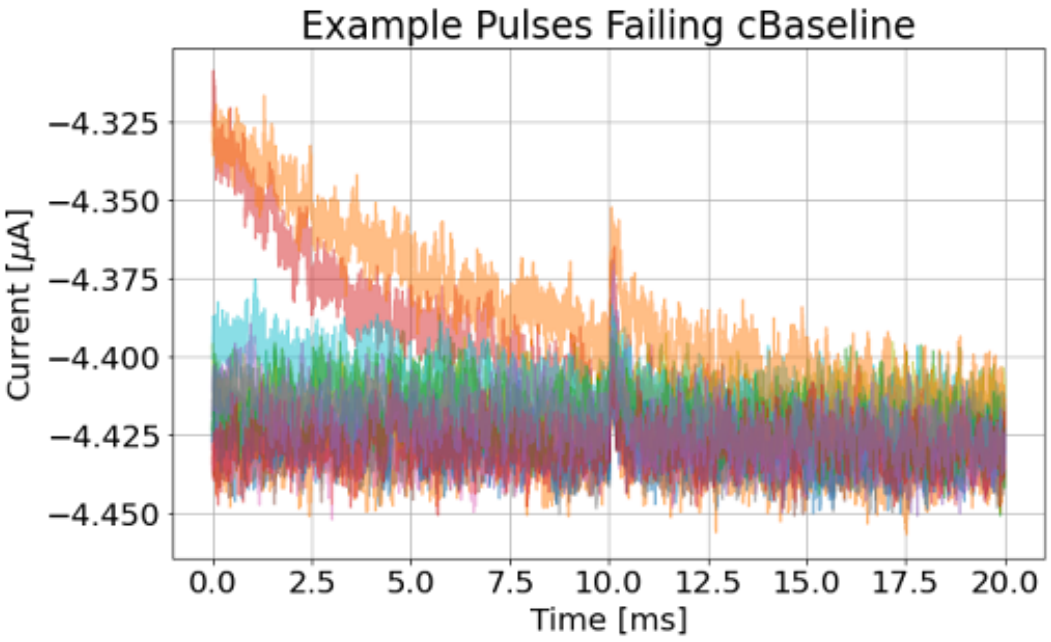


Figure 8.10: Collection of example pulses removed by prepulse baseline cut. The largest population removed are events reconstructed on the tails of high energy events.

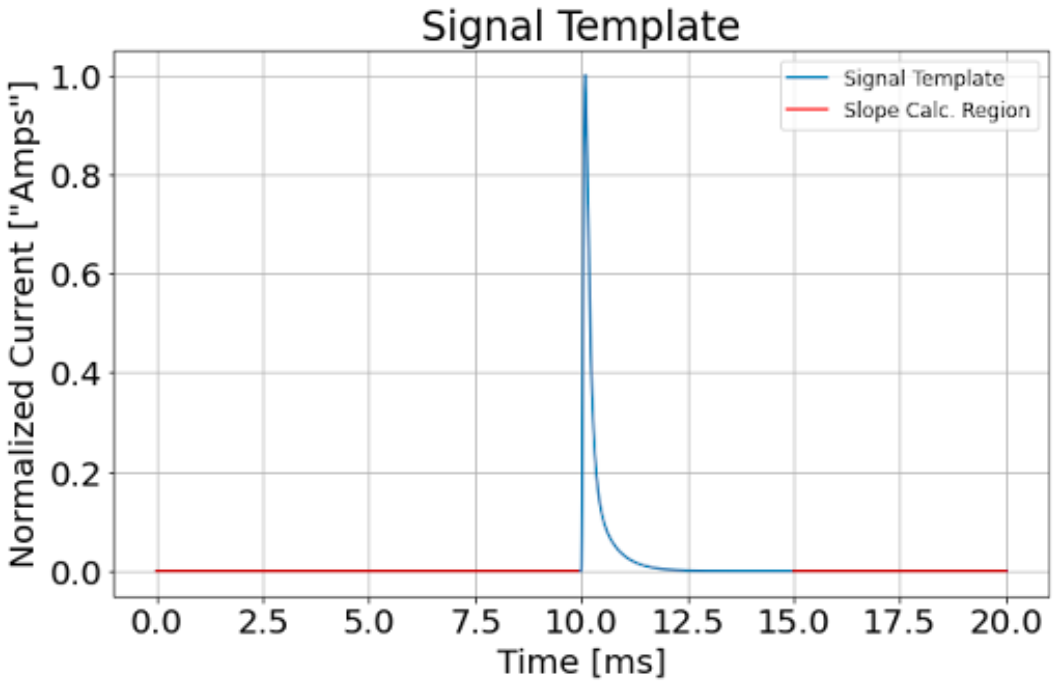


Figure 8.11: Signal template with region used to define pulse slope highlighted. The slope is calculated as the simple "rise over run" determined by averaging over each region.

the assumption that the observed trace is well described by the template, slope scales linearly with the pulse amplitude.

In order to determine our cut bounds for the general slope cut, we look at the slope of randoms (after the pre-pulse baseline cut is applied) and fit a gaussian to this distribution. We find the sigma from this fit is 0.092 uA/s, and 96.5% of randoms which pass the baseline cut pass a 3-sigma cut.

The simplest cut definition for triggered events would be to use these bounds and remove any events with an outlying slope. As saturation occurs at higher energies, the post-pulse region is typically elevated above the template. So if we take no energy dependence into account for our cut definition, the cut will remove saturated events. In Figure 8 we see how the slope changes with energy. We want to extend our cuts to preserve the 55Fe peaks so that we can compare our calibration between different fridge runs. We perform an energy cut to select the peaks, then fit a Gaussian to the slope of these events. Then define a linear correction to the slope as a function of energy. Even events at very low energy don't follow this scaling, we see from the figure that the unsaturated regime is narrow on this energy scale, and our acceptance region is wide enough to include the main band of unsaturated events. At higher energies, the slope scales quadratically with the

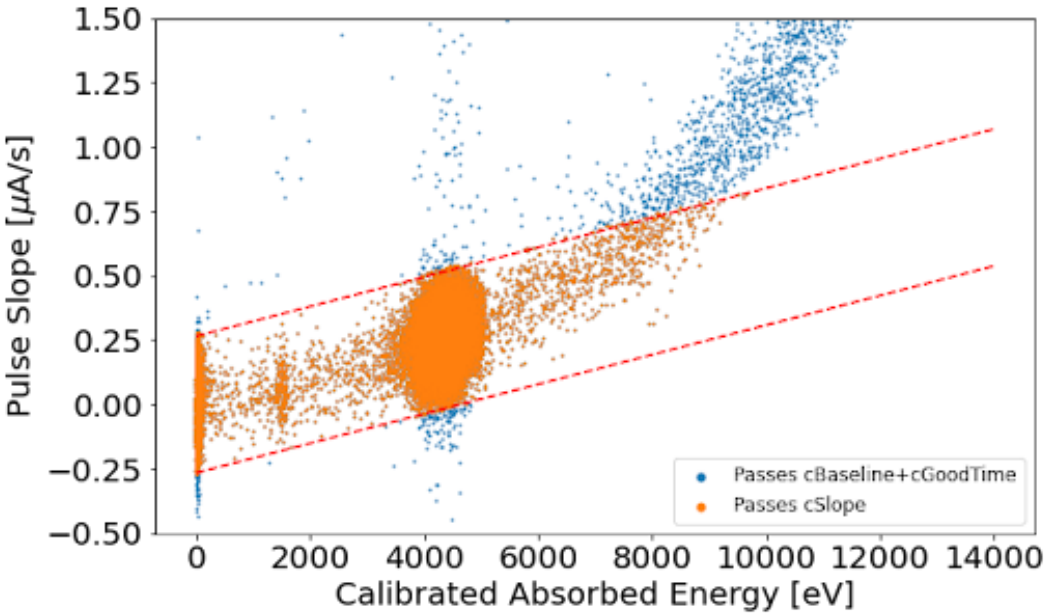


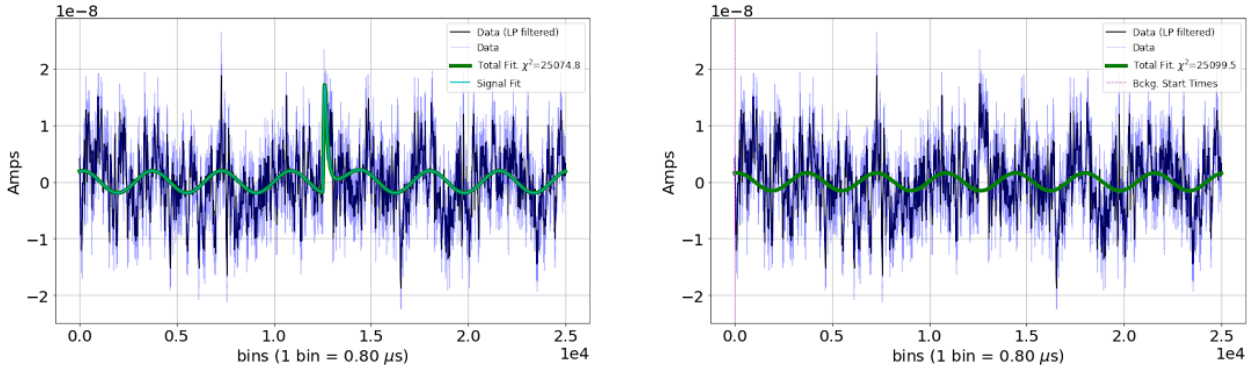
Figure 8.12: Distribution of slope parameter with respect to integrated energy. Highlighted orange events in the region between dashed red lines are events which pass the cut. The cut is constructed to preserve the ^{55}Fe calibration peak.

absorbed energy, but we don't make any attempt to include these events in our spectrum.

8.3.4 $\Delta\chi^2_{sine}$ Cut

It was observed during UMass Run 28 that there seems to be excess noise at ∼350 Hz that may be coupling into the sensor/detector. These events first appeared in example traces removed by the cGoodTime and cBaseline cuts. The sensitivity to cGoodTime indicated that this noise was not stationary in time, but rather coupled and decoupled to our detector intermittently. However, it was determined that the timescale of this behavior was too short for us to resolve with cGoodTime. The sensitivity to cBaseline is interesting because it suggests that the observed oscillations are not symmetric about the baseline. However the amplitude of the oscillations is typically consistent with noise fluctuations, and hence can't be reliably with this cut. With these considerations, it was determined that we should perform fits to the 350 Hz background in addition to our signal template and then compare the resulting χ^2 values to remove these events.

Using the infrastructure of the NsMb (N-signals, M-backgrounds) optimum filter in QETpy, the pulse template was used as the "signal" template, and the background templates were made as a cosine and a sine wave with frequency of 350 Hz. The NsMb optimum filter in QETpy is capable of time shifting the



(a) (b)

Figure 8.13: Example of a pulse removed by the $\Delta\chi^2_{sin}$ cut. Left plot shows the signal+background fit to the pulse, and right shows the background only fit. The cut is defined by comparing the difference in χ^2 resulting from each fit.

signal template, but not the background templates, so by defining the background as a sine and a cosine, it is possible to effectively time shift the templates through the relative phases of the sine and cosine.

Then pulses which satisfy the following inequality are defined to pass the cut:

$$\chi^2_{pulse} - \chi^2_{sine} + 1000A^2 - 10A^4 - 10A^6 < 10 \tag{8.10}$$

Where χ^2_{pulse} is the unconstrained χ^2 of the signal, and A is the unconstrained optimum filter signal amplitude in μA. Note that only the quadratic terms are relevant to our region of interest, and the higher order terms are corrections to include the calibration peaks. To qualitatively understand the cut, we just need to recall that a smaller χ^2 indicates a better fit. So a larger value of $\chi^2_{pulse} - \chi^2_{sine}$ indicates that the event is better described by the 350 Hz background, and hence should be removed by the cut.

8.3.5 χ^2_{LF} Cut

In order to remove any anomalous pulse shapes which happen to pass our previous cuts, we employ a cut on the low frequency χ^2 (evaluated over frequencies <10 kHz). We don't include the higher frequency components when evaluating the quality of the fit as they are sufficiently above our signal bandwidth, which is determined by our pulse fall time. Also note that we define the goodness of fit where the signal template is not constrained in time, which will remove events where a glitch occurs anywhere in the trace. Then pulses

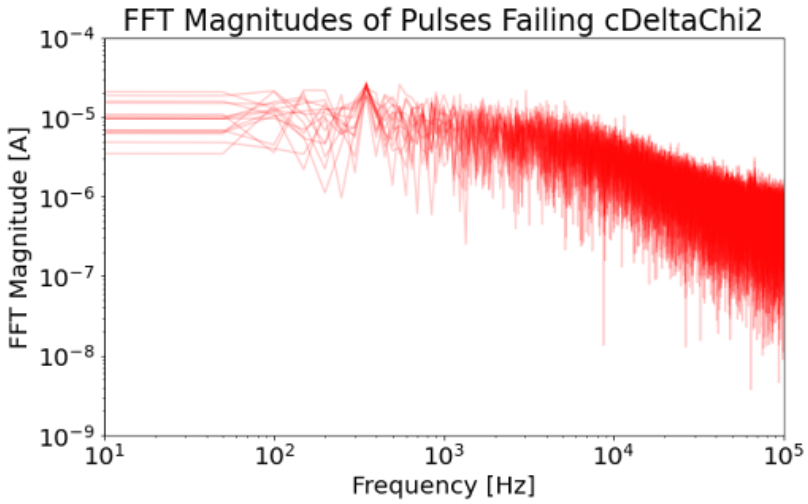


Figure 8.14: Comparison of FFTs from events which fail the $\Delta\chi^2_{sin}$ cut. We see that removed events have a simliar amplitude in the 350 Hz bin, regardless of their other frequency content.

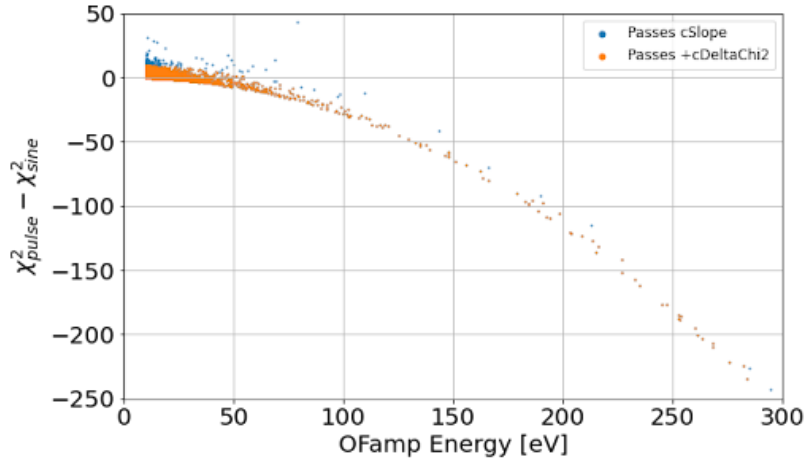
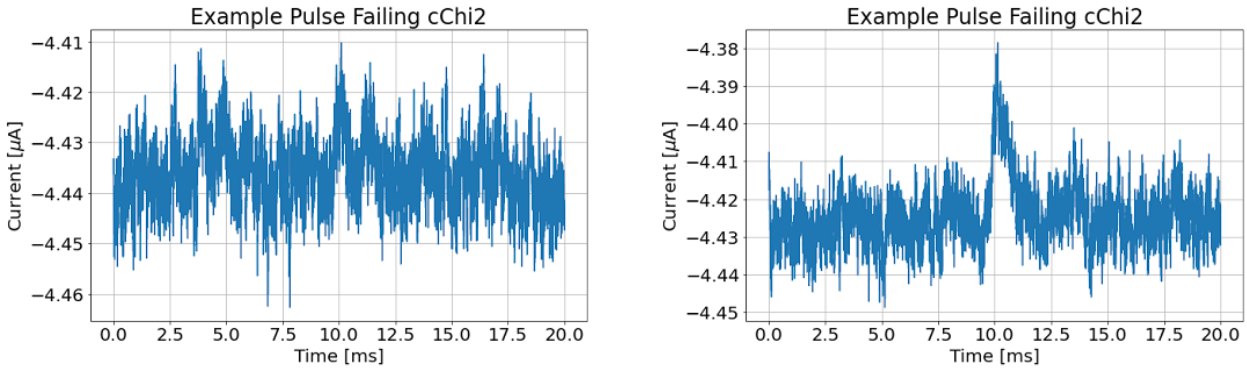


Figure 8.15: Application of $\Delta\chi^2_{sin}$ cut to the data. Events in orange pass the cut. The cut boundary is parabolic, which follows the distribution with energy. The bulk of events removed occur at low energy.



(a) (b)

Figure 8.16: Example of pulses removed by the χ^2_{LF} cut. The pulse on the left appears to contain a train of "two humped" glitches. The pulse on the right is also identified as a glitch, as its rising edge is much slower than the signal template.

which satisfy the following inequality are defined to pass the cut:

$$\chi^2 < 500 + 1000A^4 + 100A^6 \tag{8.11}$$

Where A is the unconstrained OF amplitude in uA. With a low frequency cut off of 10 kHz and a deltaF of 50 Hz, the center of the distribution in the limit of small amplitude should be at about 400, and the chosen cutoff of 500 in this regime corresponds to $\sim 3.5\sigma$.

8.3.6 Noise Cuts

In addition to the cuts described above, an additional two cuts are used to produce the noise PSD. These cuts remove events that show evidence of a pulse, and evidence of a 350 Hz sine wave, respectively termed the "no-pulse" and "no-sine" cut. The χ^2 of either the pulse template or the sine wave template are calculated, and the $\Delta\chi^2$ between those and the noise only χ^2 is taken. For the no-pulse cut, the cut line is set at the 40th-percentile of the $\Delta\chi^2$ distribution. This is to say, we automatically remove this fraction of random triggers which look most pulse-like before constructing the PSD. This is akin to removing large noise fluctuations, and we are in some sense purposefully underestimating our noise. For the no-sine cut the cut line is set at the 1st-percentile of the $\Delta\chi^2$ distribution, which the no-sine cut lowers the peak in the PSD observed at 350 Hz. After applying these cuts in addition to the other cuts described above, the noise PSD is calculated. The noise PSDs for Run 28 is shown in Fig. 8.18.

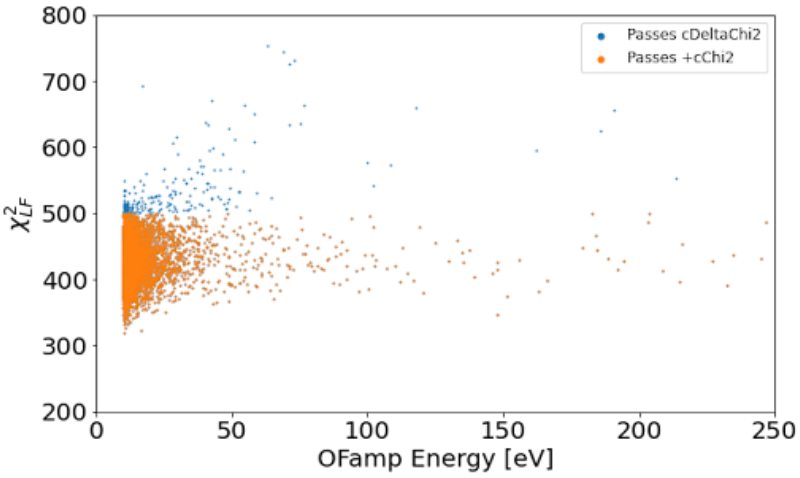


Figure 8.17: Application of the χ^2_{LF} cut to events passing subsequent cuts. At low energy, the cut is characterized by a flat threshold at $\chi^2_{LF} < 500$.

Since these cuts affect the PSD, which is a basic input of both our triggering and reconstruction algorithms, these can have subtle and wide reaching effects on our result. We discuss these implications in greater detail in 8.6.2, but the upshot is that lowering our PSD gives us a better understanding of how noise appears in our spectrum at the cost of degrading our resolution. This trade-off is desirable for this analysis, as one of the main thrusts is to better understand the interplay between the noise distribution and our signal.

8.4 Pulse Simulation and Signal Model

In order to perform a dark matter search, we need an accurate description of the spectrum a dark matter signal will produce in our detector. The differential spectrum of recoil energy imparted by a galactic DM particle on a terrestrial detector is well understood following the analysis based off of [5], and is more or less universal for any detector technology. Any detector used to search for DM on the other hand will have a unique response to a particular deposited energy depending on the noise conditions during the search and the trigger algorithm used. This effect is especially relevant when we're analyzing events which are near the trigger threshold of our detector. For example, we could consider that a DM particle deposits an energy just below the threshold required to issue a trigger, but which happens to be in coincidence with a noise fluctuation which boosts the observed amplitude above threshold, an effect termed "noise boosting". We expect to observe these events in our spectrum, and hence we should account for them in our analysis.

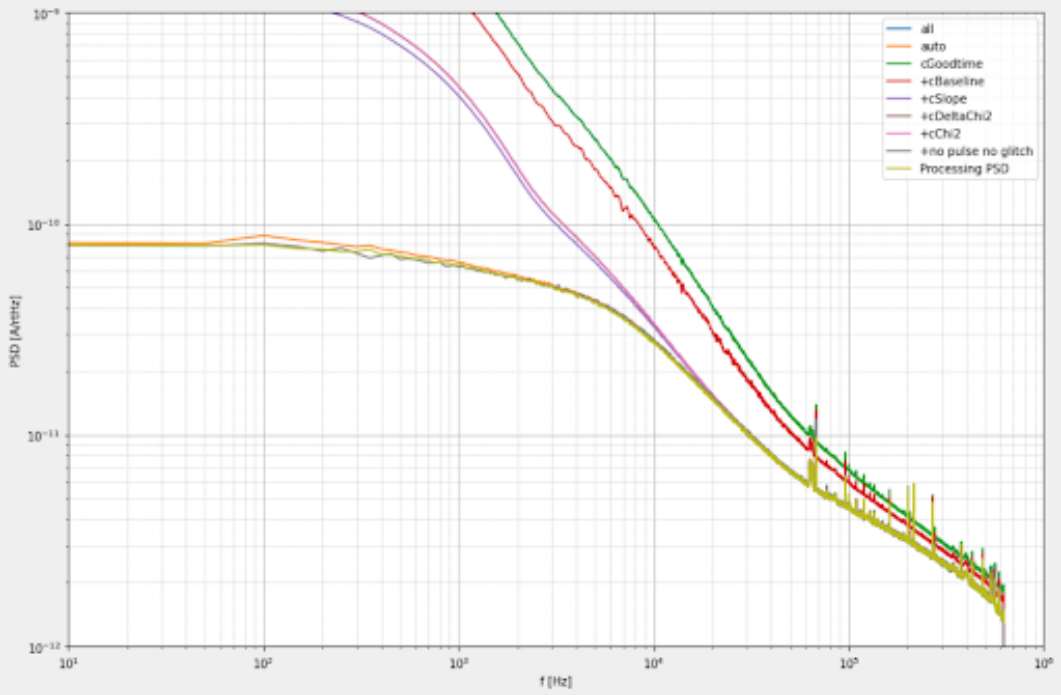


Figure 8.18: Resulting PSD after applying each cut sequentially. The Effect of the slope, baseline and χ^2 cuts is mostly to remove pileup, and hence they make the PSD look less "pulse-like". After applying our noise cuts, we arrive at a PSD which looks similar to the one determined by the QETpy.autocuts() algorithm [119], but with slightly less low frequency noise.

On the other hand, if we don't do this carefully we risk not being conservative in our analysis by claiming sensitivity to DM masses lighter than we're entitled.

For an example of how we could fail to be conservative, consider the approach used in [120]. To characterize the detector response, we define the function $f(E'|E_{in})$ as the probability to reconstruct a true energy deposition E_{in} at energy E'. Under the simplifying assumption that the template is not allowed to shift in time while reconstructing the data, the ideal detector response is simple Gaussian smearing

$$f(E'|E_{in}) = \frac{1}{\sqrt{2\pi\sigma_E^2}} \exp\left((\frac{E' - E_{in}}{\sqrt{2}\sigma_E})^2 \right) \tag{8.12}$$

Where σ_E is the baseline energy resolution. Then the expected signal is

$$\frac{dR}{dE'} = \epsilon(E') \int_0^\infty f(E'|E_{in}) \frac{dR}{dE_{in}} dE_{in} \tag{8.13}$$

where $\epsilon(E')$ is the energy dependent signal acceptance efficiency, and dR/dE_{in} is the rate of true energy depositions in the detector. Consider a simple case where the DM is bosonic and deposits all its energy $E_{DM} = m_\chi$ in the detector via absorption.

$$\frac{dR_B}{dE_{in}} = R_0 \delta(E_{in} - E_{DM}) \tag{8.14}$$

Now the signal expected to be reconstructed in the detector is

$$\frac{dR_B}{dE'} = \frac{\epsilon(E')R_0}{\sqrt{2\pi\sigma_E^2}} \int_0^\infty \exp\left((\frac{E' - E_{in}}{\sqrt{2}\sigma_E})^2 \right) \delta(E_{in} - E_{DM}) dE_{in} \tag{8.15}$$

$$= \frac{\epsilon(E')R_0}{\sqrt{2\pi\sigma_E^2}} \exp\left((\frac{E' - E_{DM}}{\sqrt{2}\sigma_E})^2 \right) \tag{8.16}$$

Then consider the limit as $E_{DM} \to 0$

$$\lim_{E_{DM}\to 0} \frac{dR_B}{dE'} = \frac{\epsilon(E')R_0}{\sqrt{2\pi\sigma_E^2}} e^{-E'^2/2\sigma_E^2} \tag{8.17}$$

Now, there is some threshold E_T baked into $\epsilon(E')$, and if is is sufficiently large it can make the rate vanishingly small, but the crucial observation is that we do not expect a non-zero rate above threshold in the limit of DM with zero mass. This is actually somewhat intuitive. If we turn the detector on and wait long enough,

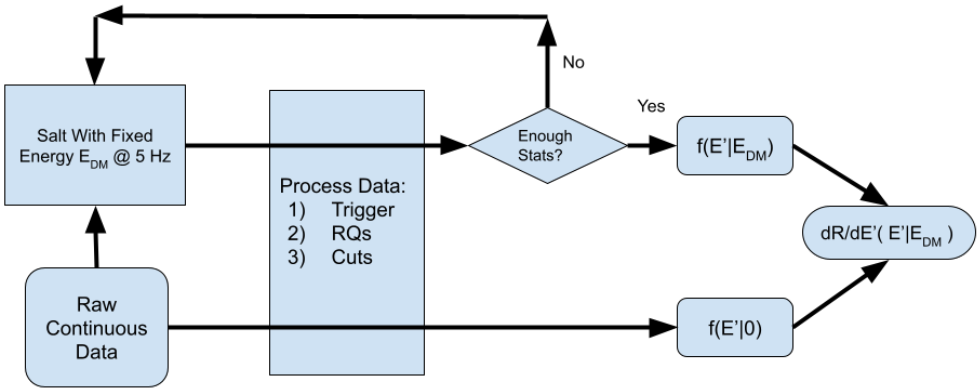


Figure 8.19: Flowchart outlining the salting process and construction of the signal model. A copy of our raw data stream is made before the salt is simulated. Then the salted and raw data are processed in parallel, subjected to the same sequence of triggering, energy reconstruction and quality cuts. At the end of this process, we compare the two spectra to determine our detector response to the injected signal.

we expect some small but nonzero rate of noise fluctuations to appear above threshold - but it would be erroneous to attribute this rate to the DM signal.

