SIMULATING THE SUPERCDMS DARK MATTER DETECTOR RESPONSE AND READOUT

By

ROBERT AGNESE

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2017

To my wife, Brandi

ACKNOWLEDGMENTS

Over the course of my PhD program I have had the opportunity to meet and work with many amazing individuals. Many of them were part of the SuperCDMS collaboration, some were my classmates, some were family, some friends, and some were staff and faculty here at the University of Florida. There are so many people who have assisted me along the way and I sincerely apologize if I leave anyone out of this acknowledgment.

First I would like to thank my wife, Brandi Phillips. Without her support it is very unlikely that I would have been able to complete this journey. She single-handedly has allowed (sometimes forced) me to try to strike a healthy work-life balance.

I would also like to thank Dr. Tarek Saab for being my advisor. His patience was much appreciated while I sometimes wrestled with concepts that later would seem obvious to me. I'm sure that sometimes it was frustrating for him, but he never let it show. His advice was always spot-on, even if I was too stubborn to heed it fully. His management style very much allowed for the freedom to explore a problem in whatever way I chose- allowing me to learn the invaluable lessons that come from such explorations, even if it wasn't always the quickest way to the solution.

As part of the incoming physics grad class of Fall 2010, I feel incredibly lucky. We had an hilarious and unique set of characters in our class- a feeling that I'm sure is felt by every class, but which I'm equally sure is only true for ours. I would like to thank everyone with whom I commiserated over coursework, who helped me with homework, or who invited me and Brandi over for board games or movies.

I would also like to thank my long-term lab mate, Brad Welliver, who graduated a little more than one year before me. His encyclopedic knowledge of the SuperCDMS experiment constantly amazed me. While I sometimes struggled to remember what "iZIP" stood for, Brad would be able to tell me every detector that had some short or excessive noise (and on which channel), how every analysis quantity was calculated, exactly how the

triggering logic looked, and just about anything one could want to know about MATLAB. Brad's wealth of knowledge was both inspiring and incredibly useful.

In addition, I need to thank the staff at the University of Florida Physics Department. In particular I'd like to acknowledge Jay Horton. While, officially, he was there to help us repair and maintain equipment in the lab, I found his amusing stories to be very valuable to my morale. I'd also like to acknowledge Pam Marlin who had to constantly nag me to get my paperwork in on time each semester. I'm sorry, Pam.

Finally I would like to thank the members of my family who gave me love, encouragement, and support during this time of my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

SIMULATING THE SUPERCDMS DARK MATTER DETECTOR RESPONSE AND READOUT

By

Robert Agnese

May 2017

Chair: Tarek Saab
Major: Physics

Dark matter is one of the remaining great mysteries of modern physics. Cosmological and astrophysical evidence hints strongly at a new particle, or set of particles, unknown to the Standard Model, comprising some 80% of all of the matter in the universe.

SuperCDMS is one of the world-leading direct-detection experiments attempting to directly measure interactions between dark and regular matter. SuperCDMS uses solid state crystalline detectors in a nearly-zero background environment to attempt to measure the particle properties of dark matter.

The detectors themselves, however, are quite complex; and a very detailed understanding of the microscopic physics is helpful in analyzing the very rare events that occur within them. Furthermore, better understanding and modeling of the detectors can aid in the design and optimization of future iterations of the experiment.

The SuperCDMS detectors are well-understood in ideal cases. However, an accurate understanding of the detector response to outlier events or under non-ideal conditions is important for a low-statistics experiment.

This work describes the design and implementation of a low temperature condensed matter physics simulation library built on top of the popular Geant4 particle tracking framework. The library, named "Geant4 Condensed Matter Physics" or G4CMP, introduces several solid state concepts to the Geant4 framework such as crystal lattices, phonon

quasiparticles, non-scalar effective masses, and implements several physics processes relevant to cryogenic temperature crystals.

In addition to the physics library, which is intended for general use, this work also describes a full Monte Carlo simulation package for the SuperCDMS iZIP detectors which utilizes G4CMP at its core and also fully simulates the detector readout sensors.

The document will begin with the canonical motivations for the cold dark matter hypothesis, followed by a description of the SuperCDMS experiment. It will then describe the microscopic physics inside of the SuperCDMS detectors, as well as details on their implementation in G4CMP. The full detector simulation with corresponding sensor readout will follow, along with plans for future work; including both scientific studies and improvements, as well as computational improvements and added efficiencies.

CHAPTER 1
EVIDENCE AND MEASUREMENTS OF DARK MATTER

1.1 Introduction

Evidence for the existence of a particle-like dark matter is vast and well covered in the literature. The intention of this chapter is to provide a brief overview of some of the canonical results and arguments for the cold dark matter hypothesis.

The mystery around dark matter is not in its existence, but rather in that its particle properties must be unlike any known Standard Model particle. In fact, dark matter is well-cemented into the standard model of cosmology: the ΛCDM model, where the "CDM" stands for "Cold Dark Matter" [1].

In the following sections, some of the intriguing astrophysical and cosmological evidence for dark matter will be explored. That evidence will lead to the conclusion that dark matter is a new fundamental particle with a particular set of properties.

1.2 Astrophysical Evidence

Credit for the first dark matter hypothesis is often attributed to Fritz Zwicky [2]. In the 1930s, Zwicky was studying the movement of galaxies in the Coma cluster. He calculated the total mass of the galactic cluster in two ways: first, he calculated the velocities of the galaxies within the cluster and applied the virial theorem to calculate the mass; he also estimated the mass by using the luminosity of the cluster and an estimated count of the number of galaxies within. The former calculation returned a much larger amount of mass than the latter (about two orders of magnitude difference). If the majority of mass in a galaxy was from stars and cosmic gases, the numbers should have been quite close. Zwicky's conclusion from this discrepancy was that there must be a significant amount of mass in the galactic cluster that is "dark" and thus not included in the luminosity calculation.

1.2.1 Galactic Rotation Dynamics

Just like Zwicky's calculations on galactic clusters, individual galaxies also seem to have much more mass than is directly observable. This can be concluded by looking at the rotational velocity profile. If the angular motion of the stars in a galaxy obey Newtonian dynamics, it is easy to relate the mass profile of the galaxy to its velocity profile,

$$v(r) = \sqrt{\frac{GM(r)}{r}}, \tag{1–1}$$

where G is the universal gravitational constant, r is the distance from the center of the galaxy, and $M(r)$ is the mass distribution of the galaxy.

If the mass profile is roughly constant at high radius, as it appears to be from luminosity measurements, then we would expect the speed of the galaxies to be $v(r) \sim r^{-1/2}$. However, measurements reveal that this is not the case, as seen in Figure 1-1.

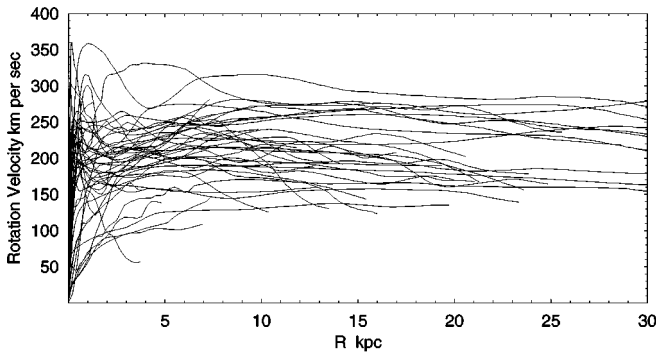


Figure 1-1. Average rotational speed vs. distance from galactic center for a sample of spiral galaxies [3].

Thus, it appears that there must be more mass hidden throughout these galaxies, contributing to the total mass profile without being detectable by various spectroscopic techniques. The needed shape of these mass distributions is often described as a "halo." A successful model for the dark matter distribution in galaxies is given by the Navarro-Frenk-White (NFW) profile [4],

$$\rho(r) = \frac{\rho_0}{\frac{r}{R_s}\left(1 + \frac{r}{R_s}\right)^2}, \tag{1–2}$$

where R_s is a parameter that needs to be tuned to a particular halo. Figure 1-2 shows the contributions of dark matter halos to the velocity profiles of several galaxies.

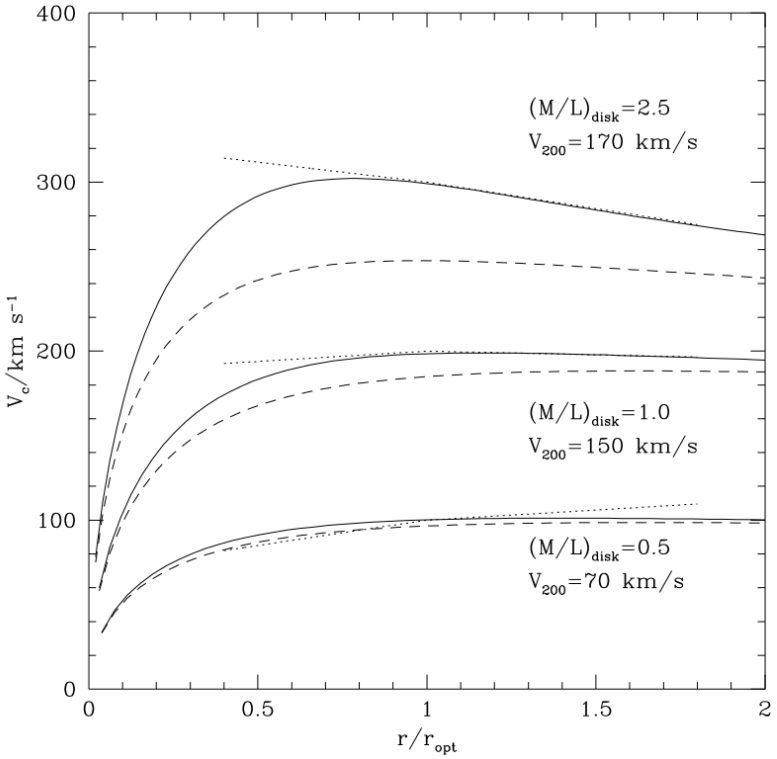


Figure 1-2. Rotation curves for several galaxies. The y-axis is the circular velocity and the x-axis is the distance from the center of the galaxy, normalized to the radius that contains most of the luminous matter. M/L is the mass to luminous matter ratio and v_{200} is the circular velocity at a distance that defines a sphere encompassing a mean overdensity factor of 200. The dotted lines are measured data, dashed lines are the NFW halo, and the solid black lines are the halo + luminous mass disk [4].

From this evidence we can conclude that dark matter particles must not interact via the electromagnetic force- otherwise, these particles would be visible to our telescopes. This also implies that dark matter is electrically neutral. Dark matter is also non-relativistic. If dark matter had been traveling at relativistic speeds in the early universe, it would have dissipated before being able to clump together to form galaxies. Dark matter must also be stable on astrophysical timescales.

1.2.2 Gravitational Lensing

Another piece of evidence of dark matter comes from gravitational lensing measurements. The quintessential case study is the Bullet Cluster. The Bullet Cluster is actually one of two colliding galactic clusters, though we often refer to both of the clusters together as "the Bullet Cluster." As seen in Figure 1-3, the center of gravitational mass of each cluster is displaced from the center of baryonic mass as seen from X-rays.

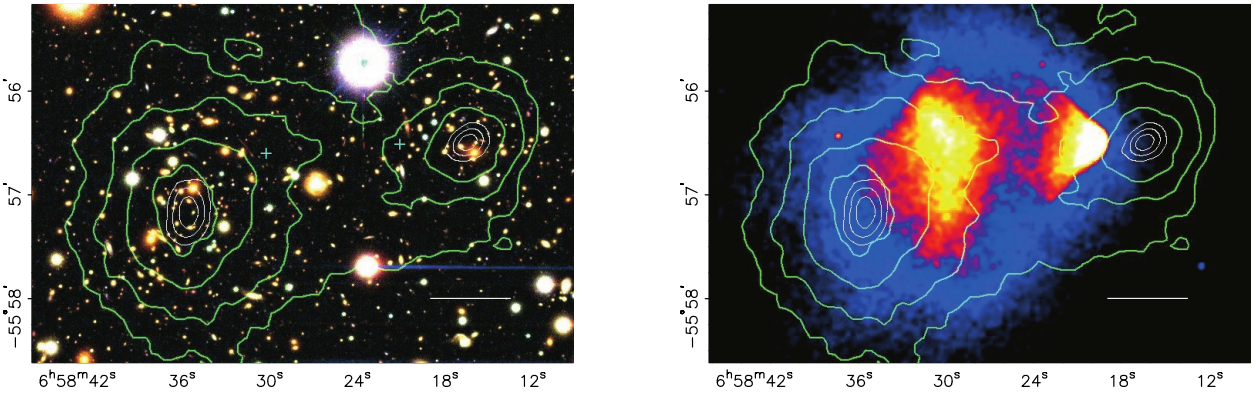


Figure 1-3. Two images of the Bullet Cluster. On the left: Color image from Magellan. On the right: Chandra X-ray image. The green contours are the weak lensing contours which show the concentration of mass from gravitational lensing. The Magellan image shows the center of mass is mostly coincident with the visible distribution of galaxies. The Chandra image shows the baryonic plasma is well separated from the centers of masses of the clusters [5].

The separation of the two centers is explained by the baryonic matter being much more likely to interact than the dark matter component of each cluster. Thus, as the clusters pass through each other, the dark matter moves unimpeded, while the baryonic gas is greatly slowed due to electromagnetic interactions. This separation of the baryonic component from the gravitational center of mass is difficult to explain with a modified gravitation theory.

From the unimpeded motion it appears that dark matter must be weakly interacting.

1.3 Cosmological Evidence

Support for the ΛCDM model of cosmology is strong. Much of the science that has gone into studying and verifying the ΛCDM also has the direct effect of validating the specific claims it makes about dark matter. Numerical simulations show that the ΛCDM model will reproduce the observed large scale structure of the universe [6]. Likewise, baryon acoustic oscillation (BAO) measurements from the spectroscopic Sloan Digital Sky Survey (SDSS) also show strong agreement with the ΛCDM model when analyzing the clustering of galaxies. And the cosmic microwave background (CMB) is yet another way to see evidence of the effects that dark matter has had in the formation of our universe. While all of these topics are quite exciting, for the sake of brevity only two of them will be expanded in this section.

1.3.1 Cosmic Microwave Background

The cosmic microwave background is the remnant noise from the early formation of the universe. Its study provides valuable insight into what must have occurred in the very early universe. In the first moments after the Big Bang, most of the energy in the universe was in the form of a very hot ionized plasma. As the universe expanded, it also cooled. This cooling caused much of the plasma to clump together into hydrogen atoms, allowing leftover photons to escape the plasma and spread, nearly uniformly, throughout the ever-expanding universe.

As the universe continued to expand, these photons became more and more red-shifted, and today have an average energy that corresponds to a temperature of about 2.73 Kelvins. The CMB is nearly uniform. However, small anisotropies reveal hints at asymmetries in the processes that were occurring in the very early universe.

The most precise observation of the CMB at the time of this writing is from the Planck telescope, which released its expanded data results in 2015. One analysis tool for the CMB anisotropies is to expand the power spectrum of the CMB in spherical

harmonics. The peaks that occur above a multipole moment of $\ell \sim 100$ are caused by competing effects.

In the early universe, baryonic matter experienced gravitational attraction, which would cause the particles to be drawn closer together and clump. That same matter also experienced radiation pressure from high energy photons in the plasma, which would cause over-dense regions to smooth out. Dark matter would not experience the latter effect, and thus would only clump. The ratio of the second and third peaks in Figure 1-4 indicates the relative abundance of non-baryonic matter to baryonic matter.

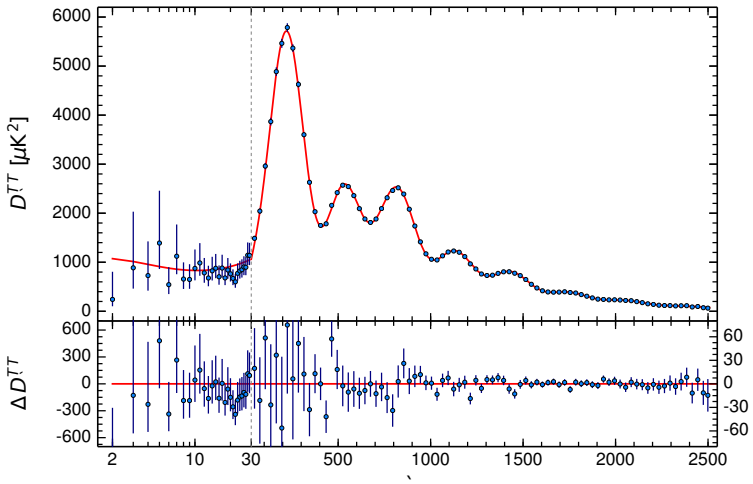


Figure 1-4. The CMB power spectrum multipole expansion. Blue dots are data from the Planck mission. The red line shows the best-fit parameters for the ΛCDM model, indicating very strong agreement [7].

The Planck CMB study further validates the ΛCDM model, which claims that the majority of matter in the universe is non-baryonic. In fact, "regular," non-dark matter makes up only $\approx 5\%$ of the mass-energy, as shown in Figure 1-5.

1.3.2 Thermal Relic

One attractive theoretical candidate for dark matter is the "Weakly Interacting Massive Particle" or WIMP. For Supersymmetry models that conserve R-parity, the lightest supersymmetric particle (LSP) is a natural WIMP candidate as long as it is also

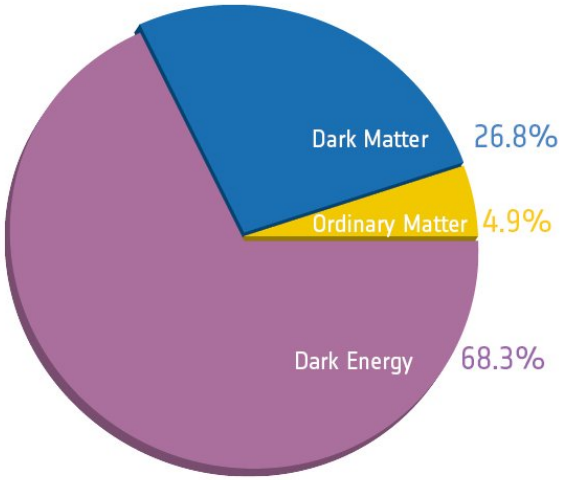


Figure 1-5. Distribution of mass-energy in the universe. Most of the universe's energy is in forms that are not well understood. Data from [7].

neutrally charged, such as a neutralino. The R-parity conservation means that the LSP would not be able to decay into Standard Model particles.

A particle of this nature and with the right interaction cross-section could "freeze out" in the early universe at roughly the right time (temperature) needed to provide the correct amount of dark matter we observe in the universe today. This is often referred to as the "WIMP miracle."

In the very early universe, when the temperature was very hot, the dark matter would be in thermal equilibrium, meaning that the rate of annihilation ($\chi\bar{\chi} \Rightarrow X\bar{X}$) is equal to the rate of production ($X\bar{X} \Rightarrow \chi\bar{\chi}$). Since the WIMP is heavier than standard model particles, as the universe cools it becomes less likely that particle-antiparticle pairs would be energetic enough to produce WIMP-antiWIMP pairs. But the WIMP-antiWIMP pairs would still be able to annihilate into standard model particles, causing an asymmetry ($\chi\bar{\chi} \nRightarrow X\bar{X}$). Eventually the universe expands enough that the WIMP-antiWIMP pairs can no longer annihilate and the WIMP is said to have "frozen out." Once frozen out, the relative abundance of the WIMPs is constant. It turns out that such a particle that had a mass in the range of 1–1000 GeV would freeze out to today's measured relic abundance if

it has a velocity-averaged interaction cross-section on the same order as most weak-scale interactions (as opposed to strong force or EM interactions) as seen in Figure 1-6.

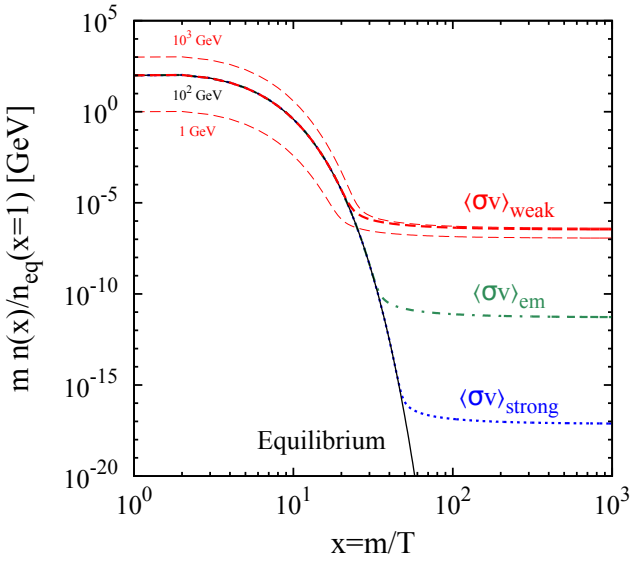


Figure 1-6. The evolution of the abundance of a 100 GeV WIMP for different cross-section interaction scales. The solid black line is the abundance if the WIMP was always in thermal equilibrium. Extra dashed-red lines are given for M = 1 GeV and M = 1000 GeV [8].

1.4 WIMP Dark Matter Properties

From the previous sections it can be concluded that dark matter particles very likely have certain properties:

- Electrically neutral
- Non-baryonic
- Non-relativistic
- Stable

The WIMP makes an excellent candidate for dark matter. The goal of the Super-CDMS collaboration is to directly measure WIMP particles. The typical hypothesized WIMP candidate is usually the lightest supersymmetric (SUSY) particle, which would satisfy the stability requirement, as long as it is difficult to decay into Standard Model particles. Being a supersymmetric particle, it must also be very massive, since it has not

yet been detected by particle collider experiments. Being very massive would satisfy the non-relativistic requirement. The non-baryonic condition is obviously satisfied by being a fundamental particle. And being electrically neutral simply depends on the details of the SUSY model being used, but is satisfied by being the hypothesized neutralino.

1.5 Dissertation Roadmap

The remainder of this work describes the creation of a software library for adding condensed matter physics to the Geant4 simulation framework (G4CMP) and an application that utilizes said library to simulate the dark matter detectors used by the SuperCDMS experiments (G4DMC).

To that end Chapter 2 will describe the SuperCDMS experiment. This involves a discussion of the local dark matter kinematics, the backgrounds expected in the experiment, and the detector's response to an event. To understand the response, an overview of the detector design and the solid state physics that occurs inside the detector volume in response to a recoil event. A description of the sensor readout that converts those microscopic physics into an electronic signal is also given. While most of the detector design discussion will focus on the previous generation of the experiment, Chapter 2 will conclude with an explanation of the upgrades planned for the next generation experiment at SNOLAB.

With the background knowledge from Chapter 2, Chapter 3 will present a more detailed discussion of the condensed matter physics that is relevant to SuperCDMS, as well as how they are implemented in the G4CMP library. This will require some detailed understanding of Geant4- the framework on which G4CMP depends, so the chapter will begin with a brief discussion of Geant4's design and nomenclature.

Chapter 4 will discuss the proof-of-concept detector Monte Carlo simulation based on G4CMP called G4DMC. Whereas G4CMP describes the condensed matter physics, G4DMC models the SuperCDMS detector phonon and charge sensors while utilizing G4CMP for the condensed matter physics that occurs inside the detector.

The final chapter includes a small description of some of the ongoing work in G4CMP, G4DMC, and the overall detector Monte Carlo effort for SuperCDMS.

CHAPTER 2
THE SUPERCDMS DIRECT DETECTION EXPERIMENT

2.1 Introduction

The first Cryogenic Dark Matter Search experiment, CDMS I, was conducted in the Stanford Underground Research Facility at Stanford University. CDMS I collected data until 2002. The next iteration of the experiment, CDMS II, was the first of two major iterations of the experiment to be conducted in the Soudan Underground Laboratory in northern Minnesota. CDMS II collected data from October 2006 to September 2008. The most recent version of the experiment was originally named SuperCDMS. Coincident with the experiment name's upgrade, the collaboration would also be referred to as "the SuperCDMS Collaboration." The first SuperCDMS experiment was operated at the Soudan Underground Laboratory (SUL) from March 2012 until the final warm up of the cryogenic system at the end of 2015.

The next experiment is to be located at the Sudbury Neutrino Observatory (SNOLAB). The new experiment's name is "SuperCDMS SNOLAB;" thus retroactively renaming the first SuperCDMS experiment to "SuperCDMS Soudan."

This work is about physics that is relevant to both the Soudan and SNOLAB experiments. There are, however, some key differences that will be covered in more detail in the following sections. So, while the work on G4CMP was conducted during the operation and analysis of the Soudan experiment, the work is directly applicable to the future experiment as well. Being flexible enough to be used for future experiments was in fact an explicit goal of G4CMP.

2.1.1 Direct Detection Methods

In order to detect a particle, it must deposit energy in a detector. This energy deposition can take several forms: ionization, scintillation, heat, or freed nuclei. It is common in dark matter experiments to measure one or two of these forms of energy. See Figure 2-1.

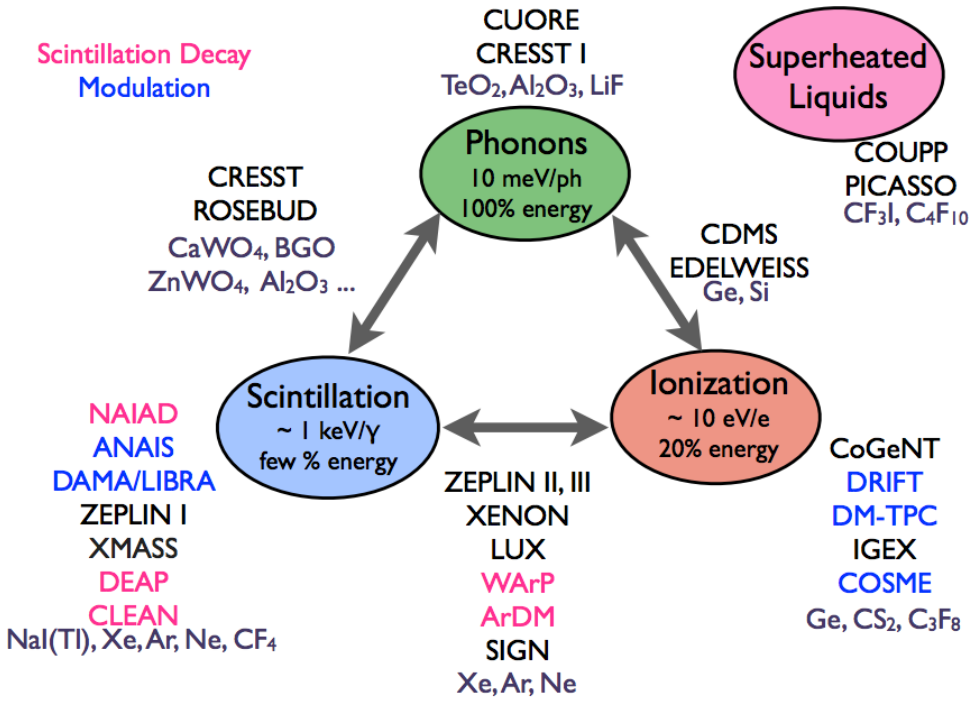


Figure 2-1. Summary of dark matter detection experiments. Measuring multiple energy channels improves the ability to discriminate against different kinds of backgrounds [9].

