# Advancing the Search for Dark Matter: from CDMS II to SuperCDMS

by

Scott A. Hertel

B.S., The College of William & Mary (2003)

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2012

Author ..................................................
Department of Physics
August 31, 2012

Certified by ..................................................
Enectalí Figueroa-Feliciano
Associate Professor of Physics
Thesis Supervisor

Accepted by ..................................................
John Belcher
Professor of Physics
Associate Department Head for Education

# Advancing the Search for Dark Matter: from CDMS II to SuperCDMS

by

Scott A. Hertel

Submitted to the Department of Physics
on August 31, 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

## Abstract

An overwhelming proportion of the universe (83% by mass) is composed of particles we know next to nothing about. Detecting these dark matter particles directly, through hypothesized weak-force-mediated recoils with nuclear targets here on earth, could shed light on what these particles are, how they relate to the standard model, and how the standard model fits within a more fundamental understanding.

This thesis describes two such experimental efforts: CDMS II (2007-2009) and SuperCDMS Soudan (ongoing). The general abilities and sensitivities of both experiments are laid out, placing a special emphasis on the detector technology, and how this technology has evolved from the first to the second experiment. Some topics on which I spent significant efforts are described here only in overview (in particular the details of the CDMS II analysis, which has been laid out many times before), and some topics which are not described elsewhere are given a somewhat deeper treatment. In particular, this thesis is hopefully a good reference for those interested in the annual modulation limits placed on the low-energy portion of the CDMS II exposure, the design of the detectors for SuperCDMS Soudan, and an overview of the extremely informative data these detectors produce.

It is an exciting time. The technology I've had the honor to work on the past few years provides a wealth of information about each event, more so than any other direct detection experiment, and we are still learning how to optimally use all this information. Initial tests from the surface and now underground suggest this technology has the background rejection abilities necessary for a planned 200kg experiment or even ton-scale experiment, putting us on the threshold of probing parameter space orders of magnitude from where the field currently stands.

Thesis Supervisor: Enectalí Figueroa-Feliciano
Title: Associate Professor of Physics

# Acknowledgments

I have had a wonderfully rewarding time in grad school, and for that I owe a great debt to a great many people.

My advisor, Enectalí Figueroa, is of course a wealth of knowledge; what made him a great advisor (and stand out among the crowd of brains at MIT) was that he clearly cared in the success of others and invested his own time and efforts to help make me become successful. I count myself extremely lucky to have knocked on Talí's office door five and a half years ago, and can't imagine what my life would be like now if I hadn't.

The others here at MIT working on CDMS were just as supportive. We worked as a group, and much of this thesis is rightfully theirs: Steve Leman, Patrick Wikus, Kevin McCarthy, and Adam Anderson.

I worked most closely with fellow grad student Matt Pyle, across the continent at Stanford. What I have learned in grad school, and the contents of this thesis, is largely just the fraction of Matt's patient explanations I could get down in writing.

Then, there are the many other members of the CDMS collaboration, with whom it has truly been both a privilege and a pleasure to work with. I feel confident most collaborations are not nearly as warm and supportive as this one. I won't simply repeat the CDMS author list, but I should particularly thank those people I worked closest with: Paul Brink, Astrid Tomada, and Matt Cherry, Zeesh Ahmed, David Moore, Todd Doughty, Danielle Speller, Jeff Filippini, Joseph Kiveni, Oleg Kamaev, Bruno Serfass, Lauren Hsu, Dan Bauer, Jodi Cooley, Richard Schnee, Bernard Sadoulet, and Blas Cabrera.

Thank you all!

# Contents

I Dark Matter 12

1 Observations of Dark Matter 13

1.1 Observations of Astrophysical Objects . . . . . . . . . . . . . . . . . . 13

1.1.1 Dwarf Galaxies . . . . . . . . . . . . . . . . . . . . . . . . . . . 13

1.1.2 Spiral Galaxies . . . . . . . . . . . . . . . . . . . . . . . . . . . 14

1.1.3 Clusters of Galaxies . . . . . . . . . . . . . . . . . . . . . . . . 14

1.2 Cosmological Observations . . . . . . . . . . . . . . . . . . . . . . . . 16

1.3 Summary of what we know . . . . . . . . . . . . . . . . . . . . . . . . 18

2 Probing the WIMP Hypothesis 21

2.1 The Assumption of a Thermal Abundance . . . . . . . . . . . . . . . . 21

2.2 Collider-based Detection . . . . . . . . . . . . . . . . . . . . . . . . . 24

2.3 Indirect Detection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

2.4 Direct Detection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29

2.4.1 Expected Signal . . . . . . . . . . . . . . . . . . . . . . . . . . 29

2.4.2 The status of Direct Detection . . . . . . . . . . . . . . . . . . 32

2.4.3 CoGeNT . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35

2.4.4 CRESST-II . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35

2.4.5 DAMA/LIBRA . . . . . . . . . . . . . . . . . . . . . . . . . . . 37

2.4.6 Where do we stand? . . . . . . . . . . . . . . . . . . . . . . . . 38

**II CDMS Detector Physics** 40

**3 Energy loss by a recoiling particle** 43

**4 Electrons and Holes** 55

4.1 Charge Loss Processes . . . . . . . . . . . . . . . . . . . . . . 60

4.1.1 Sidewall trapping . . . . . . . . . . . . . . . . . . . . . 60

4.1.2 Top-Bottom surface interactions . . . . . . . . . . . . . . 61

4.1.3 Bulk trapping . . . . . . . . . . . . . . . . . . . . . . . 62

4.1.4 Crystal neutralization . . . . . . . . . . . . . . . . . . . 62

4.2 Sensing charges: The Shockley Ramo theorem . . . . . . . . . . . 63

**5 Phonons** 67

5.1 Luke Phonons . . . . . . . . . . . . . . . . . . . . . . . . . . 72

5.2 Recombination Phonons . . . . . . . . . . . . . . . . . . . . . 74

**6 Concentrating and Sensing Phonon Energy** 75

6.1 Bogoliubov Quasiparticles . . . . . . . . . . . . . . . . . . . . 76

6.2 The Quasiparticle-Phonon Cascade . . . . . . . . . . . . . . . . 79

6.3 Concentrating Quasiparticles I: Gap Energies . . . . . . . . . . . 80

6.4 Concentrating Quasiparticles II: Diffusion . . . . . . . . . . . . 83

6.5 Transition Edge Sensors . . . . . . . . . . . . . . . . . . . . . 87

6.6 Now we are ready . . . . . . . . . . . . . . . . . . . . . . . . 95

**III CDMS II** 96

**7 Detectors for the CDMS II Experiment** 97

7.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . 98

7.2 Charge . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 99

7.3 Phonons . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 105

7.4 QET Efficiency . . . . . . . . . . . . . . . . . . . . . . . . . 116

**8 CDMS II Analysis 119**

8.1 CDMS II Pulses . . . 119

8.2 Optimal Filtering . . . 122

8.3 Energy and Ionization Yield . . . 126

8.4 Phonon-Based Position . . . 128

8.5 Normalization by position . . . 132

8.6 Improving the delay-based position measures . . . 136

8.7 Turning energy, yield, and pulse shape into a dark matter limit . . . 144

 8.7.1 Summary of WIMP selection criteria . . . 144

 8.7.2 Nuclear recoil selection . . . 146

 8.7.3 Surface event rejection . . . 146

 8.7.4 Signal efficiency . . . 147

 8.7.5 Data blinding . . . 148

 8.7.6 Results . . . 149

 8.7.7 Background estimates . . . 150

 8.7.8 Expected surface event background, after unblinding . . . 152

 8.7.9 Discussion of results . . . 152

8.8 What have we learned from CDMS II? . . . 153

**9 CDMS II at Low Energies 161**

9.1 keVee and keVnr . . . 162

9.2 Analysis of the low-energy spectrum . . . 166

**10 Search for Annual Modulation in CDMS II 171**

10.1 Motivation . . . 171

10.2 The CDMS II dataset . . . 174

10.3 Event selection . . . 175

10.4 Statistical Treatment . . . 177

10.5 Conclusions . . . 180

10.6 Testing other event populations . . . 185

10.7 Results for smaller energy ranges . . . 186

IV SuperCDMS Soudan 190

11 Detectors for the SuperCDMS Soudan Experiment 191

11.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . 192

11.2 The Interleaved Strategy . . . . . . . . . . . . . . . . . . . . 193

11.3 Evolution of the concept . . . . . . . . . . . . . . . . . . . . 195

11.4 Channel Partition . . . . . . . . . . . . . . . . . . . . . . . . 200

11.5 Charge Channels . . . . . . . . . . . . . . . . . . . . . . . . 201

11.6 Shaping the drift field at high radius . . . . . . . . . . . . . . 206

11.7 QET Arrays 1: continuing the goals of CDMS II . . . . . . . . . . . 207

11.8 QET Arrays 2: new constraints brought on by interleaving . . . . . . 210

11.9 Fabrication Robustness . . . . . . . . . . . . . . . . . . . . . 213

11.10The Array of Detectors for SuperCDMS Soudan . . . . . . . . . . . . 215

12 Analysis of iZIP4/5 Data 223

12.1 Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . 224

12.2 Charge-Based Fiducial Volume Definition . . . . . . . . . . . . . . 225

12.3 Phonon Energy . . . . . . . . . . . . . . . . . . . . . . . . . 228

12.3.1 The Phonon Absorption Rate . . . . . . . . . . . . . . . . . 228

12.3.2 The Nonstationary Optimal Filter . . . . . . . . . . . . . . . 230

12.3.3 Recoil Energy and Yield . . . . . . . . . . . . . . . . . . . 235

12.4 Data Quality . . . . . . . . . . . . . . . . . . . . . . . . . . 236

12.5 Basic Discrimination Ability . . . . . . . . . . . . . . . . . . . 237

12.6 Position from phonon energy partition . . . . . . . . . . . . . . . 239

12.7 Ionization yield from phonon pulse shape . . . . . . . . . . . . . . 253

12.7.1 Review of Phonons in a iZIP Detector . . . . . . . . . . . . . 253

12.7.2 Time-Domain Pulse Fits . . . . . . . . . . . . . . . . . . . 257

12.7.3 Discrimination Ability Using Pulse Shape . . . . . . . . . . . 258

12.8 Surface Event Yield . . . . . . . . . . . . . . . . . . . . . . . 261

12.9 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . 266

V Outlook 268

13 Detectors for SuperCDMS SNOLAB 269

14 The Future of Direct Detection 275

14.1 Towards Larger Targets . . . . . 275

14.2 Towards Lower Thresholds . . . . . 278

14.2.1 TES Sensitivity . . . . . 278

14.2.2 Phonon Absorption Rate and QET Efficiency . . . . . 279

14.2.3 Luke Amplification . . . . . 280

14.3 Final Thoughts . . . . . 284

# Part I

# Dark Matter

# Chapter 1

# Observations of Dark Matter

It is traditional to begin with a list of evidences for dark matter. But by now, dark matter is a concept nearly eighty years old, and sits as a central pillar to our current understanding of the universe. It would be as useless to defend dark matter's existence as to defend the spherical earth, the heliocentric solar system, or the big bang. Instead, this chapter starts from the foundation that dark matter is central to the workings of the universe at large scales, and lays out a broad-brush answer to the question 'what do we know about dark matter, and how do we know it?'

## 1.1 Observations of Astrophysical Objects

### 1.1.1 Dwarf Galaxies

Dark matter has been observed to form gravitationally-bound structures. Starting at the smallest observed scale of such structures, our galaxy (and presumably all galaxies) are surrounded by smaller objects, so-called mini-halos or dwarf galaxies. These smallest galaxies are interesting in that they are the most dark-matter-dominated objects known. Due to their low masses, cooling of the baryons and resulting star formation are both suppressed, leaving the baryons in the form of diffuse gas. Through repeated interactions with the central galaxy (and presumably also with other dwarf galaxies), the diffuse baryonic component is easily stripped away, leaving nearly-bare dark matter bodies. Such 'ultra-faint' dwarf galaxies have been seen [102] using data from the Sloan Digital Sky Survey (SDSS) [114]. The velocities of the few remaining stars can be used to infer the strength of the gravitational potential (and thus the mass of the dwarf galaxy), as

$$\langle KE \rangle = -\frac{1}{2}\langle PE \rangle \tag{1.1}$$

For the faintest of the ultra-faint galaxies discovered thus far, the mass-to-light ratio is ~100 times that of typical spiral and elliptical galaxies, and ~1000 times that of a stellar population alone.

### 1.1.2 Spiral Galaxies

Dark matter's effect on the kinematics of larger galaxies (such as our own) is no less clear. The rotational velocities of spiral galaxies as a function of radius are easy to measure, using Doppler shifting of spectral lines. Such a method can be extended far beyond the radial extent of the stellar population through observations of the galactic neutral hydrogen (observing the 21 cm transition). Such measurements have been performed for a large number of galaxies, and these measurement show that galaxies (universally) exhibit nearly flat rotation curves, *i.e.*, orbital velocity varies only slightly with radius (as in Figure 1.1.2). Such a rotation curve is indicative of an enclosed mass linearly proportional to radius, or a density varying as $r^{-2}$. Because this trend extends to the outer reaches of the available luminous matter (and presumably beyond), we are left with the clear understanding that the luminous portion of each galaxy sits at the center of a much larger, much more massive dark matter 'halo'.

### 1.1.3 Clusters of Galaxies

We progress to the scale of the largest bound objects: clusters of galaxies. Here, the virial theorem again allows us to translate from the average velocity of luminous matter to the strength of the gravitational potential. At these scales, we measure the proper motions of entire galaxies (rather than individual stars or nebulae) and yet

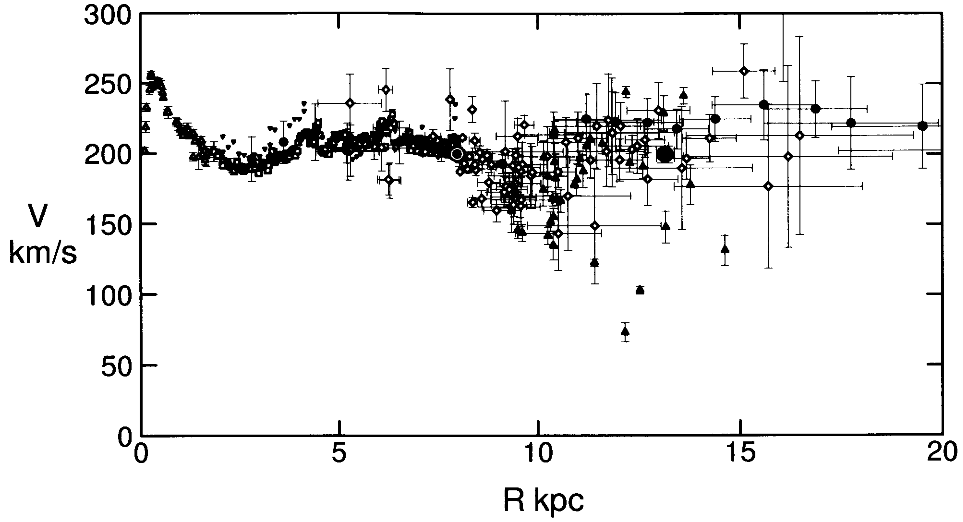

Figure 1-1: The orbital velocity of stars in the Milky Way as a function of galactic radius (a so-called 'rotation curve'). For reference, the sun is at ~8.3 kpc. Due to our disadvantageous observation point inside the distribution, the rotation curve of our own galaxy is actually one of the least-constrained rotation curves, particularly at higher radii. This plot compiles the work of many surveys, and was compiled by *Sofue et al* [103]. Dips are real, and represent the result of local spiral structure.

again map enormous amount of non-luminous matter binding the galaxies together. It was in fact these proper motions within clusters which provided the first clear evidence for dark dark matter (through the work F. Zwicky in 1933 [116]). These proper motions are not our only observational handle at cluster scales. The extremely large halos (and deep potential wells) associated with these scales imparts enough kinetic energy to the intergalactic medium to cause it to glow in x-rays (the temperature of which can be used, like the galactic velocities, as a probe of the gravitational potential). A third observational handle is the lensing of background galaxies, and a fourth observational handle is the SZ effect, the inverse Compton scattering of CMB photons from hot electrons in the intracluster gas. All these methods of observation have unrelated systematics, and all are consistent as to the amount and distribution of dark matter at these scales.

By the way, dark matter halos exhibit an interestingly scale-independence. Dark matter halos of all scales, from the dwarf galaxies to the largest clusters of galaxies, appear to follow the same radial density function, as

$$\rho(r) = \frac{\rho_0}{\frac{r}{R_S}\left(1 + \frac{r}{R_S}\right)^2} \tag{1.2}$$

where $\rho_0$ is the overall scaling in density, and $R_S$ is a radial scale. This distribution (the 'NFW profile', after Navarro, Frenk, and White, who discovered the behavior through simulation) varies as $r^{-1}$ at small radius, and $r^{-3}$ at high radius.

## 1.2 Cosmological Observations

There is one larger scale: that of the universe as a whole. The matter density, $\rho_m$, is typically specified in terms of the density parameter, $\Omega_m = \rho_m/\rho_c$ where $\rho_c = 3H^2/8\pi G$ is the critical density corresponding to a flat universe (and also happens to be at least extremely close the actual density of the universe). The total matter density consists primarily of cold dark matter and baryonic matter whose corresponding densities are denoted by $\Omega_c$ and $\Omega_b$, respectively, such that $\Omega_m = \Omega_c + \Omega_b$. A combination of cosmological measurements indicates that $\Omega_m = 0.267 \pm 0.025$, with $\Omega_c \approx 5\Omega_b$ [50]. An explanation of these many measurements is beyond the scope of this thesis, but here we include just the briefest of summaries:

**Big-bang Nucleosynthesis** (or BBN) is a theory of how light nuclei (D, $^3$He, $^4$He, and $^7$Li) are formed in the hot dense early universe. The amount of these nuclei present in the current universe places tight constraints on the duration of this nucleosynthesis epoch, in turn placing tight constraints on the universe's expansion history and mass. As an example, one recent measurement of deuterium (using Lyman-$\alpha$ absorption of light from background quasars in high-redshift, metal-poor systems) indicates a deuterium abundance of log(D/H) = -4.56±0.40, corresponding to a baryon density of $\Omega_b h^2 = 0.0213 \pm 0.0010$ [90].

**The Cosmic Microwave Background** gives us a window to the dynamics of universe at a somewhat later epoch than BBN(370,000 years after the big bang), a

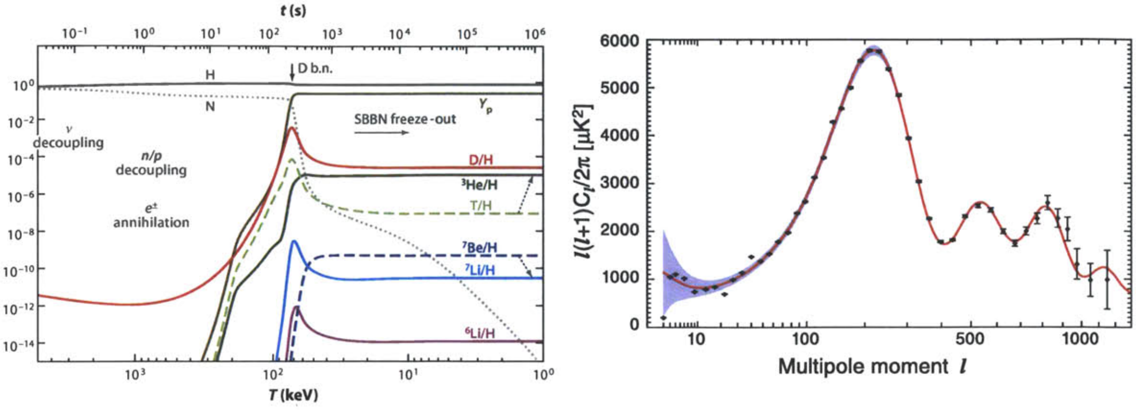

Figure 1-2: LEFT: Big Bang nucleosynthesis, showing the species fraction as a function of time (top axis) and temperature (bottom axis) according to the best-fit model to observational constraints. Taken from [91]. RIGHT: The temperature power spectrum from the seven-year WMAP data set, from [64]. Here, the most relevant fitted cosmological parameters are: $\Omega_b h^2 = 0.02260 \pm 0.00053$ and $\Omega_c h^2 = 0.1123 \pm 0.0035$.

time when the universe had cooled to exactly the temperature at which neutral hydrogen could form, and the density of the universe was still nearly perfectly uniform. Observed density fluctuations at this epoch (on the level of 1 part in $10^5$) provide a rich data set from which $\Omega_m$, $\Omega_c$, and $\Omega_b$ can be inferred.

**Large-Scale Structure** tests models of the collapse of these early density fluctuations into the galaxies and galaxy clusters we see today. Large data sets from surveys such as the SDSS [114] can be compared with ever-more-detailed simulations of structure formation, and the observations and simulations agree remarkably well. These structural arguments are one of the main observational clues to the mass of the dark matter particle; assuming dark matter is a thermal relic (as we will discuss shortly), light particles would have higher kinetic energies and would not be able to form the small structures (such as dwarf galaxies) that we observe.

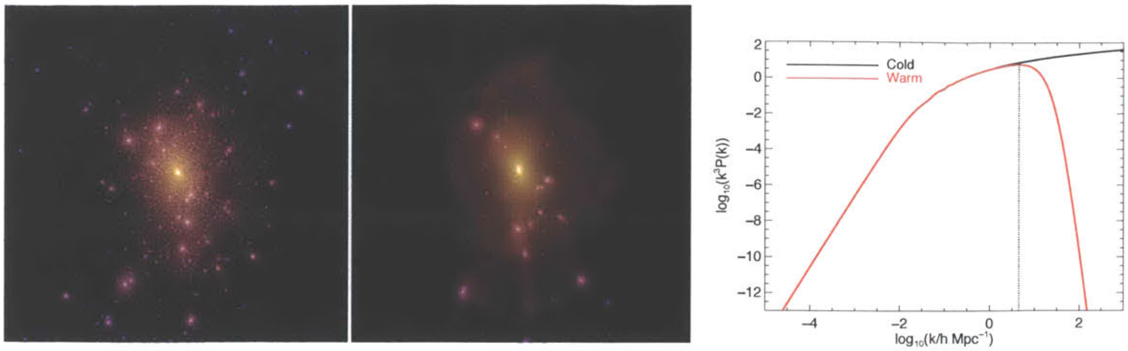

Figure 1-3: Two simulations showing varying amounts of power at small scales (cold dark matter (left) and warm dark matter (center), along with a comparison of the matter power spectrum of the two cases, again showing the lack of power at small scales in the warm dark matter case. All plots taken from [78].

## 1.3 Summary of what we know

Combining all the observations, we here list all our knowledge about dark matter. It is quite sobering to see that this sum total can easily fit on a single page.

**Dark matter ...**

**is the majority of the mass of the universe** , accounting for $\sim 83\%$ of the total matter density of the universe today. The stars of our galaxy reside within a much larger, more massive halo of dark matter (with local density $0.3 \pm 0.1$ GeV/cm$^3$ [30]).

**is at least primarily 'cold'** , meaning that its velocity is primarily non-relativistic. Only slow-moving dark matter can form the gravitationally-bound small-scale structures that we observe.

**has no easy method to dissipate energy.** Baryonic matter collapses to form disks and nebulae and stars precisely because of baryons' ability to radiate energy and loose kinetic energy. The size of dark matter halos indicates that dark matter lacks this ability to loose energy gained when falling to a potential well.

**is at least nearly collissionless** with both itself or baryonic matter. This is most clear in the example of the Bullet Cluster, but the morphology of dark matter halos themselves are also dependent on this (at least *near*) lack of collisions.

**is at least primarily non-baryonic.** The matter power spectrum at both the CMB epoch and today demands that dark matter must have started collapsing into overdensities prior to recombination, indicating Dark matter was decoupled from the photon fluid before baryons were decoupled. BBN synthesis enforces similarly tight constraints on the baryonic content of the universe.

**is stable over cosmological timescales.** No observation supports a significantly varying total dark matter mass fraction; on the contrary, all observations are consistent with a non-varying dark matter mass.

# Chapter 2

# Probing the WIMP Hypothesis

## 2.1 The Assumption of a Thermal Abundance

Nearly all particle abundances that we see in the universe today are the result of these species being initially in a thermal equilibrium (in a so-called 'soup' of particles), balanced at a certain amount by equal rates of creation and annihilation with other species. As the universe expanded and cooled, the density became so low that nearly all such creation and annihilation ceased, and, and the abundances set by the equilibrium in the early universe were 'frozen in'. This is exactly what happened in BBN, for example, though with so many species freezing out at once, this process is one of the more complicated examples of the formation of thermal relic abundances.

It is natural to assume that dark matter, too, had its present-day abundance set by such a thermal freeze-out process, so let's describe this process is some detail. First, let's give away the punchline: the assumption that the dark matter density is the result of thermal freeze-out gives us a clue as to the expected annihilation cross section, which turns out to be in range of magnitudes that seem natural for the Weak Force. For the rest of this thesis, then, we will be discussing 'WIMPs', Weakly Interacting Massive Particles.

In the early universe, when the temperature was much higher than the mass of the WIMP, $T \gg m_\chi$, creation and annihilation processes were in equilibrium, and the comoving number density of WIMPs, $n_\chi$, was constant. As the universe cooled

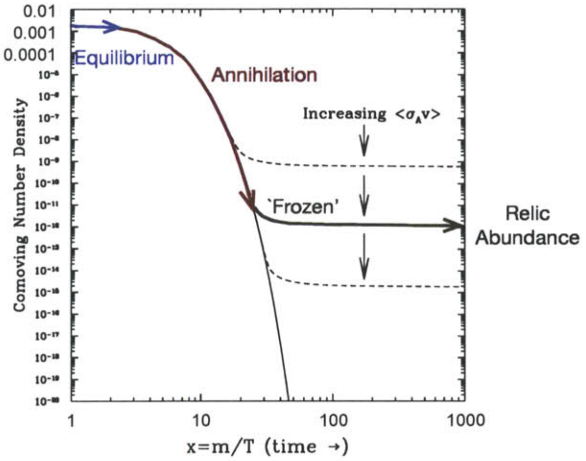

Figure 2-1: A schematic of the freeze-out of dark matter in the early universe, adapted from [60].

and the temperature fell below $T = m_\chi$, creation processes became inefficient, and the number density began to annihilate away exponentially, following the Boltzmann factor, $n_\chi \sim e^{m_\chi/T}$. See the shape of the red curve in Figure 2.1. Annihilation, however, requires two particles to 'find' each other, and becomes extremely inefficient below some threshold density. As the universe expands, the annihilation rate, $\Gamma_{ann} = \langle \sigma_{ann} v \rangle n_\chi$ falls below the Hubble expansion rate, H, and annihilation processes become inefficient. Here $\sigma_{ann}$ is the WIMP annihilation cross section and $v$ is the relative velocity of WIMPs. At this point, when the annihilation can no longer occur, the relic density of WIMPs is 'frozen'. Provided the WIMP is stable on cosmological time scales, this density remains today.

In such a model, the annihilation cross section, $\sigma_{ann}$, determines the relic abundance, $n_{eq}$. This abundance is exponentially sensitive to the cross section, with higher cross sections corresponding to lower relic abundances because the WIMPs can efficiently annihilate for a longer period, delaying freeze out. An approximate calculation of this relic abundance can relate $\sigma_{ann}$ to the measured quantity $\Omega_c$ as

$$\Omega_c h^2 \approx \frac{3 \times 10^{-27} \mathrm{cm}^3 \mathrm{s}^{-1}}{\langle \sigma_{ann} v \rangle} \tag{2.1}$$

For the observed relic abundance from cosmological measurements ($\Omega_c h^2 \approx 0.1$) we roughly estimate annihilation cross sections of $\sigma_{ann} \sim 10^{-37}\mathrm{cm}^2 = 0.1\mathrm{pb}$, where we have used that the freeze-out temperature is $T_{fo} \approx m_\chi/20$ , implying a typical velocity at freeze-out of $v_{fo} = 3T_{fo}/2m_\chi \approx 0.3c$. Thus the observed relic density is reproduced for cross sections $\sigma_{ann}$ ~0.1 pb and masses $m_\chi \sim$ 100GeV/c$^2$, both typical of weak scale interactions.

Note that this thermal relic abundance could have given us *any* scale, but out of the range of scales from $-\infty$ to $+\infty$, the relic abundance points us towards the weak scale. This is a good hint that the assumption of a thermal freezeout hypothesis is correct, first, just because at least the model spat out a *physically possible* scale, but also because we have many independent reasons to expect new physics to appear at the weak scale. For example, new physics at the weak scale can address the so-called 'hierarchy problem', the surprising fact that the Higgs mass (and by extension the mass of all Standard Model particles) is not pulled up to the cutoff scale at extremely high masses due to quadratic divergences in its radiative corrections. Supersymmetry is one solution to the hierarchy problem, and supersymmetry additionally provides a somewhat natural candidate for the dark matter particle: the lightest supersymmetric particle (LSP), the endpoint of the decay of any other supersymmetric particle, prevented from decaying further (into lighter non-supersymmetric particles) by conservation of so-called $R$-parity. Such supersymmetric models are infinitely variable, however (there is no supersymmetric model that is any more 'natural' than the others, especially not that the LHC has recently ruled out much of the cMSSM space), so assuming supersymmetry does not significantly help confine our wide-open parameter space.

Because this thesis is entirely targeted at either proving or disproving the WIMP hypothesis, we omit here any discussion of the many dark matter candidates that are *not* thermal freeze-out relics. Suffice it to say that there are in fact many other models, including Axions, which solve the hierarchy problem in an entirely different manner.

Generally speaking, there are three possible modes of WIMP detection, as illus-

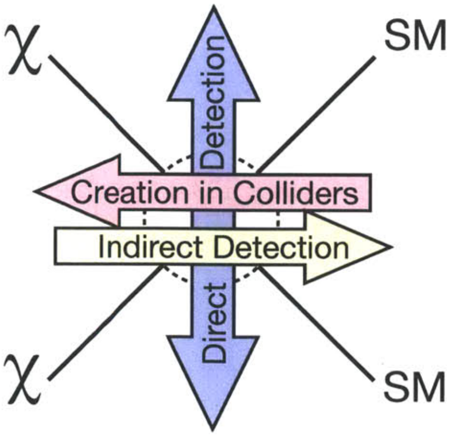

Figure 2-2: The three general detection mechanisms for WIMPs, where some level of weak interaction is assumed.

trated in Figure 2.1. All three of these pathways presumably have some cross section: creation from standard model particles (at accelerators), annihilation into standard model particles (in astrophysical regions of overdensity), and recoils off standard model particles (searched for by CDMS and other direct detection experiments). The key word here is 'complementarity', meaning that all three of these approaches provide useful constraints on the space of possible WIMP models, constraints which are to some degree orthogonal.

## 2.2 Collider-based Detection

Creation in colliders probe WIMP models up to a particular mass scale, and are therefore nicely complementary to direct detection (which looses sensitivity at low masses). Collider searches are complicated, however, by the fact that the WIMP will not interact with the detector, but will appear only as 'missing' momentum in the event reconstruction. There are two basic categories of WIMP events to search for. The lowest-energy event is the direct production of a WIMP pair, with the remaining balance of energy dissipated immediately before hand as a standard model particle. A higher-energy event type being searched for is production first of so-called WIMP 'siblings' (or perhaps more aptly, 'parents'), which decay into a WIMP pair and

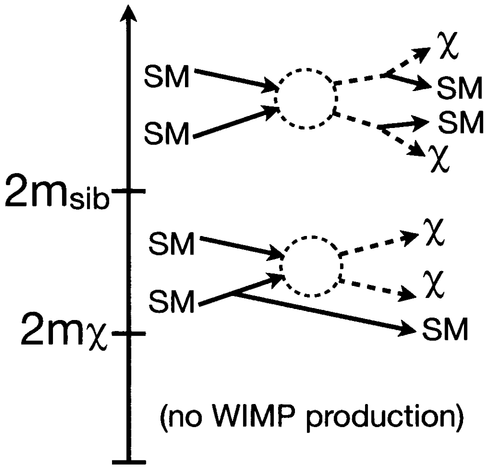

Figure 2-3: The two basic pathways by which dark matter ($\chi$) can potentially be probed with particles of the standard model(SM): through 'sibling' creation above $2\mathrm{m}_{sib}$, and through mono-jet creation with a lower threshold of $2\mathrm{m}_{\chi}$. The mono-jet signal is one of the easiest tags for a dark matter creation event, and is therefor being actively searched for at the LHC.

standard-model particles. These ideas are presented visually in Figure 2.2.

The task of the collider experimentalist (interested in discovering WIMPs) is to list all such possible WIMP signatures, define software triggers to be sure to record such events, and measure whether the observed rate agrees with the expected rate from non-WIMP processes, taking into account all the detector systematics. (When looking for a 'missing' signal, detector systematics are especially tricky). As this is being written, an army of physicists are busy performing these tasks using ATLAS and CMS data, and their limits on various production channels become tighter day by day.

These LHC results already rule out many standard WIMP models up to WIMP masses of hundreds of GeV and WIMP-nucleon cross sections orders of magnitude below where direct detection experiments have probed. On the other hand, many standard WIMP models are very difficult to probe at the LHC. And of course, there is the third category of models: those which have not yet been thought of. We turn then, to indirect and direct detection, two routes which are model-independent, in that they are probing a *generic* WIMP space.

## 2.3 Indirect Detection

Just as coannihilation occurred in the early dense universe (and set the thermal relic abundance we see today), such coannihilation processes should be continuing even now in regions where the dark matter density is highest. The rate of such a 2-body process scales as the density squared, so a search for coannihilation is greatly benefited by observing those objects where the WIMP density is highest. Unfortunately, these regions are typically also overdensities of baryonic matter, where numerous complicated radiatiative processes typically occur, potentially masking or imitating a dark matter coannihilation signature. Any indirect detection search, then, must necessarily depend on a complete understanding the baryonic physics of the search region.

The easiest way to be sure the baryonic radiation effects are understood is to avoid baryons altogether. This has been done, for example, in recent work by the Fermi-LAT Collaboration [4], looking at gamma ray emission from dwarf spheroidal galaxies (galaxies of extremely high dark matter content relative to baryonic content, as mentioned previously). See Figure 2.3. Such studies have ruled out low-energy WIMPs for certain decay modes but not others.

A second way to avoid the uncertainties inherent in baryonic physics is to constrain a search to sharp spectral features, because baryonic radiative processes at high energies are typically expressed as broad continuum emission. Recently, a hint of a narrow spectral feature at 130 GeV has been reported [110] (also using Fermi-LAT data). Greater statistics are necessary before being sure there is a spectral feature, and then a greater understanding of the relevant baryonic background physics would be required before such a signal could be confirmed. See Figure 2.3. It should be noted that the field of indirect detection is punctuated every few years by a new unexplained spectral hump or peak, and that it is very hard to ever find such a feature that is unexplainable through baryonic processes.

The third main strategy to avoid baryonic physics is to avoid photons altogether, and observe dark matter overdensities using the other long-distance messenger par-

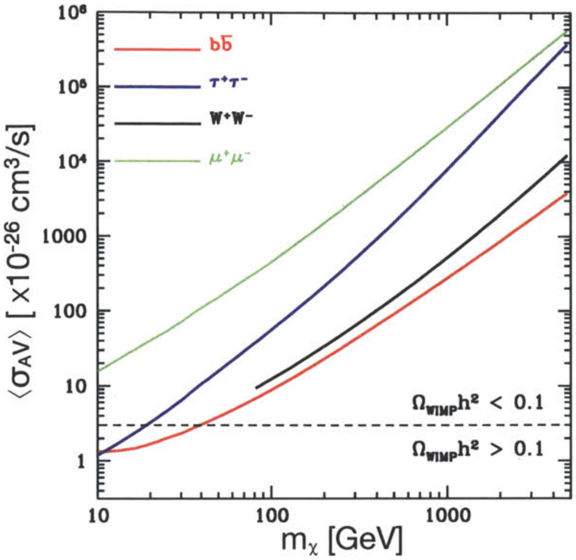

Figure 2-4: An example of dark matter coannihilation cross section limits, produced by measuring the lack of significant gamma ray flux from dwarf spheroidal galaxies (using Fermi-LAT). Note that the vertical axis is of units specific to the coannihilation cross section, and note that different hypothesized decay channels produce limits of different strengths but the same general scale. The dashed line at $3\times10^{-26}\text{cm}^3/\text{s}$ represents a canonical annihilation cross section value, of the scale necessary to produce the observed thermal relic abundance. Taken from [4].

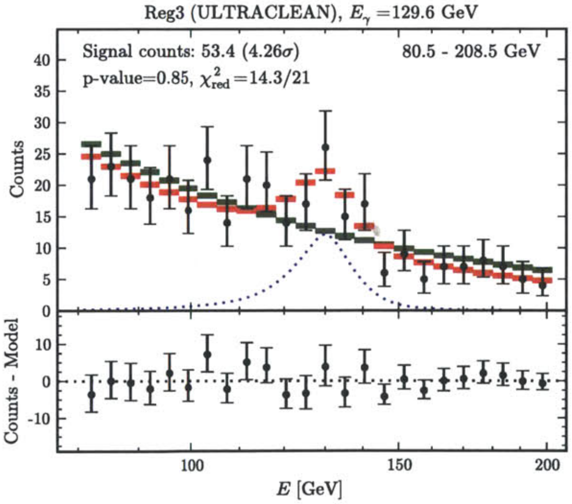

Figure 2-5: An example of a possible dark matter coannihilation spectral signature, taken from *Weniger, 2012* [110]. Shown is a gamma-ray spectrum using photons from the region surrounding the galactic center, and the spectrum appears to have an unexplained excess at ∼130 GeV.

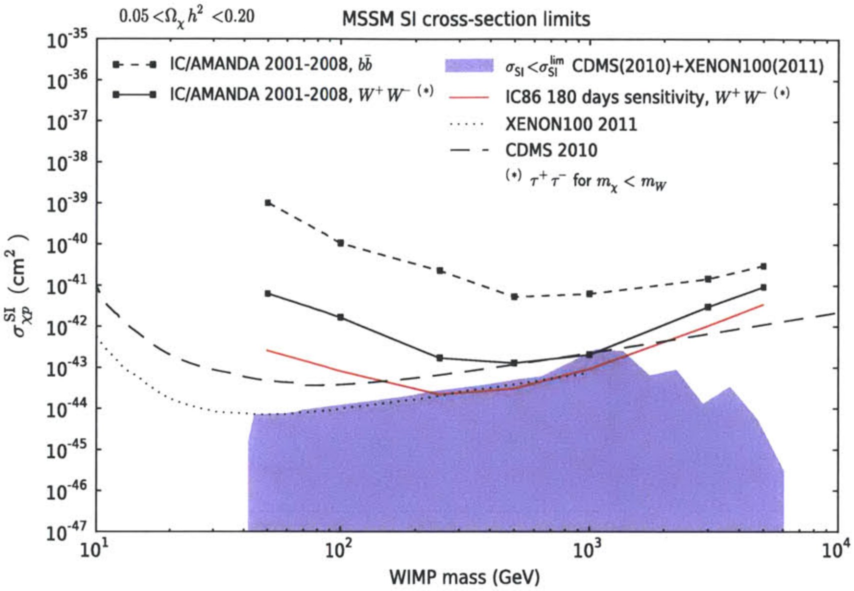

Figure 2-6: An example of spin-independent WIMP-nucleon cross section limits placed using neutrino rates from the sun. Taken from [3].

ticle: neutrinos. Neutrino-producing processes are far rarer than photon-producing processes, and therefor a annihilation signature should be much clearer. The IceCube experiment, with its enormous cubic kilometer detection volume, can place interesting limits on a multi-step process: WIMPs are captured by the sun (through the WIMP-nucleon cross section), collapse to the center of the sun through repeated interactions over millions or billions of years, and then coannihilate at the sun's center to produce heavy standard model particles (such as $W^+W^-$ pairs or $b\bar{b}$ pairs) which have neutrinos in their decay chain. Note that both the coannihilation and nuclear recoil cross sections are involved in this process. When compared with direct detection experiments (as in Figure 2.3), the limits are comparable, but include additional uncertainties and model-dependences. In the special case of a *spin-dependent* WIMP-nucleon cross section, the enormous number of H nuclei in the sun pushes these spin-dependent limits to be much more constrictive than any direct detection method.

## 2.4 Direct Detection

Direct detection, the goal of CDMS, is the detection of (typically, elastic) recoils of halo WIMPs with particles (typically, atomic nuclei) in detectors here on earth. Given a generic halo model (consistent with measurements of stellar motions), and a generic WIMP mass of ~100 GeV, we expect a flux of $\sim 5 \times 10^4$ particles per $\mathrm{cm}^2$ per second at the earth. Of course, this extremely high flux is countered by an extremely small WIMP-nucleon scattering cross section (by symmetry, assumed to be similar to the better-constrained coannihilation cross section), so that this flux has a negligible effect. However, the balance between this flux and expected WIMP-nucleon cross section is such that we would expect a rare but observable rate of nuclear recoils in macroscopic target masses, on the order of one recoil per kg per year.

### 2.4.1 Expected Signal

Assuming WIMPs have some cross section for elastic scattering off nuclei, and assuming a generic halo model (again, consistent with measurements of stellar motions in the milky way), we can make generic predictions for the recoils we expect to observe.

The WIMP velocity distribution is characterized by orbital motions, with v$\approx 10^{-3}$c. The recoil energy for elastic scattering is takes the simple non-relativistic form

$$E_{recoil} = \frac{\mu^2 v^2}{m_N}(1 - \cos\theta_c) \tag{2.2}$$

where $m_N$ is the target nucleus mass, $v$ is the WIMP velocity (in the target rest frame), $\theta_c$ is the scattering angle in the center of mass frame, and $\mu$ is the reduced mass $(m_\chi m_N)/(m_\chi + m_T)$ of the WIMP-target pair. Notice that if we look instead for WIMP recoils with electrons ($m_e$ ~0.5 MeV), WIMPs with masses much greater than a GeV cannot efficiently transfer energy. WIMP-induced electron recoils would have energies less than a single eV, and be nearly impossible to observe. On the other hand, the rough matching of the WIMP mass to the nuclear mass favors kinematically the efficient transfer of energy from a WIMP to a nucleus, giving the nucleus energy of keV scales. The minimum velocity WIMP which can produce a recoil of energy

$E_{recoil}$ is

$$v_{min} = \sqrt{\frac{m_N E_{recoil}}{2\mu^2}} \tag{2.3}$$

The differential WIMP-nucleus scattering rate, typically described in the units [keV kg day]$^{-1}$ ('differential rate unit', or 'dru') can be written as

$$\frac{dR}{dE_{recoil}} = \frac{\rho}{m_N m_\chi} \int_{v_{min}}^{\infty} v f_{(v)} \left[ \frac{d\sigma_{\chi T}}{dE_{recoil}}(v, E_{recoil}) \right] dv \tag{2.4}$$

where $\rho$ is the local dark matter density and $f(v)$ is the dark matter velocity distribution in the detector rest frame. The differential cross section $d\sigma_{\chi T}/dE_{recoil}$ encodes the particle physics of the WIMP-target interaction. The differential rate directly depends on the local WIMP density. A value of $\rho = 0.3\mathrm{GeV/cm^3}$ is canonical; it is within the measurement uncertainty from astrophysics, and it is convenient for all experiments to use a shared halo model (the 'standard halo model', or 'SHM'). Similarly, a Maxwellian velocity distribution (in the galactic rest frame) is assumed, truncated abruptly at the galactic escape velocity $v_{esc}$, as

$$f(v) = \begin{cases} Ae^{v^2/v_o^2} & v < v_{esc} \\ 0 & v > v_{esc} \end{cases} \tag{2.5}$$

where $A$ is simply a normalizing factor to ensure the integral of the probability distribution is 1, and $v_o$ is the characteristic (most probable) velocity, characteristic of the circular orbital velocity at the sun's galactic radius. In the SHM, $v_o = 220\mathrm{km/s}$. There is not a similarly canonical escape velocity, but some recent results have employed a value of $v_{esc} = 544\mathrm{km/s}$. The velocity of the earth relative to the galactic rest frame (*i.e.*, relative to the WIMP distribution) is not simply the sun's orbital velocity, but the vector sum of the sun's orbital velocity around the galactic center and the earth's orbital velocity around the sun. Combining these two velocities together one can write the earth's time-dependent velocity relative to the halo as

$$v_E = 232 + 15\cos\left(2\pi\frac{t - t_0}{365.25\mathrm{days}}\right) \mathrm{km\ s^{-1}} \tag{2.6}$$

where $t_0 = 152.5$ days (June 2nd). This few-percent annual modulation in the earth's velocity with respect to the WIMP velocity distribution should produce a corresponding variation in the recoil spectrum.

In the differential rate equation, then, we are going to assume astrophysical parameters (which are consistent with measurement), and phrase the results of direct detection experiments in terms of constraints on the term with the largest uncertainty: the differential scattering cross section. The non-relativistic WIMP-nucleon scattering differential cross section $d\sigma_{\chi N}/dE_{recoil}$ can be written as the sum of spin-dependent and spin-independent terms as

$$\frac{d\sigma_{\chi N}}{dE_{recoil}} = \frac{m_N}{2\mu^2 v^2}\left[\sigma_0^{SD} F_{SD}^2(E_{recoil}) + \sigma_0^{SI} F_{SI}^2(E_{recoil})\right] \tag{2.7}$$

In direct detection, we can amplify the cross section of a nucleon by coherently scattering off entire atomic nuclei, for which the cross section scales as the number of nucleons squared. The energy-dependent form factors ($F_{SD}(E_{recoil})$ and $F_{SI}(E_{recoil})$) contain the dependence of the scattering cross section on the momentum transfer, $q = \sqrt{2m_N E_{recoil}}$. The two cross sections ($\sigma_0^{SD}$ and $\sigma_0^{SI}$) can be written as

$$\begin{aligned} \sigma_0^{SD} &= \tfrac{32\mu^2 G_F^2}{\pi}\tfrac{J+1}{J}[a_p\langle S_p\rangle + a_n\langle S_n\rangle]^2 \\ \sigma_0^{SI} &= \tfrac{4\mu^2}{\pi}[Zf_p + (A-Z)f_n]^2 \approx \tfrac{4\mu^2}{\pi}[Af_{p,n}]^2 \end{aligned} \tag{2.8}$$

where $Z$ and $A$ are the atomic number and atomic mass number of the target nucleus, $f_p$ and $f_n$ are the spin-independent coupling strengths to protons and neutrons, $J$ is the nuclear spin, and $\langle S_p\rangle$ and $\langle S_n\rangle$ are the expectation values of the proton and neutron spin for the nucleus. In generic WIMP models, it is typically assumed that $f_p \approx f_n$.

For the spin-independent case, and assuming $f_p \approx f_n$, the nuclear form factor is given by the Fourier transform of the nucleon density, and typically parameterized ini terms of the momentum transfer $q$ as

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

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

We omit a more detailed discussion of the spin-dependent interactions, except to say that in this case the $A^2$ scaling is not available, and the majority of generic WIMP models can be much more easily directly tested through spin-independent means.

Figure 2.4.1 shows the resulting rates for two spin-independent WIMP models for a variety of nuclei and detector thresholds. Several important points:

- The observed scattering rate is dramatically increased by lowering the threshold.
- Given a threshold above several keV, rates are highest when the nuclear mass and the WIMP mass are similar.
- Given a threshold below several keV, rates are highest when the nuclear mass is highest.
- The rate for heavy target nuclei falls off more quickly in energy due to the suppression of the form factor (the interaction starts becoming non-coherent).

### 2.4.2 The status of Direct Detection

The field of dark matter direct detection is a vibrant one, in particular WIMP searches, with many experiments around the world competing to be the first to observe halo WIMPs through coherent nuclear scattering. In 2010, CDMS II produced a world-leading limit on the spin-independent cross section, which has since been superseded by Xenon100. Both these experiments follow the same basic strategy:

1. Large target mass (exposed to the dark matter flux for long times)
2. Low energy threshold
3. Low electron-recoil background rate
4. Efficient rejection of the remaining electron-recoil backgrounds

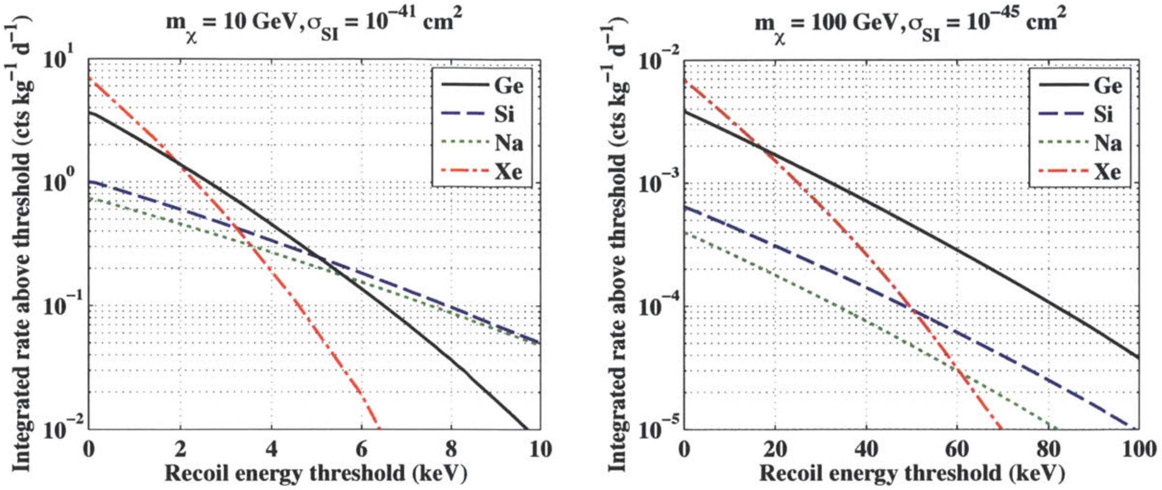

Figure 2-7: The integrated WIMP scattering rate (in counts per kg per day) for a variety of nuclei, as a function or detector threshold (assuming spin-independent interactions). On the left, we show a generic 'light WIMP' model (with mass and cross section chosen to approximately mimic several experimental hints), and on the right we show a more typical-mass WIMP (100 GeV), with cross section chosen to be in the range of current experimental limits. Figure by D. Moore.

If this strategy is implemented perfectly, the result is a large exposure (measured in mass×time, as kg-days, kg-years, ton-years, etc) in which the only events (remaining after rejection of electron recoils) are nuclear recoil events proportional to the WIMP-nucleus cross section. Neither CDMS II nor Xenon100 have seen more than a handful of events in such exposures, and so both experiments phrase such an exposure's scientific result as an upper limit on the WIMP-nucleon cross section. Such an upper limit is highly dependent on the assumed WIMP mass. An experiment is most sensitive assuming a WIMP mass similar to the nuclear mass, and less sensitive both above and below as a result of simple kinematics. Additionally, experiments loose all sensitivity to WIMP masses producing recoil spectra primarily below the detector's energy threshold.

As will soon be discussed in much more detail, CDMS II can distinguish electron recoils from nuclear recoils (and therefor reject them as non-WIMP-induced background events) by measuring the ratio of ionization and phonon production for each event. Similarly, Xenon100 distinguishes electron recoils from nuclear recoils by mea-

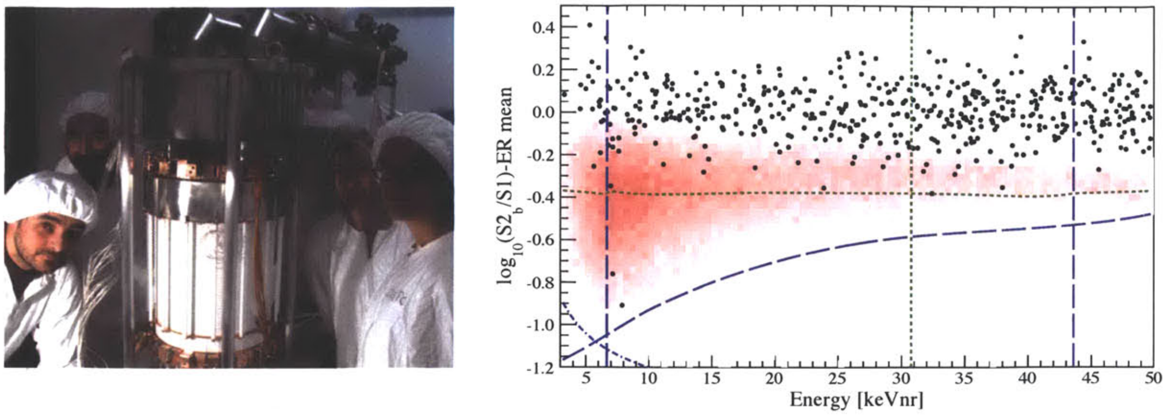

Figure 2-8: On the left, a view of the Xenon100 detector, and on the right, the events observed in the experiment's most recent exposure, displayed in the discriminator vs. energy plane. The discrimination here is in terms of S2:S1, where S2 is a signal due to ionization and S1 is a measure of the event's scintillation light. The measured distribution of nuclear recoils (from calibration data) is shown in pink, and the events of the source-free exposure are shown in black and match the expectation from electron recoil backgrounds. Plot taken from [112].

suring the ratio between ionization and scintillation production for each event. In both cases, the ionization signal is significantly larger in the electron recoil case than the nuclear recoil case. Other discrimination strategies are used in other experiments, including a scintillation:phonon ratio and recoil-type-dependent phase changes.

CDMS technology offers extreme background rejection abilities, but experiments using scintillation have the significant advantage of easily creating more massive target masses. Xenon100 [112] is now the world-leading experiment, surpassing the limits of CDMS II for all WIMP masses. At the lowest energies, Xenon100 has yet to publish a dedicated analysis, but the result from Xenon10 (a smaller predecessor) already surpassed the level of CDMS II in 2011 [14].

It is highly beneficial for an experiment to have as small a background rate as possible (remaining after discrimination). The sensitivity of a zero-background experiment scales simply with the exposure. If the experiment allows undiscriminated backgrounds ('leakage'), then the leakage will presumably scale with the exposure, and sensitivity to a dark matter signal will scale only as the square root of the exposure.

In addition to the robust null results of CDMS II and Xenon100 (and, in fact many other experiments), three experiments have seen event rates they interpret as inconsistent with backgrounds, and here we briefly summarize each of these results.

### 2.4.3 CoGeNT

The CoGeNT collaboration operates a 440 g P-type point-contact germanium detector at Soudan Underground Laboratory. This detector offers no electron recoil discrimination, because it uses employs only one measurement channel (ionization). The geometry of the point-contact produces a detector with an impressively low capacitance, leading to an exceptionally low threshold of $\sim$2keV, ideal for detecting low-mass WIMPs. CoGeNT has an ability to reject events occurring near the detector surfaces, using the fact that such events have a slower rising pulse edge (though this discrimination becomes less and less possible near threshold).

CoGeNT initially reported a large excess of events near threshold, exponentially increasing at lower energies as would be expected from a low-mass WIMP recoil spectrum [1]. Further data taking confirmed this excess, and further, gave statistically-limited evidence ($\sim 2.8\sigma$) for an annual modulation in the event rate [2]. Subsequent to the initial understanding, is has been suggested that the majority of the exponential excess could be due to surface event leakage [67][42][44]. Additionally, the modulation signature is inconsistent with realistic halo models (an order of magnitude too high), inconsistent with the lack of modulation seen in the CDMS II exposure (a topic of this thesis), and now appears to be inconsistent with the rate seen after doubling the CoGeNT exposure [43].

### 2.4.4 CRESST-II

The CRESST-II experiment consists of an array of $CaWO_4$ crystals, and combines a scintillation measurement with a phonon (more precisely, a temperature) measurement for discrimination of electron recoils. Like CoGeNT, this experiment also sees a roughly exponentially rising spectrum at low energies (shown in the center of Figure 2-

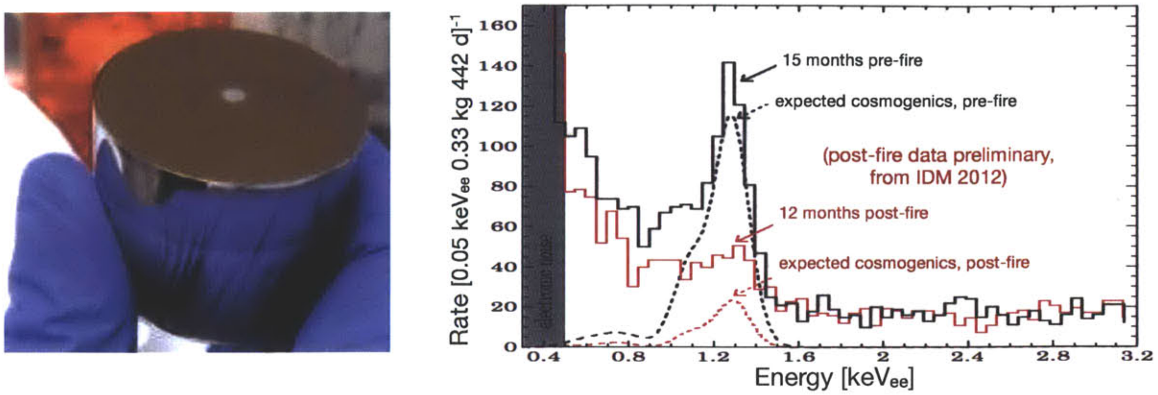

Figure 2-9: The CoGeNT spectrum at low energies, both before (black) and after (red) a break in data-taking. Note that the energy axis assumes an electron recoil energy scale. In the second half of the exposure, the background peak has decreases (as expected from the expected half-life values), but the high rate below this peak (in the region 0.5-1.0keV$_{ee}$ remains high. Taken from [43].

10), and further, the ionization yield of these events strongly implies that they are of nuclear recoil origin [16]. The CRESST collaboration has performed a combined likelihood analysis, combining all background distributions in the ionization yield vs. energy plane (extrapolated from higher energies), and arrive at a $> 4\sigma$ confidence on the presence of a WIMP-like nuclear recoil excess.

The question in such an analysis (as it is, similarly, for the CoGeNT case or any non-null result), is 'Are the background rates and distributions known?' It appears unlikely (from distribution of events in yield) that the excess is a result of electron recoil events leaking in the ionization yield measurement. Instead, the primary non-WIMP hypothesis is that the events are true nuclear recoils, but caused by radon contamination of the surrounding materials (metal clamps that hold the scintillating crystals in place) and the associated $^{210}\text{Po} \rightarrow \alpha + ^{206}\text{Pb}$ decays in which the high-energy $\alpha$ may not be detected, and the $^{206}$Pb nucleus may recoil into the target volume. Simulations of this process performed by the CRESST collaboration predict a flat spectrum of such events, whereas independent simulations including sputtering and surface roughness effects indicate that the $^{206}$Pb spectrum could easily be expected to rise at low energies [72]. Interestingly, SuperCDMS Soudan currently has Pb sources installed as calibration sources, and we see a similar high rate at low energies in

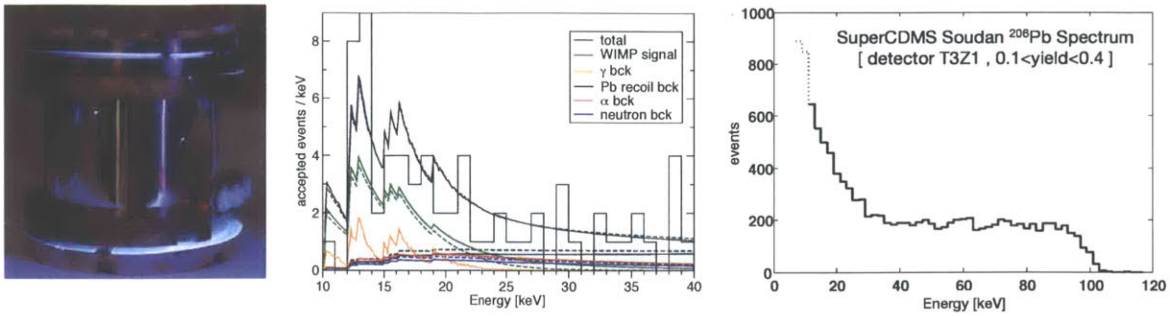

Figure 2-10: From left to right, a photograph of a CRESST II scintillation and temperature detector, the spectrum of low-ionization-yield events from the most recent exposure of CRESS II, combining several such detectors (shown along with a best-fit model including both a WIMP spectrum and backgrounds, including a flat-spectrum $^{206}$Pb contribution) from 2011 [16], and on the right, a possible mechanism for the low energy excess: a $^{206}$Pb recoil spectrum that exponentially increases at low energies (as seen here in the low-yield portion of $^{206}$Pb calibration data at SuperCDMS Soudan).

these recoils (shown in Figure 2-10). Such a spectrum is largely an expression of the material properties of the material in which the $^{210}$Pb radon daughters are embedded, and we wouldn't expect to reproduce the CRESST II situation exactly.

### 2.4.5 DAMA/LIBRA

The DAMA collaboration has been steadily improving their detector for more than a decade. The current iteration of the experiment (DAMA/LIBRA) consists of ~250 kg of thallium-doped NaI scintillator crystals of extreme radiopurity. Thanks to this large mass, the experiment has by far the largest exposure of any of its competitors, with nearly 1.2 ton-years of exposure collected over 13 annual cycles [27]. Like CoGeNT, this detector uses only one detection channel (scintillation) through which to observe events, and therefor has no discrimination between electron recoils and nuclear recoils. However, thanks to the enormous exposure and long operation, DAMA searches for (and finds) an annual modulation in the overall rate of events, with a spectrum exponentially increasing at low energies, consistent with the WIMP hypothesis. DAMA's technology is impressive, both in its extremely low event rate in the low energy region (<1 event per keV$_{ee}$ per kg per day) and its low energy threshold of 2 keV$_{ee}$. The presence of light nuclear species (Na) in the target mass make the experiment

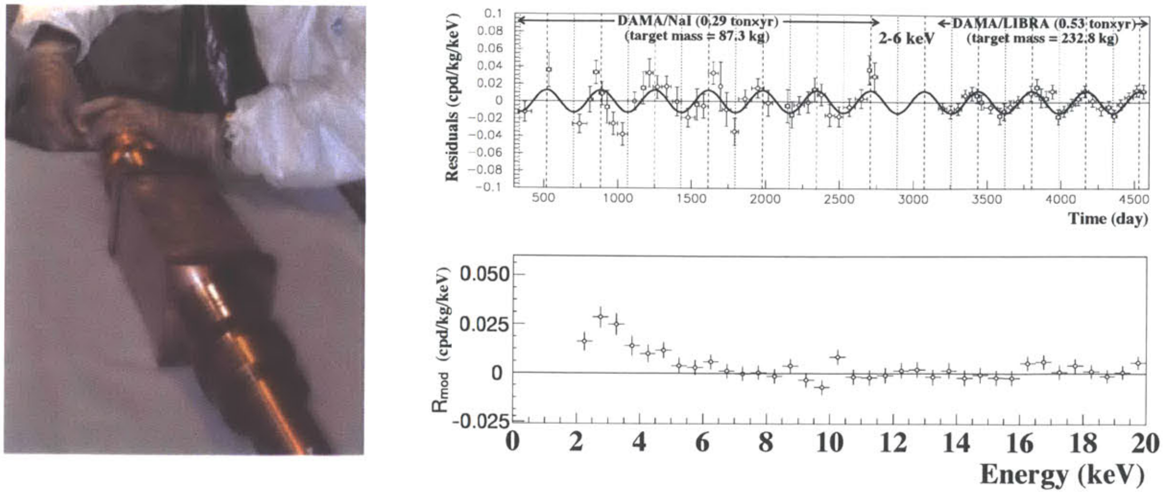

Figure 2-11: A view of one of the many detector modules that make up the DAMA/LIBRA detector (left, encased in Cu), and two views of the observed modulation over many annual cycles (right).

particularly sensitive to light WIMP kinematics.

As seen in Figure 2-11, the modulation effect is clear, with a statistical significance of $8.9\sigma$. Further, the modulation has a phase consistent with the expected phase for a WIMP-induced modulation signal ($t_0 = 144 \pm 8$ days for DAMA [27]; $t_0 = 152.5$ days according to the SHM). The modulation amplitude peaks at the lowest energies and is absent at higher energies. It occurs only for events interacting in single detectors (like CDMS and CRESST II, the target mass is made up of multiple independent detectors) and is absent for events interacting in multiple detectors. To date, no background model can fully explain the observed modulation.

### 2.4.6 Where do we stand?

Collider experiments, indirect detection experiments, and direct detection experiments are all advancing steadily through the WIMP parameter space, scanning the space of possible WIMP masses and weak interaction strengths. Taking just the direct detection experiments alone, we find severe disagreement between several strong null results and several positive observations, ranging from mere 'hints' to DAMA's clear ringing signature. CRESST II and CoGeNT could both be easily dismissed as misunderstood backgrounds, DAMA/LIBRA is harder to escape, and the rough

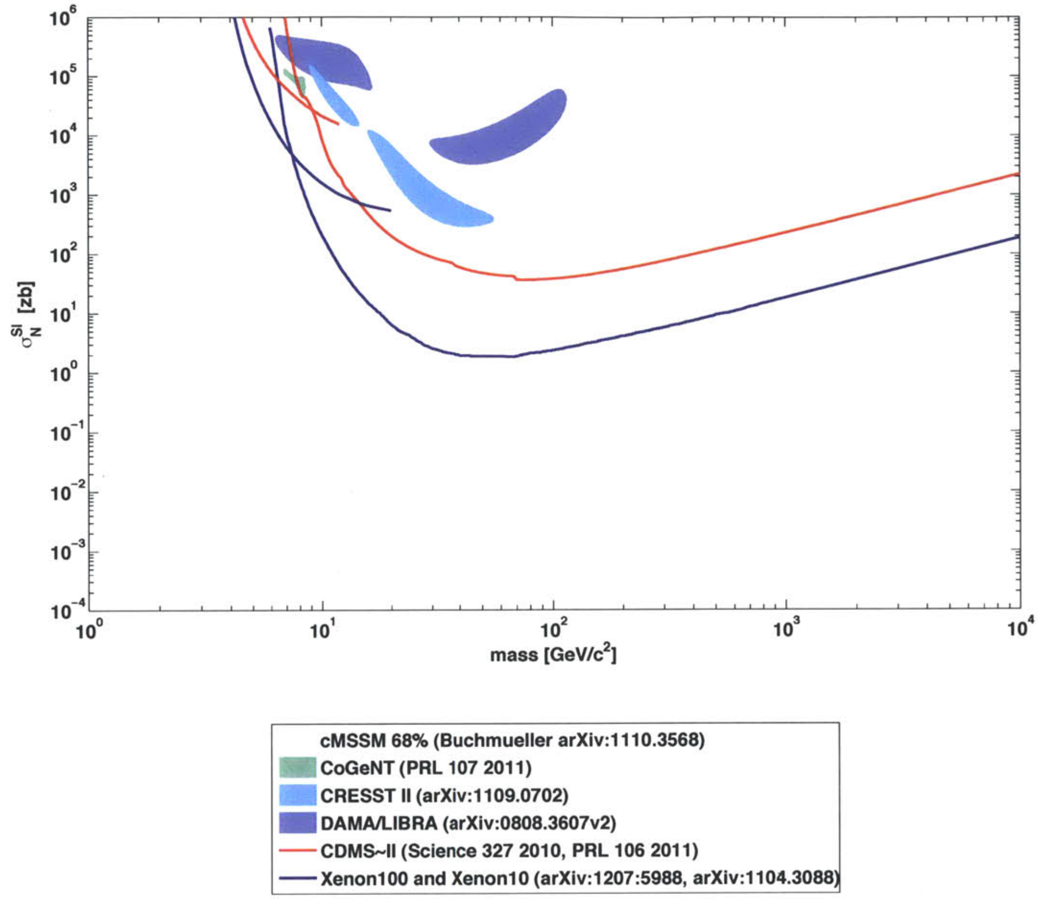

Figure 2-12: A view of constraints on WIMP models in the space of spin-independent WIMP-nucleon cross section (vertical axis) vs. WIMP mass (horizontal axis). Included here are only those results mentioned in the text; many experiments are omitted for clarity. The 68% bayesian confidence level region for one particular supersymmetric model (cMSSM) is shown simply to point the reader's eye towards the general region of space favored more generally by supersymmetric models. Both CDMS II and Xenon have had dedicated analyses (shown separately) for the low-mass regime.

similarity of the allowed regions at light masses is interesting.

Figure 2-12 shows a few of the direct detection limits mentioned in the text, many additional experiments have been omitted for clarity.

# Part II

# CDMS Detector Physics

In these chapters, we will follow the flow of energy through a CDMS detector in general terms, from a recoiling electron or nucleus all the way to a an electrical signal. We will see how the initial recoiling particle imparts energy to both phonons and e-h pairs, and how these energy carriers propagate through the crystal and are sensed at the crystal surfaces. These general explanations are designed to lay the groundwork for understanding specific applications and optimizations in later chapters.

# Chapter 3

# Energy loss by a recoiling particle

As a particle passes through matter, it gives off energy. The physical processes responsible for this energy loss are not simple, but warrant some discussion because their resulting partition of deposited energy between charge excitations and lattice vibrations lies at the very heart of CDMS.

Stopping power $S_{tot}$ (energy loss per unit length, $-dE/dx$) can be described as the sum of two terms: the electronic stopping power $S_e$ and the nuclear stopping power $S_n$.

$$-\frac{dE}{dx} = S_{tot}(E) = S_e(E) + S_n(E) \tag{3.1}$$

Let's take first the case of a recoiling electron. Electrons find it kinematically impossible to impart significant energy to an atomic nucleus, and therefor $S_n(E)$ becomes negligible. The electronic stopping power $S_e$, then, dictates the energy loss of a recoiling electron, and is illustrated in Figure 3-1. We can use this plot of stopping power to gain a better physical picture of the stopping process. Imagine an electron with 10 keV of energy. It loses energy at ~10 keV per $\mu$m, which already gives us some physical understanding: the energy of recoiling particles at these low energies is spread over tiny volumes of the material. As the electron loses energy to other electrons, eventually the identity of the starting electron is lost and the initial energy is shared among a population of energetic electrons. This cascade of electrons, each

Table 3.1: Partition of Energy from an Electronic Cascade in Si and Ge

| | $E_{gap}$ | $E_{create}$ | e-h fraction $= \frac{E_{gap}}{E_{create}}$ |
|---|---|---|---|
| Si | 1.21 eV | 3.81 eV | 0.318 |
| Ge | 0.785 eV | 2.96 eV | 0.265 |

electron exciting multiple other electrons, gradually spreads the initial energy out over many electrons, ending only when each electron no longer has enough energy to excite another (there is some minimum energy an excited charge can have, described as either a gap energy or ionization energy, depending on the medium). At this point, the electrons can no longer lose energy to electronic stopping, and either stay at this last energy, or lose energy in some other fashion. 'Some other fashion' is strongly dependent on the material. In the case of CDMS, the electrons lose their last remaining kinetic energy to the crystal lattice in the form of phonons. The end result of the electronic cascade, then, is some population of electron-hole pairs at the gap energy (1.21 eV for Si, 0.785 eV for Ge) and some population of phonons. Notice that electronic stopping power alone dictates the cascade until energies approach eV scales. This means that as long as the initial recoil energy is much greater than the minimum charge excitation energy, we should expect the final fractional partition of energy between charge excitations and phonons to be constant with recoil energy. At its bare essentials, then, the entire electronic cascade process can be stripped down to one number: the fraction of initial energy that ends up as charge excitations. This fraction can be expressed as $E_{gap}/E_{create}$, where the "creation energy" $E_{create}$ is the average energy necessary to create an e-h pair (of gap energy $E_{gap}$). The remaining energy $(1 - E_{gap}/E_{create})$ goes into phonon production. Numerical values are given in Table 3.1. For an electron recoil in Ge, 26.5% of that energy will end up in the form of e-h pairs, or in other words, ~338 e-h pairs per keV of initial recoil energy.

Now let's turn to the case of a recoiling nucleus. The nucleus will cause a similar cascade of excited particles, but in the nuclear case, the cascade involves not only excited electrons but also excited nuclei, and at lower energies the nuclear stopping power dominates. This interesting competition between electronic and nuclear stop-

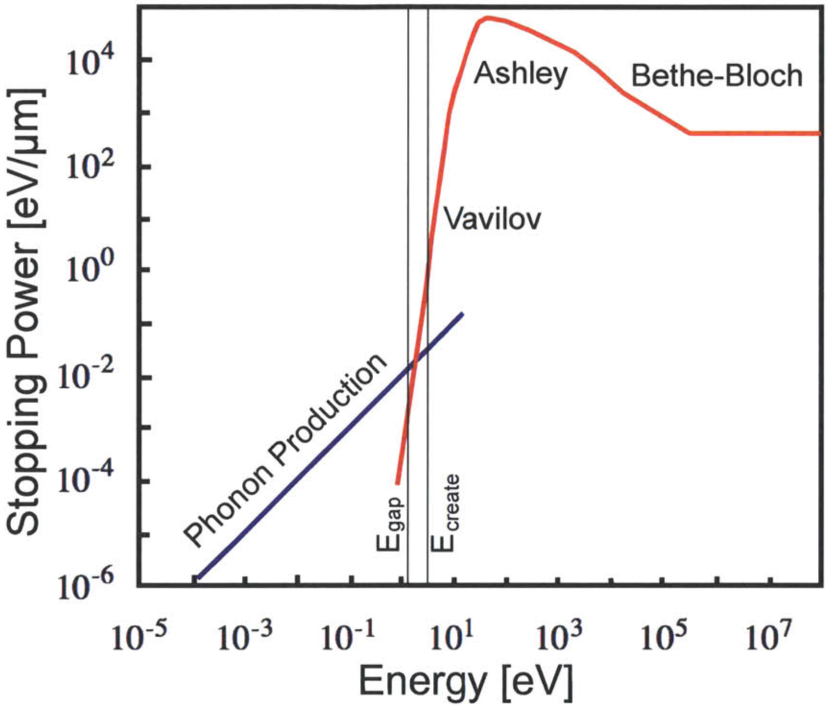

Figure 3-1: Stopping power as a function of energy, for an electron passing through Si. The main features are the so-called "minimum ionizing" energy at ~500 keV, the maximum stopping power energy at ~50 eV, and the transition from electronic stopping power to phonon stopping power at energies similar to the gap energy. Adapted from [36]

ping powers was first well-described by Jan Lindhard in the 1960's[48]. He found that, in the low energy limit, a nuclei's electronic stopping power is proportional to velocity $v$ as

$$S_{e,Lindhard}(E) \approx 8\pi\hbar a_o \left[ \frac{Z_1^{7/6} Z_2}{(Z_1^{2/3} + Z_2^{2/3})^{3/2}} \right] v \tag{3.2}$$

where $a_o$ is the Bohr radius, $Z_1$ and $Z_2$ are the recoiling and target material nuclear charges, respectively, and $v$ is the velocity of the recoiling nucleus. It is convenient when discussing the stopping of nuclei to rephrase discussions in terms of a scaled dimensionless energy

$$\epsilon = E \frac{a M_2}{Z_1 Z_2 e^2 (M_1 + M_2)} \tag{3.3}$$

and a scaled dimensionless distance

$$\rho = R N M_2 \cdot 4\pi a^2 \frac{M_1}{(M_1 + M_2)^2} \tag{3.4}$$

where $a = 0.8853 a_o (Z_1^{2/3} + Z_2^{2/3})^{-1/2}$ and $N$ is the atomic number density. Rephrasing stopping power in terms of these new dimensionless quantities, we have a dimensionless stopping power $d\epsilon/d\rho$ which is nearly universal for all nuclei and all stopping materials. Plugging in numerical factors, and setting $Z_1 = Z_2$, we arrive at a convenient description of electronic stopping power

$$\left(\frac{d\epsilon}{d\rho}\right)_e \approx 0.133\ Z_1^{2/3} A_1^{-1/2} \sqrt{\epsilon} = k\sqrt{\epsilon} \tag{3.5}$$

The electronic stopping power is proportional to the square root of the energy (*i.e.*, the velocity, as previously stated), and all the complexities of the electronic stopping are now described by a single parameter $k$. Plugging in appropriate values for $Z_1$ and $A_1$, we arrive at $k_{Si} = 0.146$ and $k_{Ge} = 0.157$ for Silicon and Germanium, respectively.

We have already mentioned that the stopping of a nucleus involves a competition between the electronic and nuclear stopping powers. $S_n$ follows from the atomic cross section, which is significantly screened by atomic electrons. At high energies, the Rutherford cross section is applicable, but at lower energies screening weakens the cross section significantly. Bohr suggested that the nuclear stopping potential was roughly constant with energy, an approximation that is still useful. More accurate descriptions of the atomic cross section (and the resulting nuclear stopping power) can only be fully described numerically. We skip, therefor, to a visual representation of $S_n$, shown along with $S_e$ in Figure 3-2.

Let us again interpret a plot of stopping power vs. energy, this time tracing the story of a recoiling nucleus from Figure 3-2. If the nucleus is initially of a very high energy, we see that the electronic stopping power dominates. In this high-energy case, then, the nucleus gives its energy almost entirely to electrons, and these high-energy electrons interact only with other electrons (as seen from Figure 3-1), and the end result of the cascade is practically indistinguishable from the electron-recoil case. If, on the other hand, the nucleus is initially of a lower energy, then the nuclear stopping potential dominates, and the end result of the nuclear cascade is (in analogy with the electronic cascade) nuclei which no longer have enough energy to excite other nuclei. The tipping-point energy (where $S_n = S_e$) occurs at $\epsilon \approx 6A_2Z_2^{-4/3}$ (assuming a generic value of $k$ =0.15), or an energy of 0.1-10 MeV, depending on the nuclei involved.

It is important to be clear that the nuclear stopping power is in no way related to the lattice. In Figure 3-2, one can see that nuclear stopping involves energy scales as high as an MeV, much higher than lattice binding energies. Nuclear stopping results from nucleus-nucleus recoils, in which a significant amount of energy is given off to a single nucleus. In a crystal like Ge or Si, these nuclei are entirely freed from the crystal lattice. These freed nuclei then excite other nuclei, and contribute to a nuclear cascade. The nuclear cascade is not as pure as the electronic case, though. Note that throughout the relevant energy range, there is some non-negligible chance of an electronic excitation (as seen in Figure 3-2). This excited electron will never excite a nucleus and contribute back to the nuclear cascade, so there is a small but significant *one-way* flow of energy from the excited nuclear population into an excited electron system.

At the end of the electronic cascade, electrons loose their remaining energy in the

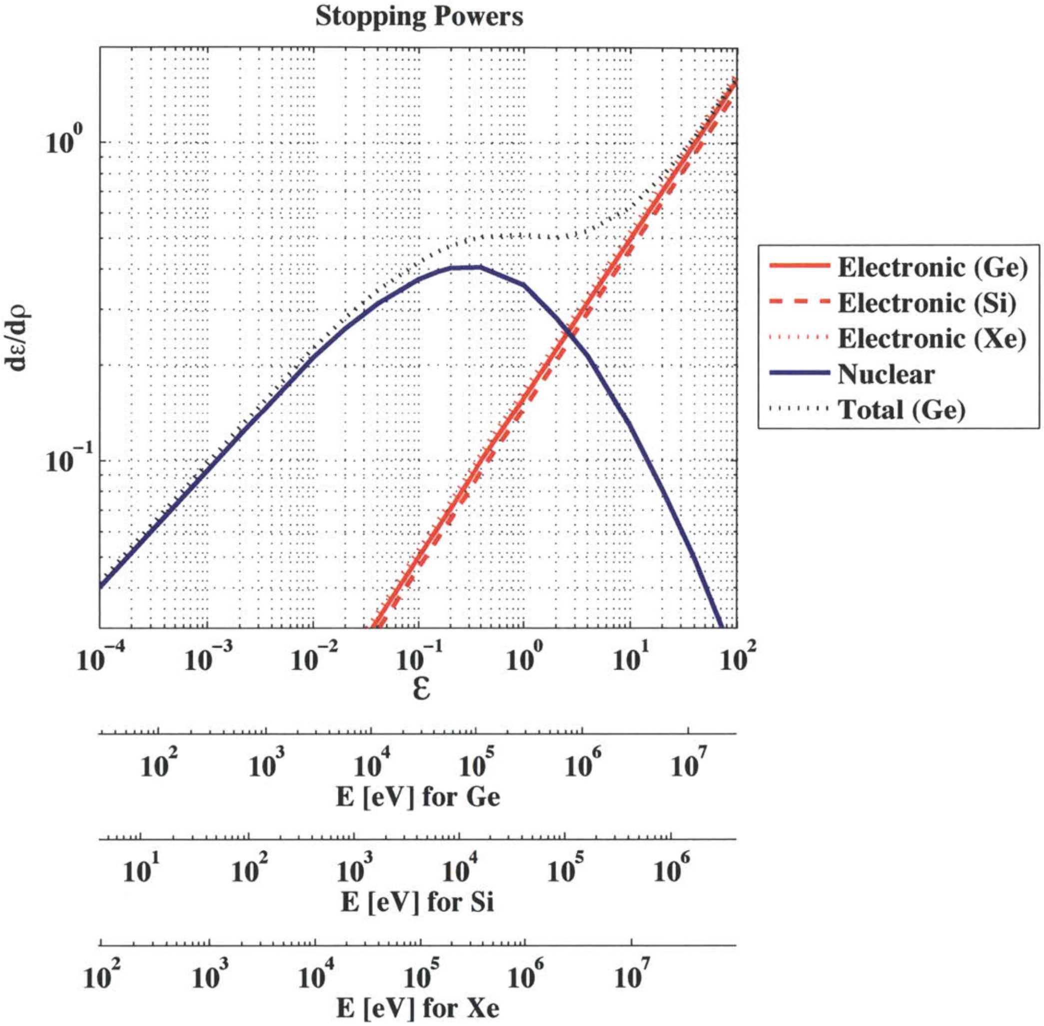

Figure 3-2: Stopping powers as a function of energy, experienced by a nucleus passing through a material. Using the scaled energy $\epsilon$ and scaled distance $\rho$, the nuclear stopping power becomes nearly universal. The electronic stopping power depends weakly on atomic species. In addition to the semiconductor target materials used in CDMS (Si and Ge), xenon is also shown, to emphasize the generality of these stopping power expressions. Note that the lowest energy scales, where electronic and nuclear stopping gives way to phonon stopping, are not displayed.

form of phonons as they lower to the gap energies. A similar process occurs at the end of the nuclear cascade, but with the difference that there is no $E_{gap}$, meaning that the conversion of energy from eV-scale recoiling nuclei into phonons is much more efficient. The only inefficiency in converting the nuclear cascade's energy into phonons is the storing of energy in the form of lattice defects, to be discussed shortly.

