Contents

Chapter 1

The dark matter problem.

1.1 Introduction.

Lots of gravitating material that doesn't emit or absorb light seems to be required in all sensible accounts of the dynamics of large-scale structures in the universe. The nature and extent of this mysterious "dark matter" has been one of the central puzzles in cosmology over the last decade. This dissertation describes an experiment that tests one possibility, that the dark matter is in the form of undiscovered Weakly Interacting Massive Particles (WIMPs) produced as a thermal relic of the big bang. In this chapter, we will review the most important observations that suggest the dark matter must exist and discuss the forms it could take.

1.2 The Friedmann equations.

Modern cosmology is rooted in the Friedmann solution of Einstein's field equations for a universe filled with an isotropic, homogeneous fluid of density ρ and pressure p:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda}{3} \tag{1.1}$$

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} + \frac{\Lambda}{3} - \frac{kc^2}{a^2} \tag{1.2}$$

Here Λ is the cosmological constant, G is Newton's constant, and $a(t)$ is the fundamental distance scale parameter found in the metric $ds^2 = c^2dt^2 - a^2(t)dl^2$. The distance scale is related to Hubble's constant H_0 by the definition

$$H_0 \equiv \left(\frac{\dot{a}}{a}\right)_{t=t_0} .$$

The time t_0 is the present. The parameter k can be made equal to +1, 0, or -1 by a rescaling of the metric coordinates. It describes the global curvature of the universe, which can be "closed" (k=+1), "flat" (k=0), or "open" (k=-1). There is a strong theoretical prejudice in favor of $k = 0$, because universes that are not flat would quickly evolve to a state with either negligible (for $k = -1$) or very large (for $k = +1$) matter density and would not look much like our own in the current epoch. The popular class of theories known as inflation include a mechanism to drive the universe into a state with $k = 0$. If $k = 0$ and $\Lambda = 0$ then Eqn. 1.2 can be solved for the mass density in terms of the Hubble constant:

$$\rho = \rho_c \equiv \frac{3H_0^2}{8\pi G}$$

Since we don't know Hubble's constant very well, it's often written in terms of a unknown constant h of order unity,

$$h = \frac{H_0}{100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}}.$$

A Mpc, or 10^6 pc, is the standard unit of distance in cosmology, with 1 pc $= 3.1 \times 10^{16}$ m.

In terms of h, the numerical value of the critical density is

$$\begin{aligned} \rho_c &= 2.8 \times 10^{11}\, h^2\, M_\odot/\mathrm{Mpc}^3 \\ &= 1.9 \times 10^{-29}\, h^2\, \mathrm{g/cm^3} \\ &= 1.1 \times 10^{-6}\, h^2\, \mathrm{GeV/c^2cm^3} \end{aligned}$$

There has been a big push over the last 5 years to increase the accuracy to which h is known, driven largely by the availability of the Hubble space telescope. As this dissertation goes to press, the favored range is $h = 0.65 \pm 0.1$. This range seems to accommodate a large fraction of recent results, which are often mutually incompatible within estimated errors.

It's conventional to write the actual matter density of the universe as a fraction of the critical density, $\Omega_m = \rho_0/\rho_c$. To include the possibility of a non-zero cosmological constant, the total density can be written $\Omega_{\mathrm{total}} = \Omega_m + \Omega_\Lambda$, with $\Omega_\Lambda \equiv \Lambda/3H_0^2$. A flat universe ($k = 0$) must have $\Omega_{\mathrm{total}} = 1$ to be consistent with the Friedmann equations.

The values of Ω_Λ and Ω_m are currently a subject of intense debate [1-6]. In general, mainstream astronomers favor $0.1<\Omega_m<0.4$ on the basis of observations that will be described below, and are agnostic on the question of Λ. Cosmologists with particle physics backgrounds prefer $\Omega_m=1$ and $\Omega_\Lambda=0$, on the grounds that Λ, which receives a contribution from the vacuum energy density of the fundamental quantum fields, should either be zero (in the case of exact cancellations) or impossibly large. However, in the face of recent evidence that $\Omega_m < 1$ and $\Omega_\Lambda > 0$, most seem willing to settle for a flat universe with $\Omega_m + \Omega_\Lambda = 1$ in order to save inflation. A small part of the community has taken a different approach, attempting to modify inflation to evade the Ω constraint ("open inflation").

1.3 Measuring Ω_m with mass-to-light ratios.

One way to estimate Ω_m is to measure the density of large objects by counting the number in a region of space and multiplying by the average mass. These objects could be individual galaxies or clusters of galaxies. It's possible to infer galaxy and cluster masses by a number of methods discussed below. In practice, the extrapolation from the masses of the few objects that are measured to the density of the universe as a whole is done, not by counting, but by multiplying the average galaxy and cluster mass/luminosity (or "mass-to-light") ratio by the galaxy luminosity density L_g, which has a measured [2] value,

$$L_g = 3.3 \times 10^8 h\, L_\odot/\mathrm{Mpc}^3.$$

The critical density mass-to-light ratio is

$$\left(\frac{M}{L}\right)_{\rm critical} \simeq 1400h\frac{M_\odot}{L_\odot}. \tag{1.3}$$

The symbols $M_\odot$ and $L_\odot$ refer to the mass and luminosity of the Sun. The large numerical value of the critical density mass-to-light ratio in solar units already hints at the magnitude of the dark matter problem: If most of the mass in the universe is in objects like our Sun, then $\Omega_m \simeq 1/1400h \simeq 0.1\%$.

1.3.1 Galaxy masses.

There is some evidence for a large dark matter component in every class of galaxy. This evidence is most convincing in spiral galaxies, such as our Milky Way, because the net rotational velocity as a function of distance from the center, often referred to as the "rotation curve", can be used to directly trace the mass density distribution out to large radii. Spiral galaxies consist of two luminous components, a spherical central nucleus of densely packed stars, called the "bulge", and an extended disk of stars with a surface brightness that decreases exponentially with radius, $I(r) = I_0 \exp -3.2r/R_{\rm opt}$. The parameter $R_{\rm opt}$ is the "optical radius", within which 83% of the light is contained. Typically $R_{\rm opt} \simeq 10$ kpc and $I_0 \simeq 140 L_\odot$ pc^{-2}. In spiral galaxies, the disk dominates the bulge in mass and luminosity.

Measurements of rotation velocities of material in the disk can be used to trace the mass distribution under the assumption that the rotation is counterbalanced by gravity, with

$$\frac{V^2(r)}{r} = \frac{GM(r)}{r^2},$$

where $V(r)$ is the rotation velocity at radius r and $M(r)$ is the mass inside the radius. Inside the optical radius, the rotation can be traced through measurements of Doppler shifts in stellar absorption lines. Beyond the optical radius, the radio emission line of neutral hydrogen can be used to trace the rotation out to about $2R_{\rm opt}$ [7]. At larger distances, it's sometimes possible to measure the rotation using satellite galaxies as test particles.

If mass followed light in spiral galaxies, the velocity curve beyond the optical radius would be

$$V(r) \propto 1/\sqrt{r} \quad r > R_{\rm opt}$$

What is actually observed is that these curves tend to flatten out at a constant velocity near $R_{\rm opt}$ and remain flat out to the largest radii where measurements are possible. This implies that there is an invisible mass component contributing $M(r) \propto r$. A natural possibility is that this mass is in a spherically symmetric dark halo with density $\rho(r) \propto 1/r^2$. This is the expected mass distribution for a set of dynamically relaxed particles supported in hydrostatic equilibrium against gravity.

Many researchers have successfully fit observed rotation curves with plausible mass models. For example, Persic and Salucci have analyzed data from 1100 spiral galaxies and claim that they can all be described by a "universal rotation curve", which gives the shape of the velocity curve for any galaxy in terms of a single input parameter, the galaxy luminosity [8] . This curve is constructed by assuming that the underlying mass distribution consists of two components, (1) a thin disk with an exponentially decreasing radial surface mass density (following the light), and (2) a spherical halo with mass $M(x) \propto \frac{x^3}{x^2+a^2}$, where $x \equiv r/R_{\rm opt}$ and $a = 1.5\left(\frac{L}{L_\star}\right)^{1/5}$. Here $L_\star$ is the mass of the Milky Way, and a is known as the "core radius". Inside the core radius, the density is approximately constant, and outside it drops as

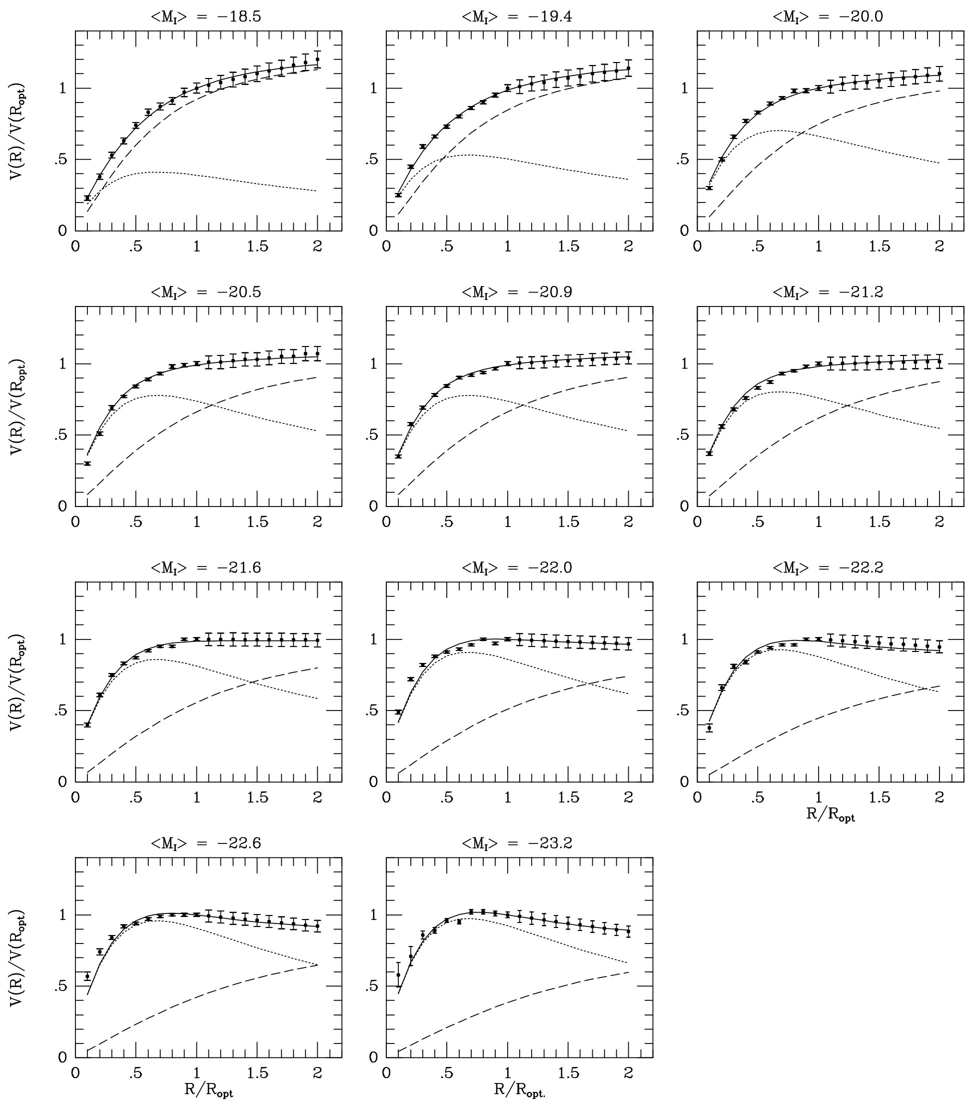


Figure 1.1: Rotation curves of 11 galaxies from Salucci and Persic, 1997 [8]. The velocity normalized to the value at the optical radius $R_{\rm opt}$ is plotted as a function of r/$R_{\rm opt}$. Data is shown with error bars. The contributions of the 2 components of the mass model are shown as a dashed line (halo), dotted line (disk), and solid line (sum).

$1/r^2$. Figure 1.1 shows 11 of the Persic and Salucci rotation curves for galaxies with varying luminosities. The dark matter component and the luminous matter component are usually both contributing significantly to the rotation velocity even at $r = R_{\rm opt}$ and the dark matter increasingly dominates beyond this.

The existence of massive dark halos seems to have been convincingly demonstrated. It is still an open question how far the halos extend. Beyond about $2R_{\rm opt}$, the emission from neutral hydrogen is too weak to trace the velocity, and we must rely on measured positions and redshifts of a small number of satellite galaxies. In the Milky Way, the most conservative estimate is that the halo extends at least to $r \simeq 34$ kpc, with a total mass of $4 \times 10^{11} M_\odot$, but if one accepts that the globular cluster Leo I at 230 kpc is bound to the galaxy, then this goes up to $1.25^{+0.8}_{-0.3} \times 10^{12} M_\odot$ [9]. In other galaxies similar to the Milky Way, there is evidence from satellites that the rotation curve remains constant out to 400 kpc. If the dark halo of a spiral galaxy extends out to a radius R , the mass-to-light ratio will be [10]

$$\frac{M}{L} \simeq 60 \left(\frac{R}{100\, kpc} \right) \frac{M_\odot}{L_\odot}.$$

Could the dark matter be in the galactic disk rather than the halo? This question can be answered in our own galaxy by measuring the velocities and positions of nearby stars that are slightly out of the disk plane. These stars see a gravitational potential $\Phi = 2\pi G \Sigma z$ at a height z above the disk due to the disk column density Σ. From statistical mechanics and the virial theorem, it follows that their positions in equilibrium will be distributed as $dn/dz \propto Exp(-\Phi/\langle v^2 \rangle)$ [5]. Recent measurements of position and velocity distributions of stars near us in the disk give

$\Sigma = 48 \pm 9\, M_\odot/pc^2$ [11]. This mass density can be directly accounted for in terms of known populations of stars and gas, which sum to $\Sigma = 48 \pm 8\, M_\odot/pc^2$, leaving no room for a significant dark matter component in the disk.

1.3.2 Galaxy cluster masses.

Galaxy clusters are groups of up to several thousand gravitationally bound galaxies. The largest clusters, the so-called "rich" clusters, have diameters up to about 5 Mpc. In addition to the member galaxies, clusters contain an "intercluster medium" (ICM) consisting of X-ray emitting gas. The gas is believed to have originated in the member galaxies and to have been swept out by tidal forces. The X-rays are made by bremsstrahlung of electrons undergoing collisions at typical temperatures of $10^6 - 10^7$ degrees.

The masses of clusters can be measured by three methods described below: (1) applying the virial theorem to measured velocities and positions of member galaxies to determine the depth of the potential well; (2) fitting the measured X-ray temperatures and luminosity of the hot gas to a model assuming hydrostatic equilibrium; and a relatively new method, (3) inferring the cluster mass by observing gravitational distortion of background galaxy images.

Virial mass estimates.

The virial theorem tells us that an individual particle in a closed mechanical system will have time averaged potential (U) and kinetic (T) energies related by $< U >$ $+2 < T >= 0$. To apply this to galaxy clusters, we make the assumption that the time average kinetic and potential energies of the cluster galaxies are equal to the average of these quantities over all the galaxies in the system as it is observed now.

In other words, the cluster must be in a state of equilibrium.

By writing the mean potential and kinetic energies in terms of the mean galaxy separation $< r >$, velocity dispersion$< v^2 >$, and the cluster mass M, it's easy to show that

$$M \simeq \frac{2 < v^2 >< r >}{G}$$

In practice, we only have experimental access to the velocities of the member galaxies along the line of sight (from redshifts) and the mean radii perpendicular to the line of sight. This is not enough to uniquely determine the mass of an individual cluster unless it is spherically symmetric. It is possible, however, to find the mean mass of a sample of clusters, since the errors will cancel. An additional complication arising from the fact that we can only measure 3 of the 6 relevant degrees of freedom for each galaxy is the possibility that foreground or background galaxies will accidentally be counted as cluster members, increasing the measured velocity dispersion, and leading to a mass over-estimate.

The assumption of dynamical equilibrium may be dangerous, since many clusters show signs that they are undergoing mergers or have substructure that is obviously not gravitationally stable. On the other hand, a significant fraction of clusters do seem to be relaxed into a pressure supported equilibrium, as indicated by smooth X-ray maps or galaxy number density counts. Virial mass estimates can be made more accurate by imposing selection criteria that include only the most relaxed objects. The availability of high resolution X-ray maps for large number of clusters has recently made this much easier.

The virial mass estimate was first made for clusters by Zwicky in 1933, who examined the Coma cluster and found a mass to light ratio of about 500 $M_\odot/L_\odot$. This result is often credited as the first discovery of the "dark matter problem" on extragalactic scales. An important example of the modern application of the technique is the work of Carlberg *et al.* [12] (the CNOC Cluster Survey), who studied 16 clusters at redshifts $\sim$1/3 selected from an X-ray sky survey. Observations were made with a multi-object spectrometer, giving redshifts accurate to about 100 km/s for 20-200 objects in each cluster or 2600 redshifts in all. They were able to study the statistical properties of background galaxies in their sample to eliminate the possibility that redshift-space interlopers were significantly raising the measured velocity dispersion. Their result is a mean cluster mass-to-light ratio 294 $\pm$ 50 $h\, M_\odot/L_\odot$, corresponding to an average member galaxy mass of 4.0 $\pm$1.1 $\times\ 10^{12}\ h^{-1} M_\odot$. To account for galaxy luminosity evolution from $z = 1/3$, they calculate the critical luminosity density using galaxies at the same redshift, with a result that is 35% lower than that given in Equation 1.3 for $z = 0$. Using their luminosity density, they find $\Omega_m = 0.23 \pm 0.1$.

X-ray mass estimates.

The X-ray method for measuring cluster masses depends on the assumption that radiating cluster gas is in hydrostatic equilibrium. In that case, the mass contained within radius r is

$$M(r) = -\frac{kT(r)}{Gm} r \left[\frac{d \log \rho(r)}{d \log r} + \frac{d \log T(r)}{d \log r} \right],$$

where $\rho(r)$ and $T(r)$ are the density and temperature profiles and m is the mean molecular weight of the gas. Many mass estimates are based on the "isothermal β

model", which assumes a density profile for the gas $\rho(r) = \rho_0 \left[1 + (\frac{r}{a})^2\right]^{-3\beta/2}$ and a constant temperature T_0 [13]. In this case, the observed surface brightness is $\Sigma(r) = \Sigma_0 \left[1 + (\frac{r}{a})^2\right]^{-3\beta+1/2}$, and the parameters a and β can be extracted by fitting $\Sigma(r)$ to an X-ray intensity map. The mass in a given radius is then

$$M(r) = 1.13 \times 10^{15} \beta \left(\frac{kT_0}{10\,keV}\right) \left(\frac{r}{Mpc}\right) \left(\frac{(r/a)^2}{1 + (r/a)^2}\right) M_{\odot}.$$

The assumption of hydrostatic equilibrium is similar to the equilibrium assumption that was needed to employ the virial theorem, but this method has the advantage that the X-ray maps themselves contain detailed information about the dynamical state. Many clusters show a smooth, azimuthally symmetric gas emission with a smooth radial temperature profile. Evrard *et al.* show that N-body simulations of clusters can reproduce the observed X-ray maps in good detail [14]. In their simulations, clusters are typically supported by pressure out to at least the radius where the density has fallen to $500\rho_c$, or $\sim$1 Mpc for rich clusters. The β model applied to simulated X-ray images of the N-body clusters gives masses accurate to about 30% when evaluated at this radius. Of course, these simulations only include the physics that is put in to them, so all possible systematic errors are not accounted for. In particular, any physics that causes clumping of the gas can lead to errors, since the X-ray luminosity is proportional to the density squared. One known problem is the existence of low temperature regions in the centers of some clusters caused by radiative cooling. This leads to errors in the β model fit and an underestimate of the mass. Another potentially serious problem is the departure from hydrostatic equilibrium caused by recent cluster-cluster mergers.

Application of the β model leads to cluster mass-to-light ratios between 100- 300

$h\, M_\odot/L_\odot$. For example, David, Jones, and Forman studied 7 clusters and groups selected for high luminosity and smooth morphology with the ROSAT satellite and found ratios of 100-150 $hM_\odot/L_\odot$ [15]. These measurements extend only to a radius of $\sim$1 Mpc, where simulations and temperature maps indicate that the assumption of equilibrium begins to break down. At the largest probed radii, the mass-to-light ratio for the richer clusters in the sample is still climbing steeply.

Mass estimates from gravitational lensing.

Since virial and X-ray cluster mass estimates both suffer from possible systematic errors due to departures from equilibrium, it would be very useful to check them against a method that doesn't share this source of systematic wrror. Recently, new techniques based on the gravitational lensing of faint background galaxies by clusters have offered hope of doing this [17]. To 30th magnitude, there are about 10^{11} galaxies in the sky, or 10^3 per square arc minute. Near the centers of clusters or individual galaxies, it's possible to observe "strong lensing", in which the background galaxies form multiple images. On larger length scales ($\sim$ 1 Mpc), it's possible to see "weak lensing", the systematic distortion of galaxy shapes by the gravitational field of the cluster. This is a statistical effect, which requires averaging over many galaxies. At each test point in the image plane, a gravitational shear strength is measured by computing the average component of the quadrupole of the individual background galaxies along the line connecting the test point to the galaxy center. This shear strength field is the fundamental observable quantity. Mass models can be constructed that reproduce the observed shear. Typically, a mass model of an imaged cluster is made of hundreds of mass "blobs", with floating parameters describing the shape and density. The best fit parameters are found by an iterative process that involves varying the parameters to reproduce the observed shear pattern. In some cases, strong and weak lensing can both be used to constrain the models.

The great advantages of the lensing methods are that they do not rely on the assumption of virial equilibrium in the lensed object and do not rely on the observation of bright test particles to trace the mass. Unfortunately, lensing methods have their own set of serious potential systematic errors. Distortions of the image caused by imperfections in the telescope optics or warping of the CCD plane can lead to false signals or reductions in the observed signal size. In the case of ground based observations, atmospheric seeing reduces the signal strength by a factor of 2-3 [18]. Another serious potential problem is that the observed signal depends on the redshifts of the lensed galaxies, which are not directly measured. The uncertainty in the estimation of the magnitude of such effects is an important limitation. However, gravitational lensing studies are still in their infancy and much improvement can be looked forward to as new, special purpose instrumentation is brought on-line and new calibration techniques are developed. Recently, deep, wide-field images made with the Hubble space telescope have given the field a big boost, raising confidence that seeing-related errors can be controlled.

At present, the lensing estimates of cluster masses seem to produce a greater scatter in the measured mass to light ratios than the other methods. To cite a few examples, Smail *et al.* studied 10 clusters with the Hubble space telescope and found mass to light ratios between 290 and 2480 $h\,M_\odot/L_\odot$ in the central 0.5 h^{-1} Mpc [18] . Fischer and Tyson studied the brightest known X-ray cluster with a ground-based instrument and found a ratio of 330$\pm$110 $h M_\odot/L_\odot$, which is a factor of 2 higher than the β model X-ray mass for this object [16].

1.4 Constraints on Ω_m from the cluster baryon fraction.

As we have discussed, mass estimates of galaxy clusters can be used to estimate Ω_m by assuming that the total luminosity of the universe comes from structures with cluster-like mass-to-light ratios. There is a second way to use clusters to measure Ω_m, which follows from the observation that the matter content of clusters should be a fair sample of the matter content of the universe as a whole. That is, clusters are claimed to be so large that they must have the universal ratio of baryonic to non-baryonic matter. Since the theory of big bang nucleosynthesis gives us an apparently accurate measurement of the baryonic contribution to the mass density of the universe, Ω_b, the total matter density should simply be

$$\Omega_m = \frac{\Omega_b}{f_b} \tag{1.4}$$

where f_b is the fraction of cluster mass that is known to be baryonic. In practice, we may measure only a lower limit on f_b, since it is possible to hide baryons in the cluster. In that case, the result is an upper limit on Ω_m rather than a measurement.

Do clusters really fairly sample the matter content of the universe? According to the conventional gravitational instability scenario for structure formation, clusters form from an initially nearly-homogeneous medium by accretion into regions that happen to have a small overdensity. In the absence of extra mechanisms for energy dissipation or pressure support, the baryonic matter will fall into the cluster at the same rate as non-baryonic matter. Deep inside the cluster, where densities and temperatures are high, hydrodynamic and radiative processes should lead to significant segregation of the baryons. However, at large distances from the cluster center, no segregation is expected, since the material there is nearly in free fall.

Numerical models of the infall process have been used to determine the distance from the cluster centers where segregation can be significant. The results show that the baryon fraction measured in the outer regions of a cluster (typically 1.5 h^{-1} Mpc, known as an "Abell radius") should be nearly unbiased. For example, White *et al.* have attempted to measure the maximum possible baryon enhancement by exploring a "infinitely dissipative" model of the Coma cluster, in which the baryons simply fall into the cluster center and stick, while non-baryonic material can lose energy only through gravitational interactions [19]. In the spherically symmetric case, the maximum baryon over-density inside 1.5 h^{-1} Mpc is 1.4. Departures from spherical symmetry always seem to reduce this factor, so this should be considered an upper limit. In fact, there is observational evidence suggesting that non-baryonic dark matter is more, not less gravitationally relaxed than baryons, probably because of energy injection into the intercluster gas by galaxies [15].

X-ray studies reveal that the intercluster gas has much more mass than the visible stars. The gas mass can be extracted by integrating over the β model density profile after corrections for light extinction by the cluster and our galaxy. White and Fabian [20] found gas mass fractions between 10-22% in a sample of 13 clusters. Mohr *et al.* obtained similar results for 45 clusters, with average $f_{gas} = (0.13 \pm 0.01) \times (h/0.65)^{-3/2}$ [22]. These gas fractions can be checked by observing distortions in the cosmic microwave background caused by Compton scattering of microwave photons on the electrons in the gas (the Sunyaev-Zel'dovich effect). These measurements have somewhat larger errors, but are generally consistent, with Carlstrom *et al.* finding $f_{gas} = (0.09 \pm 0.01) \times (h/0.65)^{-1}$ [21] .

As explored in some detail in section 1.8, big bang nucleosynthesis predicts $\Omega_b = (0.05 \pm 0.01) \times (h/0.65)^{-2}$. Using Equation 1.4, the measured cluster gas fractions imply

$$\Omega_m = \Omega_b / f_b \simeq 0.4 \pm 0.1$$

for a range $h = 0.65 \pm 0.1$.

1.5 Large scale flows.

Important constraints on Ω_m can be derived from comparisons of data on large scale matter flows with what is predicted for a FRW universe filled with a nearly-homogeneous fluid that is moving in response to its own gravitation. The response of the fluid to small density perturbations can be calculated by linearizing the FRW and fluid equations [5]. For perturbations in the density field $\delta(\boldsymbol{x}) = \frac{\rho(\boldsymbol{x})-\rho_0}{\rho_0}$, expressed in comoving coordinates, the corresponding fluid velocity field is determined by the equation

$$\nabla \cdot \boldsymbol{v}(\boldsymbol{x}) = -aH_0 f \delta(\boldsymbol{x}) \tag{1.5}$$

where $f \simeq \Omega_m^{0.6}$. This equation has the solution

$$\boldsymbol{v}(\boldsymbol{x}) = a\frac{fH_0}{4\pi} \int \frac{\boldsymbol{y}-\boldsymbol{x}}{|\boldsymbol{y}-\boldsymbol{x}|^3} \delta(\boldsymbol{y}) d^3 y \tag{1.6}$$

Comparing these equations to observations in order to infer the value of f (and therefore Ω_m) is hampered by the fact that neither $\boldsymbol{v}(\boldsymbol{x})$ nor $\delta(\boldsymbol{x})$ can be directly measured. What can be measured is the spatial distribution and line-of-sight velocity components for a sparse set of test particles, which usually are galaxies. These measurements can then be spatially "smoothed" to construct continuous fields. It's necessary to make some assumption about how the measured density of visible galaxies, $\delta_v(\boldsymbol{x})$ relates to the true mass density $\delta(\boldsymbol{x})$. A typical assumption is that $\delta_v(\boldsymbol{x}) = b\delta(\boldsymbol{x})$, where the multiplicative constant b is known as the "bias". In that case, the measurable quantity is not f, but $\beta = f/b \simeq \Omega_m^{0.6}/b$.

There is a clever trick which allows us to infer the 3 dimensional velocity field from only the line-of-site components. If $\nabla \times \boldsymbol{v} = 0$, the velocity can be written as the gradient of a scalar field $\Phi(\boldsymbol{x})$,

$$\boldsymbol{v}(\boldsymbol{x}) = \nabla\Phi(\boldsymbol{x})$$

This may be a reasonable assumption, because gravitational instability itself is expected to produce little vorticity on large scales and the Hubble expansion damps out what rotational motion there is. The potential $\Phi(\boldsymbol{x})$ can be determined by measuring $v_r = \nabla\Phi(\boldsymbol{r}) \cdot \hat{\boldsymbol{r}}$ using redshifts and forming the integral

$$\Phi(r, \theta, \phi) = -\int_0^r v_r(r', \theta, \phi) dr'$$

The gradient of Φ then gives us the "unobservable" components of v.

There are many difficulties in reliably measuring the line-of-sight velocities of the large numbers of galaxies which are needed to construct large-scale velocity fields [23]. To measure the comoving (or "peculiar") velocity of a galaxy, one needs to know (1) the galaxy's redshift z, and (2) its redshift-independent luminosity distance d_l. Then the velocity is determined by subtracting the Hubble-flow component from the redshift-derived velocity, $v_r \approx cz - \frac{H_o}{c} d_l$. Redshifts can now be determined accurately for large numbers of galaxies with relative ease, but distances are still hard to measure. Distance measurements depend on the existence of light sources of predictable absolute luminosity, known as "standard candles". Measuring the luminosity of a standard candle at the position of the Earth allows us to calculate d_l under the assumption that the light is attenuated by $1/4\pi d_l^2$. These measurements are difficult because there are few high-quality standard candles that can be observed at extragalactic distances. There are persistent suspicions that all known standard candles suffer from various sorts of position or age dependent biases. These problems have been extensively studied as part of the effort to determine an accurate value for the Hubble constant H_0 and are also the main deterrent to mapping the large scale structure of space-time (thus determining Ω_m , Ω_λ , and Ω_k) by measuring z vs. distance on very large scales.

Existing catalogs of peculiar velocity measurements contain $\sim$3000 velocities, which have mostly been obtained by the use of the Tully-Fisher and Faber-Jackson relationships between the absolute luminosity and galactic rotation velocity (for spirals) or dispersion velocity (for ellipticals). These methods produce an intrinsic scatter of at least 20% in the measured distances, resulting in an RMS velocity error of $\sim$ 600 km/s for objects within 100 Mpc. This error is very comparable in size to the expected magnitude of the bulk flow velocity from gravitational instability, so there is very little information in the measured velocities of individual galaxies. Useful information comes only from smoothing over large numbers of galaxies. Of course, this smoothing only cancels errors which are statistical in nature and does not eliminate systematic biases.

A number of efforts have been made to compare the observed velocity fields to the observed density fields using Equation 1.5. This procedure seems to work, since high densities of galaxies are found at points where the velocity field converges and low densities at points where it diverges. Willick *et al.* [24] find $\beta = 0.45 \pm 0.05$

or $\Omega_m = 0.31 \pm 0.06$ for $\beta = 1$. However, there are other results based on similar methods and data that do not agree [4], with a range $0.5 < \beta < 1.2$.

An interesting constraint on Ω_m that is independent of β comes from considering the maximum possible outflow velocity from regions of low density called "voids". A low density region acts like a negative mass, since matter will be attracted away from it. But a void can not be more empty than empty, so there is a limit to how negative the effective mass can be. Dekel and Rees [28] have used a generalization of Equation 1.5 to derive the constraint $\Omega_m \geq 0.3$ (90% CL) based on a velocity map around one large void.

1.6 Measuring Λ with supernovae.

Since the discovery of the expansion of the universe and the parameterization of the recent expansion rate by the linear Hubble law, it has been realized that Ω_m and Ω_Λ could be measured by measuring deviations from linearity of the law at high redshift. The exact relationship between luminosity distance d_l and redshift z can be obtained by integrating the FRW equations. For objects at different redshifts, this gives d_L as a different linear combination of Ω_m and Ω_Λ (See e.g. [29] for details).

Various attempts to constrain the cosmological parameters have been made over the years by fitting the calculated $d_L(z; \Omega_m, \Omega_\Lambda)$ relationship to observations of high z objects, without much success until very recently. The problem is a recurring one in observational cosmology: it is hard to find objects bright enough to be visible at high z which have a well-known absolute luminosity. Interest was revived in this problem in the 1980's when it was discovered that Type Ia supernovae, which are extremely bright, have a very small luminosity dispersion. These supernovae are believed to be caused by the gravitational collapse of white dwarf stars which have gradually accreted matter from a binary companion until they reach the point of collapse at 1.4 $M_{\odot}$, known as the Chandrasekhar mass. They are thought to be well understood theoretically and their luminosity is not expected to be effected strongly by their environments (e.g. by differences in metalicity between early and late times) in a way that would cause a bias at high z. The problem has simply been to discover enough of them, since they are rare and only bright for a short time. This problem has been overcome by the use of large CCD arrays and automated photometric surveys of thousands of galaxies per night. The supernovae take about 20 days to reach maximum brightness, at which point they may be as bright as their host galaxy, so they can be detected efficiently by comparing images taken at intervals of 2-3 weeks. Supernovae detected this way are then studied in detail with large telescopes, allowing the remainder of the light curve to be followed over time and high-resolution spectra to be obtained.

At the time of writing (Jan. 1999), this strategy has resulted in the discovery of ~ 100 Type Ia supernovae with $z > 0.2$ by two groups using similar methods, the Supernovae Cosmology Project [29] and the High Z Supernovae Search Team [30]. The results of the two teams are consistent and amazing: The high z supernovae are farther away than they would be in a decelerating or coasting universe. Taken at face value, this implies the need for a cosmological constant to provide acceleration. The Supernovae Cosmology Project fit to the Hubble diagram implies

$$0.8\Omega_m - 0.6\Omega_\Lambda = -0.2 \pm 0.1$$

or

$$\Omega_m = 0.3 \pm 0.1 \quad \text{for } \Omega_m + \Omega_\Lambda = 1$$

1.7 Summary: what is Ω_m?

What conclusions can we draw from the various Ω_m measurements? Table 1.1 summarizes the results for the methods we have discussed.

An important issue is whether or not there is a component of the dark matter density distribution that is homogeneous on the dynamical scale probed by the test particles that are used, since the motion of test particles (whether stars, gas, or galaxies) is not effected by a uniform mass density. This is strongly related to the famous question of whether "mass follows light", since we are not allowed to make measurements using dark test particles.

Figure 1.2 shows the data from dynamical estimates of Ω_m as a function of scale. There is clear evidence for dark matter clumped up on scales of a few hundred kpc. However, there is not much evidence for any increase in M/L ratios measured on scales larger than this. It is striking that galaxy clusters do not need to have more dark matter than can be accounted for in the extended (~ 200 kpc) halos of their constituent galaxies. On the largest scales probed by dynamical methods, the results are ambiguous because of large systematic errors, with some measurements hinting that M/L continues to increase towards values compatible with $\Omega_m = 1$.

Luckily, we have two Ω_m estimates available that do not depend on scale, from SnIa plus inflation and the cluster baryon argument, and these support the claim that we live in a sub critical-density universe.

Table 1.1: Selected recent estimates of Ω_m.

Method	Scale [Mpc]	Ω_m	Comments.
Visible mass		0.003	
Galaxy M/L (optical)	0.01	>0.004	
Galaxy M/L (radio)	0.03	>0.02	
Galaxy M/L (satellites)	0.1	>0.15	
Cluster M/L (virial)	1	0.15-0.35	
Cluster M/L (X-ray)	1	0.10-0.35	
Cluster M/L (lensing)	1	0.2-0.5	
Velocity- density comparison.	100	>0.25	assume b>1
Void outflow velocity	100	>0.30	
Cluster baryon fraction	∞	0.3-0.5	
Type 1a Supernovae	∞	0.2-0.4	

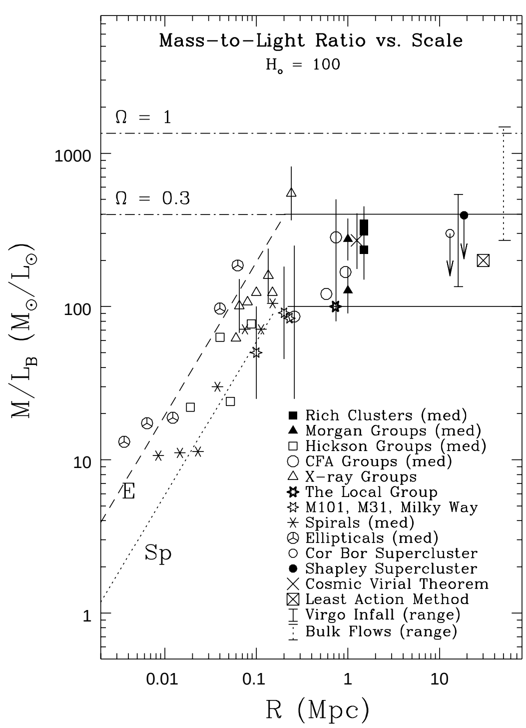


Figure 1.2: Mass-to-Light ratio as a function of dynamical scale. Figure from Bahcall *et al.*, 1995 [10].

We draw the following conclusions:

- The best fit to the data is $\Omega_m \simeq 0.3$ with $\Omega_\Lambda \simeq 0.7$. The evidence for non-zero cosmological constant comes solely from the SnIa measurements and should not be trusted until confirmed by another method.

- A remarkable diversity of observational evidence points to $\Omega_m > 0.2$. Although each of the available measurement methods suffers from some possible systematic error, this diversity makes it difficult to believe that the lower bound could be evaded.

- The upper bound on Ω_m is less firm, but the evidence suggests $\Omega_m < 0.5$. There is little evidence for matter clumped on length scales beyond ~ 1 Mpc. However, the evidence that Ω_m *doesn't* increase on larger scales is not yet completely compelling, so it's still possible that $\Omega_m = 1$.

1.8 Big bang nucleosynthesis.

The theory of big bang nucleosynthesis (BBN) powerfully constrains the amount of dark matter that can be in the form of baryons. This theory predicts the ratio of light elements (^{2}H, ^{3}He, ^{4}He,^{7}Li) to photons in the early universe based on just a few input parameters, such as the initial baryon density, the number of neutrino species, the weak interaction coupling, and the cross sections for nuclear reactions among the light elements at energies of $\sim$1 MeV. Of these input parameters, only the baryon density has significant uncertainty, so the observed light element abundances can be used to measure it. Since the BBN constraint on the baryonic dark matter density is the most compelling reason to search for non-baryonic dark matter particles, it is worth exploring this result in some detail.

The initial condition for the formation of the light elements is the hot universe at $t \sim 1$ second after the big bang. The universe is assumed to contain a net baryon and lepton number, created by some high energy process, the details of which are unimportant. At this time, the temperature is $kT \sim 1$ MeV, and compound nuclei can not exist since they would be disintegrated by black body photons.

The baryonic particles will be in the form of free neutrons and protons, which are exchanged by weak interactions such as $p + \overline{\nu}_e \Leftrightarrow n + e^+$. The ratio of neutrons to protons is determined by statistical equilibrium and is simply

$$\frac{n}{p} = \exp -\frac{Q}{kT} \simeq \exp -1.3\sqrt{\frac{t}{sec}}$$

Here Q =1.3 MeV is the $n - p$ mass splitting, and we have used the time evolution of the temperature,

$$\frac{kT}{MeV} \simeq \left(\frac{t}{sec}\right)^{-1/2}, \tag{1.7}$$

which can be found by solving the Friedman equations for the case where the expansion is driven by the energy density of the black body photons, electrons, and neutrinos.

If the neutrons could stay in thermal equilibrium with neutrinos as temperatures fell, there would soon be no neutrons left. However, the weak interaction cross sections which maintain equilibrium are proportional to E^2 and therefore have a time dependence,

$$\sigma_{n\nu} \sim G_F^2 E^2 \sim G_F^2 (kT)^2 \sim \frac{5 \times 10^{-44}}{t/\mathrm{sec}} \mathrm{cm}^2 \tag{1.8}$$

The small cross section and the dilution of the number density as space expands cause the neutrons to go out of equilibrium with the neutrino bath before the neutrons are all gone. This happens roughly when the mean free time of a neutron to neutrino scattering becomes equal to the time it will take the universe to double in size, or

$$(<\sigma_{n\nu} v> n)^{-1} = (\dot{a}/a)^{-1} \tag{1.9}$$

Here n is the neutrino number density, which is decreasing as $1/a^3 \propto 1/T^3$, and $v \simeq c$ is the relative velocity of the neutrons and the neutrinos. The brackets $<>$ denote a thermal average. Solving the equation, we find that equilibrium holds only to ~1 MeV or ~1 second after the Big Bang. More careful calculations show that the $n-p$ ratio is frozen in when $kT \simeq 0.7\ MeV$ in the ratio $n/p = 16\%$.

These neutrons are the starting material for nucleosynthesis, beginning with the reaction $n+p \Rightarrow ^2 H$. The final products are ^2H , ^3He, ^4He, and ^7Li. The fact that there are no stable nuclei with mass 5 or 8 inhibits the buildup of more massive nuclei during the ~10 minute period when the temperature is high enough for the reactions to occur. Almost all the available neutrons are ultimately burned into ^4He, with only trace amounts of the other elements left over. The left-over densities of the incompletely burned elements depend on the density of initial baryons, so we can infer the value of the initial baryon density from the measured light element abundances.

The dominant uncertainty in determining the baryon density from BBN comes from uncertainty in the measurements of the primordial abundances, rather than from ignorance of the correct nuclear physics or cosmological input parameters. It has long been recognized that a good measurement of the deuterium abundance would provide the best constraint on the baryon density Ω_b , since the deuterium abundance varies the most as Ω_b changes. The measurement itself is difficult, since deuterium is easily destroyed by stellar processing, and, until recently, only a lower limit on the primordial density was available. In the last few years, this situation has changed with the arrival of the Keck 10 m telescope, which is able to observe the splitting between ^2H and ^1H Ly-α absorption in very high redshift clouds. Since 1994, there have been a number of detections between z=2.5 - 3.5, with the result $D/H = (3.4 \pm 0.25) \times 10^{-5}$. This seems to be a true measurement of virgin deuterium, since the clouds are very old and have very low metalicity, indicating little destructive stellar processing. These measurements imply [31]

$$\Omega_b = (0.045 \pm 0.006)(h/0.65)^{-2}. \tag{1.10}$$

This is consistent with the less stringent constraints coming from ^3He, ^4He, and ^7Li.

