TESTING AND CHARACTERIZATION OF SUPERCDMS DARK MATTER DETECTORS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Benjamin Shank
May 2014

This dissertation is online at: http://purl.stanford.edu/mr247nw3394

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Blas Cabrera, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Peter Graham

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Douglas Osheroff

Approved for the Stanford University Committee on Graduate Studies.

Patricia J. Gumport, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Preface

The Cryogenic Dark Matter Search (SuperCDMS) relies on collection of phonons and charge carriers in semiconductors held at tens of milliKelvin as handles for detection of Weakly Interacting Massive Particles (WIMPs). This thesis begins with a brief overview of the direct dark matter search (Chapter 1) and SuperCDMS detectors (Chapter 2). In Chapter 3, a ^{3}He evaporative refrigerator facility is described. Results from experiments performed in-house at Stanford to measure carrier transport in high-purity germanium (HPGe) crystals operated at sub-Kelvin temperatures are presented in Chapter 4. Finally, in Chapter 5 a new numerical model and a time-domain optimal filtering technique are presented, both developed for use with superconducting Transition Edge Sensors (TESs), that provide excellent event reconstruction for single particle interactions in detectors read out with superconducting W-TESs coupled to energy-collecting films of Al.

This thesis is not intended to be read straight through. For those new to CDMS or dark matter searches, the first two chapters are meant to be a gentle introduction for experimentalists. They are by no means exhaustive. The remaining chapters each stand alone, with different audiences. As is becoming increasingly common, many of the figures included here assume that you are reading a color pdf. If you have somehow acquired a black-and-white copy and want a color version, look on http://cdms.berkeley.edu/dissertations.html or email me at helium3ben@gmail.com.

Acknowledgements

To a large degree, my accomplishments in the past seven years have been to build new pieces of much bigger structures which existed before I arrived and will carry on after I leave. Through this process, I have been surrounded by innumerable extraordinary people, whom I will nevertheless now attempt to enumerate, or at least categorize.

Doug Osheroff took me in as a first year student and got me started on a project very quickly. Although that project did not work out, I am grateful to Doug and his students, Qiang Qu and Lidiya Polukhina for everything they taught me about low temperature physics and laboratory practice.

It is very easy in academia to value yourself based on the science you are producing at the moment. This is a dangerous tendency for a (s)low temperature physicist. I am eternally grateful to Betty Young who, when I was producing nothing, helped me reassess and regroup.

Blas Cabrera has been an extraordinary research advisor. He gave me incredible freedom to choose the research I wanted to pursue while as the same time making sure that I stayed on project and produced a meaningful and thorough body of work.

I have been very blessed to be surrounded by colleagues in the Stanford CDMS group who are not only good friends, but patient teachers: my fellow grad students, Jeff Yen, Robert Moffatt and Kristi Schneck; the fabrication team, Paul Brink, Matt Cherry and Astrid Tomada; and Daniel Nagasawa who, while nominally my student, made me think about the world in ways I would never have considered. Between shifts at the mine and phone and email conversations, I am indebted to more CDMS collaborators than can be named here. Plugging into such an accepting and helpful group has been an amazing way to grow as a scientist. When we borrowed the

[3]He fridge from Sarah Church's group, Keith Thompson went above and beyond professional courtesy to make sure we came up to speed on its cryogenic and readout capabilities quickly.

Navigating graduate school involves much more than learning and doing science. I am grateful to the staff at Varian, especially Maria Frank and Marcia Keating, for demystifying everything from purchase orders to graduation forms. Karlheinz Merkle and everyone in the Physics Machine Shop have provided invaluable advice and worked wonders turning my rough drawings into precisely crafted objects. Lucy Zhou and Liz Palmquist have been very patient with me as I learned the ins and outs of chemical safety, inventory management and all the habits and hygienes needed to run a safe and secure laboratory.

Three months after getting married, my wife Kati set out for California with me in the first stage of our new life together. In the past seven years we have struggled and grown together in ways neither of us could have foreseen. Through everything she has been an incredible support and patiently put up with far too many late nights and weird living situations. We want to thank Chi Alpha Christian Fellowship, especially Glen and Paula Davis, for supporting us and praying with us through good times and bad. We are also indebted to friends from Nanosolar and Escondido Village for helping us along as we all figured out life and parenting together.

We are both indebted to our parents for helping us choose meaningful paths in life and supporting us at every stage along the way. Finally, we thank God for His continued sustaining and guiding hand in our lives, both in invisible and in very direct ways.

Contents

List of Tables

List of Figures

Chapter 1

Dark Matter

1.1 The Problem

At the close of the 1920s, the theory of gravitation as elucidated by Newton and Einstein was perhaps the most successful scientific theory to date, the major structures of the universe were largely understood and the art of lens-making had advanced to such a degree that astronomers could really get down to the nitty-gritty bookkeeping of cataloguing stars and galaxies. It therefore came as something of a shock in the early 1930s when various astronomers (Jan Oort[1] and Fred Zwicky[2] get the credit) pointed their powerful telescopes at, respectively, our own motion within the Milky Way and galactic nebulae and declared that they were moving too fast. That is, the observable mass within each system was insufficient, often by several orders of magnitude, to keep the structures gravitationally bound given their reckless speeds. Later measurements using gravitational lensing (the bending of light by super-massive objects) confirmed that most objects larger than star systems contain far more gravitational mass than the sum of their stars, free atoms and other known masses can explain. For those who saw it as their job to describe the Universe in detail, failure to account for 80% of its mass was considered unacceptable. The excess mass was termed 'dark matter' and the search for it has been on for nearly a century.

1.2 Cold Dark Matter In Galaxies

Space is very big indeed. Saying that there is excess mass somewhere in the universe is theoretically very interesting, but practically not so helpful. In 1980, Vera Rubin, Kent Ford, and Norbert Thonnard published a number of galactic rotation curves[3] in which they measured the relative speed of stars as a function of distance from the center of their galaxies. Any star traveling in a circular orbit betrays by its speed the amount of mass enclosed in its orbit according to Newton's Law of Gravitation.

$$F_c = \frac{mv^2}{r} = \frac{GMm}{r^2}$$
$$v = \sqrt{\frac{GM}{r}} \tag{1.1}$$

The Virial Theorem[4] tells us that any three dimensional object in a closed orbit should have kinetic energy equal to $^{-1}/_{2}$ of its potential energy. Since kinetic energy is a function only of speed, the speed calculated in Equation 1.1 holds for non-circular orbits as well when v is replaced by its time average $\langle v \rangle$. Figure 1.1 shows a typical rotation curve along with a predicted curve if the observed mass of the galaxy were the only source of gravity. (Aside: Locating sub-galactic objects at inter-galactic distances is hard enough without tracking their motion over millions of years. Velocity distributions rely on redshift of known spectral lines to determine edge-on velocities.) Rubin et. al. observed that stars in outlying regions travel at roughly the same speed instead of falling off to lower speeds as the radius of their orbits increased without incorporating much more mass. This suggests that dark matter is distributed concentrically within a galaxy and maintains substantial mass density well beyond the observable stars. Observations of satellite galaxies, such as globular clusters orbiting our own Milky Way, indicate that the dark matter halo usually extends to many times the radius of the 'baryonic galaxy' it supports.

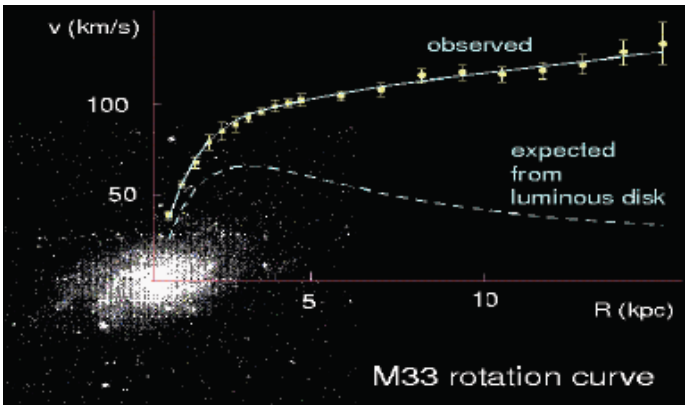


Figure 1.1: Rotation curves, velocity versus radius. The discrepancies between observed and expected velocities imply substantial invisible mass at high radius.

1.3 Weakly Interacting Massive Particles

As happens in such cases, there is a proliferation of theories to explain the missing mass. The simplest is that telescopes have limited resolution and there are whole classes of objects which they are simply unable to observe. If galaxies were packed with black holes or even Earth-sized rocks, they would be difficult to detect with existing telescopes. A number of recent cosmological observations limiting the fraction of baryonic mass in the early Universe make this hypothesis increasingly hard to defend. For example, relic abundances of light elements limit the fraction of baryons in the early Universe to about 5% of the critical density while observations of Cosmic Microwave Background anisotropies measured by the Wilkinson Microwave Anisotropy Probe (WMAP) demand some form of additional matter density around 23% of the critical density. For a more complete and readable discussion of cosmological evidence for dark matter, see [6]. A non-baryonic component to the energy density of the Universe almost demands that some new particle account for the locally-observed excess mass without interacting in a noticeable way with normal baryonic matter.

The early Universe was so hot and so dense that even very massive particles were freely created and absorbed by the thermal bath regardless of their cross-section of interaction. As the Universe cooled, first heavier particles and then successively lighter ones began to 'freeze out', fixing their abundances. However, the anti-particles

also froze out, annihilating the particles and contributing to the photon background. In the case of baryons, a slight asymmetry between particles and anti-particles led to the remainder we see today. In the absence of a similar mechanism for dark matter, the only way to avoid complete annihilation is to never encounter the anti-particles. This requires particles to have such a low cross-section that most of them pass through the universe from freeze-out to present without interacting. The cross-section needed to protect particles in the 10-1000 GeV/c^2 range is approximately the weak force cross-section ($< 10^{-40}$ cm^2). This is sometimes referred to as the WIMP Miracle and provided strong motivation for early WIMP (Weakly Interacting Massive Particle) searches. As we push the maximum cross-section for likely particles far below 10^{-40} cm^2, the WIMP Miracle becomes less convincing. It does, however, continue to motivate searches at very low cross-sections.

Many theories of supersymmetry predict such WIMPs and assign them suitably large masses. Supersymmetric models pair each observed fermion (boson) with an as-yet-undiscovered bosonic (fermionic) partner. The Minimal Supersymmetric Standard Model (MSSM) was originally proposed to provide a fermionic super-partner to the Higgs boson. Large boson masses are unstable to quantum corrections. Introducing a fermionic Higgsino that must share a mass scale with the Higgs boson sets a natural energy scale for all particles, which many theorists are counting on to solve the hierarchy problem.[1] If quantum corrections to the Higgs mass are to be on the same order of magnitude as the observed mass of 126 GeV[5], the masses of the super-partners need to also be on this scale. The most commonly cited supersymmetric models call for 100 GeV WIMPs, but it takes only a few additional parameters to allow for super-partners on the 10 GeV scale. This scale has been probed by (among others) the Tevatron and LHC, but super-partner production from standard particles would be quite rare and the resulting WIMPs would only be detectable by the energy they carried away, making them hard to distinguish from neutrinos.

MSSM cannot conserve baryon and lepton numbers *per se* since it requires couplings between particles of different spin. But clearly any extension of the Standard

[1]That is: Why do the four fundamental forces have radically different strengths? Possibly they represent different couplings to some energy scale far higher than common energies observed in the present epoch.

Model must have a property that reduces to baryon and lepton conservation in the SM limit. This is accomplished for MSSM by maintaining R-parity with

$$R = 3B + L + 2S \tag{1.2}$$

where B is baryon number, L is lepton number and S is spin. MSSM requires that R is even for all standard particles and odd for all supersymmetric particles. If R-parity is true, the lightest supersymmetric particle (LSP) cannot decay into anything and so must be stable. If the lightest super-particle matches the lightest particle, it will be the super-partner of the neutrino, or neutralino. This provides a natural mechanism to have a stable particle in the 10-1000 GeV/c^2 range with weak coupling to standard particles, a heavy WIMP.

Figure 1.2: The Bullet Cluster

1.4 Modified Newtonian Dynamics?

Before we expend large amounts of time and money (not to mention the rest of this thesis) searching for a particle that may not exist, it is worth questioning the theoretical underpinnings of dark matter. After all, rotation curves and supercluster

dynamics are modeled using Newtonian gravity. Newton's equations have proved very successful at interstellar scales, but cosmology necessitates extrapolating well beyond any conceivable previous measurement. One could imagine that $F = ma$ may not hold at enormous length scales or for tiny accelerations. For example, it has been proposed by Mordehai Milgrom[7] that

$$\frac{F}{m} = \frac{a^2}{a_0} \tag{1.3}$$

below some critical acceleration $a_0 \approx 10^{-10}$ m/s^2. Plugging in $F = {}^{GMm}/_{R^2}$ from Newtonian gravity, we get

$$\frac{GM}{R^2} = \frac{a^2}{a_0} \rightarrow a = \sqrt{\frac{GMa_0}{R^2}} \tag{1.4}$$

$$a = \frac{v^2}{R} \rightarrow v = \sqrt{R\sqrt{\frac{GMa_0}{R^2}}} = (GMa_0)^{1/4} \tag{1.5}$$

This solves the problem of flat rotation curves and deals nicely with a lot of supercluster dynamics in the bargain, even if it does require stretching the concept of momentum conservation a bit. Figure 1.2 shows a galactic supercluster known as the Bullet Cluster. The Bullet Cluster was formed when two superclusters collided. The star systems passed by each other, forming the outer purple regions in the figure. The interstellar gas from one supercluster collided with the gas from the other supercluster and formed a big X-ray-emitting ball at the center of impact (the red region). Even though free hydrogen makes up 80+% of the baryons in most galaxies, gravitational lensing measurements have shown that the majority of the mass in the Bullet Cluster is in the purple regions. In this case, the lensing mass modeled by dark matter has been physically separated from most of the baryons. If there is a Modified Newtonian Dynamic that accounts for dark matter, it must treat hydrogen in stars and free hydrogen differently. On the other hand, WIMPs previously confined to the two superclusters would pass right through the cloud of hydrogen and end up in the purple regions as observed.

1.5 The Hunt Is On!

Assuming that dark matter is primordial, a model is needed to explain the co-evolution of visible galaxies with their dark matter halos. In the beginning (well, after nucleosynthesis and recombination), atoms and dark matter clumped together around local density fluctuations. These roughly spherical proto-galaxies had low angular momentum for their size, but (for many galaxies, including ours) the atoms, slowed by frequent collisions, collapsed into a spinning galactic plane and further clumped into stars. WIMPs, which may pass through a star without a single interaction, would not collapse. Matching flat rotation curves requires that $^{dM}/_{dr} = 4\pi r^2 \rho$ be constant. The standard model for spiral galaxies calls for the observed spinning disc of stars with density e^{-r/r_D} where r_D is the 'optical radius' embedded in a much larger spherical halo of dark matter with density $\rho \sim {}^1/_{r^2}$ and no appreciable spin. (Typically the dark matter density is modeled as $(r^2 + a^2)^{-1}$ where a is small to avoid infinite density at the center of the galaxy.) While the dark matter, on average, has no velocity with respect to the galactic center, individual particles may be traveling quite fast, with an upper bound set by the galactic escape velocity (about 550 km/s near Earth).

So where does this leave prospective dark matter hunters? We strongly suspect that our planet is traveling at 230 km/s through a stationary cloud of invisible particles each moving at about that speed in a random direction. We know from rotation curves that the local density of dark matter is $0.008 \pm 0.002\ M_{sun}/pc^3 = 0.3 \pm 0.1$ GeV/cm^3[8], but our best estimate of particle mass is 10-1000 GeV (maybe, we think) so we have no way to calculate the particle flux. By construction, light-collecting telescopes cannot be used to detect WIMPs directly. Several schemes have been proposed to indirectly detect them by assuming a decay or interaction model, but their success is entirely model dependent. A more robust detection scheme would be to somehow detect the WIMPs directly as they went by.

Chapter 2

CDMS

If WIMPs are anywhere, they are everywhere, which means they are here. All we can definitely say about them is that they ought to interact with atomic nuclei at the scale of the weak nuclear force (hence the name). So detecting them should just be a matter of gathering together a large number of atomic nuclei and watching them very, very closely for unexplained impacts. This goal is somewhat complicated by the rather speculative nature of our knowledge regarding the relevant particle physics.[1] Various MSSMs motivate compelling regions of interest in WIMP mass and cross-section with nucleons, but those regions do not lie particularly near each other. Figure 2.1 shows the state of WIMP-search theory and experiment at the end of 2013. It will be incomplete by the time this is published, but gives some sense of the search ahead.

Figure 2.1 assumes that WIMPs interact with all nucleons coherently, although it is easy to construct novel particles with some spin-dependent cross-section. Even ignoring more exotic possibilities like axions and gravitinos, predicted masses span four orders of magnitude, making it difficult to say which nucleus will make the best target. Predicted cross-sections span eight orders of magnitude, so we have no idea how many nuclei to collect or how long we have to watch them to get a hit. To further complicate matters, all experiments for the foreseeable future must occur on a radioactive planet under a hail of ionizing radiation from the atmosphere which

[1] As always, no rule limits the sophistication of our speculation.

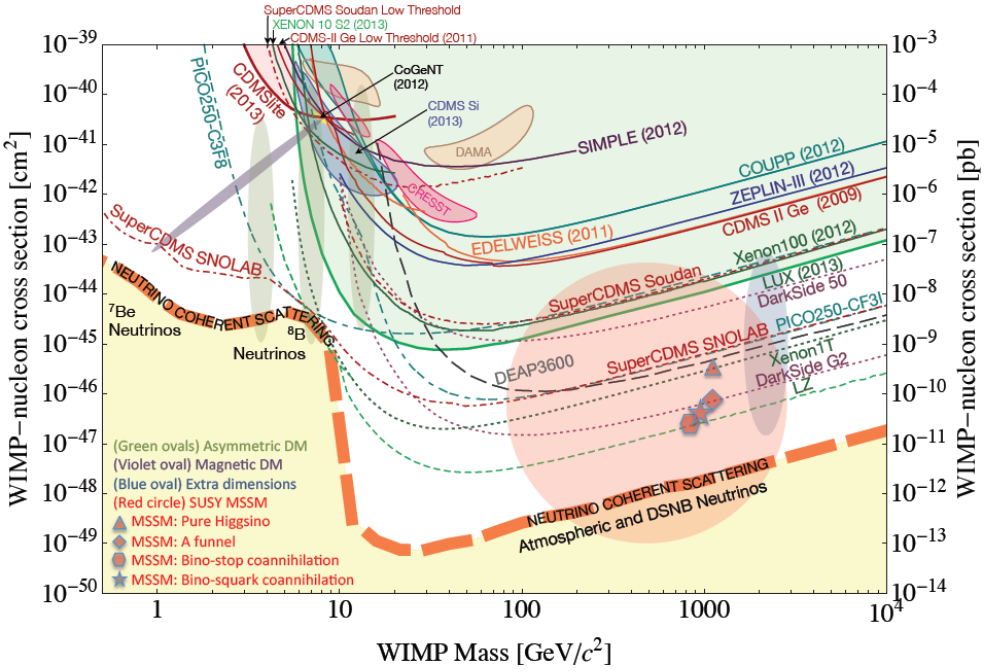


Figure 2.1: Limit plot showing the state of spin-independent dark matter detection as of late 2013. Solid open curves exclude the region above and to their right. Solid closed curves represent alleged WIMP detections with 90% confidence. Dashed lines represent estimated exclusion limits of proposed future experiments. Un-bordered shaded regions represent WIMPs predicted by supersymmetric and other theories. The heavy orange line at the bottom represents the local neutrino background. It is impossible to conduct a WIMP search on or near Earth which probes below this line with < 1 expected background event.

must be scrupulously shielded or vetoed to isolate the WIMP signal. And if we look too closely, we run up against a background of neutrinos, for which it is hard to imagine an effective shield.

2.1 Getting Started

On that cheerful note, let's dig in. Would-be WIMP hunters have two choices to make before they start. First, they must select a type of nucleus to watch and, second, they must decide what kinds of energy to watch for. These two choices are usually

correlated. Atomic impacts at non-relativistic energy scales generate three kinds of readily accessible signals: ionization (electron-hole pairs), light (if the material is a scintillator) and heat (phonons). To be measured, these signals must travel through the detector unimpeded, so the choice of target material tends to determine what gets measured. Figure 2.2 shows the detection triangle described above and notes collaborations around the world attempting every single type of measurement or pair of measurements.[2] This level of disorganization is actually an important safeguard against our ignorance. When you don't know anything, try everything. The best designed experiment is still susceptible to the chosen target atom not having any cross-section to WIMP impacts.

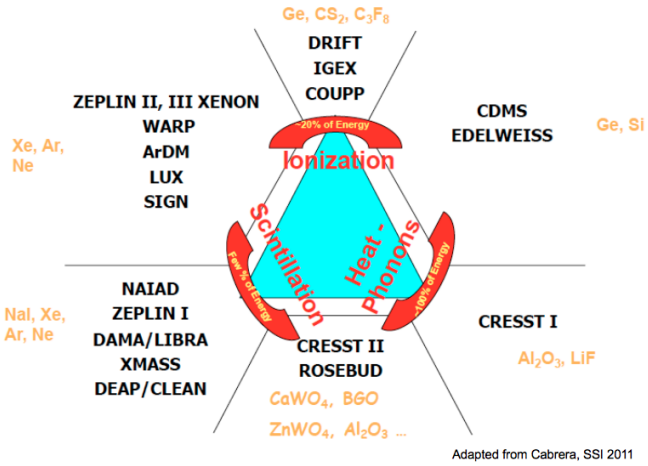


Figure 2.2: Direct WIMP searched collect many forms of energy from many different target nuclei.

With this important caveat, germanium has a number of advantages as a target atom. With a density of 5.32 g/cm^3, solid germanium crystals pack a large mass (many nucleons) into a small volume. Models of weak interactions predict cross-sections per nucleon proportional to the number of nucleons per atom at zero energy transfer and proportional to the square of the nuclear form factor. Lewin and Smith[12] show the expected spin-independent cross section for a handful of common

[2]Figure 2.2 does not show experiments which measure heat that induces a phase transition in a bubble chamber such as PICASSO.

target atoms. Germanium is not explicitly included, but can be derived from their Fig 9. This has been done in Figure 2.3. With an atomic mass of 72.6 it falls short of xenon and iodine cross-sections at low energy, but the first node of its form factor occurs well above any reasonable kinetic transfer, making it a good choice for detecting particles in the mass range of interest. Germanium-73 (7.7% of the natural abundance) has a nuclear spin of $^9/_2$. If it turns out that WIMPs have spin-dependent cross-sections, germanium will still provide some data.

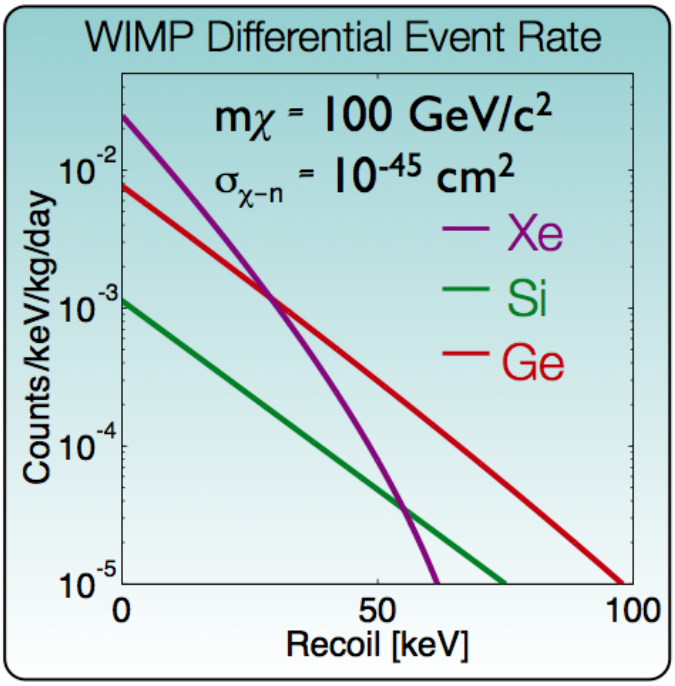


Figure 2.3: Expected event rate for xenon, silicon and germanium targets calculated from [12] assuming a mass of 100 GeV/c^2 and cross-section of 10^{-45} cm^{-2}.

Germanium is not much of a scintillator, so this choice of material leads automatically to collection of phonons and ionized carriers as a proxy for detecting WIMPs. Since we are trying to collect small amounts of energy with great precision, this is also a promising choice. The First Law of Thermodynamics ensures that the energy of impact quickly ends up in the form of heat. As much as 20% of the energy may contribute to ionized electron-hole pairs. By comparison, even in very good scintillators,

only a few percent of the energy of impact ends up in the form of photons.

Naively, if an experiment could be completely isolated from outside influences, it would be sufficient to detect the heat generated by nuclear recoils. However, as previously discussed, the Earth is a complicated and radioactive place. Any experiment with single-recoil sensitivity should expect to see many signatures from non-WIMP sources, both terrestrial and cosmogenic. These background events need to be rejected with great confidence before a small number of events over a long period can be attributed to WIMPs. For this reason, it is desirable to collect as much information as possible about each event. To do this, the CDMS collaboration needs germanium crystals which propagate phonons and free charge over macroscopic distances in a predictable, reproducible fashion. We must be able to measure small amounts of deposited energy with high precision and we must reduce the number of background events as much as possible.

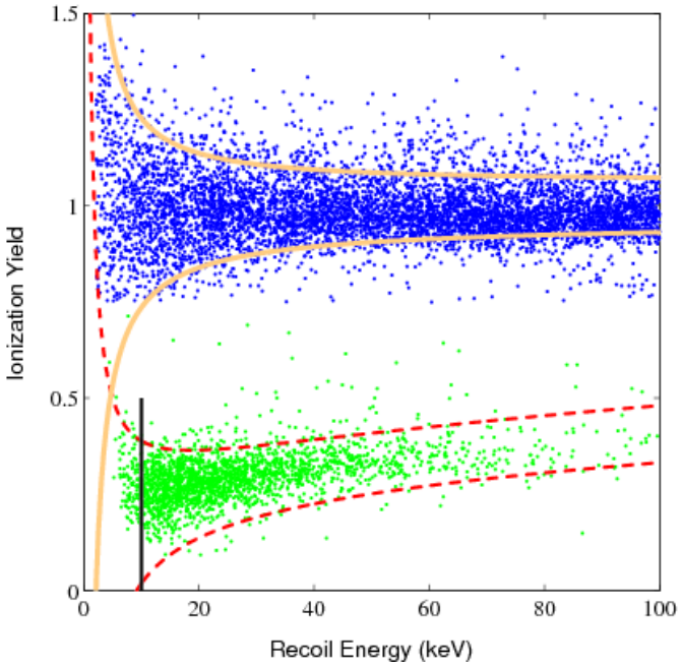


Figure 2.4: Ionization yield normalized to total energy distinguishes between electron (background, blue) and nuclear (signal, green) recoils

WIMPs, by definition, are electrically neutral and will only interact with the

nuclei of atoms. Most background particles (muons, alphas, betas, gammas) will hit the much larger atomic-scale cross-section of the electron clouds, overwhelmingly producing electron recoils. These events also create ionization and phonons, but in different ratios from a nuclear recoil. By collecting both electrons and phonons, we can reject electron recoils by comparing the yield of ionized particles per total energy. Figure 2.4 shows a yield plot versus total energy from the CDMSII experiment. The average yield for electron recoils is scaled to 1. The 2-sigma confidence limit around yield=1 for electron recoils and yield$\sim$0.25 for nuclear recoils is shown. This technique provides powerful rejection of electron recoils, but it also requires extensive position reconstruction of each event to make sure all energy is captured and accounted for.

2.2 Phonon Collection

A superconductor is often described as a material which switches suddenly from a finite electrical resistance to zero resistance as the temperature is slowly decreased. In fact, there is a narrow transition edge in which the resistance makes the switch at a rapid, but finite, rate. Figure 2.5 shows a typical superconducting R(T) curve through the transition edge. A superconductor held in the transition edge displays measurable changes in resistance for very small changes in temperature ($\frac{dR}{dT}$ large), which makes transition edge sensors (TES) very accurate thermometers (albeit over very narrow temperature ranges). The sharpness of the transition is characterized by a dimensionless parameter $\alpha \equiv \frac{d(logR)}{d(logT)} = \left(\frac{T_c}{R}\right)\left(\frac{dR}{dT}\right)$.

CDMS uses tungsten TESs with transition temperatures between 70 and 90 mK on the flat top and bottom surfaces of our crystals. Sputtered tungsten forms admixtures of two phases with wildly different critical temperatures depending on exact conditions. Tuning these to reach 70-90 mK is an interesting metallurgical problem in its own right.[10, sec 2.4] Rather than try to keep thousands of TESs in the same 2 mK wide transition edge, electro-thermal feedback (ETF) is used to maintain each TES in its transition region. The germanium substrate is held somewhat lower than the desired operating range, typically 50-60 mK. At constant applied voltage the TES dissipates power V^2/R. This leads to rapid warming of the superconducting (R$\approx$0)

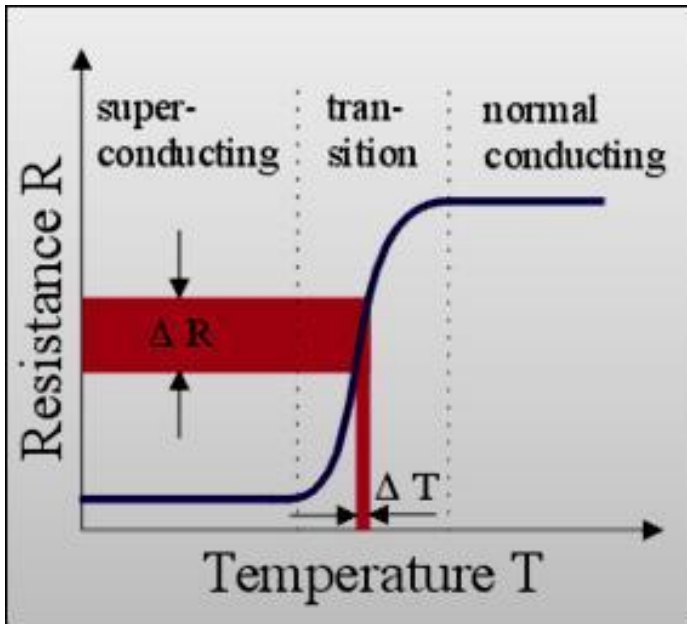


Figure 2.5: Cartoon superconducting transition curve highlighting the relatively linear region with large $^{dR}/_{dT}$ which makes transition-edge sensors extremely sensitive thermometers.

phase until it reaches the transition edge. Somewhere in the middle of the transition, the joule heating matches the power loss to the substrate and the TES reaches equilibrium. Increases in temperature result in higher resistance which decreases the joule heating and cools the TES back down. Temporary decreases in temperature have the opposite effect.

Compared to CDMS operating temperatures, phonons generated in the aftermath of a particle collision are extremely hot; the Debye temperature for phonons in germanium is 360 K. To facilitate collection and ensure that we get only hot phonons instead of the ambient thermal bath, TESs are attached to aluminum fins. To give a sense of scale, the TES highlighted in red in Figure 2.6 is 220 um long and 2.4 um wide while the fins collectively might be 650 um by 300 um. Aluminum super-conducts at 1.2 K. Normally superconductors do not couple electrically to phonon systems, but a phonon hotter than twice the superconducting energy gap (about 4 K for aluminum) can break Cooper pairs. If these freed quasiparticles can drift back to the

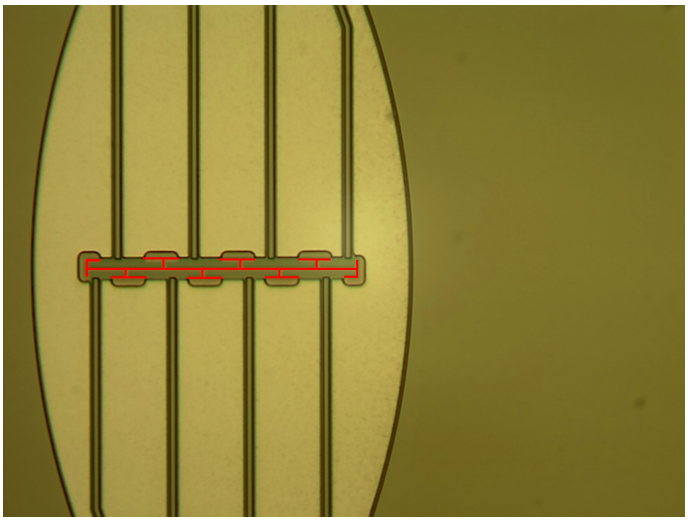


Figure 2.6: Close-up of a TES with aluminum fins. Red highlight covers tungsten TES strip with eight spurs to wide overlap regions with aluminum fins. Quasiparticles liberated in the aluminum by phonons from the substrate cascade down to the lower band gap of the tungsten, producing a measurable heat pulse.

TES before recombining, they warm the tungsten, decreasing the current across our voltage-biased devices and creating a measurable pulse. See [13] and [14] for studies on the diffusion length of quasi-particles in aluminum. Coupling hot phonons in this manner effectively turns very sensitive thermometers into very sensitive microphones.

2.2.1 Why So Cold?

It might be worthwhile at this point to explain why we have taken on the complication of running our detectors at 50 mK. After all, many materials superconduct above 4 K; why not run lead TESs in a bath of liquid helium? Besides the inherent reduction in phonon and charge noise, TESs have energy sensitivity proportional to the cube of their critical temperature. This can be seen by supposing that a resistive sensor is limited by thermal noise. The energy sensitivity σ is ultimately the Johnson noise limit $E_{noise} = 4k_BT$, but if the sensor is active an additional factor of $\sqrt{N}$ is needed where N is the equilibrium power dissipation in units of Johnson noise.

$$\sigma = E_{noise}\sqrt{\frac{E_{equil}}{E_{noise}}} = \sqrt{E_{equil}E_{noise}} \tag{2.1}$$

In thermal equilibrium, the classic balancing equation becomes

$$c_v\frac{dT}{dt} = \frac{V_{bias}^2}{R_{TES}} - \kappa(T^n - T_s^n) = 0 \tag{2.2}$$

where c_v is the heat capacity, T is the temperature of the sensor, T_s is the temperature of the substrate and n is the conductivity exponent. Since TES's must be run in their transition edge, T is always the critical temperature T_c. Below 100 mK, thermal impedance to the substrate is dominated by electron-phonon decoupling, so $n = 5$. Provided the substrates can be kept well below the 70-90 mK target T_c range, the T_s term may be safely dropped. This should be no problem for a dilution refrigerator.[3] The equilibrium energy dissipated during a typical phonon pulse is

$$E_{equil} = P_{equil}\tau_{ph} = \kappa T_c^5\tau_{ph} \tag{2.3}$$

where τ_{ph} is the characteristic phonon arrival time, which is independent of T_c. Putting this all together, we get

$$\sigma = \sqrt{(\kappa T_c^5\tau_{ph})(4k_BT_c)} = (\text{constant})T_c^3 \tag{2.4}$$

A 10 GeV particle traveling at 500 km/s carries about 14 keV of kinetic energy. We expect to collect maybe half of that. Using typical values for our TESs gives an energy sensitivity of 500 keV at 1 K. This is great for detecting relativistic neutrons, not great for detecting cold dark matter. Our best detectors, running at 70-75 mK, boast energy thresholds in the 2-3 keV range.

It should be noted that Equation 2.4 is only valid if the sensor is able to dissipate energy much faster than it arrives, that is if $\tau_{ETF} << \tau_{ph}$ where τ_{ETF} is the C/G thermal time constant enhanced by a factor of $\frac{1}{1+\alpha/n}$ due to ETF[15]. Once T_c is decreased to the point where this is no longer the case, the resolution improves at the slower rate of $T_c^{3/2}$ because the sensor must be increased in volume to ensure enough

[3]Standard experimental physics caveat: 'should be' does not mean 'is'.

heat capacity to prevent saturation[16]. The quasi-particle diffusion device discussed in Chapter 5 is designed with fast τ_{ph} in mind, so Equation 2.4 should not be carried beyond the end of Chapter 4.

2.3 Charge Collection

In addition to hot phonons, energetic particles ionize electron-hole pairs near the point of impact. Typically the triggering particle only hits one electron or nucleus, but the collision imparts enough energy to set off a cascade of chain ionizations. Although the band gap of germanium is only 0.744 eV, significant energy is lost to acoustic phonons. Any electron with insufficient energy to ionize another electron and remain free itself ends the cascade. In effect this means the number of electron-hole pairs generated is equal to the primary electron recoil energy divided by 3 eV.[10] To capture these carriers, dedicated aluminum rails carry ± 2 V to the top and bottom surfaces of the crystal, setting up a small ($1.6\ ^{\mathrm{V}}/_{\mathrm{cm}}$) electric field. Electrons and holes drift to opposite faces and are trapped there until they can leak through a layer of amorphous silicon. Their motion sets up a Ramo potential in the electrodes. Drift energy is supplied entirely by the external field, so the resulting current is a reliable measure of impact ionization. All of the 3 eV per electron-hole pair is converted to acoustic phonons, so the phonon energy is also a reliable measure of total impact energy. As electron-hole pairs are separated across the potential difference V, they acquire additional energy qV, which is converted to Luke-Neganov phonons caused by charges traveling faster than the speed of sound. The additional energy can be subtracted off to first order using the charge measurement and the known potential drop, so this is not immediately fatal to total energy reconstructions. The 4 V potential is chosen so that the Luke phonon energy is about equal to the intrinsic energy. Higher fields would increase charge collection efficiency to a point, but at the expense of phonon signal-to-noise ratios.

2.3.1 Neutralization

For a semi-conductor at 50 mK, thermodynamic equilibrium means charge states are entirely determined by the Fermi level. All donors are D^+ and all acceptors are A^-. If we aim to detect $\sim 10^4$ charges moving more than 1 cm through a crystal with $\sim 10^{10}$ impurities/cm^3 under a weak electric field, this is a problem. We should aim not for equilibrium, but for neutrality (D^0 and A^0). After cooling in equilibrium, detectors are "neutralized" at base temperature by repeatedly flashing infrared LEDs ($\lambda = 940$ nm) mounted close to the detector faces. Because the 1.32 eV LED photons exceed the 0.744 eV band gap, a cloud of electrons and holes is created in the first 1 μm of the detector surface. With the faces of the crystal grounded together, carriers are free to diffuse until they reach a region of non-negligible space-charge. The internal field created by these trapped charges draws charges liberated by the LEDs in so they are rapidly captured, returning traps to a neutral state. With the detector thusly prepared, it can be biased and operated in WIMP-search mode.

2.4 Detector Technologies

When I joined the CDMS collaboration in 2010, data had just been released from the CDMSII experiment which used Z-sensitive Ionization and Phonon (ZIP) detectors, sometimes retroactively referred to as oZIPs or BMW-style detectors. Each ZIP was photolithographically patterned onto a germanium or silicon disc 3" in diameter and 1 cm thick, allowing a 4 $^{\mathrm{V}}/_{\mathrm{cm}}$ field. Four phonon channels in one face split the volume into quadrants. Each phonon channel consisted of hundreds of TESs wired in parallel and sharing a single SQUID[4] output. Two concentric charge channels on the opposite side split the volume radially into an inner 'fiducial' volume and an outer guard ring as shown in Figure 2.7. Energy splitting between the phonon channels allowed reconstruction of where the event occurred along the crystal face (x-y plane). The timing difference between the ionization and phonon signals allowed depth (z) determination. All channels were read out on a single Digital Interface Board (DIB)

[4]Superconducting QUantum Interference Device

built into the copper housing. CDMSII set the best spin-independent cross-section limits in the world at the time. Given its location in the Soudan Underground Laboratory in Soudan, MN, it could have run longer than its scheduled two-year live time before reaching 1 event of expected background. Instead, the collaboration decided to install a larger payload with an improved detector design before proceeding.

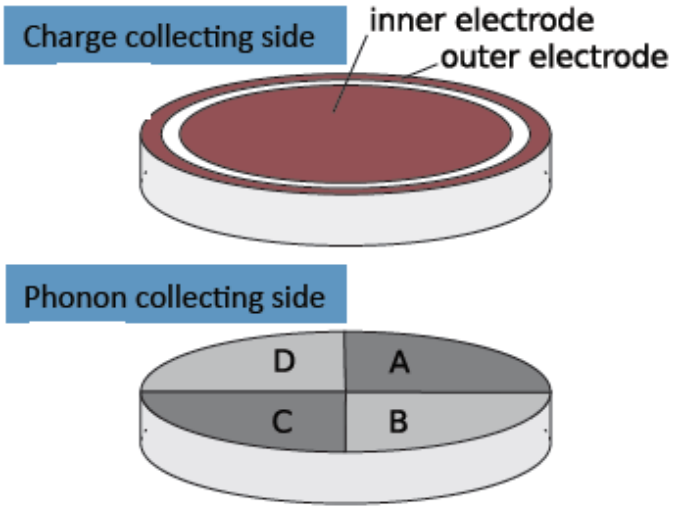


Figure 2.7: oZIP schematic. Four quadrant phonon channels on one face and two concentric charge channels on the other allowed position reconstruction of events.

In 2010, work was underway to make the next generation of Mercedes-style ZIPs (mZIPs) for SuperCDMS Soudan. These heavier 1" thick detectors would provide more target mass for the same fabrication effort as CDMSII, albeit at the cost of weaker charge separation. As shown in Figure 2.8, mZIPs had 4 phonon channels on one side and 2 concentric charge channels on the other, just like oZIPs, but with the phonon channels rearranged. Three inner channels split the angular space and the outer 'guard' channel broke radial degeneracy, allowing better position reconstruction. A 2010-2011 engineering run at Soudan revealed that mZIPs were susceptible to misidentifying surface events. Ionized carriers from events right at the surface did not have space to diffuse spherically outward before being separated by the weaker electric field. This led to many carriers of the wrong type being collected at the nearby surface,

decreasing the apparent yield and creating a potential for electron recoils (yield $\simeq$1) to be mis-identified as nuclear recoils (yield $\simeq$ 0.25). Considering that some charged particles (e.g. betas) have low penetration depth, mis-identification of surface events posed a major barrier to larger scale experiments.