8.4.1 Data Salting Process

The key difference between the method used for this analysis and similar ones used in past DM searches is the observation that if the DM signal is in coincidence with a noise fluctuation (or background event for that matter), this not only implies a increased rate in the the observed spectrum at some energy higher than that deposited by the DM, but also a decrease in the estimate of the noise. That is, we attempt to replace the detector response function considered in Eq. 8.12 with one which is aware of the noise distribution

$$f(E'|E_{in}) \rightarrow f(E'|E_{in}) - f(E'|0) \tag{8.18}$$

As we will argue in the next section, this additional term guarantees that our total expected rate goes to zero as $E_{in} \rightarrow 0$.

with this, our problem becomes one of determining $f(E'|E_{DM})$ and $f(E'|0)$. To do this, we inject simulated signals, dubbed "salt", into our data at energy E_{DM} and then compare the resulting spectrum to the one derived from the unsalted data. Since we were able to read out the data continuously for the analysis and trigger offline, the salting itself is relatively straightforward. All we need to do is take the signal template, scale it by the desired energy and then randomly inject it into the continuous data stream.

Once we've salted the data, we put it through the same process of triggering, calculating reduced quantities[1] (RQs) and applying cuts that we do for the unsalted data stream. This ensures that any difference between the two resulting spectra is due to the salt and not an artifact of processing.

The salt is injected at random locations with an average frequency of 5 Hz. This rate is somewhat arbitrary, but chosen to be comparable to the net trigger rate in the unsalted data. A reasonable question to ask is how this rate compares to the rate of DM interacting in our detector. Of course, this can't be determined unless we assume a particular cross section. The noise boosting formalism presented here is well behaved when the DM rate is sufficiently low that we don't expect significant pile-ups from DM events. Note that this is also our ideal operating conditions. We would like to have as low a background as possible, which will give us sensitivity to DM models with smaller cross sections and hence smaller total rates. However, this analysis is above ground and sees a large background, so we expect to only be sensitive at larger cross sections and can't guarantee that the pile-up rate is small. In the grand scheme of things, this isn't really a problem. The main takeaway of this analysis is that CPD has sensitivity to DM at lower masses than similar detector technologies have yet been able to achieve. If we can put the detector in a low background environment, the sensitivity to smaller cross sections will follow. In the short term, this is a bigger problem. We want to actually draw an exclusion curve for this experiment, and to do so we need to really understand all the implications of our analysis. The proper way to extend the noise boosting formalism to the high rate limit is still under development at the time of this thesis, and will be discussed in detail in an upcoming publication [121]. For now, we will treat this as a caveat. The methods presented here only strictly apply to the low rate limit, but we expect the high rate case to be closely analogous.

8.4.2 Relation Between Salting And Efficiency

For now, let's pretend we're searching bosonic for dark matter with mass M_{DM}. The true energy deposition in this model will always be equal to the dark matter mass $E_{DM} = M_{DM}$. Then if the dark matter interacts with our detector at an average rate R_{DM}, we expect to observe the spectrum

$$\frac{dR}{dE'}(R_{DM}, M_{DM}) = \frac{dR}{dE'}(0,0) + R_{DM}\big[f(E'|E_{DM}) - f(E'|0)\big] \tag{8.19}$$

[1] e.q. energy estimators, pulse shape parameters.

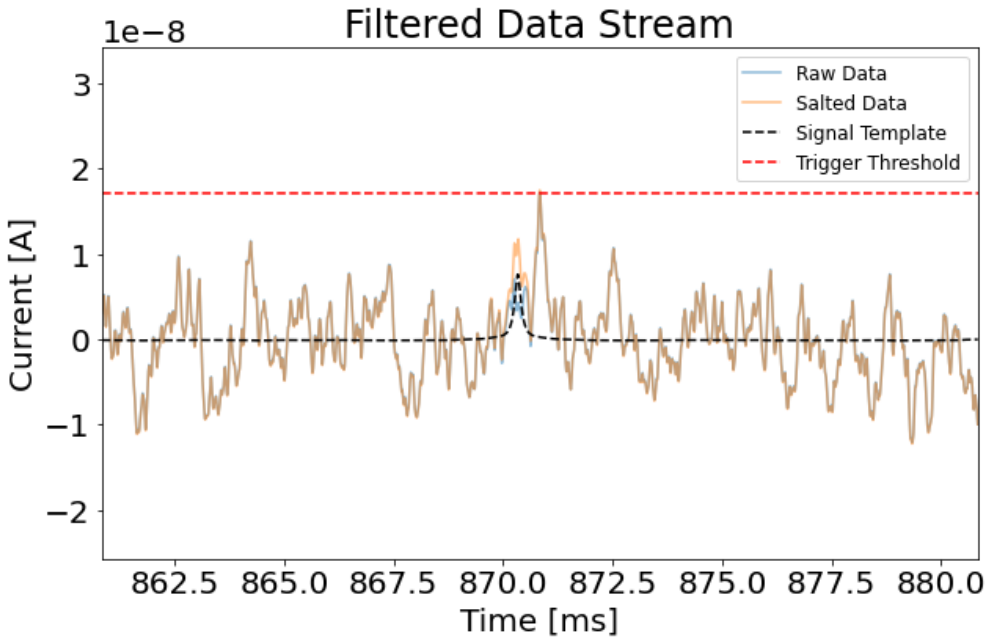


Figure 8.20: Example of a filtered trace before and after salt is injected. The trace is shown in blue before salting, and orange after. The dashed black line shows the filtered, injected signal event, and there is a small deviation between the orange and blue curves where this occurs. The red-dashed, horizontal line represents the trigger threshold. This particular event is just above threshold before the salt is injected.

Where $\frac{dR}{dE'}(0,0)$ is the spectrum we would observe in the absence of DM interactions. We make the assumption that there is no DM present in our background. This is counter intuitive, because the motivation for performing a DM search is that we suspect it could appear in our spectrum. But this assumption is very computationally convenient as it allows us to directly estimate $f(E'|0)$ from our data. Moreover, it is conservative, since any DM signal which is present will simply be treated as an elevated background in this formalism. Under this assumption we have

$$f(E'|0) = \frac{1}{R_{DM}} \frac{dR}{dE'}(0,0) \tag{8.20}$$

When we salt our data with events at a known energy and rate, we are simulating the dark matter scenario described above. Hence when we analyze our spectrum after salt is injected, we are effectively estimating

$$f(E'|E_{DM}) = \frac{1}{R_{DM}} \frac{dR}{dE'}(R_{DM}, M_{DM}) \tag{8.21}$$

Now let's consider how this formalism relates to our estimate of the efficiency. We define the total signal acceptance efficiency as the fraction of good signal events (e.g. the injected salts) which are issued a trigger and pass all the quality cuts

$$\epsilon(E_{DM}) \equiv \frac{n(E_{DM})}{N(E_{DM})} \tag{8.22}$$

When we compare the salted to the unsalted data, we know that the only difference can be to the injected salt, and that the injected salts represent good signal events. Hence the number of accepted signal events is given by difference between the integral of salted and unsalted spectra

$$n(E_{DM}) = X \int_{E_T}^{\infty} dE' \big[\frac{dR}{dE'}(R_{DM}, M_{DM}) - \frac{dR}{dE'}(0,0) \big] \tag{8.23}$$

Where E_T is our trigger threshold and X is the exposure of our search. Then using equation (1) we have

$$n(E_{DM}) = X R_{DM} \int_{E_T}^{\infty} dE' \big[f(E'|E_{DM}) - f(E'|0) \big] \tag{8.24}$$

And the number of injected signal events is simply $N = X R_{DM}$, so altogether we have

$$\epsilon(E_{DM}) = \int_{E_T}^{\infty} dE' \big[f(E'|E_{DM}) - f(E'|0) \big] \tag{8.25}$$

So we see that the function $f(E'|E_{DM}) - f(E'|0)$ is closely related to our total signal acceptance efficiency, but they are not identical. Rather it is a kind of *derivative* of the efficiency. We also see that in the framework of our salting method, it is straightforward to evaluate the signal efficiency by comparing the integrated rate of our spectrum before and after the salt is injected.

As we will see soon, sometimes $f(E'|E_{DM}) - f(E'|0) < 0$. If we directly identify $f(E'|E_{DM}) - f(E'|0)$ as the efficiency, this is clearly concerning as a negative efficiency is non-physical. However, the above argument shows that we don't have to worry about a negative efficiency as long as the integral quantity is positive.

8.4.3 Signal Efficiency

Once we've salted the data at a fixed set of energies, we can evaluate $f(E'|E_{DM}) - f(E'|0)$ for each salted data set. An example is shown in 8.21a where we've injected the salts at $E_{DM} = 20$ eV. Note that the regions where $f(E'|E_{DM}) - f(E'|0) < 0$ on the log-scale are represented by the shaded bins. From Eq. 8.25, we know what we're really interested is the integral of this expression with respect to E' at an arbitrary E_{DM}. Clearly we can't do this in general since the values of E_{DM} are fixed when we carry out the simulation, so we need a way to interpolate between intermediate E_{DM} values. To this end, we define the interpolation

$$f(E'|E) - f(E'|0) = f(E' - \delta E|E - \delta E) - f(E' - \delta E|0) \tag{8.26}$$

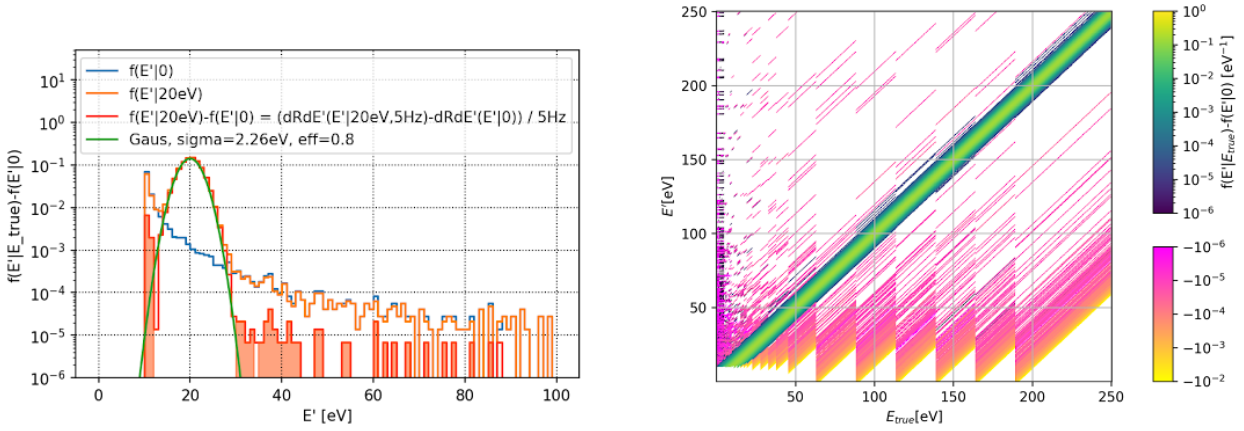
Where δE is the difference between E and the nearest salting energy used in the simulation. The result of this smearing is shown in 8.21b, with the lower color bar indicates negative amplitudes on the log-scale.

After the interpolating map is defined, we're free to carry out the integration to derive the signal efficiency. The result of this is shown in 8.22, where we've also shown how each successive cut affects the curve. We see that at high energies we have an approximately constant efficiency of $\epsilon \approx 0.8$, and that the efficiency reaches 50% around our trigger threshold of 10 eV.

8.5 NRDM Sensitivity

8.5.1 Can You Do Better With a Worse Detector?

Before we proceed with using the noise boosting formalism to estimate the sensitivity of our DM search, we should take a moment to make sure that our proposal is sane. The prospect of claiming sensitivity to



(a) (b)

Figure 8.21: (Left) Comparison of the spectrum before and after salt is injected at an energy of 20 eV. Red histogram represents the difference of the spectra, which determines our efficiency. Shaded bins indicate a negative amplitude on the log-scale. (Right) Resulting efficiency map from interpolating between discrete salting energies. The yellow-pink gradient represents negative amplitudes.

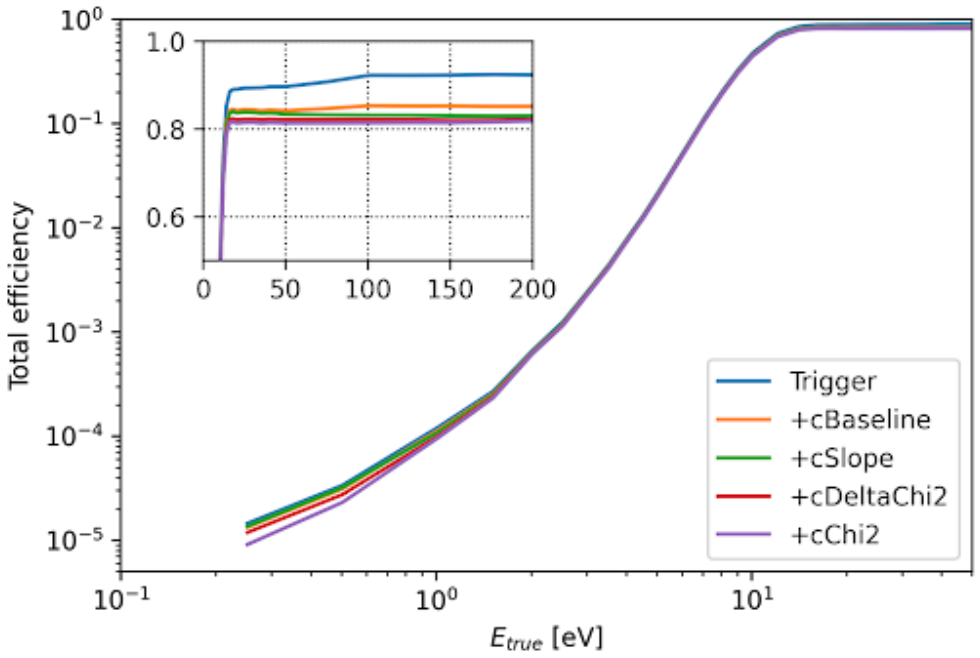


Figure 8.22: Final efficiency curve, shown after each cut is applied. The final efficiency curve (purple) resembles an error function, which is what we roughly expect. Above the ~ 10 eV trigger threshold, we have an approximately constant efficiency of 81%.

events which lie below your trigger threshold warrants a bit of skepticism at first glance. The idea is that we gain this sensitivity from the fact that we expect occasional coincidences between noise fluctuations and signal events. If the noise fluctuation is large enough, it can reasonably push a real physics event which would otherwise not be triggered on above our trigger threshold. Given this line of reasoning, one might expect that one could maximize your sensitivity to events below threshold by increasing the size of our noise fluctuations. In other words, we might suspect that we could actually get more sensitivity to LDM by operating a detector with worse resolution. This is clearly paradoxical, especially once we consider that our initial motivation for performing a LDM search with CPD was its excellent resolution. Here we will argue that this is actually not the case. Our noise boosting implementation will always result in worse sensitivity when we have a worse resolution.

To see this, we start by integrating Eq. 8.19 above our trigger threshold. Keep in mind that this assumes we are looking for a DM model which produces a mono-energetic recoil in our detector, but easily generalizes to more complex models. This gives us the expected rate of events in our spectrum

$$R = \int_{E_T}^{\infty} dE' \frac{dR}{dE'}(R_{DM}, M_{DM}) = R_0 + R_{DM} \int_{E_T}^{\infty} dE' \big[f(E'|E_{DM}) - f(E'|0) \big] \tag{8.27}$$

Where R_0 is the hypothetical rate we would observe if there were no DM present in our detector. Then we compare this to Eq. 8.25, and we find

$$R = R_0 + \epsilon(E_{DM}) R_{DM} \tag{8.28}$$

This result is obvious in hindsight. The rate we expect to see is just the background rate plus the DM interaction rate scaled by our signal efficiency. Furthermore let us recall that our sensitivity scales with R - the more signal events we expect to see, the better chance we have of observing DM. Next, we consider an idealizing assumption about our efficiency - that it takes the form

$$\epsilon(E) = \frac{C}{2} \left[1 + \text{erf}\left(\frac{E - E_T}{\sqrt{2}\sigma_E} \right) \right] \tag{8.29}$$

Where σ_E is our baseline energy resolution, and $0 < C < 1$ is the high energy limit of the efficiency. Though the efficiency doesn't strictly need to follow this form, it is a common model and we see in (relevant figure) that it is a good approximation to our data. Additionally, let us assume that we define the trigger threshold

E_T with respect to our resolution rather than setting it to a fixed value.

$$E_T \equiv n\sigma_E \tag{8.30}$$

Where n is chosen by the experimenter. Typically $n \sim 4-6$, and in the case of CPD we have $n = 4.5$.

With this, we can evaluate how we expect the number of observed events to change with the resolution

$$\frac{\partial R}{\partial \sigma_E} = R_{DM}\frac{\partial}{\partial \sigma_E}\epsilon(E_{DM}) \tag{8.31}$$

$$= -R_{DM}\frac{C}{2}\frac{E_{DM}}{\sigma_E^2}e^{n/2}e^{-E_{DM}/(2\sigma_E)} < 0 \tag{8.32}$$

From this we see that increasing (i.e. making it coarser) our resolution will always decrease our observed signal event rate. Hence we cannot gain better sensitivity in our DM search by operating with worse resolution, and our salting framework passes the proposed sanity check.

8.5.2 Sensitivity Estimation

Mass Reach

Our salting procedure assumes that the DM spectrum is mono-energetic for ease of computation, but we know that DM interacting through elastic NRs in our detector will produce an approximately exponential spectrum. To find the expected reconstructed rate for a generic dR_{DM}/dE, we can convolve the signal with our efficiency map $f(E'|E_{DM}) - f(E'|0)$. Comparing to 8.19 we have

$$\frac{dR}{dE'}(E'|M_{DM}) = \frac{dR}{dE'}(E'|0) + \int \frac{dR_{DM}}{dE_{DM}}(E_{DM}|M_{DM})\big[f(E'|E_{DM}) - f(E'|0)\big]dE_{DM} \tag{8.33}$$

To illustrate we evaluate this for several DM masses at a fixed cross section and find the total integrated rate. We also compare to the case where we don't apply any smearing, and simple Gaussian smearing. This is shown in Fig. 8.23. Notice that the result of this analysis nicely interpolates between the Gaussian and no-smearing cases. Noise boosting gives us sensitivity to DM masses below what would normally be kinematically accessible, but there is a definite cutoff that is not present in the Gaussian case. We see that we can claim sensitivity to DM with masses $m_\chi \sim 20$ MeV, compared to $m_\chi \sim 100$ in the unsmeared case.

To complete our sensitivity estimate, we must take the differential rate derived from Eq. 8.33 and compare

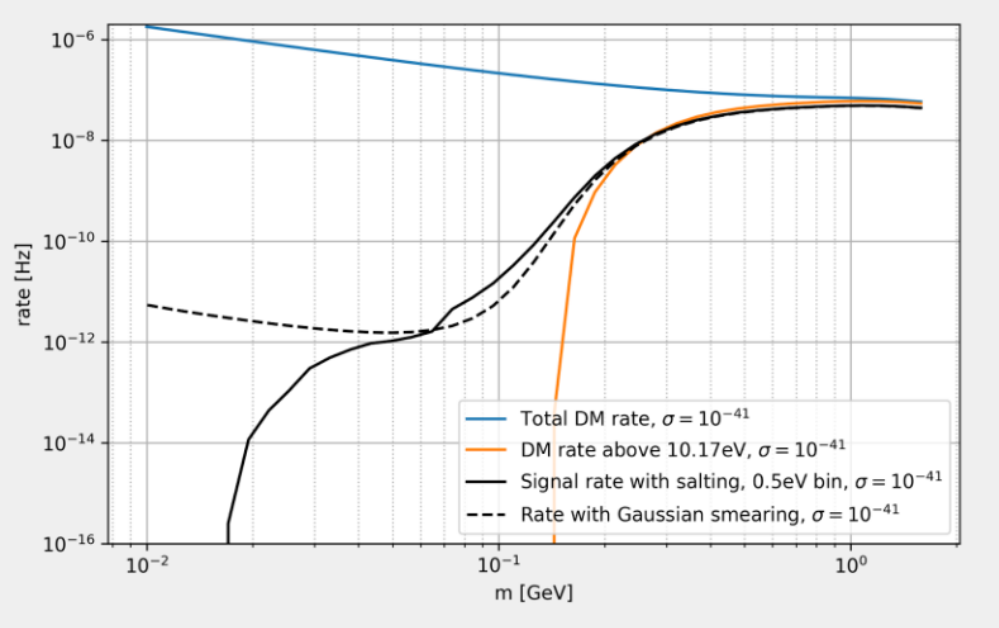


Figure 8.23: Integrated DM signal rates using our noise boosting model. The blue curve shows the total expected rate in the limit of a zero-energy trigger threshold, and the orange curve shows the expected rate above our $\sim$ 10 eV trigger threshold without any smearing. The dashed black curve shows the result of simple Gaussian smearing, which would imply sensitivity to arbitrarily small DM masses. The solid black curve is the result of applying noise boosting to this analysis, and interpolates between the orange and black-dashed curves. It imposes a definite mass cutoff at $m_\chi \sim 20$ MeV, which is below the cutoff without smearing.

it to our background. Using the Optimum Interval (OI) method, this comparison yields a lower bound on the DM-nucleon cross section at a given mass [122]. Some care must be taken when making this comparison, in that our smearing the efficiency map $f(E'|E_{DM}) - f(E'|0)$, which has negative fluctuations, with the signal model can result in negative fluctuations in the derived signal. These negative fluctuations ultimately arise from the finite statistics in our original spectrum. If a particular bin in the histogram of the observed background happens to have more events closer to one edge than the other, the injected salt is more likely to push these events to the neighboring bin. Subtracting the binned unsalted spectrum from the salted one then systematically features a negative amplitude, which are clearly non-physical and cannot be resolved by injecting more salt events. The OI method assumes that the signal model is positive everywhere. To rectify this with any negative fluctuations in our empirical signal model, we set the negative amplitudes to zero by hand. This is conservative, since we don't expect to gain any sensitivity by setting the expected signal to zero in a particular region. In fact, this is actually in the spirit of the OI method – since it prefers to select a region where the expected signal is low compared to the background and the background can never be negative.

An implication of the above argument is that our signal model depends on our choice of binning. If there are non-uniform fluctuations within a given bin, we could consider using wider bins to average these out. To quantify the effect of our bin choice on the resulting sensitivity, we repeated the signal construction and limit setting procedure with the both 1 eV and 0.5 eV bins. This resulted in sensitivities which agreed within 20%, with the larger bin size being slightly more conservative at low masses. As the low mass region is the most interesting for the DM search, we opted to use the 1 eV bin size for the final analysis.

Exclusion Limit

To complete our sensitivity estimate, we must take the differential rate derived from Eq. 8.33 and compare it to our background. Using the Optimum Interval (OI) method, this comparison yields a lower bound on the DM-nucleon cross section at a given mass [122]. Some care must be taken when making this comparison, in that our smearing the efficiency map $f(E'|E_{DM}) - f(E'|0)$, which has negative fluctuations, with the signal model can result in negative fluctuations in the derived signal. These negative fluctuations ultimately arise from the finite statistics in our original spectrum. If a particular bin in the histogram of the observed background happens to have more events closer to one edge than the other, the injected salt is more likely to push these events to the neighboring bin. Subtracting the binned unsalted spectrum from the salted one

then systematically features a negative amplitude, which are clearly non-physical and cannot be resolved by injecting more salt events. The OI method assumes that the signal model is positive everywhere. To rectify this with any negative fluctuations in our empirical signal model, we set the negative amplitudes to zero by hand. This is conservative, since we don't expect to gain any sensitivity by setting the expected signal to zero in a particular region. In fact, this is actually in the spirit of the OI method – since it prefers to select a region where the expected signal is low compared to the background and the background can never be negative.

An implication of the above argument is that our signal model depends on our choice of binning. If there are non-uniform fluctuations within a given bin, we could consider using wider bins to average these out. To quantify the effect of our bin choice on the resulting sensitivity, we repeated the signal construction and limit setting procedure with the both 1 eV and 0.5 eV bins. This resulted in sensitivities which agreed within 20%, with the larger bin size being slightly more conservative at low masses. As the low mass region is the most interesting for the DM search, we opted to use the 1 eV bin size for the final analysis.