The CDMS experiments have been designed to measure heat and ionization energy. This design decision provides several advantages, perhaps the most important of which is extremely good background discrimination by calculating the ratio of ionization energy to heat energy. For WIMP interactions, specifically, we would only expect nuclear recoils, since WIMPs do not interact electromagnetically and are kinematically unable to exchange a weak force boson with an electron. Therefore, if a CDMS-style detector sees a large amount of ionization energy compared to heat produced, it is likely to have come from an electron-recoil. This will be discussed in further detail in the upcoming sections.

A WIMP event inside a detector can deposit a range of energies, so it can be difficult to predict exactly what a detection signal may look like. There are basically three unknowns: the interaction cross-section between the WIMP and the detector material, the mass of the WIMP, and the energy distribution of incoming WIMPs. However, a few assumptions are often made when making predictions or analyzing possible data results.

2.1.2 Expected WIMP Scatter Kinematics

When trying to understand the dark matter signal in a direct detection experiment such as SuperCDMS, the key quantity is the differential scattering rate [10].

$$\frac{\mathrm{d}R}{\mathrm{d}E_r} = N_T \frac{\rho_0}{m_\chi} \int_{v_{min}}^{v_{max}} \frac{\mathrm{d}\sigma}{\mathrm{d}E_r} v f(\vec{v}, \vec{v}_e) \mathrm{d}^3 v \,, \tag{2–1}$$

where N_T is the number of target nuclei, ρ_0 is local average WIMP density- usually taken to be around 0.3 GeV/c^2/cm^3, m_χ is the WIMP mass- traditionally thought to be between 1 GeV and 1 TeV, $\mathrm{d}\sigma/\mathrm{d}E_r$ is the differential cross-section, and $f(\vec{v}, \vec{v}_e)$ is the velocity distribution of the WIMPs with respect to Earth. It is the differential cross-section and the velocity distribution that are difficult to describe, but that work has fortunately been carried out already.

In the standard halo model, dark matter surrounds a galaxy in an isothermal, isotropic, spherical "halo" with a density profile of the form $\rho(r) \sim 1/r^2$. The velocity distribution of the WIMPs is often assumed to be a Maxwell-Boltzmann distribution that is truncated at the escape speed of whichever galaxy is being modeled.

$$f(\vec{v}, \vec{v}_e) = \alpha \mathrm{e}^{-((\vec{v}+\vec{v}_e)/v_0)^2} \Theta(v_{esc} - v) \,, \tag{2–2}$$

where $\vec{v}$ is the velocity of a WIMP particle, $\vec{v}_e$ is the velocity of the Earth relative to the WIMPs, v_0 is the root-mean square of the halo's circular speed, v_{esc} is the escape speed for the galaxy, and α normalizes the distribution to unity. It has been shown that the standard halo model, with a Maxwellian velocity distribution is perhaps not the best model for fitting N-body simulations [11, 12].

The differential cross-section is usually written in the form

$$\frac{\mathrm{d}\sigma}{\mathrm{d}E_r} = \frac{\sigma_0}{E_r^{max}} F^2(q) \,, \tag{2–3}$$

where $F(q)$ is the nuclear form factor as a function of momentum transfer between the WIMP and the nucleus and σ_0 is the point-interaction cross-section for a WIMP-nucleus interaction. The form factor is usually taken to be the Helm parameterization, which takes the convolution of a uniform charge density within the nucleus: $\rho_1(r) = 3Ze/4\pi R^3$, $r \leq R$, and a charge distribution that adds "fuzziness" to the outer edge: $\rho_2(r) = \exp\{(-r^2/2s^2)\}/(2\pi s^2)^{3/2}$. One can then deduce the form factor to be [13]

$$F(q) = \frac{3j_1(qR)}{qR} \mathrm{e}^{-(qs)^2/2}\,, \tag{2–4}$$

where j_1 is a spherical Bessel function of the first kind, s is the nuclear skin depth, and R is the effective radius of the nucleus. It is common to assume a spin-independent interaction, for which the point-interaction cross-section looks like

$$\sigma_0 = \frac{4\mu_{\chi,N}}{\pi}(Zf_p + (A-Z)f_n)^2\,, \tag{2–5}$$

where $\mu_{\chi,N}$ is the reduced mass of the WIMP-nucleus system, f_p and f_n are the WIMP-nucleon coupling constants for protons and neutrons, respectively, Z is the atomic number, and A is the atomic mass. It is commonly assumed that $f_p \sim f_n$, in which case Equation 2–5 gives an A^2 dependence, thus making the choice of detector material very important.

The last part of Equation 2–3 is the maximum recoil energy, E_r^{max}. From kinematic considerations, the recoil energy of a WIMP-nucleus scatter is

$$E_r = \frac{\mu_{\chi,N}^2 v^2}{m_N}(1-\cos\theta)\,, \tag{2–6}$$

where m_N is the nucleus mass, v is the speed of the WIMP in the target's frame, and θ is the scattering angle. The maximum energy recoil is then

$$E_r^{max} = \frac{2\mu_{\chi,N}^2 v^2}{m_N}\,, \tag{2–7}$$

which can also be used to find v_{min} for a given recoil energy from Equation 2–3.

$$v_{min} = \sqrt{\frac{m_N E_r}{2\mu_{\chi,N}^2}} \tag{2–8}$$

For high values of the WIMP mass, Equations 2–3 and 2–5 indicate that having a heavier nucleus as the target material will greatly increase the scattering rate. However, for smaller WIMP masses, Equation 2–7 kinematically limits the maximum recoil energy as the target mass gets larger, thus limiting the scattering rate between heavy nuclei and light WIMPs.

2.1.3 Backgrounds at Soudan

The first of the SuperCDMS experiments took place at the University of Minnesota's Soudan Underground Laboratory in Soudan, Minnesota. The underground lab is 2,341 ft below ground in a defunct iron mine.

The key reason for the experiment being held in an underground mine is to reduce backgrounds caused by cosmic rays. Because WIMP events are so rare, and the energy spectrum (Equations 2–3, 2–4) is roughly exponential with no outstanding features, it would be very hard to distinguish a WIMP signal from background noise. WIMPs are only going to interact in the detectors via nuclear recoil. Electron recoils are easy to cut from the data with the detectors' powerful discrimination abilities. Thus, the main background of concern is anything that will recoil with a nucleus in the detectors. It is effectively impossible to distinguish a neutron interaction in the detector from a WIMP interaction.

Thus, SuperCDMS must reduce the neutron background as much as possible. The previous generation experiment's home at Soudan was approximately 713 meters below ground, which provided the equivalent shielding of 2090 meters of water (often shielding measurements are reported in units of "meters of water equivalent" or m.w.e). This overburden provided strong shielding from cosmogenic muons, which can produce secondary neutrons as they interact with rock. See Figure 2-2.

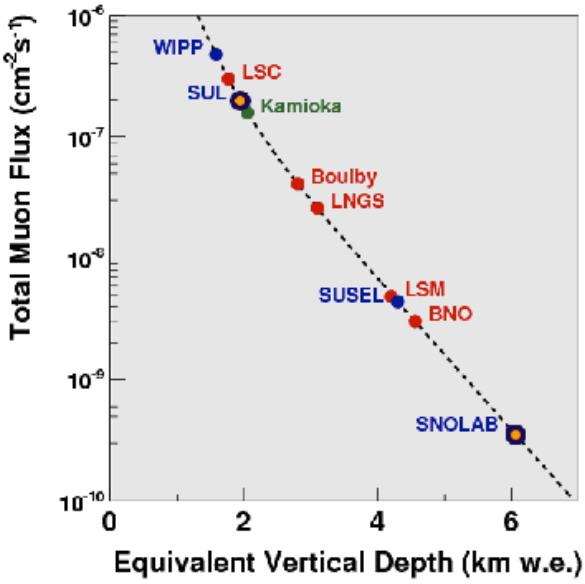


Figure 2-2. Comparison of the muon shielding ability of various underground laboratories. Note that the next generation SuperCDMS experiment will take place at SNOLAB in Sudbury, with some two to three orders of magnitude better muon shielding than the Soudan Underground Lab (SUL). Figure from [14], which was adapted from [15].

The SuperCDMS experiment at Soudan also had several layers of shielding around the detectors. The first line of defense was the active muon veto, which was a set of plastic scintillator panels that completely enclosed the experiment. A muon veto is not shielding per se, but does allow an experiment to reject events that coincide in time with a veto trigger.

Working inward from the veto layer was a 40 cm thick layer of polyethylene to block neutrons, then a 17.8 cm layer of lead to block gammas. After the first lead layer is another 4.4 cm of very old lead in which most of the ^{210}Pb has decayed. This inner lead layer shielded the decay activity of the less-stable outer lead layer. Then there was another 10 cm of polyethylene which will block neutrons that were not blocked by the other layers. A schematic representation of the shielding can be seen in Figure 2-3.

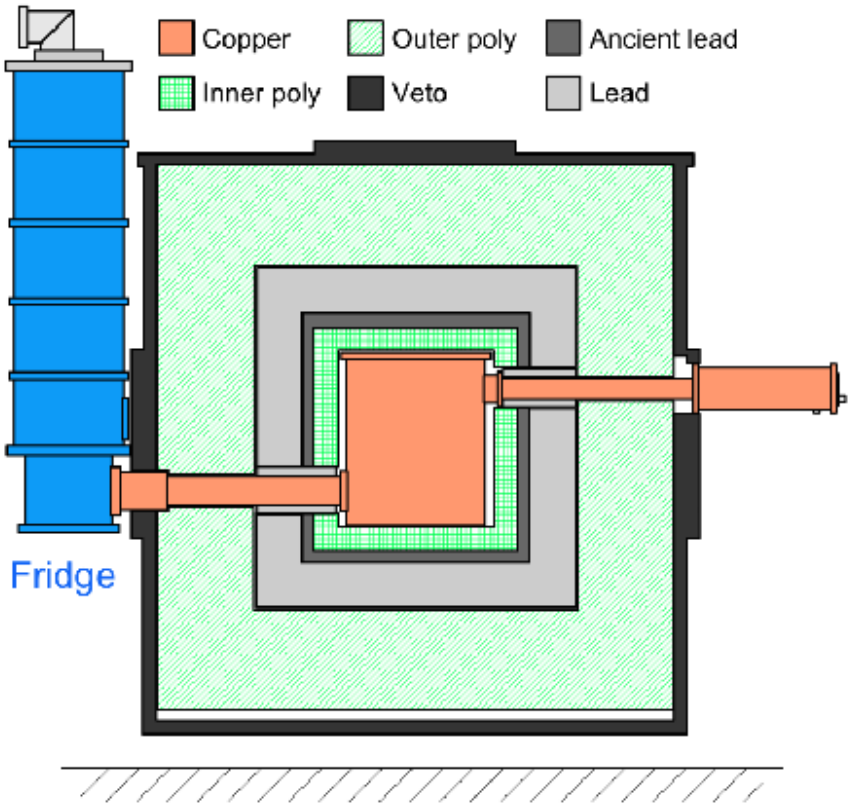


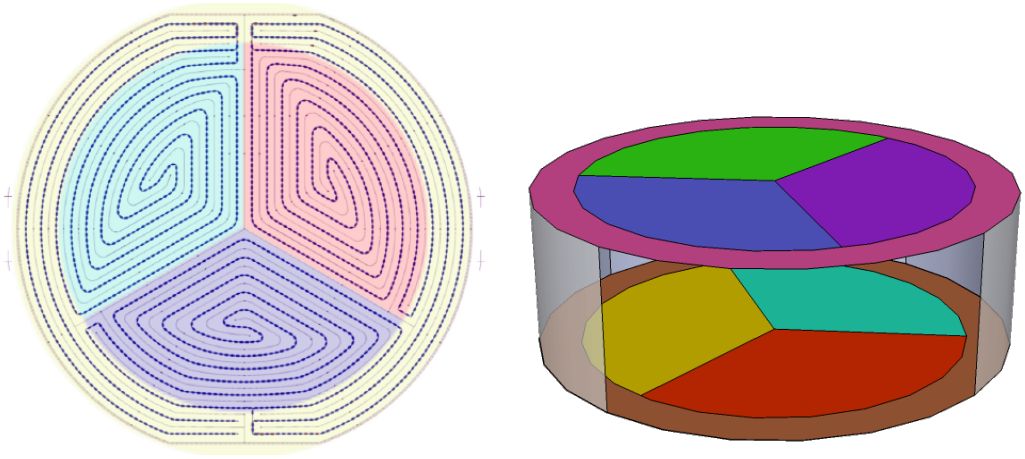
Figure 2-3. The SuperCDMS experiment shielding. The various layers of polyethylene and lead as well as the copper "ice-box" which holds the copper detector towers and is cooled to around 50 mK. To the right is the helium dilution refrigerator. The readout electronics pass through the copper tube in the right of the figure. Figure from [14].

2.2 iZIP Detectors

2.2.1 Introduction

The dark matter detectors that were used by SuperCDMS Soudan are called "interdigitated, z-sensitive, ionization and phonon detectors" or "iZIPs." Half of the detectors at SNOLAB will be iZIP-style detectors. The iZIP detectors of the SuperCDMS Soudan experiment were cylindrical pucks of Germanium with a radius of 3.81 cm, and a height of 1 inch. Each face of an iZIP contains phonon and charge sensors in an interdigitated pattern. Each Soudan detector had about 600 grams of mass. The charge sensors were read out in four channels (two on each face) via FETs, while the phonon sensors were transition edge sensors (TES) that were read out in eight channels (four on each face).

SNOLAB iZIPs will also have a total of four charge channels, but will increase the phonon channel count to 12. The phonons will still be read out via TESs, but the charge signal will be read out via high-electron-mobility transistors (HEMTs). The rest of this section refers specifically to the Soudan iZIP design, but much of the discussion translates directly to the SNOLAB iZIPs as well. Figure 2-4 shows the channel configuration of a Soudan iZIP.



A B

Figure 2-4. Schematics of a Soudan iZIP detector illustrating the sensor configurations. A) Schematic of the surface of an iZIP detector. The thin dashed lines are the charge lines and the thicker, beaded, lines are the phonon lines. The color shading shows the shapes of the four phonon channels. B) 3D model of an iZIP detector, showing the phonon channels on each face. Note the relative rotation of the phonon channels of each side. This gives better x-y position resolution. There are also two charge channels on each face: one shares the same area as the outer phonon channel, and the other shares the area of the three inner phonon channels.

The charge readout sensors are voltage biased on each face to create a weak electric field across the z-direction of the detector. The phonon sensors on both faces are grounded, producing strong, scallop-shaped electric fields near the surfaces of the detector which is crucial for the iZIP's z-sensitivity as will be discussed in Section 2.2.3.

When a particle interacts with the detector, it is either via an electron-recoil or a nuclear-recoil. Both of these kinds of events create ionization, which is read out by the

FETs; and athermal phonons, which are measured by the TESs. As the charge carriers are accelerated by the electric field, they are slowed by interactions with the crystal lattice which manifest as emitted phonons [16, 17]. This effect is referred to as Neganov-Luke phonon emission, or just "Luke phonons" for brevity. As the charge carriers come to rest near an electrode, they relax from the band gap, releasing their energy back to the lattice as phonons. The larger the field, the larger the acceleration of the carriers, and thus the larger the emitted phonon energy. These effects will be discussed in much more detail in Chapter 3. What this means is that the phonon sensors will always detect energy equal to the total energy of the recoil plus an additional amount that is proportional to the strength of the electric field across the detector.

The primary discriminator between electron-recoils and nuclear-recoils is a quantity referred to as the ionization yield, or just "yield" and will be described in detail in the next section.

2.2.2 Recoil Physics

When a particle interacts in one of the SuperCDMS detectors, it either scatters from an electron or a nucleus. When an electron recoil occurs, a valence electron is freed from its electron orbital state, leaving behind a hole. If the recoil energy is much higher than the gap energy of the semiconductor (0.74 eV for Ge, 1.18 eV for Si), then the electron will have a large amount of kinetic energy. It is kinematically unfavorable for the electron to be able to impart significant energy to a nucleus. However, it is quite likely that the freed electron will recoil with another valence electron, giving up some of its kinetic energy. The exact rate and kinematics depends on the stopping power of the material. This process continues for each freed electron until it has given up enough energy that it is no longer kinematically likely to be able to free more valence electrons. At this point, the freed charges eventually dissipate the remainder of their kinetic energies to the lattice via phonons.

It turns out that if the recoil energy is larger than some value for a material ($\sim$ 10 eV for Ge and Si [18]), then the overall number of electron-hole pairs created per unit energy is constant, giving us a value for the average energy needed to create an electron-hole pair in a material. Due to this effect we can calculate the total energy that will end up in charge carrier pairs after the initial cascade.

$$E_{charge} \approx \frac{E_{recoil}}{E_{pair}} E_{gap}\,, \tag{2–9}$$

where E_{recoil} is the energy initially deposited by the recoil, E_{pair} is the average energy needed to create an electron-hole pair, and E_{gap} is the semiconductor gap energy of the material. This also tells us that the *proportion* of the recoil energy that ends up in charge carriers is constant for a material: E_{gap}/E_{pair}. Therefore, it must also be the case that the ratio of charge energy to phonon energy initially created is

$$\frac{E_{charge}}{E_{phonon}} = \frac{E_{gap}}{E_{pair} - E_{gap}}\,. \tag{2–10}$$

A nuclear recoil works similarly to an electron recoil except that the initial impact can free both a nucleus and electrons. The electrons, of course, undergo the cascade just described, but the nucleus is subject to the electron stopping power of the material as well as the nuclear stopping power. As the nuclei all come to rest, the energy is released as phonons. These two effects compete to determine the final population ratio of freed electrons to phonons. Lindhard developed a model for the electronic stopping power of nuclei in solids [19],

$$\frac{\mathrm{d}\epsilon}{\mathrm{d}\rho} \approx 0.133\ Z^{2/3}\ A^{-1/2}\sqrt{\epsilon}, \tag{2–11}$$

where $\epsilon = 11.5\ Z^{-7/3}\ E_{recoil}$ (in keV). The nuclear stopping power can only be described numerically, but overall the Lindhard model describes a ratio of the ionization energy over the total phonon energy,

$$Y(\epsilon) = \frac{k \cdot g(\epsilon)}{1 + k \cdot g(\epsilon)}, \tag{2–12}$$

where we will refer to Y as the "ionization yield", or just "yield." In the previous equation,

$$k = 0.133\ Z^{2/3}\ A^{-1/2}, \tag{2–13}$$

and $g(\epsilon)$ is a complicated function. Lewin and Smith [20] approximated

$$g(\epsilon) \approx 3\epsilon^{0.15} + 0.7\epsilon^{0.6} + \epsilon, \tag{2–14}$$

leaving us with an easy to calculate factor for the amount of recoil energy that should go into the phonon sector from a nuclear recoil. It is also important to note that the yield is calibrated such that an electron recoil should produce a yield near 1. That means that in materials such as germanium and silicon, the yield of a nuclear recoil will be ~ 0.3.

2.2.3 Discrimination

Because of the difference in physics between an electron recoil and a nuclear recoil, the type of recoil can be deduced from the ratio of the collected charge signal and the collected phonon signal. This yield-based discrimination is a major feature for a dark matter search, since we expect all electron recoils to be background, and thus should be rejected.

One of the other key designs of the iZIP to aid in background discrimination is the electrode pattern described in Figure 2-4. Since yield is such an important background discriminator for the experiment, it is especially troubling that certain events can have reduced charge collection. This can have the effect of an electron recoil having a yield that is closer to the typical nuclear recoil. As Wolfgang Pauli said: "God made the bulk; surfaces were invented by the devil" (as quoted in [21]). Events that are either near the sidewalls or near the faces will have reduced charged collection. In the case of the high-radius events, it's because the sidewalls are not instrumented, but can collect and trap charge carriers. For surface events it is because the electronic cascade does not have time to complete before some of the carriers are collected and many of the high energy charge

carriers will also diffuse quickly into the wrong sensor; the field is not strong enough to slow and repel the opposite charged carriers.

The sidewall events are easily removed by detecting which collection channel saw the majority of charge collection (See Figure 2-4)- if it is one of the outer ring channels, then the event was too high-radius for us to be confident about its energy, so we reject it. The iZIP is also able to confidently reject surface events. Because of the shape and strength of the electric field near the surface, charge carriers that are produced near the faces will only be collected on that side of the detector as illustrated in Figure 2-5. For example, if we have an event very close to the electron side of the detector, the electrons will still be collected by the charge sensors on that side. However, the holes will not see the favorable potential on the other side of the detector, but will instead be strongly pushed toward the grounded phonon sensors on the electron side. The phonon sensors do not read out the charge energy collected, so the resulting signal of such an event appears with one side's inner charge channel registering an output pulse that is much larger than the other side's. We refer to these kinds of events as asymmetric charge events and they are also rejected.

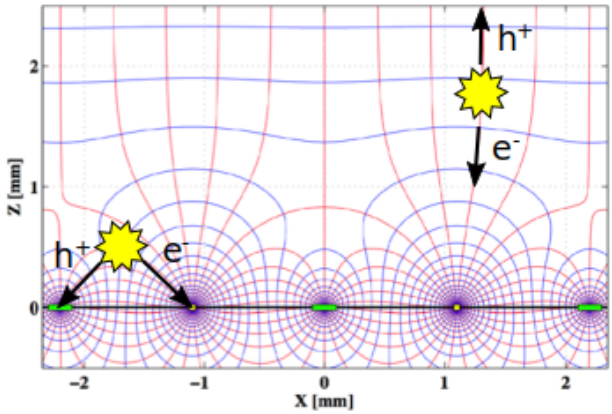


Figure 2-5. Electric potential and field near the surface of an iZIP detector. The blue lines are equipotentials and the read are the corresponding electric field lines. When electron-hole pairs are created by events in the bulk, each specie will be collected by opposite sides. When electron-hole pairs are created near a surface, only half of the charges will be collected by the ionization sensors.

In Figure 2-6 we can see the discrimination power of the iZIP in practice. These plots show around 900 hours of data collected by one of the iZIP detectors at Soudan. This detector has a ^{210}Pb source adjacent to the top face (Side 1). As the lead decays, several of its descendants interact with the detector, including ^{206}Pb, betas, and gammas (See Figure 2-7). In panel (a) of Figure 2-6 is a plot of the charge energies collected by Side 1 vs. Side 2. Because the source is near Side 1, there are many surface events in which charge energy is only collected on that side, as can be seen by the many red points near the Side 1 axis. A noteworthy feature of the data is the pillar of data points from just below 50 keV of Side 1 energy. This feature is the 46.5 keV gamma signal from the lead decay chain in Figure 2-7. In panel (b) is a plot of the ionization yield vs recoil energy (total phonon collected energy). Events that penetrated far enough into the iZIP to have symmetric charge collection also tended to be electron recoils (yield ≈ 1). This is because the heavy daughter nuclei of the lead source cannot penetrate the germanium enough to have symmetric charge collection. The two sigma nuclear recoil band is drawn in green. The vertical black line is the recoil threshold of the iZIP (8 keV). The hyperbolic black line is the ionization threshold (2 keV electron recoil equivalent, or "keVee"). Also seen are two very low yield outliers that did pass the quite generous charge symmetry cut. In panel (c) is a comparison of the yield to the charge collection partition. This view shows a clear separation of populations, with the asymmetric events clearly separated from good events. The green points are data from Cf calibration data, which fall well into the symmetric nuclear recoil region.

2.2.4 Phonon Sensors

The phonon signal in an iZIP is read out via eight channels. Each with 458 phonon sensors, so that each channel has the same impedance. The phonon sensors themselves are of physical interest. The readout is done by exploiting the sharp phase transition for a metal becoming superconducting. These kinds of devices are called "transition edge sensors," or TESs. Using an electro-thermal feedback mechanism, the SuperCDMS

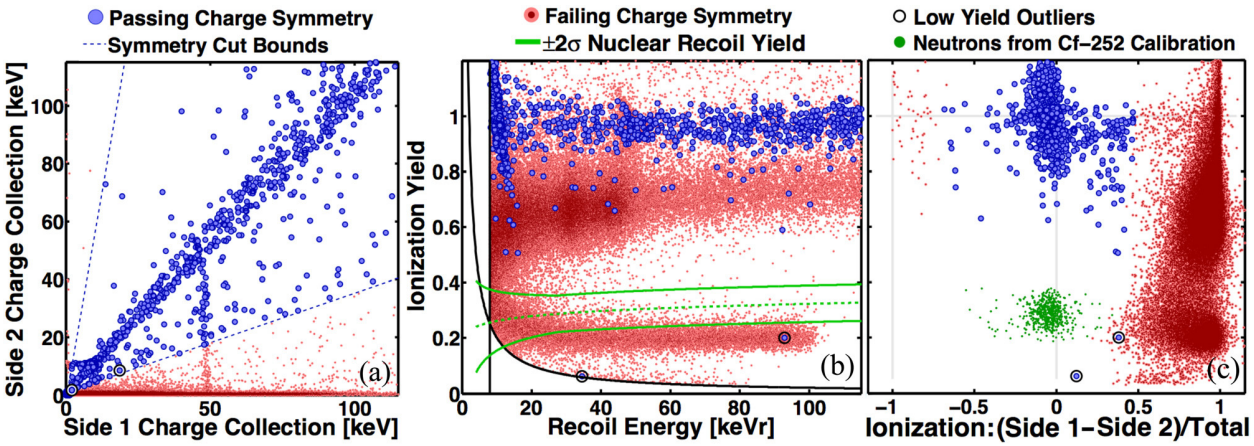


Figure 2-6. Data from a ^{210}Pb source on one face of an iZIP [22]. (a) For events in the bulk nearly 100% of created charge carriers should be drifted to and absorbed by the biased sensors on each face. (b) Ionization yield vs. recoil energy. Asymmetric events will have lower yield than bulk events. (c) Ionization yield vs. Ionization partition.