To this point, the discussion of electronic and nuclear stopping powers is entirely generic to any nucleus losing energy to any material, as long as energies above $\sim$10 eV are considered. At the very end of the electronic and nuclear cascades, energy dissipation is dependent on the electron shells and inter-atomic bondings of the material, characteristics that have not entered into our discussion of $S_n$ and $S_e$ at all. At these *molecular* energy scales, a multitude of physical processes become available to the experimentalist wishing to distinguish nuclear and electron recoils.

Unfortunately, any derivation connecting the intuitive stopping powers to the final resulting partition between phonons and e-h pairs depends on numerically solving integral equations (which already contain numerical functions to describe the nuclear cross section), and does not add any significant physical intuition. Such a derivation is skipped, then, in favor again of a visual representation, shown for both Ge and Si in Figure 3-3. "Ionization yield" is a commonly used metric, describing the fraction of recoil energy energy appearing in the form of charge excitations, normalized such that the ionization yield of an electron recoil equals 1. Nuclear recoils, then, have a ionization yield somewhere between zero and one, and a simple fit to the numerical solution can be written as

$$Y(\epsilon) = \frac{k \cdot g(\epsilon)}{1 + k \cdot g(\epsilon)} \tag{3.6}$$

where $k$ and $\epsilon$ are defined as before, and

$$g(\epsilon) \approx 3\epsilon^{0.15} + 0.7\epsilon^{0.6} + \epsilon \tag{3.7}$$

Before going further, let us take a moment to specify the energy range over which such a description is valid. At the high energy range, the $k\epsilon^{1/2}$ description

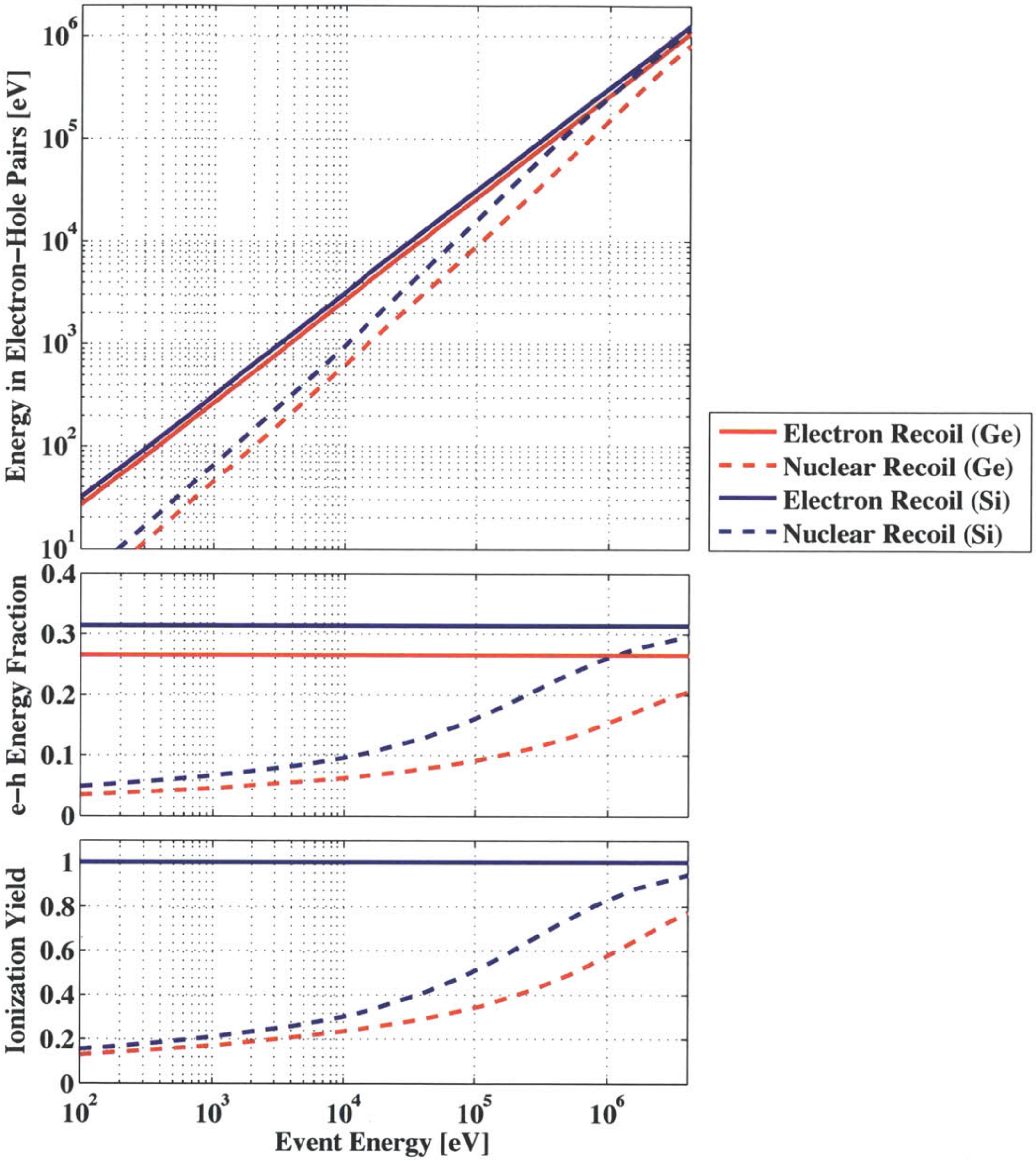

Figure 3-3: The fraction of event energy finally being expressed in the form of e-h pairs is here shown in several different ways. In the top frame, the resulting e-h pair energy is shown as a function of event energy, for both electron and nuclear recoils. In the middle frame, the same information as plotted, but as a fraction of total energy deposited. In the bottom frame, this fraction has been normalized by the electron recoil case, as is the CDMS convention.

of electronic stopping power fails when electronic stopping reaches a maximum near $v \approx (e^2/\hbar)Z_1^{2/3}$. This velocity corresponds to $\epsilon \approx 10^3$ (or higher) and occurs at energies well above the range interesting for dark matter. At the other energy extreme, the ionization yield expression given above fails at low energies only as we approach the threshold energies for creating free electrons and nuclei (at $eV$ energy scales). This is in fact a limiting threshold for the best ionization-based detectors, where a low-energy event may only excite a single electron. In the semi-conductor targets of CDMS, however, such physical thresholds occur well below the noise thresholds of the charge and phonon sensors themselves.

A complexity that has received much attention in recent years is that the effect of a crystal lattice on the competition between nuclear and electronic stopping powers. Such spacial correlations between atoms were ignored completely in the derivation of both the nuclear or electronic stopping powers. The most important result of these spacial correlations is the so-called "channeling" effect, wherein a recoiling ion travels along a channel in the lattice, and only low incidence-angle nuclear stopping occurs, boosting the electronic stopping fraction and the resulting ionization yield. As originally pointed out by Lindhard, the channeling effect can be expected to play a significant role for the case of a free ion incident on a material (when the ion has some probability of starting its path in a channel) but a much smaller role for the case of a nucleus recoiling away from its original position at a lattice site, when the ion is necessarily starting its path at a significant angle to the channel direction. Interestingly, a detailed theoretical treatment has recently found that such effects (in addition to being of negligible importance) are also strongly dependent on the temperature of the lattice, as seen in Figure 3-4.

We have emphasized how the partition of energy between the nuclear cascade and the electronic cascade is generic, described for all materials by Lindhard theory. The material-dependent properties occur when the cascades reach the atomic and molecular energy scales. The cascades in Si and Ge deposit the vast majority of their energy into the phonon and e-h pair systems. However, the nuclear cascade energy is not converted entirely into phonons; some amount is stored permanently

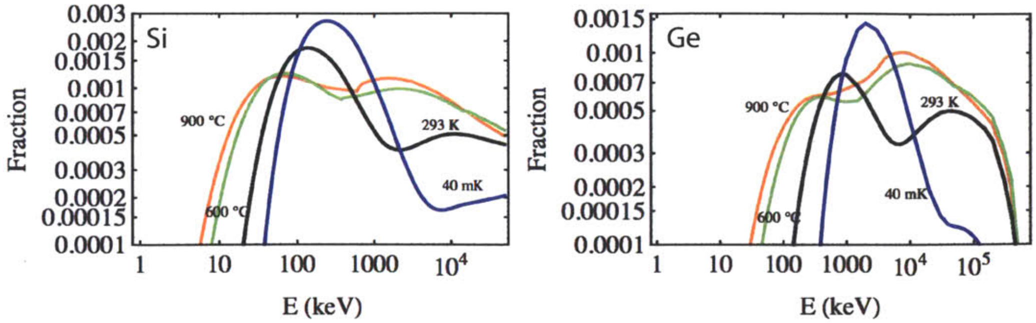

Figure 3-4: Fraction of nuclei involved in channeling, measured by passing some threshold traveling distance, as a function of energy and temperature for Si (left) and Ge (right). Note that a very small fraction of the nuclear cascade is involved ($< 1\%$). Note also that the trend with temperature is not monotonic. There are two competing effects: a vibrating lattice more easily allows a bounded nucleus to be in a channel when it recoils, but a non-vibrating lattice more easily allows channeling once the nucleus recoils. This plot is from Bozorgnia *et al.* [31].

in the crystal as defects in the crystal lattice. Lattice defects have higher potential energy than a perfect lattice, but can be stable. Just like e-h pairs, these defects can be thought of has having creation energies (typically several eV) and final stable ('gap') energies (typically ~eV). The fraction of nuclear recoil energy ending up in the form of lattice defects is not well known. Simulation work by Nordlund *et al.* [89] and measurements by the CDMS and Edelweiss collaborations [26] seem to suggest that as much as 10% of the energy lost to nuclear stopping power ends up as lattice defects (rather than phonons). Traditionally, CDMS thinks of the recoil energy as being initially partitioned between electron-hole pairs and phonons; understanding this third energy sink (and its energy dependence) could be vital to better calibrating the energy of nuclear recoil events.

The nuclear recoil can be thought of as *melting* a small volume of the crystal, and some portion of this small volume will remain amorphous (and thus, storing some small amount of energy). A 10 keV nucleus, initially melts thousands of atoms (lattice binding energy is a measly 0.156 eV per atom) surrounding the initial recoil position. Simulations by Nordlund *et al.* [89] (shown in Figure 3-5) show that this initial melted lattice persists ~5 ps in Ge (~1 ps in Si).

One final point. The goal of Lindhard's work was to create a general description of

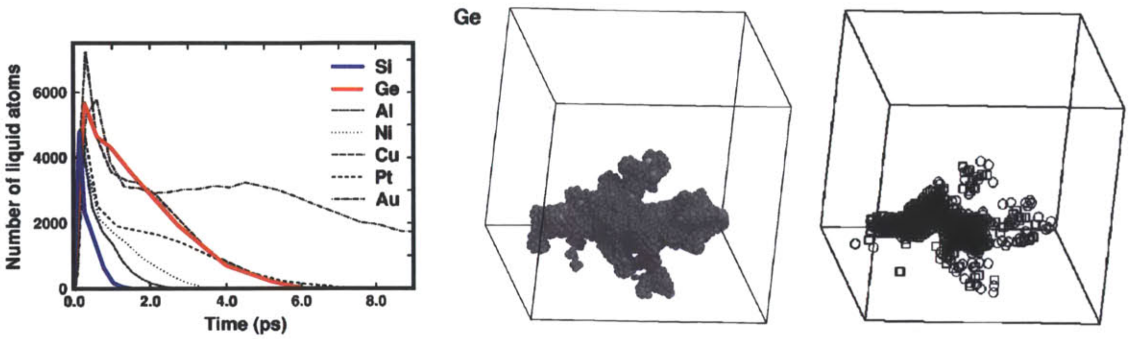

Figure 3-5: The results of atomic collision cascade simulations, adapted from Nordlund *et al.* [89]. On the left, the number of above-melting-energy atoms as a function of time is shown for a 10 keV nuclear recoil. In the center, the spacial distribution of above-melting-energy atoms is shown for an example 10 keV Ge nuclear recoil, at the time of peak melt volume. The bounding box is ~20 nm on a side. On the right, the location of the resulting crystal lattice defects for the same event is shown. Squares show the locations of vacancies and circles show the locations of interstitials (two types of defects). It can be seen that in some region, the an amorphous state dominates. Some later diffusion and merging of these lattice defects may occur.

nuclear stopping processes in all materials, not to treat the specifics of our particular circumstance, crystaline Si and Ge. It is reassuring, then, to know that Lindhard's description has been robustly experimentally verified, as seen in Figure 3-6.

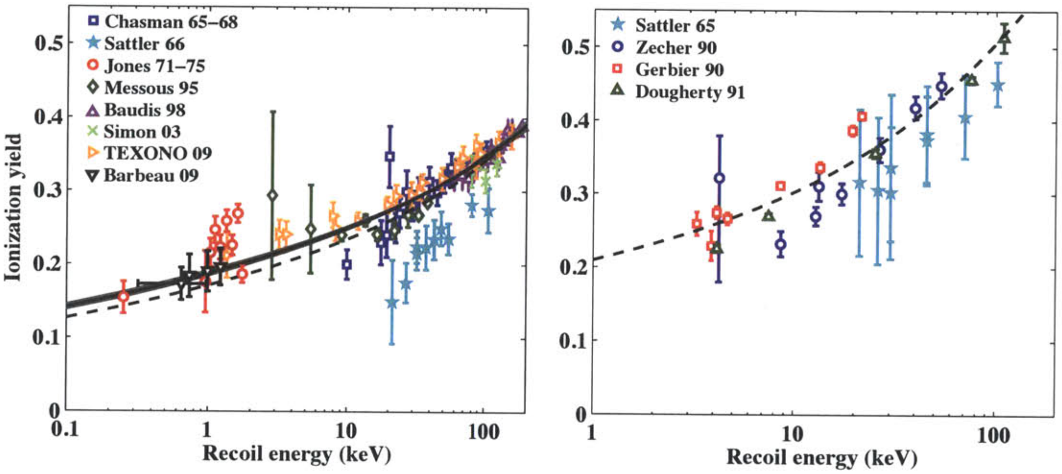

Figure 3-6: Nuclear recoil charge yield measurements in both Ge (left) and Si (right). Dashed lines represent the expectation from the Lindhard description. For Ge, a slightly better agreement with the measurements in the literature can be described using a fitted function (solid black line, for which $k$=0.235 and a linear slope parameter $Y = (0.795)Y_{Lindhard}$ is added). These measurements are taken from many authors[39, 40, 38, 65, 66, 84, 97, 23, 20, 101, 100, 25, 96, 115, 57, 45], and the plot is from a CDMS paper on nuclear recoil ionization yield currently in preparation, to appear in NIMA.

# Chapter 4

# Electrons and Holes

We have discussed how many e-h pairs have been created by the recoil of a particle, without stopping to discuss the nature and properties of this resulting excited state. Ge and Si are both semiconductors, meaning that there exists a small energy gap between the electron ground state energy and the first excited state (and the start of the conduction band). These gap energies are a reflection of the interatomic spacing, and thus vary with the temperature of the material, as seen in Figure 4-1, reaching maxima at T=0. Another important result of near-zero temperatures is that the occupancy of the excited state approaches zero. In the CDMS situation, then, the carriers are 'frozen out'. With no charges above the Fermi level, the semiconductors behave as insulators, because no charge carriers propagate (except those that have been excited to the valence band by an event).

The allowed electron states are functions not only of momentum magnitude, but of momentum *vector* with respect to the crystal lattice. Both Si and Ge form diamond cubic lattices, with similar primitive cell dimensions of 0.543 nm (Si) and 0.566 nm (Ge). The repeating spacial patterns produce repeating patterns in the momentum space available to excited electrons and holes. The resulting momentum space can be thought of as the Fourier transform of the spacial lattice, and is called the 'reciprocal' lattice.

Just as there is a primitive cell in the spacial description, there is a primitive cell in the momentum space description, called the first Brillouin zone. The Brillouin zone

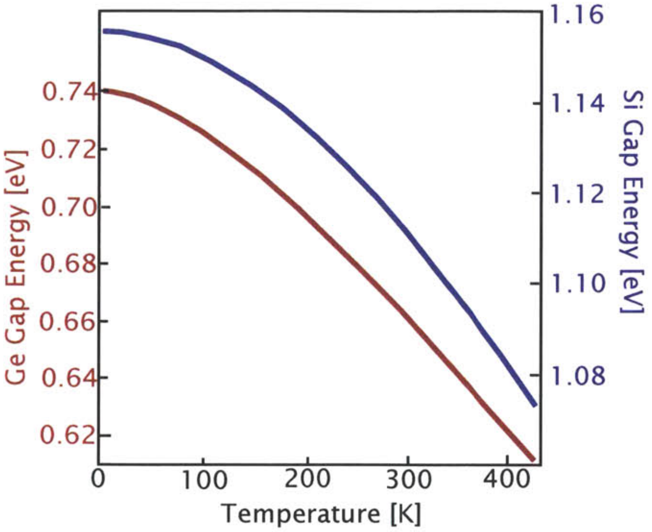

Figure 4-1: Gap energy as a function of temperature for Si and Ge, adapted from Varshni [106].

for a diamond cubic lattice is shown in Figure 4-2, and can be used as a directional version of the more familiar E-vs-k representation of a semiconductor's energy levels (the dispersion relation) shown in Figure 4-3. Here, $k$ is the usual wave vector, defined as $k = 2\pi/\lambda$. An interesting difference occurs between the otherwise similar Ge and Si cases; the minimum gap energy in Ge occurs at point $L$, whereas in Si it occurs near point $X$.

In both Ge and Si, the lowest-energy excited state for electrons happens to be at $\vec{k} \neq 0$, whereas the lowest-energy excited state for holes happens to be at $\vec{k} = 0$. This electron-hole difference has several important ramifications, the first being that holes and electrons typically have different momenta, and electron-hole recombination will require a difference in momentum be carried away by some 3rd party (a phonon). Because this is a 3-body interaction, the cross section for this process is accordingly quite low, with the important result that losing charge carriers to recombination (at least, in the crystal bulk) is a process so rare as to be completely ignored.

The second important ramification of the electron's non-zero $\vec{k}$ first excited state is that electron propagation follows certain preferred directions (while hole propagation follows simply from the direction of the force). An electron's non-isotropic behavior

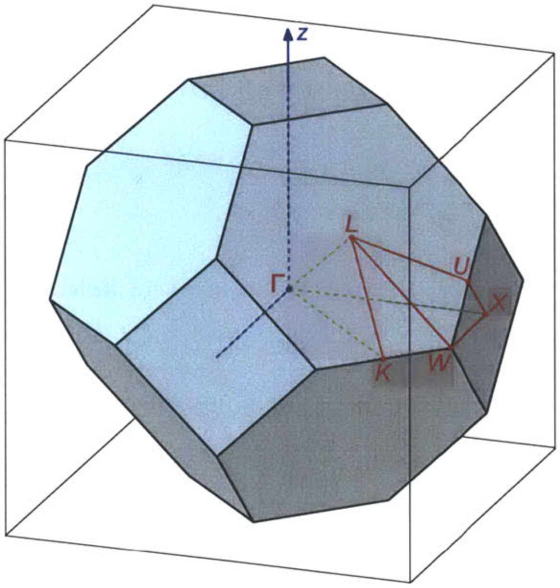

Figure 4-2: The Brillouin zone for a diamond cubic lattice, with points of special symmetry labeled according to the standard notation. Note that only one instance of each of these points is labeled (but, for example, there are six $X$ points and eight $L$ points). The outline of the diamond cubic primitive cell in the more intuitive spacial sense is also shown as a surrounding cube. Finally, the Z direction (typically, the direction of the electric field) is labeled according to the CDMS convention.

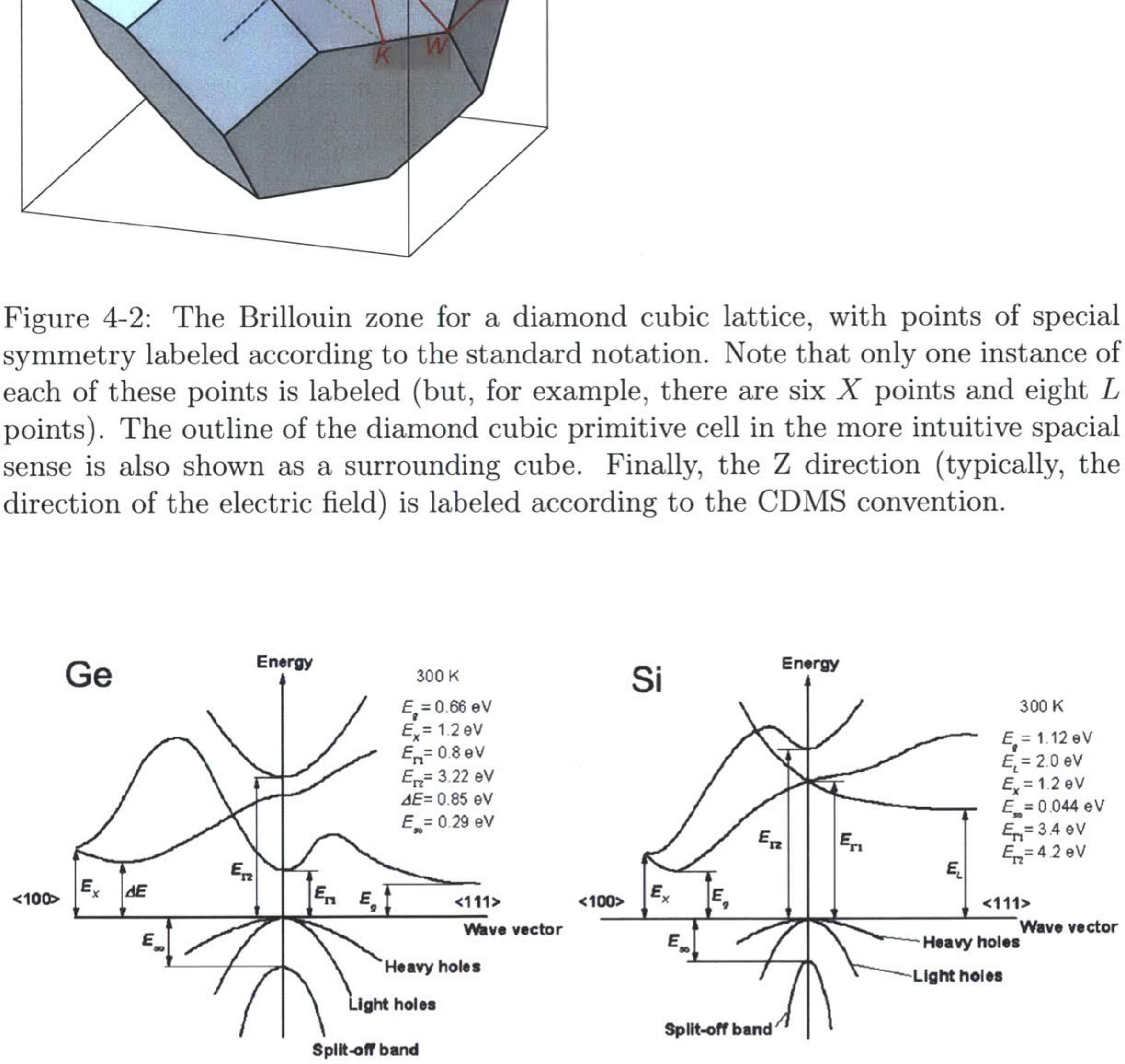

Figure 4-3: The allowed states for Ge and Si in an E vs k plane. Note that labeling of $\vec{k}$ directions from Figure 4-2 is similarly used here. The minimum gap in Ge lies along the $L$ (otherwise known as [111]) directions, and the minimum gap in Si lies near the $X$ (otherwise known as [100]) directions. From[61].

can be treated most simply by retaining $F = ma$ but changing mass $m$ from a scalar mass (as in the hole case) to a 3×3 tensor mass, corresponding to a different mass in different directions. The standard notation for this tensor *effective* mass is $m^*$, defined as

$$[m^{*-1}]_{ij} = \hbar^{-2} \frac{\partial^2 \epsilon}{\partial k_i \partial k_j} \tag{4.1}$$

where we will now denote energy as $\epsilon$ to avoid confusion with the electric field. One can see that this mass definition flows directly from taking derivatives of deBroglie relation $\epsilon = \hbar^2 k^2/2m$ in various directions. The vector basis for this tensor notation is typically chosen to be three orthogonal lattice directions (note that these directions are not typically orthogonal in real space). If the coordinates are chosen to be the $L$-valley [111] direction and two lattice directions orthogonal to it (and call this crystal basis $[\|, \perp, \perp]$), then the velocity of an electron in a Ge lattice with an external field applied can be written

$$\begin{pmatrix} v_{\|} \\ v_{\perp} \\ v_{\perp} \end{pmatrix} = \frac{2e[\epsilon\tau(\epsilon)]}{3k_B T} \begin{pmatrix} m_{\|}^{-1} & 0 & 0 \\ 0 & m_{\perp}^{-1} & 0 \\ 0 & 0 & m_{\perp}^{-1} \end{pmatrix} \cdot \begin{pmatrix} E_{\|} \\ E_{\perp} \\ E_{\perp} \end{pmatrix} \tag{4.2}$$

where, for electrons in Ge, $m_{\|}$=1.58$m_e$ and $m_{\perp}$=0.081$m_e$. Here an energy-dependent 'relaxation' rate $\tau(\epsilon)$ is used to represent the resistivity of the material. The practical effect of the non-isotropic mass (note that the degree of this non-isotropicity is ≈1.58/0.081≈19.5) is that the first excited state of the electron has certain preferred directions of propagation. A force in one direction adds momentum preferentially along certain lattice directions. Adding momentum to the perpendicular directions requires a proportionally much larger amount of energy. At the low field strengths of CDMS, electrons can be assumed to be in the lowest-energy conduction-band *only*, meaning that Ge electrons strongly prefer the momentum directions prescribed by what are called the $L$ "valleys".

As field strength increases, the momentum given to the electron at each scattering event increases, increasing the probability of a scattering event that may scatter the

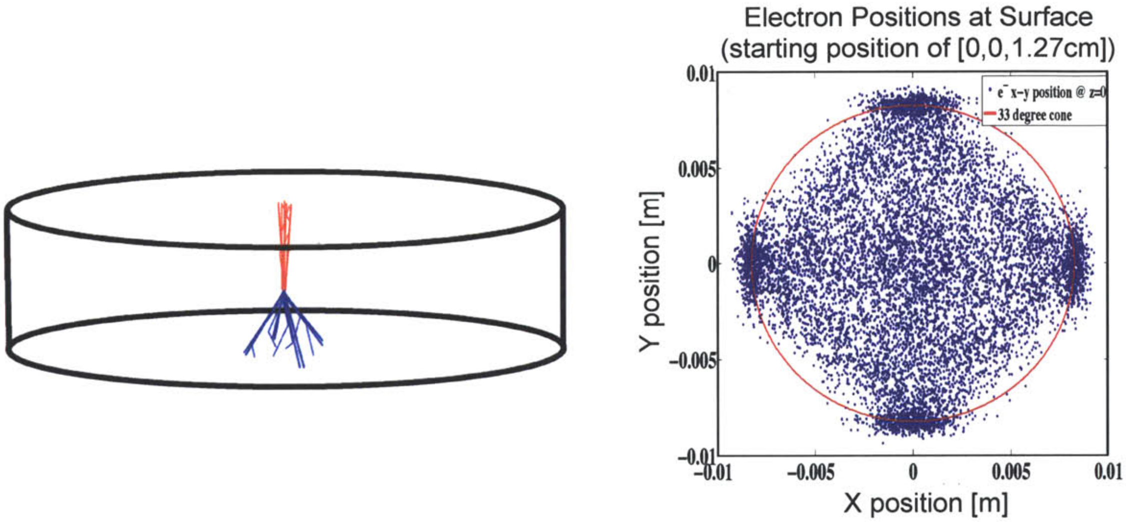

Figure 4-4: Here we see show one practical implication of the effective mass tensor discussed in the text. On the left is a cartoon of the resulting effect of this valley structure on the trajectories of charges in a Ge detector with electric field applied in the -Z direction. Here, electron trajectories are shown in blue and hole trajectories are shown in red. The plot on the right (courtesy of K. McCarthy) is the result of an electron propagation monte carlo simulation, showing the position of electrons as they reach the surface. The four main $L$ valley populations can be seen, as well as a diffuse cloud representing electrons that have undergone at least one scatter from one $L$ valley to another.

electron from one valley to another (most likely, from one $L$ valley to another $L$ valley). At field strengths greater than about 10V/cm, inter-valley scattering becomes so common that the electrons can be thought of as simply following the electric field lines (at least at macroscopic scales).

We have emphasized the situation of electrons in a Ge crystal with a field oriented in the [100] direction. If the field is instead oriented in the [111] direction, the propagation (still in the $L$ valleys) is much more similar to the field direction (the propagation is less oblique). Electrons in Si propagate preferentially in the [100] directions. The practical effects of using either [111] or [100] crystals have been explored by the author in [74].

## 4.1 Charge Loss Processes

In order to accurately measure ionization yield, charges need to propagate to the top and bottom surfaces. If some portion of the charge does not reach the surface, then the ionization yield will be underestimated, and an electron recoil can leak into the yield typical of nuclear recoils. This is obviously a big deal! In fact, understanding and tagging electron recoils with poor charge collection has been a primary focus of the CDMS collaboration for many years.

There are three main types of charge loss: sidewall, top-bottom surface, and bulk.

### 4.1.1 Sidewall trapping

The energy band structure of a semiconductor is only easily described when the lattice can be treated as infinitely repeating in all directions. At the bare cylindrical surfaces forming the sidewalls of the substrates, this is clearly not the case, and the energy levels become much harder to predict. The dangling bonds of the lattice effectively become an extremely thin layer of amorphous, rather than crystalline, structure. Amorphous Ge and Si typically have higher gap energies than their crystalline counterparts. The extreme irregularity and thinness of the side wall surfaces however, results in a highly irregular band structure (and spacial topology) in which drifting holes and electrons can very easily find a trapping site. For this reason, one can think of the bare sidewalls as effectively 100% efficient at trapping. Any charge that touches the side wall (remember that electrons in Ge have significant oblique propagation) are effectively lost, and any event at high radius will have a significantly reduced yield (even down to $Y$=0, as we shall see in later chapters). Presumably, the trapping efficiency of the sidewall varies with the lattice orientation, *i.e.*, the [110] portions of the sidewall may trap more or less than the [100] portions, but since *any* trapping ruins the yield measurement, this is an unstudied subtlety.

### 4.1.2 Top-Bottom surface interactions

The top and bottom surfaces of the substrate, where sensors are laid down, have been highly polished in preparation for fabrication, and are much less prone to trapping. Where bare polished Ge (or Si)surfaces exist, the monte carlo requires a significant amount of charge transport along this surface in order to match the data, confirming that these surfaces are very different from the rough sidewalls. The trapping on the top and bottom surfaces is dominated instead by interactions with the deposited metal layers of the sensors.

In the bulk, the initial plasma of holes and electrons is quickly separated by the external applied field. At the very initial moments after an event, however, self-interactions within the plasma dominate, and the propagation is not directional but nearly purely diffusive. If this diffusive plasma is formed close to a metal layer (within a ~10 $\mu$m length scale), carriers of *both* charges can diffuse to the surface metal, where there is no energy gap, and nothing stopping the carriers from recombining. Recombination is the enemy; we need to collect net charge not both holes and electrons. This back-diffusion has been an active field of study in the collaboration for many years. (T. Shutt famously observed, when such back-diffusion was first observed in the early 1990's, that it would be a significant problem, and may take 'months' to solve.)

The main method for reducing the back-diffusion is to add an amorphous layer (of slightly higher bandgap) between the crystal substrate and the metal layers. This higher bandgap represents a potential barrier to the diffusion process, and has been seen to prevent some amount of the back-diffusion. Given enough time, the charges may tunnel through the potential barrier due to the external field. For Ge substrates, an $\alpha$Si layer is employed for this purpose, which has the added benefit of assisting in the metal layers' adhesion.

This back-diffusion barrier unfortunately reduces but does not eliminate the effect of the so-called 'dead layer'. The tagging of events near the top and bottom surfaces, has been a major focus of both CDMS analysis efforts and design efforts.

### 4.1.3 Bulk trapping

In addition to trapping at the surfaces of the crystal, charge carriers can also be trapped within the bulk of the crystal itself. This trapping is only important when charges are asked to drift long distances by a very weak electric field, but this is precisely the situation we find ourselves in (especially with the newest detector designs). Extensive work on understanding these bulk trapping processes has been done by collaboration member K. Sundkvist, and I will not attempt to summarize his work here, only to that there are many possible trapping processes in the bulk and that this is an area of active study.

CDMS Ge bulk trapping rates as a function of electric field strength are shown in Figure 4-5.

### 4.1.4 Crystal neutralization

When charges have collected in either the bulk or external surfaces, the drift field of the crystal is significantly reduced, and also altered in direction. A weakened field promotes more trapping in a runaway process, and the detector will become unusable ('deneutralized'). The time scales for these processes can range from minutes to days. How, when, and where the charges accumulate depends (in ways that are subtle and not well understood) on the substrate, its fabrication history, and its operation history. Of course, we would like to maximize the useful time before the

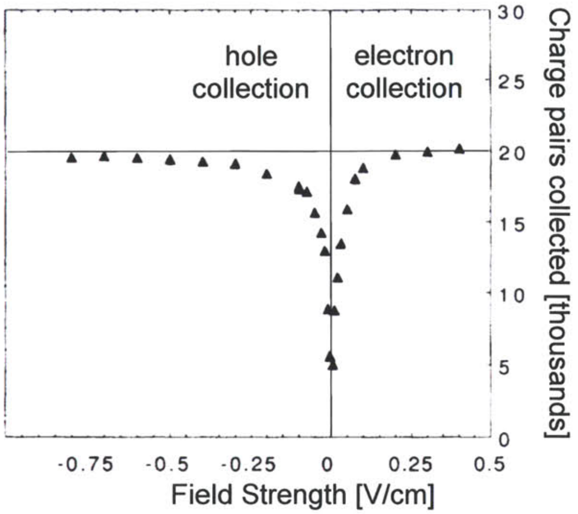

Figure 4-5: Results from a CDMS Ge trapping rate measurement, from [100]. Note that trapping becomes most important at low fields, and that holes and electrons have differing trapping rates.

crystal becomes deneutralized, so understanding these processes is another area of active research.

Both Si and Ge substrates can be neutralized through the use of infrared photons. For this purpose, all CDMS detectors have LEDs installed near the substrates, and the substrates are illuminated on a regular basis (or whenever effects of charge trapping are observed). It is unknown exactly how the photons neutralize the crystal. Especially mysterious is that the photons penetrate only very slightly ($<$ 1mm), but the photons somehow neutralize the bulk of the crystal. The working theory is that the photons excite charge carriers, which then diffuse and cause the neutralization process. Neutralization is always performed in a grounded state, so that these charge carriers can efficiently drift towards charged trapping sites and efficiently neutralize the crystal.

## 4.2 Sensing charges: The Shockley Ramo theorem

The naive (in fact, false) understanding of how charges are sensed in a detector like CDMS is that the biased electrodes attract charge to the electrode, the charge enters the metal, and a current is produced. In reality, what the electrodes sense is the movement of the charges as they drift, and in fact whether the charge enters the metal at the end of this drift process is largely immaterial to the signal amplitude.

The Shockley Ramo theorem describes the electric current induced by a charge moving near an electrode. Such a current can be thought of as the instantaneous change in the number of of electric field lines which terminate on the electrode. Quantitatively, the instantaneous current $I$ induced on a given electrode due to an electron's motion is given by:

$$I = e(\vec{E}_{electrode} \cdot \vec{v}) \tag{4.3}$$

where $\vec{v}$ is the electron velocity and $\vec{E}_{electrode}$ is the electric field at the electron's position, *due only to the sensing electrode* (setting all other conductors to ground). This field is importantly *not* the drift field, but is simply a description of the electrode's *sensitivity* as a function of position, which happens to have the same units and equation as an electric field. To calculate this sensitivity, we set the electrode bias to unity voltage and all other conducting surfaces to ground, and then calculate the electric field of the resulting situation. This "field" describing sensitivity is sometimes called the "weighting field", or more commonly in the CDMS collaboration, the "Ramo field".

In CDMS, we read out the electrode current on timescales (slightly) too slow for the instantaneous current to be useful. Instead, we simply measure the total (integral) current induced by the total drift process. Importantly, each electrode senses the movement of both holes and electrons (whereas in an ionization detector, the ions' drift is of a completely different single microsecond time scale, and typically goes unobserved). This makes the charge signal approximately independent of z-position. For a single charge carrier drifting from point $b$ to point $c$, the integral version of the Shockley Ramo theorem becomes

$$\int I = (-e)[V_{ramo}(b) - V_{ramo}(c)] \tag{4.4}$$

In the CDMS situation, where holes and electrons are created and sensed in pairs, we can imagine an event starting at point $b$, where a hole drifts to point $a$ and an electron drifts to point $c$

$$\int I = (-e)[V_{ramo}(b) - V_{ramo}(c)] + (+e)[V_{ramo}(b) - V_{ramo}(a)] = e[V_{ramo}(c) - V_{ramo}(a)] \tag{4.5}$$

The final charge signal, then, is seen to be simply proportional to the number of charge carriers created, and to the proportion of the Ramo voltage the charge carriers successfully cross without trapping.

# Chapter 5

# Phonons

We have already mentioned that phonons are copiously produced at the endpoint of both the electronic cascade (in the final collapse to the gap energy) and the nuclear cascade (where nearly all the nuclear cascade energy is transferred). Just as in the case for a semiconductor's electrons and holes, phonons exist at energies and momenta prescribed by the repeating structures of the lattice. In fact, because we are dealing with the same diamond-cubic lattice as before, we can carry over our understanding of the reciprocal (momentum) space primitive cell, and define a first Brillouin zone as before, with directions of special symmetry again denoted $\Gamma$, $L$, $X$, etc.

Phonons can be categorized as either acoustic or optical based on what type of mode is excited.

**Acoustic phonons** are the more familiar vibration modes, characterized by the fact that neighboring atoms vibrate in phase with each other. In the long-wavelength limit they are simply sound waves (hence the name). Acoustic phonons come in both longitudinal and transverse modes.

**Optical phonons** involve modes of vibration *within* the unit cell, in which neighboring atoms can be out of phase. Because they involve sub-cell structure, they are typically higher in frequency than acoustic phonons. In certain ionic crystals (but not so in the covalently bonded Si and Ge), such optical phonon modes are easily excitable by incident photons, hence the name.

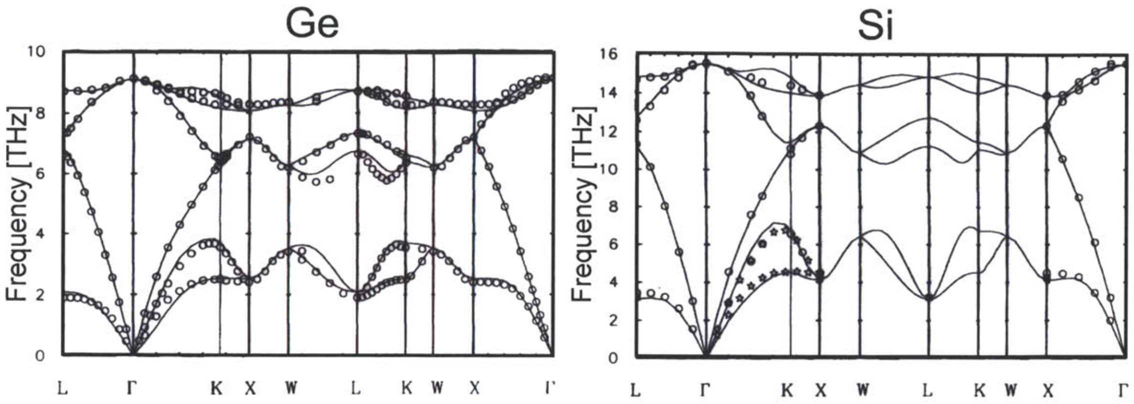

Figure 5-1: Dispersion relations for a variety of phonon modes in both Ge and Si. Theory curves are from [109], data points are from [88]. Approaching the Γ point, we see that there are three acoustic modes (corresponding to the one longitudinal and two transverse directions) along with higher-energy optical modes.

Phonon group velocity $v_g = \partial\omega_k/\partial k$ depends on both momentum and momentum direction. At low energies ($k$ low, wavelength $\lambda$ long), the dispersion relation is linear, and the phonon group velocity is $\omega a$ (where $a$ is the lattice spacing), independent of energy. Because group velocity changes slowly with frequency, a 'packet' of low-energy phonons can propagate long distances through the lattice without losing their phase relations. At high energies ($k$ high, $\lambda$ approaching $a$), this coherence fails.

Because Ge and Si share the same lattice structure, their dispersion relations appear nearly identical in Figure 5-1, simply scaled by a factor reflecting their differing atomic masses and lattice spacing. Phonon frequencies are sometimes written in a dimensionless form $\Omega$ which normalizes for these differences, as

$$\Omega = 2\pi\sqrt{\frac{\mu a^3}{e^2}}\nu \tag{5.1}$$

where $\mu$ is the reduced mass of the lattice's unit cell and $a$ is the lattice spacing as before. This expression comes from the work of Kucher, see [71]. It can be seen that low-energy phonons in Si have a group velocity approximately four times that of Ge.

Now that we have some background on the nature of the phonons in our situation, let us continue the story of an event. Both the nuclear and electronic cascades end with the production of copious phonons, and both chains dump their power mostly into the highest-energy (optical) phonon modes.

It is the nearly perfect *electrostatic* symmetry of the lattice that supports the existence of e-h excited states over long time scales. The same lattice, however, exhibits a much lower level of symmetry to vibrational excitations, leading to shorter lifetimes. Both Ge and Si are elements composed of many naturally-occuring isotopes of slightly different masses, and these isotopes are arranged randomly within the lattice. This imperfection creates a scattering cross section, strongest for the shortest wavelengths. High energy phonons exhibit high isotopic scattering rates ($\tau_i^{-1} = [36.7 \times 10^{-42}]\nu^4$ for Ge)[105], resulting in a highly diffusive propagation, with short mean free path lengths (at long wavelengths, $\ell_{mfp} = [5.4\ km/s]\tau_i = [1.5 \times 10^{44}\ m/s^4]\nu^{-4}$). There is an additional manner in which the lattice can be imperfect: if the lattice is perturbed by a second phonon. This opens up the door to phonon-phonon scattering processes (Umklapp scattering), which most likely plays a significant role at the very initial stages of an event, when energy density is still high.

Phonons not only scatter, but as they scatter they constantly decay, splitting into two lower-energy phonons. In Ge, such anharmonic decay occurs on a time scale given by $\tau_a^{-1} = [1.61 \times 10^{-55}]\nu^5$ [85]. In other words, we are seeing a third cascade process (after the nuclear and electronic cascades): the phonon cascade. There are two main differences here in the phonon situation: 1) this cascade involves macroscopic length scales, dispersing energy far from the event location, and 2) the only bottom to the phonon cascade lies at zero energy. As can be seen from the dispersion relation in Figure 5-1, there is no 'gap'; phonons can relax all the way to zero energy. The only limitation to the phonon cascade is the (enormous) heat capacity of the crystal itself. Left to their own devices, the phonon energy of the event will finally reach a thermal equilibrium (at phonon energies of $\mu$eV scales). Of course, we do hope to sense the phonon energy before it has been thermalized; and the strategy to do this is the topic of the next chapter.

Combining the energy-dependent mean free path and the energy-dependent decay rate, we can forget about the *number* of phonons and think of the phonon *energy den-*

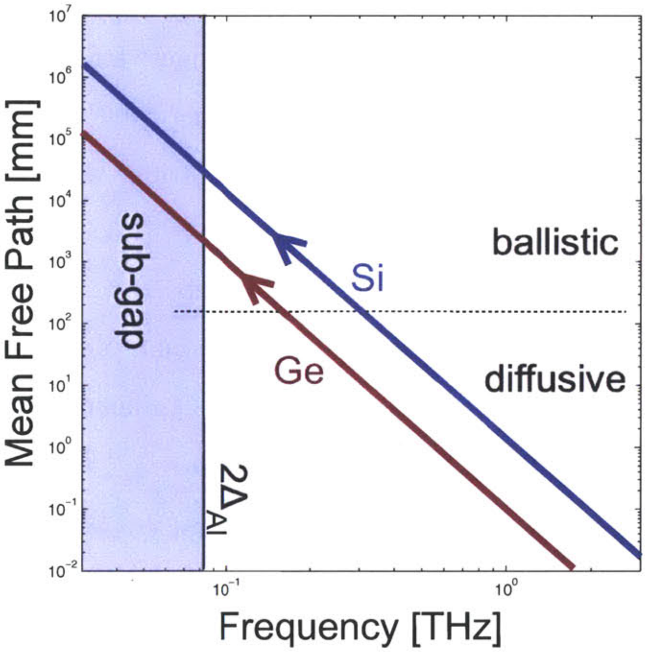

Figure 5-2: The phonon cascade and quasi-diffusion, shown as mean-free-path vs. frequency. Phonons start in the lower right corner (high energy, short mean free paths) and cascade to the upper left corner (low energy, mean free path limited by crystal size). The meaning of $2\Delta_{Al}$ will be explained in the next chapter; for now simply understand that phonons with energy below this level will not be sensed. Phonons are largely sensed within a specific energy window: between the beginning of ballistic propagation and the crossing to sub-Al-gap energies.

*sity* as diffusing out from the event location, with a mean free path (and propagation velocity) that is increasing with time. This diffusion with a time-dependent mean free path length is typically termed ‘quasi-diffusion’. To give a sense of scale, optical phonons decay to 1 THz acoustic phonons in several microseconds, and in these first several microseconds the energy diffuses ≈1 mm. Then, in the following tens to hundreds of microseconds, the mean free path between scatters approaches the crystal dimensions. At this point, reflection off of the crystal boundary is the dominating kinematic effect, and we refer to the phonons as having reached the ballistic limit.

The phonon evolution, then, can be thought of in three main stages:

1. In the immediate vicinity of the event location, high-energy (optical) phonons rapidly downconvert to lower-energy phonons, held in a tight volume (but gradually expanding) by their short (but gradually lengthening) mean free path lengths.

2. Phonons are released from this tight volume as they reach lower energies and longer mean free path lengths.

3. These ballistic phonons propagate freely through the entire crystal, reflecting spectrally off the crystal walls, and spreading the phonon energy density (nearly) uniformly throughout the crystal, as the downconverting slowly continues, eventually reaching a thermal distribution.

One last general comment about phonons should be made, and that is that the direction of $\vec{k}$ and the direction of propagation are not generally the same. As seen in the example of Si in Figure 5-3, the differing group velocities in different directions can cause so-called ’caustics’ to appear, where in phonon wavefronts from a range of $\vec{k}$ angles pile up on each other in real space. Such effects are no doubt occurring in CDMS, focusing the ballistic phonons in certain directions. Such effects, however, have not been clearly observed in CDMS, due partly to poor position resolution, and partly due to the fact that the majority of CDMS’s phonons are sensed long after the first reflection off the crystal surface.

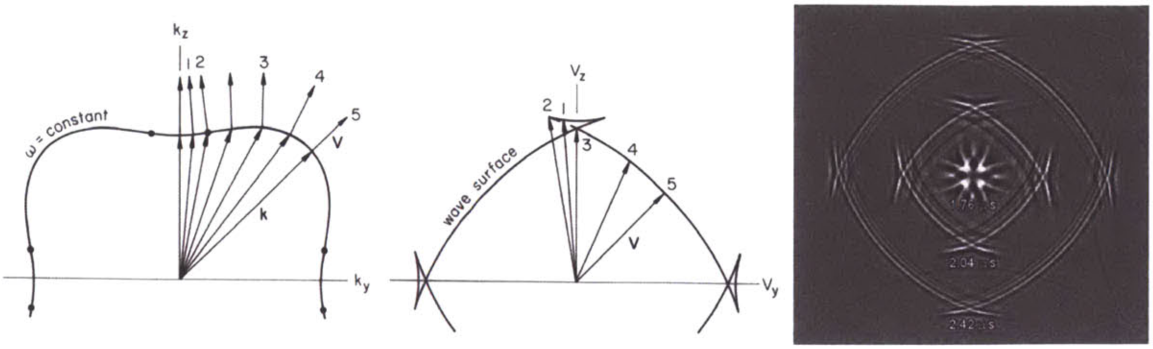

Figure 5-3: Caustics in ballistic phonons in Si, as understood and measured by Wolfe *et al.* [111]. One can see how variations in group velocity with lattice angle can cause pile-ups in real space. The Wolfe group has made many interesting measurements of such effects, including the experimental result on the right, in which three pulses of ballistic phonons were created (0.38$\mu$s appart) on one side of a [100] Si substrate, and these phonons were sensed with a high spacial resolution on the opposite side. Such images are constructed not by a highly pixelated readout, but by a single sensor and a varied event position.

## 5.1 Luke Phonons

We have discussed how phonons are sourced at the end of the nuclear and electronic cascades. In fact, there is a quite significant second source of phonons as yet unmentioned, produced by the so-called "Neganov-Luke effect".

As holes and electrons are pulled through the crystal by the electric field, *work is being done.* The charge excitations are constantly interacting with the lattice as they drift. In a classical view, one can think of the charges as being freely accelerated by the field, and then recoiling off the lattice in a way as to give off momentum, and then accelerated by field again. Given a particular field strength, the combination of external field acceleration and lattice-induced *de*celeration produce a stocastic equilibrium drift state, characterized by an average drift velocity $v_d$ (a sort of terminal velocity), an average drift power, and a distribution of resulting phonon frequencies. To give some sense of scale, Censier *et al.* [37] have measured drift velocities in our situation (high-purity Ge, low field strength, [100] orientation, 10s of mK) to be $v_d = 1.67\times10^4$ m/s for holes, and $v_d = 2.37\times10^4$ m/s for electrons [at 20 mK and 0.5 V/cm].

The potential energy of a field would normally be transferred to a charge's kinetic energy. Here, because the kinetic energy is kept low (typically 10s of meV) by the lattice interactions, we can assume that all the potential energy lost by the charge through drift has been expressed as phonons. Thus, if we imagine $n_h$ holes and $n_e$ electrons drifting, starting from the middle of some total potential difference $\Delta V_{tot}$,

$$E_{luke} = n_e e \Delta V_e + n_h e \Delta V_h = ne(\Delta V_e + \Delta V_h) = ne\Delta V_{tot} \tag{5.2}$$

Note that the total power is independent of where in the crystal the event charges start. If holes and electrons are created in equal number, then only that number and the *total* voltage difference matter. (This is a similar situation to the *charge* signal discussed earlier).

Although they release the same amount of total phonon power, holes and electrons produce differing phonon spectra as they propagate (a direct result of their differing effective masses and terminal velocities). These Luke phonon initial spectra (remember, Luke phonons will downconvert just like any phonon) are note well-measured, and have instead been modeled from first principles. The Luke spectrum from electron propagation will be somewhat lower in energy than the spectrum from hole propagation (given the same field strength). K. Sundqvist [104] has produced simulations predicting the acoustic phonon fraction for holes and electrons, and G. Wang [107] has published analytic descriptions of the initial spectrum of acoustic phonons, as seen in Figure 5-4.

Luke phonons are different from the phonons originating with the nuclear and electronic cascades (we will call these the 'primary' phonons) in several ways. Luke phonon production depends not just on the energy of the event, but on the potential difference of the detector (which can be tuned by the experimenter). Luke phonons are created along the path of the charge drift, rather than at the recoil position. They are created significantly in the *ballistic* regime rather than the diffusive regime, meaning that they largely skip the initial slowly-propagating diffusive stage (more quickly reaching phonon sensors). And, that portion of Luke phonons that are initially ballistic will retain their original momentum direction (until reaching a surface),

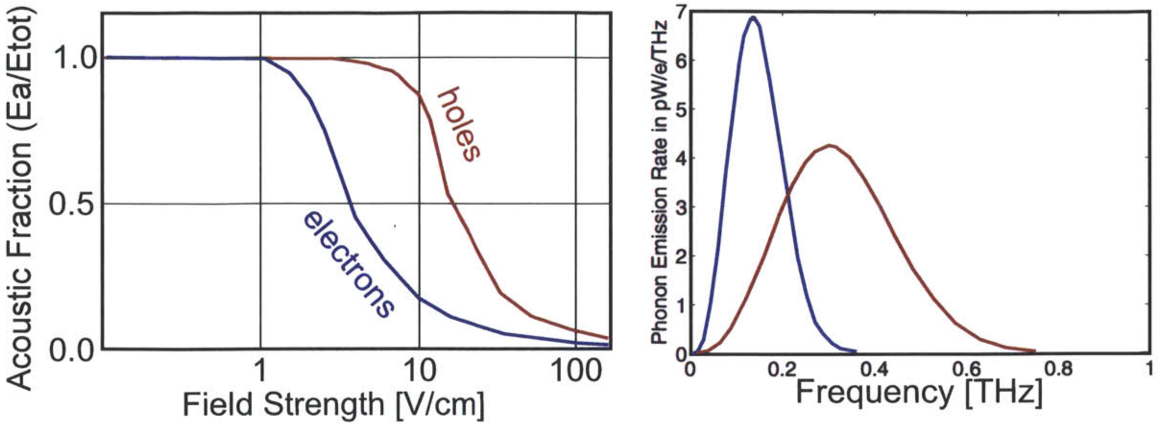

Figure 5-4: On the left, adapted from K. Sundqvist [104] is displayed a prediction for the fraction of Luke phonon energy initially expressed as acoustic (lower-energy) phonons, as a function of drift field, for both holes (red) and electrons (blue) in germanium. On the right, at a specific low field strength (3 V/cm), we see an analytic approximation for the initial Luke phonon energy spectrum, again for both holes and electrons, from the work of G. Wang [107].

meaning that if a charge is propagating toward a top surface, the Luke phonons from that propagation will similarly be directed (generally) towards the top surface. Luke phonons are clearly very information-rich (more on that in later chapters).

## 5.2 Recombination Phonons

When electrons and holes are first created, they have two types of potential energy: the energy of the gap and the energy imparted by the externally applied voltage. The energy of the external field is transferred to Luke phonons during the drift process. The gap energy is transferred to *recombination* phonons when the charge carriers reach the metal layers at the surface and drop in energy. Thus, the energy stored in the charge carriers is eventually transferred to the phonon system, emanating from the surfaces. This means the entire recoil energy (minus some small lattice defect energy) eventually ends up in the phonon system. The total phonon energy released by an event can be grouped as

$$E_{Ptot} = E_{primary} + E_{recombination} + E_{Luke} = E_{recoil} + E_{Luke} \tag{5.3}$$

# Chapter 6

# Concentrating and Sensing Phonon Energy

As discussed in the previous chapter, both primary and Luke phonons gradually down-convert, and would eventually settle into a thermal distribution. Several dark matter detection efforts (EDELWEISS, and the early BLIP version of CDMS) work by sensing this temperature change in the target mass using a thermometer in contact with the target. This strategy has the advantage of simplicity (not to be overlooked!) but has two main disadvantages:

1. Given an event of a particular energy, a large target volume (with a large heat capacity $C$) will have a correspondingly low final temperature increase, as $\Delta T = E/C$. *Both* a large target mass and a low energy threshold are vital to direct detection, and the thermal approach puts these two goals in opposition.

2. Thermalization is a process of taking energy of high information content, spreading that energy over as many available degrees of freedom as possible, and to the highest-entropy (lowest information-content) state possible. If we could measure the phonons before the second law of thermodynamics takes over, then we could hope to extract all kinds of information, including the position of the event, the precise timing of the event, and ideally event the recoil type (nuclear or electron recoil) of the event.

This chapter describes the method used by CDMS detectors to absorb the phonon energy *athermally*, before the huge target heat capacity plays any role, and before much of the event information is lost.

The basic requirements of any such strategy would be

1. a long-lived excited state, with a characteristic energy in a specific range: much higher than the thermal phonon energies, but low enough as to be excitable by the majority of ballistic phonons.
2. an efficient process of converting phonons into that excited state.
3. an efficient process of transporting and concentrating that converted energy into some appropriate sensor.

These are the topics of the next three sections.

## 6.1 Bogoliubov Quasiparticles

The excited states used by CDMS are so-called Bogoliubov quasiparticles (sometimes referred to as Bogoliubov-deGennes quasiparticles), a discussion of which will require first some explanation of superconductivity; the realm in which these quasiparticles exist. A theoretical understanding of superconductivity is a rather recent development, first laid out by Bardeen, Cooper, and Schrieffer in 1957 [21].

In a normal metal at absolute zero, energy is minimized by eliminating all phonons, and lowering all electron momenta to the minimum allowed by the Pauli exclusion principle, the Fermi momentum. In a superconductor, a different state minimizes the total energy, in which pairs of electrons are locked together through a phonon exchange force into a quasiparticle referred to as a Cooper pair. These electrons have momenta above the Fermi surface, the total kinetic energy is certainly not minimized, but the total energy is (non-intuitively) lower than the alternative. The electrons have lowered their *potential* energy by falling into this phonon-mediated bonding. It is still rather unintuitive to think of a state with non-zero kinetic energy to be the

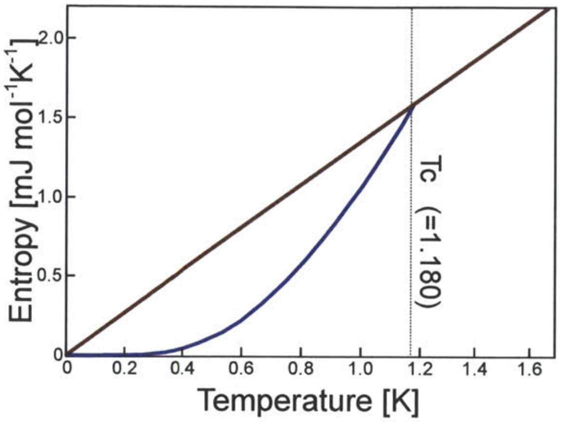

Figure 6-1: Entropy of aluminum, as a function of temperature, for both the normal and superconducting phases, after [68]. (The normal phase can be achieved below $T_c$ through the application of a magnetic field.)

ground state of a system. More intuitive is thinking in terms of entropy, as pictured in Figure 6-1: due to the additional bonds present (in this case, phonon-mediated electron-electron bonds), the superconducting state is a more ordered state compared the normal one, and if a material possesses a lattice capable of joining two electrons into a Cooper pair, it will. It so happens that a high degree of order must already be present in the system for such a delicate electron-lattice dance to occur at all. The first Cooper pair bonds together at the material's "critical" temperature $T_c$, and more and more Cooper Pairs are paired as the temperature decreases. This understanding also tells us that the phase transition occurring at $T_c$ is a second order transition, where no latent heat is required.

The occupation of energy levels in a metal is dictated by their Fermi statistics. A Cooper pair on the other hand is a quasiparticle of net spin zero, and thus obeys instead a bosonic statistics, meaning that an infinite number of Cooper pairs can occupy a common ground state. This collective many-particle ground state can be described by a single wavefunction, and it is the phase coherence of the pairs that is the origin of the perfectly zero resistance. It so happens that this Cooper pair ground state energy is equal to the Fermi energy. Imagine electrons in a superconductor with slightly more or slightly less momentum than the Fermi momentum. Such states

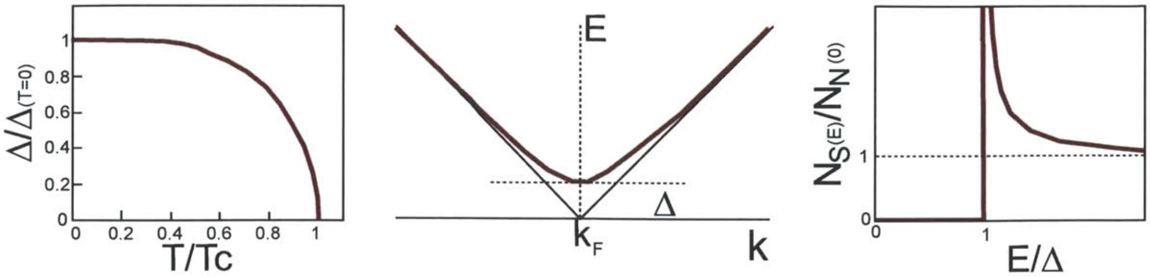

Figure 6-2: Some description of the superconducting gap Δ. On the left, the gap energy temperature dependence, which in these generalized dimensionless units takes the form for all superconductors. In the middle, the 3rd significant dispersion relation in this thesis, that of single electron states in a superconductor, showing the gap energy (very analogous here to the semiconductor gap). On the right, the density of single electron states, above and below the gap energy (normalized by the normal-metal density of states).

are quickly swept up by the many-particle ground state; in fact some range $\pm\Delta$ is a *forbidden* energy range, producing a gap of $2\Delta$ (straddling the Fermi energy). As seen in Figure 6-2, $\Delta_{(T)}$ rises rapidly from zero just below $T_C$, while at the lowest temperatures, $\Delta_{(T)}$ is nearly constant. At $T = 0$, the value of $\Delta$ is proportional to the $T_c$ of the material, or more specifically: $\Delta_{(T=0)} = 1.764 k_B T_c$. This relation makes some intuitive sense, because only electrons with an energy of order $k_B T_c$ could be expected to play a role in a phenomenon that has a threshold of $T_c$. For Al near T=0 (*i.e.*, in the CDMS context), $\Delta = 170\mu\text{eV}$, and the ground state energy of a Bogoliubov quasiparticle is $2\Delta = 340\mu\text{eV}$.

Bogoliubov quasiparticles are not simply electrons with a certain amount of energy, but a coherent superposition of a hole and an electron state. The superposition is characterized by a mixing angle (the 'Bogoliubov angle'). As the energy approaches the $E_F$, this mixing becomes a maximum. Δ is small, and so Bogoliubov quasiparticles in CDMS (near the gap energy) are best considered neutral particles. The lifetime of the quasiparticle ($\tau_{qp}$) is proportional to the density of quasiparticles in the material (remember, the ground state is the Cooper pair), and is therefore quite long at near-zero temperature, where the density of thermally-excited quasiparticles is similarly near-zero.

## 6.2 The Quasiparticle-Phonon Cascade

In CDMS, some portion of the Ge surface is covered in thin Al films, and phonon energy incident on these films is transferred to quasiparticle excitations, with some efficiency depending on the energy of the phonon, the minimum excitation energy ($2\Delta_{Al} = 340\mu$eV), and the thickness of the film (in CDMS, Al thicknesses are typically hundreds of nanometers).

The process of quasiparticle creation is yet another cascade process. Here, the incoming phonon typically has an energy significantly higher than the minimum excitation energy. The initial quasiparticles relax to lower energies through phonon emission, and these phonons can easily have enough energy for additional quasiparticle creation. The cascade, then, is a combined quasiparticle-phonon cascade (phonons have a quasiparticle 'stopping power', and quasiparticles in turn have a phonon 'stopping power'), resulting in some population of ground-state quasiparticles and some population of phonons of sub-gap energies.

Using the quasiparticle density of states

$$\rho_{(E)} = \frac{E}{\sqrt{E^2 - (\Delta)^2}} \tag{6.1}$$

we can write an expression for the initial energy distribution of quasiparticles created by phonons as

$$P_{QP}(E) = \rho_{(E)}\rho_{(\Omega-E)}\left(1 + \frac{(2\Delta)^2}{E(\Omega - E)}\right) \tag{6.2}$$

where $E$ is the quasiparticle energy, $\Omega$ is the phonon energy. $\Omega$ and $E$ are subject to the constraint that $2\Delta \leq E \leq \Omega - E$. Similarly, quasiparticles relax through phonon emission, and these phonons follow an initial energy distribution given by

$$P_{\phi}(\Omega) = \Omega^2\rho_{(E-\Omega)}\left(1 - \frac{(2\Delta)^2}{E(E - \Omega)}\right) \tag{6.3}$$

where $0 \leq \Omega \leq E$.

Throughout this quasiparticle-phonon cascade, phonons can 'leak' from the aluminum back into the Ge and propagate through the crystal. If the film is thicker, this leakage is smaller. 'Thick' and 'thin' are relative to the characteristic phonon interaction length in Al at T=0, which is $\approx$ 720 nm [32], which can be thought of as the approximate scale for the volume in which the cascade occurs.

Assuming the incident phonon energy is much greater than $2\Delta_{Al}$, and assuming zero phonon leakage back into crystal, Klein [69] finds that a quasiparticle cascade in Al divides the energy as 55% quasiparticles and 45% sub-gap phonons. P. Brink [32] has extended this towards lower energy incident phonons (where the quasiparticle efficiency goes to zero). In the CDMS case, where phonon leakage is significant, Al quasiparticles capture something less than half of an incident phonon's energy.

## 6.3 Concentrating Quasiparticles I: Gap Energies

As mentioned, Bogoliubov quasiparticles are essentially chargeless, which means these excited states are much harder to manipulate than holes and electrons. Varying the voltage in the crystal is what draws holes and electrons towards the electrodes; how can we vary the potential energy of the Bogoliubov quasiparticles, to pull them into a sensor, if they have no charge?

Remember that the superconducting gap energy depends on the transition temperature of the material, specifically $2\Delta_{(T=0)} = 3.528 k_B T_c$. If we could somehow vary the $T_c$, then we could create a potential energy gradient, much like an electric potential gradient, and this is in fact the trick CDMS uses to pull quasiparticles from the Al film into the sensor.

The gradient is formed by using two materials: Al (an initial material with a high $T_c$) and W (a 'trap' material with a low $T_c$).

**Al** $T_c = 1.180K \rightarrow 2\Delta_{Al} = 340\mu eV$

**W** $T_c \approx 80mK \rightarrow 2\Delta_W \approx 20\mu eV$

The basic idea is that diffusion propagates the Al quasiparticles to the Al-W interface, some of the quasiparticles cross the interface, lower in energy to $2\Delta_W$, and cannot cross back into the high-gap Al. Of course, this is a simplistic picture.

First, a word of explanation about tungsten. Tungsten exists in phases: $\alpha$ and $\beta$. $\alpha$-tungsten is a body-centered cubic structure, and is the most stable form (it has the lowest potential energy). $\beta$-tungsten forms what is called an 'A15 cubic' structure, in which each cubic face is filled with two central atoms instead of only one. $\beta$-tungsten is of a slightly higher potential energy, but is metastable, and the two phases can coexist in a mixture (at least, in the thin-film case of CDMS, not so in the bulk case), with material properties dependant on the $\alpha$:$\beta$ ratio. $\alpha$-tungsten has a low $T_c$ (15mK), while $\beta$-tungsten has an unusually high $T_c$ (~600mK for a 40nm film), and CDMS delicately balances the $\alpha$:$\beta$ ratio during fabrication to achieve the desired tungsten $T_c$ of $\approx 80mK$. For details of this tuning, see [62]. This tuning remains the most delicate step of the CDMS fabrication process.

Tungsten with two phases in a mixture behaves like one material. It has *one* $T_c$, because the particles of superconductivity (Cooper pairs and Bogoliubov quasiparticles) are extended objects with characteristics reflective of material properties larger than the W crystal grain size. This scale is the coherence length, and extends to ~100 nm in Al [33] and ~340 nm in W [62]. These scales are largely a function of the material's Fermi velocity (remember that both Cooper pairs and Bogoliubov quasiparticles are at approximately the Fermi energy). For our purposes, the coherence length has two main practical effects:

- In W, the $T_c$ is a single value for the mixture of lattice types.
- A W-Al bilayer will have an intermediate $T_c$ (at least, within $\pm \sim 100$ nm of the interface).

With W, Al, and a W-Al bilayer, then, we can easily achieve three distinct quasiparticle energies, and we can arrange them so that the quasiparticles get trapped in sequentially lower-gap materials. At each step down in gap energy, the quasiparticles give off phonons, and the phonon-quasiparticle cascade discussed earlier is repeated (with some phonon leakage, as before).

In addition to the material properties, we can vary the gap energy by varying the

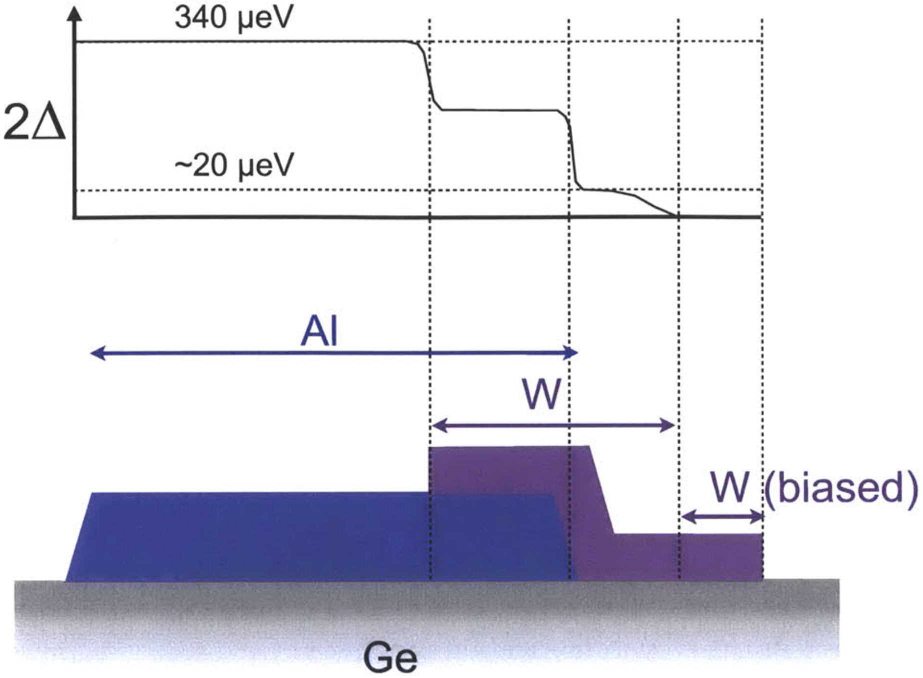

Figure 6-3: A depiction of the physical structures of the quasiparticle gap (bottom) and their resulting quasiparticle energies (top). Quasiparticles in the aluminum film gradually diffuse to regions of lower gap energy (right) and become trapped.

temperature of the material. In CDMS, the sensor consists of W with a bias current flowing through it that heats the electron system up to the tungsten $T_c$. Referring back to Figure 6-2, we see that as we increase temperature to $T_c$, the gap energy goes smoothly to 0. This positive thermal gradient serves as the last step down in quasiparticle energy.

Note that more gradual the gradient, the more adiabatic the transition (quasiparticles lower in energy without releasing phonons and starting a quasiparticle-phonon cascade) and the steeper the gradient, the more energy-conserving the transition.

These trapping ideas are displayed schematically in Figure 6-3.

What is accomplished by the creation and concentration of Bogoliubov quasiparticle excited states is something quite remarkable: phonon energy from a huge volume has been concentrated into a sensor volume $\sim 10$ *billion* times smaller.

## 6.4 Concentrating Quasiparticles II: Diffusion

Here, we lay out a simple model for the diffusion of energy through the Al and into the W, following closely the understanding of M.Pyle in [92] and [93].

As previously stated, quasiparticles *diffuse*, and the quasiparticle density $n_{qp}$ therefor follows the diffusion equation

$$\frac{\partial n_{qp}}{\partial t} = D_{Al} \nabla n_{qp} \tag{6.4}$$

, characterized by a single number, the diffusion constant $D_{Al}$, with units of [length$^2$ time$^{-1}$]. Similarly, there exists a diffusion constant $D_W$ for the tungsten case. Unfortunately, this is an overly optimistic model, for two reasons:

**Trapping** The quasiparticle recombination rate is small enough as to be negligible, but quasiparticles can find film defects (with lower gap energies), where the quasiparticle has some probability of trapping. (Presumably, the trapped quasiparticle eventually recombines, but this is beside the point.) Our Al, W, and Al-W bilayer films each have some trapping time scale ($\tau_{trapAl}$, $\tau_{trapW}$, and $\tau_{trapAl/W}$). The simple diffusion equation becomes $\frac{\partial n_{qp}}{\partial t} = D\nabla n_{qp} - \frac{n_{qp}}{\tau_{trap}}$. A more natural quantity in practical terms is the trapping length scale, $l_{trap} = \sqrt{D\tau_{trap}}$.

**Crossing impedence** If a quasiparticle is successful enough to diffuse to the Al-W interface, there is still a barrier to crossing over this interface. The physical origins of this barrier are unclear, but the most likely explanation is a layer of oxide. CDMS goes to great lengths to avoid an oxide layer growing on the Al before the W is deposited. Specifically, all Al is covered with a thin W layer immediately after deposition. The 'active' W layer (which forms the TESs) is then laid on top of this initial layer. The current best-guess is that there is in fact an oxide layer separating these two W layers. We can quantify this crossing impedance using a dimensionless transmission probability $f_{Al/W}$ (where the number of crossing particles is equal to the number of incident particles times $f_{Al/W}$). The boundary condition prescribed by this description

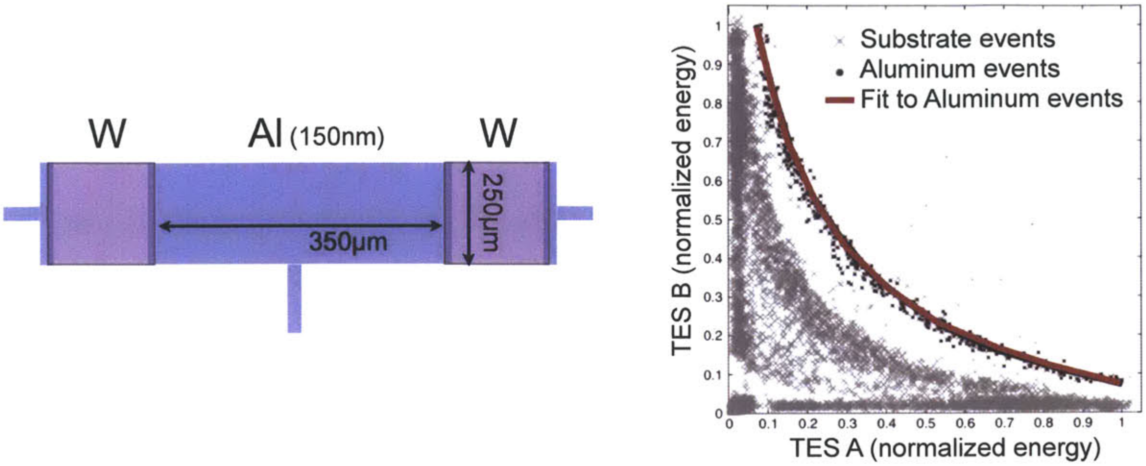

Figure 6-4: The 1998 test device for determining diffusion parameters, described more fully in the text.

is $-D_{Al}\nabla n_{qp} \cdot \hat{s} = [f_{Al/W} v_{qp}/2] n_{qp}$, where $\hat{s}$ is the unit vector pointing into the W and $v_{qp}$ is the mean quasiparticle velocity.

In 1998, CDMS designed and ran a special test device (shown in Figure 6-4) specifically to measure the diffusion characteristics of our Al and W films. This device has been called a 'banana' device, due to the shape of the energy partition plot. When an event occurs in the Al but very close to one W sensor, quasiparticle diffusion deposits much of the energy into that one sensor. When an event occurs more in towards the middle of the Al, the energy is more equally partitioned between the two sensors, and the overall amount of energy is reduced by quasiparticle trapping in the Al. Notice, also, that even when an event occurs very close to one of the sensors, there is some amount of sharing. This sharing is only possible because of a significant crossing impedance.

The diffusion parameters that best fit the data were $l_{trapAl} = 180 \pm 10\mu m$ and $f_{Al/W} = 0.0025 \pm 0.0005$, and with the addition of timing information, $D_{Al} = 100 \pm 10\frac{cm^2}{s}$, $\tau_{trapAl} = 3.2 \pm 1\mu s$. Applying these results to understanding and designing other sensors is difficult, though, for the following reasons:

1. It is not known whether the quasiparticle mean free path in Al is limited by internal scattering or surface scattering. The *guess* is that diffusion is limited

by film thickness, meaning that the diffusion constant scales as film thickness $h_{Al}$, and $l_{trapAl} = \sqrt{h_{Al}/150nm}[180 \pm 10\mu m]$.

2. Even normalizing for thickness, there may be little correlation between the $D_{Al}$ of the 1998 test device and the $D_{Al}$ of other Al films. It is known from RRR measurements (comparing resistivity at two drastically different temperatures) that CDMS Al films (and presumably the Al-W interface as well) are of widely variable quality.

3. The diffusion characteristics of both the Al-W overlap region and the W itself are poorly constrained by the 1998 device, which was designed primarily to test Al properties. It is thought that $l_{trapW} << l_{trapAl}$ and that $l_{trapAl/W}$ should be between the two, but this is far from certain. A comparison of two detector designs that varied the Al-W overlap dimension provides a lower limit that $l_{trapAl/W} \gtrsim 25\mu\text{m}$, with large uncertainties.

Another important thing to understand is that the Al of the 1998 device appears to have a much shorter quasiparticle trapping timescales than what is achieved by other research groups. As one example, the CRESST collaboration fabricated similar test devices and measured $\tau_{trapAl} \approx 100\mu s$ [49] (thirty times better). The main difference here could be the fact that CDMS deposits Al through a sputtering technique (with smaller crystal grain sizes), whereas other groups use e-beam deposition (with larger crystal grain sizes). We are limited to sputtering by our desire to deposit the Al and the W in the same (CDMS-dedicated) machine (the 'Balzers').

M. Pyle has spent considerable effort understanding the effects of material properties and geometries on the efficiency of the diffusion process, and I encourage the interested reader to his thesis. Such models are of only limited utility on future design work, unfortunately, until the model parameters are more reliably known.

We have mentioned many reasons to be depressed about quasiparticle diffusion in CDMS: the diffusion parameters are largely unconstrained, and those that we do know are not encouraging. So that we don't leave this topic still feeling so negative, here are the positives:

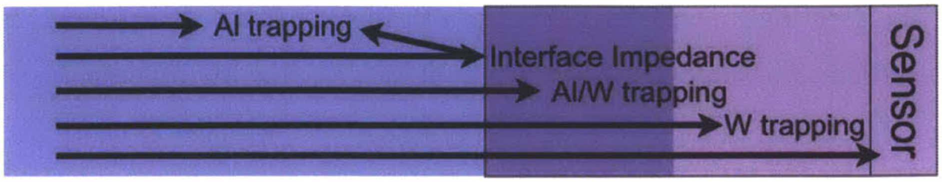

Figure 6-5: The main constraints on the quasiparticle diffusion process: trapping in Al, the Al-W bilayer, and the (superconducting portion of the) W. The transmission impedance is not an energy loss itself, but its effect is to increase trapping in the Al.

Figure 6-6: A scanning electron microscope image of an Al-W interface from CDMS II.

1. The only way to go is up! The 1998 device was fabricated before the Balzers was a CDMS-dedicated facility and before the W etch had been better optimized. It seems plausible (although not proven) to assume that there are now fewer impurities in our Al, and the W is now removed from the Al more completely. Any pocket of W on the Al surface could lower the local $T_C$, creating a trapping site.

2. Material properties (and the properties of various geometrical effects) are currently being thoroughly studied by Jeff Yen at Stanford. The results of these studies will hopefully not only pin down the unknown material constants, but also test many geometrical effects.

3. Even with all the trapping, the diffusion process is observed to be fairly efficient (more on this in later chapters).

## 6.5 Transition Edge Sensors

We have followed the flow of energy from recoiling electron or nucleus all the way to quasiparticles and electrons cascading towards thermal equilibrium in microscopic tungsten calorimeters. These sensors are amazing devices themselves, and deserve far more discussion that we will present here. In its simplest description, a Transition Edge Sensor (TES) is a superconductor which is kept in the middle of its (sharp) transition between zero and finite resistance, meaning that a tiny amount of extra heat can dramatically change the superconductor's resistivity. This technology is one of the most sensitive calorimetry techniques available. Understanding and optimizing TESs for a variety of energy ranges and environments is a complex and fruitful area of research, but here we lay out only the barest of descriptions. Excellent resources for a deeper discussion include a review article by Irwin*et al.* [63], the thesis of E. Figueroa [54], and the recent thesis M. Pyle [92]. (Because the following discussion follows entirely from these references, citations are omitted.)