The ^4He abundance provides a particularly good check of the overall BBN picture, because its value is well measured and fairly insensitive to Ω_b. The theoretical value, $^4He/^1H = 0.246 \pm 0.0014$ [31] is currently in very mild contradiction with some of the published measurements (e.g. 0.232 ± 0.003 (stat) ± 0.005 (sys) [32]). However, there are other measurement that do agree with the theory, which suggests that the magnitude of systematic errors has been underestimated.

The multiple constraints from observed abundances have so far defeated serious attempts to undermine the BBN constraint on the fraction of the dark matter that could be baryonic. Many such proposals, including modifications to the FRW

Table 1.2: Inventory of known baryonic dark matter [34].

Location	$\Omega_b(h = 0.65)$
Stars in spiral galaxies	0.0007
Stars in ellipticals	0.0015
Gas in rich clusters	0.0011
total	0.0033

expansion rate by alternative theories of gravity, the existence of new particle species, or inhomogeneity in the initial distribution of nucleons, seem to have failed. It seems that the BBN prediction that $\Omega_b \simeq 0.05$must be taken quite seriously. Since this is less than the lower limit on Ω_m, the dark matter has to be mostly non-baryonic.

1.9 Baryonic dark matter.

The nucleosynthesis constraint $\Omega_b \simeq 0.05$ implies that there must be a lot of baryonic dark matter, as well as non-baryonic matter, since, as shown in Table 1.2, this is about an order of magnitude more material than can be accounted for in known form. This situation has often prompted the observation that there are really two dark matter problems, the problem of the missing non-baryonic matter and the problem of the missing baryonic matter.

Where are the dark baryons? They could be either in the form of gas or in "dark stars", such as brown dwarfs, white dwarfs, or black holes. In both cases, there are significant observational constraints, but also big loopholes that could accommodate the missing material.

Gas can be observed by looking for the Ly-α emission or absorption from neutral hydrogen (HI), or the X-rays from ionized hydrogen (HII). Non-observation of the emission from HI and HII constrains the contribution of these forms of gas to be less than 1% of the mass of our galactic halo. It is still possible, however, that the halo has a lot of cold molecular (H_2) gas condensed into clumps [33].

Beyond the halo, constraints on the amount of Ly-α absorption in quasar light (the Gunn- Peterson effect) tell us that smoothly distributed HI contributes no more than $\Omega_b \sim 10^{-8}$. Ionized intergalactic gas at high temperatures could be detected by its imprint from Compton scattering on the cosmic microwave background. It could be the dominant contributor to Ω_b only if its temperature is below about 10^8K. Several authors argue that medium temperature gas ($10^5 - 10^6$K) is therefore the best bet for being the baryonic dark matter [35].

The constraints on compact objects as dark matter in our halo have recently become extremely tight due to aggressive searches for gravitational microlensing [44], in which an intervening object causes apparent magnification of a background star. The probability that this will occur at a given time (with magnification >1.34, corresponding to a star's image passing inside the Einstein radius of a lens) for a given background star at distance R is known as the optical depth τ and is related to the density $\rho(r)$ along the line of site by [44]

$$\tau = \frac{4\pi G}{c^2} \int_0^R \rho(r) \frac{r(R-r)}{R} dr.$$

So far, microlensing searches have found lensing of stars in the galactic bulge, the Large Magellanic Cloud (LMC), and the Small Magellanic Cloud (SMC). Observations towards the LMC and SMC are much more sensitive to dark objects in the halo (known as Massive Compact Halo Objects or MACHOs), than lensing towards the bulge, since our line of sight to the bulge passes largely through the disk.

We will discuss only LMC results, since they are better established than the more recent SMC observations. Two collaborations, known as MACHO and OGLE, have detected microlensing events towards the LMC. For example, in an analysis of 2.3 years of data, the MACHO collaboration found 8 events in a population of 8.5 million stars [36]. They calculate an optical depth $\tau = 2 \times 10^{-7}$, and expect $\tau = 5 \times 10^{-7}$ for an all-MACHO halo. The simplest explanation of the data is that the halo is around 50% MACHOs. However, uncertainty in the parameters that describe the halo, bulge, and disk complicate the interpretation of the data. Gates *et al.* [37] have attempted to take these uncertainties into account by generating a Monte Carlo distribution of galaxy models that reproduce the measured LMC and bulge optical depths while producing the correct rotation curve. They obtain a likelihood function for the halo MACHO fraction f peaked at f =0.2, with a long tail that does not completely exclude f =1.

The duration of a lensing event depends on the mass, velocity, and position of the lens. If the assumption is made that the lenses are halo objects, with velocities typical of the halo, then the observed durations of the events, which were between 34 and 145 days, suggest a mass of about 0.5 $M_\odot$. If these objects are stars, they are massive enough to burn hydrogen and should be directly observable. In fact, direct observations with the Hubble space telescope have completely ruled out the possibility that the halo consists of stars over the 0.08 $M_\odot$ hydrogen burning limit and these limits can be extended to cover brown dwarfs ($< 0.08 M_\odot$) by using star-formation theory to extrapolate the observed number density to lower mass [38]. This leaves open the possibility that the MACHOs are stellar remnants, possibly white dwarfs, neutron stars, or black holes. In this case, it's difficult to hide the heavy elements that would have been produced before the stars collapsed, since most of the material in the star is expected to be ejected before or during gravitational collapse. This would lead to contamination of the halo, which, in fact, seems to have a very low heavy element content.

The problem of explaining what causes the long-duration lensing events has led to suggestions that the objects may be an unexpected population of ordinary stars located somewhere besides the galactic halo, possibly in the LMC itself. It seems that further observations including more viewing angles, higher statistics, and some number of binary lenses (which can be used to break the degeneracy of mass, velocity, and distance) will eventually resolve this question.

Perhaps the most important result of the microlensing searches, based on the non-observation of short duration events, is that MACHOs with masses in the range $10^{-7} - 10^{-2} M_\odot$ can not make up more than 20% of the halo. This limit is complementary to the limit imposed by the fact that objects smaller than $10^{-7} M_\odot$, known as "iceballs", do not gravitate strongly enough to avoid evaporation.

1.10 Light neutrinos.

The BBN constraint that $\Omega_b \approx 0.05$ together with the evidence for $\Omega_m \geq 0.3$ tells us that there must be non-baryonic dark matter. Theories beyond the Standard Model of particle physics suggest a number of particle candidates. Perhaps the simplest possibility is that one or more of the three known neutrino flavors, ν_e , ν_μ , or ν_τ, has a mass. These particles have a big advantage in that they are already known to exist! There is no fundamental reason for them to be massless. As long as the masses are significantly less than $kT \sim 1$ MeV at the time of the

Table 1.3: Neutrino mass limits.

	Mass Limit	Experiment
ν_e	< 8 eV	Tritium β decay endpoint.
ν_e	$<$15 eV	SN 1987A
ν_e (Majorana only)	< 0.7 eV	$^{76}Ge\ \beta\beta$ decay.
ν_μ	< 160 KeV	$\pi \rightarrow \mu\nu_\mu$ decay at rest.
ν_τ	< 24 MeV	$\tau \rightarrow \nu_\tau +$ hadrons

weak interaction decoupling, they will have a present day density independent of mass and will contribute

$$\Omega_\nu \simeq \frac{m_\nu}{40\ eV}(h/0.65)^{-2} \tag{1.11}$$

The current limits on the masses of the 3 known neutrinos are given in Table 1.3. They are not very stringent from the point of view of cosmology, with only the electron neutrino having any useful constraint. In fact, the best limit on the masses of ν_μ and ν_τ comes from the observation that $\Omega < 1$, which implies $m_{\nu_\mu} + m_{\nu_\tau} < 40\ (h/0.65)^2$ eV.

If the neutrinos do have small masses and the mass eigenstates are not the same as the weak eigenstates, then oscillations between the three flavors can occur. A number of neutrino oscillation experiments have reported positive results. These experiments fall into three classes [39]:

1. *The solar neutrino deficit.* The composition, temperature profile, and neutrino production mechanisms of the Sun are believed to be well understood. There are 8 principle nuclear reactions that produce ν_e's, with energies extending up to 19 MeV and a total expected flux $\sim 10^{11}$ cm^{-2} s^{-1} . The

predicted energy spectrum is believed to be known to better than 10% accuracy. However, five experiments of three types have measured neutrino fluxes that are significantly below these predictions. There seems to be no plausible modification of the solar model that could simultaneously accommodate all the observations, so attention has turned to the possibility that the ν_e's produced in the Sun are oscillating into some other species on the way to the Earth. This other species could be ν_μ , ν_τ , or ν_s, a "sterile" neutrino that has no weak couplings (e.g. a right-handed state). A likely possibility is that the oscillation is enhanced by coupling to matter in the Sun. In that case, the mass difference between the two neutrinos should be $\Delta m^2 = \left| m^2_{\nu_e} - m^2_{\nu_X} \right| \sim 10^{-5}\mathrm{eV}^2$. A vacuum oscillation mechanism is also possible, with $\Delta m^2 \sim 10^{-10}\ \mathrm{eV}^2$.

2. *The atmospheric neutrino deficit.* The dominant neutrino production mechanism in the atmosphere is pion decay via the reactions $\pi^- \to \mu^- + \overline{\nu}_\mu$ and $\mu^- \to e^- + \overline{\nu}_e + \nu_\mu$ (+ charge conjugates), with the pions produced in hadronic showers initiated by high energy primary cosmic rays. Without oscillations, the ratio of ν_μ to ν_e detected at ground level (or below!) should therefore be about 2:1. Four experiments show that there are too few ν_μ's and $\overline{\nu}_\mu$'s. The recent Super-Kamiokande results are quite convincing [40]. In a 33 kiloton-year exposure, they find only 0.65 ± 0.13 of the expected ratio of $\nu_\mu + \overline{\nu}_\mu$ to $\nu_e + \overline{\nu}_e$. More significantly, they find a very strong dependence of the neutrino deficit on the zenith angle of the final state muons, indicating that the effect depends on the neutrino propagation distance. A good fit is obtained for $\nu_\mu \leftrightarrow \nu_\tau$ oscillations and $\Delta m^2 = 2 \times 10^{-3}\mathrm{eV}^2$. Explaining the data as $\nu_\mu \leftrightarrow \nu_e$ is difficult. In principle, $\nu_\mu \leftrightarrow \nu_s$ is a possibility, but this may cause

problems for ^4He nucleosynthesis [39].

3. *The LSND experiment.* The LSND experiment has found evidence for $\nu_\mu \to \nu_e$ oscillations. This experiment looks for the appearance of ν_e and $\overline{\nu}_e$ in a beam of ν_μ and $\overline{\nu}_\mu$ from pion decay. In three years of data, they see 22 events consistent with $\overline{\nu}_e p \to e^+ n$, with an expected background of 4.6 events. The LSND results, when combined with other accelerator and reactor neutrino limits, seem to permit 0.2 eV$^2 <\Delta m^2<$ 15 eV^2 for small mixing angles [41].

How can we synthesize all the data into a coherent picture? The simplest solution is to ignore the LSND result and assume the mass hierarchy of neutrinos imitates that for e , ν , and τ , with $m_{\nu_e} \ll m_{\nu_\mu} \ll m_{\nu_\tau}$. In that case, the atmospheric neutrino deficit implies $m_{\nu_\tau} \sim$0.1 eV and the solar neutrino deficit implies $m_{\nu_\mu} \sim 10^{-3}$- 10^{-5} eV. The ν_e mass is presumably even smaller. In this scenario, even the τ neutrino is not a very significant form of dark matter, accounting for only

$$\Omega_{\nu_\tau} \sim 0.001. \quad \text{(atmospheric + solar } \nu \text{ anomalies.)}$$

An alternative possibility is to invoke a sterile neutrino to explain the solar neutrino shortage and assume LSND is seeing $\nu_\mu \to \nu_e$. In that case, the ν_μ and ν_τ are degenerate in mass, with $m_{\nu_\tau} \approx m_{\nu_\mu} = 0.5 - 4$ eV. This is an interesting mass range for cosmology, implying a substantial neutrino mass density,

$$0.03<\Omega_\nu< 0.2. \quad \text{(atmospheric + solar + LSND)}$$

If the LSND experiment is to be believed, there is still a small window open for neutrinos to be the non-baryonic dark matter. In fact, light particles such as neutrinos (generically known as “hot dark matter”) are considered unlikely to be the dominant dark matter component because they are still relativistic during the epoch when galaxy-sized density perturbations cross into the horizon and their pressure will inhibit gravitational collapse of protogalaxies. The observed structures in the universe seem more likely to have been formed by heavy dark matter particles, known as "cold dark matter", which can begin to collapse earlier. If the dark matter is 4 eV neutrinos and $\Omega_m = 0.3$, new physics will be needed to help make structures. On the other hand, if in fact $\Omega_m = 1$, as may still be possible, a mixture of neutrino dark matter ($\Omega_\nu \sim 0.2$) with cold dark matter ($\Omega_{\rm CDM} \sim 0.8$) gives an good fit to observed structure [42].

1.11 Weakly Interacting Massive Particles (WIMPs)

The first heavy particle candidate suggested as cold dark matter was a 4th generation heavy neutrino. As we have discussed, neutrinos that were relativistic at the time of the decoupling of the weak interaction ($t \simeq 1$ sec, $kT \simeq 1$ MeV) must have masses less than ~ 40 eV to be consistent with the observation that $\Omega_m \leq 1$. This is the case because the present day abundance of particles which were relativistic at the time they decoupled is determined by the decoupling time, which is fixed by the annihilation cross section rather than the mass. However, this would not be true for a neutrino that was heavy enough to be non-relativistic at decoupling, because the abundance is then suppressed by the Boltzmann factor $\exp -m_\nu/kT$. Naively, one might expect that this leads to a number density after decoupling approximately equal to the light neutrino (or photon) density at the decoupling time times the appropriate Boltzmann factor. In fact, this is an under-estimate for the likely case where the heavy neutrinos carry a conserved quantum number and are only allowed to annihilate in pairs, because the Boltzmann suppression of the neutrino number density shifts the decoupling to an earlier time.

In this case, the number density n can be found by solving a simple differential equation, first suggested by Lee and Weinberg [43],

$$\frac{dn}{dt} = -\frac{3\dot{a}}{a}n - < \sigma_A v > n^2 + < \sigma_A v > n_0^2. \tag{1.12}$$

In this equation, n_0 is the density at thermal equilibrium. The first term on the right hand side is the dilution of n due to the expansion of space. The terms proportional to $< \sigma_A v >$, the thermal average of the product of the annihilation cross section and the velocity, represent annihilation and production of new particles. These cancel when $n = n_0$. Numerical solution of this equation shows that at late times, n approaches the equilibrium value for $kT \simeq m_\nu c^2/20$. The corresponding cosmological bound on the heavy neutrino mass, often referred to as the "Lee - Weinberg" bound, is $m_\nu > 2$ GeV/c^2.

The heavy neutrino could not be any of the known neutrinos because the limits on the mass (See table 1.3) are already too low. New neutrinos with masses below or slightly above $M_Z/2 \simeq 35$ GeV are excluded because they would make a visible contribution to the width of the Z_0 resonance. Above this mass range, searches for interactions with detectors on Earth have already ruled out the possibility that the dark matter is made of neutrinos with standard weak couplings and masses from 10 GeV/c^2 up to a few TeV/c^2 [56]. However, a number of theories beyond the Standard Model predict the existence of new particle species which would have cross sections, masses, and therefore abundances similar to heavy neutrinos, but would not have been discovered yet because the cross sections for interaction with ordinary matter are smaller. These particles are generically called Weakly Interacting Massive Particles or WIMPs.

We will use the symbol χ to represent the WIMP. To a first approximation, the contribution of χ to the dark matter will be [51]

$$\Omega_\chi h^2 \simeq \frac{3 \times 10^{-27} \mathrm{cm}^3 \mathrm{s}^{-1}}{< \sigma_A v >}.$$

It's an interesting fact that typical weak-scale interactions and masses lead naturally to $\Omega\chi \sim 1$. This has been interpreted by many people as a hint that the dark matter problem is related to the remaining problems of particle physics at the weak scale.

The leading candidates to be the WIMP are the neutral superpartners of Standard Model particles predicted by supersymmetry (SUSY), a theory in which every known fermion has a undiscovered boson partner and every boson has a fermion partner. For example, the neutrinos (spin 1/2) are paired with sneutrinos (spin 0), and the photon (spin 1) with a photino (spin 1/2). Clearly, this symmetry must be broken at currently accessible energy scales, since we haven't yet observed any of these extra particles.

Perhaps the most important reason to believe that supersymmetry is a description of the real world is that it appears to be an essential step on the road to unification of gravity in a quantum theory with the other forces. For this purpose, SUSY could be broken at an almost arbitrarily high energy scale (less than $m_{\mathrm{planck}} \sim 10^{19} \mathrm{GeV}/c^2$) and would not have to produce new weak-scale phenomena such as WIMPs. However, weak-scale supersymmetry would solve one of the last remaining problems of the standard model, known as the "gauge hierarchy problem".

The problem is that the mass of the Standard Model Higgs particle has radiative corrections at each order of perturbation theory that pull it up its mass to the scale of the more fundamental high energy theory in which it is embedded (whatever it is). In turn, the Higgs pulls up the mass of all the other Standard Model particles through its vacuum expectation value. To keep the Higgs light, the radiative corrections must be compensated for by readjusting the masses and couplings at each order of perturbation theory, which is aesthetically very unattractive. Supersymmetry solves this problem because the superpartners contribute diagrams with opposite signs, exactly canceling the radiative corrections. To preserve a weak-scale Higgs, the splitting in energy between ordinary particles and superpartners can not be much bigger than about 1 TeV/c^2. This implies the existence of new particles with the right masses and interaction cross sections to be WIMPs.

Most versions of SUSY contain a multiplicative quantum number called R parity which is +1 for all Standard Model particles and -1 for all superpartners. Conservation of this quantum number implies that the decay products of superpartners must always include other superpartners, so the lightest superpartner (known as the LSP) must be stable and is a natural dark matter candidate.

Much of the interest in supersymmetric dark matter has been focused on the lightest neutralino predicted by the simplest possible realization of SUSY, the Minimally Supersymmetric Standard Model (MSSM). The neutralinos are linear combinations of the four electrically neutral, color neutral bosons required by the MSSM, the wino, bino (or equivalently photino and zino), and two higgsinos. The contribution of each of these particles to the physical (mass eigenstate) neutralino depends on a number of model-dependent parameters of the theory that describe how supersymmetry is broken.

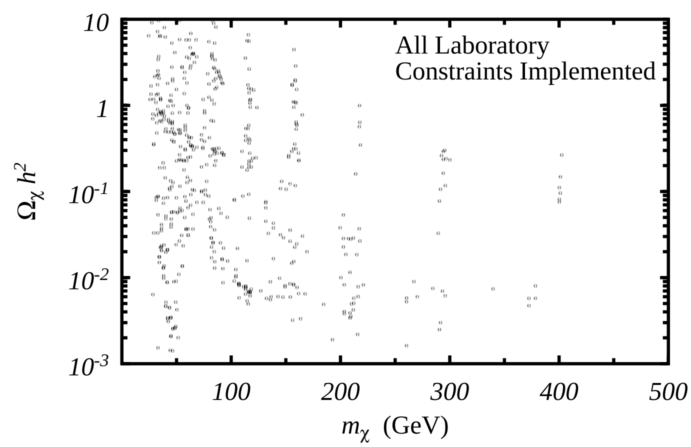


Figure 1.3: Scan of MSSM parameter space for WIMP dark matter from Jungman, Kamionkowski, and Griest, 1995 [51]. The authors reduce the MSSM parameter space to 5 parameters: the mass of the pseudoscalar Higgs, the ratio of Higgs vacuum expectation values, a gaugino mass, the higgsino mass scale, and a common slepton and squark mass scale. These parameters are scanned to produce 2×10^5 models, which are cut to eliminate ones which violate known (circa 1995) particle physics constraints. The early universe annihilation cross sections for the lightest neutralino are calculated for each model and used to calculate relic abundances. The figure shows the distribution of Ω_χ and m_χ for models that survived the cuts. The "spikes" evident at higher masses are an artifact of the discrete sampling of the parameter space.

Since the interactions of the wino, bino, and higgsinos can be exactly calculated in a given model, the neutralino interactions can also be calculated. The lightest neutralino can have quite different properties depending on what these parameters turn out to be, so the neutralino is not guaranteed to be the dark matter even if the MSSM is true. If the MSSM turns out to be true, it will be possible to measure all the unknown parameters with accelerator experiments and directly calculate the neutralino dark matter density. In the absence of these measurements, the best we can do is scan through possible parameter values and calculate the amount of dark matter that results. An attempt at this is shown in Figure 1.3, which demonstrates that $\Omega_\chi \sim 1$ for a large fraction of the parameter space. In fact, in many cases supersymmetry predicts $\Omega \gg 1$, which is inconsistent with observations, so cosmology can be used to constrain the SUSY parameters.

Accelerator experiments do place increasingly stringent limits on the mass of the LSP, through direct searches for the neutralinos (via a missing energy signature) or through chargino searches, which constrain the underlying mass matrix parameters. The best results, which currently come from the LEP collider, imply that $m_\chi \geq 32$ GeV/c^2 [46]. The LSP could be discovered in the next few years at LEP, or in Run II of the Tevatron.

Chapter 2

WIMP detection.

In Chapter 1, we introduced the evidence for the existence of non-baryonic dark matter and discussed a few of the forms it could take, including the possibility that it is made of Weakly Interacting Massive Particles (WIMPs), such as those predicted by supersymmetry. In this chapter, we discuss the possibility of detecting these particles by observing their interactions with detectors on the Earth.

2.1 Interaction of WIMPs with detectors on Earth.

As we have discussed, spiral galaxies generically have large mass -to-light ratios that could be explained by the presence of a dark halo. The Milky Way is no exception; there is evidence from a wide array of observations that our galaxy has a flat rotation curve, with a tangential velocity

$$v_0 = 220 \pm 20 \text{ km/s}$$

between radii of 2 and 15 kpc. Measurements of satellite galaxies imply that the curve remains flat out to 230 kpc, suggesting a mass-to-light ratio of ~100 [9]. Whatever the dark matter is, there seems to be plenty of it in our own galaxy. If it is WIMPs, it may be possible to measure their interactions with detectors in a laboratory.

2.1.1 The WIMP- nucleon interaction.

Although the WIMPs may interact with either electrons or nucleons in a detector, conservation of energy and momentum prohibit the transfer of significant amounts of energy to electrons because of the large disparity in mass. We will be interested, therefore, only in the nuclear interaction.

If the WIMPs are the neutralinos, they have neutral current reactions with ordinary particles via the exchange of Z^0's, Higgs particles, and squarks. These interactions can be incorporated into an effective neutralino-quark Lagrangian, which has the general form

$$L = \frac{g^2}{M_W} \sum_q \left(\bar{\chi}\gamma^\mu \gamma_5 \chi \bar{\psi}_q \gamma_\mu [V_q + A_q \gamma_5] \psi_q + \bar{\chi}\chi S_q \bar{\psi}_q \psi_q + \bar{\chi}\gamma_5 \chi P_q \bar{\psi}_q \gamma_5 \psi_q \right) . \quad (2.1)$$

The coefficients V_q, A_q, S_q, and P_q can be calculated within a given SUSY model by summing over Feynman diagrams. There is no vector coupling for χ, because $\bar{\chi}\gamma^\mu\chi$ vanishes for Majorana fermions such as the neutralino. There are a few general constraints on the relative strengths of the terms in the Lagrangian coming from the fact that both the neutralinos and the quarks are non-relativistic, with $v/c \sim 10^{-3}$for the neutralino and $\sim 10^{-1}$ for the quarks . In this case, the dominant terms are the ones proportional to A_q and S_q, the axial vector and scalar interactions [45].

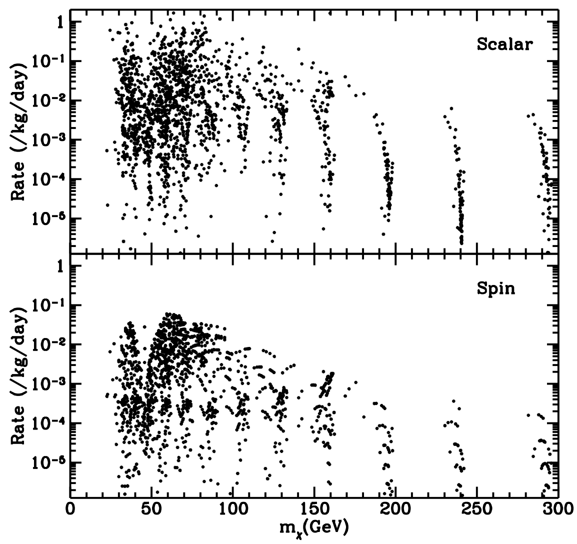


Figure 2.1: Rates for WIMPs scattering on a ^{73}Ge target for the MSSM models of Figure 1.3. From Ref. [51].

The axial vector term comes from Z^0 and squark exchange, and the scalar term comes from Higgs and squark exchange. The A_q and S_q couplings depend on the masses of the exchanged particles and the composition of χ in terms of winos, binos and Higgsinos. Complete expressions are given for many models in Refs. [51, 52, 45].

In the non-relativistic limit, the axial vector interaction becomes an interaction between the neutralino and nucleon spin ($\propto \mathbf{s}_\chi \cdot \mathbf{s}_\mathrm{N}$), while the scalar interaction is spin-independent. The de Broglie wavelength of the neutralino is

$$\lambda = \frac{h}{m_\chi v_\chi} \simeq 0.9\,\mathrm{fm}\ \cdot \left(\frac{m_\chi}{100\,\mathrm{GeV}}\right)^{-1} \left(\frac{v_\chi}{220\,\mathrm{km}/s}\right)^{-1}. \tag{2.2}$$

This is similar to a nuclear radius, $R_N \simeq 0.3 + 0.9A^{1/3}$ fm, which implies that the scattering amplitudes for each of the nucleons in the nucleus can add coherently.

For the spin-dependent interactions, coherence makes relatively little difference, because the amplitudes for opposite- spin pairs of nucleons cancel. For the spin-independent interaction, coherence can increase the cross section by a factor of up to A^2. This compensates for the fact that the single nucleon spin-independent matrix elements are smaller, and generally leads to spin-independent rates that are an order of magnitude larger than the spin-dependent ones. This is illustrated in Figure 2.1, which compares the rate of spin-independent and spin-dependent scattering expected in a ^{73}Ge target for a set of MSSM parameters considered in Ref. [51].

Because of this advantage for the spin-independent interaction, it has attracted the most experimental interest. In the remainder of this chapter, we will discuss only this case. It is important to keep in mind, however, that the spin-independent rates are heavily suppressed for some varieties of WIMP. Searches for spin-coupled WIMPs will become more interesting if the next generation of spin-independent experiments find nothing.

2.1.2 Interaction rates and spectra- general formulae.

In terms of the WIMP - nucleus cross section $\sigma_{\chi N}$, the rate of interactions expected in a detector is

$$R_0 = \left(\frac{M_\mathrm{d}}{m_N}\right)\left(\frac{\rho_h}{m_\chi}\right)\sigma_{\chi N} < |v| > . \tag{2.3}$$

In this equation, M_d is the detector mass, m_N is the mass of the detector nuclei (for a detector assumed to be made of just one element), ρ_h is the halo mass density, m_χ is the WIMP mass, and $< v >$ is their mean velocity.

It is often more useful to know the energy spectrum of nuclear recoils in a detector $dR(E_R)/dE_R$. This is

$$\frac{dR(E_R)}{dE_R} = \left(\frac{M_{\rm d}}{m_N}\right)\left(\frac{\rho_h}{m_\chi}\right)\int_0^\infty f(v)v\frac{d\sigma_{\chi N}(E_R,v)}{dE_R}dv. \tag{2.4}$$

The function $f(v)$ is the velocity distribution of halo particles ($v = \sqrt{\mathbf{v}^2}$ and $\int_0^\infty f(v)dv = 1$), and $d\sigma_{\chi N}/dE_R$ is the differential cross section for WIMP- nucleus scattering.

2.1.3 The local WIMP flux.

Unfortunately, the density of the dark matter halo at the location of the Earth is not very tightly constrained by the rotation curve, because the mass inside the Earth's radius (8.5 kpc) may have significant contributions from matter in the disk and bulge. Determining the local halo density involves fitting noisy data on galactic structure with a mass model that has many free parameters, so its not surprising that there is considerable uncertainty. Estimates for the local density range from 0.05 to 1.0 GeV/c^{2}cm^3. Moreover. as we saw in Chapter 1, the halo is not necessarily made entirely of WIMPs; there is an outside chance it has a substantial MACHO fraction [37].

For the purpose of calculating event rates in this dissertation, we will use the standard value of the halo density ρ_h =0.3 GeV/c^{2}cm^3 suggested by the Particle Data Book [47], and assume that the halo is entirely made of WIMPs.

Theories of galaxy formation indicate that particles in the dark halo should be relaxed into a thermal (Maxwell-Boltzmann) velocity distribution, with a mean particle velocity $< v_\chi >= (2/\sqrt{\pi})v_0$, and an r.m.s. velocity $< v_\chi^2 >^{1/2}= (3/2)v_0$.

The velocity distribution in the galactic rest frame is

$$f(v_\chi) = \frac{4v_\chi^2}{\sqrt{\pi}v_0^3}\exp -v_\chi^2/v_0^2. \tag{2.5}$$

This basic picture is subject to two important modifications. First, the velocity distribution can not follow Equation 2.5 up to $v = \infty$, because we do not expect to find particles with velocity greater than the galactic escape velocity $v_{\rm esc} = 550 \pm 100$ km/s in the halo. This effect can be approximated by truncating to $f(v) = 0$ for $v > v_{\rm esc}$. Since only $\simeq$0.2% of the distribution of Equation 2.5 falls above $v_{\rm esc}$, this has little effect for experiments that have an energy threshold for detection comparable to the mean WIMP kinetic energy. It can be important, however, for experiments that are only sensitive to the highest-energy particles.

For most experiments, a much more important effect is the motion of the Earth with respect to the galactic rest frame. The Sun moves on a circular orbit around the galactic center with a velocity $\simeq v_0$ and the Earth moves around the Sun with a velocity of 30 km/s in an orbit inclined 60° relative to the galactic disc. The velocity of the Earth as a function of time in the galactic frame is

$$v_{\rm E} \simeq v_0\left[1.05 + 0.07\cos\frac{2\pi}{(1\ \text{year})}(t - t_0)\right]\ \text{km/s}. \tag{2.6}$$

where t_0 is June 2nd ± 1 day [48]. The resulting velocity distribution in the Earth's rest frame is

$$f_E(v_\chi) = \frac{4v_\chi^2}{\sqrt{\pi}v_0^2}\left(\exp - \left(v_\chi^2/v_0^2 + v_E^2/v_0^2\right)\frac{\sinh 2v_\chi v_E/v_0^2}{2v_\chi v_E/v_0^2}\right). \tag{2.7}$$

A detector on Earth will see a non-isotropic flux, with a daily modulation in the average incident WIMP angle and a yearly modulation in the velocity. The $\pm 7\%$ velocity modulation results in a modulation in the average number of scattering events and in the average scattering energy [55]. This modulation could be useful for subtracting backgrounds, which hopefully have no (or at least different) time dependence. The daily modulation in the incident angle could be detected by measuring the nuclear recoil direction [49], or through the fluctuations in rate introduced by the shadowing effect of the Earth [53].

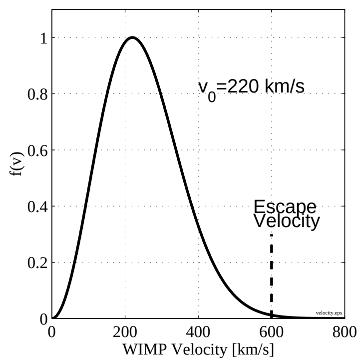
2.1.4 Explicit calculation of spectra.

In the limit where the de Broglie wavelength of the WIMP is much larger than the nuclear radius, the interaction becomes both coherent and isotropic. This is true because the coherence condition (Equation 2.2) guarantees that the scattering is entirely S-wave, since $m_\chi v_\chi R_N < \hbar$ is the semi-classical angular momentum. In this limit, the differential cross section and spectrum of interactions in a detector can be determined from kinematics alone, with no particle or nuclear physics inputs (except, of course, for the overall rate). Let's first explore this limit, before introducing the corrections that apply when coherence breaks down. These corrections turn out to be fairly mild in the situations that are of practical interest.

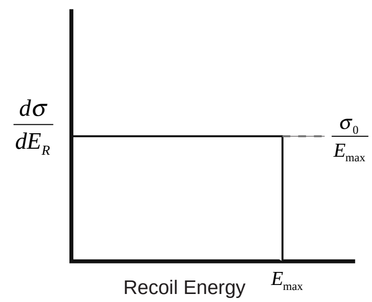
We will write the CM momenta before the WIMP- nucleon collision as $\mathbf{p}_\chi = -\mathbf{p}_\mathrm{N}$ and after the collision as $\mathbf{p}'_\chi = -\mathbf{p}'_N$. The magnitudes of all these momenta are equal to a common momentum p. The kinetic energy is

$$E_\mathrm{cm} = \frac{\mathbf{p}_\chi^2}{2m_\chi} + \frac{\mathbf{p}_N^2}{2m_N} = \frac{p^2}{2\mu},$$

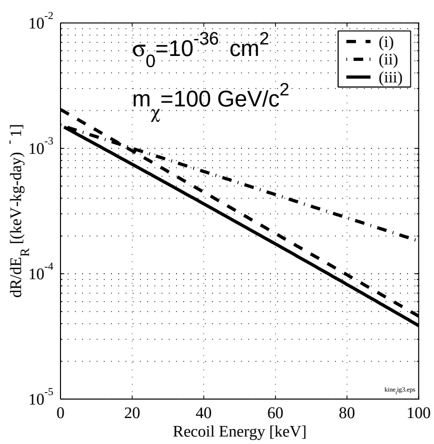
where $\mu \equiv m_\chi m_N/(m_\chi + m_N)$. In terms of the lab frame WIMP velocity v_χ,



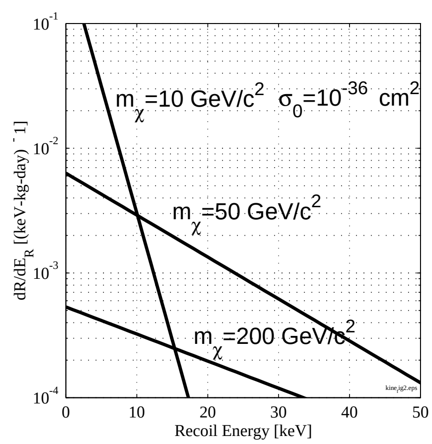
(a) Velocity distribution of halo WIMPs in the galaxy rest frame.



(b) Recoil spectrum for WIMPs with a single velocity v_χ.



(c) Spectra in a germanum detector. (i) is for a detector stationary in the galactic frame. (ii) includes the average motion of the Earth, and (iii) includes also the effect of a form factor.



(d) Spectra in a germanium detector expected for WIMPs with m_χ=10, 50, and 200 GeV$/c^2$, and a common cross section on Ge of $10^{-36}cm^2$..

Figure 2.2: WIMP velocity distribution and recoil spectra.

$$E_{\rm cm} = \frac{1}{2}\mu v_\chi^2.$$

The momentum transfered to the nucleus is

$$q^2 = (\mathbf{p}_\chi - \mathbf{p}'_\chi)^2 = \left(\mathbf{p}_\chi^2 + \mathbf{p}'^2_\chi - 2\left|\mathbf{p}_\chi\right|\left|\mathbf{p}'_\chi\right|\cos\theta_{\rm cm}\right)$$

$$= 4\mu E_{\rm cm}\left(1 - \cos\theta_{\rm cm}\right).$$

The momentum transfer is frame-invariant, so the recoil energy in the lab frame is simply

$$E_R = \frac{q^2}{2m_N} = \frac{\mu^2 v_\chi^2}{m_N}\left(1 - \cos\theta_{\rm cm}\right).$$

The maximum amount of energy is transfered when the WIMP backscatters ($\cos\theta = -1$), resulting in a recoil $E_{\rm max} = \frac{2\mu^2 v_\chi^2}{m_N}$. This can be as much as the full kinetic energy of the incoming particle if $m_N = m_\chi$. To get a recoil of a given energy E_R, there is a minimum possible velocity $v_{\rm min} = \sqrt{\frac{m_N E_R}{2\mu^2}}$.

Since the cross section is isotropic, we can write it as $d\sigma_{\chi N}/d\cos\theta_{\rm cm} = \sigma_0/2$. Then the recoil energy spectrum is

$$\frac{d\sigma_{\chi N}}{dE_R}(E_R, v_\chi) = \frac{d\sigma_{\chi N}}{d\cos\theta_{\rm cm}} \cdot \left(\frac{dE_R}{d\cos\theta_{\rm cm}}\right)^{-1} = \begin{cases} 0 & v_\chi < v_{\rm min} \\ \frac{\sigma_0}{E_{\rm max}} & v_\chi > v_{\rm min} \end{cases}$$

This spectrum from mono-energetic WIMPs is illustrated in Figure 2.2b. To get the spectrum for the velocity distribution present in the halo, we substitute this cross section into Equation 2.4 using the velocity distribution of equation 2.5:

$$\frac{dR}{dE_R} = \frac{2M_d\rho_h\sigma_0}{\sqrt{\pi}\mu^2 m_\chi v_0^3}\int_{v_{\min}}^{\infty} v\exp -v^2/v_0^2 dv \tag{2.8}$$

$$= \frac{2M_d\rho_h\sigma_0}{\sqrt{\pi}\mu^2 m_\chi v_0}\exp -E_R/E_0. \tag{2.9}$$

Here $E_0 = \frac{2m_\chi^2 m_N v_0^2}{(m_\chi+m_N)^2}$. The integrated rate is

$$R_0 = \int_0^\infty \frac{dR}{dE_R}dE_R = \frac{2}{\sqrt{\pi}}\left(\frac{M_d}{m_N}\right)\left(\frac{\rho_h}{m_\chi}\right)\sigma_0 v_0. \tag{2.10}$$

This is the answer we expect from Equation 2.3.

These equations are only valid for perfectly coherent scattering and for a detector at rest in the galactic frame. We will now discuss the modifications that occur when these assumptions are dropped. To account for the velocity of the Earth, we need to use the velocity distribution of Equation 2.7 instead of Equation 2.5. The resulting spectrum is

$$\frac{dR}{dE_R} = \frac{\sqrt{\pi}v_0}{4v_E}\frac{R_0}{E_0}\left[\mathrm{erf}\left(\frac{v_{\min}+v_{\mathrm{E}}}{\mathrm{v}_0}\right) - \mathrm{erf}\left(\frac{v_{\min}-v_{\mathrm{E}}}{\mathrm{v}_0}\right)\right].$$

If we are mainly interested in the spectrum at some particular instant in time, this can be approximated by [61]

$$\frac{dR}{dE_R} = c_1\frac{R_0}{E_0}\exp -c_2\frac{E_R}{E_0}.$$

The coefficients c_1 and c_2 appropriate for a given month of the year are tabulated in Ref. [61]. The average values are $c_1 = 0.751$ and $c_2 = 0.561$.

The breakdown in coherence of the interaction over the nucleus that occurs when $qR_N > \hbar$ can be described by the inclusion of a nuclear form factor $F^2(q)$ [50, 61, 51], so that

$$\sigma = F^2(q)\sigma_0.$$

The origin of the form factor can be understood from the Born approximation, where the differential scattering cross section of two particles interacting through a central potential $V(r)$ has the form

$$\frac{d\sigma}{d\mathbf{q}^2} \propto \left| \int e^{-i\mathbf{q}\cdot\mathbf{r}/\hbar} V(\mathbf{r}) d^3\mathbf{r} \right|^2.$$

Since the weak interaction is point-like within the nucleus, the potential $V(\mathbf{r})$ can be written as the product of the probability of finding a nucleon at position $\mathbf{r}$ in the nucleus times the strength of the WIMP-nucleon interaction, or $V(\mathbf{r}) \propto f_s \rho_N(\mathbf{r})$. Since we know that $\sigma \to \sigma_0$ as $q \to 0$, we absorb all factors that do not depend on q into σ_0 and define

$$F^2(q) = \left| \int e^{-i\mathbf{q}\cdot\mathbf{r}/\hbar} \rho(\mathbf{r}) d^3\mathbf{r} \right|^2.$$

A commonly used form factor is derived from a nuclear density that has the form $\rho(\mathbf{r}) = \int d^3r' \rho_0(\mathbf{r}')\rho_1(\mathbf{r} - \mathbf{r}')$, with ρ_0 a constant inside a sphere of radius $R_0 = \sqrt{R_N^2 - 5s^2}$ and $\rho_1(\mathbf{r}) = \exp -\frac{1}{2}(r/s)^2$ [50]. The parameter s is the nuclear skin thickness, ~ 1 fm. This distribution is similar to the Woods-Saxon parameterization of the nuclear density, but has the advantage that it results in a form factor that is analytic:

$$F^2(q) = 3\frac{j_1(qR_N/\hbar)}{qR_N/\hbar}e^{-(qs)^2/2}.$$

Putting together all the pieces, the expression we will use to calculate the nuclear recoil spectrum is

$$\frac{dR}{dE_R} = c_1\frac{R_0}{E_0}F^2(E_R)\exp\left(-c_2E_R/E_0\right). \tag{2.11}$$

In Figures 2.2c and 2.2d, we show the contributions of the various factors to the nuclear recoil spectrum of a typical WIMP, and what the spectrum would be for particles of mass 10, 100, and 1000 GeV/c^2 incident on a germanium target.

2.2 Experimental strategies.

In a sense, WIMPs are not very difficult to detect, since their scattering on nuclei would produce signals in many types of conventional radiation detectors. For example, scintillation counters, semiconductor detectors, and gas counters are all capable of detecting nuclear recoils of a few keV. Unfortunately, these instruments are also very efficient detectors of environmental radiation, such as cosmic rays and gamma rays from trace radioisotopes present in construction materials. Detectors exposed to environmental radiation in an unshielded room typically register at least 10^5 events a day per kg of detector mass and WIMPs can produce no more than 1 event per day. The energy deposited, $\sim$10 keV, is low relative to the energies of environmental photons, which have a large flux up to 2.6 MeV (the endpoint of ^{208}Tl, the highest emitter in the ^{238}U decay chain). In the region around 10 keV, the differential background counting rate of an unshielded detector is typically $\sim 10^3$ events/ (keV-kg-day). This can be reduced to ~ 0.1 event/(keV-kg-day) in an underground laboratory with carefully designed shielding. Recently, there have been a number of efforts, including the one described in the later chapters of this dissertation, to increase sensitivity by several more orders of magnitude by using detection methods that can discriminate between nuclear recoils and the electron recoils caused by background photons.

