Abstract of "Cryogenic Dark Matter Search (CDMS II) - Application of Neural Networks and Wavelets to Event Analysis", by Michael J. Attisha, Ph.D., Brown University, May 2006.

The Cryogenic Dark Matter Search (CDMS) experiment is designed to search for dark matter in the form of Weakly Interacting Massive Particles (WIMPs) via their elastic scattering interactions with nuclei. This dissertation presents the CDMS detector technology and the commissioning of two towers of detectors at the deep underground site in Soudan, Minnesota. CDMS detectors comprise crystals of Ge and Si at temperatures of 20 mK which provide $\sim$keV energy resolution and the ability to perform particle identification on an event by event basis. Event identification is performed via a two-fold interaction signature; an ionization response and an athermal phonon response. Phonons and charged particles result in electron recoils in the crystal, while neutrons and WIMPs result in nuclear recoils. Since the ionization response is quenched by a factor $\sim 3(2)$ in Ge(Si) for nuclear recoils compared to electron recoils, the relative amplitude of the two detector responses allows discrimination between recoil types. The primary source of background events in CDMS arises from electron recoils in the outer 50 μm of the detector surface which have a reduced ionization response. We develop a quantitative model of this 'dead layer' effect and successfully apply the model to Monte Carlo simulation of CDMS calibration data. Analysis of data from the two tower run March-August 2004 is performed, resulting in the world's most sensitive limits on the spin-independent WIMP-nucleon cross-section, with a 90% C.L. upper limit of $1.6 \times 10^{-43}\,\mathrm{cm}^2$ on Ge for a 60 GeV WIMP. An approach to performing surface event discrimination using neural networks and wavelets is developed. A Bayesian methodology to classifying surface events using neural networks is found to provide an optimized method based on minimization of the expected dark matter limit. The discrete wavelet analysis of CDMS phonon pulses improves surface event discrimination in conjunction with the neural network analysis, giving a 20% improvement to the expected and final WIMP-nucleon cross-section upper limits.

Cryogenic Dark Matter Search (CDMS II) - Application of Neural Networks and Wavelets

to Event Analysis

by

Michael J. Attisha

M. Phys. (Hons), Warwick University, UK, 1999

Sc. M., Brown University, 2004

A dissertation submitted in partial fulfillment of the

requirements for the Degree of Doctor of Philosophy

in the Department of Physics at Brown University

Providence, Rhode Island

May 2006

This dissertation by Michael J. Attisha is accepted in its present form by

the Department of Physics as satisfying the

dissertation requirement for the degree of Doctor of Philosophy.

Date ______________ ________________________________

Professor Richard J. Gaitskell, Advisor

Recommended to the Graduate Council

Date ______________ ________________________________

Professor Ian Dell'Antonio, Reader

Date ______________ ________________________________

Professor Robert Lanou, Reader

Approved by the Graduate Council

Date ______________ ________________________________

Sheila Bonde, Dean of the Graduate School

Vita

Michael Attisha was born in Beverley, England on May 30th 1977. He studied at Warwick University and graduated with a First Class Masters Degree with Honors in July 1999. After spending a year in Australia working as a consultant, Michael began graduate school at Brown University in August 2000. During the first two years of graduate school Michael taught basic physics and astronomy, and he began working with Prof. Gaitskell and the CDMS collaboration in May 2002.

Acknowledgements

I would first like to thank my readers Profs. Bob Lanou and Ian Dell'Antonio for their valuable input on this dissertation. I would also like to extend my heartfelt thanks to my advisor Prof. Rick Gaitskell. A more enthusiastic and dedicated advisor one could not find. His insights into the many problems that we have studied over the last few years have been enormously useful in my development as a scientist.

In a collaborative experiment such as CDMS, there are many people to thank and many people to admire. I would like to thank the CDMS 'management' - Dan Bauer, Bernard Sadoulet, Blas Cabrera, Dan Akerib and David Caldwell - for their tireless efforts in running this experiment through thick and thin. My thanks to Richard Schnee, who is probably getting tired of being referred to a role model for those of us who do data analysis, but who has my admiration and gratitude nonetheless.

There's a strange bond that we share with those colleagues whom we 'served' with at the underground site in Soudan, and to that end I'd like to thank the following people for making the time we spent in the mine together a whole lot more enjoyable: Angie, Betty, Cathy, Clarence, Darren, Dennis, Erik, Gensheng, Jeff, Jodi, Joel, Kyle, Long, Mike, Nader, Paul, Samuel, Sharmila, Ray, Rupak, Vuk and Walter. Special thanks to Jim Beaty who works so hard on everything at Soudan.

Thanks also to John-Paul Thompson, with whom I worked on much of the surface event analysis described in Chapter 5 of this thesis. He and I shared that strange $200 cab ride on our first trip to the Soudan mine and it's fair to say that nobody but John-Paul would have spent the entire journey with his GPS attached to the roof of the taxi.

Even though she didn't thank me in the Acknowledgements of her thesis, I wish to thank my sister Helen for her support. I also want to make her feel slightly guilty.

Finally, I wish to thank my parents. All that I have achieved, I share with you. Without your endless support, love, and understanding I would not be writing these words.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

At the time of writing in late 2005, the field of cosmology finds itself in a peculiar situation. On one hand, there is a tremendous amount of concordance; the ΛCDM model provides an unprecedented fit to experimental data, yet on the other hand, the model includes two components that we know little about; dark matter and dark energy.

The fact that the luminous matter in the universe comprises $\lesssim 1\%$ of the total energy density was first suggested by studies made by Zwicky in the early 1930's [1]. The observed velocity dispersion relations of eight galaxies in the Coma cluster suggested that the luminous matter comprised $\sim 0.5\%$ of the total mass based solely on gravitational dynamics.

Analysis of rotation curves of spiral galaxies also reveals the presence of an unseen mass component within each galaxy. Based on the luminous mass distribution, we would expect that stars at radii where the luminous matter ends to have velocities of the form $v(r) \sim 1/\sqrt{r}$. Yet, repeated observations of many galaxies show that the stellar velocities asymptote to $v \sim 100 - 300$ km/s for arbitrarily large radii, suggesting the presence of a large halo of unseen mass.

Given these observations, the 'missing mass' could well be baryonic in nature, but studies of Big Bang Nucleosynthesis (BBN) put strong limits on primordial production of light elements given the observed abundances today. This suggests that the baryonic component is $\sim 4\%$ of the critical energy density ($\Omega_b \approx 0.04$) - clearly too low for baryons to form the majority of the matter in the universe. Furthermore, precision measurement of the Cosmic Microwave Background (CMB) radiation, which provides a snapshot of the baryonic matter distribution at $\sim 300,000$ years after the Big Bang, strongly supports the BBN conclusion that $\Omega_b \approx 0.04$.

Further studies of the CMB, combined with surveys of Type-Ia supernova, gravitational lensing, and clusters/superclusters imply that the total matter energy density, $\Omega_m \approx 0.26$ and hence, a non-baryonic matter component $\Omega_{nbm} \approx 0.22$. Studies also show that $\Omega_{\text{total}} \approx 1$ which includes a non-matter energy density of the universe ('dark energy'), $\Omega_\Lambda \approx 0.74$, associated with the cosmological constant; i.e. the accelerating expansion of the universe.

The lines of evidence that point to these conclusions will be reviewed in more detail in subsequent sections.

1.1 Evidence for Non-Baryonic Dark Matter

1.1.1 Galactic Rotation Curves

In a spiral galaxy, we would expect stars at radii beyond the visible matter distribution to have a velocity distribution that scales as $v(r) \sim 1/\sqrt{r}$. However, analysis of the rotation curves (RCs) of spiral galaxies reveals that stellar velocities are constant at radii well ouside the region in which most of luminous material resides. The RCs are constructed by measurement of the 21cm H lines, and suggest the presence of an unseen dark matter halo that extends out beyond the luminuous matter. Early measurements of galaxy RCs used nebular emission lines (Hα) or stellar absorption lines (NaI, CaFe).

A comprehensive study of about 1100 optical and radio RCs, covering absolute magnitudes from -16 to -22, and extending RC data out to 1.5-2 optical radii, was performed by Persic, Salucci and Stel [2] in 1996. They show that spiral RCs can be parameterized by a single parameter such as the galaxy's luminosity, resulting in a 'universal' RC curve. This implies several interesting scaling properties between the dark and luminous galactic structure parameters: 1) The dark matter to luminous matter (M_{dm}/M_L) mass ratio scales inversely with luminosity; 2) The total halo mass scales as $L^{0.5}$, 3) The central halo density scales as $L^{-0.7}$. These relationships have interesting implications for the role that the dark matter plays during the formation of galaxies. Figure 1.1 shows the fitted RCs for galaxies from the survey for eleven different absolute magnitudes.

Since elliptical and lenticular (S0) galaxies do not have such ordered orbital motions, one cannot use rotation curves to examine their mass distribution. Instead, the galaxies can be approximated as a gas in thermal equilibrium, allowing the comparison of the velocity dispersion ('temperature') of stars σ^2, with the temperature of the interstellar gas T. In the absence of dark matter, these two populations reside in the same gravitational well, and therefore we would expect $T \sim \sigma^2$. Therefore, the amount of dark matter can be estimated by examining a divergence from this relationship. Typically, it is found that $T \sim \sigma^{1.45}$ [4].

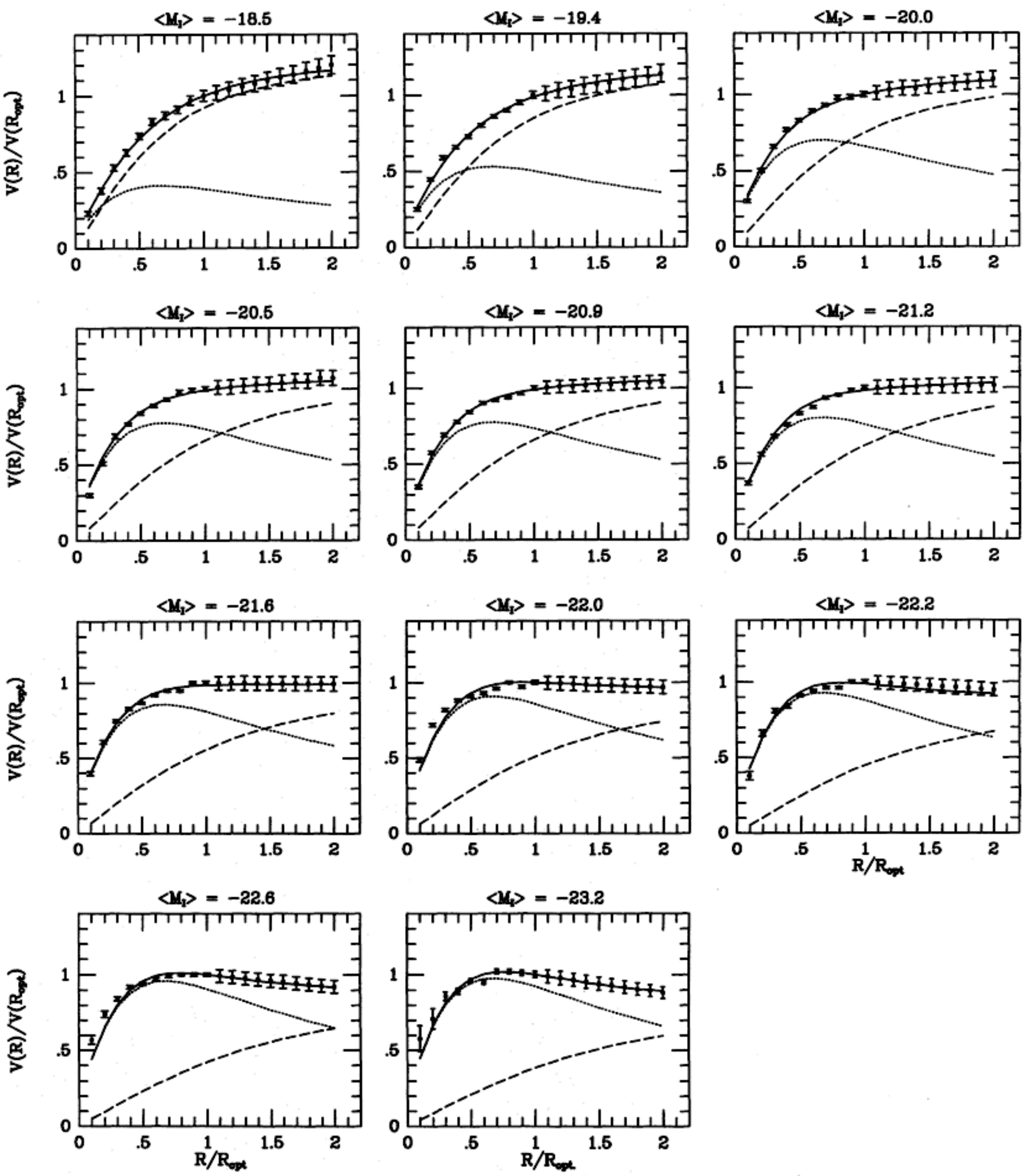


Figure 1.1: Best two-component fits to the universal rotation curve, taken from [2], for a range of absolute magnitudes M_I (each luminosity bin contains 50-100 galaxies). Dotted lines show the disc (luminous material) contribution, and dashed lines show the implied dark halo contribution. The optical radius R_{opt} is defined as the radius encompassing 83% of the total integrated light.

1.1.2 Galactic Clusters

In the study of clusters and superclusters of galaxies, one assumes that the matter is gravitationally relaxed; i.e. that the matter is virialized:

$$2E_k + U = 0 \tag{1.1}$$

where E_k is the mean kinetic energy estimated from the velocity dispersion of the galaxies, and U is the potential energy.

Both methods allow the calculation of the *mass-to-light ratio* of a galaxy cluster, which in turn allows one to calculate the mean density of luminous matter:

$$\rho_g = \mathscr{L}_g \left\langle \frac{M}{L} \right\rangle \tag{1.2}$$

where ρ_g is the density of the luminous matter in the galaxy as a fraction of the critical density $\rho_0 = 3H_0^2/8\pi G$, $\mathscr{L}_g$ is the mean luminosity per unit volume produced by the galaxies, and $\left\langle \frac{M}{L} \right\rangle$ is the mass-to-light ratio.

Typically, studies of galactic clusters find mass-to-light ratios in the hundreds of solar units. For example, Faber and Gallagher (1979) studied seven clusters and reported a median value of $M/L \sim 580\,h\,(M_\odot/L_\odot)$ [5]. In 1998, the ESO Nearby Abell Clusters Survey (ENACS) found $M/L \sim 400\,h\,(M_\odot/L_\odot)$ for 29 clusters [6]; Girardi *et al.* (2000) analyzed a sample of 105 nearby ($z \leq 0.15$) clusters and found a typical mass-to-light ratio of $M/L \sim 250\,h\,(M_\odot/L_\odot)$ [7].

Generally, it is found that M/L for clusters is much higher than that for individual galaxies by roughly a factor of ten. This suggests a great deal of hidden mass residing somewhere in the cluster (in the galaxies themselves, or in the intergalactic regions, or both). It is known from x-ray observations that around 10% of the mass of a galactic cluster is in the form of hot gas. However, the presence of hot gas is not sufficient to explain the apparent dynamical mass. Furthermore, as we shall see below, there are problems reconciling the presence of the hidden matter with nucleosynthesis predictions, were all the cluster mass baryonic.

The mass of a cluster can also be estimated using the same method as for elliptical and lenticular galaxies described in the previous section. That is, by measuring the velocity dispersion of the galaxies and the temperature of the intergalactic gas via its x-ray emission. Clusters are, in fact, some of the most luminous x-ray sources in the sky and show x-ray emission over extended, rather than point-like regions. A study performed by Girardi *et al.* found good agreement between estimates based on the virial approach and the x-ray approach [8], although these results generally agree to within a factor of two [9].

Clusters can also be studied via gravitational lensing (see Figure 1.2) [10]. This offers the possibility of measuring the entire cluster mass (both dark and luminous parts) without having to make the equilibrium and/or symmetry assumptions that are required for the virial and x-ray methods described above.

Although strong lensing can place powerful constraints on the potential well of a cluster, it is relatively rare. Weak lensing, however, will occur for all clusters, but is frequently difficult to quantify, and sometimes impossible. Some recent studies have been able to demonstrate the efficacy of the weak lensing technique on known clusters [11], and the existence of several 'dark clumps' - where mass concentration is seen with no optical or x-ray detection of a cluster - have been reported (for an example of a confirmed cluster, see [12]).

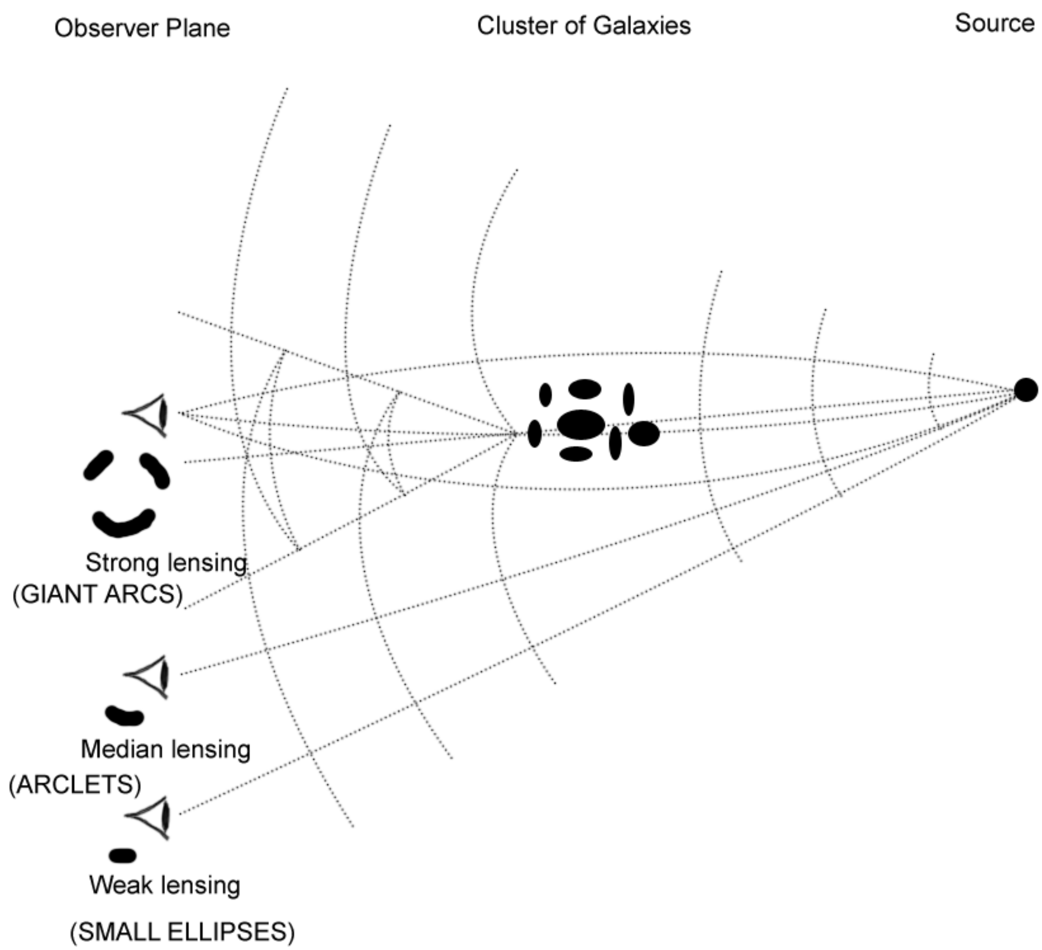


Figure 1.2: Schematic of wavefronts in the presence of a galaxy cluster. Depending upon the lens configuration, the observer may see multiple strongly distorted images (strong lensing arcs), single elliptical distorted images (arclets), or weakly distorted images with almost no individual elongation (weak lensing). Weak lensing is often difficult (and sometimes impossible) to quantify, but will occur at some level for every galaxy cluster. Figure taken from [10].

The combined observations of galactic clusters suggest masses that point to $\Omega_m \sim 0.2 - 0.3$ (some results are summarized in Table 1.1 later in the chapter).

1.1.3 Big Bang Nucleosynthesis (BBN)

The temperature evolution of the very early universe, combined with our knowledge of particle physics, tells us that nucleosynthesis of the light elements (A$\leq$ 7) occurred within the first few hundred seconds after the Big Bang such that by $t \approx 1000$ s, the light element abundances were fixed. Some of these abundances are subject to change during the evolution of the universe (eg. by construction/destruction in stellar populations), but the measurement of the present light elemental abundances can tell us a great deal about the baryonic matter distribution in the moments shortly after the Big Bang.

At $t \gtrsim 0.1$ s after the Big Bang, the neutrinos had decoupled and the relevant matter consisted of protons and neutrons in thermal equilibrium via the weak interactions:

$$\mathrm{n} + \nu_{\mathrm{e}} \rightleftharpoons \mathrm{p} + \mathrm{e}^-, \qquad \mathrm{n} + \mathrm{e}^+ \rightleftharpoons \mathrm{p} + \bar{\nu_e} \tag{1.3}$$

While in thermal equilibrium, the neutron/proton ratio $n/p \simeq 0.2$. This equilibrium was maintained until $T \simeq 1.3 \times 10^9$ K ($t \simeq 20$ s), at which point the neutrons can only transform into protons via β-decay, n $\rightarrow$ p + e$^-$ + $\bar{\nu_e}$, which has a mean lifetime of ~ 900 s. Once the protons and neutrons dropped out of equilibrium the nucleosynthesis of the light elements could begin.

The first link in the nucelosynthetic chain, howevever, p + n $\rightarrow$ D + γ, is ineffective at high temperatures, since the photodestruction rate of the Deuterium ($\propto n_\gamma e^{(2.2\mathrm{MeV})/T}$) is much larger than the production rate ($\propto n_B$) due to the large photon to baryon ratio

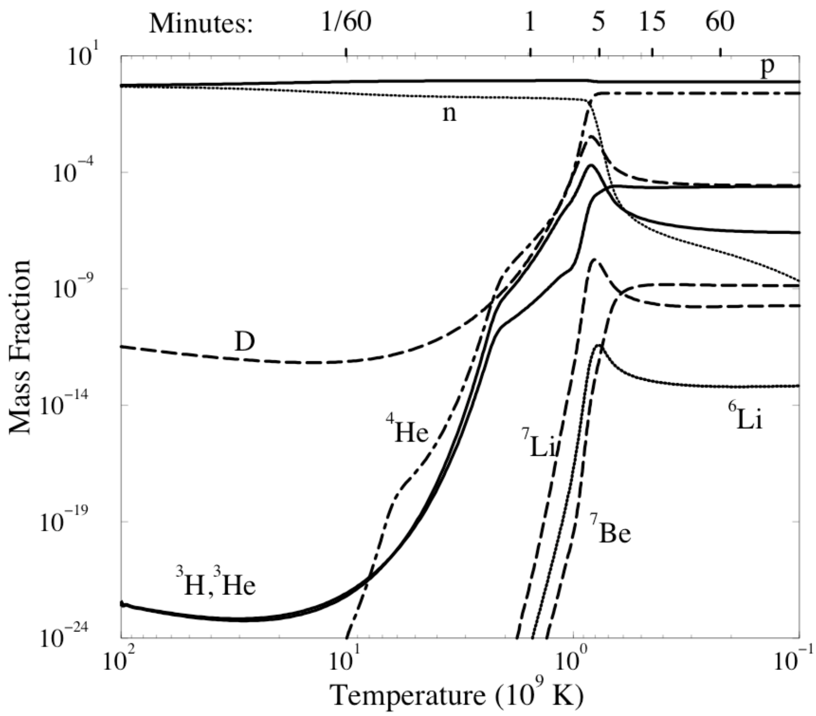


Figure 1.3: Relative mass fractions of the light elements as a function of temperature for a photon to baryon ratio of 2×10^9. Figure taken from [14].

$(n_\gamma/n_B \gtrsim 10^9)$ [13]. Since, to build any element with $A \geq 3$ Deuterium is needed, nucleosynthesis cannot proceed until the production rate becomes stable against photodissociation (this is sometimes referred to as the *Deuterium bottleneck*).

The net production of Deuterium becomes significant at around $T \simeq 2 \times 10^9$ K ($t \simeq 50$s). Once the Deuterium begins to form, the following reactions (which both have large cross-sections) quickly mop up remaining neutrons into ^{4}He:

$$\mathrm{D} + \mathrm{D} \rightarrow {}^3\mathrm{He} + \mathrm{n}, \qquad {}^3\mathrm{He} + \mathrm{D} \rightarrow {}^4\mathrm{He} + \mathrm{p} \tag{1.4}$$

Figure 1.3 shows the evolution of the relative elemental abundances as a function of temperature (and time). It can be seen that approximately 1000 seconds after the Big Bang, the initial relative abundances of the light elements are fixed.

The light element abundances all depend upon the baryonic density, and hence by accurate measurement of the relative abundances, the baryonic matter density of the universe can be obtained. In particular, the Deuterium abundance is very sensitive to the total baryonic energy density, making it an ideal 'baryonometer'.

One can perform a numerical integration of all the rate equations which results in the predictions shown in Figure 1.4. The colored horizontal bands are the predictions based on primordial synthesis. Experimental observations need to take into account that abundances will have changed since primordial synthesis. For example, Hydrogen is processed into Helium in stars, modifying the initial abundance into what we observe today. It can sometimes be difficult to prove that extrapolated abundances from observation are truly primordial, even when heavy element abundances are low.

The ^{4}He abundance is everywhere close to 25% - one of the main predictions of the BBN model described above. In order to obtain the ^{4}He abundance more accurately, it is necessary to take stellar production into account. This is usually done by examining the metallicity of stars in HII regions (usually using the O or N abundance) since stars with a higher metallicity have a slightly higher ^{4}He abundance. The ^{4}He mass fraction is then obtained by extrapolating to zero metallicity; the resulting value is actually within the observed range for some individual HII regions [15].

A similar method is used to estimate the primordial abundance of Deuterium. It is very easily destroyed in stars, but cannot be made there; in fact there are no known astrophysical sites for the production of Deuterium, making any observed D abundances robust lower limits. The observed D abundances in the ISM and the solar system are assumed to be somewhat less than primordial, and indeed higher abundances have been observed in

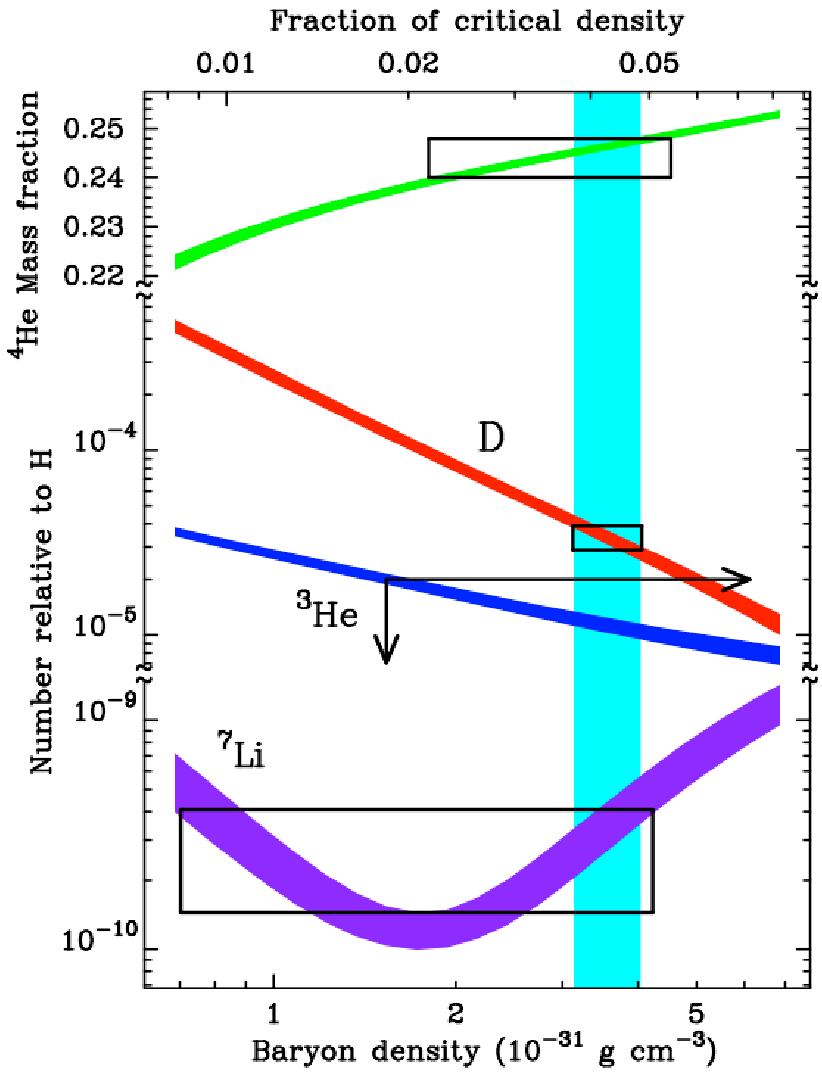


Figure 1.4: Primoridial abundance of light elements as a function of the baryonic energy density. The rectangles denote 95% confidence intervals for the experimental data, and the vertical widths of the predictions denote the 95% confidence interval of the theoretical predictions. The fraction of the critical density is calculated for a Hubble constant of 65 km s^{-1} Mpc^{-1}. Figure taken from [14].

Deuterium absorption in the spectra of high redshift quasars. These values should be close to primordial since they occur at high z and in low-metallicity systems.

Lithium is somewhat more problematic. The ^{7}Li abundance in old hot stars is found to be close to uniform [16], but the presence of ^{6}Li in some of the stars suggests that both ^{6}Li and ^{7}Li were created prior to these stars' formation. It thus becomes harder to extrapolate to zero metallicity and consequently the estimate of the ^{7}Li primordial abundance has an error of $\sim 50\%$ (see Figure 1.4).

The primordial abundance of ^{3}He is very hard to estimate, since stars both make and destroy this isotope. Complex models of stellar evolution are required in order to relate observations to the primordial values. Often the combined abundance of (D + ^{3}He) is used, since most of the ^{3}He that is created in stars is formed from Deuterium. In practice, stars of different masses manage this conversion differently, but since all stars destroy Deuterium to some extent and at least some ^{3}He survives stellar processing, the observed (D + ^{3}He) abundance can be used to place an upper limit on the ^{3}He abundance.

The above combined estimates of BBN currently predict that $\Omega_b h^2 = 0.022 \pm 0.003$ [17], where h is the Hubble constant in units of $100\,\mathrm{km\,s^{-1}Mpc^{-1}}$. Thus there is a strong constraint upon the amount of baryonic matter in the universe, which, coupled with the observations of $\Omega_m h^2 \approx 0.15$ detailed in the previous two sections, argue strongly for the existence of non-baryonic dark matter.

1.1.4 Type-Ia Supernovae

The search for suitable cosmological distance markers is an important pursuit, and a host of 'standard candles' have been utilized over the years. A particular method that has recently proven to be highly effective is the observation of Type-Ia supernovae (SNe Ia), which are highly luminous. SNe Ia events are used as standard candles since their luminosity is known from the SN Ia model, and hence the distance can be inferred by the apparent magnitude. They are also bright enough to be seen at high redshift where there is an appreciable dependence upon cosmological parameters.

Type-Ia supernova are thought to occur for stars which do not exceed the Chandrasekhar limit during their initial gravitational collapse ($M < 1.4M_\odot$), and which have at least one companion star. During collapse, the primary star collapses into a white dwarf which subsequently accretes material from the companion. Eventually, this accretion will accumulate enough mass for the star to exceed the Chandrasekhar limit for the star to undergo gravitational collapse into a neutron star. However, the energy released by turning $1.4M_\odot$ of Carbon to Iron instantly is enough to blow the star apart.

This particular set of events results in a collapse that occurs under relatively similar conditions each time, since by the time of collapse the star's H and He will have been exhausted. The mass is well-defined due to the Chandrasekhar limit being reached and so the total energy released is close to uniform (although there are slight differences depending upon the way that the star burns during the final collapse).

Although SNe Ia events are relatively homogeneous, their use as standard candles became particularly useful when the correlations between peak magnitude and/or light curve shape and spectral features were fully realized [18] [19] [20]. For example, SNe Ia events from stars with more ^{56}Ni are hotter, brighter and have broader light curves [21]. SN Ia events can also look very similar to other SNe events that are not truly Type-Ia (see Figure 1.5). Before SNe Ia events were categorized in this way, the systematic variations in peak

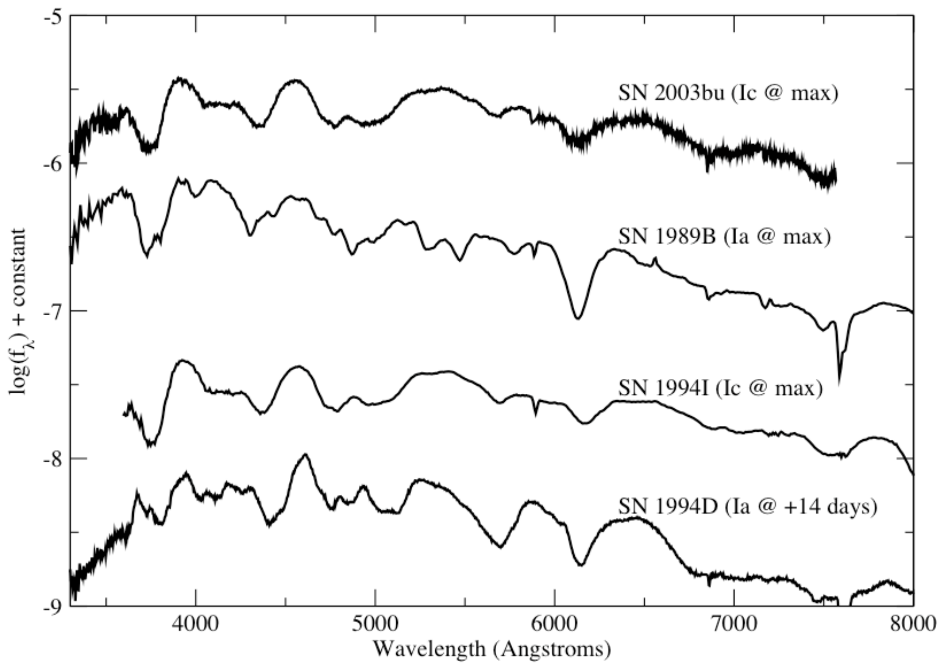


Figure 1.5: Comparison of nearby SNe Ia and Ic events at maximum light. SNe Ic can closely resemble SNe Ia with minor spectral differences (top pair), and Ia spectra after 2 weeks can look like Ic spectra at max (bottom pair). Figure taken from [22].

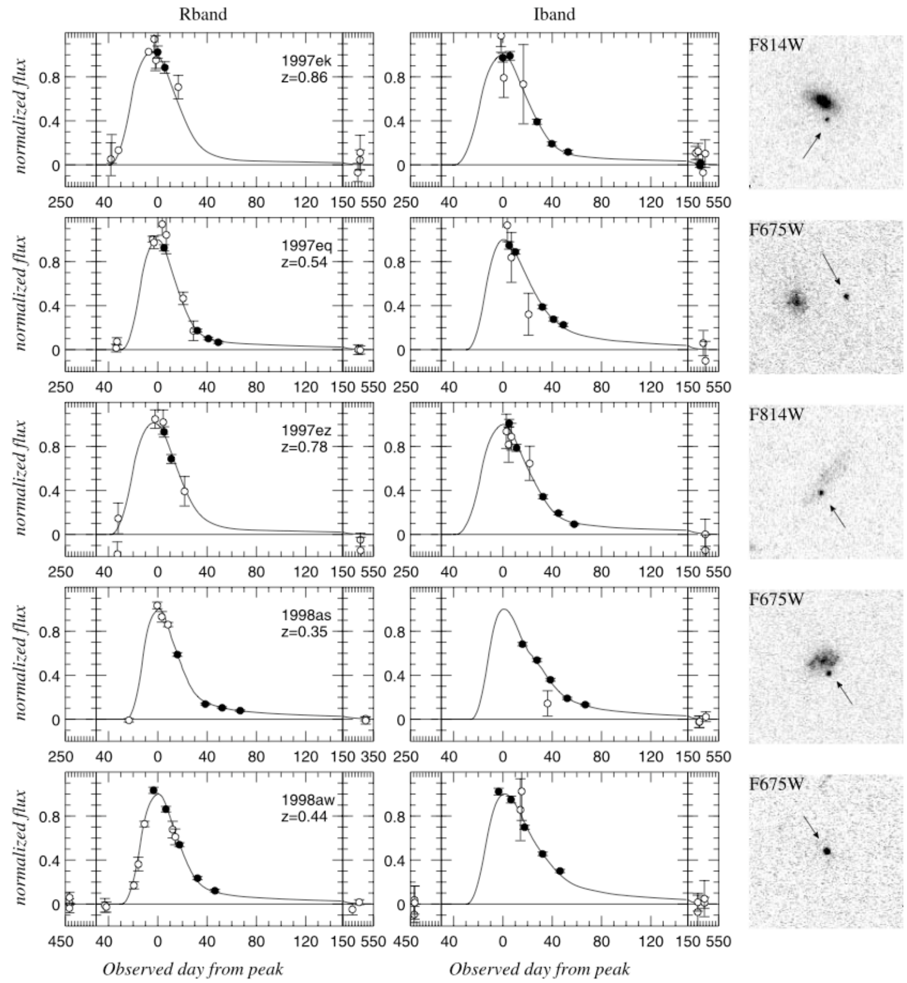


Figure 1.6: Light curves and images for 5 of the galaxies imaged using the HST and reported in [25]. Open circles represent ground-based data points; filled circles represent HST (WFPC2) data points.

brightness were comparable to experimental errors.

Two teams have been compiling data on SNe Ia events; the Supernova Cosmology Project (SCP) and the High-Z Supernova Search Team (HZSNS). Both study SNe Ia events for a typical redshift range $0.15 < z < 1.2$, allowing the direct study of H_0, Ω_m and Ω_Λ. The combined results of these studies show a strong preference for models with a positive cosmological constant ($\Lambda_\Omega > 0$) [23] [24].

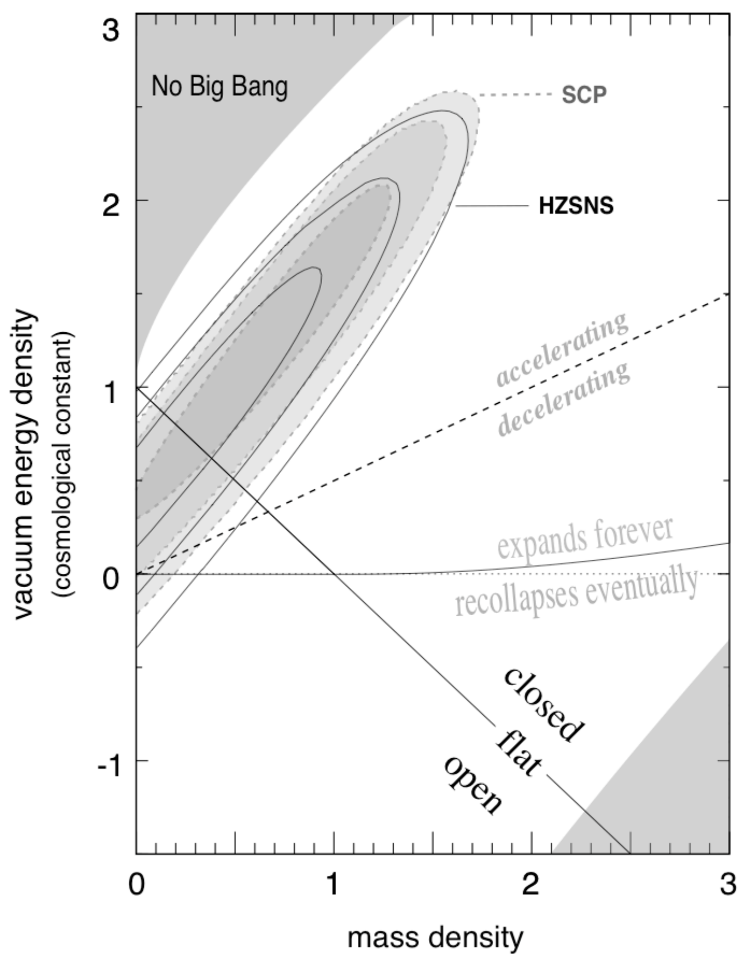


Figure 1.7: Confidence region for HZSNS and SCP data, in the Ω_m, Ω_Λ plane. The SCP result is based on 42 high-z SNe Ia, and the HZSNS result is based upon 16 SNe Ia, of which 2 are present in the SCP 42 SNe sample. A sample of $z < 0.15$ objects are used to constrain the fit and are taken from [19]. Figure taken from [26].

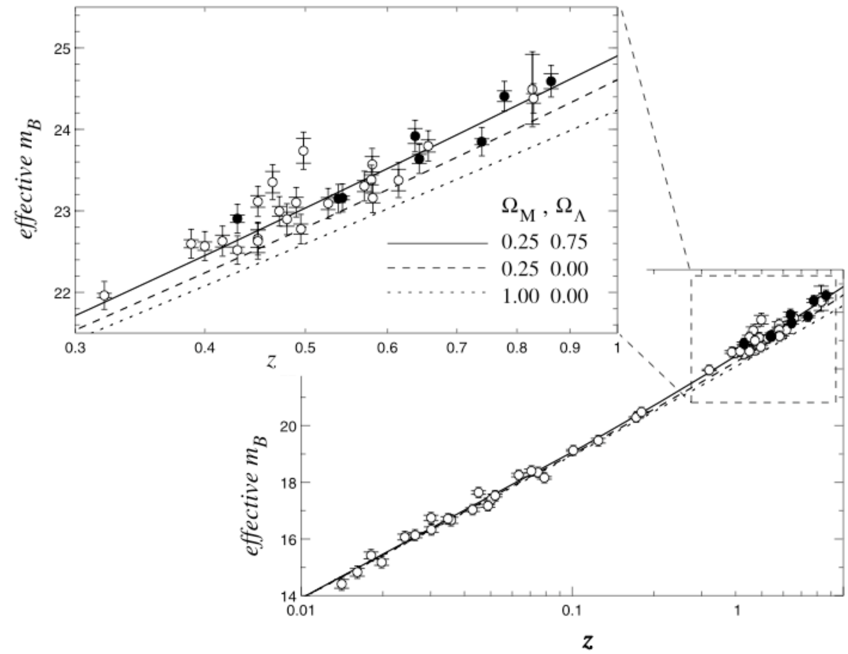


Figure 1.8: Magnitude-redshift diagram for high-z SNe Ia measured by the SCP group. Filled circles represent the 11 SNe reported in [25]. The solid line is the best-fit to the data, assuming a flat universe. This result shows a strong preference for $\Omega_\Lambda > 0$.

A recent result from the SCP experiment [25] reported on 11 SNe at $0.35 < z < 0.86$ measured using the Hubble Space Telescope (HST) (light curves for 5 of the galaxies shown in Figure 1.6). Combined with earlier SCP data, the SNe yield measurements of $\Omega_m = 0.25^{+0.07}_{-0.06}(\text{stat}) \pm 0.04(\text{sys})$ and $\Omega_\Lambda = 0.75^{+0.06}_{-0.07}(\text{stat}) \pm 0.04(\text{sys})$, under the assumption of a flat universe ($\Omega_{total} = 1$).

Figure 1.8 shows the magnitude of SCP-measured SNe Ia as a function of redshift, again showing the strong preference for a cosmological constant. Figure 1.7 shows the confidence regions for HZSNS and SCP data, taken from a 2003 review of their results [26].

1.1.5 Cosmic Microwave Background

The Cosmic Microwave Background (CMB) is a remnant of the photon-baryon decoupling that occurred around 300,000 years after the Big Bang. At this time, the photons are hot enough ($T \gtrsim 3000$ K) that they can ionize hydrogen, and the electrons 'glue' the baryons to the photons via Compton scattering and electromagnetic interactions. The growth of structure has already begun at this epoch, but infalling baryons in high-density regions are resisted by photon pressure. Consequently, the primordial gravitational perturbations create acoustic oscillations within the photon-baryon fluid due to compression in high-density regions.

When the temperature drops below $T \approx 3000$ K, the free electrons are captured, forming atomic hydrogen. The photons at this 'time of last scattering' have a characteristic temperature, but also have an increased (decreased) temperature when originating from a region of compression (rarefaction) at the time of last scattering. The photons experience gravitational redshifts from climbing out of the potential wells on the last scattering surface and from any time variation in potentials along their trajectory (Sachs-Wolfe effect).

Hence, at the time of last scattering the acoustic oscillations are frozen, resulting in photons originating from extrema of the oscillations. Usually, the temperature distribution on the sky is expanded as a sum over spherical harmonics:

$$\frac{\Delta T(\hat{\mathbf{n}})}{T} = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=+l} a_{lm} Y_{lm}(\theta, \phi) \tag{1.5}$$

where $\Delta T(\hat{\mathbf{n}})/T$ is the temperature fluctuation when pointing at an arbitrary direction in the sky $\hat{\mathbf{n}}$. The CMB power spectrum is therefore defined as the autocovariance function of

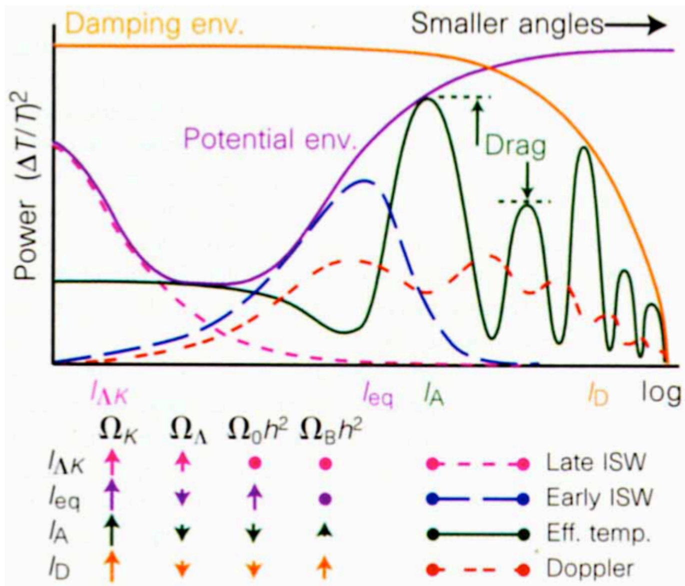


Figure 1.9: Decomposition of the CMB spectrum and each component's dependence upon cosmological parameters. Ω_K is the overall curvature of the universe ($\Omega_K = 1 - \Omega_0 - \Omega_\Lambda$) and Ω_0 is the energy-density of the matter ($= \Omega_m$). The source of each component is discussed in the text. Figure taken from [27].

the temperature fluctuations:

$$C(\Theta) = \left\langle \frac{\Delta T(\hat{\mathbf{n}}_1)}{T} \frac{\Delta T(\hat{\mathbf{n}}_2)}{T} \right\rangle \tag{1.6}$$

where $\cos\Theta = \hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2$. The angular power spectrum can be written in terms of Legendre polynomials P_l and their coefficients C_l:

$$C(\Theta) = \frac{1}{4\pi} \sum_{l=2}^{\infty} (2l+1)\, C_l P_l(\cos\Theta) \tag{1.7}$$

The power spectrum can be expressed as a function of either l or Θ, since $l \propto \Theta^{-1}$. The sum ignores the first two terms, since the $l = 0$ term is a monopole correction that is unmeasurable, and the $l = 1$ term is a dipole term caused by our motion through space relative to the frame in which the CMB is isotropic. These are generally treated separately from the $l > 1$ terms which represent true fluctuations in the CMB.

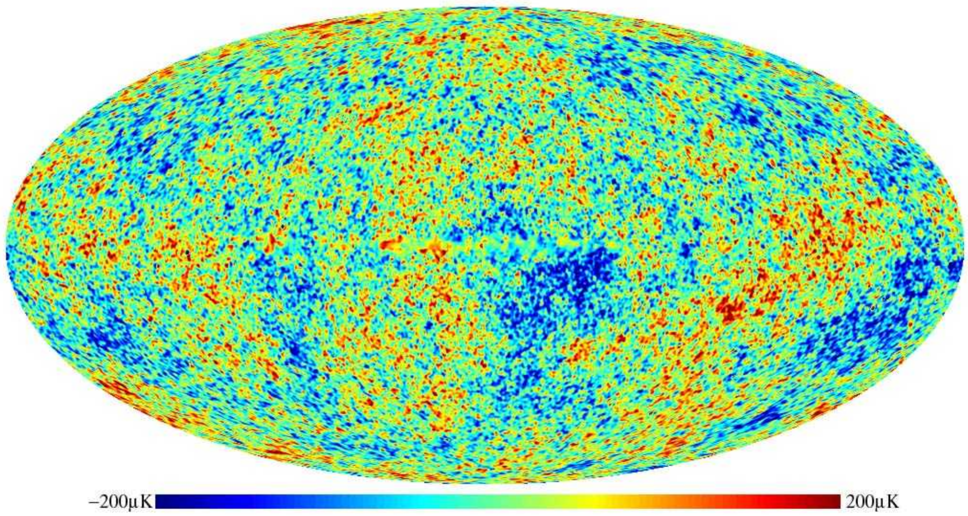


Figure 1.10: Sky map of the CMB as measured by the WMAP satellite [29]. This is an 'internal linear combination' map which combines data from each of the five frequency bands in order to maintain unity response to the CMB while minimizing foreground contamination.

There are several components to the power spectrum, summarized in Figure 1.9 which also shows each component's dependency upon cosmological parameters. As already discussed, acoustic peaks are created at last scattering due to the oscillations being caught at extrema. These peaks form a harmonic series in wavenumber based upon the sound horizon at last scattering; $k_m = m\pi/s_*$ for the mth peak, where s_* is the sound horizon (these acoustic peaks are shown in Figure 1.9 as the 'effective temperature'). Since the physical size of the acoustic horizon is set by simple physics, the acoustic peaks have an angular size on the sky which depends upon the curvature of the universe. Therefore, the acoustic spectrum is sensitive to both Ω_m and Ω_Λ.

The baryonic contribution to the effective mass of the baryon-photon fluid affects the frequency of its oscillations, and changes the amount of gravitational compression experienced by the fluid in a potential well (baryonic drag). Hence, the compression peaks are enhanced with increasing baryonic matter content, and consequently the relative peak

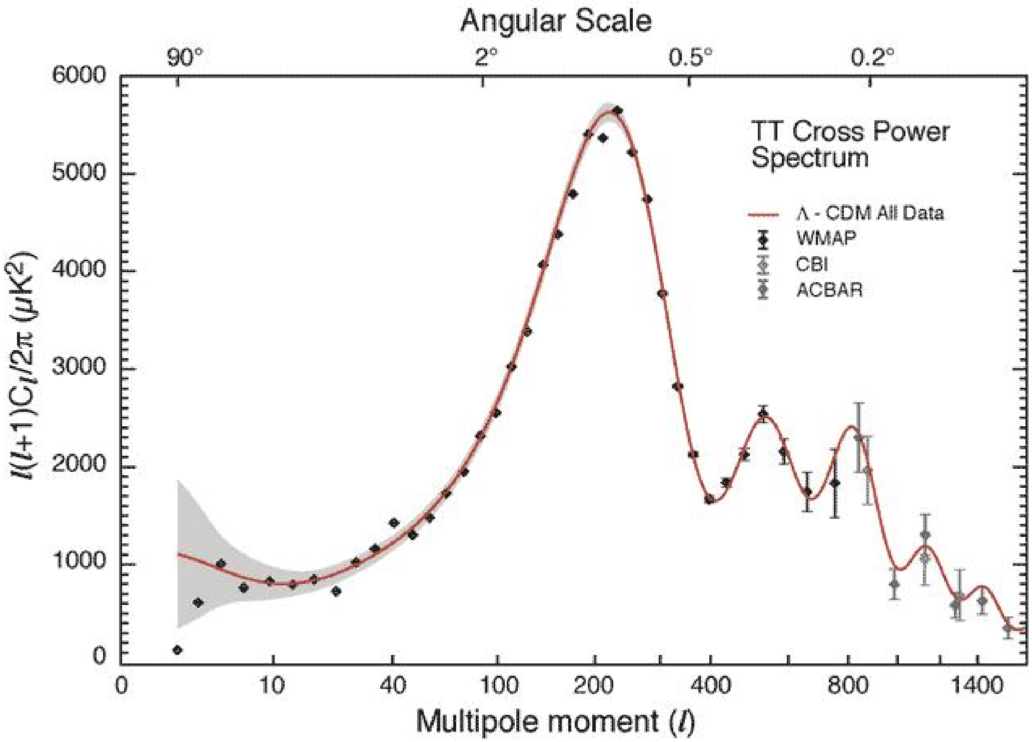


Figure 1.11: WMAP angular power spectrum from first year data. The curve shows the best fit model based on a combined data analysis as reported in [29]. The WMAP data are consistent with the ACBAR and CBI measurements, as shown. The grey area at the left is uncertainty due to cosmic variance.

heights of the acoustic spectrum can be used to determine the baryonic matter density Ω_b.

The motion of the fluid also Doppler-shifts the frequency of the photons, creating oscillations in the power spectrum that are out of phase with the temperature oscillations. Hence, the Doppler effect makes the acoustic peaks less prominent, especially when the baryonic mass content is small. As the baryonic mass increases, the relative contribution of the Doppler effect decreases and therefore we have another check on Ω_b.

Thermal fluctuations are damped at small angular scales ('damping' in Figure 1.9). The photon-baryon fluid is imperfect since the photons can random-walk through the fluid, allowing diffusion between hot and cold regions at small scales. The damping scale depends only upon the thermal history and the baryonic matter density.

The trajectory of a photon leaving the surface of last-scattering experiences gravitational redshifts and blueshifts as it traverses gravitational potentials. As the structure of the universe evolves, these potentials also evolve and therefore the redshift that is experienced by a photon leaving a potential well is not necessarily the same as the blueshift it experienced when entering. This leads to a temperature fluctuation that is sensitive to the evolution of structure, and is referred to as the integrated Sachs-Wolfe effect (ISW). This effect is usually broken down into 'early' and 'late' components (see Figure 1.9). These represent potential decays brought about by the background changing from radiation to matter dominated (early ISW), and from the background changing from matter to Λ or curvature dominated (late ISW).

The overall power spectrum is therefore very sensitive to Ω_m, Ω_b, Ω_Λ and Ω_{total}, and an accurate measurement of the spectrum provides strong limits on these parameters. The CMB sky was first mapped by the COBE satellite in 1992 [28] with a resolution of $\sim 7^\circ$ FWHM. More recently, the WMAP satellite mapped the CMB to $\sim 0.5^\circ$ resolution in 2003 (see Figure 1.10 and 1.11).

Using the WMAP data only, the cosmological parameters under the ΛCDM model are (with 1σ errors): $(\Omega_m h^2,\, \Omega_b h^2,\, h) = (0.14 \pm 0.02,\, 0.024 \pm 0.001,\, 0.72 \pm 0.05)$ and the age of the universe $t_0 = 13.4 \pm 0.3$ Gyr [30]. This result implies that $\Omega_b \neq \Omega_m$ at 5.8σ.

WMAP three year data released just prior to publication of this dissertation gives best fit values for the ΛCDM model as $(\Omega_m h^2,\, \Omega_b h^2,\, h) = (0.127^{+0.007}_{-0.013},\, 0.0223^{+0.0007}_{-0.0009},\, 0.73^{+0.03}_{-0.03})$ [31]. Assuming a flat universe, WMAP three year data combined with the Supernova Legacy Survey (SNLS) data yields $w = -0.97^{+0.07}_{-0.09}$ for the dark energy equation of state.

Limits based on the combination of WMAP one year data with other observations are detailed in the next section.

1.2 Summary of Current Limits on the ΛCDM Model

The WMAP results place strong constraints on the parameters of the ΛCDM model, and since non-CMB measurements are now less reliable and precise than the CMB, they become the limiting factor in setting combined limits on the model. In a combined analysis by Tegmark *et al.* (2003) [32], WMAP data was combined with (200,000 galaxy) data from the Sloan Digital Sky Survey (SDSS), the Two Degree Field Galaxy Redshift Survey (2dFGRS) and with observations of galaxy clusters and of weak lensing.

Some results of this analysis are shown in Figures 1.12 to 1.14. The analysis concludes that the data is consistent with the flat adiabatic ΛCDM model, but still provides variation in terms of tensor fluctuations and/or massive neutrinos. It will be left to future observations to provide stronger constraints on these values. However, the best fit 'vanilla-lite' model, which includes only 5 parameters (others are fixed, such as $w = -1$, $n_s = 1$) has the following values:

$$(\tau, \Omega_\Lambda, \omega_d, \omega_b, A_s) = (0.17, 0.72, 0.12, 0.024, 0.89) \tag{1.8}$$

where τ is the reionization optical depth, $\omega_d \equiv \Omega_d h^2$, $\omega_b \equiv \Omega_b h^2$ and A_s is the scalar fluctuation amplitude. The analysis also examines the fit when one allows these parameters plus say, w or n_s, to vary.

It is particularly interesting that one finds substantial agreement between data sets (WMAP, SDSS, 2dFGRS, etc.) even when one of the data sets is removed from the analysis.

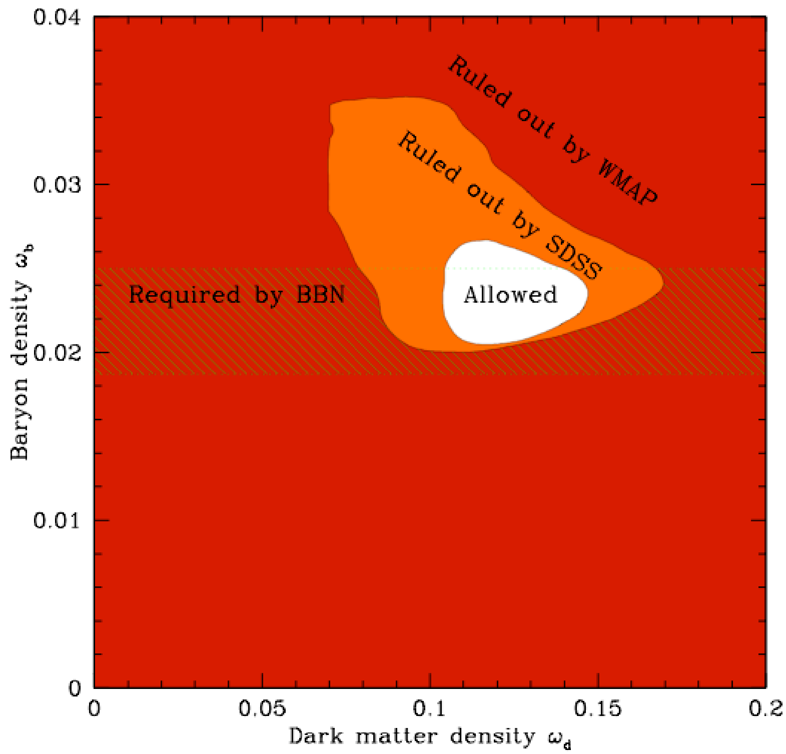


Figure 1.12: 95% constraints in the $(\omega_d, \omega_m) \equiv (\Omega_d h^2, \Omega_m h^2)$ plane, taken from the analysis in [32]. The acceptance regions are defined by the best fit to a 6-parameter 'vanilla' model $(\tau, \Omega_\Lambda, \omega_d, \omega_b, A_s, n_s)$.

i.e. no data set is indispensable, and the addition of any one of these data sets to the others does not significantly improve the χ^2 fit that results.

The existence of dark energy is now strongly supported by SNe Ia measurements, power spectrum analyses such as [32] and study of late ISW [33]. The present dark energy density, as measured by these observations, is $\Omega_\Lambda = 0.70 \pm 0.04$. This result is consistent with a dark energy density that remains constant in time ($w = -1$), but the uncertainty in the value of w is at the 20% level.

Measurement of the Hubble parameter h results in remarkable agreement between WMAP, SDSS and the HST key project measurement; $h = 0.70^{+0.04}_{-0.03}$ (WMAP + SDSS),

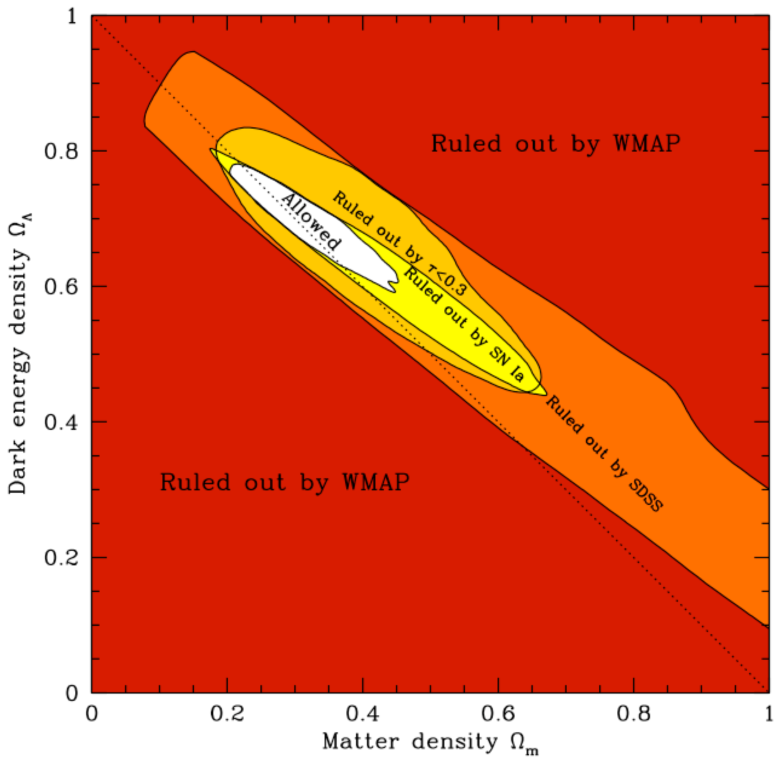


Figure 1.13: 95% constraints in the $(\Omega_m, \Omega_\Lambda)$ plane, taken from the analysis in [32]. The regions result from a 7-parameter fit $(\tau, \Omega_\Lambda, \omega_d, \omega_b, A_s, n_s, w)$.

$h = 0.72 \pm 0.07$ (HST key project [34]). The SDSS measurement of Ω_m also agrees with the pure WMAP value, which respectively are $\Omega_m = 0.30 \pm 0.04$ [32] and $\Omega_m = 0.26 \pm 0.05$ [30].

The agreement between the ΛCDM model and the wealth of experimental data is astounding, yet there remain important outstanding questions. Firstly, we do not know why the dark matter, dark energy and near-scale invariant seed fluctuations should exist, nor do we know the nature of the physics responsible for their existence. The standard model does not explain why baryogenesis occurred in the amount that it did (since the observed CP violation seems to be too small), nor does it provide candidates for dark matter particles, and nor does it explain why the neutrino masses are nonzero and small.

Limits on other parameters in the ΛCDM model also need to be further refined. For

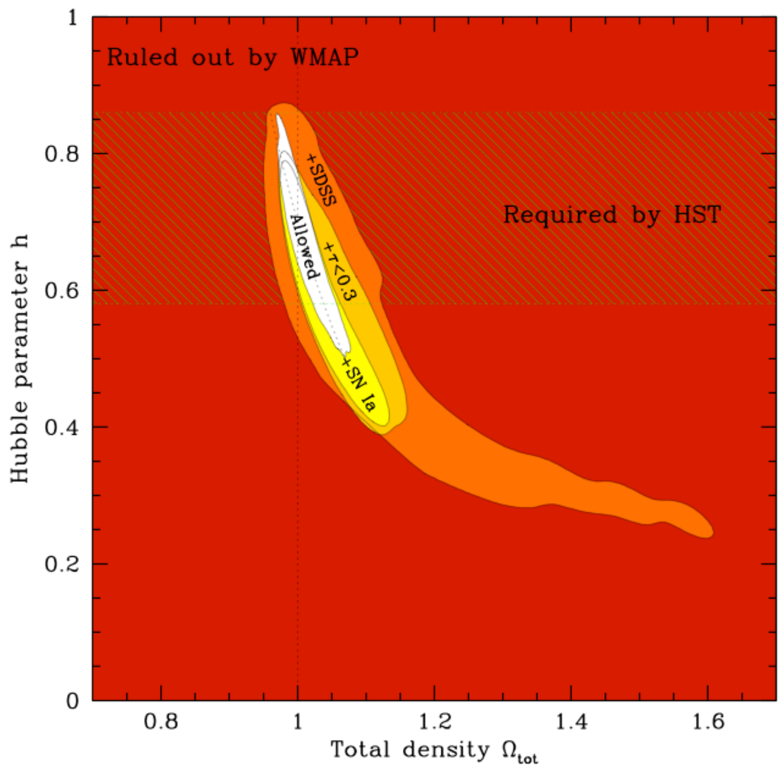


Figure 1.14: 95% constraints in the (Ω_{total}, h) plane, taken from the analysis in [32]. The regions result from a 7-parameter fit $(\tau, \Omega_\Lambda, \omega_d, \omega_b, A_s, n_s, w)$. The dark-yellow region results from adding a prior that $\tau < 0.3$ to the fit, and the hatched band is required by the Hubble key project data [35].

example, we still do not know whether $w = -1$ (which would certainly be preferred from a theoretical perspective).

1.3 Non-Baryonic Dark Matter

That the majority of the matter in the universe is non-baryonic is supported by a wealth of experimental data, summarized in Table 1.1 which shows a selection of recent measurements of Ω_m and Ω_b. Given that this non-baryonic dark matter exists, we naturally wonder what form it takes. What are the underlying physics that give rise to this matter? Can current

Observation	Measurement
Clusters (18 clusters, SZE) [36]	$\Omega_m \sim 0.25$
Clusters (17 clusters, baryon mass function) [38]	$\Omega_m = 0.24 \pm 0.12$ (for $\Omega_{total} = 1$)
Clusters (based on σ_8 at $z = 0.5 - 0.8$) [39]	$\Omega_m = 0.17 \pm 0.05$
Weak Lensing (75 deg^2 survey) [40]	$\Omega_m \sim 0.3$
Type-1a SNe (42 SNe, $z = 0.18 - 0.83$) [23]	$\Omega_m = 0.28^{+0.09}_{-0.08}$ (for $\Omega_{total} = 1$)
Type-1a SNe (8 SNe, $z = 0.3 - 1.2$) [37]	$\Omega_m = 0.28 \pm 0.05$ (for $\Omega_{total} = 1$)
2dFGRS (power spectrum fit) [41]	$\Omega_m = 0.26 \pm 0.03$ *
2dFGRS (power spectrum fit inc. WMAP data) [41]	$\Omega_m = 0.31 \pm 0.05$
CMB (WMAP only) [30]	$\Omega_m = 0.27 \pm 0.04$ *
SDSS + WMAP + 2dFGRS joint analysis [32]	$\Omega_m = 0.30 \pm 0.04$
BBN (QSO D/H abundance) [42]	$\Omega_b = 0.042 \pm 0.004$ *
BBN (review) [43]	$\Omega_b = 0.039 \pm 0.005$ *
2dFGRS (power spectrum) [41]	$\Omega_b = 0.044 \pm 0.006$ *
CMB (WMAP only) [30]	$\Omega_b = 0.046 \pm 0.002$ *
SDSS + WMAP + 2dFGRS joint analysis [32]	$\Omega_b \sim 0.049$ *

Table 1.1: Experimental measurements of Ω_m and Ω_b. * means $h = 0.7$ has been assumed.

physics provide a suitable candidate particle, or are new physics needed?

Neutrinos

One of the first proposed solutions to the dark matter problem were neutrinos. Relic neutrinos left over from the Big Bang would have been moving with relativistic velocities when they dropped out of thermal equilibrium. Obviously, neutrinos would have to be massive to constitute any or all of the dark matter, but it has now been established to a high degree of certainty that neutrinos change from one flavor to another [44] and therefore must be massive.

In order to be a significant part of the dark matter, the relic neutrinos would have to exist in massive quantities (their number density would be similar to that of the photons). Since the number of CMB photons is known, it can be deduced that [45]:

$$\Omega_\nu \approx \frac{m_{tot}}{30\,\mathrm{eV}} \tag{1.9}$$

where m_{tot} is the total mass due to all flavors of neutrinos. Clearly, if neutrinos have mass $m_\nu \leq 1\,\mathrm{eV}$ they are unable to solve the dark matter problem.

Current cosmological limits (based on observed structure formation) suggest that the neutrino fraction of dark matter $f_\nu < 0.12$, although a zero value is preferred by some analyses [32]. Atmospheric neutrino data suggests that $f_\nu \gtrsim 0.004$, since the neutrino oscillations occur with a lower mass limit of $\sim 0.05\,\mathrm{eV}$ [46].

A more massive neutrino would provide sufficient mass, but due to their relativistic nature, a universe populated by a large number of relic neutrinos would not generate the observed small-scale structure in the universe. Neutrinos are therefore often referred to as being a type of 'Hot' Dark Matter (HDM) - where 'hot' refers to their relativistic nature. A universe dominated by neutrinos would form large structures first, with smaller structures then resulting from the fragmentation of the larger objects. Thus it seems that there is a need for non-relativistic, or 'Cold' Dark Matter (CDM).

Axions

Axions are neutral scalar (spin 0) particles associated with the breaking of the $U(1)$ Peccei-Quinn symmetry, which was introduced to solve the strong CP problem [47]. That is, the QCD Lagrangian includes a term $\frac{\theta}{16\pi^2}\mathrm{Tr}\left(F_{\mu\nu}\tilde{F}^{\mu\nu}\right)$ which violates CP symmetry. There is no *a priori* reason that we expect the coupling θ to be zero, and experimental observations of the neutron dipole electric moment imply that $\theta < 10^{-9}$. This CP violation has not been observed, and it is not yet understood why θ is so small (given that we might expect a strong interaction parameter to be $\mathcal{O}(1)$).

The global Peccei-Quinn symmetry was introduced to dynamically solve this problem, by introducing a new $U(1)$ symmetry which is broken at some scale f_a, resulting in a boson (the axion) via the Goldstone theorem. One alternative solution is that θ may be zero at the Planck scale, but small and nonzero at the weak scale.

Axions are good dark matter candidates, since if they exist they would have been produced in large numbers in the Big Bang, would be nonrelativistic, and would never have been in thermal equilibrium with the rest of the matter.

The mass of the axion depends upon the scale of the $U(1)$ symmetry breaking:

$$m_a \sim 10^{-5}\,\mathrm{eV} \times \frac{10^{12}\,\mathrm{GeV}}{f_a} \tag{1.10}$$

It turns out that excessive emission of axions would over-cool stars, and observation of the supernova SN1987 suggests an upper bound to the symmetry-breaking scale of $f_a \gtrsim 10^9\,\mathrm{GeV}$. Also, the requirement that $\Omega_{total} \leq 1$ results in a lower bound of $f_a \lesssim 10^{12}\,\mathrm{GeV}$. Thus the axion would have a mass in the range $10^{-5}\,\mathrm{eV}$ to $10^{-2}\,\mathrm{eV}$.

Axions can couple to photons via fermion vacuum loops, so one way to detect their presence is to use microwave cavities in magnetic fields. The axions convert to microwave photons within the cavity, which is tuned to probe potential frequencies of the emitted photons (and therefore to probe potential axion masses). Current experiments have probed only a small amount of the potential mass range, but new experiments are forthcoming. For a review, see [48].

WIMPs

The Weakly Interacting Massive Particle (WIMP) is a particularly interesting candidate for dark matter in the universe. One problem with hypothesizing a new massive particle

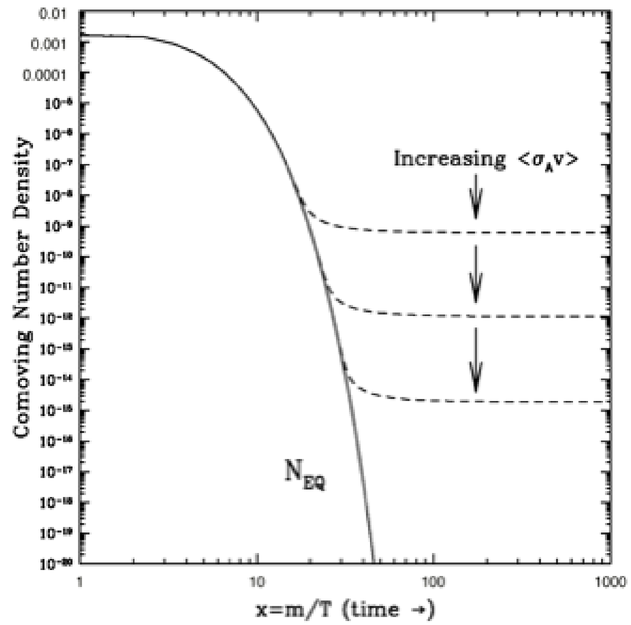


Figure 1.15: Schematic of the freeze out process that occurs when WIMPs drop out of thermal equilibrium. The solid line represents the number density of the WIMPs were they to not freeze out of thermal equilibrium; the dashed line shows the outcome of freeze out, which depends upon the WIMP annihilation cross-section and the WIMP mass. Figure taken from [49].

species is that if the particles are still in thermal equilibrium at present, their abundance, $n \sim (m/T)^{3/2} \exp(-m/T)$ would be negligible due to the exponential factor.

One is forced either to suppose the existence of a particle that was never in thermal equilibrium (as in the case of the axion), or to suppose a species whose interactions 'freeze out' at a temperature such that m/T is not much greater than 1. It is this latter case that leads to the WIMP, and it allows the species to have a significant relic abundance today.

In the early universe, the WIMP and its antiparticle (which we denote by χ and $\bar{\chi}$ respectively), are created and annihilated by one or more species X. i.e. $\chi\bar{\chi} \leftrightarrow X\bar{X}$. When the temperature of the universe is larger than the mass of the WIMP, the number of WIMPs and photons is roughly the same, due to constant creation and annihilation of the WIMP.