The derived 90% exclusion curves are shown in 8.25. We see that our sensitivity in terms of cross section is relatively high at $\sim 10^{-30}$ cm^2, we have sensitivity down to DM masses of $m_\chi \sim 20$ MeV, a considerable improvement over the original CPD analysis which reached masses of ~ 90 MeV [105]. Note that original analysis also attempted to model noise boosting by applying Gaussian smearing, but preserved conservativeness by imposing an arbitrary cutoff on the smearing. This result improves on the previous limit not only because the detector resolution improved, but also because our noise boosting method allows us to fuller advantage of the smearing effects without becoming unconservative.

8.6 Evaluation of Systematics

8.6.1 Calibration Uncertainty

When we determined the net efficiency from the aluminum fluorescence line, we assumed that thermal energy the TES loses by its coupling to the bath is negligible. This is a good approximation for a fast TES which remains in transition. For our detector, we did see evidence in the deviation of the pulse shape for events in the aluminum fluorescence peak which indicates that there was local saturation starting to occur, which means that some of the TES elements in the array were leaving transition. Hence there was some systematic uncertainty introduced on our measurement of the net efficiency.

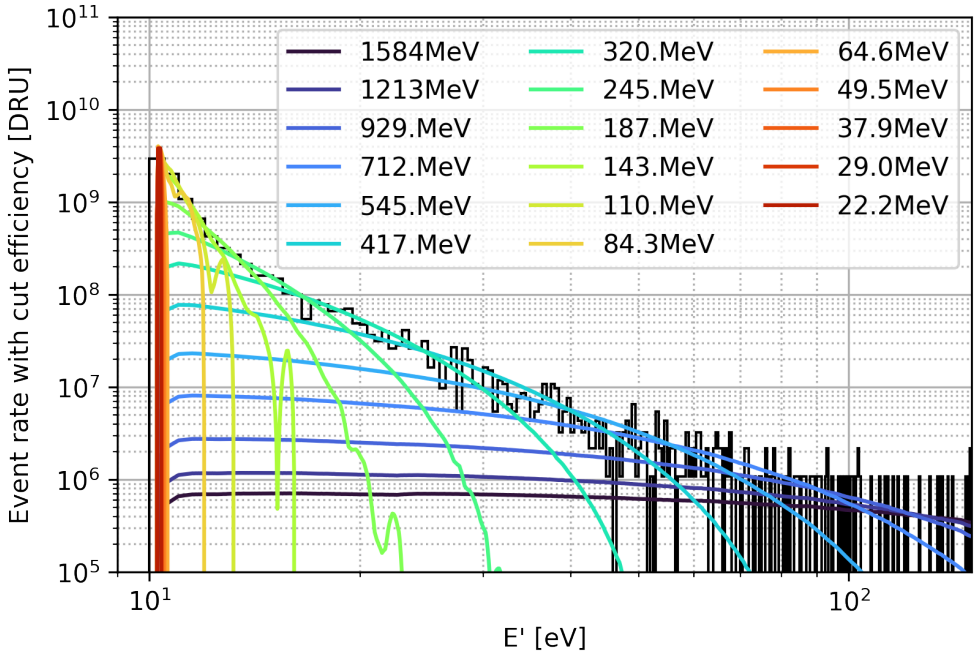


Figure 8.24: DM signal at various masses fit to observed CPD background at UMass. Black histogram is the efficiency-corrected background spectrum. Solid curves are signal models reconstructed by our noise-boosting technique. At low masses ($m_\chi < 50$ MeV), only the lowest bins in the histogram are fit. At intermediate energies ($m_\chi < 100$ MeV), we see spurious peak like features in the signal, which are a consequence of fluctuations in the in the background spectrum being propagated through the signal model. At higher energies, the signal largely resembles the canonical WIMP spectrum and is mostly unaffected by noise boosting.

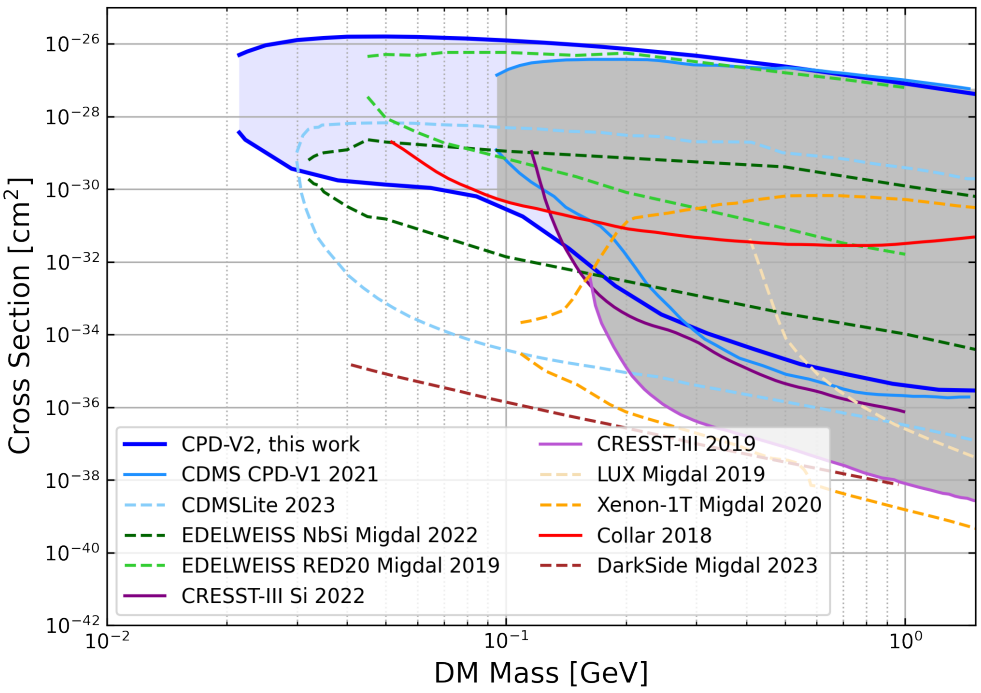
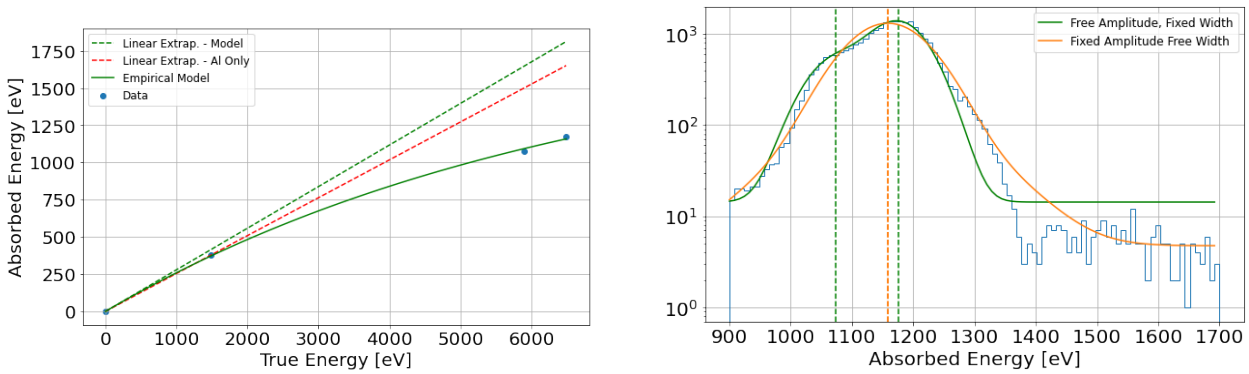


Figure 8.25: Resulting 90% exclusion curve from the UMass DM search [123]. Blue shaded curve are the results of the analysis presented here. Solid curves represent direct NR detection searches, and dashed are ones which utilize the Migdal effect. This analysis extends down to DM masses $m_\chi \sim 20$ MeV. Notably, this is the only direct NRDM search presently extending to masses this low. Other experiments in a similar mass range exploit the Migdal effect.



(a) (b)

Figure 8.26: (Left) Fit of the non-linear, empirical model relating the deposited and absorbed energy. The extrapolation of this model at low energy estimates a higher net efficiency than the linear extrapolation used in the final analysis. (Right) Attempt to fit the ^{55}Fe x-ray peaks observed in the data. The peaks show clear non-Gaussian features which prevent us from having much confidence in our fit, but we use it as a rough estimate.

To quantify this systematic, we compared the assumption that the linear extrapolation of the net efficiency to the empirical one used in [117]

$$E_{abs} \approx E_{ETF} = a(1 - e^{-E_{dep}/b}) \tag{8.34}$$

where a and b are constants with units of energy to be determined. The two interesting limits of this expression are

$$E_{abs} \approx \begin{cases} \frac{a}{b} E_{dep}, & E_{dep} \ll b \\ a, & E_{dep} \gg b \end{cases} \tag{8.35}$$

By comparing this to Eq. 8.25 we see that $a/b = \epsilon_{net}$. We emphasize that this model is strictly empirical and not theoretically motivated. It was chosen based on the fact that it has desirable behavior in the limiting cases with a minimal number of degrees of freedom. Hence we considered this to be more of a toy model than anything else.

To determine the constants a, b, we first determined the position of the ^{55}Fe x-ray peaks. This turned out to be more difficult than in the original CPD analysis. A consequence of our improved baseline resolution is that saturation effects onset earlier, hence the peaks were not well resolved in our data. The spectrum and attempted fits are shown in Fig. 8.26b. We took two approaches to the fit: one where we let the width of each peak float and fix the ratio of the peaks to what we expected from Table 7.1, and another where we constrained each peak to have the same (but still floating) width and allow the heights to float. The fixed ratio approach resulted in the line positions being degenerate, and we discarded this result. The fact that the peak ratios were not consistent with expectations was attributed to the fact observed peak shapes were not Gaussian. Clearly we don't have sufficient understanding of the peak structure, but we proceeded despite this to obtain a rough estimate.

The resulting fit of this model to our data is shown in 8.26a. We found $a = 1.6 \pm 0.2$ keV, $b = 6 \pm 1$ keV and $\epsilon_{net} = 0.28 \pm 0.2$. Given the shortcomings of model we started with and the poor fit to our data we do not use this result in our final analysis. Rather we used it to argue that our linear extrapolation of the net efficiency is likely an underestimate on the order of a few percent, and hence is a conservative assumption. By inspection of the resulting curve, we argue that it is impossible for the linear extrapolation to overestimate ϵ_{net} as long as the true model is monotonically increasing and its derivative is monotonically decreasing. This is a reasonable assumption, since the onset of saturation can only decrease the amount of

energy removed by electrothermal feedback. We attempted to derive a more realistic model in A, though our treatment of local saturation effects, which are the most important in the intermediate energy region we're interested in, are not yet sufficient to apply to real data.

8.6.2 Definition of χ^2_{LF} Cut

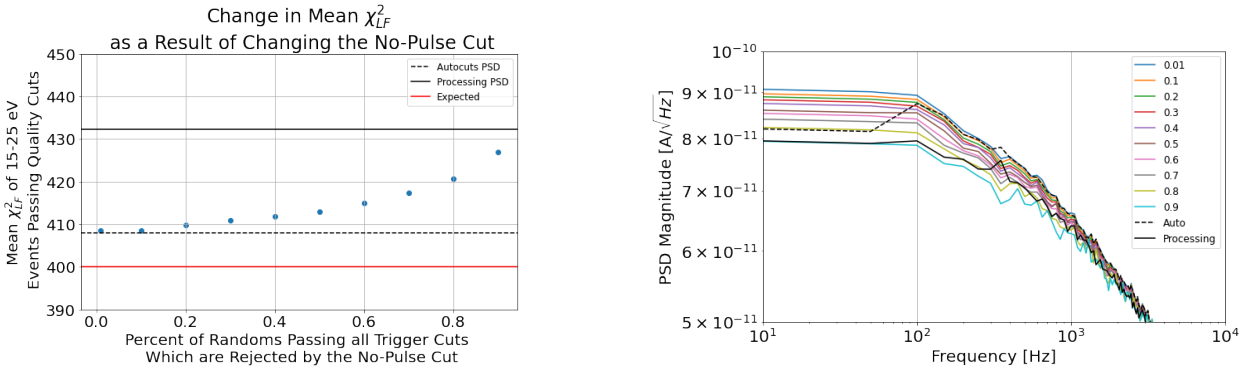
In order to perform our event reconstruction, we must define what constitutes an event by implementing a triggering algorithm. The principle behind our triggering is fairly simple. We used a characterization of the noise and the signal to define a filter, which was then applied to the continuous data stream. The resulting amplitude indicateed how "signal like" that portion of the trace was. Then we selected short (20 ms) sections of the long trace centered around peaks which exceeded a certain threshold in the amplitude (4.5 σ). This was implemented in our analysis with the **pytesdaq** package [124]. Once the trigger was carried out we were left with a collection of events to reconstruct. The central step of this reconstruction is to fit the template to each event using the OF algorithm introduced in 8.2.2.

What is crucial here is that we must choose a representation of the noise before triggering. If the noise was stationary for the entire run it would be well approximated by a single PSD, but we don't ever expect our representation to be perfect. When we determined the PSD we applied a "no-pulse" cut to remove large noise fluctuations from our estimate, and we made some choice of the threshold for this cut which affected the shape of the PSD. We see this in Fig. 8.27b, where the integral of the PSD decreased as we made the threshold more aggressive. There are two consequences of imposing such a cut

1. A more aggressive cut means that PSD was "smaller" and there was less noise deweighting in our trigger. This means that we expected to see more triggers on noise in our spectrum.

2. A more aggresive cut means that we underestimated the noise present when we performed an OF fit to our data, resulting in an increased χ^2 for the fit.

The second point here is illustrated in Fig. 8.27a, where we saw that making the no-pulse cut looser results in a pulse-χ^2 distribution which is centered closer to the expected number of degrees of freedom. This suggests that our optimal filter was not really "optimal" when we applied the cut, in the sense that our PSD did not give the most accurate representation of the noise present in our data. We expected both of these consequences to present themselves by degrading our resolution. We imposed this cut in any case for our analysis, since increasing the number of noise triggers in principle increased our sensitivity to noise boosting



(a) (b)

Figure 8.27: (Right) Estimated PSD resulting from applying a more aggressive no-pulse cut. A more aggressive cut results in smaller estimated PSD amplitudes. (Left) Effect of using these different PSDs on the distribution of χ^2 determined from good pulses slightly above threshold. The more aggressive cut underestimates the noise present in the traces, and skews the mean χ^2 to higher values.

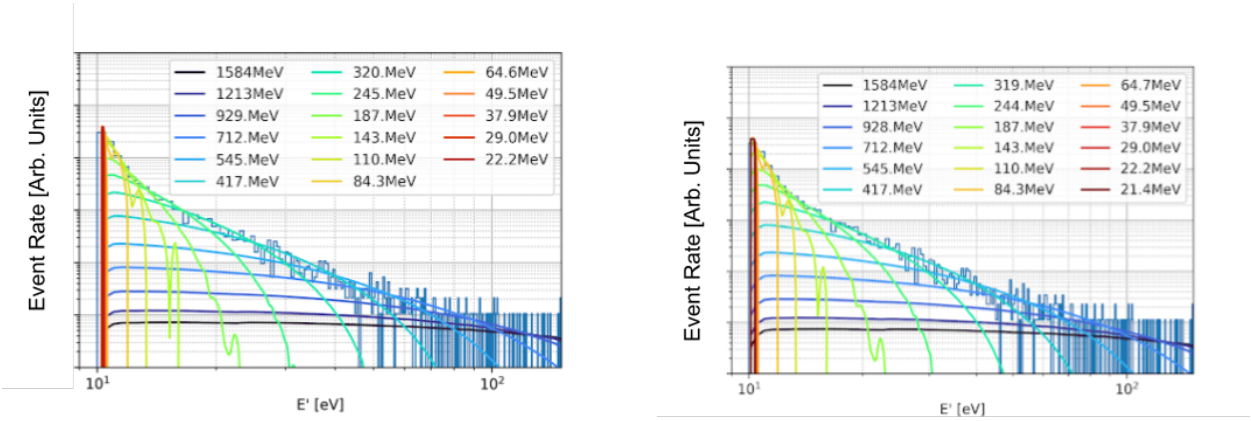
effects. Since this came at the cost of a worse resolution, and hence a worse threshold, we considered this choice to be conservative.

To estimate how this may have changed our result, we considered the effect of loosening the χ^2 cut by 1.75σ. The result demonstrated that loosening the cut only increased the efficiency by $\approx 1\%$, and we concluded that the we were relatively insensitive to our choice of the cut threshold.

8.6.3 Atmospheric Shielding

The UMass DM search was carried out at the surface with no dedicated shielding. Hence we expected our result to be background limited and our sensitivity will be in the large cross-section regime. At sufficiently large interaction cross-section, there is a high probability that the incoming DM particles will scatter off the Earth's atmosphere before reaching the detector, attenuating the velocity distribution included in the signal model. This will affect both the proper upper limit on the cross-section we could sensitivity to, as well as introduced an upper bound to the exclusion curve - above which there would be a negligible DM flux which reached the detector.

To account for these effects, we followed closely the method used in the original CPD DM search analysis [105]. We used the VERNE package to model the average energy loss of the DM for a given mass and cross-section, assuming that it undergoes continuous stopping while traveling in a straight line to the detector [125]. We modeled the detector at depth of 1 m, roughly corresponding to the thickness of concrete above



(a) (b)

Figure 8.28: Signal models with (a) and without (b) atmospheric shielding. There is no noticeable effect on the shape of the signal, the only difference is in the normalization.

the UMass lab. We also included shielding from the fridge and detector housing by modeling the detector as surrounded by 3 cm of copper.

VERNE calculates the velocity distribution and number of expected events in the detector as a function of DM mass and interaction-cross section. Since the signal shape is a function of the cross-section, it is not directly compatible with the OI method. We calculated a 90% Poisson confidence limit by comparing the expected number of events as a function of cross-section to the number of events observed in the spectrum. This was sufficient for determining the critical cross-section, above which we expected to see no events in the spectrum. This imposed a lower mass cutoff independent from the kinematic cutoff, below which no DM was expected to reach the detector. We found this occurs at $m_\chi \sim 25$ MeV.

For the upper limit on the cross-section, we wanted to take advantage of the "optimal" aspects of the OI method. To do this in a manner which accounted for the DM attenuation in the atmosphere, we performed both the OI and VERNE Poisson methods iteratively. We started by calculating the upper limit via the OI with no attenuation. The returned cross-section was then passed to the VERNE code to calculate the damped velocity distribution. We then used this velocity distribution to determine an updated signal model, which we used to once again calculate the OI upper limit. This process was repeated until the limit changed by <2% between iterations. From this process we saw very little change in the shape of the signal, and the dominant effect was a reduction in the DM flux at energies sufficient to appear in our spectrum.

8.7 Sensitivity to Inelastic Channels

An alternative strategy to our noise boosting technique when searching for LDM is to consider interaction models where the energy transfer to the detector for a light particle is more efficient than an elastic NR. These models typically involve some excitation of the electron system, since the light electrons can receive much more kinetic energy than the heavy nucleus. Two such examples which have been studied are the Migdal effect and DM scattering off electrons. When considering the Migdal effect, the DM interacts primarily via recoils off the target nucleus, but these recoils have a small probability to induce an electronic transition. For DM which interacts primarily by scattering off target electrons, we need to introduce some mediator to facilitate the exchange (e.g. a dark photon which weakly couples to the SM photon).

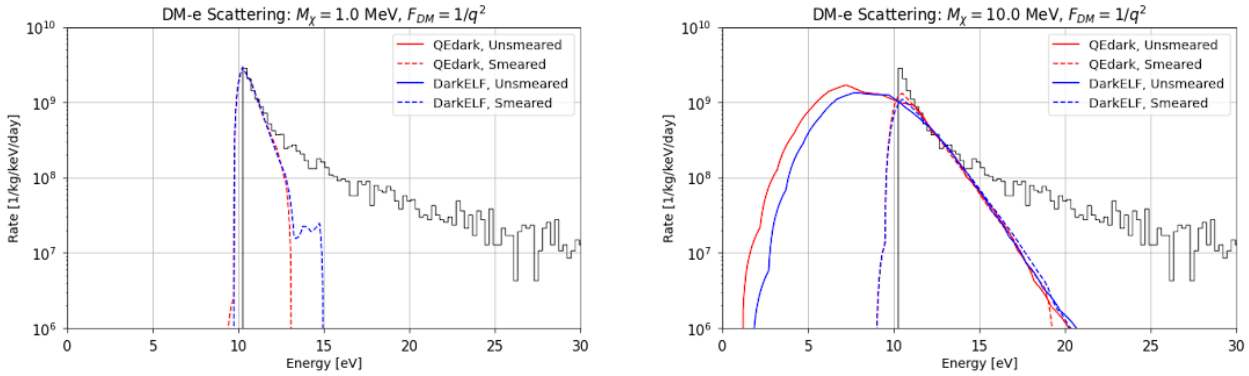
Though these inelastic searches consider different physics than our elastic NR search with noise boosting, they have a similar end result in that both strategies are looking for a signal from DM at masses below what would otherwise be kinematically accessible in the elastic NR search. Furthermore, if the inelastic signal model predicts a significantly larger rate below our detection threshold than above, then we can employ our noise boosting technique to extend our search just as we did in the elastic case.

Note that we looked at these inelastic models without making use of NTL gain. Hence our detector was not optimized to search for ionization signals, and we didn't expect to automatically gain much sensitivity by looking at these channels. Nevertheless, since noise boosting and inelastic interactions have a similar net effect, it is interesting to compare the two.

8.7.1 DM-Electron Scattering

Our detector is also sensitive to models of dark matter which interact by coupling to electrons. Once again, CPD search is not optimized for this channel. The detector being above ground results in a high background and we didn't apply an external voltage to amplify electron recoils. In any case, the fact that we applied noise boosting means that we searched for masses below what we'd normally expect. This made it worthwhile to consider a search in this channel, since we didn't know a priori whether we could probe masses which were previously accessed by other experiments.

We calculated the DM-e scattering rates for the case of a light mediator. We compared the rates calculated with QEdark model and DarkELF. We saw that applying noise boosting gives us sensitivity to DM at $m_\chi = 1$ MeV, which was not the case in the absence of boosting. Note that we didn't consider the overburden effect



(a) (b)

Figure 8.29: DM-electron scattering models fit to observed spectrum. (Left) shows the expectation for $m_\chi = 1$ MeV, and the (Right) $m_\chi = 10$ MeV. Solid line depicts fit without noise boosting, and dashed with. Notably, application of noise boosting suggests sensitivity at 1 MeV which would not be achieved without.

here, as the calculation is substantially different for the electron recoil channel. In Fig. 8.30, we compare the expected sensitivity from the UMass CPD search to the published results of DAMIC-M [49] and DAMIC-SNOLAB, another silicon detector. The sensitivity of CPD was roughly three orders of magnitude less in cross section. This is a consequence of the search being performed above ground, whereas both DAMIC runs were done at deep sites. Remarkably, when applying noise boosting in the electron-recoil channel, we expect CPD to have sensitivity down to $m_\chi \sim 0.5$ MeV. This gives both detectors a similar mass reach. Note that the DAMIC curves also do not account for the overburden, which becomes even more important for a below ground experiment.

8.7.2 Migdal Ionization

A number of DM searches have made use of the Migdal effect to probe lower masses. The Migdal effect occurs when the nucleus probed by the DM recoil causes a sudden perturbation in the Coulomb potential such that an electron is excited.