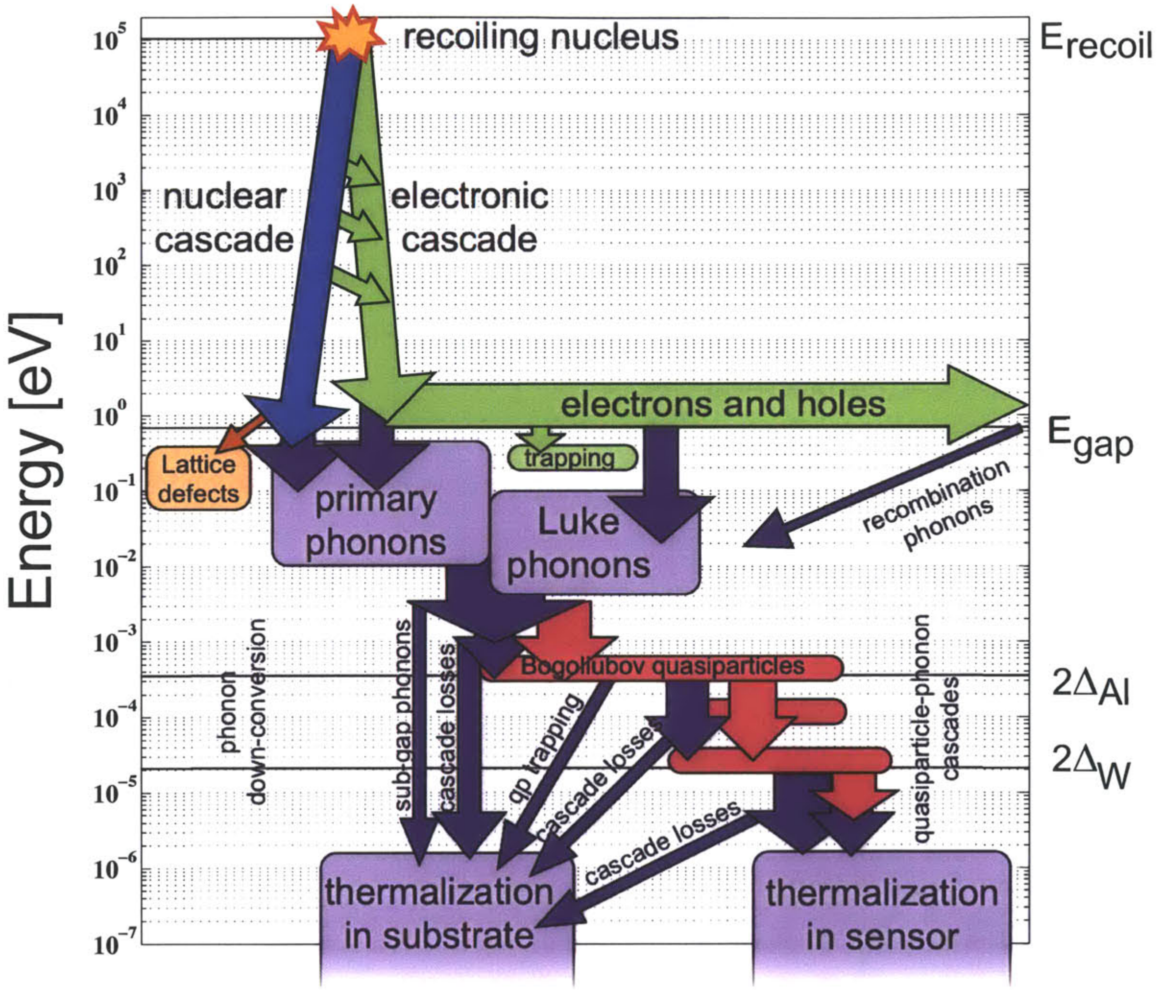

Figure 6-7: A cartoon summarizing the flow of energy through a CDMS detector, highlighting the cascade processes and the long-lived excited states. Note that after all the flowing of energy, there are essentially four places the energy can flow *to*: thermalization in the substrate (most of the energy), thermalization in the TES, the creation of lattice defects (in the case of a nuclear recoil), and the trapping of e-h pairs in lattice defects. Notice, too, that all these processes have run their course on a millisecond timescale after the recoil.

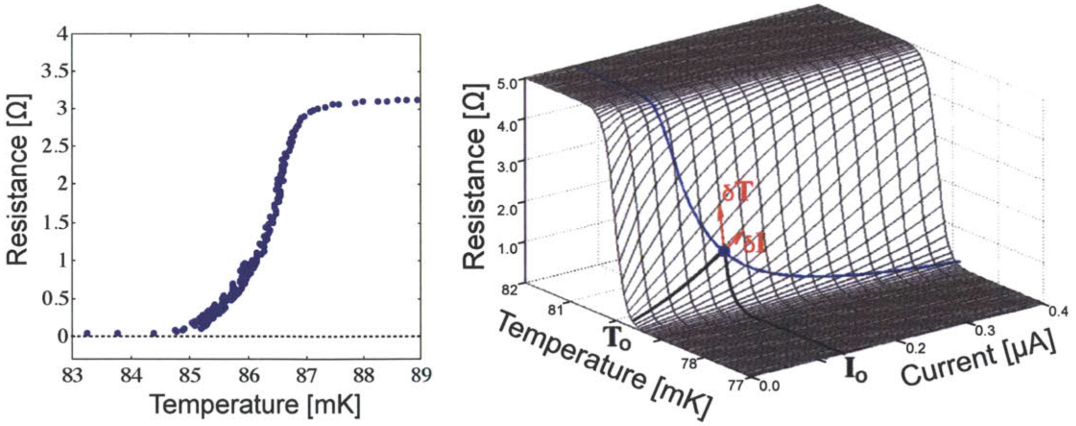

Figure 6-8: Two representations of a tungsten transition edge. On the left, we show a measurement (from the Cabrera Group) of a sensor resistance as a function of temperature (at near-zero bias current). On the right, we illustrate the $R_{(T,I)}$ surface for a typical CDMS channel. For a given substrate temperature, the blue line represents those bias conditions available to the system as the bias current is varied. For a given bias condition (blue dot), we label $T_0$, $I_0$, $\delta T$, and $\delta I$, as referred to in the text.

First, we should intuitively understand the magic of how a TES is kept in the middle of its phase transition. As mentioned earlier, the TES (or at least, the electrons in the TES) are heated above the temperature of the surrounding thermal bath (*i.e.*, the substrate it is sitting on). This heat comes from a current which is flowing through the TES, which in turn is caused by the application of a *constant potential* placed across the TES (the TES is 'voltage-biased'). Imagine placing a small amount of heat $\delta P$ into the TES. The resistance will go up, which (at our fixed voltage) will then reduce the current, which will reduce the electrical heating and return the TES to its original resistance. The other direction is also true: a small amount of heat is removed, the resistance will decrease, the electrical heating will increase, and the resistance will recover its original value. This basic idea is called 'electro-thermal feedback'.

Voltage biasing (and the resulting electro-thermal feedback) naturally biases a TES within its transition in a stable manner. Even more impressive, it can easily accomplish this trick for many sensors *with difference* $T_c$ biased in parallel with the same voltage (the CDMS situation). We have mentioned how our tungsten $T_c$ is

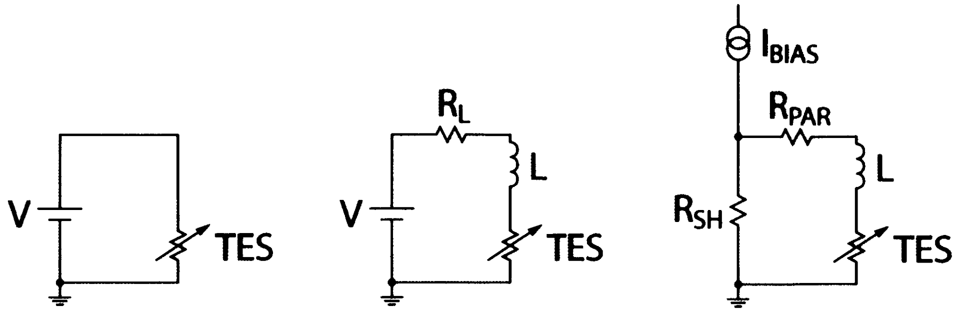

Figure 6-9: The TES input circuit, in sequentially more realistic representations. On the left is shown the most basic idea: a TES is voltage-biased in its transition. In the center, we see the introduction of an inductor to enable an inductively-coupled readout circuit. And on the right, we add in the mechanism by which we achieve a voltage-biased state: we current-bias with a low-resistance shunt in parallel to the TES (and inductor). The center illustration is the Thevenin-equivalent representation of the realistic circuit to the right, with load resistance $R_L = R_{SH} + R_{PAR}$, and bias voltage $V = I_{BIAS}R_{SH}$.

delicately balanced during fabrication (through the mixture of two phases). In fact, there always exists some amount of $T_c$ variation across the face of a detector (frequently, the gradient is radial from center to edge of the substrate). Assuming the difference in $T_c$ is not *too* great, given a constant voltage, each TES will naturally settle into a current specific to that sensor's $T_c$. By the way, the melding of quasiparticle trapping and an electrothermally-stabilized TES as in CDMS is referred to as a 'QET' (Quasiparticle-trap-assisted Electrothermal-feedback TES), at least within the collaboration.

Let's add a little meat to that qualitative description of the feedback mechanism, and describe both the thermal and electrical natures of the TES as a pair of coupled differential equations.

First, a simple heat description will read

$$C\frac{dT}{dt} = -P_{bath} + P_{Joule} + P_{input} \tag{6.5}$$

where $C$ is the heat capacity of the TES, and the three powers are the heat flowing from the TES to the bath, the Joule heating from the current through the TES, and the heat flowing into the TES from the quasiparticle cascades. It should be stressed that the thermal system we are discussing is the *electron* system, which is not at all in equilibrium with the lattice of the TES at these temperatures. The $P_{bath}$ is really the heat flowing within the TES from the electron system to the phonon system (and then to the substrate lattice), and can be written as $P_{bath} = K(T^5 - T^5_{bath})$.

Second, a simple electrical description will read

$$L\frac{dI}{dt} = V - IR_L - IR_{(T,I)} \tag{6.6}$$

where $L$ is the inductance, $V$ is the bias voltage, $I$ is the current through the TES, and $R_{(T,I)}$ describes the TES resistance as a function of temperature and current.

Let us look at the solutions to these coupled equations in the small signal limit, where we linearize about the stable bias point described by $[R_0, T_0, I_0]$.

To expand $R_{(T,I)}$ about the equilibrium position, we simply write

$$R_{(T,I)} \approx R_0 + \left.\frac{\partial R}{\partial T}\right|_{I_0} \delta T + \left.\frac{\partial R}{\partial I}\right|_{T_0} \delta I \tag{6.7}$$

where the two derivatives (that describe the steepness of the transition with temperature and current) are commonly referred to as $\alpha$ and $\beta$, defined as

$$\alpha = \left.\frac{T_0}{R_0}\frac{\partial R}{\partial T}\right|_{I_0} \qquad \beta = \left.\frac{I_0}{R_0}\frac{\partial R}{\partial I}\right|_{T_0}$$

so that $R_{(T,I)}$, expanded about the equilibrium point, now reads

$$R_{(T,I)} \approx R_0 + \alpha\frac{R_0}{T_0}\delta T + \beta\frac{R_0}{I_0}\delta I \tag{6.8}$$

Continuing the linearization, the heat flow into the bath can be written as $P_{bath} \approx P_{bath0} + G\delta T$, where $G = 5KT^4$. Similarly, the Joule heating power becomes

$$P_{Joule} = I^2R \approx P_{Joule0} + 2I_0R_0\delta I + \alpha\frac{P_{Joule0}}{T_0}\delta T + \beta\frac{P_{Joule0}}{I_0}\delta I \tag{6.9}$$

It will serve us later to group together the dimensionless $\frac{P_{bath}\alpha}{GT}$ as the 'low frequency loop gain parameter', $\mathcal{L}$, which tells us the relative magnitude of Joule heating power vs. bath cooling power. In other words, given a small heat input, $\mathcal{L}$ tells us if the subsequent cooling is largely the effect of an increased resistance and lowered Joule heating (large $\mathcal{L}$) or is largely the effect of conduction to the bath (small $\mathcal{L}$).

Plugging in the linearizations, Equations 6.6 and 6.5 become

$$\begin{aligned} \tfrac{d}{dt}\delta I &= \left[\tfrac{-R_L-R_0(1+\beta)}{L}\right]\delta I + \left[\tfrac{-\mathcal{L}G}{I_0 L}\right]\delta T + \left[\tfrac{1}{L}\right]\delta V \\ \tfrac{d}{dt}\delta T &= \left[\tfrac{I_0 R_0(2+\beta)}{C}\right]\delta I + \left[\tfrac{G(\mathcal{L}-1)}{C}\right]\delta T + \left[\tfrac{1}{C}\right]\delta P_{input} \end{aligned} \tag{6.10}$$

Several time scales are evident here. The simple thermal time scale $\tau = C/G$ appears, telling us the rate of temperature decrease, assuming that the other powers (the input power and the electrothermal feedback) are zero. Another timescale, the electrical timescale (basically, $L/R$), can be seen by setting $\mathcal{L}$ to zero, where only electrical properties matter, and the current decays with time constant

$$\tau_{el} = \frac{L}{R_L + R_0(1+\beta)} \tag{6.11}$$

Another decay constant can be seen in the limit of perfect current biasing ($\delta I \to 0$), where the temperature decays with the time constant

$$\tau_I = \frac{C}{G}\frac{1}{1-\mathcal{L}} = \frac{\tau}{1-\mathcal{L}} \tag{6.12}$$

Notice that $\tau_I$ can easily be negative ($\mathcal{L} > 1$), indicative of a non-stable system.

Rewriting the system of equations once more, this time in a matrix format,

$$\frac{d}{dt}\begin{pmatrix} \delta I \\ \delta T \end{pmatrix} = \begin{pmatrix} -\tau_{el}^{-1} & -\frac{\mathcal{L}G}{I_0 L} \\ \frac{I_0 R_0(2+\beta)}{C} & -\tau_I^{-1} \end{pmatrix}\begin{pmatrix} \delta I \\ \delta T \end{pmatrix} + \begin{pmatrix} \frac{\delta V}{L} \\ \frac{\delta P_{input}}{C} \end{pmatrix} \tag{6.13}$$

Note that Equation 6.13 becomes homogeneous when $\delta P_{input}$ and $\delta V$ are zero. TES dynamics, as described even by this clearly simplified small-signal limit in Equation 6.13, are rich with complexity. Instead of duplicating derivations easily found in the literature, we here simply quote some results directly useful to the CDMS circumstance, without derivation.

First, given a $\delta$-function $P_{input}$, the electrical and thermal time constants combine to prescribe an electrothermal feedback time constant given by

$$\tau_{etf} = \frac{\tau}{\mathcal{L}+1} \tag{6.14}$$

In the limit of weak feedback (small $\mathcal{L}$), the electrothermal feedback time scale is just the thermal time constant $\tau$ ($=C/G$). In the limit of strong feedback (large $\mathcal{L}$), the electrothermal feedback time scale is fast, greatly boosting the bandwidth of the device.

Second, we state the primary beauty of a TES, its extreme energy resolution. The resolution, described using the full width half maximum, can be written as

$$\Delta E_{FWHM} = 2\sqrt{2ln2}\sqrt{4k_BT_0^2C\frac{1}{\alpha}\sqrt{\frac{5}{2}}} = 2\sqrt{2ln2}\sqrt{4k_BT_0P_{Joule0}\tau_{etf}\sqrt{\frac{5}{2}}} \tag{6.15}$$

This expression assumes a pulse that is shorter than the sensor time constant ($\tau_{etf}$). This is not the case in CDMS, where the event energy reaches the TES over time scales in the hundreds of microseconds, longer than the electrothermal feedback time scale, which is typically on the order of tens of microseconds or less. In the CDMS situation, where $\tau_{input} > \tau_{etf}$, Equation 6.15 becomes

$$\Delta E_{FWHM} = 2\sqrt{2ln2}\sqrt{4k_BT_0P_{Joule0}\tau_{input}\sqrt{\frac{5}{2}}} \tag{6.16}$$

We see that we pay a resolution price for extending the signal over a longer timescale (intuitively, we have had more time for more noise to enter into the measurement). Matt Pyle points out in his recent thesis that by reducing the $T_C$, we can drastically increase $\tau_{etf}$, better matching the sensor time scale to the signal time scale, and that for $\tau_{input} > \tau_{etf}$, resolution scales as $\Delta E_{FWHM} \propto T_C^3$. This is a promising direction for future work, as can be seen from the extrapolation in Figure 6-10.

One significant caveat should be discussed before we wrap up our discussion. We have assumed that the TES electron system can be described by some temperature

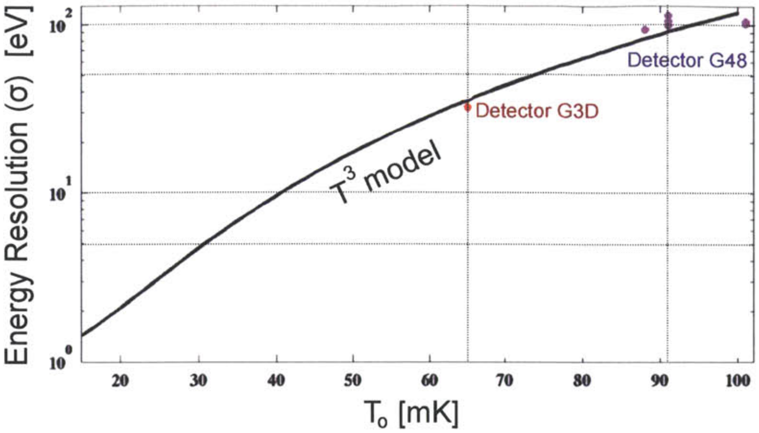

Figure 6-10: Resolution scaling with operating temperature. From the thesis of M. Pyle.

$T_0$, but in fact some amount of temperature difference exists within the TES, the magnitude of which depends largely on the thermal conductivity within the TES. . If TES sensors are longer than some maximum length

$$\ell_{max} = \sqrt{\frac{\pi^2 L_{Lor}}{\alpha \Sigma T_C^3 \rho_n}} \tag{6.17}$$

where $L_{Lor} = 25$ nW$\Omega$/K$^2$ relates electrical conductivity to (electrical) thermal conductivity (in the Wiedemann-Franz Law), $\Sigma$ is the tungsten electron-phonon coupling constant ($\sim$0.32 nW/$\mu$m$^3$K$^5$), and $\rho_n$ is the normal resistivity of W, then the thermal coupling from one end of the TES to the other is so weak as to result in phase separation. Within the TES, there will exist some boundary between normal and superconducting portions, and the situation is much more complicated than as described by all the discussions up to this point. Although the voltage biasing still supports the TES in a stable way, a heat input will increase the resistance by increasing the spacial extent of the normal portion boundary. Similarly, fluctuations in this phase boundary introduce a new and quite significant noise term.

## 6.6 Now we are ready

We have, in as brief a fashion as possible, summarized the main physical processes relevant to a CDMS detector. A recoiling electron or nucleus deposits energy in the target in some ratio of phonons and e-h pairs. The e-h pairs are drifted through the use of an electric field and this movement is sensed by surface electrodes. The phonons, including Luke phonons created by the drifting charges, decay in energy, lengthen in scattering length, and eventually propagate entirely ballistically. When these phonons encounter an Al surface, they create quasiparticle excitations of the superconducting state, which then diffuse (aided by an energy gradient) into transition edge sensors, where the energy is finally thermalized, and where even a tiny input energy creates a significant and measurable signal. Let us now start discussing the specifics of CDMS II, and its successor, SuperCDMS.

# Part III

# CDMS II

# Chapter 7

# Detectors for the CDMS II Experiment

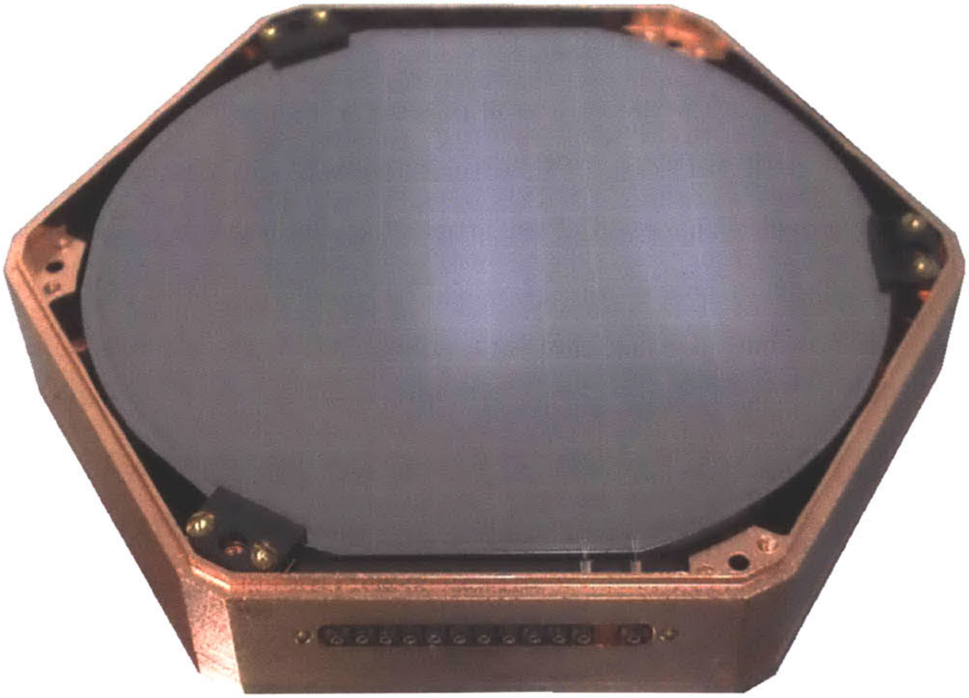

Figure 7-1: The charge-sensing surface of a mounted CDMS II detector (ZIP).

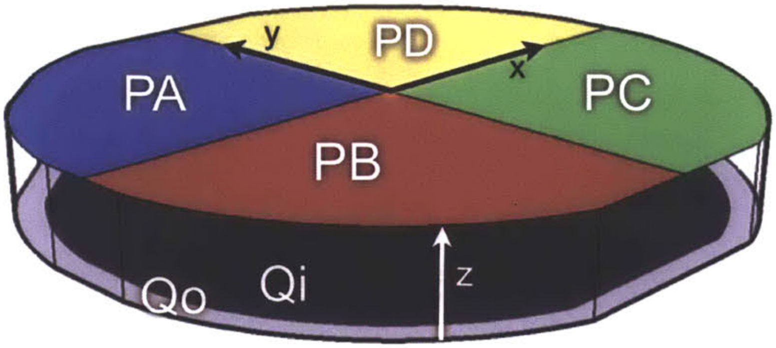

Figure 7-2: The general shape of a CDMS II crystal and channel divisions. Also indicated are the coordinate directions x, y, and z, referred to throughout this thesis.

## 7.1 Introduction

The design and testing of the detectors for CDMS II was before my time. We discuss the design here in order to understand the stepping-off point for designing the SuperCDMS detectors several chapters from now, and also because a basic understanding of the detector is of course necessary in order to understand the CDMS II analysis work.

The general layout of the detector is illustrated in Figure 7-2. Four phonon channels (named 'PA', 'PB', 'PC', and 'PD') divide the 'top' surface into four quadrants. Each phonon channel is made up of an array of QETs in parallel, as we will soon discuss. On the 'bottom' side of the detector, two charge electrodes ('Qi' and 'Qo' for 'inner' and 'outer') sense drifting charges and help veto high-radius events. The top surface is at 0 Volts, the bottom surface at +3 Volts, creating a drift field of 3 V/cm.

The crystal thickness is approximately 1 cm (although polishing and sometimes re-polishing led to some variation), and the crystal radius is 3.81 cm (1.50 inches). The cylindrical boundaries are marked by several 'flats' (the largest are at ±y) for ease of handing during the fabrication process, and to indicate the crystal lattice orientation. All crystals were oriented such that [100] pointed in the z direction (the easiest orientation to grow the huge boules from which these substrates were cut).

Each detector is held by a hexagonal copper housing, which holds the detector and a digital interface board ('DIB') to which the detector channels are wirebonded

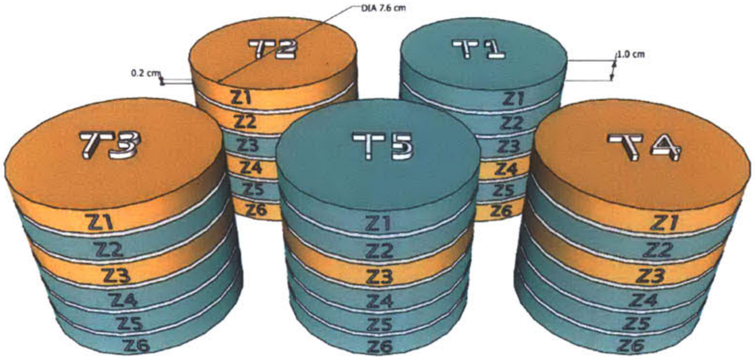

Figure 7-3: The arrangement of Si (orange) and Ge (green) detectors into five towers of six detectors each. Courtesy of Matt Fritts.

and from which the channels are connected to readout boards some distance away at higher temperature stages.

Both Si and Ge substrates were used, and the complete CDMS II array consisted of five towers of six detectors each, as illustrated in Figure 7-3. The detectors are referred to by their position, for example, the 3rd detector from the top of tower 5 is 'T5Z3'.

What is that 'Z'? CDMS detectors are often called 'ZIPs' for Z-sensitive Ionization and Phonon detectors, emphasizing their ability to determine an event's position in Z. Various versions of ZIPs have been run over the years, the CDMS II ZIPs are sometimes referred to as 'oZIPs'.

## 7.2 Charge

The Qi and Qo electrodes were attached to FET amplifier circuits, shown schematically in Figure 7-4. The noise expected from this circuit (and additional components at room temperature) is shown (blue curve) in Figure 7-5, along with the measured noise for a particular representative detector.

For a given charge channel, pulses are all of identical shape, prescribed by the L/R time constant of the readout circuit (much slower than the charge propagation time). In frequency space, these pulses have most of their power below ~5 kHz (the distribution in frequency space matches that of Johnson noise in Figure 7-5).

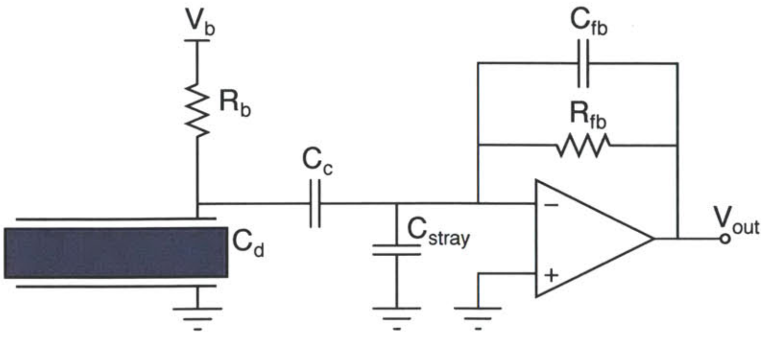

Figure 7-4: Simplified schematic of the CDMS charge amplifier. Values for the particular components are Rb = $R_{fb}$ = 40 MΩ, $C_{fb}$ = 1 pF, $C_{stray} \approx$ 75 pF, and $C_c$ = 300 pF. The detector capacitance $C_d$ is 93 pF for channel Qi and 36 pF for Qo.

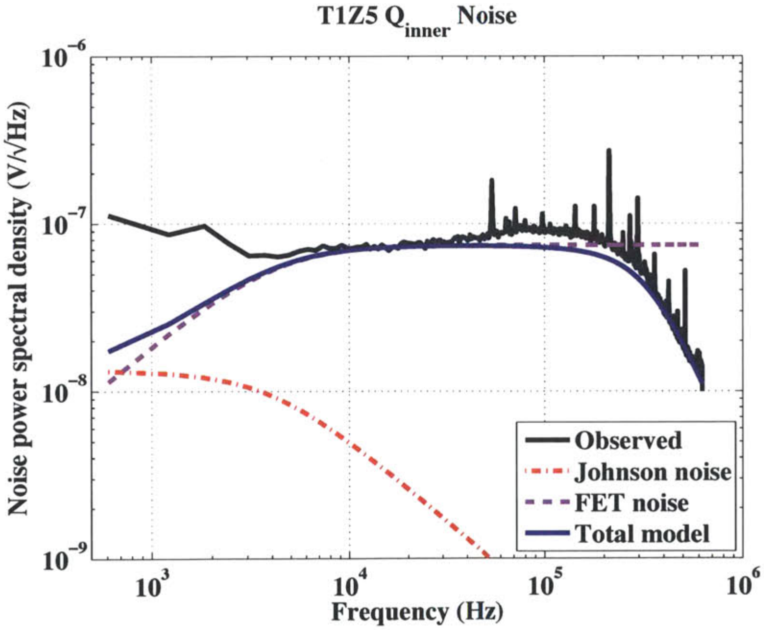

Figure 7-5: Representative charge noise for the Qi channel of detector T1Z5 during r123. The FET noise amplitude has been tuned to 0.44 nV/$\sqrt{Hz}$ to fit the data. Notice that the model fails significantly at low frequencies ($\lesssim$5 kHz). Courtesy of Jeff Filippini.

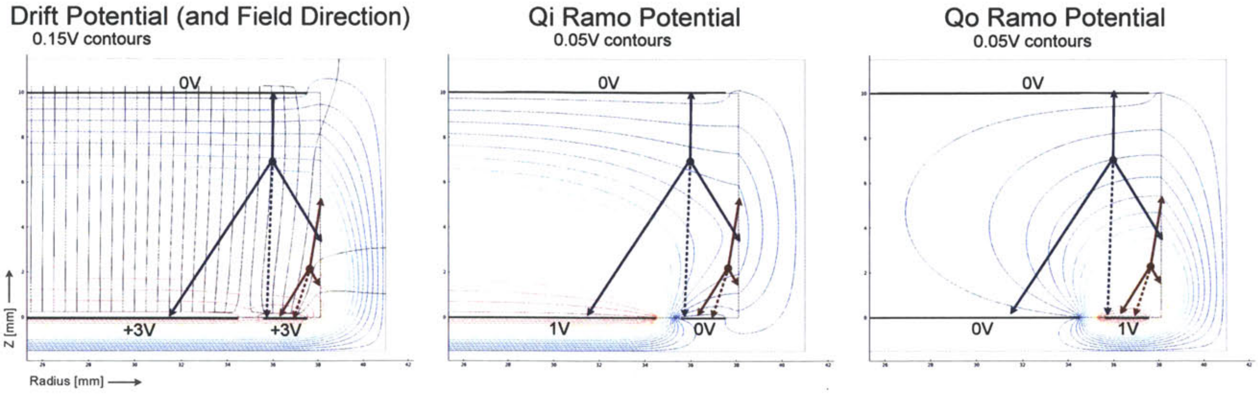

Figure 7-6: The drift potential, the Qi Ramo potential, and the Qo Ramo potential, zoomed in to the outermost ~cm of radius. Two events are shown, an 'ear' event in red, and a 'funnel' event in blue. Electrons propagate down, holes propagate up. The dashed lines represent the path of electrons in Si, where there electrons propagate in the [100] direction. For further discussion, refer to the text.

Unfortunately, there is also large (and unexplained) noise plateau in this same low frequency range. This excess low frequency charge noise is presumably part of the readout electronics design, and has been a persistent issue, significantly degrading our charge energy resolution.

There are two complications to the measured charge amplitudes. First, the drift field and the gradient of an individual electrode's Ramo potential can point in different (even opposite) directions, leading to a negative signal in the electrode. Second, the oblique propagation in Ge means that electrons follow neither the drift potential nor the ramo potential, but some third direction prescribed by the [100] orientation of the crystal lattice. This is illustrated in Figure 7-6.

The most common type of event is full charge signal on Qi and no signal on Qo. As you move to higher radius, the charge signal (the amount of drift work done by the electrode) is handed off to Qo. At high radius and high Z (blue event in Figure 7-6), there is significant electron trapping on the sidewall, leading to a reduced signal in both electrodes. Within the collaboration, we call these events 'funnel' events. At very high radius and low Z, the electrons propagate in the correct direction when viewed from the Qo electrode's potential, but in the wrong direction when viewed from the Qi electrode's potential. This means the electrons produced a negative Qi signal (the work done by Qi on the electrons is negative). If there is significant hole trapping on the sidewall, then the total Qi signal is negative (and the Qo signal is suppressed). We call such events 'ear' events.

Looking at Figure 7-7, we see in the upper right the Qo signal vs. the Qi signal for some $^{133}$Ba calibration data (*i.e.*, all events would be of yield 1 if the charge collection were complete). Ignore the mismatches axes units. To the right, these inner and outer charge energies have been normalized by the measured phonon energy, which collapses most events onto the straight line of complete charge collection (Qi+Q0=1). Two populations become clear in this plot, however, and the correspond to the two cases illustrated in Figure 7-6. Below, we have performed a simple charge propagation monte carlo simulation (the simulations have become much better since this plot was made) to better understand where in the crystal these ear and funnel events originate. Note that the creation of a funnel population is dependent on significantly oblique electron propagation, which produces significant electron sidewall trapping. For this reason, the funnel population only appears in Ge detectors; Si detectors exhibit an ear but no funnel.

The two channels are (weakly) capacitively coupled, which means that a large signal in one channel will create a small response in the other (having nothing to do with charges propagating and inducing currents). This effect can easily be corrected for, though, one just needs to square off the plot shown in the upper right of of Figure 7-7, so that the pure Qi and pure Qo populations lay along the axes.

That is all there is to the CDMS II charge measurement. One final remark: looking at Figure 7-6, one can see that the effect of the electron trapping would have been much reduced if we had biased the electrodes negatively, so that the holes (which simply follow the field lines) were attracted to the electrodes, instead of the electrons (which have a high probability of drifting into the sidewall and being trapped). This is fundamentally from the asymmetry of the ramo potential: much of the ramo potential difference occurs close to the electrode. Far from the electrode, if a charge traps or doesn't trap, the signal is affected very little. In fact, this negative (hole-attracting) bias state was the intention all along, and how the detectors were run at test facilities.

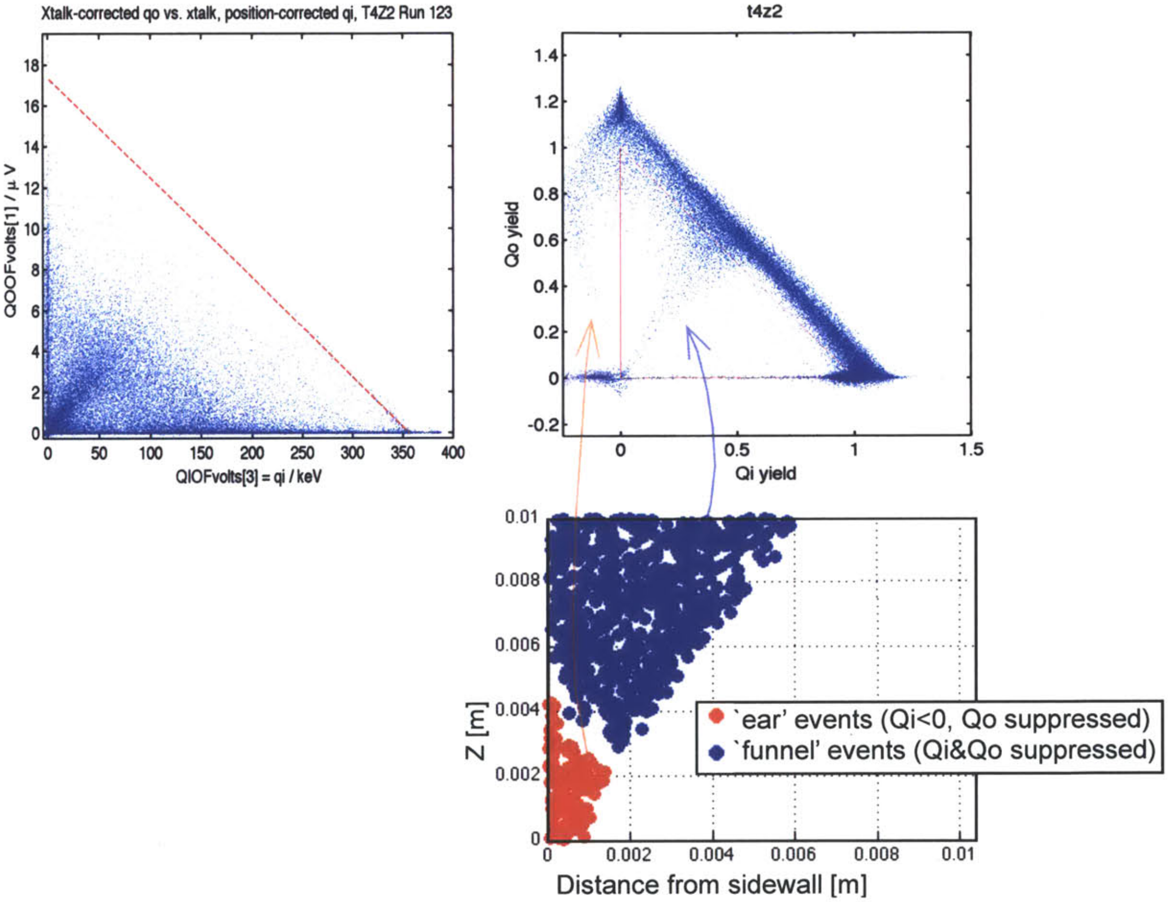

Figure 7-7: Charge partition plot for Ba133 calibration data of T4Z2 during r123. The expected position of the 356 keV line is highlighted in red. This same charge information, normalized by phonon energy, is displayed to the right, where various features indicative of specific physical regions are apparent. A monte carlo simulation was performed, and these specific physical regions were mapped out below. For further discussion, refer to the text.

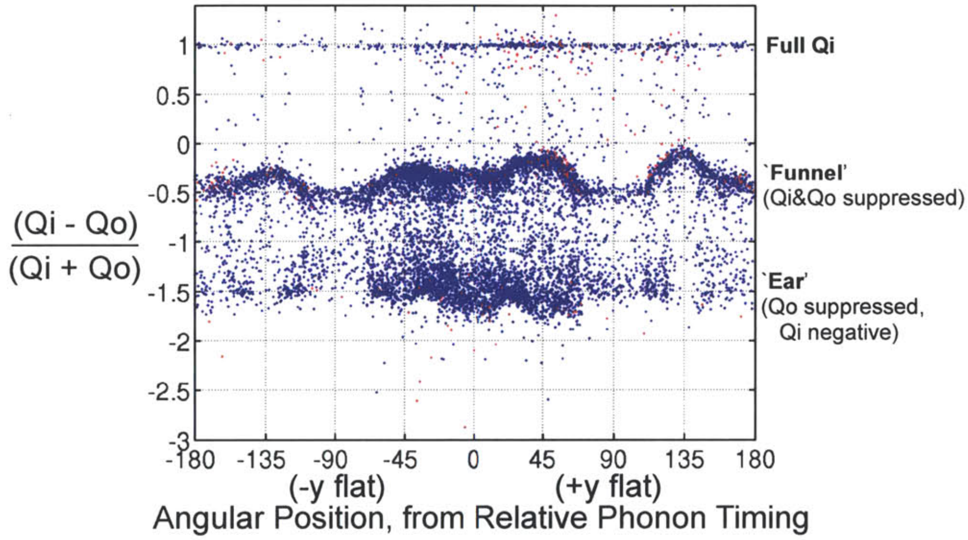

Figure 7-8: Because the funnel is a result of the oblique propagation into the sidewall, the angle of the sidewall (with respect to the crystal lattice) varies the signal seen as charges drift in the Ramo potentials. Effectively, the angle formed by the electrons in Figure 7-6 is a minimum when the sidewall and the lattice are aligned (-180, 90, 0, 90), and a maximum when the sidewall and the lattice are 45 degrees offset. The effect of the (lattice-aligned) major flats at ±90 degrees is clearly seen. Data selection for this plot of T1Z5 $^{133}$Ba data was simply quality cuts and that ionization yield was between 0.3 and 0.75. Plot courtesy of Walter Ogburn.

For multiple years, the big mystery was 'what is this funnel structure, and why do we see it at Soudan but not the test facilities?'. If the bias state had been correct during the CDMS II exposure, then we would have had noticeably higher fiducial volume fractions (a small improvement $\delta r$ at high radius adds as $r^2 + \delta r$) and greater exposure. This is just the first of several errors we'll come across... hindsight is 20-20!

## 7.3 Phonons

The design of the array of QETs in such a detector has several design drivers. Here we list the most practically relevant:

**Low Heat Capacity** Intuitively, if we put the same amount of heat into a smaller piece of W, it will have a larger affect on the resistance, and thus make for a better signal-to-noise. Additionaly, the $\tau_{etf}$ time scale lengthens with low heat capacity, bringing the sensor time scale more in line with the (comparatively slow) phonon arrival time scale and further increasing the signal-to-noise of the device (eliminating signal and noise at high frequencies where we have no signal). Heat capacity scales linearly with the W volume. More precisely, it scales with the volume of electrons in the W sample that are free to accept some small energy. For this reason, superconducting W does not count towards this heat capacity; W heat capacity depends only on the non-superconducting volume, and the boundary between superconducting and 'transition' W lies somewhere in the thermal gradient between the cold Al-W bilayer and the biased TES. In practical terms, the design goal here is then to minimize the volume of W between the TES and the Al fin.

**Impedance Matching** of the TES array to the the readout circuit (the SQUID circuit). Given the particular SQUIDs in the CDMS electronics, this dictates a resistance *when biased in the transition* of ~200 m$\Omega$. We typically bias at about 1/3 of the way up the transition edge (for the most linear response and so as to avoid saturation), so the normal resistance of the array can be about 3×

higher, or ~600 m$\Omega$. Again, this is specific to the SQUID circuit and nothing more general. Also, note that there is some small amount of wiggle room here since we can vary the resistance simply by varying the bias point (realizing that $\alpha$ will decrease if we push that point too far down the transition).

**Maximize the resistance of each TES** so that, given a specific array resistance of TESs in parallel, that array can contain a maximum number of TESs. The more TESs are in the array, the easier it is to well cover the crystal surface.

**The TES phase separation length scale** $\ell_{max}$ should not be exceeded.

**The geometry of the Al-W overlap** should maximize quasiparticle transmission.

**The Al fins should be large** so as to maximize phonon absorption in order to quickly absorb the phonons while they are most information-rich, before they randomize through multiple surface interactions.

**The Al fins should be short** (*i.e.*, no Al should be much farther than $\ell_{trapAl}$ from the Al-W overlap) in order to not absorb phonon energy somewhere where it will not be measured. The ratio between 'active' and 'passive' metal coverage directly feeds into the energy resolution of the device.

**The array inductance should be low** because it sets an L/R electrical timescale, which should be fast enough to capture all the timing information that we are interested in. Typically, L/R is much shorter than the rising edge time scale of the phonon pulses, so this design constraint is typically not a challenge. Note that typically L/R$< \tau_{etf}$, so L/R sets the limit primarily on rising edge timing information and $\tau_{etf}$ sets a limit primarily on falling edge timing information.

**The design must be robust** to small fabrication flaws. This is actually the most important of all the design motivators: of course, the design absolutely *must* be easy to fabricate and fabricate well. Practically, this sets constraints on the smallest feature size, and the minimum spacings between features.

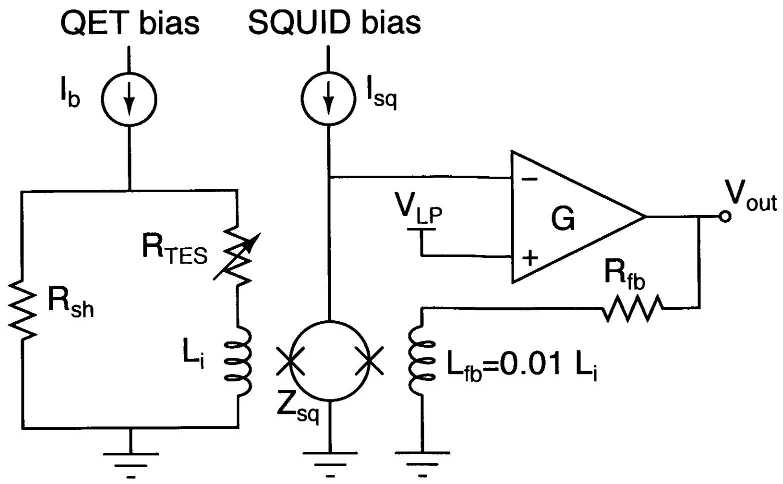

Figure 7-9: Simplified circuit diagram of a CDMS II phonon channel. The TES biasing circuit on the left serves to voltage-bias the TES array, keeping the array in its transition. The readout circuit on the right amplifies the TES current signal through the use of an inductively-coupled SQUID. Specific values for CDMS II are $R_{TES} \approx 200\text{m}\Omega$, $R_{sh} = 25\text{m}\Omega$, $R_{fb} = 1200\Omega$, and $L_i = 250\text{nH}(=100L_{fb})$.

We will now continue to lay out these design drivers, both their motivation and how to achieve them. To aid in this discussion, we show in Figure 7-10 a series of steps in the evolution of the CDMS II QET array, showing how these design drivers have been implemented.

The very first QET designed by the collaboration is shown on the left, and illustrates many mistakes that were avoided in later designs. First and foremost, the TES was of extreme length (800 $\mu$m), and therefor severely phase-separated. Plugging the W normal resistivity measured in CDMS films ($\rho_n = 1.2 \times 10^{-}6\Omega m$), a likely value for the transition steepness ($\alpha = 300$), the measured value of the electron-phonon coupling constant in W ($\Sigma = 3.2 \times 10^8 Wm^{-3}K^{-5}$), and a typical $T_C$ of 80 mK, we arrive at a phase separation length scale of

$$\begin{aligned} \ell_{max} &\approx \sqrt{\frac{\pi^2 L_{Lor}}{\alpha \Sigma T_C^3 \rho_n}} \\ &\approx \sqrt{\frac{\pi^2(2.44\times10^{-8}W\Omega K^{-2})}{300(3.2\times10^{8}Wm^{-3}K^{-5})(0.080K)^3(1.2\times10^{-7}\Omega m)}} \\ &\approx 200\mu m \end{aligned} \tag{7.1}$$

One can see that later designs kept this general scale in mind.

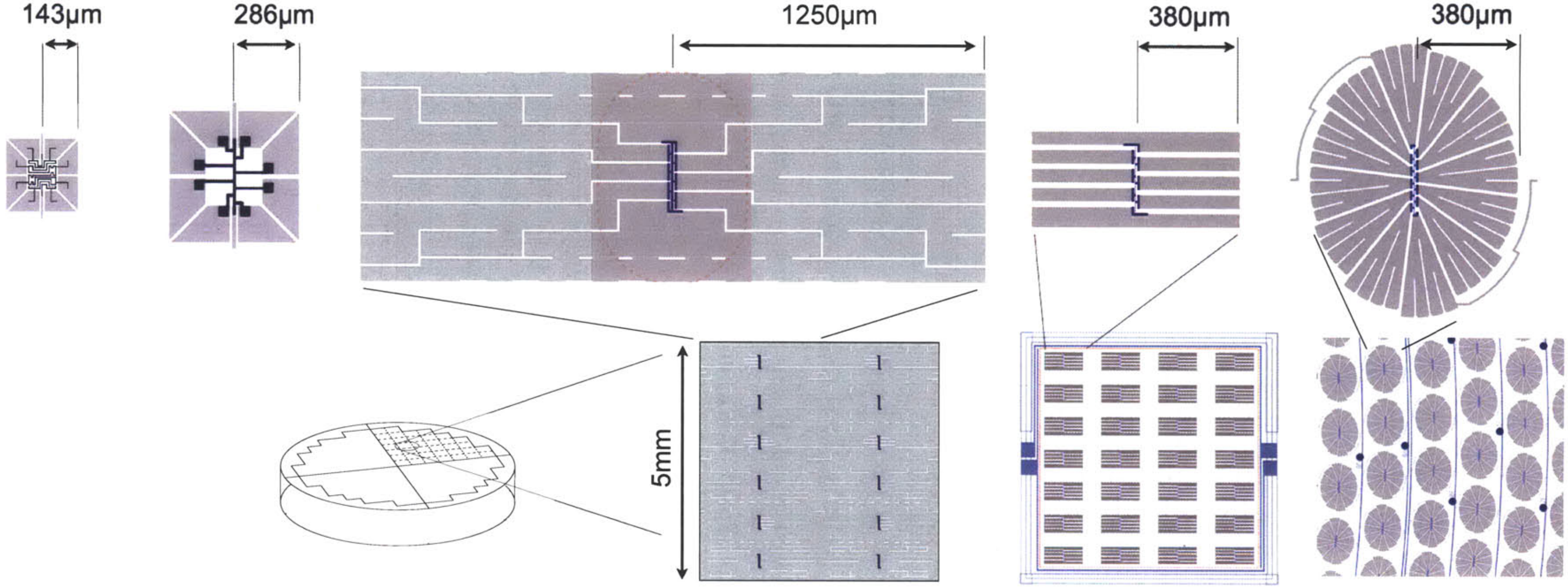

Figure 7-10: Five versions of a CDMS-II-style QET layout array, tracing the evolution of the concept. Grey is Al, dark blue or black is W. A distance from the Al-W overlap of $1.5\ell_{trapAl}$ is shown in pink, assuming the Al quality matches the 1998 banana device, and $\ell_{trapAl}$ scales with the square root of the thickness. For these five designs, the Al film thickness was [60,100,150,300,300]nm, corresponding to $1.5\ell_{trapAl}$ distances of [170,220,170,380,380]$\mu$m, respectively. All QETs (upper row) are to the same relative scale, as are the three views of QET arrays (lower row). From left to right are illustrated the first QET design (Stanford runs 9-15), a later QET design (Stanford runs 31-34), a preliminary oZIP design, the final oZIP design (as used in CDMS II), and lastly a further evolution of the design ('mZIP'). Further discussion in the text. Significantly drawn from the theses of M.Pyle and T.Saab.

Naively, one would think that the thermal phase separation length scale would depend entirely on how strong the thermal link is between one end of the TES and the other; in other words, a thicker fatter TES should have a shorter $\ell_{max}$. In reality, however, the stronger thermal coupling is offset by the larger current biasing the larger TES. The phase separation length is a material property rather than geometric property, and which can be increased only through the following strategies:

1. Reducing $\alpha$. Unfortunately, because this strategy is essentially just lowering the sensitivity of the TES to changes in temperature, reducing $\alpha$ also pay a significant penalty in threshold and resolution. In fact, I should mention here that currently our $\alpha$ values are very poorly known. They could be anywhere from $\sim 125$ to $\sim 500$, depending on what method one uses to measure them, which puts a similarly large uncertainty in our predicted $\ell_{max}$. Depressingly, our collaboration has never once made a non-phase-separated device, and we have *always* had reduced sensitivity as a result.

2. Reducing the $T_C$. Because $\ell_{max} \propto T_C^{-3/2}$, and because thermal Johnson noise goes down with $T_C$ also, reducing $T_C$ is clearly a powerful strategy. We are limited here in the design by the operating temperature of the fridge. In order to have low thermal fluctuation noise between the bath and the TES, we want the W electron temperature to be significantly above the bath (W lattice) temperature, which is typically ~40 mK at Soudan.

Referring again to the first-ever QET, we can see a second major flaw. The Al-W overlap region was efficient at *trapping* quasiparticles because no Al was very far from the Al-W overlap, but it was extremely inefficient at *transmitting* those quasiparticles from the Al-W overlap to the TES. The thin linear overlap presumably has a quite variable gap energy, which could easily trap quasiparticles at local minima. Additionally, any Al-W bilayer will have a significantly shorter $\ell_{trap}$ than a simple Al. If you want to diffuse a quasiparticle a certain distance, it is most efficient to have as much of that distance be pure Al as possible.

Skipping over some intermediate designs, the second QET shown in Figure 7-10 is much improved. The TES is now short, much less phase-separated. The Al-W overlap is shorter, too, for less trapping. Here, the largest clear flaw is the extreme length of the W connection between the Al-W overlap and the TES. It is true that this W contributes needlessly to an increased heat capacity, but the main problem is just the wasted space near the TES. Matt Pyle has called the area surrounding the TES "beachfront property", because this is the most valuable area on the detector face, area where any Al present would be maximally efficient. In this second design, the beachfront property is left entirely empty.

The third (huge) QET has fully developed this beachfront property; the inner edge of the Al is pushed as close to the TES as possible, and the W connector between the overlap and the TES is as short as possible, both given fabrication constraints. This third QET has Al fins so large that an array of such QETs is 'close-packed', resulting in a maximized Al surface coverage of near 100%. Unfortunately, given the finite $\ell_{trapAl}$ (here, the Al thickness is 150nm, so $\ell_{trapAl} \approx 180\mu$m), only a fraction of this Al is *active* Al, and a great majority of the Al area is passive. A bare metal-free surface would in fact result in an array with much superior energy resolution, because a phonon incident on this bare surface would reflect and then later be absorbed (on active Al) on its second (or third, or fourth...) interaction with a surface.

Assuming we know $\ell_{trapAl}$, what is the optimal length for an Al fin? In Figure 7-10, the pink dashed lines represent a distance of $1.5\ell_{trapAl}$, because this was the rough rule of thumb that was understood for many years. Note that the $1.5\ell_{trapAl}$ distance is seen to constantly increase in the evolution of CDMS II-style QETs; here we are assuming $\ell_{trapAl}$ is proportional to the film thickness (though this hasn't been proven in the CDMS case). Matt Pyle has a quantitative argument[92] for an optimal fin length of $1.1\ell_{trapAl}$, assuming energy resolution is the figure of merit. The energy-resolution optimization serves as a minimum-length case, then. As we increase the length from there, we very gradually lose energy resolution while linearly increasing the absorption rate, and thus the timing and position information content. How we optimize this fin length, then, depends on what information we are trying to extract from the phonon signal. If energy is all-important, then the fin length should be perhaps $1.1\ell_{trapAl}$, if instead we want to capture as many phonons on their first surface interaction as possible, then the fin length should be close-packed fin length, and if we want both energy *and* other information, then the length should be intermediate. Thus, we arrive back at something like the $1.5\ell_{trapAl}$ rule of thumb, as a way to balance Al efficiency and Al coverage.

This optimization between coverage and efficiency is the improvement between the third QET design and the fourth QET design; in the fourth design the Al coverage is much decreased ($\sim 100\% \rightarrow \sim 35\%$) but the active-to-passive ratio is vastly increased ($\sim 3:10 \rightarrow \sim 9:10$). An incident phonon only has a 35% chance of finding Al on its first first surface interaction, but this phonon's energy has a *high* probability of *eventually* reaching the TES (or, at least, higher than the earlier design). The 4th design in Figure 7-10 was that used for the CDMS II experiment, the topic of the next several chapters.

While CDMS II collected data, design work continued on making the CDMS II idea better and better. This strategy reached its apex with the so-called 'mZIP' illustrated at the right of Figure 7-10. The 'm' stands for 'maximized', because the Al fins were finally placed such that each TES had a maximum amount of active Al coverage. Note that, in CDMS II, only about one third of the area within the $1.5\ell_{trapAl}$ distance was filled in with Al. Again taking advantage of our 20-20 hindsight, the entire CDMS II experiment could have had better energy resolution, 3× better timing information, and all-around significantly improved scientific reach if we (continuing the beachfront property philosophy) squeezed as *much* aluminum as *close* to the TES as possible, as the mZIP design does. There is, in fact, still a small design flaw even in the mZIP QET. Because the width of these fins decreases towards the TES, the quasiparticle diffusion *towards the TES* is suppressed by that width, and the fin diffusion efficiency is somewhat suppressed, implying a fin length somewhat shorter than the $1.5\ell_{trapAl}$ rule of thumb.

We have just spent quite some time discussing the fin length. Any such discussion should be taken with the huge grain of salt, however, that we only measured

$\ell_{trapAl}$ once, and it presumably varies significantly from depending on subtle details of the deposition (sputtering) process. So, we shouldn't think too hard about this optimization until we have a better idea of what $\ell_{trapAl}$ actually is.

One of the design drivers mentioned above is that the array resistance has a lower bound set by the SQUID impedance. With this *lower* limit on the array's (biased) resistance of ~200 m$\Omega$, we in effect have created an *upper* limit on the number of TESs we can place in parallel in a single channel. In fact, it is this limit on the density of TESs on the detector face that forces us to think about Al fin length at all. Ideally, we would just have a low-impedance SQUID loop, a matched low-impedance QET array composed of a huge number of QETs, and then we could have extremely short (zero-trapping) Al fins that still provide near complete Al coverage. In other words, if the SQUID impedance were low enough, we would have significantly increased energy resolution *and* timing information, without compromise.

But, this is not the situation we find ourselves in. The SQUID loop should be lower-impedance in the future (at SNOLAB), but for now the only way to increase the QET density is to increase the resistance of each TES. Unfortunately, we are up against a wall here, too. The TES dimensions are bounded in all three dimensions. In the direction of current flow, it is the phase separation length (~200$\mu$m) that sets the length. The film thickness is set by fabrication (limited by the shortest possible deposition in the Balzers) to be no thinner than 40nm. This film thickness results in a sheet resistance of ~ 3.3$\Omega$/sq, by the way. The last dimension, width, is similarly constrained by fabrication capabilities. In CDMS II and before, optical stepper photolithography was used to pattern the W, resulting a smallest feature size of ~ 1$\mu$m. After CDMS II, the stepper strategy was discarded for a full-wafer strategy using with lower patterning resolution, resulting in a smallest feature size of ~ 2.4$\mu$m. This is why there are fewer QETs per area between CDMS II (the 4th design in Figure 7-10) and the mZIP (5th design). The TESs have become wider, thus of lower resistance, and so fewer can be placed per channel.

In general then, we have no freedom at all in the dimensions of the TES. We push the design in all three directions towards higher resistance, till we meet one of the

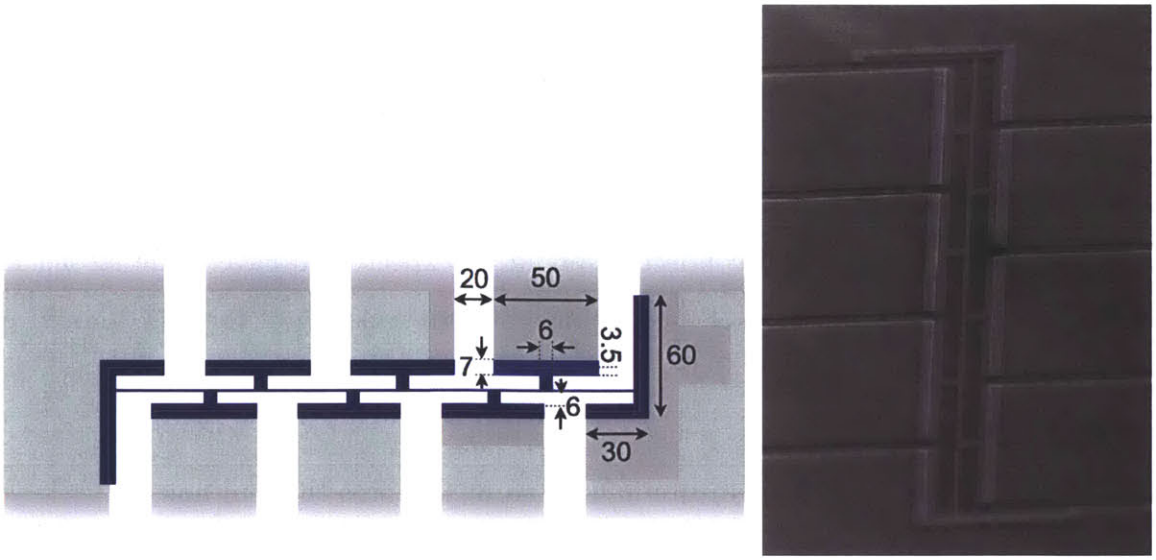

Figure 7-11: Diagram and SEM of the CDMS II QET, to illustrate the interface between the TES and the fins. Dimensions are in microns.

afor-mentioned upper or lower limits. we push TES array designs against all three of these constraints, in order to maximize the resistance per TES, thus maximizing the number of TESs per channel, and thus maximizing the instrumented (active Al) surface area. We push the fabrication towards thin films and narrow patterning, and we tempt fate by hoping the $T_C$ and $\alpha$ of the film result in an $\ell_{max}$ slightly longer than the design length. In CDMS II, the TES is 40nm$\times$1$\mu$m$\times$250$\mu$m.

Figure 7-11 shows a zoomed in view of the CDMS II QET, where we can better see the interface between the Al fins and the TES. One can see the extremely narrow (1$\mu$m) TES is actually dwarfed by the tungsten of the fin overlap and the fin connector. The idea is that we want to minimize the (electrical) conductivity of the TES, but we want to maximize the (quasiparticle) conductivity of the overlap and connector. The overlap is the full width of the fin (50$\mu$m) in order to have the maximum length to the step down in gap energy (to boost the probability of quasiparticle finding that step). Once the quasiparticle is in the Al-W overlap, it is very easy to trap, so no portion of the overlap is very far from the next step down (only 3.5$\mu$m to the monolayer W with a lower gap energy). The length of the 'neck' of the connector (6$\mu$m) is as short as possible, set by fabrication constraints.

In other words, the fin connector sounds reasonably well designed to maximize quasiparticle diffusion. Remember, that the trapping length scale in both the Al-W overlap and the bare W are largely unknown, so a detailed optimization is impossible.

On the other hand, because we have made the fin connector large and strongly coupled to the TES, the fin connector should be expected to have some strong influence on the TES dynamics. For example, the heat capacity of the TES should be increased by that portion of the W which is close enough to the TES to still have some unpaired electrons (and hence some electron heat capacity). The electron system of the fin connector has a temperature influenced by its three thermal links:

1. As just stated, given the wide 'neck' (and lack of any material interface barrier) the connector is strongly thermally coupled to the TES.

2. The W and the Al fin are both metal, so presumably their electron systems are in thermal contact to some degree (though we learned from the banana device that there is some very significant interface barrier here).

3. The electron system of the fin connector and the bath (the lattice system of the fin connector) have a thermal link to the bath $g_b$ of magnitude $V\Sigma(T^5_{electron} - T^5_{lattice})$, where $V_{Wns}$ is the volume of the (non-superconducting) tungsten.

Importantly, the volume of the fin connector actually dominates the total W volume. This makes the phase separation issue much worse, a fact that was not fully understood in the CDMS II design process. Intuitively, a localized Joule heating in the TES has two pathways through which to flow: it can spread out within the TES through the thermal conductivity of the TES's electron system with itself $g_W$ or it can exit the electron system into the lattice system through the thermal conductivity of the electrons to the lattice (bath) $g_b$. The phase separation length scale is proportional to the ratio of these two thermal conductivities. If the heat goes to the lattice rather than spreading out within the TES, then the phase separation effect is amplified. Notice that both of the relevant conductivities are here expressed as material quantities of the tungsten (rather than total quantities for a given TES).

$g_W$ can be obtained from tungsten's resistivity, following the Wiedemann-Franz law $g_W = L_{Lor}T/\rho_n$, where $\rho_n$ is the normal state resistivity of tungsten ($1.2 \times 10^{-7}\Omega m$) and $L_{Lor}$ is the Lorenz constant ($\pi^2 k_B^2 3^{-1} e^{-2} = 25nW\Omega K^{-2}$). Assuming $T_e^5 >> T_b^5$, then $g_b$ can expressed near the bias point as $\Sigma T_e^{5-1}$, where the electron temperature $T_e$ is $T_C$ at the TES, and gradually lower with distance from the TES.

The phase separation length scale can be written as

$$\ell_{max} = \pi \left[\frac{g_W}{g_b(\mathcal{L}-1)}\right]^{\frac{1}{2}} = \pi \left[\frac{L_{Lor}}{\rho_n \Sigma T_e^3(\mathcal{L}-1)}\right]^{\frac{1}{2}} \approx \pi \left[\frac{L_{Lor}}{\rho_n \Sigma T_e^3 \alpha}\right]^{\frac{1}{2}} \tag{7.2}$$

assuming that we are in the strong electrothermal feedback regime ($\mathcal{L} >> 1$) and $T_b$ is significantly lower than $T_e$. Adjusting for the fins, $g_b \rightarrow g_b \frac{V_{TES}}{V_{Wns}}$ (where $V_{TES}$ is the volume of the TES and $V_{Wns}$ is the total total (non-superconducting) W volume (mostly in the fins, in CDMS II). Note that $g_W$ does not change, because this conductivity is only along the TES. Adjusting Equation 7.2, then, we have

$$\ell_{max} \rightarrow \ell_{max}\sqrt{\frac{V_{TES}}{V_{Wns}}} \tag{7.3}$$

Quickly running through the dimensions from Figure 7-11, we see that the TES is only ~1/8 of the non-overlap W volume. Some large fraction $f_{ns}$ of the fin connector is non-superconducting, and the baseline phase separation length at 80 mK was 200$\mu$m, so

$$\ell_{max} = [200\mu m]\sqrt{\frac{1}{8f_{ns}}} \tag{7.4}$$

Worst case ($f_{ns}$ ~1), the CDMS II fin connectors reduce the phase separation length at 80 mK to only ~70$\mu$m! Of course, if we realize that some amount of W with be superconducting from the proximity effect, and nearly all of the W will have a lower temperature than 80 mK, our $f_{ns}$ fudge factor should probably be significant, perhaps 1/2, pushing the phase separation length back up to ~140$\mu$m. In any case, this all goes to show that we were certainly phase separated in CDMS II, and we certainly need to fight more aggressively for more thermally uniform TESs in future designs.

## 7.4 QET Efficiency

As energy flows from the phonon system into quasiparticles in the fins, and then finally cascading into the TES, there are many inefficiencies. Phonons can be absorbed by passive metal, particularly metal on the electrode surface, where the energy will go unmeasured. If the phonon is absorbed in an active Al fin, the initial phonon-quasiparticle cascade only converts about 55% of the energy into quasiparticles. Similar losses occur in the phonon-quasiparticle cascades when quasiparticles diffuse to lower-gap regions, and drop from $2\Delta_{Al}$ to $2\Delta_{Al-W}$, and then again when they drop to $2\Delta_W$. Also significant are trapping losses in both the Al and the Al-W bilayer. All these efficiencies are independent and the final efficiency is the simple product. The overall efficiency, then, must be quite low.

Walter Ogburn used CDMS II phonon pulse amplitudes to estimate this efficiency. Tracing the total combined amplification of the room temperature electronics, the SQUID circuit, and finally the turns ratio in the inductive coupling, one can calculate the actual current flowing through the TES, which then can be converted into an input power into the TES. In the first 12 detectors fabricated for CDMS II, the best Ge detector had a total measured efficiency of 2.7%, and the best Si detector had a total measured efficiency of 3.7%. In retrospect, this difference between Ge and Si, which is still seen in current devices, is real, and is most likely due to the transmission efficiency of phonons into the Al (which is apparently higher for Si).

Efficiencies in the single digits may sound depressing, but let's think again of what we have done. The phonon energy from a ~250g object has been absorbed before it thermalizes, in a few hundred $\mu$s, and transferred to an array of *microscopic* calorimeters with a total calorimeter mass of $\sim 1\mu$g, approaching a concentration factor of one *billion*. Achieving this level of energy concentration is something to be amazed by, even with a 2.7% efficiency.

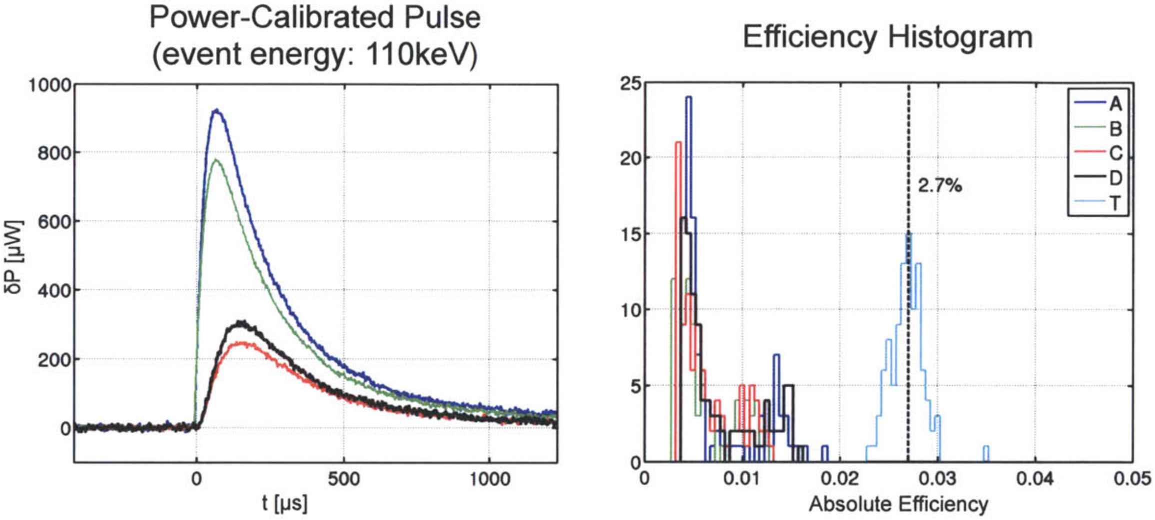

Figure 7-12: Plots by Walter Ogburn, showing the an event in all four phonon channels, after the absolute calibration has been applied (such that the units are thermal power into the TES). Taking the integral of such pulses to find the total energy input, one can compare the event energy (for this study, all events were ~100 keV), with the total heat in the TES, which turned out to be ~2.7 keV. Both plots here are for T2Z5, the most efficient Ge detector from the first two towers.

# Chapter 8

# CDMS II Analysis

This is the first of several chapters discussing CDMS II analysis (this chapter, then the main analysis, then the low energy analysis). Many theses have already been written that give thorough general discussions of these topics. My goal, rather than to simply rehash the same material, will be to highlight things I am proud of, things I think are interesting, or things that are relevant to understanding the situation of the later SuperCDMS detector (Part 3).

The basic questions for this chapter are

1. What information do we expect to be encoded in the pulse data of a CDMS II event (four phonon pulse and two charge pulses)?
2. How can we best extract useful reduced quantities from that pulse data?
3. How can we manipulate those reduced quantities to best get at the fundamental physical description of an event (position, energy, charge yield)?

## 8.1 CDMS II Pulses

We start with charge, the far simpler case. As previously mentioned, every charge pulse has exactly the same shape. We do not sample the channels at high enough of a frequency to watch the charges drift, we only see the total integral amount of induced current in the electrodes. For the two channels, then, we have only three real pieces of information: amplitude in Qi, amplitude in Qo, and the charge starttime (the two channels will have nearly identical starttimes, again from the low sampling rate).

The four phonon channels, on the other hand, contain a wealth of information. In fact, *the primary challenge of CDMS II analysis was disentangling the many overlapping types of information coded into the phonon pulses.*

Phonons propagate slowly enough (particularly in the initial diffusive stages of an event) that there should be some delay after the (for our purposes, instantaneous) charge start time before the phonon pulses start to rise out of the noise floor. Not only would we expect a charge-phonon delay, but we should expect similar arrival time delays from channel-to-channel.

Phonons are absorbed quickly enough at the surface that we expect the total amplitude of the phonon pulse in nearby ('primary') channels to be significantly larger than the phonon pulses in the other three channels. This does not depend on the phonons being absorbed in their first or second surface interaction, because of the diffusive nature of the phonons. Imagine an event occurring near a surface with partial Al coverage. If the phonons have a short mean free path, they will interact with the local surface many times, eventually being absorbed. Because of the quasidiffusive propagation, 'solid angle' arguments don't dictate relative phonon absorption by different channels, rather the question is largely 'how long is the mean free path (inversely, how frequent are surface interactions) when the phonons reach that channel?'

Luke phonons should be expected to make their presence known, also. Firstly, the production of Luke phonons will add extra energy to the phonon signal that we measure, so that

$$P_{tot} = P_{primary} + P_{Luke} = P_{primary} + [N_{eh}]eV_{bias} \tag{8.1}$$

with $V_{bias}$ in CDMS II being 1V (over a thickness of 1 cm). So, the amount of Luke phonon production depends on the charge yield of the event. Mapping from $P_{tot}$ (what we measure) to $P_{primary}$ (the energy of the recoil, what we are interested in knowning) requires that $N_{eh}$ (*i.e.*, the ionization yield) be measured.

Taking the example of an electron recoil in Ge, a 10 keV events will produce ~3,380 e-h pairs (with some variation predicted by the Fano statistics of the electronic cascade process). With $V_{bias} = 3V$, then

$$P_{totER} = 10\ keV + [3,380]e3V = 10\ keV + 10\ keV \tag{8.2}$$

and we see that an electron recoil (of in fact any energy) has $P_{primary} = P_{Luke}$. This balance is not a coincidence; the 3V bias state was actually chosen specifically such that the charge trapping would be minimized without allowing $P_{Luke}$ to dominate the phonon signal. Note that for a nuclear recoil,

$$P_{totNR} = 10\ keV + Y_{(10keV)}[3,380]e3V = 10\ keV + Y_{(10keV)}10\ keV \tag{8.3}$$

where $Y_{(10keV)}$ is the ionization yield for a nuclear recoil at 10 keV. Remember that $Y = 1$ for electron recoils (by definition).

Luke phonons have more effects than simply adding on to the total phonon signal. Because they are emitted along the entire length of the charge propagation path, presumably some Luke phonons will be created very close to the phonon sensors and will therefor 'jump start' the observed pulse. Luke phonons are also emitted at lower energies than primary phonons, and thus should be expected to reach the surfaces first for that reason as well.

We have already mentioned several effects that alter the rising edge slope: the closeness of the event to the channel (in x, y, and z), and the amount of luke phonons produced. Surface events are a special case in two ways: the phonon cascade is accelerated here, both by the surface of the Ge (because the lattice no longer exists, the phonon must change, and necessarily in a downconverting direction) and any metal layer present (where the phonon downconversion process can similarly proceed much faster). The $\alpha$-Si layer, too, possibly speeds up the downconversion. In short, if high-energy diffusive phonons are created near a surface, they become ballistic faster than in the bulk, so we expect surface events to have different phonon pulse shapes. (We also, because of charge trapping at the surface, would expect the Luke contribution and thus $P_{tot}$ to be somewhat suppressed.

The rising edge is determined by proximity to sensors, the Luke contribution, and surface effects. What determines the falling edge shapes? During the running of CDMS II, it was thought that TES dynamics (*i.e.*, $\tau_{eft}$) played a significant roll in shaping the falling edge of the pulse. In hindsight, though, it seems that the falling slope in CDMS II was defined largely by the phonon input power (like the rising edge). At late times (hundreds of $\mu s$), when the phonons are entirely ballistic, this input power is a simple exponential (the rate of phonons being absorbed by the Al is proportional to the total rate of phonon surface interactions, which is simply proportional to the total number of phonons). At earlier times, while the phonons are diffusive, the absorption rate is proportional not only to the number of phonons, but also inversely proportional to the square root of the mean free path of those phonons. So, we should expect an initially steep but mellowing exponential decay (while surface interactions per phonon are extremely frequent) stabilizing to some fixed slower exponential decay once the downconversion has reached the ballistic state.

As we have emphasized several times previously, the phonon system contains the most information at the start of an event, and information is gradually lost as the phonons diffuse, interact, and homogenize inside the crystal. The rising edge of the pulse, therefor, is where we will pull all the useful phonon timing information. The combined result of the position, yield, and surface-interaction effects is hard to predict, and we won't try... let's just define some quantities that describe pulses, and see what the observable results of all these physical effects turn out to be.

## 8.2 Optimal Filtering

We have already mentioned that we will need some measures of phonon timing, particularly on the rising edge, but let's take a step back for a moment and just ask 'what is the optimal way to measure the *amplitude* of a pulse'?

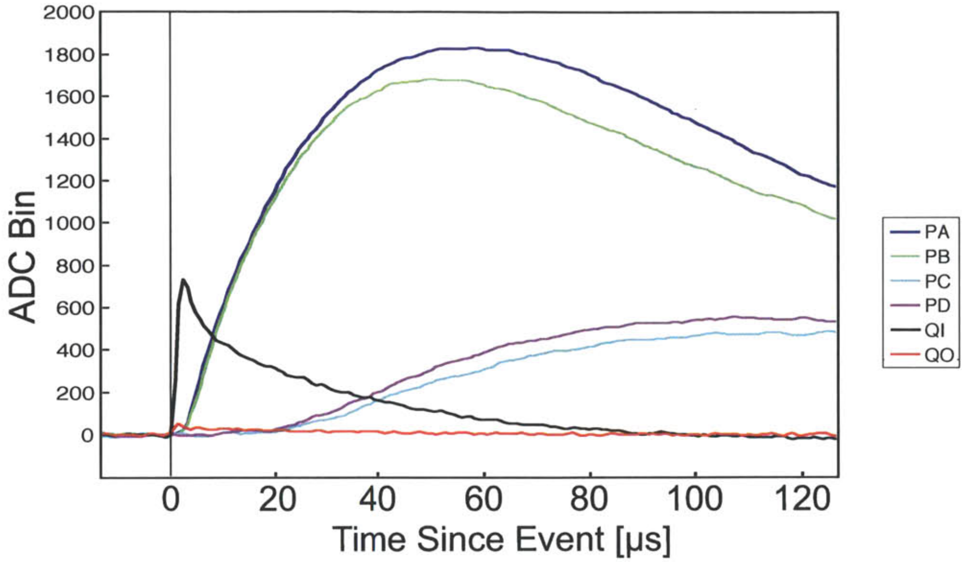

Figure 8-1: A typical CDMS II event. Notice the short timescale of the charge pulse rise (representing the drift time of the charge carriers), the significant delay before the primary phonon channel(s) start to rise (representing the initial slow quasidiffusion of phonons) and the quite significant delay and smaller amplitude of the non-primary phonon channels. Adapted from a plot by Jeff Filippini.

CDMS makes frequent use of so-called 'optimal filtering' to measure pulse amplitudes. The basic idea here is that pulse information inhabits certain frequencies, the noise that gets in the way of precise measurement inhabits certain frequencies, and we can maximize energy resolution by weighting frequencies according to the expected signal-to-noise of that frequency. Suffice it to say that CDMS II optimal filtering consists of

**Inputs**

- A raw trace containing both noise and signal.
- A template for the noise, which is typically just some measured noise traces. The frequencies are assumed to have no phase correlations, so this noise template can be simply the PSD of the noise.
- A template for the signal, which is typically just some averaged combination of pulses (averaged, in order to eliminate the noise). Note that here, the frequen-

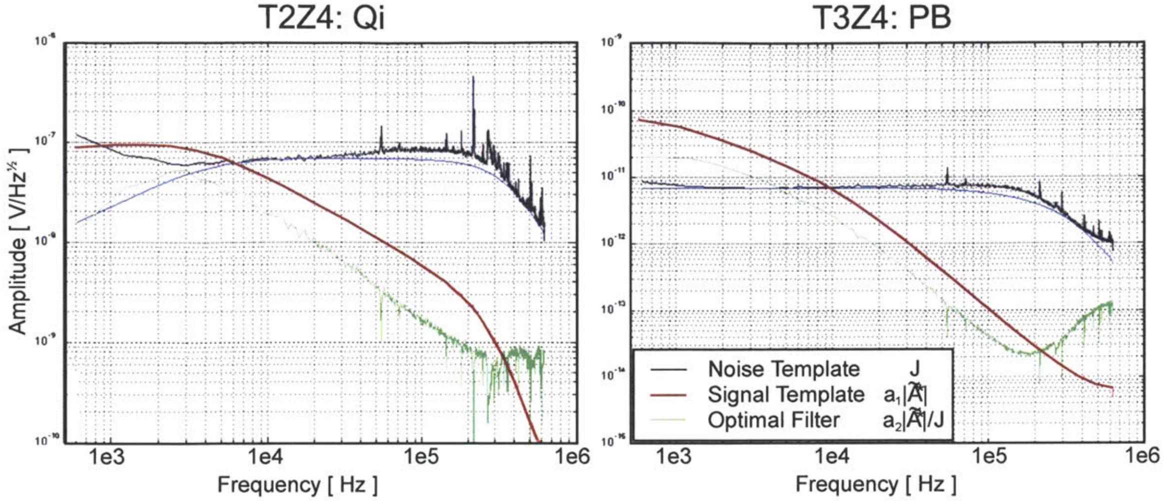

Figure 8-2: Signal, noise, and optimal filter spectra for a typical charge channel (left) and a typical phonon channel (right). Plotted by Jeff Filippini.

cies are importantly correlated, to give the pulse its overall shape in the time domain.

### Outputs

- A pulse amplitude.
- A pulse start-time.
- A chisq measure of the similarity of the data trace to the pulse template.

Given a signal template $A(t)$ and a noise power spectral density $J(f)$, the signal takes the form $S(t) = aA(t) + N(t)$, where $a$ is the signal amplitude and $N(t)$ is noise in the time domain satisfying $J(f) = \langle \tilde{N}^2(f) \rangle$ (where the notation $\tilde{N}$ represents the Fourier transform of $N$). We want to vary $a$ to find the description that best fits the data. The quantity we want to minimize in this fitting procedure is a frequency-domain $\chi^2$, written as

$$\chi^2(a) = \sum_n \frac{|\tilde{S}_n - ae^{-2\pi i t_o f_n \tilde{A}_n}|^2}{J_n} \tag{8.4}$$

which, when minimized, results in a best-fit amplitude $\hat{a}$ of

$$\hat{a} = \frac{\sum_n \frac{\tilde{A}_n^* \tilde{S}_n}{J_n}}{\sum_n \frac{|\tilde{A}_n|^2}{J_n}} \tag{8.5}$$

Notice that this equation for $\hat{a}$ is simple to compute. We are essentially applying a 'filter' of the form $\sum_n \frac{\tilde{A}_n^*}{J_n}$ (divided by a constant).

This finds the optimal value of $a$, but we actually hope to minimize $\chi^2$ in the 2D space of $a$ and $t_o$. We won't prove this fact here (see J. Filippini's or S. Golwala's theses) but the *values of* $t_o$ *which extremize* $\chi^2(a, t_o)$ *are also the values of* $t_o$ *that extremize* $\hat{a}(t_o)$. It turns out (again, we won't show this either) that computing $\chi^2(a, t_o)$ is quite computationally difficult, so instead, CDMS has traditionally computed $\hat{a}$ for many values of $t_o$, and then choose the maximum $\hat{a}(t_o)$ (rather than the minimum $\chi^2$. Foreshadowing: notice that in the case of multiple extrema, the maximum $\hat{a}(t_o)$ can correspond to any of the multiple minima of $\chi^2$ (not necessarily the smallest minimum).

Another important caveat: notice that we have specifically mentioned that we expect phonon pulse shape variation, but that the optimal filter strategy depends on a pulse template of a specific shape. Pulses of slightly different shapes (but the same total integral) will be assigned somewhat different amplitudes by the optimal filter. This is not the case for charge, where the pulse shapes are nearly identical for each pulse, shaped by the readout electronics rather than the detector. Rather than averaging together a variety of different pulse shapes, the CDMS II phonon pulse template was an analytical function: a double-exponential defined by a rising and falling time constant, as $A(t) = A_0(1 - e^{t/\tau_{rise}})e^{-t/\tau_{fall}}$.