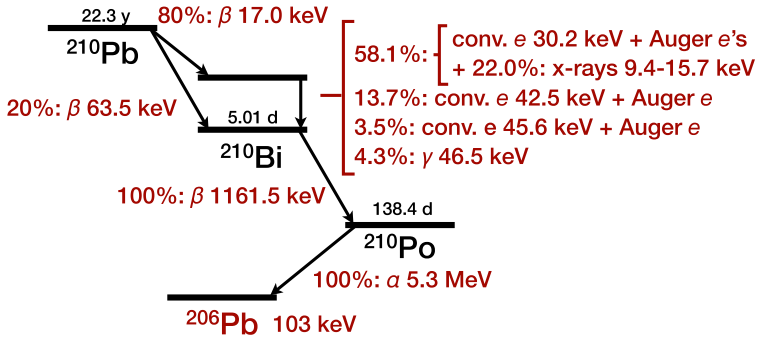


Figure 2-7. ^{210}Pb decay chain.

detectors keep the TES at the low end of the superconducting transition region (R $\sim$ 0.3 R_{normal}). As phonons are collected by the TES, the small amount of heat causes a sharp increase in resistance, which in turn causes a sharp drop in the output current. That change in current changes the magnetic field through the an inductor. That change in magnetic flux is amplified with a low-noise SQUID amplifier circuit. This means that the phonon readout can detect very low amounts of added heat in the crystal.

Figure 2-8 shows a schematic of a phonon readout channel. The TES acts as a variable resistor. At equilibrium, most of the current will flow through the TES. As the

TES heats, its resistance increases. This increase drives more current through the shunt resistor that is wired parallel to the TES. The drop in current through the TES side causes a change in the magnetic field in the inductor, which is detected by the SQUID and amplified through the sister inductor's circuit, which is otherwise isolated from the TES circuit. This results in a very low noise, and very precise temperature sensor.

Once the flux of heat into the TES has subsided, the TES's contact with the thermal bath will cause its temperature to be reduced. As the temperature is reduced, the increasing current through the TES gradually increases until the ohmic heating exactly balances the cooling from the bath. Figure 2-9 illustrates the changes that take place in the circuit over the course of an event in the detector. For a much more in depth analysis of TES dynamics, see Matt Pyle's dissertation [23].

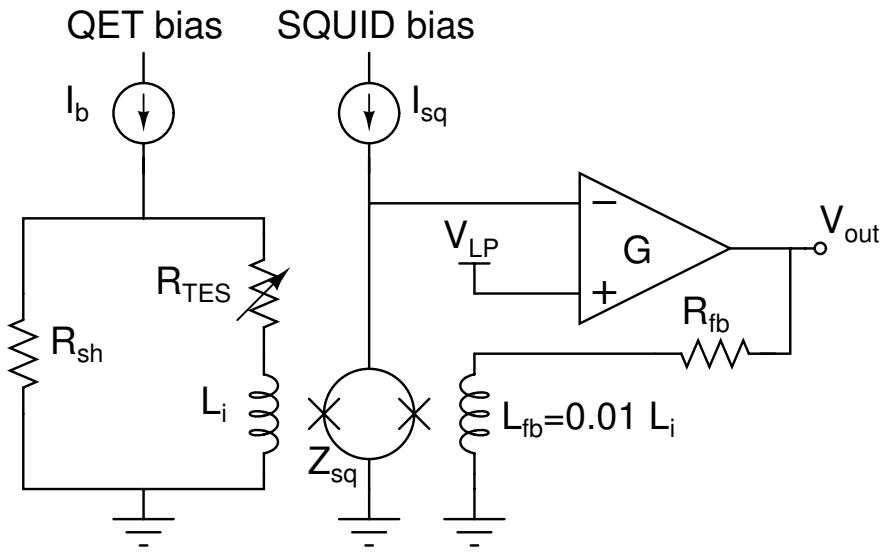


Figure 2-8. The SuperCDMS phonon readout circuit. Note that the TES (R_{TES}) is kept at cryogenic temperatures. Diagram from Jeff Filippini [24].

To use a TES as a phonon detector, one has to grapple with competing effects. The energy resolution of the detector is going to scale inversely with the heat capacity ($\Delta T = CV/Q$), so it is better to have a small TES volume, all other things equal. However, small TESs are less likely to absorb a phonon that is chaotically bouncing around inside the crystal, rendering their resolution moot.

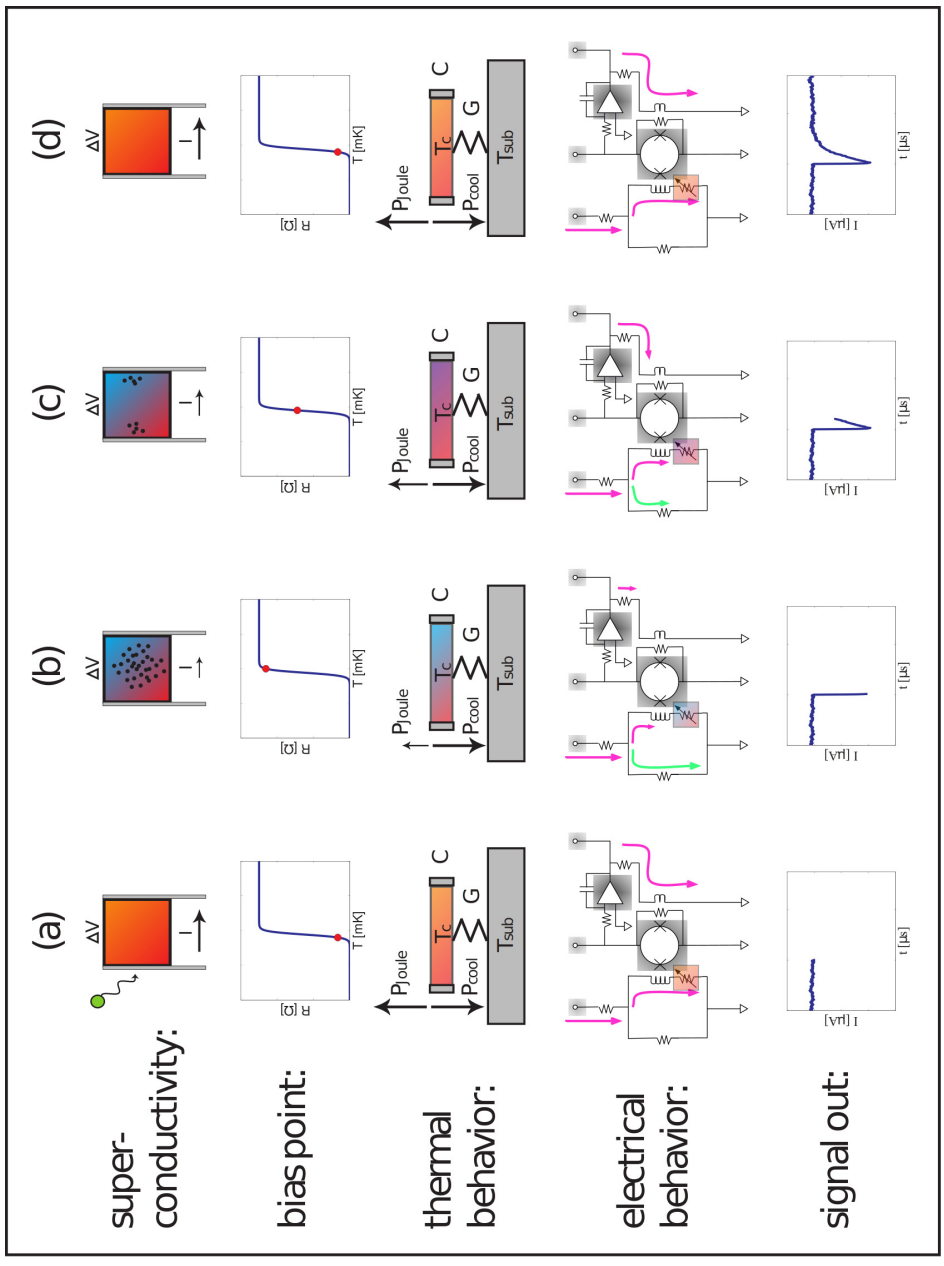


Figure 2-9. The response of the TES circuit to an event. The takeaway is that at equilibrium, Joule heating and substrate cooling cancel each other out, but when the TES is heated, the substrate cooling increases while the Joule heating decreases, leading to a rapid cooldown back to equilibrium temperature. Diagram from Jennifer Burney [25].

In order to increase the phonon collection efficiency without destroying the energy resolution of the detectors, SuperCDMS detectors utilize quasiparticle-assisted electrothermal feedback transition edge sensors (QETs). We have already covered what a transition edge sensor is, and electrothermal feedback was described in the previous section. The "quasiparticle-assisted" part of the device refers to large aluminum fins that are attached to the small tungsten TESs. This provides an effective "funnel" for the phonon energy to the TES while not significantly changing the overall heat capacity of the TES itself.

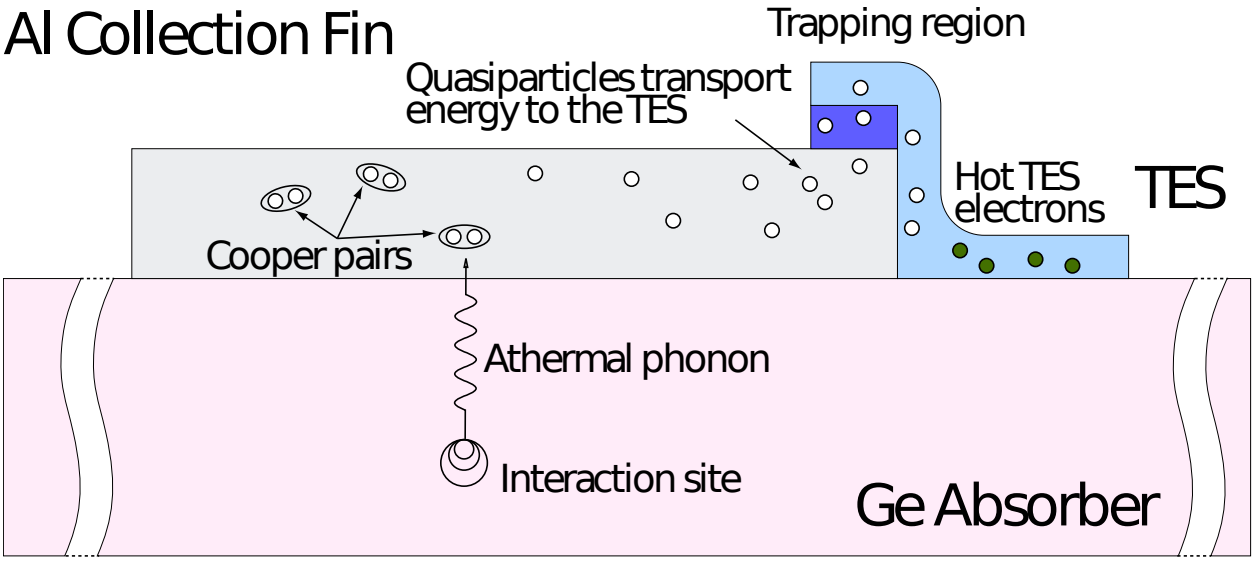


Figure 2-10. Diagram of how the QET fin aids in heat collection in one of SuperCDMS's detectors. The aluminum is in its superconducting phase, so the incoming phonons deposit energy by breaking Cooper pairs into quasiparticles. Similar devices have electric fields applied to help drift the quasiparticles to the detector, but the SuperCDMS QETs rely on diffusion.

2.2.5 Charge Sensors

The ionization measurement is read out via FETs. The charge readout circuit is much more simple than the phonon readout, as can be seen in Figure 2-11. The iZIP detector is effectively a capacitor in the FET circuit, which feeds into a current integrator. A key aspect of this circuit is the fall time of the feedback loop on the op-amp. For the values of R_f and C_f, the fall time is ~ 40 μs. The time it takes for the charge carriers from an interaction event to propagate across the detector volume is between $\sim$100 ns and ~ 1 μs,

depending on the voltage applied. Since the charge channel readout is digitized in 0.8 μs bins, this means that all of the charges arrive at electrodes within one time bin. Thus, the rise time of all charge pulses is the same (one bin). This means that the shape of the pulse coming out of the charge channel is effectively uniform for all events, sans noise effects. Only the pulse height will be different, depending on how much charge was collected. This fact will be important when simulating the charge response of the detector in Section 4.3.

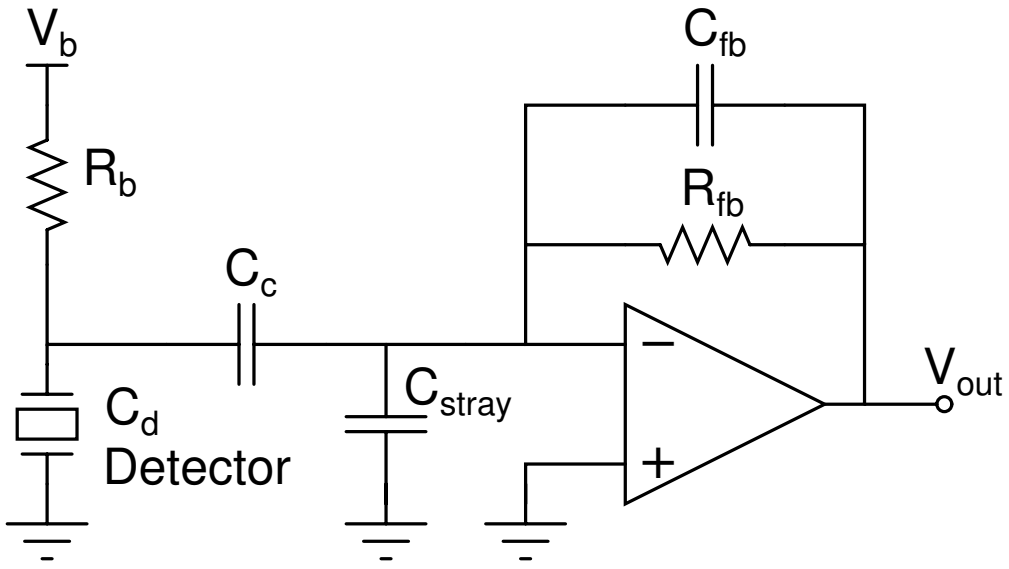


Figure 2-11. A simplified iZIP charge sensor readout circuit. See Table 2.2.5 for circuit values.

Table 2-1. The charge sensor circuit parameters for SuperCDMS Soudan. Note that the inner and outer charge channels have different capacitances.

Parameter	Value
R_b	40 MΩ
C_d (inner)	93 pF
C_d (outer)	36 pF
C_c	300 pF
C_{stray}	≈ 75 pF
R_{fb}	40 MΩ
C_{fb}	1 pF

2.3 High Voltage Mode

The previous sections describe the original intended operation of the iZIP detectors. Recall from Section 2.2.1 that due to the Luke effect, the phonon sensors will actually detect an energy that is *larger* than the energy of the initial recoil.

$$P_t = \left(1 + Y\frac{V}{E_{pair}}\right) P_r, \tag{2–15}$$

where P_t and P_r are the total phonon energy and the initial, or recoil, phonon energy, respectively; Y is the yield, which depends on the type of recoil; E_{pair} is the average charge carrier pair production energy; and V is the voltage applied across the detector.

It was realized that by simply increasing the voltage applied to the detector, we could significantly decrease the energy threshold for detection due to the Luke amplification. However, there are two main difficulties with doing so. If we increase the voltage too far, the leakage current in the detector will dominate the inherent sensor noise, which is counter-productive to lowering the energy threshold. Also, the sensor layout is such that if we increase the voltage much higher than a few volts, it means we can no longer ground the phonon sensors, which are only a few microns away from the biased charge sensors. It would also be pointless to collect charge signal as it would be dominated by noise. This means that we lose out on the discrimination power described in Section 2.2.3. However, even though event-by-event discrimination is impossible in the high voltage mode, it turns out that statistical discrimination is still possible [26]. This is because electron recoils will produce proportionally more charge carrier pairs than nuclear recoils, and therefore will produce a higher Luke gain. So the electron recoils backgrounds will tend to be push into higher total energy bins. Thus, the relative electron recoil background is suppressed relative to the WIMP signal for low energies.

The first science runs utilizing this technique are referred to as the "CDMS low ionization threshold experiment" or CDMSlite [27, 28]. The high voltage mode provides a

great advantage for searching for very low mass WIMPs ($\sim 1 - 10 GeV/c^2$) that half of the 24 detectors planned for SNOLAB are going to be operating in high voltage mode.

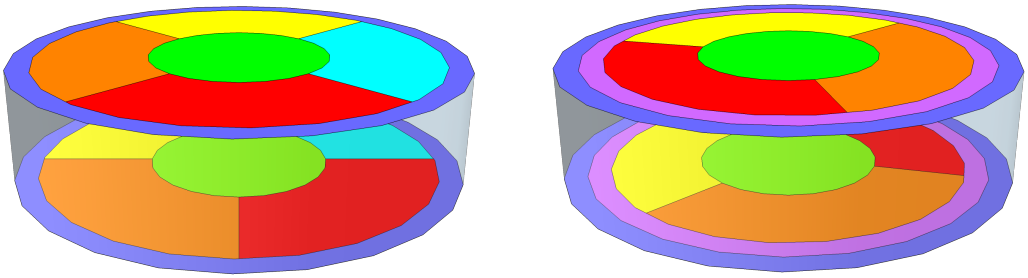
2.4 SuperCDMS SNOLAB

In the next generation experiment there will be several relevant changes to the experimental design. First, the target payload is larger. Specifically, there are more detectors, and each detector is larger. Whereas the Soudan iZIPs were each 1 inch (25.4 mm) thick with a 38.1 mm radius, the SNOLAB detectors will be 33.3 mm thick and have a radius of 50 mm radius. There will also be silicon detectors at SNOLAB, whereas all of the Soudan detectors were germanium (CDMS-II also had a mix of germanium and silicon). The silicon detectors should be more responsive to lower energy recoils due to the lower mass of the atoms.

Furthermore, there will be two different sensor layouts used at SNOLAB. One, which will still be called iZIP, is similar to the iZIP design used at Soudan, but with more phonon channels and a different sensor pattern, intended to achieve more uniform phonon collection. The other, referred to as HV (high voltage), is specifically designed for a CDMSlite-style experiment. Since there is no ionization collection for those detectors, there are no charge sensors or readout. They also have two outer ring channels as that design offers better rejection of low yield events at high radius, as revealed by Monte Carlo experiments. See Table 2-2 for the summary of the detectors planned for SNOLAB.

2.5 Direct Detection Experimental Results

Because the recoil energy and rate of WIMP interactions will depend both on the mass and interaction cross-section of the WIMP, dark matter experimental results are often reported as a limit curve or acceptance region in the parameter space of cross-section vs. WIMP mass. An example is given in Figure 2-13, which shows the last several generations of dark matter experiment results plus the published projections of several of the next-generation experiments. Some interesting trends and results can be interpreted from this plot. As can be seen, there have been several experiments that have seen signals



A B

Figure 2-12. SNOLAB detector channel layouts. A) iZIP detector channels. Six phonon channels on each surface: one "core", four "wedges", and one outer ring. There are two ionization channels on each face. The outer ionization channel shares the same area as the outer phonon ring. The inner shares the area of the other five phonon channels. B) High voltage detector channels. Same number of phonon channels as the iZIP detector, but two outer phonon rings to reject high radius events that will produce lower Luke gains. No ionization channels. Note that for each detector design, the phonon channels on one face are rotated relative to the channels on the opposite face to maximize x-y position resolution.

Table 2-2. The planned detector parameters and exposures for SuperCDMS SNOLAB.

	iZIP		HV	
	Ge	Si	Ge	Si
Number of detectors	10	2	8	4
Total exposure (kg·yr)	56	4.8	44	9.6
Phonon resolution (eV)	50	25	10	5
Ionization resolution (eV)	100	110	N/A	N/A
Voltage bias (V)	6	8	100	100

The exposures are calculated under the assumptions of an 80% live time over five years. The phonon resolution represents the root mean square (RMS) of the sum of all phonon sensors for a single detector. The ionization resolution represents the RMS of just the inner channel. HV denotes the high voltage detectors and iZIP denotes the lower voltage, interdigitated sensor design. See [29].

that they could not rule out as background, including CDMS-II silicon data. The CDMS experiments have consistently been world-leading dark matter experiments over the full range of possible WIMP masses until the liquid noble gas experiments (Xenon100, LUX) began to dominate the high-mass range. This is in part because it is relatively easier, from an engineering perspective, to rapidly scale the payload of the liquid noble gas experiments, whereas it would be very difficult to manufacture a functional iZIP that is much larger than a kilogram. Thus, CDMS experiments must produce more detectors to increase the target mass, which is expensive and adds some overhead to analysis.

However, as can be seen by the projections, the SuperCDMS experiments will completely dominate the low-mass WIMP region. This is largely because of the lighter mass of silicon and germanium atoms relative to Xenon as well as the incredible sensitivity of TESs to very small vibrations. In particular, the high voltage operating mode, as discussed in Section 2.3, make a substantial improvement in very low mass WIMP detection capabilities.

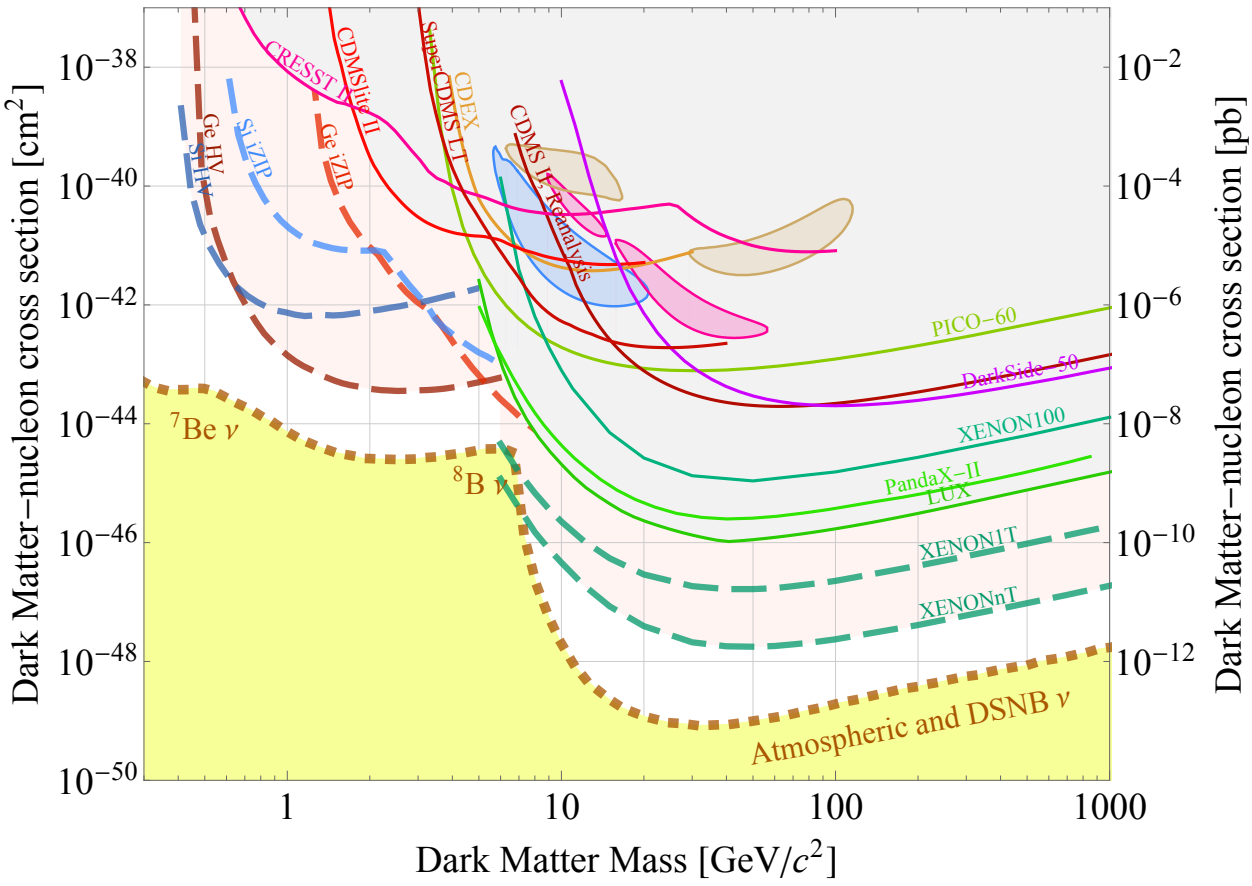


Figure 2-13. Dark matter experimental results including projections for next generation experiments. Solid lines are rejection limits, meaning that any parameter space above the line is ruled out. Filled contours are acceptance regions where a positive signal was detected. The pink region is results from CREST-II, the blue region is from CDMS-II reanalysis of the germanium data, and the tan region is from the DAMA/LIBRA experiment. Dashed lines are projected sensitivities for future experiments. Also present is the "neutrino floor," at which experiments will being detecting recoils from background solar neutrinos, presenting new difficulties for dark matter experiments [28–43].

CHAPTER 3
A CONDENSED MATTER LIBRARY FOR GEANT4 SIMULATIONS

3.1 Introduction

The behavior of the SuperCDMS iZIP detectors is generally well-understood in the nominal cases. But pen and paper, and calibration data can only go so far in characterizing the detector behavior. When using calibration data, we can know the energy spectrum of events created, but we do not get to know the energy of a specific event, so we can only study the detector response statistically. A Monte Carlo simulation can also study events in trouble regions, such as places where the electric field magnitude is extremely high, or at a saddle point. We can also study the response of non-ideal detectors, such as ones with shorted readout channels or other maladies. There are also logistic concerns, such as needing to wait for the detectors to neutralize after being activated by radioactive sources, thus decreasing live time.