When considering the Migdal effect in CPD, we were in a unique position compared to previous searches considering this model on account of the fact that we were looking at a silicon detector and our primary region of interest was below ~ 100 eV. All Migdal searches published to date have been performed on either a Ge or liquid noble target. Furthermore, their regions of interest are restricted such that they only consider Migdal excitation of the core electrons and any excitation of valence electrons is excluded. For silicon the

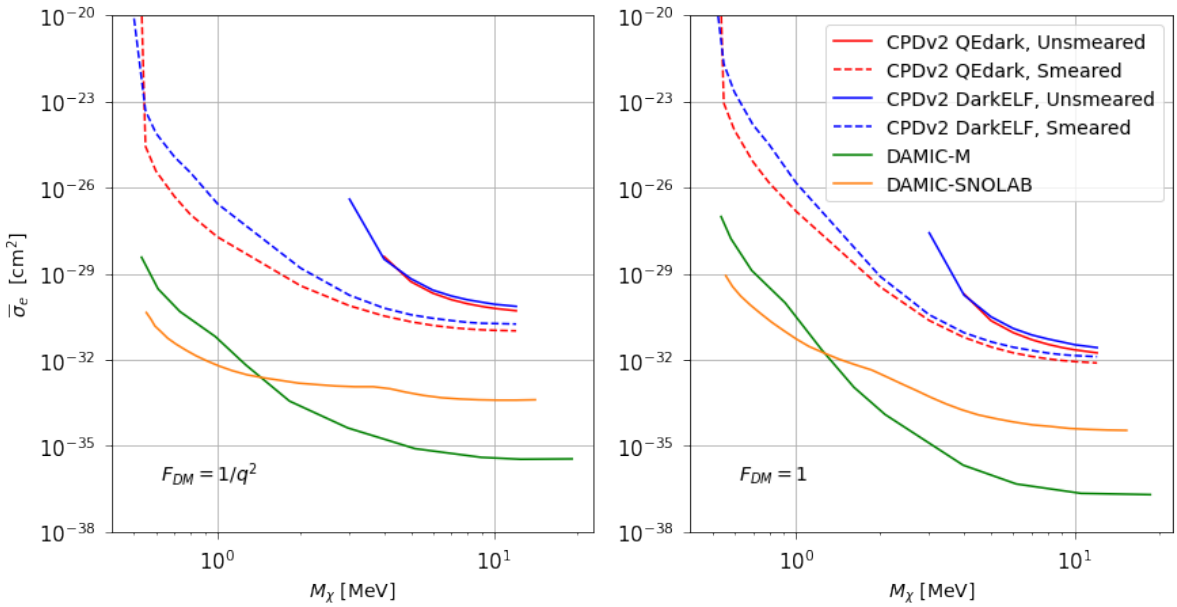
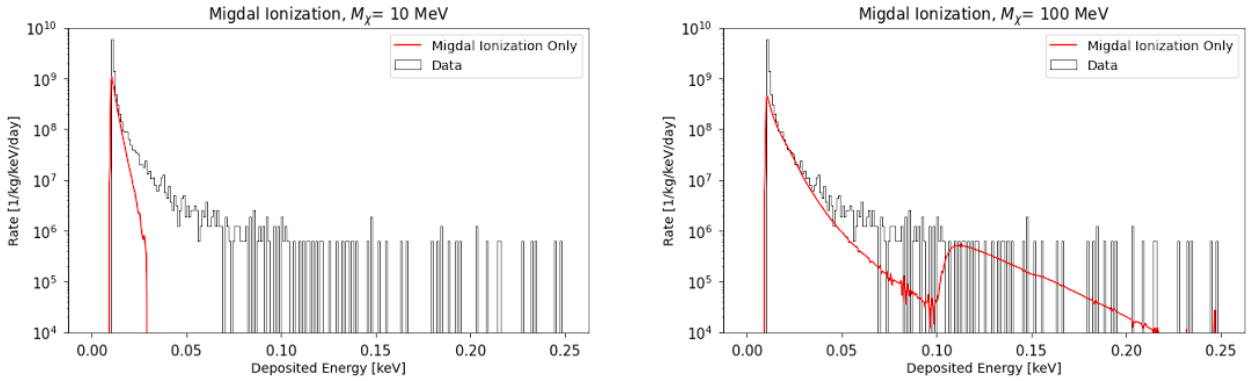


Figure 8.30: Comparison of CPD sensitivity to DM-electron scattering compared to DAMIC-M [49] and DAMIC SNOLAB [126]. Left plot assumes that the mediator of the interaction is light, and right assumes a heavy mediator. Dashed curves indicate that we have applied noise boosting, while solid curves only assume a standard efficiency correction. We present the sensitivity estimated for the UMass CPD search (here labeled "CPDv2") calculated both with QEdark (red) and DarkELF (blue). DAMIC results use QEdark. The DarkELF results are more conservative, as they model screening effects between valence electrons. We see that applying noise boosting effects gives extends the CPD mass reach from $m_\chi \sim 5$ MeV down to $m_\chi \sim 0.5$ MeV.



(a) (b)

Figure 8.31: Signal model from Migdal ionization including noise-boosting (red curve) fit to the UMass CPD background (black histogram). Note that here we only consider the ionization signal, but in general this would appear on top of the nuclear recoil spectrum. To calculate the ionization signal, we use **GPAW** as described in [46]. This calculation is only valid < 70 eV due to the prohibitive computational burden of the underlying DFT calculations at this energy. At masses high enough to excite the L-shell ($\sim$100 eV), the Migdal process should be well described by [56]. At intermediate energies, we use the Lindhard analytic model of the ELF. The left plot shows a 10 MeV DM particle, and all the ionization is from valence electrons. The right shows a 100 MeV particle, which has sufficient energy to excite the L-shell electron.

weakest bound core electron is the L-shell with an ionization energy of ~ 100 eV. Hence if we want to search for the Migdal effect in CPD in a way which takes advantage of our low threshold, we necessarily need to include valence excitations in our model.

To calculate the Migdal effect including the valence electrons, we used the DarkELF package [46]. The derived the rate in the "soft limit" using the Energy Loss Function (ELF) of the material, which was in turn calculated from a DFT formalism using the GPAW package. We made use of the free-ion approximation in our calculation, and didn't apply any low energy cutoff to the phonon excitations, as recommended by [45]. This calculation is only valid < 70 eV due to the prohibitive computational burden of the underlying DFT calculations at this energy. At masses high enough to excite the L shell (~ 100 eV), the Migdal process should be well described by [56], as was done in previously published Migdal results. At intermediate energies, we used the Lindhard analytic model of the ELF.

We see in Fig. 8.32 that, when comparing the Migdal ionization signal to the elastic NR signal with noise boosting, we potentially gain sensitivity at masses below ~ 20 MeV. This is below our proposed mass-reach cutoff of $\sim$22.5 MeV, imposed by the shielding from the atmosphere. Further uncertainty is introduced to the Migdal result in this regime, since we haven't considered how the effect of the overburden will attenuate the Migdal signal. It is worth doing a dedicated calculation of atmospheric shielding due to Migdal interactions

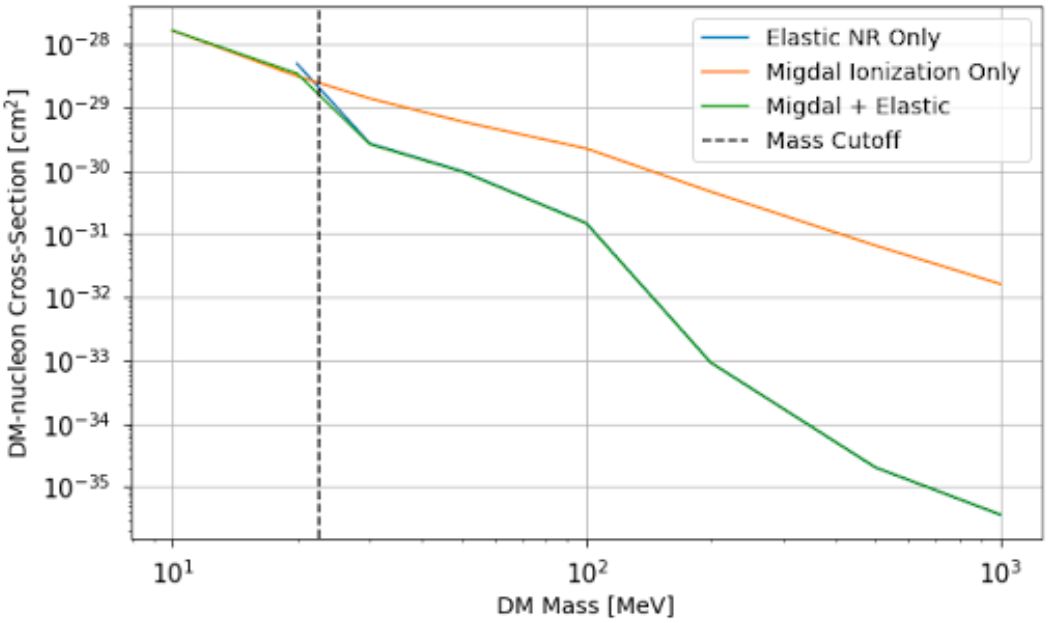


Figure 8.32: Estimated sensitivity to DM interactions including the Migdal effect for the UMass CPD experiment. The blue curve is a sample of the elastic NR search results with noise boosting shown in Fig. 8.25. The orange curve only considers Migdal ionization, also with noise boosting. The green curve is then the sum of these channels. The vertical black, dashed line represents the approximate mass cutoff imposed on Fig. 8.25 by atmospheric shielding of the DM. Though inclusion of Migdal ionization can in principle extend the mass reach of CPD down to $m_\chi \sim 10$ MeV, the region where this adds sensitivity is below the proposed mass cutoff, which brings into question the validity of extending our reach in this way.

to further explore when the channel can present us with additional sensitivity. Our calculation also assumes that the DM interaction only scatters off a single lattice site. At lower masses, the DM has a wavelength comparable to the lattice constant [45], and this may no longer be a valid assumption.

Chapter 9

Summary and Outlook

Observations of our Universe on many different scales provide ample evidence for the existence of dark matter. Historically, searches attempting to detect this dark matter directly have focused on weakly-interacting particles in the GeV-TeV mass scale, but new theoretical models have provided guidance for searches at lower masses. The SuperCDMS collaboration is installing a new underground experiment at SNOLAB which will obtain unprecedented sensitivity to models of dark matter in the 0.5-10 GeV mass range which recoil off nuclei, while developing the next generation of detectors sensitive to sub-GeV dark matter interacting through electron-recoil signatures. Both goals are accomplished by advances in sensitivity, combined with reduction of backgrounds.

For the main SuperCDMS SNOLAB experiment, reduction of the environmental background is primarily accomplished by operating deep underground at SNOLAB to protect from cosmic rays and associated high-energy neutrons. In order to capitalize on the improved overburden, additional shielding is required to mitigate radiogenic backgrounds from the cavern walls and associated radon. The design and efficacy of the shield was determined by simulating the neutron background originating from radioactive decays in the rock surrounding the experiment. Single nuclear-recoils most resemble a WIMP signal and are a particularly dangerous background. The contribution from this source in the iZIPs was found to be $\sim 300 \times 10^{-6}$ counts/kg/keV/year, which is subdominant to the CEνNS background and that from radioactive decays in the shield itself. Since energetic neutrons from the cavern were shown to reach the detectors primarily through the stem penetrations in the shield, additional high density polyethylene shielding around the outside of the stems provides an additional order of magnitude reduction of cavern backgrounds. Additional background

rejection is provided by the detector technology itself, in particular the deployment of iZIP detectors which can discriminate between nuclear and electronic recoils using independent phonon and ionization channels.

The second goal, that of achieving a high sensitivity to small energy depositions, is partly the result of using silicon and germanium solid state target materials. The detection threshold in these materials for electronic excitations, which are coupled to nuclear excitations, is determined by the $\mathcal{O}(1\text{ eV})$ band gap. Sensitive transition edge sensors record these excitations with resolutions determined by their heat capacity, critical temperature, and efficiency to absorb phonons. The expected baseline resolutions for these devices at SNOLAB are $\mathcal{O}(10\text{ eV})$. Further sensitivity is achieved by operating detectors under high voltage, which allows us to exploit the Neganov-Trofimov-Luke effect to further amplify the resulting signal.

Extending our reach to lower masses below 0.5 GeV must confront the challenge of inefficient kinetic energy transfer from a light particle to a heavy target. SuperCDMS has demonstrated two detector technologies with remarkable sensitivity sufficient to probe sub-GeV dark matter models. Both achieve this by reducing the overall TES heat capacity while increasing the total energy efficiency. The CPD is operated at zero bias, and is well suited to search for light dark matter interacting via nuclear recoils. The HVeV is designed to operate at a bias of ~ 100 V, and is best suited to search for light dark matter interacting via electron recoils. Results are presented from a dark matter search above ground using an upgraded CPD detector at the University of Massachusetts at Amherst. This analysis was a joint effort between the SuperCDMS and SPICE/HeRALD collaborations, both of which are interested in searching for light dark matter using CPD style detectors. The UMass CPD analysis achieves sensitivity to dark matter at masses as low as $\sim$25 MeV. This is accomplished by its excellent 2.3 eV energy resolution, coupled with a novel treatment of near-threshold events.

Over the next few years, SuperCDMS will take data with HV and iZIP detectors in a large shielded cryostat in the SNOLAB location. The results of that experiment will improve the sensitivity to low mass dark matter by orders of magnitude. Currently, in preparation for the 4-tower payload, one tower of silicon and germanium HV detectors is running underground in CUTE, with the expectation that the 6-month engineering run will already provide new physics results. Meanwhile, development of transition edge sensor technology is rapidly advancing, and demonstration of sub-eV resolutions from CPD detectors is expected in 2024. Once data-taking for SuperCDMS SNOLAB is complete, additional towers composed of a mosaic of CPDs and HVeVs are proposed for the underground facility. There are many exciting avenues for TES-based detectors to reach unprecedented sensitivity to the tiny energy depositions predicted by new theoretical

models, and plenty of parameter space to explore before approaching the neutrino fog at such low masses. SuperCDMS is partnering with other solid state dark matter collaborations to develop these new technologies and its SNOLAB facility can provide the low-background environment for deploying them in the future.

References

[1] N. Aghanim *et al.*, "Planck 2018 results," *Astronomy & Astrophysics*, vol. 641, A6, Sep. 2020. DOI: 10.1051/0004-6361/201833910. [Online]. Available: https://doi.org/10.1051%5C%2F0004-6361%5C%2F201833910.

[2] D. J. Fixsen, "The temperature of the cosmic microwave background," *The Astrophysical Journal*, vol. 707, no. 2, 916–920, Nov. 2009. DOI: 10.1088/0004-637x/707/2/916. [Online]. Available: https://doi.org/10.1088%5C%2F0004-637x%5C%2F707%5C%2F2%5C%2F916.

[3] W. Hu, N. Sugiyama, and J. Silk, "The physics of microwave background anisotropies," *Nature*, vol. 386, no. 6620, 37–43, Mar. 1997. DOI: "10.1038/386037a0. [Online]. Available: https://doi.org/10.1038%5C%2F386037a0.

[4] V. Rubin, W. Ford, and N. Thonnard, *Astroph. J.*, vol. 238, 471, 1980.

[5] J. Lewin and P. Smith, *Astroparticle Physics*, vol. 6, 1996.

[6] E. W. Kolb and M. S. Turner, *The Early Universe*. 1990, vol. 69, ISBN: 978-0-201-62674-2. DOI: 10.1201/9780429492860.

[7] Y. Du, F. Huang, H.-L. Li, Y.-Z. Li, and J.-H. Yu, "Revisiting dark matter freeze-in and freeze-out through phase-space distribution," *Journal of Cosmology and Astroparticle Physics*, vol. 2022, no. 04, 012, Apr. 2022, ISSN: 1475-7516. DOI: 10.1088/1475-7516/2022/04/012. [Online]. Available: http://dx.doi.org/10.1088/1475-7516/2022/04/012.

[8] L. J. Hall, K. Jedamzik, J. March-Russell, and S. M. West, "Freeze-in production of fimp dark matter," *Journal of High Energy Physics*, vol. 2010, no. 3, Mar. 2010, ISSN: 1029-8479. DOI: 10.1007/jhep03(2010)080. [Online]. Available: http://dx.doi.org/10.1007/JHEP03(2010)080.

[9] F. Elahi, C. Kolda, and J. Unwin, "Ultraviolet freeze-in," *Journal of High Energy Physics*, vol. 2015, no. 3, Mar. 2015, ISSN: 1029-8479. DOI: 10.1007/jhep03(2015)048. [Online]. Available: http://dx.doi.org/10.1007/JHEP03(2015)048.

[10] R. Essig, J. Mardon, and T. Volansky, "Direct detection of sub-gev dark matter," *Physical Review D*, vol. 85, no. 7, Apr. 2012, ISSN: 1550-2368. DOI: 10.1103/physrevd.85.076007. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.85.076007.

[11] Y. Hochberg, E. Kuflik, T. Volansky, and J. G. Wacker, "Mechanism for thermal relic dark matter of strongly interacting massive particles," *Physical Review Letters*, vol. 113, no. 17, Oct. 2014, ISSN: 1079-7114. DOI: 10.1103/physrevlett.113.171301. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.113.171301.

[12] Y. Hochberg, E. Kuflik, H. Murayama, T. Volansky, and J. G. Wacker, "Model for thermal relic dark matter of strongly interacting massive particles," *Physical Review Letters*, vol. 115, no. 2, Jul. 2015, ISSN: 1079-7114. DOI: 10.1103/physrevlett.115.021301. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.115.021301.

[13] E. Kuflik, M. Perelstein, N. R.-L. Lorier, and Y.-D. Tsai, "Elastically decoupling dark matter," *Physical Review Letters*, vol. 116, no. 22, Jun. 2016, ISSN: 1079-7114. DOI: 10.1103/physrevlett.116.221302. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.116.221302.

[14] Y. Hochberg, "SIMP Dark Matter," *SciPost Phys. Lect. Notes*, 59, 2022. [Online]. Available: https://scipost.org/10.21468/SciPostPhysLectNotes.59.

[15] D. E. Kaplan, M. A. Luty, and K. M. Zurek, "Asymmetric dark matter," *Physical Review D*, vol. 79, no. 11, Jun. 2009, ISSN: 1550-2368. DOI: 10.1103/physrevd.79.115016. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.79.115016.

[16] S. P. MARTIN, "A SUPERSYMMETRY PRIMER," *Perspectives on Supersymmetry*, WORLD SCIENTIFIC, Jul. 1998, 1–98. DOI: 10.1142/9789812839657_0001. [Online]. Available: https://doi.org/10.1142%2F9789812839657_0001.

[17] F. Chadha-Day, J. Ellis, and D. J. E. Marsh, *Axion dark matter: What is it and why now?* 2022. arXiv: 2105.01406 [hep-ph].

[18] S. R. Golwala, "Exclusion limits on the WIMP nucleon elastic scattering cross-section from the Cryogenic Dark Matter Search," Ph.D. dissertation, UC, Berkeley, 2000. DOI: 10.2172/1421437.

[19] J. Ellis, K. A. Olive, V. C. Spanos, and I. D. Stamou, "The CMSSM survives planck, the LHC, LUX-ZEPLIN, fermi-LAT, h.e.s.s. and IceCube," *The European Physical Journal C*, vol. 83, no. 3, Mar. 2023. DOI: 10.1140/epjc/s10052-023-11405-1. [Online]. Available: https://doi.org/10.1140%2Fepjc%2Fs10052-023-11405-1.

[20] X. Chu, T. Hambye, and M. H. Tytgat, "The four basic ways of creating dark matter through a portal," *Journal of Cosmology and Astroparticle Physics*, vol. 2012, no. 05, 034–034, May 2012, ISSN: 1475-7516. DOI: 10.1088/1475-7516/2012/05/034. [Online]. Available: http://dx.doi.org/10.1088/1475-7516/2012/05/034.

[21] J. Jaeckel, *A force beyond the standard model - status of the quest for hidden photons*, 2013. arXiv: 1303.1821 [hep-ph].

[22] M. J. Strassler and K. M. Zurek, "Echoes of a hidden valley at hadron colliders," *Physics Letters B*, vol. 651, no. 5–6, 374–379, Aug. 2007, ISSN: 0370-2693. DOI: 10.1016/j.physletb.2007.06.055. [Online]. Available: http://dx.doi.org/10.1016/j.physletb.2007.06.055.

[23] J. Cooley *et al.*, *Report of the topical group on particle dark matter for snowmass 2021*, 2022. arXiv: 2209.07426 [hep-ph].

[24] R. Bernabei *et al.*, "Final model independent result of dama/libra–phase1," *The European Physical Journal C*, vol. 73, no. 12, Nov. 2013, ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-013-2648-7. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-013-2648-7.

[25] R. Agnese *et al.*, "Silicon detector dark matter results from the final exposure of cdms ii," *Phys. Rev. Lett.*, vol. 111, 251301, 25 Dec. 2013. DOI: 10.1103/PhysRevLett.111.251301. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.111.251301.

[26] G. Angloher *et al.*, "Results from 730 kg-days of the cresst-ii dark matter search," *The European Physical Journal C*, vol. 72, no. 4, Apr. 2012, ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-012-1971-8. [Online]. Available: http://dx.doi.org/10.1140/epjc/s10052-012-1971-8.

[27] R. Agnese *et al.*, "Low-mass dark matter search with CDMSlite," *Physical Review D*, vol. 97, no. 2, Jan. 2018. DOI: 10.1103/physrevd.97.022002. [Online]. Available: https://doi.org/10.1103%5C%2Fphysrevd.97.022002.

[28] E. Aprile *et al.*, "Dark matter results from 225 live days of xenon100 data," *Physical Review Letters*, vol. 109, no. 18, Nov. 2012, ISSN: 1079-7114. DOI: 10.1103/physrevlett.109.181301. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.109.181301.

[29] G. Bertone, D. Hooper, and J. Silk, "Particle dark matter: Evidence, candidates and constraints," *Physics Reports*, vol. 405, no. 5-6, 279–390, Jan. 2005. DOI: 10.1016/j.physrep.2004.08.031. [Online]. Available: https://doi.org/10.1016%2Fj.physrep.2004.08.031.

[30] E. Hooper and T. Linden, "Origin of the gamma rays from the galactic center," *Phys. Rev. D*, vol. 84, 123005, 12 Dec. 2011. DOI: 10.1103/PhysRevD.84.123005. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.84.123005.

[31] A. Hayrapetyan *et al.*, "Search for dark matter particles in W^+W^- events with transverse momentum imbalance in proton-proton collisions at $\sqrt{s} = 13$ TeV," Oct. 2023. arXiv: 2310.12229 [hep-ex].

[32] R. W. SCHNEE, "INTRODUCTION TO DARK MATTER EXPERIMENTS," *Physics of the Large and the Small*, WORLD SCIENTIFIC, Mar. 2011. DOI: 10.1142/9789814327183_0014. [Online]. Available: https://doi.org/10.1142%2F9789814327183_0014.

[33] D. Baxter *et al.*, "Recommended conventions for reporting results from direct dark matter searches," *The European Physical Journal C*, vol. 81, no. 10, Oct. 2021. DOI: 10.1140/epjc/s10052-021-09655-y. [Online]. Available: https://doi.org/10.1140%2Fepjc%2Fs10052-021-09655-y.

[34] M. Markevitch *et al.*, "Direct constraints on the dark matter self-interaction cross section from the merging galaxy cluster 1e 0657-56," *The Astrophysical Journal*, vol. 606, no. 2, 819–824, May 2004, ISSN: 1538-4357. DOI: 10.1086/383178. [Online]. Available: http://dx.doi.org/10.1086/383178.

[35] A. K. Drukier, K. Freese, and D. N. Spergel, "Detecting cold dark-matter candidates," *Phys. Rev. D*, vol. 33, 3495–3508, 12 Jun. 1986. DOI: 10.1103/PhysRevD.33.3495. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.33.3495.

[36] N. W. Evans, C. M. Carollo, and P. T. de Zeeuw, "Triaxial haloes and particle dark matter detection," *Monthly Notices of the Royal Astronomical Society*, vol. 318, no. 4, 1131–1143, Nov. 2000. DOI: 10.1046/j.1365-8711.2000.03787.x. [Online]. Available: https://doi.org/10.1046%2Fj.1365-8711.2000.03787.x.

[37] M. Vogelsberger *et al.*, "Phase-space structure in the local dark matter distribution and its signature in direct detection experiments," *Monthly Notices of the Royal Astronomical Society*, vol. 395, no. 2, 797–811, Apr. 2009, ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2009.14630.x. eprint: https://academic.oup.com/mnras/article-pdf/395/2/797/4880273/mnras0395-0797.pdf. [Online]. Available: https://doi.org/10.1111/j.1365-2966.2009.14630.x.

[38] F. An *et al.*, "Seasonal variation of the underground cosmic muon flux observed at daya bay," *Journal of Cosmology and Astroparticle Physics*, vol. 2018, no. 01, 001–001, Jan. 2018. DOI: 10.1088/1475-7516/2018/01/001. [Online]. Available: https://doi.org/10.1088%2F1475-7516%2F2018%2F01%2F001.

[39] J. Amaré *et al.*, "Long term measurement of the ^{222}Rn concentration in the Canfranc Underground Laboratory," *Eur. Phys. J. C*, vol. 82, no. 10, 891, 2022. DOI: 10.1140/epjc/s10052-022-10859-z. arXiv: 2203.13978 [physics.ins-det].

[40] E. Barberio *et al.*, "Simulation and background characterisation of the SABRE south experiment," *The European Physical Journal C*, vol. 83, no. 9, Sep. 2023. DOI: 10.1140/epjc/s10052-023-11817-z. [Online]. Available: https://doi.org/10.1140%2Fepjc%2Fs10052-023-11817-z.

[41] N. Kurinsky, T. C. Yu, Y. Hochberg, and B. Cabrera, "Diamond detectors for direct detection of sub-gev dark matter," *Physical Review D*, vol. 99, no. 12, Jun. 2019, ISSN: 2470-0029. DOI: 10.1103/physrevd.99.123005. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.99.123005.

[42] S. Verma *et al.*, "Low-threshold sapphire detector for rare event searches," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1046, 167634, Jan. 2023, ISSN: 0168-9002. DOI: 10.1016/j.nima.2022.167634. [Online]. Available: http://dx.doi.org/10.1016/j.nima.2022.167634.

[43] R. Anthony-Petersen *et al.*, "Applying Superfluid Helium to Light Dark Matter Searches: Demonstration of the HeRALD Detector Concept," Jul. 2023. arXiv: 2307.11877 [physics.ins-det].

[44] T. Trickle, Z. Zhang, K. M. Zurek, K. Inzani, and S. M. Griffin, "Multi-channel direct detection of light dark matter: Theoretical framework," *Journal of High Energy Physics*, vol. 2020, no. 3, Mar. 2020, ISSN: 1029-8479. DOI: 10.1007/jhep03(2020)036. [Online]. Available: http://dx.doi.org/10.1007/JHEP03(2020)036.

[45] K. V. Berghaus, A. Esposito, R. Essig, and M. Sholapurkar, "The migdal effect in semiconductors for dark matter with masses below 100 mev," *Journal of High Energy Physics*, vol. 2023, no. 1, Jan. 2023, ISSN: 1029-8479. DOI: 10.1007/jhep01(2023)023. [Online]. Available: http://dx.doi.org/10.1007/JHEP01(2023)023.

[46] S. Knapen, J. Kozaczuk, and T. Lin, "Python package for dark matter scattering in dielectric targets," *Physical Review D*, vol. 105, no. 1, Jan. 2022, ISSN: 2470-0029. DOI: 10.1103/physrevd.105.015014. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.105.015014.

[47] R. Essig, M. Fernandez-Serra, J. Mardon, A. Soto, T. Volansky, and T.-T. Yu, *Direct detection of sub-gev dark matter with semiconductor targets*, 2016. arXiv: 1509.01598 [hep-ph].

[48] *Direct detection of sub-gev dark matter*, 2018. [Online]. Available: https://github.com/tientienyu/QEdark.

[49] I. Arnquist *et al.*, "First constraints from damic-m on sub-gev dark-matter particles interacting with electrons," *Physical Review Letters*, vol. 130, no. 17, Apr. 2023, ISSN: 1079-7114. DOI: 10.1103/physrevlett.130.171003. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.130.171003.