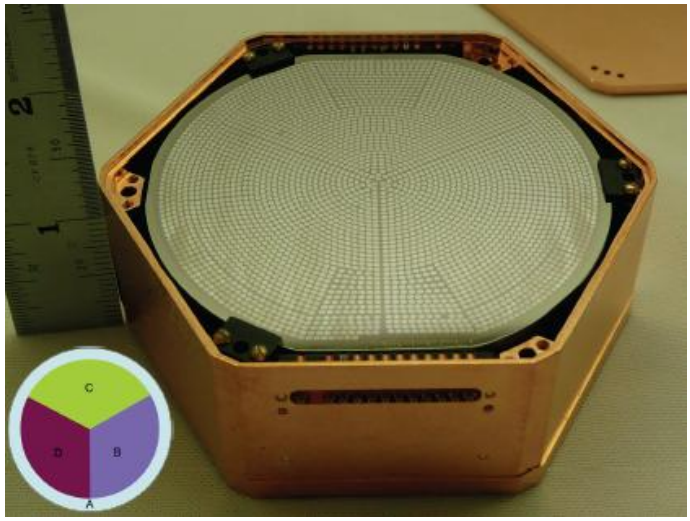


Figure 2.8: mZIP picture and schematic. The phonon channels are facing up. Note the single DIB on the front face, with two pins to read out each of the four phonon channels and a single pin for each of the two charge channels.

To reach the next level of sensitivity, SuperCDMS needed to find a way to reject surface events with high confidence. In early 2011, we started making Interleaved ZIPs (iZIPs), shown schematically in Figure 2.9. Each surface has the same 4 phonon channels as the mZIP, rotated by 60° to improve position reconstruction. In addition, each face has the two charge channels previously on the opposite face of the mZIP interdigitated with the phonon channels. This geometry is impossible in two dimensions, but the wire bonds used to attach the channels to the DIBs can also be used to jump one channel over another. Since the number of channels is doubled over the mZIP, iZIPs read out onto two DIBs[5]. Phonon channels are held near zero volts while charge lines on opposing sides are held at $\pm$2 V. This still creates a uniform field

[5]In this section, 'iZIP' refers to the 3" iZIP versions 4 and 5 developed for Soudan. The 100 mm iZIPs version 7+ being developed for SNOLab may be very different. Look for papers and theses covering SNOLab R&D in 2014 and beyond.

through the bulk of the detector, but near the surfaces field lines flow from charge to phonon lines across the surface (see Figure 2.10) Instead of sending a fraction of the charge across the detector, no carriers escape the nearby surface. Surface events can be identified with a charge symmetry cut since unequal amounts of charge reach the opposing surfaces. Of course, there are still regions of low field between the 'surface' and 'bulk' regions, but these are well away from the surface, so that low energy betas or gammas do not reach them. Carriers that pass through those locations have sufficient velocity to pass through without trapping.

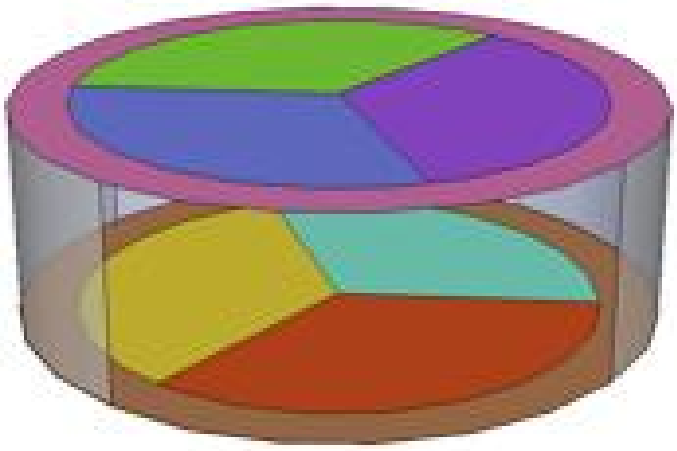


Figure 2.9: iZIP phonon schematic. Rotating each face by 60° maximizes position sensitivity.

2.5 Fabrication

iZIPs give us a great deal of information about each event, but they are also very complicated to fabricate. We start with ultra-pure boules of single-crystal germanium. The crystals are grown by the Czochralski process with a seed crystal setting the structure as the boule is spun and pulled out of molten germanium. The quality of the crystals is important, although the exact requirements are not well understood. Dopants and dislocations can trap free charge, creating barriers to transmission and disrupting our position reconstruction. Line dislocations that thread the detector

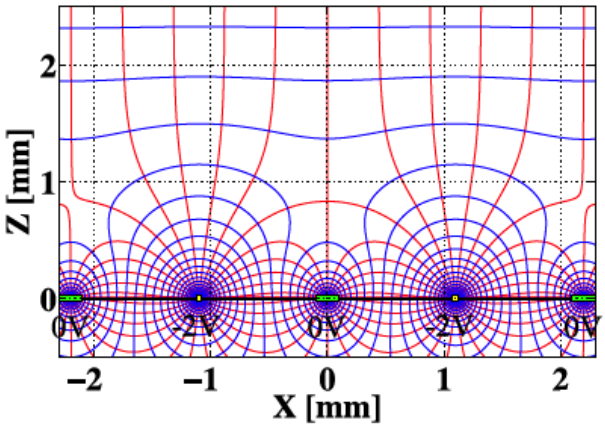


Figure 2.10: Close-up of iZIP surface field taken from [11].

may additionally create conduction paths, which swamp our charge readout with stray current. Our specs, which replicate the gamma spectroscopy Ge crystals used at 77K, currently call for $< 10^{11}$ cm^{-3} impurities (about 2.5 ppt) and < 8000 cm^{-2} etch pits.

Rough boules are cut into detector-sized slabs, shaped and polished at the Stanford Crystal Shop. To get clean lithography features in 400 Å-thick films, the surfaces must be smooth at the nanometer scale. After polishing, crystals are sonicated and chemically etched in hydrofluoric acid. Thin films of amorphous silicon (aSi), aluminum and tungsten are sputtered onto the flat surfaces in the Varian Physics building. Amorphous silicon protects the germanium from the etching chemicals used to remove unwanted metal films. Finally the films are photolithographically patterned in the Stanford Nanofabrication Facility (SNF), shaping out the TESs and rails. Figure 2.11 shows a step-by-step cartoon of the fabrication procedure. The bare HPGe (1) is sputtered with a tri-layer of 400 ÅaSi, 3000 ÅAl and 400 ÅW, first on one side (2) and then the other (3). The crystal is spin-coated in photoresist (PR) (4), which is masked and developed (5) in preparation for an Al etch (6) which also etches W. This defines the Al rails. The PR is stripped (7) and another coat of W is applied, first to one side (8), then the other (9). More PR is spun on (10) and developed (11) to prepare for the W etch (12) which defines the TESs. A third coat of PR (13) is

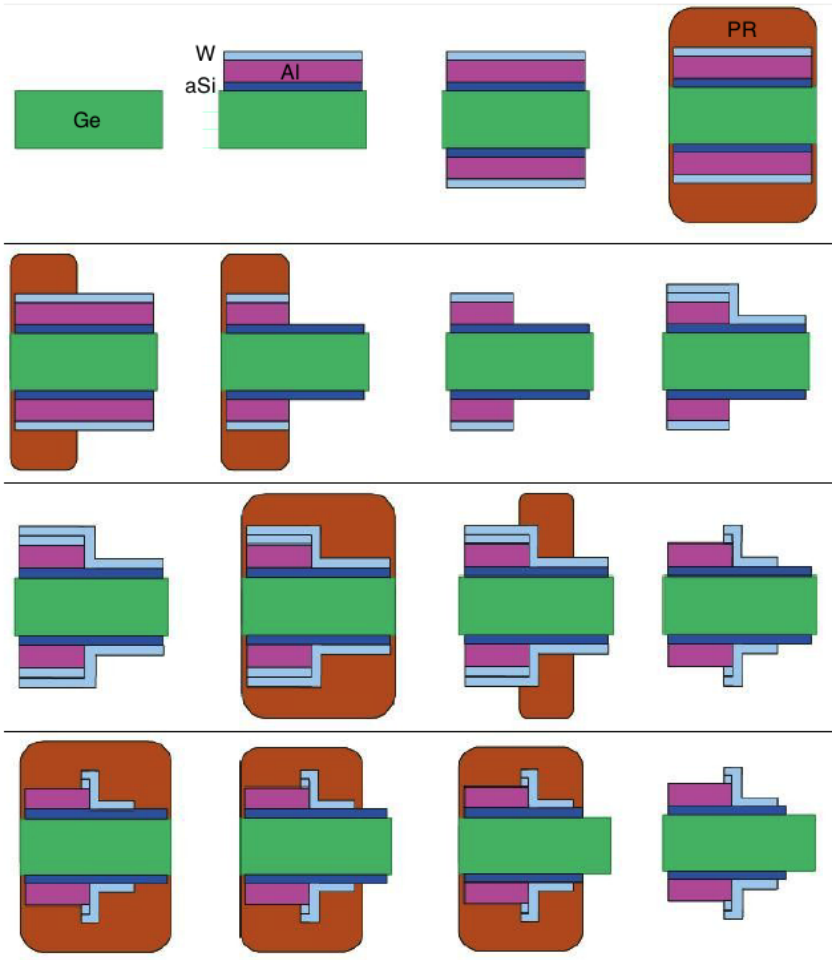


Figure 2.11: Cartoon roadmap of iZIP fabrication.

developed (14) to expose only the aSi between channels, which must be etched away (15). The aSi etch also attacks Ge, so care must be taken to ensure the correct etch time. When the final PR coat is removed (16), the detector is ready for inspection. At the end of this process, each surface has roughly 2000 TESs and line features as narrow as 2 μm. Finished detectors are returned to Varian, to the Radon Scrubbing Facility (RSF, a Class 100 cleanroom with Class 100 workbenches) for inspection, mounting in housings and wire-bonding to the DIBs.

In order to endure cross-country travel with wire-bonds intact, detectors must be firmly mounted in their housings in such a way that they do not slip when tilted

or shaken. On the other hand, we eventually want to set up a uniform electric field throughout the crystals, so electrical contact with the grounded housing must be minimized. In Figure 2.8, three Cerlex clamps can be seen just barely gripping the crystal. The three empty posts rotated by 60° support clamps on the bottom face. The contact surface of each clamp is typically 5 mm by 1 mm. Cerlex is very thermally and electrically insulating all the way down to sub-Kelvin temperatures. This is not that important once the detectors are deployed, but it must be taken into consideration when planning cryogenic testing.

From the time they arrive at SNF, detectors should never again be exposed to unfiltered air to avoid radon contamination. Radon is the only gaseous element in the uranium decay chain. Rn-222 is continuously replenished in the atmosphere and decays by alpha emission with a half-life of 3.8 days. Its daughters (Po-218, Pb-214, Bi-214, Po-214) have half-lives measured in minutes until they decay to Pb-210, which has a half-life of 22.3 years. Any clean surface exposed to air slowly accumulates Pb-210 until it comes into equilibrium with the local Rn-222 density. If Pb-210 plates out on the detector surfaces, there is no way to shield the experiment from the resulting radiation. Worse, the momentum from previous alpha decays may drive the Pb-210 a few nanometers into the crystal, making it impossible to wipe off. If this happens, the crystals must be re-polished to remove the affected layer and re-fabricated.

2.6 Signal Readout

Phonon and charge channels are very different sensors and demand different signal processing. The low impedance phonon channels lend themselves well to SQUID readout. A SQUID is a very sensitive magnetometer that, with the application of an inductor, can measure small currents very precisely. A loop broken by two Josephson junctions in parallel must contain a whole number of magnetic flux quanta. This makes the voltage across the two junctions a periodic function of the flux coupling in from the input coil, which in this case is proportional to the TES current. The function is periodic because adding a whole flux quantum to the SQUID causes a 'SQUID jump' where the system re-locks on a different equilibrium state. Input coils

must be designed so that the expected flux is small compared to the flux quantum of the SQUID. Current is sent down a feedback loop also coupled to the SQUID to keep it at the same voltage. The feedback current is thus linearly tied to the input coil current, producing an output signal with modest voltage amplification and much higher output impedance. In this way significant amplification is achieved without directly coupling a large power supply to the TES. Figure 2.12 shows the SQUID circuit of a phonon channel along with an electro-thermal picture of a typical pulse. Changes in TES resistance change the splitting of the bias current between the SQUID and the shunt resistor, transferring pulse data to the higher-power feedback line. SQUIDs rely on superconducting coils and wires and dissipate minimal heat, so they can be mounted right near the detectors. All the SQUIDs needed for one DIB are placed on a single 'SQUET'[6] card mounted on the tower assembly. The SQUIDs provide enough stability to protect against noise on the lines to the outside world, so the second-stage amplifiers can be run at room temperature.

Charge readout on a single wire necessarily has high output impedance. Just as SQUIDs must be mounted near the detectors to reduce stray input inductance, charge amplifiers have similar constraints due to stray capacitance. The charge signal arrives quasi-instantaneously but it is trapped on a capacitor and released with a characteristic time set by the feedback RC circuit of a standard FET amplifier. One major problem with this arrangement is that the silicon JFETs needed to minimize noise start to freeze out around 140 K. To get around this, FETs are supported on insulating windows to maintain the required temperature while minimizing heat load onto the cold stages. Heaters on the window panes warm them up to operating temperature; afterward, the FETs are self-heating. In the current experiment in Soudan, MN, each FET dissipates ∼5.5 mW onto the 4 K stage. This works for SuperCDMS Soudan, with 60 FETs (15 detectors x 4 charge channels), but would be fatal to any reasonable cryogenic system intended to run hundreds of CDMS-style detectors. (1 Watt boils off up to 4 L/hr from a helium bath.) In order to reach the 100-kg scale, SuperCDMS is characterizing High Electron Mobility Transistors (HEMTs)[17] which dissipate ∼0.1 mW and operate at 4 K.

[6]SQUET = SQUID + FET. The FETs on the SQUET card are for charge measurements.

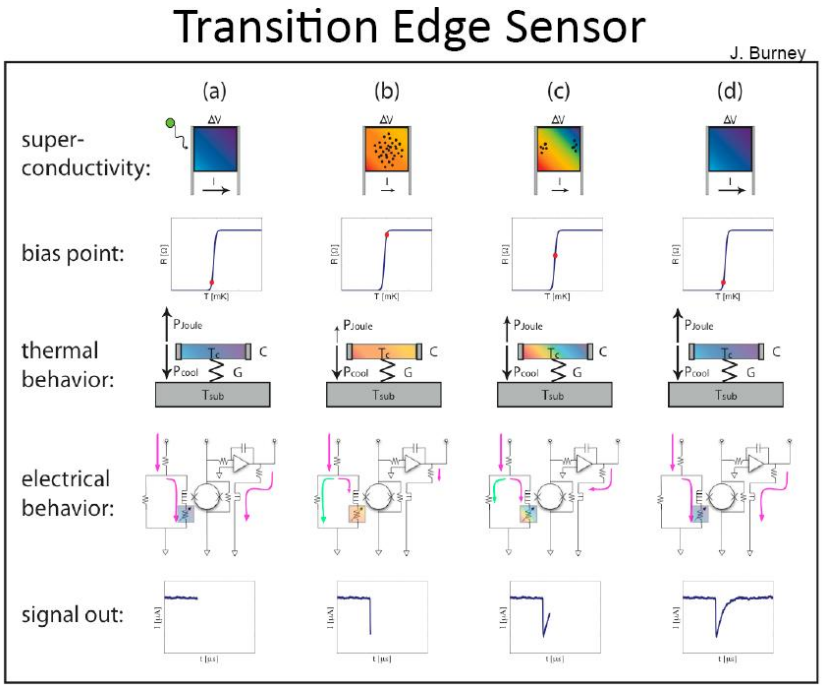


Figure 2.12: Negative feedback keeps TES circuits in narrow operating temperature range. Image from [9]

A more detailed overview of SuperCDMS signal readout and cold hardware including mechanical and thermal considerations can be found in [18].

2.7 Testing and Characterization

The ultimate destination for CDMS detectors is the underground lab at Soudan, MN where we operate a dilution refrigerator 2,340 feet below the surface to screen out cosmogenic particles. Before we invest the time and money to commission detectors at Soudan, we need to check that they are working as expected and collect calibration information needed for the final analysis. To do this, SuperCDMS operates test facilities at UC Berkeley, University of Minnesota, Queens University and University of Florida. Creating the testing schedule is a competition between several motivations. On the one hand, we want the best information possible about our detectors and

hardware, so we need to split up the burden of testing between multiple facilities. On the other hand, detectors are fragile and shipping them takes coordination and time.

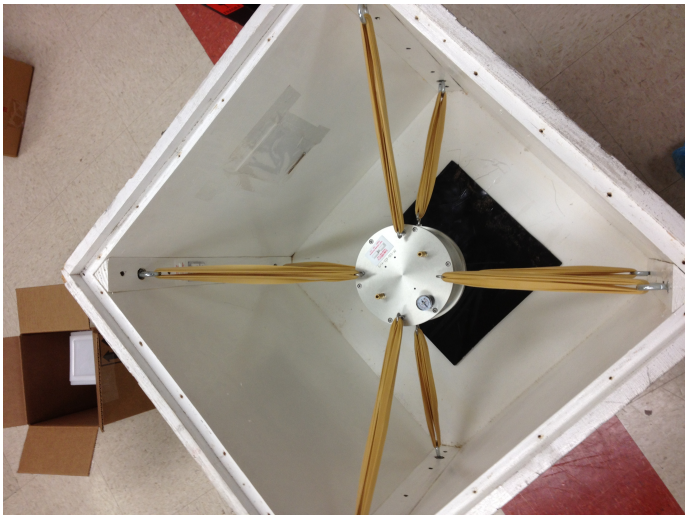


Figure 2.13: Strong elastic bands absorb shocks to detectors in transit. The inner vessel is over-pressurized with dry nitrogen gas to maintain radio-purity between cleanrooms.

For short transits (say, from Stanford to UC Berkeley) the preferred shipping method is to personally deliver detectors to one's counterpart in the other lab. For longer transits, special shipping crates have been constructed to deal with the shocks of commercial ground freight. Figure 2.13 shows a sealed inner container suspended by heavy elastic bands to absorb the shocks of loading in and out of trucks.

Traveling also poses a number of radio-purity threats. As previously mentioned, exposure to radon-laden air will render detectors unusable in a matter of days. This is easily prevented by shipping in containers over-pressurized with dry nitrogen gas. Preventing activation of the germanium and copper by cosmic rays is more difficult. Mostly we are worried about hadrons (protons and neutrons) converting Ge-70 and Ge-72 into Ge-68 and Ge-71, which decay by electron capture into gallium with half-lives of 271 and 11.3 days, respectively. Both decays emit gamma radiation within the bulk of the detectors. Lead blocks external gamma radiation, but is ineffective at stopping high-energy neutrons. Gammas collide inelastically and vanish,

but quark conservation requires that hadrons scatter elastically. Lead is too heavy to absorb much energy in an elastic collision, so hadrons bounce around rather than stopping. Hydrogen-rich plastics are much better shields because the single-proton nuclei can absorb up to half of the energy of a cosmogenic hadron per impact. In RSF, detectors are stored in a sealed cart with continuous nitrogen purge and 4" of high density polyethylene shielding. Even this is not a good solution for the long term, so crystals not being actively worked on are kept in a similar purge cart in the Stanford Underground Facility (SUF) with 25 feet of earth overhead.

Since shipping containers containing multiple feet of polyethylene shielding are impractical, it is important to keep the hadron flux down during shipping. This means CDMS detectors never fly. The hadron flux at cruising altitude is roughly 100 times the flux at sea level, so an hour in the air generates the same activation as a week in a truck. More to the point, the 8-10 hours of round-trip airtime needed to get a detector across the country and back generate the same activation as several months of ground travel. As a result, it takes almost a week to ship a detector from RSF to a remote test facility.

Each test facility has a dilution refrigerator capable of cooling one or more detectors to TES operating temperature (typically < 30 mK) and a copy of the cold electronics used at Soudan to carry signals from the cold stage to room temperature. The first test for every detector is determination of TES critical temperatures (T_c). On cooldown, a 5 μA triangle wave is fed into the bias line of each SQUID circuit (I_b in Figure 2.14). A small current is used so that Joule heating is negligible and the TESs are held at the same temperature as the crystal. Well above T_c, the resulting voltage is a triangle wave with amplitude IR_n where R_n is the normal resistance of the tungsten. Below T_c, the voltage is also a clean triangle wave with amplitude IR_p where R_p is the parasitic resistance in the SQUID circuit. Right around T_c, the waveform is distorted as the TES resistance is current-dependent. By measuring the distortion of the waveform while varying the crystal temperature very slowly, the upper and lower edges of the transition region can be determined with great precision.

Once the transition temperature is measured and the crystal is brought well below T_c, characterization of the TESs continues by measuring bias current versus SQUID

current (I_{bias} vs. I_{SQUID} = IbIs). Since the shunt resistor in parallel with the SQUID input coil is constant (see Figure 2.14), $(I_b - I_S)R_{shunt} = V_{TES}$, making IbIs a form of I-V curve as long as $I_S << I_b$. A little rearrangement gives

$$\frac{I_S}{I_b} = \frac{R_{shunt}}{R_{TES} + R_{para} + R_{shunt}} \tag{2.5}$$

Figure 2.14 shows a typical IbIs curve taken below T_c. At low bias current, the slope of the IbIs line is nearly one. Typically, $R_p \approx 4m\Omega$ and $R_{shunt} \approx 20m\Omega$ so the slope $\frac{I_S}{I_b} \approx 0.8$. Since the entire voltage drop in the TES branch occurs in the parasitic resistor, this state persists until some nucleation event drives a small portion of the TES into the transition edge. The TES 'snaps' to its normal state and a rapid drop in SQUID current leads to a slope dominated by R_n. A typical R_n value of 0.8 Ω makes this slope about 0.05. Whether we decide to bias all channels at a fixed resistance or a fixed fraction of R_n, the IbIs curves will allow us to convert a desired resistance into the appropriate voltage bias.

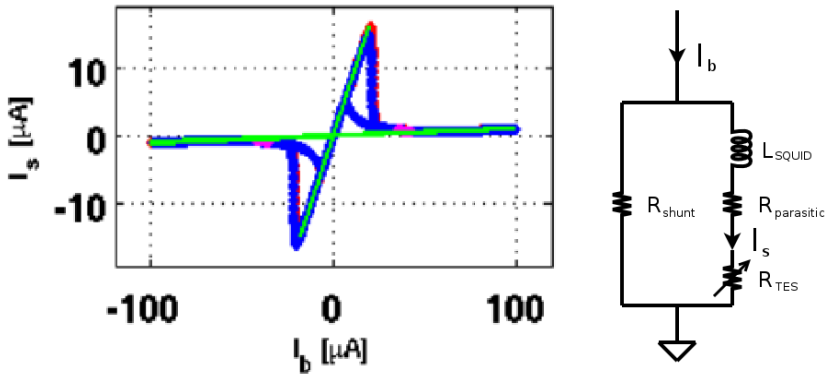


Figure 2.14: Sample Ib-Is curve with circuit schematic. Inductor shown as L_{SQUID} connects readout to SQUID circuit shown in Figure 2.12.

Below 1 K, single-crystal germanium should be a perfect insulator. Our charge collection scheme relies on the crystal in equilibrium having no free carriers of its own. However, as mentioned before, dopants and dislocations can inject carriers into the lattice. These carrier concentrations will be field dependent, so test facilities ramp up the bias on the charge lines slowly. The same circuitry that collects pulse data

can also measure the noise spectrum, at least in the frequencies that are relevant to pulse data. In theory, the noise should be constant with increasing bias. In practice, some detectors have a sudden increase in noise level at a bias point well below 4 V. Usually by the time these detectors reach the full 4 V bias, the noise is bad enough to swamp the pulses. A noise test on startup lets test facilities identify bad detectors before too much time is invested in them.

Once a detector is shown to be fully operable, test facilities can measure its response to different types of signal using radioactive sources. Barium-133 emits gamma radiation (electron recoils only) at 300-400 keV. By the time they penetrate copper vacuum shields, many of these have down-converted into X-rays. This allows for measurement of the electron recoil yield band across a broad energy range, with sufficient spectral resolution left to calibrate absolute energies. Californium-252 emits neutrons by spontaneous fission along with a strong gamma signal. The nuclear recoils from neutron impacts map out the yield regions where WIMP signals will occur. Very few detectors get californium testing because of the increased radio-activation.

Besides mapping out detector performance, test facilities test all the cold hardware that will be used at Soudan. SQUID circuits especially need to be fully characterized to understand their gain and shunt resistance before deployment. All of these tests depend on pulses coming from the detectors, so it is a big waste of time for a test facility to run a detector without full charge and phonon output. If even a single channel is damaged, barium calibration data will be mis-calibrated (i.e. useless) and the missing channel will not get IbIs testing. The repaired detector will have to undergo the full shipping and characterization process again. Several weeks and up to 200 L of liquid helium will have been spent from the time detectors are shipped from RSF to the time a fault is discovered. To build large scale CDMS experiments, a facility near the fabrication center is needed to look for simple problems that prevent test facilities from doing full characterizations.

Chapter 3

Helium-3 Fridge

3.1 Motivation

The He-3 system at Stanford was originally commissioned by SuperCDMS to provide a flexible platform for cryogenic pre-screening. As a part of the fabrication pipeline, its primary tasks are to ensure the quality of the lithographic films and crystal substrates before detectors are sent to dilution-fridge testing facilities for more thorough characterization. The screening facility does not need to reach the transition temperature of tungsten, but it does need to bring multiple detectors below 1 K quickly. He-3 evaporation refrigerators bottom out around 0.25 K but are simple to operate and more reliable than dilution refrigerators.

The Stanford He-3 screening facility is a Chase Cryogenics "He-10" refrigerator (a He-4 stage pre-cooling a pair of He-3 stages) mounted in a 14" 4K cryostat from Infrared Laboratories with a single vacuum system. The system previously housed an infrared telescope weighing 20 lbs, the typical payload of an Oxford 300 or 400 dilution unit. (The system is on extended loan from Sarah Church and the SuZIE collaboration.) It has the payload volume to cool a dozen detectors, but started with only 102 readily usable wires from the 4K work surface to the outside (a pair of MDM 51pin subD connectors). Four wires per phonon channel[1] and one per charge channel (4x4 + 1x2 = 18 wires per DIB) are needed, so that's enough for 5 DIBs. We started

[1] See Section 4.1

by wiring up 4 DIBs through twisted-pair ribbon cables to support 2 iZIPs. Instead of 150-200 liters of liquid helium (at ∼$10-20/L) to cool a comparable dilution unit, the He-3 system uses 35-40 L per run. Instead of a gas handling system with external mix kegs, the He-3 system is self-contained and controlled automatically by a PICS[2] chip.

3.2 Room To Work

Even in testing, radio-cleanliness is essential. For a while, we wheeled the entire cryostat into the anteroom of RSF to keep detectors clean while we installed the radiation shields. The anteroom is not especially clean in terms of dust measurements, but it is in the pressurized escape path for air from RSF so it contains *clean* dust. The radiation shields are sealed against infrared leaks, which also reduces the diffusion of air by enough that it is safe to move the cryostat to a more open space before installing the vacuum jacket. The anteroom was not really big enough to work around the cryostat, so we built a Class 100 cleanroom in the next room with more space (see Figure 3.1).

Two HEPA filters[3] in the ceiling scrub the radon-heavy basement air and maintain positive pressure under the walls made of of plastic strips. The concrete floor is coated in a base adhesive and tiled out to two feet beyond the edge of the screen. The walls are painted with an epoxy resin which absorbs dust. Being able to wheel the cryostat in and out for loading and unloading is an advantage over trying to keep all the dewars and electronics clean. The disadvantage is that care must be taken to prevent the cryostat and cart from carrying particles into the clean space. We also must be careful about spilling liquid nitrogen on the floor of the He-3 room because it will pick up all the dirt in the room and carry it under the screens. Fortunately we only handle liquid nitrogen while the vacuum jacket is sealed, and tile is easy to clean.

[2]Process Instrumentation and Control System

[3]Note to future students: The HEPA filters need to be changed every six months. Jim Perales (jperales@stanford.edu) helped install the clean room and knows how to change the filters.

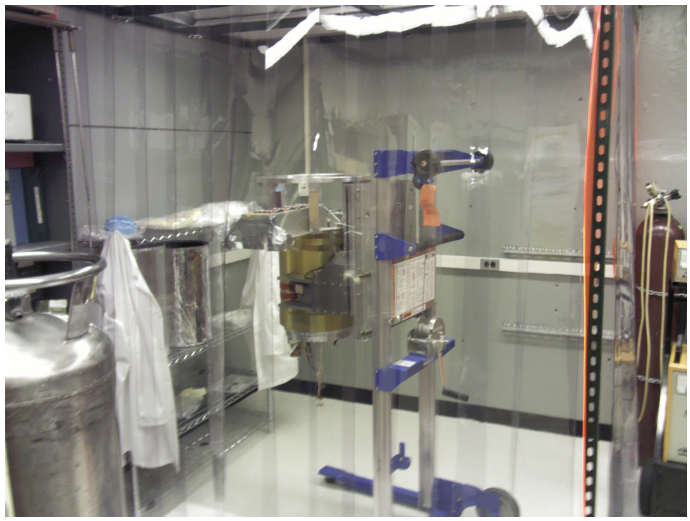


Figure 3.1: A new Class 100 cleanroom in the corner of the He-3 fridge's room makes loading and unloading detectors much more convenient.

3.3 The Work Surface

For a rapid screening facility in a time-sensitive production pipeline, simple is good. A single vacuum system saves us from replacing low temperature seals every run and requires only one pumping line to maintain, but it also means no exchange gas. That means something else must carry heat from the warm payload to the work surface during the initial cool-down and then shut off before the sub-4K stages run. Various active heat switches allow conductivity at the whim of the operator, but they require wiring and often a driver. A gas gap heat switch consists of two concentric cylinders very close in diameter (typically about 1 mm apart). One cylinder touches only the cold plate, the other touches only the payload. Between them is a sealed container of helium (-3 and -4 mix) and a small getter. When warm, the helium carries heat efficiently between the cylinders. When the getter gets down to 15-20 K, it starts to absorb the helium, naturally shutting off conduction. Of course, the stainless steel walls that contain the helium still conduct some heat. The addition of a second gas gap raised the base temperature of the He-3 stage from 385 mK to 415 mK. Given the power versus temperature curve shown in Figure 3.2[19], this corresponds to a heat

leak of about 200 μW. Gas gaps exist with tabs to shunt the heat leak to the He-4 stage, but since the goal of this He-3 fridge is to get below 1 K this was not necessary. Note that extrapolation of the zero-power slope in Figure 3.2 to the power axis gives the irreducible intrinsic heat load for this fridge, about 100 μW in this case. This is the power lost just by attaching the refrigerator to 4 K with no external devices installed.

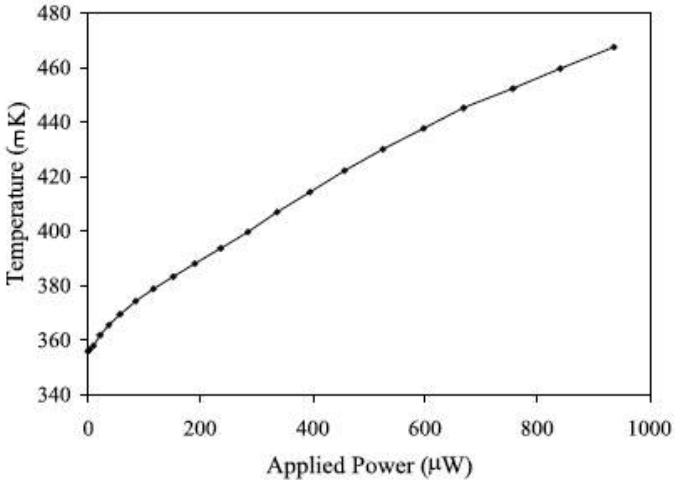


Figure 3.2: Cooling power of the intermediate He-3 stage as a function of temperature taken from [19].

Figure 3.3 shows the He-3 work surface as viewed from below. The banks of wires shown at the bottom of the figure plug into a pair of 51-pin sub-D connectors that go up to the top plate of the vacuum jacket through heat sinks at 4 K and 77 K.[4] After experimenting with various mounting schemes, we installed 4 Hydlar struts to hold the detector payload away from the work surface. This was necessary to keep our hands below the level of the 51-D connectors during installation, preventing accidental breakages. The struts bolt into a 1/2" grid of 4-40 tapped holes in the work surface and have 4-40 blind-tapped holes in their bottom ends. They can simply be unscrewed and moved to different holes to run larger detectors or different-shaped payloads. The original 1/2" diameter struts shown in Figure 3.3 were recently (2013) replaced

[4]In the event of catastrophe, these are part number MWDM1L-51PBRP-.110, but they are expensive and very difficult to find, so be careful with them!

with 1/4" struts with the same 4" length to minimize thermal conduction. Hydlar is essentially nylon with impregnated Kevlar thread, so use the thermal properties of nylon when designing replacement parts.

Just to the right of the top right Hydlar strut, a silicon diode thermometer can be seen epoxied to a bit of copper plate on the work surface. Two similar thermometers hang from wires below the work surface. For standard detector pre-screening, these are best used to monitor the housing temperature and an intermediate point between the fridge and the detectors.

Although full dilution facilities have rigid heat sinks to optimize cooling, we opted to use 0.02" thick by 0.5" wide copper straps to connect objects that we wanted to share heat. The cross-section was empirically chosen to be easily bent by hand while not being the dominant source of thermal impedance. It is important that heat straps have some flex because they thermally contract more than 1 mm during cooldown. If they do not bend they will deform the gas gap switches until the two cylinders touch, creating a thermal short to 4 K. For the same reason, extreme care should be taken to avoid deforming the gas gaps when altering the copper straps. Ideally, the straps should be removed, bent to the correct shape and reattached.

Flexible heat straps and flexible ribbon cables allow the system to run non-detector payloads at a moment's notice. SuperCDMS still does a fair amount of low temperature R&D and many projects require more sophisticated cooling than a helium dunk. Another advantage of the He-3 cryostat over a traditional dilution unit is the enormous payload volume (12" diameter by 22.5" deep). Oddly shaped payloads can be installed without geometric complications. Every time SuperCDMS upgrades to larger crystals, a lot of hardware has to change at the test facilities. While we are qualifying possible new crystals, it is handy to have a screening facility that is a bit more forgiving.

The main SuperCDMS experiment has a custom-made tower designed to hold multiple detectors in the right position and orientation to connect to cold electronics via rigid striplines. Since the dilution-unit test facilities also test cold electronics, they use the full tower assemblies. In the He-3 cryostat this is an unnecessary complication. All that is needed is something to hold the detectors in place that does not impede

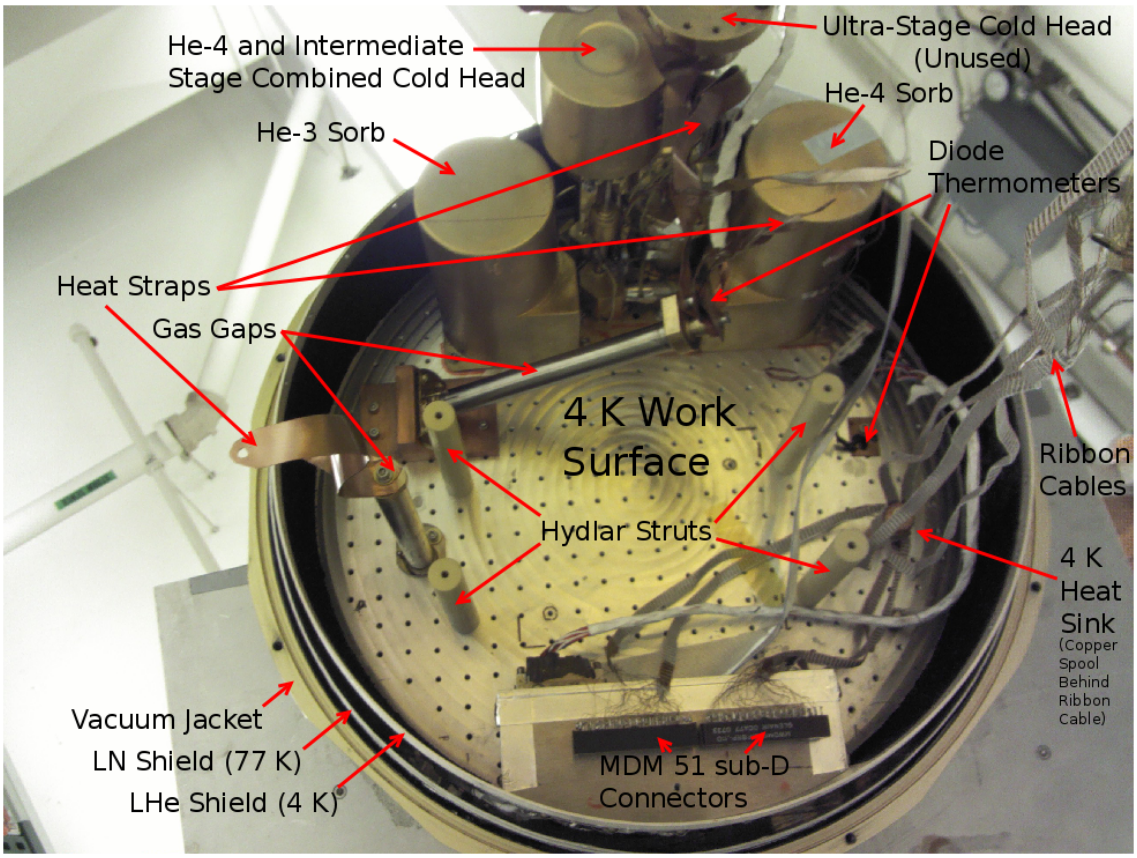


Figure 3.3: He-3 work surface as viewed from below. See text for detailed description of each item. A fiber optic cable has been added since this picture was taken with a pair of white spools bolted to the work surface between the Hydlar struts to take up excess fiber.

the installation of DIBs and heat straps. We decided to re-purpose the circular spacers with hexagonal insets that we already had for shipping detectors between test facilities. We had several of these cut down to expose the DIB faces and drilled a bunch of 4-40 tapped holes to create heat sinking options. The result is affectionately referred to as a 'burger', although in practice it works more like a bun, keeping the detector housings apart while maintaining good thermal contact and holding them in place. Compression is maintained by nuts on four stainless steel 6-32 rods that thread the burgers without touching the housings. The rods continue upward to a stainless mounting plate with through-holes lined up with the Hydlar struts. The

result, known as a 'mini-tower', is shown in Figure 3.4. The mini-tower is assembled in RSF and carried to the He-3 room in a sealed bag, making assembly quick and clean. Figure 3.5 shows a typical mini-tower installation with heat straps to both gas gaps and the He-3 stage.

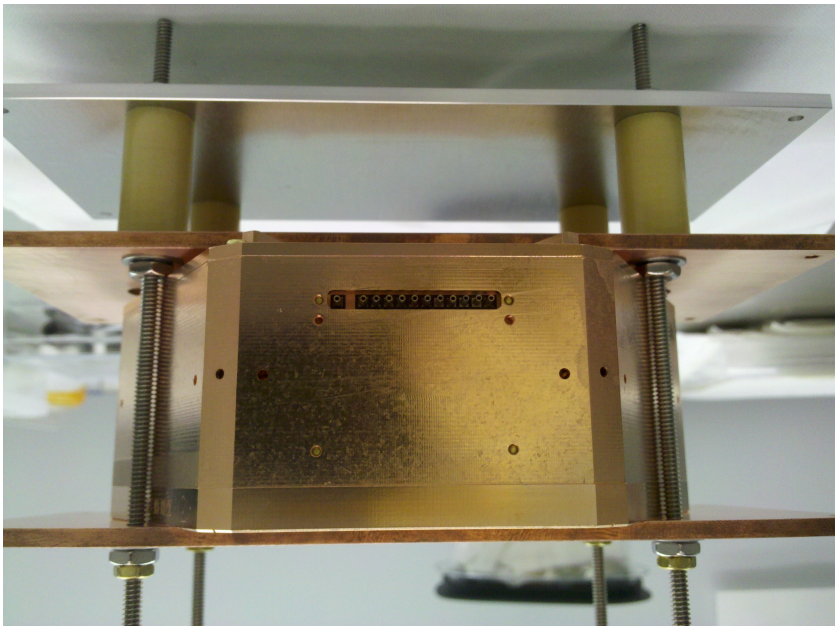


Figure 3.4: Mini-tower waiting for fridge installation. This image is an impromptu assembly for the new (at the time) 100 mm detectors. Mini-tower is shown in installed orientation; it is assembled upside-down on the workbench in RSF.

As shown in Figure 3.3, a diode thermometer monitors the temperature of the work surface. The system came with two extra unattached diodes which can be bolted to the payload at various points, or to multiple payloads if required. Five additional diodes monitor the fridge sorbs and heat switches, but diodes cannot measure below a few Kelvin accurately. To monitor the base temperature, we use a germanium resistance thermometer. All thermometers read out to a SIM 900 mainframe and are recorded once per minute by a program called 'simlogger'[5]. This started out as

[5]Open a new terminal in the 'suzie' login and type 'simlogger'. This will create a file called '/data/sim/sim_yearmoda_000.txt' where yearmoda is today's date in GMT. A 'tail -f ...' of this file on another new terminal will display the latest temperature data to the screen, making you much more likely to spot problems as they occur.