Of course, all of these concerns are well-known to experimentalists, and thus SuperCDMS does have a detector Monte Carlo package that is not what is described by this work. The old package is written in MATLAB and will therefore be referred to as the MATLAB Detector Monte Carlo package, or MATLAB DMC. There are significant shortcomings with the MATLAB DMC that are hopefully ameliorated by the current effort. The main shortcoming of the MATLAB DMC is that much of the physics and detector geometry is hard-coded. In fact, the MATLAB DMC was written with CDMS-II detectors in mind, which had phonon sensors on one side and charge sensors on the other side, as opposed to the SuperCDMS detectors that have both types of sensors on each side. In an amusing parallel to the physical world, while the SuperCDMS detector towers reused the old CDMS-II electronics in the laboratory by pretending each iZIP was actually two of the CDMS-II ZIPs, the MATLAB DMC was kludged in a similar fashion to accommodate the new detector type. Someone who does not know this bit of history would be quite confused if they were to study the MATLAB DMC code. Of course, the confusion about

channels is not reason enough to abandon a piece of software. Some of the more difficult problems are in the architecture of the MATLAB DMC. For example, it hard-codes the crystal orientation for germanium and that assumption permeates the entire code base, implicitly. Similarly, the physics processes that are implemented are not modular, so it would be difficult to replace implementations with different models. The MATLAB DMC also only simulates one detector at a time, meaning that it is impossible to simulate the effect of some background on the entire experiment (15 detectors in SuperCDMS Soudan). One more practical concern is that MATLAB licenses are not gratis, so it is not easy for all collaborators to develop and test locally.

So, it was clear that the DMC needed a rewrite to something a little more future-proof and modular. One of the main things that the MATLAB DMC does is basic particle tracking- it calculates the trajectory of charge carriers and phonons, finds the acceleration due to the applied electric field, figures out when they arrive at boundaries, etc. Geant4 is an already-existing, very well-optimized, particle tracking Monte Carlo framework. There would be no need to reinvent the wheel if we leveraged Geant4. Furthermore, SuperCDMS has a backgrounds Monte Carlo simulation package based on Geant4 already, called SuperSim. Thus, SuperCDMS could perform an end-to-end simulation all the way from modeling cosmic rays to the detector readout pulses using one software package, instead of two completely different, and somewhat incompatible, pieces of software.

The missing piece from Geant4, however, is the necessary condensed matter physics processes that are relevant at low energies inside of crystals. The goal of G4CMP is to write a generic, extensible, library of the tools needed to include very low energy crystal physics in Geant4. That library can then be used by SuperCDMS and others to perform detailed detector Monte Carlo simulations. In doing so, we keep the physics separate from the device geometry, as well as having the individual physics processes as independent and modular.

3.2 Understanding Geant4 Semantics

Before diving into the microscopic detector physics and accompanying simulation algorithms and implementations, it makes sense to describe some of the details and philosophy of Geant4, since G4CMP's design is obviously constrained by Geant4. In this section we will cover the aspects of Geant4 that are helpful to understanding how G4CMP works.

3.2.1 Particle Tracks

The "fundamental unit" of a Geant4 simulation is the G4Track object. A G4Track holds many pieces of information relevant to simulating a particle's trajectory through a volume. There is as much documentation as one could desire for the internal workings of Geant4 [44–46]. For reference I list a simplified (relevant) subset of the parameters in a G4Track in Table 3-1.

Table 3-1. A non-exhaustive list of information inside a G4Track.

Parameter	Description
Position	Global x,y,z coordinates
Global Time	Time in lab frame since the beginning of the event.
Local Time	Time in lab frame since the track was created.
Track ID	ID number of the track.
Parent Track's ID	ID number of the track which spawned this one.
Velocity	The track's current speed (scalar).
"useGivenVelocity" flag	Whether the track has a set speed (e.g., photons), or should be calculated from its energy.
Dynamic Particle	Description of the particle itself (mass, charge, etc).
Auxiliary Information	Arbitrary data associated with the track.
Track Weight	Used for Geant4's track biasing feature.
Current Volume	The current shape and material of the volume where the track is located.

Several of these parameters will be discussed as needed, however there is some up-front clarification that is worthwhile. Most of the kinematic quantities are actually stored in the dynamic particle (see Table 3-2).

Table 3-2. A non-exhaustive list of information inside a G4DynamicParticle.

Parameter	Description
PDG ID	The Particle Data Group Monte Carlo ID for the particle.
Kinetic Energy	Kinetic energy of the particle.
Momentum Direction	Unit vector pointing along momentum vector.
Polarization	Polarization 3-vector.
Proper Time	The particle's current proper time.
Mass	The particle's current mass (times c^2).
Charge	The particle's electric charge in units of e.
Spin	The particle's spin.
Magnetic Moment	The particle's magnetic moment.
Decay Time	The particle's lifetime.

This is an important clarification to make because it is easy to forget whether a parameter is part of the track object or part of the particle object, when, as physicists we may be tempted to treat them interchangeably. It is therefore quite likely that this work may also inadvertently conflate the two in later sections. Almost all of the quantities in the Dynamic Particle can change, except for the PDG ID. The PDG ID is the particle data group identification number of the particle. This is what describes whether something is actually an electron, Z-boson, or a Uranium atom; for example. Notice that the kinetic energy and the momentum direction are the only kinematic quantities stored in the dynamic particle object. This is because the velocity/speed of a particle can easily be calculated from those two quantities.

$$E^2 = p^2 + m^2 \tag{3–1}$$

$$\vec{p} = E\vec{v} \tag{3–2}$$

So,

$$|\vec{p}\,| = \sqrt{E^2 - m^2} \tag{3–3}$$

$$|\vec{v}\,| = \frac{\sqrt{E^2 - m^2}}{E} \tag{3–4}$$

And therefore Geant4 can get by with only storing the kinetic energy of a particle track and the direction of the momentum vector. However, recall that G4Track has members called "velocity" and "useGivenVelocity" 3-1. These were introduced into Geant4 in order to bypass using the dynamic particle's methods for calculating the velocity. The useGivenVelocity flag is set to true for particle tracks that should not use the default method of calculating a velocity. Instead they will use the "velocity" that is stored in the G4Track. At the time it was added specifically for optical photons, which otherwise would be treated like regular photons with speed equal to 1 in natural units. G4CMP can exploit this mechanism for phonons, which are also massless and have a different relationship between $\vec{p}$ and $\vec{v}$ than is assumed by Geant4.

Another aspect of Geant4 that we must keep in mind is that tracks are completely independent from one another. Therefore, tracks cannot directly interact. In fact, since Geant4 uses a last-in-first-out stack for tracks, there is no guarantee that tracks are simulated in chronological order, either. This makes some potentially desirable effects more-or-less impossible to implement, such as temperature fluctuations of a material from phonons, distorted electric fields from freed charge carriers, true electron-hole recombination, etc. Luckily, most of these processes either will not contribute significantly to our application or are easy enough to approximate in some other way.

3.2.2 Transportation

Geant4 simulates the evolution of a particle track in steps, as one would expect from a numerical simulation. However, one may naïvely expect the steps to be in time, which is not the case. Geant4 steps the simulation in length. This turns out to be more natural than time because of the way geometry boundaries are handled. Specifically, it's more

simple to calculate the distance to a boundary from a particle's current position and direction than it is to calculate the time it would take to get to the boundary, saving some CPU time. This means that physics processes, covered in the next section, must calculate a mean free path rather than a mean time so that Geant4 can decide if the particle will undergo some scatter or decay before reaching a boundary. It also means that the equation of motion of a particle under the influence of a force field is a little less familiar looking than that to which we are accustomed. In particular, in a time-based simulation we may expect to write

$$\Delta\vec{p} = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t}\Delta t. \tag{3–5}$$

However, the change in momentum over a length step could be written

$$\Delta p_i = \frac{\partial\vec{p}}{\partial x_i}\Delta x_i, \tag{3–6}$$

which we can rewrite to be a little easier calculate:

$$\Delta p_i = \frac{\mathrm{d}p_i}{\mathrm{d}t}\frac{1}{v_i}\Delta x_i, \tag{3–7}$$

where we have taken the liberty of approximating a partial derivative as a full derivative, since we are dealing with discrete steps. In general, Geant4 will take care of transporting the tracks and users of the library will not have to think about solving the equation of motion explicitly. However, as we will see in following sections, G4CMP needs to specify the equation of motion for electrons with non-scalar mass.

3.2.3 Physics Processes

In Geant4, a step size is equal to the distance between the track's current position and the location of the next boundary in its line of sight unless a physics process tells the simulation that the track will experience a significant change before crossing the boundary. For each step, the track is "teleported" to its new position and the kinematic parameters are updated if needed- e.g., the energy may increase if the track corresponds to a charged particle being accelerated in an electric field. To determine whether an interaction would

happen before crossing into the next boundary, all physics processes must calculate and report a "physical interaction length." It is again useful to refer to the many sources of documentation for Geant4, but for the sake of completeness, we will include a very brief description of some of the basics [44–46].

In the Geant4 process model, each process has an "AtRest" component, an "AlongStep" component, and a "PostStep" component. Often, a physics process only needs to implement one of these components and so are referred to as an "at-rest process," a "continuous process," or a "discrete process," respectively.

The at-rest process mechanism is useful for particles that may spontaneously decay after some time and calculating a length does not make sense. Recall that Geant4 steps are based on lengths, and if a particle is at rest, it would never experience another step if it were not for the at-rest mechanism. If a particle comes to rest and does not have any at-rest processes registered to it, the track is killed.

A discrete process always defines a physical interaction length. That is the length at which some kind of event occurs and alters the track. That could kill the track, change its momentum, or any number of things. For most processes there is not an exact distance until an event. Rather, most things in nature are probabilistic, and since it is very common for processes to be Poisson-like, Geant4 allows a process to simply define a mean free path. Geant4 will then throw a random number from an exponential probability distribution with that mean free path to find the true interaction length.

When Geant4 is calculating the length of the next step, it will choose the distance to the next boundary, as mentioned earlier, unless any of the discrete processes associated with the track return a physical interaction length (PIL) that is less than that distance. If there are multiple PILs that qualify, then Geant4 will choose the shortest one.

To further optimize computation, Geant4 will, by default, not recalculate the interaction length for every discrete process for every step. Rather, it takes the process with the shortest interaction length, triggers its process, and subtracts that length from all of

the other PILs for the track. Then, for the next step, it will recycle those lengths for the processes that did not trigger on the previous step. This way it only needs to recalculate the PIL for the process that had actually triggered.

A continuous process differs from a discrete process in a couple of ways. First of all, all continuous process effects are applied for every step. The second way in which they differ is that a discrete process applies its changes after the position, time, and energy of the track have been updated by the step. The continuous processes only have access to the track's properties at the beginning of the step and the step length, from which it will calculate its overall effect on the track during that step. Once all continuous processes calculate the changes they apply to the track, they are all applied to the track's initial properties simultaneously. It is that new set of properties that any discrete processes would see if they triggered that step.

So, a continuous process is good for processes that can have averaged effects over some length scale. Or, said another way, a continuous process is good for processes that should happen many, many times within some characteristic length of the track's trajectory. A discrete process is for any process for which that is not the case or for which it is difficult to predict the average result of several instances of the effect of the process.

3.3 Crystal Lattices

To be able to simulate something like a SuperCDMS iZIP with Geant4, we must add awareness of the crystal lattice to Geant4's understanding of materials. An understanding of crystal lattices can be gleaned from any standard textbook on the subject (such as Reference [47]) so we will not go through all of the painstaking details here. Rather, we will simply describe the sufficient and necessary parameters to describe a crystal and how that is added in G4CMP.

In a crystal, a lattice is created by repeated translations of a unit cell. There are exactly 14 distinct types of these lattices in three dimensions, to which we refer as Bravais lattices. The lattices can be grouped according to their conventional unit cells into seven

groups: Triclinic, Monoclinic, Orthorhombic, Tetragonal, Cubic, Trigonal, and Hexagonal. Each lattice group imposes constraints on the unit cell's axis lengths and axis angles. By knowing the lengths, angles, and stiffness tensor, we can describe the lattice vibration modes (Section 3.5). The resulting crystal structure also gives rise to the electronic band structure. Table 3.3 shows the crystal symmetry groups supported by G4CMP and the variables that must be defined.

Table 3-3. Crystal symmetry groups and their parameters.

Group	Arguments	Stiffness Components Cpq
cubic	a	C11 C12 C44
tetragonal	a c	C11 C12 C13 C16 C33 C44 C66
orthorhombic	a c	C11 C12 C13 C22 C23 C33 C44 C55 C66
hexagonal	a b c	C11 C12 C13 C33 C44 C66
rhombohedral	a α `deg`|`rad`	C11 C12 C13 C14 C15 C33 C44 C66
monoclinic	a b c α `deg`|`rad`	orthorhombic + C45 C16 C26 C36
triclinic	a b c α β γ `deg`|`rad`	All 21 Cpq
amorphous	none	C11 C12

We include amorphous as an option for G4CMP even though that is not a crystal symmetry group.

Implementing in G4CMP: Geant4 already has awareness of what it refers to as materials. Geant4 includes a NIST catalog of common material properties, such as densities, isotopes, natural abundances, etc. G4CMP adds the concept of a crystal lattice to Geant4. However, it is orthogonal to the included concept of materials for purely technical reasons. Specifically, when creating geometry for a simulation in Geant4, one must use a *G4Material* object in defining the volume object. *G4Material* was not extensible until Geant4 version 10.3- released in December 2016. When G4CMP began it was decided that lattices should be attached to volumes in much the same way that a *G4Material* is. Of course, this orthogonality of features means that sloppy programming could lead to a bug where a germanium lattice is accidentally added to a volume that is

filled with a liquid Helium material. So care should be exercised to make sure that the lattice matches the material.

The relevant G4CMP class are *G4LatticeLogical* and *G4LatticePhysical*. The distinction between "logical" and "physical" attempts to be consistent with Geant4's own nomenclature of logical and physical volumes. The physical lattice is simply the logical lattice, a set of Miller indices, and a rotation angle to indicate the crystal's orientation when attached to a physical solid. In the case of the SuperCDMS Soudan iZIP detectors, the Miller indices of the germanium lattice are $(1, 0, 0)$ aligned along the z-axis of the detector. The logical lattice object stores all of the other inherent information pertaining to the lattice, such as the crystal symmetry group, the lattice constants, Debye temperature, etc.

While it is theoretically possible to derive more properties of the crystal from the lattice structure, G4CMP only uses the lattice symmetry to determine the kinematics of phonons in the crystal (See Section 3.5). For example, calculating the conduction bands of the charge carriers in some general way is a daunting task, thus G4CMP leaves it to the experimentalist to add the relevant band parameters to the configuration file for the crystal (See Table 3-4).

More recently, the G4CMP team has been working with upstream Geant4 developers, which is part of the reason that the new *G4ExtendedMaterial* class was added in Geant4.10.3. This class is also able to be added to a volume because it contains a pointer to the old *G4Material* class, thus allowing us to add new properties to materials without breaking backwards compatibility with the old *G4Material* code. Also added is the new *G4LogicalCrystalVolume* class which is a combination of G4CMP's *G4CMPLatticeLogical* and upstream's *G4LogicalVolume*, meaning that we will no longer have to explicitly attach the lattice to a volume- the lattice properties become part of the volume itself. These new classes are an improvement over G4CMP's approach, which were not possible to do

in G4CMP itself because it required changes to other Geant4 classes. Work is planned to convert G4CMP to the improved upstream solution.

In order to be able to create a lattice and attach it to a volume, G4CMP requires users to supply a configuration file with the parameters they would like to use for the lattice. Table 3-4 shows the parameters covered by the lattice configuration file. Most of the entries will be covered in the following sections as we describe the exact physics that G4CMP attempts to model.

3.4 Electric Fields

3.4.1 Introduction

Geant4 has support for electric, magnetic, and even gravitational fields via the *G4Field* class. It also has pre-implemented uniform fields, which makes attaching a uniform field to a volume very simple. When a non-uniform field is needed, one must create a subclass of *G4Field* and simply define a function that calculates the value of the field at a given position and time. For the SuperCDMS detectors, the electric field near the surface is so intricate that a closed form description of the field in untenable (Refer to Figure 2.2.1). However, software packages exist that can numerically calculate a potential map given a set of constraints. Once we have access to a potential map, we can create a Delaunay triangulation to divide the map. This gives us a tetrahedral mesh such that if we know the containing tetrahedron of some test point, we know that those are the four nearest neighbors to the test point. We can then do a 3D interpolation with the known potentials at those four points and calculate the corresponding field. The next section goes through some of the details. G4CMP depends on a software package called Qhull [53] to do the Delaunay triangulation of the potential map.

3.4.2 Barycentric Coordinates

As mentioned in the previous section, in order to utilize the tetrahedral mesh to calculate the field at a point we need to be able to determine the containing tetrahedron of a point of interest. To be able to complete that search we need to know how to test

Table 3-4. G4CMP crystal parameters.

Parameter	Description
Crystal group	Lattice system (See Reference [47])
Lattice constants	Depending on the lattice system, may require lengths and angles.
Stiffness tensor	Also known as moduli of elasticity. Measures how stiff the lattice is in various directions. For cubic lattices, it is a nine element matrix. (See Reference [47])
Phonon decay parameters (β, γ, λ, μ)	Used in phonon kinematics (See Reference [48])
Phonon isotope scatter rate coefficient	Used to calculate isotope scattering (See Reference [49])
Phonon anharmonic decay rate coefficient	Used to calculate spontaneous phonon decay (See Reference [49])
Density of states for phonon modes	Equilibrium phonon mode distributions (See Reference [50])
Debye energy	The Debye energy for the crystal. Used to limit energy of incoming phonons.
Semiconductor band gap energy	The energy it takes to take an electron from the valence band into the lowest energy conduction band.
Average electron-hole production energy	The average energy needed to free an electron from the valence band. Typically larger than the above.
Hole effective mass	Hole charge carrier effective mass.
Electron effective mass	Electron charge carrier effective mass (Expects tensor elements).
Speed of sound	Isotropic speed of sound for the crystal.
Conduction valley direction	Euler angles that rotate z-axis into direction of the electron's conduction bands. May enter as many lines in config file as needed.
Fano factor	Used to convert true ionization energy to measured ionization energy [51].
Inter-valley scattering field constant	Constant used in inter-valley scattering rate fit [52].
Inter-valley scattering coefficient	Constant used in inter-valley scattering rate fit [52].
Inter-valley scattering rate exponent	Constant used in inter-valley scattering rate fit [52].

whether a tetrahedron contains a point or not. One method to determine whether a point resides inside of a tetrahedron is to transform the point's coordinates into the barycentric coordinates of the tetrahedron. To calculate the barycentric coordinates of a test point, $\vec{r}$, relative to a tetrahedron made up of four fixed points, $\vec{r}_i$,

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \mathbf{T}^{-1} \left(\vec{r} - \vec{r}_4 \right), \tag{3–8}$$

$$\lambda_4 = 1 - \lambda_1 - \lambda_2 - \lambda_3, \tag{3–9}$$

where

$$\mathbf{T} = \begin{pmatrix} x_1 - x_4 & x_2 - x_4 & x_3 - x_4 \\ y_1 - y_4 & y_2 - y_4 & y_3 - y_4 \\ z_1 - z_4 & z_2 - z_4 & z_3 - z_4 \end{pmatrix}. \tag{3–10}$$

So the barycentric coordinates are just a weighted, normalized, measure of how close a point is to each fixed point. In other words, if $\vec{r} = \vec{r}_2$, then the barycentric coordinates are $\vec{\lambda} = (0, 1, 0, 0)$. A value of one indicates the point is at a vertex. A value of zero indicates that the point is as far away as possible from the vertex, while still being inside the tetrahedron. Thus, if any of the barycentric coordinates come out to be negative (or greater than one, because one implies the other), then the point does not lie within the tetrahedron. Note that the numbering of the tetrahedron's vertices is arbitrary.

3.4.3 Searching the Tetrahedral Mesh

Now that we know how to test whether a point is in a tetrahedron, we can develop a method to search the mesh for the unique tetrahedron that contains a point. It turns out that finding a negative barycentric coordinate does not only tell us that a point is not inside the tetrahedron, but it actually tells us *where* the point lies relative to the tetrahedron. Recall that a barycentric coordinate equal to zero means that the point lies as far away from the corresponding vertex as is possible while still being inside the

tetrahedron; i.e., it lies on the face of the tetrahedron that is opposite of the vertex. Therefore, a negative barycentric coordinate implies that the correct tetrahedron lies somewhere in a direction pointed outward from that opposing face.

This means that the next tetrahedron to test is the neighbor that shares that face with the current tetrahedron. So, when we construct our tetrahedral mesh, we must store a map of the neighbors of each tetrahedron, and also encode that information such that we know which neighbor corresponds to which face. This can be done with a simple array of ID numbers, with sufficient care. We are still limited by not being able to "skip over" tetrahedra in this search, but this gives us a clear directed walk algorithm that is quite fast in practice. Combining this search algorithm with caching the last correct tetrahedron, and the electric field method has to test very few tetrahedra.

Algorithm 3.1 Tetrahedral search algorithm.

```
if Previous Tetrahedron exists then
    CurrentTetra = PreviousTetra
else
    CurrentTetra = Tetra[0]
end if
while Not Found do
    Get λ⃗ from r⃗ and CurrentTetra                    ▷ Eq. 3–8
    if Any λ_j <= 0 then
        k = min(λ_j)
        CurrentTetra = Get k neighbor from CurrentTetra
    else
        Found
    end if
end while
```

3.4.4 Calculating the Electric Field

Now that we can find the containing tetrahedron for a given point, getting the electric field is simple a matter of computing the gradient of the potential at the point. This can be done by a trilinear interpolation of the known potentials at the vertices of the containing tetrahedron. Luckily, finding the interpolated potential inside of a tetrahedron

is made trivial by use of the barycentric coordinates:

$$V(\vec{r}) = V_1\lambda_1 + V_2\lambda_2 + V_3\lambda_3 + V_4\lambda_4. \tag{3–11}$$

And thus calculating the field is a matter of taking the difference between several interpolated potential values. The G4CMP class, *G4CMPMeshElectricField*, enables users to simply use a text file potential map to specify the electric potential definition and does the trilinear interpolation over the tetrahedral mesh without any further configuration from the user.

3.5 Phonon Physics

3.5.1 Introduction

Phonons are quantized, vibrational excitations of the periodic crystal lattice. They can be characterized by the straight-forward Hamiltonian,

$$H = \sum_i \frac{p_i^2}{2m} + \sum_{i \neq j} \frac{m\omega^2}{2} (x_i - x_j)^2 , \tag{3–12}$$

where m is the mass of the vibrating atom and ω is the frequency of oscillation between the i-th atom and its nearest neighbors [47]. This is, of course, making the assumption that the ions in the crystal tend to be associated with a particular Bravais lattice site. Meaning that G4CMP would not be well-suited, in its current state, to simulation phonons in a material near its melting point or of a quantum solid.

3.5.2 Transportation

In order to understand the propagation of phonons in the lattice we must get expressions for the group and phase velocities:

$$v_p = \frac{\omega}{k}, \tag{3–13}$$

$$\vec{v}_g = \frac{\partial\omega}{\partial\vec{k}}, \tag{3–14}$$

where $\vec{k}$ is the wave vector and ω is the phonon frequency. The relationship between ω and $\vec{k}$ can be found by using the Hamiltonian to get an equation of motion. G4CMP currently

only supports acoustic mode phonons, and thus the equation of motion solutions that interest us take the form

$$\rho\omega^2\epsilon_\mu = \sum_\tau \left(\sum_{\sigma\nu} c_{\mu\sigma\nu\tau} k_\sigma k_\nu \right) \epsilon_\tau, \tag{3–15}$$

where ρ is the mass density of the crystal, ϵ is the polarization vector of the phonon, and c is the elasticity tensor of the crystal [50]. Luckily, for the case of cubic lattices, the elasticity tensor reduces to a mere three independent components [47]. Therefore, equation 3–15 reduces to solving for the eigenvalues and eigenvectors of

$$\rho\omega^2\vec{\epsilon} = \begin{pmatrix} c_1k_x^2 + c_3(k_y^2+k_z^2) & (c_2+c_3)k_xk_y & (c_2+c_3)k_xk_z \\ (c_2+c_3)k_yk_x & c_1k_y^2 + c_3(k_x^2+k_z^2) & (c_2+c_3)k_yk_z \\ (c_2+c_3)k_zk_x & (c_2+c_3)k_zk_y & c_1k_z^2 + c_3(k_x^2+k_y^2) \end{pmatrix} \vec{\epsilon}, \tag{3–16}$$

where we have collapsed our former notation to represent the reduction in independent terms of the elasticity tensor:

$$c_1 = c_{xxxx} = c_{yyyy} = c_{zzzz} \tag{3–17}$$

$$c_2 = c_{xxyy} = c_{yyzz} = c_{zzxx} \tag{3–18}$$

$$c_3 = c_{xyxy} = c_{yzyz} = c_{zxzx} \tag{3–19}$$

The resulting eigenvectors of equation 3–16 are the polarization vectors and the eigenvalues are the relationships between ω and $\vec{k}$ for the three different phonon modes in a cubic lattice: longitudinal, fast transverse, and slow transverse.

Implementing in G4CMP: Geant4 used to not have phonons. So the initial G4CMP developers created three new particles: G4PhononLong, G4PhononTransFast, and G4PhononTranSlow, which correspond to the three acoustic phonon modes in the cubic lattice. The particles, like optical photons, should not let Geant4 calculate their velocity from Equations 3–3. Rather, Equations 3–13 and 3–14 need to be solved to calculate the velocity for a given phonon's wave vector. So, G4CMP solves the eigenvector equation for

a wide range of values at initialization and generates a table so that each phonon physics process can correctly set the track's new velocity after any change in momentum.

G4CMP contains code to solve Equation 3–15 for its eigenvectors and eigenvalues via the *G4CMPPhononKinematics* class. It was attempted to utilize the solver during the simulation to calculate the exact values for a given wave vector. However, this "in-vivo" method slowed the simulation by an approximate factor of three compared to the look-up table method, with no significant difference in output. That does leave the door open for a hybrid approach where the table can be calculated as-needed throughout the simulation.