However, once $T \leq m_\chi$, the WIMP number density falls until the WIMP annihilation rate drops below the expansion rate of the universe. At this point (typically $T_F \approx m_\chi/20$) the WIMPs freeze out; they cannot annihilate and their density asymptotes to a constant value.

The resulting number density of WIMPs depends upon their annihilation cross-section σ_A and velocity v, as shown in Figure 1.15. This leads to the freeze out condition $\langle \sigma_A v \rangle n = H$ (where H is the Hubble constant). The current WIMP energy density turns out to be [45]:

$$\Omega_\chi h^2 \simeq \frac{3 \times 10^{-27}\,\mathrm{cm}^3\,\mathrm{s}^{-1}}{\langle \sigma_A v \rangle} \tag{1.11}$$

where the number in the numerator comes from considering $\Omega_\chi = (\rho_\chi/\rho_c)$ and the above freeze out condition.

Thus, if a new particle with weak-scale interactions were to exist in nature its cross section would be $\sigma \simeq \alpha^2/m_{weak}^2$ which is $\sigma \approx 10^{-9}\,\mathrm{GeV}^{-2} \approx 1\,\mathrm{pb}$. At the freeze out temperature, the velocity would be a significant fraction of the speed of light ($v \sim c/5$). Putting all this into (1.11) we obtain $\Omega_\chi h^2 \approx 0.3$ - close to the value that we expect for the dark matter. Therefore, by postulating new physics at the weak scale we are provided a reliable solution to the dark matter problem.

This prescription puts limits upon the characteristics of the WIMP particle, but there are many areas of new physics that could supply a possible WIMP candidate. Perhaps the most natural comes from Supersymmetry (SUSY), which introduces new physics at precisely the weak scale, and for which there is much motivation from particle physics. Essentially, the Standard Model is extended to provide a supersymmetric partner for each particle in the Standard Model. The electron is partnered by the bosonic 'selectron', the gluon is partnered by the fermionic 'gluino', etc. The supersymmetric particles all have much higher masses than those in the standard model, due to this new symmetry being broken at some energy scale.

The parameter space for SUSY is huge (there are hundreds of free parameters in the complete model), but under models that conserve R-parity (a new quantum number invoked to ensure non-observed events such as fast proton decay do not occur) one typically finds the lightest (and therefore most stable) supersymmetric particle, or LSP, is a good WIMP candidate.

The requirement of R-parity conservation is usually extended to the minimal supersymmetric model (MSSM), in which the LSP is the lightest 'neutralino'. The neutralino is a mix of four supersymmetric states: the bino (supersymmetric B), the wino (supersymmetric W) and the two neutral Higgsinos. For recent reviews of SUSY dark matter, see [50] [51] [52].

1.4 WIMP Direct Detection

CDMS is one of many WIMP direct detection experiments that have been, or are, in operation (for a review, see [52]). All of these experiments search for WIMP-nucleon scattering events, which occur with recoil energies of $< 100\,\mathrm{keV}$. In general, the detection rate of dark matter recoils is of the form:

$$\frac{\mathrm{d}R}{\mathrm{d}E_R} = \frac{R_0}{E_0 r} e^{-E_R/E_0 r} \tag{1.12}$$

where R_0 is the total event rate, E_0 is the most probable incident kinetic energy of a WIMP of mass M_χ, E_R is the recoil energy of the target particle, r is a dimensionless form of the reduced mass; $r = 4M_\chi M_T/(M_\chi + M_T)^2$, and M_T is the mass of the target particle [53]. The derivation of this spectrum assumes a stationary detector sitting in a Maxwell-Boltzmann distribution of WIMPs.

This basic spectrum, however, must be modified to take into account the following effects:

1) The Earth-Sun system is moving relative to the dark matter halo, and the Earth moves relative to the Sun.

2) There is an upper limit on the kinetic energy of a WIMP due to the escape velocity of the galaxy.

3) A suppression of the nuclear cross section can occur due to the finite size of the target nucleus (form factor suppression).

4) The WIMP-nucleon coupling may be dependent upon the spin of the nucleus, which needs to be taken into account for nuclei with nonzero spin.

Taking into account the typical WIMP escape velocity of $v_{esc} \sim 600\,\mathrm{km\,s^{-1}}$ and the Earth's velocity through the galaxy v_E yields

$$\frac{\mathrm{d}R(v_E, v_{esc})}{\mathrm{d}E_R} = \frac{1}{k}\left[\frac{\mathrm{d}R(v_E, \infty)}{\mathrm{d}E_R} - \frac{R_0}{E_0 r} e^{-v_{esc}^2/v_0^2}\right] \tag{1.13}$$

with

$$k = \left[\mathrm{erf}\left(\frac{v_{esc}}{v_0}\right) - \frac{2}{\sqrt{\pi}}\frac{v_{esc}}{v_0} e^{-v_{esc}^2/v_0^2}\right] \tag{1.14}$$

where $v_0 \approx 230\,\mathrm{km\,s^{-1}}$ is the mean WIMP velocity. Since the effective v_0 changes over the year due to the Earth's orbital motion ($v_0 \approx 215\,\mathrm{km\,s^{-1}}$ in December, $v_0 \approx 245\,\mathrm{km\,s^{-1}}$ in June), one way to search for a WIMP signal would be to look for annual modulation of the event rate. Since the modulation is $\sim 5\%$ this requires high statistics and a way to distinguish between the variation of the WIMP signal and natural variation in the backgrounds.

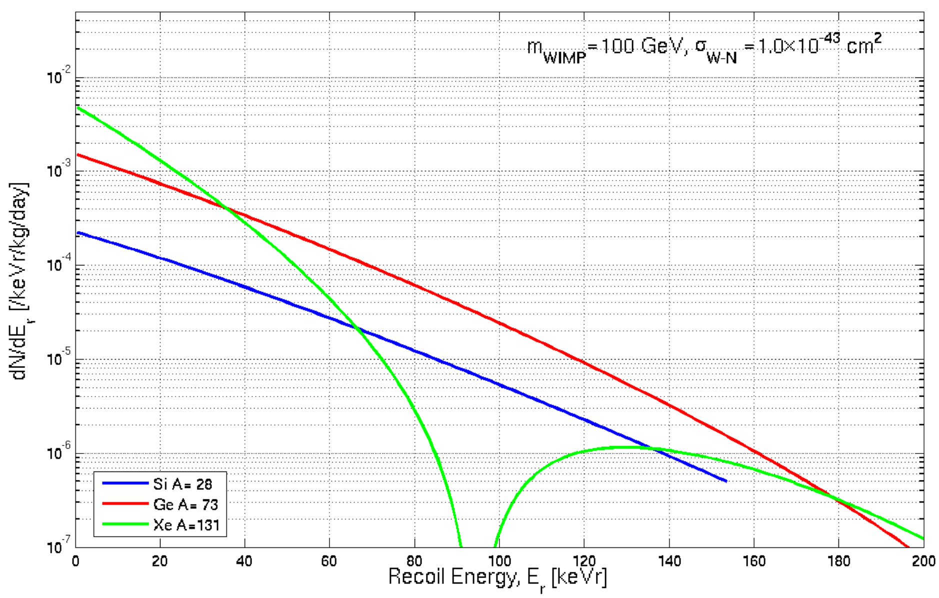


Figure 1.16: Differential recoil spectra for a 100 GeV WIMP incident on Si, Ge and Xe targets, assuming a WIMP-nucleon cross section of $\sigma_{W-n} = 10^{-43}\,\mathrm{cm}^2$. Calculated using formalism described in [53].

The detection rate equation (1.13) is well approximated by a form similar to that of (1.12):

$$\frac{\mathrm{d}R(v_E,\infty)}{\mathrm{d}E_R} = c_1 \frac{R_0}{E_0 r} e^{-c_2 E_R / E_0 r} \tag{1.15}$$

where c_1 and c_2 both vary by 1-2% over the year and have mean values, $c_1 = 0.751$ and $c_2 = 0.561$.

The rate spectrum (1.15) still assumes zero-momentum transfer, which necessitates the inclusion of a nuclear form factor that depends upon the type of interaction (eg. spin-dependent or spin-independent). It is also more instructive to use the WIMP-nucleon cross section in order to compare event rates for different target nuclei.

Figure 1.16 shows example recoil spectra and integrated event rates in Si, Ge and Xe targets. The dramatic decrease in event rate that occurs around 90 keV in Xe is a direct result of Helm form factor supression.

1.5 WIMP Direct Detection Experiments

The following sections review the different technological approaches to WIMP direct detection experiments.

1.5.1 Ge-Based Experiments

The first detectors used in dark matter searches were Ge-based detectors which grew from the technology developed to look for neutrinoless $\beta\beta$-decay. A strong electric field across a high purity Ge detector was used to separate and collect the free charges that result from an interaction in the crystal. These detectors were unable to discriminate between nuclear and electron recoils and were severely limited by backgrounds. Nonetheless, the preliminary limits on the existence of dark matter were set using these detectors, ruling out the Dirac neutrino and cosmions as potential dark matter candidates. The most sensitive limits in this regime were set by the Heidelburg-Moscow collaboration [54] and the IGEX collaboration [55].

Cooling Ge crystals down to temperatures of 20-50mK allows the measurement of the phonon signal that is also created in the crystal during an interaction event. The subsequent generation of Ge detectors again used an electric field to extract and measure the free charge created during a particle interaction, but also measured the lattice vibrations that occurred. The measurement of both allows discrimination between electron and nuclear recoil types, since the ionization is quenched for nuclear recoils in Ge. The recoil energy for both recoil types can also be extracted on an event-by-event basis, and the energy resolution is typically on the order of a keV.

CDMS-I [56] and Edelweiss [57] comprised the first generation of such experiments. The Edelwiss experiment has been operating 3×320 g Ge cryogenic detectors at the Modane Laboratory, Fréjus, France since 2002. The experiment uses neutron transmutation doped (NTD) thermistors to read the thermal phonon signal. New search results with an upgraded 7kg of detectors are expected in the near future.

CDMS-I was operated from 1996-2002 at the Stanford Underground Facility (SUF) at a depth of 10m (muon flux ~ 5 times lower than surface), and also used Ge NTDs to measure the thermal phonon response. CDMS-II uses transition-edge-sensors (TESs) to measure the athermal phonon signal; these detectors will be described in detail in Chapter 2.

At present, Ge-based experiments set the most sensitive limits on the spin-independent WIMP-nucleon scattering cross section.

1.5.2 NaI-Based Experiments

WIMP detectors based on NaI began to be used in the mid 1990s. These detectors use NaI crystals as scintillators attached to PMTs, which measure the scintillation light from a particle recoil with scintillation yields of $\sim 0.3(0.09)$ for Na(I). The crystals must be high purity in order for backgrounds to be manageable. Recoil types can be distinguished via pulse shape discrimination (PSD), although below $\sim 30\,\mathrm{keV_{ee}}$ the distributions of the two recoil types overlap making discrimination via PSD impossible.

Several experiments have used this technology: DAMA (Gran Sasso, Italy) used a PSD approach on data from ~ 100 kg of NaI detectors in 1996 [58]. The UK DM collaboration operated a range of NaI crystals underground at the Boulby Mine from 1994-2004, and reported on results from a 1.3kg crystal in 1995 [59], and subseqently a 6kg crystal using PSD in 1996 [60].

Since it is relatively easy to build very massive detectors from NaI, this makes the technology ideal in searching for evidence of an annual modulation signal. To this end, the DAMA experiment operated a total of $\simeq 100\,\mathrm{kg}$ of NaI scintillator from 1995-2002 for a total of 108,000 kg-days [61]. Analysis of this dataset shows an annually modulated variation in the signal at 6.3σ CL, and is dominated from data from the lowest reported energy bins (2-4 keV). The number of counts in this bin is clearly modulated by $\pm 1.6\%$ and is at maximum in June and mimimum in December. However, the presence of modulation would also be expected in the 2-3keV and 3-4keV bins and independent fits to the counts in these bins have not been performed. There are also many potential sources of systematics that would be annually modulated, and it is not clear that these have been fully accounted for. For example, the muon flux is known to modulate with such a period (as observed by MACRO).

Perhaps more importantly, the DAMA signal region has been explored by CDMS, Edelweiss and ZEPLIN, which all found no signal. There are systematic differences between the experiments, but studies indicate that these differences cannot account for the observed discrepancy. Since the ratio of the annual modulation amplitude to the direct WIMP rate is dependent upon the choice of halo model, the model can be changed in an attempt to reconcile these observations. However, this does not appear to resolve the conflict other than for particularly extreme models and assumptions [62]. Since the DAMA experiment is more sensitive to spin-dependent interactions (since both Na and I have nonzero nuclear spin), it is possible that the WIMP-nucleon interaction is spin-dependent. This, however, would also seem to be ruled out by recent results [63].

1.5.3 Xe-Based Experiments

Experiments based on liquid Xe (LXe) have several important advantages. Firstly, the large atomic number of the Xe nucleus provides a large event rate for WIMP-nucleon interactions, and although the event rate suffers from significant form factor supression above $60\,\mathrm{keV_r}$, this is mitigated by the rate at lower energies assuming that the threshold can be driven below $20\,\mathrm{keV_r}$. Secondly, recoil types can be identified on an event-by-event basis as in Ge detectors, but LXe detectors can be operated at 77K as opposed to 20mK. Thirdly, the technology is eminently scalable, meaning that it is not much more difficult to build a 100kg LXe detector than it is to build a 10kg LXe detector. Ge detectors, for example, suffer in scalability due to the difficult fabrication process and the need to keep the detectors contaminant-free.

When an interaction occurs in LXe, electron excitation of the Xe leads to scintillation light, and the Xe becomes ionized. When there is no electric field present, the Xe ions recombine to create a secondary scintillation signal. The first LXe-based experiments operated in such a manner, using PSD to discriminate between nuclear and electron recoils, since the scintillation signals have different characteristic timings for each recoil type. ZEPLIN I was such an experiment, had a fiducial mass of 3kg (total of 6kg) and ran for 90 live days in 2001-2002 [64]. The DAMA group also operated a LXe detector at the Gran Sasso laboratory [65].

Subsequent LXe experiments, which are currently in production, will operate in a dual phase mode. That is, both the scintillation and ionization signals will be extracted separately. A strong (kV/cm) electric field across the liquid drifts the Xe ions to the liquid surface, which are then extracted into a Xe gas phase by a 4-10 kV/cm field. The extraction field allows the electrons in the gas to undergo electroluminescence which creates another scintillation signal. The relative timing of the two scintillation signals gives position (depth) information, and although the scintillation and ionization signals are both quenched for nuclear recoils, they are quenched by different amounts such that the relative signal heights can be used to identify the recoil type.

The dual phase approach is currently being initiated by ZEPLIN II [66], and by the XENON collaboration [67].

1.5.4 Other Technologies

Several other WIMP detector technologies are currently being investigated.

The CRESST experiment operates $CaWO_4$ crystals which yield both phonon and scintillation signals. These detectors do not have the problem with surface recoil events that Ge detectors have (see Chapter 2), but a low scintillation yield makes the technique challenging. The presence of three different nuclei for a potential WIMP to interact with also makes operation difficult. CRESST II aims to build a 10kg detector consisting of 300g $CaWO_4$ crystals [68].

The PICASSO experiment operates a chamber of superheated liquid fluorinated halocarbons to search for WIMP interactions. When an interaction occurs, it triggers a phase transition and the formation of a gas bubble. The experiment is still in its preliminary stages, but results with a 40g prototype operated in the SNO observatory have been published [69]. The deadtime of the experiment can be rather large since the chamber has to be reset after each event, and the background is currently a limiting factor (there exist radioactive contaminents in the gel in which the fluorinated halocarbons are dispersed, which leads to internal alpha particles).

The DRIFT experiment using a time-projection chamber containing CS_2 gas at 50 Torr and in a 1kV/cm electric field [70]. Interactions within the gas create free electrons which subsequently create tracks. Since electron recoils typically extend further than nuclear recoil tracks, discrimination of recoil types is possible. The experiment is also naturally suited to search for annual modulation of the WIMP signal, since the directionality of the event can be discerned. However, since the gas is at low pressure one would need a huge chamber to have a significant detector mass. DRIFT is currently operating a 1 m^3 prototype chamber, with plans for a 10 m^3 detector.

Chapter 2

CDMS Detectors

2.1 The ZIP Detector

CDMS employs low-temperature crystals of Ge and Si as detectors, providing $\sim$keV energy resolution and an ability to perform particle identification on an event by event basis. The detectors, known as 'ZIP's, perform this identification by determining the type of recoil that a particle undergoes with the detector. Photons and charged particles interact by recoiling from the atomic electrons of the crystal, whereas neutrons and WIMPs recoil from the nuclei of the crystal. Hence, discrimination of WIMP events is reduced to identifying the recoil type of an interaction within the detector.

An electron recoil within the crystal results in the creation of electron-hole pairs in direct proportion to the energy deposited (ionization response), and a population of high energy (THz) athermal phonons (phonon response). As the charge carriers recombine, whether in the bulk or at the crystal surface, their total energy is released as a second population of high energy phonons. Hence, an interaction with the detector results in a collection of

Figure 2.1: A single ZIP detector inside it's Cu housing. Each detector is cylindrical, measuring 76mm in diameter and 10mm deep. In the photograph, the phonon sensor side of the detector is uppermost, showing the Aluminum and Tungsten system used to measure the detector's phonon response (see Section 2.2)

athermal phonons that have a combined energy equal to the recoil energy of the interaction. A nuclear recoil interaction results in the same process, except that the ionization response is supressed by a factor ∼0.3 (0.5) in Ge (Si). The phonon response remains the same for both recoil types. Hence, measuring the phonon and ionization reponse of the detector for each event provides a determination of the recoil energy and recoil type.

The CDMS ZIP detectors are cylindrical, 1cm thick with a diameter of 76mm. One face of the detectors contains the electronics used to measure the phonon response. This system consists of Aluminum (Al) and Tungsten (W) films patterned into structures known as Quasiparticle-Assisted Electrothermal-Feedback Transition-Edge-Sensors (QETs). Both the ionization and phonon measurement systems are described below.

2.2 Phonon Response

A particle recoil in the detector creates a population of high energy athermal phonons with a spectrum peaked near to the Debye frequency (∼10 THz in both Ge and Si). These phonons travel through the crystal undergoing anharmonic decay (a process in which a single phonon spontaneously splits into two lower frequency phonons), and isotopic scattering. Both processes conserve the total energy of the phonon population and have highly energy-dependent rates:

$$\Gamma_{AD} \propto \left(\frac{\nu_{ph}}{1\ \mathrm{THz}}\right)^5, \qquad \Gamma_{IS} \propto \left(\frac{\nu_{ph}}{1\ \mathrm{THz}}\right)^4 \tag{2.1}$$

where ν_{ph} is the phonon frequency [71].

Consequently, high frequency phonons will undergo many scatterings and decays before they are able to travel significant distances. This quasidiffuse propagation occurs until the phonon frequencies drop sufficiently for their mean free path to be larger than the mean dimensions of the crystal, which is at ∼0.6(1.0)THz in Ge(Si). At this point the phonons travel ballistically at $\sim 1-2$ cm/μs until a surface is encountered. Phonons impinging upon a bare crystal surface (the detector sides) are simply reflected back into the crystal with no change in energy, whereas a phonon encountering the Al film that is part of the QET system will be absorbed into the phonon measurement system if it has sufficient energy to create a Cooper pair in the Al ($E_{ph} > 2\Delta_{Al}$). Hence, the initial phonon population will diffuse into the Al film system on a timescale of a few μs.

The free charge carriers created by the particle recoil event are drifted across the crystal by a potential difference of a few volts. As the charges drift they are accelerated by the electric field and scatter with the lattice, and hence deposit energy into the phonon system.

This is the Neganov-Luke effect [72] [73]. Furthermore, when the charge carriers recombine at the surface electrode they deposit their energy into the phonon system, creating 'relaxation phonons'.

The fraction of the interaction's energy present in each population depends upon the electric field applied across the crystal, the recoil type and the position of the interaction within the crystal. The former two dependencies affect the energy of the phonon populations via the Neganov-Luke effect, but the latter dependency is due to an interaction that occurs between the quasidiffuse phonon population and the metal films for events sufficiently close to the surfaces of the crystal. In this case, surface events result in a very rapid conversion of the phonons into the ballistic regime, and hence the time scale upon which energy is injected into the QET system is measurably faster for surface events than for bulk events. This has important consequences for the rejection of events arising from surface contamination.

The QET sensor system consists of Al fins that transfer energy from the phonon system into W Transition Edge Sensors (TESs). A high energy phonon arriving at the phonon sensor side of the crystal is able to break a Cooper pair in the Al film, creating two quasiparticles (CDMS detectors operate at 20 mK, well below the Al superconducting T_c of 1.2 K). A fraction of the phonons arriving at the Al film have energy smaller than $2\Delta_{Al} = 340\,\mu$eV and are unable to create quasiparticles. The energy from these sub-gap phonons is not collected by the QET system and represents an unavoidable energy collection loss, estimated to be ~10%(5%) in Ge(Si) [75]. The initial pair of quasiparticles created by an incident phonon with sufficient energy decay down to the gap edge by shedding phonons, some of which are able to create further quasiparticles (if $E_q > 2\Delta_{Al}$). This cascade process transforms one or more quasiparticles into a collection of quasiparticles all with energy equal to the Al

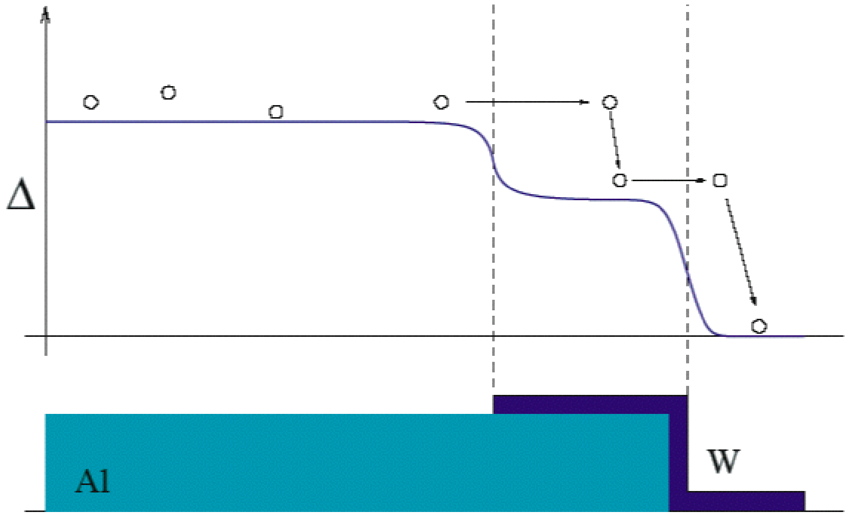


Figure 2.2: Electronic band diagram of the QET system. Once quasiparticles in the Al inelastically scatter into the overlap region and subsequently into the W system, they become trapped due to the lower energy gap.

band gap. However, the cascade also results in another collection of sub-gap phonons, and approximately half the energy in the initial quasiparticles is lost in this way.

The quasiparticles diffuse through the Al fins until they reach an Al-W interface, whereby they diffuse into the overlap region and into the W TES, where they become trapped due to the lower band gap in W. The small increase in heat in the TES gives rise to a large change in resistance, which is measured via a SQUID readout circuit inductively coupled to the TES (described below).

The quasiparticle diffusion length depends on both the mean free path of the particles and the thickness of the Al film; if the diffusion length is limited by the mean free path, increasing the film thickness results in a linear increase in the diffusion length. However, the Al-W film connection becomes less reliable as the thickness of the Al film increases, so a balance between film reliability and losses from quasiparticles that do not reach the

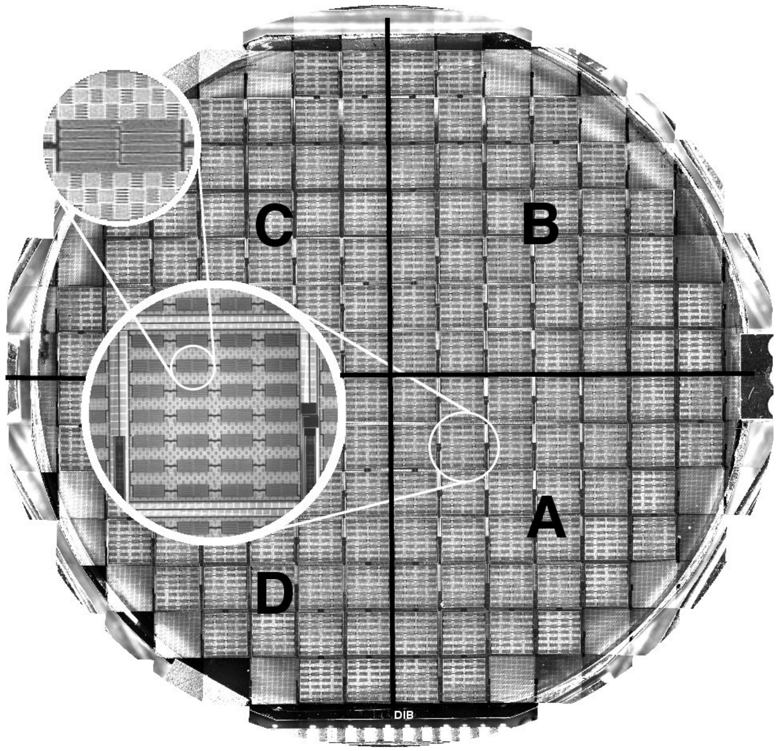


Figure 2.3: The arrangement of the Al films and W QET system, as described in the text. The upper-left insert shows the 380 μm×50 μm Al fins feeding the 250 μm×1 μm W TES. These are arranged into a 7×4 grid to form a 5 mm 'unit cell' (center insert). There are 37 unit cells in each quandrant, 148 unit cells on each ZIP detector. Each quadrant reads out an independant signal.

W TES must be found [75]. The current generation of ZIPs use an Al film thickness of 300 nm. A further issue is that a greater coverage of Al on the surface results in higher quasiparticle collection efficiency, but the distance that the quasiparticles have to travel to become trapped in the W system becomes greater and consequently fewer particles are absorbed. Diffusion simulations were performed for various numbers of Al fins of different sizes [74] [75], resulting in the current generation of ZIPs with 380 μm × 50 μm Al fins funnelling quasiparticles into a 250 μm × 1 μm W TES (See Figure 2.3). 28 of these structures are arranged into a 7×4 array to form a 5mm 'unit cell'. There are 37 unit cells in each quadrant, giving a total of 148 unit cells on each ZIP detector.

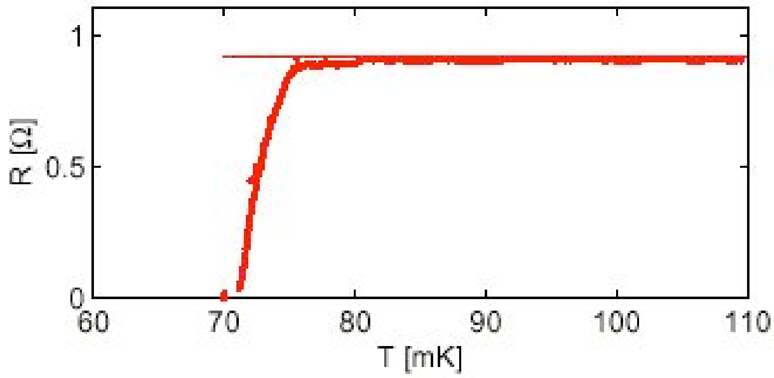


Figure 2.4: Resistance of the W TES as a function of temperature for a Ge ZIP. The T_c and transition width are typical for the detectors used in CDMS.

2.3 The Phonon Measurement System

Electrothermal Feedback

The energy that is collected by the Tungsten is read out by a TES operating in electrothermal feedback (ETF) mode. This means that the W TES is voltage biased, and is kept on it's superconducting transition. In this setup, the TES is essentially a very sensitive thermometer, able to convert a change in the W temperature into a change in resistance, and hence into a change in electric current by virtue of the bias voltage. Figure 2.4 shows the resistance as a function of temperature for a typical TES.

Figure 2.5 depicts the thermal 'circuit' that the TES sensor forms with it's surroundings. The thermal conductivities between the W, crystal, and the cryostat are sufficiently large that the phonon system of the W TES can be considered to be at the same temperature as the fridge base temperature. i.e. the main thermal impedance comes from the W electron-phonon thermal coupling. The thermal conductivity of this coupling is estimated to be

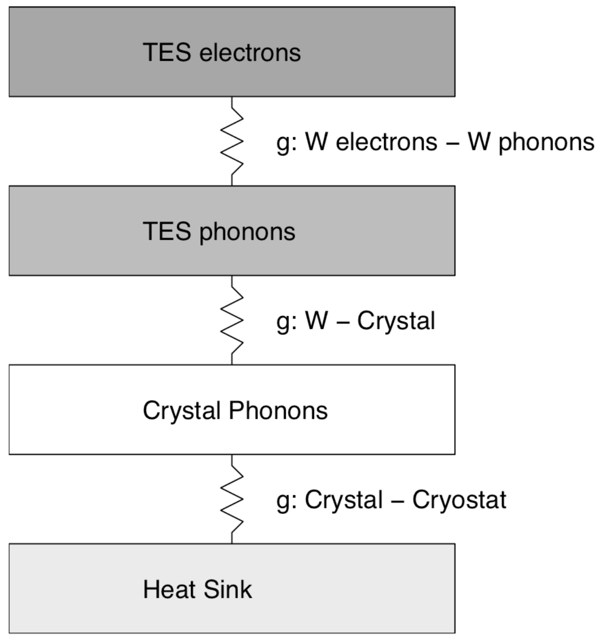


Figure 2.5: Schematic of the phonon sensor's thermal model

$g_{ep} = n\kappa T^{n-1}$. The heat flow in the TES can then be written as

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

where c_v is the heat capacity of W, T is temperature of the W electron population, V_{bias} is the bias voltage, $R_{TES}(T)$ is the resistance of W as a function of temperature (i.e. as shown in Fig. 2.4), κ is a coupling coefficient, and T_s is the temperature of the crystal and cryostat. The power n must be empirically determined, and has been found to equal 5 [76].

In order to keep the TES in equilibrium ($\frac{dT}{dt} = 0$), the net power flow through it must be equal to zero. Equation 2.2 shows how this is achieved by means of the bias voltage, which via Joule heating, allows the TES to maintain an equilibrium temperature on it's superconducting transition edge at $\sim T_c$. This power balance also underlies the negative-feedback mechanism by which the TES measures energy.

Consider the event of a small change in the TES's temperature when it is at equilibrium. Equation 2.2 yields, via a first-order expansion

$$c_v \frac{d(\delta T)}{dt} = -\frac{V_{bias}^2}{R_{TES}(T)} \frac{dR}{dT} \delta T - 5\kappa T^4 \delta T \tag{2.3}$$

where δT is the change in temperature of the TES. This equation has a solution of the form

$$\delta T(t) \sim e^{-t/\tau_{ETF}} \tag{2.4}$$

where τ_{ETF} is

$$\tau_{ETF} = \frac{\frac{c_v}{5T^4\kappa}}{1 + \frac{1}{5}\frac{T}{R}\frac{dR}{dT}(1 - \frac{T_s^5}{T_0^5})} = \frac{\frac{c_v}{5T^4\kappa}}{1 + \frac{\alpha}{5}(1 - \frac{T_s^5}{T_0^5})} \tag{2.5}$$

Here T_0 is the equilibrium temperature of the TES electron system, and $\alpha = \frac{d(\log R)}{d(\log T)}$ is a (unitless) measure of the slope of the $R(T)$ curve at the superconducting transition.

Therefore, if a small peturbation in temperature occurs, the TES responds by cooling down in an exponential fashion. For $\delta T > 0$ the cool-down becomes faster for higher temperatures since τ_{ETF} becomes larger as the temperature moves away from it's equilibrium value, and hence the temperature signal in the TES is a convolution of the incoming quasiparticle flux and a decaying exponential with time constant τ_{ETF}.

SQUID Readout Circuit

The current flowing in the TES circuit is read out by means of a SQUID readout circuit inductively coupled to the TES circuit. SQUIDs (**S**uperconducting **Qu**antum **I**nterference **D**evices) are low impedence, low-noise devices ideally suited for the TES readout. Figure 2.6 shows the setup of the TES and SQUID circuits.

The SQUID circuit is connected to a feedback amplifier that is locked to a voltage drop across the SQUID. When the current in the TES changes, the magnetic flux in the input

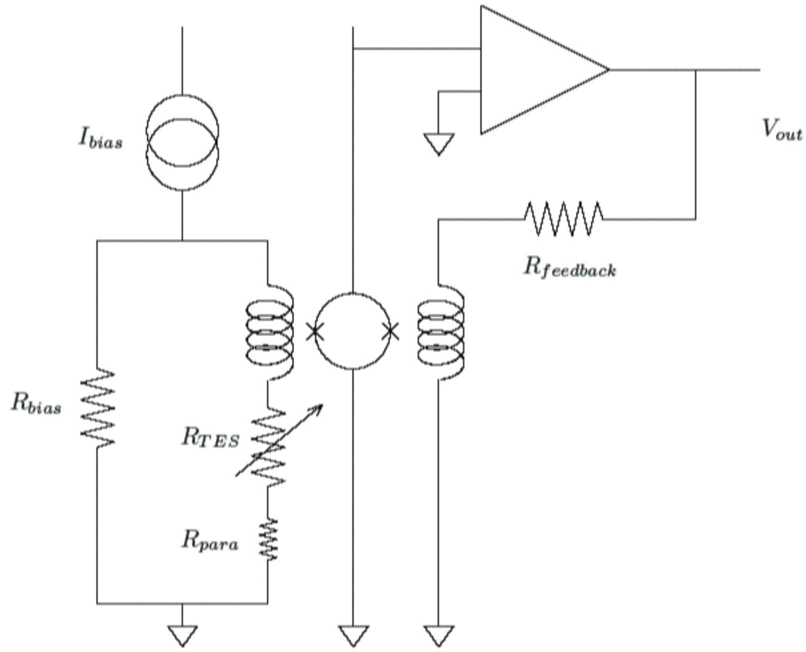


Figure 2.6: Schematic of the TES (left) and SQUID (right) circuits used to measure the phonon signal.

coil also changes, resulting in a change in the magnetic flux, and hence the voltage, across the SQUID. This change in voltage drives the feedback amplifier which feeds a current back into the feedback coil in order to cancel the changing magnetic flux in the SQUID.

The feedback to input coil turn ratio is 10, and the feedback resistance of $R_{feedback} = 1k\Omega$, giving a measured output voltage across the SQUID of $V_{out} = 10 \times I_s \times 1k\Omega$, where I_s is the current through the TES. A small parasitic resistance R_{para} is also present in series with the TES. In general

$$I_s = \frac{R_{bias}}{(R_{TES} + R_{para} + R_{bias})} I_b \tag{2.6}$$

althought the exact behaviour of I_s will depend on the TES biasing condition. i.e. it's temperature relative to the superconducting transition.

Typical values of the resistances are $R_{bias} = 20\, m\Omega$, $R_{para} = 4\, m\Omega$. The resistance of the TES obviously depends on its temperature and the bias current's value, but in it's normal state $R_{TES} \sim 1\,\Omega$, and when biased $R_{TES} \sim 200\,\mathrm{m}\Omega$.

Figure 2.7 shows the behaviour of an ideal TES for (a) normal, (b) biased, and (c) superconducting conditions.

(a) Normal. If the bias current I_b is sufficiently large, the W will be driven to a normal state with a resistance of $\sim 1\,\Omega$. Since in this case the TES resistance is much greater than either the bias resistance or parasitic resistance, Equation 2.6 reduces to

$$I_s = (R_{bias}/R_{TES})I_b \tag{2.7}$$

(b) Biased. The bias current I_b is tuned such that the TES is held on it's superconducting transistion, and it is in this regime that electrothermal feedback occurs. As described above, the power supplied through Joule heating equals the power lost to the cold substrate, and hence the power dissipated in the TES remains constant in the biased region.

(c) Superconducting. Once the bias current I_b drops below a critical current, the W becomes superconducting. In this regime $I_b \approx I_s$ (they are not quite equal due to the small parasitic resistance).

Phonon Pulses

Each ZIP detector has four phonon sensors, such that each quadrant of the ZIP measures a phonon signal individually (Figure 2.3, marked A, B, C, and D). i.e. the signal in each quadrant respresents the integrated signal over all of the sensors in that quandrant. This gives important position information on each event, by virtue of the relative timing of each

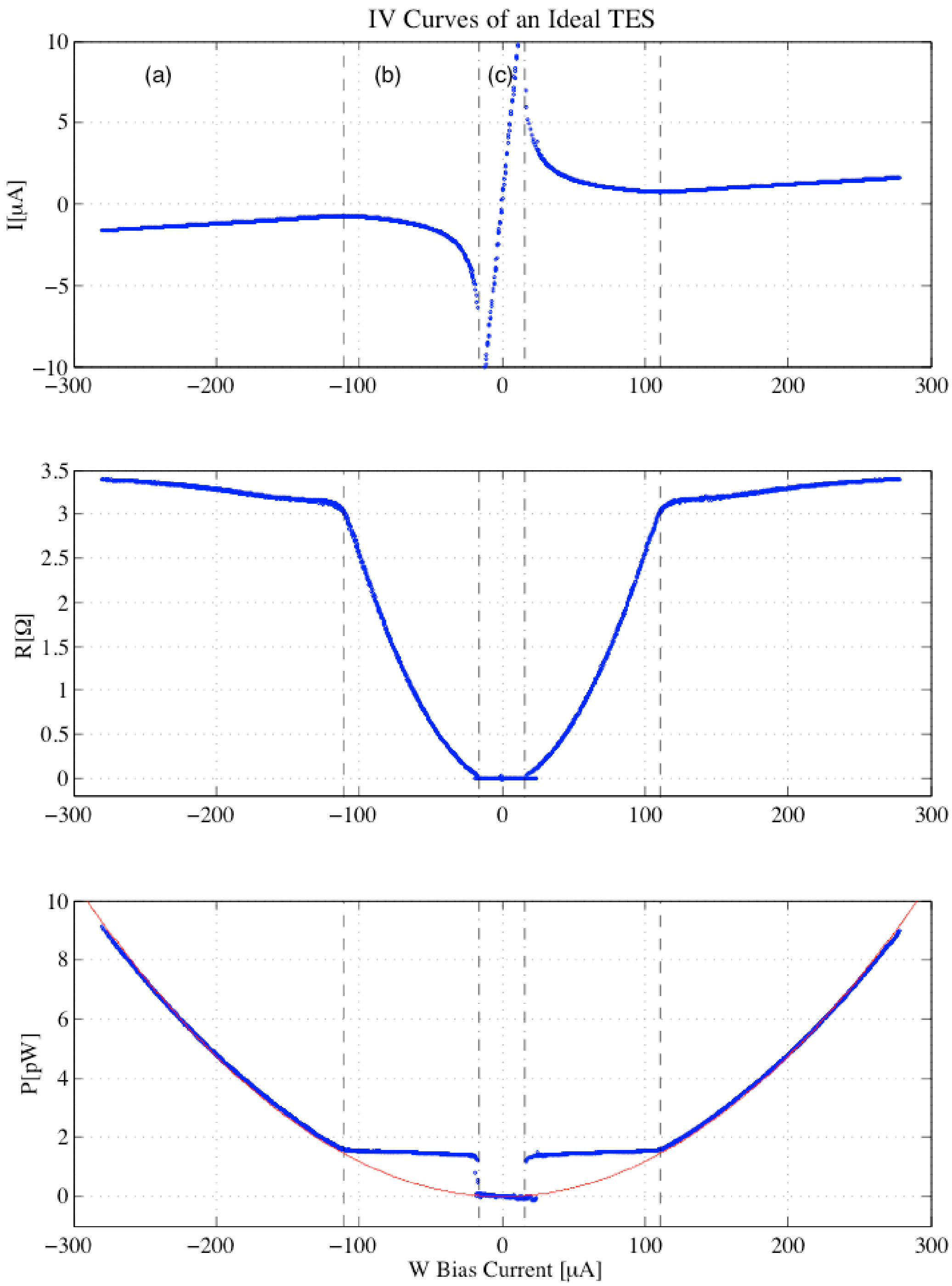


Figure 2.7: Measured Ib-Is curves for an ideal TES. The panels show (top to bottom) the current through the TES (I_s), the resistance of the TES (R_{TES}), and the power dissipated by the TES (P), as a function of the bias current (I_b). The regions shown are (a) normal, (b) biased, and (c) superconducting.

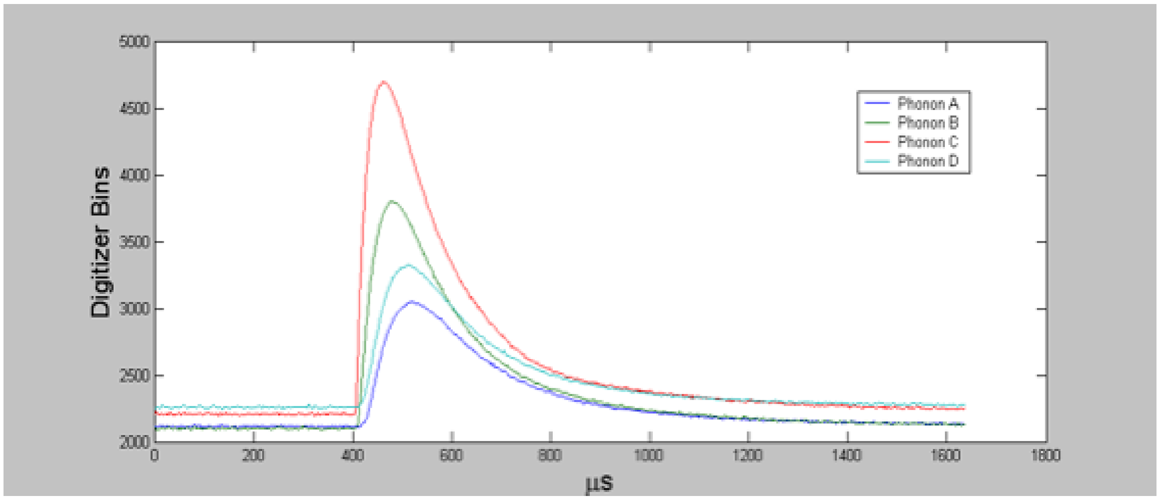


Figure 2.8: Example phonon pulses for a single event taken from real CDMS data. The relative amplitudes and timing for each quadrant's signal imply that the event's x-y position is situated in quadrant C.

of the phonon pulses. Figure 2.8 shows the four measured phonon pulses for a single event, as measured by the CDMS data acquisition (DAQ) system (see Chapter 4 for more details on the DAQ).

The combined timing and relative amplitude information in the four phonon pulses can be used to extract position information about the event. For example, the event shown in figure 2.8 has the largest amplitude in channel C, and the smallest in channel A (which is diagonally opposite to C on the detector). The rise time of the channel C pulse is also shorter than for channel A. It can then be seen that in terms of it's x-y position, the event occurred in quadrant C, slightly closer to quadrant B than quadrant D. We also have two ways of assessing the position; one based on the relative delays, and one based on the relative amplitudes.

Figure 2.9 shows the calculated position of gamma-calibration events for a Ge ZIP using the two methods of position determination, which are defined as follows:

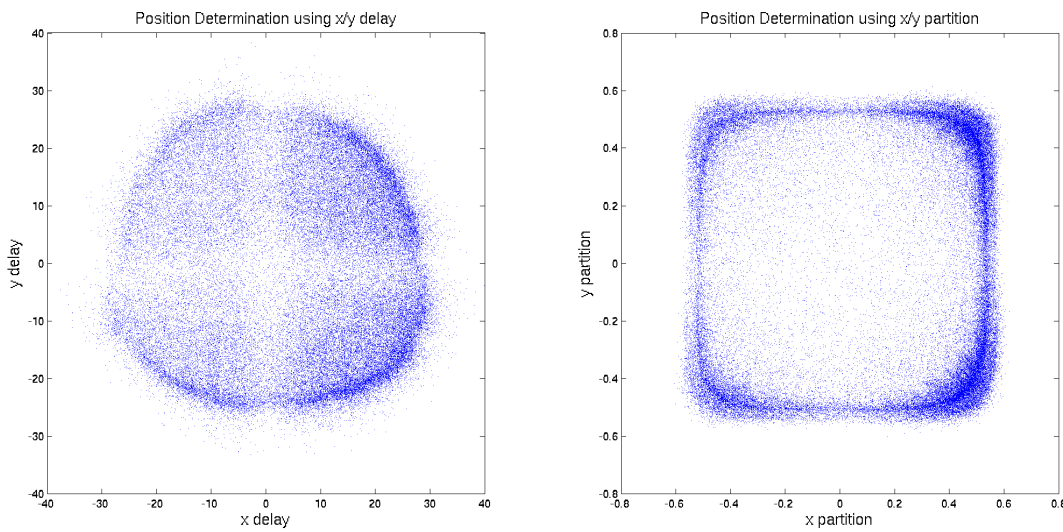


Figure 2.9: Calculated position values for bulk electron recoil events from calibration data. There are two methods by which position reconstruction can be performed: using relative delays of the four phonon signals (left), or by using the relative amplitudes of the four phonon signals (right). The source for this dataset was located such that the $+x$ side of this detector was more illuminated.

x-delay / y-delay The x-delay (y-delay) is defined as the time difference between the start of the phonon pulse in the primary channel and the start of the pulse in the horizontally (vertically) adjacent channel. e.g. the x-delay for the event in Fig. 2.8 is the difference between the start times of the channel B pulse and the channel C pulse. The y-delay for the same event is the difference between the start time of the channel C pulse and the channel D pulse The start time is defined as the time taken for the phonon pulse to rise to 20% of it's final amplitude.

x partition / y partition The partition values are defined based on the amplitudes of the pulses in each of the four channels:

$$\text{x partition} = \frac{(PC + PD) - (PA + PB)}{PA + PB + PC + PD}$$

$$\text{y partition} = \frac{(PA + PD) - (PB + PC)}{PA + PB + PC + PD}$$

where PX is the total energy deposited into channel X.

Although both methods provide useful position information, there is an inherent non-linearity present in the detector's position response, resulting in a degeneracy in timing information for events occuring close to the edge of the detector (i.e. at large radii). This is primarily due to the path that phonons created at large radii take on their way to the phonon sensors, although this effect is still far from completely understood (for more details see [77]). There is a further non-linearity arising from the fact that the 4 TES sensors on a single detector will all have slightly different values of T_c.

In order to standardize the phonon response, CDMS performs corrections to a selected set of parameters in order to homogenize their values across the crystal. This is done via a lookup table generated from bulk electron recoil events taken from calibration data. This phonon correction process is more completely described in Chapter 4.

2.4 Ionization Response

A particle recoil event in the crystal causes free charge carriers to be created, since the high-purity Ge and Si crystals become electrically insulating at temperatures below 4 K (for Ge, 10 K for Si). It takes 3.0 eV (3.82 eV) to create an electron-hole pair in Ge (Si), and so the ionization response can be measured by determining the total amount of charge created in the interaction. An electric field of a few V/cm is applied across the detector, causing the electrons and holes to drift to their respective electrodes. The charges are then collected at electrodes located on each face of the detector. There are two potential problems encountered during this process.

Firstly, although the crystals are high-purity, there will inevitably be acceptor and donor impurities present. In the CDMS ZIP detectors, impurities are about $10^{11}\,/\mathrm{cm}^3$ in Ge and $10^{14}\,/\mathrm{cm}^3$ in Si. At room temperature, these sites are fully ionized, and remain so as the crystal is cooled down. They form shallow energy levels, effectively acting as traps for electrons and holes drifting through the crystal. Hence, these sites can pose a significant problem to the collection of a charge signal. To resolve this problem, the detectors are 'baked' prior to operation - that is, photons from blue LEDs are shined onto the crystals for a period of hours for Ge, days for Si (this is done over the course of many small baking sessions, since extended periods of baking can warm up the cryogenic system significantly). This process of baking creates many electron-hole pairs within the crystal which drift into and become trapped by the impurity sites. In this way, the impurity sites become filled, and the detector is effectively neutralized. However, the neutralization state is unstable during regular running due to the electric field applied across the detector, and consequently the impurity sites gradually become empty again. Hence, the detectors must be flashed (with the same LEDs) for a much shorter span of time during regular running. The actual rate that the de-neutralization occurs depends on the event rate that the detectors see, but in regular background (WIMP search) running we flash the detectors for a few minutes every 3-4 hours.

The second complication that arises is due to recoils that occur within the top few μm of the crystal surface. These events can create charge carriers that are able to diffuse, against the electric field, to the incorrect electrode. This results in a suppressed measurement of the ionization energy, which, since a reduced ionization signature is representative of a nuclear recoil, can lead to an electron recoil being misidentified as a nuclear recoil.

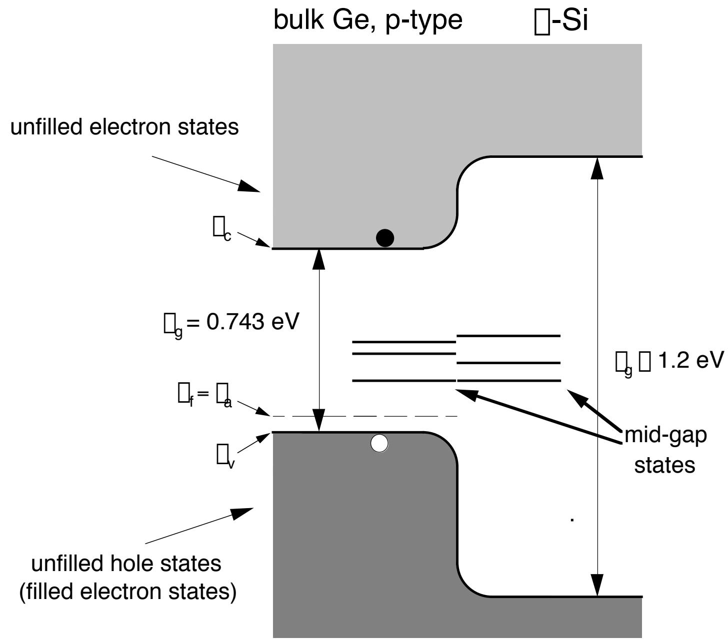


Figure 2.10: Band structure at the Ge-α-Si interface. The α-Si provides a barrier against charges created as a result of surface events that may diffuse against the field to the incorrect electrode. The band gap of the α-Si ($\sim 1.2\,\mathrm{eV}$) is greater than that for Ge ($0.743\,\mathrm{eV}$)).

This 'dead-layer' effect can be reduced significantly by the addition of a thin layer (~40 nm) of amorphous Si (α-Si) to the surface of the detector [78]. This provides a barrier against charge diffusion into the incorrect electrode, since the band gap of the α-Si ($\sim 1.2\,\mathrm{eV}$) is greater than that for Ge ($0.743\,\mathrm{eV}$) or Si ($\sim 1.17\,\mathrm{eV}$). This technique requires that the bands be well centered (as shown in Figure 2.10) in order to provide the same potential barrier to both electrons and holes.

Ionization Signal Measurement

The ionization signal is measured by means of a charge amplifier circuit, shown in Figure 2.11. The bias voltage $V_b = 3\mathrm{V}$ causes charges to drift across the detector, represented in

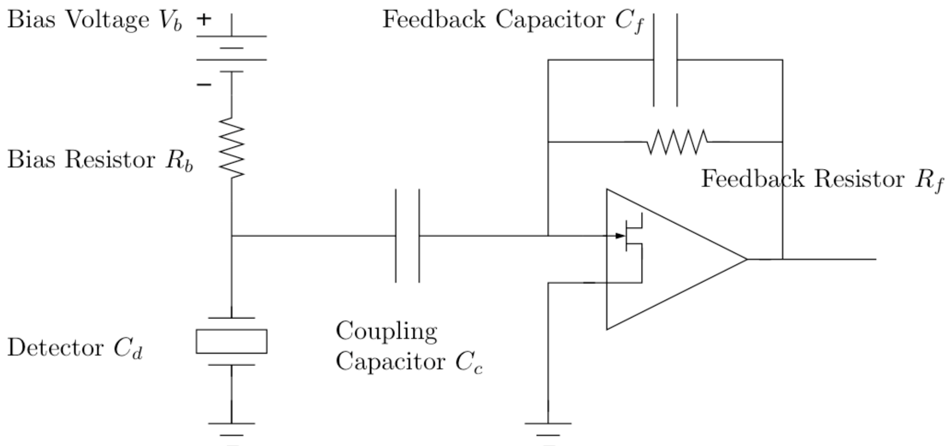


Figure 2.11: The charge amplifier circuit used to measure the ionization response in a ZIP detector. The detector is represented by the capacitor C_d.

the figure as C_d=300 nF. The drifting charges create a voltage across C_d, and the amplifier responds by lowering it's output voltage in order to maintain the JFET gate voltage at a virtual ground. The output signal is given by $V_f = Q/C_f$, i.e. the circuit produces a voltage pulse in which the pulse height depends linearly upon the total charge drifted across the crystal. The feedback capacitor then discharges through the feedback resistor with a time constant $\tau = R_f C_f \sim 40\mu s$. The coupling capacitor C_c is present in order to isolate the JFET gate from the bias voltage, although it results in a fraction of charges

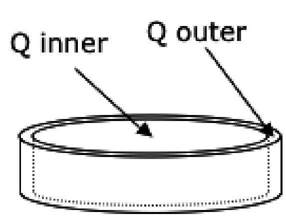


Figure 2.12: The inner and outer ionization channels on a ZIP detector. The inner and outer radii are 35mm and 38mm respectively with a 1mm gap between the sections. The inner electrode represents 85% of the volume.

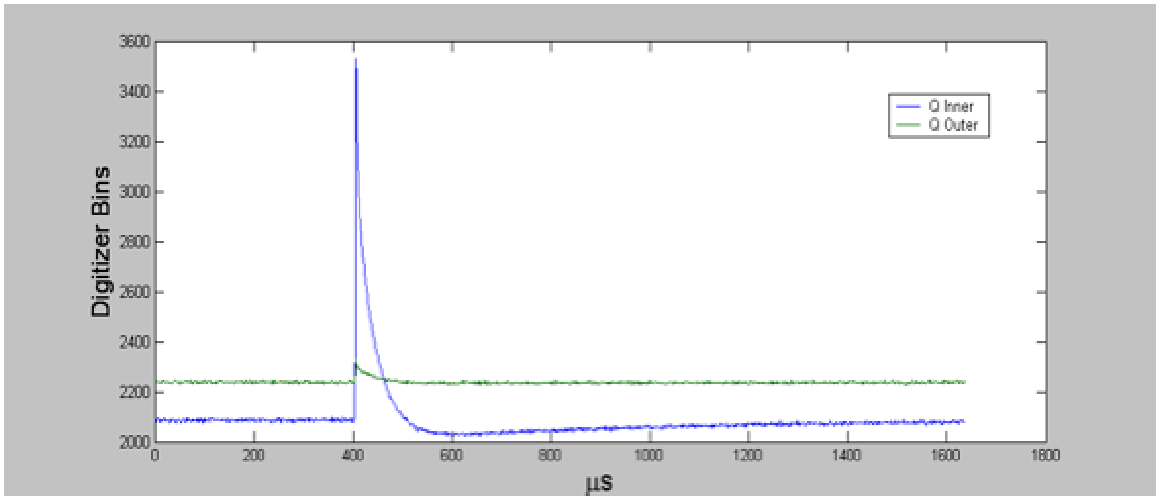


Figure 2.13: Example ionization pulses for a single event taken from real CDMS data.

$C_c/(C_c + C_d)$ not appearing on the feedback capacitor. This is minimized for a large value of the coupling capacitance; a value of 300 pF is used in order to avoid breakdown of the capacitors while operating at cryogenic temperatures.

Each ZIP detector has two ionization measurement channels; an inner and outer electrode (see Figure 2.12). This allows a fiducial volume cut to be made upon those events with energy deposited primarily into the inner electrode's volume. The outer electrode region suffers from an decreased charge collection efficiency due to a non-uniform electric field, and hence the analysis selects only events for which $> 90\%$ of the total ionization energy is measured in the inner electrode.

Figure 2.13 shows ionization traces taken from a single event. Note that there is no shape information present in the leading edge of the ionization pulses. The best estimate of the event start time comes from the ionization pulse start time.

2.5 Event Identification

As already described, a single detector event creates both phonon and ionization signals. How the event traces are digitized and their energies extracted is described in Chapter 4; here we define the quantities that CDMS uses in order to perform event identification and discrimination.

As described above, the total energy contained in the phonon signal represents the recoil energy of the event plus the phonons originating from the Neganov-Luke effect. Hence, the recoil energy can be written as

$$E_r = P - \frac{V_b}{E_{eh}} Q = P - \alpha Q \tag{2.8}$$

where E_r is the recoil energy, P is the total measured phonon energy, Q is the total measured ionization energy, V_b is the bias voltage applied across the crystal (= a few volts), $E_{eh} = 3.0$ (3.82) is the energy in eV required to break an electron-hole pair in Ge (Si) and $\alpha = V_b/E_{eh}$.

Since it is known that the creation of free charges is $\sim \frac{1}{3}$ $(\frac{1}{2})$ as efficient for nuclear recoils compared to electron recoils in Ge (Si), we create a parameter known as the *ionization yield*

$$y = \frac{Q}{E_r} = \frac{Q}{P - \alpha Q} \tag{2.9}$$

hence for electron recoils $y \sim 1$, and for nuclear recoils $y \sim \frac{1}{3}$ $(\frac{1}{2})$ in Ge (Si).

The exact value of the ionization yield depends not only upon the target atoms, but also on the recoil energy of the events [79]. In general, we expect the recoil types to form bands, for which the mean and 1σ-width take the form

$$\text{BandCenter} : y_{band} = \frac{aE_r^b}{E_r} \tag{2.10}$$

$$\text{BandWidth}(1\sigma) : \ \Delta y_{band} = \frac{c(aE_r^b) + d}{E_r} \tag{2.11}$$

In practice, the bands are formed by fitting this functional form to calibration data on a detector-by-detector basis (this process will be detailed in Chapter 6).

Chapter 3

The CDMS-II Experiment

3.1 The Soudan Underground Laboratory

The CDMS-II experiment is located on the 27th level of the Soudan Underground Laboratory in northern Minnesota, about 100 miles north of Duluth (see Figure 3.1). The former iron mine ceased operation in 1963, and is now run by the state's Department of Natural Resources (DNR) as a joint experimental physics and tourist site. Level 27 contains two primary chambers, with the CDMS experiment located in one and the MINOS collaboration in the other. Tourism is a large part of the site; thousands of tourists visit the mine each year for historical mine tours and science tours of both experiments. A yearly open house event sees around 500 visitors visit the site on a single day to learn about both experiments. Figure 3.2 shows the author presenting to a group of visitors during the 2005 open house event.

CDMS-II began full operation at Soudan in mid-2003. The first data run was in operation from October 2003 - January 2004, used a single tower of 6 ZIP detectors (4 Ge, 2

Figure 3.1: The Soudan mine, MN, where the CDMS-II experiment is located

Figure 3.2: The author, describing the CDMS-II experiment to a group of visitors on the open house event, May 2005

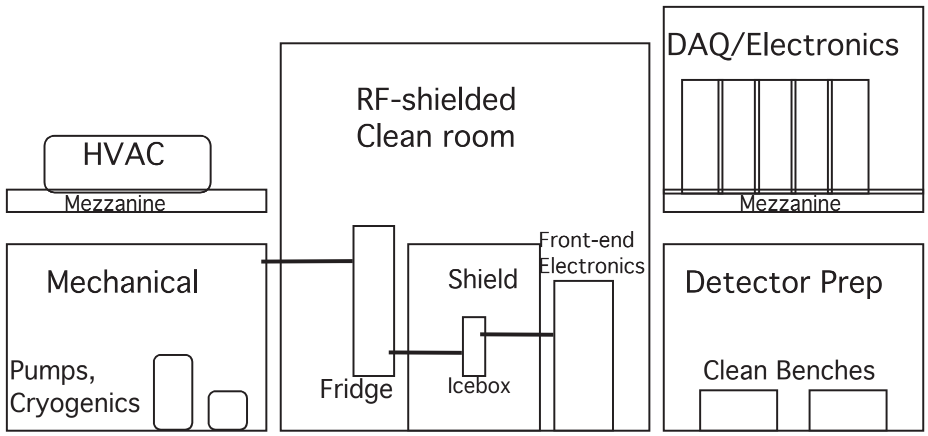


Figure 3.3: Layout of the CDMS-II experimental chamber and support areas.

Si) and will be referred to henceforth as 'Run 118'. The second data run ran from March 2004 - August 2004, used two towers for a total of 12 ZIP detectors (6 Ge, 6 Si), and will be referred to as 'Run 119'. From late 2004, CDMS began to install an additional 3 towers, making 30 ZIPs in total. Associated cryogenic improvements were also installed, and operation of the final Soudan data run is currently scheduled to begin in Spring 2006. This work will focus on the 2-tower data run, Run 119.

Figure 3.3 shows the layout of the CDMS-II experimental chamber. The detectors, associated shielding, refrigerator cryostat and front-end electronics are located in an RF-shielded, Class-10,000 clean room. The Data Acquisition (DAQ) hardware and software is located on the upper level in a mezzanine room with electronics connected down through the upper wall of the clean room. Cryogenics and monitoring software for operation of the dilution refrigerator are located on a separate cryogenics pad just outside the RF-shielded room.

In addition, CDMS has the use of a building at the surface of the mine, where a cluster of computers is used to run the analysis software and backup data to tape. There is also remote access to the cryogenic and data acquisition system from this building, which is extremely important since access to the mine is limited during off-hours. The surface building and the analysis farm are further described in Chapter 4.

3.2 Cryogenics

The ZIP detectors are held at a temperature of 40 mK by use of a Oxford Kelvinox S-400 dilution refrigerator. The refrigeration system consists of a series of concentric radio-pure copper (oxygen-free electronic (OFE) copper) cans at different temperature stages, the inner cans being progressively colder (see Figure 3.4).

The system nominally provides $\sim$ 400 mW of cooling at 100 mK, increasing quadratically with temperature. The main refrigerator housing is located outside of the experimental shield (due to the difficulty of creating a radiopure fridge); the thermal connection to the icebox is provided by a $\sim$2m coldfinger (shown in Figure 3.4). Radioactive background produced by fridge components include ^{60}Co from the stainless steel, and isotopes located in the Indium vacuum seals and Silver solder.

Once the system has been cooled down to base temperature (which is nominally 10 mK but in practice has been 40 mK), regular supplies of liquid Helium and liquid Nitrogen are needed to keep the fridge cold (i.e. they are used to keep the Helium and Nitrogen baths filled). Since mine access is limited to 10 hours a day, 6 days a week, it is necessary to have the ability to monitor and control the fridge remotely. This is done via a software

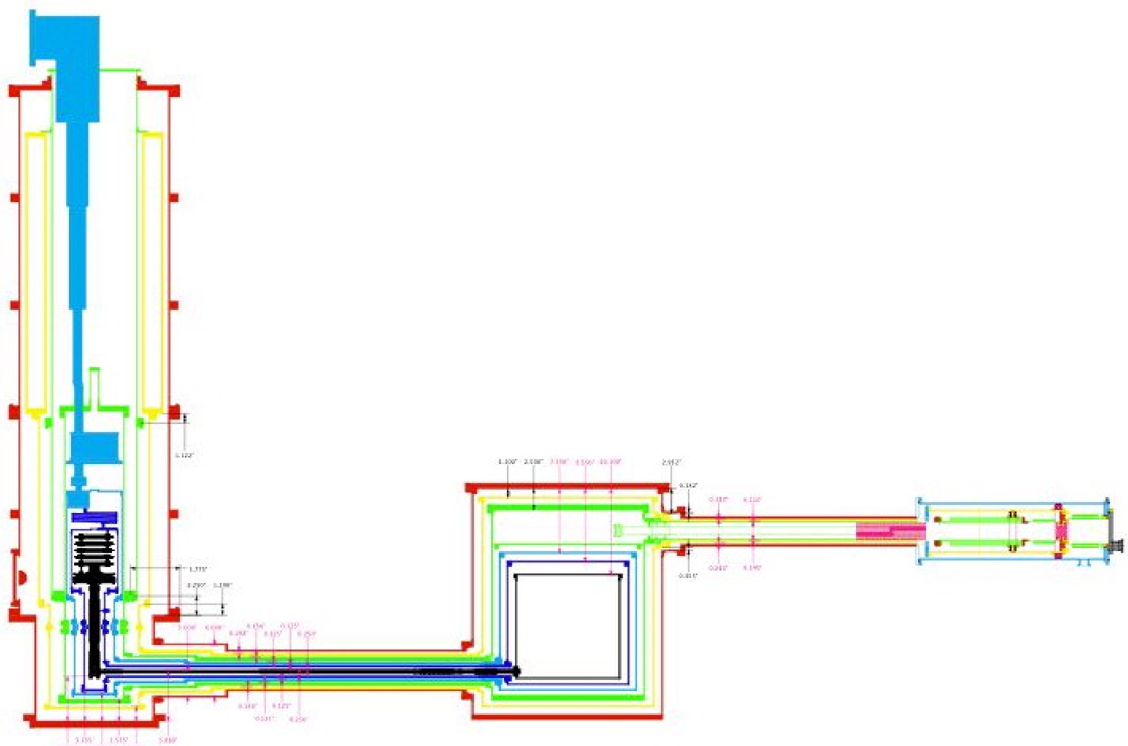


Figure 3.4: Outline of the CDMS dilution refrigerator setup. The main refrigerator housing containing the He circulation loop is on the left, the icebox is shown in the center, and the detector electronics are fed out through the 'E-stem' section shown on the right. The temperature stages are shown with nominal temperatures of 300K (red), 77K (yellow), 4K (green), 600mK (cyan), 50mK (dark blue), 10mK (black).

interface called 'Intellution' (Intelligent Solution), accessible from both the mine and the surface building computers. This allows for many cryogenic operations (including liquid transfers) to be performed from the surface. An 'Intelligent Gas Handling', or 'IGH' system controls the cryogenics below 4K, that is, the systems that deal primarily with circulation of the ^{4}He/^{3}He mixture.

3.3 Backgrounds and Shielding

Sources of Background

The dominant source of background is the natural radioactivity of the rock walls of the experimental cavern and the materials surrounding the detector assembly. Photon backgrounds arise from the decay chains of ^{238}U, ^{232}Th and ^{40}K, which includes an airborne component of ^{222}Rn gas. The air Radon levels in the mine are substantially higher than are present at the surface; the measured rate of ^{222}Rn decays in the mine varies from 200 Bq/m^3 in December to 600-800 Bq/m^3 in June.

The Soudan Underground Laboratory sits at a depth of 2341 ft below the surface (911 ft below sea level), equivalent to 2090m of water overburden. This effectively reduces the incident muon flux by a factor 5×10^4 compared to sea level rates, yielding a measured rate of $2.21 \pm 0.03 \times 10^{-3}\,\mathrm{muons\,m^{-2}\,s^{-1}}$ [80].

The cavern and experimental assembly act as a source of neutrons, primarily resulting from (α, n) reactions induced by alpha-decays from Uranium and Thorium chain elements. The neutron rate from the rock due to such processes is $2.1 \pm 0.2 \times 10^{-8}$ neutrons/g/s. The spontaneous fission of ^{238}U and ^{232}Th contributes a further 2.7×10^{-9} neutrons/g/s.

Cosmic ray muons also act indirectly as a further source of neutrons. These neutrons are generated by muon-induced spallation, muon capture, or by muon-generated electromagnetic or hadronic showers. Muon-induced neutrons represent a substantially more problematic background than neutrons coming from radioactive decays, since they typically have much greater energies and so more readily penetrate the shielding [81]. The flux of muon-induced neutrons, however, is much lower than the (α, n) neutron flux.

Shielding

Figure 3.5 shows the shield setup for the CDMS experiment. The primary shielding is cylindrically symmetric, and consists of three layers:

i) An inner layer of 8.6cm of polyethylene,

ii) A 22.5cm thick layer of lead, the innermost 4.5cm of which is 200 year-old (and hence low-activity) lead, and

iii) An outer layer consisting of 40cm of polyethylene.

The inner layer of polyethylene is present since it helps to mediate neutrons that are created within the shield. This shield reduces incident flux of neutrons (arising from natural radioactivity) by a factor 1×10^6, and effectively neutralizes electromagnetic backgrounds emanating from outside of the shield.

Surrounding the primary shield on all sides are forty 5cm-thick scintillator paddles, used to reject detector events coincident with muon events occurring inside the paddles. Each paddle contains a piece of plastic scintillator, a single Hamamatsu R329-02 photomultiplier tube, and an acrylic light guide. The panels each measure approximately 76cm$\times$114cm$\times$5cm

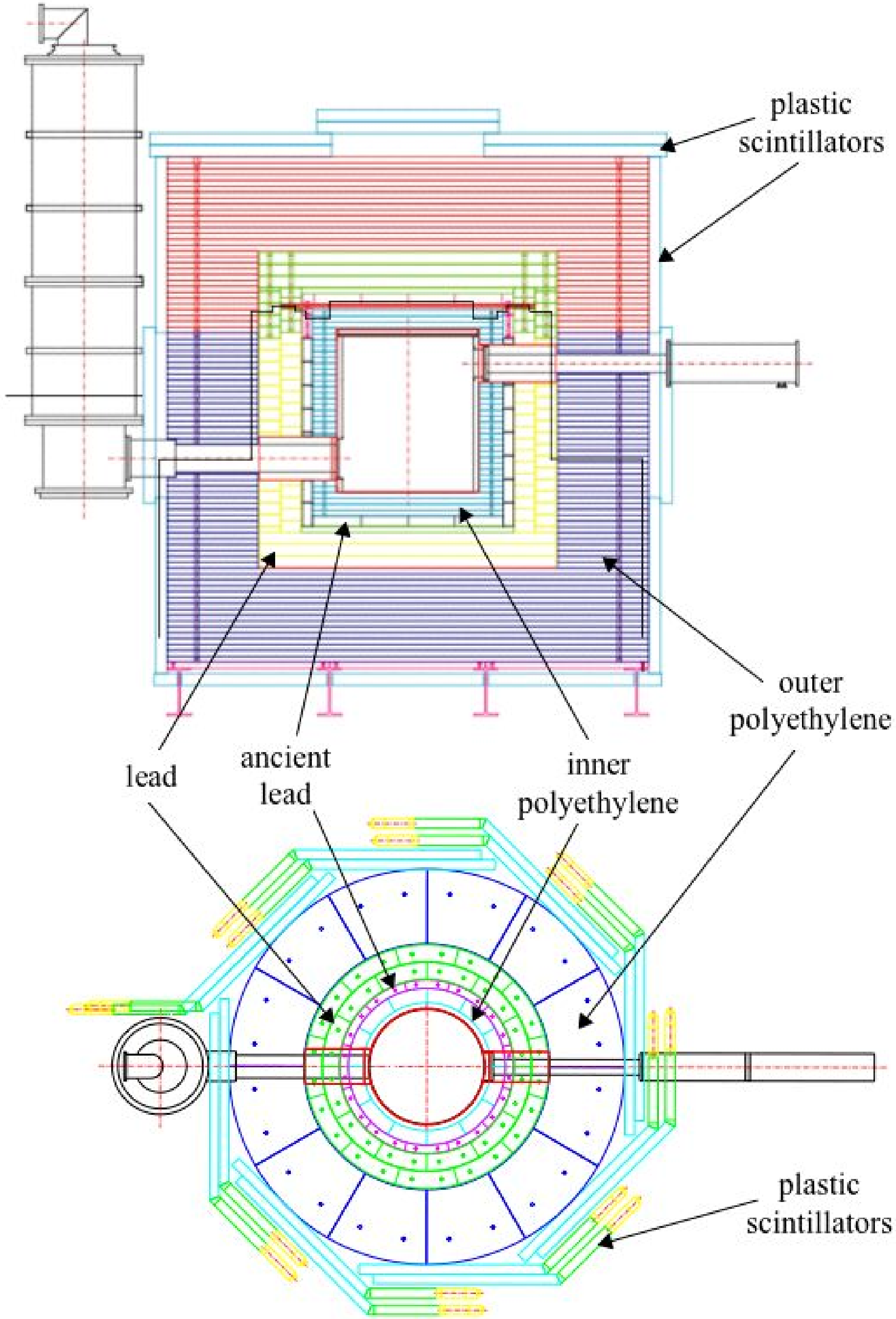


Figure 3.5: Schematic of the CDMS detector shielding. The uppermost image (side view) shows the central icebox surrounded by the polyethylene and lead shields, with dilution refrigerator affixed. The upper sections of the primary shield are colored differently in the diagram since these are the parts that are lifted away in order to access the icebox (eg. for detector installation).

Figure 3.6: Top view of the (closed) CDMS icebox. The upper sections of the shield have been removed to facilitate access (cf. Figure 3.5).

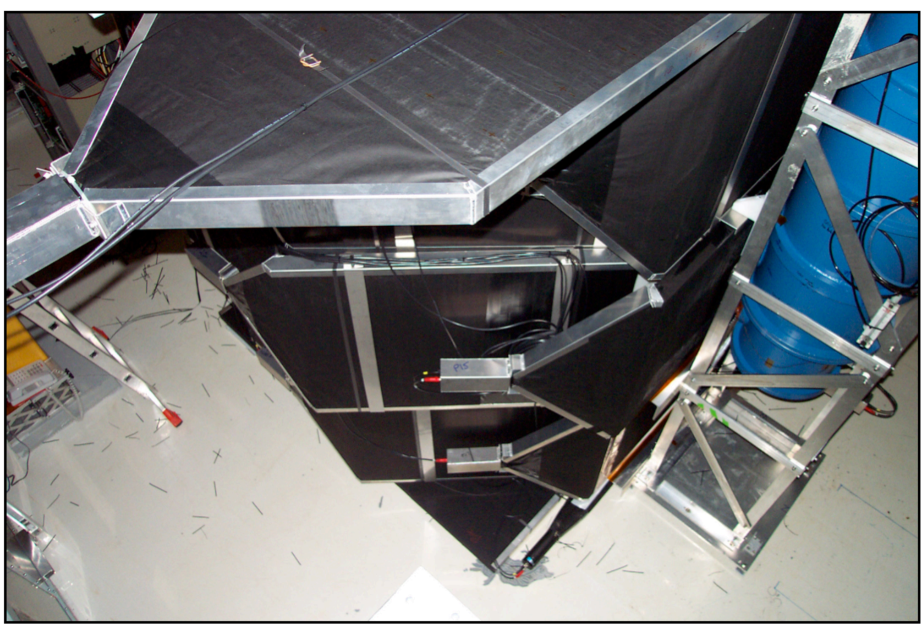


Figure 3.7: The muon veto scintillator panels surrounding the primary lead and polyethylene shielding.

