Title
Techniques for Inelastic Effective Field Theory Measurements with the Large Underground Xenon Experiment

Author
Hogan, Daniel Patrick

Publication Date
2018

Peer reviewed|Thesis/dissertation

Techniques for Inelastic Effective Field Theory Measurements with the Large Underground Xenon Experiment

by

Daniel Patrick Hogan

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Robert Jacobsen, Chair
Professor Marjorie Shapiro
Professor Kai Vetter

Summer 2018

Techniques for Inelastic Effective Field Theory Measurements with the Large Underground Xenon Experiment

Daniel Patrick Hogan

Abstract

Techniques for Inelastic Effective Field Theory Measurements with the Large Underground Xenon Experiment

by

Daniel Patrick Hogan

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Robert Jacobsen, Chair

Cosmological evidence indicates that nonbaryonic dark matter makes up a quarter of the energy density of the universe. One hypothesis for the particle nature of dark matter is the weakly-interacting massive particle (WIMP). The Large Underground Xenon (LUX) experiment is a dual-phase xenon WIMP search experiment with a 250kg active volume. Computational tools developed to support LUX analyses include data mirroring and a data visualization web portal.

Within the LUX detector, particle interactions produce pulses of scintillation light. A pulse shape discrimination (PSD) technique is shown to help classify interaction events into nuclear recoils and electron recoils based on the time-structure of the pulses. This approach is evaluated in the context of setting limits on inelastic effective field theory (IEFT) dark matter models. Although PSD is not found to provide significant improvement in the limits, LUX is nevertheless able to set world-leading limits on some IEFT models, while limits for other IEFT models are reported here for the first time.

Felix qvi potvit rervm cognoscere cavsas

–Virgil

Contents

List of Figures

List of Tables

Acknowledgments

I am grateful to my fellow graduate students who helped me throughout my years at Berkeley. I am also grateful to Carlos Faham, who was my guide to the world of LUX.

Most of all I am grateful to my family for all of their love and support. This dissertation is dedicated to them.

Introduction

Concurrent lines of cosmological evidence point to invisible, nonbaryonic dark matter as making up a quarter of the energy density of the universe. Supporting observations come from targets as diverse as supernovae, the cosmic microwave background, galactic structure, and gravitational lensing. Understanding the particle nature of the dark matter is a daunting task. Out of an almost limitless range of plausible hypotheses, many have appealing theoretical features. The theory of the weakly-interacting massive particle (WIMP), for instance, first garnered attention because of a seemingly-natural production mechanism. The most straightforward models for WIMP-nucleon scattering come directly from the physics of neutral weak scattering. But in recent years, a new focus has been placed on surveying much broader families of interaction models.

A key player in the field of WIMP dark matter direct detection is the Large Underground Xenon (LUX) experiment. LUX is a dual-phase xenon detector with a 250kg active volume. Energy deposition in the liquid xenon produces scintillation light as well as free electrons, which in turn produce electroluminescent light in the gas overhead. The light is detected by photomultiplier tubes. A key capability of LUX is injecting radioactive sources into the xenon itself for calibration purposes. LUX set world-leading limits on spin-independent dark matter in 2013 and again in 2016.

Effectively managing LUX's half petabyte of raw and processed experimental data requires special attention to computational infrastructure. The data is safeguarded and made more accessible for analysis by mirroring the full dataset from a primary storage location to an Energy Department computer center. A data visualization web portal allows collaborators worldwide to better inform their analyses through exploratory interaction with the measured data and low-level processing output.

Inside the detector, energy depositions in the xenon are of two types: electron recoils and nuclear recoils. Since electron recoils are a major background to the nuclear recoils of WIMP-nucleon scattering, effectively distinguishing the two populations is of critical importance. While using the relative sizes of the scintillation and electroluminescent pulses is a fairly effective method for doings so, the classification can be strengthened by looking at the shape of the scintillation pulse itself. In this pulse shape discrimination (PSD) approach, a figure of merit based on prompt fraction is used to characterize the shape of a scintillation pulse. Then, using calibration data, an empirical model of pulse shape is developed for electron and nuclear recoils as a function of pulse size and depth. Alternate methods for calculating

the figure of merit are also considered.

To provide a testbed for the PSD work while simultaneously investigating a novel theoretical approach to dark matter, a study is made of inelastic effective field theory (IEFT) dark matter. This framework subsumes almost all of the leading WIMP-nucleon interaction models (including elastic effective field theory) as limiting cases, and it also produces other hypothetical models that are entirely new. A generalized version of LUX's profile likelihood ratio limit-setting code is developed, incorporating PSD and arbitrary IEFT nuclear recoil spectra. With this, new limits are set on a number of previously-studied dark matter models, as well as other models that are investigated here for the first time.

Although PSD ultimately does not provide a significant improvement on limits in the low-energy range that has been the subject of LUX's previous full analysis, the tools developed here may yet be useful in the ongoing extension of that analysis to higher energy. In any case, the exploration of IEFT shows the power of LUX to continue to set world-leading limits on a wide range of dark matter models.

Chapter 1

Physics of Dark Matter

1.1 Cosmological Motivation for Dark Matter

1.1.1 Early Clues

The first clue to the existence of dark matter came more than eighty years ago from examining the motion of galaxies. Then in the 1970's, rigorous measurements of the motion of stars within galaxies would bring the dark matter theory to new prominence. By the end of the 20th century, dark matter had become central to an expansive new understanding of cosmology.

Galaxy Clusters

The first major experimental evidence for large quantities of non-luminous matter in the universe came in 1933 from observing the motion of galaxies within a galaxy cluster. Zwicky[145] studied the Coma Cluster, a cluster of some 800 galaxies.[1] What he found was an apparent discrepancy among the astronomical data.

For a bound state of interacting point particles in equilibrium, given a two-body interaction potential following a power law $U(r) \propto r^{-1}$, the virial theorem reduces to

$$E_k = -\frac{1}{2}E_U \tag{1.1}$$

where E_k is the total kinetic energy and E_U the total potential energy. A gravitationally-bound galaxy cluster is just such a system. Approximating a galaxy cluster as a sphere of uniform mass, its potential energy is

$$E_U = -\frac{3}{5}\frac{GM^2}{R} \tag{1.2}$$

[1] As estimated in 1933; current estimates for the number of galaxies are greater.

where G is the gravitational constant, M the cluster's mass, and R its radius. Furthermore, total kinetic energy is proportional to the average of the square of velocity, as

$$E_k = \frac{1}{2} M \overline{v^2} \,. \tag{1.3}$$

Equations 1.1, 1.2 and 1.3 together estimate cluster mass M in terms of its size and typical relative galaxy velocities. Zwicky measured the velocity dispersion among eight galaxies in the cluster, determined the cluster's size from its distance and angular size, and estimated the mass from the number of galaxies and typical galaxy mass. Finding a factor 400 discrepancy between his mass estimate and the mass implied by the galaxies' velocities, Zwicky observed ([145], translated in [117]):

> *If this should prove true, the surprising result would emerge that dark matter is present in very much greater density than luminous matter.*

Although Zwicky's numbers were off, his method – and conclusion – were sound. In fact, Zwicky had based his work off a distance to the cluster a factor of 7.5 too small. This error in measuring intergalactic distances (which also accounts for the over-estimate of the Hubble Constant in Hubble's work[86] from a few years before) traces to early issues with data needed for the absolute calibration of the period-luminosity relation in Cepheid variable stars, used as standard candles[87]. This, along with a baryonic mass estimate neglecting the dominant contribution from gas between the galaxies, distorted the result.[2] Recent measurements indicate that nonbaryonic dark matter makes up 84% of the matter in the universe in general[118], and some 88% of the matter in the Coma Cluster specifically[130].

Four years after Zwicky's publication, another astronomer replicated the key result by finding a similar discrepancy in the Virgo Cluster[46]. And in a modern adaptation of early galaxy cluster velocity measurements, the same conceptual technique can be applied to the intra-cluster gas. Using its x-ray emissions to measure the gas's temperature, and hence velocity distribution, the virial theorem is applied as above, leading to similar dark matter results.[131]

Contemporaneously with Zwicky, Jan Oort was investigating possible evidence for dark matter based on the motion of Milky Way stars relative to the galactic plane[114], expanding on earlier work by Jeans[89]. The following decades saw more studies of galaxy clusters and of stars within the Milky Way[46, 119]. The most dramatic results would come from studying galactic rotation curves.

Galaxy Rotation Curves

An alternative way to learn about matter distribution from the motion of astrophysical objects is to look at a galactic rotation curve, that is, a plot of the orbital speed of stars about their galaxy's center as a function of distance from that center.

[2] However, at this time "dark matter" was an established term for any non-luminous matter[46], whether baryonic or not, so the latter is more a discrepancy of language than of physics.

In 1914, the rotation of the Andromeda Galaxy (M31) was first detected by comparing the redshift of different parts of the galaxy. Subsequently, rotation curves were produced for Andromeda in the visible band starting in the 1930s and in the radio band starting in the 1950s. In the 1960s and 1970s, Vera Rubin, using a spectrograph developed by her collaborator Kent Ford, began producing galaxy rotation curves of higher quality than previously available, and with this technique studied first M31 and then dozens of other galaxies.[46, 128, 129] Rubin's results showed that each galaxy must have large amounts of non-luminous matter.

In a typical spiral galaxy, most of the mass in the form of stars is found in the galaxy's core. So for a star in the galactic disk, it would be expected that its orbital speed around the core is

$$v = \sqrt{\frac{GM_{\rm encl}}{r}} \tag{1.4}$$

where G is the gravitational constant, r is the star's distance from the core, and $M_{\rm encl}$ is the mass enclosed by the star's orbit, which is roughly the galactic core's mass. Although the velocity would deviate slightly because the other stars in the disk lack spherical symmetry in their distribution, the deviation is negligible.

What Rubin observed, however, was that the velocities did *not* fall off with the square root of the radial distance. In fact, the velocity curves stayed constant to the largest distances that would permit measurement (see Fig. 1.1). This indicated that the galaxies must also contain a non-luminous component with a very different radial distribution than the visible matter.[3] These results triggered a renewed interest in studying and understanding dark matter.

1.1.2 Composition of the Universe

Following the rapid advances of the "Golden Age of Cosmology" beginning in the early 1990s, there emerged a coherent picture of the abundance of dark matter in the universe and its role in the evolution of the universe over cosmological time.

Table 1.1 shows the constituents of the universe and the percentage of total energy density contributed by each at present. Some 70% of the universe is dark energy, a poorly-understood contribution characterized by its constant energy density and negative pressure, which drives an accelerating expansion. Another 30% is non-relativistic matter. That includes a 5% contribution from so-called baryonic matter, meaning matter made of any known Standard Model particles that are nonrelativistic (of which protons and neutrons dominate by mass, hence the name). Another quarter of the universe's content behaves as nonrelativistic particles (no pressure, thus energy density falling as the cube of the scale factor) but cannot be any Standard Model particle – this is the dark matter. Finally, relativistic particles, termed

[3]Relating this to the dark matter halo density profiles discussed in Sec. 1.3.7, a galaxy rotation curve that's perfectly flat in the large r limit is produced by a pseudo-isothermal halo model. More complicated density profiles can account for more complicated velocity curve shapes.

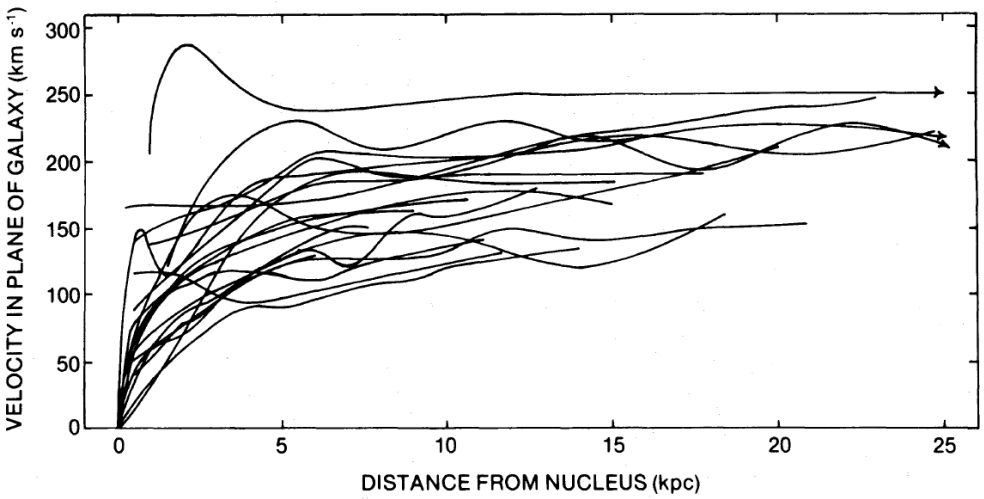


Figure 1.1: Velocity curves for 21 galaxies. Reprinted from [129].

Component	%	Scales as
Dark Energy	69.2 ± 1.2	a^0
Matter	30.8 ± 1.2	a^{-3}
Cold Dark Matter	25.8 ± 1.1	
Baryonic Matter	4.84 ± 0.10	
Radiation	$(9.04 \pm 0.25) \times 10^{-3}$	a^{-4}
Photons	$(5.38 \pm 0.15) \times 10^{-3}$	
Neutrinos	$(3.66 \pm 0.10) \times 10^{-3}$	

Table 1.1: Present composition of the universe, from the Planck Collaboration's 6-parameter fit to Planck satellite data[5] and from the Particle Data Group[118]. The "%" column gives the current percentage of total energy density. The variable a is the scale factor, a unitless factor which characterizes the expansion of the universe and by definition equals one at the present time.

"radiation," have an energy density falling off with the fourth power of the scale factor. The extra factor, compared to matter, is due to the cosmological redshift. This radiation, a small contribution at present but dominant at early times, now consists mostly of CMB photons and neutrinos. The total energy density of the universe is not detectably different from the critical density, making ours a "flat" universe to within experimental uncertainty.[118, 130]

This remarkable picture is supported by many converging lines of evidence. We will consider the two questions most germane to the dark matter line item: how is the 70% - 30% split between dark energy and matter measured, and how is it known that much of the matter cannot be baryonic?

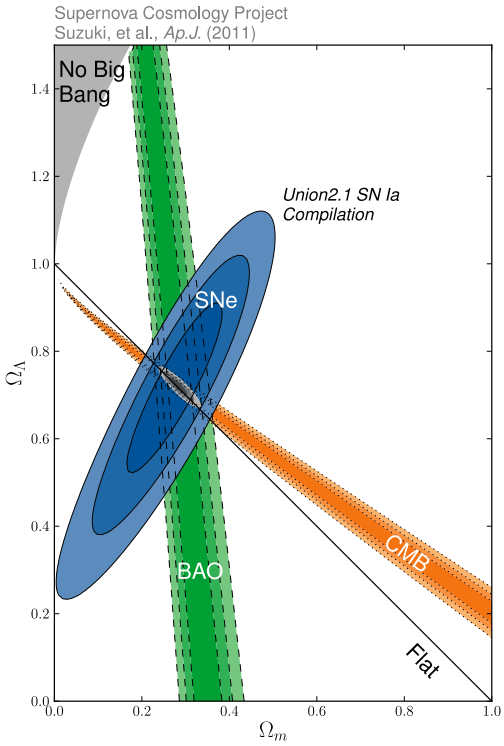


Figure 1.2: Three techniques for measuring Ω_m and Ω_Λ. SNe stands for Type-Ia supernovae, CMB for cosmic microwave background anisotropy, and BAO for baryon acoustic oscillations. Reprinted from [135].

1.1.3 Matter and Dark Energy

Fig. 1.2 shows graphically how multiple lines of evidence reinforce each other in arguing for a universe that is $\approx 70\%$ dark energy and $\approx 30\%$ matter. The horizontal axis is Ω_m, the ratio of matter density to critical density, and the vertical axis is Ω_Λ, the ratio of dark energy density to critical density.

The shaded region labeled "SNe" indicates the most likely phase space according to supernova measurements from the Supernova Cosmology Project and others[135]. This technique is based on Type-Ia supernovae, which can occur in a binary star system with a white dwarf. If the white dwarf orbits sufficiently close to its partner star, it can draw in stellar material. For white dwarfs below the Chandrasekhar limit (the maximum mass that can be supported by electron degeneracy pressure), this merely results in an occasional flare-up of hydrogen fusion on the surface, visible as a classical nova. But should the white dwarf exceed the Chandrasekhar limit, a cataclysmic explosion can entirely break it apart in a so-called Type-1a supernova.

Type-1a supernovae can serve as standard candles because they are all of approximately equal brightness, as all stars undergoing them are necessarily of the same mass. Although there is some brightness variability, this can be addressed because actual peak luminosity

is correlated with other features of the luminosity pulse shape. The brightness scale is calibrated using Type-1a supernovae in galaxies with distances that have already been measured from Cepheid variable stars.[130]

To first order, the distance to a star is just

$$d \approx \frac{cz}{H_0} \tag{1.5}$$

where z is the star's cosmological redshift, H_0 the Hubble constant, and c the speed of light. But to *second* order,

$$d_L \approx \frac{cz}{H_0}\left(1 + \frac{1-q_0}{2}z\right) \tag{1.6}$$

where d_L is specifically the luminosity distance, and the deceleration parameter q_0 measures the rate of deceleration of the universe; the number is thus negative. The deceleration parameter in turn is related to the present-day abundance of the various constituents of the universe by[130]

$$q_0 = \Omega_r + \frac{1}{2}\Omega_m - \Omega_\Lambda\,. \tag{1.7}$$

Here Ω_r, the ratio of radiation density to critical density, is negligible compared to the other terms. A measure strictly of q_0 would then constrain $\frac{1}{2}\Omega_m - \Omega_\Lambda$. In fact, supernovae measurements are made out to sufficiently high redshifts that higher order terms come to bear, and the best fit actually lies along a line that has a somewhat different slope (with a factor of around 1.3 instead of the factor of $\frac{1}{2}$). This explains the general shape of the region of parameter space allowed by supernova measurements – they most strongly constrain the value of $1.3\Omega_m - \Omega_\Lambda$.

We next consider the results from the CMB anisotropy, labeled "CMB" in Fig. 1.2. The cosmic microwave background is microwave radiation that closely follows a blackbody spectrum with a temperature of 2.7260 ± 0.0013K[75]. After subtracting out a comparatively large dipole in the CMB, due to the local motion of the solar system and observer, the remaining anisotropy projected onto its multipole moments is shown in Fig. 1.3. This anisotropy is caused by density perturbations at the time of recombination 373,000 years after the Big Bang[118], when falling temperatures led to the formation of neutral atoms, causing CMB photons to become decoupled from the previously-existing plasma.

Conceptually, the shape of the spectrum can be explained as follows: We assume a spectrum that initially has the same amount of anisotropy on all scales, as would result from inflation. As soon as enough time has passed that structures on a given size scale are in causal contact, perturbations begin growing on that scale under gravitational instability as matter falls into areas of higher density. Eventually, a maximal amount of inhomogeneity is achieved, but the rising pressure causes the matter to bounce back, only to begin the cycle anew. This leads to the oscillatory shape of the power spectrum – larger multipole moments correspond to smaller distance scales, which have more time in causal contact prior to freeze-out, allowing the oscillatory process to proceed further. Eventually, photon-mediated energy transfer damps out the oscillation, and the peaks get smaller.[53]

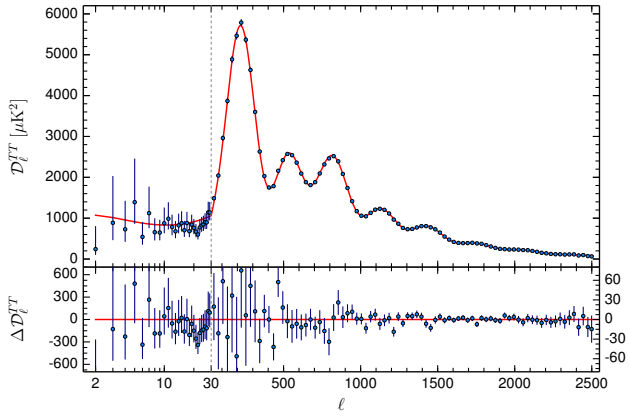


Figure 1.3: Cosmic microwave background anisotropy power spectrum, as measured by Planck. The horizontal axis is multipole moment. The upper plot is the actual result, and the lower plot is deviation from the fitted model. Reprinted from [6].

One consequence of this is that the location of the first peak is related to the curvature of the universe. In a universe with positive curvature, structures at a given distance appear larger (have a larger angular size) that the same structures at the same distance would appear in a flat universe. For a universe with negative curvature the reverse is true. So the location of the first peak in the anisotropy spectrum moves farther to the left the more positive the curvature. Curvature is related to how much the energy density of the universe exceeds its critical density today. This explains the general shape of the region of parameter space allowed by CMB measurements – they most strongly constrain the curvature, which scales with $\Omega_m + \Omega_\Lambda$.

Finally, we consider the results from baryon acoustic oscillations, labeled "BAO" in Fig. 1.2. In this approach, the positions of many galaxies are measured and the two-point correlation function is computed (see Fig. 1.4). Galaxy formation was seeded by the same density perturbations in the early universe that are responsible for the CMB anisotropy. Therefore, various features of the correlation function have cosmological significance. For example, the exact shape of the correlation function at low separations is related to the total amount of matter.[4][65, 85] Since dark energy was a negligible constituent of the universe in early times, the correlation function has less to say about it; the precise matter measurement is the technique's strength. This explains the general shape of the region of parameter space allowed by BAO measurements – they most strongly constrain the matter density Ω_m.

These three techniques look at three entirely different sets of objects: supernovae, the cosmic microwave background, and galaxies. Because the likely phase space for each experimental technique lies in a narrow ellipse and the three ellipses are oriented at different angles,

[4]To be specific, a plot of correlation function times separation squared will show a peak where the comoving separation equals the horizon size at the time of matter-radiation equality. This is caused by a slight suppression of perturbation formation on scales within causal contact during the radiation-dominated era.[65]

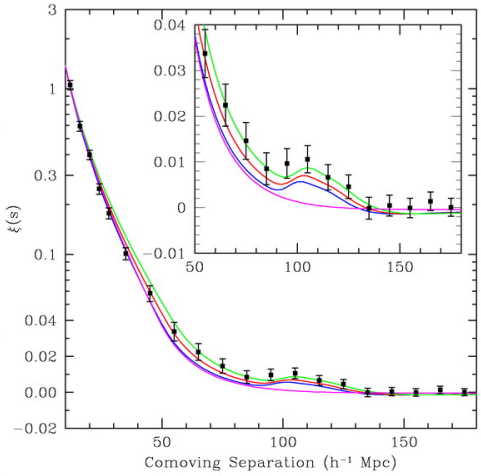


Figure 1.4: Baryon acoustic oscillation correlation function, based on data from the Sloan Digital Sky Survey. The horizontal axis is the comoving separation in units of h^{-1} Mpc, where $h \approx 0.72$ is related to the Hubble Constant; the vertical axis is the correlation function. The lines show the predictions of various models. Reprinted from [65].

any two techniques are sufficient to fix Ω_Λ and Ω_m to respectable precision, and having all three makes for an overdetermined result that is all the more robust.

1.1.4 Baryonic and Dark Matter

Having argued that matter adds up to some 30% of the density of the universe, we next survey how to determine that only about one-sixth of that matter can be baryonic. A variety of measurements now concur that the remaining matter cannot be explained away as non-luminous baryonic matter but must instead be something fundamentally different.

Some of the evidence comes from the techniques discussed in the previous section. Consider once more the CMB anisotropy spectrum (Fig. 1.3). Once a distance scale comes into causal contact, matter flows into overdense regions, then bounces back due to the resulting pressure, only to repeat the process. Because larger multipole moments correspond to distance scales that were in causal contact for longer prior to recombination, each multipole moment provides a photograph of a different instant in this process. In particular, the first peak corresponds to the first infall into initially overdense regions; the second peak corresponds to the maximum "bounce" away from those regions; the third peak represents the second infall, similar to the first in all but amplitude; and so on. In general, the amplitude of each peak is smaller than the one before it, but the exact amplitudes of the peaks provide a clue to dark matter. For dark matter would be expected *not* to bounce back from initially overdense regions because it does not strongly couple to the photons mediating that process. Therefore, having dark matter along with baryonic matter should suppress the size of the second peak (when the baryonic matter is fighting against the dark matter's potential well),

and it should enhance the size of the third peak (when the baryonic matter and dark matter are moving in concert). A careful inspection of Fig. 1.3 reveals that the third peak is almost the same size as the second one – a strong sign of nonbaryonic dark matter.[53]

Next, we return to baryon acoustic oscillations, and the two-point correlation function for the comoving distances between galaxies (Fig. 1.4). The correlation function is expected to show a small peak at the "sound horizon," the maximum comoving distance that an acoustic wave can travel between the Big Bang and recombination. This peak is in fact observed at around $100h^{-1}$ Mpc. The sound horizon is affected differently by the amount of total matter and the amount of baryonic matter, providing another handle on the difference between the two.[65]

Finally, we consider a very different clue to the disparity between baryonic matter and total matter: nucleosynthesis. This is the process by which primordial nuclei were formed in the first minutes after the Big Bang. The stage is set within one second after the Big Bang, when the weak interactions mediating transitions between neutrons and protons freeze out at temperatures around 1MeV. As the temperatures continue to drop (and with some neutrons decaying in the interim), the temperature eventually falls below the binding energy of deuterium around 100keV and nuclei can form. Calculating the exact primordial fractions of protium (hydrogen-1), helium-4, deuterium, and lithium-7 requires complicated modeling. The key point for our purposes is that the results depend sensitively on the baryon-to-photon ratio, and hence on the amount of baryonic matter.[53, 118] Fig. 1.5 shows how the primordial nuclear abundances require the same low concentration of baryonic matter as do the CMB measurements.

Gravity

In the above discussion, we have seen instances where the spatial distribution of dark matter differs from that of baryonic matter; these include the early universe's density perturbations and present-day galactic dark matter halos. Another notable example of the two types of matter not being in the same place can be found in the Bullet Cluster. Here, two galaxy clusters have collided. As shown in an x-ray image (Fig. 1.6), the two blobs of intercluster gas have been slowed by their interaction and have only moved a little ways past the collision site. However, using gravitational lensing (the bending of visible light from more distant sources due to gravity), the mass distribution can be traced out. This reveals that most of the mass is in two large lobes beyond the cluster gas. These in fact are the two clusters' dark matter, which continued apace after the collision.[58] Various attempts have been made to propose a theoretical model of gravity that might eliminate the need for dark matter, but explaining a structure like the Bullet Cluster through such an approach is all but impossible.