These new phonon particles were taken upstream into Geant4. However, G4CMP still supplies its own overridden versions because of developments and improvements that have not made it upstream yet.

3.5.3 Anharmonic Decay

Because of nonlinear terms in the elastic coupling constants, higher energy phonons are able to decay into two lower energy phonons. At fairly low energies and in crystals that are mostly isotropic, taking the isotropic approximation for anharmonic decay is appropriate [54]. This approximation leads to the conclusion that only longitudinal mode phonons can decay, for kinematic reasons [55]. The allowed decay channels are $L \to L + T$ and $L \to T + T$, and the overall rate is

$$\Gamma = A\nu^5, \tag{3–20}$$

where A is a parameter of the lattice. Treating each decay channel independently, one can calculate the relative proportion of decays into each channel for a material [48].

The remaining piece for fully modeling a low energy phonon decay is the kinematic distributions of the daughter phonons. The calculations for the three-body problem are messy, even in the isotropic approximation, but luckily have been carried out for us already [48, 50], which I will quote.

For the $L \to L + T$ channel

$$\Gamma \sim \frac{1}{x^2}(1-x^2)^2[(1+x)^2 - \delta^2(1-x)^2][1+x^2-\delta^2(1-x^2)^2]^2, \tag{3–21}$$

$$\cos\theta_L = \frac{1+x^2-\delta^2(1-x^2)}{2x}, \tag{3–22}$$

$$\cos\theta_T = \frac{1+x^2-\delta^2(1-x^2)}{2x}, \tag{3–23}$$

where $x = E_{daughterL}/E_{parentL}$, $\delta = v_L/v_T$, and $\frac{\delta-1}{\delta+1} < x < 1$. For the $L \to T + T$ channel

$$\Gamma \sim (A + B\delta x - Bx^2)^2 + \left[Cx(\delta - x) - \frac{D}{\delta - x}\left(x - \delta - \frac{1-\delta^2}{4x}\right)\right]^2, \tag{3–24}$$

$$\cos\theta_{T_1} = \frac{1-\delta^2(1-x)^2+\delta^2x^2}{2\delta x}, \tag{3–25}$$

$$\cos\theta_{T_2} = \frac{1-\delta^2x^2+\delta^2(1-x)^2}{2\delta(1-x)}, \tag{3–26}$$

where $x = \delta\frac{E_{T_1}}{E_{parentL}}$, $\delta = v_L/v_T$, $\frac{\delta-1}{2} < x < \frac{\delta+1}{2}$, $A = \frac{1}{2}(1-\delta^2)[\beta + \lambda + (1+\delta^2)(\gamma+\mu)]$, $B = \beta + \lambda + 2\delta^2(\gamma+\mu)$, $C = \beta + \lambda + 2(\gamma+\mu)$, and $D = (1-\delta^2)(2\beta + 4\gamma + \lambda + 3\mu)$; β and γ being third order elastic constants, and γ and μ are the Lamé constants. And of course, because this is the isotropic approximation, the other scattering angle of the first daughter phonon, relative to the original phonon wave vector, is randomly distributed: $\phi_1 = [0, 2\pi]$ and $\phi_2 = \phi_1 - \pi$.

Implementing in G4CMP: Despite the ugliness of the calculations, implementing the results in G4CMP is straight-forward. We simply define a discrete process where the mean free path is given by

$$\lambda = \frac{v_L}{\Gamma} \tag{3–27}$$

that is only registered to apply to longitudinal phonons. Then, upon the process triggering, we perform inverse transform sampling to get the correct decay channel and each of the initial parameters for the daughter phonons.

3.5.4 Isotope Scattering

Phonons also scatter off of defects in the crystal lattice with a bulk rate

$$\Gamma = B\nu^4, \tag{3–28}$$

where B is a constant for a given material[49]. However, an individual phonon scatters anisotropically:

$$\gamma \propto \frac{|\vec{\epsilon}\cdot\vec{\epsilon}\,'|^2}{\nu'^3}, \tag{3–29}$$

where the primed variables denote the values after the scatter. In this process, the phonon may switch modes, and the equation may be used to determine the long-term population density of the different modes [50]. It is also reasonable to approximate isotope scattering as isotropic as long as the detection time is larger than the ballistic [49].

Implementing in G4CMP: For G4CMP, we take the most simple approach, as that is still accurate in the relevant domains. Isotope scattering is implemented as a discrete process where we use equation 3–28 and equation 3–27 to determine the mean free path. Then we take the predetermined mode populations and isotropically create a new phonon track with the same energy as the original, but with a randomly selected direction and mode. Figure 3-1 shows a visualization of a G4CMP phonon simulation.

3.5.5 Phonon Interactions in Superconductors

The majority of the physics studied and implemented in G4CMP is for particles in semiconductors. However, SuperCDMS collects phonon energy via their interaction in superconducting aluminum, as was discussed in Section 2.2.4.

When a phonon enters a superconductor, if it has an energy that is greater than twice the superconducting gap energy, then it can break Cooper pairs into its constituent quasiparticles. Those quasiparticles that are produced can, themselves, produce phonons by scattering in the solid. Those "downconverted" phonons can then break more Cooper pairs if they have high enough energies, and so on.

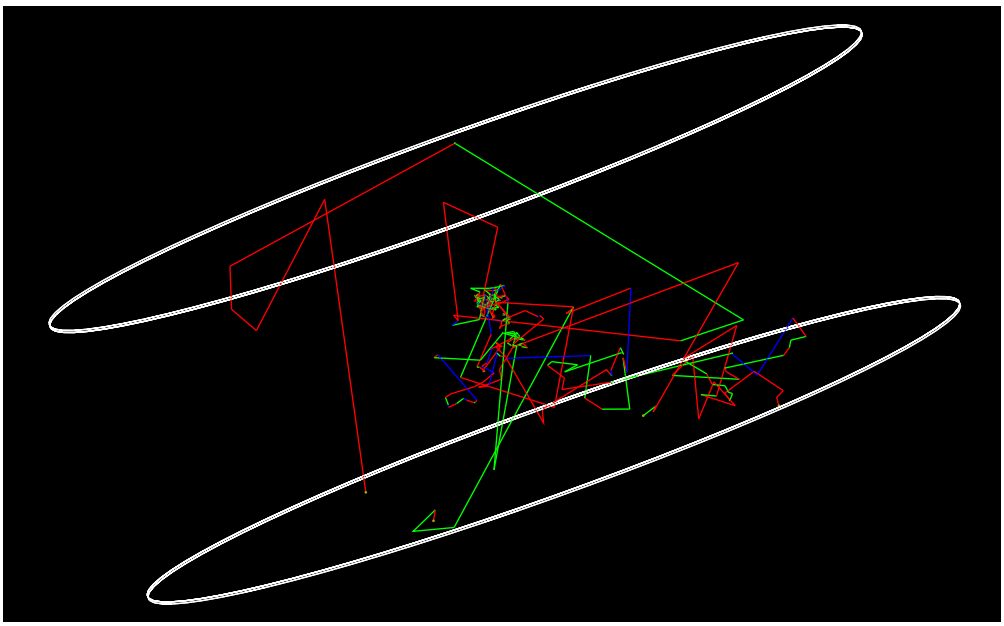


Figure 3-1. G4CMP phonon demo. In the above figure, we simulate a single, 1 meV, longitudinal phonon being created at the center of a Soudan iZIP detector. It creates daughter phonons via anharmonic decay and all of the phonons are able to mix modes via isotope scattering. Longitudinal phonon tracks are drawn in blue, fast transverse phonons are green, and slow transverse are red. Notice that most scatters happen in a small volume around where the initial phonon is created, then the phonons become ballistic.

According to Kaplan, et al.[56], the rate of Cooper pairs being broken by a phonon of energy Ω is

$$\Gamma_{p\to qp} = \frac{2\pi\rho(E_F)\alpha^2(\Omega)}{\hbar I}\int_{\Delta}^{\Omega-\Delta}\frac{\mathrm{d}E}{\sqrt{E^2-\Omega^2}}\frac{E\,(\Omega-E)+\Delta^2}{\sqrt{(\Omega-E)^2-\Delta^2}}, \tag{3–30}$$

where $\rho(E_F)$ is the density of states of the electrons at the Fermi energy, $\alpha^2(\Omega)$ is the matrix element of the electron-phonon interaction squared, and I is the ion number density. The rate at which quasiparticles of energy E decay by phonon emission is

$$\Gamma_{qp\to p} = \frac{2\pi}{\hbar Z_1(0)}\int_0^{E-\Delta}\mathrm{d}\Omega\alpha^2(\Omega)F(\Omega)\mathrm{Re}\left\{\frac{E-\Omega}{\sqrt{(E-\Omega)^2-\Delta^2}}\right\}\left\{1-\frac{\Delta^2}{E(E-\Omega)}\right\}, \tag{3–31}$$

where $Z_1(0)$ is a factor to renormalize the electron-phonon interaction. We can work backwards from Equations 3–30 and 3–31 to obtain the energy distribution of one quasiparticle

of the broken Cooper pair [50].

$$P_{qp}(E) = \frac{E}{\sqrt{E^2 - \Delta^2}} \frac{E - \Omega}{\sqrt{(E-\Omega)^2 - \Delta^2}} \left(1 + \frac{\Delta^2}{E\,(\Omega - E)}\right) \tag{3–32}$$

The quasiparticles are constrained to conserve energy, so that $E_2 = \Omega - E_1$. As above, we can get the energy distribution of emitted phonons from quasiparticles [50]:

$$P_{phonon}(\Omega) = \Omega^2 \frac{E - \Omega}{\sqrt{(E-\Omega)^2 - \Delta^2}} \left(1 - \frac{\Delta^2}{E\,(E - \Omega)}\right), \tag{3–33}$$

which assumes that $\alpha^2(\Omega)F(\Omega) \propto \Omega^2$ at low energies [56]. We also assume that quasiparticle recombination happens over length scales that will tend to be much larger than phonon emission and thus are suppressed.

Implementing in G4CMP: G4CMP does not generally support superconductor materials at the time of this writing, though it does implement the Kaplan quasiparticle model just described. The way G4CMP makes use of the down converting physics is by making several approximations about the quasiparticles rather than actually generating true particle tracks to simulate. This makes it possible to simulate the heat response of a thin film superconductor to absorbed phonons.

First, we assume that the phonon travels along the thin dimension of the film, so that if the calculated scatter distance is larger than twice the thickness, d, of the film, no Cooper pairs are broken and the phonon escapes back into the crystal. Second, we assume all further Cooper pair breaking happens at $z = d/2$, so that the escape conditions are always $|\vec{x}| > n/2 \times d$ for $n = \{1, 3\}$. Then it is straight-forward to implement the full cascade process, as described in Algorithm 3.2.

Notice that we are assuming that any quasiparticles that are energetic enough to emit phonons will do so before being absorbed by the sensors (indeed, our algorithm does not permit the quasiparticles to travel at all, except to finally be absorbed).

Algorithm 3.2 Phonon-quasiparticle down conversion.

```
repeat
    for all phonons do
        if Ω < 2Δ then
            Remove phonon from simulation
        end if
        Calculate distance to next Cooper pair breaking, |x⃗|
        Randomly choose direction: away from crystal (n = 3), toward crystal (n = 1)
        if |x⃗| < n/2 × d then
            Create quasiparticle pair according to Eq. 3–32
        else
            Phonon escapes back into crystal
        end if
    end for
    for all quasiparticles do
        if E < 3Δ then
            Absorb quasiparticle energy, then remove from simulation
        end if
        Create phonon according to Eq. 3–33
    end for
until all phonons have Ω < 2Δ and all quasiparticles have E < 3Δ
```

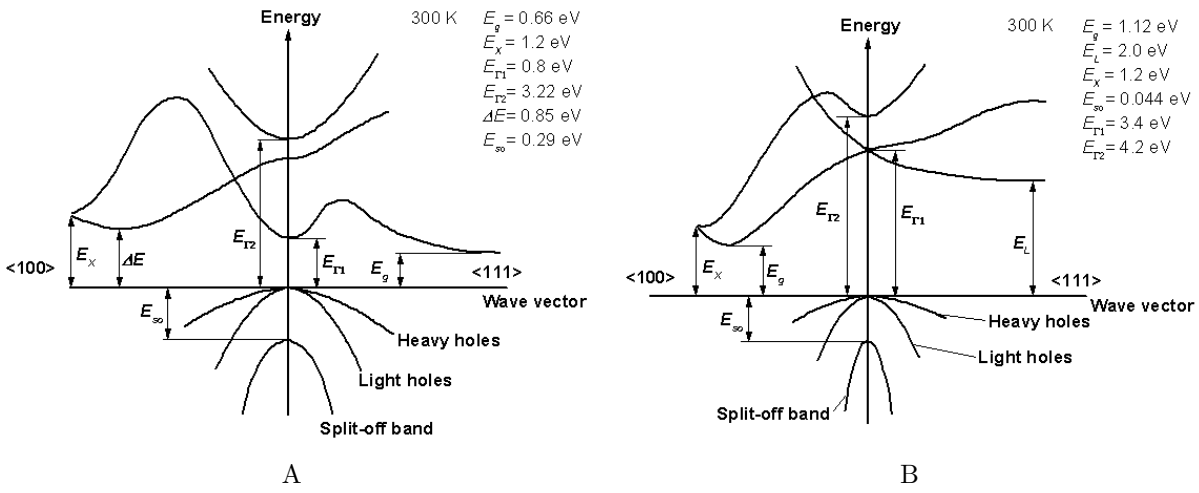
3.6 Charge Physics

3.6.1 Introduction

Germanium and silicon are both indirect-gap semiconductors with diamond cubic crystal lattice structures (cubic symmetry group in Table 3.3). Figure 3-2 illustrates what indirect-gap means. SuperCDMS dark matter detectors operate at cryogenic temperatures. As a result of these two things, the physics detailed in this section are relevant to the specific overlap of these domains, even if not explicitly stated. As a secondary result, G4CMP only currently implements physics that models this particular regime. It is noted, of course, that the goal of G4CMP is to be flexible enough that extending the physics to go to higher energies or other types of semiconductors is quite feasible.

3.6.2 Transportation

One consequence of the low temperature of the crystals, as well as the relatively weak electric field applied applied across the detectors (in "standard iZIP mode"), is that the



A B

Figure 3-2. Conduction bands of germanium and silicon. A) Germanium. B) Silicon. Notice that for both germanium and silicon, the lowest conduction band energy is away from the origin. In germanium the L band is the minimum, and in silicon the X band is the minimum. Electrons will tend to prefer to stay in those conduction bands because they are the most energetically favorable.

electrons that are freed from the valence band in either material are most likely to spend their time in the lowest energy conduction bands. As a result, we have only implemented the lowest energy set of conduction bands for the conduction electrons. It should also be noted that we will sometimes refer to the conduction bands as valleys because of the tendency of the charge carriers to stay near the parabolic bottom of the conduction band's energy.

We find that the Drude-Sommerfield model does not work well for modeling the individual electron kinematics in these materials [47] due to the anisotropy of the conduction bands at the lowest energies. However, since the valence bands are parabolic and isotropic, the free-electron model does work well for describing the conducting holes. Further, we can perform some coordinate transformations to regain some of the simplicity of free-electron physics for non-transportation physics as we will see in Section 3.6.3.

Hole Transportation: Transportation of hole charge carriers is very simple. Holes in germanium and silicon have an effective mass that is scalar, like most particles with which we are familiar. It is also conventional to treat the holes as positively charged with

a positive mass (or can be treated as having negative charge and negative mass which semantically seems more correct in some sense, as the absence of an electron). So, their equation of motion in an externally applied electric field is simply

$$\ddot{\vec{x}} = \frac{e\vec{E}}{m^*}, \tag{3–34}$$

where e is the positron charge, $\vec{E}$ is the applied electric field, and m^* is the effective mass of the hole.

Electron Transportation: Transportation for the electron at these low energies, in an indirect gap semiconductor, is a little more difficult compared to a free electron or even a hole quasiparticle. Because the constant energy contours of the low energy conduction bands in both silicon and germanium are ellipses, rather than spheres, the effective mass can no longer be represented as a single value. Said another way: the electron will more readily accelerate in directions perpendicular to the valley principle axis than in the parallel direction.

In germanium, there are eight L conduction bands. An electron's effective mass can thus be represented by a tensor, where one of the indices corresponds to the L band in which it is propagating. By symmetry, four of those bands are redundant, i.e., going "backwards" in band #5 is indistinguishable from going "forward" in band #1. So, the tensor effective mass of the electron in Ge, is basically a set of four 3×3 matrices- one for each band it may occupy.

We often find it much more simple to work in the coordinate system of the conduction band in which the electron is propagating. If we perform a rotation from some global coordinate system to one in which the first axis is aligned with the primary axis of the constant energy ellipse of the band, the mass matrix is always

$$\mathbf{M} = \begin{pmatrix} m_{\parallel} & 0 & 0 \\ 0 & m_{\perp} & 0 \\ 0 & 0 & m_{\perp} \end{pmatrix}. \tag{3–35}$$

So the kinetic energy of an electron traveling along one of the [111]-symmetric conduction bands would be (non-relativistic)

$$E_k = \frac{p_1^2}{2m_\parallel} + \frac{p_2^2}{2m_\perp} + \frac{p_3^2}{2m_\perp}. \tag{3–36}$$

This causes the equation of motion for the electron to be

$$\ddot{\vec{x}} = -e\mathbf{M}^{-1}\vec{E}, \tag{3–37}$$

where $\mathbf{M}$ is the effective mass matrix of the electron.

This also leads to the very important, but sometimes glossed over, result of the velocity vector not being in the same direction as the momentum vector of an electron traveling through one of these anisotropic bands. It forces us to revisit our physics 101 classes and recall whether we should be describing a velocity or a momentum. For example, it is better for us to think of Newton's second law as

$$\sum \vec{F} = \frac{d\vec{p}}{dt}, \tag{3–38}$$

than as we sometimes see it:

$$\sum \vec{F} = m\ddot{\vec{x}}, \tag{3–39}$$

because in the latter form, we must also account for the fact that the mass is now a matrix ($m \rightarrow \mathbf{M}$) and will change the direction of $\ddot{\vec{x}}$ to no longer point in the same direction as the applied force (in our case, the electric field). The fact that the electron's momentum will tend to be anti-parallel to the field, but its velocity will not, is referred to as oblique propagation. This idea that the velocity and momentum vectors can point in different directions has subtle implications that change the implementation of almost every aspect of our Monte Carlo package.

Implementation in G4CMP: A fundamental problem for us implementing this kind of condensed matter physics in the Geant4 framework is that the effective mass is not always a scalar number.

In particular, Geant4 assumes that 3–1 holds for all particles, where m is a scalar number. During a simulation, Geant4 does not even store the momentum magnitude of a particle, opting to instead calculate it as needed from the above relativistic equation from the particle's energy, mass, and momentum direction, which we will see is also an issue for us.

Because G4CMP has the practical goal of being useful to as many researchers as possible, it was decided to not fork and customize Geant4, but rather to attempt to work "on top of" Geant4, so that researchers who are already using Geant4 can add G4CMP to their workflow with minimal conflicts.

Before describing the approach that G4CMP takes to solve the effective mass issue, we will provide an argument that there is no correct way to transform the effective mass of a particle in an oblique conduction band into a single number for use by Geant4's core particle tracking engine.

From Equation 3–36 it becomes clear that there is no constant, scalar, value one could assign to the mass to get the correct kinematics. This is because two particles with the same total momentum magnitude will have different energies depending on the direction of the momentum vector. The only way to reconcile this equation with equation 3–1 would be to have the mass become an explicit function of the momentum:

$$m^* = m^*(\vec{p}) = |\vec{p}|^2 \left(\frac{m_\parallel}{p_1} + \frac{m_\perp}{p_2} + \frac{m_\perp}{p_3} \right). \tag{3–40}$$

Geant4 does support dynamic changing of the particle mass from its defined value (for e.g., off-shell bosons and ionized nuclei). However, those mass updates can only happen discretely (even in a so-called "continuous process"), whereas this would need to be continuous due to the possible acceleration of a particle in a field.

Thus, in order to accurately describe electron propagation in Geant4, G4CMP has to implement a layer over Geant4's track to translate between what the scientist understands as the momentum, velocity, mass, energy, etc., and what information we can give to

Geant4's core engine to produce the correct tracking. At the end of the day, we need Geant4 to have the correct value of the velocity's magnitude and direction so that it can correctly asses when the track will cross a boundary and so that it can properly use the physical interaction lengths of the electron physics processes to determine and report the step sizes.

This is simple to do for abrupt changes in energy, such as from emitting a Luke phonon. When we discover what the new velocity and energy of the particle are, we simply calculate what energy value Geant4 will need to have to recreate the correct velocity. We can either solve Equation 3–3 for E, or solve the non-relativistic version,

$$E = \frac{1}{2} m_{G4} |\vec{v}\,|^2, \tag{3–41}$$

where m_{G4} is the scalar mass that is given to Geant4 in the particle's definition, which can be any number at all as far as G4CMP cares, as long as it's constant.

This, of course, means that *only* the track's velocity will be correct. Its energy, momentum, mass, and momentum direction are incorrect. Luckily, Geant4 has a feature called auxiliary track information. With this feature, a user can attach arbitrary data to a G4Track object. This means that we can attach the charge carrier's true mass to the track (actually, we attach a pointer to the lattice of the material, which holds information about the charge carriers' effective masses) and then our G4CMP physics processes can access the correct mass and velocity. From those it can calculate the correct energy and momentum and all is well. One frustrating clash of nomenclature is that since Geant4 assumes the velocity and momentum vectors point in the same direction, it uses the "velocity," which is actually the speed, and the "momentum direction" of the track to propagate the track and find distances to boundaries, etc. This means that our translation layer not only has to set an incorrect energy value, but also has to set the momentum direction to be the velocity direction. Luckily, this only had to be done once, and now anyone who wants to work with G4CMP or add physics to it only has to remember to go

through G4CMP's track helper layer and does not have to worry about what Geant4 is doing underneath.

One more place where this issue arises is in the equation of motion for the charge carriers, since they are being accelerated by any present electromagnetic field. In Geant4, we can supply our own "right-hand side" of the equation of motion. Specifically, Geant4 will give the position and "momentum" (velocity times mc) as well as the electric and magnetic field components as input. Then it is our job to return the velocity direction unit vector, $\mathrm{d}\vec{p}/\mathrm{dx}$, and the lab time of flight per step length.

The velocity unit vector is easy because Geant4 and G4CMP both mean the same thing when they refer to velocity. However, the momentum that Geant4 gives as input is not the correct momentum, but just the velocity times the scalar mass that Geant4 has for the track. So, for the velocity unit vector we just have to manipulate the given momentum:

$$\vec{v}_{out} = \frac{\vec{p_{in}}}{|\vec{p_{in}}|}, \tag{3–42}$$

where *in* and *out* denote quantities that Geant4 gave as input to the right-hand side and quantities that we return from the function, respectively.

Getting $\mathrm{d}\vec{p}/\mathrm{d}x$ is a little trickier. Using the field input values, we can easily obtain the true force on the particle:

$$\vec{F} = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = q(\vec{E} + \vec{v} \times \vec{B}), \tag{3–43}$$

where q is the charge of the particle and $\vec{v}$ can be calculated from the input momentum as above. To go from $\mathrm{d}\vec{p}/\mathrm{d}t$ to $\mathrm{d}\vec{p}/\mathrm{d}x$ we refer to Equation 3–7.

$$\frac{\mathrm{d}\vec{p}}{\mathrm{d}x} = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t}\frac{1}{|\vec{v}|} = \frac{1}{|\vec{v}|}q(\vec{E} + \vec{v} \times \vec{B}) \tag{3–44}$$

However, we are still not done. Recall that this is the *true* momentum change for the step. But G4CMP cannot give the true momentum change to Geant4- it must report a momentum change that will correspond to the correct *velocity* change for the step. This must be done by transforming the result of Equation 3–44 with the particle's true tensor

mass and the scalar mass that Geant4 has stored:

$$\left.\frac{\mathrm{d}\vec{p}}{\mathrm{d}x}\right|_{out} = m_{G4}\mathbf{M}^{-1}\left.\frac{\mathrm{d}\vec{p}}{\mathrm{d}x}\right|_{true}, \tag{3–45}$$

where *out* and *true* denote the value to return to Geant4 and the value obtained from Equation 3–44, respectively; and $\mathbf{M}$ is the tensor mass of the particle.

3.6.3 Neganov-Luke Phonon Emission

When a charge carrier moves through a crystal faster than the speed of sound of the material, it may emit phonons in a process that is analogous in many ways to Cerenkov radiation for light[16, 17]. We refer to these emitted phonons as Neganov-Luke phonons, or Luke phonons for brevity. As we will see, use of Fermi's Golden Rule to calculate the emission rate is complicated by the kinematic differences between electrons and holes. Fermi's Golden Rule states that the transition probability per unit time from one energy eigenstate into another eigenstate is given by

$$\Gamma_{i\to f} = \frac{2\pi}{\hbar}|\langle f|H|i\rangle|^2 g(E), \tag{3–46}$$

where i and f are the initial and final eigenstates, respectively; H is the perturbed Hamiltonian; and $g(E)$ is the final density of energy states.

We treat the emission of a phonon as an elastic scatter because of the low energies involved, thus ignoring Umklapp processes which do not conserve momentum.

