First Results from the Cryogenic Dark Matter Search Experiment at the Deep Site

by

Vuk Mandic

B.S. (California Institute of Technology)

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Physics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Bernard Sadoulet, Chair
Professor Adrian T. Lee
Professor Wm. J. Welch

Fall 2004

The dissertation of Vuk Mandic is approved:

6/11/04

Chair Date

6/3/04

Date

Wm J Welch 6/11/04

Date

University of California, Berkeley

Fall 2004

First Results from the Cryogenic Dark Matter Search Experiment at the Deep Site

by

Vuk Mandic

Abstract

First Results from the Cryogenic Dark Matter Search Experiment at the Deep Site

by

Vuk Mandic

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Bernard Sadoulet, Chair

The Cryogenic Dark Matter Search (CDMS) experiment is designed to search for dark matter in the form of the Weakly Interacting Massive Particles (WIMPs). For this purpose, CDMS uses detectors based on crystals of Ge and Si, operated at the temperature of 20 mK, and providing a two-fold signature of an interaction: the ionization and the athermal phonon signals. The two signals, along with the passive and active shielding of the experimental setup, and with the underground experimental sites, allow very effective suppression and rejection of different types of backgrounds.

This dissertation presents the commissioning and the results of the first WIMP-search run performed by the CDMS collaboration at the deep underground site at the Soudan mine in Minnesota. We develop different methods of suppressing the dominant background due to the electron-recoil events taking place at the detector surface and we apply these algorithms to the data set. These results place the world's most sensitive limits on the WIMP-nucleon spin-independent elastic-scattering cross-section. Finally, we examine the compatibility of the supersymmetric WIMP-models with the direct-detection experiments (such as CDMS) and discuss the implications of the new CDMS result on these models.

6/1/04

Professor Bernard Sadoulet
Dissertation Committee Chair

To my family

Contents

List of Figures

List of Tables

Acknowledgments

I thank my family: Mina, Pinar, Milan, mom, dad, and Rosemary - your constant support over the years has made this remarkable experience possible. This dissertation is dedicated to you. Thank you!

I also thank my dissertation committee, Profs. B. Sadoulet, A.T. Lee, and W.J. Welch, for valuable comments on this dissertation. Special thanks goes to my advisor, Bernard Sadoulet, who has set a remarkable example of a true intellectual and a scientist. His insight on a number of issues, ranging from the technical details on the Soudan grounding scheme to various statistical techniques and global issues in cosmology have been invaluable to my education. I particularly appreciate the freedom he has given me to pursue my own interests and to follow my own intuition.

I also thank: Rick Gaitskell who introduced me to many aspects of CDMS and was willing to spend hours explaining them to a starting graduate student; Dennis Seitz who taught me much about electronics and hardware; Richard Schnee who taught me (and so many others) how to do data analysis properly; Blas Cabrera for the numerous insights on detectors, cryogenics etc (and for showing that taking a hammer to the fridge is not always a bad thing); Dan Bauer whose remarkable control over so many issues at Soudan has kept me sane during my stays there; Dan Akerib, Paul Brink, and Patrizia Meunier who taught me so much about the dilution refrigerators and the low temperature physics; Wolfgang Rau whose remarkable clarity of thought was so helpful on so many occasions; Tarek Saab for all his help during the Run21 and for so many good chess-games; Clarence Chang for sharing with me the coordination challenges of the analysis of the first Soudan data; Walter Ogburn for sharing with me the cryogenic horror of May 22, 2003, and for introducing me to the Settlers.

I also thank to everyone else in the Berkeley group, who have made my stay at Berkeley inspiring and enjoyable: Judith Alvaro-Dean, Scott Armel, Elizabeth Arscott, Miguel Daal, Caryl Esteves, John Emes, Jeff Filippini, Sunil Golwala, Jochen Hellmig, Chris Hernikl, Maria Isaac, Al Lu, Mary MacCready, Nader Mirabolfathi, Ron Ross, Bruno Serfass, Garth Smith, Tony Spadafora, Kyle Sundqvist and many others. Special thanks to Caryl Esteves, Mary MacCready, Anne Takizawa, and Donna Sakima for helping me find my way through the university's bureaucratic maze.

Finally, I thank all those with whom I shared the Soudan (and other) struggles: Mike Attisha, Laura Baudis, Jim Beaty, Ray Bunker, David Caldwell, Roland Clarke, Rodney Choate, Prisca Cushman, Mike Crisler, Mike Dragowsky, Don Driscoll, Long Duong, Ron Ferril, Bryan Johnson, Sharmila Kamat, Lou Kula, Bruce Lambin, Rupak Mahapatra, Harry Nelson, Robert Nelson, Stan Orr, Thushara Perera, Angie Reisetter, Joel Sander, Chris Savage, Rich Schmitt, J.P. Thompson, Gensheng Wang, James Williams, Steve Yellin, Betty Young and many others.

Thank you all for a memorable experience!

Chapter 1

Introduction

1.1 Introduction

Understanding the content of the Universe has been a long lasting problem of cosmology. Recently, exciting developments in the fields of Cosmic Microwave Background (CMB), Large Scale Structure (LSS) observations, and Type Ia supernovae observations allowed more accurate measurements of the various parameters of the "Standard Model" of cosmology. In particular, the Universe appears to be dominated ($\sim 70\%$) by the vacuum energy density, or Dark Energy, while the remaining $\sim 30\%$ is made up of matter clumps. Furthermore, most of the matter part of the Universe appears to be non-luminous (i.e. dark) and non-baryonic. The composition of Dark Matter itself is still an open question and subject to intense research activity.

The Dark Matter problem is an old one - the first evidence dates back to Zwicki in 1933 [1]. Since then, many observational pieces of evidence have been found in support of the existence of Dark Matter. Similarly, many candidates, with very different characteristics, have been proposed to account for the missing mass of the Universe. The remainder of this Chapter summarizes the current situation in the field. Section 2 introduces the Standard Model of Cosmology. Section 3 summarizes the various types of evidence supporting the existence of Dark Matter and of its non-baryonic component. Section 4 gives an overview of the different proposed Dark Matter candidates. Sections 5 and 6 give an overview of the experiments searching for Dark Matter in the form of Weakly Interacting Massive Particles (WIMPs).

Note that the Ph.D. dissertation by Sunil Golwala [2] gives an excellent review of the field as it was in 2000. We will occasionally refer to this review.

1.2 Cosmology Framework

This Section develops the framework in which the problem of Dark Matter will be introduced and discussed. We follow the discussion of standard textbooks, such as [3].

We start from the Einstein field equations:

$$R_{\mu\nu} - \frac{1}{2}\mathcal{R}g_{\mu\nu} = 8\pi G T_{\mu\nu} + \Lambda g_{\mu\nu}, \tag{1.1}$$

where $R_{\mu\nu}$ and $\mathcal{R}$ are the Ricci tensor and the scalar curvature corresponding to the metric $g_{\mu\nu}$, G is the gravitational constant, $T_{\mu\nu}$ is the stress-energy tensor, and Λ is the cosmological constant. Although there is no intrinsic motivation for the cosmological constant, this term is allowed in the Einstein equation. It represents the possibility that there is a density and a (negative) pressure associated with the "empty" space. In particle physics, this is explained by the zero-point energy - particles and antiparticles are continuously created out of vacuum, only to shortly annihilate into each other. Such processes give vacuum a non-zero potential energy, hence the cosmological constant. However, theoretical estimates of the vacuum zero-point energy (such as [4]) are larger than the observations (further discussed in the Section 1.3) by a factor of 10^{120}! Hence, the value of the vacuum zero-point energy is one of the most important unresolved problems in the particle physics today.

Assuming that the Universe is homogeneous and isotropic leads to the Friedmann-Robertson-Walker metric,

$$ds^2 = dt^2 - R^2(t)\left(\frac{dr^2}{1-kr^2} + r^2\, d\theta^2 + r^2\ \sin^2\theta\, d\phi^2\right). \tag{1.2}$$

Here, t is time, (r, θ, ϕ) are comoving spherical coordinates, k sets the curvature ($k =$ -1, 0, or 1 for open, flat, or closed geometry, respectively), and R is the scale factor. Furthermore, to be consistent with the symmetries of the metric, the stress-energy tensor must also be diagonal, with spatial components equal. Hence, we can write the diagonal terms of $T_{\mu\nu}$ as $[\rho, -p, -p, -p]$. Usually, one assumes for simplicity $p = w\rho$, where w depends on the type of matter (e.g. 0 for matter dominated Universe, 1/3 for radiation dominated, and -1 for vacuum dominated). Then, the 0-0 (i.e. the time-time) component of the Einstein equation yields the Friedmann Equation:

$$\frac{\dot{R}^2}{R^2} + \frac{k}{R^2} = \frac{8\pi G\rho}{3} + \frac{\Lambda}{3}. \tag{1.3}$$

Define the Hubble parameter $H = \dot{R}/R$ and

$$\Omega_m = \frac{8\pi G}{3H_0^2}\,\rho_0 = \frac{\rho_0}{\rho_c} \tag{1.4}$$

$$\Omega_k = -\frac{k}{H_0^2 R_0^2} \tag{1.5}$$

$$\Omega_\Lambda = \frac{\Lambda}{3H_0^2}, \tag{1.6}$$

where the subscript 0 denotes the current values and ρ_c is the current critical density. Assuming that the matter part is dominated by regular matter, so that the matter density scales as R^{-3} (alternatively, the matter part could be dominated by radiation, in which case it would scale as R^{-4}), we can rewrite the Friedmann equation as

$$\frac{H^2}{H_0^2} = \Omega_m \left(\frac{R_0}{R}\right)^3 + \Omega_k \left(\frac{R_0}{R}\right)^2 + \Omega_\Lambda. \tag{1.7}$$

For the present time $R_0 = R$, so the equation becomes

$$1 = \Omega_m + \Omega_k + \Omega_\Lambda. \tag{1.8}$$

Hence, $\Omega_{tot} = \Omega_m + \Omega_\Lambda$ is directly related to the curvature: $\Omega_{tot} = 1$ for the flat geometry, $\Omega_{tot} > 1$ for the closed geometry, and $\Omega_{tot} < 1$ for the open geometry.

Similarly, the $i - i$ (i.e. space-space) component of the Einstein equation leads to

$$2\frac{\ddot{R}}{R} + \frac{\dot{R}^2}{R^2} + \frac{k}{R^2} = -8\pi G p + \Lambda. \tag{1.9}$$

Together with Equation 1.3, this yields an equation on the acceleration of the Universe:

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}. \tag{1.10}$$

Or, in terms of Ω's (again, assuming matter domination and $w = 0$):

$$\frac{\ddot{R}}{RH_0^2} = -\frac{1}{2}\Omega_m \left(\frac{R_0}{R}\right)^3 + \Omega_\Lambda. \tag{1.11}$$

Intuitively, the Friedmann equation 1.7 describes the energy budget of the Universe at any point in time: the kinetic energy of the expansion is equal to the (negative of) gravitational potential energy due to matter, curvature and vacuum energy. Similarly, the acceleration equation 1.10 describes the dynamics of the Universe: the acceleration of the expansion is determined by the decelerating effect due to matter and the accelerating effect due to the vacuum energy. Both equations, and, therefore, the behavior of the Universe, are determined by the two parameters Ω_m and Ω_Λ.

The model outlined above is often referred to as the Standard Model of Cosmology. It is "standard" because its framework is remarkably successful in explaining a number of observations, ranging from the abundances of light nuclei to the structure formation and to the radiation remnants from the early Universe. We should note, however, that the model is considerably more complex than outlined above. For example, the matter part of the Universe is composed of different types of matter, relativistic or non-relativistic, baryonic or non-baryonic etc. All these components behave differently with time and affect the evolution of the Universe (and its structure that we observe today) in different ways.

Solving the Equations 1.3 and 1.10 gives the history of the Universe. In this model, therefore, the Universe starts as a very hot, expanding mixture of all elementary particles. Initially, due to high temperature, most particle interactions are allowed and dynamic equilibria are established. However, the Universe cools as it expands. With time, rates of certain interactions become significantly smaller than the expansion rate of the Universe. As a result, some equilibria are not maintainable any more, and asymmetries appear. We mention some of these events here, as we will refer to them later in this Chapter:

- When the Universe was ~ 1 s old, its temperature dropped below ~ 1 MeV, at which point the weak interactions (such as $e\nu \leftrightarrow e\nu$) became too weak (compared to the Universe expansion) to keep the neutrinos in the equilibrium. Hence, the neutrinos decoupled from the primordial plasma. These primordial neutrinos survived until today, and since they are believed to have non-zero mass, they account for some fraction of the dark matter (although not a significant fraction, as we will see later).

- At about the same time, the weak interactions that kept the neutrons and protons in an equilibrium (such as $n\nu \leftrightarrow n\nu$) froze out - the protons and neutrons became relatively stable and the first light nuclei started to form. The Big Bang nucleosynthesis was finished when the Universe was about 1-3 minutes old. Understanding this process provides one way of deducing the density of baryons in the Universe today.

- When the Universe was about 100,000 years old, its temperature dropped below a few eV. Until this point, the photon-electron interactions were significant and above 13.6 eV, so neutral atoms were not stable. After this point, the photons could not ionize the atoms any more, so they decoupled from the primordial plasma, which, in turn, became a mixture of stable neutral atoms. This process is known as the recombination. The photons that decoupled at this point represent the Cosmic Microwave Background (CMB) today, and carry much information about the parameters of the Standard Model of Cosmology. In fact, the CMB observations reveal small anisotropies which are believed to be the seeds that started the formation of the structures we observe today (stars, galaxies, clusters etc).

Finally, we mention the concept of redshift, as we will use it later. Redshift z is defined as:

$$1 + z = \frac{R_0}{R}. \tag{1.12}$$

It can be shown that redshift is also the factor by which the wavelength of light (emitted at scale R) is stretched due to the expansion of the Universe.

1.3 Evidence for Existence of Dark Matter and Dark Energy

In this Section, we summarize different observational pieces of evidence that support existence of dark matter and dark energy. Much of the earlier (and more intuitive) observations were focussed on studying the gravitational effects of dark matter at various scales: galaxies, clusters etc. More recently, accurate measurements of the CMB anisotropy were made, and more complete surveys of galaxies and clusters were completed. Along with the success of the Big Bang nucleosynthesis and the supernovae Ia studies, these developments allowed more accurate estimates of the various parameters of the Standard Model of Cosmology, including Ω_Λ, Ω_m, and the various components of Ω_m.

1.3.1 Spiral Galaxies

Maybe the most intuitive argument for existence of dark matter comes from the rotational curves of spiral galaxies. The rotational curve (RC) depicts the rotational velocity as a function of distance from the galactic center. The velocity is usually measured using atomic lines for stars or the 21 cm H line for the hydrogen clouds around the galaxy. It is well established that these curves do not follow the expected Keplerian fall-off based on the luminous matter, hence suggesting the existence of an invisible component of matter.

As an example, we cite the work of Persic, Salucci, and Stel [5]. They performed a careful study of about 1100 spiral galaxies ranging over a factor of 75 in luminosity, and split into 11 luminosity bins. In order to compare galaxies of different sizes, they define the optical radius R_{opt} containing 83% of the integrated light. They point out that as the luminosity decreases, the luminous matter is progressively unable to trace the observed RCs. As shown in Figure 1.1, the discrepancy is measurable even within R_{opt}, but it becomes very obvious at distances $> R_{opt}$, where the luminous matter cannot explain the flattening of the RCs. Furthermore, the distribution of the dark component is different from the distribution of the luminous matter - it is much less concentrated and it extends beyond the visible disc. However, maybe the most striking feature of the RCs is that they can be described by a "Universal Rotation Curve" which depends on a single global parameter, such as luminosity. Figure 1.1 shows that such curve fits the data well in all 11 luminosity bins.

Other studies, such as [6], [7], and [8], have examined in more detail the distribution of the dark component, which can be extracted from the RCs. They found that the dark matter haloes have a roughly constant central density region that extends beyond the luminous disc. These halo profiles disagree with the cuspy density distribution models typical of the Cold dark matter (CDM) models, but are well described, for example, by the empirical Burkert density profile:

$$\rho(r) = \frac{\rho_0 r_0^3}{(r + r_0)(r^2 + r_0^2)}. \tag{1.13}$$

The two free parameters of the Burkert model, the central density ρ_0 and the core radius r_0, turn out to be correlated: $\rho_0 \sim r_0^{-2/3}$.

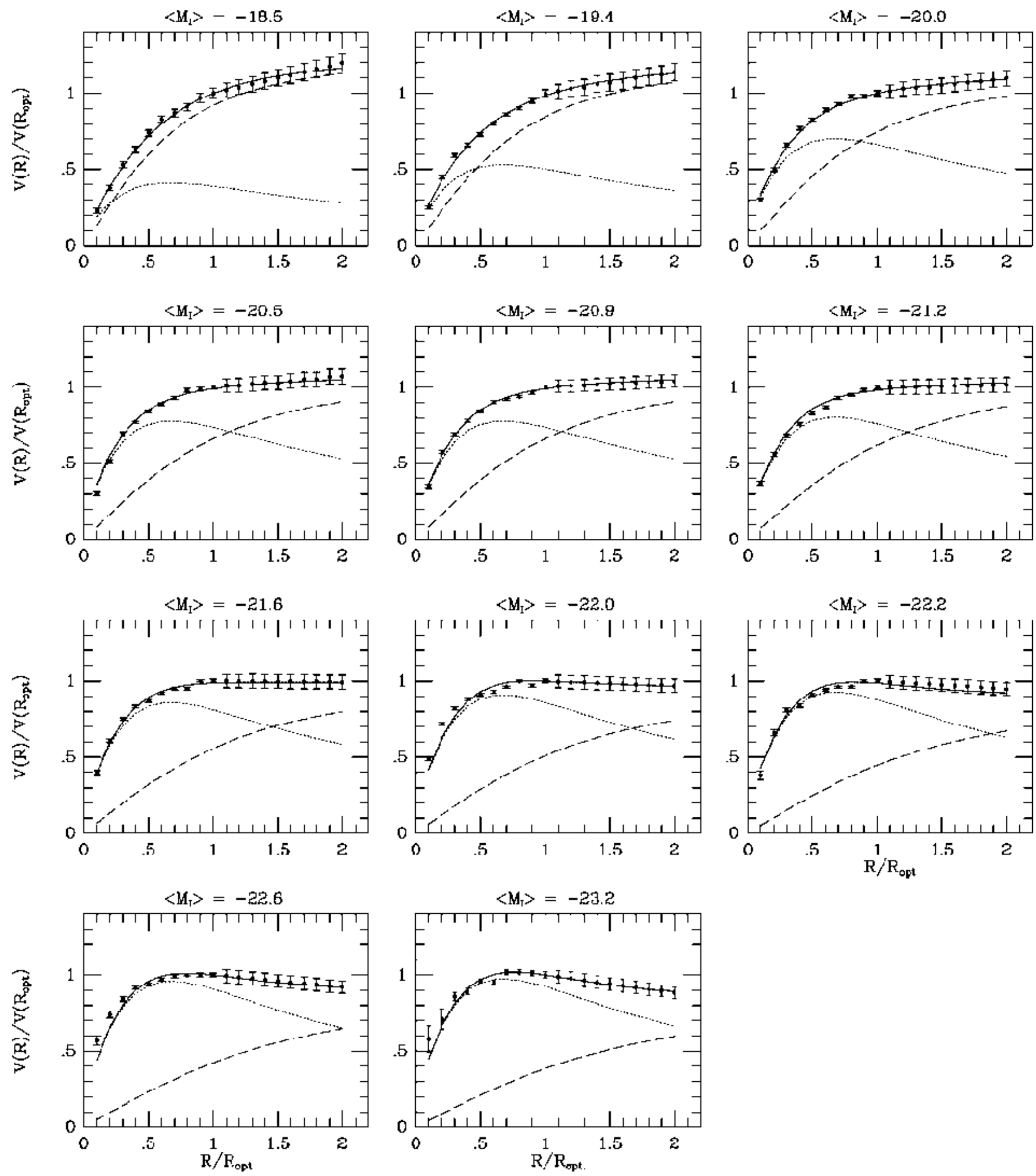


Figure 1.1: Average rotation curves and the universal rotation curve fits. The galaxies are binned in 11 luminosity bins, each bin containing 50-100 galaxies. The dotted line shows the contribution of the disc, the dashed is the dark halo component, and the solid is the sum of the two (in quadrature). Figure taken from [5].

1.3.2 Elliptical Galaxies

Elliptical galaxies also show evidence for existence of dark matter. Elliptical galaxies have anisotropic triaxial ellipsoidal shapes and contain ionized X-ray-emitting gas in the interstellar medium. The stars in elliptical galaxies do not follow any directionality patterns of the kind observed in spiral galaxies. They can be approximated as a gas in thermal equilibrium. The argument for existence of dark matter comes from comparing the temperature (or, rather, velocity dispersion) of stars with the temperature of the X-ray-emitting gas.

Assuming hydrostatic equilibrium in the gas implies that the gas pressure resists the gravitational collapse:

$$\frac{dp(r)}{dr} = -\frac{GM(r)\rho(r)}{r^2}, \tag{1.14}$$

where $M(r)$ is the total mass enclosed within distance r from the galactic center, and $\rho(r)$ is the density of gas at r. The ideal gas law provides the relationship between the pressure and temperature of the gas: $p = \rho k_B T/\mu m_p$, where m_p is the proton mass and μ is the average atomic mass of the ions in the gas. Hence,

$$\frac{d}{dr}\left(\frac{\rho(r) k_B T}{\mu m_p}\right) = -\frac{GM(r)\rho(r)}{r^2}. \tag{1.15}$$

If there is no dark matter, then $M(r)$ is determined by the stars alone (the contribution of the gas to the mass of the galaxy is negligible). The stars and the gas would move in the same gravitational well, and would, therefore, have the same velocity dispersion σ (following virial theorem):

$$\sigma^2 = \frac{k_B T}{\mu m_p}. \tag{1.16}$$

The goal, then, is to check if the stellar velocity dispersion σ and the gas temperature T follow Equation 1.16. The difficulty lies in the fact that the position-resolved

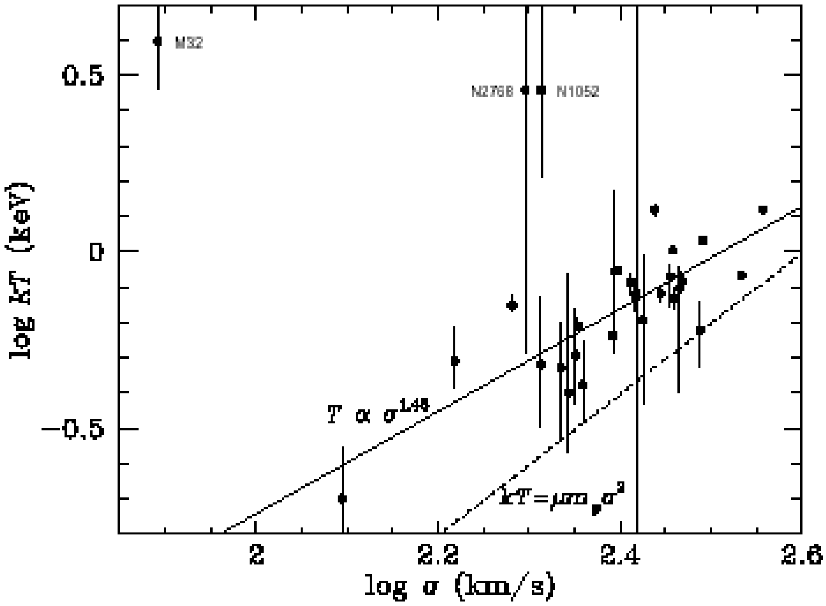


Figure 1.2: X-ray temperature vs stellar velocity dispersion for a set of elliptical galaxies. The solid line represents the best fit with the slope of 1.45, and the dashed line represents the no-Dark-Matter case with slope 2. Figure taken from [9].

X-ray luminosity measurements for small objects (such as galaxies) are non-trivial. However, the values of σ and T have been determined for a number of galaxies in the galactic centers. Davis and White [9] have performed such study on a sample of 30 elliptical galaxies and have observed $T \sim \sigma^{1.45}$, rather than $T \sim \sigma^2$. Their results are shown in Figure 1.2. Loewenstein and White [10] performed a more detailed study in terms of

$$\beta_{spec} = \frac{\mu m_p \langle\sigma\rangle^2}{k_B \langle T\rangle}. \tag{1.17}$$

They examined in detail different stellar and dark halo models and concluded that, without the presence of dark halo, β_{spec} is never less that 0.75. Since the typical observed value is $\beta_{spec} \approx 0.5$ (implying that the stars have $\sim 1/2$ the velocity dispersion they should have), they argue that dark matter is very common in elliptical galaxies. Intuitively, in their argument, the discrepancy is explained by the fact that the temperature of the hot, extended, gas is determined by the whole dark halo, while the stellar velocity dispersion at the center is characteristic of a small stellar-dominated component because the halo is sufficiently flat. This is particularly true for small galaxies - as the galaxy mass increases, the stellar and gas temperatures converge.

More recently, Loewenstein and Mushotzky [11] have used the observations by Chandra X-ray observatory and other X-ray and optical data to determine the interstellar medium temperature profile (and, therefore, the mass profile) of the elliptical galaxy NGC4336. The study was made possible by the unprecedented angular resolution of Chandra. They found that, within the experimental errors, both models with dark matter cusps and with dark matter cores are acceptable for this galaxy. Their results are shown in Figure 1.3. Finally, we should mention that there are arguments questioning reliability of the stellar kinematic evidence for elliptical galaxies. For example, Baes and Dejonghe [12] performed a Monte Carlo simulation of the absorption and scattering of light from the interstellar dust in elliptical galaxies. They observe that the attenuation by the interstellar dust has very similar kinematic signatures as a dark matter halo. However, they do not check for the IR reemission in their simulation, which could modify their result. Such degeneracy would clearly complicate the use of stellar kinematics to trace the mass distribution in elliptical galaxies.

1.3.3 Clusters of Galaxies

Clusters of galaxies are among the largest gravitationally bound objects in the Universe. They emerge from exceptionally high peaks of the primordial density perturbations, and they typically include hundreds of galaxies, along with the gas that fills the space

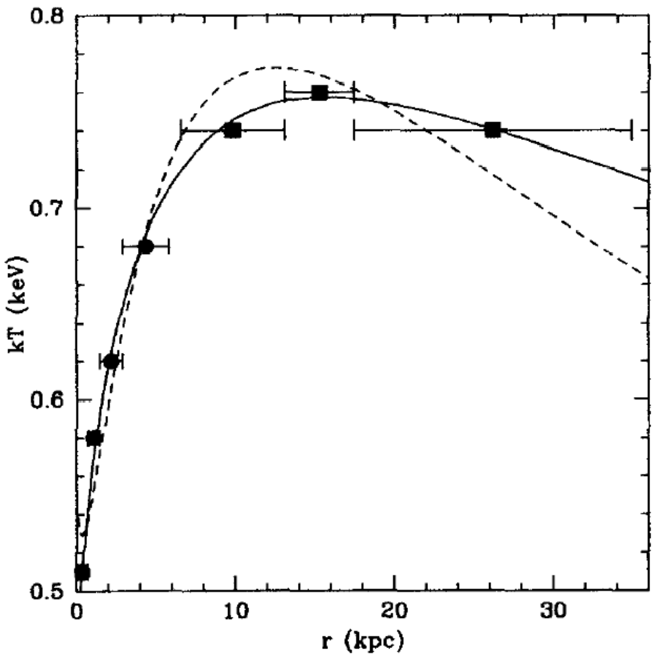


Figure 1.3: Chandra (filled circles) and XMM Newton (filled squares) temperatures and profiles for the overall best fit model with r^{-1} dark matter density cusp (solid) and the best fit model with dark matter density core (dashed). Figure taken from [11].

among them. Clusters gave the first hint of dark matter [1]. Today, there are several ways of estimating the amount of dark matter in the clusters.

Virial Approach

If the cluster is gravitationally relaxed, one can measure velocities of individual galaxies in the cluster and estimate the total cluster mass from the virial theorem:

$$\sum_i \frac{1}{2} m_i v_i^2 = G \sum_{i \neq j} \frac{m_i m_j}{r_{ij}}, \tag{1.18}$$

where i and j run over all galaxies in the cluster. The virial radius and mass are defined by:

$$\frac{1}{2} \langle m \rangle \langle v^2 \rangle = G \frac{M_{vir}}{R_{vir}} \langle m \rangle, \tag{1.19}$$

where $\langle v^2 \rangle = \sigma$ is the velocity dispersion of galaxies, $\langle m \rangle$ is the average galaxy mass (weighed by v_i^2), R_{vir} is the virial radius (inside which the assumption of virialization holds), and $M_{vir}(R_{vir})$ is the virial mass - the mass within one virial radius. Since both R_{vir} and M_{vir} are unknown, one can procede iteratively: assume some R_{vir}, determine the corresponding M_{vir}, check Equation 1.19, then assume a new R_{vir} etc. The procedure is successful if it converges to some value of R_{vir}.

This procedure was followed by Girardi *et al.* [13] on a sample of 170 clusters, each containing more than 30 galaxies with available redshifts. The clusters were collected from various sources, including the ESO Nearby Abell Cluster Survey (ENACS). The analysis was performed as outlined above, and it included a correction for the fact that galaxies far from the cluster center were not included in the observations. They also performed a detailed comparison with the X-ray estimates (see below) and found a good agreement. In a later paper, Girardi *et al.* [14] combined these results with optical luminosities of the galaxies and estimated the mean mass-to-light ratio of $\Upsilon \approx 250h$. The mass-to-light ratio is defined as the ratio of the mass to the luminosity of the system in solar units. One can define the critical mass-to-light ratio by dividing the critical density with the luminosity density of the Universe in the B-band (centered at 445 nm):

$$\begin{aligned} \Upsilon_c &= \frac{\rho_c}{j} = \frac{1.88 \times 10^{-29}\, h^2 \text{ g cm}^{-2}}{1.0 \times 10^8\, e^{\pm 0.26}\, h\, L_\odot \text{ Mpc}^{-3}} \\ &= 2800\, e^{\pm 0.3}\, h\, \frac{M_\odot}{L_\odot}. \end{aligned} \tag{1.20}$$

Hence, the measurement of $\Upsilon \approx 250h$ implies $\Omega_m \approx 0.09$. This is only a lower bound on Ω_m, as it includes only the mass within the virial radius.

More recently, Boviano and Girardi [15] studied about 100,000 galaxies from the

Two Degree Field Galaxy Redshift Survey (2dFGRS), from which they extracted 43 non-interacting clusters. They found that both density profiles with cusps and with cores can fit the measurements, although profiles with very large cores are ruled out.

X-ray Emission

Similarly to elliptical galaxies, one can estimate the cluster mass using the temperature of the X-rays emitted by the intergalactic gas, and assuming the hydrostatic equilibrium. The clusters are sufficiently large that one can measure the X-ray luminosity profile out to the virial radius (this is much more difficult to do in galaxies). Many studies of the cluster X-rays have been performed (see, for example, [16], [17], [13]). In particular, a large fraction of the clusters analyzed by Girardi *et al.* [13] in the virial approach have also been analyzed by the X-ray method. Girardi *et al.* find a good agreement between the two estimates.

Gravitational Lensing

The methods described above rely on understanding the dynamics in the clusters. In particular, assumptions of virialization or hydrostatic equilibrium are made, which are never exactly true. The gravitational lensing allows a direct estimate of the cluster mass, hence avoiding such assuptions. The idea is that a cluster gravitationally lenses the light emitted by a background galaxy.

The lensing effect varies significantly in magnitude, depending on the (cluster) lens mass and on the relative positions of the lens and the background galaxy. In some cases, the lensing is so strong that multiple images of the background galaxy appear, often distorted

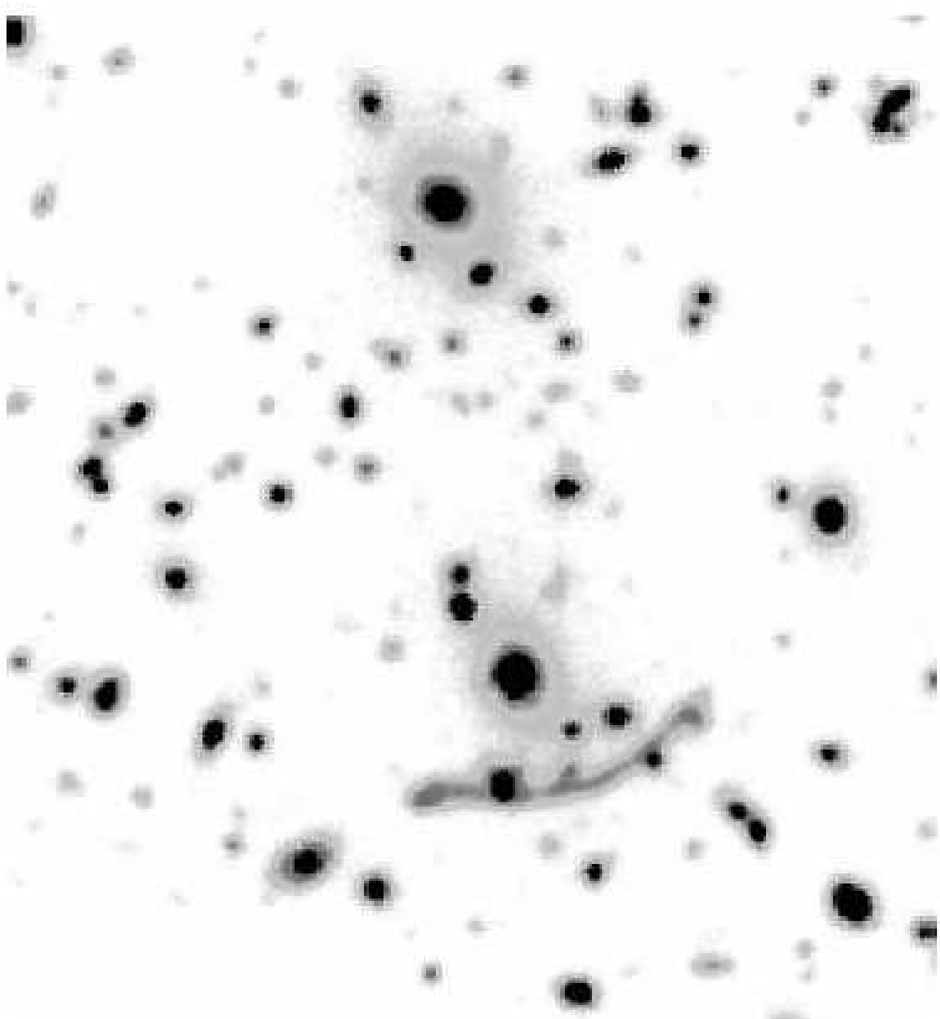


Figure 1.4: The galaxy cluster Abell 370, where the first gravitationally lensed arc was discovered. Figure taken from [18].

into arcs or rings. The first example of such gravitationally lensed arc was detected in Abell 370, and is shown in Figure 1.4. Typically, the strong lensing observations are difficult to analyze due to high non-linearity of the effect, so each detection has to be analyzed separately. The weak lensing is much simpler, although observationally more involving. In this case, the effect is observed as a shape deformation or increased brightness of a galaxy. Since ellipticity of each galaxy is not known a priori, the analysis is usually done statistically.

Alternatively, one can study the number density of galaxies above certain threshold - the weak lensing effect would increase such density.

Bartelmann and Schneider [18] review a number of clusters for which the weak lensing effect was observed. The cluster masses estimated with this method typically agree well with estimates based on the virialization or hydrostatic equilibrium assumptions.

As Weinberg and Kamionkowski [19] argue, the future weak-lensing surveys are expected to have the sensitivity to identify the galaxy clusters in the field, including the "dark lenses" - clusters which have no signal in the optical or in the X-ray regime as they have not started the gravitational collapse yet. Combined with the redshifts of the detected clusters, such surveys would directly probe the evolution of the total mass and, hence, should provide new estimates of σ_8 and Ω_m (see Section 1.3.6 for more detail on this type of approaches).

Baryon Fraction and Sunyaev-Zel'dovich Effect

If the fraction of the total cluster mass carried by the baryons can be estimated, then the matter density of the Universe can be determined using an estimate of the baryon density in the Universe, Ω_b. The baryon density, in turn, can be obtained from, for example, the Big Bang Nucleosynthesis (discussed below) or from the CMB anisotropy measurement (also discussed below).

The baryon density in the cluster can be estimated using the fact that most of the baryons $\sim$ 90%) are in the form of a gas in the intracluster medium. For a given temperature, the X-ray luminosity profile gives directly the electron density in this gas, and hence the proton density. Alternatively, one can use the Sunyaev-Zel'dovich effect (SZE),

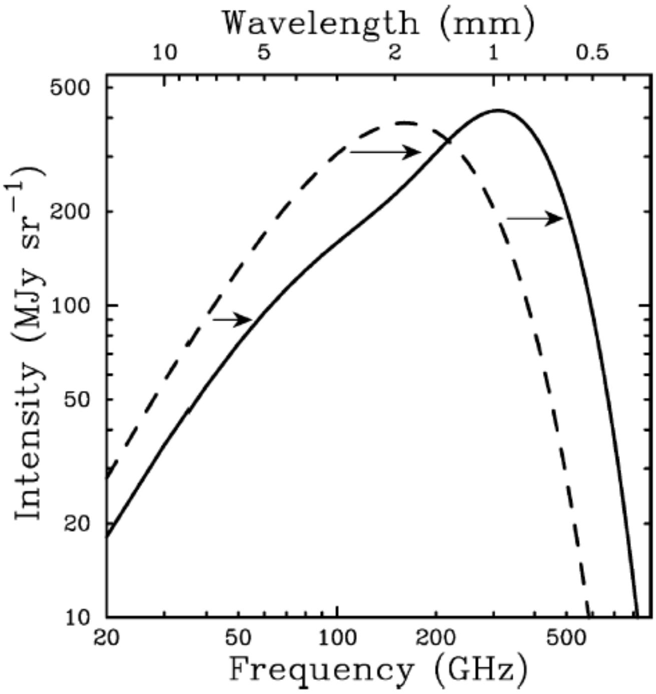


Figure 1.5: SZE: undistorted CMB spectrum (dashed) and distorted by the SZE (solid), for a cluster 1000 times more massive than a typical cluster. Figure taken from [20].

as described in detail in [20]. The SZE is the inverse Compton scattering of the cosmic microwave background photons off of electrons in the gas. As they pass through the center of the cluster, the CMB photons have about 1% probability of interacting with the hot electrons in the gas. If they do, they are preferentially boosted in energy, causing a small distortion in the spectrum ($\sim$ 1 mK or less). This is the thermal SZE, and it is illustrated in Figure 1.5. The effect has a unique signature that makes its detection somewhat easier: it appears as a decrease in the intensity of the CMB below $\sim$ 218 GHz, and an increase above this frequency. If the cluster is moving with respect to the CMB frame, there will be an additional distortion of the spectrum due to the Doppler effect. This is the kinetic SZE. It is typically much smaller than the thermal SZE [21].

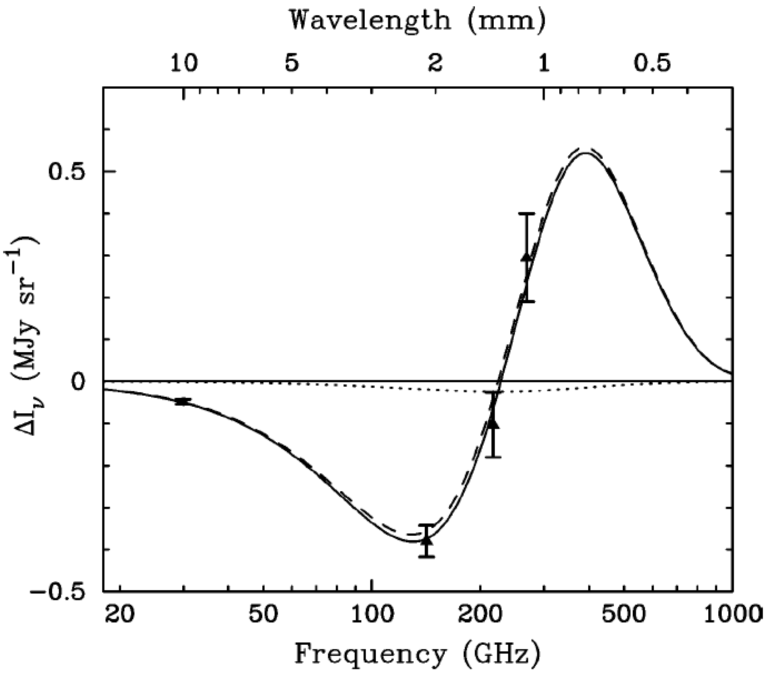


Figure 1.6: Measurement of the SZE of Abell 2163. The measurements were combined from different sources. The solid line is the best fit, and the dashed and dotted lines show the best fit thermal and kinetic components. Figure taken from [20].

The SZE has indeed been observed, both in single-dish observations (Owens Valley Radio Observatory (OVRO), Sunyaev-Zel'dovich Infrared Experiment (SuZIE)) and in interferometric observations (Ryle telescopes, OVRO+BIMA telescopes etc) - see [20] for a review of these observations. Figure 1.6 shows the observed SZE effect for the cluster Abell 2163.

As mentioned above, about 90% of the baryons are measured to be in the form of a gas in the intracluster medium. If the temperature of the electrons in the gas is known (and it can be estimated from the X-ray measurements), the SZE can be used to estimate the mass of the gas (and, hence, of baryons in the cluster). The total gravitating mass of the cluster has to be estimated by one of the other methods described above. One can then estimate the fraction of the cluster mass stored in baryons, and use an Ω_b estimate (from BBN or CMB) to estimate Ω_m. Several such analyses have been completed - as an example we cite Grego *et al.* [22]. They analyzed 18 distant clusters observed interferometrically with OVRO and BIMA system. They modelled the gas density directly from the SZE data and used the electron temperatures from the X-ray measurements. They estimated the gas fraction to be $f_g h = 0.081^{+0.009}_{-0.011}$ for $\Omega_m = 0.3, \Omega_\Lambda = 0.7$. After including estimates of the amount of baryons in galaxies and of baryons lost during the cluster formation, they obtained the baryon fraction in clusters to be $f_b = f_g(1 + 0.2h^{3/2})/0.9$. Combined with the BBN estimate of Ω_b (see Section 1.3.4), this implies the matter density $\Omega_m \sim 0.25$, as shown in Figure 1.7.

Finally, we should mention one remarkable property of the SZE - while the CMB spectrum itself suffers from cosmological dimming with redshift, the ratio of SZE to CMB does not. Hence, SZE measurement is redshift independent, which can be exploited to make a direct measurement of the evolution of the number density of clusters. The upcoming SZE surveys are intended to do exactly that.

1.3.4 Big-Bang Nucleosynthesis

The Big Bang Nucleosynthesis (BBN) model predicts the abundances of the light elements produced in the early stages of the Universe (first 2-3 minutes), in the framework of the hot Big Bang. The model yields strong constraints on the density of the baryonic matter in the Universe, Ω_b. Combined with estimates of the total amount of matter in the Universe, Ω_m, the BBN provides strong evidence for existence of the non-baryonic dark matter.

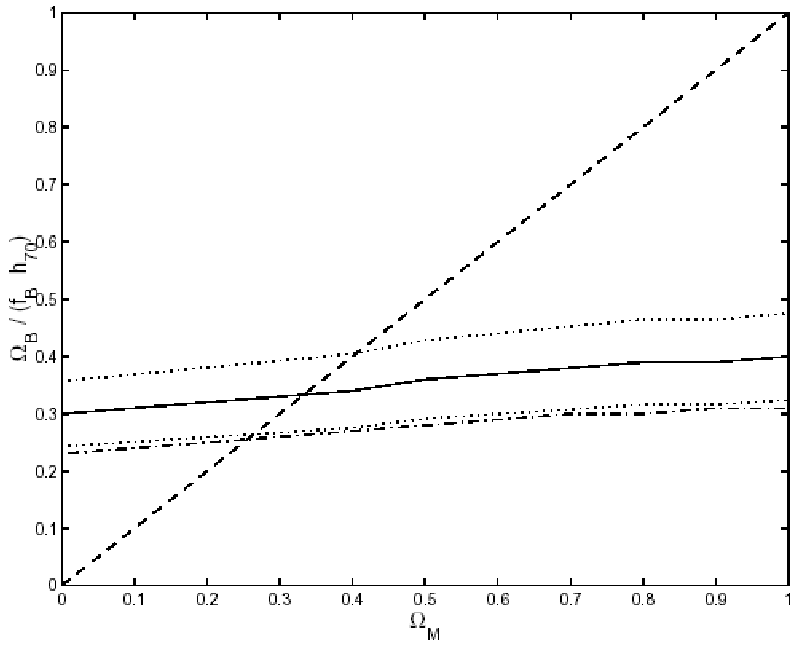


Figure 1.7: Upper limit on $\Omega_m < \Omega_b/f_b h$ is shown as a solid line, and the 68% confidence region as dotted lines, as a function of cosmology ($\Omega_\Lambda = 1 - \Omega_m$). The intercept of the upper dotted line and the dashed line gives the 68% upper limit on Ω_M. The dot-dashed line shows the matter density when the estimates of baryons in the galaxies and those lost during the cluster formation are included. The intercept of the dot-dashed line and the dashed line gives the best estimate of $\Omega_m \approx 0.25$, assuming flat Universe with $h = 0.7$. Figure taken from [22].

Discussions of the BBN can be found in any standard textbook, such as [3]. At early times (at temperatures > 1 MeV, when the Universe was less than 1 second old), the weak interactions such as

$$p + e^- \leftrightarrow n + \nu_e \tag{1.21}$$

keep the neutrons and protons in equilibrium. As the Universe expands and cools below ~ 1 MeV, the weak interactions interconverting neutrons and protons freeze out, and the n/p ratio is given by its equilibrium value before freeze-out, $\sim 1/6$.

At this time, the reactions producing light nuclei are still in equilibrium with the reactions of photodissociation of the same nuclei. For example, for deuterium:

$$p + n \leftrightarrow D + \gamma \tag{1.22}$$

As the Universe cools and expands further, the temperature of photons becomes low enough that the dissociation reactions halt. The protons and neutrons then proceed to form deuterium and the light nuclei, until all neutrons are used up. About 20% of the neutrons decay prior to being incorporated in the nuclei. Roughly 100 seconds after the Big Bang (at the temperature of about 0.03 MeV), the abundances of the light nuclei D, ^{3}H, ^{3}He, ^{4}He, ^{7}Li and ^{7}Be are set. The heavier nuclei are not formed because of the large Coulomb barrier to the necessary reactions.

Two cosmological parameters are crucial in determining the final abundances of the light elements [23]. First, the expansion rate (relative to the rate of the weak interaction) sets the n/p ratio. Second, the baryon density ρ_b affects the relative abundances of the elements: if ρ_b is larger, the nucleosynthesis starts earlier and more nucleons end up in the stable element ^{4}He (the abundances of D and ^{3}He are accordingly smaller). Usually, the baryon-to-photon ratio $\eta = n_b/n_\gamma$ is used instead of ρ_b, and the photon density n_γ is obtained from the CMB measurement. Figure 1.8 shows the evolution of the light-element abundances as the Universe cooled, as predicted by the BBN model.

The goal, therefore, is to measure the abundancies of light nuclei in the Universe, and from them infer the density of baryons ρ_b (or, equivalently, Ω_b). Typically, a ratio of abundancies of two elements is measured (one of which is usually hydrogen, whose density is usually the easiest to measure). The major difficulty, however, lies in estimating the departures from the primordial abundances.

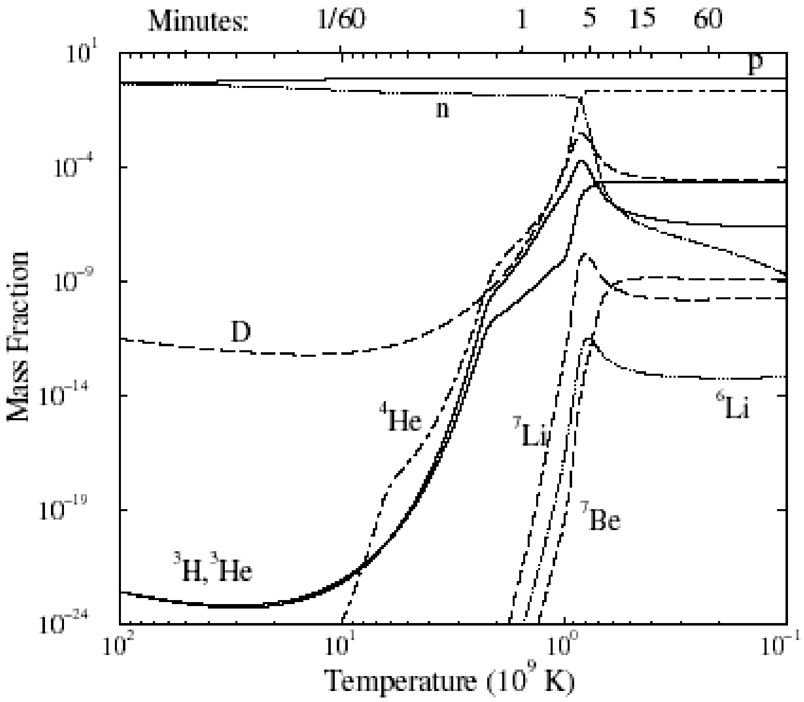


Figure 1.8: Evolution of the light-nuclei abundancies in the early Universe. Figure taken from [23].

The most reliable primordial abundance measurement is believed to be the one of deuterium D - see Figure 1.9. The measurement is made using low metalicity absorption line systems in the spectra of high-redshift quasars. The gas in these systems is in the outer regions of galaxies or in the intergalactic medium, and it is not related to the quasars. The low metalicity of the absorption system is required as it implies that no significant amount of D was destroyed in stars. A high redshift is required as the Universe was then too young for the low-mass stars to eject large amounts of gas. There are no other known processes that can produce or destroy significant amounts of D. One complication, however, is that the obsorption by D is often contaminated (if not obscured) by the absorption by H. Hence, careful selection of the absorption system is needed. Four such systems have been found

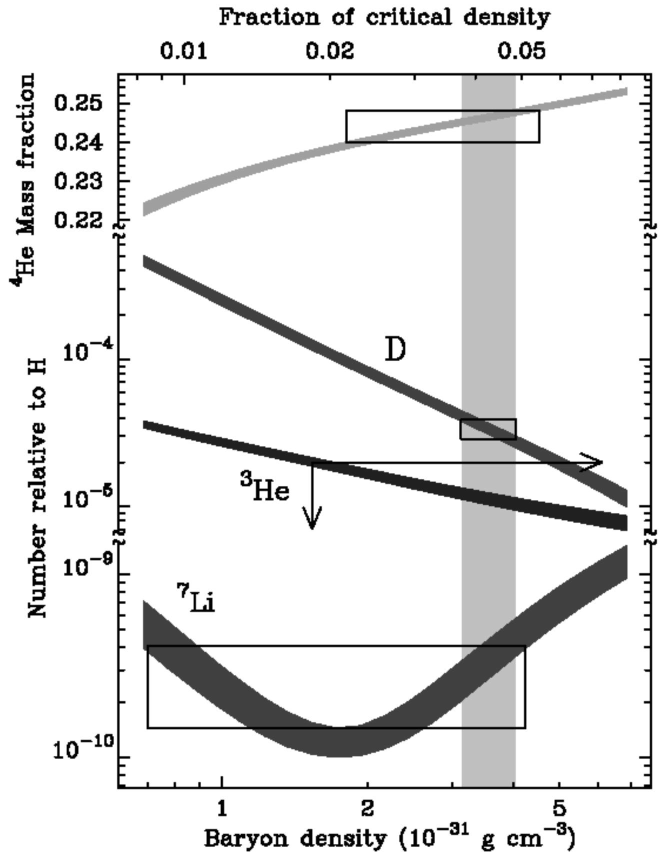


Figure 1.9: Expected abundances (calculated in the standard BBN) for the light elements are shown as gray bands. The rectangles show the 95% confidence intervals for the measurements cited in the text. The deuterium measurement is the most constraining on the only free parameter of the model, ρ_b. Figure taken from [23].

with low D/H ratio (see [23, 24, 25, 26]), giving the estimate of $\Omega_b h^2 = 0.019 \pm 0.0012$ (at 68% C.L.). We note that D was also found in the interstellar medium [23], but its abundance there is reduced due to the destruction in stars (and hence, smaller than the primordial abundance by a factor of ~ 2).

The abundance of ^{4}He has also been measured. In this case, one relies on the emission spectra of the ionized gas surrounding hot young stars, such as in blue compact galaxies. Such stars are chosen as they provide the least stellar "contamination". The observed ^{4}He mass fraction is plotted against the abundance of O (or N) in the gas, and then extrapolated to zero O (or N). Such analyses were performed [27, 28], yielding the mass fraction of baryons which is in ^{4}He, $Y_p \approx 0.244$. However, the dependence of the ^{4}He abundance on ρ_b is weaker than the dependence of D (see Figure 1.9) - hence, the deuterium abundance measurement is more constraining on ρ_b and Ω_b.

The measurement of the abundance of ^{7}Li is more difficult. The measurement is made in the old halo stars, formed from gas with low iron abundance. These "Spite-plateau" stars show approximately constant $^7\mathrm{Li/H} \approx 1.6 \times 10^{-10}$ [23], relatively independent of the abundances of the heavy elements. Hence, this value should be close to primordial. Some doubts still remain - for example, detections of ^{6}Li indicate that some of the ^{7}Li was created prior to the formation of these stars. Also, the amount of depletion is difficult to estimate. Nevertheless, the measurement is expected to be accurate to within a factor of ~ 2.

Finally, the abundance of ^{3}He is the most difficult to estimate, because stars are expected to both make and destroy it [23]. The abundance of ^{3}He in the Galactic H II regions was measured [29], $^3\mathrm{He/H} \approx 1.6 \pm 0.5 \times 10^{-5}$, but it is not clear how this value relates to the primordial one.

1.3.5 SNIa Observations

The supernovae type Ia (SNe Ia) are an example of standard candles, as their intrinsic luminosity varies little from case to case. Measurement of the redshift of a SN Ia, along with the observed flux, can be used to constrain the cosmological parameters. The argument is as follows.

For an object of luminosity L and redshift z, the flux observed by a detector is given by:

$$F = \frac{L}{4\pi d_H^2 (1+z)^2}, \tag{1.23}$$

where d_H is the proper distance between the object and the detector at the present time. If there is no expansion, $z = 0$ and the equation reduces to the familiar, intuitive, form. One factor of $(1+z)^{-1}$ in the Equation 1.23 comes from the decrease in energy due to the red-shifting of each individual photon that travels the distance d_H, and the other comes from the fact that the photons are more spread out in time when they arrive at the detector than they were when they were emitted by the object (this spreading is also due to the expansion). One often defines the luminosity distance as $d_L = d_H(1+z)$, in order to express the flux in a simpler form $F = L/4\pi d_L^2$. The proper distance itself can be calculated from the metric and the Friedmann equation. First, from Equation 1.2,

$$\begin{aligned} d_H &= \int_0^{r_H} \sqrt{g_{rr}} dr = R_0 \int_{t_1}^{t_0} \frac{dt'}{R(t')} \\ &= R_0 \int_{R(t_1)}^{R_0} \frac{dR(t')}{\dot{R}(t')R(t')}, \end{aligned} \tag{1.24}$$

where t_1 and t_0 are the emission and detection times, g_{rr} is the rr component of the metric g, and R_0 is the scaling factor at the present time t_0. From Equation 1.7,

$$\dot{R}^2(t) = H_0^2 R^2(t) \left[\Omega_m \left(\frac{R_0}{R(t)} \right)^3 + \Omega_k \left(\frac{R_0}{R(t)} \right)^2 + \Omega_\Lambda \right]. \tag{1.25}$$

Inserting this in the Equation 1.24, setting $\Omega_k = 1 - \Omega_m - \Omega_\Lambda$, and changing variables gives:

$$d_H = H_0^{-1} \int_0^z [y(y+1)^2\Omega_m - y(y+2)\Omega_\Lambda + (y+1)^2]^{-1/2} dy \tag{1.26}$$

Hence, measuring the flux of the given standard candle and its redshift directly probes the allowed values of H_0, Ω_m, and Ω_Λ (although, clearly, these parameters cannot be estimated independently). Moreover, by considering the ratios of d_H (or, equivalently, of d_L) for different objects, the dependence on H_0 can be removed.

Few more words about supernovae. The consistency of the SN Ia luminosity (and its appearance as a standard candle) is explained by its production mechanism. In general, supernovae are large explosions that occur when stars run out of the fusion fuel. In these situations, the pressure is no longer able to oppose the gravitational collapse, which in turn causes an explosion releasing large amounts of energy and ejecting large amount of material from the star's outer layers. The fusion first stops in the core of the star, but it may still continue in the outer layers. As the outer layers run out of the fusion fuel, the size of the fusion-less core increases.

If the star is massive enough, the core mass will eventually exceed the Chandrasekhar limit of 1.4 $M_\odot$, at which point the gravitational collapse of the core occurs. This is believed to be the production mechanism for the type II supernovae. Depending on the mass of the star, a neutron star or a black hole is produced. There is much variation in the final state of the star, just prior to the explosion: the total mass of the star can vary significantly, the outer-most layers may still be undergoing fusion of different elements, the core may contain different ammounts of heavy elements (anything between C/O and Fe) etc. This kind of variability certainly prevents the "standard candle" behavior.

However, the stars with mass $< 1.4\ M_\odot$ cannot exceed the Chandrasekhar limit. The core of these stars contracts, but it does not gravitationally collapse - these are white dwarfs. The outer layers of these stars continue burning H and He (or are ejected as planetary nebulae), until they run out. If a white dwarf is in a binary system with a normal star, it can continuously accrete material from its companion star. Eventually, its mass would exceed the Chandrasekhar limit, and it would undergo gravitational collapse. Note, however, that by this time, its H and He have run out, the fusion in the outer layers has mostly or completely stopped, so the explosion takes place in relatively similar conditions every time. This mechanism of gravitational collapse of white dwarfs is believed to explain the type Ia supernovae - in particular, the uniformity of the luminosity and the lack of H and He lines. We note that there are also type Ib supernovae - these are essentially Type II supernovae, whose stellar wind has blown off the hydrogen by the time of the explosion. Hence, type Ib supernovae also do not have the H lines in their spectra.

Two experiments have systematically searched and observed SNe Ia at various redshifts - the Supernova Cosmology Project (SCP) [30] and the High-z Supernova Search Team (HZT) [31]. The SCP has observed 42 supernovae at redshifts between 0.18 and 0.83. They corrected the observed light curves for the time dilation ("stretch" factor) by fitting to the template light curves from a low-redshift calibration set. This procedure further improved the uniformity of the supernovae. Finally, they fitted Ω_m and Ω_Λ to the flux-redshift relation (they actually used the apparent magnitudes, rather than flux, to avoid the H_0 dependence) and obtained the estimate of $\Omega_m = 0.28^{+0.10}_{-0.09}(\text{stat})^{+0.05}_{-0.04}(\text{syst})$ (for a flat Universe) [30].

The HZT observed 16 high-redshift SNe Ia. They performed similar analysis, but they applied a multi-wavelength correction method to remove the time-dilation effect. Their result is consistent with the one from SCP: $\Omega_m = 0.28 \pm 0.10$ (stat) (for a flat Universe) [31].

More recently, the HZT has observed 8 more SNe Ia with redshifts 0.3-1.2 [32]. This extended the redshift-range of the observed SNe Ia to $z \approx 1$, where the various systematic effects are expected to be of opposite sign compared to the cosmological effects. They combine these observations with the previously observed SNe Ia from both SCP and HZT (and others) - 230 SNe Ia overall. These observations are shown in Figure 1.10. The data was fitted to the various cosmological models - the $\Omega_m - \Omega_\Lambda$ contours are shown in Figure 1.11. They obtain $\Omega_\Lambda - \Omega_m = 0.35 \pm 0.14$ [32]. Assuming a flat Universe and the equation of state $w = -1$, this gives $\Omega_m = 0.28 \pm 0.05$. As shown in Figure 1.11, this result is very consistent with the large-scale structure estimate. It is also consistent with the earlier estimates [30] and [31].

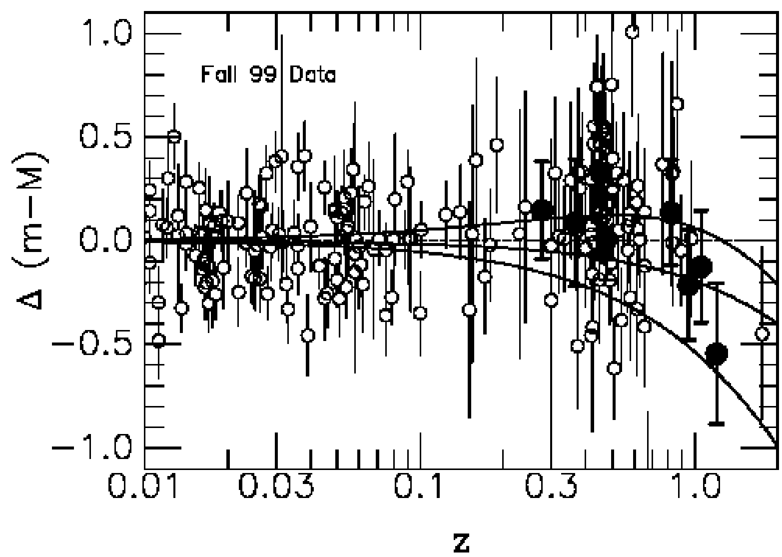


Figure 1.10: Residual Hubble diagram relative to an empty Universe, showing the residual apparent magnitude vs redshift. The eight latest high-z SNe Ia are shown as filled circles. The lines (top-to-bottom) denote $(\Omega_m, \Omega_\Lambda) = (0.3, 0.7)$, $(0.3, 0.0)$, and $(1.0, 0.0)$ models of the Universe. Note the inconsistency of the (1.0,0.0) model. Figure taken from [32].

Finally, we mention some of the remaining issues with observations of SNe Ia.

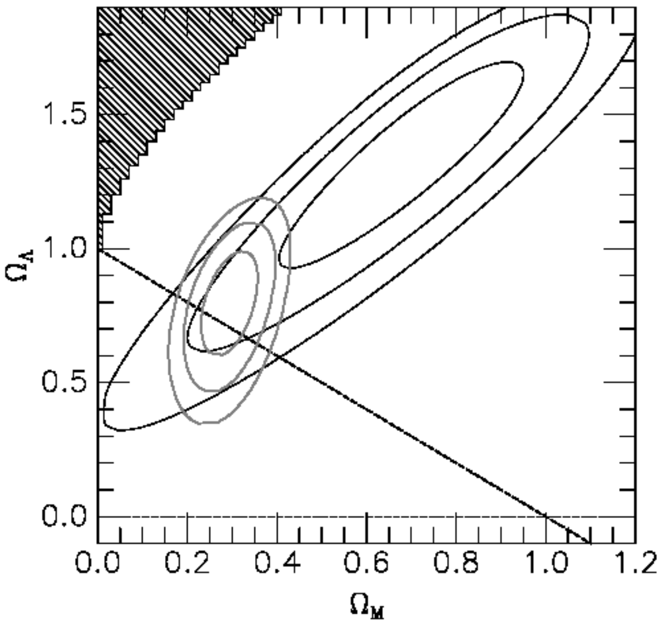


Figure 1.11: Probability contours for $\Omega_m - \Omega_\Lambda$ at 1, 2, and 3 σ are shown for the equation of state $w = -1$ (black). The gray contours are the corresponding results if a large-scale structure prior of $\Omega_m h = 0.20 \pm 0.03$ [33] is used. Figure taken from [32].

These observations are essentially measurements of the dimming of the SNe Ia, which could, in principle, be caused by dust. However, the dust would typically decrease the flux differently at different wavelengths, typically causing "reddening" of the objects. By comparing nearby and distant supernova, one can place limits on the effect of the reddening dust - Perlmutter *et al.* in [30] estimate this effect to cause < 0.03 systematic error in the measurement of Ω_m.

It is possible that the dust is "gray" - i.e. affects all wavelengths equally. Since the lines-of-sight of the observed supernovae are very different, one would expect the gray dust to introduce additional dispersion in the apparent magnitudes of the supernovae. Such excess is not observed. Furthermore, failed searches for X-rays emitted by the smoothly distributed intergalactic gray dust indicate that the amount of such dust would not be sufficient to account for the observed dimming.

Gravitational lensing is another possible "contaminant". It would cause occasional changes of supernovae brightness. Perlmutter *et al.* estimate in [30] that only 1% of the supernovae in their analysis would be rejected as outliers due to the gravitational lensing effects.

Another issue is the possible evolution of the supernovae (i.e. the high and low redshift supernovae might be systematically different). However, the observed high-z and low-z light curves are very consistent with each other. Hence, the impact of such effect is not expected to be large.

Finally, we mention that the upcoming Nearby Supernova Factory and the SNAP satellite are expected to follow light-curves of hundreds or thousands of SNe, and are expected to significantly reduce the uncertainties in the measurements of Ω_m and Ω_Λ.

1.3.6 Large-Scale Structure Formation

Formation and growth of the large-scale structure provide another way of studying the matter content in the Universe. The basic problem is to understand the evolution from the small perturbations in the very homogenous early Universe, to galaxies, clusters etc that we observe in the clumpy Universe today.

To handle this problem quantitatively, one defines the fluctuation in the matter density as (we follow the derivation in [3]):

$$\delta(\vec{x}) = \frac{\rho(\vec{x}) - \bar{\rho}}{\bar{\rho}}, \tag{1.27}$$

where $\bar{\rho}$ is the average matter density and $\rho(\vec{x})$ is the matter density at position $\vec{x}$. One can then calculate the auto-correlation function $\xi(\delta x) = \langle \delta(\vec{x})\delta(\vec{x}+\delta\vec{x})\rangle$, where the average is taken over all $\vec{x}$. The isotropy and homogeneity of the Universe imply that ξ is only a function of δx (and not of $\vec{x}$ or $\delta\vec{x}$). The Fourier transform of the auto-correlation function gives the power spectrum $P(\vec{k})$:

$$P(\vec{k}) \;=\; |\delta(\vec{k})|^2 \;=\; \frac{4\pi}{k}\int d^3x\,\xi(x)\sin kx. \tag{1.28}$$

While it is not possible to directly measure the density auto-correlation function, it can be approximated by measuring the galaxy-galaxy number correlation function (or even galaxy-velocity or velocity-velocity correlation functions), assuming that the number density of galaxies follows the matter density in the Universe.

The next step is to determine the time evolution of the primordial density fluctuations (and of the corresponding power spectrum). Classically, the non-expanding Universe can be modeled as a perfect fluid, whose evolution is described by the Eulerian equations:

$$\begin{aligned} \tfrac{\partial \rho}{\partial t} + \vec{\nabla}\cdot(\rho\vec{v}) &= 0 \\ \tfrac{\partial \vec{v}}{\partial t} + (\vec{v}\cdot\nabla)\vec{v} + \tfrac{1}{\rho}\vec{\nabla}p + \vec{\nabla}\phi &= 0 \\ \nabla^2\phi &= 4\pi G\rho, \end{aligned} \tag{1.29}$$

where ρ is the matter density, p is its pressure, $\vec{v}$ is the velocity field of the fluid, and ϕ is the gravitational potential. Linearizing gives a single second-order differential equation:

$$\frac{\partial\delta\rho}{\partial t^2} - v_s^2\nabla^2\delta\rho = 4\pi G\rho\delta\rho, \tag{1.30}$$

where $v_s = \sqrt{\partial p/\partial\rho}$ is the adiabatic speed of sound ($\sim c/3$ in the radiation-dominated era), and $\delta\rho$ is the density fluctuation around ρ. The solution is of the form:

$$\delta\rho = A\exp(-i\vec{k}\cdot\vec{x} + i\omega t)\rho, \tag{1.31}$$

with $\vec{k}$ and ω satisfying the dispersion relation

$$\omega^2 = v_s^2 k^2 - 4\pi G\rho. \tag{1.32}$$

Clearly, if $k < k_J = (4\pi G\rho/v_s^2)^{1/2}$ then we observe exponential growth of the initial perturbation (k_J is the Jeans wavenumber). If, however, $k > k_J$, the solution is oscillatory. This solution applies in the radiation-dominated era, when the matter pressure is important. However, in the matter-dominated era the pressure can be ignored, and the solution of the Equation 1.30 becomes purely exponential ($k_J \to \infty$).

Of course, this solution is wrong as it is not relativistic and it does not account for the expansion of the Universe. The proper calculation is fairly complex, so we only outline it here (for details, see [3] or [34]). One has to start from a perturbation in the metric:

$$g_{\mu\nu} = g^0_{\mu\nu} + h_{\mu\nu}. \tag{1.33}$$

One then chooses a gauge, such as the synchronous gauge $h_{00} = h_{i0} = 0$. After this, one calculates the perturbations in the Christoffel coefficients $\Gamma^\alpha_{\mu\nu}$, in the Ricci tensor $R_{\mu\nu}$, in the scalar curvature $\mathcal{R}$, and in the stress-energy tensor $T_{\mu\nu}$. Inserting these in the Einstein equation 1.1 yields differential equations analogous to Equation 1.29. The solutions are power-laws for the growth of perturbations:

$$\delta\rho \sim t^{2/3} \sim R, \text{ Matter Dominated Case, } v_s = 0, \tag{1.34}$$

$$\delta\rho \sim t \sim R^2, \text{ Radiation Dominated Case, } v_s = c/3. \tag{1.35}$$

Furthermore, the Jeans wavenumber is replaced by k_h, the wavenumber corresponding to the horizon size. The solutions presented above, therefore, apply to the super-horizon-size modes $k < k_h$. The sub-horizon-size modes do not grow. Hence, proper treatment of the density perturbations leads to power-law growth of structure, rather than the exponential growth obtained in the classical treatment. This is due to the expansion of the Universe, which effectively moderates the growth.

Finally, let us note that the primordial spectrum is expected to be determined by the quantum fluctuations of the scalar field responsible for inflation. It is commonly assumed that the power spectrum of density fluctuations as they enter the horizon is scale invariant, $k^3 P(k) = const$. Since $P = |\delta(\vec{k})|^2 \sim R^4$ for the radiation-dominated case, it follows that scale-invariance implies the primordial power spectrum $P(k) \sim k$.

To summarize, the primordial power spectrum is expected to be $P(k) \sim k$. Initially, the Universe is radiation-dominated, so density fluctuations on the scales $k > k_h$ oscillate (i.e. do not grow), while fluctuations on the scale $k < k_h$ grow as R^2. However, the horizon size increases with time (and, thus, k_h decreases). Hence, modes which were initially $< k_h$ are reached by k_h and stop growing. The power spectrum, therefore, decreases with k at large k (roughly as k^{-3}), with the exact shape determined by the interplay of the growth law and the dependence of k_h on R. As the Universe becomes matter dominated, the structure at all scales starts growing $\sim R$. The shape of the spectrum is, therefore, frozen at the time of matter-radiation equality - this sets a characteristic scale in the power spectrum, k_{eq}.

The power spectrum can be used to constrain the matter content of the Universe. In particular, different types of matter have different impacts on the power spectrum. For example, hot (i.e. relativistic) dark matter would undergo "free-streaming" - it would freely move from high-density to low-density regions, hence dampening the fluctuations and the growth of the structure. On the other hand, baryons alone could not start the structure-growth described above until they decoupled from the photons at the surface of last scattering. Since they are tightly coupled to the photons, they could not freely fall into their own gravitational wells prior to decoupling. This would give them a growth factor of $\sim 10^3\Omega_b$, not enough to explain the structure observed today [3]. Adding some amount of non-baryonic cold dark matter (i.e. dark matter that becomes non-relativistic early enough, so that its free-streeming is not important) allows the structure to start growing earlier (before the time of baryon-photon decoupling), providing gravitational wells for baryons to fall into when they decouple. This could, indeed, explain the structure observed today.

We now shift to the observations. This was a very active field of study over the years, and we only mention some of the most recent results.

The 2-degree Field Galaxy Redshift Survey (2dFGRS) has produced a 3D map of 221,000 galaxies, using a 2-degree field fiber spectrograph on the Anglo-American Telescope (AAT). Compared to previous surveys, 2dFGRS represents an order of magnitude improvement in the survey volume and in the sample size. The power spectrum measured by 2dFGRS is shown in Figure 1.12. The 2dFGRS team performed a maximum likelihood fit to the measured galaxy-galaxy power spectrum (in the linear regime, $0.02 < k < 0.15\ h\ \mathrm{Mpc}^{-1}$), over a $40 \times 40 \times 40$ grid in $(\Omega_m h, \Omega_b/\Omega_m, h)$. Their preliminary results [35] for the whole sample of 221,000 galaxies give $\Omega_m = 0.26 \pm 0.05, \Omega_b = 0.044 \pm 0.016$, and are shown in Figure 1.13 along with estimates from other methods. Their previous results [33], based on a subset of 160,000 galaxies, gave $\Omega_m = 0.29 \pm 0.07, \Omega_b = 0.044 \pm 0.021$ (extracted from their estimates of $\Omega_m h = 0.20 \pm 0.03$ and $\Omega_b/\Omega_m = 0.15 \pm 0.07$, and assuming $h = 0.7 \pm 0.07$).

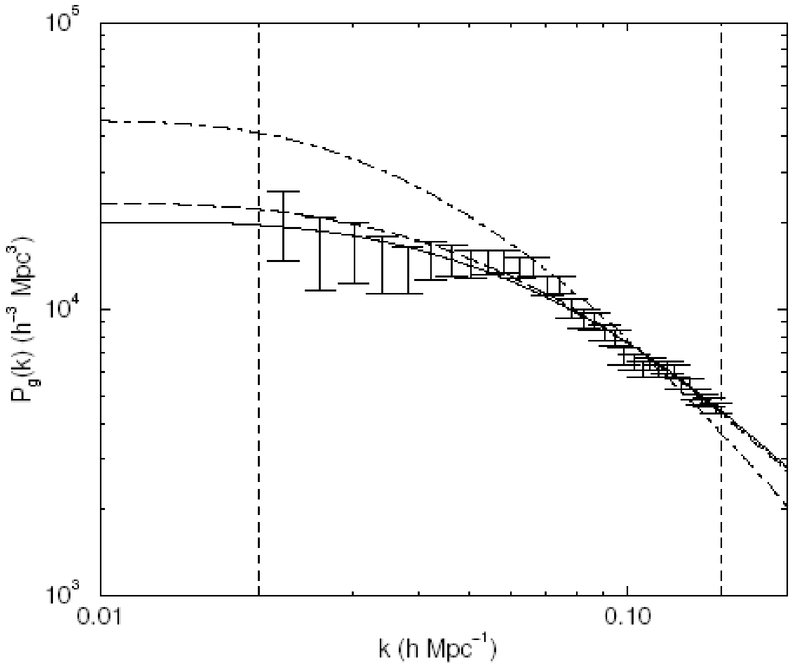


Figure 1.12: Fits to the 2dFGRS power spectrum assuming $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $h = 0.7$ for three different neutrino densities: $\Omega_\nu = 0$ (solid), 0.01 (dashed), and 0.05 (dot-dashed). Figure taken from [35].

The Ω_b estimate was possible because the models with baryon oscillations were mildly preferred by the data (over the featurless power spectra). Alternatively, it is possible to estimate the matter density from the measurement of the redshift-space distortion parameter $\beta \sim \Omega_m^{0.6}/b$ in the quasi-linear region (b is the galaxy-matter bias, defined as the ratio of the galaxy-number fluctuations and the true mass fluctuations). One such analysis gives $\Omega_m = 0.23 \pm 0.09$ [36]. One can also perform a combined analysis of 2dFGRS and CMB data - this will be discussed below.

More recently, the Sloan Digital Sky Survey (SDSS) has produced a three-dimensional map of over 200,000 galaxies with average redshift of ≈ 0.1. The SDSS relies on a mosaic CCD camera and a dedicated 2.5 meter telescope at the Apache Point Observatory in New

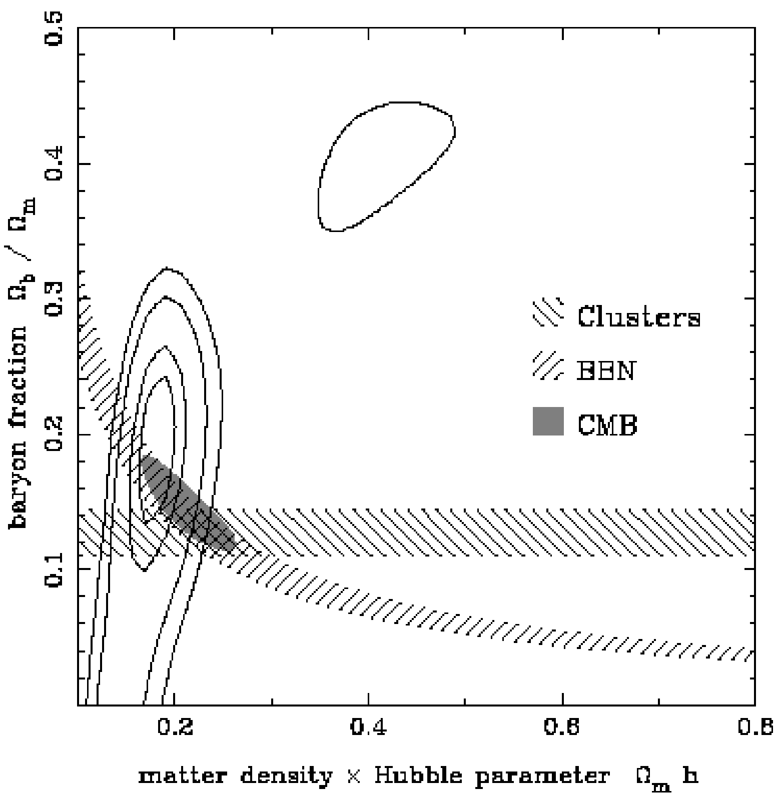


Figure 1.13: Likelihood surfaces obtained by fitting the 2dFGRS power spectrum for $\Omega_m h$, and the baryon fraction, Ω_b/Ω_m. The fit assumes a prior on the Hubble constant of $h = 0.7 \pm 0.07$. Estimates from the X-ray cluster analysis, the Big Bang Nucleosyntesis, and the CMB are also shown. Figure taken from [35].

Mexico, to image the sky in five photometric bandpasses. Tegmark *et al.* [37] performed the analysis of this data, in which they extract 3 power spectra: galaxy-galaxy, galaxy-velocity and velocity-velocity. They use a matrix-based method with eigenmodes to get uncorrelated minimum-variance measurements in 22 k-bands. They also perform a correction of the non-linear redshift distortions. The SDSS power spectrum extracted in this way is shown in Figure 1.14.

A.C. Pope *et al.* [38] used this measurement of the power spectrum to constrain the cosmological parameters. They fit for $\Omega_m h$ and $f_b = \Omega_b/\Omega_m$, assuming a prior of

$H_0 = 72 \pm 8$ km/s from the Hubble key project [39]. They find $\Omega_m = 0.375 \pm 0.089$ and $\Omega_b = 0.087 \pm 0.039$, as shown in Figure 1.15. The uncertainties are significantly reduced if they assume a prior on Ω_b from WMAP, indicating that the SDSS data alone is not yet capable of breaking the degeneracy between $\Omega_m h$ and f_b, as the baryon oscillations are not yet resolved. Again, significant improvements can be obtained by combining the SDSS data with the CMB data, as discussed below.

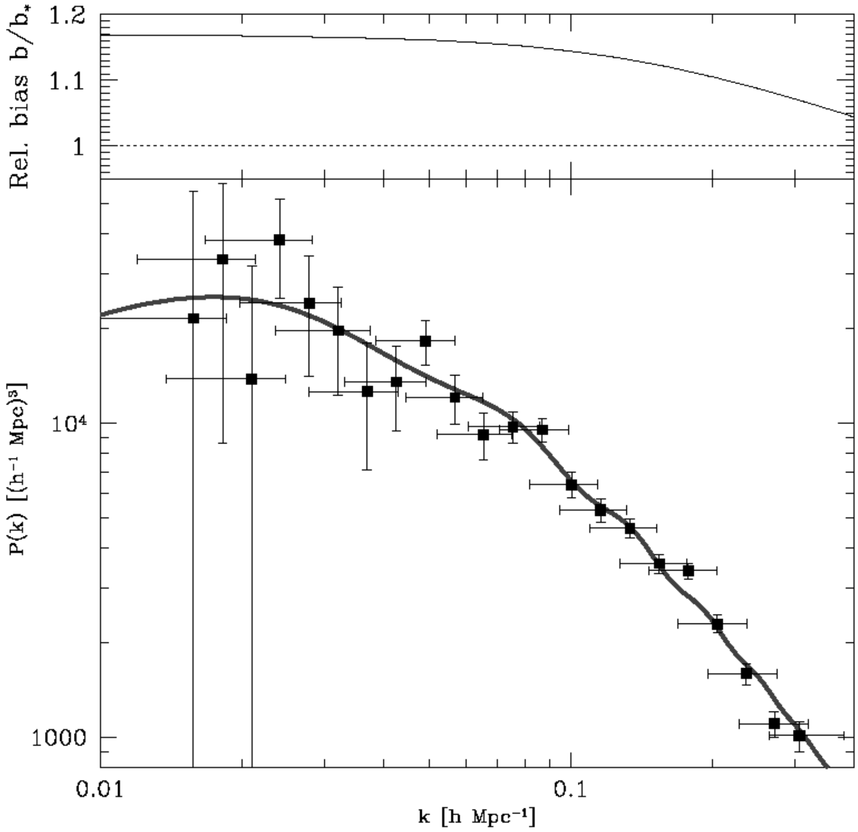


Figure 1.14: SDSS results: the decorrelated real-space galaxy-galaxy power spectrum is shown (bottom panel). To remove scale-dependent bias caused by luminosity-dependent clustering, the measurements have been divided by the square of the curve in the top panel, which shows the bias relative to L* galaxies. The solid curve (bottom pannel) is the best-fit model with $\Omega_m = 0.3$ for the priors $\Omega_b h^2 = 0.024$ (favored by the WMAP) and $h = 0.72$ (favored by the Hubble key Project [39]). Figure taken from [37].

Finally, we note that constraints on the power spectrum can also be obtained

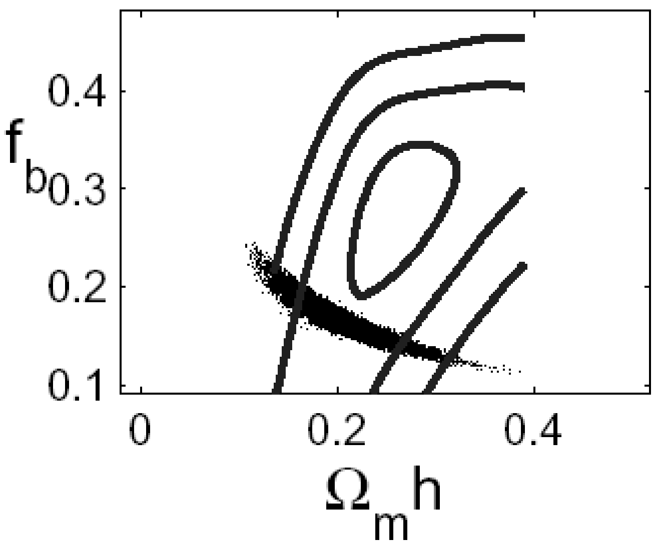


Figure 1.15: SDSS results: Two-dimensional 1, 2, and 3 σ contours in the $\Omega_m h - f_b$ plane are shown. The points denote Monte Carlo Markov Chains from WMAP (alone). Figure taken from [38].

from the clusters of galaxies. Clusters are believed to have collapsed from $\sim 3\sigma$ density fluctuations. Their abundance and evolution in time are dependent on the amount of matter in the Universe. Intuitively, this is because for large Ω_m the expansion is faster, making it more difficult for the density fluctuations to start collapsing. As a result, the larger Ω_m, the later the clusters will start forming.

This is another very active field, with a number of approaches applied to a number of observations. We mention some of these studies. Bahcall *et al.* [40] use the dependence of the observed cluster abundance on redshift to constrain Ω_m (and σ_8, the rms density variation in a sphere of radius 8 h^{-1} Mpc, which sets the scale of the power spectrum): $\Omega_m = 0.34 \pm 0.13$ for $\Omega_\Lambda = 1$ models. More recently, Bahcall *et al.* [41] used the mass function of clusters (the relationship of the number density of clusters and cluster mass), obtained using preliminary SDSS data, to constrain the combination $\sigma_8 \Omega_m^{0.6} = 0.33 \pm 0.03$.

Using the shape of the mass function, they partially break the degeneracy between the two parameters, yielding $\Omega_m = 0.19^{+0.08}_{-0.07}$.

Borgani *et al.* [42] performed a study of 103 clusters from the ROSAT Deep Cluster Survey (RDCS) with redshifts up to $z \approx 0.85$ and 4 clusters with redshifts up to $z = 1.26$. Their approach was to model the evolution of the cluster X-ray luminosity distribution. They assumed three free parameters in their models: the matter density Ω_m, the rms fluctuation amplitude at the $8h^{-1}$ Mpc scale (σ_8), and the power-spectrum shape parameter Γ. Model predictions for the cluster masses were converted into the X-ray temperature assuming hydrostatic equilibrium (similarly to elliptical galaxies). To convert the X-ray temperature into the X-ray luminosity profile (which is what they observe), they used a $L_X - T_X$ relationship based on a number of measurements including those from the Chandra observations. Figure 1.16 shows the results of their maximum-likelihood fit over $(\Omega_m, \sigma_8, \Gamma)$. Fairly stringent constraints can be placed on Ω_m - the overall estimates are:

$$\Omega_m = 0.35^{+0.13}_{-0.10},\ \sigma_8 = 0.66^{+0.06}_{-0.05}, \tag{1.36}$$

and the errors are even smaller if a particular value of Γ is chosen. No significant constraints were found on Γ because the sampled range of L_X does not correspond to large enough mass range to probe the shape of the power spectrum. Varying the $L_X - T_X$ relation does affect the estimates somewhat, but the estimate of Ω_m is always below 0.6 at least at 3σ confidence level.

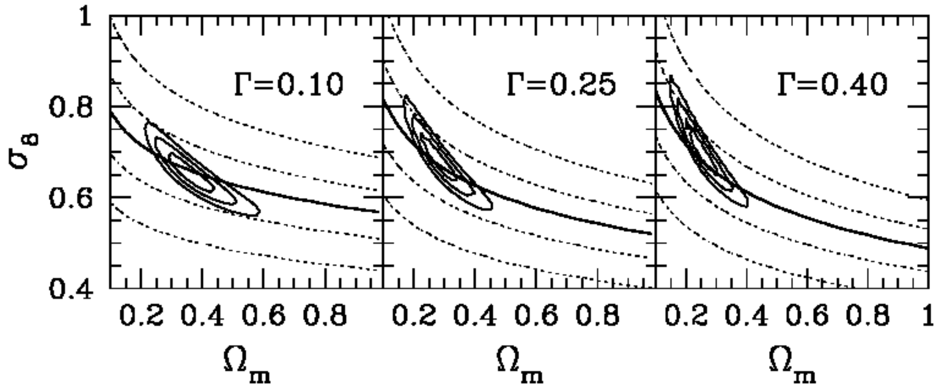


Figure 1.16: 1, 2, and 3 σ confidence regions in the $\sigma_8 - \Omega_m$ plane for different choices of Γ are shown as thin solid lines. The dotted lines are $\sigma_8 - \Omega_m$ relations corresponding to 0.1, 1, 10, and 30 high-redshift clusters. The thick solid line correspond to the observed 4 high-redshift clusters. Figure taken from [42].

1.3.7 Cosmic Microwave Background

Over the last several years, measurements of the anisotropy and the polarization of the Cosmic Microwave Background (CMB) have led to very stringent constraints on a number of cosmological parameters. We summarize here the estimates of the matter content of the Universe based on these measurements.

The CMB photons come from the surface of last scattering, at which (some 300,000 years after the Big Bang) the baryons and photons decoupled from each other. As discussed above, the growth of structure has already begun by that time, but the infall of baryons into the existing gravitational wells was opposed by the photon-pressure. Due to the tight coupling between the baryons and the photons, the density fluctuations present in the baryon-photon plasma prior to decoupling are expected to have left their signature in the photon temperature after decoupling, which we can measure today.

The measurement approach is somewhat similar to that in the Large-Scale Structure Formation. Namely, one measures the temperature $T(\hat{x})$ of the CMB radiation in different directions $\hat{x}$ in the sky. One then computes the temperature fluctuation in every direction, $\delta T(\hat{x})$, and expands in spherical harmonics:

$$\frac{\delta T(\hat{x})}{T} = \sum_{l=2}^{\infty} \sum_{m=-l}^{+l} a_{lm} Y_{lm}(\theta, \phi). \tag{1.37}$$

The angular power spectrum $C(\theta)$ is defined using the auto-correlation function:

$$\begin{aligned} C(\theta) &= \left\langle \frac{\delta T(\hat{x}_1)}{T} \frac{\delta T(\hat{x}_2)}{T} \right\rangle \\ &= \frac{1}{4\pi} \sum_{l=2}^{\infty} (2l+1) \, \langle a_{lm} a^*_{l'm'} \rangle \, P_l(\hat{x}_1 \cdot \hat{x}_2) \\ &= \frac{1}{4\pi} \sum_{l=2}^{\infty} (2l+1) \, C_l \, P_l(\hat{x}_1 \cdot \hat{x}_2). \end{aligned} \tag{1.38}$$

The last line defines the coefficients C_l (of the Legendre polynomials P_l), which are usually quoted in the measurement reports. Note that we have left out the dipole component ($l = 1$), as it is dominated by our peculiar velocity.

From the theoretical side, the sizes of various cosmological effects of the angular temperature spectrum must be determined. This is a fairly complex calculation, so we only outline the main effects following W. Hu *et al.* [43] - for more detail, see [44] or [45]. Figure 1.17 shows qualitatively how different cosmological parameters affect the shape of the spectrum - we discuss these effects now.

As mentioned above, the primordial density fluctuations try to gravitationally compress the baryon-photon fluid, while the photon pressure tries to resist this compression. The result is an acoustic oscillation in the fluid. At a given scale, the phases of different oscillation modes are synchronized. The oscillation mode that reaches its first compression by the last scattering is given by the wavenumber $k_A = \pi/s$, where s denotes the sound horizon at that time. Hence, k_A corresponds to the first peak (largest angle, smallest l) in

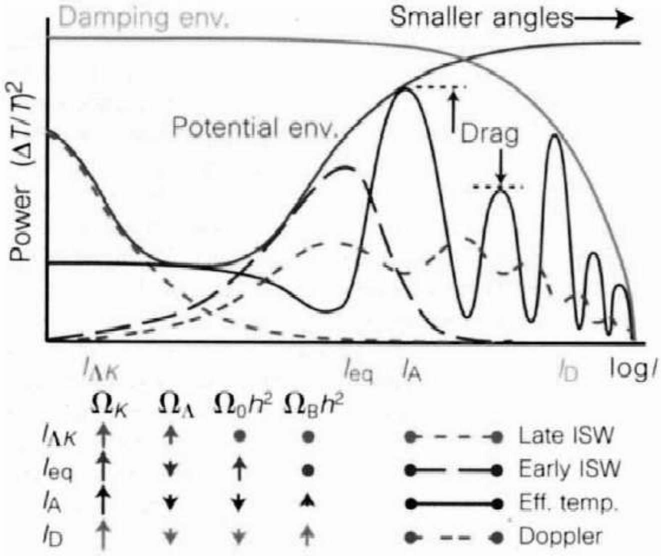


Figure 1.17: Schematic decomposition of the anisotropy spectrum. The four characteristic scales and their relation to cosmological parameters are discussed in the text. Figure taken from [43].

the angular spectrum. Of course, there will be a harmonic series of temperature fluctuation peaks, with $k_m = mk_A$ for the m^{th} peak. Since we are considering the squared temperature fluctuations, the compression and rarefaction peaks will alternate in the spectrum.

The actual angular size of the horizon size s, observed today, strongly depends on the geometry of the Universe. For example, in a closed Universe, a larger angle is needed to capture a given physical size than in a flat Universe. Hence, the position of the first peak, l_A (corresponding to the oscillation mode k_A), is very sensitive to the geometry of the Universe, or to Ω_k, or to $\Omega_{tot} = \Omega_m + \Omega_\Lambda$.

The baryon density can be extracted from the relative sizes of the compression and the rarefaction peaks. Essentially, baryons increase the effective mass of the fluid, causing a greater gravitational compression of the fluid in the potential well and shifting the zero-point of the oscillation. This increases the amplitude of the oscillation, and it changes the absolute (rms) value of the maxima vs. minima of the effective temperature fluctuation. Compressions are enhanced over rarefactions of the fluid inside potential well. This effect is known as the “baryon drag”.

The spectrum at small scales (high l) is suppressed by diffusion damping - the photons diffuse through the fluid, allowing the hot and cold regions to mix, thereby suppressing the fluctuations. This effect introduces another scale in the anisotropy spectrum, l_D, at which the diffusion damping becomes significant. This particular scale is sensitive to the density of the baryons Ω_b, as the Compton scattering of photons off of baryons sets the mean-free-path of the photons in the fluid.

The photons undergo gravitational redshifts and blueshifts as they traverse gravitational potential wells. Since the potential wells evolve, the blueshifts do not necessarily exactly cancel with the redshifts. This effect is known as the integrated Sachs-Wolfe (ISW) effect. If the Universe is not matter dominated at the last scattering, the potential will decay for the modes that cross the sound horizon between the last scattering and the matter domination: $k_{eq} < k < k_A$. Hence, the photons corresponding to these modes will be effectively redshifted, and the spectrum at these scales will increase. This is known as the early ISW effect. Note that since Ω_m strongly affects the time of matter-radiation equality, the early ISW effect can be used to constrain Ω_m.

Similarly, in an open or in a Λ-dominated Universe, the Universe will start rapidly expanding when the matter-domination stops (and the curvature or Λ domination starts, respectively). Hence, the potential will decay for modes below the horizon-size at the time when this effect takes place, $k > k_{K\Lambda}$, and the spectrum at these scales is expected to increase. This is known as the late ISW effect - it can be used to constrain Ω_Λ.

We also mention the Doppler-shift of the photon frequency due to the motion of the fluid. This effect essentially smooths the anisotropy spectrum by making the peaks less prominent. The Doppler effect is sensitive to Ω_b because the baryons increase the effective mass, hence causing the velocity to decrease.

In addition to the angular spectrum, the CMB is also characterized by its polarization. The CMB polarization arises from the Thompson scattering by electrons off of a radiation field with a local quadrupole moment. Since the quadrupole moment is suppressed until decoupling, at which time the largest contribution comes from the Doppler shifts due to the motion of the fluid, the CMB polarization directly probes the dynamics at decoupling. Such polarization modes are expected to be curl-free, and are, hence, named "E", in analogy with the electric field. Since the level of polarization depends on the fluid velocity, and highest velocities correspond to rarefactions in the density, the E-mode power spectrum is expected to be out-of-phase with the temperature spectrum. In other words, high level of TE correlation is expected. Measuring the E-mode polarization and the TE correlation is, therefore, a critical test of the underlying theory, and can be further used to constrain various cosmological parameters. We note that primordial gravitational waves could also induce E-mode polarization, along with a B-mode (curl-component) of the polarization, which could be used to directly probe inflation. B-modes can also be introduced by the gravitational lensing of the CMB photons as they pass by the clusters. We will not discuss this here.

To summarize, the CMB anisotropy and polarization carry much information on the cosmological parameters. However, there are clearly degeneracies among the different effects, so one cannot expect to constrain all of the parameters by the CMB alone. A combination with the power spectrum measurement, for example, is expected to further improve those constraints.

We now turn to the observations. Over the last several years, a number of experiments observed the CMB temperature anisotropy and polarization, with increasing presicion and using different techniques. We mention some of them:

- Satellite based COBE [46, 47], measured the temperature anisotropy at ∼10 degrees angular resolution ($l \sim 10$).
- Balloon-based BOOMERANG [48, 49], one of the first two experiments to measure the first acoustic peak.
- Balloon-based MAXIMA [50, 51], the other of the first two experiments to measure the first acoustic peak.
- Ground-based DASI [52, 53], measured the first acoustic peak, the first to measure the CMB polarization.
- Ground-based ACBAR [54], measured the temperature anisotropy out to $l \sim 3000$.

Although all of these are remarkable experiments, which measured the acoustic peaks and established that we live in a flat Universe ($\Omega_{tot} \sim 1$), the current "state-of-the-art" measurement comes from the satellite-based WMAP experiment. Much information about this experiment can be found in the volume 148 of the Astrophysical Journal Supplement, which was entirely devoted to the WMAP design, operation, and the results from the first year of running. We mention here three of these papers.

WMAP was designed to measure the CMB anisotropy on the scales from the full sky to several arcminutes, at five frequencies in the 23-94 GHz range. It relies on two 1.4 m telescopes separated by $\sim 141^\circ$ degrees, and on the differential radiometer to measure the difference in the sky brightness between the two pixels. Hinshaw *et al.* [55] estimate the anisotropy spectrum in a variety of ways - their final result is shown in Figure 1.18. The measurement of the first two acoustic peaks (at $l = 220$ and $l = 540$) is very clean.

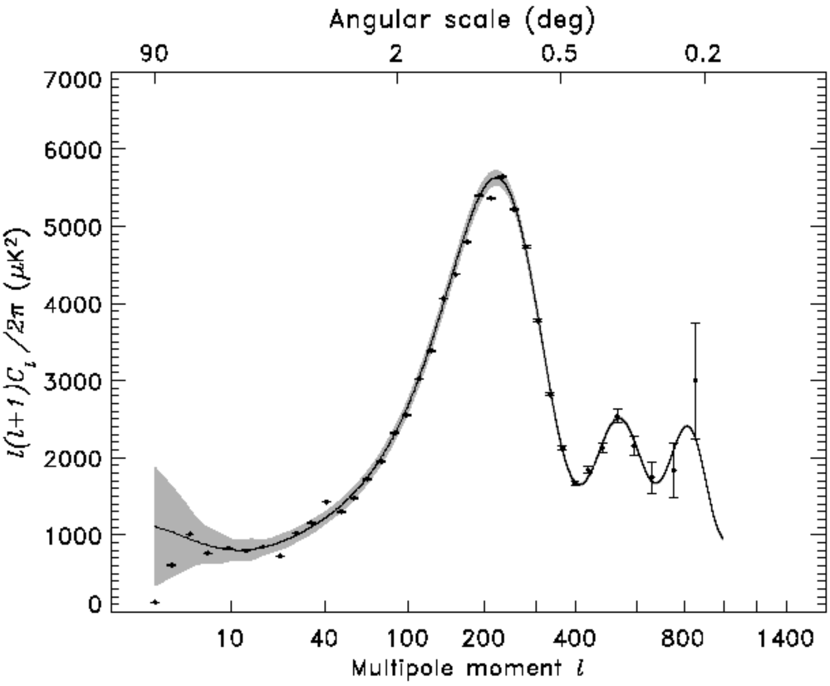


Figure 1.18: WMAP temperature anisotropy angular spectrum. The curve denotes the best-fit model from [56] (discussed in the text). The gray area around the curve is the 1σ uncertainty due to cosmic variance. Figure taken from [55].

Kogut *et al.* [57] detect correlations between the temperature and polarization maps, significant at more than 10σ. On small angular scales ($< 5^\circ$), the WMAP data show the temperature-polarization correlation expected from adiabatic perturbations in the temperature power spectrum - no additional parameters are needed. On large angular scales ($> 10^\circ$), they detect excess power, compared to predictions based on the temperature power spectra alone. The excess power is well described by reionization at redshift $11 < z < 30$ at 95% confidence. Their polarization spectrum is shown in Figure 1.19.

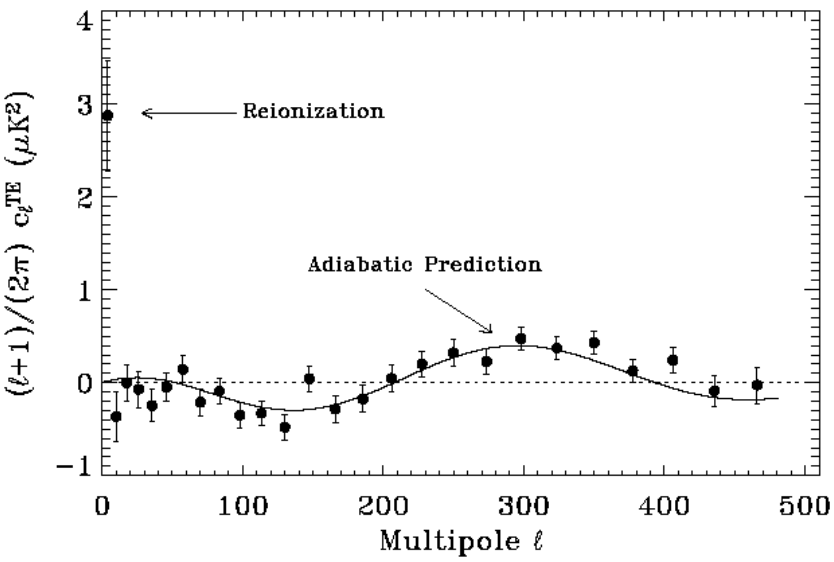


Figure 1.19: Polarization cross-power spectrum c_l^{TE} for the WMAP first-year data. The TE correlation on degree angular scales ($l > 20$) is in excellent agreement with the signal expected from adiabatic CMB perturbations. Figure taken from [57].

Spergel *et al.* [56] perform the fits of various cosmological models to the temperature and the temperature-polarization angular power spectra discussed above. They approach the problem with the goal of finding the smallest number of parameters that could fit the data. Assuming a flat Universe, they find that six parameters are sufficient, and estimate $h = 0.72 \pm 0.05$, $\Omega_m h^2 = 0.14 \pm 0.02$, and $\Omega_b h^2 = 0.024 \pm 0.001$ (from which they estimate $\Omega_b = 0.047 \pm 0.006$, and $\Omega_m = 0.29 \pm 0.07$; the other three parameters are amplitude $A = 0.9 \pm 0.1$, optical depth $\tau = 0.166^{+0.076}_{-0.071}$, and the spectral index $n_s = 0.99 \pm 0.04$).

This result implies that $\Omega_b \neq \Omega_m$ at 5.8σ.

Including other anisotropy measurements that probe finer scales, along with the power spectrum measured by 2dFGRS (discussed earlier), they further reduce the uncertainties: $h = 0.73 \pm 0.03$, $\Omega_m \approx 0.252 \pm 0.024$, and $\Omega_b \approx 0.043 \pm 0.004$ (Ω_m and Ω_b are extracted from their estimates of $\Omega_m h^2$, $\Omega_b h^2$ and h).

After the first SDSS data was published (as discussed above), it was possible to include it in the constraining of the various cosmological parameters. Tegmark *et al.* [58] perform the combined analysis of the WMAP temperature and temperature-polarization angular spectra and the SDSS power spectrum. Their approach was different from the one by Spergel *et al.* [56] - they perform a more extensive analysis, allowing up to 13 free parameters, and they study the impact of introducing these parameters to the original 6 studied in [56]. Their results are consistent with [56]. They find that combining the SDSS data (or other data, such as SNe Ia) with the WMAP data leads to reductions on uncertainties of estimated cosmological parameters (as compared to the WMAP-alone analysis), primarily because WMAP does not constrain the power spectrum significantly. The combined analysis of WMAP and SDSS, with the original six parameters only, gives $h = 0.695^{+0.039}_{-0.031}$, $\Omega_m = 0.301^{+0.045}_{-0.042}$, and $\Omega_b \approx 0.048 \pm 0.0061$ (Ω_b is estimated from their estimates of $\Omega_b h^2$ and h). Figure 1.20 shows the complementarity of the WMAP and SDSS data in determining the matter content of the Universe. For completeness, we also show their contours in the $\Omega_m - \Omega_\Lambda$ plane, Figure 1.21, where adding the SNe Ia information is also important.

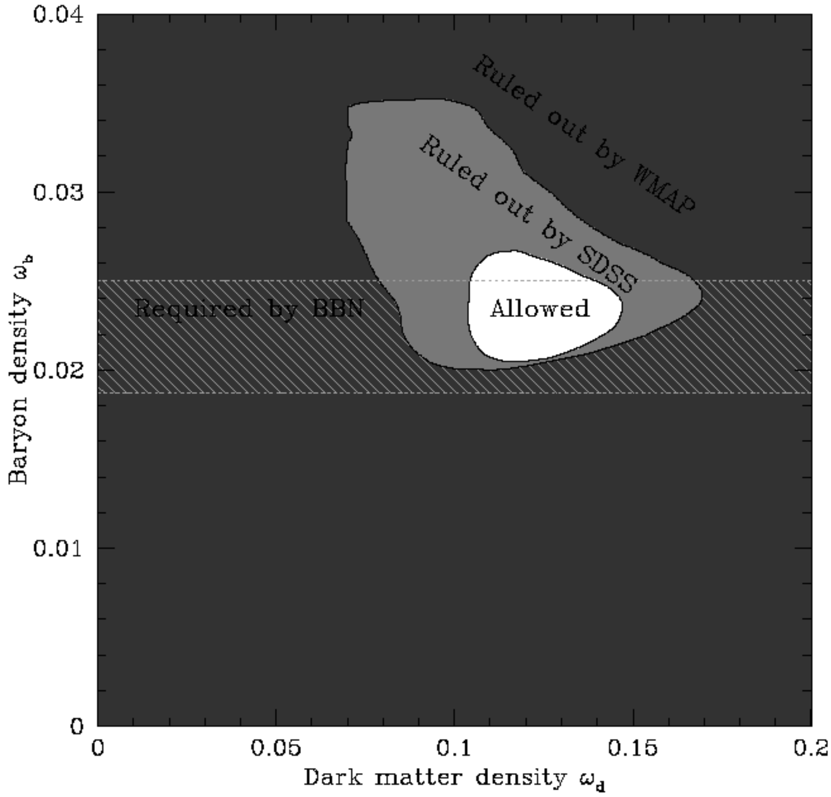


Figure 1.20: 95% C.L. constraints in the $\omega_b = \Omega_b h^2$ vs $\omega_d = \Omega_d h^2$, where Ω_d is the density of the cold dark matter component. Dark-gray region is excluded by WMAP alone, the light-gray region is excluded when adding the SDSS information, and the hatched band is required by BBN. Figure taken from [58].