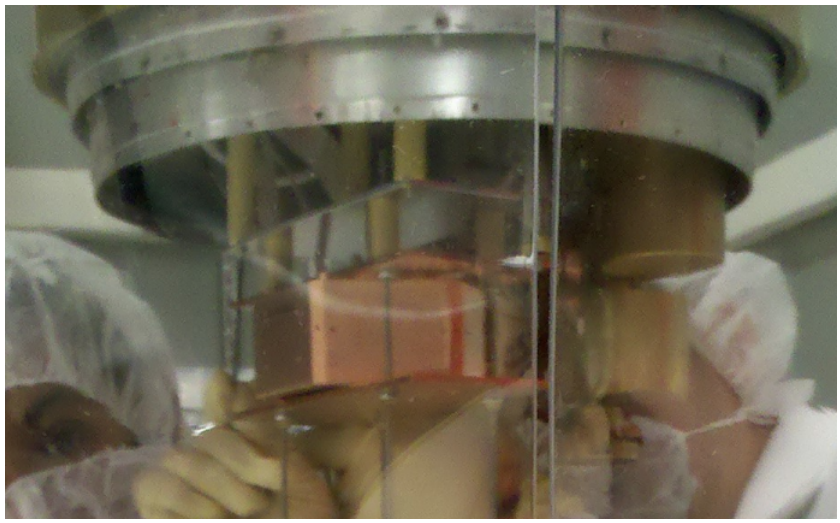


Figure 3.5: Mini-tower with 100 mm detector after fridge installation.

a convenient debugging tool, but it has become increasingly important as work has progressed.

3.4 Cryogenics and Vacuum Systems

The two radiation shields and the outer vacuum jacket labelled in Figure 3.3 represent, starting from the inside, 4 K, 77 K and room temperature isotherms. All three are made of aluminum, which makes not over-tightening the steel bolts very important. The two cold shields are covered on their outside surface with aluminized mylar to reduce the radiation load. The gold-colored coating on the outside of the vacuum jacket is anodization to prevent oxidization. The innermost shield connects to a 9.0 L liquid helium bath directly above the work surface. The outer shield connects to an 8.5 L liquid nitrogen bath located above the helium bath. These bath volumes allow for $\geq$ 24 hour hold time between fills.

As pictured, the top halves of the shields as well as the cryogen cans extend into the top half of the vacuum jacket. The layers are held about ½" apart by permanent G10 (fiberglass) struts in the top of the jacket and aligned by more G10 struts put in during installation. Cryogen fill tubes exit the top of the vacuum jacket, sealed

with Buna-N o-rings compressed against the inside wall. The helium fill tube passes through the liquid nitrogen can, adding an extra source of cooling power from cold helium vapor. To protect against thermal contraction, the helium fill tube has a section of flexible bellows between the top of the nitrogen can and the top of the vacuum jacket.

After mini-tower installation, the bottom half of each shield is bolted on with wide bolts and sealed against infrared radiation leaks. This is necessary to achieve low temperatures, but makes the initial removal of air difficult. The cryostat came with a tiny hole in the side of each can threaded with a winding Teflon tube that lets air out without letting radiation in. This slow leak rate necessitated an overnight pump-out, so we drilled a pair of 1/4" holes in the bottom of each can and covered them with radiation baffles made of aluminized tape . Although not a perfect radiation seal, this had no noticeable effect on cryogen hold time, but the pump-out now takes only 4-6 hours to get below 5 mtorr. It typically takes 20-30 minutes of slowly opening the black valve on the vacuum jacket to rough it out without overwhelming the ACP 40 dry pump. A smaller roughing pump may be used, but it needs to be oil-free to avoid contaminating the turbo pump or coating the reflective surfaces in the cryostat with an absorbent layer of oil. Once the pressure at the thermocouple gauge is below 1 torr, it is safe to start the in-line turbo pump.

The inside of the bottom half of the 4 K shield is coated in Bock black, a carbon-based foam intended to absorb infrared photons. Outgassing from this large surface area is responsible for the high base pressure at room temperature, but as soon as cooling begins the Bock black acts as an absorber, effectively removing any exchange gas.

3.5 Thermal Schematic

Figure 3.6 shows a thermal layout of the He-3 fridge. With the whole system at 4 K, a heater warms the He-4 sorb up to 54 K. This drives off the adsorbed gas, which travels to the He-4 condensation point held at 4 K, condenses into liquid and falls into the He-4 still. Once the still is full, the heater turns off and a heat switch

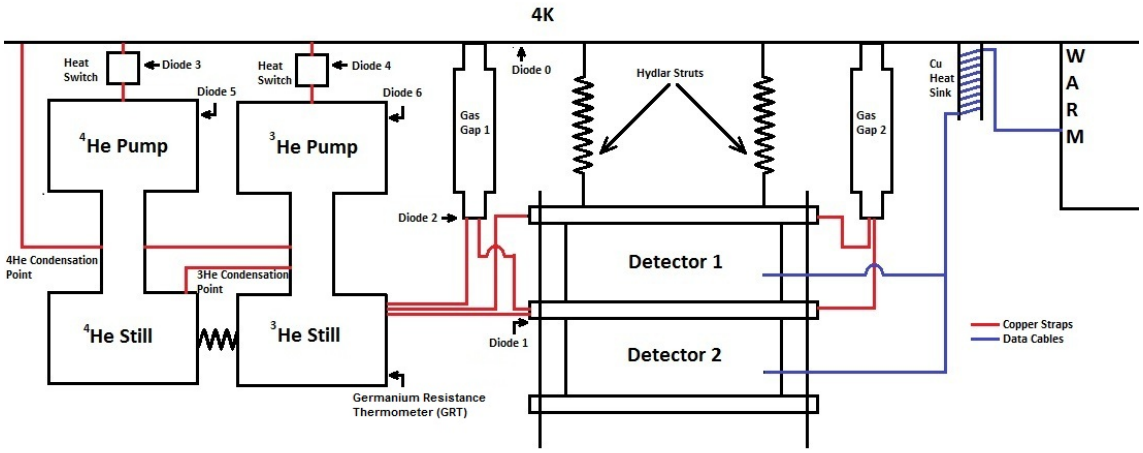


Figure 3.6: Thermal schematic of the He-3 fridge including payload and connections to 4 K

connecting the sorb to the work surface closes. (Both heat switches shown in Figure 3.6 work by the gas gap principle with a small getter loosely heat sunk to 4 K.) The sorb cools, pumping the still down to about 2 K before the He-4 heater kicks in. The 4 K work surface is insufficient to liquefy He-3, so cooling power must be taken from the He-4 still to cool the He-3 condensation point. Although not shown in any schematic, the He-4 still is actually inside the He-3 still. The entire He-4 working fluid must be consumed while filling the He-3 still to allow He-3 cooling to proceed. Once the He-3 still is filled, the He-3 heater turns off and a second heat switch starts the sorb pumping down to a theoretical base temperature of 355 mK. Figure 3.7 shows a typical pattern of diode temperatures for an initial cycle of the He-3 fridge. The heat straps that cooled both the fridge and the detectors from room temperature via the heat switch now provide a link from still to detector. The heat leak from the 'disconnected' gas gap and wiring limits the actual base temperature to about 380 mK. To speed things along and ensure even cooling a pair of additional straps are usually added directly from the still to some other point on the detector housings.

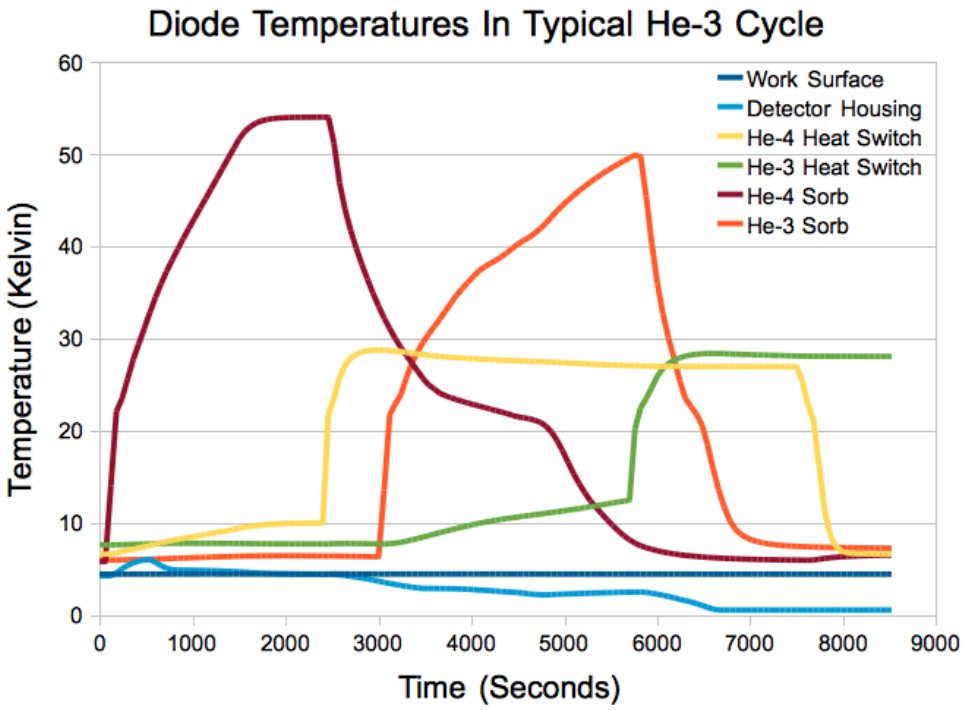


Figure 3.7: Typical pattern of diode temperatures for an initial cycle of the He-3 fridge as used in this thesis. The He-3 heat switch is left on for the duration of the run, but everything else has usually settled back to its coldest state after 2.5 hours.

3.6 Thermal Modeling

A typical He-3 run starts by pumping out the vacuum jacket and filling both cans with liquid nitrogen. Nitrogen is cheap and has high latent heat compared to helium. Ideally we would use nitrogen to cool the entire payload to 77K before adding helium. In practice the work surface and detector housing reaches ∼80-85 K in an overnight pre-cool, but the detector does not. The next morning, the liquid nitrogen is removed from the helium can. The entire work surface is held at 85 K for a few minutes to ensure that all the of the nitrogen is gone before beginning a slow helium transfer. A typical transfer will bring the work surface to 4 K in 4-6 hours. The detector housing takes a long pause at 15-20 K because the gas gaps shut off, typically reaching 4 K early the next morning. Figure 3.8 shows a plot of TES channel resistance versus

housing temperature during cool-down for a typical run.

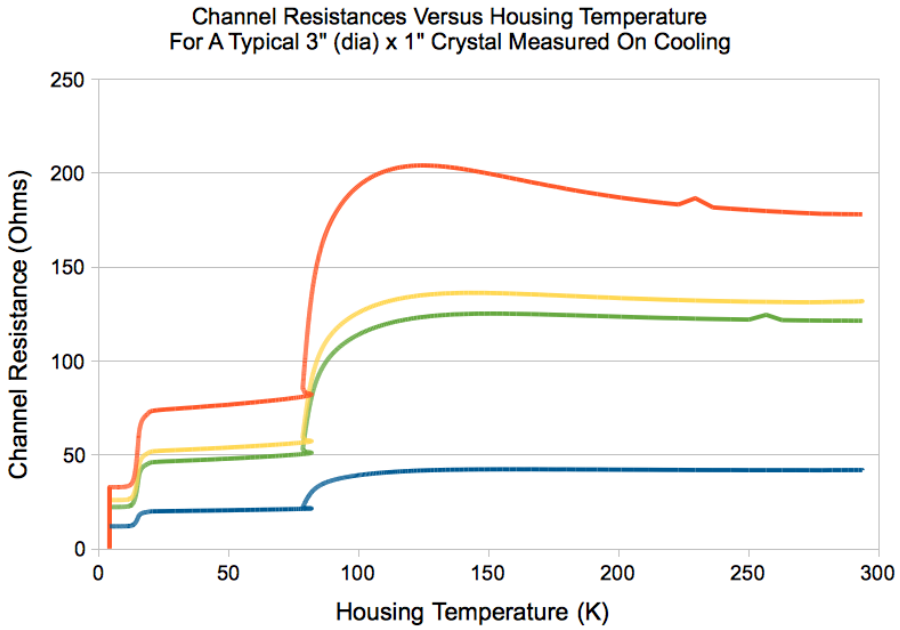


Figure 3.8: Phonon channel resistances plotted against housing temperature for a typical iZIP detector. Data were taken on cooling to accentuate features at 77 K and 15 K.

At room temperature, the germanium substrate contributes substantial conductance between channels. Below about 200 K, the germanium should become more insulating than the metal films and the remaining resistance should be roughly linear with temperature. The observed features of the R(T) curve around 77 K and 15 K correspond to long dwell times at those temperatures. This strongly suggests that the crystal temperature is lagging significantly behind the housing; the crystal changes very little in the regions where the housing is cooling rapidly, only to catch up when the housing temperature stagnates. Since great care has been taken to electrically insulate detectors from the outside world, this is not surprising. To find out how much the crystal lags, we installed a 1" iZIP with an extra 1 cm housing on top. Through the open DIB port of the 1 cm housing, we brought one of the spare diode thermometers and glued it to the crystal surface with GE varnish. This of course destroyed some of the lithography, but the detector was going to be re-polished anyway.

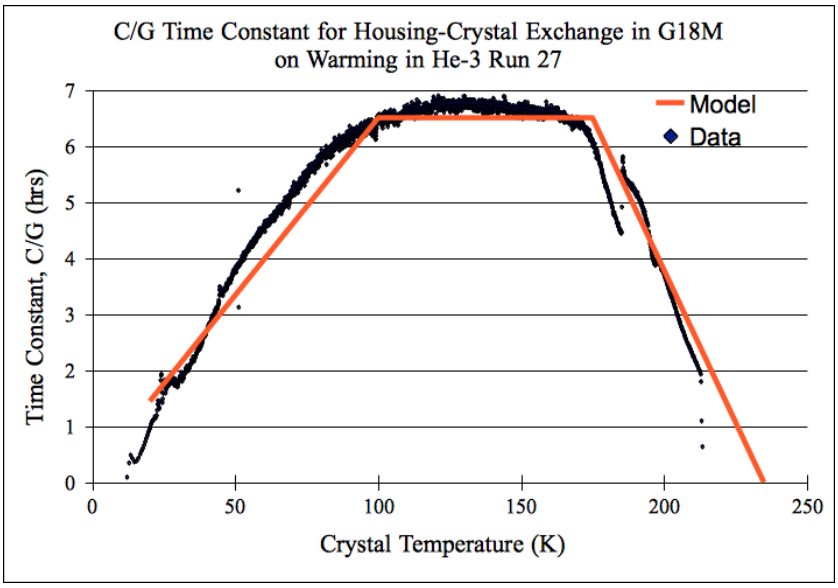


Figure 3.9: Time constant of crystal-housing interaction on warming.

The other spare diode was bolted to the housing. To prevent extraneous radiation leaks, we sealed the open DIB port with aluminized tape. We brought the fridge down to base temperature and warmed it up slowly over the course of two days to prevent thermal lags across the heat straps or burgers from affecting the measurement. We repeated this process twice, once pumping the vacuum jacket with the turbo and once with no pumping. We found that pumping significantly increased the thermal lag, but not pumping led to unpredictable lag. It is likely that outgassing from the Bock black inside the 4 K shield dominates conduction unless it is pumped away. The amount of outgassing depends on the details of the run history, so if the exact temperature of the crystal is important[6], we need to pump on the jacket during warm-ups.[7] The thermal time constant $\tau = {}^{\Delta T}/_{\dot{T}} = {}^{C}/_{G}$ for the run with pumping is shown plotted against crystal temperature in Figure 3.9. Since we only trust the diode thermometers to ± 0.5 K, there is no point in advanced thermal modeling here.

[6]See Section 4.3

[7]Safety tip: To avoid catastrophic vacuum failure, all unsupervised warm-ups should include some form of pumping above $\sim$50 K. A drum heater set to $\sim$ 50°F wrapped around the midline o-ring will prevent it from cracking in the event of such a failure.

We use a simple model of

$$\tau = \begin{cases} \frac{T}{90K}\tau_{max} & 20K < T < 90K \\ \tau_{max} & 90K < T < 170K \\ -\frac{\tau_{max}}{65K}(T - 175K) + \tau_{max} & 170K < T < 235K \end{cases} \quad (3.1)$$

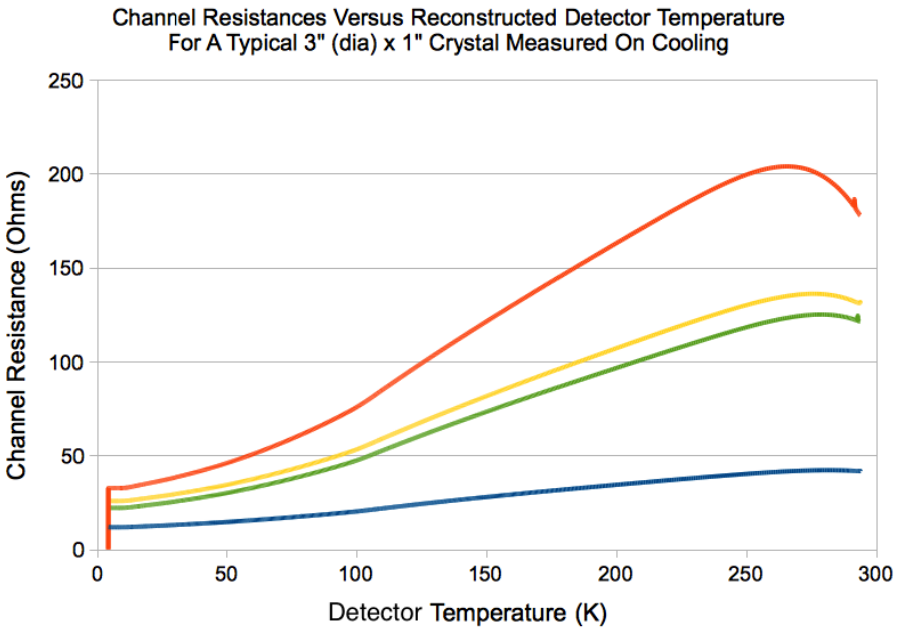


Figure 3.10: Phonon channel resistances from Figure 3.8 plotted against crystal temperature reconstructed using Equation 3.1.

For standard 3" iZIP crystals, the maximum time constant τ_{max} is about 6.5 hours. Figure 3.10 shows the $R_{TES}(T)$ curve from Figure 3.8 adjusted for thermal lag using 6.5 hours as τ_{max}. Note the lack of any features at 77 K in either graph. It is probable that the steep drop-off in thermal lag around 180-200 K is due to radiation ($P \propto T^4$) becoming important, which would invalidate this model for faster or slower warm-ups. The fact that this knee occurs at lower temperatures on cool-down supports such a claim.[8] Since the germanium substrate resistance starts to compete with the resistance of the aluminum at this temperature, it is difficult to make meaningful

[8]Note: The equation given in 3.1 is intended for use while warming the crystal slowly. In general,

detector measurements which need a calibrated temperature above 200 K anyway.

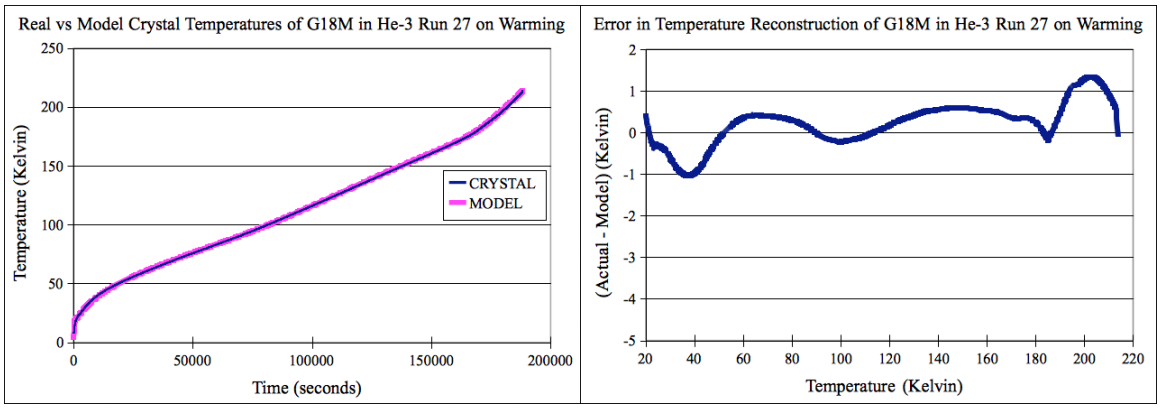


Figure 3.11: Actual crystal temperature plotted alongside temperature reconstructed using the simplistic model in Equation 3.1. The diode thermometers in the He-3 fridge are accurate to about 0.5 K, so there is no point in going beyond this level of sophistication for the study in the 100-180 K range in Chapter 4.

3.7 Data Readout

Data from one or two iZIPs reaches the top of the vacuum jacket in a pair of 55-pin military-style vacuum feed-throughs. Rather than try to guess ahead of time what kinds of measurements we might take and build compact cable bundles, we opted to build a long breakout box out of aluminum plates and send each wire to a separate banana plug. This supports our ethos of flexibility by allowing us to change our measurement schematic or run unusual payloads without re-wiring. One downside of a breakout box with 96 banana plugs is that setting up or changing a multiplexed measurement can be somewhat time-consuming. Figure 3.12 shows the top of the breakout box which was used for most of my time with the He-3 fridge.

Resistance measurements are taken with an LR700 AC resistance bridge. We

cooling happens much faster and less evenly than warming, so it is difficult to make temperature-calibrated measurements on cooling.

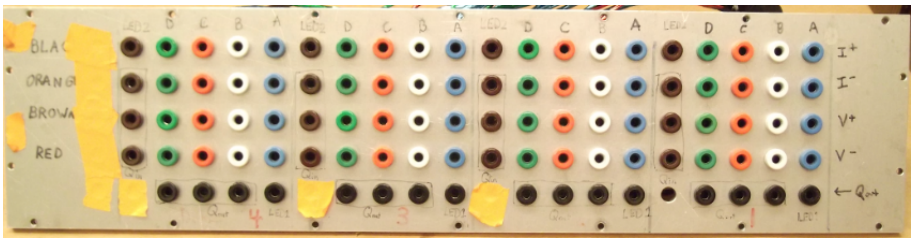


Figure 3.12: Top plate of the banana-plug-based breakout box configured to make four-wire measurements on every channel of four DIBs. Wires are provided for LED and charge channel readout as well. Three wires running to each charge channel simply provide redundancy; they are not necessary for measurements.

could not find a 16 channel multiplexer, but we do have two LR720 8 channel multiplexers.[9] Even at the expected base resistance of 0.8 Ω, a 200 μV bias is enough to get reproducible resistance values with 1 mΩ precision while dissipating only 50 nW in heat. This arrangement was not constructed with high frequencies in mind. Each lead has about 120 Ω resistance and 300 pF capacitance to ground, making measurements faster than 226 nanoseconds impossible ($T_{min} = 2\pi RC$). Ground capacitance dominates the inter-lead capacitance, except that the twisted pair coming from each phonon pin has 300 pF between wires. While we are using banana cables, all high precision measurements must be synced to the power line cycle and the environment kept as quiet as possible.

[9]IMPORTANT: If you inherit this system, do not remove either multiplexer from the rack without first removing the duct tape that seals the gap between them. We had to daisy chain the GPIB chips together by running cables between the boxes. There are no lids under that tape; its just a dust seal.

Chapter 4

Detector Characterization

4.1 Phonon Channel Connectivity

The test for which the He-3 system was originally commissioned, is for continuity faults in the metal films that make up our phonon channels. Prior to using the He-3 system, a newly minted mZIP detector would be put under a microscope in RSF and every single TES and rail would be inspected visually. This process took about three days, left the inspector cross-eyed and still missed the occasional sub-surface continuity break. A full iZIP would presumably take six days. Electrical inspection is impossible at room temperature because the resistance of phonon channels is dominated by the first few centimeters of aluminum rail. However, below 1.2 K the aluminum super-conducts and each channel becomes a collection of thin tungsten resistors in parallel. Testing channel connectivity becomes a 4-wire resistance measurement[1]. A properly connected iZIP phonon channel should have a resistance of $R_n = 0.8\ \Omega$, referred to as a 'normal' resistance because the W TESs are still in their normal state. The two brass Millmax pins in the DIB are inside the four-wire connection. They contribute a parasitic resistance R_p = 5-7 mΩ. The two Al wire-bonds from DIB to detector should be superconducting during R_n measurements.

If a break in a rail leaves some TESs unpowered, the resistance will increase and the amount of increase gives the location of the break. If a piece of aluminum has

[1] If you came here from Section 3.1, *this* is why each phonon pin needs two wires.

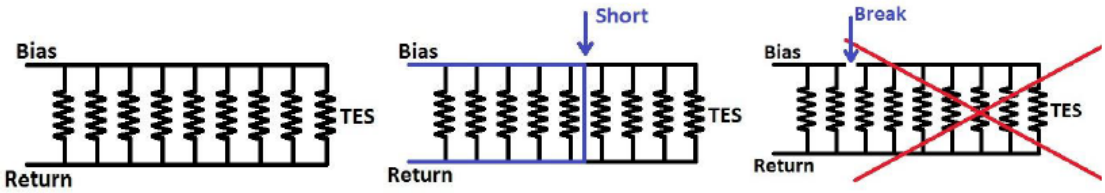


Figure 4.1: Schematic summary of possible problems with phonon channels which can be detected by a sub-Kelvin resistance measurement

escaped etching and shorted the bias and return lines together, the unusable channel will show R_p instead of R_n, a factor of 100 reduction. Figure 4.1 gives a summary of possible lithographic faults.

Since the commissioning of the He-3 fridge, a scanning microscope with $\sim$1 μm resolution was commissioned at SLAC to aid in the visual inspection process. New mask sets with the W top layer 4 μm narrower than the Al rails eased inspection immensely. The W coating is needed for protection of the fairly soft Al, but W is nearly the same color as Ge under the microscope, making defects in the underlying Al very difficult to see. Newer iZIP designs have also become more robust in the past few years. While phonon connectivity screening is still standard, the pass rate has improved and is expected to continue to improve as we gear up for SNOLab.

4.2 Dielectric Stability

Section 4.1 outlines a procedure for spending two days cooling a pair of detectors to 400 mK, thirty minutes taking data and one day warming back up to room temperature. This is faster than the visual inspection process it replaced, but most of the 16-hour hold time of the cold stage is going unused. We cannot access any of the TES properties without a dilution facility, but we should aim to make subsequent work at the characterization facilities as efficient as possible. One thing that really slows down characterization is cooling a detector only to find that it will not sustain a $\pm$2 V bias on opposite charge rails. This has nothing to do with the TESs and so is a perfect study for the He-3 fridge.

In the CDMS-II days, excess noise on biasing had to be caused by current between the two faces, either through the crystal substrate or down the side walls. With the advent of iZIPs, the crystal faces, and therefore the surface preparations, have to support transverse electric fields. It is no longer obvious when a detector breaks down what to do about it. The underlying crystal may be defective or it is possible that something about the polishing or etching process has left a stray conduction path from the charge rails directly to adjacent phonon lines. In the former case, the detector must be abandoned. In the latter, it can be re-polished and re-fabricated.

Because we want to collect charge and phonon energy with high fidelity, our crystals must be very nearly free of scattering centers and traps. As of 2011, CDMS specifications for high purity germanium (HPGe) called for $< 10^{11}$ impurities per cm^3 and < 5000 etch pits per cm^2 (etch pits are caused by linear dislocations coming to the surface). Crystals of this purity may behave very differently from more common doped germanium, so there is also intrinsic interest in studying their properties.

4.2.1 I-V Curves

Characterization facilities detect the onset of charge breakdown by observing increasing noise on their FET circuits as the bias is increased. Since the ^{3}He fridge lacks SQUET cards and the wiring to read them out, this is not an attractive option. Breakdown is a measure of charge injected by the DC bias used in operation, so the simplest possible check would be to apply gradually increasing DC voltage between various channels and measure a DC current. The flexibility of a single-wire breakout box makes this task trivial to arrange. We are free to apply bias only between channels on the same surface if we want to test the quality of the photolithography, or between channels on opposite surfaces to explore the crystal substrate.

The G9F crystal is a good example of the usefulness of I-V curves. G9F was originally fabricated as an mZIPv3 and performed well in CMDSII. During SuperCDMS fabrication, it was removed from the tower, re-polished and fabricated as an iZIPv4 in December 2010. It was again characterized as a good detector but not shipped to

Soudan for the main SuperCDMS experiment. Instead it was given to the new fabrication facility at Texas A&M University (TAMU) when they were just starting up. They wiped it clean and re-fabricated it as an iZIP v4.4, but they had not properly calibrated the correct etching time to remove the aSi layer between channels. Figure 4.2 shows the I-V curves in the ^{3}He fridge from that incarnation. The curves in Figure 4.2a represent bias through the bulk of the crystal. While some conductance is observable, 100 GΩ with no unusual features is a good benchmark for a minimally acceptable detector. However, sharp increases in slope when Side 1 is biased in Figure 4.2b show that large background currents will flow the surface when the detector is subjected to operating bias at a characterization facility. In fact, this had happened the week before at UC Berkeley, but all they could say with their fixed wiring and busy run schedule was that the detector was unusable. Had G9F been a new crystal it might have been discarded. Instead, using several iterations of this data, the TAMU facility was able to fine tune their amorphous silicon etch to consistently produce stable detectors.[20]

Consider also the G101 crystal, the first 100 mm crystal purchased for qualification in the 100 kg-scale SNOLab experiment. In the absence of any ZIP masks designed for 100 mm surfaces, four concentric charge rings of equal area were deposited on Side 1. A single grid electrode was patterned across Side 2 to allow biasing and observation of charge pulses. An accident involving a scanning microscope dragged a steel post across Side 1 shortly after fabrication. Cuts in the outer three aluminum charge lines were easily repaired with wire bonds, but the scratch left crystal damage of unknown depth. Figure 4.3 shows I-V curves taken after the scratch and repair. Voltage between any of the charge channels and the backside grid electrode encountered >100 GΩ, indicating that scratch-induced dislocations almost certainly did not propagate through the crystal. Voltages held across Side 1 between the affected surfaces encountered resistances around 25 MΩ, indicating serious sub-surface damage. The top half of Figure 4.3 shows the G101b mask alongside the G101a mask. The scratch is indicated by the red line. As shown, the scratch should have created leakage between the inner phonon and inner charge channels of the spiral G101b. However, the I-V curve showed leakage resistance greater than 100 GΩ with no breakdown up to $\pm$5 V.

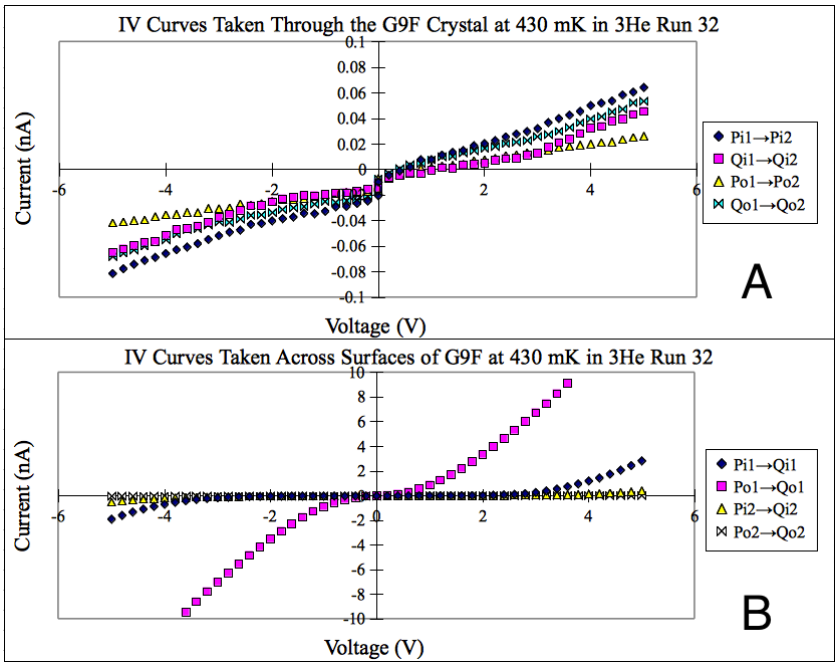


Figure 4.2: I-V curves of detector G9F after experimental fabrication. A) Voltage applied across the bulk of the crystal. B) Voltage applied across one crystal face. Note the 100x increase in scale.

This is strong evidence that the sub-surface damage was removed by polishing even though only $\sim$30 μm of material were removed.

4.2.2 Hardware Limits

Below 1K (86 μeV), HPGe should be a very good insulator because common impurities with energy bands in the 10-300 meV range will be thoroughly frozen out. In theory, I-V curves for good detectors on the 10-pA scale employed above should be horizontal lines. In practice, all of the wires to the different channels must run through the fridge together. The limiting conductance occurs at the 51-pin MDM connectors used to bring the heat-sunk wiring onto the work surface. They typically allow a few hundreds of GΩ between wires, but they are also susceptible to picking up moisture from the air. If left out too long, they may develop leaks ranging from a

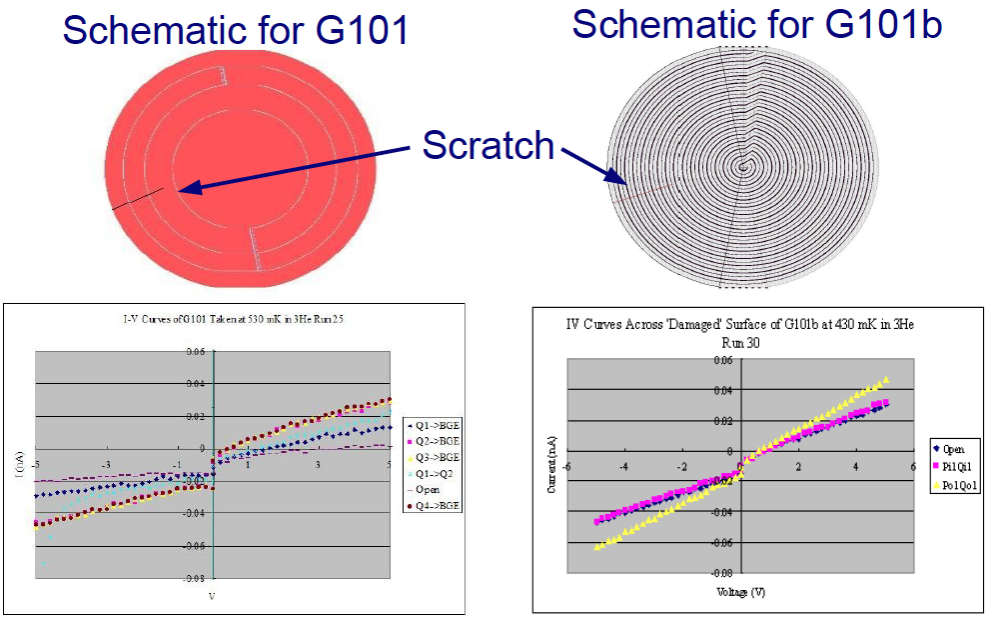


Figure 4.3: Schematic layout of detector G101.

few GΩ down to a few kΩ.[2]

To detect subtle breakdown, it would be useful to measure pico-amp scale currents with a few volt bias. Despite the quoted 300 pA noise floor, the Keithley 2400 sourcemeter in the He-3 room consistently reaches 3 pA resolution when readings are synced to the power line cycle (NPLC = 1), Resolution in this case is determined by taking five consecutive measurements half a second apart and calculating their standard deviation.

To achieve such a low current limit, care must be taken to avoid injecting noise at every stage. Figure 4.4 shows an electrical schematic of a typical I-V measurement. A single 20 Amp circuit is not enough to safely power all the electronics, the computers and the pumps. The pump cart is turned off during base temperature operation, so it makes sense to put it on a separate circuit. To prevent passive ground loops, a

[2]If you are running the fridge and this happens, dismount the connectors and gently bake them with a warm heat gun for a couple minutes. The stray conductance should go away. Do not touch the connectors with your bare hands or break any of the solder joints by handling them too much. If installed properly, the connectors cannot be removed by hand. A screwdriver inserted *carefully* between the connector and the 4K feedthrough will eventually wiggle them free.

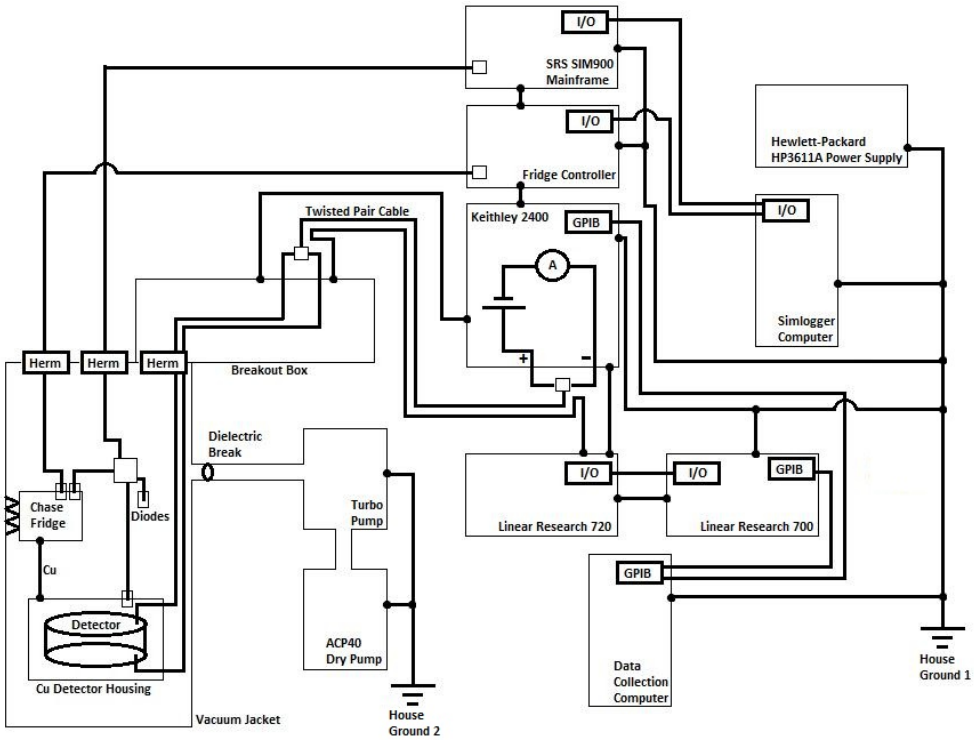


Figure 4.4: Electrical schematic of typical I-V measurement

dielectric break must be included in the pumping line. With the turbo on, magnetic coupling requires the use of a plastic 90-degree elbow joint to avoid the 40 kHz drive frequency. This is typically left in place at base temperature for safety. In theory, the computers should not need to be on the same circuit as the instruments, but when they are separated the daisy-chained GPIB cables become the dominant source of noise.

4.2.3 Hysteresis

So far the pictured I-V curves have been taken by stepping slowly from 0 V to 5 V, usually in steps of 0.2 V, waiting a few seconds and recording a current at each step. The detector is then ramped down to zero again, held there for 10-20 seconds while the surfaces neutralize and then stepped down to -5 V in the same manner. This is necessary when screening new detectors so that the process can be safely aborted

should a dangerously large current appear. However, the procedure implicitly assumes that the crystal has a single, non-history-dependent I-V profile. Figure 4.5 shows what happens when a detector is stepped up to 5 V, then stepped all the way down to -5 V and then back to zero in a discretized triangle wave. Wait times (including the 3 second measurement window) of 6, 13 and 23 seconds are shown.

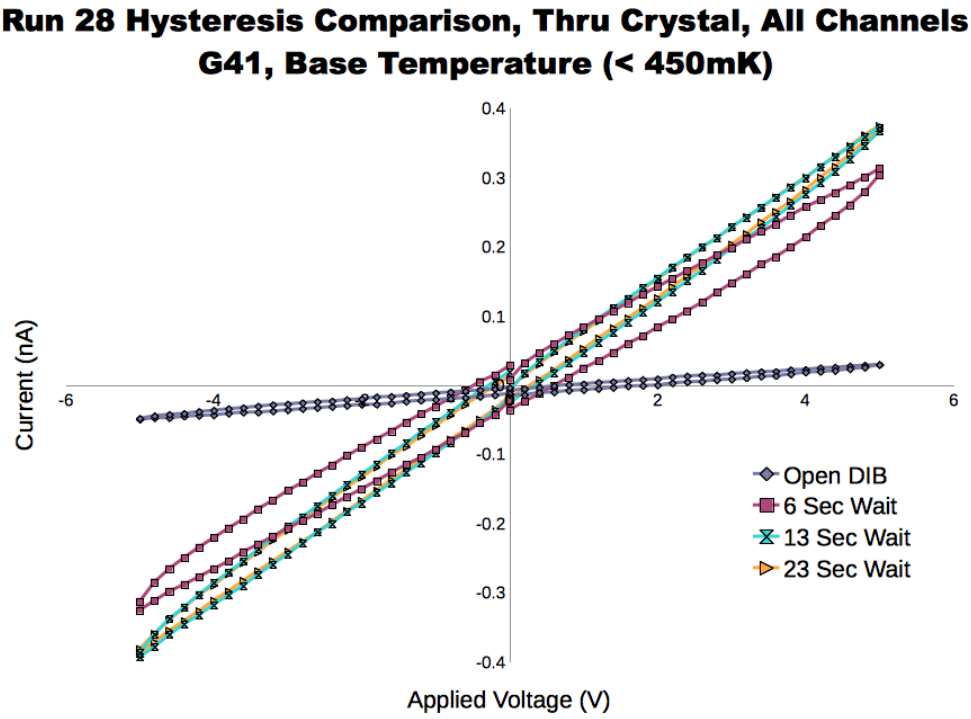


Figure 4.5: I-V curves of a typical iZIP detector showing hysteresis at different speeds. The Open curve was taken with 6 second wait times. The average resistances shown are non-reproducible, but decreasing hysteresis with increasing wait time is consistently observed.

These curves are not precisely reproducible; they change each time the detector is cooled back down from 4 K. Clearly excess current is generated at bias changes which takes some time to die away, but the effect is not well characterized by this experiment. It may eventually be useful to screen for rapid de-neutralization, but for breakdown checks it is irrelevant. Breakdown checks are typically performed with 6 second wait times.