A large variety of target nuclei can be incorporated into WIMP detectors. In order to compare the various experiments, it is useful to report results in a way that depends as little as possible on which nucleus was used. The generic form of the WIMP-nucleus cross section is

$$\sigma_0 = \frac{4\mu^2}{\pi}\left[Zf_p + (A-Z)f_n\right]^2.$$

The constants f_p and f_n are couplings to protons and neutrons. In most cases, $f_p \simeq f_n$, so $\sigma_0 \simeq (4/\pi)\mu^2 A^2 f_n^2$. One possible convention would be to reduce the results of all experiments to constraints on f_n. Another possibility is to give results in terms of the cross section which the WIMP would have if the target were a single nucleon (i.e. ^{1}H):

$$\sigma_{Wn} = \frac{m_n^2 m_\chi^2}{(m_n+m_\chi)^2}\frac{1}{A^2\mu^2}\sigma_0 \simeq \frac{m_n^2}{A^2\mu^2}\sigma_0.$$

This is the convention we will use below.

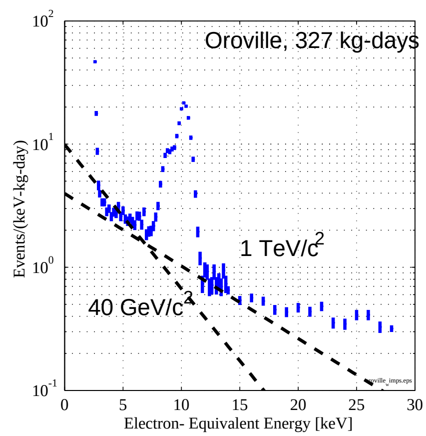
2.2.1 Methods for calculating limits.

Even if an experiment does not detect a clearly identifiable WIMP signal, the data may be useful. We can use it to set limits on the possible cross sections and masses that the particles could have. The conventional way to do this is to assume a fixed local halo mass density and mean velocity and use this to calculate the number of interactions that would occur in the detector based on the formulae given in Section 2.1. Some statistical method is used to calculate whether the observed number of counts is compatible with the expected number for a given m_χ and σ_0. The result is an "exclusion plot", which shows what values of the mass and cross section can be ruled out at a given confidence level. In this section, we will review a few of the ways to calculate these excluded regions.

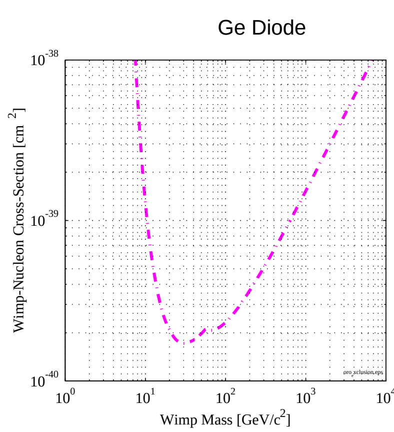
High statistics measurements.

In the high-statistics limit, where the spectrum of the WIMPs plus background radiation is known with infinite precision, calculating excluded regions is simple. Typically, the background radiation can not be safely subtracted from the measured spectrum, because its origin is unknown (or else the experimenter would have eliminated the source). WIMPs which would produce more counts than are observed in any interval of the spectrum are excluded, but nothing can be said about ones that would have made fewer events than were observed.

In practice, many existing experiments operate near the high-statistics limit. As an example, in Figure 2.3a we show data from a Ge detector that was operated under a dam in Oroville, CA by a collaboration of UC Santa Barbara, UC Berkeley, and the Lawrence Berkeley National Laboratory [56]. The Oroville experiment measured its background counting rate to 20% or better in the low-energy region.



(a) Measured spectrum shown with two Wimps that can be excluded (m_χ = 40 and 1000 GeV/c^2 and $\sigma_{Wn} = 10^{-40}$ and 10^{-39} cm^2).



(b) Combinations of m_χ and σ_{Wn} that can be excluded by the data in Figure (a).

Figure 2.3: The Oroville experiment: an example of a high statistics measurement.

Aside from the contribution of some X-ray peaks, including a large feature at 10.4 keV, the composition of the low energy spectrum is not well understood, and can not be significantly reduced by subtracting contributions from known sources (removing the X-ray peaks makes little difference). The sensitivity to WIMPs can be estimated by ignoring statistical uncertainty in the measured rate and finding the largest cross sections that make fewer counts than were observed in all energy bins. At a given WIMP mass, the limit always comes from one "most sensitive" bin. Spectra for two WIMPs that are just excluded by the data are shown in the figure. In Figure 2.3b, we show the envelope of all such curves in the m_χ - σ_{Wn} plane. This gives a reasonable approximation to results obtained with more sophisticated statistical methods.

Single bin limit- application to zero background measurements.

If there are no candidate signal events detected, calculation of excluded regions of parameter space is nearly as simple as in the high-statistics case. A limit on masses and cross sections can be found from the condition that the integral of Equation 2.11 must have no more than the Poisson upper limit on the true counting rate at some confidence level. By "true counting rate" we mean the rate that would be measured by averaging the results of an infinite number of repetitions of the experiment. Usually, we use a 90% confidence level. The lower limit on the WIMP-nucleon cross section σ_{Wn} will be called $\sigma_{90}(m_\chi)$ and the corresponding upper limit on the number of observed events $\mu_{90}(m_\chi)$. If an experiment runs for a time T and has an energy threshold δ, then the condition that defines the excluded region of parameter space is

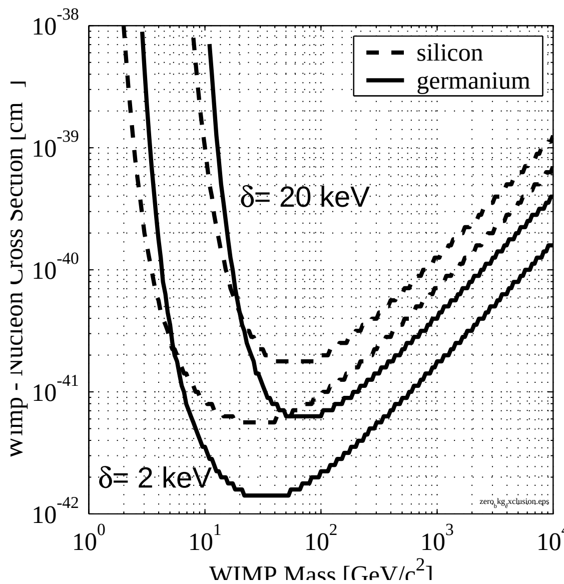


Figure 2.4: Zero background limits for a 5 kg-day exposure of germanium and silicon detectors with recoil energy thresholds $\delta = 2$ keV and $\delta = 20$ keV.

$$\int_{\delta}^{\infty} \frac{dR}{dE_R} dE_R < \frac{\mu_{90}}{T}. \tag{2.12}$$

The standard definition of the 90% confidence level upper limit for a Poisson process with N observed events is [47]

$$\sum_{n=0}^{N} \frac{e^{-\mu_{90}} \mu_{90}^{n}}{n!} = 10\%. \tag{2.13}$$

This equation must be solved numerically to find $\mu_{90}(N)$ for $N > 0$. The result for $N = 0$ is $\mu_{90} = -\ln(0.1) = 2.3$.

Figure 2.4 shows what regions of the parameter space could be excluded using Equation 2.12 for 4 hypothetical experiments: silicon and germanium detectors with energy thresholds of $\delta = 2$ and $\delta = 20$ keV are each run for 5 kg-days with no events seen. These experiments would rule out the strongest-interacting of the MSSM WIMPs ($\sigma_{Wn} \sim 5 \times 10^{-42}\mathrm{cm}^2$). For a given m_χ, the sensitivity to σ_0 is increasing linearly with the product of counting time and detector mass. This figure illustrates the importance of using large A materials such as Ge, which are more sensitive for all but the lightest WIMPs because of the A^2 enhancement in the cross section.

At some point in the future, real experiments may become sophisticated enough to discriminate perfectly between background and WIMP events, or may be so well shielded that environmental radioactivity is unimportant. In that case, Equation 2.12 will be all that we need to calculate sensitivity.

Low statistics measurements.

If the background counting rate is not measured to be zero, and it is not measured to a high degree of statistical accuracy, the problem of calculating excluded regions becomes much more complex. Many methods have been proposed for finding confidence intervals in this situation (See e.g. [57, 60]). Here we will just discuss two of the methods that are currently being used by the CDMS collaboration.

Most methods start by dividing the spectrum of observed events into energy bins. Suppose that we have done this and have K energy bins with lower and upper energy limits for each bin of δ_i^1 and δ_i^2 ($i = 1...K$). These bins have N_i counts in them. Each bin can be considered an independent dark matter detection experiment, from which it is possible to set limits on m_χ and σ_{Wn} using the K conditions

$$\int_{\delta_i^1}^{\delta_i^2} \frac{dR(\sigma_{Wn}, m_\chi)}{dE_R} dE_R < \frac{\mu_{90}(N_i)}{T}.$$

The sensitivity of the K "experiments" will vary because of variations in the ratio of the signal to background rate in each bin. For a given m_χ, some bin will be the most sensitive.

A method that has commonly been employed to make exclusion plots is to simply use the best bin for each mass, or, equivalently, to use the envelope of all the K confidence level contours. However, this is not really a correct way to combine the confidence intervals of multiple experiments. The problem is that the bin with the largest downward fluctuation in counting rate is the one that sets the limit. In cases where there are a large number of bins, each of which has only a few counts, this can produce "limits" that are much too aggressive.

To rigorously combine the results of the multiple experiments that constrain σ_{Wn} at a given m_χ, we construct a single likelihood function for the result $n_1 ... n_k$:

$$L(n_1 ... n_K; \sigma_{Wn}, m_\chi) = \prod_{i=1}^{K} \left(\frac{e^{-\mu_i(\sigma_{Wn}, m_\chi)} \mu_i^{n_i}(\sigma_{Wn}, m_\chi)}{n_i!} \right).$$

Here $\mu_i(\sigma_{Wn}, m_\chi)$ is the mean number of events expected in bin i. We define the 90% confidence level upper limit on the cross section σ_{90} to be the value which satisfies the equation

$$\frac{\sum_{n_1=1}^{N_1} ... \sum_{n_K=1}^{N_K} L(n_1 ... n_K; \sigma_{90}, m_\chi)}{\sum_{n_1=1}^{\infty} ... \sum_{n_K=1}^{\infty} L(n_1 ... n_K; \sigma_{90}, m_\chi)} = 10\%. \tag{2.14}$$

This is a straightforward generalization of the single bin confidence level given by Equation 2.13.

The problem with this method is that it is relatively insensitive when there is a large, but unmeasurable background in many of the bins. Because all the bins contribute to the numerator of Equation 2.14, we pay a penalty for looking for dark matter in parts of the spectrum where the background is high and there is no chance of seeing anything above it.

Other methods are possible which are relatively insensitive to the existence of regions with high backgrounds. One of these is the "maximal gap" method proposed in Ref. [58]. This is based on the observation that, for a given m_χ, any measured spectrum has a least-likely gap in energy between some two of the observed events. This "least-likely", or "maximal" gap is the interval where the largest number of events are expected with none observed. The probability of this largest number of missing events being less than x is

$$C_0(x,\mu) = \sum_{k=0}^{m} \frac{(kx-\mu)^k e^{-kx}}{k!} \left(1 + \frac{k}{\mu - kx}\right)$$

where m is the largest integer less than μ/x and μ is the expected number of events in the entire range of the measurement (See Ref. [58] for a derivation). Both μ and x are functions of σ_{Wn} and m_χ. To find the 90% confidence level upper limit on σ_{Wn}, we increase it until $C_0 = 0.9$.

2.2.2 Detectors with discrimination.

There is an interesting class of experiments where, in addition to the energy of scattering events, it is possible to measure some "discrimination parameter" that tends to have a different value for the signal events than for background events. This can be, for example, the decay time of the scintillation pulse in an NaI detector, which is longer for nuclear scattering than for Compton or photoelectric scattering. For the CDMS experiment, an extremely powerful discrimination parameter is the ratio of ionization to deposited heat observed in a low temperature semiconductor.

In this section, we will discuss how discrimination improves limits. For "perfect discrimination", we do not need to have this discussion at all, since all background events can be removed from the data set without removing any of the signal. In that case, limits will be given simply by Equation 2.12. Here, following Ref. [59], we will consider the case where the value of the discrimination parameter is an imperfect indicator of whether an event is signal or background.

Let's call the discrimination parameter Y, with high values indicating that an event is more likely to be background and low values indicating that it is more likely to be signal. To select signal events, we will make a cut at a value Y_c. This cut will accept events for which $Y < Y_c$ and discard events with $Y > Y_c$.

The effect of the cut can be described by the functions $\alpha(Y_c)$ and $\beta(Y_c)$, which are, respectively, the fraction of signal events and background events that will pass the cut. Hopefully, these functions can be estimated in some way, for example, by using calibration data with characteristics similar to what is expected for the signal and background. The total rate of events passing the cut will be

$$R_{\text{pass}} = \alpha R_{\text{signal}} + \beta R_{\text{background}},$$

and the rate failing the cut will be

$$R_{\text{fail}} = (1-\alpha) R_{\text{signal}} + (1-\beta) R_{\text{background}}.$$

These equations can be solved to give the rate of signal events,

$$R_{\text{signal}} = \frac{(1-\beta) R_{\text{pass}} - \beta R_{\text{fail}}}{\alpha - \beta}. \tag{2.15}$$

Unfortunately, our ability to calculate R_{signal} this way in the real experiment will be limited by two types of effects, (1) statistical uncertainties in the measured rates R_{pass} and R_{fail} and, (2) uncertainty in our knowledge of α and β .

Let's assume for the moment that we have some way to figure out α and β to good accuracy, so it will be the statistical effect (1) that dominates. We will assume that $R_{\text{background}} \gg R_{\text{signal}}$. In this case, the variance of R_{signal} from Poisson fluctuations in the number of background events that leak into the signal region is

$$\sigma_s^2 = \frac{\beta(1-\beta)}{(\alpha-\beta)^2}\frac{R_{\text{background}}}{T}.$$

The variable T is the time the experiment was running. If R_{signal} is consistent with zero, the upper limit on the number of signal events is proportional to σ_s, so it will decrease as $1/\sqrt{T}$ and increase as $\sqrt{R_{\text{background}}}$. To get the best limit, we should pick the cut Y_c to minimize the function $Q(E_R, Y_c)$, defined as

$$Q(Y_c) \equiv \frac{\beta(1-\beta)}{(\alpha-\beta)^2}. \tag{2.16}$$

The minimum possible value of Q is a useful figure of merit for the detector, since the ratio of the lengths of time two detectors of equal mass must count to get the same limit is the ratio of their Q's.

2.3 History and current status of detection efforts.

Attempts to detect WIMPs began after the publication of an article by Goodman and Witten in 1985, which first suggested the possibility [62]. This paper proposed the adaptation of a superconducting grain technology recently suggested by Drukier

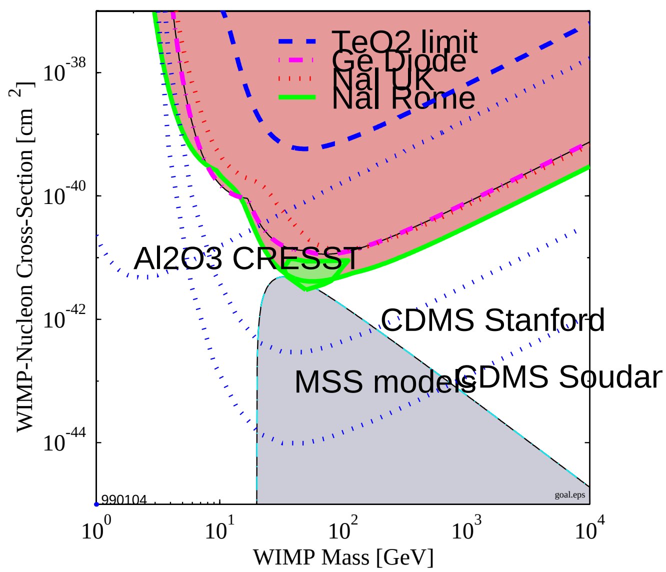


Figure 2.5: Current limits and goals for CDMS I and II.

and Stodolsky for the detection of neutrinos via neutral current scattering [63] . The principle of operation was the measurement of the collapse of the magnetic field around a small superconducting grain as it is driven into a normal state by heat input from a scattering event. This idea is still being pursued, but so far technical problems have prevented the development of detectors with enough mass to have interesting sensitivity.

The first experiments with useful sensitivity to dark matter were modified versions of existing germanium semiconductor spectrometers designed to look for neutrinoless double beta decay, such as the ones used in the Oroville experiment. These typically had a mass of 1 kg and energy thresholds as low as 3 keV for photons. Since nuclear scattering in Ge detectors produces only about 1/3 the ionization of electron scattering, the effective threshold for dark matter detection is no more than about 9 keV of nuclear recoil energy. (One usually refers to “electron equivalent energy" when discussing the amount of energy deposited in a ionization detector, since these detectors are calibrated with photon sources that produce electron scattering events.) The early efforts with these detectors, which began almost immediately after the publication of the Goodman and Witten paper, achieved background counting rates on the order of 1/(keV-kg-day) near threshold [56]. This was useful for probing WIMP- nucleon cross sections down to $\sim 10^{-40}$ cm^2 and masses down to $\sim$ 20 GeV/c^2.

There has been gradual improvement in the technology of background reduction over the decade that has passed since the initial experiments, and the most sensitive germanium experiment now has a rate of 0.05 events/keV-kg-day [65]. In addition to the Ge detectors, some successful efforts were made with smaller, lower threshold ($\sim$ 3 keV recoil energy) silicon devices that were more sensitive to lower-mass WIMPs [66].

The semiconductor detectors held the undisputed lead until the mid 90's, when several groups began to operate large NaI scintillation counters [54, 55]. These can have somewhat lower effective background levels than the Ge detectors due to the employment of discrimination methods based on differing scintillation time constants for nuclear and electron scattering. More significantly, the very massive detectors that are possible with NaI are ideal for observing the annual modulation effect, since high statistics are necessary to see small annual modulations on top of a larger constant background rate.

One NaI experiment now running sees modulation signal that is compatible with WIMPs [55]. This experiment, which is installed in the Gran Sasso underground laboratory in Italy, consists of 13 NaI modules with a total mass of 116 kg. It has a background rate of $\sim$1/(keV-kg-day) and a threshold of 2 keV in electron equivalent energy. In a 14962 kg-day exposure of the detectors, a time dependence in the counting rate is observed between 2 and 6 keV which is consistent with the effect of a WIMP of mass $m_\chi = 59^{+22}_{-14}$ GeV/c^2 and $\sigma_{Wn} = 7.0^{+0.4}_{-1.7} \times 10^{-42}\mathrm{cm}^2$. In the most significant bin (3-4 keV), this modulation is only $\sim \pm 1\%$ of the background counting rate, so there is a worry that the modulation may be due to some small systematic error. An effect of this magnitude could have numerous causes, such as variations in the phototube noise levels, or seasonal modulations in the background rate due to changing weather conditions (e.g. radon gas concentrations are known to have seasonal variations). Clearly, it is important to test this result with an experiment that has higher signal-to-noise.

Many other ideas are under investigation for WIMP detection, including large drift chambers [67], new scintillating materials, ultra-low background semiconductor arrays [68], and the cryogenic techniques that are the focus of this thesis. The new techniques offer the possibility of lower backgrounds, lower thresholds, new target nuclei, and perhaps a chance to measure the nuclear recoil direction.

Figure 2.5 shows the current limits on halo WIMPs from completed or ongoing experiments. This figure also shows the goals for the cryogenic experiments CDMS I and CDMS II, which will be discussed in much more detail in the following chapters. These experiments should be able to probe a significant fraction of the MSSM parameter space.

Chapter 3

The CDMS Experiment.

3.1 Overview.

The main challenge for WIMP detection is to achieve low background counting rates from environmental radiation at low energies in a detector that incorporates at least a few kg of target material. One approach to solving this problem is to build detectors that can discriminate between WIMP-nucleon scattering events and background events. Nuclear scattering produces less ionization in semiconductor targets per unit energy deposited than scattering on electrons. Since gamma rays, which are the dominant background source for dark matter searches, interact with electrons via Compton and photoelectric scattering, while WIMPs make ionization only through the recoil of a target nucleus, a measurement of the ratio of ionization to deposited energy permits discrimination between the two types of interaction. At low temperatures, it is possible to measure both the heat deposited by an ionizing event and the amount of ionization, making this type of discrimination technologically feasible. This is demonstrated in Figure 3.1, which shows data

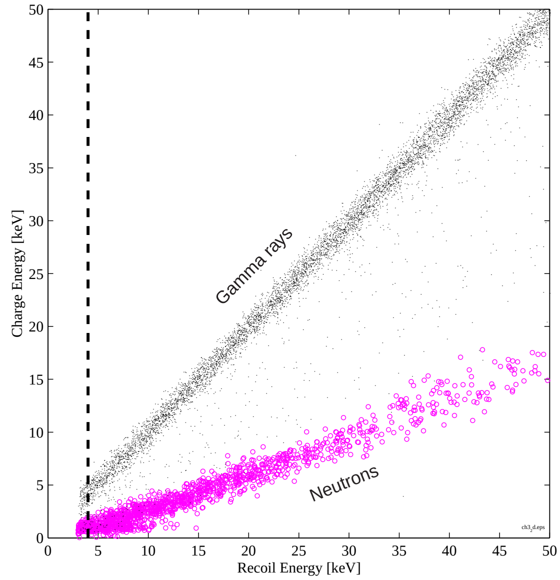


Figure 3.1: Nuclear recoil discrimination in a thermistor-instrumented calorimeter that simultaneously measures charge and heat. This figure combines data from exposures to a gamma ray source (upper line) and a neutron source (lower line). The axes are “Recoil Energy”, the amount of energy actually deposited in the detector by the scattering particle, and “Charge Energy”, the amount of energy that a gamma ray would have needed to deposit to make the amount of free charge that was observed. Gamma rays make events with an efficiency of 1 and neutrons make events with an efficiency of $\sim 1/3$. Noise and incomplete charge collection can cause departures from these ratios. The dashed lines shows the 4 keV threshold at which the detector is fully efficient.

from exposure of a detector to neutrons (which make nuclear recoils) and photons from radioactive sources.

The Cryogenic Dark Matter Search (CDMS) collaboration has built germanium and silicon detectors which operate at a temperature of 20 mK. Energy deposition is measured with two different technologies, (1) neutron-transmutation-doped germanium thermistors capable of measuring temperature changes of less than 1 μK in an attached target crystal, and (2) tungsten superconducting transition-edge

Date	
1985-1986	Cryogenic detector development begins at Berkeley and Stanford: Aluminum TES sensors at Stanford in ^{4}He fridge with FET readout. Small NTD devices in Berkeley at 1.2 K.
1988	75 μW dilution refrigerator installed at Berkeley. Demonstration of ionization/phonon discrimination. Center for Particle Astrophysics approved.
1989	First large ionization/phonon detector at Berkeley: 60 g Ge, $\sim$20 keV f.w.h.m. energy resolution ("E2").
1990	15 μW fridge at Stanford. First W TES sensors (FET readout). CDMS I infrastucture design work starts: Icebox, Tower, Shield, Veto.
1991-1992	Excavation of Stanford Underground Facility (SUF).
1992	Berkeley: NTD-instrumented detectors now down to $<$ 1 keV noise. Stanford: Ti TES devices with ionization measurement.
1994	First W TES with electrothermal feedback and SQUIDs. Later, with 2 W/Al sensors using quasiparticle trapping (QET). Icebox cools down at Berkeley. Shield and veto test assembly at SUF.
1995	NTD-instrumented detectors now $\sim$500 eV phonon noise, with 2 ionization channels having $\sim$ 1 keV noise. A 4 g Si QET device with 4 channels + a 0.25 g device with ionization. Tower cools down in '75: allows 96 wires to 20 mK. Icebox cools down at SUF. Final shield and veto assembly.
1996 Fall	First physics run at SUF (Run 13): Electron background discovered. A 60 g Ge detector with 2 NTDs and 2 ionization channels ("E5"). Run 14 adds a 100 g Si , 4 + 2 channel QET (a "FLIP") .
1997	Runs 15 and 16: E5 + a 165 g Ge detector with 2 NTDs ("BLIP 1"). Run 17: Test of tritium removal from NTD thermometers. Start Run 18: 2 BLIPs + improved FLIP with 3 keV threshold.
1998	Rise-time discrimination shown for surface events in FLIP. Start Run 19: 6 BLIPs, 4 with improved (Al/Si) ionization contacts.
1999 1999?	Demonstration of 250 g Ge detector with 4 QET sensors (a "ZIP"). CDMS II proposal approved. Work begins in Soudan. Run 20: 3 FLIPs and 3 ZIPs.
Oct. 2000?	First run of CDMS II in Soudan mine.

Table 3.1: A history of the Cryogenic Dark Matter Experiment (CDMS).

sensors (TES), which respond to the initial high-energy phonon burst associated with a particle interaction. Both types of detector have a noise threshold of ∼4 keV in recoil energy. In the TES devices, the shapes and arrival times of pulses at a grid of 4 sensors contain information about the location of the event inside the detector.

Both detector types need to be operated in an environment combining very low levels of background ionizing radiation with very low temperatures. The low temperatures are provided by a dilution refrigerator and cryostat capable of cooling down to ∼10 mK. To keep the level of background radiation low, the cryostat is constructed entirely of materials that have been found to have low levels of radioactive contamination. The cryostat is surrounded by a shield made of polyethylene and lead to block neutron and gamma radiation. To reduce radiation produced by cosmic rays, the cryostat and shield are installed in a tunnel 10.6 m underground on the Stanford University campus. A set of plastic scintillator counters surrounding the shield are used to tag events which are caused by the fraction (∼ 20%) of high-energy muons that manage to penetrate into the tunnel.

At the SUF site, it should be possible to reduce background counting rates from neutrons to 0.01 events/keV-kg-day. The long-range plan, known as "CDMS II", calls for moving the experiment to a site much deeper underground, the Soudan mine in Tower, MN. At that site, the neutron rate can be reduced to 10^{-4} events/keV-kg-day. In order to exploit this low background rate, it will be necessary to increase the detector mass to about 10 kg, which will be incorporated in 42 cryogenic detectors.

A summary of the history and planned development of the experiment is given in Table 3.1.

3.2 Phonon measurements.

In this section, we will discuss the phonon measurements made by the CDMS detectors. We will emphasize the "thermal" technology that uses NTD thermistors, rather than the "athermal" technology based on superconducting transition edge sensors, since the analysis presented in Chapter 4 of this dissertation is based on the thermal technology.

3.2.1 Thermal detectors.

The CDMS thermal detectors measure the rise in temperature and free charge produced in a germanium crystal after energy is deposited by radiation. These detectors were developed by the Sadoulet group at UC Berkeley and are often referred to by the acronym BLIP, which stands for "Berkeley Large Ionization and Phonon detector". The results presented in the later chapters of this dissertation come mainly from devices called BLIP 1 and BLIP 2. These detectors were the product of a long development effort that is described in the dissertations of N. Wang, T. Shutt, A. Cummings, and W. Stockwell [72, 73, 74, 71]. Most of the detectors that proceeded BLIP 1 and BLIP 2 were small test devices not intended to achieve meaningful sensitivity to dark matter. The exception is E5, which had a mass of 60 g and performance similar to BLIP 1 and BLIP 2. This detector was used in the first few runs at SUF and achieved impressive sensitivity to WIMPs [83].

Figure 3.2 is a drawing of a BLIP. Table 3.2 summarizes some important device parameters.

Table 3.2: Some BLIP design and performance parameters.

Mass	165 g
Thickness	1.2 cm
Radius	3.0 cm
Surface area	79 cm^2
Phonon energy resolution	375 eV (FWHM)
Charge energy resolution	1200 eV (FWHM)

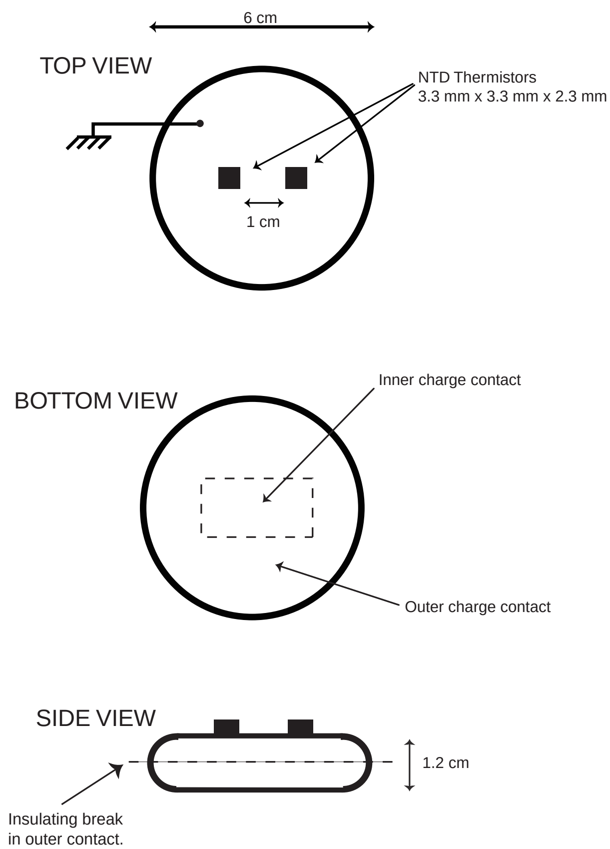


Figure 3.2: The BLIP detectors. The detectors have 2 NTD thermometers and 2 charge segments. The charge segments are created by ion implantation of the surfaces with a boron beam, followed by etching to define breaks between the bias side and the ground side (dotted line in bottom view) and between the two segments on the bias side (dotted line in center view). The top sides of the detectors are an electrical ground for currents from the NTD and charge biases.

The thermal model.

The interaction of a dark matter particle with a BLIP deposits perhaps 10 keV of energy or $\sim 10^{-15}$ J. The detector must have a tiny specific heat for this to result in a measurable temperature increase. This is possible because germanium has few available degrees of freedom at low temperature. Since it is a semiconductor, with a gap between the valence band and conduction band of 0.75 eV and $kT \simeq 5$ μeV at 20 mK, the electron system can not absorb any energy. The dominant contribution to the heat capacity comes from the lattice vibrational modes and is given by the Debye law

$$C_{\text{lattice}} = \lambda_D \left(\frac{T}{\Theta_D} \right)^3 ,$$

where λ_D= 1944 J/mole·°K and the parameter θ_D is a material dependent constant. The Debye law agrees well with measured heat capacities for many materials at low temperature. It predicts a very low heat capacity for germanium (θ_D = 374 K) at 20 mK,

$$C_{\text{ge}} = 2\,\text{keV}/(\text{mole} \cdot \mu\text{K}).$$

This implies that we can expect a temperature change of 2 μK in a BLIP struck by a WIMP depositing 10 keV.

The change of ~0.1% in the temperature of the crystal is measured with a thermistor made of neutron transmutation doped (NTD) germanium. Exposure of high purity germanium to a flux of thermal neutrons from a reactor results in the introduction of acceptor and donor impurities (Ga, As, and Se) formed as the end products of the decay of unstable Ge isotopes after neutron capture [86, 79, 74]. The ratio of donors to acceptors is 0.32. Since this is fixed by nuclear physics, and the thermal neutrons have long mean free path to capture in germanium, this procedure produces extremely low, yet uniform concentrations of impurities.

Conduction in NTD germanium comes from the jumping of charge carriers between localized impurity sites, which are separated by small gaps in space and in energy. At mK temperatures, the phonons required to make up the energy difference for a jump between gaps are far out on the high energy tail of the thermal distribution, and the transition probability is a steep function of temperature. The resistivity has the form

$$\rho = \rho_0 \exp\left(\frac{\Delta}{T}\right)^{1/2}. \tag{3.1}$$

Our thermometers are typically made from material that has $N_A - N_D = 6.28 \cdot 10^{16}$ cm^{-3}, $\Delta \simeq 7.0$ and $\rho_0 \simeq 0.55\ \Omega - \text{mm}$.

In practice, the fit to the shape of the resistance curve given by Equation 3.1 is usually very good, but the values of ρ, ρ_0 and Δ are not very reproducible between samples [74]. This is particularly true for thermistors which are glued on to germanium slabs. For glued sensors, variations of up to an order of magnitude in the resistivity have been observed. It is believed that these variations are caused by mechanical stresses put on the thermometers by contraction of the glue.

Since the method used to mount the thermometers on the target crystal may affect their performance, considerable effort was put into finding a good way to do this. The original intent of this research was to find a way to attach thermometers to detectors that maximized the transparency of the interface to phonons, but interest later shifted to increasing the reproducibility of the thermometer's operating characteristics. Because both the target and the thermometers are made of germanium, it was possible to exploit the existence of a low-melting-point (356 °C) germanium-gold alloy to fuse the thermometer and target together [73]. The procedure, called "eutectic bonding", involves pressing a heated thermometer and target crystal together with 10^3 Åof gold in between. After cooling, the thermometer and target are found to be welded together. Electron micrographs show that the germanium lattice is actually continuous between the two pieces in some regions. This bonding method results in a lower thermal contact resistance and more reproducible values of Δ and ρ_0 for the thermometer than could be achieved by other methods [74, 71]. With eutectic bonding, the resistivity ρ is reproducible between samples up to a factor of about 2.

An extremely important limitation on the performance of the NTD detectors comes from a phenomenon known as electron-phonon decoupling. In general, to measure the resistance of the thermometer as accurately as possible, one wants to apply as large a bias current as possible to maximize the signal $V = IR$. This current delivers a power $P = I^2R$ to the sensor, which will cause its temperature to rise and its resistance to decrease. We expect the function $V(I)$ to be linearly increasing for small bias currents, but to turn over and decrease at a point which is determined by the thermal conductivity between the thermometer and its heat sink. The sensitivity of the thermometer will be greater for higher values of the conductivity. Naively, one might expect that the dominant impedance for heat flow between the thermometer and target crystal is the flow of power across the physical boundary between the two materials. This should have the form

$$P = g_K A(T_s^4 - T_0^4) \tag{3.2}$$

for a thermometer at temperature T_s attached through an interface of area A to a target crystal at temperature T_0. This form of thermal impedance, which is often referred to as Kapitza resistance, is caused by reflection of phonons at the interface due to acoustic mismatch.

In fact, the Kapitza boundary resistance alone can not explain the power flows that are measured in our detectors. The dominant thermal impedance in the path connecting the thermometer to the detector is caused by decoupling of the electron system and the phonon system inside the NTD material. An electron-phonon decoupling phenomenon of this type is well-established in metals at mK temperatures, and it should be expected in doped semiconductors too [73]. The quantum mechanical explanation of this effect is that, by Fermi's golden rule, the scattering rate for electrons and phonons is proportional to the final state phase space, which is small at low temperature. A series of measurements at Berkeley have shown that the temperature dependence of the power flow between the electron and phonon systems can be fit by the expression

$$P = g_{ep} V(T_e^\alpha - T_p^\alpha). \tag{3.3}$$

A good fit to the data is found for $\alpha = 6$ and $g_{ep} = (4.0 \pm 0.5) \cdot 10^{-9}\mathrm{W/K} \cdot \mathrm{mm}^3$ [74].

The response of the BLIPs to a heat input into the target crystal can now be understood in the context of a simple model shown in Figure 3.3. This model includes three thermal baths, (1) the electron system in the NTD thermometer,

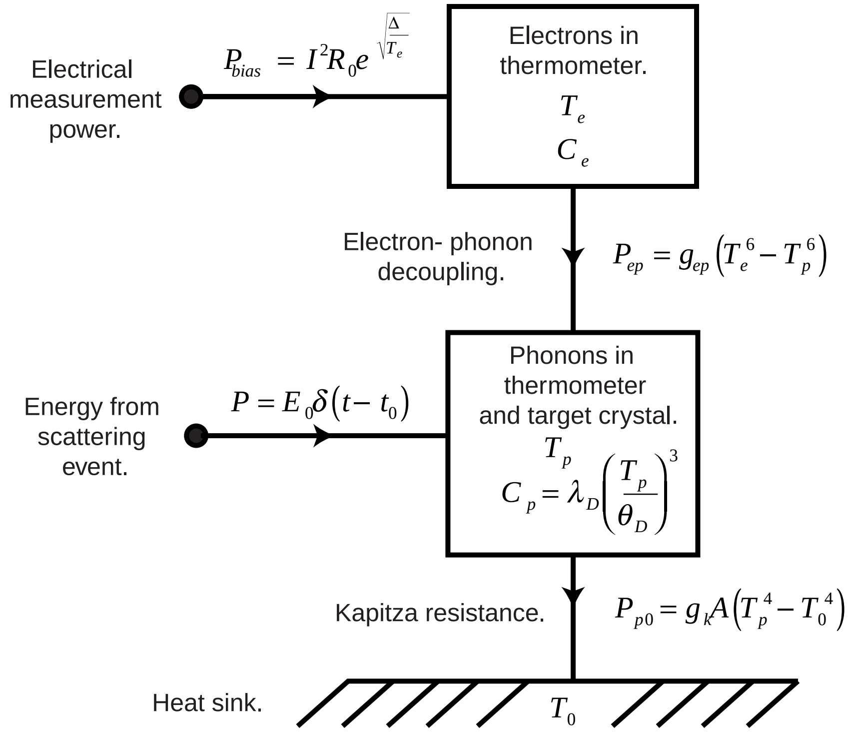


Figure 3.3: The BLIP thermal model. There are two thermal baths, with temperatures T_e (electrons in the thermometer), and T_p (phonons in the thermometer and target) and a heat sink with temperature T_0 (typically 20 mK). These baths are connected by power flows described by the laws for electron-phonon decoupling and Kapitza resistance. There are external heat inputs to the system from scattering events into the phonon system and measurement power into the electron system.

which has a temperature T_e and a heat capacity C_e[1] , (2) the lattice phonons in the thermometer and target crystal, which have a common temperature T_p and a heat capacity given by the Debye law, and (3) the heat sink, in practice the metal parts attached to the mixing chamber of the dilution refrigerator, which have an effectively infinite heat capacity at constant temperature T_0. There are also two important thermal conductivities, the conductance between electrons and phonons in the thermometer, given by Equation 3.3, and the conductance between the lattice phonons and the heat sink, which has the Kapitza form of Equation 3.2. In the BLIPs, the conductance to the heat sink is provided by a thin gold wire bonded to a gold film of area 15 mm^2, which is sputtered onto the surface of the germanium. This has a Kapitza resistance of $\sim 3\times 10^{-4}$W/K^4 .

The model of Figure 3.3 allows us to describe the behavior of the detectors with two coupled first-order differential equations. Since we are mainly interested in pulses corresponding to temperature changes of less than 0.1%, the equations can be linearized around the steady-state solution. The result for the time evolution of small pulses caused by an instantaneous heat input into the phonon system is

$$\delta v(t) = I_{\mathrm{bias}} \frac{dR}{dT_e} \frac{\partial P_{ep}}{\partial T_p} \frac{E_0}{C_e C_p (1/\tau_1 - 1/\tau_2)} (e^{-t/\tau_1} - e^{-t/\tau_2}). \tag{3.4}$$

In this expression, E_0 is the deposited energy, I_{bias} is the bias current, and R is the resistance of the NTD. All temperature-dependent parameters are evaluated for the heat-sink temperature T_0. The time constants τ_1 and τ_2 are functions of

[1] A definitive measurement of the NTD heat capacity at 20 mK has proven to be elusive , and measurements of its temperature dependance are contradictory, with some results showing a mild linear increase in temperature in the 20 - 30 mK region and others showing no dependance on temperature [74]. Estimates of the heat capacity in this temperature range are between 6×10^{12} JK^{-1}mm^{-3} and 1×10^{10}JK^{-1}mm^{-3}.

the equilibrium temperature heat capacities and thermal conductivities (Complete expressions and the derivation of Equation 3.4 can be found in Ref. [72]). They are effectively the time constants for relaxation to equilibrium of the target and heat sink system and of the target and NTD electron system. The rise time τ_2 is determined by the strength of the electron-phonon coupling and is ~ 5 ms for our detectors. The fall time τ_1 can be adjusted by changing the surface area of the heat-sink film. This constant must be kept much larger than τ_2 to maximize the pulse height, which limits these detectors to applications that involve low counting rates. Typically, we choose a heat sink that gives $\tau_1 \sim 30$ ms, about the longest time that can be tolerated in our poorly shielded, above-ground testing facility.

The performance of each detector is optimized by tuning I_{bias} to give the maximum pulse height. In practice, the thermal model tends to over-estimate the pulse height at the optimal bias current by a factor of about two, and also has some difficulty reproducing the observed pulse shapes [71, 74]. It's not clear if the problems with this model are due to missing pieces of physics or inaccuracies in the input parameters, some of which are quite difficult to measure. One interesting possibility that has been considered is that a fraction of the energy from the scattering event is transmitted quickly to the electrons in the NTD by high energy phonons [71].

The resistance measurement.

The resistance of the thermometers is measured using the circuit shown in Figure 3.4. It is essential to minimize noise in the readout circuit, since this directly effects the energy threshold at which WIMPs are detectable. The thermal pulses have a bandwidth of about 250 Hz. Unfortunately, voltage measurements made in

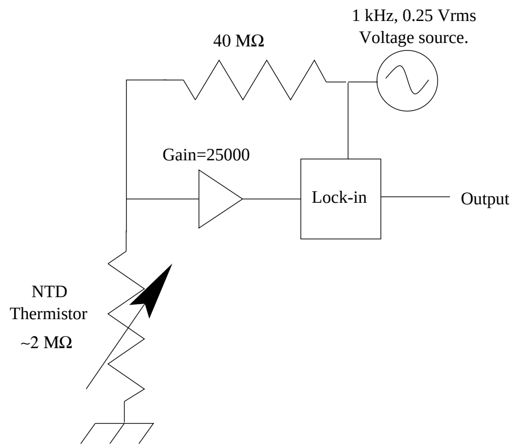


Figure 3.4: Phonon sensor circuit. A lock-in amplifier is used to boost the signal out of the noisy low frequency region.

a dilution refrigerator environment are typically relatively noisy in the 0 to 500 Hz region of the frequency spectrum where the signal power is concentrated, due to $1/f$ noise from a variety of sources and pickup of the first few harmonics of the 60 Hz AC power lines. At SUF, we measure a noise power density at the input to our voltage amplifiers of typically 5- 10 nV/$\sqrt{\text{Hz}}$ below 500 Hz, compared to 1-2 nV/$\sqrt{\text{Hz}}$ at 1 kHz. To take advantage of this decrease in noise at high frequency, we measure the resistance of the NTD thermometer using a 1 kHz bias voltage. This moves the signal into a 250 Hz band centered at 1 kHz. A standard lock-in amplifier technique is used to mix the signal back down to its normal frequency range after it has been transmitted out of the cryostat and amplified.