(some are cut down slightly in order to fit around the fridge and electronics stems). The muon veto system is able to 'tag' muons at the $99.98 \pm^{0.02}_{0.03}$ % level. A muon is considered tagged if an event above threshold is observed in one or more bottom veto scintillator paddles, and a scintillator event is simultaneously seen in one or more side or top panels. i.e. it is assumed that all muons that pass through the bottom panels must also pass through the side and/or the top panels.

The veto system also includes a set of 'pulsers', by which a source of blue light is fed into each scintillator paddle with optical fibers. This allows for periodic pulsing of the light source (controlled by the DAQ) and hence for calibration checks of the veto system.

3.4 Detector Assembly and Cold Electronics

The ZIP detectors used in CDMS are 250g (100g) cylindrical crystals of high-purity Ge (Si), with a diameter of 76 mm and a depth of 10 mm. Detectors are arranged into 'towers' of 6 detectors each, with the ZIPs in a vertical face-to-face arrangement. Figure 3.8 shows the setup of a ZIP tower along with it's associated cold electronics.

The SQUIDs and FETs are combined into a single card called a SQUET. There is a single SQUET card for each ZIP detector. Several considerations led to the setup shown in Figure 3.8.

The FET cards need to be located as close to the detectors as possible (to minimize capacitance to ground) and need to be connected via tensioned wires to minimize microphonic pickup. However, the FETs operate at 140 K and also dissipate ~ 45 mW of power, so this must be balanced with the need to have them close to the detectors. This heatload

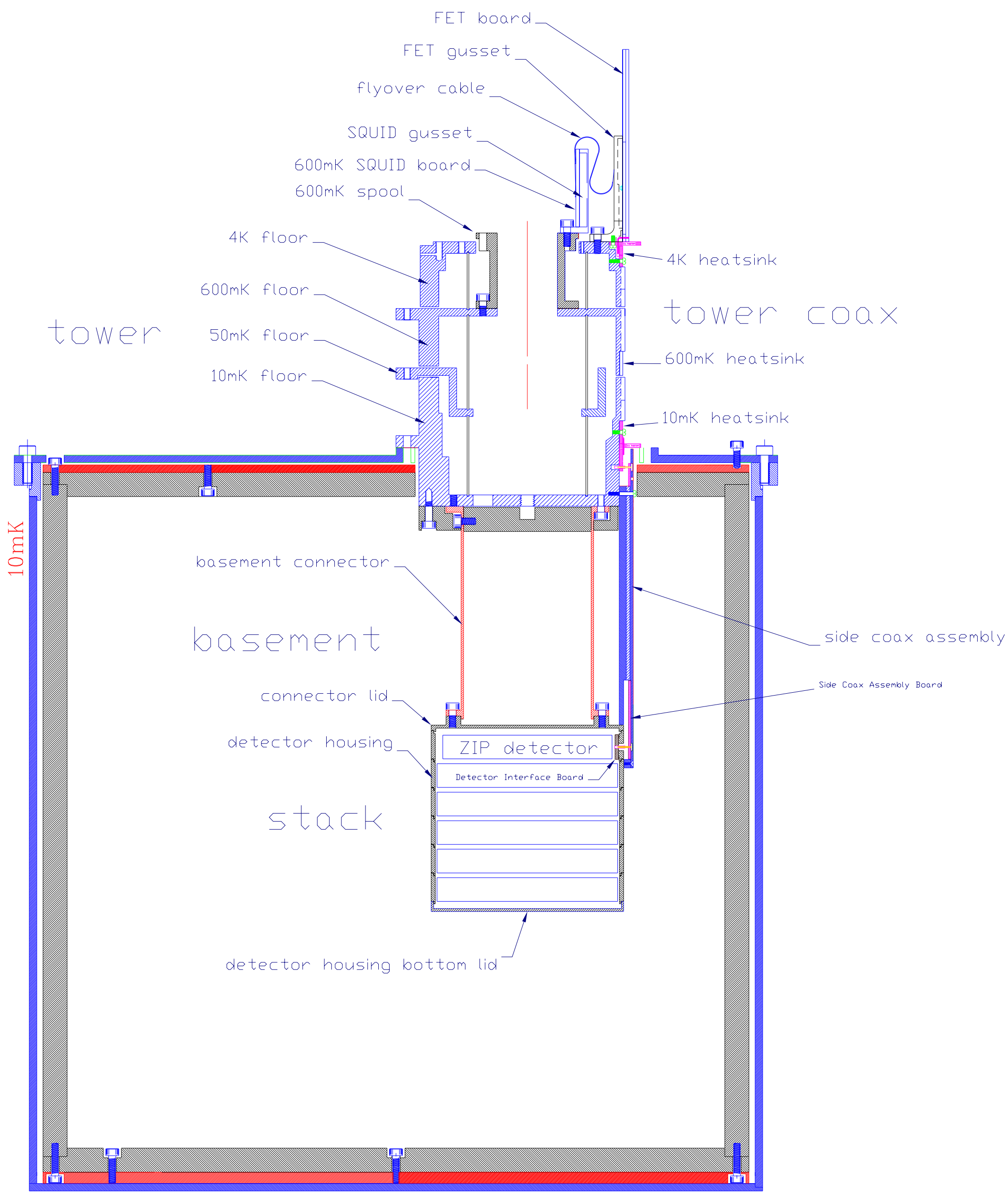


Figure 3.8: Assembly of the detector tower and cold electronics in the icebox. See text for details of the setup.

can be tolerated on the 4K stage. This setup means that the wiring between the FETs and detectors must be heat-sunk to each intervening temperature stage, and consequently the tower must bridge the various temperature stages.

The SQUIDs must be operated at $\ll 9$ K to be below their superconducting transition, and although they too could be operated at the 4 K stage, they are placed at the 600 mK stage in order to minimize the thermal noise from the associated bias resistors.

Due to radioactive backgrounds, fiberglass circuit boards cannot be used (source of ^{40}K), and nor can stainless steel (source of ^{60}Co). All circuit boards are made of multilayer Kapton laminates, and thermal isolation between tower stages is provided by thin-walled high purity carbon tubes. All solder joints are created using custom-made low-activity solder [82].

Tower

Each tower consists of four hexagonal copper sections joined by three thin-walled graphite tubes (the graphite represents the dominant heat-load between temperature stages, but is sufficiently low as to not compromise the operation of the cryostat). The top stage of the tower holds six SQUET cards, and the bottom holds six detectors. Due to the modular design of the tower, it can be completely assembled on the bench in an above ground lab and inserted into the icebox as a single unit.

The wires connecting the detectors to the SQUET cards are mounted on the sides of the tower assembly. Conventional coaxial cables with a dielectric insulator are unsuitable for this purpose, since it is possible for a buildup of charge to occur via the triboelectric effect (where a difference in work function between two neighboring materials results in a

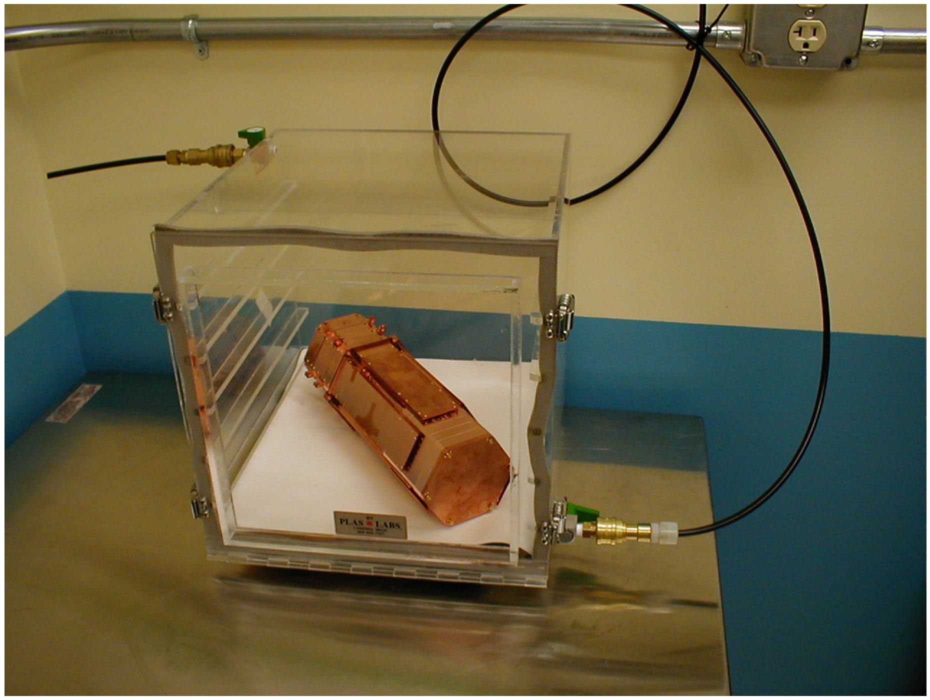


Figure 3.9: A complete tower of 6 ZIP detectors awaiting icebox installation. The tower is sitting in a purge chamber being fed old (low Radon) air. The SQUET cards have not yet been installed on the top stage.

buildup of charge at the interface). Hence, the configuration used in the tower is a 'vacuum coax' configuration where each wire is positioned inside a machined copper channel located in the tower face. The wires used are NbTi superconducting wires, and are shielded with a thin copper cover placed over the copper channel in the tower.

Striplines

The dominant considerations in designing the cables that connect to the SQUET cards to room temperature are low thermal conductivity and ease of thermal sinking. The Soudan icebox is designed to hold up to 7 towers, each containing 6 detectors, each of which require 50 wire connections to their SQUET card (20 for detector connections, 2 for LEDs used for baking, 3 for FET heaters and 25 connected to ground). The striplines that are used

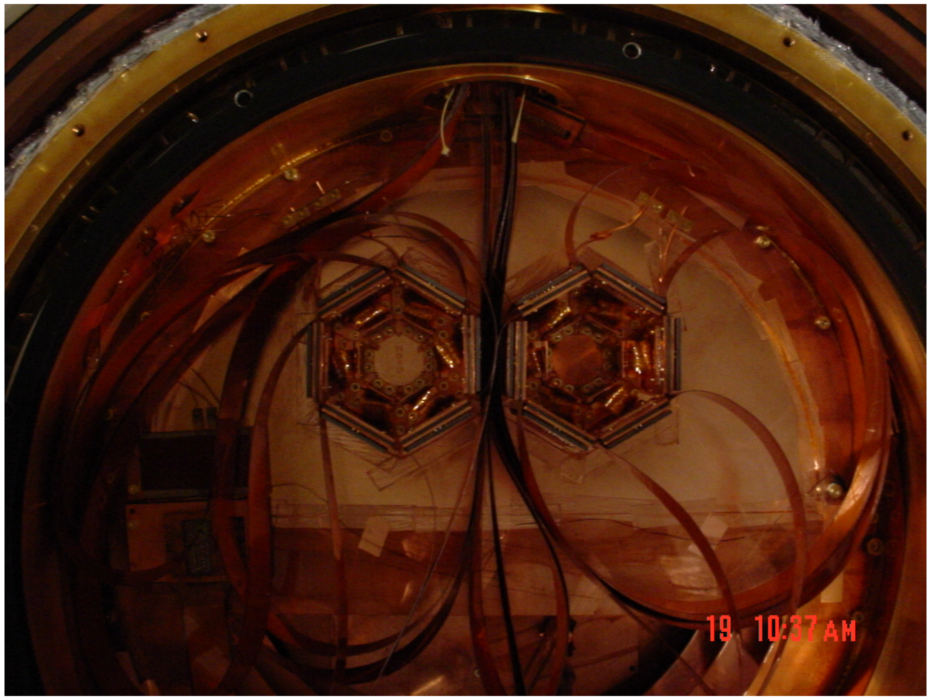


Figure 3.10: Aerial view of the icebox with 2 towers of detectors installed. The striplines can be seen entering via the fridge electronics stem, and attaching to the 12 SQUET cards present on the towers (1 per detector).

are ~300 cm in length, 2.5 cm across, and are fabricated from multiple shielded layers of Copper and Kapton, housed in a flexible cable. Each has a 50-pin connector that attaches to a SQUET card (so there is a single stripline per detector), and is fed out through several Cu pressplates and shims which provide thermal sinking at the 4 K, 77 K and 300 K stages. This design of the striplines allows a collection of up to 42 to be stacked and fed through the shims together in a relatively small bundle.

3.5 Warm Electronics and Data Acquisition

The signals from the warm end of the striplines are connected to a series of Front-End Boards (FEBs), one board per detector. A single FEB board can perform a host of operations:

- A FEB has complete control over the state of it's detector: i.e. it can perform all the SQUID biasing required to operate the phonon sensors, biasing/grounding of the charge sensors, etc.

- It is able to control the LEDs using for baking and flashing the detectors.

- It can send a large current pulse through to the SQUID to heat it locally and allow trapped flux to escape ('zapping' the SQUID).

- It can activate the detector's FET heater, required to warm the FET to 140K prior to its operation.

- The FEB receives signals from the detector's six channels (4 phonon, 2 ionization), amplifies them and passes them down the electronics chain.

The FEBs perform amplification of the detector signals such that the signals become $\sim$1V in magnitude, and hence any noise gained by subsequent transmission of the signal down the electronics chain will be much smaller than the noise already present (and amplified) from the cold electronics.

The detector signals are passed upstairs to the DAQ/Electronics room (see Figure 3.3) to a series of **R**eceiver-**T**rigger-**F**ilter (RTF) boards. A single RTF board per detector monitors the incoming signals for a pulse that is above the defined threshold value. This trigger threshold value can be user-defined, and the boards can be set to trigger on phonon signals only, charge signals only, or upon both phonon and charge signals. Trigger logic information from all of the RTF boards is sent to a Trigger Logic Board (TLB), which sets a global trigger bit based on the RTF triggers and it's own settings.

The detector signals from the RTF boards are constantly being sent though a set of digitizers. These Struck SIS3301 digitizers operate at 65MHz, and the data is oversampled at this rate to create digitized traces at 1.25MHz. The SIS3301s are 14-bit digitizers run with a maximum digitization window of 5V ($\pm$2.5V). When a global trigger bit is set to 1 by the TLB, the digitized traces from all of the Strucks are stored on disk with a suitable pre-trigger window. This is done for every detector, irrespective of which one(s) may have caused the global trigger.

The trigger logic from the RTF boards is also stored in a history buffer board (Struck SIS3400). This stores trigger logic for all detectors (i.e. was there a hit above threshold?) in 1μs binning, for the past 1 second of running. When a global trigger occurs, this history buffer information is written to disk along with the digitized event traces.

The signals from the collection of muon veto panels is read out along a completely separate electronics chain. The signals pass to a set of Joerger VTR812 digitizers, which sample at 5MHz. The digitized veto traces are also written to disk when the detector traces are, i.e. when a global trigger occurs.

A parallel signal from the veto panels is routed through a set of logic-step signal threshold boards which record trigger logic for the panels in a similar way as is done for the detector signals. These trigger logic values are sent to, and recorded by, the same Struck SIS3400 used for the history buffer; hence veto trigger history information is also written to disk along with the detector trigger history information when a global trigger occurs. Yet another parallel set of veto trigger logic values is sent to a Joerger VS64 board which records hit and trigger rates for diagnostic purposes.

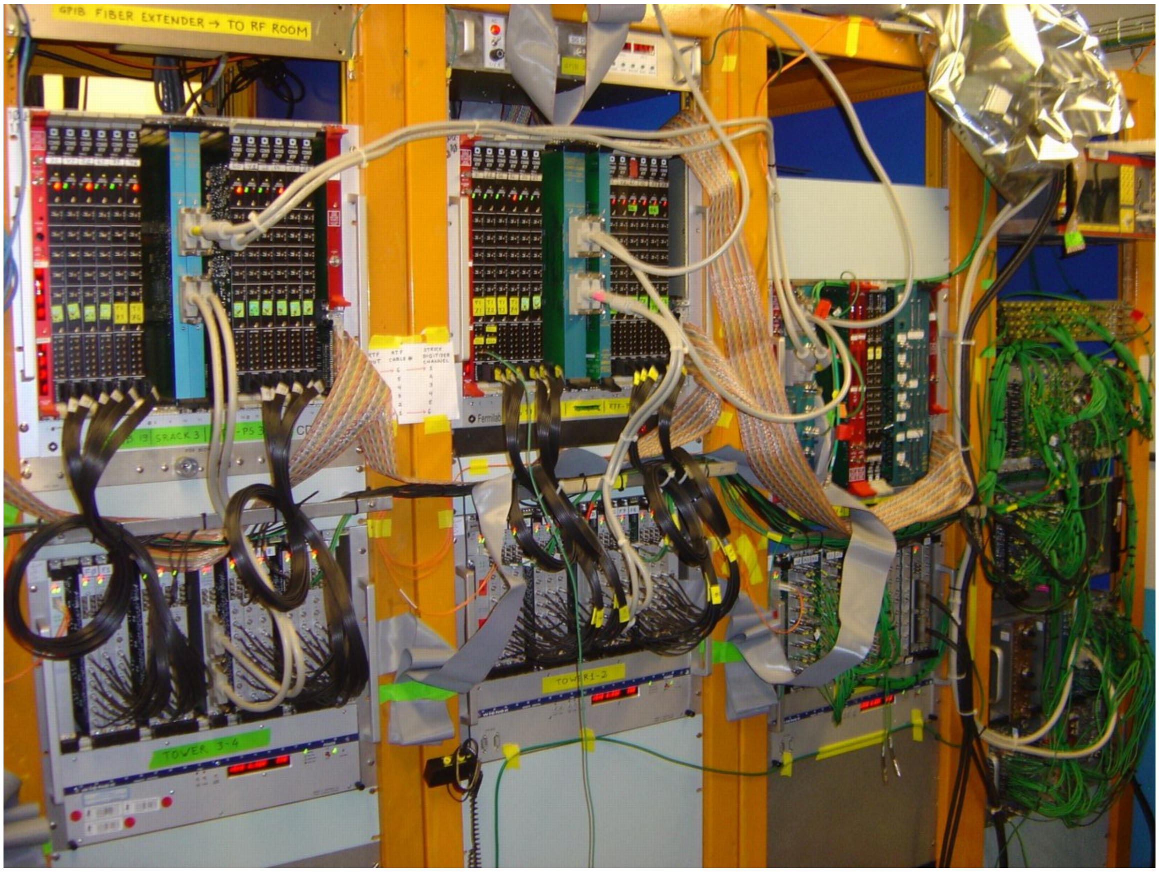


Figure 3.11: The Data Acquisition hardware in the DAQ/Electronics room. The upper-left rack row contains the RTF boards (black), with the Struck digitizers in the racks below (grey). The far-right section contains all of the veto electronics (green cables). This photo was taken post-Run 119, after new hardware had been installed for the operation of 5 towers of detectors.

The FEB, RTF and TLB boards are all controlled by computer using a General Purpose Interface Board (GPIB). The interface between DAQ software and hardware is described in Chapter 4.

Chapter 4

CDMS-II Analysis Chain

4.1 Overview

The CDMS-II Analysis Chain consists of two primary parts: the Data Acquisition (DAQ) system located in the mine (see previous chapter), and the Soudan Analysis Cluster (SAC) located in the 'surface building' a short distance from the mine.

Raw data is output to disk in the mine by the DAQ in the form of binary files with a header/detail format. A specific header value denotes the start of an event, and any number of other header values are then written to denote specific event information on a detector-by-detector basis. This information includes digitized detector traces, livetime values, event number, detector trigger history, and veto traces. Each header is followed by details relevant to that header, the format depending on the type of detail in question. The complete detector record is covered in Section 4.3.

Once raw data is written to disk, DAQ scripts queue it up to be backed up to a tape drive, and simultaneously copy the file to the SAC surface computers. Auxiliary files containing information on the current data run are also copied to the surface, along with 'status' files giving the SAC computing farm instructions on how to deal with the new data files it finds.

The SAC cluster 'farms' out jobs to machines in the cluster via a set of automatic perl scripts, based on the contents of the status file found with each event file. Each job will begin a session of *DarkPipe*, the CDMS-II analysis software (written in MATLAB). This will produce a set of output data files containing event information to be used for analysis. Once each data file in a given data series has been run through Darkpipe and has created output files, all the output files are combined by a process called *PipeCleaner*. Thus the final output is a handful of files representing quantities for the entire data series.

4.2 Involvement

Since this thesis details work done by many members of the CDMS collaboration, it is necessary to note which work was done by the author. A similar section will follow in Chapter 6 detailing the analysis work; here we note the work done on the analysis chain and related hardware/software.

Generally speaking, the author held a role in CDMS as 'processing czar' for Runs 118 and 119. This responsibility has been passed on to the newly-formed DAQ, Data Quality and Computing (DDC) group for subsequent data runs. The tasks performed from the pre-Run 118 to the post-Run 119 period were as follows.

Prior to the start of Run 118, the *DarkPipe* and *PipeCleaner* software required extensive upgrades for operation at Soudan, since the data acquisition hardware had changed from CDMS-I operation at the Stanford Underground Facility (SUF), and the software needed to function for up to 7 towers of detectors. This work was primarily done by the author. The farm scripts and online diagnostics were created and tested by the author and John-Paul Thompson. Automated diagnostics routines and the Run 119 blinding scheme (to be described) were implemented and tested by the author. Reprocessing of the Run 118 and Run 119 datasets ($\sim$ 4TB of raw data) was performed on the SUF computer cluster by the author, assisted by collaboration members at Stanford. Finally, much of the post-Run 119 SAC farm upgrade to 30 dual-CPU machines, including the farm script upgrades and necessary *DarkPipe*, *PipeCleaner* and diagnostics upgrades were handled by the author.

4.3 Data Acquisition System

Data Format

As described in Section 3.5, the occurrence of a global trigger as decided by the Trigger Logic Board results in information on the event causing the trigger being written to disk. The DAQ software writes a single event out via it's event structure (to be described below). The data contents of each event record are as follows:

SeriesNumber - SeriesNumber for the current data run, in the format 1ymmdd_hhmm. eg. for a dataset beginning on July 8th 2005, at 3:52pm, the SeriesNumber would be 150708_1552.

EventNumber - Events are written into files of 500 events per file. The Event number is equal to (file number $\times$ 10,000) + sequential number within data file. e.g. EventNumber=20056 is the 56^{th} event in file number 2, or the 556^{th} event overall.

EventTime - Time of event's global trigger in seconds since 12:00am Jan 1, 1970.

TimeBetween - Time since last event in ms.

LiveTime - Total livetime of the detectors that this event represents, in ms.

Trigger Info - A bitmask showing which detectors were above threshold and triggered for this event. Random and global trigger bits are also represented in this value (see below for details of random triggers).

Num Detectors - The number of operational ZIP detectors for this dataset.

Veto Times - A list of times for which any of the veto panels had a hit above threshold (time before global trigger in ms, taken from veto history buffer).

Veto Masks - A corresponding list of bitmasks for the veto panel hits listed as the 'Veto Times'.

Trigger Times - A list of times for which any of the detectors had a hit above threshold (time before global trigger in ms, taken from detector trigger history buffer).

Trigger Masks - A corresponding list of bitmasks for the detector hits listed as the 'Trigger Times'.

ZIP Address and Channel - DAQ board hardware addresses and channels used for each detector.

ZIP Phonon Traces - List of ADC values read out by the digitizers for each zip (2048 14-bit values read out at 1.25MHz). A trace is given for each detector (irrespective of which detector(s) caused the global trigger), for each of the four phonon channels.

ZIP Charge Traces - The same output as the ZIP Phonon Traces, except that traces are written for each detector for each of the two charge (ionization) channels.

ZIP Timing Info - Information on digitization timing (sampling rate, timing offsets, etc.)

Veto Address and Channel - DAQ board hardware addresses and channels used for the veto panels.

Veto Traces - List of ADC values read out by the digitizers for each of the 40 veto panels (1024 values read-out at 5MHz).

Veto Timing Info - Information on digitization timing (sampling rate, timing offsets, etc.)

Each piece of information is associated with a logical record header value. The data is stored as a binary file in a 'linked list' format. Figure 4.1 shows the data structure of a raw data file.

The file header contains values describing the current data format version and a flag dictating the endianness of the data (which end of a word the bits are read from, which differs on Macintosh and PC/Linux-based systems). Each unique event header value has it's own format for the subsequent data, and these are used to store the types of data detailed above.

Datafile Recording

A cluster of Linux-based PCs are used to combine the information from the various hardware boards and write the file to disk in the format described in the previous section.

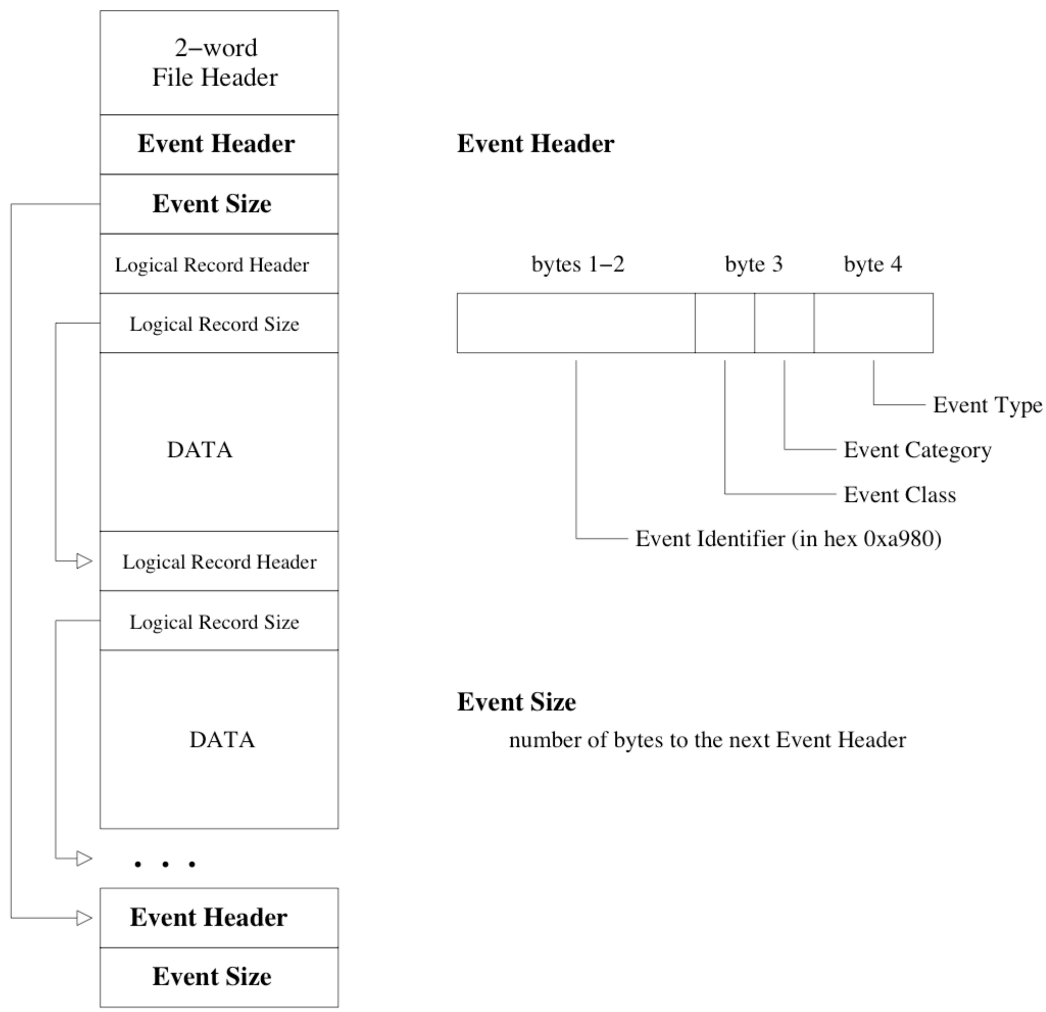


Figure 4.1: Data structure of a CDMS raw event file. Details of the header contents are described in the text.

The digitized detector signals from the Struck boards are sent to two machines, named ‘Tower 1’ and ‘’Tower 2’, via PCI-card interfaces (i.e. the signals from Tower 1 detectors go to the ‘Tower 1’ PC, etc.). Similarly, the digitized veto traces from the Joerger boards are sent to a machine named ‘Vetocrate’. A machine called ‘Builder’ combines the data from Tower 1, Tower 2 and VetoCrate, along with the information from the history buffer, to write the complete event files. Any other information needed to create the event file is obtained via the PCs that monitor and control the data-taking process (named, unsurprisingly, 'Monitor' and 'Control'). Builder, Vetocrate, Tower 1 and Tower 2 communicate via Gigabit ethernet connections.

A machine called 'DataServe' takes the datafiles and performs two operations: it backs up the files to a tape drive, and it copies them to an analysis farm computer in the surface building. Auxiliary files giving further information on the data series are also copied to the surface along with the data files. The connection to the surface was by 100 Base-T for Run 119, but has been upgraded to Gigabit ethernet for future operation.

The Control PC is responsible for running the event builder routines on Builder. It also issues all of the GPIB commands needed to bias the detectors, flash the LEDs used to bake the detectors, etc. The Monitor PC runs the Graphical User Interface used to start and stop data taking and flash LEDs. It also monitors information like trigger rates, and runs several diagnostics programs designed to give realtime feedback on the current data.

4.4 Soudan Analysis Cluster

The Soudan Analysis Cluster (SAC) consists of a set of dual-CPU Linux-based PCs that run the analysis software and perform other housekeeping tasks. The cluster has been upgraded to a total of 36 dual-CPU machines post-Run 119, but here we describe the setup of the cluster as it was for Run 118 and Run 119 - that is, 6 dual-CPU machines.

These 6 machines, named cdmsa-cdmsf, are each Pentium 1.25GHz systems with 1GB of memory. cdmsb, cdmsc, cdmsd and cdmse are primarily used to run the analysis software (*DarkPipe* and *PipeCleaner*). cdmsa and cdmsf perform various tasks required to operate the analysis farm, which are as follows:

cdmsa - Stores RQ and RRQ output files (see below), runs farm operation scripts, stores CVS code repository, runs the cluster's webserver.

cdmsf - Stores raw data files, backs data up to tape drive, runs the cluster's netServer (see below).

4.4.1 Farm Management and Data Processing

The primary operation of the analysis farm is to identify data files that need to be processed and spawn a corresponding Matlab job on an available farm machine. The farm also needs to keep track of which files have been processed successfully (and unsuccessfully). The handling of these processes is performed via a perl script named *SoudanPipe*.

Each data file that exists on disk has a corresponding status file, the contents of which are essentially a *DarkPipe* processing log that is also used to inform the analysis farm of action to take on the data file, if any. The farm script, SoudanPipe, constantly polls all of the status files present on the system, searching for the presence of particular keywords in the files. The final keyword found in a status file is taken to be the status of that data file. The status thus informs the farm as to what action to take upon the data file.

The most important status' are 'ready' and 'complete'. A status of 'ready' tells the farm to begin to process the corresponding data file via a spawned *DarkPipe* MATLAB job on an available machine. The 'complete' status does not require any immediate action on the part of the farm, but if every data file in a series has a status of 'complete', then the farm will start a *PipeCleaner* process if this has not been done already.

The pathway from raw data files to completed output is as follows:

1. Each raw data file contains 500 events, and is given a file name in the format 1ymmdd_hhmm_Fxxxx, where xxxx is the sequential number of the data file. For example, the first data file in a series beginning on March 25th, 2004 at 5:50pm would be named 140325_1750_F0001. The second file of this series would be named 140325_1750_F0002, etc.

2. Each raw data file is processed through *DarkPipe* and several output files are created containing RQ ('reduced quantity') values for each input data file. eg. processing 140325_1750_F0001 with *DarkPipe* would result in the creation of the following files:

   - 140325_1750_F0001_DPTG (RQs pertaining to detector triggering)
   - 140325_1750_F0001_DPVT (RQs containing veto quantities)
   - 140325_1750_F0001_DPZ1 (RQs containing ZIP 1 quantities)
   - 140325_1750_F0001_DPZ2 (RQs containing ZIP 2 quantities)
   - .... etc.

   So for the processing of each raw data file, there will be $(z + 2)$ RQ files created as *DarkPipe* output, where z is the total number of ZIP detectors operational. In general RQ values are not calibrated to be in the correct units for energy quantities; this is done in the next stage of processing. RQ files are also written to local machines, so that in the event of cdmsa being down, the farm nodes can continue to process data through *DarkPipe*. The SoudanPipe script is responsible for the aggregation of output RQ files onto cdmsa.

3. Once all of the raw data files have been processed in this manner, *PipeCleaner* combines all of the RQ files into what are generally termed RRQ ('relational reduced quantity') files. That is:

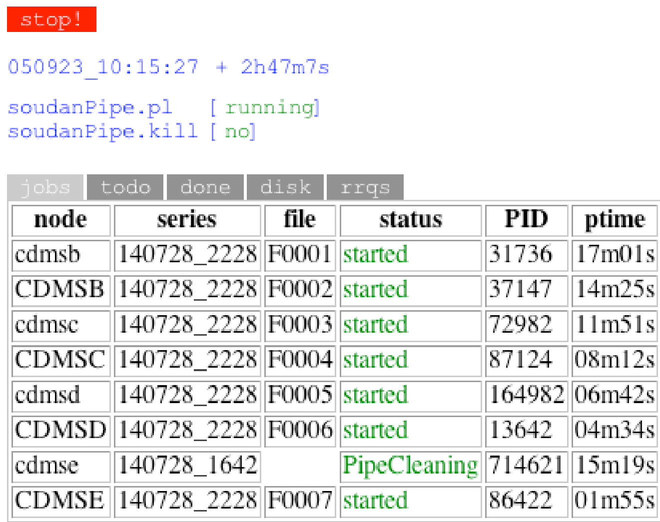
    - 140325_1750_F0001_DPTG
    - 140325_1750_F0002_DPTG
    - 140325_1750_F0003_DPTG
    - ...etc.

    are merged to create 140325_1750_DPTG and so forth. Further calibrated quantities are calculated from the RQ files and stored in a 1ymmdd_hhmm_DPR1 file. Phonon-position corrected quantities (see Section 4.7) are similarly held in a 1ymmdd_hhmm_DPM1 file. These $(z + 4)$ RRQ files represent the final output of the analysis chain.

4. The RRQ files for one or more data series' can then be loaded into a MATLAB program called *CAP* (CDMS Analysis Package). *CAP* allows for the analysis of very large datasets, and provides users with pre-defined cuts.

4.4.2 Online Monitoring Software

Since there can be, at times, more than 100 data files waiting to be processed by the analysis farm, it is extremely useful to have a way to monitor farm activity. To this end, the action of the *SoudanPipe* script is described, via a script named *FarmStatus*, on an XML-based web page.



node	series	file	status	PID	ptime
cdmsb	140728_2228	F0001	started	31736	17m01s
CDMSB	140728_2228	F0002	started	37147	14m25s
cdmsc	140728_2228	F0003	started	72982	11m51s
CDMSC	140728_2228	F0004	started	87124	08m12s
cdmsd	140728_2228	F0005	started	164982	06m42s
CDMSD	140728_2228	F0006	started	13642	04m34s
cdmse	140728_1642		PipeCleaning	714621	15m19s
CDMSE	140728_2228	F0007	started	86422	01m55s

Figure 4.2: The **jobs** tab of the farm monitoring page. This shows the current operation of the analysis farm, along with process IDs and the time since the process began.

Whenever *SoudanPipe* polls a status file, it writes the current status of that file to a log file. Similarly, whenever a *DarkPipe* processing job begins, completes, or results in an error, real-time information on these processes is written to a log file. This allows the *FarmStatus* script to describe, in close to real-time: 1) The current activity of the farm (which node is processing which file and for how long), 2) The list of data files currently waiting to be processed, 3) A log of completed *DarkPipe* and *PipeCleaner* processes, including error messages where appropriate.

This information is collated into a set of XML-generated web pages, shown in Figures 4.2 - 4.6.

The set of farm monitoring pages has proved massively useful for users from all institutions to keep track of processing at Soudan. There also exists a set of perl scripts on the SAC cluster which allow a user to change the status of a large number of data files with a stop!

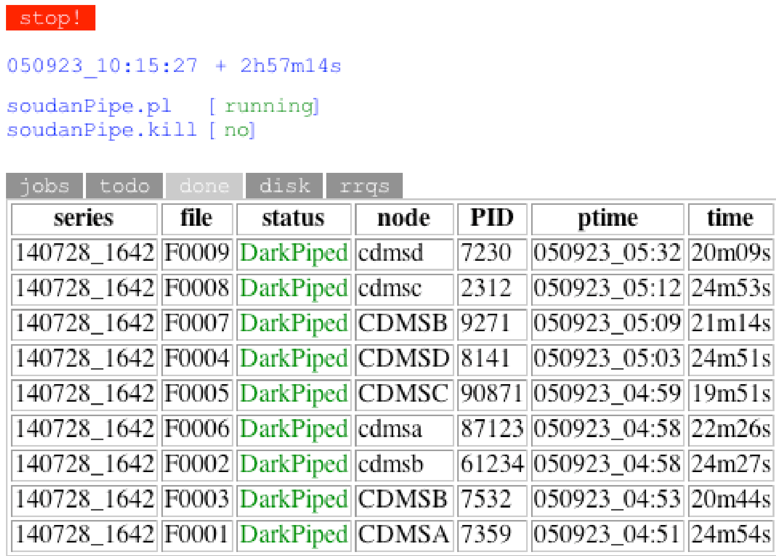
050923_10:15:27 + 2h50m12s

soudanPipe.pl [running]
soudanPipe.kill [no]

jobs todo done disk rrqs

series	file	status
140728_2228	F0008	ready
140728_2228	F0009	ready
140728_2228	F0010	ready
140728_2228	F0011	ready
140728_2228	F0012	ready
140728_2228	F0013	ready
140728_2228	F0014	ready
140728_2228	F0015	ready
140728_2228	F0016	ready
140728_2228	F0017	ready
140728_2228	F0018	ready
140619_1314	F0001	ready
140619_1314	F0002	ready

Figure 4.3: The **todo** tab of the farm monitoring page. This lists the data files which have a status of 'ready' or 'paused' (used to put processing on hold).



stop!

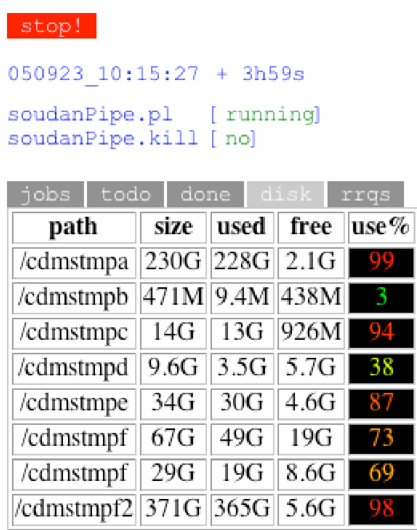
050923_10:15:27 + 2h57m14s

soudanPipe.pl [running]
soudanPipe.kill [no]

jobs | todo | done | disk | rrqs

series	file	status	node	PID	ptime	time
140728_1642	F0009	DarkPiped	cdmsd	7230	050923_05:32	20m09s
140728_1642	F0008	DarkPiped	cdmsc	2312	050923_05:12	24m53s
140728_1642	F0007	DarkPiped	CDMSB	9271	050923_05:09	21m14s
140728_1642	F0004	DarkPiped	CDMSD	8141	050923_05:03	24m51s
140728_1642	F0005	DarkPiped	CDMSC	90871	050923_04:59	19m51s
140728_1642	F0006	DarkPiped	cdmsa	87123	050923_04:58	22m26s
140728_1642	F0002	DarkPiped	cdmsb	61234	050923_04:58	24m27s
140728_1642	F0003	DarkPiped	CDMSB	7532	050923_04:53	20m44s
140728_1642	F0001	DarkPiped	CDMSA	7359	050923_04:51	24m54s

Figure 4.4: The **done** tab of the farm monitoring page. Displays recent activity of the farm, including failed processing jobs (due to errors).



stop!

050923_10:15:27 + 3h59s

soudanPipe.pl [running]
soudanPipe.kill [no]

jobs | todo | done | disk | rrqs

path	size	used	free	use%
/cdmstmpa	230G	228G	2.1G	99
/cdmstmpb	471M	9.4M	438M	3
/cdmstmpc	14G	13G	926M	94
/cdmstmpd	9.6G	3.5G	5.7G	38
/cdmstmpe	34G	30G	4.6G	87
/cdmstmpf	67G	49G	19G	73
/cdmstmpf	29G	19G	8.6G	69
/cdmstmpf2	371G	365G	5.6G	98

Figure 4.5: The **disk** tab of the farm monitoring page. This shows the available disk space for all drives on the farm machines.

stop!

050923_10:15:27 + 3h4m53s

soudanPipe.pl [running]
soudanPipe.kill [no]

jobs todo done disk rrqs

series	version	status	time
140808_1920	10.15	complete	050308_14:15:58
140808_0828	10.15	complete	050308_14:10:49
140807_2013	10.15	complete	050308_14:05:19
140807_0851	10.15	complete	050308_13:59:57
140806_1845	10.15	complete	040914_21:56:53
140803_1813	10.15	complete	041120_04:14:49
140803_1812	10.15	complete	041120_01:38:30
140803_1811	10.15	complete	041119_12:14:39
140803_1810	10.15	complete	041118_23:06:28
140803_1809	10.15	complete	041118_09:37:20
140803_1808	10.15	complete	041117_20:23:09
140803_1807	10.15	complete	050323_09:20:36
140803_1806	10.15	complete	050323_09:20:31
140803_1805	10.15	complete	050323_09:20:26
140803_0902	10.15	complete	040915_10:07:33
140803_0901	10.15	complete	040915_10:34:22

Figure 4.6: The **rrqs** tab of the farm monitoring page. This page lists the data series' that have successfully been through *PipeCleaner* and have therefore completed all processing. This list therefore shows data series' that are available to be used for analysis.

simple command.

4.5 Analysis Software

As already described, the *DarkPipe* software takes raw data and auxiliary files as input, and outputs a set of RQ files. The code is managed via Concurrent Versioning System (CVS), which allows for multiple users to write new code and 'check-in' changes to a central repository. CVS also provides the ability to 'tag' particular releases, which we have used as a way to fix our code for each round of processing. This is especially important and useful when using many machines to perform processing, since each machine must be using identical code.

An overview of the actual *DarkPipe* and *PipeCleaner* code can be found in Appendix A, but here we examine three key routines.

ReadEvents

The ReadEvents routine is responsible for reading the raw data and creating a Matlab structure which stores the same data. Since the raw data is located on a single machine, the farm needs to be able to read from a raw data file remotely, or the file must be copied across the network to be analyzed. Being able to read the data across the network has several advantages: 1) The raw data can be read in chunks (i.e. any number of specific event numbers can be read out), 2) The raw data can be compressed, saving disk space and still be read out piecemeal, 3) The alternative of copying the data file across the network would comprise a significantly greater network load.

The analysis farm is able to operate in this way by means of a network protocol called

NetServer. This provides the above functionality via the use of a small program that can be run on a single farm machine. Each netServer can handle requests from multiple machines; empirically this has a limit of around 10 nodes (more than that and the netServer can run into memory issues).

By means of the netServer, the ReadEvents routine is able to return the requested events by searching for the relevant event headers. In regular *DarkPipe* running, the analysis of a 500-event data file is performed in chunks of 50 events at a time. That is, ReadEvents is run ten times (and hence the netServer is accessed ten times), during the processing of a single raw data file.

Generation of Noise Templates

The first 500 events taken in a data series are not triggered by a global trigger, but are triggered at a random time. This allows the collection of events that represent the baseline noise of each detector channel for each data series. Around 1% of events during the rest of the data series are also randomly triggered to allow the study of changes in noise that might occur during the data series. The random trigger events are tagged in the history buffer by a specific bit in the mask.

The random triggers occurring at the start of the data series are used to generate noise templates. Each of the 500 traces is examined for each detector channel individually, and those that contain signal significantly above baseline noise are discarded from the sample. The remaining random trigger events are averaged and a power spectral density of the noise spectrum is stored.

Optimal Filtering

Optimal Filtering is a signal analysis method used to estimate the time-offset and amplitude of a signal by use of a pulse template. A real detector pulse can be written as

$$v(t) = As(t - t_0) + n(t) \tag{4.1}$$

where $s(t)$ is the expected pulse represented by a template with a pulse start time at $t = 0$ and normalized to peak height of 1 (here it is shifted to start with an arbitrary time-offset t_0), A is the amplitude to be estimated, and $n(t)$ is the noise present in the signal. Also, let $J(f)$ be the noise power spectral density of the signal. i.e.

$$J(f) = \lim_{T\to\infty} \frac{1}{2T} |\tilde{n}(f)|^2 \tag{4.2}$$

where $\tilde{n}(f)$ is the Fourier transform of $n(t)$.

The quality of the fit provided by the signal template and amplitude values is given by the χ^2 of the fit:

$$\chi^2 = \int_{-\infty}^{\infty} df \frac{|\tilde{v}(f) - Ae^{-i\omega t_0}\tilde{s}(f)|^2}{J(f)} \tag{4.3}$$

where $\tilde{s}(f)$ is the Fourier transform of $s(t)$.

We wish to determine the best-fit values of both A and t_0, which is done by minimizing the χ^2 with respect to A and t_0 respectively. It is useful to define the *optimal filter* by

$$\tilde{\phi}(f) = \frac{\tilde{s^*}(f)}{J(f)} \tag{4.4}$$

with Fourier transform $\phi(t)$. Minimizing χ^2 with respect to t_0 now yields

$$\frac{\partial}{\partial t_0} \left[\phi(f) * v(f)\right](t_0)\Bigg|_{t_0=\widehat{t_0}} = 0 \tag{4.5}$$

In other words to find the best estimator $\widehat{t_0}$ of the time offset t_0, apply the filter $\phi(t)$ to the trace and find the time of the peak. The best estimator for the amplitude is as follows:

$$\widehat{A} = \frac{\int_{-\infty}^{\infty} df\, e^{i\omega\widehat{t_0}} \tilde{\phi}(f)\tilde{v}(f)}{\int_{-\infty}^{\infty} df\, \tilde{\phi}(f)\tilde{s}(f)} \tag{4.6}$$

where $\widehat{t_0}$ is the best estimator of t_0 as described above.

Note that the numerator of $\widehat{A}$ is the convolution of the trace $v(t)$ with the optimal filter $\phi(t)$, evaluated at the time offset $\widehat{t_0}$. As already noted, $\widehat{t_0}$ corresponds to the time of the peak in the trace once the optimal filter has been applied, so $\widehat{A}$ is just proportional to the amplitude of that peak.

This method relies upon knowing the shape of the pulses that your detectors will see, along with a good estimate of the noise in the detectors. MATLAB code showing the optimal filter routine in practice can be found in Appendix B.

In Run 118-119, pulse templates for ionization signals were created by averaging over a set of normalized calibration ionization signals. This is sufficient since the shape of the ionization pulses do not change, and hence the shape of the pulse is determined by the electronics used to measure the signal. ‘Crosstalk’ templates are also created, and used to remove the small amount of signal present in the inner electrode channel due to the signal in the outer electrode channel, and vice versa.

Pulse templates for the phonon signals are created from a generalized 2-pole pulse shape, with predefined rise and fall times on a detector-by-detector, channel-by-channel basis. Recent work [84] has demonstrated the need to model the timing of the phonon pulse in more detail, providing greater phonon energy resolution and potentially greater precision in the calculation of timing parameters. It is likely that this will be the method used to calculate the amplitude for phonon pulses in the future.

4.6 Blinding Scheme

In order to perform our analysis in an unbiased way, it is necessary to develop a blinding scheme. Run 119 was the first CDMS run to be fully blinded - an overview of the scheme is as follows:

- Half of the calibration data taken during the run was designated as 'Open' (defined to be all data with an even SeriesNumber). This data was used to define cuts and recoil bands.

- The remaining half of the calibration data was designated 'Closed' (odd SeriesNumbers). This data was used to estimate the efficiency of cuts and expected backgrounds from the analysis.

- The WIMP search data had a blinding cut applied to it, such that any event that could potentially be identified as a WIMP was removed from the analysis. This allowed the WIMP search data to be studied without biasing ourselves with respect to any potential WIMP candidate events. Once all cuts have been frozen and the resulting efficiencies and backgrounds calculated, these blinded events are allowed into the analysis for the first time.

The blinding cut, applied to the WIMP search data, removed events fulfilling all of the following criteria:

- Anti-coincident with a muon veto event

- Event located in the inner ionization electrode volume
- Recoil energy between 5-130keV
- Event lies within 3σ nuclear recoil band (defined range of ionization yield values)
- Event deposited energy above threshold in one detector only (single scatter)

Any possible WIMP candidate event would have to fulfil all of these criteria.

Chapter 6 will describe in detail the results of calculating cut efficiencies from the open and closed calibration data, plus the results of 'unblinding' the WIMP search data. The fraction of WIMP search events culled by the blinding cut was $\sim 0.5 - 1\%$. The analysis was blind to this number prior to unblinding.

4.7 Data Reprocessing

There are several operations for which it is preferable to have all of the data run's calibration data available. Thus, once all data taking has ceased for a run, all of the data is reprocessed. These operations are such that the only reprocessing that is required is a re-running of *PipeCleaner* . Reprocessing is important for the following reasons:

- We would like to maximize the available statistics from the calibration data to define the energy calibration.
- It is preferable to have maximal calibration statistics to perform the phonon position correction algorithm (to be described below).
- The energy response of the detectors is non-linear, and this must be corrected (see below).

- Maximal calibration statistics is desirable for other calibration values, such as relative phonon quadrant weighting (see Chapter 6).
- Particular errors or data issues may be identified during the course of a run, and these can be corrected in the reprocessing.

For both Run 118 and Run 119, the reprocessing was performed on a cluster of computers located at the Stanford Underground Facility (SUF), which was the location of the CDMS-I experiment. This cluster also acted as the repository for the reprocessed data.

4.7.1 Phonon Position Correction

It is well-established that phonon timing parameters vary with position over the crystal [85]. Firstly, this is because T_c gradients exist across the crystal - that is, the four TES sensors on a single ZIP do not all have the same value of T_c. The size of the gradient can vary up to 20-30mK, but it is always present due to the intense difficulty in fabricating multiple TESs with the same T_c over a 76mm diameter wafer. This gradient causes variations in both the amplitude and risetime of the phonon signal. Secondly, the mean number of scatterings that a phonon will undergo before being absorbed by the phonon sensors increases with crystal radius[1]. This number increases rapidly close to the outer edge of the crystal and is responsible for a nonlinearity and subsequent degeneracy in the phonon position parameters.

Since the timing parameters of the phonon pulses are critical in rejecting surface events, it is imperative that this effect be corrected. This correction must be performed in such a way that the non-uniformities are removed, while the intrinsic depth information present in

[1]This follows from the crystal geometry and by considering that phonons are reflected from the crystal sides.

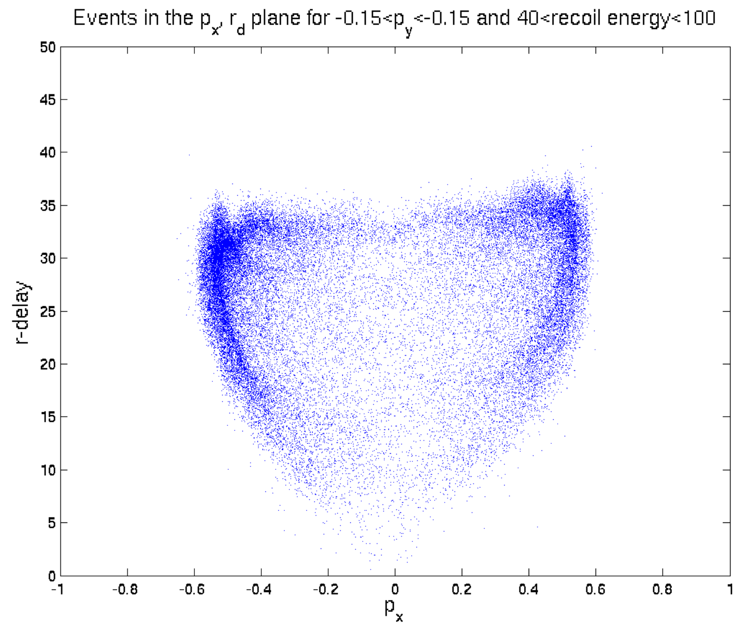


Figure 4.7: Phonon radial delay as a function of phonon x-partition for the selection of events shown in red in Figure 4.8. This shows the non-linearity in the event position reconstruction. All events are in ZIP 3 and are taken from ^{133}Ba calibration data.

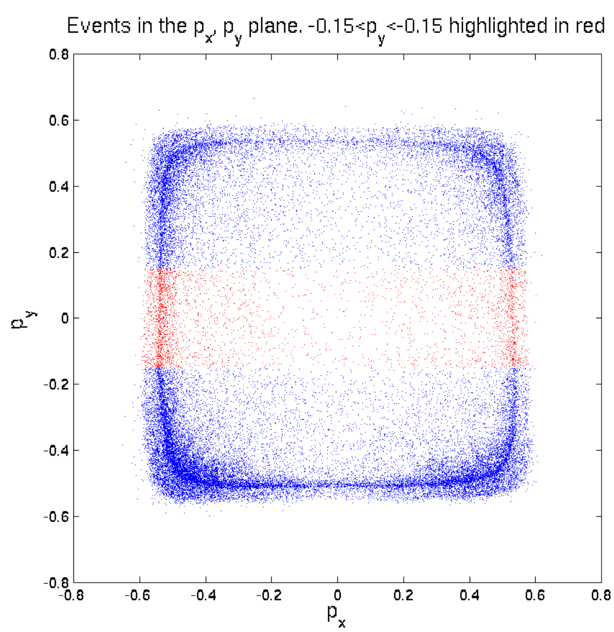


Figure 4.8: Phonon partition for ^{133}Ba calibration events in Z3. The highlighted horizontal slice of events in red is depicted in Figure 4.7.

the phonon signal remains unaffected. This is done by making the response of the detector to bulk electron recoil events independent of x-y position via a lookup table containing data on event parameters as a function of position.

Figure 4.7 shows the radial delay value (as calculated from x-delay and y-delay) as a function of the x-partition parameter. Both of these values are an estimate of position, but it can be seen that events occurring at large radii suffer a degeneracy in these parameters, irrespective of whether you use delay or partition to define the position. The position correction will require a unique way to describe the position of an event, and hence both phonon partition and phonon delay position estimations must be used.

The algorithm uses a lookup table, created from bulk electron recoil events taken from calibration data. An overview of the creation and application of the lookup table is as follows:

1. Create Lookup table using gamma calibration data:

    - Select events from calibration data that are bulk electron recoils. This selection is performed using the ionization yield parameter, which we can be sure will select only bulk recoils when the value is close to 1.
    - For each event, define it's 'position' as the location in the 3-dimensional space (p_x, p_y, d_r), where p_x and p_y are the x and y phonon partition, and d_r is the r-delay, defined as $\sqrt{d_x^2 + d_y^2}$.
    - Find N nearest neighbors in (p_x, p_y, d_r) space for the selected event. N is typically chosen to be ~ 80.
    - For each of the event parameters that are to be corrected, find the mean value of

the parameter for each of the nearest neighbor events. For example, each 'bin' in the lookup table will have a mean value of the phonon risetime for events in that bin.

- Making the detector response independent of position for the bulk electron recoil events implies that events in each bin should have the same mean value for each event parameter that is being corrected. Therefore a correction factor which will scale event parameters to the same mean is stored as a function of (p_x, p_y, d_r) for each parameter. This completes generation of the lookup table.

2. Apply Position Correction:

   - For every event in the new dataset being corrected (eg. WIMP search data), find the table entry corresponding to the closest point in (p_x, p_y, d_r) space.
   - Scale each parameter for the event according to the correction factor stored in the table derived from the gamma calibration for this position.

For more information on the phonon position correction algorithm, see [86].

4.8 Analysis Chain Upgrades

As the CDMS-II experiment is upgraded to a total of 30 ZIP detectors, improvements to the analysis farm are required. These upgrades include:

- An expanded Soudan Analysis Cluster. The SAC has been upgraded with 30 rack mounted, dual 1GHz Pentium machines donated from Fermilab. The machines that were previously used as the analysis farm now operate only the farm scripts and netServers for the new farm. Since the network throughput will be that much greater, cdmsf now operates two gigabit ethernet cards, each with it's own IP address. The farm scripts and monitoring software have been upgraded to handle the up-to-60 nodes that will be running. In particular, the farm now has a dynamic hostlist, since it is likely that nodes will be added and removed from the farm frequently.

- The farm's automated diagnostics routines have been better integrated with the DAQ's own diagnostics software.

- The increased computer power needed for reprocessing will be handled by a farm of several hundred machines at Fermilab. Even with the improved SAC cluster, running the phonon position correction algorithm on the amount of data the subsequent run is likely to generate will still take a long time. Obtaining a sufficient number of MATLAB licenses for this operation, however, would cost a great deal and a workaround has been found in the form of a compiled version of *DarkPipe* and *PipeCleaner*. This compiled analysis code, orchestrated by Walter Ogburn, can be run without a license on the Fermilab farm.

Chapter 5

Surface Event Analysis

As described in Chapter 2, the ionization yield $y = \frac{Q}{E_r} = \frac{Q}{P-\alpha Q}$ is used to discriminate between electron recoil and nuclear recoil events. Recall that for bulk electron recoil events, $y \sim 1$, and for nuclear recoils $y \sim \frac{1}{3}\,(\frac{1}{2})$ in Ge (Si). It has been found empirically that events that deposit energy within the outer $\sim 50\mu$m of the crystal surface suffer from a reduced ionization response [85]. Electron recoil events that deposit energy in the outer few μm have such a reduced ionization response that they appear in the nuclear recoil band.

Events in this 'dead layer' come from three sources: gammas interacting within the surface layer, low-energy electrons ejected from nearby material by high-energy x-rays ('ejectrons'), and electrons produced by radioactive beta decays from surface contamination. In dark matter search data it appears that the greatest contribution to the WIMP signal region is electrons (from beta emitters) that interact in a single detector. Electrons that scatter in more than one detector can be rejected since WIMP signal acceptance cuts require energy to be deposited in a single detector.

As described in Chapter 2, the dead-layer effect was reduced significantly by the addition of a thin layer ($\sim$40nm) of amorphous Si (α-Si) to the surface of the detector. The amorphous layer provides a barrier against charge diffusion into the incorrect electrode (i.e. electrons diffusing into the negative electrode, or holes into the positive electrode), since the band gap of the α-Si is greater than that of Ge or Si. This technique has reduced the physical size of the dead layer by a factor $\sim$3, but does not completely remove the effect.

The other way to combat the dead layer is to utilize the phonon signal risetime. It appears that the phonon population for an event close to a metallic layer (in this case the ionization electrodes) experiences a fast down-conversion of phonon energy to the range where propagation becomes dominated by ballistic propagation. This creates a measurably shorter risetime for the phonon signal arising from surface events [87].

5.1 Characterizing the Dead Layer

The dead layer effect is known to effectively reduce the ionization response for surface events, since free charges created in the particle interaction have a chance of diffusing, against the electric field, into the incorrect electrode. Empirically, it is also apparent that the phonon energy collection efficiency is suppressed for surface events.

The ideal calibration source to probe the suppression of the ionization and phonon responses would provide both bulk and surface events in the recoil energy range of 0-100 keV. Calibration of the Ge detector known as G31 (which eventually became T2Z5 in Run 119) was performed at UC Berkeley in 2002 using a ^{109}Cd source.

5.1.1 Radioactive Decay of ^{109}Cd

^{109}Cd decays via electron capture, which gives rise to an x-ray cascade. The resulting ^{109}Ag* excited state at 88 keV decays with a half-life of 40 s. The de-excitation of the 88 keV ^{109}Ag* state can follow several pathways.

Firstly, the excited nucleus can emit an 88 keV gamma - this occurs with a branching ratio of 3.6%. The remaining 96.4% of the time, one of the atomic electrons undergoes internal conversion. i.e. the nucleus charge distribution couples directly to the s-wave orbitals of the atomic electrons. This leads to a de-excitation of the nucleus by the emission of an atomic electron and the creation of an atomic shell vacancy. The resulting vacancy then fills with an electron from a higher energy orbital either by x-ray emission, or by an Auger process in which an additional atomic electron is emitted (this leaves another vacancy and the above process repeats).

Hence the primary processes by which the 88 keV excited state of ^{109}Ag* decays, are as follows [88]:

1) (41% decays) Internal Conversion electron - 63 keV K-shell ejection. The subsequent K-shell vacancy results in:

    a) (80%) ~22 keV photon[1], plus ~3 keV in photons and/or electrons.

    b) (~10%) ~25 keV photon.

    c) (~10%) Auger processes with total electron energies of 25 keV.

2) (45% decays) Internal Conversion electron - ~85 keV L-shell ejection. The subsequent

[1]Energies are approximate since there is a ~ 1 keV variation depending upon which sub-orbital is involved in the transition.

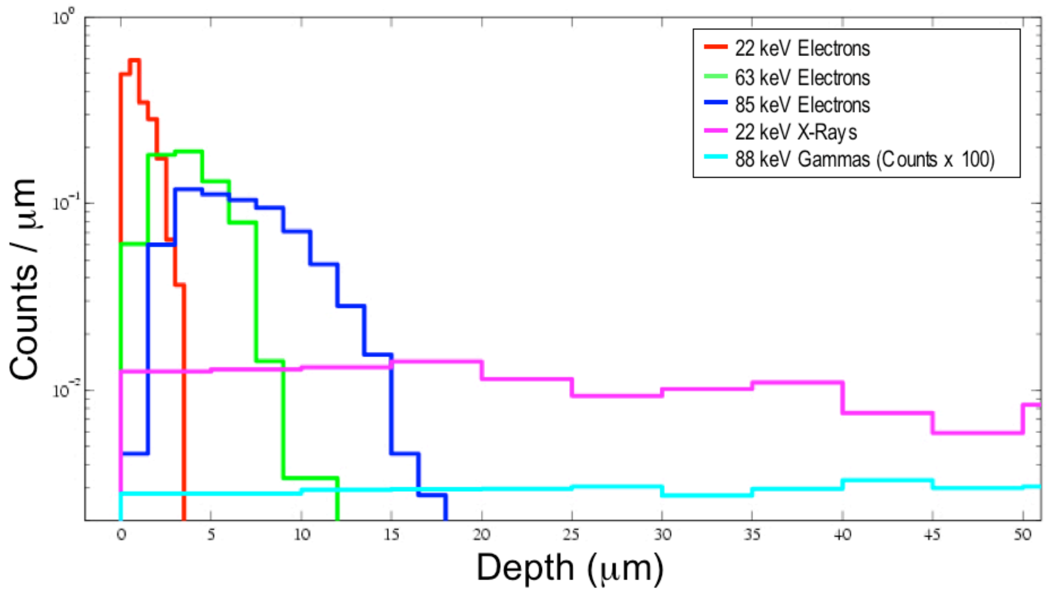


Figure 5.1: Depth profiles of the primary electron and gamma/x-ray lines from ^{109}Cd when incident upon Ge. The electrons from a ^{109}Cd source sample different depths of the detector's dead layer.

L-shell vacancy results in ~3 keV in photons and/or electrons.

3) (11% decays) Internal Conversion electron - ~88 keV (M/N/O)-shell ejection. The subsequent (M/N/O)-shell vacancy results in small amount of additional photons and/or electrons.

4) (3.6% decays) 88 keV photon.

A summary of all the photon and electron energies produced by the decay are shown in Tables 5.1 and 5.2 respectively.

Figure 5.1 shows the depth profiles of the main ^{109}Cd electron and gamma/x-ray lines when incident upon Ge, as obtained from a GEANT 4 Monte Carlo simulation.

Therefore a combination of events sampling different depths of the crystal, and with well-defined energies are emitted from the source. This provides an opportunity to quantitatively

Photon Energy (keV)	Branching Ratio	Notes
2.63	0.18	M→L transition
2.81	0.09	M→L transition
2.98	5.0	M→L transition
3.21	3.6	M→L transition
3.57	0.36	M→L transition
21.99	28.9	L→K transition
22.16	54.5	L→K transition
24.93	13.7	M→K transition
25.60	2.72	(N/O)→K transition
88.03	3.6	Nuclear de-excitation gamma

Table 5.1: Table of photon energies arising from the decay of ^{109}Cd, along with branching ratios and notes on the source of each line. Taken from [88].

Electron Energy (keV)	Branching Ratio	Notes
3	81	L-shell vac. filled by (M/N/O)-shell Auger
4	53	L-shell vac. filled by (M/N/O)-shell Auger
18	6	K-shell vacancy filled by L→L-shell Auger
19	8.3	K-shell vacancy filled by L→L-shell Auger
21	2.8	K-shell vacancy filled by L→M-shell Auger
22	2.9	K-shell vacancy filled by L→M-shell Auger
24	0.3	K-shell vacancy filled by M→M-shell Auger
63	40.8	K-shell internal conversion electron
84	3.2	L-shell internal conversion electron
85	41.5	L-shell internal conversion electron
87	9.1	M-shell internal conversion electron
88	1.7	(N/O)-shell internal conversion electrons

Table 5.2: Table of electron energies arising from the decay of ^{109}Cd, along with branching ratios and notes on the source of each line. Taken from [88].

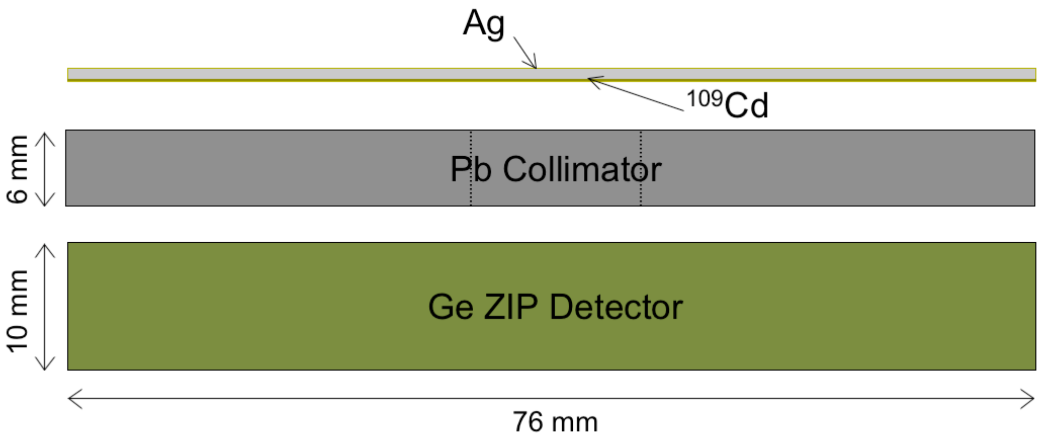


Figure 5.2: Schematic of the G31 Monte Carlo setup. Detector, collimator and source substrate are to scale. The hole in the Pb collimator is shown as a dotted line.

study the dead layer by comparing experimental data of the ^{109}Cd calibration with a Monte Carlo simulation of the calibration setup.

5.1.2 Analysis of G31 Calibration Data

This work focuses on two runs of the G31 calibration run at UC Berkeley; Run 296 which used the ^{109}Cd source to illuminate the 'charge side' of the detector (i.e. the side without the phonon sensors), and Run 297 which used the same source to illuminate the phonon sensor side of the detector. The source is a 1 mm thick, 76 mm diameter substrate with a thin layer of ^{109}Cd deposited upon it, covered in a 0.1 μm Au layer. The ^{109}Cd has silver backing. Both runs used a 6 mm thick lead collimator to illuminate a small region of the detector surface, and situated the source 18.1 mm above the detector (center to center measurement). See Figure 5.2 for a schematic of the setup.

The physical spectrum observed in the detector will differ from the ^{109}Cd source details listed in the previous section due to the Au layer, the Ag backing (which provides some backscattered events), the Pb collimator (which will Compton scatter some events onto the

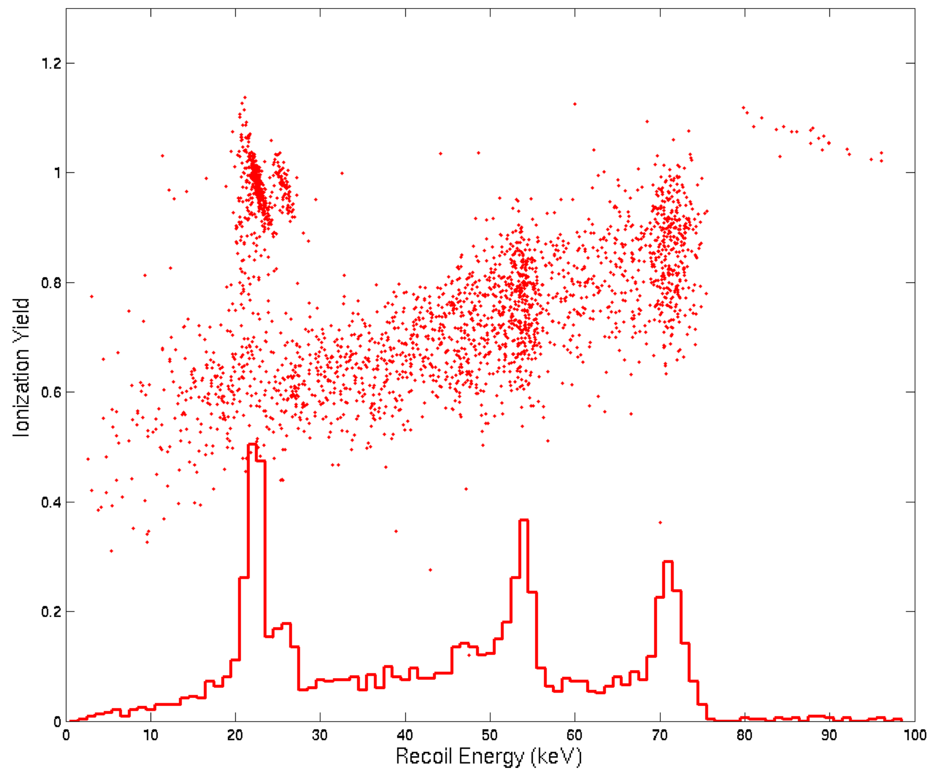


Figure 5.3: ^{109}Cd calibration data of the G31 detector, Run 296 (charge side). Overlayed histogram shows event counts on a linear scale.

detector), and the Ge detector itself, which will backscatter around 30% of the incident electrons in the energy range of interest.

Figure 5.3 shows the detector response for Run 296 (charge side). The electron energies of 63 keV and 85 keV actually appear in the detector with recoil energies at ~54 keV and ~71 keV, while the 22 keV x-ray peak appears at 22 keV recoil. The electron lines appear with ionization yield values centered below 1, while the 22 keV x-rays appear with values around 1. Both the ionization and phonon responses of the detector are suppressed for surface events. The data requires additional scaling to ensure that the 88 keV bulk gammas appear with an ionization yield of 1 and at 88 keV in recoil energy.

The suppression of the phonon response was also observed in the ^{241}Am calibration of the G32 detector at UC Berkeley. A positive correlation between the phonon energy and phonon risetime of calibration events is also present in both G31 and G32 calibration data. This is strong evidence that the reduction in phonon energy is a depth-dependent effect, since it is well-established that the phonon risetime is depth-dependent.

A GEANT4 Monte Carlo simulation mirroring the experimental setup was created, recording recoil energies and event depths for events interacting in the crystal. Where an event deposits energy in multiple locations, the depth of the event is taken to be an energy-weighted average of the sites where energy was deposited. The Monte Carlo code did not attempt to simulate the ^{109}Cd events that occur in coincidence - for example, a 22 keV photon always follows a 63 keV electron. The particle generator simply fired the particles listed in Tables 5.1 and 5.2 in the correct proportions.

The task is to find the ionization yield efficiency function $\epsilon_y(z)$, and the effective phonon efficiency function $\epsilon_P(z)$, where z is the energy weighted average event depth taken from the Monte Carlo output. We conjecture that these efficiency functions should be equal to 1 for $z \gtrsim 50\mu$m, and that they should decrease as z tends towards zero. For a given choice of $\epsilon_y(z)$ and $\epsilon_P(z)$, the analysis was performed in the following manner:

1) The effective ionization yield follows directly from the event depth. We know that all events in the calibration data are electron recoils defined to have an ionization yield of 1, but that the dead layer effect will reduce this yield for events depositing energy in the top 50 μm:

$$y = 1 \times \epsilon_y(z) \tag{5.1}$$

2) The phonon response is defined using the Monte Carlo energy-weighted average depth z, and recoil energy R:

$$P = \epsilon_P(z) \times (R + \alpha y R) \tag{5.2}$$

This follows from a) $R = P - \alpha Q$, and b) $y = Q/R$. Recall that α is the bias voltage divided by the energy required to create an electron-hole pair (see Chapter 2).

3) Noise can be added to y and P based on the detector's noise response.

4) Define the ionization energy by: $Q = yR$.

5) The final effective recoil energy is then given by: $E_r = P - \alpha Q$.

Hence, the task is to find the $\epsilon_y(z)$ and $\epsilon_P(z)$ that provide the best fit to the data (using y and E_r). We chose to model the efficiency functions using the functional forms:

$$\epsilon_y(z) = 1 - Ce^{-(Az+B)} \qquad \epsilon_P(z) = 1 - Ee^{-Dz} \tag{5.3}$$

where A, B, C, D and E are constants to be determined for each calibration run (we don't have an *a priori* reason for the efficiency functions to be the same on each side of the detector). Other functions were studied, including a direct deconvolution of the data to extract the depth dependence relation. The direct deconvolution was found to be too unstable with this method, and a generalized mapping gave results very similar to that of the functions above. The functions in (5.3) are representative of those that gave good fits to the data.