Using the optimal filter strategy, each CDMS II inner charge channel has a resolution of ~250-500 eV, depending on detector. The combined signal from the four phonon channels if of resolution ~100-500 eV. These resolutions also vary from run to run, depending on noise conditions. There is a trend that the later towers have worse phonon resolution (for unknown reasons) and silicon detectors have better phonon resolution (due to a closer match between the faster pulse shape in Si and the $\tau_{etf}$ of

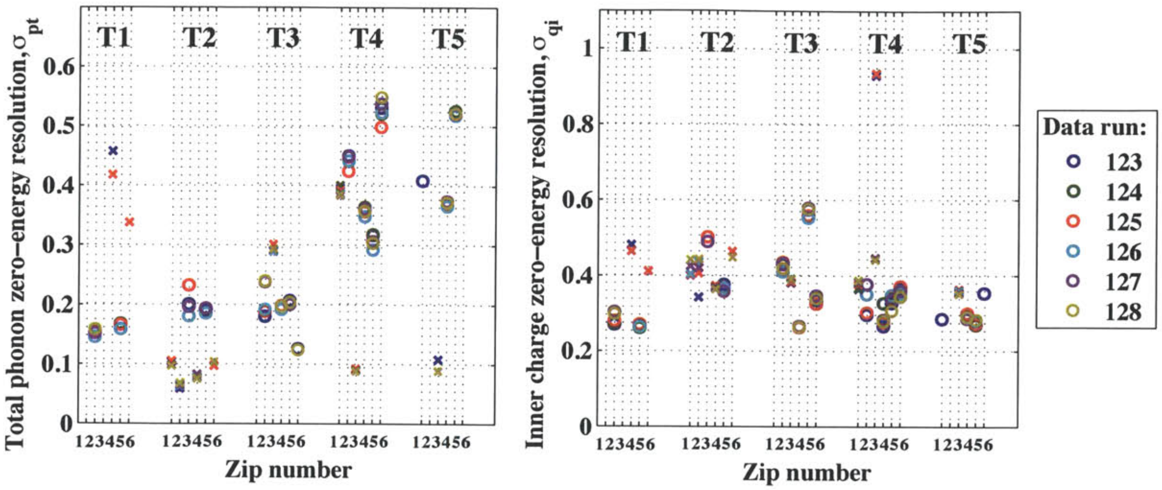

Figure 8-3: Optimal filter energy resolutions for CDMS II, for total phonon energy (left) and inner charge energy (right). Circles indicate germanium detectors, crosses are silicon detectors. Plotted by D. Moore.

the sensors). A summary of resolutions is shown in Figure 8-3

## 8.3 Energy and Ionization Yield

The two optimal filter charge amplitudes can be combined (taking account of crosstalk) into a total charge amplitude, and the four phonon amplitudes can be combined into a total phonon amplitude. For the phonon measurement, if one combines the four raw pulses together before applying the phonon optimal filter, both the standard noise and the position-variation noise are reduced.

These quantities in electrical units must be scaled by calibration factors before they become energies. The foundation of this calibration is a $^{133}$Ba gamma source, particularly its 356 keV peak. As shown in Figure 8-4, the charge signal showed some variation in amplitude with position, an effect that has never been fully understood. The effect is easily corrected for, however, using phonon-based position measures.

The phonon channels are similarly scaled such that, given a statistically significant amount of $^{133}$Ba events, each of the four channels contains roughly 1/4 of the scaled phonon energy, and the overall sum is scaled to a keV scale using the 356 keV line. The measured phonon energy is of course the sum of of the recoil energy and the

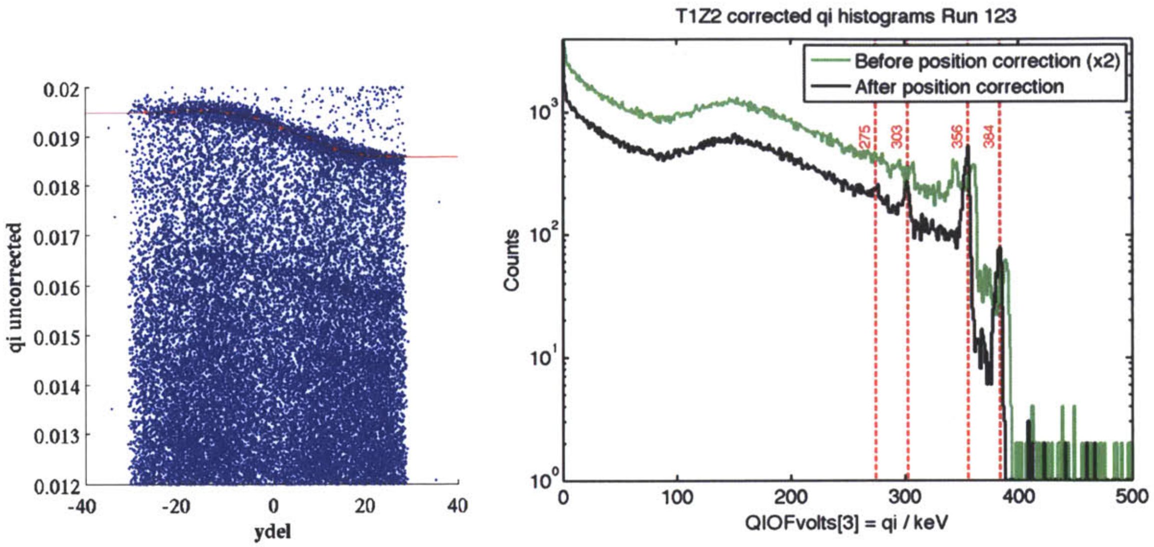

Figure 8-4: Two important steps in the calibration of the charge amplitude. On the left, we show the variation of charge amplitude with y position (determined through phonon relative timing between channels). A correction spline is fitted to the 356 keV peak as a function of position (in x also, not shown). The right plot shows the scaled $^{133}$Ba calibration data, both before (green) and after (black) this position correction, with several gamma lines indicated. Both plots by J. Filippini.

Luke phonon energy. To obtain the recoil energy from our measurements, then, we subtract $P_{Luke}$ from $P_{tot}$, as

$$P_{recoil} = P_{tot} - P_{Luke} = P_{tot} - Q_{tot}\frac{V_b}{E_{charge}} \tag{8.6}$$

where $Q_{tot}$ is the sum of the two charge channel energies, $V_b$ is the bias voltage (typically 3V), and $E_{charge}$ is an energy per Coulomb of charge. If the charge fiducial cut is applied, meaning that no energy is deposited in the outer charge channel, then the charge noise can be reduced by using the signal from only the inner channel, as

$$P_{recoil} = P_{tot} - Q_{inner}\frac{V_b}{E_{charge}} \tag{8.7}$$

## 8.4 Phonon-Based Position

The ‘Z’ in ’ZIP’ emphasizes the position-measuring ability of the four independent phonon channels to determine event position within the crystal. Of course, since the four channels are arranged in an x-y array, the measurement of position in x and y is far superior to the measurement in z, but some z-position ability does exist, as we will see.

The four individual phonon optimal filter **amplitudes** can be compared, to judge position based on phonon energy partition. For an x position, we compare the fraction in the right two channels with the left two channels, and similarly for y.

$$\text{x partition} = \frac{\text{(pc+pd) - (pa+pb)}}{\text{pa+bp+pc+pd}} \tag{8.8}$$

$$\text{y partition} = \frac{\text{(pa+pd) - (pb+pc)}}{\text{pa+bp+pc+pd}} \tag{8.9}$$

A radial partition quantity can be defined as r partition $= \sqrt{(\text{x partition})^2 + (\text{y partition})^2}$.

The arrival time of the phonon pulse at the four phonon channels (relative **delay**) can be similarly compared to measure x-y position. The 20% time on each phonon pulse was found to give the best signal-to-noise on this position measurement. In the primary channel, the 20% point seems to be determined by the Luke phonon arrival time, whereas in non-primary channels, the 20% point reflects the slow diffusion of the primary component. X and y positions based on delay were derived comparing the primary channel with the neighboring channel (in either x or y). For example, given an event with primary channel A,

$$\text{x delay} = A_{20\%} - D_{20\%} \tag{8.10}$$

$$\text{y delay} = B_{20\%} - A_{20\%} \tag{8.11}$$

(similarly for each of the other three quadrants). A radial delay can be defined as

$\text{rdelay} = \sqrt{(\text{x delay})^2 + (\text{y delay})^2}$ and a radial angle can be similarly defined.

The partition space and the delay space are illustrated in the upper portion of Figure 8-5. Notice that the outermost events (in red) are not at the outermost portions of the distributions. This scary 'foldback' in the two spacial quantities can be unfolded by combining partition and delay information, as seen in the lower portion of the figure.

Let's try to make sense of the shape we see in the lower-right of Figure 8-5. At low radius, timing and partition both trace the true radial position, as naively expected. As radial position increases, the first effect to appear is an artificial decrease in partition radius, *i.e.*, phonon energy is more shared with non-primary channels than expected. Our understanding is that the phonons from high-radius events reflect off the sidewall, thereby becoming more evenly distributed in the crystal. The folding back in delay space is a smaller effect; it starts being evident at a somewhat higher radius, and is a smaller magnitude than the foldback in partition. The understanding here is that not only are the phonons reflecting off the sidewall, but, because they reach the sidewall while still of high energy (diffusive), they downconvert due to sidewall interactions. Because they become ballistic sooner, they reach the non-primary channels sooner, and the delay between channels is suppressed. The curvature of the shape in Figure 8-5, then, involves the interplay of many aspects of phonon physics in the detector: downconversion in the bulk, downconversion at the sidewall, and varying propagation velocities with energy.

We have focused on understanding how the radial position maps onto Figure 8-5 and why. Beautifully, the third dimension of position, z, can also be extracted from this partition-delay space (though with worse fidelity). Let's compare an event near the phonon surface to one near the charge surface (but at the same x-y position). Near the phonon surface, the partition will be maximized because the phonons will be highly diffusive at the sensors and therefor highly absorbed by the Al. The phonons from the charge-side event, on the other hand, will have significantly downconverted on their 1cm voyage to the sensors, and will be fairly ballistic by that time, more easily spreading throughout the crystal and producing a lower partition effect. In other words, the phonon side is more partitioned and the charge side is less partitioned. In order to separate Z using a partition-delay space, though, these two quantities must behave differently, and in fact partition and delay are anti-correlated in Z. The phonon-side event will have a suppressed delay and the charge-side event will have an enhanced delay. In the primary channel, any Z position will have the same 20% time, because the primary pulse 20% time is is dominated by the Luke contribution (produced nearly instantaneously with the event, at all Z, because of charge propagation. The phonons from the recoil itself, on the other hand, dominate the non-primary channels rise time, and these phonons take longer to reach those phonon channels from the charge side than from the phonon side (just due to the distance of travel).

Till now, we have been discussing phonon quantities of $^{133}$Ba calibration, which, like the vast majority of CDMS background events, are gammas. The partition and timing quantities change when the luke phonon population is smaller (*i.e.*, for nuclear recoils or for surface events with significant trapping) and also for events *extremely* near the surface (within a micron or so), where the initial phonon downconversion is altered by the surface.

**Phonon-Side Surface Events** have an exaggerated x-y partition, due to their extremely high phonon energies at the sensor, leading to extremely short mean free path lengths and extremely high absorption rates. The interactions with the surface also speed up the downconversion, meaning that those phonons that are not immediately absorbed propagate quickly to the non-primary channels, shrinking the timing difference and reducing the delay radius.

**Charge-Side Surface Events** similarly have accelerated downconversion, leading to faster (ballistic) expansion. These phonons reach the non-primary channels faster, reducing the radial delay.

**Nuclear Recoil Events** have reduced Luke contribution, changing the ‘shrimp’ plot in many ways. In the most general terms, though, the width of the shrimp is increased due to an exaggeration of the Z-dependent effects. This is because Luke phonons are emitted at all Z, so suppressing this Z-independent portion

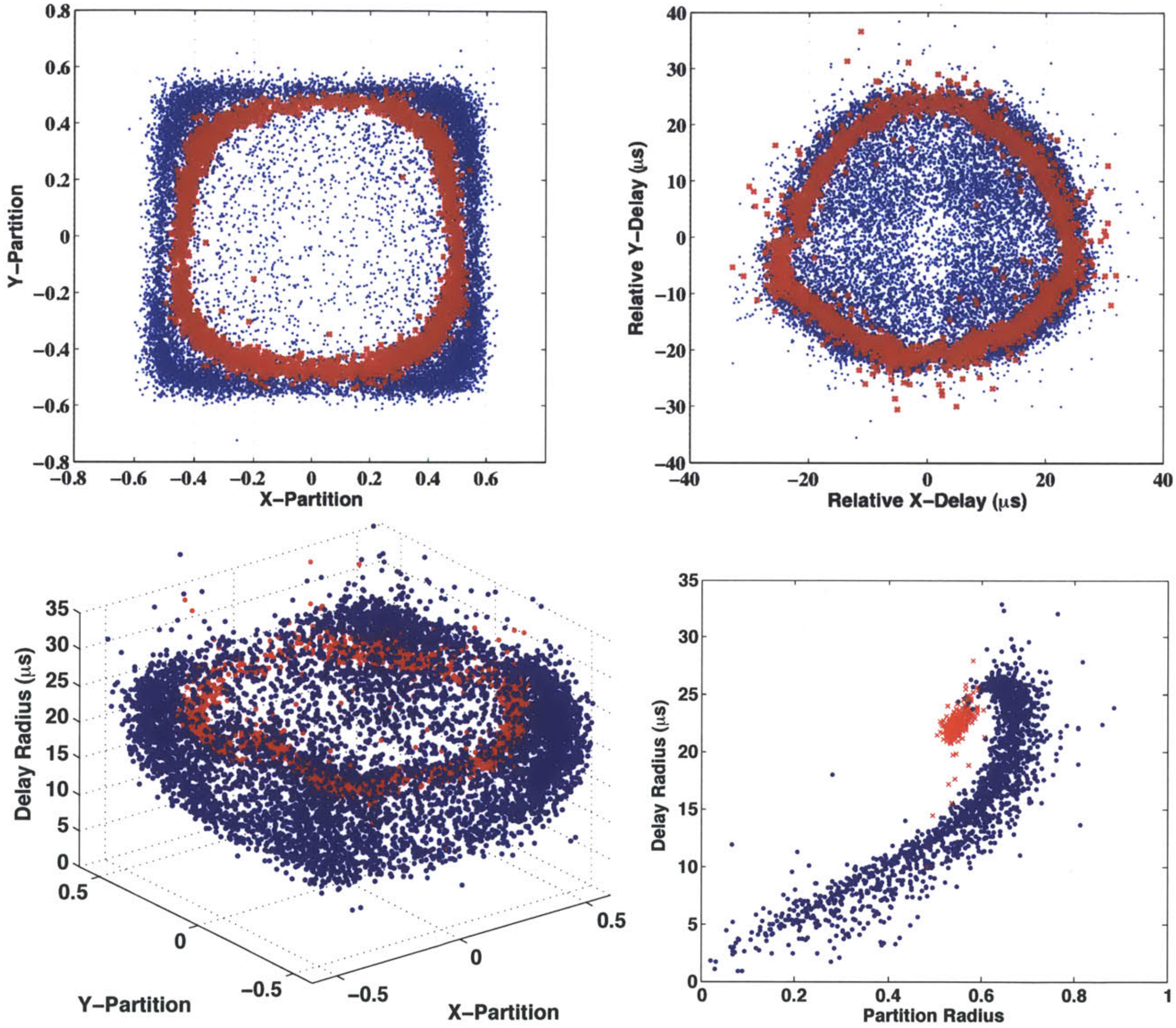

Figure 8-5: $^{133}$Ba calibration events plotted in both the partition (top left) and delay (top right) x-y spaces. Below left, the radial delay is added as a third dimension to the x-y partition space, and below right, a thin angular slice of this space is shown. Events highlighted in red have some portion of their charge signal in the outer charge channel. Plots from Z. Ahmed.

of the phonons frees the recoil phonons (emitted at a particular Z) to dominate the observed characteristics.

These effects are illustrated in a cartoon, Figure 8-6.

At no point in the data analysis do we map these partition and delay quantities to actual physical position. This is of course possible, especially with the aid of a monte carlo simulation, but adds no new information or additional discrimination potential to the analysis.

## 8.5 Normalization by position

As should be clear by now, phonon physics of diffusion and downconversion create a great many position-dependent effects within a CDMS II detector. This is both a blessing and a curse; events within several microns of the top and bottom surface in z (dangerous for their reduced charge yield) can hope to be tagged by their position-dependent phonon physics, but this difference must be deconvolved with the equally large differences in phonon physics due to x and y variations.

We have discussed how phonon and charge physics creates differences in high-radius events. There are additional (and sometimes equally important) position dependencies arising from from spacial variation in the QET array. The transition temperature of the TES sensors is a very delicate parameter, very sensitive to slight differences in the mixture of the two phases of W, the thickness of W, impurities, tension in the W film, and any other even subtler effect that can be imagined. It is impossible to keep the $T_C$ exactly the same over the surface of the detector, which means that the heat capacity, saturation energy, electrothermal feedback time $\tau_{etf}$, and other important characteristics vary over the detector surface.

Determining if an event is at high radius (and therefor potentially of suppressed yield) is easy, using the outer charge electrode as a veto. Determining if an event is of extreme Z (the only other source of yield suppression is *hard* in this detector, and the focus of much of the analysis. The only way to distinguish these extreme-Z surface events from other events is based on their pulse shape, as seen in Figure 8-7. But

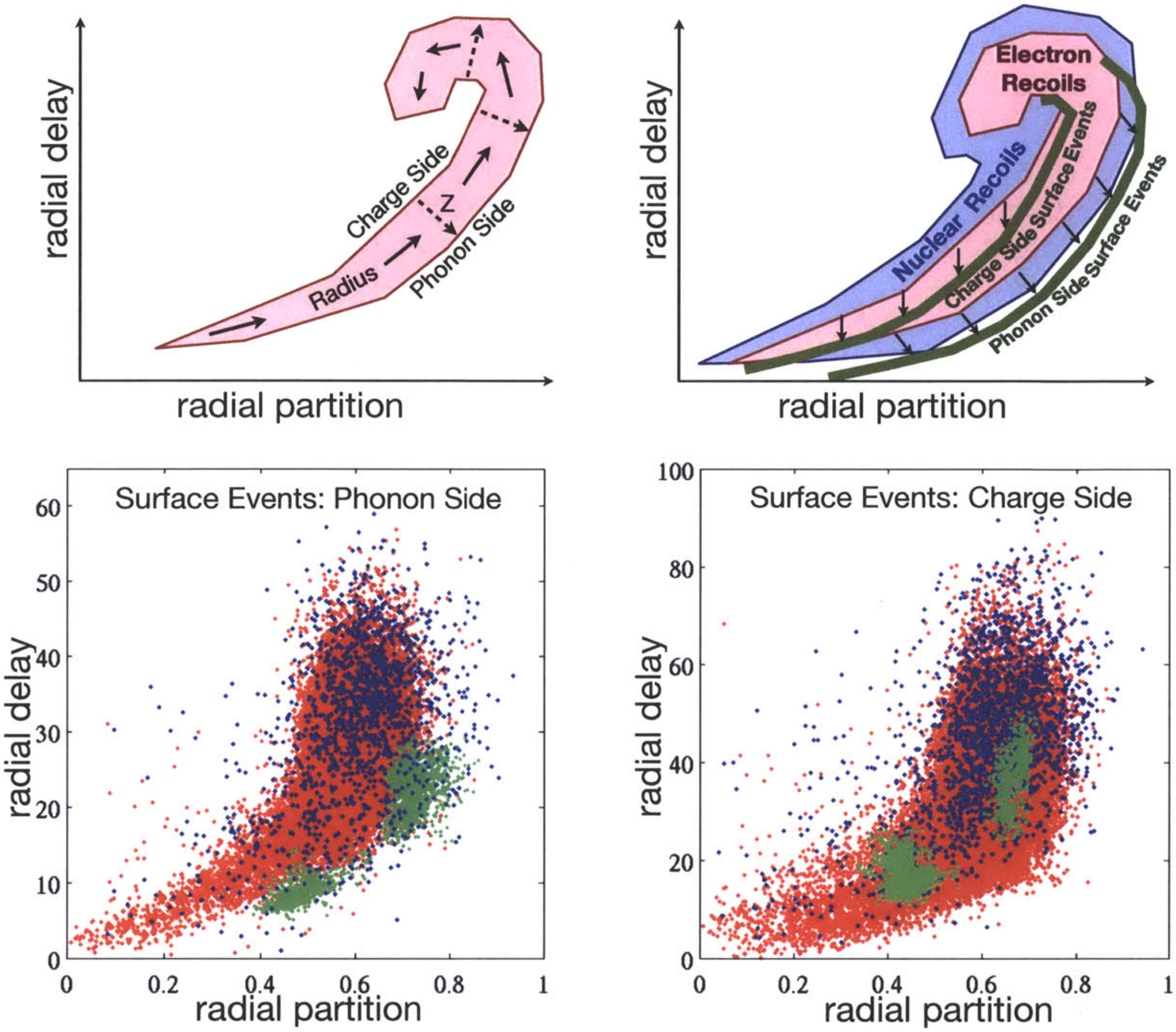

Figure 8-6: The radial delay vs. radial partition space, shown several different ways. In the upper left, the electron recoil distribution is shown, labeled by position within the detector. In the upper right, nuclear recoil and surface event populations are added to the cartoon, showing their differing distributions. The lower row shows the test results from detector G31 at Berkeley, showing electron recoils, nuclear recoils, and surface events (betas), colored as in the cartoon. The data looks unlike the cartoon for several reasons: the charge fiducial volume cut has been applied, eliminating the high-radius events, the beta source ($^{109}$Cd is colimated to only expose points at two radii, events from all angles have been shown in the same plot (blurring out the usually more distinct shapes), and the nuclear recoil events are very few in number compared to the electron recoil events (somewhat exaggerating the extent of the ER distribution and diminishing the extent of the NR distribution). Note that the two plots of of G31 data are taken from two different fridge runs at two different bath temperatures, leading to some run-to-run variation in the distributions as well in delay.

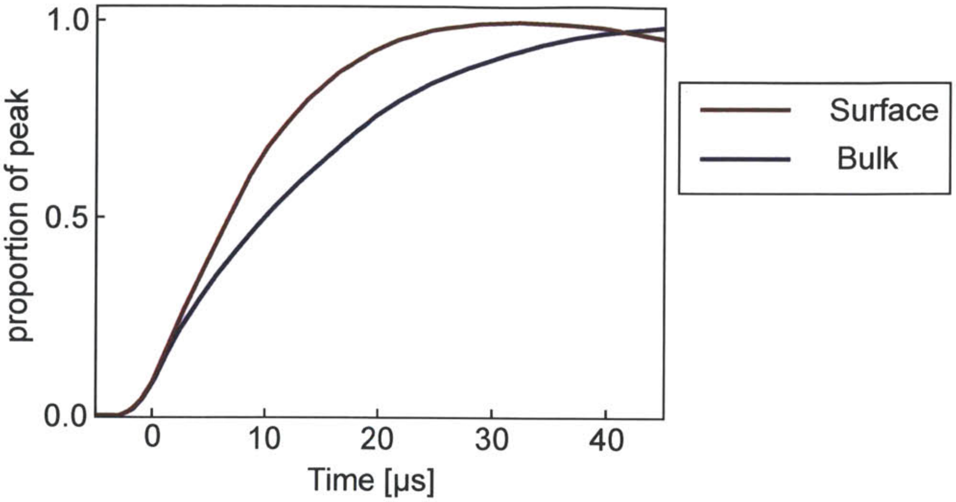

Figure 8-7: A comparison between surface and bulk events at exactly the same x-y position, using a tightly collimated 60 keV gamma source. Several events of each category were averaged together in this plot, adapted from a plot by P. Meunier and B. Serfass.

these subtle variations in pulse shape are at least as great as the differences caused by x-y positional variation. We have to normalize by x-y position, so that the subtle z-position differences are all that remain. Then, we can cut on our z-position variable to have a clean sample of non-surface events on which to perform our dark matter search.

How do we normalize by x-y position? For a thorough explanation, see the thesis of Z. Ahmed. Here, we describe the steps qualitatively:

1. Use electron recoil calibration data to populate the [x partition, y partition, r delay] space (illustrated in the lower left of Figure 8-5).

2. Define a metric in this space (a definition for distance). Basically, the idea is to define a weighting between r delay and x-y partition. If we call this weighting $\ell_{del}$, then the metric will be $d = \sqrt{(\Delta\text{xpart.})^2 + (\Delta\text{ypart.})^2 + (\Delta\text{rdel.}/\ell_{del})^2}$.

3. For each event in the calibration data, find the nearest neighbors in the [x partition, y partition, r delay] space using the defined metric.

4. For each event in the calibration data, use the nearest neighbors to define an

average value for the quantity you are trying to xy-normalize, and create a normalization factor for this [x partition, y partition, r delay] position, such that the average value of the nearest neighbors, times the normalization factor is some (arbitrary) xy-normalized value.

The result of this procedure is a mapping between each position in [x partition, y partition, r delay] space (appearing in the calibration data) and a normalization factor for that position. Each quantity that one might want to normalize (basically, our discriminators: yield and pulse shape quantities) requires its own map between position and normalization factor, but the hardest step (determining the set of nearest neighbors for each calibration event) is shared. Once these mappings have been made using ($^{133}$Ba gamma) calibration data, data from the WIMP-search (no calibration source) periods can be normalized by first finding the nearest event appearing in the calibration data, and then applying that event's (that position's) calibration factor.

There are several subtleties that deserve mention, since this normalization process was so central to the CDMS II analysis.

First, we should emphasize again that we are not trying to normalize by 3-D position, but by x-y position. If an event's nearest neighbors contain a significant number of surface events, then we will normalize out the very differences we are trying to uncover. It's delicate! For this reason, the number of nearest neighbors (relative to the total number of events in the calibration data) used to determine the calibration factors (*i.e.*, the spacial extent of the averaging) plays an important roll in determining the usefulness of the x-y calibration. The optimization of this number of nearest neighbors was performed simply by varying the number and seeing which value gave the best surface discrimination, resulting in an optimized value of ~5 events per 1000 events in the data sample. Typical calibration data samples were on the oder of 1e5 events.

A second subtlety is the improvement of the method by including energy in the nearest-neighbor space. This new metric quantity then, is

$$d = \sqrt{(\Delta \text{xpart.})^2 + (\Delta \text{ypart.})^2 + (\Delta \text{rdel.}/\ell_{del} + (\Delta \text{E}/\ell_E)^2} \tag{8.12}$$

There are clear variations in pulse shape (and also phonon energy as measured using a shape-dependent optimal filter strategy) with energy, and these energy-dependencies can be naturally folded in with the position-dependencies into one single normalization step. One of the main causes for energy-dependence in CDMS is TES saturation (typically, the very local saturation of some small portion of a channel's array), but timing parameters too can be significantly energy-dependent simply because they are harder and harder to measure at low signal-to-noise (more on this later). Altogether, then, the x-y-E normalization strategy follows from three parameters: the two weightings in the metric ($\ell_{del}$ and $\ell_E$) and the number of nearest neighbors. A description of the optimization of these three parameters will appear soon in published form.

The third and last major subtlety is the selection of the calibration data from which the calibration factors are defined. We select this sample using all available data quality cuts, as well as a selection in yield to specifically eliminate surface events from the sample. Yield is a very position-dependent quantity, so this initial selection is very rough. In fact, what is done, is the initial sample (selected using rough yield cuts) is used to create table of normalization constants, and then these normalization constants are applied to the calibration data, and then the selection in yield is further refined. We have essentially checked the calibration sample for self-consistency her, asking the question 'does each event have roughly the same yield as its nearest neighbors?", eliminating any event from the calibration sample for which this is not the case.

## 8.6 Improving the delay-based position measures

We have emphasized how normalizing the discrimination quantities by their position (and energy) is vital to the analysis, so the success of our analysis then depends on how well we can measure position. The optimal filter strategy employed for both the total charge and phonon energies, as well as the individual channel amplitudes (for position through energy partition) works extremely well. Phonon timing, on the other hand, is much harder to measure. Essentially, timing measurements are based

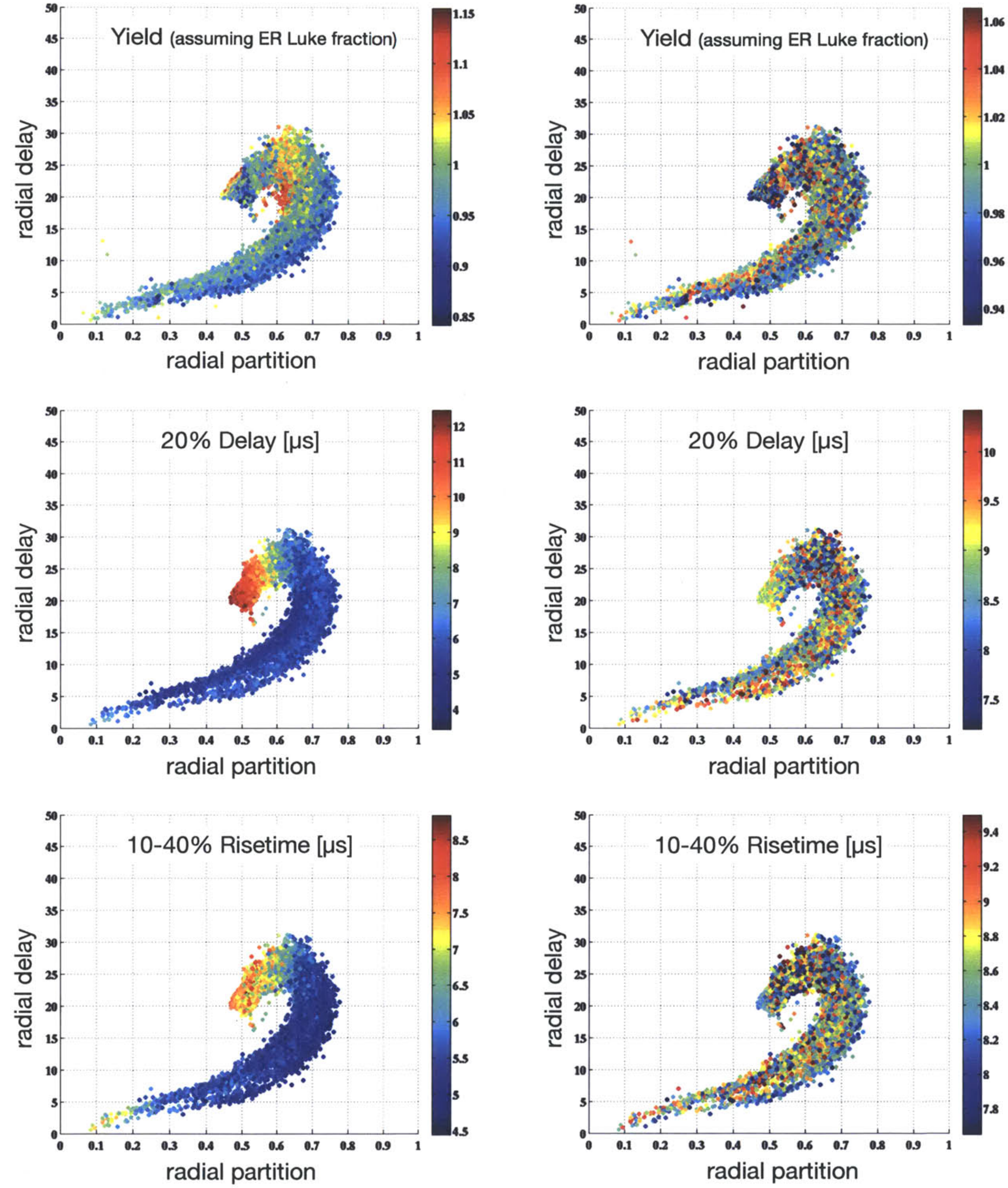

Figure 8-8: Ten degree slices (in partition angle) of the partition-delay space for r125 for detector T1Z5. Coloring is according to the three main discrimination quantities: yield (top), delay between charge start time and primary phonon channel 20% time, and the 10-40% risetime for the primary phonon pulse. The left column shows these values before normalization; the right column shows the same events after. Note that this is $^{133}$Ba data, and that the colorbars are differently scaled on the left and the right.

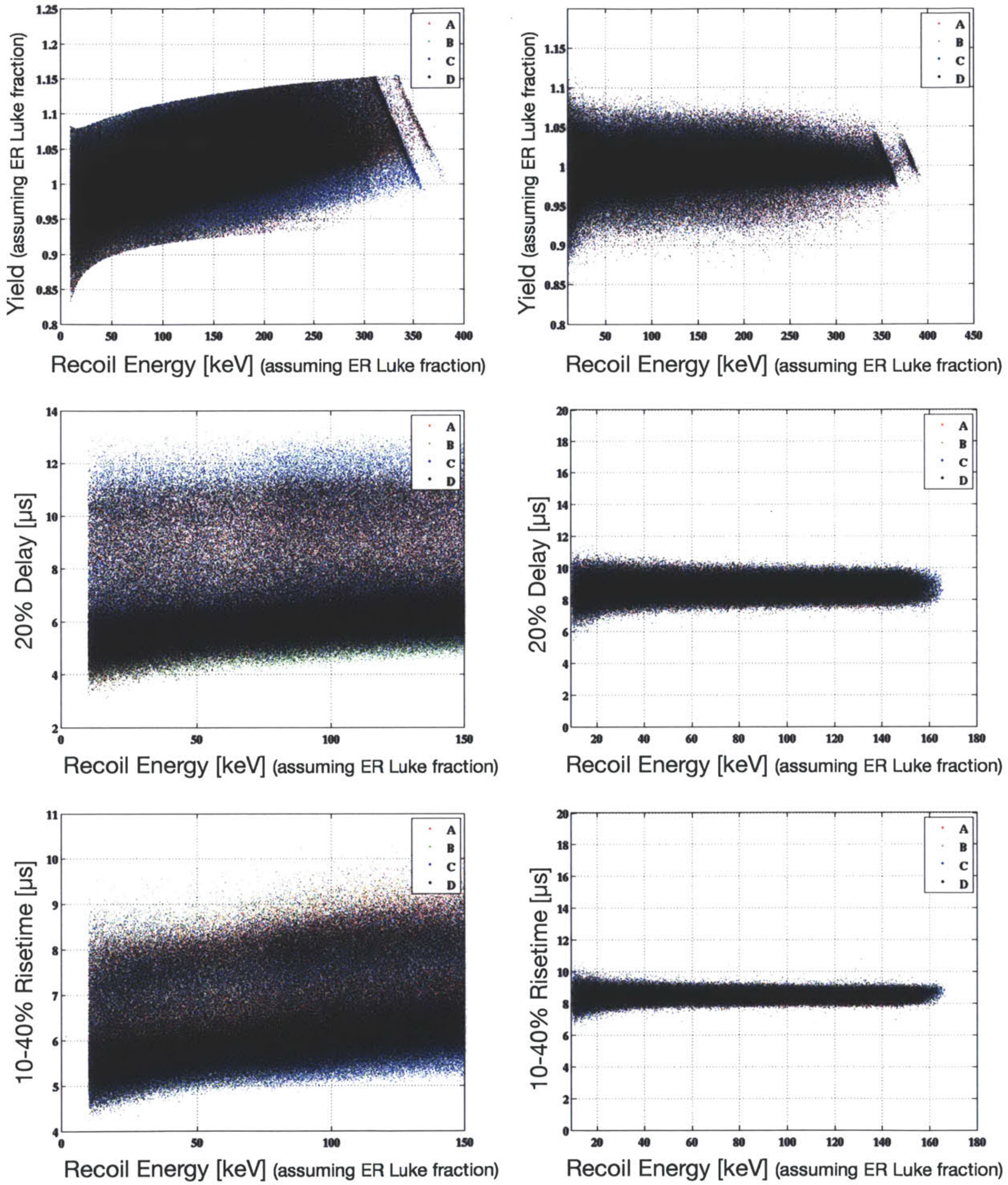

Figure 8-9: The same discrimination quantities as in Figure 8-8 (and the same $^{133}$Ba data for the same detector), but here shown for all position as a function of energy (before normalization on the left, after normalization on the right). Dot coloring is according to primary phonon channel.

on only the few time bins of the rising edge, whereas the amplitude is based on the whole pulse, so there is simply less information available upon which to make the measurement.

The charge start time is assumed to be the event time (for practical purposes), and is determined using the optimal filter strategy. The 10%, 20%, and 40% points on the four phonon pulses (usually relative to the charge start time) are derived using a very simple strategy, in the time domain, called 'walking'. First, the pulse is smoothed using a low-pass Butterworth filter. Then, starting from the peak of the pulse, an algorithm walks towards earlier times, and records the first time at which the pulse goes below each percentage (40%, 20%, etc).

This strategy works well, assuming two conditions: the amplitude of the noise is much smaller than the amplitude of the pulse, and the slope of the rising edge is relatively steep. If either of these two conditions fails, then the rising edge will be non-monotonic, and the walk algorithm will find the first instance of a 20% point at the position of a random downward fluctuation from the noise, rather than at a point more accurately representative of the rising edge timing. See Figure 8-10 for an illustration of this pathology.

Naively, this pathology is telling us to smooth the pulses more, by decreasing the cutoff to the low-pass filter. But notice that the rising edge slope of primary pulses would start to loose its distinctive (and vitally important to the analysis) steep shape if we decrease this low-pass filter for all pulses. We need, then, some strategy where we filter each trace according to a low-pass filter that is tailored to the pulse. This was done for runs 125-128 by taking a wide variety of pulses, both primary and non-primary, at high and low energies, and asking the question "what is the highest cutoff frequency for which the rising edge is monotonic, as a function of pulse amplitude over noise?" A clear trend was seen, and an empirically-defined function was created, mapping pulse amplitude (normalized by noise amplitude) to filter cutoff. These new pulse-specific filter definitions worked wonderfully on the low-energy non-primary pulses (which determine the delay position used in position normalization), and had little effect on the primary pulses of any energy (the timing of which were well-measured by both strategies).

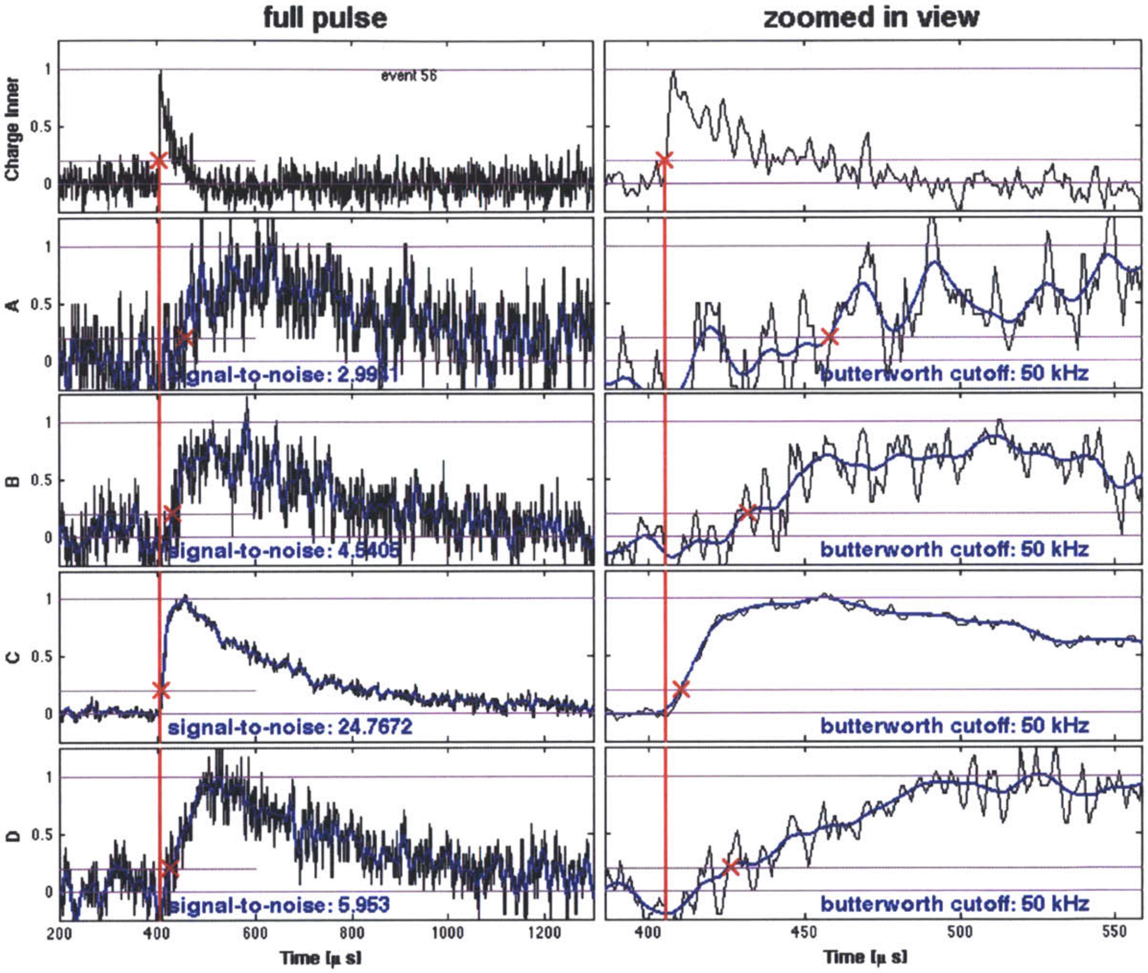

Figure 8-10: The derivation of phonon delay quantities, illustrated using a Butterworth filter with a 50kHz cutoff. At top is the inner charge channel (used to define the charge start time) and below are the four phonon channels (black is the raw data, blue is after application of the filter). The derived 20% times are indicated with red crosses. The same event is shown on both the left and the right (zoomed in on the rising edge). Channel C is the primary phonon channel (well-measured using a 50kHz cutoff), and channel A is the opposite channel (poorly-measured using a 50kHz cutoff).

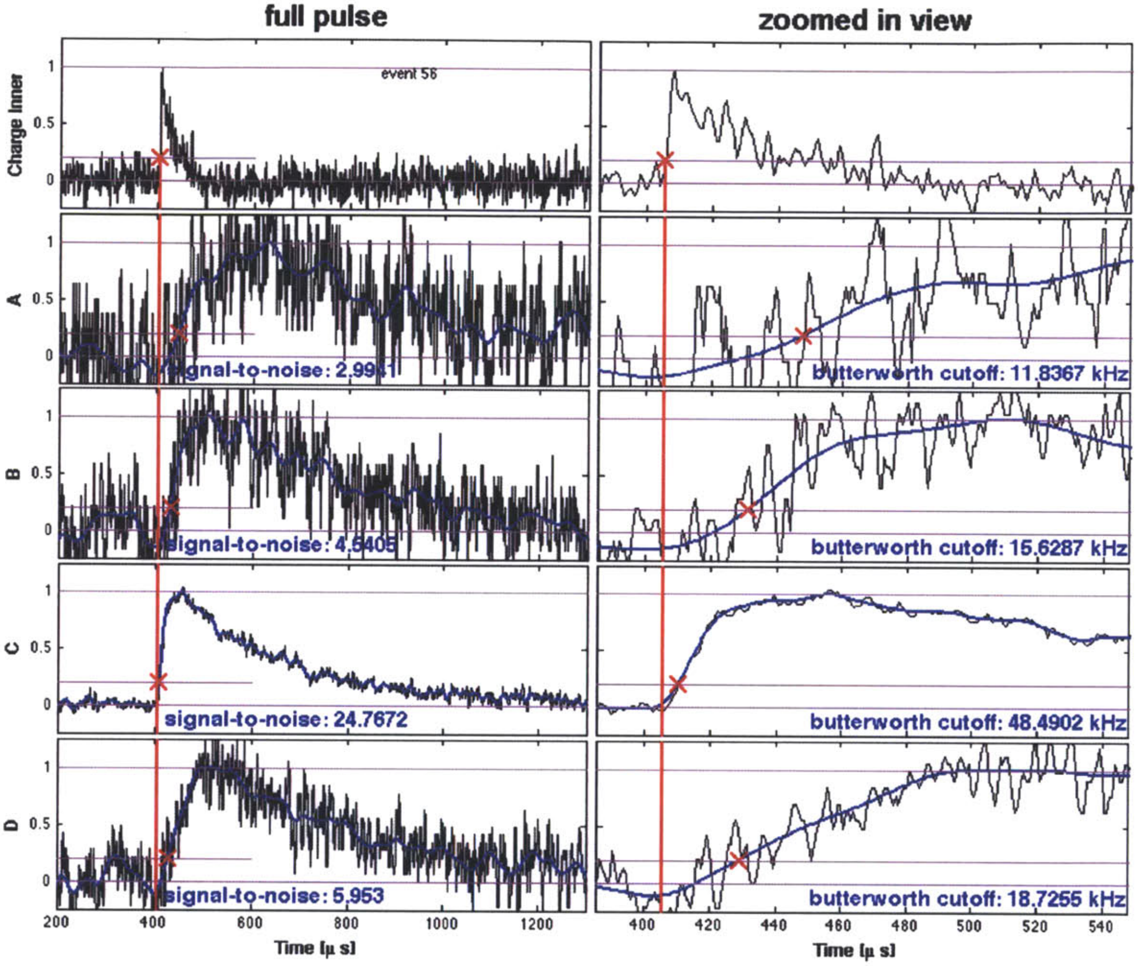

Figure 8-11: The same event as in Figure 8-10, but with the pulse-specific filter applied before finding the 20% times. The filter applied to the primary pulse is nearly unchanged; the filter applied to the non-primary channels greatly suppressed the high-frequency content. The 'signal-to-noise' quantity is simply the ratio between the pulse amplitude and the noise rms (as measured by the first 500 bins of the trace, the 'prepulse' data). This 'signal-to-noise' quantity is then mapped into a Butterworth cutoff frequency using an empirically-defined mapping function.

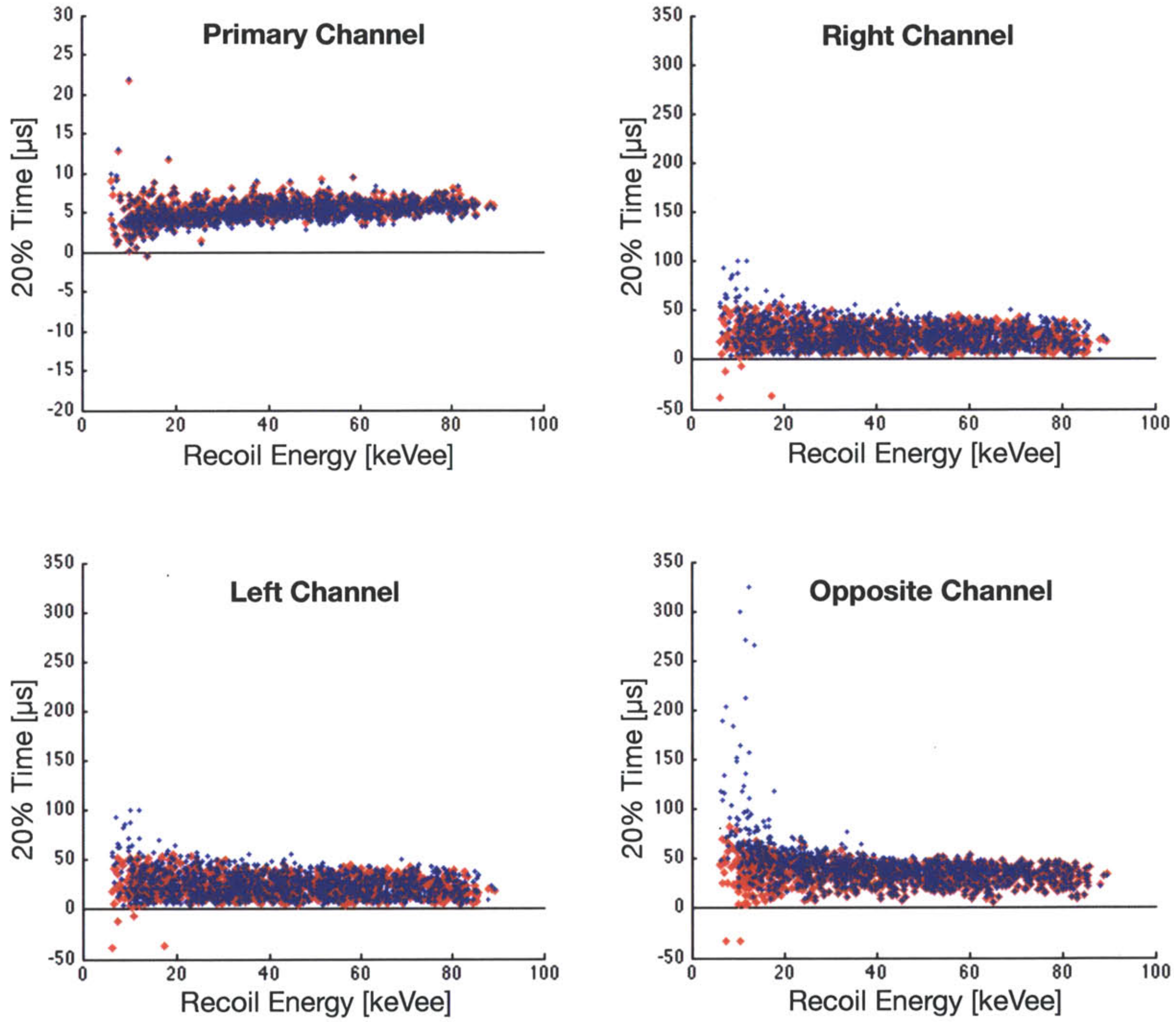

Figure 8-12: The 20% delay time as a function of energy is shown for a small population of $^{133}$Ba calibration events. Each event has four different 20% delay times, shown in the four separate plots: delay in the event's primary channel, in the event's opposite channel, and in the two 'side' channels. Blue crosses represent the delay times with a constant 50kHz cutoff Butterworth filter applied, and the larger red dots represent the same quantity as derived using a pulse-height-dependent filter cutoff. Notice there is very little difference in the primary channel (with a fast rising edge and a high signal-to-noise) but a very significant difference at low energies in the opposite channel (with a slow rising edge and a low signal-to-noise). Notice also that there is a small number of poorly measured events at low energies, but these can be cut by enforcing that they delay must be positive.

## 8.7 Turning energy, yield, and pulse shape into a dark matter limit

The data preparation is clearly a complicated procedure. The raw pulses have been digitized, filtered, optimal-filtered, calibrated, and position-normalized. We haven't even mentioned the data quality cuts, of which there are many. The huge amount of information the detectors provide for analysis is both a blessing and a curse. Once the data has been put into a polished form by the above procedures, then cuts are defined in yield and pulse shape, in such a way as to maximize the acceptance of nuclear recoils while minimizing the leakage of electron recoils, thus maximizing the scientific reach of the exposure. This final step of cut definition and optimization has been described quite fully in many other CDMS II theses (for example, by Z. Ahmed or M. Fritts), and here we simply summarize the results, based on the published report.

CDMS II operated an array of 30 detectors (19 Ge and 11 Si) in a low-radioactivity installation in the Soudan Underground Laboratory, Minnesota, USA [10]. The depth of the experimental facility (713 meters below the surface) greatly reduces the rate of background events from particle showers induced by cosmic rays. Nearly all remaining events from this source were identified using a layer of plastic scintillator surrounding the detector volume. Inner layers of lead and polyethylene further shielded the detectors against environmental radioactivity. Data taken during four periods of stable operation between July 2007 and September 2008 were analyzed for this work. Due to their greater sensitivity to spin-independent WIMP scattering, only Ge detectors were used to search for WIMP scatters. After excluding periods of poor detector performance, a total exposure to WIMPs of 612 kg-days was considered for this work.

### 8.7.1 Summary of WIMP selection criteria

The data selection criteria (cuts) that define the WIMP acceptance region were developed using calibration sets of electron and nuclear recoils obtained during regular *in situ* exposures of the detectors to $^{133}$Ba and $^{252}$Cf sources.

Candidate WIMP scatters were required to be within $2\sigma$ of the mean ionization yield of nuclear recoils and at least $3\sigma$ away from the mean ionization yield of electron recoils, have recoil energy between 10 and 100 keV, and have ionization energy at least $4.5\sigma$ above the noise level. They were further required to occur within the ionization fiducial volume and satisfy data quality criteria. An electron recoiling in the first few $\mu$m of the detector surface (a surface event) has ionization yield similar to that of a nuclear recoil, but has faster phonon timing. We thus require that a candidate event have phonon timing that is characteristic of a nuclear recoil. Finally, since WIMPs are expected to interact only once in the experimental apparatus, a candidate event was required to have energy deposition consistent with noise in the other 29 detectors (single-scatter event) and to have no significant activity in the surrounding scintillator shield from $-185$ to $+20$ $\mu s$ relative to the event trigger.

After detector calibration, we defined a series of criteria to identify candidate WIMP-scattering events. WIMP candidates were required to deposit 10-100 keV of energy in a single detector, have the ionization and phonon characteristics of a nuclear recoil and have no identifiable energy deposition in the rest of the array or in the scintillator shield.

All 30 detectors were used to identify particle interactions, but only the Ge detectors were used to search for WIMP scatters. Five Ge detectors were not used for WIMP detection because of poor performance or insufficient calibration data; four more detectors were similarly excluded during subsets of the four data-taking periods. For this exposure, we did not complete the full analysis for WIMP scatters in the Si detectors due to their lower sensitivity to coherent nuclear elastic scattering; the Si detectors are used only to identify multiple-scatter events.

Periods of poor detector performance were identified and excluded from analysis on a detector-by-detector basis. Data-quality criteria were developed using Kolmogorov-Smirnov tests performed on an array of parameter distributions. Special effort was made to exclude periods of poor detector "neutralization" by monitoring the ionization yield distribution. At the low temperatures required to operate the phonon sensors, impurities and defects in the crystal detector substrate can produce isolated charge "trapping centers." Our detectors required regular neutralization of these trapping centers [10] to maintain full ionization collection. After applying these data quality selections, the total exposure to WIMPs considered for this work was 612 kg-days.

### 8.7.2 Nuclear recoil selection

Nuclear recoils from the elastic scattering of neutrons emitted by the $^{252}$Cf source were used to define the ionization yield acceptance region. The acceptance region was taken to be the $2\sigma$ band about the mean neutron ionization yield ($2\sigma$ nuclear recoil band) and varied as a function of recoil energy (Fig. 8-15). These nuclear recoils were also used to develop the surface event rejection cut and measure the signal efficiency, as described below.

### 8.7.3 Surface event rejection

The sum of the rise time of the largest phonon pulse with its delay relative to the ionization signal was empirically found to provide the best discrimination between surface events and nuclear recoils. We optimized this cut using nuclear and surface electron recoils from the $^{252}$Cf and $^{133}$Ba calibration exposures. Surface-event rejection criteria based on this discriminator were tuned on the calibration data by maximizing the expected sensitivity for a $60\,\mathrm{GeV/c^2}$ mass WIMP. Figure 8-14 shows a key step in the setting of each detector's timing cut threshold, and Figure 8-15 demonstrates our surface-event rejection capability on the calibration data.

Despite the great discrimination power of this experiment, a small expected rate of misidentified background events remains. In the exposure considered here we expected to misclassify $0.8 \pm 0.1(\mathrm{stat}) \pm 0.2(\mathrm{syst})$ surface electron recoils as WIMP candidates. We also expect neutrons produced by cosmic rays and radioactivity to generate an average of $\sim 0.1$ nuclear recoils, which would be indistinguishable from WIMP scatters.

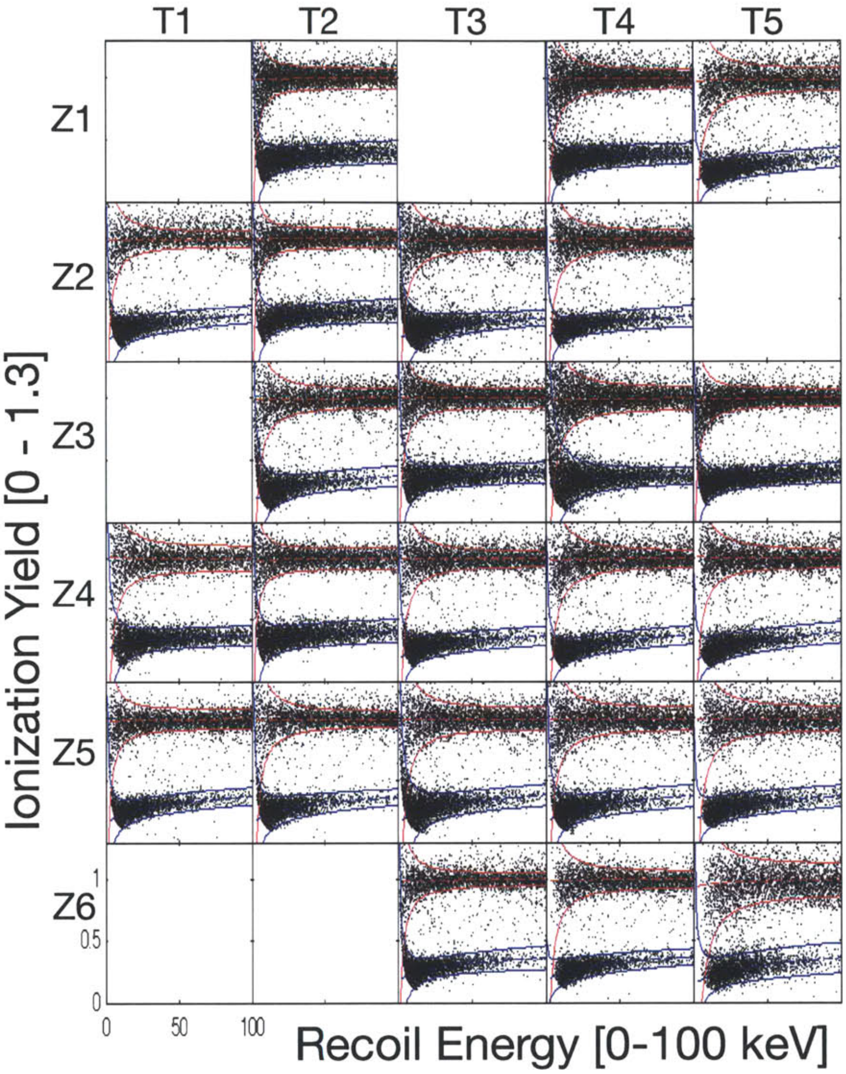

Figure 8-13: The run 123 electron recoil and nuclear recoil $\pm 2\sigma$ yield bands, overlain on $^{252}$Cf calibration data from run 123. Yield bands were defined separately for each run. Adapted from a figure by J. Filippini.

### 8.7.4 Signal efficiency

The fractional acceptance (efficiency) of our analysis cuts for nuclear recoils was measured as a function of energy using both neutron-calibration and WIMP-search data. The fiducial volume estimate was corrected for the systematic effects of neutron multiple-scattering by using Monte Carlo simulations. Our efficiency for signal events has a maximum of 34% at 20 keV. It falls to ~25% both at 10 keV, due to ionization threshold and flaring of the electron-recoil band; and at 100 keV, due to a

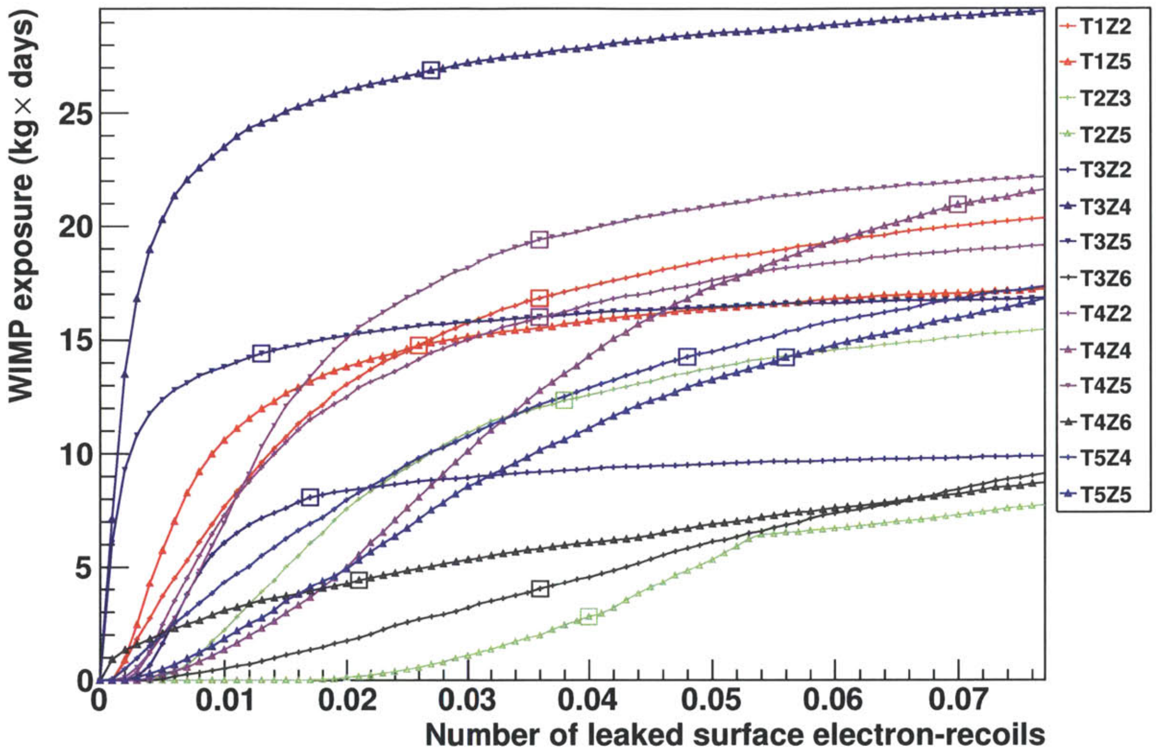

Figure 8-14: For each of the Ge detectors used to set the r125-r128 limit, the timing cut threshold was tuned separately (to maximize the *total* exposure while minimizing the *total* leakage. Notice that the cut setting (indicated with a square) for each of the detectors occurs at points of equal slope for all detectors. This is not a coincidence; the slope of the exposure-vs-leakage curve is the important quantity in optimally setting this cut position for many detectors.

drop in fiducial volume. After all selection criteria, the spectrum-averaged equivalent exposure for a WIMP of mass 60 GeV/c$^2$ is 194.1 kg-days. The efficiency at various stages of cut application is shown in Fig. 8-16.

### 8.7.5 Data blinding

To avoid unconscious bias, we performed a Òblind analysisÓ in which the exact selection criteria were defined without prior knowledge of the content of the signal region or its vicinity. Immediately following event reconstruction and calibration, we excluded from study all events in the entire exposure satisfying the following criteria: single-scatter events with no coincident activity in the scintillator shield and ionization yield within the $3\sigma$ nuclear band. All events in this "masked" group (which includes all potential WIMP candidate events) were automatically removed from the data files distributed to collaboration members, and so had no effect on the development of the analysis criteria. The masked events which did not meet the WIMP-selection criteria defined earlier in this work were restored to the analysis in the final stages of development for the surface-event cut and the surface event background estimate. Any candidate WIMP events remained hidden until all selection criteria were finalized and the choice of statistical technique for the limit calculation (Yellin optimum interval [113]) was made. On November 5, 2009, the entire data set was "unblinded": all remaining masked events were restored to analysis.

### 8.7.6 Results

We observed two candidate events at recoil energies of 12.3 keV and 15.5 keV, as can be seen in Figure 8-17. These events occurred during periods of nearly ideal experimental performance, were separated in time by several months, and took place in different detectors. These candidates match the expectations for WIMP scattering events, but the probability to have observed two or more background events in this exposure is 23%. The results of this analysis thus cannot be interpreted as significant evidence for WIMP interactions, nor can we reject either event as a WIMP scatter.

These data constrain the spin-independent scattering cross section between WIMPs and nucleons to be less than $7.0 \times 10^{-44}$ cm$^2$ ($3.8 \times 10^{-44}$ cm$^2$ when combined with CDMS II's previous results) for a WIMP of mass 70 GeV/c$^2$ [1]. While this work represents a doubling of our exposure, the observation of two candidate events leaves the combined upper limit nearly unchanged below 60 GeV/c$^2$ but allows for a modest strengthening in the limit above this mass.

---

[1] We calculate the 90% confidence level upper limit based on standard galactic halo assumptions [76] and in the presence of two events at the observed energies. We use the optimum interval method [113] with no background subtraction. A combined limit was also calculated by combining these data with all previous results from Soudan [6], including all candidates and with efficiency weighted by the exposure of each analysis. The abrupt features in these curves are consequences of threshold-crossings at which intervals containing one or more events could enter into the optimum interval computation. An improved estimate of our detector masses was used for the reported exposure calculation and applied retroactively to our previous CDMS II result.

### 8.7.7 Background estimates

#### Expected neutron background

Neutrons with energies of several MeV can generate single-scatter nuclear recoils that are indistinguishable from possible dark matter interactions. We consider two major sources of background neutron events: cosmic-ray muons interacting in or near the experimental apparatus (cosmogenic neutrons) and radioactive processes in materials making up and surrounding the apparatus (radiogenic neutrons).

The location of the Soudan Underground Laboratory, beneath a rock overburden equivalent to 2090 meters of water, greatly mitigates the background rate due to cosmogenic neutrons. Nearly all of the remaining cosmogenic neutrons are identified (vetoed) by coincident activity in the scintillator shield. Three such vetoed single-scatter nuclear recoil events were observed in this exposure, whereas none had been seen in previous CDMS II data. We used Monte Carlo simulations to estimate the remaining background from unvetoed cosmogenic neutrons. Simulations of muon-induced particle showers and subsequent neutron production were performed with Geant4 [5, 12] and FLUKA [52, 22]. Good agreement was found between both simulation packages in the predicted ratio of unvetoed to vetoed neutrons, as well as in the ratio of single- to multiply-scattering neutrons, but their predictions of the absolute event rate differed. We thus took as our background estimate the product of the observed number of vetoed single-scatter nuclear recoils (three) with the ratio of unvetoed to vetoed single-scatter nuclear recoils determined from Monte Carlo. This resulted in a more conservative estimate than that produced by either package on its own. This procedure, combined with corrections for efficiency and exposure time, predicts $0.04^{+0.04}_{-0.03}$(stat) unvetoed, cosmogenically-produced, single-scatter nuclear recoils in the reported data.

Our estimate of the radiogenic neutron background due to spontaneous fission and $(\alpha, n)$ processes was informed by measurements of samples of our shielding and detector materials, which were screened for U and Th daughters using high-purity Ge gamma counters. We also derived an independent estimate of the contamination levels of U and Th from a fit of the electromagnetic spectrum observed by our experiment to a Monte Carlo simulation. The combined estimate of contamination levels, together with the assumption of secular equilibrium, were used as input to a detailed Geant4 simulation that propagated the resulting neutrons through the experimental setup. The estimated radiogenic neutron background is between 0.03 and 0.06 events and is dominated by U spontaneous fission in the Cu cans of the cryostat. The radiogenic neutron background originating from the surrounding rock is estimated to be negligibly small in comparison to other sources.

### Expected surface event background, prior to unblinding

The number of misidentified surface events was estimated by multiplying the observed number of single-scatter events failing the timing cut inside the $2\sigma$ nuclear-recoil band by the expected ratio of events passing the timing cut to those failing it (the "pass-fail ratio"). This ratio was estimated using a combination of three methods with differing statistical and systematic errors:

1. The first method computed the pass-fail ratio from events that reside within the $2\sigma$ nuclear-recoil band and multiply scatter in vertically adjacent detectors (multiple-scatter events).

2. The second method uses multiple-scatter events surrounding the $2\sigma$ nuclear-recoil band (wide-band events). Wide-band events have different distributions in energy and in detector face (ionization- or phonon- side) from nuclear-recoil band events, affecting the pass-fail ratio. To account for these differences, the pass-fail ratio of these events was corrected using the face and energy distributions of events observed in the nuclear-recoil band that failed the timing cut.

3. A third, independent estimate of the pass-fail ratio was made using low-yield, multiple-scatter events in $^{133}$Ba calibration data, again adjusted for differences in energy and detector-face distributions.

All three estimates were consistent with each other and were thus combined to obtain an estimate prior to unblinding of $0.6\pm0.1$(stat) surface events misidentified as nuclear recoils.

### 8.7.8 Expected surface event background, after unblinding

After unblinding, a detailed study motivated by one of our candidate events revealed that an approximation made during the ionization pulse reconstruction degrades the timing-cut rejection of a small fraction of surface events with ionization energy below ~6 keV. Such events are more prevalent in WIMP-search data than in the data sets used to generate the estimate of misidentified surface events prior to unblinding. A refined calculation, which accounted for this reconstruction degradation, revised the surface-event estimate to $0.8 \pm 0.1(\text{stat}) \pm 0.2(\text{syst})$ events from the prior estimate of $0.6 \pm 0.1(\text{stat})$. The systematic uncertainty in this final estimate is dominated by the uncertainty in our assumption that the pass-fail ratio for multiple scatter events is the same as that for single scatter events.

### 8.7.9 Discussion of results

#### Additional information on the unblinded data

Figure 8-17 shows events on all detectors after all selection criteria have been applied, except the yield and timing cuts. Candidate events appear in the signal region of detectors T1Z5 (at 12.3 keV) and T3Z4 (at 15.5 keV). The dates and local times of the two candidate events are October 27, 2007 at 14:41 CDT (T1Z5 event) and August 5, 2007 at 20:28 CDT (T3Z4 event). Both occur during periods of good experimental performance. The candidate event observed with detector T3Z4 suffers from the ionization reconstruction effect described in the previous paragraph, which increases the possibility for it to be a surface event.

#### Varying the timing cut

To quantify the proximity of these events to the surface-event rejection threshold, we varied the timing cut threshold of the analysis. Reducing the revised expected surface-event background to 0.4 events would remove both candidates while reducing the WIMP exposure by 28%. No additional events would be added to the signal region until we increased the revised estimate of the expected surface-event background to 1.7 events. The corresponding limits that are obtained with these cut values are shown in Fig. 8-20. The figure also shows that while the allowed surface-event background chosen by the blind analysis gives, on average, the best expected limit over the range shown, the actual observed limit is never more than 15% higher in the range of 0.1 to 6 expected surface events.

## 8.8 What have we learned from CDMS II?

Figure 8-20 shows how (as intelligently as we could) we found a happy medium between a large exposure (but significant leakage of surface events) and confidence of having a zero-background experiment (but a much-reduced exposure). If we ran the experiment another year, or two, or ten, the exposure would not increase linearly with time, because the timing cut threshold would need to be more and more constrictive (due to the larger and larger integrated amount of surface events).

Although it is true that CDMS II was able to make background-free exposures, it is not at all true that the experiment was not back-ground limited. In fact, the total amount of exposure was severely limited by the fact that the surface event pulse shapes were only distinguishable from bulk nuclear recoils in rather subtle ways, with imperfect discrimination. Looking at Figure 8-16, one can see that the timing cut, necessary to eliminate the surface event background, cut our exposure in half. Clearly, the CDMS II strategy suffers from a significant problem, one that requires a significant technological change in order to probe smaller cross sections.

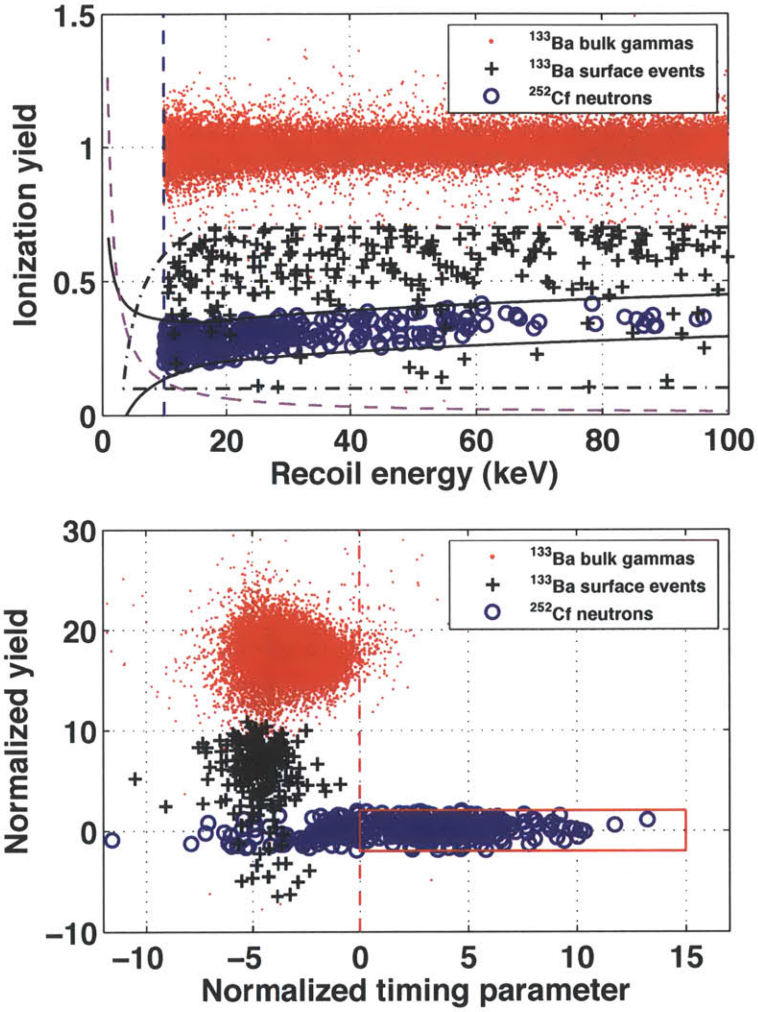

Figure 8-15: The power of the primary background discrimination parameters, ionization yield and phonon timing, is illustrated for a typical detector using *in situ* calibration sources. Shown are bulk electron recoils (red points), surface electron events (black crosses) and nuclear recoils (blue circles) with recoil energy between 10 and 100 keV. Top: Ionization yield versus recoil energy. The solid black lines define bands that are $2\sigma$ from the mean nuclear-recoil yield. The sloping magenta line indicates the ionization energy threshold while the vertical dashed line is the recoil energy analysis threshold. The region enclosed by the black dash–dotted lines defines the sample of events that are used to develop surface-event cuts. Bottom: Normalized ionization yield (number of standard deviations from mean of nuclear recoil band) versus normalized timing parameter (timing relative to acceptance region) is shown for the same data. Events to the right of the vertical red dashed line pass the surface-event rejection cut for this detector. The red box is the WIMP signal region.

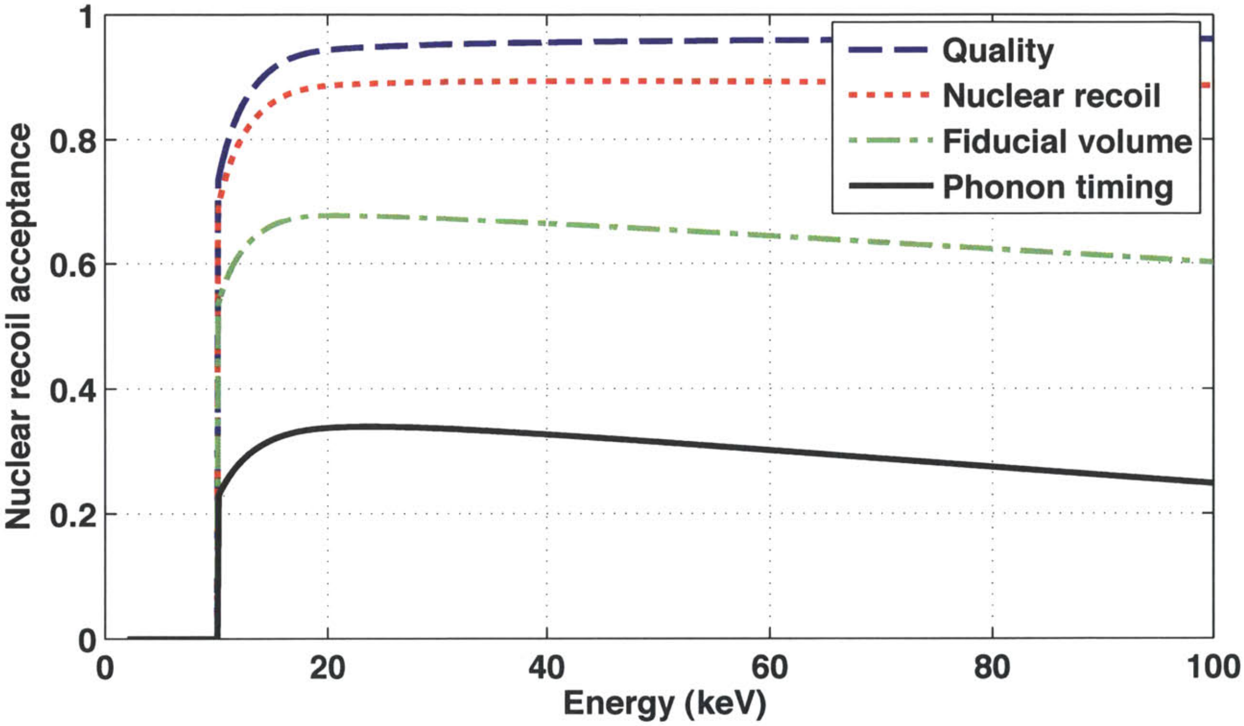

Figure 8-16: Cumulative efficiency as a function of recoil energy after application of each WIMP-selection cut shown. The solid curve shows the overall efficiency of this analysis. The drop in efficiency near 10 keV is due to the ionization threshold and the flaring of the electron recoil band. The slight drop in efficiency at high recoil energies is due to a drop in the fiducial volume.

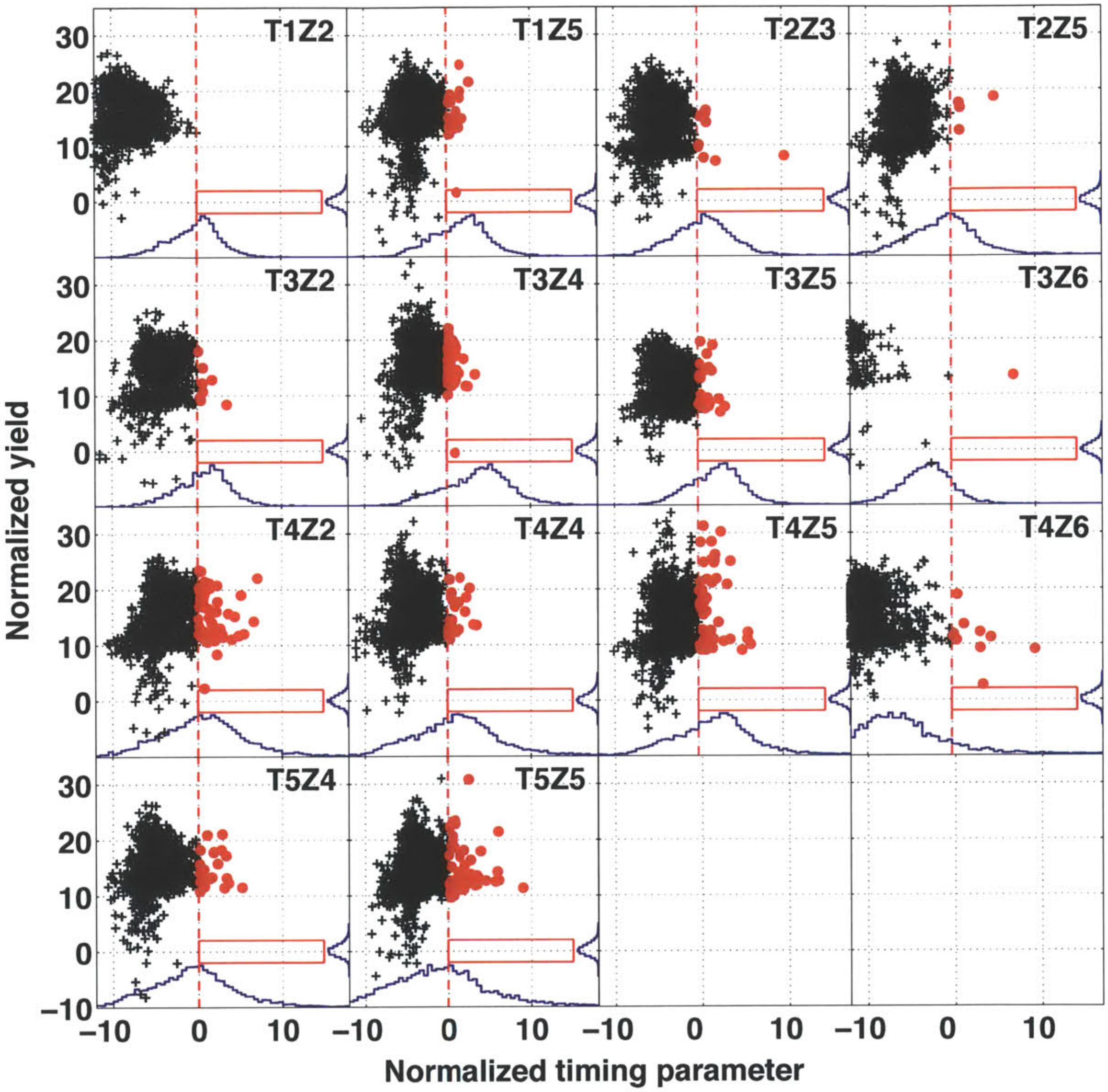

Figure 8-17: Normalized ionization yield (number of standard deviations from mean of nuclear recoil band) versus normalized timing parameter (timing relative to acceptance region) for all events and with all cuts applied except for yield and timing. Each panel shows the data taken with the indicated detector. Data for all detectors that were used in this reported WIMP search are shown. Events that pass the phonon timing cut are shown with round markers. The red boxes indicate the signal region for that detector. The blue histograms shows the expected distributions for nuclear recoils in each detector, as measured by the calibration data.