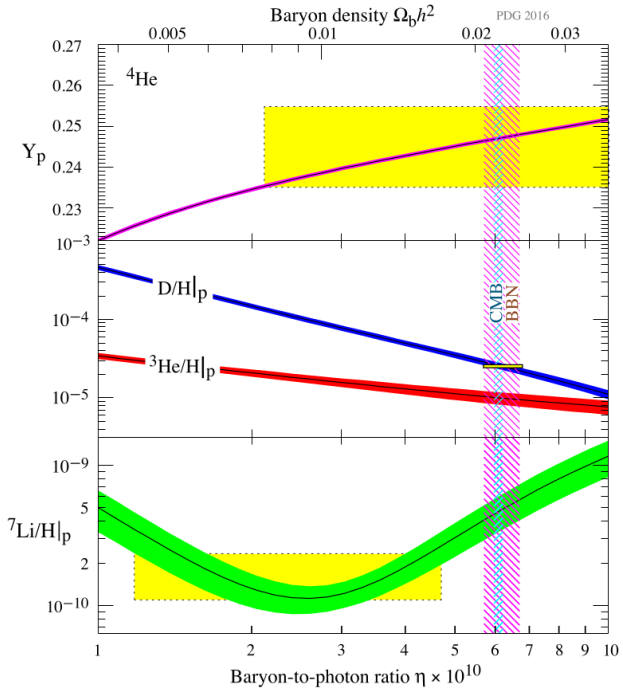


Figure 1.5: Primordial nuclei abundances from nucleosynthesis, as a function of either baryon-to-photon ratio or baryon density. Top panel shows ^{4}He mass fraction, middle panel shows D/H and ^{3}He/H ratios, and bottom panel shows ^{7}Li/H ratio. Within each panel, the diagonal or curved lines show the theoretical relationship among the variables, and the yellow boxes show the values consistent with observations. The region of upper-left-to-lower-right hatch marks shows the baryon density implied by nucleosynthesis, and the overlapping region of lower-left-to-upper-right hatch marks shows the range from CMB measurements. Reprinted from [118].

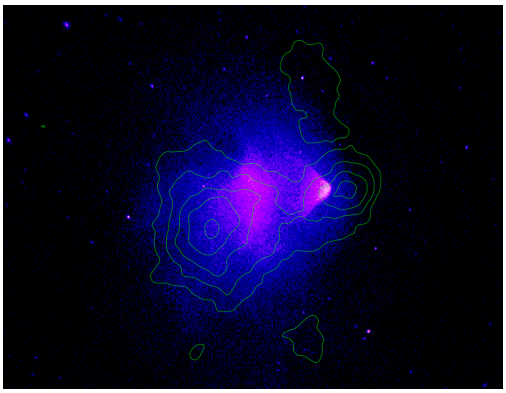


Figure 1.6: X-ray image of the Bullet Cluster, with overlaid mass contours from gravitational lensing. Reprinted from [104].

1.2 Dark Matter Candidates

The strong cosmological evidence for non-baryonic dark matter raises the question of just what that dark matter could be. Any particle candidate for dark matter must have a number of properties: It must be "matter," in the sense that it is nonrelativistic both now and far enough in the past to be consistent with observed galaxy formation. It must be "dark," in the sense that its coupling to standard model particles is weak enough to avoid direct observation or detection in all previous experiments. The particle must be stable or at least have a half life well in excess of the age of the universe, and its physics must accommodate the measured density of dark matter in the universe.[118]

We will first look at the WIMP, the dark matter particle theory that has received the most attention both historically and today, and also touch on another theory, the axion, as well as a few other proposals.

1.2.1 WIMPs

The weakly interacting massive particle (WIMP) is a hypothesized particle that could play the role of non-baryonic dark matter. In this theory, the dark matter particles are thermal relics, meaning a species of particle that was in thermal equilibrium with other constituents of the early universe only to subsequently "freeze-out," that is, fall out of equilibrium and become thermally decoupled from other particle species.[5] Importantly, the WIMP theory proposes that the dark matter particles are "cold" dark matter, meaning that they had already become nonrelativistic prior to decoupling. The theory's most striking feature, which we will derive below, is that it naturally suggests interactions between WIMPs and standard model (SM) particles that are mediated by the nuclear weak force, or at least something of similar strength to it.

WIMPs as Thermal Relics

The history of a particle species destined to become a cold relic proceeds through three stages, with the turning points fixed by the properties of the particle species and the changing properties of the universe itself. If the dark matter is made of WIMPs, it would have followed such a story.

Initially ("stage 1"), the WIMPs start out being relativistic and in thermal and chemical equilibrium. They're relativistic because the temperature of the universe, T, is substantially larger than their mass, m_χ, so the kinetic energy of a typical WIMP dwarfs its rest energy. A particle stays in equilibrium so long as its typical interaction rate Γ with other particles

[5]Note that "coupling" has slightly different meanings in particle physics and cosmology. In particle physics, two particle species are coupled if they are able to interact. In cosmology, they are coupled only if they frequently interact in practice. It is possible to meet the former definition but not the latter if particles become too dilute or low-temperature. We will use the term in the cosmological sense in section 1.2.1 and in the particle physics sense elsewhere.

is large compared to the Hubble parameter H. For relativistic, equilibrated WIMPs, the number density n_χ is[139]

$$n_\chi = \frac{\zeta(3)}{\pi^2} g_\chi T^3 \begin{cases} 1 & \text{bosonic WIMPs} \\ \frac{3}{4} & \text{fermionic WIMPs} \end{cases} \tag{1.8}$$

where ζ is the Riemann zeta function, g_χ is the number of spin states for the WIMP, and we use units where $\hbar = c = k_B = 1$. Note that dark matter decouples during the radiation-dominated era[53], so before and during freeze-out the temperature is roughly inversely proportional to the scale factor a [123].

As time progresses and the temperature of the universe decreases, eventually $T < m_\chi$ and the WIMPs stop being relativistic. The effect of this change ("stage 2") is to decimate the WIMP population. Because the WIMPs are still in chemical equilibrium and interacting with their surroundings, it is easy for them to find an appropriate antiparticle and annihilate. But because their rest energy is now much greater than the universe's ambient temperature, it has become a much rarer occurrence to produce a WIMP particle-antiparticle pair. The resulting population plunge is captured by the exponential factor in the expression for the WIMP number density in the nonrelativistic regime[139],

$$n_\chi = g_\chi \left(\frac{m_\chi T}{2\pi}\right)^{\frac{3}{2}} e^{-\frac{m_\chi}{T}} . \tag{1.9}$$

As the WIMPs become more diluted due to annihilation as well as the continuing expansion of the universe, eventually their typical interaction rate Γ falls below the Hubble parameter, and they cease non-gravitational interactions with other particles entirely ("stage 3"). At that point, they have frozen out of equilibrium, and their number density falls off purely due to expansion, as

$$n_\chi = \frac{n_{\chi f} a_f^3}{a^3} = \frac{n_{\chi 0}}{a^3} \tag{1.10}$$

where $n_{\chi f}$ and a_f are the number density and scale factor at freeze-out, respectively, and $n_{\chi 0}$ is the number density today.

Fig. 1.7 traces the evolution in WIMP number density during freeze-out, and the accompanying Table 1.2 summaries the steps.

Estimating WIMP Interactions

Having laid out the history of WIMP decoupling, we are now in a position to estimate a typical WIMP interaction cross-section. In what follows, we confine our goal to getting an answer correct to within a few orders of magnitude. As such, all constants of order unity will be disregarded. This presentation of the derivation is developed from Refs. [53, 139, 122]. Assume units where $\hbar = c = k_B = 1$.

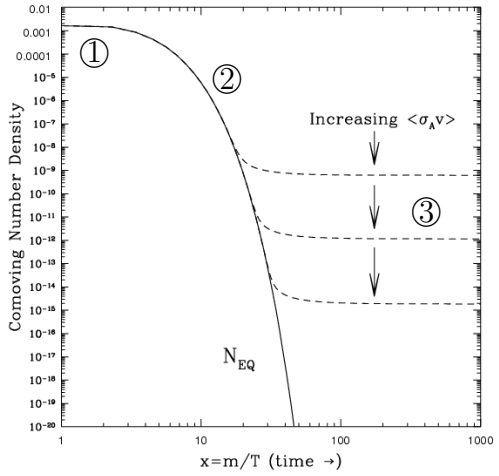


Figure 1.7: Comoving number density (na^3) of a cold thermal relic during freeze-out, as a function of m_χ/T (which is proportional to a). The three stages of this process, labeled 1, 2, and 3, are described in Table 1.2. Reprinted from [92], with added labels.

Stage ①	Stage ②	Stage ③
In Equilibrium: $\Gamma > H$ Relativistic: $T > m_\chi$	In Equilibrium: $\Gamma > H$ Non-relativistic: $T < m_\chi$	Frozen Out: $\Gamma < H$ Non-relativistic: $T < m_\chi$
$n_\chi \approx T^3$	$n_\chi \approx (m_\chi T)^{\frac{3}{2}}\, e^{-\frac{m_\chi}{T}}$	$n_\chi = n_{\chi 0}/a^3$

Table 1.2: Conceptual outline of the freeze-out of a cold thermal relic. The evolution of the comoving number density as the universe progresses through each stage is shown in Fig. 1.7.

WIMP freeze-out occurs around when the WIMP annihilation rate falls below the Hubble parameter, beyond which point the number of WIMPs no longer changes much. The annihilation rate is roughly

$$\Gamma \approx n_\chi \sigma v \tag{1.11}$$

where σ is a typical cross-section for WIMP physics, v is a typical WIMP velocity, and n_χ is the number density of WIMPs. So at freeze-out (denoted by a subscript f on the number density),

$$H \approx n_{\chi f} \sigma v\,. \tag{1.12}$$

Already, we have made a number of simplifying assumptions. An exact result would require calculating all interactions among all interacting particle species as freeze-out proceeded.[6] One would need to integrate σv over an appropriate velocity distribution to obtain the

[6]This is the approach taken when, for example, modeling nucleosynthesis.

exact average quantity. But for the purpose of an order-of-magnitude estimate, the simple multiplication will suffice.

Next, to address H we use the Friedmann Equation[53]

$$H^2 = \frac{8\pi G}{3}\rho - \frac{\kappa}{a^2} \tag{1.13}$$

where G is the gravitational constant, ρ the total energy density, and κ the curvature. In a flat universe, and dropping the unitless constants, $H^2 \approx G\rho$. Since the universe is radiation-dominated during WIMP freeze-out, $H^2 \approx G\rho_\gamma$, where ρ_γ is the energy density of photons.[7] Substituting this into Eqn. 1.12 gives

$$\sqrt{G\rho_\gamma} \approx n_{\chi f}\sigma v\,. \tag{1.14}$$

Next, taking the energy-density form of the Stefan-Boltzmann Law,

$$\rho = \frac{\pi^2}{15}T^4\,, \tag{1.15}$$

applying it to the photons at the time of WIMP decoupling, and dropping unitless factors, gives

$$\rho_\gamma \approx T_f^4\,. \tag{1.16}$$

Substituting this into Eqn. 1.14 gives

$$n_{\chi f} \approx \frac{\sqrt{G}T_f^2}{\sigma v}\,. \tag{1.17}$$

While the WIMPs are undergoing freeze-out, the photons are still in thermal equilibrium, and their number density is given by an expression analogous to Eqn. 1.8, which after dropping unitless constants is

$$n_{\gamma f} \approx T_f^3\,. \tag{1.18}$$

After WIMP decoupling, the number of WIMPs falls off as a^3, the cube of the scale factor, since the number of WIMPs changes little. Importantly, the number of photons also falls off roughly as a^3. This dependency holds for photons in the radiation-dominated era, shown by referring to the previous equation and noting that $T \propto a^{-1}$ when radiation dominates. The same dependency also holds once photons begin free-streaming after recombination, and the interlude between those two periods is not sufficient to change the photon number density by orders of magnitude. Therefore, the ratio of the number densities is roughly the same today as it was just after WIMP freeze-out:

$$\frac{n_{\chi 0}}{n_{\gamma 0}} \approx \frac{n_{\chi f}}{n_{\gamma f}}\,. \tag{1.19}$$

[7] The ratio between the energy density in photons and the energy density in all relativistic species depends on exactly how many particle species are relativistic, but in any case the difference is a factor of less than a couple orders of magnitude.

Subscript f refers to the time of WIMP decoupling, and subscript 0 refers to the present day. Substituting Eqns. 1.17 and 1.18 into Eqn. 1.19 gives

$$\frac{n_{\chi 0}}{n_{\gamma 0}} \approx \frac{\sqrt{G}}{\sigma v T_f} \,. \tag{1.20}$$

Because the typical WIMP kinetic energy just before decoupling should match the temperature,

$$\frac{1}{2} m v^2 \approx T \,, \tag{1.21}$$

it follows that Eqn. 1.20 can be expressed without explicit reference to velocity as

$$\frac{n_{\chi 0}}{n_{\gamma 0}} \approx \frac{\sqrt{G}}{m\sigma} \left(\frac{m}{T_f} \right)^{\frac{3}{2}} \,. \tag{1.22}$$

Therefore, the dark matter energy density today is

$$\rho_{\chi 0} = m n_{\chi 0} \approx \frac{\sqrt{G} n_{\gamma 0}}{\sigma} \left(\frac{m}{T_f} \right)^{\frac{3}{2}} \,. \tag{1.23}$$

Finally, we assume that freeze-out occurs when T, although smaller than m, does not differ from m by many orders of magnitude. This may seem at first to be a highly restrictive and arbitrary assumption. However, during the intermediate stage when the WIMPs are nonrelativistic and in equilibrium, their number density falls off exponentially with a factor of $e^{-\frac{m}{T}}$ (Eqn. 1.9). So restricting m/T_f to a range of, say, three orders of magnitude actually only restricts the freeze-out WIMP number density to within a range of twenty orders of magnitude, which is hardly restrictive at all.

Thus we arrive at our final result, an estimate for a typical WIMP interaction cross-section. Restoring factors of $\hbar$ and c previously left off,

$$\sigma \approx \sqrt{\frac{\hbar^3 G}{c}} \frac{n_{\gamma 0}}{\rho_{\chi 0}} \,. \tag{1.24}$$

Given the many simplifying assumptions underlying this result, the only expectation is that it should be within a few orders of magnitude of the correct value; any stronger claims of precision would not be justified. The significance of the result is that it makes an estimate at all and, notably, one that does not depend on WIMP mass – m has canceled out of the final result entirely. In fact, all the variables that do appear are known quantities.[8] Whatever else may be true of WIMP physics, WIMPs should participate in some interaction that has a typical cross-section on the order of 10^{-38}cm^2, or 10^{-11}mb.

This becomes suggestive if one considers the cross-sections of typical interactions associated with each of the known fundamental forces, as shown (aside from gravity) in Table 1.3.

[8] $G = 6.67 \times 10^{-11}$ $\text{m}^3\text{kg}^{-1}\text{s}^{-2}$, $n_{\gamma 0} = 411$ cm^{-3}, and $\rho_{\chi 0} = 2.23 \times 10^{-30}\text{g}$ cm^{-3} [118]

Force	Typical Cross-section (mb)
Strong	10
Electromagnetic	10^{-3}
Weak	10^{-11}

Table 1.3: Typical cross-sections for fundamental forces. From [80].

Our estimated WIMP annihilation cross-section falls right into the neighborhood of cross-sections associated with interactions mediated by the nuclear weak force.[9] This striking concordance, sometimes called the "WIMP miracle," motivates the final idea of the WIMP concept: not only is dark matter a cold thermal relic, but it is one that couples to the weak force and could thereby be directly detected.

Some specific theoretical ways to implement a WIMP will be discussed in Sec. 1.3.

1.2.2 Axions

In addition to WIMPs, other theories of the particle nature of dark matter have been proposed. One of these is the axion, which was originally suggested as a solution to a different particle physics problem.

The idea of axions can be outlined as follows. Consider a field with a spontaneously-broken U(1) symmetry, as produced by a so-called "Mexican hat" potential (Fig. 1.8a). This field will produce a massless Nambu-Goldstone boson, corresponding to the angular degree of freedom. The particle is massless because changing the angular coordinate, or moving "around the brim of the hat," does not change the potential, so no mass term for this particle appears in the Lagrangian.

However, if the original U(1) symmetry is itself slightly broken, such as by the tilt shown in Fig. 1.8b, then after spontaneous symmetry breaking there will instead be an axion, that is, a pseudo-Nambu-Goldstone boson with a small but nonzero mass. Fig. 1.8c shows the potential associated with the angular degree of freedom resulting from the spontaneous breaking of such an approximate symmetry. Because axions would be produced from spontaneous symmetry breaking once the temperature of the early universe became sufficiently low, they would never have been in thermal equilibrium. This production mechanism is fundamentally different from the thermal relic origin of WIMPs.[53, 118]

The axion has the merit of resolving the so-called strong CP problem. The strong force Lagrangian contains a CP-violating term. It is known from measurements of the neutron electric dipole moment that the coefficient of this term must be very small ($\lesssim 10^{-10}$) even though there is no *a priori* reason why it shouldn't be of order unity. One solution is to regard

[9] By comparison, the estimate is nowhere near the typical cross-sections for strong or electroweak processes; their presence in Table 1.3 highlights the great range of orders of magnitude over which known forces vary. There is no *a priori* reason for the thermal relic cross-section to be near any of them.

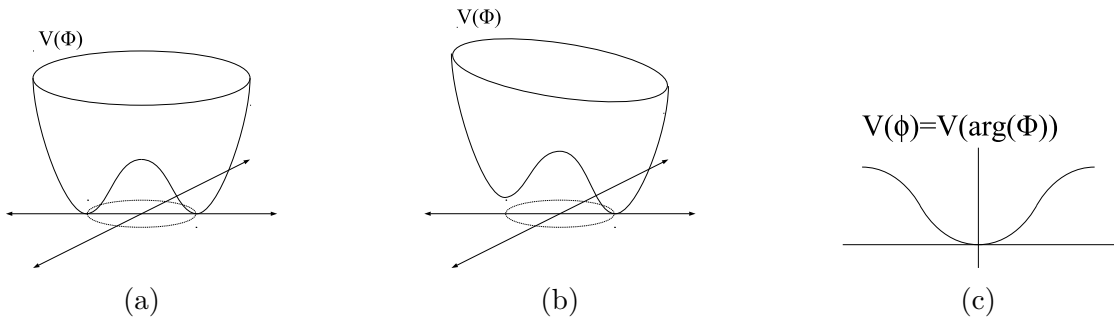


Figure 1.8: Potential with a spontaneously broken U(1) symmetry. (a) Exact symmetry. (b) Approximate symmetry. (c) Potential vs. angular coordinate in approximate symmetry case.

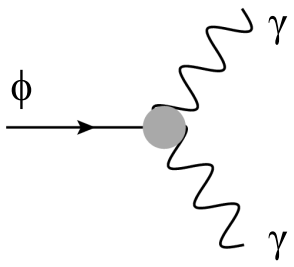


Figure 1.9: Feynman diagram of axion-photon coupling. The actual interaction is via a quark loop, symbolized by the "blob" at the vertex.

the coefficient not as a constant but as a dynamical value stemming from an underlying field of the type described above, which would allow the coefficient to evolve to zero. The axion is then an observable consequence of the symmetry.[69, 118]

Axions couple to photons via the vertex shown in Fig. 1.9, which forms the basis for many direct detection methods. One technique is to look for "light shining through walls," due to a photon radiating an axion, which can then produce a pair of photons on the other side of an opaque barrier. Another method is to look for a beam of light becoming polarized while traversing a magnetic field, since the two polarizations would have different axion production rates. Axions can also be inferred from their effect on stellar evolution and other astronomical phenomena.[118]

1.2.3 Other Dark Matter Theories

In addition to WIMPs and axions, a host of other dark matter candidates have been proposed. Some have been effectively ruled out, while others remain viable.

One early idea was that dark matter might be so-called massive compact halo objects (MACHOs), meaning astronomical objects such as brown dwarfs or dim stars (red dwarfs, neutron stars, etc.). These would be detectable through microlensing experiments, which

measure changes in the brightness of more distant stars due to gravitational lensing by more nearby MACHOs. The non-observation of this, combined with the evidence against dark matter being baryonic (see Sec. 1.1.4), largely ruled out MACHOs.[46] It remains possible for primordial black holes to serve as MACHOs if they formed prior to nucleosynthesis, although this has been characterized as "somewhat contrived" [118].

Another early theory was that Standard Model neutrinos (the usual electron-, muon-, and tau-neutrinos) could serve as dark matter. These would be "hot" dark matter, having frozen out while still relativistic. In light of the calculable number density of relic neutrinos, the existing upper bounds on neutrino masses are now strong enough to preclude Standard Model neutrinos from being the dark matter.[53] However, a different hypothesis related to neutrinos remains viable. Each ordinary neutrino species is part of a weak isospin doublet along with the other lepton of the same generation. In addition, there could be other neutrino species that are weak isospin singlets, so-called sterile neutrinos. A fairly-stable sterile neutrino in the right mass range might resolve some small anomalies seen in neutrino oscillation experiments while also serving as dark matter.[116]

Many new theories involve a more complicated picture of dark matter, positing multiple species of dark matter particles and/or dark matter particles that frequently interact non-gravitationally in the present universe. Theories that involve dark matter self-interactions may be able to resolve problems with structure formation simulations[57]. One example, the SIMP (strongly interacting massive particle) theory, proposes thermal relics that decay by a five-point interaction, where three particles go in and two come out. Under this assumption, the observed dark matter density is consistent with an energy scale typically associated with the strong force.[82] Hidden sector dark matter theories posit multiple species of dark particles with their own interactions. For example, in the dark photon theory, spin-1/2 dark matter generates a novel gauge boson through the same local gauge invariance that leads to photons in Standard Model matter[4]. Different theories allow for different types of coupling to Standard Model particles. One coupling that appears in a variety of models is the Higgs portal, where the dark sector can be detected through interactions between the Standard Model Higgs particle and its dark sector equivalent[132, 67]. Some of these theories lead to WIMP-like interactions.

Finally, scalar field dark matter posits dark matter particles that are bosons of such low mass ($\sim 10^{-31}$ GeV) that they have wavelengths the size of an entire galactic dark matter halo, making the halo a huge Bose-Einstein condensate[102].

Range of Theories

Most leading dark matter particle models are "two for the price of one" – they furnish a dark matter candidate while also solving some other outstanding physics problem.[10] Nevertheless

[10] Axions solve the strong CP problem. Dark matter self-interaction solves the core-cusp problem. Asymmetric dark matter solves the baryon asymmetry problem. Lightest superparticle WIMPs (discussed in Sec. 1.3.4) are part of a solution to the hierarchy problem. Before the solar neutrino problem was solved by the discovery of neutrino oscillations, there was a dark matter candidate (the "cosmion") to solve that, too.

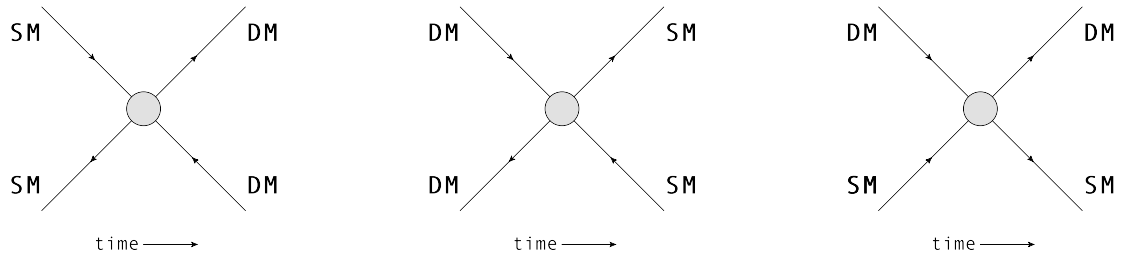


Figure 1.10: Feynman diagrams of the various approaches to detecting dark matter. (a) Dark matter pair production in an accelerator. (b) Astronomical observation of dark matter annihilation. (c) Direct detection of dark matter scattering off a target. "DM" indicates dark matter and "SM" Standard Model matter.

the range of possible models is vast. Let us consider just one parameter: the mass of the dark matter particle. We start by considering the full range of plausible masses, from the Hubble scale at 10^{-43} GeV to the Planck scale at 10^{19} GeV. There is presumably a lower limit on fermionic dark matter mass from phase space considerations at about 10^{-7} GeV. While not affected by that constraint, bosonic dark matter should require a DeBroglie wavelength less than or equal to the size of a galactic dark matter halo, setting a lower bound around 10^{-31} GeV. This still leaves some 26 orders of magnitude for a permissible fermion dark matter mass, and 50 orders of magnitude for bosonic dark matter. WIMPs are generally presumed to have masses within a few orders of magnitude of the weak scale around 10^2 GeV, and axions are generally presumed to have masses within five or so orders of magnitude of 10^{-14} GeV. This leaves a large range of other mass possibilities.[124] In a plot of dark matter cross-section versus dark matter mass (as in Ref. [127]), various leading models appear as islands in an ocean of possible parameter space.

1.2.4 Detection Approaches

There are three basic approaches to detecting individual dark matter particles, corresponding to the possible arrangements of the Feynman diagram vertex that couples dark matter (DM) and Standard Model (SM) particles. Suppose this coupling takes the form of a four-point interaction with two DM lines and two SM lines,[11] as in Fig. 1.10.

The first approach is DM pair production (Fig. 1.10a). Two SM particles annihilate, producing a pair of DM particles. This could be discovered with an accelerator. Although the dark matter itself would not be visible in the accelerator's detectors, it would show up as "missing" four-momentum, from which the rest mass of the dark matter particles could

[11]The "blob" at each intersection in Fig. 1.10 indicates that the specific form of the interaction is unspecified. For example, it might be mediated by a DM or SM force carrier instead of being a true four-point interaction. Although we focus on vertices with two DM and two SM external lines, this classification scheme is generalizable to vertices with other numbers of DM and SM lines.

be reconstructed. Dark matter searches of this kind have been conducted with detectors including ATLAS[40] and CMS[59], which have both released dark matter searches using data from the LHC's 13 TeV run.

The second approach, indirect detection (Fig. 1.10b), is the time reversal of the first. In this case, a dark matter particle and its dark anti-particle annihilate, producing SM particles. This could lead to an excess of various SM particles in astrophysical observations. For example, dark matter could accumulate inside the sun if scattering events left DM particles with too little kinetic energy to escape the solar gravitational well. This natural concentration of dark matter could lead to annihilation events that might produce neutrinos detectable in terrestrial neutrino detectors, such as SuperKamiokande[56] and IceCube[3]. Alternately, DM annihilation could produce mono-energetic photons from near the galactic center, but alternative explanations have been offered for such signals observed in recent years[120]. Detecting astrophysical dark matter annihilation is inherently challenging because it requires a high degree of understanding of other production mechanisms for the SM particle under study.

Finally, there is so-called direct detection, in which a DM particle interacts with a SM particle (Fig. 1.10c). Examples include the Axion Dark Matter Experiment (ADMX), which has placed limits on axions interacting with a SM magnetic field to produce microwave photons[84].[12] WIMP direct detection experiments, discussed in the next section, use an instrumented mass of target material to look for recoils caused by WIMP scattering.

1.3 WIMP Dark Matter Direct Detection

Having surveyed a range of dark matter models and detection mechanisms, we now focus on the model and mechanism of concern to the Large Underground Xenon (LUX) experiment. The LUX experiment's primary physics goal is to search for WIMP dark matter. It is a direct detection experiment, designed to detect galactic halo WIMPs scattering off of an instrumented xenon target. Like practically all WIMP direct detection experiments, LUX searches specifically for WIMP scattering off of nuclei. This is not because WIMPs can't interact with electrons – On the contrary, many if not most models that predict WIMP-nucleon coupling also predict WIMP-electron coupling. There are, however, two benefits to focusing on nuclear scattering. The first is that nuclear masses fall near the range of expected WIMP masses around the weak scale, while the electron mass is much smaller. That means elastic nuclear recoils can transfer more energy than elastic electron recoils.[13] The second benefit is that nuclear recoils are a signal channel in which background problems are more manageable. Numerous electron recoils can be caused by β and γ rays, while the nuclear recoil background from neutrons can be made less severe.