The phonon efficiency function is set to a constant value below a particular depth threshold, which was 15 μm for both Run 296 and Run 297 data. This improves the overall fit to the data but is also motivated by the likely source of the phonon efficiency - that is,

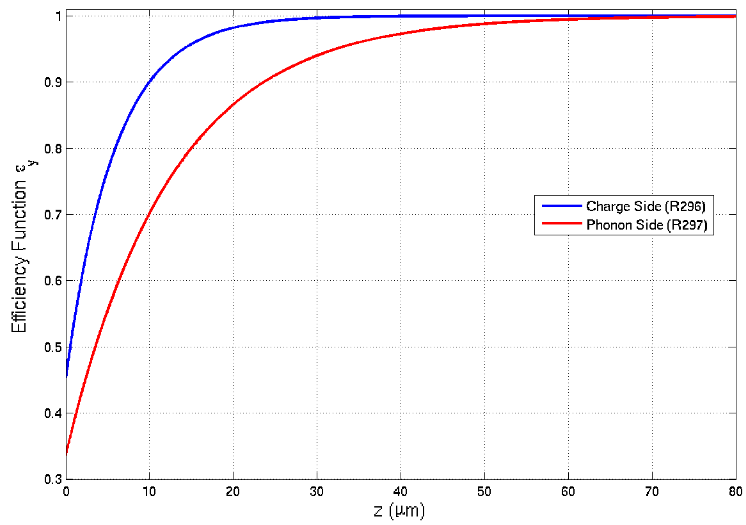


Figure 5.4: Ionization Yield efficiency functions $\epsilon_y(z)$, as derived from G31 data.

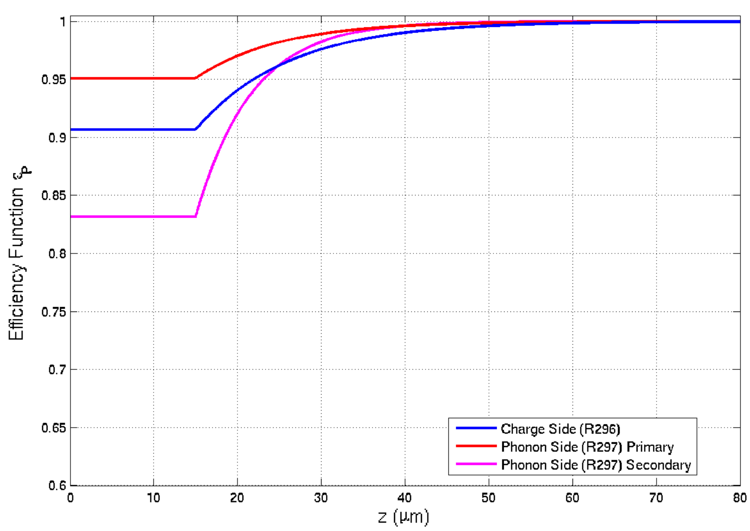


Figure 5.5: Phonon efficiency functions $\epsilon_P(z)$, as derived from G31 data. Run 297 (phonon side) data has two phonon efficiency functions; the primary function is applied to 70% of the Monte Carlo data, the secondary function is applied to the remaining 30% of the Monte Carlo data (see text).

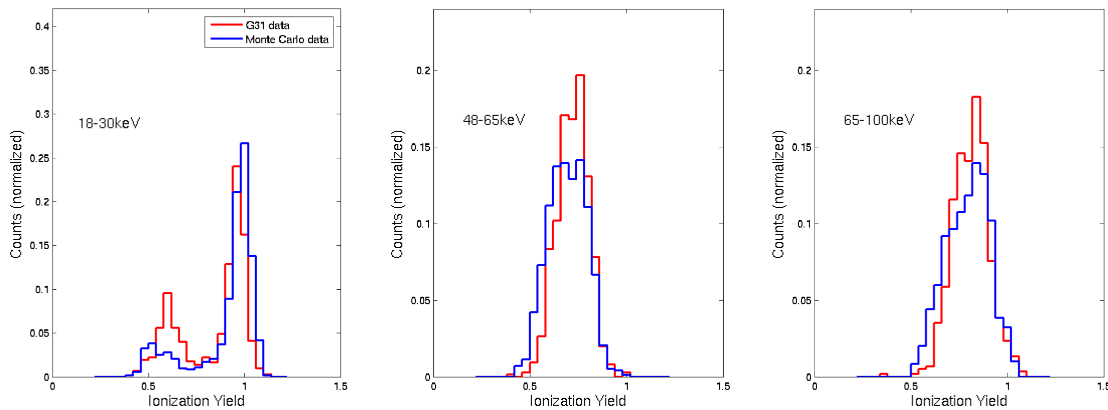


Figure 5.6: Fit of Monte Carlo data (blue) to G31 Run 296 (charge side) calibration data (red) around each of the primary peaks from the ^{109}Cd source. This fit allows the calculation of the ionization yield efficiency function, $\epsilon_y(z)$, independently of the phonon efficiency function. Each plot is normalized to the total number of events in the recoil energy window.

an interaction between the initial phonon population and the surface metallic films of the detector.

The observed 'shadowing effect' seen in the Run 297 data (see Figure 5.10) is also successfully modelled by this method by supplying two phonon efficiencies which are applied to two subsets of the data (see below).

The efficiency functions, as derived for both Run 296 and 297 calibration data, are shown in Figures 5.4 and 5.5.

G31 Run 296 (Charge Side) Results

The fit to the data is shown in Figures 5.6 to 5.8. The values used in Eqn. (5.3) to obtain these fits were $A = 0.17$, $B = 0.7$, $C = 1.1$, $D = 0.09$ and $E = 0.36$.

We have applied a non-linear correction to the recoil energy in G31 data in order to ensure that the 88 keV gamma events appear with a mean ionization yield of 1. The 88 keV events are bulk events, as shown in Figure 5.1, so the phonon energy was transformed

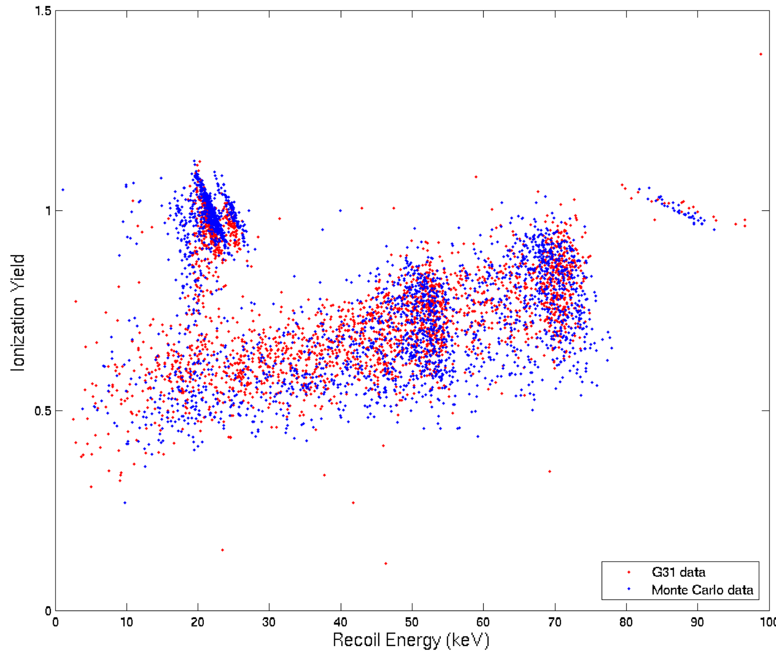


Figure 5.7: G31 Run 296 (charge side) calibration data (red), and Monte Carlo data with ionization yield and phonon efficiency functions applied (blue).

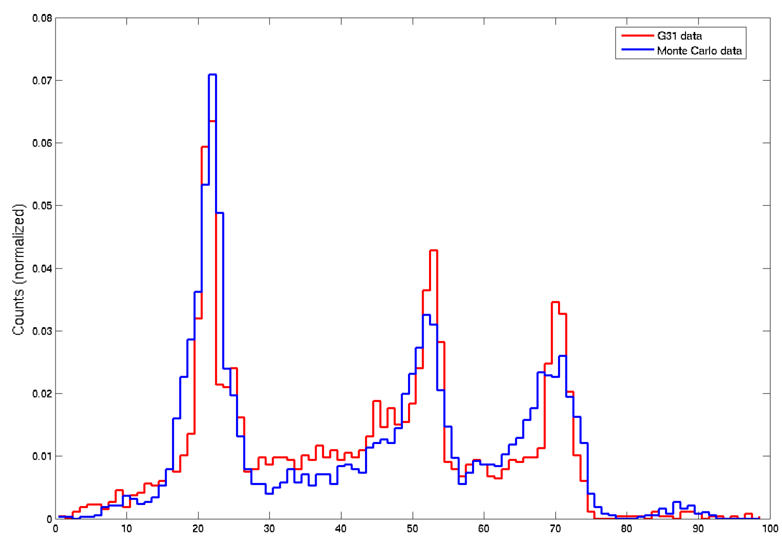


Figure 5.8: Counts by recoil energy (normalized to total number of counts in each data set) for G31 Run 296 (charge side) calibration data (red), and Monte Carlo data with ionization yield and phonon efficiency functions applied (blue).

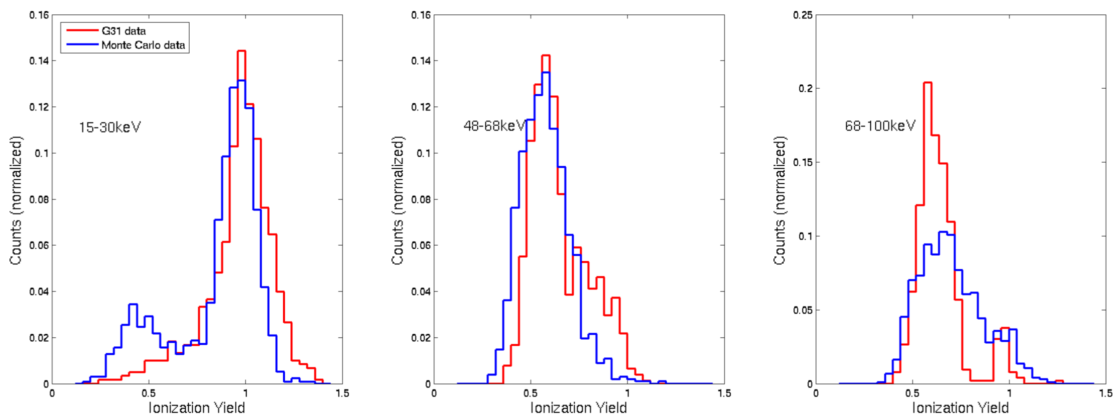


Figure 5.9: Fit of Monte Carlo data (blue) to G31 Run 297 (phonon side) calibration data (red) around each of the primary peaks from the ^{109}Cd source. This fit allows the calculation of the ionization yield efficiency function, $\epsilon_y(z)$, independently of the phonon efficiency function. Each plot is normalized to the total number of events in the recoil energy window.

by the relation $P_{\mathrm{new}} = \tau(e^{P/\tau} - 1)$ with $\tau = 800$ in order to place the 88 keV bulk recoils at an ionization yield of 1. The recoil energy was then scaled linearly to put the 88 keV bulk gammas back at a recoil energy of 88 keV.

A depth of 15 μm was used as a threshold for the phonon efficiency function - at depths below 15 μm the efficiency becomes constant. This is necessary to get a reasonable fit to the electron events while still getting agreement on the 22 keV x-ray distribution in Figure 5.7. The 15 μm depth may indicate the size of the initial phonon 'fireball' that interacts with the surface metallic films.

G31 Run 297 (Phonon Side) Results

The MC fit to the data is shown in Figures 5.9 to 5.11. Data taken in G31 Run 297 has a 'shadowing' effect, where a subset of the data appears at a higher ionization yield and at a lower recoil energy (see Figure 5.10). The shadowing effect is modelled by using two

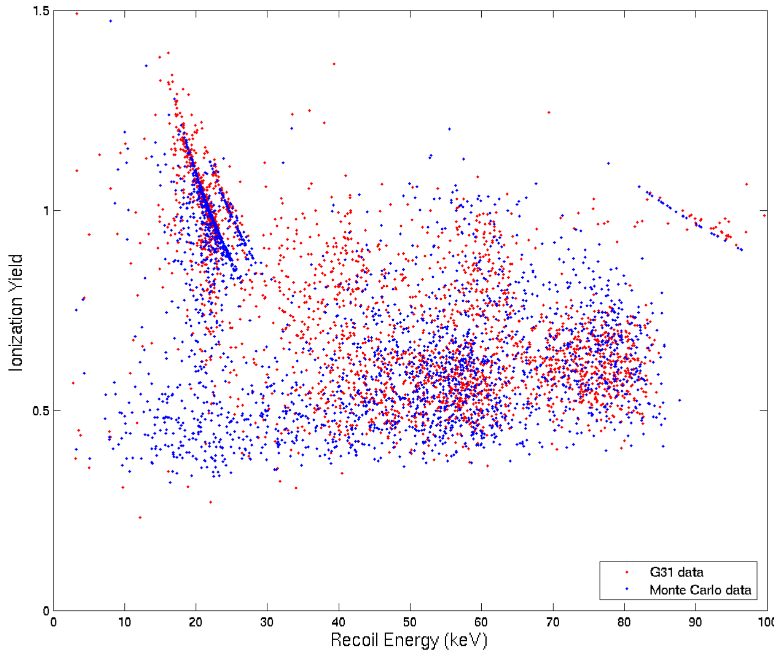


Figure 5.10: G31 Run 297 (phonon side) calibration data (red), and Monte Carlo data with ionization yield and phonon efficiency functions applied (blue).

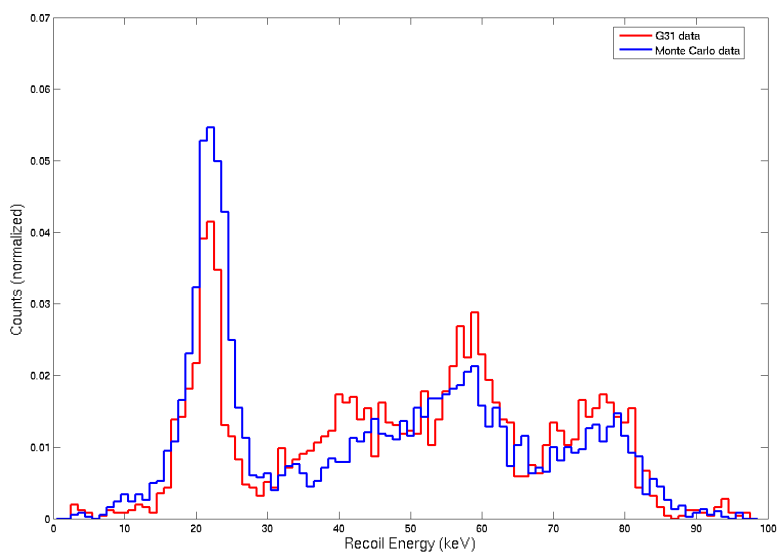


Figure 5.11: Counts by recoil energy (normalized to total number of counts in each data set) for G31 Run 297 (phonon side) calibration data (red), and Monte Carlo data with ionization yield and phonon efficiency functions applied (blue).

different phonon efficiency functions. Both used a threshold of 15 μm, below which the efficiency is constant, as shown in Figure 5.5. Monte Carlo data was randomly selected so that 70% of the events utilized the primary phonon efficiency function, and 30% used the secondary function.

Despite the shadowing effect, the G31 Runs 297 data passes data quality checks which would rule out detector problems. The phonon pulses for the shadow events have typical χ^2 values, and are not unusual aside from having a lower amplitude than expected. The shadowing effect has not been observed in subsequent ^{241}Am and ^{60}Co calibrations on this and other detectors, but these sources do not supply events that heavily sample the surface as ^{109}Cd does. It is possible that the shadowing effect is due to real detector physics - the fraction of events that suffer from this effect ($\sim$30%) is very close to the fraction of the ZIP surface that is covered by the Al fins (25%).

The values used in (5.3) to obtain the fit for the yield efficiency function were $A = 0.08$, $B = 0.55$ and $C = 1.15$. The primary phonon efficiency function uses $D = 0.10$ and $E = 0.22$ - the secondary phonon efficiency function uses $D = 0.15$ and $E = 1.60$.

5.2 Study of Beta Backgrounds

The dead layer model developed in the previous section provides a good fit to the G31 ^{109}Cd calibration data. Next we apply the analysis to simulation of ^{133}Ba calibration data.

^{133}Ba provides gamma ray lines at 303 keV, 356 keV and 384 keV, and is used to perform the energy calibration in CDMS. It also provides a population of surface events comprised of photons interacting at the surface, and ejectrons. Recall that an ejectron is a low-energy

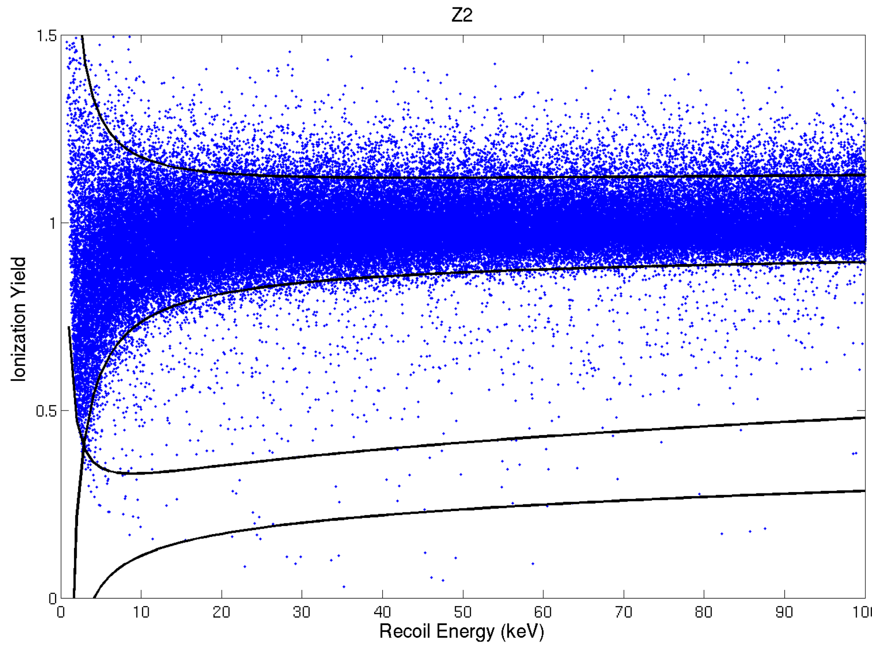


Figure 5.12: ^{133}Ba calibration data for detector Z2. The 2σ electron recoil (upper) and nuclear recoil (lower) bands are shown. Charge threshold and inner electrode cuts have been applied.

electron ejected from nearby material by an x-ray photon.

Figure 5.12 shows ^{133}Ba calibration data taken using the Ge ZIP Z2. The 2σ electron recoil and nuclear recoil bands are shown, as calculated from (2.10) and (2.11). Details of the creation of these bands can be found in Chapter 6.

The vast majority of events in Figure 5.12 appear in the electron recoil band, but that there is a population of events that 'rain down' from the electron recoil band. These are surface events that have a reduced ionization yield as a result of the dead layer effect. Generally, electron recoils appearing below the electron recoil band are referred to as 'betas', although the population can actually include betas from radioactive contaminants, ejectrons, and photons interacting at the surface. Similarly, events appearing in the 2σ electron

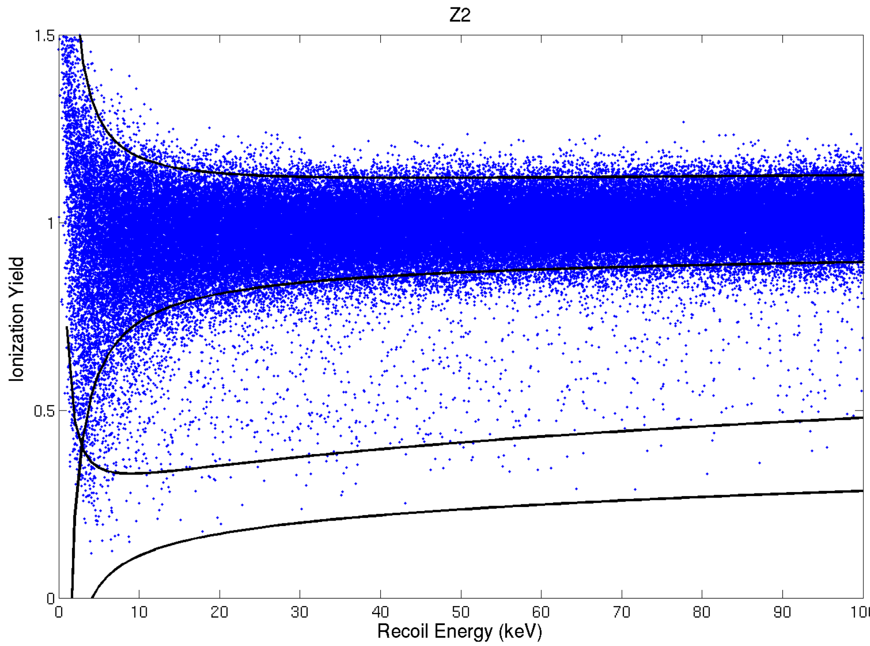


Figure 5.13: Simulation of ^{133}Ba calibration data for detector Z2. The 2σ electron recoil (upper) and nuclear recoil (lower) bands are shown. Charge threshold and inner electrode cuts have been applied.

recoil band are referred to as 'gammas' although technically this includes some x-rays.

5.2.1 Simulation of ^{133}Ba Calibration Data

A GEANT4 Monte Carlo simulation of one tower of 6 ZIP detectors was created, modeling the calibration that took place during Run 118 at Soudan. In the simulation, the stack of 6 ZIPs are surrounded by 2 mm of copper to model the detector housing. 2.5 million events from a gamma source modeled after ^{133}Ba were fired from a sphere surrounding the entire geometry, and the recoil energy and depth of each event incident upon a detector was recorded. The depth is again taken as an energy-weighted average of the depths where energy is deposited.

Events with a depth greater than $50\,\mu$m are considered to be bulk events, since we assume that beyond this depth both the yield and phonon efficiency functions are equal to one. The efficiency functions were applied using the procedure described in Section 5.1.2 to obtain the effective ionization and phonon energies. Since we will be coadding the statistics, it is assumed that an event is equally likely to be on either the charge or phonon side of the detector and is randomly determined. The efficiency functions for that side are then used.

Figure 5.13 shows the simulated data for detector Z2 using the same cuts as was used to produce Figure 5.12. The simulated dataset contains approximately twice the number of events as the calibration dataset.

5.2.2 Comparison of Surface Event Rates

We would like to examine the depth dependence of particular types of events, and compare their rates with those observed in the experimental data. The following types of events are of interest:

- **Single** scatter events. These deposit energy above threshold ($\sim 1\,\mathrm{keV}$ in ionization energy) in one detector only. Single scatters that deposit energy in the top $1\,\mu\mathrm{m}$ will contribute to the WIMP signal region due to a low yield efficiency value that puts the event in the nuclear recoil band.

- Nearest-Neighbor Double (**NND**) events. These are events for which two neighboring detectors both have simultaneous events above threshold.

- Beta-Beta Double (**BBD**) events. A subset of the nearest-neighbor double events for which both events are defined as betas.

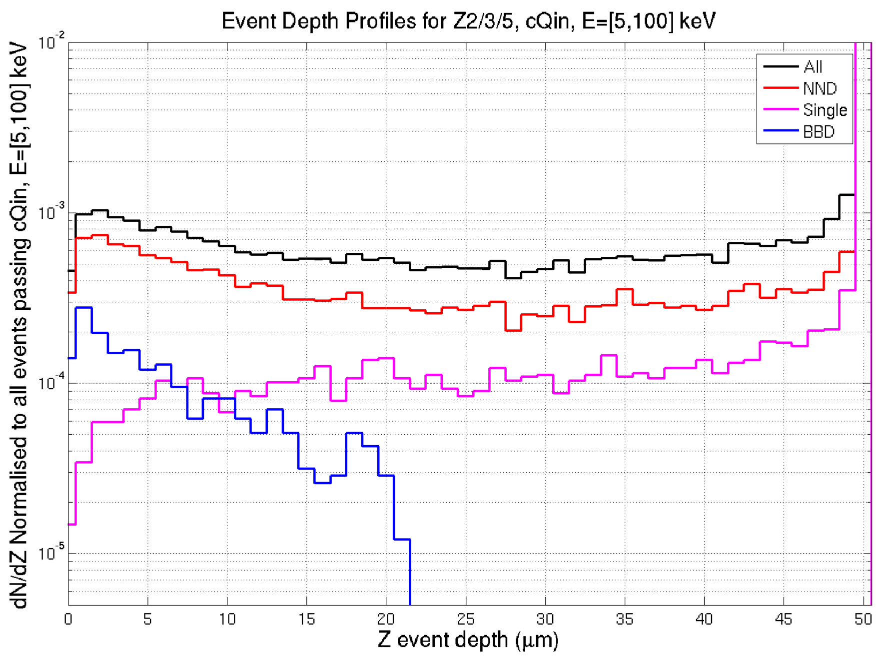


Figure 5.14: Depth profiles of 5-100 keV events from the Monte Carlo simulation of ^{133}Ba calibration data for detectors Z2/Z3/Z5 combined. See text for description of event types.

The beta definition in this case follows by finding a suitable depth cutoff which matches the observed ratio of gammas to betas found in the experimental data. A beta is hence defined for the simulation data as an event with a depth of $< 21\,\mu$m, which is consistent with the value expected from the yield efficiency function.

Monte Carlo events in these categories are shown as a function of depth in Figure 5.14 for the Ge ZIPs Z2, Z3 and Z5 combined. Events are preselected to have a recoil energy in the range 5-100 keV, and pass the inner electrode (cQin) cut. The count in the figure is normalized to the total number of events passing these cuts. The $50\,\mu$m bin contains all events with a depth $> 50\,\mu$m.

As would be expected, nearest-neighbor double and beta-beta double events predominantly sample the ZIP surface. Single scatter events show a corresponding suppression of

Event Type	Monte Carlo Simulation	Run 118 Calibration Data
β all	$1.44\%\,^{+0.04\%}_{-0.03\%}$	$1.45\%\,^{+0.06\%}_{-0.05\%}$
β singles	$0.19\%\,^{+0.01\%}_{-0.02\%}$	$0.13\%\,^{+0.02\%}_{-0.02\%}$
2σ Nucl. Recoil Band all	$0.09\%\,^{+0.01\%}_{-0.01\%}$	$0.11\%\,^{+0.01\%}_{-0.01\%}$
2σ Nucl. Recoil Band singles	$0.002\%\,^{+0.002\%}_{-0.001\%}$	$0.005\%\,^{+0.004\%}_{-0.003\%}$

Table 5.3: Percentage of events by type from simulation and calibration data [89] of ^{133}Ba source during Run 118 for detectors Z2, Z3 and Z5 combined. Events are preselected to have recoil energy in the range 5-100 keV, be above the ionization threshold, and pass the inner electrode cut. Total event numbers passing these cuts in each dataset are 361,478 (Monte Carlo) and 192,632 (Run 118 data). 1σ error values shown.

events very close to the surface.

Table 5.3 shows the percent of ^{133}Ba calibration events types observed in the simulated and the experimental data. A event in the simulation data is taken to appear in the nuclear recoil band when it has an average depth $< 1\,\mu$m. A beta is defined in the simulation data as an event which has an average depth $< 21\,\mu$m.

5.2.3 Summary

The simulated data is fit to both the total number of betas and the total number of nuclear recoil band events seen in the Run 118 calibration data (first and third rows in Table 5.3). Having set these two free parameters, the number of beta singles and nuclear recoil band singles predicted by the simulation match with the calibration data numbers within statistics at the 1.5σ level. The fraction of beta events that are singles (0.13) and the fraction of nuclear recoil band events that are singles (0.02) differ by a factor of 6 yet the simulation is able to match both these numbers within statistics.

The development of the dead layer model on G31 calibration data and it's subsequent application to ^{133}Ba data appears to vindicate the use of this model in conjuction with a

GEANT4 Monte Carlo simulation. Analyses of local beta contamination sources in CDMS, such as ^{210}Pb and ^{40}K, using this dead layer model are being performed by a group at the University of Florida. The effectiveness of the model when applied to the ^{241}Am and ^{60}Co calibrations of the G32 detector should also be studied.

Chapter 6

Run 119 Analysis

This chapter describes the results of the second CDMS-II data run taken at the Soudan Underground Laboratory. The data run, referred to as Run 119, was in operation from March - August 2004, and used 2 Towers of detectors, for a total of 12 ZIPs (6 Ge, 6 Si), and a total mass of 1.5kg Ge, 0.6kg Si.

6.1 Run Overview

A total of 77.4 live-days of WIMP search data was taken over the course of Run 119, giving a total of 116.1 kg days exposure on Ge and 46.44 kg days on Si. The 2-Tower setup is shown in Figure 6.1, and the detectors were operated throughout the run at a bias voltage of 3V for the Ge ZIPs, 4V for the Si ZIPs. Tower 1 had already been in operation for several months at Soudan in Run 118 [90], and before that was operated in CDMS-I at the Stanford Underground Facility. Hence, Tower 1 contains detectors that were already well understood at the start of Run 119, but the Tower 2 detectors, although characterized at

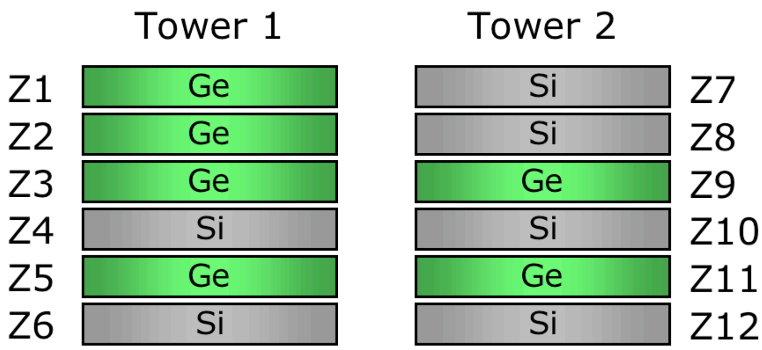


Figure 6.1: 2-Tower setup for Run 119. Tower 1 contains the detectors operated during Run 118 at Soudan.

the test facilities prior to installation, had not been in operation at a deep site before.

A few minor problems with the cryogenics system were present throughout Run 118 and Run 119 at Soudan. The main problem stemmed from a suspected Helium bath to OVC leak which led to a fridge 'burp' once every 3-4 weeks. Each of these incidents expelled some of the ^{3}He/^{4}He mixture and warmed the detectors to $\sim$1K, making data-taking impossible. It took several hours after a burp to stabilize the cryogenic system and resume data acquisition.

During regular running, routine cryogenic operations consisted of two transfers to both the liquid Helium and liquid Nitrogen bath per day. Usually, we would expect transfers to be necessary every $\sim$18 hours, instead of every 12, but the LHe to OVC leak increased the loss rate of the He and N baths. During the transfers, and every 4 hours otherwise, the detectors were flashed with the LEDs (described in Chapter 2). The several weeks prior to the start of the run consisted of many long LED baking sessions until the detectors were sufficiently neutralized.

Almost two hundred calibration datasets were taken spread over the course of Run 119. These were taken with one of the following sources:

- A 2 μCi ^{133}Ba source. The rate of this gamma source is low enough as to not create pileup events in the detectors, yet high enough that the gamma lines at 303 keV, 356 keV and 384 keV are prominent in the detectors. These are the primary lines used to perform the energy calibration of the ZIPs. Over one hundred ^{133}Ba calibrations were performed spread throughout the entire run.

- A 1 mCi ^{252}Cf source. This source provides neutrons over the 0-100 keV recoil energy range that we are interested in. It is primarily used to define the nuclear recoil band, and hence the signal region (where we would expect to see WIMPs). Four neutron calibrations were performed during Run 119. During each one thermal neutron activation will occur, and consequently we see elevated gamma and beta background rates immediately afterwards.

The ^{133}Ba also acts as a source of calibration surface events. These occur as a result of photons interacting in the surface of the detector, or (more likely) by kicking out an electron from the surface of a detector which then interacts in a neighboring detector (an 'ejectron').

Figure 6.2 shows the cumulative livetime of WIMP search data taken over the course of Run 119. A total of 7.57M events (before cuts) were taken in ^{133}Ba calibration data (average of 631k events per detector), with 214k events taken during ^{252}Cf calibrations. The ^{133}Ba calibrations were fairly evenly distributed over the course of the run, with more frequent and longer calibration runs being performed towards the end of the run (effect of ^{133}Ba calibrations on WIMP search livetime can be seen in Figure 6.2).

Table 6.2 shows the total number of calibration events passing successive analysis cuts

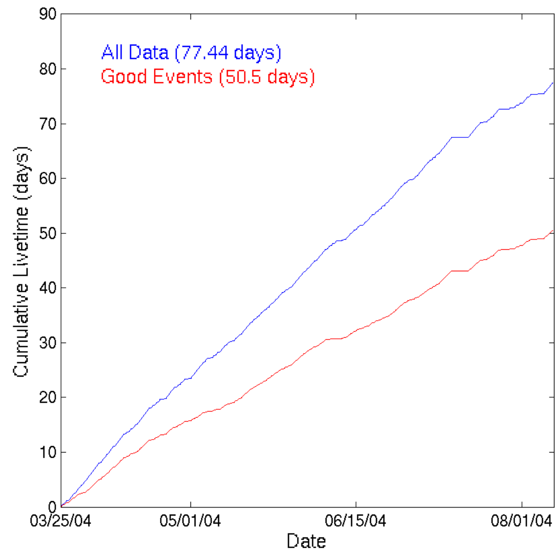


Figure 6.2: Cumulative livetime of WIMP search data over the course of Run 119. All the WIMP search data is shown in blue, and only those events that pass the good event cut (to be described) are shown in red.

(the details of which will be described in this chapter).

6.2 Involvement

Since this thesis details work done by many members of the CDMS collaboration, it is necessary to note which work was done by the author. A similar section is found in Chapter 2 detailing the author's role in the analysis chain; here we note the work done on Soudan Run 119 analysis.

All the figures in this chapter (except where noted otherwise) were generated by the author. The overall analysis is similar, but independent to the primary analyses performed by other members of the CDMS collaboration. Primary responsibility for the individual parts of the analysis, as it was first performed, was distributed throughout the collaboration

Preselection Cuts	Number of events in Gamma (^{133}Ba) calibration
Above ionization threshold in specified detector	1.26M/1.04M/967k/856k/770k (Ge) - 469k/472k/345k/315k (Si)
+ Good events, inner electrode, 7-100 keV recoil energy	172k/141k/139k/142k/106k (Ge) - 170k/200k/144k/142k (Si)
+ Anticoicident with veto	166k/136k/134k/137k/103k (Ge) - 164k/192k/139k/136k (Si)
+ Phonon timing cut	35k/15k/4k/15k/20k (Ge) - 2k/70k/2k/9k (Si)
+ Single scatter	8838/3521/2112/5205/9112 (Ge) - 369/25240/569/5132 (Si)
+ 2σ nuclear recoil	0/0/0/0/3 (Ge) - 1/0/0/0 (Si)

Table 6.1: Event numbers passing successive analysis cuts in the Soudan Run 119 gamma calibration (^{133}Ba) data, Open and Closed data coadded, for detectors Z2/3/5/9/11(Ge) and Z4/6/8/10 (Si). Data taken from March-August 2004. Cuts listed on each row are cumulative, i.e. the numbers given are for events passing all cuts up to and including that row.

Preselection Cuts	Number of events in Neutron (^{252}Cf) calibration
Above ionization threshold in specified detector	23k/22k/21k/23k/21k (Ge) - 13k/15k/13k/12k (Si)
+ Good events, inner electrode, 7-100 keV recoil energy	3549/2660/2853/4129/3296 (Ge) - 3781/5002/4227/3968 (Si)
+ Anticoicident with veto	3406/2569/2723/3974/3162 (Ge) - 3642/4797/4074/3821 (Si)
+ Phonon timing cut	1185/696/603/975/1147 (Ge) - 493/2593/694/1125 (Si)
+ Single scatter	450/246/258/393/480 (Ge) - 235/1035/315/597 (Si)
+ 2σ nuclear recoil	329/171/208/311/285 (Ge) - 231/677/296/499 (Si)

Table 6.2: Event numbers passing successive analysis cuts in the Soudan Run 119 neutron calibration (^{252}Cf) data, Open and Closed data coadded, for detectors Z2/3/5/9/11(Ge) and Z4/6/8/10 (Si). Data taken May-August 2004 and contains both gamma and neutron events. Cuts listed on each row are cumulative, i.e. the numbers given are for events passing all cuts up to and including that row.

on a voluntary basis. The author was responsible for creating the prepulse baseline cut and the recoil band cut definitions (to be described).

6.3 Detector Performance

Noise Performance

Figure 6.3 shows typical power spectral density distributions for detector noise. This distribution was created from a sample of random trigger events taken at the start of a data series. Events that contain coincident events are removed from the sample. The phonon noise falls as $1/f$ at high frequencies. The charge noise is close to expected due to the associated electronics, although there are several high frequency noise sources most of which are known to arise from connections to the fridge.

The total noise response of the detectors was also monitored by analysis of the 'noise blobs'. These are the calculated (via optimal-filtering, fixed start time) pulse amplitudes for random trigger events. We should expect these distributions to be approximately gaussian in shape. Noise blobs for Tower 1 and Tower 2 detectors are shown in Figures 6.4 and 6.5 respectively. The phonon resolution of Z1 is poor due to the TES having a variation across the surface. Gaussian fits to the noise blobs provide the appropriate values to use for trigger thresholds and to make single/multiple event cuts. Note that the noise performance of the ionization channels is somewhat worse than for the phonon channels, and the level of this noise will be the limiting factor in placing our analysis threshold.

The trend shown in Figure 6.6 demonstrates how the detector noise fluctuates over the course of the run. The values shown are the pre-pulse baseline mean and standard deviation,

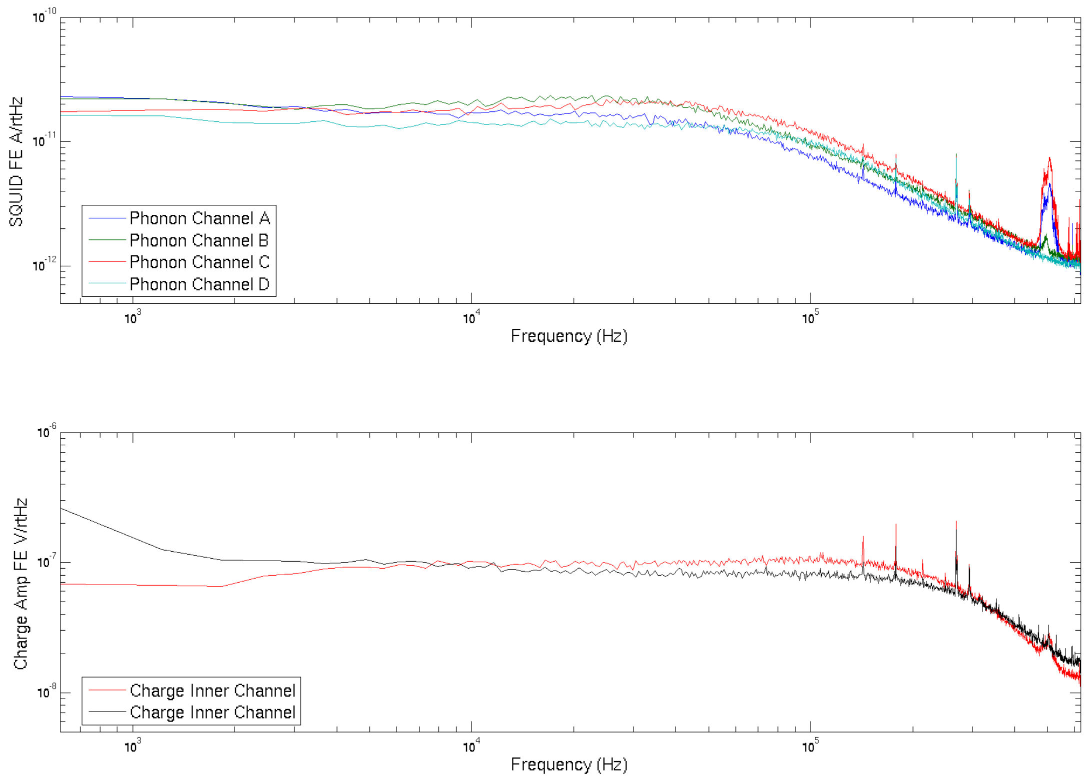


Figure 6.3: Noise power spectral densities for detector Z3, taken at the end of Run 119. The noise PSD shown here is fairly typical for the Run 119 detectors over the course of the run. The upper plot shows the noise associated with the SQUID input coil on each of the four phonon channels. The lower plot shows the inner and outer ionization electrode channel noise, associated with the noise on the feedback capacitor.

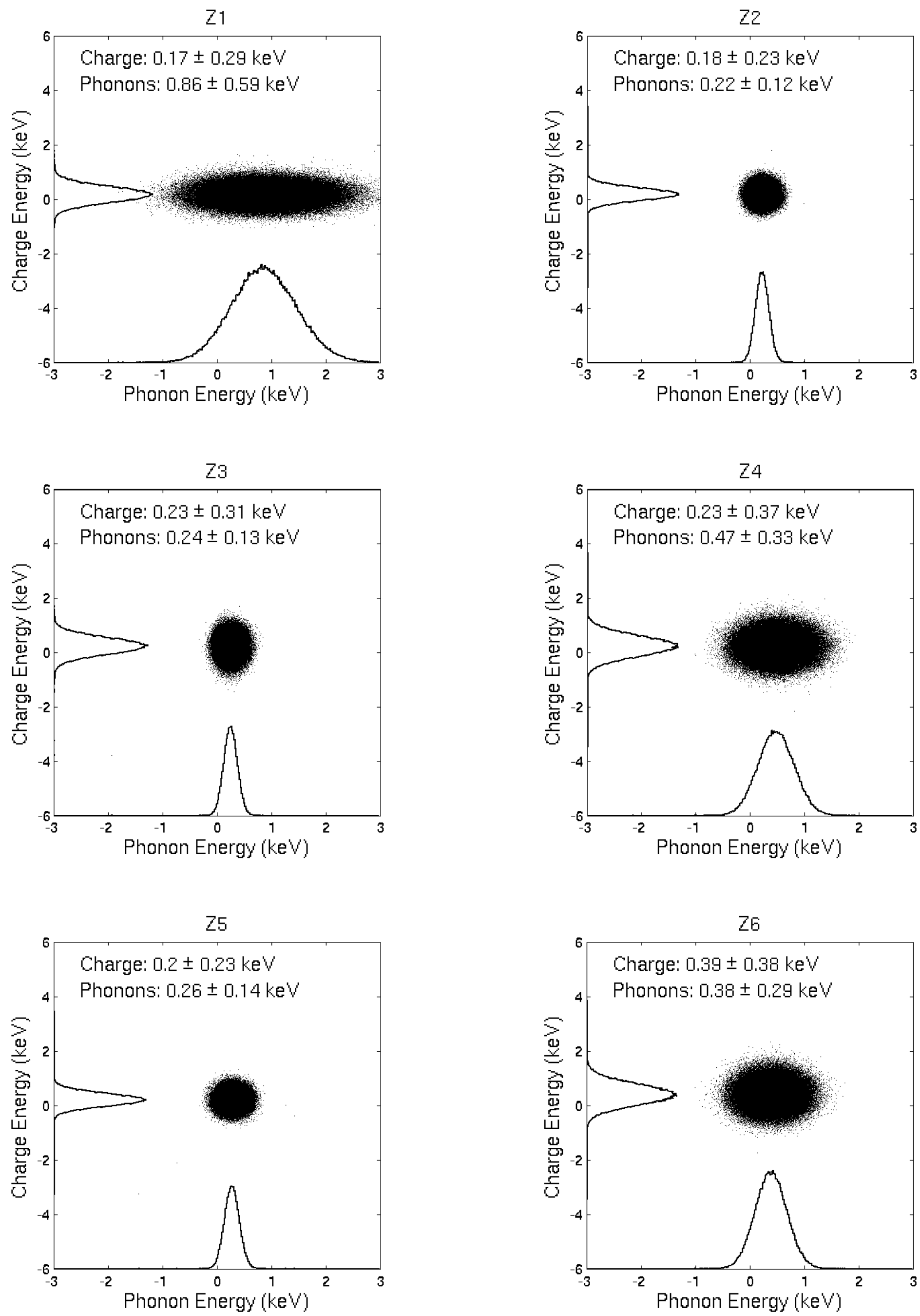


Figure 6.4: Noise blobs for the Tower 1 detectors. The events plotted are all randomly triggered WIMP search events. i.e. the histograms above represent the fitted energy to baseline noise. The figures show means and standard deviations from gaussian fits to the noise blob distributions; these values are used to set trigger thresholds and to define single and multiple cuts. The poor noise behavior of Z1 is due to it's large T_c gradient.

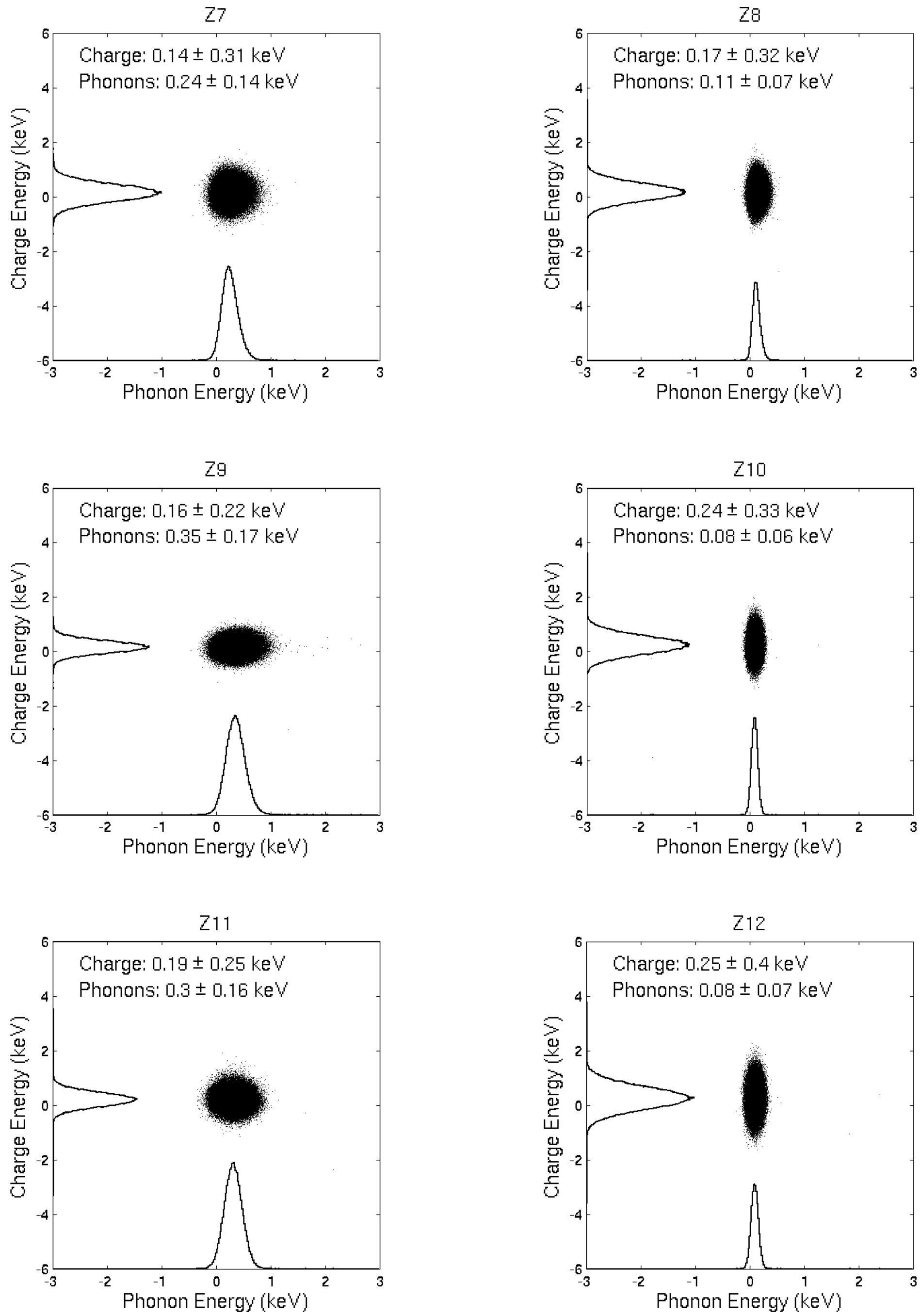


Figure 6.5: Noise blobs for the Tower 2 detectors. The events plotted are all randomly triggered WIMP search events. i.e. the histograms above represent the fitted energy to baseline noise. The figures show means and standard deviations from gaussian fits to the noise blob distributions; these values are used to set trigger thresholds and to define single and multiple cuts.

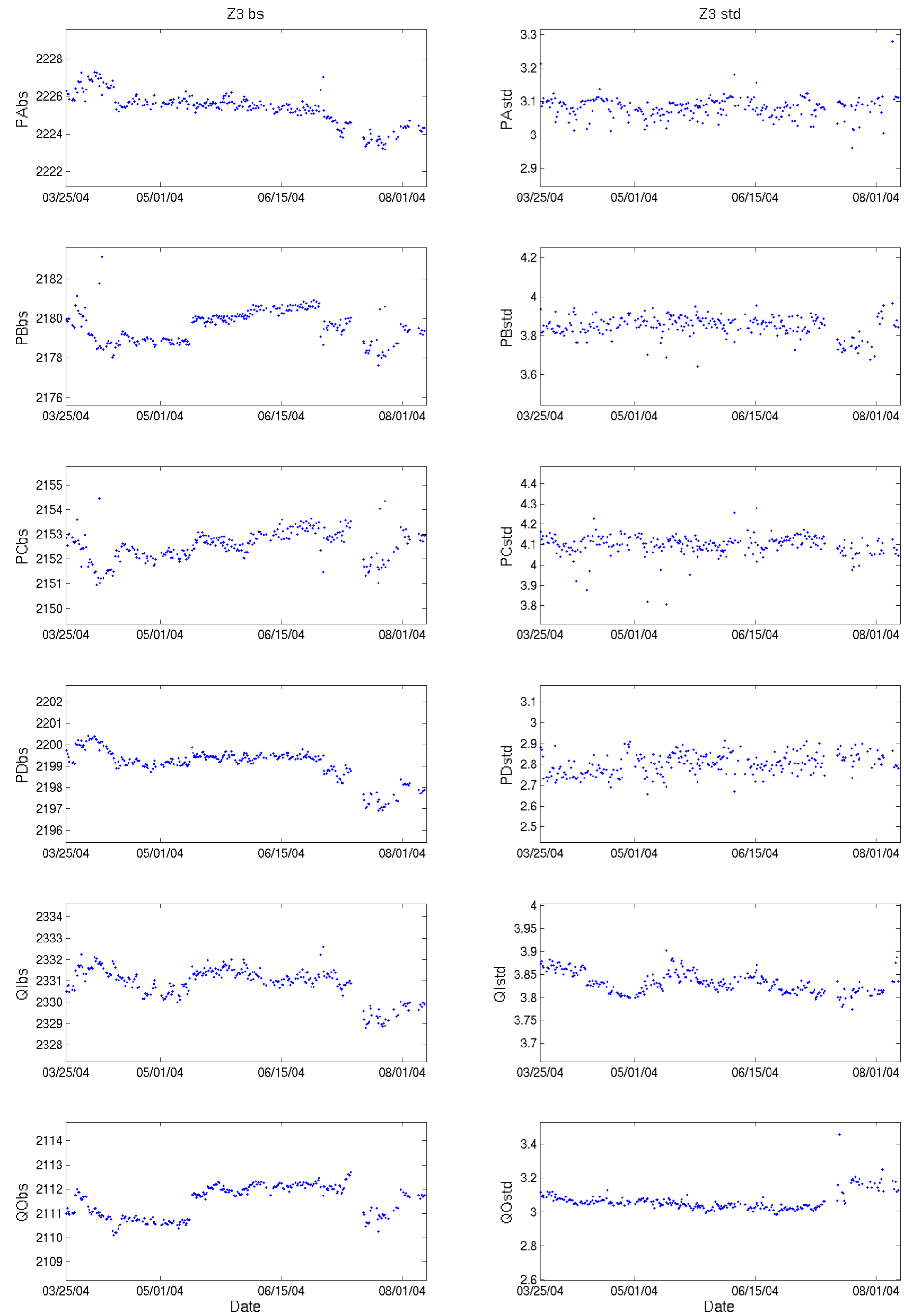


Figure 6.6: Variation of the pre-trigger baseline (bs) and pre-trigger standard deviation (std) in digitizer bins for all channels of Z3, shown for all WIMP search datasets over the course of Run 119. 1 keV $\sim$ 14 digitizer bins in the phonon channels, 1 keV $\sim$ 4 digitizer bins in the charge channels.

averaged for each data series over the run. The noise is stable over the course of the run, with an rms deviation of $\sim 0.01(0.007)$ keV from a constant baseline noise in the charge (phonon) channels. Fluctuations in the phonon noise do not seem to be correlated with fluctuations in the ionization noise, and fluctuations in the noise of a given phonon channel does not seem to be correlated with noise fluctuations in other phonon channels. Individual events that have values of the pre-pulse baseline standard deviation greater than 6σ above the mean are cut out in our analysis (see Section 6.5.3); this figure is intended to show that the noise is stable from data series to data series.

Trigger Efficiencies

Since the phonon noise is lower than the ionization noise, and the phonon baseline resolution is lower than the ionization baseline resolution, we use the sum of the four phonon channels on which to trigger all detectors. However, the RTF boards do not always flag a detector trigger even though the actual phonon signal was above threshold.

If we only look at the triggering of the detector Z, we learn nothing about the times that this detector didn't trigger. We wish to measure the deposited energy in detector Z (via the digitizer traces) and simultaneously examine whether or not the detector also triggered (via the corresponding trigger bit in the history buffer).

The trigger efficiency in a given detector Z is calculated by first preselecting all events for which a detector other than detector Z caused the global trigger. The trigger efficiency of Z is then the fraction of events that had a phonon trigger in detector Z occurring within $50\mu s$ of the global trigger, as a function of the phonon energy in detector Z.

The trigger efficiencies are a consequence of the acquisition hardware operation, and not

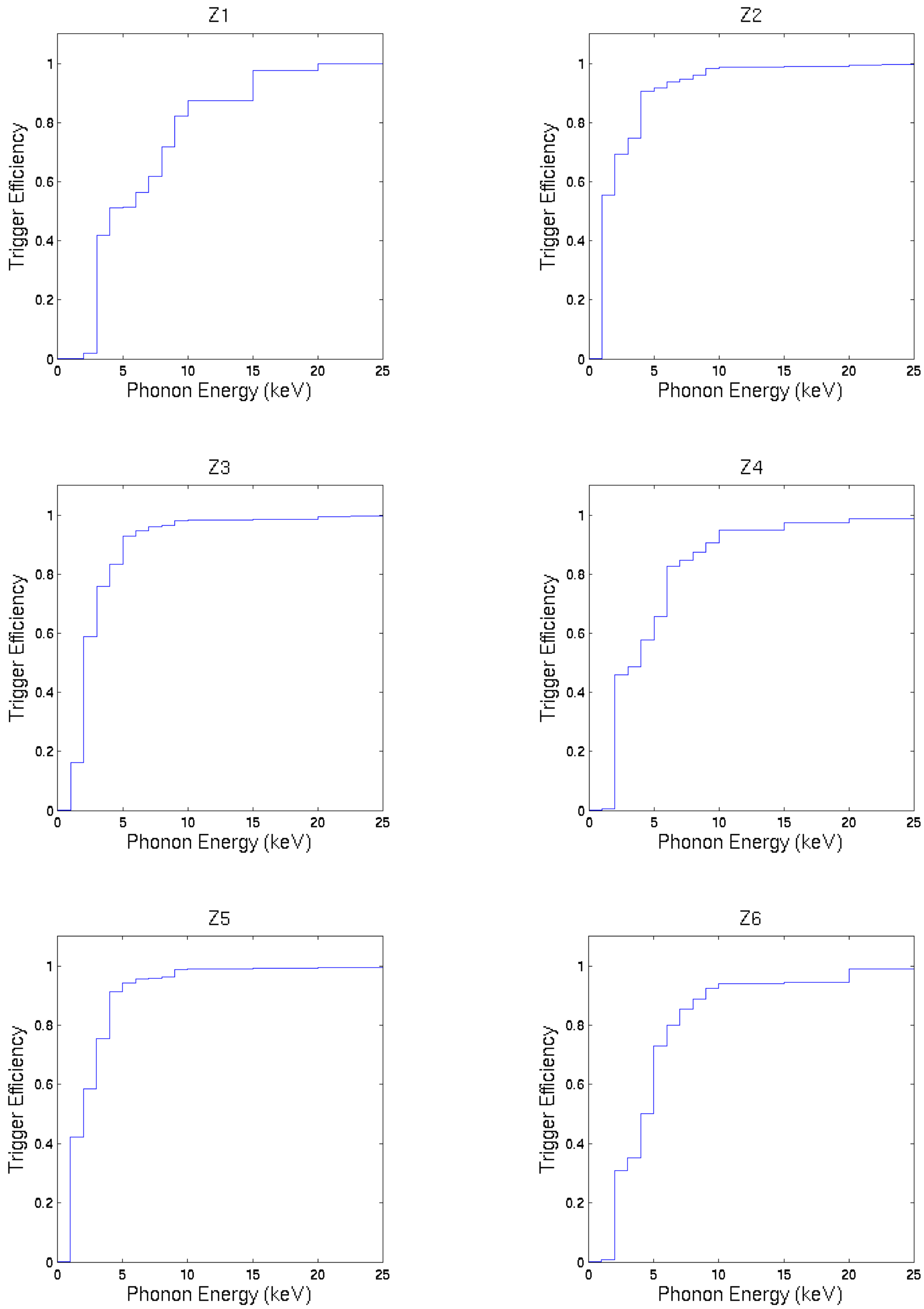


Figure 6.7: Phonon trigger efficiencies for Tower 1 detectors, as a function of the total phonon energy in the detector.

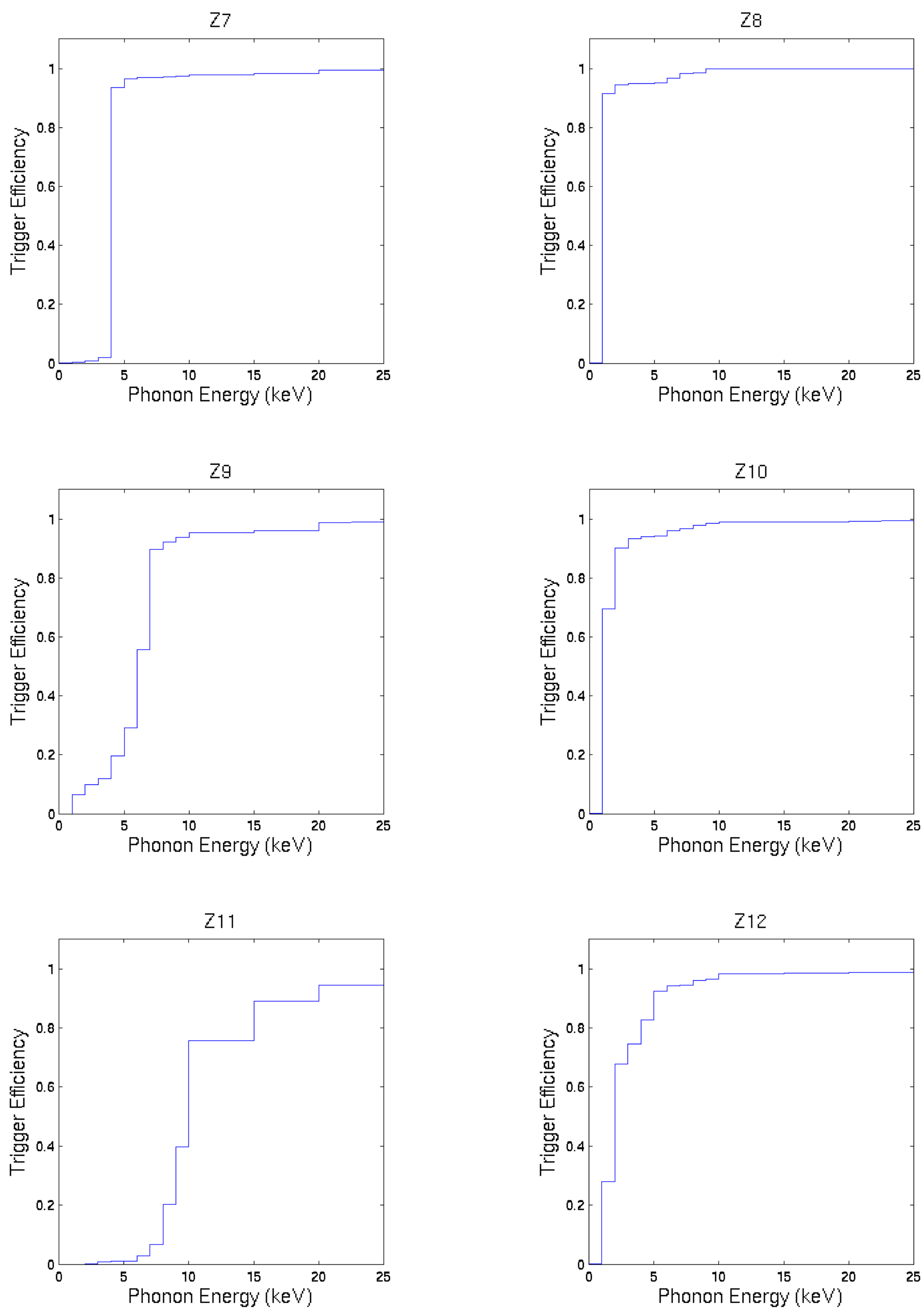


Figure 6.8: Phonon trigger efficiencies for Tower 2 detectors, as a function of the total phonon energy in the detector.

any underlying physics in the detectors. Some of the detectors have efficiencies that either rise slowly to 100%, or do not reach 100%. This is due to a known bug in the RTF boards used with these detectors; Z9 and Z11 being the worst in this respect.

The trigger efficiencies calculated from WIMP search data are shown in Figures 6.7 and 6.8. The trigger efficiencies for calibration data are calculated to be slightly higher than for the WIMP search data, but we use the efficiencies shown when calculating the net efficiency upon our WIMP signal region.

6.4 Detector Calibration

6.4.1 Energy Calibration

The energy calibration of the detectors is performed by matching the spectrum of ^{133}Ba calibration data to Monte Carlo (GEANT 4) predictions of the spectrum. Since the ionization signal has a linear response in energy, and the phonon signal does not, the calibration is performed by matching the ionization energy spectrum to the calibration data, and then calibrating the phonon spectrum to match. We can do this since we know that the mean ionization yield of the bulk gamma events is defined to be 1 (electron recoils). The non-linearity of the phonon response is corrected as part of the set of phonon corrections (see Section 6.4.3).

Figures 6.9 and 6.10 compare the results of Monte Carlo simulations to the ^{133}Ba calibration data. Both Monte Carlo and calibration data have had an inner electrode cut applied, and a noise term the same width as the corresponding noise blob has been added to the Monte Carlo data, to simulate each detector's energy resolution. In order to fit the

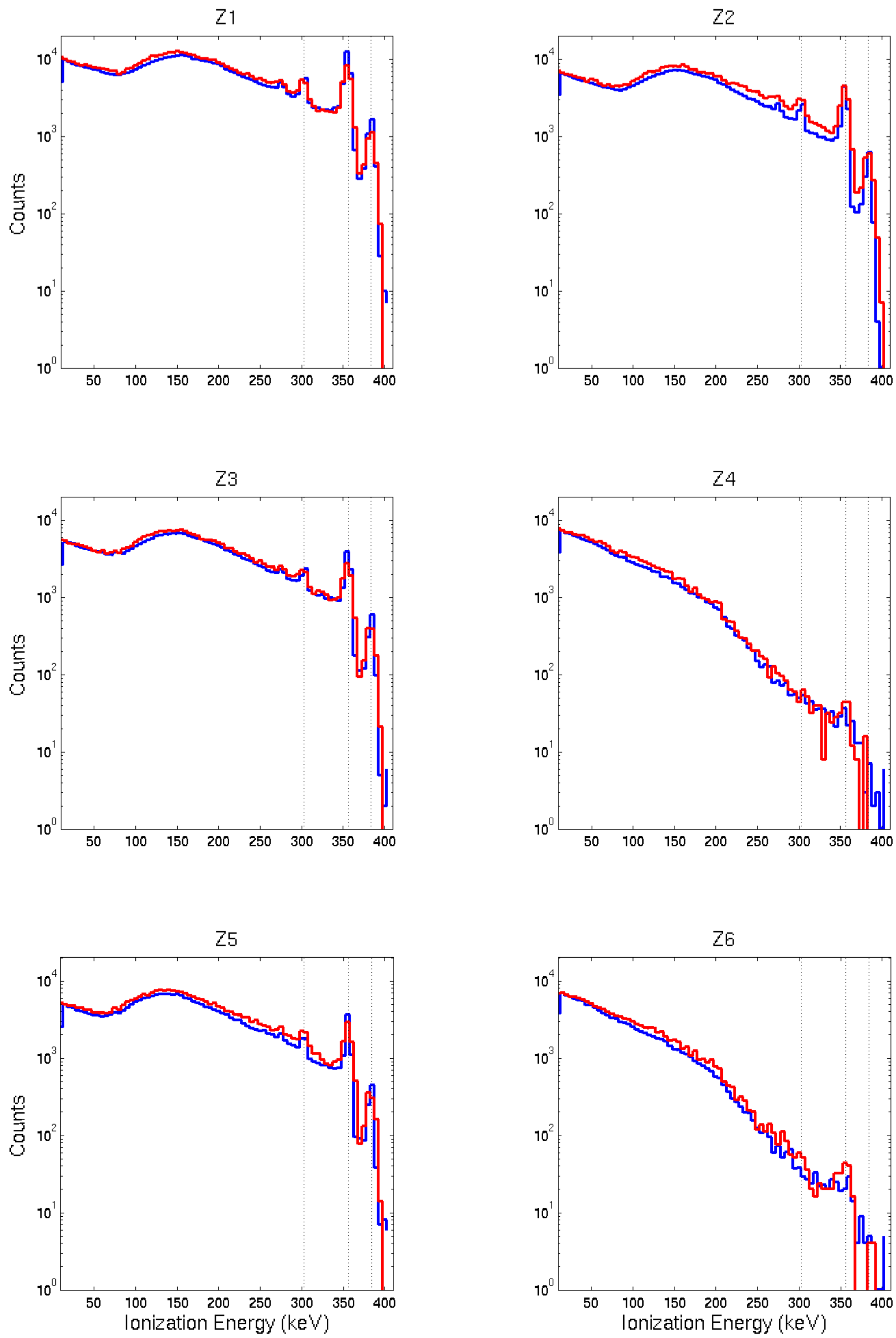


Figure 6.9: Comparison of ^{133}Ba calibration data (blue) with GEANT 4 Monte Carlo simulation (red) for Tower 1 detectors. The dotted vertical lines denote the energies of the primary ^{133}Ba gamma lines at 303 keV, 356 keV and 384 keV. Monte Carlo simulation performed by L. Baudis.

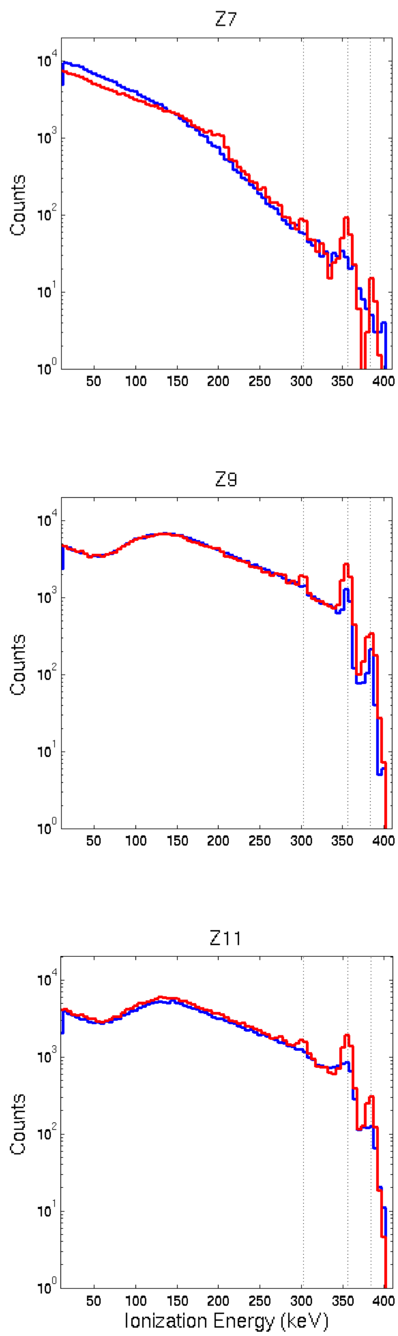


Figure 6.10: Comparison of ^{133}Ba calibration data (blue) with GEANT 4 Monte Carlo simulation (red) for Tower 2 detectors. The dotted vertical lines denote the energies of the primary ^{133}Ba gamma lines at 303 keV, 356 keV and 384 keV. Monte Carlo simulation performed by L. Baudis.

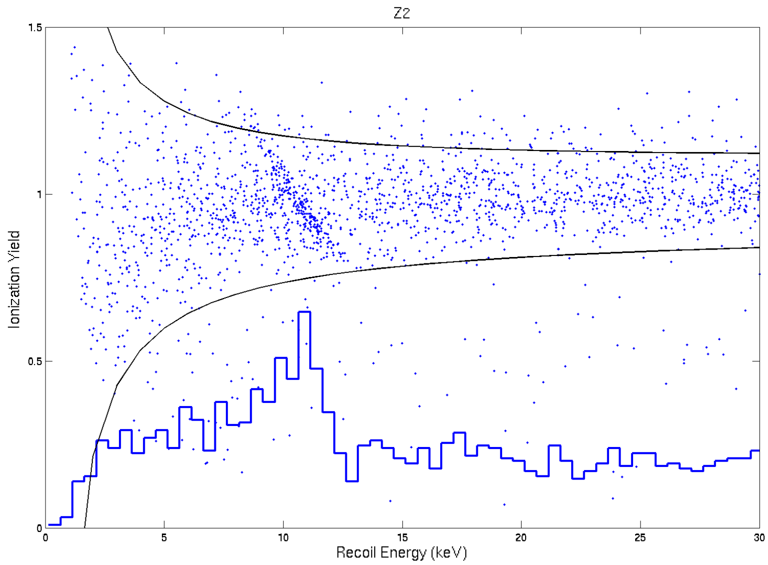


Figure 6.11: Low energy WIMP search events in Z2 that are above threshold and pass the good event cut (to be described). The 10.4 keV line results from an x-ray cascade after electron capture on ^{68}Ge. The histogram shows event counts on a linear scale.

Monte Carlo data to the calibration data, only the overall (linear) energy scale of each detector is considered a free parameter. i.e. the relative rates and detector resolutions were fixed.

The fit of the Monte Carlo to calibration data is extremely good, especially in the Ge detectors. The low photoelectric scattering cross-section to Compton scattering ratio in Si means that the primary lines are not as visible in the Si ZIPs.

A secondary check on the energy calibration can be performed by studying the activation lines of the Ge detectors. One of the activation modes has a half-life long enough for the line to be observable many months after the detectors are taken underground:

- $^{70}\mathrm{Ge} + \mathrm{n} \rightarrow {}^{68}\mathrm{Ge} + 3\mathrm{n}$
- $^{68}\mathrm{Ge} \rightarrow_{(E.C.)} {}^{68}\mathrm{Ga}$ + (x-ray cascade, total of 10.4 keV), $\tau_{1/2} = 270$ days.

Hence the 10.4 keV line can be examined in order to check the low energy calibration of the detectors. Figure 6.11 shows the low energy WIMP search events in detector Z2, where the 10.4 keV line can be seen clearly.

6.4.2 Ionization Position Correction

The ionization energy response of each detector has a position dependence at the ∼3% level. The exact reason for this is unknown, although it is likely to be associated with T_c gradients and/or the neutralization state of the detector.

We apply a simple linear correction using the phonon x- and y-delay values to remove this dependence using ^{133}Ba calibration data. Figure 6.12 shows the ionization energy of events as a function of position, and it can be seen that there is a clear variation of ionization energy with y-delay (the same variation occurs along the x-axis of the detectors). A linear position-dependent transformation is applied to the ionization energy of events in both the x and y directions, such that the ^{133}Ba 356keV line appears at the correct energy at all detector positions, as shown in Figure 6.12 for the y-axis correction.

6.4.3 Phonon Position Correction

The phonon signal must also be corrected for position variations across the crystal. The process of the phonon position correction and the physics that necessitate it are described in Section 4.7.1; here we show the results of applying the correction to Run 119 data.