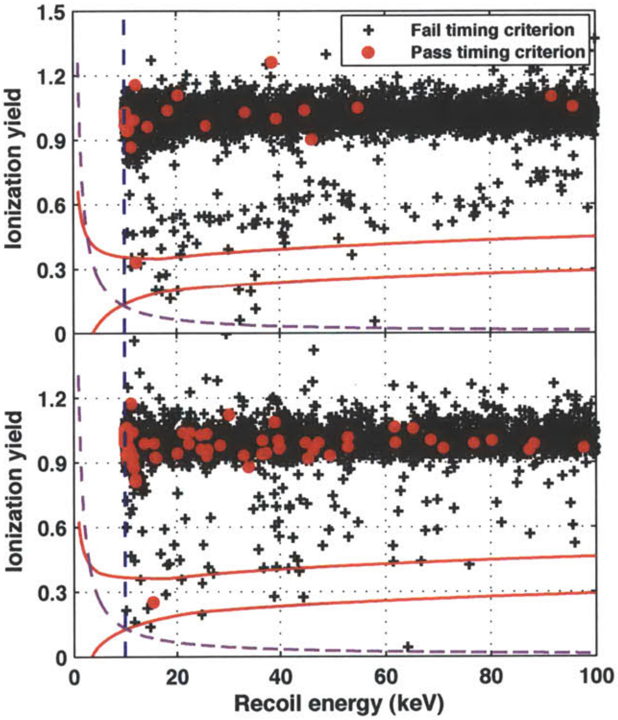

Figure 8-18: Ionization yield versus recoil energy for events consistent with all signal criteria, excluding yield and timing. The top (bottom) plot shows events for detector T1Z5 (T3Z4) (see SOM text for detector nomenclature). The solid red lines indicate the ionization yield acceptance region. The vertical dashed line represents the recoil energy threshold and the sloping magenta dashed line is the ionization threshold. Events with phonon timing characteristics consistent with our selection criteria are shown with round markers. The candidate events are the round markers between the red lines.

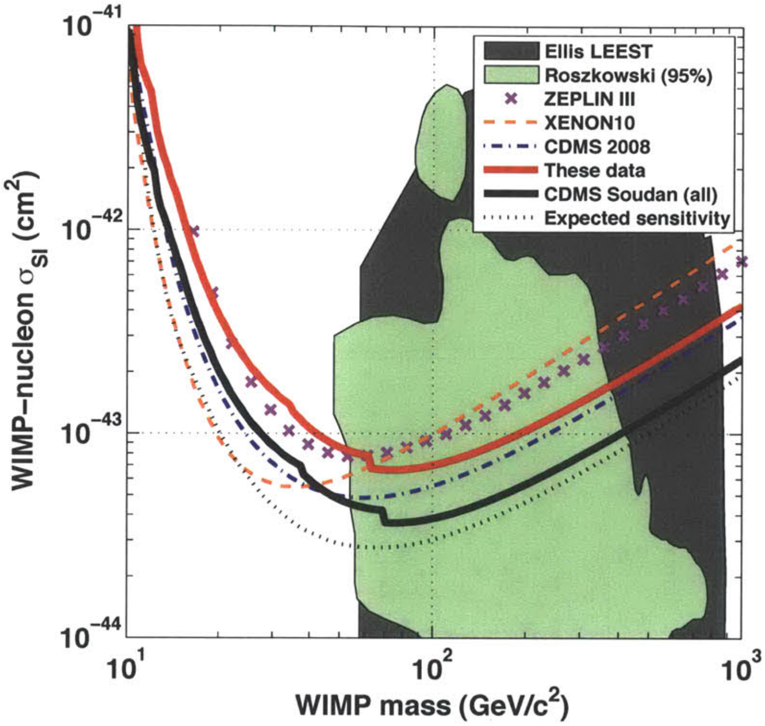

Figure 8-19: Experimental upper limits (90% confidence level) and theoretical allowed regions for the WIMP-nucleon spin-independent cross section as a function of WIMP mass. The red (upper) solid line shows the limit obtained from the exposure analyzed in this work. The solid black line shows the combined limit for the full data set recorded at Soudan. The dotted line indicates the expected sensitivity for this exposure based on our estimated background combined with the observed sensitivity of past Soudan data. Prior results from CDMS [6], XENON10 [17], and ZEPLIN III [73] are shown for comparison. The shaded regions indicate allowed parameter space calculated from certain Minimal Supersymmetric Models [47, 94].

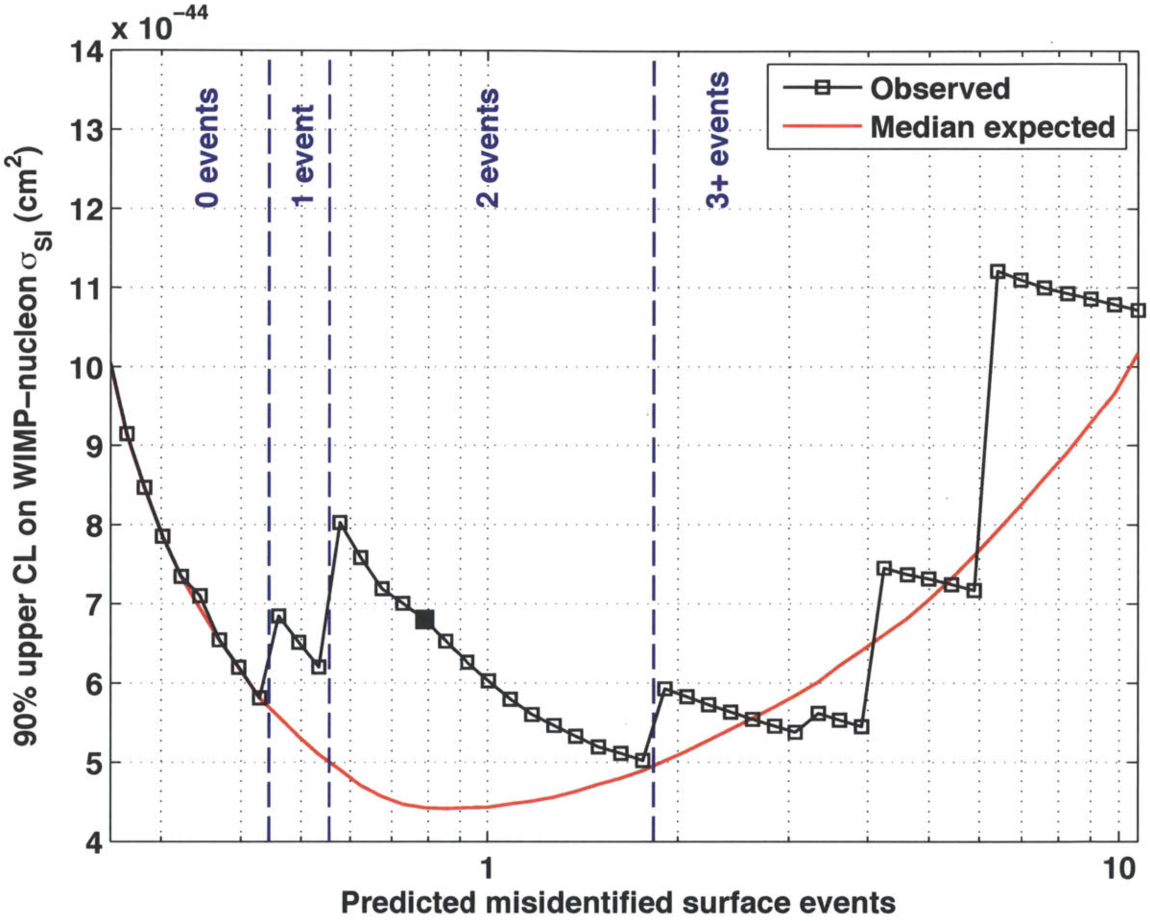

Figure 8-20: The observed 90% C.L. upper limit, at a WIMP mass of 70 GeV/c$^2$, as a function of predicted misidentified surface event background. The solid square shows the observed limit from the choice of timing cut value used in the blind analysis, which corresponds to the indicated misidentified surface event prediction. The open squares show the limits that would have been observed for other choices of timing cut value. The blue dashed lines mark the transitions at which additional background events would appear. The red curve is generated by taking the median of many limits, each of which is obtained by drawing a value from a Poisson distribution with mean value given by the corresponding value of predicted misidentified surface events.

# Chapter 9

# CDMS II at Low Energies

The 10 keV threshold used in the standard analysis of CDMS II is a threshold on the detector's ability to measure subtle timing differences between bulk and surface phonon pulses. Measuring pulse shapes is hard. The charge and phonon *energies*, on the other hand can be well-measured significantly below the 10 keV threshold, thereby giving CDMS II yield discrimination down to ~2 keV. This ~2-10 keV range represented an untapped reserve of exposure with significant (yield-only, no pulse shape) discrimination, and is in fact the *only* place to search in the CDMS II exposure for WIMPs with masses less than ~20 GeV/c$^2$. The standard 10 keV threshold serves well, therefor, for a test of the WIMP models predicted by cMSSM, but is not ideal for a test of the light-WIMP models that have been offered as explanations of the low-energy excesses in CoGeNT and CRESST-II, or the modulation signal seen by DAMA/LIBRA.

David Moore led an effort to analyze this sub-10 keV portion of the CDMS II exposure, building on similar work by Ray Bunker and Walter Ogburn. By eliminating the timing cut from the analysis, we must accept significant backgrounds. On the other hand, we gain significant exposure (remember that the timing cut cut the exposure by almost a half). Because David Moore recently wrote a thesis largely about this low-threshold analysis, we will not try to summarize his work here to any level of detail. We simply point out several aspects of the low-threshold analysis that are relevant to understanding the low-threshold modulation analysis (the topic of the next chapter).

Because the analysis was known in advance to be background limited, only those detectors with the lowest backgrounds were analyzed. In this analysis, backgrounds in the nuclear recoil yield band are largely the result of leakage of events into the band due to poorly-measured charge pulse amplitudes (*i.e.*, the result of charge noise), and so the detectors for this analysis were selected largely on the basis of charge noise amplitude. Phonon noise was also a relevant consideration, in that the trigger threshold of the detector is set by the phonon noise. Eight germanium detectors were selected, and the resulting total exposure (before fiducial volume and yield band cuts) of the low-energy analysis was 241 kg-days.

At high energies, where the largest contribution to yield uncertainty is positional variation in the phonon measurement, it is natural to present events in the yield vs. energy plane. When charge channel noise becomes the dominating contributor to yield uncertainty, the data is easier to interpret when the charge and phonon measurements are entirely separated, as in Figure 9-1, where the axes are charge energy on the y axis and phonon energy on the x axis.

## 9.1 keVee and keVnr

The largest difference in the low-energy analysis is that the event energies are scaled from the phonon energy alone, rather than combining a phonon measurement and a charge measurement to estimate the Luke phonon contribution (and subtract it off to obtain the recoil energy). This change in strategy is intended to improve the recoil energy resolution, which, if the charge signal were used, would be dominated by charge noise (in the Luke phonon estimation). The cost paid to improve the noise is that we must assume a Luke phonon contribution, based on assuming a yield for the event. At low energies, then, we have less information than at high energies, and resort to the $\mathrm{keV_{ee}}$ and $\mathrm{keV_{nr}}$ scales more typically used by single-channel experiments, assuming either the Luke contribution for yield = 1 or the energy-dependent yield of a nuclear recoil, respectively.

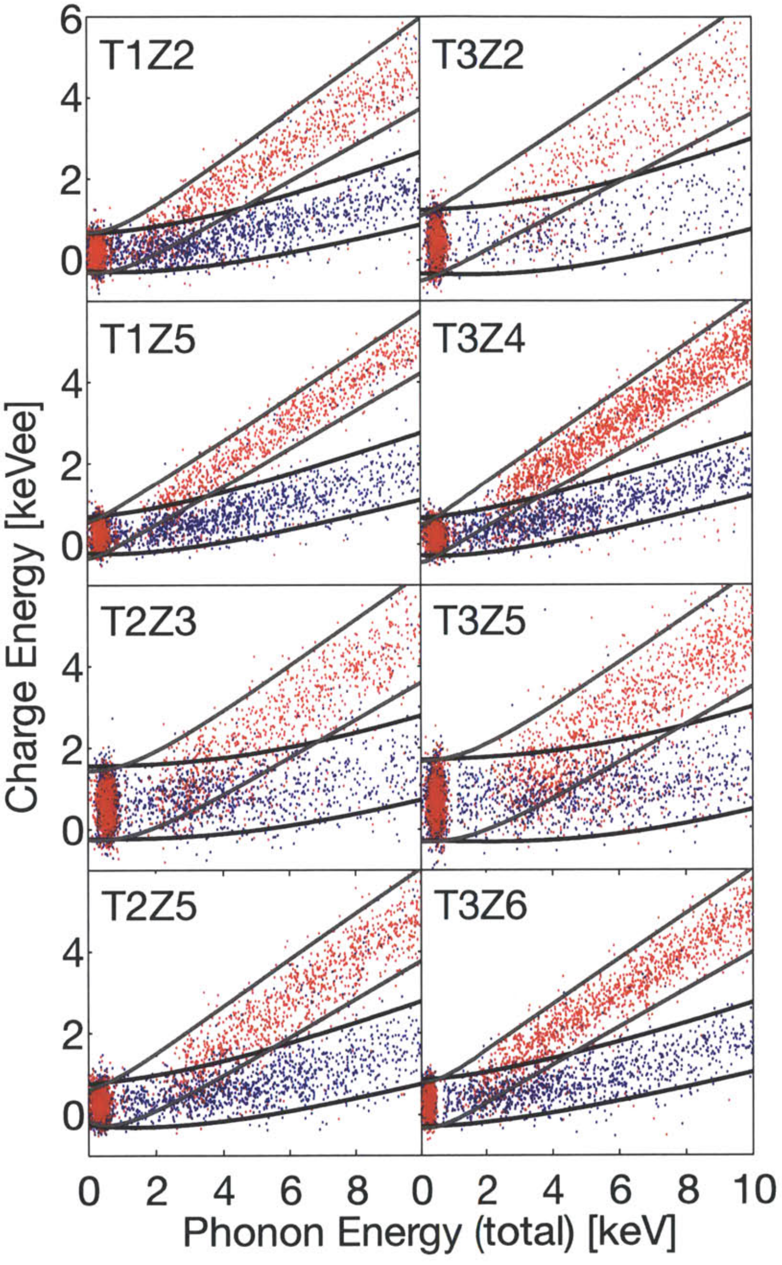

Figure 9-1: The charge energy vs. phonon energy space, with ±$2\sigma$ ER and NR band definitions overlain on electron recoils from $^{133}$Ba calibration (red) and nuclear recoils from $^{252}$Cf calibration (blue). Notice the noise 'blob' surrounding zero, and notice that the phonon energy axis is here in terms of *total* phonon energy (the measured quantity: the sum of the recoil and Luke contributions). Plot from D. Moore.

Compared with single-channel experiments, however, we do have the advantage that we can use both channels *in situ* to determine the correct $\mathrm{keV_{ee}}$ and $\mathrm{keV_{nr}}$ scalings. These scalings (the ionization yields) are difficult to determine event-by-event at low energies, but easy to determine for *distributions* of calibration data. This *measured* mean nuclear recoil yield has escaped the poor charge noise (we have overcome it with statistics) while giving us a *in situ* scaling between the measured total phonon energy and an inferred nuclear recoil energy.

Let's discuss that general strategy in a little more detail. Assuming an event is an electron recoil, the recoil energy, $E_{recoil}$, in units of $\mathrm{keV_{ee}}$ is given by

$$E_{recoil} = \frac{P_{tot}}{2} \tag{9.1}$$

where we have used that

$$E_{recoil}(P_{tot}) = P_{tot} - eV_b N_Q = P_{tot} - \frac{eV_b}{\epsilon} E_Q \tag{9.2}$$

where $eV_b/\epsilon = 1$ for $V_b = 3.0$ V. The charge energy $E_Q$ is calibrated such that the recoil energy for an electron recoil equals the charge energy (and yield has the simple expression $E_Q/E_{recoil}$).

The factor of 1/2 appearing in Equation 9.1 is different for the case of nuclear recoils, because of the reduced Luke contribution. Yield varies with energy for nuclear recoils, so this Luke contribution is similarly energy-dependent, based on measured charge yield for the nuclear recoil population. The mean ionization energy for nuclear recoils, $\mu_{Q,NR}(P_{tot})$ is determined over the energy range from 2-100 keV using the distribution of nuclear recoils in $^{252}$Cf calibration data, as fully described in the thesis of D. Moore. Assuming an event is a nuclear recoil, the measured phonon energy can be mapped to the recoil energy (in units of $\mathrm{keV_{nr}}$) as

$$E_{recoil}(P_{tot}) = P_{tot} - \mu_{Q,NR}(P_{tot}) \tag{9.3}$$

The calibration of both $E_Q$ and $P_{tot}$ at low energies is performed using the $^{71}$Ge activation lines at 1.3 and 10.4 keV. These sharp electron recoil lines both have a half-life of 11.4 days, appearing most strongly after calibration by the $^{252}$Cf neutron source. The calibration of $\mu_{Q,NR}(P_{tot})$ is performed by fits to the nuclear recoil band in the $E_Q$ vs $E_{recoil}$ plane, where $E_{recoil}$ is derived in the standard way, subtracting off the poor-resolution estimation of the Luke contribution. The fits to the nuclear recoil yield bands were parameterized by a power law of the form $\mu_{Q,NR}(E_{recoil}) = AE^B_{recoil}$, specific to each detector.

It should be noted that these fits to the measured CDMS yield bands (upon which our low-energy NR energy scale depends) are in significant tension with the literature, particularly at the lowest energies. The smaller-than-expected $\mu_{Q,NR}(E_{recoil})$ at low energies observed by CDMS could be explained in two basic ways:

**Suppressed charge collection** for nuclear recoils at low energies is the most natural explanation, but would require a charge-suppression mechanism specific to charges produced by a nuclear recoil. It could be imagined that perhaps electron-hole recombination is enhanced in the denser initial charge ball created by nuclear recoils, and that this enhancement is not evident in other experiments due to CDMS's extremely low drift field. This is very speculative idea, yet to be tested. If charge collection is suppressed, then the measured $\mu_{Q,NR}(E_{recoil})$ is the correct one to use for estimating the Luke contribution.

**Over-estimation of the phonon energy** for nuclear recoils at low energies could produce the same yield vs. energy curve. Perhaps the pulse-shape-dependent optimal filter is to blame, although preliminary work indicates this is not the case. It is hard to imagine a way to over-estimate an energy, but if this is in fact what is happening, then any statements we make using the nuclear recoil energy scale would be conservative, in the sense that they would be probing energies actually smaller than the estimated energies (by a factor of 5-20% at 2 keV, depending on detector).

For more details on this, as on all low-energy topics, refer to the recent thesis of D. Moore, and also to an article dedicated to understanding our nuclear recoil energy scale due to be published quite soon.

This disagreement between our measured yield and the expected yield from Lindhard theory results in some unknown systematic uncertainty in the the production of Luke phonons. This uncertainty, however, is not a large practical concern for two reasons:

- The two cases above can be summarized as CDMS either creating an unbiased energy scale or *over*-estimating the energy scale. In other words, the statements made my CDMS at low energies are safe: we are, at worst, making weaker statements about low energies that we might be able to.
- The uncertainties are not in the total energy scale, but in the magnitude of the Luke phonon contribution, which, for nuclear recoils at these low energies, is only ~15% of the total phonon energy. Any uncertainties in the Luke phonon contribution, therefor, are scaled by a similar percentage as applied to the total energy scale.

## 9.2 Analysis of the low-energy spectrum

As described in the thesis of D. Moore, once the NR and ER energy scales are in place, it is a straightforward process to follow the same general procedure as in the standard CDMS II analysis of defining quality cuts, defining charge fiducial volume cuts, defining nuclear recoil yield band cuts (optimized in this case at -0.5$\sigma$ to +1.25$\sigma$ to maximize exposure while minimizing leakage), measuring the total efficiency (of both the analysis cuts and the hardware trigger) and finally to arrive at a sample of low-energy 'candidate events' (illustrated in Figure 9-2) and the detector-by-detector, efficiency-weighted exposure. We can combine the candidates and exposures to create spectra for each individual detector or the combined exposure. We can also combine the candidates and the exposures within the optimum-interval method (as in the standard analysis), and arrive at a 90% confidence upper limits on the WIMP-nucleon cross section as seen in Figure 9-3.

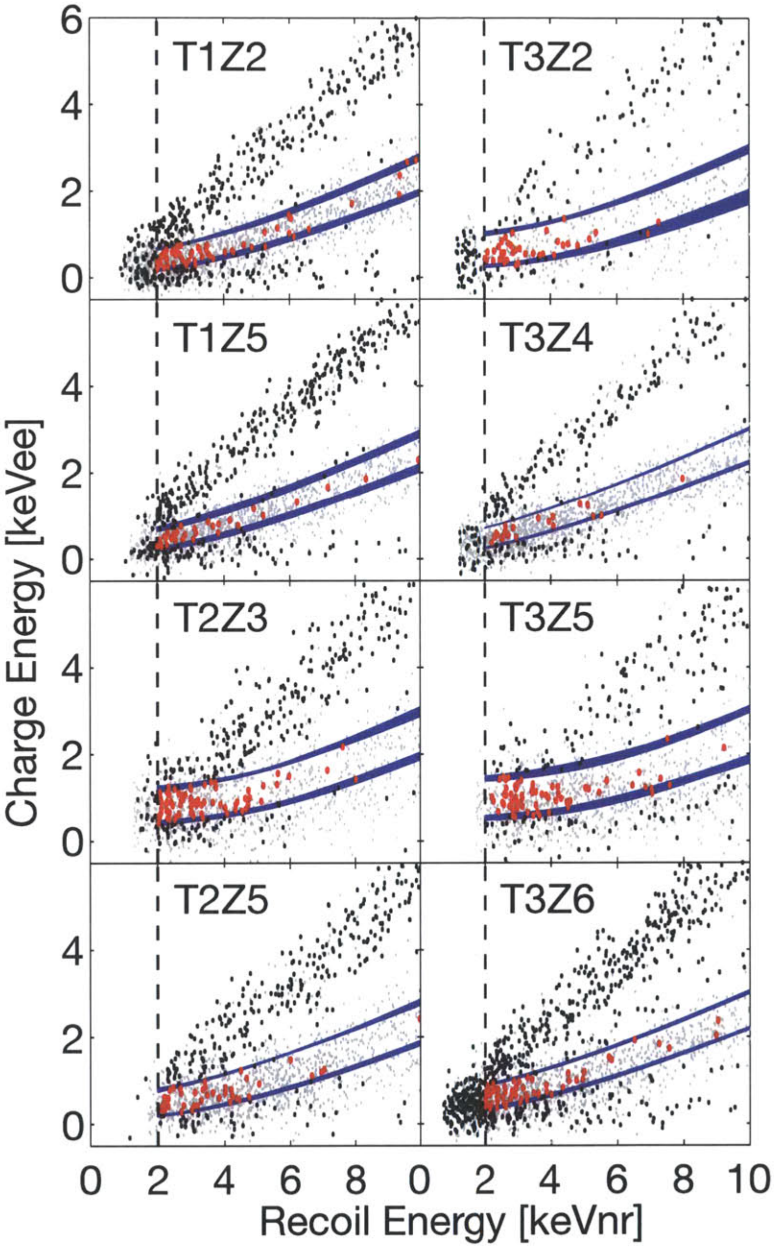

Figure 9-2: The charge energy vs. phonon energy space, with -0.5 to +1.25 $\sigma$ NR band definitions (width indicating range of definitions for various runs) overlain on the low background data of the CDMS II exposure used in the low-energy analysis. The 2 keV analysis threshold is indicated with a dashed line. Notice that the phonon energy axis is here in terms of *recoil* energy, assuming the Luke contribution of a nuclear recoil. Plot from D. Moore.

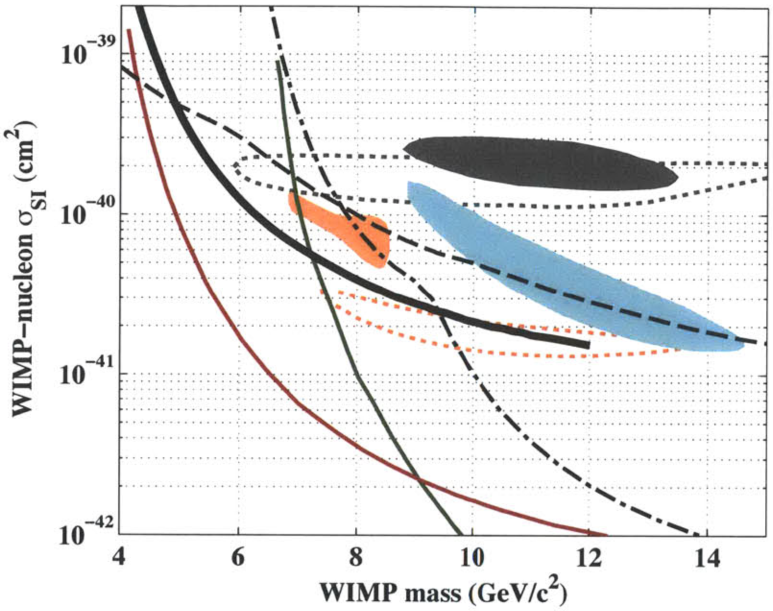

Figure 9-3: The low-mass WIMP region, in cross-section vs mass, highlighting the 90% CL exclusion limit for the CDMS II low-energy analysis (thick black curve), and how it relates to other experimental results in this mass range. Similar exclusion limits from other analyses are also shown: a CDMS II exposure with a 10 keV threshold (dash-dotted), a low-threshold analysis from the shallow site [11] (dashed), XENON100 [18] (green, solid), and a low-threshold analysis of the XENON10 data [15] (red, solid). Allowed regions are shown for CoGeNT [2] (orange, filled), DAMA/LIBRA [98] (gray, filled), and CRESST- II [16] (cyan, filled). An alternative calculation of the CoGeNT allowed region after subtracting the expected surface event background [[67],[42]] (orange, dotted) is also shown, as well as an alternative calculation of the DAMA/LIBRA allowed region which allows for larger quenching factors at low energy [58] (gray, dotted). Plot from D. Moore.

CDMS II is in strong contradiction with CoGeNT, unless (as in [67] and [42]) only some small fraction of the CoGeNT spectrum is WIMPs (the rest of the excess representing some surface background). The two limits from Xenon have some significant uncertainties in this comparison due to uncertainties in the nuclear recoil scintillation yield at low energies, and perhaps some target-dependent behaviors could make a hypothetical WIMP spectrum appear differently in CoGeNT (Ge), DAMA/LIBRA (Na and I), and CRESST (O and W), but it is difficult to reconcile the difference between the rates seen in CoGeNT and CDMS II, due to their identical target materials (unless, as previously stated, only some small fraction of the CoGeNT spectrum is from WIMP recoils).

# Chapter 10

# Search for Annual Modulation in CDMS II

## 10.1 Motivation

Possible experimental signals from DAMA/LIBRA [28, 29], CoGeNT [1, 2] and CRESST-II [16] have led to significant recent interest in WIMPs with masses ~10 GeV/$c^2$ and spin-independent WIMP-nucleon cross sections ~$10^{-41}$–$10^{-40}$ cm$^2$ (*e.g.* [59, 56, 70, 24]).

The WIMP scattering rate is expected to annually modulate due to the relative motion of the earth through the local dark-matter halo [46]. The presence of an annually modulating component in the observed interaction rate can identify a WIMP signal in the presence of significant unmodulated backgrounds. This modulation signature is especially useful for WIMPs with masses ~10 GeV/$c^2$ which would primarily produce recoils with energies just above the detection threshold, where the rejection of backgrounds that can mimic a WIMP signal is less powerful. Both the DAMA/LIBRA [28, 29] and CoGeNT [1, 2] experiments claim evidence for such a modulating signal in their data. The CoGeNT and DAMA allowed regions are illustrated in Figure 10-3, as well as several exclusion limits.

If, as recently suggested, only a small fraction of the low-energy excess events in CoGeNT are due to WIMPs, then constraints from CDMS II may be avoided [59, 42].

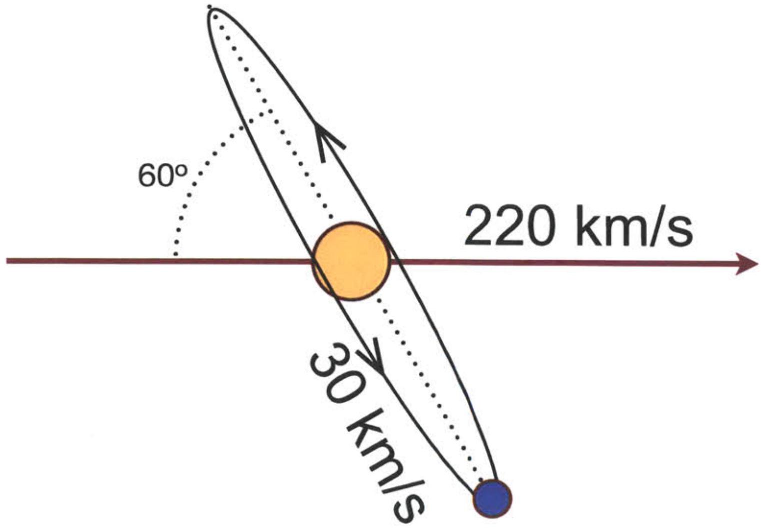

Figure 10-1: The relevant geometry and velocities relevant to the annual modulation effect. In this cartoon, earth is in the position it would be in September, a phase of 243.5 days.

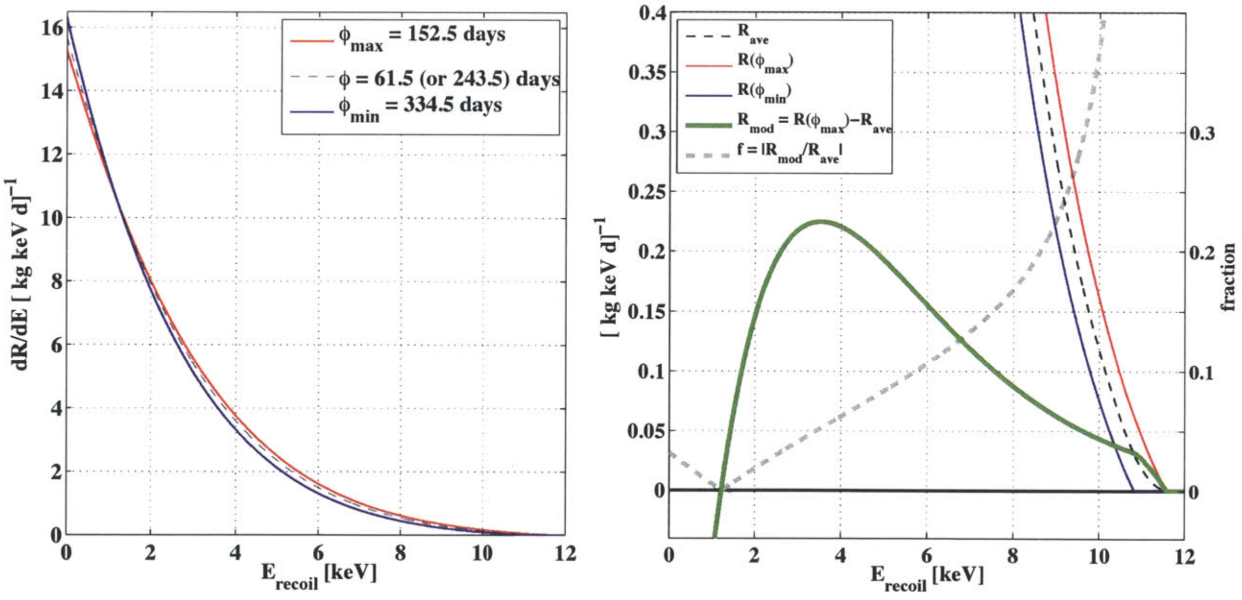

Figure 10-2: Two views of the annual modulation and its effect on a hypothetical spectrum. We assume a generic a generic light-WIMP model, with $m_\chi$ = 10 GeV/c$^2$ and spin-independent nucleon cross section $\sigma_{SI} = 1\times10^{-40}$ cm$^2$. The first view shows the expected spectrum at the maximum (red) and minumum (blue) phase times of the year (as well as the yearly average spectrum, in black dashed). Note the point at ~1 keV at which the phase of the modulation flips sign. In the lower plot, we show the same total spectrum as above, but also the amplitude of the modulation only (green). We also show the fractional modulation (grey), which approaches 100% at the highest-energy portion of the spectrum.

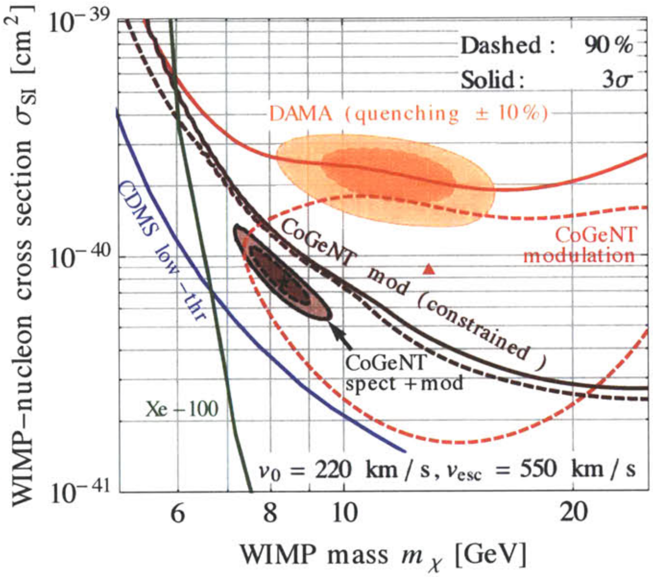

Figure 10-3: A plot from Fox*et al.* [56], showing the strong tension between the limits set by the CDMS II low-threshold analysis (blue) and the Xe-100 standard analysis (green) with the CoGeNT allowed region (solid red) and the DAMA/LIBRA allowed region (solid orange). A CoGeNT allowed region drawn using only the CoGeNT modulation is also shown (red dashed). All of these allowed regions and limits assume a spin-independent dark matter-nucleon scattering process, and a standard Maxwell-Boltzmann halo with escape velocity $v_{esc} = 550$ km/s and velocity dispersion $v_0 = 220$ km/s. Notice that the CoGeNT modulation alone prefers a higher-mass WIMP than the CoGeNT spectrum. (The spectral information dominates the combined spectrum+modulation allowed region, shown.)

In this case, if CoGeNT's annual modulation is due to WIMPs, the fractional variation is several times larger than expected for a "standard" halo with a Maxwellian velocity distribution [59, 67]. In addition, the energy spectrum of the modulation extends to higher energies than expected for a standard halo [56, 67, 99, 51]. Such large modulation fractions and hard spectra might be possible if the halo exhibits (non-standard) local substructure [59, 67, 56]. To test such a scenario, this chapter searches for a corresponding annual modulation in the CDMS II germanium data. This analysis does not cover the full energy range of the CoGeNT modulation, restricting itself to energies above 5 $\mathrm{keV_{nr}}$ (which, due to quenching [77, 76], corresponds to 1.2 $\mathrm{keV_{ee}}$ in the standard CoGeNT energy scale [41]). Because germanium serves as the target material for both CDMS II and CoGeNT, these results provide a check of whether the reported modulating signal is due to WIMPs that is less model-dependent than recent results from XENON10 [15] and XENON100 [18].

## 10.2 The CDMS II dataset

The data analyzed here were collected over nearly two annual cycles, from October 2006 to September 2008, using all 30 Z-sensitive Ionization and Phonon (ZIP) detectors installed at the Soudan Underground Laboratory [9, 6, 7]. Data-quality and detector selection criteria are identical to the previous analysis of the low-energy CDMS II nuclear-recoil spectrum in [8]. Only the 8 germanium detectors with the lowest trigger thresholds were used to search for WIMP interactions, while all 30 detectors were used to veto events with interactions in multiple detectors.

Following [8], the nuclear-recoil energy scale is based on the phonon measurement, which is corrected by ~20% to take into account the fraction of the total phonon signal arising from the Neganov-Luke phonons [86, 79] generated by the charge-carrier drift across the detectors. The Neganov-Luke phonon contribution for nuclear recoils in $^{252}$Cf calibration data was directly measured for the recoil-energy range of this analysis. As in [8], the phonon energy scale for electron recoils was conservatively calibrated, ensuring to the 90% C.L. that the 1.3 and 10.4 keV activation-line energies were not underestimated.

The maximum energy considered in this analysis was 11.9 $\mathrm{keV_{nr}}$, matching the highest energy observable by the CoGeNT "LG" (low gain) channel [41]. Because time-dependent variations in the CDMS trigger thresholds could mimic or hide a modulation in the event rate, the energy threshold for this analysis was conservatively chosen to be 5 $\mathrm{keV_{nr}}$, high enough that events are triggered with essentially perfect efficiency. To avoid bias, trigger efficiency was measured throughout the exposure, using events for which at least one other detector triggered. For the 5–11.9 $\mathrm{keV_{nr}}$ energy range, combining all detectors and all time bins yielded 4350 events in this unbiased sample, only 3 of which failed to trigger. These missed triggers were each in a different detector and were uniformly spaced throughout the considered energy range.

Because CDMS II uses a phonon-based energy scale (at these low energies), and CoGeNT uses an ionization-based energy scale, quenching causes the two experiments to exhibit different mappings between energies assuming nuclear recoils (such as the energy range of this analysis, 5.0–11.9 $\mathrm{keV_{nr}}$) and energies assuming electron recoils. For electron recoils with the same total phonon signal in the CDMS II experiment, the equivalent recoil-energy interval is 3.0–7.4 $\mathrm{keV_{ee}}$, due to the larger Neganov-Luke phonon contribution. Analogously, for electron recoils with the same total *ionization* signal in the CoGeNT experiment, the equivalent recoil-energy interval is 1.2–3.2 $\mathrm{keV_{ee}}$, where we apply CoGeNT's measured ionization yield for nuclear recoils [41, 19].

## 10.3 Event selection

Detector stability was monitored throughout data taking with quality cuts, removing periods of abnormal detector performance [6, 7]. For consistency with previous work, we followed [8] and removed data taken during the 20 days following exposure of the detectors to a neutron calibration source. The CDMS WIMP-search data are not continuous over the nearly two years of exposure considered here, but include gaps due to neutron calibrations, warming of the detectors, and data periods removed due to data-quality diagnostics. Since data-quality diagnostics remove individual detectors, different detectors serve as WIMP-search detectors for different times. Figure 10-4 shows the history of exposure for the period considered on a detector-by-detector basis. The detectors were arranged in five "towers," and are identified by their tower number (T1–T5) and by their (top-to-bottom) ordering within the tower (Z1–Z6).

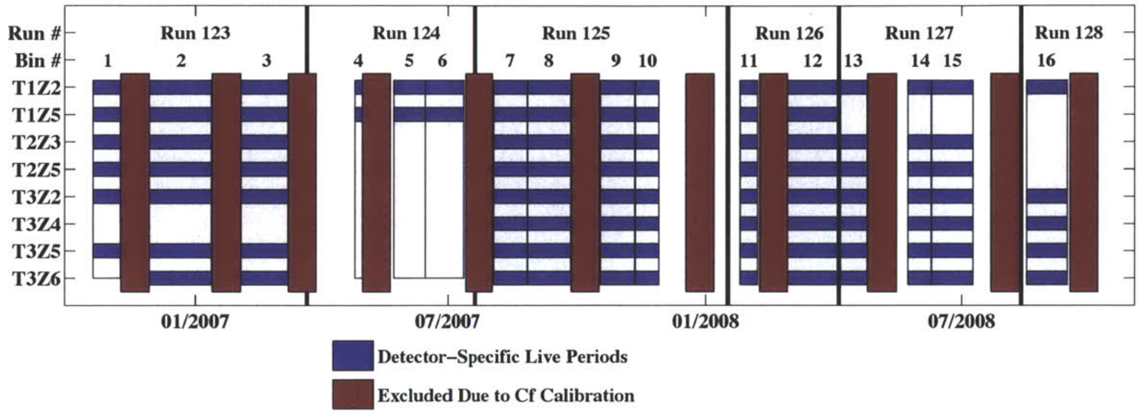

Figure 10-4: The CDMS II exposure, displaying the detectors and time bins used in this analysis. For each time bin, a detector is colored blue if this detector's data was used in this analysis. Divisions between "runs" represent at least partial warm-ups of the dilution refrigerator used to cool the detectors. In order to avoid the effects of Ge activation, 20-day periods were omitted (red) following each $^{252}$Cf calibration time.

After removing these time periods, a total of 241 kg days raw exposure were considered, as in the previously published low-energy analysis [8]. To allow checks for stability to be applied to multiple-scatter events, an additional cut was introduced to eliminate electronics "glitch" events, for which phonon pulses were detected above threshold in more than 15 detectors simultaneously.

Events inconsistent with WIMP interactions were rejected. Since modulation of data-selection cut efficiencies could mimic or hide a modulation in the event rate, selection criteria were designed to have constant acceptance with time, and any residual modulation in the cut efficiencies was constrained using events sampled throughout the data taking period. Since WIMPs have a negligibly small probability of interacting more than once in the apparatus, events with energy deposited in more than a single detector ("singles cut") or in the active scintillator veto ("veto cut") were removed. The glitch cut, veto cut and singles cut have a combined efficiency >97%, with negligible time-dependent variation.

Events were further required to have ionization signals consistent with noise in the outer charge electrode of the detector ("$Q$-inner cut"). To search for a nuclear-recoil signal of WIMP origin, the ionization energy was required to be within $\pm 2\sigma$ of the mean of the energy-dependent nuclear-recoil distribution from calibration data ("nuclear-recoil cut") in the main analysis described below. This ionization-based selection increases the sensitivity of this analysis to a modulating signal relative to the more restrictive ionization-based selection used in [8], provided that backgrounds do not modulate. These wider nuclear recoil bands, and candidate events selected for this analysis, are illustrated in Figure 10-5. To explore different physics or instrumental origins of a potential signal, we also applied our modulation analysis to two additional event samples consisting of either single-scatter or multiple-scatter events with no ionization-based nuclear-recoil cut. The quality, glitch, veto and $Q$-inner cuts were always maintained.

## 10.4 Statistical Treatment

These various cuts result in an efficiency $\varepsilon\left(t, E, d\right)$ that depends on the time $t$, the deposited energy $E$ and the detector $d$. With notations following the descriptions above, the total efficiency for our primary "WIMP-candidate" sample can be written as

$$\begin{aligned}\varepsilon\left(t, E, d\right) =& \varepsilon_{\text{glitch}}\varepsilon_{\text{trigger}}\varepsilon_{\text{singles}}\left(d\right)\varepsilon_{\text{veto}}\times \\ & \varepsilon_{\text{Qinner}}\left(t, E, d\right)\varepsilon_{\text{NuclearRecoil}}\left(t, E, d\right),\end{aligned} \tag{10.1}$$

where we have explicitly identified the dependence on time $t$, energy $E$ and detector $d$ for each of the cuts. For the event samples that remove the nuclear-recoil and singles cuts, the corresponding efficiencies, $\varepsilon_{\text{singles}}$ and $\varepsilon_{\text{NuclearRecoil}}$, are ignored.

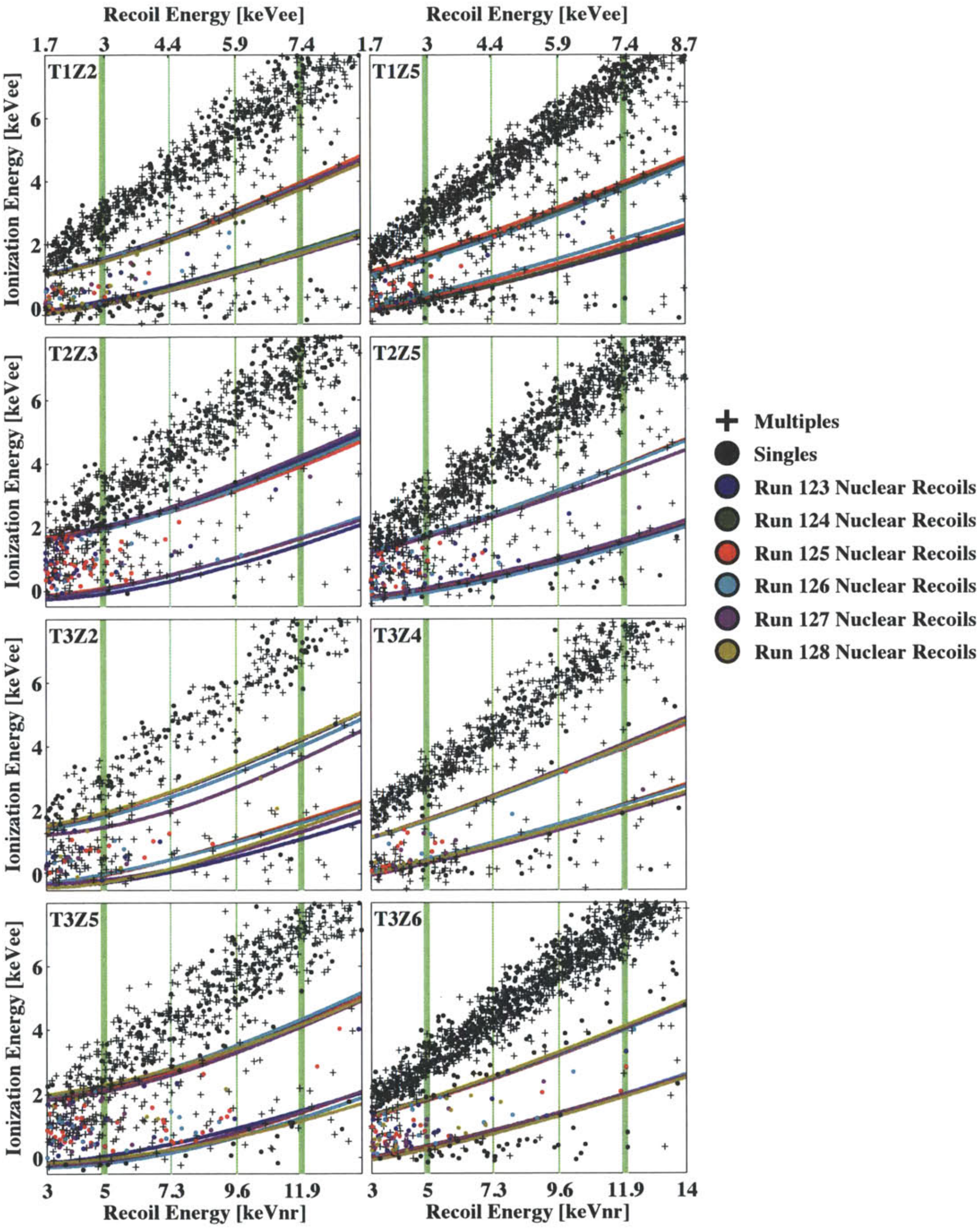

Figure 10-5: Event selection for the eight detectors employed by this analysis, shown in an ionization-energy vs. phonon-energy plane. The energy range of this analysis is indicated by thick green lines, along with borders between smaller energy bins. In this and following plots, the 5–11.9 $\mathrm{keV_{nr}}$ range has been subdivided into three equal parts: [5–7.3 $\mathrm{keV_{nr}}$], [7.3–9.6 $\mathrm{keV_{nr}}$], and [9.6–11.9 $\mathrm{keV_{nr}}$]. All events shown have passed data-quality cuts, as well as the veto cut and the $Q$-inner cut. Crosses represent events registering in multiple detectors ("multiples"), filled markers represent events registering in only one detector ("singles"), and colored markers represent singles lying within the $\pm 2\sigma$ nuclear-recoil bands, defined through $^{252}$Cf calibration independently for each detector and run. The edges of these run-by-run nuclear-recoil band definitions are also indicated.

In the lower portion (5–7.3 $\mathrm{keV_{nr}}$) of the energy range considered, the rate of nuclear-recoil candidate events measured in this analysis is $0.28\pm0.03\,[\mathrm{keV_{nr}\ kg\ day}]^{-1}$, while the maximum-likelihood estimate for the CoGeNT modulation amplitude is $0.35\,[\mathrm{keV_{nr}\ kg\ day}]^{-1}$. The corresponding numbers for the entire energy interval considered are $0.15\pm0.01\,[\mathrm{keV_{nr}\ kg\ day}]^{-1}$ for CDMS and $0.16\,[\mathrm{keV_{nr}\ kg\ day}]^{-1}$ for CoGeNT. In both cases, a modulation of the magnitude observed by CoGeNT would require a modulation fraction in CDMS of ~100%.

We test whether the $Q$-inner and nuclear-recoil cut efficiencies are sufficiently constant using calibration data collected throughout the time period used in the analysis. For each cut, for a given time interval $\gamma$, and detector $d$, we measure $P_{\gamma d}$ events passing the cut and $F_{\gamma d}$ failing. Note that the time intervals of the efficiency data are not coincident with the low-background time intervals, but are suitably distributed over the whole data-taking period. We then maximize the likelihood appropriate for a binomial distribution

$$l = \prod_{\gamma d} \varepsilon_{\gamma d}^{P_{\gamma d}} \left(1 - \varepsilon_{\gamma d}\right)^{F_{\gamma d}}. \tag{10.2}$$

The efficiencies $\varepsilon_{\gamma d}$ are written as

$$\varepsilon_{\gamma d} = \varepsilon_d \left\{1 + A \cos\left[\omega\left(t_\gamma - \phi\right)\right]\right\}, \tag{10.3}$$

where $\omega = 2\pi/365.24\,\mathrm{day}^{-1}$. For the chosen cut and energy interval we fit for the detector-dependent unmodulated efficiency $\varepsilon_d$, the detector-independent relative modulation amplitude $A$, and the phase $\phi$ (measured from Jan. 1st), while requiring the efficiency to be $\leq 1$. We generate $10^4$ artificial realizations of the model under consideration, and determine the confidence regions for the modulation and phase using the Feldman-Cousins method [53]. This analysis indicates that the maximum efficiency modulation allowed by our experimental measurement of the nuclear-recoil cut efficiency is 1.2% at the 90% C.L. In the case of the $Q$-inner cut, this upper limit is 2.3%.

In order to estimate the modulation in the observed event rate, we bin events into 16 time intervals, labeled by $\beta$, of ~25 days each. We denote the center of each time bin as $t_\beta$ and its width as $\Delta t_\beta$. The number of events observed in the time interval $\beta$ in detector $d$ is $n_{\beta d}$. We construct a likelihood using the expected Poisson distribution for the $n_{\beta d}$

$$\ell = \prod_{\beta,d} e^{-\mu_{\beta d}} \left(\mu_{\beta d}\right)^{n_{\beta d}}, \tag{10.4}$$

where factorial terms have been omitted for convenience. In this equation, $\mu_{\beta d}$ is the expected number of events

$$\mu_{\beta d} = \left\{\Gamma_d + M \cos\left[\omega\left(t_\beta - \phi\right)\right]\right\} m_d \varepsilon_{\beta d} f_{\beta d} \Delta t_\beta \Delta E, \tag{10.5}$$

where $\Gamma_d$ is the unmodulated rate in detector $d$, $M$ is the modulation amplitude, $m_d$ is the mass of detector $d$, $\varepsilon_{\beta d}$ is the appropriate efficiency using Eq. 10.1, $f_{\beta d}$ is the live-time fraction appropriate for the detector and time interval, $\Delta t_\beta$ is the time interval width, and $\Delta E$ is the energy interval width.

Since the efficiency modulation allowed by the fits to Eq. 10.3 is much smaller than the physics effect we are testing, we need not add an additional term $\ell_{\text{eff}}\left(\varepsilon_{\beta d}\right)$ in the likelihood, which would take into account such uncertainties in Eq. 10.3.

Figure 10-7 shows residual rates for WIMP candidate events, after subtracting the best-fit constant rates $\Gamma_d$ (found with modulated rate $M$ fixed at 0). Using a Feldman-Cousins approach, we test modulation models $[M, \phi]$ on the WIMP candidates, which consist of all events satisfying the data-selection cuts described above.

## 10.5 Conclusions

Figure 10-8 shows that our observed WIMP-candidate event rate is consistent with a constant value. All modulated rates in this energy range with amplitudes greater than 0.06 $[\text{keV}_{\text{nr}}\ \text{kg day}]^{-1}$ are excluded at the 99% C.L.

For comparison, a similar analysis was carried out using the publicly available CoGeNT data [41]. Our analysis of CoGeNT data is consistent with previously published

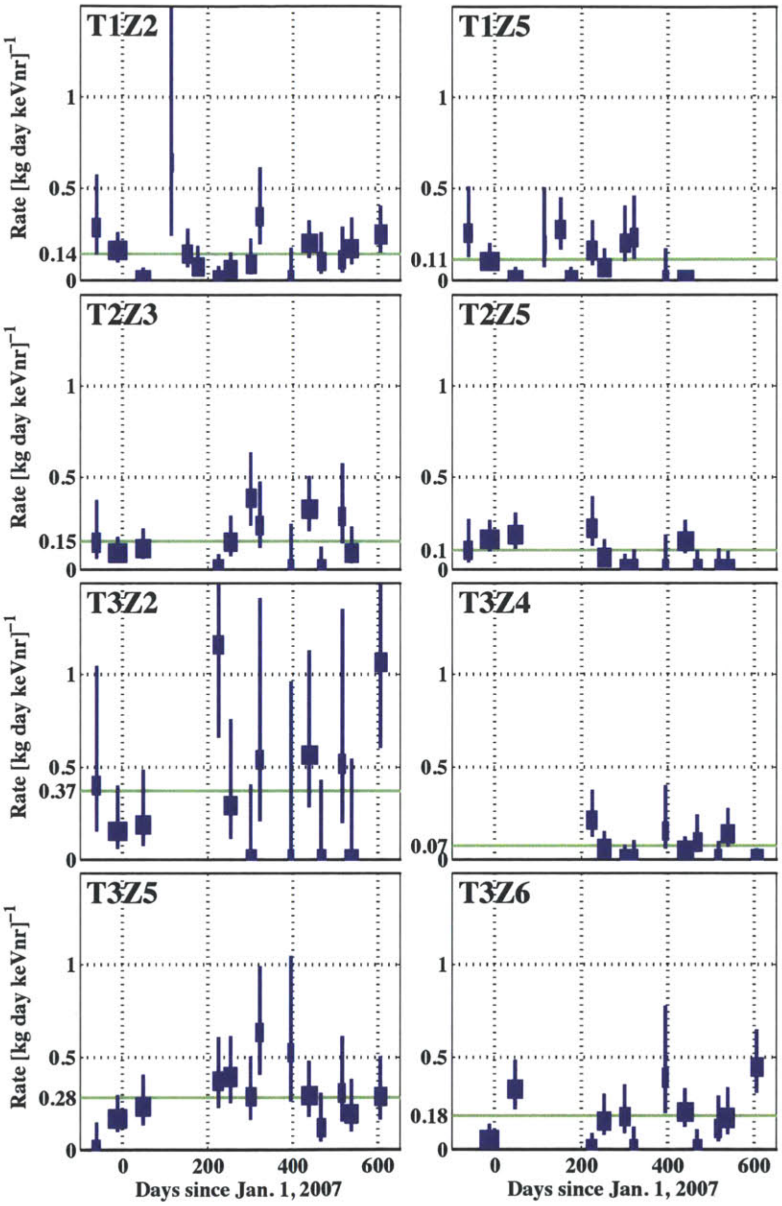

Figure 10-6: Event rate (nuclear recoil band singles, 5–11.9 $\mathrm{keV_{nr}}$) as a function of time for each of the eight detectors used in this analysis. The best-fit rates for each detector ($\Gamma_d$), assuming zero modulation amplitude ($M$), are indicated in green.

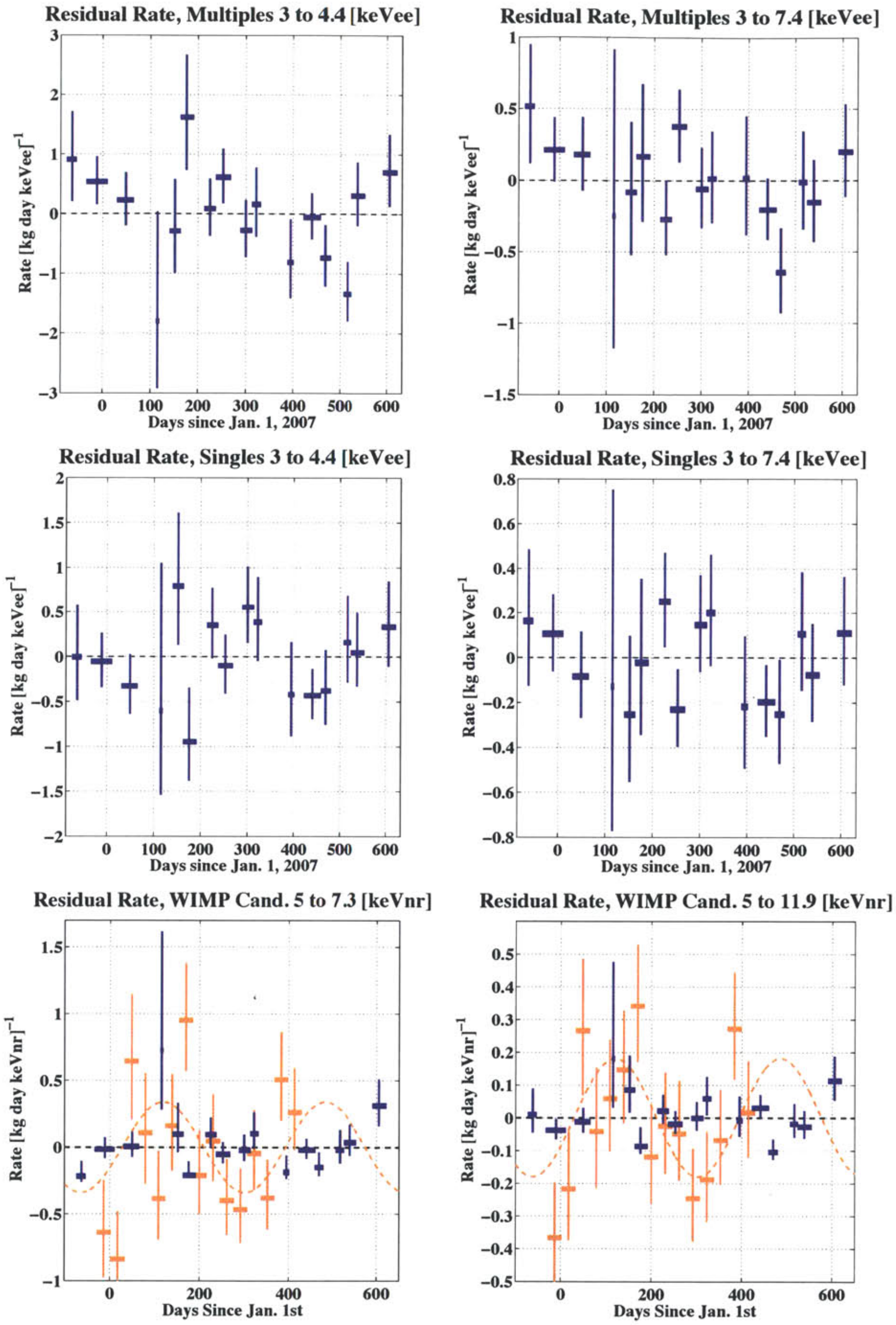

Figure 10-7: Residual event rate as a function of time, for three event populations: multiples (top), singles (middle), and singles in the nuclear-recoil band (bottom), as defined in the text. Two energy ranges are shown, in the left and right columns. Because the multiples and singles populations are dominated by electron recoils, an electron-recoil energy scale has been used for these rates. At the bottom, the rate of CDMS II nuclear-recoil band events is shown for the 5.0–11.9 keV$_{\text{nr}}$ interval (dark blue), after subtracting the best-fit unmodulated rate, $\Gamma_d$, for each detector. The horizontal bars represent the time bin extents, the vertical bars show $\pm 1\sigma$ statistical uncertainties (note that one CDMS II time bin is of extremely short duration). The CoGeNT rates (assuming a nuclear-recoil energy scale) and maximum-likelihood modulation model in this energy range (light orange) are shown for comparison. The CDMS exposure starts in late 2007, while the CoGeNT exposure starts in late 2009.

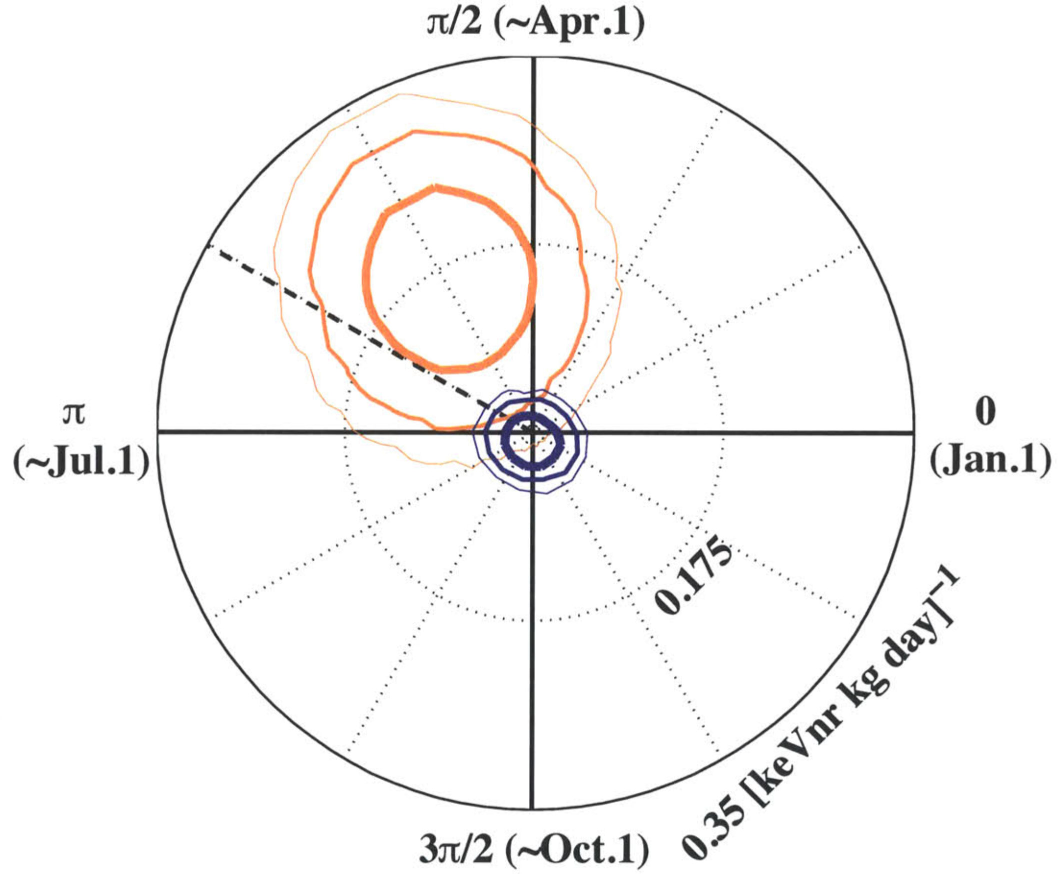

Figure 10-8: Allowed regions for annual modulation of CoGeNT (light orange) and the CDMS II nuclear-recoil sample (dark blue), for the 5.0–11.9 keV$_{\mathrm{nr}}$ interval. In this and the following polar plot, a phase of 0 corresponds to January 1st, the phase of a modulation signal predicted by generic halo models (152.5 days) is highlighted by a dashed line, and 68% (thickest), 95%, and 99% (thinnest) C.L. contours are shown.

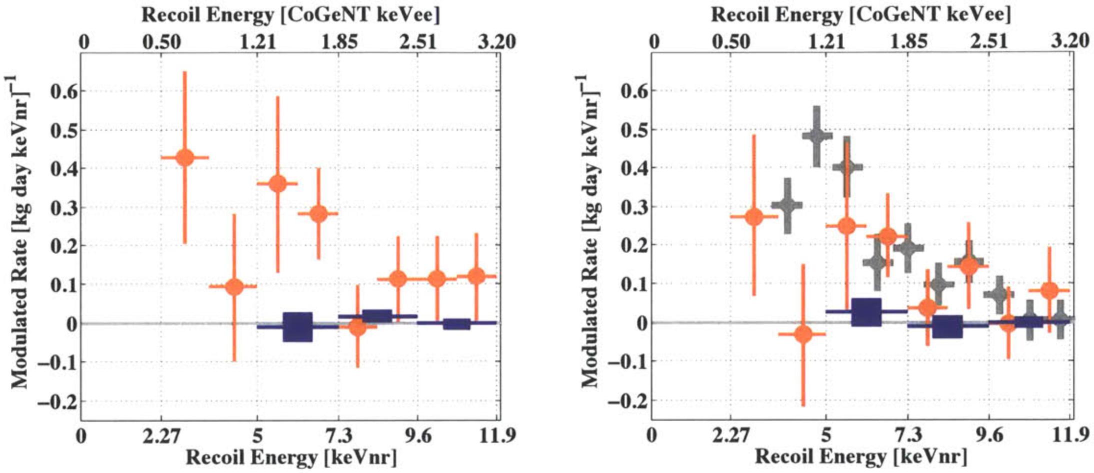

Figure 10-9: Amplitude of modulation vs. energy, showing maximum-likelihood fits where the phase has been fixed and the modulated rates $M$ have been determined for both CoGeNT (light orange circles, vertical bars denoting the 68% confidence intervals) and CDMS (dark blue rectangles, with vertical height denoting the 68% confidence intervals). The phase that best fits CoGeNT (106 days) over the full CoGeNT energy range is shown on the left; the phase expected from interactions with a generic WIMP halo (152.5 days) is shown on the right. The upper horizontal scales show the electron-recoil-equivalent energy scale for CoGeNT events. The 5–11.9 $\mathrm{keV_{nr}}$ energy range over which this analysis overlaps with the low-energy channel of CoGeNT has been divided into 3 equal-sized bins (CDMS) and 6 equal-sized bins (CoGeNT). In the right plot, we also show the DAMA modulation spectrum (small grey circles), following the method of Fox *et al.* [55], for which we must assume both a WIMP mass (here, m$\chi$=10 GeV/c$^2$) and a Na quenching factor (here, $q_{Na} = 0.3$). Lower WIMP masses or higher quenching factors can push the DAMA modulated spectrum towards significantly lower energies. No attempt has been made to adjust for varying energy resolutions between the experiments.

analyses [59, 56, 67]. The nuclear recoil energy scale followed the prescription [1] of the CoGeNT collaboration to relate the ionization energy (calibrated using electron-recoils) to the nuclear recoil energy

$$E_{\mathrm{Q}} = (0.19935) E_{\mathrm{NR}}^{(1.1204)}, \tag{10.6}$$

good over the energy range 0.2 keV$_{\mathrm{nr}}$ $< E_{\mathrm{NR}} \lesssim$10 keV$_{\mathrm{nr}}$.

Figure 10-9 shows the modulated spectrum of both CDMS II and CoGeNT, assuming the phase (106 days) which best fits the CoGeNT data over the full CoGeNT energy range. Compatibility between the annual modulation signal of CoGeNT and the absence of a significant signal in CDMS is determined by a likelihood-ratio test, which involves calculating $\lambda \equiv \mathcal{L}_0/\mathcal{L}_1$, where $\mathcal{L}_0$ is the combined maximum likelihood of the CoGeNT and CDMS data assuming both arise from the same simultaneous best-fit values of $M$ and $\phi$, while $\mathcal{L}_1$ is the product of the maximum likelihoods when the best-fit values are determined for each dataset individually. The probability distribution function of $-2\ln\lambda$ was mapped using simulation, and agreed with the $\chi^2$ distribution with two degrees of freedom, as expected in the asymptotic limit of large statistics and away from physical boundaries. The simulation found only 82 of the $5\times10^3$ trials had a likelihood ratio more extreme than was observed for the two experiments, confirming the asymptotic limit computation which indicated 98.3% C.L. incompatibility between the annual-modulation signals of CoGeNT and CDMS for the 5.0–11.9 keV$_{\mathrm{nr}}$ interval.

## 10.6 Testing other event populations

We extend this analysis by applying the same method to CDMS II single-scatter and multiple-scatter events without applying the ionization-based nuclear-recoil cut. These samples are both dominated by electron recoils. Figure 10-10 shows the confidence intervals for the allowed modulation amplitudes and phases for these two samples, both of which are consistent with no modulation. For the energy range cho-

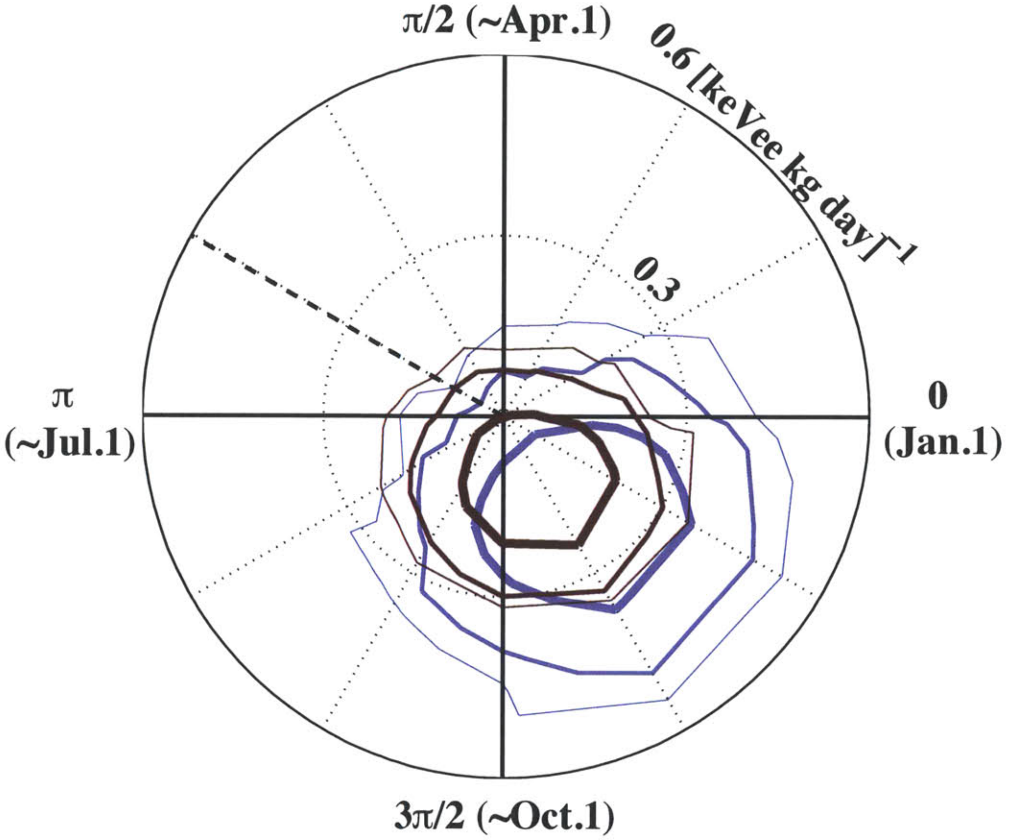

Figure 10-10: Confidence limits on the amplitude and phase of annual modulation for two electron-recoil-dominated data samples: multiple scatters (light blue) and single scatters (dark red), as defined in the text for the interval 3.0–7.4 $\mathrm{keV_{ee}}$. These events are of the same total phonon energy (recoil + Neganov-Luke) as the nuclear-recoil band events of the main modulation analysis shown in Fig. 10-8, of 5.0–11.9 $\mathrm{keV_{nr}}$.

sen for this analysis, there is not significant overlap with the corresponding CoGeNT energy range under the hypothesis of an electron-recoil modulation. Our minimum electron-equivalent energy is 3 $\mathrm{keV_{ee}}$ compared to a 3.2 $\mathrm{keV_{ee}}$ maximum energy for the CoGeNT low-energy channel. Consequently, this analysis cannot exclude the possibility of the modulation observed by CoGeNT being the result of electron recoils. The absence of modulation in the single-scatter and multiple-scatter events indicates the absence of strong systematic effects in our data.

## 10.7 Results for smaller energy ranges

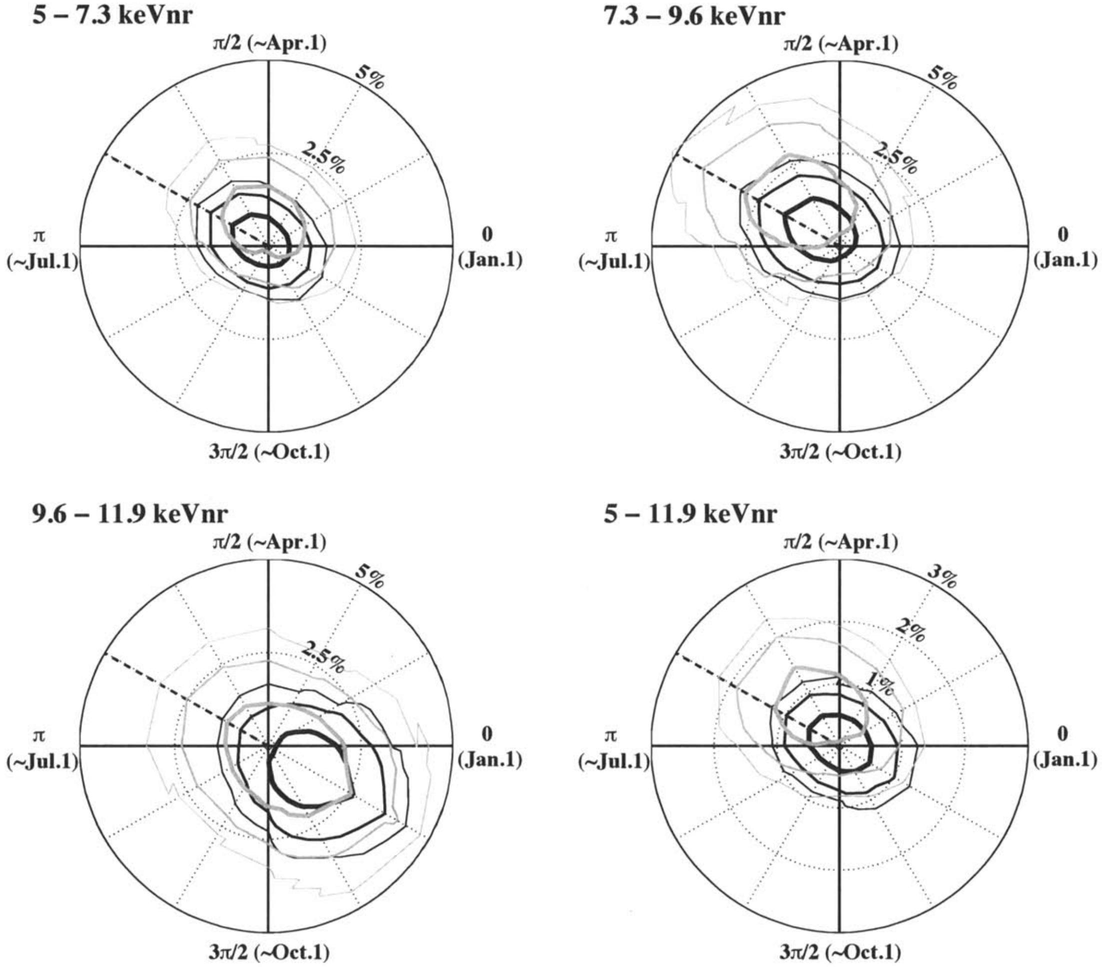

Figure 10-11: Confidence limits on the relative amplitude and phase of annual modulation in the nuclear-recoil band cut efficiency (black) and the $Q$-inner cut efficiency (grey). Three different energy bins are shown, along with the total energy range (lower right). Contours are 68, 95, and 99%.

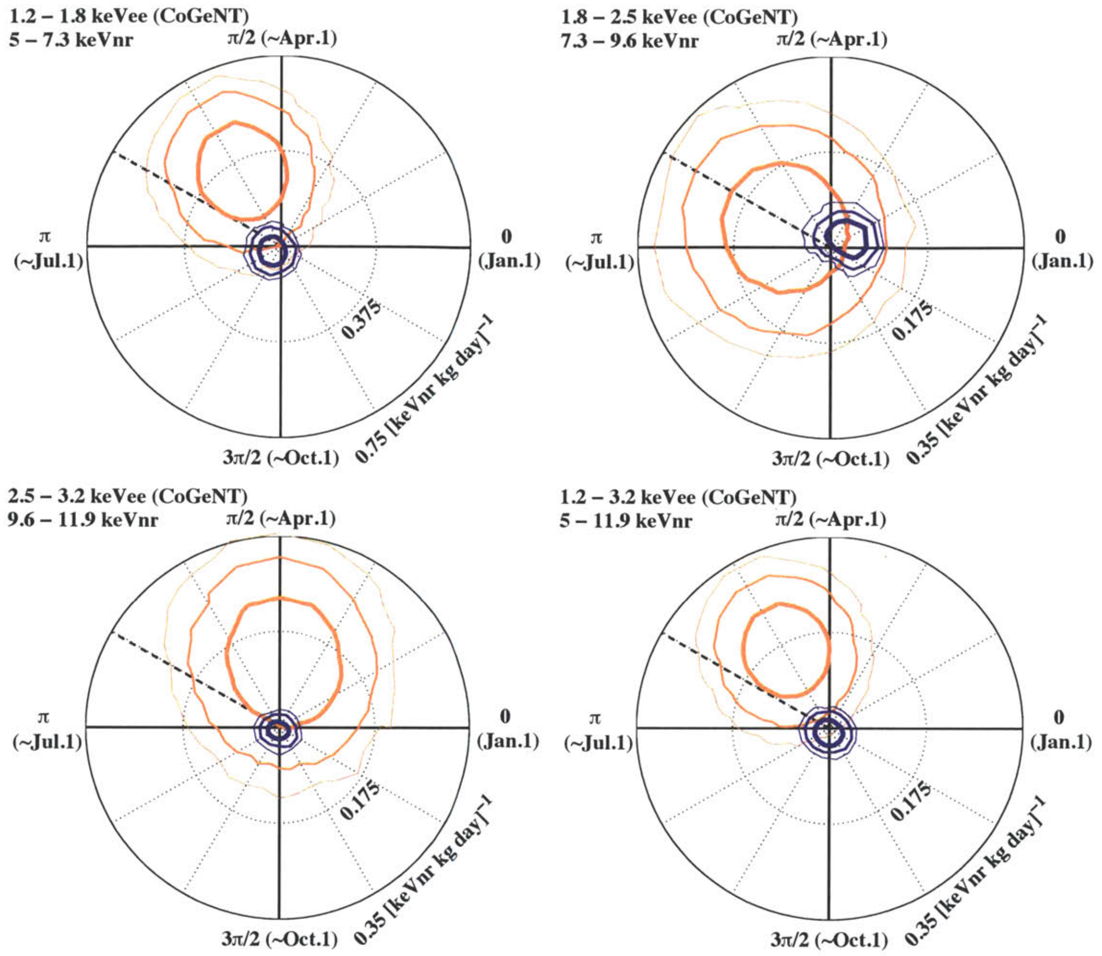

Figure 10-12: Feldman-Cousins allowed regions in a polar projection of modulated rate $M$ vs. $\phi$ for CDMS singles passing the nuclear-recoil cut (dark blue) and for the CoGeNT data (light orange). Three different energy bins are shown, along with the total energy range (lower right). Contours are 68, 95, and 99%.

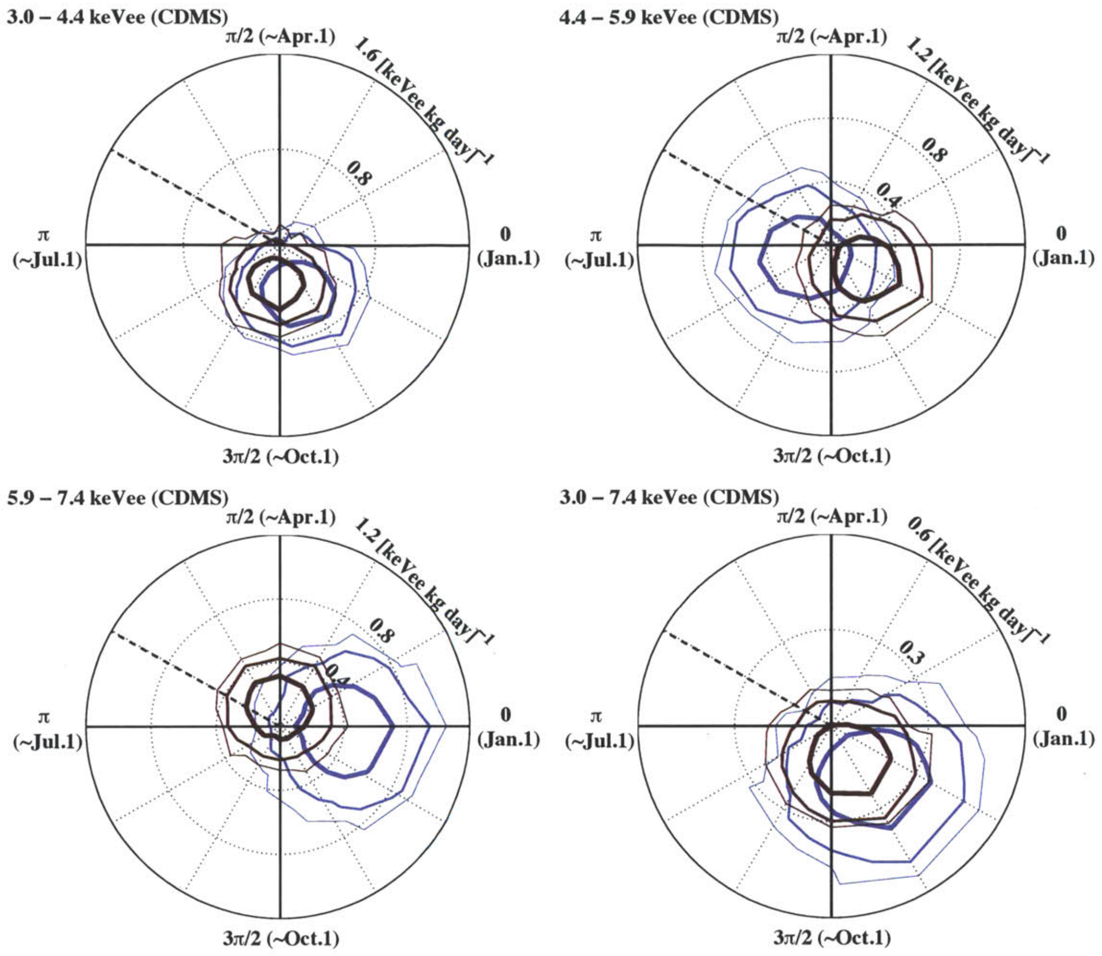

Figure 10-13: Feldman-Cousins allowed regions in a polar projection of modulated rate $M$ vs. $\phi$ for two event populations dominated by electron recoils: multiples (light blue) and singles (dark red). Three different energy bins are shown, along with the total energy range (lower right). Contours are 68, 95, and 99%.