[12] In this case, there is no outgoing DM line because the axion is destroyed, but it still falls in the category of direct detection.

[13] There is some interest in searching for light WIMPs via electron scattering[142].

In this section, we will consider direct detection of WIMPs via scattering off of nuclei, focusing on the particle physics of how such interactions take place.

1.3.1 WIMP/Nucleus Coupling: Spin-Independent and Spin-Dependent

To model WIMP-nucleus interactions, the first step is proposing a theory of WIMP coupling to an individual quark, or gluon if applicable. This can then be extrapolated to a whole nucleon and ultimately an entire nucleus, which depending on the theory and desired accuracy may require more or less information about the geometry and spin structure of the nucleons and nucleus.[90] For WIMP-nucleus scattering in the low-velocity limit, the coupling generally takes one of two forms: spin-independent or spin-dependent.

Spin-dependent coupling arises when the effective Lagrangian term coupling dark matter particles χ to quarks q takes on the axial vector form

$$\mathcal{L} = f\bar{\chi}\gamma^{\mu}\gamma^{5}\chi\bar{q}\gamma_{\mu}\gamma^{5}q \tag{1.25}$$

where f is some arbitrary coupling constant and γ^5 is the usual product of the gamma matrices, $\gamma^5 \equiv i\gamma^0\gamma^1\gamma^2\gamma^3$. In the non-relativistic limit, this axial vector coupling describes dark matter that "couples to spin," meaning that the matrix element for this interaction changes sign when either the quark's spin or the WIMP's spin is flipped. Since most nucleons in a large nucleus are in pairs, with one spin up and the other spin down, their contributions in any axial vector theory will largely cancel out. To first order then, the spin-dependent coupling of a spin-0 nucleus is zero, and the spin-dependent coupling of a nucleus with nonzero nuclear spin is due entirely to the one unpaired nucleon[78]. More rigorous calculations can give more accurate results but require knowledge of the spin structure of the nucleus in question[90].

Spin-independent coupling in the low-velocity limit can arise in a variety of ways. The simplest way is to have a scalar coupling of the dark matter to quarks, with an effective Lagrangian like

$$\mathcal{L} = f\bar{\chi}\chi\bar{q}q \tag{1.26}$$

However, there are other ways to get spin-independent dark matter. Consider vector coupling

$$\mathcal{L} = f\bar{\chi}\gamma^{\mu}\chi\bar{q}\gamma_{\mu}q \tag{1.27}$$

and even tensor coupling

$$\mathcal{L} = f\bar{\chi}\sigma^{\mu\nu}\chi\bar{q}\sigma_{\mu\nu}q \tag{1.28}$$

where $\sigma^{\mu\nu} \equiv \frac{i}{2}(\gamma^{\mu}\gamma^{\nu} - \gamma^{\nu}\gamma^{\mu})$. In general, these three types of coupling lead to different behavior. But in the low-velocity limit, all three reduce to the same form. The resulting matrix element is spin-independent in that flipping the spin of the incoming and outgoing quark does not change the matrix element, nor does flipping the spin of the incoming and outgoing dark matter particle. Because the matrix element in the non-relativistic limit is

proportional to the product of the quark and dark matter particle masses,[14] spin-independent dark matter is said to "couple to mass." [90]

For completeness, we observe that the final possible effective Lagrangian, with the pseudoscalar coupling

$$\mathcal{L} = f\bar{\chi}\gamma^5\chi\bar{q}\gamma^5 q\,, \tag{1.29}$$

gives a matrix element of 0 in the low-velocity limit.[15]

1.3.2 "Heavy Neutrino" Model

Historically, the first particle model of WIMP dark matter was the "heavy neutrino." In 1977, in the wake of the discovery of the third generation of fundamental particles, three almost-simultaneous publications[99, 143, 88] raised the idea of a neutrino with large rest mass and discussed cosmological limits on such a particle. One of these papers, by Lee and Weinberg[99], observed that the particle could serve as dark matter.[16] In 1985, Goodman and Witten[78] worked out the spin-independent and spin-dependent cross section for scattering of heavy neutrinos off of nuclei.

It is illuminating to consider in some detail the derivation of the WIMP scattering cross-section in the original Goodman-Witten model of spin-independent dark matter[78], as this illustrates not just a specific model of historical importance, but more generally the process of working from theoretical proposals to testable interaction cross-sections. Many of the features of this model appear in other models as well, such as models from supersymmetry.

In the model, dark matter is hypothesized to scatter off of nuclei by the weak interaction, and specifically by the exchange of a neutral weak mediator, the Z neutral intermediate vector boson of the Glashow-Weinberg-Salam theory of weak interactions. The dark matter particle is treated as a true heavy Dirac neutrino, with a large mass but the same vector and axial couplings to the Z as a Standard Model Dirac neutrino would have (Fig. 1.11). We will first look at Z-mediated scattering in general, and incorporate the other assumptions of the model as needed. In the following derivation we use Gaussian units and retain all factors of c and $\hbar$ for clarity.

Scattering Off a Quark

We begin with dark matter scattering off of a single quark. (The generalization to an entire nucleus will follow.) From the Feynman diagram of the quark-DM interaction (Fig. 1.11), we

[14] and also because the combined matrix element for an entire nucleus is roughly proportional to atomic mass

[15] In writing all of these Lagrangians, we have implicitly assumed a spin-$\frac{1}{2}$ dark matter particle. If the dark matter has a different spin, then the dependence of the matrix element on the dark matter's spin or mass may differ, but the conclusions about coupling to quarks remains the same, because the quark current still must be one of the five possible bilinear covariants: $\bar{q}q$, $\bar{q}\gamma^5 q$, $\bar{q}\gamma^\mu q$, $\bar{q}\gamma^\mu\gamma^5 q$, or $\bar{q}\sigma^{\mu\nu}q$. In particular, the classification into spin-independent and spin-dependent coupling in the zero-velocity limit still holds.

[16] To be exact, Lee and Weinberg noted that heavy neutrinos would "provide a plausible mechanism for closing the universe," as this work predated the discovery of dark energy.

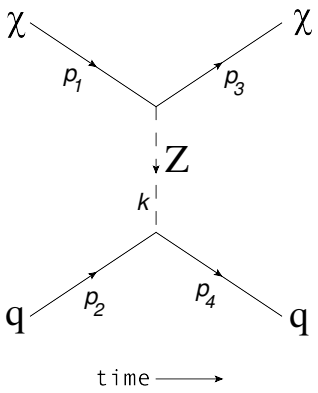


Figure 1.11: Feynman diagram of "heavy neutrino" dark matter scattering. The dark matter χ exchanges a neutral intermediate vector boson Z with a quark q.

write the factors of the scattering matrix element, which after integrating over the internal momentum k of the Z boson gives

$$\mathcal{M} = \frac{g_Z^2}{4M_Z^2c^2}\bar{u}_3\gamma^\mu\left(c_V^\chi - c_A^\chi\gamma^5\right)u_1\bar{u}_4\gamma_\mu\left(c_V^q - c_A^q\gamma^5\right)u_2 \tag{1.30}$$

where c_V^χ and c_A^χ are the vector and axial vector couplings of the dark matter particle, c_V^q and c_A^q are the same for the quark, g_Z and M_Z are the weak coupling constant and Z mass, and the various u's are the Dirac spinors of incoming and outgoing particles with subscripts matching Fig. 1.11.

When working with weak interactions, it is convenient to define the so-called Fermi coupling constant, G_F, as

$$G_F \equiv \frac{\sqrt{2}}{8}\left(\frac{g_W}{M_Wc^2}\right)^2(\hbar c)^3 = \frac{\sqrt{2}}{8}\left(\frac{g_Z}{M_Zc^2}\right)^2(\hbar c)^3 \tag{1.31}$$

and we further define G_F^o to absorb additional factors of c and $\hbar$:

$$G_F^o \equiv G_F\frac{c^2}{(\hbar c)^3}\,. \tag{1.32}$$

In terms of the latter the formula for the matrix element is

$$\mathcal{M} = \sqrt{2}G_F^o\bar{u}_3\gamma^\mu\left(c_V^\chi - c_A^\chi\gamma^5\right)u_1\bar{u}_4\gamma_\mu\left(c_V^q - c_A^q\gamma^5\right)u_2\,. \tag{1.33}$$

Multiplying this expression by its complex conjugate to get an expression for $|\mathcal{M}|^2$, and then summing over outgoing particle spins and averaging over incoming particle spins gives an expression for the spin-averaged square of the matrix element's magnitude:

$$\langle|\mathcal{M}|^2\rangle = \frac{G_F^o}{2}\sum_{\text{spins}}\left[\bar{u}_3\gamma^\mu\left(c_V^\chi - c_A^\chi\gamma^5\right)u_1\right]\left[\bar{u}_3\gamma^\nu\left(c_V^\chi - c_A^\chi\gamma^5\right)u_1\right] \cdot\left[\bar{u}_4\gamma_\mu\left(c_V^q - c_A^q\gamma^5\right)u_2\right]\left[\bar{u}_4\gamma_\nu\left(c_V^q - c_A^q\gamma^5\right)u_2\right] \tag{1.34}$$

Using the so-called Casimir's Trick[79] to rewrite the expression in terms of traces of products of gamma matrices, and evaluating the resulting expressions, gives the most general form of the tree-level matrix element for neutral weak scattering:

$$\begin{aligned}\langle|\mathcal{M}|^2\rangle = 16{G_F^o}^2 \Big[&\left({c_V^\chi}^2 + {c_A^\chi}^2\right)\left({c_V^q}^2 + {c_A^q}^2\right)((p_1\cdot p_2)(p_3\cdot p_4) + (p_1\cdot p_4)(p_2\cdot p_3)) \\ &+ \left({c_V^\chi}^2 - {c_A^\chi}^2\right)\left({c_V^q}^2 - {c_A^q}^2\right) 2m_\chi^2 m_q^2 c^4 \\ &- \left({c_V^\chi}^2 + {c_A^\chi}^2\right)\left({c_V^q}^2 - {c_A^q}^2\right) m_q^2 c^2 (p_1\cdot p_3) \\ &- \left({c_V^\chi}^2 - {c_A^\chi}^2\right)\left({c_V^q}^2 + {c_A^q}^2\right) m_\chi^2 c^2 (p_2\cdot p_4) \\ &+ 4c_V^\chi c_A^\chi c_V^q c_A^q \left((p_1\cdot p_2)(p_3\cdot p_4) - (p_1\cdot p_4)(p_2\cdot p_3)\right) \Big] .\end{aligned} \tag{1.35}$$

If the dark matter has the same couplings as a neutrino,

$$c_V^\chi = c_A^\chi = \frac{1}{2}, \tag{1.36}$$

then the scattering matrix element immediately simplifies to

$$\begin{aligned}\langle|\mathcal{M}|^2\rangle = 8{G_F^o}^2 \Big[&(c_V^q + c_A^q)^2 (p_1\cdot p_2)(p_3\cdot p_4) + (c_V^q - c_A^q)^2 (p_1\cdot p_4)(p_2\cdot p_3) \\ &- \left({c_V^q}^2 - {c_A^q}^2\right) m_q^2 c^2 (p_1\cdot p_3) \Big] .\end{aligned} \tag{1.37}$$

For two-body scattering in the center-of-mass (CM) frame,

$$\frac{d\sigma}{d\Omega} = \left(\frac{\hbar c}{8\pi}\right)^2 \frac{S\langle|\mathcal{M}|^2\rangle}{(E_1+E_2)^2}\frac{|\mathbf{p}_f|}{|\mathbf{p}_i|} \tag{1.38}$$

where $S = 1$ in this case because there are no identical particles among the ingoing or outgoing particles, and $|\mathbf{p}_i|$ and $|\mathbf{p}_f|$ (the magnitudes of the incoming and outgoing momenta, respectively) are equal to each other in this CM frame because the interaction is elastic. Substituting Eqn. 1.37 into Eqn. 1.38 and integrating over spherical coordinates gives

$$\begin{aligned}\sigma = \frac{(\hbar c^2){G_F^o}^2}{2\pi}\frac{1}{(E_\chi+E_q)^2} \Big[&(c_V^q + c_A^q)^2 (E_\chi E_q + |\mathbf{p}|^2)^2 + (c_V^q - c_A^q)^2 (E_\chi^2 E_q^2 + \frac{|\mathbf{p}|^4}{3}) \\ &- \left({c_V^q}^2 - {c_A^q}^2\right) m_q^2 c^2 E_\chi^2 \Big] .\end{aligned} \tag{1.39}$$

Since WIMPs are nonrelativistic, the result can be further simplified by considering only the limit of slow-moving particles, in which case

$$\begin{cases} c|\mathbf{p}_i| \ll E_k & \forall i,k \\ E_i \approx m_i c^2 & \forall i \end{cases} . \tag{1.40}$$

This gives a result for elastic nonrelativistic DM-quark scattering of

$$\sigma = \frac{1}{\hbar^4}\frac{G_F^2}{2\pi}\left(\frac{m_\chi m_q}{m_\chi + m_q}\right)^2 \left(c_V^{q\,2} + 3c_A^{q\,2}\right) . \tag{1.41}$$

Although calculated in the CM frame, this expression is a total cross-section without explicit momentum dependence, and hence it is a frame-invariant result.

Scattering Off a Nucleus

To generalize the "heavy neutrino" scattering result from a single quark to a nucleus will require addressing the three quark-specific quantities in Eqn. 1.41: m_q, c_V^q, and c_A^q. In the low-energy-transfer limit, the dark matter interacts with the nucleus coherently – the nucleus appears as a single point particle so far as the interaction is concerned, and the entire nucleus recoils as one. Therefore, the first modification of the quark scattering cross section is to replace the quark mass m_q by the total nuclear mass m_N. The total couplings c_V^N and c_A^N of the nucleus can be found, in the coherent limit, by summing the couplings of the constituent quarks. By summing the couplings themselves, rather than summing their squares, we are effectively adding amplitudes for scattering off of all the possible scattering centers in the nucleus.

$$\begin{aligned} c_V^u &= \frac{1}{2} - \frac{4}{3}\sin^2\theta_w \\ c_V^d &= -\frac{1}{2} + \frac{2}{3}\sin^2\theta_w \\ c_A^u &= \frac{1}{2} \\ c_A^d &= -\frac{1}{2} \end{aligned} \tag{1.42}$$

where θ_w is the weak mixing angle.

One subtlety that must be considered relates to the spins of the quarks within the nucleus. Does scattering off a spin-up quark interfere constructively or destructively with scattering off a spin-down quark? If it's the former, the coupling constants can be added as-is; if the latter then appropriate negative signs must be introduced. The answer to that question is found by calculating the quark current for spin-up and spin-down quarks. In the zero-velocity limit, the bispinors for spin-up and spin-down quarks are, respectively,

$$u_\uparrow = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} , \; u_\downarrow = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \tag{1.43}$$

leaving off an overall multiplicative constant for simplicity. We define T^i to be everything sandwiched between the outgoing and incoming quarks' bispinors in the matrix element,

$\bar{u}_\uparrow T^0 u_\uparrow = c_V$	$\bar{u}_\downarrow T^0 u_\downarrow = c_V$
$\bar{u}_\uparrow T^1 u_\uparrow = 0$	$\bar{u}_\downarrow T^1 u_\downarrow = 0$
$\bar{u}_\uparrow T^2 u_\uparrow = 0$	$\bar{u}_\downarrow T^2 u_\downarrow = 0$
$\bar{u}_\uparrow T^3 u_\uparrow = -c_A$	$\bar{u}_\downarrow T^3 u_\downarrow = c_A$

Table 1.4: Quarks currents for neutral weak scattering.

that is,

$$T^i \equiv \gamma^i \left(c_V - c_A \gamma^5\right) . \tag{1.44}$$

The outcome of the calculation, shown in Table 1.4, is that vector couplings add regardless of quark spin, but the axial vector couplings of two quarks of opposite spin cancel each other out. Cases where the incoming and outgoing quark have the same spin are the only ones included because we are assuming the scattering is elastic.[17]

In the case of vector coupling, we proceed to simply add up the quarks' vector coupling constants. The valence quark content of the proton is *uud* and that of the neutron is *udd*, so for a nucleus with Z protons and $A - Z$ neutrons

$$c_V^N = -\frac{1}{2}(A - Z) + \frac{1}{2}(1 - 4\sin^2\theta_w)Z \tag{1.45}$$

Since there is no net vector contribution from sea quarks or gluons[90], no further adjustment to the vector coupling constant is needed.

As for axial coupling, the spin dependence results in a coupling that that is nonzero only if the nucleus has a net spin, and even then is negligible compared to the vector component. An accurate calculation of the spin-dependent coupling requires detailed calculations of the spin structure of the nucleus and will not be undertaken here.

A heavy neutrino has both a spin-independent vector coupling and a spin-dependent axial vector coupling, but the later is negligible compared to the former. Neglecting the axial coupling, and substituting in the new nucleus-appropriate expressions for mass and vector coupling, we reach the final result. The cross-section for elastic scattering of heavy neutrino dark matter off of a nucleus is

$$\sigma = \frac{1}{\hbar^4}\frac{G_F^2}{2\pi}\left(\frac{m_\chi m_N}{m_\chi + m_N}\right)^2 \left(-\frac{1}{2}(A - Z) + \frac{1}{2}(1 - 4\sin^2\theta_w)Z\right)^2 . \tag{1.46}$$

After the leading constants, the first term is the square of the reduced mass of the nucleus and dark matter particle, and the second term is roughly proportional to the square of the number of neutrons.

[17]The cancellation of axial couplings can also be shown by repeating the one-quark cross-section calculation but substituting a sum of spin-up and spin-down quarks for the single quark used there. That approach would take care of the negative signs automatically, but at the expense of a more verbose calculation.

1.3.3 Normalized Cross-sections, and the Need for New Models

As derived above, the cross-section for heavy neutrino dark matter is

$$\sigma = \frac{1}{\hbar^4}\frac{G_F^2}{2\pi}\left(\frac{m_\chi m_N}{m_\chi + m_N}\right)^2 \left(\left(c_V^N\right)^2 + 3\left(c_A^N\right)\right)^2 \tag{1.47}$$

where

$$c_V^N = -\frac{1}{2}(A - Z) + \frac{1}{2}(1 - 4\sin^2\theta_w)Z\,. \tag{1.48}$$

Here, c_V^N gives the strength of the vector coupling, which is a spin-independent interaction. The analogous c_A^N gives the strength of the axial vector coupling, which is a spin-dependent interaction. For heavy neutrinos scattering off of large nuclei, c_A^N is much smaller than c_V^N. However, only Dirac heavy neutrinos have both terms; Majorana heavy neutrinos would have only an axial term[121].

When comparing various theoretical models and experimental limits for spin-independent dark matter, it is customary to express limits that are "normalized to a nucleon." The normalized cross-section is defined as

$$\sigma_{\text{norm}} \equiv \frac{\sigma}{A^2}\left(\frac{\mu_n}{\mu_N}\right)^2 \approx \frac{\sigma}{A^2}\left(\frac{(m_\chi + m_N)m_n}{m_\chi m_N}\right)^2 \tag{1.49}$$

where A is atomic mass; μ_n is the reduced mass of the dark matter and a nucleon; μ_N is the reduced mass of the dark matter and a nucleus; and m_χ, m_n, and m_N are the masses of the dark matter, a nucleon, and a nucleus, respectively. This removes most or all of the dependence on the choice of target nucleus. Doing so is advantageous because it allows for a fair comparison of spin-independent dark matter limits set with different target materials.[69, 101]

For heavy neutrinos, the normalized cross-section approximately assumes the simple form

$$\sigma_{\text{norm}} \approx \frac{1}{\hbar^4}\frac{G_F^2 m_n^2}{32\pi}(2r)^2 = 4.6 \times 10^{-40}\text{cm}^2(2r)^2 \tag{1.50}$$

where r is defined to be the fraction of nucleons that are neutrons.

The physics of neutral weak interactions of heavy neutrinos were worked out in the mid-1980s[78], and by the end of that decade heavy neutrinos were ruled out as a viable dark matter candidate[90]. The major limits came from germanium detectors operated in the Homestake Mine in South Dakota[10] and the Oroville Dam power station in California[51, 50]. The heavy neutrino model only has one quantitative free parameter (the neutrino mass), and the germanium detector limits along with other particle and cosmological considerations ruled out the mass range of interest. But at the same time, another theory that could provide a dark matter candidate was on the rise: supersymmetry.

1.3.4 Supersymmetry

Supersymmetry (SUSY) is an extension to the Standard Model (SM) that proposes a new symmetry between bosons and fermions. For every known SM boson there is a "super-partner" fermion, and vice versa. A feature of this hypothesis is that it can resolve the so-called Hierarchy Problem. The observed mass of each fundamental particle is modified by corrections due to interactions with other particles. Where these corrections diverge, the divergences are typically logarithmic and are addressed through renormalization. For the Higgs boson, however, the problem is more acute. The mass corrections grow *quadratically* with the masses of other particles that couple to the Higgs or with the energy cut-off of the theory. This makes the Higgs mass very sensitive to the masses of all other particles, including any that have not yet been discovered. Even if a novel particle did not couple to the Higgs directly, it would still couple indirectly if it coupled to any gauge bosons. In the absence of fine-tuning, all these corrections should drive up the Higgs mass to be around the energy cutoff of the theory, presumably in the neighborhood of the Planck mass, 10^{19} GeV. The mass corrections due to fermions are negative and those due to bosons are positive, suggesting a solution: if every SM boson has a superpartner fermion and vice versa, then their contributions will cancel out. In an exact supersymmetry, the SM particles and their superpartners would be of equal mass. The non-detection of superpartners thus far means that if they exist they must be of higher mass, requiring supersymmetry to be a broken symmetry.[105] Aside from solving the Hierarchy Problem, supersymmetry might also provide a route to unification of SM physics with gravity. But if the symmetry is very broken, then fine-tuning problems reemerge, and LHC results continue to rule out possible phase space.[118]

Supersymmetry introduces a conserved multiplicative quantum number, known as R-parity, in order to preserve SM baryon and lepton number conservation to within stringent experimental limits. All SM particles have even R-parity and all superpartners have odd R-parity. Because R-parity is conserved, any interaction among SM particles that creates superpartners would have to create them in pairs. More importantly for our purposes, the lightest superpartner, also called the lightest supersymmetric particle (LSP), would be stable against decay, making it an ideal WIMP candidate.[118]

The Minimal Supersymmetric Standard Model (MSSM) is the most direct viable implementation of the SUSY concept. The names and spins of the particles in MSSM are given in Table 1.5. The standard naming convention is that bosonic superpartners prepend an "s" to the particle name and fermionic superpartners append an "ino." There are some 124 free parameters in the MSSM, and the nature and nuclear scattering cross-section of the LSP depend on the choice of these parameters. One option is that the LSP could be the lightest of the four neutralinos, which are mass eigenstates that are superpositions of the gauginos (winos and bino) and higgsino. Another option, sneutrino WIMPs, has been effectively ruled out. Also, more complicated implementations of SUSY can introduce even more possibilities. Supergravity theories, for example, allow for WIMPs that are gravitinos[118].

The simplest, original supersymmetry-based calculations suggested nuclear scattering

SM Particles	Spin	SUSY Superpartners	Spin
quark	$\frac{1}{2}$	squark	0
lepton	$\frac{1}{2}$	slepton	0
Higgs	0	higgsino	$\frac{1}{2}$
gluon	1	gluino	$\frac{1}{2}$
$W^{(+,-,0)}$	1	wino	$\frac{1}{2}$
B	1	bino	$\frac{1}{2}$

Table 1.5: Fields in the Minimal Supersymmetric Standard Model (MSSM)[105]. Note that the Z boson and the photon are superpositions of the W^0 and B fields, and their corresponding superpartners are the zino and photino, respectively. Not included are the hypothetical spin-2 graviton and spin-3/2 gravitino of supergravity theory.

cross-sections very similar to the heavy neutrino model.[18] The sneutrino scattering cross-section would be exactly four times the heavy neutrino vector cross-section, and the sneutrino's axial term would be as negligible as the heavy neutrino's, meaning[121, 101]

$$\sigma_{\tilde{\nu}} = \frac{1}{\hbar^4}\frac{2G_F^2}{\pi}\left(\frac{m_\chi m_N}{m_\chi + m_N}\right)^2 \left(c_V^N\right)^2 . \tag{1.51}$$

Meanwhile, the cross-section for pure higgsinos would differ only by a multiplicative constant from the heavy neutrino axial coupling,[121]

$$\sigma_{\tilde{H}} = \frac{1}{\hbar^4}\frac{6G_F^2}{\pi}\left(\frac{m_\chi m_N}{m_\chi + m_N}\right)^2 \left(c_A^N\right)^2 \cos^2(2\beta)\,, \tag{1.52}$$

where β is related to the relative masses of the various Higgs bosons in the supersymmetric theory.

However, this is only the beginning of the story for LSP WIMPs. By varying the many free parameters of the MSSM, a vast range of LSP cross-sections and masses can be achieved. One approach to this task is to vary some free parameters (while fixing others as constants or simple functions of the varied parameters) and note the LSP mass and nuclear scattering cross-section in each case. One can then generate a density plot of the masses and cross-sections that result from varying the free parameters over plausible ranges, leaving out possibilities that have already been ruled out by accelerator or direct detection experiments. A useful simplified model for this purpose is the Constrained MSSM (CMSSM), with only six free parameters. Another is the phenomenological MSSM (pMSSM), variations of which have either 10 or 19 free parameters.[118] Fig. 1.12 shows the masses and normalized cross-sections for some typical spin-independent examples.

[18]Historically, supersymmetric dark matter candidates were being considered contemporaneously with heavy neutrinos; Goodman and Witten's original paper on WIMP neutral weak scattering[78] discusses possible superpartner candidates for dark matter.