At 1 kHz, the dominant noise sources are Johnson noise in the NTD, which has a power spectral density $\sqrt{4kTR} \simeq 1$ nV$\sqrt{(1/\text{Hz})(R/1\ \text{M}\Omega)(T/20\ \text{mK})}$ and "FET noise" in the JFET that is the first amplifying element. Both these noise sources have a flat frequency spectrum. The FET noise is due to statistical fluctuations in the number of charge carriers in the conduction channel and is typically 0.7 nV/$\sqrt{\text{Hz}}$

for good FETs. It can vary by a factor of a few between devices, so we minimize it by using hand-picked FETs that have been tested at their optimal operating temperature, about 130 K. The Johnson noise could be reduced by reducing the NTD resistance, but this would also reduce the pulse height. A careful analysis of Equation 3.4 shows that the pulse height at the optimal bias current scales as $\sqrt{R}$, so the signal-to-noise ratio has no R dependence if the Johnson noise is the dominant noise source. The value of the resistor is always chosen to be large enough that this is the case.

We measure a pulse height of $\simeq$ 150 nV/keV (before amplification) in BLIPs using the circuit of Figure 3.4. Since the r.m.s. noise on a measured pulse amplitude is approximately $\sqrt{250 \text{ Hz}} \times 1.5 \text{ nV}/\sqrt{\text{Hz}} \simeq 24$ nV, it is the equivalent of $\simeq$160 eV amplitude difference. This simple estimate agrees well with the more sophisticated treatment described in Appendix A and is roughly consistent with the measured width of low-energy X-ray peaks seen with the detector. This noise level is several times lower than that which can be achieved with ionization measurements for detectors of this size.

3.2.2 Athermal phonon measurements.

When a particle scatters in a crystal, the initial phonons created have a frequency $\sim$10 THz [78]. The phonons are subject to two competing scattering processes. One of these, isotope scattering, is elastic, and simply results in the phonon changing its propagation direction. The other process is anharmonic decay, in which the phonon splits into two lower-energy phonons. In silicon, the mean free time to isotope scattering is

$$\tau_I \sim \left(\frac{1\ \mathrm{THz}}{\nu}\right)^4\ \mu\mathrm{s}$$

and the time to anharmonic decay is

$$\tau_{\mathrm{anh}} \sim 25\left(\frac{1\ \mathrm{THz}}{\nu}\right)^5\ \mu\mathrm{s}.$$

The very steep dependence of these time constants on frequency causes a phenomenon known as “quasidiffusion”. The initial phonons have extremely short mean free paths, but as they decay in energy, they become more and more free. The phonons which ultimately arrive at the surface of the detector (some cm away from the interaction) have a mean frequency ~ 1 THz, corresponding to a temperature of ~ 10 K. They have an effective straight-line velocity of ~ 1 mm/μs. Because the temperature of these phonons is much higher than the temperature of the crystal itself, we often refer to them as “athermal phonons”.

Detecting the athermal phonons at several sensors on the surface of a crystal allows triangulation of the original event location. Unfortunately, these sensors can not be the NTD thermistors discussed in the last section, because the athermal phonons are scattered out of the NTD back into the target crystal, producing small signals. The NTDs respond strongly only after the phonons have degraded in energy to a near-thermal distribution.

The Stanford CDMS collaborators have developed detectors that can respond to the high energy phonons. The sensors are made with tungsten films deposited on the surfaces of germanium and silicon target crystals. These films are biased in the middle of their superconducting transition near 90 mK by a scheme known as electro-thermal feedback. High energy phonons produce quasiparticles in the film and cause a change in the bias current, which is detected with a SQUID current amplifier. Silicon detectors with these sensors have energy resolution on the order of 1 keV and position resolution of ∼1 mm in the plane of the phonon sensor. An important recent development has been the discovery that the shape of the pulses is different for events occurring near the detector surfaces. This allows the rejection of surface background events due to electrons, which, as discussed below, are otherwise difficult to discriminate from nuclear scattering. Progress has also been made recently on controlling the tungsten phase-transition temperature and on producing germanium detectors with high quality films.

Since this dissertation is concerned largely with the interpretation of data from BLIP detectors, we refer the reader to the recent work of R. M. Clarke [77] and S. W. Nam [78] for details on the progress of the transition-edge-sensor technology.

3.3 Ionization measurement.

In some respects, fabrication of semiconductor detectors to be operated at 20 mK is simpler than at higher temperatures. At temperatures where kT is comparable to the energy gap between donor sites and the conduction band or acceptor sites and the valence band, typically ~ 10 meV, there are significant numbers of charge carriers in the crystal. These cause a "leakage current" when an electric field is applied, and the shot noise in the leakage current competes with the much smaller currents induced by particle interactions. To overcome this noise problem, the detectors are built as p-n diodes operated with a large reverse bias. This creates a charge-free depletion region where there is no conduction. The depletion region is the active volume of the detector. It can be as thick as about 1 cm for high-purity silicon detectors and more than 10 cm thick for high-purity Ge detectors operated at 77 K.

At 20 mK, there is no thermal excitation of the impurities, so there is no leakage current. Indeed, it's possible to use relatively low purity materials that would be unacceptable in conventional 77 K detectors[2]. An electric field is necessary only to collect the charge produced by particle interactions and this can be as low as 200 mV/cm for essentially full collection of charge from events generated in the bulk of the crystal [72]. It is fortunate that high fields are unnecessary, since charges drifting in the field produce heat, which makes the measured ratio of ionization to phonons more similar for nuclear and electron scattering. The amount of heat produced by charges drifting after a particle interaction deposits energy δ is simply $eV\delta/\epsilon$, where ϵ is the amount of deposited energy needed to make an electron-hole pair, e is the charge of an electron, and V is the bias voltage. (Note that ϵ is defined for gamma rays, not neutrons or WIMPs.) For germanium at 20 mK, $\epsilon = 3.0$ eV, so in a crystal biased at 1 V, charge drift will add an extra 1/3 to the amplitude of the thermal pulse.

Naively, the extra heat may seem like an advantage, since this is a natural amplification mechanism for the signal. In fact, detectors that exploit this amplification to achieve very high gains have been built [93], and there may be an application to dark matter detection at low energy thresholds. The problem for detectors based on ionization/phonon discrimination is that the presence of noise in the charge measurement makes it impossible to determine the fraction of the measured heat

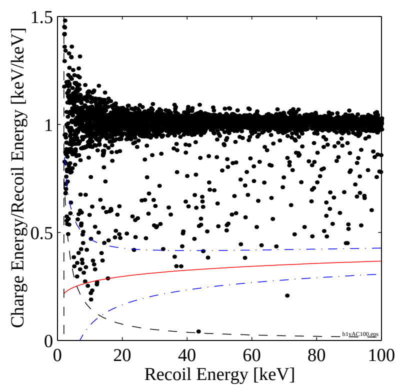
[2]We usually save money by puchasing germanium that has been rejected by a major detector manufacturer. In fact, it still has remarkably high purity, with a net impurity density $n_a - n_d \simeq 6 \cdot 10^{10}$ cm^{-3}.

that comes from the charge drift for individual events. As we show in Appendix B, the gain causes an increase in the noise in the measured ionization-to-heat ratio that reduces discrimination power. This effect begins to become important at $V \sim 3$ Volts, where the contribution of drift charge to the thermal pulse starts to dominate.

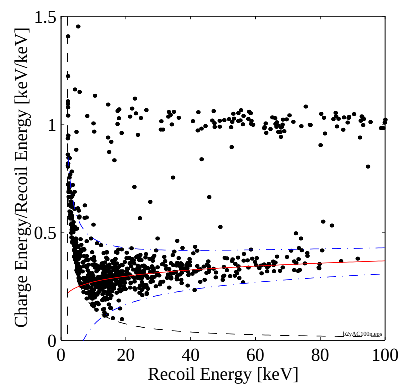
An important subtlety of operating a semiconductor detector at low temperature with low bias fields is the need to keep impurities neutral in order to avoid charge trapping. In the ground state of the crystal, donor impurities have given their electrons to acceptor impurities. This leaves the crystal full of fixed charges, which act as trapping sites for any free charge that may be produced by particle interactions.

Fortunately, at these low temperatures, the state in which all the impurities are neutral is stable, because the electrons attached to donor impurities need to jump up to the conduction band in order to find the acceptors and they can not find enough energy to do this. The problem is that cooling down the detector slowly from room temperature to the base temperature seems to be an effective way to bring the crystal to its ionized ground state. If no active steps are taken to restore neutrality after cooling down, we observe many fewer and, on average, much smaller charge pulses. This problem can be fixed by creating a large number of electron-hole pairs in the crystal while it is cold. Presumably, some of the charge carriers diffuse to the trapping sites and stay there, restoring neutrality. The electron -hole pairs can be made either with an intense radioactive source or with a burst of light from an LED mounted on the detector package. Currently, we usually use the second approach, since our detectors tend to be extremely well shielded from radiation.

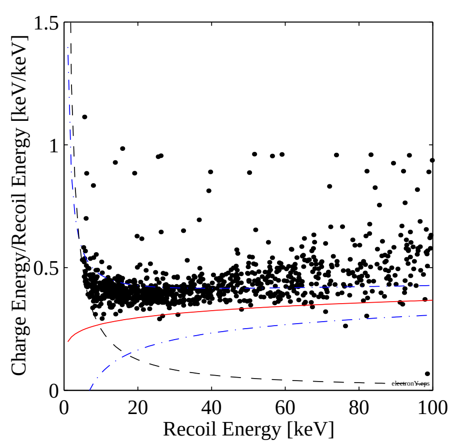
Figure 3.5: Effect of dead layer on charge collection for surface electrons. The data shown is from exposure of a BLIP to gamma rays (a) and neutrons (b), and of a small test detector similar to a BLIP to low-energy electrons (c). The ratio of ionization to recoil energy is plotted as a function of recoil energy. The dashed lines indicate the position of the nuclear recoil band and show the full-efficiency thresholds for triggering on thermal pulses and reconstructing the charge signal.



(a) Exposure of a BLIP to a gamma ray source (^{60}Co).



(b) Exposure of a BLIP to a neutron source (^{252}Cf).



(c) Exposure of a small test device to an electron source (^{14}C with Q=156 keV).

Although charges in the bulk of the detector are efficiently collected with low fields, there is dead layer of poor charge collection near the surface. We believe that the dead layer is due to back-diffusion of wrong-polarity charge (e.g. electrons onto the side where holes are normally collected) onto surface trapping sites. In this model, the thickness of the layer is determined by the diffusion length.

By exposing the detectors to low energy X-rays, we have been able to probe the depth dependence of charge collection in some detail. The fraction of collected charge $q(z)$ seems to have a dependence on depth z well fit by the equation

$$q(x) = (1 - fe^{-z/\lambda})$$

We have experimented with different types of electrical contacts on the surface of the detector in an attempt to control the parameters f and λ. For many years, our standard charge collection contacts were made by ion implantation of boron atoms to a density of 10^{19} cm^{-2}. This makes the surface metallic, and wires can be attached to it by ultrasonic wire bonding onto small sputtered gold pads. These standard contacts are roughly 1500 A thick. For these devices, which include BLIP 1 and BLIP 2, we find $f = 0.7 \pm 0.05$ and $\lambda \approx 6\ \mu m$ at an electric field of 2 V/cm [84].

The dead layer is a serious defect because electrons incident on the detector surface produce ionization-deficient signals that are similar to those made by nuclear recoils. This is illustrated in Figure 3.5, which shows the effect of illuminating detectors with gamma rays, neutrons, and electrons. Electrons produce signals with a ratio of ionization to deposited energy uncomfortably close to the ratio made by nuclear scattering.

3.4 The dilution refrigerator and cryostat.

The design of the CDMS cryogenic apparatus presented many problems. It was necessary to provide very low temperatures, a large cold volume, low levels of background radioactivity and low-noise wiring for hundreds of electronics channels. This led to a large number of engineering constraints which were difficult to satisfy simultaneously. The Ph.D. thesis of P.D. Barnes Jr. [70] gives many of the details on how this was achieved. Here we will just mention some of the most important details.

The CDMS cryogenic system is shown in Figure 3.6. The cooling power is provided by an Oxford Instruments Model 400S dilution refrigerator. Dilution refrigerators can achieve temperatures as low a few mK by mixing He-3 and He-4 in a closed cycle [**?**]. This is the only appropriate technology for achieving temperatures in the mK range for an extended period of time in kilogram-sized samples. Our refrigerator has a nominal cooling power of 400 μW at 100 mK and a base temperature below 5 mK when operated without any experimental package.

Unfortunately, commercially available dilution refrigerators employ a wide variety of materials in their construction and are inevitably too radioactive to directly host a dark matter experiment. This problem was solved by separating the experimental cold volume from the refrigeration apparatus through a horizontal cold finger called the "cold stem", which is about 1 m long. The cold stem reaches through a radiation shield and attaches to the "icebox", which is the cryostat designed to hold detector modules. Another stem, the "electronics stem", carries electrical signals to the outside world on wires and flexible printed circuits.

The icebox consists of six nested copper cans held in normal operation at tem-

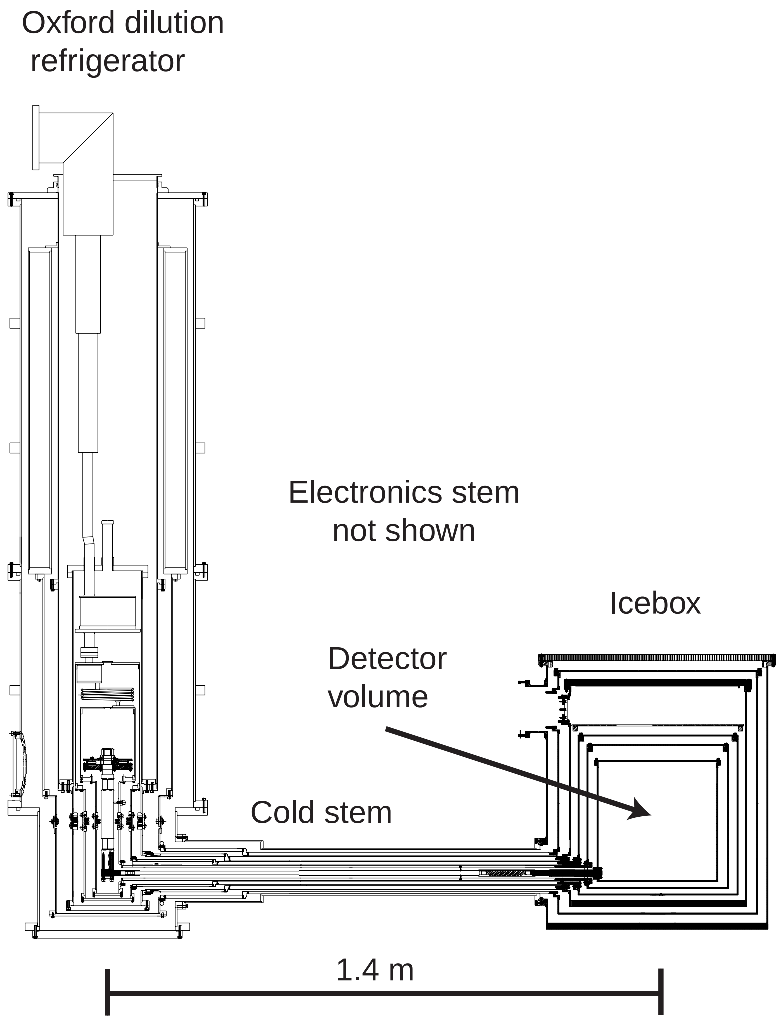


Figure 3.6: The CDMS cryostat, or "icebox", attached to the dilution refrigerator, an Oxford Model 400. The icebox consists of 6 concentric copper cans, spanning the temperature range from room temperature to 20 mK. The dilution refrigerator is separated from the icebox by the "cold stem", a concentric set of pipes surrounding a bar that connects the mixing chamber of the dilution refrigerator to the lowest temperature can. The cans are supported by flexible Kevlar straps, allowing them to tilt relative to each other as the inner layers of the stem contract during cooldown.

peratures of approximately 300 K, 77 K, 4.2 K, 600 mK, 100 mK, and 20 mK. This nested design minimizes the heat load from conduction and radiation between layers. The cans are thermally isolated from each other by hanging them on flexible straps, which allow freedom of motion as the various layers contract differentially during cool-down. Allowing for this motion was one of the principle challenges in implementing the L-shaped configuration of refrigerator and cryostat, which was demanded by the shortage of vertical space in our tunnel.

Each icebox can is attached to a corresponding layer of the cold stem, which then attaches to a flange on the Oxford refrigerator. The icebox has no cryogens inside and is cooled entirely by conduction along the layers of the cold stem.

It was much simpler to build a cold finger and cryostat made of low-radioactivity materials than it would have been to make a complete low-radioactivity refrigerator. The principle construction material used in the Icebox is Oxygen Free High Conductivity (OFHC) copper, which is a material that combines relatively high thermal conductivity with very low levels of radioactivity. Complex copper parts were made by electron-beam welding, using only copper as the fill material. Unfortunately, copper is a soft metal and not very suitable for the construction of a large vacuum system, since it tends to deform easily and develop leaks. This problem was largely overcome by using many fixtures and extreme care to reduce mechanical stress in the assembly.

A particularly severe cryogenic problem for CDMS was the need to instrument 42 detectors, each of which has 4- 6 independent sensors. To reduce noise, the first amplifying elements for these sensors needed to be inside the cryostat, as close to the detectors as possible. The SQUIDs used to amplify signals from transition-edge sensors operate at 4 K or below, while the FETs used for charge amplifiers and NTD thermistor readout operate optimally at 130 K. Each FET dissipates about 10 mW of power, an enormous amount in the environment of a dilution refrigerator, where cooling power at 20 mK is only $\sim 10\ \mu$W. Clearly, the amplifiers could not have been mounted at the same thermal stage as the detectors. However, noise concerns demanded very short, mechanically rigid (to prevent "microphonics") paths between the FETs and the detectors. This problem was solved by the invention of a modular detector package, called a "tower", which has 4 separate temperature stages, corresponding to the 4 inner stages of the icebox. The tower holds the detectors at $\sim$20 mK and amplifiers at 4 K, with wires running between them that go through heat sinks at the intermediate stages. The wires themselves are made of superconducting niobium, which has a low thermal conductivity, and are surrounded by copper on all sides to reduce pickup. The FETs are thermally isolated from the 4 K stage on an insulating standoff and self-heat to their optimal temperature. The distance between the detectors and FETs is about 10 cm.

Each detector has 25 wires connecting from the front-end amplifier card to the outside world. The icebox has to accommodate ~ 1000 electrical connections from 4K to room temperature. Conventional wiring would have created a large heat load and taken up too much space to be practical. The solution chosen was to run the wires on flexible printed circuits, which are called "striplines". A stripline carries all the signals for one detector on 1 mil copper traces deposited on a thin plastic ribbon. All the traces can be simultaneously heat-sunk by clamping the ribbon between cold copper plates. The stripline connects to the front-end amplifier cards via a removable connector, allowing towers of detectors to be easily exchanged. The strip lines run through the electronics stem and connect to a standard vacuum feedthrough connector outside the radiation shield. This provides a very compact, low-noise interconnect for the large number of signals brought out of the icebox.

The performance of the cryogenic system is monitored with ∼100 thermometers strategically positioned inside the icebox and refrigerator, and with a number of other sensors to measure cryogen levels and pressures in the gas and vacuum lines. A dedicated computer records the measurements at intervals of a few minutes and sounds alarms or broadcasts email messages when a serious anomaly is detected.

The icebox has performed as expected. It requires little maintenance other than replenishment of cryogens (25 L of liquid helium and 50 L of liquid nitrogen a day) and can run cold for months at a time. As is usual with dilution refrigerators, it tends to develop vacuum leaks over time from thermal cycling. The most serious incident of this type, a leak from the ^{3}He circulation system into the inner vacuum can, required a complete dismantling of the joint between the Oxford refrigerator and the cold stem, so that the refrigerator could be jacked up and striped of its outer cans. This took about two months, but proved that the system is not too fragile to fix.

3.5 Radiation shielding and background issues.

Since the cryogenic detectors can not perfectly discriminate between WIMP scattering and scattering from background gamma and beta radiation, we reduce these sources of background events as much as possible with shielding and by careful choice of materials that will be near the detectors. Neutrons from radioactivity and cosmic rays are an even worse problem, since neutron scattering events produce the same ratio of ionization to deposited energy as WIMP scattering events. We have taken particular care to shield the detectors from neutrons.

A detailed study of the shielding design and composition of the background radiation is published in Da Silva's thesis [75]. Here we will summarize the most important considerations and give some new information on recent developments.

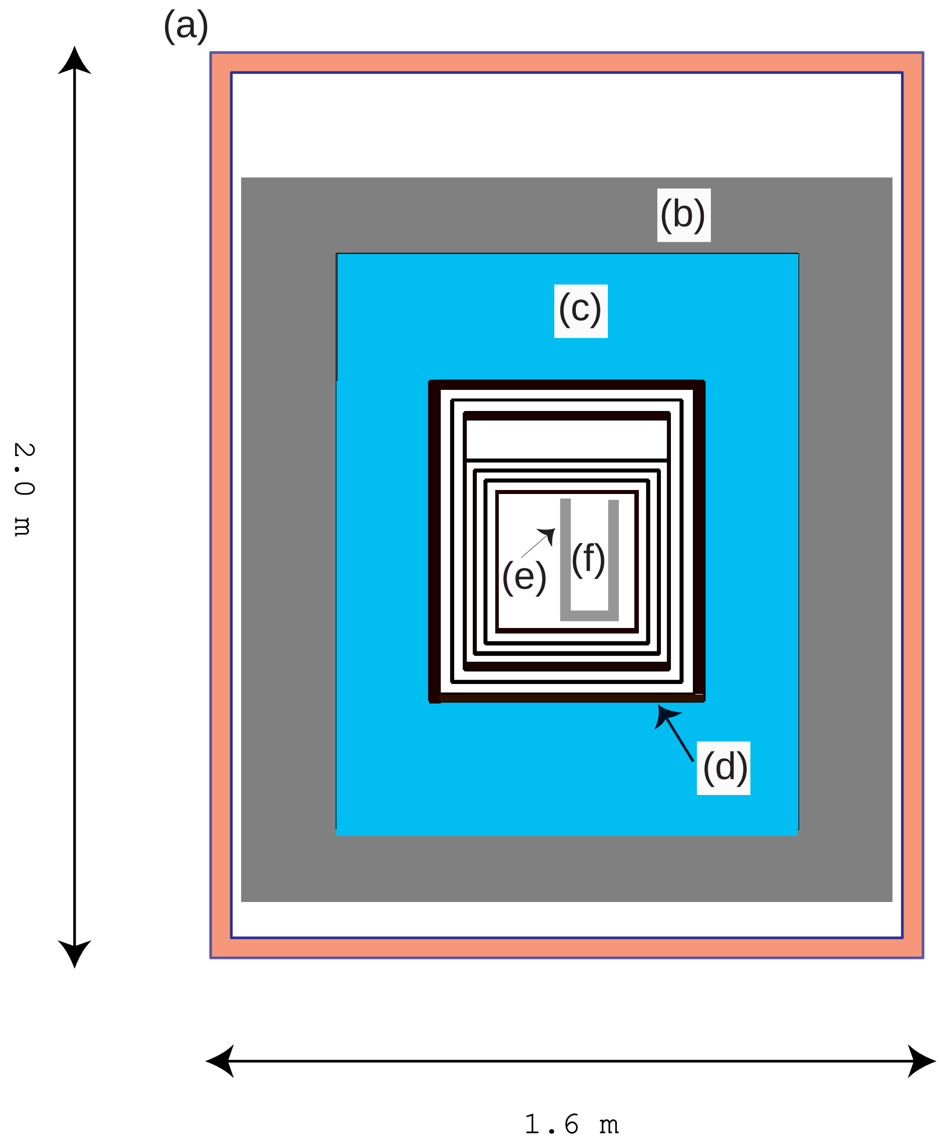
The CDMS I shielding configuration is shown in Figure 3.7. The detectors are shielded from gamma rays by at least 16 cm of lead in all directions, and from neutrons by 25 cm of polyethylene. The shield is located in a tunnel at Stanford University called the Stanford Underground Facility (SUF). This tunnel is 10.6 meters underground, which gives it the equivalent of 17 meters of water (m.w.e.) shielding from cosmic ray muons. This reduces the surface muon flux by 80% to $45/\mathrm{m}^2\mathrm{s}$. Other components of the sea-level cosmic ray flux are reduced to negligible levels by this overburden, except insofar as they are regenerated by muons. In fact, as we will discuss below, the interaction of muons with materials around the detectors is still the dominant source of ambient photons, electrons, and neutrons at this depth.

The Stanford location is a compromise between the need for a low- radiation environment and the convenience of running the detectors near the development labs at Stanford and UC Berkeley. In the future, the experiment will be moved to the Soudan mine in Tower, MN, which is at a depth of 2200 m.w.e. Construction of a clean room in this mine was begun recently, and detector deployment there is scheduled for the year 2000.

3.5.1 Neutron backgrounds.

At ground level, the ambient fast neutron flux (0.4 eV - 10 MeV) is 4×10^{-3} /cm^{2}s [88]. This is reduced to $(8.1\pm0.6)\times10^{-5}$ /cm^{2}s in the SUF tunnel [76]. The neutron flux in the tunnel comes dominantly from production of neutrons in the rock walls

Figure 3.7: The CDMS I Radiation Shield. The components of the shield are (a) a set of plastic scintillator muon veto counters ($\sim$ 5 cm), (b) low activity modern lead ($\sim$ 15 cm) , (c) polyethylene neutron moderator ($\sim$ 25 cm), (d) the OFHC copper cans of the Icebox ($\sim$ 1 cm) , and (e) an inner box of 150 year-old lead with depleted ^{210}Pb (1 cm). The top of this inner lead box is completed by a piece attached to the detector module, which is inserted into position (f). The penetrations of the cold stem and electronics stem through the shield are not shown.



via muon capture and (γ,Xn) reactions.

Because the rate of neutron production increases with the atomic number of the target, the flux inside a lead shield is larger than the ambient flux. The rate of neutron production in lead at SUF has been measured to be (243 $\pm$6)/kg-day [76]. If the neutron moderator in the shield shown in Figure 3.7 were left out, this would lead to a flux in the center of the shield about 20 times higher than the ambient flux in the tunnel and a differential scattering rate in a germanium detector at the center of $\sim 10^3$ dru at 10 keV (In this section, we will make use of the "differential rate unit", or dru, with 1 dru = 1 event/keV-kg-day. Unless otherwise noted, we will quote values at 10 keV.). With the 25 cm of polyethylene in place, we measure a rate of $\sim$10 dru, roughly in agreement with the Monte Carlo calculations of Da Silva [75, 76] and Eichblatt [90]. Eichblatt's calculations show that 70% of the neutron scatters with the polyethylene in place are caused by neutrons produced in the copper of the cryostat rather than in the lead.

The scintillation counters surrounding the shield, which are described in detail in Section 3.6 below, allow us to detect the muons that produce neutrons in the lead and copper. By using these counters in anticoincidence with the cryogenic detectors, we can effectively reduce the background from these neutrons by a factor that is limited only by the inefficiency of the counters, which has been measured to be less than 10^{-4}. Monte Carlo calculations suggest that the remaining neutron background should come from the fraction of high energy neutrons that are produced in the rock outside the shield and manage to penetrate the 16 cm of lead and 25 cm of polyethylene. These should produce a scattering rate of 0.018 dru [89]. As this dissertation is being written, it seems that we may have achieved sufficient sensitivity with the current (Run 19) detector package to measure this small rate of nuclear scattering events. However, results are still inconclusive, with only 1 multiply scattered neutron event detected so far. There is room inside the icebox for the addition of 7 cm of polyethylene, which should reduce the rate by a factor of about 5.

Monte- Carlo studies have been carried out to investigate the neutron background expected at the Soudan site for CDMS II [89]. The goal is to design a shield to reduce the neutron scattering rate to 10^{-4} dru. At Soudan, most of the neutrons will come from (α, n) reactions on ^{18}O nuclei in the rock, where the α originates in the U-Th decay chain elements present at trace levels. These neutrons are relatively easy to shield, because they are low in energy. Work completed so far shows that 50 cm of polyethylene (the maximum that can be accommodated without redesigning the icebox) is enough to reduce the flux from this source to well below the required level. As in CDMS I, the high energy neutrons made by cosmic rays outside the shield are a concern. Preliminary results show that these will cause a rate of 1.4×10^{-4} dru. A veto system will be used to suppress neutrons made in the lead shield, which would otherwise contribute 5×10^{-4} dru.

It's important to note that the neutron background is not indistinguishable from a rate of WIMP scattering events. Although individual neutron events may appear to be caused by WIMPs, since they will have the appropriate ratio of ionization to deposited energy, a substantial fraction of neutrons ($\sim 10\%$, depending on the number of detectors) will scatter in more than one detector of our array, providing a clear signal for the presence of a neutron population. Additional information will be given by the presence of silicon detectors along with the germanium, since silicon has a larger cross section for neutron scattering, and the mean recoil energy is higher. Therefore, neutrons can not "fake" a WIMP signal in these detectors, given a large enough data sample, though they can limit sensitivity.

3.5.2 Gamma ray backgrounds.

Spectra accumulated with an unshielded germanium detector in the SUF tunnel have a differential rate due to scattering of ambient gamma rays of 2 $\times 10^4$ dru in the low energy (<100 keV) region. Since we want to measure WIMP rates that are at least 6 orders of magnitude lower than this, and the cryogenic detectors are imperfect at discriminating between gamma rays and nuclear scattering, it's clearly important to reduce the gamma ray flux as much as possible with conventional shielding.

The gamma ray shield for CDMS I, which is shown in Figure 3.7, consists of 15 cm of lead bricks outside the polyethylene and 1 cm additional lead inside the cryostat. This reduces the external gamma ray flux by at least a factor of 10^4. The exact attenuation factor is hard to measure, because at this point the flux inside the shield is dominated by internal sources. The thickness was chosen on the basis of test measurements that showed no improvement in the background rate for greater thicknesses [75]. On two sides of the shield, the thickness is increased to 20 cm of lead for mechanical reasons.

Although the low background community has experimented with alternative gamma ray shielding materials, lead still seems to be the best one [88]. Lead is favored because of its high atomic number, good mechanical properties, and low cost. In addition, the standard methods used by industry to purify lead naturally result in a product that is relatively free of radioactive impurities. There are no important long-lived activities in lead that can be induced by cosmic ray or neutron exposure. There is just one flaw - lead tends to be contaminated at the time of its manufacture with the isotope ^{210}Pb, which has a 22-year half life and can not be chemically removed. The contamination is caused by the presence of trace amounts of ^{238}U, a parent of ^{210}Pb, in the lead-bearing ore. Typically, lead smelted in modern times has a ^{210}Pb content of ~100 Bq/kg (Note: 1 Bq= 1 decay/ s). At a cost a few times higher than ordinary lead, it's possible to get lead from special mines that produce material with an order of magnitude less contamination. To get to lower activity levels than this, it's necessary to obtain lead that has either been isotopically purified or has been "stored" (e.g. in a sunken ship) for many multiples of the ^{210}Pb half life. This lead can have undetectably low ^{210}Pb content ($< 10^{-3}$ Bq/kg)[3] .

The ^{210}Pb decay itself emits mostly soft β particles (Q =17 keV), which do not propagate far enough in lead to cause a problem, but the daughter ^{210}Bi has an energetic β (Q = 1.16 MeV) that produces high energy gamma rays via bremsstrahlung. Da Silva measured the gamma ray flux inside a box made with lead that had an activity of 33 Bq/Kg and found that this produced a background of 8 dru [75].

Since lead with low ^{210}Pb content is expensive, we built the CDMS I shield with layers of progressively higher quality lead towards the inside. The outer 10 cm is commercial lead, with an activity of 250 Bq/kg, left over from previous physics experiments at Stanford. The 5 cm of lead directly inside this is made from 33 Bq/kg lead that comes from a low activity mine. Finally, the inner 1 cm, which is inside the icebox, is lead from a ship that sank off the coast of France an estimated 150 years ago, which has an activity of 10^{-3} Bq/kg. Monte Carlo calculations show that the radiation from ^{210}Bi bremsstrahlung is reduced in intensity by about an

[3] Recently, a small industry has emerged to supply this material to the microelectronics industry. Demand is driven by the possibility that alpha particles coming from lead solder deposited directly on silicon will cause chips to malfunction. This has driven up the cost to levels that are prohibitive for large radiation shields.

order of magnitude after passing through 1 cm of lead, so the inner layer should bring the background down below 1 dru.

The measured rate of events in the cryogenic detectors that are not coincident with muons inside the CDMS I shield is 1.3 $\times 10^3$/(kg-day) (See Chapter 3). The differential rate at low energy, ~ 2 dru is typical of what other groups have achieved in similar counting setups. The remaining activity is almost certainly due to sources inside the 15 cm lead walls, but the location of the isotopes responsible is unknown.

Great care was taken to screen all the materials that make up the cryostat and detector mounting hardware. This screening was done by our collaborators at Lawrence Berkeley National Laboratory with conventional Ge detectors in low background shields similar to the one used for the main experiment. The isotopes of greatest concern, because they are the most common, are the ones in the uranium and thorium decay chains, and ^{40}K. Radioisotope contamination of construction materials by these isotopes is hard to avoid, since they are ubiquitously present in the natural environment. These isotopes are often present in manufactured materials at levels comparable to the levels in "natural" materials, since no special care is taken by industry to eliminate them. On the other hand, some industrial purification methods have very general separatory powers, and their use naturally results in a final product that is both chemically pure and free of radioisotope contamination.

An important example of a low-activity industrial product is the Oxygen Free High Conductivity (OFHC) copper that the icebox, tower, and and detector holders are made from. This is a high grade of copper which is refined by electrolysis to improve its electrical conductivity. The electrolysis produces copper that has chemical purity greater than 99.99% and, immediately after production is probably free of any measurable radioactive impurities. Unfortunately, there are a number of long-lived radioisotopes that can be made in copper by cosmic ray neutrons. The most important of these is ^{60}Co, which has a half life of 5.2 years and an equilibrium activity in copper at the Earth's surface of 1.7 $\times 10^{-4}$Bq/kg [75]. Since there are 872 kg of copper in the icebox, there may be a rate of $\sim 10^4$/day decays from this source in the shield.

Besides copper, the other material present inside the shield in large quantities (about 1000 kg) is polyethylene. Plastics usually have extremely low contamination levels from the U-Th chain in bulk material and have no significant cosmogenic activity. The main worry is that the surface of the plastic may be contaminated with dust or other foreign matter may be included in the plastic pieces as they are formed into finished products.

Another important source of gamma rays inside the shield is the capture of thermal neutrons on materials near the detectors, and on the detectors themselves. These neutron captures produce prompt gamma rays and the addition of the extra neutron often destabilizes the nucleus, leading to delayed activity (for example, the β emitters ^{64}Cu and ^{65}Cu are made by capture on copper). About 100 neutrons/(kg- day) are made in copper at SUF [75], leading to a total neutron production rate of $\sim 10^5$/day inside the shield.

The relative contribution of the sources we mention above to our measured background rate has not yet been carefully calculated.

3.6 The Muon veto.

The radiation shield is surrounded with a muon detector, which detects the muons that make neutrons in the lead and copper. The goal is to detect at least 99% of them, so that the neutron rate can be brought down from $\sim$1 to $\sim 10^{-2}$ dru by using the muon detector in anticoincidence (i.e. as a "veto") with the cryogenic detectors. In addition to having high efficiency for muons, it is important to have only a small chance of triggering on gamma rays or tube noise, so that the veto trigger rate will not lead to a large dead time for the dark matter search.

The muon veto is made of 13 pieces of 5-cm-thick plastic scintillator, which surround the lead shield on all sides. The counters on the side closest to the dilution refrigerator have clam shell type penetrations for the cold stem and electronics stem, and the counters under the lead have gaps for steel support beams. Overall, coverage on the top and sides is >99% and coverage on the bottom is about 70%.

The scintillator used is a standard plastic mixture known as NE110 (from Nuclear Enterprises), which was chosen because it has a long attenuation length, over 3 m in bulk material. This property is important for insuring uniformity of light collection, which directly effects energy resolution and the ability to discriminate between muons and ambient gamma rays. The NE110 also has a reasonably high light output (60% of anthracene) and was available in the large pieces we wanted. The emission of the scintillation light is peaked at 430 nm.

The light from the scintillator is detected with 26 2-in.-diameter phototubes (a combination of Burle models 8850 and 8875). Since there is limited space in the SUF tunnel, the light is collected into the tubes with waveshifter bars, which can be much more compact than conventional light guides. The waveshifter bars run

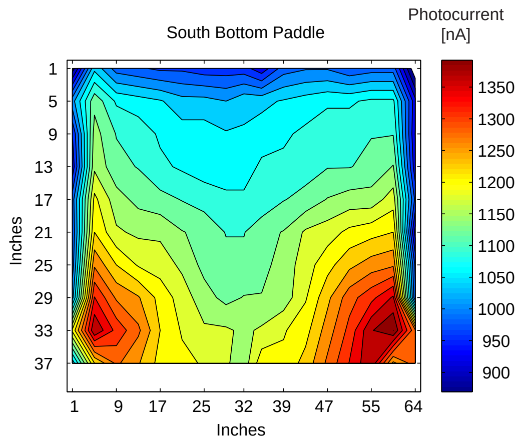


Figure 3.8: Spatial variation of light collection efficiency in a typical counter (the "South Bottom" counter). This counter has two phototubes mounted on waveshifter bars running along the bottom and sides, with the bottom bar shared by both tubes. The contour plot shows the summed photocurrent induced by a collimated source of gamma rays from a ^{241}Am source that was scanned along the surface. This current has maxima at the corners closest to the tubes and a broad minimum near the top center. The ratio of strongest to weakest response is 1.5. The minima along the side edges of the counter are probably a measurement artifact caused by the finite width of the beam.

along the edges of the scintillator panels, with an air gap between the scintillator and waveshifter defined by thin nylon shims. The waveshifters for each paddle run into either 2 or 4 glued-on phototubes, which have their signals summed together to increase the uniformity of response. The scintillator panel - waveshifter assemblies are wrapped loosely with aluminized Mylar, a good reflector, and with a light barrier of black construction paper and tape outside the Mylar.

The principle of operation of the veto is extremely simple. Muons travel at least 5 cm through the scintillator and deposit at least 10 MeV by ionization loss (roughly 2 MeV/cm for a minimum-ionizing particle), producing $\sim 10^5$ scintillation photons. In an ideal slab geometry, with perfect surfaces and no attenuation, 11%

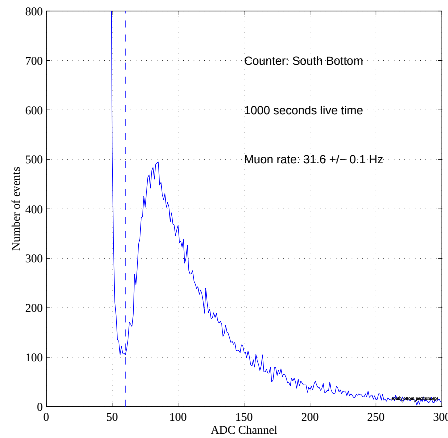


Figure 3.9: A spectrum taken with the counter shown in Figure 3.8. The peak at high amplitudes is caused by muons passing through the counter, which deposit > 10 MeV by ionization. There are a large number of small amplitude events below channel 50 caused by ambient gamma rays and phototube noise.

of the light would escape from each of the 6 sides of the slab, and 34% would be trapped inside the slab by total internal reflection (for index of refraction $n = 1.58$) [87]. The light emerging from the sides of the slab crosses an air gap and is absorbed in the waveshifter material (Bicron 428A), which absorbs light near 420 nm and re-emits it at 444 nm. Since the re-emission is isotropic, a fraction of the re-emitted light will be trapped by total reflection inside the bar and propagate to the ends where it may be detected by a phototube. The phototubes have Be-O cathodes, with a quantum efficiency of about 25% for the frequencies of light emitted by the waveshifter.

In principle, arrangements of this sort can have efficiencies of up to about 2% for conversion of photons into photoelectrons per instrumented scintillator face [87]. In practice, it is difficult to achieve efficiencies this high due to non-ideal behavior, such as attenuation in the scintillator and waveshifter bars, warping of the panels (which are very heavy) and imperfections in the scintillator surfaces caused by abrasion. In some of our counters, we get only ~ 20 photoelectrons per muon, implying an efficiency of $< 10^{-3}$.

Environmental gamma rays from radioactivity can have energies up to only 2.6 MeV (from ^{208}Tl) and yield at most 1/3 the number of photoelectrons that muons do. In fact, the bulk of the ambient gamma ray flux is below 500 keV. Muon hits can then be distinguished from gamma ray events on the basis of pulse height. Variations in the efficiency of light collection across the counter must be kept small, so that gamma rays interacting in high efficiency regions will not be confused with muons interacting in low efficiency regions. At the same time, the absolute light collection efficiency must be kept high, so that the statistical fluctuations in the numbers of photoelectrons for the two types of events do not cause their pulse height distributions to merge. At the single photoelectron level, there is a large background of noise pulses in the phototube comparable to the number of gamma ray events. The rate of such "dark events" (unfortunately, not the dark events we are looking for!) varies quite a bit between tubes and also seems to be effected strongly by temperature. The origin of these pulses is uncertain; they may be caused by thermionic emission in the tube, or by dielectric breakdown in the phototube glass or base components.

The phototube gains can be adjusted by changing the high voltage, which typically is between -1400 V and -2000 V for gains of $\sim 10^6$. The gains of tubes attached to the same counter must be balanced relative to each other so that the summed signal height will depend as little as possible on the location of the scintillation. Balance is achieved by adjusting the high voltage to equalize the average amplitude of muon pulses seen in individual tubes.

Recently, we have measured the relative light collection efficiency of the counters with a spatial resolution of ~ 2 cm by scanning them with a strong, collimated ^{241}Am source. This method works very well, because the 60 keV photons from ^{241}Am easily penetrate the < 1 mm thick Mylar and paper wrapping, but are fully absorbed locally in the scintillator. The 60 keV photons each have only a small chance to make a photoelectron, so the response of the tube must be measured by monitoring the total photocurrent induced by the source. We find that the maximum ratio of highest to lowest light collection is about 3 in the worst counters. The results are shown in Figure 3.8 for a typical counter, which has a ratio of 1.5. The quality of separation between muon events and events from gamma rays can be judged from Figure 3.9, which shows the pulse height distribution in the counter of Figure 3.8.

The veto electronics is quite simple. The outputs of all the tubes from a counter are summed together with a fast linear amplifier and the sum is fed into a discriminator circuit, which has an adjustable threshold for firing. The output of the discriminator is stretched into a 1-μs-long logic pulse, which is fed into a dual port memory that records the output of all discriminators at 1 μs intervals. More details are given in Section 3.7.