4.3 Resistive Impurity Studies

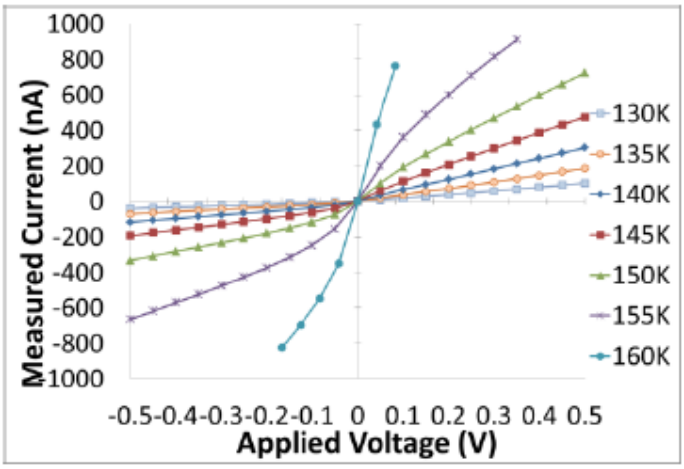


Figure 4.6: I-V curves through the bulk of the first 100 mm test detector (G101a) depicted in Figure 4.3 (top left). Selected curves are from the central charge ring to the backside grid electrode.

Figure 4.6 shows a series of I-V curves taken as a function of temperature for the simplified, 100 mm diameter by 33.3 mm thick, 1.4 kg HPGe R&D detector shown in Figure 4.3. The top face was patterned into four equal-area concentric electrode rings. The bottom face was covered by a single electrode. Data are shown for the central electrode, through the substrate. As expected, a strong temperature dependence was observed.

Common models[21] [22] of the carrier density n in semiconductors use

$$n = 2\left(\frac{2\pi\sqrt{m_e^* m_h^*}k_B T}{h^2}\right)^{3/2} e^{-E_g/2kT} \approx (\text{const})e^{-E_g/2kT}$$

$$R = \frac{1}{\sigma} = \frac{1}{ne\mu} \approx (\text{const})e^{+E_g/2kT} \tag{4.1}$$

for an intrinsic semi-conductor with band gap E_g and

$$n = n_d e^{-E_a/kT} \approx (\text{const})e^{-E_a/kT}$$

$$R = \frac{1}{\sigma} = \frac{1}{ne\mu} \approx (\text{const})e^{+E_a/kT} \tag{4.2}$$

for a semiconductor where conductivity is dominated by impurities with concentration n_d and activation energy E_a. Here μ is the carrier mobility and e is the electron charge. A plot of ln(R) vs. 1/kT (Arrhenius plot[23]) yields a slope with units of energy. This energy should correspond to half the band-gap of Ge or the activation energy of an impurity or defect depending on the temperature. For a more complete model of this particular system, see Appendix A.

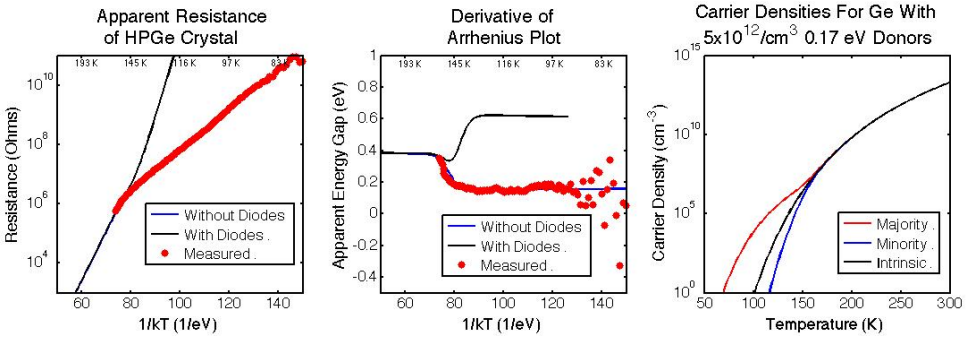


Figure 4.7: Left: Arrhenius plot of I-V curves through an early 100 mm test detector (G102) (red dots) along with model Arrhenius plot assuming 5x10^{12}/cm^3 impurities with 0.17 eV activation energy (blue line, under red dots). For consistency, the positive-going, small-bias derivative of the I-V curve is used for all resistances. Center: 'Energy' vs temperature obtained by taking the derivative of the Arrhenius plot. Right: Calculated carrier densities as a function of temperature.

Figure 4.7 shows the Arrhenius plot for I-V curves similar to those shown in Figure 4.6 (red dots). The conductances used are derivatives of the I(V) function at low bias because impact ionization creates an artificially high carrier density at high bias. Super-imposed is a simulated Arrhenius plot for a crystal with 5x10^{12}/cm^3 impurities with 0.17 eV activation energy (blue line). The black line is a model incorporating the aSi etch-stop layer, which we will discuss in a moment. This impurity density is significantly above our $<$ 1x10^{11}/cm^3 desired limit. However, as shown in the right-hand pane, it would be extremely difficult to detect this kind of contamination above intrinsic levels in common measures of carrier density such as the Hall effect measurements provided by the vendor. This suggests that we need additional tools

to identify deep traps in our crystal stock. However, it is unclear that Arrhenius measurements on already-fabricated detectors provide the information we need. For example, carrier trapping studies (see Section 4.4) suggest that the stock of HPGe in use by Ge-based dark matter searches contains 1x10^7-1x10^8/cm^3 As impurities, which at elevated temperatures form 13 meV donor states. As shown in Figure 4.8, this should result in Arrhenius plots cutting off at < 10 MΩ in this temperature range. So far, this has never been observed.

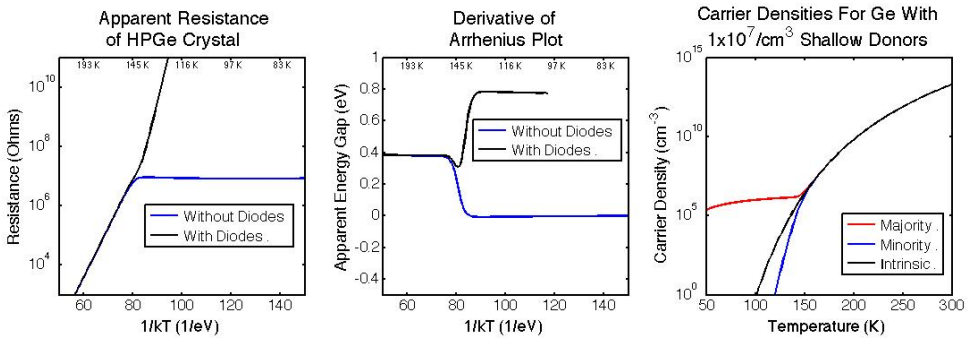


Figure 4.8: Figure 4.7 reproduced with 1x10^7/cm^3 shallow donors. This model is believed to describe most of our HPGe stock and should be readily accessible to our K2400 sourcemeter, yet we have no example for this kind of flat Arrhenius plot or the 0.8 eV slope implied by the diode model.

4.3.1 Dealing With Diode Layers

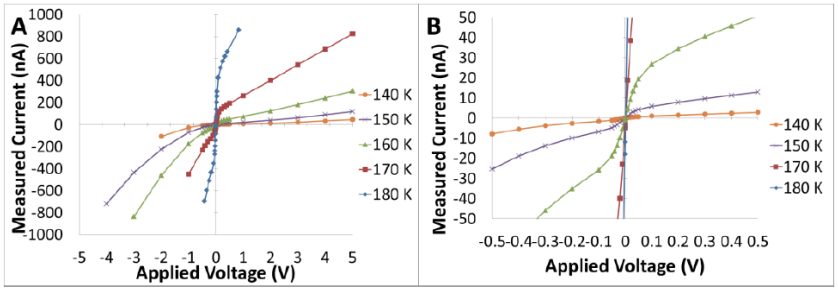


Figure 4.9: I-V curves through the bulk of a typical iZIP detector.

Figure 4.9 shows I-V curves at several temperatures for a prototype SuperCDMS detector of exactly the same design as those bound for the underground experiment. Data taken close to zero bias ($\pm$ <0.5 V) show non-linear physics, perhaps suggestive of a bipolar junction at one or both surfaces. All electrodes on CDMS detectors are underlaid with amorphous silicon, which is doped at 0.65-16 ppm = $3x10^{16}$-$8x10^{17}/cm^3$ with boron (a 45 meV acceptor) to make it conductive enough to sputter. Aluminum contacts on p-type silicon, on (usually) n-type germanium form a p-n diode if the Al-aSi contact is ohmic, or an n-p-n transistor with a floating base if the Al-aSi contact forms a Schottky barrier. A floating-base transistor will simply contribute an unknown current attenuation which can be incorporated as a resistance multiplier. The model which needs careful consideration for Arrhenius I-V measurements is a pair of diodes facing each other with a large resistor in between. This model, addressed in detail in Appendix A, contributes the black lines in the Arrhenius simulations shown in this section.

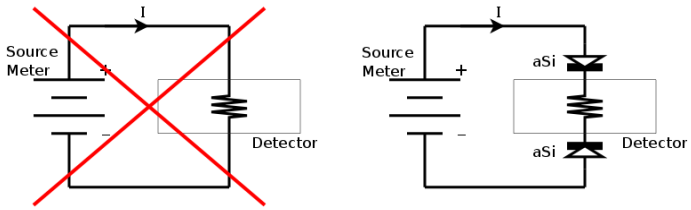


Figure 4.10: Diagram of detector model for understanding Arrhenius data which is addressed in the second half of Appendix A.

In general, the 0.8 eV activation energies predicted by the diode model are not observed. However there are a number of examples of Arrhenius plots like those shown in 4.11 where activation energies suggest intrinsic germanium through the measured temperature range but a current gain of 0.03-0.06 is needed to get the resistance scale right. In fact, all iZIPs where the intrinsic region is visible require a current gain. The two prototype detectors (G101a and G102) we have tested which have grid electrodes on one side do not require this offset.

While we have learned a great deal from resistive studies, it is clear that Arrhenius measurements on finished iZIP detectors are not a good tool for determining the

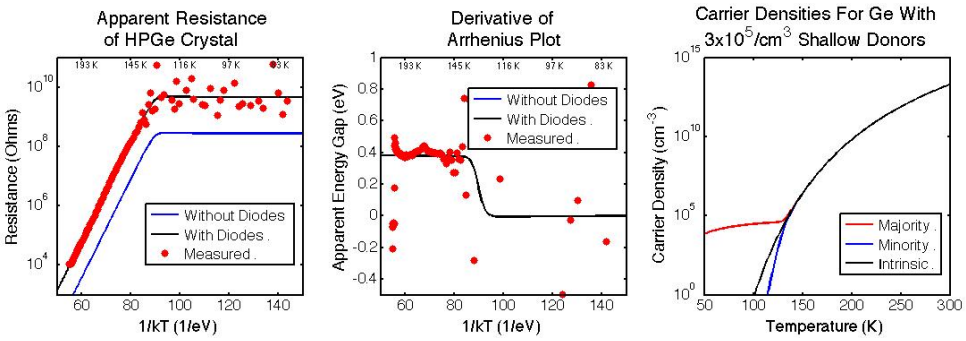


Figure 4.11: Figure 4.7 reproduced with $3x10^5/cm^3$ shallow donors and a current gain of 0.06 due to n-p-n transistors built into the iZIP fabrication. Measured points are from G23R, which was not used in SuperCDMS Soudan. The 1 GΩ limit at low temperatures is the limit of the K2400 in the low-bias measurement scheme being used here. As shown, this limits our sensitivity to shallow traps to $3x10^5/cm^3$.

chemical makeup of the underlying crystals.

4.4 Carrier Trapping

All of the above experiments were performed with the crystals starting in thermodynamic equilibrium. Recall from Section 2.3.1 that in the low temperature CDMS operating mode, electrical neutrality is established over thermodynamic equilibrium by 'baking' detectors with infrared LEDs after they have reached base temperature. Neutralization has been shown to re-set whenever the detector warms above ∼2 K, so it can be performed in the He-3 fridge. A typical neutralization routine requires flashing the LED with 1 mA of current for 100 μsec every 5 msec for 10 minutes with all surfaces grounded. An LED flash onto a neutralized crystal creates a cloud of free carriers on the illuminated surface. By illuminating different surfaces relative to a small applied field, different carriers are transported through the bulk of the crystal, allowing independent study of electron and hole trapping.

The Keithley 2400 sourcemeter (16 Hz measurement frequency, 3 pA resolution) used to make I-V measurements is ideal for studying very shallow traps that thermally emit carriers on a time scale of seconds. Figure 4.12 shows an electrical schematic

of the charge measurement setup. When the LED is flashed, the crystal acts like a current source with capacitive output. A second power supply is used to bias the channels not under measurement to ensure as uniform an electric field as possible through the bulk of the crystal.

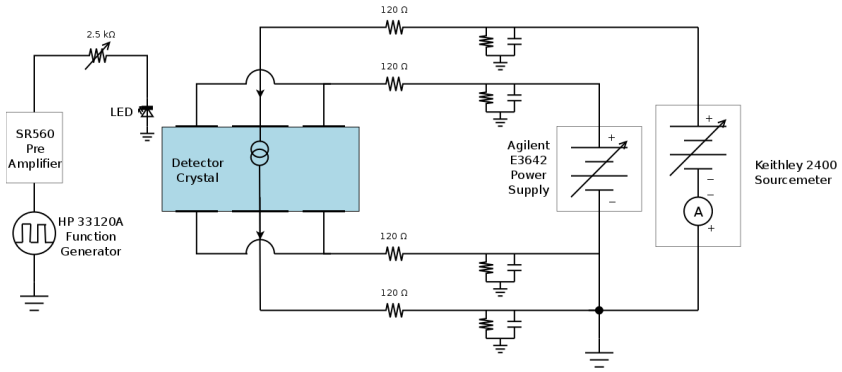


Figure 4.12: Electrical schematic of the LED flashing apparatus. Capacitors and resistors to ground represent current paths entering the measurement circuit through connectors at 4 K and room temperature. A second power supply is used to bias the channels not under measurement. Lead resistance of 120 Ω has negligible impact on current measurements through these >100 GΩ detector crystals. LED pulses were generated by sending 100 μsec long 10 V square pulses from a HP 33120A function generator through a tuned resistor (shown as 2.5 kΩ) to limit the current to 1 mA.

Figure 4.13 shows the current transients for bulk propagation resulting from single LED flashes as a function of germanium substrate bias orientation (top or bottom side positive) and LED position (above or below crystal). Measurable pulses only occurred when the LED on the side of the crystal with positive bias was flashed. This implies that some process was trapping holes and re-emitting them with an emission rate on the order of seconds. (Electrons either propagated in less than the $1/16$ second measurement speed of the sourcemeter or were trapped for longer than the 1 minute trace length.) Emission curves should follow a single exponential ($I \propto e^{-t/\tau_e}$) where $\frac{1}{\tau_e} = \sigma_n v_{th} N_c e^{-E_t/k_B T}$ is the thermally stimulated emission rate of H^+-like states that form so-called 'giant traps' with cross-sections σ_n as high as 10^{10} cm^2 at low temperatures.[25] In the above equation, $v_{th} = \sqrt{\frac{3k_B T}{2m^*}}$ is the thermal

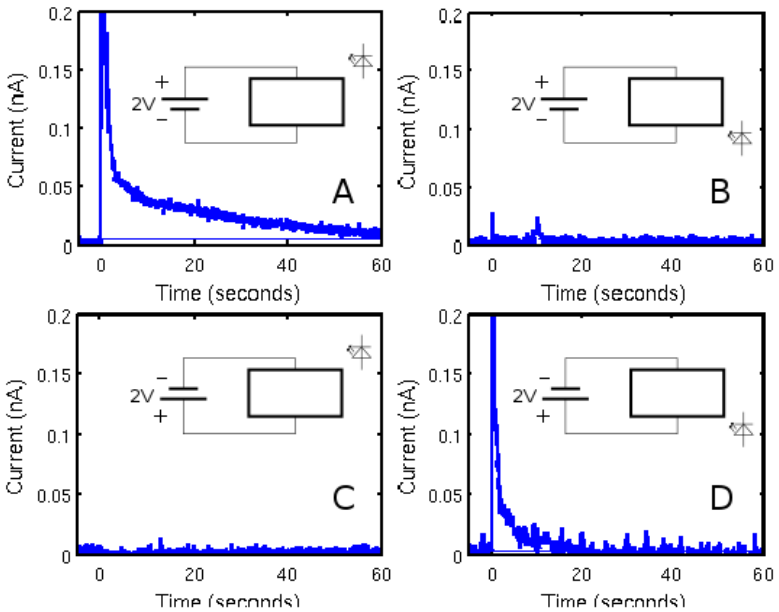


Figure 4.13: Current transients observed from CDMS detectors under various bias and flashing configurations: (A) Top Side +2V, Flash Top LED, (B) Top Side +2V, Flash Bottom LED, (C) Bottom Side +2V, Flash Top LED, (D) Bottom Side +2V, Flash Bottom LED. Measurable transients appear on time scales of 1-100 seconds when detectors are illuminated on the positively-biased face, but not when they are illuminated on the negatively-biased face. The long tail in panel A is believed to be caused by two level systems in the wiring, not semiconductor physics.

velocity, $N_c = 2\left(\frac{m^* k_B T}{\hbar}\right)^{3/2}$ is the effective density of states in the conduction band[3] and E_t is the trapping energy. The shapes of the emission curves found in our He-3 cryostat were temperature-independent in the range 400-460 mK and did not fit a single exponential as seen by Domange [26]. We know from hysteresis in the slow I-V curves that the MDM-51 connectors at 4 Kelvin trap charge and release it on a time scale of seconds. It is probable that the exact shapes of the curves in Figure 4.13 are dominated by this effect, so we are unable to calculate exact emission rates. Similarly, the integrated charge of 0.3-1.5 nC in these curves should be taken as order of magnitude estimates, representing $\sim 10^8$ carriers/cm^3 distributed through the crystal. To reduce the noise from the wiring, only the current flowing between

[3]See Section A.1

inner charge channels of an iZIP were measured. This means only half of the inner volume or 3/8 of the total charge was collected.

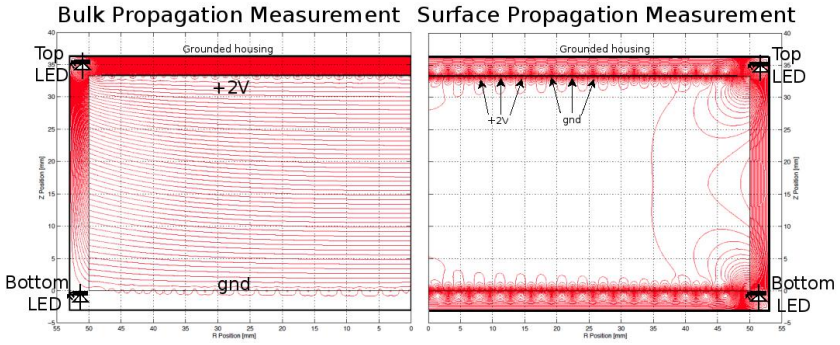


Figure 4.14: Finite element model showing equipotential lines in the through-crystal (left) and surface-propagation (right) current measurement schemes. The housing, held at chassis ground in the fridge, is included as a cylindrical wall. In the surface propagation model, electric field strength is exponentially attenuated away from the surfaces of the crystal. According to basic models [28], charges liberated e.g. on the bottom face in the surface propagation configuration should never reach the measurement channels on the top face.

Instead of fixing each face at a constant potential, the other obvious thing to try with an iZIP is to apply electric fields parallel to each face as shown in Figure 4.14 (right). Because the interleaved electrodes are separated by ∼1.4 mm, which is much smaller than the 25 mm thickness of the crystal, there should be negligible propagation through the bulk if holes are traveling along electric field lines. An LED flash on the opposite face from the measurement should therefore produce minimal detected current[28]. Instead, we observed that flashes on both sides of the crystal produced pulses comparable to those seen in Figure 4.13 when the positively-biased face was illuminated. (See Figure 4.15) The large transient seen in Figure 4.15c is the result of an error in LED illumination settings. The time-scale of charge release combined with the previous experiment indicates temporary trapping and re-emission of holes, but the lack of bulk field suggests that holes are propagating by diffusing through the bulk of the crystal either before or after trapping. This is consistent with previous studies which showed that neutralization propagates away from the LED

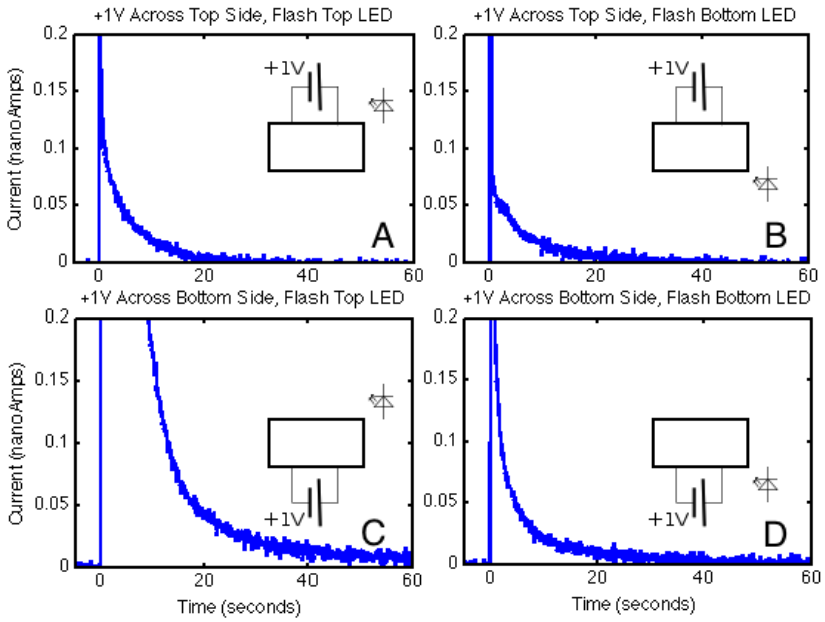


Figure 4.15: Current transients observed from CDMS detectors under various lateral bias and flashing configurations: (A) Measure Top Side, Flash Top LED, (B) Measure Top Side, Flash Bottom LED, (C) Measure Bottom Side, Flash Top LED, (D) Measure Bottom Side, Flash Bottom LED. Transients comparable to those seen in Figure 4.13 appear in all configurations. The large signal in panel C is due to more intense illumination.

like burning away a fog rather than the whole crystal neutralizing uniformly. In the absence of an electric field, this suggests that diffusion is a major contributor to free carrier propagation in the neutralized region but that charges are rapidly trapped in the charged region.

4.5 Neutralized I-V Curves

In the standard prescreening procedure, detectors are not neutralized before I-V characterization because it is time-consuming and unnecessary. Currents exceeding 1 nA rapidly restore trap populations to equilibrium, so any avalanche-like breakdown will deneutralize the crystal anyway. However when a detector **is** neutralized prior to I-V testing, it is possible to see charges moving around in response to the initial bias.

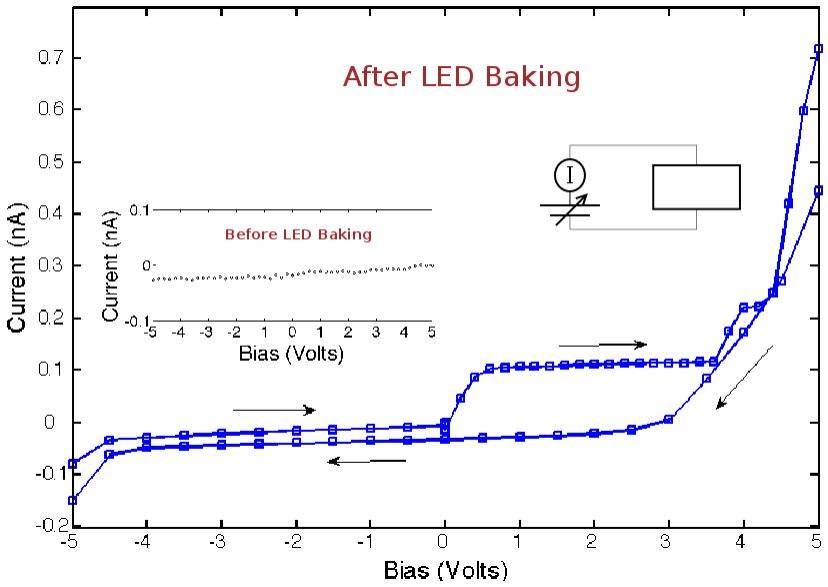


Figure 4.16: I - V curves of an iZIP detector taken at 420 mK before (inset) and shortly after LED neutralization for 90 minutes with average luminosity of 2.75 x 10^{12} photons/cm^2/sec. This detector is known to be a functional particle detector when cooled to CDMS operating temperatures of < 50 mK. Before neutralization at 420 mK, the substrate acted like a >150 GΩ resistor. Substantial charge movement was observed upon bias in the first I - V sweep after neutralization. Subsequent I - V sweeps showed that the device returned to its >150 GΩ state.

Figure 4.16 shows I - V curves taken over a period of 10 minutes through the bulk of the CDMS iZIP detector before (inset) and after neutralization. The I - V measurement steps were 0.2 V every 6 seconds. The I - V curve after LED neutralization (main frame) starts at zero with increasing voltage and proceeds clockwise around the diagram ($0V \rightarrow 5V \rightarrow 0V \rightarrow -5V \rightarrow 0V$ with each leg taking 2.5 minutes). However, when the voltage input is reversed ($0V \rightarrow -5V \rightarrow 0V \rightarrow 5V \rightarrow 0V$), the hysteresis loop appears on the negative side, as though Figure 4.16 were rotated 180°. This suggests that the excess current is caused by charges getting separated from neutral impurities under bias. The fact that excess current is observed for the duration of the first voltage sweep after neutralization but not in subsequent sweeps indicates that these charges contribute to significant trapping on a time scale of a few minutes. At this speed, the excess charge in these post-neutralization I - V curves

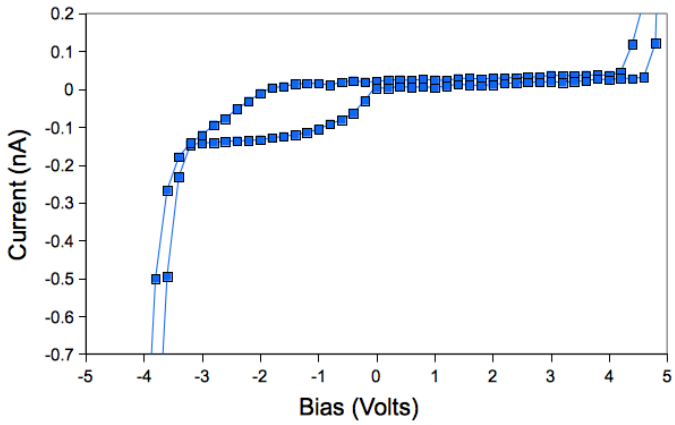


Figure 4.17: Neutralized I - V curve taken under the same conditions as Figure 4.16 but starting with negative biases. This run showed stronger breakdown at high bias than Figure 4.16 but otherwise appears to be a 180 ° reversal.

integrates to 15-20 nC. For a 3" dia x 1" thick crystal, this works out to ∼few x 10^9 cm^{-3} charges. Since the crystal substrate contains ∼10^{10} cm^{-3} impurities, this should go some way toward restoring thermodynamic equilibrium, but is not enough current to completely de-neutralize the detector. In fact, when later used as a particle detector, this detector remained neutralized for more than one hour at 4 Volts bias. It even performed well for short periods at biases as high as 9 Volts. From this, it is apparent that the detector state we think of as 'neutralized' in fact is characterized by ∼ 10% of the chemical traps being charged and that this does not have a substantial effect on detector performance.

4.6 Next Steps

The He-3 fridge at Stanford has proved a very effective platform for pre-screening detectors close to the fabrication facility. As SuperCDMS scales up for the SNOLab project, the original SuZIE fridge has inadequate wiring to test 150-200 100 mm detectors in the time required. A Chase 'He-7' (one He-4 stage, one He-3 stage) is being added to a cryogen-free cryostat at SLAC with 0.5 W at 4 K. Figure 4.18 shows the planned detector testing and characterization schedule for the SNOLab project.

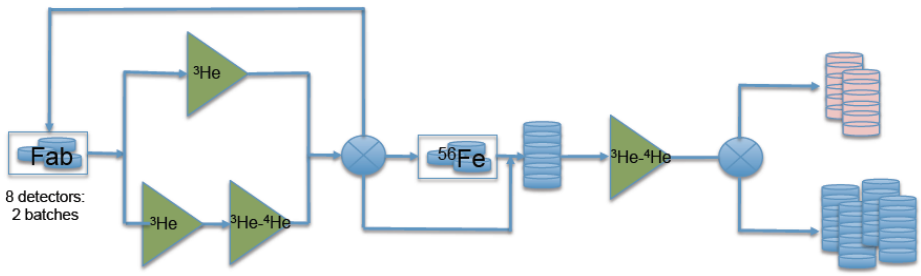


Figure 4.18: Detector testing schedule lifted from Figure 14 of the SuperCDMS G2 proposal. Detectors fabricated in batches of four will all pass through He-3 prescreening. Only one from each batch will be tested in dilution facilities to gauge the critical temperature (T_c) and Ib-Is characteristics. If necessary, the T_cs of the whole batch will be adjusted by ion implantation of iron atoms. When a tower of six detectors has been assembled, it will pass through final inspection at the UC Berkeley 400 μW dilution facility to test not only the detectors, but integration with cold hardware.

Resistive connectivity checks and un-neutralized I-V characterization in the SLAC He-3 system have become standard for all detectors, allowing faster feedback with fabrication and repair as we scale production up by an order of magnitude.

Temperature-dependent resistance studies of germanium crystals have proved not to be a sensitive probe of impurity concentrations below $10^{11}/\text{cm}^3$ after those crystals have been fabricated into detectors. The Hall effect tests provided by the vendor give some idea to the impurity concentration, but no species information. To finally nail down impurity and defect information well enough that it can be correlated to detector performance, SuperCDMS needs to run Deep Level Transient Spectroscopy (DLTS) on a number of witness samples cut from the same boules as its detector crystals. DLTS measures capacitive transients that encode information about the carriers in the depletion region behind a Schottky barrier. Not only concentration, but also band energy and capture cross-sections can be determined, allowing species identification.

Chapter 5

Modeling of Quasiparticle Devices

After a particle interaction in the bulk of a detector, hot phonons bounce around ballistically and more or less isotropically until they are captured or down-convert into thermal phonons. The energy resolution of a thermal sensor is proportional to its heat capacity (and therefore volume), but we want to capture as many particle-induced phonons as possible before they decay. As noted in Section 2.2, these competing design criteria can be partially reconciled by using superconducting aluminum fins splayed out from each TES as phonon absorbers. Phonons in excess of twice the band gap (2Δ) break Cooper pairs while lower energies are reflected. Free quasiparticles which diffuse to the Al-W interface fall down the potential well into the lower gap tungsten, producing heat. However, quasiparticles which recombine before encountering the interface or are trapped in shallow traps caused by gap variations in the Al film are lost.

In an effort to measure properties of the aluminum films produced by SuperCDMS, Jeff Yen et. al.[27] deposited aluminum films of varying lengths between a pair of 250 μm square tungsten TESs on a silicon substrate[1]. To study quasiparticle diffusion in the 300 nm thick Al films, we need low energy x-rays that are absorbed in the Al with high efficiency. We use the 2.62 keV Cl K_α xrays which fluoresce from NaCl illuminated with an ^{55}Fe source. These produce near- instantaneous energy

[1]Jeff was not the first to commission such a device, but all modeling work in this chapter references his samples unless otherwise noted.

depositions in the Al which probe the impulse response of the Al-W-Si system. Figure 5.1 shows a schematic of the quasiparticle device.

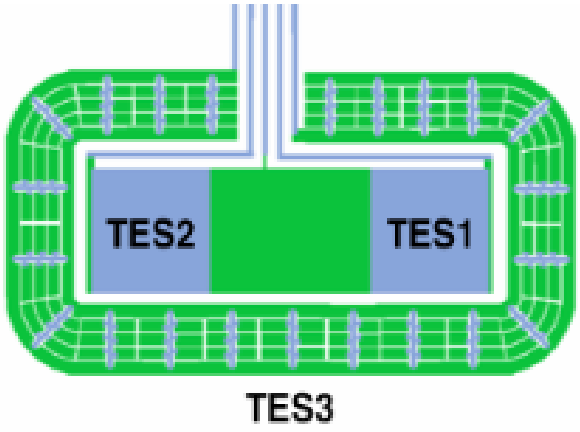


Figure 5.1: Schematic drawing of quasi-particle device.

TES1 and TES2 lay at opposite ends of an aluminum film with overlap regions to collect energy just like a regular detector. Like standard CDMS detectors, all W films are 40 nm thick and all Al films are 300 nm thick. Different devices have aluminum films of different lengths to remove device properties from calculations of material properties such as diffusion length. TES3 is distributed into a guard ring to measure how much energy escaped (or was deposited) into the silicon substrate. The negatively-biased leads on TES1 and TES2 draw current because the lead connected to the aluminum 'fin' is grounded. The leads extend as a superconducting Al rail with W overlap along the back side of each TES, ensuring uniform current flow and uniform electro-thermal feedback (ETF).

Figure 5.2 shows the sealed source, fluorescent crystal and collimators used to direct X-rays onto the quasiparticle device. ^{55}Fe decays by electron capture, leaving ^{55}Mn which emits a 5.90 keV x-ray. Collimated gammas from the sealed 10 μCi source strike the poly-crystalline NaCl sample. Another collimator over the sample restricts fluorescent X-rays to (mostly) striking the quasiparticle device. The most common energies to pass the second collimator are 2.62 keV Cl K_α emission and 5.90 keV reflected Mn K_α x-rays from the source, although 2.82 keV Cl K_β emission and

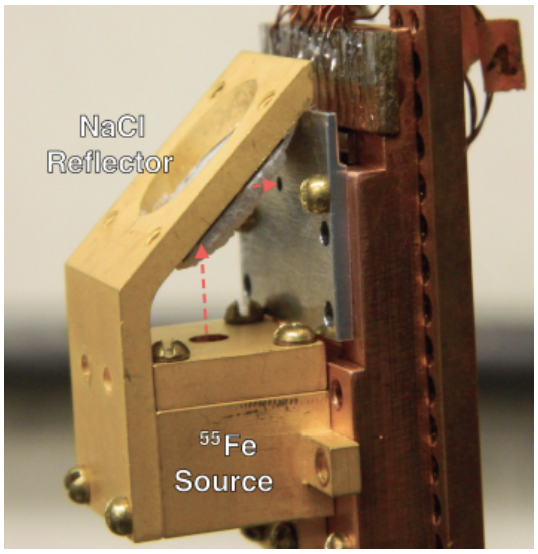


Figure 5.2: Picture of sealed source apparatus used to illuminate the quasiparticle device with collimated X-rays

1.04 keV Na K_α emission also make noticeable contributions. The absorption length for 2.62 keV x-rays in 300 nm thick Al is 3.29 μm, so that 8.7% are absorbed in the Al and the majority in the Si substrate with an absorption length of 3.06 μm.

In order to keep a TES in its transition edge, negative feedback is needed to maintain Joule heating equal to the power loss to the substrate:

$$\begin{aligned} P_{Joule} &= P_{substrate} \\ I_{TES}^2 R_{TES} &= (\kappa T_{TES}^n - \kappa T_{sub}^n) \end{aligned} \tag{5.1}$$

where κ and n define the limiting energy loss mechanism. In the case of tungsten below 100 mK, the limiting mechanism is electron-phonon coupling, with n=5.[29] Negative feedback also speeds up the return of a perturbed TES to its quiescent state. In effect, the reduction in Joule heating causes the "equilibrium temperature" to drop as the electron temperature rises. The characteristic recovery time in electro-thermal feedback is

$$\tau_{ETF} = \frac{\tau_0}{1 + {}^{\alpha}/_{n}} \tag{5.2}$$

where τ_0 is the usual ${}^{C}/_{G}$ time constant and $\alpha \equiv \frac{\partial(logR)}{\partial(logT)} = \frac{T}{R}\frac{\partial R}{\partial T}$ is the steepness parameter for the superconducting transition curve[15]. Since the TES begins and ends in the same state, in the limit of small energy deposition and near-constant T_{TES}, the energy deposited is simply the decrease in Joule heating integrated over the pulse.

In practice, the finite increase in temperature during a pulse is going to make such estimates systematically low. This is especially true if TESs are optimized to have minimal heat capacity so that direct X-ray hits on the TES drive it above its transition curve; in this case ETF is lost and the dwell time at high temperature may be quite long. Additionally, there is a competition between the desire to collect the entire pulse energy by taking longer traces and the $\sqrt{t}$ increase in uncertainty of any quantity reconstructed by integrating noisy traces. Template matching to simulated pulses eliminates both of these problems.

5.1 Working Model Of Quasiparticle Device

A naive current model of the quasiparticle measuring device would involve a uniform sheet of current flowing from the Al to the square of W, which warms as a single lump element. For small energies, this would produce single exponential pulses as the excess cooling power is linearly proportional to the deviation from equilibrium. This situation is shown in Figure 5.3a along with data from an actual device. In 2005, Pyle et. al.[30] showed that the sharp initial spikes are a result of the fast but non-instantaneous conduction of heat across the TES. In their model, a TES device was divided into strips along its length. The conduction between strips was determined using the measured TES normal-state resistance and the Weidemann-Franz Law

$$\frac{\kappa}{\sigma} = LT \tag{5.3}$$

which gives the ratio of the thermal conductivity κ to the electrical conductivity σ for a

metal at temperature T. The Weidemann-Franz Law holds whenever phonons make a negligible contribution to thermal transport, meaning electron dynamics govern both thermal and electrical transport. The Lorenz number L was theoretically derived by Sommerfeld[31] to be $L_0 = \frac{\pi^2 k_B^2}{3q^2} = 2.44 \text{ x } 10^{-8} \frac{W\Omega}{K^2}$. However later studies[32] revealed that the measured Lorenz number deviates by up to a factor for two for many metals, especially at low temperatures. Although tungsten was measured to hold to the Sommerfeld value down to 4 Kelvin, we will leave the Lorenz number as a parameter to be fit for our devices at <100 mK.

The simple TES model adjusted for non-instantaneous 1-D propagation is shown along with a sample pulse in Figure 5.3b. On examination of SEM images, Yen et. al. found that the tungsten connecting the proximitized Al-W overlap region to the TES below was not well connected. For the particular films examined, we estimate that the connecting region has 2.7% the cross-section of the TES, but it has been found that poor draping of tungsten down the side of a thicker aluminum film is common in sputtered CDMS depositions. This small constriction increases the current density, substantially lowering the critical current of the film in that region. This creates a small normal region which acts as a heater, allowing the TES to have quiescent states below the steep part of the transition curve without going fully superconducting. The reduction in electro-thermal feedback results in increased fall times as the TES approaches its quiescent state. This situation is shown in Figure 5.3c.

5.1.1 Early Success Of The Weak-Link Model

Figure 5.4 shows an SEM image of an earlier-generation quasiparticle device fabricated in 1997. It is clear that W has fallen off the edge of the Al fin in strands rather than draping down to the TES to make good contact. The bottom pane of Figure 5.4 shows a model used to create the fit shown in Figure 5.5 with weak links at the fin/TES and TES/rail interfaces. A weaker link at the rail creates an excess heat load far from the energy input coming from the fin. X-ray energy is added to the coldest part of the TES, creating a temperature inversion which replicates the sharp spike and oscillation behavior seen in the pulse. Only one of the two TESs exhibited

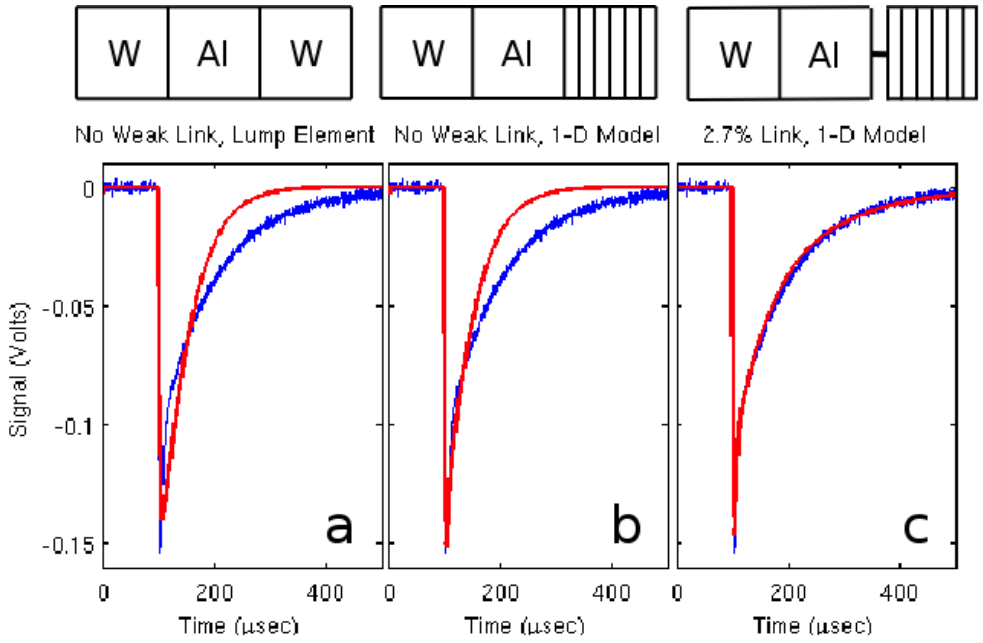


Figure 5.3: Simulated pulses (red) plotted against the same data (blue) for A) a naive lump element model, B) a 1-D model and C) a 1-D model with a weak link at the aluminum-tungsten interface

this behavior. The other is well modeled by a single weak link at the fin. The W film for this device appears to have been deposited at a slight angle, forming shadows in the lee of the taller Al films. While the two-weak-link model can be fine-tuned to reproduce any pulse, small changes in link strength can lead to large changes in pulse shape. It is clear that to move forward into energy reconstructions with confidence we need to improve our lithography to minimize the effect of poor W draping. Fortunately, recent devices have not required such fine-tuning.