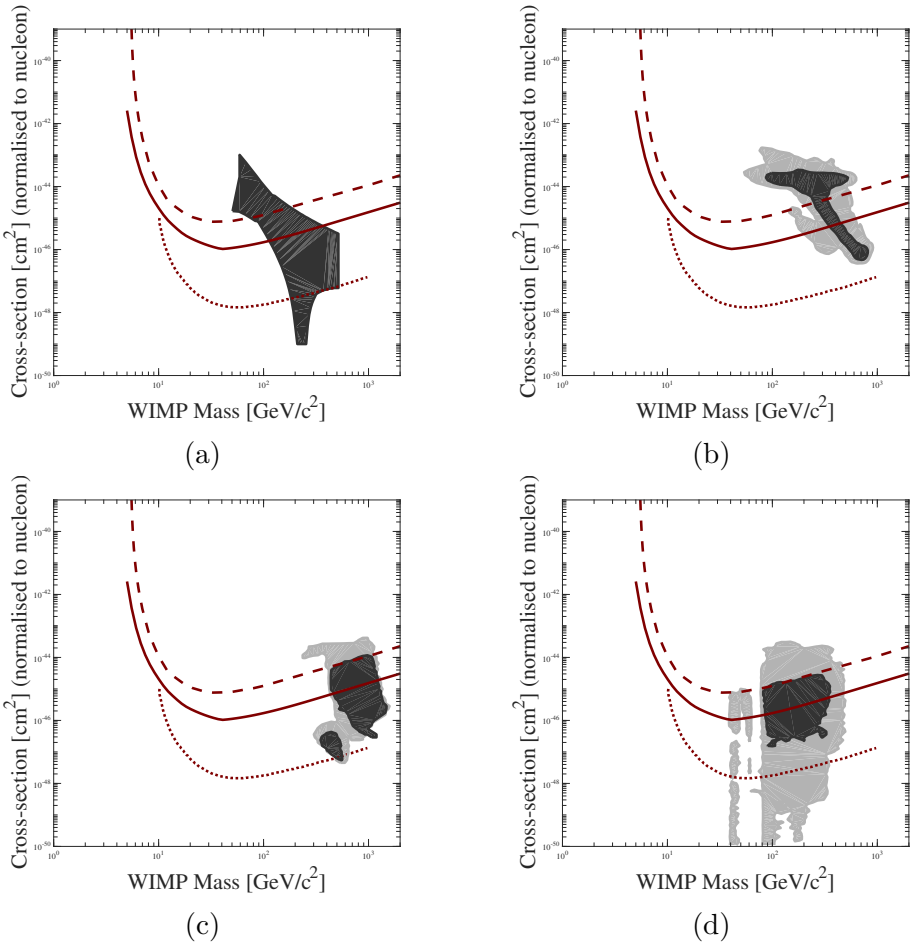


Figure 1.12: Spin-independent supersymmetric WIMP mass and cross-section phase space. Plots (a), (b), and (c) are three CMSSM analyses produced in 2000, 2008, and 2015, respectively, based on different data and assumptions, while plot (d) is a pMSSM plot based on the same recent data as plot (c). (a) [66], (b) 68% and 95% contours from [140], (c) & (d) 68% and 95% contours from [41]. The reference lines shown, from top to bottom, are the LUX Run 3 result (dashed) [14], the LUX Run 3+4 result (solid) [20], and an LZ projected result (dotted) [60]. Plots generated with DMTools[63].

1.3.5 Effective Field Theory

In the absence of any incontrovertible dark matter direct detection, attention has been garnered by a new approach to limit-setting that can complement the placing of limits on specific theoretical models. In the effective field theory (EFT) approach, one instead places limits on the coupling constants of all the different ways a dark matter particle can couple to a nucleon. In the zero-velocity limit that's easy enough, because in that limit any interaction reduces to either spin-independent coupling, spin-dependent coupling, or zero. But to include the full range of possibilities, it is useful to consider even interactions that do go to zero in the zero-velocity limit.

Fitzpatrick et al.[74] have worked out the EFT for elastic dark matter scattering in the nonrelativistic (but *not* necessarily low-momentum-transfer or low-velocity) limit. Their calculations cover interactions up to second order in momentum transfer mediated by particles of spin 1 or less. Interactions are described in terms of four Galilean-invariant quantities. Two of these are the spin of the nucleon, $\mathbf{S}_n$, and the spin of the dark matter particle, $\mathbf{S}_\chi$.[19] The other two are the momentum transfer, $\mathbf{q}$, and the relative velocity before scattering, $\mathbf{v}$. For these last two, it is more convenient to work in terms of quantities that lend themselves to building Hamiltonians that are Hermitian, so we use the quantities $i\mathbf{q}$ and

$$\mathbf{v}^\perp \equiv \mathbf{v} + \frac{\mathbf{q}}{2\mu_n} \tag{1.53}$$

where μ_n is the reduced mass of the dark matter and nucleon. The most general interaction Lagrangian for nonrelativistic elastic dark matter scattering is

$$\mathcal{L} = \sum_{n=\text{proton,neutron}} \sum_i c_i^{(n)} \mathcal{O}_i \chi^+ \chi^- n^+ n^- \,. \tag{1.54}$$

The $\mathcal{O}_i$ term is one of 14 different possible couplings, each of which is some combination of $\mathbf{S}_n$, $\mathbf{S}_\chi$, $i\mathbf{q}$, and/or $\mathbf{v}^\perp$. The first summation allows for the dark matter to couple to protons differently than to neutrons.

This same approach can be extended to the case of inelastic dark matter scattering. Inelastic dark matter[133, 49] is a model of dark matter scattering wherein the outgoing dark matter particle has a different rest mass than the incoming particle. One way this could happen is if the interaction changed one type of dark matter particle into a different type. Another way this could happen is if the dark matter is a composite particle that leaves the interaction in an excited state.

Building an EFT for inelastic scattering[43] requires surprisingly little modification of the elastic scattering EFT. The main difference is an additional term in the definition of $\mathbf{v}^\perp$. This additional term is proportional to the mass difference δ between the outgoing and income dark matter particles.[20] Inelastic EFT covers a wide swath of possible interactions. Spin-independent dark matter, spin-dependent dark matter, elastic EFT dark matter, standard

[19] For spin-0 dark matter the latter is trivially zero, but the following formalism still applies.

[20] Note that δ can in principle be negative.

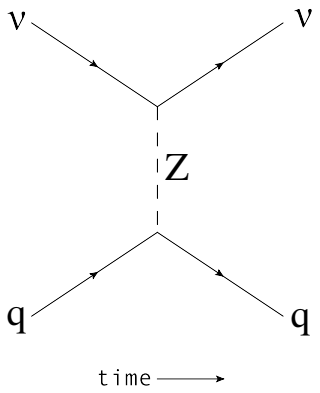


Figure 1.13: Feynman diagram of coherent neutrino scattering.

inelastic dark matter, and isospin-violating dark matter are all special cases subsumed by the inelastic EFT framework.

1.3.6 Coherent Neutrino Scattering

As advances in WIMP direct detection enable searches for ever-smaller WIMP-nucleus scattering cross-sections, an irreducible background will eventually pose a challenge to further progress. Coherent neutrino scattering (CNS, also called coherent elastic neutrino-nucleus scattering) is the scattering of neutrinos off of nuclei via Z-boson exchange (Fig. 1.13). Not surprisingly, the cross-section for this process looks much like the heavy neutrino cross-section calculated previously (Eqn. 1.46). The only differences for the Standard Model neutrinos that cause CNS are a factor of two and the replacement of the system's reduced mass by the neutrino's energy, like so:[64]

$$\sigma = \frac{1}{(\hbar c)^4}\frac{G_F^2}{\pi}E_\nu^2\left(-\frac{1}{2}(A-Z)+\frac{1}{2}(1-4\sin^2\theta_w)Z\right)^2 . \tag{1.55}$$

Like spin-independent WIMP scattering, CNS has a cross-section that scales with the square of the target's atomic mass. That is why it is called "coherent."

Recently, the first direct observation of CNS was made by using a novel approach. Instead of using a dark matter experiment or similar large detector, CNS can be measured by setting up a smaller detector in a high-neutrino environment. Ideal environments for such an experiment would be as close as possible to either a fission reactor or a stopped pion source like the Oak Ridge Spallation Neutron Source[26]. It was at the latter that CNS was detected, by using a Na-doped CsI detector[27].

Because a nuclear recoil looks the same whether it's caused by WIMP scattering or CNS, CNS events are indistinguishable from WIMP events in a detector. The exact detector sensitivity where CNS becomes a problem depends on the WIMP mass one wishes to set a limit on. As shown in Fig. 1.14, a higher background affecting low-mass spin-independent WIMP studies comes from solar neutrinos, while a lower background affecting

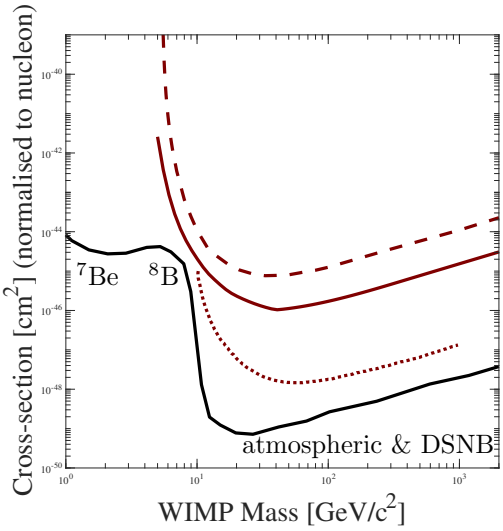


Figure 1.14: Anticipated background for spin-independent WIMP direct detection experiments due to coherent neutrino scattering (black) [47]. The regions of the neutrino spectrum are labeled; "DSNB" stands for diffuse supernova neutrino background. The reference lines shown, from top to bottom, are the LUX Run 3 result (dashed) [14], the LUX Run 3+4 result (solid) [20], and an LZ projected result (dotted) [60]. Plot generated with DMTools[63].

high-mass spin-independent WIMP studies comes from atmospheric and supernova neutrinos[60]. Spin-dependent WIMP searches will also eventually run up against the CNS background. Searching for WIMPs with cross-sections below the CNS floor is theoretically possible by using a direction-sensitive detector or by looking for small time variations in the signal, but these approaches present immense experimental challenges.

1.3.7 Experimental Limits

So far, we have discussed a variety of theoretical dark matter scattering cross-sections. In order to set a cross-section limit based on experimental data, it is necessary to be able to convert between cross-section and event rate in a detector, a topic we will now explore.

The recoil spectrum for spin-independent dark matter scattering off a target can be derived following the approach of Ref. [101]. We consider only the case where the DM-nucleus scattering cross-section σ_0 varies slightly or not at all with velocity.[21] In that case, the scattering rate in units of recoils per target mass per time is

$$R_0 = \frac{N_0}{A}\frac{\rho_\chi}{m_\chi}\sigma_0\langle v\rangle \tag{1.56}$$

[21]This condition is met by heavy neutrinos and typical supersymmetric models, but *not* by most terms of elastic effective field theory dark matter and inelastic EFT dark matter.

where N_0 is Avogadro's number, A is the target's atomic mass, ρ_χ is the dark matter mass density in the vicinity of Earth, m_χ is the mass of a dark matter particle, and $\langle v \rangle$ is the average velocity of dark matter relative to the target. Working in the low-velocity limit, we neglect any actual velocity dependence in the cross-section formula. The average velocity is

$$\langle v \rangle = \int v f(\mathbf{v}, \mathbf{v}_E) d\mathbf{v} \tag{1.57}$$

where $f(\mathbf{v}, \mathbf{v}_E)$ is the probability density function giving the probability per unit $d\mathbf{v}$ that a dark matter particle has a velocity in the vicinity of $\mathbf{v}$. In general this can be a complicated function of the Earth's velocity through the dark matter halo $\mathbf{v}_E$ and other variables. The simplest case is a Maxwell distribution with characteristic speed v_0, in which case the integral evaluates to $\frac{2}{\sqrt{\pi}} v_0$. Other modifications are discussed below.

Combining Eqns. 1.56 and 1.57 and expressing the overall rate as a derivative with respect to $\mathbf{v}$ gives

$$\frac{dR_0}{d\mathbf{v}} = \frac{N_0}{A} \frac{\rho_\chi}{m_\chi} \sigma_0 v f(\mathbf{v}, \mathbf{v}_E) \,. \tag{1.58}$$

The next step is to develop a recoil energy spectrum in term of the overall event rate R_0. Consider a WIMP of mass m_χ scattering off of a target nucleus of mass m_N, such that in the center-of-mass frame each particle's path is deflected by an angle θ. The momentum transfer, meaning the change in momentum of each particle, is

$$q = p_i^{\mathrm{CM}} \sqrt{2 - 2\cos(\theta)} \tag{1.59}$$

where p_i^{CM} is the initial momentum of either particle in the CM frame. We can express p_i^{CM} in terms of variables in the lab frame, that is, the target's rest frame. If v_χ is the initial WIMP velocity in the lab frame, then

$$q = m_\chi v_\chi \left(\frac{m_N}{m_\chi + m_N} \right) \sqrt{2 - 2\cos(\theta)} \,. \tag{1.60}$$

A change in momentum is a Galilean-invariant quantity. In the lab frame, the nucleus starts at rest, so its recoil energy E_R is just $q^2/(2m_N)$, or

$$E_R = \left(\frac{1}{2} m_\chi v_\chi^2 \right) \left(\frac{4 m_\chi m_N}{(m_\chi + m_N)^2} \right) \left(\frac{1 - \cos\theta}{2} \right) \,. \tag{1.61}$$

The first term in parentheses is the incoming WIMP kinetic energy E. The second term in parentheses is a constant, which we will call r. The third term in parentheses, which varies from 0 to 1, we'll call a. So the recoil energy E_R is $E \cdot r \cdot a$. Note that this equation mixes variables from two different reference frames: v_χ is the lab frame velocity, while θ is the center-of-mass scattering angle.

To proceed, we make a physical assumption and assert that dark matter scatters isotropically in the center-of-mass frame. This condition is satisfied automatically in the low-momentum-transfer limit, in which case the Born approximation indicates that only S-wave

scattering occurs. However, the condition is also satisfied even at higher momentum transfers by interactions mediated by heavy particles[126], as in many WIMP models.[22] Given this assumption, the recoil angular spectrum will be uniform in $\cos\theta$ and consequently will also be uniform in a. Since recoil energy is determined by E and a, this provides the information needed to work out the rate with respect to recoil energy:

$$\frac{dR}{dE_R} = \int_{E_{\min}}^{E_{\max}} \frac{d^2R}{dEda}\frac{da}{dE_R}dE = \int_{E_{\min}}^{E_{\max}} \frac{dR}{dE}\frac{da}{dE_R}dE = \int_{E_{\min}}^{E_{\max}} \frac{1}{Er}\frac{dR}{dE}dE = \int \frac{1}{Er}\frac{dR}{d\mathbf{v}}d\mathbf{v}\,. \tag{1.62}$$

Substituting in the differential rate with respect to velocity derived in Eqn. 1.58 gives a recoil spectrum of

$$\frac{dR}{dE_R} = \int \frac{1}{Er}\frac{N_0}{A}\frac{\rho_\chi}{m_\chi}\sigma_0 v f(\mathbf{v},\mathbf{v}_E)d\mathbf{v}\,. \tag{1.63}$$

With some algebra this simplifies to

$$\frac{dR}{dE_R} = N_0\frac{\rho_\chi}{m_\chi}\frac{m_n\sigma_0}{2\mu_{\chi,N}^2}\int \frac{f(\mathbf{v},\mathbf{v}_E)}{v}d\mathbf{v} \tag{1.64}$$

where m_n is the nucleon mass, $\mu_{\chi,N}^2$ is the reduced mass of the WIMP and nucleus, and all other quantities are as defined above. The bounds of the integral are understood to exclude velocities too low to permit recoils of the desired E_R.

Normalizing to a Nucleon

Our recoil spectrum (Eqn. 1.64) has so far been given in terms of the cross-section for scattering off a nucleus, but it is conventional to express the spin-independent recoil spectrum in terms of the cross-section normalized to a nucleon. Using our definition of the normalized cross-section σ_{norm} given in Eqn. 1.49, the recoil spectrum is

$$\frac{dR}{dE_R} = N_0\frac{\rho_\chi}{m_\chi}\frac{m_n\sigma_{\text{norm}}A^2}{2\mu_{\chi,n}^2}\int \frac{f(\mathbf{v},\mathbf{v}_E)}{v}d\mathbf{v} \tag{1.65}$$

where $\mu_{\chi,n}$ is the reduced mass of the WIMP and nucleon.

It should be emphasized that this recoil spectrum assumes a velocity-independent cross-section and isotropic scattering in the CM frame. As we will see in a subsequent chapter, models that do not meet these assumptions can have very different recoil spectra. However, the following discussions of event rate, form factor, and velocity distribution hold generally, even for recoil spectra that behave differently from Eqn. 1.65.

[22] In the case of the heavy neutrino, angular dependence appears in the form of terms like $E_1E_2 \pm \left(p_i^{\text{CM}}\right)^2\cos\theta$, where the E's are CM energies (including rest mass) of either the WIMP or target, while p_i^{CM} is the CM momentum of each particle. Therefore, the angular-dependent contribution is dwarfed by the isotropic contribution for nonrelativistic interactions.

Event Rate

A nuclear recoil only transfers some of its energy to electrons (as ionization and excitation), with the rest being lost to phonons. For detectors like LUX that only detect the energy transferred to electrons, it is useful to work in terms of "ee," or electron-equivalent energy. If we define E_v to be the visible, electron-equivalent energy, then

$$E_v = f_n E_R \tag{1.66}$$

where E_R is the nuclear recoil energy as before and f_n is the relative efficiency for converting nuclear recoil energy into ionization and electronic excitations measurable in LUX. Allowing for f_n to itself be E_R-dependent, the chain rule gives

$$\frac{dR}{dE_v} = \frac{\frac{dR}{dE_R}}{f_n + E_R \frac{df_n}{dE_R}} \,. \tag{1.67}$$

A well-known phenomenological model for f_n is the Lindhard model.[134, 101]

To calculate the expected number of events in a detector experiment, one need only integrate over the recoil spectrum, weighting it by a specific detector's detection efficiency G, which is energy-dependent. The expectation number of events N is then[48]

$$N = MT \int G(E) \frac{dR}{dE} dE \tag{1.68}$$

where M is detector mass, T is livetime, and the equation is equally applicable for nuclear recoil energies or for electron-equivalents.

Form Factor

For non-negligible momentum transfer, the accuracy of the above recoil spectrum can be improved including a form factor to take into account the loss of coherent scattering off of different parts of the nucleus. The form factor is computed using the Born approximation.

$$F(q) = \int \rho(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \,, \tag{1.69}$$

where $\rho(\mathbf{r})$ is the nuclear density at position $\mathbf{r}$ and $\mathbf{q}$ is the momentum transfer. In the usual case of spherical symmetry, the expression simplifies to

$$F(q) = \frac{4\pi}{q} \int_0^\infty r \sin(qr) \rho(r) dr \,. \tag{1.70}$$

One can use nuclear models of varying sophistication. One of the simplest models is to treat the density of scattering centers in the nucleus as uniform out to a radius r_0. In this hard sphere model, the form factor is

$$F(q) = \left. \frac{3(\sin x - x \cos x)}{x^3} \right|_{x=qr_0} \,. \tag{1.71}$$

A more realistic approach is using a Fermi distribution to smooth out the edge of the hard sphere. An exact Fermi distribution does not yield an analytic solution for the form factor, but there are similarly-shaped functions that do have analytic solutions. One of these leads to the analytic Helm form factor.[101]

However the form factor is calculated, its application is simple: The recoil spectrum scales with the square of the form factor.

$$\frac{dR}{dE_R} = N_0 \frac{\rho_\chi}{m_\chi} \frac{m_n \sigma_{\rm norm} A^2}{2\mu_{\chi,n}^2} F^2(q) \int \frac{f(\mathbf{v}, \mathbf{v}_E)}{v} d\mathbf{v} \,. \tag{1.72}$$

Note that q is related to E_R, so this term changes the shape of the recoil spectrum.

Density and Velocity Distribution

Two pieces of astrophysical information needed to convert a direct detection experimental result into a statement about cross-sections are the local dark matter density ρ_χ and the local dark matter velocity distribution $f(\mathbf{v})$.

The local dark matter density is much greater than the overall dark matter density of the universe, because dark matter is concentrated in galactic halos. Various models of the halos' radial density profiles have been proposed. The simplest is the pseudo-isothermal halo model, where

$$\rho(r) = \frac{\rho_0}{1 + \left(\frac{r}{r_c}\right)^2} \tag{1.73}$$

and each galactic halo is characterized by the two parameters ρ_0 and r_c. The Navarro-Frenk-White model is a similar two-parameter-per-halo model that adds an extra $1/r$ dependence. And there are many other models still[107].

For the purpose of direct-detection limit setting, the full density profile is not needed, as only the dark matter density near Earth is used. When reporting limits, this density is conventionally taken to be 0.3 GeV/cm^3 (as in, for example, Ref. [14]), although the true value may be closer to 0.4 GeV/cm^3 [54].

Velocity distribution models come in varying levels of sophistication. As noted previously, the simplest is a Maxwell distribution

$$f(\mathbf{v}) \propto e^{-\mathbf{v}^2/v_0^2} \tag{1.74}$$

where the characteristic velocity v_0 is around 220 km/s.[23] One way to build a more realistic model is to add in the velocity of Earth relative to the rest frame of the galaxy. The Earth's velocity, $\mathbf{v}_E$, has a magnitude of approximately 245 km/s, with a 6% annual modulation due to the Earth's orbit. Another improvement is to impose a cutoff at the galactic escape

[23]This equation is expressed as a proportionality so as to leave out the normalization constant, which gets ungainly in the following example.

velocity. The escape velocity $v_{\rm esc}$ for particles near Earth is around 544 km/s.[101, 14] With these modifications the velocity distribution becomes

$$f(\mathbf{v}) \propto \begin{cases} e^{-(\mathbf{v}+\mathbf{v_E})^2/v_0^2} & v < v_{\rm esc} \\ 0 & v \geq v_{\rm esc} \end{cases} \tag{1.75}$$

Other approaches involve replacing the Maxwell distribution with different functions[71].

1.4 WIMP Direct Detection Experiments

Dozens of collaborations are now engaged in the task of developing WIMP direct detection experiments. Such experiments are designed to look for WIMPs scattering off of atomic nuclei in a target mass, instigating nuclear recoils. The energy of the recoiling nucleus is ultimately divided among three main channels. One of these is ionization, where electrons freed from their atoms can be collected and counted. Another is scintillation, where excited atoms de-excite and release photons in the process. Finally, some energy isn't transferred to these electronic processes at all, but instead is transferred to the material as heat. If the material's temperature is cold enough, it is even possible to measure individual phonons of deposited thermal energy. The interplay among these processes is complex. For example, an important process is recombination, wherein a freed electron is captured on another atom, turning a would-be quantum of ionization into a scintillation signal instead. An important background in direct detection experiments is electron recoils, which can be caused by ordinary β and γ radiation. Fortunately, nuclear recoils (NR) and electron recoils (ER) tend to distribute their energy differently among the various channels (ionization, scintillation, and heat), providing a discrimination mechanism. For detectors that measure two of the three energy inputs, this provides the key to distinguishing ER and NR events. Among detectors that measure only one of the three energy inputs, other approaches must be taken.

We will briefly survey current dark matter experiments, sorting them by which types of energy deposition (again, ionization, scintillation, and/or heat) that they measure. The state of the field is summarized in Fig. 1.15, where the collaborations are sorted by what they measure and color-coded by target medium. Note that many collaborations have built multiple detectors. For example, "XENON" in the figure denotes the XENON Collaboration, responsible for completed experiments with the XENON10 and XENON100 detectors, an ongoing experiment with the XENON1T detector, and development work towards an aspirational multi-ton XENONnT.

1.4.1 Two-Channel Discrimination

The most straightforward way to distinguish nuclear recoils from electron recoils is to measure two different types of energy deposition, since NRs and ERs distribute energy differently.

One approach is to measure both scintillation light and ionization electrons. This can be done using a two-phase noble liquid/gas detector, a technique that will be discussed in detail

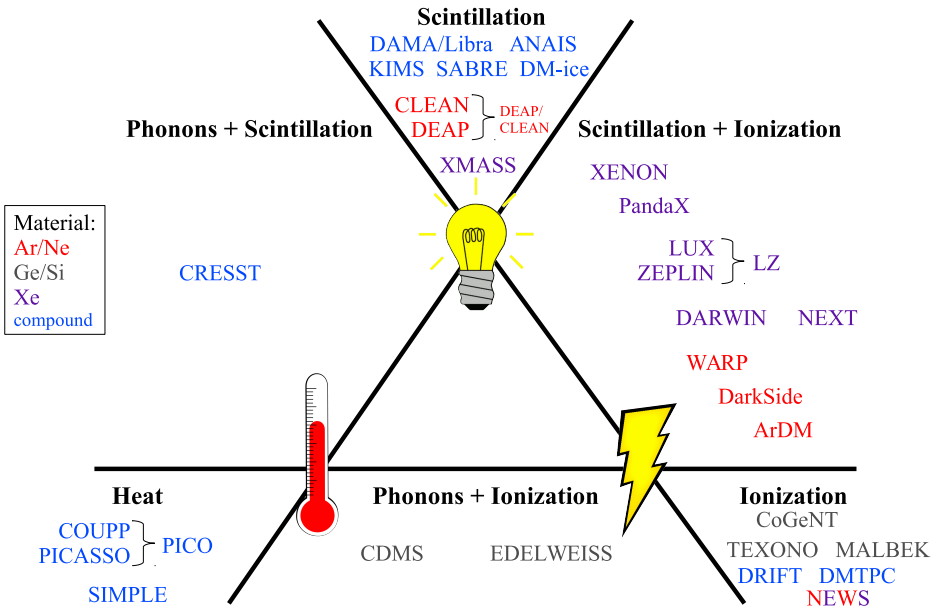


Figure 1.15: Major current, planned, and recent dark matter direct detection experimental collaborations, sorted by detection methods. Experiments in boxes adjacent to the light-bulb detect scintillation; those in boxes next to the lightning bolt detect ionization; and those next to the thermometer detect phonons or other heat-related effects. Braces indicate collaborations that have merged.

in the next chapter in the context of the Large Underground Xenon (LUX) experiment. Other dual-phase detectors that, like LUX, use a xenon target include those by XENON[38] and PandaX[137]. PandaX reported a spin-independent limit almost as strong as LUX's world-leading result in 2016, while XENON subsequently surpassed it with their latest detector[37]. The LUX and ZEPLIN collaborations have joined to form the LZ collaboration[109], which will build a multi-ton dual-phase xenon detector. There have been other proposals for future multiton detectors, such as one by DARk matter WImp search with liquid xenoN (DARWIN) collaboration[1]. Instead of xenon, argon can be used as a target material, as is the case for DarkSide[7], the WIMP Argon Programme (WARP)[144], and Argon Dark Matter (ArDM)[52].

Another approach is to measure both scintillation light and phonons of thermal energy. This is the approach of the Cryogenic Rare Event Search with Superconducting Thermometers (CRESST) collaboration, which uses a target of $CaWO_4$ at a temperature on the order of a milli-Kelvin. ER/NR discrimination is provided by the fact that ERs create more scintillation light than do NRs with the same amount of phonon energy. Although previously reporting a possible detection, CRESST itself has now ruled it out.[35]

The final way to use two-channel discrimination is to measure ionization and phonons, again requiring milli-Kelvin temperatures. For this semiconductor targets are used. EDEL-

WEISS[39] uses germanium, while the Cryogenic Dark Matter Search (CDMS)[8] mostly uses germanium and has also worked with silicon. ER/NR discrimination is provided by the fact that ERs create more ionization than do NRs with the same recoil energy. In CDMS's "iZIP" detector design, germanium crystals are cut into cylinders of 25mm thickness and 76mm diameter. The faces of the cylinders have both phonon and ionization sensors. The ionization sensors create an electric field to draw in charges. The phonon sensors are transition edge sensors (TES), containing a small amount of tungsten barely cold enough to be superconducting. The addition of a minuscule amount of heat therefore appreciably changes the tungsten's conductivity. That in turn changes the current flowing through the tungsten, creating a change in the circuit's magnetic field which is picked up and amplified by a superconducting quantum interference device (SQUID).[33]

1.4.2 One-Channel Discrimination

For detectors that only measure one type of deposited energy, other methods must be used to distinguish nuclear recoils that could be dark matter from electron recoils that are just background.