The goal of the veto is to detect as high a fraction of muons entering the shield as possible, while detecting as small a fraction of incident gamma rays as possible. This latter requirement comes from the fact that each trigger in the muon veto results in some amount of dead time for the experiment, since events which accidentally occur in coincidence with these triggers must be thrown out. (Here we use the word “trigger” to refer only to the firing of the discriminators, rather than to the higher level triggers that result in data acquisition from the cryogenic detectors.) The fraction of dead time is the product of the veto trigger rate and the length of time that must be waited after each trigger to avoid neutrons. The rate of muons entering the shield is about 350 Hz, but we often set the phototube thresholds low enough to produce a trigger rate that is an order of magnitude more than this in order to catch the few muons ($< 1\%$) in the low amplitude tail. We find that at least 99% of neutron scattering events in the cryogenic detectors occur in the first 20 μsec after a muon hit.

Some detailed information on the efficiency and dead time induced by operation of the veto can be found in Appendix B. The efficiency of the veto counters for muon detection can be monitored by identifying muon interactions in the cryogenic detectors. These interactions are the primary cause of events with energy deposition above 10 MeV, and we find that >99.99% of such high energy events are coincident with muons in the veto counters. This indicates that the veto inefficiency is less than 10^{-4} , enough to suppress the expected muon-coincident neutron flux below the level of the flux we expect from neutrons that originate outside the shield and are able to penetrate the thick moderator. In typical operation (counting rate $\sim$10 kHz and veto window of 20 μsec), the veto causes a dead time of 20%.

3.7 Data acquisition.

A block diagram of the data acquisition system is shown in Figure 3.10. The signals which must be acquired include long phonon and charge waveforms from cryogenic detectors (1000 samples for charge channels, 2000 samples for BLIP phonon channels, and 10000 samples for FLIP phonon channels), and trigger bits from the cryogenic detectors and the 13 veto counters. The main challenge is the need to

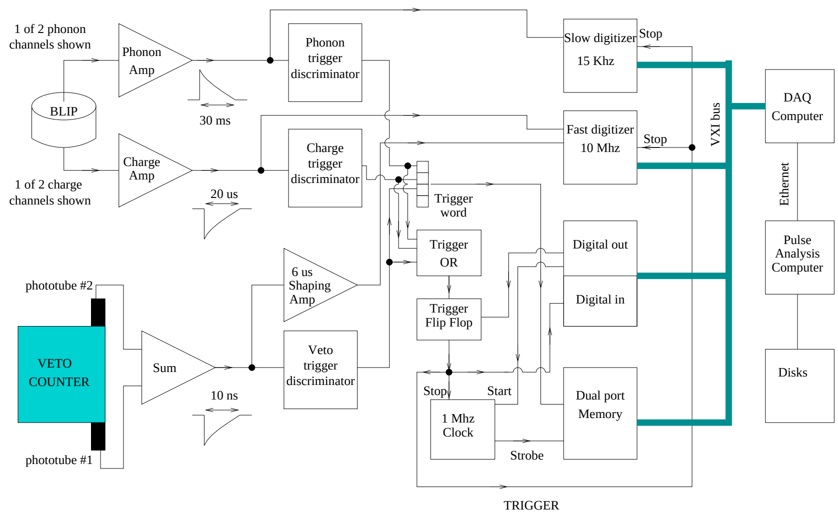


Figure 3.10: Data acquisition block diagram for Run 18. Only one cryogenic detector and one veto counter are shown. The trigger logic has been somewhat simplified.

deal with three types of signals which occur on quite disparate time scales. The cryogenic detectors trigger at a maximum rate of a few Hz, with BLIP phonon pulses rising in a few ms and falling (to $1/e$) in $\simeq$40 ms. FLIP phonon pulses and charge pulses from both detectors rise in a few μs and fall in $\simeq$ 10 μs. At the same time, veto pulses lasting only $\sim$ 10 ns are occurring at rates up to 10 kHz.

Since the veto and cryogenic detector trigger rates are so different, and events which have activity only in veto counters are uninteresting, we don't read out events which trigger just the veto. Instead, the veto hit patterns are stored in a digital buffer that overwrites itself every 32 ms if no cryogenic detector trigger occurs. In some running periods, we have also digitized the analog sum of all the phototube signals after slowing them down to speeds compatible with the 10 MHz digitizers that are used for the cryogenic detectors. In the future, there are plans to begin digitizing the veto pulse heights on a counter-by-counter basis.

The normal data acquisition sequence goes as follows: (1) The digitizers are normally free running and record all waveforms from BLIPs and FLIPs in a circular memory. (2) When a BLIP phonon or charge channel, or a FLIP phonon channel goes over a preset trigger threshold, a stop signal is sent to the waveform digitizers and digital history buffer. At the same time, a bit is set on the DAQ computer's input indicating that it is time to read out an event, and further triggers are inhibited. (3) The digitizers and history buffer continue running for a preset time after receiving the stop signal in order to record the decay of the pulses (and the rise of the BLIP phonon pulse). (4) The computer waits for long enough to ensure that the post-trigger acquisition is complete and then reads out the waveforms and digital buffer, storing all the information on disk. This usually takes a few hundred ms (e.g. 350 ms in Run 18). (5) The computer restarts the digitizers and clears the trigger condition, getting ready for another event.

The strategy of digitizing full waveforms from detectors rather than just amplitudes is a fairly new one in high energy physics that is just now becoming feasible for moderate ($\sim$ 100) numbers of high bandwidth electronics channels. This is an expensive and technologically difficult approach because it requires high speed digitizers connected to deep local memories for each channel. The large amount of data collected (e.g. $\sim$ 50 KByte/event for CDMS I with a $\sim$ 1 Hz trigger rate) must be quickly moved out of the local digitizer memories onto disks after each event and passed to off-line computers for the signal processing that results in pulse amplitude measurements. Heavy demands are placed on the networks that connect the digitizer memories to the data acquisition control computer and then to the disk farm that holds the events while they are being analyzed. The computing and mass-storage resources required to deal with this data flow are also significant.

The reward for accepting this increased complexity in the data acquisition and processing chain is the availability of substantially more information for each event. This is important, given that we are interested in selecting the few events in 10^6 that are WIMP scattering candidates. These few events involve energy depositions in just one detector at a time and the pulse amplitudes are typically close to the steady-state noise. It's common for mechanical vibrations to cause noise transients with pulse height equivalent to 10 keV or more and these could be confused with real detector pulses if only the amplitude was measured. In this situation, there is a high premium on the extra information that comes with waveform recording, which allows the use of digital filters optimized for the noise actually measured during event acquisition. As a final check, the shapes of candidate WIMP pulses can be examined by eye for any irregularities.

The pulse amplitudes are fit by a program which runs on a dedicated computer that is constantly checking to see if there are new events to analyze. Some details on the pulse processing are given in Appendix A.

The data acquisition and analysis system is constantly evolving, because it has to respond to the stress of adding more detectors. This increases both the trigger rate and the event size, so the data volume tends to increase as the square of the number of detectors. This scaling will lead to big challenges in the next few years as this number increases from 6 (in Run 19) to 42 for CDMS II.

Chapter 4

Analysis of CDMS Run 18 BLIP data.

In this chapter, we present an analysis of data taken with two BLIPs at the Stanford Undergound Facility during "Run 18", which lasted from January to June of 1998. The data collected includes several periods of "background counting", when we tried to minimize the count rate in the detectors in order to be sensitive to a small rate from WIMP scattering, and several periods of exposure to neutron and gamma ray calibration sources. Both types of data are useful to have for finding WIMPs, because we need to know what nuclear scattering and electron scattering will look like in the background data set in order to best discriminate between them.

Our main analysis goal is to set limits on the masses and cross sections of WIMPs that could populate our galactic halo. We will also make observations about the performance of the BLIP detectors and the nature of the background sources at SUF that will hopefully be useful for others as the experiment continues to evolve.

4.1 Operation of the detectors.

The analysis shown here is based on data from the two 165 g germanium BLIP detectors known as BLIP 1 and BLIP 2. The design of these detectors has been described in Chapter 3, as has the cryogenic apparatus, radiation shielding, and data acquisition system that we used. In addition to the two BLIPs, the Run 18 experimental package included a silicon FLIP detector that acquired data simultaneously. The analysis of the FLIP data is described in Ref. [77]. For the analysis presented here, the only use made of the FLIP detector signals was to reject events with energy deposition in more than one detector (either two BLIPs or a BLIP and a FLIP) as candidate WIMP scattering events.

Figure 4.1 shows the assembled detector package that was mounted in the icebox. The end of the tower which holds the detectors is inserted into a lead box on the bottom of the base temperature can, which was cooled to 25 mK. Note that BLIP 2 and FLIP are somewhat better shielded from the external environment than BLIP 1 because they are less exposed to the opening at the top of the inner lead shielding box that surrounds the tower. BLIP 2 also has the advantage that it is surrounded on both sides by other detectors which can be used in anticoincidence to suppress multiple scatters. As discussed below, we find somewhat higher counting rates for BLIP 1 than BLIP 2 in both calibration and background data.

The detectors are held individually in copper boxes that include circuit cards with the cold electronic components needed for the biasing circuits. Each copper box also has an LED, which was turned on briefly to neutralize the crystals at the beginning of the run (See Chapter 3, Section 4). The copper boxes have a minimum thickness of 1/8 in. and block all direct lines of site to the external world. The only

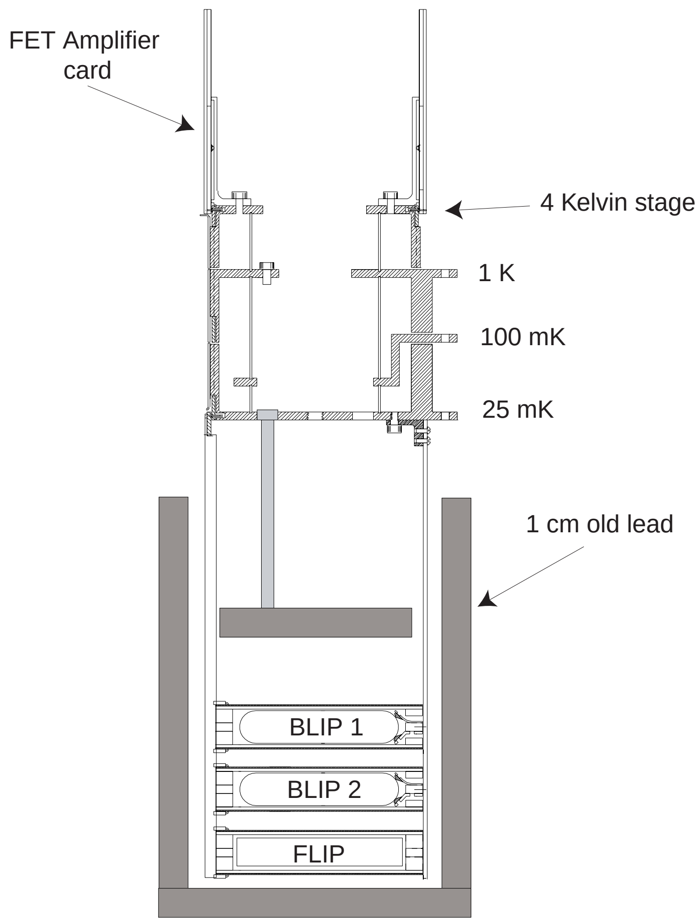


Figure 4.1: Run 18 experimental package, showing the tower with 3 mounted detectors. The inner layer of lead shielding is also shown. This shield is in two pieces, one of which (5 sides of a box) sits on the bottom of the base temperature icebox can. The other piece is mounted in the tower directly above BLIP 1. Before inserting the tower, the bottom section is wrapped tightly with copper-clad tape as a final precaution against infrared radiation leaks from the 4 K stage.

materials directly exposed to the germanium surfaces are acid-etched copper, the plastic LED casing, and the small quantity of epoxy-Kapton composite that is used for thermal isolation. All non-penetrating radiation (such as low energy X-rays and beta particles) that is seen in the detector must come from these materials.

The charge bias was 2 V/(1.2 cm) for BLIP 1 for all of Run 18 and 1 or 2 V/(1.2 cm) for BLIP 2 at various times during the run. Because a stronger electric field is believed to reduce the charge deficit for interactions that occur near the surface of the detector, we wanted to make this bias as high as possible. However, 2 V turned out to be the practical limit that could be achieved, because higher biases eventually cause surges of current through the crystal and very high noise. This "breakdown" behavior is believed to be caused by injection of charge carriers into the semiconductor at inhomogeneities in the contacts, where the electric field is strong.

BLIP 2 had an additional constraint on its maximum bias voltage for the early part of the run. The inner charge channel decoupling capacitor, which normally prevents bias currents from flowing into the charge amplifier feedback network, developed a highly resistive short across its terminals after cooldown. We were able to find a solution to this problem that worked initially only for biases up to 1 V. Later in the run, further modifications allowed biasing both detectors at 2 V.

As discussed in Chapter 3, the BLIP detectors measure both ionization and heat. Heat comes from both the recoil energy deposited by interacting particles and from the drifting of the electron- hole pairs in the electric field applied across the detector. The recoil energy can be calculated by subtracting the charge drift contribution from the amplitude of the thermal pulses. The procedure for doing this is described in Appendix A. The most useful quantities to work with in the event analysis are the recoil energy, E_R, and the ratio of collected charge to recoil energy, Y, which is called the "charge yield". The charge yield is normalized to give electron recoil events an average of $Y = 1$, so Y can be thought of as the ratio of "electron-equivalent-energy" to the actual amount of energy that is deposited. Note that we always calculate Y by dividing by the recoil energy rather than the total amount of heat (including a contribution from charge drift), which is what some other CDMS authors have done.

For each event that occurred in any detector, we recorded the full wave form from each sensor on all detectors. The amplitudes of the pulses are found by a fitting procedure described in Appendix A. The output of the fitting program includes a number of diagnostic parameters, such as the χ^2 of the fit, that are useful for rejecting noise and pileup events. The cuts we make on these parameters are described in Appendix B, which also describes how we calculate the fraction of "normal" events that pass the cuts.

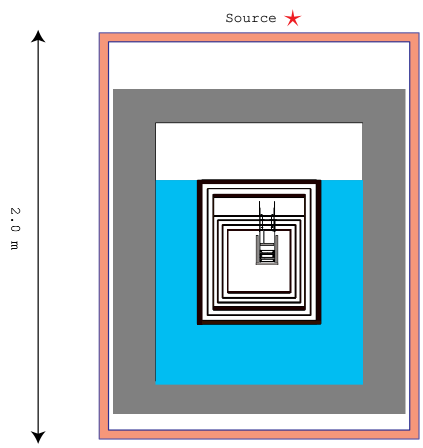
The energy threshold for triggering on an event is determined by discriminator circuits which have the phonon and summed charge pulses as input. Since these discriminators use only very simple filtering, and there tends to be a high rate of large-amplitude noise events from microphonic or "thermophonic" pickup, the trigger thresholds must be set at ~ 10 times the r.m.s. noise width. This noise width is measured to be 157 eV (246 eV) for BLIP 1 (BLIP 2) phonon channels and 432 eV (541 eV) for the charge channels. The thresholds were adjusted occasionally throughout the run to compensate for changes in the noise, which would otherwise have caused intolerably high trigger rates. For the data presented here, the maximum phonon threshold was below 4 keV of recoil energy and the maximum charge threshold was below 8 keV of charge energy. It is the lower of these two thresholds which determines whether an event is acquired, and interesting events always have a phonon signal at least as large as the charge signal, so we will always use 4 keV as the analysis threshold.

Although the charge trigger is set at 7 keV, charge traces are always acquired simultaneously with the phonon traces when a phonon trigger occurs. This allows us to search the charge time trace for small pulses that were below the trigger threshold and use them to calculate the correct ratio of ionization to phonon energy. Since the charge pulses have a width in time of $\sim$10 μs and the phonon signal has a jitter in the trigger time of a few ms, the search must be made over a time interval that is long compared to the charge pulse length. For charge amplitudes below about 1.5 keV, there is a significant chance that the real charge pulse will not be found in the noise. Instead, the algorithm finds a "noise pulse", which tends to be larger than the real one. As noted below, this causes distortions in the charge yield distribution for recoil events below about 10 keV.

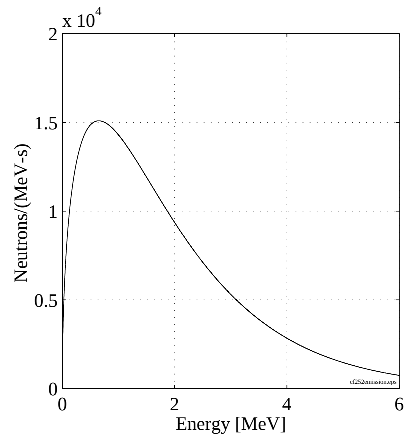
4.2 Neutron source calibration.

It's important to understand the response of our detectors to nuclear recoils, so that we can select a known fraction of candidate WIMP-nucleus scattering events in the final data. The nuclear recoil response was measured by exposing the detectors to fast neutrons from a ^{252}Cf fission neutron source.

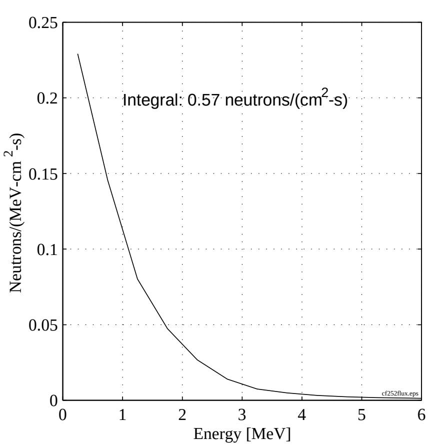
Figure 4.2a shows the location of the source and the configuration of the shield during the measurement. The source was placed on top of the Top South veto counter, 100 cm from the detectors. To increase neutron flux, the polyethylene neutron moderator normally present in the shield was removed down to the level



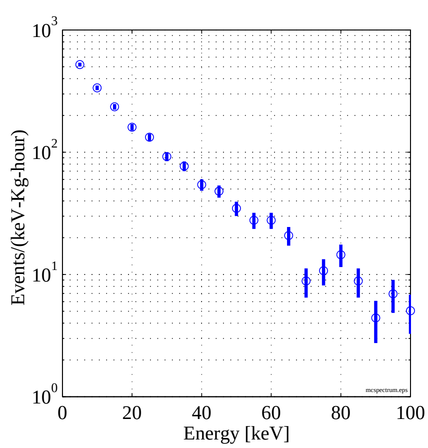
(a) Source location. The source is 28 cm south of the icebox center and 100 cm above the detectors.



(b) Spectrum of neutrons leaving the source.



(c) The neutron flux entering the top lid of the icebox.



(d) The rate of neutron scattering events in BLIP 1 and BLIP 2.

Figure 4.2: Neutron calibration with a ^{252}Cf source. The spectrum in (b) comes from Ref. [96]. Figures (c) and (d) are based on Perera's Monte Carlo calculations [97].

of the cryostat lid. The neutrons arriving at the detectors are still moderated somewhat by the their passage through the 4.5 cm thick top plastic scintillator panel and by reflections off the remaining polyethylene below the level of the lid. In addition, the total number of neutrons is reduced by reflection off the 15 cm thickness of lead at the top of the shield and the smaller amounts of lead and copper inside the cryostat.

The ^{252}Cf isotope decays 3% of the time by spontaneous fission, producing the neutron spectrum shown in Figure 4.2b. The neutrons leaving the source have a mean energy of 2.14 MeV, with 95% below 5 MeV. As shown in Figure 4.2c, which is based on a GEANT Monte- Carlo simulation that incorporates the detailed shielding geometry [97], the spectrum of neutrons arriving inside the cryostat is significantly lower in energy and also quite low in intensity. The scattering is nearly isotropic, and the mean energy loss is 3% of the energy of the incoming neutron. As shown in Figure 4.2d, the scattering produces a featureless, exponentially falling recoil spectrum, with a mean energy of 20 keV and a integrated rate of 0.5 Hz/detector.

Data from a 3.8 hour exposure to the source is shown in Figures 4.3 and 4.4. Because of the limited speed of the data acquisition system, the effective live time was only 1.2 hours. The trigger rate during this period was 3 Hz, caused mostly by gamma rays that were associated with the source. These are probably from thermal neutron capture and the inelastic scattering of fast neutrons on materials inside the shield, since direct rays from the source are heavily shielded by lead.

The gamma rays and neutrons produce two distinct populations of events, which are separated by a gap in charge yield. Gamma rays appear at $Y \simeq 1$ and neutrons at $Y \simeq 0.3$. We will separate these populations by making a cut at $Y = 0.8$.

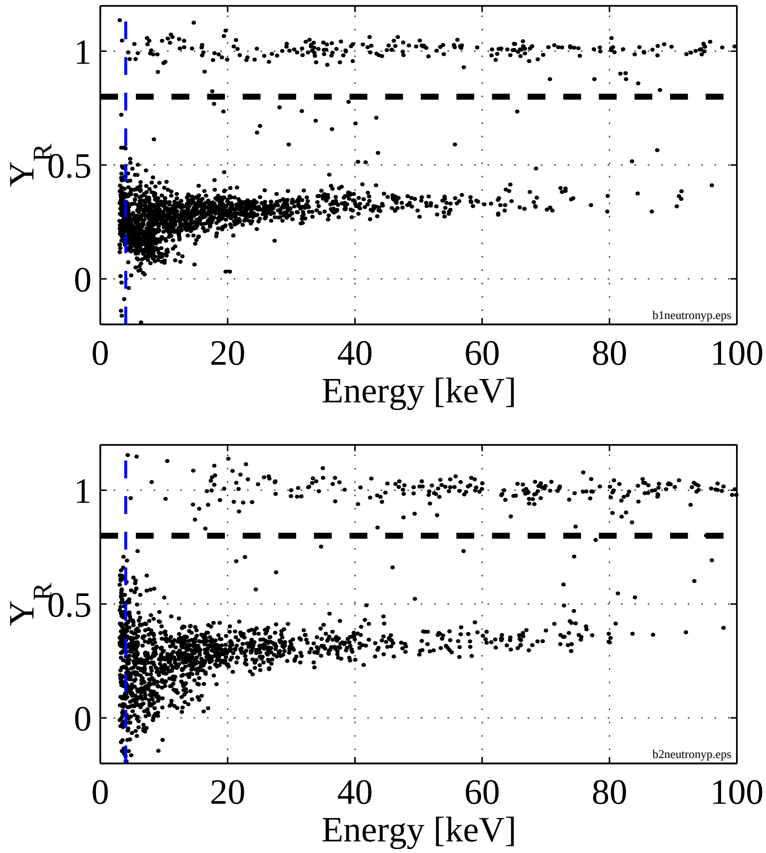
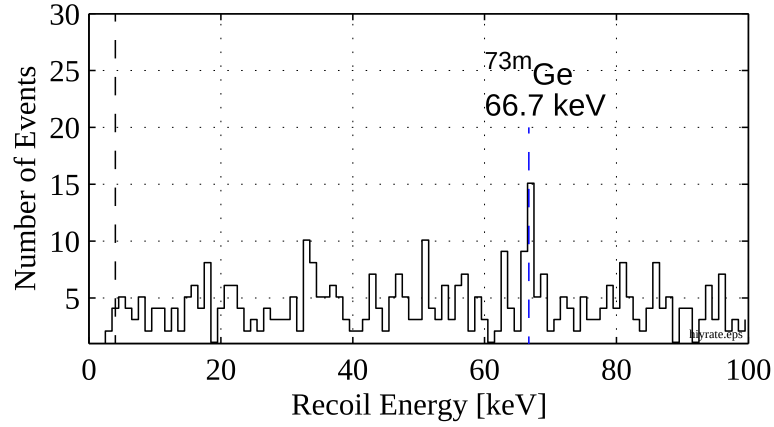
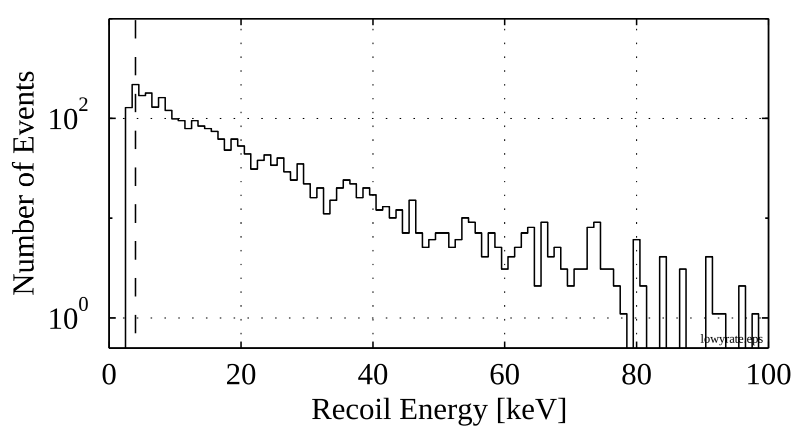


Figure 4.3: Scatter plot of Y vs. recoil energy for BLIP 1 (top) and BLIP 2 (bottom). The horizontal dotted line shows a cut at $Y = 0.8$, which we use to separate the neutron scattering events clustered at $Y \simeq 0.3$ from the gamma rays at $Y \simeq 1$. The vertical line at 4 keV indicates the threshold at which the phonon trigger is fully efficient.



(a) High charge yield events ($Y > 0.8$).



(b) Low charge yield events ($Y < 0.8$).

Figure 4.4: Energy spectra for the two populations of events separated by the dotted line in Figure 4.3. There is a line in the high charge yield data caused by relaxation of the state ^{73m}Ge, which can be excited by thermal neutron capture.

Energy spectra of events falling on the high- and low- Y sides of the cut are shown separately in Figures 4.4a and 4.4b. There are a total of 2910 events between 4 and 100 keV that pass that data quality cuts described in Appendix B. Of these, 86% have $Y < 0.8$. Above 100 keV, the fraction of events with $Y < 0.8$ falls to almost zero, because there are few neutrons with high enough energy to produce such large recoils.

Our main interest in the neutron data is to use it to learn precisely what the distribution of nuclear recoil events in Y is as a function of energy. As a first step in this direction, Figure 4.5 shows histograms of Y for three broad energy regions, 4-15 keV, 15-30 keV, and 30-80 keV. Neutrons appear in each histogram as a peak centered somewhere in the region $0.20 < Y < 0.35$, with the location depending on the energy. We fit this peak with a Gaussian curve to find the center $\bar{Y}$ and standard deviation σ_Y in each bin. These curves fit the data well in the 15-30 keV and 30-80 keV bins. In the lowest energy (4-15 keV) bin, there are some distortions due to failure of the charge search alogrithm. The Y distribution in this bin can not be relied on to remain stable in time, because it is highly sensitive to changes in the noise environment, and the noise depends on some factors that are not well understood or controlled.

As the energy increases, the width of the charge yield distribution narrows and the mean value increases. To investigate this in detail, we calculated the mean value of the charge energy and recoil energy for neutrons in 22 bins from 4 to 80 keV. As shown in Figure 4.6a, there is a linear relationship for energies between 10 keV and 80 keV. Below 10 keV, the mean charge yield begins to rise because of the charge search errors mentioned above.

According to the theory of Lindhard [95], the relationship between charge en-

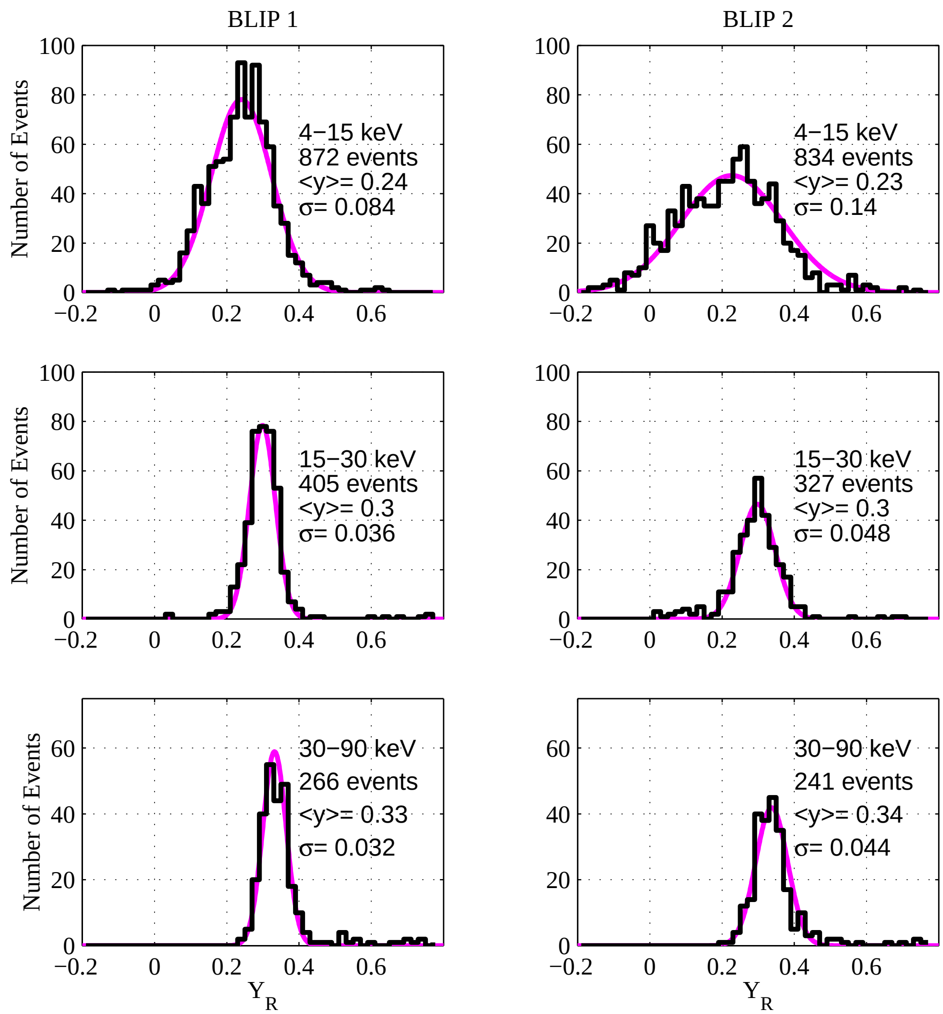
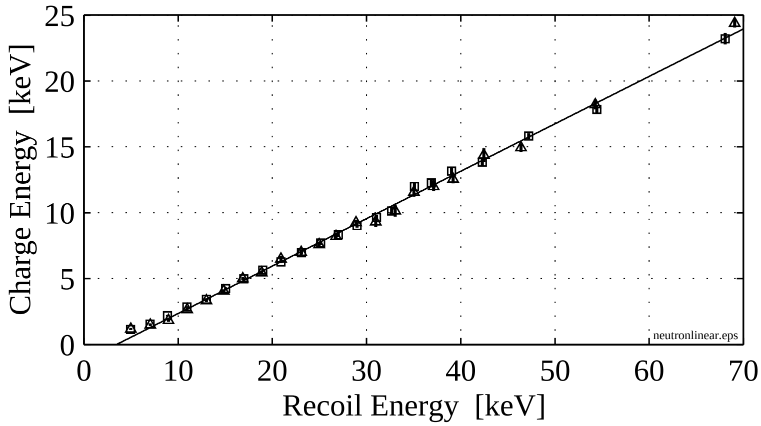
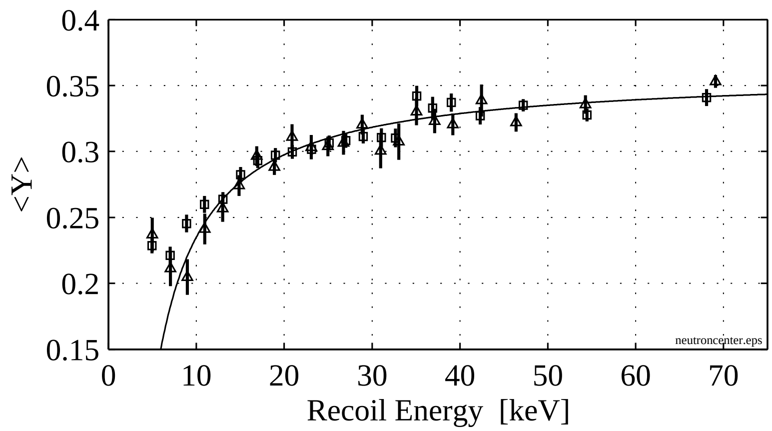


Figure 4.5: Charge yield distribution of neutron calibration data in three energy ranges: 4-15 keV (top), 15-30 keV (middle), and 30-90 keV (bottom). The curves show the result of a Gaussian fit to the data, with the best fit parameters shown to the right of each plot.



(a) The mean charge energy (E_q) as a function of mean recoil energy (E_R), shown with a best-fit line. The equation of the line is $E_q = a + bE_R$, with $a = -1.25$ and $b = 0.36$.



(b) The mean charge yield ($\bar{Y}$) as a function of the mean recoil energy, shown with the fit $\bar{Y} = a/E_R + b$, using the same constants as in Figure (a).

Figure 4.6: Variation of charge yield with recoil energy in 22 energy bins from 4 to 80 keV. BLIP 1 data is shown with triangle markers and BLIP 2 data with squares.

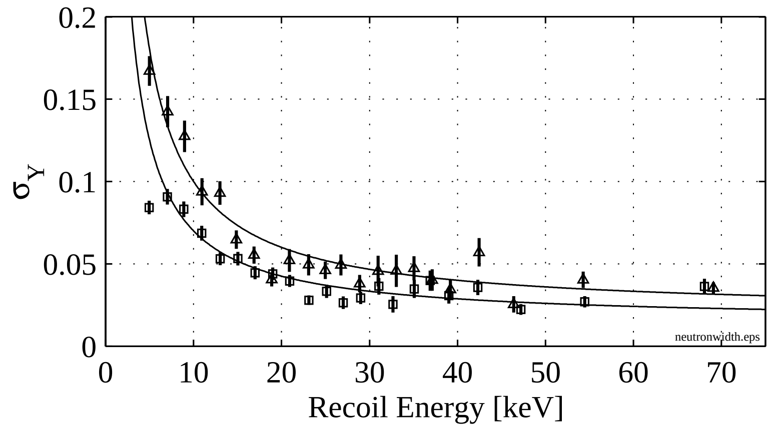


Figure 4.7: Standard deviation of Y in 22 energy bins from 4 to 80 keV. BLIP 1 data points have triangle markers and BLIP 2 points have squares. Best fit lines of the form $\sigma_Y = a + b/E_R$ are shown for each detector, with the values of the parameters a and b given in the text.

ergy and recoil energy should not remain linear at much lower or much higher energy. Indeed, our fit would be non-physical if extrapolated to $E_R = 0$, because it predicts negative charge energy. However, the linear fit seems adequate over the energy range that is important to us, except below 10 keV. In this region, where necessary, we will interpolate between the measured data points to find the expected amount of charge for a given recoil energy.

In Figure 4.7, we show the value of σ_Y in each energy bin. In Appendix A, we derive a formula for σ_Y in terms of the noise in the phonon and charge circuits:

$$\sigma_Y^2 = \frac{1}{E_R^2}\left[\sigma_p^2 \bar{Y}^2 (1+A)^2 + \sigma_q^2 (\bar{Y}A + 1)^2\right].$$

In this expression, σ_q is the r.m.s. charge noise, σ_p is the r.m.s. phonon noise, and $A = |V_b|/\epsilon$ (V_b is the charge bias voltage and $\epsilon = 3.0$ V). If σ_p and σ_q were constant with energy, we would expect $\sigma_Y \propto 1/E_R$, which clearly is not a complete description, since σ_Y appears to approach a constant at high energy. Fitting the data with an expression $\sigma_Y = a + b/E_R$ above 10 keV, we find $a = 0.015$ and $b = 0.55$ keV for BLIP 1 and $a = 0.020$ and $b = 0.81$ keV for BLIP 2 (these lines are shown in Figure 4.7). If we attribute the linear increase in noise with energy to the charge channel, we can write $\sigma_q^2 = \sigma_{q0}^2 + \sigma_{qE}^2 E_R^2$ and at high energy $\sigma_Y \simeq (\bar{Y}A+1)\sigma_{qE}$. Comparing with the results of our fit to the data, $\sigma_{qE} = 0.012$ for BLIP 1 and $\sigma_{qE} = 0.016$ for BLIP 2. This is a fairly gentle increase in the noise with energy (e.g. 1.2 or 1.6 keV at 100 keV), which would otherwise go unnoticed. The reason for this increase in noise is unknown.

4.3 Gamma ray source calibration.

By exposing the detectors to a gamma ray source, we can test how well they can discriminate between a nuclear recoil signal and gamma ray backgrounds. We do not expect the separation between the two populations of events to be perfect because of leakage of background events into the signal region. This leakage can be caused by noise or by the charge collection deficit for interactions in the surface dead layer (See Chapter 3, Section 4).

To expose the detectors to a strong gamma ray flux, we used a ^{60}Co source, which was inserted on a rod into a 1/4" hole penetrating the lead shield. This source produces primarily monoenergetic gamma rays, with energies 1332 keV and 1173 keV. The high energy photons interact with the materials of the inner shield and cryostat, producing showers of lower energy photons and electrons near the detectors.

The source position was adjusted to achieve a counting rate of 3.5 Hz, large

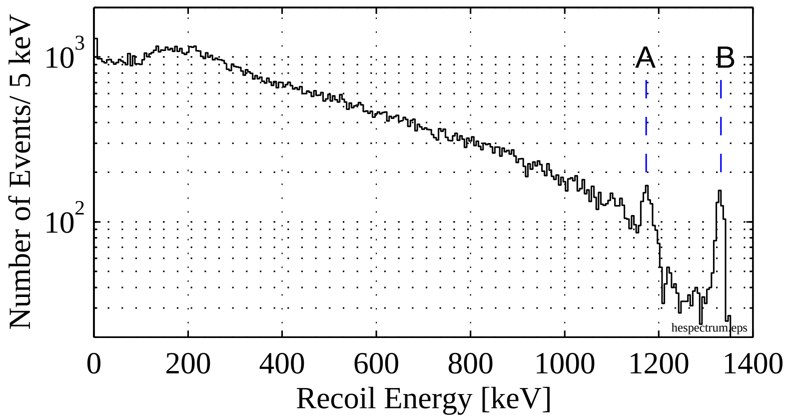


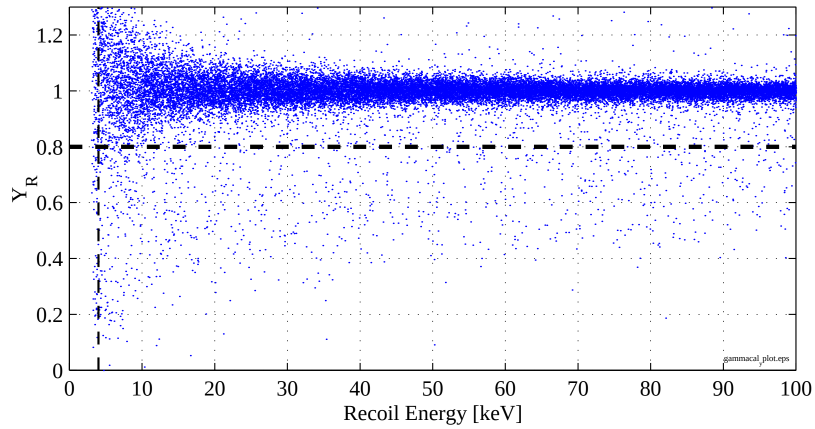
Figure 4.8: Spectrum of the sum of energy in BLIP 1 and BLIP 2 for ^{60}Co events. Full- absorption lines appear at 1173 keV (A) and 1332 keV (B).

enough to completely dominate the muon-anticoincident background rate ($\sim$ 0.01 Hz), but small enough to avoid pulse distortions from pileup in an unacceptably large fraction of events.

Figure 4.8 shows data collected in a 17 live-hour period. In order to increase the efficiency of full- absorption for high energy gamma rays, this spectrum is made by summing the energy in BLIP 1 and BLIP 2 for each event. The full-absorption peaks at 1332 keV and 1173 keV are clearly resolved. There are small peaks at 75 keV and 85 keV from Pb X-ray fluorescence, which appear only in the BLIP 1 spectrum (or in the summed - energy spectrum). These X-rays are presumably made in the inner lead shield and penetrate the thin copper box the detector is held in. Note from Figure 4.1 that BLIP 2 is better shielded from these X-rays by other detectors on the top and bottom, and by relatively thick copper on the sides of the package.

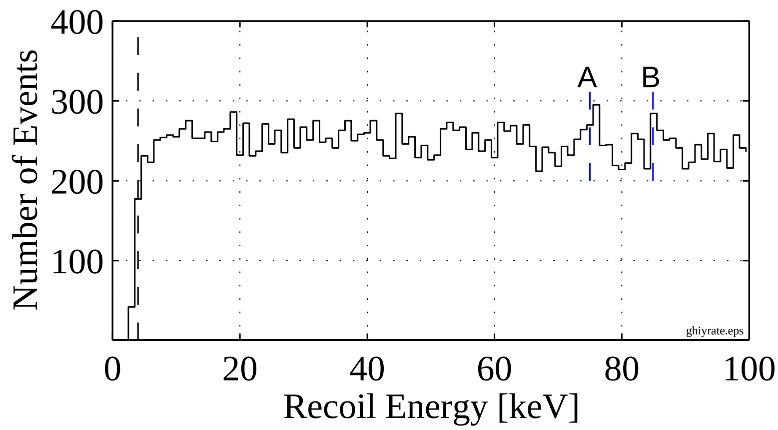
Figures 4.9, 4.10 and 4.11 show the Y and energy distribution of the events below 100 keV. Since the data from BLIP 1 and BLIP 2 is similar, we combined

Figure 4.9: Co-60 calibration data in the 4-100 keV region. There are 15884 events in BLIP 1 and 12942 events in BLIP 2 in 17 live hours.

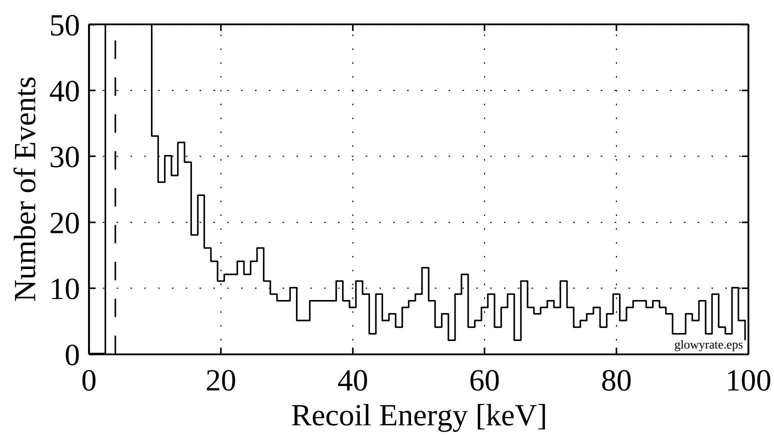


the data from the two detectors on a single plot, but here the energy we plot is the single-detector energy, not the sum of the energy in both detectors.