[50] Q. Arnaud *et al.*, "First germanium-based constraints on sub-mev dark matter with the edelweiss experiment," *Physical Review Letters*, vol. 125, no. 14, Oct. 2020, ISSN: 1079-7114. DOI: 10.1103/physrevlett.125.141301. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.125.141301.

[51] L. Barak *et al.*, "Sensei: Direct-detection results on sub-gev dark matter from a new skipper ccd," *Physical Review Letters*, vol. 125, no. 17, Oct. 2020, ISSN: 1079-7114. DOI: 10.1103/physrevlett.125.171802. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.125.171802.

[52] R. Agnese *et al.*, "Erratum: First dark matter constraints from a supercdms single-charge sensitive detector [phys. rev. lett. 121, 051301 (2018)]," *Phys. Rev. Lett.*, vol. 122, 069901, 6 Feb. 2019. DOI: 10.1103/PhysRevLett.122.069901. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.122.069901.

[53] T. Trickle, "Extended calculation of electronic excitations for direct detection of dark matter," *Physical Review D*, vol. 107, no. 3, Feb. 2023, ISSN: 2470-0029. DOI: 10.1103/physrevd.107.035035. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.107.035035.

[54] C. E. Dreyer, R. Essig, M. Fernandez-Serra, A. Singal, and C. Zhen, "Fully ab-initio all-electron calculation of dark matter–electron scattering in crystals with evaluation of systematic uncertainties," Jun. 2023. arXiv: 2306.14944 [hep-ph].

[55] L. Landau and E. Lifshitz, *Quantum Mechanics (Non-Relativistic Theory). 3rd Edition*. Pergamon Press, Oxford, 1977.

[56] M. Ibe, W. Nakano, Y. Shoji, and K. Suzuki, "Migdal effect in dark matter direct detection experiments," *Journal of High Energy Physics*, vol. 2018, no. 3, Mar. 2018. DOI: 10.1007/jhep03(2018)194. [Online]. Available: https://doi.org/10.1007%5C%2Fjhep03%5C%282018%5C%29194.

[57] E. Aprile *et al.*, "Search for light dark matter interactions enhanced by the migdal effect or bremsstrahlung in xenon1t," *Physical Review Letters*, vol. 123, no. 24, Dec. 2019, ISSN: 1079-7114. DOI: 10.1103/physrevlett.123.241803. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.123.241803.

[58] P. Agnes *et al.*, "Search for dark-matter–nucleon interactions via migdal effect with darkside-50," *Phys. Rev. Lett.*, vol. 130, 101001, 10 Mar. 2023. DOI: 10.1103/PhysRevLett.130.101001. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.130.101001.

[59] D. S. Akerib *et al.*, "Results of a search for sub-gev dark matter using 2013 lux data," *Phys. Rev. Lett.*, vol. 122, 131301, 13 Apr. 2019. DOI: 10.1103/PhysRevLett.122.131301. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.122.131301.

[60] M. Albakry *et al.*, "Search for low-mass dark matter via bremsstrahlung radiation and the migdal effect in supercdms," *Physical Review D*, vol. 107, no. 11, Jun. 2023, ISSN: 2470-0029. DOI: 10.1103/physrevd.107.112013. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.107.112013.

[61] E. Armengaud *et al.*, "Search for sub-gev dark matter via the migdal effect with an edelweiss germanium detector with nbsi transition-edge sensors," *Phys. Rev. D*, vol. 106, 062004, 6 Sep. 2022. DOI: 10.1103/PhysRevD.106.062004. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.106.062004.

[62] Z. Z. Liu *et al.*, "Constraints on spin-independent nucleus scattering with sub-gev weakly interacting massive particle dark matter from the cdex-1b experiment at the china jinping underground laboratory," *Phys. Rev. Lett.*, vol. 123, 161301, 16 Oct. 2019. DOI: 10.1103/PhysRevLett.123.161301. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.123.161301.

[63] R. Essig, J. Pradler, M. Sholapurkar, and T.-T. Yu, "Relation between the migdal effect and dark matter-electron scattering in isolated atoms and semiconductors," *Physical Review Letters*, vol. 124, no. 2, Jan. 2020, ISSN: 1079-7114. DOI: 10.1103/physrevlett.124.021801. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.124.021801.

[64] S. Knapen, J. Kozaczuk, and T. Lin, "Migdal effect in semiconductors," *Physical Review Letters*, vol. 127, no. 8, Aug. 2021, ISSN: 1079-7114. DOI: 10.1103/physrevlett.127.081805. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.127.081805.

[65] J. Xu *et al.*, *Search for the migdal effect in liquid xenon with kev-level nuclear recoils*, 2023. arXiv: 2307.12952 [hep-ex].

[66] D. Adams, D. Baxter, H. Day, R. Essig, and Y. Kahn, "Measuring the migdal effect in semiconductors for dark matter detection," *Physical Review D*, vol. 107, no. 4, Feb. 2023, ISSN: 2470-0029. DOI: 10.1103/physrevd.107.l041303. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.107.L041303.

[67] R. Essig *et al.*, *Snowmass2021 cosmic frontier: The landscape of low-threshold dark matter direct detection in the next decade*, 2023. arXiv: 2203.08297 [hep-ph].

[68] SuperCDMS Collaboration, *SuperCDMS Limit Plotter*, https://supercdms.slac.stanford.edu/science-results/dark-matter-limit-plotter, 2023.

[69] F. E. Emery and T. A. Rabson, "Average energy expended per ionized electron-hole pair in silicon and germanium as a function of temperature," *Phys. Rev.*, vol. 140, A2089–A2093, 6A Dec. 1965. DOI: 10.1103/PhysRev.140.A2089. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.140.A2089.

[70] K. Ramanathan and N. Kurinsky, "Ionization yield in silicon for eV-scale electron-recoil processes," *Physical Review D*, vol. 102, no. 6, Sep. 2020. DOI: 10.1103/physrevd.102.063026. [Online]. Available: https://doi.org/10.1103%5C%2Fphysrevd.102.063026.

[71] Y. Varshni, "Temperature dependence of the energy gap in semiconductors," *Physica*, vol. 34, no. 1, 149–154, 1967, ISSN: 0031-8914. DOI: https://doi.org/10.1016/0031-8914(67)90062-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0031891467900626.

[72] R. W. Ogburn IV, "A search for particle dark matter using cryogenic germanium and silicon detectors in the one- and two- tower runs of CDMS-II at Soudan," Ph.D. dissertation, Stanford U., 2008. DOI: 10.2172/935475.

[73] C. W. Fink *et al.*, "Characterizing TES power noise for future single optical-phonon and infrared-photon detectors," *AIP Advances*, vol. 10, no. 8, 085221, Aug. 2020, ISSN: 2158-3226. DOI: 10.1063/5.0011130. eprint: https://pubs.aip.org/aip/adv/article-pdf/doi/10.1063/5.0011130/12982982/085221_1_online.pdf. [Online]. Available: https://doi.org/10.1063/5.0011130.

[74] M. Pepin, "Low-mass dark matter search results and radiogenic backgrounds for the cryogenic dark matter search," Available at https://hdl.handle.net/11299/185144, PhD thesis, University of Minnesota, 2016.

[75] K. Irwin and G. Hilton, "Transition-edge sensors," *Cryogenic Particle Detection*, C. Enss, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, 63–150. DOI: 10.1007/10933596_3. [Online]. Available: https://doi.org/10.1007/10933596_3.

[76] Z. Ahmed *et al.*, "Search for inelastic dark matter with the CDMS II experiment," *Physical Review D*, vol. 83, no. 11, Jun. 2011. DOI: 10.1103/physrevd.83.112002. [Online]. Available: https://doi.org/10.1103%5C%2Fphysrevd.83.112002.

[77] R. Agnese *et al.*, "Production rate measurement of tritium and other cosmogenic isotopes in germanium with CDMSlite," *Astroparticle Physics*, vol. 104, Jan. 2019. DOI: 10.1016/j.astropartphys.2018.08.006.

[78] R. Agnese *et al.*, "Projected sensitivity of the SuperCDMS SNOLAB experiment," *Physical Review D*, vol. 95, no. 8, Apr. 2017. DOI: 10.1103/physrevd.95.082002. [Online]. Available: https://doi.org/10.1103%5C%2Fphysrevd.95.082002.

[79] J. L. Orrell, I. J. Arnquist, M. Bliss, R. Bunker, and Z. S. Finch, "Naturally occurring 32si and low-background silicon dark matter detectors," *Astroparticle Physics*, vol. 99, 9–20, May 2018, ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2018.02.005. [Online]. Available: http://dx.doi.org/10.1016/j.astropartphys.2018.02.005.

[80] N. J. T. Smith, "The snolab deep underground facility," *The European Physical Journal Plus*, vol. 127, Sep. 2012. DOI: 10.1140/epjp/i2012-12108-9. [Online]. Available: https://doi.org/10.1140/epjp/i2012-12108-9.

[81] J. A. Formaggio and C. Martoff, "Backgrounds to sensitive experiments underground," *Annual Review of Nuclear and Particle Science*, vol. 54, no. 1, 361–412, 2004. DOI: 10.1146/annurev.nucl.54.070103.181248. eprint: https://doi.org/10.1146/annurev.nucl.54.070103.181248. [Online]. Available: https://doi.org/10.1146/annurev.nucl.54.070103.181248.

[82] C. A. J. O'Hare, "New definition of the neutrino floor for direct dark matter searches," *Physical Review Letters*, vol. 127, no. 25, Dec. 2021, ISSN: 1079-7114. DOI: 10.1103/physrevlett.127.251802. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.127.251802.

[83] P. Abbamonte *et al.*, "Revisiting the dark matter interpretation of excess rates in semiconductors," *Physical Review D*, vol. 105, no. 12, Jun. 2022, ISSN: 2470-0029. DOI: 10.1103/physrevd.105.123002. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.105.123002.

[84] P. Du, D. Egana-Ugrinovic, R. Essig, and M. Sholapurkar, "Sources of low-energy events in low-threshold dark-matter and neutrino detectors," *Physical Review X*, vol. 12, no. 1, Jan. 2022, ISSN: 2160-3308. DOI: 10.1103/physrevx.12.011009. [Online]. Available: http://dx.doi.org/10.1103/PhysRevX.12.011009.

[85] M. F. Albakry *et al.*, "Investigating the sources of low-energy events in a supercdms-hvev detector," *Phys. Rev. D*, vol. 105, 112006, 11 Jun. 2022. DOI: 10.1103/PhysRevD.105.112006. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.105.112006.

[86] R. Anthony-Petersen *et al.*, "A Stress Induced Source of Phonon Bursts and Quasiparticle Poisoning," Aug. 2022. arXiv: 2208.02790 [physics.ins-det].

[87] S. Agostinelli *et al.*, "Geant4—a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, 250–303, 2003, ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[88] W. B. Wilson *et al.*, "Sources 4c : A code for calculating ([alpha],n), spontaneous fission, and delayed neutron sources and spectra.," Jan. 2002. [Online]. Available: https://www.osti.gov/biblio/976142.

[89] R. Calkins, *Snolab cavern neutron flux spectra*, SuperCDMS Internal Note, 2015. [Online]. Available: https://scdms.slac.stanford.edu/cdms_restricted/rcalkins/150722_rc/.

[90] K. Shibata *et al.*, "Jendl-4.0: A new library for nuclear science and engineering," *Journal of Nuclear Science and Technology*, vol. 48, no. 1, 1–30, 2011. [Online]. Available: https://doi.org/10.1080/18811248.2011.9711675.

[91] A. N. Villano, K. Harris, and S. Brown, "nrCascadeSim - a simulation tool for nuclear recoil cascades resulting from neutron capture," *Journal of Open Source Software*, vol. 7, no. 70, 3993, Feb. 2022. DOI: 10.21105/joss.03993. [Online]. Available: https://doi.org/10.21105%2Fjoss.03993.

[92] K. Harris, A. Gevorgian, A. Biffl, and A. Villano, "Neutron capture-induced silicon nuclear recoils for dark matter and cevns," *Physical Review D*, vol. 107, no. 7, Apr. 2023. DOI: 10.1103/physrevd.107.076026. [Online]. Available: https://doi.org/10.1103%2Fphysrevd.107.076026.

[93] A. Biffl, A. Gevorgian, K. Harris, and A. Villano, "Neutron capture-induced nuclear recoils as background for cevns measurements at reactors," *Physical Review D*, vol. 107, no. 9, May 2023. DOI: 10.1103/physrevd.107.092011. [Online]. Available: https://doi.org/10.1103%2Fphysrevd.107.092011.

[94] A. Barr, F. Duncan, I. Lawson, K. McFarlane, and S. Venne, *Sl-sci-res-00-001-p snolab technical reference manual rev 0*, 2016.

[95] M. Browne, "Preparation for deployment of the neutral current detectors (ncds) for the sudbury neutrino observatory," PhD thesis, 1999.

[96] A. Rindi, F. Celani, M. Lindozzi, and S. Miozzi, "Underground neutron flux measurement," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 272, no. 3, 871–874, 1988, ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(88)90772-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0168900288907723.

[97] M. Pyle, "Optimizing the design and analysis of cryogenic semiconductor dark matter detectors for maximum sensitivity," Available at https://www.osti.gov/biblio/1127926, PhD thesis, Stanford University, 2012.

[98] R. Ren *et al.*, "Design and characterization of a phonon-mediated cryogenic particle detector with an ev-scale threshold and 100 kev-scale dynamic range," *Physical Review D*, vol. 104, no. 3, Aug. 2021, ISSN: 2470-0029. DOI: 10.1103/physrevd.104.032010. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.104.032010.

[99] B. Young, T. Saab, B. Cabrera, J. Cross, and R. Abusaidi, "Tc tuning of tungsten transition edge sensors using iron implantation," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 444, no. 1, 296–299, 2000, ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(99)01400-X. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016890029901400X.

[100] J. Hofer and N. Haberkorn, "Superconductivity in nanocrystalline tungsten thin films growth by sputtering in a nitrogen-argon mixture," *Thin Solid Films*, vol. 685, 117–122, 2019, ISSN: 0040-6090. DOI: https://doi.org/10.1016/j.tsf.2019.06.014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0040609019303839.

[101] A. H. Abdelhameed *et al.*, "Deposition of Tungsten Thin Films by Magnetron Sputtering for Large-Scale Production of Tungsten-Based Transition-Edge Sensors," *Journal of Low Temperature Physics*, vol. 199, no. 1-2, 401–407, Feb. 2020. DOI: 10.1007/s10909-020-02357-x.

[102] N. A. Kurinsky, "The low-mass limit: Dark matter detectors with ev-scale energy resolution," Jan. 2018.

[103] C. W. Fink, "A Gram-Scale low-Tc Low-Surface-Coverage Athermal-Phonon Sensitive Dark Matter Detector," Ph.D. dissertation, UC, Berkeley, 2022.

[104] S. M. Griffin, Y. Hochberg, K. Inzani, N. Kurinsky, T. Lin, and T. C. Yu, "Silicon carbide detectors for sub-gev dark matter," *Phys. Rev. D*, vol. 103, 075002, 7 Apr. 2021. DOI: 10.1103/PhysRevD.103.075002. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.103.075002.

[105] I. Alkhatib *et al.*, "Light dark matter search with a high-resolution athermal phonon detector operated above ground," *Phys. Rev. Lett.*, vol. 127, 061801, 6 Aug. 2021. DOI: 10.1103/PhysRevLett.127.061801. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.127.061801.

[106] D. W. Amaral *et al.*, "Constraints on low-mass, relic dark matter candidates from a surface-operated supercdms single-charge sensitive detector," *Phys. Rev. D*, vol. 102, 091101, 9 Nov. 2020. DOI: 10.1103/PhysRevD.102.091101. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.102.091101.

[107] S. L. Watkins, "Athermal phonon sensors in searches for light dark matter," Ph.D. dissertation, UC Berkeley, 2023. arXiv: 2301.08699.

[108] R. Agnese *et al.*, "First dark matter constraints from a supercdms single-charge sensitive detector," *Phys. Rev. Lett.*, vol. 121, 051301, 5 Aug. 2018. DOI: 10.1103/PhysRevLett.121.051301. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.121.051301.

[109] D. N. M. et al, "Snowmass2021-letter of interest the tesseract dark matter project," 2020. [Online]. Available: https://www.snowmass21.org/docs/files/summaries/CF/SNOWMASS21-CF1_CF2-IF1_IF8-120.pdf.

[110] R. E. Underwood, "Dark Matter Searches with Germanium and Silicon detectors in SuperCDMS and CUTE," Ph.D. dissertation, Queen's U., Kingston, 2022.

[111] P. Camus *et al.*, "The Cryogenic Underground TEst (CUTE) Facility at SNOLAB," Oct. 2023. arXiv: 2310.07930 [physics.ins-det].

[112] M. Be *et al.*, "Table of radionuclides (vol.3 - a = 3 to 244)," Jan. 2006.

[113] M. Berger, J. Hubbell, S. Seltzer, J. Coursey, and D. Zucker, *Xcom: Photon cross section database*, en, 2010.

[114] M. Isaac, V. Vanin, and O. Helene, "The internal bremsstrahlung following the electron capture decay offe," eng, *Zeitschrift für Physik A Atomic Nuclei*, vol. 335, no. 3, 243–246, 1990, ISSN: 0930-1151.

[115] M. Biavati, S. Nassiff, and C. Wu, "Internal bremsstrahlung spectrum accompanying 1s electron capture in decay of fe55, cs131, and tl204," eng, *Physical review*, vol. 125, no. 4, 1364–1372, 1962, ISSN: 0031-899X.

[116] SNOLAB Low Background Group, *Gamma-screening with hpge vda and analysis of: Cute fe-55 calibration sample*, 2021. [Online]. Available: https://www.snolab.ca/users/services/gamma-assay/vda/CUTE/V01/V01.html.

[117] C. W. Fink *et al.*, "Performance of a large area photon detector for rare event search applications," *Applied Physics Letters*, vol. 118, no. 2, Jan. 2021, ISSN: 1077-3118. DOI: 10.1063/5.0032372. [Online]. Available: http://dx.doi.org/10.1063/5.0032372.

[118] R. Germond, "Techniques and challenges in low-mass dark matter searches using CDMS style detectors," Ph.D. dissertation, Queen's U., Kingston, Queen's U., Kingston, 2023.

[119] SPICE/HeRALD Collaboration, *QETpy*, https://github.com/spice-herald/QETpy, 2023.

[120] E. Armengaud *et al.*, "Searching for low-mass dark matter particles with a massive ge bolometer operated above ground," *Physical Review D*, vol. 99, no. 8, Apr. 2019, ISSN: 2470-0029. DOI: 10.1103/physrevd.99.082003. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.99.082003.

[121] B. S. Xinran Li Matt Pyle, "Modeling the differential rate for dark matter interactions in coincidence with noise fluctuations," *To Be Submitted*, 2024.

[122] S. Yellin, "Extending the optimum interval method," Feb. 2008. arXiv: 0709.2701 [physics.data-an].

[123] S. C. SuperCDMS Collaboration, "Light dark matter search with the cpd-v2 detector," *To Be Submitted*, 2024.

[124] SPICE/HeRALD Collaboration, *pytesdaq*, https://github.com/spice-herald/pytesdaq, 2023.

[125] B. J. Kavanagh, "Earth scattering of superheavy dark matter: Updated constraints from detectors old and new," *Physical Review D*, vol. 97, no. 12, Jun. 2018, ISSN: 2470-0029. DOI: 10.1103/physrevd.97.123013. [Online]. Available: http://dx.doi.org/10.1103/PhysRevD.97.123013.

[126] A. Aguilar-Arevalo *et al.*, "Constraints on light dark matter particles interacting with electrons from damic at snolab," *Phys. Rev. Lett.*, vol. 123, 181802, 18 Oct. 2019. DOI: 10.1103/PhysRevLett.123.181802. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.123.181802.

[127] J. Wilson, *Notes on tes simulation*, SuperCDMS Internal Note, 2017. [Online]. Available: https:%20//confluence.slac.stanford.edu/download/attachments/120202114/notes-tes-simulation.pdf.

[128] A. C. Hindmarsh and L. R. Petzold, *Lsoda, ordinary differential equation solver for stiff or non-stiff system*, Sep. 2005.

Appendix A

Estimating Energy Lost to the Bath

A.1 Introduction

When using a TES to measure energy deposited in a substrate, the measurement is performed by observing a change in the Joule power of the TES. This doesn't account for all of the deposited energy since some will always be lost to the bath.

In this note I will estimate the magnitude of the energy dissipated to the bath in an idealized system. Note that I will assume the results of Irwin-Hilton [75] as a jumping off point.

A.2 Power Balancing in the TES

From Irwin-Hilton, we have that the temperature of the TES evolves according to the power balancing equation

$$C\frac{dT}{dt} = P_J - P_b + P_s(t) \tag{A.1}$$

Where P_J is the power absorbed by the TES due to Joule heating

$$P_J = I^2 R \tag{A.2}$$

We also have the electronic equation governing the evolution of the current

$$L\frac{dI}{dt} = V_b - I\big(R_L + R(T)\big) \tag{A.3}$$

P_b is the power emitted by the TES via cooling to the bath

$$P_b = \kappa(T^n - T_b^n) \tag{A.4}$$

To make our lives easier, let's assume $T_b \ll T_c$ so we can approximate

$$P_b \approx \kappa T^n \tag{A.5}$$

And $P_s(t)$ is the signal power absorbed by the TES. Let's make a few assumptions about the signal:

1. The particle interaction deposits energy E_{dep} in the substrate
2. The system has a net efficiency ϵ such that the energy absorbed by the TES is $E_{abs} = \epsilon E_{dep}$
3. The phonons are instantly absorbed by the TES at time $t = 0$

Under these assumptions, the signal power can be written

$$P_s(t) = \epsilon E_{dep}\delta(t) \tag{A.6}$$

In the absence of a signal, the TES will remain in a steady state such that the Joule power is balanced by the bath power

$$P_{J,0} = P_{b,0} \tag{A.7}$$

We can rewrite A.1 in terms of the deviation from these steady state values

$$\Delta P_J = P_J - P_{J,0} \tag{A.8}$$

$$\Delta P_b = P_b - P_{b,0} \tag{A.9}$$

$$\rightarrow C\frac{dT}{dt} = P_{J,0} + \Delta P_J - P_{b,0} - \Delta P_{b,0} + \epsilon E_{dep}\delta(t) \tag{A.10}$$

$$= \Delta P_J - \Delta P_{b,0} + \epsilon E_{dep}\delta(t) \tag{A.11}$$

Next we integrate over $(-\infty, \infty)$. We define

$$\Delta E_J \equiv -\int_{-\infty}^{\infty} \Delta P_J dt \tag{A.12}$$

$$\Delta E_b \equiv \int_{-\infty}^{\infty} \Delta P_b dt \tag{A.13}$$

And with this we have

$$\epsilon E_{dep} = \Delta E_b + \Delta E_J \tag{A.14}$$

Note that an increase in the temperature of the TES always results in a decrease in the Joule power, hence

$$\Delta E_J > 0 \tag{A.15}$$

If we have sufficient understanding of the circuit, ΔE_J can be calculated directly from an observed trace. Furthermore, if we have the case $\Delta E_b \ll \Delta E_J$, then $\epsilon E_{dep} \approx \Delta E_J$ and we can consider the system to be"self-calibrating" (at least insofar as it allows us to measure ϵ).

A.3 The Linear Response Case

Our goal is to estimate ΔE_b by evaluating the integral in Eq. A.13. Note that once we do this, we will also have an estimate of ΔE_J from Eq. A.14 AKA conservation of energy. To do this, we need an expression for $T(t)$ derived from the relevant system of ODEs. In the case of a impulse signal which doesn't drive the TES out of transition (quantitatively this is roughly $\epsilon E_{dep}/C \lesssim T_w$) the solution is given by Eq. (27) of Irwin-Hilton. We will simplify that result a bit further by assuming we're in the limit $\tau_+ \ll \tau_-$. In this case we have

$$\delta T(t) \approx \frac{\epsilon E_{dep}}{C} e^{-t/\tau_-} \tag{A.16}$$

For $t \geq 0$. I claim (here without justification) that neglecting the fast time constant can significantly change the pulse amplitude but has a negligible effect on the integrated energy.

As long as we remain in transition, we can typically use the linearized expansion of the bath power

$$\Delta P_b \approx G\delta T \tag{A.17}$$

Where

$$G \equiv n\kappa T_c^{n-1} \tag{A.18}$$

With this approximation the integration is trivial, and we find

$$\Delta E_b \approx \frac{\tau_-}{\tau_0}\epsilon E_{dep} \tag{A.19}$$

Where τ_0 is the natural thermal time constant of the system in the absence of electro-thermal feedback.

$$\tau_0 = C/G \tag{A.20}$$

In the limit of large loop gain ($\mathcal{L} \gg 1$) we have $\tau_-/\tau_0 \sim \mathcal{L}^{-1}$. This justifies the approximation that we typically employ when calibrating the detector in the unsaturated regime

$$\epsilon E_{dep} \approx \Delta E_J \tag{A.21}$$

Including the correction we've found here, we have rather

$$\Delta E_J \approx (1 - \frac{\tau_-}{\tau_0})\epsilon E_{dep} \tag{A.22}$$

The relevant ratio of the falltimes in the limit $\tau_+ \ll \tau_-$ is evaluated in Irwin-Hilton Eq. (29). Including this result, which also depends on the intrinsic current sensitivity β, the load resistance $R_L \equiv R_{par} + R_{sh}$ and the operating resistance R_0, we have

$$\Delta E_J \approx \frac{(1 - R_L/R_0)\mathcal{L}}{1 + \beta + R_L/R_0 + (1 - R_L/R_0)\mathcal{L}}\epsilon E_{dep} \tag{A.23}$$

A.4 The Saturated Response Case

Our analysis of the unsaturated case was pretty straight-forward, which in large part is due to the fact that Irwin and Hilton already did most of the hard stuff. To handle the saturated case we're on our own, so we might as well try to make the problem as easy as possible at first. The general saturated problem is complex for a few reasons

1. The unsaturated analysis relies on the assumption that the change in the TES resistance is proportional to change in temperature, which breaks down in the saturated regime
2. The TES is really an array of $\mathcal{O}(100-1000)$ individual elements arranged in parallel, each one in general has its own distinct temperature. This leads to situations where some elements remain in transition while others are driven out.