Using conservation of energy and momentum for the three-body problem yields

$$k'^2 = k^2 + q^2 - 2kq\cos\theta \tag{3–47}$$

$$q = 2(k\cos\theta - k_s) \tag{3–48}$$

$$\cos\phi = \frac{k^2 - 2k_s(k\cos\theta - k_s) - 2(k\cos\theta - k_s)^2}{k\sqrt{k^2 - 4k_s(k\cos\theta - k_s)}}, \tag{3–49}$$

where k and k' are the initial and final wave numbers of the charge carrier, respectively; q is the wave number of the emitted phonon; θ is the angle between the initial charge carrier wave vector and the emitted phonon; ϕ is the angle between the final charge carrier wave

vector; and k_s is the wave number of the charge carrier if it were moving at exactly the speed of sound,

$$k_s = \frac{m v_s}{\hbar}, \tag{3–50}$$

where m is a scalar effective mass for the carrier (which we will discuss further for the case of particles with tensor masses in the next section), and v_s is the speed of the longitudinal mode phonons in the material. In doing the calculations for Equations 3–47–3–49 we assume a linear dispersion relation for the phonons,

$$\omega = v_s q, \tag{3–51}$$

which is approximately true at cryogenic temperatures. We can rewrite Equation 3–46 specifically to this process.

$$\Gamma_{\vec{k}\to\vec{k}'} = \frac{2\pi}{\hbar}\left|\left\langle \vec{k}' + \vec{q}\middle| H \middle|\vec{k}\right\rangle\right|^2 \delta(E - E' - \hbar\omega) g(E) \tag{3–52}$$

The matrix element for the emission can be written as

$$\left|\left\langle \vec{k}' + \vec{q}\middle| H \middle|\vec{k}\right\rangle\right|^2 = \frac{\Xi^2 \hbar}{2V\rho v_s} q(n_q + 1), \tag{3–53}$$

where n_q is the phonon occupation number given by Bose-Einstein statistics as $(\mathrm{e}^{\hbar\omega/k_B T} - 1)^{-1}$, ρ is the mass density of the lattice, and Ξ is the deformation potential. We can then integrate Equation 3–52 with respect to energy to get the angular distribution of Neganov-Luke phonon emissions. To do the integration we need the density of energy states, for which we can use the free-electron gas density of states,

$$g(E) = \frac{m}{\hbar^2 \pi^2}\sqrt{\frac{2mE}{\hbar^2}}. \tag{3–54}$$

We also assert that $n_q \to 0$ at very low temperatures, and using the free-electron density of states to get

$$P(k,\theta)\ \mathrm{d}\theta = \frac{v_s}{l_0}\left(\frac{k}{k_s}\right)^2 \left(\cos\theta - \frac{k_s}{k}\right)^2 \sin\theta \mathrm{d}\theta \tag{3–55}$$

where l_0 is a characteristic scattering length,

$$l_0 = \frac{\pi\hbar^4\rho}{2m^3\Xi^2} \tag{3–56}$$

The final piece of the puzzle comes from integrating Equation 3–55 over theta for all valid thetas: $0 \leq \theta \leq \arccos(k_s/k)$ (using Equation 3–48), giving us the overall rate of Luke phonon emission,

$$\Gamma = \frac{v_s}{3l_0}\left(\frac{k}{k_s}\right)^2\left(1-\frac{k_s}{k}\right)^3 \tag{3–57}$$

Holes: Going from Equation 3–57 to a mean free path is not as simple for charge carriers as it was for phonons. In the phonon case, we could simply divide the current speed by the rate (assuming it is Poisson-like) and obtain a mean free path. This will not work for charge carriers because the acceleration over the resulting length step will tend to change the rate significantly. There are other complications as well: If the charge carrier is traveling slower than the speed of sound at the time of computation, then the mean free path for emitting a Luke phonon would be infinite; also if the charge carrier is currently be slowed by the field, the probability distribution for emitting a phonon is not shaped like an exponential, as is typically assumed by giving Geant4 a mean free path value.

Solving this conundrum in a general way that also has reasonable computing performance is not an easy task. More work on this process is described in Section 5.4.

The approach that has been chosen for G4CMP is to force a recalculation of the mean free path at pre-defined step sizes, such that the rate change over a step of that size is small. These pre-defined step sizes are a function of the electric field strength and the typical maximum wave vector of a charge carrier under such conditions [50].

The way this must be done for use with Geant4 is that Luke emission is actually implemented across two processes: *G4CMPLukeScattering* and *G4CMPTimeStepper*. *G4CMPTimeStepper*'s only job is to always return a physical interaction length defined by the maximum wave vector for the material. The process itself does not modify the track in any way. *G4CMPLukeScattering* sets a flag that tells Geant4 to always recalculate its

mean free path at each step, even if it did not limit the step. This means that the mean free path that is calculated by *G4CMPLukeScattering* will be roughly constant over the maximum step size of the charge track simulation.

Electrons: Electrons pose a complication because of the elliptical constant-energy contours of the conduction bands. As mentioned in Section 3.6.2, this is often expressed by the effective mass of the electron being represented as a tensor. Besides needing to replace the scalar mass in the kinematics calculations above, this also changes the deformation potential from a scalar value to something anisotropic, complicating the calculations further.

To simplify these calculations, we can instead perform a coordinate transformation into a space in which the constant-energy contours are spherical[57]. This allows us to use the same prescription as the holes for calculating Luke phonon emissions of the electrons using an effective scalar mass and an isotropic deformation potential. We refer to this transformation as a Herring-Vogt transformation.

In order to apply this transformation, we must first rotate our coordinate system such that one of the axes is aligned with the direction of the conduction valley of the electron, as in Equation 3–35. Once in this frame, we can apply a transformation to "squeeze" the mass matrix into a scalar value.

$$\vec{k}^* = \begin{pmatrix} \sqrt{\frac{m^*}{m_\parallel}} & 0 & 0 \\ 0 & \sqrt{\frac{m^*}{m_\perp}} & 0 \\ 0 & 0 & \sqrt{\frac{m^*}{m_\perp}} \end{pmatrix}, \vec{k}, \tag{3–58}$$

where $3/m^* = 1/m_\parallel + 2/m_\perp$. It is left as an exercise to the reader to show that

$$E_k = \frac{\hbar^2}{2}\left(\frac{k_1^2}{m_\parallel} + \frac{k_2^2}{m_\perp} + \frac{k_3^2}{m_\perp}\right), \tag{3–59}$$

$$= \frac{\hbar^2 k^{*2}}{2m^*}. \tag{3–60}$$

The matrix in Equation 3–58 is the Herring-Vogt transform:

$$T_{HV} = \begin{pmatrix} \sqrt{\frac{m^*}{m_\parallel}} & 0 & 0 \\ 0 & \sqrt{\frac{m^*}{m_\perp}} & 0 \\ 0 & 0 & \sqrt{\frac{m^*}{m_\perp}} \end{pmatrix}, \tag{3–61}$$

and in fact most of our basic kinematics that we learned in introductory physics class comes back while we're working in Herring-Vogt space:

$$\vec{p}^{\,*} = m^* \vec{v}^{\,*} \tag{3–62}$$

$$= \hbar \vec{k}^{\,*} \tag{3–63}$$

$$\vec{F}^{\,*} = \frac{\mathrm{d}\vec{p}^{\,*}}{\mathrm{d}t}. \tag{3–64}$$

Note, however, that since

$$\hbar \vec{k} = \mathbf{M}\, \vec{v} \tag{3–65}$$

and

$$\vec{v}^{\,*} = \frac{\hbar}{m^*}\, T_{HV}\, \vec{k} \tag{3–66}$$

that

$$\vec{v}^{\,*} = \frac{1}{m^*}\, T_{HV}\, \mathbf{M}^{-1}\, \vec{v} \tag{3–67}$$

$$= T_{HV}^{-1}\, \vec{v}. \tag{3–68}$$

And so it turns out that vectors that are "momentum-like" (momentum, wave vector, force, field vectors) transform via T, and things that are "velocity-like" (velocity, displacement) transform via T^{-1}. This inversion comes from the presence of the mass in the various kinematic equations.

Now, with this framework in place, it is straight-forward to simulate the emission of Luke phonons from electron tracks in our indirect gap semiconductors. We simply perform a coordinate transformation into Herring-Vogt space and follow the same prescription

that we described for hole charge carriers previously. The caveat is that because the Herring-Vogt transformation matrix is non-unitary, back-transforming both the electron's momentum and the emitted phonon's momentum is not guaranteed to conserve momentum. The solution is to maintain the phonon's momentum magnitude from Herring-Vogt space and only use the back-transformed vector for its direction. This is justified by recalling that the energy of the electron is the same in Herring-Vogt space as in real space and that conserving the phonon's momentum magnitude will conserve the energy of the system.

3.6.4 Inter-Valley Scattering

While electrons are mostly constrained to a particular conduction band, some processes can enable the electron to jump from one band to another. Some relevant examples are scattering off of impurities and interacting with high energy phonons. While it is possible to attempt to accurately model the microscopic physics that causes inter-valley scattering, G4CMP's current approach is to use an empirical model developed by the EDELWEISS collaboration [52], which is described in more detail in Section 5.2. Therefore, the rate of inter-valley scattering is calculated as

$$\Gamma = 6.72 \times 10^{-2}\mathrm{s}^{-1}\left[(E_0^2 + |\vec{E}_{HV}|^2) \times \left(\frac{\mathrm{meter}}{\mathrm{volt}}\right)^2\right]^{3.24/2} \tag{3–69}$$

where E_0 is a constant value (with electric field units) that depends on the doping of the germanium and $\vec{E}_{HV}$ is the Herring-Vogt transformed electric field that the electron is experiencing (see Section 3.6.3). Figure 3-3 shows a visualization of a charge carrier simulation in G4CMP.

3.6.5 Comparison to Data

If the previous physics is all implemented correctly, then the charge carriers from an event should reach an equilibrium where the energy gained from the field become balanced by the Luke phonons being shed. One way to check that this balance is achieved correctly is via comparison to experimental data. Members of SuperCDMS have done exactly the

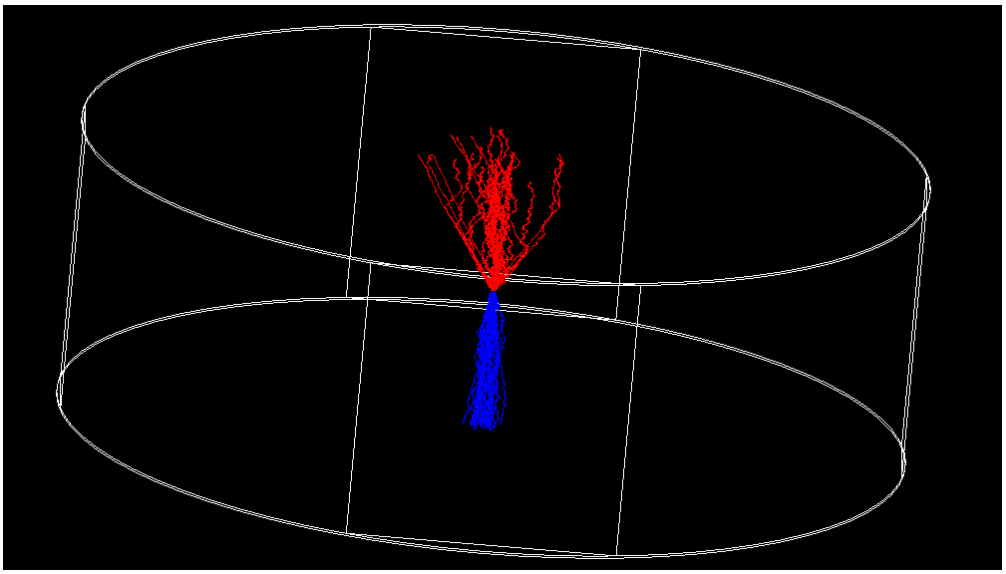


Figure 3-3. G4CMP charge carrier demo. Here we visualize the creation of 30 electron-hole pairs created at the center of the Soudan iZIP detector. Electrons are red; holes are blue. We can clearly see the oblique propagation of the electron as well as the abrupt change in propagation direction caused by occasional inter-valley scattering. The small momentum kicks from Luke phonon emissions causes the tracks to be "wiggly" rather than straight lines.

kind of measurement for this comparison. We can see in Figure 3-4 that our low energy physics models for charge carriers work quite well. However, if we increase the field to a regime that is relevant for the high-voltage mode experiments (Figure 3-5), we begin to see our low energy model diverge from data. This is because these high fields will cause the charge carriers to no longer be constrained to the minimum energy conduction/valence bands. We will need to find a better description for the effective masses of the charge carriers at these higher energies.

For inter-valley scattering specifically, one of our collaborators, Peter Redl, compared G4CMP's inter-valley scattering model with data taken by another one of our colleagues, Robert Moffatt. The result was very good agreement between the simulated and real data as seen in Figures 3-6 and 3-7.

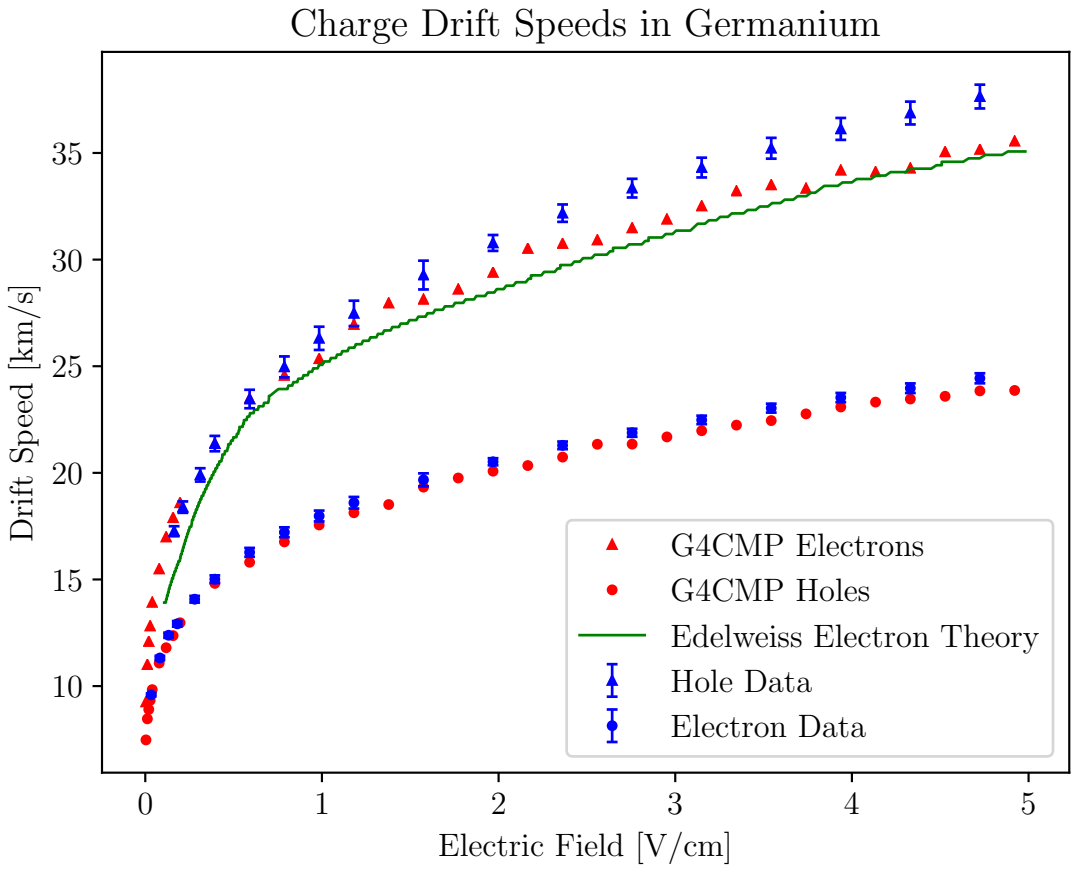


Figure 3-4. Charge drift speeds for weak electric fields. Data from Reference [58]. Theory curve from Reference [59].

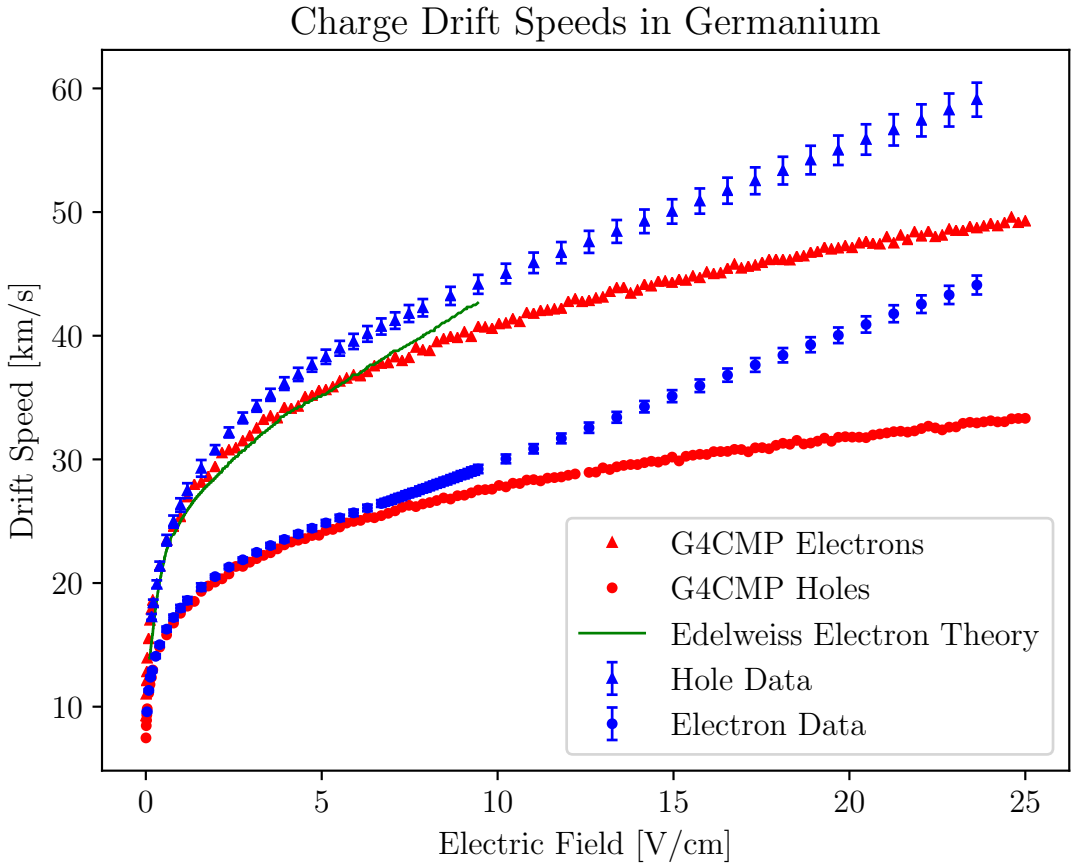


Figure 3-5. Charge drift speeds up to strong fields. Data from Reference [58]. Theory curve from Reference [59].

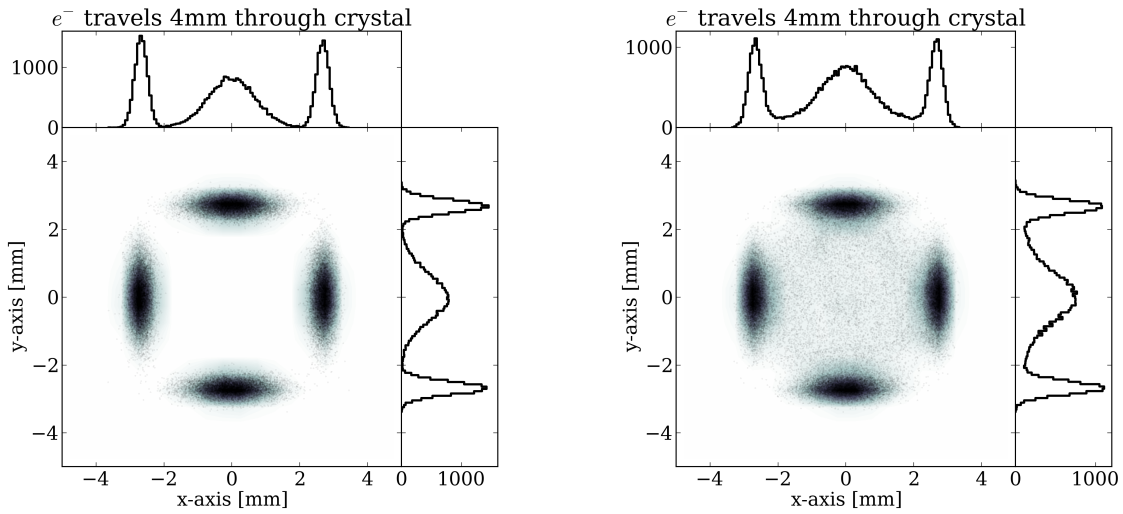


Figure 3-6. Comparison of G4CMP simulation with and without inter-valley scattering physics. Left: G4CMP simulation results with inter-valley scattering disabled. Right: G4CMP simulation results with inter-valley scattering enabled. Without inter-valley scattering, it is (nearly) impossible for electrons to land near (x, y) = (0, 0) [60].

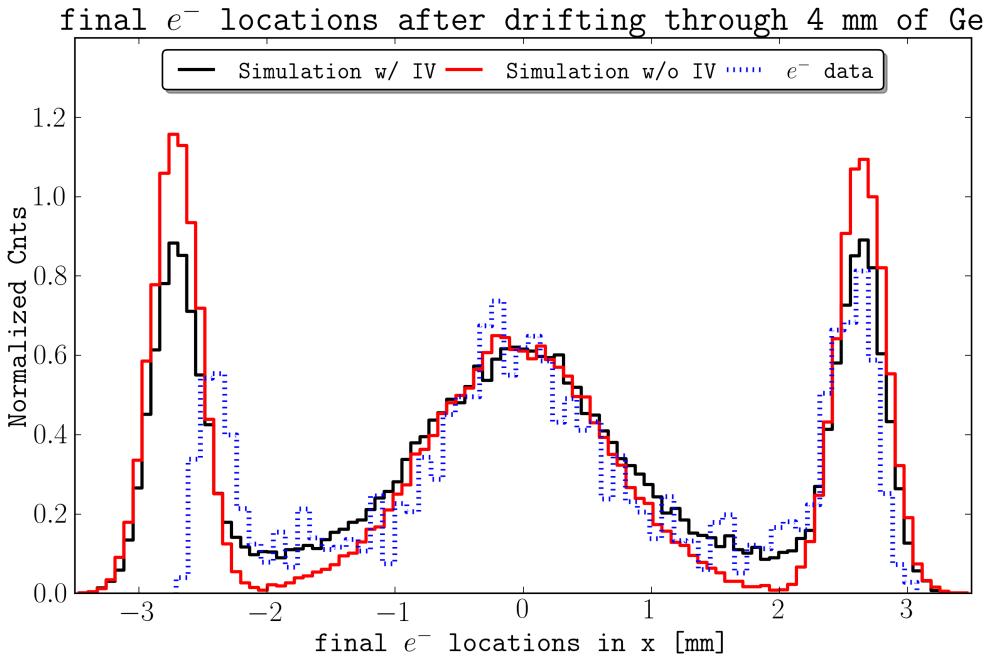


Figure 3-7. Comparison of Peter Redl's simulated data from G4CMP with Robert Moffatt's real experimental charge data. Note that the experiment was not symmetric about x = 0, so Peter chose to fit and normalize the data according to the x > 0 part of the data. We can see that enabling inter-valley scattering decreases the high-radius peaks and increases the middle-radius (x ∼ 2 mm) counts in the right proportions to match the data. [60]

CHAPTER 4
A MONTE CARLO SIMULATION FOR THE SUPERCDMS DETECTORS

4.1 Introduction

Now that we have detailed the implementation of the physics library, we can utilize it in an actual detector Monte Carlo package. This software package, named G4DMC, is something of a stopgap solution. As mentioned in Section 3.1, the long term goal of the detector Monte Carlo group is to add G4CMP to the backgrounds simulation package, SuperSim, to have an end-to-end Monte Carlo chain. One advantage there is that SuperSim already has incredibly detailed and modular models of all generations (including SNOLAB) of CDMS experimental set-ups. With that in mind, we created G4DMC as a minimal working example used to model one SuperCDMS Soudan iZIP. This way we can begin direct comparison to the old MATLAB DMC as well as keep the initial develop of the sensor simulations as expedient as possible.

4.2 Phonon Sensors

4.2.1 Modeling the QETs

Each SuperCDMS detector has eight phonon channels and 458 QETs per channel, wired in parallel. Referring back to Figure 2-8, it would be more precise to describe this as the circuit of each phonon channel, where R_{TES} is actually the resistance of the 458-TES *array*. The actual TES dynamics are complex [23]. However, in the well-behaved regime near T_c the response of an individual TES can be well-described by

$$R_i = \frac{R_N}{2}\left[1 + \tanh\frac{T_i - T_c\left(1 - \frac{|I_i|}{I_c}\right)^{2/3}}{T_w}\right], \tag{4–1}$$

where R_N is the normal resistance, I_c and T_c are the critical current and temperature of the TES, respectively, T_w is the width of the transition curve, and i is the index of the individual TES. However, $I_c = I_c(T_i)$ and $T_w = T_w(I_i)$. Under the typical operating conditions of iZIP detectors, the critical current is typically quite high compared to

the currents seen through the TES (some two orders of magnitude). Therefore, we approximate I_c and T_w as constant.

The dynamics of the full circuit offer another constraint. A standard, albeit somewhat tedious, analysis of the TES part of the circuit from Figure 2-8 reveals

$$\dot{I}_{TES}L = I_{TES}\left(\frac{R_b R_{sh}}{R_b + R_{sh}} - R_{TES} - R_p\right) + \frac{V_b R_{sh}}{R_b + R_{sh}}, \tag{4–2}$$

where V_b is the bias voltage, R_b is the bias resistor, and R_p is a parasitic resistance term from the inductor in Figure 2-8. R_{TES} and I_{TES} are the overall resistance and current of the TES array for the channel, respectively.

The heat change in the TES is related to the difference between the Joule heating power and the substrate (bath) cooling power.

$$C\dot{T} = P_{joule} + P_{sub}, \tag{4–3}$$

where C is the heat capacity of the TES, P_{joule} is the Joule heating power, and P_{sub} is the substrate cooling power.

From Equation 4–3, it is straightforward to describe the equilibrium conditions of an individual TES:

$$P_{joule} + P_{sub} = 0 \tag{4–4}$$

$$I_i^2 R_i - \Sigma\left(T_i^5 - T_{sub}^5\right) = 0, \tag{4–5}$$

where Σ is the electron-phonon coupling constant and T_{sub} is the substrate temperature.

Using Equations 4–1 – 4–5 we can model the response of the phonon channels to an event. By using the measured resistance of a phonon channel, we can set the initial, equilibrium, R_i and use Equations 4–1 and 4–5 to find the initial T_i and I_i. After an event occurs in the detector, phonon energy will begin to be collected by the QETs and funneled into the TESs. As phonon energy is absorbed by a TES, that energy is converted into a temperature change in the TES via the heat capacity and we use Equation 4–1

to calculate a new resistance. As the resistance changes, the circuit will respond to the change in R_{TES} by changing the current. We can then use Equation 4–3 to find the change in temperature in each TES for the next time bin, dt.