Among scintillation-only detectors, different experiments use different techniques to separate ER and NR events. Single-phase liquid argon detectors like DEAP[72] can use pulse shape discrimination, taking advantage of the fact that nuclear recoils populate a different distribution of excited states than do electron recoils, leading to different decay times. (Pulse shape discrimination in xenon, a more difficult experimental challenge, will be discussed in a subsequent chapter.) For NaI(Tl) scintillation detectors, the only method is to look for the annual modulation expected in the WIMP detection rate. The DAMA/LIBRA collaboration asserts a detected signal of this kind[45]. When the DAMA/LIBRA result was contradicted by experiments with other target media, attempts were made to theoretically justify the apparent contradiction (for an example, see Ref. [73]). However, the DAMA/LIBRA signal region has now been mostly ruled out by even the KIMS experiment, which also uses NaI(Tl) crystals[98]. This, along with contradictions and omissions in DAMA/LIBRA data[118], calls the positive result into question. Neutron backgrounds have been suggested as an alternative explanation[125].

Ionization-only detectors also use different techniques. In experiments like DRIFT[62] and DMTPC[9], a time projection chamber with a low-pressure gas target allows for imaging the recoil path, which not only enables ER/NR discrimination but also provides directional sensitivity. DRIFT uses a CS_2 target, while DMTPC uses CF_4. Also found in the ionization-only category are the high-purity germanium (HPGe) detectors, such as those deployed by the Coherent Germanium Neutrino Technology (CoGeNT) collaboration[2]. CoGeNT uses P-type point contact (PPC) detectors, which have lower capacitance and hence lower electronic noise than previous germanium detectors. Because Ge detectors like CoGeNT's lack a specific method to distinguish ERs and NRs, a very strong understanding of backgrounds is required. CoGeNT previously reported an excess of events beyond background, which may be due to surface events[118]. The MALBEK project[77] is a reminder of the technological overlap

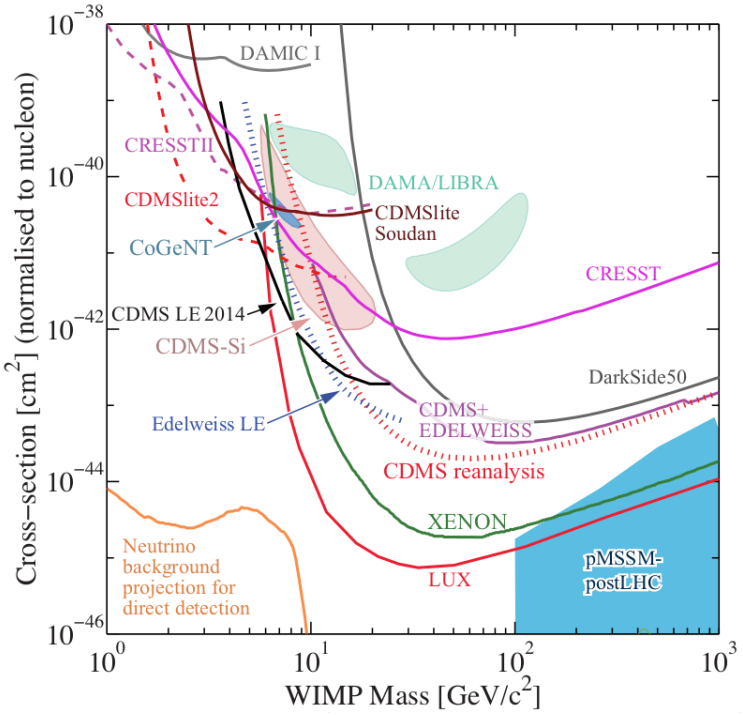


Figure 1.16: Spin-independent normalized cross-section limits and supposed detections versus WIMP mass for recent experiments. LUX's recent 300-day run is not included, but it gives a limit about 4 times stronger than the 85-day run, shown. Reprinted from [118].

between dark matter direct detection and other types of rare event searches; MALBEK uses germanium detectors developed for the MAJORANA neutrinoless double beta decay experiment.

Finally, dark matter detectors can work by measuring heat alone. This is done by building a bubble chamber tuned so that a recoil will only nucleate the formation of a bubble if the recoil's energy deposition and dE/dx both exceed certain thresholds. The dE/dx threshold prevents electron recoils from forming a bubble. The PICO collaboration's[31] 2-liter C_3F_8 bubble chamber has set the strongest direct detection limit on spin-dependent coupling to neutrons. Although stronger limits have been published by indirect detection experiments, those limits are more model-dependent[118].

This summary of detectors, although not exhaustive, shows the wide range of direct detection target types and NR/ER discrimination methods being developed. A selection of their spin-independent limits is shown in Fig. 1.16. Having now reviewed the cosmological and theoretical background for WIMP direct detection and some of the basics of the techniques, we will next look in detail at the Large Underground Xenon experiment.

Chapter 2

LUX Hardware and Operation

The Large Underground Xenon (LUX) experiment is a dark matter search based on direct detection. The LUX detector, a two-phase xenon time-projection chamber with a fiducial volume in excess of 100kg, is deployed underground to mitigate backgrounds. From 2013 through 2017, LUX placed world-leading limits on dark matter scattering cross-sections[14, 20]. In this chapter, we will discuss hardware and analysis techniques used for LUX. We start with a general introduction to LUX, before addressing hardware, signal processing, and modifications over the detector's lifetime. Information in this chapter, where not otherwise cited, comes from Ref. [24].

2.1 Detector Overview

LUX detects interactions of particles with the liquid xenon "target" at the heart of the detector. Energy deposited in an interaction is divided among three channels: scintillation light, ionized electrons, and phonons. Phonons, which are the dominant channel for nuclear recoils, are not detected by LUX, but the two electromagnetic signals are. The liquid xenon is surrounded by 122 photomultiplier tubes (PMTs): 61 immersed in the liquid xenon below the active volume, and another 61 in the xenon gas just above the liquid's surface. These PMTs immediately detect the pulse of light from the scintillation, the so-called "S1" signal. Most of this signal is measured by the lower PMTs because of total internal reflection at the xenon surface. Next, because an electric field is applied to the xenon, ionized electrons that avoid recombination are drifted away from the interaction site up towards the xenon's surface. An increased electric field in the vicinity of the surface drives the electrons to exit the liquid and energetically collide with xenon gas molecules, producing more light through electroluminescence. This second signal, the so-called "S2," is proportional to the number of ionization electrons. Because electron recoils and nuclear recoils have different S2 distributions for a given S1 size, this dual-channel method provides a tool for distinguishing the two populations. What's more, it permits a full three-dimensional reconstruction of where in the detector the event occurred. The X-Y coordinates of the event are determined

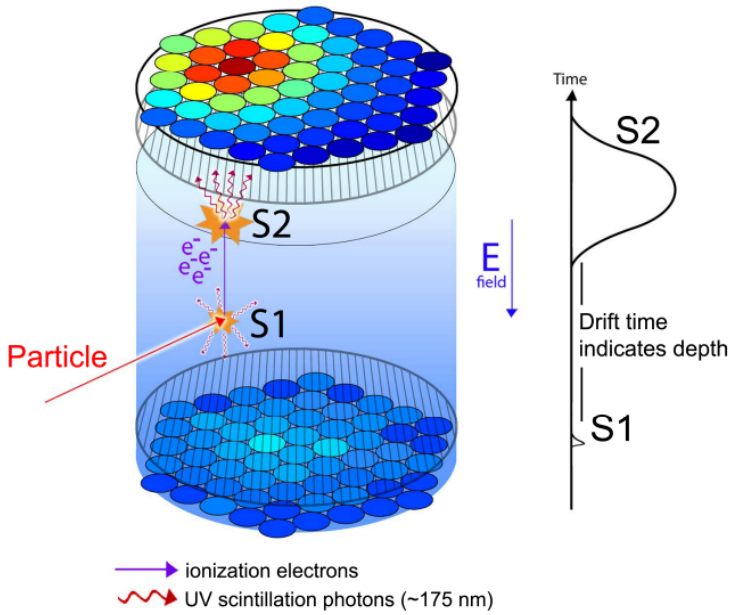


Figure 2.1: How LUX detects events. See description in text. Reprinted from [24].

to ≈1cm accuracy based on which upper PMTs were the most illuminated by the S2. The time delay between the S1 and S2, on the other hand, reveals the depth or Z-coordinate in the detector. A cartoon of this process is shown in Fig. 2.1.

A typical event in LUX is shown in Fig. 2.2. Within the waveform graphs (right side of figure), the first small signal is the S1 and the second, larger signal is the S2. The train of single ionization electron (SE) signals following the S2 is an example of complicated real-world detector behavior beyond the simple picture sketched in the previous paragraph. Single photoelectrons (such as from PMT dark current) and larger or longer-lived electron trains, although not seen here, are additional complications for pulse identification. The left side of Fig. 2.2 shows which PMTs detected the signal. This is dominated by the S2 signal detected mostly by the upper PMTs. The drift time of about 14,000 samples (140μs) indicates that this event happened at a depth about halfway between the top and bottom of the active volume.

2.2 Location

An important source of nuclear recoil backgrounds is cosmic ray muons, which can produce neutrons via spallation near the detector. In order to attenuate muons, the detector is deployed deep underground. This is done at the Sanford Underground Research Facility (SURF) in Lead, South Dakota. The facility is the former Homestake Mine, a now disused

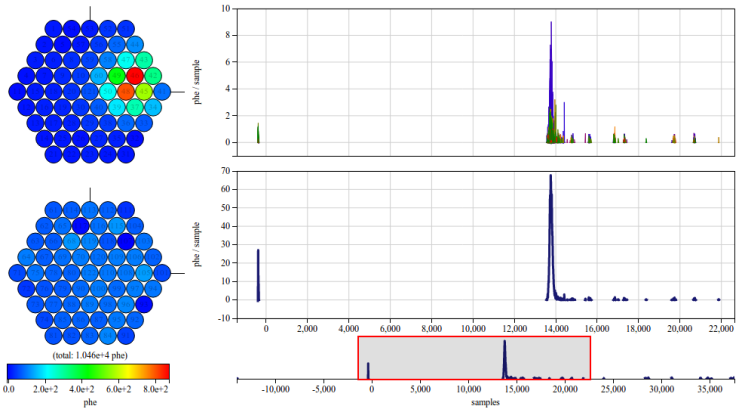


Figure 2.2: A typical event in LUX. See description in text.

gold mine in South Dakota's Black Hills.[1] LUX is located in a laboratory facility on the mine's 4850 level, meaning it is sited 4,850 feet below a reference point at the surface. This depth, 1500m or nearly a mile, corresponds to a water equivalent depth of 4300m given the local geology. Such shielding attenuates the muon flux by a factor on the order of 10^7 compared to the surface.

LUX and the other experiments now at SURF (including the Majoranna Demonstrator neutrinoless double beta decay experiment) are not the first to take advantage of this low-muon location. In the 1960's, Ray Davis set up an experiment to detect solar neutrinos in the very cavern where LUX would later be deployed. Davis's experiment[61] worked by detecting individual atoms of argon produced by neutrinos scattering off of chlorine atoms. The discrepancy between Davis's result and John Bahcall's theoretical calculation of the solar neutrino flux became known as the solar neutrino problem, which was not resolved until the discovery of neutrino oscillations decades later.

2.3 Hardware Systems

2.3.1 Shielding

The rock overburden is not the only shielding to reduce interactions of standard model particles in the detector's active volume. Surrounding the detector is a cylindrical water tank 6.1m high and 7.6m in diameter. The cryostat is positioned in the water tank so that the water barrier is at least 2.75m above the detector, 3.5m at the sides, and 1.2m below. An inverted pyramid of steel plates embedded in the floor beneath the water tank provides some additional shielding from the rock immediately beneath the tank. The water is circulated through a purification system which limits radioactive contaminants to a few

[1] Operated from 1876 to 2002, Homestake Mine became the largest and deepest gold mine in North America, with some three hundred miles of tunnels.

parts per billion.[2] The water tank attenuates gamma radiation from the rock by a factor of 8×10^{-9}. It also effectively moderates fast neutrons from nuclear decays and muon spallation in the cavern walls, with 99% of spallation neutrons moderated to low energy (<1keV) within 2m. That still leaves the risk of muon spallation in the water itself. The solution is a system of water PMTs to detect the Cherenkov radiation from muon interactions in the water, permitting the veto of any events coincident with such muons.

Immersed in the water is the detector cryostat, made of two concentric capsule-shaped titanium shells with a thermally-insulating vacuum between them. The inner cryostat has a height of 101cm and a diameter of 62cm. Inside the cryostat, oxygen-free high thermal conductivity (OFHC) copper blocks above and below the active volume provide some additional gamma shielding, in addition to their role in conducting heat. However, the main shielding within the cryostat comes from the xenon itself.

The principle of xenon "self-shielding" is this: Most backgrounds coming from outside the xenon cannot traverse very far through the xenon itself. The mean free path for gammas in liquid xenon is around 4cm at 662keV, falling to a mere 2mm around 122keV. Typical neutron scattering lengths in the same medium are on the order of 10cm.[68] Therefore, simply rejecting events near the periphery of the active volume provides powerful background rejection, and LUX's 3D position reconstruction furnishes the location information needed to do that. LUX's active volume, the area of xenon in view of the PMTs with the applied drift field, is a cylinder about half a meter in height and half a meter in diameter. In LUX's first underground run (the 85.3 live-day "Run 3"), a fiducial volume was defined by an 18cm radius and upper and lower constraints on drift time. This constraint eliminates about half the signal, as compared to using the full active volume, but it also eliminates the vast majority of background events that would otherwise be accepted (see Fig. 2.3). Filling the LUX detector takes 370kg of xenon, of which 250kg is located in the active volume in view of the PMTs. Imposing the fiducial volume cut described here leaves a fiducial volume of 118kg of xenon, with an expected background passing all cuts of only $2.6 \pm 0.2_{\text{stat}} \pm 0.4_{\text{sys}}$ events in an 85.3day run.[14]

2.3.2 Target, Grids, and PTFE

At the center of LUX is the liquid xenon target. A voltage difference is applied between a gate "grid" (actually a plane of parallel wires) five millimeters below the xenon surface and a cathode grid 49cm below the gate. The gate voltage is held near ground, while the cathode voltage, provided via a dedicated high-voltage feedthrough, is held at negative 10kV. These grids, along with a series of 48 copper field rings around the active volume linked by resistors in series, provide the nearly-uniform drift field that moves S2 electrons towards the xenon surface. Just 5mm above the xenon surface is the wire mesh of the anode grid at positive 7kV. The high electric field between the gate and anode serves as both the extraction field (driving electrons out of the liquid) and electroluminescence field (producing light as the

[2] 2ppb U, 3ppb Th, and 4ppb K

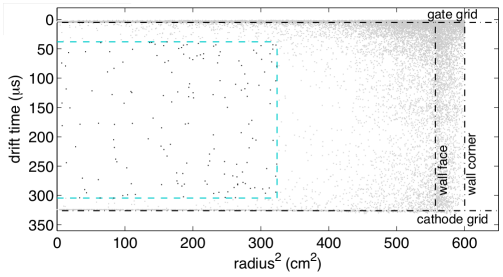


Figure 2.3: Position, in terms of drift time and radius squared, for all events in the 85day LUX Run 3. The cyan dashed line shows the boundary of the fiducial volume. Reprinted from [14].

electrons move 5mm through the xenon vapor above the liquid). Two other grids, one below the cathode and one above the anode, are held near ground to protect the PMTs from high electric fields.

The PMTs, each 2 inches in diameter, are arranged in hexagonal honeycombs of 61 upward-facing PMTs in the xenon liquid below the cathode and 61 downward-facing PMT in the xenon gas above the anode. They are housed in "PMT blocks," support structures which, like the gamma shields, are made of OFHC copper.

Interior to the copper field rings are the polytetrafluoroethylene (PTFE) walls of the instrumented volume.[3] The walls of the roughly-cylindrically shaped volume are made of twelve PTFE panels, so a horizontal cross-section is actually a dodecagon rather than a perfect circle. PTFE also lines the faces of the PMT blocks between the PMTs themselves. For xenon scintillation light, with frequencies in the vacuum ultraviolet (VUV) band, a smooth PTFE surface has about 95% reflectivity. An individual photon might be reflected many times before finding its way to the face of a PMT.

Fig. 2.4 shows the arrangement of these components in the cryostat.

2.3.3 Circulation and Purification

Deployed with the detector itself is accompanying infrastructure for storing, analyzing, and purifying the xenon. During detector operation, xenon is continuously recirculated through a getter. In the getter, gaseous xenon flows past reactive hot zirconium metal surfaces, causing contaminants such as nitrogen and hydrocarbons to bind to those surfaces. The gas handling system also includes the Storage and Recovery Vessel (SRV), which can store liquid xenon outside of the detector, and a gas sampling system. In the sampling system, a sample of xenon gas is isolated in a pipe and a xenon liquified by immersion of the pipe in liquid nitrogen. The remaining gas, in which the fractional contributions of any contaminants have been greatly increased by removal of much of the xenon, is then fed into a mass spectrometer for analysis.

[3]PTFE is commonly referred to by the trade name Teflon.

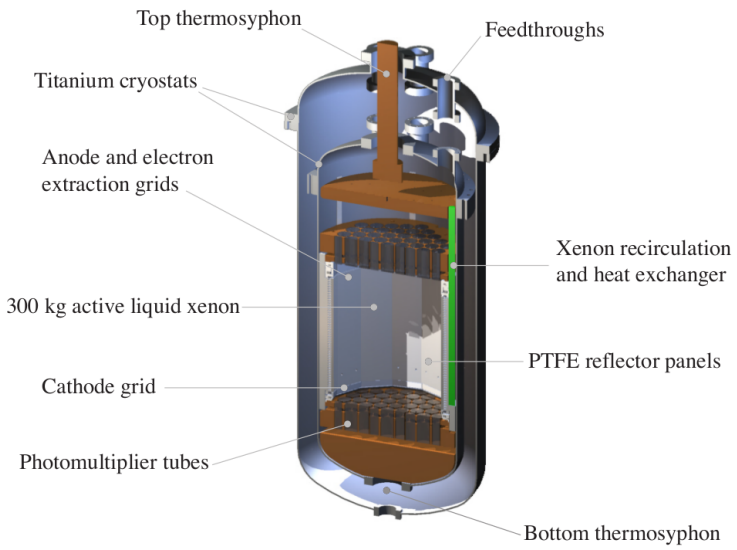


Figure 2.4: Key components inside the LUX cryostat. Reprinted from [24].

To cool the detector, thermosyphons are used. In this approach, liquid nitrogen flows down to the detector through hollow tubes. As it evaporates, gaseous nitrogen flows back up through the same tube, allowing for efficient heat transfer from the detector to a nitrogen dewar overhead.

2.3.4 Photomultiplier Tubes

Photomultiplier tubes (PMTs) turn light signals as small as a single photon into much larger electric signals that can be fed into ordinary electronic instrumentation. The LUX detector uses model R8778 PMTs manufactured by Hamamatsu Photonics K.K. of Japan. This model is specifically designed for use in liquid xenon detectors.

The operation of the PMT begins when a photon impinges on the window at the front of the PMT. Since ordinary glass is not transparent to 175nm photons generated by xenon scintillation, the R8778 PMT instead uses a window made of fused silica (artificial quartz). Coating the inside of the window is the photocathode, a thin film of material that can emit an electron when struck by a photon. To achieve the necessary low work function, many photocathode materials contain alkali metals. In the case of the R8778, bialkali (a rubidium/cesium/antimony blend) is used. The quantum efficiency (QE, the rate of conversion from photons to electrons[4]) of the R8778 is 33%.[69]

Next, the single free electron is amplified to a much larger electrical signal by a gauntlet of dinodes, electric contacts each of which is at a higher voltage than the last. The voltage

[4]To be precise, the quantum efficiency also includes the probability that the photon makes it through the window, assuming a normal angle of incidence. Quantum efficiency does *not* include the probability that the emitted electron will cause a cascade (see next paragraph), but this probability is about 90%.

difference between the photocathode and the first dinode creates an electric field accelerating the electron towards the dinode. When the electron strikes the dinode surface, multiple electrons are usually expelled from the surface, and these then begin moving towards the second dinode. The number of emitted electrons from each electron impact follows Poisson statistics; for an R8778 operated at typical voltages the mean is around three. The R8778 has 12 dinodes in total, and, as this process repeats, the electron cascade becomes exponentially larger. By the time the electric signal is collected by an anode at the base of the PMT, the signal has been amplified by a factor on the order of 3^{12}, or roughly a million.[69]

The operating voltage for an R8778 is around 1200V, with a resistor chain holding the dinodes at the correct relative voltages. For the final three resistors (i.e., those closest to the anode), each resistor has a capacitor in parallel. Without these decoupling capacitors, the ever-growing electron cascade with draw sufficient current off the dinodes to change their voltages, effectively un-biasing them. The decoupling capacitors, however, provide a reservoir of charge to mitigate that effect. In general, there are two ways to provide the voltage to a PMT: Either the photocathode can be grounded and the anode held at a high positive voltage, or the anode can be grounded and the photocathode held at a high negative voltage. LUX uses the latter case because it simplifies connecting the PMT to the next step in the signal processing chain, the preamplifier. The disadvantage is that the PMT body's voltage is far from ground, so care must be taken in designing a PTFE mount to prevent direct contact with the surrounding PMT block. A resistor connects the anode to ground, and the voltage difference across the resistor is used as the preamplifier input. A PMT's output signal from a single photoelectron has a rise time of 3.5ns, a fall time of 11ns, and a FWHM (full width at half maximum) of 7.7ns. The single-photoelectron pulse height varies from pulse to pulse by $\sigma/\mu = 40\%$. This large variability is due to the Poisson statistics with small expectation values that govern charge multiplication in the early dinodes.[69]

The pressure inside a PMT is initially on the order of 10^{-7} to 10^{-6} torr, in the high vacuum range. But if gas has entered the PMT, signals will be accompanied by afterpulsing. Afterpulsing occurs when electrons emitted by the photocathode strike and ionize a gas molecule on their way to the first dinode. The positive gas ion then drifts back towards the photocathode under the influence of the electric field, strikes the photocathode, and initiates a second cascade after the first cascade has already happened. Afterpulsing can be caused by a compromised vacuum seal, which could admit air prior to installation or xenon after installation. It can also be caused by helium, which can diffuse through the quartz window during prolonged exposure to air. Interestingly, the drift time for the ion to reach the photocathode is nearly independent of where the ion starts, because the electric field between the photocathode and first dinode scales roughly linearly with distance from the photocathode. Therefore, the time delay between pulses and afterpulses, on the order of a few microseconds, can be used to identify the exact contaminants in the PMT.[69]

2.4 Signal Processing

2.4.1 Amplification and Digitization

For each PMT, about 12m of cable conveys the signal out of the detector into the space above the water tank. There, a preamplifier amplifies the signal voltage by a factor of 5. Each preamplifier is based on a single AD8099 op-amp. Next, about 13m of cable conveys the preamplifier output to a series of electronics racks located nearby on the upper level of the Davis Cavern. Because the cables attenuate frequencies above 1MHz to a frequency-dependent extent, they temporally smear out the pulses by about a factor of two. Inside the electronics racks, the signal is fed into a "postamplifier," which temporally stretches the pulse by an additional factor of 1.5. The postamplifier outputs is fed into a digitizer card that includes an internal low-pass filter with a 30MHz cutoff.[24, 69]

The digitizer cards used in LUX's data acquisition system (DAQ) are model SIS3301 analog-to-digital converters (ADCs) manufactured by Struck Innovative Systeme of Germany. They digitize the analog postamplifier output at a rate of 100MHz, which explains the importance of the pulse stretching discussed in the previous paragraph: Without that signal shaping, a 7.7ns FWHM pulse might get missed almost entirely. The ADCs use special firmware developed for LUX that employs a "pulse-only digitization" (POD) technique, in which signal pulses are written to disk but the low-amplitude noise between pulses is not. To do that, data recording begins when the signal exceeds a "pulse start threshold" tuned to accept 95% of single photoelectron signals while excluding baseline noise 99.99% of the time. The end of the pulse is taken to be when the signal falls below a somewhat lower "pulse end threshold." These thresholds are set relative to an average baseline which is a running average measured on microsecond-intervals between pulses. That way, low-frequency noise (such as 60Hz noise) does not throw off the measurement. In addition to recording the pulse itself, the Struck cards record an additional 31 samples (310ns) after the end of the pulse. And because the digitized signal is being continuously written to a ring buffer, the cards are also able to record 24 samples (240ns) from just before the pulse starts.[5] Pulse-only digitization reduces the memory needed to store an event by roughly a factor of 50. During normal operation, this makes it possible to output nearly every signal pulse in every PMT, preserving everything for subsequent analysis.[13]

2.4.2 Data Processing

The raw output from the digitizer cards is transferred to a DAQ control computer and stored to disk as binary files which are given ".dat" filename extensions. These will be hereafter referred to as "dat files." The dat files are sent from the DAQ computer to a surface server and then sent over the internet to the primary LUX data processing facility at Brown University. There, the data is housed on a series of RAID arrays connected to a

[5] As a technicality, POD is applied to PMT channels in sets of two, so a PMT's data will also be retained if its partner PMT detects a signal pulse.

File Type	Information Therein
dat	Raw output from digitizer cards
evt	Waveforms for each event
rq	Results from analyzing each event
iq	Calibration information

Table 2.1: File types used in LUX.

LUX-maintained server called gsk-60. A WIMP search acquisition with a livetime of a few hours generates tens of thousands of individual dat files.

On gsk-60, the dat files are processed to produce "evt files," identified by their ".evt" filename extensions. The evt files for an acquisition are binary files containing a subset of the information in the dat files organized into a more usable form. The "extended event builder" software identifies events in the detector by using a software implementation of the LUX trigger. Then, all PMT signals within 500μs of the trigger are grouped together to make an entry for that event.

Next, a higher-level analysis of each event is done. The evt files are sent to the Brown University Center for Computation and Visualization (CCV), where a series of software modules analyze each event to produce a variety of descriptive information, or "reduced quantities" (RQs), which are output to "rq files." It is at this stage that the waveforms are analyzed to automatically determine what types of pulses occurred (S1, S2, a single photoelectron in a PMT, a single electron electroluminescing, etc.). Some quantities are determined for every pulse (such as various definitions of start time), while others are determined for the event as a whole.[6] To facilitate usage in many different environments, rq files are generated in several different file formats, all of which contain the same data. These include a LUX-specific binary format, as well as Matlab files, CERN Root files, and, occasionally, HDF5 files. After they are generated, the rq files are sent back from CCV and stored on gsk-60 along with the dat and evt files. All data (dat, evt, and rq files) is mirrored from gsk-60 to the National Energy Research Scientific Computing Center (NERSC), a process that will be discussed in detail in the next chapter. Table 2.1 summaries the different data files used in LUX, and Fig. 2.5 shows the flow of data among various computer systems.