1.3.8 Summary - Case for Non-Baryonic dark matter

We conclude this Section with a summary of the observational evidence supporting the existence of dark matter. The need for dark matter has been established at very different spatial scales - in spiral and eliptical galaxies, in galaxy clusters, and on the cosmological scale. Table 1.1 lists various estimates of Ω_m and Ω_b discussed in this Section. This list is nowhere near complete, but it does illustrate a very important point: experiments probing very different types of physics, including the virial theorem and the hydrostatic equilibrium techniques in galaxy clusters, the standard candle measurements in SNe Ia, the

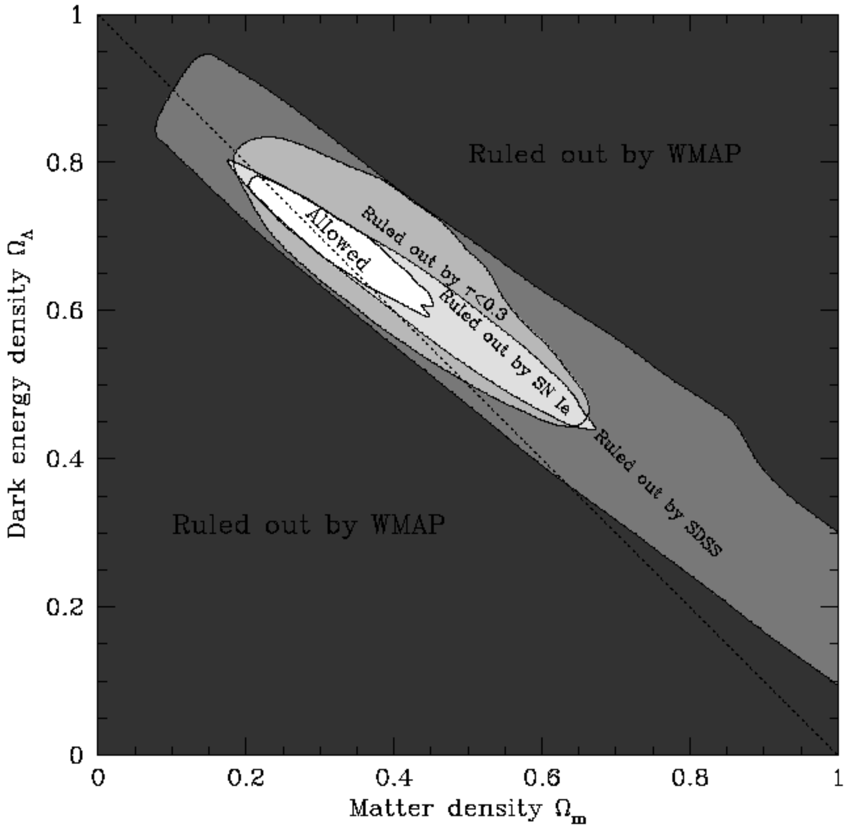


Figure 1.21: 95% C.L. constraints in the $\Omega_m - \Omega_\Lambda$ plane. Going inwards, the outer-most region (dark-gray) is excluded by WMAP alone, the next (light-gray) region is excluded when adding the SDSS information, the next region is excluded when also adding requirement $\tau < 0.3$, and the next (lightest-gray) region is excluded when also adding the SNa Ia information. The dotted line denotes flat Universes. Figure taken from [58].

galaxy-galaxy autocorrelation and the cluster abundance tests of the large-scale structure formation, the temperature and polarization anisotropy of the microwave background radiation, and the nucleosynthesis of the lightest elements - all agree that the matter makes up approximately 25% of the Universe *and* that most of it is non-baryonic! This is a truly remarkable achievement of the modern cosmology!

Particularly striking is the agreement of the baryon density estimates from the BBN (based on understanding the rates of nuclear reactions, taking place ~ 1 minute after

Approach	Measurement
Clusters - Virial [13]	$\Omega_m > 0.09$
Clusters - SZE [22]	$\Omega_m \approx 0.25$
SNe Ia - SCP [31]	$\Omega_m = 0.28^{+0.10}_{-0.09}(\text{stat})^{+0.05}_{-0.04}(\text{syst})$
SNe Ia - HZT [32]	$\Omega_m = 0.28 \pm 0.05$
LSS - 2dFGRS (g-g*; old) [33]	$\Omega_m = 0.29 \pm 0.07$
LSS - 2dFGRS (g-g*; new prelim) [35]	$\Omega_m = 0.26 \pm 0.05$
LSS - 2dFGRS (z-space distortion) [36]	$\Omega_m = 0.23 \pm 0.09$
LSS - SDSS (g-g*) [38]	$\Omega_m = 0.375 \pm 0.089$
LSS - clusters, evolution [40]	$\Omega_m = 0.34 \pm 0.13$ (for $\Omega_\Lambda = 1$)
LSS - clusters, mass function [41]	$\Omega_m = 0.19^{+0.08}_{-0.07}$
LSS - cluster X-ray evolution [42]	$\Omega_m = 0.35^{+0.13}_{-0.10}$
CMB - WMAP alone [56]	$\Omega_m = 0.29 \pm 0.07$
CMB - WMAP+2dFGRS [56]	$\Omega_m \approx 0.252 \pm 0.024$
CMB - WMAP+SDSS [58]	$\Omega_m = 0.301^{+0.045}_{-0.042}$
BBN - Deuterium** [23, 24, 25, 26]	$\Omega_b \approx 0.037 \pm 0.007$
LSS - 2dFGRS (g-g*; old) [33]	$\Omega_b = 0.044 \pm 0.021$
LSS - 2dFGRS (g-g*; new prelim) [35]	$\Omega_b = 0.044 \pm 0.016$
LSS - SDSS (g-g*) [38]	$\Omega_b = 0.087 \pm 0.039$
CMB - WMAP alone [56]	$\Omega_b = 0.047 \pm 0.006$
CMB - WMAP+2dFGRS [56]	$\Omega_b \approx 0.043 \pm 0.004$
CMB - WMAP+SDSS [58]	$\Omega_b \approx 0.048 \pm 0.006$

Table 1.1: Summary of various estimates of Ω_m and Ω_b. (*) g-g stands for galaxy-galaxy correlation. (**) Ω_b was extracted using the $\Omega_b h^2$ estimate from D abundance and $h = 0.72 \pm 0.07$ from the Hubble key Project [39].

the Big Bang) and from the CMB (based on electromagnetic interactions on the surface of last scattering, taking place 300,000 years after the Big Bang). Similarly impressive is the agreement of the CMB observations with the large-scale structure formation tests that probe processes happening on the time-scale of 10^9 years.

Finally, we re-emphasise the result of the analysis of the WMAP data alone [56]: assuming a flat Universe, $\Omega_b \neq \Omega_m$ at 5.8σ.

The evidence supporting the existence of dark matter and its non-baryonic character is indeed overwhelming. The next natural question to address is the nature (or content) of dark matter. The observational evidence discussed above gives some hints - the large scale structure formation requires some amount of cold dark matter, which decoupled early from the primordial plasma and started forming structure prior to the baryon-photon decoupling. Similarly, the CMB anisotropy spectra are well fitted with models in which most dark matter is cold.

However, it is possible that dark matter has several constituents, some of which may be baryonic or hot non-baryonic. We discuss these possibilities in the next Section.

1.4 Dark Matter Candidates

1.4.1 Baryonic Dark Matter Candidates

Although the previous Section established that the large fraction of the matter must be non-baryonic, there still remains the need for baryonic dark matter. In particular, the density of the luminous baryonic matter (observed in galaxies etc) is estimated to be $\Omega_{lum} \sim 0.005$, considerably less than the estimates listed in the Table 1.1. At high redshifts ($z > 1$), the remaining baryons are observed in the form of ionized hydrogen in the intergalactic clouds. Detailed measurements have been performed of the absorption lines in the spectra of high-redshift quasars (usually referred to as the Lyman α forest), indicating the presence of the hot ($10^4 - 10^5$ K) ionized H gas in the intergalactic medium. These measurements result in a lower limit on the matter density in such clouds [59, 2]: $\Omega_g h^2 > 0.018$ (the lower limit is due to various systematics that could not be estimated). This is consistent with the estimates listed in the Table 1.1.

However, the Lyman α forest disappears at redshifts < 1. The question is, then, what happened to the hot ionized baryons from the high redshifts? The MAssive Compact Halo Objects (MACHOs) were one of the most popular candidates proposed to solve this problem. MACHO is a generic name for dark, compact and massive objects populating the halo of our galaxy. Such objects could, for example, be brown dwarfs - gravitationally collapsed objects of sub-stellar mass that could not reach high enough pressure in the core to start fusing hydrogen. Hence the mass of such objects is bounded from above ($< 0.08 M_\odot$) by this requirement. Another possibility are Jupiter-like objects, with mass $\sim 10^{-3} m_\odot$. Yet another possibility are black holes, which may have been formed by collapsing baryonic matter.

Since these objects are dark, one way to search for them is gravitational lensing. In particular, one can observe temporary brightening of a star due to a MACHO passing near the line-of-sight between the observer and the star. Since the probability of such events is very small, millions of stars have to be monitored on a daily basis, in order to make the search plausible. The duration of such "microlensing" event is determined by the lens (MACHO) mass m, distance x, transverse velocity v, and by the distance L of the source star [60]:

$$\Delta t \sim \sqrt{mx(L-x)/(v^2 L)}. \tag{1.39}$$

Clearly, there is a degeneracy since the only two observables are the duration of the event and its probability.

Two experiments, MACHO [61] and EROS [62] searched for these objects by making nightly observations towards the Large and the Small Magellanic Clouds. Over a 5.7-year period, MACHO observed 13-17 events (depending on the classification) toward the LMC of typical duration $\sim$ 100 days, and one towards the SMC (also observed by EROS). This rate is larger than the expected background of 2-4 events due to known stellar populations, but not enough to account for a significant fraction of the halo. The absence of short-duration events ($\sim$ 20 days) implies that less than 20% of the halo can be in the form of MACHOs of mass between $10^{-4}\ M_{\odot}$ and $0.03\ M_{\odot}$. Furthermore, from the observed rate of events, the MACHO team determined that an all-MACHO halo is ruled out at 95% confidence, in the same mass range [61]. Based on the 4 observed events towards the SMC, the EROS team placed similar constraint: $< 25\%$ of the halo is in the form of MACHOs in the mass range $2 \times 10^{-7} - 1M_{\odot}$, at 95% confidence [62]. More recently [63], the MACHO team extended the excluded mass range, based on the lack of events with very long duration-times (> 150 days) towards the LMC: objects with masses $0.3 - 30M_{\odot}$ cannot account for the whole dark matter halo at 95% confidence. In particular, this result excludes objects such as black holes in this mass range.

To conclude, although the MACHOs have been shown to exist, and may even account for a part of the baryons missing at low redshifts, they cannot account for all of dark matter. This is hardly a surprise, given the wealth of evidence indicating the non-baryonic character of dark matter.

1.4.2 Non-Baryonic Candidates

We now turn to the non-baryonic candidates. Three candidates have caused the most excitement over the recent years: neutrinos, axions, and WIMPs.

Neutrinos

Neutrinos decoupled from the primordial plasma when the temperature of the Universe dropped to ~ 1 MeV. They were relativistic at the time of decoupling, and are hence an example of hot dark matter. Although a viable dark matter candidate, it is difficult to explain the large-scale structure observed today with models in which neutrinos make up most of the dark matter. This is primarily due to the effect of free-streaming: neutrinos can freely move between the more dense and the less dense regions, hence suppressing the density fluctuations. However, a small amount of neutrinos may be useful in suppressing the excess power at small scales that is characteristic for the cold dark matter (CDM) models (thereby making the CDM models more compatible with observations).

The neutrino density today can be estimated as [2]:

$$\Omega_\nu = \frac{8\pi G}{3H_0^2} n_\nu m_\nu = 0.011 \left(\frac{m_\nu}{1 \text{ eV}} \right) h^{-2}, \tag{1.40}$$

where $n_\nu = 113$ cm^{-2} is the current number density of neutrinos, and m_ν is their mass. The neutrino mass has not been measured, but limits have been placed. The Standard Model of Particle physics does not predict neutrinos to be massive. However, measurements of the atmospheric and solar neutrino fluxes (by a number of experiments) have established that neutrinos oscillate among the three flavors, implying that their masses are non-zero. For example, SuperKamiokande established the $\nu_\mu - \nu_\tau$ oscillation from the reduced flux of the atmospheric neutrinos, with $\Delta m^2 = 3.2 \times 10^{-3}$ eV^2 [64]. More recently, Sudbury Neutrino Observatory (SNO) measured the $\nu_e - \nu_\mu$ oscillation from the flux of solar neutrinos, with $\Delta m^2 = 7.1^{+1.0}_{-0.3} \times 10^{-5}$ eV^2 [65]. However, these experiments determine only the differences among the neutrino masses - the absolute scale is not known. Using Equation 1.40, therefore, the SuperKamiokande result sets a lower bound on the neutrino density today $\Omega_\nu \gtrsim 6\times10^{-4}$.

The upper bounds on the neutrino density come from the large scale structure observations. Croft *et al.* [66] use the Lyman α forest in the quasar spectra, along with hydrodynamic simulations, to measure the mass power spectrum. Combined with measurements of other cosmological parameters available at the time (such as $\Omega_b h^2$ from the BBN and σ_8 from galaxy power spectra), they constrain the neutrino mass: $m_\nu < 5.5$ eV ($\Omega_\nu \lesssim 0.1$) for all values of Ω_m.

More recently, the measurements of the CMB anisotropies by WMAP and of the power spectrum by 2dFGRS and SDSS allowed constraining many cosmological parameters, including the neutrino density. For example, Tegmark *et al.* [58] in the combined analysis of the WMAP and SDSS data estimate that $f_\nu = \Omega_\nu/\Omega_{dm} < 0.12$ at 95% confidence, corresponding to $M_\nu < 1.74$ eV at 95% confidence (M_ν is the sum of the masses of the three neutrino species). Most of this constraint comes from the SDSS measurement of the power spectrum - as mentioned earlier, the free-streaming of neutrinos tends to suppress power at small scales.

To conclude, although it has been shown that neutrinos are indeed massive, their abundance today can explain $\lesssim 12\%$ of the dark matter.

Axions

Axions were originally proposed as a solution to the strong-CP problem. Namely, the QCD Lagrangian contains the term

$$\frac{\theta g^2}{32\pi^2} G^a_{\mu\nu} \tilde{G}^{a\mu\nu}, \tag{1.41}$$

where G is the gluon field strength and g is the QCD coupling. The parameter θ depends on the quark masses, and it has no a priori reason to be zero. However, if $\theta \neq 0$, this term violates P and CP. The most stringent constraint on θ currently come from the measurement of the neutron dipole moment: $|\theta| < 10^{-9}$ [67]. Hence, the strong CP violation has not been observed yet, but it is not understood why θ is so small. This is the strong CP problem. Before we discuss axions as a solution to this problem we note that there are alternative "solutions" - in particular, if $\theta = 0$ at the Planck scale, values $< 10^{-9}$ at the weak scale are not unexpected [67, 68].

A solution to the problem was proposed by Peccei and Quinn (1977) [69] - they introduced a new $U(1)$ symmetry which is broken at some scale f_a, and the quasi-Goldstone boson associated with the symmetry-breaking is the axion. The symmetry-breaking scale is related to the vacuum expectation value v of the potential that breaks the symmetry, and is, hence, related to the mass of the axion:

$$m_a \approx 0.6 \text{ eV} \frac{10^7 \text{ GeV}}{f_a}. \tag{1.42}$$

The symmetry-breaking scale, and the mass of the axion, are *a priori* unknown. However, they are significantly constrained by accelerator experiments and astrophysical observations. In particular, the observation of the neutrino signal from the supernova 1987a, whose duration was consistent with the hypothesis that the collapsed supernova cools solely by emitting neutrinos, rules out the axion masses in the range $3 \times 10^{-3} - 2$ eV. Combined with the accelerator searches, this sets the upper limit on m_a of 3×10^{-3} eV.

On the other hand, the relic density of axions is inversely proportional to m_a. Hence, requiring $\Omega_a < 1$ imposes a lower bound on the axion mass $m_a > 10^{-6}$ eV. There are, therefore, three orders of magnitude in mass for which axions are plausible: 10^{-6} eV $< m_a < 3 \times 10^{-3}$ eV.

Cosmologically, axions appear as the Universe cools below the temperature corresponding to the Peccei-Quinn symmetry-breaking scale, f_a. The axion temperature at this scale is still high compared to the new effective potential, so the axion field can still take any value. As the temperature drops below the QCD scale (~ 300 MeV), the axion falls into the potential well and acquires mass. There are several mechanisms for axion production, ranging from the freeze-out process to the decay of the axion strings [3, 67]. Each of them could be the dominant mechanism, depending on the axion mass, inflation etc. They are too complex to discuss here, but their common feature is that the axions are produced at small velocities, and should thus behave as cold dark matter.

The favored approach used in the axion searches relies on microwave cavities in magnetic fields. In particular, axion is expected to interact in a strong magnetic field and produce photons in the microwave regime. The exact form of the interaction is

$$L_{a\gamma\gamma} = -g_\gamma \frac{\alpha}{4\pi f} \phi_a F_{\mu\nu} \tilde{F}^{\mu\nu} = -g_{a\gamma\gamma} \phi_a \vec{E} \cdot \vec{B}, \tag{1.43}$$

where α is the fine structure constant, ϕ_a is the axion field and $F_{\mu\nu}$ is the photon field strength. $\vec{E}$, and $\vec{B}$ are the electric and magnetic fields, and g_γ is a model dependent constant. Two most important axion models are Dine-Fischler-Srednicki-Zhitnitskii (DFSZ; $g_\gamma = 0.36$) and Kim-Shifman-Vainshtein-Zakharov (KSVZ; $g_\gamma = -0.97$). Hence, the experiments set limits in terms of $g_{a\gamma\gamma}$.

The most sensitive search so far has been performed by the U.S. Large Scale Search [67]. They used a cryogenically cooled, tunable microwave cavity with a heterostructure field-effect transistor (HFET) amplifier. Tuning of the cavity allows scanning through the frequency of the emitted photon, or, equivalently, through the axion mass range. Their results are shown in Figures 1.22 and 1.23. These results exclude the KSVZ axion in the axion-mass range 2.3-3.3 μeV, and are about 1 order of magnitude away in sensitivity from being able to probe the DFSZ axion. More recently [70], they extended the excluded mass down to 1.9 μeV.

We note that there remains a large fraction of the allowed mass range, which is beyond reach of the current experiments. However, significant improvements are expected in the next generation of the axion-search experiments. Two directions will be pursued [67], one using low-noise SQUID-based amplifiers and the other using a Rydberg-atom single-quantum detector. Both techniques are expected to probe the DFSZ axion in the coming years.

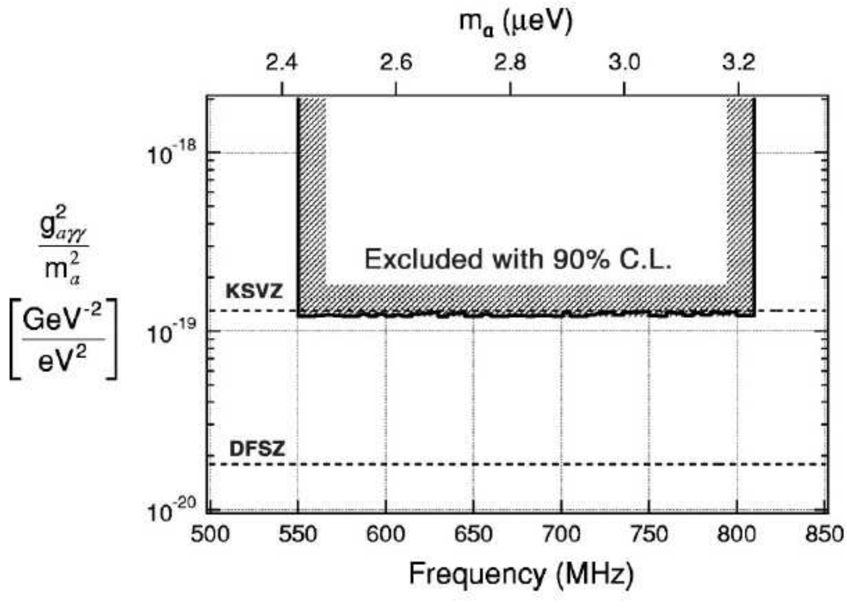


Figure 1.22: Axion sensitivity range excluded at 90% C.L. by the US Large Scale Search. The KSVZ models are excluded in the mass range $2.3 - 3.3\mu$eV. Figure taken from [67].

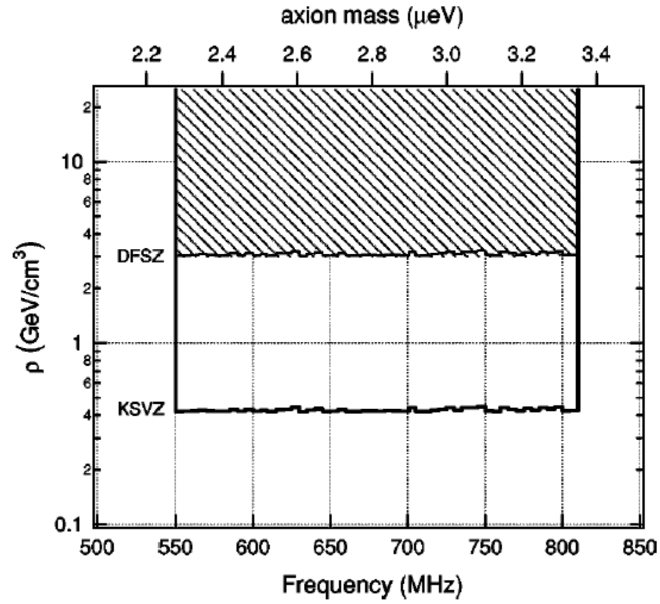


Figure 1.23: Axionic halo density excluded at 90% C.L. by the US Large Scale Search, for the two most interesting axion models. Figure taken from [67].

WIMPs

Particles that were in the thermal equilibrium in the early stages of the Universe and that decoupled from the primordial plasma when they were non-relativistic, could potentially be cold dark matter candidates. These particles would have to be massive to account for (a part of) the dark matter. They would also have to be weakly interactive, which would explain why they have not been observed yet and would allow their relic density today to be cosmologically relevant. Such particles are covered with a generic name Weakly Interactive Massive Particles (WIMPs).

To turn the argument around, it can be shown that the relic density of these particles today is inversely proportional to their annihilation cross-section. If their density today is to be of the order of the critical density, the annihilation cross-section must be of the weak-interaction scale. Note, however, that this does not imply weak interactions themselves! By the crossing argument (crossing the legs of the relevant Feynmann diagrams), the elastic scattering off of nucleons should be of the weak scale as well (although significant corrections to this argument must be made, as described in Chapter 2).

WIMPs are particularly attractive because a natural WIMP candidate is offered by supersymmetry. Namely, the lightest supersymmetric particle in the Minimal Supersymmetric Standard Model (MSSM) is expected to be stable, massive and interact with ordinary matter on the weak scale - exactly the properties of a WIMP. Moreover, supersymmetry very elegantly solves some of the remaining problems with the Standard Model of Particles, such as the mass hierarchy problem, and is, therefore, one of the most favored extensions of the Standard Model. We will discuss the theoretical aspects of the supersymmetric WIMP candidate in detail in Chapter 2.

1.5 Indirect WIMP-Detection Experiments

We now review the current state of the various WIMP search methods. In this Section, we discuss methods of indirect detection, while the following Section discusses methods of direct detection. Unless specified otherwise, we assume a supersymmetric WIMP candidate in these two Sections.

Indirect detection experiments search for products of WIMP annihilation in regions that are expected to have relatively large WIMP concentration. Examples of such regions are centers of the galaxy, the Sun or the Earth, where the WIMPs are expected to be gravitationally captured. Higher WIMP density gives larger annihilation signal, which can be manifested as a flux of gamma rays, neutrinos, or antimatter (positrons or anti-protons) produced in the WIMP-annihilation. We discuss these possibilities in some detail.

1.5.1 Gamma Rays

WIMP-annihilation can produce γ rays in several different ways. First, a continuous spectrum is produced from the hadronization and decay of π^0's produced in the cascading of the annihilation products. Second, γ-ray spectral lines are produced by the annihilation channels in which γ's are dirrectly produced, such as $XX \to \gamma\gamma$ (producing a line at M_X) and $XX \to \gamma Z$ (producing a line at $M_X(1 - M_Z^2/M_X^2)$). Observing such lines would be a clear detection of the WIMP annihilation. However, the Feynmann diagrams contributing to these two annihilation channels involve loops. Hence, the flux of these lines is model dependent and is typically much smaller than the continuum flux.

Such gamma rays could be produced close to the galactic center. Although the production rates are relatively low, a large halo density may compensate sufficiently to make such signals observable. Since these gamma rays would propagate essentially freely, ground or satellite based experiments may observe them. Indeed a number of such experiments exist or is in preparation - we mention some of them here.

The ground-based experiments rely on the Atmospheric Čerenkov Telescopes (ACTs), which detect the Čerenkov light emitted by the shower produced by the gamma ray interacting at the top of the atmosphere. Some experiments, such as CELESTE (France) [71] and STACEE (New Mexico) [72], use the large mirror areas used by the solar power plants. These two experiments are sensitive to 20-250 GeV gammas. A number of experiments uses dedicated mirrors or arrays of mirrors with a detector in the focal point: CANGOROO (Australia) [73], VERITAS (Arizona) [74], CAT (France) [75], HESS (Namibia) [76], HEGRA (Canary Islands, dismantled) [77], MAGIC (Canary Islands) [78]. Such experiments are typically sensitive to 100 GeV - 10 TeV gamma rays. They are also capable of distiguishing (usually $> 99\%$ efficiency) between the showers caused by gamma rays and those caused by cosmic rays (dominant background). Another experiment, MILAGRO (New Mexico) [79], deploys a water-Čerenkov detector. This experiment has the advantage that it is not dependent on the weather, and it can monitor the entire sky continuously, which is clearly not the case with telescopes. However, it suffers from larger background, as it does not distinguish between the showers caused by gamma rays and those caused by cosmic rays. Finally, there are satellite based experiments: EGRET [80] completed its mission and observed gamma rays in the 20 MeV - 30 GeV energy range, and GLAST [81] is scheduled to launch in 2006 and observe gamma rays of energies 10 MeV - 100 GeV. GLAST is expected to have better energy and angular resolutions than EGRET and hence identify the various sources more precisely. This experiment uses a scintillator-based calorimeter and Si-Pb tracker of electrons and positrons used to extract the direction of the incoming gamma ray.

Recently, two experiments observed excess flux of gamma rays coming from the galactic center - see Figure 1.24. VERITAS [83], operating the Whipple 10 m telescope on Mt. Hopkins, Arizona, observed an integral flux of $1.6 \pm 0.5 \pm 0.3 \times 10^{-8}$ m^{-2}s^{-1} with the energy threshold of 2.8 TeV. The high energy threshold is due to the fact that the telescope is on the northern hemisphere, so the observations of the galactic center are made close to the horizon. CANGOROO [84], on the other hand, is located in the southern hemisphere, so its energy threshold is much lower - 250 GeV. CANGOROO made a $\sim 10\sigma$ detection of

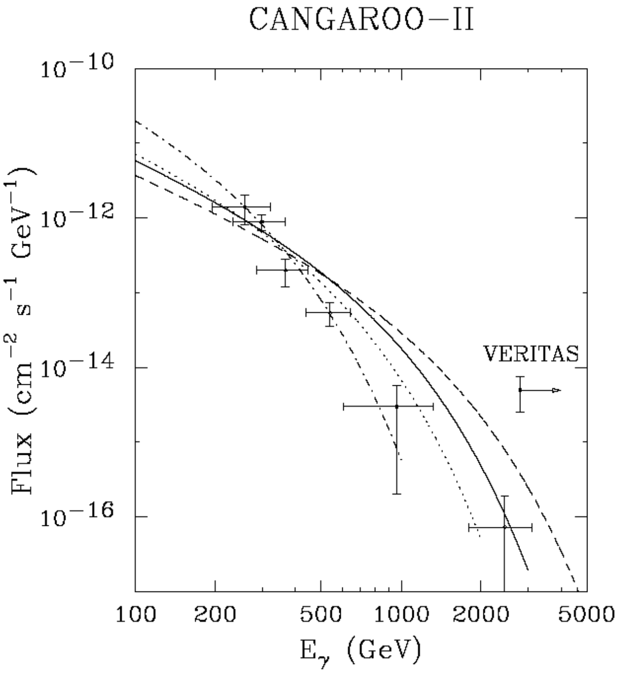


Figure 1.24: Observed gamma-ray flux from the galactic center from CANGOROO (crosses) and VERITAS (single point at 2.8 TeV). The dot-dashed, dotted, solid and dashed curves are predicted spectra for WIMP annihilation into gauge bosons for 1, 2, 3, and 5 TeV WIMPs. The VERITAS point was inferred from the integral flux assuming a 5 TeV WIMP. Figure taken from [82].

the gamma ray source in the galactic center, in six energy bins over the range 250 GeV - 2.5 TeV. Both experiments find that their measurements are consistent with a point-source, within their angular resolutions (this does not exclude cuspy distribution). In addition, EGRET also found a strong, unidentified, source of GeV gamma rays, which appears to be about 10 arminutes away from the dynamic center of the galaxy.

Hooper *et al.* [82] examine the compatibility of these observations with the possibility of the signal being generated by WIMP annihilations. They find it difficult to reconcile the results of CANGOROO and VERITAS. The spectrum measured by CANGOROO is consistent with a WIMP mass of 1-3 TeV, while VERITAS, with its energy threshold of 2.8

TeV, requires a much heavier WIMP. Moreover, very high annihilation rates are required for this signal to be explained by WIMP-annihilation. This implies very high annihilation cross-section *and* very high dark matter concentration at the galactic center. In particular, extremely cuspy models, or models with a density spike are necessary.

We note the possibility that these observations could also be explained by astrophysical sources. In particular, the black hole at the galactic center, coincident with the radio source Sgr A*, emits in other parts of the spectrum (such as infrared and X-ray). Also, the galactic center region may contain X-ray binaries emitting relativistic plasma jets, capable of producing high-energy gammas by either hadronic processes (such as π^0 production) or leptonic processes (such as inverse Compton scattering). Finally, this region may also contain supernova remnants, which are believed to produce galactic cosmic rays (including TeV-scale gamma rays).

Results from the HESS experiment are expected in the near future - with four telescopes, HESS is expected to be more sensitive in the direction of the galactic center and to have superior angular resolution.

1.5.2 Neutrinos

Although WIMPs are expected to scatter very infrequently (current upper limits for WIMP-nucleon scattering cross section reach 10^{-42} cm^2), they do scatter off of nuclei in the Sun or Earth, lose energy and become gravitationally bound. Hence, the density of WIMPs at the center of the Earth or the Sun can be considerably larger than in the halo, implying higher annihilation rates. Such effect could be significant even at the galactic center, if the density profile is sufficiently cuspy (or spiky), but the small solid angle suppresses the signal compared to the Sun or the Earth.

Neutrinos produced in such WIMP-annihilations would penetrate through the Earth, or escape from the Sun. The neutrinos can be produced both directly $XX \rightarrow \nu\bar{\nu}$ and indirectly $XX \rightarrow f\bar{f}$, where the fermion f can decay (or hadronize and decay) and emit a neutrino. Hence, the energy spectrum is expected to be continuous, rather than a line, but it is expected to extend up to the WIMP mass. Also, the neutrinos produced in annihilations in the Sun lose energy while escaping, which also causes the spectrum to be continuous.

If the neutrino interacts with the Earth-rock sufficiently close to the Earth surface, the products of the interaction may be detectable. The muon neutrinos are, hence, of the most interest, because their interactions produce muons which can travel considerable distance through the rock and reach a detector. Electron neutrinos are clearly not a good choice, because electrons are easily stopped in the rock. The tau neutrinos are produced in significantly smaller amounts than the muon or electron neutrinos, as the relevant decay processes are suppressed by the large mass of the tau.

The muon neutrinos can be detected at the surface of the Earth, usually using dedicated solar or atmospheric neutrino detectors. In particular, one searches for the upward-going muons - for the high-energy neutrinos, the produced muons are well collimated with the original neutrino direction. They also carry much of the original neutrino energy. Hence, one can search for the upward-going muon signal with a high-energy-threshold detector. The only known background are the atmospheric neutrinos produced in the interactions of cosmic rays with the atmosphere at the opposite side of the Earth - such neutrinos are of

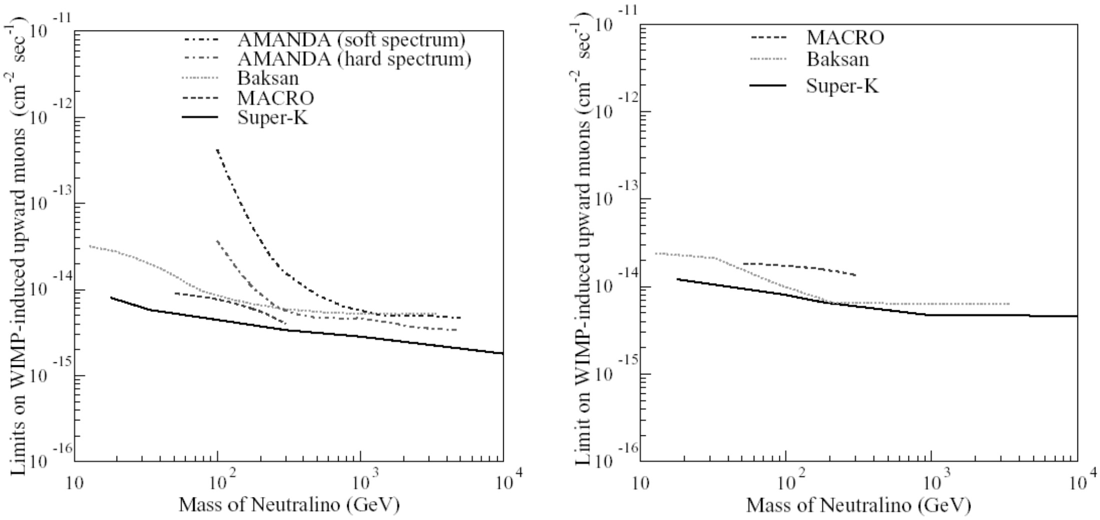


Figure 1.25: Upper limits on the upward-going muon flux due to WIMP annihilation in the Earth (left) and in the Sun (right) for different experiments. Figure taken from [85].

similar energies, and they can produce muons in the same way as the WIMP-annihilation neutrinos.

Hence, experiments designed to study solar or atmospheric neutrinos can also be used to look for the WIMP-annihilation neutrino signal. At the moment, none of the experiments has observed excess neutrinos from the Earth or the Sun, but several experiments have determined upper bounds on their flux: Baksan (neutrino experiment in Caucasus, Russia) [87], SuperKamiokande (atmospheric neutrino experiment in Japan) [85], MACRO (liquid scintillator neutrino experiment in Italy) [88], and AMANDA II (ice Čerenkov detector at the South Pole) [89]. Their limits are shown in Figure 1.25. These limits are just starting to probe the theoretically allowed region in supersymmetric WIMP models, as shown in Figure 1.26. Future experiments, such as ANTARES [90], Lake Baikal [91], as well as future runs of AMANDA II and IceCube, are expected to improve the sensitivity to the WIMP-annihilation neutrons by $\sim$2 orders of magnitude.

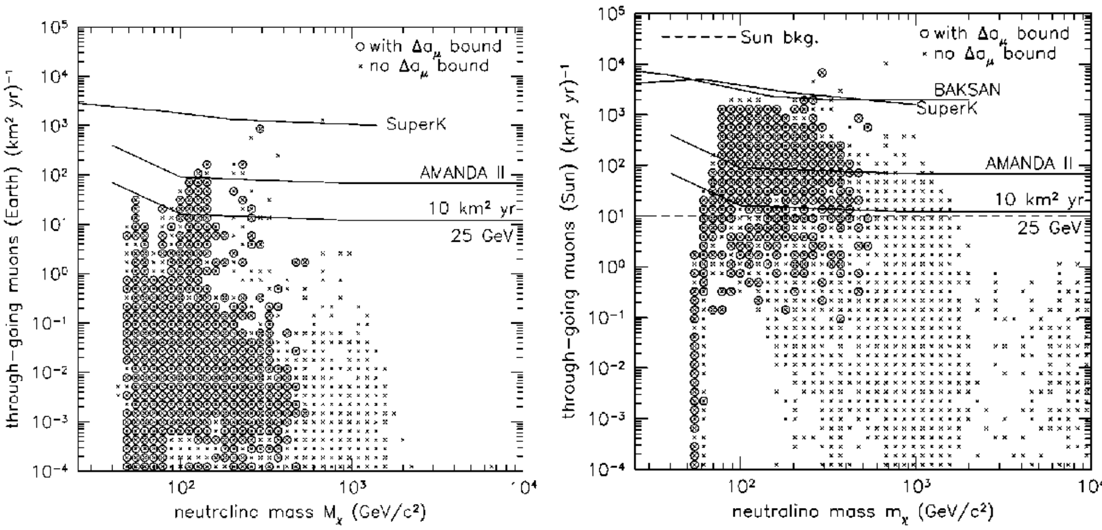


Figure 1.26: Comparison of the experimental limits on the upward-going muon flux from the Earth (left) and the Sun (right) and the theoretically allowed region for the supersymmetric WIMP models. The x's denote all allowed models, and o's denote the models that are also consistent with the recent measurement of the anomalous magnetic moment of muon. We will discuss these models in more detail in Chapter 2. Note that the SuperKamiokande limits are somewhat older than those appearing in the Figure 1.25. Figure taken from [86].

1.5.3 Antimatter

Apart from the photons and neutrinos, all other normal-matter particles produced in the WIMP annihilation would be very difficult to detect due to the very large background of normal-matter particles in the cosmic rays. However, the antimatter background is substantially smaller. Hence, detecting antimatter produced in WIMP-annihilation may actually be plausible.

The experiments searching for such particles are balloon or satellite based to avoid the atmosphere. They usually rely on a drift chamber in a magnetic field to determine the mass and charge of the particle. Furthermore, they need a calorimeter (e.g. a scintillator) to measure the energy spectrum of the particles.

The flux of antiprotons has been measured by the experiment BESS [92, 93]. BESS is a balloon-based experiment, and it performed a number of flights in 1995, 1997, 1999, and 2000. As shown in Figure 1.27, it observed a peak in the antiproton spectrum at 2 GeV, which is a characteristic feature of secondary antiprotons produced by cosmic-ray interactions in the interstellar gas. In other words, secondary antiprotons produced by cosmic-ray spallation are sufficient to explain the observed spectrum - no additional antiproton flux is needed. Although the uncertainty of the measurement will improve with time, the uncertainties in the production and propagation of these secondary antiprotons are significant, so it is not clear if antiproton flux measurement will be able to constrain the WIMP models [2]. Figure 1.27 also shows that the observations fit well the changes in the solar magnetic field.

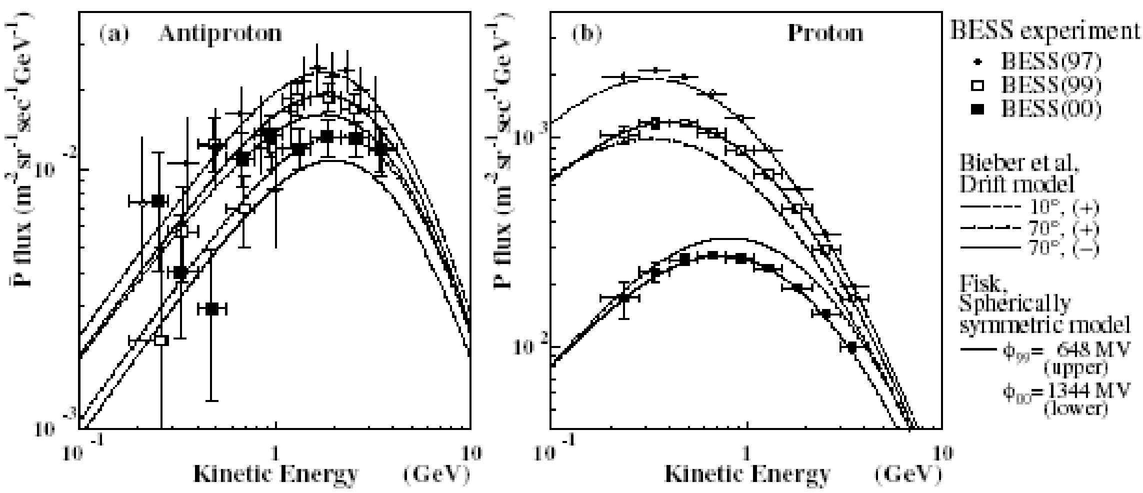


Figure 1.27: The antiproton and proton fluxes are shown, for three BESS operational years. The curves represent predictions of a drift model that takes into account the tilt angle of the Sun's magnetic polarity, corresponding to $10°(+)$ in 1997 (dashed), $70°(+)$ in 1999 (dot-dashed) and $70°(-)$ in 2000 (solid). The dotted lines correspond to a spherical model calculation with a single free parameter fitted for each year separately. Figure taken from [93].

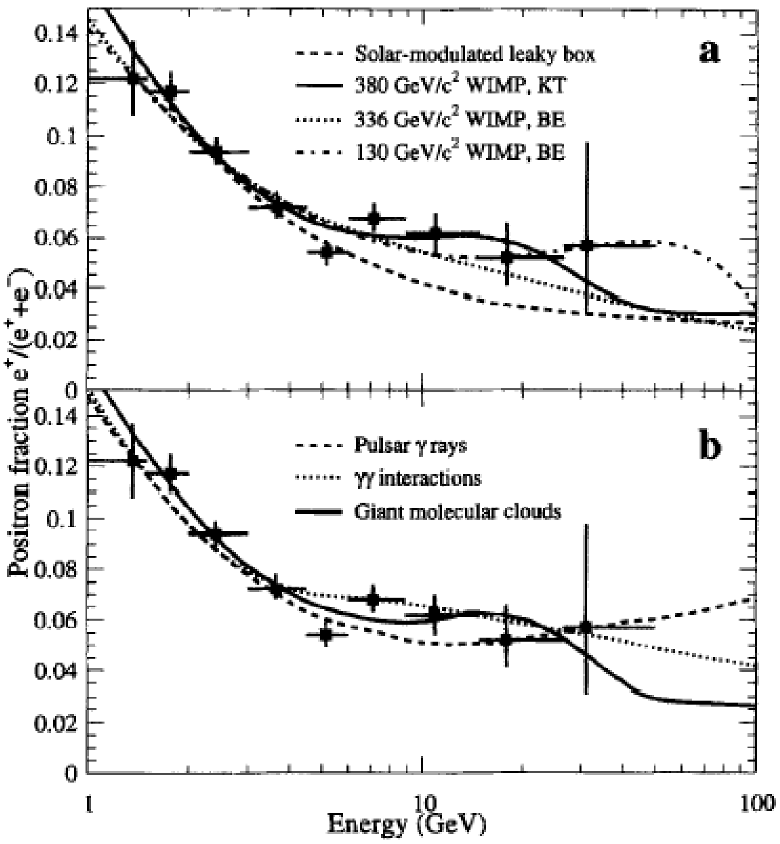


Figure 1.28: Positron spectrum measured by HEAT. The top figure shows a comparison with some WIMP models: KT and BE denote the models proposed by Kamionkowski and Turner [94] and Baltz and Edsjo [95]. The bottom figure shows some possible astrophysical explanations. Figure taken from [96].

The flux of positrons has been measured by the balloon-based experiment HEAT. In 1994 and 1995, HEAT measured an excess positron flux [97] compared to the predicted rate, as shown in the Figure 1.28. These results were confirmed by another HEAT flight in 2000 [96]. The source of these positrons is not known. It is possible to explain the observed bump at $\sim$ 20 GeV with additional astrophysical sources of positrons. It has also been argued that this excess can be explained by WIMP-annihilation.

Recently, D. Hooper *et al.* [98] tried to establish the plausibility of the excess of positrons being due to the WIMP-annihilation in the framework of supersymmetric WIMP models. In these models, positrons are produced usually in decays of gauge bosons produced in the WIMP-annihilation, although they can also be produced in cascades of other annihilation products. Hence, they are expected to have a continuous spectrum, whose shape can vary significantly depending on the mass and content of the WIMP. They find that the spectrum of positrons observed by HEAT can be fitted well with a dark-matter clump at some unknown distance, but they find that such solutions are unlikely and unnatural.

1.6 Direct WIMP-Detection Experiments

WIMPs can be detected directly by observing their (rare) interactions with the ordinary matter. The plausibility of this approach was first proposed by Goodman and Witten [99]. Since then, a number of experiments deployed different types of technology in search for the WIMP signal. In this Section, we first review the expected interaction rates and recoil-energy spectra, and then review the various technologies used so far in various experiments.

1.6.1 Recoil-Energy Spectra and Event Rates

WIMPs are expected to interact with the nucleons in the ordinary matter. The WIMP-nucleon elastic-scattering cross-section σ is model dependent, and we will discuss the theoretical expectations for the supersymmetric WIMP models in Chapter 2. Here, we establish the relationship between σ and the expected event-rates in the detector. We follow Lewin and Smith [100].

First, the WIMP distribution is given by

$$f(\vec{x},\vec{v},\vec{v}_E) \sim \exp\left(- \frac{M_W(v+v_E)^2/2 + M_W\phi(\vec{x})}{k_BT} \right), \tag{1.44}$$

where M_W and T are the WIMP mass and temperature, ϕ is the gravitational potential, and $\vec{v}$ is the WIMP velocity with respect to the Earth. The Earth velocity in the motionless halo is given by $\vec{v}_E$, and it is a sum of the Earth velocity around the Sun, the Sun velocity with respect to the galactic disk, and the velocity of the galactic disk itself. These components have been measured, so

$$v_E = 232 + 15\cos\left(2\pi\,\frac{t-152.5}{365.25}\right)\ \text{km/s}. \tag{1.45}$$

The sinusoidal behavior comes from the Earth velocity around the Sun, hence causing the annual modulation of the WIMP flux of $\sim \pm 6\%$. At a given location, the position dependent part of the WIMP distribution function is fixed, so the distribution reduces to a Maxwellian velocity distribution:

$$f(\vec{v},\vec{v_E}) \sim \exp\left(- \frac{|\vec{v}+\vec{v}_E|^2}{v_0^2} \right), \tag{1.46}$$

where v_0 is the characteristic velocity given by $k_BT = M_W v_0^2/2$. The characteristic velocity is usually set equal to the galactic rotation velocity [100], which is measured to be around 230 km/s. Hence, the WIMP density can be written as:

$$dn = \frac{n_0}{K}\, f(\vec{v},\vec{v}_E)\, d^3v, \tag{1.47}$$

where we defined

$$n_0 = \int_0^{v_{esc}} dn \tag{1.48}$$

$$K = \int_0^{2\pi} d\phi \int_{-1}^{1} d(\cos\theta) \int_0^{v_{esc}} f(\vec{v},\vec{v}_E) v^2 dv. \tag{1.49}$$

We have introduced the escape velocity v_{esc} - WIMPs with velocity larger than v_{esc} would escape the gravitational potential, so they should not contribute to the WIMP density. The escape velocity is usually taken to be 600 km/s, although significant variation in the estimates exists (depending on the modeling of the local group).

The event rate is the product of the number of target nuclei (mN_0/A, where m is the detector mass, N_0 is Avogadro's number and A is the atomic mass of the target nucleus), the incoming number flux of WIMPs (vn) and the WIMP-nucleus scattering cross-section (σ). For the moment, we consider the zero-momentum transfer, corresponding to σ_0 - we will come back to this later to include the non-zero momentum transfer. Differential rate per unit mass of the detector is:

$$\begin{aligned} dR &= \frac{N_0}{A}\,\sigma_0\,v\,dn \\ &= \frac{N_0}{A}\,\sigma_0\,v\,\frac{n_0}{k}\,f(\vec{v},\vec{v}_E)\,d^3v \\ &= R_0\,\frac{\sqrt{\pi}v}{2Kv_0}\,f(\vec{v},\vec{v}_E)\,d^3v. \end{aligned} \tag{1.50}$$

In the last line we defined

$$R_0 = \frac{2N_0}{\sqrt{\pi}A}\,n_0\,v_0\,\sigma_0. \tag{1.51}$$

One could integrate over the WIMP velocity and obtain the overall interaction rate, as discussed in [100]. However, we are more interested in deriving the differential recoil energy spectrum. The recoil energy of a nucleus of mass M_T, which is struck by a WIMP of energy E, and which is recoiling at an angle θ, is given by:

$$\begin{aligned} E_R &= E\,r\,(1-\cos\theta)/2 \\ r &= \frac{4M_W M_T}{(M_W+M_T)^2}. \end{aligned} \tag{1.52}$$

We can digress from the derivation briefly, to get a sense of the energies involved. The kinetic energy of a 50 GeV WIMP with the characteristic velocity $v_0 = 230$ km/s is about 15 keV. For a target nucleus of 73 GeV (such as Germanium), $r = 0.96$, so the recoil-energy of the nucleus is expected to be (roughly) in the 0-15 keV range. Clearly, lighter WIMPs would imply smaller recoil-energies.

We now assume that the scattering is isotropic (uniform in $\cos\theta$), so that the recoil energy is uniformly distributed in the $0 - Er$ range. To get the differential recoil-energy spectrum, we have to average over the diferential incident-energy spectrum E:

$$\frac{dR}{dE_r} = \int_{E_{min}}^{E_{max}} \frac{1}{Er} \frac{dR}{dE} \, dE, \tag{1.53}$$

where $E_{min} = E_R/r$ is the minimum energy that can give recoil energy E_R, and E_{max} corresponds to the escape velocity WIMPs. Changing variables from energy to velocity, and using Equation 1.50 integrated over angular components of velocity, gives

$$\begin{aligned} \frac{dR}{dE_R} &= \int_{v_{min}}^{v_{esc}} \frac{2}{rM_W v^2} \frac{dR}{dv} \, dv \\ &= \frac{R_0}{rE_0} \frac{2\pi^{3/2} v_0}{K} \int_{v_{min}}^{v_{esc}} v \, \exp\left(\frac{-(v+v_E)^2}{v_0^2} \right) dv, \end{aligned} \tag{1.54}$$

where $v_{min} = \sqrt{2E_R/rM_W}$ and $E_0 = M_W v_0^2/2$. In the limit $v_E \to 0$ and $v_{esc} \to \infty$, $K \to (\pi v_0^2)^{3/2}$ and the integral becomes simple:

$$\frac{dR}{dE_R} = \frac{R_0}{rE_0} \, e^{-E_R/rE_0} \tag{1.55}$$

This form, although incorrect, illustrates the typical exponential behavior of the recoil-energy spectrum. In particular, it shows that by observing the amplitude *and* the shape of the recoil-energy spectrum, it is possible to constrain two parameters: the WIMP mass and the elastic-scattering cross-section. The correct form of the recoil-energy spectrum, for non-trivial v_E and v_{esc} is [100]:

$$\begin{aligned}
\frac{dR(v_E,\infty)}{dE_R} &= \frac{R_0}{E_0 r}\,\frac{\sqrt{\pi}}{4}\,\frac{v_0}{v_E}\left[\operatorname{erf}\left(\frac{v_{min}+v_E}{v_0}\right)-\operatorname{erf}\left(\frac{v_{min}-v_E}{v_0}\right)\right] \\
k_0 &= (\pi v_0^2)^{3/2} \\
k_1 &= k_0\left[\operatorname{erf}\left(\frac{v_{esc}}{v_0}\right)-\frac{2}{\sqrt{\pi}}\,\frac{v_{esc}}{v_0}\,e^{-v_{esc}^2/v_0^2}\right] \\
\frac{dR(v_E,v_{esc})}{dE_R} &= \frac{k_0}{k_1}\left[\frac{dR(v_E,\infty)}{dE_R}-\frac{R_0}{E_0 r}\,e^{-v_{esc}^2/v_0^2}\right].
\end{aligned} \tag{1.56}$$

So far, the discussion was limitted to the zero-momentum transfer. When the momentum transfer $q=\sqrt{2M_T E_R}$ is large enough so that the wavelength h/q is comparable to the size of the nucleus, the cross-section begins to drop. This can usually be described by adding a multiplicative form factor. The form factor is usually determined empirically, and it is dependent on the type of the interaction (i.e. spin-dependent or spin-independent, see below). For the spin-independent interaction, which is usually dominant over the spin-dependent interaction, one can use the Helm form factor, which is given by:

$$\begin{aligned}
a &= 0.52 \text{ fm} \\
s &= 0.9 \text{ fm} \\
c &= 1.23A^{1/3}-0.60 \text{ fm} \\
r_n^2 &= c^2+\frac{7}{3}\pi^2 a^2-5s^2 \\
F(qr_n) &= 3\frac{j_1(qr_n)}{qr_n}\,e^{-(qs)^2/2} \\
&= 3\frac{\sin(qr_n)-qr_n\cos(qr_n)}{(qr_n)^3}\,e^{-(qs)^2/2}.
\end{aligned} \tag{1.57}$$

The cross-section is then given by

$$\sigma(q^2) = \sigma_0 F^2(qr_n). \tag{1.58}$$

In order to compare different experiments, and since different experiments use different target nuclei, it is preferable to report the recoil-energy spectrum referred to nucleons (e.g. proton) rather than referred to the target nucleus itself. If the WIMP-nucleon interaction is spin-independent, the contributions of various nucleons in the nucleus will be added coherently. Two corrections happen: the cross-section scales as $\mu_T = M_T M_W/(M_T + M_W)$ and the WIMP-nucleus coupling scales as A^2. Hence,

$$\sigma_{WIMP-Nucleus} = \sigma_{WIMP-p} \frac{\mu^2_{Nucleus}}{\mu^2_p} \ A^2. \tag{1.59}$$

If the WIMP-nucleon interaction is spin-dependent, the various nucleons will still coherently add, but paired nucleons will have opposite-sign contributions. Hence, essentially only the odd nucleon will contribute, modulo some spin factors which are nucleus-dependent [100]. Compared to the spin-independent case, it is clear that the A^2 scaling usually makes the spin-independent interaction dominant over the spin-dependent one.

Since the only place where σ_0 appears is in R_0 (Equation 1.51), and R_0 appears in both terms of $dR(v_E, v_{esc})/dE_R$, the recoil-energy spectrum can be written in the form:

$$\frac{dR(v_E, v_{esc})}{dE_R}|_{(T,q^2)} = \frac{dR(v_E, v_{esc})}{dE_R}|_{(p,0)} \ F^2(E_R) \ S, \tag{1.60}$$

where the (T, q^2) subscripts denote the WIMP interaction with the target nucleus at non-zero momentum transfer, and $(p, 0)$ denote the WIMP interaction with a proton at zero-momentum transfer, F^2 is the Helm form factor, and S is the scaling defined in Equation 1.59 ($= A^2\mu^2_{Nucleus}/\mu^2_p$ for the spin-independent interaction).

Finally, we mention that some experiments use target materials with more than one nucleus. Also, some experiments have different sensitivity to electron-recoil events and to nuclear-recoil events. Such complications can be dealt with [100], but we do not deal with them here.

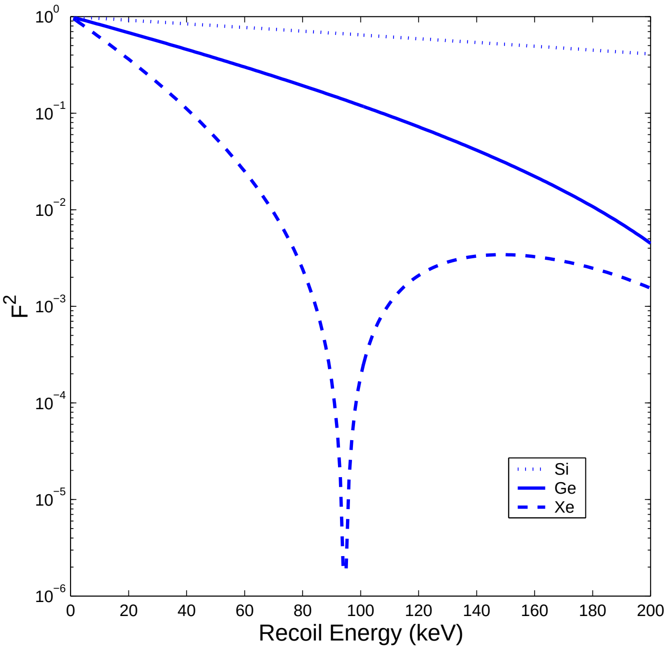


Figure 1.29: Helm form factor for Si, Ge, and Xe nuclei.

We conclude this derivation with the example calculations for several interesting target nuclei (Si ($A = 28$), Ge ($A = 73$), and Xe ($A = 131$)). We consider the spin-independent interaction of a WIMP of mass 100 GeV, and with $\sigma_{WIMP-p} = 10^{-42}$ cm^2 = 10^{-6} pb. Such value of σ_{WIMP-p} is currently interesting, as a number of experiments are begining to be sensitive to it. Figure 1.29 shows the Helm form factors for the three nuclei. Note that the form factor for Xe drops off quickly compared to Ge or Si, which partly compensates for the larger A^2 factor. Figure 1.30 shows the differential and integrated spectra

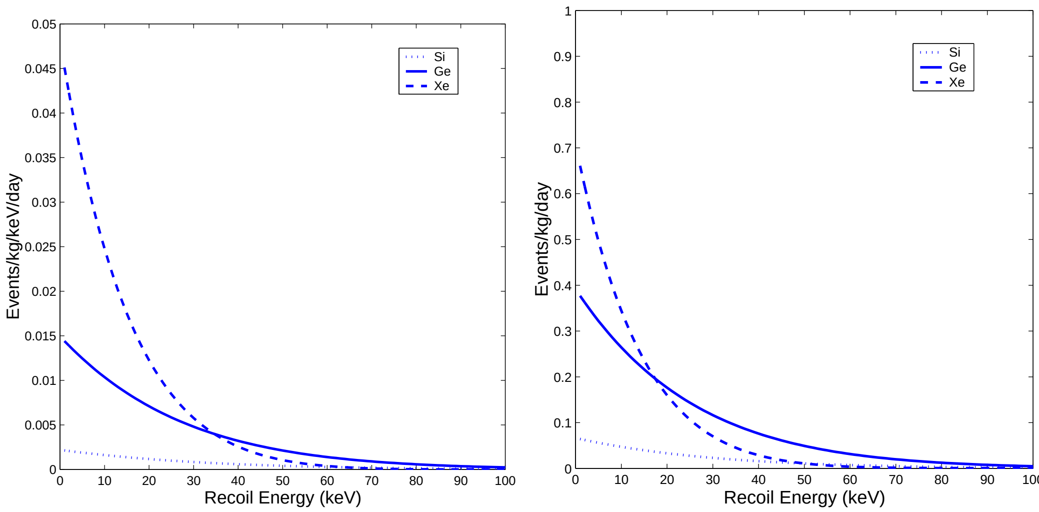


Figure 1.30: Differential (left) and integrated (right) recoil-energy spectra for the Si, Ge, and Xe nuclei.

corresponding to the three nuclei. Several important conclusions are apparent from these plots. First, the expected event rates are very low - even for the 10 keV experimental energy threshold, one expects $\lesssim 0.5$ events/kg/day. Hence, the main object of the direct detection experiments is understanding and suppressing various types of backgrounds. Second, due to the exponential nature of the spectrum, majority of the signal is at very low energies. Hence, low energy threshold is desirable. Third, despite the significant A^2 advantage of the Xe nucleus, the Helm form factor suppression makes it less optimal than Ge at energies $\gtrsim 20$ keV. In other words, Xe-based experiments would be more advantageous than Ge-based experiments if they can push the energy threshold significanly below 20 keV. Finally, Si targets are evidently much less sensitive to WIMPs than Ge or Xe ones.

1.6.2 Experimental Techniques

We now turn to various experimental techniques used in direct searches for WIMPs. Various upper limits on the WIMP-nucleon spin-independent elastic-scattering cross-section, obtained by different experiments discussed below (as of March 2004), are shown in Figure 1.31. Note that the experiments are just beginning to probe the theoretically allowed region in the supersymmetric WIMP models.

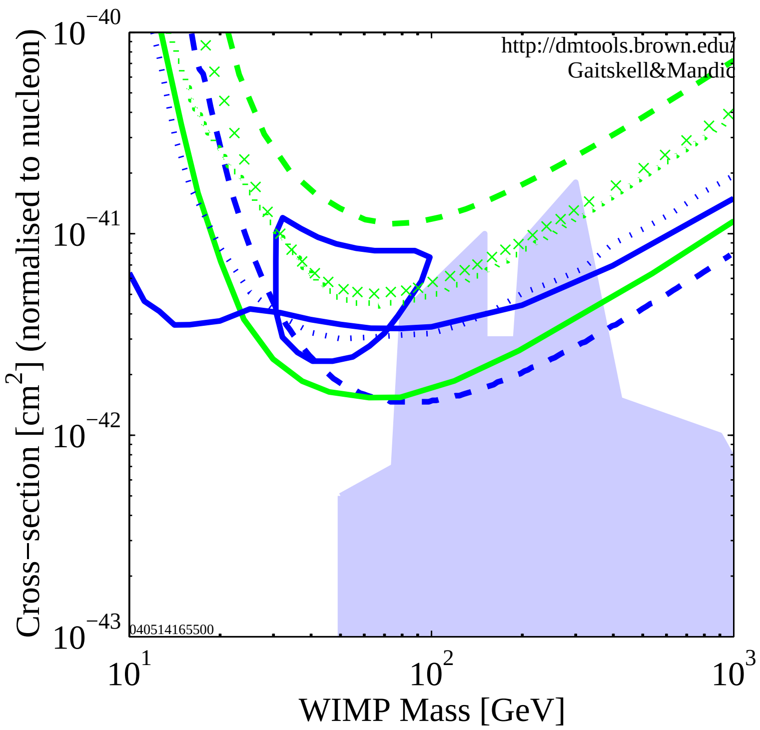


Figure 1.31: 90% C.L. upper limits of WIMP-nucleon spin-independent elastic-scattering cross-section are shown. Gray/green: DAMA 1996 exlusion limit (crosses) [101], preliminary limit from CRESST using $CaWO_4$ (private communication, dotted), Heidelberg-Moscow collaboration 1998 limit (dashed) [102], ZEPLIN I preliminary limit (private communication, solid). Black/blue: CDMS I limit (dotted) [103], CDMS II limit (solid) [104], Edelweiss limit (dashed) [105]. Black/blue closed contour is the 3σ signal region claimed by DAMA(1-4) [106]. The shaded region is an example of a calculation of a theoretically allowed region in the supesymmetric WIMP models [86].

Ge-Based Experiments

Ge-based detectors were the first detectors used in dark matter searches. Large Ge crystals were cooled to LN temperature (77 K). The approach was to use a strong electric field to separate and collect the electrons and holes created during a WIMP-nucleus interaction. The typical energy thresholds achieved were 5-10 keV, which after the ionization yield factor (ratio of the observed energy for nuclear and electron-recoil events of the same recoil-energy) implies 15-30 keV. This technique cannot discriminate against the electron-recoil events, which is the dominant background caused by gammas or betas. Hence, the sensitivity of such experiments is limitted by the cleanliness of the experimental setup and by the cosmogenically induced particles that penetrate the shielding of the apparatus. The first experiments applying this technique were the PNL-USC collaboration [107] and the UCSB/LBNL/UCB collaboration [108]. The most sensitive limits with this technique were achieved by the Heidelberg-Moscow collaboration (performed in Gran Sasso, Italy) [102] and IGEX (Canfranc, Spain) [109].

If cooled to substantially lower temperatures ($\sim 20-50$ mK) Ge crystals provide a second signature of an interaction - the phonon signal. The combination of the ionization signal and the phonon signal can be used to efficiently reject the electron-recoil background on the event-by-event basis. Hence, for such experiments, the electron-recoil background is significantly suppressed. As we will discuss for the case of CDMS II, the sensitivity is limited by the cosmogenically produced neutron background and by the electron-recoil rejection efficiency, (possibly in combination with cleanliness). CDMS I (at the shallow site at Stanford) [103] and Edelweiss (France) [105] detect the temperature changes due to the phonon signals using Ge NTD's, and CDMS II (shallow site at Stanford) [104] detects the athermal phonon signal using transition-edge-sensors. Up to March 2004, these experiments held the most sensitive limits on the WIMP-proton elastic scattering cross-section. We note that CDMS II also uses Si detectors - the difference in WIMP and neutron rates in the Si and Ge detectors allows statistical determination of the WIMP signal and of the neutron background.

NaI-Based Experiments

NaI crystals can be used as scintillators to search for WIMPs. In particular, the recoiling particle excites electrons to energy levels above their ground state, from which they decay by emitting a photon. Such photons are observed using photo-multiplier tubes (PMTs). It is possible to purify the crystals to achieve low levels of background. The scintillation yields are fairly low (0.3 for Na and 0.09 for I), leading to recoil-energy threshold of ~ 20 keV for I, which is the more interesting nucleus for the WIMP-search due to its large A^2 factor. It is also possible to discriminate against electron-recoil events using pulse-shape information, but such discrimination is not very effective at lowest energies. Furthermore, since it is easy to achive large masses, this type of detectors is ideally suited for searching for annual-modulation signal (i.e. the annual variation in the WIMP-nucleon scattering rate due to the change in the Earth velocity through the dark halo). Several experiments use this technology. DAMA (Gran Sasso, Italy) in 1996 established an upper limit using pulse-shape discrimination [101]. In 2000, using four new data-sets, DAMA observed an annual variation in the event rates that could be explained by the annual-modulation in the WIMP-proton elastic scattering [106].

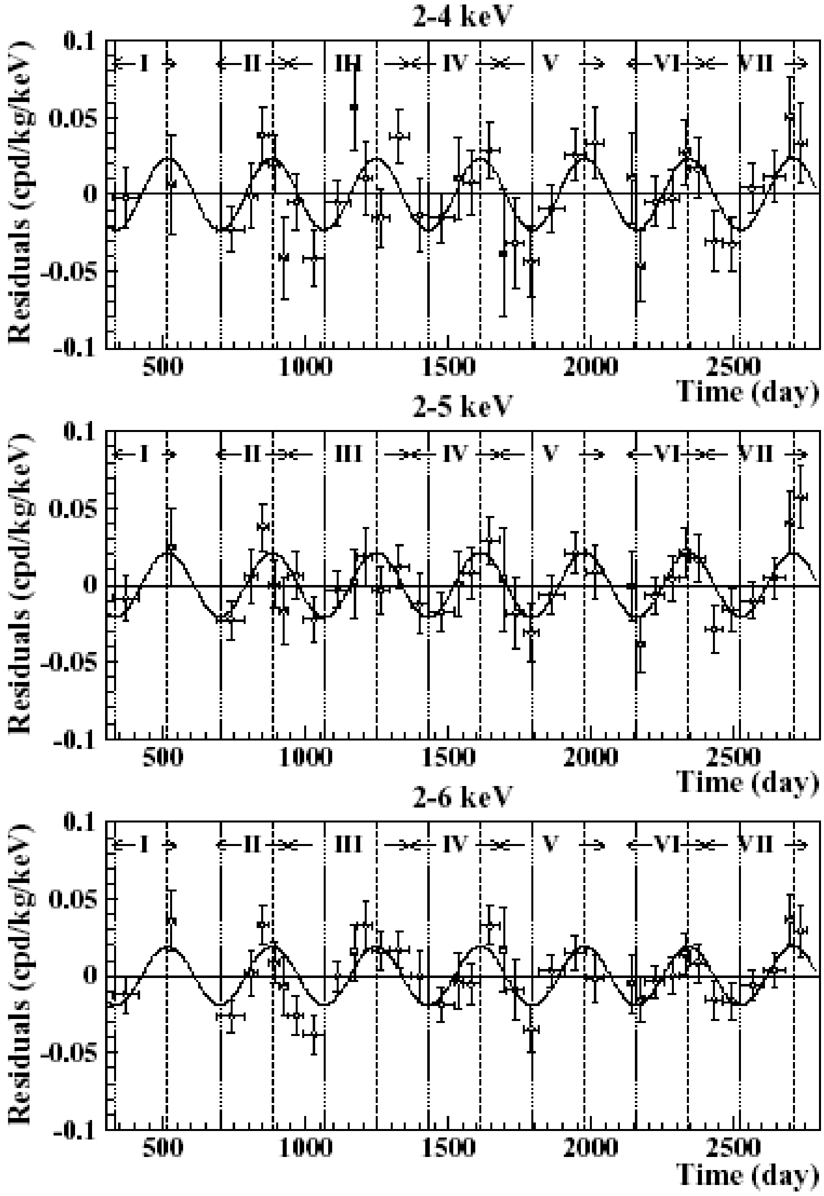


Figure 1.32: DAMA residual event rates are plotted against time for three energy bins. The residual rates were calculated from the measured rates by subtracting the constant part. The roman numbers denote the seven data-acquisition runs. The curves are the best-fits to the data points, with 1 year period and phase set at 2^{nd} of June. Figure taken from [110].

Recently, DAMA confirmed their claim by observing the anually-modulated signal over a seven-year period [110] - see Figure 1.32. While the DAMA evidence for the annual modulation is clear, its interpretation is more questionable. Most of the modulation signal comes from the lowest energy bins (2-6 keV, electron-equivalent), where understanding the efficiencies is particularly important. Although DAMA performed a study of the various possible systematic effects, some doubts remain that the signal is caused by another, less interesting, effect (such as, for example, the annual modulation in the muon flux, observed by MACRO). The doubts are further fueled by the fact that there are three experiments today (namely, CDMS (two detector-technologies), Edelweiss, and ZEPLIN I) which have explored the same parameter space and found no signal. Although there are systematic differences among these experiments (different target nuclei, total rate vs annual modulation etc), careful studies seem to indicate that such differences cannot account for the observed discrepancy.

Several upcoming experiments may help resolve the conflict. NaIAD (Boulby mine, UK) [111] has 65 kg of NaI crystals and is already acquiring data. ANAIS (Canfranc, Spain) [112] plans to run 107 kg of NaI, and has currently successfully tested a prototype. Both of these experiments should test the entire signal region claimed by DAMA in the coming 2-3 years. In addition, DAMA has upgraded its setup, under a new name (LIBRA). They plan to run 250 kg of NaI crystals, with higher-purity PMTs, stored underground for ~ 1 year. LIBRA is already operational.

Xe-Based Experiments

Liquid Xe can also be used to search for the direct WIMP signal. Xe has the obvious advantage over other materials due to its large A^2 factor. However, as shown in Figures 1.29 and 1.30, this advantage is partially compensated by the form factor suppression. In particular, the Xe based detectors would be particularly advantageous if their nuclear-recoil energy threshold can be reduced below 20 keV. The signature of an interaction in Xe is two-fold. First, there is electron excitation of Xe, which leads to scintillation. Second, there is ionization of Xe. The two signals can be used to discriminate against the electron-recoil events.

In absence of an electric field, the electron and ionized Xe recombine producing secondary scintillation. The timing of the two scintillation pulses differs for electron and nuclear recoils, so the pulse-shape discrimination can be used to suppress the electron-recoil background. This technique is used by ZEPLIN I (Boulby mine, UK) [113] - their preliminary result is comparable with the Ge-based experiments CDMS and Edelweiss, and incompatible with the signal region claimed by DAMA. However, they have not performed an *in situ* neutron calibration, which makes this preliminary result somewhat less reliable.

Alternatively, electric field can be used to extract the electrons of the ionization-signal. Such techniques are being investigated by ZEPLIN II and ZEPLIN III (both at the Boulby mine, UK).

Possibly the most important advantage of the Xe detectors is their scalability to large detector-masses. In particular, while it may be non-trivial to operate ~ 1 ton scale of a Ge-based experiment at 20 mK, it is relatively simple to increase the tank-size of liquid Xe. Such experiments (ZEPLIN IV and XENON) are in the proposal stage. However, the dependence of the basic parameters, the ionization yield and the scintilation yield, on the energy and on the type of recoil are yet to be demonstrated.

$CaWO_4$-Based Experiments

$CaWO_4$ crystals also yield two-fold signature of an interaction - the phonon and the scintillation signals. This technique has an important advantage over the Ge-based detectors in that it does not have surface-event problems (as we will see later, surface electron-recoil events in the Ge-based detectors can be mistaken for nuclear-recoil events, and are, therefore, an important source of background). However, the technique also has some difficulties. First, rather than using the PMTs to observe the scintillation signal, the current approach (taken by CRESST II [114]) is to use a second, phonon-mediated, detector adjacent to the primary detector. The light collection is relatively poor, resulting in an energy threshold of 15-20 keV. Second, there are three nuclei in the crystal, all of which could potentially interact with the WIMPs. The scintillation yield produced by the three nuclei is yet to be studied carefully. The goal of CRESST II is to build a 10 kg detector consisting of 300 g crystals.

Superheated Droplet Experiments

If a liquid is kept in a superheated state, a WIMP-nucleus interaction could trigger formation of a bubble, and an explosive phase transition can occur. The signal can be either measured using a piezo-electric sensor, or using a camera. The difficulty is that the detector has to be reset after each event, so the down-time can be rather large. The down-time can be reduced by operating in low-background environments (where event rates are low) and by building the experiment out of many modules so that only a small fraction of the experiment suffers the down-time after an interaction. Another difficulty is that only the integrated event rate can be measured, as there is no energy measurement. This difficulty may be overcome by using scintillating liquids, and such investigations are being considered. The clear advantage of this approach is the ease of scalability to 1 ton scale detector-mass. Picasso [115] uses fluorinated halocarbons in their setup at the bottom of the Sudbury mine (SNO site). This experiment is already running 40 g of effective detector mass and has first results.

CS_2-Based Experiment

The DRIFT experiment [113] uses a time-projection chamber (TPC) with CS_2 gas at 50 Torr and in a 1 kV/cm electric field. The free electrons in the ionization tracks combine with electron-negative CS_2, producing negative ions that then drift toward the end of the chamber where they are measured. This technique significantly reduces the diffusion of the tracks, without using a magnetic field. Furthermore, electron-recoil tracks extend further than the nuclear-recoil tracks, allowing electron-recoil discrimination. The clear advantage of this experiment is the measurement of the WIMP directionality, which is ideally suited for the annual modulation search. However, the gas must be kept at a low pressure, implying necessarily low detector mass. Currently, DRIFT is operating 1 m^3 prototype chamber, and plans exist for a 10 m^3 detector.

Chapter 2

Supersymmetry and Dark Matter

2.1 Introduction

The Standard Model of particle physics is certainly one of the most successful theories known today. Nevertheless, there are indications that the story of the Standard Model is not complete. In particular, the Standard Model is susceptible to the famous problem of quadratic divergences, which requires remarkable fine-tuning of parameters (1 in 10^{50}) in order to produce the particle masses we observe today. Besides the aesthetic side of this problem, it is also not easy to come up with a more complete theory (asymptoting the Standard Model at low energy), which would give a reasonable explanation of such fine tuning.

As noted in the Chapter 1, supersymmetry is currently regarded as one of the most promising extensions of the Standard Model for several reasons. First, the cancellation of the quadratic divergences (characteristic for many supersymmetric theories) offers a very natural (and elegant) solution of the fine-tuning problem. Second, supersymmetry naturally appears in the string theory, which is currently regarded as the best candidate for the Grand Unified Theory. Third, the Haag-Lopuszanski-Sohnius theorem states that the largest symmetry possessed by a unitary field theory is a direct product of some gauge group, Lorentz invariance, and possibly supersymmetry. The first two of these already appear in the Standard Model. It is, therefore, reasonable to ask whether supersymmetry can also be incorporated into a description of Nature.

Furthermore, there are experimental observations today that could fit very nicely with a supersymmetric extension of the Standard Model. For example, if the gauge couplings of the electromagnetic, the weak, and the strong nuclear force are run to high energies in the Standard Model alone, they do not converge to a single value. If, on the other hand, the running is done in the framework of the Minimal Supersymmetric Standard Model (discussed further below), the couplings unify at 2×10^{16} GeV. Another example is the measurement of the anomalous magnetic moment of muon (by the Brookhaven AGS experiment, also discussed below) which is not in good agreement with the Standard Model alone. This measurement can be easily explained in the MSSM framework.

From the point of view of dark matter, supersymmetry is interesting because it naturally produces a stable, heavy particle that interacts very weakly with the ordinary matter. Such a particle could account for the missing mass of the Universe. In the remaining part of this Chapter, we will first briefly review the construction of the MSSM. We will then focus on the neutralino, which is the lightest supersymmetric particle in most cases, and is the most likely dark matter candidate among the MSSM particles. We will study the mass and the interaction rate of the neutralino with the ordinary matter. In particular, we will try to answer two questions:

1. Which models are accessible to CDMS and other similar direct detection experiments?

2. Are there other experiments (e.g. accelerator-based) that could be complementary to CDMS?

2.2 Constructing Supersymmetric Theories

2.2.1 Supersymmetry Algebra

The detailed discussion of how to construct supersymmetric theories is given in [116]. Here, we outline the main steps. One starts from the supersymmetry algebra:

$$\begin{aligned} \{Q_\alpha, \bar{Q}_{\dot\beta}\} &= 2\sigma^m_{\alpha\dot\beta} P_m \\ \{Q_\alpha, Q_\beta\} &= \{\bar{Q}_{\dot\alpha}, \bar{Q}_{\dot\beta}\} = 0 \\ [P_m, Q_\alpha] &= [P_m, \bar{Q}_{\dot\alpha}] = 0 \\ [P_m, P_n] &= 0, \end{aligned} \tag{2.1}$$

where the Greek indices denote the two-component Weyl spinors (the dotted indices running over the conjugate components), the Latin indices denote the Lorentz four-vectors, P is the momentum operator, Q is the supersymmetry generator, and $\sigma^m_{\alpha\dot\beta}$ are the Pauli matrices. The first line of Equation 2.1 shows that the dimension of Q is $mass^{1/2}$, which is why the supersymmetry generator is sometimes thought of as square root of the momentum operator. This algebra can be rewritten entirely in terms of commutators only, using the anticommuting parameters ξ^α and $\bar{\xi_{\dot\alpha}}$:

$$[\xi Q, \bar\xi \bar{Q}] \quad = \quad 2\xi\sigma^m\bar\xi P_m$$

$$[\xi Q, \xi Q] = [\bar{\xi}\bar{Q}, \bar{\xi}\bar{Q}] = 0$$

$$[P^m, \xi Q] = [P^m, \bar{\xi}\bar{Q}] = 0, \tag{2.2}$$

where the summation convention is

$$\xi Q = \xi^\alpha Q_\alpha,\ \bar{\xi}\bar{Q} = \bar{\xi}_{\dot{\alpha}}\bar{Q}^{\dot{\alpha}}. \tag{2.3}$$

With this algebra in place, one can look for a multiplet of fields $(A, \psi \ldots)$ on which to define the infinitesimal transformations:

$$\delta_\xi A = (\xi Q + \bar{\xi}\bar{Q})\ A$$

$$\delta_\xi \psi = (\xi Q + \bar{\xi}\bar{Q})\ \psi \tag{2.4}$$

In particular, it can be shown that the following set of fields is closed (in the sense that no other field is required for the definition of infinitesimal transformations):

$$\begin{aligned}
\delta_\xi A &= \sqrt{2}\,\xi\,\psi \\
\delta_\xi \psi &= i\sqrt{2}\,\sigma^m\,\bar{\xi}\,\partial_m A + \sqrt{2}\,\xi\,F \\
\delta_\xi F &= i\sqrt{2}\,\bar{\xi}\,\bar{\sigma}^m\,\partial_m \psi.
\end{aligned} \tag{2.5}$$

Note that if A is a scalar, of mass dimension 1, it transforms into a spinor ψ of mass dimension $\frac{3}{2}$, which transforms into a tensor of even higher dimension (2) and a derivative of a lower dimension field.

2.2.2 Superfields

After defining the algebra, it is now necessary to find the relevant representations. For this purpose, we introduce superfields. Superfields are an elegant and compact way of describing supersymmetry representations.

Recall that the supersymmetry algebra can be rewritten in terms of commutators only (Equation 2.2). Hence, it can be viewed as a Lie algebra with group elements:

$$G(x,\theta,\bar{\theta}) = e^{i(-x^m P_m + \theta Q + \bar{\theta}\bar{Q})}. \tag{2.6}$$

Essentially, we have expanded the regular spacetime x^m into the "superspace" by introducing two anticommuting coordinates θ and $\bar{\theta}$. The differential operators in this space are:

$$\begin{aligned} D_\alpha &= \frac{\partial}{\partial\theta^\alpha} + i\sigma^m_{\alpha\dot{\alpha}}\,\bar{\theta}^{\dot{\alpha}}\,\partial_m \\ \bar{D}_{\dot{\alpha}} &= -\frac{\partial}{\partial\bar{\theta}^{\dot{\alpha}}} - i\theta^\alpha\,\sigma^m_{\alpha\dot{\alpha}}\,\partial_m. \end{aligned} \tag{2.7}$$

A superfield is defined to be a function in this superspace. We can expand such superfield in powers of θ and $\bar{\theta}$

$$\begin{aligned} F(x,\theta,\bar{\theta}) &= f(x) + \theta\phi(x) + \bar{\theta}\bar{\chi}(x) \\ &\quad + \theta\theta m(x) + \bar{\theta}\bar{\theta}n(x) + \theta\sigma^m\bar{\theta}v_m(x) \\ &\quad + \theta\theta\bar{\theta}\bar{\lambda}(x) + \bar{\theta}\bar{\theta}\theta\psi(x) + \theta\theta\bar{\theta}\bar{\theta}d(x). \end{aligned} \tag{2.8}$$

One can then proceed to define the infinitesimal transformations. Note that the linear combinations and products of superfields are again superfields, implying that the superfields form a representation of the supersymmetry algebra. In general, this is a reducible representation. The problem of finding irreducible representations, therefore, turns into imposing appropriate constraints on superfields. Two cases are important: chiral and vector superfields.

Chiral Superfields

Chiral superfields are defined to satisfy

$$\bar{D}_{\dot{\alpha}}\Phi = 0. \tag{2.9}$$

This equation can be solved by introducing a new variable $y^m = x^m + i\theta\sigma^m\bar{\theta}$ and θ, since $\bar{D}_{\dot{\alpha}}(y) = 0$ and $\bar{D}_{\dot{\alpha}}(\theta) = 0$. Hence, the most general chiral superfield can be written as:

$$\Phi = A(y) + \sqrt{2}\theta\psi(y) + \theta\theta F(y), \tag{2.10}$$

which can then be expanded in x. The most general Lagrangian (containing only chiral superfields) is extracted from the chiral superfields themselves:

$$\begin{aligned} \mathcal{L} &= \Phi_i^+\Phi_i \,|_{\theta\theta\bar{\theta}\bar{\theta}\ component} + \Big[\Big(\frac{1}{2}m_{ij}\Phi_i\Phi_j \\ &+ \frac{1}{3}g_{ijk}\Phi_i\Phi_j\Phi_k + \lambda_i\Phi_i\Big)\Big|_{\theta\theta\ component} + h.c.\Big] \\ &= i\partial_m\bar{\psi}_i\bar{\sigma}^m\psi_i + A_i^*\Box A_i + F_i^*F_i \\ &+ \Big[m_{ij}\Big(A_iF_i - \frac{1}{2}\psi_i\psi_j\Big) \\ &+ g_{ijk}(A_iA_jF_k - \psi_i\psi_jA_k) + \lambda_iF_i + h.c.\Big], \end{aligned} \tag{2.11}$$

where $\Box$ denotes the four-dimensional Laplace operator. Finally, the field F is clearly auxiliary and it can be removed from the Lagrangian using its Euler equations. In other words, the Lagrangian can be rewritten entirely in terms of the scalar field A and spinor ψ.

2.2.3 Vector Superfields

The vector superfields obey

$$V = V^+. \tag{2.12}$$

Again, we expand the superfield in powers of θ and $\bar{\theta}$:

$$\begin{aligned} V(x,\theta,\bar{\theta}) &= C(x) + i\theta\chi(x) - i\bar{\theta}\bar{\chi}(x) \\ &+ \frac{i}{2}\theta\theta\Big[M(x) + iN(x)\Big] - \frac{i}{2}\bar{\theta}\bar{\theta}\Big[M(x) - iN(x)\Big] \\ &- \theta\sigma^m\bar{\theta}v_m(x) + i\theta\theta\bar{\theta}\Big[\bar{\lambda}(x) + \frac{i}{2}\bar{\sigma}^m\partial_m\chi(x)\Big] \\ &- i\bar{\theta}\bar{\theta}\theta\Big[\lambda(x) + \frac{i}{2}\sigma^m\partial_m\bar{\chi}(x)\Big] + \frac{1}{2}\theta\theta\bar{\theta}\bar{\theta}\Big[D(x) + \frac{1}{2}\Box C(x)\Big]. \end{aligned} \tag{2.13}$$

The component fields are chosen in this particular way because fixing the gauge becomes particularly simple. We define the (generalization of) gauge transformation:

$$V \rightarrow V + \Phi + \Phi^+, \tag{2.14}$$

where Φ and Φ^+ are chiral superfields. By working out the transformation of component fields of V, one can show that C, χ, M, and N can be set to zero. Then, the vector superfield becomes:

$$V = -\theta\sigma^m\bar{\theta}v_m(x) + i\theta\theta\bar{\theta}\bar{\lambda}(x) - i\bar{\theta}\bar{\theta}\theta\lambda(x) + \frac{1}{2}\theta\theta\bar{\theta}\bar{\theta}D(x). \tag{2.15}$$

In other words, we have a vector field v and its corresponding fermionic field λ. D will turn out to be auxiliary. To construct the Lagrangian, define

$$W_\alpha = -\frac{1}{4}\bar{D}\bar{D}D_\alpha V, \tag{2.16}$$

$$\bar{W}_{\dot{\alpha}} = \frac{1}{4}DD\bar{D}_{\dot{\alpha}}V, \tag{2.17}$$

and note that these superfields are chiral. The Lagrangian can then be formed from the $\theta\theta$ component of $W^\alpha W_\alpha$:

$$\mathcal{L} = \frac{1}{4}\Big(W^\alpha W_\alpha\,|_{\theta\theta} + \bar{W}_{\dot{\alpha}}\bar{W}^{\dot{\alpha}}\,|_{\bar{\theta}\bar{\theta}}\Big). \tag{2.18}$$

After integrating by parts, this becomes the usual gauge boson Lagrangian:

$$\int d^4x \mathcal{L} = \int d^4x \Big(\frac{1}{2} D^2 - \frac{1}{4} v^{mn} v_{mn} - i\lambda\sigma^m \partial_m \bar{\lambda} \Big), \tag{2.19}$$

where $v_{mn} = \partial_m v_n - \partial_n v_m$.

2.3 Minimal Supersymmetric Standard Model

We can apply the construction described in the previous Section to obtain the smallest supersymmetric extension of the Standard Model of particles. The Minimal Supersymmetric Standard Model (MSSM), by definition, contains the minimal number of particles necessary to incorporate the whole SM. Hence, each particle in the SM corresponds to a superfield (containing the particle and its supersymmetric partner) in the MSSM. Unfortunately, the cancellation of quadratic divergences does not occur in models with a single Higgs field. It is necessary to include (at least) two Higgs fields. The full construction of supersymmetric Lagrangians can be found in [116], [117], [118] or [119]. With this procedure, the Lagrangian of the MSSM can be compactly written in the following form:

$$\begin{aligned} \mathcal{L} \;=\; & -D_m A_i^* \, D^m A_i \;-\; i\overline{\chi}_i \, \overline{\sigma}^m \, D_m \chi^i \\ & -\tfrac{1}{4} \, F_{mn}^{(a)} \, F^{mn(a)} - \; i\overline{\lambda}^{(a)} \overline{\sigma}^m \, D_m \lambda^{(a)} \\ & - \, i\sqrt{2} \, g \overline{\lambda}^{(a)} \overline{\chi}_i \, T^{(a)i}{}_j \, A^j + \; i\sqrt{2} \, g \, A_j^* \, T^{(a)j}{}_i \chi^i \lambda^{(a)} \\ & -\tfrac{1}{2} \tfrac{\partial^2 P(A)}{\partial A^i \partial A^j} \chi^i \chi^j - \tfrac{1}{2} \big(\tfrac{\partial^2 P(A)}{\partial A^i \partial A^j} \big)^* \overline{\chi}^i \overline{\chi}^j \\ & - \mid \tfrac{\partial P(A)}{\partial A_i} \mid^2 - \tfrac{1}{2} \, g^2 \, (A_j^* \, T^{(a)j}{}_i \, A_i)^2 \end{aligned} \tag{2.20}$$

with the following definitions:

- There are seven chiral superfields: $\widehat{Q}$, $\widehat{\overline{U}}$, $\widehat{\overline{D}}$, $\widehat{L}$, $\widehat{\overline{E}}$, $\widehat{H_1}$, $\widehat{H_2}$, corresponding to quark doublet, up and down quark singlets, lepton doublet, lepton singlet, and two Higgs fields, respectively. Each of these fields contains a scalar field A and a (Dirac) fermionic field χ. These appear in the same representations as in the SM.

- There are three vector superfields, corresponding to SU(3), SU(2) and U(1) gauge bosons in SM, each containing the appropriate (SM) bosonic gauge field and the corresponding (Majorana) fermionic gaugino field λ.

- σ^m are Pauli matrices, $T^{(a)}$ are generators of the appropriate gauge symmetry groups, D_m is the appropriate covariant derivative, F_{mn} is the usual field-strength tensor and P is the superpotential defined by

$$P = \mu \widehat{H_1} \widehat{H_2} + h_U \widehat{Q} \widehat{\overline{U}} \widehat{H_2} + h_D \widehat{Q} \widehat{\overline{D}} \widehat{H_1} + h_E \widehat{L} \widehat{\overline{E}} \widehat{H_1}, \tag{2.21}$$

where h's are the Yukawa couplings. Notice that the superpotential is defined to be R-parity invariant, where $R = (-1)^{3(B-L)+2S}$ (B and L are baryon and lepton numbers and S is spin). With this definition, $R = 1$ for ordinary particles and $R = -1$ for their superpartners. Furthermore, in the last two lines of Equation 2.20, only scalar parts of the superpotential are differentiated.

- In Equation 2.20 we implicitly sum over all gauge bosons and gauginos (via the index a) and fermions and sfermions (via the indices i and j), taking into account the compatibility of their representations and using the appropriate gauge coupling constants g.

The given Lagrangian is by construction invariant under supersymmetry transformations, so it is in clear disagreement with the experimental results. The only solution is to break the supersymmetry, which is achieved by assuming that MSSM is only an effective theory and that there are possible terms in the Lagrangian which are not supersymmetry invariant. These additional terms must preserve the cancellation of quadratic divergences, which is why they are called 'soft'. The most general soft symmetry-breaking terms have been found by Girardello and Grisaru in [120]. For the MSSM they are as follows (in two component notation):

$$\begin{aligned} V_{soft} \;=\; & \epsilon_{ij}(\widetilde{e}^*_R A_E h_E \widetilde{l}^i_L H^j_1 + \widetilde{d}^*_R A_D h_D \widetilde{q}^i_L H^j_1 \\ & -\widetilde{u}^*_R A_U h_U \widetilde{q}^i_L H^j_2 - B\mu H^i_1 H^j_2 + h.c.) \\ & +\widetilde{q}^{i*}_L M^2_Q \widetilde{q}^i_L + \widetilde{l}^{i*}_L M^2_L \widetilde{l}^i_L + \widetilde{u}^*_R M^2_U \widetilde{u}_R + \widetilde{d}^*_R M^2_D \widetilde{d}_R + \widetilde{e}^*_R M^2_E \widetilde{e}_R \\ & +H^{i*}_1 m^2_1 H^i_1 + H^{i*}_2 m^2_2 H^i_2 \\ & +\tfrac{1}{2}(M_1 \widetilde{B}\widetilde{B} + M_2 \widetilde{W}^a \widetilde{W}^a + M_3 \widetilde{g}^b \widetilde{g}^b), \end{aligned} \tag{2.22}$$

where i and j are SU(2) indices, $\epsilon_{12} = 1$, λ's are the Yukawa couplings, A's are trilinear couplings, and M's are the appropriate mass matrices for squarks and gauginos. B is a soft bilinear coupling and L and R refer to the chirality of the fields.

In the Higgs sector, both neutral Higgs bosons get vacuum expectation values. As in the SM, three degrees of freedom are Goldstone bosons, and the remaining five are: two charged Higgs fields (H^+ and H^-), one neutral parity odd field (A) and two neutral parity even fields (heavier H^0 and lighter h^0) (for more detail see, for example, [121]). In addition, the charged SU(2) gauginos and charged higgsinos have the same quantum numbers, so they can mix and form 'charginos'. Similarly, the neutral wino, bino and two (parity even) neutral higgsinos also have the same quantum numbers and they can mix. These states are called neutralinos.

2.4 LSP as a Dark Matter Candidate

Having defined the MSSM, we can now examine the implications of this model on dark matter. Clearly, if any of the particles in the MSSM is to be a candidate for 'cold dark matter', it must be stable, i.e. it should not decay into lighter particles. The lightest of the newly introduced supersymmetric partners is stable, since the R-parity forbids its decay into the ordinary particles. The question, therefore, is - which particle could be the Lightest Supersymmetric Particle (LSP).

Any charged candidate would have been discovered in the searches for anomalously heavy protons. The electrically neutral hadrons, formed from squarks and gluinos are a possibility. However, the Grand-Unified models predict relations between the masses of supersymmetric particles. The usual outcome is that the gluino is heavier than the neutralino and squarks are heavier than sleptons, which effectively rules out the possibility of electrically neutral hadrons. Another possible candidate is a sneutrino. However, in most cases there is a slepton slightly lighter than a sneutrino. As a result, neutralino is the most favored possibility.

As mentioned above, the neutralinos are mixtures of the bino ($\widetilde{B}$), wino ($\widetilde{W}$), and two higgsino ($\widetilde{H}_1^0$ and $\widetilde{H}_2^0$) states. In this basis, the neutralino mass matrix is given by:

$$\mathbf{M} = \begin{pmatrix} M_1 & 0 & -m_Z s_\theta c_\beta & +m_Z s_\theta s_\beta \\ 0 & M_2 & +m_Z c_\theta c_\beta & -m_Z c_\theta s_\beta \\ -m_Z s_\theta c_\beta & +m_Z c_\theta c_\beta & \delta_{33} & -\mu \\ -m_Z s_\theta s_\beta & -m_Z c_\theta s_\beta & -\mu & \delta_{44} \end{pmatrix} \tag{2.23}$$

where $s_\theta = \sin\theta_W, c_\beta = \cos\beta$ etc, $\tan\beta = v_2/v_1$ is the ratio of the vacuum expectation values for the two Higgs bosons, $M_{1,2}$ are the gaugino masses (introduced in the soft supersymmetry-breaking terms) and m_Z is the mass of the Z boson. The δ's are radiative corrections important when two higgsinos are close in mass. They can be found in [122]. The eigenvalues of this matrix are the masses of the four physical neutralino states. The neutralino states are the corresponding eigenvectors and they are given by:

$$\widetilde{\chi}_i^0 = \widetilde{\chi} = N_{i1}\widetilde{B} + N_{i2}\widetilde{W}_3 + N_{i3}\widetilde{H}_1^0 + N_{i4}\widetilde{H}_2^0 \tag{2.24}$$

where N's are the mixing coefficients. Furthermore, we define the gaugino fraction of the neutralino by

$$z_g = N_{i1}^2 + N_{i2}^2 \tag{2.25}$$

and then define the following:

- If $z_g < 0.01$ the model is considered higgsino-like.
- If $z_g > 0.99$ the model is considered gaugino-like (generally bino-like due to renormalization group equations (RGEs)).
- If $0.01 < z_g < 0.99$ the model is considered mixed.

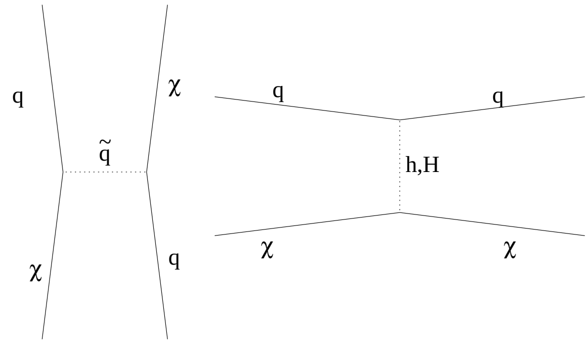


Figure 2.1: The leading diagrams for direct detection. Note that there is also a u-channel diagram for squark exchange. There are also diagrams where the neutralino scatters off of gluons in the nucleon through heavy squark loops.

Note the importance of the parameters μ and M_2 (M_1 is related to M_2 by the RGEs): they set the masses of the four neutralino states (up to small corrections) and they determine the content of the lightest neutralino state. In particular, $|\mu| >> |M_2|$ and $|M_2| >> |\mu|$ yield the pure bino and the pure higgsino models, respectively, while the case $|\mu| \sim |M_2|$ corresponds to the mixed models.

As discussed in the Section 1.6.1, the spin-independent WIMP-nucleus interaction is dominant for the Ge and Si target nuclei, used by the CDMS experiment. We will discuss this interaction in detail in the Section 2.6 (detailed discussions can also be found in [123] and [124]), and here we mention several important points.

- The calculation of the elastic-scattering cross-section is done in several steps. First, one calculates the cross-sections for interactions of neutralino with quarks and gluons. Then, using the matrix elements of the quark and gluon operators in a nucleon state, one evaluates the cross-sections for the neutralino-nucleon interaction. Finally, one must add the calculated components coherently to get the neutralino-nucleus elastic-scattering cross-section - as discussed in the Section 1.6.1, this addition is different for the spin-dependent and the spin-independent cases. Each of these steps introduces uncertainties (particularly the last two), which makes the calculation less reliable. The last step is usually omitted and the neutralino-nucleon cross-section is usually quoted, for a given theoretical model. However, the last step must be performed in order to compare with the experimental measurements - usually, it is the experimental measurements of event rates in various target materials that are converted to the WIMP-nucleon cross-section.

- The spin-independent elastic scattering occurs through two main channels (shown in the Figure 2.1). First, there is scattering via the Higgs boson exchange. By examining the Equation 2.20 it is clear that vertices $\widetilde{B}\widetilde{B}H$ and $\widetilde{H}\widetilde{H}H$ (where H is a parity even Higgs boson) do not exist, but $\widetilde{B}\widetilde{H}H$ does. Second, there is elastic scattering via squark exchange. In this case, $\widetilde{B}q \to \widetilde{q} \to \widetilde{B}q$ and $\widetilde{H}q \to \widetilde{q} \to \widetilde{H}q$ are spin-dependent (and thus suppressed), but $\widetilde{B}q \to \widetilde{q} \to \widetilde{H}q$ is spin-independent. Both of these channels, therefore, favor the mixed neutralino models (compared to pure bino or pure higgsino). Since the neutralino content is determined by the parameters μ and M_2 (as mentioned above), these parameters have a crucial role in determining the neutralino-nucleon elastic scattering cross-section.

- Since the Higgs boson mass is usually smaller than the squark masses, the Higgs boson exchange usually dominates the spin-independent interaction. Furthermore, since the lightest Higgs boson mass is expected to be $\sim$ 120 GeV (i.e. slightly heavier than the Z boson), the spin-independent elastic scattering cross-section is expected to be of the order of the weak scale.

- In addition to the tree-level diagrams discussed above, there are also several one-loop diagrams allowing scattering of neutralinos off of gluons. The importance of these channels depends on the details of the model.

At this point we can examine the cosmological significance of the lightest neutralino.

2.5 Relic Density

The neutralino relic density is determined by the neutralino annihilation cross-section and by the expansion rate of the Universe. Intuitively, larger annihilation cross-section implies that more neutralinos would be annihilated by the freeze-out, and, hence, the relic density today would be smaller. The neutralino relic density has been calculated in the past to various degrees of accuracy. As argued by Griest and Seckel [125] and Edsjo and Gondolo [126], the coannihilation processes must be taken into account in these calculations. The coannihilations are annihilations of neutralinos with slightly heavier particles, for example squarks or charginos. Coannihilation into charginos is particularly important since there is a mass degeneracy between the lightest two neutralinos and the ligthest chargino (coannihilation into squarks happens only accidentally when the squark mass is close to neutralino mass). Occasionally, coannihilations with the lightest sfermion (usually stau) are also important. We reproduce the calculation of the relic density given in [126].

Consider annihilation of N supersymmetric particles χ_i with masses m_i and internal degrees of freedom g_i. Assume also that the masses obey $m_1 < m_2 < ... < m_N$. Then, the abundance of the N particles is governed by a system of N Boltzmann equations:

$$\begin{aligned}
\frac{dn_i}{dt} &= -3Hn_i - \sum_{j=1}^{N} \langle \sigma_{ij} v_{ij} \rangle \, (n_i n_j - n_i^{eq} n_j^{eq}) \\
&\quad - \sum_{j \neq i} [\langle \sigma'_{Xij} v_{ij} \rangle \, (n_i n_X - n_i^{eq} n_X^{eq}) - \langle \sigma'_{Xij} v_{ij} \rangle \, (n_X n_j - n_X^{eq} n_j^{eq})] \\
&\quad - \sum_{j \neq i} [\Gamma_{ij}(n_i - n_i^{eq}) - \Gamma_{ji}(n_j - n_j^{eq})]
\end{aligned} \tag{2.26}$$

The first term on the right hand side depicts the expansion of the Universe (H is the Hubble parameter), the second describes the $\chi_i \chi_j$ annihilations with cross section

$$\sigma_{ij} = \sum_X \sigma(\chi_i \chi_j \to X), \tag{2.27}$$

the third term describes $\chi_i X \to \chi_j X'$ scattering with cross section

$$\sigma'_{ij} = \sum_{X'} \sigma(\chi_i X \to \chi_j X'), \tag{2.28}$$

and the fourth term represents decays of χ_i with decay rate

$$\Gamma_{ij} = \sum_X \sigma(\chi_i \to \chi_j X). \tag{2.29}$$

In the above expressions, X and X' are any standard model particles (possibly sets), v_{ij} is the relative velocity and n^{eq} is the equilibrium number density of the particle χ_i given by

$$n_i^{eq} = \frac{g_i}{(2\pi)^3} \int d^3 p_i f_i \tag{2.30}$$

with f_i being the equilibrium distribution function, in Maxwell-Boltzmann approximation given by

$$f_i = e^{-E_i/T}. \tag{2.31}$$

Assuming that all of the heavier particles decay eventually into the lightest one, we get that the final density n of the LSP to be the sum of n_i's. Then, summing the Boltzmann equations (Equation 2.26) we get

$$\frac{dn}{dt} = -3Hn - \sum_{i,j=1}^{N} \langle\sigma_{ij}v_{ij}\rangle \left(n_i n_j - n_i^{eq} n_j^{eq}\right) \tag{2.32}$$

(the last two terms in Equation 2.26 cancel out in the summation). Furthermore, assuming that the scattering off of the thermal background is much faster than the annihilation (since the background density is much larger than the density of each particle individually) implies $n_i/n = n_i^{eq}/n^{eq}$, so

$$\frac{dn}{dt} = -3Hn - \langle\sigma_{eff}v\rangle \left(n^2 - n_{eq}^2\right) \tag{2.33}$$

where we define

$$\langle\sigma_{eff}v\rangle = \sum_{ij}\langle\sigma_{ij}v_{ij}\rangle \, \frac{n_i^{eq}}{n^{eq}} \, \frac{n_j^{eq}}{n^{eq}}. \tag{2.34}$$

The equilibrium density is given by

$$\begin{aligned} n^{eq} &= \sum_i \frac{g_i}{(2\pi)^3} \int d^3p_i e^{-E_i/T} \\ &= \tfrac{T}{2\pi^2} \sum_i g_i m_i^2 K_2\left(\frac{m_i}{T}\right) \end{aligned} \tag{2.35}$$

where K_2 is the modified Bessel function of the second kind of order 2. The numerator in the Equation 2.34 is given by

$$\begin{aligned} \sum_{ij}\langle\sigma_{ij}v_{ij}\rangle \, n_i^{eq} n_j^{eq} \;&=\; \sum_{ij}\frac{g_i g_j}{(2\pi)^6} \int d^3p_i \; d^3p_j \; f_i f_j \; \sigma_{ij} \; v_{ij} \\ = \sum_{ij} \int W_{ij} \; &\frac{g_i \; f_i \; d^3p_i}{2(2\pi)^3 E_i} \; \frac{g_j \; f_j \; d^3p_j}{2(2\pi)^3 E_j} \end{aligned} \tag{2.36}$$

where W_{ij} is related to the unpolarized cross section through

$$W_{ij} = 4E_i E_j \sigma_{ij} v_{ij} \tag{2.37}$$

The integral in Equation 2.36 can be evaluated by changing integration variables into $E_+ = E_i + E_j$, $E_- = E_i - E_j$ and $s = m_i^2 + m_j^2 + 2E_iE_j - 2|\mathbf{p_i}||\mathbf{p_j}|\cos\theta$, where θ is the relative angle between the momenta 3-vectors. After performing integrations over E_- and E_+ this becomes

$$\sum_{ij} \langle \sigma_{ij} v_{ij} \rangle \, n_i^{eq} n_j^{eq} \;=\; \frac{T}{32\pi^4} \sum_{ij} \int_{(m_i+m_j)^2}^{\infty} ds \, \frac{v_{ij} E_i E_j}{\sqrt{s}} \, g_i g_j W_{ij} K_1 \left(\frac{\sqrt{s}}{T} \right) \tag{2.38}$$

where K_1 is the modified Bessel function of the second kind and of order 1. Define

$$W_{eff} = \sum_{ij} \frac{v_{ij} E_i E_j}{v_{11} E_1 E_1} \, \frac{g_i g_j}{g_1^2} \, W_{ij}, \; p_{eff} = \frac{v_{11} E_1^2}{\sqrt{s}} \tag{2.39}$$

Then using Equation 2.35 and substituting into Equation 2.34 we get

$$\langle \sigma_{eff} v \rangle = \frac{\displaystyle\int_0^{\infty} dp_{eff} \, p_{eff}^2 \, W_{eff} \, K_1 \left(\frac{\sqrt{s}}{T} \right)}{m_1^4 T \left[\displaystyle\sum_i \frac{g_i}{g_1} \, \frac{m_i^2}{m_1^2} \, K_2 \left(\frac{m_i}{T} \right) \right]^2}. \tag{2.40}$$

Furthermore, it is convenient to define $Y = \frac{n}{S}$, where S is the entropy density. Assuming that there is no production of entropy, SR^3 is constant, so $\frac{dS}{dt} = -3HS$. Then,

$$\frac{dY}{dt} = \frac{1}{S}\frac{dn}{dt} + 3H\frac{n}{S}. \tag{2.41}$$

Defining $x = m_1/T$, the Equation 2.33 becomes

$$\frac{dY}{dx} = -\frac{m_1}{3Hx^2} \, \frac{dS}{dT} \, \langle \sigma_{eff} v \rangle \, (Y^2 - Y_{eq}^2). \tag{2.42}$$

Next, using the Friedmann equation for the flat radiation-dominated Universe ($H^2 = 8\pi G\rho/3$) and the usual parametrization of the energy and entropy densities in terms of effective degrees of freedom ($\rho = g_{eff}(T)\pi^2 T^4/30$ and $S = 2h_{eff}(T)\pi^2 T^3/45$, see [3] for more detail) we get

$$\frac{dY}{dx} = -\sqrt{\frac{\pi}{45G}} \, \frac{g_*^{1/2} m_1}{x^2} \, \langle \sigma_{eff} v \rangle \, (Y^2 - Y_{eq}^2) \tag{2.43}$$

where

$$Y_{eq} = \frac{n_{eq}}{S} = \frac{45x^2}{4\pi^4 h_{eff}(T)} \sum_i g_i \left(\frac{m_i}{m_1}\right)^2 K_2 \left(x\frac{m_i}{m_1}\right), \tag{2.44}$$

and

$$g_*^{1/2} = \frac{h_{eff}}{\sqrt{g_{eff}}} \left(1 + \frac{T}{3h_{eff}} \frac{dh_{eff}}{dT}\right). \tag{2.45}$$

To get the relic density, we have to integrate (numerically) Equation 2.43 from $x = 0$ to $x = m_\chi/T_0$, where $T_0 = 2.726K$ is the temperature today. Name the result of this integration Y_0. Then the relic density is given by

$$\Omega_\chi = \frac{\rho_\chi}{\rho_{crit}} = \frac{m_\chi S_0 Y_0}{\rho_{crit}}, \tag{2.46}$$

where ρ_{crit} is the critical density and S_0 is the entropy density today. Using the values from [127] for g_{eff}, h_{eff} and $g_*^{1/2}$, we finally get

$$\Omega_\chi h^2 = 2.755 \times 10^8 \frac{m_\chi}{\text{GeV}} Y_0. \tag{2.47}$$

The problem is, therefore, reduced to evaluating all possible σ_{ij}'s, since they appear in the integration of Equation 2.43. Edsjo and Gondolo have performed these calculations in [126]. Further information on the relic density can be found in [125], [128], [129], [123]. We will examine the results of this calculation in detail in the following Section.

2.6 Direct Detection of Neutralino WIMPs

2.6.1 Motivation

In the previous Sections we have shown how the Standard Model of particles can be extended to include supersymmetry. In the minimal such extension, MSSM, the lightest neutralino turns out to be massive and stable. The neutralino relic density could be cosmologically important, in the sense that it could account for much of the dark matter. Maybe more importantly, the neutralino dark matter would be cold and non-baryonic, which is expected to be the dominant component of dark matter, as discussed in detail in Chapter 1. Finally, the interactions of neutralinos with the ordinary matter are roughly of the order of the weak scale, implying that it may be possible to detect the occasional interactions of the relic neutralinos in carefully designed detectors.