Approximately 97% of events caused by the source appear distributed symmetrically around $Y = 1$. A long tail containing $(3.2 \pm 0.1)\%$ of the events in BLIP 1 and $(2.9 \pm 0.1)\%$ of the events in BLIP 2 extends to $Y \simeq 0.2$. This tail is due to the existence of a "dead layer" near the surface of the detector, where charge collection is poor. As discussed in Chapter 3, data from measurements made with low energy X-rays incident on devices similar to the BLIPs can be fit with a model in which the fraction of charge collected at depth x varies as $f(x) = (1 - f_0 e^{-x/\lambda})$, with a good fit to the data for $f_0 = 0.7$ and $\lambda = 6$ μm. A Monte Carlo simulation of the ^{60}Co calibration based on this formula predicts a distribution in Y very similar to what we observe, with 3.32% of all events appearing in the low Y tail [98]. The results of this simulation, scaled to the number of events in our data set, are shown with the data in Figure 4.11. The model is fairly successful at reproducing the shape of the low Y distribution as well as the ratio of high- to low- Y events.



(a) Events with $Y > 0.8$. Positions of Pb K_α (A, 75 keV) and K_β (B, 85 keV) X-ray lines are shown.



(b) Events with $Y < 0.8$.

Figure 4.10: ^{60}Co spectra in the low energy region for BLIP 1 and BLIP 2 combined (17 live hours/ detector).

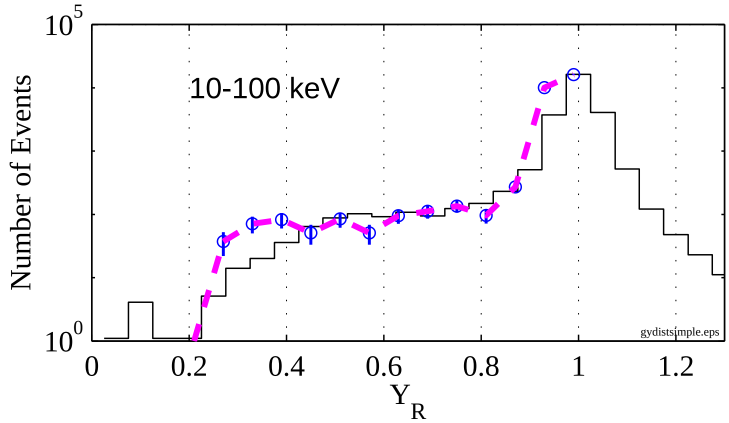


Figure 4.11: Charge yield distribution of Co-60 events from 10 - 100 keV. Solid line shows data. Dashed line shows Monte Carlo results of Ref. [98], based on the dead layer model discussed in the text.

In Figure 4.12, we show the spillover of the ^{60}Co gamma rays into the region of the nuclear recoil band. We plot the charge yield distribution of the ^{60}Co data in the three energy intervals of Figure 4.5, along with curves derived from our fits to the location and width of the neutron distribution. The number of nuclear recoil events in each plot is normalized to be equal to 1% of the number of gamma ray events. Above 30 keV, only 0.2% of the gamma rays spill into the recoil region (defined as the region that includes $\bar{Y} - 2\sigma_Y < Y < \bar{Y} + 2\sigma_Y$). As the energy decreases, the separation between the two distributions is degraded. In the 4-15 keV bin, 3% of the gamma events fall into the nuclear recoil region in BLIP 1 and 4% in BLIP 2.

How much of the leakage into the recoil region is due to noise, rather than to the dead layer? It's clear that the noise is relatively unimportant. For example, at 10 keV, the mean value of Y for neutron events is $\simeq 0.25$, and the standard deviation from noise is $\sigma_Y \simeq$0.1 in the worst detector. The noise width of the gamma ray

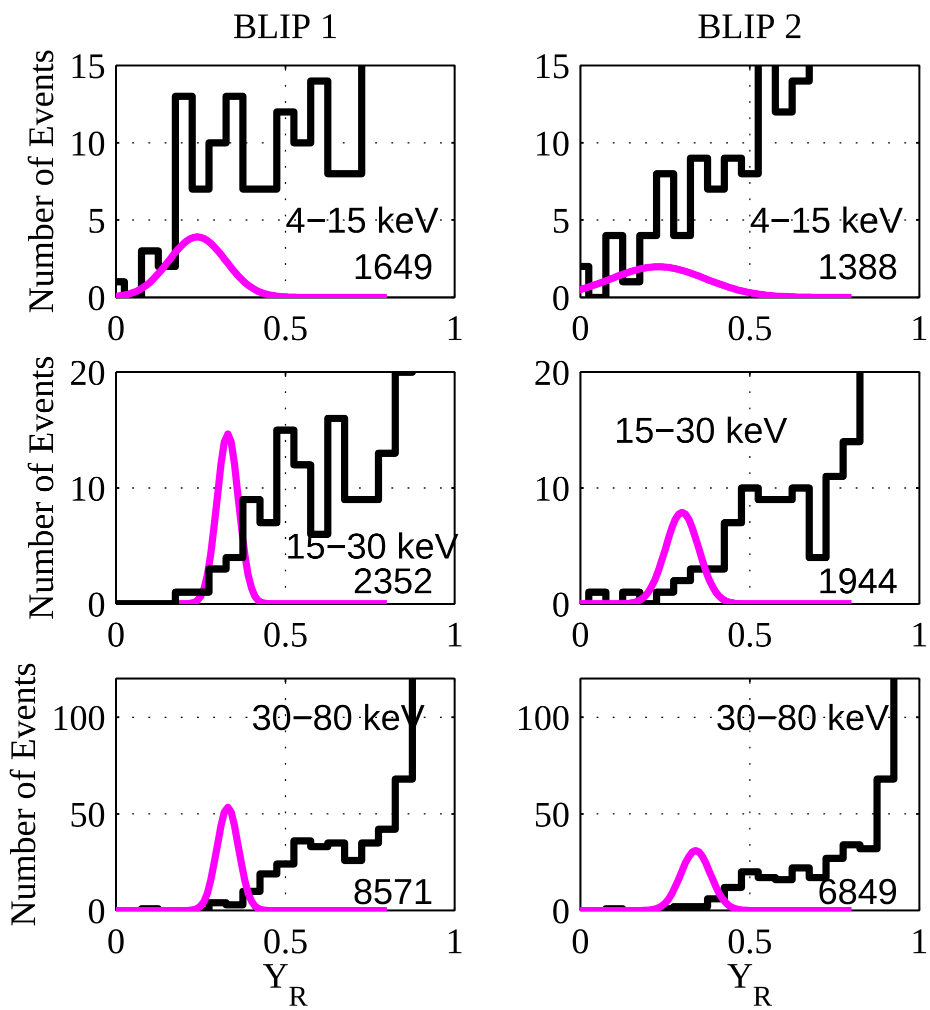


Figure 4.12: Charge yield distribution of ^{60}Co events in the three energy intervals of Figure 4.5. The total number of events in each interval (including events beyond the range of the histogram) is printed in the lower right corner of each plot. The Gaussian curves show the location and width of the measured neutron distribution, with the number of neutron events normalized to be equal to 1% of the number of ^{60}Co events.

distribution is presumably similar and the mean value is $\bar{Y} = 1$. The center of the neutron and gamma ray distributions are separated by ~ 7 standard deviations, implying an overlap of $\sim 10^{-4}$, which is 2 orders of magnitude less than what we observe.

4.4 Optimization of charge yield cut.

In the final data set, we will use a cut on Y to separate candidate WIMP scattering events from background events. In Chapter 2, Section 2.2, we discussed a formalism which we can use to decide where to make the cut. There we showed that to maximize sensitivity to WIMPs, we should pick the cut $Y < Y_c$ to minimize the function Q, defined as

$$Q \equiv \frac{\beta(1-\beta)}{(\alpha-\beta)^2}. \tag{4.1}$$

Here $\alpha(E_R, Y_c)$ and $\beta(E_R, Y_c)$ are the fraction of signal and background events which are expected to pass the cut. We will attempt to construct these functions from the neutron and gamma calibration data.

One problem we have in measuring $\alpha(E_R, Y_c)$, $\beta(E_R, Y_c)$, and $Q(E_R, Y_c)$ is that our data sets do not have enough events for us to construct these functions over small energy intervals. The problem with β is particularly severe, since only a tiny fraction of the ^{60}Co photons interact in the dead layer. This forces us to use large energy bins. However, we want to be sure that the functions are not changing very much from one side of the bin to the other.

To solve this problem, we will change coordinates from Y to a new coordinate Y^*, defined by

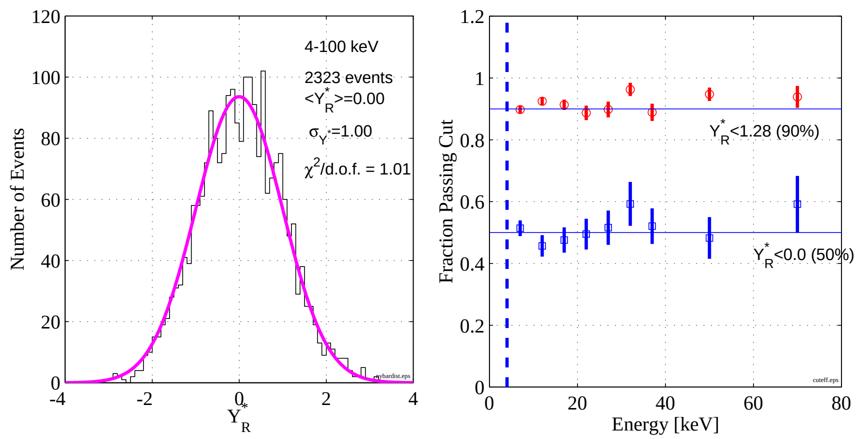


Figure 4.13: (left) Distribution of neutron source events (4-100 keV) in Y^* with a Gaussian fit. (right) The measured efficiency of the cuts $Y^* < 0$ (50 % nominal efficiency) and $Y^* < 1.28$ (90% nominal efficiency) as a function of energy.

$$Y^*(Y, E_R) = \frac{Y - \bar{Y}(E_R)}{\sigma_Y(E_R)}.$$

The signal events have a Gaussian distribution around $Y^* = 0$, with a standard deviation of 1, independent of energy. Cuts made at constant Y^* will have an efficiency that is energy-independent. This is demonstrated in Figure 4.13, which shows that the distribution of neutron events in the Y^* coordinates is very nearly Gaussian, and cuts at $Y^* < 0$ and $Y^* < 1.28$ have efficiencies of $\simeq$50% and $\simeq$90% at all energies.

In Figure 4.14, we show $\alpha(Y_c)$ and $\beta(Y_c)$ in our three standard energy intervals for the two detectors. As expected, α has a very regular sigmoidal shape in the new coordinate system. The β function decreases as energy increases and becomes $< 10^{-3}$ in both detectors above 30 keV for cuts at $Y_c = 0$. In both detectors, β is

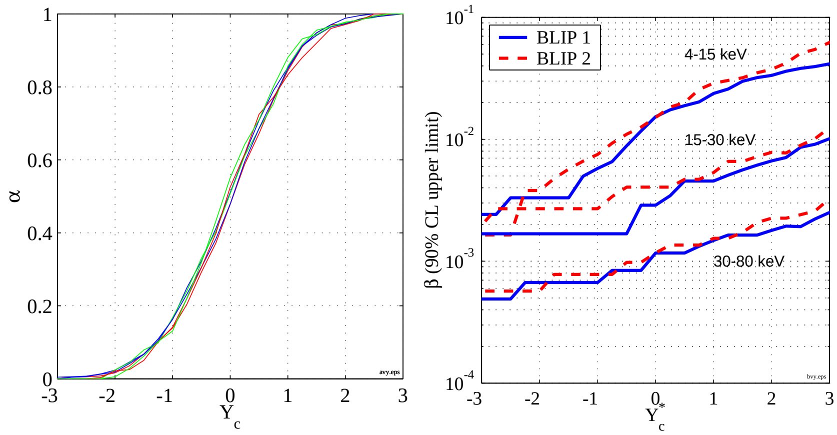
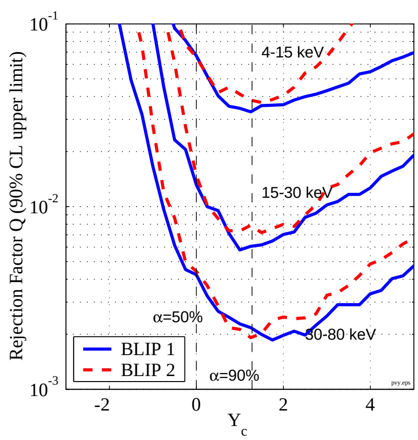
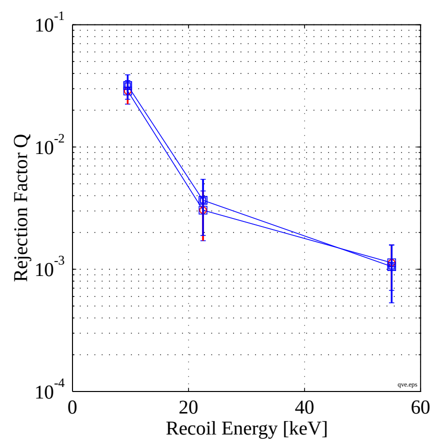


Figure 4.14: The $\alpha(E_R, Y_c)$ and $\beta(E_R, Y_c)$ functions.



(a) $Q(Y_c)$ in the three energy intervals 4-15 keV, 15-30 keV, and 30-80 keV.



(b) $Q(E_R, Y_c)$ at the optimal choice of Y_c as a function of energy for BLIP 1 and BLIP 2.

Figure 4.15: The $Q(E_R, Y_c)$ function.

low enough at small Y_c that it is difficult to measure due to a shortage of gamma calibration events. Because of this problem, we plot the 90% CL upper limit on β instead of the central value.

Using these values of α and β, we can calculate Q from equation 4.1. The results are shown in Figure 4.15. Conveniently, it seems that Q has a minimum at a value of Y_c that is nearly energy independent. This minimum corresponds to $\alpha \simeq 0.9$. There is a clear trend of decreasing Q with energy. At the highest energies, the BLIPs have a background rejection factor for gamma rays of 10^{-3} or better. This factor is nearly equal in the two detectors and has the same energy dependence.

4.5 Background data.

Now that the response of the detector to nuclear recoils and to gamma rays has been measured, and we have gained some insight into how best to separate the two using cuts on the charge yield parameter, we can look at the data taken with no intentional radiation sources present and use it to look for WIMPs.

The data used for this analysis was collected between April 8 and June 11, 1998. This is the subset of Run 18 data for which both the detectors and muon veto system were operating optimally. There were 33.0 days of data acquisition live time out of 64.4 real days. After the elimination of bad running periods, there are 28.9 live days for BLIP 1 and 26.5 live days for BLIP 2, or 4.9 kg-days for BLIP 1 and 4.5 kg-days for BLIP 2. Cuts that remove muon-coincident, noisy, or other types of bad events also remove an estimated 30.0% of good events, leaving the equivalent of 6.5 kg-days of germanium exposure between the two detectors.

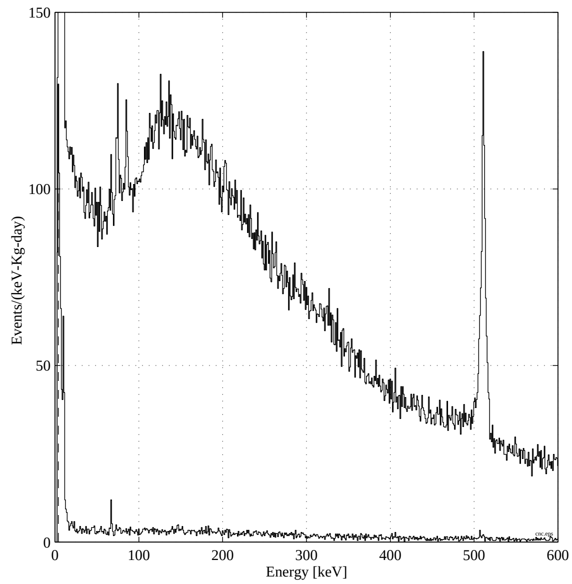


Figure 4.16: Veto-coincident (top line) and anticoincident (bottom line) event rates in the region below 600 keV for BLIP 1 and BLIP 2 combined (6.5 kg-days). In the top line, Pb K_α and K_β X-ray peaks are visible at 75 keV and 85 keV, and a line from positron annihilation at 511 keV. In the anticoincident data, there are lines from ^{73m}Ge at 66.7 keV and X rays at low energy.

The first level of the analysis is described in the appendices, which include descriptions of the pulse fitting and energy calibration (Appendix A), the cuts to remove noise, pileup, and muon coincident events (Appendix B), and the selection of the best quality subset of the data (Appendix C).

4.5.1 Muon coincident data.

The counting rate is dominated by events caused by cosmic-rays. Figure 4.16 shows the effect of using the veto system to select events that are not muon-coincident. Cutting muon-coincident events removes 99% of the 1.3×10^5 events/(kg-day) above 4 keV.

The veto-coincident data has the overall shape typical of a gamma ray spectrum that has been processed by lead, with an edge near 150 keV caused by the rapid increase in the photoelectric cross section as the Pb K-shell binding energy is approached. Interactions in the lead produce K-shell fluorescence lines at 75 keV and 85 keV.

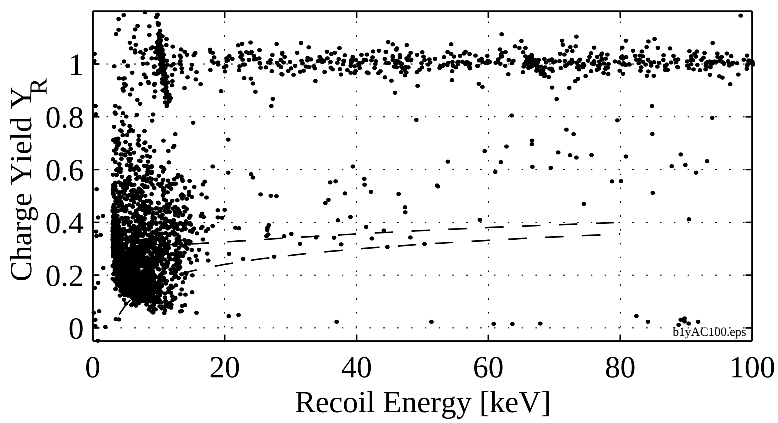
The dominant line feature in the muon-coincident spectrum, at 511 keV, comes from the annihilation of positrons created in muon-induced electromagnetic cascades by pair production. As discussed in Appendix A, this line is used to fix the energy calibration.

4.5.2 Anticoincident spectrum.

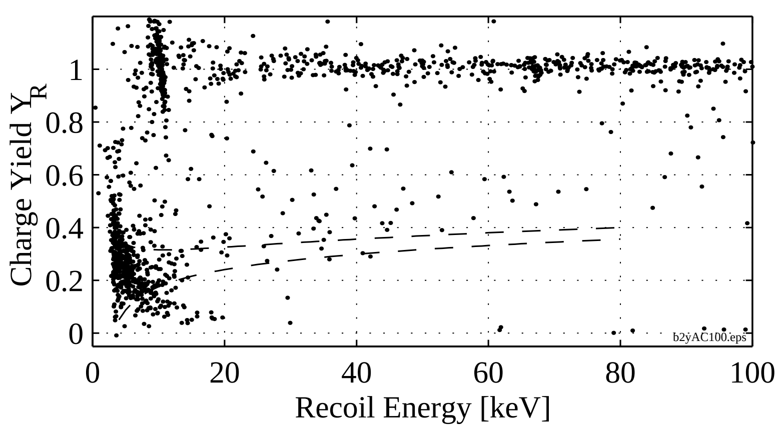
Figures 4.17 and 4.18 show a close-up view of the low energy region of the veto-anticoincident data, and Figure 4.19 shows the measured event rates. In the gamma ray band ($Y > 0.8$) , there is a flat counting rate of 1.4$\pm$0.1 (keV-kg-day)$^{-1}$, with lines at 9.0 keV, 10.4 keV, and 67 keV.

At low Y, there are a large number of events in the region 4-18 keV, with 1794 events in BLIP 1 and 383 events in BLIP 2. As we will show below, the shape of the spectrum of these events is a good match to a ^{3}H beta decay spectrum between 8 keV and the 18.6 keV ^{3}H endpoint. At higher energy, there is a wide, sparsely populated band of events running diagonally in the plots of Figure 4.17 from low Y at low energy to high Y at high energy. Approximately equal numbers of these events appear in each detector. As shown in Figure 4.19b, the rate decreases with energy, from about 1/3 of the gamma band rate at 20 keV to 1/10 at 100 keV.

We believe that the low charge events are due interactions in the dead layer. A similar band has been observed in exposures of small test devices to electrons

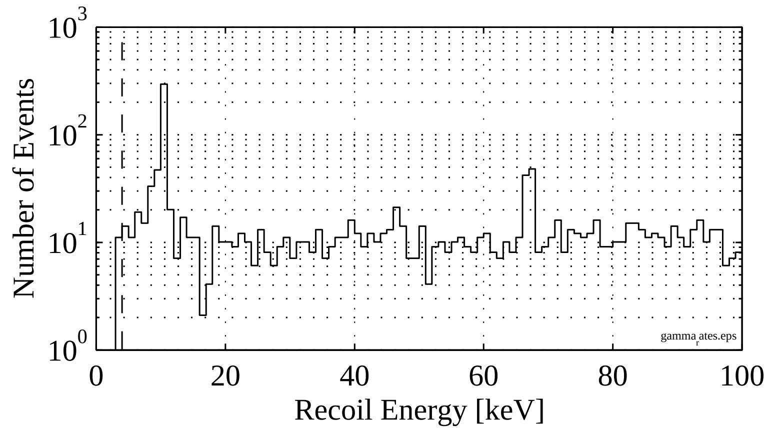


(a) BLIP 1 with 3.4 kg-days exposure.

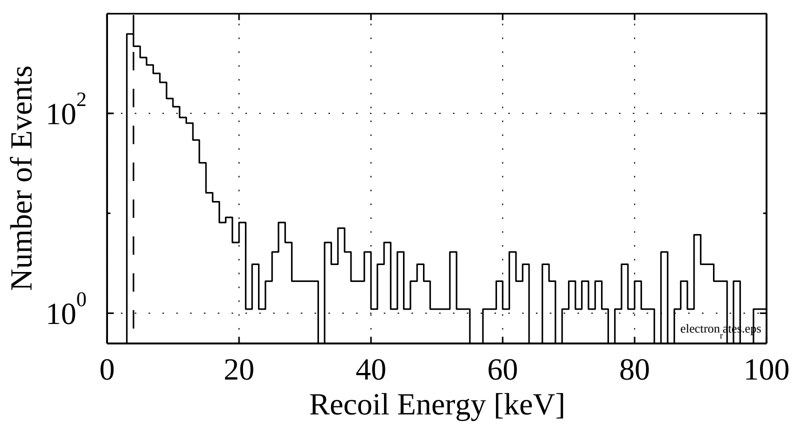


(b) BLIP 2 with 3.2 kg-days exposure.

Figure 4.17: Scatter plots of Y vs. energy below 100 keV. Dashed lines show the $\pm 1\sigma$ bounds of the nuclear recoil band. Only veto-anticoincident, single-detector events are shown.

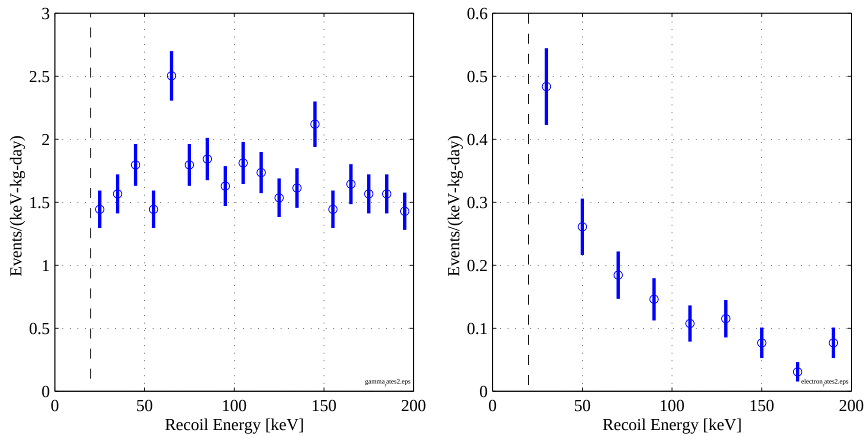


(a) High charge yield events ($Y > 0.8$).



(b) Low charge yield events ($Y < 0.8$).

Figure 4.18: Spectra of combined BLIP 1 and BLIP 2 data for Run 18, with a total exposure of 6.5 Kg-days. Only veto-anticoincident, single-detector events are shown.



(a) Gamma ray spectrum ($Y > 0.8$).

(b) Electron spectrum ($Y < 0.8$).

Figure 4.19: Rates of high- and low- charge yield events in BLIP 1 and BLIP 2 combined.

from ^{14}C beta decay. Data from one of these exposures are shown in Chapter 3, Figure 3.5. Although the distribution of electron calibration events is qualitatively similar to what we see in the background data, it is narrower in Y. This may be due to differences between the test device and a full scale BLIP, or perhaps simply to the fact that the electrons striking the small device were highly collimated. It is not unreasonable to expect variations in the properties of the dead layer across the surface of the detector, which could lead to dispersion in the fraction of collected charge under conditions of uniform illumination. Unfortunately, we do not yet have data from full-scale detectors exposed to an electron source.

It is clear that the bulk of the low Y events are not caused by neutrons or WIMPs, since they fall outside the bounds of the nuclear recoil band. Unfortunately, no clear WIMP scattering signal appears in this data.

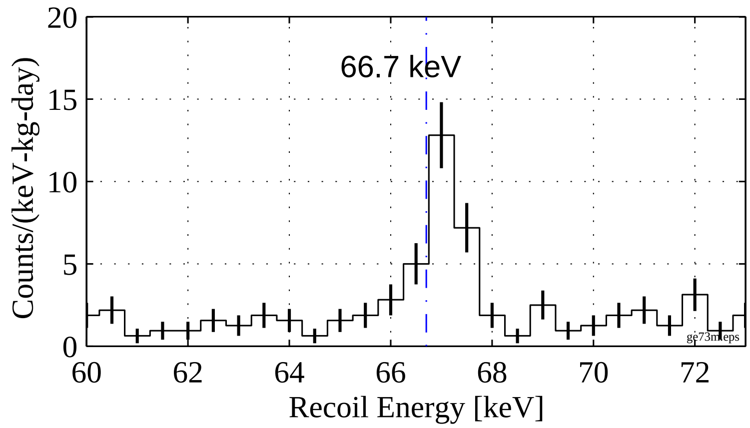


Figure 4.20: The region near the 67 keV peak in BLIP 1 and BLIP 2 combined. The integrated rate in the peak after stripping the background is 10.4 $\pm$1.6 (kg-day)$^{-1}$. A line is drawn to indicate the position of the expected 66.7 keV line from relaxation of ^{73m}Ge.

In the sections below, we will review the physics behind the features in the background data that are understood, before using it to calculate excluded regions of the m_χ, σ_{Wn} parameter space for the halo WIMP population. There are two reasons for examining these data closely. First, it is important to build confidence that the detectors are functioning as expected by explaining the features of the data in terms of phenomena which are understood. Second, a study of these features may lead to ideas for reducing the background level in future incarnations of the experiment.

Thermal neutron capture.

The 67 keV peak, which is shown in Figure 4.20, is probably due to the production of the metastable state ^{73m}Ge, which is 66.7 keV above the ground state. This state has a half-life of 500 ms, and decays to a lower excited state at 13.3 keV ($t_{1/2} = 3.0$ μs) and then to the ground state. Since the second decay is very fast compared to the rise-time of a BLIP phonon pulse, the energy of the two decays would be integrated in the detector. A likely way to make the ^{73m}Ge state is by thermal neutron capture on ^{72}Ge [99]. The capture would be followed by a cascade of short-lived decays with a total energy release of 6.8 MeV. Other possible populating reactions for ^{73m}Ge are ^{74}Ge(n,2n), with neutrons above the 10.1 MeV threshold, or decay of ^{73}As ($t_{1/2}$ = 80 days), which might be produced in germanium by cosmic rays when the detector is above ground. The latter two explanations seem less attractive, since the flux of high-energy neutrons that are over the threshold for ^{74}Ge(n,2n) is believed to be too low, and the cosmogenic production rate of ^{73}As is probably very low.

Let's explore the most straight-forward interpretation, that the state is populated entirely by thermal neutron capture. The ^{72}Ge isotope is 28% of natural germanium and the cross section for the capture reaction is 810 mb for thermal neutrons, which implies that the detectors each have a cross section for this reaction of 0.32 cm^2.

The rate in the 67 keV peak is 10.4 $\pm$1.6 (kg-day)$^{-1}$. The real decay rate must be higher than this, because the long data acquisition dead time following the capture event that populates the state makes acquisition inefficient. By convolving the measured distribution of Run 18 data acquisition times with the expected distribution of decay times, we find that the efficiency is 53%. This implies a real decay rate of 20 (kg-day)$^{-1}$ and a thermal neutron flux of 10 neutrons/(cm^2- day) in the center of the shield. This flux would lead to the production of a number of other isotopes that may be of interest for future background studies. For example, scaling by cross section and abundance, it implies production rates of 322 (kg-day)$^{-1}$ and 69 (kg-day)$^{-1}$ of the beta emitters ^{64}Cu and ^{65}Cu in the copper parts

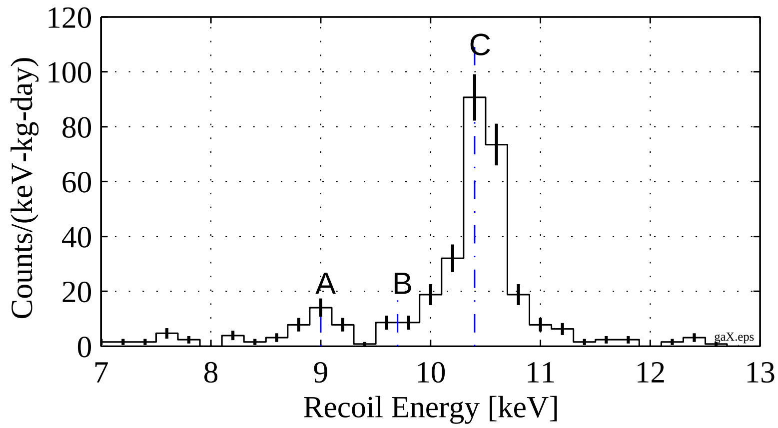


Figure 4.21: Low energy X-ray peaks from internal activity of Ge crystals. Peaks due to (A) ^{65}Zn$\rightarrow$^{65}Cu at 8.98 keV, and (C) ^{68}Ge$\rightarrow$^{68}Ga + ^{71}Ge$\rightarrow$^{71}Ga at 10.4 keV are clearly resolved. A third peak (B) from ^{68}Ga$\rightarrow$^{78}Zn at 9.66 keV may contribute to the low energy tail of the 10.4 keV peak. Peaks occur with energy equal to the full K- shell binding energy, because all emitted X-rays are fully absorbed. This figure combines the data from BLIP 1 and BLIP 2.

near the detectors.

Low energy X-ray lines.

The inferred neutron flux would also lead to a capture rate of 54 (kg-day)$^{-1}$ on ^{70}Ge , which has a capture cross section 2900 barns and a 21% abundance. This reaction produces ^{71}Ge, which decays to ^{71}Ga by K - shell electron capture with an 11 day half-life. The electron capture is followed by relaxation of the electrons to fill the K- shell vacancy, with a total energy release of 10.4 keV. Thus, this reaction will produce a line in the spectrum at 10.4 keV.

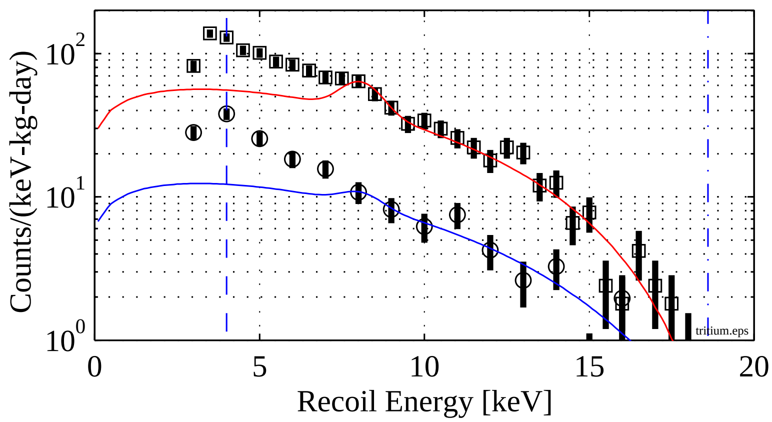
The region around 10.4 keV is shown in Figure 4.21. The peak at 10.4 keV, which has an integral rate of 49 $\pm$ 3 (kg-day)$^{-1}$ can also have a contribution from ^{68}Ge $\rightarrow$^{68}Ga. The isotope ^{68}Ge ($t_{1/2}$ = 270 days) is made by spallation from cosmic ray neutrons when the detectors are above ground. Avignone *et al.* have estimated the production rate to be 30 $\text{kg}^{-1}\text{day}^{-1}$[94]. This should be approximately the decay rate in our detectors, since they were underground for only a fraction of the 270 day half-life before this measurement was made and the material they are made of was above ground for several years.

The sum of the expected rate at 10.4 keV from ^{68}Ge and the inferred rate from ^{71}Ge due to neutron capture is 84 $(\text{kg-day})^{-1}$, significantly higher than the measured rate. This is an interesting puzzle, which deserves further study, since a possible implication is that the high-energy neutron flux is greater than we had supposed.

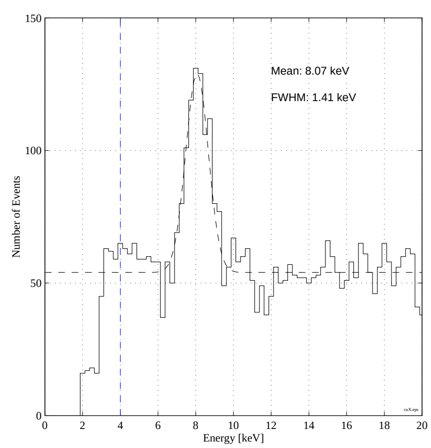
Tritium surface contamination.

In the region 4-20 keV, there are a large number of events with low charge yield in both detectors. The rate in BLIP 1 is higher by a factor of 4.7. As shown in Figure 4.22a, the spectrum is a good match to ^{3}H (tritium) β decay near the endpoint (Q=18.6 keV). The fact that the tritium events all have a low charge yield indicates that this contamination is external to the detector crystals, with all the β particles falling into the most absorbing fraction of the surface dead layer. Even the highest energy β particles from tritium penetrate less than 1 μm of germanium.

Below 9 keV, the data departs increasingly from a pure tritium spectrum. There is bump at 8 keV in BLIP 1 which appears at the location of a Cu X-ray line. This line is clearly visible in veto-coincident data, as shown in Figure 4.22b. We fit the data with two parameters, (1) the amplitude of a fully-absorbed tritium spectrum, and (2) an 8 keV Cu X-ray fluorescence with a Gaussian shape and a width measured from the veto-coincident data. Since it is impossible to fit the low



(a) Low-energy, low-Y spectrum in BLIP 1 (top) and BLIP 2 (bottom), with best-fit lines.



(b) The 8.05 keV copper K_α X-ray in veto-coincident data, with a Gaussian fit.

Figure 4.22: Fit of low energy spectrum to tritium + copper X-rays (a). The veto-anticoincident spectra are fit with a combination of a tritium β decay spectrum plus a copper X-ray line, the shape of which is measured in the veto-coincident data (b).

energy part of the spectrum very well with just these two parameters, we used a least - squares fit to the data over the interval $7 < E_R < 19$. There is no significant evidence for a Cu X-ray in the BLIP 2 data, but it clearly helps the BLIP 1 fit.

The excess of events at low energy relative to what would be seen in an ideal tritium spectrum suggests that the source is external to the detectors. Excess events at low energy would be caused by electron backscattering from the surface of the germanium, or by self-absorption in an external source. In Run 19 (which began 5 months after Run 18 ended, in November 1998), we tested the hypothesis that the tritium was on the copper parts surrounding the detectors by switching the BLIP 1 and BLIP 2 detector mounting hardware. Preliminary results show that the tritium has followed the hardware, with a larger counting rate now observed in BLIP 2 than in BLIP 1. It is not clear how the copper became contaminated with the $\sim 10^6$ atoms of ^{3}H that are needed to explain the rates we measure. It is possible that this contamination is generically present in OFHC copper, but has not been noticed before.

Comparison with Oroville experiment event rates.

Figure 4.23 compares the BLIP spectrum with the result of the UCSB/UCB/LBL Oroville experiment, one of the first-generation dark matter searches. Since the Oroville detectors were conventional germanium ionization devices, we compare them with the BLIPs on the basis of the charge signal alone. The main effect of this for the BLIPs is to suppress the low-charge yield portion of the spectrum, which is dominated by tritium beta decay counts. It's interesting that the Oroville spectra rise gradually below the tritium endpoint in much the same way the BLIP spectra do. In fact, at the time of the experiment, this was attributed to a low-level

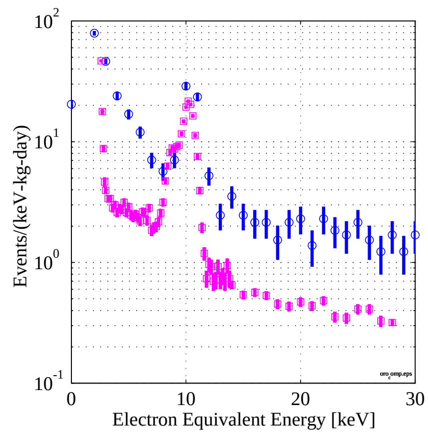
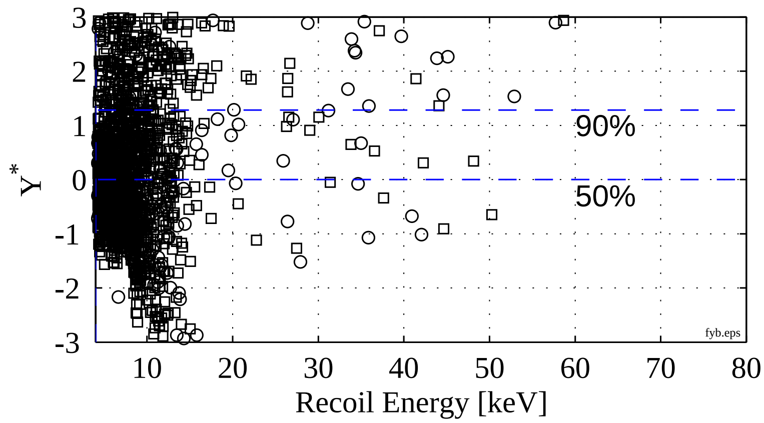


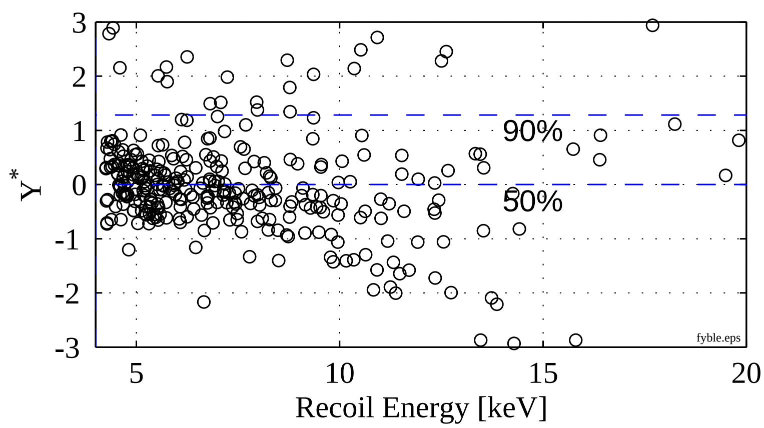
Figure 4.23: Comparison to Oroville spectrum. The upper line with error bars is the BLIP spectrum and the lower line is the spectrum from 327 kg-days exposure of 2 0.9 kg detectors at Oroville. Energy for the BLIPs is plotted based on the ionization signal only, since the Oroville detectors could not separately measure the recoil energy.

internal contamination of the germanium by tritium. However, as we have noted, the tritium in the BLIPs must be near the germanium surface rather than in the bulk. The Oroville detectors would have been insensitive to surface contamination, because they had a charge collection dead layer about 1 mm thick.

The direct comparison with Oroville data shows quite dramatically that CDMS has not achieved background levels equivalent to the older experiments based on conventional technology. This can be attributed to the necessity of operating at the relatively shallow SUF site and to the complexity of the cryogenic apparatus, which makes maintaining radiopurity difficult. Unfortunately, some of the discrimination power of the detectors must be used just to “pull even” with other experiments.



(a) BLIP 1 (squares) and BLIP 2 (circles) from 4-80 keV.



(b) BLIP 2 only from 4-20 keV.

Figure 4.24: The Y^*distribution of the Run 18 BLIP 1 and BLIP 2 data. Dotted lines at $Y^* = 1.28$ and $Y^* = 0$ show the regions that should include 90% and 50% of nuclear recoil events.

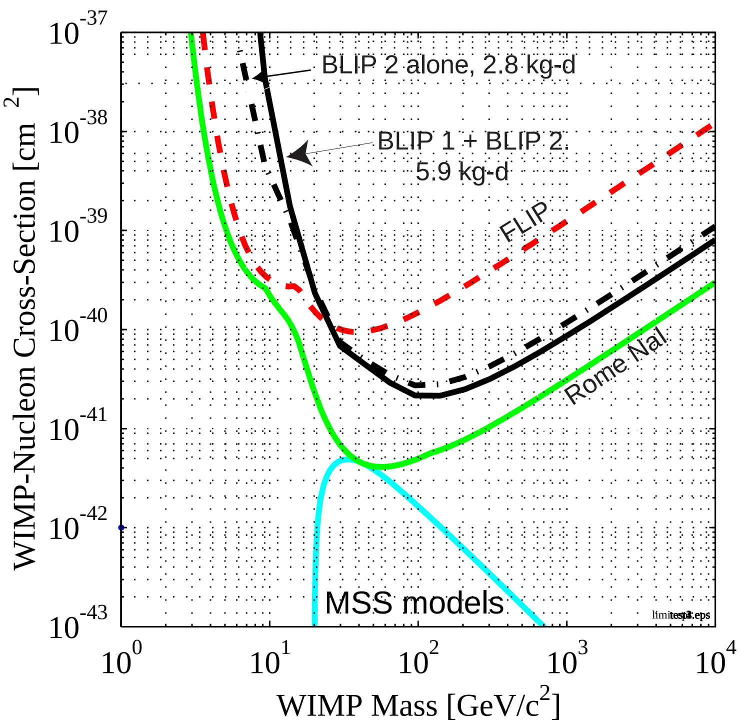


Figure 4.25: Limits (90% CL) on halo WIMP masses and cross sections from Run 18 BLIPs calculated using the "maximal gap" method. Also shown for comparison are a set of MSSM models from Ref. [51], the preliminary Run 18 limits from the silicon FLIP, and the limits from the DAMA (Rome) NaI experiment, which are currently the most restrictive [55].