Note that these complications only apply to a kind of *intermediate* case where we have partial saturation. If enough heat is deposited in the array ($\epsilon E_{dep}/C \gg T_w$) the resistance of the array becomes arbitrarily close to the normal resistance (R_N) and no longer changes with temperature. This simplifies the problem by effectively decoupling the thermal and electric behavior of the system (electro-thermal feedback is turned off when sufficiently saturated). In this case we can rewrite the power balancing equation

$$C\frac{dT}{dt} \approx I_N^2 R_N - \kappa T^n \tag{A.24}$$

Where I_N is the current through the TES when driven normal

$$I_N = \frac{V_b}{R_L + R_N} \tag{A.25}$$

With this in mind, we'll proceed by sweeping the aforementioned complications under the rug and promise to come back to them later. To get around the temperature gradient in the TES array, we'll artificially assume that all of the individual elements share a common temperature.

To account for the nonlinear response of the TES, I propose the following crude model of the temperature response. Rather than trying to account for the intermediate region where the slope of the the resistance curve is changing, I approximate the response by stitching together the well-understood regions. In this model the temperature response is exactly linear between $[T_c - T_w, T_c + T_w]$ and perfectly flat outside this

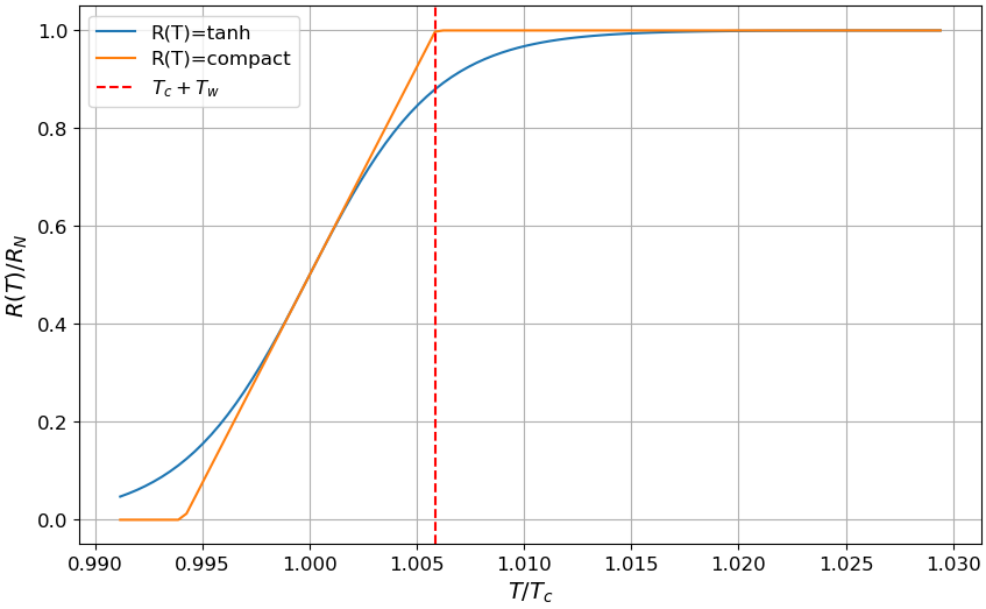


Figure A.1: Comparison of our "compact" TES model to the conventional "tanh" parametrization.

region. I will refer to this parametrization as the "compact" model of the TES.

Using this model with our assumptions of extremely fast current response and an impulse energy deposition, we arrive at the following picture of a saturated event

1. We have some energy deposition $E_{dep} > \frac{C}{\epsilon}(T_w + T_c - T_0)$ which quickly drives the TES out of transition.

2. While out of transition, the temperature evolves according to Eq. A.24 for some time t_{sat} at which $T(t_{sat}) = T_c + T_w$. Note that for $t < t_{sat}$ the change in Joule power is constant

$$\Delta P_{J,N} = P_{J,N} - P_{J,0} \equiv I_N^2 R_N - I_0^2 R_0 \tag{A.26}$$

3. For time $t > t_{sat}$, the temperature evolution is the same as the linear case. Hence we can use our previous result to find the energy lost to the bath during this time using Eq. A.19 with the replacement

$$E_{dep} \to E_{sat} \equiv \frac{C}{\epsilon}(T_w + T_c - T_0)$$

The key take away here is that the Joule power is constant while the device is saturated and already known.

Hence we can write the change in energy due to the change in Joule heating for a saturated event as

$$\Delta E_J = -\Delta P_{J,N} t_{sat} + (1 - \frac{\tau_-}{\tau_0})\epsilon E_{dep} \tag{A.27}$$

Where the only unknown is t_{sat}. Note that the negative sign in front of $\Delta P_{J,N}$ follows from our definition in Eq. A.12.

A.4.1 Saturation With Linear Bath Power

We were able to get a simple result for the energy transferred to the bath in the unsaturated case by assuming that the temperature deviation is small enough that the change in bath power can be approximated as linear (Eq. A.16). This approximation is valid as long as $\delta T \ll T_c$ and breaks down at larger temperature deviations. We can consider this the "barely saturated" case. With the approximation of linear bath power, the temperature when saturated evolves according to

$$C\frac{d\delta T}{dt} = \Delta P_{J,N} - G\delta T \tag{A.28}$$

It is straight forward to separate variables in Eq. A.28 and integrate to find t_{sat}

$$\frac{d\delta T}{\Delta P_{J,N} - G\delta T} = \frac{dt}{C} \tag{A.29}$$

$$\frac{t_{sat}}{C} = \int_{\epsilon E_{dep}/C}^{T_c+T_w} \frac{d\delta T}{\Delta P_{J,N} - G\delta T} \tag{A.30}$$

$$= \frac{1}{G}\ln\left(\frac{\Delta P_{J,N} - G\epsilon E_{dep}/C}{\Delta P_{J,N} - G(T_c + T_w)}\right) \tag{A.31}$$

$$\rightarrow t_{sat} = \tau_0 \ln\left(\frac{\epsilon E_{dep} - \tau_0 \Delta P_{J,N}}{E_{sat} - \tau_0 \Delta P_{J,N}}\right) \tag{A.32}$$

Recall $\Delta P_{J,N} < 0$. From this we see that the fundamental timescale in the saturated regime is the natural thermal time constant τ_0 (what else could it be!), and that the saturation time increases like the log of the deposited energy in this regime. Then our estimate of ΔE_J from Eq. A.22 is

$$\Delta E_J = -\tau_0 \Delta P_{J,N} \ln\left(\frac{\epsilon E_{dep} - \tau_0 \Delta P_{J,N}}{E_{sat} - \tau_0 \Delta P_{J,N}}\right) + (1 - \frac{\tau_-}{\tau_0})\epsilon E_{sat} \tag{A.33}$$

A.4.2 Saturation With Arbitrary Bath Power

If we want to consider an extremely saturated event, then we should relax our assumption that the actual temperature deviations are small compared to the critical temperature. Solving Eq. A.24 in general for $n \approx 4-6$ is not exactly trivial. Rather than focus on the particular solution it is perhaps more informative to investigate the limiting behavior in this case. Heuristically we can consider the following hypothesis:

"A larger energy deposition will create a pulse which saturates for longer"

This is clearly true in the case of Eq. A.33, but at first glance it is not obvious that this continues to arbitrarily large energies. To address this question I'll start be rewriting Eq. A.24 in terms of the characteristic time and temperature scales

$$C\frac{d\delta T}{dt} = P_{J,N} - P_{b,0}(1+\delta T/T_0)^n \tag{A.34}$$

$$= P_{b,0}\big(P_{J,N}/P_{b,0} - (1+\delta T/T_0)^n)\big) \tag{A.35}$$

$$\rightarrow \frac{d}{dt}\delta T = \frac{T_0}{n\tau_0}\big(q - (1+\delta T/T_0)^n)\big) \tag{A.36}$$

Where I've defined

$$q \equiv \frac{P_{J,N}}{P_{b,0}} = \frac{P_{J,N}}{P_{J,0}} = \frac{R_0+R_L}{R_N+R_L} \tag{A.37}$$

Stable operation of the device requires the magnitude of the Joule power to decrease as the TES temperature increases, hence

$$|P_{J,N}| < |P_{J,0}| \rightarrow 0 < q < 1 \tag{A.38}$$

Notice that in the limit $R_L \rightarrow 0$, q reduces to the bias point

Next I introduce the rescaled variables

$$T' \equiv \frac{\delta T}{T_0}, \quad t' \equiv t/\tau_0 \tag{A.39}$$

and the relevant separated ODE becomes

$$\frac{dT'}{q-(1+T')^n} = \frac{dt'}{n} \tag{A.40}$$

To integrate this, we make use of the fact that $q < 1$ and hence $q < (1+T')^n$, so the left side of (39) can be

written in terms of a geometric series as follows

$$\frac{1}{q-(1+T)^n} = -\frac{1}{(1+T')^n}\frac{1}{1-\frac{q}{(1+T')^n}} = -\frac{1}{(1+T')^n}\sum_{k=0}^{\infty}\frac{q^k}{(1+T')^{nk}} \tag{A.41}$$

With this we have

$$-dT'\sum_{k=0}^{\infty} q^k\big(1+T'\big)^{-n(k+1)} = \frac{dt'}{n} \tag{A.42}$$

$$\rightarrow t'_{sat} = -n\sum_{k=0}^{\infty} q^k\int_{T'_1}^{T'_{sat}} dT'\big(1+T'\big)^{-n(k+1)} \tag{A.43}$$

Where

$$T'_1 = \frac{\epsilon E_{dep}/C - T_0}{T_0} \tag{A.44}$$

$$T'_{sat} = \frac{T_c + T_w - T_0}{T_0} \tag{A.45}$$

Carrying out the integration

$$t'_{sat}(T'_1) = n\sum_{k=0}^{\infty}\frac{q^k}{1-n(k+1)}\big[\big(1+T'_1\big)^{1-n(k+1)} - \big(1+T'_{sat}\big)^{1-n(k+1)}\big] \tag{A.46}$$

If needed, Eq. A.46 can be used as an explicit numerical evaluation of the saturation time. We're interested in the limiting behavior of this expression. Making the approximation $T'_{sat} \ll 1$ and taking the limit $T'_1 \rightarrow \infty$ gives

$$t'_{max} \equiv \lim_{T'_1\rightarrow\infty} t'_{sat}(T'_1) = -n\sum_{k=0}^{\infty}\frac{q^k}{1-n(k+1)}\big(1+T'_{sat}\big)^{1-n(k+1)} \tag{A.47}$$

$$\approx -n\sum_{k=0}^{\infty}\frac{q^k}{1-n(k+1)}\big(1+(1-n(k+1))T'_{sat}\big) \tag{A.48}$$

$$= -n\sum_{k=0}^{\infty}\frac{q^k}{1-n(k+1)} - nT'_{sat}\sum_{k=0}^{\infty} q^k \tag{A.49}$$

$$= -n\sum_{k=0}^{\infty}\frac{q^k}{1-n(k+1)} - \frac{nT'_{sat}}{1-q} \tag{A.50}$$

With some help from Mathematica, the remaining sum evaluates to

$$S \equiv -n\sum_{k=0}^{\infty}\frac{q^k}{1-n(k+1)} = \frac{n}{n-1}\,{}_2F_1(1,\frac{n-1}{n};\frac{2n-1}{n};q) \tag{A.51}$$

Where ${}_2F_1$ is the hypergeometric function. From a computational perspective we've solved the problem, but I have very little intuition about ${}_2F_1$. So let's take one more step and find an approximate evaluation of the sum. To do this, we note

$$\frac{1}{n(k+1)-1} \leq \frac{1}{n-1}\frac{1}{k+1} \tag{A.52}$$

Hence

$$S \leq \frac{n}{n-1}\sum_{k=0}^{\infty}\frac{q^k}{k+1} = -\frac{n}{n-1}\frac{\ln(1-q)}{q} \tag{A.53}$$

For $q < 1$. We could put in some more legwork to rigorously demonstrate that this doesn't just serve as an upper bound, but a decent approximation - but if we compare to our numerical results it becomes clear from inspection. So we claim

$$S \lesssim -\frac{n}{n-1}\frac{\ln(1-q)}{q} \tag{A.54}$$

Then putting this together we have

$$t'_{max} \lesssim -n\big(\frac{\ln(1-q)}{q(n-1)} + \frac{T'_{sat}}{1-q}\big) \tag{A.55}$$

And plugging this into Eq. A.22 gives our estimate for the maximum possible Joule energy

$$\Delta E_{J,max} \lesssim n\tau_0\Delta P_{J,N}\big(\frac{\ln(1-q)}{q(n-1)} + \frac{T'_{sat}}{1-q}\big) + (1-\frac{\tau_-}{\tau_0})\epsilon E_{sat} \tag{A.56}$$

A.4.3 Simulation

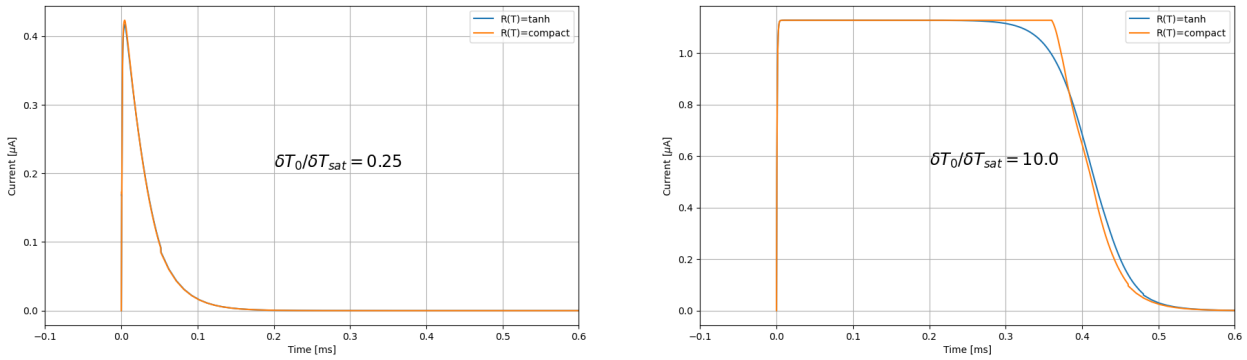
To verify that this result is valid under the given assumptions, we perform a simple pulse simulation to evaluate the energy partition. This is conceptually straightforward, numerically integrating a system of ODEs, in our case Eq. A.1 and Eq. A.3, is an elementary physics problem. Some care must be taken however since this particular is stiff. Following the development of the official SuperCDMS TES simulation [127] and use the **LSODA** solver to handle the integration [128]. To evaluate the validity of our "compact TES" model, we evaluate the system with both the compact and more conventional "tanh" parameterizations of the TES resistance.

The characteristic TES parameters used in the simulation are given in Table A.1. These are chosen to approximately represent CPD. for simplicity we assume $\epsilon = 1$ in the simulation.

We draw the impulse response pulses for each parametrization at different energy scales in A.2. When

Parameter	Value	Units
Normal Resistance	176	mΩ
Load Resistance	4.1	mΩ
Critical Temperature	51	mK
Transition Width	0.3	mK
Bath Temperature	13	mK
TES-Bath Conductance	46	pW/K
Inductance	130	nH
TES Heat Capacity	40	fJ/K
Bias Point (R_0/R_N)	0.5	–

Table A.1: TES parameters used in bath energy simulation



(a) (b)

Figure A.2: Two simulated pulses with different parametrizations of the TES resistance. The left plot shows an energy deposition which doesn't drive the TES into transition, while the right plot is well in the saturated regime. There are clear differences in the pulse shape between parametrizations which suggests that the compact model is not sufficient for describing the pulse shape

the TES remains in transition, there is relatively good agreement in the resulting pulse shape between different parametrizations. Once the pulses start to saturate, the differences between the models become more apparent.

For now, we just use Fig. A.2 to demonstrate that our simulation is giving something that looks like reasonable pulses. Though there are differences in the resulting pulse shapes when we introduce the compact model of the TES resistance, we haven't actually claimed that we expect them to agree in the time domain. Rather, what we want to demonstrate is that each model gives approximately the same result when we integrate the pulse to estimate the energy. We perform this integration over a wide range of energies.

We remark on some of the features observed in this comparison. First and foremost, there appears to be very good agreement in the ETF energy between the two models, which supports our claim about the utility

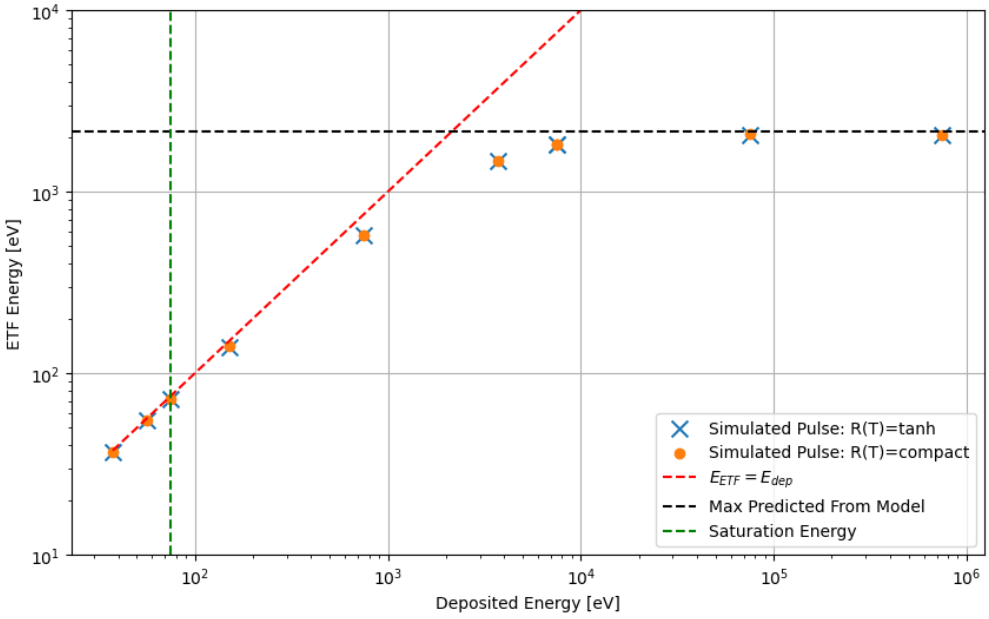


Figure A.3: Comparison of the energy removed by electrothermal feedback for each TES parametrization across a wide energy range. We see that the two models largely agree on the amount of energy removed.

of the compact parametrization. Since the TES spends a relatively small amount of time in the intermediate region between the middle of the transition and normal, it is a good approximation to simply neglect the non-linearity in this region for the purposes of estimating the integrated energy. As expected, we also see that the ETF energy is well approximated by the deposited energy in the limit of small deposited energy, with a slight correction arising to the non-zero response time of the TES. Our final observation is that the maximum ETF energy we observe is in fact less than, and on a similar scale to, our prediction Eq. A.56.

A.5 Local Saturation in the TES

A.5.1 Qualitative Description and Toy Phonon Model

Our discussion in the previous sections should give us a good description for a single fast TES. In reality, we don't look at a single TES, but rather an array of $\sim$ 1000 TES elements connected in parallel. The propagation of phonons from the detector bulk to the distributed array introduces position sensitivity to the detector. Recall that the mediating the phonon collection through Al fins makes our primary signal *athermal* phonons. To first order, athermal phonons propagate ballistically and isotropically across the

crystal originating from the interaction point[1]. Once these ballistic phonons reach the surface of the detector, we expect them to either be absorbed by the sensor or reflected. Upon reflection, some of the phonons will begin to thermalize, with the exact probability strongly depending on the density of crystal defects at the surface. The population of thermal phonons have insufficient energy to break Cooper pairs in the Al fin, but in principle can still be directly absorbed by the small coverage of W on the detector surface.

With this picture we propose a toy model for how the athermal phonons will be distributed in time across the TES array. In our next section we will take consider how the TES responds to this distributed power input to estimate how local saturation affects the detector response.

Consider a cylindrically symmetric detector with height h. For now, we'll only consider point like events which are sufficiently far from the edge of the detector such that we can model the radius as effectively infinite. Then let us call the speed of sound in the material ${v_s}^2$. The detector has a single instrumented surface, which has probability ϵ to absorb an athermal phonon to the TES.

Then suppose we have a point-like interaction which occurs at the surface of the detector, which expect to approximately describe and x-ray calibration event. For now, we will assume that all athermal phonons which reach a surface without being absorbed instantly thermalize and are lost to the bath. Then, assuming that the initial population of athermal phonons are generated isotrpically, the fraction which have reached the instrumented surface is simply

$$f(t) = \frac{1}{2}\big(1 - \frac{h}{v_s t}\big) \equiv \frac{1}{2}\big(1 - \frac{t_{min}}{t}\big) \tag{A.57}$$

Notice that this implies that half the phonons are always lost, since they reach the uninstrumented surface before the instrumented one. In our simple model, this population of phonons will trace out a circle on the instrumented surface, with time dependent radius given by

$$\rho(t) = \sqrt{(v_s t)^2 - h^2} = h\sqrt{(t/t_{min})^2 - 1} \tag{A.58}$$

[1]For a photoabsorption event (e.g. an x-ray from ^{55}Fe), the event really should be point-like. Higher energy interactions (e.g. muons) will create an extended track

[2]This should be a stand in for the *average* speed of sound in the material. In reality, each phonon mode has its own non-isotropic dispersion relation.

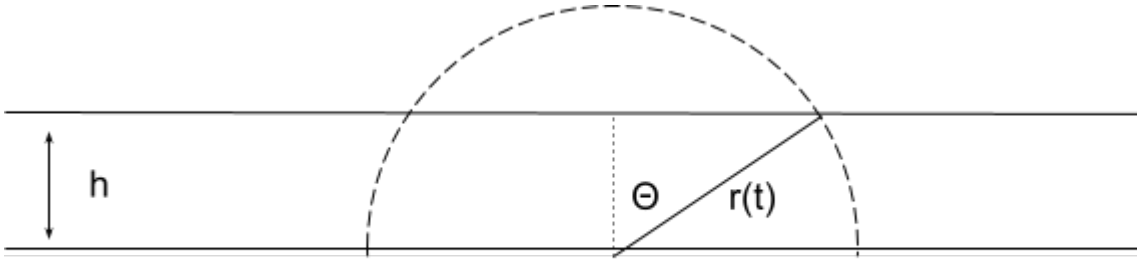


Figure A.4: Cartoon illustrating our toy phonon model. After time t, the athermal phonons have traveled a distance $r(t) = v_s t$. We are interested in the fraction which have crossed the surface a distance h from the interaction point, and their density along the surface.

With this, we can find the energy weighted radius of phonon depositions in the TES array

$$\langle \rho \rangle(t) = \frac{\int_{t_{min}}^{t} \frac{df}{dt'}(t')\rho(t')dt'}{\int_{t_{min}}^{t} \frac{df}{dt'}(t')dt'} \tag{A.59}$$

Where

$$\frac{df}{dt} = \frac{t_{min}}{2t^2} \tag{A.60}$$

and evaluating the integral in the denominator is simply

$$\int_{t_{min}}^{t} \frac{df}{dt'}(t')dt' = f(t) \tag{A.61}$$

The numerator is then

$$\int_{t_{min}}^{t} \frac{df}{dt'}(t')\rho(t')dt' = \frac{ht_{min}}{2} \int_{t_{min}}^{t} \frac{dt'}{t'^2} \sqrt{(\frac{t'}{t_{min}}^2) - 1} \tag{A.62}$$

$$= \frac{h}{2} \int_{1}^{\tau} \frac{d\tau'}{\tau'^2} \sqrt{\tau'^2 - 1} \tag{A.63}$$

Where $\tau \equiv t/t_{min}$. Then we make the substitution $\tau' = \cosh u$ and the integral becomes

$$= \frac{h}{2} \int_{0}^{\cosh^{-1}(\tau)} du \tanh^2(u) \tag{A.64}$$

$$= \frac{h}{2} \left[\cosh^{-1}(\tau) - \tanh\left(\cosh^{-1}(\tau)\right)\right] \tag{A.65}$$

$$= \frac{h}{2}\left[\cosh^{-1}(\tau) - \sqrt{1 - 1/\tau^2}\right] \tag{A.66}$$

Then altogether we have

$$\langle\rho\rangle(\tau) = \frac{h}{\tau - 1}\left[\tau\cosh^{-1}(\tau) - \sqrt{\tau^2 - 1}\right] \tag{A.67}$$

Notice that in the limit $\tau \to \infty$, this grows as $\langle\rho\rangle(\tau) \sim \ln(\tau)$, and hence grows without bound. This is a consequence of us ignoring the finite radius of the detector. Realistically, this sets an upper bound on the distance traveled by the phonons. Let's now introduce the finite radius R. So we have $\tau_{max} = t_{max}/t_{min} \lesssim R/h$. So over the course of the entire event we have

$$\langle\rho\rangle \sim h\big(\ln(2R/h) - 1\big) \tag{A.68}$$

For CPD, we have $h = 1$ mm and $R = 38$ mm, hence $\langle\rho\rangle \sim 3.3$ mm. Stated differently, we can roughly characterize the phonons across the entire event as being concentrated within this radius, which corresponds to just 0.8% of the area of the sensor array. Now that we've estimated an approximate length scale, let's estimate the fraction of the phonons which are absorbed within that radius. We define the characteristic time t_0 such that

$$\rho(t_0) = \langle\rho\rangle \tag{A.69}$$

We then find

$$t_0 \sim t_{min}\sqrt{\big(\ln(2R/h) - 1\big)^2 + 1} \tag{A.70}$$

Then the fraction of the phonons which have been collected at this time is

$$f(t_0)/f(\infty) \sim 1 - \left[\big(\ln(2R/h) - 1\big)^2 + 1\right]^{-1/2} \tag{A.71}$$

Plugging in our values for CPD, we find that of all the phonons which get absorbed, 58% are collected within the characteristic time.