# Part IV

# SuperCDMS Soudan

# Chapter 11

# Detectors for the SuperCDMS Soudan Experiment

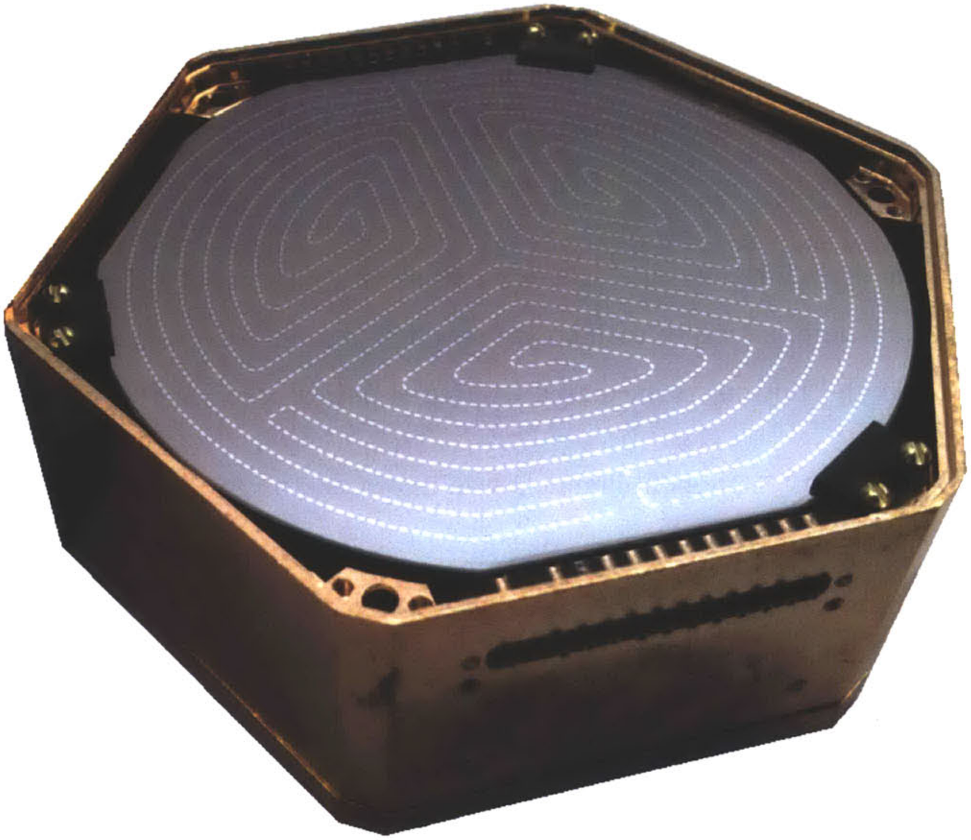

Figure 11-1: A mounted iZIP4 detector.

## 11.1 Introduction

We saw from the discussion of CDMS II that the main limitation of CDMS technology is the suppressed ionization yield of near-surface events. When an event occurs within a micron (or so) or a surface (either a bare surface or metal surface), the electrons and holes have the opportunity to recombine (either in the amorphous lattice of the bare surface or in the metal layer which has no energy gap at all), thus reducing the measured ionization yield, and making the surface electron recoil look very much like a nuclear recoil. In CDMS II, subtle differences in phonon pulse shape were used to offer some amount of surface event rejection ( 1:200 leakage for surface events in the NR band), but in order to gain this (still imperfect) rejection ability, we had to harshly cut on the data, eliminating half of the exposure and thus half of the scientific reach.

While CDMS II collected data underground at Soudan, detector design work continued at the surface. The basic detector design remained unchanged, and three incremental changes were implemented into what became known as the 'mZIP' (for 'maximized'):

1. The crystal thickness was increased from 10 mm to 25.4 mm (1 inch). The strategy here is simply to increase the bulk volume (exposure) while keeping the top and bottom surface area (source of potential leakage) unchanged.

2. The partition of phonon channels was improved to give better radial position information. A four-channel design was still employed, but instead of four quadrants, one channel became an outer ring (and the remaining three channels split the inner region into equal thirds). With inner-outer phonon information, the radial event position was much easier to measure (no more 'shrimp').

3. The QETs were better optimized for larger Al surface coverage, potentially resulting in better signal-to-noise on the subtle phonon pulse shape quantities that drove the surface-event discrimination. This QET improvement was tempered, though, by a necessary increase in TES width (from the switch form an

EV-align to full-wafer photolithography), which decreased the total number of QETs allowed per channel (to match the desired impedance).

The improvements of the mZIP over CDMS II detectors were significant, but the experiment was still be left largely in the same position: exposures would be limited in their reach by the necessity of a harsh surface-event cut based on dangerously subtle phonon pulse shape characteristics. This chapter is about a major leap in technology, which attacks the problem of low-yield surface events using a completely different strategy and succeeds in eliminating these surface events from the exposure completely, increasing the detector efficiency while dramatically decreasing the danger of surface event leakage.

## 11.2 The Interleaved Strategy

Let's begin by looking at Figure 11-2, a simplified cross-section through this new type of detector, with the top and bottom surfaces at the top and bottom of the figure. Notice the alternating voltages placed at both top and bottom surfaces. One surface alternates between *positively*-biased electrodes and ground lines, and the other surface alternates between *negatively*-biased electrodes and ground lines). Near the surface, field lines terminate on the biased electrode and the grounded metal (on the *same* side), whereas in the bulk, the electric field is as it was in CDMS II: smooth and uniform, connecting the top surface and the bottom surface.

This is the foundation of the iZIP ('i' for 'interleaved): Events in the bulk field region produce nearly *symmetric* charge signals (comparing charge signals from the top and bottom surfaces), whereas events in the surface field region will produce extremely *asymmetric* charge signals. For an idealized surface event in this detector, no charge at all will leak out of this surface field region and the opposite side will record no charge signal at all. This change in the electric field (and change from a one-sided to a two-sided charge sensing strategy) results a near-perfect surface event veto, solving the major difficulty of CDMS II.

In addition to the surface event veto, the interleaved strategy has many other

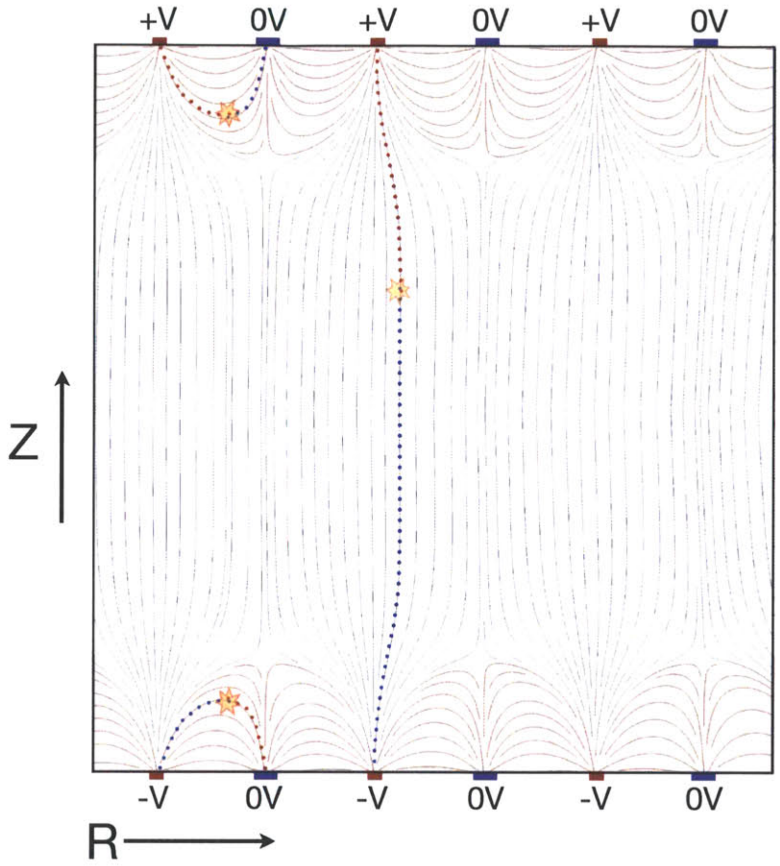

Figure 11-2: The general idea of the interleaved strategy, illustrated here in schematic form for clarity. We are looking at a cross-section through a detector, with the Z axis running top-bottom. Spaced at intervals across the top and bottom surfaces is a pattern of alternating biased and grounded metal surfaces (opposite bias on the two surfaces). The overall effect is a smooth uniform bulk field running from one side to the other, terminating on the biased electrodes, with a surface field region (colored pink here) with significantly radial direction, and field lines terminating at both ends on the same side. When charge is deposited in the bulk, it produces nearly symmetric charge signals in the electrodes on the opposite sides; when charge is deposited near the surface, it produces a charge signal on one side but no charge escapes the surface field to produce a signal in the opposite-side electrode.

benefits, which we get essentially for free:

**Two-sided phonon readout** has long been a goal of CDMS, but setting up a circuit which both biases TESs (in their transitions) and biases the QET aluminum (to produce a bulk drift field) is a challenging task. In the interleaved design, we have QETs at ground on both top and bottom surfaces (these are the grounded metal layers between the bias lines in the cartoon), and yet still have a bulk drift field (produced entirely by the biased, interleaved electrodes).

**Strong, horizontal surface fields** should raise the yield of surface events by partially overcoming the effects of surface trapping.

**Localized Luke phonon production** now occurs predominantly in the near-surface (high field) regions, rather than uniformly along the charge drift paths as in CDMS II. The fact that Luke phonons are always produced in the same place, irrelevant of event Z position, and the fact that Luke production is spatially separated from the recoil position, make the Luke contribution separable from the recoil contribution in phonon pulse shapes, and makes both contributions more informative.

We will come back and discuss each of these points more in this and the next chapter.

## 11.3 Evolution of the concept

The interleaved strategy is intuitive enough to have been thought of long ago (and perhaps it was), but CDMS is using this strategy only after learning of the work by P.N. Luke, at Lawrence Berkeley Laboratory. Beginning with a paper published in 1995 [81], Luke laid out the basic strategy of interleaving surface electrodes to create electrodes which are sensitive to whether an event is in a smooth bulk field region or in a surface field region. A figure from this first paper is shown in Figure 11-3. Luke based his concept on the idea of a Frisch grid, in which the bulk Ramo field is weakened (and the surface Ramo field is strengthened) by the placing of a

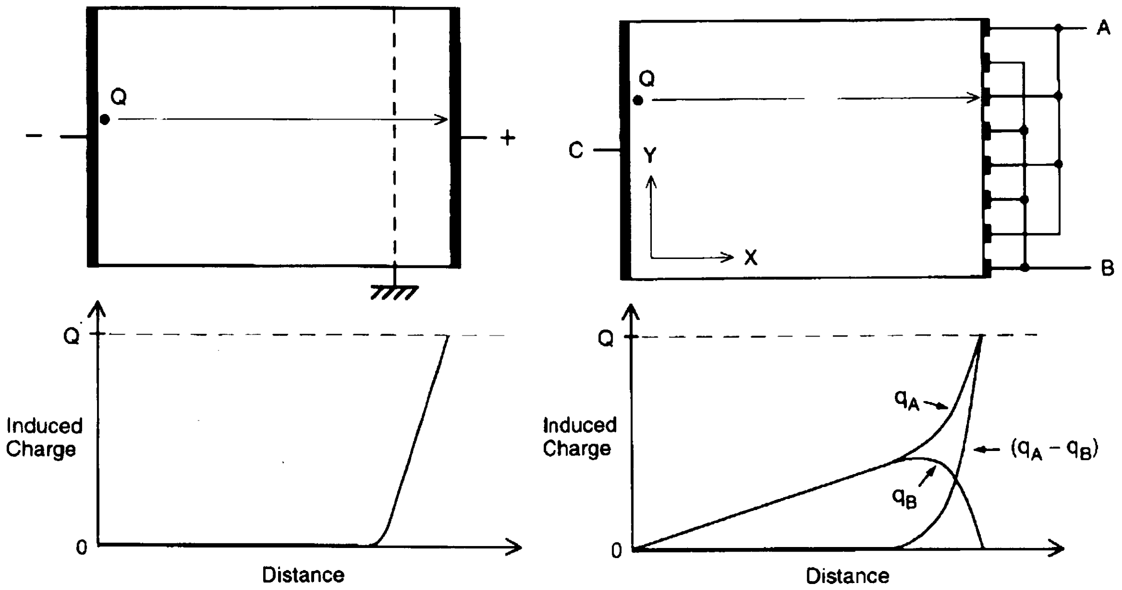

Figure 11-3: Figures from Luke*et al.* [81], showing two device schematics (top) in which charge (Q) drifts in the x direction towards a sensing electrode (or two) on the far right. On the left, the Frisch grid strategy is illustrated, in which a grounded metal grid distinguishes the bulk region (low drift field, low ramo potential, low induced charge on the electrode) from the surface region (high drift field, high ramo potential, high induced charge on the electrode). On the right, the Frisch grid is essentially just moved to the surface, but also instrumented, forming two interleaved electrodes (A and B). The charge induced on A and B ($q_A$ and $q_B$) by the drift is shown below, as well as a difference signal ($q_A$- $q_B$).

grounded mesh some distance into the detection medium (allowing only a portion of the electrode's field lines to reach the bulk). Of course, the Frisch grid strategy is much easier to implement in a gas or liquid detector (such as a TPC) than in a solid-state device like CDMS, but the basic idea of moving the Ramo sensitivity from the bulk to the surface by terminating field lines on an extra shaping element are the same.

The CDMS collaboration saw great potential in this strategy for tagging surface events (long ago recognized as the limiting factor for larger exposures), and altered Luke's design to place electrodes on both surfaces, interwoven with strings of (0V) QET sensors. This switch from a device in which two electrodes are on one side to a device in which the electrodes are on opposite sides (but still interwoven with something to alter their Ramo potentials) is the only way in which the Frisch-Luke basic strategy was altered.

The first CDMS device with interwoven electrodes and QET 'ribbons' was fabricated and tested in 2006, presented at the Low Temperature Detectors 11 conference that year [34], and one side of the mask design is shown in Figure 11-4. This device was constructed during the collaboration's transition from the stepper process (a grid of many small exposures) and the existing full-wafer exposure process. The Al was put patterned using a full-wafer mask, whereas the TESs were placed using a small 5mm square mask as in CDMS II. The device had the same number of channels as a standard CDMS II detector (or the mZIP), but arranged in such a way as to cover both sides: the two charge channels corresponded to the two sides, and each side was split in half into two phonon channels (where the sides were rotated at 90 degrees to each other, to give something like the quadrant layout of CDMS II when the two sides are combined).

The 2006 device was important in that it proved the basic concept of combining the yield discrimination of CDMS II with a charge-based fiducial cut in Z. Experience with this design also pointed pointed out the several weaknesses of the interleaved strategy:

- The interleaved design is inherently fragile. In practice, the 2006 device suffered from both charge and phonon opens and shorts that significantly degraded the device's usefulness. It is basically a problem of lowering the dimensionality; in CDMS II and in the mZIP design, both the charge and phonon sensors were essentially *planes*, and in the iZIP design, both the charge and phonon sensors became essentially *lines*. In retrospect, the iZIP1 QET ribbon (see FIgure 11-5) was well designed in theory, but terribly designed in practice in that it did not take into account the fabrication errors that unavoidably occur when patterning an entire wafer face. The ribbon placed Al elements at both the TES bias voltage and grown in close proximity to each other, leading to a multitude of shorts within the ribbon. But, even with an improved QET design, the linearity of the structures poses a hurdle.

- The capacitance of the charge channels is increased dramatically as a result of being interleaved with the grounded QET ribbons, increasing the charge noise and decreasing the charge energy resolution (and hence, the usefulness of yield as a discriminator). This capacitance constraint is why there must be some significant bare substrate surface separating the interleaved structures.

- In order to lower the capacitance of the electrodes, the surface area of the QETs, and hence the detector's ability to measure phonon timing and pulse shape, is significantly reduced. In defining the width of the QET ribbon, there is an unavoidable choice to make between better phonon signal-to-noise or better charge signal-to-noise.

- A two-sided charge readout and a two-sided phonon readout necessitates more channels than CDMS II or the mZIP. With one charge channel on the top and one on the bottom, there was no 'outer' charge channel on either side, which let in so many high-radius backgrounds that the vetoing of extreme-z backgrounds was overshadowed. Additionally, the 'outer' phonon channels of the mZIP had to be discarded in order to obtain at least some level of x-y position information.

As with the benefits of the interleaved strategy, these weaknesses too will be discussed more in the coming pages.

One imaginary weakness that de-motivated the collaboration from pursuing the iZIP strategy more actively was the fear of the 'saddle points' in the fields (as in Figure 11-2), where the field of the opposite side electrodes and the near side electrodes and QET ribbons exactly cancel each other out, leading to a region of zero field directly under each QET ribbon. It was thought that this was an extremely dangerous situation to allow in a detector, potentially leading to both electron-hole recombination and charge trapping. It was the Edelweiss collaboration, inspired by the CDMS device presented in 2006, which fabricated the first device to show that these saddle points are not a significant issue in 2008 [35], showing a surface event rejection better than 1:$10^4$. It was at this point that CDMS decided to forcefully pursue the interleaved strategy.

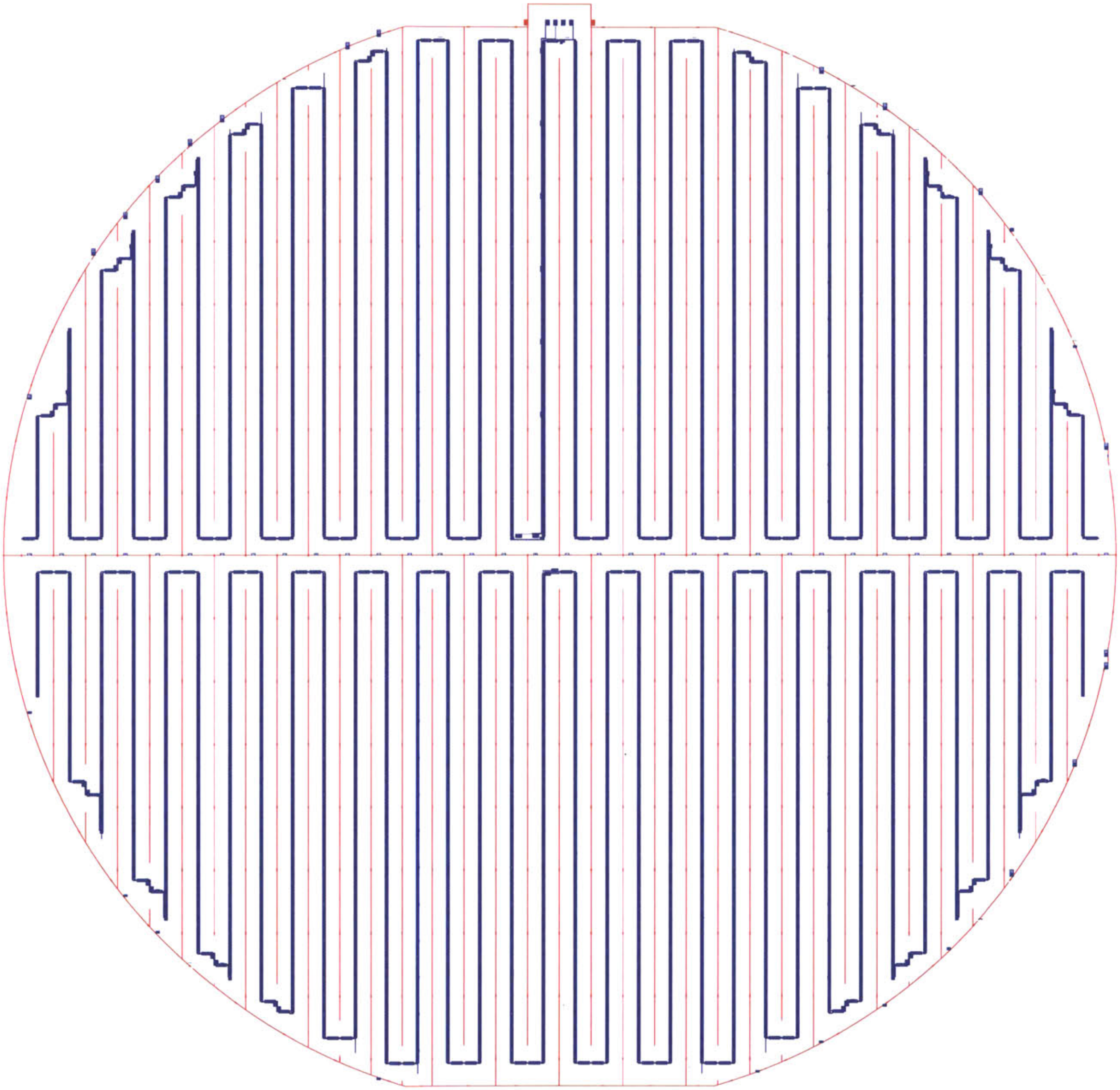

Figure 11-4: Side 1 of the iZIP1, with Al in blue and W in purple. Some Al features are apparent here that were temporary alignment marks, later removed during the $\alpha$-Si etch.

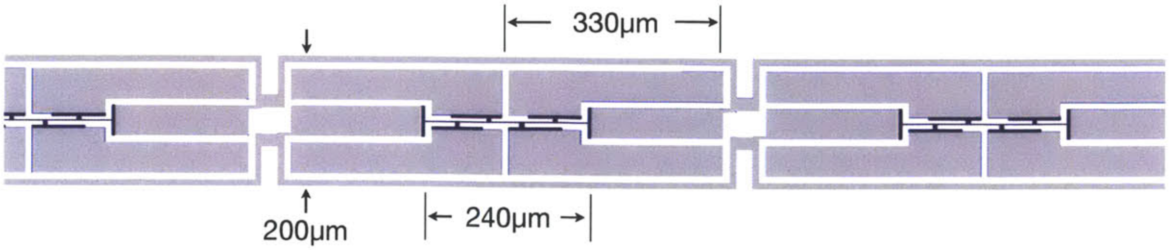

Figure 11-5: The first QET ribbon, as designed for the iZIP1 in 2006. Al is grey, W is dark blue.

We will skip the intermediate designs (no offense intended!) and discuss the motivations and resulting design decisions represented in the current generation of iZIPs (versions 4 and 5), currently running at Soudan.

## 11.4 Channel Partition

How many charge and phonon channels are necessary for an iZIP, and how should they be arranged on the detector faces? For SuperCDMS Soudan, we were constrained to use the existing CDMS II readout electronics (an issue that will come up again) which assumed each detector consisted of four phonon channels and two charge channels. From the 2006 device, it was clear this was not enough. The minimum number of charge channels per detector is 3 (a veto outer channel on a single side, preferentially the hole collection side, as was intended in CDMS II). This necessarily means involving the readout electronics of two CDMS II detectors, and for simplicity of triggering, it was decided to not share CDMS II readout groupings between multiple detectors. This leaves us with four charge channels and 8 phonon channels worth of readout per detector, so we might as well take advantage of the glut.

As illustrated in Figure 11-6, the charge channels essentially just duplicated the CDMS II (or mZIP) partition on the top and bottom (we now have two outer charge sensors) and duplicated the mZIP phonon partition on both the top and bottom (we now have two outer phonon sensors). Already, we can see that this design will have as one of its main strengths an extreme position-sensitivity. This partition gives us four channels measures of radial energy partition (inner-outer on the top and bottom in both charge and phonon systems). The inner three phonon channels on one side give us x-y position information, and in this design we have that measurement, too, duplicated on both sides (notice the 60 degree turn between the top and bottom side channel divisions to further boost this x-y sensitivity).

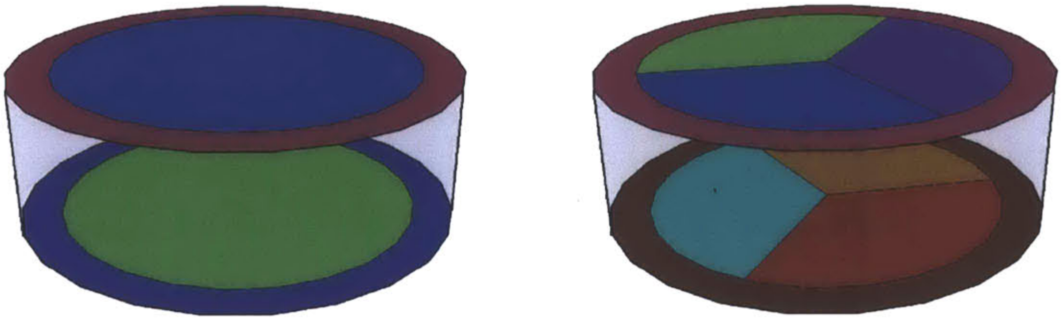

Figure 11-6: The partitioning of the sensors into channels. The left plot shows the arrangement of the four charge channels (inner/outer on side 1 and side 2) and the right plot shows the arrangement of the eight phonon channels (inner/outer, with the inner channels divided into thirds). Importantly, the 8 phonon channels are constrained to each contain the same number of QETs (and hence have the same resistance), which, if we want roughly constant Al surface coverage, leads to a constraint that the phonon channels be of nearly equal area.

## 11.5 Charge Channels

The charge channels are simply metal lines on the surface, connected to charge amplifying FETs (some distance away, at a higher temperature stage). The charge channels are defined almost entirely by a single attribute: capacitance. In the limit where the detector capacitance dominates the total capacitance of the channel (*i.e.*, detector capacitance is larger than the capacitance of the striplines connecting the detector to the FETs), charge channel resolution scales as the square root of the channel capacitance. The three variables relevant to the design, then, are just the three geometrical quantities: the width of the electrode lines, the width of the QET ribbon, and the spacing between them.

Note that capacitance is not the only attribute that varies with these geometrical variables. As the electrode width increases (the area of the biased metal), the surface field weakens and the bulk field strengthens. At the same time, the depth and strength of the ramo potentials (and resulting sensitivities to drifting charges) shifts from being nearly entirely in the surface region (thin electrodes) to being more evenly distributed (fat electrodes). And of course, the electrodes are phonon sinks (passive Al), so we pay a price in phonon signal to noise simply from phonon losses if the electrodes are a significant fraction of the Al surface area.

Back to capacitance for a moment. Capacitance is simply a description of the geometry, and the geometry of the iZIP is just slightly too complex to easily write down an analytical description. If we simplify the geometry slightly (as shown in Figure 11-7) by assuming that both the electrodes and QET ribbons have circular cross sections (instead of being extremely flat), then we can write an approximate description.

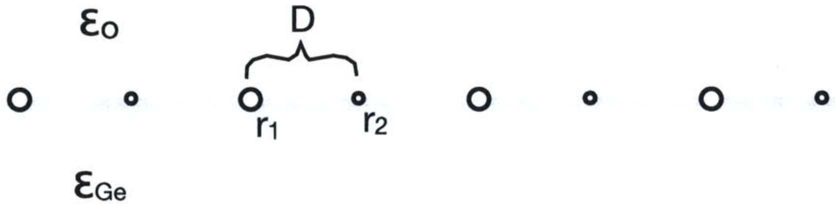

Figure 11-7: The simplified surface geometry, used as a starting point for writing down an approximate expression for iZIP capacitance. $r_1$ and $r_2$ are radii.

By Gauss' Law, the electric field outside a simple wire can be written as $E = \lambda/(2\pi\epsilon r)$, where $\lambda$ is the linear charge density along the wire. If we approximate the iZIP field as the superposition of the field resulting from placing $+\lambda$ on the charge rails and the field resulting from placing $-\lambda$ on the ground rails (this approximation should hold in our case, where $r_1$ and $r_2$ are much smaller than $D$), then the voltage difference between the two lines becomes simply

$$\Delta V = -\int_{r_1}^{r_2} E \cdot dr = -\int_{r_1}^{r_2} \frac{\lambda}{2\pi\epsilon r} dr = \frac{-\lambda}{2\pi\epsilon} ln\left(\frac{r_2}{r_1}\right) \tag{11.1}$$

In our case, we need to ad up the $\Delta V$ components not just in an electrode-ground pair, but between all electrodes and all grounds. We choose a starting pair (labeling them as "0") and add up the contributions of both the electrodes and grounds to the left and the electrodes and grounds to the right, as

$$\begin{aligned} \Delta V &= \frac{\lambda}{2\pi\epsilon} ln\left(\frac{D - r_2}{r_1}\right) \\ &+ \frac{\lambda}{2\pi\epsilon} ln\left(\frac{D - r_1}{r_2}\right) \\ &+ \sum_{m=1}^{\infty} \frac{-\lambda}{2\pi\epsilon} ln\left(\frac{2mD - r_1}{(2m-1)D - r_2}\right) \\ &+ \sum_{m=1}^{\infty} \frac{\lambda}{2\pi\epsilon} ln\left(\frac{(2m+1)D - r_1}{2mD - r_2}\right) \\ &+ \sum_{m=-1}^{-\infty} \frac{\lambda}{2\pi\epsilon} ln\left(\frac{(2|m|+1)D - r_2}{2|m|D - r_1}\right) \\ &+ \sum_{m=-1}^{-\infty} \frac{-\lambda}{2\pi\epsilon} ln\left(\frac{2|m|D - r_2}{(2|m|-1)D - r_1}\right) \end{aligned} \tag{11.2}$$

where the individual lines of the expression represent the voltage drop due to charge on the "0" ground line, on the "0" electrode line, on all ground lines to the right, on all electrodes to the right, on all ground lines to the left, and to all electrodes to the left. Collecting terms and simplifying, we arrive at

$$\Delta V = \frac{\lambda}{2\pi\epsilon}\left[ln\left(\frac{D}{r_1}\right) + ln\left(\frac{D}{r_2}\right) + 4\sum_{m=1}^{\infty} ln\left(\frac{2m+1}{2m}\right)\right] \tag{11.3}$$

which can be converted easily into a capacitance per area as

$$C = \frac{2\pi\epsilon}{ln(D/r_1) + ln(D/r_2) + 4\sum_1^\infty ln[(2m+1)/2m]} \tag{11.4}$$

The foundation on Gauss' Law allows us to set $\epsilon$ to be simply an equal share of the two materials, as $\epsilon = (\epsilon_o + \epsilon_{Ge})/2$. If we set the charge line width to some minimum (set by fabrication) and the QET ribbon width to some approximate minimum (set by our desire to have *some* Al coverage and the rough scale of the $\sim 200\mu$m TESs), then the main free parameter is $D$, the separation distance. Figure 11-8 shows the capacitance of both electrodes on one side of a 76.2 mm crystal. A spacing of 1.6mm was chosen, to give a rough scale of $\sim$ 80pF for the iZIP4 ($8\mu$ electrode width) and $\sim$ 105pF for the iZIP5 ($40\mu$ electrode width).

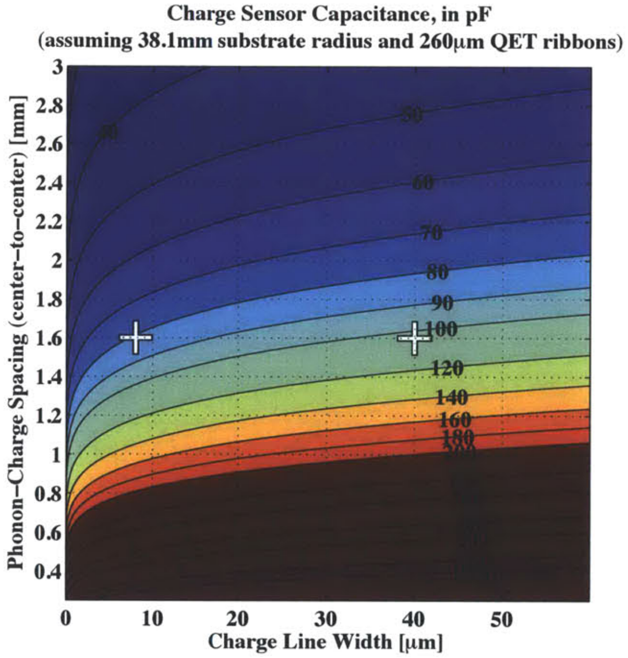

Figure 11-8: The total charge sensor capacitance (inner channel + outer channel) for one side of 38.1mm-radius iZIP, as calculated by the analytic approximation discussed in the text. It is seen that expected capacitances are in the range of tens to hundreds of pF. The design parameters for the iZIP4 ($8\mu$m charge lines) and iZIP5 ($40\mu$m charge lines) are indicated.

By the way, this increase in electrode width is the main difference between the iZIP v4 and v5, and was intended to both boost the bulk electric field strength (to lower bulk trapping rates) and allow the Ramo potential of the electrode to have greater sensitivity to charges drifting in the crystal bulk (to increase sensitivity to charges even if they do trap before reaching the surface). With our operating experience so

| Electrode Width [$\mu$m] | $E_{max}$ [V/cm] | $A_{pol.}$ [$\mu$m$^2$] | $E_{bulk}$ [V/cm] | $C$ [pF] | $f_{passive}$ |
|---|---|---|---|---|---|
| 8 | 3700 | 100 | 0.46 | 127 | 0.01 |
| 20 | 2500 | 56 | 0.50 | 140 | 0.07 |
| 40 | 2100 | 28 | 0.55 | 151 | 0.16 |
| 100 | 1600 | 13 | 0.62 | 175 | 0.35 |

Table 11.1: Several design characteristics as a function of electrode width, assuming the top and bottom electrodes are biased at ±2V (typical for the iZIP), the electrode-ribbon spacing constant at 1.6 mm, and the ribbon width constant at ~200$\mu$m. The meaning of the design characteristics is briefly described in the text.

far, the practical difference achieved by this electrode widening seems to have offered only slight practical value.

Assuming that the separation distance $D$ as a fixed parameter, the optimization of the bias electrode width is a complicated business, involving

**The maximum field strength $E_{max}$** , which occurs at the electrode edge, and determines the voltage onset of charge breakdown (*i.e.*, the maximum bias voltage).

**The cross-sectional area of polarized Ge $A_{pol.}$** , surrounding the biased electrode, which shields the electrode and lowers the induced signal.

**The strength of the bulk field $E_{bulk}$** , which may be very important to avoid charge trapping and promote charge stability.

**The total side capacitance $C$** , which causes a noise term that scales as $\sqrt{C}$, as previously mentioned.

**The fraction of the surface Al that is in the electrodes $f_{passive}$** , and therefor robbing the phonon sensors of valuable phonons.

A summary of these effects as a function of electrode width is summarized in Table 11.1, based on a finite element model by M. Pyle.

The finite element model used by M. Pyle suffered from a low spacial resolution, and therefor thicker electrodes (4$\mu$m instead of the real value of 350nm). An improved finite element method was employed (using COMSOL modeling software) without

|  | Qi1 | Qo1 | Qi2 | Qo2 |
|---|---|---|---|---|
| Qi1 | 69.8 | 2.81 | 1.48 | 0.648 |
| Qo1 | 2.81 | 36.9 | 0.648 | 0.646 |
| Qi2 | 1.48 | 0.648 | 69.8 | 2.81 |
| Qo1 | 0.648 | 0.646 | 2.81 | 36.9 |

Table 11.2: The capacitance matrix for the iZIP5 (electrode width of 40$\mu$m), in pF. The *total* capacitances are 69.8 pF for each inner channel and 36.9 pF for each outer channel (leading to a combined side total electrode capacitance of 101 pF).

this limitation, in order to calculated channel by channel capacitances. The resulting capacitance matrix is shown in Table 11.2, and agrees surprisingly well with the the basic analytic approach, and somewhat less well with the low-resolution finite element method.

In the end, we have perhaps over emphasized the importance of electrode capacitance, because we are likely dominated by other sources of charge noise, further up in the channel electronics. At the Berkeley test facility, iZIP4 detector G48 was observed to have a charge resolution (using an optimal filter) of 288±20 eV, quite similar to the value obtained for the CDMS II and mZIP charge channels, of radically different geometry.

## 11.6 Shaping the drift field at high radius

We have seen that we expect the bulk drift field to be uniform and of amplitude ~0.5 V/cm (much weaker than the CDMS II field strength of 3 V/cm). This field is uniform in the center of the detector, but at high radius can include some radial component as a result of the edge of the crystal, the ending of the alternating surface patter, and the geometry of the surrounding materials (primarily a grounded copper housing, but partly too any detectors immediately above or below).

The radial components of the high-radius field are clearly important, because they can either increase or decrease the amount of sidewall trapping depending on how these fields are shaped. The two geometrical parameters available to CDMS in shaping this high-radius field are the distance between the crystal and the grounded housing (which tends to push field lines *into* the crystal) and the width of the outermost electrode line (which tends to boost the high-radius field and create a field more dominated by the Z direction). Some illustration of this basic idea is shown in Figure 11-9, where the spacing between the crystal and the grounded housing is slowly increased, keeping all other parameters fixed.

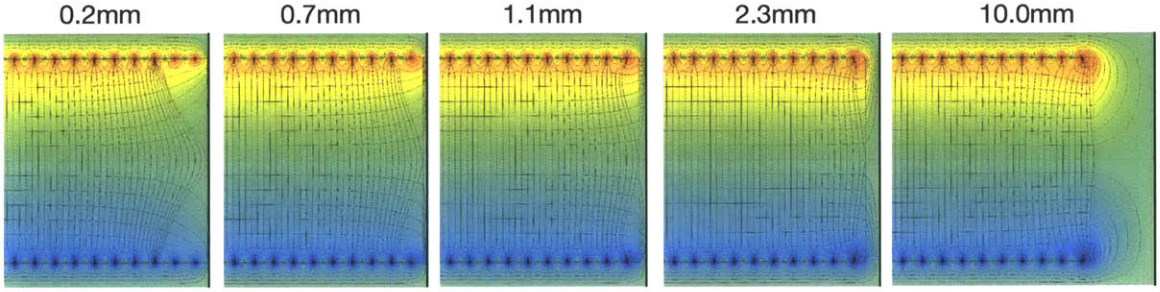

Figure 11-9: A cross-sectional view through the same iZIP design as the radial size of the housing is increased from left to right (distance between the crystal and the housing is indicated above each view). The farther the housing, the more z-oriented the high-radius field lines.

## 11.7 QET Arrays 1: continuing the goals of CDMS II

We move from the geometry of the charge sensors to the geometry of the phonon sensors, the QET ribbon. Much of the QET ribbon optimization remains the same from CDMS II. We want to maximize active Al area, while minimizing passive Al area. Relatedly, we want to keep the Al fin lengths short ($< 1.5\ell_{trapAl} \approx 380$ $\mu$m assuming Al thickness of 350 nm). We want the TES length to be just barely less than the phase separation length ($\sim 200\mu$m for a $T_C$ of 80 mK), which is suppressed from that value by the volume of W in the Al-W overlap region. We want the volume of the W in the Al-W overlap region to be small so that the TES does not extend larger than the phase separation length ($\ell_{max} \sim V_W^{-1/2}$), and additionally so that the W has a small heat capacity. On the other hand, we would want the Al-W overlap to be large to increase the transmission probability for quasiparticles from the Al into the W, known to be one of the major sources of inefficiency in the QET.

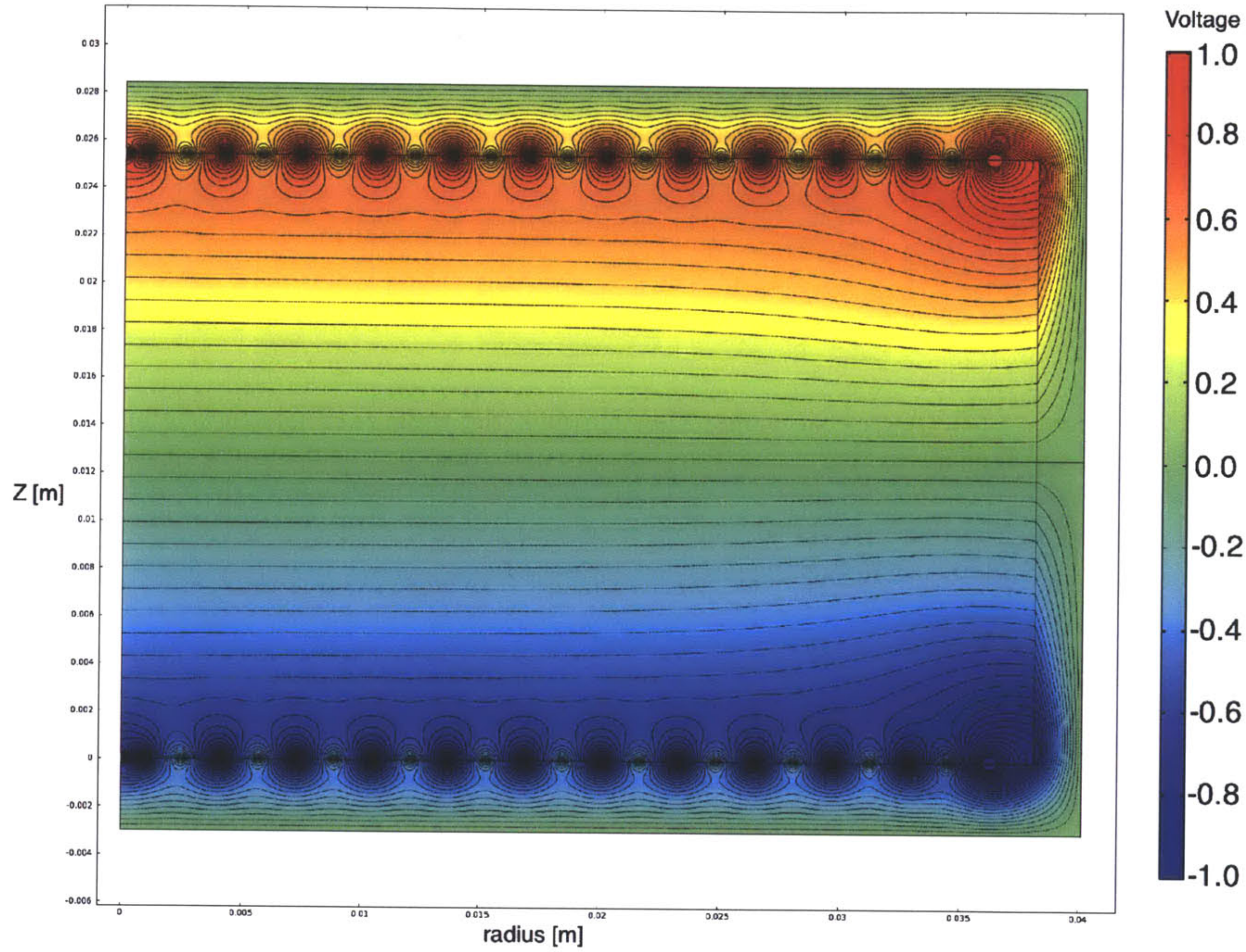

Figure 11-10: A cross-section through an iZIP5, showing the voltage in the crystal (and between the crystal and a grounded Cu housing). The geometry is assumed to be axisymmetric here to reduce the calculation down to two dimensions. Note the smoothness of the bulk field region, and also that the detector is biased at ±2V, but the coloring and equipotential contours only extend to ±1V. Much of the voltage difference occurs very near the electrodes. The outermost charge lines are significantly thicker than the others, pushing the equipotential contours deeper into the crystal at higher radius.

This thesis does not contain an in-depth optimization of all these competing interests, largely because we are still lacking the measurements necessary to understand the optimization. We do not know what $\ell_{trapAl}$ is with any certainty, or any of the other quasi-particle diffusion parameters, so we do not really know the optimal fin length, the optimal Al-W overlap length, etc. We also do not know how much of the Al-W overlap contributes to the heat capacity and phase separation length degradation (though something along the lines of 'half' seems to fit all the data). We have not learned much about these fundamental design parameters since the design work for CDMS II a decade ago, which means that the iZIP QETs will look much the same as before. The only difference from the CDMS II optimization is that the minimum feature size has increased (due to the full-wafer patterning) to $2.4\mu$m, decreasing the resistance of a single TES, and decreasing the number of parallel QETs per channel (keeping the biased-state resistance fixed at $\sim$200m$\Omega$ to match the readout electronics).

With a maximum TES length $\ell_{max}$ of $\sim 220\mu$m, a minimum TES width of $2.4\mu$m, a W film sheet resistance of $\sim 3.3\Omega$/sq, and an operating channel resistance of $\sim$200m$\Omega$, we arrive at the number of QETs per channel: 458. Given the spacing decisions made based on reducing the electrode capacitance, the total length of each phonon channel turns out to be 41 cm (impressively long!) which results in a QET spacing along the ribbon of just under a millimeter, at $895\mu$m. Note that this is not *too* far distant from twice the $1.5\ell_{trapAl}$ fin length of $325\mu$m, meaning that the ribbon largely be composed of QETs themselves, rather than sparsely spaced QETs connected with long bias rails.

We are largely running blind when optimizing the Al-W overlap region, but this region has gradually evolved through trial and error since CDMS II. Most significantly, the area of the overlap has been minimized for two reasons: so as to minimize contributions to phase separation, and so as to minimize trapping in both the W and the Al-W overlap region, both of which have shorter (but unknown) trapping lengths as compared to Al. The overlap width was set to $4.5\mu$m.

The absolute efficiency of the iZIP QETs (fraction of the total phonon energy reaching the TES) was calculated by H. Chagani and M. Pyle, and found to be $\sim$ 12% for the iZIP4 (detector G48). This is remarkably better than the $\sim$ 3% of CDMS II, despite the similarly-sized Al fins and significantly lower total active surface coverage ($\sim$ 6.2% per side). The size of this improvement in efficiency is somewhat surprising, therefor, and should probably be attributed to either the reduced Al-W overlap area (resulting in reduced trapping in that region) or improved fabrication procedures (or, most likely, some combination of the two).

In addition to the efficiency of the quasiparticle diffusion, another measurable effect of reducing the Al-W overlap area is the changed TES electrothermal feedback time constant, $\tau_{etf}$. In an iZIP2 (G3D) which happened to be fabricated with sides exhibiting very different $T_C$ values ($\sim$46 and $\sim$107 mK), $\tau_{etf}$ was measured by observing the phase-offset portion of the device's complex impedance. At the very lowest frequencies, electrothermal feedback is perfect, and the input increase in power is offset by an equal and opposite (phase = $-\pi$) response. As the frequency increases, the current response starts to lag behind the input power, and the frequency at which this lagging starts is a measure of the electrothermal feedback time constant. (At higher frequencies, inductances of the readout loop dominate, and the phase offset becomes positive.) For this iZIP design, $\tau_{etf}$ was found to be $\sim$25$\mu$s and $\sim$40$\mu$s for the low-$T_C$ and high-$T_C$ sides, respectively.

## 11.8 QET Arrays 2: new constraints brought on by interleaving

The only phonon channel design goals that are fundamentally new for the iZIP are

- The QET ribbon width, which should be small so as not increase the charge channel capacitance.
- The QET ribbon length and internal fill factor, which should be short and 'filled in' respectively, to lower the ribbon's inductance (so as not to increase the L/R electrical time constant of the SQUID readout).

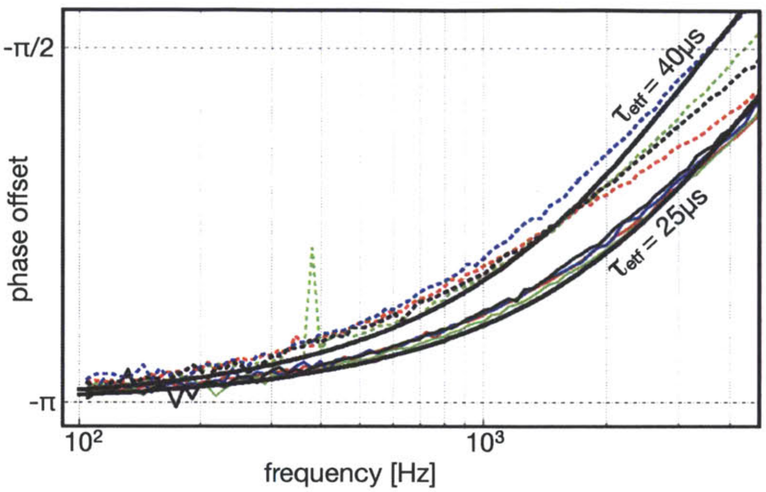

Figure 11-11: The estimation of $\tau_{etf}$ by looking at the low-frequency complex impedance (here, the phase offset) of the phonon channels. Channels with two significantly differing $T_C$ values are shown here, leading to the differing values of $\tau_{etf}$ despite identical geometries.

Lets discuss each of these issues a little more.

We have emphasized both the importance of the general surface geometry (charge line width, QET ribbon width, charge-QET separation distance) in setting the electrode capacitance, and thus the fundamental noise limit of the charge measurement. Of course, we have also mentioned that other noise issues upstream from the detector appear to be dominating, but let's still treat the question of electrode capacitance seriously. The width of the QET ribbon should be minimized in order to lower the electrode capacitance, and should be maximized in order to increase the active Al surface area and maximize the phonon pulse shape signal-to-noise. In practical terms, we decided on an acceptable electrode capacitance (this is the priority, because it is charge information that defines the all-important fiducial volume in both R and Z) and then set the QET ribbon width accordingly. Unfortunately, this means the QET ribbon suffers from the same problem as in CDMS II: there are areas near the TES (where super-efficient 'beachfront property' Al could be placed) that remain uninstrumented due to the ribbon width constraint. The maximum ribbon width (near the TES itself) is 308$\mu$m and in an effort to further lower capacitance, the ribbon is much narrower in the gaps between QETs. For calculating capacitance, the 'average width' can be used, ~200$\mu$m. The constraint of low capacitance has effectively forced us to reduce the total Al surface area on the top and bottom surfaces down to only ~ 6.2%.

Just as for capacitance, inductance is a purely geometrical characteristic. The inductance of an iZIP phonon channel is naturally high, because the inductance of parallel lines increases as the square of the length (and remember that the phonon channel is a surprising 41 cm long). We can write down an analytic expression for inductance per length if we first make the simplifying geometrical assumption (as we did for calculating the electrode capacitance) that the phonon rails are simple wires of circular cross section (radius $r$), separated by a distance $w$ from surface to surface. In this case, the flux per unit length becomes

$$\Phi = \ell I = 2\int_r^w B dy = 2\int_r^w \frac{\mu_o I}{2\pi y} dy = \frac{\mu_o I}{\pi} ln\left(\frac{w}{r}\right) \tag{11.5}$$

and the inductance per unit length is simply

$$\ell = \frac{\mu_o}{\pi} ln\left(\frac{w}{r}\right) \tag{11.6}$$

To give a sense of magnitude, this inductance-per-length becomes ~10 nH/cm after rough numbers are put in.

We have an inductance per length, but we do not simply multiply this by the channel length. Some signal is put into this channel at the maximum length, some at the minimum length (where the effective detector inductance for that input is zero). We already see a good reason to minimize detector inductance: different areas of the detector will be put through a low-pass filter with different cutoffs, determined by where along the length QET ribbon the even happens to be. If we forget about local information for the time-being, and imagine the phonons hitting the entire channel at once, then the effective inductance goes as length squared.

Two geometrical strategies can be employed to reduce the inductance: decrease the length, and decrease the gap width $w$. First, the length can be shortened by changing how the 41 cm is arranged. Instead of a 41 cm line, the inductance can be reduced simply by covering the same surface area with two 20.5 cm 'branches'. This is effectively what was done in the iZIP4/5, except that the two branches were joined at their tips (for added fabrication robustness) so that each channel is actually a loop of QETs, effectively cutting the inductance in half ($L \sim \ell^2 \rightarrow L \sim 2(\ell/2)^2 = (1/2)\ell^2$). Additionally, the inductance is proportional to the amount of flux passing between the rails, which is proportional to the non-metal 'white space' area between the rails. Efforts were made to push the rails together and fill in the inter-rail area with conductor (Al). This is totally consistent with both our desire to push the rails together for capacitance reasons and our desire to fill the region between the rails to obtain high phonon absorption. The resulting inductance per phonon channel in the iZIP4/5 is estimated at ~20nH, meaning that the inductance of the readout electronics dominates. The channel inductance will be much more important once the SNOLAB readout electronics (with a lower inductance) are used, but in the future the non-metal area between rails will also be significantly reduced (as will be discussed in the SNOLAB section).

## 11.9 Fabrication Robustness

The iZIP4/5 was designed not as a test device, but for actual mass production and deployment at Soudan. Fabrication robustness was one of the main goals of the design, so that detectors could be quickly and easily fabricated for the experiment.

Laying down the device layers and patterning them is a difficult job. Several types of microscopic problems can occur, including isolated bubbles in the photoresist , isolated scratches on either a substrate or a fabricated layer, or a patch can simply flake off. Aluminum, in particular, is a fragile material and has poor adhesion qualities, making it difficult to work with. For this reason, whenever possible, aluminum in the design was overlain with tungsten as a sort of rigid support. This was done on the bond pads, on the electrode lines, and even on the rails between QETs (even though this deadens some small portion of otherwise active Al).

On the other hand, if the W completely covered the Al rail, any gaps in the Al

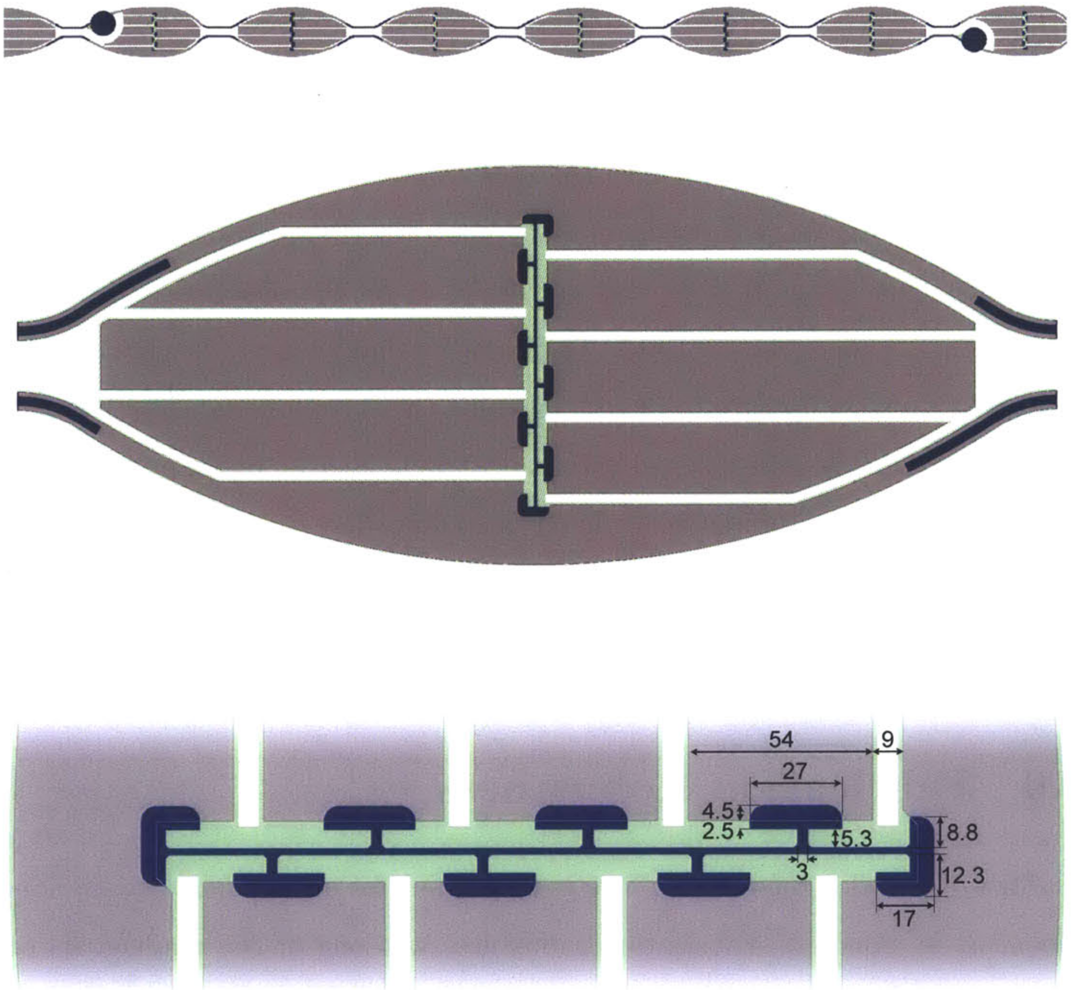

Figure 11-12: Several views of the iZIP4.3 phonon sensors. At top is shown the 'ribbon' of QETs. Included are two 'safety' bond pads, to repair opens in case of fabrication errors. Also, note the use of W to strengthen the Al rails between QETs. At center is shown a single QET (length of central fin: $327\mu$m). At bottom is shown the TES ($\sim 2.4\mu\text{m}\times 220\mu\text{m}$) and the fin-TES interface structures. Compare to the similar figure from the CDMS II detector chapter. Grey is Al, dark blue is W, and light green is the underlying $\alpha$-Si layer.

rail would no longer be visible. For the phonon channels, it is the superconducting path through the Al that really biases and reads out the QETs, so the Al continuity is extremely important to inspect after fabrication is complete. The phonon channel rails had an Al width of $14\mu$m and a W width of only $7\mu$m, to try and balance the competing goals of strength and inspectability. The electrodes were similarly half-covered, though the superconducting Al continuity was less important.

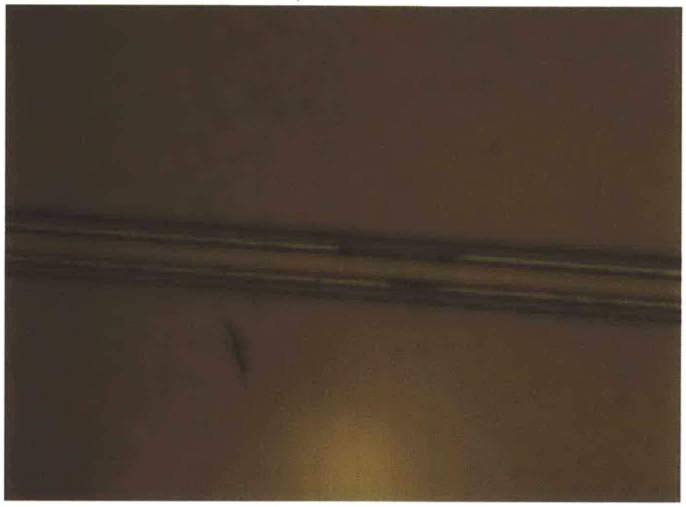

Figure 11-13: A photomicrograph, showing the importance exposing a significant amount of Al. All three layers of the line are visible here: $\alpha$-Si, then Al, then W. Here, the Al layer has a small gap (a fabrication error), which can only be seen where the W is not overlain. Because the Al open can be seen, it is easy to fix with an Al wirebond (between two 'safety' pads, visible at the top of Figure 11-12).

The QET underwent a revision from v4.3 to v4.4 in which the far ends of the Al fins were arranged such that voltage differences between adjacent structures was minimized, as in Figure 11-14.

## 11.10 The Array of Detectors for SuperCDMS Soudan

To take advantage of the existing readout electronics at Soudan, consisting of 30×4 phonon channels and 30×2 charge channels, 15 detectors were prepared for installation (15×8 phonon channels and 15×4 charge channels). These 15 detectors were arranged into five towers of three detectors each, leading to a total Ge mass of 8.95 kg

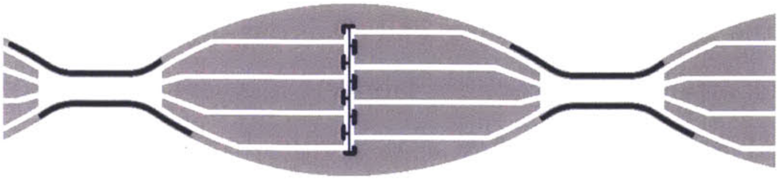

Figure 11-14: An improved fin design (starting in iZIP4.4) in which Al structures (fins and rails) have only small voltage differences with adjacent structures. Compare to Figure 11-12.

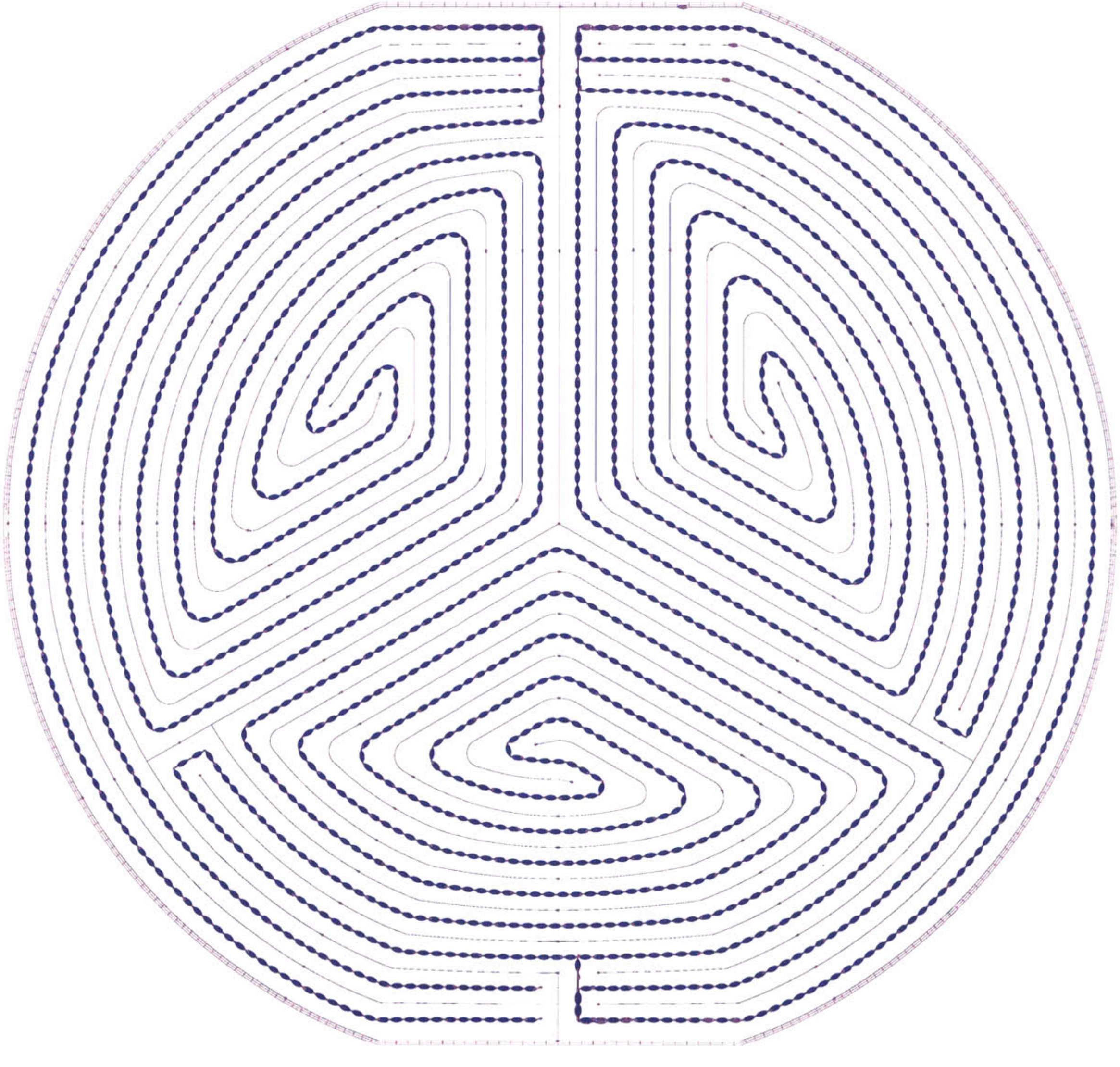

Figure 11-15: Side 1 of the iZIP4. Here, Al is blue, W is purple, and $\alpha$-Si is green.

(masses vary from 580 g to 610 g depending on level of surface polishing).

In a display of confidence in the detectors' surface event rejection capabilities, two weak $^{210}$Pb sources are installed within the cryostat so that we have a large exposure of surface events, similar to a ton-scale experiment, and can make measurements of beta rejection which are limited neither by muon backgrounds (as is the case at the surface) or by statistics. These two sources are at the top and bottom of Tower 3, one above T3Z1 and one below T3Z3. These sources were produced by exposing Si wafers to radon-contaminated air over the course of weeks, letting the $^{210}$Pb gradually accumulate, while regularly measuring the Si wafer radioactivity. The obtained rate of beta emission is $\sim$1 decay per minute. Although the rate is astronomically higher than any we expect in an actual WIMP search run, these radon-daughter sources otherwise exactly replicate the most dangerous backgrounds we expect to see.

The next chapter is a first look at the rejection ability of these surface recoils.

|  | $\mathrm{T}_C$ [mK] | $\mathrm{R}_{Normal}$ [$\Omega$] | Elect. Width [$\mu$m] | Impurity Level [$\mathrm{cm}^{-3}$] | Etch Pit Density [$\mathrm{cm}^{-2}$] | Mass [g] $\pm$4.9g |
|---|---|---|---|---|---|---|
| **T1Z1** | ?, 87, 87, 87 | >1, 0.65, 0.65, 0.70 | 40 | $4\times10^{10}$ | 5000 | 609.5 |
| (G41) | 114, 100, 101, 100 | 0.75, 0.60, 0.70, 0.65 |  |  |  |  |
| **T1Z2** | 75, 78, 78, 78 | 0.85, 0.75, 0.70, 0.85 | 8 | $4.4\times10^{10}$ | 3890 | 609.5 |
| (G42) | 82, 80, 78, 80 | 0.80, 0.80, 0.80, 0.70 |  |  |  |  |
| **T1Z3** | 86, 87, 87, 87 | 0.70, 0.70, 0.70, 0.65 | 40 | $6\times10^{10}$ | 3890 | 597.4 |
| (G43) | 113, 100, 99, 98 | 0.75, 0.65, 0.65, 0.65 |  |  |  |  |
| **T2Z1** | 76, 80, 79, 80 | 0.48, 0.50, 0.41, 0.51 | 40 | $4\times10^{9}$ |  | 597.4 |
| (G24S) | 98, 94, 95, 94 | 0.58, 0.51, 0.54, 0.56 |  |  |  |  |
| **T2Z2** | 70, 79, 77, 78 |  | 40 |  |  | 591.3 |
| (G53) | 93, 85, 86, 85 |  |  |  |  |  |
| **T2Z3** | 83, 85, 85, 85 | 0.65, 0.61, 0.51, 0.57 | 40 | $8\times10^{10}$ | 3132 | 579.1 |
| (G2E) | 101, 94, 95, 94 | 0.62, 0.54, 0.58, 0.53 |  |  |  |  |
| **T3Z1** | 88, 91, 91, 91 | 0.70, 0.65, 0.70, 0.65 | 8 |  |  | 603.4 |
| (G48) | 109, 101, 101, 101 | 0.65, 0.55, 0.55, 0.60 |  |  |  |  |
| **T3Z2** | 91, 92, 93, 91 | 0.75, 0.60, 0.65, 0.70 | 8 | $7\times10^{10}$ | 5900 | 591.3 |
| (G47) | 107, 96, 99, 99 | 0.60, 0.50, 0.50, 0.55 |  |  |  |  |
| **T3Z3** | 87, 89, 89, 89 | 0.75, 1.00, 0.65, 0.65 | 8 | $2\times10^{10}$ | 6720 | 609.5 |
| (G52) | 103, 97, 96, 98 | 0.75, 0.60, 0.80, 0.60 |  |  |  |  |
| **T4Z1** | 88, 90, 88, 90 | 0.55, 0.45, 0.45, 0.45 | 40 | $1.1\times10^{10}$ | 5110 | 597.4 |
| (G21P) | 104, 97, 97 97 | 0.60, 0.85, 0.70, 0.70 |  |  |  |  |
| **T4Z2** | 89, 92, 91, 92 | 0.85, 0.75, 0.80, 0.80 | 8 | $8\times10^{10}$ | 4312 | 597.4 |
| (G50) | 106, 101, 101, 100 | 0.65, 0.64, 0.55, 0.70 |  |  |  |  |
| **T4Z3** | 83, 86, 86, 86 | ?, 0.70, ? 0.70 | 40 | $2.5\times10^{9}$ | 5290 | 594.3 |
| (G20O) | 100, 95, 95, 94 | 0.70, 0.65, 0.60, 0.63 |  |  |  |  |
| **T5Z1** | 71, 75, 75, 75 | 0.65, 0.55, 0.60, 0.60 | 8 |  |  | 579.1 |
| (G51) | 82, 79, 77, 78 | 0.65, 0.60, 0.55, 0.60 |  |  |  |  |
| **T5Z2** | 63, 70, 68, 71 | 0.80, 0.65, 0.75, 0.75 | 8 | $3\times10^{9}$ | 6470 | 606.5 |
| (G19N) | 96, 88, 88, 88 | 0.70, 0.75, 0.50, 0.75 |  |  |  |  |
| **T5Z3** | 83, 84, 84, 84 | 0.64, 0.62, 0.62, 0.63 | 40 |  |  | 591.3 |
| (G7F) | 100, 91, 91, 90 | 0.60, 0.62, 0.49, 0.60 |  |  |  |  |

Table 11.3: Summary of the basic measured physical properties of the 15 SuperCDMS Soudan detectors (transition temperature, normal resistance, electrode line width, substrate impurity level, substrate etch pit density, and substrate mass). The $\mathrm{T}_C$ and $\mathrm{R}_N$ values are listed for each of the eight phonon channels, as Side 1 (A,B,C,D) then Side 2 (A,B,C,D). The electrode width was changed from the iZIP4 (8$\mu$m width) to the iZIP5 (40$\mu$m width) in an effort to increase the charge stability (and the resulting time between LED flashing). The impurity levels and etch pit densities are two measures of crystal purity (an etch pit density test reveals the density of certain lattice dislocations). The mass was inferred from the crystal thickness, itself measured by how many spacers were necessary when mounting the crystal in a housing. Thickness (and mass) varies as a result of varying amounts of polishing.

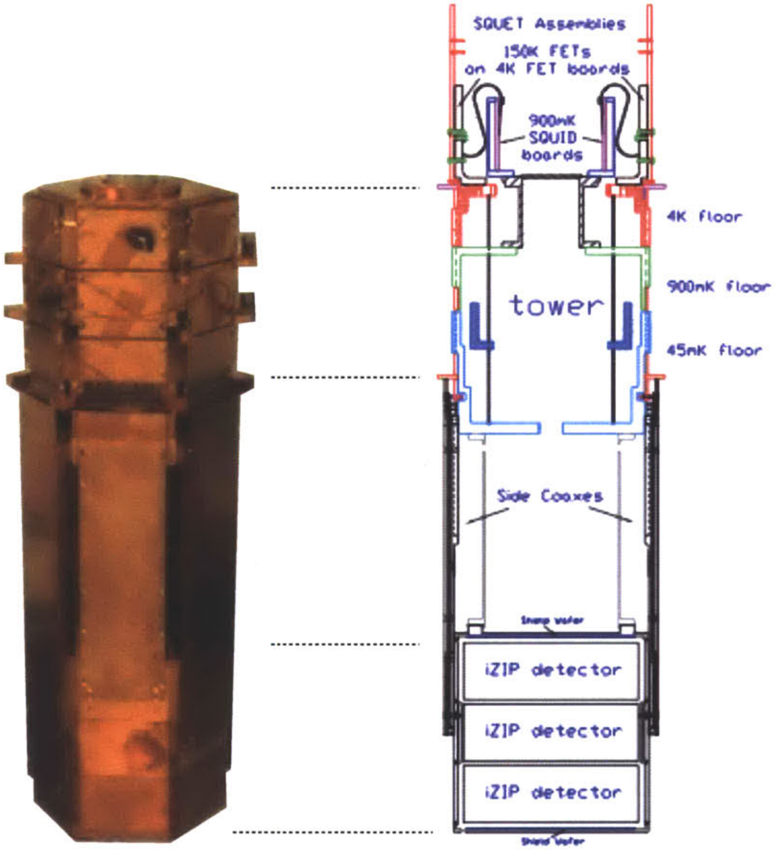

Figure 11-16: The SuperCDMS Soudan 'tower' structure, holding three iZIPs (base) and connecting them to the first stage of readout electronics (SQUIDs and FETs) at higher temperature stages. This structure is suspended from the upper portions during operation.

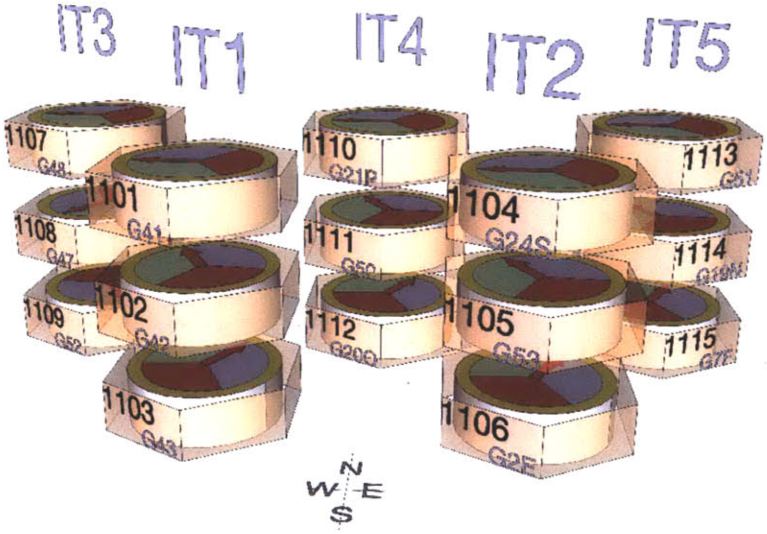

Figure 11-17: An exploded view of the 15-detector array making up the SuperCDMS Soudan experiment. There are 5 towers of 3 detectors. Within each tower, the detectors are rotated 60 degrees relative to each other. Tower 3 is unique in that $^{210}$Pb sources are installed at the top and bottom of tower, exposing the top surface of the top detector (T3Z1) and the bottom surface of the bottom detector (T3Z3).

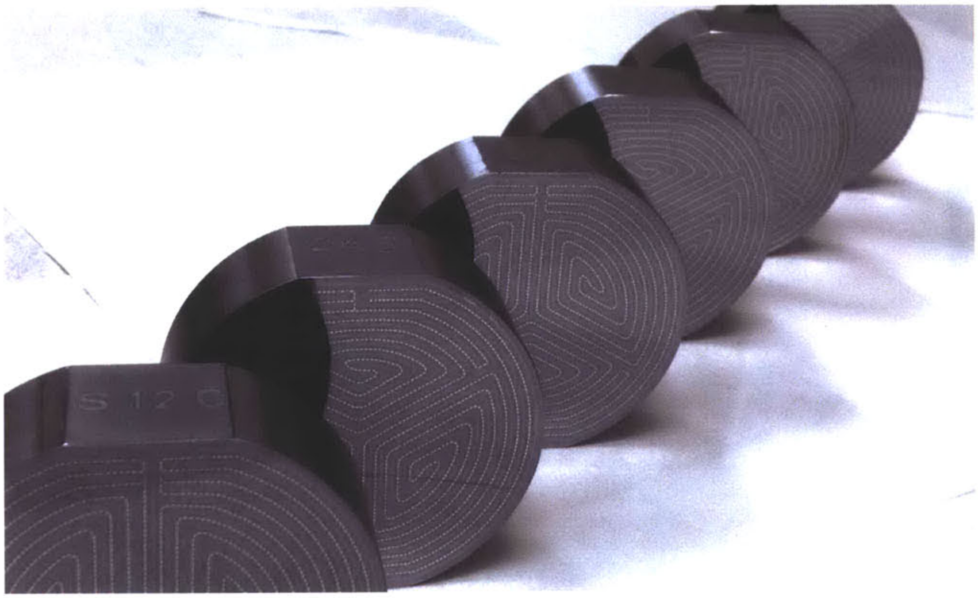

Figure 11-18: A batch of 6 iZIPs, fabricated specifically to test the fabrication throughput. These detectors were fabricated using Si substrates (as indicated by the sidewall labeling), and are interesting to study in their own right for comparison with the Ge detectors.

# Chapter 12

# Analysis of iZIP4/5 Data

This chapter draws primarily on the analysis of one particular iZIP4 detector (G48, aka T3Z1), sometimes using data from testing at Berkeley, and sometimes using data obtained after its installation underground at Soudan. This detector makes an excellent example case because it has been more thoroughly inspected than any other iZIP, partly because it was in the first fabrication batch of iZIP4/5 detectors for SuperCDMS Soudan.

In addition to $^{133}$Ba gamma sources and $^{252}$Cf neutron sources external to the cryostats at both sites, this detector has been exposed to rigorous surface event rejection tests, necessitating the installation *inside* the cryostats of sources of non-penetrating radiation. At Berkeley, this consisted of a $^{109}$Cd beta source, facing the top surface of the detector, collimated through a regularly spaced grid of holes. At Soudan, the surface events are the result of a non-collimated Si wafer that had been exposed to radon-contaminated air, imparting a $^{210}$Pb contamination into the substrate. As illustrated in Figure 12-1, the $^{210}$Pb decay chain has many products. Most important for this analysis are the betas at 15 keV, 61 keV, and 1.1 MeV, as well as the $^{206}$Pb nuclei which recoil at 103 keV and below as a result of alpha emission. The overall rates are not high for this source; the rate in T3Z1 is ~71 betas/hour and ~16 $^{206}$Pb nuclei per hour.

At the time of writing (summer, 2012), the Soudan installation is busy collecting more data and greater statistics, and the collaboration is busy creating improved

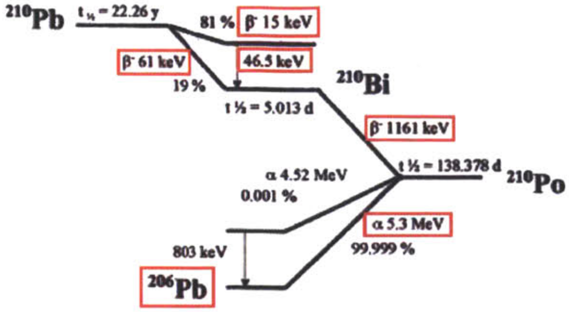

Figure 12-1: The $^{210}$Pb decay chain.

analysis methods, so this chapter should be thought of as something like a 'first look', rather than anything like a final word.

## 12.1 Overview

Much of the analysis infrastructure from CDMS II carries over quite well to the iZIP case. All eight phonon channels and all four charge channels are interpreted using an optimal filter to obtain raw pulse amplitudes. The charge pulse amplitudes are then calibrated (while correcting for capacitively-coupled crosstalk) using $^{133}$Ba calibration data. The eight phonon optimal filter amplitudes are first calibrated relative to each other by aligning the eight $E_{chan}/E_{tot}$ distributions, such that the average deposition in each channel is approximately 1/8 of the total. A small improvement to this relative calibration strategy can be achieved if instead of aligning the means of the distributions, we align the low- $E_{chan}/E_{tot}$ edges. This means setting equal each channel's fractional energy deposition *when the channel is non-primary*, which should avoid the effects of local saturation and channel geometry (each phonon channel has the same collection area, but we would still expect a primary event in the outer channels to absorb less energy than a primary event in the inner channels). This relative calibration consists of 7 scalars. An eighth number provides the absolute scaling, converting the optimal filter amplitudes (in electrical units) to energy units (keV). This last step is performed by minimizing the difference between the phonon energy and the charge energy (which has already been calibrated using $^{133}$Ba lines), implicitly using the assumption that the $^{133}$Ba events (gammas) are of fixed ionization yield. With the four charge amplitudes (two from each side), we can select only those events in the crystal bulk, and with a charge energy and a phonon energy we can construct an ionization yield, and then the job is done: a well-measured yield and a well-defined fiducial cut are the two ingredients to a perfect detector. Notice that no timing information is used at any stage, only charge and phonon pulse *amplitudes*. This makes the analysis significantly easier and more robust than the CDMS II analysis.

Now that we have listed the two key ingredients to a clean nuclear recoil sample: a fiducial volume definition and a yield definition, let's discuss each of these ingredients in turn.

## 12.2 Charge-Based Fiducial Volume Definition

In CDMS II, we did have a charge-based fiducial volume, but only in the radial direction. We eliminated all events that registered any amount of induced charge on the outer charge channel. The iZIP fiducial volume definition consists of three simultaneous cuts (an event must pass all three cuts to be considered in the fiducial volume), and two of these cuts are basically the same idea as CDMS II: **"did the event produce a significant amount of induced charge in the outer charge channel?"** The ramo potentials of the new geometry (which describe the electrode's sensitivity to drifting charges) now are largely limited to the near-surface field region (meaning drifting charges must reach this region in order to produce a significant signal). On the other hand, we now have the luxury of asking this radial question twice for each event, once on the electron side and once on the hole side. The third fiducial volume cut is the so-called 'symmetry' cut, as laid out in Figure 11-2, asking the question **"did the event register approximately equal amounts of induced charge on the top and bottom sides?"**. There are two ways in which an event could be side-asymmetric: either the event occurs in the surface field region (where no charge escapes to drift to the opposite side) or the event could suffer from significant asymmetric *trapping*. Of course, it's always possible for an event to suffer from equal trapping of holes and electrons, but this is a rare coincidence. In summary: To be considered in the the fiducial volume, an event must satisfy all of the following three criteria:

- Most of its hole-side collection must be in the inner charge channel.
- Most of its electron-side collection must be in the inner charge channel.
- Charge collection must be roughly equal on the hole side and the electron side.