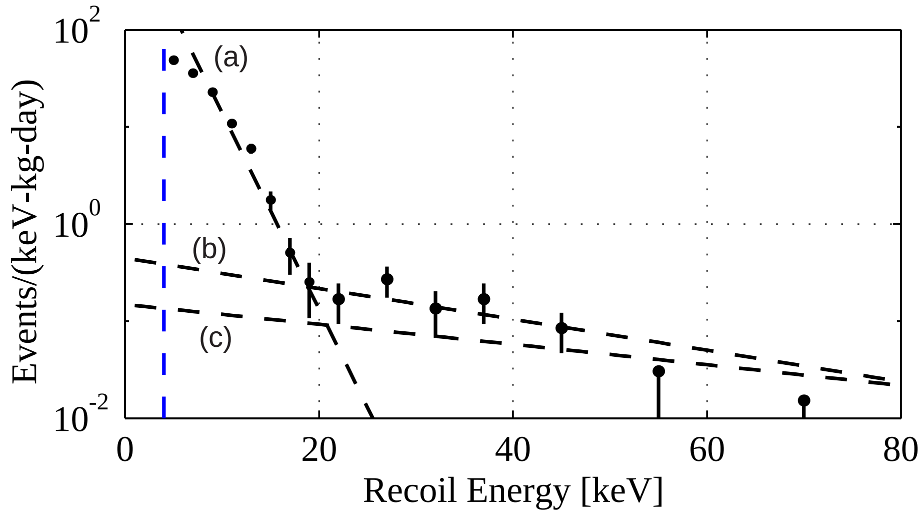


Figure 4.26: BLIP 1 and BLIP 2 summed spectrum with 5.9 kg-days effective exposure after a 90% efficiency nuclear recoil cut. Error bars show 1σ Gaussian errors. In the last two bins, there are no events, so we show the 84% CL upper limit on the rate. Some WIMP spectra that are excluded by the "maximal gap" method are shown as dotted lines: (a) $m_\chi = 10$ GeV/c^2, $\sigma_{Wn} = 1.0 \times 10^{-38}$ cm^2, (b) $m_\chi = 100$ GeV/c^2, $\sigma_{Wn} = 3.0 \times 10^{-41}$ cm^2, and (c) $m_\chi = 1000$ GeV/c^2, $\sigma_{Wn} = 1.0 \times 10^{-40}$ cm^2.

4.6 Looking for WIMPs.

Now we are close to our goal of setting limits on the rate of events that could be caused by WIMPs. Unfortunately, it is clear that the gamma ray calibration data presented in Section 3 has a very different Y distribution than the background data shown here. Although it seems likely that some electrons are contributing to the events we see in the "nuclear recoil" region, we certainly don't have the right calibration data to accurately estimate how many. In fact, given the possibility of spatial variation in the fraction of collected charge over the surface of the detector, obtaining calibration data that could be used to make this estimation might be a practical impossibility.

However, our neutron calibration data has allowed us to estimate the $\alpha(Y^*)$ function. We can use this knowledge to reduce the background rate by looking for dark matter in the smallest possible Y interval. Our analysis of the calibration data suggested that the best results would be obtained for $\alpha \simeq 0.9$, corresponding to a cut that includes events with $Y^* < 1.28$. This cut was optimal for reducing the effect of statistical fluctuations after subtraction of the expected number of background events. We will use the cut here even though we do not intend to subtract anything. This is reasonable under the assumption that the Q function for the real signal and background resembles the Q of the calibration data, even if it is not exactly the same.

Figure 4.24a shows the distribution of the data in E_R and Y^*, with lines corresponding to cuts at $\alpha = 0.5$ and $\alpha = 0.9$. Looking at the events above the 18.6 keV ^{3}H endpoint, it seems that the cut $Y^* < 1.28$ (90% efficiency) is a quite reasonable choice, since the counting rate would double if we tried to accept events up to

$Y^* < 3$ (99.8% efficiency). Also, dropping the cut to $Y^* < 0$ (50% efficiency) does not help, because the number of background events decreases no faster than the number of signal events.

Since BLIP 2 has many fewer events in the low energy region, the BLIP 2 low energy data is presented separately in Figure 4.24b. Below 10 keV, it is questionable whether we really know what the efficiency of cut at low $Y^*$is, because of the possibility that $\bar{Y}$ and σ_Y are different in this data set than they were in the calibration data. Although it appears that we could get a substantial reduction in the background rate by moving the cut to lower Y_c, this seems very risky, since we can't calculate the fraction of signal events that would pass the cut with confidence. The cut at $Y^* < 1.28$ is safe, however, because events that are rejected by the cut always have a charge signal significantly above noise.

To calculate our sensitivity to WIMPs, we will use the "maximal gap" method discussed in Chapter 2 and in Ref. [58] to calculate 90% CL upper limits on the cross section of possible halo WIMPs for scattering on a single nucleon as a function of WIMP mass. Producing these limits involves a number of assumptions about the galactic halo and the form of the WIMP-nucleus interaction, which are also discussed in Chapter 2. For high-mass WIMPs, we should get the best results by combining the BLIP 1 and BLIP 2 data, since the detectors have similar background rates above the ^{3}H endpoint and the increased statistics help reduce uncertainty in the measured spectrum. For low masses , the BLIP 2 data alone should be used, because the counting rate at low energy is less.

The results for BLIP 1 combined with BLIP 2 (5.9 kg-days) and BLIP 2 alone (2.8 kg-days) are shown in Figure 4.25. The two data sets yield limits that are equivalent for practical purposes. As expected, there is slightly more sensitivity at low mass with the BLIP 2 set, and slightly greater sensitivity at high mass with the combined set. In Figure 4.26, we compare the measured background spectrum with the expected spectra for WIMPs that are excluded by the maximal gap method.

Conclusions.

The sensitivity we have achieved in this measurement is close to, but not quite equal to the sensitivity of the best conventional germanium experiment [65], or of the best NaI experiment [55]. The main importance of this work clearly lies in its future promise, rather than in the limited Run 18 result. These measurements can be regarded as a "proof of concept", since we have demonstrated that the gamma ray backgrounds which have limited dark matter search experiments in the past can be suppressed by at least 2-3 orders of magnitude by using the simultaneous ionization and phonon measurement technique. We have also proved that it is possible to operate an array of large cryogenic detectors for a period of months in an environment that has reasonably low background radiation fluxes.

In the future, sensitivity could be dramatically improved by either decreasing the flux of electrons striking the surface of the detectors or by making the detectors less sensitive to them. Since the end of Run 18, the CDMS collaboration has made progress on both fronts. For example, in Run 19, the background from ^{3}H has been eliminated by lining the detector holders with high-purity germanium wafers. The Run 19 BLIPs also use a new ionization contact technology, which significantly lowers the maximum charge-loss for surface events. Another exciting development is the demonstration that surface events can be identified in FLIP-type detectors using pulse shape discrimination [77]. Preliminary Run 19 data shows that CDMS has now achieved greater sensitivity than all other experiments for some WIMP masses.

Appendix A

Waveform reduction.

In this appendix, we describe how the raw data recorded by the phonon and charge channel digitizers is turned into a few simple parameters (e.g. energy and charge yield) that describe each event.

A.1 Pulse fitting.

For each event, we recorded two charge traces for each detector, one each for the inner and outer charge electrodes, and two temperature traces, one for each of the two NTD thermometers. The digitization rate and number of samples recorded has varied from run to run as we experimented with various compromises between the need for more information and less data volume. For Run 18, the charge traces were 4 ms long, with 1000 samples (nominally 250 pretrigger and 750 posttrigger) at 250 kHz. The temperature traces were 128 ms long, with 2000 samples (750 pretrigger and 1250 posttrigger) at 15.6 kHz.

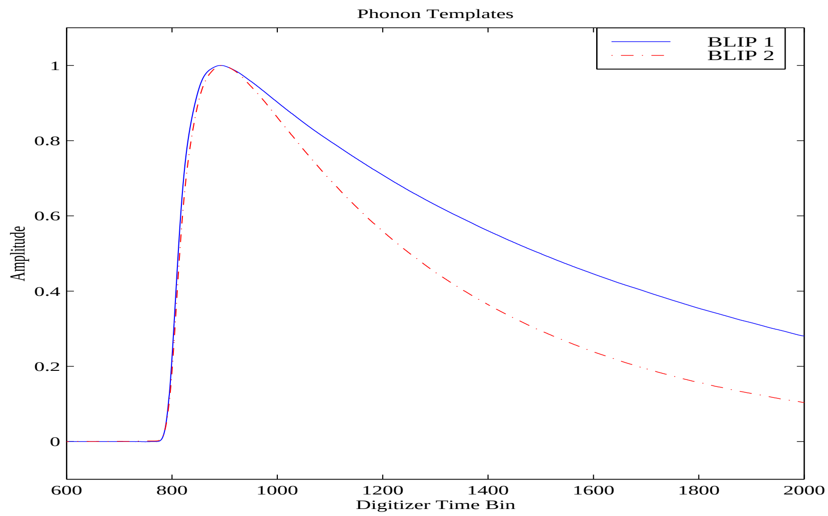
Figure A.1 shows the charge and thermal pulse shapes determined by averaging

~ 100 measured pulses. The charge pulse shape is determined by the bandwidth of the front end amplifier, the resistance and capacitance of the amplifier feedback network, and the noise filtering we do before digitization. The thermal pulse shape is set by the heat capacities and thermal conductivities of the electron and phonon systems in the detector. The electrical constants are relatively easy to control, so the charge pulses from different charge sensors have nearly the same shape. The phonon pulses, on the other hand, tend to vary substantially between sensors and over time as the temperature of the refrigerator changes. An example of this is the longer fall time for BLIP 1 vs. BLIP 2 seen in Figure A.1a. The cause of this was never investigated in detail, but it is likely due to differences in the quality of heat sinking between the detectors and the low temperature thermal bath.

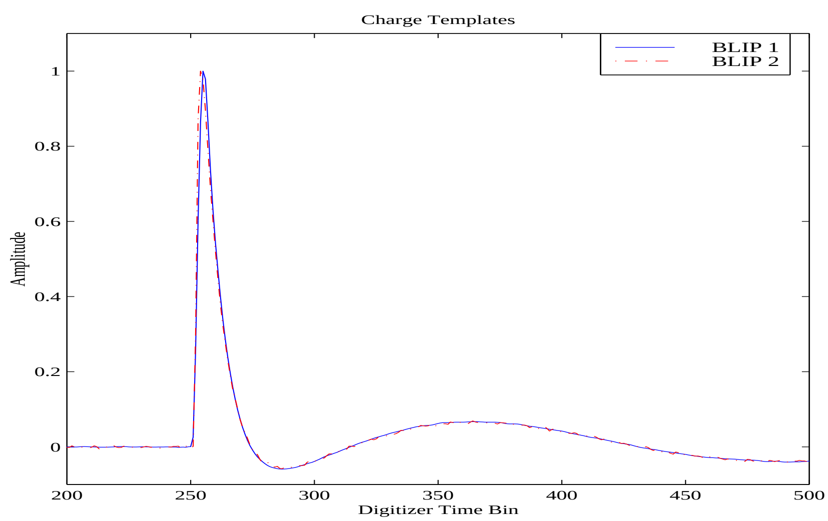
The handling of the data from the digitizers depends on whether data recording is triggered by the charge channels (a "charge trigger") or NTD channels (a "phonon trigger"). In each case, the trigger signal comes from a discriminator circuit that fires at a fixed pulse height after some simple filtering. The filters include a single pole low pass filter to suppress high frequency noise, typically from power-line glitches, and a notch filter to remove crosstalk into the charge channels from the 1 kHz thermometer bias. The discriminator thresholds were adjusted during the run to be as low as possible without incurring a large rate of noise triggers and the consequent dead-time penalty. They were the equivalent of 3-4 keV for the phonon trigger and 7-8 keV for the charge trigger.

A charge trigger always occurs within 1 μs of the actual interaction time. In this case, data acquisition is relatively simple. The acquisition computer waits for long enough for the full waveforms to have been recorded in all sensors, downloads them, and starts acquisition again (See Chapter 3 for some more details).

Figure A.1: Run 18 Pulse Templates.



(a) Phonon templates for BLIP 1 and 2. The time scale is 64 μs per bin.



(b) Charge traces. The time scale is 4 μs per bin.

Since the charge trigger for a given event will always occur a few ms before the phonon trigger, phonon triggers are only important for events that deposit small amounts of charge. Phonon triggered events may have an associated sub-threshold charge pulse, which is stored in the charge digitizer's pretrigger memory. The attempt to recover this signal involves three steps: (1) The data acquisition computer convolves the observed phonon pulse with a phonon pulse template to determine the time at which the event occurred with sufficient accuracy to choose the right 4 ms long piece of charge digitizer memory to download. (2) In the off-line pulse analysis, a convolution of the downloaded charge trace with a second template is used to narrow the start time to a few μs. (3) After the time of the charge pulse is computed, the phonon trigger data can be treated just like charge trigger data, and its amplitude determined using the fits discussed below.

Problems occur if either step (1) or (2) gives the wrong result. This happens increasingly for charge pulses which have amplitudes close to the frequency-averaged noise in the charge channel, typically $\simeq 1.5$ keV f.w.h.m. The consequence is that the measured charge for these events becomes uncorrelated with the charge that was really made in the interaction. Another serious problem is that it is impossible to check if there was an associated hit in the muon veto system, since the time of the event isn't known well enough.

The signal amplitudes are determined using a computer program called "dark-Pipe" , which minimizes the χ^2 of a multi-parameter fit to each trace. The charge fit has three parameters for each of the two channels on each detector, (1) a constant baseline offset, (2) the amplitude of the charge pulse, and (3) the amplitude of a crosstalk pulse from a possible charge signal in the other channel of the same detector. The shapes of the signal and crosstalk pulses are given by pre-recorded templates, such as those shown in Figure A.1.

For the temperature traces, no crosstalk component is needed in the fit, but an additional component is added to reduce the effects of pulse pileup, since these pulses are much slower than charge pulses and pileup occurs more frequently. The extra component is a falling exponential in time, with a 1/e time constant equal to the fall time of a real phonon pulse. The amplitude of this exponential at the start of the trace is left as a free parameter in the fit. This procedure accounts well for the effect of a pulse that started before the beginning of the pretrigger range, but poorly for pulses that start inside the pretrigger range, since no attempt is made to fit the rise at the start of the early pulse. Also, no attempt at all is made to compensate for pulses which start in the posttrigger period. The effect of pretrigger and posttrigger pulses is to badly distort the fit, yielding high χ^2 values. Piled-up pulses can be rejected by cutting out events with high χ^2, as described in Section B.5 below.

As usual, our χ^2 parameter is calculated from an expression of the form

$$\chi^2 = \frac{1}{M-n}\sum_{i=1}^{M}\left[\frac{S_i - E_i(\alpha_1 ... \alpha_n)}{\sigma_s}\right]^2 \tag{A.1}$$

where S_i is the signal, E_i is the expected signal for the fit parameters $\alpha_1 ...\ \alpha_n$, and σ_s is the expected statistical variation of $S_i - E_i$. For σ_s, we use the r.m.s. noise per bin in the relevant sensor channel, calculated using empty digitized traces. For a fit consistent with statistical fluctuations, we expect $\chi^2 \sim 1$ independent of the values of the parameters $\alpha_1 ...\ \alpha_n$. What we actually find is a correlation between the amplitude of the fit and the χ^2, with $\chi^2 \gg 1$ for large amplitude pulses. This tells us that the pulses are not statistically consistent with our template model.

The variation of χ^2 with pulse amplitude is not of fundamental importance, since we are not interested in using the χ^2 to prove the validity of the pulse model, only to identify misshapen events. For this, we just have to keep track of what the normal χ^2 range is for pulses of a given energy. It is useful, however, to consider the possible causes of the amplitude variation we observe, if only to help understand what the normal χ^2 range should be at a given amplitude.

At least two effects could contribute to the deterioration of the fit quality as amplitude increases,

1. The real pulse shape is amplitude-dependent, since the heat capacities and thermal conductivities that determine it are functions of temperature. For convenience, however, we use a single template to fit pulses of all amplitudes.

2. The pulse-shape templates we use are not perfect representations of the real pulse, since they are calculated by averaging over a finite number of noisy measured pulses.

The effect we expect from (2) can be calculated. Let's consider the behavior of a simple 1 parameter pulse fit to a shape template in the case where the template has some noise in it. Let αS_i be the true form of a pulse with amplitude α. This pulse is observed in a noisy system, resulting in a measured pulse $M_i = \alpha S_i + n_i$, where n_i is a random variable with $\langle n_i \rangle = 0$ and $\langle n_i^2 \rangle = \sigma^2$. The pulse will be fit to a template $T_i = S_i + \epsilon_i$, which has a noise ϵ_i. If the template was found by averaging over m measured pulses of amplitude γ, then $< \epsilon_i >^2 = \sigma^2/\gamma^2 m$. The χ^2 for a fit of the measured pulse to a template-shaped pulse of amplitude β is

$$\chi^2(\beta) = \frac{1}{K}\sum_{i=1}^{K}\left[\frac{M_i - \beta T_i}{\sigma}\right]^2 = \frac{1}{K\sigma^2}\sum\left(\alpha S_i + n_i - \beta S_i - \beta\epsilon_i\right)^2$$

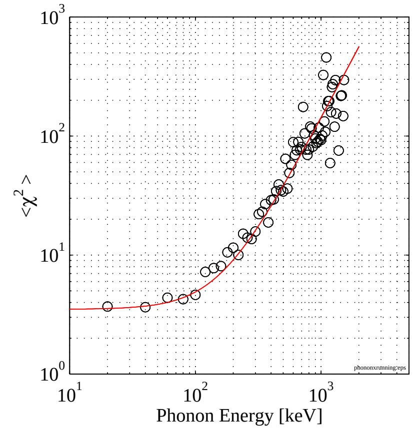


Figure A.2: Running of $< \chi^2 >$ with amplitude for BLIP 1 phonon sensor 1. There is a fit line $< \chi^2 >= 3.5 \times (1 + 4 \times 10^{-5} E_p^2)$.

For a large number of samples K, minimizing χ^2 should result in a fit error $f = \alpha - \beta$ that has an absolute value which is much less than the error σ in an individual bin. The expectation value of χ^2 is then

$$\left\langle \chi^2 \right\rangle = \frac{1}{K\sigma^2} \sum \left\langle (fS_i + n_i - \beta\epsilon_i)^2 \right\rangle \simeq 1 + \frac{\beta^2}{\gamma^2 m} \tag{A.2}$$

Does this explain the observed scaling of $\langle \chi^2 \rangle$ with pulse amplitude? For phonon pulses in Run 18, we used templates formed from about 200 measured pulses with energy ~ 100 keV, which leads to the prediction $\langle \chi^2 \rangle \simeq 1+(E_p/1 \text{ keV})^2 \times 5 \times 10^{-7}$. For the charge pulses, we used 20 pulses of mean energy $\sim$50 keV, which suggests $\langle \chi^2 \rangle \simeq 1 + (E_q/1 \text{ keV})^2 \times 2 \times 10^{-5}$. These equations predict much lower absolute values of χ^2 than we observe. This tells us that the running of the χ^2 is not simply due to statistical errors in the template. It may well be, however, that there are other errors in the template that are larger than the statistical errors. Equation A.2 is useful because the observed energy dependence does have the form

	BLIP 1	BLIP 2
PHONONS (P1+P2)/2	157±1	246
CHARGE (QIN+QOUT)	432	541
RECOIL ENERGY	251	315

Table A.1: The r.m.s. baseline noise in BLIP 1 and BLIP 2 determined by fitting data from random triggers. All numbers are in units of electron volts. Only the lowest noise periods (as defined in Appendix C) were included.

$< \chi^2 > \propto 1 + \alpha E^2$. This is illustrated in Figure A.2, where we plot $< \chi^2 >$ vs. energy for a BLIP 1 phonon channel with a fit of this form. We will use curves with this shape to define cuts on χ^2.

There is an overall amplitude-independent error in the normalization of χ^2 caused by errors in σ_s in Equation A.1. Our estimate of σ_s is never very accurate, since we neglect its time variation, which can be caused both by changes in the external noise environment and by our own adjustments to the electronics. A more general statement of this problem is that our noise is not flat in the frequency domain, which results in noise correlations between time bins. In the future, we plan to abandon the use of equation A.1 in favor of a χ^2 calculated in the frequency domain, with the correct noise weight given to each frequency bin.

The noise in the fit can be measured by fitting empty, randomly-triggered phonon and charge traces. It's necessary to cut out data with accidental pulses or noise glitches and to avoid periods where the noise "signal" is not in the digitizer's dynamic range. This results in the numbers shown in Table A.1.

A.2 Time dependent energy calibration.

The pulse shapes from our detectors changed as the base temperature of the cryostat drifted during the run. The templates used in the fits were periodically upgraded as the temperature drifted, but this does not completely remove the effect. It's necessary to calibrate the data in a time dependent way to get the best possible energy estimate for each event.

There are three line features in the BLIP background spectrum that can be clearly identified, a 10.4 keV Ga X-ray, a 67 keV gamma ray from neutron capture on Ge, and a 511 keV line from positron annihilation. The 511 keV line has by far the highest intensity, which makes it the best one to use for the energy calibration. Its high energy is also helpful since the error in determining the location of the peak leads to a low systematic error in the overall energy scale. As illustrated in figure A.4, the 511 keV line is easily distinguished in data sets with ~ 0.5 live days exposure. This allows us to calibrate the detectors separately for each file series of raw data, which is very convenient. It was "standard operating procedure" during Run 18 to start a new file series whenever any significant change occurred in operating conditions, so calibration jumps are more likely to occur between series than within series.

The calibration procedure assumes that the amplitude obtained from pulse fitting is linear in energy and that a zero energy pulse produces zero amplitude. At some level, both of these assumptions will break down, but they seem adequate for determining the energy scale to a precision of $\sim$1% up to at least 1 MeV.

The following procedure was followed within each file to find the calibration constants:

1. For each NTD thermometer, the fit amplitudes were histogrammed in bins with width corresponding to $\sim$ 5 keV. This yielded the spectrum N_i , with bin index i. The phonon amplitude of an event in bin i is P_i.

Figure A.3: Time variation of calibration constants during Run 18. This is a plot of $(K - \langle K \rangle)/\langle K \rangle$ for $K_{\rm ntd1}$, $K_{\rm ntd2}$, and $K_{\rm qouter}$ in BLIP 1 for files with more than 0.2 live days.

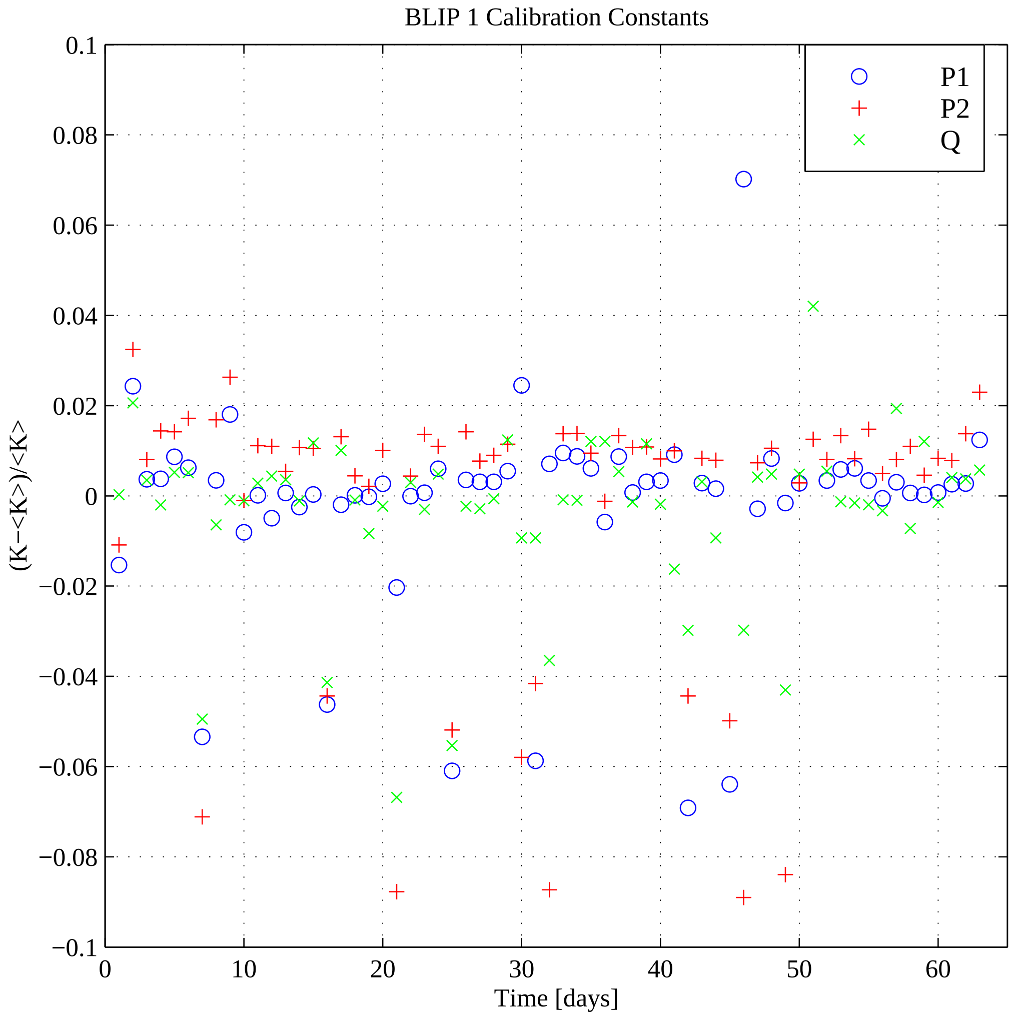


Figure A.4: Histograms summarizing the calibration of the file series 980414_1508, which had 0.43 live days of data. The top two sets of plots show the energy regions of the 10.4 keV gallium X-ray, and the 511 keV line after the calibration. The bottom row of plots shows the Y_R distribution of veto anticoincident data in the range 3 - 500 keV.

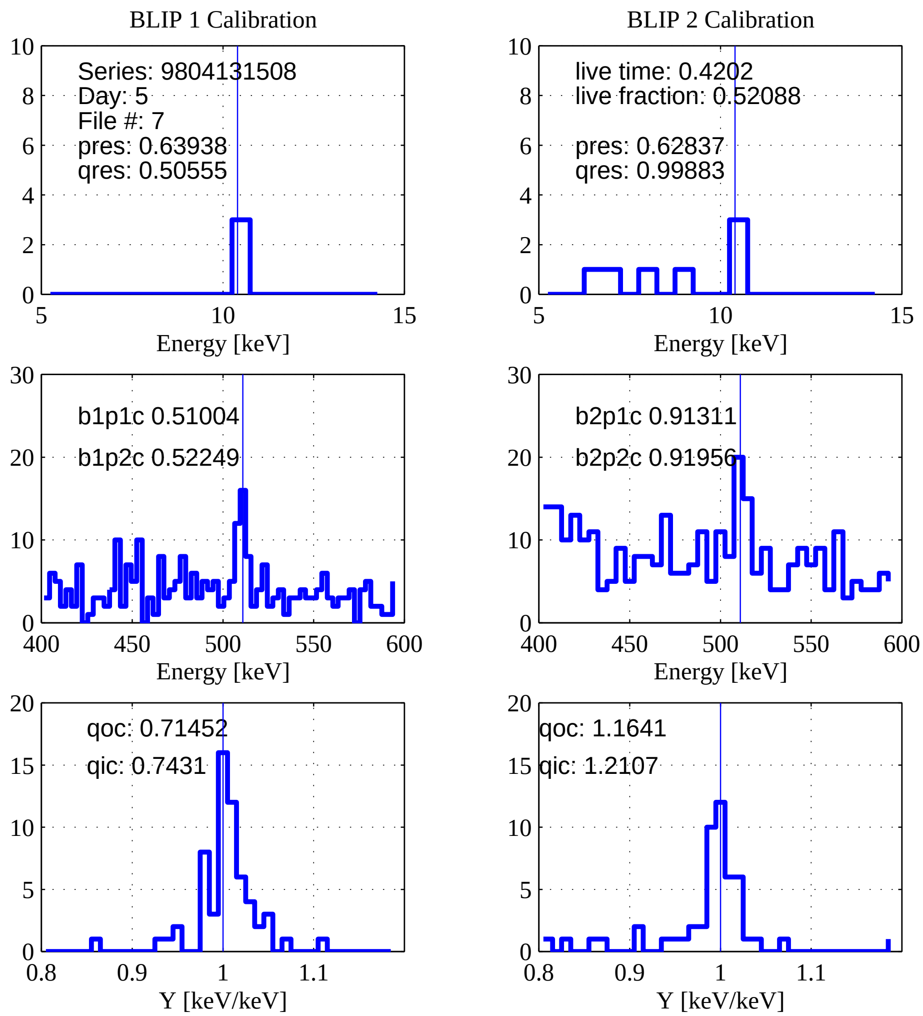
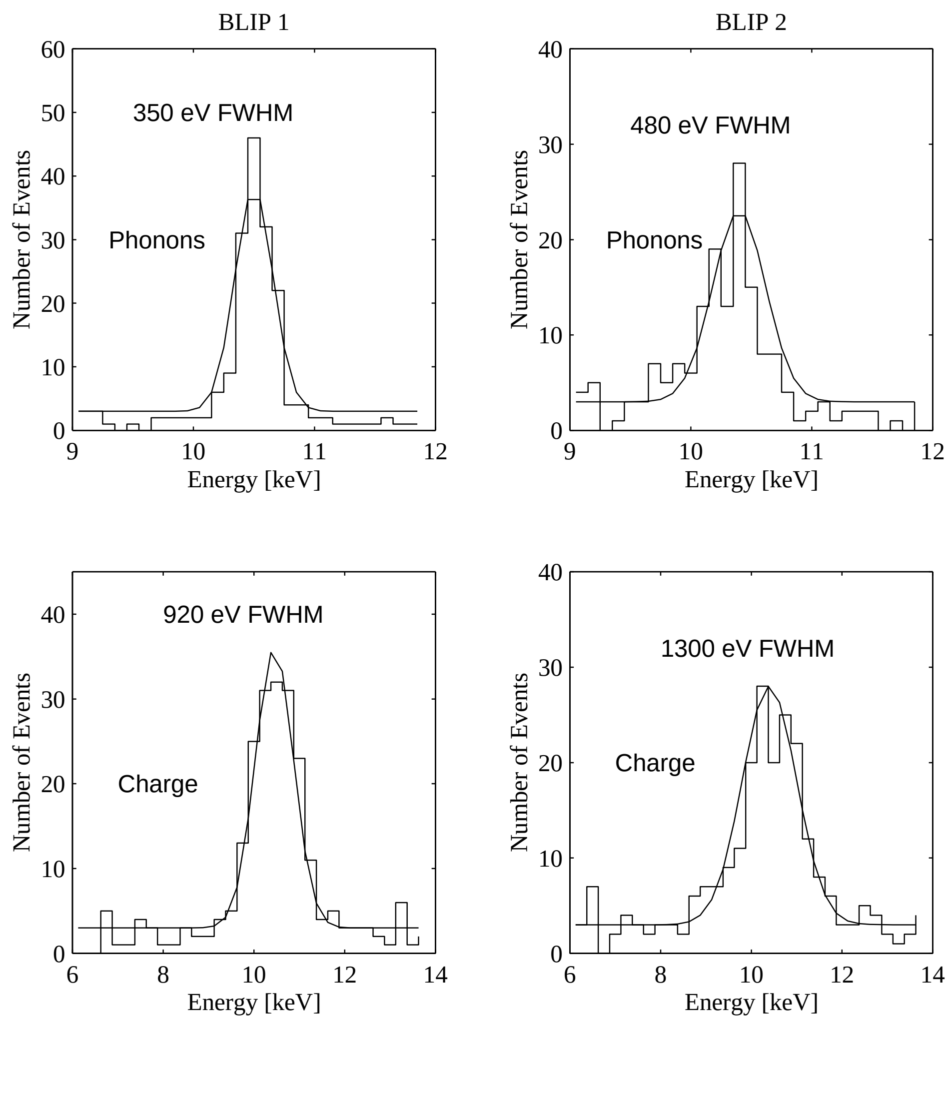


Figure A.5: Energy resolution of BLIPs. The 10.4 keV Ga-68 X-ray line is shown in phonon and charge channels for BLIPs 1 and 2 after file-by-file energy calibration based on the 511 keV line.



2. The bin with the largest number of counts, m is found in the ~100 keV region around where 511 keV should be.

3. To estimate the center of the peak, at bin m', use the weighted average

$$m' = \frac{(m+1)N_{m+1} + mN_m + (m-1)N_{m-1}}{N_{m+1} + N_m + N_{m-1}}$$

4. The calibration constant for this sensor is $K_{\text{phonon}} = 511\ \text{keV}/P_m$. This establishes the energy scale for the thermal measurement, $E_p = K_{\text{phonon}} \times P$. A 5 keV error in the determination of the peak location will produce a 1% error in the energy scale.

5. To determine the charge energy scale factor K_{charge}, from the charge amplitudes Q, histogram the quantity $z = E_p/Q$. This histogram has a peak at $z = K_{\text{charge}}$ corresponding to "gamma" events, which have an electron equivalent energy equal to their thermal energy.

This procedure was applied to each of the 87 files in the low background data set. The spectra for each file were inspected to make sure that the 511 keV peak and the 10.4 keV peak were correctly positioned in the calibrated data. Figure A.3 shows the time behavior of the resulting calibration constants. Figure A.5 shows that this calibration results in an energy resolution near the 10.4 keV line that is close to the limit imposed by baseline noise.

A.3 Luke effect corrections.

The discussion above of the energy scale determination ignores the effect of heat produced by charges drifting in the crystal, the "Luke effect". In fact, the heating of the crystals we measure has a contribution from both the energy deposited by the particle recoil and the energy produced by the drifting charges, which is neV_b for n electrons with a charge e drifting under the influence of the bias voltage V_b. At 20 mK in germanium, 3.0 eV of energy is required to create an electron hole pair. For the 2 volt charge bias common in Run 18, a 511 keV event actually heated the crystal by, in principle, $(1 + 2/3) * 511$ keV $= 852$ keV. There is some debate in the literature over whether all the interaction energy really shows up as heat or whether a fraction of it, perhaps 10%, gets stored up in long-lived deformations of the lattice [100]. Here we always assume that 100% of the energy becomes heat.

What we are really interested in knowing is the amount of energy deposited in the crystal by the scattering event, E_R, regardless of the amount of charge that's produced and drifted. This is related to the calibrated phonon energy E_p and charge energy E_q by the formula

$$E_R = E_p(1 + A) - AE_q \tag{A.3}$$

where $A = |V_b|/\epsilon$ and $\epsilon = 3.0$ V. Note that the Luke effect correction involves adding a factor AE_p to the energy as well as subtracting a factor AE_q. This is necessary because the data has been calibrated using photon lines that have already been boosted in energy by a factor $(1 + A)$. We need to add AE_p to correct the resulting error in the energy scale.

A.4 Summary of useful event parameters.

Its helpful to define a few other parameters for further analysis of events. The simplest is the "charge sum", the total charge collected in the inner and outer charge electrodes,

$$E_q = E_{\mathrm{qinner}} + E_{\mathrm{qouter}}.$$

The "charge yield", or ratio of collected charge to deposited thermal energy, is defined

$$Y = E_q / E_R.$$

This is the parameter that allows us to distinguish nuclear interactions, such as those of WIMPs, from electron interactions. It is often important to know the error in Y caused by errors in E_q and E_p. Using the standard method for propagating errors,

$$\sigma_Y^2 = \left(\frac{\partial Y}{\partial E_q}\right)^2 \sigma_q^2 + \left(\frac{\partial Y}{\partial E_p}\right)^2 \sigma_p^2$$

$$= \frac{1}{E_R^2}\left[\sigma_p^2 Y^2 (1+A)^2 + \sigma_q^2 (YA+1)^2\right].$$

Finally, it's very useful to know the partitioning of signal between the two NTD thermometers and the two charge electrodes, so we define "P partition" and "Q partition" variables,

$$P_{\mathrm{ppart}} = (E_{\mathrm{ntd1}} - E_{\mathrm{ntd2}}) \,/\, (E_{\mathrm{ntd1}} + E_{\mathrm{ntd2}})$$

and

$$Q_{\mathrm{qpart}} = \left(E_{\mathrm{qinner}} - E_{\mathrm{qouter}}\right) / \left(E_{\mathrm{qinner}} + E_{\mathrm{qouter}}\right)$$

An event depositing entirely in NTD #1 would have $P_{\rm part} = +1$ and an event in NTD #2 will have $P_{\rm part} = -1$ (in practice, all events cause some heating of both thermometers so $|P_{\rm part}|$ is never has high as this). Similarly, an event entirely in the inner charge region will have $Q_{\rm part} = 1$ and $Q_{\rm part} = -1$ for the outer region.

Appendix B

Cut definitions and efficiency estimates.

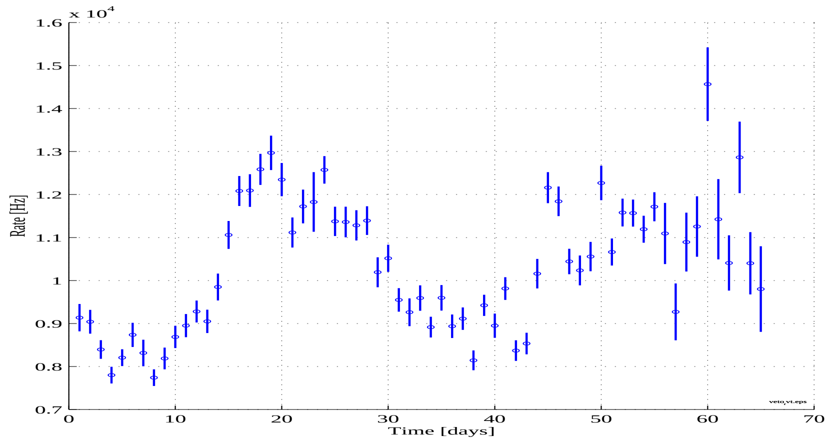
In this appendix, we define the basic cuts which are used in the BLIP analysis and show how we calculate the efficiency of these cuts for signal events.

B.1 Muon anticoincidence cut.

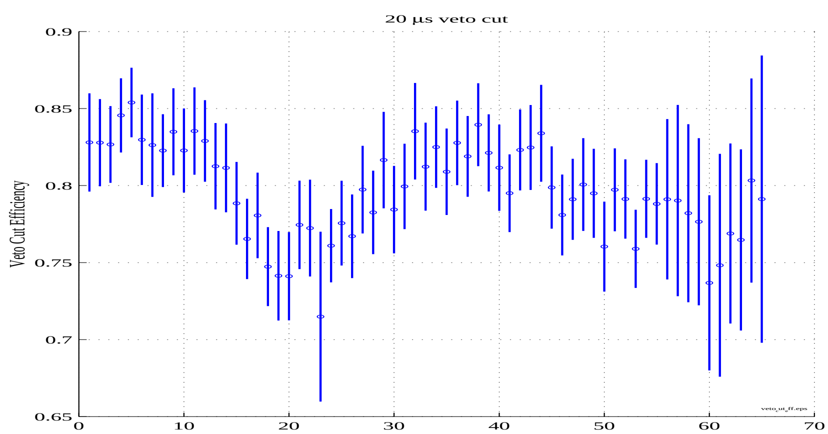
Muons passing through our shield make neutrons and other particles. To remove events due to the interactions of the muons and secondary particles in our detectors, we cut events which are coincident in the surrounding muon veto counters. The most important events to get rid of are those due to elastic neutron scattering of neutrons off Ge nuclei, since the resulting events look like WIMP interactions.

Neutrons produced in the copper and lead by muons have a half-life in the shield of $\sim$180 μs. Once the neutrons have scattered down to thermal energies, they are captured on a nucleus in the shielding materials or cryostat, usually producing

Figure B.1: Time variation of Run 18 muon veto trigger rate and cut efficiency.



(a) The veto trigger rate.



(b) Fraction of random trigger events accidentally coincident within 20 microseconds of a veto trigger.

prompt capture gamma rays. An important example is the 2.2 MeV gamma which comes from capture on the hydrogen in our polyethylene neutron moderator. This line was observed in the shield test measurements made with a relatively large (1 kg) conventional Ge detector and the events in the line followed the expected time distribution. Our cryogenic detectors have too low an efficiency for full containment of the 2.2 MeV photons to see this line, but we do see a small excess of continuum events with this time distribution. We usually don't attempt to veto these events, since they are a very small contribution to the nuclear recoil background rate.

Neutron scattering, of course, does produce a high recoil background. Neutrons are kinematically only allowed to transfer about 6% of their energy in a single collision and we only care about energy depositions over our dark matter detection threshold, which was 4 keV in Run 18. This implies that neutrons reduced in energy by collisions with H atoms in the moderator to below 60 keV are harmless. As shown in figure B.2, this occurs in much less than the 180 μs thermal capture time. This figure shows the time distribution after the most recent muon veto hit of neutron-like events arriving in B1 or B2 and FLIP. For the BLIP, imperfect separation between neutrons and electrons results in an electron background which has a flat time distribution. The FLIP cuts are cleaner, but the sample of neutrons is smaller and the time distribution is somewhat distorted by phonon propagation delays, since the FLIP is always triggered on the ballistic phonon signals. The BLIP data shown here is from charge triggers only. This data shows a reduction of about an order of magnitude in neutron recoil rate for every 5 μs we wait after the veto signal. There are no signs that the BLIP-triggered data has any time correlated neutrons arriving after more than 20 μs, which is the size of the time window we usually use for our anticoincidence cut.