Before the phonon lookup table can be defined and the phonon timing parameters corrected, the phonon energy must be linearized. Figure 6.13 shows the recoil energy of ^{133}Ba calibration events as a function of their ionization energy for one detector. Since

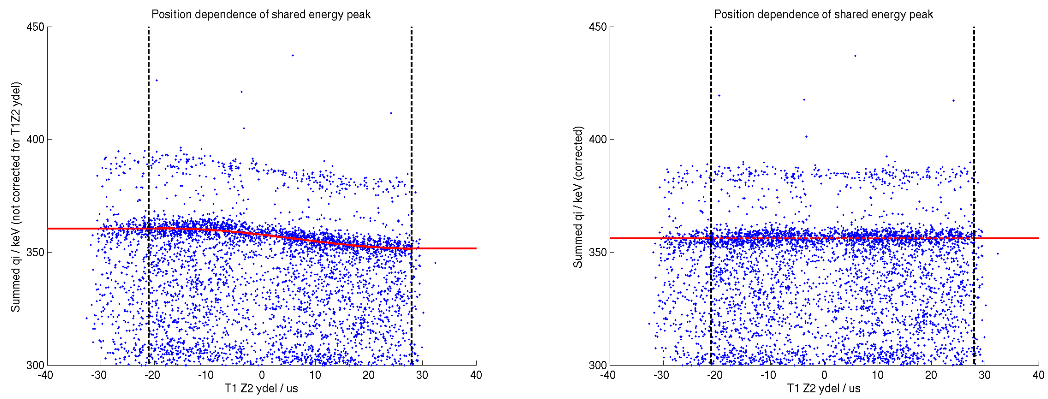


Figure 6.12: Position correction of the ionization signal for Z2 calibration events. The ionization energy is plotted against y-delay position, and it can be seen that the energy of the primary ^{133}Ba line is dependent upon position (uncorrected, left) at the $\sim$3% level. A simple linear transformation is applied to remove this position dependence (corrected, right). Figure by R. W. Ogburn.

(ionization energy) / (recoil energy) = (ionization yield) is defined to be equal to one for electron recoil events, this curve is linearized to have a slope of one using the scaling $P_{\mathrm{new}} = \tau(e^{P/\tau} - 1)$ with $\tau \sim 10^3$. The ^{133}Ba data is used to define the transformation, which is then applied to all data.

A similar transformation is applied to the phonon timing parameters prior to calculation of the lookup tables, since the timing parameters also have an non-linear energy dependence. This is shown for the phonon risetime in detector Z2, phonon channel B in Figure 6.14. Here each timing parameter is corrected using a scaling of the form $a + c(E_r)^b$ for every detector's phonon channel independently.

6.5 Cut Definitions

Since there are small differences between each of our detectors (variations in T_c gradients, etc.), our analysis cuts are usually defined on a detector-by-detector basis. The cut usually

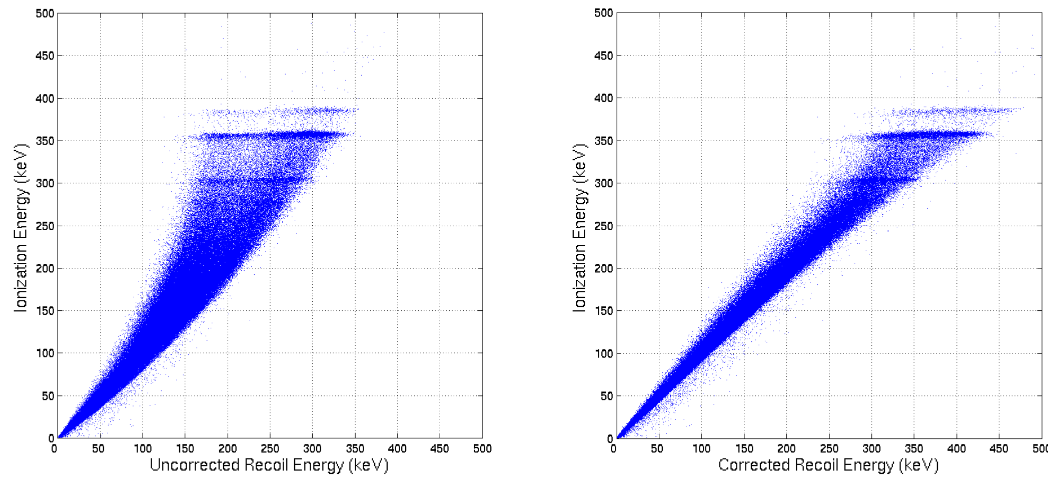


Figure 6.13: Correction of the phonon energy nonlinearity in detector Z2 for ^{133}Ba calibration events. Since these events are known to be electron recoils, the slope (Ionization Energy)/(Recoil Energy) = (Ionization Yield) is defined to be equal to 1.

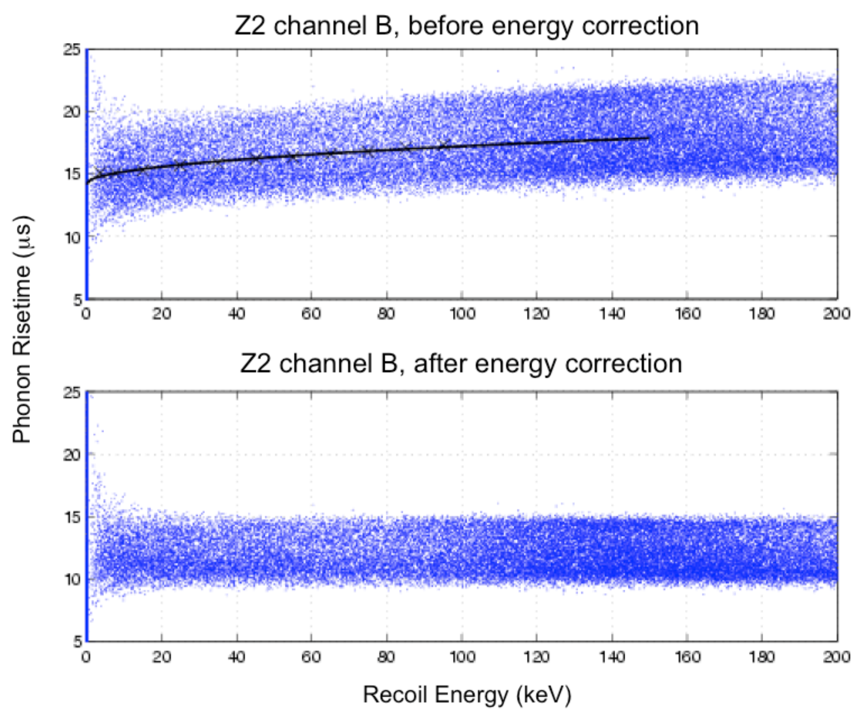


Figure 6.14: Correction of the phonon risetime's dependence upon energy in detector Z2, phonon channel B, for ^{133}Ba calibration events. Each timing parameter is corrected using a scaling of the form $a + c(E_r)^b$. Figure by B. Serfass.

has a general form, but the exact values will vary from ZIP-to-ZIP. The subsequent sections describe our primary analysis cuts. Note: Where events are referenced as 'passing a cut', this by convention means that these events will remain in the analysis when the cut is applied. Events 'failing a cut' are removed from subsequent analysis.

As described in Section 4.6, half of the calibration data is used to set cuts while blind, with the remaining calibration data being used to calculate cut efficiencies.

With the exception of the bad data cut, which necessarily must be defined for all data *a priori*, the analysis cuts were all defined using a pre-selected subset of the calibration data. This data (referred to as the 'open' calibration data) was simply chosen as being all those calibration datasets with even-numbered data series numbers (the remaining odd-numbered series were defined as the 'closed' set). This results in a near-equal split of the data, with both the 'open' and 'closed' groups of data being fairly evenly spread across the data run.

The statistics in the neutron calibration data were low enough that it was decided not to break up this dataset into open and closed subsets. This will introduce bias into the cuts that are set using the neutron calibration data (the nuclear recoil band and phonon timing cuts), but the reduction in statistics that would result by splitting up the dataset would introduce uncertainties of a greater magnitude.

The open calibration data was used to define our analysis cuts and the closed calibration data used to determine initial cut efficiencies and estimate expected backgrounds. The WIMP search data with the blinding cut applied (i.e. with events in the signal region removed) was used to calculate further expected cut efficiencies and to examine any systematics between the calibration and WIMP search data. We shall return to the unblinding process and calculation of background event leakage in Section 6.8.

6.5.1 Data Quality and the Bad Data Cut

The identification of datasets that had one or more problems was performed in the following way. A set of diagnostics were run on the entire dataset (blinded WIMP search & calibration data). These diagnostics performed Kolmogorov-Smirnov (K-S) tests, using a template from a subset of the relevant search data (i.e. a ^{133}Ba data template was used to test ^{133}Ba data, etc.). Any dataset that significantly diverged from the template dataset on the basis of these tests was flagged. Datasets that were flagged for several problems were dropped from the list of good datasets. Analysis was also performed on the blinded WIMP search data during the run, so that serious problems could be highlighted and/or addressed.

This approach highlighted the following:

- Five WIMP search datasets were removed from the good set list due to a) missing veto traces ($\times$2), b) bad triggering ($\times$1), c) very low event count with poor K-S test values ($\times$1), d) very poor ionization noise in K-S test ($\times$1).

- Four WIMP search datasets had higher than usual chi-squared values for the optimal filter fit to the phonon pulses. The noise PSDs for these datasets, however, were typical and in fact the phonon noise blob was narrower than usual. The decision was made to keep these datasets in the good data list.

- Four WIMP search datasets had poor K-S test values for phonon delay and phonon partition distributions. They were cut from the good dataset list.

- A handful of events spread over the run are missing global trigger values (due to a problem with the DAQ). These individual events are cut based on their trigger mask values.

- Three ^{133}Ba datasets had trouble with one or more Z6 phonon channels. Events from those datasets in that detector were cut out.

- A WIMP search dataset was inadvertently still running when the fridge transfer started (and hence the fridge temperatures became slightly elevated). The last few hundred events of this series were cut.

- The end of a WIMP search dataset suffered from having the SQUIDs lose lock due to a power trip on a pump. The events during this period were cut.

This selection process resulted in a complete list of good datasets which was used for all subsequent cut definitions. A 'bad data cut' was also defined for the good datasets containing a subset of events to be excluded from the analysis (i.e. the final four items in the above list).

Previously quoted values of the total Run 119 livetime haven taken account of the datasets listed above that were thrown out in their entirety, for example in Figure 6.2. Application of the bad data cut to the remaining good datasets reduces the total livetime of the good datasets from 77.44 days to 77.28 days.

6.5.2 Charge χ^2 Cut

The ChiSq cut is designed to remove events for which a noise spike occurred in either of the ionization traces. The optimal filter fit to the ionization pulse template provides a χ^2 value which is used to perform this cut, using a simple cut of the form $aQ^2 + b$, where Q is the ionization energy, and a and b are constants to be determined on a detector-by-detector basis. Figure 6.15 shows the form of the cut for an example detector.

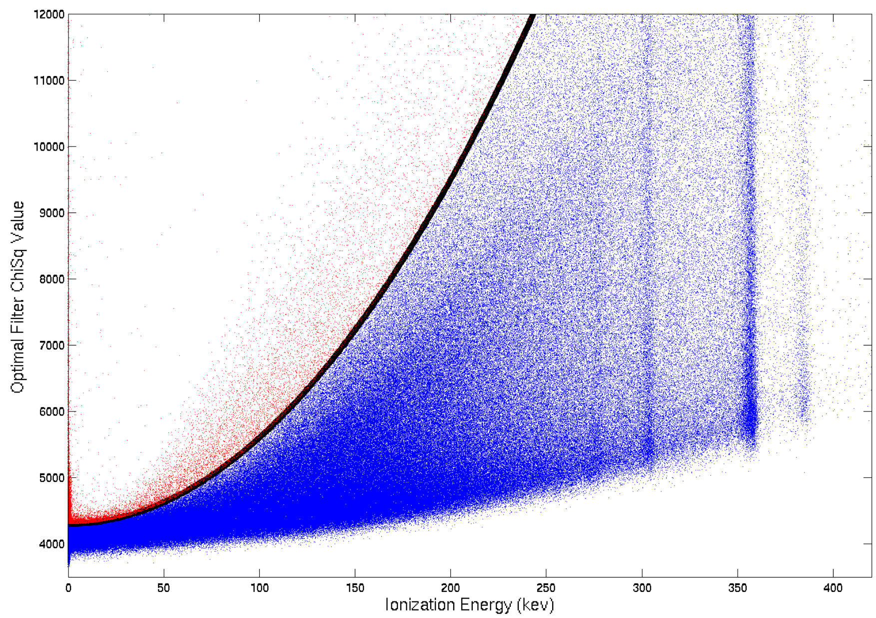


Figure 6.15: Definition of the Charge ChiSq cut for detector Z2. Open ^{133}Ba data used to define the cut is shown. Those events in red fail the cut, those in blue pass the cut.

6.5.3 Phonon Pretrigger Cut

Another cut designed to remove events which suffer noise glitches is the phonon pretrigger cut. This cut removes events which have anomalously high values of the standard deviation of the phonon pretrigger baseline (the 'Pstd' value). We expect the Pstd distribution to be close to Gaussian, and use a value 3.85σ above the mean of a fitted Gaussian with which to cull events. These values are defined for each phonon channel of each detector. Figure 6.16 shows an example phonon sensor's Pstd values along with the location of the cut for this sensor.

The combination of bad data cut, charge chi-squared cut and phonon pretrigger cut is usually referred to as the 'Good Event Cut' (cGoodEv), in that only events that pass all three cuts make it through into the analysis. Application of cGoodEv is a prerequisite for

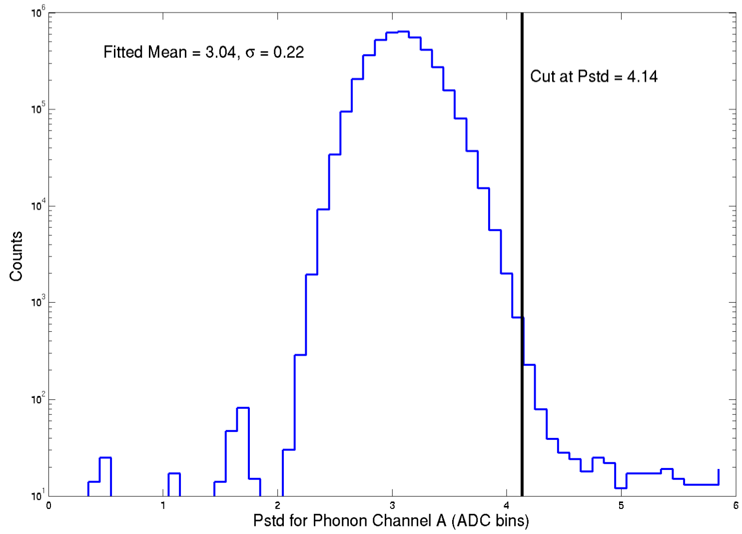


Figure 6.16: Definition of the Phonon Pretrigger Cut for detector Z3, Phonon channel A. Open ^{133}Ba data used to define the cut is shown. Events with higher Pstd values than the vertical black line fail the cut.

performing any analysis of the data.

6.5.4 Charge Threshold Cut

The charge threshold cut removes events that could potentially be noise in the ionization signal. It is set by considering the noise blobs in the inner electrode ionization signal and cutting events for which this signal has an amplitude less than 6 standard deviations above the fitted gaussian mean of the noise blob (see Figures 6.4 and 6.5). These cutoff values vary in magnitude from best of 1.51keV in Z9 to worst of 2.85keV in Z12. Figure 6.17 shows the application of the charge threshold cut in the (y, E_r) plane. Thus it can be seen that the charge threshold cut removes both events with low energy and events with low ionization yield.

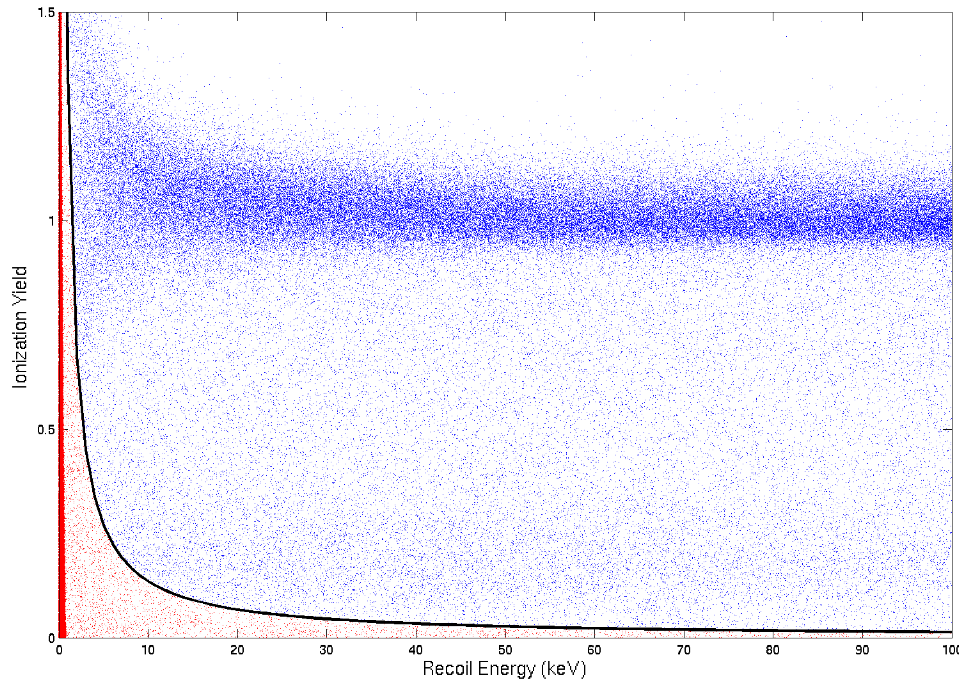


Figure 6.17: Application of the Charge Threshold Cut for detector Z3. Open ^{133}Ba data used to define the cut is shown - the cut removes all events with an inner electrode ionization energy of less than 1.35keV in this detector (the cut value is different but similar for other detectors). This appears in the (y, E_r) plane as a curved line (shown in black), below which events are cut (blue events pass the cut, red events fail the cut). A Good Event Cut has been applied to the data (= $\sim$cBad & cChiSq & cPstd)

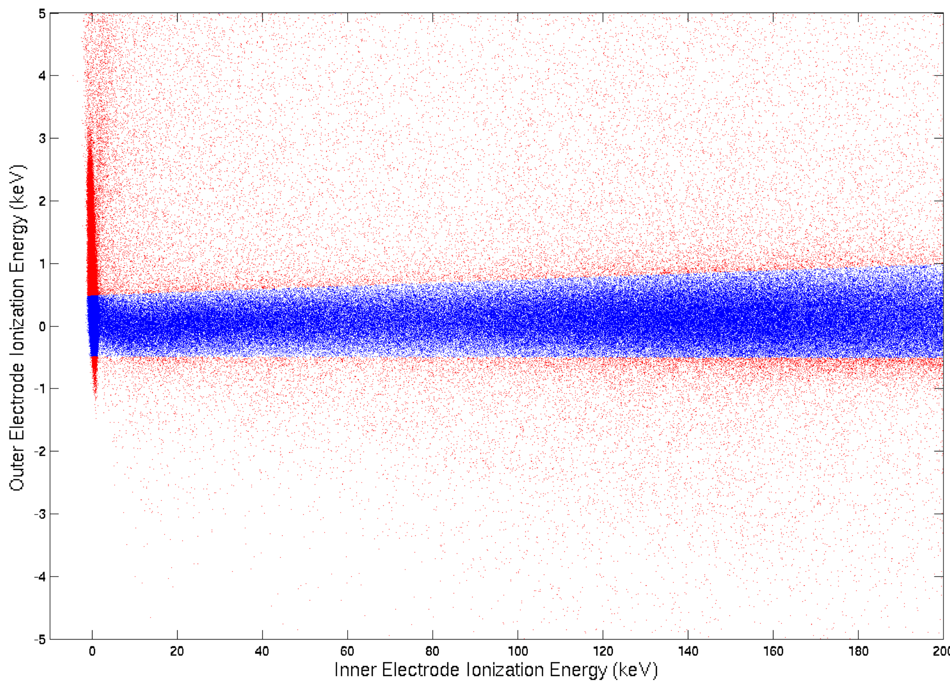


Figure 6.18: Application of the Inner Electrode Cut to Z3 for ^{133}Ba data with cGoodEv applied. Events that pass the cut are in blue, events in red fail the cut.

6.5.5 Inner Electrode Cut

The outer electrode region of the detectors, which represents 15% of the detector volume, suffers from poor charge collection and thus it is advantageous to remove events which deposit energy primarily in this region. The inner electrode cut, cQin, typically removes events that deposit greater than 0.5keV-1keV in the outer electrode, although the cut is energy dependent as shown in Figure 6.18 for an example detector.

6.5.6 Saturated Pulse Cut

This cut removes events for which the phonon pulse is saturated. i.e. the signal is clipped by the digitizer. Events for which this occurs could have sufficiently large ionization energies to result in a low estimation of the recoil energy. These events typically occur at energies > 1MeV and are removed from our analysis.

6.5.7 Muon Veto Anticoincident Cut

Events that occur in the detectors within a particular time window of an event above threshold in the muon veto are cut in our primary analysis, since they are assumed to be caused by muon interactions. The size of the time window is chosen based on the rate of such events; Figure 6.19 shows the times of the most-recent above-trigger veto events relative to the global trigger (this information is easily retrieved from the veto history buffer). The times of the most recent veto events correspond to a trigger rate of ~ 0.6 kHz.

The veto anticoincident cut, cVT, removes events for which there was a hit above threshold in the muon veto in the $50\mu s$ before the global trigger. This value is chosen to safely cover potential signals prior to the phonon pulse rising above threshold.

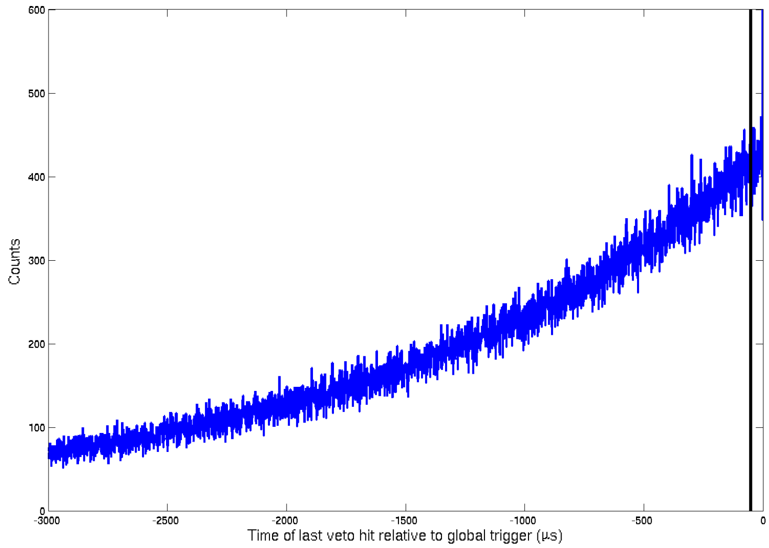


Figure 6.19: Time of events above threshold occurring within the muon veto, relative to the global trigger at $t = 0$, for WIMP search data. The veto anticoincident cut removes events where the time of the last veto hit occurs in the $50\mu s$ before the global trigger (vertical black line).

6.5.8 Singles and Multiples Cuts

A 'hit', for both the single and multiple hit definitions, is defined as a 6σ outlier in the noise blob distribution (see Figures 6.4 and 6.5). i.e. a single-scatter event is defined as one for which the phonon energy is greater than 6σ above the mean of the noise blob, in one and only one detector. A multiple event is similarly defined as an event for which the phonon energy is greater than 6σ above the mean of the noise blob in more than one detector. Note that this means that not all events are necessarily singles or multiples - there are some low energy events above threshold that do not fit either cut definition.

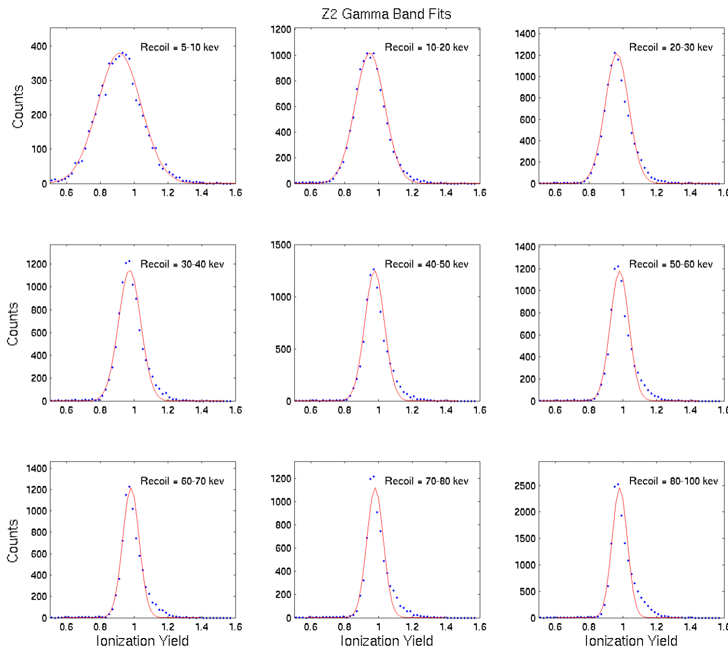


Figure 6.20: Gaussian fits to the ionization yield distribution for Open ^{133}Ba data in Z2. The fitted means and σs are used to fit the band centers and widths (see Fig. 6.22 and 6.23 respectively).

6.5.9 Electron and Nuclear Recoil Band Cuts

The definition of the electron and nuclear recoil bands is based on the theory of Lindhard [79]. In general we expect each set of recoil types to form bands for which the mean and 1σ-widths take the forms shown in Equations (2.11) and (2.10). The calculation of the constants a, b, c and d for each detector is performed as follows.

The data being used to define the band (^{133}Ba data in the case of the electron recoil band, ^{252}Cf data in the case of the nuclear recoil band), is binned by recoil energy, and a gaussian is fit to the ionization yield distribution for each energy bin (Figures 6.20 and 6.21). The means of each bin is fit to the relation in Equation (2.10) to provide estimates for the values of a and b (Figure 6.22). Similarly, the σs of the fitted gaussians are fit to

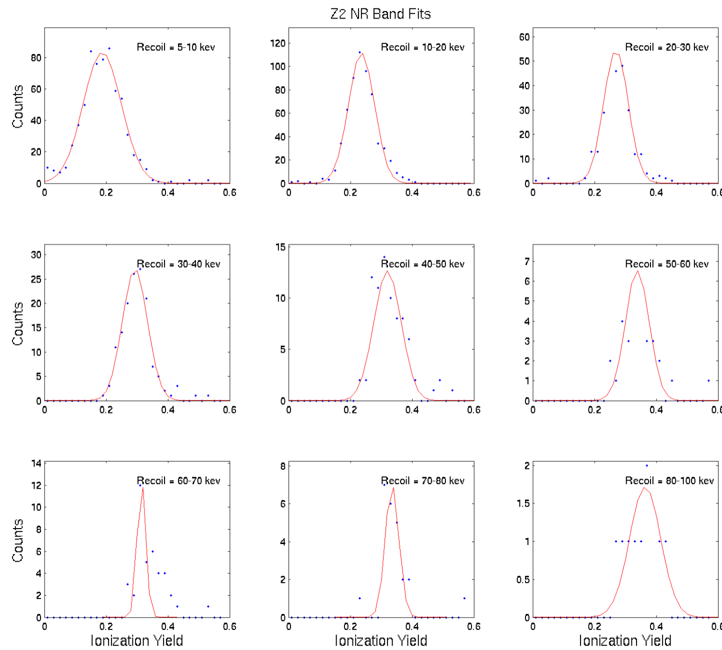


Figure 6.21: Gaussian fits to the ionization yield distribution for all ^{252}Cf data in Z2. The fitted means and σs are used to fit the band centers and widths (see Fig. 6.22 and 6.23 respectively).

Equation (2.11) to give values for c and d (Figure 6.23).

As with the other cuts, the open ^{133}Ba data was used to define the electron recoil band, but due to low statistics, the entire set of ^{252}Cf data was used to define the nuclear recoil band [1]. Thus we are unable to perform an unbiased efficiency calculation on the nuclear recoil band (since we used all of the data to create the cut), but the alternative would clearly be much worse since the statistics would be useless at higher energies (see number of counts in Figure 6.21).

This procedure of calculating the band center and width distributions is followed for all 12 detectors individually. Figures 6.24 and 6.25 show data from Tower 1 and 2 ZIPs respectively, along with the 2σ bands. The potential WIMP signal region is usually taken

[1]For the neural network analysis to be presented in Chapter 9, the training of the network necessitated the splitting of the ^{252}Cf data into two sets

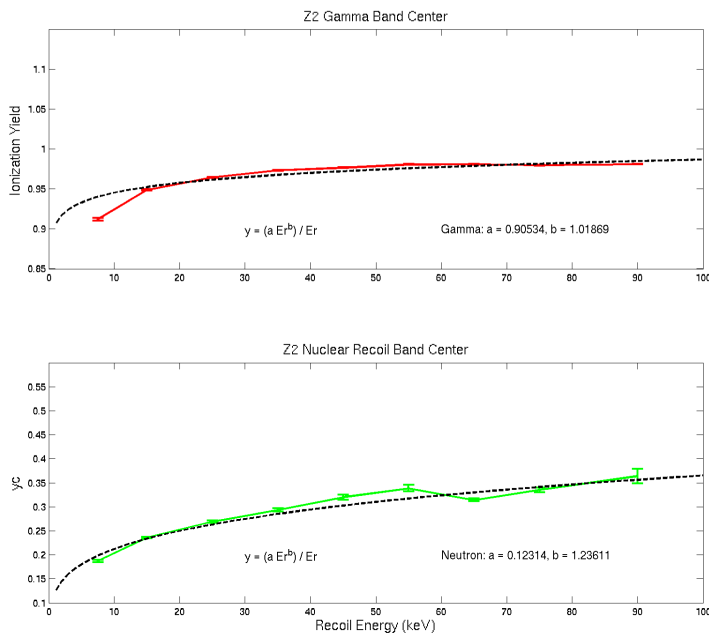


Figure 6.22: Fit of the band means to Eq. (2.10). The red line attempts to fit the means of the gamma band shown in Fig. 6.20 to a non-linear function of the recoil energy. Similarly, the green line fits the means of the nuclear recoil band in Fig. 6.21 to the same function with different parameters. The error bars are taken from the error in the gaussian fits in Fig. 6.20 and 6.21.

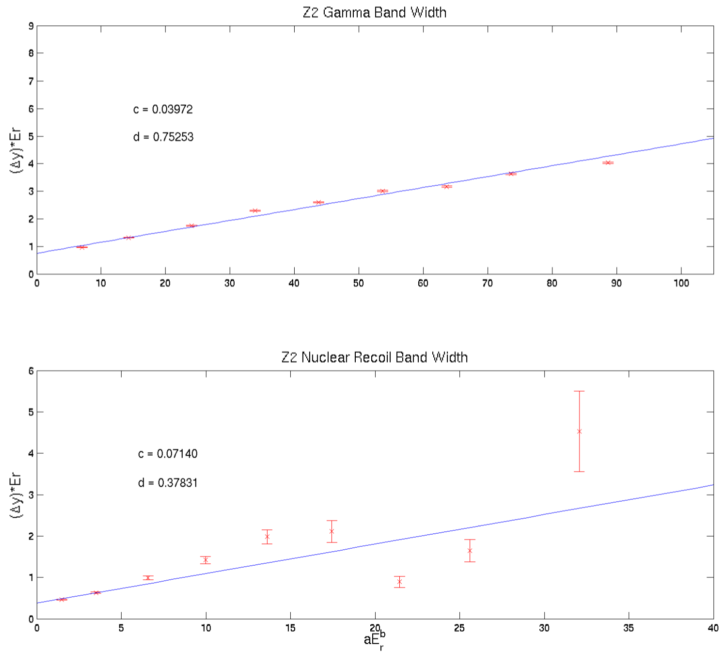


Figure 6.23: Fit of the band widths to Eq. (2.11). The error bars are taken from the error in the gaussian fits in Fig. 6.20 and 6.21.

to be the 2σ nuclear recoil band.

When calculating the recoil bands, only the inner electrode ionization energy was used. Previously the summed energy of the ionization channels and an inner electrode cut would have been used, but now we use only the inner electrode energy. The outer electrode suffers from elevated noise, so the outer electrode contribution to the total ionization energy for an event passing the inner electrode cut will be close to noise (see Figure 6.18).

Thus, quantities such as the ionization yield and recoil energy are now defined as (compare with definitions in Eq. (2.9) and (2.8)):

$$E_r = P - \frac{V_b}{E_{eh}} Q_{inner} = P - \alpha Q_{inner} \tag{6.1}$$

$$y = \frac{Q_{inner}}{E_r} = \frac{Q_{inner}}{P - \alpha Q_{inner}} \tag{6.2}$$

Subsequent references to E_r and y will consider these definitions to be implicit.

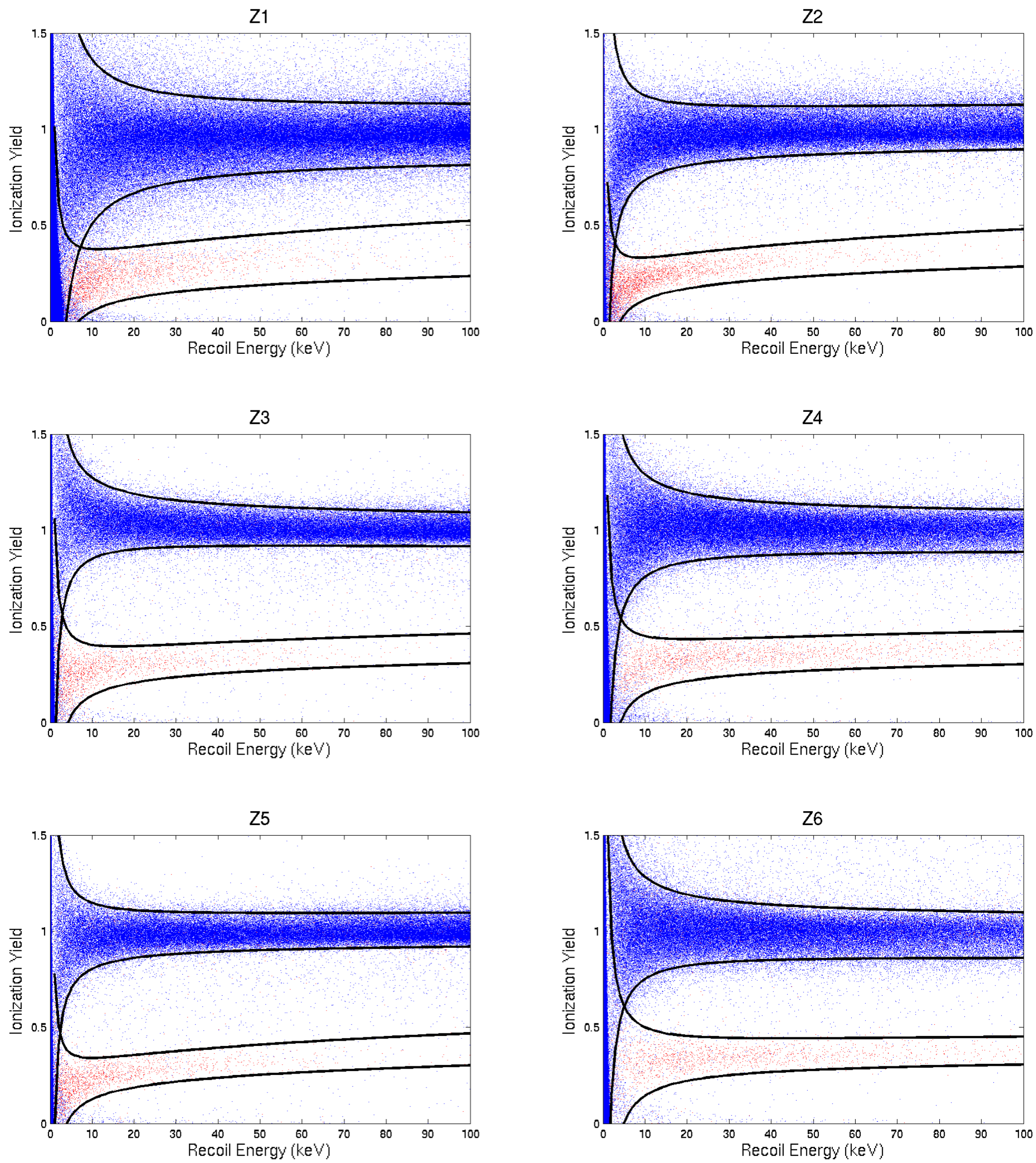


Figure 6.24: Calculated 2σ electron recoil and nuclear recoil bands (both shown as black lines) for Tower 1 detectors. Closed ^{133}Ba calibration data is shown in blue, all ^{252}Cf data is shown in red. A good event cut has been applied.

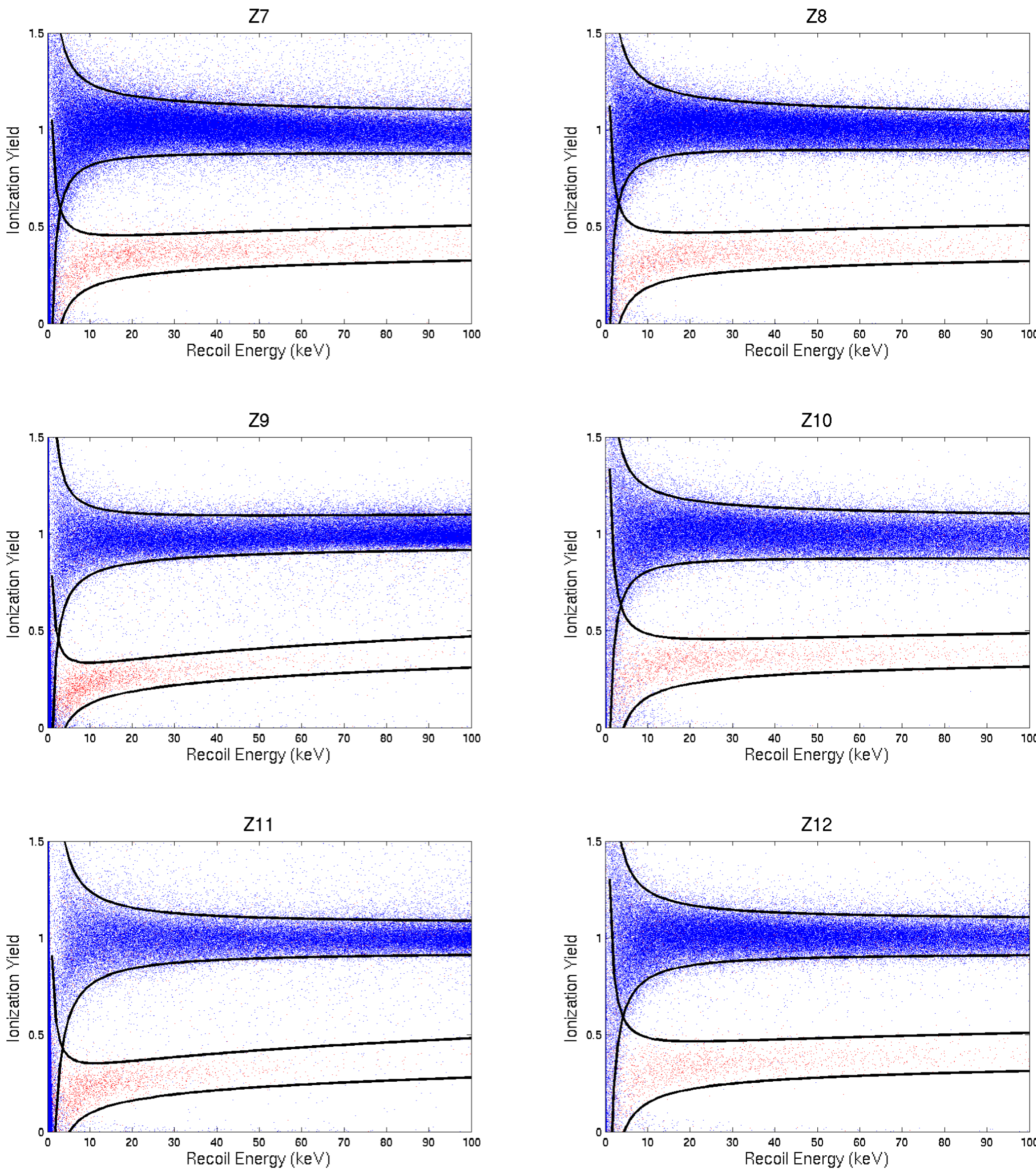


Figure 6.25: Calculated 2σ electron recoil and nuclear recoil bands (both shown as black lines) for Tower 2 detectors. Closed ^{133}Ba calibration data is shown in blue, all ^{252}Cf data is shown in red. A good event cut has been applied.

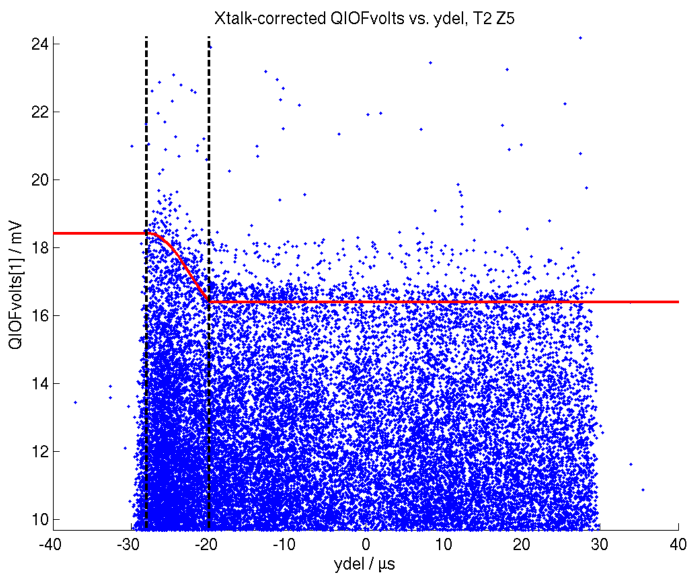


Figure 6.26: Amplitude of the inner ionization electrode signal in mV as a function of y-delay for Z11 ^{133}Ba events, showing the region of unusual charge collection at ydel $< -20\,\mu$s. The red line highlights the ^{133}Ba 356keV peak.

6.5.10 Z11 Cliff Cut

During calibration of the detectors, it was discovered that Z11 has a region in x-y that shows a peculiar ionization response. This behavior was not present during testing of this detector at the UC Berkeley testing facility, and it is unclear how this pathology could have arisen during assembly and transportation of Tower 2.

As Figure 6.26 shows, the ionization signal is clearly behaving differently in the ydel $< -20\,\mu$s region of the detector, with a $\sim$15% stepped increase in the ionization signal in this region. This area of the detector also has a significant number of calibration events that have very far outlying values in the phonon timing parameters. We opted to perform a cut that removed all events for which ydel < -20 μs in this detector.

6.5.11 Phonon Timing Cut

The phonon timing cut is designed to discriminate between 1) nuclear recoil events and 2) electron recoil events that deposit their energy in the top 20 μm of the detector ('surface events'), and hence have a reduced ionization yield such that they may appear in the nuclear recoil band (see Chapter 5). This discrimination is performed by examination of the phonon pulse shape, since surface events typically have a shorter phonon risetime than bulk events.

The general strategy applied in creating the timing cut is to maximize our expected sensitivity to a WIMP signal. The cut is created using open calibration data, and the resulting number of nuclear recoils and low-yield electron recoils passing the cut in the closed calibration data is used to extrapolate to the expected sensitivity and background estimates in the WIMP search data. Since we know the number of low-yield events in both the calibration and WIMP search data, and are blinded to potential WIMPs and/or neutrons, we can scale the observed sensitivity and background estimates on the calibration data to the expected sensitivity and background on WIMP search data.

The primary cut for the Run 119 analysis is constructed using a combination of the primary phonon pulse 10-40% risetime and the primary phonon pulse delay (ionization start time to phonon 20% time). The cut is a simple linear selection in the risetime-delay plane, as shown in Figure 6.27 for detector Z2, and is created using 2σ nuclear recoils events from ^{252}Cf data, and 'wide beta' events from ^{133}Ba data, where a wide beta is defined as an event for which $0.1 < y <$(lower bound of 5σ electron recoil band).

During Run 119, several alternative timing analyses were performed, and Chapter 9 describes one such attempt to utilize a wavelet-based analysis of the phonon pulses to

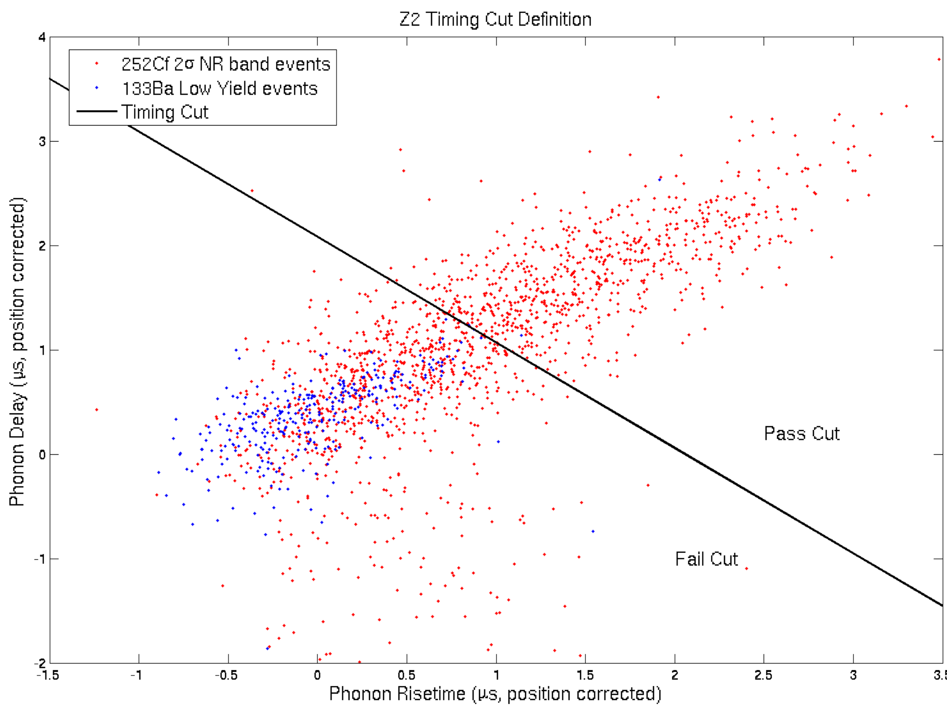


Figure 6.27: Phonon timing cut for Z2. Low yield (below 5σ electron recoil band) events from open ^{133}Ba calibration are shown in blue, events in 2σ nuclear recoil band from ^{252}Cf calibration are shown in red. All data has the following cut applied: cGoodEv & cQin & cQThresh & cVT & $(0 < E_r < 100\text{kev})$.

discriminate between bulk and surface events.

6.6 Cut Efficiencies

Except where stated otherwise, all cut efficiencies are calculated using neutron ^{252}Cf calibration events since this is how we define our WIMP signal region. The cuts were all defined (using open calibration data) and discussed before calculating efficiencies upon the closed data. In this way we sought to avoid biasing the cuts based on their performance on the closed calibration data. The exceptions are the nuclear recoil band and phonon timing cuts, which were defined using the entire neutron calibration dataset.

We treat Ge and Si detectors as separate entities in the calculation of efficiencies and subsequent calculation of limits. Hence we will finally arrive at two results - an estimate of the number of Dark Matter candidate events for WIMPs interacting with Ge, and an estimate of the number of candidate events for WIMPs interacting with Si. Three detectors are excluded from the main analysis due to poor operation; Z1 has a very large T_c gradient and consequently has poor recoil band discrimination, Z6 is known to be contaminated with ^{14}C, and Z7 has a poor noise response. The efficiencies below will be quoted for the 'Good Ge detectors', by which we mean (Z2, Z3, Z5, Z9, Z11 = 1.25kg), and for 'Good Si detectors', referring to (Z4, Z8, Z10, Z12 = 0.4kg).

Where efficiency values are quoted and shown in Figures, the efficiency is defined as the number of events passing the cut divided by the total number of events present prior to application of the cut. Error bars shown are based on binomial statistics of the cut passing fraction and do not include any systematics that may be present.

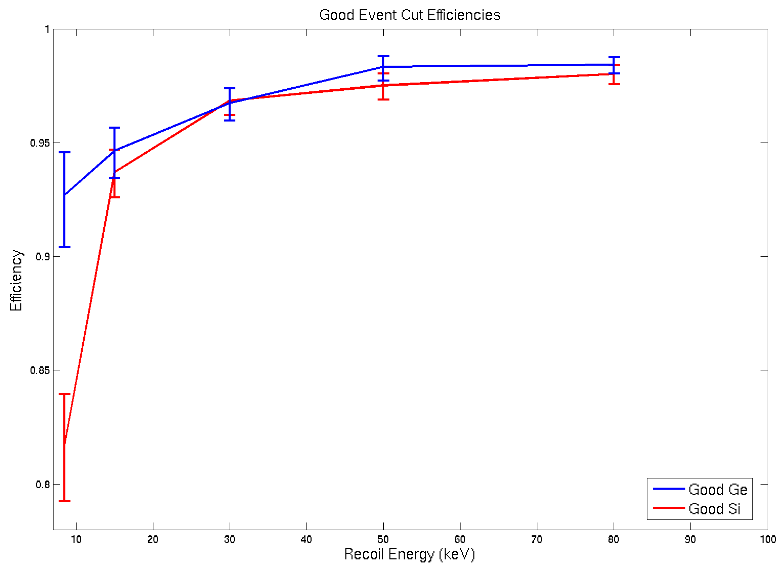


Figure 6.28: Efficiency of the Good Event Cut (= Bad data & Charge χ^2 & phonon pretrigger cuts) for WIMP search events occurring in good Ge (Z2, Z3, Z5, Z9, Z11) and good Si (Z4, Z8, Z10, Z12) detectors.

6.6.1 Bad Data, Charge χ^2 and Phonon Pretrigger Cut Efficiencies

The total WIMP search data livetime for Run 119 is 77.44 days. Application of the Bad Data Cut (see Section 6.5.1) results in a net livetime of 77.28 days.

Application of the remaining Good Event cuts (charge χ^2 and phonon pretrigger, Sections 6.5.2-6.5.3) to the WIMP search data results in 61.84 livedays (77.30 kg-days) on the good Ge detectors, and 68.03 livedays (27.21 kg-days) on the good Si detectors. The overall efficiency of the Good Event Cut upon the WIMP search data as a function of energy is shown in Figure 6.28. The efficiency drops with decreasing recoil energy since the cut includes the charge χ^2 and phonon pretrigger cuts, which are sensitive to the signal to noise ratio of the ionization and phonon pulses.

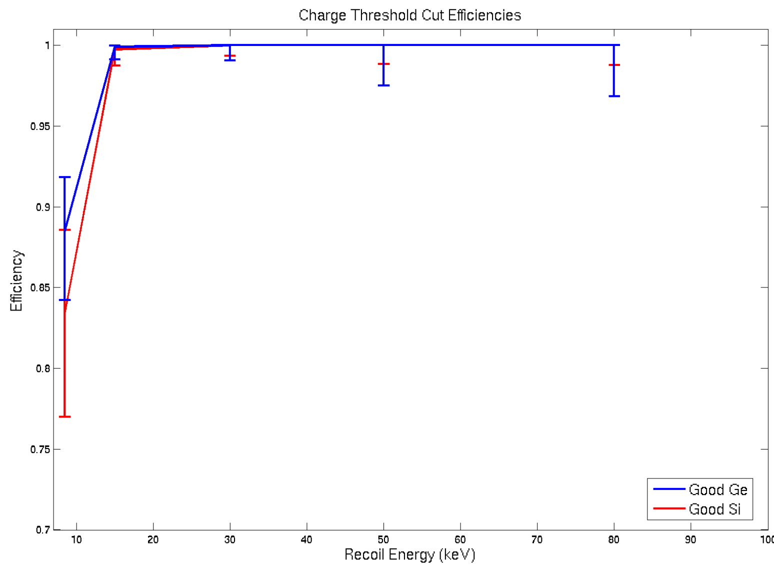


Figure 6.29: Efficiency of the Charge Threshold Cut for good Ge and good Si detectors. Events are preselected to pass the Good Event Cut and be located within the 2σ nuclear recoil band.

6.6.2 Charge Threshold and Inner Electrode Cut Efficiencies

The efficiencies of the cQThresh and cQin cuts when applied to data with cGoodEv already applied are shown in Figures 6.29 and 6.30 respectively.

6.6.3 Muon Veto Cut Efficiency

The efficiency of the muon veto cut for passing WIMPs can be inferred from the spectrum of events above threshold in the veto. As Figure 6.19 shows, the timing of veto hits above threshold follows a falling exponential as predicted by Poisson statistics. Therefore, the efficiency of the veto at cutting WIMP events is simply the probability of having a veto hit falling by chance in the $50\mu s$ window prior to an event's global trigger.

The efficiency of the veto on passing WIMP events is then calculated to be 0.9697 ± 0.0001.

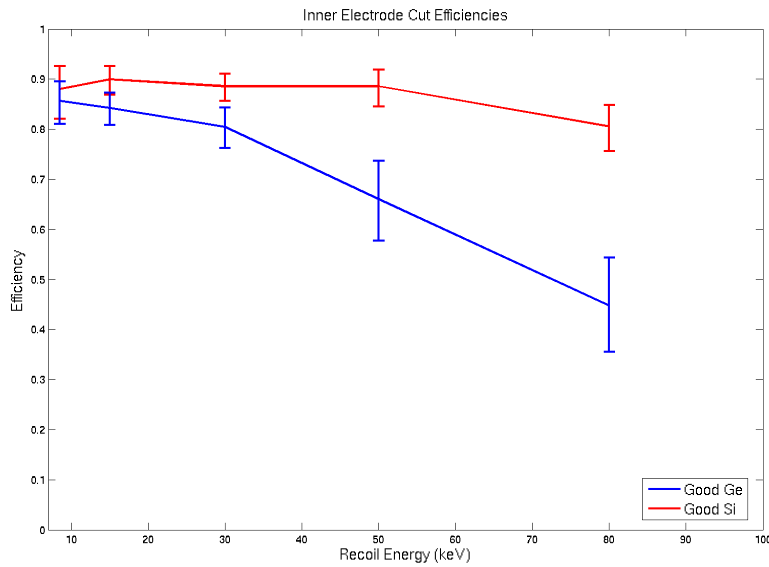


Figure 6.30: Efficiency of the Inner Electrode Cut for good Ge and good Si detectors. Events are preselected to pass the Good Event Cut and be located within the 2σ nuclear recoil band.

6.6.4 Nuclear Recoil Band Cut Efficiency

The neutron calibration also contains gamma events from the ^{252}Cf source. Therefore to get a more accurate estimation of the nuclear recoil band's efficiency upon neutrons, events in the electron recoil band are excluded from the calculation.

The efficiency of the 2σ nuclear recoil band at passing events from the ^{252}Cf data that lay outside the 3σ electron recoil band is shown in Figure 6.31. The efficiency on the neutrons is consistent with the 95% passing fraction that we would *a priori* expect from a 2σ selection.

6.6.5 Phonon Timing Cut Efficiency

Efficiency of the timing cut on nuclear recoil events is shown in Figure 6.32. An ideal timing cut would obviously remove all surface electron recoil events whilst keeping all nuclear

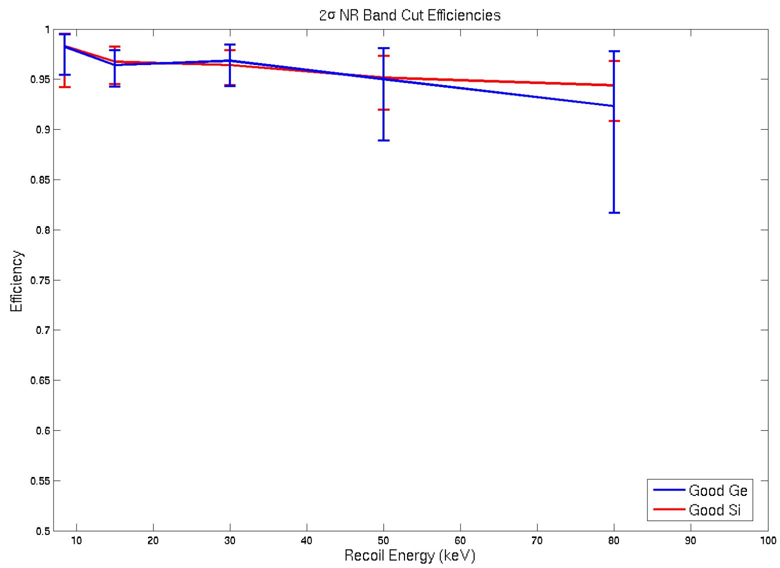


Figure 6.31: Efficiency of 2σ nuclear recoil band for events in ^{252}Cf calibration data. Events are preselected to pass the Good Event Cut and be outside the 3σ electron recoil band (see text).

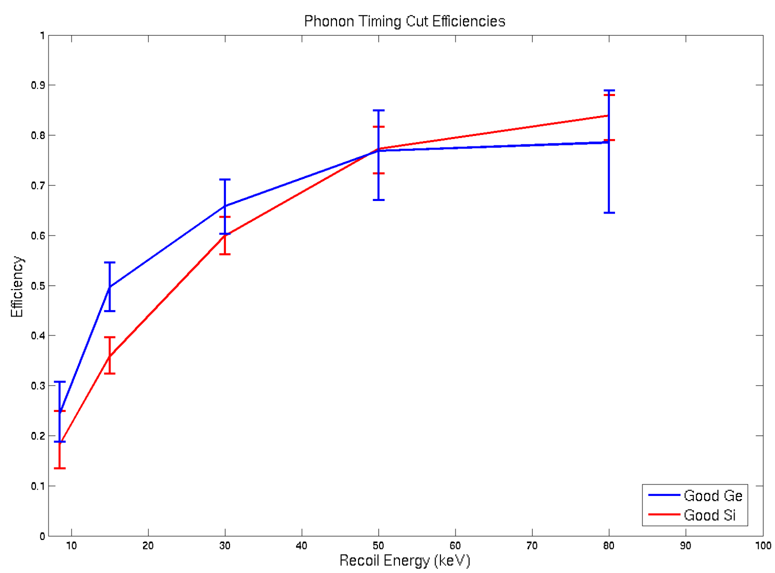


Figure 6.32: Efficiency of the phonon timing cut on 2σ NR events in the ^{252}Cf calibration data. Events are preselected to pass the Good Event Cut, the charge threshold cut, the inner electrode cut and the veto anticoincidence cut.

recoils, and hence this cut is one of the few pieces of the analysis for which it is, at least in principle, possible to improve our net sensitivity to WIMPs via improving the cut analysis. Again, see Chapter 9 for one such attempt to improve the timing cut efficiency beyond that discussed here for the CDMS primary analysis.

The efficiency of the cut on nuclear recoils is not of course, the whole story. The expected number of surface electron recoils that will 'leak' past the cut is also of paramount importance. The number of these 'leakage' events in the WIMP search data can be estimated by knowing the relative number of low yield events that appear in the ^{133}Ba data to the number of low yield events in the WIMP search data. Since our WIMP search data is blinded (potential WIMP events are removed from the data), we can do this without biasing ourselves. However, one potential problem lies in the presence of systematic differences between the low yield events in the calibration data and in the WIMP search data. More on this can be found in Section 6.7.

6.6.6 Overall Efficiency

The overall efficiency represents the combination of (Good Event Efficiency & Trigger Efficiency & Inner Electrode Efficiency & Charge Threshold Efficiency & Muon Anticoincidence Efficiency & Nuclear Recoil Band Efficiency & Phonon Timing Cut Efficiency) applied to 2σ Nuclear Recoil events. This combined efficiency is shown in Figure 6.33.

The combined efficiency, along with the expected number of leakage events for a given energy range will inform our choice of analysis threshold. The analysis threshold cut is simply a binary cutoff below a particular energy. Our overall efficiency will drop to zero at some energy, but our aim is to maximize the expected sensitivity to WIMPs. Hence we

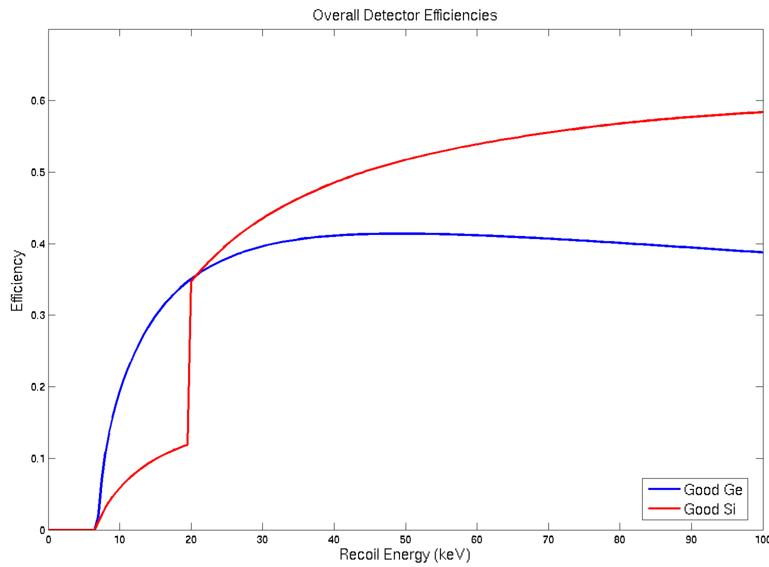


Figure 6.33: Overall detector efficiencies for good Ge and good Si detectors. The discontinuity in the Si efficiency is due to the combination of detectors with different analysis thresholds.

may wish to set the analysis threshold higher if the expected number of leakage events in a particular detector is high with a low analysis threshold energy.

We set our analysis threshold to 7 keV in all Ge detectors, which is where the overall efficiency drops to zero. Z4, Z8 and Z10 had finite efficiencies down to 7 keV, but also have higher expected leakages with a 7 keV analysis threshold. The threshold is therefore set to 20 keV in Z4, Z8 and Z10, and 7 keV in Z12, resulting in the observed discontinuity in the Si combined efficiency distribution.

The combined efficiencies yield a final spectrum-averaged WIMP exposure of 34 kg-days on Germanium and 14 kg-days on Silicon. By spectrum-averaged exposure, we mean the exposure that would be needed to expect the same number of WIMPs if the detectors were 100% efficient. Literally, it is the convolution of the efficiency with the expected WIMP spectrum, times the raw exposure.

6.7 Overall Rates and Expected Background

6.7.1 Gamma Rates

The mean gamma rate is calculated from WIMP search data to be 135.7 ± 2.9 evts/kg/day in the good Ge detectors in the energy range 20-60 keV. However, this includes a period of around 9 days where the old air purge to the icebox was functioning at a decreased level, and the rise in environmental Rn inside the shield increased the background rates by a factor 2-3 during this period. A fairer estimate is therefore given by the 68.35 livedays for which the old air purge was functioning correctly, which results in a mean rate of 108.8 ± 2.8 evts/kg/day in good Ge ZIPs, and 450.0 ± 8.9 evts/kg/day in good Si ZIPs, for recoil energies in the range 20-60 keV. The energy range is chosen this way in order to not count events that come from the cosmogenic activation of Ge that results in lines at 10.4 keV and 66.7 keV.

GEANT Monte Carlo studies of the gamma background suggest that the majority of the gamma background ($\sim$ 75%) comes from ^{222}Rn chain decays from outside the mu-metal shield due to an imperfect air purge. The remaining gammas most likely come from U/Th/K chain decays from the copper cans and/or the inner polyethylene [90].

6.7.2 Beta Rates

Using the 'wide beta' definition ($0.1 < y <$ lower bound of 5σ electron recoil band), the mean beta rates in good Ge and Si detectors are 2.55 ± 0.21 evts/kg/day and 3.27 ± 0.18 evts/kg/day for recoil energies in the range 4-100keV. As a subset of these values, the mean rates of single-scatter betas are 0.48 ± 0.09 evts/kg/day and 0.42 ± 0.09 evts/kg/day in good Ge and Si detectors respectively, 4-100 keV. These rates are calculated for the 68.35 livedays for which the old air purge was functioning correctly.

By examining the gamma and beta rates during the periods with the old air purge active and inactive, it can be estimated that 40 - 50% of the betas in the WIMP search data come from ejectrons, of which $\sim$ 80% multiply scatter. 10 - 40% of the beta background is thought to come directly from ^{210}Pb decays, where the ^{210}Pb is distributed across the Cu cans and the towers themselves via plating from Radon. More Monte Carlo simulations of ^{210}Pb contamination is needed to improve these estimates, and to compare directly with the energy distribution of background beta events.

6.7.3 Alpha Rates

Alpha particles are identified in the analysis as events with $y < 0.5$ and $E_r > 1500\,\mathrm{keV}$. The WIMP search data has an observed mean alpha rate of 1.24 ± 0.24 evts/kg/day in the good detectors. Since it is expected that the overwhelming majority of these events should come from ^{210}Pb decays, the expected number of beta decays from this contamination can be calculated, and was used in part to estimate the beta background from ^{210}Pb described in the previous section.

6.7.4 Neutron Rates

One veto-coincident neutron was observed in the entire WIMP search data; this event multiply-scattered between Z11 and Z12 and was coincident with a muon in the veto system. It is expected that with 2 towers of detectors a neutron will multiply-scatter $\sim$ 20% of the time. The time between the muon veto event and the first detector event was less than 1μs.

No veto-anticoincident neutron events were observed in Run 119.

The background due to single-scatter muon veto anticoincident neutrons is estimated to be 0.06 events in Ge and 0.05 events in Si, for all WIMP search Run 119 data [99].

Several neutron Monte Carlo simulations have been performed using both GEANT and FLUKA Monte Carlo simulation packages. The different simulations predict that Run 118 and 119 combined should have seen between 5-10 veto-coincident neutrons, where the greatest discrepancy occurs between the two simulation packages. Yet, in the combined data sets only one anticoincident neutron was observed. The simulations are still being refined, so the number of observed neutrons in future runs will have to be weighed against future predictions.

6.7.5 Expected Background after Cuts

The expected background is the number of WIMP-like events (that are not WIMPs) leaking into the signal region that remain after cuts are applied. The expected background leakage due to gammas is negligible based upon the ^{133}Ba gamma calibration data. The cuts remove $> 99.99\%$ of electron recoil band events, and the recoil bands for all detectors are all well separated above the analysis thresholds.

The expected background due to low-yield (nuclear recoil band) surface events that pass the phonon timing cut is calculated by considering the efficiency of the timing cut upon beta events. The gamma calibration data tells us, for a given number of betas (using the 'wide beta' definition), what fraction we should expect to see located in the nuclear recoil band. Since we also know the relative number of betas in the calibration data and WIMP search data sets, we can scale the expected number of leakage events that pass the timing cut in the calibration data to an estimate of the expected leakage in the WIMP search data. We can also use the single/multiple ratio in the calibration data as a cross-check on the expected number of single-scatter leakage events, since multiple scatter events are not removed from the WIMP search data in the blinding cut.

This assumes no systematic difference between the betas in the calibration and WIMP search data, which is not the case. Systematic differences arise due to 1) different sources of low yield events (eg. much larger fraction are ejectrons in calibration data), and 2) Due to the calibration source location, the phonon sensor sides of the detector are predominantly 'lit' by the source, which results in a systematic effect due to the differing dead layer effects on the phonon and charge sensor sides (see Chapter 5).

The expected background due to betas from local contamination sources described in Section 6.7.2 is calculated to be 0.5 ± 0.2(stat) $\pm\, 0.2$(sys) in Ge, 7-100keV; 1.2 ± 0.6(stat) $\pm$ 0.2(sys) in Si, 7-100keV [99].

6.8 Results from Unblinding

The presence of the blinding scheme, which removed any potential WIMP-like events from the open analysis, allowed the estimation of the above expected backgrounds and cut efficiencies in an unbiased way. The cuts were all frozen, with efficiencies and expected leakage calculated before the 'box was opened'; that is, the blinding protocols were removed and all events were included in the final analysis.

All unblinded WIMP search data is shown in Figures 6.34 and 6.35, with the following cuts applied: (Good Event) & (Charge Threshold) & (Inner Electrode) & (Muon Veto

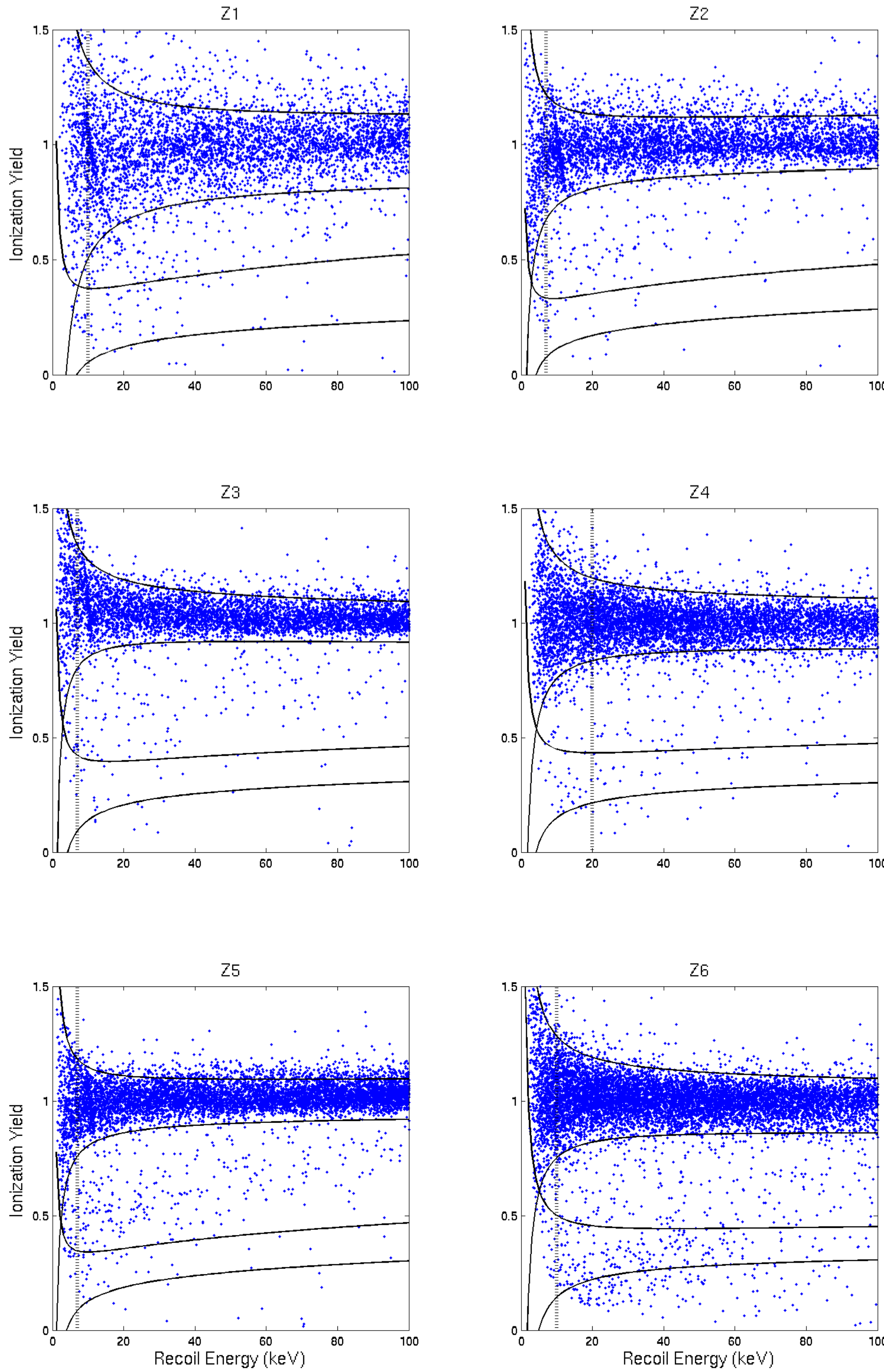


Figure 6.34: Unblinded WIMP search data in Tower 1 detectors with all cuts except for the phonon timing and singles cuts applied. 2σ recoil bands are shown as black lines, and each detector's analysis threshold is shown as a vertical dotted line.

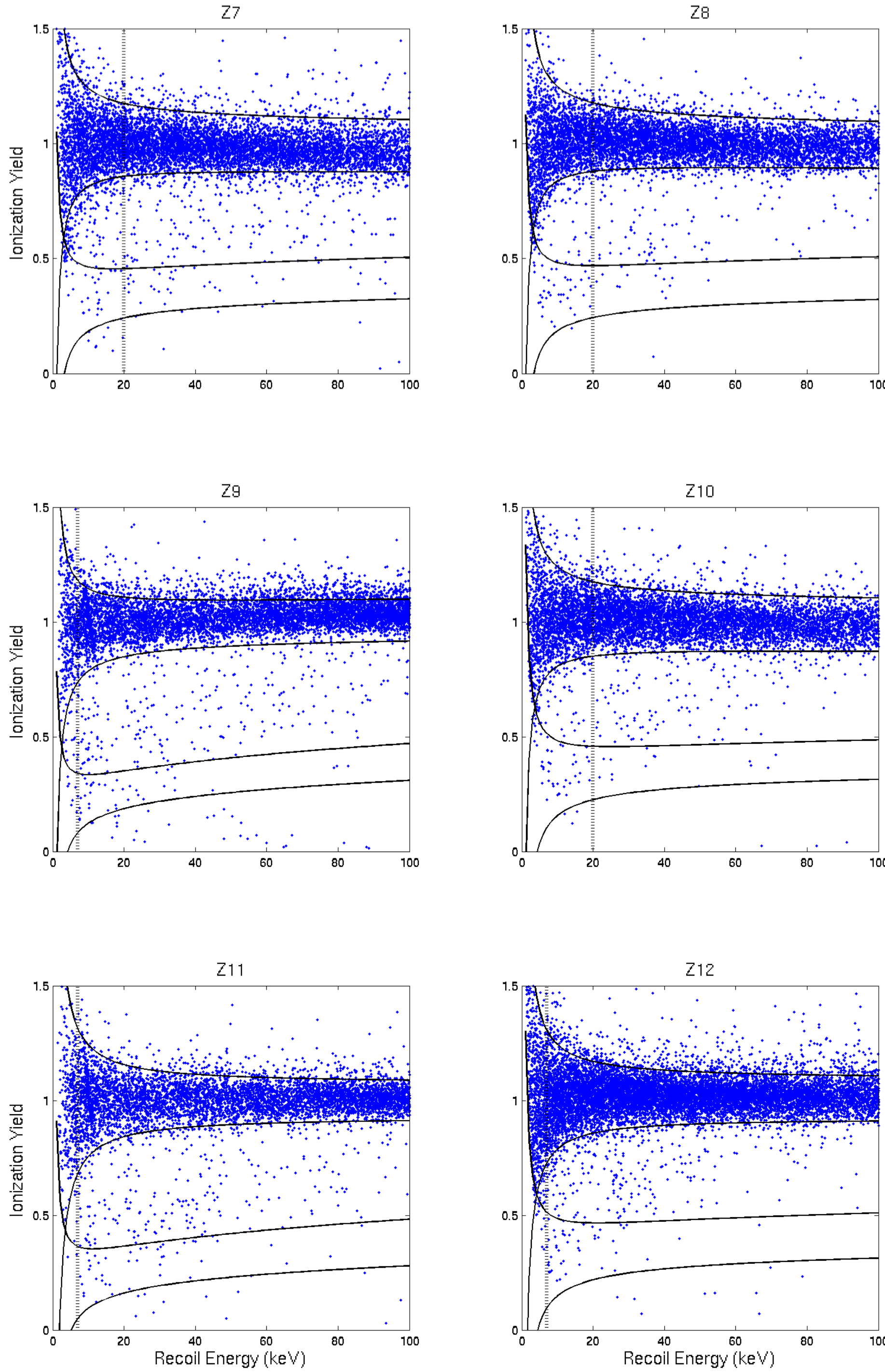


Figure 6.35: Unblinded WIMP search data in Tower 2 detectors with all cuts except for the phonon timing and singles cuts applied. 2σ recoil bands are shown as black lines, and each detector's analysis threshold is shown as a vertical dotted line.