This can be made into a naïve integration algorithm as long as we choose an appropriate size for dt. We will discuss choosing dt in Section 4.2.3.

Algorithm 4.1 TES pulse simulation.

```
Set initial R_i
Solve Eq. 4–5 and Eq. 4–1 for I_i and T_i
for all t in timebins do
    for all i in number of TESs do
        T_i = T_i + E_phonon,i / C
        R_i = R_i(I_i, T_i)                          ▷ Eq. 4–1
    end for
    R_TES = 1/∑_i (1/R_i)                            ▷ Parallel resistors
    I_TES = I_TES(I_TES, R_TES)                      ▷ Eq. 4–2
    for all i in number of TESs do
        I_i = I_TES * R_TES/R_i
    end for
end for
```

4.2.2 Heat Diffusion Along the TES

Each TES is much longer in one dimension than the others. As a result, a phonon that is collected by an aluminum fin will tend to heat the TES unevenly. G4DMC accounts for this by implementing each TES as a series chain of resistors along the long dimension and allowing heat to flow from resistor to resistor via the one-dimensional heat diffusion equation:

$$C\frac{\partial T}{\partial t} - k\nabla^2 T = 0, \tag{4–6}$$

where k is the Boltzmann constant. In order to perform a standard numerical integration of Equation 4–6, we need to be sure to choose a dt that satisfies the Courant-Friedrichs-Lewy (CFL) condition [61]:

$$\mathrm{d}t < \frac{C\,(\mathrm{d}x)^2}{2k} \tag{4–7}$$

4.2.3 Choosing dt for TES simulation

As explained in Section 4.2.1, we need to choose a proper dt in order to calculate the time-evolving response of the phonon circuit. Naturally, the upper bound on dt would be the digitization bin size for the real hardware readout. We also discussed the CFL condition in the previous section, which also puts an upper bound on dt. However, there is one more constraint to consider. If the time step is too large, then the substrate cooling power can cause the temperature to be less than the substrate's temperature, which is unphysical. This leads to our last constraint by rearranging Equation 4–3:

$$\mathrm{d}t < \frac{C\left(T_i - T_{sub}\right)}{I_i^2 R_i - \Sigma\left(T_i^5 - T_{sub}^5\right)} \tag{4–8}$$

Of course, this is somewhat frustrating because it means that we must find the minimum dt from a set of 460 (458 TESs + the CFL condition + digitization bin) for each step of the simulation.

Figures 4-1 and 4-2 show the results of the phonon sensor simulation on a 1 keV test event in an iZIP detector. The channels that are closest to the event's location in the detector absorb more heat initially, but very quickly the phonons become uniformly distributed in the detector volume and all of the channels sync as they return to equilibrium.

4.3 Charge Sensors

The Shockley-Ramo theorem can be used to determine the induced current on an electrode by charges moving nearby. It is usually written

$$I_{induced} = q\vec{E}(\vec{x}) \cdot \vec{v}, \tag{4–9}$$

where $\vec{E}$ is a special, unitless, field that is calculated by setting the electrode's potential to unity (no units), grounding any other electrodes, and removing all free charges [62, 63].

This would allow us to calculate the response of each charge channel as time passes in the simulation. However, we can actually simplify the calculation greatly. As mentioned in Section 2.2.5, the charge carriers will arrive at their final positions "instantaneously"

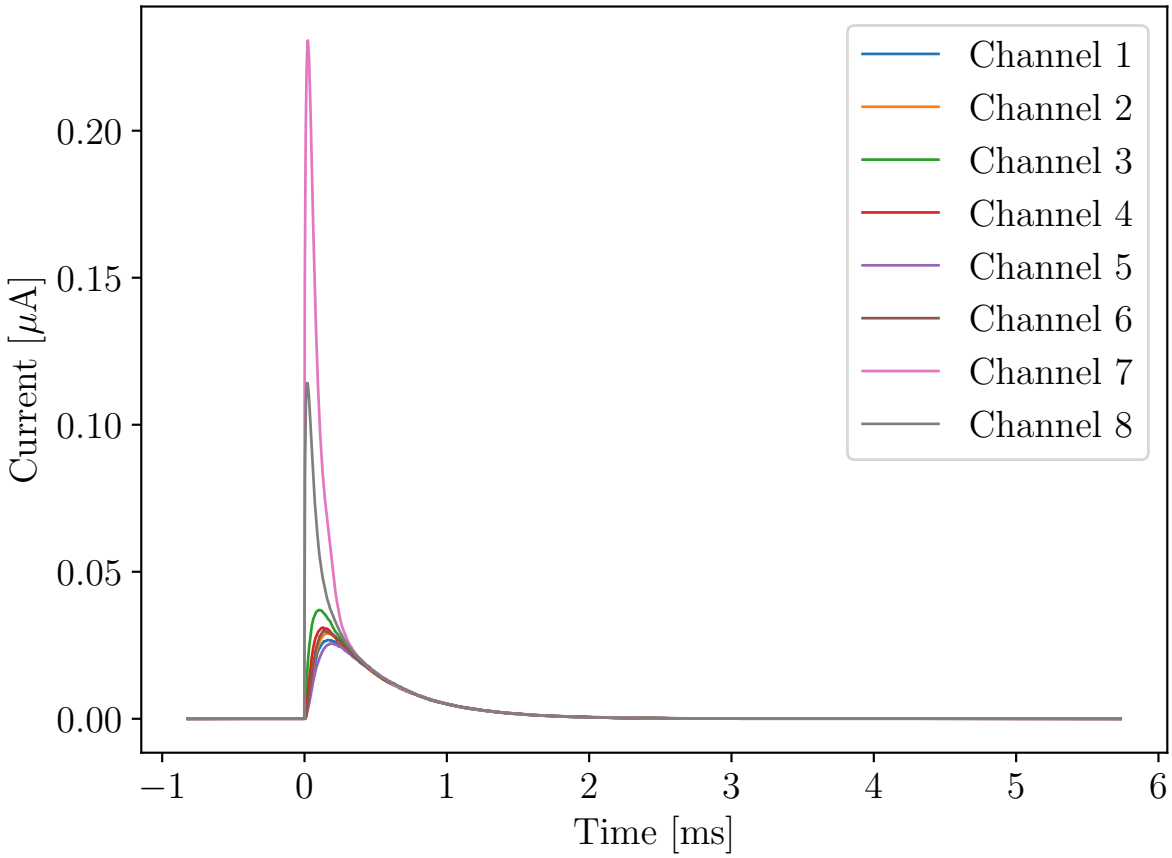


Figure 4-1. Simulated phonon channel response. Response of all eight phonon channels to an electron-recoil at $\vec{x} = (1.0\text{cm}, 1.0\text{cm}, 2.0\text{cm})$. Notice that despite two of the channels absorbing quite a bit more phonon energy than the others, the fall times become synchronized quickly. This is because once the phonons become ballistic, they will reflect from surfaces several more times before being absorbed. Thus, after some time, the remaining phonon energy is randomly distributed in the detector volume, and is absorbed uniformly by the channels. Note that the current is baseline subtracted and inverted for the sake of visualization.

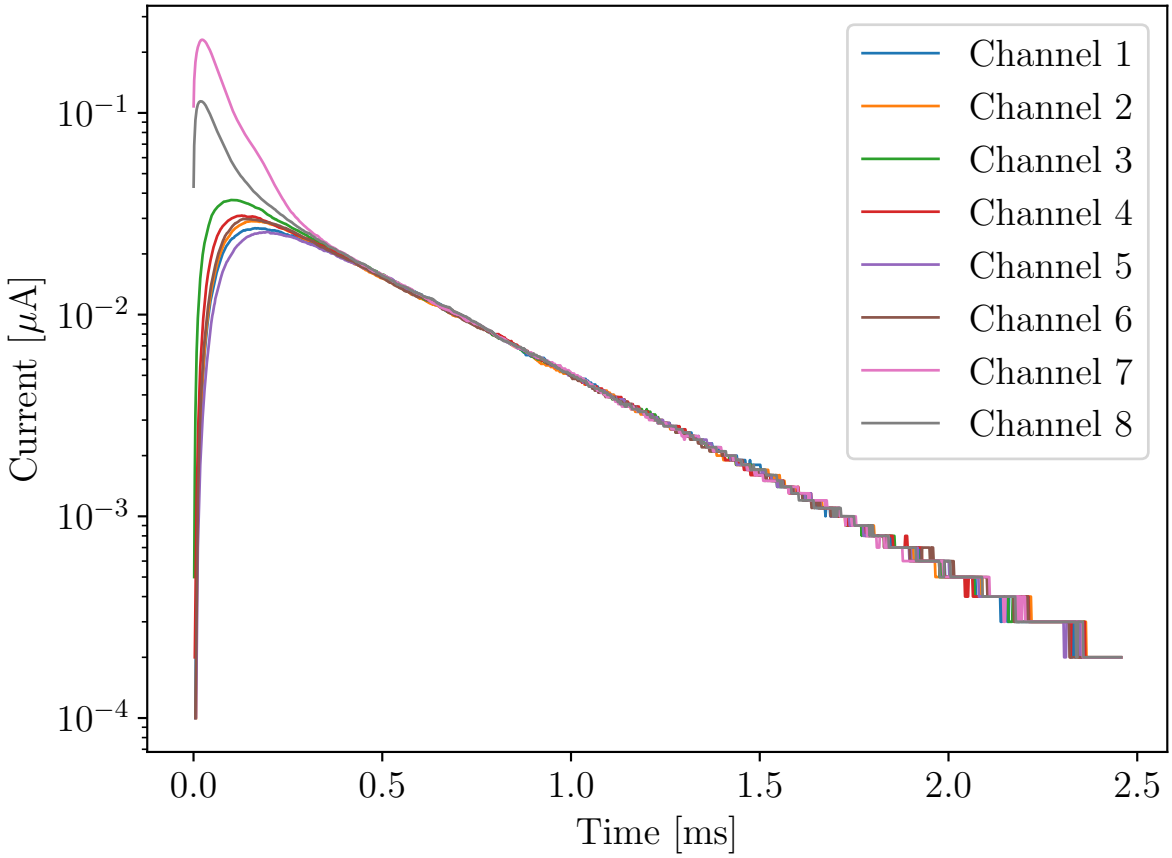


Figure 4-2. Simulated phonon channel response on a log scale. Response of all eight phonon channels to an electron-recoil at $\vec{x} = (1.0\text{cm}, 1.0\text{cm}, 2.0\text{cm})$. This "zoomed-in" view allows us to see the point in time at which the phonons are uniformly distributed in the detector volume, causing all eight channels to behave identically.

relative to the electronic response. So, rather than spend the effort computing the induced current at each step of the simulation, it makes more sense to simply use the start and end information of the charge tracks to determine the FET pulse height. To do that, we can refer to either Shockley's derivation ([62]) or Ramo's derivation ([63]) of Equation 4–9 and simply undo their final step by integrating Equation 4–9 with respect to time, giving us the form we will actually use for the FET simulation:

$$Q_{induced} = -q\phi(\vec{x})\,, \tag{4–10}$$

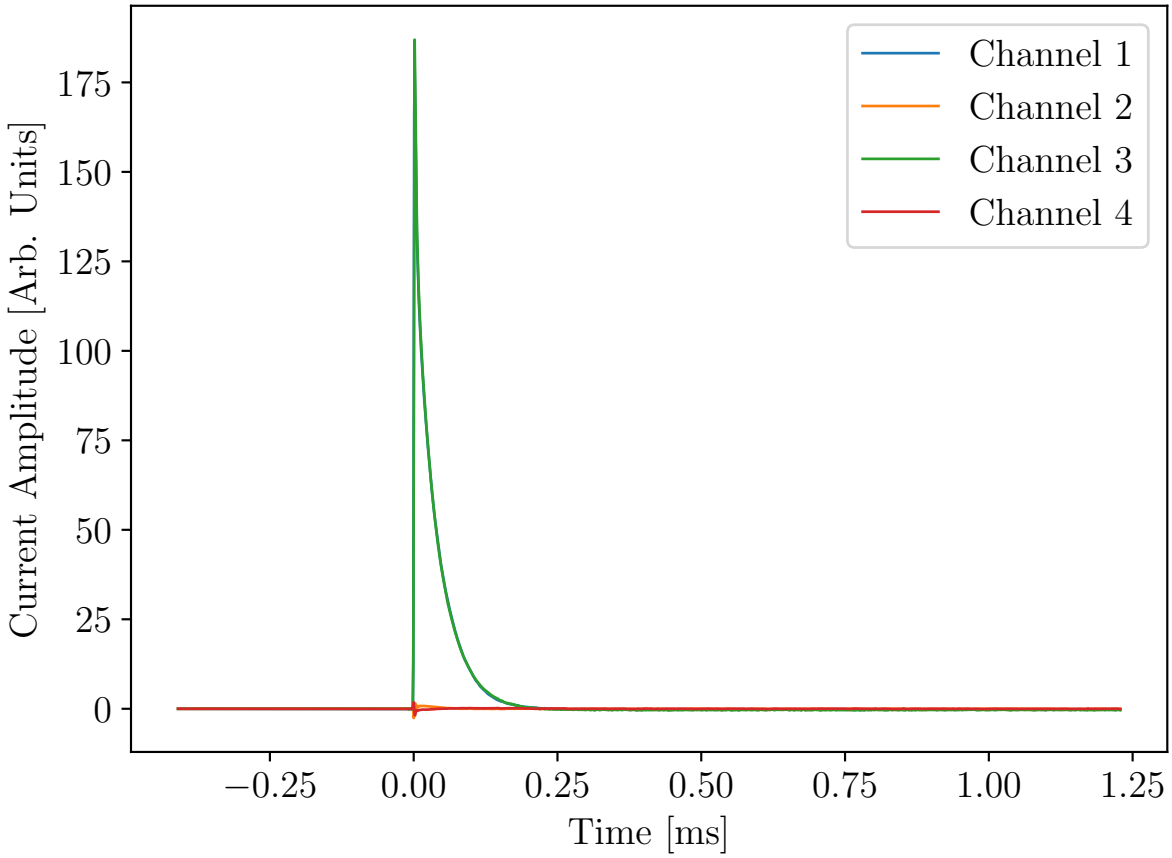


Figure 4-3. Simulated charge channel response. The figure shows a typical output of applying the Shockley-Ramo theorem to the resulting charge carriers collected from an event. This event was near the radial center of the detector, thus the inner charge channels (Channels 1 and 3) collected most of the charge. Both inner channels also collected similar amounts of charge, revealing that the event was not a surface event.

where we refer to $\phi(\vec{x})$ as the Ramo potential. Using Equation 4–10, we can simply calculate $Q_{induced}$ for each charge carrier at its initial position (at t = 0 for the event) and at its final position, and use the difference to scale a template charge pulse's height. This will insure that the relative pulse amplitudes between events are correct, and it is simply a matter of calibrating the absolute height by comparing to real data.

4.4 The Life of an Event

With the physics models outlined in the previous chapter, and the sensor models described in the previous sections of this chapter, we can give an overview of exactly what happens during the simulation of an event in a SuperCDMS detector.

4.4.1 Before the Event

As described in Section 3.3, G4CMP (and therefore G4DMC) requires a configuration file for the crystal to be used in the simulation. G4CMP currently ships default configuration files for germanium and silicon with parameters based either on literature or experimental data. In order to model the intricate electric field of an iZIP, we employ the COMSOL multiphysics software package. This will allow G4DMC to accurately model surface and high-radius events.

4.4.2 Type of Recoil

When a particle deposits energy in a volume which contains a lattice, the first thing that must be done is to decide if the energy deposit came from a nuclear recoil or an electron recoil. The method currently employed in G4CMP is to refer to the particle data group (PDG) ID of the particle that has deposited the energy. If the particle ID is either 2112 (neutron) or larger than 10,000 (stable nuclei will have ID >10,000) we register the hit as a nuclear recoil. This method is not fool-proof and in the future, a more robust look-up table mechanism should be developed. In particular, there are many exotic particles with PDG codes >10,000 and neutrinos have PDG IDs between 10–20. However, for the backgrounds we expect for the SuperCDMS experiment, this simple approach works well.

4.4.3 Energy Partitioning

Once the type of recoil is determined, G4DMC, via G4CMP's *G4CMPEnergyPartition* class, will partition the recoil energy into phonon energy and charge energy.

Algorithm 4.2 Energy partition logic.

if recoil type == nuclear **then**
  $E_{phonon} = Y(E_{recoil}) * E_{recoil}$ ▷ Eq. 2–12
  $E_{charge} = E_{recoil} - E_{phonon}$
else
  $E_{phonon} = 0$
  $E_{charge} = E_{recoil}$
end if

4.4.4 Applying the Fano Factor

After the energy partitioning step, we know exactly how much energy should be used to create phonons at the interaction vertex. However, the charge energy is not simply the exact remainder of the total recoil energy.

The so-called Fano factor is an intrinsic property of a material that determines the dispersion of the distribution of ionized atoms. In other words, the same amount of deposited energy in a detector will lead to differing numbers of electron-hole pairs being created (in the case of a solid state detector such as the SuperCDMS iZIP).

Naïvely, one may expect that the distribution of charge pairs should follow Poisson statistics. However, experiments show that to not be the case [51]. In fact, the fluctuation tends to be less than a purely Poisson process would be (F = 1 would correspond to a Poisson distribution), meaning that the individual ionizations are not statistically independent.

In G4CMP, the Fano factor is used to calculate the resolution of the measurement,

$$\sigma_E = \sqrt{E_{recoil}\ F\ \Delta_E} \tag{4–11}$$

where F is the Fano factor (0.1057 for Ge), and Δ_E is the average energy per charge pair (2.96 eV for Ge). So, G4DMC adjusts the recoil charge energy by sampling from a Gaussian distribution with the mean set to the true recoil charge energy and the width set by the Fano factor.

4.4.5 Creating the Initial Particle Tracks

Now that we have the total energy for phonons and the total energy for charge carriers, we are able to spawn the initial populations. Many details of the initial populations are not well understood. However, the phonon downconversion rate and a reasonable assumption of directional isotropy makes generating the initial particles straight-forward.

For the phonon population, the initial energy distribution is not known. However, recall from Equation 3–28 that the isotope scattering rate is proportional to the frequency to the fourth power. So, the mean free path to a decay, for a given frequency is

$$\lambda = \frac{v_s}{\Gamma} \tag{4–12}$$

where v_s is the speed of sound in the crystal. If we plug in the values for germanium, we get

$$\lambda \approx 5 \times 10^{44} m/s^4 \ \frac{1}{\nu^4} \tag{4–13}$$

For a high energy phonon near the Debye frequency ($\sim 2THz$ in Ge), the mean free path is on the order of 30 microns. Since the iZIP's smallest dimension is 25.4 mm (33.3 mm for SNOLAB) and events are rejected if they are within $2mm$ of a surface, it is arguably safe for us to set the initial population of phonon modes to be equal to the population we expect at the end of the simulation (See Section 3.5.4 and Reference [50]). Likewise, it should also be safe for the directional distribution to be initially isotropic and for the energy distribution to be flat.

So the initial phonon population is isotropic, monoenergetic, with a mode distribution equal to the final expected distribution.

For charge carriers, we don't need to perform quite as much approximation and hand-waving. One reason is that there are fewer variables (no modes). The other reason is that the energy of the initial particles is well-understood.

For input energies above $\sim 10eV$, the energy per freed charge pair is constant for a given crystal [18].

It is somewhat more difficult to argue that charge carriers should be created with an isotropic velocity distribution.In the first place, the ionization cascade happens first with one electron being freed from a valence band; it then travels a short distance before recoiling with another valence electron, freeing that one, which frees others, etc. Overall, there should be a total momentum in the same direction as the initial freed electron.

Also note that charges will be dragged into particular directions by the applied electric field in the detector volume. So, unlike the phonons, the charge carriers will not end up isotropic or uniformly spread throughout the volume.

In the most naïve and pessimistic calculation, we assume that there is no Luke phonon emission and that a particular charge carrier is supposed to be directed exactly parallel to the force from the field, but has mistakenly been directed exactly antiparallel to it. Basic kinematics tells us that the distance before turning around will be,

$$d_{turn} = \frac{E_{k0}}{|\vec{E}|q} \tag{4–14}$$

where E_{k0} is the initial kinetic energy and $\vec{E}$ is the applied electric field. Using typical values for a hole and a standard SuperCDMS iZIP field, the distance comes out to around 7 mm. That number is large enough to be of concern, since a Soudan iZIP will only reject events within 2 mm of the surface. However, if we consider Luke phonon emission, it will bring that number down substantially.

Recall from Equation 3–57 that the highest order term on the rate is

$$\Gamma \approx \frac{1}{3}\frac{v_s}{l_0}\left(\frac{k}{k_s}\right)^2 \tag{4–15}$$

so an approximate mean free path to a scatter would be

$$\lambda \approx \frac{v_0}{\Gamma} \tag{4–16}$$

$$\approx 3l_0 \frac{v_0}{v_s} \left(\frac{k_s}{k_0}\right)^2 \tag{4–17}$$

$$\approx 3l_0 \frac{k_s}{k_0} \tag{4–18}$$

and in the worst case ($k_0 \sim k_s$) would be, in germanium, 771 μm for electrons and 324 μm for holes. However, the initial kinetic energy of a freed electron is around 1.11 eV for germanium, which is roughly a factor of 10,000 larger than $E(k_s)$. Do to the stopping power of germanium at that energy, it will only be able to travel a few microns before settling down below k_s, at which point the electric field and the Luke phonon emission dictate the dynamics (See Reference [64] for more detail on the stopping power of germanium). Thus, it would appear that we are safe to assume the charge carriers begin isotropically, since their kinematics will become determined by the applied field and Luke phonon physics after very short distances.

4.4.6 Phonons Traveling in the Bulk

As described in Section 3.5, there are two main processes that low energy acoustic phonons will experience while traveling through the bulk in a CDMS detector: anharmonic decay and isotope scattering. Thus, most phonon tracks that are created wont actually make it to being absorbed– rather, some of their descendants will. Since the details of these processes are covered in an earlier section and the implementation is a simple translation of the physical model into computer code, I will save the wear on my keyboard.

4.4.7 Charges Traveling in the Bulk

Charges are a little more complex than the phonons in G4CMP. In particular, because our model for Luke phonon emission calculates an instantaneous time rate of emission. That rate depends on the wave vector of the charge carrier. Obtaining a closed-form expression for the actual mean free path is difficult, so we take an approximation approach

whereby we limit the maximum simulation step size such that the rate calculated by our model is approximately constant over a time step. That allows the charge carrier to release the correct amount of Luke phonon energy over the course of the simulation.

4.4.8 Phonons at the Surface

For the iZIP, we have different behavior if the phonon is arriving at a face or at the cylindrical sidewall. Most importantly, since the sidewall is not instrumented, the phonon is not likely to be absorbed. Thus, all phonons that hit a sidewall are reflected. If the phonon arrives at an instrumented surface, there is a probability to be absorbed. That probability is equal to the total aluminum coverage of the face times the average probability of being absorbed upon hitting a QET's aluminum fin, as determined by calibration against data. Note that a more accurate method would be to actually define the position, size, and orientation of each of the $458 * 8$ QETs on the detector and only absorb if the phonon arrives within one of those areas. However, implementing the pattern is nontrivial and it is not likely that such specific location information would ever cause a noticeable effect in the phonon pulses.

When a phonon is absorbed into a QET, we simply pass the energy to G4CMP's *KaplanQP* helper functions (described in Section 3.5.5) which perform the Monte Carlo calculations to determine a final number of phonons to reintroduce into the crystal, with their corresponding energies, as well as a total amount of energy to add to the TES.

4.4.9 Charges at the Surface

Charge interactions at the surface are mostly quite simple. Surfaces are defined with probabilities for absorption, and reflection (and transmission, but that is not used for G4DMC). When a charge arrives at the surface, we throw a random number to determine if it should be absorbed or reflected.

If it is absorbed, its kinetic energy is converted into phonons, which are radiated back into the crystal. This approximates the charge carrier rattling around near the minimum of the potential well until it eventually recombines with the valence band. Then, the

charge carrier releases half of the semiconductor band gap energy as phonons. Releasing half of the gap energy is not entirely physical- rather, an electron should find a hole and the two may interact, releasing the gap energy back to the crystal. However, since tracks cannot directly interact with each other in Geant4, the closest approximation is to have each charge carrier do half of the job, so that the total energy is correct. Once all of the kinetic and potential energy is converted to phonons, the charge carrier track information is recorded as a "hit" and the track is killed.

If a hole is reflected, we perform a simple specular reflection of the momentum direction. Electrons, however, are a little different. Since the momentum direction is not aligned with the velocity, simply reflecting the momentum may leave the velocity still pointing outward from the crystal. So, for electrons we make the adjustment that we reflect the velocity vector and then calculate a new momentum from that.

CHAPTER 5
FUTURE WORK

SuperCDMS does have an old detector Monte Carlo simulation that dates back to the CDMS-II style detectors, which had a very different design than the SuperCDMS iZIP detectors- namely that the phonon signal was read out on one face, and the ionization signal on the other. That code was "retrofitted" to accommodate the newer interdigitated readout design of the iZIPs, but it was not easy to do. Likewise, the old Monte Carlo code made many assumptions and hardcoded many properties of the physics it was used to model, making it often very difficult to modify or test new models.

It was the explicit goal of G4CMP and G4DMC to be designed as modularly as possible, making it possible to interchange and enable/disable physics processes, to model many detectors at once, and to be fast while doing so. The modular nature makes many of the studies outlined in this chapter much more feasible than they would have been with the MATLAB-based simulation code.

5.1 Study TES Saturation Effects

Take, for example, the results published by the SuperCDMS Collaboration in 2014[43]. In the analysis, three of the 11 candidate events were from the same detector (out of 15 detectors). It turns out that this was a known-faulty detector in that the outer charge channel was shorted on one side. The analysis took this into account by having a severe radial cut on the fiducial volume. Obviously, this cut was not severe enough. Kristi Schneck did a detailed study of this detector using the old detector Monte Carlo in her dissertation[65].