5.1.2 Heat Capacity

Since pulses are linked to device temperature, energy reconstructions are potentially sensitive to the heat capacity of the tungsten composing the TES. We will adopt the BCS form of the heat capacity for the normal and superconducting state[33]

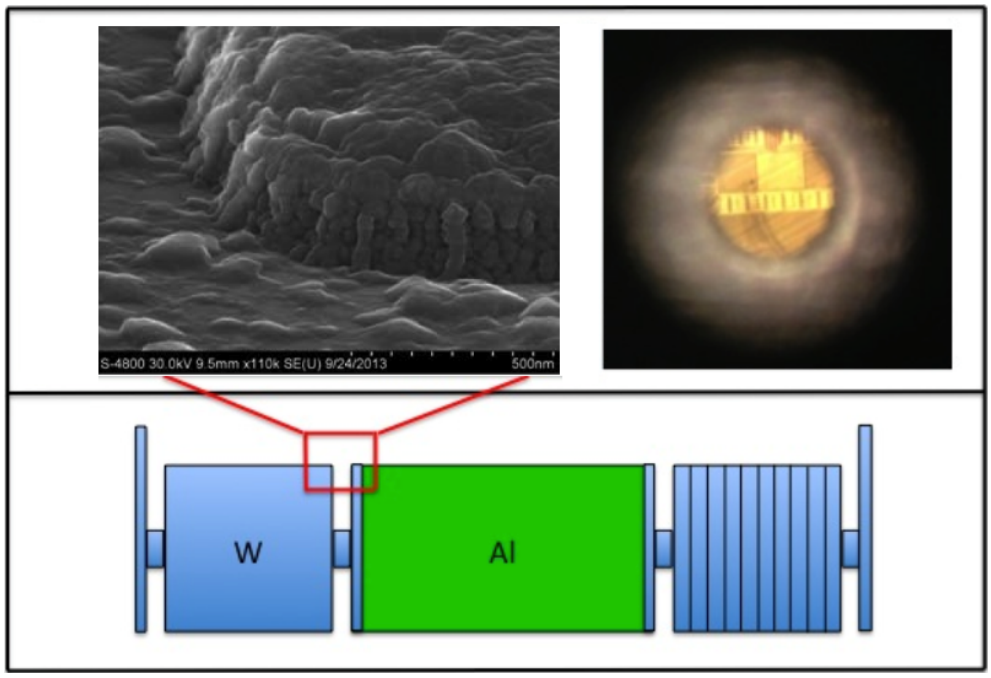


Figure 5.4: SEM image (left) of a 1997 quasiparticle device, optical image (right) through collimator and schematic model (bottom) used in successful simulations.

$$C_{norm}(T) = \gamma T \tag{5.4}$$

$$C_{super}(T) = AT^{-3/2}e^{\Delta/k_BT} \tag{5.5}$$

where $\Delta = 1.76k_BT_c$ is the superconducting band gap. In the normal state, $\gamma = 0.85meV/mK^2/\mu m^3$ for tungsten. The constant A that sets the scale for superconducting heat capacity was computed from Ginzburg-Landau theory. The usual approach is to hold the wave number constant through the transition and minimize the free energy. (see, e.g., [34]) We find that $C_s(T_c) = 2.43C_n(T_c)$, so A is set from the beginning to enforce this condition[2]. To avoid a discontinuity which clearly does

[2]In this chapter, we use 'T_c' to refer to the *middle* of the transition region. The actual phase transition occurs at the *top* of the transition region, so this is where theory papers invariably set 'T_c'. To adjust for this, the precise condition being enforced in the code is $C_s(T_c + T_{width}) = 2.43C_n(T_c + T_{width})$ where T_{width} is the 10-90% transition width so that R('T_c') = 0.99R_n.

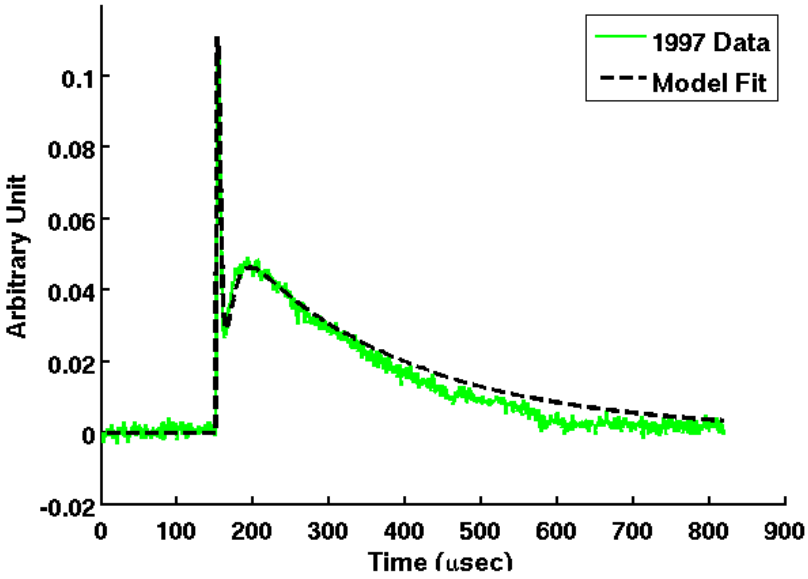


Figure 5.5: A moderately small pulse (green) taken from the previously-unexplained 1997 data and model fit (black dashed).

not appear in the data, we adopted a two-fluid model. Taking the normal fraction f_n to be some function of the resistance, we write

$$C_{TES}(T) = f_n C_n(T) + (1 - f_n) C_s(T) \tag{5.6}$$

In a uniform-current, large-device approximation (say, in a vortex-induced resistance model[35]), $f_n = \frac{R_{TES}}{R_n}$ since each line of current independently sees normal regions proportional to the overall density of distributed normal regions. We find that energy reconstructions are relatively insensitive to the shape of f_n as long as other device parameters are fit to data after that choice is made. Unless otherwise specified, we will use $f_n = \frac{R_{TES}}{R_n}$.

5.1.3 Computing Time Steps

Although different regions of the TES have the same f_n function, temperature gradients along a narrow transition width may cause substantial differences in heat capacity. When numerically moving energy between regions of variable heat capacity, care must be taken to use sufficiently small time steps that adding and then subtracting a typical pulse energy returns the device to essentially its original temperature. On the other hand, computing time is a finite resource. We need to record the current every 0.2 μsec to match the data coming from the DAQ. We divide each of these recorded time steps into a number of sub-steps calculated such that neither the Joule heating nor the incident energy is sufficient to heat any spatial bin of the TES by more than $T_w/4$. Empirically, it has been found that more stringent subdivision than this does not change the energy reconstruction to better than one part in 10^4. Since the number of sub-steps is typically less than four, there is no profit in employing a fourth order Runge-Kutta technique, which would otherwise be our go-to method. A linear Euler calculation is sufficient.

5.1.4 Resistance

Although many recent efforts have been made to empirically map out $R(T, I)$ for superconducting transitions, we used the standard Ginzburg-Landau model[9, Sec 2.1]

$$R(T, I) = \frac{R_n}{2}\left(1 + \tanh\left(\frac{T - T_c + (I/A)^{2/3}}{2\ln(2)T_w}\right)\right) \tag{5.7}$$

for a TES with normal resistance R_n, critical temperature T_c and 10-90% transition width T_w. The constant A denotes $\frac{I_{c,0}}{T_c^{3/2}}$, the strength of the suppression of T_c by non-zero current density in the film. The critical current at zero temperature, $I_{c,0}$ is given as[36]

$$I_{c,0} = 1.76\sqrt{\frac{k_B C_n}{\hbar R_n}}T_c \tag{5.8}$$

although sources[37] disagree about the pre-factor, with some increasing it by up to

a factor of two. This model does not depend heavily on the current-coupling, so we adopt an A value midway between the extremes and do not include it as a fitting parameter.

5.1.5 Electron-Phonon Coupling

In order to determine the electron-phonon coupling constant, we return to the quiescent power balance equation.

$$P_Q = I_{TES}^2 R_{TES} = \kappa(T_{TES}^n - T_{substrate}^n) \tag{5.9}$$

Recall from Figure 2.14 (brought forward to Figure 5.6) that we can take I-V curves of devices by feeding a bias current I_b into the SQUID circuit. The pickup coil L_{SQUID} measures the current I_s flowing through the TES and any parasitic resistance in the TES branch, but for slow input signals the coil itself is invisible. We can then calculate the TES resistance as follows:

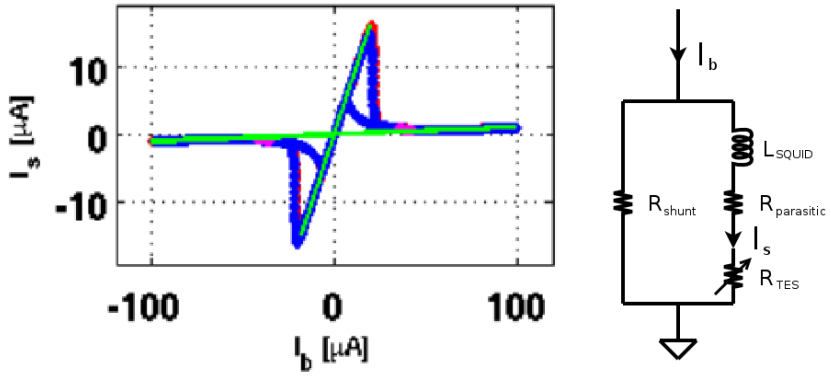


Figure 5.6: Figure 2.14 repeated for reference.

$$\begin{aligned} V_{TES} + V_{para} &= V_{shunt} \\ I_s(R_{TES} + R_{para}) &= (I_b - I_s)R_{shunt} \\ R_{TES} &= \frac{I_b - I_s}{I_s} R_{shunt} - R_{para} \end{aligned} \tag{5.10}$$

The TES resistance is sometimes labeled R_W to indicate that it is tungsten. Application of Equations 5.9 and 5.10 gives us:

$$R_W = \left(\frac{I_b}{I_s} - 1\right) R_{shunt} - R_{para} \tag{5.11}$$

$$P_W = I_s^2 R_W \tag{5.12}$$

$$T_W = \left(\frac{P_W}{\kappa} - T_{base}^n\right)^{1/n} \tag{5.13}$$

Figure 5.7 uses these equations to convert a sample IbIs curve into TES resistance, power and temperature curves. The temperature curve requires κ and n, which we will derive presently. Note the long flat region in the power curve. This occurs because ETF holds the TES in its narrow transition edge for a range on bias currents. When the device is operated in particle-detection mode, the quiescent power will be fairly constant regardless of the exact bias point. To preserve precision, we select the region around T_c as the 'quiescent region'. The pink highlight in Figure 5.7 indicates points with R_W = 40-60% of R_n. To avoid extreme non-equilibrium effects in the snap point, the selected region is chosen to be only where the absolute bias current is decreasing. The average power dissipated in the quiescent region is the 'quiescent power' P_Q.

Equation 5.9 gives the quiescent power $P_Q = I_s^2 R_W$ as a function of substrate temperature[3]. If we define a quiescent power at zero temperature $P_0 = \kappa T_c^n$, we can do a χ^2 fit of $P_Q(T_b) = P_0 - \kappa T_b^n$ for best-fit values of P_0, κ and n as shown in Figure 5.8. A free n fit as shown in Figure 5.8a gives κ values in different units, which is clearly non-physical. Once we have gained confidence that $n \approx 5$ for a large number of samples, we can go back and fix $n \equiv 5$, yielding plots like Figure 5.8b. With κ, n and T_c values now in hand, we can fill in the bottom left pane of Figure 5.7.

[3]Warning: It is extremely important when making this measurement that the substrate temperature be the same as the fridge base temperature. It is also very easy for stray power to elevate the substrate temperature in a way that is difficult to measure. If you are not very, very certain of your heat sinking, do not attempt this measurement with a radioactive source in the fridge!

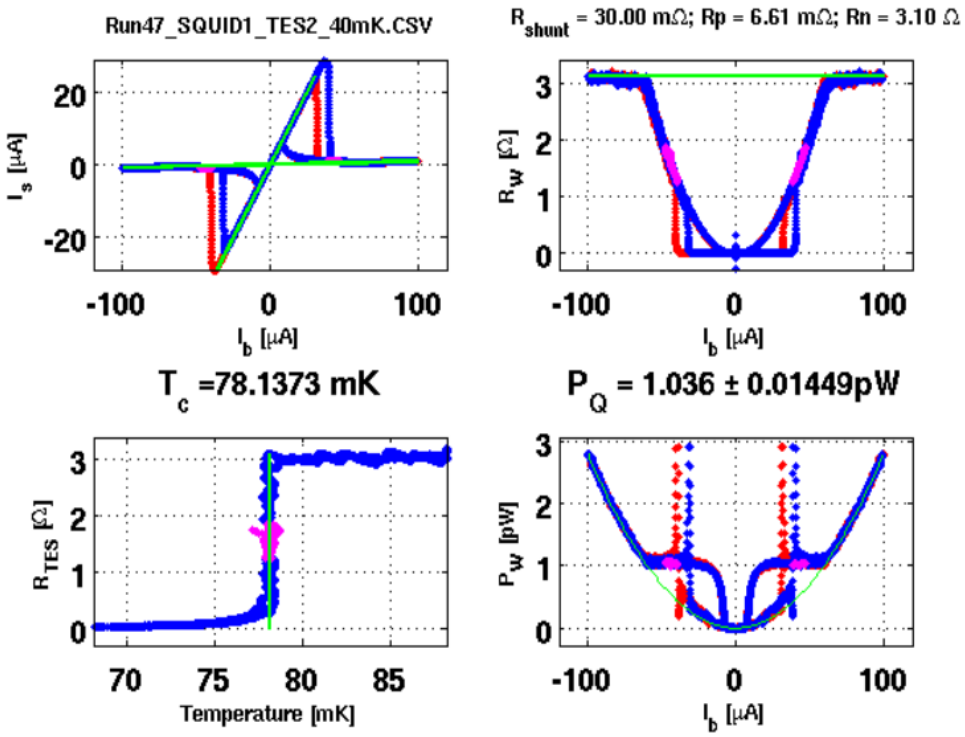


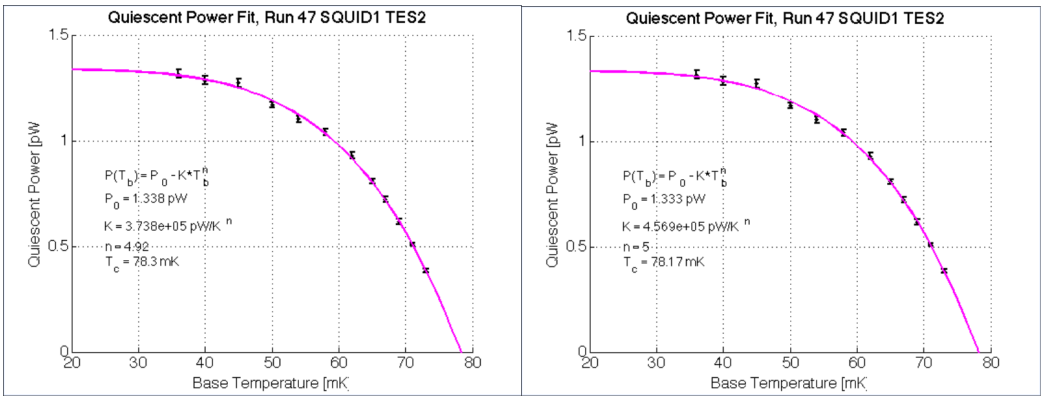
Figure 5.7: IbIs curve (top left) converted to resistance (top right), power (bottom right) and temperature (bottom left) curves.

5.1.6 Noise

Since we are headed for a χ^2 fit, we should make sure that our model provides good estimates of expected noise from the beginning. In a TES system like ours, typical sources of noise include Johnson noise V_{rms} in the TES and other resistors R_i in the circuit, thermal fluctuation P_{rms} from the link to the thermal bath and usually some empirical constant noise term to account for the SQUID, the amplifiers and anything else that doesn't depend on TES physics. These are commonly modeled as:

$$V_{rms} = \sqrt{4k_B(\sum_i R_i T_i)f} \tag{5.14}$$

$$P_{rms} = \sqrt{4k_B T^2 g f}, g \equiv \frac{dP}{dT} = n\kappa T^{n-1} \tag{5.15}$$



(a) (b)

Figure 5.8: A best-fit of the quiescent power P_Q versus base temperature T_b provides information about the electron-phonon coupling constant κ, the conduction exponent n and the critical temperature $T_c = \left(\frac{P_0}{\kappa}\right)^{(1/n)}$. In order to compare κ values between different films in the same units, n is fixed to the nearest integer as shown in (b).

where k_B is the Boltzmann constant and f is the bandwidth, or the inverse of twice the sampling rate ($f = \frac{1}{2\Delta t}$). In our TES experiments[27], the output is measured as a current, so these need to be converted to current-noise values. Remembering that the voltage V is (essentially) constant and $I = {}^{V}/_{R}$:

$$I_{V,rms} = \frac{V_{rms}}{R_{TES}} \tag{5.16}$$

$$\begin{aligned} I_{P,rms} = \frac{\partial I}{\partial E}\delta E &= \frac{\partial I}{\partial T}\frac{\partial T}{\partial E}P_{rms}\Delta t \\ = \frac{\partial I}{\partial R}\frac{\partial R}{\partial T}\frac{P_{rms}\Delta t}{C_e} &= \frac{V}{R}\left(\frac{1}{R}\frac{\partial R}{\partial T}\right)\frac{P_{rms}\Delta t}{C_e} \\ I_{P,rms} &= \frac{I\alpha}{T_e}\frac{P_{rms}\Delta t}{C_e} \end{aligned} \tag{5.17}$$

Here, C_e is the electron heat capacity of the TES, $\alpha \equiv \frac{T}{R}\frac{\partial R}{\partial T}$ is the transition steepness parameter and Δt is the sampling rate. When we go to fit real pulses, we will need to calculate $\delta I^2 = \delta I_V^2 + \delta I_P^2 + \delta I_{other}^2$. Our model obviously cannot give us δI_{other}. The other terms are minimized at the top of the transition curve, so we can get a good estimate of δI_{other}^2 by taking $\langle (I - \langle I \rangle)^2 \rangle$ on a few randomly-triggered

pulse-free traces with the TES just barely in its normal state.

5.1.7 Filtering

For traditional χ^2 template matching, it is sufficient to know the RMS value of the expected noise given the bandwidth of the recorded measurement. For applications in which we must add random noise into the model, we need to consider the bandwidth of the model as well as the measurement. In the experiment, the 0.35 μH input coil and the $\sim 1\ \Omega$ TES operating resistance give an $^{L}/_{R}$ cut-off at $\tau \approx 0.3$ μsec. This gives a bandwidth around $\frac{1}{2\pi\tau} = 0.5$ MHz. The simulation, on the other hand, has a nominal Nyquist frequency at 2.5 MHz, but noise is added in much shorter time steps. Not only is noise applied multiple times per recorded bin, but its value is amplified by the higher bandwidth of computation. To cut out the extra power from high-frequency noise, we need to apply a digital filter with a cut-off at the Nyquist frequency of recording (i.e. 2.5 MHz). Since we are mimicking an LR filter, we can grab the standard recursive single-pole low-pass filter from any electronics textbook. Rapid fluctuations are damped by feeding back a portion of the previous measurement into each new value.

$$I_{n,recorded} = (1-\beta)I_{n,calculated} + \beta I_{n-1} \tag{5.18}$$

The strength of the feedback β is a function of the ratio between the desired cut-off frequency f and the sampling/calculation time T.

$$\beta = e^{-2\pi fT} \tag{5.19}$$

Passing this algorithm forward through the data recursively ensures that each point is exponentially less affected by points further in the past and not affected at all by points in the future; this is a causal filter. If the calculation and recording rates were identical ($f = \frac{1}{2T}$, this would result in a constant $\beta = e^{-\pi}$. In Section 5.1.3, we created a mechanism to divide each recorded time step into N sub-steps, meaning $T_{calc} = \frac{1}{2f}/N$. To keep the simulated noise performance independent of a changing N, we need to keep track of

$$\beta = e^{-\pi/N} \tag{5.20}$$

and apply the recursive filter at every sub-step. There is a subtle difference between filtering the output by running the simulation then filtering and limiting the bandwidth of the model circuit by applying the filter at each step of the simulation. In the former process, rapid current fluctuations are allowed to create large temperature fluctuations, the effect of which are then filtered out. In the latter process, which is the correct model for a limiting inductor, temperature and resistance fluctuations are suppressed along with current noise.

5.2 Energy Reconstruction

For a voltage-biased TES with a resistance curve characterized by $\alpha \equiv \frac{\partial(lnR)}{\partial(lnT)}$ and $\beta \equiv \frac{\partial(lnR)}{\partial(lnI)}$ it can be shown[28] that the total derivative $\frac{d(lnR)}{d(lnT)}$ is given by

$$\gamma \equiv \frac{d(lnR)}{d(lnT)} = \frac{\alpha}{1+\beta} \tag{5.21}$$

In fact, our devices are not perfectly voltage biased. Using Thevenin equivalence to incorporate the shunt resistor, they can be modeled as a perfect voltage bias with $R_s + R_p$ in series. When we replace $I = \frac{V}{R}$ with $I = \frac{V}{R+R_s+R_p}$ in [28], Equation 5.21 becomes

$$\gamma \equiv \frac{d(lnR)}{d(lnT)} = \frac{\alpha(1+r)}{1+r+\beta} \tag{5.22}$$

where we quantify the softness of the voltage bias by the dimensionless parameter $r \equiv \frac{R_s+R_p}{R}$.

Assume the film starts in equilibrium ($P_{Joule} = P_{e-ph}$) at temperature T_0 and resistance R_0 while dissipating power P_0. If it then absorbs a small external power P_{ext}, we can linearize the changes to its energy as

$$\begin{aligned} \delta P_{e-ph} &= n\kappa T_0^{n-1}\delta T \equiv G\delta T \quad (5.23) \\ \delta P_{Joule} &= -P_0\frac{\delta R}{R_0} = -\gamma P_0\frac{\delta T}{T_0} \quad (5.24) \end{aligned}$$

Adding up the total power, we get:

$$\begin{aligned} \delta P &= \delta P_{Joule} - \delta P_{e-ph} + P_{ext} \\ C\delta\dot{T} &= -\left(\gamma\frac{P_0}{T_0} + G\right)\delta T + P_{ext} \quad (5.25) \\ \delta\dot{T} &= -\frac{G}{C}(L+1)\delta T + \frac{P_{ext}}{C} \end{aligned}$$

where we have adopted from [28]

$$L \equiv \frac{\gamma P_0}{T_0 G} = \frac{\gamma\kappa(T_0^n - T_s^n)}{n\kappa T_0^n} = \frac{\gamma}{n}\left(1 - \frac{T_s^n}{T_0^n}\right). \quad (5.26)$$

If the temperature of the film (and thus the power loss to the substrate) remained constant through the pulse, then according to Equation 5.25 the external energy could be determined exactly by integrating the loss of Joule heating. Instead an amount of energy ΔE is lost equal to the excess δP_{e-ph}.

$$\begin{aligned} \Delta E &= \int_0^\infty \delta P_{e-ph} dt = \int_0^\infty G\delta T dt \\ \Delta E &= \int_0^\infty G\left(\frac{E_{ext}}{C}\right) e^{-t/\tau_{ETF}} dt \\ \Delta E &= \frac{G}{C}\tau_{ETF}E = \frac{1}{L+1}E \quad (5.27) \end{aligned}$$

We have assumed here a delta-function for P_{ext}, but as long as the temperature response is linear, Equation 5.27 can be proved for arbitrary P_{ext} by adding up impulse responses. This means that for small energy inputs to a TES with perfect voltage bias, the energy captured by integrating the loss of Joule heating is:

$$E_{int} = E - \Delta E = \frac{L}{L+1} E \tag{5.28}$$

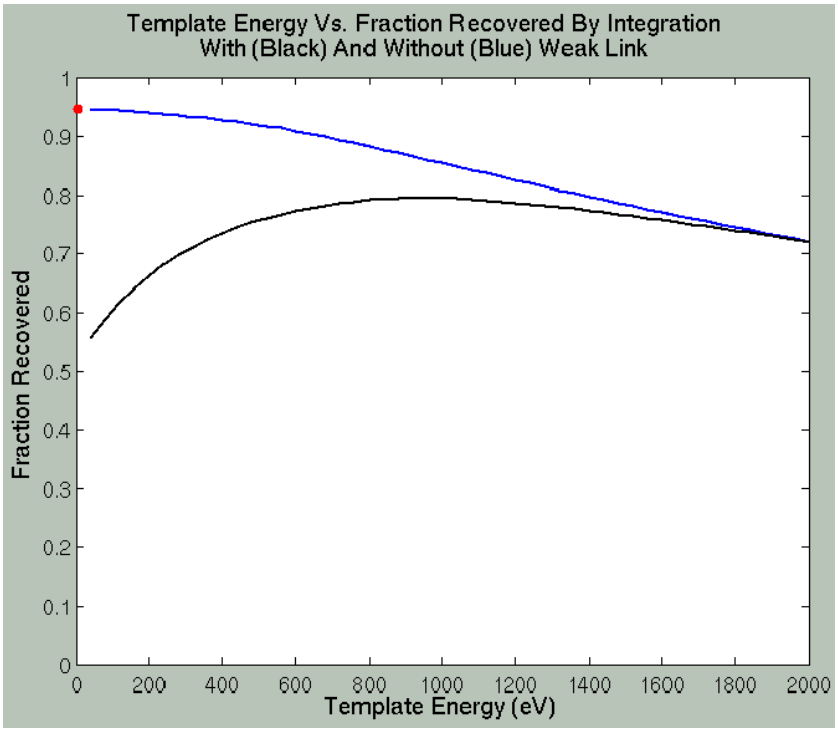


Figure 5.9: Fraction of energy recovered by integration as a function of E_{ext} for a sample TES model with (black) and without (blue) a 2.7% connected weak link. The red dot indicates the small-signal efficiency predicted by Equation 5.28.

Figure 5.9 shows in black the calculated efficiency of energy reconstruction using the integral method for a sample quasi-particle device with a particularly poor (but still connected) Al/W link. All integrals are performed on templates, so the exact input energy is known. The blue line shows the same calculation for another set of templates which have identical properties except for a perfect Al/W link. The red dot shows the small-signal efficiency predicted by Equation 5.28.

Because it dissipates energy but has negligible volume for electron-phonon coupling, the weak link is always warmer than the rest of the TES. Further, its constricted cross-section increases the current density, effectively suppressing T_c. Recall

from Equation 5.7 that the coefficient A indicates $\frac{I_{c,0}}{T_c^{3/2}}$ and the effective T_c is reduced by $\left(\frac{I}{A}\right)^{2/3}$. A constriction of cross-sectional area by f_{link} reduces A proportionally. For our devices, a poor link will have $T_{c,eff}$ reduced by 2-5 mK in the quiescent state. Since the typical transition width is $T_w \approx 1$ mK, the link spends most or all of its time in its normal state. With no electrothermal feedback, it acts like an additional parasitic resistor, albeit one that delivers its energy to the TES. (Electron energy is unable to travel into the higher band gap of the proximitized W pad, so all energy dissipated in the weak link is captured.) Additional heat delivered from one side allows the TES to sit at a cooler quiescent state without going fully superconducting, reducing α. If only one link is weak, the part of the TES farthest from the extra heat may fall almost entirely off the transition curve. This especially hurts integral reconstruction of small pulses, which are unable to heat the device up into a region of strong ETF.

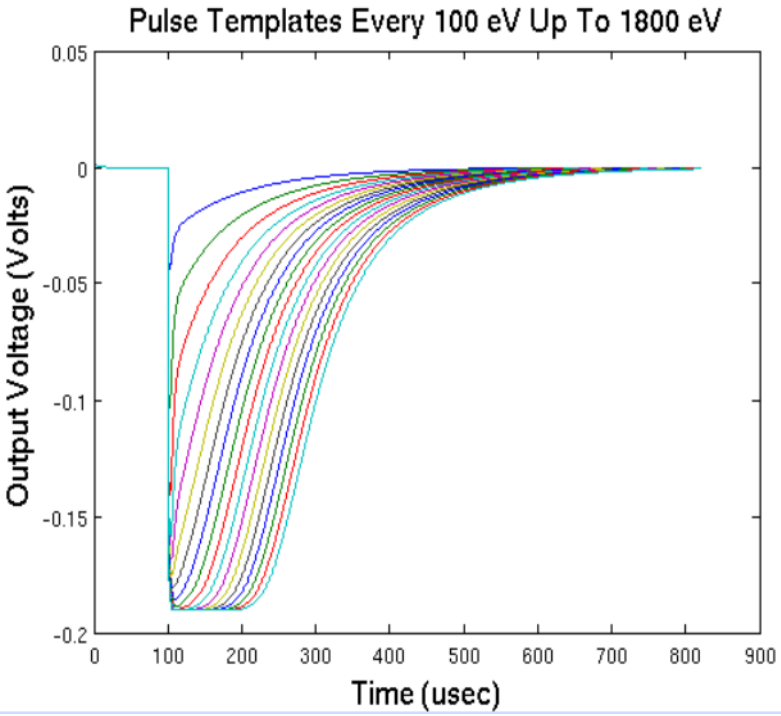


Figure 5.10: Templates from the weak-link model used in Figure 5.9. Templates are shown for every 100 eV of input energy up to 1800 eV.

Figure 5.10 shows the templates for the weak-link model used in Figure 5.9. The quiescent current is set to zero and a gain of 10^5 V/A is applied to match the output of the AC-coupled readout circuit. At low energies, pronounced initial spikes indicate that conduction across the TES places an important role. Around 900 eV, parts of the device start to briefly reach their normal state. This is when integral energy reconstruction is most effective because the entire dynamic range of the transition curve is used (minimizing the effect of the additional weak-link heat) without losing too much energy to saturation. Beyond this level, both weak- and strong-link models start to lose energy because the effective α is zero in saturation; ETF cannot remove any more energy than when the current is shut off.

The weak-link model has proved very effective at modeling the behavior of real quasiparticle devices. Knowing that a device has a weak link also improves our ability to operate it effectively. TESs are normally biased so that $R_{quiescent} \approx 0.2$ R_n to increase the saturation energy without sacrificing α. Since parasitic heating of a weak link will reduce the temperature of the main device, such devices should be supplied a higher bias current to keep them in their sensitive regimes. Although matching of long tails is an effective method for pinning down the strength of a single weak link, it is very difficult in most cases to distinguish between a single very weak link and a pair of moderately weak links on opposite ends of the device. One of the things we have learned from the weak-link model is that it is worth the effort to improve the draping of W over the thicker Al films.

5.3 Template Matching

When matching a signal S_i to a series of templates $T_{i,j}$, physicists tend to default to χ^2 minimization as in:

$$\chi_j^2 = \sum_i \left(\frac{S_i - T_{i,j}}{\sigma_{i,j}} \right)^2 \tag{5.29}$$

where $\sigma_{i,j}$ is the expected rms noise at the ith point of the jth template.

5.3.1 Non-Stationary Noise

Defining a variance $\sigma_i^2 = \langle (S_i - T_i)^2 \rangle$ is fine if the actual value of the noise varies rapidly; that is, if the noise at one measurement is uncorrelated to the noise in the previous measurement. However, the possibility of large thermal fluctuations that dissipate on a time-scale $\tau_{ETF} >> \Delta t$ calls for a correlation, or covariance, matrix with its corresponding goodness of fit metric

$$\Sigma_{i,j}^2 = \langle (S_i - T_i)(S_j - T_j) \rangle \tag{5.30}$$

$$\chi^2 = (S - T)^T W (S - T) \tag{5.31}$$

where $W \propto 1/\sigma^2$ is the inverse of the covariance matrix Σ^2. In the same way that rms noise from different sources are added in quadrature ($\sigma^2 = \sum_i \sigma_i^2$), covariance matrices from different sources, e.g. the TES and the amplifier, can be added linearly ($\Sigma^2 = \sum_i \Sigma_i^2$). In principle, each element of the simulation could have an independently calculated covariance matrix, but once a simulator with all the relevant physics and all the noise terms is created, it is less costly in terms of computing time to make a noiseless template T_i and a few thousand noisy pulses $S_{i,j}$ at each of a comb of energies and calculate the weighting matrices by Monte-Carlo.

Since the energy of real pulses will fall between energies on the comb, it is necessary to find the minimum χ^2 by parabolic or third-order fitting to $\chi^2(E)$. Figure 5.11 shows covariance matrices for sample small, medium and large templates. As expected, the 0 eV template is nearly diagonal with a width of about 100 μsec $\approx \tau_{ETF}$. As the pulse approaches saturation, the diagonal is suppressed somewhat. This is the quantity that would be used in a traditional χ^2 calculation. Far more strikingly, the off-diagonal elements are suppressed during saturation. Off the transition curve, power fluctuations have no coupling into the current readout, so correlations on the scale of τ_{ETF} are essentially absent.

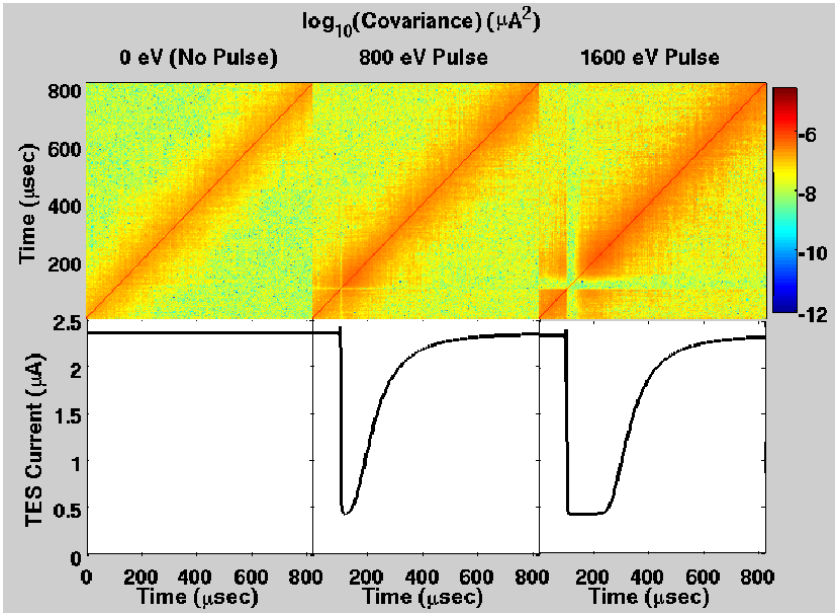


Figure 5.11: Covariance matrices (top) for pulses (bottom) of different energies. Color map is shown on a log scale. Long range correlations are suppressed during saturation. A standard χ^2 match would use only the diagonal elements.

5.3.2 Amplifier Covariance Matrix

As noted in Section 5.1.6, the effect of the amplifier (or, more generally, everything except the TES circuit) on covariant noise cannot be supplied by a TES simulator. Again, the closest thing to a measurement excluding TES noise we can take is a large number of pulseless traces $\vec{S}$ taken with the TES just barely in its normal state. Rather than the scalar RMS average used in the standard χ^2 calculation, we could use the covariant formulation

$$\Sigma^2_{amp} = \langle (\vec{S} - \langle S \rangle)^T (\vec{S} - \langle S \rangle)^T \rangle \tag{5.32}$$

Unfortunately, this method requires at least four random traces per time bin to get a non-singular covariance matrix with minimal error. Pulseless random traces are slow to collect compared to the time for which small dilution refrigerators can be held at a stable elevated temperature. If only a few hundred 'randoms' are available,

a serviceable covariance matrix can be constructed in the frequency domain. For each pulse $y(t)$, take the Fourier transform[4] $Y(f) = \int_{-\infty}^{\infty} e^{i2\pi ft} y(t)dt = \text{fft}(y)$. Phase information is not required for this exercise, so construct $\bar{Y}(f) = \langle |Y(f)| \rangle$ and place the values on the diagonal of an N x N matrix, where N is the length of each random trace. This frequency-domain variance may be returned to the time domain using the inverse Fourier matrix operator:

$$\Sigma^2_{amp} = (iFF)^T \bar{Y} (iFF) \tag{5.33}$$

where iFF is the inverse Fourier function applied to the identity matrix of size N (in Matlab: $iFF = ifft(eye(N))$).

5.4 Energy Resolution

The energy resolution of TESs is limited by Johnson noise in the link to the phonon bath. In this regime, the fundamental resolution for calorimetric detectors in general is $\Delta E_{RMS} = \sqrt{4k_B T^2 C}$, where C is the heat capacity of the sensor.[38] In fact, a more detailed analysis of TESs[15] shows that ETF improves this by nearly a factor of α.

$$\Delta E_{RMS} = \sqrt{\frac{4k_B T_0^2 C_0}{\alpha} \sqrt{\frac{n}{2}}} \qquad (E < E_{sat}) \tag{5.34}$$

T_0 and C_0 refer to the temperature and heat capacity of the device in its quiescent state. Equation 5.34 is only valid for small pulses under the quasi-equilibrium assumption that the device *has* a single temperature at quiescence. At some point, resolution is limited by the ability of ETF to remove energy from the TES. The maximum energy removed in one fall time is $E_{sat} = P_0 \tau_{ETF} = \frac{C_0 T_0}{\alpha}$. For energies above E_{sat}, we replace $E_{sat} \to E_{xray}$ in Equation 5.34, giving

[4] If pulse data will be filtered or down-selected prior to fitting, these operations must also be performed on the randoms before taking the fft.

$$\Delta E_{RMS} = \sqrt{4k_B T_0 E \sqrt{\frac{n}{2}}} \qquad (E > E_{sat}) \tag{5.35}$$

Figure 5.12 shows the FWHM energy resolution achieved by the methods described above when attempting to reconstruct the energy of simulated noisy pulses. (Assuming a gaussian distribution, $E_{fwhm} = 2\sqrt{2ln(2)}E_{rms} \approx 2.355E_{rms}$) The black line marks the theoretical best possible resolution. For the integral method, resolution is scaled up as if we had adjusted real pulse integrals for the known energy loss computed from the model. All data uses a 1-D model for conduction across the TES and assumes no amplifier noise. In the perfect connection model (left pane) the model slightly outperforms the theory in the small-pulse limit. This is a reflection of the fact that for these parameters, almost no heat reaches the far side of the TES from the aluminum film, reducing the effective heat capacity of the tungsten.

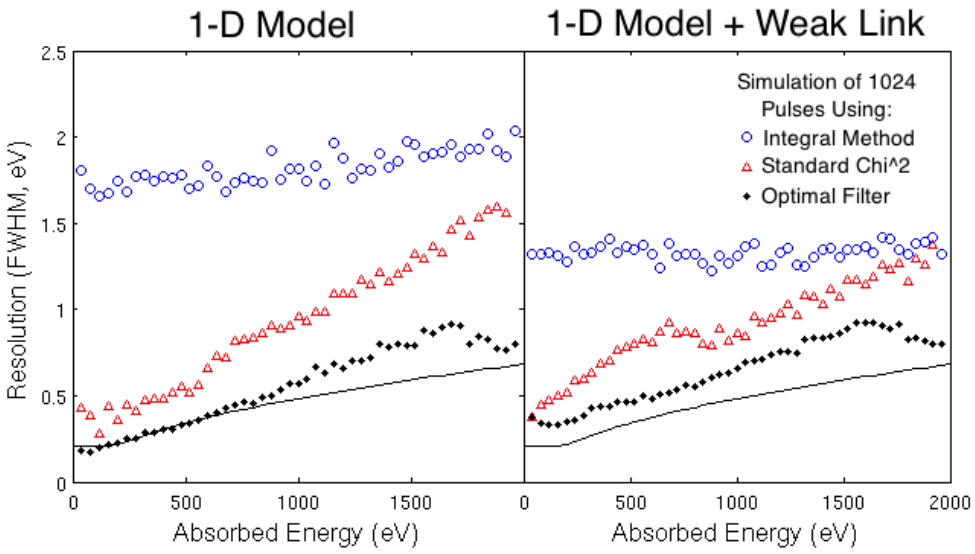


Figure 5.12: Energy resolution (FWHM) of covariance χ^2 (black dots), standard χ^2 (red triangles) and integral method (blue circles) reconstructions of 1024 simulated noisy pulses as a function of pulse energy. The theoretical best resolution is shown as a black line. The left pane shows results for a model with a perfect connection from the proximitized tungsten collector to the TES. The right pane shows results for the weak link model described in Figure 5.3.

In order to process data through the optimal filter in a reasonable time, both real and simulated pulses are reduced from 4096 to 256 time bins and all weighting matrices are 256 bins square. The deviation of the optimal filter performance from the theoretical curve at high energies is thought to be related to the loss of high frequency information in the down-select, which limits our ability to detect the end of the saturation region.

In the weak link model used to calculate the right pane of Figure 5.12, the increase in current density through the link is sufficient to drive it normal even in the quiescent state. This little normal section acts as a series resistor of $\sim 0.2\ \Omega$, effectively a heater that creates a quiescent temperature gradient across the TES. The excess heat is dumped into the part of the TES which should have the most suppressed Joule heating immediately after an event. This degrades the effect of electro-thermal feedback, especially damaging estimates of smaller pulse energies where the shape of the peak is important. The extra heat also allows the TES to sit lower in its transition without going fully superconducting. The reduced transition steepness (α) reduces the effect of P_{RMS} on the integral method, although this is partially offset in the resolution plots by the increased energy loss, particularly at low energies.

5.5 Event Identification

Figure 5.13 shows integral estimates of the energy deposited in the three TESs for a large number of X-rays incident on the quasiparticle device. The horizontal axes are TES1 and TES2. The vertical axis is TES3, the guard ring. Although X-rays are mostly collimated onto the device, many hit the surrounding bare Si and of those that do hit the Al, 91.3% continue into the substrate. These events still send hot phonons up into the films, but the energy is more evenly distributed between the three sensors. The cloud of events with high TES3 energy are identified as substrate events. They will be very useful for studying phonon absorption from the substrate to Al films, but for our purposes here they are not used to determine film properties.