In this Section, we will perform more detailed numerical calculations of the neutralino relic density and the neutralino-proton spin-independent elastic scattering cross-section. Such calculations can be performed with sophisticated numerical packages, such as DarkSUSY and micrOMEGAs. In particular, we are interested in two questions:

- Exactly which supersymmetric models are accessible to the direct detection experiments such as CDMS?
- Are there other experiments (e.g. accelerator based) that could be complementary to the direct detection method?

In the remainder of this Section, we first describe the numerical packages and the theoretical frameworks that we will work with. We then discuss the neutralino-proton spin-independent elastic scattering cross-section, $\sigma_{\chi-p}$, with the particular emphasis on the existence of its lower bound. We examine the lower bound on $\sigma_{\chi-p}$ in two different frameworks of the supersymmetric models. Finally, we investigate the impact of the measurement of the anomalous magnetic moment of muon, and its complementarity with the direct detection approach.

2.6.2 Numerical Packages

Two numerical packages are used in this analysis, DarkSUSY and micrOMEGAs.

DarkSUSY [130] is a Fortran package by P. Gondolo *et al.* . It calculates the spin-independent and the spin-dependent elastic scattering cross-sections of the neutralino off of the neutron and the proton, as well as various particle masses and widths. It was the first package to properly solve the Boltzman equations in the calculation of the neutralino relic density. It includes neutralino coannihilations with charginos, squarks, and sfermions in the calculation of the neutralino relic density. Furthermore, this package calculates various accelerator-based bounds on the supersymmetric models, and determines whether the models pass such bounds. It also calculates the supersymmetric component of the anomalous magnetic moment of the muon, which can also be constrained by the experimental measurement. Finally, it evaluates various types of indirect detection signals (such as neutrino fluxes from the Sun and the Earth and the gamma flux from the galactic center, due to the neutralino annihilations) but we do not use these calculations here.

MicrOMEGAs [131] is a C-based numerical package by G. Belanger *et al.* . It includes over $2,800$ processes, including all coannihilation channels, in the calculation of the neutralino relic density. It tends to be computationally somewhat faster than DarkSUSY, but it does not calculate all of the other parameters provided by DarkSUSY.

The relic density calculations by the two packages typically agree to within 5% or better. In the analysis to follow we will use them to verify our analytic conclusions, and to compare the various models to the direct detection experiments, such as CDMS.

2.6.3 Theoretical Frameworks

We will examine two frameworks of supersymmetric models: the minimal supergravity (mSUGRA) and the general MSSM frameworks. We will generate models in both frameworks and use the numerical packages to study neutralinos as WIMP candidates. A large fraction of these models has first been described in [132].

The mSUGRA Framework

The mSUGRA framework is defined by the following assumptions:

- There exists a Grand Unified Theory (GUT) at some high energy scale. The gauge couplings $\alpha_{strong}, \alpha_{weak}$, and α_{EM} unify (that is, are equal) at the GUT scale. The value of the couplings at the weak scale determines the GUT scale to be $\approx 2 \times 10^{16}$ GeV. The gaugino mass parameters unify to $m_{1/2}$ at the GUT scale.

- The scalar mass parameters unify to m_0 and the trilinear couplings unify to A_0 at the GUT scale. Using the MSSM renormalization group equations (RGEs), we evaluate all parameters at the weak scale. We choose to do this using the one loop RGEs that can be found, for example, in [133] or [134].

- Radiative electroweak symmetry breaking (REWSB) is imposed: minimization of the one-loop Higgs effective potential at the appropriate scale fixes μ^2 and m_A (we follow the methods of Refs. [135, 136, 137, 138]). For completeness, we reproduce the equation for μ^2 at tree level:

$$\mu^2 = \frac{m_{H_1}^2 - m_{H_2}^2 \tan^2\beta}{\tan^2\beta - 1} - \frac{1}{2}M_Z^2. \tag{2.48}$$

With these assumptions, the mSUGRA framework allows four free parameters (m_0, $m_{1/2}$, A_0 and $\tan\beta$), in addition to the sign of μ. Starting with these parameters at the GUT scale, we determine all of the parameters at the weak scale, and supply them to the numerical packages to calculate the relic density and the scattering cross-section. We allow the free parameters to vary in the intervals

$$\begin{aligned} 0 < m_{1/2} < 300\,\mathrm{GeV},\ 95 < m_0 < 1000\,\mathrm{GeV}, \\ -3000 < A_0 < 3000\,\mathrm{GeV},\ 1.8 < \tan\beta < 25. \end{aligned} \tag{2.49}$$

The upper bounds on these parameters come from the naturalness assumption: one of the reasons for using supersymmetry is its ability to naturally relate high and low energy scales; as a result, no parameter in the theory should be very large. Given the mass of M_Z, values of $\mu > 1$ TeV would imply fine-tuning at less than 1% in the Equation 2.48. The GUT relation between μ and $m_{1/2}$ converts the constraint $|\mu| < 1$ TeV into $m_{1/2} < 300$ GeV. We will investigate the behavior of the mSUGRA models as this assumption is relaxed. We also impose the 1 TeV bounds on the other mass parameters. The low value of $\tan\beta$ is set by the requirement that the top Yukawa coupling does not blow up before the GUT scale is reached.

The General MSSM Framework

In the general MSSM framework we relax our assumptions. The scalar masses and the trilinear scalar couplings no longer unify at a high energy scale, but the unification of the gaugino masses is still imposed. We also drop the REWSB requirement (i.e., we take $m^2_{H_{1,2}}$ as independent parameters from m_0). We assume that all scalar mass parameters are equal at the *weak scale*: m_{sq}. This assumption simplifies the calculation, and it should not affect the general flavor of our results. Furthermore, of all trilinear coupling parameters, only A_t and A_b are allowed to be non-zero. The free parameters are, therefore, $\mu, M_2, \tan\beta, m_A, m_{sq}, A_t, A_b$. We also relax the naturalness assumption, allowing the free parameters to have very large values. Besides the relatively uniform scans of the parameter space, we also performed dedicated scans exploring particular regions of the parameters space. The overall free parameter space is given by:

$$\begin{aligned} -300\,\mathrm{TeV} < \mu < 300\,\mathrm{TeV}, &\qquad 0 < M_2 < 300\,\mathrm{TeV}, \\ 95\,\mathrm{GeV} < m_A < 10\,\mathrm{TeV}, &\qquad 200\,\mathrm{GeV} < m_{sq} < 50\,\mathrm{TeV}, \\ -3 < \frac{A_{t,b}}{m_{sq}} < 3, &\qquad 1.8 < \tan\beta < 65. \end{aligned} \tag{2.50}$$

2.6.4 Spin-Independent Elastic Scattering Cross-Section

As mentioned in the Section 2.4, the dominant contribution to the neutralino-proton spin-independent elastic scattering cross-section, $\sigma_{\chi-p}$, is usually the Higgs boson exchange. Figure 2.2 illustrates this in our results. The nearly perfect 45 degree line in the Figure indicates good agreement between the total cross-section as evaluated by DarkSUSY and the cross-section calculated including the exchange of the Higgs bosons only (in the approximation explained below). We will, therefore, concentrate on the Higgs boson exchange.

The contribution of the Higgs boson exchange can be found in the literature [123, 139, 140, 141, 142, 143]. It is of the following form:

$$\sigma_{h,H} \sim |(f_u + f_c + f_t)A^u + (f_d + f_s + f_b)A^d|^2, \tag{2.51}$$

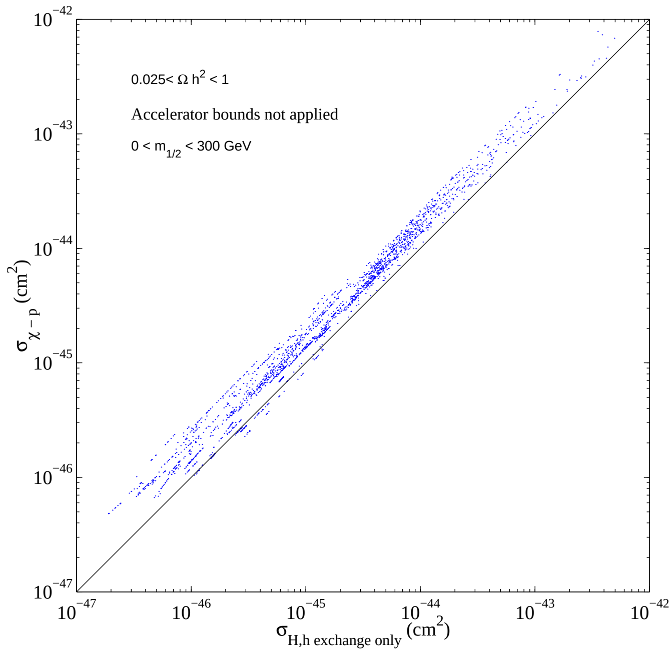


Figure 2.2: The cross-section for spin-independent χ-proton scattering: the complete (DarkSUSY) calculation is shown on the y axis and the contribution to the cross-section from the Higgs bosons exchange alone is shown on the x axis. A relatively weak relic density cut is applied, $0.025 < \Omega_\chi h^2 < 1$. Figure taken from [132].

where $f_u \approx 0.023, f_d \approx 0.034, f_s \approx 0.14, f_c = f_t = f_b \approx 0.0595$ parametrize the quark-nucleon matrix elements with

$$A^u = \frac{g_2^2}{4M_W}\Big(\frac{F_h}{m_h^2}\frac{\cos\alpha_H}{\sin\beta} + \frac{F_H}{m_H^2}\frac{\sin\alpha_H}{\sin\beta}\Big), \tag{2.52}$$

$$A^d = \frac{g_2^2}{4M_W}\Big(-\frac{F_h}{m_h^2}\frac{\sin\alpha_H}{\cos\beta} + \frac{F_H}{m_H^2}\frac{\cos\alpha_H}{\cos\beta}\Big), \tag{2.53}$$

and

$$\begin{aligned} F_h &= (N_{12} - N_{11}\tan\theta_W)(N_{14}\cos\alpha_H + N_{13}\sin\alpha_H) \\ F_H &= (N_{12} - N_{11}\tan\theta_W)(N_{14}\sin\alpha_H - N_{13}\cos\alpha_H). \end{aligned} \tag{2.54}$$

The N's are the coefficients appearing in the Equation (2.24) and α_H is the Higgs boson mixing angle (defined after radiative corrections have been included in the Higgs mass matrix). A^u represents the amplitude for scattering off of an up-type quark in a nucleon, while A^d represents the amplitude for scattering off of a down-type quark in the nucleon. For bino-like neutralino, $\mu > M_1$ and $\mu > M_Z$ are satisfied. Then, following [142], we can expand the N_{1i}'s out in powers of $\frac{M_Z}{\mu}$. We reproduce their result here:

$$\begin{aligned}
N_{11} &\approx 1, \\
N_{12} &\approx -\frac{1}{2}\frac{M_Z}{\mu}\frac{\sin 2\theta_W}{(1-M_1^2/\mu^2)}\frac{M_Z}{M_2-M_1}\left[\sin 2\beta + \frac{M_1}{\mu}\right], \\
N_{13} &\approx \frac{M_Z}{\mu}\frac{1}{1-M_1^2/\mu^2}\sin\theta_W \sin\beta \left[1+\frac{M_1}{\mu}\cot\beta\right], \\
N_{14} &\approx -\frac{M_Z}{\mu}\frac{1}{1-M_1^2/\mu^2}\sin\theta_W \cos\beta \left[1+\frac{M_1}{\mu}\tan\beta\right].
\end{aligned} \tag{2.55}$$

We are particularly interested in the existence of the lower bound on $\sigma_{\chi-p}$. We identify five cases in which the Higgs boson exchange channels are suppressed (we follow the argument in [132]).

- Case 1: $N_{12} - N_{11}\tan\theta_W = 0$ would make both F_h and F_H vanish. Intuitively, this condition implies that the neutralino is a pure photino, and the tree level Higgs coupling to the photino vanishes. Using Equation 2.55, this condition becomes:

$$\mu(M_2 - M_1) = -M_Z^2\cos^2\theta_W\left(\sin 2\beta + \frac{M_1}{\mu}\right). \tag{2.56}$$

Since $M_2 > M_1 > 0$, the last equation can be satisfied only if $\mu < 0$. If we use the GUT relationship betweeen M_1 and M_2, we get

$$\mu M_1\left(\frac{3\cos^2\theta_W}{5\sin^2\theta_W} - 1\right) = -M_Z^2\cos^2\theta_W\left(\sin 2\beta + \frac{M_1}{\mu}\right) \tag{2.57}$$

or, equivalently, for μ and M_1 in units of GeV

$$M_1 = \frac{-6449\sin 2\beta}{\mu + \frac{6449}{\mu}}. \tag{2.58}$$

For $\sin 2\beta = 1$, Equation 2.58 implies $M_\chi \approx M_1 < 40$ GeV, and the constraint is even more severe for other values of $\sin 2\beta$.

- Case 2: $N_{14} \sin\alpha_H - N_{13}\cos\alpha_H = 0$ would make only F_H vanish. Using Equation 2.55, this becomes:

$$M_1/\mu = -\cot\alpha_H - \cot\beta. \tag{2.59}$$

At the tree level, the following relation holds:

$$\cot 2\alpha_H = k\cot 2\beta,\ k = \frac{m_A^2 - M_Z^2}{m_A^2 + M_Z^2}, \tag{2.60}$$

which, after some trigonometric manipulations, gives

$$\begin{aligned} \cot\alpha_H &= \frac{k}{2}\left(\frac{1}{\tan\beta} - \tan\beta\right) \\ &\quad -\sqrt{\frac{k^2}{4}\left(\frac{1}{\tan^2\beta} + \tan^2\beta - 2\right) + 1}. \end{aligned} \tag{2.61}$$

Both terms on the right hand side of the Equation 2.61 are negative because $\tan\beta > 1$. Hence, the minimum of $|\cot\alpha_H| = 1$ occurs when $k = 0$ (or, equivalently, when $m_A = M_Z$). Finally, since $\tan\beta > 1.8$, the Equation 2.59 gives:

$$\frac{M_1}{\mu} > 0.5. \tag{2.62}$$

- Case 3: $N_{14}\cos\alpha_H + N_{13}\sin\alpha_H = 0$ would make F_h vanish. Using Equation 2.55 gives:

$$M_1/\mu = \tan\alpha_H - \cot\beta. \tag{2.63}$$

Since $\tan\alpha_H < 0$, the condition for the vanishing of the light Higgs boson contribution can be satisfied only for $\mu < 0$. Note that if this condition is satisfied, $\sigma_{\chi-p}$ becomes dominated by the heavy Higgs boson exchange (along with other channels such as the squark exchange). An upper bound on the heavy Higgs boson mass would, therefore, place a lower bound on $\sigma_{\chi-p}$ in this case. Such a bound can be introduced by imposing naturalness assumptions (for example, in the mSUGRA framework).

- Case 4: Another way to make both F_h and F_H small, is to allow very large values for μ (N_{12}, N_{13}, and N_{14} all contain μ in denominator). A priori, there is no upper bound on μ.

- Case 5: It is possible to achieve a complete cancellation of terms in the Equation 2.51. To show this, we examine the relative signs of F_h, F_H and μ. Since $\tan\beta > 1.8$ and $|M_1| < |\mu|$ for bino-like models, Equation 2.55 implies $|N_{13}| > |N_{14}|$. Further, since $\cot\alpha_H < -1$ (and $\sin\alpha_H < 0$), the N_{13} term dominates over the N_{14} term in F_H (Equation 2.54). Hence, $F_H/\mu > 0$ always, consistent with the analysis of Case 2 above. The situation is more complicated in the case of F_h. For $\mu > 0$, a similar analysis gives $F_h/\mu > 0$. In this case, the two terms in A^u (Equation 2.52) are of opposite signs and the two terms in A^d (Equation 2.53) are of the same sign. Moreover, the pre-factor of A^u in the Equation 2.51 is much smaller than the pre-factor of A^d. As a result, A^d strongly dominates over A^u in the Equation 2.51, preventing $\sigma_{\chi-p}$ from vanishing.

  On the other hand, if $\mu < 0$ and if $\mu > -M_1\tan\beta$, N_{14} can change sign relative to μ. This change of sign can propagate through Equations 2.54, 2.53, and 2.51, so that the A^u and A^d terms in the Equation 2.51 are of opposite sign. If the parameters are tuned properly, a complete cancellation of these terms can be obtained.

We now examine these cases in the mSUGRA and in the general MSSM frameworks.

2.6.5 mSUGRA Results

Before we present the results in the mSUGRA framework, a few remarks are in order. First, the $b \to s + \gamma$ constraint eliminates large portions of the $\mu < 0$ parameter space, in agreement with [144, 145, 146, 147]. Second, for both $\mu > 0$ and $\mu < 0$, we find no higgsino-like LSP models that are cosmologically important, in agreement with [146, 147].

We scan the parameter space defined in the Equation 2.49 and use DarkSUSY and micrOMEGAs to evaluate the neutralino mass M_χ, the neutralino relic density, the neutralino-proton elastic scattering cross-section $\sigma_{\chi-p}$, and the bounds from various accelerator based experiments. Figure 2.3 shows the allowed region in the mSUGRA framework in the $\sigma_{\chi-p} - M_\chi$ plane. Note that the allowed region is split into two parts by the annihilation channel into W^+W^-, which affects the neutralino relic density. We impose a relic density constraint based on the WMAP measurement of $\Omega_m h^2 = 0.14 \pm 0.02$ [56] (see also Chapter 1 for more detail). At 2σ, this becomes $\Omega_\chi h^2 < 0.18$. We use the estimate of the relic density calculated by micrOMEGAs - using the DarkSUSY estimate gives similar results. The upper bound on $\sigma_{\chi-p}$ and the lower bound on M_χ (of the theoretically allowed regions) come from the accelerator-based constraints. The upper bound on M_χ is a combination of the upper bound on the neutralino relic density and of the bounds on the free parameters. These results agree with other calculations, such as [148, 149], although different scans of the parameter space are performed and different constraints (cosmological and accelerator-based) are applied throughout the literature.

In the mSUGRA framework, with the parameter space defined by the Equation 2.49, $\sigma_{\chi-p}$ has a lower limit of $\sim 10^{-46}$ cm^2 - we examine the origin of this bound in detail below. Assuming a ^{73}Ge target, the local dark matter density $\rho = 0.3$ GeVc^{-2}cm^{-3}, the WIMP characteristic velocity $v_0 = 230$ km s^{-1} and following Ref. [100], this lower bound corresponds to the event rate $R > 0.1$ ton^{-1}day^{-1}. A large fraction of this allowed region is within reach of the current, or future proposed, direct detection experiments, as shown in Figure 2.3. However, relaxing the naturalness assumption by allowing $m_{1/2} < 1$ TeV leads to larger neutralino masses and to significantly lower $\sigma_{\chi-p}$ values. In this case, much of the allowed region is out of reach of the direct detection experiments.

We now go through the five cases of low $\sigma_{\chi-p}$ in the mSUGRA framework. Figure 2.4 illustrates which regions of the M_1-μ plane satisfy the conditions of the five cases.

- Case 1: As $\sin 2\beta$ ranges from 0 to 1, Equation 2.58 spans the dashed regions marked "Case 1" in Figure 2.4. Such low values of M_χ are excluded by the accelerator-based constraints. We conclude that this condition cannot be satisfied in the mSUGRA framework.

- Case 2: We formulate a relationship between μ and M_1 resulting from the RGEs and REWSB assumptions. An approximate solution to the RGEs for the Higgs mass parameters, m_{H_1} and m_{H_2} (based on the expansion around the infrared fixed point) is given in [134]:

$$\begin{aligned} m_{H_1}^2 &\approx m_0^2 + 0.5 m_{1/2}^2, \\ m_{H_2}^2 &\approx -0.5 m_0^2 - 3.5 m_{1/2}^2. \end{aligned} \tag{2.64}$$

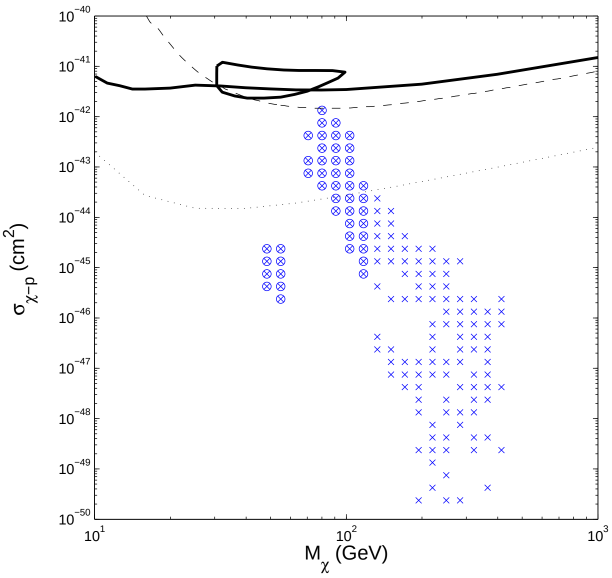


Figure 2.3: The cross-section for the spin-independent χ-proton elastic scattering is plotted against the neutralino mass. Accelerator bounds, including $b \to s + \gamma$, are imposed through the DarkSUSY code. We do not impose a constraint on the anomalous magnetic moment of muon here. The constraint on the neutralino relic density from the WMAP measurement of $\Omega_m h^2$ is imposed (at 2σ): $\Omega_\chi h^2 < 0.18$. The o's denote the models in the parameter space defined in the Equation 2.49. Note that in this parameter space, there is a lower bound on $\sigma_{\chi-p}$ of 10^{-46} cm^2. The x's denote models allowed when the upper bound on $m_{1/2}$ is relaxed to 1 TeV - in this case, much lower values of $\sigma_{\chi-p}$ are allowed. The solid line denotes the CDMS result achieved at the shallow site at SUF [104], the dotted line is the projection for the CDMS II experiment in Soudan, Minnesota, and the dashed line denotes the recent result from the Edelweiss experiment [105]. The closed contour denotes the DAMA (1-4) 3σ signal region [106, 110].

These equations, coupled with the Equation 2.48, yield a value for μ^2 in terms of $m_{1/2}, m_0$, and $\tan\beta$. Using the GUT relationship:

$$M_1 = m_{1/2}\frac{\alpha_1(m_Z)}{\alpha_{GUT}}, \tag{2.65}$$

we get a roughly linear relationship between μ and M_1:

$$M_1 = (0.3|\mu| - 60) \pm 40. \tag{2.66}$$

The spread ± 40 comes from the variation in m_0 and $\tan\beta$ and the linearity breaks

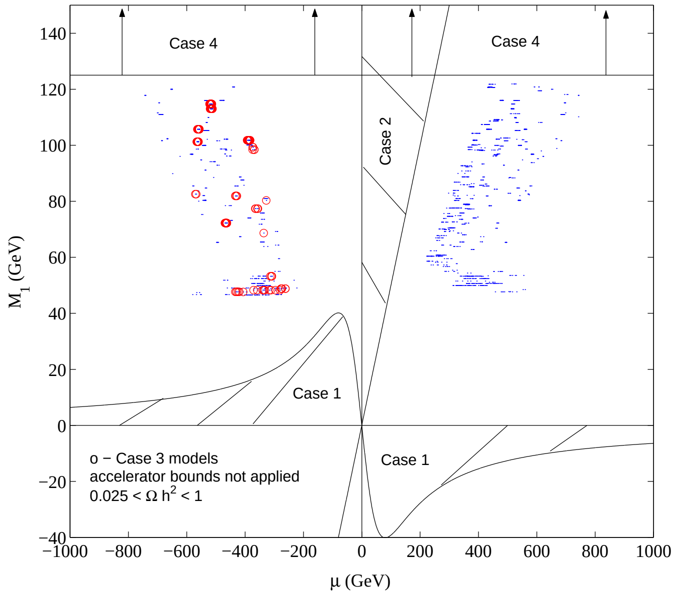


Figure 2.4: This figure illustrates the regions of $M_1 - \mu$ plane (as defined by the Equation 2.49) in which the different cases discussed in the text can be satisfied. For the models shown, the accelerator constraints were not applied, but a relatively weak relic density cut ($0.025 < \Omega_\chi h^2 < 1$) was applied. Note that almost all of the models of Case 3 shown in this plot are ruled out by the accelerator constraints. Also, there are no models satisfying the Case 5 for the parameter space defined in the Equation 2.49. Figure taken from [132].

down somewhat at the low values of M_1. The empirical relationship that we obtain from running the DarkSUSY code (see Figure 2.4) is very similar:

$$M_1 = (0.3|\mu| - 40) \pm 25. \tag{2.67}$$

The smaller spread comes from the application of the relic density constraint and of the accelerator-based constraints. In any case, we conclude that $M_1/\mu > 0.3$ is not allowed in the mSUGRA framework. This is in conflict with the Equation 2.62, implying that the Case 2 cannot be satisfied in the mSUGRA framework.

- Case 3: Figure 2.5 shows that when this condition is (approximately) satisfied, the contribution of the heavy Higgs boson exchange to the elastic scattering cross-section

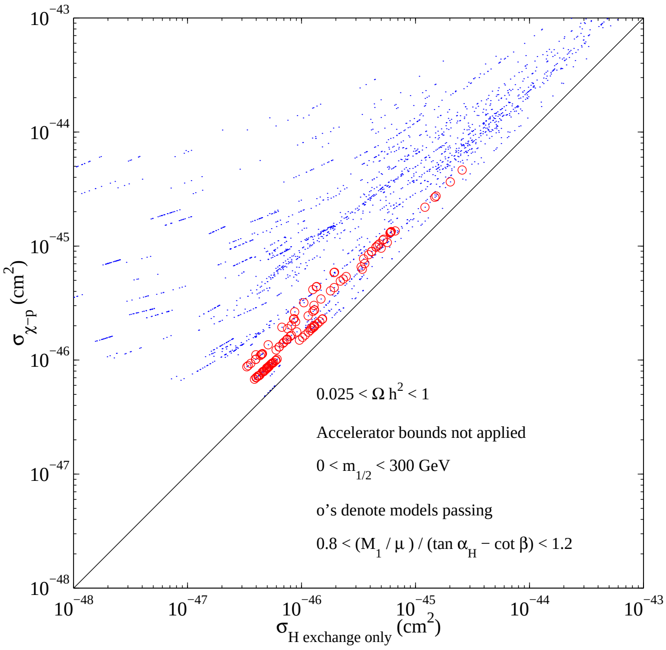


Figure 2.5: The neutralino-proton scattering cross-section for models where the condition in Case 3 is approximately satisfied (in particular, models satisfying $0.8 < \frac{M_1/\mu}{\tan\alpha_H - \cot\beta} < 1/2$ are denoted by o's). The complete (DarkSUSY) calculation is shown on the y-axis and the contribution to the cross-section due only to the exchange of the heavy Higgs boson, is shown on the x-axis. A relatively weak relic density cut is applied, $0.025 < \Omega_\chi h^2 < 1$. Most of the models denoted by o's (for which the Equation 2.63 is approximately satisfied) are excluded by the accelerator bounds on the Higgs boson masses. Figure taken from [132].

is the largest. Since the parameter space is bounded, the heavy Higgs boson mass cannot be arbitrarily large, implying that $\sigma_{\chi-p}$ cannot be arbitrarily small.

- Case 4: The naturalness assumption keeps $\mu < 1$ TeV in the mSUGRA framework. Hence, Case 4 is not satisfied in this framework.

- Case 5: The complete cancellation of terms in the Equation 2.51 indeed happens in the mSUGRA framework, but only if $M_\chi(\approx M_1) > 120$ GeV. Such neutralino masses are not allowed due to the upper bound on $m_{1/2}$ given in the Equation 2.49, but they are allowed if this upper bound is relaxed to $m_{1/2} < 1$ TeV. Figure 2.6 shows that

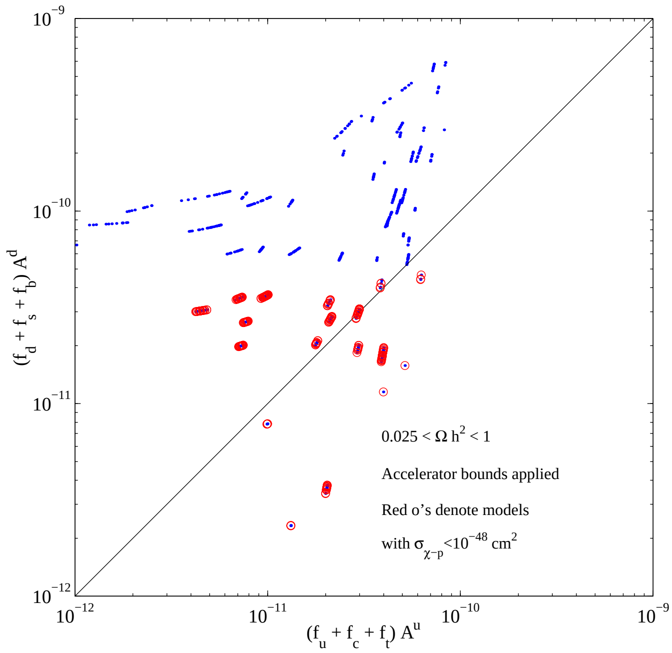


Figure 2.6: The absolute values of the A^d and the A^u terms of the Equation 2.51 are shown. Only $A^d < 0$ (for which $\mu < 0$) are shown, satisfying a weak relic density constraint $0.025 < \Omega_\chi h^2 < 1$, passing the accelerator bounds, and with $m_{1/2} < 1$ TeV. The o's denote the lowest $\sigma_{\chi-p}$ models ($< 10^{-48}$ cm^2). Clearly, the lowest values of $\sigma_{\chi-p}$ are achieved when the A^d and A^u terms cancel each other out. Figure taken from [132].

the lowest values of $\sigma_{\chi-p}$ in the mSUGRA framework are reached exactly when the A^d and A^u terms (roughly) cancel each other. Hence, the lower bound on $\sigma_{\chi-p}$ in the mSUGRA model exists only if the naturalness assumption is sufficiently constraining (assuming that the constraint on the anomalous magnetic moment of muon is not applied - we will see below how such constraint modifies our conclusion).

2.6.6 General MSSM Results

We start with a couple of comments. First, in the general MSSM framework we observe the higgsino-like (as well as the bino-like) lightest neutralino. In agreement with [150], we find very few light higgsino-like models, which will probably be explored soon by the accelerator experiments. Most of the higgsino-like models (with gaugino content $z_g < 0.01$) have $M_\chi > 450$ GeV, implying very large values of $m_{1/2}$. In particular, in higgsino-like models $M_1 > \mu \approx M_\chi$; our results give $M_1 > 700$ GeV or, equivalently, $m_{1/2} > 1700$ GeV, which can be considered unnatural. For these reasons, we choose not to analyze the higgsino case. Second, $b \to s + \gamma$ is less constraining than in the mSUGRA case, but our results are still consistent with [144, 145, 146, 147].

Figure 2.7 shows $\sigma_{\chi-p}$ versus M_χ plane in the general MSSM framework. As in the case of mSUGRA, we impose accelerator-based constraints, including $b \to s+\gamma$, and the relic density constraint based on the WMAP result, $\Omega_\chi h^2 < 0.18$ (we use the DarkSUSY calculation of the relic density). We do not impose a constraint on the anomalous magnetic moment of muon. Although our parameter-scans do not map out the complete allowed region in this plane, Figure 2.7 does depict some generic differences from the mSUGRA case. Namely, we can obtain much larger values for the neutralino mass because of the size of the parameter space. In addition, much lower values of $\sigma_{\chi-p}$ (than in the mSUGRA case) are allowed, even at the lowest M_χ values. We discuss the details below. The results presented here are in good agreement with other calculations, such as [151, 152, 153, 154, 86], although the parameter space definitions and the applied constraints (cosmological and accelerator-based) vary throughout the literature.

We concentrate only on the bino-like lightest neutralino. All results in this Section are presented with this assumption in mind. In particular, we demand $z_g = (N_{11}^2 + N_{12}^2)/(N_{13}^2 + N_{14}^2) > 10$. In this case, we can rely on the approximations outlined in the Section 2.6.4. In particular, the expansions in the Equation 2.55 are valid. Hence, we can

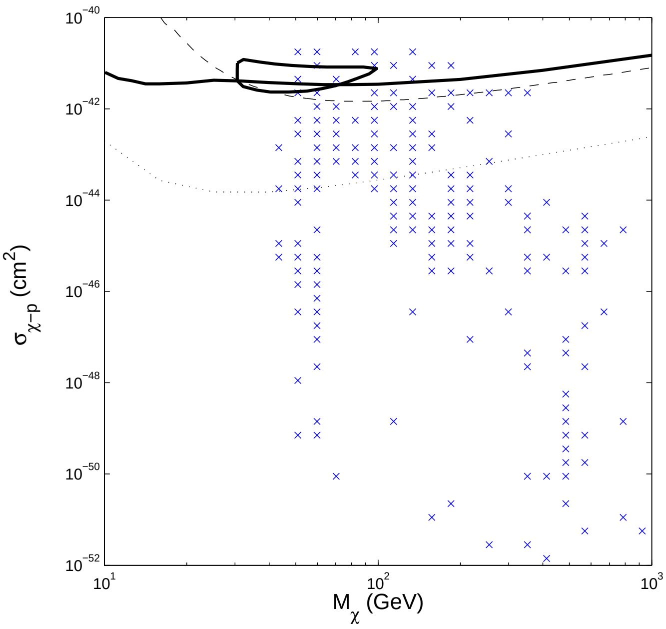


Figure 2.7: The cross-section for spin-independent neutralino-proton scattering in the general MSSM framework is shown versus the neutralino mass. The relic density constraint $\Omega_\chi h^2 < 0.18$ is imposed. The constraint on the gaugino fraction, $z_g > 10$, is imposed. Accelerator bounds, including $b \rightarrow s + \gamma$, are imposed through the DarkSUSY code. As in the Figure 2.3, the solid line denotes the CDMS result achieved at the shallow site at SUF [104], the dotted line is the projection for the CDMS II experiment in Soudan, Minnesota, and the dashed line denotes the current result from the Edelweiss experiment [105]. The closed contour denotes the DAMA (1-4) 3σ signal region [106, 110].

revisit the five different cases of low $\sigma_{\chi-p}$.

- Case 1, as noted in the Section 2.6.4, requires $M_\chi < 40$ GeV, which is ruled out by the accelerator-based bounds (as shown in the Figure 2.7). The models that get close to satisfying this condition have very low contributions due to the Higgs bosons exchange, so this is one of the rare situations where the squark exchange is important.

- Case 2 is possible in the general MSSM framework, because μ and M_1 are not related. We indeed observe that the heavy Higgs boson exchange contribution is very small in this case, but since the light Higgs boson exchange dominates, $\sigma_{\chi-p}$ is kept relatively high.

- Case 3 is also possible to satisfy in the general MSSM framework. As in the mSUGRA framework, the light Higgs boson exchange is small and the heavy Higgs boson exchange dominates in this Case, so $\sigma_{\chi-p}$ cannot be arbitrarily small.

- Case 4 is also more important in the general MSSM framework. Since the naturalness constraint has been relaxed, μ is allowed to have very large values. The neutralino content coefficients N_{12}, N_{13}, and N_{14} can take much smaller values, which in turn can make the contribution of the Higgs bosons exchange small. Intuitively, as discussed in the Section 2.4, large $|\mu|$ implies that the neutralino is a very pure bino, for which the Higgs boson scattering channels vanish. If the squark masses are kept large, the squark contribution will be small as well, making the total elastic scattering cross-section very small. This is illustrated in Figure 2.8 - the lowest values of $\sigma_{\chi-p}$ are obtained for the largest values of $|\mu|$.

- Case 5 is also possible to satisfy in the general MSSM framework. In fact, since M_1 and μ are not related, it is relatively easy to make $(\frac{M_1}{\mu}\tan\beta)$ negative and large, which can lead to the complete cancellation of terms in the Equation 2.51. Hence, such cancellation can happen even at low neutralino masses (recall that, in the mSUGRA case, such cancellation was possible only for $M_\chi > 120$ GeV). In particular, this effect can lead to models with $M_\chi \sim 50-60$ GeV and $\sigma_{\chi-p} < 10^{-50}$ cm^2. Such models were not allowed in the mSUGRA framework due to the $M_1-\mu$ relationship (determined by the RGE's and the REWSB assumption), which does not exist in the general MSSM framework.

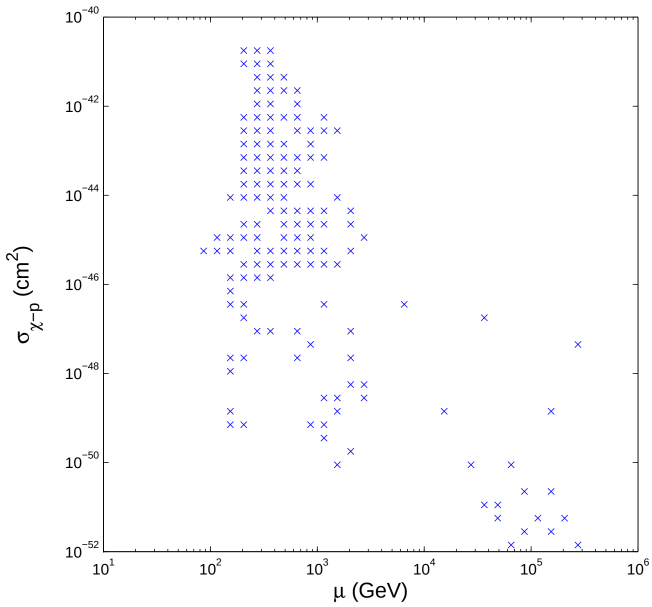


Figure 2.8: The cross-section for the spin-independent neutralino-proton scattering in the general MSSM framework is plotted versus $|\mu|$. The relic density constraint, $\Omega_\chi h^2 < 0.18$, is imposed, along with the constraint on the gaugino fraction, $z_g > 10$, and with the accelerator bounds, including $b \rightarrow s + \gamma$ (calculated by the DarkSUSY code).

In other words, in the general MSSM framework it is possible to obtain low values of $\sigma_{\chi-p}$ in several different ways: either by tuning parameters to suppress the contribution of the heavy or the light Higgs boson exchange, or by allowing parameters (such as μ) to be very large, which suppreses both the heavy and the light Higgs boson exchange channels, or by fine-tunning parameters to achieve a complete cancellation of terms in the Equation 2.51. As a result, much lower values of $\sigma_{\chi-p}$ are allowed even at the lowest neutralino masses, beyond the reach of the present and the proposed direct-detection experiments.

2.6.7 Anomalous Magnetic Moment of Muon

In early 2001, the Brookhaven AGS experiment 821 measured the anomalous magnetic moment of the muon, a_μ, [155]. Their result was larger than the Standard Model prediction by more than 2.6σ. However, a sign error was made in the calculation of the hadronic light-by-light contribution to a_μ. After correcting the error, the disagreement with the Standard Model was only at 1.6σ. More recently, the AGS experiment produced a new result with much smaller statistical error [156]. Similarly, new calculations of the Standard Model contributions have been made (such as [157]). However, the calculations using data from the hadronic e^+e^- cross-section and from the hadronic τ decays produce different results. We summarize these here:

$$\begin{aligned} a_\mu(\mathrm{exp}) &= 11\,659\,203(8) \times 10^{-10}, && (2.68)\\ a_\mu(\mathrm{SM}, \mathrm{e^+e^-}) &= 11\,659\,169.1(7.8) \times 10^{-10}, && (2.69)\\ a_\mu(\mathrm{SM}, \tau) &= 11\,659\,186.3(7.1) \times 10^{-10}. && (2.70) \end{aligned}$$

The e^+e^- estimate disagrees with the AGS experimental result at 3σ, while the τ result only at 1.6σ. The e^+e^- result is likely more reliable, so we will use it here, but we note that this issue has not been resolved yet and that there are likely going to be modifications in the future. Hence, the discrepancy between the AGS measurement and the Standard Model prediction is:

$$a_\mu(\mathrm{exp}) - a_\mu(\mathrm{SM}) = (33.9 \pm 11.2) \times 10^{-10}. \qquad (2.71)$$

As shown in [158, 159, 160, 161], this discrepancy can possibly be explained by including the supersymmetric corrections to a_μ. Hence, the supersymmetric contribution to a_μ is constrained (at 2σ):

$$11.5 \times 10^{-10} < \Delta a_\mu(\mathrm{SUSY}) < 56.3 \times 10^{-10}. \qquad (2.72)$$

Several recent studies examined the implications of the Brookhaven experiment for the neutralino dark matter searches [162, 163, 86, 164] and for the accelerator experiments [164, 165]. Here, we want to investigate the impact of this result on the neutralino-proton spin-independent cross-section and, in particular, on its lower bound. In other words, we will investigate if the conclusions reached in the previous Sections are modified by the AGS result.

The crucial point here is that the sign of the supersymmetric contribution to a_μ is strongly correlated to the sign of μ, as discussed in [159]. The Equation 2.72 then implies that only $\mu > 0$ models could account for the a_μ discrepancy. This is very important: as discussed in the Section 2.6.4, Cases 1, 3, and 5 of low $\sigma_{\chi-p}$ values can happen only if $\mu < 0$. Hence, by prohibiting the most severe cancellations (Cases 3 and 5), the result of the Brookhaven experiment strongly favors high values of $\sigma_{\chi-p}$. We found this to be true both in the general MSSM and in the mSUGRA frameworks.

Furthermore, the Δa_μ(SUSY) constraint (Equation 2.72) also imposes upper bounds on the mass parameters M_2, μ etc. The origin of these bounds can be traced back to the fact that the dominant diagrams contributing to Δa_μ(SUSY) have neutralino-smuon and chargino-sneutrino intermediate states. Hence, Δa_μ(SUSY) is strongly related to the masses of these particles, and, therefore, to the more fundamental mass parameters (see, for example, Equations 30-34 in [159] for more detail). Figure 2.9 shows this effect on the mSUGRA and on the general-MSSM models. These results agree with Baltz and Gondolo [86] - they allow somewhat larger masses, which could be explained by the their higher upper bound on $\Omega_\chi h^2$, or possibly by different scanning of the parameter space.

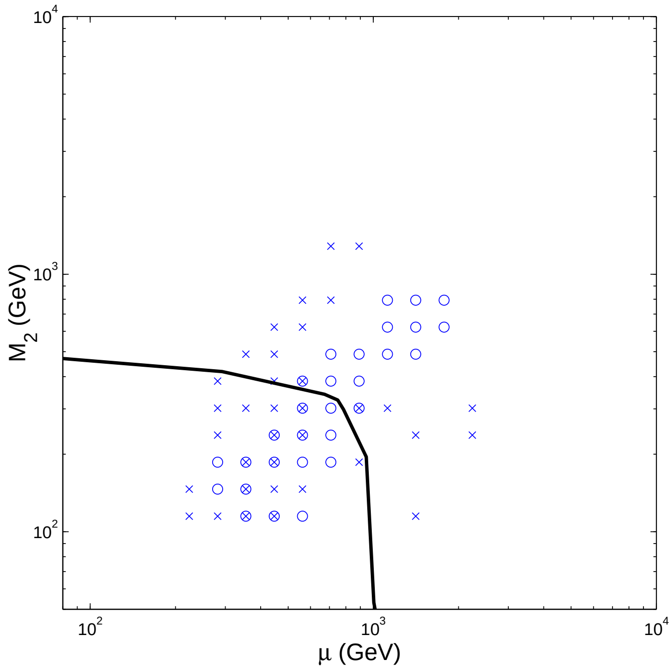


Figure 2.9: M_2 vs. μ plane for both the mSUGRA (o's) and the general MSSM (x's) models defined by the Equations 2.49 and 2.50, respectively. We relax the mSUGRA upper-bound on $m_{1/2}$ to 1 TeV. We impose the accelerator-based constraints and the relic density constraint $\Omega_\chi h^2 < 0.18$. The models passing the Δa_μ(SUSY) constraint (Equation 2.72) are below the thick solid line.

We conclude that the Δa_μ(SUSY) constraint has a similar effect to the naturalness assumption. In particular, it prohibits the Case 4 discussed in the Section 2.6.4, in both the general MSSM framework and in the mSUGRA framework (where this was achieved earlier by imposing the naturalness constraint).

Hence, of the five cases of low $\sigma_{\chi-p}$ values, the Δa_μ(SUSY) constraint allows only Case 2 to happen. We saw earlier that this Case was not allowed in the mSUGRA framework, but it was in the general MSSM framework. Nevertheless, in the Case 2, only the heavy Higgs boson contribution is small - the light Higgs boson contribution is unaffected, so the $\sigma_{\chi-p}$ cannot take very low values.

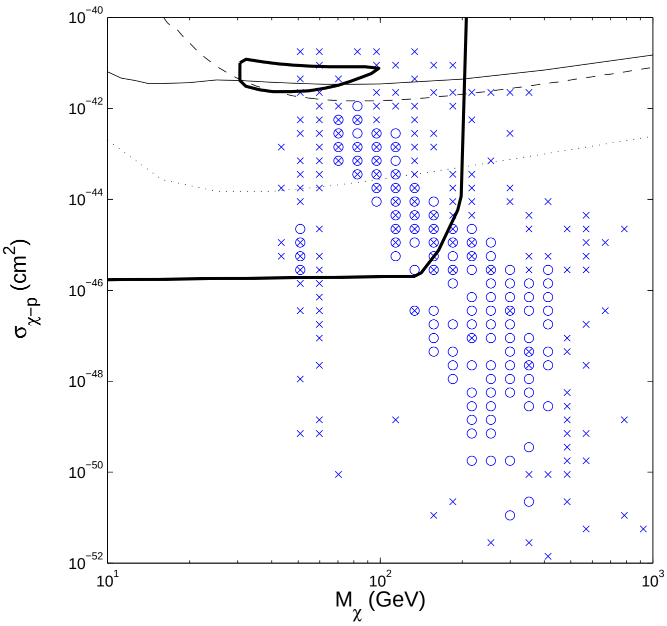


Figure 2.10: The regions allowed in the mSUGRA (o's) and in the general MSSM (x's) frameworks are shown. We relax the mSUGRA upper bound on $m_{1/2}$ to 1 TeV. We impose the accelerator-based constraints and the relic density constraint $\Omega_\chi h^2 < 0.18$. The models passing the Δa_μ(SUSY) constraint fall above the thick solid line. The thin solid line denotes the CDMS result achieved at the shallow site at SUF [104], the dotted line is the projection for the CDMS II experiment in Soudan, Minnesota, and the dashed line denotes the current result from the Edelweiss experiment [105]. The closed contour denotes the DAMA (1-4) 3σ signal region [106, 110].

Figure 2.10 depicts the impact of the Δa_μ(SUSY) constraint in the $\sigma_{\chi-p}-M_\chi$ plane for the mSUGRA and the general MSSM frameworks. As discussed above, low neutralino masses and high $\sigma_{\chi-p}$ values are preferred. In particular, much of the allowed region is within reach of the current or proposed direct-detection experiments. It is interesting that the region preferred by the Δa_μ(SUSY) constraint is similar to the region preferred by the naturalness assumptions.

We conclude, therefore, that the Δa_μ(SUSY) constraint given in the Equation 2.72 places a lower bound on $\sigma_{\chi-p}$. We emphasise again, however, that the estimate of the Standard Model contribution to a_μ is still uncertain. Hence, if the future modifications of this contribution allow $\Delta a_\mu(\mathrm{SUSY}) < 0$, then the $\mu < 0$ models (and the Case 5) will be allowed again, implying much lower values of $\sigma_{\chi-p}$.

Chapter 3

CDMS Setup

3.1 Introduction

Cryogenic Dark Matter Search experiment was designed to directly search for Dark Matter in the form of Weakly Interactive Massive Particles (WIPMs). The main objective of experiments such as CDMS is suppresssing and understanding various types of backgrounds. In CDMS experiment, this is achieved in several different ways:

- CDMS is operated in underground sites (shallow Stanford Underground Facility and deep at the Soudan mine in Minnesota). This allows passive suppression of various cosmogenic backgrounds.

- CDMS also benefits from the lead shield that suppresses the ambient gamma background, and from the polyethylene shield that suppresses the ambient neutron background.

- CDMS uses scintillator-based muon veto, to discriminate against muon coincident events (i.e. events caused by muons interacting with the material surrounding the detectors).

- The detectors designed by CDMS are based on crystals of Ge or Si, operated at the temperature of about 20 mK. These detectors give a two-fold signature of an interaction inside the crystal. One signature is the ionization generated by the recoiling nucleus or electron after the interaction in the crystal. The other signature is the athermal phonon signal generated by the recoiling particle. A particle entering the crystal could interact either with an electron (as in the case of incoming gammas or betas) or with a nucleus (as in the case of incoming neutrons and possibly WIMPs) inside the crystal. The two interaction signatures allow discrimination against events where the recoiling particle is an electron (events predominantly caused by gammas or betas) - this discrimination is achieved on event-by-event basis.

- One difficulty with these detectors is the reduced ionization signal for events that take place on the surface of the crystal. In particular, the surface electron-recoil events could look like nuclear-recoil events, and therefore fake the WIMP-signal. With the CDMS detectors, we have a way of handling this problem - in particular, the pulse shape of the athermal phonon signal contains event position information, so it can be used to discriminate against the surface events.

- The items listed above leave out only one type of background - the muon-anticoincident neutron-caused nuclear-recoil events. CDMS handles this problem in several ways. First, the interaction cross-sections of Ge and Si nuclei with neutrons and WIMPs are different. In particular, the Si detectors can be used to measure the ambient neutron background, since the WIMPs are expected to interact rarely with Si (as compared to Ge). Second, the CDMS experiment runs stacks of 6 closely packed detectors, which allows studying multiple-scatter events. In particular, the nuclear-recoil events that happen in two or more detectors should be caused by neutrons, since the probability of WIMPs double-scattering is negligibly small. The multiple-scatter nuclear-recoil events are, therefore, another handle for estimating the muon-anticoincident neutron background. Third, besides the multiple-scatter events that take place in different detectors, one can identify the multiple-scatter nuclear-recoil events that take place in a single detector using the relative amplitudes in the four phonon channels (we discuss this in Chapter 8).

The goal of this Chapter is to introduce some aspects of the CDMS experimental setup. In particular, we will focus on the detector readout (Sections 4, 5, and 6), including the hardware design and the operational aspects, while leaving the detailed discussion of detectors themselves for the Chapter 4. For completeness, we also briefly describe other aspects of the experiment Sections 2 and 3, and we come back to them in more detail in Chapter 7.

3.2 Experimental Sites

Two experimental sites are used by CDMS: the shallow site at the Stanford Underground Facility (SUF) and the deep site at the Soudan mine in Minnesota. The Soudan site is described in detail in Chapter 7.

The Stanford Underground Facility (SUF) consists of 3 tunnels at the south end of the End Station III of the Hansen Experimental Physics Laboratory on Stanford University campus. Tunnel A has been extended and widened to house the CDMS experimental setup. Tunnel B contains the pumps and gas canisters needed for running the CDMS cryostat. Tunnel C houses a clean room where the detectors, cold hardware, and counted materials are stored, away from the cosmic hadronic showers that could activate them.

The Tunnel A has a 10.7 m rock-overburden (16 meters-water-equivalent), which suppresses the hadronic compononent of cosmic rays by a factor of 1000 and the muon flux by a factor of 5. The walls and ceiling of the tunnel are covered with shotcrete, a waterproofing sealant, and paint, retaining the wall of the tunnel, while the floor is covered with concrete which provides a flat platform for the experiment. An air-conditioning unit is used to feed air from the surface through standard and electrostatic filters into the end of the tunnel. In this way, the end of the tunnel is over-pressurized, so the air flows out of the tunnel, reducing the build-up of radon and of dust particulates. The Icebox of the cryostat (see below for a description of the cryostat) is placed at the clean end of the tunnel, and is separated from the rest of the tunnel with a transparent plastic wall, to maintain a higher level of cleanliness. The end part of the tunnel is considered as level 1000 clean room. Detectors and most sensitive parts of the cold hardware are handled in this area, so complete clean suits are required when entering this area.

The cryostat consists of the Oxford 400 dilution refrigerator, capable of achieving 10-20 mK base temperature, and the custom-made Icebox, designed to host the detectors and the cold hardware. Both the fridge and the Icebox have several concentric “cans”, corresponding to different temperature stages: room-temperature, LN temperature, LHe temperature, 600 mK, 50 mK, and 10 mK. The fridge and the Icebox are connected via the fridge-stem, consisting of several concentric cylinders that connect the appropriate stages of the fridge and the Icebox. The Icebox has another extension, called the electronic-stem, through which the signals are brought from the detectors to the room temperature. A more detailed description of the cryostat and the cryogenic system at the Soudan site is given in Chapter 7.

3.3 Passive Shield and Muon-Veto

The CDMS passive shield at the shallow Stanford Underground Facility, consists of several components (the CDMS shield at Soudan is similar, and we describe it in Chapter 7).

The outermost layer is the 5 cm thick plastic scintillator veto, which identifies muons passing into the Icebox. Each side of the Icebox is covered by one or two scintillator panels, with waveshifter bars glued to at least three edges of each panel; the waveshifter bars are then viewed by phototubes at their ends. The only exception is the bottom side of the Icebox, which does not need to be fully covered, as most of the muons are downward going. Hence, only about 70% of the bottom side of the Icebox is covered by four slabs of scintillator, each with one phototube at one end. Finally, a single slab of scintillator, with one phototube at each end, is used to cover the joint between the two slabs covering the top side of the Icebox.

Note that the muons interacting inside the detectors are not a problem, since they typically deposit about 2 MeV of energy, far above the energy region of interest for the dark matter search. It is the muon-induced background that this shield is intended to identify - namely the particles produced in the interactions of muons with the material inside the Icebox.

Note, also, that photons from interactions outside the shield can also leave a signal (up to about 2.6 MeV) in the scintillator. The muons typically deposit about 10 MeV, but the deposition depends on the angle of incidence. In order to keep the high efficiency on muon identification (about 99.95% is needed for the goal of 0.3 kg^{-1}d^{-1}), the threshold on muon veto has to be set relatively low, which implies that some photons could also trigger the muon veto.

Inside the veto is a 15 cm thick lead shield, which suppresses the photon flux from the external sources - these include the rock, as well as various parts of the apparatus:

- The dilution refrigirator used to cool the Icebox contains parts made of stainless steel which contain high levels of ^{60}Co.
- Indium used for various cryogenic seals has an isotope that decays by beta emission.
- Silver in the silver solder (used for many joints in the dilution refrigirator) also has many long lived isotopes.
- The veto phototubes are made of glass.

The lead used for this layer of the shield is off-the-shelf, implying that it contains the isotope ^{210}Pb. This isotope decays into ^{206}Pb via two beta decays - one of the produced betas is very energetic (1.16 MeV), so it emits bremsstrahlung as it passes through lead. This effect limits the thickness of the lead shield to 15 cm.

Going inwards, the next layer is 25 cm thick polyethylene, designed to reduce the flux of neutrons. There are several sources of neutrons [166]:

- Muon interactions in the lead shield, or in the surrounding rock.
- Low energy photo-nuclear interactions in the lead shield, or in the surrounding rock.
- Fission neutrons and (α,n) reaction products.

The thickness of the polyethylene is fixed by the neutron production in the poly itself, which becomes significant at $>$ 25 cm thickness. This layer is very effective for the low energy neutrons (below about 50 MeV), but the neutrons of higher energies are capable of punching through this shield. These higher energy neutrons are produced primarily by the muons interacting in the surrounding rock (and are, therefore, expected to be suppressed at the deep site in Soudan), and (to a smaller extent) in the radioactive processes in the rock.

The Icebox, sitting inside the polyethylene shield, is composed of several concentric cans, operated at gradually decreasing temperatures. Much attention was paid to use only radioactively clean materials for the Icebox. Each can was made of high-purity copper (oxygen-free electronic (OFE) copper), and each piece of copper used in the Icebox was checked for radioactive backgrounds in the LBNL's Low-Background Testing Facility. Furthermore, only brass screws were used for fastening inside the Icebox (copper screws would break, stainless-steel screws are much more contaminated) and kevlar strings are used to suspend the cans inside each other (the standard materials used for cryogenic supports, such as fiberglass, are known to have high contamination levels). Indium, so standard for the cryogenic seals, also has a radioactive isotope; instead copper guskets were used for the seals. The Icebox cans are connected to the cans of the dilution refrigerator through the "fridge-stem". Since the fridge is known to have radioactive isotopes, the F-stem is designed so that there is no direct line of sight between the fridge and the Icebox.

Inside the inner-most Icebox can, operated at the lowest temperature of about 20 mK, there is a 1 cm thick layer of ancient lead. This lead was extracted from a French ship that sunk in the 1800's. Hence, the troubling isotope ^{210}Pb has decayed away and little intrinsic gamma background is produced in this lead. Hence, the main purpose of this lead is to further reduce the gamma background produced in the outer layers of the shield.

Finally, surrounding the detectors themselves is another layer of polyethylene, which further reduces the neutron background by a factor of 2.

3.4 Cold Hardware and Cold Electronics

Cold hardware and cold electronics refer to all the hardware and electronics components mounted inside the cryostat. Good descriptions are given in [168] and [2] - we summarize these here, focussing primarily on the developments in the recent several years.

Because of the unusual combination of requirements in CDMS — low noise, low background, high channel count, and low temperature — new designs of the detector packaging were developed, using new types of materials. The Figure 3.1 depicts the cold hardware design of the CDMS. The heart of the design is the "Tower", consisting of several temperature-stages, and providing coaxial wiring for all the necessary readout lines. The detectors are packaged in "detector holders", and are mounted at the bottom of the Tower in stacks of six. The electrical connections from detectors to the Tower are achieved via

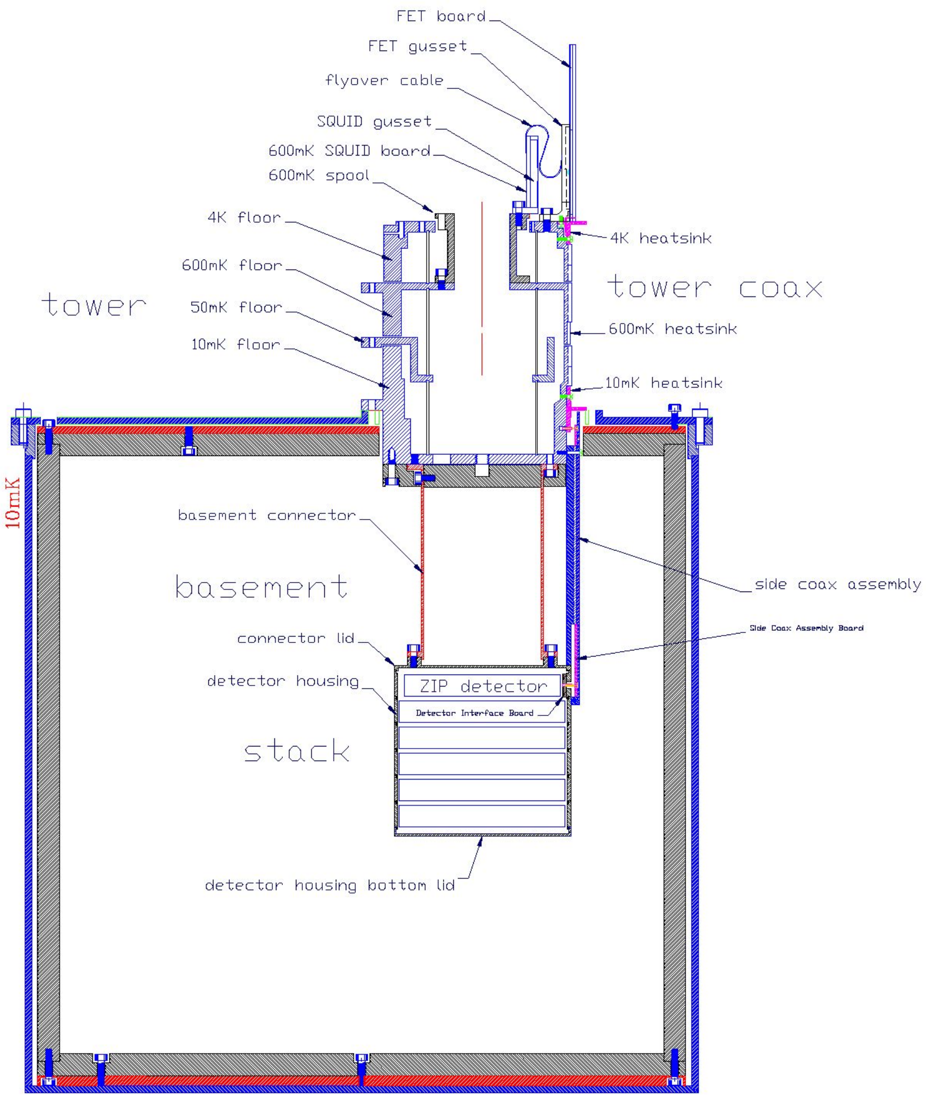


Figure 3.1: Schematic of the cold hardware. The components shown here are described in detail in the text. Figure obtained from [167].

“side-coaxes”. The Tower also hosts “SQUET-cards”, which contain the SQUIDs and the FETs (hence the name) needed for the detector readout. SQUET-cards connect to the room-temperature electronics via “striplines”. There is a one-to-one correspondence between a detector, a side-coax, a Tower face, a SQUET-card, a stripline, and an external “warm” electronics front-end board.

Some of the most important requrements of the experiment are:

- The detectors, made in the form of short cylindrical 250-gram germanium and 100-gram silicon targets, need to be operated at around 20 mK.

- The ionization measurement is accomplished with amorphous silicon and aluminum electrodes on either side of the cylinder faces. The readout is performed using a FET-based, low-noise charge amplifier (see Section 3.6 for more detail). The FETs need to be as close as possible to the detectors and connected through tensioned wire to minimize the capacitance (to ground) and the susceptibility to microphonic pickup.

- Since the FETs dissipate approximately 4.5 mW of power, their heatload can be tolerated on the 4 K stage. The wiring between the FET and the detectors must be heatsunk at each temperature stages. This requirement implies that the Tower needs to bridge the various temperature stages.

- Microphonic pick-up is important for the gate lines of the FETs. Extensive studies indicated that a “vacuum coax” design (where the central conductors were suspended in a shielded channel free of dielectric material) greatly reduces microphonic pickup.

- The connection to the source line of the FET must be of low impedance compared with the transconductance of the FET, but with low thermal loading.

- The measurement of the athermal phonon signal uses a superconducting transition-edge sensor (TES) operated in electrothermal feedback (ETF) mode. These sensors are voltage biased which results in a current signal that is read out with a SQUID-based amplifier. The SQUIDs superconduct below 9 K, so they could be operated at the 4 K stage, but operating them at 600 mK reduces the SQUID noise.

- Given the above requirements, the first amplification stages (for both ionization and phonon signals) had to be inside the dilution refrigerator. After the first amplification, the signals need to be taken to the room temperature electronics through a 10-foot flex cable. Since CDMS II intends to operate 42 detectors, the thermal load of these cables also needs to be kept under control.

- Low background requirement demanded using radiopure materials. In particular, we could not use fiberglass circuit boards (contains ^{40}K) or stainless steel structural elements (contains ^{60}Co). All circuit boards were composed of multi-layer Kapton laminates. Thermal isolation between the stages of the Tower is made with thin-walled high-purity carbon tubes. All solder connections were made with custom-made low-activity solder [169].

- To further minimize contamination with a line-of-sight to the detectors, the detector holders are nearly hermetic copper boxes and the detectors within a stack are mounted close together. The material between detectors is minimized so that background photons and electrons have maximum chance of multiple scattering.

- Finally, since we were designing the detector package to accommodate a full system of 42 detectors with 16 wires per detector, modularity and manufacturability were also important considerations.

Except for the warm end of the stripline, which is outside the radioactive shielding, all of the components of the Towers, stripline, electronics cards and detector packages are made from materials that have been prescreened for U/Th isotopes, with the goal of having < 0.1 ppb of the mass of the material surrounding the detector package (this converts to radioactivity of roughly < 1 μBq/g).

3.4.1 Tower

The Tower consists of four hexagonal copper sections, joined by three thin-walled graphite tubes, and is shown in Figure 3.1. Each copper section is heat-sunk to a stage of the dilution refrigerator: 10 mK (mixing chamber), 50 mK (cold plate), 600 mK (still), and 4.2 K (LHe bath). The graphite tubes are responsible for the dominant heat-load between the different temperature stages. However, the heat-load is sufficiently low that it does not compromise the operation of the cryostat. The top stage of the Tower can host up to six SQUET-cards, and up to 6 detectors can be mounted at the bottom of the Tower. The faces of the Tower carry the wires from the SQUET-cards toward the detectors. The modular design of the Tower allows it to be completely assembled on the bench and then installed as a unit into the cryostat.

An important aspect of the Tower design is the microphonic pick-up in the wire leading from the detector to the gate of the FET [170]. The microphonics comes from the triboelectric effect, where a difference in the work function of two materials in contact results in the charge-buildup at the interface. Such situation occurs, for example, between a conductor and an insulator in a coaxial cable. If there is a microscopic motion of the conductor relative to the insulator, some of the charge at the interface can couple directly or capacitively into the conductor. As a result, the conventional coaxial cable with dielectric insulator cannot be used. Instead, a "vacuum coax" configuration is made by positioning each wire in a machined copper channel in the Tower face. The shield is completed with a thin copper cover, placed over the copper channel in the Tower face. We use copper cladded NbTi superconducting wires, since they provide low impedance without thermally shorting the various stages of the Tower. The connectors at the top and at the bottom of the Tower are made using the MillMax pin/socket combination.

3.4.2 Detector Holder

Two considerations were the most important in the detector holder design. First, the holders are made of counted copper and as thin as possible, in order to minimize the muon-induced neutrons in the material close to the detector. Second, the detectors are packed close together, in order to minimize the solid angle for external sources of radiation to reach the detectors and to maximize the chance that surface contamination on the detectors themselves will produce coincidence events between nearest neighbors. The holders are hexagonal rings, each face corresponding to one face of the Tower. Up to six detector modules are assembled in an interlocking stack, forming a closed copper cavity containing the detectors, with 3 mm spacing between them.

The electrical connection between the detector and the Tower face is made using a copper-based side-coax assembly. Similarly to the Tower faces, each signal in the side-coax is brought by a copper cladded superconducting NbTi wire. The clad is etched in the central region of the wire to provide thermal isolation. Each channel goes through a vacuum coax (similar in design to the vacuum coaxes in the Tower faces), in order to minimize the amount of cross-talk between channels, as well as to avoid the microphonics that could be generated in insulators. In addition to this, the side-coax carries some of the electronic elements needed for the ionization signal read-out. On both ends of the side-coax, the NbTi wires end on solder-coated copper traces from which the connection is made using MillMax pin connectors.

On the detector side, the side-coax connects to a Detector Interface Board (DIB), which is mounted on the detector holder. The main purpose of this board is to receive the signals from the side-coax through a MillMax connector and lead them (via solder-coated copper traces) to the aluminum pads, from which wire-bond jumps are made onto the surface of the detector. The DIB is made of cirlex with 1-ounce copper traces on each side. Due to the limitted space, it was convenient to place the aluminum pads on the horizontal edge of the board, which was a bit of a technological challenge since the tinned copper traces had to extend around the edge.

In addition, the DIB hosts two infrared LEDs needed for flashing (i.e. neutralizing) the detector. The detector itself is held in place by six cirlex feet screwed onto the holder (three on each side). The grip is made tight at room temperature and it only strengthens as the detector is cooled due to the thermal shrinking of the holder.

3.4.3 Cold electronics

In the following two Sections we describe the cold electronics cards that contain the FETs for the ionization and the SQUIDs for the fast phonon measurement. The FETs are placed on the FET-card residing on the 4 K stage of the Tower, while the SQUIDs are placed on the SQUID-card mounted on the 600 mK stage of the Tower for better noise performance. The FET-card and the SQUID-card are connected via flyovers, and are commonly referred to as the SQUET-card.

FET Card

The primary role of the FET-card is to host FETs needed to read out the ionization signal. In addition, this card also serves as a through-path from the stripline to other connections and components on the detector holder or side-coax, such as bias resistors, coupling capacitors, and LEDs (used to neutralize the detector; see Chapter 4). Also, the FET-card interfaces with the SQUID-card, which hosts SQUIDs used to read out the transition-edge sensors.

Figure 3.2 is a photograph of the FET-card and Figure 3.3 is a drawing showing different features on the FET-card. The FET-card is effectively a small circuit board made of 0.062" thick cirlex (including three layers of the 1-ounce copper), with bracing for mechanical support. The electrical connections on the surface of the card are of 1-ounce copper, patterned using standard techniques. The card is mounted at the top of the 4 K stage of the Tower using a copper bracket that makes an effective thermal connection to the bath. The upper section of the FET-card has a 50 pin array of MillMax connectors

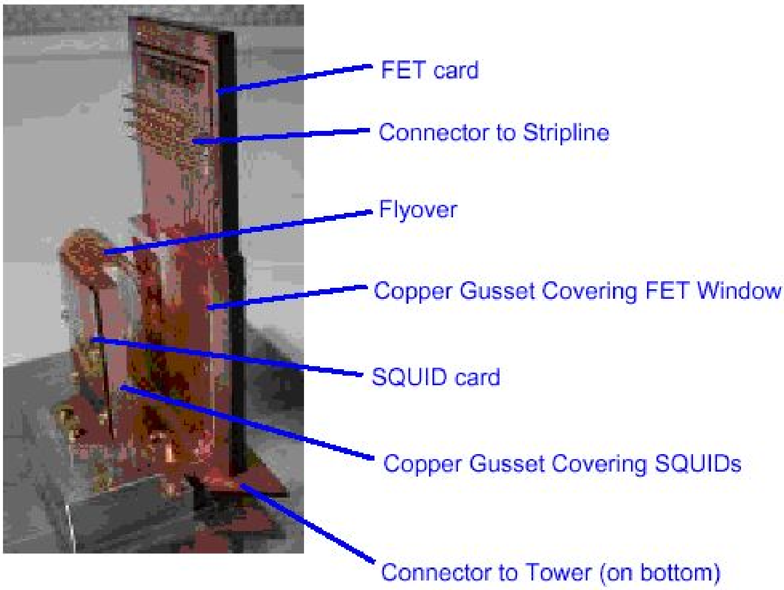


Figure 3.2: Photograph of the SQUET-card used in the CDMS II readout. The SQUET-card is composed of the FET-card (larger, on the right) and SQUID-card (smaller, on the left), connected by the flyover. The Kapton window is under the copper gusset on the FET-card.

that mates with a connector paddle on the stripline, which leads the signals to the warm electronics at room temperature. The bottom edge of the card has a 16 pin array that mates with the Tower face leading toward a detector.

The central part of the FET contains a cavity where the FETs are operated. The cavity is coated with infrared absorber to absorb the radiation from FETs. Since FETs are optimally operated at 150 K (see below), the two FETs needed to readout one detector are mounted in the center of 0.006"-thick kapton window. The dimensions of the window are carefully chosen so that the power dissipated by the FETs (about 4.5 mW each) self-heats them to the optimal temperature. Since the FETs are not turned on when cold, a startup

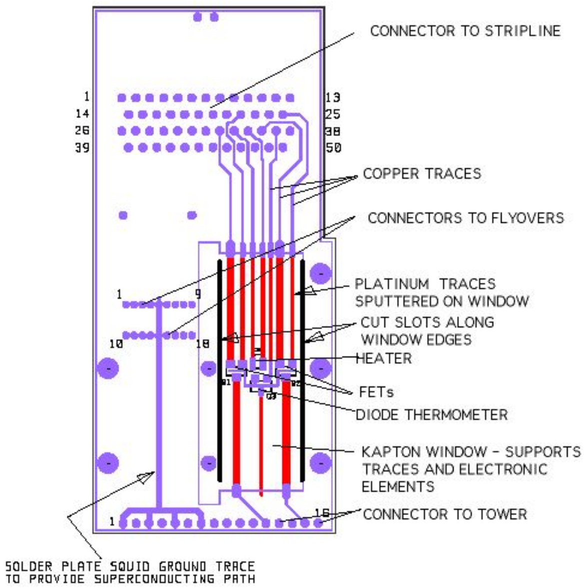


Figure 3.3: This drawing of the FET-card describes the geometrical relationships between the Kapton window and the stripline, flyover, and Tower connectors. It also shows the design of the Kapton window. Figure obtained from [167].

heater (a 1 kΩ resistor), as well as a diode thermometer, are included at the center of the window. All elements are connected to copper pads in the kapton window using silver-filled epoxy.

The FETs, the heater, and the thermometer are connected to the copper circuit board on the surface of the FET-card via copper traces of approximately 3 μm thickness. The width of the traces is carefully chosen so that they do not thermally short the center of

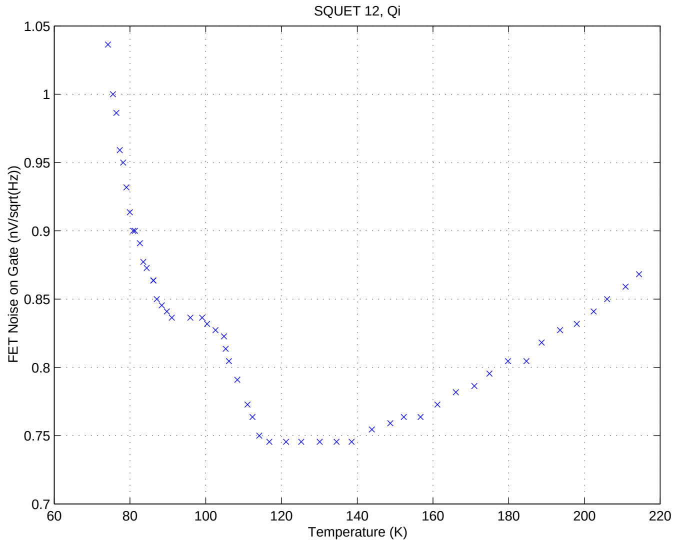


Figure 3.4: JFET noise as a function of temperature. The noise measurements were taken at 10 kHz. At 120-140 K, the JFET noise is the lowest (~ 0.75 nV/$\sqrt{\text{Hz}}$). At 50 kHz, the trough reaches 0.5 nV/$\sqrt{\text{Hz}}$. The design of the FET-card window forces the FETs to self-heat at 150 K.

the window to the FET-card itself. The Kapton window is treated as a separate assembly, and it is glued onto the FET-card. The electrical connections between the FET-card traces and the window traces are achieved with solder joints.

We verified that this design yields optimal noise performance of the FET. In particular, we measured the noise of a FET mounted on a FET-card window as a function of temperature (see Figure 3.4). Indeed, the FETs self-heat to an optimal temperature of 150 K, where the FET noise is at its minimum. Furthermore, the FETs are sufficiently far from the freeze-out temperature (~ 115 K).

We observed occasional difficulties with the kapton window sagging during the processing. To avoid these difficulties, we cut out the kapton window and instead glue (along the horizontal edges) a separate piece of 0.004" kapton to the board. This design does not change the noise considerations discussed above, so the FETs are still operated at the optimal temperature of 150 K.

The remaining space of the central part of the FET-card is then used for an interface with the SQUID-card (see Figures 3.2 and 3.3). The two cards are connected via two “flyovers”, each corresponding to a pair of SQUIDs. Each flyover is composed of pairs of twisted (to reduce cross-talk) copper-nickel cladded, formvar-insulated, niobium-titanium (NbTi) wires of 0.009” diameter (with 0.005” core diameter) sandwiched between two 0.001” thick layers of kapton. Since the NbTi wires are superconducting at these temperatures, the flyover thermally isolates the two cards. Copper-nickel clad is used to make soldering possible, without creating a thermal short. Furthermore, the kapton makes the design relatively robust. On both ends of the flyover, the wires are integrated into the circuit via solder joints.

SQUID-Card

The role of the SQUID-card is to host four SQUIDs necessary to read out the transition-edge sensors. This card is made of 0.032” cirlex with 1-ounce copper circuit boards on each side. The card is connected to the 600 mK stage of the Tower, which allows operating SQUIDs at low noise (see below for more detail). As described above, the signals are brought to the SQUID-card by the flyover. Since one flyover corresponds to two SQUIDs, each signal wire in the flyover is twisted with a ground wire in order to reduce the amount of cross-talk between channels. The ends of the flyover wires are soldered to the solder-coated copper traces on the surface of the SQUID-card. The copper traces lead to the aluminum spuddered pads from which wire-bond jumps are made to the SQUID chips. The SQUIDs are glued to the 0.002” thick niobium foil, which serves as a magnetic shield, using Hardman epoxy. We find that this epoxy is reliable under thermal stress, while it can still be removed without damaging the SQUID chip. The niobium foil is, in turn, glued to the surface of the board, without shorting to the copper layer on the surface of the board. The SQUIDs are mechanically prottected using a copper gusset, which is also used as a heat-sink.

3.4.4 Stripline

The signals from the FET-card are brought to the room temperature using a stripline. Since CDMS II is expected to run 42 detectors, each requiring 50 wires to be brought out to the room temperature, the ease of manufacturability and the thermal sinking of the wires were the most important consideration in the design of the stripline. The final design uses a multi-layer shielded flex circuit of approximately 1-inch width and 10-foot length made of copper and kapton. This design allows a large stack of striplines to be heat-sunk using copper press-plates and shims, while still allowing a relatively straighforward production and installation.

The stripline's signal layer is made up fifty 0.005-inch wide 1/2-ounce thick copper traces on 0.015-inch centers. It is sandwiched between two 0.005-inch thick kapton layers, two sputtered copper ground planes, and two 0.001-inch thick kapton layers. The cold end of the stripline ends with a connector paddle, holding an array of 50 MillMax sockets that mate with the similar array of MillMax pins on the FET-card. The warm end of the stripline is soldered to a standard 50-pin D-connector. The striplines are heatsunk at the LN temperature (~ 15 cm from the room temperature) and at the LHe temperature (~ 100 cm from the LN heat-sink). The lengths are a compromise between thermal loading and increased resistance. Furthermore, to minimize the infrared leakage, the striplines pass through a set of copper baffles (painted with the IR-absorbing epoxy mixture) on their way from the room-temperature to the FET-cards.

3.5 Warm Electronics

In this Section, we briefly describe what happens to the signals as they come out of the cryostat.

The signals from a detector, brought out by a stripline, end at a bulk-head connector at the end of the electronics-stem of the cryostat. From this connector, they are taken via the detector I/O cable to a Front-End Board (FEB). The FEB performs a number of functions:

- It has complete control over the state of the detector: biasing of the SQUIDs and the phonon sensors, biasing or grounding of the charge electrodes etc - see Section 3.6 for more detail on the sensor operation.

- It can perform flashing/baking of the LEDs, necessary to keep the detectors neutralized (see Chapter 4 for more detail on the LED flashing).

- It can perform "zapping" of a SQUID, by sending a large current pulse through the SQUID, allowing the possible flux trapped in the SQUID to be released.

- It can heat the FETs by sending current through the heater-resistor on the Kapton window of the relevant SQUET-card.

- It receives the signals from the detectors, amplifies them, and sends them further down the electronics chain.

The detector signals are further transfered from the FEB, via a detector cable, to a Receiver-Trigger-Filter (RTF) board. The main function of the RTF board is to create the trigger signals based on the (filtered) detector signals. The trigger thresholds are tunable. Triggers are created on the sum of the four phonon signals of the given detector and on the sum of the charge signals of the given detector. The trigger signals are then sent to the Trigger Logic Board (TLB) which creates the global trigger based on the triggers from different RTF boards, corresponding to different detectors. The global trigger created by the TLB activates the digitizers, which then record the traces supplied by the RTF boards. The digitized traces are saved on a computer hard-disk.

The FEB, RTF board, and the TLB are all controlled by a computer using a General Purpose Interface Bus (GPIB). This is achieved using the data-acquisition software, which, in turn, has evolved over the years. The WIMP-search runs at the SUF used Labview software running on Macintosh computers, while the Soudan setup relies on the Java-based software runing on Linux PC's. In addition to controlling various electronics boards, the data-acquisition software also monitors various aspects of the experiment, such as the trigger rates, various DC offsets, cryogen levels in the cryostat etc. The data-acquisition system at Soudan is described in more detail in Chapter 7.

3.6 Detector Readout Performance

In this Section we discuss the performance of the ionization and the phonon readout system.

3.6.1 Ionization Channels

In this Section we describe the ionization-signal readout system. A more detailed description of this system is given in [171], and here we briefly describe the principles of operation and the noise performance. The circuit used to read out the ionization signal can be simplified as shown in Figure 3.5.

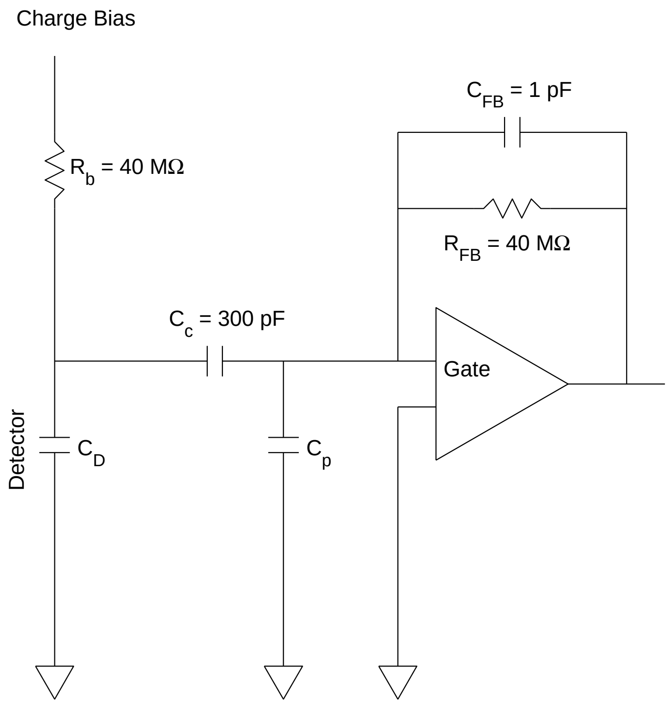


Figure 3.5: Charge readout circuit. The heart of the amplifier is a JFET. The parasitic capacitance has been measured to be $C_p = 100$ pF and the detector capacitance is typically $C_D = 50$ pF. "Gate" denotes the gate of the JFET inside the amplifier.

A voltage bias (typically around 3 V) is used to separate the electron-hole pairs that are created in the Ge (Si) crystal when an interaction takes place. This charge Q creates a voltage across the coupling capacitor C_c (and at the gate of the JFET). The amplifier reacts by charging the feedback capacitor C_F to bring the voltage level at the gate to zero. The feedback capacitor is subsequently discharged through the feedback resistor R_F. The voltage ionization signal is, therefore, created at the feedback capacitor and its amplitude (to zeroth order) is $V = Q/C_F$. The fall-time of the ionization pulses is set by the $R_F C_F$ circuit and it is typically 40 μs. The rise-time of the ionization pulses is determined by the details of the amplifier and it is typically around 1 μs. The noise at the input of the amplifier (that is, at the gate of the FET) is amplified by the ratio of the total gate-to-ground capacitance to the feedback capacitance C_F. The total gate-to-ground capacitance is a combination of the detector capacitance (typically $C_D \approx 50$ pF), coupling capacitance ($C_c = 300$ pF), and the effective parasitic capacitance (measured $C_p \approx 100$ pF).

The heart of the amplifier is a JFET. The primary advantage of the JFET is the high input impedance, which allows good charge collection efficiency. Moreover, the JFET is characterized by low noise performance. The total noise of the circuit shown in Figure 3.5 is given by the following expression [170]:

$$e_o^2 = |A(f)|^2 \left\{ e_{FET}^2 \left((C_D + C_p + C_F)^2 (2\pi f)^2 + \left(\frac{1}{R_F} + \frac{1}{R_b}\right)^2 \right) + 4kT\left(\frac{1}{R_F} + \frac{1}{R_b}\right) \right\}, \tag{3.1}$$

where

$$A(f) = \frac{R_f}{1 + 2\pi i f R_F C_F}, \tag{3.2}$$

k is the Boltzmann constant, $T = 20$ mK is the temperature of the bias and feedback resistors, and e_{FET} is the voltage noise of the JFET (it is typically 0.5 nV/$\sqrt{\text{Hz}}$ at 50 kHz, as discussed in Section 3.4.3). Equation (3.1) predicts that at frequencies above roughly 1 kHz the JFET noise dominates the total output noise of the system. Furthermore, at these frequencies the noise spectrum should be flat at $\sim$ 100 nV/$\sqrt{\text{Hz}}$, which has been experimentally confirmed (see Figure 3.6). The bandwidth of the system is set by the amplifier; the -3 dB point is at $\sim$ 160 kHz.

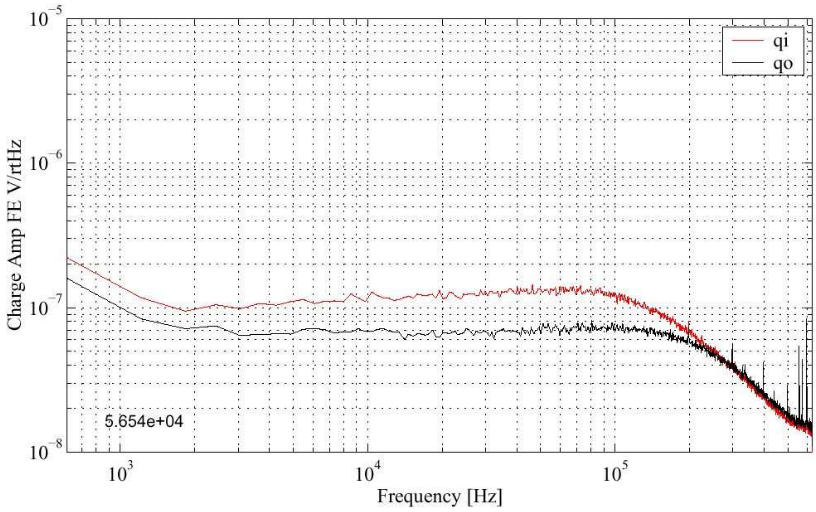


Figure 3.6: Typical noise spectra obtained while operating ZIP detectors. Two charge channels on the same detectors are shown, corresponding to the inner and outer charge electrodes (see Chapter 4 for the description of the charge electrodes).

To estimate the lowest signal observable by this readout system, one can integrate the power spectral density to obtain the variance of noise of $\sigma^2 \approx 2 \times 10^{-9}$ V^2. Define the lowest observable signal to have amplitude of $2\sigma \approx 0.1$ mV, corresponding to 0.1 fC of charge on the feedback capacitor. Since 3.8 eV is needed to create an electron-hole pair in Si (3 eV in Ge), the lowest energy that can be observed by the ionization channel on a Si detector is $\sim$ 2.5 keV ($\sim$ 2 keV for a Ge detector). We will confirm this estimate in Chapter 8 in the first Soudan WIMP-search data.

3.6.2 QET phonon channels

Principles of Operation

The CDMS II experiment uses the QET sensors to measure the athermal phonon signal. The readout system designed for this type of sensors is based on SQUIDs. The low noise characteristics of the SQUIDs make them suitable for amplification of very small current signals. However, the read-out system used for the phonon signal is consequently different from the one used for the ionization signal. The simplified schematic is shown in Figure 3.7.

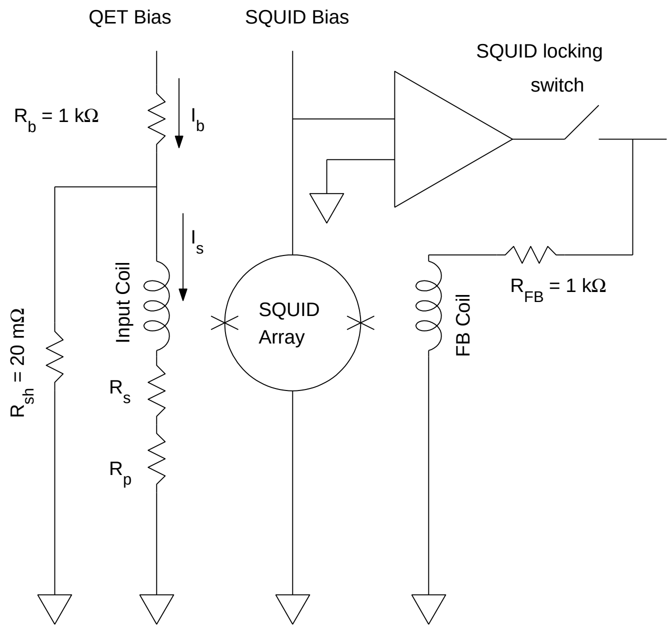


Figure 3.7: The phonon channel readout system. The system is in the open-loop mode if the switch is open and in the closed-loop mode is the switch is closed.

As mentioned above, the phonon signal is created using a tungsten based Transition Edge Sensor (TES). Briefly, the phonons generated in an event in the crystal cause an increase in the temperature of the sensor, which is kept on the superconducting transition edge. Hence, its resistance (denoted R_s in the Figure 3.7) increases. Since the typical operating value of R_s is ~ 100 mΩ, the shunt resistor $R_{sh} = 20$ mΩ keeps the sensor voltage biased. It follows, then, that a change in R_s is equivalent to a change in the current flowing through the input coil of the SQUID, which is, in turn, equivalent to the voltage across the SQUID (that is, at the input of the amplifier on Figure 3.7). Furthermore, the ratio of the number of turns in the input coil to the number of turns in the feedback coil of the SQUID is 10. Hence, the current signal is amplified by a factor of 10 at this stage, in the closed-loop mode.

A phonon pulse is characterized by a rise-time of typically 10 μs and a fall-time of typically 200 μs. However, these characteristic times strongly depend on the type of the detector (the phonon signal is about two times faster in Si than in Ge) and of the event itself, as well as on the charge bias voltage which affects the energy carried by the Luke phonons. We will discuss the phonon pulse-shape in detail in Chapters 5, 6, and 8.

SQUIDs

The SQUIDs used by CDMS are produced by NIST at Boulder, Colorado. They are characterized by:

- Turns ratio: ratio of the number of turns in the input coil to the number of turns in the feedback coil. It is designed to be 10.

- Current-per-Φ_0: current in input or feedback coil corresponding to one quantum of flux in the SQUID. It is 25 μA in the input coil or 250 μA in the feedback coil.

- Self-inductance of the input coil: $L = 0.25$ μH.

- V-Φ curve: variation of the voltage across the SQUID with the change of flux through the SQUID. The maximum peak-to-peak amplitude of the $V - \Phi$ curve (also called modulation depth) is typically 5 mV.

Another important characteristic of the SQUIDs is the responsivity. Responsivity $r = dV_{sq}/dI$ is defined as the change in voltage across the SQUID for a given change in current (either in the input or in the feedback coil). Equivalently, it is given by the slope of the $\mathrm{V} - \Phi$ curve. For the NIST SQUIDs, responsivity of $500 - 1000\ \Omega$ referred to input (RTI) is easily achievable.

The responsivity is important for two reasons. First, the input noise of the amplifier (used to amplify the SQUID voltage) is amplified by the factor R_F/r, where $R_F = 1000\ \Omega$ is the feedback resistance. Hence, large responsivity reduces the impact of the amplifier noise and increases the signal-to-noise ratio. Second, the open-loop gain of the amplifier (which includes responsivity as one of the gain factors) sets the bandwidth of the amplifier chain in the closed loop mode. The amplifier is designed to give ~ 2 MHz bandwidth for $r = 1000\ \Omega$ RTI, but this value varies for different values of responsivity.

Another issue of importance are the resonances of the SQUIDs. The resonances appear as distortions of the $\mathrm{V} - \Phi$ curve and they tend to increase the SQUID noise significantly, so it is desirable to avoid them. The SQUID resonances are well known and have been studied (see, for example [172]). There are two types of resonances that appear in the SQUIDs used by the CDMS experiment.

First, there is a “feedback”-type resonance. In this case, the current through the tunnel junctions capacitively couples to the coils, which then feeds back inductively into the SQUID. On one slope of the $V - \Phi$, the two effects will cancel each other out. On the other slope, however, the feedback will be positive and a resonance will be created. Note that this kind of resonance can appear only on one slope of the $V - \Phi$ curve.

Second, there is a “junction”-type resonance. In this case, the bias of the SQUID introduces a voltage drop across the SQUID, which is, in turn, related to the frequency of the currents flowing through the SQUID loops. Changing the SQUID bias changes this frequency. If the frequency matches the natural frequency of the system, a resonance appears. Hence, this kind of resonance appears only at a particular, usually relatively large, SQUID bias voltage.

Hence, both types of resonances can be avoided by applying relatively low SQUID bias and by locking the SQUID on the side of the $V - \Phi$ which has no feedback-type resonances.

Few words about the dynamic range. In the closed loop mode the system is locked at a single point on the $V - \Phi$ curve, or, equivalently, at a particular value of the current in the input coil. In practice, however, the amplifier is not infinitely fast, so the error signal, defined as the difference of a signal I_i in the input coil and the amplifier response I_r (as seen in the input coil), $e_i = I_i - I_r$, is non-zero. In particular, if the error signal during a phonon pulse is too large, it could cause losing lock and the SQUID would jump one cycle of the $V - \Phi$ curve *during the pulse.* This, in turn, would distort the current pulse and affect the calorimetric measurement.

For the circuit shown in Figure 3.7, the error signal is given by

$$e_i = \frac{I_i}{1 + \frac{G_{OL}}{10^4 \Omega}}, \tag{3.3}$$

where G_{OL} is the frequency-dependent open-loop gain of the amplifier (including the responsivity RTI) and it is given by

$$G_{OL} = \frac{6000\ r}{\sqrt{1 + (\frac{f}{2.4\ \mathrm{kHz}})^2}}. \tag{3.4}$$

For example, for responsivity $r = 1000\ \Omega$ RTI and at 100 kHz (which is higher than the relevant frequency range for the signal), the error signal amounts only to about 6% of the input signal. Given that the typical phonon signal has amplitude of $\sim 1 - 10\ \mu$A, the error signal is not expected to exceed 1 μA, which has been experimentally verified. Furthermore, since the input coil current corresponding to Φ_0 (or, equivalently, to one $\mathrm{V} - \Phi$ cycle) is $\sim 25\ \mu$A, it is clear that the error signal is sufficiently small to keep the feedback stable.

SQUID Noise

There are several contributions to the total output noise of the phonon channels. We separate the SQUID noise as particularly interesting.

The noise of the SQUID comes from the Johnson noise of the shunt resistors used to dampen the resonant behavior. Since in the CDMS experiment the SQUIDs are operated at 600 mK, it is expected that they have low noise performance.

The noise of the SQUID can be thought of as the magnetic flux noise or, equivalently, as the current noise in the input coil. In order to measure this noise, we made $R_s = \infty$ (that is, we opened the circuit on the input side of Figure 3.7). Then, the total output noise has only two contributions: the SQUID noise itself and the input amplifier noise. The input amplifier noise has gain of R_F/r (referred to feedback), whereas the SQUID noise is amplified by the turns ratio and the R_F resistor. Hence, by measuring the total output noise at different values of responsivity, we can separate the two noise contributions.

We changed the value of responsivity by locking the SQUID at different points on the $V - \Phi$ curve (both slopes). At each point, we measured the noise in the flat region of the spectrum ($5 - 10$ kHz). The results are shown in the Figure 3.8.

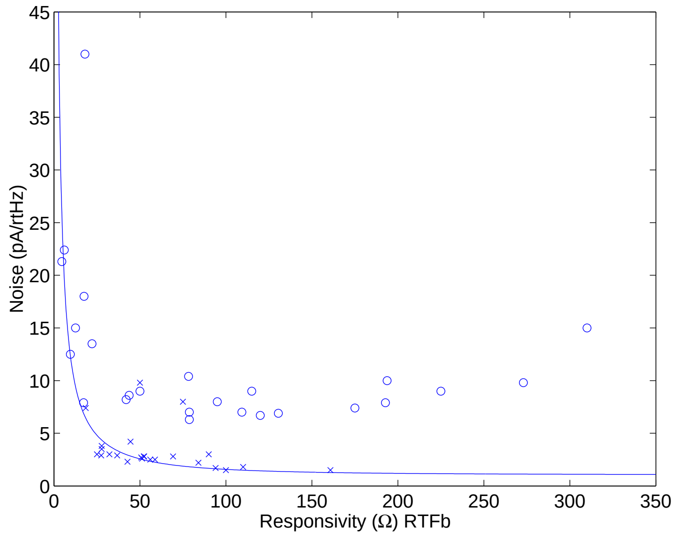


Figure 3.8: Noise vs responsivity plot. See text for detailed explanation.

The o's correspond to the measurements taken on the "bad" slope of the $V - \Phi$ curve (containing the feedback type resonances). The $\times$'s correspond to measurements taken on the "good" side of the curve. It can be observed that the noise on the "bad" side of the $V - \Phi$ curve is typically larger than that on the "good" side. Several points taken on the "good" side of the $V - \Phi$ curve also exhibit large noise level. However, all of these measurements were taken with SQUIDs locked very close to the junction-type resonance.

Keeping only the points far from the resonances on the "good" side of the $V - \Phi$ curve, the least-square fit gives the SQUID noise of 1 pA/$\sqrt{\text{Hz}}$ RTI and the input amplifier noise of 1.2 nV/$\sqrt{\text{Hz}}$ (this estimate is consistent with the amplifier noise measurement on the bench).

Further notice that the responsivity has a strong impact on the contribution of the amplifier noise. In particular, $r = 100\ \Omega$ RTFb (or $1000\ \Omega$ RTI) makes the amplifier noise contribution comparable to the SQUID noise contribution.

Total Noise

The total output noise of the phonon channel has several contributions. As shown in Figure 3.7, the resistances R_s, R_{sh}, and R_p contribute with their Johnson noise; the SQUID and the amplifier also must be taken into account. The total current noise in the input coil is given by [173]

$$i_n^2 = 4k\,\frac{R_sT_s + R_{sh}T_{sh} + R_pT_p}{(R_s + R_p + R_{sh})^2} + i_{SQUID}^2. \tag{3.5}$$

Here, $T_s = 65$ mK, $T_{sh} = 600$ mK and T_p are the temperatures of the sensor, the shunt, and the parasitic resistances respectively and k is the Boltzmann constant. The exact temperature of the parasitic resistance depends on its physical position in the cryostat.

The total output voltage noise is then given by

$$v_n = \sqrt{i_n^2(10R_F)^2 + v_e^2\left(\frac{R_F}{r}\right)^2}, \tag{3.6}$$

where v_e is the input voltage noise of the amplifier. Typical parasitic resistance is $R_p \approx$ 15 mΩ and the shunt resistance is $R_{sh} = 20$ mΩ. In the usual mode of running $R_s =$ $100 - 200$ mΩ. This value of the sensor resistance is determined by the maximization of the signal-to-noise ratio, which requires biasing the sensors at the optimal-slope point on the transition edge. Furthermore, the shunt resistance is the dominant voltage noise source, although R_s strongly affects the total current noise in the input coil. For $r = 1000\ \Omega$ RTI, the Equations (3.5) and (3.6) imply the total output noise of 10 pA/$\sqrt{\text{Hz}}$ RTI in the flat region of the spectrum. This has been also observed, as shown in Figure 3.9.

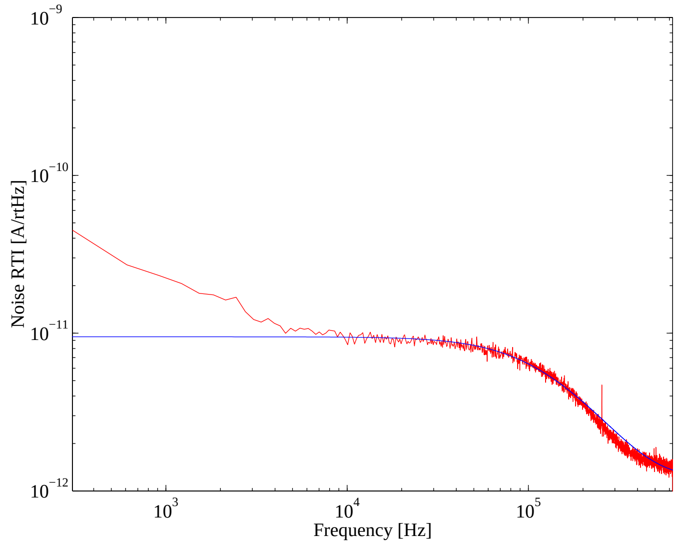


Figure 3.9: Total noise in the phonon channel observed in the usual mode of running. The thin noiseless line is the theoretical prediction based on $R_s = 120$ mΩ and $L = 0.25$ μH. The observed roll-off is faster than $1/f$ due to the elements of the electronics chain at the room temperature.

Bandwidth

When discussing the bandwidth, it is useful to separate two noise components. The current noise in the input coil of the SQUID is rolled off by the L/R filter on the input side of the SQUID. Here $L = 0.25$ μH is the self-inductance of the input coil and R is the combined resistance on the input side, dominated by R_s in the usual mode of operation (as discussed above, optimization of the signal-to-noise ratio leads to $R_s = 100 - 200$ mΩ). Therefore, the current noise in the input coil is frequency dependent at the output and it has the -3 dB point at 65 - 130 kHz, after which it has $\sim 1/f$ dependence (other steps in the electronics chain make this roll-off somewhat faster). The second noise component is frequency independent in the output and it is composed of the amplifier noise and the digitizer noise. It contributes 5 pA/$\sqrt{\text{Hz}}$ or less, depending on the digitizer. Hence, the total output noise is dominated by the amplifier and digitizer noise components at frequencies much above 130 kHz. However, at these frequencies, the signal-to-noise ratio is typically already negligible for most phonon pulses. This shape of the noise spectrum is experimentaly confirmed (see Figure 3.9).

Note, also, that the roll-off described above is dynamic. In other words, during an interaction R_s first increases (since the sensor is kept at its superconducting transition edge) and then falls back to its original value. Hence, the -3 dB point of the L/R filter first increases during a phonon pulse and then falls back to its original value.

Furthermore, the amplifier response is not exactly frequency independent. Its response depends on the responsivity and for reasonable $r = 1000\ \Omega$ RTI, the -3 dB point of the amplifier is at 2.1 MHz. This is far above the frequency range of the signal, so the frequency dependence of the amplifier can be neglected.

As in the case of ionization channels, one can attempt to estimate the lowest energy observable by the phonon readout system. Integration of the power spectral density gives the noise variance of $\sigma^2 \approx 10^{-17}$ A^2. Hence, the smallest observable signal has amplitude of $2\sigma \approx 6$ nA. For a typical voltage bias of 5 μV, this corresponds to 30 fW power pulse. And, for a pulse with characteristic decay time of ≈ 300 μs, this pulse corresponds to the energy of ≈ 60 eV. Hence, the lowest signal observable by the setup described above has energy of 60 eV. Note, however, that this is not necessarily the lowest observable recoil-energy of an event. The observed phonon signal strongly depends on the value of the charge bias (which affects the energy carried by the Luke phonons) and on the phonon collection efficiency - both of these issues are discussed in Chapter 4. Typically, the lowest observable recoil-energy is a few hundred eV, as discussed in Chapter 8 for the first WIMP-search data taken at the deep site in Soudan, MN.

Chapter 4

CDMS II Detectors

4.1 Introduction

The previous Chapter described the CDMS II experimental setup and the detector readout system. In this Chapter we focus more on the detectors.

CDMS II uses the Z-sensitive Ionization and Phonon (ZIP) detectors. These detectors are based on crystals of Ge or Si, cooled down to about 20-30 mK. At these temperatures, an interaction inside the crystal leaves a two-fold signature. First, the interaction breaks the electron-hole pairs of the semiconductor crystal. An electric field through the crystal then separates these electrons and holes before they could recombine - this is the ionization signal. Second, the interaction inside the crystal "shakes" its lattice. We detect the modes of these oscillations, or phonons, using Transition Edge Sensors (TES), operated in the Electro-Thermal Feedback (ETF) mode.

This Chapter is organized as follows. In Section 2 we discuss in more detail the physics inside the ZIP detectors, in order to point out several important issues. Section 3 describes the actual fabrication and testing procedures that each detector passes through before it is chosen for use in the actual dark matter experiment. The remainder of the Chapter describes several techniques we have developed in order to extract various information from the detectors. In particular, Section 4 describes algorithms for event-position reconstruction, and Section 5 describes different ways of studying the phonon physics inside these detectors.

4.2 ZIP Detectors

4.2.1 Phonon Signal

Phonons in the Crystal

A particle interaction in the crystal causes either an electron or a nucleus to recoil, which in turn disturbs the crystalline lattice. The particle representation of this acoustic phenomenon are phonons.

Phonons are created during an interaction at the Debye energy (about 10 THz in the Ge and Si crystals). Their subsequent behavior is named quasi-diffusion and it is dominated by two processes: the anharmonic decay and the isotope scattering. In the anharmonic decay, a single phonon decays into two phonons of smaller energy. The rate of this process goes as the -5^{th} power of frequency:

$$\tau_d \propto \left(\frac{1\mathrm{THz}}{\nu}\right)^5. \tag{4.1}$$

At $\nu = 1$ THz, the anharmonic decay time is roughly 10μs (for Si) [174]. The isotopic scattering is an elastic process occurring because the Ge and Si crystals are not mono-

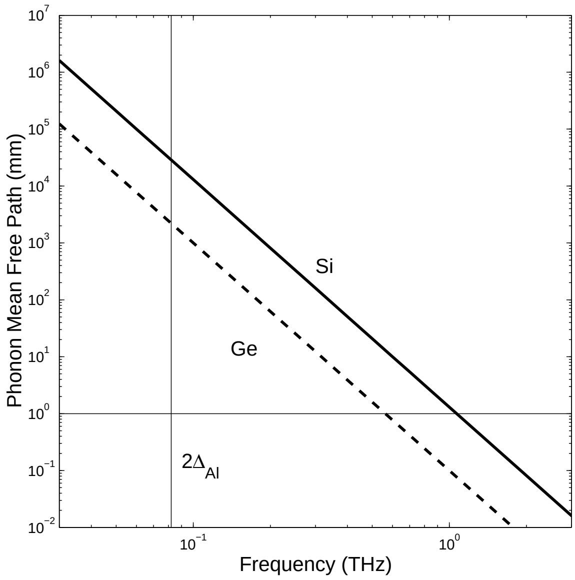


Figure 4.1: Phonon mean free path as a function of phonon frequency in Ge and Si crystals. Figure reproduced following [176] and [177].

isotopic. The time constant for this process goes with the -4^{th} power of frequency [175]:

$$\tau_I \propto \left(\frac{1\text{THz}}{\nu}\right)^4. \tag{4.2}$$

At $\nu = 1$ THz, the time constant is 0.41μs (for Si). Hence, the phonons created during an interaction are of very high energy and their mean free path is very short as the time constant for the isotopic scattering is very short. However, they very quickly undergo a sequence of anharmonic decays - the produced lower-energy phonons have longer mean free paths. Within a few microseconds, phonons have decayed to energy below 1 THz. As shown in Figure 4.1, such phonons have the mean free paths similar to the dimensions of the detector. Hence, we refer to such phonons as ballistic. These phonons can freely travel to the surface of the crystal. If they hit the bare surface, they simply reflect; if they hit the superconducting Al fins on the surface of the detector, they can contribute to the final phonon signal (see below).

Another class of phonons is created by the electrons and holes drifting through the crystal. As described in the Section 4.2.2, the interaction in the semiconducting crystal also creates electrons and holes; a small electric field across the crystal then separates these electrons and holes. As they drift through the crystal, they are accelerated by the electric field and they scatter with the lattice, thereby depositing some of their energy into the phonon system. This effect has been independently discovered by Neganov-Trifomov [178] and by Luke [179]. The remaining energy carried by the electrons and the holes is also deposited in the phonon system when the charges reach the electrodes - these are referred to as the relaxation phonons. Hence, the energy deposited in the phonon system in this way is simply the product of the total charge Q (released in the initial interaction) and of the electric potential V. Since it takes about 3 eV of energy to break an electron-hole pair in Ge (3.8 eV in Si), and we typically use electric potential of about 3 V, the energy carried by the Luke phonons is very similar to the energy carried by the initial interaction phonons. Furthermore, the energy spectrum of these phonons can be calculated [180] - it is predicted to be ≈ 0.25 THz for the electric field values we use. Hence, these phonons are ballistic.

Two aspects of the phonon signal are particularly interesting. First, since the ballistic phonons are created within a few microseconds after the interaction, one can attempt to cover the whole detector surface with phonon sensors and then use the timing information to reconstruct the position of the event. This will be discussed in detail in the Section 4.4. Second, for events that happen very close to the surface of the crystal, the surface imperfections can accelerate the down-conversion of phonons. In other words, the ballistic phonons are produced more quickly for surface-events than for the bulk-events.