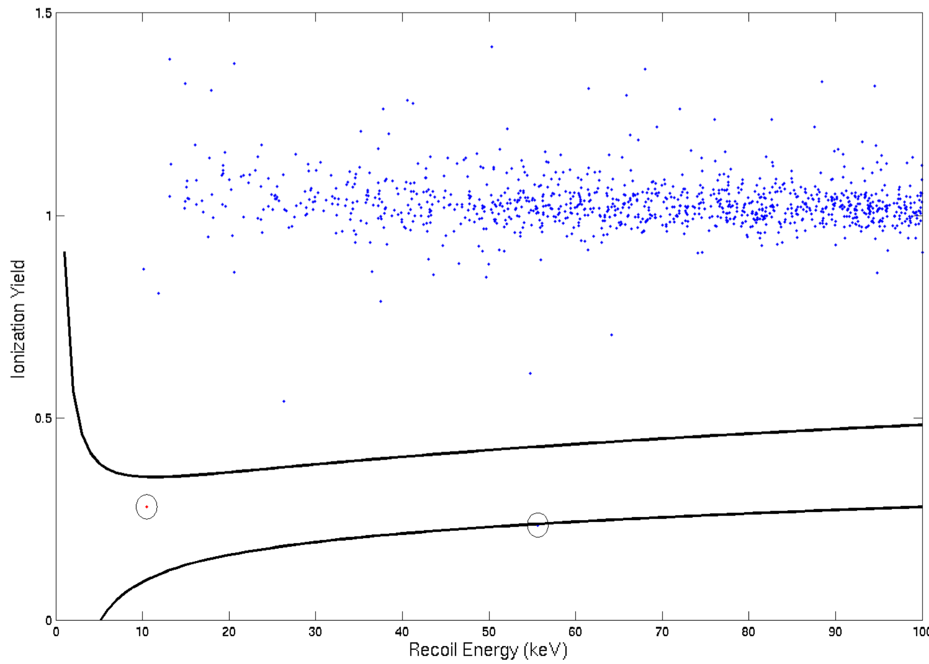


Figure 6.36: Unblinded WIMP search data in Z2/3/5/9/11 with all cuts, including the singles cut, applied. The red circled event at 10.15keV is the only event that passes all cuts and appears in the 2σ nuclear recoil band (Z11 NR band shown as black lines). Another 'near-miss' event appears (circled) at 56keV and lies within the 3σ nuclear recoil band.

Anticoincident) & (Z11 Cliff Cut).

The number of events in each good detector passing successive cuts is given in Table 6.3.

When we apply the analysis to unblinded Run 119 WIMP search data, a single event remains in the 2σ nuclear recoil band. i.e. this event passes all cuts and therefore is a candidate for a WIMP event. The event, shown in Figure 6.36 occurred in detector Z11 with a recoil energy of 10.5 keV. Further analysis of the event reveals that Z11 exhibited some anomalous behavior during the data series in which this event occurred, which was not flagged by the original set of blind cuts.

Figure 6.37 shows ionization yield vs. event time for all events passing cGoodEv in

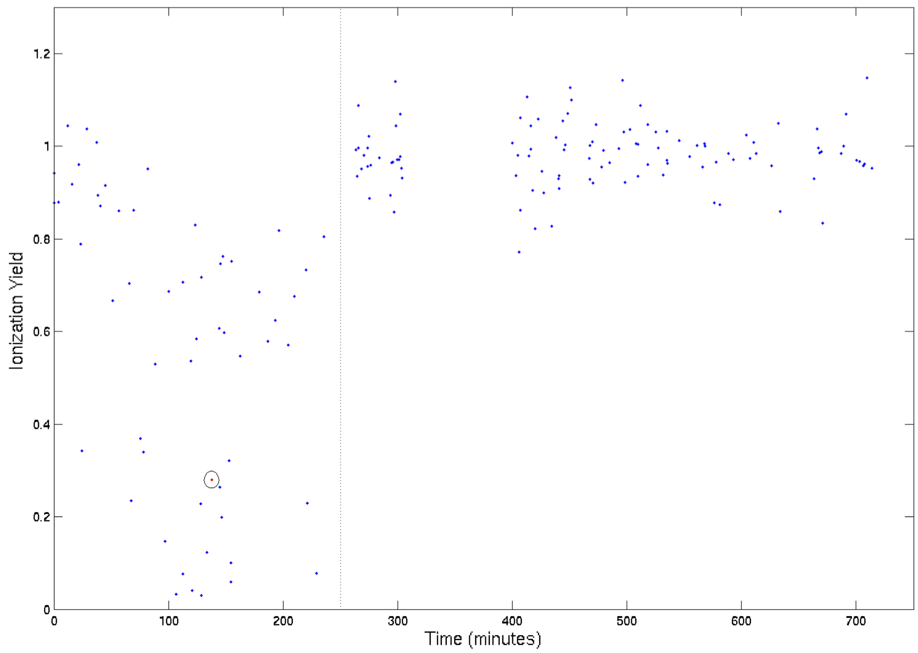


Figure 6.37: Ionization yield as a function of event time for Z11 good events in the data series which contains the candidate event (in red, circled). The drooping of the electron recoil band over time is characteristic of detector neutralization loss; the recoil band returns to its normal position when the regularly scheduled detector flashing is performed towards the end of the run (dotted line). The subsequent data run is shown to the right - the time between the datasets has been reduced for clarity.

Z11 for the (310 minute) dataset containing the candidate event. It can be seen that the yield of the events begin to 'droop' as time goes on, which is characteristic of a detector losing neutralization. It seems likely that this event only appears in the nuclear recoil band because its ionization yield is artificially reduced due to lack of neutralization in Z11.The presence of this event, however, is included in our calculation of the dark matter limit in order to maintain the integrity of the blind technique. It is not known why this problem occurred but it was not repeated in any other parts of Run 118 or 119.

Figure 6.36 also shows a 'near miss' event just outside the nuclear recoil band at $\sim$ 56 keV. This event is not considered a WIMP candidate since our bands were defined whilst blind and this event, although close, does not appear in the 2σ nuclear recoil band.

A single event is consistent with our background leakage estimate and cannot be considered statistically significant evidence of a WIMP.

6.9 Dark Matter Limits

The Run 119 result can be used to calculate three different dark matter limits; 1) Upper limit on spin-independent WIMP-nucleon cross section (as a function of WIMP mass) for both Ge and Si, 2) Upper limit on spin-dependent WIMP-neutron cross section on Ge and Si, 3) Upper limit on spin-dependent WIMP-proton cross section on Ge and Si.

Spin-Independent Limits

Calculation of the spin-independent Ge and Si WIMP recoil spectra follows the method of Lewin and Smith [53] using standard assumptions for the galactic halo (WIMP characteristic velocity $v_0 = 220\,\mathrm{km\,s^{-1}}$, mean Earth velocity $v_E = 232\,\mathrm{km\,s^{-1}}$, $\rho = 0.3\,\mathrm{GeV\,c^{-2}\,cm^{-3}}$). In addition, we use the Optimal Interval method developed by S. Yellin [91], which sets a limit on the signal cross-section by examining the interval in recoil energy that is least contaminated by background. This method reduces to the Poisson limit in the case of no observed events.

The spin-independent upper (90%) limits on the WIMP-nucleon cross-section obtained from Run 119 are shown in Figure 6.38. The combined R118 + R119 Ge data sets a limit on the WIMP cross-section that is a factor ~ 10 better than any other direct WIMP-detection experiment [99].

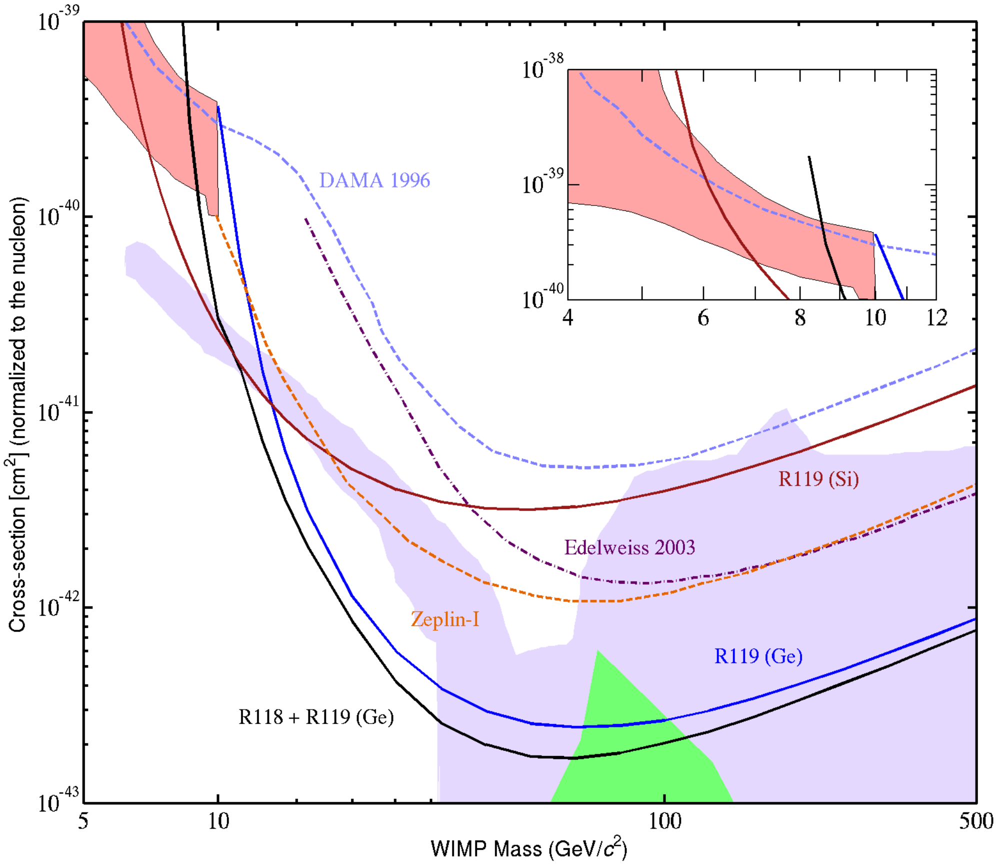


Figure 6.38: Spin independent 90% upper limits on WIMP-nucleon cross-section obtained from Run 119 in Ge (solid blue) and Si (solid red). Also shown are the combined Ge limit from Run 118 and Run 119 combined (solid black) and limits from Zeplin-I (dashed salmon) [92], Edelweiss (dashed dk. red) [93] and DAMA (dashed blue) [94]. The shaded low mass region shown in more detail at the upper-right is from DAMA [95]. The two block regions are predictions on supersymmetry taken from [96] (lavender) and [97] (green). Figure generated by [98].

Spin-Dependent Limits

Spin-dependent interactions only occur where the target nuclei have an odd number of nucleons (and hence have nonzero nuclear spin). Although both Ge and Si are predominantly made from even-A isotopes, both contain an isotope with nonzero nuclear spin: ^{73}Ge (spin-9/2) makes up 7.73% of natural Ge, and ^{29}Si makes up 4.68% of natural Si. Since both of these isotopes contain a single unpaired neutron, this makes CDMS much more sensitive to spin-dependent interactions with neutrons than with protons [63].

The spin-dependent limits obtained from a combination of Run 118 + Run 119 data on both Ge and Si are shown in Figures 6.39 and 6.40 [63]. Again, standard halo model assumptions are made (WIMP characteristic velocity $v_0 = 220\,\mathrm{km\,s^{-1}}$, mean Earth velocity $v_E = 232\,\mathrm{km\,s^{-1}}$, $\rho = 0.3\,\mathrm{GeV\,c^{-2}\,cm^{-3}}$).

6.10 The Future

The next CDMS run at Soudan, using 30 ZIP detectors - 19 Ge (4.75kg) and 11 Si (1.1kg) - will begin running in summer 2006, and is expected to have ~ 4 times the livetime of Run 119. Thus the limits set on this data are expected to be ~ 15 times greater than those set for Run 119, based on scaling the mass and exposure.

It is expected that the overall efficiency of the analysis will improve by the time the subsequent run's data is in hand. The efficiency of the phonon timing (surface event) cut has a significant effect on the final efficiencies, and hence one of the improved timing analyses should be able to provide significant improvement in this area (see Chapter 9 for discussion of one possible approach).

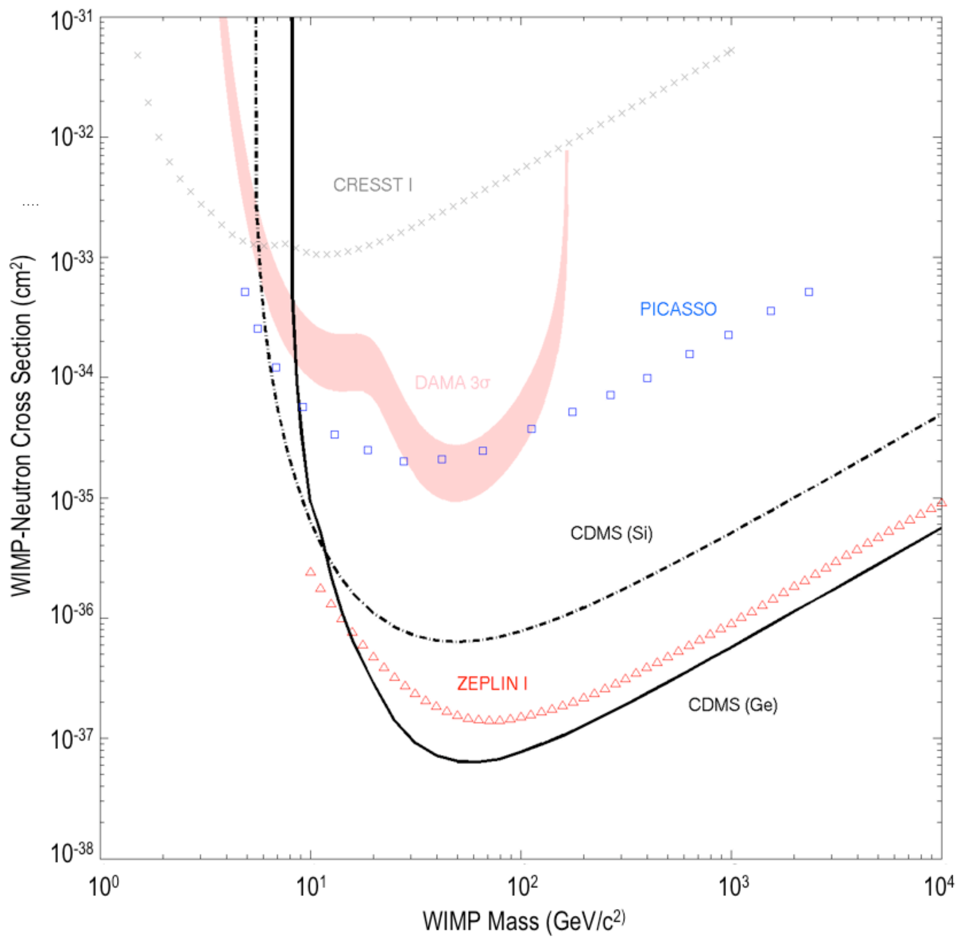


Figure 6.39: Spin dependent 90% upper limits on WIMP-neutron cross-section. CDMS limits for R118 and R119 are shown for Ge (solid black) and Si (dash-dot black) [63]. Also shown are the DAMA/NaI annual modulation signal (red filled region) [100], CRESST I (grey crosses) [101], PICASSO (blue squares) [102] and ZEPLIN I (red triangles) [103]. Figure generated by [98].

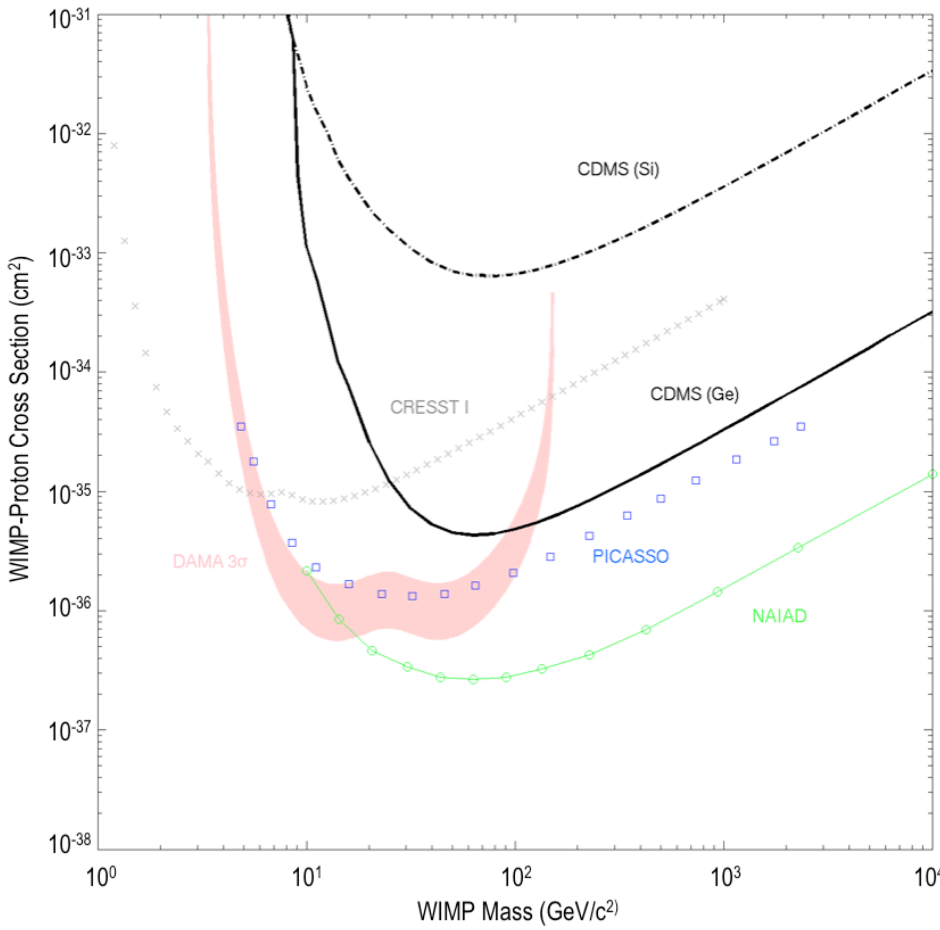


Figure 6.40: Spin dependent 90% upper limits on WIMP-proton cross-section. CDMS limits for R118 and R119 are shown for Ge (solid black) and Si (dash-dot black) [63]. Also shown are the DAMA/NaI annual modulation signal (red filled region) [100], CRESST I (grey crosses) [101], PICASSO (blue squares) [102], ZEPLIN I (red triangles) [103], and NAIAD (green circles) [104]. Figure generated by [98].

Preselection Cuts	Number of events in WIMP search data
Above ionization threshold in specified detector	37k/41k/54k/45k/57k (Ge) - 25k/24k/26k/33k (Si)
+ Good events, inner electrode, 7-100 keV recoil energy	4668/5281/7483/7412/7969 (Ge) - 6773/7297/7627/11037 (Si)
+ Anticoicident with veto	4497/5071/7230/7141/7727 (Ge) - 6534/7051/7360/10666 (Si)
+ Phonon timing cut	888/470/208/1022/1515 (Ge) - 75/2445/74/625 (Si)
+ Single scatter	212/111/77/314/575 (Ge) - 9/744/18/268 (Si)
+ 2σ nuclear recoil	0/0/0/0/1 (Ge) - 0/0/0/0 (Si)

Table 6.3: Events numbers passing successive analysis cuts in the WIMP search data, for detectors Z2/3/5/9/11(Ge) and Z4/6/8/10 (Si). Cuts listed on each row are cumulative, i.e. the numbers given are for events passing all cuts up to and including that row.

Chapter 7

Wavelet Technology

Analyzing signals using Fourier theory gives no direct indications as to when transients occur in a signal since the information is encoded in both the phase and amplitude coefficients over a wide range of frequencies. One can window the signal in the time domain and analyze each section separately, but information is limited by the size of the window chosen. i.e. a small window gives good timing information, but is not able to resolve low frequency components, whereas a large window resolves low frequency components but is unable to resolve small time structure.

The wavelet transform is a solution to these problems. It uses a scalable function, shifted along the signal in the time domain, and for every position the function is compared to the signal. A level of correlation is determined, which is recorded as a coefficient. This process is repeated many times with a differently scaled function, producing a set of time-frequency representations of the signal, all with different resolutions.

Since, in analyzing CDMS detector signals, we are particularly interested in time-frequency information associated with the leading edge of the pulse, the wavelet approach

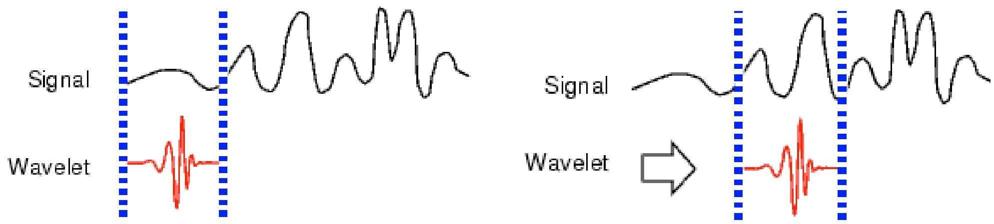


Figure 7.1: The Continuous Wavelet Transform (CWT) analyses a signal using all translations and scalings of the mother wavelet.

is a natural way to proceed.

7.1 Continuous Wavelet Transformation

The wavelet transform described above is known as the continuous wavelet transformation, or CWT. It can be written as:

$$\gamma(s, \tau) = \int f(t)\, \Psi^{*}_{s,\tau}(t)\, dt \tag{7.1}$$

where $\gamma(s, \tau)$ (the 'wavelet coefficient') represents the level of correlation between the function being analyzed $f(t)$ and the wavelet $\Psi^{*}_{s,\tau}(t)$. The variables s and τ represent the scale and translation of the wavelet respectively, as derived from the 'mother wavelet' described by:

$$\Psi_{s,\tau}(t) = \frac{1}{\sqrt{s}}\, \Psi\left(\frac{t-\tau}{s}\right) \tag{7.2}$$

Note that at this point we haven't specified what $\Psi(t)$ is. The theory specifies only general properties, and it is left to the problem at hand to determine which particular family of wavelets is appropriate. $\Psi(t)$ appears as a complex conjugate in 7.1 by convention, i.e. this could be changed by a suitable redefinition of $\Psi(t)$.

7.2 Wavelet Properties

All wavelet functions must satisfy a set of conditions to be considered a suitable wavelet. Firstly, it can be shown that if a square integrable function $\Psi(t)$ satisfies the *admissibility condition*:

$$\int \frac{|\tilde{\Psi}(\omega)|^2}{|\omega|}\, d\omega < +\infty \tag{7.3}$$

where $\tilde{\Psi}(\omega)$ is the Fourier transform of $\Psi(t)$, then it can be used to analyze and then reconstruct a signal without loss of information [107]. The admissibility condition also implies that:

$$|\tilde{\Psi}(\omega)|^2\Big|_{\omega=0} = 0 \tag{7.4}$$

And hence a zero at the zero frequency means that the time-averaged value of the wavelet must be zero:

$$\int \Psi(t)\, dt = 0 \tag{7.5}$$

The definition of (7.1) shows that the wavelet transform of a one-dimensional function will be two-dimensional; the wavelet transform of a two-dimensional function will be four-dimensional, etc. Since this is an undesirable property, one imposes additional constraints upon the wavelet functions in order to make the wavelet transform decrease quickly with decreasing scale s.

These conditions, known as the *regularity conditions* essentially state that the wavelet function should be concentrated in both time and frequency domains (‘compactly supported’) [107]. The concept of regularity is a complex one[1], but as an overview consider the

[1]For more information on the regularity of wavelets, see [109] and [111].

moments of the wavelet, defined by

$$M_p = \int t^p \, \Psi(t) \, dt \tag{7.6}$$

It can be shown that if the wavelet transform (7.1) is Taylor expanded in s about $t = 0$, then the moment M_n is the coefficient of the $(n+1)$th power of the scale s. i.e.

$$\gamma(s, 0) = \frac{1}{\sqrt{s}} \left[f(0) M_0 s + \frac{f^{(1)}(0)}{1!} M_1 s^2 + \frac{f^{(2)}(0)}{2!} M_2 s^3 + ... + \frac{f^{(n)}(0)}{n!} M_n s^{n+1} + O(s^{n+2}) \right] \tag{7.7}$$

where $f^{(p)}$ is the pth time derivative of $f(t)$.

Using (7.5), we know that $M_0 = 0$, so if we can manage to make the other moments up to M_n zero (or very small) as well, then the wavelet transform coefficients $\gamma(s, \tau)$ will decay as fast as s^{n+2} for a smooth signal $f(t)$. Study of these *vanishing moments* suggests that the ideal number of them present in the wavelet will depend heavily on the application [108].

So to summarize, the admissibility condition specifies that Ψ must be a wave, and the regularity conditions and vanishing moments specify that it must be compactly supported. The result is a localized wave, or a wavelet.

7.3 The Discrete Wavelet Transformation

The wavelet transformation as it stands is somewhat impractical. The continuous transformation requires the calculation of coefficients over every possible scale by continuously shifting a continuously scalable function over a signal, and these scaled functions will not necessarily form an orthonormal basis [107].

Even without this redundancy, there are still an infinite number of wavelets in the wavelet transformation and in order to be able to compute wavelet coefficients we would like to reduce this number. It is also clear that the wavelet transformation will not, in general, have an analytic solution, requiring the resulting algorithms to be calculable given a reasonable amount of computer power.

To overcome the problem of redundancy, discrete wavelets are introduced. These are just like regular wavelets, except they can only be scaled and translated by discrete steps:

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}}\,\Psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j}\right) \tag{7.8}$$

where j and k are integers, and the usual choice for the scaling and translation factors are $s_0 = 2$ ('dyadic sampling'), and $\tau_0 = 1$, so the discrete wavelet definition becomes:

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\,\Psi\left(\frac{t - (2^j)k}{2^j}\right) \tag{7.9}$$

Thus the signal is now only sampled at discrete intervals in time. However, it can be shown that such a decomposition still allows for perfect reconstruction of the original signal, so long as the total energy of the wavelet coefficients lies between two positive bounds A and B [109]:

$$A\|f(t)\|^2 \leq \sum_{j,k} |\langle f(t), \Psi_{j,k}\rangle|^2 \leq B\|f(t)\|^2 \tag{7.10}$$

where $\|f(t)\|^2 \equiv \int_0^\infty |f(t)|^2\,\mathrm{d}t$ is the energy of $f(t)$, $A > 0$, $B < \infty$ and A, B are independent of $f(t)$. When this equation is satisfied, the family of basis functions $\Psi_{j,k}(t)$ with $j, k \in \mathbb{Z}$ is referred to as a 'frame' with frame bounds A and B.

When $A = B$ the frame is *tight*, and the discrete wavelets form an orthonormal basis. i.e.

$$\int \Psi_{j,k}(t)\Psi^*_{m,n}(t)dt = \delta_{jm}\delta_{kn} \tag{7.11}$$

Hence the original signal can be reconstructed by summing the orthogonal wavelet basis functions, weighted by the wavelet coefficients:

$$f(t) = \sum_{j,k} \gamma(j,k)\Psi_{j,k}(t) \tag{7.12}$$

When $A \neq B$, the frame is *dual*, and although reconstruction of the original signal is still possible, the decomposition wavelet is different from the reconstruction wavelet. Orthogonality, however, is not an essential quality in the representation of signals, and it is sometimes useful to use a collection of wavelets that do not obey the orthogonality relation.

One of the disadvantages of the discrete wavelet transform is that that it is not shift invariant. That is, if one takes the discrete wavelet transform of a signal and the discrete wavelet transform of the same signal shifted in time, the results will not in general be the same. The results will also not, in general, be shifted versions of one another.

7.4 The Scaling Function

The discrete wavelet transform still requires a very large number of scalings and translations in order to be calculated for a sampled signal. The key to reducing the number of wavelets is to consider the wavelet transform as a type of filter. That is, a wavelet transform at a particular scale produces information over a limited range of frequencies, and in effect is acting like a band-pass filter.

Fourier theory tells us that scaling the wavelet by a factor of two will stretch the frequency spectrum of the wavelet by a factor of two, and will also shift the frequency components up by a factor of two. Hence, we can cover the whole frequency range of the signal by using a set of dyadically dilated wavelets.

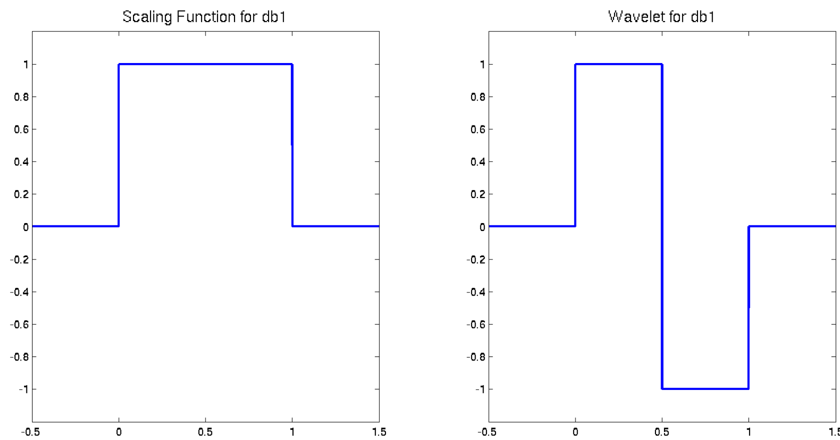


Figure 7.2: Wavelet and Scaling Functions for the db1 (also known as the haar) wavelet, which belongs to the Daubechies family of wavelets and has 1 vanishing moment. The functions are shown at an arbitrary scale.

However, in order to cover all frequencies down to zero, an infinite number of dilated wavelets is still needed, because each time the wavelet is dilated it's bandwidth is reduced. However one can cover the frequency range below some minimum value by use of a *scaling function*. This method, first introduced by Mallat [110], allows the analysis of a signal by using a finite number of wavelets, and a scaling function that covers the (low) frequency space not covered by the wavelets. i.e. the scaling function is essentially a sum of the (infinite number of) wavelets that encompass the low-frequency part of the spectrum:

$$\phi(t) = \sum_{j,k}^{n} \gamma(j,k)\Psi_{j,k}(t) \tag{7.13}$$

i.e. The frequency space is covered by the scaling function up to some scale n. The scaling function has the condition

$$\int \phi(t)\, dt = 1 \tag{7.14}$$

Figure 7.2 shows the wavelet and scaling functions for the Daubechies wavelet with 1 vanishing moment, also called *db1* or equivalently the *haar* wavelet. Figure 7.3 shows the

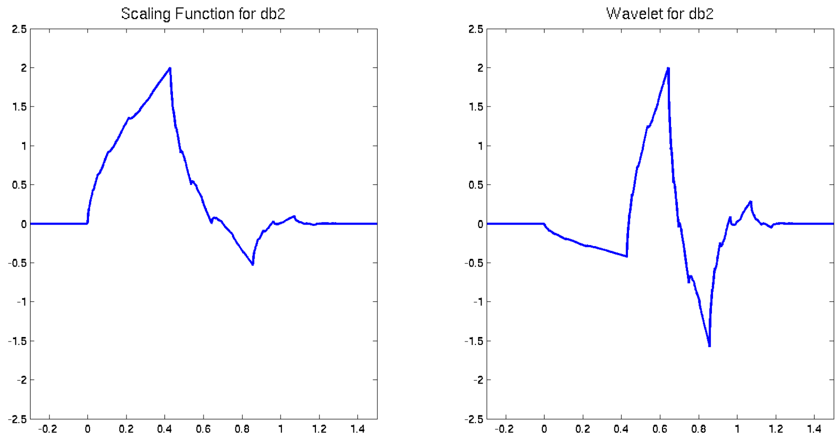


Figure 7.3: Wavelet and Scaling Functions for the db2 wavelet, which belongs to the Daubechies family of wavelets and has 2 vanishing moments. The functions are shown at an arbitrary scale.

wavelet and scaling functions for the Daubechies wavelet with 2 vanishing moments, also called *db2*.

7.5 Band-Pass Filters

One way to implement the multiresolution analysis that follows from the wavelet and scaling function is to view these functions as a filter bank. In this way, the signal is passed through a series of filters and the outputs of the different filter stages are the wavelet and scaling function transform coefficients.

That is, we can express the signal in the following way [111]:

$$A_j(k) = \sum_m A_{j+1}(m)\,\bar{h}(2k-m) + \sum_m D_{j+1}(m)\,\bar{g}(2k-m) \tag{7.15}$$

where A_0 is identified as the signal $f(t)$, the coefficients $A_j(m)$ and $D_j(m)$ can be expressed as

$$A_{j+1}(k) = \sum_n h(2k-n)A_j(n) \tag{7.16}$$

$$D_{j+1}(k) = \sum_n g(2k-n)A_j(n) \tag{7.17}$$

$g(k)$ is a high-pass *wavelet* decomposition filter, $h(k)$ is a low-pass *scaling* decomposition filter, $\bar{g}(k)$ and $\bar{h}(k)$ are the corresponding recomposition filters, and j is referred to as the *level of decomposition*.

The total frequency range is now being covered at low frequencies by the low-pass (scaling) function, and at high frequencies by the high-pass (wavelet) function. The signal now can be expressed in terms of the coefficients A and D, which we call 'Approximation' and 'Detail' coefficients respectively. The only remaining problem is how to define the filters.

The filters are uniquely defined from the wavelet and scaling functions by a set of conditions which define a *quadrature mirror filter* [112]. For an orthonormal wavelet, the high and low pass filters are related by

$$h(L-n+1) = (-1)^n g(n) \qquad 1 \le n \le L \tag{7.18}$$

$$\bar{h}(L-n+1) = (-1)^n \bar{g}(n) \qquad 1 \le n \le L \tag{7.19}$$

and reconstruction filters are related to deconstruction filters by

$$\bar{g}(L-n+1) = g(n) \qquad 1 \le n \le L \tag{7.20}$$

$$\bar{h}(L-n+1) = h(n) \qquad 1 \le n \le L \tag{7.21}$$

where the filters are all of length L. Given one of the four filters, the rest follow from these relationships.

We shall describe the case where the wavelet is not orthonormal below. Now, the mother wavelets (and 'father' scaling functions) are described by some polynomial, and the requirement that the high and low pass filters are able to reproduce this polynomial leads to a prescription to perform this calculation, which also uniquely defines the filters for a given wavelet.

One such prescription is the *cascade algorithm*: Take $\bar{h}$, upsample it by a factor of 2 and convolve the result with $\bar{h}$. Continue to upsample and convolve with $\bar{h}$ and the result will converge to the scaling function. Similarly, take $\bar{g}$, upsample and convolve with $\bar{h}$ repeatedly and the result will converge to the wavelet function. Figures 7.4 and 7.5 shows iterations of this process on the db1 filters to compare with the functions shown in Figure 7.2. Figures 7.6 and 7.7 also show iterations of this process on the db2 filters to compare with the functions shown in Figure 7.3.

A wavelet in the Daubechies family *dbX* has X vanishing moments. The *db2* wavelet therefore has 2 vanishing moments, which is why it extends beyond a dyadic range in terms of the number of parameters present at each iteration of the cascade algorithm. Similarly, wavelets with more vanishing moments will have a greater increase in the number of parameters with each successive iteration of the cascade algorithm.

These filters can also be used to directly calculate the approximation and detail coefficients at a given level without first calculating the previous level's approximation coefficients. If the high-pass reconstruction filter after i iterations of the cascade algorithm is given as $\bar{g}_i(n)$, having length L, then for a signal $x(t)$ an individual detail coefficient at level j can

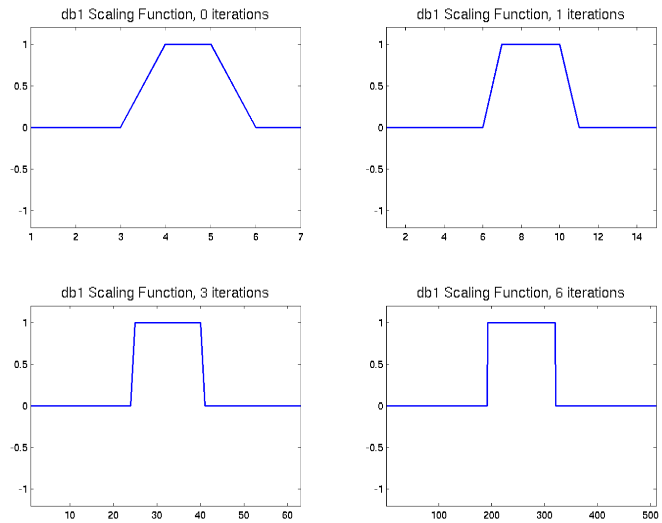


Figure 7.4: Successive iterations of the cascade algorithm for the db1 scaling function. The upper-left figure shows $\bar{h}$, which is iteratively upsampled and convolved with itself to approach the exact scaling function shown in Figure 7.2.

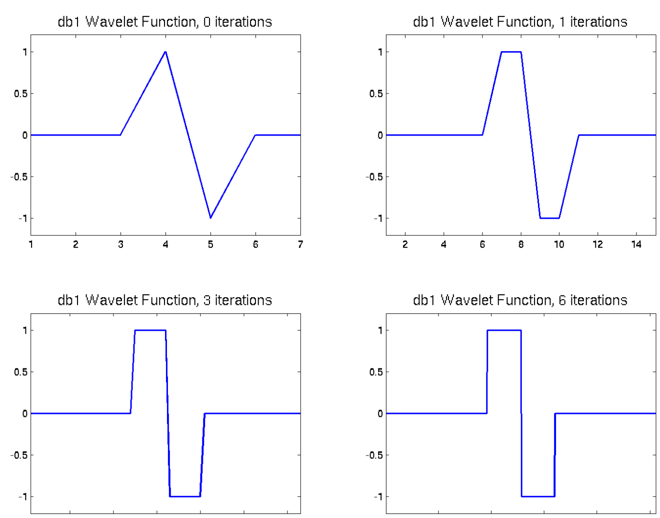


Figure 7.5: Successive iterations of the cascade algorithm for the db1 wavelet function. The upper-left figure shows $\bar{g}$, which is iteratively upsampled and convolved with $\bar{h}$ to approach the exact wavelet function shown in Figure 7.2.

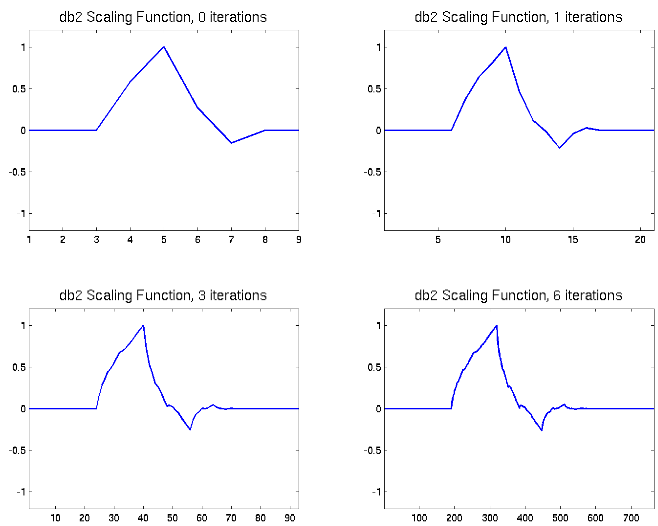


Figure 7.6: Successive iterations of the cascade algorithm for the db2 scaling function. The upper-left figure shows $\bar{h}$, which is iteratively upsampled and convolved with itself to approach the exact scaling function shown in Figure 7.3.

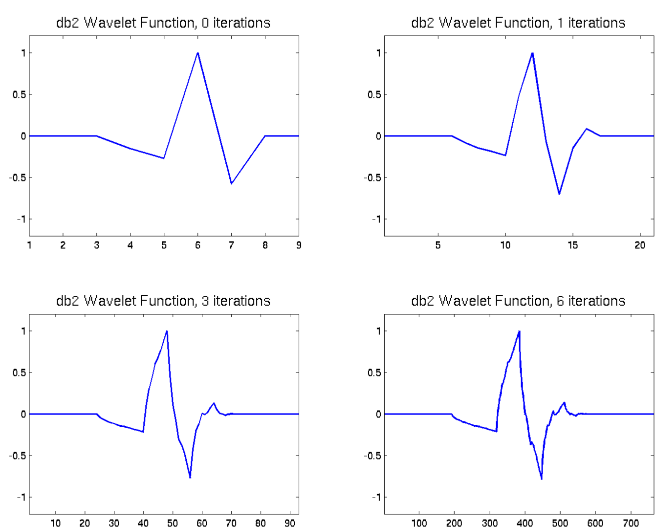


Figure 7.7: Successive iterations of the cascade algorithm for the db2 wavelet function. The upper-left figure shows $\bar{g}$, which is iteratively upsampled and convolved with $\bar{h}$ to approach the exact wavelet function shown in Figure 7.3.

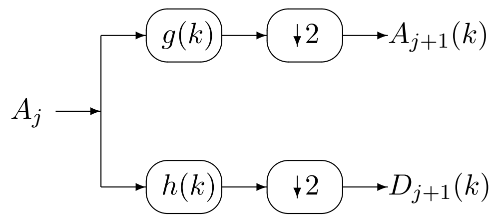


Figure 7.8: A signal is filtered into approximation and detail coefficients via high-pass and low-pass filters, and are subsequently downsampled by a factor of 2.

be written

$$D_j(k) = \sum_{n=1}^{L} x\left(2^j(k+1) + n - L\right) \bar{g}_{j-1}(n) \tag{7.22}$$

In the case where the wavelets are not orthonormal, there are now two scaling functions and two wavelet functions - one pair is used for signal decomposition and one pair is used in signal reconstruction. The high and low pass filters in this case are still uniquely determined by the quadrature mirror filter conditions since relations similar to those in (7.18) - (7.21) still hold.

Once the high and low pass filters are defined, the signal can be filtered into two parts - one that contains high-frequency information (the detail coefficients), and one that contains the low-frequency information (the approximation coefficients). Unfortunately this also means that the output data rate is equal to twice the input data rate, since each coefficient type will yield the same number of values as are contained in the input signal. However, without loss of information, we can take advantage of the factor of $2k$ that appears in (7.16) and (7.17) to downsample each output collection of coefficients by a factor of 2, so that the output data rate is equal to the input data rate.

Figure 7.8 shows the filtering of a set of approximation coefficients into approximation and detail coefficients of the next lowest level. Since $f(t) \equiv A_0$, the signal can now be

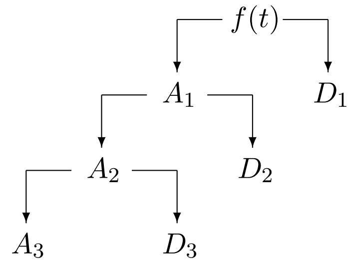


Figure 7.9: The filtering of a signal $f(t)$ into approximation and detail coefficients. If the signal contains n values, then each set of coefficients at level j contains $n/(2^j)$ values. The filtering can continue at successive levels until $2^j \geq n$.

filtered down to any desired level as a collection of approximation and detail coefficients, as shown in Figure 7.9.

Since the number of coefficients at each detail level is progressively halved, the iteration limit will occur where only one value is present in the output coefficients. i.e. If the signal contains n values, then the highest level j occurs when $2^j \geq n$. The width of the scaling function spectrum can be inferred from this limit; usually the length of the longest possible filter is used to determine the width.

Although the mechanics of the process will not be discussed here (for example, see [112]), it is possible to reconstruct the original signal exactly from any given level by using that level's approximation coefficients, and all the detail coefficients down to that level. i.e.

$$
\begin{aligned}
f(t) &\equiv A_1 \oplus D_1 \\
&\equiv A_2 \oplus D_2 \oplus D_1 \\
&\equiv A_3 \oplus D_3 \oplus D_2 \oplus D_1
\end{aligned}
$$

etc...

The decomposition of a signal in this way naturally lends itself to compression and noise removal, which are much the same process. Both use some method of *thresholding* to discard detail coefficients below some value, which removes high-frequency transients in the signal (denoising) and results in a smaller number of values with little loss of information (compression). See Figure 7.10 for an example of the compression of a CDMS phonon signal.

The detail and approximation coefficients of the uncompressed signal are calculated (shown in blue in Figure 7.10) and detail coefficients with a magnitude below a threshold value ($= 1.88$) are set to zero. The coefficient values after thresholding are shown in red; all of the detail coefficients at level 1 and 2 are set to zero. All coefficients except for one at level 3 are set to zero. The signal is reconstructed using the lowest level approximation coefficients and all levels of the detail coefficients (red, top right). The reconstructed signal has an energy ($\equiv \int f(t)^2 \, dt$) which is 99.6% of the original signal's energy, yet it is constructed from a set of approximation and detail coefficients (the same size as the original signal) of which 88% are identically zero.

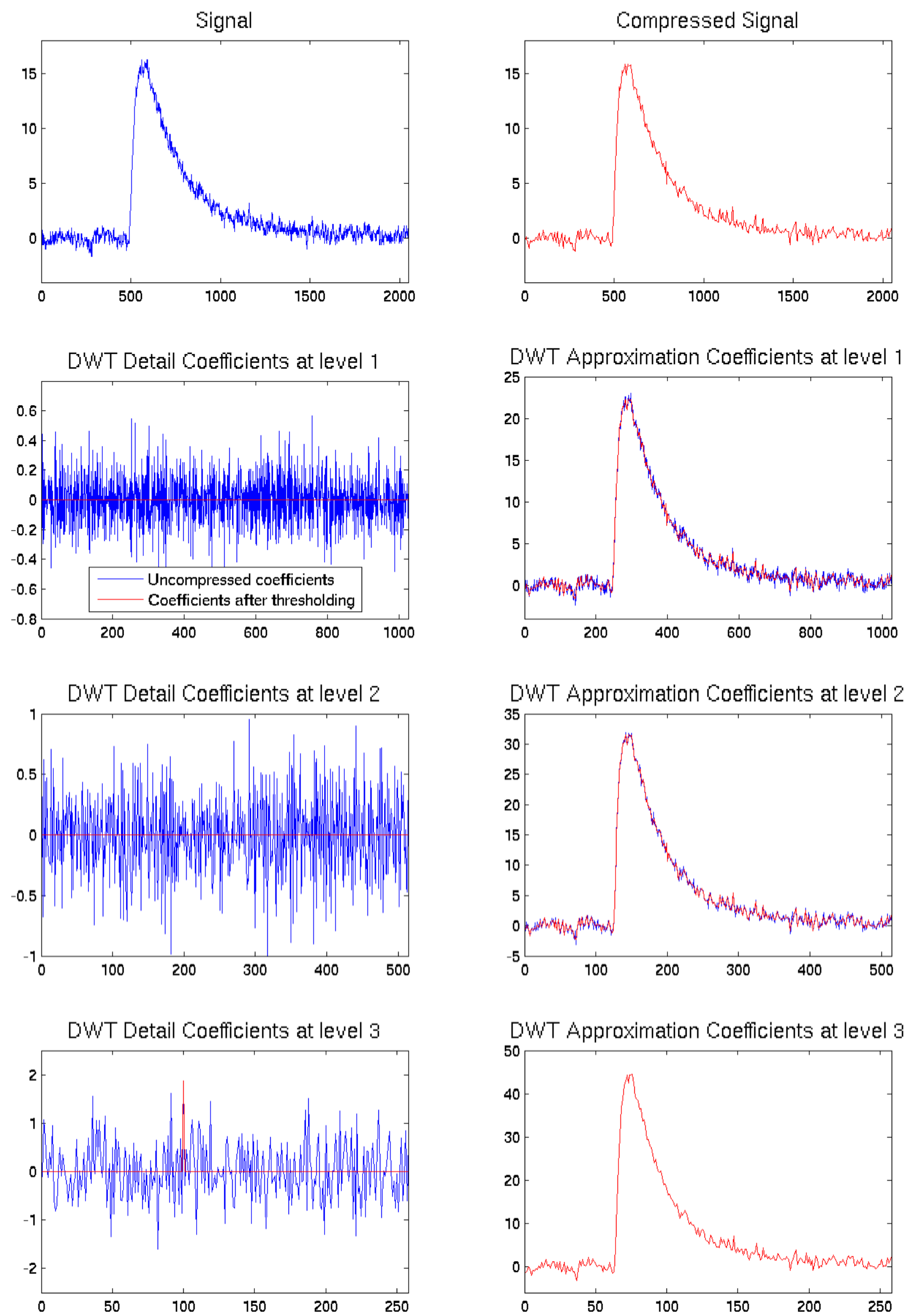


Figure 7.10: Example of performing compression on a CDMS phonon pulse using the *db2* wavelet. See text for explanation.

7.6 Wavelet Families

Figures 7.11 - 7.15 show examples of the most commonly used wavelet families. These are the wavelet functions; scaling functions are also necessary for a multiresolution analysis but are not shown here. Generally speaking, the coefficient for a wavelet decomposition will be larger when the shape of the wavelet physically resembles the shape of the signal. For example, this makes the step-function-shaped 'haar' wavelet particularly useful for finding discontinuities in signals.

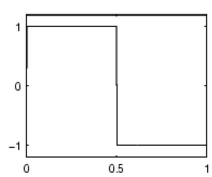


Figure 7.11: The **haar** wavelet, the simplest of all the wavelets. It is particularly useful at finding discontinuities in signals, by virtue of it's own discontinuous nature.

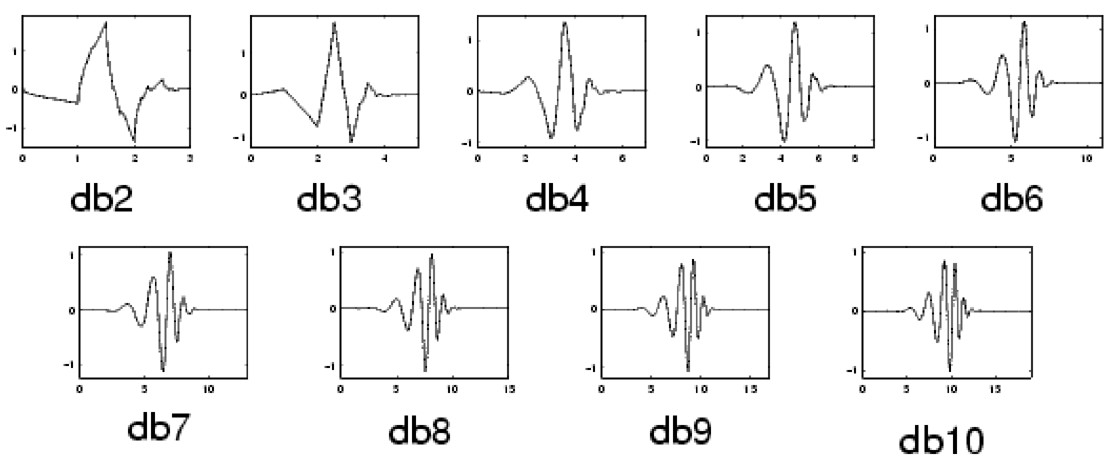


Figure 7.12: The **Daubechies** wavelet family, which are orthonormal. db1 is the same as the haar wavelet.

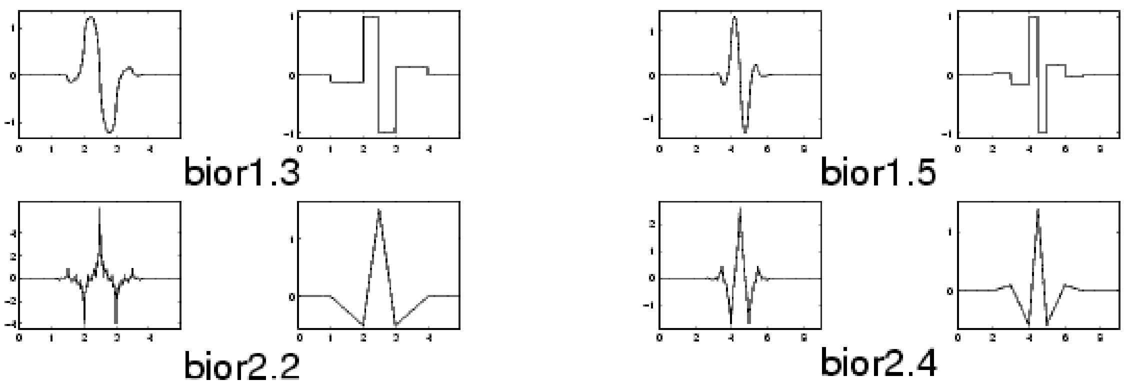


Figure 7.13: A selection of wavelets from the **Biorthogonal** family. These wavelets form a dual frame, so that the wavelet used for decomposition is different than the one used for reconstruction. Hence these are shown in pairs - the right hand figure in each pair is the decomposition filter. The **Reverse Biorthogonal** or **rbio** family is obtained by exchanging the decomposition and reconstruction filters in the Biorthogonal family.

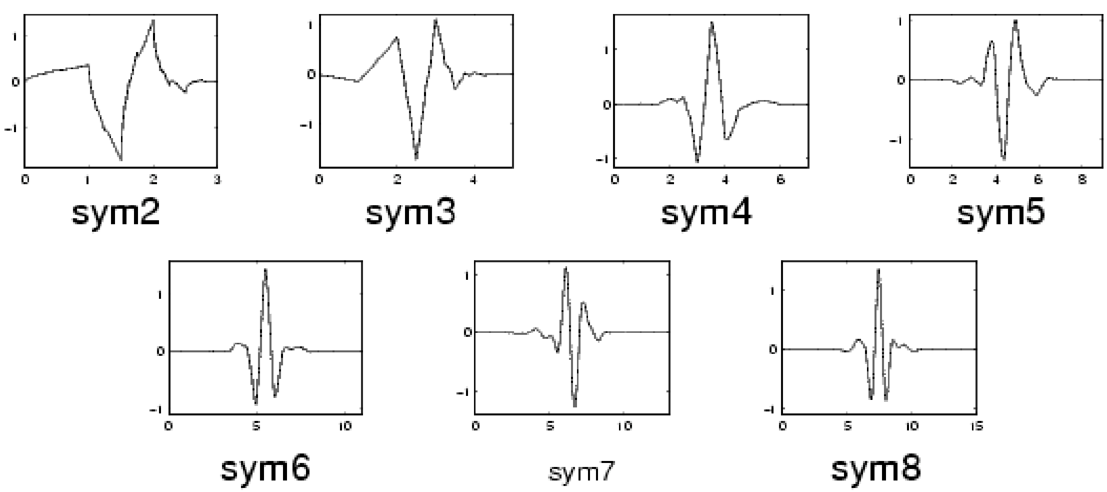


Figure 7.14: The **Symlet** family of wavelets. These have similar properties to the Daubechies wavelets.

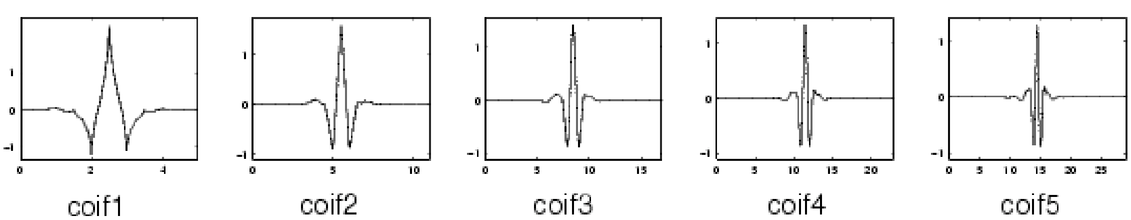


Figure 7.15: The **Coiflet** family of wavelets.

7.7 DWT Example I: Discontinuous Signal

Consider the signal in Figure 7.16. It is a sine wave + white noise with a discontinuity in the sine component part way through the signal (at bin 650):

The decomposition of this signal using the *haar* wavelet, up to level 4, is shown in Figure 7.17. This Discrete Wavelet Transform (DWT) analysis follows the use of a filter bank described in Section 7.5. Notice how, particularly at the higher levels, the detail coefficients show a large negative value where the discontinuity occurs (negative because the signal has a step up, and the haar wavelet is a step down). The choice of wavelet for a particular task depends heavily on the nature of the task, and there is no analytic way to choose the wavelet. This is just a simple example to show how the analysis works in practice.

7.8 DWT Example II: CDMS phonon pulse analysis

Although the wavelet analysis of a CDMS phonon pulse will be covered in more detail later in Chapter 9, here we show the discrete wavelet decomposition of an actual CDMS detector phonon pulse taken from the primary channel for a single event. The main part of a CDMS pulse is shown in Figure 7.18, with the decomposition of this signal, using the *rbio1.5* (reverse biorthogonal) wavelet shown in Figure 7.19.

The 'high pass filters' shown in the third column of Figure 7.19 are those applied to the original signal to obtain the approximation and detail coefficients at a given level, as per Eqn. (7.22). The scales of the high-pass filters as shown are correct to obtain the coefficients at each level from the original signal.

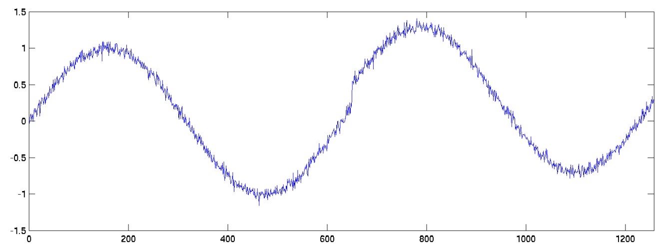


Figure 7.16: Noisy sine function with discontinuity at bin 650.

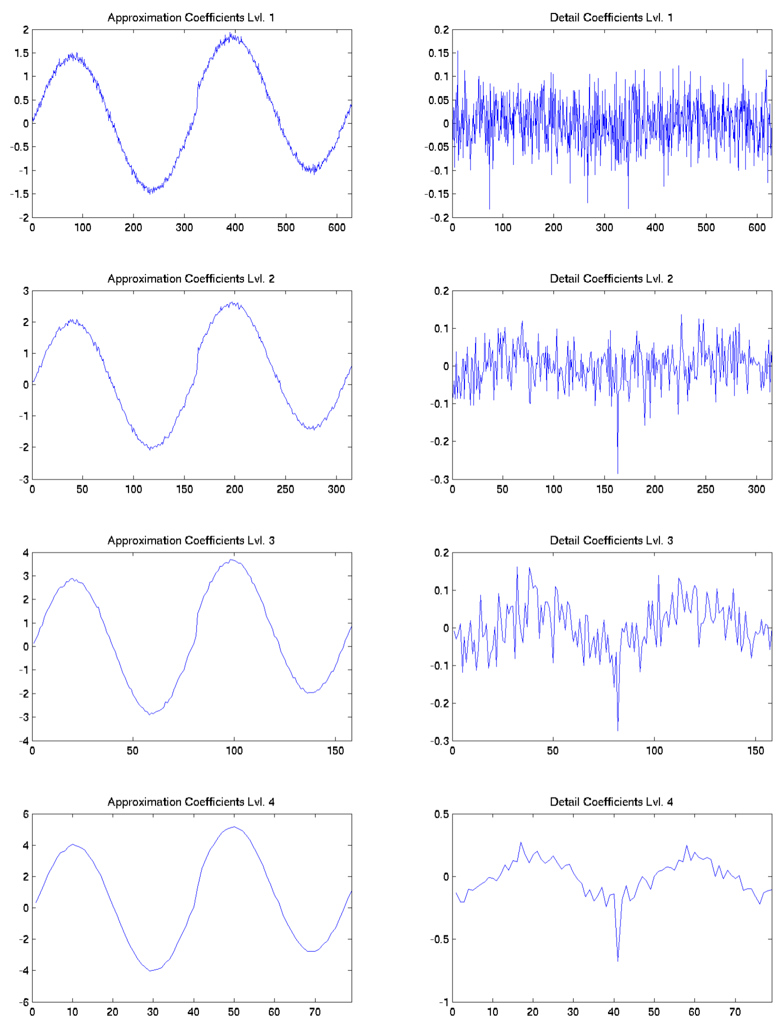


Figure 7.17: Discrete Wavelet Decomposition of the signal shown in Figure 7.16 using the *haar* wavelet, up to level 4.

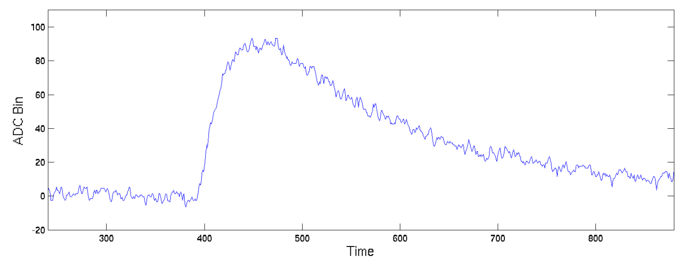


Figure 7.18: Section of a CDMS detector phonon pulse with recoil energy $\sim 50\,\text{keV}$.

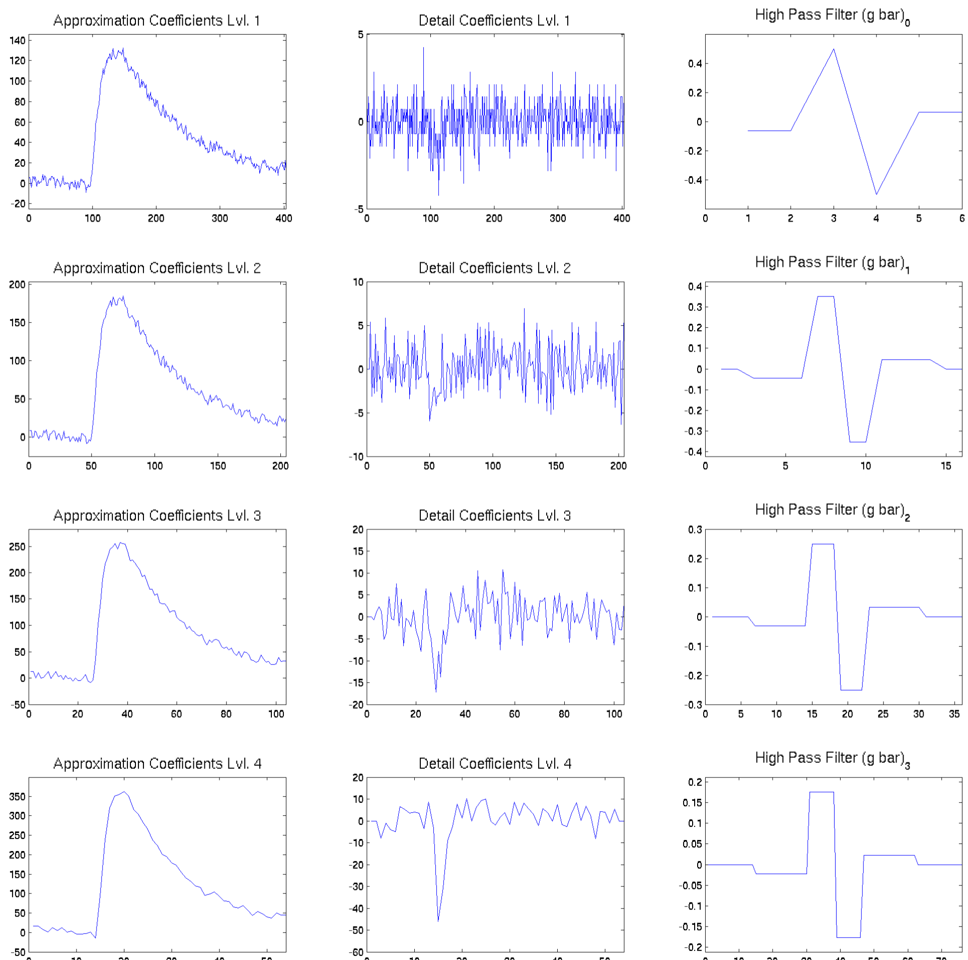


Figure 7.19: Discrete Wavelet Decomposition of the signal shown in Figure 7.18 using the *bior1.3* wavelet, up to level 4. The filters $\bar{g}_j(n)$ used in Eqn. (7.22) are shown at levels $j+1$.

All wavelet families, when applied to the CDMS phonon pulse, show some sharp change in detail coefficient values around the leading edge of the pulse, although the *rbio* family of wavelets have a particularly noticeable change in values at this point. Whether the values of these leading edge coefficients provide information on recoil type is a question we shall return to in Chapter 9.

Chapter 8

Neural Network Technology

The creation of a surface event cut is essentially a classification problem. That is, we would like to know, for each point in some n-dimensional space - where n is the number of detector parameters that we're considering - what the probability is of a surface electron recoil or a nuclear recoil being found at that point. Once we know how likely we are to find a given class of event at each location, we can place our cut based on this set of probabilities.

Since neural networks based on perceptrons are particularly good at analyzing classification problems, we examine their use applied to this problem.

8.1 The Perceptron

A perceptron is a mathematical object that takes some collection of input values $x_1, x_2, ...x_d$ and produces output values a_j:

$$a_j = \sum_{i=1}^{d} w_{ji} x_i + b_j \qquad j = 1, 2, ..., M. \tag{8.1}$$

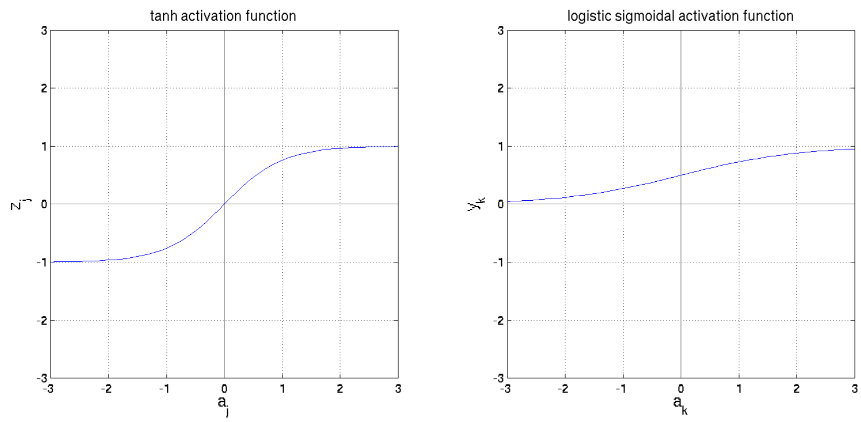


Figure 8.1: Example perceptron activation functions. Use of *tanh* for the hidden layer function in a multilayer perceptron network is very common. The output layer activation function is usually either a logistic sigmoidal function as shown, or a simple linear function.

where w_{ji} is a matrix of *weight* values and b_j is an array of *bias* values. The number of output values, M, depends upon the relation of this perceptron to the desired output values, and to other perceptrons, if they exist.

The values a_j are usually transformed by some activation function to give the final output of the perceptron. This is typically some non-linear function such as tanh.

$$z_j = \tanh(a_j) \qquad j = 1, 2, ..., M. \tag{8.2}$$

Two example activation functions are shown in Figure 8.1.

So in summary, the perceptron takes input values x_d and creates output values z_j. This operation alone isn't particularly spectacular. However, perceptrons can be combined into a network which can then be 'trained' on model data - that is, the weight and bias values are modified based on some comparison of the output values and the desired output values, until the values converge and the network produces the desired behavior.

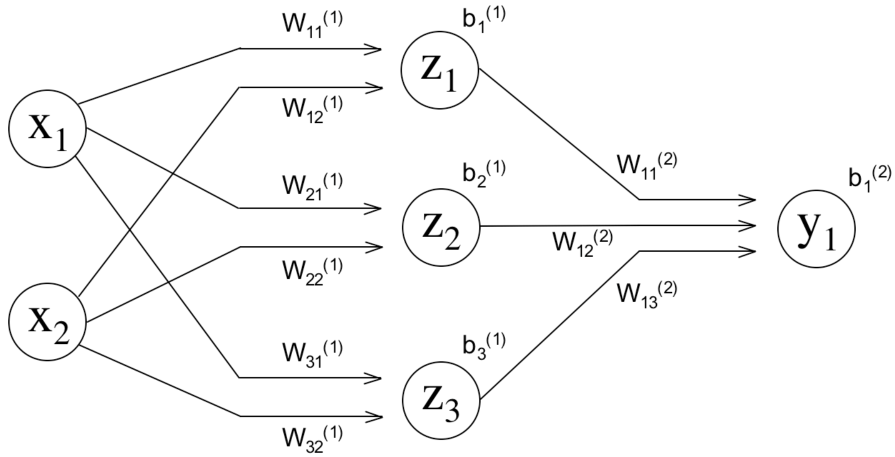


Figure 8.2: Example MLP network with 2 input perceptrons, 1 output perceptron and 3 hidden perceptrons.

8.2 Multilayer Perceptron Networks (MLPs)

It has been shown that one layer of suitable nonlinear perceptrons followed by a linear or logistic layer (to be described) can approximate any nonlinear function with arbitrary accuracy, given enough nonlinear perceptrons [113].

The multilayer perceptron network has d input values x_i, M 'hidden layer' units z_j, and c network outputs y_k. The input layer and hidden layer have associated weights matrices of size $(M \times d)$, and $(c \times M)$ respectively. The hidden layer and output layers have M and c associated bias values, respectively. The hidden layer transfer function is tanh, and the output layer activation function is a logistic sigmoidal function, $y_k = (1 + \exp(-a_k))^{-1}$ (see Figure 8.1) although linear functions are also commonly used.

An example network, for which $d = 2$, $M = 3$ and $c = 1$ is shown in Figure 8.2. The number of input and output vectors is defined by the problem at hand - for example, this network could be used to classify 2-dimensional data points as being one of two different types (where the output $y_1 = 0$ for a data point of type 1, and $y_1 = 1$ for a data point of type 2). The number of units in the hidden layer is empirically chosen to be sufficient for network conversion, since computation time rises exponentially with the number of hidden units.

The layer outputs are given by:

$$z_j = \tanh\left(\sum_{i=1}^{d} w_{ji}^{(1)} x_i + b_j^{(1)}\right) \qquad j = 1, 2, ..., M. \tag{8.3}$$

$$y_k = \left(1 + \exp\left(-\sum_{j=1}^{M} w_{kj}^{(2)} z_j + b_k^{(2)}\right)\right)^{-1} \qquad k = 1, 2, ..., c. \tag{8.4}$$

The network needs to be *trained* on sample data in order to perform in the desired way. That is, a set of sample inputs are given along with correct network outputs and the network is run multiple times. Each time the network runs, the outputs are compared with the correct values and the error is propagated back to the weights and biases, which are modified, and the network runs again. This process is repeated until the weights and biases have converged.

The network converges via minimization of an error function. For example, a network with c outputs (index k) is supplied with d data points (index n), outputs values y_k^n, and has target output values t_k^n. The sum-of-squares error function is given by

$$E = \frac{1}{2} \sum_{n=1}^{d} \sum_{k=1}^{c} (y_k^n - t_k^n)^2 + \epsilon \tag{8.5}$$

where ϵ is a stochastic term neccessary to stop the network from getting stuck in a local minimum of the error function.

The y_k^n values are effectively a function of the input values and the network weights and biases. There are various ways to minimize a smooth function of many parameters, such as

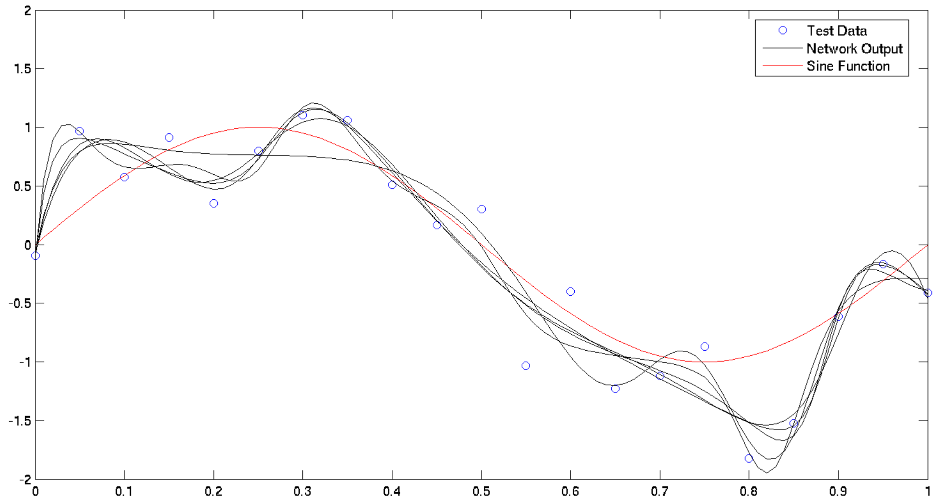


Figure 8.3: Example of overfitting with an MLP network. 20 data points (circles) are taken from a noisy sine function (red line), and used to train five separate MLP networks (black lines). Although this particular example is simple enough that overfitting could easily be avoided by manipulating the amount of noise added to the error calculation, we don't know what this number should be *a priori*.

scaled conjugate gradient and quasi-Newton methods [115]. These will not be discussed in detail here, but scaled conjugate gradient is used predominantly in the subsequent analysis.