X-rays incident on the 40 nm-thick W TES films are, tautologically, collected primarily in one sensor. However there is substantial phonon leakage into the substrate[41],

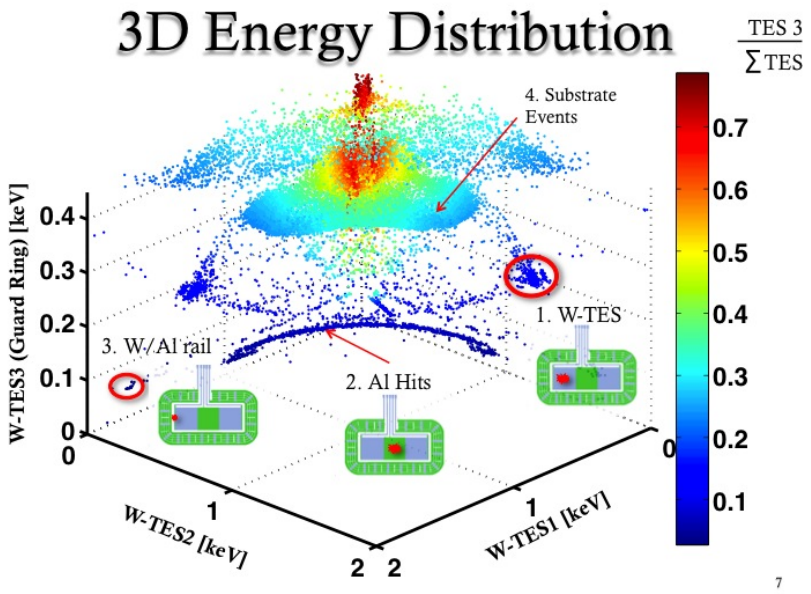


Figure 5.13: 3-D plot of events incident on quasiparticle device color coded by fraction of energy in the guard ring.

which triggers the guard ring. Circled region #1 on the right side of Figure 5.13 shows events that hit one TES directly. The mirror region on the left side is direct hits on the other TES. Note a small population of events (#3) for each TES with more collection in the primary sensor and less collection in the guard ring. These indicate hits on the W-covered Al rail opposite Al fin. Instead of being lost into the substrate, hot phonons escaping the W break quasiparticle pairs in the Al, which return most of their energy to the W sensors. Figure 5.14 shows direct hits on TES1 with the TES2 axis suppressed. In the top right, Cl K_α and K_β lines can be easily distinguished. A continuous low energy spectrum leading up to the K-lines represents Auger electrons striking the W film. Each event should shed energy into the substrate (for which the guard ring is a good proxy) proportional to the event energy. We interpret the fact that the Auger tail does not quite reach the origin as a sign of some subtle non-linearity remaining in our templates. The blob of events in the bottom right of Figure 5.14 is caused by x-rays striking the W-coated Al rail away from the TES. While there is presumably an Auger tail associated with these events, its intensity is

too low to detect here. The horizontal tail of events which all deposit similar energy into the guard ring is a signature of K-shell x-rays striking the rail farther from the TES. Imperfect quasiparticle diffusion along the rail explains how the TES receives less energy while the guard ring is unaffected.

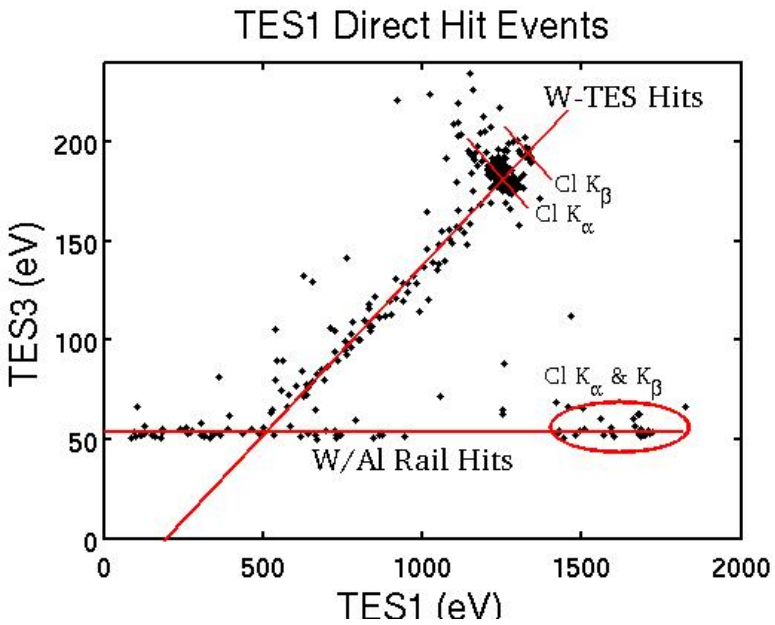


Figure 5.14: Optimal filter energies of TES1 direct hits with the TES2 axis suppressed. Selection criteria includes $E_2 < 20$ eV.

Figure 5.15 shows direct hits on a 350 μm aluminum 'fin' with the TES3 axis suppressed. These events are labeled as region 2 in Figure 5.13. The red line indicates how much energy would be collected if the fin offered no barrier to propagation. Instead, quasiparticle trapping reduces the apparent energy, yielding the characteristic 'banana plot'.

5.6 Linearity Concerns

Before we get to the punchline of this analysis, a moment to cross-check our reconstructed energy scale would be prudent. We argued in Section 5.5 that direct x-ray

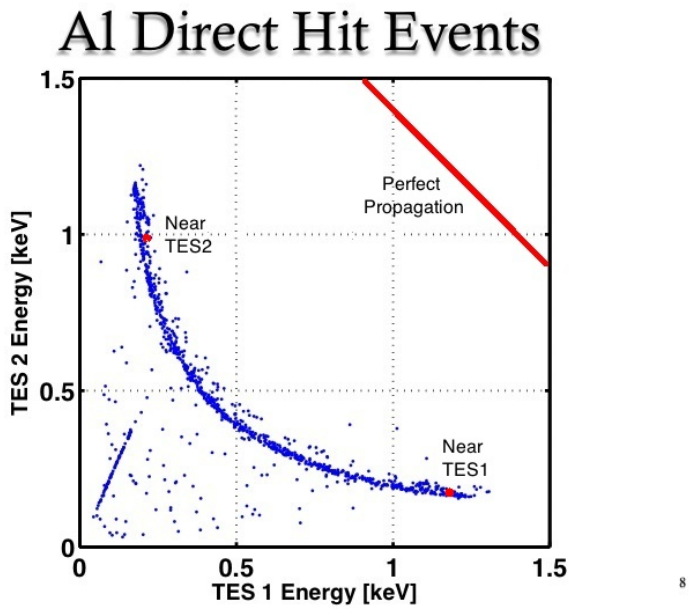


Figure 5.15: Aluminum direct hits with the TES3 axis suppressed. Selection criteria includes 37 eV $< E_3 <$ 49 eV.

hits on a TES should shed energy into the guard ring proportional to the energy absorbed in the primary TES. In other words, the direct W hits in Figure 5.14 should form a straight line through the origin. In fact, despite measurement of many of the model parameters and best-fit solutions for large and small pulses, this usually turns out not to happen automatically. Figure 5.14 is taken from a TES with a critical temperature of 78.2 mK, which is fairly high for our fabrication process. When struck by a Cl K_α x-ray, this device saturated for about 80 μsec, less than its 100 μsec recovery time. The W direct hits have a slight non-linearity, but it appears to be mostly at lower energy scales than the events in which we are interested. This can be corrected by playing with the shape of the two-fluid heat capacity function discussed in Section 5.1.2 but in the absence of some theoretical motivation or better data, we should refrain from too much fine-tuning.

However as the TES goes further into saturation, this non-linearity becomes more pronounced. Figure 5.16 shows direct hits on a TES with a critical temperature of T_c = 66.7 mK using the optimal filter as we have constructed it above. With a critical

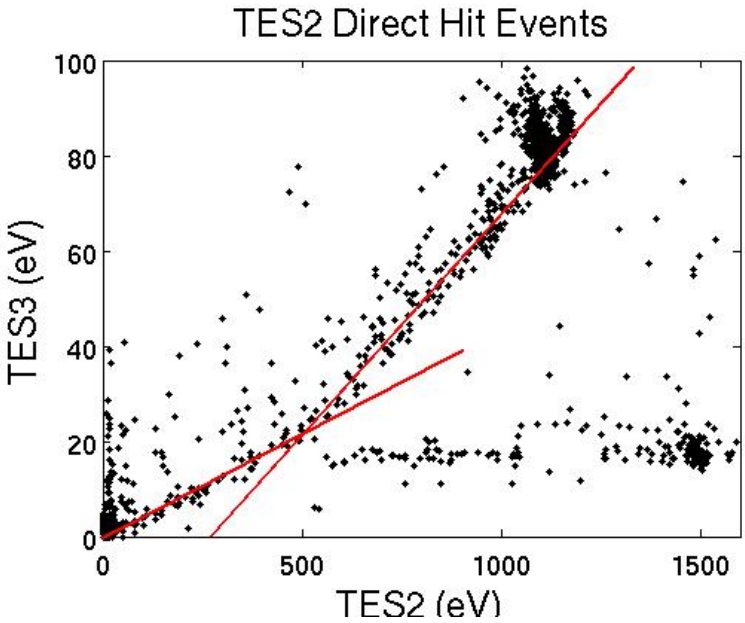
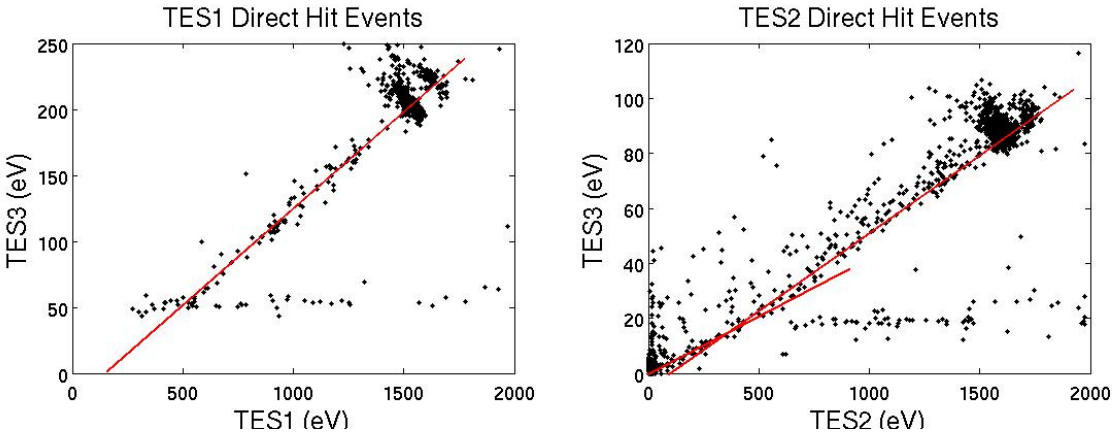


Figure 5.16: Standard optimal filter energies of direct hits on a TES with lower critical temperature.

temperature 85% of the previous device, its quiescent power ($P_Q = \kappa T_c^5$) is reduced by a factor of two. The same Cl K_α x-ray, sends this device into saturation for 250 μsec, but its recovery time is still about 100 μsec. The red lines in Figure 5.16 show how the energy collection in the primary TES changes abruptly around 500 eV, which is close to the energy of a barely saturated pulse. Further, the energy reconstructed for Cl K_α events is reduced in this model compared to the previous model. It is clear that somehow we are mishandling the reconstruction as the TES goes from linear to saturated operation.

Our knowledge of the normal resistance and saturated pulse heights is fairly precise, so the quiescent operating point for each run is known very well. Given this, the amount of energy needed to saturate the sensor is directly proportional to the heat capacity in the superconducting state and the transition width ($Q = C\Delta T$). Unfortunately, the narrow transition shown in Figure 5.7 (bottom right) is only a guideline. The device is not in perfect thermal equilibrium during IbIs measurements. Since, the electron-phonon coupling constant κ is empirically measured, the TES recovery



(a) Figure 5.14 remodeled (T_c = 78.2 mK) (b) Figure 5.16 remodeled (T_c = 66.7 mK)

Figure 5.17: W direct hit models from above reanalyzed with half the normal state heat capacity given in Kittel[33]. To preserve the Ginzburg-Landau argument, the superconducting heat capacity is similarly reduced.

time is also largely a function of superconducting heat capacity and transition width. Models must fit measured fall times, so any change in superconducting heat capacity will be cancelled by a change in the best-fit transition width. In effect, changing the superconducting W heat capacity has no effect on the linearity of the model.

Normal state heat capacity, on the other hand, has no such balancing mechanism. Once the TES goes into saturation, the transition width is irrelevant to further heating but the amount of time spent in saturation is highly dependent on exactly how hot the sensor gets. (Remember $P_{cooling} \propto T^5$ but $P_{warming}$ is constant in saturation.) This means that the normal state heat capacity has a large effect on how long a pulse of a particular energy with remain saturated in the model and, conversely, how much energy a pulse with a particular saturation time will be reconstructed to have. The low temperature values for the heat capacity of W given in Kittel[33] and by the CRESST experiment[39] are derived from measurements of very high quality single phase crystals. Our sputtered films are known to be admixtures of α and β-phase W with comparatively small grain sizes and non-trivial vacancy density. It is likely that the heat capacity of our films is reduced from the value for single crystals.

Figure 5.17 shows the linearity checks presented in Figures 5.14 and 5.16 remodeled

with all heat capacities halved. Note that both direct hit Auger lines now pass much closer to the origin, although neither model is perfect. Note also that Cl K_α events are reconstructed at 1.5 keV in both models, as opposed to 1.05 and 1.2 keV, respectively, in the original models. (The fact that we do not recover 2.62 keV is discussed in Section 5.8.) To completely solve the linearity issue, we should make complex impedance measurements to directly assess the heat capacity of our W films as a function of bias point after the fashion of Lindeman et. al.[40]. Look for future publications by J. J. Yen on this subject. Without precise heat capacity data, further refinements of the existing model are not well motivated. However we may now return to our discussion of quasiparticle diffusion in the Al "collection fin" confident that future modification of the results due to linearity corrections will be minimal.

5.7 Modeling Quasiparticle Diffusion

In any moment of time, quasiparticles generated in the Al fin might experience one of three processes. They might diffuse to a different part of the fin; they might combine with another quasiparticle or get trapped in a local minimum of the band gap and be lost; or they might cross the boundary into one of the W pads and be collected[5]. The exact diffusion constant D is not particularly important; as long as energy arrives at the TESs on a scale fast compared to τ_{ETF} we have no further need to improve it. We do see arrival time differences between TES1 and TES2 of about 2 μsec which do probe D directly, but for now we are not using that information. The major concerns for our Al films are loss and boundary impedance at the Al/W interface. To model these, we will define an rms diffusion length $L_{diff} = \sqrt{\frac{D}{g}}$ where g is the $1/e$ loss rate (or the inverse of the quasiparticle lifetime) and an absorption length $L_{abs} = \frac{D}{A*dx}$ where A is the probability per time that a quasiparticle near the TES will be captured and dx is an arbitrary setting to dictate how close the quasiparticle has to be before A applies. L_{abs} can be thought of as the distance through which quasiparticles would have to diffuse to achieve the same capture rate as the boundary impedance. Since we intend to use fins of different length and the diffusion constant D is arbitrary, we

[5]aka 'the goal'

can normalize all of the lengths (L_{diff}, L_{abs} and dx) to the fin length L_{Al}.

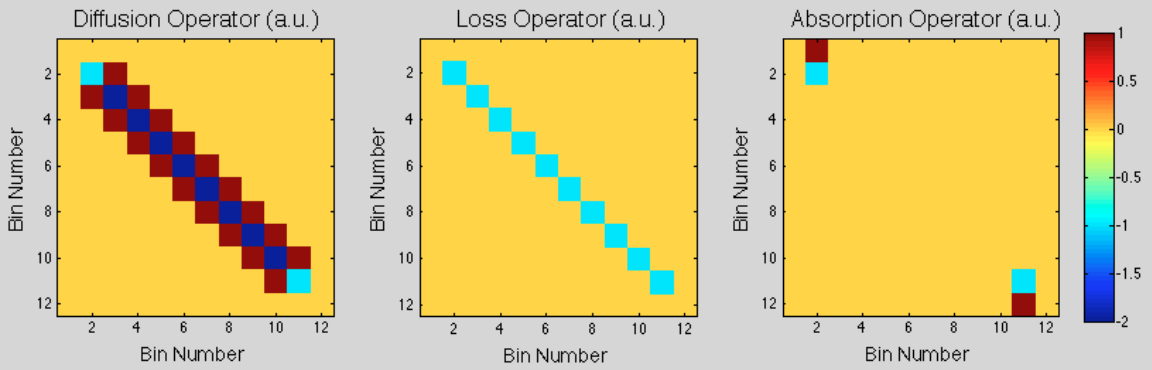


Figure 5.18: Three operators needed to model quasiparticle transport in Al films. The Al fin is divided into N=10 bins for this graphic. N=100 is a more typical working precision.

The Al films are essentially two-dimensional, but with collectors only on the ends a one-dimensional model will do.[6] If we divide the fin into N strips, we need to keep track of $N+2$ energies, one for each strip and one for each TES. This sets a natural 'closeness' scale $dx = 1/N$. Figure 5.18 summarizes the transport operators that map the quasiparticle density in Bin (Row #) now to the density in Bin (Column #) in the next time step. The diffusion operator **D** carries two units of energy out of the current bin, one to the left and one to the right. No energy is carried into Bin 1 or Bin N+2 since these represent the TESs, which cannot be reached by diffusion. The energy scale for diffusion is set by the diffusion rate $d = \frac{D}{dx^2}$. (The energy scale is also set by the amount of energy in the bin, but these are all proportional operators so this is already implied.) The loss operator **L** simply removes one unit of energy, set by the loss rate $-g$, from each bin. The absorption operator **A** moves one unit of energy from Bin 2 to Bin 1 and another from Bin N+1 to Bin N+2. The energy scale for absorption is set by the absorption rate A. Note that **A** allows the TESs to capture energy, but no process allows it to escape. Once it crosses the W interface, energy is recorded by the SQUID, so its subsequent loss to electron-phonon coupling

[6] This will produce diffusion lengths that are low by a factor of $\frac{1}{\sqrt{2}}$, which is only relevant if we need to compare our results to someone else's 2-D diffusion model.

is not relevant to the QP diffusion model. The lower band gap of the W is considered to completely block back-diffusion of quasiparticles.

To determine what fraction of x-ray energy incident on Bin n of the Al fin would end up in each TES, we could set up an energy vector with all zeros and a one in the $(n+1)^{th}$ cell, an $(n \vec{+} 1)$ vector. If we hit this vector repeatedly with the Hamiltonian $\mathbf{H} = \mathbf{D} + \mathbf{L} + \mathbf{A}$, gradually accumulating energy in the TESs until it stopped changing, we could then read off the value in the 1^{st} and $(N+2)^{th}$ cells. Until the CS folks finish their Infinite Iteration Machine, we could instead characterize $\mathbf{H}$ by its eigenvectors $\vec{\mathbf{V}}$ and eigenvalues λ. The TES bins will create a pair of degenerate eigenvectors which we can identify by their duplicated eigenvalues and set to unit vectors in the $\vec{1}$ and $(N \vec{+} 2)$ directions. Since $\vec{\mathbf{V}}$ contains an eigenvector for each bin of the Al, the collection fraction for the left TES can be taken from $\vec{1}/\vec{\mathbf{V}}$ and for the right TES by $(N \vec{+} 2)/\vec{\mathbf{V}}$. These left *vs* right vectors can be multiplied by an energy scale factor E_0 and compared to measured banana plots. We are then free to pick a goodness-of-fit metric and optimize L_{diff}, L_{abs} and E_0.

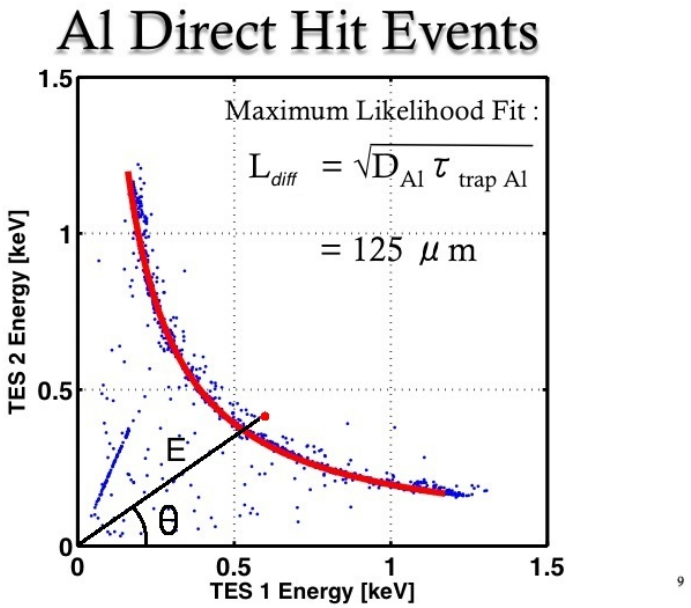


Figure 5.19: Maximal likelihood fit of the banana plot shown in Figure 5.15

Figure 5.19 shows a maximal likelihood fit assuming gaussian energy uncertainty for the direct Al hits from Figure 5.15. Selection of the events to be fit is very important because accidental selection of an event which is far from the 'real banana' compared to the assumed uncertainty will contribute extremely low likelihood. To get around this, an initial cut is made by eye and the model is run iteratively, with events outside 5σ being discarded after each round. The resulting values of $L_{abs} = 130\ \mu$m, $L_{diff} = 125\ \mu$m and $E_0 = 2400$ eV are characteristic of all three devices fabricated on this chip, even though their Al fins are different lengths (250, 350 and 500 μm).

Treating the best fit curve in terms of operators in position space results, not simply in an energy *vs.* energy curve, but in an energy *vs.* position curve for each TES. This means each event can be assigned an angle θ and a distance from the origin which can be readily transformed to a position and energy as shown in Figure 5.20.

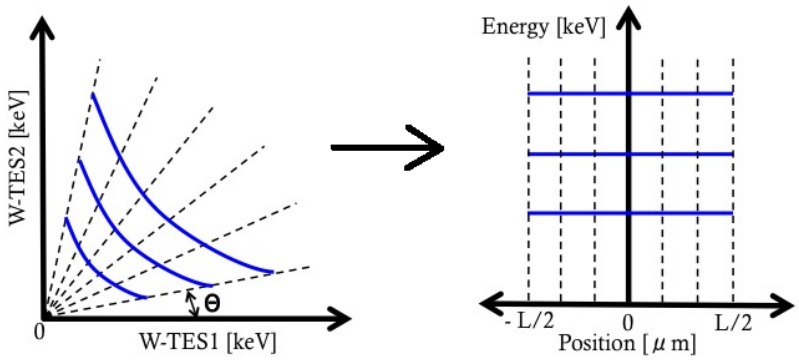
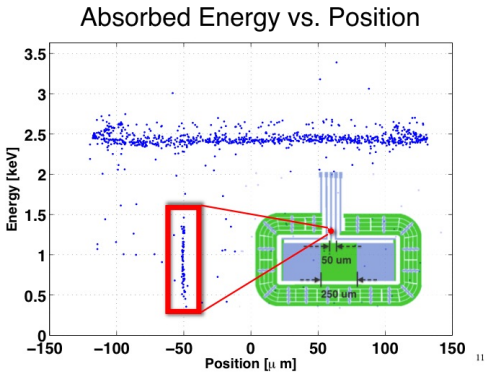
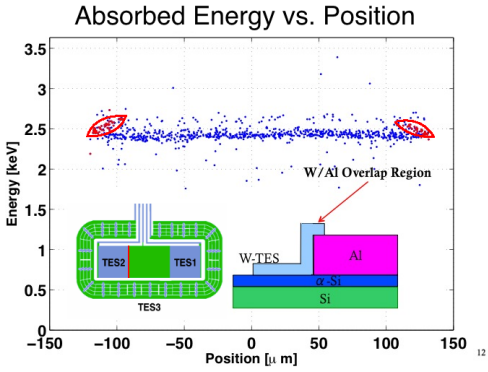


Figure 5.20: Angle from TES1 axis transforms to position on the Al film by interpolating from best-fit curve. Distance from the origin transforms to total absorbed energy (before trapping losses) by taking the ratio of the event to the best-fit curve along the line at θ and multiplying by the scale factor.

Figure 5.21 shows the events in Figure 5.19 transformed according to the algorithm given in Figure 5.20. The banana itself becomes the horizontal line at 2.4 keV. The 'spur' coming from the origin becomes a vertical line offset from the center of the Al film by 50 μm. This corresponds with the location of the W/Al ground lead. The covering of highly absorbent W ends a few microns before the fin and the Al



(a) Events on the ground rail show the correct position, but reduced energy.



(b) Events in the TES/fin overlap region are slightly more efficient at capturing x-ray energy.

Figure 5.21: Banana plot from Figure 5.19 transformed as shown in Figure 5.20

underneath necks down from 10 μm to 2 μm. Although most of the events in the red box in Figure 5.21a are Cl K_α events, this bottleneck impedance explains why the spur of ground-rail events does not reach the main K_α line. Figure 5.21b highlights events striking the overlap region between the W-TES and the Al fin. The two films combined reabsorb phonons more efficiently, allowing these events to be selected by a decrease in guard ring energy. In further analysis of direct Al hits, these events are cut.

5.8 Accounting For Energy Loss

If the above discussions, we have focused on 2.62 keV Cl K_α x-rays striking our films. In no case have we recovered 2.62 keV of energy. It would be easy at this point to say in passing that the remaining energy must have escaped as phonons into the Si substrate, but a recent paper by Kozorezov et. al.[41] gives us a framework to determine *how much* energy should be carried away in this manner. This turns out to be a bit complicated. Quoting Kozorezov:

> "The energy down conversion following the absorption of an energetic photon in a metal can be viewed as evolving through three distinct stages.

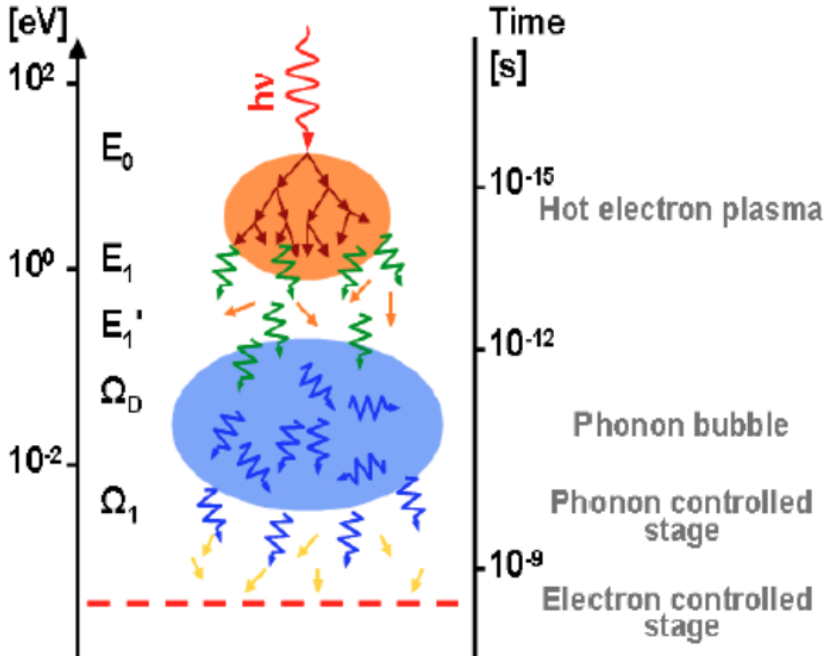
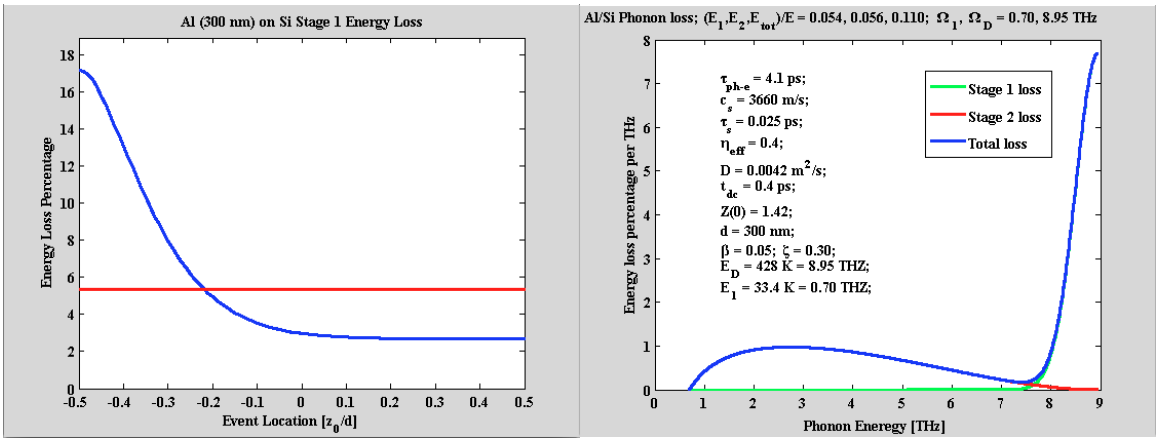


Figure 5.22: Summary of phonon losses for three stages of down-conversion. Image from Aug 2006 talk by A. Kozorezov.

Stage I starts from the ejection of a photoelectron and proceeds through electron-electron (e-e) interactions, finishing with the formation of a highly non-equilibrium distribution of electrons and holes in the conduction band occupying a tiny volume in the vicinity of an absorption site. During Stage I the entire photon energy is transferred from photoelectron to the ensemble of secondary, ternary, etc. electronic excitations with rapidly diminishing mean energy. Stage II takes over when the mean energy of electronic excitations becomes sufficiently low for the electron-phonon (e-ph) interaction to determine further momentum and energy relaxation processes. This happens below a threshold E_1^* , i.e., $\epsilon \leq E_1^*$ where ϵ is the electron energy. For typical metals E_1^* far exceeds the Debye energy: $E_1^* >> \Omega_D$. During this stage electrons and holes mostly relax via emission of phonons. As a result, a bigger fraction of photon energy is transformed into the energy of the evolving phonon distribution. Stage



(a) Stage I energy loss versus depth in our 300 nm Al films.

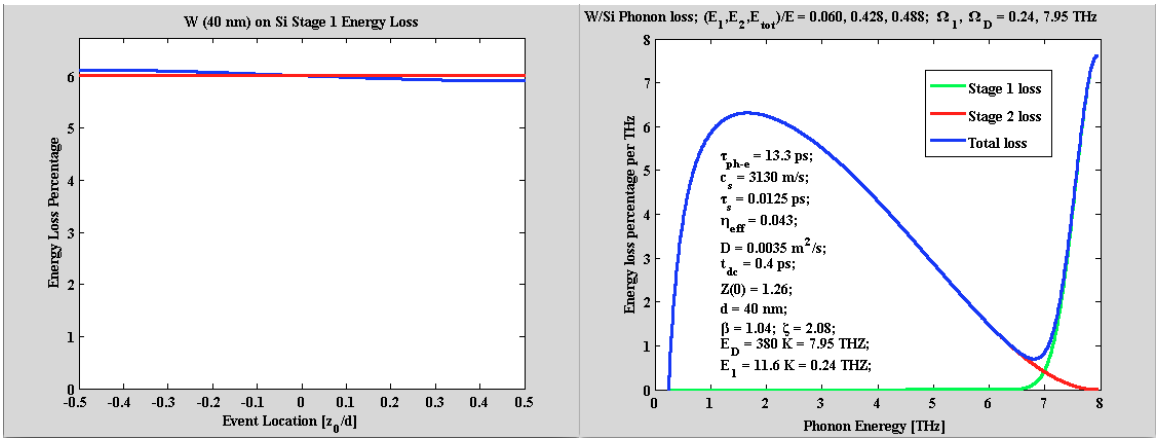
(b) Total energy loss in Al assuming mean Stage I loss.

Figure 5.23: Numerical estimates of energy losses in our 300 nm Al and 40 nm W films.

> II is completed once the mean excitation energy reaches the low-energy threshold Ω_1, which is determined from the relation $\tau_{e-ph}(\Omega_1) = \tau_{ph-e}(\Omega_1)$. Below this threshold (Stage III) a rapid energy conversion into predominantly electronic excitations takes place." (Reference [41])

Stage I and Stage III consist entirely of electron-electron interactions, so there is no cumulative energy loss from the quasiparticle system. However [41, Eqn3] assesses the energy of primary phonons generated in the x-ray interaction which ejects the initial photoelectron. We will refer to this as 'Stage I loss'. The Stage I loss is generally not as much as in Stage II, but it may depend on how far the x-ray penetrated through the film before being absorbed. Since most x-rays pass entirely through our films (i.e. the entire depth is uniformly sampled), this will create an intrinsic spread in our energy histograms. Energy loss in Stage II is frequency and dwell-time dependent, but position independent since the size of the absorption site very quickly becomes much larger than the film thickness. The upshot is given in [41, Eqn9]. The details are discussed in [42], which is an instructive read but not for the faint of heart.

Figures 5.23 and 5.24 show numerical simulations of the energy losses in our films[43]. Our comparatively thick Al films contain plenty of volume where the Stage



(a) Stage I energy loss in our 40 nm W films. (b) Total energy loss in W.

Figure 5.24: Numerical estimates of energy losses in our 300 nm Al and 40 nm W films.

I loss is a constant 2-3%, but high diffusion lengths in Al mean that near the Al/Si interface the loss of first generation phonons increases substantially. Total predicted losses range 8-22% depending on depth of absorption. Thus for a monochromatic 2.62 keV incoming beam, we expect a peak at 2.41 keV of reconstructed energy, but with a tail all the way down to 2.04 keV. Fortunately, the excess phonon energy in 'tail' events would turn up in the guard ring, increasing its collected energy by nearly a factor of three. The TES3 cut used to isolate Al direct hit events from substrate events also cuts out most of this tail, allowing us to clearly distinguish the 2.62 keV K_α line from the 2.82 keV K_β line, even with the complication of non-linear energy loss due to quasiparticle diffusion after Stage III settles out.

Despite its thinness, the high density of our W films ensures a nearly-constant 6% energy loss in Stage I. Largely due to Stage II shedding of secondary phonons, we expect 48.8% losses on absorption in our W films. It might seem strange that after going on about the weak e-ph coupling in W, we now say that the energy loss in the e-ph relaxation stage is large. In fact, the emission of phonons by high energy electrons is enhanced by the high density of W. The weak ph-e coupling ensures that the Stage II cut-off energy Ω_1 (defined as the energy at which $\tau_{e-ph} = \tau_{ph-e}$) is low. It is *because* of its weak low-energy electron-phonon coupling that Stage II takes a long

time and sheds a large amount of electron energy in W. If this model were exact, we should expect the 2.62 keV line to show up in our direct W hits as a 1.357 keV peak with an intrinsic width of < 10 eV.

5.9 Total Effect Of Optimal Filter

In Section 5.4, we saw that the optimal filter clearly beats out other methods of energy reconstruction in the magical world of simulation where fridges are all perfectly stable and the environment is completely noiseless. How did we do in the real world? To answer that question it makes sense to start with Cl K-shell events in one of the W-TES films for two reasons. First, we will be dealing entirely with saturated events, which makes detecting drifts in base temperature easier. Second, the Kozorezov factor computed above is constant through the depth of the film, meaning the inherent line width is negligible. The T_c=78.2 mK model in Figure 5.17a has been selected for this analysis because the environmental noise was more stable for that run.

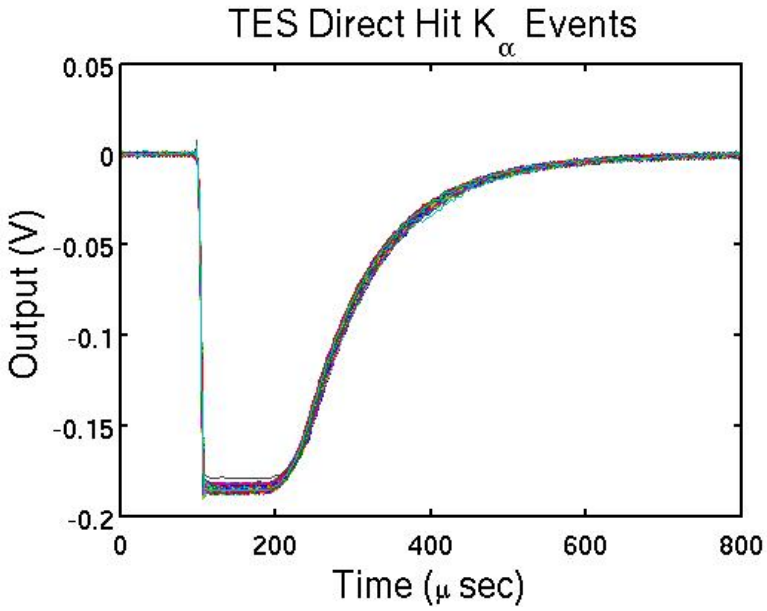


Figure 5.25: Raw K_α pulses. The variation in saturation height is a measure of base temperature instability.

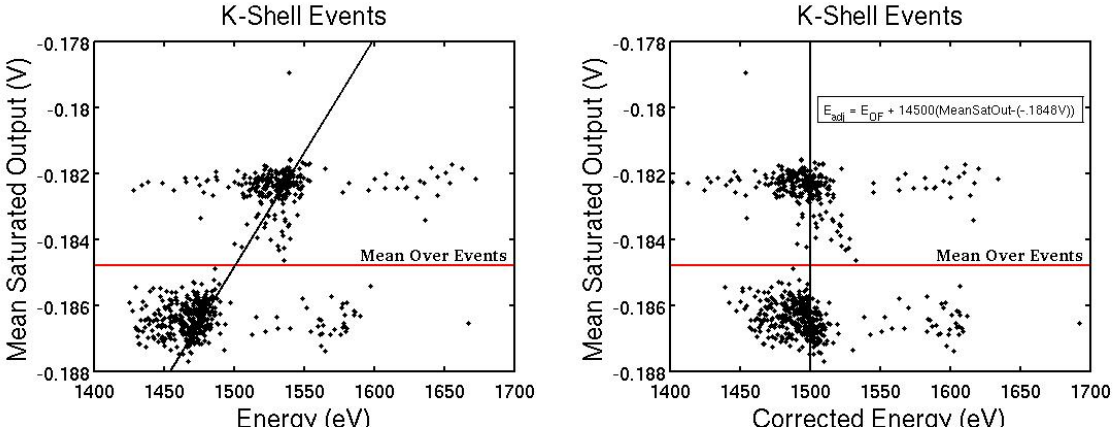
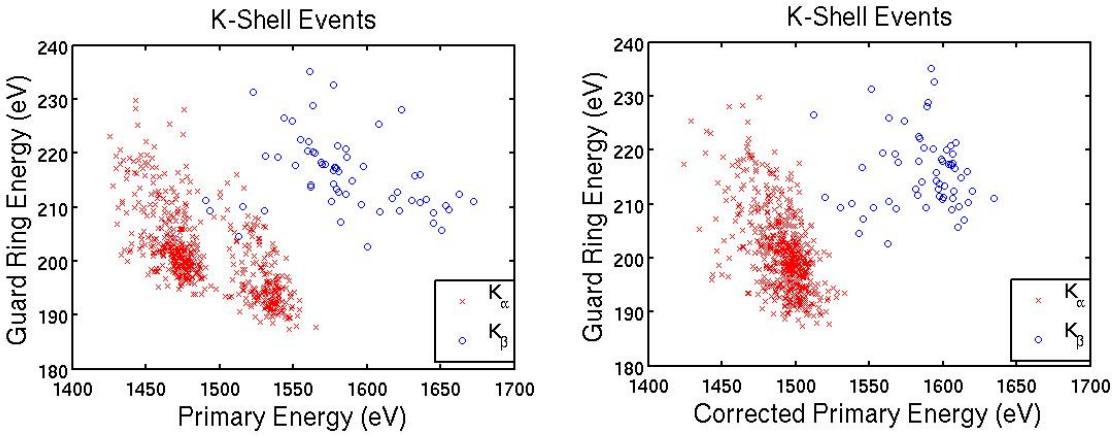


Figure 5.26: Reconstructed energy of K-shell events (a) before and (b) after amplitude correction

One of the advantages of integral energy estimation which is lost in the template matching scheme is a comparative insensitivity to changes in base temperature. Small increases in base temperature drive up the quiescent bias point, reducing the amplitude of pulses in a non-linear way. In the case of saturated pulses, this also increases the saturation time. The two effects partially cancel when an integral is taken, but a χ^2 minimizer simply assigns poor fitting to the inaccurate pulse top for *all* templates and interprets the increased saturation time as an increase in energy. This can be seen most clearly in the selection of Cl K_α pulses shown in Figure 5.25. Although the refrigerator base temperature stayed within a 2 mK zone (39-41 mK) throughout the run, pulse heights are not consistent.[7]

When we plot out the mean saturated heights of pulses vs. their reconstructed energy as in Figure 5.26a, we see not only that there is a strong correlation, but that in this particular run there was a sharp shift in base temperature about halfway through. This was actually known beforehand; we are merely confirming its effect. The red line shows the average amplitude of all saturated pulses, which is the level to which large-energy templates were fit. Since the changes in pulse height and energy

[7]It is worth noting that the presence of a waterfall region increases the sensitivity to base temperature fluctuations. We add this to the growing list of reasons to make better draping of W over our Al fins a fab R&D priority.



(a) (b)

Figure 5.27: K_α (red x's) and K_β (blue o's) events in the W-TES (a) before and (b) after amplitude correction.

are comparatively small, we can remove at least the linear portion of their dependence as shown in Figure 5.26b. To preserve the overall energy scale of the model, pulses which have the same mean saturated height as the templates should not change their energy. Figure 5.27 shows the effect of amplitude correction on the K-shell events selected from Figure 5.17a. Because we have selected a run with a sharp temperature transition, the merging of two populations to form each group of events can be clearly seen.