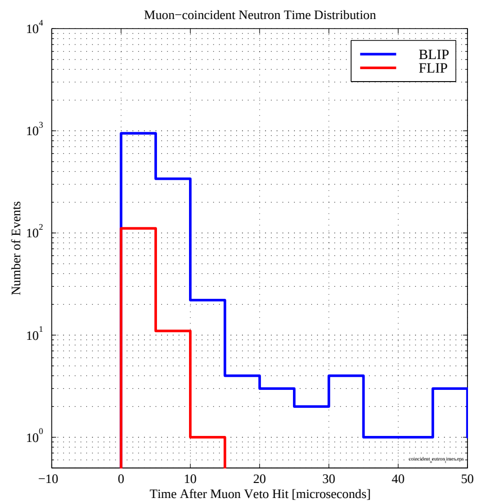


Figure B.2: Time distribution of neutrons relative to muon veto triggers. The BLIP events are all charge-triggered and therefore above about 20 keV in recoil energy. The tail of events in the BLIPs with times beyond 20 μs is due to contamination of the neutron sample by electrons.

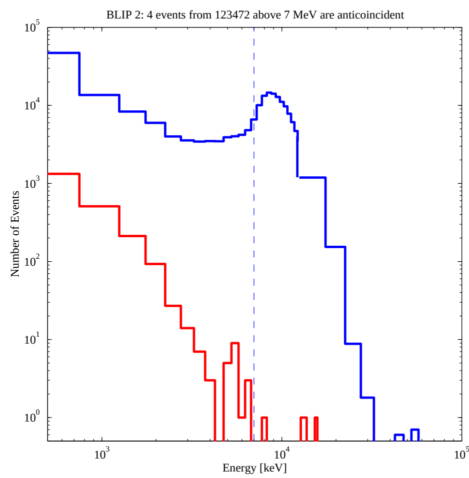


Figure B.3: Veto efficiency determined from high energy events in BLIP 2. After a 20 μs veto anticoincidence cut, 4 events are left above 7 MeV out of 123472 initial events. The peak centered at 9 MeV is due to muons with near-vertical trajectories that pass through the 1 cm thickness of the detector. There is an peak at 5.3 MeV from Po-210 alpha particles.

Muons enter the veto counters at a rate of perhaps 350 Hz, but the veto counters triggered much more often then this, at about 10 kHz in Run 18. This high trigger rate, which is the price we had to pay to get the highest possible efficiency for real muons, includes a large contribution from gamma rays and tube noise. As shown in figure B.1, there were significant time variations in the rate. These seem to correlate with humidity and temperature drifts in the tunnel, which presumably cause gain changes that allow dark current pulses to leak above our trigger threshold. In the future, the proportion of these spurious events will hopefully be reduced by repairs and recalibration of the veto (Preliminary Run 19 data shows that it has been reduced by a factor of 2). For Run 18 data, the high rate resulted in a 19.8% time-averaged probability for events to be accidentally coincident with veto triggers within a 20 μs window.

The efficiency of the veto cut for removing real muon coincident events can be estimated by applying the cut to a sample of high energy BLIP events due to direct muon interactions. The effect of this is shown in figure B.3. Above 7 MeV, the veto cut removes $>$99.993% of all events at a 90% confidence level. This is probably a good limit on the veto efficiency for events caused by muons passing through the central region of the shield and penetrating as far as the cryogenic detectors, but it is an open question whether the efficiency is so high for muons that interact in the corners of the shield or near the stems. We should not assume that this high a fraction of neutrons will be vetoed. A solid lower limit on the efficiency for cosmogenic neutrons comes from the Run 18 FLIP data, where 2 out of 131 nuclear recoil band events survive the veto cut, implying an efficiency $>$98%. Of course, the remaining events may be unvetoable electrons or WIMPs rather than neutrons.

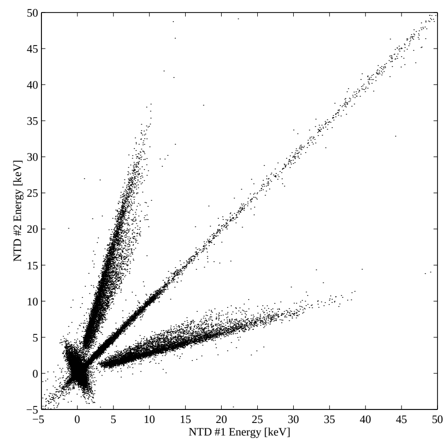
B.2 NTD direct hit cuts.

The BLIPs have two NTD thermistors because it was realized early on that interactions which occurred in or near the thermometers themselves might be easily confused with dark matter signals, since they would have deficient ionization relative to the size of the thermal signal. The hope was that direct thermometer hits could be recognized by a large asymmetry in the response of the two thermometers. As we will see, this turned out to work very well. It happens that direct hits also have significantly distorted pulse shapes because the ratio of thermometer heat capacity to thermal conductivity into the detector, which determines the fall time of the pulse, is much lower than the corresponding ratio for the bulk detector/ heat sink system. This results in very poor pulse fits when a template derived from events in the detector bulk is used to fit a NTD direct hit pulse. In fact, we could probably dispense entirely with the second thermometer and use the difference in pulse shape alone to reject these events. But for now, it is helpful to make use of this extra information if for no other reason that it helps us understand what types of events are being rejected by pulse shape (χ^2) cuts.

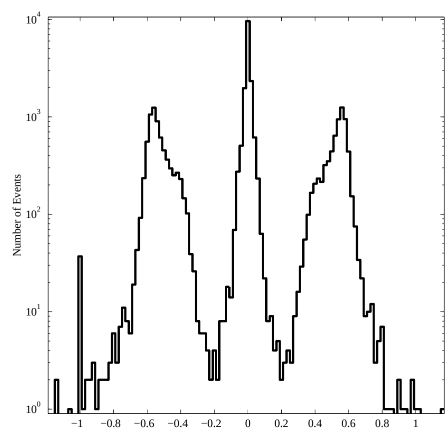
Figure B.4a shows a plot of $E_{\rm ntd1}$ vs. $E_{\rm ntd2}$ for BLIP 2 in the 0-100 keV region. At low energies, there is clearly a population of events with a large asymmetry in the response of the two thermometers. These events come from contamination of the NTD material by tritium, probably produced by reactions of the type Ge(n,tritium) during exposure to reactor neutrons. Since our calibration of the detector was made for events in the detector bulk, it is not surprising that the direct hit events appear to extend in energy far above the 18.6 keV tritium endpoint. In fact, experiments done with isolated pieces of NTD material show that these events are distributed in energy exactly as one expects for tritium. At energies above the tritium endpoint, there is a long, sparse tail of events which presumably includes the interactions of gammas rays and other particles originating outside the thermometers themselves as well as any other intrinsic NTD activity.

To define a cut that removed the direct hit events, it's useful to look at the distribution of $P_{\rm part}$ shown in Figure B.4b. There is a complete separation between events in the bulk, distributed around $P_{\rm part} = 0$ and events in the two thermometers at $P_{\rm part} \sim \pm 0.5$. Our normal "direct hit" cut removes all events with $|P_{\rm part}| > 0.2$. This cut removes a negligible fraction of normal events, even near the 4 keV energy threshold.

Figure B.4: Partitioning of thermal signals between the two NTD thermometers for veto anticoincident events at low energy in BLIP 2. The side lobes are due to tritium contamination of the NTD germanium.



(a) The "dagger plot" of NTD1 energy vs. NTD2 energy.



(b) A plot of Ppart for events between 3-50 keV.

B.3 Saturation cut.

The NTD temperatures are subject to slow drifts caused by thermal relaxation following various types of transient heating events that warm up the cryostat and detector mounting hardware. This includes shocks due to helium transfers, activity in the tunnel, and even boiling in the liquid nitrogen tank. These events result in slow drifts in the temperature that easily exceed 1 mK in amplitude, much larger than the μK changes associated with the particle interactions we want to measure. However, our amplifiers and digitizers only function in a limited dynamic range. To solve this problem, the measurement of the NTD resistances is AC coupled at the front end amplifier, with a baseline restoration time constant of about 10 seconds. This helps, but does not completely solve the problem of keeping the signal in a range where it can be properly recorded. The system has been designed to stay linear for signals that are within the range of our 12 bit digitizers, so the first sign that the dynamic range of the electronics has been exceeded is that the waveforms saturate at the minimum or maximum possible digital values, 0 or 4095.

Its possible to extract useful information from pulses that are partially saturated. For example, the rate of rise for a pulse that saturates before reaching its peak can be used to estimate the amplitude. However, it's often safer to just throw these events away and accept the appropriate efficiency penalties. To make it possible to discard events with saturating pulses, the "darkPipe" code informs us when a pulse has come within 2 ADC bins of the limits for more than 3 samples in a trace. For the Run 18 low energy analysis, we cut away all events which satisfy this condition in any of the sensor channels. This cut can not be used for events above about 1 MeV, since all these high energy events saturate.

During Run 18, the BLIP and FLIP LEDs were flashed at the beginning of grounding periods, which usually lasted 10 minutes and occurred every 25 minutes. The LEDs heat up the detectors and send all the thermistor channels into saturation. Often, they did not quite recover by the end of the grounding period. Another cause of saturation was "thermophonic" activity, usually caused by bubbling in the LN bath. We minimized this by pressurizing the bath during normal bias periods and releasing the pressure during grounding. This technique turned out to be surprisingly effective, but there were still occasional thermophonic periods, which often caused saturation.

The combination of these two effects resulted in the detectors being totally saturated (i.e. the whole phonon trace is off scale) about 4% of the time. A substantially larger fraction, 12% of random triggers, have some saturation in one of the detectors. Therefore, about 8% of events are "partially saturated", which can result either from accidental coincidences of random triggers with large events or from baseline wandering near the saturation threshold. In either case, a real event occurring at the same time as the random trigger would have been thrown away by the saturation cut, so the saturation cut does impose a 12% efficiency penalty. Many of the events removed in this way would also be removed by a pulse shape (χ^2) cut, so one must be careful not to directly add the two penalties.

B.4 Charge pileup.

If two events occur within a few ms of each other, they may produce a thermal pulse that looks exactly like the pulse from a single event, since the rise time of the thermal pulse is longer than the time difference. The charge for these near-coincident events will not be added together in general, so it is possible to make fake events with low Y this way. This is likely to occur with some frequency, since there are excited nuclear states produced, for example, by neutron capture, that decay within a few ms. Even without the existence of such states, about 1 in 30 events would be piled up within 10 ms of another event, assuming a 0.3 Hz trigger rate.

To help identify these events, we store the time of all charge triggers in the digital history buffer with $\sim 1\ \mu$s time resolution. Our usual charge pileup cut eliminates events with a second trigger within 8 ms before or 5 ms after the triggering pulse.

B.5 Pulse shape (χ^2) cuts.

As discussed briefly above in section A.1, the value of χ^2 for the template fits to the charge and thermometer pulses can be used to identify poorly shaped events. Unusual pulse shapes can be the result of

- Pileup of two real events.
- Pileup of a real event with a noise glitch.
- A pure noise event caused by an electronic or thermal transient or by electrical breakdown of the crystal.
- Crosstalk between detectors.
- An event with energy deposited somewhere other than in the bulk germanium detector. For example, in an NTD thermometer.

- Movement of the signal outside the linear range of the electronics ("saturation").

- Changes in the temperature of the thermal bath.

All of these distorting effects have been observed. The χ^2 test is very powerful because it is so non-specific. On the other hand, it can often be difficult to know what exactly one is throwing away when a χ^2 cut is applied and this can lead to problems in determining the efficiency for detection of the events we are interested in (See section B.6 below). Therefore, it is best to use the χ^2 cut "sparingly" in the sense that we try to get rid of bad events by every other possible means before considering what effect a χ^2 cut will have on the remaining data. Almost all of the events that are legitimately removed by a χ^2 cut can be gotten rid of in some other way.

In this spirit, consider Figure B.5, which shows phonon log χ^2 vs. energy distributions for events that have survived the following cuts,

- Veto anticoincident

- No NTD asymmetry.

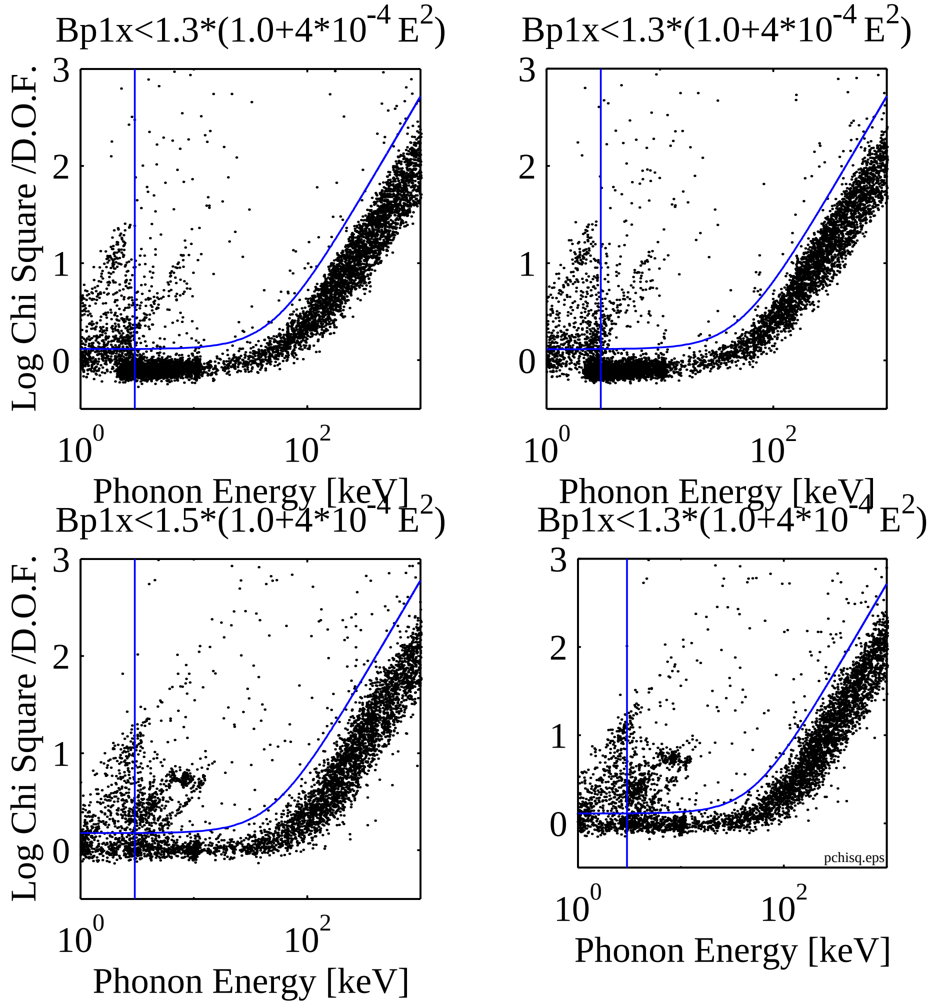
- No saturation of charge or thermal pulses.

- No charge pileup in the history buffer.

- No signal above noise in the FLIP detector.

These cuts result in relatively clean χ^2 distributions for the remaining events. Well-fit events form a band with $\chi^2 \approx 1$ (or $\log \chi^2 \approx 0$) at low energy, increasing at high

Figure B.5: Phonon Template Fit χ^2 distributions. The curved lines, which are described by the equations given at the top of each plot, define a cut. The vertical line shows the approximate trigger threshold.



energy. As discussed in section A.1, this increase seems to be due to errors in the templates we use to make the fit.

The lines in figure B.5 show our choice of χ^2 cut. Inspection of individual pulses shows that events with $log\ \chi^2$ well above this line have some defect of the sort listed above, while events below the line appear to be clean. There is some ambiguity in the choice of the cut threshold at a given energy which could be resolved in the future by calculating the χ^2 distribution for Monte-Carlo events convolved with the measured noise in our system.

B.6 Estimation of efficiency.

The goal of our experiment is to calculate the rate of some rare events as a function of energy. To calculate a rate, we must know the number of events that have been seen in the detector as a function of energy, $N(E)$, the time we were looking for events, T_L (the "live time"), and the probability that an event would have been properly counted if it had occurred, $\xi(E)$, which we call the "efficiency". Then, the rate of events can be written

$$R(E) \equiv \left(\frac{1}{\xi(E)T_L}\right)\frac{dN}{dE}$$

The data acquisition system tells us on an event-by-event basis how long it was armed and waiting between triggers with a precision of 1 ms. Since the event rate is less than 1 Hz, the sum of these times should give us T_L to better than 0.1% if there are truly no biases in the recorded times bigger than the time resolution of the clock. We always assume that this is the case. For Run 18, the live time turned out to be about 80% of the elapsed real time, not counting inactive periods due to maintenance, cryogen fills, or crystal neutralization. If one counts these periods, the experiment was live for 33.0 days out of 64.4 or 51% of the time.

The best way to find the function $\xi(E)$ would be to use imitation pulses of varying energy injected into the data throughout the run. In that case, $\xi(E)$ would just be the fraction of pulses at each energy that pass all the cuts. In the future, we should have the capability to do this, but for Run 18 the technology was not ready, so we have to resort to less direct methods. One possible approach is to assume that the fraction of valid events thrown away at a given energy by each cut is nearly the same as the fraction of all events that are thrown away. This turns out to be a very bad assumption for low energy (0-100 keV) BLIP data for two reasons: First, there are a lot of low energy events due to noise and NTD direct hits which are very efficiently eliminated by cuts that would pass a large fraction of valid events. Second, badly reconstructed high energy events tend to be shifted down into this energy region. As a result, the fraction of "low energy events" failing saturation and pulse shape cuts has little relationship to the fraction of valid events that would survive.

A better approach is to enumerate the possible causes of an event's failure to appear in the spectrum after cuts and then estimate the probability that one of the failure modes will apply. Unfortunately, it's difficult to do this as a function of the event energy, so we assume that the efficiency has little energy dependence. This is probably true except very close to the noise threshold.

The inefficiency $1 - \xi$ should have contributions from the following possible causes of failure to count an event:

1. One of the detectors was in saturation. Random triggers show saturation 12% of the time, which includes a 8% contribution from pileup-related saturation.

2. An accidental coincidence with a muon veto hit occurred. Of random trigger events, 19.8% pass our standard 20 μs veto cut.

3. An accidental coincidence occurred with a real or noise event in another detector, leading to rejection by the charge trigger pileup cut or a simple energy threshold cut (usually phonon energy <3 keV, which includes about 10 σ of the Gaussian noise tail). This occurs 5.3% of the time in the random trigger set.

4. An accidental coincidence occurred with a real or noise event in the same detector, leading to rejection by either the (a) saturation, (b) charge trigger pileup, or (c) pulse shape cuts. This should happen about 2.6% of the time.

It is incorrect to simply sum these contributions to the inefficiency, since, for example, some of the saturation events are veto coincident. What one really wants to know is the joint probability P(1|2|3|4a|4b|4c). A big fraction of this, P(1|2|3|4a|4b), can be calculated from the random trigger data and is equal to 29.8% for Run 18. The ease with which this number can be obtained is the main advantage of this approach.

The probability that 4c will apply can't be directly estimated from random triggers, since we don't know precisely what effect the pileup will have on the χ^2 of a real pulse. It is clear, however, that the contribution of 4c is small, since very few of the ~2.6% of events with same-detector pileup will pass the other cuts. A good way to calculate it would be to compute χ^2 values for pulse templates of various energies added to real random trigger traces and see how many pass cuts (i.e. we are asking what would have happened had an ideal pulse occurred at the same time as a random trigger). This hasn't been done yet.

Another way to estimate P(4c) is to use the assumption that any event with a phonon fit amplitude above a few keV in the random trigger set would have distorted a real event occurring at the time of the random trigger enough to cause χ^2 to go out of bounds. This assumption is probably valid in the low-energy region, where the pileup pulses always have a significant fraction of the amplitude of real pulses. Using this assumption, we find P(1|2|3|4a|4b|4c)-P(1|2|3|4a|4b)=0.2% for pileup pulses above 3 keV. Our best estimate of the total efficiency of the cuts is 1-0.298-0.002=0.700.

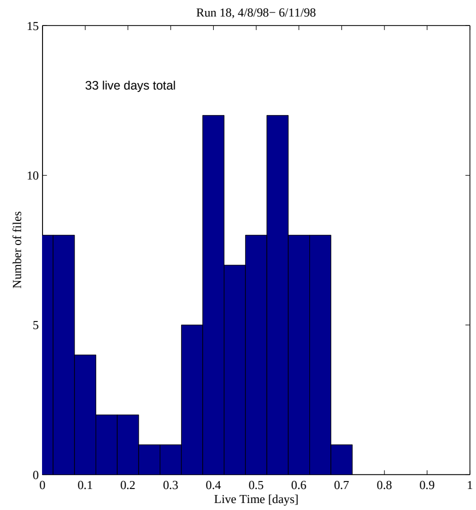
Appendix C

Selection of good Run 18 data.

C.1 What data was collected?

This note describes mainly the BLIP data from the period of "low background" data collection between April 8 and June 11 of 1998. This is the highest quality data from Run 18 because it was taken after improvements in the veto system which greatly decreased the fraction of undetected muons and unvetoed neutrons. There were 87 file series of data (87 files after merging) taken during this period of 64.4 real days. Adding the live times recorded by the data acquisition system on an event by event basis yields 33.0 live days. Figure C.1 shows the distribution of live times, which is peaked at about 0.5 live days per file. There are quite a number of files with low live time due to runs which were aborted because of problems such as excessive noise or triggering. The data from these files is not necessarily unusable, since appropriate cuts can remove noise events with high efficiency. However, it probably makes sense to eliminate the most pathological files to impose uniformity on the data, provided the cost in lost detector exposure is not too great. One

Figure C.1: Distribution of live times in 87 Run 18 files.



serious problem with the short files is that the energy calibration is uncertain. A list of "good" files is presented below in section C after the tools for making the selection are described.

The noise in both the phonon and charge channels varied quite a bit during this run. This resulted in dramatic changes in the trigger rates which were often compensated for by changing the trigger thresholds slightly. For some files, the trigger rates and noise are many times higher than the average. In a few cases, the event rates are inexplicably lower than they should have been. I found it useful to impose a measure of homogeneity on the data sets by making the following cuts on quantities averaged within each of the 87 files:

- live time $>$ 0.2 live days.
- phonon trigger rates $<$ 1500 / day.
- charge trigger rates $<$ 30,000/ day.

- Between 550/day and 950/day 20-100 keV events passing phonon χ^2 and direct hit cuts.
- phonon resolution $\sigma <$ 750 eV.
- charge resolution $\sigma <$ 750 eV.

The rates above are for each detector separately and are calculated using the sum of on-line live times. These cuts are quite gentle, but they eliminate some really pathological periods of data collection. The cuts yield the following lists of "good" files.

For BLIP 1 (57 files):

9804082355 9804092008 9804101301 9804111932 9804121923 9804131508 9804161258 9804171126 9804181519 9804191758 9804201155 9804211151 9804221159 9804241405 9804251503 9804261746 9804271330 9804281755 9804291726 9805011917 9805021814 9805032027 9805041247 9805051618 9805061154 9805081558 9805091455 9805101441 9805111412 9805121604 9805131644 9805141424 9805151601 9805161456 9805171446 9805181754 9805191257 9805201437 9805211647 9805221634 9805230918 9805241011 9805250929 9805281831 9805291526 9805301717 9805311733 9806021824 9806032119 9806041828 9806051518 9806062240 9806072107 9806082357 9806091443 9806101455 9806111923

For BLIP 2 (51 files):

9804101301 9804161258 9804171126 9804181519 9804191758 9804211151 9804221159 9804231852 9804241405 9804251503 9804261746 9804271330 9804281755 9804291726 9805021814 9805032027 9805041247 9805051618 9805061154 9805071729 9805081558 9805091455 9805101441 9805111412 9805121604 9805131644 9805141424 9805151601 9805161456 9805171446 9805181754 9805191257 9805201437 9805211647 9805221634

9805230918 9805241011 9805250929 9805261603 9805271834 9805281831 9805291526 9805311733 9806011726 9806051518 9806062240 9806072107 9806082357 9806091443 9806101455 9806111923

The sums of live times are 28.9 live days for BLIP 1 and 26.5 live days for BLIP 2. The cuts remove 12% of the BLIP 1 live time and 20% of the BLIP 2 live time. Fewer BLIP 2 files are acceptable because of a week-long period at the beginning of the run when we were experimenting with a new amplifier configuration which resulted in quite poor charge resolution. For most purposes, it is easiest to ignore the BLIP 2 data from this period before April 16.

C.2 Summary statistics.

It's interesting to look at histograms of summary statistics calculated by averaging over all events in each file. Let's define a few such statistics:

- RQTRIG - The rate of charge triggers [/(detector live-day)].
- RPTRIG - The rate of phonon triggers.
- RGOOD20 - The rate of 2-20 keV events passing phonon χ^2 and asymmetry cuts.
- RGOOD80 - The rate of 20-100 keV events passing phonon χ^2 and asymmetry cuts.
- RNV20 - The rate of 2-20 keV veto anticoincident (vtpretimefast<-20) events.
- RNV80 - The rate of 20-100 keV veto anticoincident events.

- RLY20 - The rate of 2-20 keV low Y (Charge Energy/Phonon Energy$<$0.8) events.

- RLY80 - The rate of 20-100 keV low Y events.

Figures C.2 and C.3 show the time dependence of these quantities for the files passing the cuts we have defined.

Figure C.2: B1 Rates.

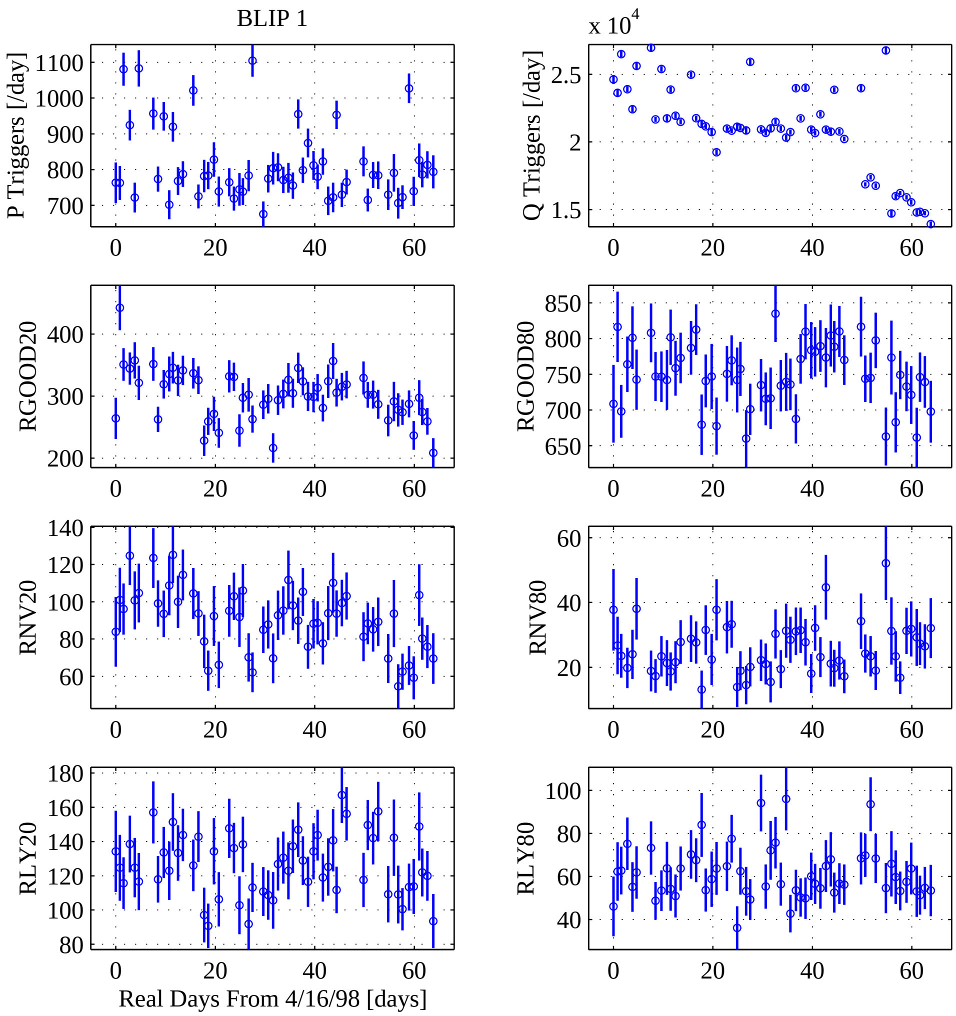
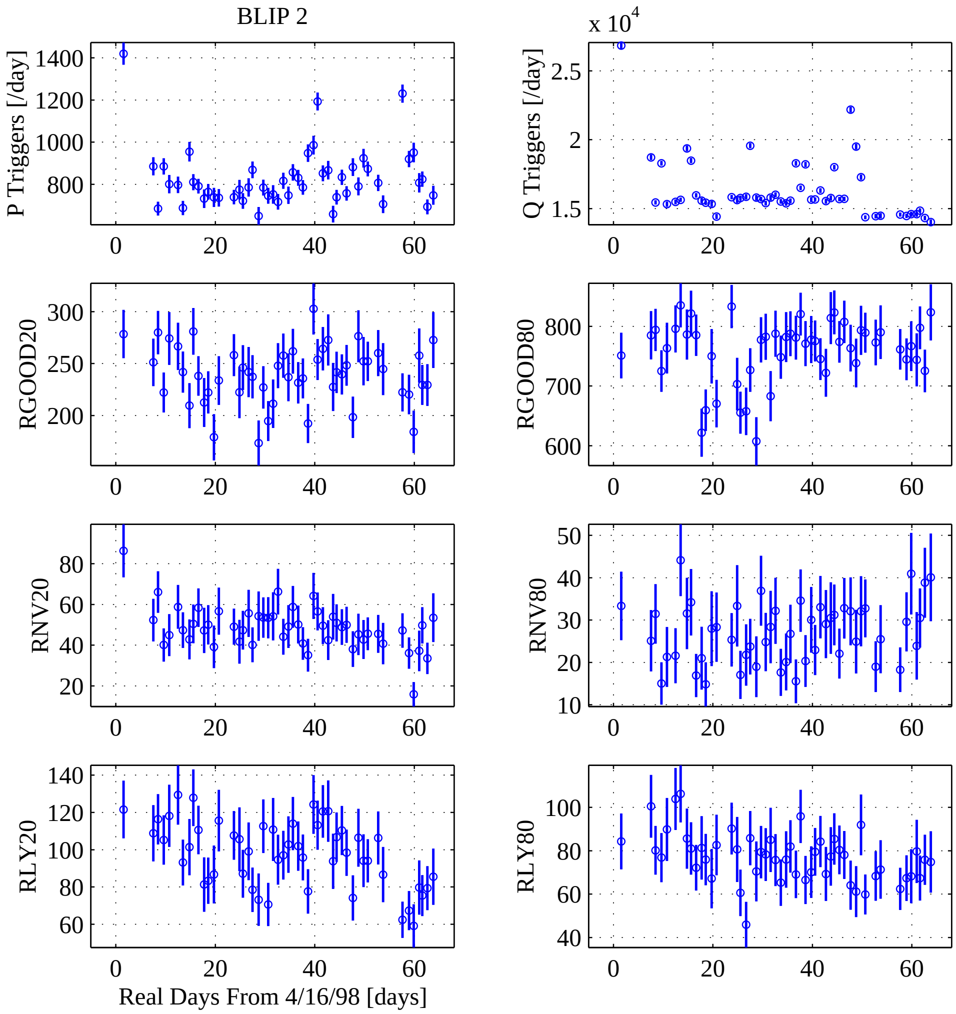


Figure C.3: B2 Rates.



Bibliography

[1] J.R. Primack in *Formation of Structure in the Universe*, Proceedings of the Jerusalem Winter School, 1996. Ed. A. Dekel and J.P. Ostriker, (Cambridge, 1996).

[2] P. Coles, and G.F.R Ellis, *Is the Universe Open or Closed?*, (Cambridge, UK, 1997).

[3] V. Trimble, *Annu. Rev. Astron. Astrophys.*, **25** (1987).

[4] A. Dekel, D. Burstein, S.D.M White, in *Critical Dialogues in Cosmology*, ed. N. Turok (World Scientific, 1996).

[5] P.J.E. Peebles, *Principles of Physical Cosmology*, (Princeton, NJ, 1993).

[6] E. Kolb, and M. Turner, *The Early universe*, (Addison-Wesley, 1990).

[7] R. Sancisi, and T.S. van Albada in *Dark Matter in the Universe*, Ed. J. Kormendy and G.R. Knapp, (Dordrecht, Holland, 1987).

[8] P. Salucci, and M. Persic, in *Dark and Visible Matter in Galaxies*, Proceedings of the Sesto DM 1996 Conference, Eds. M. Persic and P. Salucci, Astro-ph/9703027.

[9] M. Fich and S. Tremaine, *Annu. Rev. Astron. Astrophys.* **29** 45 (1991).

[10] N.A. Bahcall, L.M. Lubin, and V. Dorman, *Ap. J.* **447**, L81-L85 (1995).

[11] G. Gilmore, in Proceedings of the First International Workshop on the Identification of Dark Matter (IDM 96), Sheffield, UK, 1996. Ed. N.J.C. Spooner (World Scientific, 1997).

[12] R.G. Carlberg *et al., Ap. J.* **462** 32 (1996).

[13] A. Cavaliere and R. Fusco-Fumiano, *Astron. Astroph.* **70** 677 (1978).

[14] A.E. Evrard, A. Metzler, and J.F. Navarro, *Ap. J.* **469** 494 (‘996).

[15] L.P. David, C. Jones, C., and W. Forman *Ap. J.* **445** 578 (1995).

[16] P. Fischer, and T. Tyson, *Astron. J.* **113** 521 (1997).

[17] T. Tyson in SLAC Summer Institute proceedings, 1998.

[18] I. Smail *et al.*, *Ap. J.* **479** 70 (1996).

[19] S.D.M. White *et. al.*. *Nature* **366** 429 (1993).

[20] D.A. White, and A.C. Fabian, *Mon. Not. R. Astron. Soc.* **273** 72 (1995).

[21] J.E. Carlstrom *et al.*, Astroph/9905255.

[22] J.J. Mohr *et al., Ap. J.* **517** (1999). Astroph/ 9901281.

[23] M.A. Strauss and J.A. Willick, *Phys. Rep.* **261** 271 (1995).

[24] J.A. Willick *et al., Ap. J.,* **486** 629 (1997).

[25] J. Einasto in *Dark Matter in the Universe*, Ed. P. Galeotti and D.N. Schramm, (Kluwer, 1990) 195.

[26] D. Zaritsky, R. Smith, C. Frenk, and S.D.M. White, *Ap. J.*, **405**, 464 (1993).

[27] A. Dekel, *Annu. Rev. Astron. Astroph.* **32**, (1994).

[28] A. Dekel, and M.J. Rees, *Ap. J.*, **422**:L1-L4 (1994).

[29] S. Perlmutter *et al.*, Astroph/9812133.

[30] A. Filippenko and A.G. Reiss in Proceedings of the 3rd International Conference on the Sources and Detection of Dark Matter in the Universe (DM98), Ed. D. Cline. Astro-ph/9807008.

[31] S. Burles *et al.*, *Phys. Rev. Lett.* **82** 4176 (1999). Astroph/9901157.

[32] K. Olive *et al., Ap. J.* **483** 788 (1997).

[33] O. Gerhard, and J. Silk *Ap. J* **472** 34 (1996).

[34] B. Carr, *Annu. Rev. Astron. Astrophys.* **32** 531 (1994).

[35] M. Turner in Proceedings of the 3rd Stromlo Symposium, 1998, (Eds. B.K. Gibson, T.S. Axelrod, and M.E. Putnam, 1998. Astroph/ 9811454.

[36] W. Sutherland *et al.*, Proceedings of the 1st International Conference on the Identification of Dark Matter, Sheffield, UK, September 1996, (Ed. N.J.C. Spooner).

[37] E. Gates, G. Gyuk, and M. Turner, Astroph/9411073.

[38] K. Freese, in Proceedings of the Workshop on Aspects of Dark Matter and Particle Physics in Heidelberg, Germany, July 1998, Astroph/9901178.

[39] D.O. Caldwell, in Proceedings of the 2nd International Conference on the Identification of Dark Matter, Buxton, UK, September 1998. (Ed. N.J.C. Spooner), Astroph/9812026.

[40] Y. Fukuda *et al.* (The Super-Kamiokande collaboration), *Phys. Rev. Lett.*, **82** 1810 (1999).

[41] S. Yellin, Proceedings of the International Workshop on Particle Physics and the Early Universe (COSMO 98), Monterey, CA, 1998, (Ed. D. O. Caldwell).

[42] J. Primack, J. Holtzman, A. Klypin, and D.O. Caldwell, *Phys. Rev. Lett.* **74** 2160 (1995).

[43] B.W. Lee and S. Weinberg, *Phys. Rev. Lett.* **39** 165 (1977).

[44] P. Jetzer, Astroph/9901058.

[45] J. Engel, S. Pittel, and P. Vogel *Int. Journal. Mod. Phys.* Vol 1., No. 1 (1992) 1-37

[46] Ellis, J. in Proceedings of the International Workshop on Particle Physics and the Early Universe (COSMO 98), Monterey, CA, 1998, Ed. D. O. Caldwell.

[47] The particle data book, C. Caso *et al.*, *European Physical Journal* **C3**, 1 (1998).

[48] A.K. Drukier, K. Freese, and D.N. Spergel, *Phys. Rev.* **D33**, 3495 (1996). K. Freese, J. Frieman, A. Gould, *Phys. Rev.* **D37** 3388 1988.

[49] D. Spergel, *Phys. Rev.* **D37**, 1353 (1988).

[50] J. Engel, *Phys. Lett.* **B264** 1 115 (1991).

[51] G. Jungman, M. Kamionkowski, K. Griest, *Phys. Rep.* **267**, 195 (1996).

[52] K. Griest, *Phys. Rev.* **D38** 2357 (1988).

[53] J.I Collar and F.T. Avignone, *Phys. Lett.* **B275** (1992) 181-185. J.I. Collar and F.T. Avignone, *Phys. Rev.* **D47** 5238. E. Garcia *et al.*, *Phys. Rev.* **D 51** 1458 (1995).

[54] N.J.C. Spooner *et al.*, *Nucl. Phys.* **B** (Proc. Suppl.) **48** 64 (1996).

[55] R. Bernabei *et al.*, Phys. Lett. **B 389**, (1996) 757-766. R. Barnabei *et al.*, Phys. Lett. **B424** (1998) 195-201. R. Bernabei *et al.*, Preprint ROM2F/98/34.

[56] D.O. Caldwell, R.M. Eisberg, D.M. Grumm, M.S. Witherell, B. Sadoulet, F.S. Goulding, and A.R. Smith, *Phys. Rev. Lett.* **61** (1988) 510.

[57] M. Grey, Ph.D. thesis, University of California, Santa Barbara, 1994.

[58] S. Yellin, CDMS note 9708004, August 26, 1997.

[59] R.J. Gaitskell, P.D. Barnes, Jr., A. Da Silva, B. Sadoulet, and T. Shutt, *Nucl. Phys.* B Proc. Supp., 1996.

[60] J. I. Collar, Ph.D. Thesis, University of South Carolina, 1992.

[61] J.D. Lewin and P.F. Smith, RAL-TR-024.

[62] M.W. Goodman, E. Witten, *Phys. Rev.* **D31** 12 3059 (1985).

[63] A. Drukier, L. Stodolsky, *Phys. Rev.* **D30** 11 2295 (1984).

[64] P.F. Smith, and J.D. Lewin, *Phys. Rep.* **187**:5 203 (1990).

[65] M. Beck *et al. Phys. Lett.* **B336**, 141-146 (1994). L. Baudis *et al.*, *Phys. Rev.* D, 1999.

[66] D.O. Caldwell *et al.*, *Phys. Rev. Lett.* **65** (1990) 1305.

[67] D.P. Snowden-Ifft, C.J. Martoff, and J.M. Burwell, Astro-ph/9904064.

[68] The GENIUS experiment, Hep-ex/9811040.

[69] O.V. Lounasmaa, "Experimental Principles and Methods Below 1 K", (New York: Academic Press,1974).

[70] P.D. Barnes Jr., Ph.D. Thesis, University of California, Berkeley, 1996.

[71] W.K. Stockwell, Ph.D. Thesis, University of California, Berkeley, 1996.

[72] T. Shutt, Ph.D. Thesis, University of California, Berkeley, 1993.

[73] N. Wang, Ph.D. Thesis, University of California, Berkeley, 1991.

[74] A. Cummings, Ph.D. Thesis, University of California, Berkeley, 1997.

[75] A. Da Silva, Ph.D. Thesis, University of British Columbia, 1996.

[76] A. Da Silva et al., NIM A 354 (1995) 553-559.

[77] R. M. Clarke, Ph.D. Thesis, Stanford University, 1999.

[78] S.W. Nam, Ph.D. Thesis, Stanford University, 1999.

[79] N. Wang et al. *Phys. Rev.* **B41** 3761 (1990) .

[80] E. Aubourg et al. *J. Low Temp. Phys.* **93** 289 (1993) .

[81] T. Shutt et al., *Phys. Rev. Lett.*, **69** 3531 (1992).

[82] T. Shutt et al., *Phys. Rev. Lett.*, **69** 3531 (1992).

[83] T. Shutt *et al.*, Proceedings of the 18th Texas Symposium on Relativistic Astrophysics, 16-20 December, 1996, Chicago IL.

[84] T. Shutt et al., Proceedings of the VIIth International Workshop on Low Temperature Detectors, 27 July - 2 August, 1997, Munich, Germany (Ed. S. Cooper) 224-226 (1997).

[85] T. Shutt, "E5 Design", CDMS internal note, June 23, 1994.

[86] E. E. Haller, *J. Appl. Phys.* **77:**7 1 (1995).

[87] G. Kiel, *NIM* **89** 111 (1970).

[88] G. Heusser, *Annu. Rev. Nucl. Part. Sci.* **45**:543-90 (1995).

[89] S. Yellin, personal communication.

[90] S. Eichblatt, CDMS Note 9802001.

[91] S. Okada and N. Momoshima, *Heath Phys.* **65** 595 (1993).

[92] M.J. Wood et al, *Health Phys.*, **65** 610 (1993).

[93] P.N. Luke, *J. Appl. Phys.* **64** 6858 (1988).

[94] F. Avignone *et al.*, *Nucl. Phys.* **B** (Proc. Suppl.) *28A* (1992) 280.

[95] J. Lindhard, V. Nielsen, M. Scharff, and P.V. Thomsen, *Mat. Fys. Medd. Dan. Vid. Selsk.*, 33 (1963) 1.

[96] E. Lorch, *Int. J. Appl. Rad. Isotop.*, 1973, **24** 585.

[97] T. Perera, personal communication.

[98] T. Perera, and D. Akerib, CDMS Note 9808001.

[99] R. Gaitskell CDMS note 9805001.

[100] Alessandrello *et al., Phys. Lett.* **B 384** (1996).