As discussed in the Section 4.2.2, the surface electron-recoil events can have incomplete charge collection, and they can mimic the nuclear-recoil events. Hence, the start-time and the rise-time of the phonon signal can be very useful in discriminating against the surface events - this is discussed in more detail in Chapters 5, 6, and 8.

Quasiparticle Trapping

The energy carried by the phonons that reach the surface of the crystal has to be captured and converted into electric signals. This is achieved by W transition-edge sensors (TES) operated in the electro-thermal feedback (ETF) mode, and assisted by Al quasiparticle traps.

The idea of quasiparticle trapping was first proposed by N. Booth [181]. In the ZIP detectors, it is implemented using the superconducting Al fins. A schematic of quasiparticle creation and diffusion is given in Figure 4.2. The phonons entering the Al fins can break the Cooper pairs and create quasiparticles if their energy is larger than $2\Delta_{Al} = 340\ \mu$eV, or equivalently, if their frequency is larger than 82 GHz. Since the phonons in the Si and Ge crystals are ballistic below ~ 1 THz, the phonons reaching the surface of the Si or Ge crystals are sufficiently energetic to produce quasiparticles in the Al at the surface of the crystal. Of course, energy in phonons below 82 GHz is lost (this is about $5 - 10\%$ of the initial energy, [182]). The quasiparticles then must diffuse towards the edge of the Al fin, where a region with reduced gap is created. As the initial quasiparticles have a few meV of energy, they shed phonons as they diffuse, which in turn can create new quasiparticles. However, some of the produced phonons will have energies $< 2\Delta_{Al}$, so they will be lost. In fact, about 50% of the initial energy is lost in this way [182], [176].

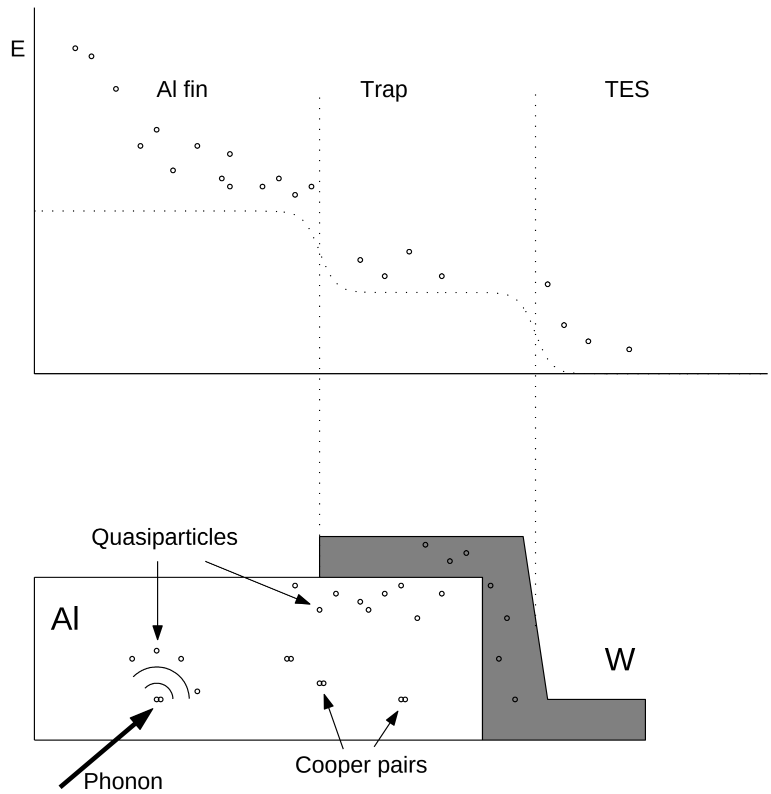


Figure 4.2: Schematic of the quasiparticle production and diffusion.

The diffusion length of quasiparticles is determined by the diffusivity (D) and the quasiparticle lifetime (τ):

$$l_{diff} \sim \sqrt{D\tau}, \tag{4.3}$$

which in turn is determined by the mean free path of the quasiparticles and by the thickness of the Al fin. The lifetime is very sensitive to the thickness of the Al fin, as shown in [183]. In particular, thicker Al fin would allow longer lifetime and diffusion length (and hence longer fins). However, it was determined empirically that if the thicknesses of Al and W films are very different from each other, the Al-W overlap region becomes less reliable. On the other hand, one cannot increase the thickness of the W film at will either, since its heat capacity must be kept low. These considerations lead to the current design of the Al and W films on the surface of the ZIP detectors. In particular, the dimensions of the Al fins are 380 μm $\times$ 50 μm $\times$ 300 nm. As discussed in [182], one can estimate that such dimensions lead to the quasiparticle collection efficiency of 24%, and quasiparticle diffusion times (from production to trapping) of order 1-2 μs.

The edge of the Al fin overlaps with W TES's. This region of proximitized tugsten has lower gap than most of the fin. To be trapped in this region, the quasiparticle must lose some of its energy, so that its energy becomes smaller than the Al gap. This can happen either by scattering off of Cooper pairs, or by emitting a phonon (in which case, energy is again lost). Once the quasiparticle is trapped, it has to pass through the Al/W interface. In principle, energy loss is possible in this step too, but it has been shown that this transmission probability is fairly good [176]. In the W, the electron-electron scattering is about 100 times more frequent than electron-phonon scattering, so most of the quasiparticle energy is deposited in the electron system of the W.

Hence, the overall phonon energy collection efficiency is dominated by the loss due to sub-gap phonons and by the quasiparticle diffusion losses. It amounts to about $5-10\%$.

ETF-TES

The phonon/quasiparticle signal described above is converted into an electric signal in the tungsten transition-edge sensor (TES) operated in the electrothermal feedback (ETF) mode. This means that the W TES is voltage biased, and kept on its superconducting transition edge.

Figure 4.3 shows a schematic of the thermal model. The thermal conductivities

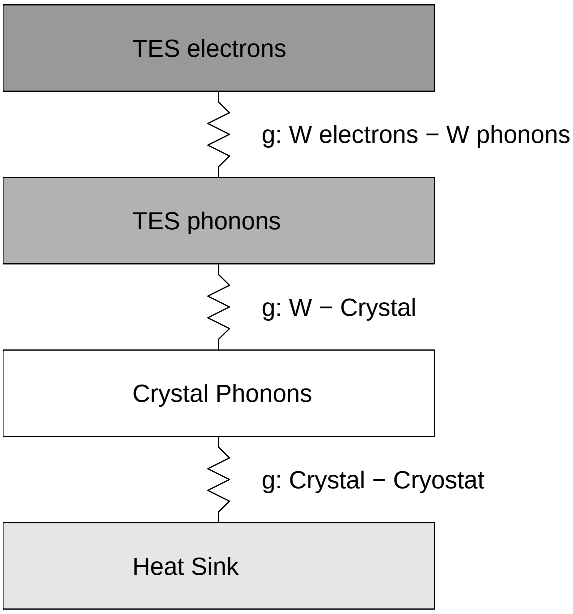


Figure 4.3: Schematic of the thermal model of the phonon sensors in ZIP detectors.

from W phonons to the crystal and from the crystal to the cryostat are sufficiently large, so that the cryostat and the W phonon system are at the same temperature. Therefore, the main thermal impedance comes from the W electron-phonon coupling. The electron-phonon decoupling theory estimates this thermal conductivity to be: $g_{ep} = n\kappa T^{n-1}$. The heat-flow in the TES is then described by the following equation:

$$C_V \frac{dT}{dt} = \frac{V_b^2}{R(T)} - \kappa(T^n - T_s^n). \tag{4.4}$$

Here, C_V is the heat capacity of W, T is its electron temperature, t is time, V_b is voltage bias, $R(T)$ is the temperature dependent resistance of W, κ is the coupling coefficient, and T and T_s are the W and the substrate (i.e. the crystal, or the cryostat) temperatures. The power n is empirically determined to be 5, by the Cabrera group at Stanford [184] .

Intuitively, in the equilibrium, $dT/dt = 0$, so the Joule heating and the cooling into the substrate are equalized. When energy (quasiparticles from Al fins) is injected into the TES, the temperature of the electron system changes slightly, corresponding to a relatively large change in the sensor resistance (which is what we measure). The increased resistance, however, implies that the Joule heating is reduced (since the sensor is voltage biased, $V_b = const$), which brings the TES back into the original equilibrium. In particular, consider a small perturbation δT around the equilibrium, in the extreme case of exact voltage biasing. The Equation 4.4 yields the following differential equation describing the time evolution of δT:

$$C_V \frac{d\delta T}{dt} = -\left(\frac{V_b^2}{R^2(T)} \frac{dR}{dT} + 5T^4 \kappa \right) \delta T. \tag{4.5}$$

This equation can easily be solved:

$$\delta T(t) \sim e^{-t/\tau_{ETF}}, \tag{4.6}$$

where τ_{ETF} can be written as (assuming $V_b^2/R(T) = \kappa T^5$, which is true if the sensor is operated reasonably far from the substrate temperature)

$$\tau_{ETF} = \frac{\frac{C_V}{5T^4\kappa}}{1 + \frac{1}{5} \frac{T}{R} \frac{dR}{dT}} = \frac{\tau_0}{1 + \frac{\alpha}{5}}. \tag{4.7}$$

In the last equation we have defined $\tau_0 = C_V/(5T^4\kappa)$ and $\alpha = (T/R)\, dR/dT$. In particular, α is a measure of the steepness of the transition curve.

Therefore, if a small temperature perturbation is introduced, the TES responds by exponential cool-down, with the time constant τ_{ETF}. Or, in other words, the temperature signal in the TES is a convolution of the initial energy (i.e. quasiparticle) flux and a decaying exponential with the time constant τ_{ETF}. Empirically, τ_{ETF} was determined to be about 50μs, although this value depends on the operating point on the transition edge. We will further discuss various aspects of the phonon pulse shape in the Section 4.5.

4.2.2 Ionization Signal

The ionization signal is described in considerable detail in [2]. Here we outline the basic principles and some of the most important details.

The principles behind the ionization signal are simpler than those behind the phonon signal in the ZIP detectors. The crystals of Ge or Si of high purity become electrically insulating at temperatures below 4 K (for Ge, 10 K for Si), because there are no thermal excitations that would populate the conduction and valence bands. A particle interacting in the semiconductor crystal causes electron-hole pairs to be broken. Empirically, it takes about 3 eV to break an electron-hole pair in Ge and 3.8 eV in Si. We then use a small electric field (usually 3 V/cm) to separate these electrons and holes, and then collect them in the electrodes covering the two flat surfaces of the crystal. There are two complications, however.

The first is the impurities. Namely, the crystals will always have a number of acceptor and donor impurities (about $10^{11}/\mathrm{cm}^3$ for Ge and $10^{14}/\mathrm{cm}^3$ for Si). These are fully ionized at room temperature. As the cryostat (and the detectors) is first cooled down, these impurities are left ionized. Hence, they form very shallow energy levels, just above the valence band (or below the conduction band), and effectively act as traps for electrons and holes drifting by after an interaction. This effect can seriously impact the size of the ionization signal. To resolve this problem, we perform "LED baking" - we shine infrared photons at the detectors for an extended period of time (several hours for Ge, several days for Si) while grounding the charge electrodes. This creates plenty of electrons and holes, which then drift toward the impurity sites and populate them (i.e. they fill the traps).

After this procedure, the detector is effectively neutralized and the signal-to-noise of the ionization signal is optimized. However, this situation is not stable when the detector is biased (i.e. when an electric field is created). Some of the traps are emptied during the particle interactions, so the charge collection efficiency degrades with time. The time span over which this degradation becomes important depends on the event rate. Hence, at the test facilities (not underground, so event rate is about 50 Hz) this is of order 10 minutes, while at the underground sites at SUF or at Soudan this is of order 8 hours (for Si, much longer for Ge). To recover from these degradations, we again use LEDs, but we only flash them for a few minutes. This restores the neutralized state and we can resume data acquisition at the optimal charge collection efficiency.

The second complication are the surface events. Events that happen sufficiently close to the surface of the crystal can have a fraction of the electrons (or holes) travel against the electric field and be collected on the "wrong" electrode. This effectively reduces the ionization signal, and makes the ionization yield of these events smaller. The effect can be sufficiently large that the surface electron-recoil events may look like the nuclear-recoil events. Hence, the surface events are one type of background in the WIMP-search runs (in fact, dominant at Soudan) and we discuss different ways of handling such events in Chapters 5, 6, and 8.

The effect can be significantly reduced by depositing a layer of amorphous Si (α-Si) on the surface of the crystal [170, 2]. The α-Si has a larger gap (~ 1.2 eV) than Ge (0.743 eV) or Si (~ 1.17 eV). If the gaps of the α-Si and of the crystal are centered, potential barriers (of the same size) for the electrons and holes are created at the interface between

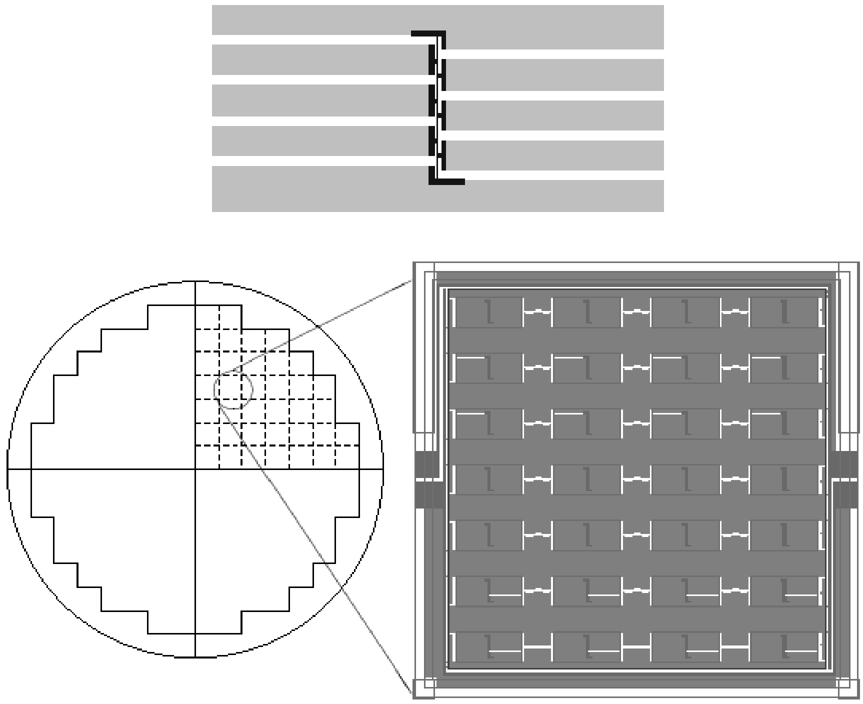


Figure 4.4: Schematic of the phonon-sensor design. The top drawing shows the W TES connected to eight superconducting Al fins that supply the quasiparticles. An array of 7 × 4 of such W-Al modules forms a 5 mm unit cell. Each of the four quadrants contains 37 such cells. The space between the W-Al modules is covered by the Al grid, completing the ground-electrode for the charge channels. See text for more detail. Figure taken from [182].

the crystal and the α-Si. Such barriers reflect back the carriers diffusing in the wrong direction. If the gaps are not well-centered, the barriers for the electrons and holes will be of different sizes, and the two sides of the detector will have assymmetric ionization response. We discuss the effectiveness of this technique in Chapters 5, 6, and 8.

4.3 Fabrication and Detector Testing

4.3.1 Initial Fabrication

The ZIP detectors are based on crystals of Si or Ge of thickness 1 cm and diameter 7.6 cm. Both sides are first covered by an amorphous Si film, which is necessary to improve the charge collection efficiency for the surface events, as discussed in the previous Section. One side of the crystal is covered by four phonon sensors, each covering one quadrant, and each consisting of 37 cells, as shown in Figure 4.4. Each cell has an array of 7×4 tungsten TES's, and each TES is "fed" by eight Al fins of dimensions 380 μm $\times$ 50 μm $\times$ 300 nm. An Al bias rail goes through all the cells of one quadrant in a snake-like pattern, and it connects all the TES's (1036 of them) in parallel. Each W TES is 250 μm $\times$ 1 μm, and its thickness is variable, as described below. Between the Al fins, an Al grid is formed from 2 μm wide lines 20 μm apart - together with Al fins and W TESs, this grid represents the ground for the charge signal. For more information on the design of the phonon sensors in the ZIP detectors see, for example, [185, 186, 187]. The other side of crystal has two charge channels. They are in the form of a similar grid of W and Al, but split into two electrodes: an inner one in the shape of a disk and an outer one in the shape of a ring.

The crystals themselves are cut from the original boules, supplied by the manufacturers (Perkin Elmer for Ge and Topsil for Si), which satisfy our requirements of chemical and structural purity. The crystals are stored in the shallow underground clean room (at SUF) whenever they are not being processed. Most of the fabrication is performed at the Stanford Nanofabrication Facility (SNF), which is housed at the Center for Integrated Systems (CIS) on Stanford University campus.

The substrates are first cleaned by stripping and regrowing the surface oxide layer. The actual fabrication is fairly involving (see [188] for more detail), so we outline here only the main steps, depicted in Figure 4.5. First, the amorphous Si, Al and W films are deposited on both sides of the crystal (in two separate depositions), using a DC magnetron sputtering machine, Balzers 450. On the phonon side the film thicknesses are 40 nm, 300 nm, and 35 nm, respectively, and on the charge side, they are 40 nm, 20 nm, and 20 nm respectively. The second step is to cover both surfaces with a photoresist, which is then exposed in the Ultratech stepper-aligner, modified to handle 1 cm thick wafer. This machine can only expose one 5×5 mm^2 square, which is why each phonon sensor consists of 37 cells. Hence, to define the Al fins, Al bias rails etc, a sequence of masks has to be used. After developing the photoresist, the exposed regions (of both W and Al) are etched away. At this point the Al fins, grid and bias rails are created. The third step is another W deposition (in Balzers). In the fourth step, the photoresist is again applied on the phonon side, exposed and developed, and the exposed part of the W film is etched. At this point the quasiparticle traps have been formed, and most of the W film covering the Al fins has been removed. The phonon side is now done, and it is covered with the photoresist again until the charge side is completed. In the fifth step, the photoresist on the charge side is applied, exposed and developed, after which the exposed parts of the W and Al films are etched away. Most of the charge side pattern is now complete. The last step is to introduce the break between the inner and outer charge electrodes. So, another dose of photoresist, and the gap in the amorphous Si between the two electrodes is created. The fabrication procedure is now complete.

On average it takes about 2 months to complete these fabrication steps. Usually, a group of substrates is processed at the same time, typically with a few thin test-wafers. The test-wafers are useful because they can be used to get an idea of the W critical temperature (T_c) distribution across the surface. Namely, we have observed non-uniformities in the T_c distribution, but since the W TES's are wired in parallel it is very difficult to study such non-uniformities in the detectors themselves. The test wafers, however, can be cut in pieces and T_c can be measured for each piece.

After the fabrication is complete, the detectors are sent to Test Facilities (at UC Berkeley or at Case Western Reserve University), where detailed tests are performed (see below for more details). These tests, combined with the test-wafer measurements, give some information on the T_c distribution across the surface. Typically, the T_c's fall in the $130-150$ mK range, about 50 mK higher than desired. This is intentional - we routinely use ion-implantation procedure to bring the T_c distribution down to the desired range ([189]). In particular, the detectors are sent to have ^{56}Fe ions implanted. The implantation is done in several steps, and different parts of the detector are exposed to different doses, depending on their starting T_c values.

4.3.2 Detector Characterization at Test Facilities

The Test Facilities (one at UC Berkeley and one at Case Western Reserve University) are designed to perform a sequence of tests on the newly fabricated ZIP detectors. They rely on dilution refrigerators, specifically modified to accept one tower with up to three ZIP detectors, and all the necessary cold hardware. They also have the necessary room-temperature electronics needed to operate the detector, including the digitizers to

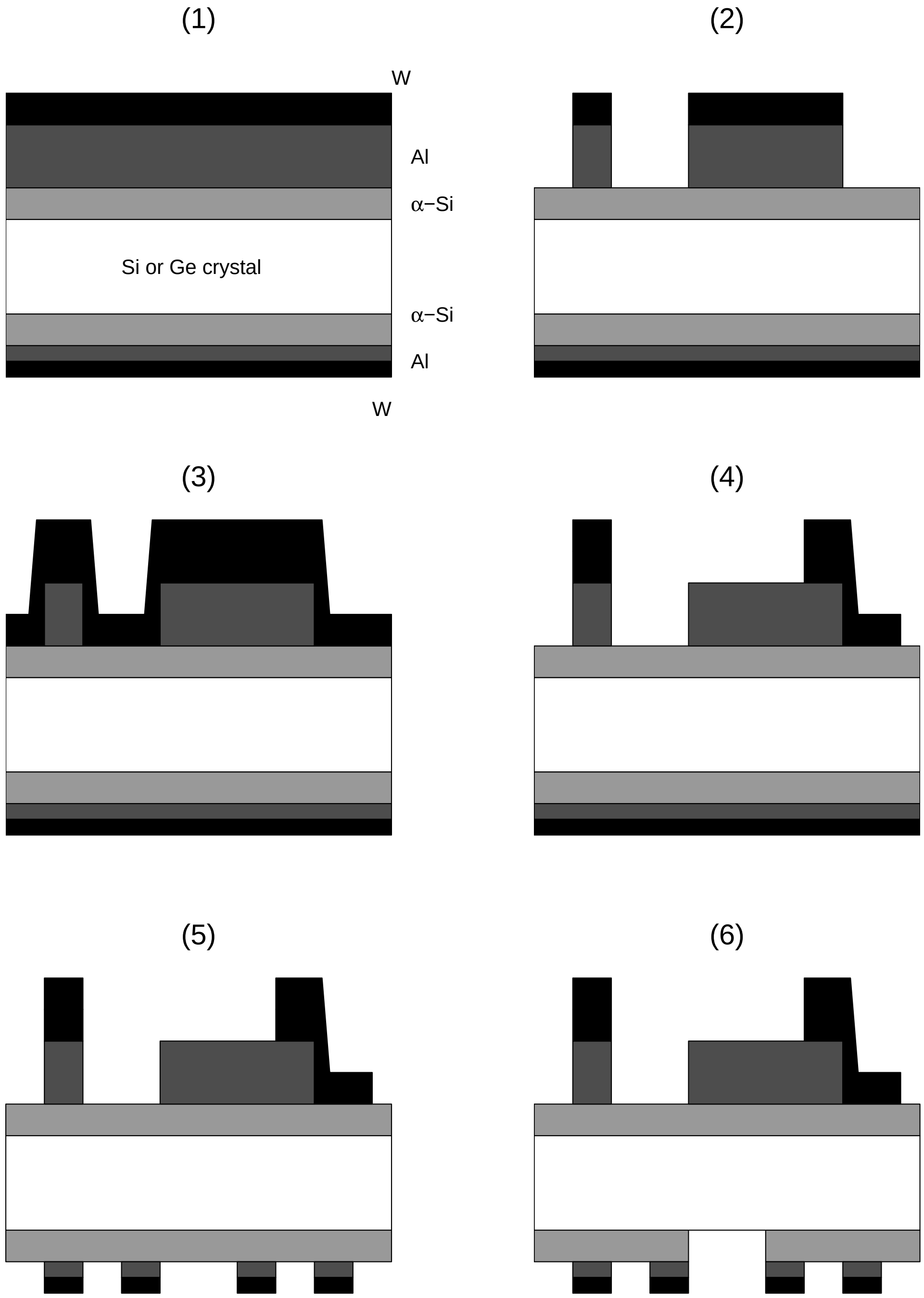


Figure 4.5: Schematic of ZIP fabrication. Si/Ge crystal is shown in white, amorphous Si in lighter gray, Al in darker gray and W in black. See text for explanations of the six steps depicted here.

record calibration datasets.

The tests performed in these facilities fall in three groups. First, there is a number of immediate tests, geared to verify that all six channels of a detector are operational: we check for electronic shorts or opens circuits, verify that all channels have reasonable pulse shapes, and check for the break-down voltage for the charge channels (it was observed that detectors from some of the boules could not be biased more than a few Volts/cm; above this break-down voltage, the crystal would start to conduct). The second group includes a set of measurements on the phonon channels: critical temperature T_c, critical current I_c, and the $I_b - I_s$ measurement. We will discuss these in more detail in this Section. Finally, the last set of tests is acquiring calibration datasets. The datasets taken predominantly depend on the available time - this varies from case to case, and in Chapter 6 we describe one of the detailed calibrations we performed at the UCB Test Facility.

Critical Temperature

In principle, this is a straightforward measurement - one measures the resistance of a phonon sensor, while changing the substrate temperature (i.e. temperature of the fridge). The resistance is usually measured by sending a small triangle wave (a few μA) through the input side of the phonon readout circuit (see Figure 3.7), while measuring the amplitude of the output triangle wave in the SQUID-locked (i.e. closed-loop) mode.

The complication arises from the fact that one phonon sensor consists of 1036 TESs, which can have very different critical temperatures. The measurement described above is, therefore, sensitive only to the upper range of the T_c distribution. Essentially, this is because if one of the TES's gets close to being superconducting, it shorts the remaining

TES's.

To understand this more quantitatively, one can perform simple simulations. The steady state solution of the Equation 4.4, for a single TES, obeys:

$$\frac{V^2}{R(T)} = \kappa(T^5 - T_s^5). \tag{4.8}$$

One can then solve these equations simultaneously for several TES wired in parallel. As an example, we consider the case where $T_s = 30$ mK, $\kappa = 2.5 \times 10^{-5}$ J/K^5, $R_b = 20$ mΩ, and the transition of each TES is given by:

$$R_{TES} = \frac{R_n}{2}\left[\tanh\left(\frac{T - T_c}{w}\right) + 1\right], \tag{4.9}$$

where $w = 0.002$ K determines the width of the transition, T_c is the critical temperature of the TES, and R_n is the normal resistance of the TES (we assume the resistance of the whole sensor is 1Ω, so $R_n = N\ \Omega$, where N is the number of TES's in the sensor). We assume the phonon sensor consists of 11 TES with $T_c = 65..75$ mK, 1 mK apart. The top plot in Figure 4.6 shows the transitions of the individual TES's in this simulation, along with the steady state equilibrium for the typical biasing current of $I_b = 120\ \mu$A. Note that despite the fact that the T_c distribution is wide, the individual TES's are biased at about the same point on the transition edge. This is because the currents through the individual TES's are not equal - rather, they are distributed so that the highest T_c TES gets the largest current, as shown in Figure 4.7 (left). To understand this intuitively, consider the case where all TES's have the same T_c (and the same current). Then, if T_c of one TES is increased, it's resistance at the given biasing current would be smaller than other TES's - it would then start shorting the other TES's, which would increase its current, and therefore its resistance.

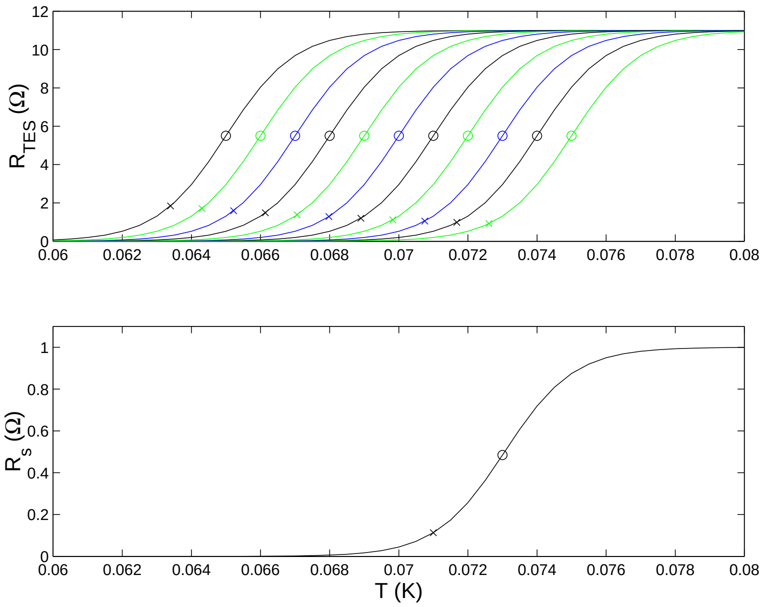


Figure 4.6: The top figure shows the transitions of the individual TES's. The bottom plots shows the transition of the whole sensor. In both plots, o's denote the T_c, and x's denote the biasing points for $I_b = 120\ \mu$A.

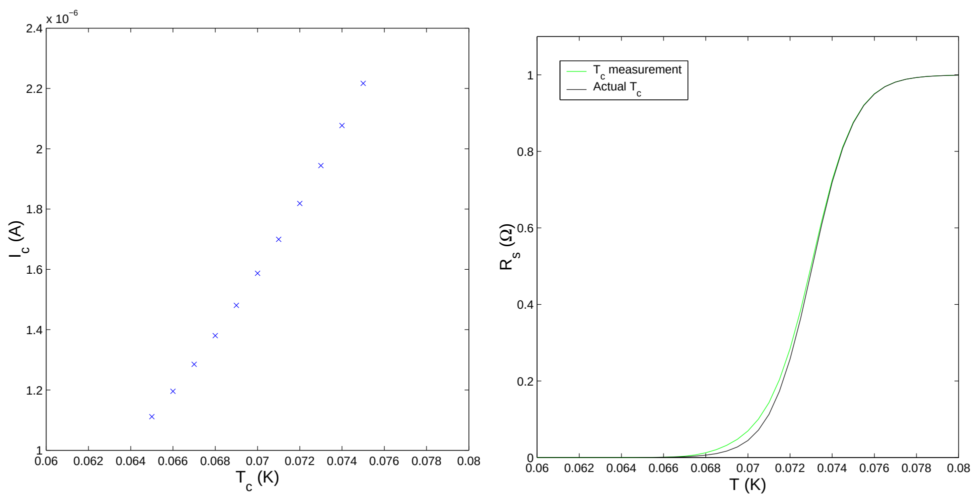


Figure 4.7: Left: Current through individual TES's as a function of their T_c. Right: Comparison of the T_c measurement (green/gray) to the true transition curve (black).

This point is very important, as it significantly dampens the effect of T_c non-uniformity on the response of the sensor. Of course, it does not remove the effect - the size of the effect depends on the magnitude of the non-uniformity, transition width of individual TES's etc.

The bottom plot in Figure 4.6 shows the resistance of the whole sensor as a function of temperature. Comparing the top and bottom plots of Figure 4.6 shows that the transition curve of the whole sensor captures only transition of the high T_c TES's.

Finally, one can simulate the T_c measurement described above: keep the bias current small, $I_b = 10\ \mu$A, and vary the substrate/fridge temperature through the T_c range. The Figure 4.7 (right) shows that the superconducting transition curve measured in this way is very close to the true transition curve - in fact, we routinely use input currents smaller than 10 μA, which makes the difference even smaller.

Critical Current

The critical current measurement is also relatively simple: one measures the current needed to drive the whole (superconducting) phonon sensor into the transition, while varying the temperature of the substrate (i.e. of the fridge). This measurement gives another handle on determining the T_c distribution in the phonon sensor. The analysis method has been developed by Paul Brink, and it is described in [190]. Here we briefly outline it.

The total supercurrent through the phonon sensor, at the substrate/fridge temperature T_s, is maximally equal to the sum of the supercurrents through the individual

TES's, indexed by k:

$$I_c(T_s) \leq \sum_k I_{c_k}(T_s, T_{c_k}). \tag{4.10}$$

The individual TES supercurrents are given approximately by the Ginzburg-Landau theory:

$$I_{c_k}(T_s, T_{c_k}) \approx I_{c_k}(0)\left(T_c^{3/2} - T_s^{3/2}\right). \tag{4.11}$$

One then starts with a group of TES's with high T_c and attempts to fit the I_c vs T_s data in the region of highest substrate temperature. At lower temperatures such fit is not good, so one invokes a second group of TES's with somewhat lower T_c to improve the fit, and so on. The fit is, of course, not unique, but it gives a simple, discrete, distribution that fits the data. In particular, the combination of the T_c and I_c measurements gives a fairly reliable description of the upper portion of the T_c distribution.

$I_b - I_s$ Measurement

The $I_b - I_s$ measurement is another handle on determining the T_c distribution of the phonon sensor. The input side of the phonon readout circuit (see Figure 3.7 for a schematic) can be solved to give:

$$I_s = I_b \, \frac{R_{sh}}{R_s + R_p + R_{sh}}. \tag{4.12}$$

As we scan I_b from large values ~ 2 mA to zero, we encounter three regions:

- Normal region: In this case, I_b is large enough to drive the sensor completely normal, $R_s \approx 1\ \Omega >> R_p, R_{sh}$. From Equation 4.12,

$$I_s \approx (20\ \mathrm{m}\Omega/1\ \Omega)I_b = I_b/50. \tag{4.13}$$

Since $R_s = const$ in this region, the power dissipated in the sensor is $P = I_s^2 R_s \approx I_b^2 R_s / 50^2$ - i.e. it follows quadratic behavior.

- Biased region: In this case, the sensor is in its superconducting transition. The power dissipated in the sensor is given by Equation 4.8. Since the sensor temperature T does not change much in the transition, the power dissipation in the sensor is roughly constant in this region.

- Superconducting region: In this case, I_b is small enough that the sensor is fully superconducting. Hence, $R_s = 0$, so the Equation 4.12 reduces to

$$I_s = I_b \frac{R_{sh}}{R_p + R_{sh}} \approx \frac{I_b}{2} \tag{4.14}$$

  because $R_p \approx R_{sh} = 20$ mΩ. No power is dissipated in the sensor.

The Figure 4.8 shows the result of the $I_b - I_s$ measurement for one of the ZIP sensors. The top plot shows I_s vs I_b - note the linear dependence in the superconducting and normal regions. The middle plot shows the sensor resistance R_s as a function of I_b - note that the measurement returns the normal and parasitic resistances from the normal and superconducting regions respectively. Finally, the thick solid line on the bottom plot shows the power dissipated in the sensor as a function of I_b - note that the power is constant in the biased region. It turns into the square law behavior in the normal region, and it drops to zero in the superconducting region. The thin solid line depicts the quadratic dependence of a simple resistor.

Clearly, the measurement immediately gives much information about the phonon sensor. However, one can push this analysis further - with some assumptions regarding the

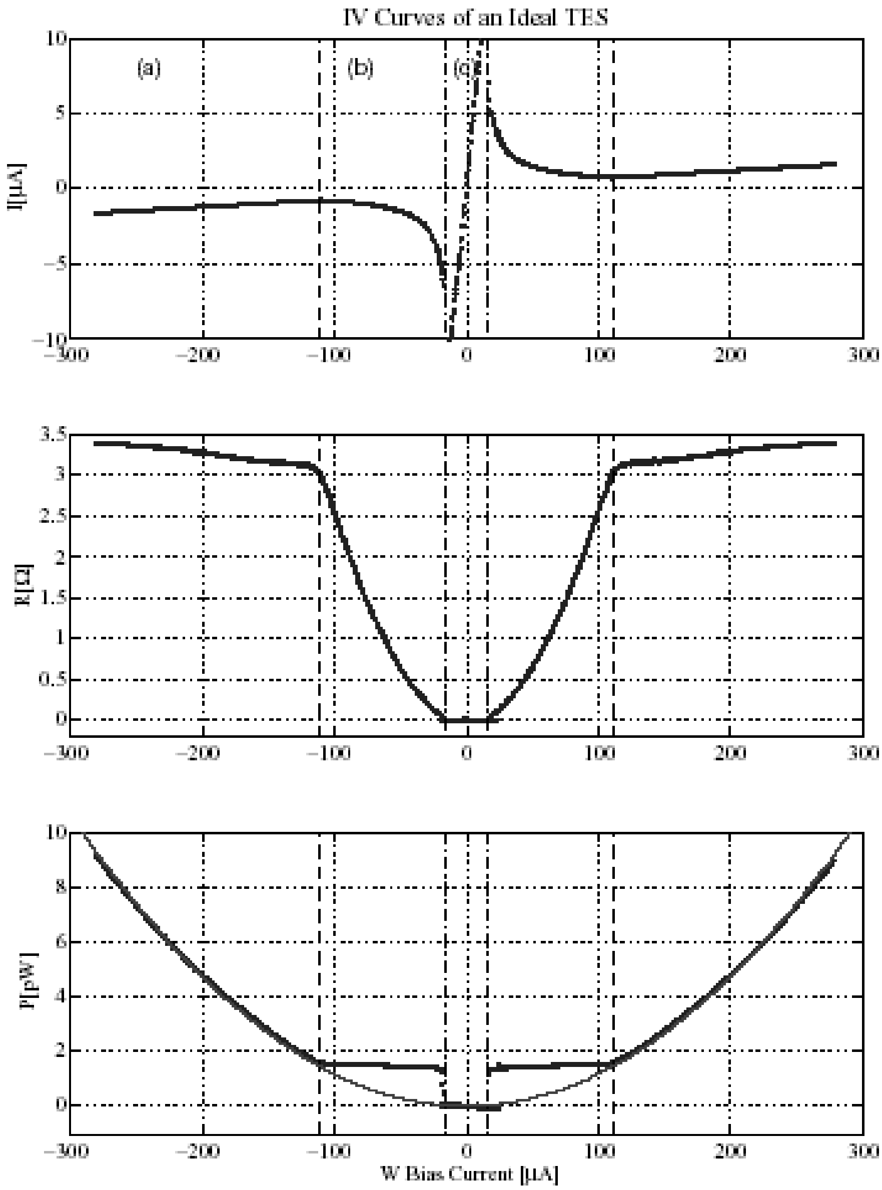


Figure 4.8: $I_b - I_s$ measurement, taken from [182]. See text for details.

T_c distribution one can use the $I_b - I_s$ measurement to calculate the R_s vs T curve. In Figure 4.9, we use a simple simulation to show that the power vs I_b curve changes for different transition widths. Hence, the $I_b - I_s$ measurement is sensitive to the T_c non-uniformity, and it is complementary to the T_c and I_c measurements, as it probes somewhat lower range of the T_c distribution than the I_c measurement, and it returns a more detailed $R_s - T$ curve than the T_c measurement.

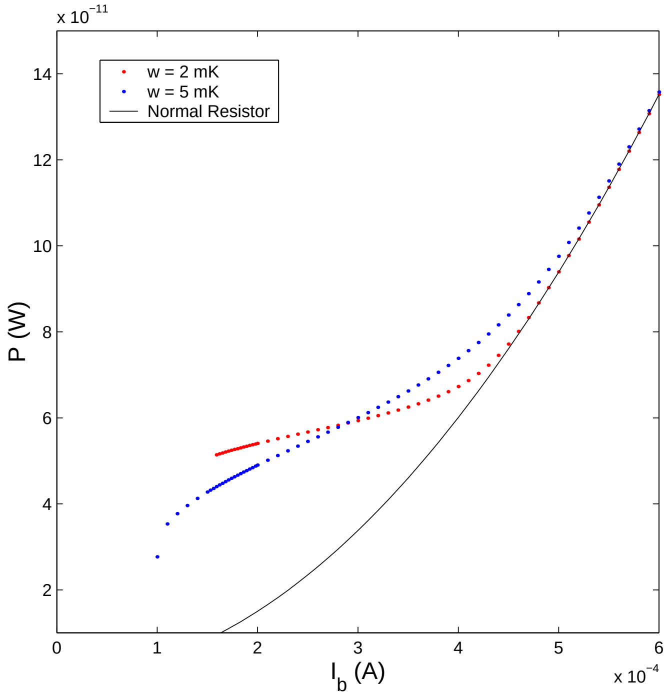


Figure 4.9: Results of a numerical solution to the heat-flow equation. Note that the shape of the $P - I_b$ curve changes depending on the transition width.

One can, therefore, start with the results of the T_c and I_c measurement, which give some information regarding the T_c distribution, and check how well they fit the $I_b - I_s$ data. In some cases, a modification of the assumed T_c map is required to fit the data well. Figure 4.10 shows an example of how the data can be fitted by this model.

The case discussed so far assumes that the whole TES is at the same temperature. It is, however, possible to find solutions to the heat-flow Equation 4.4 in which a part of the TES is superconducting and a part is normal. The criterion for phase separation is

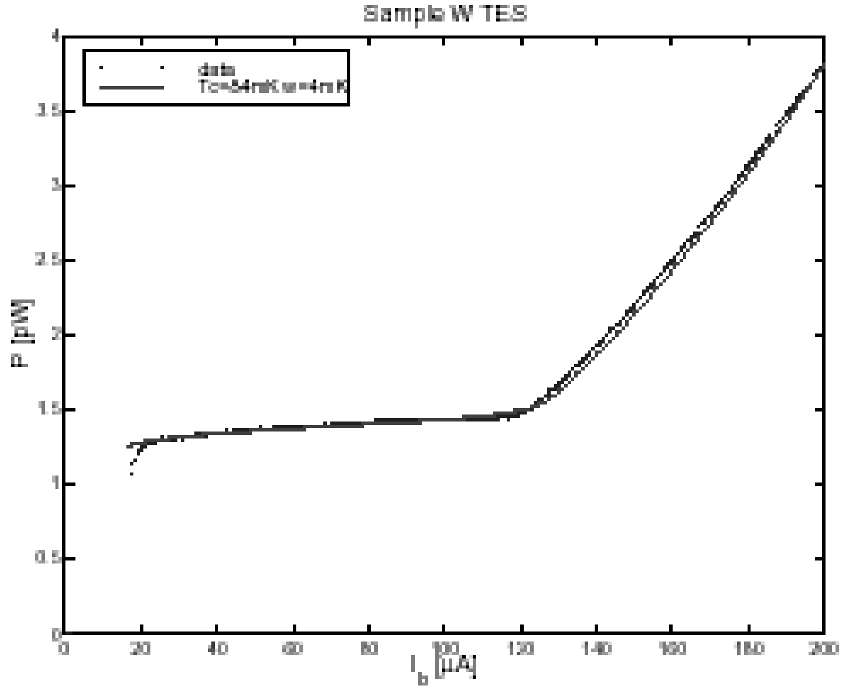


Figure 4.10: $I_b - I_s$ data are very well fitted by the model. Hence, one can extract transition parameters (T_c, transition width etc). Taken from [182].

obtained by balancing the heat flow along the TES with the heat flow into the substrate, as discussed in detail in [173]. In particular, the criterion for phase separation is:

$$\frac{g_{wf}}{g_{ep}} \geq \frac{\alpha/n}{\pi^2}, \tag{4.15}$$

where g_{wf} is the Wiedeman-Franz thermal conductivity along the TES and g_{ep} is the thermal conductivity into the substrate via the electron-phonon interactions. $\alpha = (T/R)(dR/dT)$ is the measure of the sharpness of the transition.

Such phase-separated solutions make the $I_b - I_s$ analysis somewhat more complicated. This has been described in detail in [182], so we will not repeat it here. Figure 4.11 demonstrates the effect of phase separation on the $I_b - I_s$ measurement.

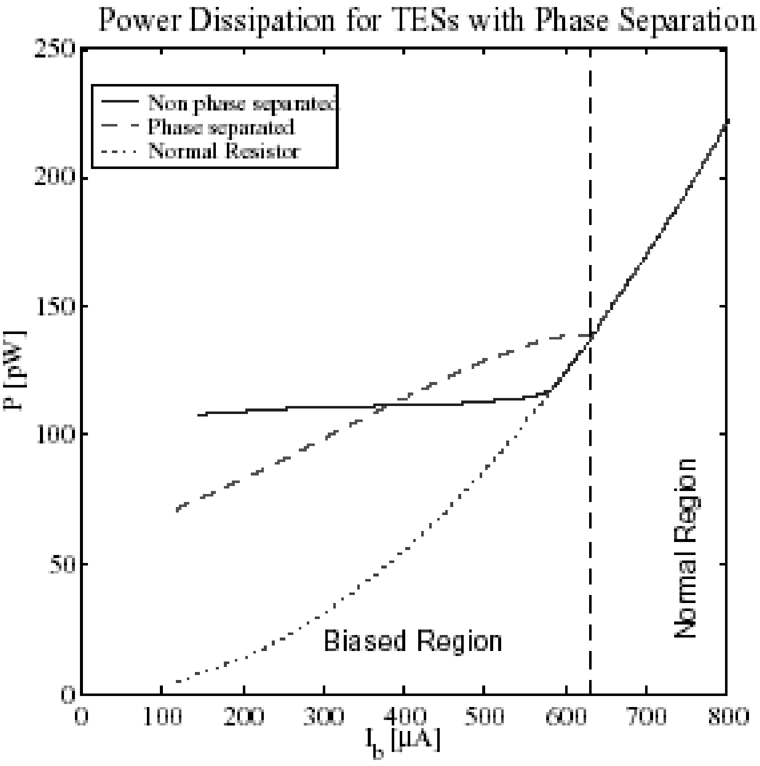


Figure 4.11: Effect of phase separation on the shape of the $P - I_b$ curve. Taken from [182].

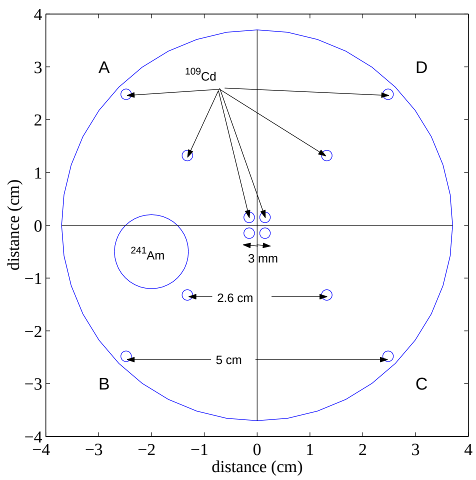
4.4 Event Position Reconstruction

Since the ZIP detectors have 4 phonon sensors, each covering a quadrant of the Si/Ge crystal, we can attempt to reconstruct the position of the interaction using the 4 phonon signals. One of the test runs performed at the UC Berkeley test facility was specifically designed to study different algorithms for event position reconstruction. In particular, data was collected with a Si ZIP detector exposed to a 12-hole collimator ^{109}Cd source, covering one surface of the detector, and a single-hole collimator ^{241}Am source. Figure 4.12 (left) shows a schematic of the setup. The collimator hole for the Am source was covered with Al foil to block the emitted αs. ^{109}Cd is a source of both γs and βs, and was used extensively to study the surface events in a separate run (see Chapter 6).

One approach to extracting the event position is to use amplitudes of the four signals, since the largest signal occurs in the primary quadrant (in which the event occurred). We define phonon partitions

$$p_x = \frac{(C+D)-(A+B)}{A+B+C+D}, p_y = \frac{(A+D)-(B+C)}{A+B+C+D}, \tag{4.16}$$

where A, B, C, D denote the pulse amplitudes in the four sensors (marked in counterclockwise direction). Figure 4.12 (right) shows the event distribution on the detector surface generated by applying this procedure to a Si dataset - such plot is usually referred to as the "box plot". This plot is not a realistic position map - it has a rectangular shape instead of the expected circular one. This is because a part of the phonon signal is lost for events that happen toward the edge of the crystal. Hence, the p_y-p_x plot folds on itself and the edges of the plot are actually formed by the events from the detector interior.



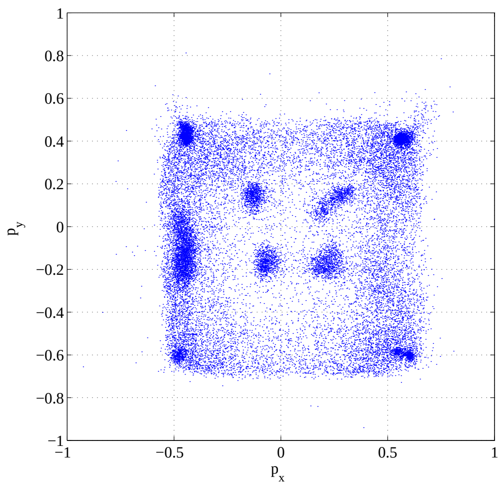


Figure 4.12: Left: Schematic of the experimental setup. The outer-most collimator holes are too close to the edge of the crystal and will be discarded in the analysis. Right: Map of the detector surface using phonon partitions. Figure taken from [191].

Another approach, as discussed in [191], is to use the delay of the phonon pulses.

Since the ionization signal has almost no delay, the start of the ionization signal is a measurement of the event time. The phonon pulses are delayed with respect to the ionization signal depending on their position.

We measure the delay of the 20% of the leading edge of the phonon pulse with respect to the start of the ionization signal. The shortest delay indicates the primary quadrant. Subtracting this shortest delay from the delays of the neighboring sensors gives the x and y estimates. As an example, if the shortest delay is in channel A, then the x-coordinate is given by $\mathrm{delay}_\mathrm{A} - \mathrm{delay}_\mathrm{D}$ and the y-coordinate is given by $\mathrm{delay}_\mathrm{B} - \mathrm{delay}_\mathrm{A}$ (see Figure 4.12 for the definitions of A, B, and D). We also define the delay radius r_d as the Euclidean distance of an event from the origin in the delay x-y plot. We found that this simple algorithm gives the best x-y position map, both in terms of the non-linearity of the map and in terms of the discontinuities of the map at the edges between sensors. The result of applying this algorithm to the Si data is shown in Figure 4.13. Although this plot does not "fold on itself" and it is a more realistic map of the detector surface, it is not free of non-linearities. In particular, the inner four Cd blobs in this plot correspond to collimator holes that are 0.3 cm apart, while the outer four are 2.6 cm apart (recall that the diameter of the crystal is 7.6 cm). Hence, as one would expect, the position resolution of this map is the best in the center of the detector and it worsens toward the edge. The magnitude of this non-linearity can be seen in the Figure 4.13 (right), where we used the physical position of the Cd collimator holes to radially transform the delay x-y plot into a "physical" x-y plot. In the corrected plot, the Cd blobs are stretched in the radial direction, but they are centered appropriately. This suggests that two blobs per sensor are not sufficient to fully understand the non-linearity of the delay x-y plot.

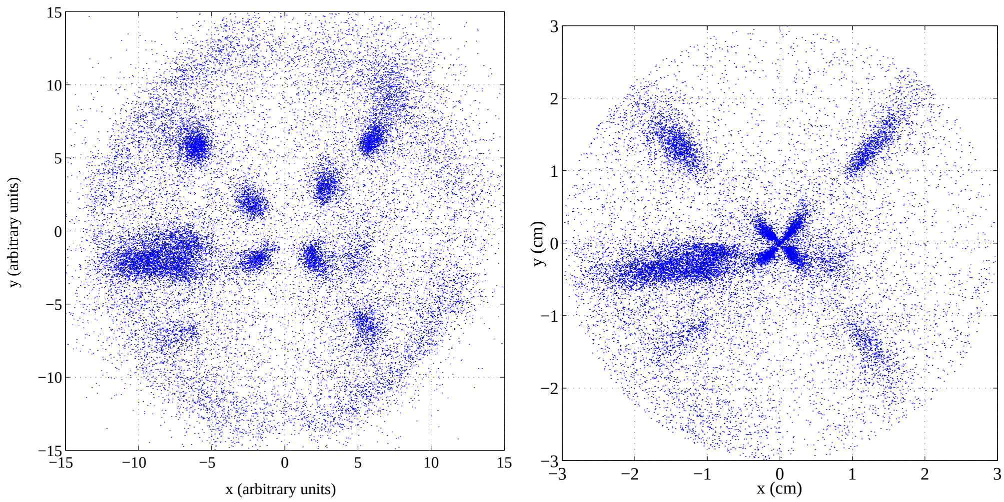


Figure 4.13: Left: Delay x-y plot. Right: Physical x-y plot, obtained using the delay x-y plot and the physical positions of the Cd collimator holes. Due to the low resolution, only events within 3 cm from the center will be considered in this study. Figure taken from [191].

4.5 Understanding the Phonon Signal

Although the production and the read-out of the phonon signal in the ZIP detectors are fairly well understood, some aspects of the phonon signal remain relatively poorly understood. In this Section, we discuss some of the peculiarities of the phonon signal and outline some of the issues to be addressed in the future.

One very important aspect of the phonon signal is its position dependence. The position dependence is caused by two effects. First, as discussed in detail in Section 4.3, the non-uniformity in the critical temperature of the TESs can cause variations in the phonon-sensor response. For a given recoil-energy event, the phonon-pulse amplitude and shape may depend on the position of the event, as the TESs in one region of the detector may be more sensitive than those in another region. Second, the surface-coverage by the TESs becomes worse towards the edge of the crystal. Similarly, reflections off of the cylindrical surface of the crystal become more important close to the edge of the crystal, simply due to the larger solid angle. Hence, regardless of the possible T_c non-uniformity, the phonon signal in the ZIP detector intrinsically suffers from the radial position dependence.

In practice, we can correct for the position dependence using large gamma calibrations, as discussed in Chapters 5, 6 and 8. However, the quantitative understanding of the position dependence has not been established yet. To understand the position dependence from "first principles", one can perform simulations of the phonon production and propagation through the crystal. Such simulations have been done in the past (e.g. [173]). More recently, G. Wang's work along these lines has yielded very interesting preliminary results [192].

Another interesting aspect of the phonon signal is the very long decay-time, of order 300-400 μs. As discussed in Section 4.2, such time-scales are not expected in the phonon-signal production and propagation: the balistic phonons are expected to be produced in the first few μs after an interaction, the quasi-particle diffusion time is believed to be of order 1-2 μs, and the TES response time, τ_{ETF}, is expected to be of order 50 μs. Several possible explanations of the long phonon tails have been proposed, such as secondary phonon-production mechanisms (e.g. by quasiparticle recombinations in Al fins) or modified models of quasiparticle diffusion in Al fins. However, without detailed simulations of the relevant processes, it is difficult to accept or reject any of these explanations.

To study the position-dependence of the phonon signal, we fit the phonon pulses with the following functional form:

$$f(t) = A(e^{-t/\tau_2} - e^{-t/\tau_1}) + B(e^{-t/\tau_3} - e^{(t-t_0)/\tau_4}). \tag{4.17}$$

Intuitively, the functional form is a sum of two pulses - the fast and the slow pulse. The fast pulse, defined by the rise-time τ_1, fall-time τ_2, and the amplitude A, is meant to capture the rising edge and the peak of the phonon pulse. However, this term is not sufficient to fit the tail of the phonon pulse well. Hence, we introduce the second, slow, pulse, with rise-time τ_3, fall-time τ_4, amplitude B, and offset t_0, which is designed to fit the tail of the phonon pulse.

We fit this functional form to the dataset taken with a Si detector, described in the Section 4.4. Examples of the fitted pulses are shown in Figure 4.14. As mentioned above, it is clear that the fast component of the functional form, having a single fall-time, is not sufficient to explain the observed tail of the phonon-pulse. Figure 4.15 shows the dependence of the most interesting parameters on the distance from the relevant sensor. One can observe that the phonon-pulse rise-time, τ_1, in a given phonon channel, increases as the event position moves further away from the corresponding phonon-quadrant. The typical values of the rise-time are 10-40 μs. The first phonon fall-time, τ_2, is much less dependent on the position of the event, and its typical values are 40-70 μs, consistent with the expectation for τ_{ETF}. The second fall-time, τ_4, which describes the behavior in the tail of the phonon pulse, also seems to be position-independent, but its typical values are much larger: 300-400 μs. The pulse-amplitude is also position dependent, as expected, being the largest for events that take place in the quadrant corresponding to the examined phonon channel. Of course, the overall phonon energy of an event is obtained by summing over all

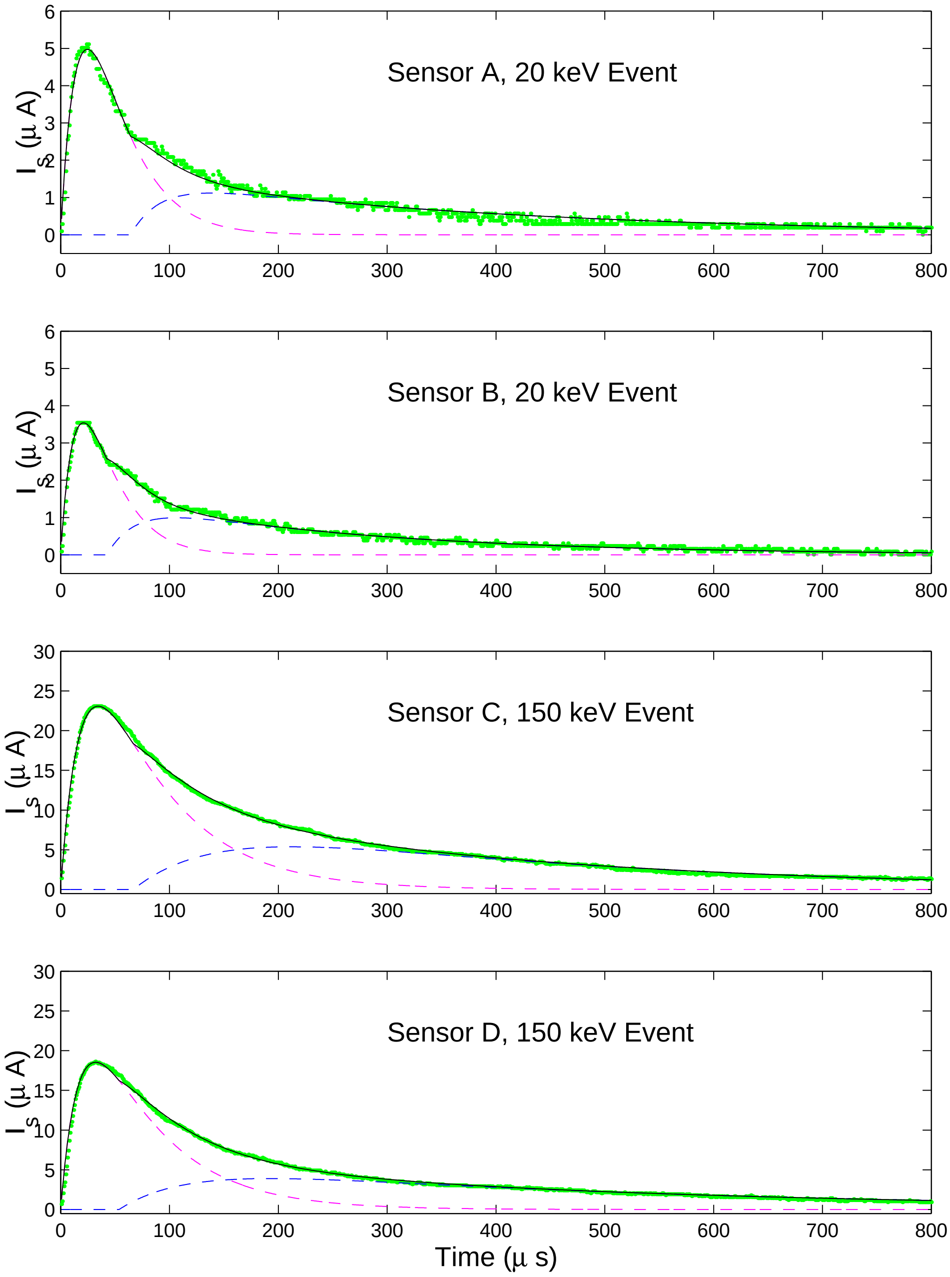


Figure 4.14: Examples of fitted pulses in the four phonon sensors. The gray/green dots denote the digitized trace, the gray/magenta and the dark/blue dashed lines denote the fast and the slow fit-pulse components respectively, and the black solid line is the overall fit.

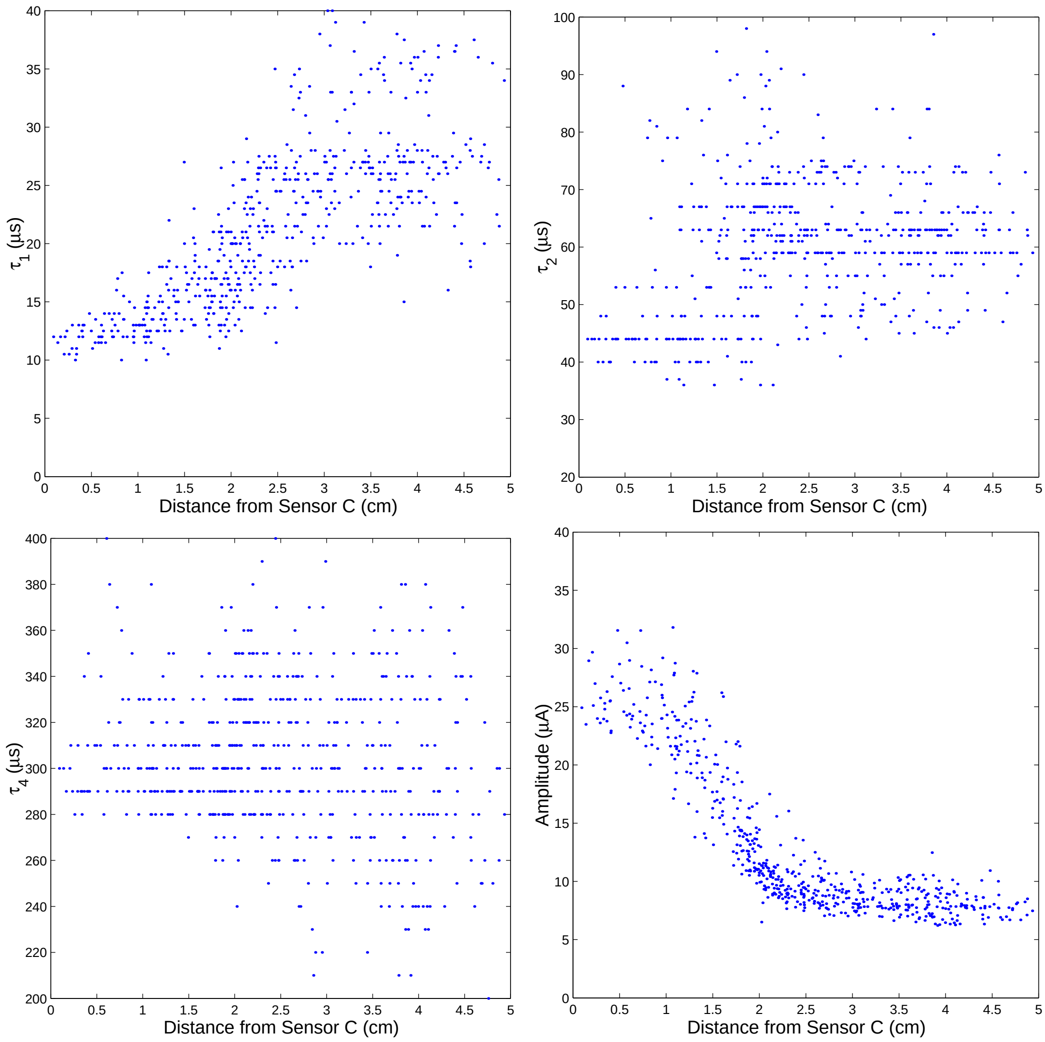


Figure 4.15: Upper Left: τ_1 estimate for channel C pulses is plotted versus the distance from sensor C. Distance from sensor C is defined as the distance from the point (1.4 cm,-1.4 cm) in the plot shown in Figure 4.13 (right). Hence, distance of ~ 2 cm corresponds to the origin. Upper Right: similar to the above, but for τ_2. Lower Left: similar to the above, but for τ_4. Lower Right: similar to the above, but for phonon-pulse amplitude in channel C.

four phonon channels.

Although this pulse-fitting method performs well in terms of reproducing the phonon pulse-shapes, it is not used in the standard CDMS analysis due to processing speed constraints. However, the method could be applied to dedicated calibration runs. In particular, one could use this method to study the variations of phonon pulse-shape for different charge-bias values (which would yield information on the Luke-Neganov-Trofimov component of the phonon flux), or for different TES voltage-bias values (which would yield more information on the effect of changing τ_{ETF}). Furthermore, one could use the fit to study the surface-events, which are one of the dominant types of backgrounds in the WIMP-search runs, as described in detail in Chapters 5, 6, and 8.

An alternative approach is to use the TES heat-flow Equation 4.4, which can be modified as follows:

$$C_V \frac{dT}{dt} = \frac{V_b^2}{R(T)} - \kappa(T^n - T_s^n) + \delta(x, y, t). \tag{4.18}$$

We have introduced the perturbation $\delta(x, y, t)$, representing the flux of the quasiparticles that enters the W TES. This perturbation is dependent on the position (x, y) of the interaction and it varies with time t. One can then model the dependence of δ on x, y, and t, and use the observed pulses, along with the event position estimates (discussed in Section 4.4), to constrain the free parameters of the δ-model. This approach requires some assumptions about the various parameters in the heat-flow equation, such as C_V or κ, as well as the relation $R(T)$. Also, the T_c non-uniformity may have significant effects. The advantage, however, is that the flux of the quasiparticles can be directly probed, before it gets convolved with the TES response time.

Yet another possible approach is to directly deconvolve the response of the TES from the observed phonon pulse. However, this requires the knowledge of τ_{ETF}, which again requires some simulation of the heat-flow equation.

Although somewhat involving, these studies could lead to a better understanding of the phonon pulse-shape and of its position-dependence. This could, in turn, lead to a better position-correction algorithm, to a better rejection of the surface-event background, and it could even suggest modifications in the detector design.

Chapter 5

Surface Events

5.1 Introduction

As discussed in the Section 4.2.2, the events that happen at the surface of the crystal, within $\sim 5\ \mu$m from the surface, can have incomplete charge collection due to the imperfections of the electric field. In particular, the degradation of the charge channel could be sufficiently large that a surface electron-recoil (ER) events can look like a nuclear-recoil (NR) events. Hence, this is one type of background for the CDMS II experiment. Furthermore, this type of background is expected to become very important (and likely dominant) at the deep site, where the neutron background will be significantly suppressed.

Such surface ER events are predominantly caused by betas and, to a smaller extent, low-energy gammas. High-energy gammas can more easily penetrate the crystal to the depths where the charge collection is complete. For example, a 10 keV photon has the attenuation coefficient of 37 cm^2/g, and can easily penetrate to depths of $\sim 50\ \mu$m in germanium. Hence, even at low energy, only a small fraction of gammas would cause surface events.

ZIP detectors provide a good handle for identifying the surface events. Namely, the phonon spectrum generated by the surface events is different from that generated by the bulk events. In particular, the rise-time and the start-time of the phonon pulse are shorter for the surface events than for the bulk events.

There are several directions that can be pursued in order to understand and suppress this type of background:

- Develop statistical methods that would optimally incorporate the information provided by the phonon rise-time/start-time parameters, the ionization yield parameter, and any additional parameters carrying information on the type of recoil.
- Perform detailed beta calibration of ZIP detectors, in order to identify the optimal operating conditions (e.g. charge bias), the optimal rise-time or start-time parameters etc.
- Improve handling, storage, and fabrication of detectors, so that their exposure to air (and Radon) and other contamination is minimized.
- Possibly modify the fabrication of the detectors to improve the information in the phonon signal on the depth of the interaction.
- Pursue detailed simulations of the phonon physics in the detector, in order to determine optimal ways of extracting information from the phonon timing parameters.

The first item will be discussed in this Chapter. The second item is the subject of the Chapter 6. The remaining items are long-run improvements that should be pursued in the future.

The remaining Sections of this Chapter are organized as follows. In Section 2 we introduce the dataset used in this analysis, we describe the standard discriminating parameters used to handle the surface ER events, the ionization yield and the phonon rise-time, and we discuss how these parameters can be used for estimating the NR signal. In Sections 3 and 4 we describe how the NR and surface ER event distributions are fitted in the yield vs rise-time parameter space. In Section 5 we use the fitted distributions to calculate the rejection efficiency of the surface electron-recoil events. In Section 6 we describe a maximum likelihood method for directly extracting the number of "true" NR events using the fitted distributions, and we discuss the results of applying this method to the low background data.

5.2 Data-set and Discrimination Parameters

The method discussed in this Chapter was developed using the data from the low-background ZIP run at the shallow Stanford Underground Facility. This run, the SUF Run21, was described in detail in the theses of Tarek Saab [182] and Don Driscoll [193]. Here, we only mention several details.

Run21 took place largely uninterrupted from November 2001 to June 2002. We collected 66 live-days with charge bias of -3 V and 52 live-days with charge bias of -6 V. After cuts, this becomes 28 kg-days and 21 kg-days respectively. The run was performed with Tower 1, which contains 4 Ge detectors and 2 Si detectors. The run was interrupted with several calibrations with ^{60}Co gamma source and ^{252}Cf neutron source. The sources were placed outside of the cryostat, so the gammas and neutrons had to penetrate through the various cans of the Icebox and through the internal shielding (lead and polyethylene - see Sections 3.3 and 3.4). In particular, it turns out that about 1% of the events taken during the ^{60}Co calibration are surface events, predominantly caused by betas created in interactions of gammas with the material surrounding the detectors. Hence, the ^{60}Co calibration could be used as a surface-event calibration as well.

In this Chapter, we focus on the data taken at -3 V charge bias. We will not discuss the cuts, cut efficiencies, background levels, or dark matter limits that this run produced - see [182], [193], [104] for such details. Instead, we focus on the surface electron-recoil events, both in the calibration data and in the low background data, and we discuss a method that can be used to understand and effectively suppress this type of background.

As discussed in Chapter 4, the ZIP detectors provide a two-fold signature of an interaction: the ionization and the phonon signals. The phonon signal contains three components: 1) the phonons created during the initial interaction, 2) the Luke phonons created by the charges drifting towards the electrodes, and 3) the relaxation phonons produced when the drifting charges reach the electrodes. The drifting charges are accelerated by the electric field, they collide with the lattice, and they emit the Luke phonons as they travel. When they emit phonons, they slow down to be accelerated again by the electric field. Eventually, they reach the electrodes, at which point their remaining energy is also released into the phonon system. Hence, the energy released by the drifting charges into the phonon system is:

$$P_L = NeV = \frac{Q}{\epsilon} eV, \tag{5.1}$$

where N is the number of drifting charges, e is the electron charge, and V is the charge bias voltage. It takes $\epsilon = 3.0$ eV to create an electron and a hole in Ge (3.8 eV in Si), so one can determine N from the energy Q measured by the charge channel and the parameter ϵ. Hence, the recoil energy (i.e. energy dumped into the phonon system during the initial interaction) can be extracted from the total phonon signal P_t and the Luke signal P_L:

$$E_r = P_t - P_L = P_t - \frac{Q}{\epsilon}eV. \tag{5.2}$$

In practice, we require energy calibration. In the SUF Run21, this was achieved using a ^{137}Cs source which a gamma line at 662 keV. These gammas are sufficiently energetic to punch through the internal shields, and the line can be observed in Ge detectors. In Si detectors the line is not easy to observe as very few of the 662 keV gammas are stopped in the detector - instead, the energy calibration is done by matching the data to the Monte Carlo simulation of the ^{137}Cs calibration. Furthermore, for the germanium detectors, the activation during neutron calibrations yields additional lines at 10 keV and 67 keV. These lines (and the Monte Carlo's) allow conversion of the measured voltage pulse in the charge channel into energy units. The phonon channels are then calibrated so that the recoil energy for the bulk electron-recoil events matches the energy measured by the charge channels.

The ionization signal is about 3 times smaller for a nuclear-recoil event than for the same recoil-energy electron-recoil event. We, therefore, define the ionization yield as a parameter distinguishing between ER and NR events:

$$y = \frac{Q}{E_r}. \tag{5.3}$$

Figure 5.1 shows a typical $y - E_r$ plot, made for the Ge detector Z5 in Tower 1, using the gamma and neutron calibrations. The dashed lines denote the 2σ electron and

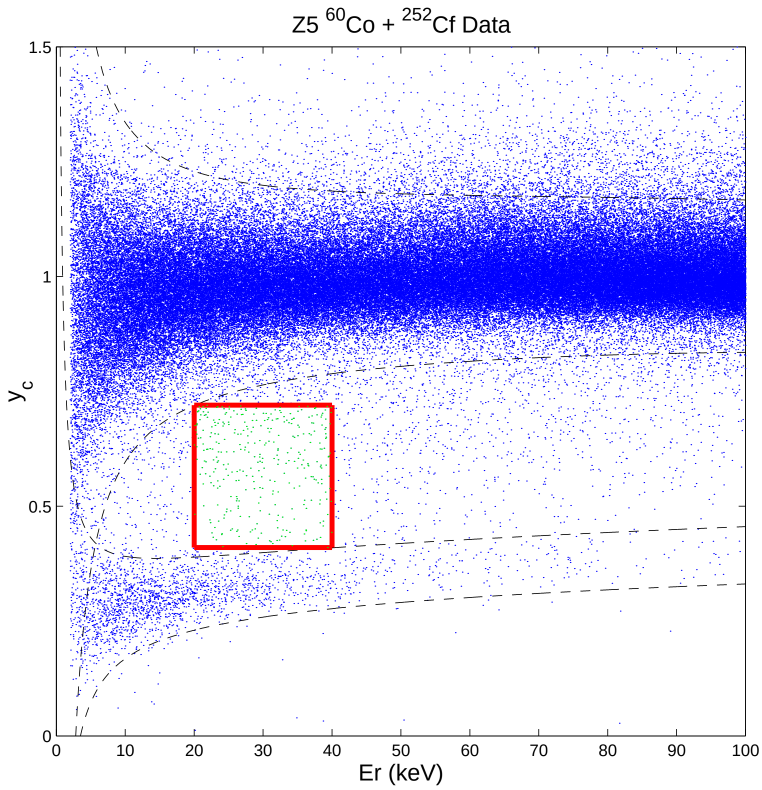


Figure 5.1: Yield vs Recoil Energy for the Ge detector Z5. The subscript "c" in y_c indicates that the phonon signal has been position corrected, following the procedure described in Chapter 6. See text for details.

nuclear-recoil bands. The means of bands are calculated by fitting a gaussian to the y-distribution in several energy bins. The means are then fitted using the following functional form:

$$y = aE_r^{b-1}, \tag{5.4}$$

and the 1σ points are fitted in the form:

$$\sigma E_r = c(aE_r^b) + d, \tag{5.5}$$

where a, b, c, and d are the free parameters. Clearly the bands are separated at $> 2\sigma$ down to below 5 keV - in fact the gamma rejection is $> 99.98\%$ in the 5-100 keV bin [182].

However, one can also observe a number of events falling between the two bands. The number of such events is larger than statistically expected, given the numbers of events in the 2σ bands. These events are the surface electron-recoil events, whose charge signal is degraded due to incomplete charge collection. Clearly, the degradation of the charge signal for such events can be sufficiently large that these events can fall into the NR band!

As mentioned above, the rise-time of the phonon signal is sensitive to the depth of the interaction. In practice, we define the phonon rise-time to be the time between 10% and 40% of the rising edge of the phonon signal. Furthermore, in the SUF Run21, we used the rise-time of the total phonon signal - the four phonon pulses added together (in the following Chapter, we will discuss other possible parameters that can be used for determining the interaction depth).

We select the events falling between the two bands in the Figure 5.1 in the 20-40 keV energy bin, and we plot them in the rise-time vs yield plane in the Figure 5.2. The neutron "blob" is centered around $y \approx 0.3$ and the gamma "blob" is centered around $y \approx 1$ in this plot. The surface ER events form a tail from the gamma "blob" into the neutron "blob". Note that the surface ER events typically have shorter (faster) rise-times compared to the neutron (i.e. NR) events.

The yield and the rise-time parameters clearly have the discriminating power (against the ER events). The question becomes: how exactly do we use these parameters? This is a non-trivial question, and it is strongly depends on how the final Dark Matter limit will be calculated and on the dominant background. The simplest (and until now, standard) approach is to define cuts in yield and rise-time separately. One can

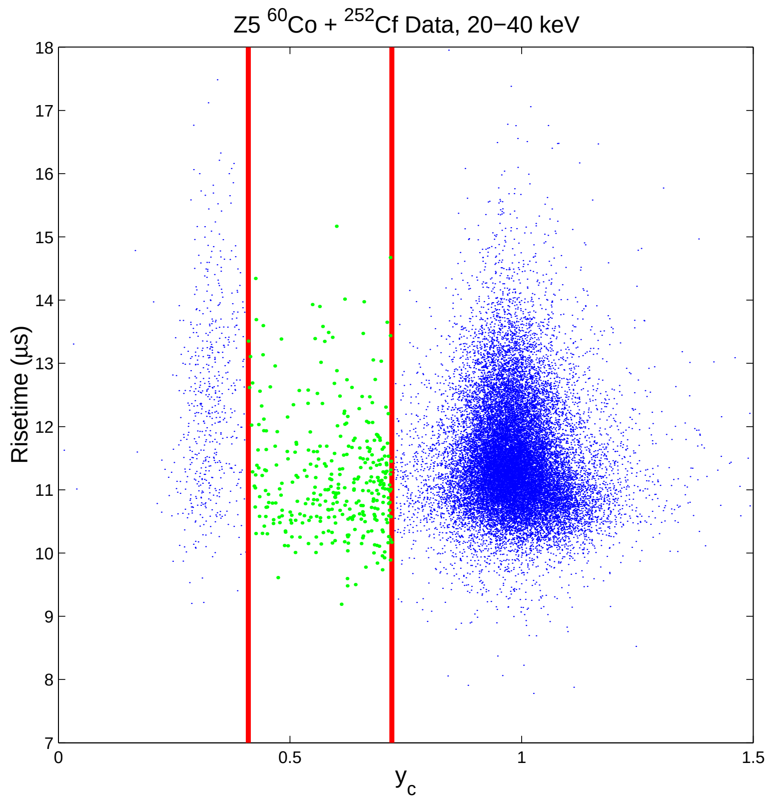


Figure 5.2: Phonon rise-time vs yield for the Ge detector Z5. Both the phonon amplitude and the rise-time parameter have been position corrected, following the procedure described in Chapter 6. See text for details.

then choose a relaxed or a harsh cut in rise-time. The relaxed cut can be chosen if the dominant background is due to neutrons, in which case a small leakage of the surface ER events is not important - this was indeed the case at the shallow site at SUF. The harsh cut could be chosen if the surface ER events are the dominant background and if we do not want to perform a statistical subtraction of these events, but rather cut them all out. This may indeed be the case at the deep site in Soudan, MN, where the neutron background is expected to be strongly suppressed. In either case, one should estimate how large the remaining surface-event background is, which is also non-trivial.

The alternative approach developed in this Chapter is based on the idea that neutron and surface ER distributions in the yield vs rise-time plane can be fitted using a maximum likelihood technique. The fitted functional forms can then be used in several different ways:

1. Integrate the functional form for the surface ER events (with or without rise-time cut) to estimate the leakage of these events into the NR band.

2. Use these functional forms to optimize the cut in the yield vs rise-time plane, and to possibly include other parameters (such as the relative position of two events in the nearest-neighbor double scatter events). This method is described in detail in [193].

3. Use a maximum likelihood method again to extract directly the number of "true" nuclear-recoil events in the low background data (this method was originally proposed by Bernard Sadoulet).

In particular, method 3) avoids cuts in the ionization yield and the phonon pulse-shape parameters and it uses the information stored in these parameters optimally. We will show that this method yields better sensitivity to WIMPs than the cut-based methods.

5.3 Fitting Surface Electron-Recoil Events

5.3.1 Risetime-Yield Correlation

We begin this section with a discussion of the correlation between the rise-time and yield parameters for the surface ER events. A priori, such correlation would not be surprising, as both parameters are sensitive to the depth of the interaction. Figure 5.3 (left) shows with a simple straight-line fit to the rise-time vs yield parameters for the surface events that such correlation indeed exists. Furthermore, we compared the distributions of rise-time for two slices in yield: the Kolmogorov-Smirnov test indicates that the distributions are indeed different, but they seem to be simply offset with respect to each other, as shown in Figure 5.4 (left). This is true for different detectors and for different energies. If the rise-time is corrected against the yield, such that the "straight-line correlation" is taken out, as shown in Figure 5.3 (right), the KS test indicates that distributions of such, corrected, rise-time in different yield slices agree, as shown in Figure 5.4 (right). This is a very convenient results, because it implies that the rise-time part and the yield part of the distribution function of the betas in the yield vs rise-time plane are effectively decoupled. In other words, one can essentially write

$$f(r, y) = A(r) \times B(y), \tag{5.6}$$

and one only needs to add the yield-dependent offset for the rise-time in the form of a straight line!

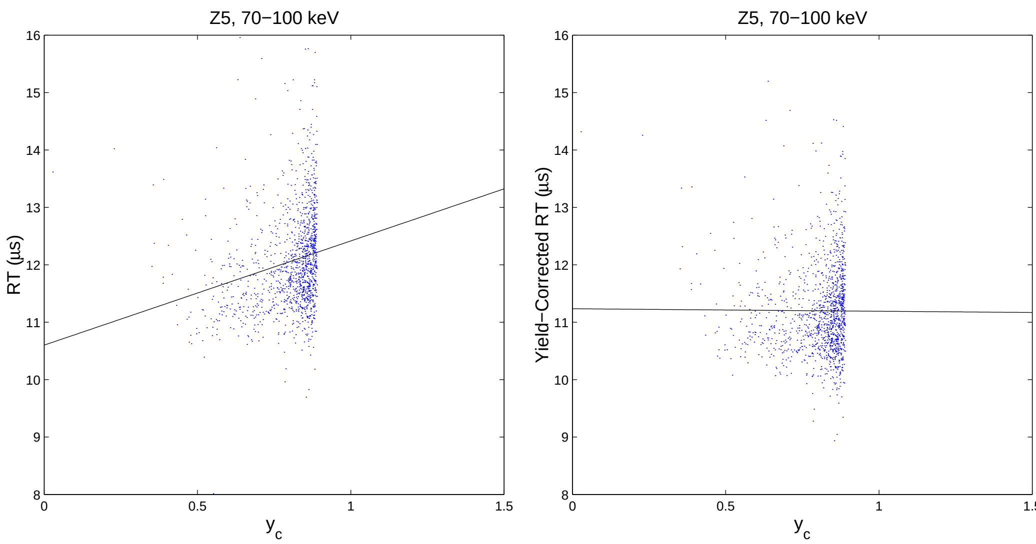


Figure 5.3: Phonon rise-time vs yield for the Ge detector Z5. Left: Fit to yield and rise-time parameters indicates some correlation. Right: yield vs rise-time correlation taken out.

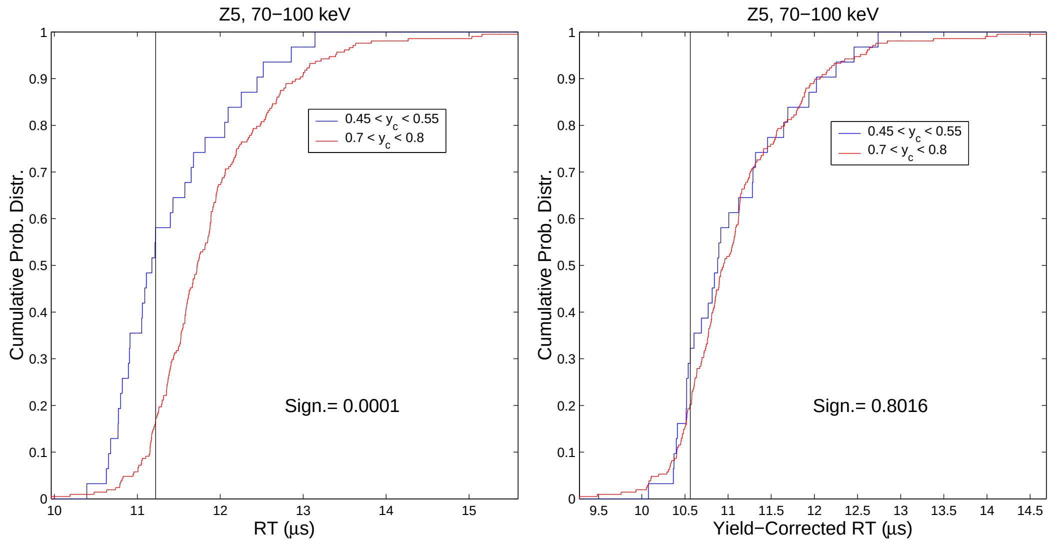


Figure 5.4: Kolmogorov-Smirnov test applied to the rise-time distributions for two slices in yield. Left: test performed on the original rise-time parameter. Right: test performed on yield-corrected rise-time parameter (i.e. corrected for the yield vs rise-time correlation).

5.3.2 Functional Form and Covariance Matrix

We choose the following functional form for fitting the surface ER event distribution in the rise-time vs yield plane:

$$f_b(y,r;\theta) = \frac{1}{N}(y-y_p)^2 \int_{x_0}^{\infty} dr' e^{-(r'-x_0)/\tau} e^{-(r-r')^2/2\sigma^2}, \tag{5.7}$$

where $x_0 = a(y-p)+b$ and

$$N = \sqrt{2\pi}\sigma\tau\Big(\frac{(y_{max}-y_p)^3}{3} - \frac{(y_{min}-y_p)^3}{3}\Big). \tag{5.8}$$

The parameters a and b take care of the yield vs rise-time correlation discussed above. The yield part of the functional form is taken to be quadratic, with a single free parameter y_p - there is no physical motivation for such choice, it was empirically determined to work well. The rise-time part of the functional form is taken to be a convolution of a gaussian (with parameter σ) and an exponential (with parameter τ). While the gaussian reflects the resolution in the rise-time, the exponential is chosen to reflect the low rise-time vs high rise-time asymmetry in the distribution. Again, there is no physical motivation to use the exponential, but it was determined to work well empirically. The parameter p is a constant usually set to 0.6 to remove the correlation between a and b (this simplifies the likelihood maximization). The parameters y_{max} and y_{min} determine the maximum and minimum yield values over which the fitting will be performed. These are essentially set by the lower 3σ bound of the ER band and the upper 2σ bound of the NR band, respectively. Finally, θ in $f_b(y, r; \theta)$ denotes the vector of the free parameters discussed above.

The likelihood function is then given by

$$\mathcal{L} = \prod_i f_b(y_i, r_i; \theta), \tag{5.9}$$

where i runs over all events to be fitted. The covariance matrix for the free parameters in the maximum likelihood method can be calculated by inverting the expected value of the Hessian matrix for the logarithm of the likelihood function:

$$Cov = -E\Big(\frac{\partial^2(\log \mathcal{L})}{\partial\theta_i\partial\theta_j}\Big)^{-1}. \tag{5.10}$$

However, calculating the second derivatives of our functional form would be quite involving. Instead, we use the fact that

$$\begin{aligned} E\Big(\frac{\partial^2(\log \mathcal{L})}{\partial\theta_j\partial\theta_k}\Big) &= E\Big(\sum_i \frac{\partial^2(\log f_b(y_i, r_i; \theta))}{\partial\theta_j\partial\theta_k}\Big) \\ &= -E\Big(\sum_i \frac{\partial f_b(y_i, r_i; \theta)}{\partial\theta_j}\frac{\partial f_b(y_i, r_i, \theta)}{\partial\theta_k}\Big). \end{aligned} \tag{5.11}$$

The last equality can be easily shown by differentiating twice

$$\int f_b(y, r; \theta) dy dr = 1. \tag{5.12}$$

Hence, we can get away with calculating only the first derivatives. They are given by:

$$
\begin{aligned}
\frac{d\log f_b}{dy_p} &= \frac{-2}{y-y_p} + \frac{3[(y_{max}-y_p)^2-(y_{min}-y_p)^2]}{(y_{max}-y_p)^3-(y_{min}-y_p)^3}, \\
\frac{d\log f_b}{da} &= \frac{y-p}{\tau} - \frac{y-p}{I_0} e^{-(\sigma^2/\tau+a(y-p)+b-r)^2/2\sigma^2}, \\
\frac{d\log f_b}{db} &= \frac{1}{\tau} - \frac{1}{I_0} e^{-(\sigma^2/\tau+a(y-p)+b-r)^2/2\sigma^2}, \\
\frac{d\log f_b}{d\tau} &= -\frac{\sigma^2}{\tau^3} + \frac{r-a(y-p)-b}{\tau^2} - \frac{1}{\tau} + \frac{\sigma^2}{I_0\tau^2} e^{-(\sigma^2/\tau+a(y-p)+b-r)^2/2\sigma^2}, \\
\frac{d\log f_b}{d\sigma} &= \frac{\sigma}{\tau^2} - \frac{1}{\sigma^2} - \frac{2\sigma}{I_0\tau} e^{-(\sigma^2/\tau+a(y-p)+b-r)^2/2\sigma^2} + \frac{1}{I_0}\int_0^\infty dr' \frac{(r'-\mu)^2}{\sigma^3} e^{-(r'-\mu)^2/2\sigma^2}, \\
I_0 &= \int_0^\infty dr' e^{(r'-\mu)^2/2\sigma^2}, \\
\mu &= r - \sigma^2/\tau - a(y-p) - b. \qquad (5.13)
\end{aligned}
$$

5.3.3 Results and Systematic Errors

The fits are performed for each detector separately, to allow for the detector-to-detector variations, and for several energy bins, to allow for the energy dependence of the free parameters. The energy bins could not be made too small, otherwise there would be too few events to fit. A good compromise is to choose the bins $5-10, 10-20, 20-40$ and $40-100$ keV - this combination captures most of the energy dependence, while providing enough events per bin to perform the fits.

We also performed the fits to three different sets of events, in order to get handles on various sources of systematic error:

1. Events from the ^{60}Co calibration, between the electron and the nuclear-recoil bands.

2. Events from the background data, between the electron and the nuclear-recoil bands.

3. Events from the ^{60}Co calibration, below the electron-recoil band (i.e. including the

nuclear-recoil band events).

Including events in the nuclear-recoil band of the ^{60}Co calibration (case 3) is meaningful because the calibration was relatively short, so the number of true nuclear-recoils is expected to be very small. Furthermore, since most nuclear-recoils are coincident with hits in the muon-veto, the number of these nuclear-recoils is further reduced by the muon-veto cut. In other words, most of the events in the nuclear-recoil band of the ^{60}Co calibration are leaked surface events. Hence, in this way we can directly measure the leakage of surface events into the nuclear-recoil band.

Comparing results for cases 1) and 3) indicates the size of the systematic error due to extrapolating the fit based on events between the two bands into the nuclear-recoil band. As shown below, we found that the cases 1) and 3) yield consistent results, so the systematic error due to our choice of the functional form is reasonably small. Comparing case 2) to cases 1) and 3) tests whether there are any systematic differences in distributions of the ^{60}Co calibration surface events and the background surface events. Again, as shown below, the estimates for the free parameters in all three cases agree well, indicating that this source of systematic error is also reasonably small.

The Figure 5.5 demonstrates the performance of the fitting algorithm for the Ge detector Z5 in the 20-40 keV bin. We generated and inspected such plots for all detectors, for all energy bins, and for several slices in rise-time and in yield. This allowed visual verification that the fit was reasonable. We further verified in each $(detector, energy\ bin)$ case that the maximum of the likelihood was indeed found, by examining the behavior of the likelihood function around the estimated maximum point. Finally, we verified in each

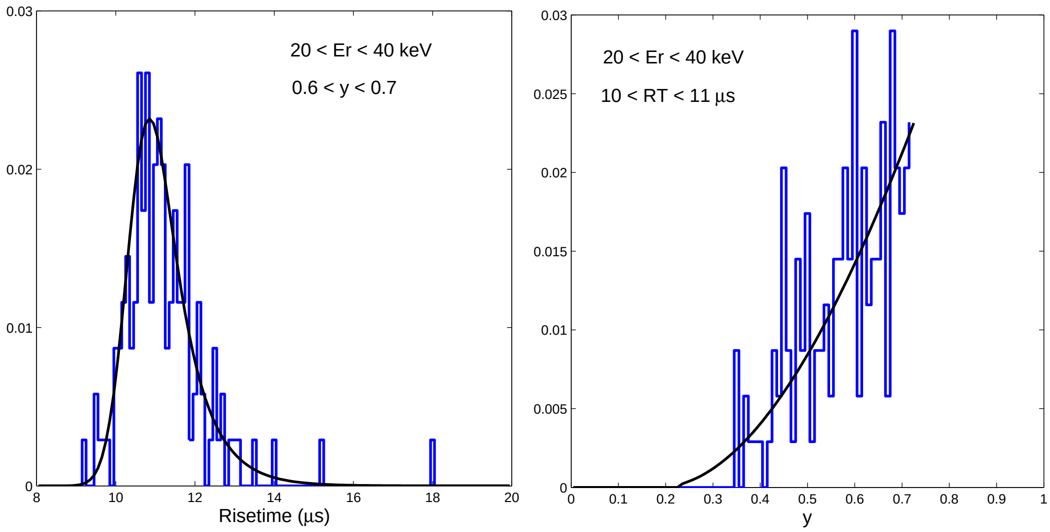


Figure 5.5: Fitting surface events: The plot on the left compares the histogram of rise-time in a yield slice to the maximum likelihood fit. The plot on the right compares the histogram of the yield in a rise-time slice to the maximum likelihood fit. Both plots use recoil energies between 20 keV and 40 keV, and are made using the ^{60}Co beta/gamma calibration. Histograms have been scaled by the total number of events used in the fit.

case that the required tolerances on the change in the likelihood function value for the last iterative step have been reached (i.e. that the algorithm actually converged to a solution).

Figure 5.6 shows the estimate of the free parameters for the Ge detector Z5, and for the 3 sets of events considered. Note that the three estimates of the free parameters agree well (within the estimated errors on the free parameters, in most cases). Hence, the systematic errors due to our choice of the functional form, or due to differences in distributions of the surface events for the background data and for the ^{60}Co calibration, are not expected to be larger than the estimated errors on the free parameters.

Note also that most of the free parameters have some energy dependence. While σ is the resolution in rise-time, and is expected to have $1/E_r$ dependence, the rest of the parameters were chosen empirically, and their energy dependence is more difficult to

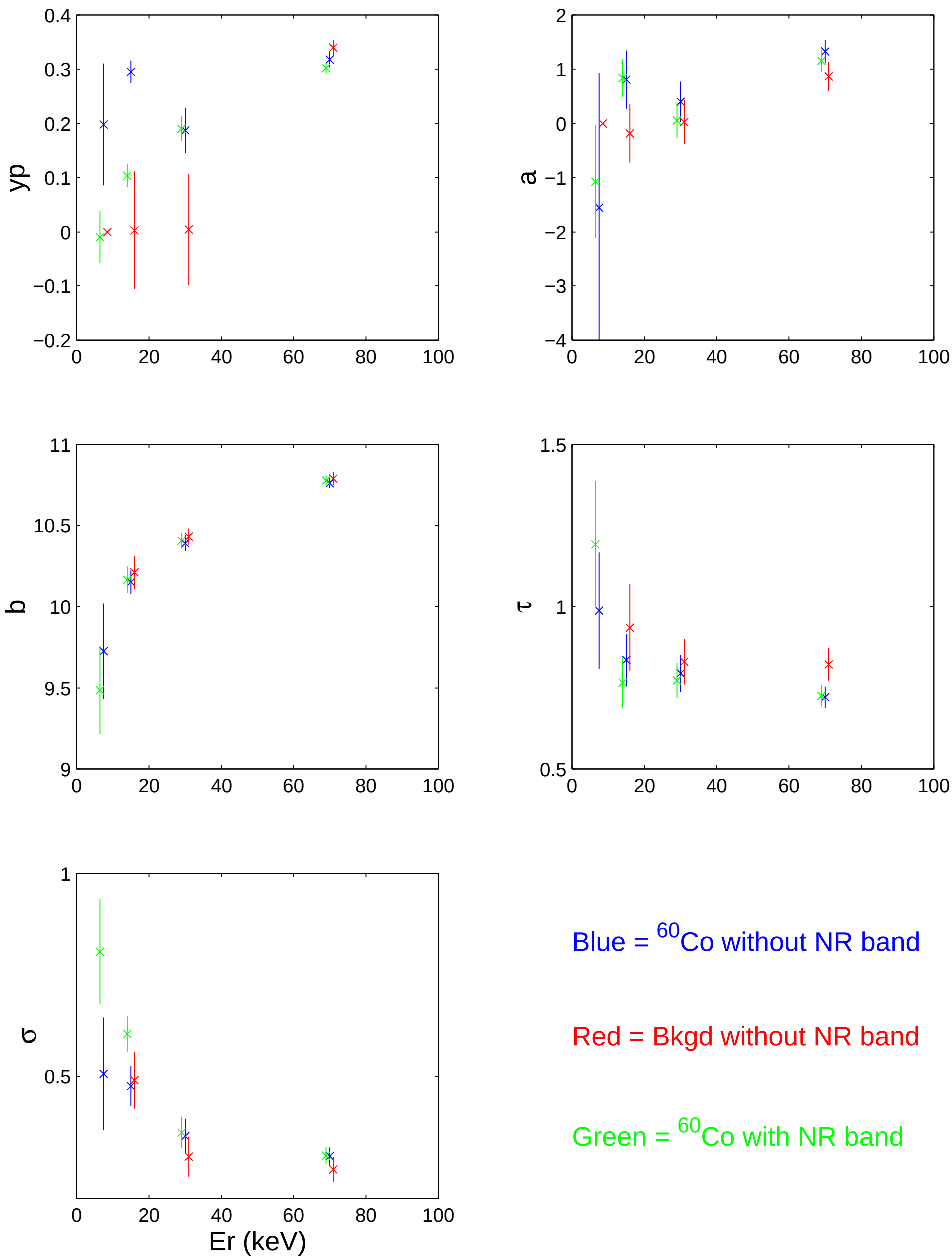


Figure 5.6: Estimates of the free parameters of the surface event distribution function for the Ge detector Z5.

understand intuitively. Nevertheless, one could attempt to define these parameters to be energy dependent, and to perform the fit to the data for all energies at once (i.e. in the 5-100 keV bin). We do not attempt this here.

5.4 Fitting Neutrons

5.4.1 Functional Form

We choose the following functional form for fitting the distribution of the NR events in the rise-time vs yield plane:

$$f_n(y,r;\theta) = \frac{1}{2\pi\sigma_r\sigma_y} e^{-[y-a(r-p)-b]^2/2\sigma_y^2} \int_{x_0}^{\infty} e^{-(r'-x_0)/\tau} e^{-(r-r')^2/2\sigma_r^2} dr'. \tag{5.14}$$

The yield part of the functional form is now a gaussian with a rise-time dependent mean (free parameters are a and b, while p is a constant introduced to remove the correlation between a and b) and standard deviation σ_y. While it is possible to use a gaussian for the rise-time part of the functional form as well, we choose the convolution of a gaussian and an exponential as that fits the data a bit better and it is consistent with the functional form chosen for the surface events. The free parameters are τ, x_0, and σ_r. Similarly to the case of the surface events, we will calculate the covariance matrix using the Equation 5.11. The derivatives are given by:

$$\begin{aligned}
\frac{\partial \log f_n}{\partial a} &= \frac{(r-p)[y-a(r-p)-b]}{\sigma_y^2} \\
\frac{\partial \log f_n}{\partial b} &= \frac{y-a(r-p)-b}{\sigma_y^2} \\
\frac{\partial \log f_n}{\partial \sigma_y} &= -\frac{1}{\sigma_y} + \frac{[y-a(r-p)-b]^2}{\sigma_y^3}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial \log f_n}{\partial \tau} &= -\frac{1}{\tau} + \frac{r - x_0}{\tau^2} - \frac{\sigma_r^2}{\tau^3} + \frac{\sigma_r^2}{I_0 \tau^2} e^{-(x_0 - r + \sigma_r^2/\tau)^2/2\sigma_r^2} \\
\frac{\partial \log f_n}{\partial x_0} &= \frac{1}{\tau} - \frac{1}{I_0} e^{-(x_0 - r + \sigma_r^2/\tau)^2/2\sigma_r^2} \\
\frac{\partial \log f_n}{\partial \sigma_r} &= \frac{\sigma_r}{\tau^2} - \frac{1}{\sigma_r} - \frac{2\sigma_r}{I_0 \tau} e^{-(x_0 - r + \sigma_r^2/\tau)^2/2\sigma_r^2} \\
&+ \frac{1}{I_0} \int_0^\infty \frac{(r' - r + x_0 + \sigma_r^2/\tau)^2}{\sigma_r^3} e^{-(r' + x_0 - r + \sigma_r^2/\tau)^2/2\sigma_r^2} dr' \\
I_0 &= \int_0^\infty e^{-(r' + x_0 - r + \sigma_r^2/\tau)^2/2\sigma_r^2} dr'. \qquad (5.15)
\end{aligned}
$$

5.4.2 Results

The functional form given above is fitted to the neutron events in the ^{252}Cf calibration. Similarly to the case for surface events, the fits are performed for each detector and for each energy bin separately. The energy bins chosen are $5-10, 10-20, 20-40$, and $40-100$ keV.

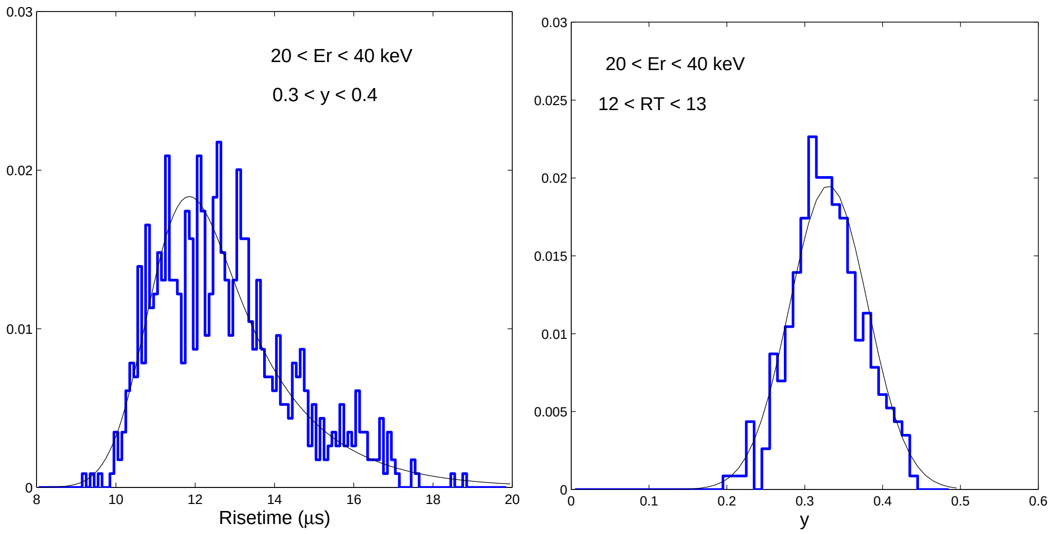


Figure 5.7: Fitting neutrons: The plot on the left compares the histogram of rise-time in a yield slice to the maximum likelihood fit. The plot on the right compares the histogram of the yield in a rise-time slice to the maximum likelihood fit. Both plots use recoil energies between 20 keV and 40 keV, and are made using the ^{252}Cf neutron calibration. Histograms have been scaled with the total number of events used in the fit.

Figure 5.7 demonstrates the performance of the fitting algorithm for the neutron events in the Ge detector Z5. The same checks discussed for the case of surface events were repeated for the case of NR events (for every detector and every energy bin) in order to verify that the algorithm indeed converged to a maximum of the likelihood function and that the fits are reasonable.

Figure 5.8 shows the estimated free parameters of the neutron distribution function. Similarly to the case of the surface events, many free parameters are energy dependent. The $\sim 1/E_r$ dependence of σ_y and σ_r is expected, since these are resolutions in yield and in rise-time. The other parameters were chosen empirically, so their energy dependence is not as intuitive.

5.5 Surface Events Rejection Efficiency

Once the free parameters of the functional form $f_b(y, r; \theta)$ have been determined, one can use the functional form to estimate the leakage of the surface events into the NR band (with or without a cut in rise-time). To do this, we integrate the functional form over three regions:

- A: between the ER and NR bands in yield and over all rise-time.
- B: over the NR band region in yield and over all rise-time.
- C: over the NR band region in yield and over $r_0 - \infty$ in rise-time, where r_0 is the value at which the cut in rise-time is placed.

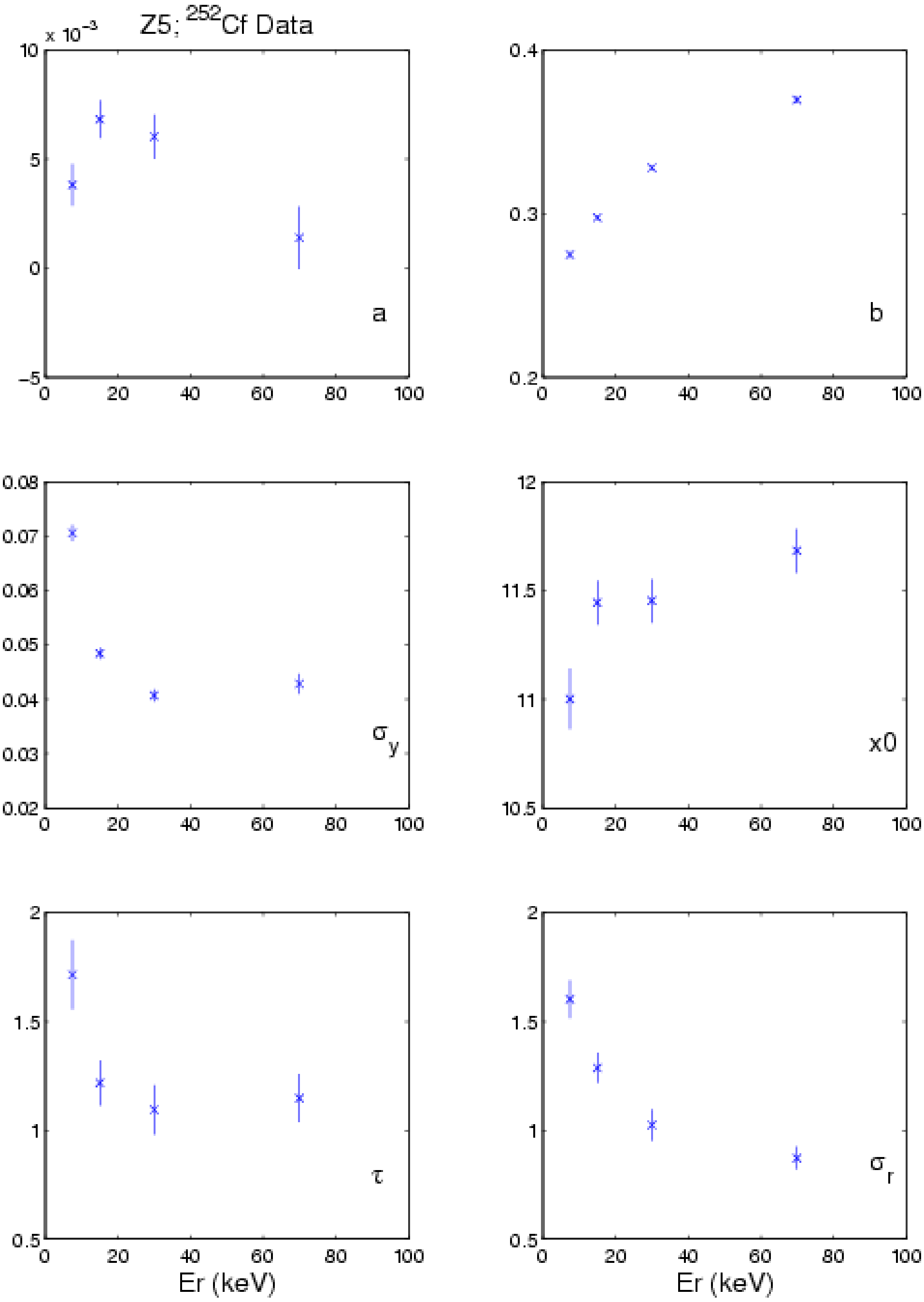


Figure 5.8: Estimates of the free parameters of the neutron (i.e. nuclear-recoil) distribution function for the Ge detector Z5.

The ratio of integrals over regions B and A (C and A), along with the number of events observed in the region A in the low background data, gives an estimate of the number of surface events that leaked into the NR band before (after) the rise-time cut.

The integration over regions A and B is straightforward, as the rise-time part of the function integrates out:

$$\begin{aligned} V &= \int_{y_1}^{y_2} dy \int_0^{\infty} dr f_b(y, r; \theta) \\ &= \frac{(y_2 - y_p)^3 - (y_1 - y_p)^3}{(y_{max} - y_p)^3 - (y_{min} - y_p)^3}. \end{aligned} \tag{5.16}$$

The only free parameter uncertainty that affects this integral is the one of y_p.