2.4.3 Data Analysis

To identify potential dark matter nuclear recoils, a number of cuts are used to reject events that are not good candidates. For one, RQs for XY position and drift time are used to reject any event not occurring in the fiducial volume. Another important cut concerns the types of

[6]Previously, we have used the term "pulse" to refer to either a pulse of light (such as an S1 or S2) or a pulse of signal in a single PMT, depending on the context. Where it is necessary to distinguish the two, we will refer to a signal in one PMT as a "pod" (after the POD technique discussed in Sec. 2.4.1) and continue using "pulse" to describe a pulse of light as identified by the pulse-finder RQ algorithm.

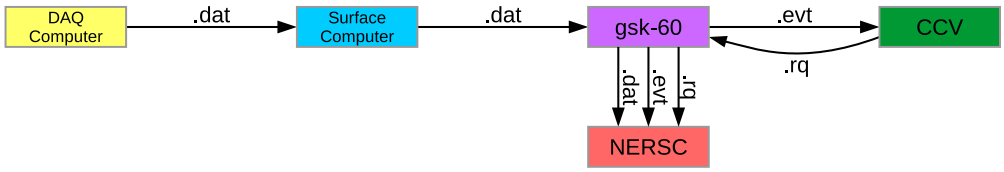


Figure 2.5: LUX data transfers. The mirroring of data to NERSC is discussed in the next chapter. Based on a figure from [138].

pulses making up the event. Since any real dark matter interaction would be a single scatter, candidate events should have one S1 followed by one S2, and no other S1's or S2's. These are referred to internally as "golden events."[7] Other events, which include double scatters, trains of single electrons following earlier large energy depositions, etc., are cut.

The fiducial volume cut, golden event cut, and restrictions on the sizes of the S1 and S2 are the main data analysis cuts. A few minor cuts also make small contributions. For each Run 3 event passing all cuts, four key RQs are collected for use in the next step: S1 size, S2 size, location depth, and location distance from center axis[110]. The distribution of events in this four-dimensional space is compared to expected distributions for electron and nuclear recoils to set probabilistic limits on nuclear recoils and, hence, on the dark matter interaction cross-section. The most important difference between nuclear and electron recoil populations is that for the same S1 signal size, nuclear recoils have lower S2 signals, on average. This is because nuclear recoils have more recombination due to their shorter track lengths[68].

Having now summarized the chain of signal processing from the first photons impinging on a PMT to the final analysis, we turn next to calibrations underpinning some of the analysis steps.

2.5 Detector Calibration

To determine a number of key properties underlying the analysis described above, calibration measurements were taken with a variety of sources.

2.5.1 Kr-83m

It is important to measure the S1 and S2 signal as accurately as possible, for several reasons. For one, the different distributions in S2 for a given value of S1 are used to distinguish electron recoils and nuclear recoils. For another, the S1 and S2 signals give the total energy

[7]For some analyses, the definition of golden events is slightly broadened to permit, for example, extra S1's after the S2. This is to address certain mistakes in pulse identification; plausible candidate events have only one actual S1 and one actual S2.

deposition of the interaction, according to

$$E = W\left(n_\gamma + n_e\right) = W\left(\frac{S1}{g_1} + \frac{S2}{g_2}\right). \tag{2.1}$$

Here, E is the energy deposition, W is the xenon W-value of 13.7 eV/quantum, n_γ is the number of scintillation photons created, and n_e is the number of free electrons created. $S1$ and $S2$ are the S1 and S2 signal, in units of "phd" or photoelectrons detected in a PMT, while g_1 and g_2 are their respective gain factors. The factor g_1 is 0.117 ± 0.003 phd/photon, reflecting quantum efficiency, photon absorption on grids, and other considerations, while the factor g_2 is 12.1 ± 0.8 phd/electron, due to the signal amplification from electroluminescence, an extraction efficiency at the surface of about 50%, and other considerations.[25]

There is a complication, however. The values given above hold for events right at the center of the detector, but the detector's true efficiency at measuring S1 and S2 light is position-dependent. In other words, it varies based on where in the detector the event occurs. There are two reasons for that: First is purely geometric considerations – some points are more favorably positioned relative to the PMTs. This effect is ameliorated by the reflective PTFE walls, which substantially improve the light collection of photons not originally headed towards a PMT. But the effect is still present, and applies to both S1 and S2. A second concern is specific to the S2. As the electrons are drifting through the xenon, not all of them make it to the surface without being captured. Average drift lengths in the vicinity of 100cm are typical in LUX. For an active region half a meter tall, this would cause significant error in the measured S2 size if not corrected.[68]

Given that the analysis will require using position-corrected S1 and S2 values to cancel out these effects, how can the size of the effects be determined? Some insight can be gained through simulations.[8] It is most desirable to have actual measurements of a monoenergetic interaction from different places in the detector, with which these effects can be measured directly. LUX includes six source tubes, vertical tubes running near the outside of the cryostat which make it easy to lower a sealed source into place. Sources such as Cs-137, with its monoenergetic 662 keV gamma rays, have been used with LUX, but the detector's xenon self-shielding (discussed in Sec. 2.3.1) makes it difficult to get many single scatters in the fiducial volume. LUX has developed an innovative solution: a calibration source deployed in the xenon itself.

The basis of LUX's xenon purity monitoring calibration is the decay of Kr-83m. This metastable state decays with a half life of 1.83 hours to another excited Kr-83 state, emitting a 32.1 keV gamma-ray. This other state then decays to the ground state of Kr-83 with a brief half life of 154ns, emitting a 9.4 keV gamma-ray. Because the half life of the second decay is so brief, the event appears almost as a single 41.5keV decay. The peaks of the two S1s are often separated, but the two S2s overlap entirely. In LUX, Kr-83m is fed directly

[8]Simulations of photon paths can be made accurate if the reflectivity of PTFE is well-known. Simulating average drift length based on, say, measures of xenon contaminants would be a far more difficult task. Simulations informed by calibrations are much more effective than simulations from first principles alone.

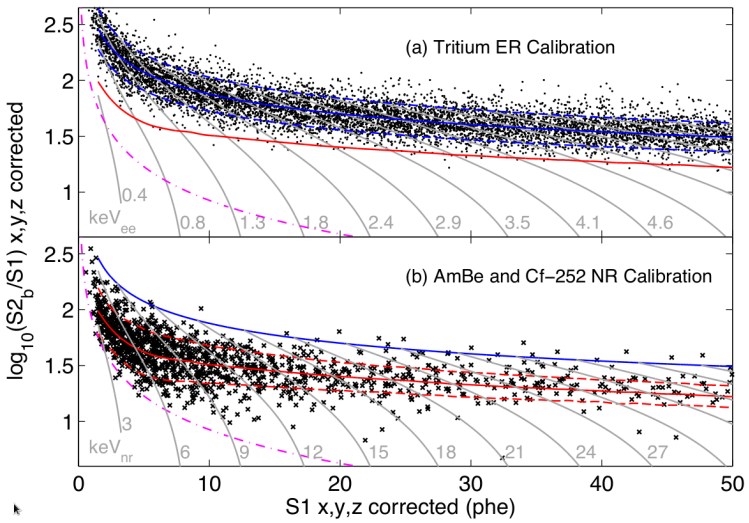


Figure 2.6: Electron recoil (top) and nuclear recoil (bottom) bands. Horizontal axis is position-corrected S1, vertical axis is base-10 logarithm of the ratio of S2 (at bottom PMTs) to S1, also position-corrected. ER band is from tritium calibration; NR band is from simulations based on sealed-source calibrations. Reprinted from [14].

into the xenon circulated through the detector, creating a uniform calibration source within the active volume itself. The Kr-83m that's used is produced by a Rb-83 source with a half life of 86.2 days. No action is taken to remove the ground-state Kr-83 from the xenon, as Kr-83 is both stable against nuclear decay and chemically inert.[91]

2.5.2 Tritium

As mentioned, ER vs. NR discrimination hinges on electron recoils showing a different distribution of S2s for a given S1 than nuclear recoils show. This can be seen in Fig. 2.6, where the x-axis is S1 size, and the y-axis is the log of the S2-to-S1 ratio. In this figure, position-corrected S1 and S2 values are used. Also, as in the analysis, the S2 signal is based on only the bottom PMTs, to reduce signal variations stemming from whether the electroluminescence happens directly below a top PMT's face or directly below a space between PMTs. In order to characterize these populations for analysis and simulation purposes, it is necessary to have calibration data for both ERs and NRs. We discuss ERs first.

To understand the behavior of electron recoils over the whole electron-equivalent energy range where LUX searches for dark matter candidates, we need a calibration source that produces electron recoils over a wide range of energies. For this, the Kr-83m with its monoenergetic electron recoils will not suffice. A source that will work well is tritium. Tritium (^{3}H, also designated T) decays via the beta decay

$$^3\mathrm{H} \rightarrow {}^3\mathrm{He} + e^- + \bar{\nu}_e \,. \tag{2.2}$$

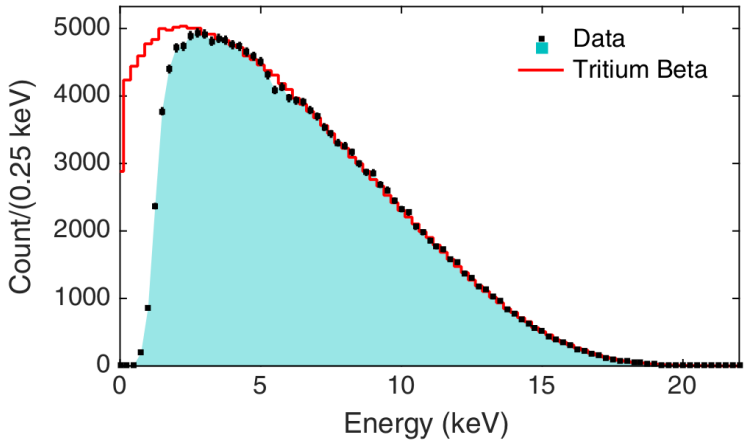


Figure 2.7: Tritium energy spectrum. Block dots with blue shaded region are from data; red line is from theory. Aside from low detection efficiency below ≈ 3keV, the match is quite good. Reprinted from [25].

Because there are three decay products, the electron's kinetic energy is not kinematically constrained to a single value, leading to the range of electron recoil energies shown in Fig. 2.7. While most events are low energy, a tail extends up to a maximum of 18.6 keV.[9][25]

As with the Kr-83m source, it is desirable to introduce the tritium into the xenon itself to create a calibration source that's uniform throughout the detector and not subject to xenon self-shielding. To do that, tritiated methane (CH_3T), in which one ordinary hydrogen atom in each molecule is replaced by a tritium atom, is introduced into the circulating xenon. One reason for using tritiated methane instead of pure tritium (T_2) is that the larger molecules do not adsorb onto surfaces as well. Also, the tritiated methane is removed by the getter very effectively, which is important because tritium's 12.3-year half life precludes simply waiting for it to decay away.[25]

2.5.3 D-D Neutrons

Nuclear recoil calibration presents more challenges than electron recoil calibration. Because neutrons interact with nuclei and not electrons, neutrons do induce nuclear recoils no different than the NRs which a WIMP scatter would cause. However, there is no neutron source that can be introduced into the xenon and then easily removed. One option is to place a neutron-emitting sealed source, such as AmBe or Cf-252, in a source tube. Another solution, which provides a beam of monoenergetic neutrons, is to use a neutron generator.[19]

LUX's neutron generator (or colloquially, "neutron gun") contains a tabletop linear accelerator in which deuterium fusion occurs by the interaction

$$^2\mathrm{H} + {}^2\mathrm{H} \rightarrow {}^3\mathrm{He} + \mathrm{n}\,. \tag{2.3}$$

[9]That's 18.6 keVee, since electron recoil energy loss to phonons is minimal.

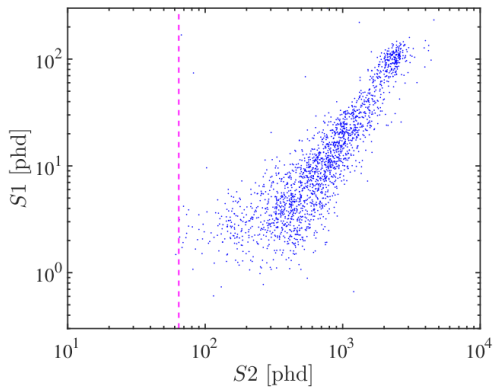


Figure 2.8: S1 and S2 signals from D-D neutrons, in units of detected photoelectrons. Reprinted from [19].

This "D-D" process emits neutrons with a kinetic energy of 2.45 MeV. The neutron generator is located just outside LUX's water tank, and an air-filled tube in the tank provides a passage through which neutrons can more easily reach the detector. Once inside the liquid xenon, the neutrons have a 12.6cm mean free path. Although the neutrons are monoenergetic, only a fraction of a neutron's energy is transferred in a scattering event, and the recoils are not monoenergetic. Instead, nuclear recoils occur over a range of energies, up to a kinematically-imposed maximum of 74 keVnr. Fig. 2.8 shows the range of S1 and S2 signals for nuclear recoils from D-D neutrons in LUX.[19]

A number of useful measurements come from neutron calibration data. Using single scatters, one can map out the S2-vs-S1 spectrum of NRs, for use in distinguishing NRs from ERs. Using double scatters, one can calculate the energy of the first scatter and thereby measure nuclear recoil signal sizes as a function of recoil energy.[19]

2.6 Run 3 and Run 4

The LUX detector was assembled at a surface lab at SURF. After a commissioning run at the surface[23] in 2011-2012, it was moved underground, opened one final time in an underground cleanroom, and then installed in the water tank. LUX conducted two WIMP search campaigns. The first, "Run 3," incorporated a live time of 85.3 days of WIMP search data, collected between 21 April 2013 and 8 August 2013. Run 3 data formed the basis for an initial spin-independent limit[14] and a stronger spin-independent limit based on reanalysis of the same data[16]. The next WIMP search campaign, "Run 4," incorporated a live time of 332.0 days of WIMP search data, collected between 11 September 2014 and 2 May 2016. A combined spin-independent limit based on both Run 3 and Run 4 data has been published[20]. Extensive calibration data was collected before, during, and after the periods of WIMP search during both runs. Fig. 2.9 shows the increasingly-strong spin-independent WIMP cross-section limits set over the course of LUX operation. A number

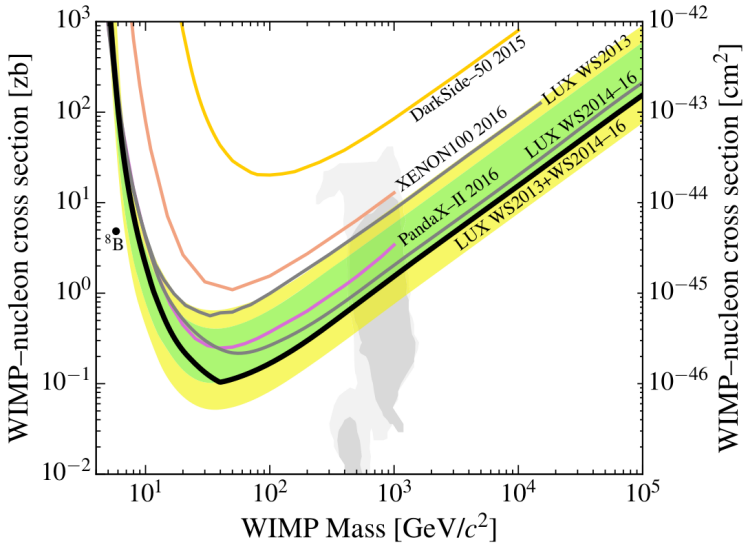


Figure 2.9: Spin-independent WIMP cross-section limits from LUX, including the Run 3 reanalysis (labeled "LUX WS2013"), Run 4 (labeled "LUX WS2014-16"), and the combined Run 3 + Run 4 result (labeled "LUX WS2013+WS2014-16"). Reprinted from [20].

of instrumentation papers[24, 25, 12, 19, 22], limits on spin-dependent dark matter[21, 17], and limits on other dark matter theories[15] have also been published.

While many aspects of LUX operations remained the same over the course of the instrument's life, Run 4 introduced some important changes compared to Run 3. We will survey the key differences here.

Grid Conditioning. Between Run 3 and Run 4, a campaign of "grid conditioning" was undertaken. In this process, the detector was filled with xenon gas and the voltages of the cathode, anode, and gate grids were deliberately set to cause electrical breakdown for days at a time. As expected, the result was an increase in the voltage differences necessary to cause breakdown in the detector, presumably due to ablation of any rough spots on the grids during conditioning. Using a larger gate-anode voltage difference in Run 4 increased the electron extraction efficiency from 49% to 73%. However, the conditioning campaign also had the undesirable effect of electrically charging the active volume's PTFE walls. This is thought to have happened when VUV photons induced electron-hole pairs in the PTFE. Because electrons in PTFE are less mobile than holes under an applied electric field, a negative charge built up. As a result, the mostly-uniform electric field of Run 3 was replaced by a highly position-dependent field in Run 4, complicating the analysis.[20]

Salting. A scientific analysis is "blind" if most or all of the data is not reviewed until after the exact analysis steps have all been finalized. This allows researchers to develop the details of the analysis without being biased by knowing the results during the development process. The Run 3 analysis was deliberately not a blind analysis. For Run 4, an approach was needed to allow for a true blind analysis, while still leaving it possible to study any

unexpected backgrounds prior to unblinding. The solution was to "salt" the data – inserting fabricated WIMP events when building evt files to camouflage any real WIMP events that could be present.[10] Fabricated WIMP events were created by pasting together pulses of the desired sizes drawn from dedicated tritium calibration data.[115] Because these salt events are added during evt-building, it was necessary to restrict access to dat files prior to unblinding, as the identity of salt events could be easily revealed by just checking which events in the evt files are missing from the dat files. An example of designing data transfer/storage systems to restrict dat access will be seen in the next chapter.

Other. Various other analysis improvements were introduced in the Run 3 reanalysis and in Run 4. To give just one example, one change is a migration from reporting PMT signals in units of photoelectrons (phe) to using units of detected photons (phd). The two quantities are not quite the same because of the possibility of one photon striking the photocathode and causing two electrons to be emitted[70].

Over the course of the LUX experiment, large amounts of data were generated. The development of some tools to handle that data is the topic of the next chapter.

[10]Salting has been described as "blinding with a flashlight."[115]

Chapter 3

Data Storage and Visualization

The large amount of data produced in the course of the LUX program presents special challenges. With more than half a petabyte of dat, evt, and rq files, developing ways to effectively store, manage, and access this data is paramount. In this chapter, we discuss two software tools for working with LUX data. The first of these is "Crossing Guard," a package of scripts for building and maintaining a full mirror of the LUX experimental data at NERSC[1]. The second of these is "Visualux," an interactive website for viewing LUX data. This chapter will discuss the design of these software tools. Details of how to administer the software can be found in an internal LUX technical report[83].

3.1 LBL NERSC Data Mirror

3.1.1 Why a Mirror?

The primary repository for LUX experimental data is the gsk-60 server at Brown University (full DNS name: `gsk-60.het.brown.edu`). The project to establish a mirror of this data at Lawrence Berkeley National Laboratory's NERSC computer facility is motivated by two considerations. First is the need for offsite data backup. Although the RAID arrays used in gsk-60 protect against data loss from a single failed hard disk, having the data and its only backup in the same room leaves other vulnerabilities. The second motivation is to make the data easier to use. Prior to the establishment of the mirror, many LUX analyses required the tedious step of downloading large amounts of data to use on one's local computer. Having all the data at NERSC permits users to simply log into NERSC and run analyses with the facility's extensive computing resources. Visualux, described in Sec. 3.2, is an example of software running at NERSC using the mirror copy of LUX's experimental data.

[1] A Lawrence Berkeley National Lab computing facility, described in Sec. 3.1.2

3.1.2 Infrastructure

NERSC

The National Energy Research Scientific Computing Center (NERSC) is "the primary scientific computing facility for the Office of Science in the U.S. Department of Energy" and "one of the largest facilities in the world devoted to providing computational resources and expertise for basic scientific research" [111]. A division of E.O. Lawrence Berkeley National Laboratory (LBL), NERSC recently located to a new facility on the main LBL site.

NERSC currently operates three supercomputer systems: Cori, Edison, and PDSF. These allocate computing time in different ways. For Cori and Edison, would-be users apply for grants of computing time. For PDSF, collaborations can contribute to the cost of the machine in exchange for a proportionate share of its computing time. NERSC also has extensive data storage capabilities. Cori and Edison have associated "scratch space," machine-specific temporary storage to support computations. Permanent storage is provided through the "eliza" system, accessible only from PDSF, as well as the "project" and "projecta" drives, accessible from any machine. NERSC's tape archive, the High Performance Storage System (HPSS), has been in operation for nearly twenty years and houses more than 80 petabytes of scientific data.[112] NERSC provides various networking resources as well.

NERSC resources allocated to LUX include 440TB of storage space, divided among eliza, projecta, and project. LUX's annual HPSS allocation (which accounts for both storage and input/output) is sufficient to write an additional 440TB and store it for a year. The collaboration has computing time on all three supercomputers, although most work is done on PDSF. For the LUX NERSC data mirror, rq and evt files are stored on projecta and eliza. The dat files are stored on tape in HPSS.

Globus

The Globus transfer service is used to copy data from gsk-60 to NERSC. Compared to well-known tools for transferring data over the internet, such as SCP or rsync, Globus provides a number of advantages. For one, if the origin or destination machine goes offline, the transfer cleanly resumes once the machine is back. The most important Globus advantage is speed. Typical SCP or rsync transfers from gsk-60 to projecta run at speeds in the neighborhood of 5MB/sec. However, Globus transfers from the gsk-60 Globus endpoint to the NERSC Data Transfer Nodes' endpoint can easily exceed that rate by a factor of five or more. The reason for the Globus transfer service's celerity is that it's an implementation of GridFTP, a file transfer protocol optimized for large transfers[28].

3.1.3 Software Layout

To populate the LBL NERSC mirror of LUX's data and keep it up to date, a collection of scripts called "Crossing Guard" was developed. Central to Crossing Guard are two Python 2.7 scripts called finddata.py and movedata.py. Run at regular intervals on PDSF, the former

updates a record of all data available at gsk-60, while the later issues Globus calls to copy data to NERSC. The code is described below, and a schematic of how it works is shown in Fig. 3.1. The code fully automates the mirroring process, and supersedes the manual Globus requests previously used on an as-needed basis to bring data to NERSC.

Keeping Track of Data

The finddata.py script uses a series of SSH calls to gsk-60 to find out what data is available to be mirrored from gsk-60's RAID arrays.[2] On each RAID array, files are organized into folders by type (dat, evt, rq) and then by which acquisition they come from. An acquisition is a period of continuous data-taking with unchanged detector settings, which may last from a few minutes up to eight hours. We'll use the term "dataset" to describe a folder of data of a given type (dat, evt, rq) from a given acquisition.

The files to be mirrored include all dat files, as well as all evt and rq files generated from them. There are also evt and rq files produced from simulations, but they are not populated to the mirror. That's because there's less need to back them up (if lost, they can always be regenerated later) and because they are more easily handled on an ad hoc basis, since the production and tracking of simulation output is not centralized in the same way as with actual detector data. Additionally, there are special steps for restricting access to dat files designated to have salt added during evt-building. The extra precautions for these "locked" dat files will be discussed separately in Sec. 3.1.3.

For each type of data (dat, evt, & rq), a Python Pandas dataframe is used to keep track of which datasets are available and whether they've been mirrored yet. The appropriate dataframe is loaded into RAM whenever finddata.py or movedata.py is run and is saved to the NERSC filesystem as a Python Pickle object between runs. Each time finddata.py finds a new dataset to mirror, a corresponding new entry is added to the dataframe. Then as movedata.py supervises the data transfer, it updates that entry to reflect which steps have been completed. Whenever a script writes the dataframe to disk, it also generates a human-readable text file version for use by the mirror administrator.

Moving Data

Mirroring data to NERSC takes place in four steps: (1) submitting a Globus job, (2) confirming the completion of the Globus job, (3) changing file location/permissions at NERSC, and (4) confirming the completion of those changes. Each time movedata.py is run, it executes one step for each dataset that's in the process of being moved.

Before submitting a Globus job, movedata.py must determine whether a given dataset on gsk-60 is complete and ready for transfer. For the most part, this is done through the use

[2]The RAID arrays are named after mythical horses and other fantastic beasts: Arion, Brag, Centaur, Dragon, Epona, Fenrir, Gringolet, Hippocamp, Ifrit, Ipotane, Jackalope, Kelpie, Leviathan, Morvark, Nuckelavee, Orochi, Pegasus, Quetzalcoatl, Rhaebus, Sleipnir, Svadilfari, Tikbalang, Unicorn, Vedfolnir, Wyvern, Xanthos, Yeti, & Ziz.