Before moving on to discuss how we expect an array of TESes to respond to such a distribution of phonons, let's summarize the results of this section. First, we should acknowledge the shortcomings of our phonon model. Perhaps the most significant is that we assume phonons thermalize as soon as they reach a surface and are not absorbed. Realistically, the phonons may bounce several times of the detector walls

before thermalizing - though the exact distribution depends on details of the surface which are difficult to directly characterize. Taking this into account, we would expect the characteristic radius of the event to increase somewhat. Additionally, we expect the fact that phonons don't propagate isotropically and that the sensors are not positioned uniformly across the detector surface to alter our result. These complications motivate the need for an accurate detector Monte Carlo. But the argument presented in this section provides a useful estimate of the scale of the problem. For CPD, we expect $\mathcal{O}(50\%)$ of the recorded energy to be absorbed by only $\mathcal{O}(1\%)$ of the TES elements! From A.68, we see that this is behavior is dominated by the thickness of the detector. That is to say, the thin geometry of CPD results in increased local saturation, and for a full sized SNOLAB detector there should be a more uniform distribution of the energy across the array[3].

A.5.2 Response of a Partitioned TES

In the previous section we derived an approximate scale for local saturation effects. Next we will continue down the path of toy models, and estimate how an array of TES elements would evolve given asymmetric energy depositions on this scale. First, recall the basic system of ODEs governing the thermoelectric behavior of a single TES.

$$C\frac{dT}{dt} = I^2R(T) - \kappa T^5 \tag{A.72}$$

$$L\frac{dI}{dt} = V_b - I(R_L + R(T)) \tag{A.73}$$

Where C is the heat capacity of the TES, I is the current drawn from the source, $R(T)$ is the temperature dependent resistance of the TES, κ is the thermal conductance, V_b is the applied bias voltage, L is the inductance and R_L is the load resistance. We've made a few simplifying assumptions above, mainly that we can neglect the intrinsic current dependence of the TES resistance and that we can neglect the reduction in the bath power due to the non-zero temperature of the bath.

So let's sketch our model of the TES array. We will assume that our array is composed of N identical TES elements in parallel. The entire array should have the same heat capacity C, thermal conductance κ and normal resistance R_N as the single TES considered in the previous section. This implies that the

[3]Though keep in mind that we've assumed the detector is operated at $0V$. Our picture will change drastically if an external voltage is applied, since NTL phonons are not emitted isotropically to zeroth order

respective characteristic values for an individual element are C/N, κ/N and NR_N.

It's straightforward to show that we need $N+1$ differential equations in order to describe the behavior of an array composed of N individual TES elements - one for the temperature of each individual TES and an additional one to describe the total current. We expect our array to contain ~ 1000 elements, so if our goal is to gain intuition about the system it's unlikely that we can do so by analyzing a system of this many equations. Instead we will model local saturation effects in the array by assuming that the energy deposited by some incident particle is distributed equally among a individual TES elements (where $0 < a \leq N$), and the remaining elements absorb no energy. With this partition in mind, we can replace the array with two equivalent resistors composed of a and $N-a$ individual elements. This is of course a simplifying assumption. Even if we can reasonably approximate a large array of TES elements as being described by two temperatures, in general we expect the fraction belonging to each temperature group to change over the course of the event.

With this decomposition, we can write a system of three ODEs which describe the evolution of the dynamic variables. Introducing the parameter $x = a/N$ we have

$$xC\frac{dT_1}{dt} = I_1^2 R(T_1)/x - x\kappa T_1^5 \tag{A.74}$$

$$(1-x)C\frac{dT_2}{dt} = I_2^2 R(T_2)/(1-x) - (1-x)\kappa T_2^5 \tag{A.75}$$

$$L\frac{dI}{dt} = V_b - I(R_L + \frac{R(T_1)R(T_2)}{(1-x)R(T_1) + xR(T_2)}) \tag{A.76}$$

Where I_i, T_i denote the temperature and current through branch i and I is the current through the inductor. Applying Kirchhoff's laws we have $I = I_1 + I_2$ and $I_1\frac{R(T_1)}{x} = I_2\frac{R(T_2)}{1-x}$ and we can eliminate I_1, I_2 from the above system. For ease of notation we introduce the expression for the equivalent resistance of the two branch system at different temperatures

$$\mathbf{R}(T_1, T_2) = \frac{R(T_1)R(T_2)}{(1-x)R(T_1) + xR(T_2)} \tag{A.77}$$

With this we find that the system becomes

$$C\frac{dT_1}{dt} = I^2\frac{\mathbf{R}(T_1, T_2)^2}{R(T_1)} - \kappa T_1^5 \tag{A.78}$$

$$C\frac{dT_2}{dt} = I^2\frac{\mathbf{R}(T_1,T_2)^2}{R(T_2)} - \kappa T_2^5 \tag{A.79}$$

$$L\frac{dI}{dt} = V_b - I(R_L + \mathbf{R}(T_1,T_2)) \tag{A.80}$$

Notice that all the dependence on x has been stashed away in our definition of $R(T_1,T_2)$. Pretty neat! Now just as we did in the single TES case, we want to expand the nonlinear terms to first order. The only terms we haven't seen before involve the resistance of the equivalent array. Note that at equilibrium all the elements of the array should have the same temperature, and the problem of solving for the equilibrium value is the same as in the single TES case.

$$\mathbf{R}(T_0,T_0) = R(T_0) = R_0 \tag{A.81}$$

Then we can write the expansion of the equivalent resistance about equilibrium

$$\mathbf{R}(T_1,T_2) \approx R_0 + \frac{\partial \mathbf{R}}{\partial T_1}\delta T_1 + \frac{\partial \mathbf{R}}{\partial T_2}\delta T_2 = R_0 + \frac{\partial \mathbf{R}}{\partial \delta T_1}\delta T_1 + \frac{\partial \mathbf{R}}{\partial \delta T_2}\delta T_2 \tag{A.82}$$

Where $\delta T_i = T_i - T_0$ and the partial derivatives are evaluated at $\delta T_1 = \delta T_2 = 0$. To evaluate this expression, we rely on the typical expansion of the single TES. Note that we'll retain our approximate parametrization of the TES, where the response is exactly linear within the transition region.

Partition Without Saturation

As long as we neither TES is driven normal we have

$$\mathbf{R}(T_1,T_2) = R_0\frac{(1+z_1)(1+z_2)}{(1-x)(1+z_1)+x(1+z_2)} \tag{A.83}$$

where we've introduced for convenience

$$z_i = \alpha\frac{\delta T_i}{T_0} \tag{A.84}$$

Writing out the full evaluation of the derivatives is a bit tedious, but we arrive at a simple result which we probably could have guessed

$$\mathbf{R}(T_1, T_2) \approx R_0(1 + \frac{\alpha}{T_0}[x\delta T_1 + (1-x)\delta T_2]) \tag{A.85}$$

We also need to evaluate the resistance terms which appear in the thermal equation of each TES, and we find

$$\frac{\mathbf{R}(T_1, T_2)^2}{R(T_1)} \approx R_0(1 + \frac{\alpha}{T_0}[(2x-1)\delta T_1 + 2(1-x)\delta T_2]) \tag{A.86}$$

$$\frac{\mathbf{R}(T_1, T_2)^2}{R(T_2)} \approx R_0(1 + \frac{\alpha}{T_0}[2x\delta T_1 + (1-2x)\delta T_2]) \tag{A.87}$$

And we arrive at the linearized system

$$C\frac{d\delta T_1}{dt} = [(2x-1)\frac{\alpha P_0}{T_0} - G]\delta T_1 + 2(1-x)\frac{\alpha P_0}{T_0}\delta T_2 + 2I_0R_0\delta I \tag{A.88}$$

$$C\frac{d\delta T_2}{dt} = 2x\frac{\alpha P_0}{T_0}\delta T_1 + [(1-2x)\frac{\alpha P_0}{T_0} - G]\delta T_2 + 2I_0R_0\delta I \tag{A.89}$$

We can simplify this system by considering a change in variables to describe the mean temperature of the array and the temperature difference between the two components, rather that track each individual temperature

$$Y_1 = x\delta T_1 + (1-x)\delta T_2 \tag{A.90}$$

$$Y_2 = \delta T_1 - \delta T_2 \tag{A.91}$$

The system then transforms to

$$C\frac{dY_1}{dt} = (\frac{\alpha P_0}{T_0} - G)Y_1 + 2I_0R_0\delta I \tag{A.92}$$

$$C\frac{dY_2}{dt} = -(\frac{\alpha P_0}{T_0} + G)Y_2 \tag{A.93}$$

$$L\frac{dI}{dt} = -\alpha\frac{I_0R_0}{T_0}Y_1 - (R_L + R_0)\delta I \tag{A.94}$$

So we find that in the case where we partition the energy in the TES array such that neither branch is

driven normal, the two branches will reach a common temperature on a time scale independent of either the current or the mean temperature. Then the mean temperature couples to the current just as it does in the single TES case. Furthermore, the evolution of the system does not depend on our choice of partition. This makes sense! If no TES is driven normal then there are no saturation effects to worry about (local or otherwise) and the system is well described by the single TES model.

Partition With Saturation

If we assume that branch 1 is driven out of transition, then we have $R(T_1) \approx R_N$. Exploiting our hyperbolic trig identities, we can show that in this limit

$$\mathbf{R}(T_1, T_2) = \frac{R(T_1)R(T_2)}{(1-x)R(T_1) + xR(T_2)} \approx R(T_2 + \Delta T) \tag{A.95}$$

Where

$$\Delta T \equiv T_w \operatorname{arctanh}\left(\frac{x}{2-x}\right) \tag{A.96}$$

It is an interesting result that putting a fraction of the TES array out of transition introduces a new effective temperature scale which is determined by the transition width. For local saturation, we have argued that we expect $x \ll 1$, and hence

$$\Delta T \approx \frac{x}{2} T_w \tag{A.97}$$

To proceed, we apply the approximation

$$I \approx \frac{V_b}{R_L + \mathbf{R}(T_1, T_2)} \approx \frac{V_b}{R_L + R(T_2 + \Delta T)} \tag{A.98}$$

And the system governing the temperature evolution becomes

$$C\frac{dT_1}{dt} = \frac{V_b^2}{\left(R_L + R(T_2 + \Delta T)\right)^2} \frac{R(T_2 + \Delta T)^2}{R_N} - \kappa T_1^5 \tag{A.99}$$

$$C\frac{dT_2}{dt} = \frac{V_b^2}{\left(R_L + R(T_2 + \Delta T)\right)^2} \frac{R(T_2 + \Delta T)^2}{R(T_2)} - \kappa T_2^5 \tag{A.100}$$

Notice that the equation for T_2 is now independent of T_1, but this is only true while branch 1 remains out of transition. After it falls back into transition, ETF couples the two branches again to bring them back into

thermal equilibrium. Next, we notice that in the extreme case $R_L = 0$ both branches become uncoupled, and the evolution of each branch reduces to the case of the uniform TES in the respective saturated/unsaturated limits. Hence we argue that we can arrive at a rough estimate of the bath power lost by treating each branch as a uniform TES array, in which case the bath power is given by our previous results.

A.6 Discussion

We have presented some encouraging results here, but we should take stock of all the simplifying assumptions we've made before reading too much into them.

1. We have completely neglected critical current effects
2. We have assumed that there is a single thermal conductance between the TES in the bath. A more realistic model would include an intermediate "absorber" which is coupled to both the TES and the bath, and hence there are two relevant conductances.
3. We have only considered the impulse response case where the deposited energy is instantly absorbed by the TES. In reality, we expect the energy collection to occur on (at least one) non-zero time scale.
4. For the partitioned TES, we assumed that there is a single partition in temperature which is constant in time. In reality, we might expect many, infinitesimal partitions, and the fraction of elements belonging to each may vary over time.
5. We assumed that the TES array is described by a single critical temperature T_c, while in reality there is likely a gradient in T_c across the array.

These simplifications mean that it is premature to directly compare our results to real data. Despite this, there are still a number of useful take-aways that we can take from this discussion.

1. $E_{ETF} \approx \epsilon E_{dep}$ is a good approximation for a fast TES which remains in transition, with corrections occurring on the scale τ_{ETF}/τ_0.
2. At higher energies, there is an upper limit on E_{ETF}. We can estimate that this is on the scale of $E_{ETF,max} \sim \tau_0 \Delta P_{J,N}$
3. The compact model of the TES resistance gives good agreement with the conventional parametrization when calculating E_{ETF}.

4. When we introduce the partition to the TES array, there should be no change in the energy lost to the bath when neither branch leaves transition.

5. When one branch does leave transition, it introduces a new effective temperature scale ΔT into the evolution of the system.

With all of these considerations, there is much we leave to future work. One can go back to the simplest conclusions and evaluate the effect of relaxing our simplifying assumptions. We can also approach the problem from the other direction, and study local saturation effects by comparing our simple analogs between the partitioned array and uniform TES to results from the more complete SCDMS DMC.

Appendix B

The Constant Flux Problem

B.1 Introduction

The total interaction rate (in Hz) in our detector under the assumption $v_E = 0$, $v_{esc} = \infty$ is

$$R_0 = \frac{2}{\sqrt{\pi}} N_T \frac{\rho_\chi}{m_\chi} \sigma_{0,N} v_0 \tag{B.1}$$

Where N_T is the total number of targets contained within our detector. Note that $\sigma_{0,N}$ is the DM-*nucleus* cross-section. It's easier to compare different targets by looking at the DM-*nucleon* cross section, σ_0. For a detector made of a species with atomic number A, the two cross-sections are related by

$$\sigma_{0,N} \approx A^2 \sigma_0 \tag{B.2}$$

$$\rightarrow R_0 \approx A^2 \frac{2}{\sqrt{\pi}} N_T \frac{\rho_\chi}{m_\chi} \sigma_0 v_0 \tag{B.3}$$

Note that this is just a rough estimate - - we're playing fast and loose with galactic velocities and the coherence enhancement. But the order of magnitude is robust. Now let's use B.3 to estimate the rate we expect to see in the UMass CPD analysis for a typical point on our sensitivity curve. We have a 10.6 g Si detector, and are aiming for sensitivities on the scale $m_\chi \sim \mathcal{O}(100 \text{ MeV/c}^2)$, $\sigma_0 \sim \mathcal{O}(10^{-30} \text{ cm}^2)$. Then,

taking $\rho_\chi = 300$ MeV/c^2, $v_0 = 230$ km/s, we expect

$$R_0 \approx 1.4 \times 10^4 \text{ Hz} \tag{B.4}$$

This is a huge rate! Now, the mass is sufficiently light that very few of these interactions will appear above our trigger threshold - so our DM search is still a rare event search. However, with our noise-boosting formalism we aim to characterize our detector response by injecting synthetic signal events into our background, which we assume is free. This rate suggests that our background would in fact rampant with DM if we take the models we're probing at face value. We might not be able to resolve these interactions, but they may collectively have some measurable effect on our observed noise or the steady-state current of the TES. We should then take some care to make sure our framework is self-consistent.

At the limit we want to set in the CPDv2@UMass analysis, there is a huge total interaction rate ($\sim 10^5$ Hz) in our detector. Note that this isn't the triggered rate, most of these interactions are below threshold and we only expect a small fraction to trigger. But these persistent interactions in the detector would act as a constant source of power into the TES. So how can we accurately interpret our baseline if there is an external power source?

Let's evaluate whether the power produced by the DM is enough to actually throw us off. We already have our estimate of the rate from (B.3). Then we estimate the average energy per interaction as

$$E_0 \approx \frac{1}{2}\frac{m_\chi^2}{m_T} v_0^2 \tag{B.5}$$

Where m_T is the mass of the target nucleus and we've evaluated the recoil energy in the limit $m_\chi \ll m_T$. With these two pieces we can evaluate the average power input to the detector by DM scattering

$$P_{DM} = \epsilon R_0 E_0 \tag{B.6}$$

Where M_T is the total mass of the detector and ϵ is the net energy efficiency. So altogether we have

$$P_{DM}(m_\chi, \sigma_0) = \frac{\epsilon}{\sqrt{\pi}} A^2 \rho_\chi N_T \frac{m_\chi}{m_T} \sigma_0 v_0^3 \tag{B.7}$$

Consider that our DM is characterized by parameters $m_\chi = 100$ MeV/c^2, $\sigma_0 = 10^{-30}$ cm^2. With this we

estimate the power in our detector as

$$P_{DM} \approx 0.4 \text{ fW}\left(\frac{m_\chi}{100 \text{ MeV/c}^2}\right)\left(\frac{\sigma_0}{10^{-30} \text{ cm}^2}\right) \tag{B.8}$$

Where we've assumed $\rho_\chi = 0.3$ GeV/c^2/cm^3, $v_0 = 230$ km/s and $\epsilon = 25\%$. This power and the associated total interaction rate can have profound implications for our DM search, depending on how we carry out the analysis. Triggering for pulses in our data requires us to accurately characterize our noise, which we characterize by collecting random triggers. But given a large enough rate, we would expect the DM signal to contaminate every random period of the data. Hence our DM signal would be closely entwined with the noise.

For the CPDv2@UMass analysis, we have attempted to account for this before we had even considered this specific problem. When constructing the noise power spectral density (PSD), which is our characterization of the noise used to trigger on the data and perform the optimal filter (OF) fit to each event, we aggressively remove any fluctuations in the randoms which resemble a pulse[1], the implementation of which is referred to as the "no-pulse cut" (NPC). But even in this case, we could run into certain contradictions in our analysis. If we expect every random to have a large fluctuation, we would similarly expect every random to fail the NPC - hence we could never construct an estimate of the noise for DM models with sufficiently large rate.

For now, let's ignore the subtleties of the NPC, since characterizing the effect requires us to make some assumptions about the noise in the absence of the DM signal. Let us instead focus on how the DM would appear in our PSD.

B.2 Estimating a PSD from One or Many Signals

Let's take a simplified look at how pileup events appear when we look at their power in the frequency domain. Before we even start, let's look at a single normalized (in amplitude) pulse with a single fall time τ and the length of the trace is T. Let us also assume that $\tau \ll T$.

$$s_1(t) = \Theta(t) \exp\left(-t/\tau\right) \tag{B.9}$$

[1] Arguably, considering that the amplitude of the fluctuations we're removing are small compared to our baseline resolution - we're not really evaluating how much they resemble a pulse. We're really just throwing out pulses where a sufficiently large fluctuations are observed.

Where Θ is the Heaviside step function and τ is the fall time. When we say "PSD", we really mean the normalized magnitude of the Fourier transform

$$j_1(f) \equiv \frac{1}{T}\left|\tilde{s_1}(f)\right|^2 \tag{B.10}$$

Where T is the length of the trace. We will use $j(f)$ to indicate the "normalized" PSD derived from the template, with units $1/Hz$, and $J(f)$ for the PSD describing actual pulses in units A^2/Hz. Now we use the following convention for our transform

$$\tilde{s_1}(f) \approx \int_{-\infty}^{\infty} dt\Theta(t)e^{-(i2\pi f+1/\tau)t} \tag{B.11}$$

$$= \int_{0}^{\infty} dt\Theta(t)e^{-(i2\pi f+1/\tau)t} \tag{B.12}$$

$$= \frac{\tau}{1+i2\pi f\tau} \tag{B.13}$$

Hence

$$j_1(f) = \frac{1}{T}\frac{\tau^2}{1+(2\pi f\tau)^2} \tag{B.14}$$

Next, consider that we have a series of N pulses (all normalized to unit amplitude) in our trace, distributed randomly and uniformly over the interval at start times $(t_1, t_2, ..., t_N)$

$$s_N(t) = \sum_{j=0}^{N} \Theta(t-t_j)\exp\left(\frac{1}{\tau}(t-t_j)\right) \tag{B.15}$$

$$\tilde{s_N}(f) = \int_{-\infty}^{\infty} dt e^{-i2\pi ft}\sum_{j=0}^{N} \Theta(t-t_j)\exp\left(\frac{1}{\tau}(t-t_j)\right) \tag{B.16}$$

$$= \sum_{j=0}^{N} e^{t_j/\tau}\int_{t_j}^{\infty} dt\exp\left(-t(i2\pi f+1/\tau)\right) \tag{B.17}$$

$$= \frac{\tau}{1+i2\pi f\tau}\sum_{j=0}^{N} e^{-i2\pi ft_j} \tag{B.18}$$

Then

$$\left|\tilde{s_N}(f)\right|^2 = \frac{\tau^2}{1+(2\pi f\tau)^2}\left|\sum_{j=0}^{N} e^{-i2\pi f t_j}\right|^2 \tag{B.19}$$

The sum over the exponential is just a random walk in the complex plane with unit step size, so we know

$$\langle\left|\sum_{j=0}^{N} e^{-i2\pi f t_j}\right|^2\rangle = N \tag{B.20}$$

$$\rightarrow \left|\tilde{s_N}(f)\right|^2 \approx N\frac{\tau^2}{1+(2\pi f\tau)^2} \tag{B.21}$$

And from this we conclude

$$j_N(f) = N \cdot j_1(f) \tag{B.22}$$

This is an idealized approximation, since we really estimate the PSD from a discrete FFT rather than a continuous Fourier tranform over all time, and the pulses vary in amplitude. But this argument is sufficient to tell us how the power supplied by the DM scales in our PSD as a function of the interaction rate.

B.3 The PSD Induced by Dark Matter

Now, let's use the result of the previous section to construct a PSD from our DM signal. For any particular set of characteristic DM parameters (m_χ, σ_0), we can write the total interaction rate $R_0(m_\chi, \sigma_0)$ and mean recoil energy $E_0(m_\chi)$ as we did in (B.3) and (B.5) respectively. Then, given a trace length T, we expect $\langle N\rangle = R_0 T$ pulses to occur in any random trigger. The mean amplitude of these pulses will be

$$\langle A\rangle = \frac{\epsilon E_0}{\frac{dP}{dI}\int_{-\infty}^{\infty} s(t)dt} \tag{B.23}$$

Here, $s(t)$ is the normalized signal template, which should be empirically determined from the data. ϵ is the net energy efficiency, or the fraction of the energy deposited in the crystal which can actually heat the TES. Then $\frac{dP}{dI}$ is the power to current ratio, which characterizes the effective gain of our detector. We won't make explicit use of it here, but it's useful to keep in mind that for sufficiently small signal amplitudes, we have

$$\frac{dP}{dI} \approx V_b - 2I_0R_L \tag{B.24}$$

where V_b is the bias voltage, I_0 is the steady-state current and R_L is the load resistance.

With the number of pulses in the trace and average amplitude, we can use (B.22) to construct the expected PSD (in units A^2/Hz).

$$J_{DM}(f) \approx 2\langle N\rangle\langle A\rangle^2 \frac{1}{T}|\tilde{s}(f)|^2 = 2\langle N\rangle\langle A\rangle^2 j_1(f) \tag{B.25}$$

Where $j_1(f)$ is the PSD estimated from the normalized signal template. We acknowledge two shortcomings in our derivation of this result

1. We calculated the PSD in the continuous case, while our actual data is discrete.
2. We approximated all the DM events as having a fixed amplitude, rather than following a continuous spectrum.

Then

$$J_{DM}(f) = 2(R_0 T)\left[\frac{\epsilon E_0}{\frac{dP}{dI}\int_{-\infty}^{\infty} s(t)dt}\right]^2 j_1(f) \tag{B.26}$$

$$= 2\frac{T}{R_0}(P_{DM})^2\big(\frac{dP}{dI}\big)^{-2}\frac{j_1(f)}{\left[\int_{-\infty}^{\infty} s(t)dt\right]^2} \tag{B.27}$$

Let us perform a Monte Carlo simulation to evaluate the plausibility of the hypothesis that we can approximate the PSD induced by the DM signal by scaling the PSD of a single pulse.

At this stage of our study, we will not model any of the noise components independent of the DM signal. Our simulation follows the following steps

1. Specify our detector parameters to match those describing CPDv2. These include: T, $\frac{dP}{dI}$, ϵ, $s(t)$, $M_{detector}$.
2. Specify characteristic parameters m_χ, σ_0 of the DM model we want to simulate.
3. Calculate the mean number of events $\langle N\rangle$ which occur in the interval T given our DM model.
4. Draw the number of pulses N to simulate from a Poisson distribution with $\lambda = \langle N\rangle$.
5. Sample N amplitudes from the spectrum of our DM model, along with uniformly distributed random start times across the interval T. Add together N templates scaled and shifted according to these values to construct a sample pulse.

6. Repeat steps 4-5 a number of times (in our case, say 100) to build a collection of simulated random triggers. Calculate the PSD from this collection as we would with random triggers in real data[2].

7. Repeat steps 2-6 for a number of different DM models, to confirm that the PSD scales with the interaction rate and mean energy as expected.