A noise term is usually added to the error value since without this term there is a risk that the network will converge to a local minimum. However, if the noise term is too large, the system can end up in a state with a loss of generality. The iteration of the system can be thought of that of a 'noisy ball' in a landscape of potentials.

If the analysis is not a simple one, the network can *overfit* the training data - meaning that the performance of the network is good on training data, but is poor on subsequent test with unseen data. i.e. the network is fitting the noise on the training data, rather than the underlying behavior of the data. Figure 8.3 shows an example of overfitting, where the network outputs follow the noise details rather than the underlying function. This example is intended to be illustrative only, since it could easily be improved by manipulation of the noise term in the calculated error, although this could only be done were we to examine the behavior of the network on unseen data and adjust accordingly. That is, overfitting could not, in principal, be avoided in even this simple example if we needed to fix the network behavior whilst being blind to unseen data. Since that is indeed the case when setting a surface event cut, this is especially important.

8.3 Bayesian Techniques

Running the network on the same set of inputs twice will, in general, result in two different sets of output data because of the addition of noise in the iterative convergence of the error function. We also do not know, in principle, what the size of the noise should be. This makes it difficult to train a network without overfitting on the training data unless one is allowed to observe the effectiveness of the network upon unseen data.

The solution to this problem is to use a Bayesian approach that accounts for uncertainty in the network's weights and biases. The Bayesian method offers a number of important benefits; 1) By taking account of parameter[1] uncertainty, overfitting is generally not a problem, 2) The relative importance of different input variables can be determined via automatic relevance determination (ARD).

In this approach, an iterative procedure is used to dynamically adjust the network's sensitivity to each input value based on repeated observations of the network output value(s). At first, we do not know which weight values associated with each input node are likely or unlikely. Each input node is assigned a prior probability distribution to express the initial

[1]By parameters, we mean the combined collection of network weight and bias values

uncertainty in the value of that input's weights to the hidden layer. After the network undergoes training in the usual fashion, Bayes' theorem is used to update each input's prior probability distribution on the basis of the network output values. For example, if it's clear from the output values that a particular input is not contributing much to the outputs then it's prior probability distribution will become narrower, constraining it's weight values to be small. Conversely, if an input node is contributing significantly to the output values then it's prior distribution will be widened to allow large weight values.

This method quickly tranquilizes input nodes that are only providing noisy input values. The weights of such an input are dynamically constrained to be small, no matter what effect on the input minimizing the error might have. Without this procedure the network is free to set the weights to any values in order to minimize the error function at the cost of overfitting. By constraining unimportant inputs to have small weights, overfitting is avoided.

We start with a *prior* probability distribution $p(\boldsymbol{\omega})$ which expresses knowledge of the parameters (network weights and biases) before the data is observed (where $\boldsymbol{\omega}$ is a vector containing all of the network weight and bias values). The network is trained to create a set of output values, $\mathcal{D}$, where $\mathcal{D}$ is used to denote the set of network output values y_k^n. Bayes' theorem can be used to update the prior probability distribution.

Bayes' Theorem gives

$$p(\boldsymbol{\omega}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\omega})p(\boldsymbol{\omega})}{p(\mathcal{D})} \tag{8.6}$$

where $p(\boldsymbol{\omega})$ is the prior probability distribution of the network parameters (weights and biases), $p(\mathcal{D}|\boldsymbol{\omega})$ is the likelihood of getting the observed set of output values $\mathcal{D}$ given the network parameters $\boldsymbol{\omega}$, and $p(\mathcal{D})$ is the probability of seeing the set of output values $\mathcal{D}$.

$p(\mathcal{D})$ is called the *evidence* and is given by the identity:

$$p(\mathcal{D}) = \int p(\mathcal{D}|\boldsymbol{\omega}')p(\boldsymbol{\omega}')\,\mathrm{d}\boldsymbol{\omega}' \tag{8.7}$$

Assuming that we are able to calculate this integral, the prediction for new inputs $\boldsymbol{x}$ is given by

$$p(\boldsymbol{y}|\boldsymbol{x},\mathcal{D}) = \int p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\omega})p(\boldsymbol{\omega}|\mathcal{D})\,\mathrm{d}\boldsymbol{\omega} \tag{8.8}$$

where $\boldsymbol{y}$ is the vector of network outputs using the inputs $\boldsymbol{x}$, and $p(\boldsymbol{y}|\boldsymbol{x},\mathcal{D})$ means the probability of the output $\boldsymbol{y}$ given the inputs $\boldsymbol{x}$, as a function of $\mathcal{D}$.

Unfortunately, integrals like (8.7) and (8.8) can only be calculated analytically for a very small class of prior distributions. The calculation of the integrals must be performed either by the use of approximations - known as the *evidence* procedure - or by numerical methods such as Monte Carlo or Markov Chain sampling. Here we concentrate on application of the evidence procedure.

The choice of the prior probability distribution for the weights and biases of our network should reflect the fact that we expect the underlying generator of the dataset to be smooth. We also tend to favor small values for the network weights, since large weight values suggest that the network is performing an unstable mapping.

These requirements suggest a Gaussian prior distribution of the form

$$p(\boldsymbol{\omega}) = \left(\frac{\alpha_i}{2\pi}\right)^{W/2} \exp\left(-\frac{\alpha_i}{2}\|\boldsymbol{\omega}\|^2\right) \tag{8.9}$$

where α_i represents the inverse variance of the expected distribution and is known as a *hyperparameter* (since it parameterizes the distribution of the other parameters), W is the total number of relevant weights and biases (the length of $\boldsymbol{\omega}$), and

$$\|\boldsymbol{\omega}\|^2 \equiv \sum_{i=1}^{W} \boldsymbol{\omega}_i^2 \tag{8.10}$$

This is equivalent to adding an extra term to the error function (8.5) which penalizes large weight values:

$$\begin{aligned} E &= \frac{1}{2} \sum_{n=1}^{d} \sum_{k=1}^{c} (y_k^n - t_k^n)^2 + \frac{\alpha}{2} \sum_{i=1}^{W} \boldsymbol{\omega}_i^2 + \epsilon & (8.11) \\ &= E_S + E_W + \epsilon & (8.12) \end{aligned}$$

where E_S is the sum-of-squares error and E_W is the weight-based error term.

In the automatic relevance determination (ARD) framework, we associate a hyperparameter α_i with each input variable x_i. Each α_i represents the inverse variance of the prior distribution of the weights from that input to the hidden layer nodes.

There is also a single α_i for the hidden layer bias, an α_i for the set of hidden layer to output weights (one per output), and a final α_i for the output layer bias. During the training of the network, it is possible to modify the hyperparameters and find their optimal value. Because the hyperparameters represent the inverse of the variance of the weights, a small α_i for a network input means that weights significantly different from zero are favored and thus it is likely that the corresponding input is important. Conversely, a large α means that the weights are constrained to be close to zero, and thus the corresponding input is apparently less important. In order be able to compare inputs fairly, each input from the data is independently scaled to have zero mean and unit variance.

8.4 The Evidence Procedure

The evidence procedure is an approximation used to solve the integrals (8.7) and (8.8). We wish to find the probability $p(\boldsymbol{\omega}|\mathcal{D})$ - for a given set of network outputs $\mathcal{D}$ what is the probability distribution of the network weight and bias values?

Now that the prior probability distribution has been parameterized by the α_i hyperparameters, in principle it is possible to find $p(\boldsymbol{\omega}|\mathcal{D})$ by integrating over all the unknown hyperparameter values:

$$
\begin{aligned}
p(\boldsymbol{\omega}|\mathcal{D}) &= \int p(\boldsymbol{\omega}, \boldsymbol{\alpha}|\mathcal{D})\, \mathrm{d}\boldsymbol{\alpha} && (8.13)\\
&= \int p(\boldsymbol{\omega}|\boldsymbol{\alpha}, \mathcal{D})\, p(\boldsymbol{\alpha}|\mathcal{D})\, \mathrm{d}\boldsymbol{\alpha} && (8.14)
\end{aligned}
$$

where $\boldsymbol{\alpha}$ is a vector of the α_i hyperparameter values.

The approximation made in the evidence procedure is to assume that the function $p(\boldsymbol{\alpha}|\mathcal{D})$ is sharply peaked around $\boldsymbol{\alpha_{MP}}$, the most probable values of the hyperparameters:

$$
\begin{aligned}
p(\boldsymbol{\omega}|\mathcal{D}) &\approx p(\boldsymbol{\omega}|\boldsymbol{\alpha_{MP}}, \mathcal{D}) \int p(\boldsymbol{\alpha}|\mathcal{D})\, \mathrm{d}\boldsymbol{\alpha} && (8.15)\\
&\approx p(\boldsymbol{\omega}|\boldsymbol{\alpha_{MP}}, \mathcal{D}) && (8.16)
\end{aligned}
$$

Hence, if we find the hyperparameter values that maximize the weight posterior probability, then we can perform all calculations that involve $p(\boldsymbol{\omega}|\mathcal{D})$ using the distribution with the hyperparameters fixed at those values.

It can be shown [116] that by considering $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\alpha}} \ln p(\mathcal{D}|\boldsymbol{\alpha})$, the most probable value of the hyperparameter $\boldsymbol{\alpha_{MP}}$ is expressed as a function of it's previous value $\boldsymbol{\alpha}$:

$$
\boldsymbol{\alpha_{MP}} = \frac{1}{\|\boldsymbol{\omega}\|^2} \left(\sum_{i=1}^{W} \frac{\lambda_i}{\lambda_i + \boldsymbol{\alpha}} \right) \tag{8.17}
$$

where λ_i are the eigenvalues of the network's Hessian matrix. In this framework, the Hessian matrix is defined as the second derivative of the error function with respect to the network's first and second layer weights (see Figure 8.2):

$$\boldsymbol{H} \equiv \frac{\partial^2 E}{\partial \omega_{ji}^{(1)} \partial \omega_{lk}^{(2)}} \tag{8.18}$$

so that $\boldsymbol{H}$ is a square matrix of dimension $W \times W$, and E is given by (8.11).

So the evidence procedure is iterative; given some initial choice of $\boldsymbol{\alpha}$, the network is trained, the new $\boldsymbol{\alpha}$ is given by (8.17) and the network is trained again, $\boldsymbol{\alpha}$ is re-estimated, etc. until $\boldsymbol{\alpha}$ converges.

To implement the evidence procedure in practice the following steps are performed:

1. Choose initial values for the hyperparameters α_i (one per input) in the network. The final outcome is relatively insensitive to the initial hyperparameter values. The weights and biases are also set to nominal values, typically on the order of 10^{-2}, but again the final outcome is empirically close to independent of these values.

2. Train the network to minimize the error function (8.11). This is done via some calculation of the error function gradient with respect to the weights and biases, such as quasi-Newton or scaled conjugate gradient [115]. The operation of a network not using the evidence procedure would stop here.

3. The evidence procedure described above is used to set new values for the hyperparameters, using (8.17).

4. Steps 2 and 3 are repeated, until the hyperparameters converge.

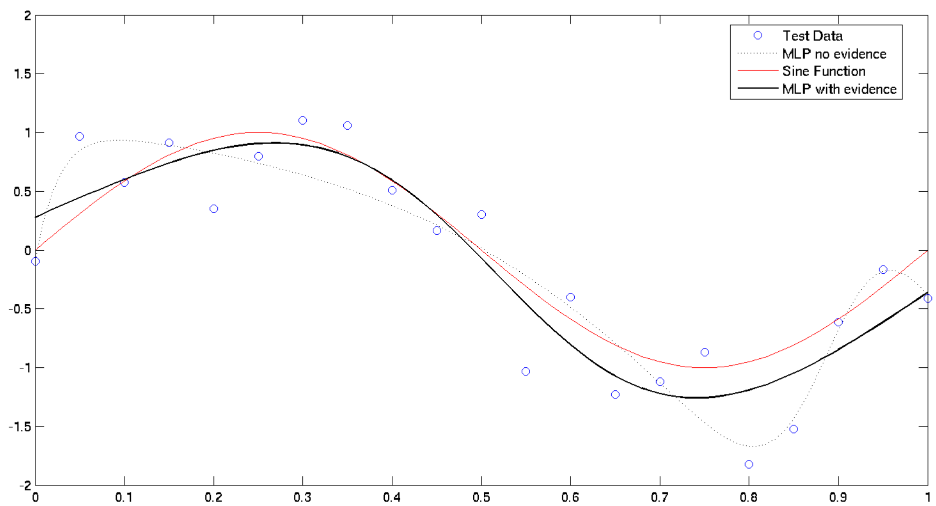


Figure 8.4: Example of applying the evidence procedure to a regression problem (the same as shown in Figure 8.3). 20 data points (circles) are taken from a noisy sine function (red line), and used to train an MLP network without using the evidence procedure (dotted black line), and the same MLP network with the evidence procedure applied (solid black line). Network shown in Figure 8.5.

Figure 8.4 shows an example of applying the evidence procedure to the same problem as shown in Figure 8.3. Although the initial hyperparameter and weight values are selected manually, the output of the network is not sensitive to them, given enough iterations to converge the hyperparameters and enough hidden perceptrons. Thus an optimal solution can be reached without overfitting and with no sensitivity to initial conditions.

There is no analytical way to select the number of units in the hidden layer, although a good method is to start with the same number as there are input units and increase the number of hidden units by one, running the network each time. For most applications it becomes clear how many hidden nodes are appropriate in this way. The example shown in Figure 8.4 uses 1 input node, 3 hidden nodes and 1 output node (Figure 8.5).

We can also apply the evidence procedure to classification problems, such as the one we

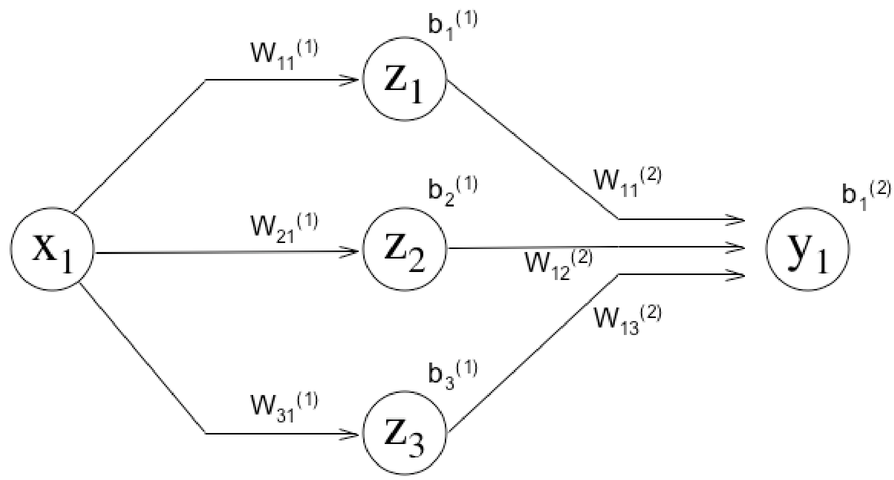


Figure 8.5: MLP network used to produce Figure 8.4 with 1 input, 1 output and 3 hidden perceptrons.

will face in making a surface event cut. Figure 8.6 shows the result of applying the evidence procedure to such a test problem, where the data is generated from a set of three Gaussian Mixture Models (GMMs). The advantage in using GMMs is that, via knowledge of the underlying generator(s) of the data, they allow the analytic calculation of the optimal Bayes result - that is, the result that would be obtained if the neural network is working perfectly (calculation of the Bayes result takes account of the limited statistics in the dataset).

The example in Figure 8.6 shows the result of using two different networks on the same classification problem. The regularized network uses the evidence procedure, whereas the unregularized network does not. Both networks have 2 input nodes, 8 nodes in the hidden layer, and 1 output node.

This toy model allows us to examine how well a cut set using an MLP and the evidence procedure would perform on unseen data, since the Bayesian decision boundary exactly describes the optimal performance of a network. An MLP using the evidence procedure on a binary classification problem is able to produce a result very close to the optimal solution independently of the initial network parameters.

8.5 Key Summary

It can be seen that the evidence procedure gives a smoother decision boundary, closer to the optimal Bayes boundary, than the unregularized (no evidence procedure) MLP. **Although the unregularized MLP can be improved by tweaking the magnitude of the noise term, we would be unable to do this whilst being blind to the results.** The evidence procedure empirically provides results that do not depend upon any initial conditions (assuming enough hidden perceptrons and sufficient iterations of the hyperparameters).

Further studies of test cases also show that the evidence procedure is particularly good at not being 'distracted' by unimportant inputs. For example, running the above classification problem again while adding another input variable that is generated from noise does not affect the result of the network. However, the unregularized MLP performs worse when given this extra variable, since it lacks the ability to dynamically make the weight values for this input small without overfitting. Conversely, the evidence procedure gives almost exactly the same result as it did without the noise input, even when adding several such inputs. Furthermore, it provides an easy way to weigh the importance of each input via ARD - examination of the hyperparameter value associated with each input variable allows a direct comparison since a smaller α represents a larger prior variance, and consequently a more relevant input.

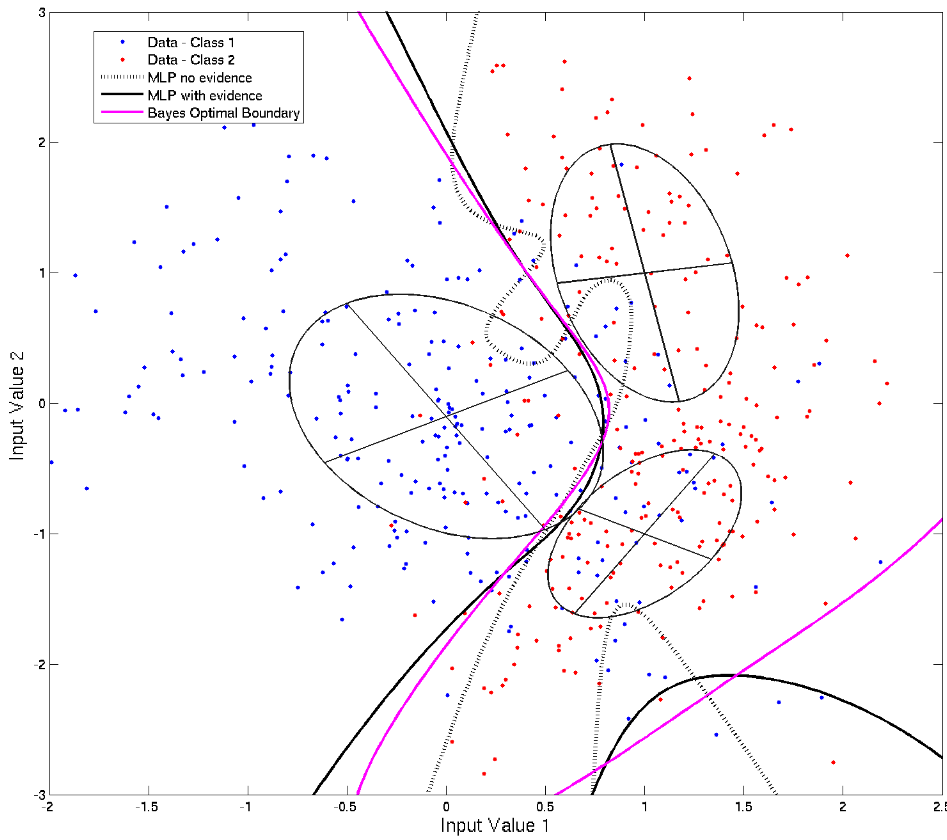


Figure 8.6: Example of applying the evidence procedure to a classification problem. Dummy data is generated from gaussian generators at (0,-0.1) (Data Class I) and (1,1) & (1,-1) (Class II), shown as grey ellipses. The data is used to train an unregularized MLP network (result in dotted black) and an MLP network using the evidence procedure (result in solid black). The results shown are the optimal decision boundaries - the points at which the network output claims a 50/50 probability of a Class I or Class II data point. The use of Gaussian generators allow the analytic calculation of the Bayesian optimal decision boundary (solid magenta).

Chapter 9

Run 119 - Wavelet / Neural Net Based Analysis

This chapter describes an alternate approach to the surface event cut that is so central to the CDMS analysis. Recent work in CDMS has suggested that phonon pulse timing parameters can be combined to provide even greater discrimination between electron and nuclear recoils [105] [106]. It is also interesting to ask whether these timing parameters, such as the phonon signal's risetime and delay, represent all of the recoil discrimination information contained within the phonon signal.

The alternate analysis discussed here uses a wavelet-based approach in order to extract the maximum amount of information from the leading edge of the pulse, and maximizes discrimination based on this wavelet information combined with existing analysis parameters by using a neural network. All of the analysis in this chapter, unless otherwise noted, is the sole work of the author.

9.1 Creating an Optimal Cut

Once we have characterized a parameter space in terms of the expected location of signal and background events in that space, we must decide where to place a cut. This will depend on our goals with the experiment - for example, an experiment that hopes to see very few signal events would want to have a much lower expected background that an experiment that expects to see many signal events. In general, however, we wish to optimize our choice of cut such that our analysis gives the best expected limit on the signal.

Let us suppose that the fraction of signal events that pass the cut is given by $\alpha(p)$ and the fraction of background events that pass the cut is given by $\beta(p)$, where p is a value that parameterizes these distributions. Thus the cut is set by deciding upon a suitable choice of p, and a perfect cut would have $\alpha = 1, \beta = 0$.

We wish to minimize the expected upper limit of the WIMP-nucleon cross-section in the absence of a signal, and we know that this limit will improve as $\alpha(p)$ increases and the experiment's exposure increases. However, the limit will also depend upon the upper limit of the expected background. Hence, we can write:

$$\text{Limit} = \frac{\mathrm{P}_{90}(N_\beta\, \beta(p))}{\alpha(p) MT} \tag{9.1}$$

where N_β is the number of background events present before the cut is applied, MT is the total exposure (mass $\times$ time), and $\mathrm{P}_{90}(x)$ is the Poisson 90% upper limit of x ($\mathrm{P}_{90}(0) = 2.3, \mathrm{P}_{90}(1) = 3.9$ etc.). This expression assumes no attempt to subtract the background and no signal events. The dark matter limit is in units of (90% confidence limit of signal events) / kg / day.

Therefore to find where to place the cut, we need only find the value of p that minimizes

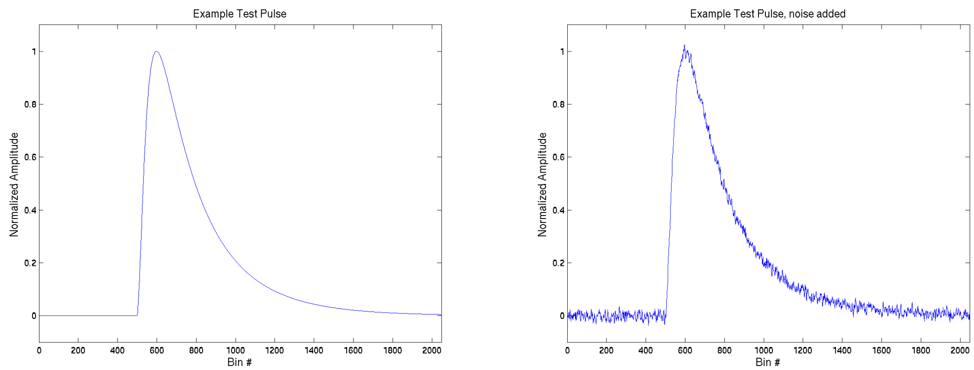


Figure 9.1: Test pulses used for initial investigation into wavelet representation of pulse parameters, modeled after CDMS phonon pulses.

the function (9.1).

9.2 Pulse Parameters via Wavelets

The first thing we would like to know in using wavelets to represent the phonon pulse shape is whether there is a simple way to represent familiar phonon pulse characteristics using wavelets. For example, can the wavelets easily represent quantities such as the signal start time and risetime?

The problem was examined first for a simple error-function-shaped distribution, and then generalized to the CDMS phonon pulse shape. Figure 9.1 shows an example test signal used for the analysis, with and without noise added. The test signal is created from the form

$$s(t) = A\left(e^{-t/\tau_2} - e^{-t/\tau_1}\right) \tag{9.2}$$

where τ_1 and τ_2 are constants used respectively to set the rise and fall time of the pulse, and A is a normalization constant used to set the peak height of the signal to 1 before adding noise. The signals are modeled after real CDMS phonon pulses (see Figure 2.8), and use a

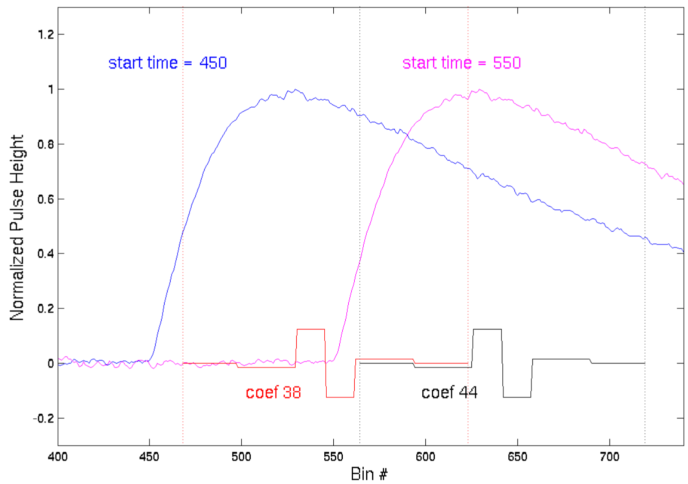


Figure 9.2: Schematic showing the calculation of two discrete wavelet detail coefficients shown in Figure 9.3. Two phonon pulses are shown, with start times of 450 (blue) and 550 (magenta). Coefficient 38 for each signal follows from the correlation of the filter (red) and the signals - hence we obtain expect a small positive value for coefficient 38 when the start time is 450, and a large negative value when the start time is 550. The values of coefficient 44 obtained through correlation of the filter (black) signal are almost identical for both signals (see Figure 9.3).

noise spectrum similar to that present in Soudan data.

A set of 2500 test signals was created using many possible combinations of start time and risetime values, whilst keeping the amplitude and fall times fixed. We examine the 10-40% risetime values since this has been historically used in CDMS analyses, but this could easily be extended to some other part of the leading edge of the pulse. The discrete wavelet detail coefficients from a DWT analysis of all of these signals were then examined using a variety of wavelet families and decomposition levels. The aim was to find one or more coefficients that represent one of the two parameters (start time, risetime), yet are not dependent upon the other parameter.

By examining Discrete Wavelet Transformations of the test pulses, it was found that the *bior1.3* wavelet at a decomposition level of 4 (see Figure 7.19) is particularly good at

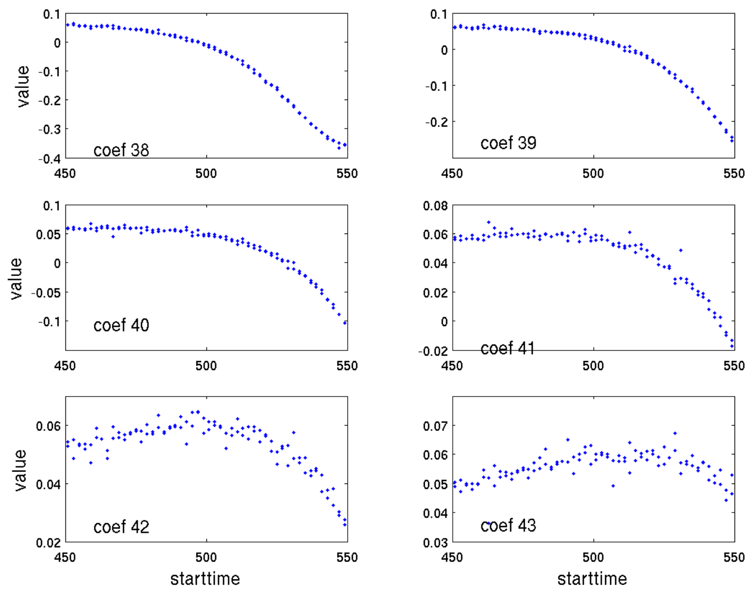


Figure 9.3: Selection of DWT coefficients of noisy training pulses, using the *bior1.3* wavelet at level 4, plotted against start time values of the test pulses. There is enough dependence of these coefficients upon the start time that a neural network can be trained to calculate the start time given these coefficients.

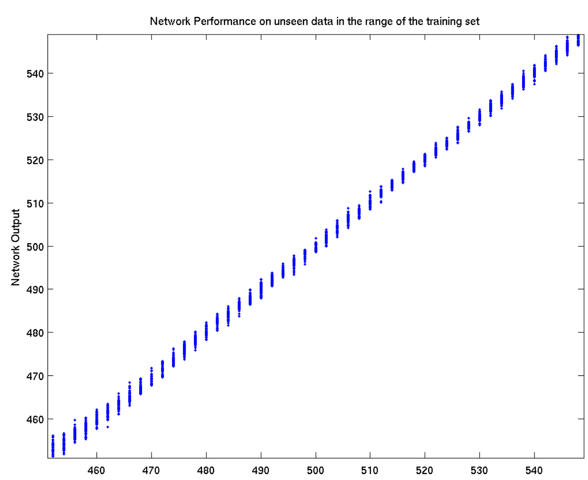


Figure 9.4: Performance of a neural network trained using the coefficients shown in Figure 9.3 on pulses with start times different from those in the training data set.

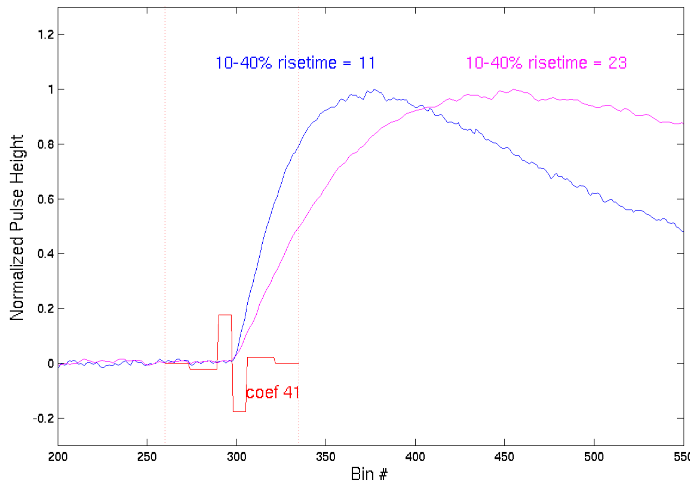


Figure 9.5: Schematic showing the calculation of the discrete wavelet detail coefficient 41, as shown in Figure 9.6. Two phonon pulses are shown with common start times but different risetimes. Coefficient 41 for each signal is calculated by correlating the filter (red) with the signal. When the risetime is smaller, the coefficient value has a more negative value, as can be seen in Figure 9.6.

representing the start time of these pulses. Figure 9.2 is a schematic showing the calculation of these coefficients. Figure 9.3 shows the value of several detail coefficient values for the set of training pulses, as a function of the start time bin. A neural network with 3 input nodes, 5 hidden nodes and 1 output node was trained to calculate the start time value using detail coefficients 38-40. Figure 9.4 shows the performance of the network on unseen data.

There are also coefficients that represent the rise time values, but there is a dependence upon the start time value. This dependence can be removed by using the network-derived value of the start time, and shifting the pulses to a common start time.

A DWT analysis of the shifted pulses showed that detail coefficients using *bior1.3* at a decomposition level 3 (see Figure 7.19) have a well-defined dependence upon the rise time. Figure 9.5 is a schematic showing the calculation of the coefficients. Figure 9.6 shows the value of several detail coefficients for the training pulses, as a function of the risetime value.

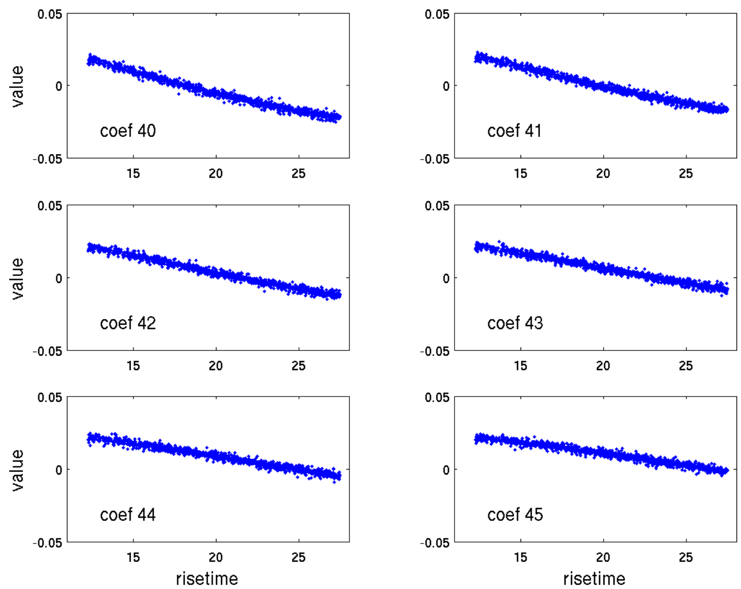


Figure 9.6: Selection of DWT coefficients of noisy training pulses, using the *bior1.3* wavelet at level 3, plotted against rise time values of test pulses that have been shifted back to a common start time using the network derived value of the start time. Again, a neural network can be trained to calculate the rise time given these coefficients.

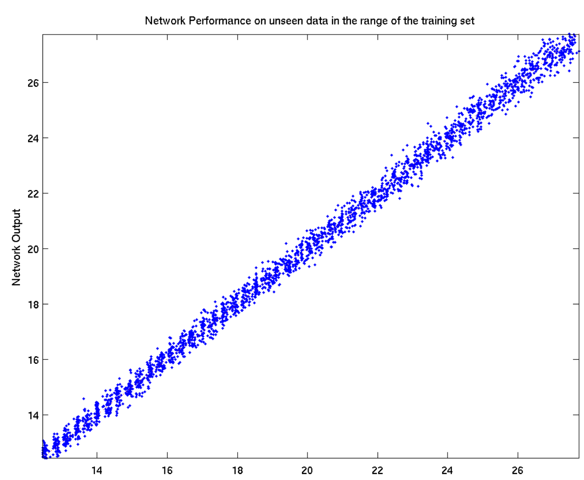


Figure 9.7: Performance of a neural network trained using the coefficients shown in Figure 9.6 on pulses with rise times different from those in the training data set.

A neural network similar to that used to calculate the start time was used to calculate the rise time. Network performance on unseen data is shown in Figure 9.7.

The work presented in this section is intended as a 'proof of concept' - that wavelet coefficients can be used to represent pulse parameters that we know are useful in performing discrimination between surface and bulk events. Hence it follows that the wavelet coefficients themselves should be useful in performing discrimination, although it is still unclear whether this information will give an improvement over using only timing RQs. Therefore we develop a cut based only upon phonon RQs before incorporating wavelet information.

9.3 Cut Based on Phonon RQs

Here we report on the first attempt at using a neural network approach on RQ values from CDMS Run 119 at Soudan. The cut described below has since been improved upon (see Section 9.6). It is reported on here since this analysis was developed blind, in parallel to the primary Run 119 analysis, and was designated as a secondary Run 119 analysis.

Firstly, we define our training and testing datasets. The data is selected from ^{133}Ba and ^{252}Cf calibration data, using the cuts cGoodEv & cQin & cQThresh & cVT & $7\,\text{keV} < E_r < 100\,\text{keV}$ (see Chapter 6). The remaining data is classified according to the following definitions:

- *Betas* are events from ^{133}Ba gamma calibration data that lie below the 5σ electron recoil band in ionization yield, and have an ionization yield $y > 0.1$.

- *Neutrons* are events from ^{252}Cf neutron calibration data that lie within the 2σ nuclear recoil band.

Subsequent discussion of betas and neutrons within this chapter will consider these definitions to be implicit.

Now that we have a collection of events that fall into one of two classes, we must select a subset of events with which to train the network, while the remaining subset will be used to test the network output. Since we wish to set the phonon timing cut in an unbiased way, it is imperative that our methodology allows the creation and 'freezing' of the cut before it's effectiveness on unseen data is observed. i.e. we wish to create the cut blind. This is one of the reasons that the evidence procedure is so useful here - without it we would not know if we were overfitting on the training set.

Since we already have a selection of open and closed data defined for other cuts on the ^{133}Ba calibration data, we used the same definitions to select our training set. That is, datasets with even SeriesNumbers form the training set, whilst datasets with odd SeriesNumbers form the testing data set. The situation is a little trickier with ^{252}Cf calibration data, since there was no defined open and closed datasets. Here, we select ^{252}Cf data with even EventNumbers as part of the training set, and take the odd numbered events to be in the testing data set. Now we have training and testing data sets of comparable size.

Three parameters were chosen with which to perform the cut; the phonon delay, phonon risetime and phonon partition. Two of these have been described previously, but for completeness here are the definitions:

Phonon Delay (pdelc) The time from the start of the ionization pulse to the time at which the primary phonon pulse has reached 20% of it's amplitude (the primary phonon pulse is the quadrant that measures the greatest phonon energy).

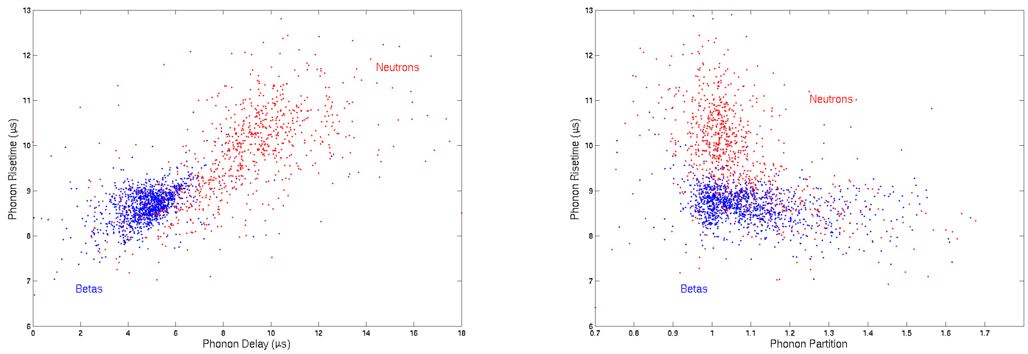


Figure 9.8: The parameters phonon delay, phonon risetime and phonon partition used to make the RQ-based cut, shown for training data in detector Z3.

Phonon Risetime (pminrtc) The time taken for the primary phonon pulse to rise from 10-40% of it's amplitude.

Phonon Partition (pfracc) The phonon energy in the primary quadrant divided by the phonon energy in the opposing quadrant. This provides a rough measure of radial position, since an event close to the center of the detector will have a phonon partition value close to one, whereas an event at a higher radial position will have a higher value.

Figure 9.8 shows the distribution of these values for training set beta and neutron events in a sample detector.

A multilayer perceptron network consisting of 3 input nodes, 8 hidden nodes and 1 output node, and using the evidence procedure, is then applied to the training data. The network inputs are the three phonon parameters, whilst the output is 1 when the event is a neutron, and 0 when the event is a beta. The resulting network effectively maps a point in the 3-dimensional input parameter space into a value q with values from 0 to 1, which represents the probability of a neutron being found at that location (the probability of finding a beta at that point is $(1-q)$). By allowing all events with a value of q greater

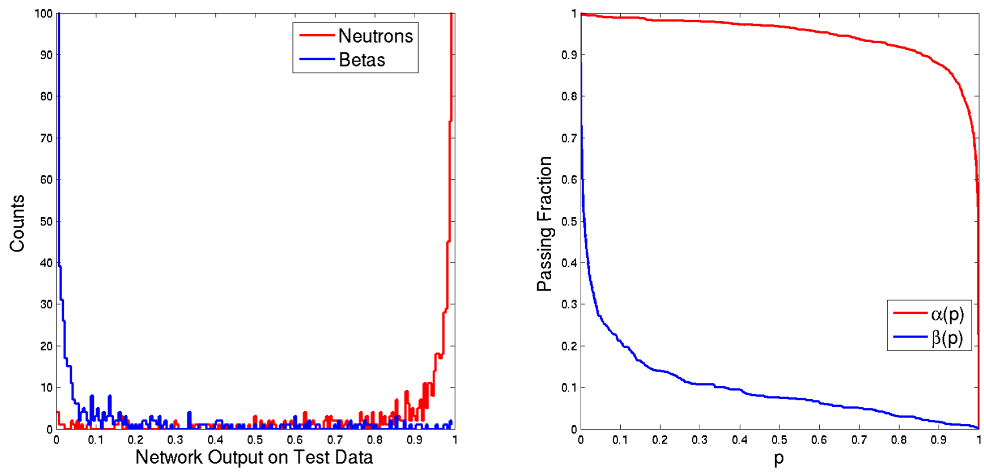


Figure 9.9: Network output values on test data and the resulting passing fractions of neutron and beta events as a function of the network output cutoff p. The network gives values close to 1 for neutron events, and values close to 0 for beta events. The cut is placed by choosing a value of the cutoff p which gives the best expected limit, as described in Section 9.1. Typically, $p > 0.9$.

than some cutoff p to pass the cut, we can tune p to create an optimal cut. i.e. We choose a value of p, which will pass some fraction $\beta(p)$ of the betas, and some fraction $\alpha(p)$ of the neutrons. We can therefore use (9.1) to choose the appropriate value of p.

Figure 9.9 shows the network output values for test data from a single detector, and the resulting passing fractions $\alpha(p)$ and $\beta(p)$.

The cut is set in this way for the good Ge detectors (Z2/3/5/9/11) resulting in neutron efficiencies of $\sim 80\%$ above 20keV on the training data. The training data results in 28 beta events leaking past the cut in the 5 good Ge detectors. Before the cut there are 5235 neutron events and 2728 beta events in the good Ge detectors. After the cut the same detectors have 3316 neutron events (63% passing fraction) and 28 beta events (1.0% passing fraction). Although neutron efficiencies of 80% are obtained at higher recoil energies, the total neutron passing fraction is 63% since the neutron efficiency drops to 40-50% at low

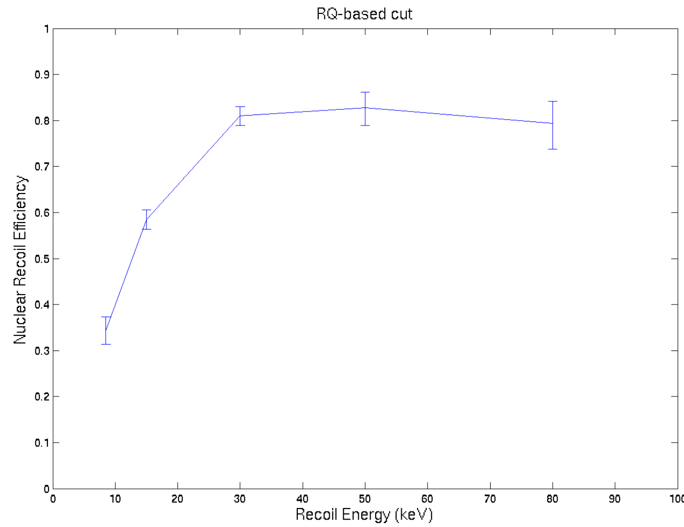


Figure 9.10: Efficiency of the RQ-based cut on 2σ nuclear recoil testing (unseen) data events summed over all 5 good Ge detectors (Z2/3/5/9/11). Error bars are statistical and do not reflect any systematics (see text).

energies, and there are more neutrons at lower recoil energies.

As described in Section 6.7.5, since we know the number of low yield events in the WIMP search data and the number of them that reside within the signal region, we can scale the timing cut's leakage value to an expected leakage upon the WIMP search data, resulting in an expected background of 0.5 ± 0.4 events in good Ge detectors from 7-100 keV (error includes both statistics and systematics).

The cut is frozen at this point and the efficiency and leakage calculated for the testing data. This results in the efficiencies shown in Figure 9.10 for unseen data, which are the same as the efficiencies obtained on the training data within statistics. Applying the cut to the testing data results in 48 beta leakage events, almost a factor of two greater than was obtained for the training data. This represents a 3σ fluctuation above the expected leakage. Subsequent analysis showed that there were two reasons for this.

Firstly, it is worse to miscatagorize a beta as a neutron than it is to miscatagorize a neutron as a beta, and the analysis did not account for this. A false positive nuclear recoil degrades our expected limit by more than a false negative nuclear recoil. This problem is discussed in Section 9.4.

Secondly, the value p that parameterizes the cut is typically > 0.9 since the cut is being set in a region of the parameter (RQ) space which is at the tail of the beta population. The neural network has the greatest generalization to unseen data when one is interested in the regions of parameter space with the highest data point density. Hence, the region where we set the cut in practice can suffer from a loss of generalization since it is dealing mostly with outliers, and the network training optimizes for dense regions of the data distribution. It was found that training on all data in a single energy bin for a given detector increased the statistics sufficiently to regain generality for the outliers.

Incorporating the number of beta leakage events seen in the testing data, the expected background estimate becomes 0.9 ± 0.6 events in good Ge detectors from 7-100 keV. The Run 119 primary timing cut has an expected background of 0.5 ± 0.4 events in the same energy range, so although the nuclear recoil efficiency is improved in this analysis the expected background is higher.

The testing data set has around 10% more events than the training data set, so this fluctuation is unlikely to be purely statistical. Timing analyses developed by other research groups using the same parameters did not see a significant systematic increase in the number of leakage events when comparing open to closed data. The conclusion is that there were other systematic effects in the neural network method including overfitting on the training dataset.

The timing cut discussed in this section was designated a secondary analysis prior to unblinding Run 119 data. That is, it was developed blind and frozen before unblinding in the same way as the primary analysis. Of the secondary analyses developed during this period for the WIMP search data, this analysis gives the lowest expected limit across almost all of the WIMP mass range. When this cut is applied to the WIMP search data, no candidate events remain in the 2σ nuclear recoil band between 7-100 keV, whereas the primary analysis saw a single candidate event at 10.5 keV.

9.4 Logistic Error Function with Risk

The analysis above has ignored one crucial piece of information: it's worse to miscatagorize a beta event as a neutron or WIMP candidate than it is to miscatagorize a neutron as a beta. Letting a single beta event cross the decision boundary of the neural network output changes the expected limit of the cut far more than it does to let a single neutron cross the decision boundary.

In order to account for the unbalanced nature of this classification we replaced the sum-of-squares term E_S in the error function (8.12) with a *logistic error function* based on a Bernoulli random variable for each target output:

$$E_L = -\sum_{n=1}^{d} \left(l_1 t^n \ln y^n + l_2 (1 - t^n) \ln(1 - y^n)\right) \tag{9.3}$$

where the network has a single output node and d data points are being fed into the network, t^n are the target output values (either 0 or 1), y^n are the actual network output values and l_1, l_2 define the relative risk of misclassification. When the target output is 1 (in our case a neutron), the error goes like $-l_1 \ln y$; when the target output is 0 (a beta) the error goes

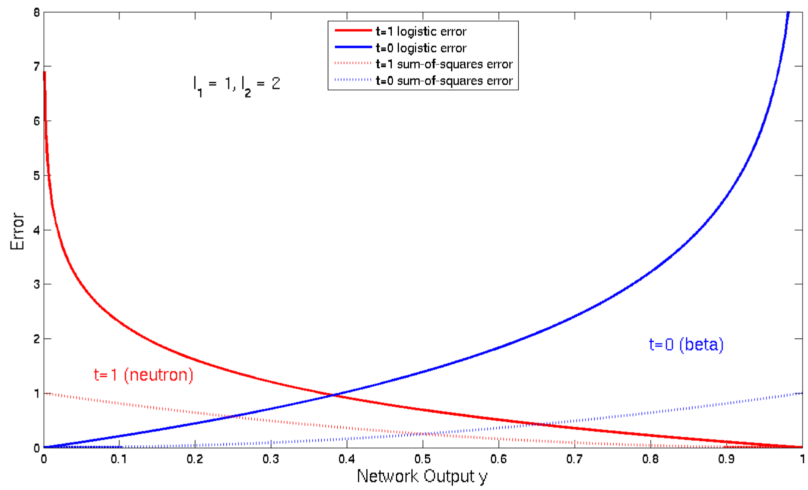


Figure 9.11: The logistic error function for a single data point in a binary classification problem. The error function for betas rises more quickly with network output since the risk of misclassifying a beta as a neutron is higher than the risk of misclassifying a neutron as a beta.

like $-l_2 \ln(1-y)$.

Therefore we can account for the unbalanced classification by making $l_2 > l_1$ which penalizes the network more for mistaking a $t = 0$ event for a $t = 1$ event than it does for mistaking a $t = 1$ event for a $t = 0$ event. Figure 9.11 shows the error as a function of the output for both target values.

The logistic error function also penalizes the network for getting an output value wrong by more than the sum-of-squares error does. As Figure 9.11 shows, the gradient of the error function with respect to the network output is greater in the logistic case.

The relative risk coefficients l_1, l_2 were set based by optimizing the expected limit for the classification problem at hand, and were used in the analysis in Section 9.6.

9.5 Discrete Wavelet Analysis of CDMS Phonon Pulses

Next we look at wavelet information from the same set of neutron and beta events that were selected for the analysis described in Section 9.3, so that we can add these parameters to the neural net analysis.

For each event, the phonon pulse from the primary quadrant was selected. The pulses were each scaled by the optimal filter-calculated amplitude in order that all the pulses be normalized to the same pulse height. Finally, each event was shifted back to a common start time using the inner electrode ionization pulse start time. Recall that the phonon delay used for the RQ based cut is the time from the ionization pulse start to the time at which the phonon pulse has reached 20% of its final amplitude. Shifting the phonon pulses back allows the wavelet analysis to focus on their shape relative to the ionization pulse start time.

All of the major wavelet families shown in Section 7.6 were used to perform discrete wavelet transforms on the CDMS pulses and the coefficients were examined in order to see which might allow determination of the recoil type. The conclusions were: firstly, analysis with most wavelets does show some level of difference between beta and neutron pulses, and this difference is situated around the leading edge of the pulses.

Secondly, it is clear that the pulse shape differences that exist between surface and bulk events are such that a decomposition level between 3 and 5 is the most effective scale to see these differences. Since each time bin in a CDMS phonon pulse is $0.8\,\mu$s, successive wavelet coefficients taken at level j, step by $(0.8 \times 2^j)\,\mu$s. Typical phonon pulse risetimes are 10-20 μs so we might have expected that a level of ~ 4 would best represent this feature.

The *sym3* and *rbio1.3* wavelets also have good signal to noise but show a greater difference between the leading edge wavelet coefficients for each recoil type.

Figure 9.12 shows some example discrete wavelet decompositions for a variety of wavelet families and decomposition levels, along with the wavelet filter used to calculate the coefficients. At the leading edge of the neutron and beta phonon pulses there is a corresponding change in the discrete wavelet coefficient values. The *haar* and *db2* wavelet analysis have good signal to noise at the leading edge but the neutron and beta wavelet coefficients are too similar to perform discrimination based on these values.

Empirically it was found that the *rbio1.3* and *sym3* wavelets at a decomposition level of 4 give the greatest discrimination power between recoil types in Ge ZIPs. Figure 9.13 shows the value of the leading edge DWT coefficients for all the preselected calibration events in detector Z2 using these wavelets. Use of multiple coefficients from the same wavelet decomposition was examined, but it was found that other wavelet coefficients from around the leading edge do not provide additional discrimination power. These coefficients either provide the same information as the leading edge coefficients, or do not provide useful discrimination information.

Although the neutron and beta populations shown in Figure 9.13 are well separated, there are a small number of beta events that lie in the neutron region. It would not be possible to make a very effective cut for WIMP search data on these parameters alone in their present state.

The wavelet coefficients are derived from raw pulses, and as such have no ability to account for the known variation of pulse shape with detector position. Even a ZIP detector with TESs that exhibited no variation of T_c values across the phonon quadrants would

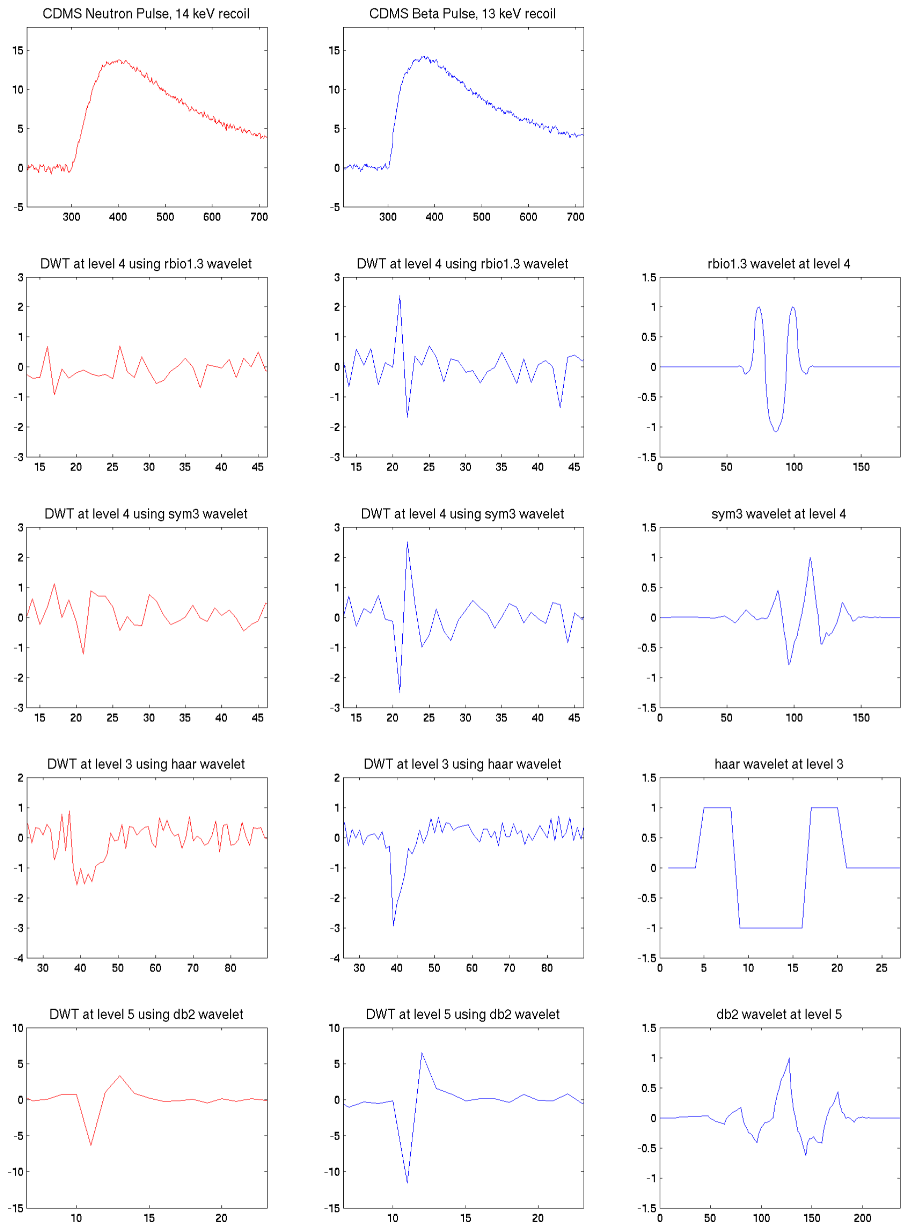


Figure 9.12: Sample Discrete Wavelet Transformation coefficients for example neutron and beta phonon pulses taken from CDMS calibration data. The leading edge of the primary phonon pulse is shown and the wavelet coefficients have been similarly windowed in order to show the leading edge behavior. The scales of the high-pass filters as shown in the third column are correct to obtain the coefficients at each level from the original signal, as per Eqn. (7.22).

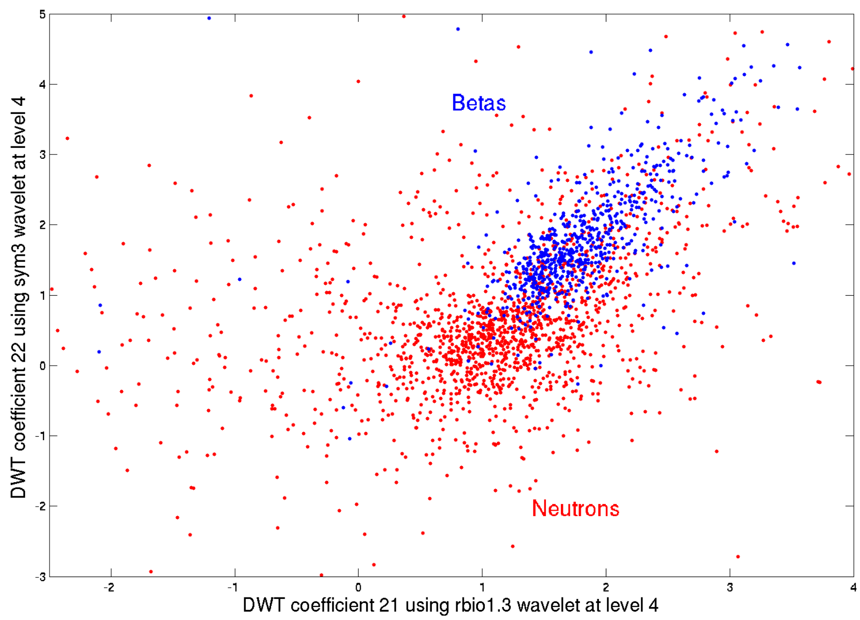


Figure 9.13: Leading edge DWT coefficients using the *rbio1.3* and *sym3* wavelets at a decomposition level of 4. The beta and neutron populations are generally well separated, but there are a significant number of beta outliers situated within the neutron population.

exhibit pulse shape variation across the crystal due to geometric effects. The majority of these beta outliers have position (partition and delay) values that place them at outer crystal radii, and therefore we might expect that some of them would move closer to the beta population after a phonon position correction (see Section 4.7).

These outliers appear no matter which wavelet we use, since the actual phonon pulse shape is clearly neutron-like for these events. It can also happen for low-yield events below ~ 10 keV that the ionization pulse start time is poorly calculated (due to poor signal to noise for ionization pulses), leading to a phonon pulse in our analysis that is shifted to a position not congruent with the other pulses. In this case the wavelet coefficient will not represent the leading edge of the pulse.

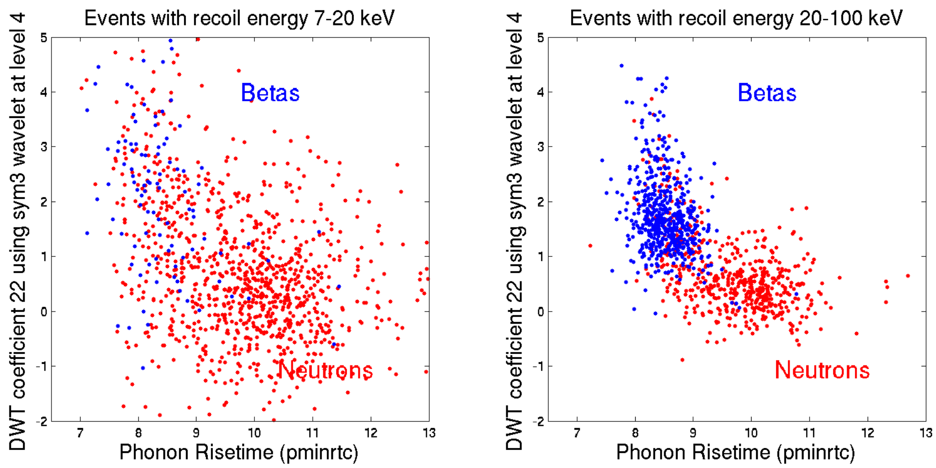


Figure 9.14: Phonon risetimes and leading edge wavelet coefficients using *sym3* wavelet, for analysis events in detector Z2. The wavelet representation suffers from an abundance of outliers but events with recoil energy above 20 keV give greater separation between the two populations.

Within the timeframe of this work it was not possible to use the computer farm for a complete phonon position correction of the wavelet parameters, however a naïve position correction was performed using the correction table calculated for the phonon risetime. This did not aid in discrimination between recoil types using the wavelet values.

Since the neural network approach allows the use of many input parameters, it may be possible to combine the wavelet information with RQ values in order to identify these outliers within the network. Figure 9.14 shows the *sym3* leading edge wavelet value with the (position-corrected) risetime RQ value. The risetime value provides greater discrimination power between recoil types than the wavelet value for events 7-20 keV in recoil energy. However, the discrimination power of the two parameters becomes comparable when we consider events in the range 20-100 keV in recoil energy.

Since the neural network framework is able to operate with many parameters, we create a timing cut using these wavelet parameters and a collection of RQ parameters. The use of the Automatic Relevance Determination (ARD) framework will allow us to observe how much each network input contributes to the output.

9.6 Final Analysis

This section describes an approach to the surface event cut using a combination of RQ values and discrete wavelet coefficients. The data is preselected from Soudan Run 119 ^{133}Ba gamma calibration and ^{252}Cf neutron calibration data as described in Section 9.3. A multilayer perceptron network with 7 input nodes, 14 hidden nodes and 1 output node was used to train the data using the evidence procedure (Section 8.4). The logistic error function described in Section 9.4 was used with values $l_1 = 1$, $l_2 = 2$. The number of hidden nodes to use was determined by increasing the number while observing the effect on the network's performance on training data. Having more hidden nodes than is necessary only impacts the training time and not the network output. 14 hidden nodes were used since increasing to 15 did not affect the network's performance on the training data.

Two of the network inputs were taken from discrete wavelet analysis using the *rbio1.3* and *sym3* wavelets. These were both derived from phonon pulses in the primary channel that were normalized and shifted to a common start time using the ionization pulse start time, as described in Section 9.5. These wavelet-derived inputs were:

- Coefficient 21 taken from a discrete wavelet analysis of the scaled and shifted primary phonon pulse using the wavelet *rbio1.3* at a decomposition level of 4 (see Figures 9.12 and 9.13).

- Coefficient 22 taken from a discrete wavelet analysis of the scaled and shifted primary phonon pulse using the wavelet *sym3* at a decomposition level of 4 (see Figures 9.12 and 9.13).

The remaining network inputs were CDMS RQ values, as follows:

Phonon 10-40 Risetime (pminrtc) The time taken for the primary phonon pulse to rise from 10-40% of its final amplitude.

Phonon Radial Delay (rdel) The radial position as determined by the relative delays of the four phonon pulses.

Phonon Partition (pfracc) The phonon energy in the primary quadrant divided by the phonon energy in the opposing quadrant. This provides a rough measure of radial position, since an event close to the center of the detector will have a phonon partition value close to one, whereas an event at a higher radial position will have a higher value.

Ionization Inner Electrode Start Time (QIst) Start time, relative to the beginning of the digitized trace, of the ionization signal in the inner electrode.

Recoil Energy (pric) Recoil energy, using inner ionization electrode quantities (see Section 6.5.9).

Figure 9.15 shows the values of these parameters for the set of neutrons and betas in detector Z2.

The open and closed datasets for this analysis were determined randomly, by assigning each event to be in the open or closed data set with equal probability. This is to avoid systematic effects that may be present in the primary analysis' choice of open and closed

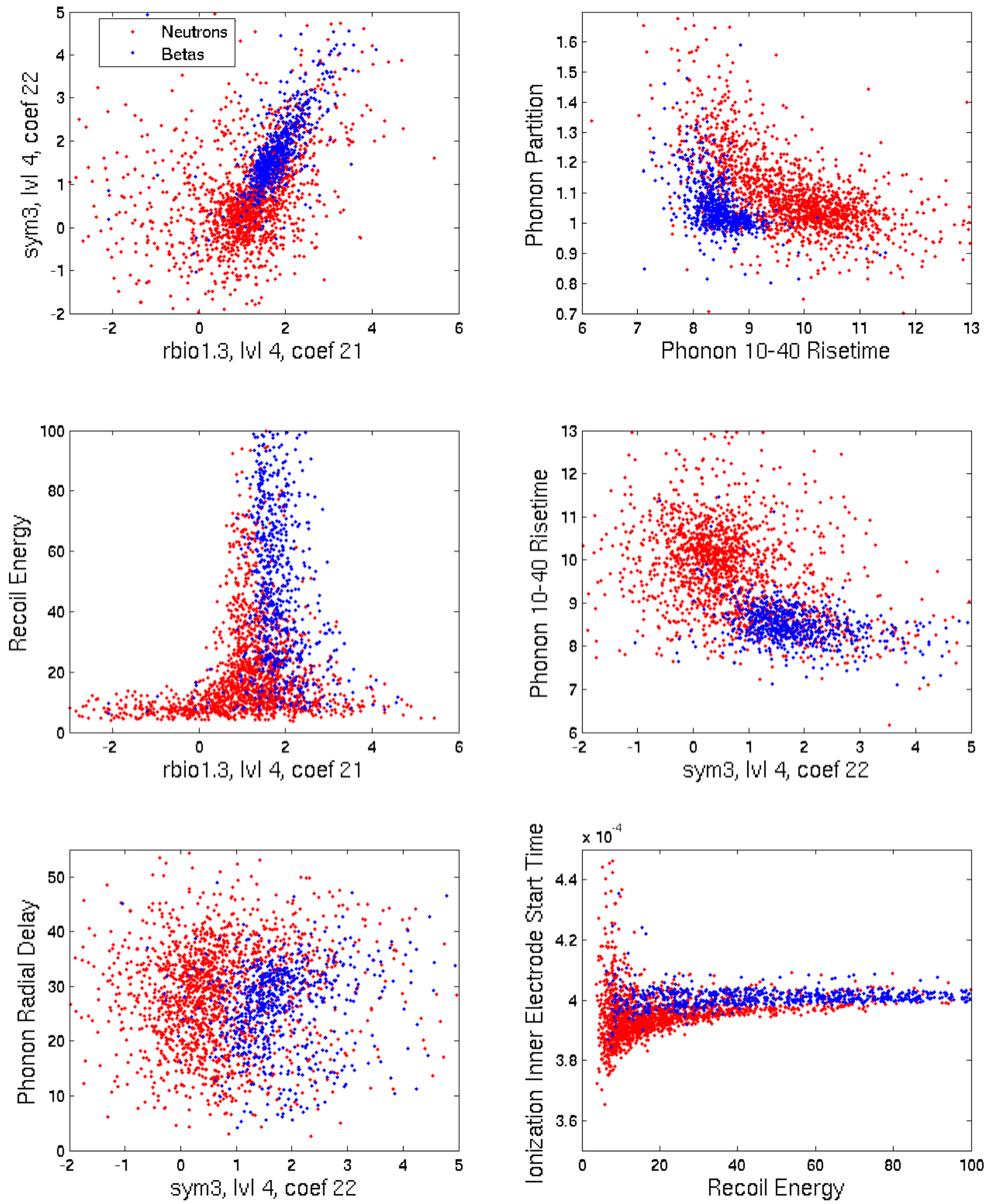


Figure 9.15: Values of the 7 parameters used as inputs to the multilayer perceptron network for detector Z2 (see text for parameter descriptions). All events in Z2 from 7-100 keV are shown - betas from gamma calibration are shown in blue, neutrons from neutron calibration are shown in red.

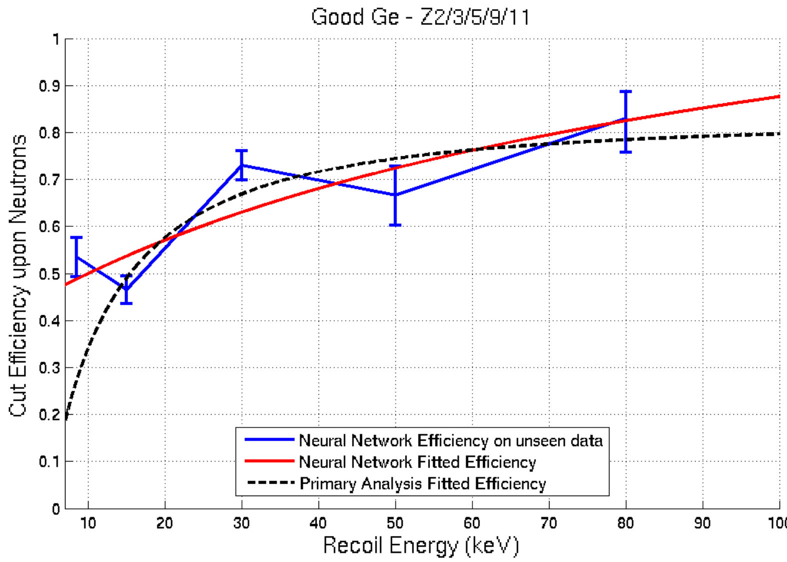


Figure 9.16: Efficiency of the neural network surface event cut upon neutrons from the closed (unseen) dataset for the good Ge detectors. 2σ error bars shown. The neutron efficiency of the primary analysis surface event cut is shown for comparison.

Detector	Open Beta Leakage	Closed Beta Leakage	Total Beta Leakage
Z2	5 / 344	5 / 313	10 / 657
Z3	4 / 594	3 / 594	7 / 1188
Z5	1 / 519	2 / 474	3 / 993
Z9	1 / 678	3 / 692	4 / 1370
Z11	1 / 361	0 / 343	1 / 704
Z2/3/5/9/11	12 / 2496	13 / 2416	25 / 4912

Table 9.1: Number of beta events leaking past the neural network surface event cut for the open and closed datasets in the recoil energy range 7-100 keV on good Ge detectors.

data sets. The Open dataset of input values for each parameter, for each detector, were scaled to have a mean of zero and unit variance. The same scaling parameters were then applied to the Closed data. Open data from 7-100 keV recoil energy in each detector was used to train MLP networks with the evidence procedure in three bins: 7-20 keV recoil, 20-60 keV recoil, 60-100 keV recoil. There are therefore three MLP networks per detector.

The cut for each detector was set by minimizing the expected limit based on the performance of the network on the training data, as described in Section 9.1. The networks were then applied to the closed data on each detector.

For each MLP network and for all detectors, the input hyperparameter values α_i have the same values to within a factor ~ 2. This implies that each network input gives a comparable contribution to the output values. The reduced network performance that resulted when an input was removed and the network retrained also supports this conclusion. Removal of the wavelet inputs results in a $\sim 20\%$ drop in overall neutron efficiency, and beta leakage numbers remain the same within statistics.

Figure 9.16 shows the efficiency of the surface event cut upon the closed neutron data for the good Ge (Z2/3/5/9/11) detectors. The efficiency for neutrons has been fit to a function of the form $\epsilon = a + \frac{b}{(E_r - c)^d}$ where ϵ is the efficiency and E_r is the recoil energy. Table 9.1 shows the number of beta events that leak past the cut in the good Ge detectors in the open and closed data.

The expected number of background events, by which we mean WIMP search single-scatter events in the 2σ nuclear recoil band after application of the surface event cut, is calculated as follows. Before application of the surface event cut, we calculate for WIMP search data the ratio of single-scatter events found in the 2σ nuclear recoil band to the total number of low-yield multiple scatters. This tells us the number of events that contribute to the signal region in the absence of the surface event cut, for a given number of low-yield multiple scatters. Therefore the expected contribution to the signal region can be determined by examining the number of low yield multiple scatters that remain after application of the surface event cut.

For this analysis, the expected background due to betas is 0.3 ± 0.2 events for 7-100 keV recoil energy. Recall that the primary Run 119 analysis had an expected background due

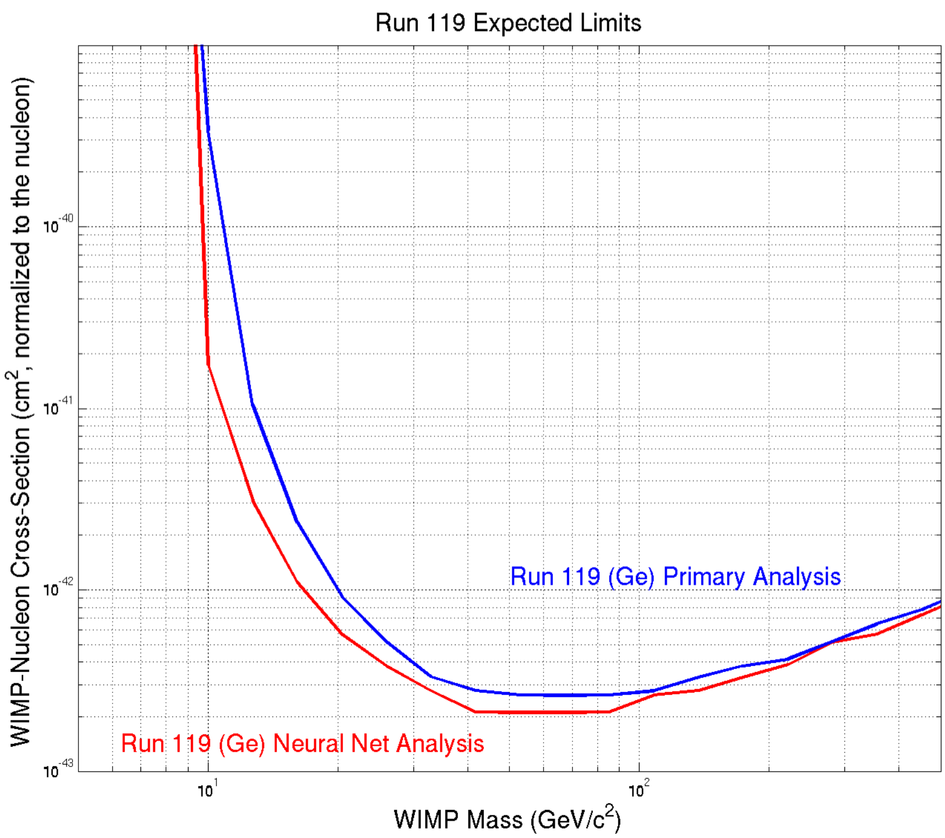


Figure 9.17: Expected spin independent 90% upper limits on the WIMP-nuclear cross-section for the neural network analysis described in Section 9.6 (red), and for the Run 119 primary 7-100 keV$_r$ analysis (blue).

to betas of 0.5 ± 0.2 events over the same energy range, by passing a total of 51 beta events in the ^{133}Ba calibration data.