The integration over region C is more complicated because the rise-time part does not integrate out. The integral is then evaluated numerically:

$$\begin{aligned} V &= \int_{y_1}^{y_2} dy \int_{r_0}^{\infty} dr f_b(y, r; \theta) \\ &= \frac{1}{N} \int_{y_1}^{y_2} dy (y - y_p)^2 \int_0^{\infty} dr' e^{-r'/\tau} \int_{r_0}^{\infty} dr e^{-(r - r' - x_0)^2/2\sigma^2}. \end{aligned} \tag{5.17}$$

Furthermore, errors of all free parameters affect the result, and most of derivatives necessary for the error propagation have to be evaluated numerically. The calculations of derivatives are straightforward and not very illuminating, so we avoid them here.

We perform such integrations to calculate the ratios B/A and C/A, for all detectors, energy bins, and for the three different cases described above. This gives us the fractional leakage of surface events into the NR band, with respect to the region A. We then count the number of events in the low background data in the region A, and multiply with the appropriate ratio (B/A or C/A) to estimate the number of surface events that leaked into the NR band. Table 5.1 compares the results (integrated over 10-100 keV) for the 3 cases described above, for the SUF Run21 -3V dataset without the cut in rise-time, for four Tower 1 detectors (Z2, Z3, Z4, and Z5). It also compares these results to the "direct" estimate of the leakage, which calculates the fractional leakage by dividing the numbers of events in the regions B and A in the ^{60}Co calibration (and accounting for the "neutron contamination" in region B). The results of the four estimates agree within the error bars in most cases, although the case of ^{60}Co data without NR band tends to give somewhat smaller estimates than other methods.

Detector	Direct	^{60}Co w/o NR	Bkgd w/o NR	^{60}Co w/ NR
Z2	0.5 ± 0.2	0.1 ± 0.1	0.2 ± 0.1	0.4 ± 0.1
Z3	0.6 ± 0.2	0.6 ± 0.3	1.3 ± 0.6	0.8 ± 0.3
Z4	1.0 ± 0.3	0.4 ± 0.2	1.0 ± 0.4	1.4 ± 0.5
Z5	2.6 ± 0.3	1.9 ± 0.6	5.1 ± 1.3	2.9 ± 0.5

Table 5.1: Estimates of the number of the surface events that leaked into the NR band for the SUF Run21 -3V dataset (without the cut in rise-time): column 2 - Rough estimate based on ^{60}Co data; Column 3 - Estimate based on fits on ^{60}Co data without NR band; Column 4 - Estimate based on fits on background data without NR band; Column 5 - Estimate based on fits on ^{60}Co data with NR band.

Another way to look at these results is to calculate the rejection efficiency for the surface events. The surface-event rejection, for the case before the rise-time cut, in terms of integrals over regions A and B is defined as:

$$rej = 1 - \frac{\mathrm{B}}{\mathrm{A}} = \frac{\mathrm{A}}{\mathrm{A}+\mathrm{B}}, \tag{5.18}$$

and similarly for the case after the rise-time cut. One can also define the effective surface-event rejection as:

$$Q_{rej} = 1 - Q = 1 - \frac{\beta(1-\beta)}{(\alpha-\beta)^2} \tag{5.19}$$

where Q is the quality factor (as defined in [194]) and α and β are the efficiencies of the cut (with or without the rise-time cut) on the NR events and on surface events respectively. α

is estimated from the ^{252}Cf calibration and $\beta = 1 - rej$. Intuitively, Q_{rej} measures how well betas are rejected for the given efficiency α on the NR events. Figure 5.9 shows these rejection factors as a function of energy, before and after the rise-time cut. Note that we have integrated the 3 Ge detectors used in this analysis (Z2, Z3 and Z5).

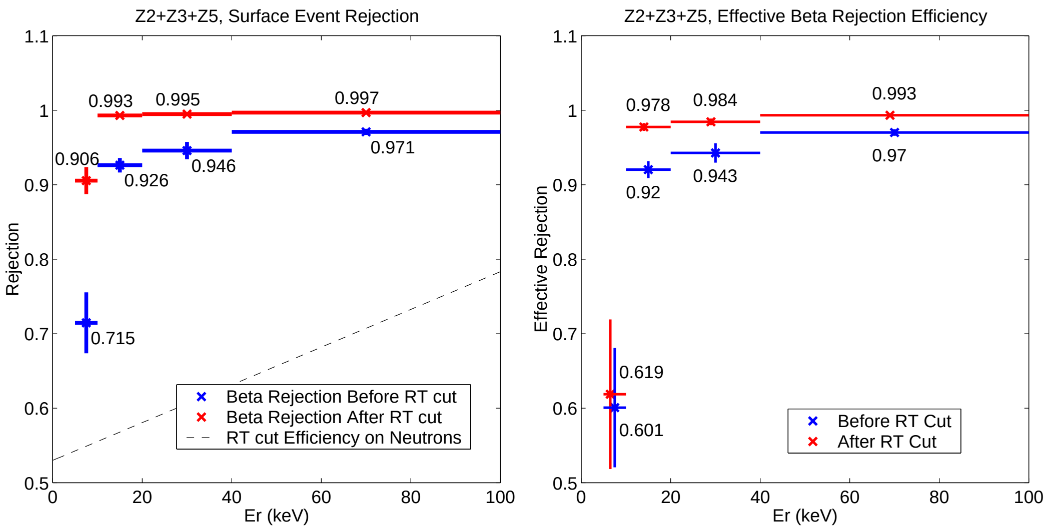


Figure 5.9: Rejection efficiency (left) and effective rejection (right) as a function of energy, integrated over 3 Ge detectors of Tower 1, for the SUF R21 -3V dataset, before (blue) and after (red) the rise-time cut. The dashed line denotes the efficiency of the rise-time cut on the NR events.

5.6 Maximum Likelihood Method

5.6.1 Maximum Likelihood Estimate

The method described here was originally proposed by Bernard Sadoulet. We will develop the method for a general setting, and then apply it for our case of yield and rise-time parameters.

Consider a detector that provides the energy of the recoiled particle and P parameters $p_k,\ k = 1..P$, that provide information about the nature of the interaction. Assume also that there are M signals and backgrounds, that all of them can be studied independently, and that their distributions in the discriminator space $\vec{p}$ can be estimated. Denote these distribution functions as $f_m(\vec{p}, \theta_m)$, where $m = 1, 2, ..., M$, and θ_m is the set of parameters determining the m^{th} distribution function. The parameters θ_m can be estimated from some underlying theory that describes the distribution of events in the given parameters, or from empirical fits to detector calibrations. Note, also, that the number of distribution functions M and the number of discriminating parameters P need not be related. With these assumptions, the likelihood function if given by:

$$\begin{aligned} \mathcal{L} &= e^{-N_T} \frac{N_T^N}{N!} \prod_{i=1}^{N} \left[\sum_{m=1}^{M} \frac{B_m}{N_T} f_m(\vec{p}_i, \theta_m) \right] = \frac{1}{N!} e^{-N_T} \prod_{i=1}^{N} \left[\sum_{m=1}^{M} B_m f_m(\vec{p}_i, \theta_m) \right], \\ N_T &= \sum_{m=1}^{M} B_m. \end{aligned} \tag{5.20}$$

In this expression, B_m are the amplitudes of each of the signals and backgrounds - they are to be estimated. The term outside the product is Poissonian - it fixes $\sum_{m=1}^{M} B_m = N$, where N is the total number of observed events, and it automatically incorporates the Poisson error estimates into the final estimate of the errors on B_m's. The second term is the product over all events of the value of the distribution function for each event. The distribution function is taken to be the weighed sum of the individual distribution functions corresponding to the various signals and backgrounds.

The likelihood function can be maximized by solving the system of M equations:

$$\frac{\partial(\log \mathcal{L})}{\partial B_m} = \sum_{i=1}^{N} \frac{f_m(\vec{p}_i, \theta_m)}{\sum_{n=1}^{M} B_n f_n(\vec{p}_i, \theta_n)} - 1 = 0. \tag{5.21}$$

The covariance matrix can be estimated by inverting the Hessian matrix of the (logarithm of the) likelihood function:

$$E\left(-\frac{\partial^2(\log\mathcal{L})}{\partial B_m \partial B_n}\right) = \sum_{i=1}^{N} \frac{f_m(\vec{p}_i,\theta_m) f_n(\vec{p}_i,\theta_n)}{(\sum_{k=1}^{M} B_k f_k(\vec{p}_i,\theta_k))^2}. \tag{5.22}$$

Hence, maximization of the likelihood yields estimates of the parameters B_m (i.e. the amplitude of each of the M signals and backgrounds), and the errors on these parameters, including the Poisson statistical uncertainties.

5.6.2 Systematic Errors

The method described above relies on the distribution functions, defined by the parameters θ_m, which would have to be estimated previously. The errors on θ_m are not propagated. One can approach this problem in several different ways.

First, if N, is small, which is often the case, the Poisson error dominates over all other sources of error. In this case, one can usually neglect the errors on θ_m.

Second, if N is large, one could include θ_m in the formulation given above, and the method would directly estimate these parameters (and their errors), along with B_m's. The problem with this approach, however, is that the number of free parameters can become large, which could make the likelihood maximization difficult.

Third, one can calculate the change in the estimate of B_m's when each of the parameters θ_m is changed by 1σ. One can then conservatively choose the largest such change as the upper bound on the estimate of the errors on B_m's due to the uncertainty in the parameters θ_m.

5.6.3 Example of Single Dominant Background

We demonstrate the method on the simple case of one signal (S) and one background (B) distribution functions, with a single discriminator parameter: $f_s(p)$ and $f_b(p)$. Then, Equation 5.20 becomes:

$$L = \frac{1}{N!}e^{-N_T}\prod_{i=1}^{N}\left[\frac{S}{S+B}f_s(p_i,\theta_s) + \frac{B}{S+B}f_b(p_i,\theta_b)\right] \tag{5.23}$$

We further assume $S << B$, which is a common case in experiments such as direct dark matter search experiments. With this assumption it is straightforward to evaluate Equation 5.22:

$$E(-\frac{\partial^2(\log L)}{\partial B_m \partial B_n}) = \frac{1}{B}\begin{pmatrix} 1 & -1 \\ -1 & \int \frac{f_s^2(p)}{f_b(p)}dp \end{pmatrix}. \tag{5.24}$$

Inverting this matrix yields gives the covariance matrix. The variance on the signal estimate is given by:

$$\sigma_s^2 = \frac{B}{\int \frac{f_s^2(p)}{f_b(p)}dp - 1}. \tag{5.25}$$

Assume, for simplicity, that both signal and background distributions are gaussian, with means 0 and μ respectively, and with the same standard deviation σ. The last Equation then becomes:

$$\sigma_s^2 = \frac{B}{e^{\mu^2/\sigma^2} - 1}. \tag{5.26}$$

If instead we apply a cut $p < \lambda$ in the discriminator p, N_p events will pass it and N_f events will fail it. Let α be the efficiency of the cut on the signal and let β be the efficiency of the cut on the background. With such definitions, and following [194], we get:

$$\begin{aligned} N_p &= \alpha S + \beta B \\ N_f &= (1-\alpha)S + (1-\beta)B. \end{aligned} \tag{5.27}$$

Finally, assuming $S << B$, we get:

$$\begin{aligned} S &= \frac{(1-\beta)N_p - \beta N_f}{\alpha - \beta} \\ \sigma_s^2 &= B\frac{\beta(1-\beta)}{(\alpha-\beta)^2} = BQ. \end{aligned} \tag{5.28}$$

In the last equality we have defined the quality factor Q.

Finally, in Figure 5.10 we compare the maximum likelihood method to the simple cut method. The maximum likelihood method performs better than the cut method (σ_s is smaller), regardless of the value of the cut parameter λ. Note that the optimal cut value, λ_{opt}, is negative and it is pushed to lower values as μ increases. This result holds only if $S = 0$. If S is allowed to be small but non-zero, the Equation 5.28 has to be modified to include the S term as well as the B term, and λ_{opt} is quickly pushed to the more intuitive positive range. We note, however, that the result discussed here is asymptotic in nature (i.e. it holds for large $N_p + N_f$). The case of a small sample size will be studied in the future.

5.6.4 Case for ZIP Detectors

We now return to our problem of the surface event background and to the yield and the rise-time parameters. In the notation described above, we have a two-vector of discriminating parameters: $\vec{p} = [y\ r]$. We assume that the neutrons and WIMPs have identical distribution in the rise-time vs yield plane, $f_n(y, r, \theta_n)$ given by Equation 5.14. We consider only one background distribution, $f_b(y, r, \theta_b)$, of the surface electron-recoil events, given by Equation 5.7. With these notations, Equation 5.20 becomes:

$$\mathcal{L} = e^{-(S+B)}\frac{(S+B)^N}{N!}\prod_{i=1}^{N}\left[\frac{S}{S+B}f_n(y_i, r_i, \theta_n) + \frac{B}{S+B}f_b(y_i, r_i, \theta_b)\right], \tag{5.29}$$

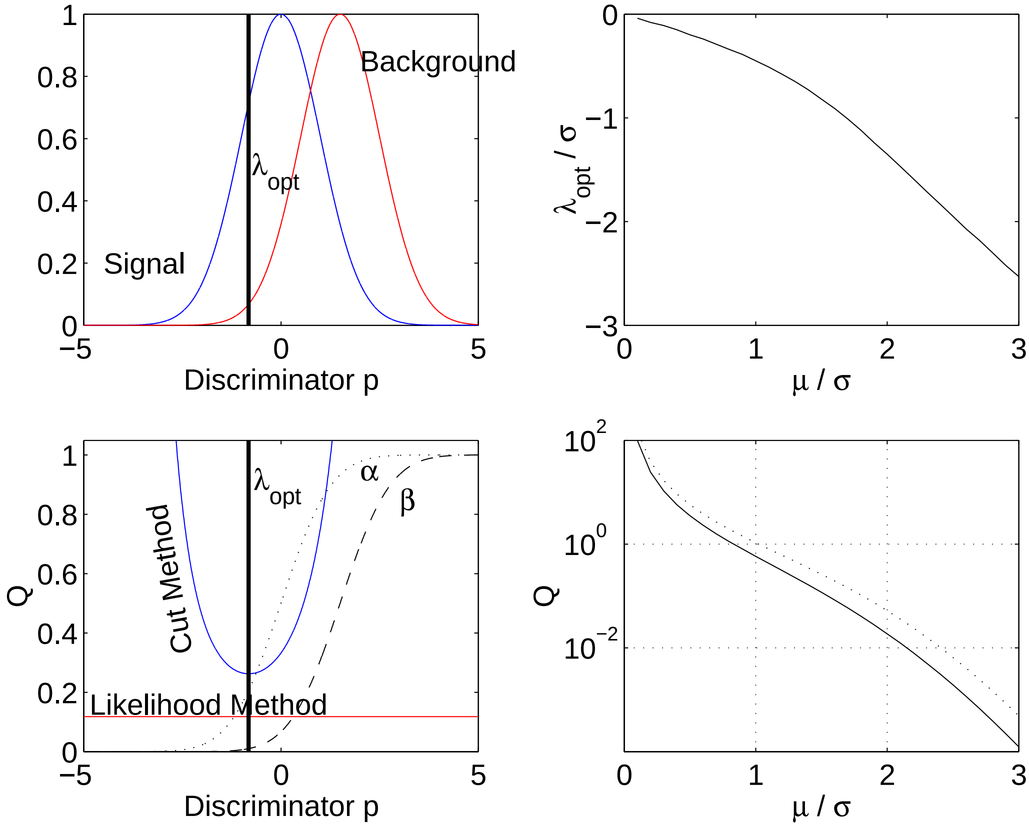


Figure 5.10: The top-left figure shows the gaussian distributions of the signal and the background normalized to 1 with means at 0 and $\mu = 1.5$ respectively, and with standard deviation $\sigma = 1$. The bottom-left figure shows that Q calculated from the cut method (solid curve) is always below the one calculated from maximum likelihood method (horizontal line). λ_{opt} is the optimal cut in the sense that it minimizes σ_s^2 for the cut method. Also, the cut efficiencies are displayed as dotted (signal; α) and dashed (background; β) lines. The top-right figure shows the dependence of λ_{opt} on μ, and the bottom-right figure shows that Q calculated for λ_{opt} (dotted line) is always larger (for any μ) than Q calculated using the likelihood method (solid line).

where S and B are estimates of the number of true nuclear-recoil events and of the number of surface electron-recoil events respectively.

The Figure 5.11 shows the result of maximizing this likelihood function for the Ge detector Z5 in the 20-40 keV recoil energy range. The events used in the likelihood maximization are all low background events below the 3σ bound of the ER band. The

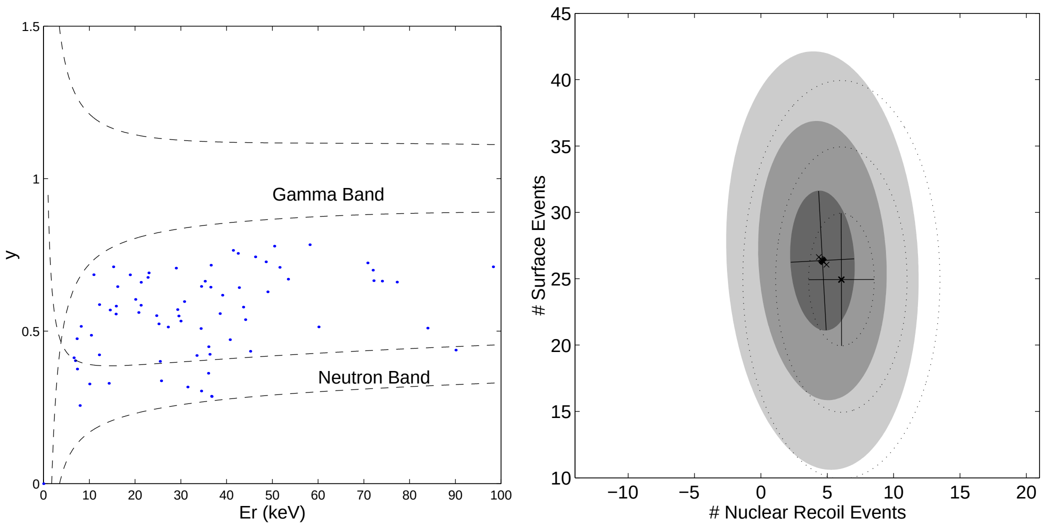


Figure 5.11: The plot on the left shows the background events used for the likelihood maximization, along with the ER and NR bands. The plot on the right shows the 1σ, 2σ, and 3σ contours of the estimate of the number of true nuclear-recoils and of the number of surface events for the Ge detector Z5 in the 20-40 keV bin. For the shaded contours, we used the surface-event distribution based on ^{60}Co data including NR events, while for the dotted contours we used the surface-event distribution based on the background data without NR events. Also, 1σ error bars are shown as solid lines. The x's denote the best estimates if one of the free parameters of the fitted functions is changed by 1σ.

dominant source of error is the Poisson error on the observed number of the low background events. Two estimates were made: one using the surface-event distribution function based on the ^{60}Co events including the NR band and one based on the background events excluding the NR band (cases 2 and 3 described in Section 5.3.3; case 1 is a subset of case 3, so we decided not to pursue it here). Note that the Poisson error is much larger than the difference between the two estimates. Hence, any systematic difference between ^{60}Co calibration data and the background data is not a significant effect.

The Poisson error is also much larger than the variation in the estimates of S and B due to changing any of the free parameters in the fitting functions by 1σ. Hence, the systematic errors due to possibly wrong functional forms are also not a concern.

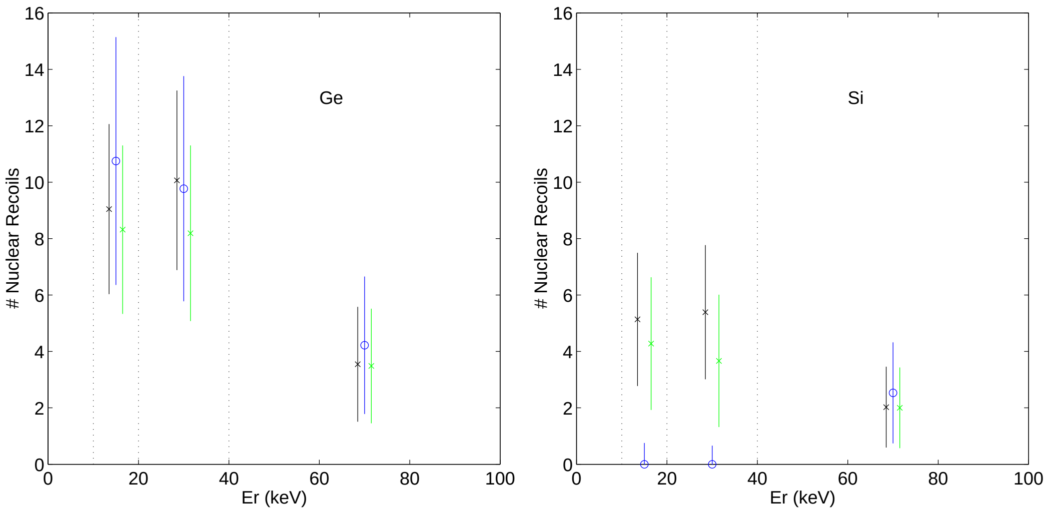


Figure 5.12: The estimate of the number of true nuclear-recoil events as a function of energy for the three Ge detectors (left) and for the single Si detector (right). The dotted lines denote the energy bins: 10-20 keV, 20-40 keV, 40-100 keV. For the green (gray) x-points (and the corresponding 1σ error bars), we used the surface event distribution based on the ^{60}Co data including NR events (case 2); for the black x-points and the corresponding error bars, we used the surface event distribution based on the background data without NR events (case 3). The o-points and the corresponding error bars correspond to the phonon rise-time cut method ([182]), adjusted for the cut efficiency. Note that the error bars on the o-points are larger than those on the x-points. Also note that for low energy bins in Si, no events pass the rise-time cut.

We perform this estimate for all detectors and in all energy bins and we sum over Ge detectors. The final results for the 10-100 keV region are shown in the Figure 5.12.

Figure 5.12 shows (similarly to Figure 5.11) that the cases 2 and 3 described in Section 5.3.3 agree well. It also shows that the maximum likelihood method described above gives a better estimate of the number of the true nuclear-recoil events (smaller error bars) than the traditional approach using a cut in the phonon rise-time. In particular, integrated over the Ge detectors Z2, Z3 and Z5 of Tower 1, and over 10-100 keV recoil-energy, we estimate: (1) 20.0 ± 4.8 events using surface event distribution based on the ^{60}Co data including NR events (case 2); (2) 22.7 ± 4.8 events using the surface event distribution based on the background data without NR events (case 3); (3) 22.4 ± 5.8 events using a rise-time cut method ([182]) and after correcting for the 67% cut efficiency. In other words, we expect improvement in sensitivity of about 17%. For reference, note that doubling the exposure (or detector mass) would improve the sensitivity by $1/\sqrt{2}$, or by 29%, so the improvement that this method offers is significant. This expected improvement in the sensitivity to the WIMP signal is illustrated in Figure 5.13.

Note, however, that the main motivation for developing this statistical method is not to make the marginal improvements in the sensitivity shown in Figure 5.13. Rather, we expect the surface event background to be much more significant, if not dominant, at the deep site at Soudan, MN, where the neutron background is expected to be significantly suppressed. We expect, therefore, that the method described in this Chapter will eventually become useful in the analysis of the long-exposure Soudan data, although, as discussed in Chapter 8, the cut methods were sufficient for the analysis of the first WIMP-search run at Soudan.

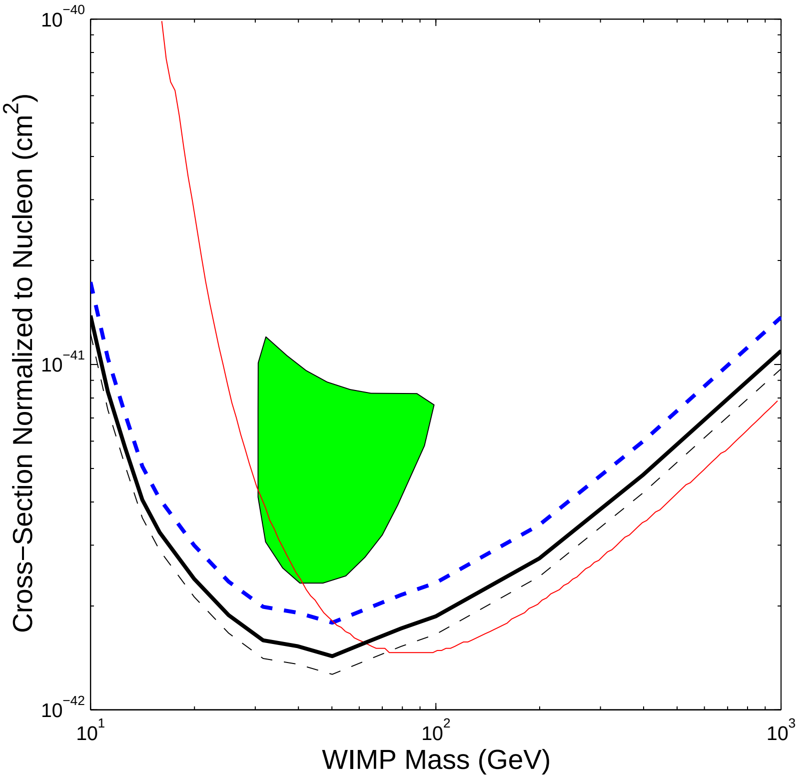


Figure 5.13: WIMP-nucleon interaction cross section vs WIMP mass. The thick dashed curve is the expected sensitivity calculated for the SUF Run21 -3V dataset using the rise-time cut method. The thick solid curve is the expected sensitivity for the same dataset using the maximum likelihood method. The thin dashed curve would be the expected sensitivity for the same dataset, with the rise-time cut, if the exposure were doubled. For reference, we also include the DAMA (1-4) 3σ allowed region (shaded; [106, 110]) and the Edelweiss limit (thin solid curve; [105]).

Chapter 6

Beta Calibration

6.1 Introduction

As discussed in the last two Chapters, the surface electron-recoil events can have incomplete charge collection. In fact, their charge signal can be so poor, that these events can look like nuclear-recoil events. Hence, the surface electron-recoil events are one type of background for the CDMS II experiment. Moreover, we expect that this will be the dominant background at Soudan, where the neutron background (dominant at the shallow site at SUF) will be strongly suppressed.

The last Chapter discussed statistical methods developed to more optimally use the information in the ionization yield and phonon rise-time parameters. In this Chapter we approach the problem of surface events in a different way. In particular, we will try to determine how to optimally extract the interaction-depth information from the detectors. For this purpose, we have performed a detailed calibration of a Ge detector with a radioactive source of betas. Since betas cannot penetrate deeply into Ge, the electron-recoil events caused by betas will be dominantly surface events. Hence, this calibration is essentially geared to study the surface events. In particular, we will try to determine the optimal charge bias, study different phonon timing parameters, study differences in performance of the two sides of the detector etc.

The rest of this Chapter is structured as follows. In Section 2, we describe the experimental setup, the calibration run itself, and the data-sets taken. In Section 3, we describe some aspects of the analysis performed on these data-sets, such as cuts, corrections etc. In Section 4, we discuss some preliminary results, such as electronic noise performance, position reconstruction, electron and nuclear-recoil band calculation etc. In Section 5, we discuss the energy calibration and energy spectra. In Section 6, discuss different phonon timing parameters (start-time, rise-time) and make a rough comparison of them in terms of beta discrimination. In Section 7, we calculate the surface events rejection efficiency and we compare the performance of the detector for different values of the charge bias and for different discrimination parameters. We conclude the Chapter with Section 8, which describes discrimination power in yet another parameter, based on the phonon amplitudes.

6.2 UCB Beta Calibration Run

6.2.1 Experimental Setup

The beta calibration was performed at the UC Berkeley CDMS II test facility. This facility hosts a system very similar to the full CDMS II setup at SUF (or Soudan), with a few differences.

The facility hosts an Oxford 75 μW cryostat, which can achieve base temperature of ~ 30 mK, sufficiently low to run the ZIP detectors. The cryostat was modified to host a single tower of six ZIP detectors. However, only four striplines are installed, allowing only three detectors to be read out in a single run (one stripline is used for the thermometry read-out).

The striplines take the signals out to the room temperature connectors, from where the detector I/O cables take them to the front-end board (FEB). In the beta calibration, we used a "modified-modified version 1" FEB (see Section 3.5 for more detail on the FEBs). Although this particular FEB was somewhat outdated (compared to version 3 FEBs used at Soudan), many of the deficiencies typical for version 1 FEBs were removed manually by Dennis Seitz. Hence, this board was fully functional and stable, and it provided good noise performance at the time of the calibration run. The cryostat and the FEB are protected from the ambient EM radiation by a Faraday cage, although it was demonstrated that the Faraday cage is not of crucial importance for the electronic noise performance.

The amplified signals are then taken from the FEB outside of the Faraday cage, to the Receiver-Trigger-Filter (RTF) board, which further filters the signals and generates the trigger edges (see Section 3.5 for more detail). The experiment was set to trigger on the sum of the four phonon signals. Also, both low and high trigger thresholds were used. The low trigger thresholds were set very low, to trigger on the smallest possible pulses. The high trigger thresholds (when used) were set to trigger on large pulses (usually around 200 keV). In the beta calibration run, we used a trigger logic board (built at UC Santa Barbara, and borrowed from the SUF setup) to combine the phonon low and high triggers and create the "gated" triggers. Gated triggering is very useful if one is trying to ignore the very high energy pulses (i.e. if one wants to trigger only on events in an energy window).

The trigger logic board, therefore, creates a global trigger, based on the low and high phonon triggers. The global trigger signal is used to start the digitizers recording. At the UCB test facility, Joerger digitizers are used, allowing up to 10 MHz sampling speed. We run the digitizers at 1.25 MHz, which is sufficient for tracking the rising edge of the phonon signals. The system includes 8 Joerger channels, allowing digitization of all six channels of a single detector. However, multiple detectors cannot be digitized in the present setup.

The FEB, RTF board, trigger logic board, and the digitizers are controlled by a Macintosh computer running Labview software. The software includes dedicated Labview vi's, designed to control the boards via GPIB commands. These vi's allow the user to study each channel in detail, and to optimize its settings (e.g. by varying the SQUID bias, QET bias, trigger thresholds etc). The software also includes the DAQ Labview package, which performs complete control of the system during the data-acquisition periods. This includes biasing detectors, flashing LEDs for neutralization purposes, setting trigger thresholds, recording digitizer traces, and piping the data over the local network to a Linux-based PC. This Labview DAQ package was used in the previous runs at SUF ([182]). However, a completely new, java based, DAQ package was written for the Soudan setup, where a larger volume of data is expected to be acquired (from a larger number of detectors).

As mentioned above, the data is finally recorded on a Linux-based machine. The dataflow flow in the beta calibration was much larger than what the test facility was originally designed for - usually, for detector characterization purposes, a single dataset with $\sim 50,000$ events is sufficient. Instead, the beta calibration was meant to include many

$\sim 150,000$-event datasets. Hence, a new dual-processor Linux machine was installed, along with a 120 GB hard-drive, which more than tripled the processing speed. Along with the existing machine (totalling 54 GB of hard-drives) the two machines were sufficient to handle the data flow.

The data analysis package applied to the beta calibration data-sets was the standard, Matlab based, CDMS II analysis package (known as the DarkPipe). This package is described in more detail in Chapter 8 and in [2, 182] - it includes several algorithms returning a number of "reduced quantities" (RQs):

- Optimal Filter: applied to both the charge and phonon pulses, it returns the pulse amplitude and start-time (see [2] for more detail).

- F5 Algorithm: Fits charge pulses with a 5-parameter fit, and returns the amplitude. Useful for large pulses, which exceed the digitizer range.

- RtFt Walk Algorithm: applied only to phonon pulses, finds the maximum of a pulse, and then "walks" down the rising and falling edges to determine the times of 10%, 20% etc of the pulse amplitude, interpolating between points if necessary. The pulses are smoothed out prior to the measurements, to minimize algorithm failures.

The RQs discussed above are not necessarily in "physical" units, such as keV of energy. Hence, new, "reduced-reduced quantities" in "physical" units are calculated from these RQs, using another Matlab program (known as the pipeCleaner).

6.2.2 Calibration Runs

Since betas cannot penetrate through the walls of the cryostat, the source of betas had to be placed inside the cryostat. For this purpose, we have purchased a large surface ($\sim$ 3 inch diameter) ^{109}Cd source. ^{109}Cd decays into ^{109}Ag$_{\rm m}$ by electron capture, which is 88 keV above the ground state for ^{109}Ag. The emitted 88 keV gamma is strongly internally converted into electrons and accompanying X-rays. The yield of unconverted 88 keV gammas is only 3.6%. The rest of the gammas lead to 63 keV (41%), 84 keV (45%) and 87 keV (9%) electron lines. The highest X-rays are at 22 keV and 25 keV; there are also lines below 3 keV, which will not be visible to us.

We built a dedicated copper holder to host the source, such that it can be mounted on a Tower along with detectors. With such, modular, design of the source holder, compatible with the design of the Tower, we could expose either surface of the detector to the betas. In fact, our calibration run actually consisted of two runs (UCBRun296 and UCBRun297), in which one side of the Ge detector G31 was exposed to the betas (charge and phonon side respectively), as shown in the Figure 6.1.

The source holder also included a lead collimator, which allowed the betas (and gammas) to pass only through several collimator holes. The collimator was necessary both in order to keep the event rate reasonably low, and to allow us to check if our algorithms for position reconstruction and position corrections are working well. The geometry of the lead collimator is shown in the Figure 6.2. In each of the four phonon quadrants (denoted by A, B, C, and D in the counter-clockwise direction, as depicted in the Figure), there were three collimator holes, the outer-most one being so far that it faced the outer charge electrode.

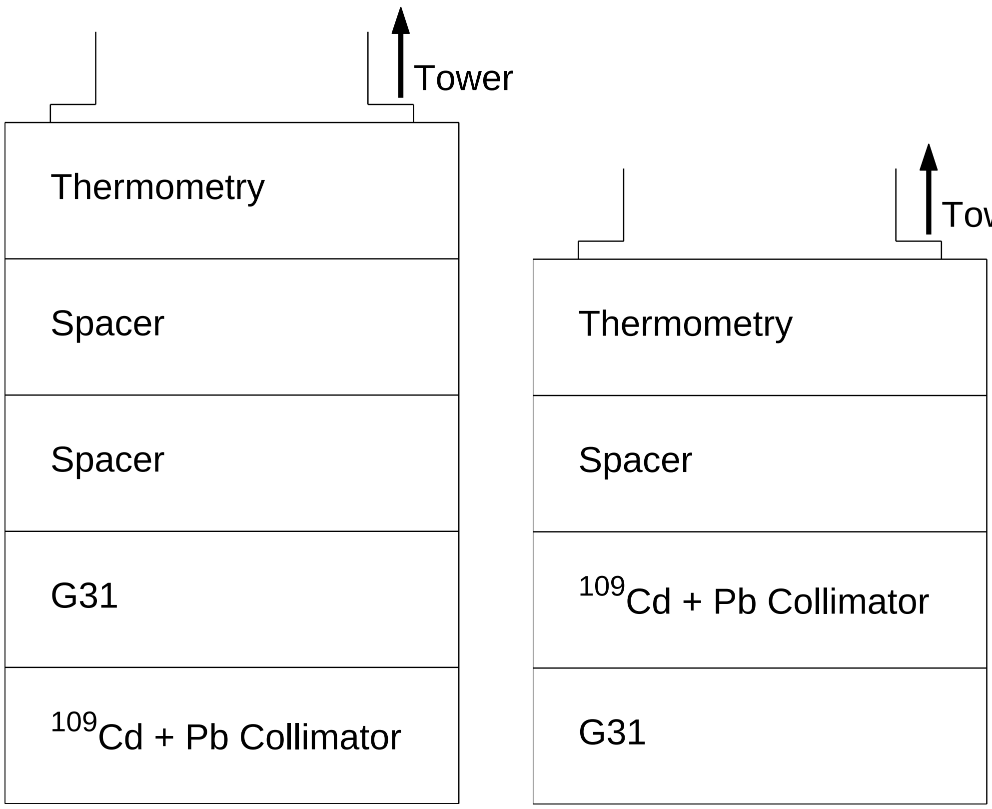


Figure 6.1: Schematic of the Tower configuration for UCBRun296 (left) and UCBRun297 (right).

Since events under the outer charge electrode are removed with the fiducial volume cut, only two holes per sensor were useful in the analysis. Two of the inner-most holes were covered with Al foil (A and D in UCBRun296 and B and C in UCBRun297). The foil allowed only gammas from the Cd spectrum to pass, while all the betas were stopped. The events in these positions are, therefore, clean gammas that can be used to, for example, study the depth profile of the ionization signal.

Exposing both sides of the detector to the betas allows us to address the issues of possible asymmetries between the two sides of the detector. However, the approach also introduces possible systematic errors due to possibly different conditions in the two runs

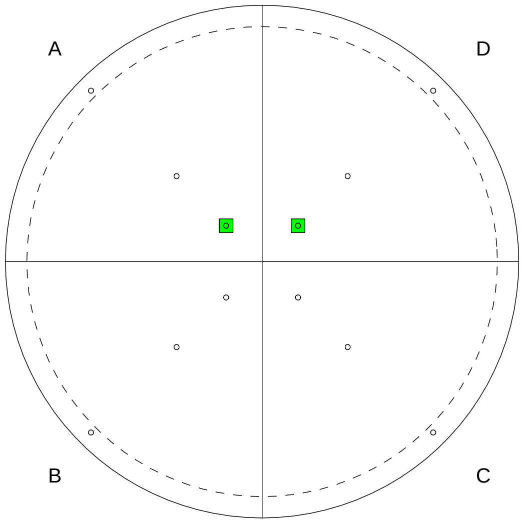


Figure 6.2: Schematic of the lead collimator used in the UCBRun296 (Cd facing charge side of the detector). The dashed line denotes where the two charge electrodes are separated. The shaded regions denote the two holes covered with Al foil (inner-most holes in quadrants A and D). In UCBRun297 (Cd facing phonon side of the detector), inner-most holes in quadrants B and C were covered.

(e.g. base temperature, various bias values etc).

Furthermore, for each of the two runs, we acquired data at four different charge bias values: ±3V and ±6 V. The goal is to determine which charge bias is optimal in terms of the beta rejection, at either side of the detector. Note that 6 V bias will clearly allow better charge collection for the surface events, hence improving the rejection in the yield parameter, but it would also increase the Luke component of the phonon spectrum, hence washing out the fast component of the initial phonon spectrum (which is believed to give shorter rise-time to the surface events). Hence, these are opposing effects, and it is not clear a priori which charge bias would be better in terms of beta rejection.

Finally, to be able to perform the analysis fully, for each run and each charge bias voltage, we took three different datasets:

- 1) Dataset with ^{109}Cd internal source only.

- 2) Dataset with ^{109}Cd internal source and ^{60}Co external source. This Co source was placed outside the cryostat for this particular dataset only. Its gamma spectrum extends to high energies (MeV scale), so it was used to obtain a uniform gamma exposure of the detector. Such calibration, with a large number of bulk gamma events, is needed for calculating correction tables, used for position-correcting various parameters (discussed in detail in the following section).

- 3) Dataset with ^{109}Cd internal source and ^{252}Cf external neutron source. This source was also placed outside of the cryostat for this dataset only, and it provided the neutron-caused nuclear-recoil (NR) events, needed for calculating the position of the NR band (and, therefore, for estimating the beta leakage into the NR band).

Note, also, that due to significant presence of lead inside the cryostat, we could also observe the lead scintillation lines. These appear at 10.5 keV (12.8%), 72.82 keV (27.8%), 74.97 keV (46.8%), and 84.9 keV (10.7%).

6.3 Position Correction and Cuts Definitions

6.3.1 Position Correction

As mentioned in Chapter 4, the phonon signal in the ZIP detectors tends to be non-uniform across the detector surface. There are essentially two reasons for this. First, the phonon signal for an event close to the center of the crystal is different from the phonon signal for a similar event close to the edge of the crystal (simply because of the phonon reflections effects, different coverage of the surface by the TES's etc). Second, the critical temperatures of the many TES's on the detector surface are not uniform (the variations can be as large as 20-30 mK). Such non-uniformities cause variations in the phonon signal, both in amplitude and in the rise-time/start-time. Hence, position correcting the phonon signal is expected to improve both the energy and the rise-time/start-time resolutions.

As mentioned above, for each of the 8 cases (two sides times four charge bias values) we took a dedicated ^{60}Co calibration dataset. The (predominantly bulk) gamma events from this dataset (and from the other two datasets of the same case) were used to construct the position correction tables. The technique we deploy was developed by Blas Cabrera and Clarence Chang at Stanford University. It can be summarized in the following steps:

1. Select bulk electron-recoil events, usually using the non-corrected ionization yield. Events of all energies can be used, as most of the low recoil energy events are expected to be Compton scatters (i.e. caused by higher energy gammas, which penetrate sufficiently deep into the crystal).

2. Define the x-y position of an event using three position parameters: two phonon partition parameters p_x and p_y, and the radius of the delay x-y plot, r_d (see Section 4.4 for more details on these parameters). The necessity for using 3 parameters for estimating 2D position comes from the fact that both the phonon partition plot ("box" plot) and the delay plot become non-linear towards the crystalline edge, or even have a "fold-over" (i.e. events that happen very close to the crystalline edge appear closer to the center in the box/delay plot, than events that actually take place closer to

the center). Hence, the third parameter is needed to break this degeneracy. This non-linearity is shown in the Figure 6.3.

3. For each of the selected events, determine the nearest N neighbors in the $p_x - p_y - r_d$ space described above (N is usually chosen to be 20). Average the parameter in question (e.g. ionization yield, or phonon rise-time) over these neighbors. Enter the three coordinates of the event and the averaged values into a "table". This completes the calculation of the correction table.

4. When applying the table to a dataset, for each event of the dataset find the table entry the closest to the event (in the $p_x - p_y - r_d$ space described above). Then correct the phonon amplitude and rise-time parameter(s) of the event according to the "average values" of the chosen table entry, as determined in the previous step.

We position-correct three phonon timing parameters, with the hope of determining which one (or a combination of them) gives the best rejection of the surface events:

- ptrt: $10\% - 40\%$ rise-time for the sum of all four phonon pulses. This is the standard parameter used in the SUF Run21 analysis (discussed in detail in the previous Chapter).

- pminrt: $10\% - 40\%$ rise-time of the largest phonon pulse (usually corresponding to the shortest rise-time of the four phonon sensors).

- pmindel: delay (with respect to the charge pulse) of the largest phonon pulse. The start time of the phonon pulse was taken to be at 20% of leading edge.

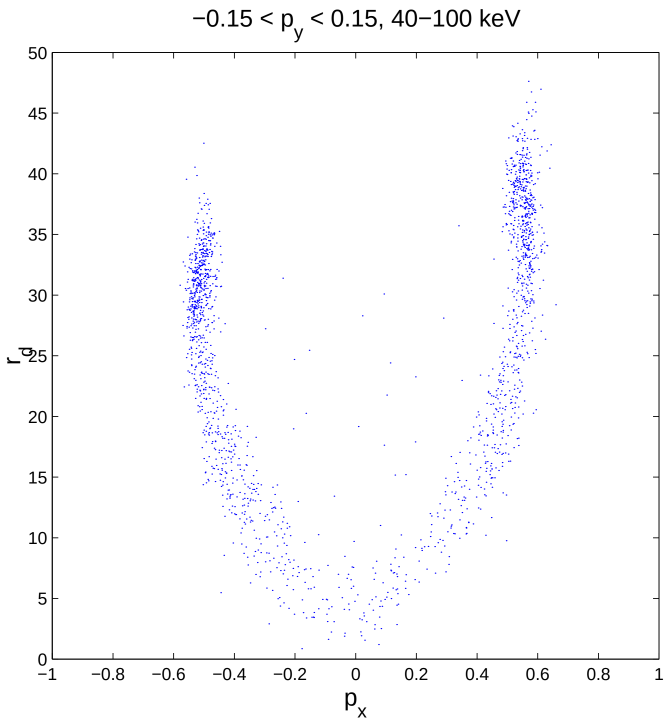


Figure 6.3: A horizontal slice in the box plot, plotted in the $r_d - p_x$ plane indicates the non-linearity in the event position reconstruction in the box plot. The case for UCBRun297 -3V, 40-100 keV is shown.

We emphasize again that the position correction is meant to remove the non-uniformities only across the detector surface. It must not affect the depth information in our parameters! Hence, it is a priority to use bulk events only, which do not have dependence on depth in either yield or rise-time parameters.

It is, therefore, very important to avoid the surface events when calculating the correction table, as they can significantly distort the table. This is particularly important for the rise-time parameters, as they are sensitive to depths of order 1 mm (while yield is sensitive to only a few μm depths). Moreover, 22 keV gammas from the ^{109}Cd source will also behave as surface events in rise-time - 22 keV is here the incident (not the recoil) energy! To avoid these events, we use the fact that they are localized - they all fall in the blobs in delay and box plots corresponding to the positions of the collimator holes - see Figures 6.5 and 6.6. In particular, work of Bruno Serfass [195] has shown that it is sufficient to remove the events below 90 keV in recoil energy in these blobs - this removes all betas (^{109}Cd beta spectrum ends below 90 keV) and the 22 keV gammas. Events above 90 keV in the blobs, and events of all energies outside the blobs were used for the correction table calculation.

The results of applying the correction will be described in the following Section.

6.3.2 Cuts

The following cuts are applied to all data.

- Outer electrode cut: removes events with significant energy deposited in the outer electrode. It requires that the charge partition: $qpart = (inner - outer)/(inner + outer) > 0.75$. We have verified that varying the cutoff value does not introduce low-yield events, implying that the fiducial volume defined by this cut is sufficiently far from the edge effects. The efficiency of the cut is $\sim 85\%$.

- Charge saturation cut: requires that the digitizers do not rail (i.e. that the whole pulse is within the digitizer range).

- Charge χ^2 cut: requires that the χ^2 of the charge optimal filter fit is well behaved. This cut removes majority of the pile-up events (i.e. situations in which more than one event happens in a single digitized 2 ms trace).

- Standard deviation of the pre-trigger: requires that the pre-trigger part of each phonon and charge pulse is well behaved. In particular, the standard deviation of the pre-trigger part of the pulse is required to be within 5 σ from the estimated mean. This cut also removes pile-up events.

- Bad data cut: removes some datasets with known problems. In particular, some datasets had slightly distorted box-plots, causing the position correction to be unsuccessful. The cause of these distortions is unknown, and is probably related to the room-temperature electronics. Unfortunately, for UCBRun296 -6V and UCBRun297 +6V, the problematic datasets were neutron calibrations, so the estimates of the neutron bands were unreliable. Hence, these two cases will not be examined in detail in the further analysis.

- Bad correction cut: for UCBRun296 -3 V and +3 V, we have observed that the correction was not successful in the lower part of the detector (bottom of sensors B and C). Careful checks did not reveal any obvious problem with the correction. We believe, however, that the three position parameters were simply not able to reconstruct event position reliably in this region. More fundamentally, the problem may have been in the bias points of the phonon sensors for this run.

We also mention here that the phonon channel A in the UCBRun296 was not optimally biased. For this reason, the pulse shape parameters for channel A were significantly different from the other three channels. However, the position correction removed most of this discrepancy, so we chose not to impose any cuts on the channel A events.

6.4 Detector Performance

In this Section we summarize some preliminary results from the beta calibration.

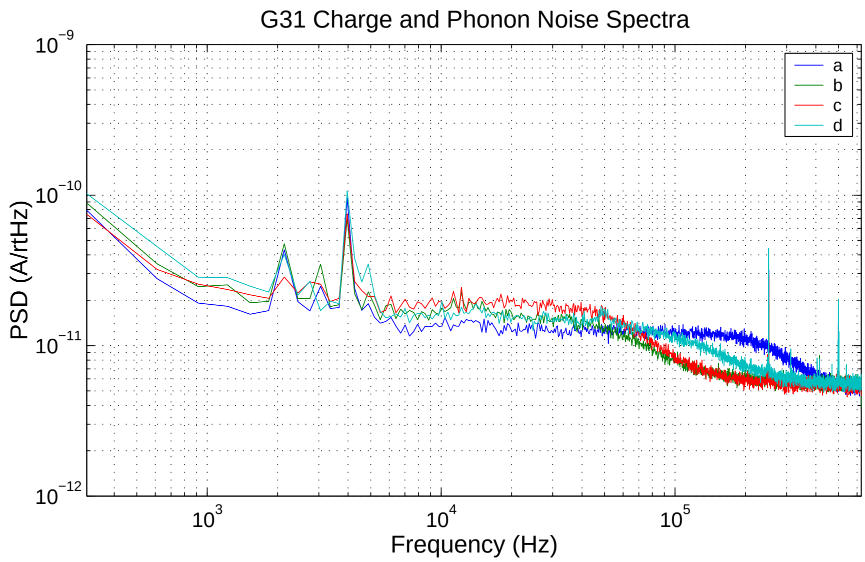
6.4.1 Noise Performance

The electronic noise performance is shown in Figure 6.4. The discrete peaks in the spectra are believed to come from the room temperature electronics, and they do not affect the signal-to-noise significantly.

6.4.2 Position Reconstruction

As discussed in Chapter 4, we can reconstruct the event position using either the relative timing of the phonon channels, or the relative amplitudes - these are the standard delay and box plots. Figures 6.5 and 6.6 show these plots for the cases when charge bias was -3 V and the ^{109}Cd source was facing the charge side (UCBRun296) and the phonon side (UCBRun297) respectively. Events with 15-40 keV recoil energy were selected, which includes both the 22 keV gamma line from ^{109}Cd and the beta spectrum at these energies (we will discuss the energy spectra more later). The four phonon sensors were scaled relative to each other in order to symmetrize the plots.

The large (red) blobs that can be observed in all of these plots correspond to the collimator holes (see Figure 6.2). Hence, the positions of holes can be reconstructed fairly well - two holes per quadrant, along the diagonal. Note that the third (outer-most) collimator hole in each quadrant was above the outer (guard) electrode, so the events corresponding to these holes were removed with the fiducial volume cut.



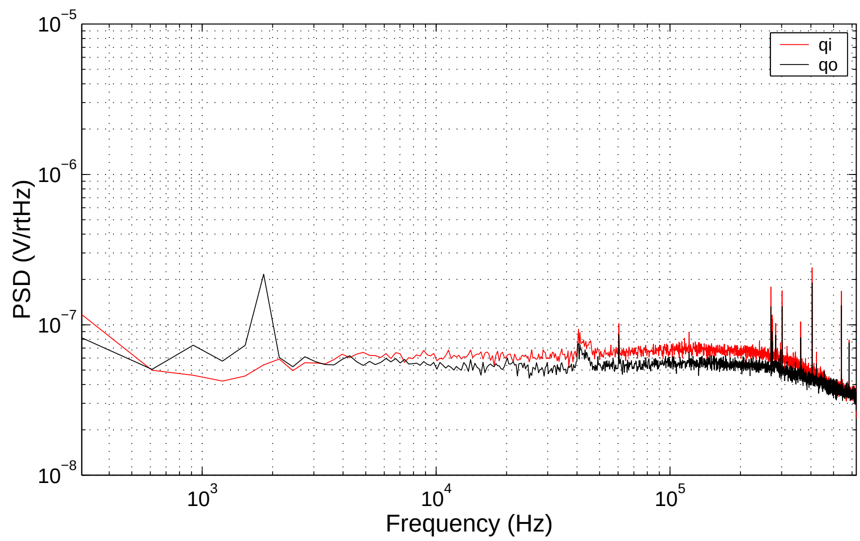


Figure 6.4: Electronic noise for the G31 phonon and charge channels. For phonon channels, the noise is referred to the input side of the SQUID, and for charge channels it is referred to the feedback capacitor (after the first stage amplification). The peaks at 4 kHz in the phonon spectra are usual in the UCB test facility setup - they do not affect the signal-to-noise significantly.

6.4.3 Position Correction Performance

As discussed in the previous Section, we can use the position information stored in the box and delay plots to correct for any non-uniformities in the detector response across

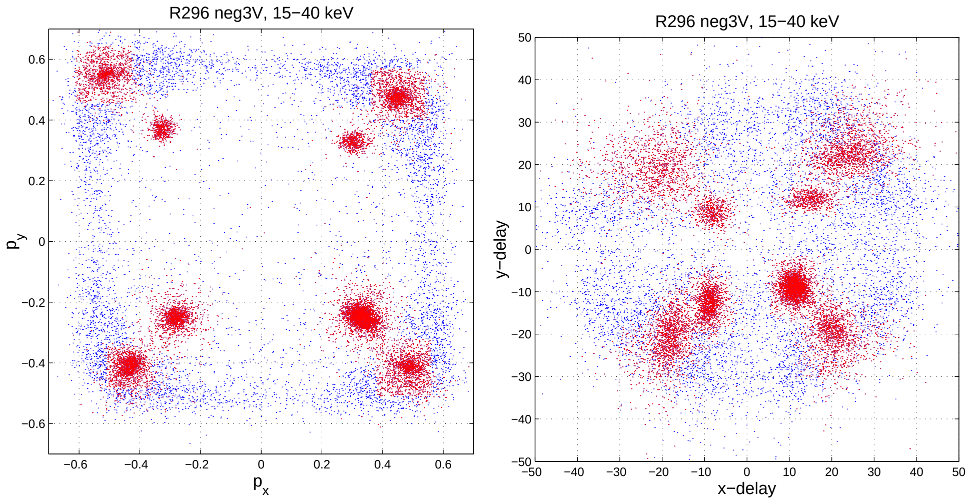


Figure 6.5: The phonon partition (box) plot (left) and delay plot (right) for the UCBRun296 (Cd facing charge side of G31) at -3 V charge bias.

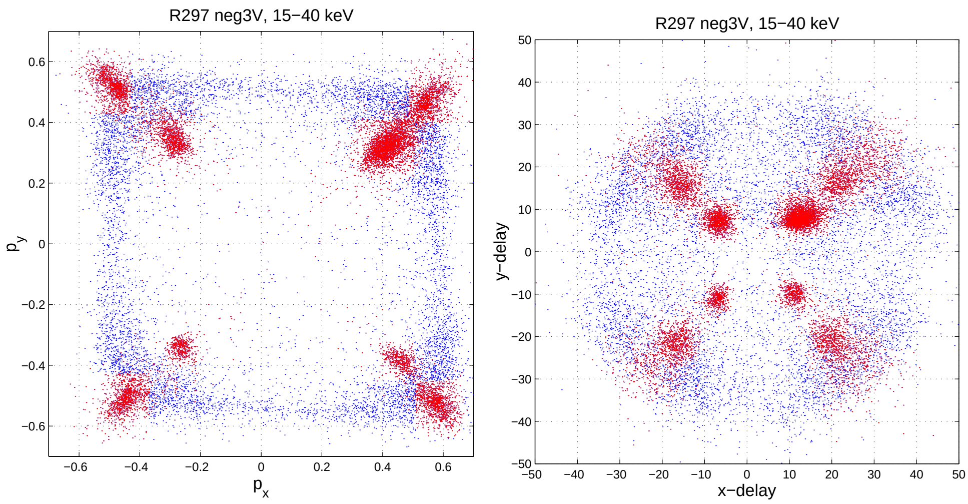


Figure 6.6: The phonon partition (box) plot (left) and delay plot (right) for the UCBRun297 (Cd facing phonon side of G31) at -3 V charge bias.

the detector surface. This includes the non-uniformities in both phonon amplitudes and in phonon pulse shape (i.e. rise-time/start-time). Here, we briefly discuss how well the position correction algorithm performed.

Figure 6.7 depicts the improvement (i.e. tightening) of the electron-recoil band when the phonon signal amplitude is position corrected. For this Figure, we excluded events in the Cd blobs to suppress the surface events (although some still sneak in). The position correction is particularly important for the UCBRun296, where the phonon sensors were not biased optimally.

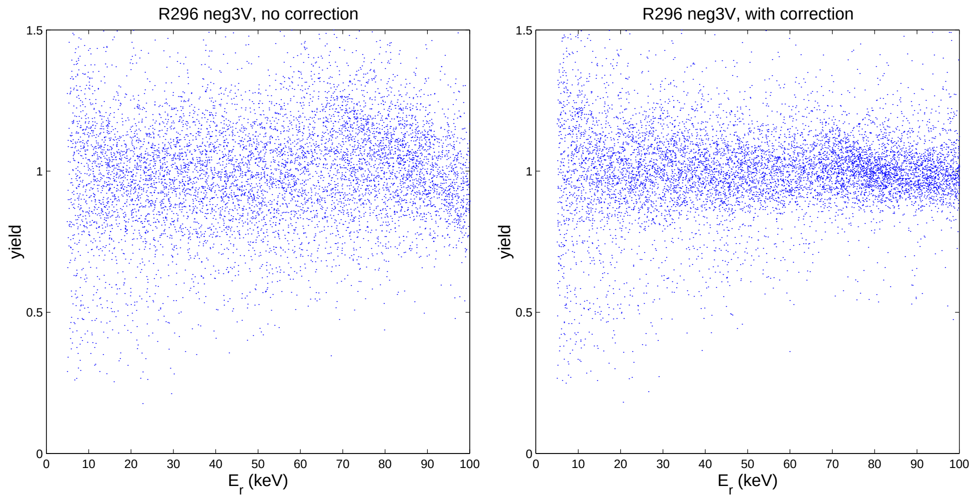


Figure 6.7: The yield-recoil energy plots before (left) and after (right) position correction of the phonon signal amplitude, for UCBRun296 (Cd facing charge side of G31), and charge bias of -3 V.

A more careful test of the position correction performance is to look at what happens to the distributions of the corrected parameters in the Cd blobs. Since blobs are localized, one expects the means of the distributions in the various blobs to be brought closer together by the correction, but the widths of the distributions in individual blobs should not be affected by the position correction.

Figure 6.8 examines, as an example, the phonon rise-time parameter ptrt in the 50-100 keV energy bin (other parameters and energy bins behave similarly). Before the position correction, ptrt distributions in different blobs are centered around very different values (sensor A is the most off, due to its non-optimal bias point). The position correction brings the distributions much closer together. The corrected distributions do not align perfectly, implying that the correction removes most but not all of the original position dependence.

Figure 6.9 verifies that the widths of distributions in individual blobs are not affected by the correction. Examples of the inner Cd blob in sensor B and outer blob in D are shown (other blobs and parameters have also been verified). After manually scaling the distributions to match the means, it is clear that the widths of the distributions are not affected by the position correction.

As mentioned in the previous Section, the only place where the position correction failed was the bottom part of sensors B and C in UCBRun296 $\pm$ 3 V. We believe this was due to poor position reconstruction in this region, and we have excluded this region from the analysis.

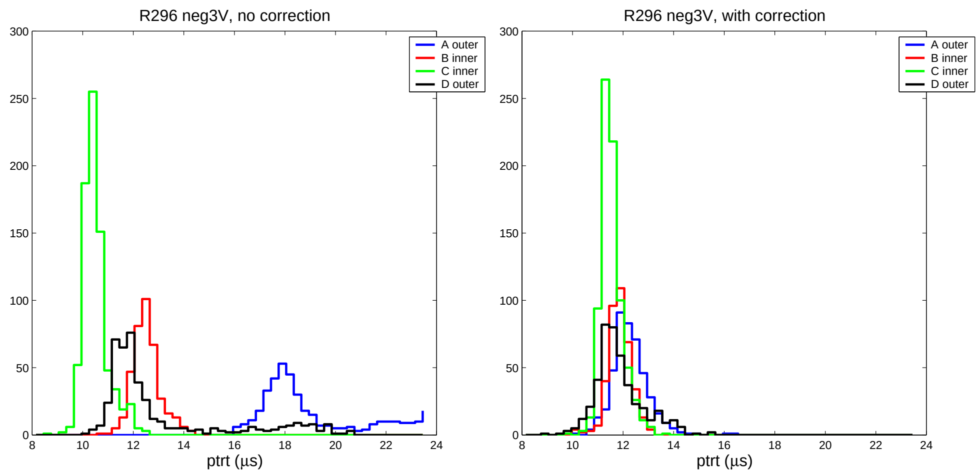


Figure 6.8: The distributions of the phonon rise-time parameter ptrt in four of the Cd blobs before (left) and after (right) the position correction. Data is shown for the UCBRun296 -3 V case.

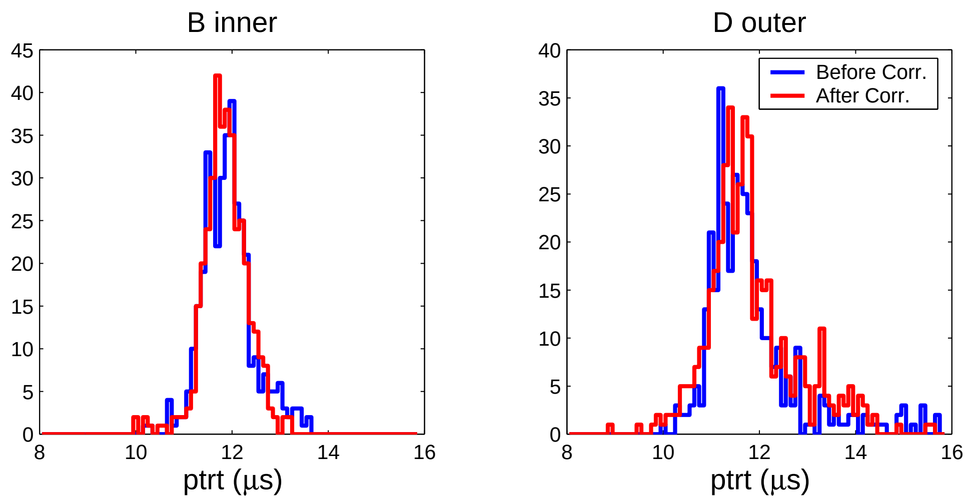


Figure 6.9: The distributions of the phonon rise-time parameter ptrt before and after the position correction, for the inner Cd blob in sensor B (left) and the outer blob in sensor D (right). The distributions were manually scaled to match the means. Data is shown for the UCBRun296 -3 V case.

6.5 Energy Calibration and Spectra

The energy calibration is done using the 22 keV gamma line from the ^{109}Cd source. We first calibrate the charge signal against this line. To be consistent with the SUF/Soudan energy calibrations, the phonon signal is calibrated against the charge signal, rather than against the 22 keV line, by requiring that the ionization yield for the bulk electron-recoil events is centered around 1. This is because at SUF/Soudan the calibration lines are typically at high energies ($\sim$ 350 keV or higher), at which the phonon sensors become saturated and have non-linear response (so only charge channels can be calibrated).

Figure 6.10 shows examples of the charge spectra for events in the Cd blobs. The 22 keV line is clear and easy to calibrate against. The spectra also show hints of the 88 keV gamma line of ^{109}Cd, but they are too small for reliable calibration. The lines at $\sim$ 10 keV and $\sim$ 75 keV are believed to be the lead scintillation lines. Note that the beta peaks, expected at 63 keV and 84 keV, are not visible at all, because the charge signal for these surface events is severely damaged.

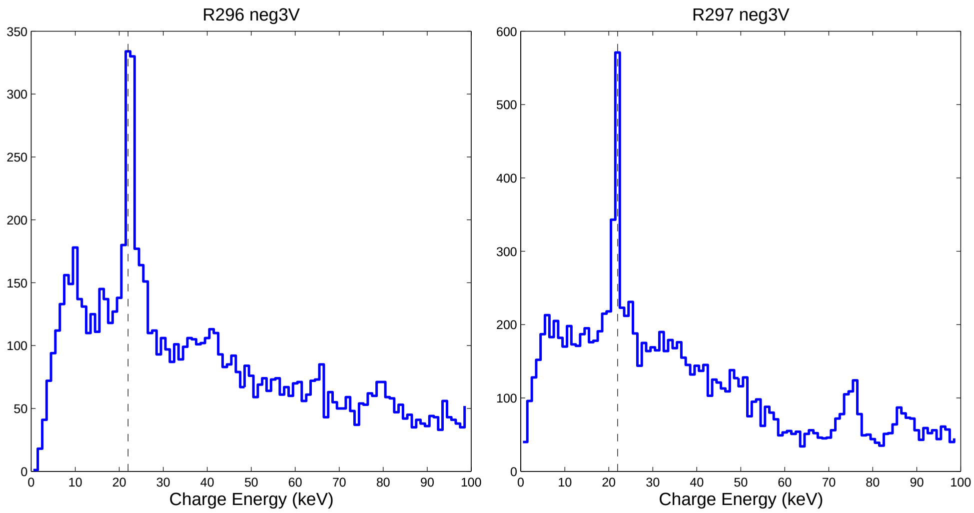


Figure 6.10: Charge energy spectra for UCBRun296 (left) and UCBRun297 (right) at -3 V charge bias. Only events from Cd blobs were used. The dashed line denotes the 22 keV line used for calibration.

Figure 6.11 shows the corresponding recoil energy spectra - the expected beta peaks are very clear. The energy resolution is typically ~ 1 keV for the charge, ~ 1.5 keV for the phonons, and $\sim 2 - 2.5$ keV for the recoil (at 22 keV). These values vary a bit from case to case, mostly depending on how well the position correction worked. They are also similar to the resolution estimates from the low background run at SUF (SUF Run21, described in the previous Chapter and in [182]).

Finally, for completeness we show the Pb scintillation lines in the Figure 6.12, for the case of UCBRun297 -3V. We exclude the ^{109}Cd blobs and plot only the events in the ER band to extract these lines more clearly (otherwise they would be mixed with the electron peaks). These lines are not as clear in all cases, possibly because of slightly different operating conditions from case to case (such as orientation of the source with respect to Pb

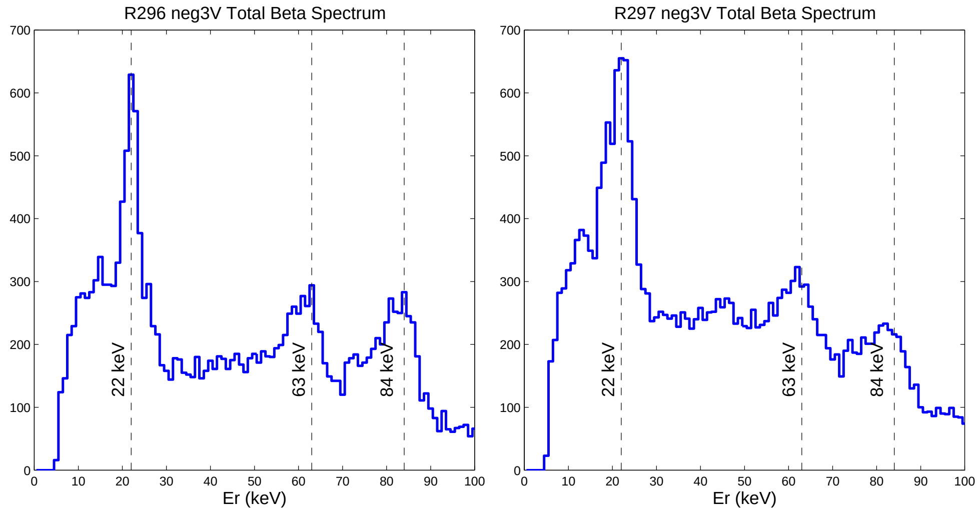


Figure 6.11: Recoil energy spectra for UCBRun296 (left) and UCBRun297 (right) at -3 V charge bias. Only events from Cd blobs were used. The dashed lines denote the 22 keV gamma line and the 63 keV and 84 keV beta peaks.

etc).

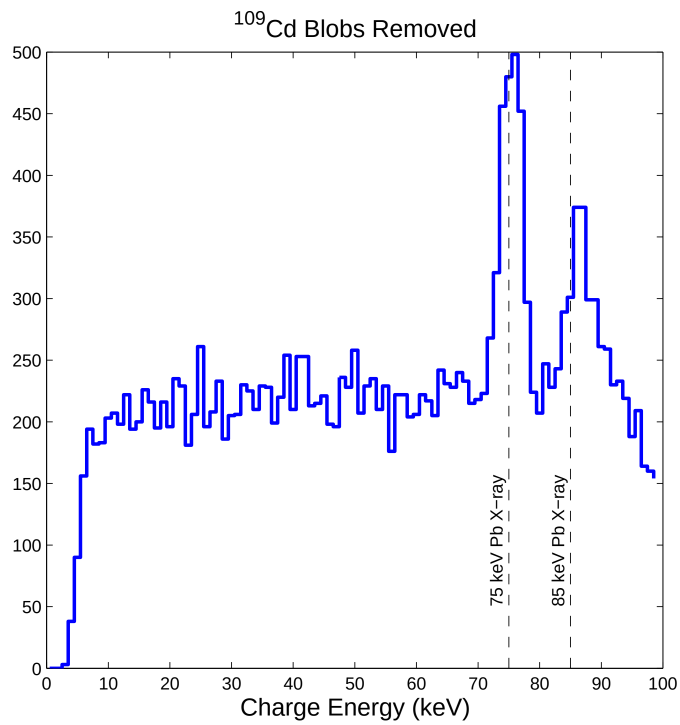


Figure 6.12: The Pb scintillation lines for the case of UCBRun297 -3V. The ^{109}Cd blobs were removed.

6.5.1 Electron and Nuclear-Recoil Bands

Having calibrated the charge and the phonon channels, we are ready to calculate the electron and nuclear-recoil bands. The bands are calculated following the procedure in Section 5.2. For calculating the ER band, we selected events outside the Cd blobs to avoid the beta contamination. Similarly, for calculating the NR band, we used the Cd+Cf datasets, selecting events outside the Cd blobs. The bands and their relation to the data are shown in the Figure 6.13. A more careful comparison of the various bands is given in the Figure 6.14.

Several comments and observations can be made based on these Figures:

- All ER bands have centroids properly centered around 1.
- The widths of bands vary from case to case, but they are not correlated with the number of events used to calculate the correction table or to calculate the bands. Hence, these are not artifacts of the different statistics in different cases.
- The ER bands corresponding to positive charge bias polarity are somewhat wider than those corresponding to negative charge bias polarity.
- Similarly, the NR bands corresponding to the cases with positive charge bias polarity tend to be wider than those with the negative polarity. However, the centroids of NR bands for positive charge bias are somewhat lower than for negative charge bias, so that the upper 2 σ bounds are similar for all cases.
- There seems to be no significant difference between the 3V and the 6V cases in terms of band positions and widths.

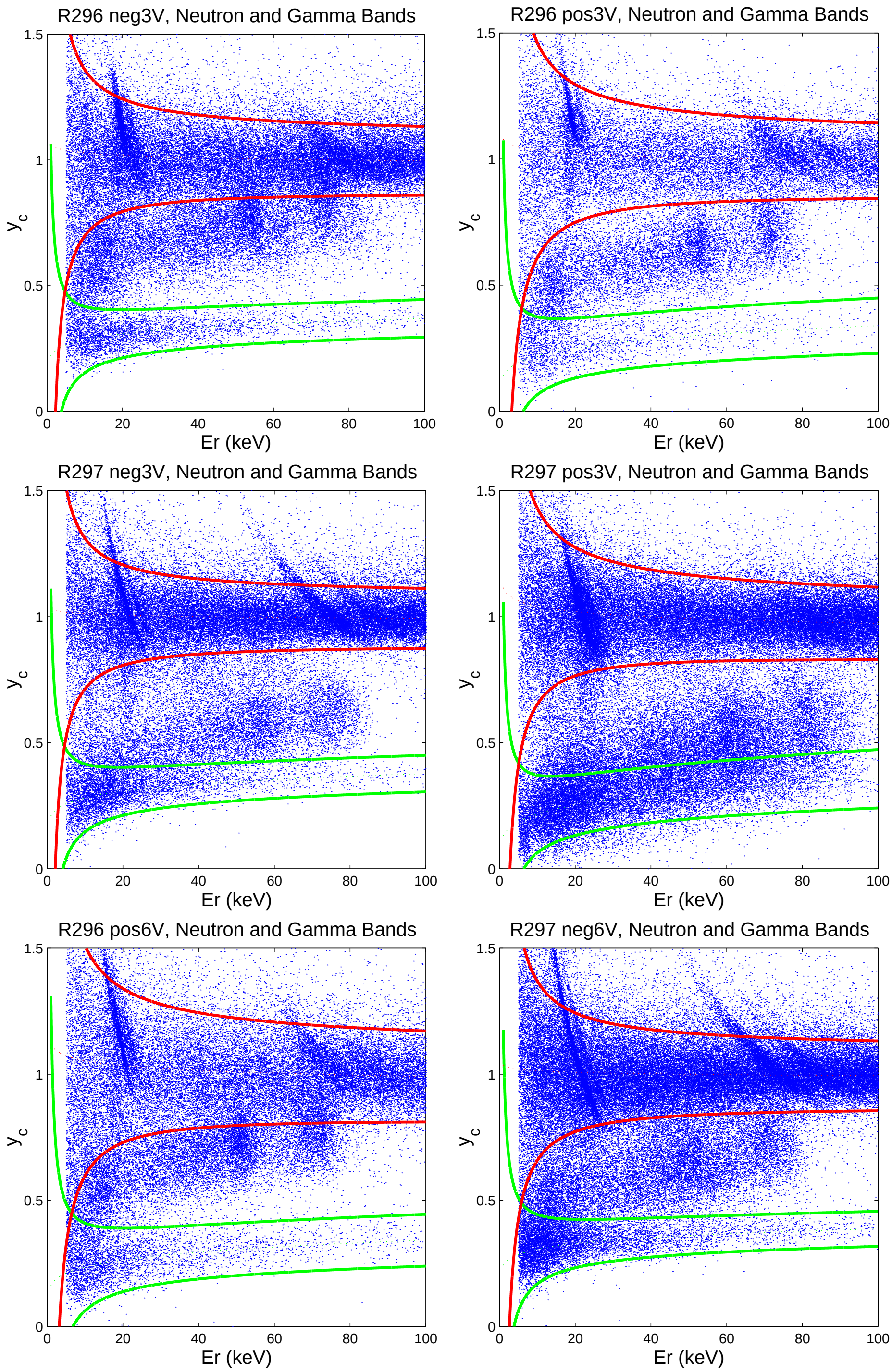


Figure 6.13: Electron and nuclear-recoil bands for the six cases.

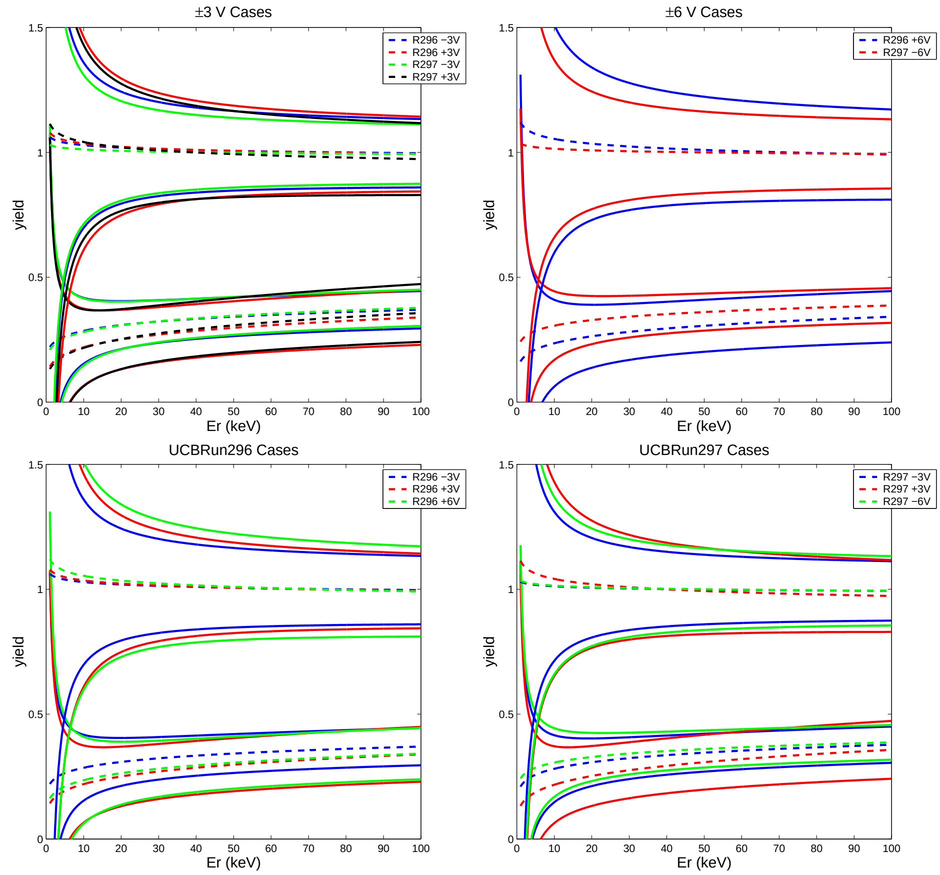


Figure 6.14: Comparisons of ER and NR bands

- There is no obvious difference between UCBRun296 and UCBRun297 in terms of the band positions and widths, indicating that the bulk of the crystal was about equally well neutralized in the two runs.

- The band of betas is much higher in yield for the UCBRun296 (Cd facing the charge side of G31) than for the UCBRun297 (Cd facing the phonon side of G31). This asymmetry appears to be real, but we do not yet understand its physical origin. One possibility is the asymmetry in the fabrication of the two sides of the detector - the Al and W layers on the two sides have different thicknesses and shapes. We have made extensive checks in both runs to verify that the detector was indeed neutralized, and that it's state did not degrade over time during data acquisition.

- Note that the numbers of events in the scatter plots are not the same.

Note that these plots are not conclusive in terms of beta leakage, since we did not use phonon timing information yet. Nevertheless, it is clear that the beta leakage into NR band is larger on the phonon side than on the charge side of the detector. As pointed out above, it is not clear what is causing such asymmetry, but there are asymmetries in the lithographic depositions on the two sides. Another interesting (although small) effect is that the positive charge bias tends to give wider bands. This hints that the processes of electron (and hole) relaxation on the charge electrodes on the two sides might not be identical - e.g. the phonons released during electron relaxation may be different depending on where the relaxation happens.

6.6 Phonon Timing Parameters

In this Section, we introduce three phonon timing parameters that we may use for rejecting the surface events. We briefly demonstrate their surface event rejection ability and show how they compare and relate to each other. In the following Section we will combine these parameters with the ionization yield, to calculate the overall surface-event rejection efficiency.

As mentioned above, the parameters we consider are:

- ptrt: $10\% - 40\%$ rise-time for the sum of all four phonon pulses.

- pminrt: 10% – 40% rise-time of the largest phonon pulse.

- pmindel: delay (with respect to the charge pulse) of the largest phonon pulse (the start time of the phonon pulse was taken to be at 20% of the leading edge).

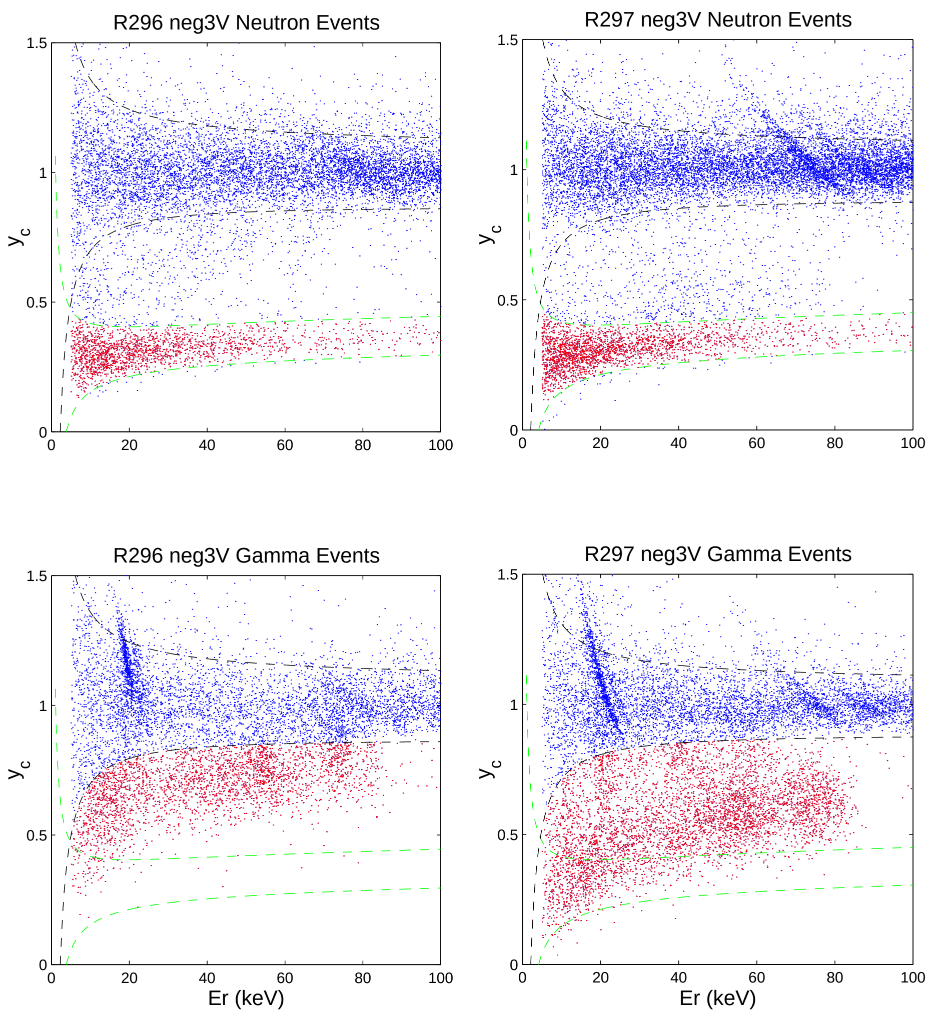


Figure 6.15: Definition of the neutron events (upper row, in red) and beta events (lower row, in red). Recall that the beta source was facing the charge side in the UCBRun296 and the phonon side in the UCBRun297. See text for more detail.

The neutron (i.e. nuclear-recoil) events are defined as events in the NR band from the Cd+Cf dataset, outside of the Cd blobs to avoid contamination by the surface electron-recoil events. The beta (i.e. surface electron-recoil) events are defined as the events in the Cd blobs of the Cd-alone and the Cd+Co datasets (i.e. no neutrons from Cd+Cf dataset), below $-2\ \sigma$ bound of the ER band. Figure 6.15 shows how neutron and beta events are chosen for the cases of UCBRun296 -3V and UCBRun297 -3V.

Next, we look at the time-yield plane, where time is one of the three phonon timing parameters. Since there is considerable energy dependence, we split our analysis in several energy bins: 5-15 keV, 15-30 keV, 30-50 keV and 50-100 keV. Figures 6.16 and 6.17 are examples of such planes for the cases UCBRun296 -3V and UCBRun297 -3V. Note that all of the phonon timing parameters have some rejection power. Note also, as demonstrated in the last column in these Figures, that these parameters are somewhat correlated. It is worthwhile examining if a combination of these parameters could improve the overall rejection efficiency.

The question of which of these parameters is "the best" is related to the question of how the final result (i.e. the limit or the allowed region in the WIMP-nucleon scattering cross-section vs WIMP-mass plane) will be calculated. There are at least two options.

Option 1: one can make a stringent cut in a phonon timing parameter, rejecting most of the surface electron-recoil events. One can then estimate the remaining background of these events and ignore it in the calculation of the limit, either because it is negligibly small or because it is dominated by another type of background (which was the case with the neutron background at the shallow site at SUF). In this case, one is essentially looking for a cut that would reject most of the betas, while keeping as large fraction of neutrons as possible.

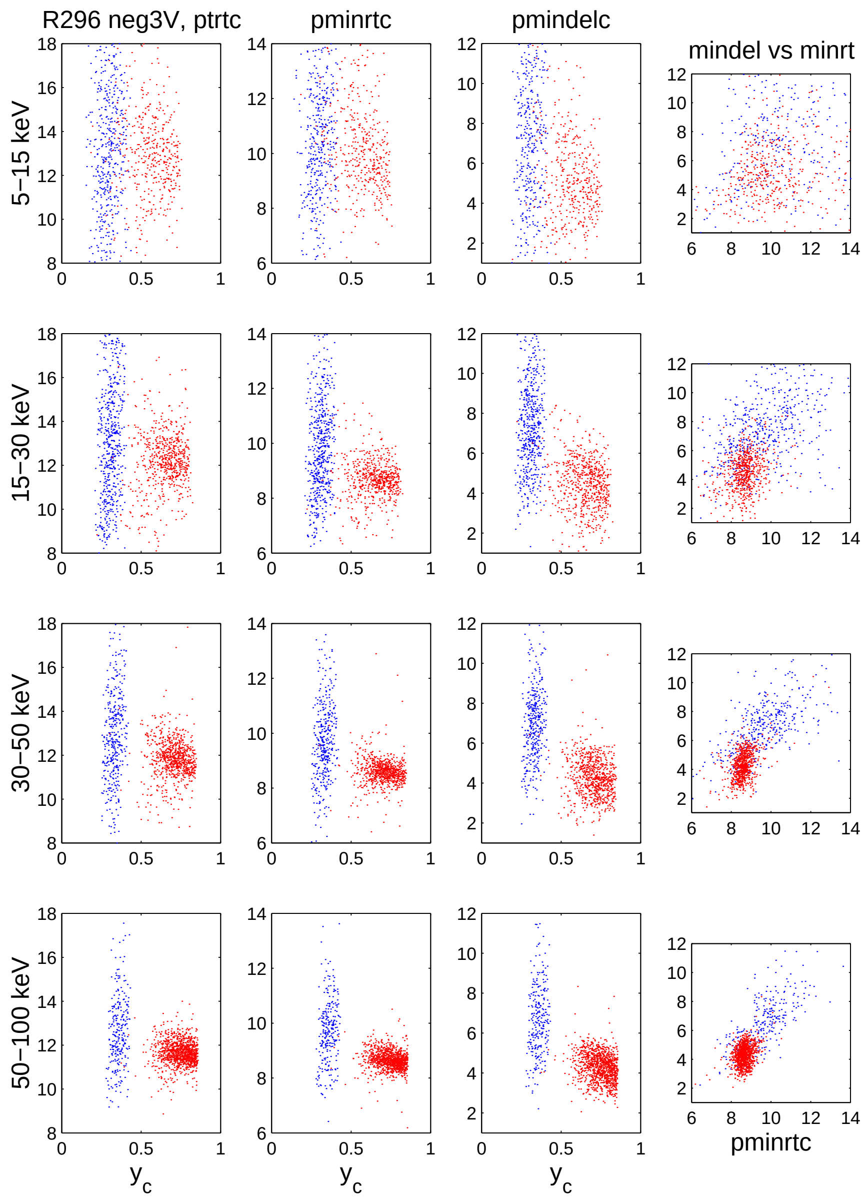


Figure 6.16: Time-yield plane for UCBRun296 -3V (beta source facing charge side), for the parameters ptrtc (column 1), pminrtc (column 2) and pmindelc (column 3). Column 4 demonstrates correlation of the parameters pminrtc and pmindelc. The four rows correspond to four energy bins. Neutrons (in blue) and betas (in red) are shown. Note that "c" in the parameter names denotes that the parameters have been position-corrected.

Option 2: one can try to estimate the background after the rise-time cut due to surface electron-recoil events and then statistically subtract it from the overall population

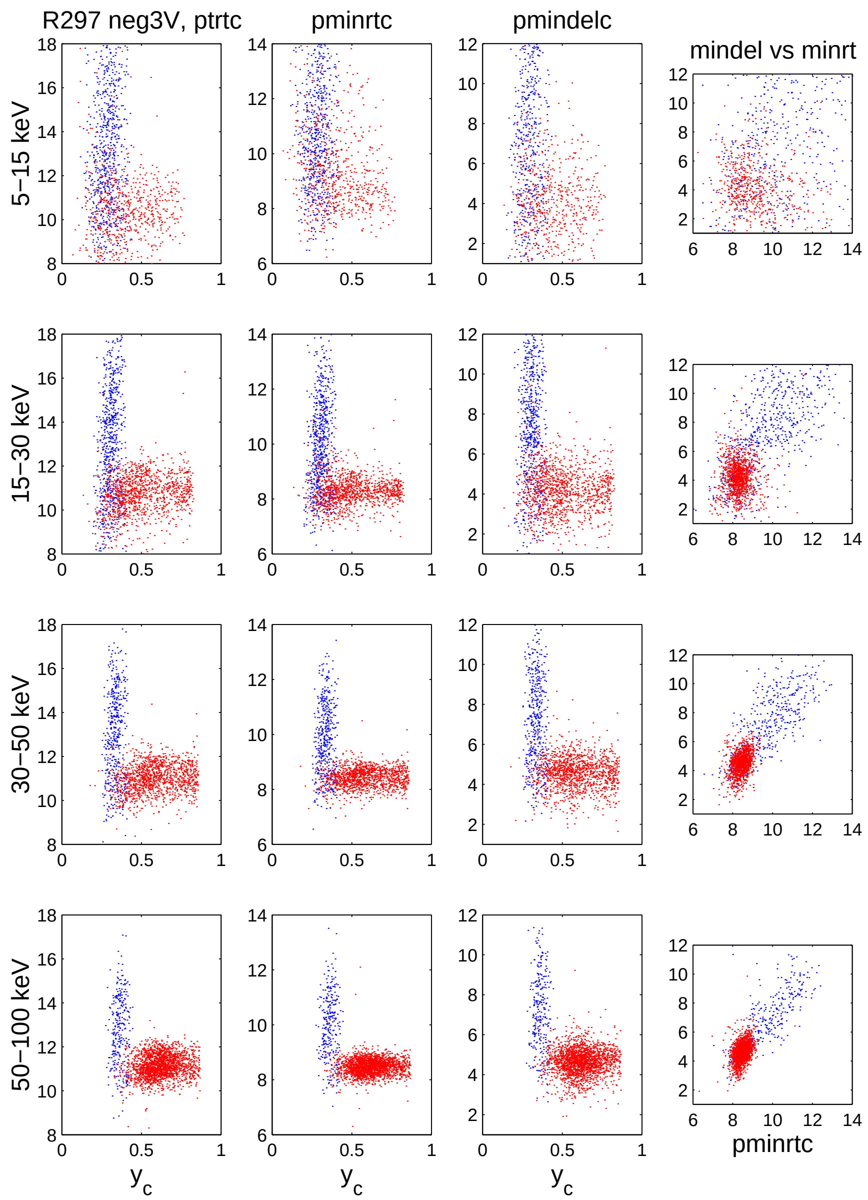


Figure 6.17: Same as Figure 6.16, but for UCBRun297 -3V.

of events observed in the NR band. In this case, one essentially wants to minimize the uncertainty on the signal estimate, which implies minimization of the quality factor Q (defined in 5.28). This approach will likely be useful in the analysis of the Soudan data, where the surface electron-recoil events are expected to be the dominant background.

Here we make a quick comparison of the three parameters by estimating the fraction of neutrons that survive the timing cut, when 10% of the betas survive such timing cut. Such comparison is more along the lines of Option 1 above - it hints which parameter would be best for a harsh cut. Figures 6.18 and 6.19 show the results for the cases of UCBRun296 -3V and UCBRun297 -3V, but the general conclusions hold in other cases as well.

In most cases, the delay of the largest phonon pulse (pmindelc) allows the largest fraction of neutrons to survive (when 10% of the betas survive). In some cases, the rise-time of the largest phonon pulse is close, or even better. The rise-time of the sum of the four phonon signals (ptrtc), which was used in the previous analyses (including the one discussed in the last Chapter), usually gives smaller fractions of neutrons that survive. Based on these results we decided to include pmindelc and pminrtc among the parameters to be position-corrected in the analysis of the first Soudan data.

6.7 Rejection of Surface Events

In this Section, we try to use both the ionization yield and the phonon timing parameters to study the surface events rejection efficiency. We are primarily interested in (1) determining which charge bias value provides the best rejection efficiency and (2) determining if one timing parameter (or a combination of them) gives particularly good rejection efficiency.

To calculate and compare the rejection efficiencies for different cases, we need to determine a way of defining a cut in phonon timing parameters. As before, we define α

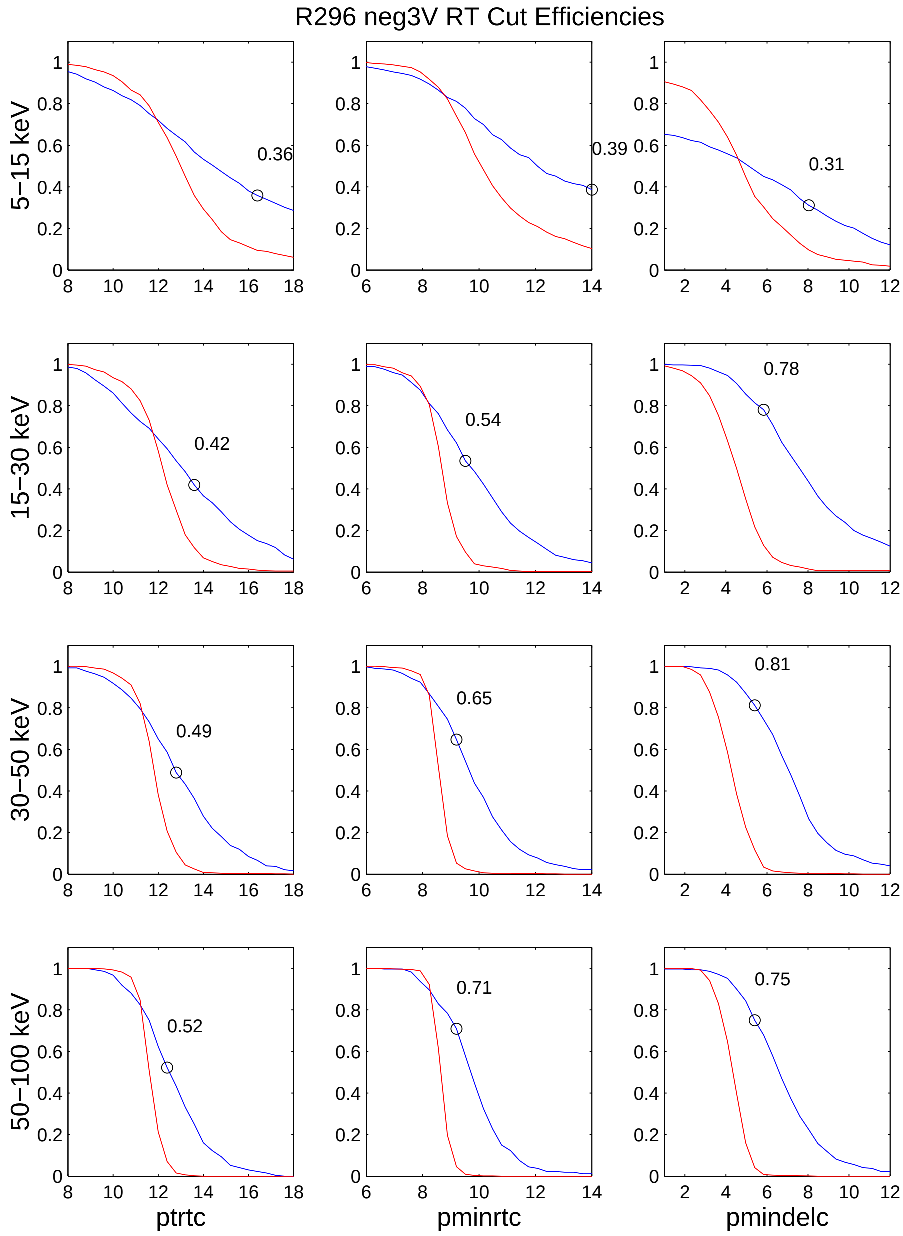


Figure 6.18: Efficiency of the phonon timing cut on neutrons (blue) and betas (red) for the case of UCBRun296 -3V, as the cut value scans through a range of values. The three columns correspond to the three timing parameters, and the four rows correspond to four energy bins. The numbers on the plots denote the fraction of neutrons surviving the cut when 10% of the betas survive it.

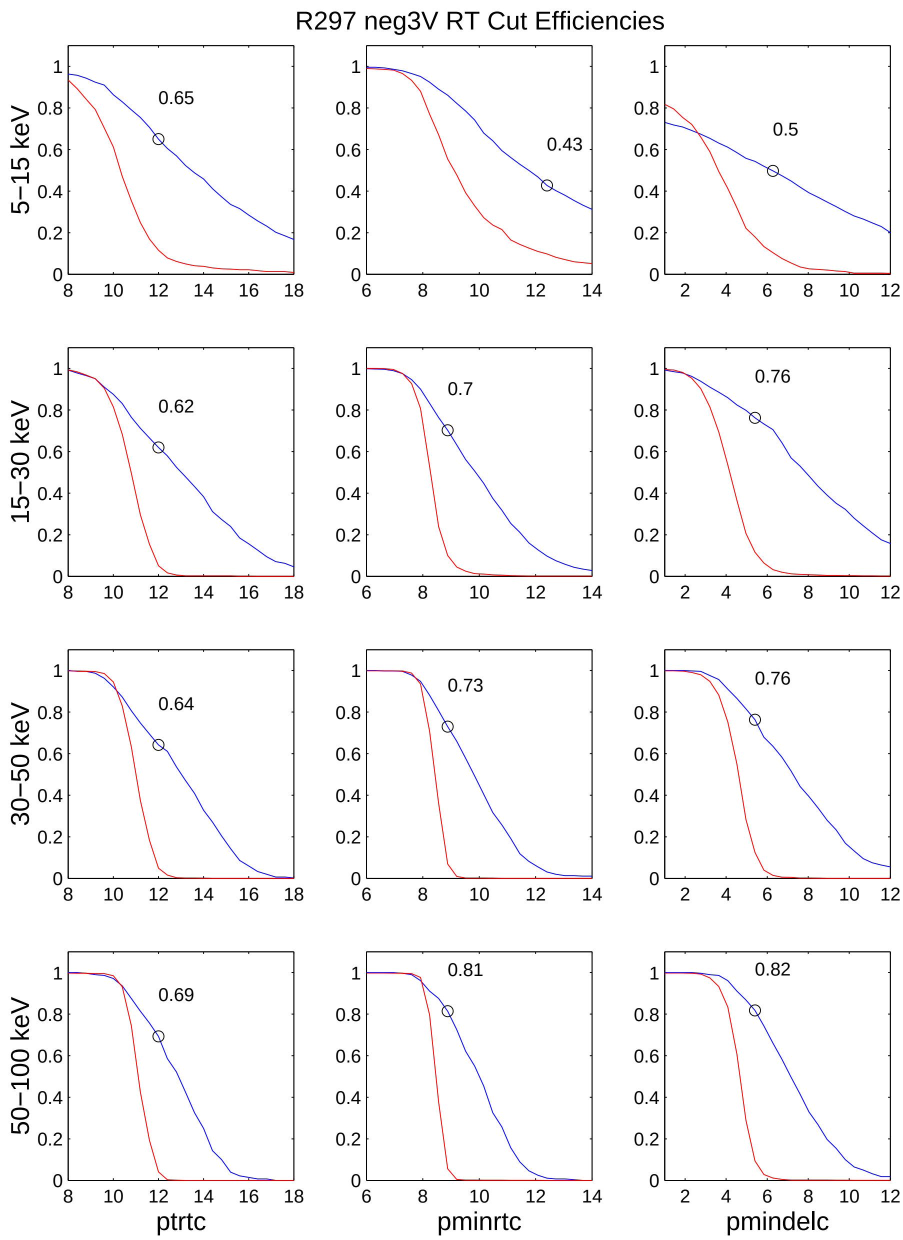


Figure 6.19: Same as Figure 6.18, but for UCBRun297 -3V.

and β to be the efficiencies of the cut in a phonon timing parameter for neutrons and betas respectively, and let γ be the efficiency of the combined cuts in the yield and the phonon timing parameter on betas. We will assume that statistical subtraction will be used in the calculation of the final limit, which implies that the cut should minimize the quality factor (see Section 5.6):

$$Q = \frac{\gamma(1-\gamma)}{(\alpha-\gamma)^2}. \tag{6.1}$$

The rejection efficiency of the surface events is therefore $1-\gamma$. We also define the effective rejection $Q_{rej} = 1 - Q$. In practice, minimizing Q is very sensitive to removing the last beta-event (at which point γ and Q drop to zero). Hence, we instead maximize $\alpha - \beta$ - for slowly varying β, which is true in our case, this is similar to minimizing Q. Figure 6.20 shows how the cut value is calculated for the example of the UCBRun297 -3V.

Having defined the cut value, we can proceed with the calculation of the surface event rejection efficiency and the effective rejection. In particular, we will combine the cases of UCBRun296 -3V and UCBRun297 -3V, taking into account the different statistics, in order to estimate the rejection efficiency of the detector if the two sides were illuminated equally (with betas). We do the same for the +3V cases. We directly compare the rejection efficiencies and effective rejection for the "two-sides-combined" -3V and +3V cases in the Figures 6.21 and 6.22. Since UCBRun296 -6V and UCBRun297 +6V are excluded from the analysis (problems with box plots, see Section 6.3), the only other possible comparisons are UCBRun296 +3V vs +6V cases, and UCBRun297 -3V vs -6V cases. These are shown in Figures 6.23, 6.24, 6.25, and 6.26.

Several conclusions can be made based on these plots:

- The -3V case gives slightly better effective rejection efficiency than other bias values. However, depending on the phonon timing parameter, +3V is not much worse.
- For the -3V case, parameters ptrtc and pmindelc perform very similarly (within error

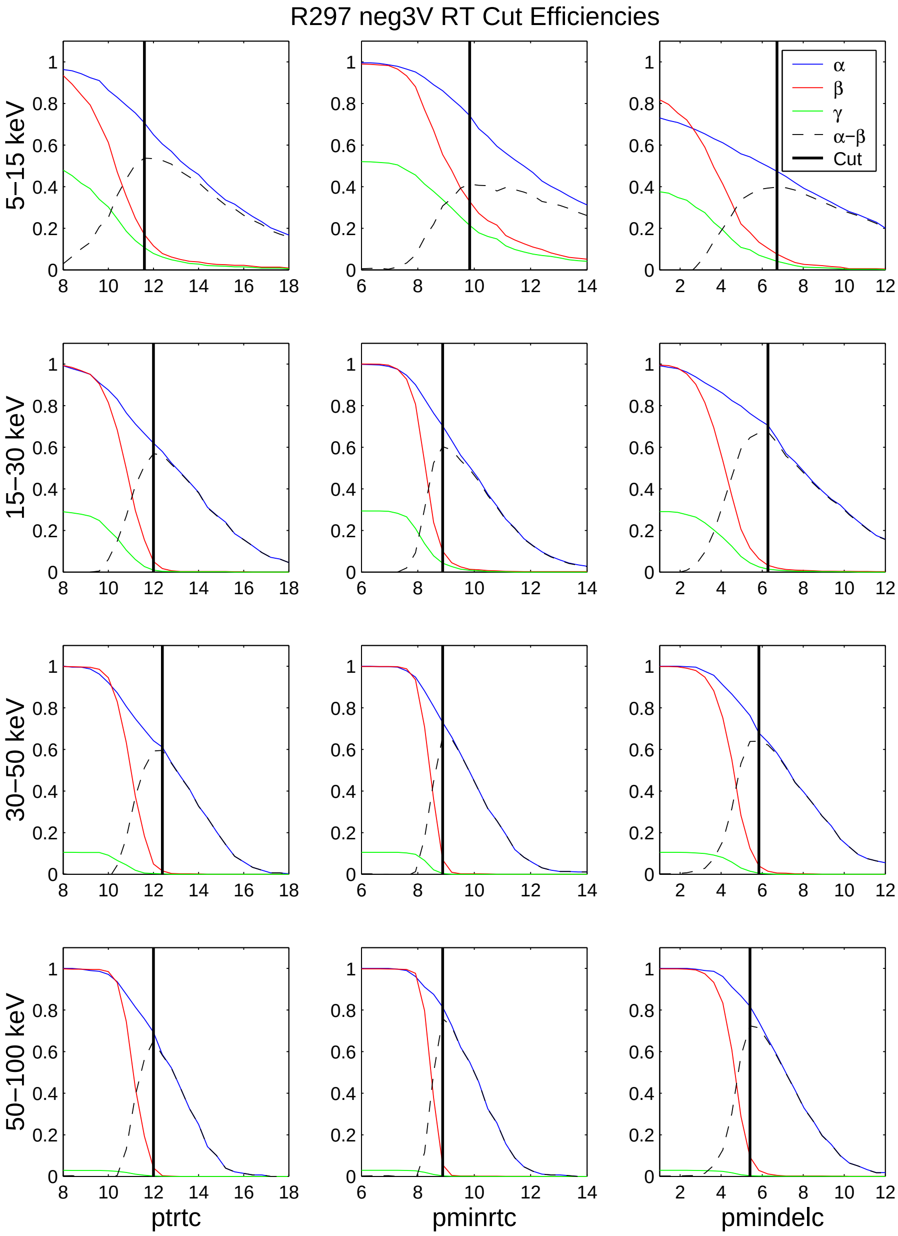


Figure 6.20: Efficiencies of the cut in phonon timing parameters for UCBRun297 -3V. The three columns correspond to the three parameters, the four rows correspond to the four energy bins. α and β are efficiencies of the cut for neutrons and betas respectively, and γ is the efficiency of the combined cuts in yield and in phonon-timing parameter on betas.

bars), while pminrtc is somewhat worse at low energies.

- The performance with ptrtc is very similar to the estimate made for the SUF Run21

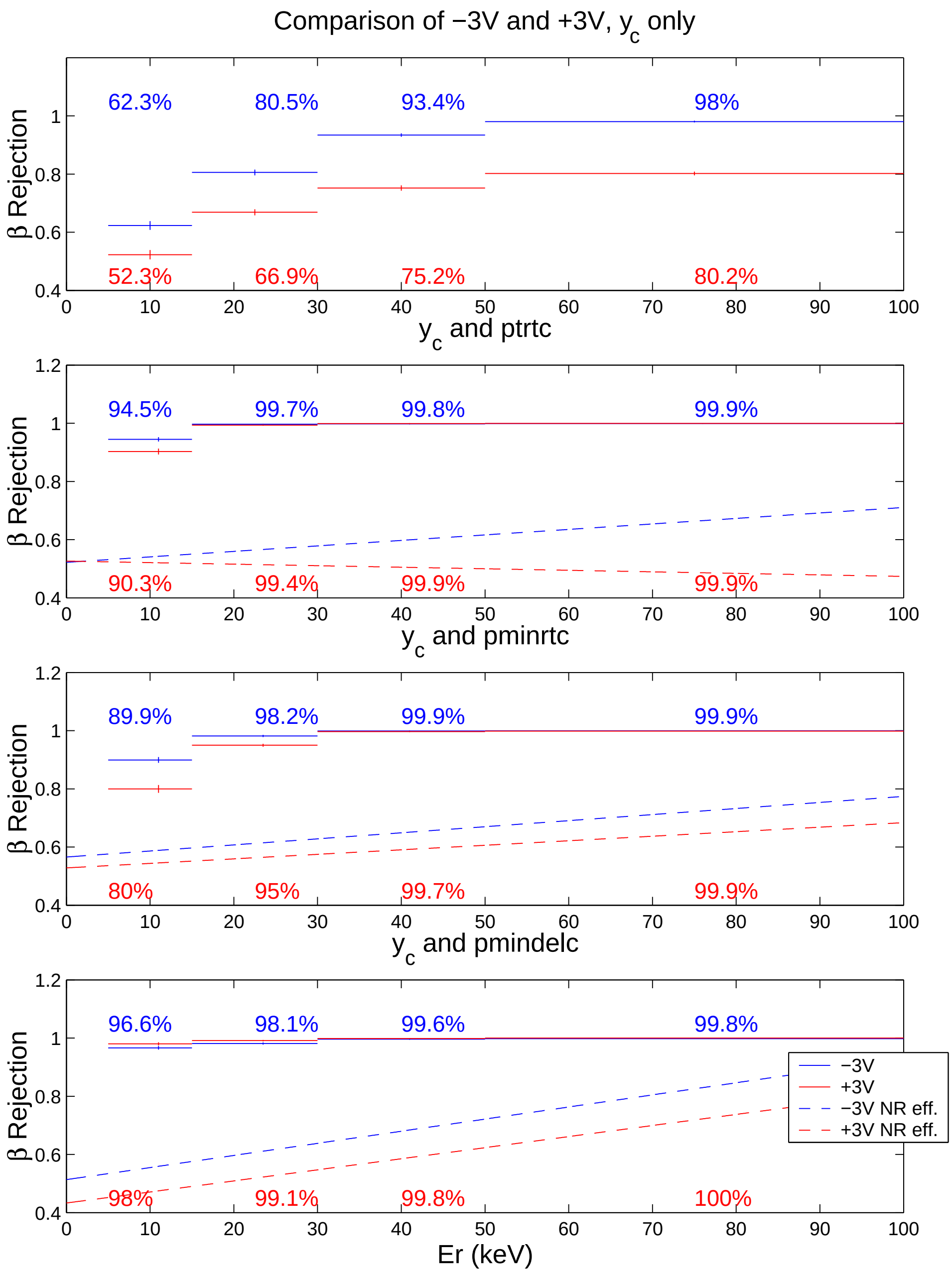


Figure 6.21: Comparison of surface events rejection efficiency for the -3V and +3V cases (two sides of the detector combined), for energy bins 5-15, 15-30, 30-50, and 50-100 keV. The numbers denote the rejection efficiencies in the corresponding energy bins.