We need now to define 'most' and 'roughly'. More specifically, the three-cut fiducial volume definition depends on setting four cut thresholds: a threshold between inner and outer on the hole side, a threshold between inner and outer on the electron side, and two thresholds separating the side-symmetric population from the side 1 population and the side 2 population. All four of these thresholds can be thought of as angles and as (sometimes fuzzy) thresholds in real space, as seen in Figure 12-2.

Notice how, in Figure 12-2, a symmetric cut of any threshold will eliminate a large population of events near the outer sidewall, for which one of the charge carriers is trapped (on the side wall) more than the other. In a way, these are almost exactly the events we want to be cutting: not just based on position, but based on whether significant trapping has occurred. Of course, it is always possible that trapping could be symmetric, and this population (that we earlier described as depending on a coincidence to exist) is the only low-yield population not cut by the otherwise extremely powerful symmetry cut. The two radial cuts, then, can be made much looser than in CDMS II (thereby gaining fiducial volume efficiency), because we are now eliminating high-radius events in three ways simultaneously.

The setting of these four fiducial volume cut thresholds was performed (detector-specific, for T3Z1) by scanning the 4-dimensional cut position space, and finding the point in that four-parameter space which maximizes the fiducial volume (measured by the rate of acceptance of neutrons from $^{252}$Cf calibration) while keeping leakage at zero

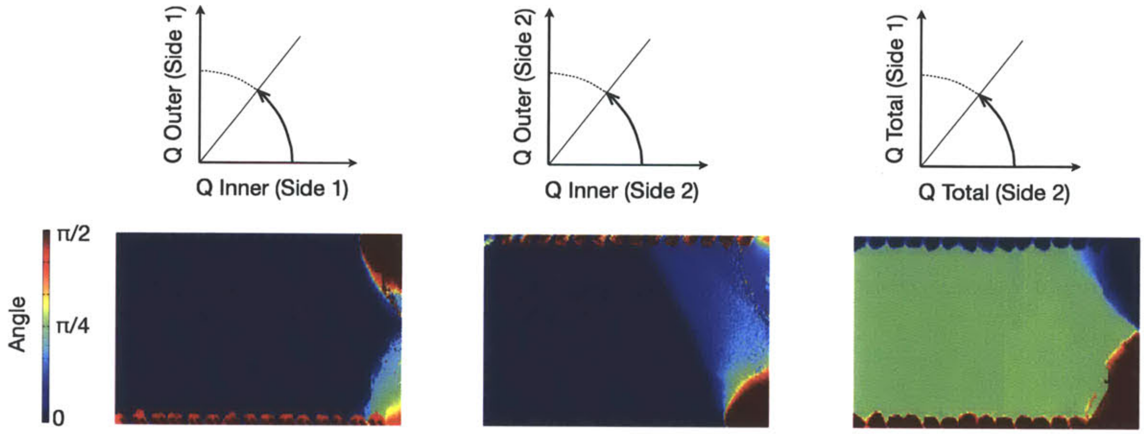

Figure 12-2: The basic fiducial volume definition for an iZIP. From left to right, we define a Side 1 angle, a Side 2 angle, and a Symmetry angle. The lower row displays results from a Monte Carlo study (slices through the crystal, R on the horizontal axis, Z on the vertical axis), showing the predicted measured values for these angles at different positions. The crystal slice is through the crystal lattice direction of maximally-oblique electron propagation, meaning that the electrons spread towards both low and high radius maximally. In this simulation, the detector is biased such that holes propagate towards Side 1 and electrons propagate towards Side 2 (and this difference in carrier type is clearly visible in the simulation plots).

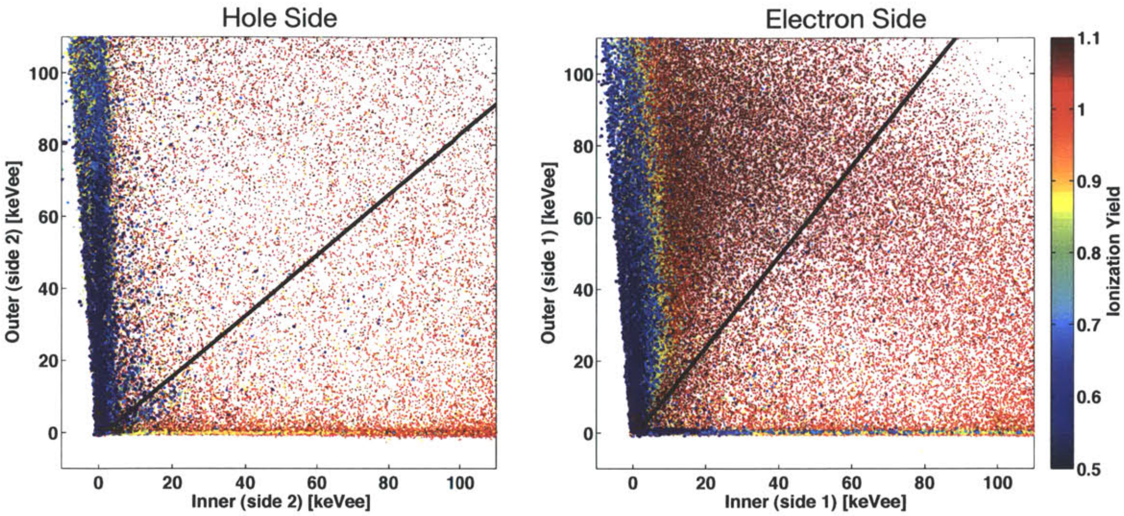

Figure 12-3: The radial charge fiducial cuts (black lines), superimposed on $^{133}$Ba and $^{210}$Pb calibration data, colored by ionization yield. Events below the black lines pass each cut. Note that *most* events lie exactly along the horizontal axis. Note also the higher population of 'shared' events as measured using the electron side, as expected by the electrons' oblique propagation.

(measured by the number of combined $^{133}$Ba and $^{210}$Pb electron recoil events leaking into the wide $\pm 3\sigma$ nuclear recoil yield band). The result of this optimization can be seen in Figure 12-3 (radial cuts for both sides) and Figure 12-4 (the side-symmetry cut).

## 12.3 Phonon Energy

### 12.3.1 The Phonon Absorption Rate

The phonon pulses of an iZIP are dramatically different from the phonon pulses we saw in CDMS II, as can be seen in the top portion of Figure 12-5. iZIP pulses are characterized by a dramatic channel-to-channel variation in the pulse shape in the first $\sim 200\mu$s or so, paired with an equally dramatic channel-to-channel similarity in amplitude during a long (much longer than CDMS II) slow decay. These pulse shapes are an expression of newly position-dependent behaviors in the iZIP design, and are a direct result of 1) the production of luke phonons almost entirely in the near-surface

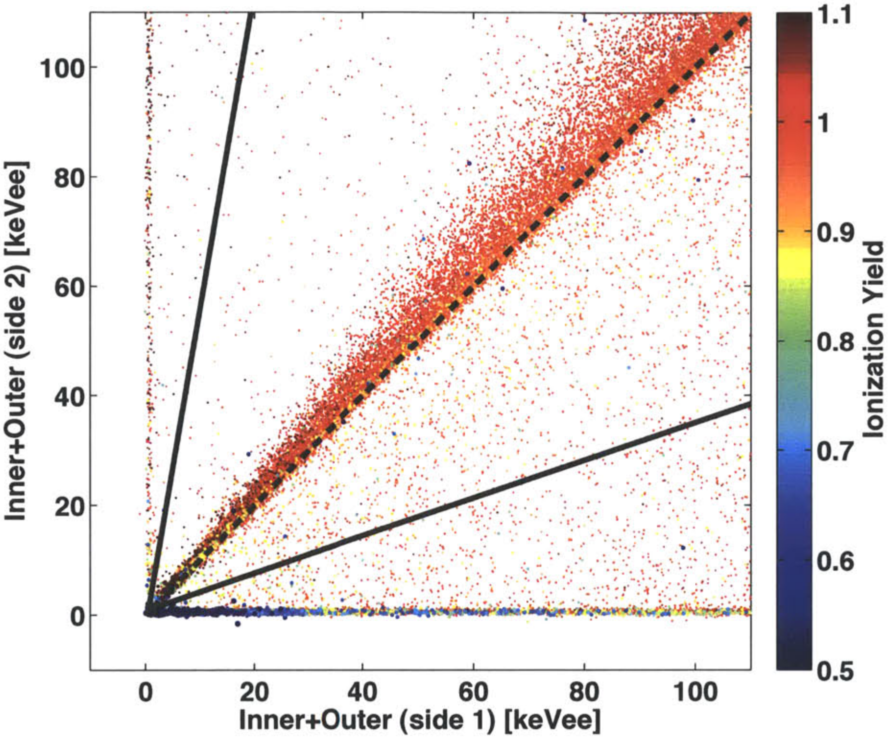

Figure 12-4: The side-symmetry fiducial cut (black lines), superimposed on $^{133}$Ba and $^{210}$Pb calibration data, colored by ionization yield. Events between the black lines pass each cut. To highlight the symmetric band, both radial fiducial cuts have been forced in the event selection. Note that the edge of the symmetric band is sharp (and near perfect symmetry) on one side, but falls off more gradually (and farther from perfect symmetry) on the other side. This is a result of the asymmetric trapping rates of holes (collected near-perfectly on side 2 here) and electrons (collected with some variable amount of trapping on side 1 here). The dashed black line at $45^o$ is shown simply to highlight this slight asymmetry. Some low-yield events pass this fiducial volume definition (as can be seen in this plot), but none of the low-yield events are as low in yield as the nuclear recoil band.

high-field region, and 2) the significantly reduced Al coverage (and therefor phonon absorption rate) in the iZIP (necessary to reduce the electrode capacitance).

In Figure 12-6, we show dozens of channel-summed pulses (signals from all eight channels are combined, after being relatively calibrated) from positions all over the detector. It can be seen that the huge pulse shape variation seen channel-by-channel is impressively canceled out when all the channels are summed, leaving only a small residual variation either upwards or downwards during the first $\sim 100\mu$s.

The long fall time of these pulses is simply a reflection (pun intended) of the high likelihood of a phonon incident on the surface to encounter a bare Ge surface (94% on the top and bottom surfaces) instead of an Al surface, where the phonon could be absorbed (6%). The taller (1 inch rather than 1cm) metal-free sidewalls only decrease this absorption rate further. We are left with a phonon absorption time constant in the iZIP4/5 of $\sim 720\mu$s (this number was closer to $\sim 200\mu$s in CDMS II). Note that this absorption rate only applies after the phonons have reached ballistic frequencies (more on this later).

### 12.3.2 The Nonstationary Optimal Filter

What happens when we naively apply the optimal filter strategy of CDMS II to these pulses? The lower half of Figure 12-5 gives us some intuition. The PSDs for several iZIP pulse shapes are shown (scaled arbitrarily), along with the phonon noise. Notice that the PSDs for the 'sharper' or 'smoother' pulses are substantially different in the amount of power around 10 kHz, but that the total power for all pulse shapes is still highest at the lowest frequencies. It is these low frequencies that supply power to the long $\tau_{fall} = 720\mu$s decay, and it is thus this decaying portion of the pulse that we want to find the amplitude of. Looking at the noise, though, there is a similar rise at low frequencies (terrible 1/f noise from the readout electronics is a problem both at the Berkeley test facility, where this data was taken, and at Soudan). Thinking of the optimal filter naively as signal/noise in this space, we would expect the filter to de-emphasize the lowest frequencies (where most of the pulse power lies), instead emphasizing the frequencies between 1-10 kHz, where the signal-noise amplitude is

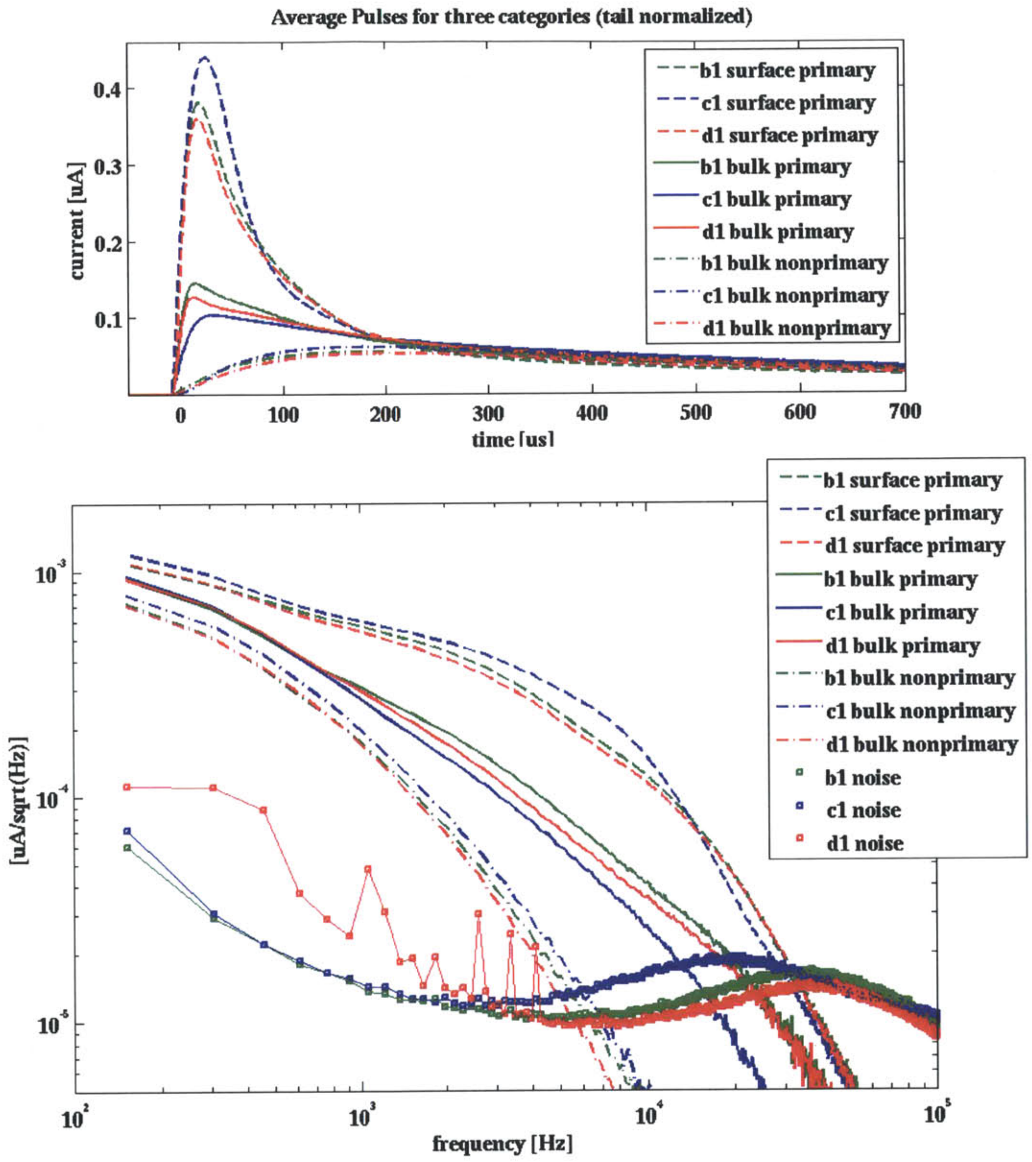

Figure 12-5: Phonon pulses from the three inner channels of side 1 of G48, selected to be at low energies (20-30 keV) so that TES saturation would not play a role, grouped by position using phonon partition quantities, normalized according to amplitude at late times, and then finally averaged to form noise-free pulse shapes. These pulse shapes are shown in the time domain at top, and as a PSD below, along with the measured noise PSD in each channel.

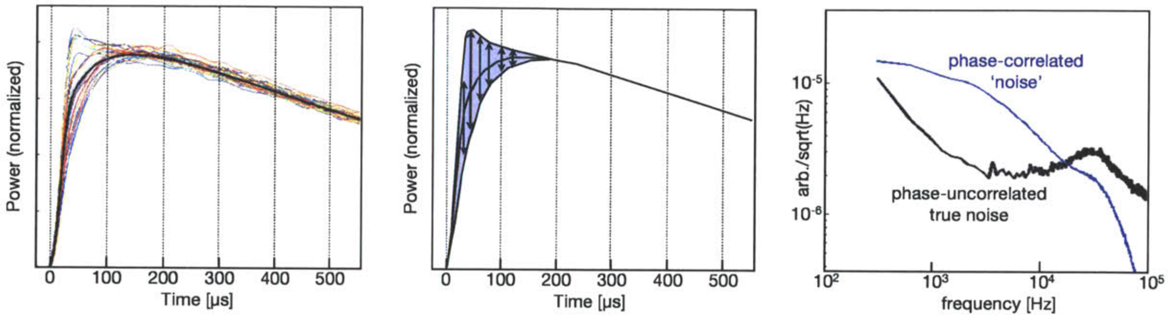

Figure 12-6: On the left is shown the total phonon pulse (sum of all phonon channels) for many events from detector G9F, an iZIP2 with somewhat reduced surface coverage than an iZIP4/5, and therefor longer falltimes. Note the extreme similarity of all total pulses after $\sim 200\mu$s, and the small but significant variations before that, both above and below a modeled pulse shape (black solid curve). In the central plot, these two pulse attributes are separated into a 'signal' template (black) and a 'coherent noise' template which can take either positive or negative amplitudes (blue). On the right is shown the power spectral distribution of the standard phase-coherent (stationary) noise (black) and this new phase-coherent (non-stationary) 'noise' from position dependence (blue). The non-stationary noise is here scaled for a 100 keV event, relative to the stationary noise. The PSD is all the information needed to describe the stationary noise, but all the phase information necessary to describe the non-stationary noise is not displayed. Note that the non-stationary noise has a significantly different shape than the stationary noise, and will cause an optimal filter to de-weight higher frequencies (at a specific time in the pulse, as determined by the template phase information). Plots after M. Pyle.

the highest.

I have tried to motivate what we have seen practically: the simplest optimal filter strategy weights the peaked portion of the pulse (which is extremely variable and tells us very little about the pulse energy) and deweights the lowest frequencies of the pulse (where the real energy information lies). Optimal filters are best suited to a signal which always matches the signal template; here in the iZIP, the pulse shapes vary widely, and the optimal filter strategy seems to fail. As recently noted (and shown in Figure 12-6), the total summed phonon pulses suffer this position-dependent shape variation much less than the channel-by-channel pulses, but the amount of variation still remaining causes significant (and unnecessary) degradation in the phonon energy resolution.

The elegant solution to applying an optimal filter to this variable signal (applied to CDMS by M. Pyle) is to tell the optimal filter what that position-dependent variation looks like, and call it a phase-correlated noise. In the CDMS II optimal filter, all noise was assumed to be uncorrelated in phase, but here we are trying to describe a 'noise' which is really a pulse in time (all contained within that first $\sim 100\mu$s). The optimal filter for this strategy, then, has three inputs: a stationary (phase-uncorrelated) noise template which describes the real noise, a non-stationary (phase-correlated, higher frequencies) noise template which describes the position-dependent variation, and a non-stationary (phase-correlated, lower frequencies) noise template which describes the position-independent pulse template, the best-fit amplitude of which is the end goal.

Introducing this non-stationary 'noise' term significantly complicates both the description of the optimal filter (I will not reproduce the expression here) and the computation time necessary for its application to raw traces. But, given the iZIP situation (energy information primarily at lower frequencies, noise primarily at lower-frequencies, position-dependence in the high signal-noise higher frequencies), the non-stationary optimal filter has been seen to significantly improve both the resolution and the pulse-shape-independence of the phonon energy measurement (as can be seen in Figure 12-7).

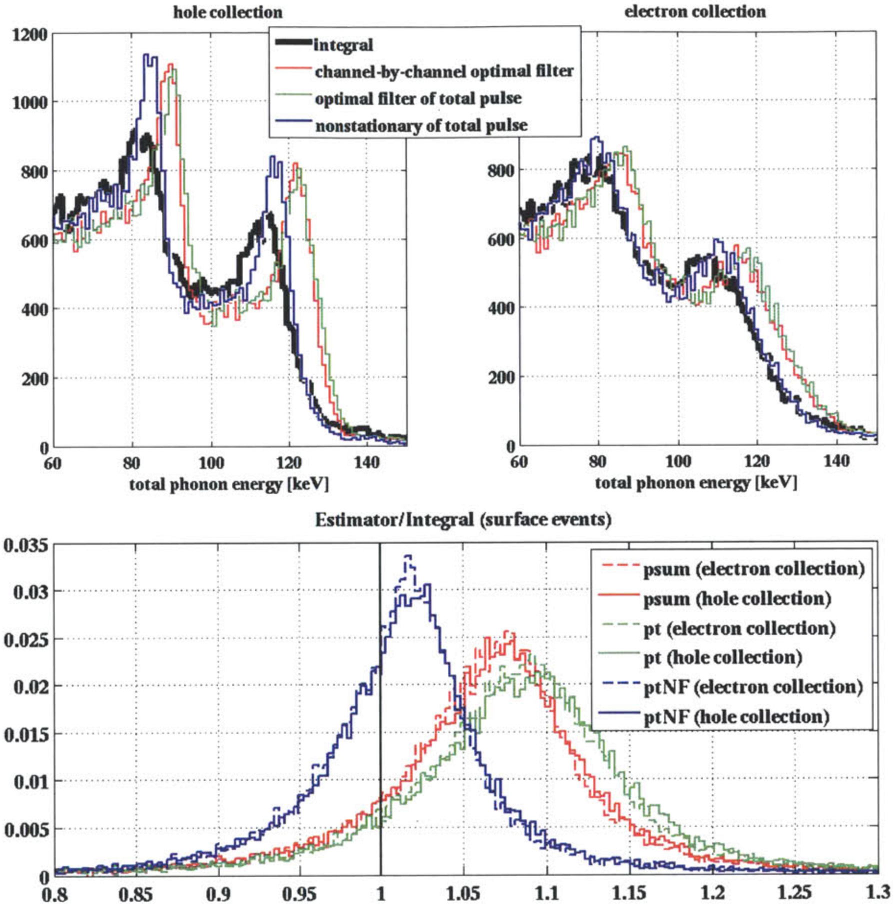

Figure 12-7: At top, $^{109}$Cd spectra (surface events selected using an asymmetric charge cut), showing both the peak widths and positions, deriving an amplitude from a raw pulse in several different ways: integral, channel-by-channel stationary-noise optimal filter (and summing the resulting energies), summed-pulse stationary-noise optimal filter (*i.e.*, applying an optimal filter to the total trace), and the summed-pulse *non*-stationary optimal filter. Notice that the nonstationary optimal filter not only gives the sharpest peaks, but also best reproduces the mean value of the integral (pulse-shape-independent) energy estimator. The offset is caused by simpler optimal filters over-emphasizing the high frequencies and therefor over-estimating the amplitude of highly peaked pulses. This is emphasized in the lower plot, where various energy estimators are compared with the (pulse-shape-independent) integral quantity (coloring is the same).

By the way, the traditional (stationary) optimal filter, which accidentally emphasizes the higher frequency portions of the pulse (*i.e.*, the peak of the pulse) is actually ideal for constructing channel-by-channel partition quantities, where the peakiness of a pulse exaggerates differences between primary and non-primary channels, and thus boosts the partition signal-to-noise.

### 12.3.3 Recoil Energy and Yield

We now have four well-measured charge energies (one for each channel) and one well-measured phonon energy (the non-stationary optimal filter value for the total detector). How do we combine them into a recoil energy and a yield?

We have effectively made the charge energy measurement twice, once with the sum of the electron-side channels and once with the sum of the hole-side channels. Which measurement is the right one? The side with the *maximum* charge amplitude will always be the better measurement, suffering from less charge trapping, and it is this maximum side measurement that is used to measure the charge energy. Of course, the maximum side is usually the hole side, but not always.

We have a charge energy measurement, now we need to use the charge signals to estimate the Luke phonon production (so that we can subtract it from the phonon energy to obtain the recoil energy). Imagine first the case of a bulk symmetric event. Its Luke production will be something between the value assuming the hole side charge measurement and the value assuming the electron side charge measurement. The portion of an event's energy that is symmetric (and crosses the full 4V difference) is the minimum of the two side measurements, and the Luke energy produced by these symmetrically propagating charges is:

$$E_{LukeSym} = 2 \times \min[\mathrm{Q1},\ \mathrm{Q2}] \left| \frac{V_b}{3\mathrm{eV}} \right| \tag{12.1}$$

On the other hand, if the event occurs in the near-surface region, less Luke production occurs, because the total voltage difference transited by the charge carriers is only half as great (for example, ±2V to 0V). We estimate the contribution of Luke energy from asymmetrically-propagating charges as

$$E_{LukeAsym} = |\mathrm{Q1} - \mathrm{Q2}| \left| \frac{V_b}{3\mathrm{eV}} \right| \tag{12.2}$$

and the total Luke energy is simply the sum of the symmetric and asymmetric estimations. Most events are either symmetric or asymmetric, but some events have contributions from both terms.

Finally, we arrive at the yield quantity

$$\mathrm{Y} = \frac{\text{charge}}{\text{recoil}} = \frac{\max(\mathrm{Q1},\mathrm{Q2})}{E_{phonon} - E_{LukeSym} - E_{LukeAsym}} \tag{12.3}$$

## 12.4 Data Quality

Because the rejection of electron-recoil events using yield and charge fiducial volume cuts is so easy, any leakage events seen will most likely be the result of failures in the data quality cuts. This is exactly where we want to be: all the information needed for perfect discrimination is in the pulse data, we just need to make sure that we are correctly turning those pulses into reduced quantities. The definition of what makes an event well-measured is where all the subtlety of the analysis lies. Here, we list the various quality cuts applied so far at Soudan, but these will surely be added to and further refined in the future.

**Good Series** During datataking, onsite and offsite shift members tag each series as 'good' or 'questionable' or 'bad' based on data quality diagnostics plots or separate knowledge as the health of the DAQ. In recent analyses, all series listed as either 'good' or 'questionable' have been included.

**Flashtime** It is important to ensure that the detector has been flashed with LEDs recently so that charge collection (*i.e.* yield) is not in a degraded state. In recent analyses, a 3-hour time limit since last LED flashing has been enforced.

**Charge bias** For simplicity, all data has been taken recently with only a single bias direction ('+/-' rather than '-/+')

**Charge and Phonon $\chi^2$** The quality of fit for the optimal filters (for phonon, both stationary and nonstationary versions) offer a powerful rejection of any pulse that does not match the standard event pulse shape (pathologies caught by these cuts include various types of noise triggers, electronic glitches, and multiple-scatters).

**Charge and phonon prepulse** The standard deviation of the portion of the trace before the trigger can be used to flag either unusual noise conditions or a pileup event (*i.e.*, a trigger while the baseline is still falling after a previous event).

**Glitch** The glitch cut, as it's currently implemented, removes events in which a non-physical number of detectors contain pulses above a certain amplitude. This is usually the result of an electronic glitch, rather than a real many-detector scatter. The definition of this cut is still in flux. Ideally, we will eliminate glitches through their pulse shape, perhaps by defining an optimal filter template for glitches and cutting on its associated fit $\chi^2$.

**Veto Coincidence** Events within $50\mu s$ of a veto trigger are removed. 'Veto' here refers to the experiment-surrounding scintillator veto, intended to tag muon events.

**Multiples** Dark matter will never recoil in more than one detector, so this is an important discrimination cut. It also is a data quality cut, which has been seen to catch certain types of detector-to-detector cross talk.

## 12.5 Basic Discrimination Ability

Plots visually representing the discrimination of surface events (based on combined yield and chage-based fiducial volume) in T3Z1 at Soudan are shown in Figures 12-8,12-9,12-10,12-11, and 12-12.

After application of the charge-based fiducial volume cut (radial side 1 & radial side 2 & side symmetric), the nuclear recoil passage fraction was $\sim 0.63$, integrated over the 8-100 keV energy range. Only a very small amount of $^{252}$Cf calibration data has so far been taken, so there are large statistical uncertainties in this number. There is also a systematic uncertainty, caused by some unknown amount of contamination of low-yield electron recoils in the nuclear recoil sample. These near-surface electron recoils will certainly be cut by the fiducial cuts, artificially lowering the nuclear recoil sample's passage efficiency.

Estimating the leakage rate of electron recoils through the combined yield and fiducial volume cuts is hard for the happy reason that no leakage has yet been seen. Yield discrimination alone for bulk electron recoils can be taken as perfect. The beta and $^{206}$Pb nuclei surface events similarly show no leakage through the fiducial volume cuts. As can be seen most clearly in the yield vs. charge z partition plane of Figure 12-12, the $^{206}$Pb population appears to be approaching the symmetric nuclear recoil 'box' in the charge partition axis, but this is to be expected (surely, once more statistics are taken, some $^{206}$Pb recoils will be correctly measured to be charge-symmetric events). The expected $^{206}$Pb background rate is extremely low, however (the associated alpha rate at Soudan has been measured at 0.07±0.01 per inner channel per day), so any leakage (again, not yet seen) from $^{206}$Pb nuclei is not a real concern.

The real question we want to ask is 'what is the leakage rate of betas?' So far in T3Z1, in 37.6 days of exposure to the $^{210}$Pb source, we have obtained 79,059 surface events, of which an estimated 64,511 are believed to be betas based on their yield. To distinguish betas from $^{206}$ recoils, we are simply drawing a line in the sand at yield=0.4 (betas above, $^{206}$Pb recoils below 0.4).

**With a recoil energy threshold of 8 keV, and 79,059 surface events collected, zero leakage into the signal region was seen, implying an upper limit on the beta leakage rate (above 8 keV) of $< 2.9 \times 10^{-5}$ at the 90% confidence level.**

Although a longer exposure at Soudan may push this discrimination limit further down, the confusion between $^{206}$Pb recoils and beta recoils will eventually restrict the experiment's capacity to measure a beta-only discrimination ability. It is expected that many events currently appearing close to the signal region will be excluded with future quality cuts.

## 12.6 Position from phonon energy partition

In the previous section, we stated the conclusions of a simple iZIP analysis. Given a definition of ionization yield and fiducial volume, we have an efficiency (~63%) and a leakage ($< 2.9 \times 10^{-5}$). We call this a *simple* analysis because we have used only a fraction of the information available to us about each event. We have essentially performed the thermal measurement, where all the many types of phonon information available to us have been ignored, and only one phonon-derived number (the total phonon energy) has been used. But, we went to so much trouble to design a detector which collects phonons athermally. The rest of this chapter discusses the measured information content of these athermal phonons, and how phonon information alone can be used to determine both the fiducial volume and yield. We have so much information for each event, in other words, that we can perform the entire discrimination analysis twice, once based on the charge measurement for position and yield, and once based on the phonon measurement for position and yield.

The low Al surface coverage of the iZIP (6.2%) is both a blessing and a curse. It is a curse because it draws the pulse shape out to millisecond time scales, far longer than the ideal range of the TES (also making pile-up more likely). It is a blessing because the low probability of phonon absorption per phonon surface interaction acts as a natural filter, separating phonons that are of extremely short mean free path near the surface (which, will recoil against the surface over and over again, and therefor be efficiently absorbed) from low-energy ballistic phonons (which will recoil against a surface only once, most likely recoiling off a bare Ge surface, and then only much later, after propagating the macroscopic distance of the crystal thickness, finding another surface and another chance for absorption). The low surface coverage, then, leaves the highly-efficient absorption of high-energy phonons unchanged, but dramatically slows down the absorption of low-energy ballistic phonons. This separation of the two types is clearly seen in Figure 12-5.

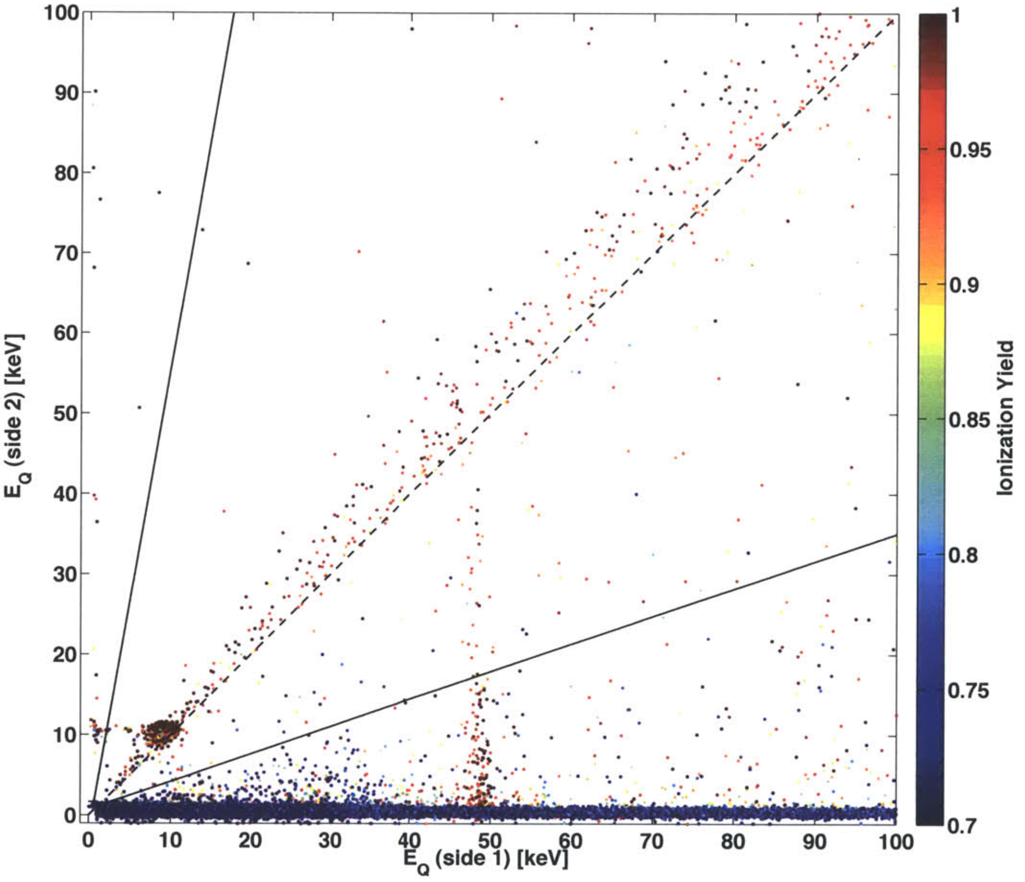

Figure 12-8: Charge Side 2 vs Side 1, showing the result of Side 1 exposure to a $^{210}$Pb source. Along with the Side 1, Symmetric, and (sparse) Side 2 populations, several other features are visible, including the 45.6 keV gamma population penetrating from the Side 1 surface into the bulk, and the 10 keV (bulk) Ge activation line. The two radial fiducial cuts have been applied here, to visually accentuate these features.

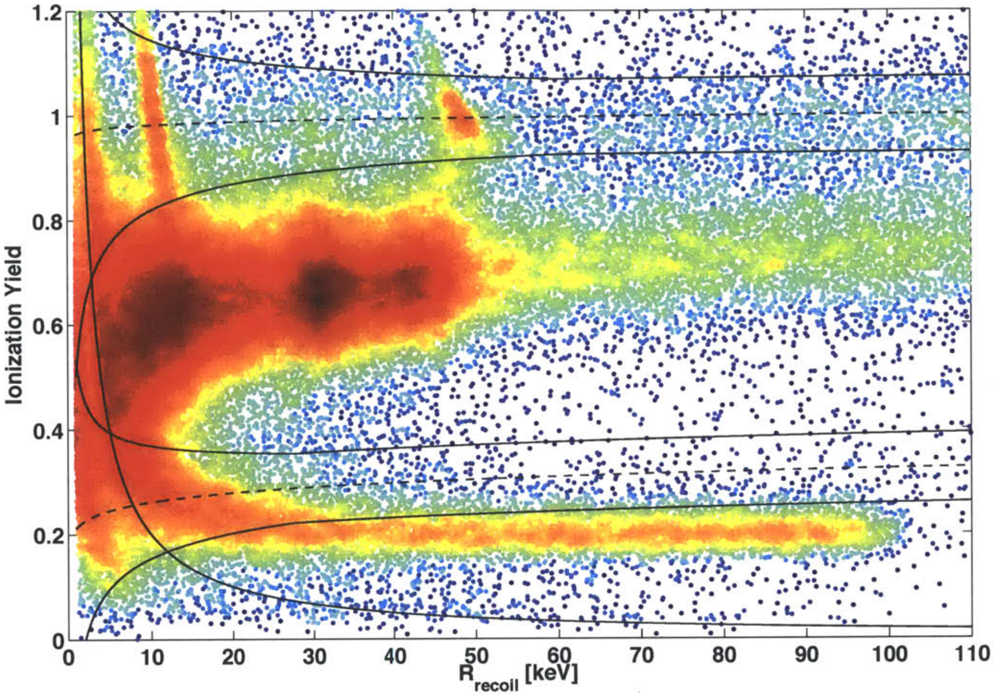

Figure 12-9: The yield vs. recoil energy plane for an iZIP at Soudan for which Side 1 is exposed to a $^{210}$Pb source (exactly the same data as in Figure 12-8, with the radial fiducial cuts similarly applied). Many expected features from the source are seen: the tail of the 1.1 MeV beta population, a 103 keV $^{206}$Pb population and its tail (at yield of ~0.2), the 45.6 keV gamma population (at yield of 1), beta lines at 61 and 15 keV, and the 10.3 keV Ge activation line (at yield of 1). Charge-symmetric electron recoil and nuclear recoil $\pm 2\sigma$ bands are shown in black, as fitted to $^{133}$Ba and $^{252}$Cf calibration data, respectively. Coloring in this plot is by (log) density.

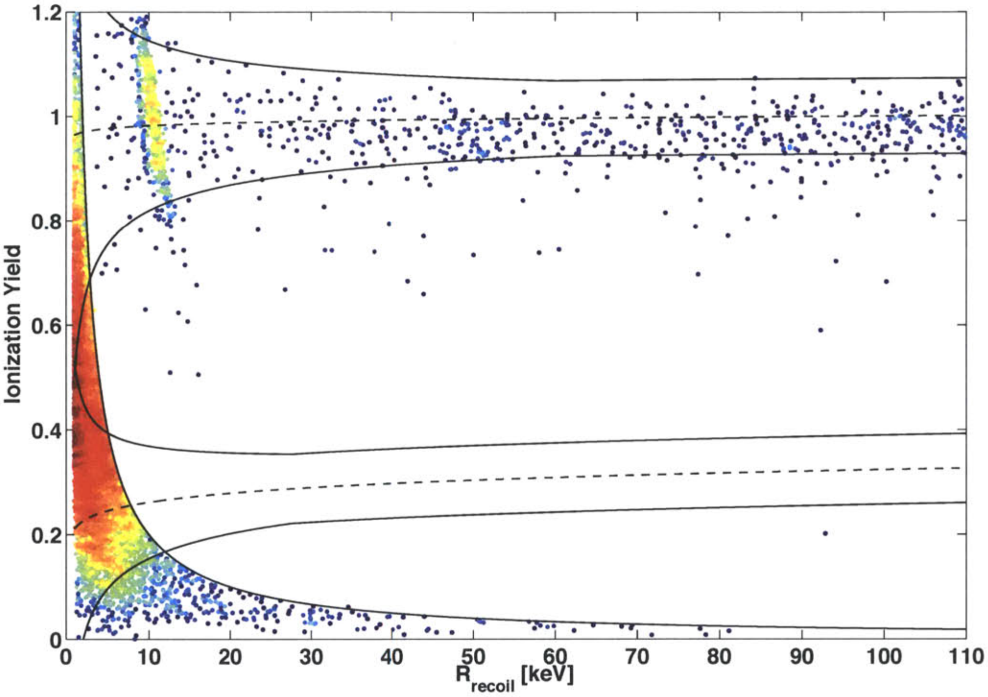

Figure 12-10: Same as Figure 12-9, but with all three charge-based fiducial volume cuts defined (inner side 1, inner side 2, and side-symmetric). The surface event populations have all but disappeared, and no $^{206}$Pb events (either betas or nuclei) appear with in the $\pm 2\sigma$ nuclear recoil band. The 2 keV charge threshold is abrupt here because the charge fiducial volume cuts are defined such that they pass all events below that threshold.

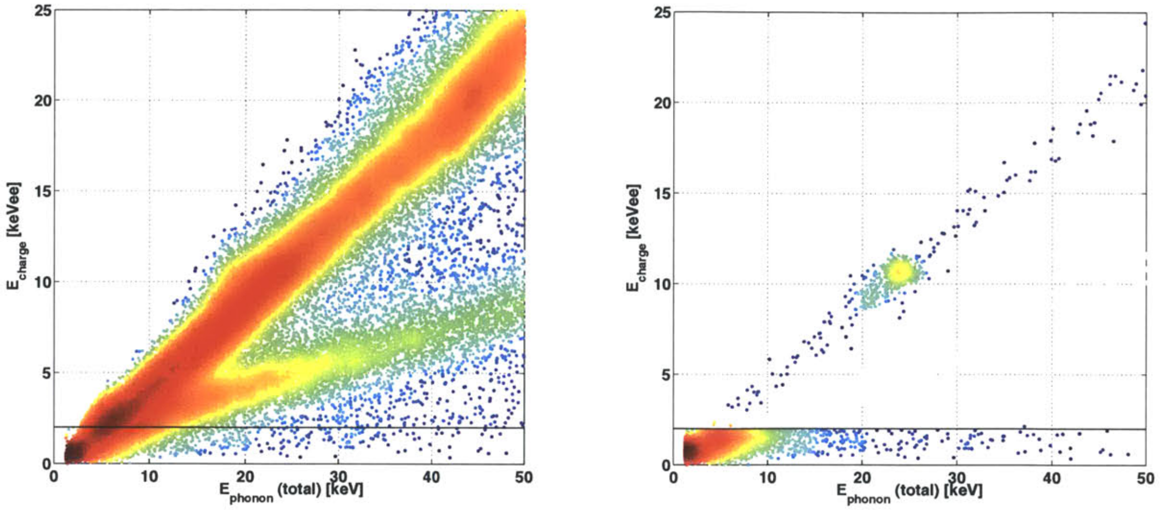

Figure 12-11: These plots are the same as Figures 12-9 and 12-10, but in the space of charge energy vs. phonon energy (and zoomed in to lower energies). The neutron calibration population is shown underneath in light grey.

Coupled with these naturally different absorption rates, the iZIP surface field produces a population of highly energetic, highly diffusive Luke phonons right near the surface, right where they will have a minimal mean free path lengths and therefor a maximal absorption rate. Luke phonon energy is produced in proportion to the voltage traversed, and in an iZIP this voltage difference occurs largely in the near-surface (~1mm) region (as seen in Figure 12-13. Not only is the majority of Luke energy produced in this region (near the QETs), but the Luke phonons given off in a high-field region are of significantly higher energy (and therefor much shorter mean free path length) than the Luke phonons given off in the low-field bulk. This is shown graphically in Figure 12-14.

Because charges propagate either in side-symmetric or side-asymmetric paths, that means Luke Phonons, which result directly from charge propagation, must similarly be produced in either side-symmetric or side-asymmetric proportions. Just as we define a symmetry band in a space of *charge* side 1 signal vs side 2 signal, we can define a symmetry band in a space of *phonon* side 1 vs side 2 signal, as seen in Figure 12-15. Luke phonons are only a fraction of the total phonon energy, so no events are entirely single-sided and entirely asymmetric (as was the case for charge

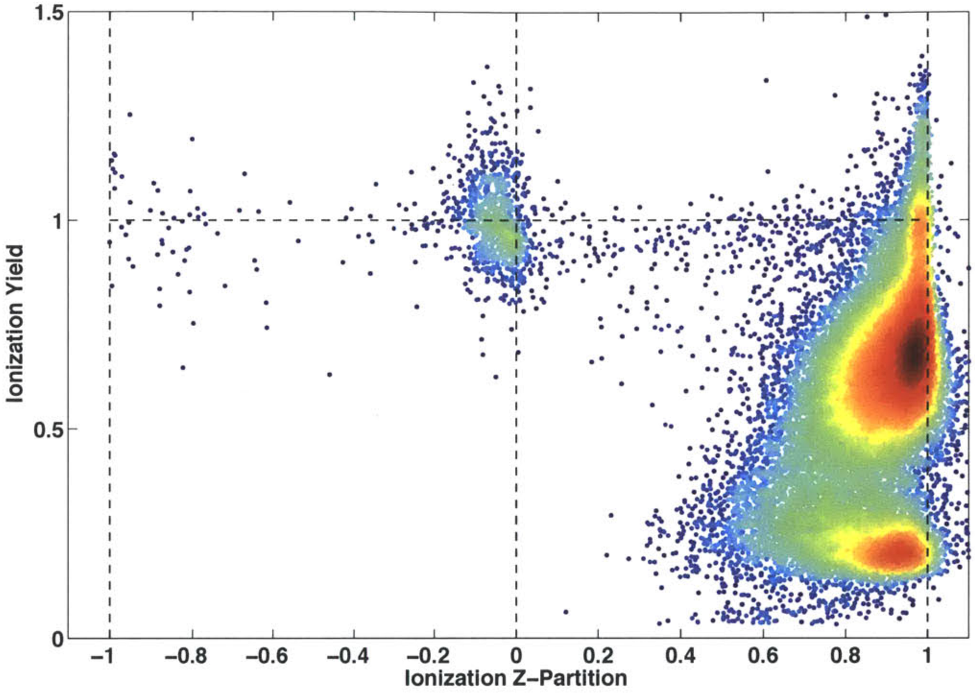

Figure 12-12: The plane of the two primary discriminators of an iZIP: ionization yield on the vertical axis vs. charge symmetry on the vertical axis. A low-energy threshold of 7 keV (recoil) is enforced here. The nuclear recoil yield varies in this view (combining a wide range of energies), so the 'box' is shown simply through a population of neutrons from calibration, in light grey. As in previous plots, color indicates simply the (log) density of events in the plane of the plot. The high-yield symmetric electron recoil population is seen at top center, along with the far-dominating surface event populations (near z-partition of 1) of betas (yield∼0.7) and lead recoils (yield∼0.2). Notice that the 'stragglers' in charge z-partition are of high yield. Charge energy (and thus, signal-to-noise of the z-partition measurement) decreases with yield, leading to the much wider surface event z-partition blob at lower yield than higher yield.

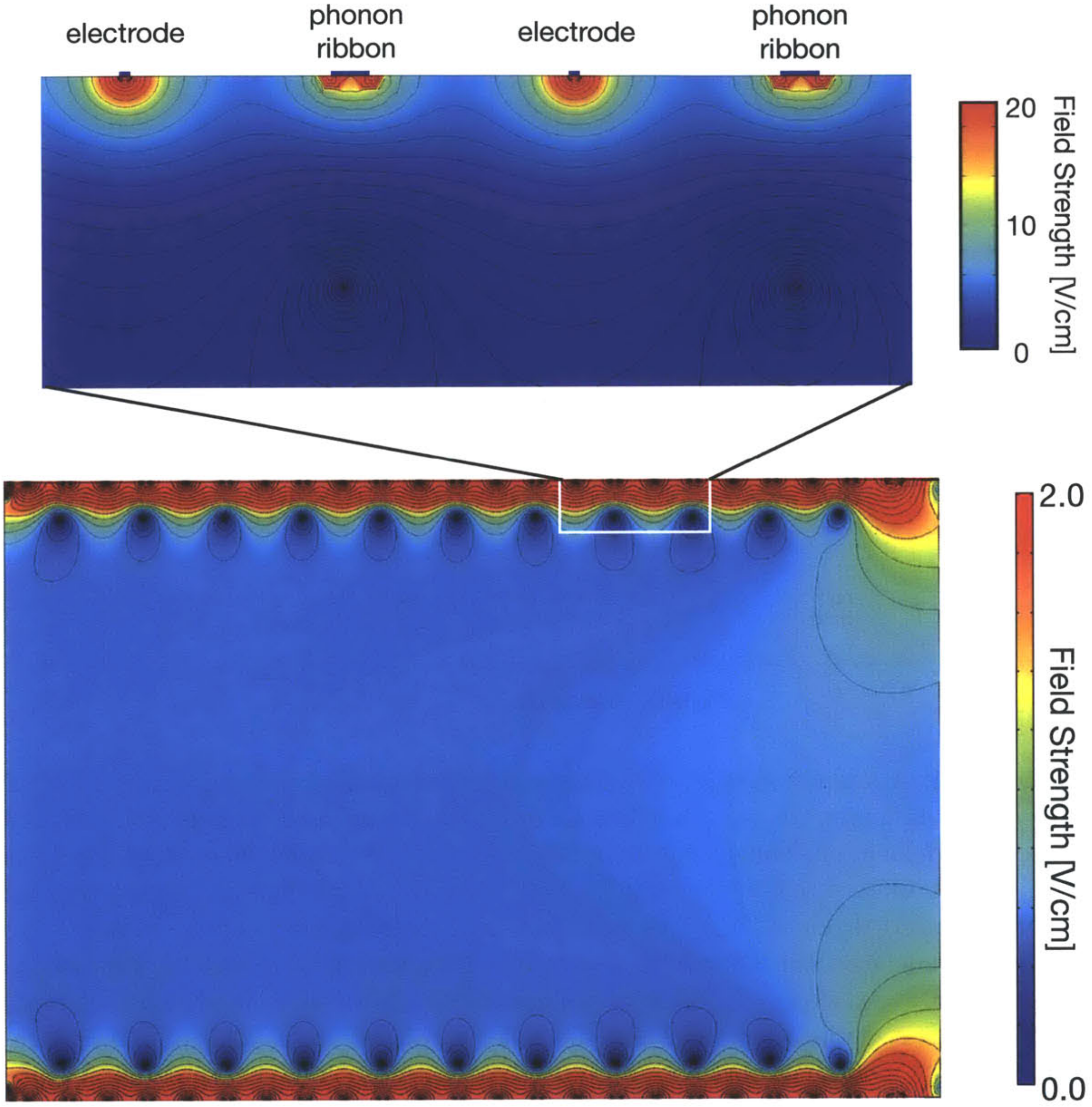

Figure 12-13: The electric field strength as a function of position within an iZIP5 (40$\mu$m electrode line widths). Regions with electric field strengths greater than 2 V/cm are uniformly colored red. In the region immediately surrounding the electrodes, this electric field strength can reach a hundred times that value. These plots can be thought of as showing where in the crystal the Luke phonons of an event are produced.

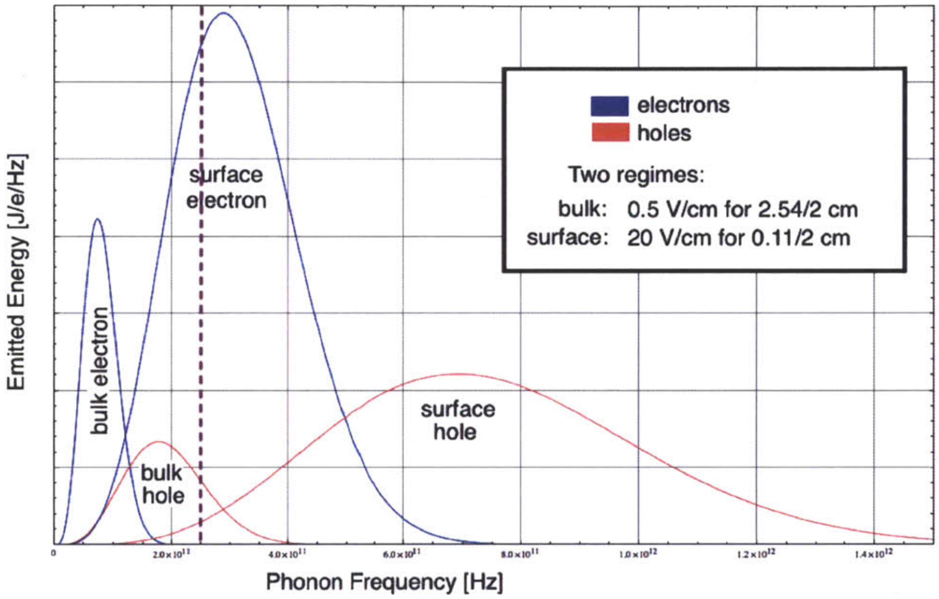

Figure 12-14: A simplified view of Luke phonon production for a symmetric event. A dashed line indicates the approximate boundary between short-wavelength diffusive phonons and long-wavelength ballistic phonons. In this model, holes and electrons propagate through a 0.5 V/cm bulk for half the width, and then propagate through a thin (0.55mm) region of 20 V/cm near the surfaces (of course, as seen in Figure 12-13, the real surface field is far from constant). This simplified model, though, shows several features we expect in the real situation: the electrons emit phonons at lower frequencies (*i.e.* more ballistic) than the holes emit, the total Luke phonon energy emitted in the near-surface region dominates the Luke phonon energy emitted in the bulk, and the Luke phonons emitted in the surface region are largely diffusive at creation (whereas the Luke phonons emitted in the bulk region are largely ballistic at creation). These Luke phonon spectra were derived using expressions from Wang *et al.* [108].

events), but there is still a clear gap in the distributions between the bulk-field events and the surface-field events.

Interestingly, the fidelity of this surface vs. bulk separation in Z actually improves at lower energies (as seen especially in the bottom side events in Figure 12-16). Surface events produce Luke phonons over a very small volume of the crystal (the electrode-ground spacing is only 1.2 mm), and this focused local energy tends to saturate the QETs near the event location, flattening the pulse shape and suppressing the measured phonon pulse amplitude and the measured phonon z-partition. At lower energies, where QETs saturate less, the measured phonon energy is somewhat higher, and the separation between surface and bulk in z-partition is somewhat greater. Already we are seeing that the strength of the phonon-based approach lies at the low-energy regime, where the extreme sensitivity of TESs can be taken advantage of.

Notice that we are simply remeasuring exactly the same propagation of charges twice now: once by observing their induced charge on the electrodes, and again by observing their resulting Luke phonons. As expected by the fact these two effects are produced by exactly the propagation of exactly the same charges, the charge measurements and the (Luke) phonon measurements are highly correlated, as seen in Figure 12-17. There is one significant hitch to the clean observation of Luke phonons: they are only a fraction of the total phonon energy created by an event, and the presence of primary phonons to some extent masks the beauties of the Luke phonon position measurement. This is particularly the case when the yield is low (and Luke phonons are a smaller fraction of the total). With a reduced Luke energy, the phonon z-partition for nuclear recoils is based largely on primary phonons, and therefor the z-partition distribution for bulk events is less symmetric (*i.e.*, wider in z-partition), the beautiful bulk-surface separation seen in Figure 12-15 and Figure 12-16 is no longer present, and the bulk nuclear recoil population overlaps (somewhat) with the surface electron recoil population (as seen in FIgure 12-19).

So far we have been discussing the phonon information of z-position, and we have mentioned how low-yield events have a broader partition distribution (are less side-

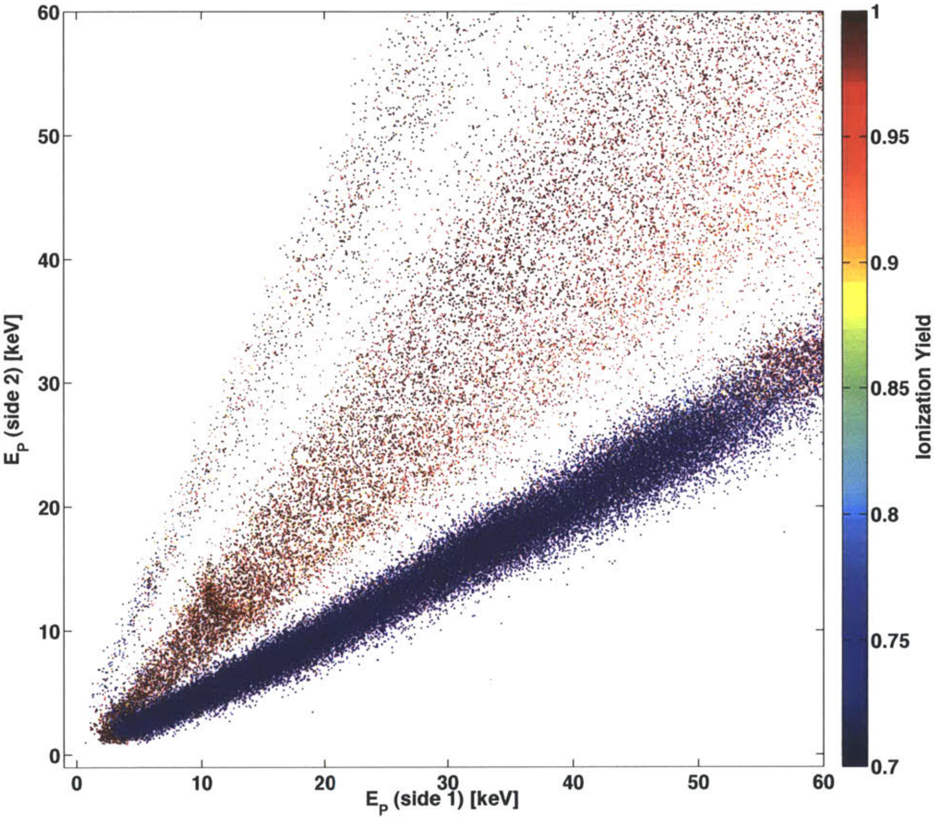

Figure 12-15: The phonon energy measured in Side 1 vs. Side 2. In this and the following plots, Soudan $^{210}$Pb data has been combined with Soudan $^{133}$Ba data (gammas in both the bulk and the surface) to show the full distributions of electron recoils. Note that a yield cut (yield must be above 0.4) has been imposed on both samples to eliminate the surface $^{206}$Pb nuclear recoils. Note also that only one bias (Side 1: +2V) is shown in these plots. The 10.3 keV activation line is the clearest feature observed here. Compare with the charge-based versions of this plot in Figure 12-4 and Figure 12-8.

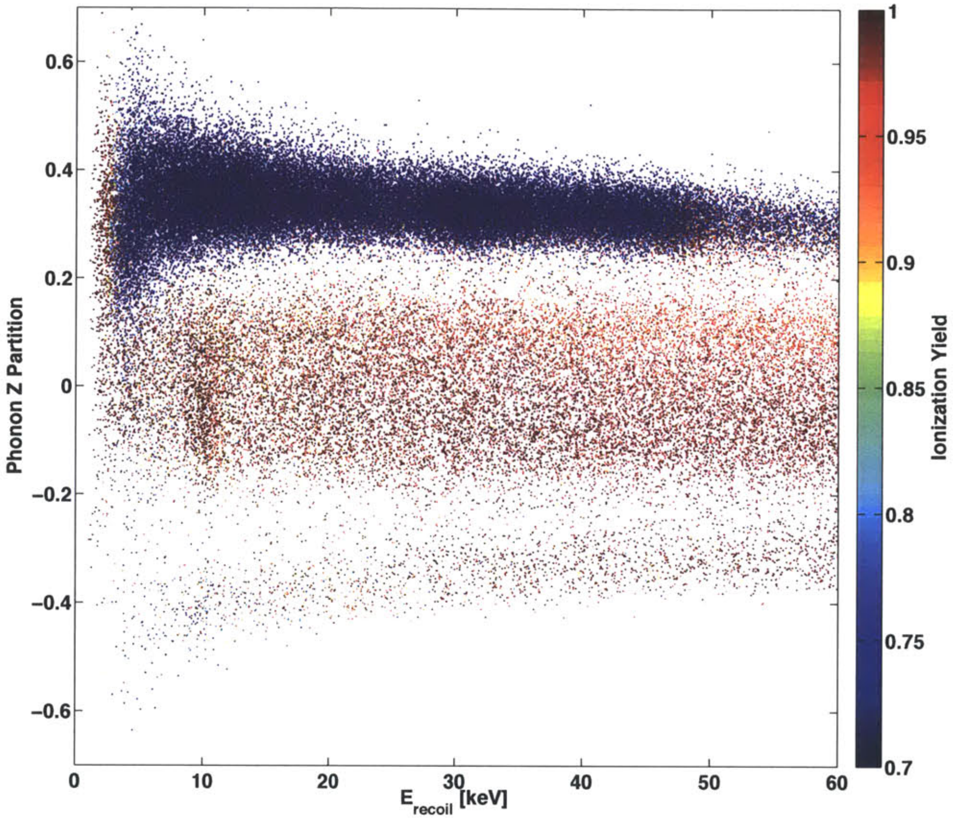

Figure 12-16: The same information as in Figure 12-15, but projected differently. Here, the horizontal axis is the recoil energy (estimated using the nonstationary optimal filter phonon energy subtracting off the Luke energy estimated from the charge measurement) and the vertical axis is phonon z-partition, defined as (Side1-Side2)/(Side1+Side2). Notice how the separation of the surface population increases at lower energies, indicating that the higher energy surface events are suffering from underestimated primary-side amplitudes caused by local saturation of the TESs.

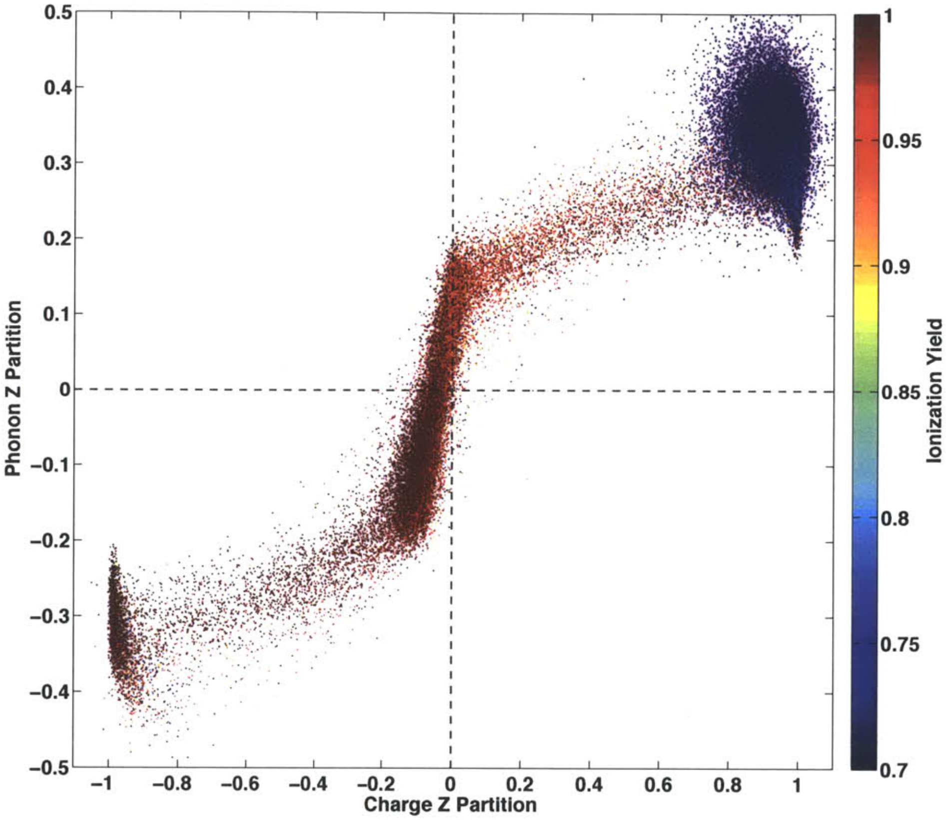

Figure 12-17: Here, the vertical axis is the same as Figure 12-16, but the horizontal axis is the *charge* z-partition, defined similarly as (Side1-Side2)/(Side1+Side2). Notice the strong correlation between the two measurements, but with a higher degree of separation in charge. Notice how the charge signal is perfectly symmetric only when the z position (as measured using the phonon z-partition) is strongly towards the electron-collection side. Notice, too, how the 'straggling' population of events between the symmetric and asymmetric population is common to both charge and phonon z-partition. Lastly, notice how (most clearly visible in the low-statistics Side 2 population), surface events vary only slightly in charge partition but vary significantly in phonon partition (presumably indicative of different rates of phonon absorption depending on event location within the surface structures).

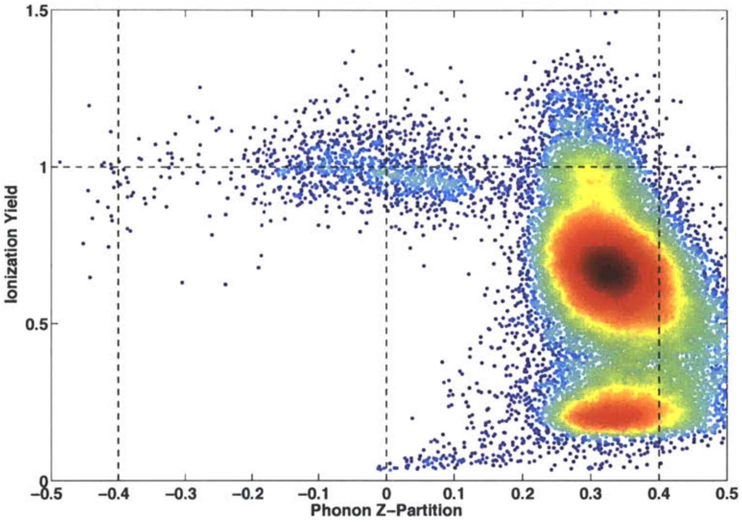

Figure 12-18: The same as Figure 12-12, but this time using phonon z-partition instead of charge z-partition. In broad strokes, the phonon distributions are the same as the charge distributions, but less distinct due largely to the presence of non-Luke phonons. The bulk nuclear recoil population has a much wider phonon z-partition distribution than the bulk electron recoil population, because the Luke phonons (which the side-symmetric collection of which pushes the bulk events towards symmetric partition values) are much reduced in the nuclear recoil case. Here, as in Figure 12-12, a loose radial cut in charge has been applied, the surface event tails toward symmetry are reduced under tighter radial cuts. Further, if loose versions of both phonon and charge fiducial volume cuts are applied, outliers due to mismeasurement can be reduced.

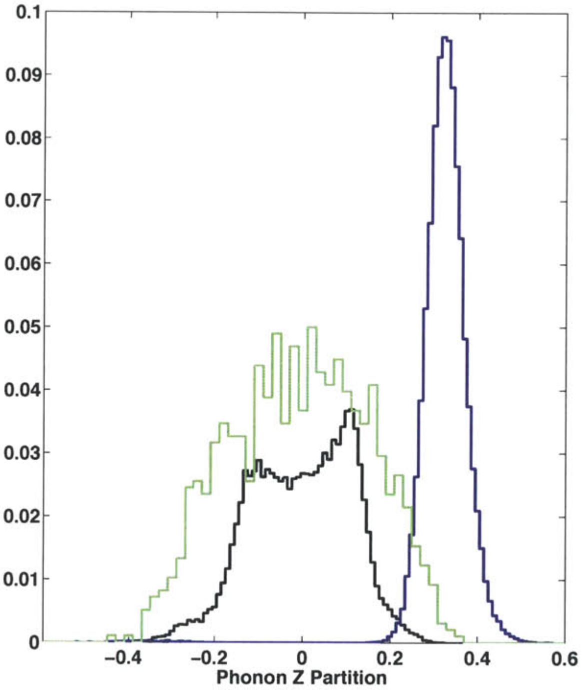

Figure 12-19: Charge-symmetric Ba calibration data: black, Charge-symmetric Cf calibration data from the NR band: green. (Charge-asymmetric Ba calibraion data: blue) Notice the wider phonon Z partition for nuclear recoils, resulting from their fractionally smaller Luke phonon population.

symmetric). The opposite is true in radial partition, as seen in Figure 12-20. In phonon radial partition, the Luke phonons tend to exaggerate the partition quantities, because the Luke contribution focuses the phonon energy in the channels directly above and below the events (the 'primary' channel), while the recoil phonons spread out more spherically (ignorant of the field direction). The lower the Luke contribution, the more radially 'shared' the total phonon energy is, and the less extreme the measurement of radial (or x-y) partition. This is unfortunate, because it means the events at highest radius (with low yield due to charge trapping on the sidewall) will have a suppressed phonon radial partition measurement.

## 12.7 Ionization yield from phonon pulse shape

The partition of energy between channels has been seen to clearly distinguish bulk events from surface events in both radius and Z. In this section, we show how the other half of the measurement, ionization yield, can be measured directly using the rising edge slope of the pulse, which in an iZIP is a direct measure of the Luke:Primary ratio.

### 12.7.1 Review of Phonons in a iZIP Detector

- **Primary phonons** produced by the recoil event itself are initially highly energetic ($\nu >$1 THz, ie $\ell < \sim 100\mu$m). If the event occurs far from a sensor surface, the diffusive behavior slows the arrival at the surface and lengthens the eventual mean free path, slowing the absorption rate at that surface. If the event occurs near a sensor surface, there is little delay in arrival, and the comparatively short mean free path increases a the rate of absorption at that surface.

- **Neganov-Luke phonons**[87, 80] are created as charge carriers are drifted by the electric field. Such phonons are created very soon after the event time (the charge drift time is $\sim$ 1 $\mu$s). In the low field of the detector bulk, Neganov-

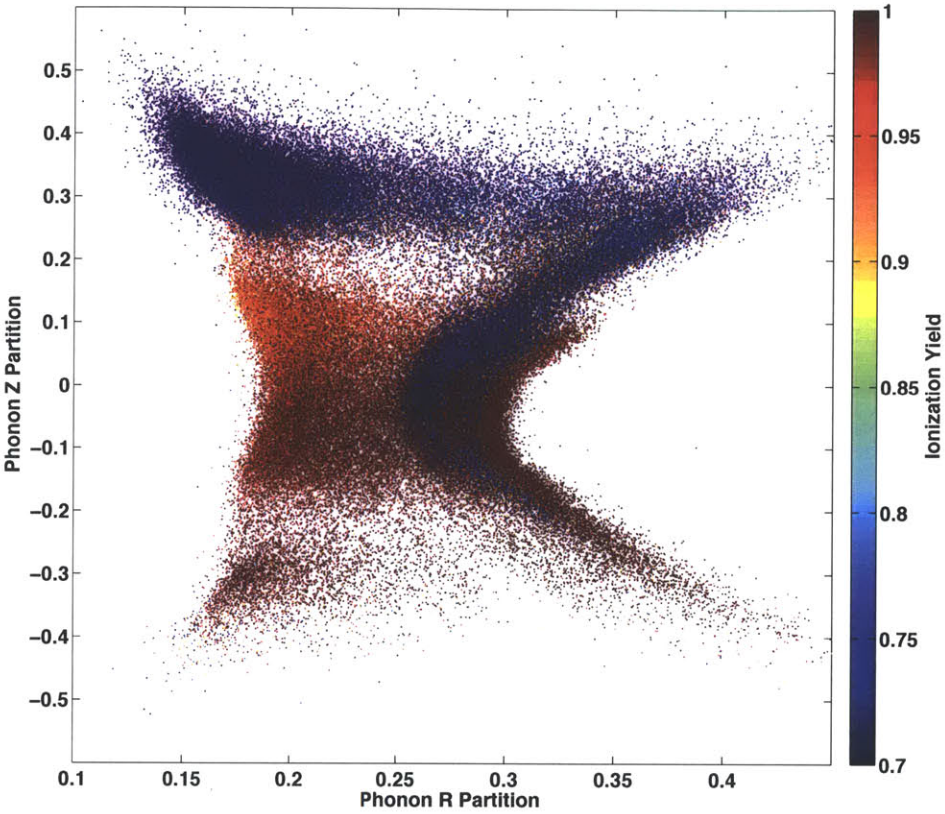

Figure 12-20: Yet another plot showing phonon z-partition on the vertical axis, this time with phonon radial partition on the horizontal axis. This is something like a cross-sectional view through the detector, with the central fiducial (high yield) portion of the crystal at the left center, surrounded at top, right, and bottom, by the top surface, outer sidewall, and bottom surface (all regions of suppressed yield). The location of the $^{210}$Pb source is clearly visible through its associated population of side 1 and side-wall betas. Notice how the sidewall betas, however, are at an unfortunately reduced radial partition value (compared with their high yield neighbors at slightly lower event radius), indicative of how suppressed Luke phonon production weakens phonon radial (or x y) position information. Radial partition here is defined as (outer s1 + outer s2)/(sum of all channels).

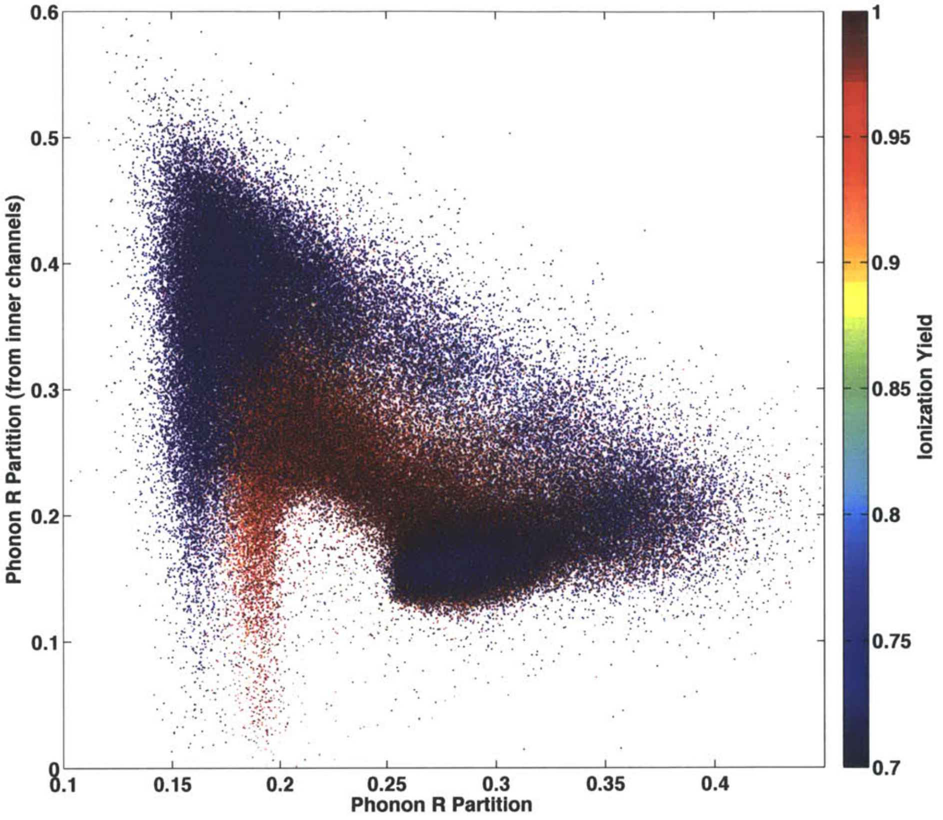

Figure 12-21: Here, two different types of radial information are plotted vs each other. The horizontal axis is the same as in Figure 12-20: the fraction of the phonon energy absorbed by the outer channels. The vertical axis is a radius as measured by the partition only among the inner six phonon channels, similar as was done among the quadrants of CDMS II. Going out from the center, first the inner-channel partition becomes more and more extreme, then, when events are at a high radius, going to higher radius makes the inner channel partition less extreme (the inner channels all end up equally non-primary). Starting at high radius and moving in, the inner-outer partition gradually becomes more and more inner, then stays fixed at the same value (the 'inner event' value) for all positions within some inner radius. Notice that here, just like in CDMS II, surface events always have a more extreme value than bulk events, either extremely inner or extremely outer.

Luke phonons are emitted with low energies and correspondingly ballistic free paths, whereas the majority of Neganov-Luke phonons are produced in the strong near-surface fields at higher frequencies (~300 GHz and ~700 GHz for electrons and holes) with correspondingly short mean free paths (~1.8 cm and ~600 $\mu$m)[108, 75] and fast absorption rates.

These three phonon populations combine together to produce the observed phonon pulses and their associated rising and falling timing characteristics. A recoil occurring in the crystal bulk will produce a roughly side-symmetric signal of near-surface Neganov-Luke phonons, which will arrive at both the top and bottom surfaces at fast times ($\sim 1 < t < \sim 5$ $\mu$s) and are absorbed rapidly by these surfaces. Primary phonons will reach the top and bottom surfaces more gradually, as quasi-diffusive propagation allows ( $\sim 1 < \mathrm{t} < \sim 15$ $\mu$s). These primary phonons will have shorter mean free paths (and will be more quickly absorbed) at the closer surface, and will have longer mean free paths (and will be more slowly absorbed) at the further surface.

The initial phonon energy spectrum and subsequent rate of absorption in a QET surface depend then on three major event characteristics: 1) the distance between the event and the surface as compared with the downconversion rate of the primary phonons, 2) the initial quantity of charge carriers produced, and 3) the starting position of these charges within the varied field geometry. Here we describe a simple toy model for relative scales of the three phonon populations. There is a large uncertainty (parameterized here as $\alpha$, where $0 \leq \alpha \leq 1$) in the creation efficiency of relaxation phonons, due to an unknown Ge-Al crossing transmittance (which is perhaps quite low for charges at drift velocity) and also due to the sensitivity of the subsequent cascade process to subtleties of the local Al and Ge environment. The total phonon energy can be described in terms of an event's recoil energy through a simple model,

$$\begin{aligned} E_{phonon} &= E_{primary} && + E_{luke} \qquad (12.4) \\ &= (1 - E_{gap}/\epsilon)E_{recoil} && + (e\Delta V/\epsilon)E_{recoil} \end{aligned}$$

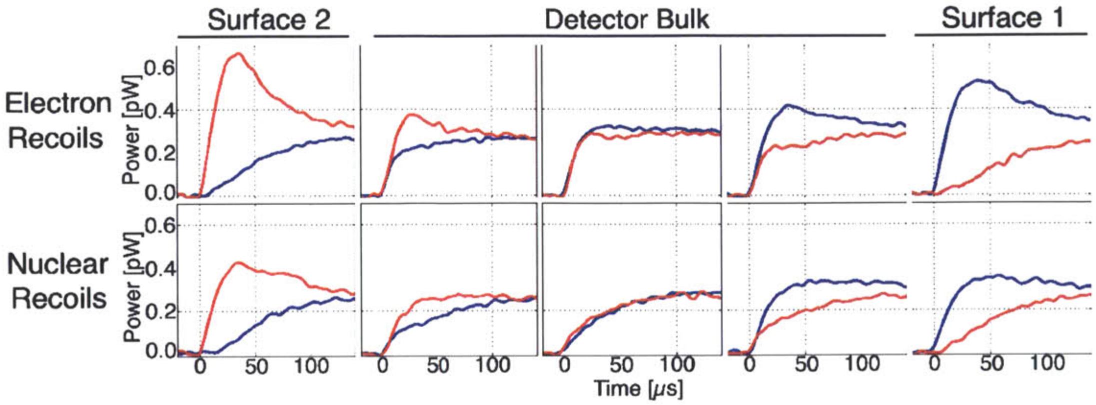

Figure 12-22: Example phonon pulses from calibration runs with a SuperCDMS Soudan detector. The four side 1 QET channels are summed and shown in blue (dark grey); the side 2 channels are summed and shown in red (light grey). All eight events shown are of similar total phonon energy (~50 keV), and are located at different positions along the central axis of the cylindrical crystal (as determined by relative weighting among the eight phonon channels). Surface 1 and surface 2 events were selected using an additional a requirement that the charge signals from the two sides be strongly asymmetric.

Plugging in the appropriate charge carrier creation efficiencies ($\epsilon$), voltage differences, germanium gap energy of 0.75 eV, and what essentially amounts to a guess for the relaxation phonon creation efficiency ($\alpha$) of 0.1, we arrive at three different [ballistic:diffusive] phonon ratios for three different event types: bulk electron recoils [1:1.8] ; surface electron recoils [1:0.9] ; and bulk nuclear recoils [1:0.5]. These differences in phonon population magnitudes, convolved with a reweighting should produce observable effects on the shape and timing of SuperCDMS phonon pulses.

### 12.7.2 Time-Domain Pulse Fits

The understanding of phonon populations given in the previous section predicts certain shape characteristics as a function of both proximity to the absorbing surface and quantity of charge carriers (and resulting Neganov-Luke production). Looking at recent calibration data (Figure 1), we do see such behaviors, confirming our understanding.

To make use of these phonon pulse shape variations as event type discriminators(distinguishing electron recoil and near-surface recoils from nuclear recoils), we first reduce the raw traces to a small number of simple quantities, and then combine these quantities in a way that accommodates the large position-dependent variation. First, each of the raw traces from the detector's eight QET channels were fit using the following functional form:

$$P(t) = A_{fast}\left(1 - e^{\frac{-(t-T_{fast})}{R_{fast}}}\right)e^{\frac{-(t-T_{fast})}{F_{fast}}} + A_{slow}\left(1 - e^{\frac{-(t-T_{slow})}{R_{slow}}}\right)e^{\frac{-(t-T_{slow})}{F_{slow}}}$$

where the total pulse is treated as the sum of a "fast" pulse and a "slow" pulse, with rising and falling time constants $R_{fast}$, $R_{slow}$, $F_{fast}$, and $F_{slow}$, and start time offsets $T_{fast}$ and $T_{slow}$. The fast pulse can be thought of as the contribution of highly-diffusive quickly-absorbing phonons, while the slow signal can be thought of as the contribution of more ballistic slowly-absorbing phonons, but the quasi-diffusive nature means that in reality there is not a simple two-category distinction. The falling time constant $F_{slow}$ of the slow pulse was observed to be nearly identical for every event, and was set to 755 $\mu$s, the observed rate of absorption of the late-time uniform bath of low energy phonons.[83] Each event's resulting 7 fit parameters for each of 8 phonon channels were further reduced by summing the fits for each side and then finding key points (10%, 20%, 30%, etc.) along the rising and falling edges of these side-summed fits.