Figure 9.17 shows the expected WIMP cross-section upper limits for the neural network and wavelet analysis, and for the Run 119 primary analysis. The calculation of the expected WIMP limit follows the Optimal Interval method described in Section 6.9, and uses standard assumptions for the galactic WIMP halo. The expected number of background events at a given WIMP mass is obtained by fitting an exponentially falling spectrum of background events to the multiple scatter events that pass the surface event cut in the WIMP search

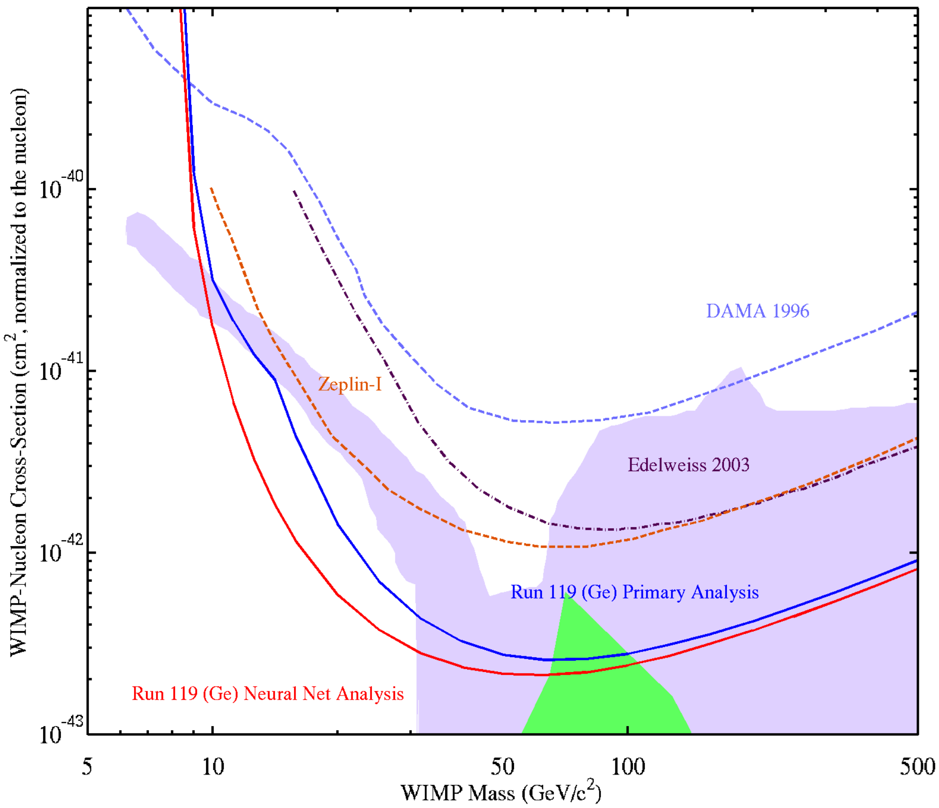


Figure 9.18: Spin independent 90% upper limit on the WIMP-nucleon cross-section obtained from the neural network analysis described in Section 9.6 (red). Also shown are the Ge limit from the Run 119 primary 7-100 keV$_r$ analysis (blue) and limits from Zeplin-I (dashed salmon) [92], Edelweiss (dashed dk. red) [93] and DAMA (dashed blue) [94]. The two block regions are predictions on supersymmetry taken from [96] (lavender) and [97] (green).

data. The only difference between the neural network and primary analyses for the purposes of setting a dark matter limit is the surface event cut. All other cuts and efficiencies remain as described in Chapter 6.

When this analysis is applied to the WIMP search data no events remain in the signal region. The spin-independent upper limit on the WIMP-nucleon cross-section that results is shown in Figure 9.18. The Optimal Interval method of Section 6.9 is again used along with standard assumptions for the galactic WIMP halo. The limit set by the primary analysis includes one WIMP candidate event that passed all cuts.

9.7 Conclusions

The expected number of Run 119 background events leaking into the WIMP signal region is $\sim 40\%$ smaller for the neural network analysis compared to the primary analysis. The neural network analysis' signal region efficiency is comparable to that of the primary analysis above 10 keV, and is twice as efficient as the primary analysis below 10 keV. This results in an expected limit for the neural network and wavelet analysis that is 20-40% lower than the expected limit for the primary analysis.

The comparative differences in expected limits become larger when the experiment becomes limited by backgrounds. The subsequent CDMS II run at Soudan will be using 19 Ge detectors (compared to the 6 Ge detectors used in Run 119), so the expected background due to local beta contamination will most likely exceed 1 event. The improvement in expected limit that results from a $\sim 40\%$ reduction in expected background will be of even greater importance in this situation.

When applied in a neural network approach, the discrete wavelet detail coefficients from the leading edge of the phonon pulse are clearly providing additional discrimination power against surface events. Since the hyperparameters for the network inputs have similar values, this implies that the two wavelet inputs are as useful as, but no more useful than, the phonon risetime and phonon partition values traditionally used to reject surface events in CDMS.

This is not entirely surprising, as Sections 9.2 and 9.5 show, the wavelet coefficients that we use for the analysis are simple filters that provide some measure of the leading edge slope. The wavelet approach is also susceptible to error when the ionization pulse start time is uncertain, as can happen with low energy, low yield events.

A more effective way to use wavelets to study the phonon pulse shape would be to use the continuous wavelet transform, since this would provide information at all scales and time offsets, and would be independent of the phonon pulse start time. This would, however, generate a great deal of information and some way to reduce the number of wavelet coefficients obtained from a signal may be necessary. The neural network approach would likely shed a great deal of light on this problem. It was not possible in the timescale of this thesis to fully explore the potential application of the continuous wavelet transform.

The study of phonon signals other than the primary channel's should also prove fruitful, since quantities like the phonon partition show that the relative timing of the four quadrants can prove useful in surface event discrimination.

The use of Multilayer Perceptron Networks with the evidence procedure and some measure of relative classification risk has proven very powerful in performing a surface event cut while blind to the results. The methods described in the last two chapters allow the setting of an optimized cut on a arbitrary number of input parameters based on minimization of the expected dark matter limit.

No doubt there are undiscovered CDMS parameters that will provide even greater surface event discrimination that those known to date. Whatever these turn out to be, and however many of them there are, the neural network methodology described here should be able to make an optimized cut upon them (assuming the training time is manageable).

However, the neural network approach is more sensitive to differences in parameter distributions between the calibration and WIMP search datasets. If the input parameter distributions within the WIMP search dataset are different from the calibration dataset values (and with more input parameters this is increasingly likely), the neural network trained upon the calibration data will not generalize well to the WIMP search data. If these shifts in distributions are dependent upon recoil energy for example, the data can be split into several recoil energy bins and each can be trained by a separate neural network (as was done for the analysis in Section 9.6). This way, each neural network is less susceptible to classification error as a result of subsequent shifts in the input parameter's distributions.

Chapter 10

Soudan Low Background Counting Facility (SOLO)

The SOudan LOw Background Counting Facility (SOLO) has been in operation at the Soudan mine in Minnesota since March 2003. The SOLO screening facility was constructed by Brown University, in conjunction with PNNL, the University of Minnesota and the Soudan Mine, and has been, and is currently being used to screen materials for the Majorana, CDMS and XENON experiments.

The facility uses HPGe detectors and shielding previously deployed in Double Beta Decay (DBD) Experiments at the Homestake Mine, although recently, in conjunction with the University of Florida, one of the DBD detectors has been upgraded to a newer, larger, HPGe detector.

Figure 10.1: The SOLO experiment (red box) is located at the back of the MINOS hall in the Soudan Underground Laboratory.

10.1 SOLO Installation

The SOLO experiment was built in January 2003 at the back of the MINOS hall in the Soudan Underground Laboratory (see Figure 10.1). The enclosure is built from ∼10 tons of Doe run Pb (2"×4"×8" bricks) with activity of 50 Bq/kg. Additional low-activity lead is used for the innermost layer (i.e. the lead that was to be closest to the detectors), which is ∼200 year old German lead with an expected activity of 50 mBq/kg.

The entire enclosure is built upon a (very strong!) steel table measuring approximately 4'×4'. One side of the tabletop contains a platform that can be raised and lowered by a drill mechanism. This platform is used as the door, such that lowering the platform allows access to the chambers containing the detectors.

The initial SOLO setup contained the detectors 'TWIN' and 'Diode M', which are high-purity Ge (HPGe) diode detectors with crystal masses of 1.05kg and 0.6kg respectively. Both detectors had previously been used in Double Beta Decay Experiments at Homestake, so their expected sensitivity was already known.

Figure 10.2 shows the (partially constructed) setup of the chambers when SOLO was put together in Jan 2003. The two detectors are housed in separate chambers, each surrounded by at least 12" of Pb, the innermost 2" of which is the low ^{210}Pb-activity (German) Pb. A 2" wall of German Pb separates the chambers, each of which houses a detector. The entire enclosure is sealed up with two airtight layers of 50μm thick white mylar, and radon gas is purged using boiloff gas from a LN dewar fed in through the shield. At present when samples are introduced no airlock is employed, but gamma lines due to Rn disappear less than 3 hours after the shield is closed and the N_2 gas purge (at 3 cfh) begins (airborne Rn rates in the Soudan lab vary from 200 Bq/m^3 in December to 600-800 Bq/m^3 in June).

A simple data acquisition system is used, consisting of Spectroscopy Amplifiers with 2μs shaping times, and pulse height ADCs connected to a PC running acquisition software. Spectra are integrated in 4 hour intervals, and transferred automatically to a web accessible archive. Both detectors have independent data acquisition systems.

10.2 Preliminary Experiences

Upon first operation in June 2003, the TWIN spectrum showed $>$ 10 events/keV/kg/day up to $\sim$ 1 MeV, the same detector having operated at $<$ 1 event/keV/kg/day in the past (see Figure 10.3). At the same time, Diode M observed a background rate $\sim$ 1 event/keV/kg/day

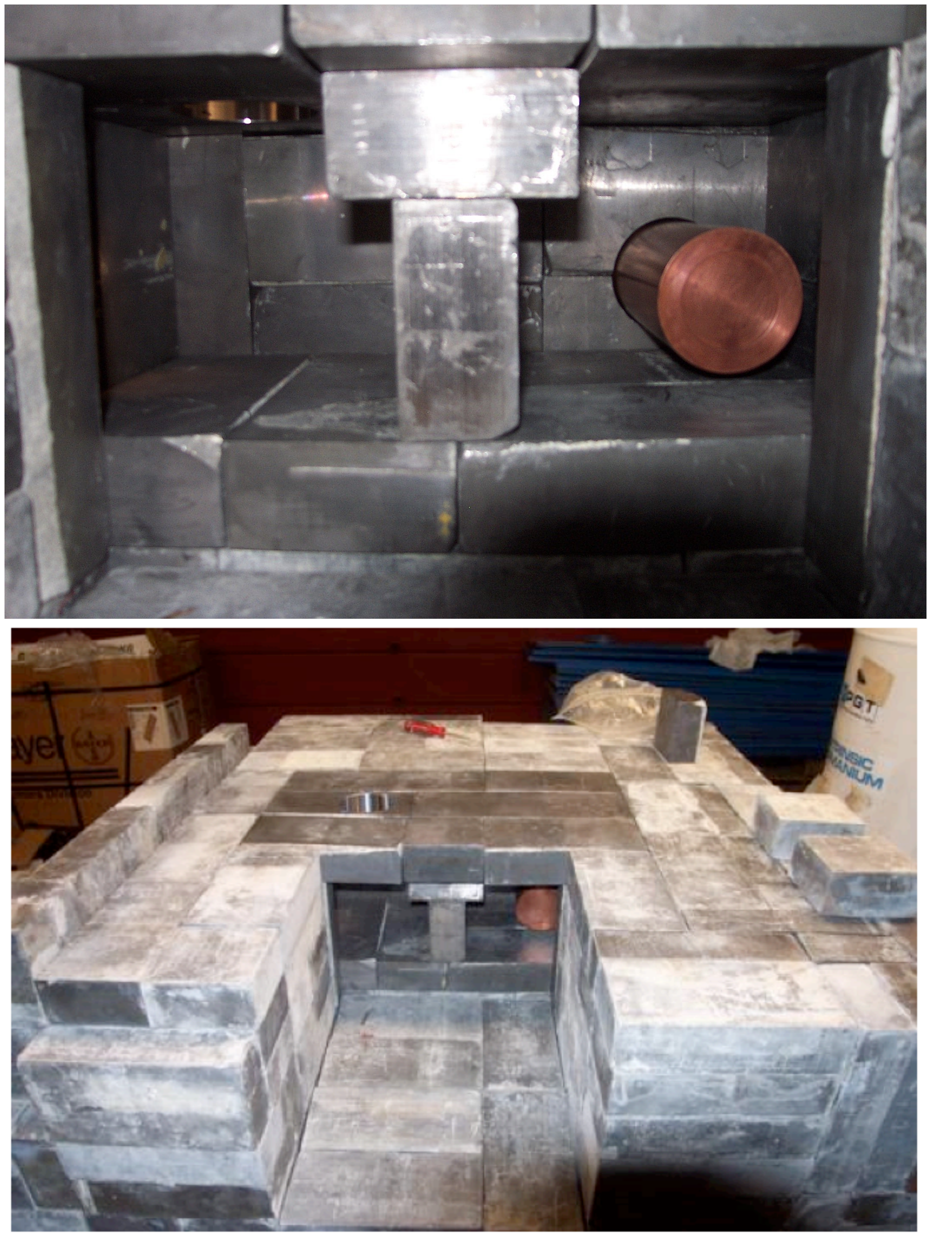


Figure 10.2: The initial chamber setup of the SOLO experiment (partially constructed) showing the open door. The Diode M detector is shown on the right, and the TWIN detector although not pictured, protrudes into the left hand chamber through the circular hole in the ceiling. The innermost layer of both chambers is the low-activity German lead, which includes the inner side of the door that faces into the chamber when the door is closed.

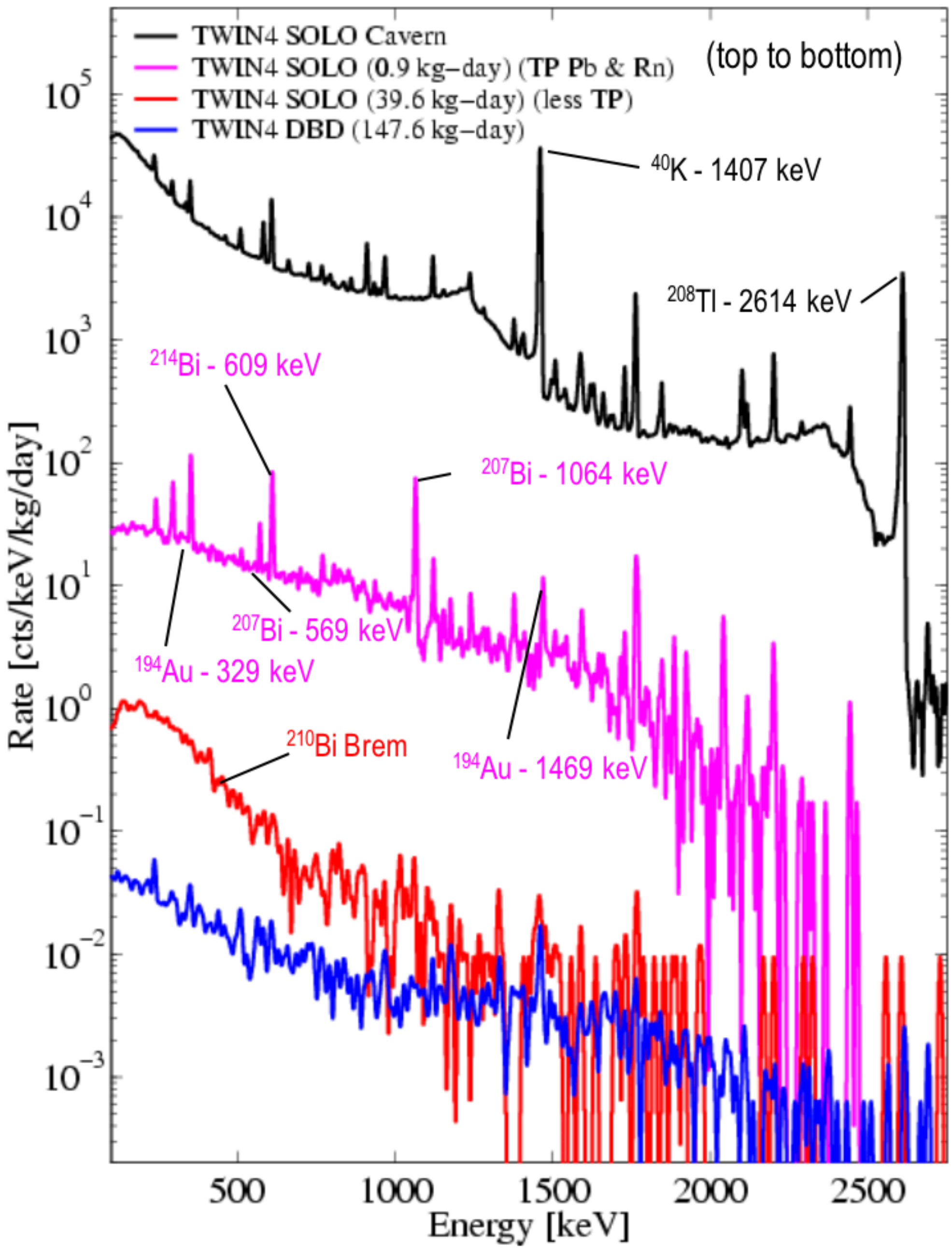


Figure 10.3: Spectra obtained from the 1.05kg TWIN detector, from top to bottom: data taken with door open (i.e. no Rn purge) (black), data taken with Rn purge active but with 'TP' bricks present (see text) (magenta), data taken with TP bricks removed (red), and data taken at Homestake during earlier operation of the detector with same shielding (blue).

up to $\sim$ 500 keV, suggesting that the TWIN detector was subject to a source that Diode M was not. Analysis of the TWIN spectral lines in the data indicated the presence of two unexpected isotopes: ^{207}Bi (31.6 years half-life) and ^{194}Hg (444 years half-life). The relative strengths of the lines at different energies appeared to indicate the presence of 3.5-4.5cm of Pb between the source contaminant and the detector. Both isotopes are very rare, but are known to be spallation products arising from high energy protons on Pb.

In order to test the hypothesis that contaminated bricks were causing the high background levels on TWIN, the upper section of the SOLO Pb shield was disassembled to examine the bricks surrounding the chambers. Bricks stamped with the letters 'TP' were found directly above the TWIN chamber, behind the low activity German Pb layer, in a location consistent with our calculations. We understand that these bricks possibly came from Fermilab, where they may have been used as a proton beam stop. Unfortunately, at some stage they must have been combined with the gamma shielding stock of Pb at Soudan. Removal of these 'TP' bricks substantially reduced the background on the TWIN detector, as can be seen in Figure 10.3 (magenta and red lines).

In May 2004, SOLO was counting several blocks of polyethylene for the CDMS experiment, in order to establish the intrinsic levels of U/Th/K. The polyethylene data showed a large number of unexpected gamma lines (Figure 10.4). The polyethylene spectrum was $\sim$5 events/keV/kg/day above background across all energies. (On closer inspection the no–sample backgrounds also appeared elevated.) When the polyethylene lines could not be matched to the patterns from any known combination of radioactive contaminants, it was realized they were characteristic of gamma cascades arising from thermal neutron capture in Ge. The low energy neutron flux coming from the rock at Soudan was insufficient to be

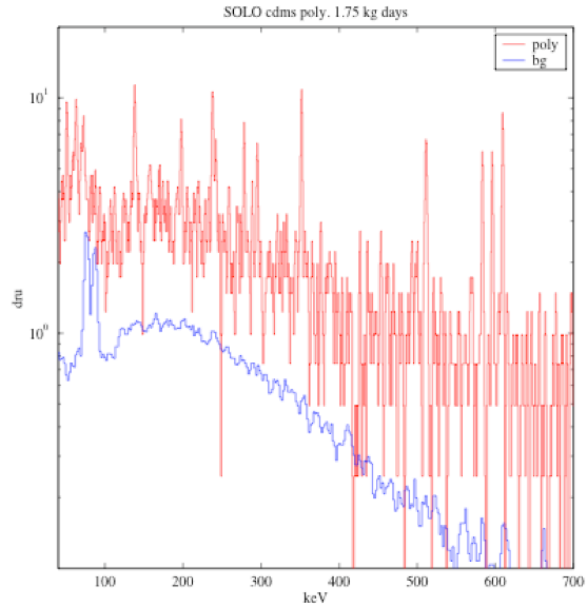


Figure 10.4: Elevated count rate in TWIN due to moderation of neutrons by polyethylene sample (red), compared to background in the absence of a sample (blue).

generating such pronounced features, however, it became clear that the polyethylene was moderating neutrons from a calibration source (10^5 n/s) had been placed in storage 20ft from the SOLO experiment. Needless to say, the neutron source was moved to another storage location.

The TWIN detector is currently offline due to a failure in its front–end FET. A low background/low noise FET is currently being prepared by PNNL as a replacement. In conjunction with the University of Florida, a 2kg HPGe detector (named 'GATOR') was installed in the shield in June 2005, and is now performing screening alongside Diode M.

10.3 Background

As Figure 10.3 shows, the SOLO background below 1 MeV is 5-10 times greater than when these detectors were in operation at Homestake, while the background above 1 MeV is very similar to the levels achieved previously. The shape of the excess background is consistent with that expected from ^{210}Pb bremsstrahlung. As a general rule of thumb, a brem spectrum with a rate at low energies (< 100 keV) of 1 event/keV/kg/day would arise from Pb with an intrinsic ^{210}Pb level of 1 Bq/kg. The German Pb liner has an intrinsic activity of <50 mBq/kg so it is unlikely that this is the cause of the elevated background. However, the non-ancient lead spent >10 years underground in non-airtight containers with background airborne Rn levels between 200-800 Bq/m^3, so it is entirely plausible that the brem arises due to Rn plating on the Pb bricks.

Hence in April 2005 the surfaces of the inner Pb liner were cleaned using dilute nitric acid. Unfortunately, this intervention has had little effect on the level of this background. We now believe that the brem could be arising from the ancient Pb bricks themselves. This would have arisen during recasting when the bricks were inadvertently contaminated with ^{210}Pb.

10.4 GATOR Installation and Shield Reconstruction

GATOR is a 2kg HPGe detector purchased direct from Canberra, along with a data acquisition system. The upper section of the chamber was removed and rebuilt in April 2005 in order to install the GATOR detector in the chamber previously occupied by the (non-functioning) TWIN detector (and to perform the nitric etching described above). This resulted, however, in two complications.

Firstly, the GATOR detector is considerably larger than TWIN, which required the chamber to be increased in size. At the time there was an insufficient amount of ancient

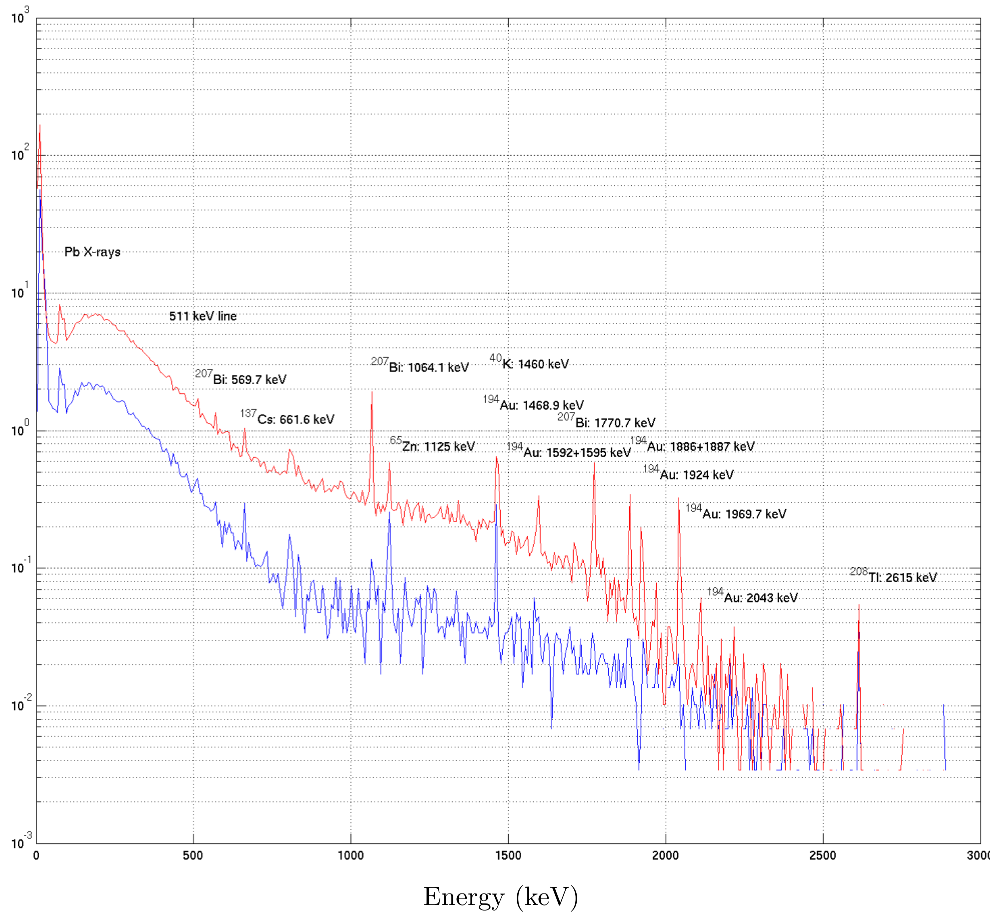


Figure 10.5: Spectra from the 2kg GATOR detector before (red) and after (blue) the November 2005 repair operation, when all TP bricks were removed from the shield and two 50Bq/kg bricks on the inside of the GATOR chamber were replaced with ancient bricks. Vertical axis shows counts/keV/kg/day.

lead to accommodate the inner layer of the new, larger chamber, so two ‘regular’ (50Bq/kg) lead bricks were used to form part of the inside chamber wall. Secondly, the bricks were naturally moved around in the process, which led to some TP bricks that were previously lower in the stack being placed close to the chambers.

Thus, upon first operation, the GATOR chamber saw an elevated low energy background level (due to proximity of the non-ancient lead) and a collection of ^{194}Hg and ^{207}Bi lines (although these appeared to be attenuated by several inches of lead). A hint of the ^{207}Bi

Figure 10.6: Partially reconstructed GATOR (left) and Diode M (right) chambers during the November 2005 repair operation.

line at 1064keV also appeared in the Diode M data.

A repair operation was therefore planned to once and for all, remove all of the TP bricks from the stack, and to replace the regular lead on the inside of the GATOR chamber with ancient lead bricks. This did not occur until November 2005, although Diode M was able to continue counting samples in the interim. The operation resulted in the removal of 140 'TP' bricks from the stack, which were forevermore banished from the land.

Figure 10.5 shows the GATOR spectrum before and after the November 2005 repair operation. Figure 10.6 shows the partially reconstructed GATOR and Diode M chambers.

10.5 Sample Counting Results

Over the past two years SOLO has screened many samples. The TWIN detector unfortunately was unable to provide screening, since shortly after the polyethylene/neutron source counting described above, it's front-end FET failed. Diode M has henceforth provided the majority of the sample screening results; GATOR is also now providing screening as of Dec 2005.

Below we describe results from two sample counting runs. The procedure for calculating sample activity upper limits is based on lines at 511keV and 585keV for ^{232}Th chain contamination, 352keV and 609keV for ^{238}U chain contamination, and 1461keV for ^{40}K contamination.

The number of counts above background S is given by $S = N - B$, where N is the number of counts in the sample data and B is the number of counts in the background data. The signal S can also be expressed as $S = \alpha T \epsilon$, where α is the source activity, T is the livetime and ϵ is the peak detection efficiency. ϵ is defined as the ratio of detected peak events to total number of decays, which is obtained via Monte Carlo simulation of the SOLO geometry. Values of ϵ are shown in Table 10.1.

Given that the number of counts N is Poisson distributed, the source activity for a given

Element	Energy (keV)	ϵ
^{232}Th	511	3.3×10^{-4}
^{232}Th	582	2.2×10^{-4}
^{238}U	352	9.0×10^{-5}
^{238}U	609	1.6×10^{-4}
^{40}K	1461	5.6×10^{-4}

Table 10.1: Peak Detection Efficiencies (see text) of the primary lines used to set limits on ^{232}Th, ^{238}U and ^{40}K contamination, determined from Monte Carlo simulation.

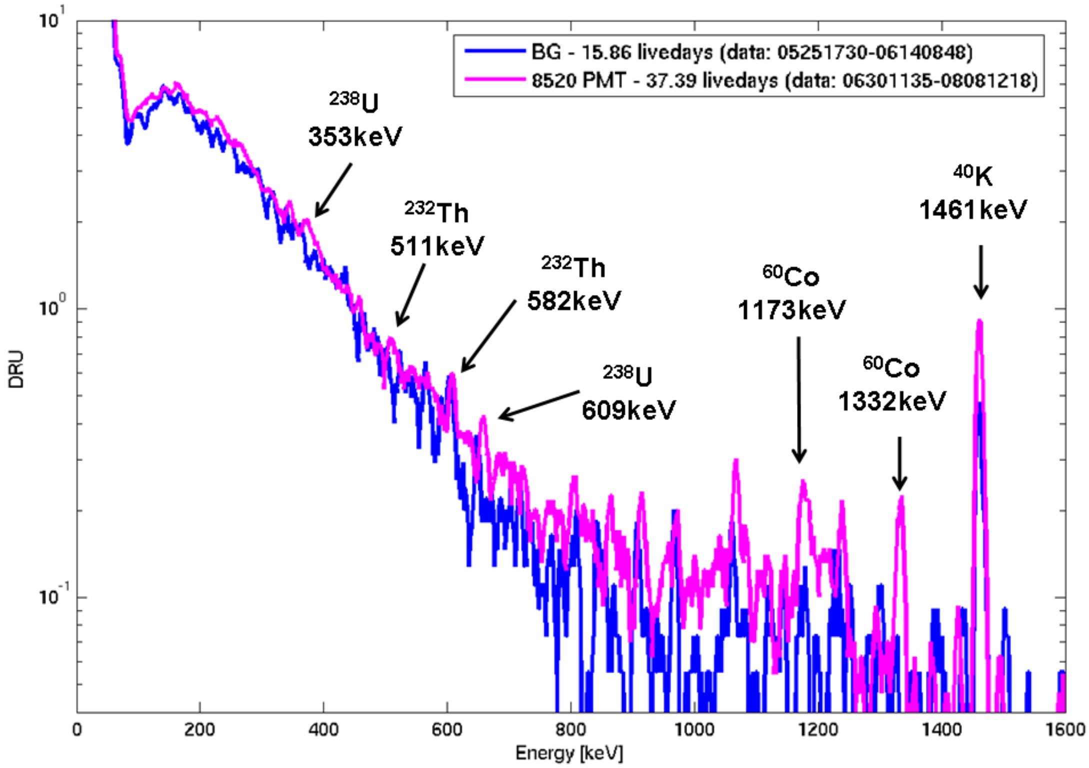


Figure 10.7: Results from sample screening of 2 Hamamatsu PMTs. The y-axis shows the 'differential rate unit' (DRU) = counts/keV/kg/day.

confidence level can therefore be determined.

Hamamatsu R8520 PMTs

2 Hamamatsu PMTs were screened for the XENON collaboration using Diode M, with a total livetime of 37.4 livedays (22.4 kg days). Figure 10.7 shows the results of the counting.

The resulting upper limits for activities at the 90% confidence level were calculated to be 6/18/5 mBq per PMT for ^{232}Th/^{238}U/^{40}K.

With Diode M for sample runs of $\sim$2 weeks, sensitivities better than 0.1 ppb level for ^{238}U/^{232}Th, and 0.25 ppm level for ^{40}K have been achieved.

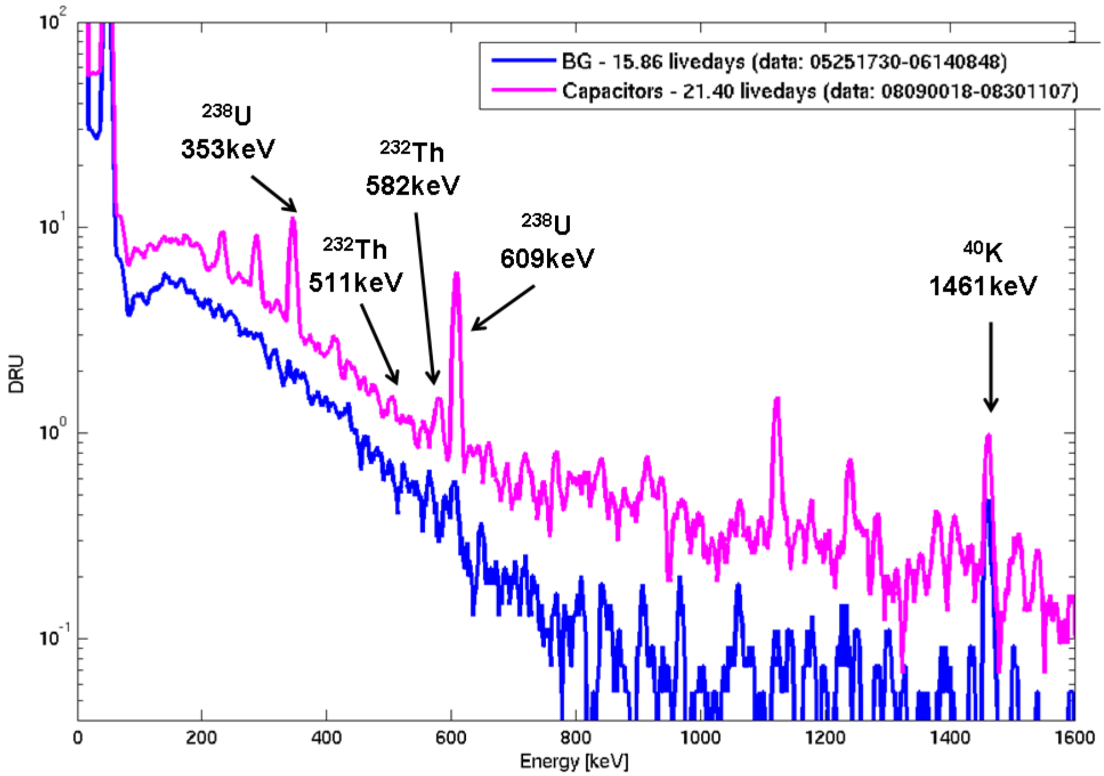


Figure 10.8: Results from sample screening of a collection of 135 capacitors. The y-axis shows the 'differential rate unit' (DRU) = counts/keV/kg/day.

135 Capacitors

A collection of 135 capacitors were screened using Diode M, with a total livetime of 21.4 livedays (12.8 kg days). Figure 10.8 shows the results of the counting.

The majority of the activity of this sample comes from the U/Th chains. The resulting upper limits for activities at the 90% confidence level were calculated to be 48/462/12 mBq for ^{232}Th/^{238}U/^{40}K, for the whole batch.

Appendix A

Analysis Software Documentation

A.1 DarkPipe

In a nutshell, Darkpipe takes the raw data as is output from the DAQ, and creates a collection of 'Reduced Quantity' (RQ) files which contain values calculated from the raw event information. RQ values are uncalibrated in quantities like energy; timing values are usually correct in some units.

As well as the raw input file (which can be gzipped or not), Darkpipe also reads a selection of auxiliary files which contain further diagnostic information from the DAQ. If some information does not pertain to a specific single event, it's in the aux files. Examples are samplerate, pretrigger values, charge bias values, thresholds, etc. More details are givin in the Auxiliary Files section below.

The most important routines in DP are:

- DP_config_v10 - contains all the hard coded parameters used by DP that are not found in the aux files

- Dark_Pipe_Control - this is the main loop. When you run Darkpipe, you run this script.

- Dark_Pipe_v10 - does all the hard work. Dark_Pipe_Control is just a wrapper that handles most of the file I/O, but Dark_Pipe_v10 contains or runs all the routines that calculate and save the RQ values.

Another important routine is readEvents. This is a Matlab mex file, written in C and compiled to make readEvents.mexglx. The routine is responsible for reading the DAQ data format and outputting a Matlab array that contains the same information. This is usually done by using the netServer protocol, which allows raw data that may, or may not, be gzipped, to be read from across the network. Hence a raw data file can be loaded by a machine that does not direct access to the raw data directory.

Stublists

The stublist files are used to store values that might change on a run by run basis. These include such things as the input and output data locations and which detectors to run. They can also be used to change file locations when running on different networks via the getLocation function.

Auxiliary Files

- .info file - Contains details on which detectors are part of the current run, their digitizer samplerates, pre/posttrigger values, trace lengths.

- .isr file - Whenever the DAQ writes a GPIB command, it is echoed into this file. The isr file is therefore hard to parse by eye since it contains hex commands. Darkpipe extracts this information along with timing data.

- .dmm file - Contains SQUID offset and trigger threshold values.
- gpib_states_changed.log - records information when the state of each detector changes. Used to extract time since last flash.

Running DP Manually

This is usually done when new code is being tested. One can checkout a copy of Darkpipe on the same system as the running version, and run it in an interactive way in order to examine changes or find bugs. Of course, you could also run the 'running' version by hand if you so choose, but should not edit the files since all updates are to be handled via CVS updates.

1. Start up Matlab in DP_root
2. Create a string variable, pathraw, which is a full path to the file you wish to load

   e.g. pathraw = '/cdmstmpf2/Soudan_data/150425_0732/150425_0732_F0001'
3. Run Dark_Pipe_Control

A useful way to debug is to enter a 'keyboard' command into a particular Darkpipe routine, run Darkpipe as outlined above, and it will stop in the debug mode at the keyboard command. This will allow you to examine variable values and see what's happening at that point.

Adding a new RQ

The process of adding a new RQ value is fairly straightforward, and requires more work to calculate the new value (i.e. write the new code) than it does to merely tell Darkpipe to start storing a new variable.

Let's say we want to create a new RQ that calculates some zip-specific quantity, called *UsefulNumber*.

1. Add a corresponding entry to BigList. When this is run (does not run as part of Darkpipe), it creates a 'namefile' .mat file. This contains a list of all the RQ names that Darkpipe is to calculate. Therefore, to add our RQ we would need to ensure that BigList will now have an extra entry for UsefulNumber. This is stored in the Namelist as UsefulNumberZ1, UsefulNumberZ2, etc. It's not until the main DP loop that Darkpipe drops this naming convention. Take a look at Biglist to see how it's done for an existing RQ and follow that prescription.

2. Add code (to Dark_Pipe_v10 or one of it's subroutines) that stores the above RQ quantitiy in the appropriate RQ structure. For ZIP quantities, you'd store it in submZ(iZ).m(iev,Z(iZ).UsefulNumber), where iZ is the ZIP number, iev is the event number as it appears in the file (i.e. 1-500). Again, compare to existing RQs.

   For Trigger quantities, you'd store the value (or array of values) in something like submTG(:,NewTGRQName).

   Similarly for Veto quantities, you'd use submVT(:,NewVTRQName).

   The way this works is that Darkpipe assigns a unique integer value to each RQ name. So you're actually accessing a column of the submTG matrix, you just don't have to worry about getting the correct value - Darkpipe assigns them based on the Namelist.

3. Now when you load up CAP on one of the output RQs (or merged RRQs) with the new variable, it should automatically create an FCCS function for the new RQ.

RQ file structure

An RQ file contains the following variables (** = TG, VT, Z1, Z2, etc.):

- DP**_Version - Darkpipe code version. Set in DP_config_v10
- DP**_start - Date/time the Darkpipe processing that created this file started
- DP**_stop - Same as above, for time processing completed
- Names** - List of all the RQ names (each row is a string containing an RQ name)
- **_RQVersion - An entry containing the RQ version, stored on an event-by-event basis. This value comes from the one in BigList
- m** - Array containing all of the RQ values. It will have as many columns as Names** has rows; each column corresponding to the relevant RQ quantity. eg. m**(:,23) will return all the RQ values for the quantity in row 23 in Names**

ZIP RQ files also have the variable

- **T_filename - Stores the filename of the charge template file used in the processing

DarkPipe Code Overview

```
Variable 'pathlist' must be set prior to calling Dark_Pipe_Control
```

```
Dark_Pipe_Control
Reads and Parses stublist file
Set up output path and RQ structures
Read aux files

 Dark_Pipe_v10

  Set up file to be read

  DP_config_v10

   Load up all the preset variables and structures
   Optimal filter noise threshhold values, setup phonon templates
   Set pulse window
   Define regions of interest (baseline section, pretrigger, posttrigger)
   Define 2-pole Butterworth filter
   Define risetime thresholds of interest
   Load charge templates as defined in template_list.txt
   Setup veto/trigger mask/time structures
  end DP_config_v10

  Setup netserver path
  Load raw data file using readEvents to number of events (num_events)
  Allocate memory for RQ matrices
  num_events is broken into parcels (eg. every 50 events) for memory management

  Loop over parcels (eg. 1-10)

   Load up events in current parcel using readEvents (into 'ev' structure)
   Allocate memory for RQ sub-matrices
   Store RQs: EventNumber, SeriesNumber, EventTime,
   Store RQs: LiveTime, TimeBetween, TrigTime

   # TRIGGER ANALYSIS SECTION

   Read TrigInfo value for each tower (bits are rearranged from raw data->RQ)

   map_all_ISR - Store bias values taken from the .isr file (also stores times)

   Initialize all TGTime and TGMask variables to their error values

   Loop over each event in parcel

    calcTrig
     Extract trigger bit information from history buffer (trigger/veto mask/time)
     Bit manipulation to make RQs appear in a more logical form than raw data
     vetorq - returns RQ values for interrelated Trigger/Veto RQs
    end calcTrig

   end loop
```

```
calcZZflt (filter creation routine) - Only if filters do not exist

flt_GetTraces

 Set number of noise events
 Read in noise candidate events (eg. first 500 events in run)
 Generate MinMax noise cuts (to exclude pulses)
 Generate Pileup cuts
 Make list of events that pass noise cuts

flt_CalcFlt
 Take noise traces and make FFTs of events
 Generate template pulse and form the PSD from it
 Calculate amplitude of fft noise
 Calculate conj of noise with template
 Store all filter information in the filter file and save it as a mat file

end calcZZflt

# ZIP ANALYSIS SECTION

Loop over each event in parcel (eg. 1-50)

 Loop over detectors

  calcZbasic - baseline, baseline std dev & saturation calculations
  calcZbsub - create baseline subtracted pulses
  calcZpoptflt - calculate amplitude and delay for phonon pulses
                      using optimal filter
  calcZqoptfiltX - calculate amplitude and delay for charge pulses
                       using optimal filter
  calcZFitsC - perform multiparameter fit (F5) for charge channels
  calcZint - find integral values of phonon channel pulses
  calcZrtft - find phonon pulse risetime values

  Store values taken from .dmm file & gpib states file

 end loop over events

end loop over detectors

# VETO ANALYSIS SECTION

Initialize veto RQs with error values

Loop over each event in parcel

 calcVT - fills veto RQs with pretrigger and posttrigger times & masks
 calcNearVT - find times of largest Q and P signals relative to veto hits
 calcVTtrace - calculates amplitudes & times from raw veto pulse
```

```
    end

    Move data from sub-matrices into main matrices

   end loop over parcels

   Save all the main matrices as the RQ files: VT, TG, Z1, Z2, etc.

   Update .status file associated with this raw data file
                          to note that processing is complete

  end Dark_Pipe_v10

end Dark_Pipe_Control
```

A.2 PipeCleaner

Overview

PipeCleaner takes the Darkpipe output (RQ files) and performs the following operations:

1. All the RQ files are merged together to form a single VT file, a single TG file, etc. eg. if we processed a 15k dataset we would have 30 raw data files, each of which would produce each of the RQ file types. The corresponding RRQ files are created my concatenating the m?? matrices in the RQ files. xxxxxx_xxxx_F0001_DPTG.mat + xxxxxx_xxxx_F0002_DPTG.mat + ... = xxxxxx_xxxx_DPTG.mat etc. These merged files are usually referred to as RRQ files, even though they are technically just merged RQ files.

2. A new file consisting of RRQ (relative reduced quantities) is created via the application of calibration values and other corrective operations. These are some of the most important analysis quantities such as the recoil energy, the ionization yield and the phonon risetime. The filename has the form xxxxxx_xxxx_DPR1.mat. The exact

contents of the file are described below.

3. Subject to the state of a flag value in the PipeCleaner call, a further calculation may be performed in order to correct the phonon quantities based on position. This utilizes .mat files that have been created for each detector describing nearest neighbour parameters. This process creates a file with the form xxxxxx_xxxx_DPM1.mat that contains all the phonon corrected quantities.

When running at Soudan, PipeCleaner also performs two diagnostic routines:

- A small set of plots (6) is created for each detector and sent to a web directory. This is intented for fast feedback.

- A more detailed trend diagnostic is performed that uses a template file created by a seperate piece of code. This template file contains reference information for many quantities for all zips. The current dataset is compared with the reference data and K-S test significance values are produced. In this way, each dataset is given a rating that reflects the extent to which it is similar to the reference dataset.

Datum Files

Datum files contain the following variables (z is the number of zips that info is stored for):

- epg (1xz) - keV needed to create electron-hole pair. 3 in Si, 3.82 in Ge.

- overall (1xz) - scaling to get from PxOFeV →px (x = A,B,C,D,T)

- pcorrect (1xz) - flag for phonon correction (linearlization)

- qadjust (1xz) - flag for charge correction (linearization)

- qia (1xz) - scaling to get from QIOFvolts → qi
- qipos (10xz) - values needed to perform the charge position correction (Note)
- qix (1xz) - qi crosstalk values (see PC_first_pass.m)
- qoa (1xz) - scaling to get from QOOFvolts → qo
- qox (1xz) - qo crosstalk values (see PC_first_pass.m)
- relative (4xz) - Box plot balancing factors for each detector.

The datum file should be set with a best guess at the start of a run, using older detectors as a guide. Once the detectors are suitably neutralized, a new datum can be created with the energy calibration and box plot balancing factors set more accurately. The final values will need to be created along with the phonon position correction at the end of the run during reprocessing.

Running PipeCleaner Manually

This is usually done as new code is being tested. One can checkout a copy of the Darkpipe package on the same system as the running version, and run it in an interactive way in order to examine changes or find bugs. Of course, you could also run the 'running' version by hand if you so choose, but should not edit the files since all updates are to be handled via CVS updates.

A useful way to debug is to enter a 'keyboard' command into a particular Darkpipe routine, run Darkpipe as outlined above, and it will stop in the debug mode at the keyboard command. This will allow you to examine variable values and see what's happening at that point.

To run PipeCleaner Manually:

1. Start up Matlab in DP_root

2. PipeClean(outdir,Series,dgZ,doDiag,doCorrectP,storeFilePath,numfiles) where

    - outdir : (string) directory in which to output RRQ files
    - Series : seriesnumber as an integer
    - dgZ : array of detector numbers to pipeclean
    - doDiag : flag to do diagnostic plotting. default = 0 (no plots)
    - doCorrectP : flag to do phonon correction. default = 1 (do correction)
    - storeFilePath : (string) full path to where the RQ files reside
    - numfiles : file number of the highest numbered file (needed for merging memory management)

Adding a New RRQ

Adding a new RRQ value is fairly straightforward. The easiest way to proceed is to add the name of the new RRQ to the 'rrq_ZIP' variable within PC_first_pass.m, then simply calculate the named RRQ quantity from within this code (or from a subroutine). Taking a look at PC_first_pass.m should make this clear.

RRQ File Structure

The DPTG, DPVT, DPZ1, DPZ2, etc. files have the same structure as their respective RQ files. The R1 and M1 files contain only the following variables (** = R1 or M1):

- DP**_Version
- Names**
- m**

The description of these variables match the ones described for RQs above.

PipeCleaner Code Overview

```
PipeClean(outdir,Series,dgZ,doDiag,doCorrectP,storeFilePath,numfiles)

 Use numfiles to determine whether the list of files needs to be split
 If we need to split up the list into batches

  Create additional RQ directories, incrementing the seriesnumber by 1 each time
  Move RQs to new directories and rename them

 end

 Loop over number of batches

  Create output directory if it doesn't exist

  mergeRQfiles (concatenates individual RQ files and creates resulting RRQ files)
  PipeCleaner

   If xxxxx_xxxx_DPR1.mat file already exists, move it to xxxxxx_xxxx_DPR1.mat.old
   Startup CAP with unique (random) filelist name

   PC_first_pass

    Use the datum_list.txt file to find out which datum to use for this series
    Load the datum file
    Initialize arrays
    Loop over dgZ

     Determine delays in x,y,z
     Calculate Luke factor
```

```
    Find saturated events
    Apply calibration values to charge amplitudes
    Use optimal filter quantities for non-saturating charge events
    Use F5 fit for saturating charge events
    Perform qi position correction

    Calculate phonon amplitudes from optimal filter quantities
    Calculate box plot balancing values
    Calculate xppart, yppart
    Calculate recoil energy
    Calculate risetime values (eg. ptrt = PTr40 - PTr10)

   end loop over dgZ

   Save RRQs in xxxxxx_xxxx_DPR1.mat file

  end PC_first_pass

  Reinitialize RRQs by switching filelist in CAP

  If doDiagPlots=1 and at Soudan

   Create Diagnostic plots directory
   AutoDiagPlots (creates basic diagnostics, sends to web directory)

  end

  if doCorrectP=1

   PC_correct_phonons

    Use the datum_list.txt file to find out which datum to use for this series
    Load the datum file
    Initialize arrays

    Loop over dgZ
     phonon_corr (performs position correction)
    end

    Save phonon corrected RRQs in xxxxxx_xxxx_DPM1.mat file

   end PC_correct_phonons
  end
 end
end

If this is a background set
 Blind_Data
end
If at Soudan
 diagnostics
```

```
 end
end
```

Appendix B

Optimal Filter Code

This is an adaptation of the routine that is actually used in the DarkPipe code, but essentially performs the same operation. The MATLAB routine below uses *time2psd* and *psd2fft*, both of which perform the operation that their names suggest whilst being careful to keep the units correct.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% OPTIMAL FILTER ROUTINE %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Parameters that need to be defined:
%
%% trace - signal to be fit (1 x M)
%% template - template of signal, normalized to amplitude of 1 (1 x M)
%% samplerate - samplerate of the signal in Hz
%% noise - power spectral density of noise at frequencies (1 x 0.5M)
%%   samplerate/M : samplerate/M : samplerate/2

M = length(trace);

% Fourier Transform of signal
trace_ft = ft(trace');

% Calculate noise power spectral density J
noise = [inf noise];        % DC noise assumed to be infinite
noise_fft = psd2fft(noise');
J = abs(noise_fft).^2;
```

```
% calculate fft of template
[ftp ftp_angle freq_tp] = time2psd(template', samplerate);
template_fft = psd2fft(ftp,ftp_angle);

% calculate optimal filter
optimalfilter = conj(template_fft)./J;

%%%%%%%%%%%%%%
%%% DELAY %%%%
%%%%%%%%%%%%%%

% Calculate amplitude estimator
t_hat = real(fi(trace_ft .* optimalfilter/sqrt(samplerate)));

% delay is given by maximum value of the estimator
[amp,delay] = max(t_hat);

% delay is now relative to start/end of window.
% i.e. a pulse that starts -N bins before the template start will have
% a delay value that is M-N. A pulse starting +N bins after the
% template start will have a delay of N.

% Change delay value so that the value is relative to pulse start
delay = -1 + delay - M*(delay>M/2);

%%%%%%%%%%%%%%%%%
%%% AMPLITUDE %%%
%%%%%%%%%%%%%%%%%

clear j         %so j = sqrt(-1)
phase_factor = exp(2*pi*j*(0:M-1)'/M*delay);

% calculate top and bottom of amplitude estimator
Atop =  sum(real(trace_ft .* phase_factor .* optimalfilter)) / sqrt(samplerate);
Abottom = sum(real(template_fft .* optimalfilter));

% calculate amplitude
A = Atop/Abottom;
```

Appendix C

Wavelet and Neural Network Code

C.1 Calculation of DWT Coefficients

This MATLAB function returns approximation and detail coefficients from the discrete wavelet transformation of a signal.

```
function [d,a] = getdwtcoef(x,wname,lvl,coef)
% CALCULATE SPECIFIC DETAIL AND APPROXIMATION COEFFICIENTS FROM DISCRETE WAVELET
% TRANSFORMATION
%
% function [d,a] = getdwtcoef(x,wname,lvl,coef)
%
% - d is the detail coefficient(s) requested
% - a is the approximation coefficient(s) requested
%
% - x is the signal to be analyzed
% - wname is a string containing the wavelet family name (eg. 'rbio1.5')
% - lvl is the decomposition level
% - coef is a coefficient value, or a range of continuous integers.
%     eg. coef=5, coef=10:20 are OK.
%   The number of coefficients = floor(length(x)/2^lvl) + N/2
%       where N is the length of the wavelet filter to be used (eg. N=6 for
%       sym3, N=2 for haar, etc.)

switch wname
    case 'sym1'
        Hi_D = [-0.707106781186548    0.707106781186548];
```

```
    case 'sym2'
        Hi_D = [-0.48296291314469     0.836516303737469  -0.224143868041857 ...
                -0.129409522550921];
    case 'sym3'
        Hi_D = [-0.332670552950957    0.806891509313339  -0.459877502119331 ...
                -0.135011020010391    0.0854412738822415  0.0352262918821007];
    case 'sym4'
        Hi_D = [-0.0322231006040427  -0.0126039672620378   0.0992195435768472 ...
                 0.297857795605277   -0.803738751805916    0.497618667632015  ...
                 0.0296355276459985  -0.0757657147892733];
    case 'rbio1.3'
        Lo_D = [ 0                    0                     0.707106781186548 ...
                 0.707106781186548    0                    0];
        Hi_D = [ 0.0883883476483184   0.0883883476483184   -0.707106781186548 ...
                 0.707106781186548   -0.0883883476483184   -0.0883883476483184];
    case 'rbio1.5'
        Lo_D = [ 0                    0                     0                 ...
                 0                    0.707106781186548     0.707106781186548 ...
                 0                    0                     0                 ...
                 0                    ];
        Hi_D = [-0.0165728151840597  -0.0165728151840597    0.121533978016438 ...
                 0.121533978016438   -0.707106781186548     0.707106781186548 ...
                -0.121533978016438   -0.121533978016438     0.0165728151840597 ...
                 0.0165728151840597   ];
    case 'bior1.3'
        Lo_D = [-0.0883883476483184   0.0883883476483184    0.707106781186548 ...
                 0.707106781186548    0.0883883476483184   -0.0883883476483184  ];
        Hi_D = [ 0                    0                    -0.707106781186548 ...
                 0.707106781186548    0                     0                  ];
    case 'haar'
        Hi_D = [-0.707106781186548    0.707106781186548];
    otherwise
        disp('Unsupported wavelet');
        d=0;a=0;
        return;

end

if ~exist('Lo_D')
    % get low pass from quadrature mirror filter condition
    Lo_D = Hi_D(end:-1:1);
    Lo_D(2:2:end) = -Lo_D(2:2:end);
end

for n=1:lvl

    % Pad function so that filter can be applied across entire signal
    I = getSymIndices(length(x),length(Lo_D)-1);
    y = x(I);

    % Compute coefficients of approximation.
```

```
    z = conv2(y,Lo_D,'valid');
    % downsample by factor of 2
    CA = z(2:2:length(x)+length(Lo_D)-1);

    if n==lvl
        % Compute coefficients of detail.
        z = conv2(y,Hi_D,'valid');
        % downsample by factor of 2
        CD = z(2:2:length(x)+length(Lo_D)-1);
    end

    x = CA;

end

if coef==0
    d = CD;
    a = x;
else
    d = CD(coef);
    a = x(coef);
end

%-------------------------------------------------------------------------%
function I = getSymIndices(lx,lf)

I = [lf:-1:1 , 1:lx , lx:-1:lx-lf+1];

if lx<lf
    K = (I<1);
    I(K) = 1-I(K);
    J = (I>lx);
    while any(J)
        I(J) = 2*lx+1-I(J);
        K = (I<1);
        I(K) = 1-I(K);
        J = (I>lx);
    end
end
%-------------------------------------------------------------------------%
```

C.2 MLP Classification Example

The MATLAB code below was used to create Figure 8.6. It shows the use of the MLP network to solve a classification problem both with, and without, the evidence procedure.

```
% EXAMPLE OF NETWORK CLASSIFICATION WITH MLP
%
% Uses: gmm, gmmsamp, gmmactiv, gmmpost, mlp, foptions, netopt, mlpfwd
%       evidence, mlpevfwd from the NETLAB toolbox
%               (http://www.ncrg.aston.ac.uk/netlab/index.php)
%
% Data generated from gaussian mixture model is trained using MLP network
% with and without the evidence procedure.

clear all

n=500;  % number of data points
randn('state',684);     % set pseudo random number generator state
rand('state',684);

% Set up gaussian mixture model: 2d data with three centres
% Class 1 is first centre, class 2 from the other two
mix = gmm(2, 3, 'full');
mix.priors = [0.5 0.25 0.25];
mix.centres = [0 -0.1; 1 1; 1 -1];
mix.covars(:,:,1) = [0.625 -0.2165; -0.2165 0.875];
mix.covars(:,:,2) = [0.2241 -0.1368; -0.1368 0.9759];
mix.covars(:,:,3) = [0.2375 0.1516; 0.1516 0.4125];

% generate the data from the mixture model
[data, label] = gmmsamp(mix, n);

% label==0 is given a target value of 0
% label==1 or 2 is given a target value of 1
target=(label>1);

% plot data
figure(1);clf;
plot(data(target==0,1),data(target==0,2),'b.')
hold on
plot(data(target==1,1),data(target==1,2),'r.')

% determine data range parameters
x0 = min(data(:,1)); x1 = max(data(:,1));
y0 = min(data(:,2)); y1 = max(data(:,2));
dx = x1-x0; dy = y1-y0;
expand = 10/100; % Add on 5 percent each way
x0 = x0 - dx*expand; x1 = x1 + dx*expand;
y0 = y0 - dy*expand; y1 = y1 + dy*expand;
resolution = 100;
step = dx/resolution;
xrange = [x0:step:x1];
yrange = [y0:step:y1];

% create a grid of points in data range
[X Y]=meshgrid([x0:step:x1],[y0:step:y1]);
```

```
% Set network parameters
nin = 2;        % number of inputs
nhidden = 8;    % number of hidden units
nout = 1;       % number of network outputs

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% Train using unregularized MLP (no evidence) %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% noise parameter (added to error function to avoid local minima)
alpha = 0.01;

% initialize network
net = mlp(nin, nhidden, nout, 'logistic',alpha);

options = foptions;     % default options
options(14)=60;         % set 60 steps of error reduction

% Train network using quasi-Newton error minimization
[net]=netopt(net, options, data, target, 'quasinew');

% put all points in the grid through the network so we have a network
% output at all values in the data range
yg = mlpfwd(net, [X(:),Y(:)]);
yg = reshape(yg(:,1),size(X));

% plot the optimal (50/50) decision boundary for this network
[cN, hN] = contour(xrange, yrange, yg, [0.5 0.5],'k:');
set(hN,'LineWidth',1)


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% Train using evidence-based MLP %%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Using same number of inputs, hidden units and outputs as before

% Initialize 1st and 2nd layer weights and biases to nominal small values
aw1 = 0.01*ones(1,nin);
ab1 = 0.01;
aw2 = 0.01;
ab2 = 0.01;

% define priors based on the current network parameters
prior = mlpprior(nin, nhidden, nout, aw1, ab1, aw2, ab2);

% initialize network
net = mlp(nin, nhidden, nout, 'logistic',prior);

% set number of times to perform evidence procedure
```

```
nouter = 5; ninner = 2;
options = foptions;
options(2)=1e-5; options(3)=1e-5; options(14)=100;

% train network, re-evaluate hyperparameters via the
% evidence procedure and retrain...

for k=1:nouter
    net = netopt(net, options, data, target, 'quasinew');
    [net, gamma] = evidence(net, data, target, ninner);
end

% put grid points through the regularized MLP network
[yg, ymodg] = mlpevfwd(net, data, target, [X(:),Y(:)]);
yg = reshape(ymodg(:,1),size(X));

% plot the optimal (50/50) decision boundary for the regularized MLP network
[cE, hE] = contour(xrange, yrange, yg, [0.5 0.5],'k-');
set(hE,'LineWidth',1)

% calculate Bayesian posteriors from gaussian mixture model
px_j = gmmactiv(mix, [X(:) Y(:)]);
px = reshape(px_j*(mix.priors)',size(X));
post = gmmpost(mix, [X(:) Y(:)]);
p1_x = reshape(post(:, 1), size(X));
p2_x = reshape(post(:, 2) + post(:, 3), size(X));

% plot the optimal Bayesian decision boundary
[cB, hB] = contour(xrange, yrange,p1_x,[0.5 0.5],'m-');
set(hB,'LineWidth',1)

% plot gaussian mixture model centers
plot(0,-0.1,'k+','markers',20)
plot(1,1,'k+','markers',20)
plot(1,-1,'k+','markers',20)

axis([-2 2.5 -3 3]);
legend('Data - Class 1','Data - Class 2','MLP no evidence', ...
    'MLP with evidence', 'Bayes Optimal Boundary',0);
xlabel('Input Value 1','FontSize',12)
ylabel('Input Value 2','FontSize',12)
```

Bibliography

[1] F. Zwicky, *Helv. Phys. Acta*, **6**, 110 (1933).

[2] M. Persic, et al., *Mon. Not. R. Astron. Soc.*, **281**, 27 (1996).

[3] P. Coles, F. Lucchin, *Cosmology: The Origin and Evolution of Cosmic Structure*, John Wiley & Sons, Ltd. (2002).

[4] D. S. Davis, R. E. White III, *Ap. J.*, **470**, L35 (1995).

[5] S. M. Faber, J. S. Gallagher, *ARA & A*, **17** 135 (1979).

[6] C. Adami et al., *Astron. Astroph.*, **336**, 63 (1998).

[7] M. Girardi et al., *Ap. J.*, **530**, 62 (2000).

[8] M. Girardi et al., *Ap. J.*, **505**, 74 (1998).

[9] M. Bartelmann, P. Schneider, *Phys. Rep.*, **340**, 291 (2001).

[10] B. Fort, Y. Mellier, *Astron. Astrophys. Rev.*, **5**, 239 (1994).

[11] X. P. Wu et al., *Mon. Not. R. Astron. Soc.* **301**, 861 (1998).

[12] D. Wittman, et al., *Ap. J.***557**, L89 (2001).

[13] K. Olive, G. Steigman, T. P. Walker, *Phys. Rep.*, **333**, 389 (2000).

[14] D. Tytler, et al., astro-ph/0001318.

[15] K. A. Olive, *19th Texas Symposium on Relativistic Astrophysics and Cosmology*, Paris (1998); astro-ph/9903309.

[16] P. Molaro, P. Bonifacio, F. Primas, *Memorie della Societa Astronomia Italiana*, **66**, 323 (1995).

[17] M. Tegmark, et al., *Phys. Rev. D***69**, 103501 (2004).

[18] S. Perlmutter, et al., *Ap. J.*, **440**, L41 (1995).

[19] M. Hamuy, et al., *Astron. J.*, **109** No. 1669, 1 (1995).

[20] A. Filippenko, *Annu. Rev. Astron. Astrophys.*, **35**, 309 (1997).

[21] D. Branch, *Annu. Rev. Astron. Astrophys.*, **36**, 17 (1998).

[22] I. M. Hook, et al., astro-ph/0509041, accepted for publication in *Ap. J.*.

[23] S. Perlmutter, et al., *Ap. J.*, **517**, 565 (1999).

[24] A. Reiss, et al., *Astron. J.*, **116**, 1009 (1998).

[25] R. A. Knop, et al., *Ap. J.*, **598**, 102 (2003).

[26] S. Perlmutter, B. Schmidt, in *Supernovae and Gamma Ray Bursts*, Springer Lecture Notes in Physics (IP), astro-ph/0303428.

[27] W. Hu, N. Sugiyama, J. Silk, *Nature*, **386**, 37 (1997).

[28] G. F. Smoot, et al., *Ap. J.*, **396**, L1 (1992).

[29] C. L. Bennett, et al., *Ap. J.*, **148**, 1 (2003).

[30] D. N. Spergel, et al., *Ap. J. S.*, **148**, 175 (2003).

[31] D. N. Spergel, et al., *submitted to Ap. J.*, astro-ph/0603449.

[32] M. Tegmark, et al., *Phys. Rev. D***D69**, 103501 (2004).

[33] S. Boughn, R. Crittenden, *Nature*, **427**, 45 (2004).

[34] S. S. McGaugh, *Ap. J.*, **541**, L33 (2000).

[35] W. Freedman, et al., *Ap. J.*, **553**, 47 (2001).

[36] L. Grego, et al., *Ap. J.*, **552**, 2 (2001).

[37] J. L. Tonry, et al., *Ap. J.*, **594**, 1 (2003).

[38] A. Vikhlinin, et al., *Ap. J.*, **590**, 15 (2003).

[39] N. A. Bahcall, P. Bode, *Ap. J.*, **588**, L1 (2003).

[40] M. Jarvis, et al., *Astron. J.*, **125**, 1014 (2003).

[41] J. A. Peacock, *AIP Conf. Proc.* **666**, 275 (2003).

[42] J. M. O'Meara, et al., *Ap. J.*, **552**, 718 (2001).

[43] S. Burles, K. M. Nollett, J. W. Truran, M. S. Turner, *Phys. Rev. Lett.*, **82**, 4176 (1999).

[44] S. N. Ahmed, et al., *Phys. Rev. Lett.*, **92**, 181301 (2004).

[45] C. Munoz, *Int. J. Mod. Phys.*, **A19**, 3093 (2004).

[46] S. F. King, *Rept. Prog. Phys.*, **67**, 107 (2004).

[47] R. D. Peccei, H. R. Quinn, *Phys. Rev. D***16**, 061791 (1977).

[48] R. Bradley, et al., *Rev. Mod. Phys.*, **75**, 777 (2003).

[49] E.W. Kolb and M.S. Turner, *The Early Universe*, Addison-Wesley, Redwood City, (1989).

[50] A. Corsetti, P. Nath, hep-ph/0005234.

[51] H. Baer, C. Balázs, A. Belyaev, J. O'Farrill, *J. Cosmol. Astropart. Phys.*, **09**, 007 (2003).

[52] R. J. Gaitskell, *Annu. Rev. Nucl. Part. Sci*, **54**, 315 (2004).

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

[54] L. Baudis, et al., *Phys. Rev. D***D59**, 022001 (1999).

[55] J. Morales, et al., *Nucl. Phys. B Proc. Suppl.*, **118**, 524 (2003).

[56] R. Abusaidi, et al., *Phys. Rev. Lett.*, **84**, 5699 (2000).

[57] A. Benoit, et al., *Phys. Lett. B*, **545**, 43 (2002).

[58] R. Bernabei, et al., *Phys. Lett. B*, **389**, 757 (1996).

[59] J. Quenby, et al., *Phys. Lett. B*, **351**, 70 (1995).

[60] P. F. Smith, et al., *Phys. Lett. B*, **379**, 299 (1996).

[61] R. Bernabei, et al., *Riv. Nuovo Cim.*, **26**, 1 (2003).

[62] C. Copi, L. Krauss, *Phys. Rev. D***67**, 103507 (2003).

[63] D. S. Akerib, et al., *Phys. Rev. D***73**, 011102(R) (2006).

[64] J. C. Barton, et al., *Proc. 4th Int. Workshop on Identification of Dark Matter*, p. 302, Singapore: World Sci. (2003).

[65] R. Bernabei, et al., *Phys. Lett. B*, **436**, 379 (1998).

[66] D. B. Cline, et al., *Proc. 4th Int. Workshop on Identification of Dark Matter*, p. 363, Singapore: World Sci. (2003).

[67] E. Aprile, et al., *New Astron. Reviews*, **49**, 289 (2004), astro-ph/0407575.

[68] M. Altmann, et al., astro-ph/0106314.

[69] M. Barnabe-Heider, et al., (PICASSO Collaboration), *Nucl. Phys. B*, **138**, 160 (2005).

[70] S. Hart, et al., *Nucl. Phys. B Proc. Suppl.*, **110**, 91 (2002).

[71] R. Clarke, Ph.D. Thesis, Stanford University (1999). http://cdms.berkeley.edu/Dissertations/

[72] B. Neganov, T. Trofimov, *Pis'ma v Zhurnal Experimental'noi i Teoreticheskoi Fiziki*, **28**, 356 (1978).

[73] P. Luke, *J. Appl. Phys.*, **64**, 6858 (1988).

[74] S. Nam, Ph.D. Thesis, Stanford University (1998). http://cdms.berkeley.edu/Dissertations/

[75] T. Saab, Ph.D. Thesis, Stanford University (2002).

[76] A. Miller, Ph.D. Thesis, Stanford University (2001).

[77] V. Mandic, et al., *Ninth International Workshop on Low Temperature Detectors*, AIP Conf. Proc., **605**, 509 (2002).

[78] T. A. Shutt, Ph.D. Thesis, University of California, Berkeley (1993). http://cdms.berkeley.edu/Dissertations/

[79] J. Lindhard, *Phys. Lett.*, **12**, 126 (1964).

[80] R. Abusaidi, et al., *Phys. Rev. D***84**, 5695 (2000).

[81] S. Kamat, Ph.D. Thesis, Case Western Reserve University (2005). http://cdms.berkeley.edu/Dissertations/

[82] D. S. Akerib, et al., *Nucl. Instrum. Meth.*A, **400**, 181 (1997).

[83] S. Golwala, Ph.D. Thesis, Appendix A, University of California, Berkeley (2000). http://cdms.berkeley.edu/Dissertations/

[84] J. Yoo, et al., *PipeFitter*, Internal CDMS notes.

[85] V. Mandic, et al., *Nucl. Instrum. Meth.*A, **520**, 171 (2004).

[86] C. L. Chang, Ph.D. Thesis, Stanford University (2004). http://cdms.berkeley.edu/Dissertations/

[87] R. M. Clark, et al., *Appl. Phys. Lett*, **76**, 2958 (2000).

[88] E. Browne, R. B. Firestone and V. S. Shirley, Editor, *Table of Radioactive Isotopes*, John Wiley & Sons (1986).

[89] D. Grant, CDMS Internal Note (2005).

[90] D. S. Akerib, et al., *Phys. Rev. D*72 052009 (2005).

[91] S. Yellin, *Phys. Rev. D*66, 032005 (2002).

[92] G. J. Alner et al., (UK Dark Matter Collab.), *Astropart. Phys.*, **23**, 444 (2005).

[93] V. Sanglard et al., (EDELWEISS Collab.), *Phys. Rev. D*71, 122002 (2005).

[94] R. Bernabei et al., *Phys. Lett.*, **B389**, 757 (1996).

[95] P. Gondolo and G. Gelmini, *Phys. Rev. D*71, 123520 (2005).

[96] A. Bottino et al., *Phys. Rev. D*69, 037302 (2004).

[97] J. Ellis et al., *Phys. Rev. D*71, 095007 (2005).

[98] R. J. Gaitskell, V. Mandic, *Dark matter direct detection results plotter*, http://dmtools.brown.edu, http://dmtools.berkeley.edu (2004).

[99] D. S. Akerib, et al., *Phys. Rev. Lett.*, **96**, 011302 (2006).

[100] R. Bernabei et al., *Riv. Nuovo Cim.*, **26N1**, 1 (2003).

[101] G. Angloher et al., (UK Dark Matter Collab.), *Astropart. Phys.*, **18**, 43 (2002).

[102] M. Barnabe-Heider et al., (PICASSO Collab.), *Phys. Lett.*, **B616**, 17 (2005).

[103] V. A. Kudryavtsev (UK Dark Matter Collab.), presented at the Fifth International Workshop on the Identification of Dark Matter, Edinburgh, Scotland (2004).

[104] G. J. Alner et al., (UK Dark Matter Collab.), *Phys. Lett.*, **B616**, 17 (2005).

[105] V. Mandic, Ph.D. Thesis, University of California, Berkeley (2004). http://cdms.berkeley.edu/Dissertations/

[106] G. Wang. Ph.D. Thesis, Case Western Reserve University (2005). http://cdms.berkeley.edu/Dissertations/

[107] Y. Sheng, *The Transforms and Applications Handbook*, Ed. by A. D. Poularikas, CRC Press (1996).

[108] A. R. Calderbank, I. Daubechies, W. Sweldens, B.-L. Yeo, *Applied and Computational Harmonic Analysis*, **5**, 332-369 (1998).

[109] I. Daubechies, *Ten Lectures on Wavelets*, 2nd Ed., Siam (1992).

[110] S. G. Mallat, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11** No. 7, 674 (1989).

[111] C. S. Burrus, R. A. Gopinath, H. Guo, *Introduction to Wavelets and Wavelet Transforms, A Primer*, Prentice Hall (1998).

[112] D. F. Walnut, *An Introduction to Wavelet Analysis*, Birkhäuser (2002).

[113] K. Hornik, M. Stinchcombe, H. White, *Neural Networks*, **2** No. 5 (1989).

[114] D. J. C. MacKay, *Neural Computation*, **4** No. 3 (1992).

[115] I. T. Nabney, *NETLAB: Algorithms for Pattern Recognition*, Springer-Verlag (2002).

[116] C. M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford (1995).