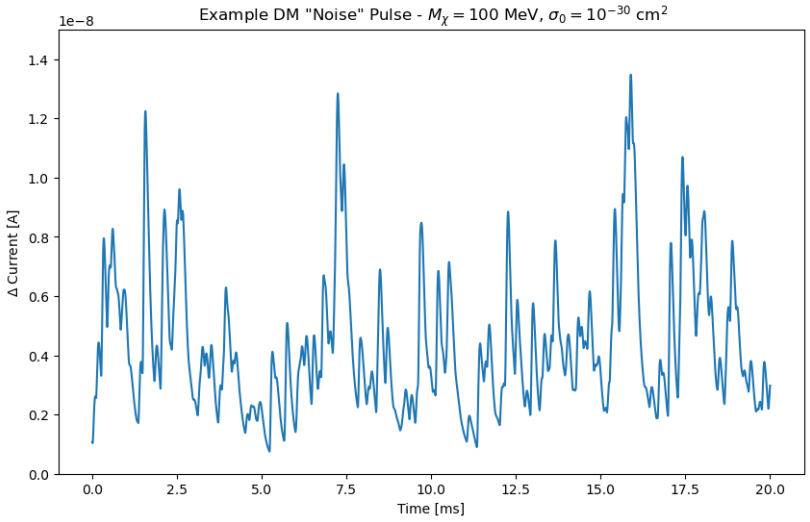


Figure B.1: Trace generated for our MC simulation. In addition to the numerous pileup pulses from the DM present in the trace, we also observe a net deviation from the steady-state baseline

Looking at our comparison between the scaled and simulated PSDs, we note that the shapes agree well below ~ 10 kHz, which roughly corresponds to our signal bandwidth

Now, let us consider how we might actually use this to set a limit on the DM interaction. Consider that we estimate some PSD $J_0(f)$ from our actual data. We want to integrate this in order to estimate the number of DM events which could be present in the PSD given a value of m_χ. Since the experimental PSD is likely contaminated by electronics noise at high frequency, we carry out this integral over some frequency range $[-f_0, f_0]$ which is determined by our signal bandwidth. Then our estimated observed number of events N_0 is can be written

$$N_0(m_\chi) = \frac{1}{2}\left[\frac{\frac{dP}{dI}\int_{-\infty}^{\infty} s(t)dt}{\epsilon E_0(m_\chi)}\right]^2 \frac{\int_{-f_0}^{f_0} J_0(f)df}{\int_{-f_0}^{f_0} j_1(f)df} \tag{B.28}$$

Given the estimated number of events, we can set a 90% upper limit on the cross section by choosing σ_0

[2]Well, not exactly the same yet. For real data, we apply a number of quality cuts to our randoms in an attempt to remove pulses and glitches. At this stage we don't apply these cuts to our simulation, mostly because this step won't give us much meaningful information unless we also simulate the intrinsic detector noise.

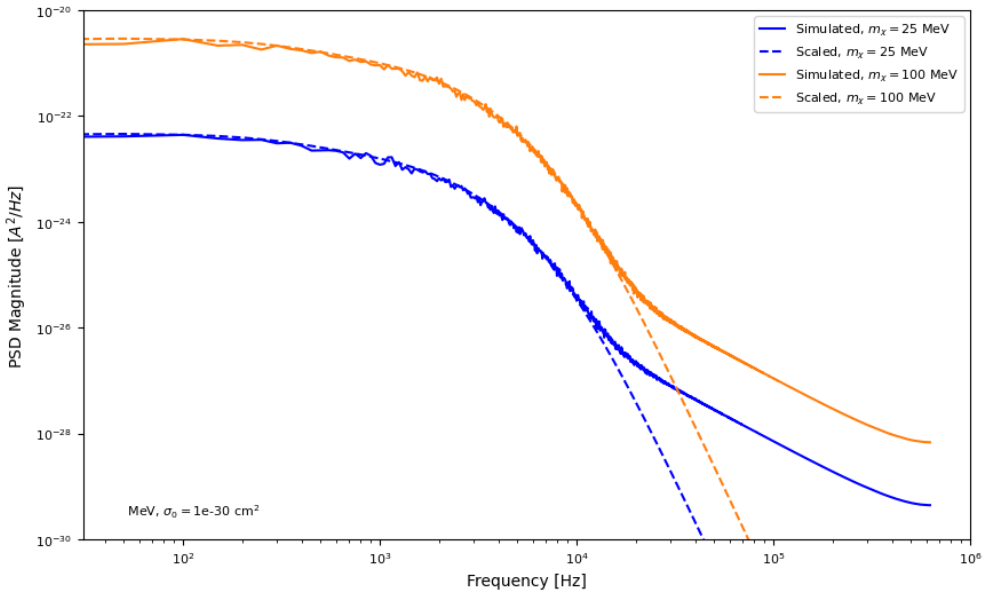


Figure B.2: PSD induced by frequent DM interactions from our simulation (solid) and the scaling estimate (dashed). We make the comparison for two different masses of LDM at a fixed cross section. We see good agreement between the simulation and the scaling law within our signal bandwidth < 10 kHZ

such that

$$\langle N \rangle (m_\chi, \sigma_0) \leq N_0(m_\chi) + 1.28\sqrt{N_0(m_\chi)} \tag{B.29}$$

Some care must be taken in that our above argument assumes that we have a signal acceptance efficiency of unity. In reality, when we think of our PSD as a measurement of the number DM events, there is some sub-unity efficiency introduced by the cuts we apply to remove pileup events. We leave the study of this effect and the resulting sensitivity estimation to future work.

Appendix C

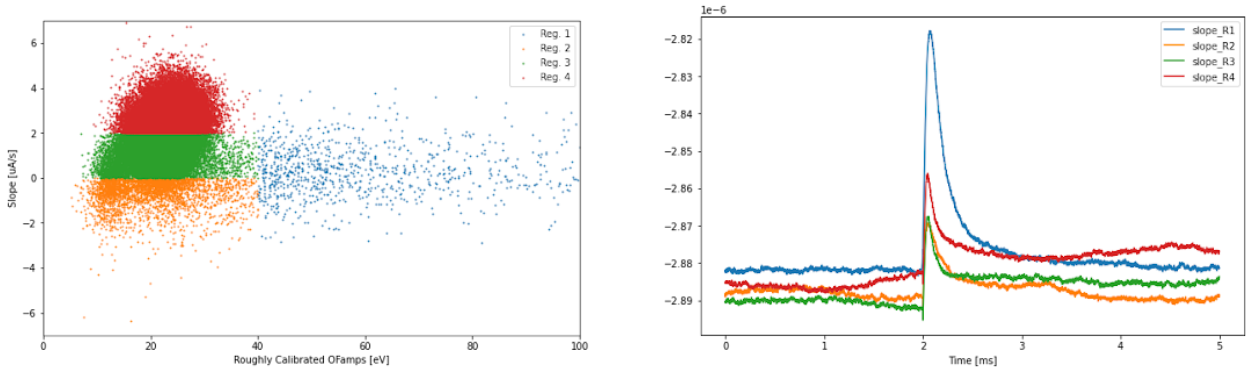
UMass R26 Low Frequency Noise Template

C.1 Introduction

This appendix describes a study of the low frequency noise observed in the DM search data collected during UMass fridge run 26 (R26). Many of the noise issues were resolved in later runs, in large part due to a heroic effort by Doug Pinckney to wrap sensitive electronic elements in foil. The favorable noise conditions in later runs made R28 a better candidate for performing a DM search, and the R26 analysis was subsequently put on hold. Regardless of the ultimate fate of the R26 data, I believe that the results of these studies are still relevant for the following reasons:

1. These studies motivated our choice to adopt a longer 20 ms trace length for the R28 analysis.
2. The techniques employed in this study can in principle be used to identify low frequency noise in other contexts, which is a topic of historical [27] (and likely future) significance for SuperCDMS.

In the course of studying the pulse slope we found a population of events near threshold with elevated slope. The slope distribution of these events overlaps with the slope of randoms. While they could be removed with a simple cut on the slope, this will necessarily come with a penalty on the efficiency. If we define a template can be which describes the glitch better than it does a good signal event, we could remove



(a) (b)

Figure C.1: (Left) Distribution of slope parameter with energy near threshold in R26. All events shown pass pre-pulse baseline cut. Note that a 5 ms trace length was used in this run. A clear blob of elevated slope events can be seen, which overlaps with the band of good events extending to higher energy. (Right) Averaged pulses from each corresponding region in the left plot. It is appears that the elevated slope parameter corresponds to an offset in the baseline after the trigger is issued. With this picture, we can't distinguish whether this offset corresponds to a lower baseline before the trigger or an higher baseline after.

these events without hurting our efficiency by comparing the resulting χ^2 when fitting each template. When we look at the distribution of these events in time, we see that they arrive in 13 Hz intervals.

C.2 Template Construction

Since the experiment is performed with continuous readout, we can look at a relatively long section of data ($\sim$ 1 s) and attempt to characterize any features which appear at 13 Hz. As shown in Fig. C.2, there are regular downward fluctuations in the trace detectable by eye. The feature is reasonably well described by a train of negative-going square waves with period 13 Hz and an on-time of 5ms.

Fig. C.2 demonstrates that we can take a 1s trace and fit a phase to a test signal with our assumed shape. We use this as a starting point in order to collect a number of suspected glitches to average into a template. We start by looping over 1s segments of the data between hours 20-21 of data taking. A peak fitting procedure is run on each segment, and the trace is rejected if any pulses or baseline steps are found. If the segment is found to be pulse and step free, we fit for the phase of the assumed pulse train. Then we extract a 25 ms region centered around each square pulse as our suspected glitch.

After detecting the edges, we can evaluate the width of each suspected glitch. We find that there is a well defined Gaussian distribution centered at 5ms with a flat background. We define events within 3σ of

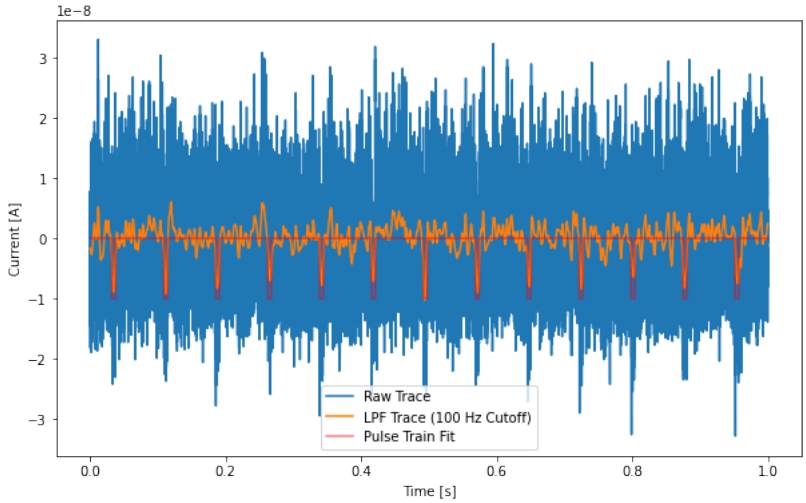
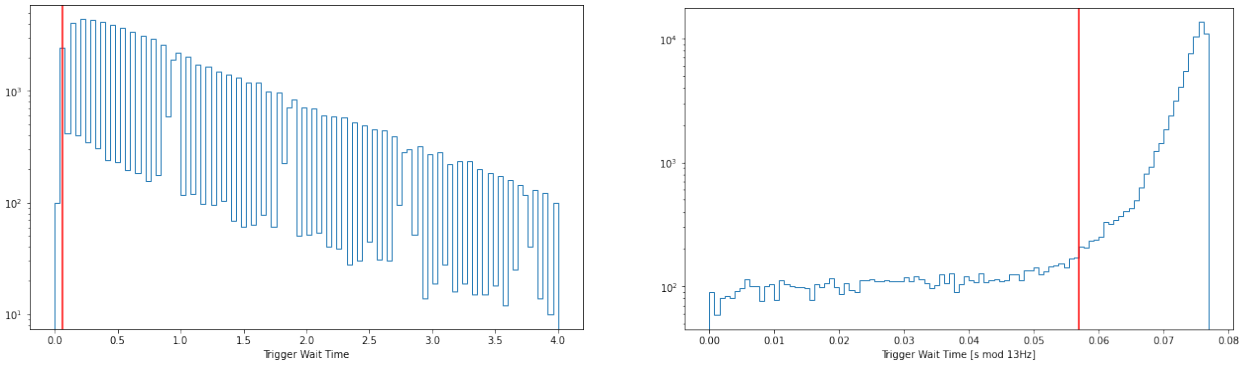


Figure C.2: Example 1s trace with 13Hz pulse train fit.



(a) (b)

Figure C.3: Δt distribution observed in UMass R26. (Left) shows the raw distribution, with many peaks visable. (Right) shows the the distribtuion modulo 13 Hz, where there is a single peak visable.

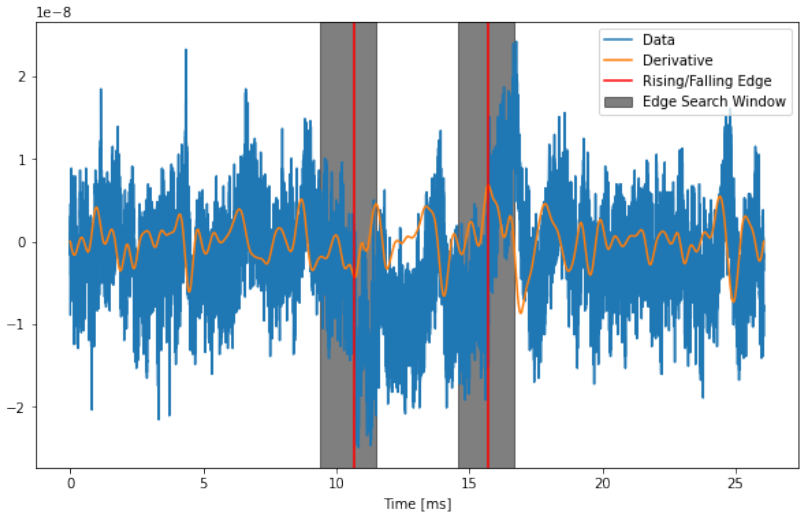


Figure C.4: Edge detection on a suspected glitch. We find the edges by the extrema in the Gaussian derivative filter applied to the trace, which allows us to estimate the width of the glitch.

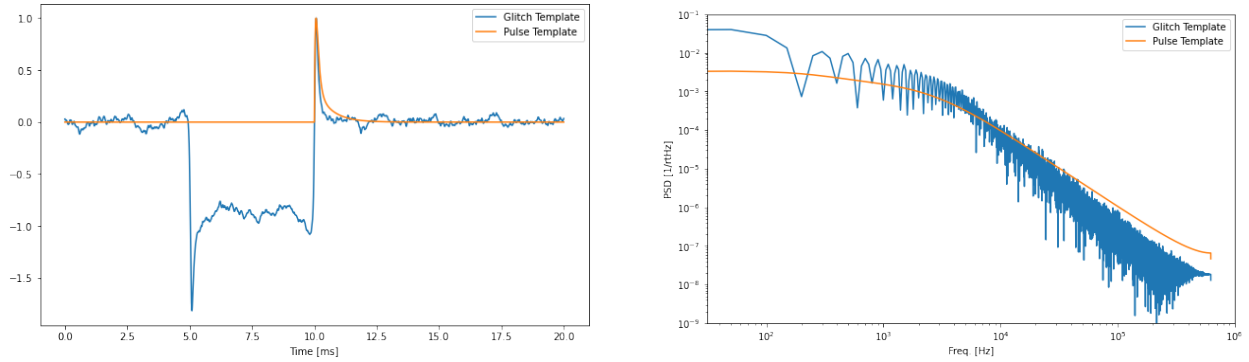
the Gaussian peak to be examples of actual glitches, and only include these when averaging the traces.

Taking the traces with well-defined width, we now align one of the detected edges to a defined location. We choose to align the rising edges, since this is presumably the edge which gets triggered on. This is an assumption, and one can potentially resolve a small rising bump preceding the falling edge. A future study could consider aligning the falling edge to determine if the resulting template shape is significantly

C.3 Alternate Construction Algorithm

We note that this is in some sense the easiest implementation of this analysis on account of having continuous readout. We learn the time-scale from the inter-pulse wait time distribution, and then we can identify the pulse train by eye from simply looking at a long section of data. We can also imagine applying a similar technique to the case where we have a regular (or semi-regular) train of low frequency noise pulses in data where we have an online trigger. To do this, we propose the following algorithm:

1. Make a collection of random triggers.
2. Apply QETpy.autocuts() to remove traces which contain a pulse, or have bad slope/skewness.
3. Pad ends of each trace with zeros to prevent wraparound effects when we shift the traces.
4. Define a “Running Mean” that we will add each trace to, and will eventually become our template.



(a) (b)

Figure C.5: Reconstructed glitch template resulting from aligning leading edge of the suspected glitches. We also compare to the signal template. (Left) shows the comparison in the time domain, and (Right) in frequency.

5. Go through each trace and calculate the cross correlation between it and the Running Mean. Where the cross correlation is maximal defines the delay to be applied to the trace.

6. Shift the trace by the delay and add it to the Running Mean.

The idea here is that, if the low frequency noise has a consistent shape, we our shifting of the traces will act as constructive interference for this noise source while the underlying uncorrelated noise will average out destructively. Of course, if there is more than one correlated low frequency noise component, the resulting template will be some superposition of these. The validity of this template can be tested by evaluating the χ^2 of triggered pulses and comparing it to usual χ^2 of the signal template. If there really is an underlying low frequency noise present in the data and the constructed template is valid, we expect to see distinct peaks in the $\Delta\chi^2$ distribution. This is a similar procedure to what we've done for UMass R26, but is more general in the sense that it doesn't require any prior knowledge of the timing distribution or access to arbitrarily long trace lengths

C.4 Results

Once we have a template, we can attempt to discriminate between events triggered on glitches and good pulses by performing an OF fit with both our glitch and pulse templates and looking at the difference in chi2 between the fits. We define this $\Delta\chi^2$ quantity as $(\chi^2_{glitch} - \chi^2_{pulse})$. Since a smaller χ^2 indicates a better fit, a more positive $\Delta\chi^2$ indicates a more pulse-like event following this convention.

In order to accommodate the 5ms width of the glitch, we use a longer trace length. Here we choose a 20ms trace with 10ms pretrigger length, so that the entire glitch is located in the pretrigger.

We define the cut as a simple line in the no-delay OF-energy vs $\Delta\chi^2$ plane which separates the glitch blob from the band of good events. Noting the intercept in this definition, it's clear that this cut will only discriminate between pulses and glitches for triggered events, and will not have discrimination power for randoms. We can consider a non-linear cut boundary where the curve passes through the origin in an attempt to remove randoms containing the glitch, but note that our chi^2 evaluation relies on the rising edge of the glitch being at the center of the trace, which we don't expect to apply to randoms.

At this point we have two cuts designed to remove the 13 Hz glitches, the $\Delta\chi^2$ cut defined above and the glitch slope cut described in 20ms Slope Cuts. With two cuts performing the same function, it's natural to assess whether both are necessary or if the problematic events can be removed with a single cut. One simple way to assess this is to apply both cuts but change the order in which they're applied, and observe how the spectrum changes. We test this in Fig 13. In each case we first apply the baseline cut and general slope cut. In the top plot the glitch slope cut is applied before the $\Delta\chi^2$ cut and in the bottom we swap this order. When applying the glitch slope cut first, we see a remnant shoulder in the spectrum ~ 20 eV which is subsequently removed by the $\Delta\chi^2$. On the other hand, when we apply $\Delta\chi^2$ first the shape of the spectrum is not noticeably affected by the glitch slope cut. This suggests that the cuts are redundant, with the $\Delta\chi^2$ giving the best performance.

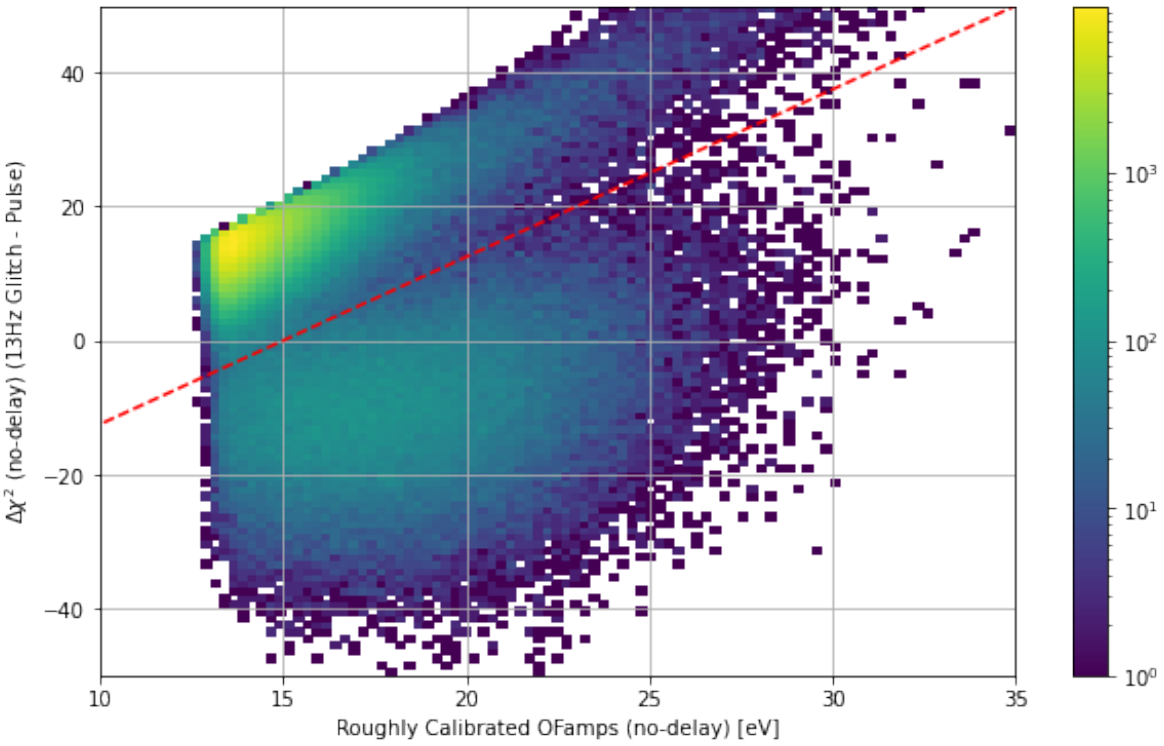


Figure C.6: Definition of the proposed glitch cut. The blob of glitch events is clearly centered below the dashed line.

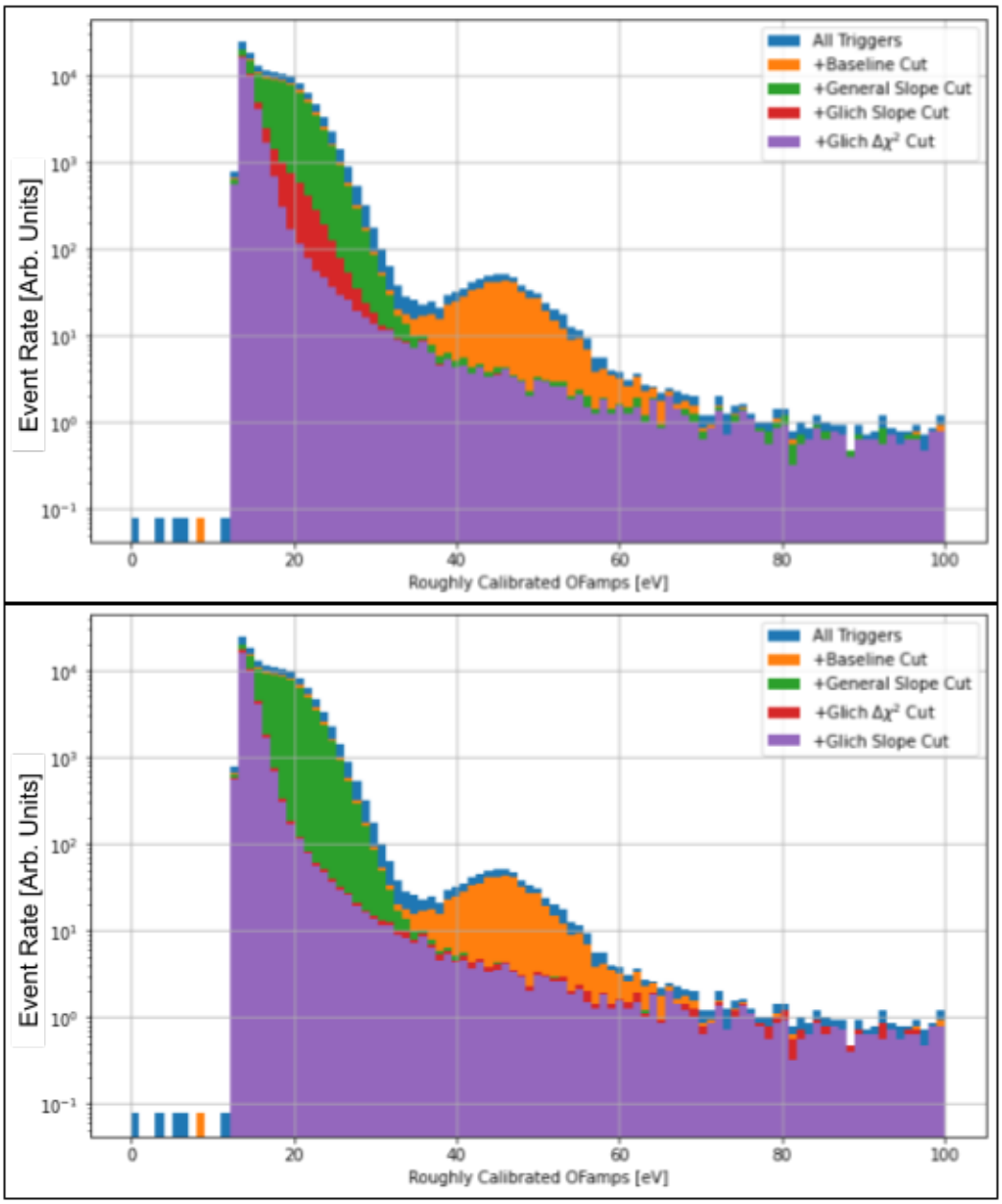


Figure C.7: Effect of $\Delta\chi^2_{glitch}$ cut on R26 spectrum. We compare this to the "glitch slope" cut, which is effectively a tighter version of our conventional slope cut to more aggressively target glitches. We see that the glitch slope cut removes very few events preserved by the $\Delta\chi^2_{glitch}$ cut, but the converse is not true. Furthermore, the $\Delta\chi^2_{glitch}$ cut to have a higher efficiency than its counterpart.