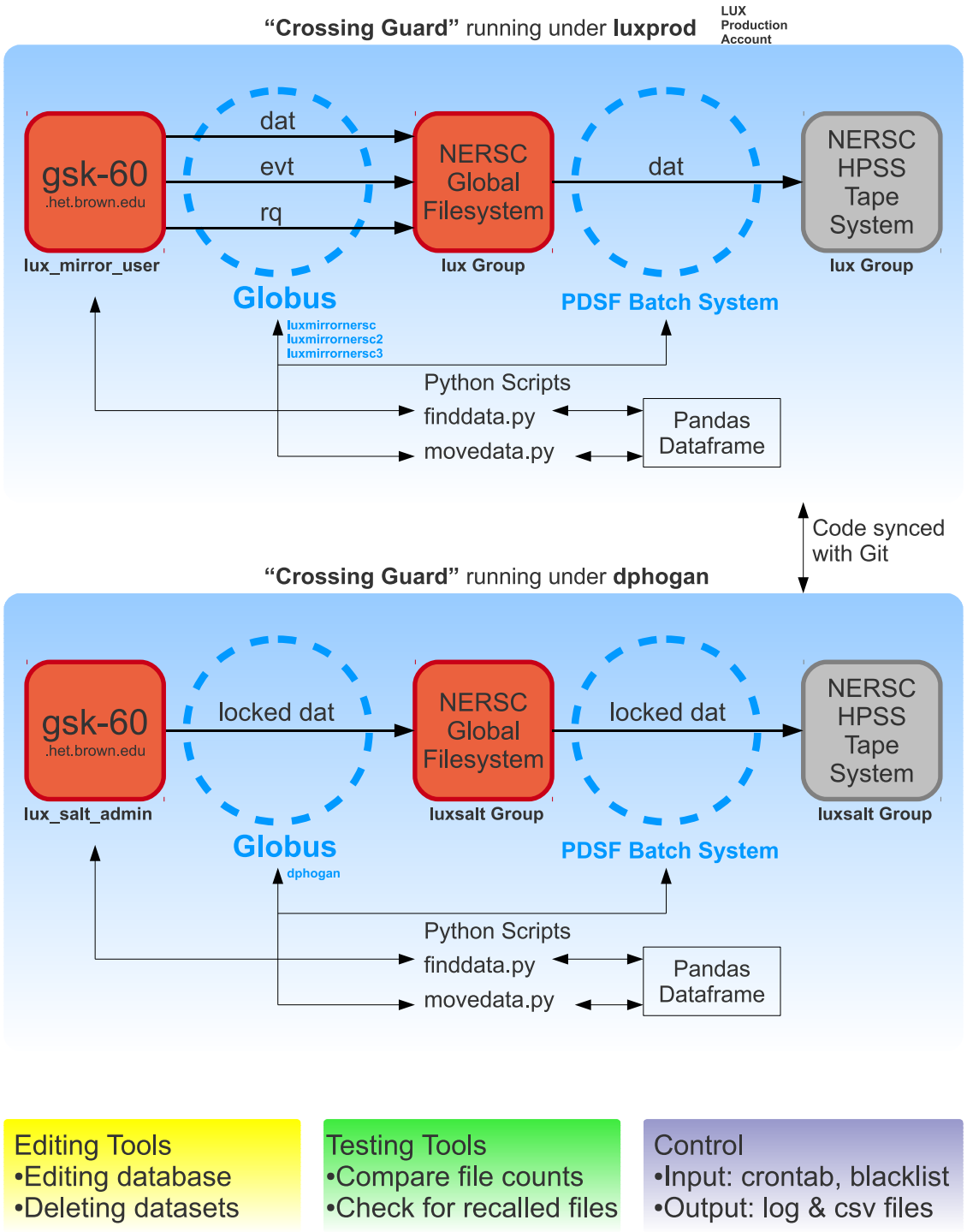


Figure 3.1: Schematic of hardware and software systems involved in transferring data to the NERSC mirror.

of "flags," empty files added to each dataset's folder as certain file operations on gsk-60 are completed. For dats and evts, flags indicating the completion of file compression mean the data is ready to move, while for rqs the cue is a flag indicating completion of data transfer from CCV to gsk-60. In practice, the presence of pathological incomplete datasets (such as empty folders, or even folders empty aside from flags) requires a somewhat more complicated chain of if-then questions to determine if a dataset is really ready to be copied, as described in Ref. [83]. Once a dataset is found to be ready for mirroring, the transfer is started by submitting a request to Globus's command line interface (CLI).

After the data's been transferred and the transfer verified by checksums, movedata.py will find out through a CLI query on its next run and in the following run will initiate the final file processing. In the case of evt and rq datasets, which stay on the file system, nothing more is required, aside from a file permission adjustment. But for dat files, which are stored on HPSS, another step is needed. For each dat dataset, a script is submitted to the PDSF batch system. That script, when run on a PDSF batch node, instructs the system to transfer the data to HPSS and then delete it from the filesystem. After movedata.py runs again and finds the batch job or permissions change has concluded without errors, the data transfer is complete.

Scheduling

Both finddata.py and movedata.py are run at regular intervals by cron on a PDSF login node. Typically, finddata.py is run four times per day per data type (dat, evt, rq), and movedata.py is run twice per hour per data type. If many datasets are available for mirroring (as was the case when initially setting up the mirror), trying to move them all at once would overwhelm various system resources. To avoid that, movedata.py respects limits on the maximum number of datasets to work with at once; limits are typically set at around 50 datasets per data type.

When first setting up the mirror, some datasets were more urgently needed for immediate analysis at NERSC than were others. Given a list of top-priority datasets, such transfers could be prioritized by setting an "Ignore" boolean in the Pandas dataframe entries for all lower-priority datasets, until such time as the high-priority transfers had completed. Another tool for manual fine control of transfers is the "blacklist" and "whitelist" files. For datasets named in these files, movedata.py will ignore its usual decision mechanisms about whether a dataset is ready to be mirrored and instead mirror any files in the whitelist while permanently ignoring those in the blacklist.

Salt

For locked dats (Run 4 dat files earmarked to have salt added to their corresponding evt files during evt-building), access to these dat files by the broader collaboration had to be restricted prior to unblinding. To back up these files from gsk-60 to NERSC, a parallel implementation of the mirroring system was set up. The code was cloned, using Git, from

a shared LUX NERSC account to an individual user's NERSC account. The locked dat datasets were accessed through a special gsk-60 Globus node, and once at NERSC were assigned to a Linux file group with only one user as a member. Thus a full backup was achieved without compromising the integrity of the salting protocol.

Other Tools

In addition to the main finddata.py and movedata.py programs, a number of supporting tools have been created. One tool allows for making changes to the dataframes, such as turning "Ignore" booleans on and off. Another allows for cleanly deleting datasets, removing the dataset from the filesystem and/or (if applicable) HPSS, as well as deleting the dataset's dataframe entry. One diagnostic tool compares the number of files for each dataset at NERSC and gsk-60, looking for discrepancies. On two occasions, salting details changed and some previously-generated evt and rq datasets had to be withdrawn; another diagnostic tool checks that none of this superseded data is present on the mirror.

Transferring the bulk of the data took place over a period of a few months during Run 4. The LUX NERSC data mirror currently houses 291TB of dat files, 208TB of evt files, and 184TB of rq files.

3.2 Visualux

Visualux is web-based interface for viewing information about events recorded by the LUX detector. Visualux draws on evt files to plot the signals measured by the PMTs, and it overlays additional information drawn from the corresponding rq files. A screenshot of Visualux's output is shown in Fig. 3.2.

Located at URL `https://portal.nersc.gov/project/lux/visualux/`, Visualux uses password authentication to restrict data access to within the collaboration. The online approach has many advantages: The system can be used from anywhere, and users don't have to install any software on their own computers. Users also don't have to download any datasets, nor do they have to locate their event(s) of interest within the data files. Collaboration members have requested to view event(s) in Visualux more than 15,000 times, and most requests ($> 80\%$) result in all requested events being found by the system. In total, 2.2 billion unique events are available for viewing in Visualux.

Visualux is made up of two parts. One part is the "frontend," which is the website itself running in a user's web browser. The other part is the "backend," which is software running on PDSF to gather, format, and return the data requested by the frontend. The two communicate through "NEWT," an application programming interface (API) developed by NERSC. The Visualux system makes use of a variety of languages and utilities, including Python, C++, shell scripting, HTML, PHP, JavaScript, CSS, AJAX, Git, and MySQL.

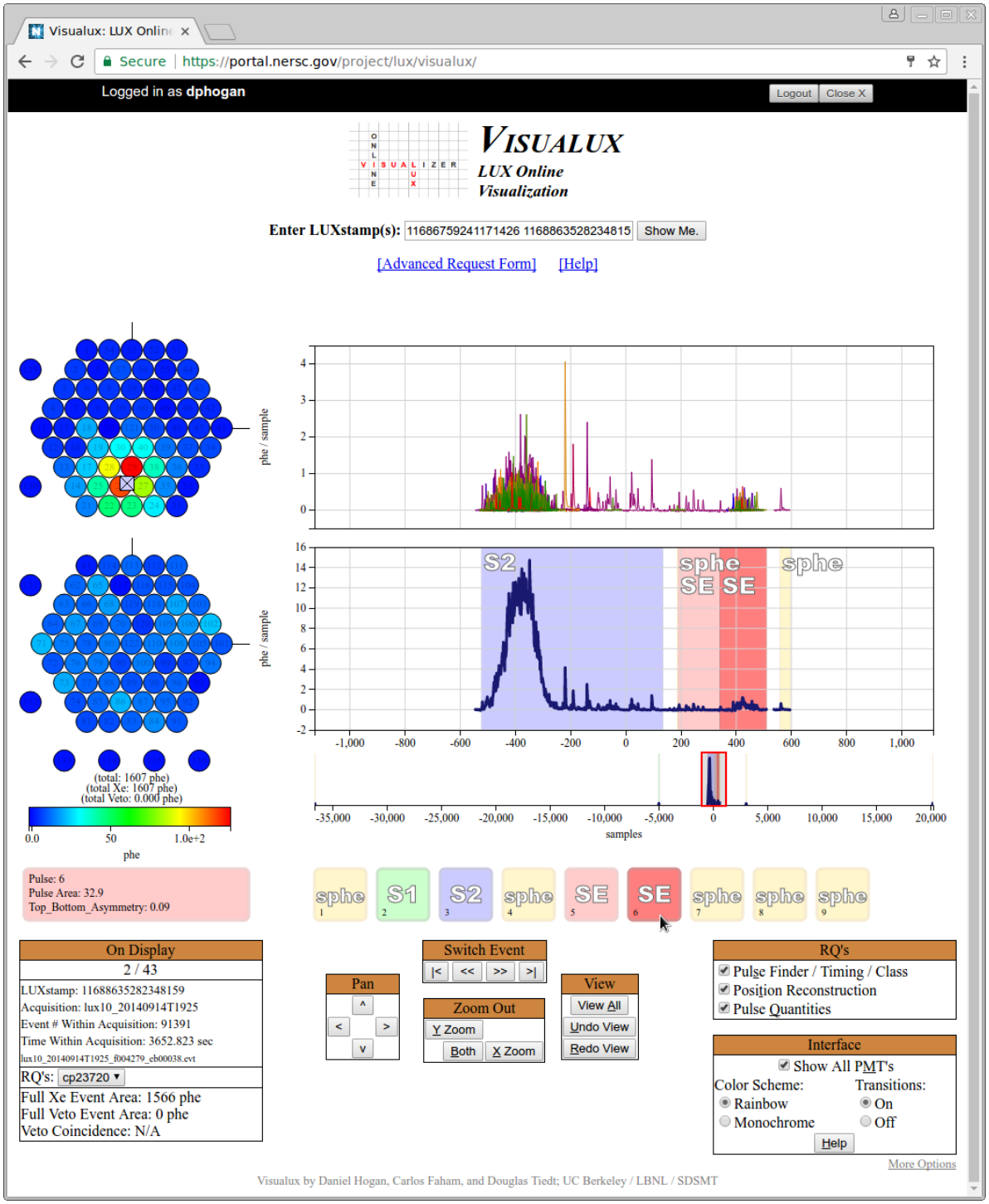


Figure 3.2: Screenshot of Visualux frontend webpage. The view is of an S2 from a Run 4 WIMP search electron recoil.

3.2.1 Communication

Communication between the frontend (webpage) and backend (software on PDSF) is handled through the Nice and Easy Web Toolkit (NEWT). To use the NEWT API, a webpage must include the NEWT JavaScript library. That library defines a function for AJAX-style calls to communicate with any of NERSC's supercomputers. With this API, a user of a webpage can communicate with NERSC systems and execute any command just as if he/she logged in on a terminal and typed the command himself/herself.[113]

For that to work, NEWT has to know which user is using the webpage. A feature of the NEWT JavaScript library is that it adds an authentication login bar at the top of each webpage from which it's called. Unless a user successfully logs in with their NERSC credentials, NEWT will not honor their requests. Therefore, a website using NEWT does not require its own database of users and passwords, as this service is provided through NERSC. Any commands run via NEWT will run with the logged-in user's privileges, meaning that using a NEWT page doesn't elevate any user's privileges beyond what that user can already do. In particular, non-LUX researchers with NERSC accounts cannot access LUX data, because NEWT respects the file permissions of the data.

NERSC's web hosting service supports both unsecure (HTTP) and secure (HTTPS) protocols. To increase Visualux's cyber security, a server configuration file is used to redirect all HTTP page requests to secure HTTPS connections.

3.2.2 Backend Code

When the backend code receives a request from the webpage, the first thing it must do is figure out where the data can be found. An event in the detector can be labeled in one of two ways. The first way is to the give its "luxstamp," which is an identification number unique to that event.[3] Luxstamps can be calculated from the contents of an evt file, and they're also given directly in rq files. The second way a user can specify an event is to give its acquisition and order in the acquisition. So a request can sound like "Show me event 11688635282348159" or it can sound like "Show me the 91391st event in the acquisition called lux10_20140914T1925." Either way, the code must be able to quickly find that event in hundreds of terabytes of data, and that calls for a database.

Visualux's event database contains an entry for every event in every evt dataset on the LUX NERSC data mirror. Each event's entry includes:

- luxstamp
- acquisition name (also called "filename_prefix")
- event's order in the acquisition

[3] A luxstamp is a large integer. It's defined as the amount of time from New Year's Day 2011 to the event's trigger, measured in units of 10ns, with a known offset from the fact that the Data Acquisition System records times relative to the acquisition start time, which itself is only recorded to 1-minute accuracy.

- file number
- order within the file
- which version of the event builder was used to generate the evt files

Using this information, the particular file in which any given event is stored, as well as its order within the file, can be looked up. Although the database is built using evts, a second database for rqs is not required, because rq datasets distribute their data among files in the same arrangement as the evt files from which they are generated.

The event database has gone through two iterations. Originally, the database was stored as a MySQL table, and MySQL queries were used to retrieve the information. However, having many entries in a single table adversely affected performance when loading additional entries (although the use of indices kept data retrieval times manageable).

To solve that problem, the database has been reimplemented with a two-tiered file-based approach. First, if the user requests a luxstamp, a text file with the minimum luxstamp of each acquisition is used to match the user's requested luxstamp to the acquisition it came from. Second, with the acquisition now identified, an acquisition-specific file with the same information as in the bulleted list above is consulted to locate the event. These text files are plain ASCII text, so the shell script that looks up the events can use standard Linux stream-processing commands. The text files are generated en masse by the PDSF batch system and saved on eliza, so they're already in place and ready for use whenever a user requests an event. Most of the database-building code is written in Python; however, a C++-based evt file reader is used. For tasks of this scale,[4] the time saved from using C++ instead of Python is measured in days.

Once the backend code has used the event database to locate the desired event(s), it's time to look up the needed data and organize it in a convenient file format. The backend code is written in Python, so it uses LUX's evt-reading Python module to read the relevant evt file. To read the relevant rq file, no specialized code is needed. That's because the Matlab-formatted rq files can be read by the Matlab file reader bundled into the standard SciPy library.

Having opened the data files, the backend code then collects the desired subset of that data in a Python dictionary. From the evt file, the code reads in all PMT signals recorded by the digitizers. The code uses a preset IQ (calibration) file to convert ADC counts to more useful physical units (phe/sample). A frontend interface option allows the user to select a different calibration file, but in practice this is rarely needed. If the user does request a different set of calibration data, a call to the MySQL database underlying LUX's electronic log[69] retrieves it. From the rq file, the code reads in information about the event's pulses. This information includes each pulse's start time, stop time, type, etc. Oftentimes there are multiple rq datasets for the same evt dataset, because the data has been re-processed with

[4]Building the full event database requires reading every evt file.

a revised set of rq-building modules. In this case the information from all applicable rqs is included. A dropdown menu on the frontend page enables toggling between them.

Next, the backend code must format the data in a way that's easy for the frontend to handle. The event's Python dictionary is therefore saved as a JSON file. JSON (JavaScript Object Notation) is a human-readable JavaScript-friendly file format that is widely used for transferring information in AJAX-style web design. If the user requests more than one event, each event is saved in its own JSON file, to limit the size of each file transfer.[5] As each JSON file is completed, the backend code writes its name to a "request file." Meanwhile, the frontend code continuously checks this request file to know when each JSON file is ready for transfer. Once a JSON file is ready, the frontend sends a NEWT request to get the data. Each request is assigned a unique number, so there are no problems if two users are using the website at the same time.

3.2.3 Frontend Code

In describing the website that serves as the frontend client to Visualux, we will consider the same subject from two perspectives. First will be a look at the website's user interface as experienced by the user. Next will be a discussion of the code that powers that interface.

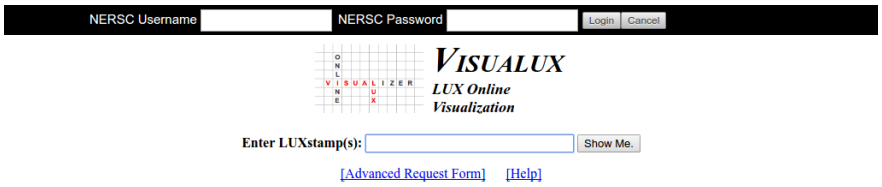
User Interface

When a user first navigates to `https://portal.nersc.gov/project/lux/visualux/`, the user enters his/her NERSC username and password in the appropriate fields and presses "Login" to authenticate through NEWT. The page (Fig. 3.3a) also has a single, simple text box for entering a list of luxstamps to view.[6]

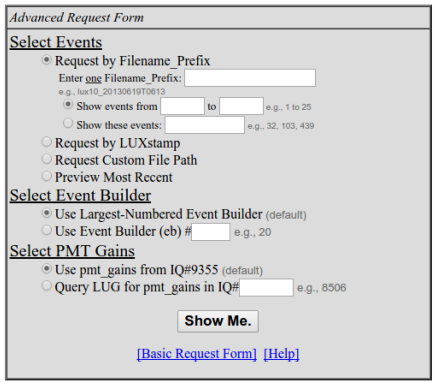
That simple user interface is sufficient for the most common queries, but additional options for requesting data can be found by clicking the link to the "Advanced Request Form" (Fig. 3.3b). Here, a user may request events by luxstamp or by giving an acquisition's name and the events' order in that acquisition. Another Advanced Request Form option lets the user manually specify file paths to their own evt and rq files anywhere at NERSC. This seemingly-obscure feature in fact plays an important roll: It allows any user to view his/her own simulation output on Visualux, extending the site's applicability from experimental data to simulations as well. The form also allows the user to modify certain Visualux defaults. Sometimes, one acquisition will have multiple evt datasets because it has been processed with different versions of the event builder. In that case, Visualux defaults to the largest-numbered version (which is usually the most recent), but the user can specify a different version if desired. Another option, mentioned in the backend discussion above, lets the user pick a different IQ file of PMT gains.

[5] Even so, there have been cases of individual JSON files exceeding NEWT's transfer size limits. However, this happens only for the largest muon-related events and does not affect any dark matter candidates.

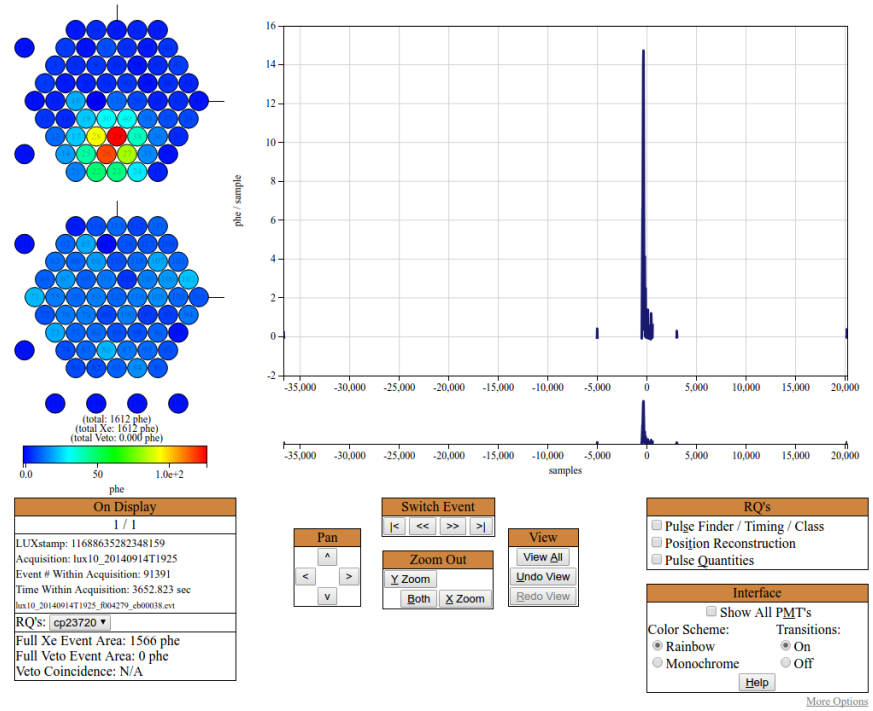
[6] Presumably, the user will usually cut-and-paste these from an analysis output file or other document, as typing luxstamps by hand gets tedious.



(a) Login bar and basic request form



(b) Advanced Request Form



(c) Plots and controls with optional features turned off, in contrast with Fig. 3.2

Figure 3.3: Various Visualux on-screen elements

Once the user requests data, a "Processing" message appears as NEWT communication gets underway. This becomes a "Transferring" message when data transfer to the frontend begins. As soon as the first event is available, the set of graphs and control buttons for exploring data appear. When a user requests multiple events, he/she can begin interacting with the data from the first events even as subsequent events are still loading. The user does not have to wait for all data to transfer before getting started.

When the graphs first appear, not all of the optional features are turned on (Fig. 3.3c). The main graph (upper right) is a plot of the sum of the waveforms recorded by all the PMTs (this sum is sometimes referred to as the "sumpod"). The x-axis is the time before or after the trigger, measured in samples (10ns intervals). The y-axis is the calibrated signal height measured in phe/sample. To the left of this plot is the PMT "hitmap." It shows the integrated signal measured by each PMT, as indicated by the PMT's color according to the labeled colorbar below. The upper hexagon is the detector's upper PMTs, the lower hexagon is the detector's lower PMTs, and the ten separate circles represent the ten output channels from the water PMTs in the water tank. Hovering the mouse over a PMT gives the numerical value corresponding to its color.

To zoom in on part of the waveform plot, the user clicks and drags to select the region he/she wishes to see up close. The plot then zooms in on the selected region in a smooth, animated transition. Every time the view of the waveform plot changes, the PMT hitmap updates to show the integrated signal for just the range of samples being viewed in the waveform plot. So, for example, if the user selects just an S1 pulse, then the PMT hitmap will show how much light was received by each PMT during just that pulse, regardless of what happened at other times during the event.

Below the waveform plot is a second, tiny plot of the sumpod called the "context bar." When the user zooms in, the section of samples they're looking at is highlighted on the context bar by a "brush," rendered as a gray box with a red border. This helps the user stay oriented as he/she zooms in on small details. However, the context bar is also a powerful navigation tool in its own right. The user can click and drag in the context bar to move the brush or either of its edges. Whenever that happens, the waveform plot and PMT hitmap update to reflect the newly selected region. The user can also click the brush to scale the y-axis of the waveform plot to fit the range of the data.[7] Clicking the context bar outside of the brush, on the other hand, will zoom back out to look at the event as a whole. Through a combination of clicking-and-dragging on the waveform plot and using the context bar, the user can very quickly and very naturally navigate around the event to view different parts up close and see their hitmap patterns.

So far, we've only seen how to look at evt information, but Visualux also overlays rq information which the user can add by clicking checkboxes in the "RQ's" menu. Clicking the first checkbox highlights the pulses found by the Pulse Finder rq module (which identifies up to ten pulses per event) and gives the pulse types in a series of tiles below the plots. Clicking the second checkbox shows reconstructed S2 pulse locations, when available, on the

[7] This rescaling also happens automatically when the user moves the brush or either of its edges.

PMT hitmap. Mousing over a pulse's tile or PMT hitmap marker will highlight the pulse in a darker color, and clicking either will zoom in on that pulse. The mousing over will also bring up a shaded box with more rq values if the third checkbox in the RQ's menu is also selected.

If the user wants to see waveforms from individual PMTs, clicking the "Show All PMT's" checkbox creates a second waveform plot with all 100+ individual waveforms. Passing the mouse over a PMT in the hitmap will remove all waveforms except the one for the PMT in question. When the user changes the view on one waveform plot (either the sumpod plot or the individual PMT plot), the other waveform plot will also change its x-axis to keep the two plots lined up with each other.

The "On Display" menu below the graphs provides information about the event (luxstamp, time, etc.) and provides a dropdown menu for toggling between different rq datasets if more than one is available for the event. Additional menus provide traditional navigation buttons for panning, zooming, and undoing view changes. There are also buttons for switching between events if more than one have been loaded.

Visualux includes features to enhance accessibility. The PMT hitmap can be toggled from its default rainbow color scheme to grayscale, to make it colorblind-compatible. Keyboard shortcuts are bound to all of the menu button actions via the Keypress JavaScript library to make the webpage more ergonomically-friendly.

Behind the Scenes

Behind the scenes, the website runs JavaScript code in the user's browser to access the data and respond to the user's interactive requests. Because JavaScript natively supports JSON files, the incoming data can be converted directly to JavaScript variables for subsequent use.

The plots are rendered as two SVG (Scalable Vector Graphics) images: one for the waveform plots and one for the PMT hitmap. The pulse tiles and rq values box are generated as two additional SVG graphics, while the control buttons below are regular HTML form elements. Plotting the data and creating animated transitions between views is handled through the D3 (Data-Driven Documents) JavaScript library. The D3 library works by taking elements of a webpage's DOM (Document Object Model), binding arrays of data to them, and then changing those objects' properties (color, location, etc.) as the associated data is changed. For example, the PMT hitmap colors are created by binding an array of integrated signals to the list of SVG circles representing the PMTs, and then changing the circles' colors to match the data. That process, managed through D3, is repeated every time the user changes the view. D3 permits these changes to occur as smooth animated transitions, by repeating the process many times in sequence with intermediate values.

It's worth emphasizing that these graphics are all generated within the user's browser. Once the data is successfully transferred to the browser, communication with the backend at NERSC ceases and everything else happens locally. This requires the frontend code to be able to run smoothly in a typical browser. Users with many events to view are encouraged to look them up in blocks of no more than 100, to avoid overwhelming JavaScript (or even the

computer's RAM) with too much data at once. For large events, there's another problem to address: Such events can have millions of individual datapoints, and waveform traces that include every point would overwhelm the browser with an abundance of graphical elements. To solve that problem, the data is intelligently downsampled before plotting the waveforms. If multiple points have x-values (sample times) lying within one pixel-width of each other, then only two of those points are retained: the one with the largest y-value, and the one with the smallest y-value. This step reduces the memory size of the waveform graphics and improves performance, with no visible difference in the graphics' appearance.[8]

To allow for testing and developing new features without affecting uptime, clones of the code (including frontend as well as backend) are set up in separate NERSC directories. The codebase is tracked through a Git repository, which makes it easy to propagate changes from these test versions of the system to the production version that end users actually see.

3.2.4 Dataset Preview

In addition to the main Visualux site, a couple of specialty applications of the system were developed. One of these is the "Dataset Preview." During Run 4 WIMP search data-taking, it was arranged for collaboration members to take shifts checking the incoming data for any signs of problems. These were called "offsite shifts" to distinguish them from the "onsite shifts" spent in South Dakota operating the detector. Collaboration members conducting offsite shifts checked various detector-monitoring tools for unusual output. As part of this effort, it was desired that a few representative waveforms from each acquisition be checked, even before the acquisition had finished running, let alone wound its way through the whole data-processing framework. To that end, a script was put in place on the surface server to generate just the very first evt file of an evt dataset and send that one file to NERSC. By going to a modified Visualux page at a different address, users could see a menu to let them view the first 100 events from any of these preview evts. The records from offsite shifts were later used to determine which datasets to use for setting the Run 4 limit.

In fact, the "Dataset Preview" URL actually points to the same Visualux page used for viewing normally-processed evts and rqs. The only difference is an added HTML query string. The Visualux frontend page includes PHP code to store the value of this argument such that it can be read by the JavaScript code, which then displays the correct menus depending on whether the user wants the Dataset Preview or the regular version of Visualux. This process, like various other parts of the frontend code, relies on HTML attributes that render elements invisible without actually removing them from the DOM.