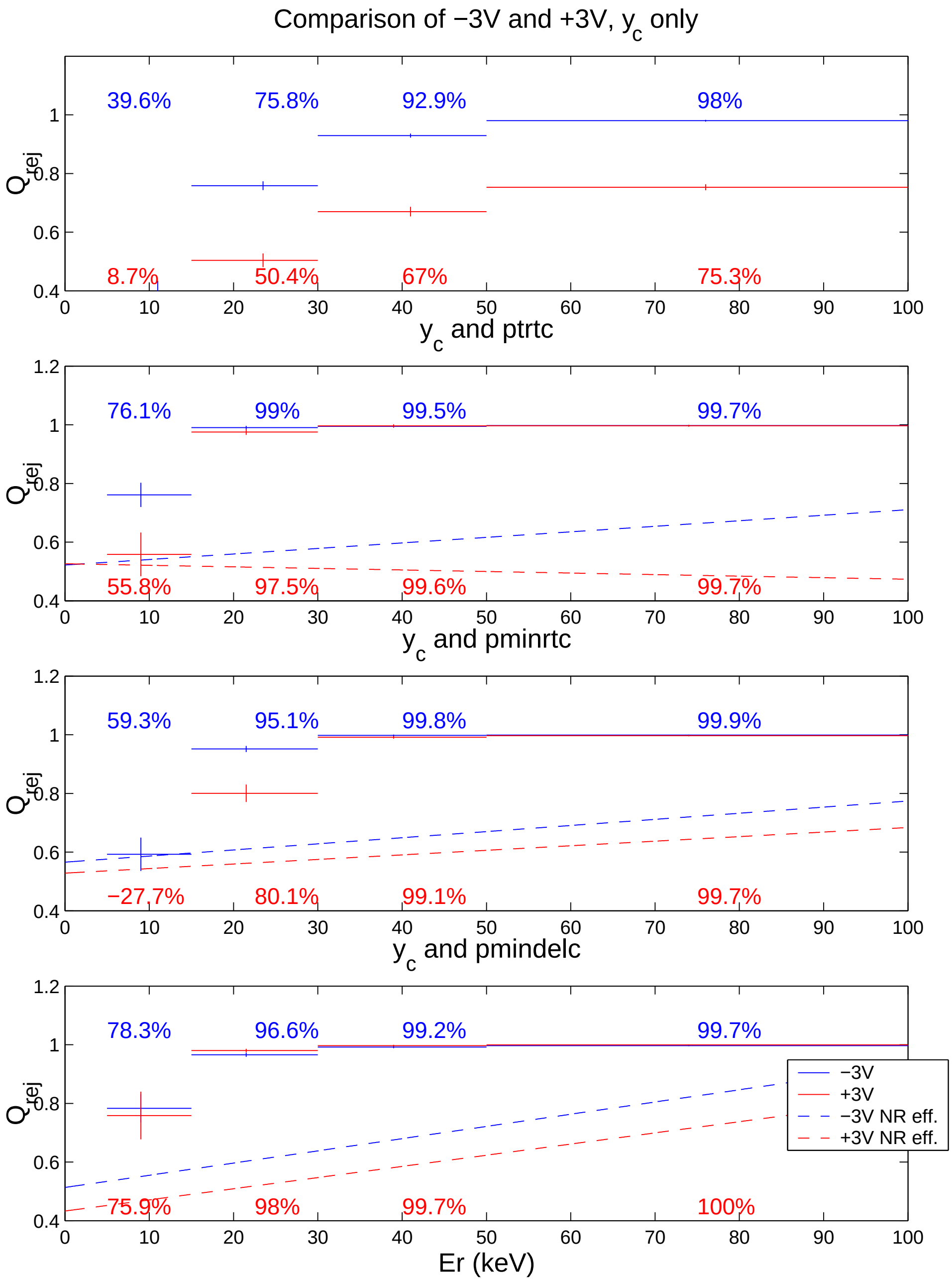


Figure 6.22: Comparison of effective rejection of surface events for the -3V and +3V cases (two sides of the detector combined), for energy bins 5-15, 15-30, 30-50, and 50-100 keV. The numbers denote the effective rejections in the corresponding energy bins.

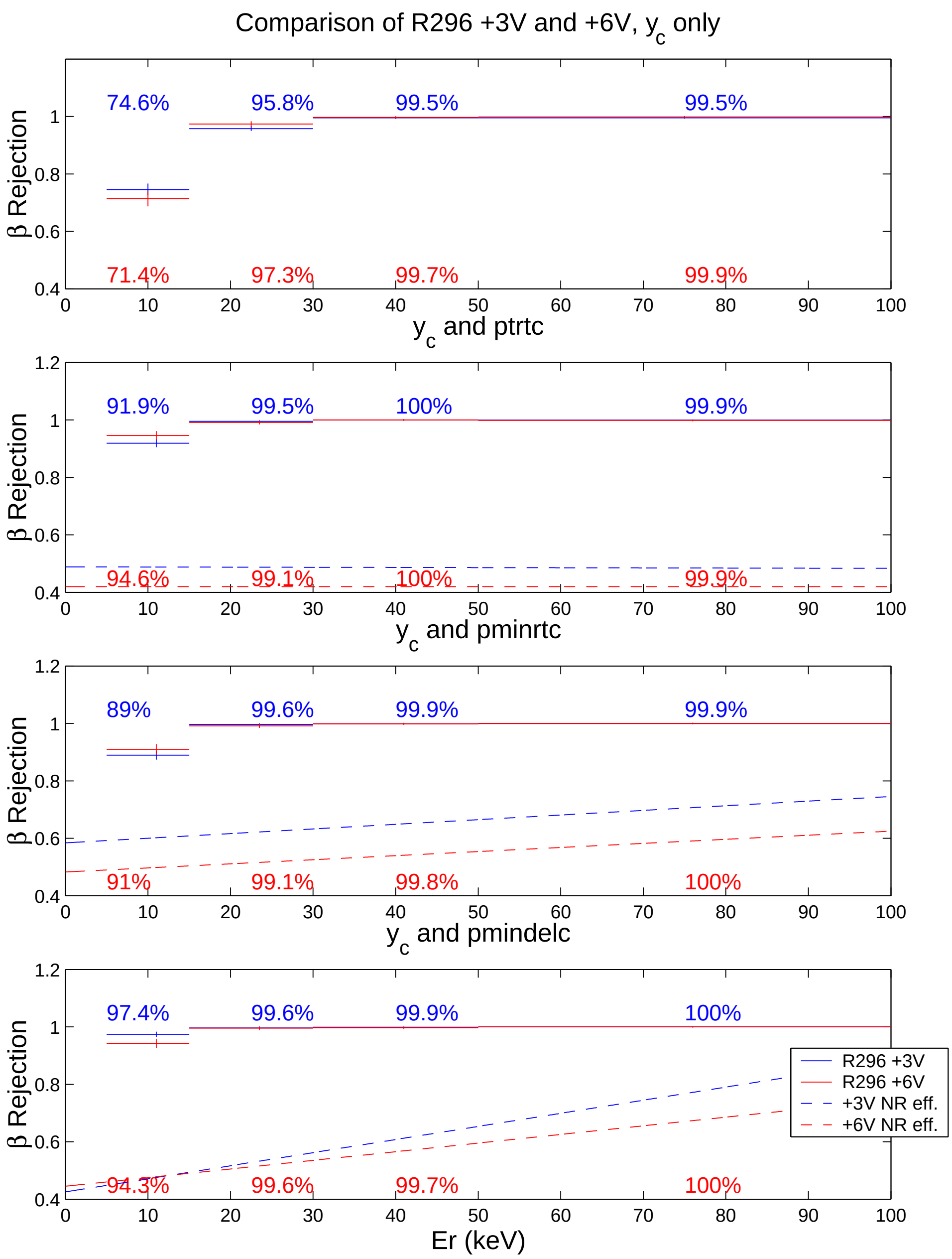


Figure 6.23: Comparison of surface events rejection efficiency for the UCBR296 +3V and +6V cases, for energy bins 5-15, 15-30, 30-50, and 50-100 keV. The numbers denote the rejection efficiencies in the corresponding energy bins.

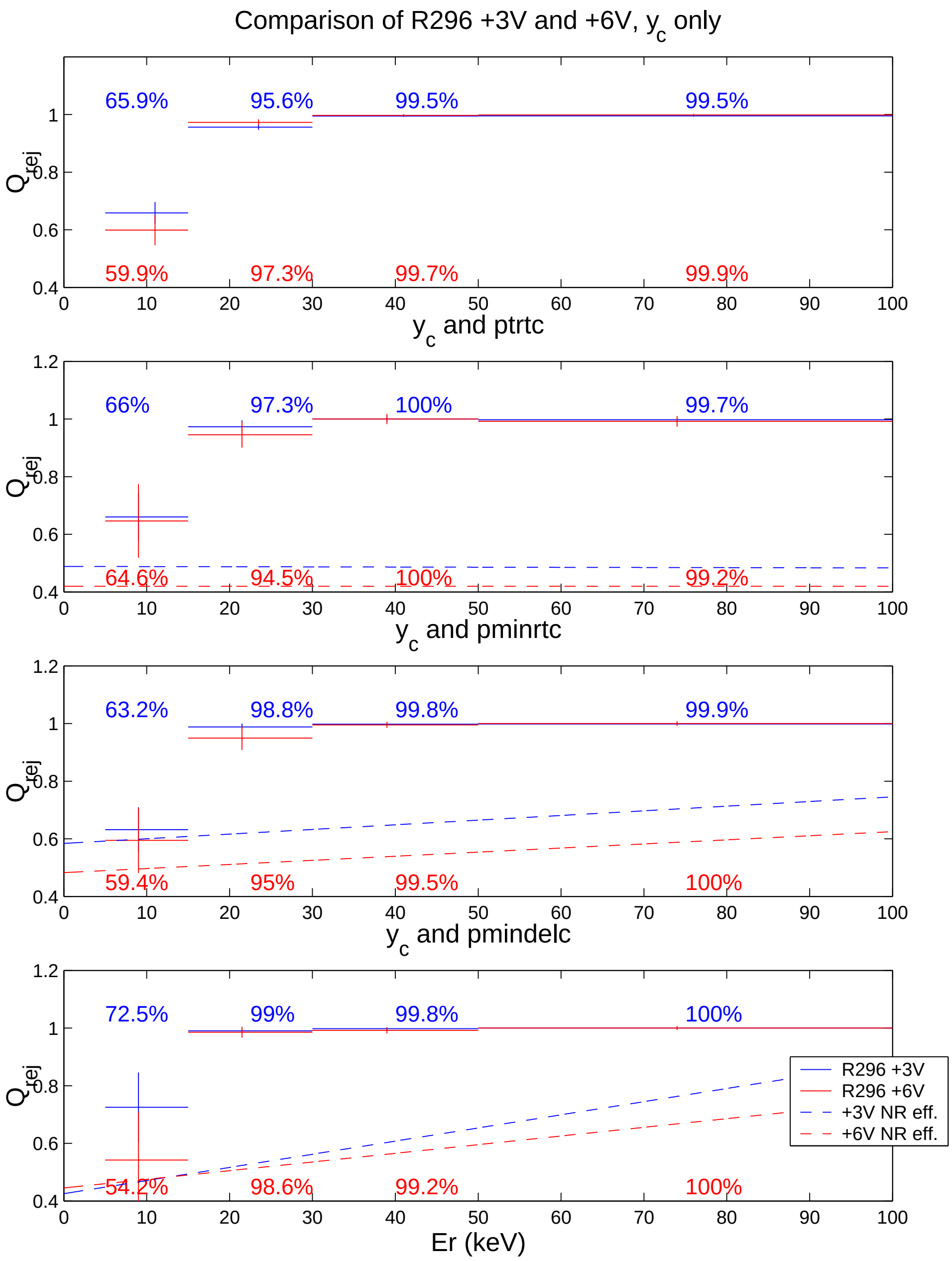


Figure 6.24: Comparison of effective rejection of surface events for the UCBRun296 +3V and +6V cases, for energy bins 5-15, 15-30, 30-50, and 50-100 keV. The numbers denote the effective rejections in the corresponding energy bins.

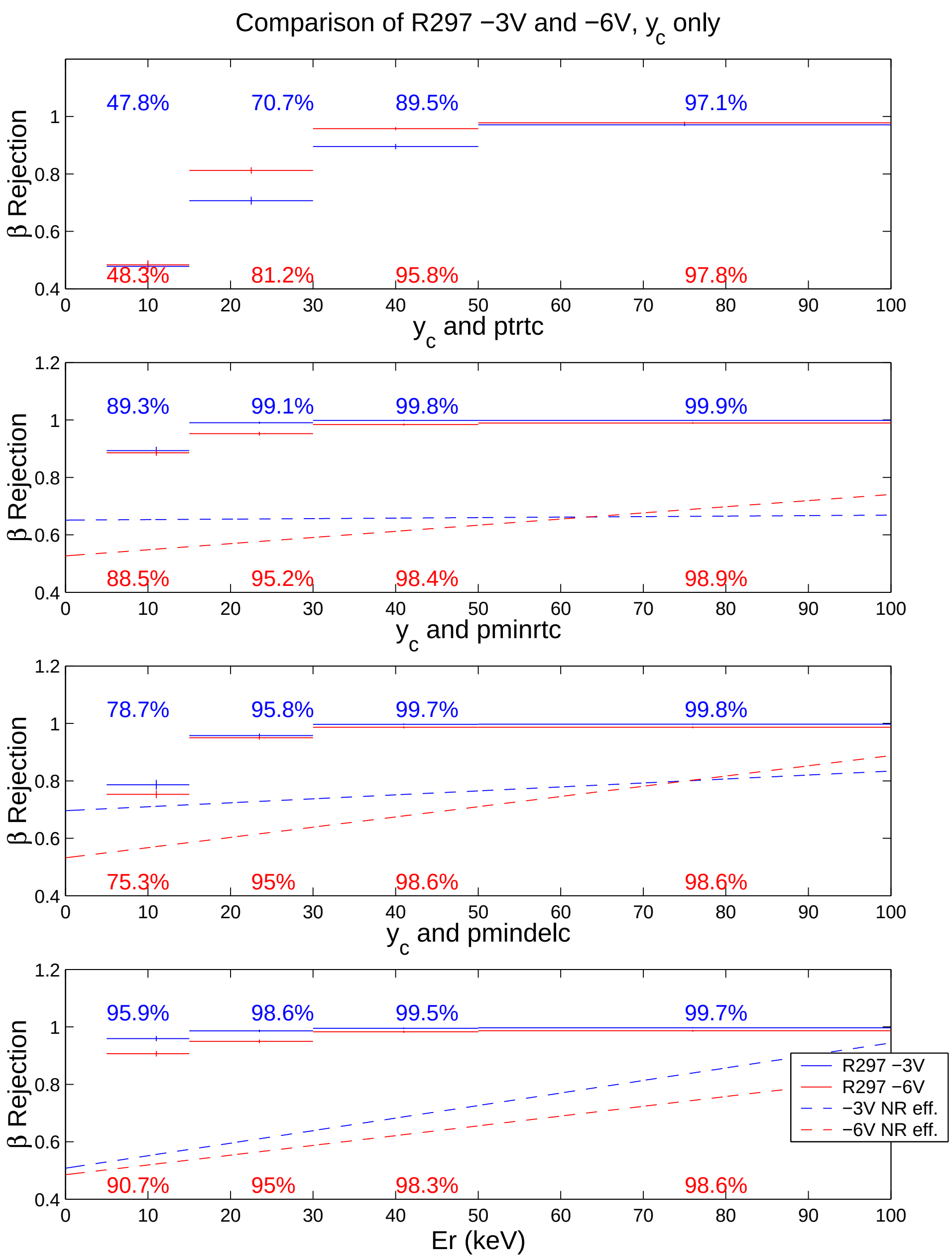


Figure 6.25: Comparison of surface events rejection efficiency for the UCBR297 -3V and -6V cases, for energy bins 5-15, 15-30, 30-50, and 50-100 keV. The numbers denote the rejection efficiencies in the corresponding energy bins.

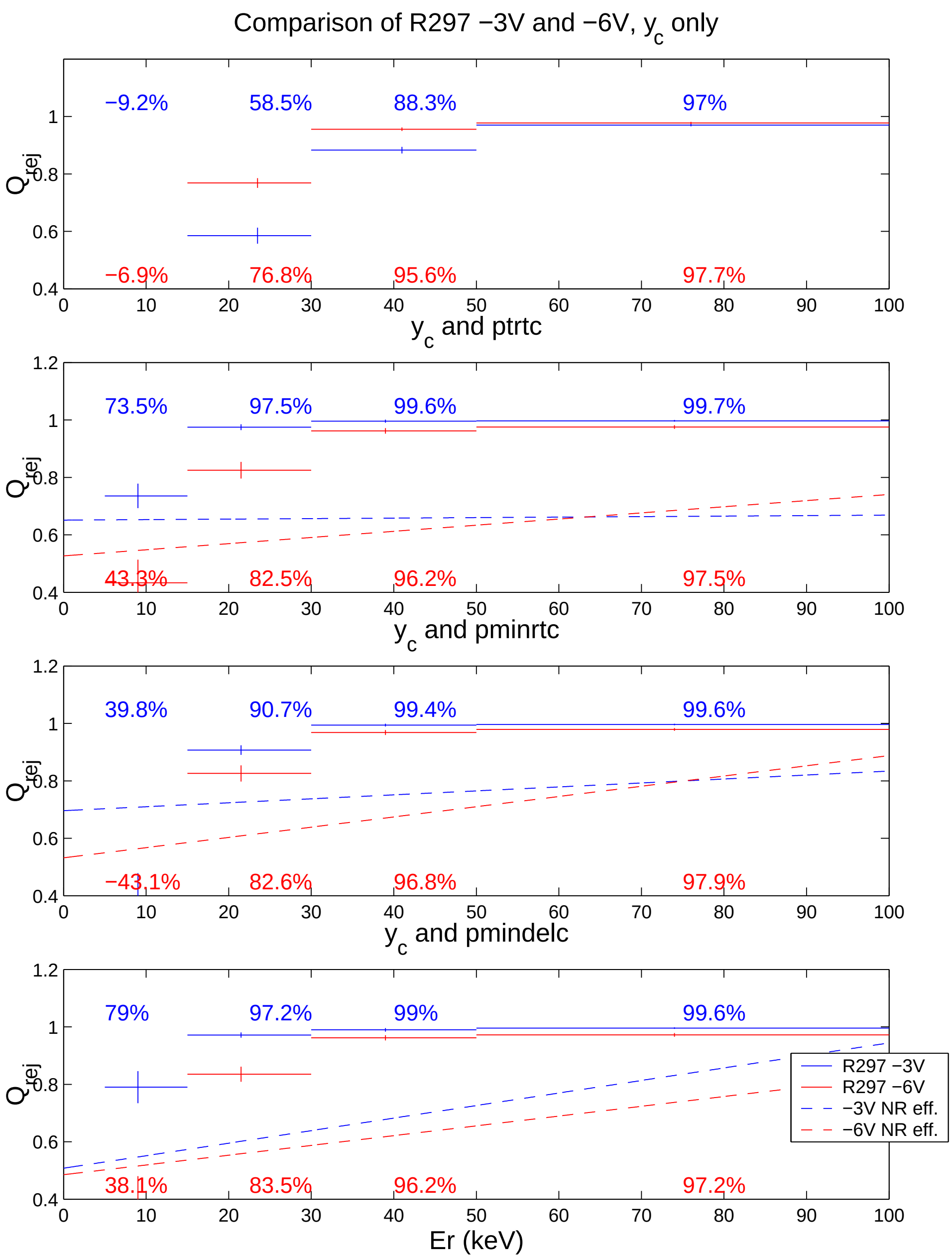


Figure 6.26: Comparison of effective rejection of surface events for the UCBRun297 -3V and -6V cases, for energy bins 5-15, 15-30, 30-50, and 50-100 keV. The numbers denote the effective rejections in the corresponding energy bins.

-3V, shown in Figure 5.9. Note, however, that that estimate was done in a different way. At lowest energies, the G31 beta calibration estimate of the rejection seems a bit better than the Run21 estimate, but the rise-time cut for Run21 was not defined in the same way.

It is, however, possible that the three phonon timing parameters carry complementary information. Indeed, if one attempts cutting in two of the parameters, one can achieve even better rejection efficiencies. Figure 6.27 shows how such a cut can be made for the -3V case (both sides of the detector (i.e. UCBRun296 and UCBRun297) combined). The cut is again defined by maximizing $\alpha - \beta$, and the cut efficiencies are shown in Figure 6.28. Figure 6.29 shows the corresponding surface event rejection efficiency and effective rejection. Note that the results are better than those based on the single parameter cuts, shown in the Figures 6.21 and 6.22.

6.8 Yet Another Discrimination Parameter

We conclude this Chapter by pointing out yet another parameter that can be used for rejecting the surface events. This parameter has not been used in the analysis above, primarily because we thought of it after the position correction was completed. However, in the future passes of position correction of the G31 data, it would be very good to include this parameter.

Unlike the parameters studied above, which were based on phonon timing, this parameter is extracted from phonon amplitudes. Namely, the total phonon signal created in an interaction is distributed among the four phonon sensors. It is not unreasonable to

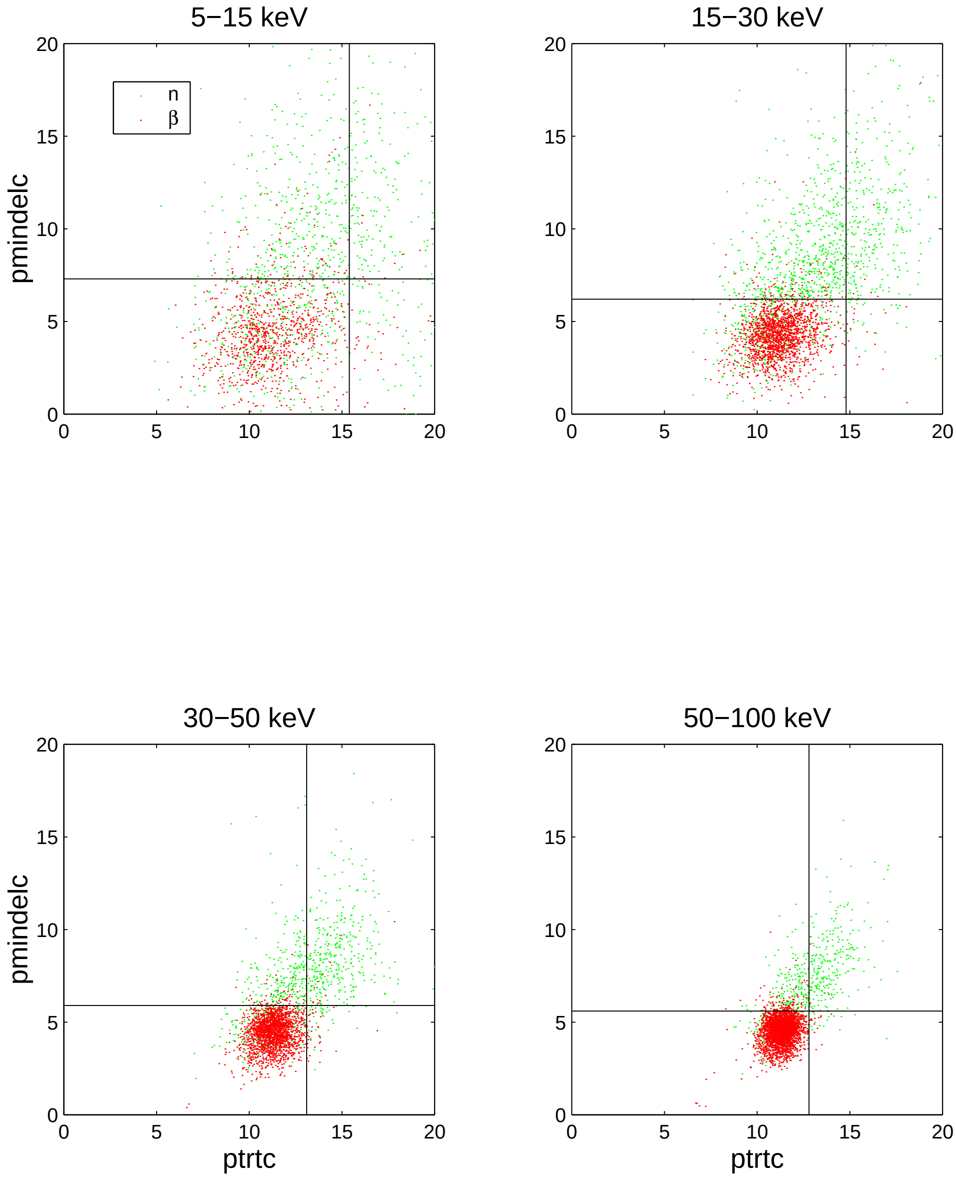


Figure 6.27: Defining the cut in two phonon timing parameters (ptrtc and pmindelc) for the -3V cases combined. The lower left corner in each plot is rejected.

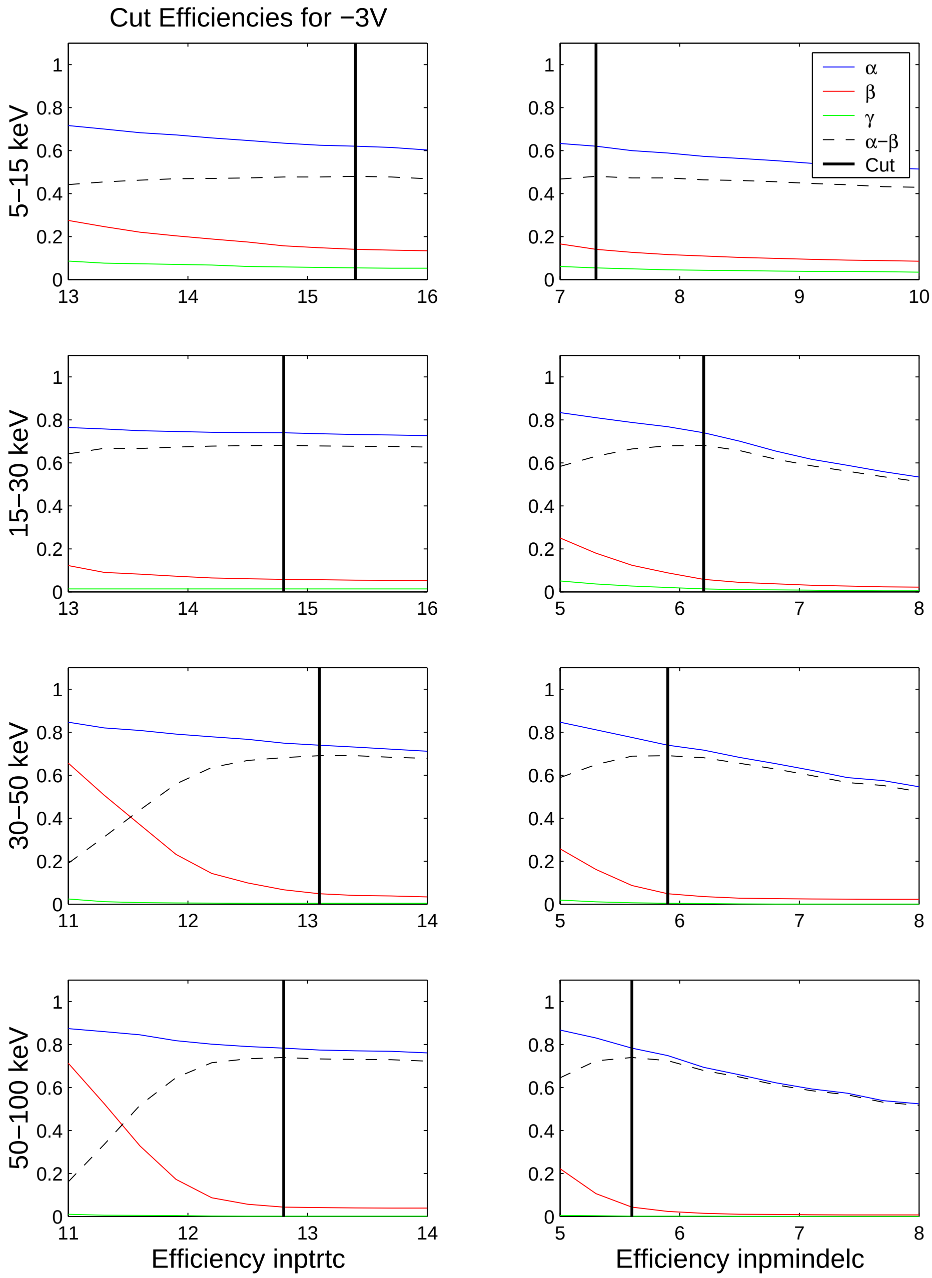


Figure 6.28: Efficiency of the cut in ptrtc-pmindelc plane around the optimal point for the -3V cases combined. The first column fixes the cut value in pmindelc and scans the cut value in ptrtc (vice versa for the second column). The four rows correspond to the four energy bins.

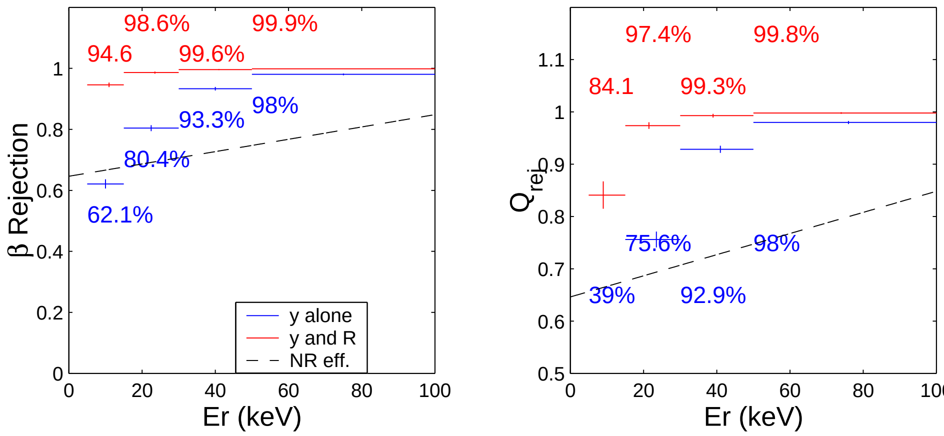


Figure 6.29: Rejection efficiency and effective rejection for the cut in the ptrtc-pmindelc plane for the -3V cases combined.

expect that (for events taking place on the phonon side) a larger fraction of the total phonon signal will be "absorbed" by the primary sensor (under which the event happened) for the surface events than for the bulk events.

To test the idea, we define the new parameter, $pfrac$, to be the ratio of the amplitudes in the primary sensor and the sensor directly opposite to it. Clearly, such parameter cannot be very useful at the center of the crystal, where the signal distribution is determined by the geometry (e.g. an event at the center would be equally distributed among the four phonon sensors, regardless of whether the event was on the surface or in the bulk of the crystal). But, it may become useful further away from the center (possibly as close as 1 cm away from the center) where majority of the events takes place.

To test this parameter, we look at the UCBRun297 -3V case, and we consider the four outer Cd blobs. We do not consider the inner blobs, as we do not expect the parameter to be useful there due to their closeness to the center. We select the bulk (gamma) and surface (beta) events in the 40-60 keV bin. Figure 6.30 shows that the phonon distribution

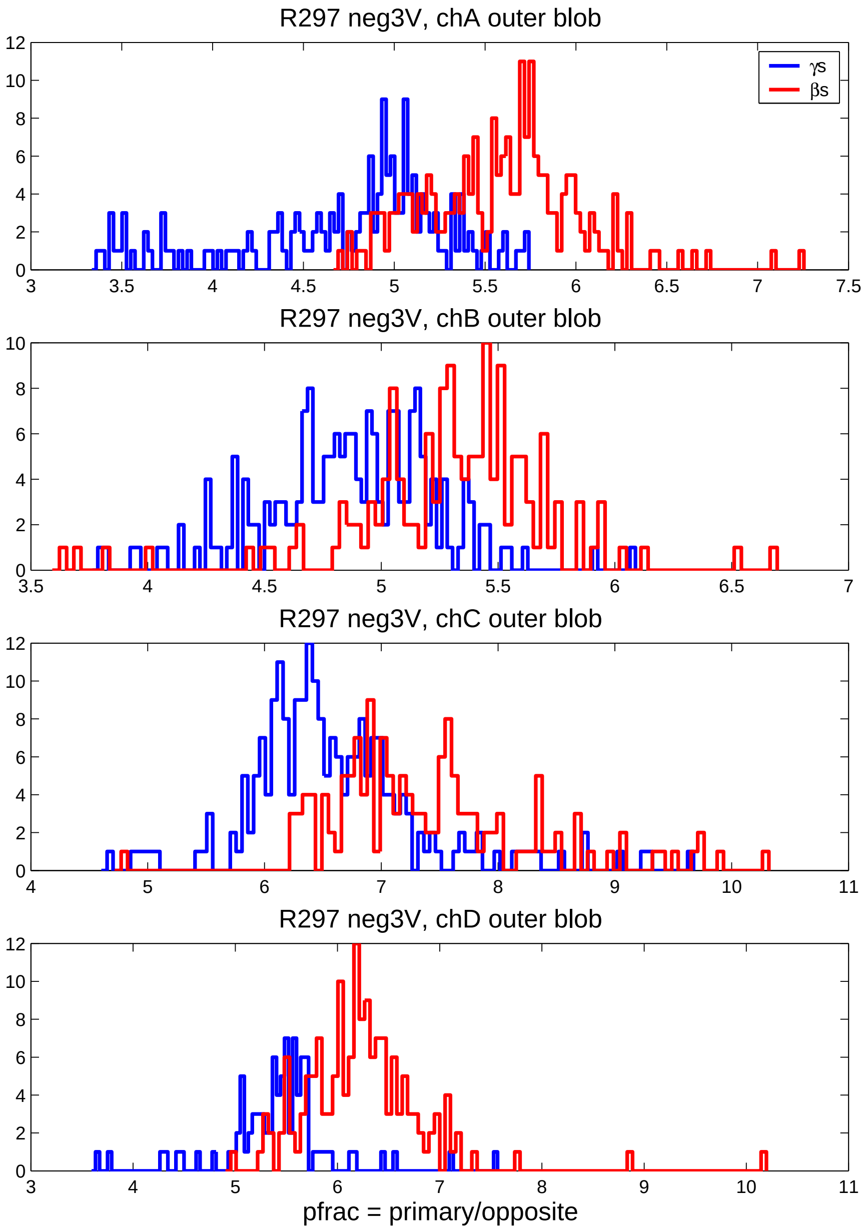


Figure 6.30: Distribution of the phonon distribution parameter $pfrac$ for the bulk and for the surface events for the case UCBRun297 -3V. The four plots correspond to the four outer Cd blobs.

parameter $pfrac$ tends to be larger for the surface events than for the bulk events. Note that, since the parameter has not been position corrected, it is not meaningful to treat the detector as a whole, or even to combine the four blobs into one. However, the distributions of the surface and of the bulk events can be compared at each local point, such as for each individual Cd blob.

Although Figure 6.30 is quite convincing in demonstrating the ability of the phonon distribution parameter to reject surface events, a more careful analysis would certainly be in order. Besides the position correction, one may want to construct a more complex parameter that would capture information in all four phonon amplitudes, rather than just two. In addition, position dependence, energy dependence, and detector-side dependence of this parameter should be studied. Indeed, the first preliminary results indicate that the parameter is far less useful for rejecting surface events on the charge side of the detector (i.e. in the UCBRun296).

Chapter 7

Soudan Commissioning

7.1 Introduction

In this Chapter we attempt to describe the various difficulties and problems (some of which were truly incredible), encountered while commissioning the CDMS II experiment at the Soudan mine in northern Minnesota. Commissioning CDMS experiment at Soudan was a very difficult process, by far the most difficult aspect in my involvement with CDMS as a graduate student. Having survived this phase (and having recovered from it for several months) I can appreciate more fully the learning curve we have gone through. Two aspects have left a particularly strong impression on me.

First, getting CDMS experiment to work in Soudan was a true team-effort. Almost all members of the collaboration cycled through the Soudan site and spent much time (often several weeks at a time) fighting various issues. Looking back at the variety and complexity of the problems we have encountered, I am convinced that a single person, or a small group of people, would not have succeeded. The importance of working in a collaboration is certainly more clear to me than it ever was before.

Second, several members of the collaboration (including the most senior ones) have set striking examples of the level of dedication required to get an experiment to work. To young scientists, such as myself, such examples are great lessons, and they influence our own attitude towards science.

The problems encountered in Soudan varied in nature. Section 2 covers the general issues and difficulties regarding the work in the Soudan mine. Section 3 gives an overview of the CDMS II setup at Soudan. Sections 4-9, cover a variety of cryogenic problems we have dealt with, in a roughly chronological order. Section 10 describes various electronic noise issues. Section 11 discusses the remaining aspects of the commissioning phase, including the passive shield, the muon veto system, the data acquisition system, and the data analysis software.

7.2 Working in Soudan

7.2.1 Preliminaries

Soudan is a very small town in northern Minnesota, about 3.5 hours of driving from the Minneapolis airport. The size of Soudan is reflected in a single store/gas station. About two miles away is Tower, a somewhat larger town with a reasonable size store and several bar/restaurants. About twenty miles away is Ely, an even larger town with a bit more variety in restaurants and things to do after work.

The CDMS II experiment was constructed on the 27^{th} level of the old iron mine in Soudan, at the depth of 2341 feet. The mine itself is not active any more, and it has been converted into a tourist attraction. In fact, every summer tens of thousands of tourists visit the mine and take tours to the bottom of the mine, where a (very good!) re-creation of the mining conditions from many decades ago was made.

The access to the mine is provided via a single, two-cage shaft, which allows direct travel from the surface to the 27^{th} level. The cages themselves have intentionally not been modernized, in order to depict the conditions of when the mine was active. They are operated by an ancient engine, also preserved to be a part of the Soudan-mine tours. In the winters of 2002 and 2003, work was going on to replace the aged rails. Typically, our work-day would start at 7:30 am, when the cage would go down to the 27^{th} level, and end at 5:30 pm, when the cage would go back to the surface. In the summers, there was some more flexibility since one could go down with the tours. In the winters, however, that was much harder, particularly because of the work on the rails and because of the commissioning of the MINOS experiment, which was also constructed at the 27^{th} level and it required large steel plates to be hauled down day-after-day (MINOS construction finished toward the end of the summer of 2003).

Much of the construction of the RF-room, the ante-room, and the electronics room, the preliminary assembly of the cryostat/Icebox, and the assembly of the shield was completed in 2000 and 2001. In December 2001, we started a more organized rotation schedule of physicists, with the goal of resolving a number of cryogenic and other issues. At first, we had only one or two physicists present in the mine at a time. Very quickly, it became obvious that a more substantial presence was necessary, and much attention had to be paid to overlapping people's stays, to achieve the continuity in the work. By the summer of 2002, we routinely had 3-4 physicists in the mine, along with 2-3 technicians/engineers. At first, most of us were staying at the “Tubs” in Tower - decently equipped apartments, but located just above a laundromat. By the summer of 2002, CDMS has rented two houses, one in Soudan and one in Ely, both of which were nicely furnished and very pleasant to stay.

7.2.2 Difficulties

Commissioning the CDMS experiment in the Soudan mine was difficult for several reasons.

Remoteness

The remoteness of the site was a major hurdle in two different ways. First, one had to plan very carefully what kind of equipment one would need, and make sure that the equipment was available at Soudan. This was true for both electronic equipment (oscilloscopes, volt-meters etc) and for different types of hardware (screws, metal pieces etc), as the supply of such items in Soudan was very limited. Second, travel to Soudan is not trivial at all. For people travelling from California, it requires at least a 4-hours flight (if one decides to fly direct), followed by a 3-4 hours drive. From personal experience, such trips have significant impact on productivity, as it would usually take a couple of days to recover. Furthermore, they are quite exhausting if one is on a weekly or bi-weekly rotation schedule.

Limited Access

The access to the experiment was largely determined by the dedicated cage-rides at 7:30 am and at 5:30 pm (3:30 pm on Saturdays). Cage trips at other times were not always possible due to the work on the rails or due to hauling of the MINOS equipment. Staying after 5:30 pm was possible during a large part of the CDMS II commissioning (since MINOS construction was being done in two shifts), but it usually meant staying until very late due to the limited number of cage-rides. Such limited access meant three things. First, items forgotten on the surface would very likely not be available until the next day. Second, going out for lunch was not a possibility, so one had to prepare a lunch box (usually a day in advance) and bring it to the mine. Third, in the winters, one would not see the daylight from Sunday-to-Sunday - an experience that certainly does not help one's mood or work attitude.

Noise

Once in the mine, one is constantly surrounded by various machines, pumps, air-conditioners, fans etc. All of these continuously produce background noise, that makes an already difficult work place even less comfortable.

Weather

Winters in northern Minnesota are known to be very cold. While some may find that advantageous, it is certainly not trivial to alternate between sub-zero temperatures in Soudan and the sixties in northern California. Plus, of course, winter does not make the long drives any easier.

On the Good Side

The difficulties listed above significantly affect the overall productivity at Soudan - some estimates indicate that it is 3-4 times more costly and more time-consuming to accomplish an experimental task at Soudan than, for example, at a University campus. However, to end this Section on a positive note, there are aspects of the northern Minnesota that, at least partly, make up for the difficulties.

For example, the summers in Ely and the Boundary Waters are very beautiful, with many lake or hiking oriented things to do. The peaceful nature of that region really does have a soothing effect after a noisy day in the mine. The people (and, of course, especially our dedicated technician Jim Beaty) are always very welcoming and willing to help. And, introducing the ping-pong table at the bottom of the mine, while it raised the background noise level somewhat, certainly made the atmosphere in the mine more colorful.

7.3 Overview of the CDMS II Setup

The CDMS II experimental setup is schematically shown in Figure 7.1.

The heart of the experimental setup is a class-1000 clean room, whose walls are grounded, thereby creating a good Faraday cage. The cryostat and the first-stage room temperature electronics are installed in this room. All the plumbing from the cryostat, along with thermometry and other wiring, goes out through the RF wall into the cryopad area. The cryopad area hosts all the pumps, the circulation system plumbing, the automatic cryogen-transfer system, and all the electronics needed to run the system automatically. The signals coming from the detectors inside the cryostat, go through the first stage of the room-temperature electronics (the Front End Boards, FEBs), after which they are taken outside of the clean-room and into the Electronics Room on the second floor. The Electronics Room hosts the remaining room-temperature electronics, including the Receiver-Trigger-Filter (RTF) boards, digitizers, muon veto hardware, data acquisition hardware etc. The last room in the setup is the Ante-room. This room is the "changing room", before entering the clean room itself, and it also serves as the assembly room, where various materials are cleaned and assembled before being taken into the clean room. On the surface of the mine, the setup is complemented with a "surface trailer" (later substituted by a "surface building"), where the CDMS II Soudan Analysis Cluster (SAC) resides.

7.4 Cryogenics: How Things are Meant to Work

The Oxford Kelvinox 400-S dilution refrigerator is designed to provide 400 μW of cooling power at 100 mK. To this cryostat, we attach a custom-made extension, named Icebox, whose purpose is to provide a clean environment for the detectors (Icebox is made of radioactively clean materials, unlike the cryostat itself, which contains steel, indium etc). The Icebox is designed such that there is no direct line of site from the cryostat to the detectors. Similarly to the fridge itself, the Icebox consists of several "concentric" layers between the room temperature wall and the base stage, which mate with the corresponding layers in the fridge. The setup is meant to be very similar to the one at the Stanford Underground Facility, except that the electronic stem (E-stem), through which the signals are brought to the room temperature, is coming out opposite from the cold-stem (C-stem), which connects the Icebox and the fridge (at SUF the E-stem and the C-stem come out on

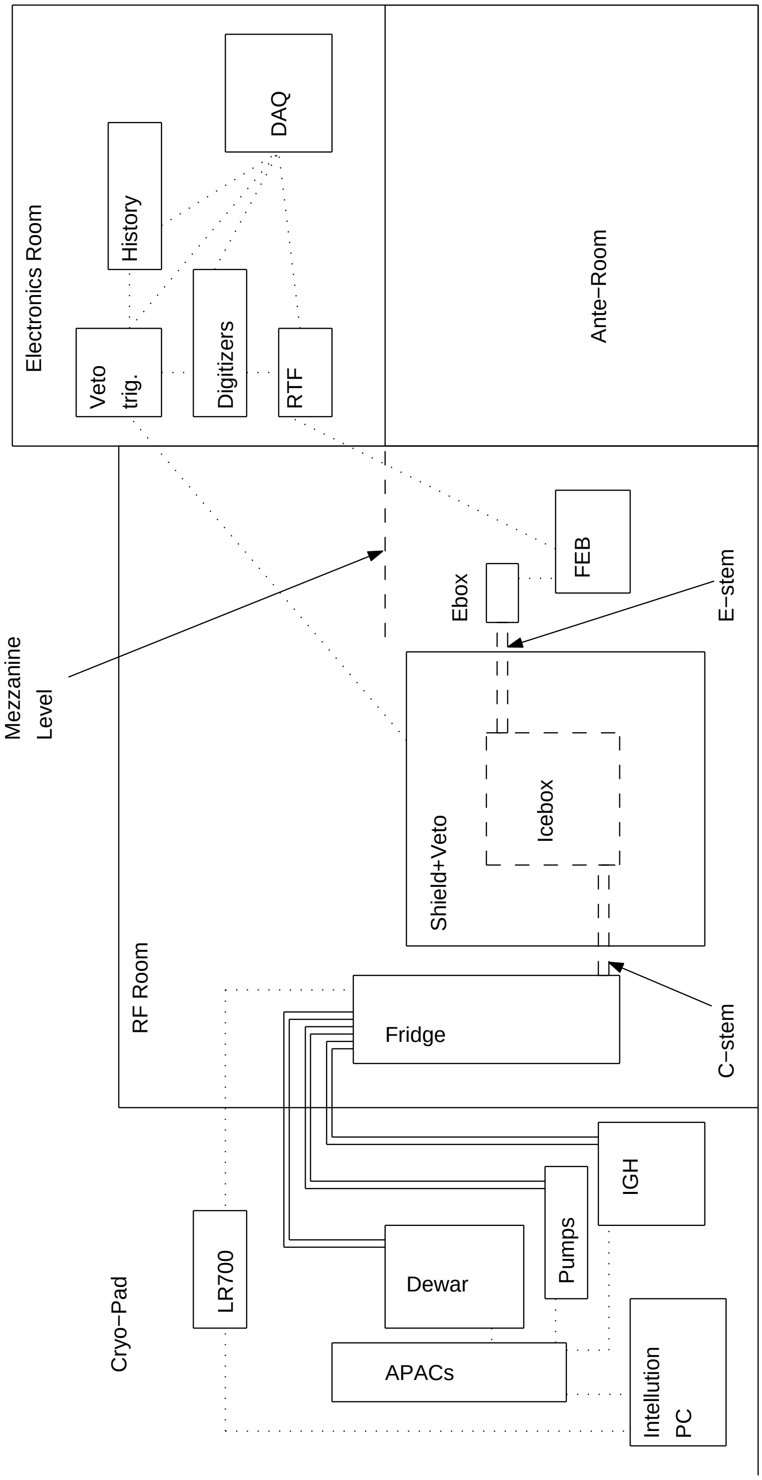


Figure 7.1: Schematic of the CMDS II experimental setup in Soudan mine.

the same side of the Icebox; see Chapter 3 and [2] for more detail on the SUF setup).

A serious complication of the Soudan cryogenic system, with respect to the one at SUF, comes from the fact that our access to the experiment is limited. In practice, this means that the system should be able to survive for a couple of days on its own, and that it should be able to recover from the occasional power trips or outages, which are common in northern Minnesota, especially in the summers. This requirement essentially implies that a computer (powered from an uninterruptable power supply, UPS) should observe the level of cryogens in the fridge and automatically perform transfers as necessary. Given the rates of consumptions of the cryogens, it is necessary to have two LHe dewars and two LN dewars plugged into the transfer system at any given time, and the computer must be able to recognize when one dewar runs out during a transfer and switch to the other one. Besides the automatic transfers, this computer should also monitor various pumps, restart them after power glitches, monitor the various pressures and temperatures in the system etc.

To achieve such automatic operation, several systems are deployed. The standard (non-dilution unit) cryogenic systems are monitored and controlled using an industrial-strength control and monitoring unit (APACs), with Intellution (for "intelligent solution") software interface. The Intellution is running on a PC with the Windows NT operating system. In addition to APACs/Intellution, we also use the Intelligent Gas Handler (IGH), provided by Oxford Instruments, designed to control the ^{3}He/^{4}He circulation. Oxford Instruments also provide Labview software to operate the IGH, which should allow the fridge to cool down to base temperature with 3 mouse clicks! At least for the CDMS II setup at Soudan, this is not true - it is very hard to cool down to base temperature even manually, so we use the Labview software for monitoring purposes only. The APACs/Intellution/IGH setup controls all the pumps: the turbo and rotary (roughing) pumps on both the Inner Vacuum Can (IVC) and the Outer Vacuum Can (OVC), the Roots Blower and the Pfeiffer (backing) pumps in the circulation system, and the rotary pump on the 1 K pot. The APACs/Intellution also track the levels of cryogens in the cryostat, the pressure gauges on the IVC, OVC , 1 K pot, and other lines, the flow rates in the LHe and LN exhaust lines, the temperatures across the cryogen-transfer system, they hand-shake with the data-acquisition system etc. Finally, all thermometry in the cryostat is read out using the LR700 (from Linear Research) with an internal multiplexer. Eventually, these readings were also supplied to the Intellution PC.

7.5 Cryogenics: Initial Problems

By December 2001, the dilution refrigerator and the Icebox have been assembled, but not attached together. The Icebox had not been cooled down, and the fridge was cooled to 4 K, but attempts to cool it further failed. Furthermore, there was an incident in which the LHe bath was pumped while the outer vacuum can (OVC) was at 1 atm pressure, which caused significant crumpling of the upper bath wall - Figure 7.2 schematically shows the LHe bath structure, along with the crumple zone. At that time we started a more regular rotation of physicists, with the goal of dealing with the various cryogenic problems at Soudan.

In early 2002, the fridge was indeed cooled down to below 150 mK. However, a number of problems started to appear. Probably the largest problem was a clear leak from

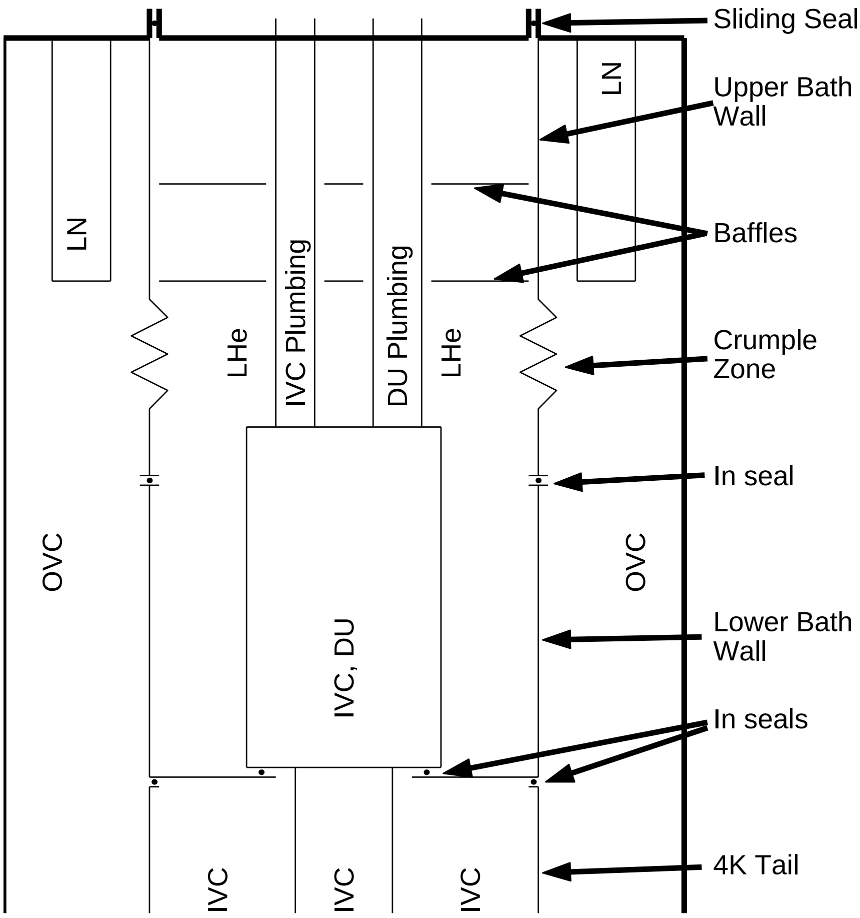


Figure 7.2: Schematic of the LHe bath in the cryostat (not drawn to scale).

the LHe bath into the OVC. Although it was usually observable at room temperature, this leak was different in magnitude from cooldown-to-cooldown and occasionally it would cause catastrophic failures of the fridge by thermally shorting the inner layers of the fridge to the room temperature (RT) wall. Such catastrophic failures were indeed quite dramatic: within a couple of minutes, the whole LHe bath would boil away, the mix would rush out of the fridge and into the mixture dump, pressures would dramatically increase etc. We have performed a test in which the LHe bath is filled with LN, which is then blown out slowly using He gas, while monitoring the leak rate in the OVC. We repeated this test many times, and every time the result was the same - the leak would start appearing when the level of LN drops below the In seal (usually by a few inches) between the upper and the lower bath walls. We also measured the profile of the upper bath flange, and found that it has a shape of the potato chip, most likely caused by the crumpling. Hence, we suspected that the problem was at the In seal, although we could not rule out the possibility that there was a crack in the upper bath wall caused by the crumpling. To make things harder, we could not get a good look at the crumple zone without pulling out the insert (containing the dilution unit), which in itself was risky because we did not know how large the crumpling was (i.e. it could have been large enough that the insert could not be pulled out). And, of course, the leaks in the external plumbing did not make the diagnostics any easier. We decided to keep trying to run, with the hope that we would learn more about the problem. In particular, we experimented with a new lower bath wall, but without success.

Another problem was in the circulation loop itself. Essentially, the pressures in the loop after the circulation pumps were oscillating between 150 and 300 mbar, and the LHe cold trap would occasionally get plugged. This indicated a possible air leak in the circulation loop, or hydrogen gas from the pump oil.

Yet another problem was with the inner vacuum can (IVC). By the end of April 2002 we were convinced that there was a He film developing on the walls of IVC (both by observing the He leak rate in the IVC and by observing the heat load on the inner stages of the fridge). However, we could not find such a leak at RT, neither from the bath nor from introducing He gas into the dilution unit. Moreover, by pumping on the He bath we could produce the superfluid conditions for the He in the bath, which would significantly increase the leak rate of He into IVC, if such leak existed - we could observe no increase in the leak rate with such test. This left the possibilities of the OVC-IVC leak, air-to-IVC leak, or a superfluid leak from the dilution unit into the IVC.

The next problem on the list was the automatic transfer system, which was unreliable at best. The Intellution software had a number of bugs, so that valves would not open when they should etc. It took a long time to sort such issues out, partly because the software was updated/modified remotely. Moreover, the dewars themselves, especially those for LN, were very different from each other in terms of their pressures, relief valves etc. It was simply hard to make the system perform well with such variations, and, again, it took time until we managed to get consistent dewars from Praxair. Furthermore, the transfer lines were made very long, so relatively high pressure was needed to precool them. This, in turn, caused other problems (e.g. the top of the fridge would get very cold ~ -100 C). Eventually, significant improvements were made by introducing double-jacketed vacuum transfer lines.

Finally, there were many other smaller problems, such as the He gas back-diffusing through the OVC pumps, leaky leak-detectors, leaks in the IVC/OVC plumbing outside the cryostat, virtual leaks in the various gauges, leaky valves in the IGH etc. We also discovered a leak from LN shield into the OVC, but since this was not a He leak, it was not a serious issue. Despite all the problems, however, by the end of April 2002, we managed to keep the fridge at the base temperature of about 24 mK for five days.

We also noticed that the dilution unit was not centered with respect to the IVC.

With help from Oxford Instruments, we established that the Still pumping line imposes a large force on the insert, causing it to move a bit. We installed a new bellows for the Still line - unfortunately this only made things worse. The sliding seal, which is meant to close the LHe bath by a set of o-rings between the top plate and the insert (see Figure 7.2), could not be closed any more, so one could observe a large plume of He gas coming out from the LHe bath at the top of the fridge.

Towards the end of June 2002, Paul Brink led a sequence of tests on the dilution unit showing that the leak rate in the IVC was significantly higher when the mix was condensed in the dilution unit (at 1-2 K) than when we had ^{4}He at 4 K in dilution unit (or when the dilution unit was empty). The tests were done carefully, while trying to keep the temperature profile of the fridge roughly equal. This proved that there was a small leak from the dilution unit into the IVC!

In July 2002, we attempted to cool down again, after redoing some of the In seals on the mixing chamber. However, after a couple of attempts, and another catastrophic failure on July 16, we have finally decided that we had to change our strategy.

7.6 Cryogenics: Turning Point

On July 17, 2002, we successfully pulled the insert out of the Oxford dewar. This event marks a turning point in the Soudan commissioning phase, because we finally had a feeling that we are making progress. We also finally had a closer look at the crumple zone - the ordinarily cylindrical surface was transformed into a fairly regular hexagon, with deviations from the cylindrical shape as large as a couple of centimeters at places. The further collapse of the bath was probably stopped by the baffles.

Many problems still remained, though. Regarding the problem of the bath-OVC leak (and crumple zone), we decided to fabricate a new upper bath wall, that would then have to be welded to the top plate in place of the existing one. This is a non-trivial procedure, as it is very easy to change the alignment of the 4 K can. In parallel with this effort, Prof. Blas Cabrera pursued the option of straightening the existing upper bath wall.

Regarding the problem of the dilution unit-IVC leak, we decided to try running the dilution unit in a different dewar. This approach would allow us to separate the two major problems and work on them independently. For this purpose, a separate dewar was shipped from Stanford University to Soudan. Since the dilution unit leak was so small, we were afraid that we may not be able to find it and fix it. Hence, we pursued yet another path: we borrowed another dilution unit from the Princeton University, which was similar (although not identical) to the existing one. This dilution unit was shipped to Stanford and was to be tested for leaks in a special dewar setup built by Betty Young. The test was successful, so this dilution unit was indeed a viable back-up option.

Finally, as another backup option, we started considering moving the SUF setup (or parts of it) to Soudan, either in place of, or in addition to, the existing setup. This was particularly because we feared that our efforts to fix the cryostat may not be successful. More to the point, even if we managed to weld a new upper bath wall, there was a danger that we would affect the alignment of the cans, which would make it difficult to match them with the Icebox cans. Furthermore, the first looks at the alignment of the Icebox cans were not encouraging. Nevertheless, the procedure of moving the SUF setup to Soudan is also very complex and it has its own risks. For the time being, however, we decided to get more information regarding the alignment of the existing Icebox in Soudan.

By the end of August 2002, we have made several breakthroughs. First, at the end of July 2002, the upper wall of the bath was straightened (first with a rubber mallet, and then a polyethylene close-fitting plug was tapped through the entire length of the upper bath). The potato-chip distortion in the bath flange was removed in this process, and the Oxford dewar passed the LN test with a blank-off plate.

Second, the dilution unit was inserted into the newly arrived Stanford dewar. After an episode where the OVC of the Stanford dewar was punctured and blew all the cryogens, the dilution unit was finally cooled down (after redoing some In seals on the mixing chamber) and on August 22, 2002, it passed the superfluid test!

Third, careful measurements of the positions of tails on the Icebox were made (led by Profs. Bernard Sadoulet and Blas Cabrera). While some problems were discovered (will be discussed later), on August 22, 2002, the Oxford dewar OVC and LHe bath were successfully mounted on the Icebox. In other words, there were no major problems with the Icebox alignment!

Fourth, the lower bath wall and a blank-off plate was installed on the Oxford dewar, and another attempt to cool down was made. This time, there was again a leak from the LHe bath to the OVC! And, the LN test revealed that the leak is at the same place as usual - a few inches below the In seal between the upper and the lower bath walls. After disassembling the lower bath wall, a likely cause of the bath-to-OVC leak was found: the upper-bath lower weld protrudes into the indium seal region. Such lump in the weld could easily introduce a mid-flange leak, whose magnitude would depend on how exactly the In seal was made!

So, at the end of August 2002, things looked very different indeed! Needless to say, the weld lump was removed, the lower bath wall was reinstalled, and the system passed the LN test. The insert was then placed back into the Oxford dewar. Another important improvement was fixing the sliding seal leak. We replaced the sliding seal collar (as it was bent out-of-round) and introduced a T-bellows configuration to dampen the force exerted by the Still plumbing line. We also installed the LN cooling loop and fixed the 1 K Pot siphon line, which was bent out of shape. By October 11, 2002, the Oxford fridge reached the base temperature of 11 mK and remained stable for more than a week, after which it was intentionally warmed up!

7.7 Cryogenics: The First Icebox Cool-down

As mentioned above, after the careful studies in July and August, no serious problems were found with the Icebox alignment. By adjusting the Kevlar strings it was possible to make all the cans move freely inside each other. The only exception was the LN-OVC touch in the tail. At first, we believed this was due to the fact that the Icebox (i.e. the OVC can) itself was not horizontal. However, after inserting the shims and making the OVC horizontal, the LN-OVC touch was still there. The next suspect was the E-stem. Indeed, the bellows connecting LHe and LN stages in the E-stem was twisted, but disconnecting it did not remove the touch. However, if the LN can is placed at an angle, the tail becomes free! Hence, there seems to be a small design error in the LN level of the Icebox.

Of course, tilting the LN can caused the inner cans to touch. Since the Kevlar strings do not affect the sidewise position of the bottom of the can, shims were placed behind the Cu hinged joint at the bottom of the Kevlar fibers. This allows sidewise motion of the bottom of the can by half the thickness of the shim. Hence, in the end, we have introduced small tilts in the LN and LHe levels to avoid the tail touches, and have introduced shims for the inner cans to avoid the bottoms of the cans from touching the neighbor cans.

Couple of other problems worth noting: (1) there was a also rotational error in the design of the LN C-stem - this was corrected; (2) the vertical cold finger (10 mK stage) was too short in order to mate with the horizontal cold finger - a spacer was fabricated to allow mating.

In parallel with the alignment studies, work was taking place on the thermometry of the fridge and the Icebox. Many of the thermometry wires were broken, and some connectors had to be replaced. Also, the thermometry wiring outside of the cryostat was not well organized - there were many paths for different thermometers, many of the cables were not properly shielded/grounded etc. It was determined that microphonics in such cables could induce significant noise in the temperature reading. Similarly, any high-frequency filtering on the most sensitive thermometers significantly affected the measurements (although the measurements are made using a 16 Hz signal). Some of these issues were dealt with prior to the first Icebox run, and some have remained throughout the first data run.

On December 7, 2002, we started the first Icebox cool-down. Although by design, the IVC in the Icebox is sealed using a Cu gasket and raised seals on the IVC lid and the IVC can flange (to avoid radioactive materials such as In), due to small scratches on the can flange, we decided to grease the gasket. This worked well, and by December 14 2002, the Icebox 10 mK can reached 25 mK (eventually it reached 21 mK). The temperature measurement was done using the ^{60}Co source.

7.8 Cryogenics: Detector Installation and Cool-Down

In January 2003 we proceeded with the installation of the striplines (which bring the signals from detectors to room temperature) and the electronics box (Ebox) which contains 50-pin D connectors mating the striplines inside the cryostat with the outside cables. Two striplines were wired incorrectly, so special adapters were built and installed. We also started preparation for detector installation.

Since the radon level in the mine is significant ($\sim$ 250 Bq/m^3), we decided it is important to install the detectors and Towers in a radon free environment. To achieve this, an old air purge was designed, such that the detector volume in the cryostat is kept at a low radon level of 2.5 Bq/m^3. The Towers 1 and 2 were transported separately from Stanford, by a company specialized for transportation of sensitive equipment. We also installed an accelerometer on Tower 1 casing and observed that the casing did not go through accelerations larger than 4.9 g. By early March 2003, both towers and all SQUETs were installed in the Icebox and no broken channels were found! Only one SQUET failed after installation and was replaced. We have also installed thermometers on the Tower 4K floors and on the 4K stage of the Icebox, to be able to monitor heating effects as we turn on the Towers.

We also used this time to make several improvements to the plumbing outside the cryostat. We have introduced several manifolds (for leak checking, for condenser lines, for cold traps etc) which made many procedures simpler. This also introduced much flexibility into the system, so that we could easily put more pumping power on any particular part of the cryostat.

By March 21, 2003 we finally finished all installations and closed the Icebox. Very quickly we discovered that many of the connectors on the Ebox, as well as the Ebox itself, leaked. We used much Vacseal and Apiazon Q to bring the overall leak level to $< 10^{-6}$ mbar l/s, but then had to clean the Vacseal from the connector pins to allow the electric connections to be made once the cables are plugged in.

On April 8 2003, the first detector cool-down started. We observed the LHe bath-OVC leak (again!), but this time it was different - the leak seemed to be at the bottom of the bath. It was reasonably small however, $\sim 5 \times 10^{-7}$ mbar l/s at LN temperature. By April 17, 2003 we reached the base temperature and could observe first pulses in Soudan!

At this point the cryostat was relatively stable. The bath-OVC leak was manageable, but we continuously had to pump on the OVC. The IVC seemed stable, and as the IVC turbo pump was found to introduce noise (see Section 7.10), we turned it off and we valved off the IVC. The Still and Cold Plate temperatures were somewhat elevated (850 and 75 mK respectively), but they were reasonably stable. The problem with the circulation loop pressures, which existed even in the very early runs of the fridge, was still present.

A number of bugs still existed in the automated system. In particular, the algorithms of when the transfer should start, how to switch the dewars etc, were being fine-tuned. Moreover, the handshaking with the data acquisition had to be implemented: we do not want to take data while transferring, the LN level meter should not be on during the data acquisition (for noise reasons) etc.

A new problem appeared in early May: LHe dewars started blowing out overnight, causing problems with the automatic transfers. Although Blas Cabrera promptly warned us about the Taconis oscillations (when a gas-filled tube reaches from the room temperature to ~ 4 K, the gas can start oscillating with a large heat transport from room temperature to the 4 K stage), we believed that the problem was with the automatic valves, so we waisted days working on the valves without success. Eventually, we confirmed that the Taconis oscillations were indeed the culprit, and we resolved the problem by connecting the vapor side of the dewar to the transfer line in the manifold (while not transfering). This, of course, further complicated the automatic transfer algorithms.

On May 14, 2003 we had to clean the LHe cold trap as it was plugging up. The problems continued, and we had to clean it again on May 20.

On May 22, 2003 we attempted cleaning it for the third time, and this time we tried to be more thorough and heat the trap longer. In such state, a catastrophic failure of the cryostat occured: the LHe in the bath was gone in minutes, the mix rushed out of the dilution unit, the temperatures started rapidly climbing etc. At the time, we were concerned that there was a drastic failure of the IVC or the OVC, so we opened the IVC (IVC was valved off for a long time) and connected it to the OVC - the concern was that a large differential pressure between IVC and OVC would damage the bellows in the E-stem that connects the two. In the process we measured a large pressure in the IVC (about 20 mtorr)! However, gas in the IVC by itself could not explain why the LHe bath boiled away

- otherwise there would have been some frosting at the top of the fridge or at the Ebox!

Given our later experience with the bath-OVC leak, the likely scenario for this failure was as follows. Because of the way plumbing was set up, while cleaning the LHe cold trap we had to stop pumping on the OVC for a long time. As a result, He (from the bath-OVC leak) built up in the OVC and thermally shorted the 4 K stage of the cryostat with the RT wall. This would make the LHe bath boil away quickly. It would also boil away any He gas that was plated out on the inner surface of the 4 K wall (we indeed observed this gas when we opened the IVC). Such gas would short the inner stages of the cryostat to the 4 K stage (which was already shorted to RT, and hence warmer than 4 K), and hence the mix would be boiled out of the dilution unit.

In the end, this was probably the first failure of the kind we were about to encounter in the next cool-down, and it probably revealed the He gas in the IVC which could have caused problems with phonon signals (He film on the detector surface would reduce the size of the phonon signals). During the extensive leak tests following this failure, we indeed found a large air-IVC leak at one of the Fischer connectors at the top of the fridge, close to the LHe Cold Trap port. The leak was very easy to miss if one did not try to specifically test that particular connector using bags of He gas. The leak was fixed by simply tightening the connector.

We also used the opportunity to make further improvements in the plumbing (so that one could clean the LHe cold trap while pumping on the OVC!), thermometry etc.

7.9 Second Cool-down of Detectors

In early June 2003, the second cool-down of detectors started. The bath-OVC leak was larger this time, so much that a LHe transfer was necessary every few hours. Moreover, the cryostat was going through a very interesting oscillatory behavior - see Figure 7.3. Every hour (or so), the pressure and He leak rate in the OVC would increase dramatically, a large fraction of the LHe bath would boil away, and the 4 K stage of the cryostat would heat up to as high as 15-20 K. Within a few minutes, however, the system would come back to normal.

For a long time, we could not come up with a mechanism for such behavior. Eventually, Prof. Blas Cabrera proposed a mechanism, which we now believe is the likeliest description of this phenomenon. The OVC is continuously being filled with He through the bath-OVC leak. Since much of the OVC-IVC wall is at ~ 5 K, much of the He gets plated on the walls, whose surface in the Icebox is fairly large. Eventually, when there is no more room for plating (or when there is a strong disturbance), the He gas starts thermally connecting the RT stage and the 4 K stage. The plated He also evaporates, causing a large thermal load on the 4K stage, which in turn causes the bath to evaporate. The inner stages of the fridge probably get shorted to the heated 4 K stage by the remaining He gas in the IVC. The He gas in the OVC gets pumped out at this point, the various stages recover to their original temperatures, and the system starts the next cycle.

Clearly, with such mechanism, pumping stronger on the OVC is expected to improve the situation. Hence, we started pumping with both available turbo pumps on the OVC, and we improved the plumbing to use the largest diameter pipes available. We also

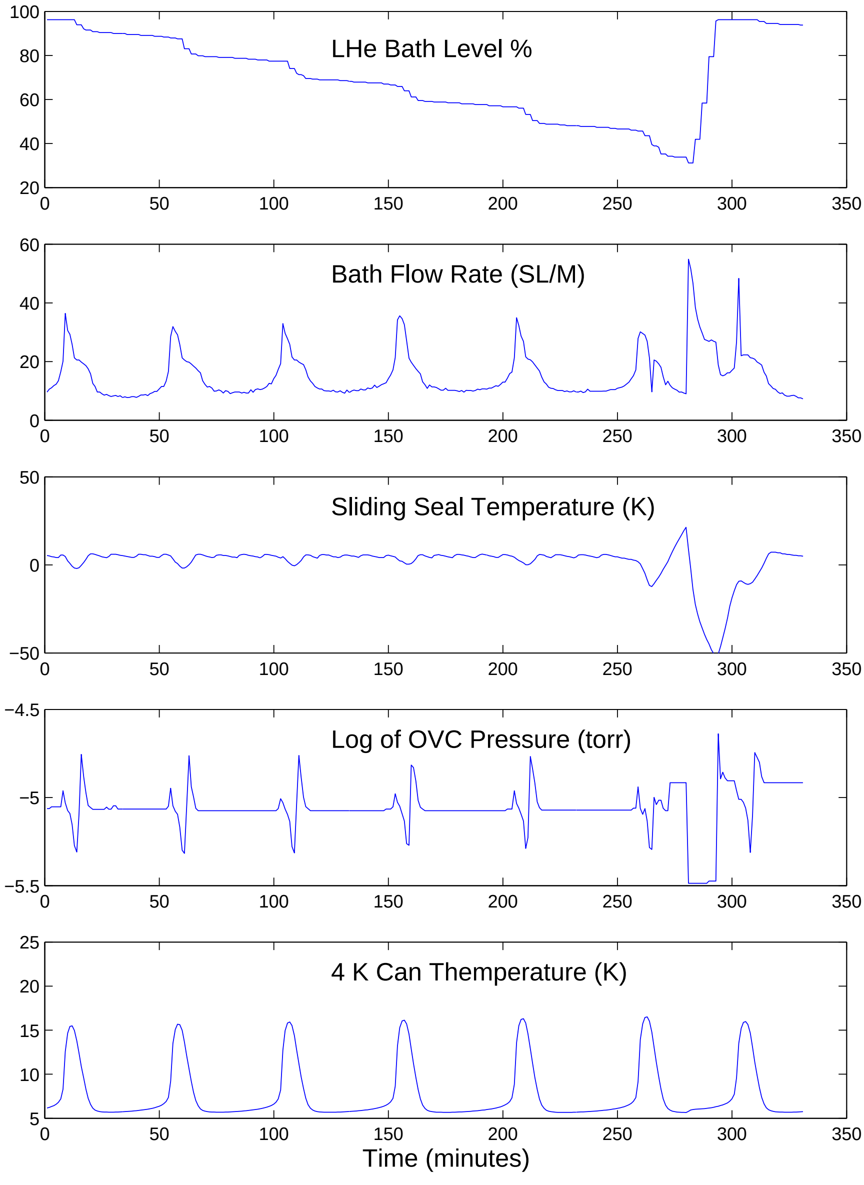


Figure 7.3: The Soudan cryostat oscillations, seen in several parameters in the system.

started pumping on both sides of the IVC (there is an IVC port on the E-stem). Such improvements reduced the number of LHe transfers to 2-3 per day, but still we had such large "burps" during most of the transfers. Ordinarily, we could run with such a problem, but the temperature excursions were so large that the SQUIDs would go through their transition with every burp. This would be a serious operational difficulty, as the SQUIDs would have to be re-tuned after every transfer, which would be very impractical, if not impossible.

One possible explanation for these transfer burps, is related to the fact that the LHe transfers were done at very high pressures. Hence, after the precooling of the transfer line is completed, the transfer line is opened to the LHe bath. One would expect a lot of "warm" He gas to rush into the bath, possibly causing some of the gas plated on the outside of the bath wall (on the OVC side) to evaporate, which in turn would start thermally shorting the RT stage and the 4 K stage, and cause a run-away. A solution, proposed by Dan Bauer, could be to do the transfers at as low pressure as possible. Indeed, with pressures of 7-8 psi, we were able to routinely transfer LHe without the burps. This was a significant success in achieving stable running.

Another serious problem was being dealt with in parallel. Namely, the LHe cold trap continued plugging up after the second detector cool-down. We found strong evidence that much of the plug was hydrogen, indicating that the problem might be in the cracking of the pump oil. We changed the oil for the circulation backing pump (the Pfeiffer pump) and started using the Fomblin oil, which is not prone to hydrogen cracking (and it comes in beautiful packages). Combined with many attempts to clean the LHe cold trap, the circulation system eventually recovered sufficiently that running was plausible. We still had to clean the trap several times during the first data run, but it was certainly manageable.

At the same time, much effort was made to improve the Intellution logic, so that the transfers are done as smoothly as possible (for example, the dewars need to be vented before the transfer starts, to avoid burps), and that the hand-shaking with the data acquisition is reliable.

In mid-August 2003, the cryostat was just beginning to be stable, but we had another crash. This time, it was due to a forgotten sinter in the dewar venting line, which did not allow the dewar pressure to drop sufficiently low for the transfer to start. Of course, this happened at night, so the transfer never started, and the LHe bath ran dry. The fridge warmed up only to 77 K, so during the next cool-down to 4 K, we tried to freeze some LN at the bottom of the LHe bath, with the hope of reducing the bath-OVC leak. The trick indeed worked well - the leak was four times smaller than prior to the crash. Hence, the pumping speed was sufficient to keep up with the leak, and we could transfer "only" twice per day.

By the end of August 2003, the base temperature was reached again, and the operation was finally stable enough that we could start working on detectors and data.

7.10 Electronic Noise

In parallel with the commissioning of the cryogenic system, work was going on to understand and improve the electronic noise performance of the system. In late January 2003, the warm electronics boards (front-end boards (FEBs) and receiver-trigger-filter (RTF) boards) were brought into the mine. In addition, special "Partly Implemented System Test" (PIST) boards were installed and connected to the rest of the warm electronics. These boards were meant to mimic the real detectors, so that noise performance of the warm electronics chain alone could be tested (before the cryostat and detectors are brought in).

This was a very useful strategy, as it allowed much improvement to be made quickly, with a simpler system. In particular, contrary to the original design, the shields of the detector cables (connecting RTF and FE boards) were originally connected both to the RF wall and to the RTF/FE boards. This setup introduced large noise because there were currents flowing through the wall. Hence, the cable shields were disconnected from the wall. Similarly, routing the power supply and sense lines so that they are close together (hence minimizing electro-magnetic pickup) was also very important. These two changes, in particular, reduced the noise pickup by 100-200 times, especially in the 60 Hz regime. In the end, the noise performance of the warm electronics alone (with PIST boards) was very close to ideal, with FEB-related peaks at 26 and 60 kHz in the charge channels. These peaks remained during the first data run as well - they did not affect our signal-to-noise or the trigger thresholds. Furthermore, a parallel setup of the warm electronics chain, for a single detector, (with a PIST board) was put together at Fermilab, with the intention of studying potential problems that show up at Soudan.

We encountered two major problems regarding the noise performance of the complete setup (including the fridge/Icebox). The first problem was that the grounding philosophy was not followed when the cryostat was assembled. The second problem was in the fact that wiring for the thermometry and other instrumentation in the cryogenic system was not done very carefully, so many different types of noise were introduced into the system.

Figure 7.4 schematically shows the grounding status of the CDMS II setup during the first data run (i.e. after all modifications discussed below were implemented). The grounding philosophy of the CDMS experiment starts with the front-end ground (FEGND) as the star ground. The front-end ground is essentially the back-plane on the FE crate. This ground is meant to be connected to the real (earth) ground via a single line, typically a very large surface conductor, such as a braid. All other grounds in the system should be connected to the FE ground, each through a single line. This includes the Icebox/fridge chasis, and the other two ground references on the FE boards - the analogue ground (AGND) and the digital ground (DGND).

The Icebox chasis ground is particularly important. All of the lines going to detectors along the Tower faces and the side-coaxes are vacuum-coaxes, the shield part of which is the Icebox chasis. Hence, any voltage drops across the fridge/Icebox could be capacitively picked up by the (very sensitive) charge amplifiers. Similarly, any currents induced into the phonon sensors could be picked up by the SQUID-based amplifiers. Hence, to avoid currents through the Icebox chasis, it is a priority to connect the Icebox to the FEGND in one place, with a good conductor.

In February 2003, the situation was very different from the one described above. All of the plumbing leaving the fridge passed through the RF wall, and it was well connected to it. The fridge stand was resting on the floor, and was separated from the steel floor by several layers of paint (which pealed of in places!). The Ebox was resting on the steel support, which was well-bolted to the steel floor. Many of the thermometry cables leaving

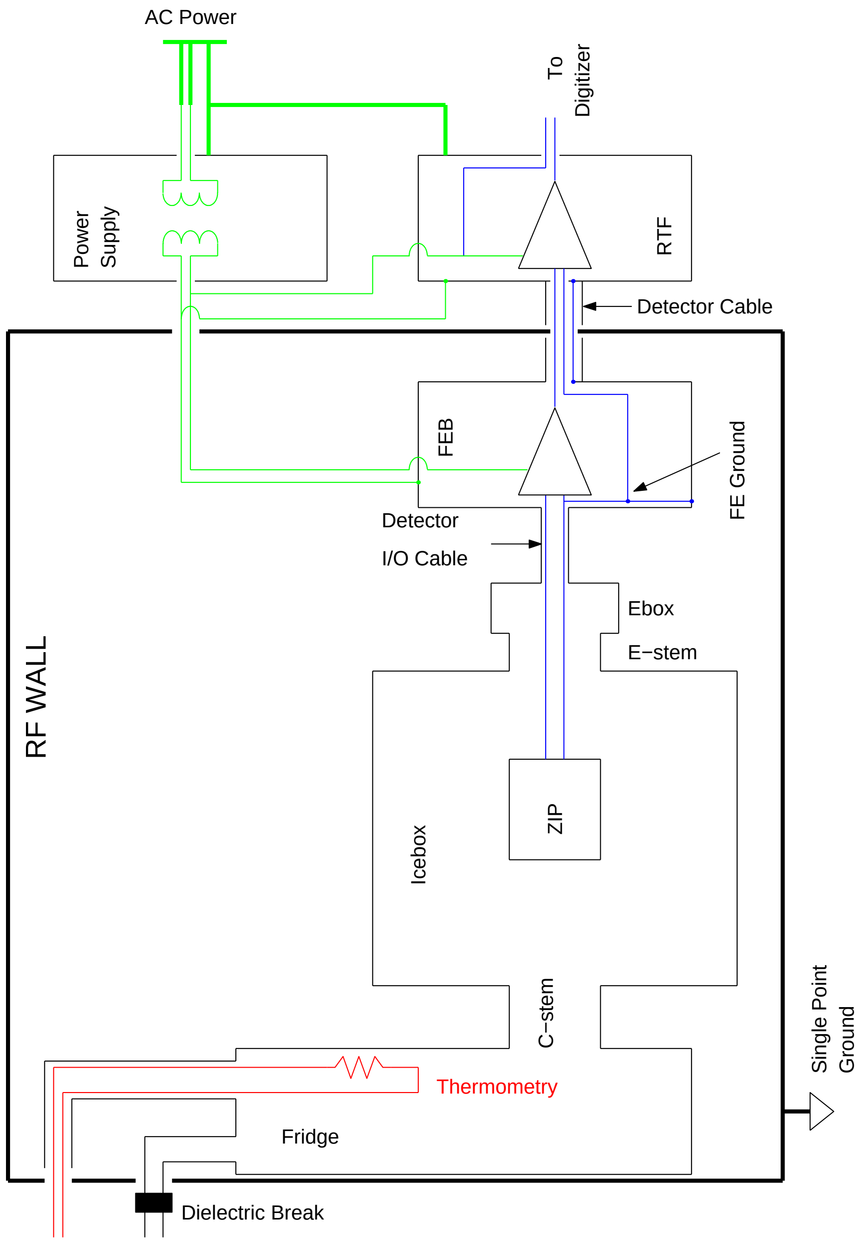


Figure 7.4: Schematic of the grounding for the CDMS II setup.

from the fridge had the ground wire connected both to the fridge chasis and to the wall (at the penetration point). Since the wall was grounded in a separate way, all these wall-fridge connections introduced many loops through which currents could flow.

Hence, all of these connections had to be broken. Dielectric breaks were introduced to all the plumbing as soon as it passed through the wall (outside of the RF room) to disconnect the fridge from the pump grounds, and layers of mylar were placed around the pipes (at penetration points) to isolate them from the wall. Layers of mylar were also slid under the fridge support stand, the Ebox was rested on the Ebox support stand using fiberglass screws, the condenser manifold was isolated from the wall with plastic stand-offs, the ground wires in thermometry cables were broken etc.

By late March 2003, the fridge/Icebox and the wall were disconnected. As mentioned above, the best way to connect the Icebox chasis to the FEGND was via a single conductor. We indeed prepared a large surface Cu sheet to establish this connection, but after much experimenting we determined that the best performance of the system (especially at 60 Hz) was achieved when all six detector I/O cables (corresponding to the six detectors of Tower 1, and connecting the Ebox connectors to the FE boards) were plugged in, with shields connected both to the Icebox chasis and to the FEGND. We did not understand this result at the time. It was determined later, that all detector cables (connecting FE and RTF boards) had ground wires connecting the AGNDs in the two boards (see Figure 7.4). This was also a violation of the grounding philosophy, as it introduced a number of ground loops. We suspect that small currents were flowing through these loops - the configuration of all six detector I/O shields connected on both sides probably distributed such currents in a way that the pick up (in the detector channels) was the smallest. The issue of the ground wires in the detector cables did not get resolved until after the first data run, and it was also responsible for some noise pickup at 1 MHz scale.

The second major problem (in addition to the violations of the grounding philosophy) had to do with the various instrumentation of the cryogenic system. We list some of the issues here.

- The LN ILM level meter produced large peaks at 16 kHz harmonics (the frequency varied a bit with the level of LN, as the coupling was likely capacitive in nature). The peaks were observed both in charge and in phonon channels, and the solution was to turn the level meter off during the data acquisition. We also noticed that some of the thermometers had higher readings when this meter was on, probably due to self-heating by the large noise.

- The IVC turbo pump controller (made by Varian) produced large 12 kHz harmonics (actually, high frequency signal with 12 kHz repetitive rate) observed in the phonon channels. This one was particularly difficult to find: it was being picked up by two thermometry cables (used for thermometers at the Icebox base stage and in the fridge itself), even if the other end of the cable was unplugged! We searched for the source using a wire-loop antenna and a spectrum analyzer but without success. At some point, we realized that the best antenna we had was the problematic thermometry cable itself, combined with one of the detectors as an amplifier. With such antenna, we found that the signal was brought into the clean/RF room through some of the pipes, after which the trace lead to the turbo pump. The problem was fixed by replacing the controller with one that was known not to have such problems.

- Cables L and E (used to read out the base stage thermometry in the Icebox and the thermometry in the fridge) introduced several discrete peaks in the phonon spectrum (L: 36, 54, and 72 kHz; E: 54 and 66 kHz). These were fixed by shielding the cables, and grounding the shields to the fridge chasis.

- Cables B and J (used to read out the Icebox thermometers at the LN stage and the 50 mK stage respectively) introduced a forest of peaks around 50 kHz in the charge channels. The peaks were variable in time, and likely introduced by the electronics in the APACs. Since this was not crucial thermometry, we decided to run without it.

- Cable T (for the 4 K thermometry in the Icebox and on the Towers) was determined to bring in a forest of large peaks at the 1 MHz scale in the charge channels. The source was traced to be a commercial module in the APACs (Moore EAM). For the first data run, we decided to run without these thermometers. The problem was fixed after the run.

- Many of the thermometry cables introduced relatively small peaks at the 1 MHz scale. This was judged to be reasonable, and it was further suppressed by a more aggressive roll-off in the RTF boards at 250 kHz.

- A high-pass filter was also added to the RTF boards with the knee at 150 Hz, in order to further improve on the 60 Hz pickup (and on detector triggers).

- Another interesting problem was the ground current in the walls of the RF room. Stan Orr identified a defective component in the line filter (mounted on the RF room wall), which caused about 700 mA of current to flow through the RF room wall! The 19^{th} harmonic was particularly large, and it was observed as ~ 1 kHz peak in the detectors. The defective component was, of course, replaced, but this took about 2 months due to a very poor response of the manufacturer.

- GPIB box (with the GPIB fiber extender), as well as the AC fans for the FE boards, were found to induce pickup in the boards. The GPIB box was moved far away, and the fan was replaced with a DC one.

A couple of items were missed and discovered after data acquisition started: (1) sliding seal heater (at the top of the fridge) introduced time-varying noise; (2) many charge channels had $1/f^2$ noise at frequencies below 10 kHz - this was tracked down to be an artifact of low frequency peaks (and their harmonics) and was significantly reduced after the first data run was finished, by improving the grounding scheme (particularly by cutting the numerous lines connecting the AGNDs in FE and RTF boards).

7.11 Shield, Veto, DAQ, and DarkPipe

In this Section we describe some of the remaining aspects of the Soudan commissioning. In particular, we briefly describe the passive shield, the veto system, the data acquisition system, and the data analysis software, DarkPipe. All of these systems have been tested before (shield, veto and DAQ on the UC Santa Barbara campus, and a somewhat older version of DarkPipe was used in previous runs at SUF), so their commissioning at Soudan went fairly smoothly.

7.11.1 Passive Shield

The passive shield was first assembled at UC Santa Barbara, and then reassembled at Soudan. The shield can be thought of as having three parts:

- The base, on which the Icebox rests.
- The annulus, around the sides of the Icebox.
- The lid, covering the top of the Icebox.

In each direction, the Icebox is surrounded by several layers. Outside-to-inside, these layers are [196]:

- 16 inches of outer polyethylene.
- 7 inches of outer (Glover and Doe Run) lead.
- 1.75 inches of inner (ancient, French) lead. This lead was obtained from a sunken ship, and hence is radioactively clean in terms of ^{210}Pb contamination.
- 4 inches of inner polyethylene (at the bottom and around the sides; 3 inches at the top).
- mu-metal shield, meant to provide low magnetic field environment inside the Icebox, particularly important for the SQUID operation.

This ordering was determined by Monte Carlo simulations to be optimal in terms of suppression of neutron background - essentially, the inner polyethylene suppresses the cosmic-ray-induced neutrons in the lead, but it is not very massive, so it does not increase the gamma flux significantly.

The lid pieces of the outer polyethylene layer were assembled in a single piece that can be easily mounted on the rest of the outer polyethylene using a 5-ton crane in the clean/RF room. Similar was done with the lead lid. This greatly simplifies assembly of the shield after the Icebox is closed and the cool-down is proceeding.

The base of the shield was completed prior to the cryostat assembly. The side walls of the shield were completed in 2002, in parallel with commissioning the cryostat. The lids were installed after the first detector cool-down started in April 2003.

7.11.2 Scintillator Muon Veto

The whole passive shield is surrounded by 40 muon veto paddles. Each paddle consists of a BICRON BC-408 plastic scintillator piece, acrylic light guide and a Hamamatsu R329-02 photo-miltiplier tube.

The scintillator paddles are cut in different sizes - most of them are 31" $\times$ 54" $\times$ 2" (top), 24.75" $\times$ 43" $\times$ 2" (bottom) and 30" $\times$ 38" $\times$ 2" (side), with a few smaller pieces to close a few cracks.

The acrylic light guide is epoxied on the edge of the plastic scintillator and the whole assembly is wrapped in mylar for light isolation. For the side paddles, the scintillator piece has an edge at an angle, so that the light guide can be epoxied at an angle. This allows the paddles to "curve" around the shield.

The PMTs are epoxied at the end of the light guide. The high voltage power supply, needed for the PMTs, is CAEN SY2527.

The veto system also includes a source of blue light, transported to each scintillator paddle with an optical fiber. Periodic pulsing of the blue light (between data acquisition runs) allows frequent calibration checks of the PMTs.

7.11.3 Data Acquisition and Data Analysis System

The data acquisition (DAQ) system was designed by the UC Santa Barbara group. This includes both the hardware and the Java/Corba software written to run the hardware and to interface with the user. Much of the system was tested in detail at UC Santa Barbara, Fermilab, Stanford and UC Berkeley, prior to installation at Soudan, so commissioning of this system was relatively smooth at Soudan. The installation was complete by the end of March 2003, and the system was ready (modulo a few bugs) to acquire data as soon as the detectors were operating in April 2003. The first data with the fully operational muon veto was taken in August 2003. Here we briefly outline how the DAQ system works.

The detector signals from the RTF boards are recorded with Struck digitizers (SIS3301). These digitizers can sample at rates up to 65 MHz sampling so we oversample to get the sample rate of 1.25 MHz. This sample rate is sufficient to capture the rising edge of the phonon signals well.

The muon veto signals branch into two. One set of signals is stretched in time using custom made "stretcher boards", and then recorded using Joerger digitizers (model VTR812). This is done because the veto signals are very short in time (ns scale), while we digitize at 1.25 MHz. The second set of veto signals goes to discriminators which generate logic step signals if the veto signals exceed a threshold (the threshold is adjusted to trigger mostly on muons). Custom boards were made that allow OR-ing the veto hits from all the paddles. The logic step-signals then proceed to the history buffer (SIS3400) where they get time-stamped. In this way, we record both the time of the veto hit and the pulse shape that can be used for determining if the veto hits were caused by gammas or muons.

From discriminators, another set of signals is taken to the scalers (Joerger VS64) that record the veto hit rates, along with detectors trigger rates etc. These are used for quick diagnostics of possible problems.

The digitized traces are then recorded in a data acquisition (DAQ) computer cluster. The cluster consists of six Linux based PCs: Builder, Tower1, VetoCrate, Monitor, Control, and Datasrv.

The traces from Struck digitizers (recording detector signal) first go to the Tower1 PC via a Struck interface (between the VME crate hosting the digitizers and the PCI slot on the PC). Similarly, the traces from Joerger digitizers (recording the stretched out veto signals) are taken to the VetoCrate PC (again, via a Struck VME interface). These two PC's then send the traces to the Builder PC, which puts the traces (along with the history buffer information) in the CDMS II event format. The three computers (Builder, Tower1, and VetoCrate) communicate via a local fast Gigabit ethernet. The Builder PC then records the events on a large fiber disk. Builder PC is also responsible for handshaking with the Intellution PC that controls the whole cryogenic system.

Datasrv PC accesses the fiber disk, records the event files on a tape, and it copies the files to a Linux-based PC (cdmsf) on the surface of the mine. The files are written to tape again on the surface. The Control and Monitor PC's, as their names suggest, control the whole data acquisition process, handle the GPIB communication with the room temperature electronics (including biasing the detectors, flashing LEDs etc), monitor the behavior of the system (such as trigger rates, pulse display etc) and handle the Java-based interface with the user. All six DAQ machines, and cdmsf on the surface, are on a local 100-base-T network allowing fast data flow (this is in addition to the fast Gigabit ethernet for the Builder-Tower1-VetoCrate). Finally, some of the DAQ machines have a second ethernet card, allowing the whole cluster to be connected to the outside world via another 100-base-T network.

The cdmsf PC is part of the analysis cluster on the surface. Hence, this is the link between the data acquisition and data analysis systems. The Soudan Analysis Cluster (SAC) includes 5 dual-processor Linux PC's, totalling to 10 analysis nodes. These machines run the standard CDMS II analysis package, known as DarkPipe, which returns a number of reduced quantities (RQs) capturing the charge and phonon pulse amplitudes, start-times, rise-times, fall-times, trigger information, veto hit information, live-time etc. Some of these RQs are not in physically meaningful units, so they are processed with "pipeCleaner" to generate new, "reduced-reduced" quantities (RRQs), which reflect the pulse amplitudes in keV, start-times in μs etc. The SAC is capable of processing 10 events per second, sufficient to keep up with the incoming data even for the largest gamma calibration datasets. For more detail on the DarkPipe and the pipeCleaner, see Section 8.3 and [2, 182].

Chapter 8

First CDMS Results at Soudan

8.1 Introduction

By October 2003, the CDMS II setup at Soudan was relatively stable. The cryogenic system was still occasionally misbehaving - the "burps" described in the previous Chapter would still occasionally show up during the cryogen transfers, but the temperature excursions could usually be contained by timely interventions. After the new controller was installed for the IVC turbo pump, the noise performance was also brought to acceptable levels. Hence, we could finally start searching for WIMPs.

This Chapter describes the results from the first WIMP-search run performed by CDMS at Soudan. In Section 2 we give an overview of the run. In Section 3, we discuss the various algorithms and corrections applied in the analysis. In Section 4, we discuss the detector performance during the run. In Section 5, we discuss the energy calibration. In Section 6, we describe all cuts used in the analysis. In Section 7, we calculate the efficiencies of all cuts. In Section 8, we discuss the various types of backgrounds and their possible misidenitification as the WIMP-signal. In Section 9, we derive the limit on the WIMP-proton elastic-scattering cross-section. In Section 10, we discuss preliminary results of an alternative analysis method. We conclude with Section 11, discussing the implications of the new CDMS result and the future expectations.

8.2 Run Overview

The first WIMP-search run started on October 11, 2003 and it lasted until January 11, 2004. We acquired 52.6 live-days of low background data with Tower 1 detectors: Z1 (Ge), Z2 (Ge), Z3 (Ge), Z4 (Si), Z5 (Ge), and Z6 (Si). The same tower, with identical detector configuration, was used in a WIMP-search run at the shallow site at Stanford University: SUF Run21. This run was described in detail in [182] and [193], and briefly described in Chapter 5. The run was interrupted by short neutron calibrations with ^{252}Cf source on November 25, 2003, December 19, 2003, and January 5, 2004. In addition, several extensive gamma calibrations with ^{133}Ba sources were performed. Most of them were taken in the first half of December 2003, with occasional earlier and later calibration runs for checks of performance. The complete listing of all data-sets, including the calibrations, is given in the Appendix.

To keep the experiment running, periodic "maintenance" was required:

- Cryogen Transfers: typically twice per day LHe had to be transferred to keep the cryostat stable. The LN transfers were done at the same time as LHe transfers (to both the LN shield in the cryostat and to the LN traps in the circulation system), although one transfer per day would have been sufficient.

- LN Level Reading: as discussed in the previous Chapter, the ILM reading the LN level in the cryostat introduced a large amount of $\sim$ 17 kHz noise in both the charge and the phonon channels of the detectors. Hence, we decided to turn it off during the data acquisition, and to take readings only during the transfers. Between the transfers, we used the LN exhaust flow rate to estimate the amount of LN left in the cryostat.

- Needle valve: Occasionally, tweaking the needle valve was necessary for the stable operation of the 1 K pot. The time scale involved here is longer than 12 hours, so any tweaking of the needle valve could also be done during the transfers.

- LED flashing: during the data acquisition, the impurity sites in the detectors are emptied and can cause degradation of the charge signal (as discussed in Section 4.2.2). To fill the impurities (and neutralize the detectors) we flash LEDs periodically, while grounding the charge electrodes. Flashing every 12 hours for $\sim$ 1 minute, then waiting for $\sim$ 45 minutes for the cryostat to cool was sufficient to provide continuously good performance of all detectors of Tower 1. Hence, LED flashing could also be done during the transfers. However, after the run was completed, we reexamined the neutralization behavior of the detectors and realized that Z6 required more frequent LED flashing. Hence, this detector was not operated optimally and it had an unusually large number of electron-recoil events with incomplete ionization signal (and low ionization yield). Furthermore, in the SUF Run21, this detector was determined to be contaminated, probably by ^{14}C source of betas. As a result, Z6 was not used in the analysis of the WIMP-search data.

Hence, all aspects of the periodic "maintenance" of the experiment were performed together, every $\sim$ 12 hours, in order to minimize the down-time. We outline below some of the other minor issues that have come up during the course of the run.

- Two small burps occured early in the run - October 14, 2003 and October 23, 2003. The base temperature did not exceed $\sim$ 500 mK, so these burps had no effect on the running of the experiment.
- Early in the run ($\sim$ October 20) the front-end boards (FEBs) for detectors Z2 and Z6 started misbehaving. The FEBs were replaced, and later analysis showed that this did not introduce an observable difference in the performance of the two detectors.
- Early in the run we observed that some phonon channels were developing large DC offsets. These offsets could potentially heat the feedback resistors (increasing their resistance) and affect the overall gain values of the phonon channels. To handle the problem, the SQUIDs were automatically relocked by the DAQ whenever the DC offsets exceeded certain threshold values.
- On October 28, 2003, large low-frequency noise appeared in the charge channels. This was eventually traced to the routing of the IVC turbo pump power cable, but it resulted in 3 days of down-time.
- On November 5, 2003, two datasets were corrupted because of a loose cable.
- On November 6, 2003, we started flushing old air around the Icebox in order to reduce the amount of radon inside the passive shield. This was motivated by the observation of high gamma rates in our detectors, which indeed dropped with the flushing.

- On November 9, 2003, phonon channel C of detector Z6 started misbehaving - it was essentially impossible to bias the sensor. This was eventually resolved by a combination of re-seating and re-powering the FEB, but the problem was never fully understood. However, we suffered 3 days of down-time.

- On November 20, 2003, the Intelligent Gas Handler (IGH) shut down (its power cord was loose). As a result, the circulation was stopped and the base temperature rose to above 1 K, so the event was followed with a period of detector neutralization and a down-time of 1-2 days. Later analysis showed that Z6 was not properly neutralized after this warm-up. This was yet another reason not to use Z6 in the WIMP search analysis.

- On November 22, one phonon channel of detector Z4 started having episodes of very large noise. The problem was eventually resolved with a better sequence of configuring the detector prior to data acquisition. However, several datasets taken until the problem was resolved suffered from anomalously high trigger rates.

- On December 11, 2003, we started triggering on multiple hits in the muon veto. This was motivated by the fact that there were too few muon-coincident events with the original triggering (on ZIP detectors), which made the study of the veto performance more difficult.

- On December 16, 2003, the cryostat went through a large burp, with the base temperature rising above 1 K. In this case, the temperature excursion was large enough that the detectors had to be neutralized again in an extended period of LED flashing, resulting in 2 days of down-time.

8.3 Analysis: Algorithms and Corrections

In this section we briefly outline the various algorithms and corrections applied in the standard CDMS analysis.

8.3.1 DarkPipe

The main analysis package used in analysis, called DarkPipe, includes several algorithms used to extract the pulse amplitude and shape information. They are:

- Optimal Filter: This algorithm fits a template pulse to the real data pulses in frequency space, with appropriate weighing by the noise spectrum. The algorithm returns both the amplitude and the start-time of the pulse, and it is applied both to the charge and to the phonon channels. Since the charge pulse-shape is determined by the electronics rather than by detector physics, the charge pulses always have the same shape. Hence, the template is easily calculated by averaging real charge pulses. The phonon pulses vary in shape, depending on the event position and the type of interaction. Hence, we use a generic template, which is not ideal but still gives very good energy resolution. For more information on how the optimal filter works, see [2].
- F5 Algorithm: This is a fit to charge pulses only, in the time domain. The algorithm yields somewhat worse resolution than the optimal filter, and it is meant to be used for saturated events - events which exceed the digitizer range. The algorithm uses

the rising and falling edges to estimate the amplitude of the pulse - this is possible because the charge pulses do not vary in shape.

- RtFt Algorithm: This algorithm starts at the maximum of the pulse and "walks" along the rising and falling edges of pulse, picking out the times of 10%, 20% etc of the maximum. Interpolations are performed if necessary. The algorithm is applied to phonon pulses only, and its output is used for studying the start-time and rise-time of the phonon pulses. Pulses are slightly filtered prior to the measurements, to minimize algorithm failures.

DarkPipe also returns a number of other reduced quantities (RQs), containing the information on trigger timing and type, timing and amplitudes of muon-veto hits, live-time, etc.

8.3.2 PipeCleaner

The reduced quantities (RQs) returned by the DarkPipe are usually not in physically meaningful units. PipeCleaner is a second-pass routine which calculates pulse amplitudes in energy units (i.e. keV), start-time and rise-time in μs etc. It also calculates the event position quantities (phonon partition and delay parameters, defined in the Section 4.4). Finally, PipeCleaner applies a correction for the cross-talk between the two charge electrodes of a given detector, as well as position corrections to both phonon and charge channels.

The phonon position correction is described in some detail in the Sections 6.3 and 6.4, so we only mention it here. The correction relies on the large gamma calibrations (of the order of 100,000 events per detector) and the event position reconstruction to smooth out the variations in the phonon amplitude and in the phonon pulse-shape parameters across the surface of the detector.

The charge position variation is a much smaller effect. It is not well understood yet, and it is possibly related to the charge electrode design. To correct it, we used the 356 keV line in the ^{133}Ba calibration as shown in Figure 8.1.

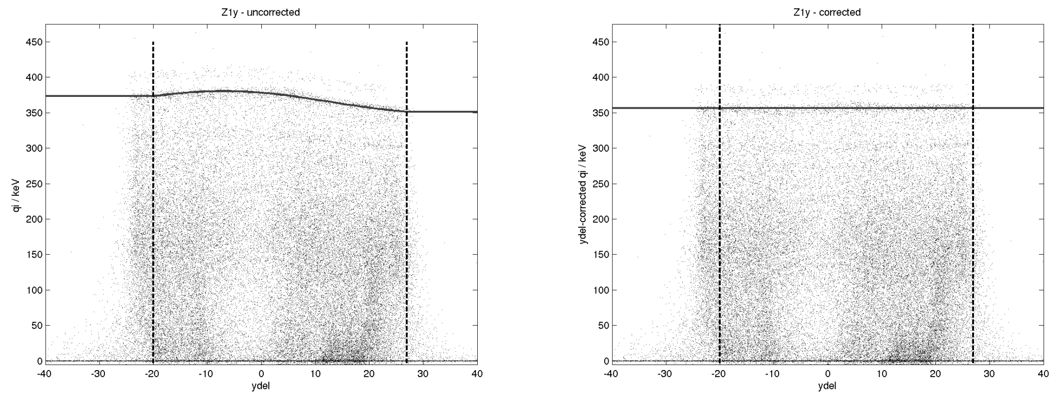


Figure 8.1: Position correction of the charge signal for Z1. The amplitude in the inner (disc) electrode is plotted against the y-position. The (roughly) horizontal line denotes the fit to the variation of the 356 keV line in the ^{133}Ba calibration. The left figure shows the situation before the correction and the right after the correction. Figure taken from [197].

8.4 Detector Performance

In this Section we discuss various aspects of the detector performance.

8.4.1 Detector Biasing

We used -3 V bias for the charge channels of Ge detectors and -4 V for the Si detectors. The -3 V bias was determined to be slightly better than +3 V or ±6 V bias in the studies of the beta calibration of a Ge detector at UCB (see Chapter 6). For the Si detectors, we decided to use a slightly higher bias as their charge performance is typically worse than for Ge due to the larger number of impurities.

The SQUIDs were biased following the discussion in Section 3.6, in order to avoid various resonances. The voltage bias values of the TESs were determined iteratively. In the first step, the bias of each sensor was reduced to slightly above the point were the sensor would snap into the superconducting state - this optimizes the signal-to-noise of the given sensor. In the second step, calibration data was then taken with such bias values, and analyzed. If the phonon partition and phonon delay plots (defined in the Section 4.4) revealed asymmetries or discontinuities among the four phonon channels, the bias values were adjusted and the the second step was repeated. This procedure allows balancing the sensitivity of the four phonon sensors. Although, in principle, phonon position correction would smooth out the differences among the sensors, removing the large non-uniformities from the start makes the position correction simpler. Figures 8.2 and 8.3 show the phonon partition and delay plots for all six detectors in the gamma calibration - note that there are no significant discontinuities at the edges between different sensors. Figure 8.4 shows the charge-vs-phonon plots for all six detectors for the gamma and the neutron calibrations. Note that the separation between the electron and nuclear-recoil events is good down to very low energies (we will discuss this more later).