The partition of energy between the eight channels also contains significant discrimination information (in addition to position information). The amplitude of each channel's pulse was obtained using an optimal filter, in which a template pulse of fixed shape was scaled to best match the amplitude of the (variable shape) phonon pulse. These measured amplitudes, then, are shape-dependent. Distributions of some example partition and pulse shape quantities are shown in Figure 2.

### 12.7.3 Discrimination Ability Using Pulse Shape

As a very first look at discrimination based on phonon pulse shape characteristics, we here show a simple example of bulk electron recoil (ER) vs bulk nuclear recoil (NR) discrimination using only one pulse shape characteristic: the 40%-to-70% risetimes.

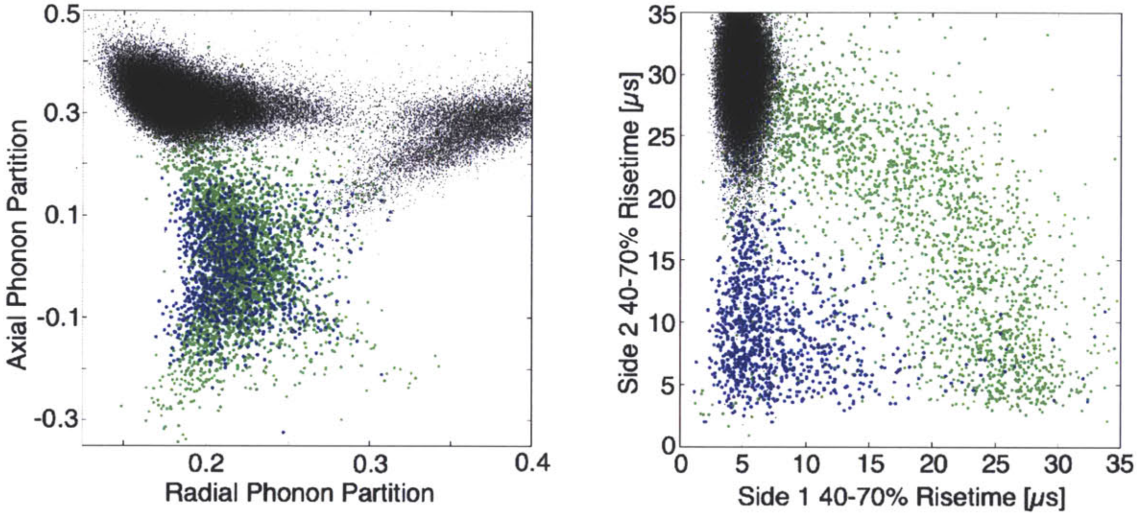

Figure 12-23: Example phonon discrimination quantities. On the left, an example of partitioning between channels is shown. The vertical axis is an axial partition quantity (side1 - side2)/(side1 + side2), and the x axis is a similarly defined radial partition quantity. On the right, the time difference (in $\mu$s) between the 40% and 70% points on the rising edge of the side 1 summed pulse (x axis) and side 2 summed pulse (y axis) are shown. In both plots, calibration events are colored using a charge-signal-based categorization as either bulk electron recoils (blue or dark grey), bulk nuclear recoils (green or light grey), or side 1 electron recoils (black). The total phonon energy for these events are between 7 and 20 keV recoil energy, where recoil energy has been scaled from total phonon energy using a nuclear recoil assumption for all events.

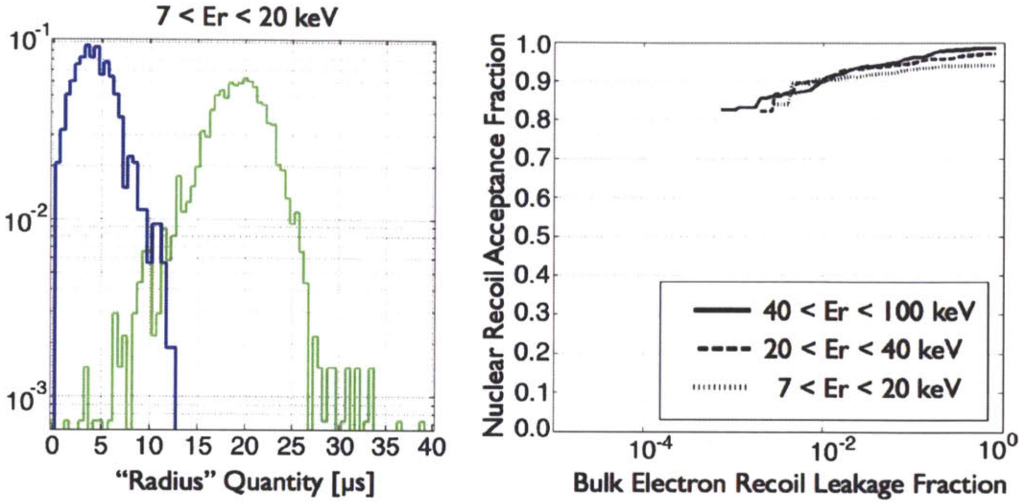

Figure 12-24: A simple look at pulse shape discrimination between bulk nuclear recoils and bulk electron recoils, using 40%-to-70% risetimes. Nuclear recoils are green (light grey); electron recoils are blue (dark grey). For discussion, see text.

This simple discrimination example serves as a lower bound on the abilities one might expect when using a combined analysis of all phonon pulse quantities.

We first construct ER and NR populations. Identical charge-based fiducial volume cuts are enforced on both populations, and measured [charge:phonon] energy ratios are used to categorize events as either ER or NR. After these two populations have been defined, a new pulse shape discrimination quantity is constructed using the 40%-to-70% risetimes (for side 1 and side 2) for each event: "radius" = ([40%-to-70% side $1]^2$ + [40%-to-70% side $2]^2)^{1/2}$. One can see in Figure 2 that this combined quantity is largely position-independent.

Figure 3 shows a histogram of this timing quantity, plotting only the lowest energies inspected in this analysis ($7 < \mathrm{E}_{recoil} < 20$ keV, defined by assuming that all events are nuclear recoils and scaling the total phonon energy accordingly). Although the ER and NR distributions overlap somewhat, there are no slow ER outliers to the statistics available in the calibration dataset. The right panel of Figure 3 shows the NR acceptance fraction vs ER leakage fraction as one varies a 1D cut threshold. Discrimination better than $1{:}10^3$ is seen, and it is further seen that this discrimination shows no degradation with energy down to at least 7 keV $\mathrm{E}_{recoil}$.

Although significant pulse shape discrimination ability has been demonstrated, phonon-only discrimination is only beginning to show its promise. The simple analysis performed here can be improved in many ways, and looking further on the horizon, this rough analysis using a detector optimized for charge-electrode-based discrimination suggests that a detector optimized specifically for phonon-only discrimination could take advantage of the extreme sensitivity of transition edge sensors to extend event-by-event discrimination well below the typical CDMS low energy threshold of ~10 keV (set by charge readout noise). Such phonon-only discrimination capabilities should be possible if detectors are specifically optimized for this goal, through the reduction of TES internal thermal fluctuation noise[13], and also through an increase in pulse shape differences themselves by increasing the total phonon-absorbing Al area of the QETs.

This pulse shape analysis was very preliminary in many ways, but shows tremendous promise. There is an enormous amount of extremely useful information in the shape of iZIP phonon pulses, newly available with the spacial separation of primary and Luke phonon production. There is a lot of low-hanging fruit here, waiting for another graduate student.

## 12.8 Surface Event Yield

Even if we threw away the iZIP's ability to define a perfect fiducial volume, the yield of surface betas alone is high enough for much better rejection than all handles combined could do for CDMS II. This has been something of a surprise, and there are tantalizing hints that we could boost the surface event yield even higher if we ever have need to.

Let's begin by looking at Figure 12-25, which shows the results of $^{210}$Pb calibration (with no fiducial volume cuts applied) in two very similar detectors: the an Edelweiss II detector, and T3Z1 at Soudan. We have not mentioned Edelweiss yet at all, but they are a European collaboration progressing very much in parallel to CDMS. Their detectors are Ge as well, and similarly have a surface event rejection

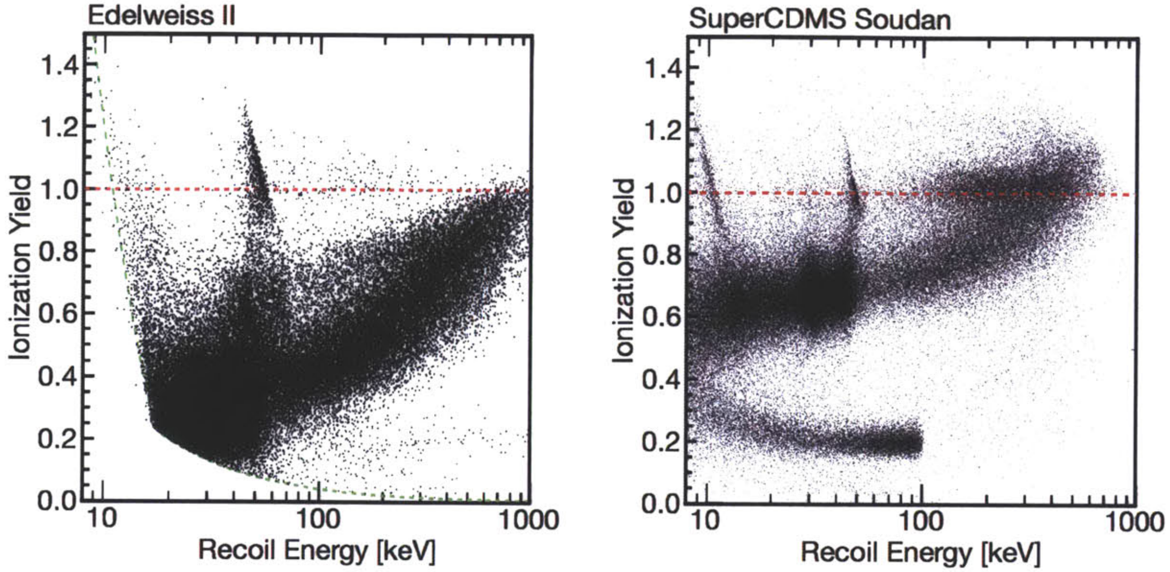

Figure 12-25: Surface event yield for two different detectors, an Edelweiss II detector on the left, and a SuperCDMS Soudan detector (T3Z1) on the right. Both detectors have been exposed to a $^{210}$Pb source. Edelweiss II plot from [95]. No fiducial volume cuts are placed on either set of data. In CDMS, some TES saturation occurs in the high energy range (starting around 200 keV), leading to somewhat underestimated recoil energies and somewhat overestimated ionization yield. Events below this do not seem to suffer from significant saturation effects.

through interleaved electrodes. The main difference between Edelweiss and CDMS is that instead of using QETs, Edelweiss makes a *thermal* yield measurement, simply using a thermometer (a very sensitive NTD thermometer) to measure the tiny temperature increase of the large Ge substrate. But this difference just means the two experiments collect differen amounts of phonon *information*; they both collect the same amount of phonon *energy*, and thus we would naively not expect the two experiments to differ substantially in their ionization yield measurements.

Instead, we see dramatic differences. The (penetrating) gamma line at 45.6 keV is of yield=1 in both detectors, but the the suppression of surface beta yield in CDMS is roughly *half* the level we see in Edelweiss, and in Edelweiss, the $^{206}$Pb recoils exhibit *complete* charge trapping. The difference between the two experiments (it is assumed) is largely due to differences in the surface. In the space between the electrode lines and the QET ribbons in CDMS, there is only a bare polished Ge surface. Between the electrode lines of Edelweiss II, there is a thin layer of amorphous Si. It seems we have strong evidence, in comparing the two experiments, that a bare polished crystaline surface has the least amount of charge trapping associated with it (or, equivalently, a thin amorphous layer is quite effective at encouraging either recombination or trapping).

Now that we have made many iZIPs, we can compare various detectors of our own. Several iZIPs were tested at Berkeley using a $^{109}$Cd beta source (exposing only one side, as at Soudan). It was observed that all three detectors had nearly identical surface yield when biased such that holes were collected on the electrodes (yield$\approx$0.85), but that the detectors showed varying beta yield when biased such that electrons were collected on the electrodes. One detector exhibited an electron-biased surface yield that was nearly unchanged (at y$\sim$0.8), and the other two detectors exhibited an electron-biased surface yield that was significantly degraded (at y$\sim$0.7). Given these small statistics (three detectors), we cannot know for certain, but it is interesting to note that the one detector with impressively high electron-biased surface yield was also the one detector with a P-type substrate (the others were N-type). It would not be at all surprising for such a difference in doping to significantly affect the polarization of the Ge immediately surrounding the electrodes, or to significantly affect any barrier potential present near the semiconductor-metal interface. We await tests of additional P-type crystals to be sure.

An additional (and equally unexpected) determiner of beta yield perhaps should have been expected. It was observed that the angle of the electrodes and ribbons (equivalently, the orientation of the surface field) with respect to the crystal lattice is a strong determiner of surface event yield, as shown in Figure 12-27 and Figure 12-28. Depending on the angle, hole-biased surface yield varies from $\sim$0.75 to $\sim$0.8, and electron-biased surface yield varies even more dramatically, from $\sim$0.65 to $\sim$0.75. I say this effect perhaps should have been expected, because as soon as the effect was seen, it became clear what was happening (an illustration is shown in Figure 12-29). The oblique propagation of electrons is either in line or not in line with the near-surface field direction. Yield is maximized when the L-valley propagation and the

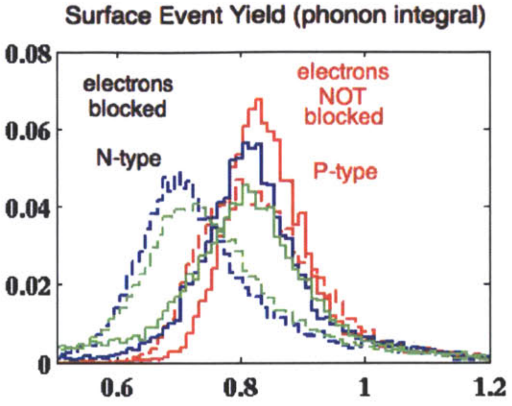

Figure 12-26: Surface event yield for three different detectors, using $^{109}$Cd calibration data taken at Berkeley. Dashed lines indicate the source side +2V bias state (*i.e.*, electron collection) and solid lines indicate the source side -2V bias state (*i.e.*, hole collection). Red is G3D, blue is G41, and green is G48. G41 and G48 show significantly lower yield when biased in the electron collection state, while G3D (the only P-type crystal) shows only slight suppression.

electric field are maximally aligned. In future designs, we could imagine boosting surface yield by thinking carefully about the orientation of the surface field with respect to the crystal lattice.

Lastly, it should also be mentioned that there exists a population of surface events for which the phonon measurement is significantly suppressed (in all channels), leading to a yield greater than 1 (as high as 1.3). This can be seen, for example, in the side 1 surface event population in Figure 12-12. We haven't seen such phonon suppression previously in CDMS II-style devices, perhaps because we simply haven't previously looked at surface events with such high statistics, or perhaps in previous designs our surface yield was so low that the combination of both effects (suppressed phonon measurement and suppressed charge measurement) meant the yield was not high. We don't really need to understand effects that push backgrounds *away* from our signal region, but it would always be nice to understand our detectors. There are two (not mutually-exclusive) explanations for this surface event phonon-undermeasurement:

- These events are very shallow and directly under an electrode line, leading to significant phonon absorption in the tiny (passive) Al electrode rail (while the

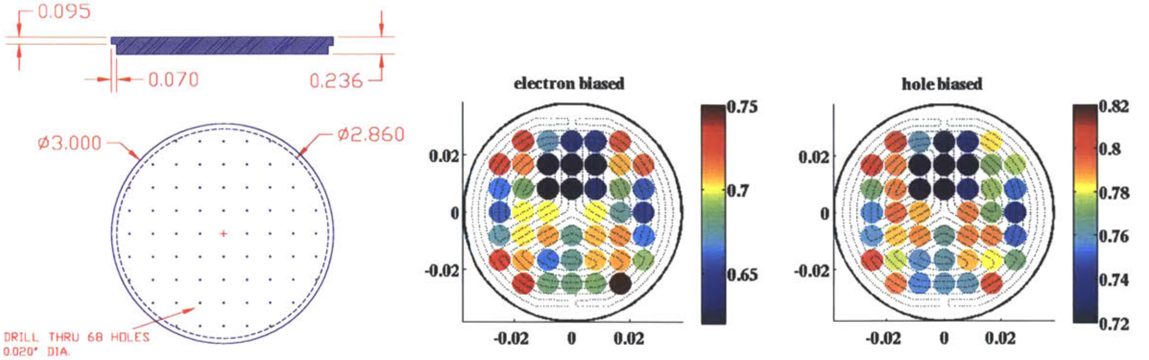

Figure 12-27: On the left, we show the collimator used in testing at Berkeley to produce a grid of $^{109}$Cd exposure points. On the right, we show the average yield of surface events (events failing the charge-symmetry cut), by position, for the two different bias states (note the differing coloring scales).

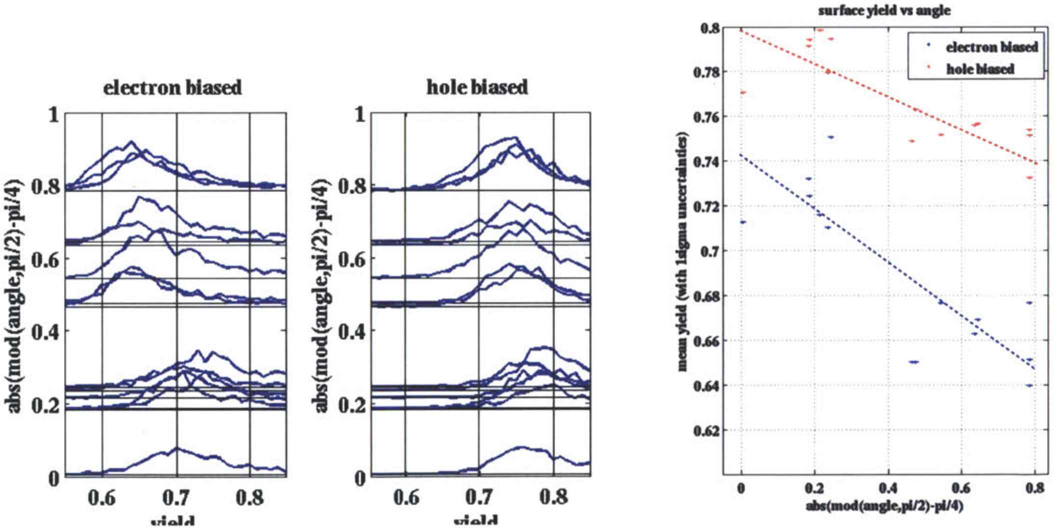

Figure 12-28: A strong correlation is seen between the relative angle between the crystal lattice and the yield of charge-asymmetric events. On the left, histograms of yield for surface events are shown, for both bias states, by angle, and the mean yield values are shown for both biases on the right.

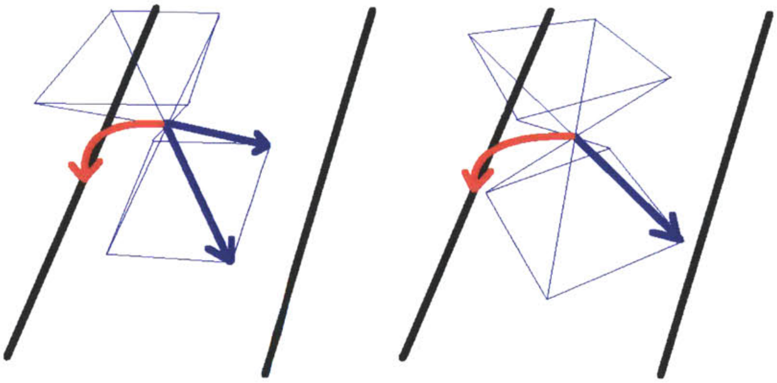

Figure 12-29: A cartoon illustrating the relative lattice-electrode orientation at minimum yield (left) and maximum yield (right). Red lines indicated the path of holes (simply following the electric field lines), and blue lines indicate the path of electrons (assuming the field is weak enough such that they propagate obliquely).

phonons are still in their extremely short mean-free-path state, and therefor highly likely to be absorbed).

- These events are very shallow and directly under a QET, leading to a significant phonon absorption the single QET and an extremely saturated single TES.

Both possibilities represent a way in which some fraction of phonon energy could be trapped in passive Al (in the second case, the Al is passive only because it is connected to a saturated TES that contributes very little to the signal).

## 12.9 Summary

The iZIP represents a significant leap in CDMS detector technology. The CDMS II demonstrated a leakage-free rejection of bulk electron-recoils, and now this freedom from leakage events has been extended to *all* electron recoils, bulk or surface. CDMS II depended on subtle pulse-shape differences to produce final results; the iZIP analysis is comparatively *easy*. The entire analysis can be performed using only *amplitude* quantities, and the cuts can be placed in natural gaps in the event distributions, so that neither the background leakage nor the nuclear recoil exposure are delicately dependent on the placement of these cuts. Simply put, the interleaved strategy has succeeded above and beyond our most optimistic expectations.

# Part V

# Outlook

# Chapter 13

# Detectors for SuperCDMS SNOLAB

SuperCDMS Soudan can be viewed as simply a proof-of-principle of the extreme background rejection abilities of the iZIP design. Full implementation of this strategy in a large target mass is the goal of the next version of CDMS, SuperCDMS SNOLAB, the planning for which is well under way.

The iZIP4/5 represents a great leap forward, but there are several comparatively small ways in which the detector could still be improved for SNOLAB. First and foremost, bigger is better when it comes to direct detection. The mass goal for SuperCDMS SNOLAB is 200 kg, which, if composed of iZIP4/5s, would require the fabrication and operation of more than 300 detectors. We are no longer afraid of the large surface area this represents, but the complexity of the experiment grows with the number of detectors (and channels). The SNOLAB detectors, then will be larger, with a radius of 100 mm, a thickness of 33.3 mm, and a resulting Ge mass of 1.38 kg (compared with the iZIP4/5: 76 mm radius, 25 mm thickness, and 0.62 kg mass). This more than doubling of individual detector mass will mean

- Fewer detectors need to be fabricated and tested (~144 instead of ~320)
- Fewer channels will need to be connected to room temperature (1728 phonon, 576 charge, compared with 2576 phonon, 1288 charge using the iZIP4/5)

- A higher fraction of the target mass will be in the fiducial volume ($\sim 73\%$ instead of $\sim 63\%$)

The SNOLAB iZIP is more than a simple scaling up of the iZIP4/5, however. With the freedom of essentially starting over from the ground up, the readout electronics have been redesigned. Noise on the charge measurement should be significantly reduced by replacing the JFET amplifier with newly available high-electron-mobility transistors (HEMTs), which exhibit significantly reduced $1/f$ noise.

On the phonon readout, the SQUID array for SNOLAB has a significantly reduced input inductance, which allows the efficiency of the phonon measurement to be dramatically improved. Remember that in order to cover a detector surface, we want to maximize the number of QETs (connected in parallel) in each channel, but that as we thus lower the channel resistance, we eventually run up against the constraint that the $L_{squid}/R_{channel}$ time constant cannot grow too large. The new low-inductance SQUID arrays allow us to lower the channel resistance, and increase the number of QETs per channel, thus significantly increasing the signal amplitude. Remember, too, that we we made the TESs extremely long (perhaps *too* long, longer then their phase separation lengths) in an effort to increase the resistance of each TES, thereby being able to pack more TESs (more active surface coverage) per channel. The low-inductance SQUID arrays, then allow us to *both* pack more QETs in per channel *and* reduce the TES length to phase-homogeneous lengths.

There is even a third advantage to the lower resistance of the new SQUIDs. Remember that in previous iZIPs, the QETs are spaced almost a millimeter apart along the ribbon, much longer than the quasiparticle trapping length scale. Because the rails that connect the QETs together are basically phonon sinks (where the energy can be trapped but cannot reach a TES), there is a strong motivation to reduce the size of these rails (so that only a small fraction of the phonon energy is lost to them). This produces a tension between optimizing the detector for lowest-possible phonon losses (narrow rails), and optimizing the detector for lowest-possible rate of phonon 'opens' (wide, robust rails). We no longer must choose between these two options at SNOLAB. As seen in Figure 13-1, the optimal QET ribbon geometry undergoes a 'phase change' when the spacing of TESs along the ribbon is lower than the $1.5\ell_{trapAl}$ rule-of-thumb fin length. At SNOLAB, no portion of the phonon ribbon will be passive; the rails have become a shared, highly efficient fin, and there are no longer any narrow Al features. The ease of producing consistent Al continuity is extremely important for SNOLAB, with the much larger number of detectors.

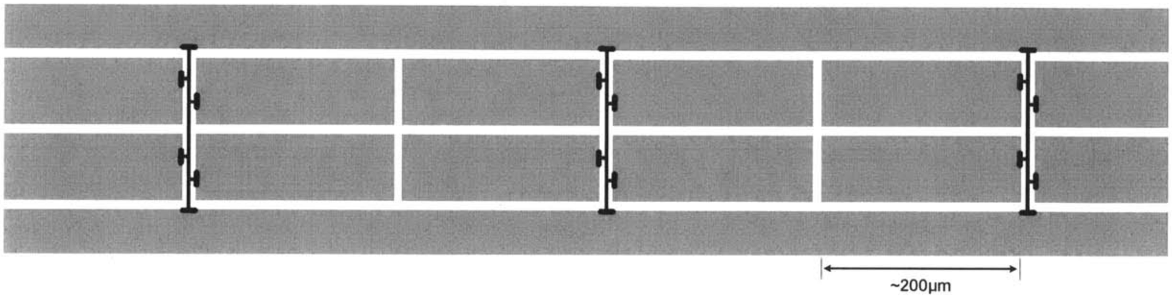

Figure 13-1: A preliminary version of the SNOLab detector's QET ribbon, showing the newly 'close-packed' design.

The last major change for SNOLAB is the phonon channel partition. To retain a excellent position sensitivity over the now much larger volume, the baseline design for SNOLAB has 6 phonon channels per side, arranged as an outer ring, a central bull's eye, and the remaining gap divided into fourths (as can be seen in Figure 13-2). The two sides are offset by 45 degrees, such that the boundaries of the four middle channels lie in the middle of the opposite side channels (similar to the 60 degree side-offset in the iZIP4/5).

The operating resistance allowed by the new SQUID array is 50m$\Omega$, a little more than three times lower than at Soudan. How exactly to partition this factor of three between more TESs (*i.e.*, shorter fins) and lower-resistance TESs (*i.e.*, lower TES noise) is still being worked out, but one possible design could be to set the TES length at 125 $\mu$m and the electrode-ribbon spacing at 1.6 mm, which would result in a ~1,000 QETs per channle, a QET spacing along the ribbon of ~300 $\mu$m, and a fin length of 150$\mu$m (less than half that of the iZIP4/5).

Although the efficiency of the fins should be extremely high, the total active Al surface area will be no greater than in the past, due to the same electrode capacitance

Figure 13-2: Side 1 of the iZIP6: a preliminary version of the SNOLab detector (top) and a photo of the first iZIP6 fabricated (bottom). The spacing of the TESs along the QET ribbon structures is much larger than it will be in the final SNOLAB detector, increasing the detector resistance to match the existing readout electronics.

constraint as in previous designs. In fact, with the larger surface area, the electrode capacitance is similarly (unavoidably) increased. Figure 13-3 shows the capacitance of the combined charge channels of one side of a 100 mm detector, as a function of electrode-to-ribbon spacing and the electrode width. A spacing of 1.6mm keeps the phonon relatively homogenous across the surface, while limiting the electrode capacitance to less than 200 pF (only weakly dependent on electrode width), nearly double the iZIP4/5 value.

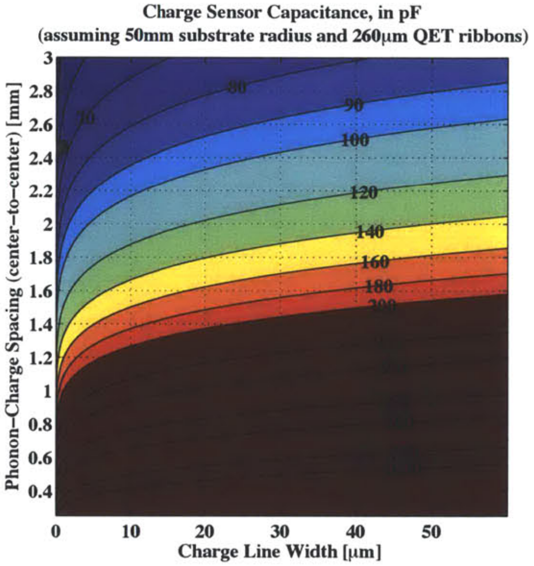

Figure 13-3: The capacitance of one side of a 100 mm detector, as a function of electrode width and electrode-ribbon spacing, assuming a ribbon width of 260 $\mu$m.

The detector design shown in Figure 13-2 is called the iZIP6, and is not the SNOLAB design, but instead the a mockup of the SNOLAB design that is of phonon channel resistance suitable to the existing readout electronics (meant to test the detector concept while the SNOLAB electronics are still being perfected). Presumably, the SNOLAB design (iZIP7) will look much the same, only spacing the QETs more densely along the ribbons.

| | **CDMS II** | **SuperCDMS Soudan** | **SuperCDMS SNOLAB** |
|---|---|---|---|
| Mass per Detector [kg] | 0.25 [Ge] | 0.62 | 1.38 |
| Number of Detectors | 19 [Ge] | 15 | ~144 |
| Total Ge Mass [kg] | 4.75 | 8.98 | ~200 |
| Fiducial Mass Fraction | 0.37 | 0.63 | ~0.73 |
| Fiducial Mass [kg] | 1.75 | 5.66 | ~150 |
| Phonon Channels per Det | 4 | 8 | 12 |
| TES Length [$\mu$m] | 250 | 220 | ~150 |
| Al Fin Length [$\mu$m] | 380 | 327 | ~150 |
| QET Efficiency | ~ 3% | ~ 12% | ~ 20%? |
| Phonon Energy Resolution [eV] | 180 | 200 | <100? |
| Trigger Threshold [eV] | ~2000 | ~3000 | ~500? |
| Capacitance (One Side) [pF] | 130 | 100 | ~180 |
| Charge Energy Resolution [eV] | 300 | 450 | < 300 |
| **Bulk ER Leakage** | | | |
| Yield only | $< 1 \times 10^{-6}$ | $< 1 \times 10^{-6}$ | $< 1 \times 10^{-6}$ |
| Phonon only | $\sim 1 \times 10^{-1}$ | $< 2 \times 10^{-4}$ | $\sim 1 \times 10^{-5}$? |
| **Surface ER Leakage** | | | |
| Yield only | $\sim 2 \times 10^{-1}$ | $\sim 1 \times 10^{-3}$ | $\sim 1 \times 10^{-3}$ |
| Charge only | $NA$ | $< 2 \times 10^{-5}$ | $< 2 \times 10^{-5}$ |
| Phonon only | $\sim 2.5 \times 10^{-2}$ | $< 1 \times 10^{-4}$ | $< 1 \times 10^{-4}$ |
| Total | $\sim 5 \times 10^{-3}$ | $< 2 \times 10^{-5}$ | $< 2 \times 10^{-5}$ |
| Surface rejection threshold [keV] | 10 | ~8 | ~2? |
| **Experimental Reach** [zb] ($\sigma_N^{SI}$ at $m_\chi = 60$ GeV) | 40 (~2 years) | ~5 (~2 years) | ~0.08 (~4 years) |

Table 13.1: Overview comparison of detector characteristics for CDMS II, SuperCDMS Soudan, and (plans for) SuperCDMS SNOLAB. Some of the SuperCDMS Soudan numbers are still tentative, and *most* of the SuperCDMS SNOLAB numbers are tentative. Some SuperCDMS Soudan leakage rates are taken from testing at Berkeley, where higher-statistics surface event samples were collected (with the downside that muon-induced neutron events occur). Note that many of the expected SNOLAB improvements are the result not only of detector improvements, but significant improvements to the readout electronics.

# Chapter 14

# The Future of Direct Detection

## 14.1 Towards Larger Targets

Sensitivity to generic WIMP models, such as those predicted by supersymmetric theories (10 GeV$< m_\chi <$ 1 TeV), is largely a function of target mass. The larger the target mass, the greater the sensitivity to such 'high mass' models (assuming a near-zero background rate). Figure 14-1 shows the dramatic improvements in this scaling up of background-free targets, as well as projections into the future for several experiments currently being developed.

Because the CDMS (and the similar Edelweiss) detectors are modular in nature, a thorough testing of background rejection (such as is being performed at Soudan right now) can be performed on a single detector, and this same rejection ability can be safely assumed for the set of many such detectors. In other words, if we were to measure a bulk ER rejection of $1 \times 10^{-7}$ and a surface ER rejection of $< 1 \times 10^{-5}$ in T3Z1 at Soudan (as we expect to achieve in the coming months), then it is we can feel very confident building 200 kg experiment, knowing the rejection abilities (and remaining fiducial volume) of the 200 kg in advance. The SuperCDMS projections in Figure 14-1, then, are very safe bets. It will, however, take a substantial period of time ($\sim$2 years) to fabricate the $\sim$144 detectors necessary for the payload.

The biggest unknown for scaling up the CDMS mass is not whether the plans are technologically possible, but whether they are *practically*. CDMS has a stated

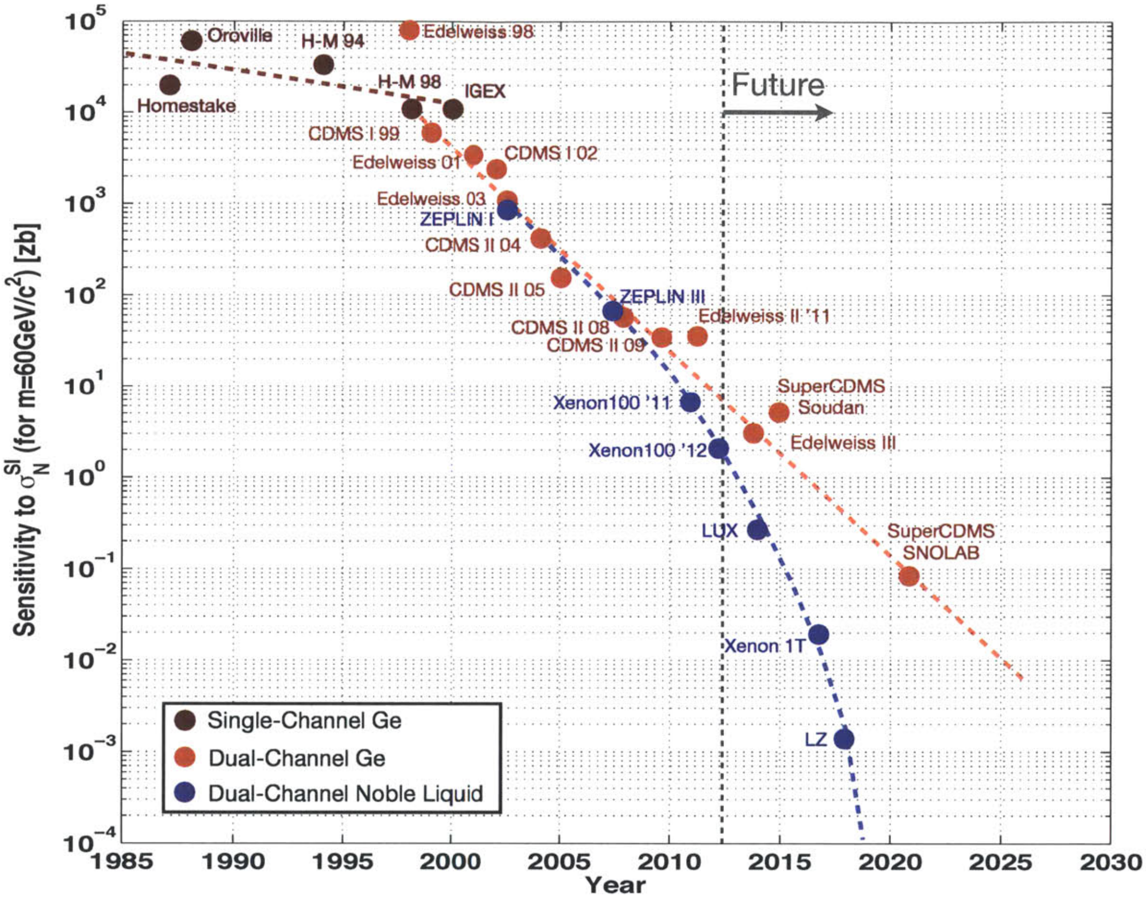

Figure 14-1: A general overview of the sensitivity of direct detection experiments over time to the generic WIMP mass of ~60 GeV, grouped into general categories of 'single-channel Ge' (germanium targets with charge measurement only), 'dual-channel Ge' (germanium targets like CDMS, with ER discrimination), and 'Dual-channel noble liquid' (like Xenon and LUX, experiments that have weak ER discrimination but large target volumes). General historical trends for the three categories are indicated, along with claims made by several collaborations as to their projected sensitivities. Notice that the noble liquid collaborations anticipate an enormous gain in sensitivity from extreme radio-purity Xe and the exponential benefits of shielding.

goal for SuperCDMS SNOLAB of fabricating six working detectors per month, a rate which has recently been demonstrated (see Figure 11-18). Additionally, a new fabrication pipeline has been developed by collaborators at Texas A&M, which has recently demonstrated an ability to fabricated detectors not only quickly, but *better* (in terms of both $T_C$ homogeneity across the surface and between detectors, and in terms of Al film quality). The detector fabrication rate and quality necessary for a larger target mass has been effectively proven.

Noble liquid targets posses the advantage of a single massive target volume, which makes the scaling up to larger masses exponentially easier. Noble liquids have the additional advantage that the surface area (where the majority of backgrounds lie) scales more slowly than the target mass. Particularly in liquid xenon, electron recoils from the surrounding vessel (the highest sources are typically found near the photomultiplier tubes) penetrate only cm length scales into the target, meaning that the central volume is well shielded from the surface contaminants. On the other hand, the electron-recoil rejection abilities of noble liquid experiments is many orders of magnitude worse than in CDMS. In discrimination space, the centroid of the nuclear recoil population lies on the tail of the electron recoil distribution. This means eliminating the tiny rate of electron recoil backgrounds in the central volume is of extreme importance. In the most recent exposure of Xenon100, for example, had an impressively low background rate of $5.3 \pm 0.6 \times 10^{-3}$ [keV kg day]$^{-1}$, but even this low rate necessitated severe discrimination cuts to produce a background-free nuclear recoil sample. The remaining fiducial volume was only 34 kg, and the nuclear recoil efficiency of this volume (after discrimination cuts) was only $\sim 50\%$ at 10 keV (better below this energy, worse above) [112].

With poor background discrimination, then, the biggest unknown for noble liquids heading forward is not whether a high-mass target volume can be successfully built and instrumented, but the *technological* question of whether the xenon purity necessary for such scales is achievable. For the a LZ experiment, with a 7 ton (fiducial) xenon target mass, an intrinsic background of $\sim 1 \times 10^{-7}$ [keV kg day]$^{-1}$ is assumed in the projection shown in Figure 14-1, dominated by the $\beta$-decay of xenon itself (the isotope $^{137}$Xe) [82]. For these large volumes, high purity is also essential for efficient drifting of the electrons through the length of the detector.

## 14.2 Towards Lower Thresholds

The main thrust of direct detection experiment R&D is towards larger and larger target volumes, while keeping the energy threshold fixed at $\sim 5 - 30$ keV. But with no hints of supersymmetry yet observed by the LHC, one of the primary motivations of this 'high mass' ($m_\chi$ >10 GeV) hypothesis is significantly weaker, and there is growing interest in exploring the 'low mass' ($m_\chi$ <10 GeV) parameter space. For low masses, large swaths of parameter space can be explored for the first time by relatively low-mass targets, if the energy thresholds could be significantly reduced.

CDMS-style technology is *uniquely* suited to the low energy regime, thanks both to the extreme sensitivity of the TES sensors themselves, and to the low energy-per-quantum of both the charge (semiconductor instead of ionization) and phonon channels. It is actually somewhat embarrassing, then, that CDMS is not dominating the field at low energies. We now discuss three ways in which CDMS-style technology could be optimized for lower-energy reach.

### 14.2.1 TES Sensitivity

As previously mentioned, the energy resolution of a TES scales as the operating temperature to the third power. Phonon energy resolution in the iZIP4/5 is ~200eV (at $T_C \approx$ 90mK). One side of an iZIP was fabricated with a $T_C$ of ~65 mK (accidentally), and surface testing showed a one-sided energy resolution of ~40eV, already quite impressive. Reducing the $T_C$ further (purposefully, this time) can produce extreme rewards. Even a moderate reduction in $T_C$ to 30 mK (still well above the temperatures attainable by commercially available dilution refrigeration) should result in energy resolutions <10 eV. This $T^3$ scaling is by far the easiest way to improve TES sensitivity; we have not yet considered here the incremental but important improvements easily attainable by shortening the TESs below their phase separation length or

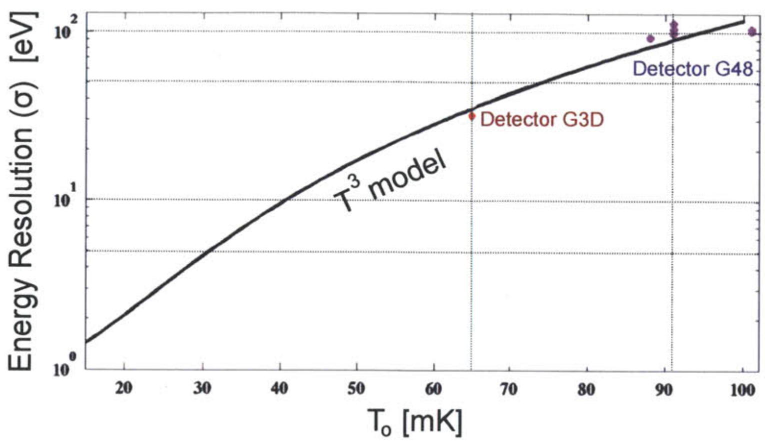

Figure 14-2: TES energy resolution as a function of $T_C$. From the thesis of M. Pyle.

increasing the absorption rate so that the pulse length better matches the TES time constants. Dramatically reduced thresholds through moderate TES improvements seem easily within reach.

### 14.2.2 Phonon Absorption Rate and QET Efficiency

Because TES noise can be so dramatically reduced, it is tempting to imagine a detector designed as a phonon-only device, free of the irreducible noise of the charge measurement (from unavoidable capacitive couplings). We have shown how, in the iZIP4/5, both yield and fiducial volume can already be defined with some success using phonon measurements alone. If we are to design a detector specifically designed for this strategy, QET efficiency becomes even more important.

First, there is the Al surface coverage. In a phonon-only device, the surface field region would still be useful for defining a fiducial volume (as in Figure 12-15 or Figure 12-16), but tight constraint on capacitance would be released, allowing for larger QETs and a higher Al surface coverage. A faster transfer of the phonon energy into the TESs would mean that much more information could be captured. In the iZIP4/5 (and presumably the iZIP6/7), only about 1/5 of the phonon energy is captured before the phonons form a homogeneous bath in the crystal. If our goal is not just phonon energy measurement, but to measure all the information contained in phonon timing (event position, event yield, event multiplicity), then faster absorption

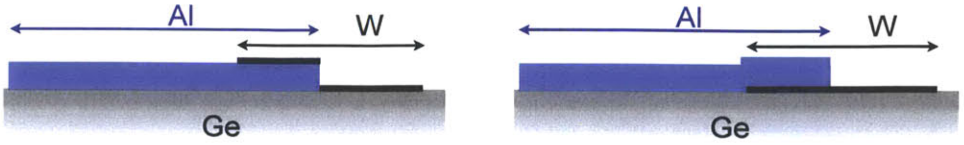

Figure 14-3: A cartoon showing (left) the manner in which W step coverage limits the Al fin thickness (and the Al fin's resulting phonon absorption efficiency and quasiparticle trapping length $\ell_{trap}$), and (right) the 'W-first' strategy, which allows Al of any thickness.

through larger Al surface coverage is vital.

Second, the current QET technology is still far short of the theoretically possible efficiency of $\sim 50\%$ (set by the phonon-quasiparticle cascade efficiency). One of the main limitations to our current technology is the fabrication necessity of placing the W layer on top of the Al fin layer. This limits the Al thickness (and hence the phonon absorption rate, the efficiency of the phonon-quasiparticle cascade, and the quasiparticle mean free path), because the thin W layer must have good step coverage between the Al-W overlap region and the W monolayer region. If the Al is too thick (thicker than the current ~300nm), then the W is discontinuous between the overlap and the monolayer, and no quasiparticles can diffuse from the overlap to the TES. The solution to this is simple in concept: eliminate the W step coverage altogether by depositing (and patterning) the W first and the Al second, as in Figure 14.2.2. Of course, there are very good reasons we don't do this already, principally the fear of accidentally leaving some residual contaminants on top of the TES after fabrication (the current fabrication path lays down the TES tungsten last). With a new fabrication pathway at Texas A&M, however, much in the fabrication recipe deserves a second look.

### 14.2.3 Luke Amplification

The last technique for extending the low-energy reach of CDMS technology unfortunately means giving up any ER-NR rejection ability (at least, on an event-by-event basis), and essentially only measures the charges produced by an event. The extreme sensitivity of the TESs can be paired with our ability to scale Luke phonon production (through scaling the drift field). Typically, we scale the drift field amplitude so that the Luke phonons for an electron recoil are equal in energy to the 'prompt' or 'primary' phonons produced at the recoil site itself. If we boost the drift field from where it currently stands (~1 V/cm) to strength orders of magnitude higher (perhaps ~100 V/cm), we increase (linearly with total voltage difference) the production of Luke phonons. Although it is essentially a one-channel measurement (the prompt phonons are swamped by the Luke phonons), a tiny energy deposited in the form of e-h pairs can be amplified into a much larger, much easier to sense, *phonon* signal.

This strategy could be implemented in specially-optimized detectors, or it could be implemented in currently-existing detectors. Tests of this strategy are currently taking place at Soudan using the installed iZIP4/5 detectors, where stable running has been achieved up to 30 V/cm (for such tests, the QET ribbons and electrodes are at the same voltage, meaning there is only a bulk field strength, no surface region). At this bias, noise in the phonon measurement roughly doubles, but the signal has been amplified by a factor of ~24 compared to the prompt signal (as compared with the CDMS II factor of 2). For clarity, we can write the phonon energy of an event as

$$E_{phonons} = E_{recoil} \times \left[1 + \frac{V_{bias}}{\epsilon}\right] = E_{recoil} \times \left[1 + Y\frac{V_{bias}}{\epsilon_{ER}}\right] \tag{14.1}$$

where $\epsilon$ is the creation energy of an electron-hole pair, $\epsilon_{ER}$ is that value for an electron recoil in Ge (3 eV), and $Y$ is the ionization yield for the event (1 for electron recoils, energy dependent for nuclear recoils).

With the observed noise and amplification at Soudan, we should be able to run existing detectors with a single-channel threshold of an impressively low ~85 $\text{keV}_{\text{ee}}$. Notice that we are using the language of an electron-equivalent energy scale here, where we have assumed that the event is an electron recoil. The equivalent nuclear recoil threshold is significantly higher (applying a yield of 0.15, this threshold rises to ~460 $\text{keV}_{\text{nr}}$. Still quite low, though!

We have lost event-by-event discrimination power, but notice two things that help

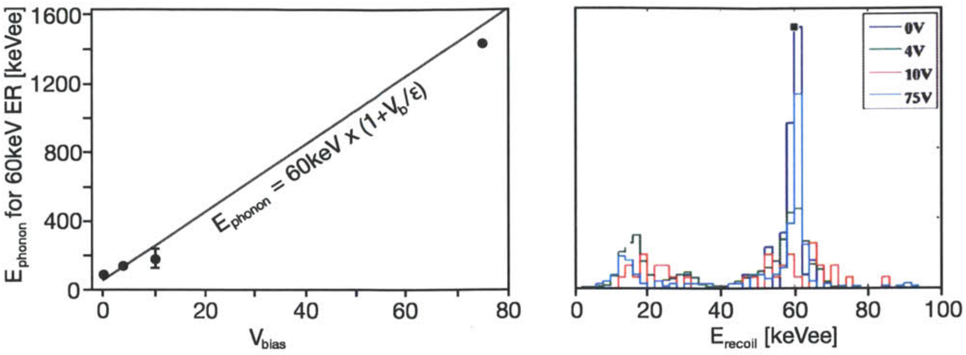

Figure 14-4: Recent measurements at Soudan, using the 60 keV line from a $^{241}$Am source to show the Luke-effect scaling at four $V_{bias}$ values (left) and the retention of excellent energy resolutoin even at high Luke-amplification values (right). Some TES saturation occurs at the highest $V_{bias}$ values, leading to some non-linearity in the measurement (though not in the Luke phonon production itself). Data and plots taken from R. Basu Thakur.

us:

- Because electron recoils are amplified by a factor of $\sim 7$ ($1/\mathrm{Y}_{NR}$) more than nuclear recoils, the rate of electron recoils *within a particular energy window* is now ~7 times less than before (compared to the rate of nuclear recoil events). In a sense, not only has the Luke effect amplified all events, it has 'swept up and away' the electron recoil events.

- Although event-by-event discrimination is lost, a beautiful *statistical* discrimination technique has been gained. By varying the bias voltage and observing how the spectrum scales (either as $V_{bias}/\epsilon_{ER}$ or $V_{bias}/\epsilon_{NR}$), it can be determined if the distribution represents an electron recoil or nuclear recoil population.

It is tentatively planned to run some fraction of SuperCDMS Soudan in this 70V Luke-amplification mode for some fraction of the exposure. A projection of the sensitivity of a relatively small exposure (2 iZIPs for 2 months) is shown in Figure 14.2.3.

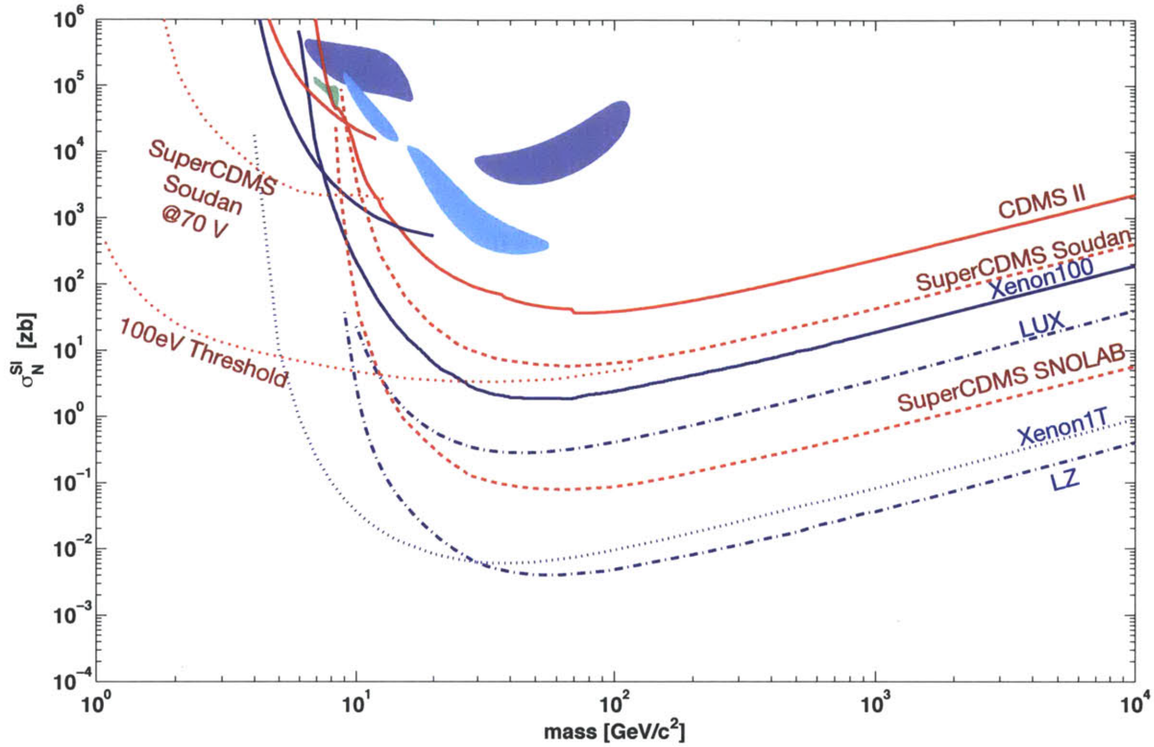

Figure 14-5: The expected sensitivities of some future experiments (together with existing limits and allowed regions, as in Figure 2-12). The results of the two main strategies are evident: increasing target mass pushes sensitivities toward smaller cross sections, decreasing detector threshold pushes sensitivities toward lower WIMP masses. The projection labeled '**SuperCDMS Soudan @70V**' simulates the sensitivity of two iZIP4/5 detectors running at $V_b$=70V (threshold of ~460 $keV_{nr}$), a 77 kg-day exposure, including a (totally undiscriminated) electron recoil background as predicted by extrapolating electron recoil background rates from measured values. Such an exposure will likely be taken over the coming months (Fall 2012). The projection labeled '**100 eV Threshold**' assumes a hypothetical exposure of 4 kg-yr (fid.) using a low-threshold Ge detector with perfect phonon-only fiducial volume definition and electron recoil rejection. The 100 eV threshold should be obtainable through a relatively moderate drop in $T_C$, to ~40 mK. Assumptions used in the high-mass projections are taken from recent conference proceedings, and can be summarized as **LUX**: 30,000 kg-day (fid.) with a 45% efficiency and a 5 keV threshold, **SuperCDMS SNOLAB**: 420 kg-year (fid.) with a 92% efficiency and a 10 keV threshold, **Xenon1T**: 3 ton-year (fid.) with a 45% efficiency and a 2 keV threshold **LZ**: 5,000 ton-days (fid.) with a 50% efficiency

## 14.3 Final Thoughts

As can be seen in Figure 14-1, the field of direct detection has been progressing *exponentially* for the past 15 years, starting with CDMS I in 1998. Only time will say what technology will be used to first probe cross sections orders of magnitude from where we now stand, but one thing is certain: we *will* probe those cross sections, and it won't be long. Perhaps some day not too many years down the road, graduate students will be writing their theses comparing the measurement systematics of multiple observed *WIMP signals*, and finally piecing together a clearer understanding of particle physics and the nature of the universe.

# Bibliography

[1] C. E. Aalseth et al. Results from a Search for Light-Mass Dark Matter with a P- type Point Contact Germanium Detector. *Phys. Rev. Lett.*, 106:131301, 2011.

[2] C. E. Aalseth et al. Search for an Annual Modulation in a p-Type Point Contact Germanium Dark Matter Detector. *Phys. Rev. Lett.*, 107:141301, 2011.

[3] R. Abbasi, Y. Abdou, T. Abu-Zayyad, M. Ackermann, J. Adams, K. Andeen, J. A. Aguilar, M. Ahlers, D. Altmann, J. Auffenberg, X. Bai, M. Baker, S. W. Barwick, R. Bay, J. L. Bazo Alba, K. Beattie, J. J. Beatty, S. Bechet, J. K. Becker, K.-H. Becker, M. Bell, M. L. Benabderrahmane, S. BenZvi, P. Zarzhitsky, and M. Zoll. Multiyear search for dark matter annihilations in the sun with the amanda-ii and icecube detectors. *Phys. Rev. D*, 85:042002, Feb 2012.

[4] M. Ackermann, M. Ajello, A. Albert, W. B. Atwood, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, B. Berenji, R. D. Blandford, E. D. Bloom, E. Bonamente, A. W. Borgland, J. Bregeon, M. Brigida, P. Bruel, R. Buehler, T. H. Burnett, S. Buson, G. A. Caliandro, and R. A. Cameron. Constraining dark matter models from a combined analysis of milky way satellites with the fermi large area telescope. *Phys. Rev. Lett.*, 107:241302, Dec 2011.

[5] S. Agostinelli et al. GEANT4: A simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.

[6] Z. Ahmed et al. Search for Weakly Interacting Massive Particles with the First Five-Tower Data from the Cryogenic Dark Matter Search at the Soudan Underground Laboratory. *Phys. Rev. Lett.*, 102:011301, 2009.

[7] Z. Ahmed et al. Dark matter search results from the cdms ii experiment. *Science*, 327(5973):1619–1621, 2010.

[8] Z. Ahmed et al. *Phys. Rev. Lett.*, 106:131302, 2011.

[9] D. S. Akerib et al. Exclusion limits on the WIMP nucleon cross-section from the first run of the Cryogenic Dark Matter Search in the Soudan underground lab. *Phys. Rev. D*, 72:052009, 2005.

[10] D. S. Akerib et al. Exclusion limits on the wimp-nucleon cross section from the first run of the cryogenic dark matter search in the soudan underground laboratory. *Phys. Rev.*, D72(5):052009, 2005.

[11] D. S. Akerib et al. Low-threshold analysis of cdms shallow-site data. *Phys. Rev. D*, 82(12):122004, Dec 2010.

[12] John Allison et al. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53:270, 2006.

[13] A. Anderson, S. Leman, M. Pyle, E. Figueroa-Feliciano, K. McCarthy, T. Doughty, M. Cherry, and B. Young. Simulations of noise in phase-separated transition-edge sensors for supercdms. *Journal of Low Temperature Physics*, 167:135–140, 2012. 10.1007/s10909-012-0555-1.

[14] J. Angle, E. Aprile, F. Arneodo, L. Baudis, A. Bernstein, A. I. Bolozdynya, L. C. C. Coelho, C. E. Dahl, L. Deviveiros, A. D. Ferella, L. M. P. Fernandes, S. Fiorucci, R. J. Gaitskell, K. L. Giboni, R. Gomez, R. Hasty, L. Kastens, J. Kwong, J. A. M. Lopes, N. Madden, A. Manalaysay, A. Manzur, D. N. McKinsey, M. E. Monzani, K. Ni, U. Oberlack, J. Orboeck, G. Plante, R. Santorelli, J. M. F. Dos Santos, S. Schulte, P. Shagin, T. Shutt, P. Sorensen, C. Winant, and M. Yamashita. Search for Light Dark Matter in XENON10 Data. *Physical Review Letters*, 107(5):051301, July 2011.

[15] J. Angle et al. *Phys. Rev. Lett.*, 107:051301, 2011.

[16] G. Angloher et al. Results from 730 kg days of the CRESST-II Dark Matter Search. *arXiv:1109.0702v1*, 2011.

[17] E. Aprile et al. New Measurement of the Relative Scintillation Efficiency of Xenon Nuclear Recoils Below 10 keV. *Phys. Rev.*, C79:045807, 2009.

[18] E. Aprile et al. *Phys. Rev. Lett.*, 107:131302, 2011.

[19] P. Barbeau. *Ph.D. thesis, U. Chicago*, 2009.

[20] P.S. Barbeau, J.I. Collar, and P.M. Whaley. Design and characterization of a neutron calibration facility for the study of sub-kev nuclear recoils. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 574(2):385 – 391, 2007.

[21] J. Bardeen, L. N. Cooper, and J. R. Schrieffer. Microscopic theory of superconductivity. *Phys. Rev.*, 106:162–164, Apr 1957.

[22] Giuseppe Battistoni et al. The FLUKA code: Description and benchmarking. *AIP Conf. Proc.*, 896:31–49, 2007.

[23] L Baudis, J Hellmig, H.V Klapdor-Kleingrothaus, Y Ramachers, J.W Hammer, and A Mayer. High-purity germanium detector ionization pulse shapes of nuclear recoils, interactions and microphonism. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 418:348 – 354, 1998.

[24] P. Belli et al. *Phys. Rev. D*, 84:055014, 2011.

[25] A. Benoit, L. Berge, J. Blumer, A. Broniatowski, B. Censier, A. Chantelauze, M. Chapellier, G. Chardin, S. Collin, X. Defay, M. De Jesus, H. Deschamps, P. Di Stefano, Y. Dolgorouky, L. Dumoulin, K. Eitel, M. Fesquet, S. Fiorucci, J. Gascon, G. Gerbier, C. Goldbach, M. Gros, M. Horn, A. Juillard, R. Lemrani, A. de Lesquen, M. Luca, S. Marnieros, L. Mosca, X.-F. Navick, G. Nollez, E. Olivieri, P. Pari, V. Sanglard, L. Schoeffel, F. Schwamm, and M. Stern. Measurement of the response of heat-and-ionization germanium detectors to nuclear recoils. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 577(3):558 – 568, 2007.

[26] A. Benoit, L. BergÌ©, J. BlÌ?mer, A. Broniatowski, B. Censier, A. Chantelauze, M. Chapellier, G. Chardin, S. Collin, X. Defay, M. De JÌ©sus, H. Deschamps, P. Di Stefano, Y. Dolgorouky, L. Dumoulin, K. Eitel, M. Fesquet, S. Fiorucci, J. Gascon, G. Gerbier, C. Goldbach, M. Gros, M. Horn, A. Juillard, R. Lemrani, A. de Lesquen, M. Luca, S. Marnieros, L. Mosca, X.-F. Navick, G. Nollez, E. Olivieri, P. Pari, V. Sanglard, L. Schoeffel, F. Schwamm, and M. Stern. Measurement of the response of heat-and-ionization germanium detectors to nuclear recoils. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 577(3):558 – 568, 2007.

[27] R. Bernabei, P. Belli, F. Cappella, R. Cerulli, C. J. Dai, A. D'Angelo, H. L. He, A. Incicchitti, H. H. Kuang, X. H. Ma, F. Montecchia, F. Nozzoli, D. Prosperi, X. D. Sheng, R. G. Wang, and Z. P. Ye. New results from DAMA/LIBRA. *European Physical Journal C*, 67:39–49, May 2010.

[28] R. Bernabei et al. First results from DAMA/LIBRA and the combined results with DAMA/NaI. *Eur. Phys. J.*, C56:333–355, 2008.

[29] R. Bernabei et al. New results from DAMA/LIBRA. *Eur. Phys. J. C*, 67:39–49, May 2010.

[30] J. Bovy and S. Tremaine. On the local dark matter density. *ArXiv e-prints*, May 2012.

[31] Nassim Bozorgnia, Graciela B. Gelmini, and Paolo Gondolo. Channeling in direct dark matter detection ii: channeling fraction in si and ge crystals. *Journal of Cosmology and Astroparticle Physics*, 2010(11):028, 2010.

[32] P.L. Brink. Non-equilibrium superconductivity induced by x-ray photons. *PhD thesis, Oxford University*, 1995.

[33] P.L. Brink. *personal communication*, 2012.

[34] P.L. Brink, B. Cabrera, J.P. Castle, J. Cooley, L. Novak, R.W. Ogburn, M. Pyle, J. Ruderman, A. Tomada, B.A. Young, J. Filippini, P. Meunier, N. Mirabolfathi, B. Sadoulet, D.N. Seitz, B. Serfass, K.M. Sundqvist, D.S. Akerib, C.N. Bailey, M.R. Dragowsky, D.R. Grant, R. Hennings-Yeomans, and R.W. Schnee. First test runs of a dark-matter detector with interleaved ionization electrodes and phonon sensors for surface-event rejection. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 559(2):414 – 416, 2006. Proceedings of the 11th International Workshop on Low Temperature Detectors : LTD-11 (11th International Workshop on Low Temperature Detectors).

[35] A. Broniatowski, X. Defay, E. Armengaud, L. BergÃ©, A. Benoit, O. Besida, J. BlÃ¼mer, A. Chantelauze, M. Chapellier, G. Chardin, F. Charlieux, S. Collin, O. Crauste, M. De Jesus, P. Di Stefano, Y. Dolgorouki, J. Domange, L. Dumoulin, K. Eitel, J. Gascon, G. Gerbier, M. Gros, M. Hannawald, S. HervÃ©, A. Juillard, H. Kluck, V. Kozlov, R. Lemrani, A. Lubashevskiy, C. Marrache, S. Marnieros, X.-F. Navick, C. Nones, E. Olivieri, P. Pari, B. Paul, S. Rozov, V. Sanglard, S. Scorza, S. Semikh, M.-A. Verdier, L. Vagneron, and E. Yakushev. A new high-background-rejection dark matter ge cryogenic detector. *Physics Letters B*, 681(4):305 – 309, 2009.

[36] Blas Cabrera. Oblique propagation. *Presented at CDMS Detector Monte Carlo Workshop*, 2009.

[37] B. Censier. Etude et optimisation de la voie ionisation dans lÕexperience edelweiss. *Ph.D. Thesis, IPN Orsay*, 2006.

[38] C. Chasman, K. W. Jones, H. W. Kraner, and Werner Brandt. Band-gap effects in the stopping of ge$^{72*}$ atoms in germanium. *Phys. Rev. Lett.*, 21:1430–1433, Nov 1968.

[39] C. Chasman, K. W. Jones, and R. A. Ristinen. Measurement of the energy loss of germanium atoms to electrons in germanium at energies below 100 kev. *Phys. Rev. Lett.*, 15:245–248, Aug 1965.

[40] C. Chasman, K. W. Jones, R. A. Ristinen, and J. T. Sample. Measurement of the energy loss of germanium atoms to electrons in germanium at energies below 100 kev. ii. *Phys. Rev.*, 154:239–244, Feb 1967.

[41] J. Collar. *private communication*, 2011.

[42] J. I. Collar. Talk at TAUP 2011 Workshop, Munich, Germany, Sep. 5-9, 2011.

[43] J. I. Collar. Talk at IDM2012 Conference, Chicago, IL, July 2012.

[44] J. I. Collar and N. E. Fields. A Maximum Likelihood Analysis of Low-Energy CDMS Data. *ArXiv e-prints*, April 2012.

[45] Brian L. Dougherty. Measurements of ionization produced in silicon crystals by low-energy silicon atoms. *Phys. Rev. A*, 45:2104–2107, Feb 1992.

[46] Andrzej K. Drukier, Katherine Freese, and David N. Spergel. Detecting cold dark-matter candidates. *Phys. Rev. D*, 33:3495–3508, 1986.

[47] John R. Ellis, Keith A. Olive, Yudi Santoso, and Vassilis C. Spanos. Update on the direct detection of supersymmetric dark matter. *Phys. Rev.*, D71:095007, 2005.

[48] Lindhard J. *et al. K. Dan. Viderask. Selsk., Math. Fys. Medd*, 33 and 36(10 and 36), 1963 and 1968.

[49] M. Loidl et al. Quasiparticle diffusion over several mm in cryogenic detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 465(2 - 3):440 – 446, 2001.

[50] R. Keisler et al. *Astrophys. J.*, 743:28, 2011.

[51] M. Farina, D. Pappadopulo, A. Strumia, and T. Volansky. Can CoGeNT and DAMA modulations be due to Dark Matter? *JCAP*, 11:10, November 2011.

[52] A. Fasso, A. Ferrari, J. Ranft, and P. R. Sala. CERN-2005-010, 2005.

[53] Gary J. Feldman and Robert D. Cousins. Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D*, 57:3873–3889, Apr 1998.

[54] E. Figueroa-Feliciano. Theory and development of position-sensitive quantum calorimeters. *PhD thesis, Stanford University*, 2001.

[55] P. J. Fox, J. Liu, and N. Weiner. Integrating out astrophysical uncertainties. *Phys. Rev. D*, 83(10):103514, may 2011.

[56] Patrick J. Fox et al. A cogent modulation analysis. *arXiv:1107.0717v2*, 2011.

[57] G. Gerbier, E. Lesquoy, J. Rich, M. Spiro, C. Tao, D. Yvon, S. Zylberajch, P. Delbourgo, G. Haouat, C. Humeau, F. Goulding, D. Landis, N. Madden, A. Smith, J. Walton, D. O. Caldwell, B. Magnusson, M. Witherell, B. Sadoulet, and A. Da Silva. Measurement of the ionization of slow silicon nuclei in silicon for the calibration of a silicon dark-matter detector. *Phys. Rev. D*, 42:3211–3214, Nov 1990.

[58] D. Hooper et al. Consistent dark matter interpretation for cogent and dama/libra. *Phys. Rev. D*, 82(12):123509, Dec 2010.

[59] Dan Hooper. The empirical case for 10 gev dark matter. *arXiv:1201.1303v1*, 2012.

[60] Dan Hooper and Stefano Profumo. Dark matter and collider phenomenology of universal extra dimensions. *Physics Reports*, 453(2â4):29 – 115, 2007.

[61] Ioffe Physico-Technical Institute. New semiconductor materials. characteristics and properties. *Electronic archive: www.ioffe.ru/SVA/NSM/.*

[62] K. Irwin. Phonon-mediated particle detection using superconducting tungsten transition-edge-sensors. *PhD thesis, Stanford University*, 1995.

[63] K.D. Irwin and G.C. Hilton. Transition-edge sensors. In Christian Enss, editor, *Cryogenic Particle Detection*, volume 99 of *Topics in Applied Physics*, pages 81–97. Springer Berlin / Heidelberg.

[64] N. Jarosik, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, M. Halpern, R. S. Hill, G. Hinshaw, A. Kogut, E. Komatsu, D. Larson, M. Limon, S. S. Meyer, M. R. Nolta, N. Odegard, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results. *Astrophysical Journal Supplement*, 192:14, February 2011.

[65] K. W. Jones and H. W. Kraner. Stopping of 1- to 1.8-kev $^{73}$Ge atoms in germanium. *Phys. Rev. C*, 4:125–129, Jul 1971.

[66] K. W. Jones and H. W. Kraner. Energy lost to ionization by 254-ev $^{73}$Ge atoms stopping in ge. *Phys. Rev. A*, 11:1347–1353, Apr 1975.

[67] Chris Kelso, Dan Hooper, and Matthew R. Buckley. *arXiv:1110.5338v1*, 2011.

[68] C. Kittel. Introduction to solid state physics, 8th edition. 2005.

[69] C.A. Klein. Bandgap dependence and related features of radiation ionization energies in semiconductors. *Journal of Applied Physics*, 39(4):2029–2038, 1968. cited By (since 1996) 162.

[70] Joachim Kopp, Thomas Schwetz, and Jure Zupan. *arXiv:1110.2721v1*, 2011.

[71] T.I. Kucher.

[72] M. Kuźniak, M. G. Boulay, and T. Pollmann. Surface roughness interpretation of 730 kg days CRESST-II results. *ArXiv e-prints*, March 2012.

[73] V. N. Lebedenko et al. Result from the First Science Run of the ZEPLIN-III Dark Matter Search Experiment. *Phys. Rev.*, D80:052010, 2009.

[74] S. Leman, S. Hertel, P. Kim, B. Cabrera, E. Do Couto E Silva, E. Figueroa-Feliciano, K. McCarthy, R. Resch, B. Sadoulet, and K. Sundqvist. Comparison of cdms [100] and [111] oriented germanium detectors. *Journal of Low Temperature Physics*, 167:1106–1111, 2012. 10.1007/s10909-011-0427-0.

[75] S. W. Leman. Review Article: Physics and Monte Carlo Techniques as Relevant to Cryogenic, Phonon and Ionization Readout of CDMS Radiation-Detectors. *ArXiv e-prints*, September 2011.

[76] J. D. Lewin and P. F. Smith. Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil. *Astropart. Phys.*, 6:87–112, 1996.

[77] J. Lindhard et al. *K. Dan. Vidensk. Selsk., Mat.-Fys. Medd.*, 33:10, 1963.

[78] M. R. Lovell, V. Eke, C. S. Frenk, L. Gao, A. Jenkins, T. Theuns, J. Wang, S. D. M. White, A. Boyarsky, and O. Ruchayskiy. The haloes of bright satellite galaxies in a warm dark matter universe. *Monthly Notices of the Royal Astronomical Society*, 420:2318–2324, March 2012.

[79] P. N. Luke. Voltage-assisted calorimetric ionization detector. *J. Appl. Phys.*, 64:6858–6860, December 1988.

[80] P. N. Luke. Voltage-assisted calorimetric ionization detector. *Journal of Applied Physics*, 64(12):6858 –6860, dec 1988.

[81] P. N. Luke. Single-polarity charge sensing in ionization detectors using coplanar electrodes. *Applied Physics Letters*, 65(22):2884 –2886, nov 1994.

[82] D. Malling. Talk at IDM2012 Conference, Chicago, IL, July 2012.

[83] K. McCarthy, S. Leman, A. Anderson, D. Brandt, P. Brink, B. Cabrera, M. Cherry, E. Do Couto E Silva, P. Cushman, T. Doughty, E. Figueroa-Feliciano, P. Kim, N. Mirabolfathi, L. Novak, R. Partridge, M. Pyle, A. Reisetter, R. Resch, B. Sadoulet, B. Serfass, K. Sundqvist, and A. Tomada. Validation of phonon physics in the cdms detector monte carlo. *Journal of Low Temperature Physics*, 167:1160–1166, 2012. 10.1007/s10909-012-0474-1.

[84] Y. Messous, B. Chambon, V. Chazal, M. De Jesus, D. Drain, C. Pastor, A. de Bellefon, M. Chapellier, G. Chardin, E. Gaillard-Lecanu, G. Gerbier, Y. Giraud-Heraud, D. Lhote, J. Mallet, L. Mosca, M.-C. Perillo-Isaac, C. Tao, and D. Yvon. Calibration of a ge crystal with nuclear recoils for the development of a dark matter detector. *Astroparticle Physics*, 3(4):361 – 366, 1995.

[85] M. E. Msall and J. P. Wolfe. Phonon production in weakly photoexcited semiconductors: quasidiffusion in ge, gaas, and si. *Phys. Rev. B*, 56:9557–9564, Oct 1997.

[86] B. Neganov and V. Trofimov. *Otkryt. Izobret.*, 146:215, 1985.

[87] B. S. Neganov and V. N. Trofimov. *Journal of Experimental and Theoretical Physics*, 28:328, 1978.

[88] G. Nilsson and G. Nelin. Study of the homology between silicon and germanium by thermal-neutron spectrometry. *Phys. Rev. B*, 6:3777–3786, Nov 1972.

[89] K. Nordlund, M. Ghaly, R. S. Averback, M. Caturla, T. Diaz de la Rubia, and J. Tarus. Defect production in collision cascades in elemental semiconductors and fcc metals. *Phys. Rev. B*, 57:7556–7570, Apr 1998.

[90] M. Pettini, B. J. Zych, M. T. Murphy, A. Lewis, and C. C. Steidel. Deuterium abundance in the most metal-poor damped Lyman alpha system: converging on $\Omega_{b,0}\mathrm{h}^2$. *Monthly Notices of the Royal Astronomical Society*, 391:1499–1510, December 2008.

[91] M. Pospelov and J. Pradler. Big Bang Nucleosynthesis as a Probe of New Physics. *Annu. Rev. Nucl. Part. Sci.*, 60:539, 2010.

[92] M. Pyle. Optimizing the design and analysis of cryogenic semiconductor dark matter detectors for maximum sensitivity. *PhD thesis, Stanford University*, 2012.

[93] M. Pyle, P.L. Brink, B. Cabrera, J.P. Castle, P. Colling, C.L. Chang, J. Cooley, T. Lipus, R.W. Ogburn, and B.A. Young. Quasiparticle propagation in aluminum fins and tungsten tes dynamics in the cdms zip detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 559(2):405 – 407, 2006. ¡ce:title¿Proceedings of the 11th International Workshop on Low Temperature Detectors¡/ce:title¿ ¡ce:subtitle¿LTD-11¡/ce:subtitle¿ ¡xocs:full-name¿11th International Workshop on Low Temperature Detectors¡/xocs:full-name¿.

[94] Leszek Roszkowski, Roberto Ruiz de Austri, and Roberto Trotta. Implications for the Constrained MSSM from a new prediction for $b \to s\gamma$. *JHEP*, 07:075, 2007.

[95] V. Sanglard and Edelweiss Collaboration. EDELWEISS-II Dark Matter search: Status and first results. *Journal of Physics Conference Series*, 203(1):012037, January 2010.

[96] A. R. Sattler. Ionization produced by energetic silicon atoms within a silicon lattice. *Phys. Rev.*, 138:A1815–A1821, Jun 1965.

[97] A. R. Sattler, F. L. Vook, and J. M. Palms. Ionization produced by energetic germanium atoms within a germanium lattice. *Phys. Rev.*, 143:588–594, Mar 1966.

[98] C. Savage, G. Gelmini, P. Gondolo, and K. Freese. Compatibility of dama/libra dark matter detection with other searches. *Journal of Cosmology and Astroparticle Physics*, 2009(04):010, 2009.

[99] T. Schwetz and J. Zupan. Dark matter attempts for CoGeNT and DAMA. *JCAP*, 8:8, August 2011.

[100] T. Shutt, B. Ellman, P. D. Barnes, A. Cummings, A. Da Silva, J. Emes, Y. Giraud-Héraud, E. E. Haller, A. E. Lange, R. R. Ross, J. Rich, B. Sadoulet, G. Smith, W. Stockwell, C. Stubbs, N. Wang, S. White, B. A. Young, and D. Yvon. Measurement of ionization and phonon production by nuclear recoils in a 60 g crystal of germanium at 25 mk. *Phys. Rev. Lett.*, 69:3425–3427, Dec 1992.

[101] E. Simon, L. Berge, A. Broniatowski, R. Bouvier, B. Chambon, M. De Jesus, D. Drain, L. Dumoulin, J. Gascon, J.-P. Hadjout, A. Juillard, O. Martineau, C. Pastor, M. Stern, and L. Vagneron. Sicane: a detector array for the measurement of nuclear recoil quenching factors using a monoenergetic neutron beam. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 507(3):643 – 656, 2003.

[102] J. D. Simon and M. Geha. The Kinematics of the Ultra-faint Milky Way Satellites: Solving the Missing Satellite Problem. *The Astrophysical Journal*, 670:313–331, November 2007.

[103] Y. Sofue, M. Honma, and T. Omodaka. Unified Rotation Curve of the Galaxy – Decomposition into de Vaucouleurs Bulge, Disk, Dark Halo, and the 9-kpc Rotation Dip –. *Publications of the Astrophysical Society of Japan*, 61:227–, February 2009.

[104] K. Sundqvist and B. Sadoulet. Detectors of the cryogenic dark matter search: Charge transport and phonon emission in ge ?100? crystals at 40Êmk. *Journal of Low Temperature Physics*, 151:443–447, 2008. 10.1007/s10909-007-9666-5.

[105] S. Tamura and H. J. Maris. Spontaneous decay of ta phonons. *Phys. Rev. B*, 31:2595–2598, Feb 1985.

[106] Y.P. Varshni. Temperature dependence of the energy gap in semiconductors. *Physica*, 34(1):149 – 154, 1967.

[107] G. Wang. Phonon emission in germanium and silicon by electrons and holes in applied electric field at low temperature. *Journal of Applied Physics*, 107, 2010.

[108] G. Wang. Phonon emission in germanium and silicon by electrons and holes in applied electric field at low temperature. *Journal of Applied Physics*, 107(9):094504 –094504–7, may 2010.

[109] Siqing Wei and M. Y. Chou. Phonon dispersions of silicon and germanium from first-principles calculations. *Phys. Rev. B*, 50:2221–2226, Jul 1994.

[110] Christoph Weniger. A tentative gamma-ray line from dark matter annihilation at the fermi large area telescope. *Journal of Cosmology and Astroparticle Physics*, 2012(08):007, 2012.

[111] J. Wolfe. Imaging electron-phonon interactions - a retrospective. *Klemen's Award Talk at Phonons 2010.*

[112] XENON100 Collaboration, E. Aprile, M. Alfonsi, K. Arisaka, F. Arneodo, C. Balan, L. Baudis, B. Bauermeister, A. Behrens, P. Beltrame, K. Bokeloh, E. Brown, G. Bruno, R. Budnik, J. M. R. Cardoso, W.-T. Chen, B. Choi, D. Cline, A. P. Colijn, H. Contreras, J. P. Cussonneau, M. P. Decowski, E. Duchovni, S. Fattori, A. D. Ferella, W. Fulgione, F. Gao, M. Garbini, C. Ghag, K.-L. Giboni, L. W. Goetzke, C. Grignon, E. Gross, W. Hampel, F. Kaether, H. Kettling, A. Kish, J. Lamblin, H. Landsman, R. F. Lang, M. Le Calloch, C. Levy, K. E. Lim, Q. Lin, S. Lindemann, M. Lindner, J. A. M. Lopes, K. Lung, T. Marrodan Undagoitia, F. V. Massoli, A. J. Melgarejo Fernandez, Y. Meng, A. Molinario, E. Nativ, K. Ni, U. Oberlack, S. E. A. Orrigo, E. Pantic, R. Persiani, G. Plante, N. Priel, A. Rizzo, S. Rosendahl, J. M. F. dos Santos, G. Sartorelli, J. Schreiner, M. Schumann, L. Scotto Lavina, P. R. Scovell, M. Selvi, P. Shagin, H. Simgen, A. Teymourian, D. Thers, O. Vitells, H. Wang, M. Weber, and C. Weinheimer. Dark Matter Results from 225 Live Days of XENON100 Data. *ArXiv e-prints*, July 2012.

[113] S. Yellin. Finding an upper limit in the presence of unknown background. *Phys. Rev.*, D66:032005, 2002.

[114] D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, S. Burles, L. Carey, M. A. Carr, F. J. Castander, B. Chen, P. L. Colestock, A. J. Connolly, J. H. Crocker, I. Csabai, P. C. Czarapata, J. E. Davis, M. Doi, T. Dombeck, D. Eisenstein, N. Ellman, B. R. Elms, M. L. Evans, X. Fan, G. R. Federwitz, L. Fiscelli, S. Friedman, J. A. Frieman, M. Fukugita, B. Gillespie, J. E. Gunn, V. K. Gurbani, E. de Haas, M. Haldeman, Harris, and SDSS Collaboration. The Sloan Digital Sky Survey: Technical Summary. *The Astronomical Journal*, 120:1579–1587, September 2000.

[115] P. Zecher, D. Wang, J. Rapaport, C. J. Martoff, and B. A. Young. Energy deposition of energetic silicon atoms within a silicon lattice. *Phys. Rev. A*, 41:4058–4061, Apr 1990.

[116] F. Zwicky. *Helv. Phys. Acta.*, 6:110, 1933.