Interestingly, a clear inverse correlation between the primary TES and the guard ring remains after amplitude correction. There appears to be a 35-40 eV (or 1.4% of 2.6 keV) spread in the energy collected in the primary TES which is a function of where on the TES the x-ray was absorbed. Using Kozorezov's formulation for an infinite film, we predicted negligible ($< 0.2\%$) spread in Section 5.8 purely as a function of absorption depth. The observed spread gives us a good estimate of the variation in energy loss as a function of absorption location in the plane of the device. In fact, neglecting the few events which strike the corners of the square TES, we expect the TES to re-absorb half as many of the escaped phonons from an edge strike as from a center strike. Since the difference between these extremes is 35-40 eV we expect center strikes, which comprise the majority of events, to absorb 70-80 eV more

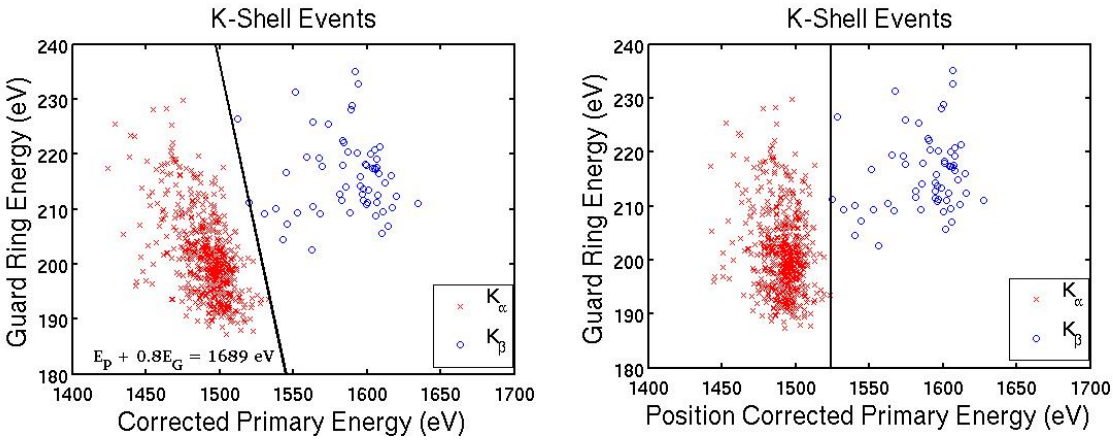
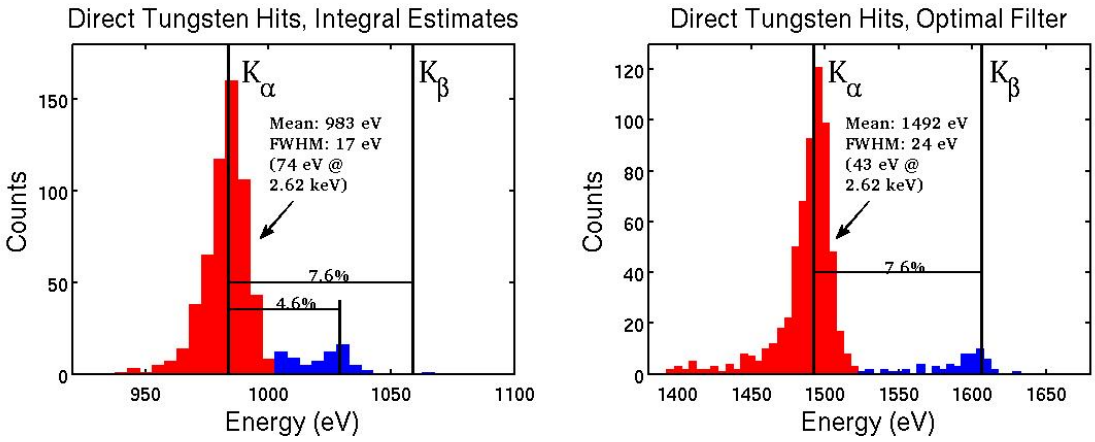


Figure 5.28: K_α (red x's) and K_β (blue o's) events in the W-TES (a) before and (b) after position correction.

energy than predicted in Section 5.8, or about 1.43 keV.

While informative, position dependence should be removed before we calculate the final energy resolution of our system. Again, the guard ring may not recover energy lost to the primary TES in a linear manner, but since the variations are small we can at least remove the linear component. This is done in Figure 5.28, being careful to keep the mean energy of K_α events constant.

Figure 5.29b shows a histogram of the amplitude- and position-corrected events in Figure 5.28b. The tail of events straggling below each peak is a continuation of the Auger electron spectrum into higher energies. For purposes of peak fitting, these events are cut. Given the mean of the K_α peak, a vertical line is placed where the K_β peak should be. Although K_β statistics are too low to get a precise peak fit over the Auger background, the bulk of the events fall near the predicted energy. The main K_β grouping contains 8% as many events as the K_α peak (44 vs. 557). This is more than the fundamental 4.5% relative emission rate, but a combination of Auger contamination and K_β's increased escape probability from the NaCl fluorescence target may explain the difference. Recall that while position-correcting, we increased our estimate of total energy captured by the W electron system in the K_α peak from 1.36 keV to 1.43 keV. The final reconstructed value of 1.49 keV is fairly close given the



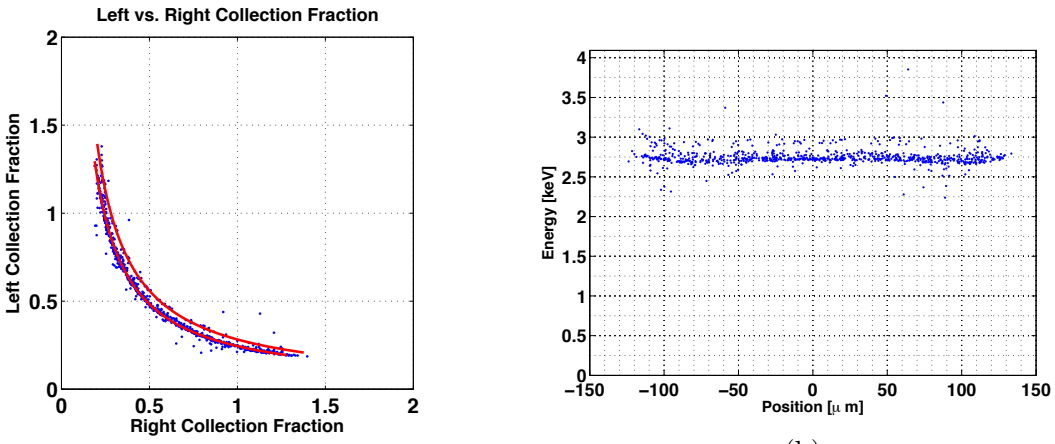
(a) (b)

Figure 5.29: Histograms of K_α and K_β peaks in the W-TES using (a) integral and (b) optimal filter reconstructions.

uncertainties in normal state heat capacity discussed in Section 5.6.

Figure 5.29a shows the amplitude- and position-corrected energy histogram starting with integral estimates. Completely new correction coefficients and group selections were made to prevent the final resolution being contaminated by decisions made on optimal filtered quantities. As expected for saturated pulses, the overall energy scale is too low and locally non-linear. On the arbitrary energy scale of the integral reconstruction both peaks appear narrower than when optimally filtered, but when adjusted to their respective local energy scales (by assuming 200 eV separation between K_α and K_β), the optimal filter wins out by nearly a factor of two.

In principle, the optimal filtering method can reconstruct 1.5 keV pulses with TES-dominated noise to 1 eV precision. Clearly we have not reached the threshold where dealing with non-stationary TES noise matters. For example, the rms noise in the quiescent state was measured to be 1.0 mV for the device described in this section, compared to a prediction of 0.62 mV for TES physics alone. Fixsen et. al.[45] recently published an excellent paper in which they use a covariant filter to detect optimal photons over a S/N ratio of a few. In their case, it was important to properly deal with low frequency noise which might be mistaken for a pulse. In our case, a standard χ^2 minimization without noise correlations performs as well as the covariant optimal



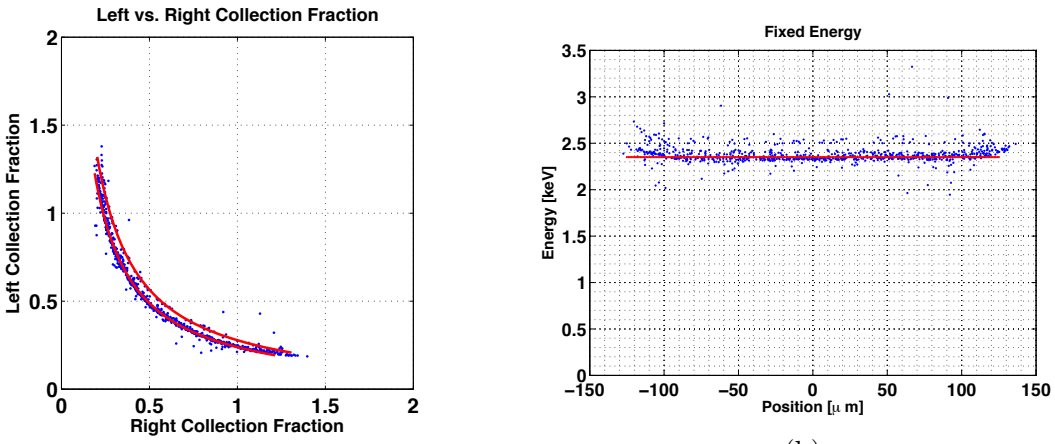
(a) (b)

Figure 5.30: (a) Left *vs.* Right energy collection for direct Al hits. Best fits for K_α and K_β shown. (b) Banana plot transformed to Position *vs.* Energy coordinates using those fits. Scale Factor = 2.78 keV, Diffusion Length = 127 μm, Absorption Length = 135 μm.

filter in terms of linearity and resolution.

5.10 So What About The Aluminum Fin?

Figure 5.30 shows the 'banana plot' of direct Al hits reconstructed using the reduced heat capacity model from Section 5.6. Maximum likelihood estimates (MLEs) of the diffusion equation from Section 5.7 for K_α and K_β are shown in red. In this case, the scale factor is found to be E_0 = 2.78 keV at K_α with diffusion and absorption lengths of L_{diff} = 127 μm and L_{abs} = 135 μm. Since the total energy of K_α radiation is 2.62 keV, having 2.78 keV of that absorbed into the film is non-physical. From Section 5.8, we expect about 11% of each x-ray incident on the Al to be lost as high frequency phonons with about 5% of the loss recovered by reabsorption. This leads to an expectation that the scale factor for the banana plot should be 2.35 keV. Removing the scale factor from the MLE algorithm and fixing it at 2.35 keV yields Figure 5.31. Particularly in the energy *vs.* position plot it is clear by eye that this is not the best fit; the tails turn up. However the model used to generate templates was shown in Section 5.6 not to be perfectly linear on the scale of the changes needed to make



(a) (b)

Figure 5.31: Figure 5.30 with a scale factor of 2.35 keV enforced in the maximum likelihood estimate. Scale Factor = 2.35 keV, Diffusion Length = 137 μm, Absorption Length = 120 μm.

Figure 5.31 the best fit. Estimates of diffusion and absorption length are consistent at the 10% level between the estimators, but the shape of the curve is too sensitive to non-linearities in the energy scale to make good estimates of of phonon energy losses.

5.11 Summary and Future Plans

Template fitting and simulation have improved our understanding of quasiparticle devices and allowed better reconstructions of energy, but there is still much work to be done. Time-domain optimal filtering has the potential to allow energy resolution of Cl K-shell events down to 1 eV, but its benefit is negated in the present experiment by high levels of comparatively white environmental noise. The existence of weakly linked waterfalls at the Al/W overlaps makes reconstructing linear energy scales difficult, particularly at low energies. Before making further refinements to the device model, we need detailed data about the temperature and current responses as well as the heat capacity of our W films. Such information is available through measurements of complex impedance as a function of bias point[40].

The present generation of Al films are found to have diffusion and absorption

lengths in the range of 120-140 μm. These values are likely to be limited by the 0.3 μm thickness of the films and band gap inhomogeneities tied to thickness variations. Improved fabrication methods which allowed better draping of W over thicker Al films and limited trapping in the Al films themselves would not only improve the energy resolution of these devices, but would allow for longer Al fins in the main SuperCDMS experiment. Greater Al coverage would decrease phonon collection times and improve yield rejection. Look for future papers on these topics by J. J. Yen.

Appendix A

Calculations For Si-Ge-Si Detector Model

This is a model of the small-bias resistance of a pair of Si-Ge diodes facing each other in series with a germanium resistor in between. See Section 4.3.1 for the motivation behind this model. To model the current for a given bias V_b, we combine the Shockley equation with Ohm's Law.

$$I = I_s \left(e^{qV_D/\eta k_B T} - 1\right) = \frac{V_b - 2V_D}{R} \tag{A.1}$$

where I_s is the saturation current, V_D is the voltage drop across the diode, q is the fundamental charge, η is the so-called 'ideality factor' of the diode, $k_B T$ is the temperature in energy units and R is the resistance. For small bias, the forward and reverse diode are treated as identical, so the total loss of voltage due to diodes is $2V_D$. A perfect diode has ideality factor 1; diode manufacturers like to keep their ideality factors to $\eta < 2$. For this explanation, we will adopt the conventional definition of a 'thermal voltage' $V_T \equiv \frac{k_B T}{q}$. Equation A.1 is transcendental, so it cannot be solved analytically for arbitrary input values. However it can be solved very quickly by a computer given specific values, so let's dive in.

A.1 Carrier Density

(This section is mostly adopted from [21].)

Semiconductor device properties are largely determined by carrier densities in various regions of the device. For an intrinsic (no dopants) semiconductor, the equilibrium density n_0 of electrons in the conduction band (and holes in the valence band) can be written

$$n_0 = \int_{E_c}^{\infty} n(E)dE = \int_{E_c}^{\infty} g_c(E)f(E)dE \tag{A.2}$$

where $g_c(E)$ is the density of states in the conduction band and $f(E) = \frac{1}{1+e^{(E-E_F)/k_BT}}$ is the Fermi function for occupation probability. In theory, the limit should stop at the top of the conduction band, but in practice the Fermi function is so small by then that ∞ may be substituted with impunity. For the density of states in an isotropic crystal, we assume each energy surface is a sphere in momentum-space

$$\begin{aligned} k &= \sqrt{2m^*E} \\ dk &= \frac{k}{m^*}dE \\ g_c &= \frac{2_{(spin)}4\pi k^2 dk}{(2\pi\hbar)^3} \\ g_c &= \frac{\sqrt{2m^*(E-E_c)^3}}{\pi^2\hbar^3}dE \end{aligned} \tag{A.3}$$

Assuming that $(E_c - E_F) >> k_BT$, we can set $f(E) \approx e^{-(E-E_F)/k_BT}$ in the integration region. Integrating Equation A.2, we now get:

$$n_0 = 2\left(\frac{m^*k_BT}{\hbar}\right)^{3/2} e^{\frac{(E_F-E_c)}{k_BT}} \equiv N_c e^{\frac{(E_F-E_c)}{k_BT}} \tag{A.4}$$

where N_c is an effective density of states in the conduction band. A similar analysis for the hole density p_0 gives

$$n_0 = 2\left(\frac{m_h^* k_B T}{\hbar}\right)^{3/2} e^{\frac{(E_v - E_F)}{k_B T}} \equiv N_v e^{\frac{(E_v - E_F)}{k_B T}} \tag{A.5}$$

where N_v is the effective density of states in the valence band. A completely intrinsic semiconductor has its Fermi energy near the middle of its band gap; for near-intrinsic devices contaminated with unknown concentrations and species, the Fermi level is essentially a random number and should be eliminated from calculations wherever possible. We therefore define an *intrinsic* carrier density n_i by arguing that in the intrinsic semi-conductor $n_0 = p_0$.

$$n_i = \sqrt{n_{0,i} p_{0,i}} = \sqrt{N_c e^{\frac{(E_F - E_c)}{k_B T}} N_v e^{\frac{(E_v - E_F)}{k_B T}}} = 2\left(\frac{\sqrt{m_e^* m_h^*} k_B T}{\hbar}\right)^{3/2} e^{\frac{(E_c - E_v)}{k_B T}} \tag{A.6}$$

The quantity $(E_c - E_v)$ is commonly called the 'gap energy' and goes by E_g in most textbooks. Note that the right-hand side of Equation A.6 is true regardless of the Fermi energy and so the whole equation is true even for non-intrinsic carrier densities. The Law of Mass Action says that

$$n_0 p_0 = n_i^2 \tag{A.7}$$

for all non-degenerate semiconductors in thermal equilibrium and fully justifies the capital letters by its usefulness.

A.1.1 Adding Impurities

In the case of a doped semiconductor, we can no longer say that $n_0 = p_0$, but we can keep the Law of Mass Action (Equation A.7) and electrical neutrality:

$$\rho = q(p_0 - n_0 + N_d^+ - N_a^-) = 0 \tag{A.8}$$

Using the Law of Mass Action to eliminate p_0 (or n_0) and solving the resulting quadratic equation, we get

$$n_0 = \frac{n_i^2}{n_0} + N_d^+ - N_a^- \quad \rightarrow \quad n_0 = \frac{N_d^+ - N_a^-}{2} + \sqrt{\left(\frac{N_d^+ - N_a^-}{2}\right)^2 + n_i^2} \quad \text{(A.9)}$$

$$-p_0 = -\frac{n_i^2}{p_0} + N_d^+ - N_a^- \quad \rightarrow \quad p_0 = -\frac{N_d^+ - N_a^-}{2} + \sqrt{\left(\frac{N_d^+ - N_a^-}{2}\right)^2 + n_i^2} \quad \text{(A.10)}$$

If we take $N_d > N_a$ as convention, this would give us majority carrier density n_0 and minority carrier density p_0 provided we knew what fraction of the impurities are ionized. This is just another Fermi function $f_{free} = \frac{P_{free}}{P_{free}+P_{bound}}$, but we need to account for having multiple bound states (namely, spin up and spin down). Consider a donor with bound state $-E_d$ (setting the bottom of the conduction band to $E_c \equiv 0$ for convenience). The probability for a bound state to be unoccupied (ionized) is

$$f_{free} = \frac{e^{\frac{0}{k_B T}}}{1e^{\frac{0}{k_B T}} + 2e^{\frac{-(-E_d)}{k_B T}}} = \frac{1}{1 + 2e^{\frac{E_d}{k_B T}}} \quad \text{(A.11)}$$

Since the donor band is partially filled, the Fermi energy will be fixed at $-E_d$ and all acceptors will be ionized. From Equation A.4, we know that $e^{\frac{-E_d}{k_B T}} = \frac{n_0}{N_c}$. Equation A.9 becomes

$$\begin{aligned} n_0 &= \frac{n_i^2}{n_0} + \frac{N_d}{1 + 2e^{\frac{E_d}{k_B T}}} - N_a \rightarrow \\ n_0 &= \frac{\frac{N_d}{1+2e^{\frac{E_d}{k_B T}}} - N_a}{2} + \sqrt{\left(\frac{\frac{N_d}{1+2e^{\frac{E_d}{k_B T}}} - N_a}{2}\right)^2 + n_i^2} \end{aligned} \quad \text{(A.12)}$$

There is a tendency among makers of highly-doped devices to assume that essentially all majority carriers will come from ionized donors and neglect $p_0 = \frac{n_i^2}{n_0}$ in this equation. Since we have gone to some effort to minimize N_d, we expect significant minority carrier concentrations. We will therefore keep p_0 and the quadratic solution given in Equation A.9.

A.2 Mobility

Most of the variation in the germanium resistance $R = \frac{1}{nq\mu}$ comes from the free carrier density n. Just to get the scale right, we need to give the mobility μ a brief look. Mobility is a measure of how easy it is to develop a large carrier drift velocity with a small electric field ($\vec{v}_d = \mu\vec{E}$). Applying the Drude Model, where each collision resets the carrier velocity to a random direction and defining a mean free path l and mean time between scatters τ:

$$\begin{aligned} \frac{l}{\tau} &= \mu\frac{F}{q} \\ F = m^*a \quad &; \quad l \approx a\tau^2 \\ \frac{a\tau^2}{\tau} &= \mu\frac{m^*a}{q} \\ \mu &= \frac{q}{m^*}\tau \end{aligned} \tag{A.13}$$

Mobility limits come from anything that deflects carriers. If there are multiple independent candidates for this role, their effects are added by summing the scattering frequencies $1/\tau$. This is formally known as Matthiesen's Rule.

$$\frac{1}{\mu} = \sum_i \frac{1}{\mu_i} \tag{A.14}$$

A.2.1 Impurities and Defects

Except for very pure crystals at sub-Kelvin temperatures, the thermal velocity $v_T = \sqrt{\frac{3k_BT}{m^*}}$ is much larger than the drift velocity v_d. To first order, carrier paths are straight lines of length $l = v_T\tau$. In three dimensions, a particle traveling through a field of hard spheres has mean free path $l = \frac{1}{n_d\sigma_s}$, where n_d is the density of impurities and $\sigma_s = \pi r_s^2$ is the scattering cross-section. The exact cross-section is impurity- and carrier-dependent, but to get a sensible value, we consider a charged particle of mass m^* approaching another heavy charged particle from a great distance through dielectric medium ϵ. To have a "significant scattering event", we might argue

that half of the initial kinetic energy must be transformed into electric potential.

$$\begin{aligned} KE_i &= 2PE(r_s) \\ \frac{1}{2}m^* v_T^2 = \frac{3}{2}k_B T &= 2\frac{1}{4\pi\epsilon}\frac{q^2}{r_s} \\ r_s &= \frac{q^2}{3\pi\epsilon k_B T} \\ \frac{1}{\pi r_s^2} &= 9\pi\epsilon^2\frac{(k_B T)^2}{q^4} \\ \mu_{impurities} = \frac{q}{m}\frac{l}{v_T} &= \frac{q}{m}\frac{1}{\pi r_s^2 n_d}\sqrt{\frac{m^*}{3k_B T}} \\ \mu_{impurities} &= \frac{3\sqrt{3}\pi\epsilon^2}{n_d\sqrt{m^*}q^3}(k_B T)^{3/2} \end{aligned} \tag{A.15}$$

Line dislocations through the crystal are thought to be a major source of scattering in our high-purity crystals. To model this, we use the same scattering radius as above, but solve a two dimensional problem where $l = \frac{1}{r_s(EPD)}$ and (EPD) is the etch pit density in cm^{-2} observed by etch-decorating the line dislocations where they leave the faces of the crystal. Grinding through the same math again gives

$$\mu_{defects} = \frac{\sqrt{3}\pi\epsilon\sqrt{k_B T}}{(EPD)\sqrt{m^* q}} \tag{A.16}$$

A.2.2 Lattice Interactions

At finite temperature, even a structurally and chemically pure crystal will have limited mobility due to electron-phonon interactions. Our semi-classical treatment gives out here because it is unclear how to assess the cross-section σ_{ph} of a phonon. A quantum mechanical treatment of electron-phonon scattering[44] gives

$$\mu_{lattice} = \frac{q\rho\lambda_T^3 v_s^2}{2\sqrt{6\pi}m^* E_{def}^2} \tag{A.17}$$

where ρ is the density, $\lambda_T \equiv \sqrt{\frac{2\pi\hbar^2}{m^* T}}$ is the thermal wavelength, v_s is the speed of

sound, and E_{def} is the deformation potential[1]. Deformation potential is notoriously sample-specific, but a typical value of 10 eV for annealed Ge gives the vendor's target mobility of 22,000 $^{cm^2}/_{Volt-sec}$. In fact, for the range of impurities and dislocations expected in SuperCDMS detector crystals, electron-phonon interactions completely dominate mobility at elevated temperatures. Figure A.1 shows the estimated mobility due to the interactions described here for typical crystal parameters ($N_d = 10^{11}/cm^3$, EPD = $8000/cm^2$, E_{def} = 10 eV). Impurity and defect mobilities are included here in case this model is used for other kinds of crystal.

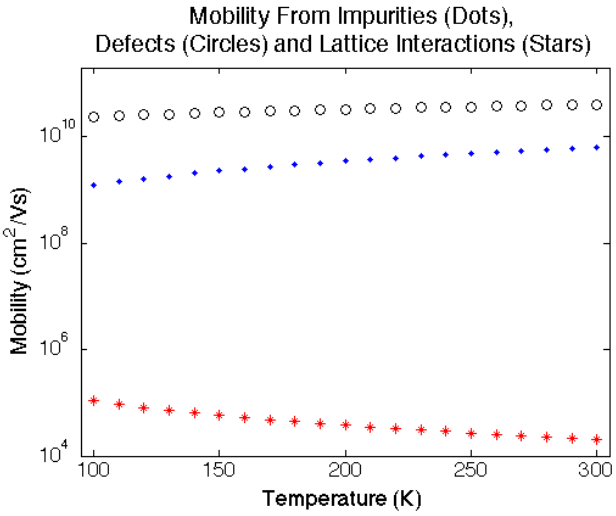


Figure A.1: Mobility estimates from various interactions in the temperature range in interest to the Arrhenius studies outlined in Section 4.3.

A.3 Germanium Resistance

We can calculate the resistance of the Ge crystal, remembering that majority and minority carriers both carry current.

$$\rho_{Ge} = \frac{1}{q(n_{maj}\mu_{maj} + n_{min}\mu_{min})} \tag{A.18}$$

[1]Note for engineers: Deformation potential is the bulk modulus normalized per atom, $E_{def} = \frac{G}{n}$.

$$R_{Ge} = \frac{\rho_{Ge} L}{A} \tag{A.19}$$

Here ρ_{Ge} is the resistivity, L is the length of the current path (equal to the thickness) and A is the surface area of the crystal face. Later, we will limit A_{diode} to the fraction of the face which is covered in aluminum, but the electric field spreads quickly compared to the crystal thickness, so $A_{resistor}$ is the entire (circular) surface area.

A.4 Diode Properties

If we were making a dedicated germanium resistance measurement, the Appendix could end here. However, the p-n diode formed at the boundary between the Ge crystal and the aSi barrier layer may drop significant voltage, increasing the apparent resistance of the device. In a doped semiconductor, the Fermi energy is fixed at the partially filled dopant band. When semiconductors of different impurity types are brought into contact, carriers move to equalize the Fermi energies[2]. This creates a built-in voltage V_{bi} equal to the difference in initial Fermi energies. Since all the free carriers near the junction have drifted to the other side, a 'depletion region' with no free charge forms. For small bias voltage, this makes the p-n diode a good insulator. Applied voltage V_D raises or lowers the diode barrier. The current density $\vec{j}$ is then a competition between carriers diffusing from high to low concentration across the depletion region and drifting in the applied electric field $\vec{E}$ set up by oppositely-polarized ions. This competition only occurs for minority carriers; the two effects conspire to keep majority carriers out of the depletion region. For simplicity, we will follow electrons traveling from the p-type aSi to the n-type Ge and add the comparable terms for holes at the end.

$$\vec{j}_e = \vec{j}_{drift} - \vec{j}_{diff} = nq\vec{v}_{drift} - eD_e \nabla n \tag{A.20}$$

where $\vec{v}_{drift} = \mu\vec{E}$ is the drift velocity and $D_e = \mu V_T$ is the diffusion coefficient. Recall that we defined a thermal voltage $V_T \equiv \frac{k_B T}{q}$ at the beginning of the chapter.

[2]Remember that Fermi energy is the energy needed to add/remove a carrier

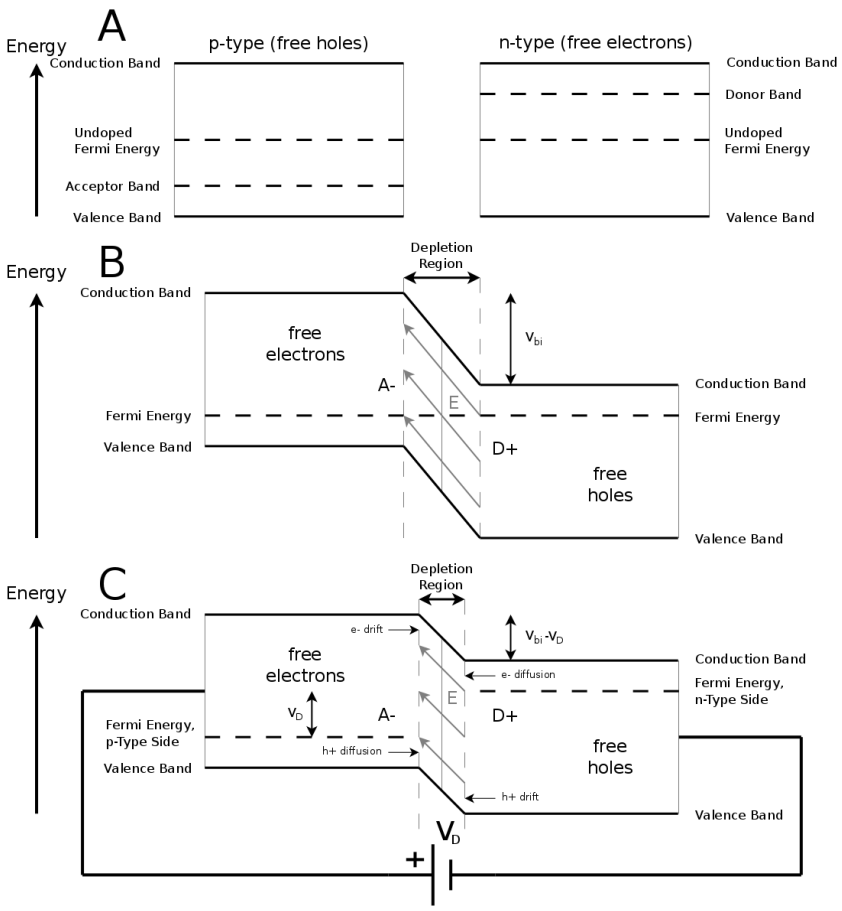


Figure A.2: Brief explanation of the diode effect. The p-type and n-type semiconductors in A are brought together in B, causing a matching of their Fermi levels which are otherwise fixed at their respective impurity bands. Gradients in the conduction and valence bands require rearranging charge to maintain the built-in voltage $V_{b,i}$. This creates a permanent charge-free region. The application of a diode voltage V_D in C raises or lowers the barrier.

Plugging in $\vec{E} = -\nabla\phi$ gives

$$\frac{k_B T}{q}\frac{\nabla n}{n} = \frac{k_B T}{q}\nabla(ln(n)) = -\nabla\phi$$
$$\frac{k_B T}{q} ln(n) = -\phi + \text{const}$$
$$n = n_0 e^{\frac{-e\phi}{k_B T}} \tag{A.21}$$

In the bulk of the material, $\phi \equiv 0$; at the boundary, $\phi \equiv V_D$ by construction. Taking a linear approximation of the carrier gradient:

$$\nabla n = \frac{n_{bulk} - n_{edge}}{w} = \frac{n_0}{L}\left(1 - e^{\frac{-eV_D}{k_B T}}\right) \tag{A.22}$$

where w is the relevant depletion width, which we will discuss momentarily. Plugging Equations A.21 and A.22 into Equation A.20:

$$\vec{j_e} = \frac{qDn_0}{w}\left(1 - e^{\frac{-eV_D}{k_B T}}\right) \tag{A.23}$$

To calculate the total current, we need to sum $\vec{j}_{e,Si}$ and $\vec{j}_{h,Ge}$, which means we need to deal with those depletion widths w. We will simplify the analysis by taking the full-depletion approximation depicted in Figure A.2; the depletion region consists of opposed regions with zero charge density on either side of the interface. Outside this region, the charge density is considered to immediately return to its bulk value. The total depletion width is the sum of Si and Ge widths.

$$w = w_{Ge} + w_{Si} \tag{A.24}$$

The net charge in these two regions will induce an electric field which increases linearly up to the interface. Since no charge is trapped at the interface itself, the field must be continuous across the join. Working Gauss' Law from both directions:

$$\begin{aligned} \vec{E}_{Ge,interface} &= \vec{E}_{Si,interface} \\ \frac{qN_d w_{Ge}}{\epsilon_{Ge}} &= \frac{qN_a w_{Si}}{\epsilon_{Si}} \end{aligned} \tag{A.25}$$

It should be noted that this derivation assumes uncompensated semiconductors for simplicity of notation. In the full analysis, replace $(N_d \rightarrow N_{d,Ge} - N_{a,Ge})$ and $(N_a \rightarrow N_{a,Si} - N_{d,Si})$. Combining Equations A.24 and A.25 gives:

$$w_{Ge} = w\frac{N_a}{N_a + \frac{\epsilon_{Si}}{\epsilon_{Ge}}N_d} \tag{A.26}$$

$$w_{Si} = w\frac{N_d}{N_d + \frac{\epsilon_{Ge}}{\epsilon_{Si}}N_a} \tag{A.27}$$

Integrating the electric fields $\vec{E}_i = \frac{qN_i|w_i - x|}{\epsilon_i}$ gives the total potential.

$$V_{bi} - V_D = \frac{qN_d w_{Ge}^2}{2\epsilon_{Ge}} + \frac{qN_a w_{Si}^2}{2\epsilon_{Si}} \tag{A.28}$$

Applying your favorite equation cruncher to this will eventually yield:

$$w_{Ge} = \sqrt{\frac{2(V_{bi} - V_D)}{q}\frac{\epsilon_{Ge}N_a}{\epsilon_{Si}N_d}\left(\frac{N_a}{\epsilon_{Si}} + \frac{N_d}{\epsilon_{Ge}}\right)^{-1}} \tag{A.29}$$

$$w_{Si} = \sqrt{\frac{2(V_{bi} - V_D)}{q}\frac{\epsilon_{Si}N_d}{\epsilon_{Ge}N_a}\left(\frac{N_a}{\epsilon_{Si}} + \frac{N_d}{\epsilon_{Ge}}\right)^{-1}} \tag{A.30}$$

Although its low impurity density may lead to Ge depletion widths measured in millimeters, there is plenty of bulk substrate to allow for this. On the other hand, there is a very real possibility of the Si depletion width exceeding its 40 nm thickness.

A.5 Shockley Equation

Bringing this all together, and combining pre-factors for convenience gives:

$$\boxed{I = I_{sat}\left(e^{\frac{qV_D}{k_BT}} - 1\right)} \tag{A.31}$$

$$\boxed{I_{sat} \equiv qA_{diode}\left(\frac{D_{Ge}p_{0,Ge}}{w_{Ge}} + \frac{D_{Si}n_{0,Si}}{w_{Si}}\right)} \tag{A.32}$$

Equation A.31 is the Shockley equation. Fortunately in the process of deriving it we have also derived the saturation current needed for our computation.

A.6 Carrier Masses, Band Gaps and Dopant Bands

Before we can construct model Arrhenius plots, we need to tie up a few loose ends. First, we have referred repeatedly to effective masses m_e^* and m_h^* without giving their values. These are given below.

Use	Quantity	Germanium	Silicon
Density	Effective Electron Mass	$0.55\mathrm{m}_e$	$1.08\mathrm{m}_e$
	Effective Hole Mass	$0.37\mathrm{m}_e$	$0.81\mathrm{m}_e$
Mobility	Effective Electron Mass	$0.12\mathrm{m}_e$	$0.26\mathrm{m}_e$
	Effective Hole Mass	$0.21\mathrm{m}_e$	$0.386\mathrm{m}_e$

Effective masses arise from taking the second derivative $\frac{d^2E}{dk^2}$ at local minima of the $E(k)$ function, to create a quantity that obeys $\frac{dE}{dx} = m^*_{eff}\frac{d^2x}{dt^2}$. The reason for using different masses for different applications is that each carrier has multiple minima which need to be accounted for cumulatively. In density of states calculations, a geometric average is needed, so $m^*_{eff} = \left(\prod_{n=1}^{N} m_n\right)^{1/N}$. In mobility applications, each band minimum contributes a separate term to $\mu = \frac{q}{m}E$, so the effective masses are added in parallel, $\frac{1}{m^*_{eff}} = \sum_n \frac{1}{m_n}$.

The following quantities will also be useful to anyone trying to replicate this work.[3]

Quantity	Symbol	Germanium	Silicon
Density	ρ	5.32 g/cm^3	2.33 g/cm^3
Atomic Density	N	4x10^{22}/cm^3	5x10^{22}/cm^3
Dielectric Constant	ϵ	16 ϵ_0	11.7 ϵ_0
Speed of Sound	v_s	5400 m/s	8433 m/s
Deformation Potential	E_{def}	10-40 eV	20-40 eV

[3]Deformation potential is highly sample-dependent. These values represent the range expressed in a brief literature search. In practice, one should set E_{def} to match the vendor's mobility measurements, assuming they are trustworthy.

The band gaps of semiconductors are typically temperature dependent, following the model:

$$E_g(T) = E_g(0) - \frac{AT^2}{T+B} \tag{A.33}$$

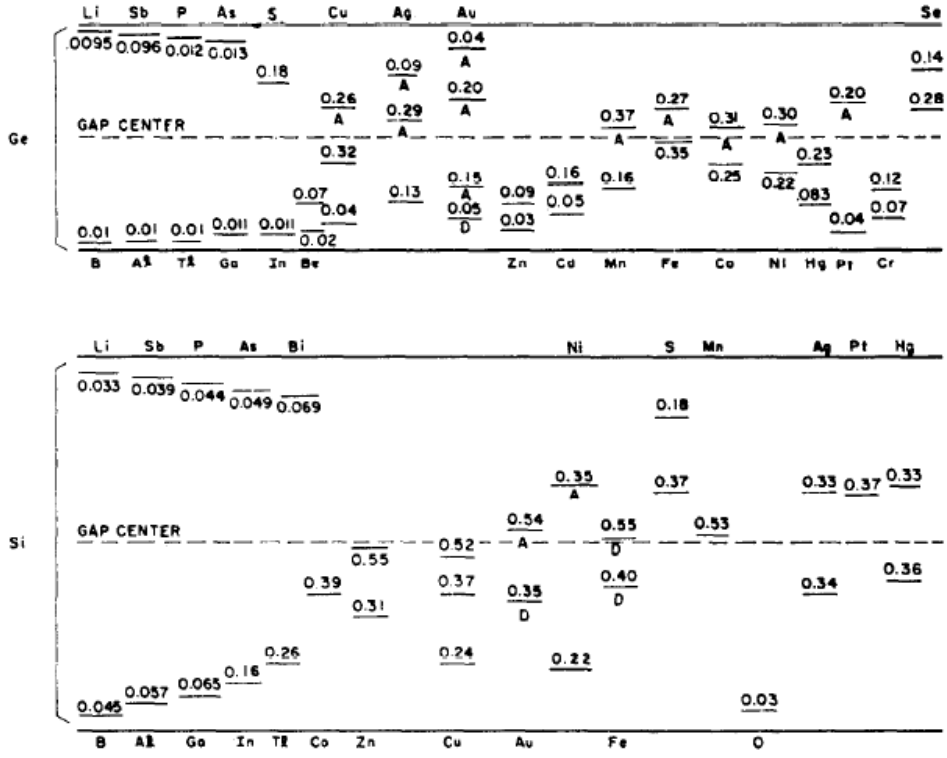


Figure A.3: Impurity energy levels in Si and Ge at 300 K taken from [24]

Empirically, these constants have been determined to be $E_g(0) = 0.7437$ eV, A = 4.77 meV/K, B = 235 K for germanium and $E_g(0) = 1.166$ eV, A = 4.73 meV/K, B = 636 K for silicon. Energy levels for a wide variety of impurities in Si and Ge at 300 K are helpfully tabulated by [24]. The relevant part of their Fig. 3 is reproduced in Figure A.3. These levels all have small temperature dependencies, but we do not have the sensitivity to detect them. Instead, this serves as a rough guide to what kinds of impurities we an rule out by our measurement. If the impurity density starts to approach the density of states N_c or N_v, these discrete levels will spread out into

bands of finite size. This is not generally significant unless (perhaps in the case of our heavily, shallowly doped aSi layer) the impurity intersects the conduction band. In this case, the material ceases to behave like a semiconductor and instead behaves like a metal. If this happens, we might see Arrhenius plots for the pure Ge despite the aSi layer.

Appendix B

Helium Liquefier

Since its inception, the ability to do low temperature physics has been tied to the availability of liquid helium. With a liquefaction temperature of 4.2 K and no freezing point at common laboratory pressures, helium allows us to store and transfer cooling power when every other element has frozen to the walls of the storage vessel. Although 'dry' cryocoolers with small, closed-cycle reservoirs of helium are growing in sophistication and popularity, acquiring liquid helium is still very much a part of the life of most low temperature physicists.

The Federal Helium Reserve, established in the 1920s when an air travel industry dominated by zeppelins was a real possibility, became an important strategic reserve for the kinds of science needed to fight the Cold War (space flight, nuclear detection, etc.). In 1996, after it was clear that the Cold War was really, truly, over for real, Congress directed the Department of the Interior to sell off the massive, indebted Reserve. This led to more than a decade of "priced to sell" helium available to anyone and an increased reliance on helium for a wide variety of research and medical applications. However, it killed helium recycling efforts at many institutions, including Stanford, and strongly dis-incentivized separation of new helium from natural gas wells, which are the primary worldwide helium source. Helium is produced in the Earth from alpha decays and accumulates in the same pockets as natural gas. Once vented to atmosphere, helium does not accumulate because it is not very strongly gravitationally bound to the Earth. The thermal velocity of a helium atom in the

upper atmosphere is about 1.2 $^{km}/_{s}$. This is much lower than the escape velocity of the Earth (11.2 $^{km}/_{s}$), but about 1 in every $e^{11.2/1.2} = 11{,}000$ escape attempts succeeds which is enough to bleed off any helium added to the atmosphere in a short time.

Although the Federal Helium Reserve is not empty at the time of this writing (2014), the authorization to keep the pumps running was set to run out in 2012. This led to extreme market pressure on the price of helium starting in late 2010. In 2011, Profs. Blas Cabrera and Hari Manoharan decided to resurrect a Linde 1400 helium liquefier belonging to Prof. Doug Osheroff shown in Figure B.1.

Figure B.1: Linde 1400 liquefier located in Varian room 015.