8.4.2 Noise Performance

Figure 8.5 shows the typical noise power-spectrum densities. Note that the phonon spectra are very flat in the whole signal region (roughly 1-20 kHz), and, as expected, they

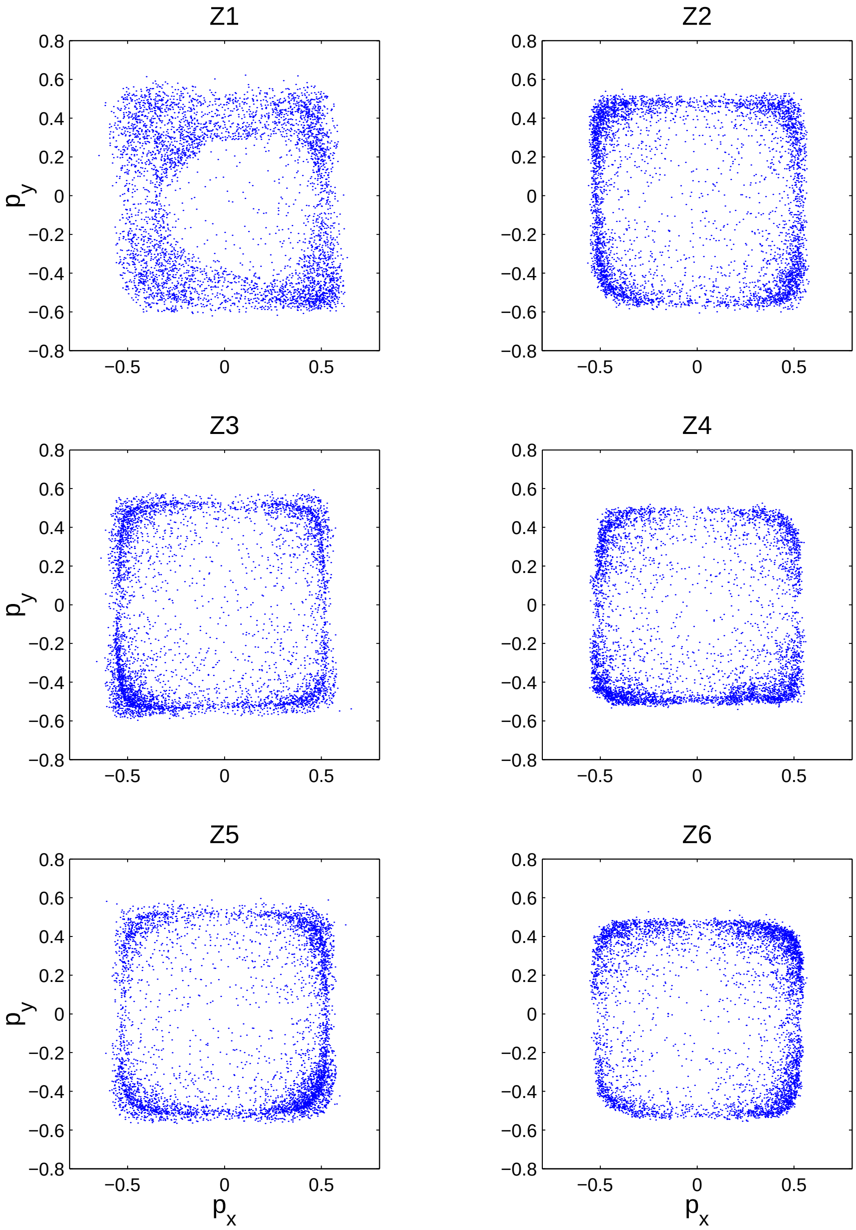


Figure 8.2: Phonon partition plots for all six detectors using 30-60 keV events in the ^{133}Ba calibrations.

fall-off above $\sim$ 50 kHz. This is typical for all detectors. Hence, the performance of the phonon channels is satisfactory.

The charge channels are not quite so simple. Most channels, such as Z3 Qinner

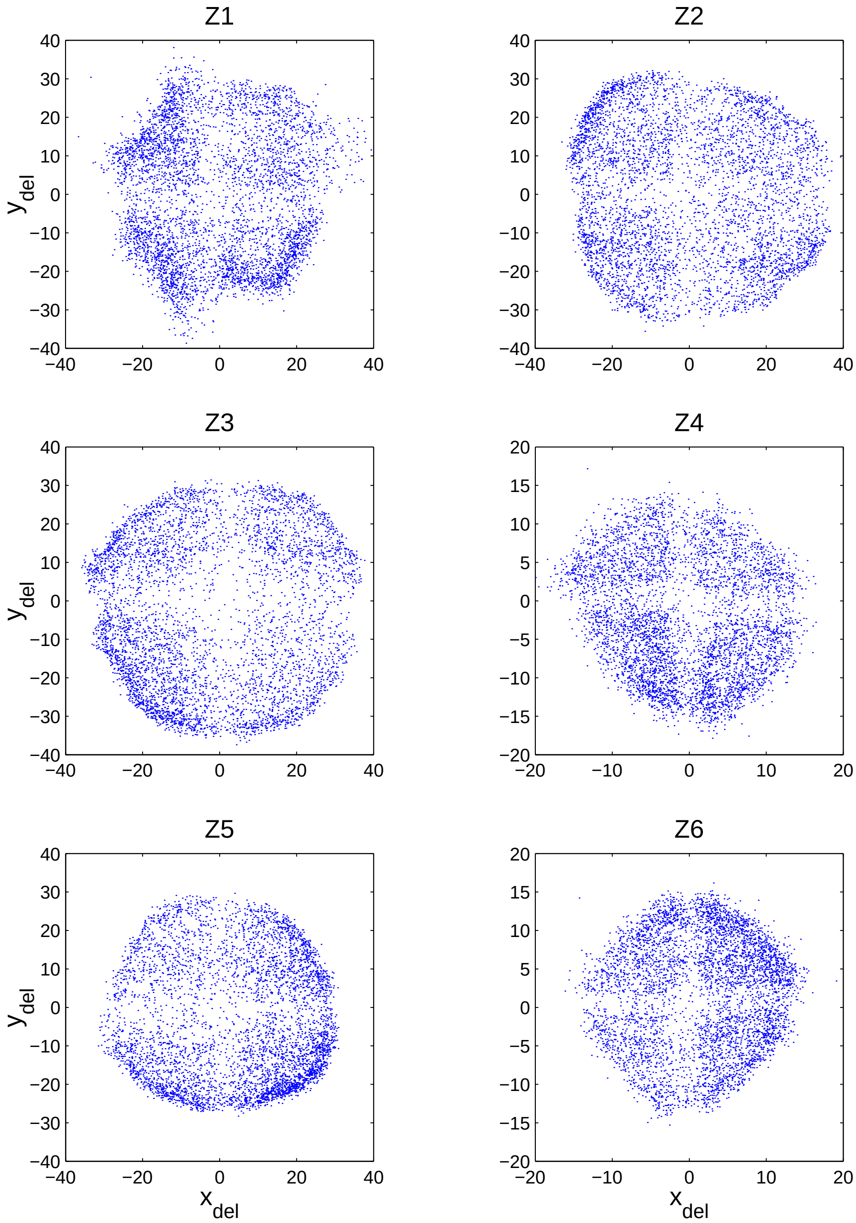


Figure 8.3: Phonon delay plots for all six detectors using 30-60 keV events in the ^{133}Ba calibrations.

displayed in Figure 8.5 are flat in the signal region (roughly 1-160 kHz), or even decrease at lower frequencies (as expected, [170]). However, most charge channels suffer from high-frequency peaks (hundreds of kHz to several MHz scale). These peaks are outside the signal

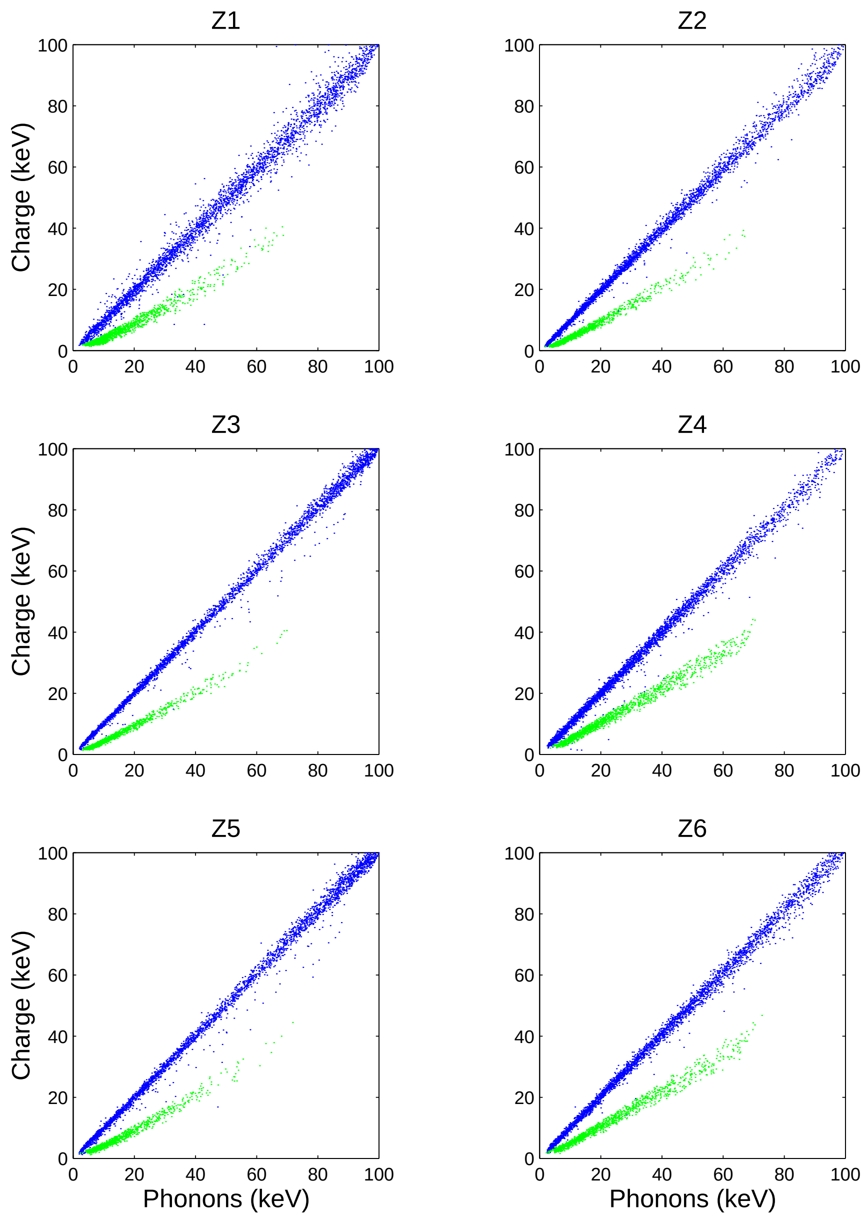


Figure 8.4: Charge energy vs phonon energy for all six detectors using events from the ^{133}Ba calibration (gammas, in dark/blue points) and from the ^{252}Cf calibration (neutrons, in gray/green points).

region, so they do not impact the signal-to-noise significantly. We have indeed verified that the signal-to-noise is similar to the values obtained in the SUF Run21, where the high-frequency end of the spectrum was clean. Furthermore, some charge channels, such as Z3

Qouter displayed in Figure 8.5, also suffer from $\sim 1/f^2$ noise at low frequencies, roughly below 10 kHz. After the run was completed, more detailed studies revealed that this was actually an artifact of a $\sim$ 150 Hz peak and its higher harmonics. Comparing with SUF Run21 results, the signal-to-noise for these channels is reduced 10-20%, with some variation among the channels. We decided that this was an acceptable loss, in the interest of completing the first WIMP-search run in a timely manner. After the run was completed, the grounding of the experiment was improved (in particular, by cutting the numerous lines connecting the AGNDs in the FE and RTF boards; see Chapter 7 for more detail on the grounding scheme). This resulted in a significant reduction of both the high-frequency noise peaks and the low-frequency $1/f^2$ noise.

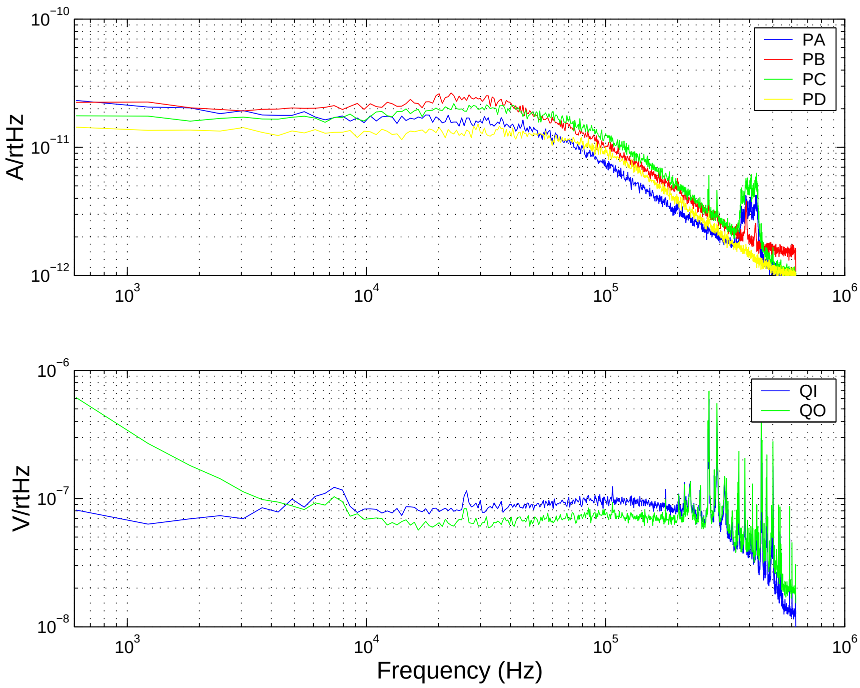


Figure 8.5: Noise power-spectrum densities for detector Z3 and for the WIMP-search dataset 131011_2015. The top plot shows the four phonon channels, referred to the input coil of the SQUID, and the bottom plot shows the two charge channels, referred to the feedback capacitor (i.e. after the first stage amplification).

8.4.3 Noise Blobs

Another way to quantify the performance of detectors is to determine the baseline resolution in the charge and phonon channels. To do this, we apply the optimal filter algorithm to pulse-less traces and study their distribution. The algorithm is applied at a fixed position in the trace, in order to avoid being biased by random noise spikes in the trace. The pulse-less traces were selected using the "random-trigger" - about 5% of the events are triggered at random times by the DAQ sofware (i.e. they are not triggered by the detectors or by muon-veto), for purposes of monitoring the noise performance during the run.

Figure 8.6 shows the "noise blobs" - the charge energy vs the phonon energy estimates for the pulse-less traces. Note that the noise performance of the phonon channels is sufficiently good to trigger on sub-keV phonon-energy events. The baseline resolution of the charge channels is somewhat worse - events with $\lesssim 2$ keV ionization energy are difficult to distinguish from the noise.

8.4.4 Phonon Trigger Efficiency

Since the phonon baseline resolution is lower than the charge baseline resolution, and since the phonon channels are less susceptible to random noise pick-up, we decided to trigger all detectors on the sum of the four phonon channels. The efficiency of the phonon trigger is shown in Figure 8.7. The trigger efficiency for a detector D is estimated by preselecting events whose global trigger (that prompted the digitizers to record) was NOT caused by the detector D. Of these events, we calculate the fraction that had a phonon

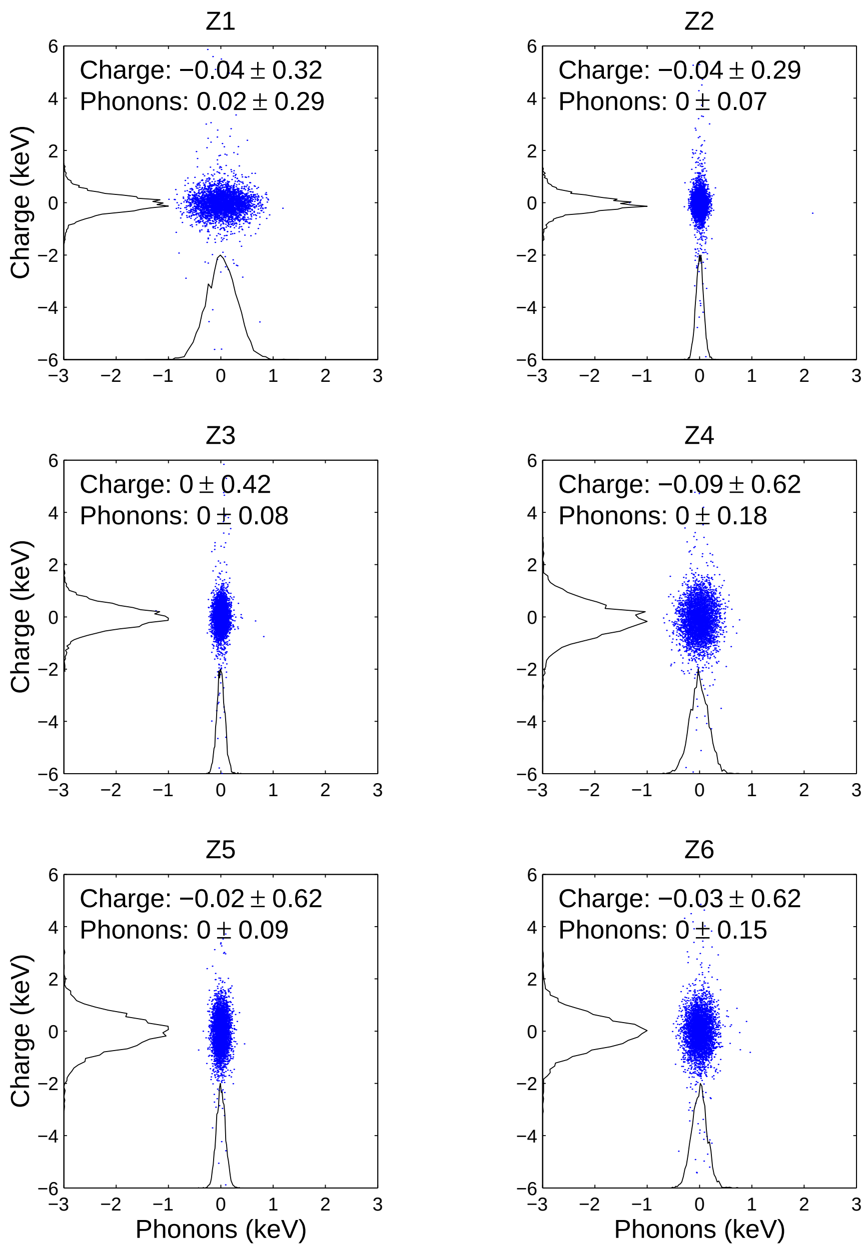


Figure 8.6: Noise Blobs: charge energy vs phonon energy estimates for the pulse-less traces. The mean and the sigma of a gaussian fit for both the charge and the phonon distributions are displayed in the Figure.

trigger in detector D recorded within 100 μs after the global trigger, as a function of the phonon energy in the detector D (this is possible to do since the history buffer keeps the time-stamps of all triggers, and not just of the global trigger). The estimate was done using all of the WIMP search data. Note that, for all detectors, the trigger efficiency reaches 100% below 10 keV and for some detectors even below 5 keV. Also, note the correlation of the trigger efficiencies and the baseline phonon resolution (Figure 8.6) among the detectors. To achieve these efficiencies, the trigger thresholds were set low, such that $\sim 30\%$ of the triggers were caused by noise.

8.4.5 Diagnostics and Long-Term Trends

Substantial effort was put into monitoring the long-term stability of the detector performance. For this purpose, an extensive off-line diagnostic package was written. In addition to the standard diagnostic plots (phonon partition and delay plots, noise spectra etc), this package performs a number of tests and makes long-term trend plots for various parameters. In particular, this package:

- Examines the baseline variation for each channel of each detector.
- Examines the standard deviation of the pre-trigger part of the traces (~ 400 μs), for each channel of each detector.
- Calculates χ^2 for the noise power-spectrum density (PSD), for each channel of each detector, using a set of template PSDs (the first WIMP-search data-set was used to calculate the templates).
- Examines the signal-to-noise ratio, for each channel of each detector, calculated using the optimal-filter algorithm and the set of pulse templates determined as described in Section 8.3.1.

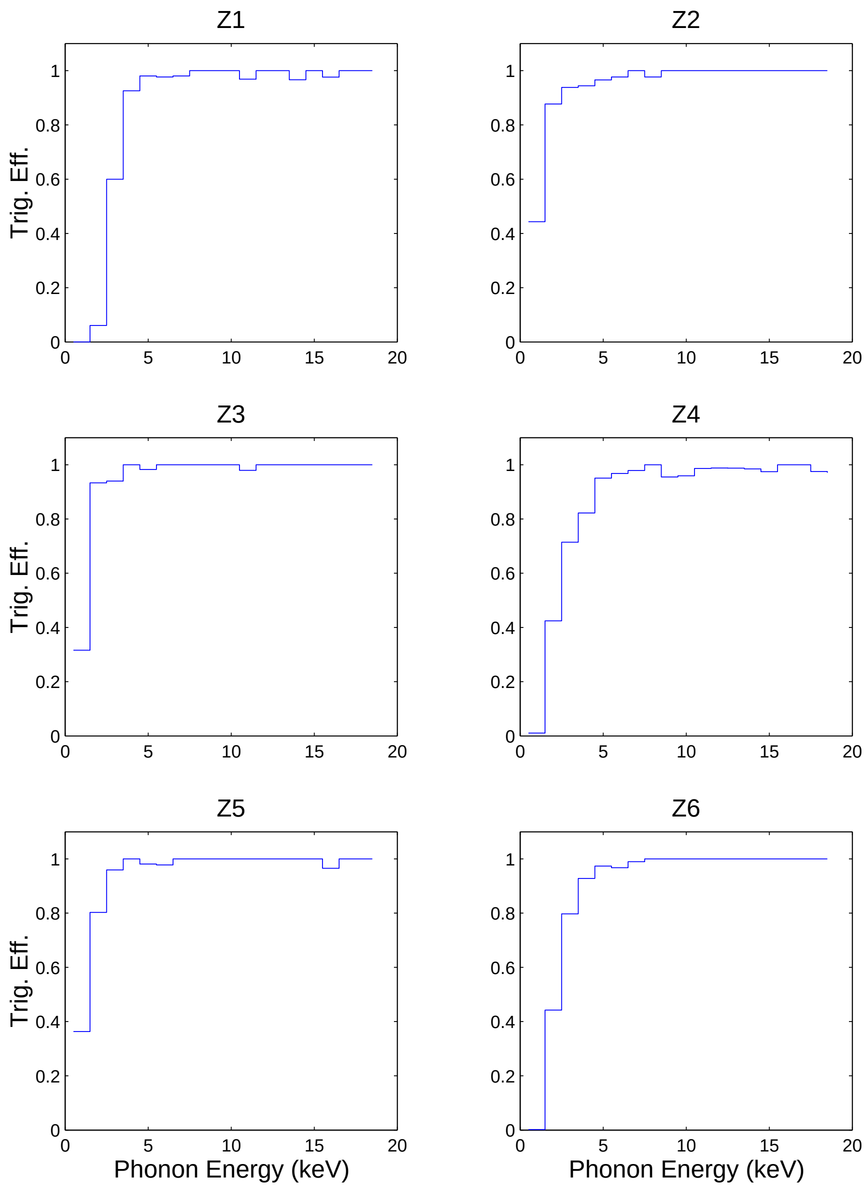


Figure 8.7: Phonon Trigger Efficiencies for all six Tower 1 detectors. See text for more detail.

- Examines the mean and the standard deviation of the noise blobs, for both the charge and the phonon signals.

- Performs a number of 1D and 2D Kolmogorov-Smirnov tests and examines their trends. In particular, the 2D KS tests are performed on the phonon partition plots, phonon delay plots, and yield vs recoil-energy plots. The 1D KS tests are performed on phonon and charge spectra etc. The templates used in all these tests were calculated from the first WIMP-search data-set.

As examples, we show the trends of some of the parameters discussed above for the detector Z3 in Figures 8.8, 8.9, and 8.10.

Several comments are in order:

- The trends showed in Figures 8.8, 8.9, and 8.10 reveal that the charge noise spectra change over time. This can be observed in the trends of the pre-trigger standard deviation or of the PSD χ^2 (these two parameters are clearly correlated). A more careful examination showed that the variation was due to the high-frequency structure of peaks (at the 0.5-1 MHz scale) shifting to somewhat lower frequencies.

- The mean and the standard deviation of the charge component of the noise blobs look quite stable over time and they are not correlated with the changes in the charge noise spectra. This is because the changes in the noise spectra are outside of the signal region, so they do not affect the resolution significantly. Furthermore, the signal-to-noise parameter is also very stable over time and it is not correlated with the changes in the noise spectra. Hence, we conclude that the high-frequency variations in the noise spectra do not impact the resolution of the charge signal.

- Some of the phonon channels also show discrete jumps in the PSD χ^2, although to a much smaller extent than those observed in the charge channels. Many of these jumps are likely due to small operational changes (such as recycling the power in the

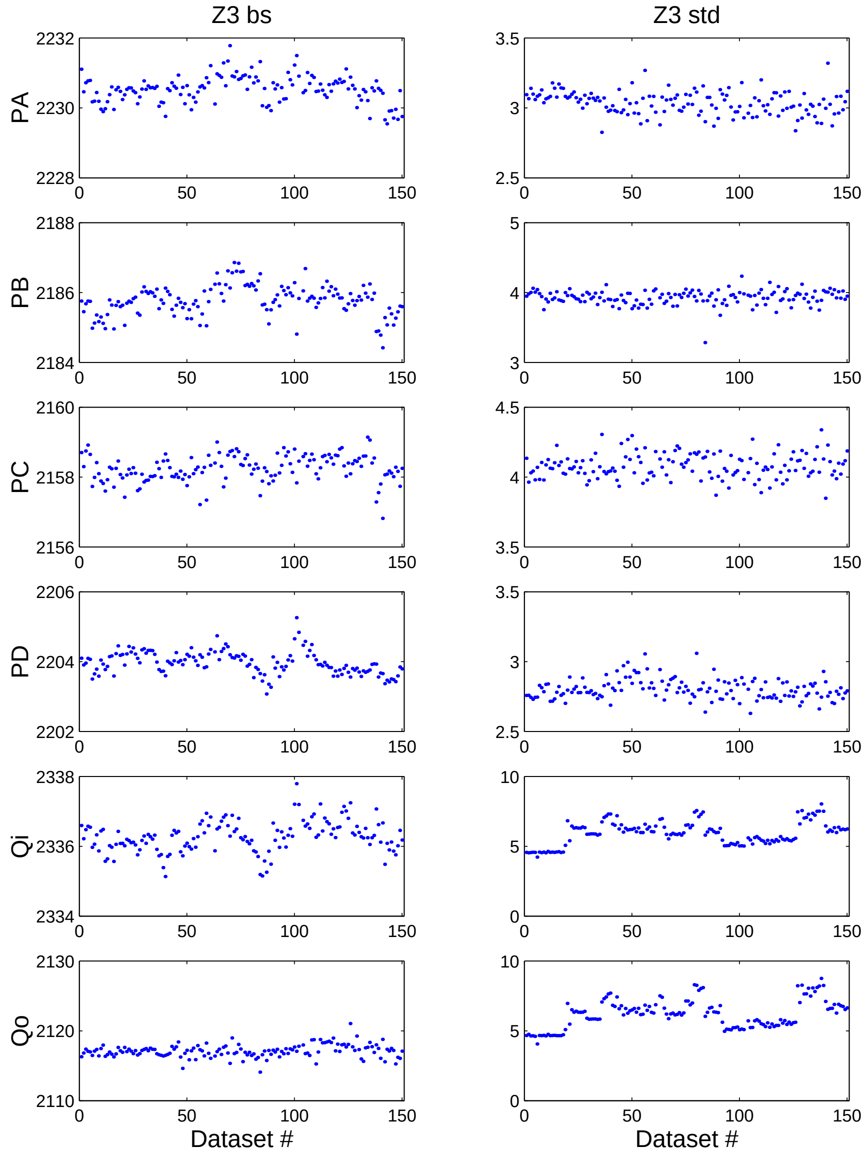


Figure 8.8: Variation of the baseline (bs) and the pre-trigger standard deviation (std) for all channels of Z3, across all WIMP-search data-sets.

electronics, reseating the boards etc), and they do not have any impact on the signal-to-noise or on the noise blob mean and standard deviation. Hence, these changes do not impact the resolution of the phonon signal.

- The mean and the standard deviation of both the charge and the phonon components

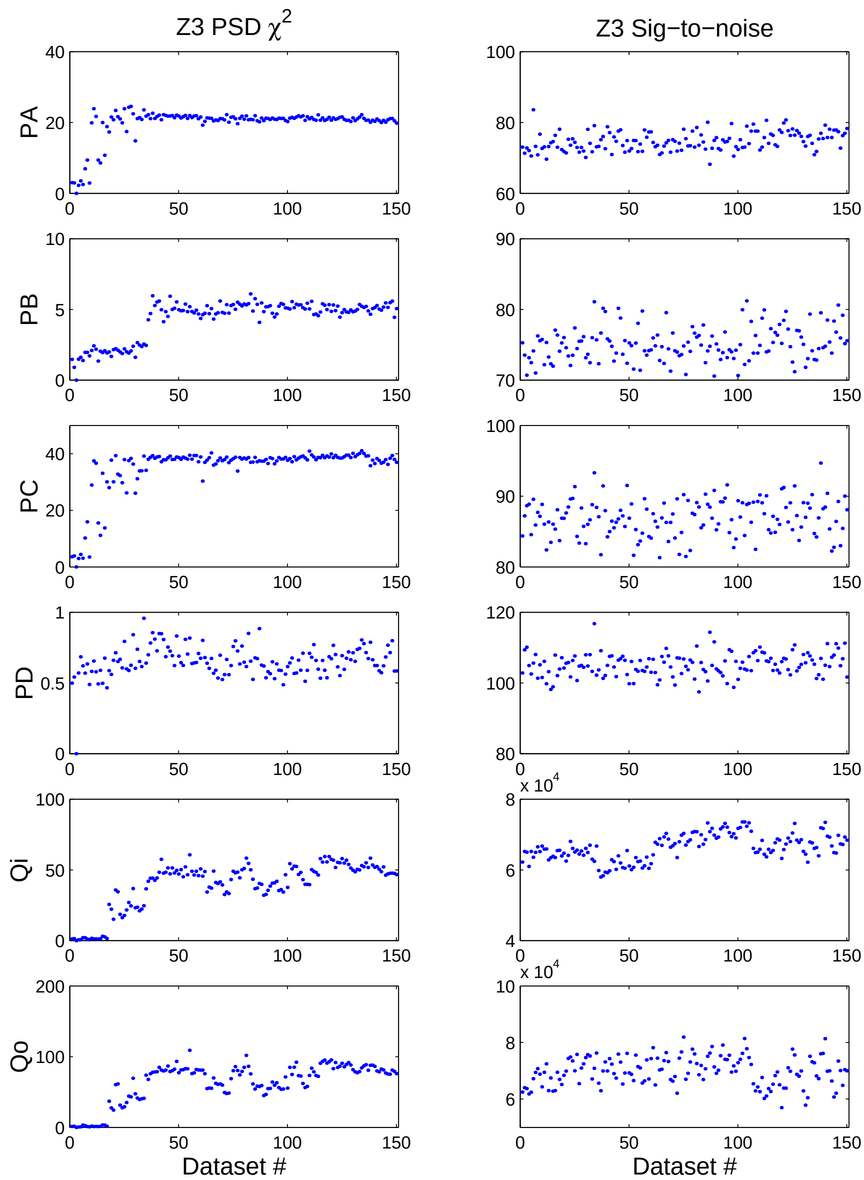


Figure 8.9: Variation of the PSD χ^2 and the signal-to-noise for all channels of Z3, across all WIMP-search data-sets.

of the noise blobs are stable over time. Similarly, the signal-to-noise is stable for all channels. We conclude that the resolution in both phonon and charge signals is stable over time.

- The trends of the significances of the 2D KS tests on the phonon partition and delay

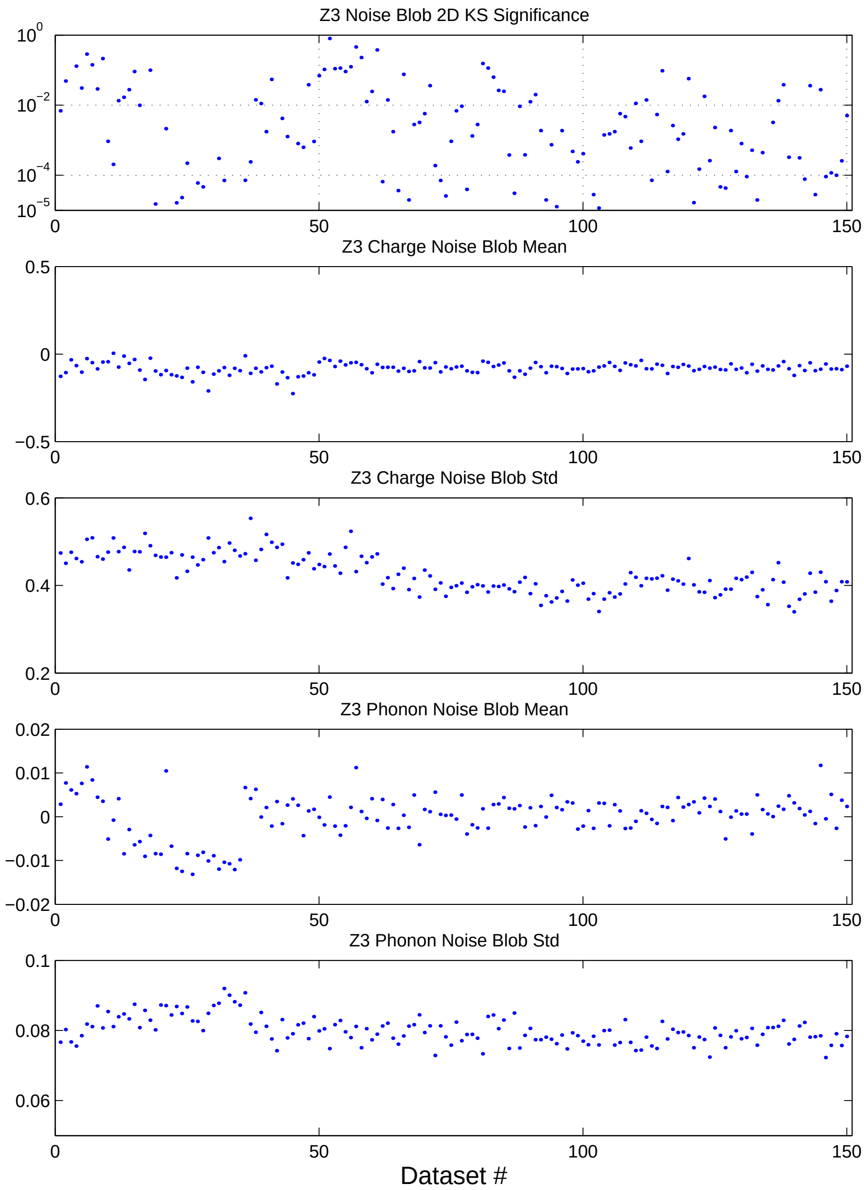


Figure 8.10: Variation of the mean and the standard deviation of the charge and phonon components of the noise blob, and of the 2D KS test statistic for the noise blob of Z3, across all WIMP-search data-sets.

plots, and on the yield vs recoil-energy plots are also stable over time. Similar holds for the 1D KS test significances on the phonon and charge spectra, implying that the overall gains of the phonon and charge signals were reasonably stable over time.

- The optimal-filter algorithm returns the χ^2 parameter for the charge fit, computed in the frequency domain. This parameter is used for rejecting the pile-up events, as well as events triggered by noise spikes. We have observed a slight variation over time in this parameter, probably due to the variations in the noise spectra of the charge channels. As noted above, no such variation was found in parameters that describe the resolution or the overall gain of the charge signal.

The comments listed above do not include occasional data-sets that were corrupted. In fact, the diagnostic routines described here were very useful in identifying these corrupted data-sets and in determining the "good data" cuts. These will be described in more detail below.

8.5 Energy Calibration

The energy calibration was performed using a ^{133}Ba gamma source. To avoid the Pb in the passive shield, the source was inserted through a special tube along the E-stem of the Icebox, which penetrates the passive shield. This particular isotope offers several distinct lines: 276 keV, 303 keV, 356 keV, and 384 keV. These lines are sufficintly energetic that the photons can punch through the copper cans of the cryostat and reach the detectors. However, only charge channels have linear response in this energy region. The phonon channels typically start saturating (and become non-linear) above 200 keV. Hence, we use these lines to calibrate the charge channels, and then calibrate the phonon channels against the charge.

Detailed Monte Carlo simulation of the ^{133}Ba calibration was performed using

Geant 3 [198]. The results of the simulation were convolved with the energy-dependent resolution of the charge channels, as estimated in [182]. The resulting prediction is compared with the actual calibration in the Figure 8.11. The lines are very clear in all Ge detectors, making the calibration straightforward. However, the penetration depth of photons in Si is larger than in Ge, which makes detection of the lines in Si detectors considerably more difficult. Hence, for the case of Si detectors, the calibration is achieved by matching the spectral shapes of the Monte Carlo simulation and the data.

For the Ge detectors, a cross-check of the energy calibration can be made using the 10.4 keV line due to the activation of Ge. The natural isotope abundances of Ge are: ^{70}Ge – 20.8%, ^{72}Ge – 27.5%, ^{73}Ge – 7.7%, ^{74}Ge – 36.3%, ^{76}Ge – 7.6% [199]. Ge can be activated by neutrons in several different ways. One possible chain is:

$$\begin{aligned}
{}^{70}\mathrm{Ge} + \mathrm{n} &\rightarrow {}^{68}\mathrm{Ge} + 3\mathrm{n} \\
{}^{68}\mathrm{Ge} &\rightarrow {}^{68}\mathrm{Ga} + \gamma(10.36\ \mathrm{keV}),\ \tau_{1/2} = 270\ \mathrm{days} \\
{}^{68}\mathrm{Ga} &\rightarrow {}^{68}\mathrm{Zn} + \gamma(9.66\ \mathrm{keV}),\ \tau_{1/2} = 67\ \mathrm{min}
\end{aligned}$$

The second reaction in this chain has a relatively long half-life. Hence, Ge activation by the cosmic rays on the surface can be important even several months after the detector is brought underground. Once underground, the dominant cause of the neutron activation is the neutron calibration by the ^{252}Cf source. A secondary cause would be neutrons produced by muons interacting in the surrounding material, but this is strongly suppressed by the rock overburden in the Soudan mine. In addition to the reaction chain given above, other

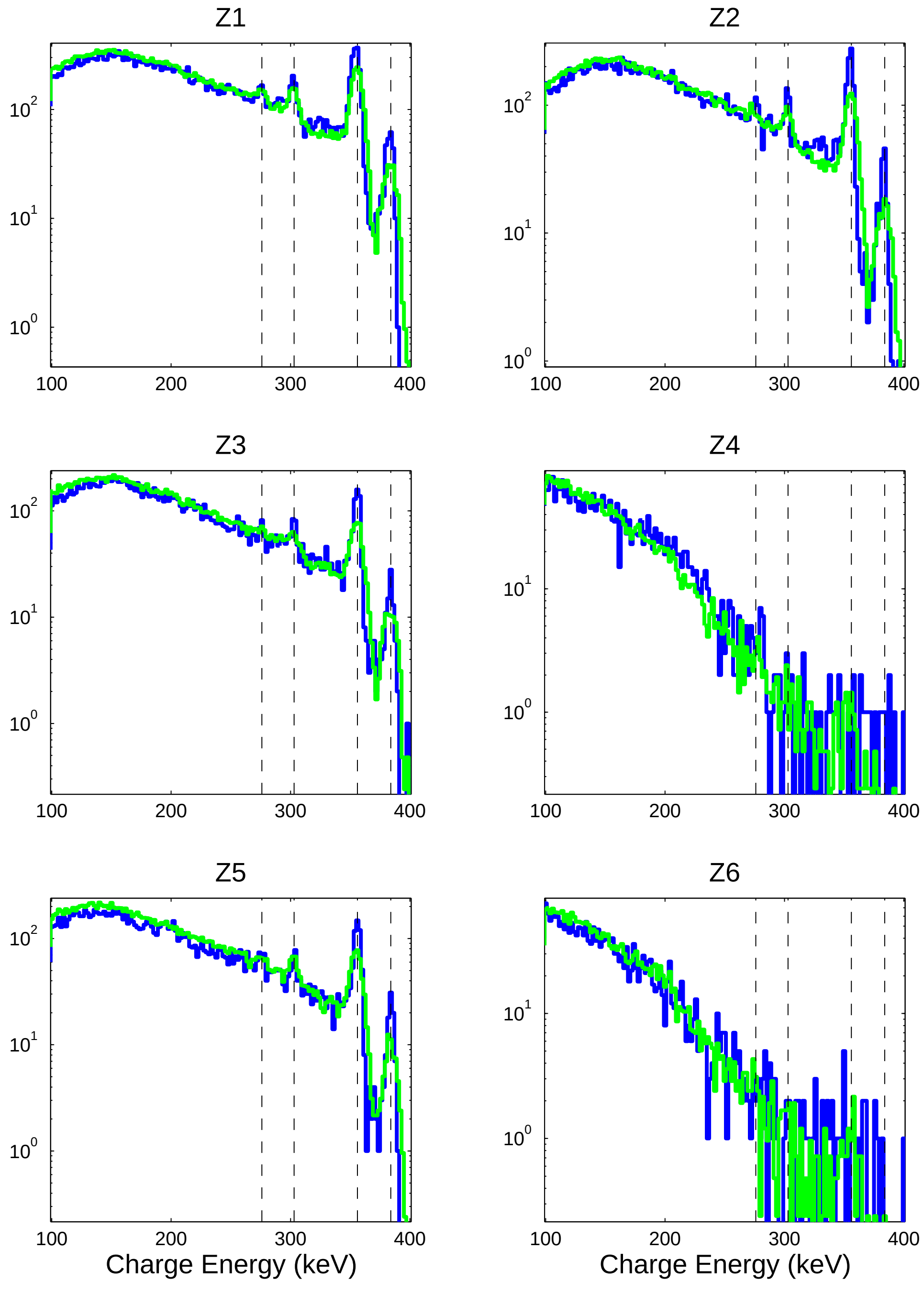


Figure 8.11: Comparison of the measured ^{133}Ba charge spectrum (dark/blue) and the Monte Carlo simulation (gray/green) for the six detectors of Tower 1. The vertical dashed lines denote the expected lines at 276 keV, 303 keV, 356 keV, and 384 keV.

possible chains are:

$$
\begin{aligned}
{}^{70}\mathrm{Ge} + \mathrm{n} &\rightarrow {}^{71}\mathrm{Ge} &&\rightarrow {}^{71}\mathrm{Ga} + \gamma(10.36\ \mathrm{keV}),\ \tau_{1/2} = 11.4\ \mathrm{days},\\
{}^{72}\mathrm{Ge} + \mathrm{n} &\rightarrow {}^{73\mathrm{m}}\mathrm{Ge} &&\rightarrow {}^{73}\mathrm{Ge} + \gamma(66.7\ \mathrm{keV}),\ \tau_{1/2} = 0.5\ \mathrm{s}.
\end{aligned}
$$

The first chain has a reasonably long half-life, so the 10.36 keV gammas can be observed. The second chain has a very short half-life. There are hints in our data of the 66.7 keV gammas corresponding to the second chain, but the statistics are too low for a reliable estimate. For more detail on the Ge activation and the corresponding lines, see [182].

Figures 8.12 and 8.13 display the position of the 10.4 keV line in the charge and phonon channels. Also shown is the resolution - for both the charge and the phonon channels, the typical resolution at 10.4 keV is about 5%, with some variation among detectors. Note that the phonon resolution of Z1 is the worst - this is expected since this detector suffers from the largest critical-temperature gradient in the phonon sensors. The resolution performance discussed here is consistent with the resolution performance of the same detectors in the SUF Run21 [182].

8.6 Blind Analysis: Cut Definitions

The first-pass analysis of the first CDMS II data from Soudan was performed *blindly*. In particular, we adopted the policy that the nuclear-recoil region of the WIMP-search data was in a "sealed box", not to be examined until cut definitions were finalized. This procedure allowed an unbiased definition of all cuts. All cuts were defined using calibration data only, *prior* to "opening the box", but their efficiencies were estimated after the box was opened. In this Section we describe all blind cuts. In the following Section, we

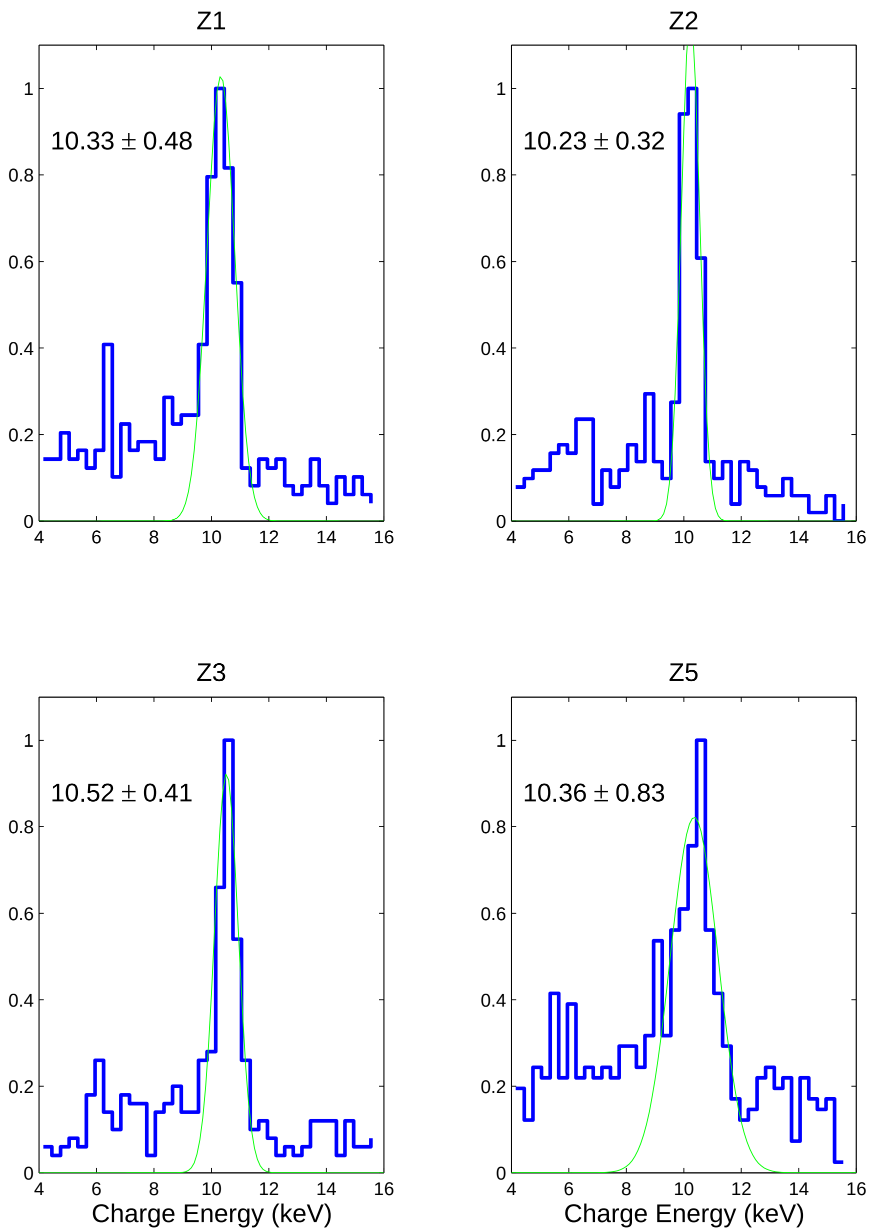


Figure 8.12: Observation of the 10.4 keV line in the charge channels of Ge detectors. Mean and standard deviation of a gaussian fit are displayed. The data histogram is in dark/blue, and the gaussian fit is in gray/green.

will estimate their efficiencies.

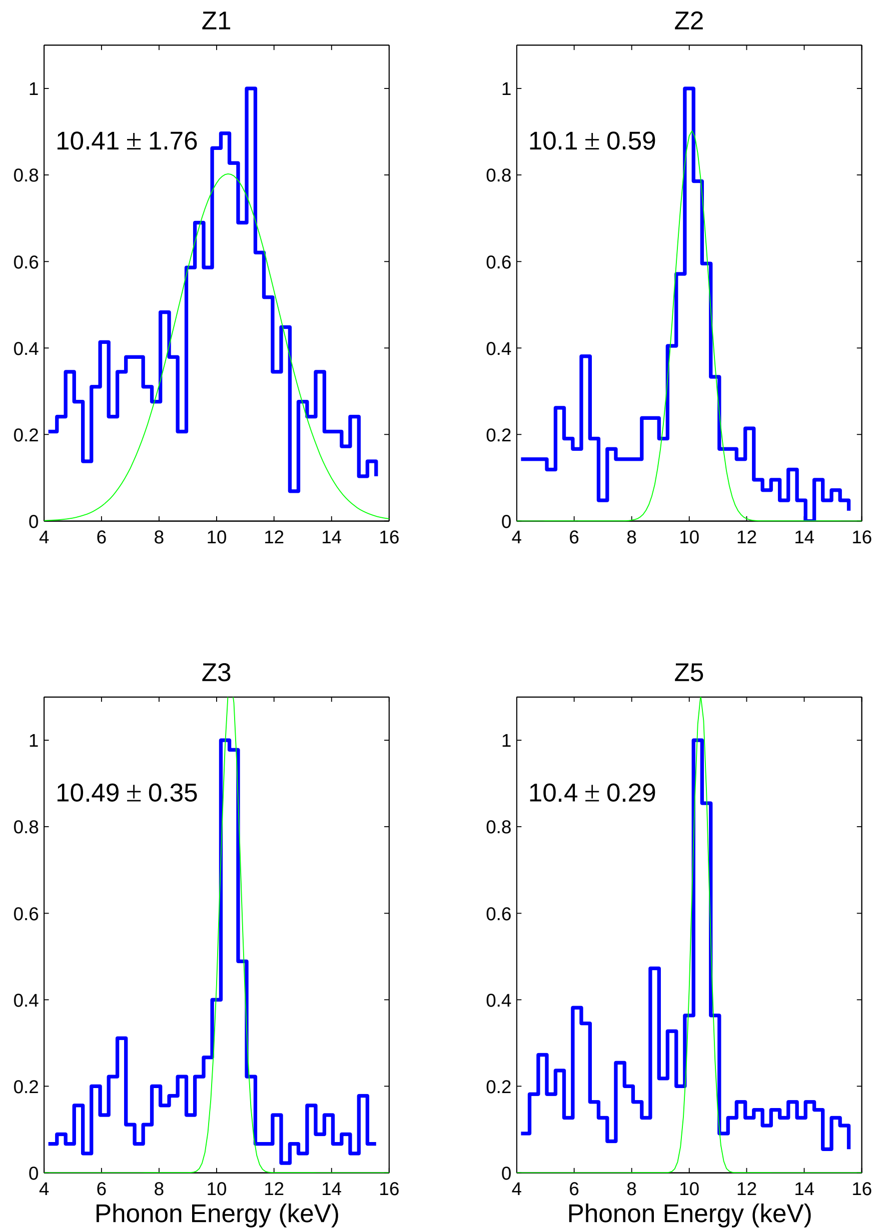


Figure 8.13: Observation of the 10.4 keV line in the phonon channels of Ge detectors. Mean and standard deviation of a gaussian fit are displayed. The data histogram is in dark/blue, and the gaussian fit is in gray/green.

8.6.1 Bad Data Cut

The complete list of WIMP-search data-sets used in this analysis is given in the Appendix. It includes all data-sets that lasted more than $\sim 1-2$ hours. Several shorter data-sets were included if they could have some special scientific interest (for example, data-sets taken right after the neutron calibrations, to study the neutron activation of Ge). Also, data-sets that had known problems (e.g. not triggering on all detectors etc) were not included in this list.

We made a detailed study of datasets on this list, primarily using the diagnostic package, logs etc, and defined a set of cuts to remove all corrupted data-sets (or parts of data-sets). We list these cuts here:

- cBadData - removes several data-sets:

  1. 131108_2202, 131109_1040, and 131109_2029 did not have operational phonon channels for Z6.
  2. 131222_0948 had bad 2D KS test significances for Z1 and Z4, for the phonon partition and delay plots, and 1D KS test significances for the phonon spectrum.
  3. 140105_1715 had a particularly bad PSD χ^2 for the phonon channel C of Z1.
  4. 131121_1412 failed processing in the diagnostics package (short data-set).
  5. 131021_2300, 131022_1739, 131102_0935, and 121116_1921 had unusual behavior of the χ^2 for the optimal-filter charge fit.

- cGlobTrig - removes a very small number of events that did not have the global trigger recorded in the history buffer.

- cQbias - removes rare events with charge bias voltage set to zero - these events took place after the DAQ unbiases detectors and before it stops acquiring the data. This error was fixed partly into the run.

- cZ2badph - removes data-set 131114_0931 which had very low significances for the 2D KS tests on the phonon partition and delay plots, and on the yield vs recoil-energy plot for Z2.

- cZ4burst - removes data-sets 131123_1128, 131127_2306, and 131128_2137. These data-sets had poor signal-to-noise for phonon channel C of Z4 - they were taken during the time period when this channel suffered from noise bursts.

- cGlitch - removes events triggered by noise glitches. These events usually triggered several detectors, so the cut was based on the high multiplicity of triggers.

- cZ5Bacalib - removes corrupted events in one ^{133}Ba calibration data-set (131211_0920). The LN ILM fuse blew in the middle of this data-set, causing a part of the data to be anomalously noisy.

- cCfbadfile - removes some spurious events in one ^{252}Cf calibration data-set (140105_1556).

- cTransfer - removes parts of two data-sets (131011_2015 and 131014_1845) for which data acquisition continued through the cryogen transfers.

- cBad - combines all of the above.

- cPurge - passes only events after the old-air purge started (data taken on November 7, 2003 and later).

- cZ6neutr - removes data-sets taken in the period between November 21 and December 3, 2003, during which Z6 is known not to be neutralized.

Hence, cBad is meant to include all the “bad data” cuts. The last two cuts were defined as “optional” and were not used in the blind analysis presented below. Figure 8.14 shows the acquired live-time (after the “bad data” cuts described above) as a function of real time. By the end of the run, 52.6 live-days of WIMP-search data was acquired (after the “bad data” cuts). The short flat regions in this figure (around October 20, October 28, November 9, November 20, and December 16) correspond to the brief periods of down-time mentioned in Section 8.2.

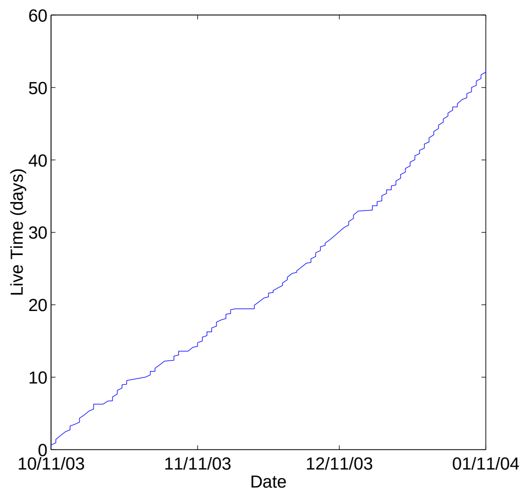


Figure 8.14: Live-time (acquired after the “bad data” cuts) is plotted versus the real time.

8.6.2 Charge χ^2 Cut

The optimal-filter algorithm applied to charge channels returns the χ^2 of the fit. This parameter can be used to reject events with poor charge pulse shapes - for example, the pile-up events (in which more than one pulse takes place in the same 2 ms trace), or the events triggered by noise glitches. The cut is implemented as a quadratic in the χ^2 vs charge energy plane, as illustrated in Figure 8.15 for detector Z3. The cuts were defined on the ^{133}Ba calibration data.

Figure 8.15 also demonstrates existance of two bands. The cause for the upper band turns out the be the sliding seal heater at the top of the cryostat. This heater periodically turns on in order to prevent ice build-up at the top of the cryostat - when it was on, it introduced additional noise peaks at a few hundred kHz scale in the charge channels. The effect was not very large, but sufficiently large to be caught by the optimal-filter algorithm. Hence, about 5% of the events, which took place while this heater was on, were slightly noisier and they formed the upper band in the Figure 8.15. We decided to reject these events in order to keep the cut simple, although these events are perfectly valid.

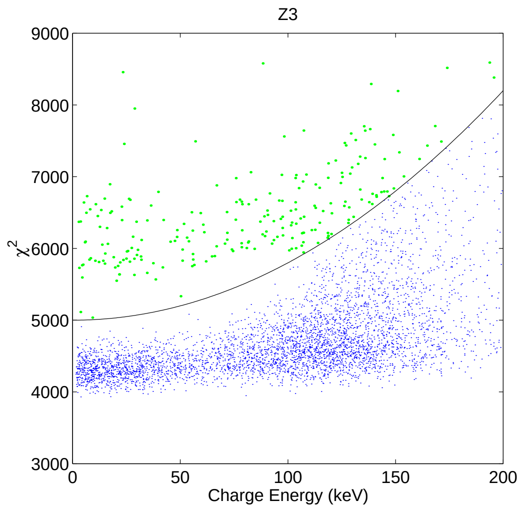


Figure 8.15: Optimal filter χ^2 is plotted against the charge energy. The χ^2 cut, shown as the solid line, rejects events shown in gray/green, and preserves events in black/blue. ^{133}Ba calibration data is shown.

8.6.3 Phonon Pre-trigger Cut

This is a cut in the standard deviation of the pre-trigger part of the phonon traces (about 400 μs long). This cut is complementary to the cut in the charge χ^2 for purposes of eliminating the pile-up events or the noise-triggered events. The distribution of this parameters was obtained from the ^{133}Ba gamma calibrations, by performing gaussian fits for every phonon channel of every detector. The cut was then set at 5σ, where σ was estimated in the gaussian fit.

8.6.4 Charge Threshold Cut

The goal of this cut is to select events that have measurable pulses in the charge channels. This cut is particularly important at lowest energies, where an event with a small phonon signal and no signal in the charge channel could be mistaken for a low-energy nuclear-recoil event.

This cut is defined essentially by using the noise blobs (in the ^{133}Ba calibration data). In particular, we examine the distribution of the optimal-filter amplitudes of pulse-less traces in the inner charge electrode. We fit a gaussian to this distribution and set the cut at $\mu + 5\sigma$, where μ and σ are the mean and the standard deviation estimated in the gaussian fit.

8.6.5 Outer Charge-Electrode Cut

The ZIP detectors contain two charge electrodes, one in the form of a disc covering the inner region of the crystal, and the other in the form of a ring covering the outer edge of the crystal. The outer electrode is used to reject events that take place close to the edge of the crystal, where both charge and phonon signals may be corrupted (since the electric field is not uniform in this region, and since this region is not well covered by the phonon sensors). The cut is defined in the $Q_o - Q_i$ plane, where Q_o and Q_i stand for the optimal-filter amplitudes in the outer and the inner charge channels respectively. ^{133}Ba calibration data is used to fit the Q_o distribution in several bins of Q_i energy. The events passing the cut are in the band centered around $Q_o \approx 0$, bounded from below by $\mu - 3\sigma$ and from above by $\mu + \sigma + 0.04Q_i$, where μ and σ are estimated in the fits. The last term is introduced in order to reduce the impact of the outer-electrode noise on the high-energy inner-electrode events. Figure 8.16 demonstrates this cut for the detector Z3.

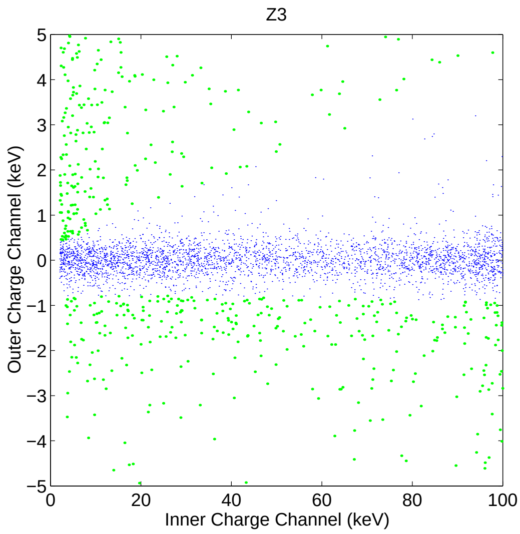


Figure 8.16: Optimal-filter amplitudes in the outer and inner charge electrodes are shown. Dark/blue events pass the outer electrode cut, while gray/green events fail it. ^{133}Ba calibration data is shown.

8.6.6 Electron and Nuclear-Recoil Bands

The electron and nuclear-recoil bands are calculated following the procedure in the Section 5.2. For the ER band we used the ^{133}Ba gamma calibrations and for the NR band we used the ^{252}Cf neutron calibrations. As an example, we show the calculation for the detector Z3; other detectors are very similar.

First, we fit gaussians to the yield (= charge energy / recoil energy) distributions for both electron and nuclear-recoil events, in energy bins, as shown in Figures 8.17 and 8.18. The estimated means and standard deviations are then fitted against energy, as discussed in the Section 5.2. The results of this fit are shown in the Figure 8.19. Finally, Figure 8.20 shows the calculated bands together with the gammas and neutrons from the ^{252}Cf calibration. Unless specified otherwise, the bands are always taken to be $\pm 2\sigma$.

8.6.7 Muon Veto Cut

For this cut we use the time between the global trigger and the last hit in the muon-veto. Events for which this time is long are likely not caused by a muon interacting inside the veto. From the SUF Run21 experience, setting the cut at 40 μs was sufficient to reject most of the muon-coincident events. Since the muon rate at Soudan is significantly suppressed, we could not perform a detailed study of what cut value we should use. Hence, we decided to be somewhat conservative and chose the cut value at 50 μs. We have verified that all of the few muon events we have observed are rejected by this cut.

8.6.8 Singles and Multiples Cuts

To define single scatter events (events in which only one detector was hit), we again use the noise blobs. In particular, single scatter events are events in which only one detector had a phonon signal larger than 6σ of the phonon component of its noise blob.

The multiple scatter events have to satisfy:

- ALL detectors have to pass the "bad data" cut, the charge χ^2 cut, and the phonon pre-trigger std cut.

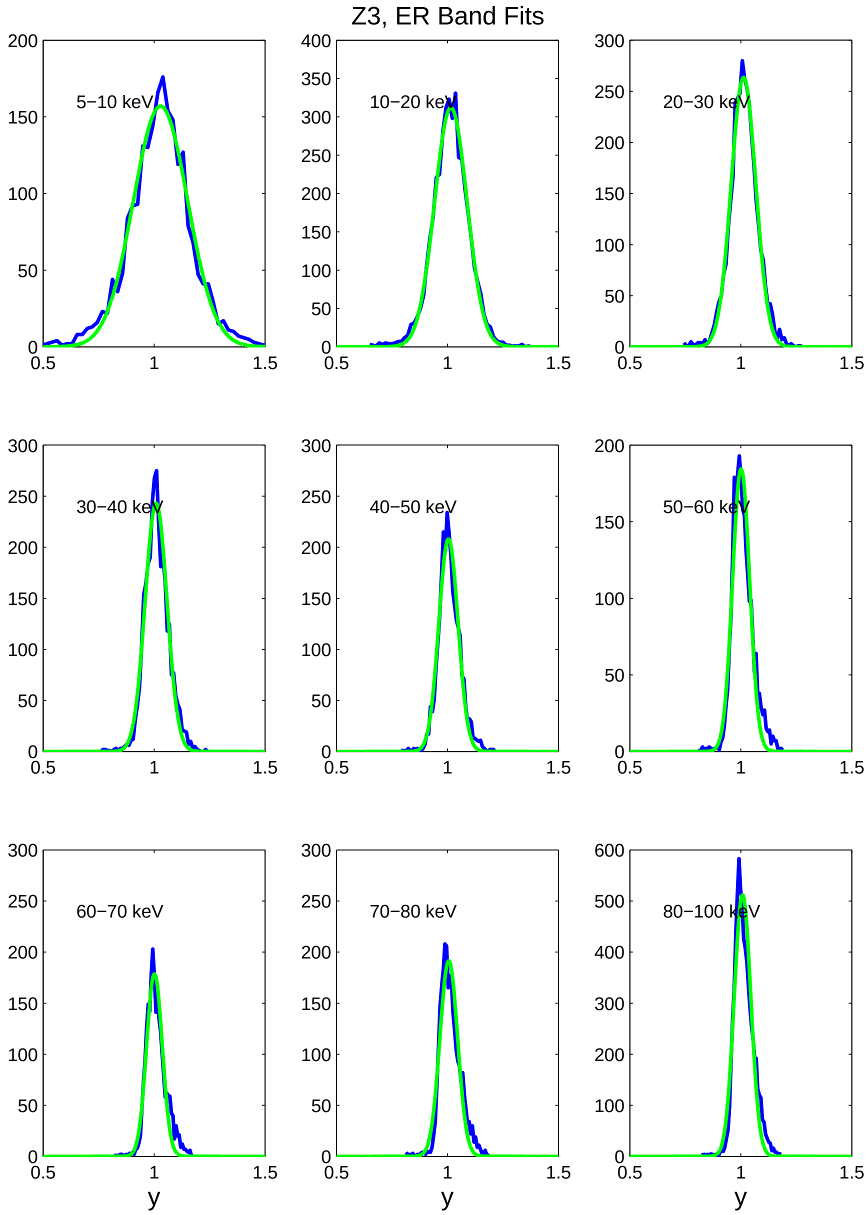


Figure 8.17: Yield y distribution (dark/blue) is fitted in energy bins for the electron-recoil events in Z3. The fits are shown in gray/green. Calibration data is shown.

- At least two detectors have to pass the outer-electrode cuts and have larger recoil energies than 5 keV.

Note that these two definitions do not account for all events. For example, an

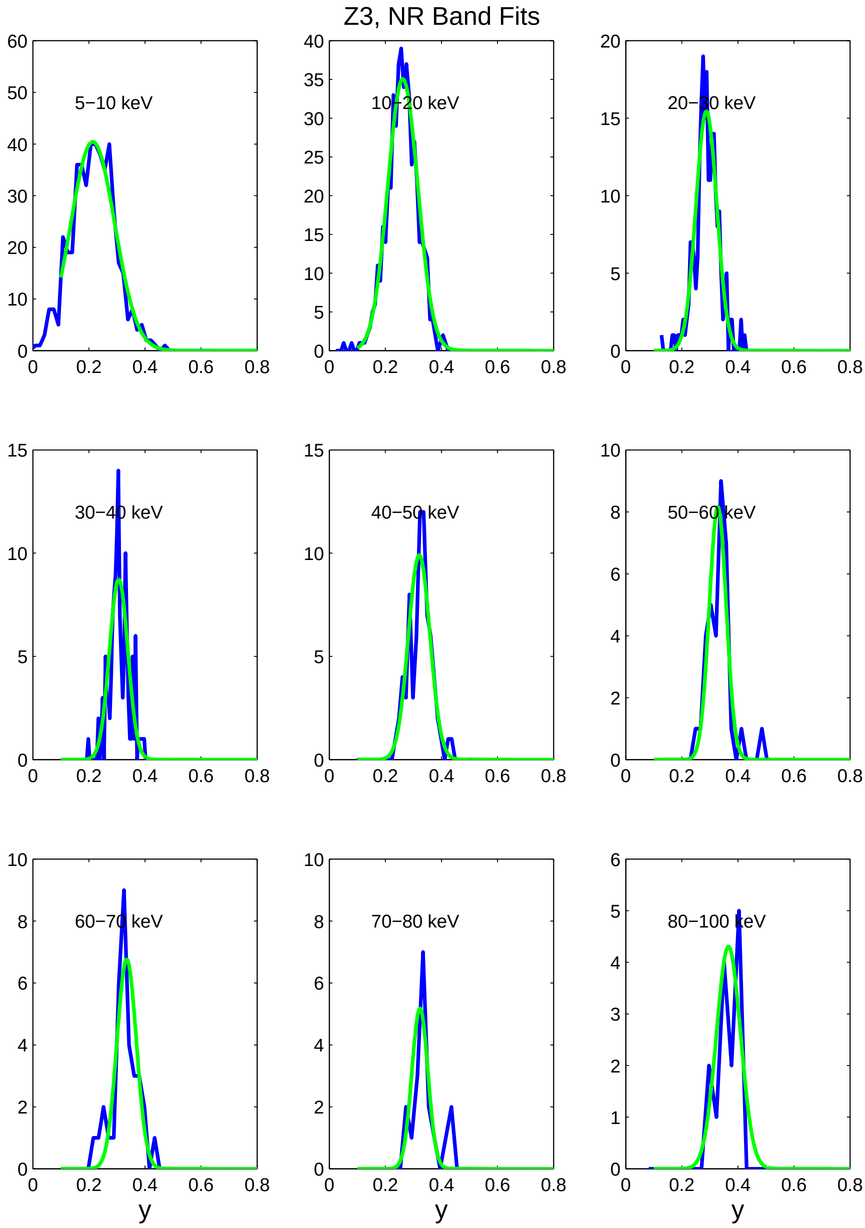


Figure 8.18: Yield y distribution (dark/blue) is fitted in energy bins for the nuclear-recoil events in Z3. The fits are shown in gray/green. Calibration data is shown.

event with 6 keV and 4 keV recoil energies in two detectors would not qualify as multiple scatter ($<$ 5 keV in the second detector), but it may also not qualify as a single scatter since both detectors may have phonon signals larger than 6σ of the phonon component of

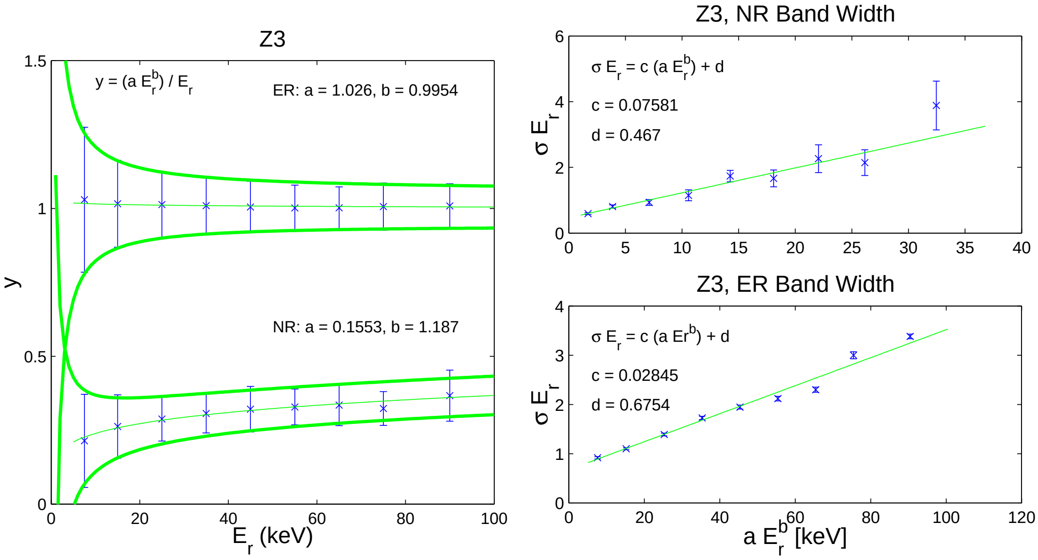


Figure 8.19: Left: Estimated means and standard deviations of the yield (y) distribution are shown as dark/blue x's and vertical bars (2σ bars are shown). The means are fitted against recoil energy E_r; the fits are shown as gray/green thin solid lines. The thick solid lines are the estimated 2σ ER and NR bands. Right: Straight-line fits of σE_r against aE_r^b, where a and b were estimated in the left plot. These fits give estimates of the bands' widths σ for a given recoil energy E_r.

their noise blobs.

8.6.9 Phonon Timing Cuts

The phonon timing cuts are meant to deal with the surface events. As discussed in detail in Chapters 4, 5, and 6, the surface events can have incomplete charge collection, which, if sufficiently incomplete, could make electron-recoil events look like nuclear-recoil events in the ionization yield. Phonon pulse-shape parameters can be used to reject the surface events - as discussed in Chapter 6, both phonon start-time and rise-time tend to be shorter for the surface events than for the bulk events.

Following the results of the beta calibration discussed in Chapter 6, several parameters have been position-corrected:

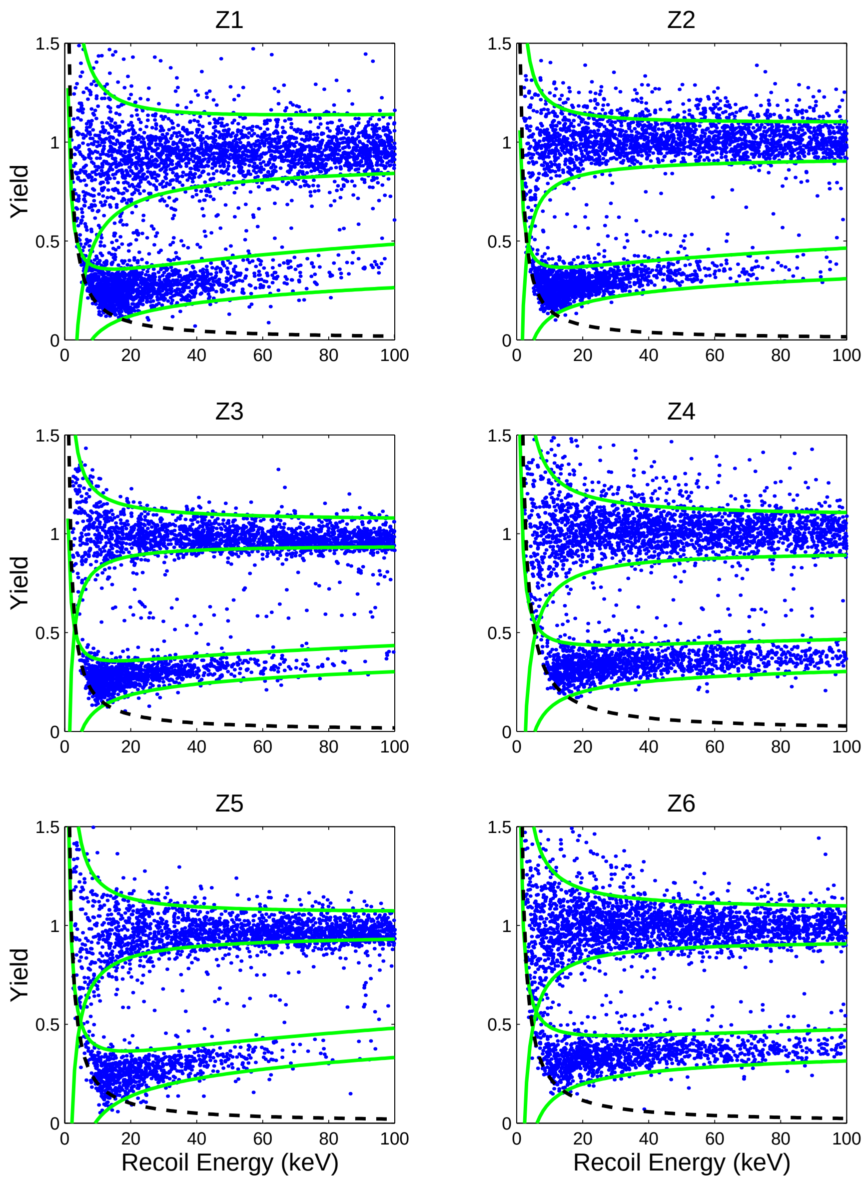


Figure 8.20: ER and NR bands are shown for all detectors (gray/green solid lines), together with the gammas and neutrons from the ^{252}Cf calibration. The dashed line depicts the effect of the charge threshold cut.

- ptrtc: 10% - 40% rise-time of the sum of the four phonon pulses.

- pminrtc: 10% - 40% rise-time of the largest of the four phonon pulses.

- pdelc: delay of the 20% point of the largest phonon pulse with respect to the charge start time.

- pfracc: ratio of the amplitudes in the primary phonon sensor (under which the event took place) and the diametrically opposite phonon sensor.

In order to study how this cut should be made, we define neutrons to be the events in the neutron calibration inside the 2σ NR band, and betas to be the events in the gamma calibration below -3σ of the ER band, and passing the charge threshold cut. It turns out that during the gamma calibration a number of electrons are ejected by gammas interacting in the material around the detectors. These electrons cause surface events in the detectors. As a result, the gamma calibration offers a decent number of surface events, and can be used as a beta calibration as well. The neutrons, therefore, are representative of the signal that should pass the phonon-timing cut and betas are representative of the surface-event background that should be rejected by the cut. We have verified that of the three phonon-timing parameters listed above, the pminrtc+pdelc combination gives the best surface-event rejection. This is consistent with the results of the beta calibration (Chapter 6). Hence, we use this pair for defining the blind phonon-timing cuts.

Since we could not look at the NR region of the WIMP-search data, it was difficult to guess how large the surface-event background will be. Hence, it was difficult to judge if the statistical background subtraction would be performed in the limit calculation or not. For this reason, we decided to define two cuts.

Cut 1 was to be used if the statistical background subtraction is performed. In this case, the cut should minimize the quality factor $Q = \beta(1-\beta)/(\alpha-\beta)^2$, where α and β are the cut efficiencies on neutrons and betas respectively (see Chapter 5). Since the limit is set in the tail of the beta distribution, β varies slowly with the cut parameters, so maximizing $\alpha - \beta$ is similar to minimizing Q. Maximizing $\alpha - \beta$ is also less susceptible to "rejecting the last beta" than Q (Q drops to zero when the last beta event is rejected, i.e. when $\beta = 0$). Hence, we maximize $\alpha - \beta$.

Cut 2 was to be used if the statistical background subtraction is not performed (this cut was used in the blind analysis, see below). In this case, the quantity to minimize is $Po90/\alpha$, where $Po90$ is the Poisson 90% upper limit on the expected number of surface events to leak into the NR band after the phonon-timing cut *in the WIMP-search data*. $Po90$ is easy to calculate for the calibration data: if no beta event passes the cut, $Po90 = 2.3$ etc. To estimate $Po90$ for the WIMP-search data, we calculate the WIMP-search-to-calibration scaling factor by dividing the numbers of events that fall between the NR and ER bands in the WIMP-search data and in the calibration data. Given the number of calibration betas that leak through, and the scaling factor, we can estimate the number of WIMP-search surface events that would leak into the NR band after the phonon-timing cut, and then use this estimate to calculate $Po90$. In practice, the scaling factor was always < 1 (because the number of events in the calibration data was much larger), so we conservatively forced the scaling factor to be 1. Finally, as another conservative step, we consider the leakage in the 4σ NR band, rather than the standard 2σ.

The optimization for both cuts is performed over a grid in the pminrtc-pdelc plane and in energy bins 5-10 keV, 10-20 keV, 20-40 keV and 40-100 keV, for each detector separately. Figure 8.21 shows the positions of both optimized cuts in the pdelc-pminrtc plane for Z3. Note that for high energies (above ~ 20 keV) pdelc is sufficient for rejecting all calibration betas. Below 20 keV, a combination of pminrtc and pdelc becomes necessary. Also note that, as expected, cut 1 is somewhat weaker than cut 2, as it allows more of the calibration betas to pass.

Figure 8.22 shows the surface-event rejection efficiency β and the effective rejection Q_{rej}, as defined in Sections 5.5 and 6.7. These values are obtained prior to the energy-smoothing of the cut-parameter values, and are directly comparable to the values shown in the Section 6.7.

The cut values estimated above were defined in four recoil-energy bins. To define a relatively smooth cut in energy, we "fitted" the cut values against recoil-energy. Figure 8.23 shows how this was done for cut 2 of Z3. Note that, as mentioned above, at energies above 20-30 keV the dominant cut is in pdelc, while at lower energies both parameters become important.

The efficiencies and leakages of the phonon-timing cuts will be discussed in detail below. However, we illustrate the impact of the cut on various types of events in Figure 8.24. Note that the cut rejects most of the surface events, as well as a large fraction of the bulk electron-recoil events, while preserving much of the signal.

8.6.10 Analysis Thresholds

The ultimate energy threshold of the experiment is set by the point where the ER and the NR bands merge. Below this energy we cannot distinguish between the NR events and the dominant ER background. This point is around 5 keV for all detectors. However, the charge threshold cut and the phonon-timing cut become very severe at these lowest energies. Hence, to be conservative, we set the analysis thresholds for most detectors to be

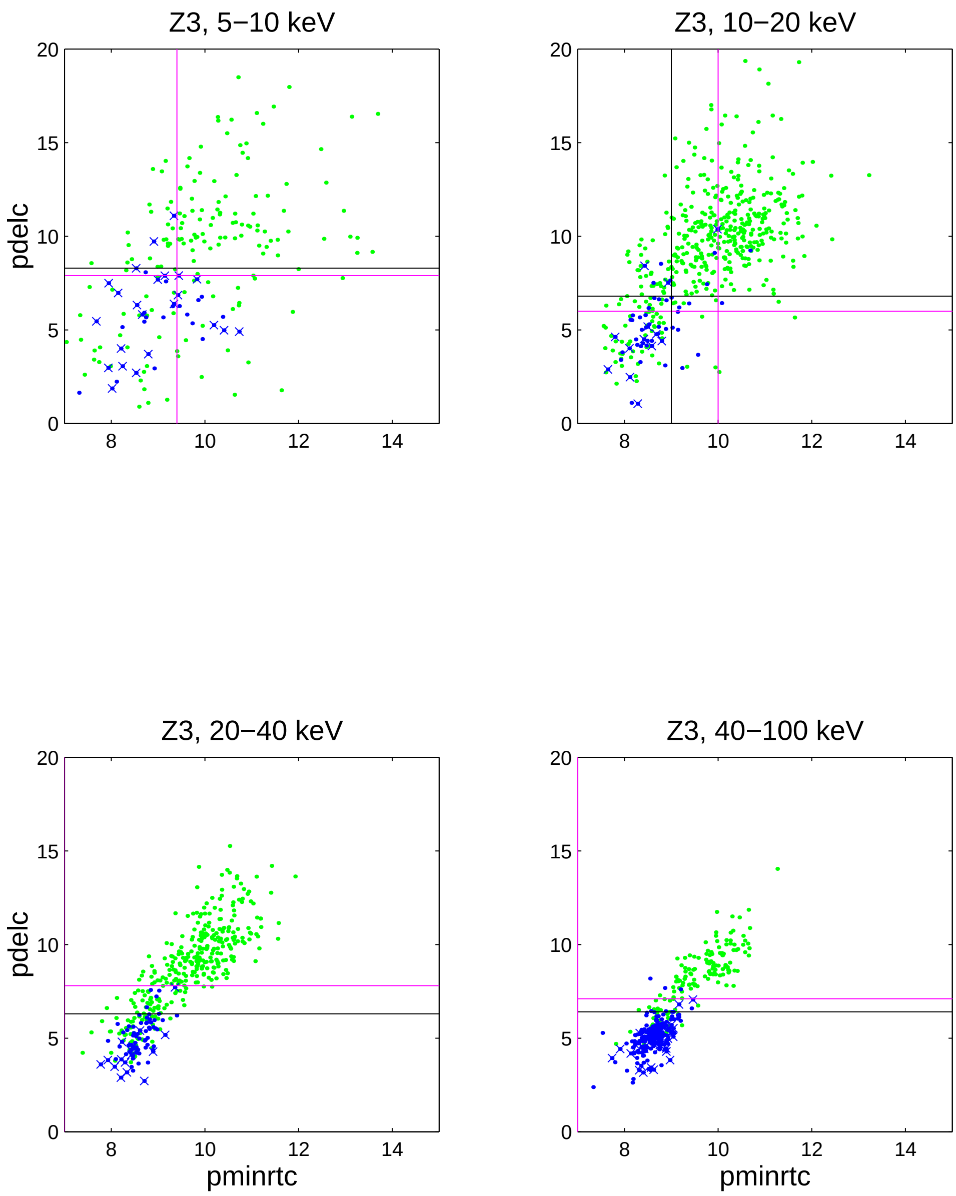


Figure 8.21: pdelc-pminrtc plane for Z3, for four recoil-energy bins. Calibration neutrons are shown as the gray/green points, all calibration betas are shown as dark/blue points, and the calibration betas within 4σ of the NR band are shown as dark/blue x's. The cut values are shown as straight lines - cut 1 in dark/black and cut 2 in gray/magenta.

10 keV. The exceptions are Z1 and Z4 - these detectors had considerable leakage of surface events in the calibration data below 20 keV, even after the phonon-timing cuts. Hence, we set the analysis threshold for these detectors at 20 keV.

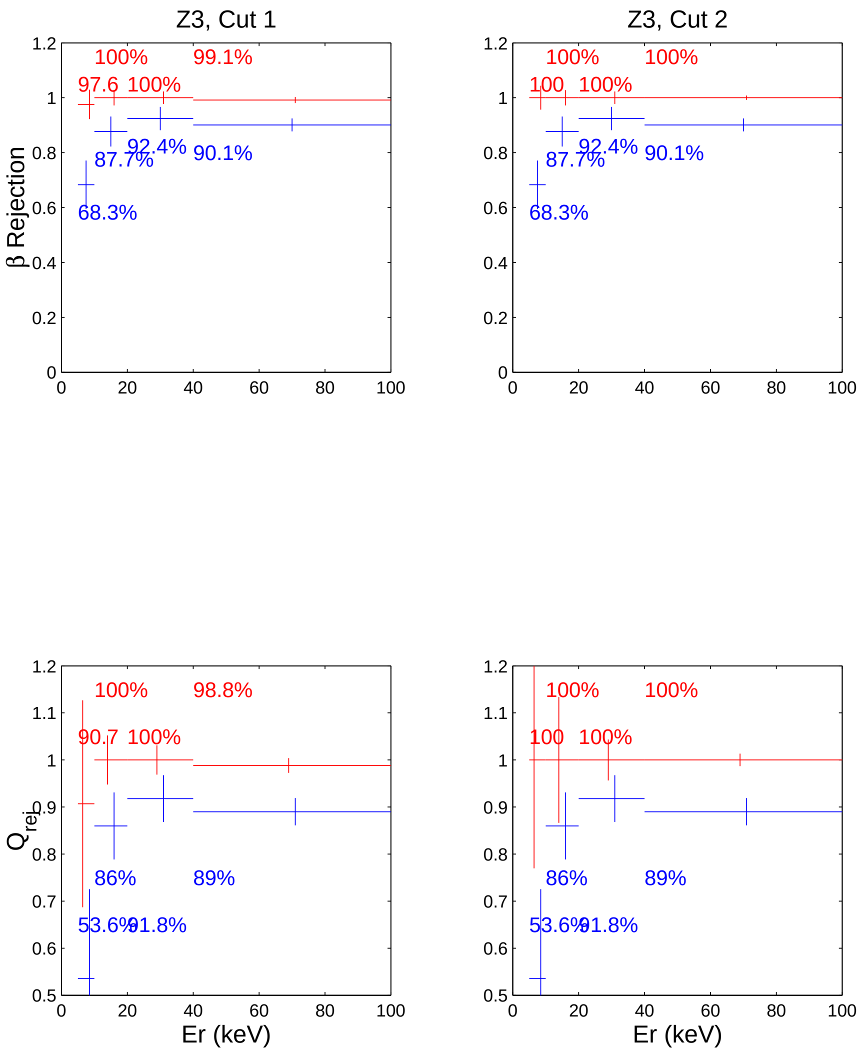


Figure 8.22: Surface event rejection: the two columns of plots correspond to two cuts, the first row shows the surface-event rejection efficiency β and the second row shows the effective rejection Q_{rej}. The lower (blue) set of points (and numbers) corresponds to using the ionization yield alone, and the upper (red) corresponds to using both the ionization yield and the phonon-timing cuts. Note that these values were obtained prior to smoothing the cut values against energy.

8.7 Cut Efficiencies

We now calculate the efficiencies of all blind cuts described above. The cut efficiencies were estimated *after* unblinding the nuclear-recoil region of the WIMP-search data.

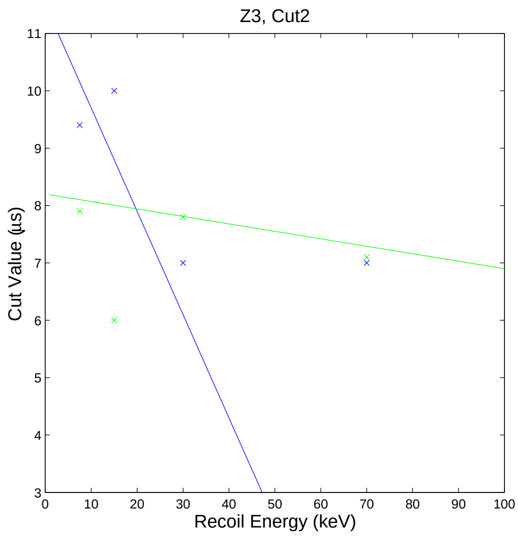


Figure 8.23: Defining phonon-timing cuts that are smooth in recoil-energy. The x's denote the optimized cut values determined in the four energy bins, and the straight lines denote fits to these values against energy. The fits were not performed rigorously - rather, we tried to be conservative and cut more harshly than suggested by optimization, and we used the fact that at energies above 20-30 keV pdelc was sufficient by itself.

Furthermore, after unblinding we discovered that the pulse-fitting algorithm (F5) designed to handle saturated charge pulses had been used to analyze much of the WIMP-search data. This algorithm gives only slightly worse energy resolution than the optimal-filter (OF) algorithm, so the effect was not large. The effect was essentially two-fold:

- The NR and ER bands were calculated using the F5 charge-amplitude estimates, so they are slightly wider than expected.
- The outer-electrode cut was designed using OF charge-amplitude estimates. When applied to F5 estimates of the charge pulse amplitudes, the efficiency of the cut becomes lower due to the worse noise performance of the F5 algorithm.

To properly account for this effect, we estimated the efficiencies of all blind cuts for both OF and F5 estimates of the charge pulse amplitudes. We then made a weighed average

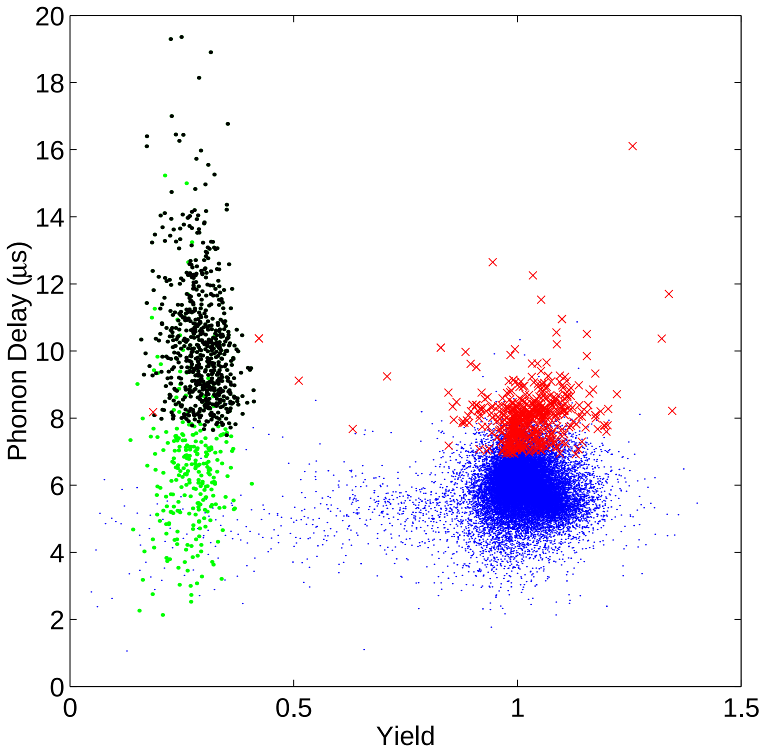


Figure 8.24: Delay of the maximum phonon pulse vs the ionization yield. The thick points show the NR events from the ^{252}Cf calibration: gray (green) for all events and black for events that pass the phonon-timing cut 2. The small (blue) points show the electron-recoil events from the ^{133}Ba calibration - note that the surface events from this calibration extend to the lowest values of the yield, and are centered arount the phonon-delay of ~ 5 μs. The x's denote the electron-recoil events that pass the phonon-timing cut 2. Note that essentially all surface events are rejected, as well as a large fraction of the bulk electron-recoil events (centered around the yield of ~ 1).

of the two, where the weights were determined from the fractions of the WIMP-search data that were analyzed by the two algorithms - see Table 8.1. All efficiencies are estimated in energy bins: 5-8, 8-12, 12-16, 16-20, 20-30, 30-40, 40-50, 50-60, 60-70, and 70-100 keV.

Detector	Z1	Z2	Z3	Z4	Z5	Z6
F5 Fraction	66.1%	57.2%	50.1%	41.0%	51.2%	37.0%

Table 8.1: Fractions (in live-time) of the WIMP-search data analyzed by the F5 algorithm, for all six detectors. The remainder of the data was analyzed by the OF algorithm.

8.7.1 Charge χ^2 and Phonon Pre-Trigger Cut Efficiencies

The combined efficiency of the charge χ^2 cut and of the cut on the standard deviation of the phonon pre-trigger baseline is estimated using the WIMP-search electron-recoil events. We use WIMP-search data rather than the calibration data because we have observed variation in the charge χ^2 parameter among the background datasets. However, the efficiency estimated in this way applies to the electron-recoil events only. We convert this efficiency into the efficiency on the nuclear-recoil events using the centroid of the estimated NR bands. As an example, Figure 8.25 shows the efficiency of these two cuts on the NR events for the detector Z3.

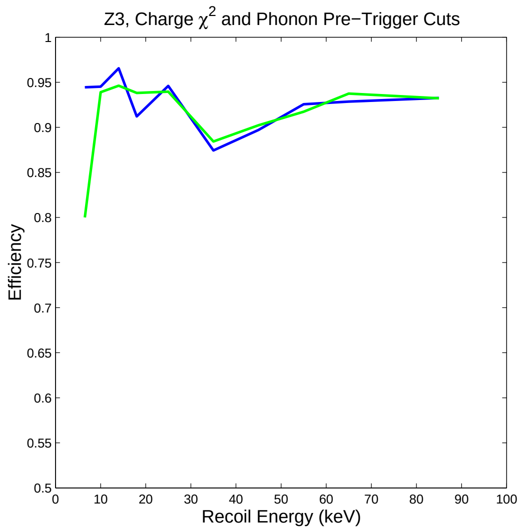


Figure 8.25: Efficiency of the charge χ^2 and phonon pre-trigger standard deviation cuts. Dark/blue shows the estimate for the F5 quantities and gray/green for the OF quantities.

8.7.2 Charge Threshold Cut Efficiency

The efficiency of the charge threshold cut can be estimated in two different ways. One approach is to assume that the yield distribution is gaussian, and then *calculate* the fraction of the gaussian rejected by this cut, which takes a hyperbolic shape in the yield vs recoil-energy plane (see Figure 8.20). This method suffers from the fact that the cut is set on the inner charge-electrode amplitude, while the yield depends on the sum of the two charge channels (so the cut is not exactly hyperbolic in the yield vs recoil-energy plane). The other approach is to use the events in the NR band in the neutron calibration with ^{252}Cf source. In this case, the efficiency is estimated to be the fraction of the NR events (triggered on phonons) that fail the charge threshold cut. The deficiency of this method is that it may be dominated by the low statistics.

The two methods give almost identical results, implying that their potential deficiencies are not significant. Figure 8.26 shows the efficiency of this cut for the example of Z3.

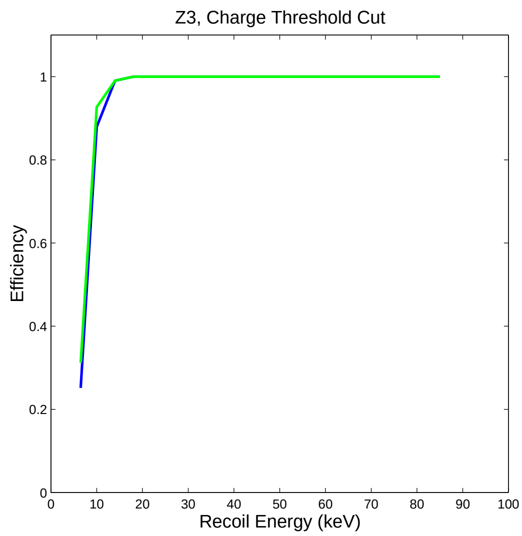


Figure 8.26: Efficiency of the charge threshold cut. Dark/blue shows the estimate for the F5 quantities and gray/green for the OF quantities.

8.7.3 Outer Charge-Electrode Cut Efficiency

The efficiency of the outer charge-electrode cut can be estimated using the NR events in the neutron calibration with ^{252}Cf source. These events are believed to be roughly uniformly distributed throughout the crystal (this is confirmed by the Monte Carlo simulations of the neutron calibration), and, therefore, can be used for determining the efficiency of the fiducial volume cut such as the outer charge-electrode cut. The events are preselected to pass the data-quality cuts (charge χ^2, phonon pre-trigger, charge threshold) and to fall in the NR band. The efficiency estimate is defined as the fraction of the preselected events that pass the outer charge-electrode cut. The efficiency estimated in this way suffers from two complications. First, the electron-recoil events that take place close to the edge of the crystal leak into the NR band, hence contaminating the sample of "true" NR events. The consequence is that the efficiency is under-estimated, and the effect becomes important at high energies (> 70 keV) where the statistics in the neutron calibration are relatively low. Second, there is a correlation between the outer charge-electrode cut and the NR band cut - namely, a small population of events is rejected by both cuts. Hence, one must be careful about preselecting events, so that this population of events is counted in the efficiency estimate of only one of the two cuts.

It is possible to correct the estimate for these two complications. First, one can impose a very weak cut on the outer charge-electrode amplitude to preselect events avoiding the leaked ER events (only a small fraction of the "true" neutron NR events is removed by this weak cut; this fraction can be neglected, or even estimated using the neutron NR events at all energies (10-100 keV) to avoid the low-statistics problem at energies > 70 keV).

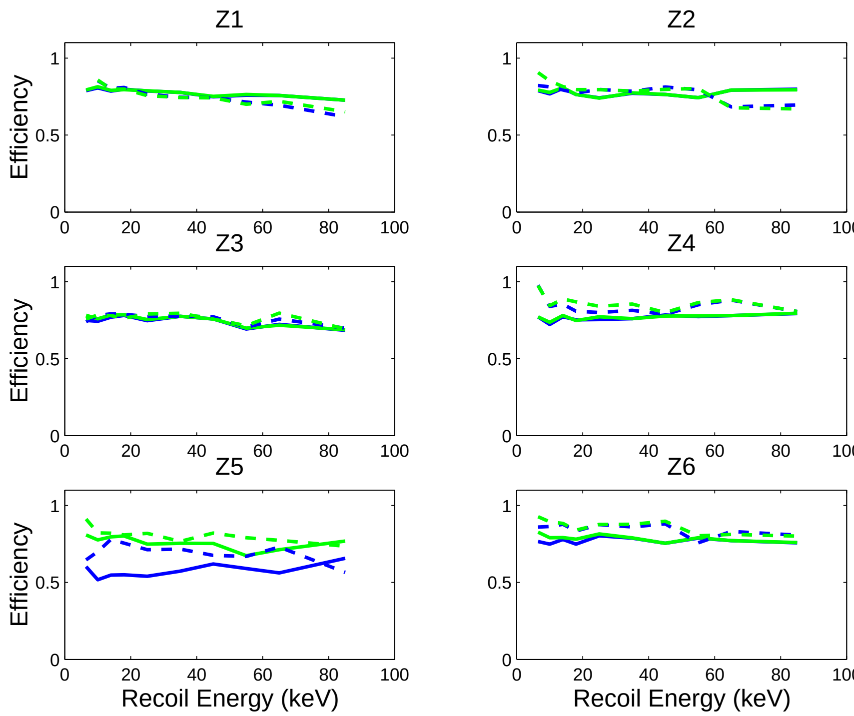


Figure 8.27: Efficiency of the outer charge-electrode cut for all Tower 1 detectors. Dark/blue shows the estimates for the F5 quantities and gray/green for the OF quantities. The solid lines correspond to the Monte Carlo estimate and the dashed lines correspond to the data estimate of the efficiencies (see text for more detail).

Second, to avoid the correlation complication, one can impose a very weak NR band cut when preselecting events - the ^{252}Cf calibration is relatively short, so the contamination by the surface events in a wide NR band is small. Hence, the correlated population of events is taken into account here, and for the NR band efficiency it can be ignored by preselecting events that pass the outer charge-electrode cut.

The efficiency of the outer charge-electrode cut can also be estimated using a Monte Carlo simulation [200]. In particular, a Monte Carlo simulation of the ^{252}Cf neutron calibration is performed. The simulated recoil-energies are converted into charge-energies using the centroids of NR bands, and are convolved with the observed noise performance of the charge channels. With such calculation, one can estimate directly the fraction of the neutrons that fail the outer-charge electrode cut.

Figure 8.27 shows the efficiencies of the outer charge-electrode cut for all detectors and for both the data and the Monte Carlo methods. Note that the data and the Monte Carlo efficiency estimates agree well for most detectors. The only exception is Z5, for which the two complications discussed above were the most substantial. For this reason, we decided to use the more conservative Monte Carlo estimates in the calculation of the WIMP limit.

8.7.4 Nuclear-Recoil Band Efficiency

The efficiency of the NR band cut is determined using the NR events in the neutron calibration with ^{252}Cf source. The events are preselected using the data-quality cuts (charge χ^2, phonon pre-trigger, charge threshold), the outer charge-electrode cut, and the 4σ NR band, which is expected to include most of the true NR events. Furthermore, we apply the phonon-timing cut to avoid possible contamination by the surface events. The efficiency estimate is defined as the fraction of the preselected events that fall in the 2σ NR band. Figure 8.28 shows the efficiency of this cut for the example of Z3.

8.7.5 Muon Veto Cut Efficiency

The efficiency of the muon veto cut is straightforward: for the 600 Hz rate of the muon veto (dominated by the ambient photons), and the 50 μs rejected window after each hit, about 3% of the data-acquisition time is rejected. Hence, the efficiency of the muon veto cut is 97%, and it is energy independent.

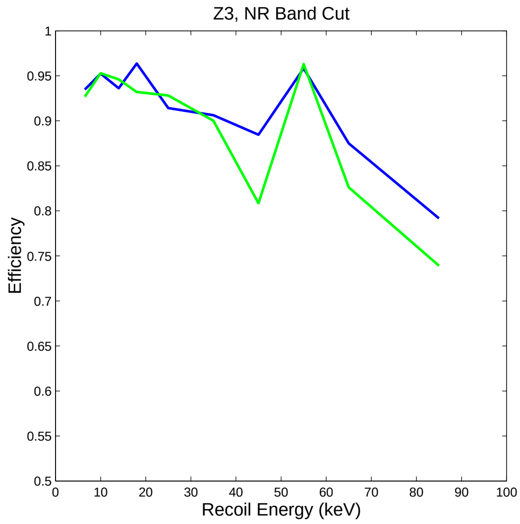


Figure 8.28: Efficiency of the NR band cut. Dark/blue shows the estimate for the F5 quantities and gray/green for the OF quantities.

8.7.6 Phonon-Timing Cut Efficiency

We will focus on the phonon-timing cut 2, as this is the cut we used in calculating the new limit (without statistically subtracting the background - see Section 8.6.9). The efficiency of the phonon-timing cut is estimated using the NR events in the neutron calibration with ^{252}Cf source. The events are preselected to pass all data-quality cuts (charge χ^2, phonon pre-trigger, charge threshold), the outer charge-electrode cut, and the 2σ NR band cut. The efficiency estimate is defined as the fraction of the preselected events that pass the phonon timing cut 2. Figure 8.29 shows the efficiencies of the phonon-timing cut 2 (case for no statistical background subtraction) for all detectors of Tower 1. We show all detectors (rather than just Z3) to indicate the non-negligible variation of the efficiency among the detectors.

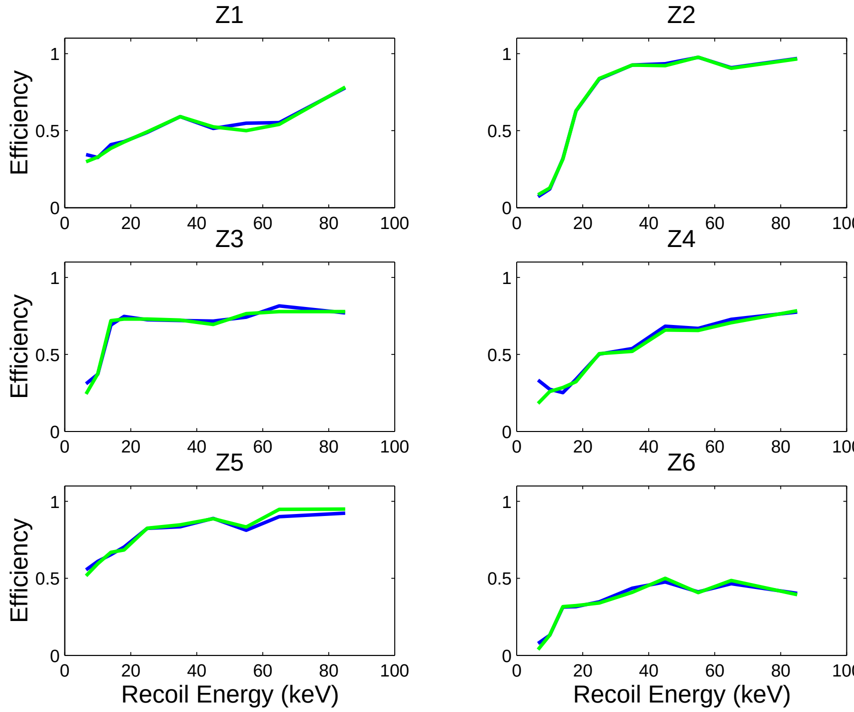


Figure 8.29: Efficiency of the phonon-timing cut 2 (case for no statistical background subtraction) for all Tower 1 detectors. Dark/blue shows the estimates for the F5 quantities and gray/green for the OF quantities.

8.7.7 Overall Cut Efficiency

The cut efficiencies described above are combined (multiplied together) to obtain the overall cut efficiency. Note that the efficiency is forced to zero below 10 keV for all detectors, and below 20 keV for Z1 and Z4, to reflect their analysis thresholds. Furthermore, Z6 is not used in this analysis because of its contamination and of its neutralization problems. We further perform the average over the Ge detectors Z1, Z2, Z3 and Z5, and we perform the weighed average of the OF and F5 estimates. Finally, we perform a fit to the efficiency estimate against energy: a straight line above 20 keV and a quadratic below 20 keV, with the slopes matched at 20 keV. Figure 8.30 shows the effect of adding each individual cut efficiency estimate, and it shows the overall efficiency estimate of this analysis. The overall efficiency estimate for the OF-only analysis is also shown. The discontinuities at 20 keV are due to the analysis threshold of Z1. Integrated over the 10-100 keV bin, the overall F5-OF averaged efficiency yields the net Ge exposure of 19.4 kg-days.

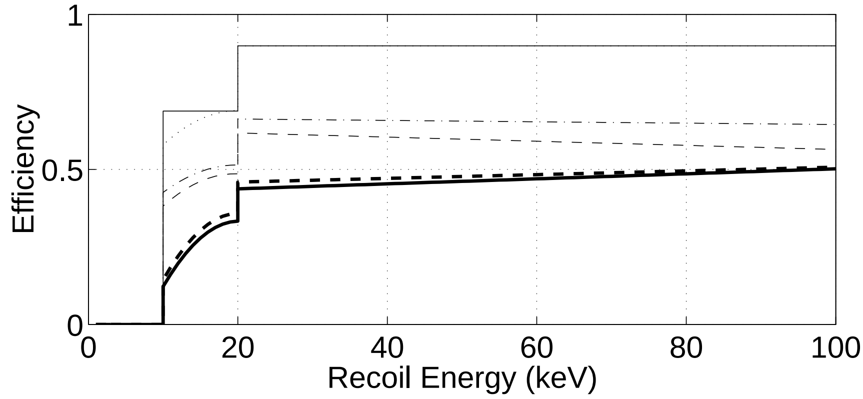


Figure 8.30: Efficiency averaged over the Ge detectors Z1, Z2, Z3, and Z5, and over the F5 and OF estimates. The solid thin line includes the charge χ^2, the phonon pre-trigger, and the muon-veto cuts. The dotted line also includes the charge threshold cut. The dot-dashed line also includes the outer charge-electrode cut. The thin dashed line also includes the NR band cut. The thick solid line also includes the phonon-timing cut 2 - it is the overall efficiency of this analysis. The thick dashed line shows the corresponding overall efficiency if OF-only estimates are used.