In her work with the old DMC Kristi bumped into a deficiency in the model that the MATLAB DMC used to simulate the TES response to phonons- that it does not correctly model TES saturation. When a TES absorbs enough energy to go normal, it can continue to absorb energy and increase its temperature. Then the TES will take longer to cool back to superconducting temperature than it would have if it had only barely become

normal conducting. The MATLAB DMC did not appear to model this behavior, as seen in Figure 5-1. What was done to adjust for the lack of proper saturation modeling was to use a technique developed by Peter Redl to post-process the phonon data, that he called "bulldozing," where phonon energy that came into a TES after it was already saturated was pushed forward in time until the it is absorbed at a time in which the sensor was not already saturated. The visualization he used gave rise to the name "bulldozing" 5-2.

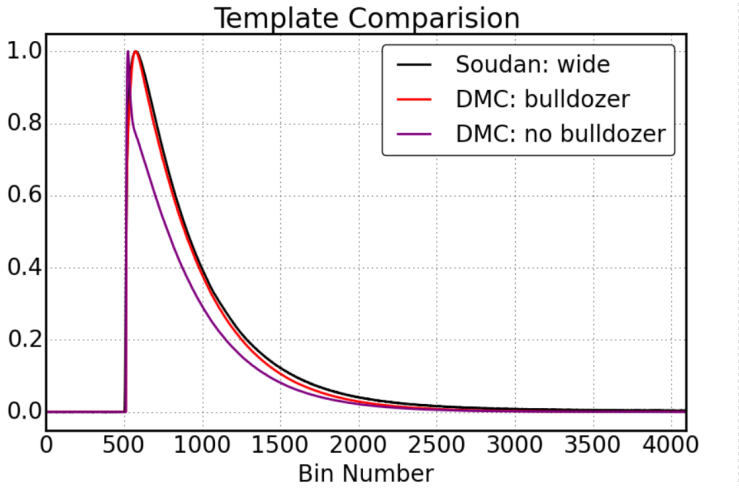


Figure 5-1. Template pulses from phonon pulses, scaled to maximum height of 1.0. We can see the result of poor accounting of local saturation in the non-bulldozed DMC pulse. After bulldozing, the DMC pulses have a shape much closer to the data. [65]

Because the new G4DMC does carefully model individual TESs and allows for simple saturation (with constant heat capacitance) of individual sensors, it would be interesting to see a re-analysis of this detector within the new framework to see if a similar result can be achieved without as much by-hand adjustment.

A very coarse test was conducted with the current G4DMC code to see if saturation effects are visible under certain circumstances. The effect can be seen in Figure 5-3.

5.2 Inter-Valley Scattering Model

The current model for inter-valley scattering of electrons comes from empirical, macroscopic data, from studies done by the Edelweiss collaboration, who have detectors quite similar to the SuperCDMS iZIPs. The experiment was carried out by having a thin

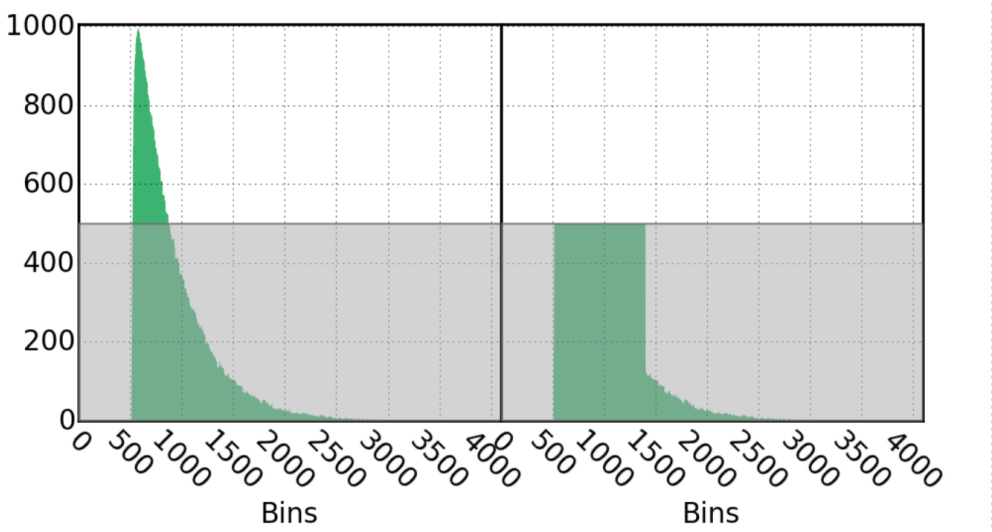


Figure 5-2. A graphic depiction of how "bulldozing" the phonon data works. The phonon energy that would bring a TES above the saturation threshold is pushed into the nearest future time bin that is still below threshold. Taken from Kristi Schneck's dissertation[65]. Original credit to Peter Redl.

(20 mm) germanium device very similar to the SuperCDMS iZIP. An adjustable electric field is applied across the detector and an Americium-241 gamma source is placed near the hole-carrier side.

The resulting gammas will create electron-hole pairs, where the electrons will drift across the full length of the detector. For very low field strength, the probability of inter-valley scatters within that path length is low. Thus, almost all of the charge signal is collected in the high-radius sensors. However, as the field strength is increased, more and more of the electrons will experience inter-valley scatters and the relative population of charges absorbed in inner charge sensors increases.

At the end they found that the rate of inter-valley scattering is

$$\Gamma = 6.72 \times 10^{-2}\mathrm{s}^{-1}\left[(E_0^2 + |\vec{E}_{HV}|^2) \times \left(\frac{\mathrm{meter}}{\mathrm{volt}}\right)^2\right]^{3.24/2}, \tag{5–1}$$

as already mentioned in Section 3.6.4.

This is the model that is implemented in G4CMP currently. However, since this is a macroscopic model of the bulk rate of inter-valley scattering, no attempt is made

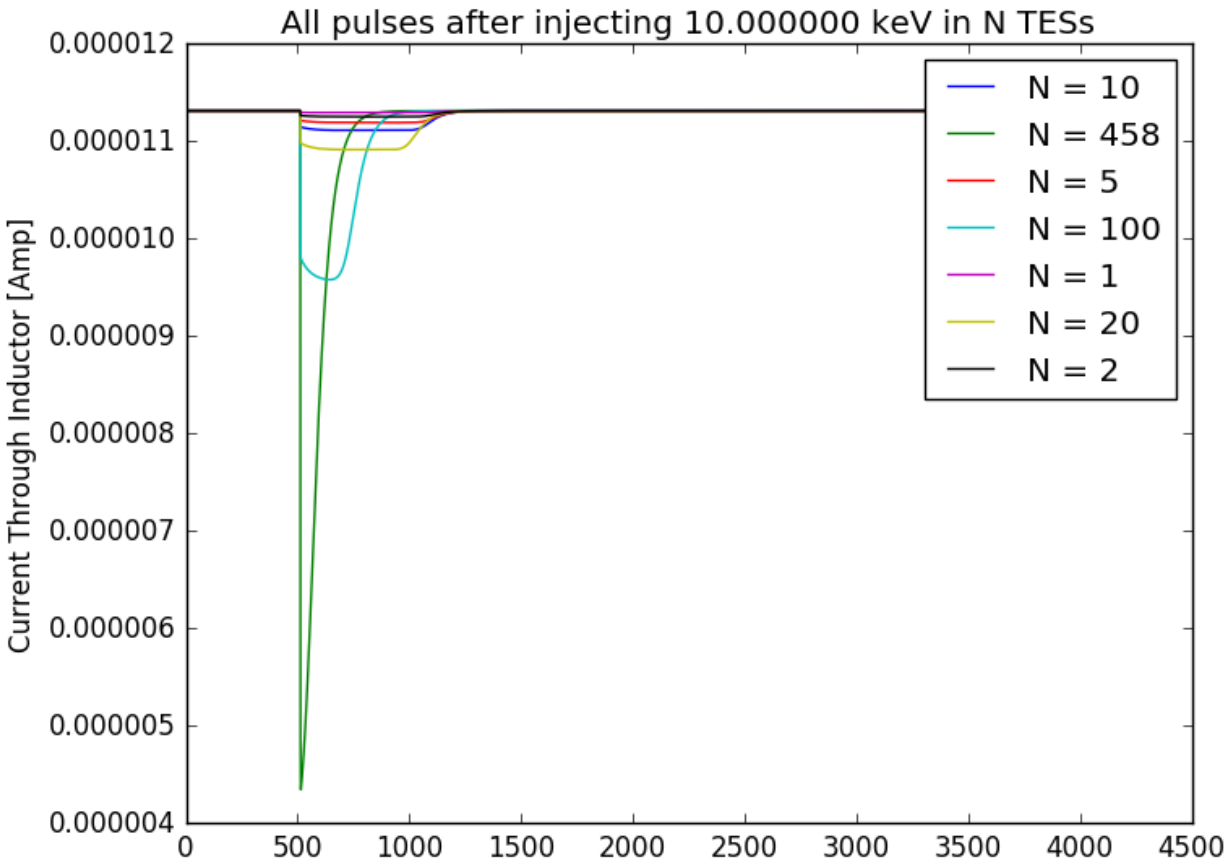


Figure 5-3. Testing G4DMC's TES model for saturation effects. This crude test was to simply inject 10 keV of phonon energy into N QETs of one channel. Each channel has 458 QETs total. As we can see from the response, some of the TESs are clearly becoming saturated when the energy is concentrated into relatively few of the QETs, taking much longer to transition back to the superconducting phase.

to correctly model the underlying microscopic physics. This is obviously a problem for a simulation that handles each electron independently. In particular, when a track in G4CMP undergoes inter-valley scattering, it simply changes the mass tensor for the track, taking no care to conserve momentum or energy. While that would seem to be quite egregious, it turns out to not have terrible consequences in practice.

There is the opportunity to improve the inter-valley scattering model used by G4CMP. By using the microscopic scattering model defined in Reference [66], Equation 5,

$$\Gamma_{IV} = \frac{3\sqrt{m_{\parallel} m_{\perp}^2}\,\Xi^2}{\sqrt{2}\pi\hbar^2\rho\hbar\omega}\sqrt{E-\hbar\omega}\sqrt{1+\alpha\left(E-\hbar\omega\right)}\left[1+2\alpha\left(E-\hbar\omega\right)\right] \tag{5–2}$$

where $m_{\parallel}$, $m_{\perp}$ are the same masses defined in Section 3.6.2, and Ξ is the deformation potential encountered in Section 3.6.3.

Armed with that information it should be straight-forward to implement a new inter-valley scattering process and compare its results to the current model.

5.3 Improve Initial Particle Distribution Models

One of the challenges with creating G4CMP is that some details of the microscopic physics are not well known. The following is a list of what distribution is not known to the author and what the current implementation does as an approximation.

5.3.1 Recoil Charges

These are the initial distribution of charge carriers from an energy deposition in the crystal. What actually occurs is that few electrons are freed from the valence shell of the lattice site that is impacted. Those few electrons have a large amount of energy and thus rapidly scatter off of nearby valence electrons, given them a large amount of energy as well. The cascade continues until all of the freed charges are kinematically unable to free more electrons.

As discussed in Section 4.4.5, G4CMP assumes that the charge carriers will very quickly slow toward the speed of sound of the crystal and continue on being accelerated by the applied electric field. This could potentially be an issue if we want to conduct studies in crystals with no electric field applied, or if the initial momentum direction actually does have significant effects in some regions of the detector.

5.3.2 Recombination and Charge-Absorption Phonons

As discussed in Section 4.4.5, phonons are created when a charge carrier is absorbed at an electrode. The phonons that are created from the kinetic energy and the

semiconductor gap energy are not well-understood. Currently, G4CMP uses the same approximation used for initial recoil phonons: uniform spatial distribution, Debye energy, equilibrium mode distribution.

5.4 Lookup Table Approach to Luke Phonons

The current implementation of Luke phonon emission from charge carriers in G4CMP has several down sides (See Section 3.6.3):

1. Multiple processes for one effect
2. Wasted computation updating the step when nothing happens
3. Ansatz-style maximum step function

The difficulty arises because the rate of Luke phonon emission depends on the instantaneous wave vector magnitude, which we repeat here from Equation 3–57.

$$\Gamma = \frac{v_s}{3l_0}\left(\frac{k}{k_s}\right)^2\left(1-\frac{k_s}{k}\right)^3 \qquad \text{(3–57 repeated)}$$

Where, under the influence of an (constant, uniform) electric field, $k = k(t) = \vec{k}(t)\cdot\vec{k}(t)$,

$$k\,(t) = \sqrt{|\vec{k}(0)|^2 + 2\frac{q}{\hbar}\left(\vec{k}(0)\cdot\vec{E}\right)t + \frac{q^2}{\hbar^2}|\vec{E}|^2\,t^2} \qquad (5\text{–}3)$$

So, we would ideally like to integrate the time-evolving emission rate, create a probability density function, and sample from that to determine the actual time at which a Luke phonon would be emitted. To create the PDF, we can approximate the process as a non-stationary Poisson process. Luke phonon emission is not a true Poisson process because the scatters are not independent. However, it is common to approximate processes as Poisson-like when we are only interested in the first event and it is unlikely to see two events in the same time window. The "zeroth" event is independent from the first event. It is non-stationary because the rate is changing. It turns out that the probability of n events occurring in a time window, t for a non-stationary Poisson process is

$$P(n,t) = \frac{\Lambda(t)^n}{n!}\mathrm{e}^{-\Lambda(t)}, \qquad (5\text{–}4)$$

where

$$\Lambda(t) = \int_0^t \lambda(\tau)\,\mathrm{d}\tau \tag{5–5}$$

is referred to as the intensity of the Poisson process, and $\lambda(t)$ is the time-evolving rate. This means that, for the probability of one Luke phonon emission occurring between time 0 and t,

$$P(1,t) = \left(\int_0^t \Gamma(\tau)\,\mathrm{d}\tau\right) \mathrm{e}^{-\int_0^t \Gamma(\tau)\,\mathrm{d}\tau}, \tag{5–6}$$

where $\Gamma(t)$ is from allowing Equation 3–57 to vary in time. Once we have the above equation, we can obtain a mean free path via the standard approach,

$$\lambda_{mfp} = \frac{\int_0^\infty |\vec{x}(\tau)| P(1,\tau)\,\mathrm{d}\tau}{\int_0^\infty P(1,\tau)\,\mathrm{d}\tau}. \tag{5–7}$$

This is obviously a huge amount of computation to do during a Monte Carlo simulation of on the order of 10,000 charge carrier tracks, each emitting some 1,000 Luke phonons as they traverse the SuperCDMS detector. The plan is to do these calculations up front and store the results in a lookup table as is done for the phonon velocity equations. In this case, the lookup table is three dimensional, with the variables being $|\vec{k}|$, $|\vec{E}|$, and θ, where θ is the angle between $\vec{k}$ and $\vec{E}$.

Work is already underway on incorporating this method into G4CMP and the scripts needed to generate the lookup tables are included in the code repository.

5.5 Downsampling Tracks

The performance profile of G4CMP-based simulations is not where we would like it to be in order to simulate a useful amount of data for rare-event experiments, like SuperCDMS. Geant4 is fundamentally well-designed to be fast. However, G4CMP requires that each track hold more data than the usual Geant4 simulation's track as well as perform more calculations per physics process than many other physics models. Neither of these should make such a large impact on runtime as the sheer number of tracks that a recoil event creates in a G4DMC event. A 1 keV electron-recoil will make on the order of a

few million phonon tracks, and around 350 electron-hole pairs. Such an event takes about one hour to run on our current workstations. This is especially depressing when we recall that a 1 keV event is so low energy that it's below the threshold for SuperCDMS Soudan detectors.

Geant4 has a track biasing feature wherein we can set a track's weight inside our processes. We should then be able to only produce tracks with some probability, P, and give the created tracks a weight of W/P where W is the parent track's weight. We have to be very careful about how this is applied, however, lest we lose too many phonon tracks. There are many "generations" of phonons produced after an initial recoil, so we cannot kill them off too quickly. Likewise, the timing information is very import in the TES response- if we bin up the incoming energy too coarsely, our phonon pulse signal will be distorted. One more concern is that process that create phonons with some complicated energy distribution are prone to lose information, or reconstruct energy incorrectly, if we do a blind downsampling on them. Work is currently ongoing to understand these effects of applying Geant4's track biasing mechanics to G4CMP processes.

REFERENCES

[1] V. Springel, C. S. Frenk, and S. D. M. White, Nature **440**, 1137 (2006), arXiv:astro-ph/0604561 [astro-ph] .

[2] F. Zwicky, Helv. Phys. Acta **6**, 110 (1933), [Gen. Rel. Grav.41,207(2009)].

[3] Y. Sofue and V. Rubin, Ann. Rev. Astron. Astrophys. **39**, 137 (2001), arXiv:astro-ph/0010594 [astro-ph] .

[4] J. F. Navarro, C. S. Frenk, and S. D. M. White, Astrophys. J. **462**, 563 (1996), arXiv:astro-ph/9508025 [astro-ph] .

[5] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, Astrophys. J. **648**, L109 (2006), arXiv:astro-ph/0608407 [astro-ph] .

[6] M. Tegmark *et al.* (SDSS), The Astrophysical Journal **606**, 702 (2004).

[7] P. A. R. Ade *et al.* (Planck), Astron. Astrophys. **594**, A13 (2016).

[8] G. Steigman, B. Dasgupta, and J. F. Beacom, Phys. Rev. **D86**, 023506 (2012), arXiv:1204.3622 [hep-ph] .

[9] T. Saab, in *The Dark Secrets of the Terascale: Proceedings, TASI 2011, Boulder, Colorado, USA, Jun 6 - Jul 11, 2011* (2013) pp. 711–738, arXiv:1203.2566 [physics.ins-det] .

[10] F. Donato, N. Fornengo, and S. Scopel, Astropart. Phys. **9**, 247 (1998), arXiv:hep-ph/9803295 [hep-ph] .

[11] S. Ahlen *et al.*, Int. J. Mod. Phys. **A25**, 1 (2010), arXiv:0911.0323 [astro-ph.CO] .

[12] A. M. Green, Mod. Phys. Lett. **A27**, 1230004 (2012), arXiv:1112.0524 [astro-ph.CO] .

[13] Y.-Z. Chen, Y.-A. Luo, L. Li, H. Shen, and X.-Q. Li, Commun. Theor. Phys. **55**, 1059 (2011), arXiv:1101.3049 [hep-ph] .

[14] R. Basu Thakur, University of Illinois at Urbana-Champaign (2015), PhD Dissertation.

[15] D. Mei and A. Hime, Phys. Rev. **D73**, 053004 (2006), arXiv:astro-ph/0512125 [astro-ph] .

[16] B. Neganov and V. Trofimov, USSR Patent No. 1037771 (1985).

[17] P. N. Luke, Journal of Applied Physics **64**, 6858 (1988), http://dx.doi.org/10.1063/1.341976 .

[18] V. S. Vavilov, Soviet Physics Uspekhi **4**, 761 (1962).

[19] J. Lindhard, V. Nielsen, M. Scharff, and P. V. Thomsen, Danske Videnskab. Selskab. Mat. Fys. Medd. **33** (1963).

[20] J. D. Lewin and P. F. Smith, Astropart. Phys. **6**, 87 (1996).

[21] B. Jamtveit and P. Meakin, *Growth, Dissolution, and Pattern Formation in Geosystems* (Springer Netherlands, 1999).

[22] R. Agnese *et al.* (SuperCDMS), Appl. Phys. Lett. **103**, 164105 (2013), arXiv:1305.2405 [physics.ins-det] .

[23] M. C. Pyle, Stanford University (2012), PhD Dissertation.

[24] J. P. Filippini, University of California, Berkeley (2008), PhD Dissertation.

[25] J. A. Burney, Stanford University (2007), PhD Dissertation.

[26] M. Pyle, D. A. Bauer, B. Cabrera, J. Hall, R. W. Schnee, R. B. Thakur, and S. Yellin, J. Low. Temp. Phys. **167**, 1081 (2012), arXiv:1201.3685 [astro-ph.IM] .

[27] R. Agnese *et al.* (SuperCDMS), Phys. Rev. Lett. **112**, 041302 (2014), arXiv:1309.3259 [physics.ins-det] .

[28] R. Agnese *et al.* (SuperCDMS), Phys. Rev. Lett. **116**, 071301 (2016), arXiv:1509.02448 [astro-ph.CO] .

[29] R. Agnese *et al.* (SuperCDMS), Phys. Rev. D (2016), 10.1103/PhysRevD.95.082002, [Phys. Rev.D95,082002(2017)], arXiv:1610.00006 [physics.ins-det] .

[30] R. Agnese *et al.* (SuperCDMS), Phys. Rev. **D92**, 072003 (2015), arXiv:1504.05871 [hep-ex] .

[31] W. Zhao *et al.* (CDEX), Phys. Rev. **D93**, 092003 (2016), arXiv:1601.04581 [hep-ex] .

[32] C. Savage, G. Gelmini, P. Gondolo, and K. Freese, JCAP **0904**, 010 (2009), arXiv:0808.3607 [astro-ph] .

[33] C. Amole *et al.* (PICO), (2017), arXiv:1702.07666 [astro-ph.CO] .

[34] A. Tan *et al.* (PandaX-II), Phys. Rev. Lett. **117**, 121303 (2016), arXiv:1607.07400 [hep-ex] .

[35] D. S. Akerib *et al.* (LUX), Phys. Rev. Lett. **118**, 021303 (2017), arXiv:1608.07648 [astro-ph.CO] .

[36] E. Aprile *et al.* (XENON100), Phys. Rev. **D94**, 122001 (2016), arXiv:1609.06154 [astro-ph.CO] .

[37] R. Agnese *et al.* (CDMS), Phys. Rev. Lett. **111**, 251301 (2013), arXiv:1304.4279 [hep-ex] .

[38] P. Agnes *et al.* (DarkSide), Phys. Rev. **D93**, 081101 (2016), [Addendum: Phys. Rev.D95,no.6,069901(2017)], arXiv:1510.00702 [astro-ph.CO] .

[39] G. Angloher *et al.* (CRESST), Eur. Phys. J. **C76**, 25 (2016), arXiv:1509.01515 [astro-ph.CO] .

[40] G. Angloher *et al.*, Eur. Phys. J. **C72**, 1971 (2012), arXiv:1109.0702 [astro-ph.CO] .

[41] E. Aprile *et al.* (XENON), JCAP **1604**, 027 (2016), arXiv:1512.07501 [physics.ins-det] .

[42] F. Ruppin, J. Billard, E. Figueroa-Feliciano, and L. Strigari, Phys. Rev. **D90**, 083510 (2014), arXiv:1408.3581 [hep-ph] .

[43] R. Agnese *et al.* (SuperCDMS), Phys. Rev. Lett. **112**, 241302 (2014), arXiv:1402.7137 [hep-ex] .

[44] S. Agostinelli *et al.* (GEANT4), Nucl. Instrum. Meth. **A506**, 250 (2003).

[45] J. Allison *et al.*, IEEE Trans. Nucl. Sci. **53**, 270 (2006).

[46] J. Allison *et al.*, Nucl. Instrum. Meth. **A835**, 186 (2016).

[47] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Harcourt, 1976).

[48] S. Tamura, Phy. Rev. **B31** (1985).

[49] S. Tamura, Journal of Low Temperature Physics **93**, 433 (1993).

[50] S. W. Leman, Rev. Sci. Instrum. **83**, 091101 (2012), arXiv:1109.1193 [physics.ins-det] .

[51] U. Fano, Phys. Rev. **72**, 26 (1947).

[52] A. Broniatowski, Journal of Low Temperature Physics **176**, 860 (2014).

[53] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, ACM TRANSACTIONS ON MATHEMATICAL SOFTWARE **22**, 469 (1996).

[54] S. Tamura and H. J. Maris, Phy. Rev. **B31** (1985).

[55] H. J. Maris, Phys. Rev. **B41**, 9736 (1990).

[56] S. B. Kaplan, C. C. Chi, D. N. Langenberg, J. J. Chang, S. Jafarey, and D. J. Scalapino, Phys. Rev. **B14**, 4854 (1976).

[57] C. Herring and E. Vogt, Phys. Rev. **105**, 1933 (1957).

[58] A. Phipps, University of California, Berkeley (2016), PhD Dissertation.

[59] V. Aubrey-Fortuna, AIP Conference Series **1185** (2009).

[60] P. Redl, Internal CDMS Note (2013).

[61] R. Courant, K. Friedrichs, and H. Lewy, Mathematische Annalen **100**, 32 (1928).

[62] W. Shockley, Journal of Applied Physics **9**, 635 (1938), http://dx.doi.org/10.1063/1.1710367 .

[63] S. Ramo, Proc. Ire. **27**, 584 (1939).

[64] S. Hertel, Massasschusetts Institute of Technology (2012), PhD Dissertation.

[65] K. Schneck, Stanford University (2015), PhD Dissertation.

[66] V. Aubry-Fortuna and P. Dollfus, Journal of Applied Physics **108**, 123706 (2010), http://dx.doi.org/10.1063/1.3520656

BIOGRAPHICAL SKETCH

As a child, Rob Agnese was always interested in intellectual challenges. He spent much of his young childhood solving puzzles: logic problems, number games, video games with a puzzle or strategy aspect. He loved school and was so thirsty for knowledge and education that on his first day of kindergarten, he ran from his mother to the school bus without looking back- a story that his mother shares with some chagrin.

He took his first physics course in high school and became somewhat obsessed. He had always excelled at mathematics and greatly enjoyed it, but for the first time he had experienced the real power of mathematics, as he saw it, which was to truly understand the world in an incredibly precise sense that was impossible by any other means.

Early in his undergraduate degree Rob became fascinated by computers as well as physics. His first research experience was working with Dr. Darin Acosta as part of the CMS experiment at CERN. His task was to write code to analyze and generate plots of some muon data coming from the detector for use in calibrating. This was before the LHC had officially been "turned on".

He began working in Dr. Tarek Saab's lab toward the end of his undergraduate career. He mostly worked on lab computing tasks, writing some useful applications for monitoring the cryogenic fridge status.

Rob completed his undergraduate work in the spring of 2010, earning a B.S. in physics and a B.A. in mathematics from the University of Florida. He decided to stay on with Dr. Saab for graduate school and begun his PhD program in fall of 2010.

As part of the SuperCDMS collaboration he attended several conferences, mostly to share progress on his PhD project: a low energy condensed matter simulation framework. He completed his PhD in the spring of 2017.