B.1 Principle of Operation

Like most refrigerators, the Linde 1410 cools gas by adiabatic expansion. A closed compressor (located in Varian 033) feeds helium gas at 230 psi into a series of heat exchangers (E30-E34 in Figure B.2). A pair of pistons (E37 and E39) provide additional compression at lower temperatures. The five heat exchangers each consist of

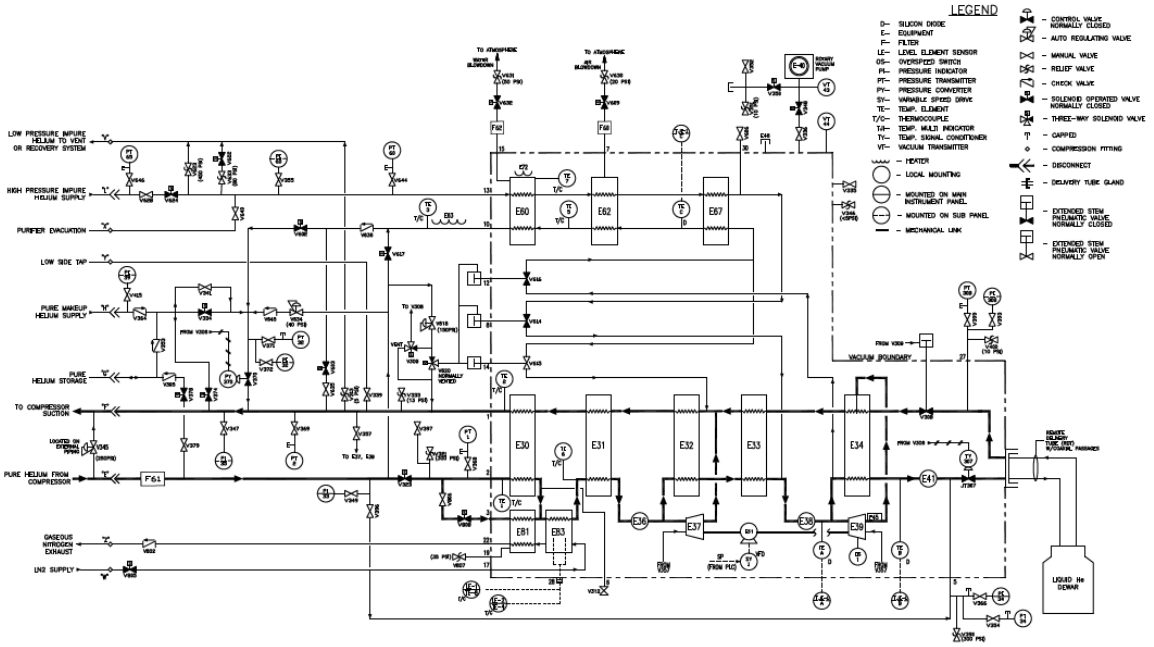


Figure B.2: Flow diagram of Linde 1410 liquefier with on-board purification. The primary pure flow path is bolded.

a long coil of metal tubing through which the incoming gas flows. Returning gas at low pressure flows through baffles around the incoming pipes, providing ample heat exchange through a low impedance return path. The heat exchangers are shown in cross-section in Figure B.3.

At the outlet of E34, the gas is cooled to about 7 K. At this point, the cooling power of reversible adiabatic expansion is nearly exhausted because Van der Waal terms in the Ideal Gas Law have become important. The cooled gas is sent through a Joule-Thompson valve (JT307, just after E41). The Joule-Thompson process is also a form of adiabatic expansion, but rather than expand isentropically, gas is forced through a narrow constriction into a region of lower pressure with constant kinetic energy. In this case, enthalpy is held constant. At low temperatures, this allows the gas to do considerable work without proportionally increasing in volume[46]. This final loss of energy mostly liquefies the helium. It then enters the central shell of a three-walled Return-Delivery Tube (RDT, sometimes called a 'triple tube'). At

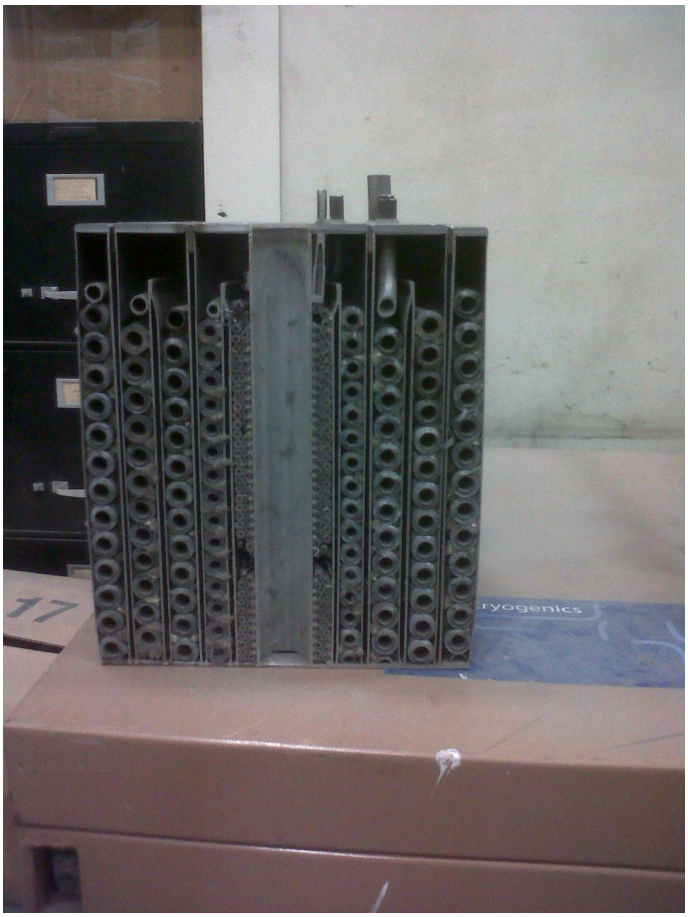


Figure B.3: Cross-sectional view of Linde 1400 heat exchangers. This piece is half of the cylindrical arrangement, but also half of the height. All heat exchangers are shown, but they are all twice this tall.

the other end of the RDT, liquid helium falls into the 500 L mother dewar, from which rolling dewars can be filled to cool experiments. Cold gas evaporated from the mother dewar returns through the second shell of the RDT through V308 to cool the heat exchangers of the liquefier. The third shell of the RDT is the standard vacuum-walled, radiation-baffled barrier used in most cryogenic applications to keep the hot surroundings at bay.

B.2 Purification

Because everything except helium freezes at liquid helium temperatures, helium liquefiers are susceptible to ice plugs. Many systems have stand-alone purifiers which cryogenically separate impurities out of the helium stream coming back from experiments. Our Linde 1410 has an on-board purifier which uses cooling power from the fourth heat exchanger to drive the separation cold traps.

All experiments that receive liquid helium from the liquefier system must have their helium exhaust lines plumbed into the recovery system which leads to a 500 SCF gas bag in Varian 032. The bag will only hold 20 liquid liters (lLeq) of boil-off, so a Bauer VGR120II 6 SCFM compressor located across the hall empties the bag to high pressure storage up to a maximum of 2000 psi. The storage tanks are located outside the Varian Machine Shop in a recessed pit. The 12 long tanks with their valves facing the Main Quad are impure. The eight long tanks on top of them and all of the upright tanks are pure buffer tanks which are only rated for 500 psi. The Linde 1410 purifier needs at least 330 psi of back pressure to function. Given the volume of impure storage, this means that up to 550 lLeq of blow-off gas can be waiting for liquefaction at any one time.

When the purifier is activated, cold gas leaves the second piston (E38) through valve V616, cools a series of heat exchangers (E60, E62 and E67) and escapes to the low-pressure side through valve V602. Impure gas from high pressure storage enters the heat exchangers through valve V624, is cooled and feeds the low-pressure side through valve V615 just before E32. Purification occurs in batches, with a new cycle automatically initiated whenever the inlet pressure to the the pure-side compressor falls below its set-point, typically about 3 psig. When it is working well, the purifier makes a *click*-(3 sec inhale)-*clonk* sound about every 15 seconds, although it often runs slower.

B.3 Current Status and Future Plans

In May 2013, oil from the Linde 1400 reciprocating pure compressor penetrated the outlet filter and contaminated all of the piping into room 015 as well as the heat exchangers in the liquefier and probably the purifier. This created additional impedance which rendered the system inoperable. As of this writing (early 2014), attempts to physically and chemically remove the oil have failed. The liquefier has been decommissioned for the foreseeable future. However, with the rising cost of helium, a regional liquefier overseen be the University to supply labs in the Varian, Moore and McCullough buildings has gained traction. Drawing on our experience with the Linde 1410 liquefier and Linde 1400 compressor, the system will likely employ a Linde 1610 liquefier with onboard purification and a pair of RS scroll compressors. With a plumbed in liquid nitrogen supply for pre-cooling (E81 and E83 in Figure B.2, unused in the current system), the new system will probably produce 70-90 L per hour, compared to 4 L/hr in the existing system when it was working well. Such increase in scope allows the hiring of a full-time support person rather than a team of grad students with very limited bandwidth.

During its period of operation as a shared facility in 2012-2013, the Varian liquefier saved over 9000 L of liquid helium, replacing over $80,000 of purchase orders. This does not include the 3000 L of helium which were lost and replaced by new dewar purchase. However, the use of inexperienced grad students on an aging system resulted in multiple trips by expert technicians to Stanford to conduct maintenance and repairs. This offset most of the savings, with most of the expense coming in the form of wages rather than parts. A regional system with expert staff on-site has the potential to operate much more efficiently while creating less drag on the science output of the groups involved. On a weekly basis, we have shown that 85% capture efficiency of boil-off gas is an achievable target across a building-wide recovery system. It is reasonable to expect, although not guaranteed, that the additional catastrophic losses which reduced our over-all collection efficiency to 75% could be mitigated by a new, professionally-built recovery system.

B.4 Instructions

The following are instruction manuals which have been helpful to new liquefier operators.

B.4.1 Startup

1. Pump down the vacuum jacket.
   (a) Engage jacket pump with **Vac Pump**'switch on side of liquefier.
   (b) Open to jacket with **V336** (back of liquefier)
   (c) Pump down to <50 mtorr on Hastings gauge below V806 in the front of liquefier. Depending on the initial jacket pressure, the process might take 20 minutes to 2 hours.
2. Make sure the **pure pressure** on the wall above the gas cylinders is > 175psi. If it is not, open the makeup tank until it is.
3. Purge the heat exchanger
   (a) Make sure the **Recovery** switch is OFF
   (b) Press and hold the **compressor start** button (green button on front of liquefier). Wait to hear the water flowing through the pipes in the ceiling.
   (c) Press the **engine start** button (black button on front, left side of the liquefier) QUICKLY. You do NOT want to get the pistons moving. The pressure read by **PI34** will rise quickly, so be prepared.
   (d) When the pressure is approaching 150 psi, switch the **Recovery** switch ON. Turn the compressor OFF.
   (e) Open **V312** to blow down the heat exchanger. Close **V312** when the pressure is around 10 psi. Switch **Recovery** OFF.
   (f) Repeat a-e two more times.

4. Press and hold the **compressor start** button. After the water begins to flow overhead, press the **engine start** button for a second or so. Manually start the flywheel and use the **speed control** knob (large black knob just under engine start button) to increase the speed to 100-120 RPM. (Sometimes you dont need to manually start the wheel, it will just spin by itself.)

5. Adjust the regulator **V370** (large vertical screw near bottom front) to put **PI35** (Magnehelic) at 45-55 inches of water (50-60 if starting cold).

6. Purge the purifier by flipping the **Purifier** switch on the right side of the liquefier, allowing it to fill with gas (a few seconds), and turning it off. Do this three times. Leave it off when done.

7. Within the next half hour, the engine speed will become faster, so you need to slow it down. If it gets over 250 RPM, the wheel triggers the **overspeed protection** switch, which shuts down the liquefier. If this happens, use a piece of metal with a slight hook or a stiff piece of wire to reach under the left side of the flywheel and pull the switch back toward you.

8. Wait for about 2 hours until the engine temperature ($\mathbf{T}_A$) is below 80K.

   (a) Open **V308** (in front of the flywheel) three and a half turns.

   (b) Open the JT needle valve (**V307**) by about 10 mil. (The JT valve doesn't close to zero, so pay attention to where you started.)

   (c) If the 500L dewar is warm, attach the surgical tubing between the **dewar vent** (thin copper tube coming from top of mother dewar) and **V339** AFTER flushing both lines (V339 is always over atmosphere). Adjust the impedance on the dewar vent to >25.

9. Now warm gas is flowing into the 500L dewar and making LHe evaporate, so the **pure pressure** will increase rapidly. If it rises to above 230 psi, open a **low-pressure tank** (upper long cylinders outside the Machine Shop) to lower the pressure. If the pressure goes above 245 psi or below 145 psi, the liquefier

will shut itself down. It is helpful to keep a marker outside on the pure tanks to record the pressure of any tank that you close.

10. Now the JT temperature ($\mathbf{T}_B$) should start to drop. It will drop much faster if $P_{pure} > 200$ psi. The engine temperature ($\mathbf{T}_A$) will rise up a little and then cool back down.

    (a) If the 500L dewar is warm, it may take several days and about 10 bottles of helium gas to cool it. Keep the dewar vent at > 25 SCFH until the level meter reads 12.2. (12.4 is cold), then gradually shut off the impedance.

    (b) If the dewar is warm, $\mathbf{T}_A$ should be ∼20K, and $\mathbf{T}_B$ should be ∼7K. Then its time to turn on the purifier.

    (c) When cold, **PI33** (right side on the front, about 2' off the ground) should not go below 220psi. Slowing the engine will bring it up somewhat, but it is better to increase the pure pressure. Often you can start the purifier anyway and PI33 will come up on its own.

B.4.2 Turning on the Purifier (manual 4.7.1)

1. Make sure **V635** (black valve with white painted line on back of liquefier) is set to about 4 full turns open and valve **V628** (stiff needle valve on back of liquefier) is set to 7½ turns open. V635 controls the flow of cooling gas to the purifier. V628 controls the impure supply.

2. Check that **PI35** is around 60-80 inches of water. If the pressure is too high, there wont be enough impure helium sent to the purifier; if the pressure is too low, the purifier heat exchanger draws too much cold gas and the engine warms up fast. At this temperature, it takes at least 10 minutes for changes in valve **V370** to affect PI35.

3. If the liquefier is running well by itself, close the JT valve **V307** (needle valve in front of the flywheel) by 1 mil. You will need to keep a little extra cooling power in the engine.

4. Purge the purifier a couple of times.

5. Switch the purifier ON. **TI61** (near the top on the backside of the liquefier) will drop slowly. Check that **PI32** (bottom right on the front) is ∼150. You may have to adjust **V370** if PI32 is far askew. After about 10 minutes, the temperature stops going down and a small brass cylinder on the backside of the liquefier (E63) will soon be coated with frost. $\mathbf{T}_A$ will climb; don't let it go above 30 K.

6. Switch the **Regen** switch (on the side) ON and use the wire to hold it. A heater (E72) is turned on in the purifier. **TI61** will rise slowly, and air will be released through valve **V609** (solenoid valve zip-tied to back of liquefier). After 40 seconds, **V609** is closed automatically and **V632** is open. This is where water will be blown out the copper tube with the Swageloc fitting in the far back near the ground.

7. Wait until **TI61** is above 275. Release the **Regen** switch. Turn off the purifier. Then purge the purifier 3 times. After each purge, wait until the gas is completely released and all the water, if there is any, drops out before starting the next purge.

8. Keep the purifier off. $\mathbf{T}_A$ will slowly decrease. When it is back to ∼20K, go back to step 4. After repeating this procedure three times, **TI61** should reach 30 and cycle between 30 and 70. This is what you want to happen.

9. Once the purifier starts cycling, the pure pressure will increase and the engine speeding will increase correspondingly. If it is not too fast (>200RPM), dont disturb it.

10. Now PI35 should be between 30 and 60, pure pressure between 160 and 190 psi, $\mathbf{T}_A$ ∼20K and $\mathbf{T}_B$ ∼7K. The liquid level of the 500L dewar will start to rise after ∼6-10 hours.

B.4.3 Manually Regenerate the Purifier ("Man Regen")

This needs to be done twice per day on the old Linde 1410 system or three times if it has been a while since the last cleanup. Newer systems can run indefinitely on automatic regenerations. To manually regenerate, follow steps 5 and 6 in the purifier startup, then restart the purifier. It should cool back down on its own. If no gas comes out when you purge the purifier, it means the line is blocked by ice. In this case, turn on the purifier, switch **Regen** on and keep it on until water flows out from the copper tube (it may take as long as an hour). Then do the purge. Since the system has been warmed up, you cant just turn on the purifier. Instead, you should go through the whole purifier startup sequence.

Appendix C

Code For Non-Linear Optimal Filter

This section contains an excerpt from the code used to generate templates and weighting matrices. Parameters for a fit of Run40 with half the Kittel heat capacity are used. This program assumes you have used a χ^2 minimization script to get best-fit parameters using templates taken from the raw data. All of the programs needed to get fit parameters (benSimPulseV1_2b.m), build templates and weighting matrices (benSimWeightV1_2b.m) and fit those templates to data (ETFTemplateMatchV1_2.m) can be found on nero in the /scratch/bshank/TES_Model directory. In order to use these scripts, data will also have to fed through DownSelectData.m, which cleans up baseline offsets and trigger delays, filters the raw pulses and down-selects to reduce computation time later.

C.1 Building Weighting Matricies

```
tic
%TES Parameters
Tc = 78.16; %Critical Temp in mK
Tw = 2.1875; %10-90% Width of Transition in mK
Rn = 3.09; %Normal Resistance in Ohms
```

```
Width = 250; %Microns
Length = 250; %Microns
Thick = 0.040; %Microns
N = 10; %Number of Strips to Model

%Waterfall Parameters
WaterFrac1 = .0453; %Fraction of Tungsten Waterfall Which Is Connected
WaterFracN = WaterFrac1;
LWater = 0.375; %Length of Waterfall in Microns

%SQUID Parameters
Bias = 0.115; %Bias Voltage in Volts
Rl = 7000; %Limiting Resistance in Ohms
Ib = Bias/Rl; %Bias Current in Amps
Rs = .029; %Shunt Resistance in Ohms
Rp = .0062; %Parasitic Resistance in Ohms
Gain = 1e5; %Effective KO15 Gain in Ohms

%OTHER Parameters
Tb = 41.892; %Fridge Base Temp in mK
Thot = 1.4; %1K Pot Temp in K
IRFrac = 0; %Fractional IR Absorption From 4.2 K
%Must Be Between 0 and 1
Qtemp = 0:2000; %Energies (eV) Of Template Comb
Qweight = 0:40:2000; %Energies (eV) Of Weighting Comb
NumPulses = 1024; %Number Of Pulses Per Weighting Matrix
dt = 0.2e-6; %Time Bin Size in sec
trace = 4096*dt; %Trace Length in sec
Select = 16; %Amount To Down Select Weighting Matrices
savefile = 'Run40Mar22CHalfWeights1.mat';
t = dt:dt:trace; %Time in sec
NumBins = length(t);

K = 3.685e-22 / N; %Electron-Phonon Coupling in W/mK^5
SB = 5.67e-20; %Stefan-Boltzmann Const in W/um^2/K^4
P_IR = SB*Width*Length*4.2^4 / N; %Blackbody IR absorption from 4.2 K bath
Pleak = IRFrac*P_IR; %Estimated IR loading on TES in Joules
```

```
L = 0.9846*2.44e-14; %Lorenz Number from Weidemann-Franz Law, WΩ/mK^2
JtoeV = 6.242e18; %Convert Joules to eV
k = 1.38e-26; % Boltzmann constant in J/mK
h_bar = 1.054e-34; % Planck constant J-s/rad


%Functions For Normal And Superconducting Heat Capacity
Delta = 1.76*k*(Tc+Tw); %Superconducting Band Gap of Tungsten in Joules
C_norm = @(T) 0.5*.00085 * Width*Length*Thick .* T / N;
SuperConst = 2.43*C_norm(Tc+Tw)*(Tc+Tw)^(3/2)*exp(1.76);
%Heat Capacity of Supercond W = 2.07 meV/mK^2/um^3
C_super = @(T) SuperConst .* T.^(-3/2) .* exp(-Delta./(k*T));
% 1/N Incorporated Into SuperConst

%Ginzberg-Landau Current-Coupling Constant
A = 3.52*sqrt(k*C_norm(Tc+Tw)/h_bar/(Rn/N)/(Tc+Tw)/JtoeV)/1.5;
AWater = A*WaterFrac1;

%Adjust Measured Parameters For Waterfall
WWater = WaterFrac1*Width;
RnWater = Rn * LWater/WWater;
if RnWater < Rn/2
Rn = Rn - 2*RnWater;
else
disp('Waterfall Resistance Exceeds Rn!!!')
end

Qstart = 500; %Time Bin of First Energy Deposit
QTime = 1.842e-6; %SECONDS! Time Constant of Short Tail

%Are You Only Calculating Templates For A White-Noise Chi^2 Matching?
TemplatesOnly = 0;

%Pre-Declare Big Matrices As Needed
FinalT = round(length(t)/Select);
simul_events = zeros(NumPulses,FinalT,length(Qweight));
if TemplatesOnly == 0 %i.e. full weighting matrices
```

```
    Templates1 = zeros(1,FinalT,length(Qtemp)); %Templates vs. Energy
    Cov1 = zeros(FinalT,FinalT,length(Qweight)); %Covariance Matrices
    Weights1 = zeros(FinalT,FinalT,length(Qweight)); %Weighting Matrices
else %i.e. compute templates only
    Templates1 = zeros(length(Qtemp),length(t));
    Errors1 = zeros(length(Qtemp),length(t)); %Error Vectors for Standard χ^2
end
Inoise = zeros(length(Qtemp),length(t));
intE2 = zeros(length(Qtemp),1); %Integral Energies of Templates (For Comparison)
intRes2 = zeros(length(Qweight),1); %FWHM Resolution of Integral Method
BestRes = zeros(length(Qweight),1); %Theoretical Optimal Resolution

%————————————————————————
%Begin Equilibrium Calculation
%————————————————————————

Tnew = (Tc-Tw)*ones(N,1); %Guess That TES Starts At Tc
Rnew = Rn/2/N * ones(N,1); %Guess That Each TES Bin Starts Halfway Up Its Transition
Inew = Ib*Rs/(sum(Rnew)+Rs+Rp); %Estimate Current
Cnew = C_norm(Tnew).*(N*Rnew/Rn) + C_super(Tnew).*(1-(N*Rnew/Rn));
%Estimate Heat Capacity
TW1new = Tc-Tw; %Leading Waterfall Temperature
RW1new = RnWater/2; %Leading Waterfall Resistance
TWNnew = Tc-Tw; %Trailing Waterfall Temperature
RWNnew = RnWater/2; %Trailing Waterfall Resistance PWater1 = K*Tc^5*RnWater/(Rn+2*RnWater);
%Estimate Quiescent Power in Waterfalls
PWaterN = K*Tc^5*RnWater/(Rn+2*RnWater);
T0 = zeros(1,1,N); %Pre-Declare Quiescent Temperature Matrix
C0 = zeros(1,1,N); %Pre-Declare Quiescent Heat Capacity Matrix

dtEquil = min([C_norm(Tc)*Rn/(4*N*L*Tc*JtoeV) dt])/2; %Time Steps For Equilibration
X = 100e-6/dtEquil; %Run Equilibration For 100 μsec At A Time
TSave = zeros(N,1); %Dummy Save Variables To Check If Values Have Changed in 100 μsec
TW1Save = 0;
TWNSave = 0;
Beta = exp(-pi); %Anti-Nyquist filter parameter is always e^-π in this step
%(Not needed, but useful for stability)
```

```
%While T Is Still Settling
while min(abs(Tnew-TSave)) > 1e-10 || abs(TW1new-TW1Save) > 1e-10 ...
  || abs(TWNnew-TWNSave) > 1e-10
   TSave = Tnew; %Reset 'Saved' Variables
   TW1Save = TW1new;
   TWNSave = TWNnew;

   for x = 1:X %Let temp settle for 100 μsec
       %Reset 'Old' Variables
      Told = Tnew;
      Rold = Rnew;
      Iold = Inew;
      Cold = Cnew;
      TW1old = TW1new;
      RW1old = RW1new;
      TWNold = TWNnew;
      RWNold = RWNnew;

      tempCond = L*Told./(Rn/N); %Temporary Conductance Matrix
      WaterCond1 = L*TW1old/RnWater; %Conductance of Waterfall Region
      WaterCondN = L*TWNold/RnWater;
      Pcond = tempCond .* (circshift(Told,-1) - Told);
        %Conduction Energy From Strip n+1 to Strip n
      Pcond(N) = 0; %No Conduction Off the End of the Strip
      PWater1 = Iold^2*RW1old; %Pre-Calculate Waterfall Heating
      PWaterN = Iold^2*RWNold;
      Pnoise = 0; %No Noise During Equilibration

      Told(1) = Told(1) + dtEquil*PWater1*JtoeV/Cold(1);
      Told(N) = Told(N) + dtEquil*PWaterN*JtoeV/Cold(N);
      Tnew = Told + dtEquil * (Pleak + Iold.^2.*Rold - K.*(Told.^5-Tb^5) ...
        + Pcond - circshift(Pcond,1) + normrnd(0,1,[N,1]).*Pnoise) .* JtoeV ./ Cold;
      Rnew = rfn(Tnew,Iold,Rn/N,Tc,Tw,A);
      %Massless Waterfall Instantly Adjusts To Change In Joule Heating
      TW1new = Tnew(1) + PWater1 / WaterCond1;
      RW1new = rfn(TW1new,Iold,RnWater,Tc,Tw,AWater);
```

```
        TWNnew = Tnew(N) + PWaterN / WaterCondN;
        RWNnew = rfn(TWNnew,Iold,RnWater,Tc,Tw,AWater);
        %Only Calculate New I_s Once Resistance of All Strips is Summed
        Inew = Ib * Rs/(sum(Rnew)+RW1new+RWNnew+Rp+Rs);
        Inew = (1-Beta)*Inew + Beta*Iold;

        if mean(Tnew) < Tc-Tw/2 %Numerical Oscillation Suppression Only Needed At Low Bias
            for i = 1:10
                Rnew = geomean([Rnew Rnew Rnew rfn(Tnew,Inew,Rn/N,Tc,Tw,A)],2);
                RW1new = geomean([RW1new RW1new RW1new rfn(TW1new,Inew,RnWater,Tc,Tw,AWater)]);
                RWNnew = geomean([RWNnew RWNnew RWNnew rfn(TWNnew,Inew,RnWater,Tc,Tw,AWater)]);
                Inew = Ib*Rs/(sum(Rnew)+RW1new+RWNnew+Rp+Rs);
                Inew = (1-Beta)*Inew + Beta*Iold;
            end
        end %Oscillation Suppression
        %Calculate New Heat Capacities Only After R, T and I Are Sorted Out
        f_norm = N*Rnew/Rn;
        Cnew = C_norm(Tnew).*f_norm + C_super(Tnew).*(1-f_norm);

    end %T_equil loop
end %While-Settling Loop

%Save Baselines For Further Calculations
I0 = Inew; %Baseline Source Current
T0(1,1,:) = permute(Tnew,[3 2 1]); %Baseline Temperature
TW1_0 = TW1new; %Baseline Waterfall Temperatures
TWN_0 = TWNnew;
C0(1,1,:) = permute(Cnew, [3 2 1]); %Baseline Heat Capacity
```

Another iteration of this physics engine starts in equilibrium and adds energy inputs to build templates on an energy comb Qtemp (typically very fine-toothed). Noise terms are calculated and saved, but not included. All finalized templates are run through a third order Butterworth filter and down selected so that any effect of the filter on real pulses is replicated.

```
Output = Gain*(Is-I0); %Shift Current To What AC-Coupled DAQ Reads Out
```

```
[Bfilt,Afilt] = butter(3,1/Select,'low'); %Filter Parameters
for x = 1:length(Qtemp) %Filter Each Pulse Separately, x Indexes Qtemp
    Dummy = filtfilt(Bfilt,Afilt,Output(x,:)); %Filter
    Templates1(1,:,x) = permute(Dummy(1,1:Select:NumBins),[3 2 1]); %Down-Select and Re-Orient
end
```

A third iteration through the same physics engine creates 'NumPulses' independent events with all noise terms included at each energy in Qweight. Qweight and Qtemp should have the same upper and lower bounds, with length(Qweight) $\approx$ sqrt(length(Qtemp)). After filtering and down-selecting, these pulses are averaged and their deviations from the average are correlated to make a covariance matrix.

```
Template = mean(Output_reduced,1); %Average Over Pulse Number To Produce Template
RepTemplate = repmat(Template,NumPulses,1); %Copy Template Into NumPulses Columns
Dev = Output_reduced - RepTemplate; %Deviation For Each Pulse
Cov = Dev'*Dev/NumPulses; %Correlate Deviations To make Covariance Matrix
W = inv(Cov); %Invert Covariance Matrix To Make Weighting Matrix
% z Indexes Qweight
Cov2(:,:,z) = Cov; %Save Covariance Matrix To Permanent Variable
Weights2(:,:,z) = W; %Save Weighting Matrix To Permanent Variable
```

After running through all values of energy in Qweight (for z = 1:length(Qweight)), all relevant quantities are saved for the template matching script.

```
WeightKey = permute(Qweight, [1 3 2]); %Make Qweight A Depth Vector To Aid Matching Later
save(savefile, 'Templates1', 'Cov1', 'Weights1', 'intE1', 'intRes1',...
  'TemplateKey', 'WeightKey', 'BestRes');
```

C.2 Matching Data To Templates

```
tic
clear all;
%Load Template File. Needs Template, Error and EnergyKey Matrices
load('Run40Mar22CHalfWeights1.mat');
```

```
load('Run40Mar22CHalfWeights2.mat');

%Load Reduced Data
load('Run40_data_reduced16_Dec13.mat');

NumPulses = size(events_reduced,1); %How Many Pulses Are There?
N = size(Weights1,2); %How Many Points Per Pulse?
OFP1 = zeros(1,NumPulses);
OFP2 = zeros(1,NumPulses);
OFP3 = zeros(1,NumPulses);
minChi2 = 1e6*ones(3,NumPulses); %If minChi2 Isn't Reset, Don't Plot Point

%Are You Only Using One TES Model?
if 0
    Weights2 = Weights1;
    Templates2 = Templates1;
    Cov2 = Cov1;
end
if 1
    Weights3 = Weights1;
    Templates3 = Templates1;
    Cov3 = Cov1;
end

%Should The Weighting Matrices Include Amplifier Noise?
if 1
    a = load('Run70Amp1Cov.mat');
    CovA1 = a.CovA;
    a = load('Run70Amp2Cov.mat');
    CovA2 = a.CovA;
    clear a;
    for z = 1:size(Weights1,3)
        Weights1(:,:,z) = inv(Cov1(:,:,z)+CovA1);
        Weights2(:,:,z) = inv(Cov2(:,:,z)+CovA2);
        Weights3(:,:,z) = inv(Cov3(:,:,z)+CovA2);
    end
end
```

```
%Assemble Interpolated Weighting Matrices
  %Pre-Declare 'FineWeights' To Be The Area of 'Weights', But the Depth of 'Templates'
FineWeights1 = zeros(size(Weights1,1),size(Weights1,2),size(Templates1,3));
FineWeights2 = zeros(size(Weights2,1),size(Weights2,2),size(Templates2,3));
FineWeights3 = zeros(size(Weights3,1),size(Weights3,2),size(Templates3,3));
for i = 1:size(Weights1,1) %(i,j) indexes a weighting matrix at a single energy
    for j = 1:size(Weights1,2)
        % Interpolation Embedded In 3-D Permutations To Preserve Matching Later
      FineWeights1(i,j,:) = permute(interp1(permute(WeightKey,[3 2 1]),...
        permute(Weights1(i,j,:),[3 2 1]),permute(TemplateKey,[3 2 1])),[3 2 1]);
      FineWeights2(i,j,:) = permute(interp1(permute(WeightKey,[3 2 1]),...
        permute(Weights2(i,j,:),[3 2 1]),permute(TemplateKey,[3 2 1])),[3 2 1]);
      FineWeights3(i,j,:) = permute(interp1(permute(WeightKey,[3 2 1]),...
        permute(Weights3(i,j,:),[3 2 1]),permute(TemplateKey,[3 2 1])),[3 2 1]);
    end
end
toc
%-------------------------------------------------
%Begin χ^2 Calculations
%-------------------------------------------------
for x = 1:NumPulses % x Indexes Pulse Number
    PulseTable = repmat(events_reduced(x,:,1), [1 1 size(Weights1,3)]);
      %Replicate Pulse To Matricize χ^2 Calculation
    [~,CoarseKey]= ismember(TemplateKey, WeightKey);
    Dev = PulseTable - Templates1(1,:,CoarseKey>0); %Deviation Matrix (Depth Indexes Energy)
    ColDev = permute(Dev, [2 1 3]); %Deviation Transpose
    Chi2 = zeros(1,1,size(Weights1,3)); %Pre-Declare χ^2
    for i = 1:size(Weights1,3) %i Indexes Coarse Energies
        Chi2(1,1,i) = Dev(:,:,i) * Weights1(:,:,i) * ColDev(:,:,i) / N;
    end
    [bestChi2, I]= min(Chi2,[],3); %Minimize χ^2 And Record Index
    minChi2(1,x) = bestChi2;

    if bestChi2 < 1000 %If We Got A Decent Fit
        % Select A Fine Energy Comb Around The Coarse Best Fit
        if I == 1%Is The Best Fit The Lowest Energy?
```

```
        fineKey = find(TemplateKey <= WeightKey(I+1));
    elseif I == size(WeightKey,3) %Is The Best Fit The Highest Energy
        fineKey = find(TemplateKey >= WeightKey(I-1));
    else %Is The Best Fit Somewhere In The Middle
        fineKey = find(TemplateKey >= WeightKey(I-1) & TemplateKey <= WeightKey(I+1));
    end
    fineEnergy = TemplateKey(fineKey);

    %Repeat χ^2 Minimization On The Selected Fine Comb
    PulseTable = repmat(events_reduced(x,:,1), [1 1 length(fineKey)]);
    Dev = PulseTable - Templates1(:,:,fineKey);
    ColDev = permute(Dev, [2 1 3]);
    Chi2 = zeros(1,1,length(fineKey));
    for i = 1:length(fineKey)
        Chi2(1,1,i) = Dev(:,:,i) * FineWeights1(:,:,TemplateKey(fineKey(i))+1)
          * ColDev(:,:,i) / N;
    end
    [bestChi2, J ]= min(Chi2,[],3);
    minChi2(1,x) = bestChi2;

    if J ~= 1 && J ~= length(fineEnergy)
        %Parabolic Fit To Find True Best Fit Between Energies On Comb
        OFP1(x) = fineEnergy(1,1,J) + ((fineEnergy(1,1,J+1)-fineEnergy(1,1,J))/2)
          *(Chi2(1,1,J-1)-Chi2(1,1,J+1))/(Chi2(1,1,J+1)+Chi2(1,1,J-1)-2*Chi2(1,1,J));
    else
        %If Best Fit Is An Endpoint, Just Select That Point (Shouldn't Happen Often)
        OFP1(x) = fineEnergy(1,1,J);
    end
else %If Not A Good Fit
    OFP1(x) = NaN; %Don't Record A Nonsense Energy
end %If Good Fit
```

Repeat this procedure to produce OFP2 and OFP3.

```
end
toc
```

Bibliography

[1] Oort, J. *The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems.* 1932BAN.6..2490 (1932, Astronomical Institute of the Netherlands).

[2] Zwicky, F. *On the Masses of Nebulae and of Clusters of Nebulae.* 1937ApJ.86.217Z (1937, The Astrophysical Journal).

[3] Ruben et al. *Rotational Properties of 21 Sc Galaxies with a Large Range of Luminosities and Radii, from NGC4605 (R=4kpc) to UGC2885 (R=122kpc).* 1980ApJ238471R (1980, The Astrophysical Journal).

[4] http://en.wikipedia.org/wiki/Virial_theorem

[5] ATLAS Collaboration. *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC* Phys. Lett. B716 (2012) 1-29

[6] Bunker, Raymond A. "A Low-threshold Analysis of Data from the Cryogenic Dark Matter Search Experiment" (2011) pgs. 24-30

[7] Milgrom, M. (1983). *A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis* 1983ApJ...270..365M Astrophysical Journal 270: 365370.

[8] Jo Bovy, Scott Tremaine, *On the Local Dark Matter Density*, arxiv:1205:4033

[9] J. Burney. *Transition-Edge Sensor Imaging Arrays For Astrophysics Applications* (2006)

[10] Ogburn, Reuben Walter. "A Search For Particle Dark Matter Using Cryogenic Germanium And Silicon Detectors In The One- And Two-Tower Runs of CDMS-II At Soudan" (2008)

[11] R. Agnese et. al. (SuperCDMS). "Demonstration of Surface Event Rejection With Interleaved Germanium Detectors in Dark Matter Searches", *Appl. Phys. Lett.* **103** 164105 (2013).

[12] J.D. Lewin and P.F. Smith. "Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil" (1996)

[13] Hart, Sean. "Quasiparticle Transport and Phase Separation in Transition Edge Sensors" (2009)

[14] J.J. Yen et. al. *Quasiparticle Diffusion in Al Films Coupled to Tungsten Transition Edge Sensors*, LTD-15, 15th International Workshop on Low Temperature Detectors, Pasadena, California 24-28 June 2013; to appear in *J. Low Temp. Phys.*

[15] K.D. Irwin, An application of electrothermal feedback for high resolution cryogenic particle detection. Appl. Phys. Lett. 66 (15), 1998 (1995).

[16] K.D. Irwin, G.C. Hilton, *Cryogenic Particle Detection*, pgs 81-97 (2005)

[17] R. Partridge for the SuperCDMS Collaboration. "Ionization Readout Electronics for SuperCDMS SNOLAB Employing a HEMT Front-End ", LTD-15, 15th International Workshop on Low Temperature Detectors, Pasadena, California 24-28 June 2013; to appear in *J. Low Temp. Phys.*

[18] D. N. Seitz et. al., "SuperCDMS Detector Readout Cryogenic Hardware", AIP Conference Proceedings **1185**, 282 (2009); doi: 10.1063/1.3292334

[19] R.S.Bhatia et al. "A three-stage helium sorption refrigerator for cooling of infrared detectors to 280 mK" Cryogenics 40 (2000) 685-691

[20] A. Jastram, R. Mahapatra et. al., "CDMS Detector Fabrication Process and Recent Improvements", submitted to *Nucl. Inst. Methods* (2014)

[21] http://ecee.colorado.edu/∼bart/book/book/chapter2/ch2_6.htm. Online textbook by B. Van Zeghbroeck

[22] K. Huang, *Introduction to Statistical Physics*, CRC Press (2001).

[23] http://en.wikipedia.org/wiki/Arrhenius_plot

[24] S. M. Sze and J.C. Irvin, Solid State Electron. Vol. **11**, 599-602 (1968).

[25] M. Lax, Phys. Rev. 119, 1502 (1960).

[26] J. Domange, E. Olivieri, N. Fourches and A. Broniatowski, *J. Low Temp. Phys.* **167**, 11311136, (2012).

[27] J. J. Yen et. al., "Measurement of Quasiparticle Transport in Aluminum Films Using Tungsten Transition Edge Sensors", to be published in Appl. Phys. Lett.

[28] M. Pyle, *Optimizing the Design and Analysis of Cryogenic Semiconductor Dark Matter Detectors for Maximum Sensitivity*, Ph.D. Thesis, Stanford University, 2012.

[29] N. Ashcroft, D. Mermin, "Solid State Physics", (1976)

[30] M. Pyle et al., "Quasiparticle propagation in aluminum fins and tungsten TES dynamics in the CDMS ZIP detector", LTD-11, 11th International Workshop on Low Temperature Detectors, 2005; NIM A 559 (2006) 405-407.

[31] A. Sommerfeld. Naturwissenschaft **15** (1927) 825.

[32] G. S. Kumar, G. Prasad, R. O. Pohl, "Experimental determinations of the Lorenz number", *J. Mat. Sci.* **28** (1993) 4261-4272.

[33] C. Kittel, *Introduction to Solid State Physics*, John Wiley & Sons, 1986.

[34] D. L. Goodstein et. al., "Heat capacity of a current carrying superconductor", Phys. Lett. A 245 (1998) 477-478

[35] J. Bardeen and M. J. Stephen, "Theory of the Motion of Vortices in Superconductors", Phys. Rev. (140): A1197-A1207

[36] M. Tinkham. *Introduction to Superconductivity.* McGraw Hill, 1975.

[37] K. D. Irwin, G. C. Hilton, D. A. Wollman, J. M. Martinis. "Thermal-response time of superconducting transition-edge microcalorimeters". *J. App. Phys.*, 83(8):39783985, Apr 1998.

bibitemmoffattarchive R. Moffatt, arvix post

[38] S.H. Moseley, J.C. Mather and D. McCammon (1984). "Thermal detectors as x-ray spectrometers". J. Appl. Phys. (56): 12571262.

[39] S. Roth, *Sputtered Tungsten Films and Composite Detectors for the Application in the Dark Matter Experiments CRESST and EURECA*, Ph.D. Thesis, Technische Univsersitat Munchen (2007), Section 3.1

[40] M. A. Lindeman et. al. "Impedance measurements and modeling of a transition-edge-sensor calorimeter". *Rev. Sci. Inst.* **75** 1283-1289 (2004).

[41] A. G. Kozorezov et. al. "Athermal energy loss from x-rays deposited in thin superconducting films on solid substrates". *Phys. Rev. B* **87**, 104504 (2013).

[42] A. G. Kozorezov, J. K. Wigmore, D. Martin, P. Verhoeve, A. Peacock, *J. Low Temp. Phys.* **151**, 51 (2008).

[43] B. Cabrera, code sharing.

[44] R. Moffatt, CDMS internal paper.

[45] D. J. Fixsen et. al., "Optimal Energy Measurement in Nonlinear Systems", *J. Low Temp. Phys.* online, DOI 10.1007/s10909-014-1149-x

[46] http://en.wikipedia.org/wiki/Joule-Thompson_effect