8.8 Background Rates and Leakages

The background for the CDMS II experiment can be divided in three categories:

- Gamma background: This background comes mostly from the radioactive contamination inside the Pb shield, and from the radon present in the air throughout the passive shield (the passive shield is not hermetically closed, but it was purged with old air for most of the run - we will discuss this in more detail). The gammas dominantly interact with electrons in the detectors, so they appear in the electron-recoil band.

- Beta background: There are two sources of this type of background. One source are interactions of gammas in the material around the detectors. Electrons ejected in such interactions can then proceed to interact in the detectors. The second source is the radioactive contamination of the detectors themselves and of the inner surfaces of the Tower and of the detector holders. Note that since the penetration depth of the betas is very small, betas produced outside of the Tower cannot reach the detectors. However, exactly because of their small penetration depth, betas usually interact with electrons at the surface of the detector. In other words, betas are the dominant cause of surface events. As discussed in Chapters 4, 5, and 6, and earlier in this Chapter, the surface events can have poor charge collection, and, hence, can mimic nuclear-recoil events. In absence of the neutron background (see below), betas become the dominant background in the experiment.

- Neutron background: Neutrons are the dominant background that interacts with the nuclei in the detector. Hence, unlike the gamma or beta backgrounds which can be substantially suppressed on event-by-event basis using the ionization yield or phonon pulse-shape parameters, the neutron background can be handled only in statistical terms - e.g. by comparison of rates in Si vs Ge detectors, multiple-scatter vs single-scatter rates etc. There are two sources of neutron background. The first is the interactions of the cosmic rays (muons) that produce neutrons. These interactions can take place in the passive shield (internal neutrons), in which case they can be significantly suppressed by the muon-veto. They can also take place in the surrounding rock (external neutrons) - if the produced neutrons are sufficiently energetic, they can punch through the polyethylene shield and interact in the detectors. There is currently no way to identify such events. The second source of neutrons are the (α,n) reactions in the rock, where α's are emitted by uranium and thorium, which then produce neutrons in interactions with Al, Na, Mg and ^{18}O. These neutrons are significantly suppressed by the polyethylene shield.

In the remainder of this Section, we discuss the observed and the expected rates of these three types of backgrounds. We also estimate the leakages of these backgrounds into the signal region, after the cuts described in the previous Section.

8.8.1 Background Gamma Rate and Leakage

The gamma-background events are single-scatter muon-anticoincident events that pass the data-quality cuts (charge χ^2, phonon pre-trigger, charge threshold), the outer charge-electrode cut and that have the ionization yield around 1. More precisely, we require the events to be inside a wide electron-recoil band:

- Above the -3σ of the ER band or above 0.8 in yield.
- Below the 3σ of the ER band or below 1.2 in yield.

Figure 8.31 shows the trend of the gamma rates over the course of the WIMP-search run. The rates are calculated for the 15-100 keV bin, and are corrected for the OF-F5 weighed average of the cut efficiencies, as discussed in the previous Section. Note that the gamma rate was initially relatively high - on November 6, 2003, we started flushing old air inside the passive shield, which reduced the amount of radon in the air around the Icebox and filling the gaps in the passive shield. As a result, we observed a drop in the gamma rate roughly by a factor of 5. Note also that the gamma rates in the inner detectors tend to be lower - the definition of the single-scatter events for Z1 and Z6 is compromised by the fact that they do not have a neighbor on one side.

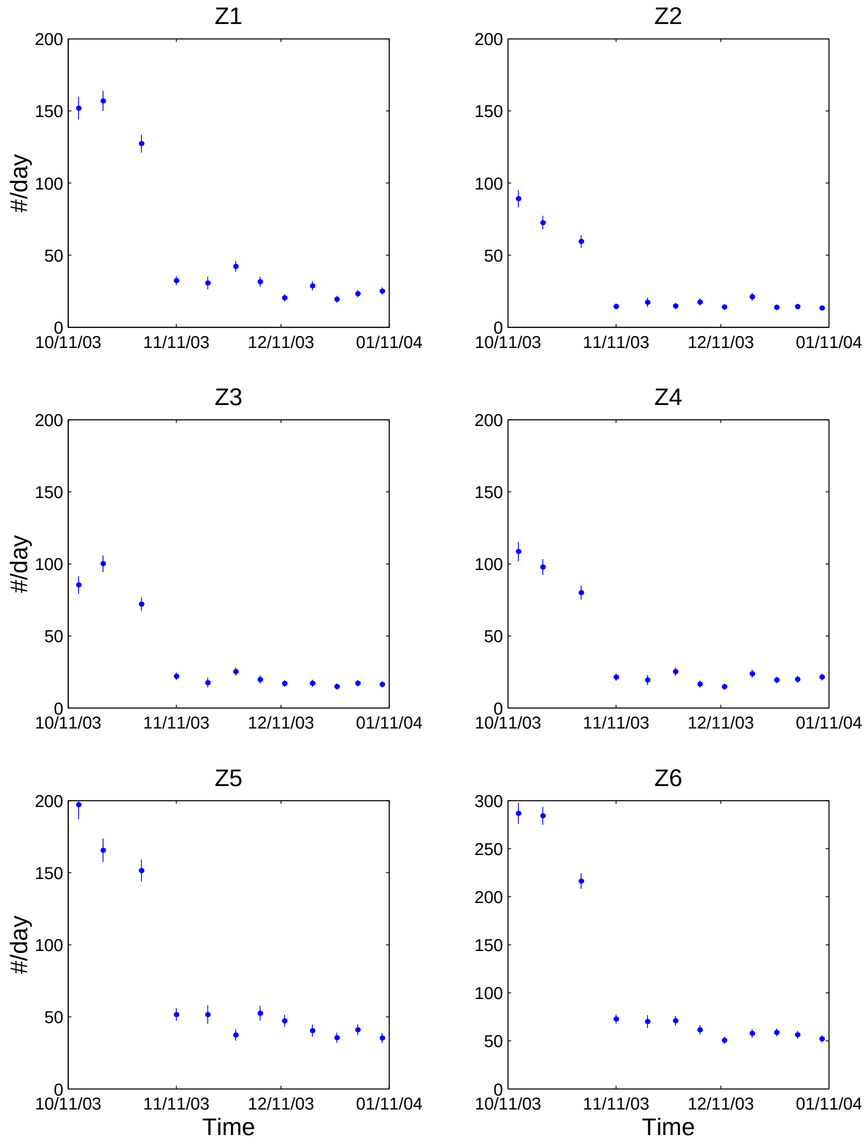


Figure 8.31: Gamma rates as a function of time. Only gammas in the 15-100 keV bin are considered (to avoid the time-variation in the 10.4 keV line due to neutron activation of Ge in calibration runs). The rates can be easily converted into #/keV/kg/day for 250 g Ge and 100 g Si detectors - we plot simply #/day as it gives a sense of the total number of events observed in the 52.6 live-days of the WIMP-search run.

Figure 8.32 shows the gamma spectra for all Tower 1 detectors after the old-air purge started. The spectra are generally flat, except for the 10.4 keV line due to the activation of Ge, and at 1 event/keV/kg/day or less.

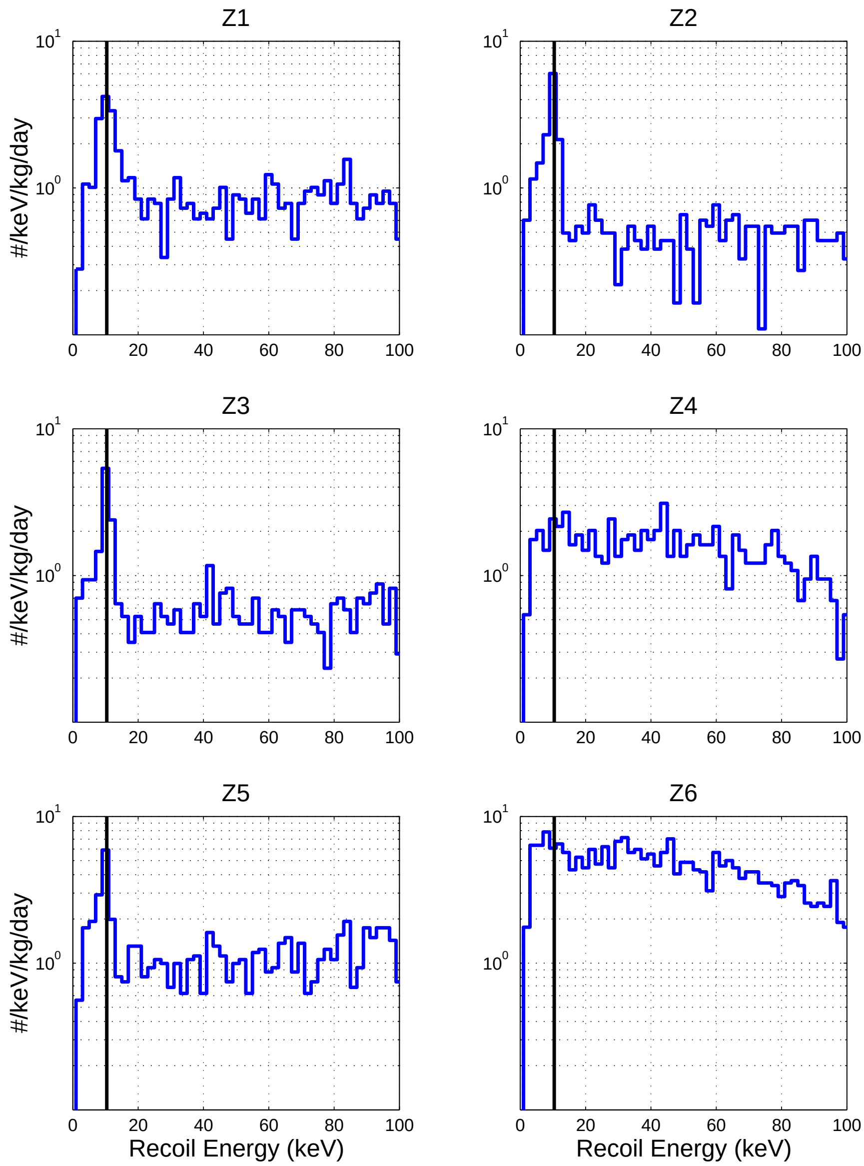


Figure 8.32: Gamma spectra are shown for all Tower 1 detectors, after the old-air purge started. The vertical lines denote the 10.4 keV line due to the activation of Ge (the line is not visible in the spectra for Si detectors).

Finally, we estimate the leakage of the gamma-background into the signal region using the gamma calibrations with ^{133}Ba. In the 10-100 keV bin for Ge detectors Z2, Z3, and Z5, and in the 20-100 keV bin for the Ge detector Z1, we observe 29,716 single-scatter events that pass the data-quality cuts and the outer charge-electrode cut. None of these events pass the NR band cut and the phonon-timing cut. Hence, the rejection efficiency of the gamma events is estimated to be $> 99.99\%$ at 90% confidence.

In the WIMP-search data, we observe the total of 2443 events in the same detectors and energy bins, that pass the gamma-event definition (given above). Hence, we expect to have < 0.25 gamma events leaking into the signal region of the WIMP-search data (at 90% confidence).

8.8.2 Background Beta Rate and Leakage

The beta-background events are single-scatter muon-anticoincident events that pass the data-quality cuts (charge χ^2, phonon pre-trigger, charge threshold), the outer charge-electrode cut, and that have low yield values. More precisely, we require the events to be:

- Above the $+2\sigma$ of the NR band.
- Below the -3σ of the ER band.

Figure 8.33 shows the trend of the beta rates over the course of the WIMP-search run. Similarly to the gamma rates, the rates are calculated for the 15-100 keV bin, and are corrected for the OF-F5 weighed average of the cut efficiencies, as discussed in the previous Section. Although the statistics are low, there does seem to be a drop in the beta rates after the old-air purge started. This is expected since the overall gamma rate dropped by a factor of ~ 5, and a considerable fraction of betas comes from the interactions of gammas with the material around the detectors. Similarly to the case of gammas, the beta rates in the inner detectors also tend to be lower. Again, this is because Z1 and Z6 do not have a neighbor on one side, so the definition of the single-scatter events is compromised. Since betas are prone to back-scattering (and, hence, tend to be double-scatters in the nearest-neighbor detectors), this effect is non-negligible.

Estimating the beta rejection efficiency and the beta leakage is more involving than estimating gamma leakage. One naive, straightforward, and incorrect way of estimating beta leakage is as follows. We count the number of beta-events in the following three cases:

- A: beta-events in the ^{133}Ba calibration that also fall in the NR band and pass the phonon-timing cut 2.
- B: beta-events in the ^{133}Ba calibration that fall outside of the NR band.
- C: Single-scatter muon-anticoincident beta-events in the WIMP-search that fall outside of the NR band.

The idea is to use the ratio of betas outside the NR band in the calibration and in the WIMP-search data, together with the number of betas in the NR band of the calibration data to estimate the number of betas in the NR band in the WIMP-search data. Hence, we are interested in $f = A/B$ (although $A/(A+B)$ is the beta rejection efficiency) - we assume f is binomially distributed, which allows an easy estimate of σ_f using, for example, the Matlab built-in function *binofit*. The beta leakage is then estimated as $fC = AC/B$ in each energy bin, and the overall leakage is summed over the detectors and over the energy

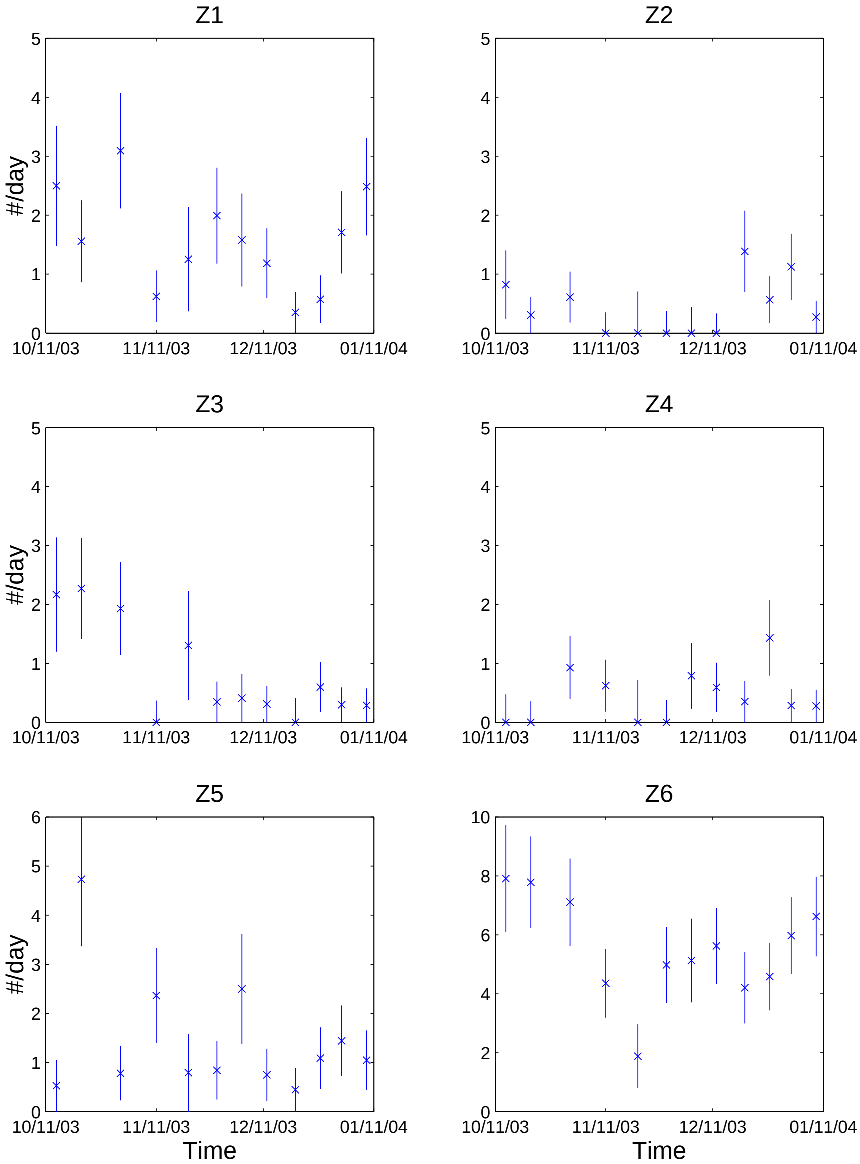


Figure 8.33: Beta rates as a function of time. Only betas in the 15-100 keV bin are considered. The rates can be easily converted into #/keV/kg/day for 250 g Ge and 100 g Si detectors - we plot simply #/day as it gives a sense of the total number of beta events observed in the 52.6 live-days of the WIMP-search run.

bins (with uncertainties summed in quadrature). Similarly to cut efficiencies, we take the weighed average of the OF and F5 estimates. The results of this approach are shown in the Table 8.2 - the overall beta-leakage estimate for the 10-100 keV bin of Z2, Z3, and Z5, and

20-100 keV bin of Z1 is 0.35 ± 0.45.

Detector	5-10 keV	10-20 keV	20-40 keV	40-100 keV	Sum
Z1	-	0 ± 0.18	0.25 ± 0.24	0 ± 0.13	0.24 ± 0.27
Z2	0.05 ± 0.08	0 ± 0.04	0 ± 0.19	0 ± 0.04	0 ± 0.19
Z	0 ± 0.18	0 ± 0.22	0 ± 0.11	0 ± 0.03	0 ± 0.24
Z4	-	0.20 ± 0.15	0.02 ± 0.06	0 ± 0.07	0.02 ± 0.10
Z5	0.48 ± 0.40	0.06 ± 0.12	0.04 ± 0.11	0 ± 0.09	0.10 ± 0.18
Z6	0.12 ± 0.36	0 ± 0.23	0.32 ± 0.77	0.73 ± 0.62	1.05 ± 1.01
Z1235	-	0.06 ± 0.25	0.28 ± 0.34	0 ± 0.16	**0.35 ± 0.45**

Table 8.2: Beta Leakage in the "naive" approach: the sum in the last column is done over 10-100 keV for Z2, Z3, Z5, and Z6, and over 20-100 keV for Z1 and Z4. The last row shows the sum over the Ge detectors Z1, Z2, Z3, and Z5. The bold-faced numbers in the lower-right corner are the overall leakage estimate for the Ge detectors in the afore-mentioned energy bins.

This estimate is incorrect for at least two reasons. First, the phonon-timing cut 2 (which we use in this analysis) is essentially defined to reject the last beta-event in the calibration data. Hence, if we use the same calibration data-set to estimate the beta leakage, f will not be binomially distributed (hence, the above method underestimates the leakage). Second, the *binofit* routine returns the confidence interval at 68%, not σ, so adding uncertainties in quadrature is also not correct. The first problem can be avoided by using calibration data that was not used for defining the phonon-timing cut. The second problem can be avoided by using the likelihood-ratio method in the Feldman-Cousins framework [201, 202]. C. Chang calculated the beta-leakage estimate with this approach using a smaller ^{133}Ba calibration data-set (about 1/10 the size of the data-set used for defining the cuts) [203] - his estimate was 0.59 (2.52 for 90% U.L.) for the F5 quantities and 0.67 (2.80 for 90% U.L.) for the OF quantities, for the same detector-bin combination as above. These results are consistent (but larger) with the "naive" approach outlined above.

An alternative approach is to estimate the probability distribution of the rejection factor f. This can be done by using the fact that the phonon-timing cut was (approximatly) defined to reject the last beta-event out of N beta-events in the calibration sample. As shown by B. Sadoulet [204], the mean and the standard deviation of f are:

$$< f >= \frac{1}{N+1},\ \sigma_f^2 = \frac{N}{(N+2)(N+1)^2} \tag{8.1}$$

One then applies the estimated rejection factor f to the number of events in the region C defined above to get the estimate of the beta leakage. With this method, the beta-leakage estimate is 0.69 ± 0.24 for the same detector-bin combination as above. Again, this estimate is consistent with the previous two.

All methods described above rely on the comparison of the calibration and the WIMP-search data in order to estimate the beta-leakage in the WIMP-search data. While they account for the statistical error fairly well (some better than others), the systematic differences between the calibration and the WIMP-search data are not taken into account. For example, it is reasonable to expect that the angular distributions of the incoming betas are not the same for the WIMP-search data and the calibration data (performed using a point-source). Also, the calibration data allows the study of betas ejected by gammas from the surrounding material - the betas from the radioactive decays due to the contamination of the detectors (or the Tower) may indeed behave differently. As shown in the following Section, there are signs that these systematic differences are non-negligible, probably even larger than the statistical error estimated above. However, appart from Monte Carlo studies which have not been performed yet, it is difficult to estimate such effects.

8.8.3 Neutron Rate

The expected neutron rates are estimated using Monte Carlo simulations [166], [205]. These simulations predict that for the 52.6 live-days of the WIMP-search run, 0.7 internal neutrons are expected to reach the detectors. The internal neutrons are produced in the interactions of muons inside the passive shield - hence, they should be tagged by the muon-veto system. However, no such muon-coincident nuclear-recoil events were observed. Similarly, the simulations predict 0.07 external neutrons to reach the detectors (external neutrons are produced in interactions of muons outside the muon-veto, and, hence, cannot be tagged by the muon-veto system). The ratio 1:10 of the external and internal neutron rates is robust, but the overall scaling has a factor of ~ 2 uncertainty due to the poor knowledge of the neutron production rates at the Soudan mine depth. Since no internal neutrons were observed, the expected number of the external (veto-anticoincident) neutrons is < 0.3 at 90% C.L. Finally, Monte Carlo simulations based on the composition of the rock in the Soudan mine, and including the polyethylene shields, predict that the contribution of the (α,n) neutrons to the overall neutron background is > 10 times smaller than the contribution of the muon-induced external neutrons.

8.8.4 Summary of Expected Backgrounds

Table 8.3 summarizes the expected background estimates for the Ge detectors Z2, Z3, and Z5 10-100 keV, and Z1 20-100 keV.

Gamma Leakage	< 0.25 events, 90% C.L.
Beta Leakage	0.69 ± 0.24 ([204])
Neutron Background	< 0.3 at 90% C.L.

Table 8.3: Summary of different types of backgrounds expected for the 52.6 live-days of the Tower 1 WIMP-search run.

8.9 Dark Matter Limit

In the previous Sections, we have defined the various cuts, estimated their efficiencies on the nuclear-recoil events, and estimated the contamination of the signal region by the various types of backgrounds. In this Section, we finally look at the single-scatter, muon-anticoincident, nuclear-recoil (SSACNR) events in the WIMP-search data. Figures 8.34 and 8.35 show the yield vs recoil-energy planes for all Tower 1 detectors before and after the phonon-timing cut was applied. All single-scatter muon-anticoincident events passing the data-quality cuts (charge χ^2, phonon pre-trigger, charge threshold) and the outer charge-electrode cut are shown. We do not observe any events in the nuclear-recoil band, after the phonon-timing cut, in the Ge detectors Z2, Z3, and Z5 10-100 keV, and in Z1 20-100 keV. We also do not observe any nuclear-recoil events in the Si detector Z4 20-100 keV. We also do not observe any multiple-scatter muon-anticoincident nuclear-recoil events that pass the phonon-timing cut in *at least* one detector. Note that the phonon-timing cuts have also a significant effect on the bulk electron-recoil events (i.e. events that fall in the electron-recoil band), as mentioned in the Section 8.6.9. The effect varies from detector-to-detector, because the phonon-timing cuts were adjusted to each detector separately.

Figure 8.36 shows the corresponding results if OF quantities only were used. The imposed cuts are the same, but they are applied to the OF estimates of the charge channel

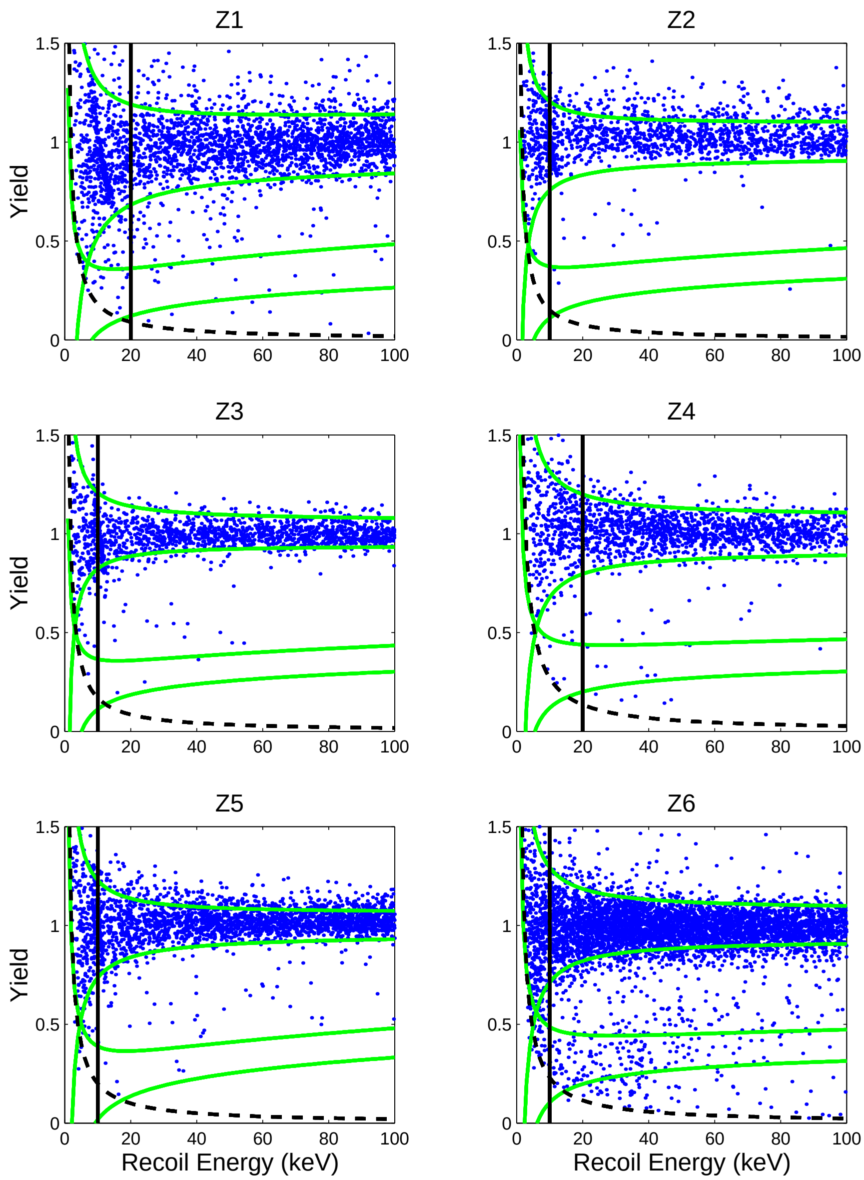


Figure 8.34: Single-scatter muon-anticoincident events observed in the WIMP-search run in all detectors of Tower 1, prior to applying the phonon-timing cut. The gray lines denote the NR and ER bands, the vertical black lines denote the analysis thresholds, and the dashed black lines denote the charge threshold cut.

amplitudes. In this case, we do observe one single-scatter muon-anticoincident nuclear-recoil event in Z5 at 64 keV recoil-energy. This event was previously analyzed with the F5 estimates of the charge channel amplitudes, and it failed the outer charge-electrode cut

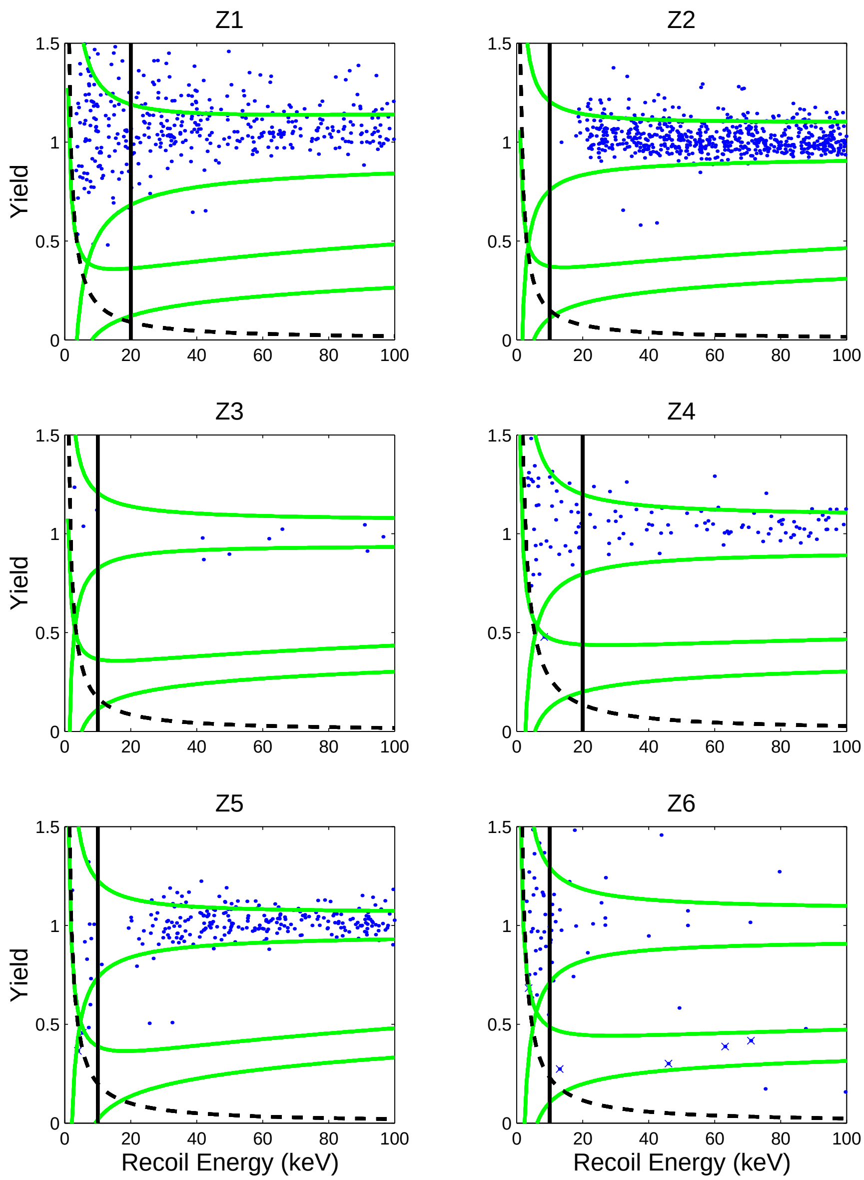


Figure 8.35: Single-scatter muon-anticoincident events observed in the WIMP-search run in all detectors of Tower 1, after applying the phonon-timing cut. The x's denote the NR events. The gray lines denote the NR and ER bands, the vertical black lines denote the analysis thresholds, and the dashed black lines denote the charge threshold cut.

(because the F5 estimates are somewhat noisier than the OF estimates, and more sensitive to the baseline drift). This single observed event is consistent with expected beta-leakage discussed in the previous Section. Note that this analysis is not blind, but it remains unbiased because the original, blind, cuts are used.

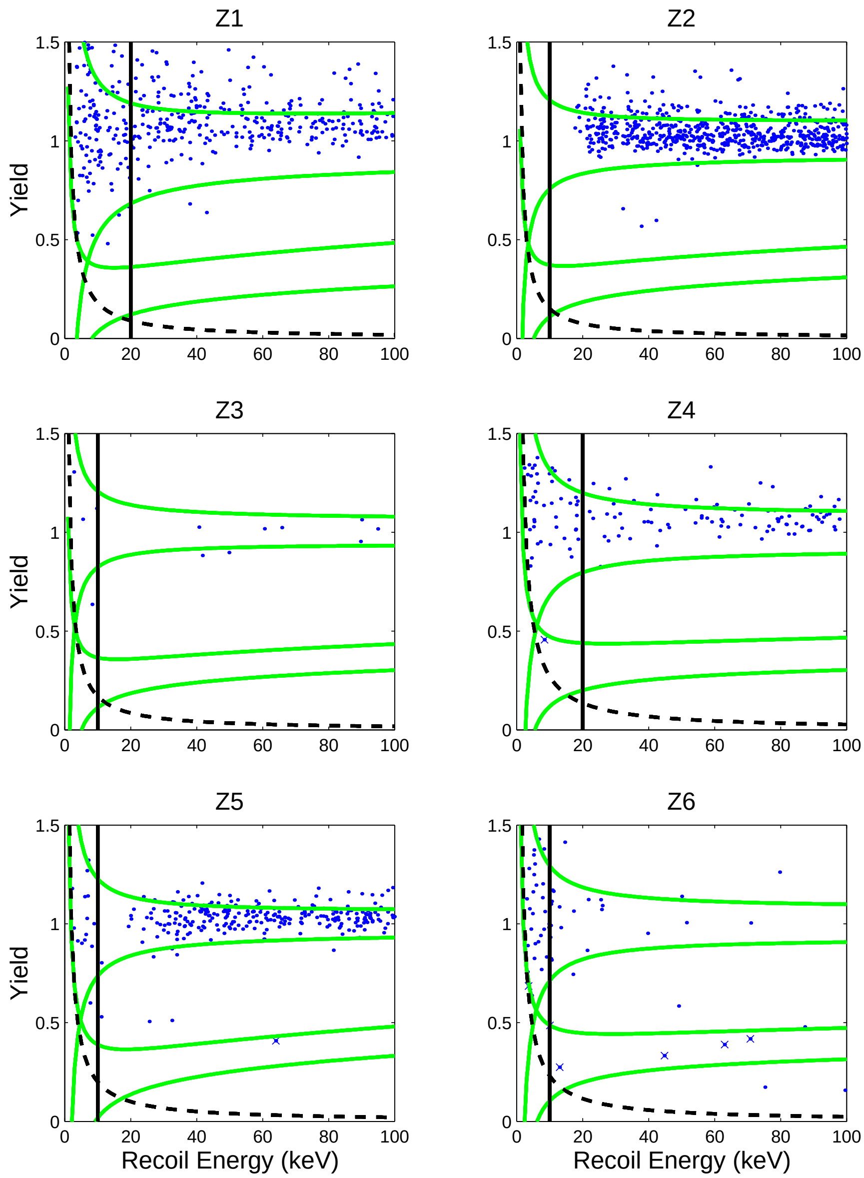


Figure 8.36: OF-only analysis: Single-scatter muon-anticoincident events observed in the WIMP-search run in all detectors of Tower 1, after applying the phonon-timing cut. The x's denote the NR events. The gray lines denote the NR and ER bands, the vertical black lines denote the analysis thresholds, and the dashed black lines denote the charge threshold cut.

Given these observations, and the estimates of the cut efficiencies shown in the

Figure 8.30, we can proceed with the calculation of the WIMP limit. The limit is calculated using the Optimal Interval method of S. Yellin [206]. In general, the Optimal Interval method relies on the idea that, given a distribution of background events in recoil-energy, the strongest limit on the signal cross-section can be placed using the interval in the recoil-energy that is the least contaminated by the background events. In the case of no observed events, the application simplifies to using the entire energy range considered, and the estimate is identical to the Poisson estimate.

Furthermore, to convert the observed rates into the WIMP-nucleon elastic scattering cross-section, we follow the calculation presented in [100] and in Chapter 1: we assume the Helm spin-independent form-factor, A^2 scaling, WIMP characteristic velocity $v_0 = 220$ km s^{-1}, mean Earth velocity $v_E = 232$ km s^{-1}, and $\rho = 0.3$ GeV c^{-2} cm^{-3}.

Figure 8.37 shows the limits based on the blind analysis with no observed events and the OF-only analysis with one observed event. Note that the two curves are very similar. They are a factor of ~ 4 lower than the previous best limit by Edelweiss [105] and a factor of ~ 8 lower than the previous CDMS result obtained at the shallow site at SUF [104]. The disagreement with the DAMA allowed region is obvious (for the spin-independent couplings and for the given halo model).

We conclude this Section with two cross-checks that support the hypothesis that most (if not all) of the events observed in the NR band prior to the phonon-timing cut are surface electron-recoil events.

First, we can use the regions A, B, and C described in the previous Section. The only exception is that we do not apply the phonon timing cut to the region A. Hence,

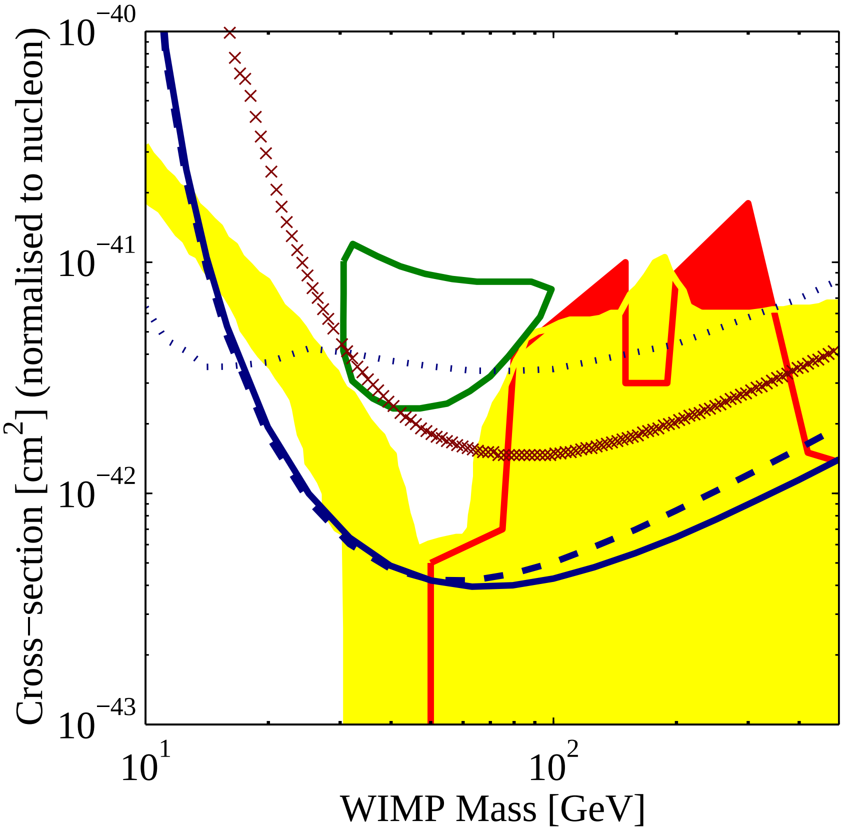


Figure 8.37: The Optimal Interval limit estimate based on the blind analysis is shown as a solid curve. The dashed curve denotes the limit corresponding to the OF-only analysis with one observed event. The dotted line denotes the CDMS result achieved at the shallow site at SUF [104], and the x's denote the result from the Edelweiss experiment [105]. The closed contour denotes the DAMA (1-4) 3σ signal region [106, 110]. Examples of allowed supersymmetric models are denoted as light gray [154] and dark gray [86] shaded regions.

AC/B is an estimate of the number of betas (i.e. surface events) that leaked into the NR band in the WIMP-search data, *prior* to imposing the phonon-timing cut. This number can be compared with the observed number of SSACNR events in the WIMP-search data. Table 8.4 compares the expected beta-leakage in the 10-100 keV bin, prior to imposing the phonon-timing cut, with the observed number of events in the NR band in the WIMP-search data. For detectors Z2, Z3, and Z5 the number of the observed events is consistent with the number of expected leaked beta-events, indicating that most (if not all) of the observed events are leaked betas. The agreement becomes worse for detectors Z1, Z4, and Z6 (Z6 had neutralization problems, so it is not clear that the test is meaningful for it). Since the observed number of events for these detectors is relatively large, and the statistical errors are relatively small, this discrepancy is suggesting a systematic error in the calculation, likely due to the systematic differences between the calibration and the WIMP-search data. As mentioned in the previous Section, there are possible differences in the angular distribution of the incoming betas in the WIMP-search data and in the calibration data. There are also possible systematic differences between the betas produced in the interactions of gammas with the material surrounding the detectors and the betas coming from the radioactive decays in the surface contamination of the detectors and the Tower.

Detector	10-20 keV	20-40 keV	40-100 keV	Expected	Observed
Z1	-	2.4 ± 0.7	4.2 ± 0.8	6.6 ± 1.1	16 ± 4.0
Z2	0.3 ± 0.2	0.4 ± 0.3	0.3 ± 0.2	1.0 ± 0.4	2 ± 1.4
Z3	1.5 ± 0.6	0.7 ± 0.4	0.5 ± 0.2	2.6 ± 0.8	3 ± 1.7
Z4	-	1.7 ± 0.6	1.6 ± 0.5	3.2 ± 0.8	9 ± 3.0
Z5	0.9 ± 0.5	1.1 ± 0.5	2.0 ± 0.6	4.1 ± 0.8	6 ± 2.4
Z6	8.6 ± 2.3	24.0 ± 3.5	18.8 ± 2.6	51.4 ± 4.9	146 ± 12.1

Table 8.4: Expected numbers of beta-events leaked into the NR band in the WIMP-search data are shown in the columns 2-5, while the observed number of SSACNR events is shown in the column 6. Columns 5 and 6 refer to the 10-100 keV bin (20-100 for Z1 and Z4). The estimate performs a weighed average of OF and F5 values.

For the second check, we calculate the fractions of the events between the NR and ER bands in the WIMP-search data (i.e. known surface events) that pass the phonon timing cut. We also calculate the fraction of the events inside the NR band (surface events or true nuclear-recoil events) in the WIMP-search data that pass the phonon-timing cut. If the set of events inside the NR band contained true nuclear-recoil events, than these two fractions would be different from each other (because the efficiency of the phonon-timing cut is different on the NR events and on the surface events). Of course, since we observe no events in the NR band after the phonon-timing cuts, we do not expect to find a discrepancy. Indeed, as shown in the Table 8.5, the agreement between the fractions calculated in this way is good.

Detector	Outside NR	Inside NR
Z1	$0.04^{+0.05}_{-0.03}$	$0^{+0.11}$
Z2	$0.21^{+0.16}_{-0.11}$	$0^{+0.6}$
Z3	$0^{+0.09}$	$0^{+0.46}$
Z4	$0^{+0.11}$	$0^{+0.18}$
Z5	$0.06^{+0.07}_{-0.04}$	$0^{+0.26}$
Z6	$0.02^{+0.02}_{-0.01}$	$0.03^{+0.02}_{-0.01}$

Table 8.5: Fractions of events in the WIMP-search data that pass the phonon-timing cut. Column 2 shows beta-events that fall between the NR and ER bands, and Column 3 shows events that fall inside the NR band. 10-100 keV bin is used for Z2, Z3, Z5, and Z6, and 20-100 keV for Z1 and Z4. 68% confidence limits are shown on the estimates.

8.10 Alternative Analysis

As we have seen in Chapters 4, 5, 6, and earlier in this Chapter, the ZIP detectors provide much information about the position and the nature of an interaction in the crystal. In particular, we have seen that the timing information stored in the phonon pulses can be effectively used to discriminate against the surface electron-recoil events, which (as shown in the last Section) are the dominant background at the deep site in Soudan mine.

However, we have not used all the information provided in the phonon channels. As mentioned at the end of Chapter 6, the phonon amplitudes can also be used to discriminate against the surface events. In this Section, we will try to incorporate this information as well. We will first describe the amplitude-based parameter pfracc, and then develop a new method of combining the various discrimination parameters, with the goal of achieving the best surface-event rejection possible. We note that the results described here are preliminary, as several important improvements are still possible.

8.10.1 Amplitude-Based Discrimination Parameter

As mentioned in Section 8.6.9, pfracc is defined as the ratio of the amplitudes in the primary phonon sensor (under which the event took place) and in the diametrically opposite phonon sensor. In Section 6.8, we have shown that this phonon-fraction parameter can be used to discriminate against the surface-events that take place on the phonon side of the detector - this result was based on the beta calibration performed at the UCB test facility, without applying the position-correction. As discussed in detail in Chapter 6, the surface-event problem is more prominent on the phonon side than on the charge side of the detector, so such parameter could be very important. We also note, however, that such parameter cannot have any rejection power at the center of the crystal - the phonon signal will be equally distributed among the four phonon sensors, regardless of the type of the interaction. However, the parameter is expected to become relevant $\gtrsim 1$ cm away from the center, where majority of the events takes place.

Figure 8.38 confirms this result in the first WIMP-search data from Soudan. In this case, position correction was applied to the phonon-fraction parameter. It is clear from Figure 8.38 that the phonon-fraction parameter is complementary to the phonon-timing information. In particular, a cut in the phonon-delay parameter pdelc, rejecting all betas, would have to be set at ~ 8 μs. However, a cut in the pfracc-pdelc plane, rejecting all betas, could preserve a large fraction of the neutrons with pdelc values as low as 4 μs.

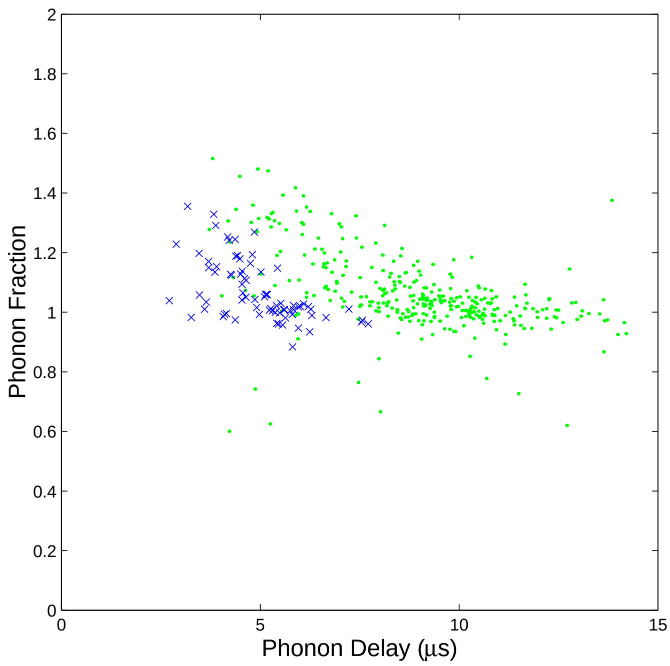


Figure 8.38: Phonon-fraction parameter (pfracc) is plotted against the delay of the maximum phonon pulse (pdelc) for the 20-40 keV bin of detector Z3. The neutrons from the ^{252}Cf calibration are shown as gray/green dots, and the betas from the ^{133}Ba calibration are shown as dark/blue x's.

It is, of course, possible to apply such a cut for all detectors and all energy bins, and then repeat the analysis discussed in the previous Sections. However, we will try to develop a new method that would incorporate this parameter with several other parameters, with the goal of optimally using all available information to reject the surface events.

Before we describe the new method, we point out another important property of pfracc. Figure 8.38 shows that there are neutron-events with particularly low values of pfracc, clearly away from the main distribution of events. Low values of pfracc are expected for the neutrons which multiply scatter inside the same detector - essentially, the phonon signal would be distributed more evenly among the four sensors than in the case of a single hit in one quadrant, so the *primary*/*opposite* ratio would be smaller. Hence, it should be possible to use pfracc to identify the internal multiple-scatter nuclear-recoil events, which would be yet another handle of distinguishing between the neutrons and WIMPs (WIMPs would not multiply scatter).

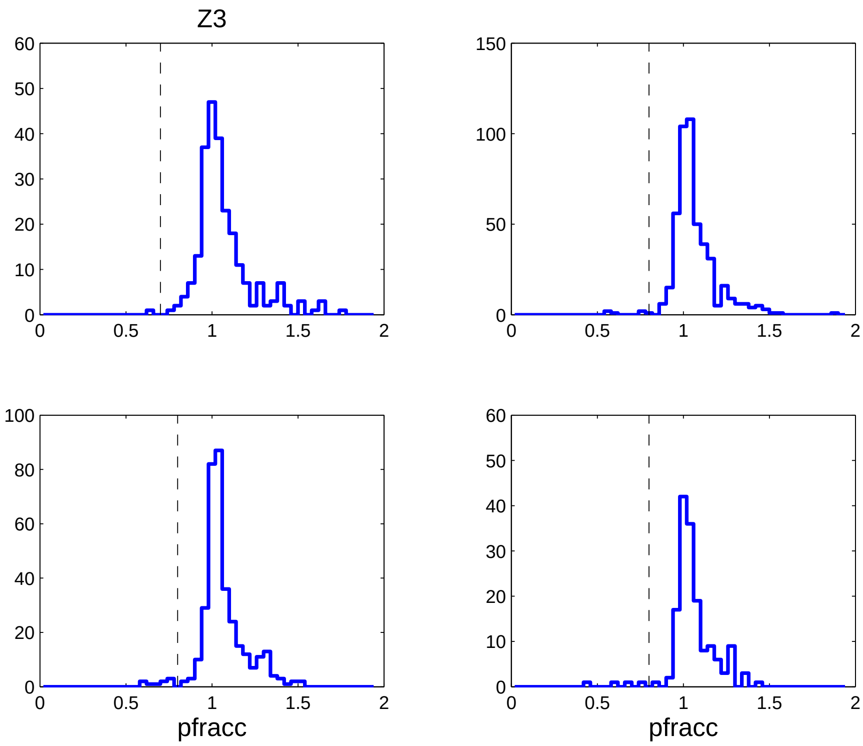


Figure 8.39: Histograms of the phonon-fraction parameter pfracc, in energy bins 5-10, 10-20, 20-40, 40-100 keV of detector Z3, for neutrons in the ^{252}Cf calibration. Events below the vertical lines are plotted in Figure 8.40.

As a cross-check, we identify the low-pfracc ^{252}Cf neutron events in four energy bins, as shown in Figure 8.39, and we plot them in the yield vs recoil-energy plane. Since the NR band is energy dependent, an internal multiple-scatter NR event is expected to have lower yield value than the single-scatter NR event of the same recoil-energy (on average). Indeed, as shown in Figure 8.40, the low-pfracc events tend to have low yield values, in agreement with the hypothesis that these are internal multiple-scatter NR events.

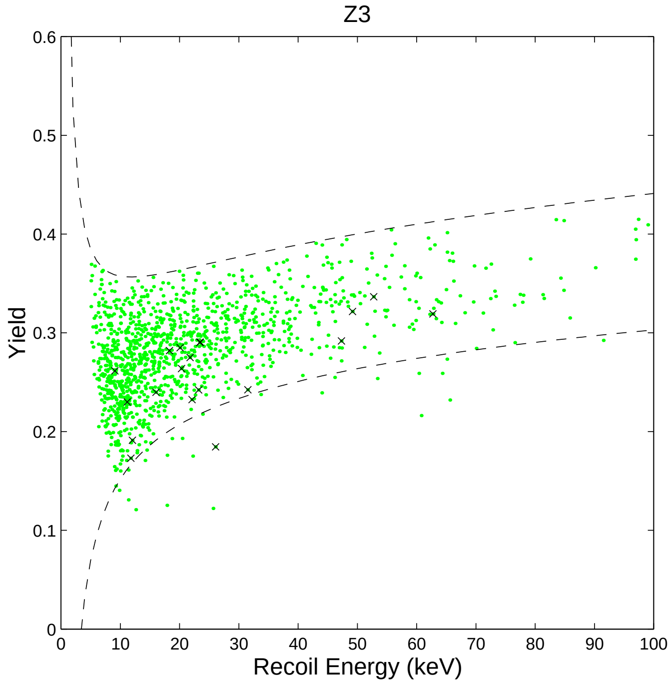


Figure 8.40: Yield vs recoil-energy for the ^{252}Cf neutrons (gray/green) of detector Z3. Events with low pfracc values, selected in the Figure 8.39, are shown as black x's.

8.10.2 The $r_b - r_n$ Method

The approach used so far in the analysis of the CDMS data can be summarized as follows:

1. Use the ionization yield to reject most of the electron-recoil events.

2. Apply a cut on the phonon pulse-shape to reject surface electron-recoil events that pass the yield cut.

However, this approach does not utilize all of the information provided by the ZIP detectors. In particular, the analysis of the first data from Soudan (presented above) did not use the pfracc parameter. Moreover, studies of the phonon-physics that determines the phonon pulse-shapes of the ZIP detectors are likely to offer additional parameters that can be used to further suppress the surface-event background. Hence, it is important to investigate how to optimally use the information stored in all these parameters.

In principle, one could take an approach along the lines of Chapter 5, where one estimates the distribution of surface events and nuclear-recoil events in all of the relevant parameters. However, this implies guessing a functional form for the distribution, which is non-trivial as many of the relevant parameters are correlated. Such approach could still be pursued if we could reduce the number of parameters to 2-3, while preserving the necessary information on the nature of the interaction.

In the approach described below, we use four parameters: the ionization yield y, the rise-time of the largest phonon pulse *pminrtc*, the delay (with respect to the charge pulse) of the largest phonon pulse *pdelc*, and the primary/opposite amplitude fraction $pfracc$. All of these parameters have been position-corrected. The idea is to define the "beta region" and the "neutron region" in this 4D parameter space. These regions can be defined using the gamma and neutron calibrations. We can then compute distances of each WIMP-search event from these two regions - r_b and r_n respectively. A beta-like event would have small r_b and large r_n, and vice-versa for a neutron-like event.

The calculation is done in several steps, and we outline them below.

1. We first need to scale all four parameters so that they have similar weights in the distance calculation. In particular, we define: $(y', pminrtc/9, pdelc/6, pfracc)$, where y' of an event is the ionization yield of the event scaled by the mean of the NR band at the event's energy. In this way, all parameters are scaled to 1.

2. We use the ^{133}Ba and ^{252}Cf calibrations to select beta and neutron events. Betas are

defined to be events in the ^{133}Ba calibration below 3σ of ER band, with $0 < y < 0.75$, that pass the charge threshold cut. We record the four parameters of these events in a table. We also record the radius r_d in the phonon delay plot for each event. We do this in the four standard energy bins (5-10, 10-20, 20-40, and 40-100 keV). Neutrons are defined to be events in the ^{252}Cf calibration that fall within 2σ of the NR band, and that pass the charge threshold cut. We record their parameters in a separate table.

3. When applying the tables, for each event we find 5 nearest neighbors in each of the (neutron and beta) tables and average the distance of the given event from those 5 neighbors - these are r_n and r_b respectively. The distance is defined to be euclidean with two subtleties. First, for events with $r_d > 15$ (6.5) for Ge (Si) we use only the table entries with $r_d > 15$ (6.5). Second, for events with $r_d < 15$ (6.5), we use only the table entries with $r_d < 15$ (6.5), and we do not use $pfracc$ in the distance calculation (i.e. we use only the other three parameters). These complications come from the fact that $pfracc$ is not a useful quantity for events with $r_d < 15$ (6.5) in terms of rejecting surface events - events that take place at the center of the crystal will have the phonon signal evenly distributed in the four sensors, regardless of the type of the interaction.

To illustrate the method, we plot r_b vs r_n for the ^{133}Ba and ^{252}Cf calibrations in the Figure 8.41. Note that all events seem to fall in a one-dimensional band - this remains to be true for all detectors, but it is somewhat less obvious at lowest energies. Also note that the $r_b - r_n$ parametrization allows directly rejecting the bulk electron-recoil events, as

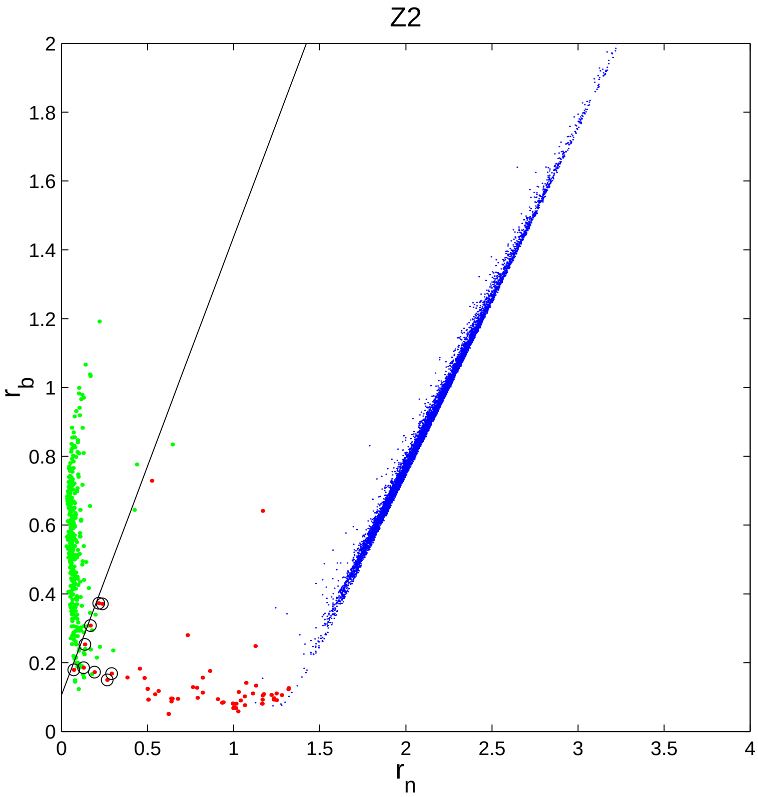


Figure 8.41: r_b vs r_n for the detector Z2 in the energy range 20-40 keV. The small dark/blue points are events above $> -3\sigma$ of the ER band (i.e. bulk electron-recoils). The large dark/red points are betas (as defined in the text) from the ^{133}Ba calibration. The circles denote ^{133}Ba calibration betas that fall in the 2σ NR band. The gray/green large points, close to the y-axis, are neutrons from the ^{252}Cf calibration. The straight line denotes a possible position of a cut.

the ionization-yield parameter is included in the analysis. In other words, it is possible to avoid calculating the NR bands, although in the analysis described below we decide to keep them. Finally, the separation between the neutrons and the surface electron-recoil events is fairly clean, although this varies among the detectors and energy bins.

8.10.3 Cuts, Efficiencies, and Leakage

Having computed the beta and neutron tables, and having defined the distance in the 4D parameter space, we can proceed to define the cut in the $r_b - r_n$ plane to reject the surface electron-recoil events. The goal, again, is to reject most of the surface events from the ^{133}Ba calibration, while preserving as much as possible of the neutron events from the ^{252}Cf calibration. Although this is similar to the objective of the phonon-timing cut 2 (discussed in Section 8.6.9), we do not perform a formal optimization - in most cases it is possible to set such cut by eye. We define the cut in the standard energy bins: 5-10, 10-20, 20-40, and 40-100 keV. We do not observe any dicontinuities in the cut efficiency on the neutrons, so we decide that smoothing the cut parameter against energy is not necessary. With such simple, non-optimized, definition of the cut, we obtain the cut efficiencies shown in the Figure 8.42. Note that the efficiency of the $r_b - r_n$ cut tends to be higher than the efficiency of the phonon-timing cut 2, especially at the lowest energies (where the WIMP-signal is expected to be the strongest).

We make several other improvements with respect to the blind analysis and the OF-based analysis discussed earlier. First, we use only the optimal-filter estimates of the charge-signal amplitudes. We then redefine the charge threshold cut. The definition of this cut is identical to the one performed blindly (see Section 8.6.4), but we use only the OF quantities. Hence, the cut parameter values are somewhat lower than for the blind case, giving somewhat higher efficiencies at the lowest energies. The comparison of the new charge-threshold cut efficiency and the blind charge-threshold cut efficiency (OF analysis) is shown in the Figure 8.43.

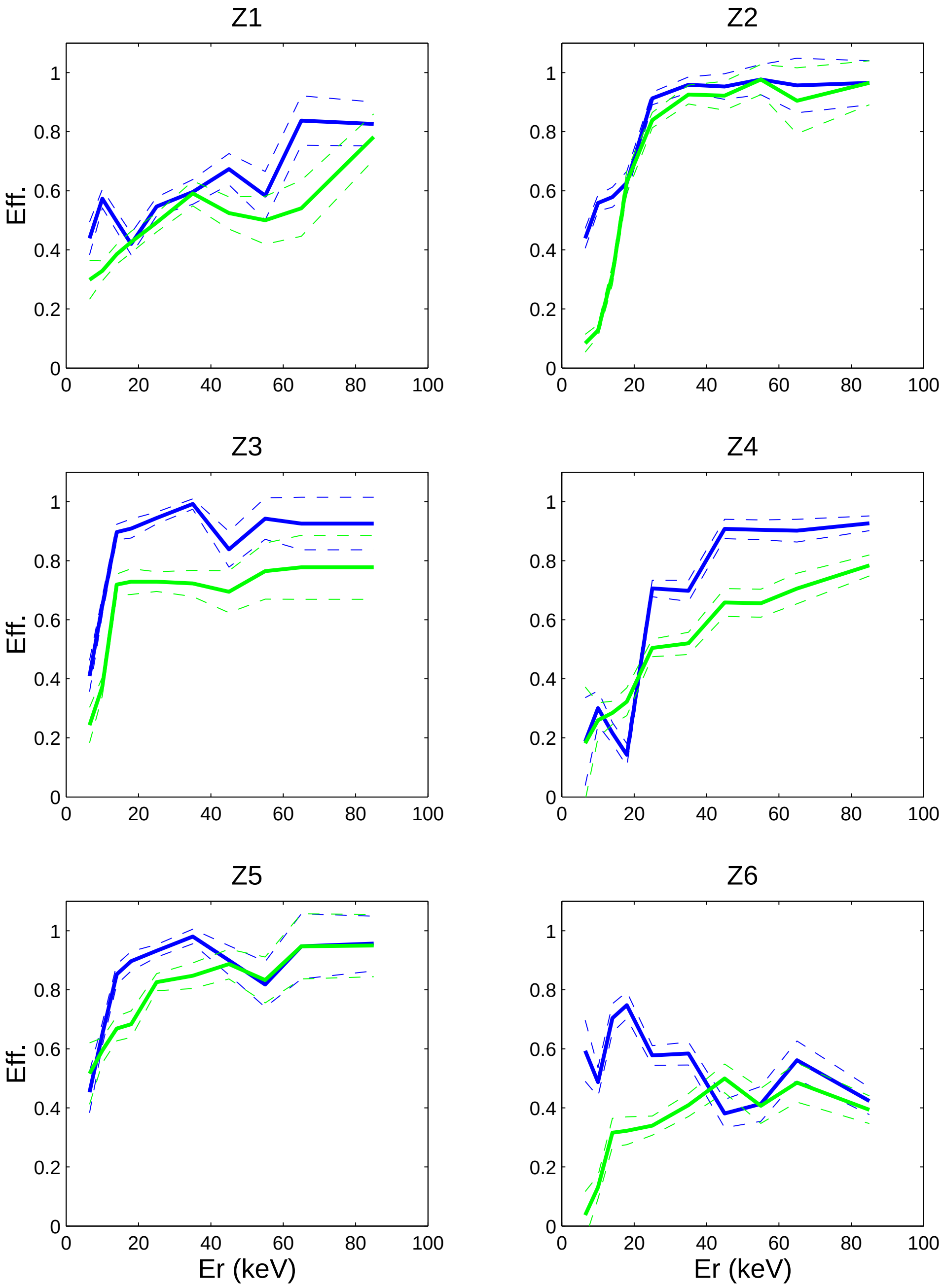


Figure 8.42: Efficiencies of the $r_b - r_n$ cut on ^{252}Cf neutrons are shown in dark/blue for all Tower 1 detectors. The gray/green lines show the efficiencies of the phonon-timing cut 2, as discussed in Section 8.7.6. Solid lines are the best estimates, and the dashed lines are 1σ statistical uncertainties.

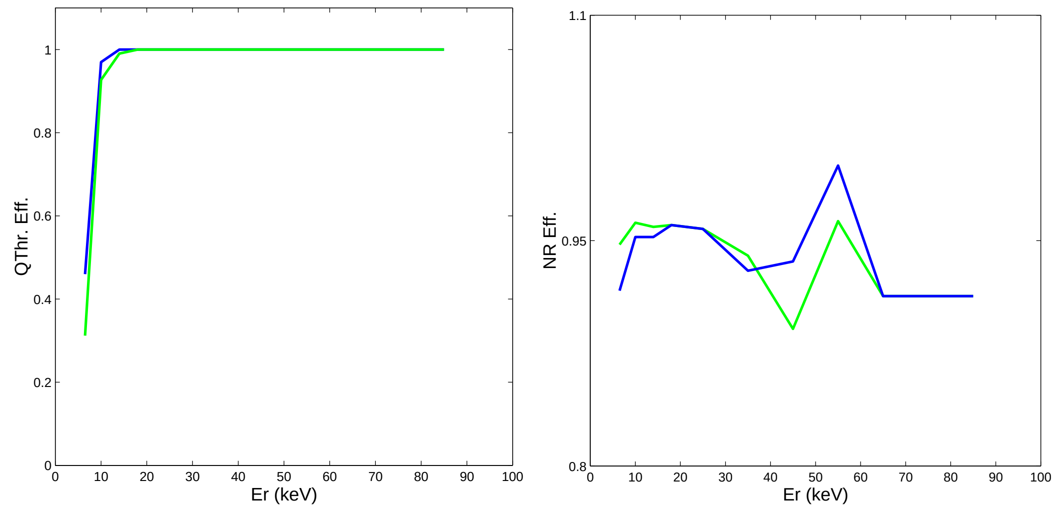


Figure 8.43: Left: Efficiencies of the charge threshold cut for the new OF analysis (dark/blue) and for the blind OF analysis (gray/green) are shown for the detector Z3. Right: Efficiencies of the NR band cut for the new OF analysis (dark/blue) and for the blind OF analysis (gray/green) are shown for the detector Z3.

We also recalculate the NR and ER bands, using the OF quantities only. The entire calculation is done in the same way as for the blind analysis (see Section 8.6.6). The comparison of the NR band cut efficiency for the new OF analysis and the blind OF analysis is shown in the Figure 8.43.

To calculate the overall cut efficiency, we use the new cut efficiencies for the charge threshold cut, the NR band cut, and the $r_b - r_n$ cut, along with the cut efficiencies estimated in the Section 8.7 for the charge χ^2 and phonon pre-trigger cuts, the outer charge-electrode cut, and the muon-veto cut (all for the OF-only case). Figure 8.44 shows the overall cut efficiency as the cuts are added one-at-a-time. This Figure is directly comparable with the Figure 8.30. Figure 8.45 compares the new overall efficiency with the previous efficiencies estimated in Section 8.7. Note that the improvement is significant, especially at the lowest energies. In particular, the 5-10 keV bin now seems to have reasonably high efficiencies, so

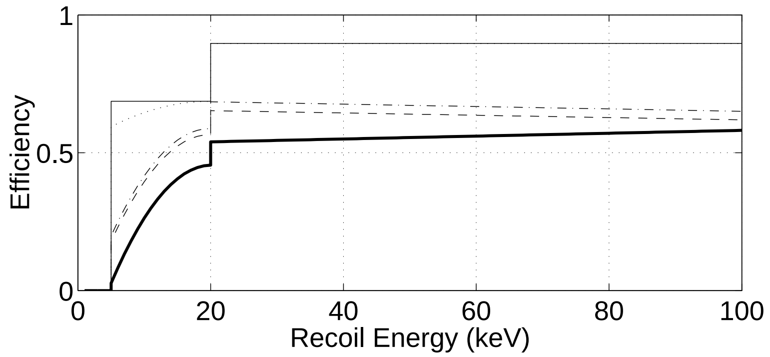


Figure 8.44: The cut efficiency for the $r_b - r_n$ analysis, averaged over the Ge detectors Z1, Z2, Z3, and Z5. Top-to-bottom: charge χ^2 cut, phonon pre-trigger std cut and muon veto cut (thin solid); also add charge threshold cut (dotted); also add the outer charge-electrode cut (dot-dashed); also add the NR band cut (dashed); also add the $r_b - r_n$ cut (thick solid) - this is the final efficiency of this analysis. The discontinuity at 20 keV is due to the analysis threshold of Z1.

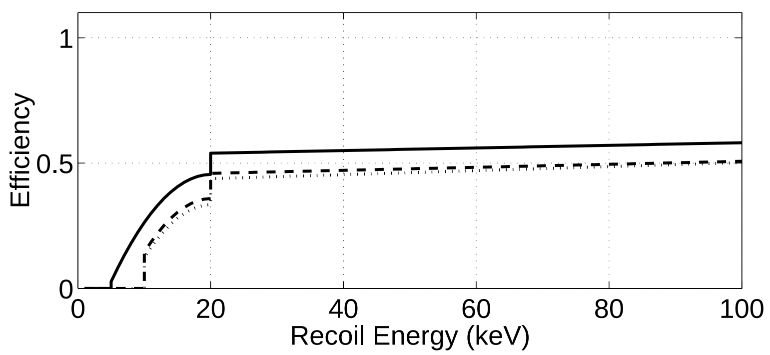


Figure 8.45: Final cut efficiencies for the three different analyses presented in this Chapter: Blind analysis (dotted), OF-based analysis (dashed), and the $r_b - r_n$ analysis (solid). Note that the $r_b - r_n$ analysis gives higher efficiency at all energies, and it allows reducing the analysis threshold to 5 keV.

we decide to lower the analysis threshold to 5 keV for Z2, Z3, and Z5. Z1 and Z4 remain at 20 keV, as the beta leakage was still a problem in these cases. Integrated over the 5-100 keV bin, this efficiency gives 27.4 days of effective exposure.

Having estimated the cut efficiencies, we can redo the estimates of the various backgrounds. The gamma background is essentially the same as for the blind analysis

($<$ 0.25 events at 90% confidence, Section 8.8.1), as still none of the gamma-calibration single-scatter events passes all the cuts. Similarly, the neutron background is very similar to the blind analysis (0.07 of external neutrons, Section 8.8.3), since the increase in the exposure is not sufficient to make a significant impact on this background. The surface-event background is the only background expected to change. The "naive" method described in the Section 8.8.2 yields 0.76 ± 0.79 surface events, expected to be misidentified as NR events, in the whole WIMP-search data. This is larger than the "naive" method prediction in the blind analysis, both because we are now including 5-10 keV bin, and because the surface-event leakage was somewhat higher in some higher-energy bins for some detectors (by chance, as we did not rigorously optimize the $r_b - r_n$ cut). We note, as discussed in detail in Section 8.8.2, that this method tends to underestimate the surface-event background, especially because it does not take into account the systematic error due to differences between the calibration and the WIMP-search data.

8.10.4 WIMP-Exclusion Limit

We now apply the $r_b - r_n$ cut to the WIMP-search data. We apply the new bands and charge-threshold cuts (defined using OF-only quantities, as discussed above). The rest of the cuts are as described in the Sections 8.6 and 8.9.

Figure 8.46 shows the WIMP-search data after applying the cuts, including the $r_b - r_n$ cut. In the Ge detectors Z2, Z3, and Z5, we set the analysis threshold at 5 keV, as discussed above. The analysis threshold of Z1 is left at 20 keV. With such thresholds, one event passes all cuts in Z5 at 64 keV - it is the same event appearing in the OF-based analysis presented in Section 8.9. Again, this event is consistent with the expected surface-

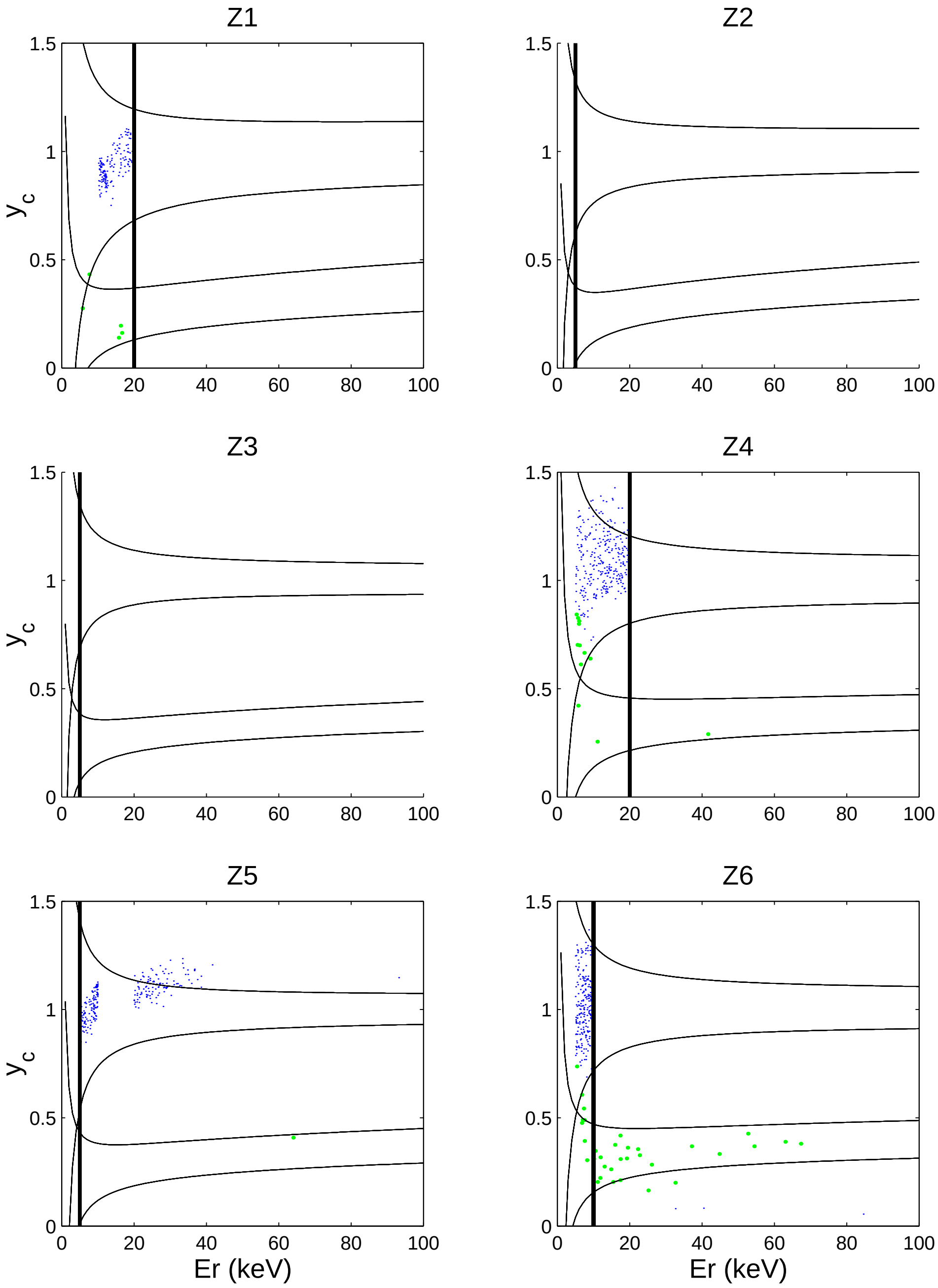


Figure 8.46: The yield vs recoil-energy plots for the WIMP-search data, for all detectors of Tower 1. We apply the new NR band and charge-threshold cuts (defined using OF-only quantities) and the $r_b - r_n$ cut.

event background, estimated to be 0.76 ± 0.79. Note, also, that for some energy bins in some detectors some of the bulk electron-recoils also pass the $r_b - r_n$ cut - this is because the slope of the straight-line cut in the $r_b - r_n$ plane (as shown in Figure 8.41) was sometimes set too low, allowing some of the electron-recoils to pass the cut. In this analysis, this is not a concern as we still impose the NR band cut - it would, clearly, be straightforward to avoid this problem (by imposing a more careful cut in the $r_b - r_n$ plane), if one wanted to avoid the bands.

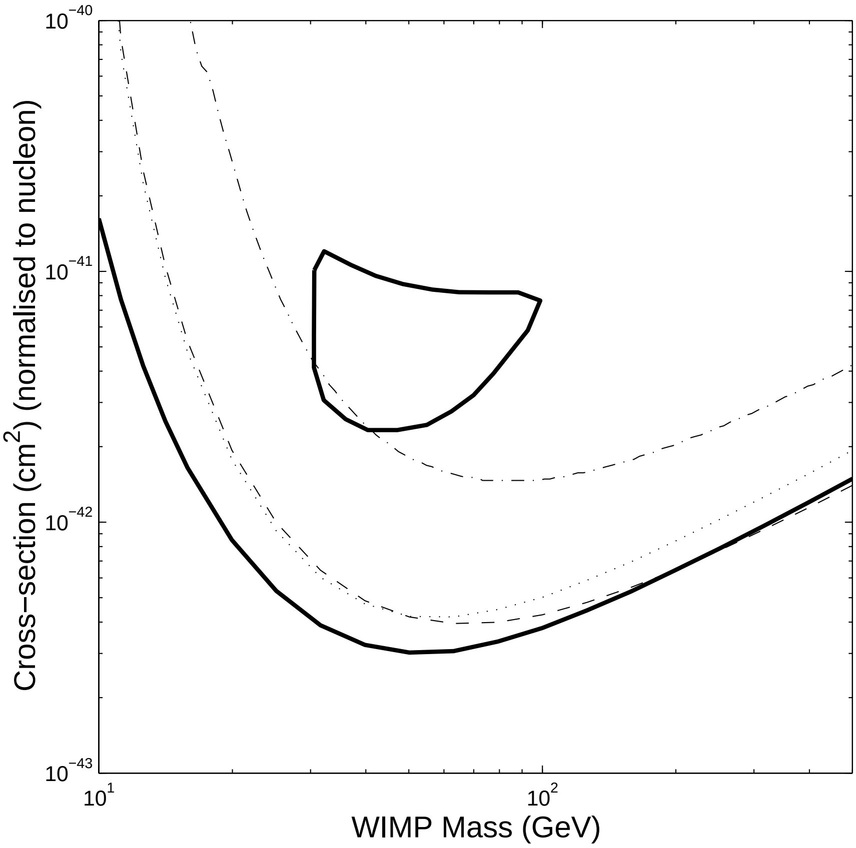


Figure 8.47: The preliminary Optimal-Interval limit based on the $r_b - r_n$ analysis is shown as a solid line. It is compared to the blind analysis limit (dashed) and the OF-only limit (dotted). Also shown are the Edelweiss result (dot-dashed) [105] and the DAMA (1-4) 3σ signal region [106, 110]. Note that the $r_b - r_n$ analysis yields a lower limit at the lowest masses, because the analysis threshold is reduced to 5 keV in most detectors.

Figure 8.47 shows the limit corresponding to the preliminary $r_b - r_n$ analysis described in this Section. Note that the limit is considerably lower at the lowest masses, primarily because the analysis threshold is reduced.

At least two aspects of the $r_b - r_n$ analysis presented here can be improved. First, no effort was made to optimize the relative weights of different parameters used in the calculation of r_b and r_n. It is likely that weighing more heavily the parameters that are more powerful in discriminating against surface events would further improve the surface-event rejection. Second, it has been observed that occasional events have anomalously large measurements in some of the parameters. Such cases are likely due to mis-estimates by the relevant algorithms, and should not be used in the calculation of r_b and r_n. However, this issue has not been addressed in the analysis so far.

One way to approach these issues, as suggested by B. Sadoulet, is to calculate the covariance matrix for the parameters used in the calculation of r_b and r_n, and then define the distance from the beta and neutron regions as the χ^2. This approach would automatically weigh different parameters appropriately, to allow optimal use of all available information.

It should be emphasized that the most important strength of the method described here lies in the ease of incorporating additional parameters, with additional information on the nature of the interaction. Furthermore, the method reduces the large number of initial parameters to only two (r_b and r_n), which significantly simplifies applications of the statistical methods described in the Chapter 5. In particular, it is difficult to guess the appropriate functional forms to describe distributions of betas and neutrons in the manydimensional parameter space (especially since many of the parameters are correlated), while it is much simpler to do it in the two-dimensional $r_b - r_n$ plane. These extensions of the $r_b - r_n$ method will be pursued in the future.

8.11 Implications and Future

We end this Chapter with a brief discussion of some of the implications of the new CDMS results. We also outline what is expected in the future of the CDMS experiment.

Figure 8.48 shows how the new limits discussed in the previous Sections compare to the supersymmetric models described in the Chapter 2. In both the mSUGRA and the general MSSM frameworks, these results rule out significant fractions of the allowed parameter space. Typically, the excluded models have large values of $\tan\beta$ (usually > 20) - such models are known to prefer relatively large values of $\sigma_{\chi-p}$. Also, the excluded models tend to have relatively low values of μ and M_2. As discussed in Chapter 2, models with large values of μ and M_2 tend to have lower values of $\sigma_{\chi-p}$. Similarly, Figure 8.37 shows that the new CDMS result excludes a large fraction of the low-neutralino-mass models, allowed if the requirement of the gaugino mass unification is dropped in the general MSSM framework [154].

CDMS started operating two Towers of detectors in February 2004. By the summer of 2004, CDMS is expected to at least double the exposure, compared to the first WIMP-search run. In absence of backgrounds, the CDMS sensitivity will improve linearly with the exposure. Eventually, however, the neutron background and the surface-event background will appear, at which point the sensitivity will improve only as the square-root of the

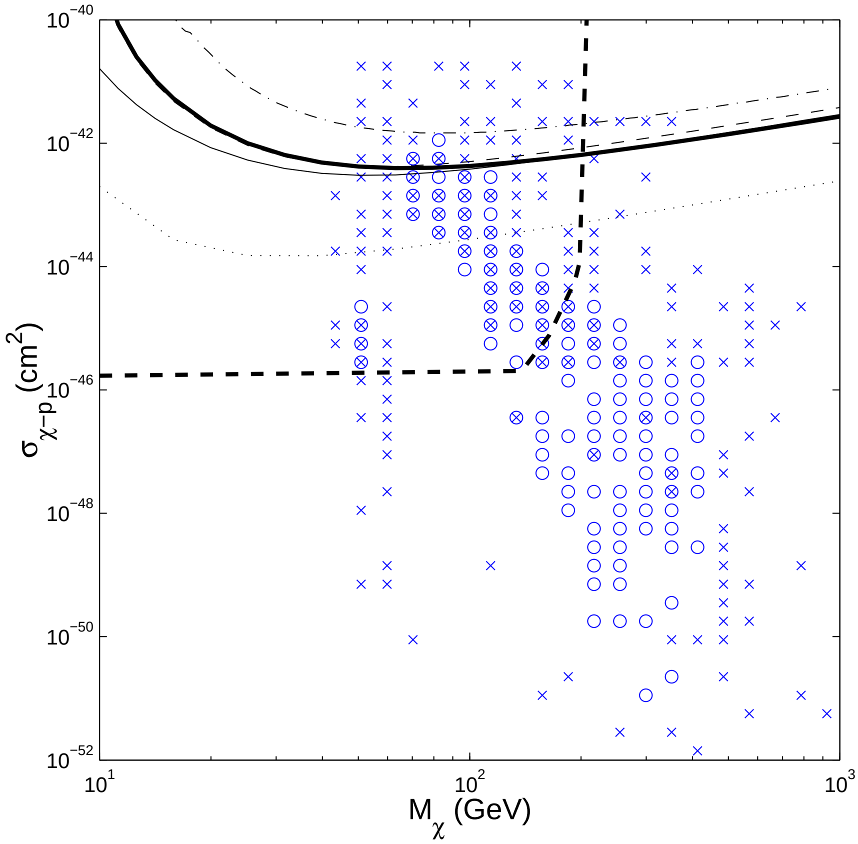


Figure 8.48: The supersymmetric models in the mSUGRA framework (o's) and in the general MSSM framework (x's), defined in the Equations 2.49 and 2.50, respectively, are shown in the $\sigma_{\chi-p} - M_\chi$ plane. The mSUGRA constraint on $m_{1/2}$ was relaxed to 1 TeV. We impose accelerator-based constraints and the relic density constraint $\Omega_\chi < 0.18$, based on the WMAP result. Models passing the constraint due to the anomalous magnetic moment of muon fall above the thick dashed line (the vertical part of this line also roughly denotes the mSUGRA region satisfying $m_{1/2} < 300$ GeV). The blind limit is shown as the thick solid line, the unblind OF-based limit is shown as the thin dashed line, the $r_b - r_n$ result is shown as the thin solid line, the Edelweiss result [105] is shown as the dot-dashed line, and the projected final CDMS limit at Soudan is shown as the dotted line.

exposure. These two types of backgrounds are expected to set the final sensitivity of CDMS II - the expected final sensitivity of CDMS II is also shown in the Figure 8.48. This sensitivity is expected to be reached by the end of 2005, by running a full complement of five Towers for more than one year.

As mentioned before, there are still some tools that could be used to battle the surface-event background. Namely, one can improve the $r_b - r_n$ method by adding new parameters and by weighing them appropriately. Additional information can be extracted by deploying methods along the lines of Chapter 5. In the very long run, to further suppress this type of background, changes in the experiment design will probably be necessary. For example, thicker detectors would reduce the fraction of the surface-to-bulk events, hence allowing stricter cuts to reject surface events. There may also be ways of improving the information on the interaction depth in the phonon signals by patterning both sides of the crystal with TESs. Another front that can be pushed is the contamination level. Studies of the detector surface contamination are already under way. Such studies could help identify the steps (or chemicals) in the detector fabrication that potentially contaminate the detectors. Also, adding a layer of ancient lead inside the cryostat would reduce the ambient gamma flux, and hence reduce the flux of betas ejected by gammas interacting in the material around the detectors.

The external neutron background is much more difficult to handle. In the long run, there are ideas to add an active neutron shield around the whole experimental setup. It may also be possible to use the existing muon-veto system to identify external neutrons, but this has not been studied in detail yet.

There are also ideas of scaling the CDMS experiment to the 1 ton scale. This would be a significant technological challenge from the points of view of the detector fabrication, the detector read-out, and the cryogenic operation. However, such increase in the target mass would allow improving the sensitivity down to $10^{-45} - 10^{-46}$ cm^2. Although in the general MSSM framework there are models with much smaller values of $\sigma_{\chi-p}$, the upper bound on $\Omega_\chi h^2$, the measurement of the anomalous magnetic moment of muon, and our sense of "naturalness", all prefer models with $\sigma_{\chi-p} > 10^{-46}$ cm^2. Hence, the 1-ton scale experiment would be able to explore a very interesting part of the supersymmetric parameter space.

Appendix A

Soudan R118 Data-sets

A.1 WIMP-search Data-sets

131011_1207	131011_1730	131011_2015	131012_1019	131012_1934
131013_1814	131014_1845	131015_1047	131015_1827	131016_1740
131017_0033	131017_1717	131018_2108	131019_0915	131020_0039
131020_0846	131020_1759	131021_2300	131022_1739	131023_2116
131024_1505	131024_1749	131025_0842	131025_1957	131026_0917
131026_1806	131027_1429	131027_1736	131031_1916	131101_0857
131101_1900	131102_0935	131102_2029	131103_1907	131104_1844
131106_1243	131106_1712	131107_0911	131107_1710	131108_2202
131109_1040	131109_2029	131110_1753	131111_1159	131111_1717
131112_0844	131112_1735	131113_1134	131113_1756	131114_0931
131114_1714	131115_0849	131115_1824	131116_0954	131116_1921

Table A.1: WIMP-search data-sets, part 1.

131117_0914	131117_1417	131117_1853	131118_1107	131118_1812
131119_1238	131121_1412	131123_1128	131123_1919	131124_1905
131125_1831	131126_0933	131126_1441	131126_1853	131127_0910
131127_1212	131127_2306	131128_0926	131128_2137	131129_1000
131129_2224	131130_0942	131130_2232	131201_1816	131202_0846
131202_1708	131203_1901	131204_1736	131205_1330	131205_1815
131206_0902	131206_1809	131207_0934	131207_1817	131208_0911
131208_2319	131209_1935	131210_1744	131211_1713	131212_1751
131213_0855	131213_1842	131214_0901	131214_2115	131215_1747
131218_1238	131218_1706	131219_1633	131219_1749	131220_0919
131220_1315	131220_2011	131221_1040	131221_1915	131222_0948
131222_1813	131223_1328	131223_1748	131224_0944	131224_2013
131225_1057	131225_2022	131226_1053	131226_2029	131227_1025
131227_1954	131228_1100	131228_2007	131229_0939	131229_1728
131230_0902	131230_1712	131231_0913	131231_2015	140101_0900
140101_2019	140102_0919	140102_2036	140103_0911	140103_2124
140104_0920	140104_2041	140105_1715	140105_2056	140106_1732
140107_1002	140107_1754	140108_0944	140108_1716	140109_0858
140109_1658	140110_0917	140110_1916	140111_0915	140111_1954

Table A.2: WIMP-search data-sets, part 2.

A.2 Neutron Calibration Data-sets

131125_1538	131219_1447	140105_1556

Table A.3: Neutron calibration data-sets, with ^{252}Cf source.

A.3 Gamma Calibration Data-sets

131209_1153	131210_1048	131210_1342	131211_0920	131211_1159
131211_1347	131212_0904	131212_1244	131215_0937	131215_1419
131216_0925	131204_1416	131205_0925		

Table A.4: Gamma calibration data-sets, with ^{133}Ba source.

Bibliography

[1] F. Zwicky, Helv. Phys. Acta **6**, 110 (1933).

[2] S. Golwala, *Exclusion Limits on the WIMP-Nucleon Elastic Scattering Cross Section from the Cryogenic Dark Matter Search*, PhD thesis, U.C. Berkeley, 2000.

[3] E. Kolb and M. Turner, *The Early Universe*, Addison-Wesley, Reading, MA, 1990.

[4] S. Carroll *et al.*, Annu. Rev. Astron. Astrophys. **30**, 499 (1992).

[5] M. Persic *et al.*, Mon. Not. R. Astron. Soc. **281**, 27 (1996).

[6] A. Borriello and P. Salucci, Mon. Not. R. Astron. Soc. **323**, 285 (2001).

[7] P. Salucci and A. Burkert, Astroph. J. Lett. **537**, 9 (2000).

[8] P. Salucci, Mon. Not. R. Astron. Soc. Lett. **320**, 1 (2001).

[9] D. Davis and R. W. III, Astroph. J. Lett. **470**, 35 (1996).

[10] M. Loewenstein and R. W. III, Astroph. J. **518**, 50 (1999).

[11] M. Loewenstein and R. Mushotzky, Nucl. Phys. B Proc. Suppl. **124**, 91 (2003).

[12] M. Baes and H. Dejonghe, Astroph. J. Lett. **563**, 19 (2001).

[13] M. Girardi *et al.*, Astroph. J. **505**, 74 (1998).

[14] M. Girardi *et al.*, Astroph. J. **530**, 62 (2000).

[15] A. Boviano and M. Girardi, Astroph. J. **585**, 205 (2003).

[16] N. Bahcall *et al.*, Astroph. J. Lett. **447**, 81 (1995).

[17] D. White *et al.*, Mon. Not. R. Astron. Soc. **292**, 419 (1997).

[18] M. Bartelmann and P. Schneider, Phys. Rep. **340**, 291 (2001).

[19] N. Weinberg and M. Kamionkowski, Mon. Not. R. Astron. Soc. **337**, 1269 (2002).

[20] J. Carlstrom, Annu. Rev. Astron. Astrophys. **40**, 643 (2002).

[21] E. Komatsu and U. Seljak, Mon. Not. R. Astron. Soc. **336**, 1256 (2002).

[22] L. Grego *et al.*, Astroph. J. **552**, 2 (2001).

[23] D. Tytler *et al.*, R. Swedish Acad. Sci. Physica Scripta T **85**, 12 (2000).

[24] S. Burles and D. Tytler, Astroph. J. **507**, 732 (1998).

[25] S. Burles *et al.*, Phys. Rev. Lett. **82**, 4176 (1999).

[26] J. O'Meara *et al.*, Astroph. J. **552**, 718 (2001).

[27] Y. Izotov and T. Thuan, Astroph. J. **500**, 188 (1998).

[28] Y. Izotov *et al.*, Astroph. J. **527**, 757 (1999).

[29] D. Balser *et al.*, Astroph. J. **510**, 759 (1999).

[30] S. Perlmutter *et al.*, Astroph. J. **517**, 565 (1999).

[31] A. Riess *et al.*, Astron. J. **116**, 1009 (1998).

[32] J. Tonry *et al.*, Astroph. J. **594**, 1 (2003).

[33] W. Percival *et al.*, Mon. Not. R. Astron. Soc. Lett. **327**, 1297 (2001).

[34] E. Bertschinger, astro-ph/9503125.

[35] M. Colless *et al.*, astro-ph/0305051.

[36] L. Verde *et al.*, Mon. Not. R. Astron. Soc. **335**, 432 (2002).

[37] M. Tegmark, Astroph. J. **606**, 702 (2004).

[38] A. Pope *et al.*, astro-ph/0401249.

[39] W. Freedman *et al.*, Astroph. J. **553**, 47 (2001).

[40] N. Bahcall *et al.*, Astroph. J. Lett. **485**, 53 (1997).

[41] N. Bahcall *et al.*, Astroph. J. **585**, 182 (2003).

[42] S. Borgani *et al.*, Astroph. J. **561**, 13 (2001).

[43] W. Hu *et al.*, Nature **386**, 37 (1997).

[44] C. Ma and E. Bertschinger, Astroph. J. **455**, 7 (1995).

[45] H. Kodama and M. Sasaki, Int. J. of Mod. Phys. A **1**, 265 (1986).

[46] G. Smoot *et al.*, Astroph. J. **360**, 685 (1990).

[47] M. Tegmark *et al.*, Astroph. J. Lett. **464**, 35 (1996).

[48] P. de Bernardis *et al.*, Nature **404**, 955 (2000).

[49] J. Ruhl *et al.*, Astroph. J. **599**, 786 (2003).

[50] A. Balbi *et al.*, Astroph. J. Lett. **545**, 1 (2000).

[51] A. Lee *et al.*, Astroph. J. Lett. **561**, 1 (2001).

[52] N. Halverson *et al.*, Astroph. J. **568**, 38 (2002).

[53] J. Kovac *et al.*, Nature **420**, 772 (2002).

[54] C. Kuo *et al.*, Astroph. J. **600**, 32 (2004).

[55] G. Hinshaw *et al.*, Astroph. J. Suppl. **148**, 135 (2003).

[56] D. Spergel *et al.*, Astroph. J. Suppl. **148**, 175 (2003).

[57] A. Kogut *et al.*, Astroph. J. Suppl. **148**, 161 (2003).

[58] M. Tegmark *et al.*, Phys. Rev. D **69**, 103501 (2004).

[59] D. Weinberg *et al.*, Astroph. J. **490**, 564 (1997).

[60] B. Sadoulet, Rev. of Mod. Phys. S **71**, 197 (1999).

[61] C. Alcock *et al.*, Astroph. J. **542**, 281 (2000).

[62] C. Alfonso *et al.*, Astron. and Astroph. **400**, 951 (2003).

[63] C. Alcock *et al.*, Astroph. J. Lett. **550**, 169 (2001).

[64] S. Fukuda *et al.*, Phys. Rev. Lett. **85**, 3999 (2000).

[65] S. Ahmed *et al.*, Phys. Rev. Lett. **92**, 181301 (2004).

[66] R. Croft *et al.*, Phys. Rev. Lett. **83**, 1092 (1999).

[67] R. Bradley *et al.*, Rev. of Mod. Phys. **75**, 777 (2003).

[68] J. Ellis and M. Gaillard, Nucl. Phys. B **150**, 141 (1979).

[69] R. Peccei and H. Quinn, Phys. Rev. Lett. **38**, 1440 (1977).

[70] S. Asztalos *et al.*, Phys. Rev. D **69**, 11101 (2004).

[71] http://www.cenbg.in2p3.fr/extra/groupes/astrop/celeste.html.

[72] http://www.astro.ucla.edu/ stacee/index.html.

[73] http://icrhp9.icrr.u-tokyo.ac.jp.

[74] http://www.astro.ucla.edu/ veritas.

[75] http://lpnp90.in2p3.fr/ cat/index.html.

[76] http://www.mpi-hd.mpg.de/hfm/HESS/HESS.html.

[77] http://www.mpi-hd.mpg.de/hfm/CT/ct_frame_start.html.

[78] http://hegra1.mppmu.mpg.de/MAGICWeb/.

[79] http://www.lanl.gov/milagro/.

[80] R. Hartman *et al.*, Astroph. J. Suppl. **123**, 79 (1999).

[81] http://www-glast.stanford.edu/.

[82] D. Hooper *et al.*, astro-ph/0404205.

[83] K. Kosack *et al.*, astro-ph/0403422.

[84] K. Tsuchiya *et al.*, astro-ph/0403592.

[85] S. Desai *et al.*, hep-ex/0404025.

[86] E. Baltz and P. Gondolo, Phys. Rev. D **67**, 063503 (2003).

[87] M. Boliev *et al.*, Proc. 24th ICRC (Rome) **1**, 722 (1995).

[88] M. Ambrosio *et al.*, Phys. Rev. D **60**, 082002 (1999).

[89] J. Ahrens *et al.*, Phys. Rev. D **66**, 032006 (2002).

[90] http://antares.in2p3.fr/Overview.

[91] http://www.ifh.de/baikal/baikalhome.html.

[92] S. Orito *et al.*, Phys. Rev. Lett. **84**, 1078 (2000).

[93] Y. Asaoka *et al.*, Phys. Rev. Lett. **88**, 051101 (2002).

[94] M. Kamionkowski and M. Turner, Phys. Rev. D **43**, 1774 (1991).

[95] E. Baltz and J. Edsjo, Phys. Rev. D **59**, 023511 (1999).

[96] S. Coutu *et al.*, Positron Measurements With the HEAT-pbar Instrument, in Proceedings of 27th ICRC (2001).

[97] S. Coutu *et al.*, Astropart. Phys. **11**, 429 (1999).

[98] D. Hooper *et al.*, hep-ph/0312076.

[99] M. Goodman and E. Witten, Phys. Rev. D **31**, 3059 (1985).

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

[101] R. Bernabei *et al.*, Phys. Lett. B **389**, 757 (1996).

[102] L. Baudis *et al.*, Phys. Rep. **307**, 301 (1998).

[103] R. Abusaidi *et al.*, Phys. Rev. Lett. **84**, 5699 (2000).

[104] D. Akerib *et al.*, Phys. Rev. D **68**, 082002 (2003).

[105] A. Benoit *et al.*, Phys. Lett. B **545**, 43 (2002).

[106] R. Bernabei *et al.*, Phys. Lett. B **480**, 23 (2000).

[107] S. Ahlen *et al.*, Phys. Lett. B **195**, 603 (1987).

[108] D. Caldwell *et al.*, Phys. Rev. Lett. **61**, 510 (1988).

[109] J. Morales *et al.*, Nucl. Phys. B Proc. Suppl. **118**, 524 (2003).

[110] R. Bernabei *et al.*, Riv. Nuovo Cim. **26**, 1 (2003).

[111] B. Ahmed *et al.*, Nucl. Phys. B Proc. Suppl. **124**, 193 (2003).

[112] J. Morales *et al.*, Nucl. Phys. B Proc. Suppl. **118**, 525 (2003).

[113] S. Hart, Nucl. Phys. B Proc. Suppl. **110**, 91 (2002).

[114] M. Altmann *et al.*, astro-ph/0106314.

[115] M. Barnabe-Heider *et al.*, hep-ex/0312049.

[116] J. Wess and J. Bagger, *Supersymmetry and Supergravity*, Princeton University Press, Princeton, NJ, 1983.

[117] J. Bagger, hep-ph/9604232.

[118] M. Drees, hep-ph/9611409.

[119] H. Haber and G. Kane, Phys. Rep. **117**, 75 (1985).

[120] L. Girardello and M. Grisaru, Nucl. Phys. B **194**, 65 (1982).

[121] W. Hollik, R. Ruckl, and J. Wess, *Phenomenological Aspects of Supersymmetry*, Springer-Verlag, Berlin, Heidelberg, 1992.

[122] M. Drees, M. M. Nojiri, D. Roy, and Y. Yamada, Phys. Rev. D **56**, 276 (1997).

[123] G. Jungman, M. Kamionkowski, and K. Griest, Phys. Rep. **267**, 195 (1996).

[124] M. Drees and M. Nojiri, Phys. Rev. D **48**, 3483 (1993).

[125] K. Griest and D. Seckel, Phys. Rev. D **43**, 3191 (1991).

[126] J. Edsjo and P. Gondolo, Phys. Rev. D **56**, 1879 (1997).

[127] P. Gondolo and G. Gelmini, Nucl. Phys. B **360**, 145 (1991).

[128] K. Griest, Phys. Rev. D **58**, 2357 (1988).

[129] M. Kamionkowski and M. S. Turner, Phys. Rev. D **42**, 3310 (1990).

[130] P. Gondolo *et al.*, astro-ph/0012234.

[131] G. Belanger, Comput. Phys. Commun. **149**, 103 (2002).

[132] V. Mandic *et al.*, hep-ph/0008022.

[133] V. Barger *et al.*, Phys. Rev. D **47**, 1093 (1993).

[134] V. Barger *et al.*, hep-ph/0003154.

[135] R. Arnowitt and P. Nath, Phys. Rev. D **46**, 3981 (1992).

[136] A. Brignole *et al.*, Phys. Lett. B **371**, 123 (1991).

[137] J. Ellis *et al.*, Phys. Lett. B **257**, 83 (1991).

[138] M. Dress and M. Nojiri, Phys. Rev. D **45**, 2482 (1992).

[139] L. Bergstrom and P. Gondolo, Astropart. Phys. **5**, 263 (1996).

[140] A. Corsetti and P. Nath, hep-ph/0003186.

[141] U. Chattopadhyay *et al.*, Phys. Rev. D **60**, 063505 (1999).

[142] R. Arnowitt and P. Nath, Phys. Rev. D **54**, 2374 (1996).

[143] R. Barbieri *et al.*, Nucl. Phys. B **313**, 725 (1989).

[144] P. Nath and R. Arnowitt, Phys. Lett. B **336**, 395 (1994).

[145] P. Nath and R. Arnowitt, Phys. Rev. Lett. **74**, 4592 (1995).

[146] F. Borzumati *et al.*, Phys. Rev. D **51**, 341 (1995).

[147] J. Ellis *et al.*, Phys. Rev. D **62**, 075010 (2000).

[148] J. Ellis *et al.*, Phys. Rev. D **67**, 123502 (2003).

[149] J. Ellis *et al.*, Phys. Lett. B **481**, 304 (2000).

[150] J. Ellis *et al.*, Phys. Rev. D **58**, 095002 (1998).

[151] J. Ellis *et al.*, Phys. Lett. B **532**, 318 (2002).

[152] J. Ellis *et al.*, Phys. Rev. D **63**, 065016 (2001).

[153] Y. Kim *et al.*, **0212**, 034 (2002).

[154] A. Bottino *et al.*, Phys. Rev. D **69**, 037302 (2004).

[155] H. Brown *et al.*, Phys. Rev. Lett. **86**, 2227 (2001).

[156] G. Bennett *et al.*, Phys. Rev. Lett. **89**, 101804 (2002).

[157] M. Davier *et al.*, Eur. Phys. J. C **27**, 497 (2003).

[158] D. Kosower *et al.*, Phys. Lett. B **133**, 305 (1983).

[159] T. Moroi, Phys. Rev. D **53**, 6565 (1996).

[160] A. Czarnecki and W. Marciano, Nucl. Phys. B **76**, 245 (1999).

[161] A. Czarnecki and W. Marciano, Phys. Rev. D **64**, 013014 (2001).

[162] M. Drees *et al.*, Phys. Rev. D **63**, 115009 (2001).

[163] E. Baltz and P. Gondolo, Phys. Rev. Lett. **86**, 5004 (2001).

[164] R. Arnowitt *et al.*, Phys. Lett. B **505**, 177 (2001).

[165] J. Ellis *et al.*, Phys. Lett. B **508**, 65 (2001).

[166] T. Perera, *The Limiting Background in Dark Matter Search at Shallow Depth*, PhD thesis, Case Western Reserve University, 2001.

[167] D. Seitz, private communication.

[168] D. Akerib *et al.*, in preparation.

[169] D. S. Akerib *et al.*, Nucl. Instrum. Meth. A **400**, 181 (1997).

[170] T. A. Shutt, *A Dark Matter Detector Based on the Simultaneous Measurement of Phonons and Ionization at 20 mK*, PhD thesis, U.C. Berkeley, 1993.

[171] D. Yvon *et al.*, Nucl. Instrum. Meth. A **368**, 778 (1996).

[172] T. Minotani, K. Enpuku, and Y. Kuroki, J. Appl. Phys. **82**, 457 (1997).

[173] S. W. Nam, *Development of Phonon-Mediated Cryogenic Particle Detectors with Electron and Nuclear Recoil Discrimination*, PhD thesis, Stanford University, 1998.

[174] S. Tamura, Monte carlo study of effective decay time of high-frequency phonons, in *Phonon Scattering in Condensed Matter V. Proceedings of Fifth International Conference*, edited by A. Anderson and J. Wolfe, page 356, Berlin, 1986, Springer-Verlag.

[175] S. Tamura, J. Shields, and J. Wolfe, Phys. Rev. B **44**, 3001 (1991).

[176] R. Clarke, *An Athermal Phonon Mediated Dark Matter Detector with Surface Event Discrimination*, PhD thesis, Stanford University, 1999.

[177] J. Wolfe, *Imaging Phonons*, Cambridge University Press, 1998.

[178] B. Neganov and V. Trofimov, Pis'ma v Zhurnal Experimental'noi i Teoreticheskoi Fiziki **28**, 356 (1978).

[179] P. Luke, J. Appl. Phys. **64**, 6858 (1988).

[180] B. Cabrera, J. Low Temp. Phys. **93**, 365 (1993).

[181] N. Booth, Appl. Phys. Lett. **50**, 293 (1987).

[182] T. Saab, *Search for Weakly Interacting Massive Particles with the Cryogenic Dark Matter Search Experiment*, PhD thesis, Stanford University, 2002.

[183] S. Kaplan *et al.*, Phys. Rev. B **14**, 4854 (1976).

[184] A. Miller, *Development of a Broadband Optical Spectrophotometer Using Superconducting Transition-Edge Sensors*, PhD thesis, Stanford University, 2001.

[185] T. Saab *et al.*, Nucl. Instrum. Meth. A **444**, 300 (2000).

[186] B. Cabrera *et al.*, Nucl. Instrum. Meth. A **444**, 304 (2000).

[187] J. Hellmig *et al.*, Nucl. Instrum. Meth. A **444**, 308 (2000).

[188] P. Brink, Detector Technical Review, Internal CDMS Note, 2001.

[189] B. Young *et al.*, J. Appl. Phys. **86**, 6975 (1999).

[190] P. Brink *et al.*, Determination of the tc distribution for 1000 transition edge sensors, in *Low Temperature Detectors, Ninth International Workshop on Low Temperature Detectors*, edited by F. Porter, D. MacCammon, M. Galeazzi, and C. Stahle, page 529, Madison, WI, 2001, AIP Conference Proceedings 605.

[191] V. Mandic *et al.*, Position dependence in the cdms ii zip detectors, in *Low Temperature Detectors, Ninth International Workshop on Low Temperature Detectors*, edited by F. Porter, D. MacCammon, M. Galeazzi, and C. Stahle, page 509, Madison, WI, 2001, AIP Conference Proceedings 605.

[192] G. Wang, *Position Information*, Internal R118 Ebook Note #92, 2004.

[193] D. Driscoll, *Development and Performance of Detectors for the Cryogenic Dark Matter Search Experiment with an Increased Sensitivity Based on a Maximum Likelihood Analysis of Beta Contamination*, PhD thesis, Case Western Reserve University, 2004.

[194] R. Gaitskell, Nucl. Phys. B Proc. Suppl. **51**, 279 (1996).

[195] B. Serfass, *Correction Tables*, Internal G31 Ebook Note #43, 2003.

[196] R. Bunker, private communication.

[197] W. Ogburn, *Calibrating and Correction qi, qo*, Internal R118 Ebook Note #114, 2004.

[198] D. Day and L. Baudis, ^{133}*Ba Source: Improved MC Simulation vs Data*, Internal R118 Ebook Note #108, 2004.

[199] http://www.webelements.com.

[200] S. Kamat, *Checking the Efficiencies of the Data Cuts Using the External Neutron Simulations*, Internal R118 Ebook Notes #192 and #193, 2004.

[201] G. Feldman and R. Cousins, Phys. Rev. D **57**, 3873 (1998).

[202] R. Schnee, private communication.

[203] C. Chang, *Estimated R118 Electron Recoil Misidentification*, Internal R118 Ebook Note #217, 2004.

[204] B. Sadoulet, *Number of Expected Betas in the WIMP Search Data and Leakage Probability*, Internal R118 Ebook Note #214, 2004.

[205] D. Akerib *et al.*, astro-ph/0405033.

[206] S. Yellin, Phys. Rev. D **66**, 32005 (2002).