Another specialty application of Visualux was testing the quality of salt (decoy WIMP events), prior to implementing salting for Run 4. A modified version of Visualux was used to show collaboration members a mix of real events and salt events, to see if they could tell

[8]During animated transitions between views on very different scales, a sharp-eyed viewer might notice the sumpod graph fade away only to be simultaneously replaced by a differently-downsampled rendering of the same data. This is the only visible sign of the intelligent downsampling happening in the JavaScript code.

the difference by eye. Nothing was found that was not trivially fixable. This exercise was dubbed the "Turing Test" because the purpose was to see whether people could distinguish something real from a computer-generated facsimile.

3.2.5 Example Pulses and Events

To close this chapter, we will put Visualux to work and look at examples of some of the different signal types seen in LUX.

- Fig. 3.4: A typical golden event in LUX. This background electron recoil from Run 4 WIMP search data has an S1 pulse followed by an S2 pulse. The period after the S2 shows an increase in single electrons (SE), occurring when an electron trapped at the liquid surface makes a delayed escape from the liquid and electroluminesces. This event also includes a few single photoelectrons (sphe), which is when one phototube detects a photoelectron, resulting from either scintillation or phototube dark current. The pulse finder algorithm identifies only the ten largest pulses in each event, which is why not all the sphe's are labeled.

- Fig. 3.5: The same event's S1 pulse. Note that most of the light is collected by the lower PMTs.

- Fig. 3.6: The same event's S2 pulse. Note that the reconstructed X-Y position is closer to the edge than the most strongly illuminated PMTs. That's because of deconvolving the inward push on drifting electrons from the charged PTFE walls.

- Fig. 3.7: A single electron (SE) from the same event.

- Fig. 3.8: A single photoelectron (sphe) from the same event. Note the waveform recorded from a second channel where there is no signal; that happens because the digitizers' pulse-only digitization process works on pairs of channels.

- Fig. 3.9: A typical salt event. Because the salt event is built from real calibration data, it has the same accompanying SEs and sphe's that would be expected in a real event.

- Fig. 3.10: S1 from Kr-83m calibration data, showing the distinctive double S1 from the two-step decay chain of Kr-83m. The Kr-83m S2 (not shown) has the same shape as in Fig. 3.6.

- Fig. 3.11: A pair of S2 pulses generated by double-scattering of a DD neutron in the detector. The S2s are well-separated in this case because the two scattering interactions happened at two different depths, giving them different drift times. Because the time interval between scatters is so small, the scintillation light of both interactions combines to make a single, normal-shaped S1 (not shown).

- Fig. 3.12: An "e-train," in which a large S2 is followed by a train of closely-spaced single electrons. A sufficiently long-lived e-train may be split into multiple events by the event builder.
- Fig. 3.13: An "electron burst," when an electroluminescent signal gradually builds up then fades away. Electron bursts happen when electrons trapped at the liquid surface enter the gas en masse. These events can occur in isolation but are often found within e-trains, as is the case in this example.
- Fig. 3.14: A baseline shift after one or more large S2s. This happens when the digitizer's running average of the baseline is influenced by the negative overshoot following a large pulse, causing the digitizer to mistake the true baseline for another pulse. As a result, the digitizer continues to record the baseline up to a maximum duration of 100,000 samples in the affected channels.
- Fig. 3.15: Electronic noise. Electronic noise can be large enough to activate pulse-only digitization. In this example, the noise occurs repeatedly in the same PMT channel.
- Fig. 3.16: A gas event. When an interaction occurs in the gaseous xenon above the liquid, there's no delay between the emission of scintillation light and the onset of electroluminescence.
- Fig. 3.17: An inverse field region event. When an interaction occurs below the cathode, its S1 light is concentrated in the closest PMT. Such events have no S2 pulse, because their freed electrons are driven downwards from the cathode and never reach the xenon gas.
- Fig. 3.18: Cherenkov radiation in the water tank, detected by the muon-veto PMTs in the water.

Now that we've surveyed the variety of pulse shapes seen in LUX, we will next turn our attention to a specific type of pulse, the S1, to see what information can be gleaned from a closer examination of S1 pulse shapes.

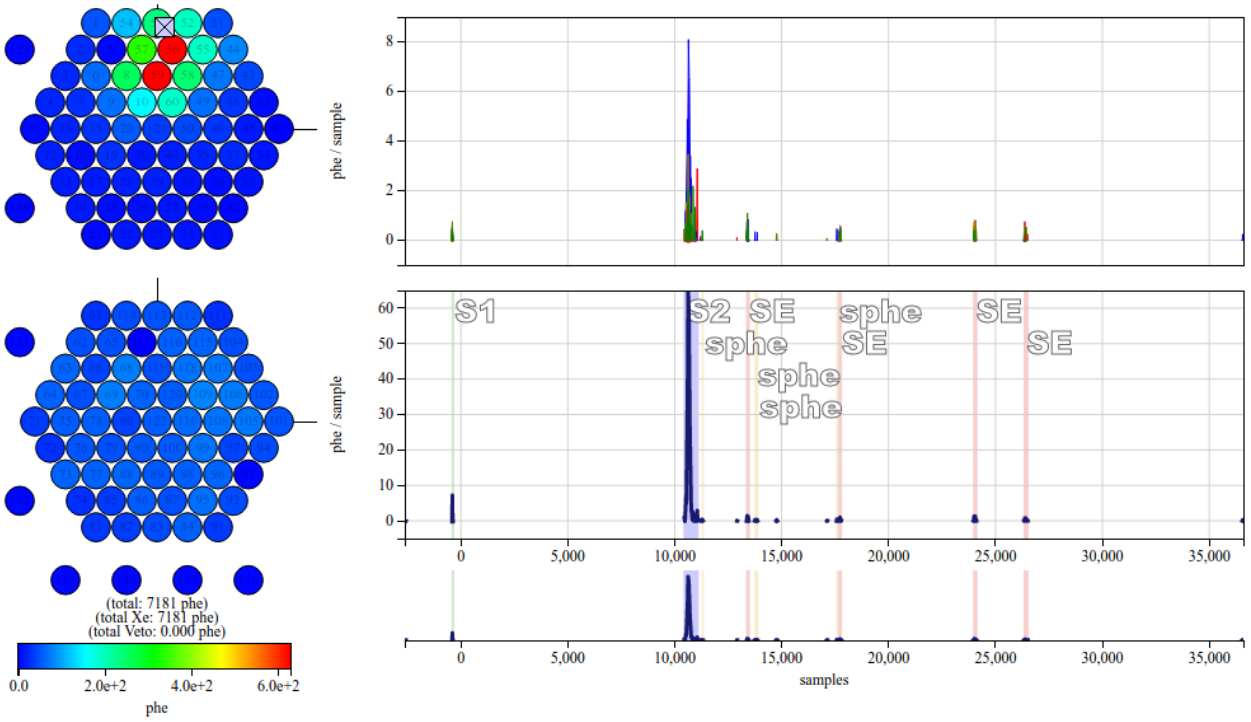


Figure 3.4: A golden event

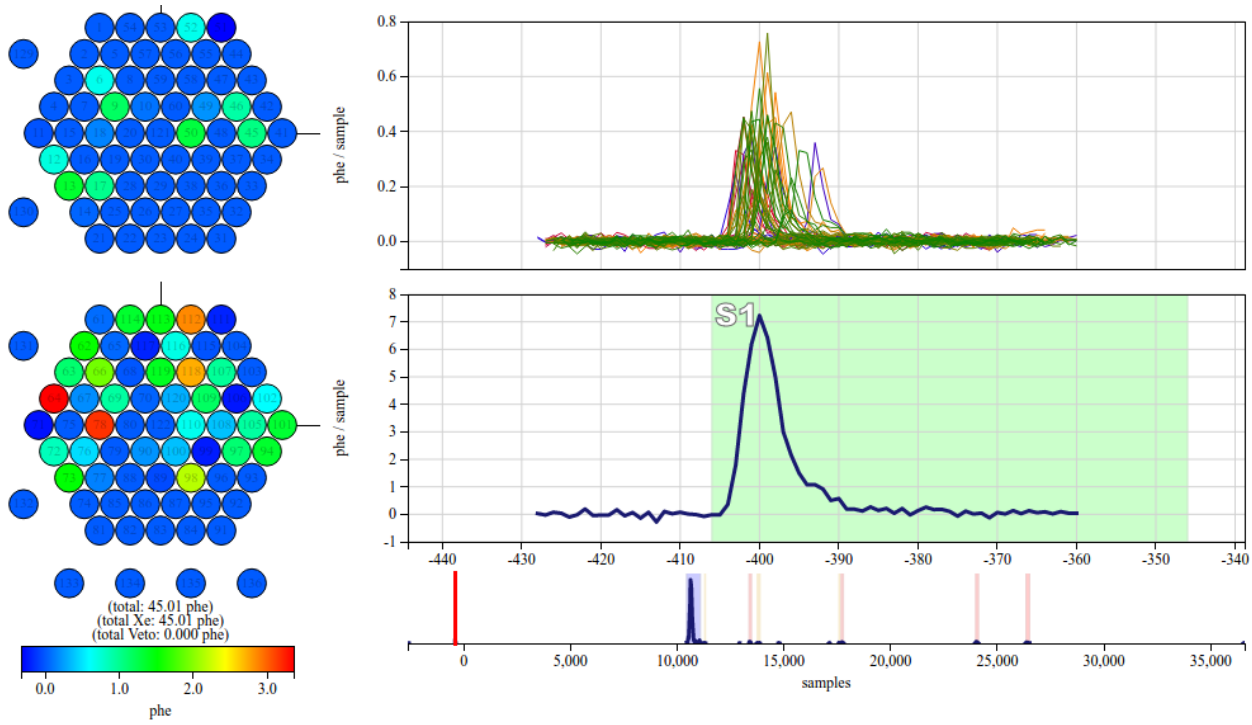


Figure 3.5: S1 pulse

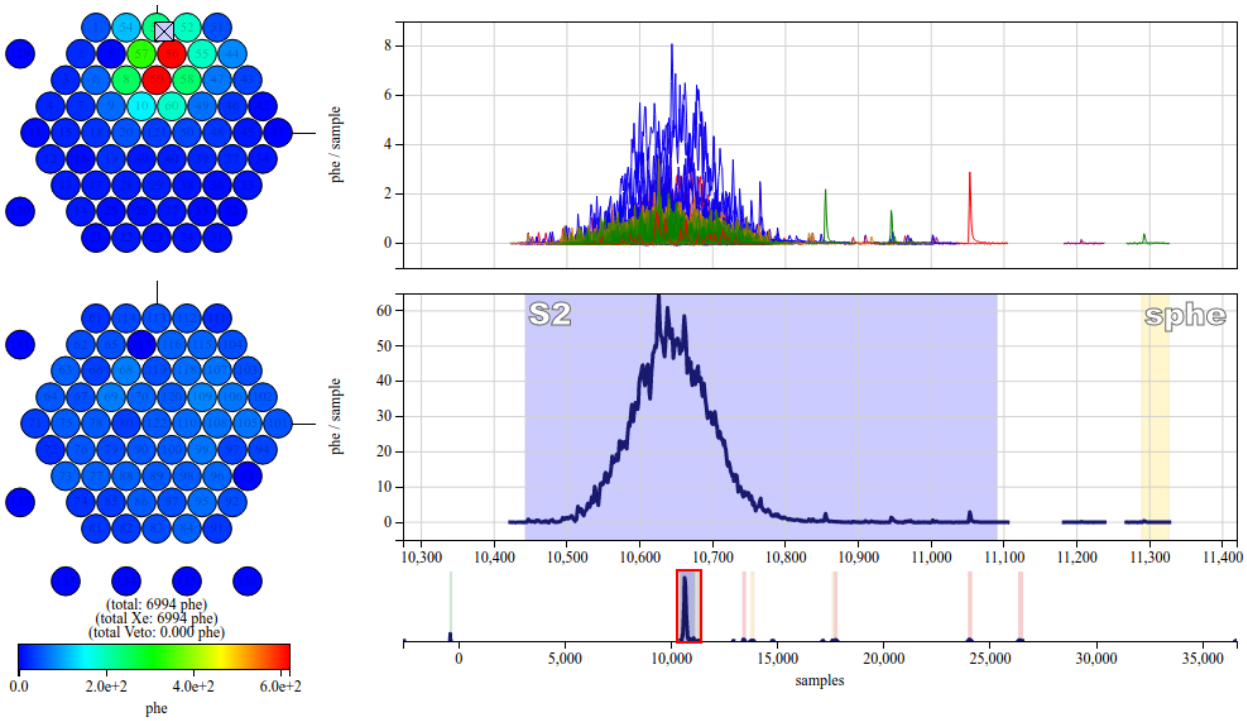


Figure 3.6: S2 pulse

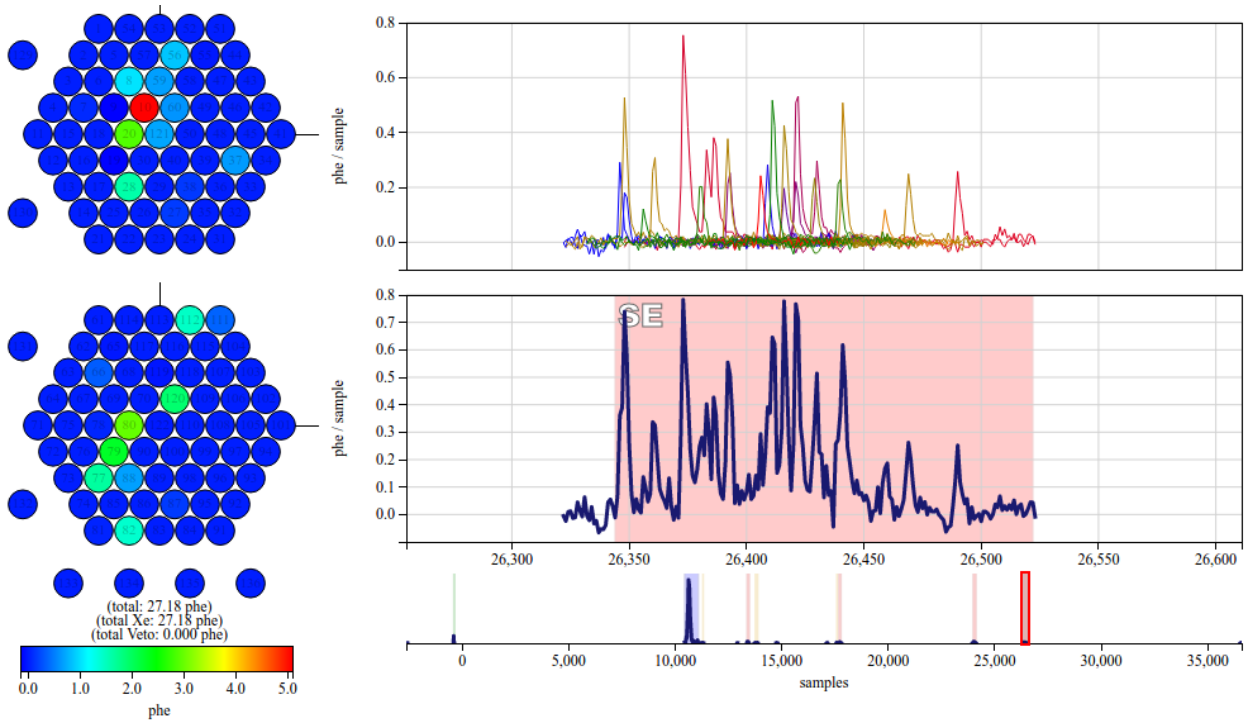


Figure 3.7: Single electron

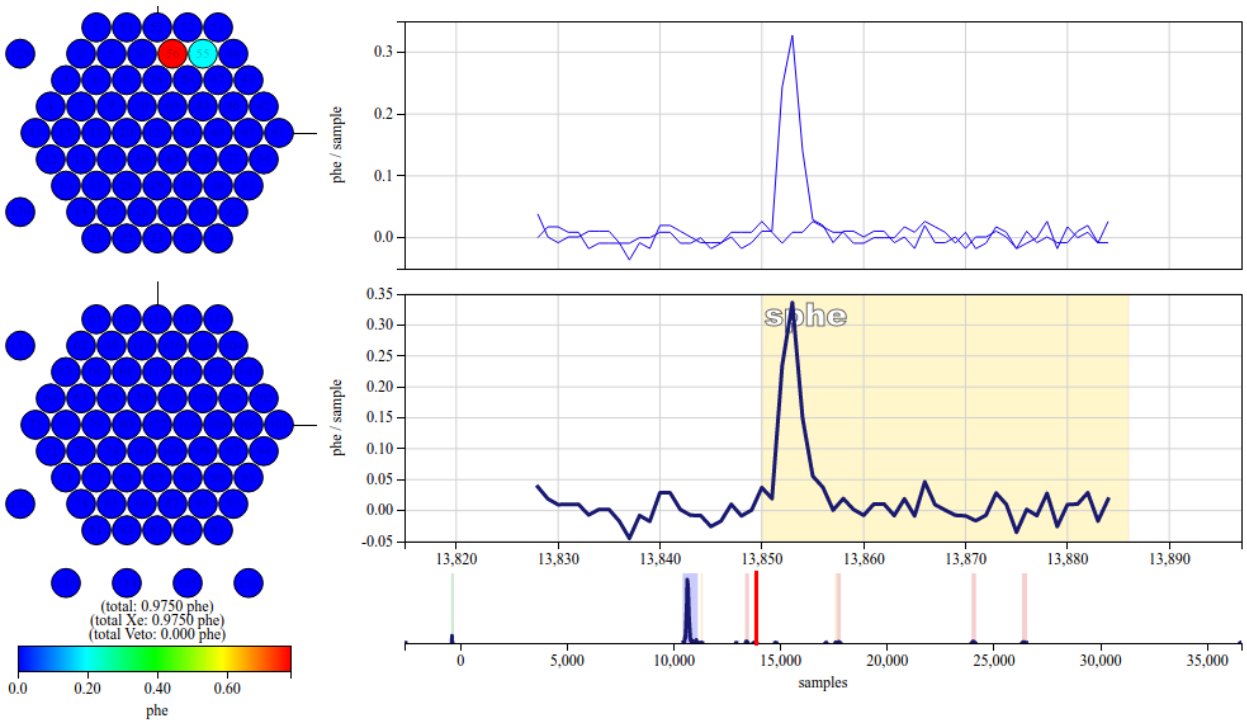


Figure 3.8: Single photoelectron

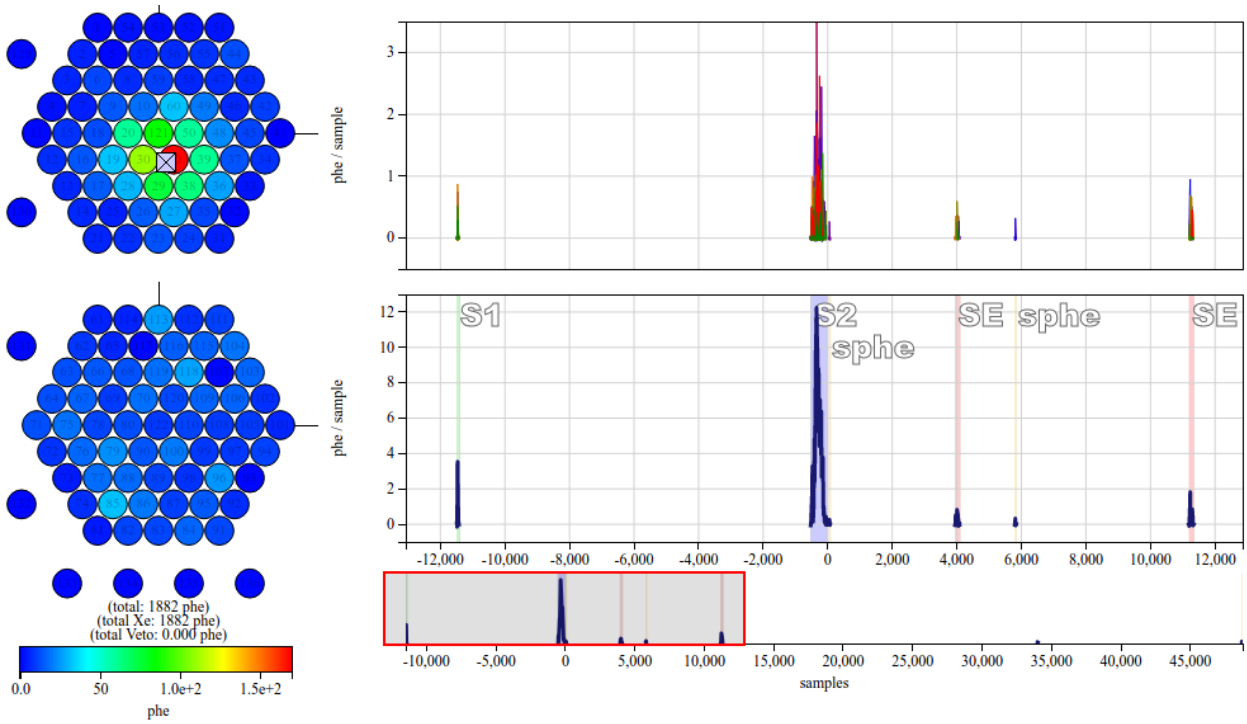


Figure 3.9: A salt event

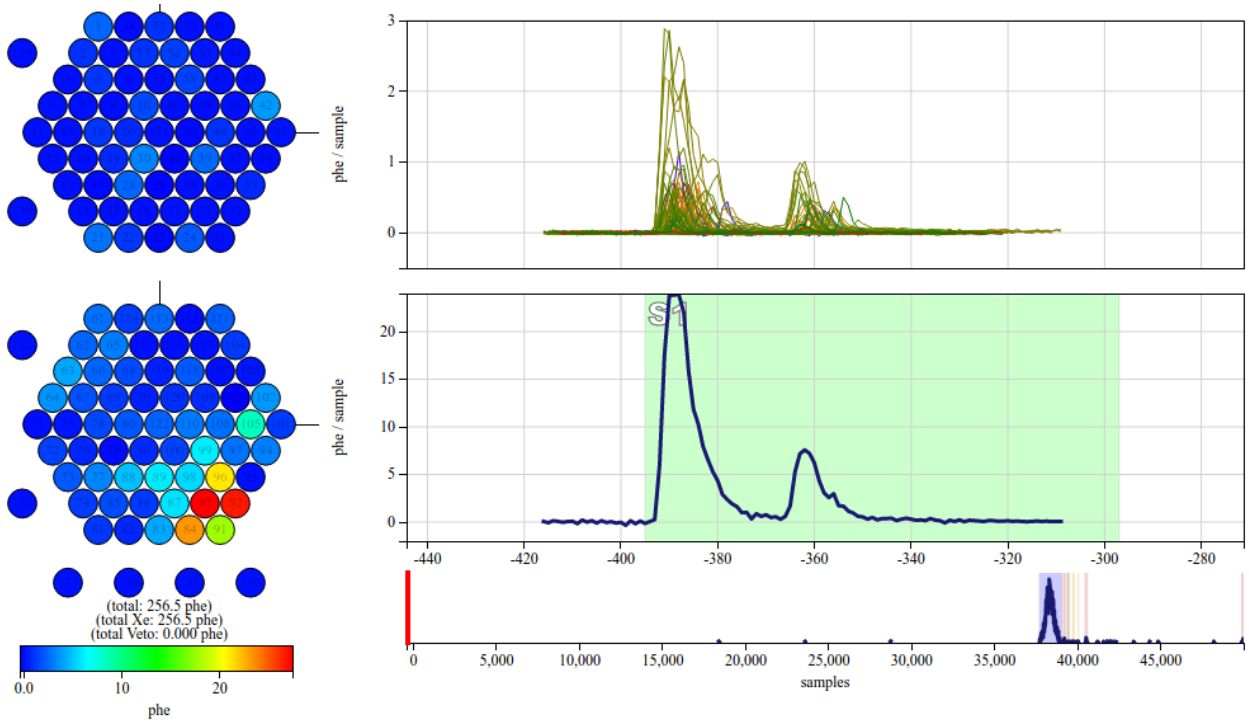


Figure 3.10: Kr-83m S1 pulse

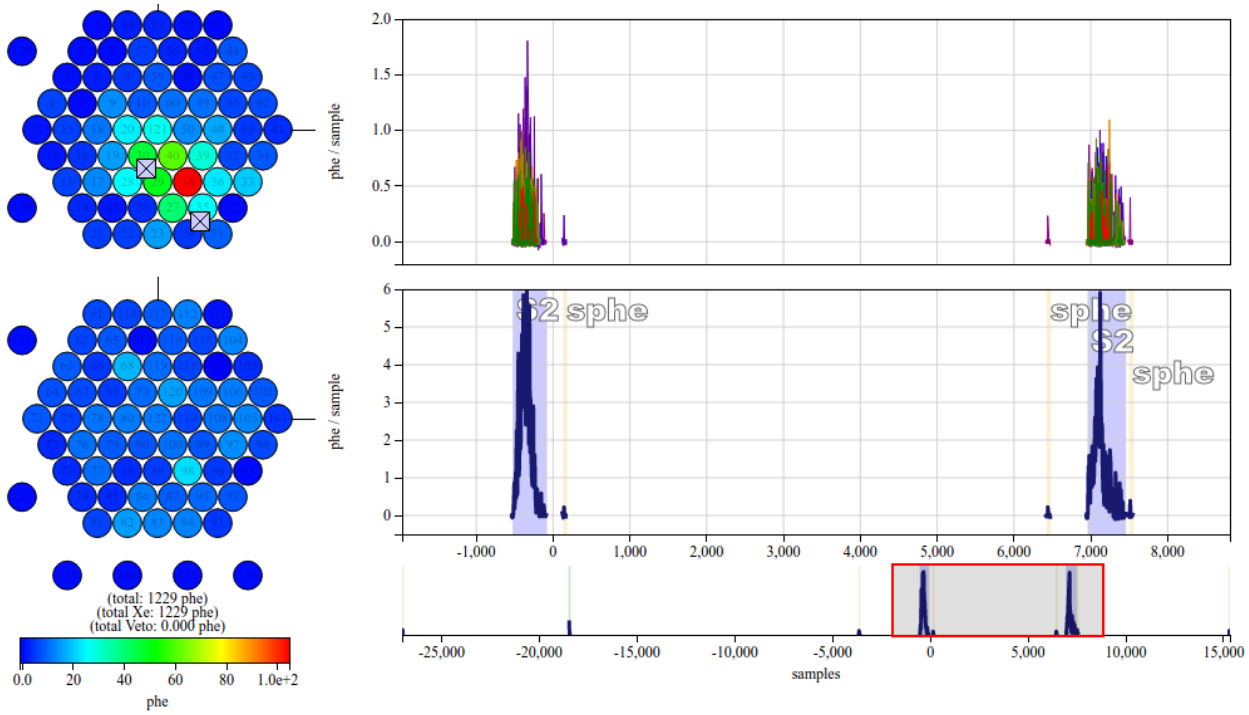


Figure 3.11: Neutron double-scatter S2 pulses

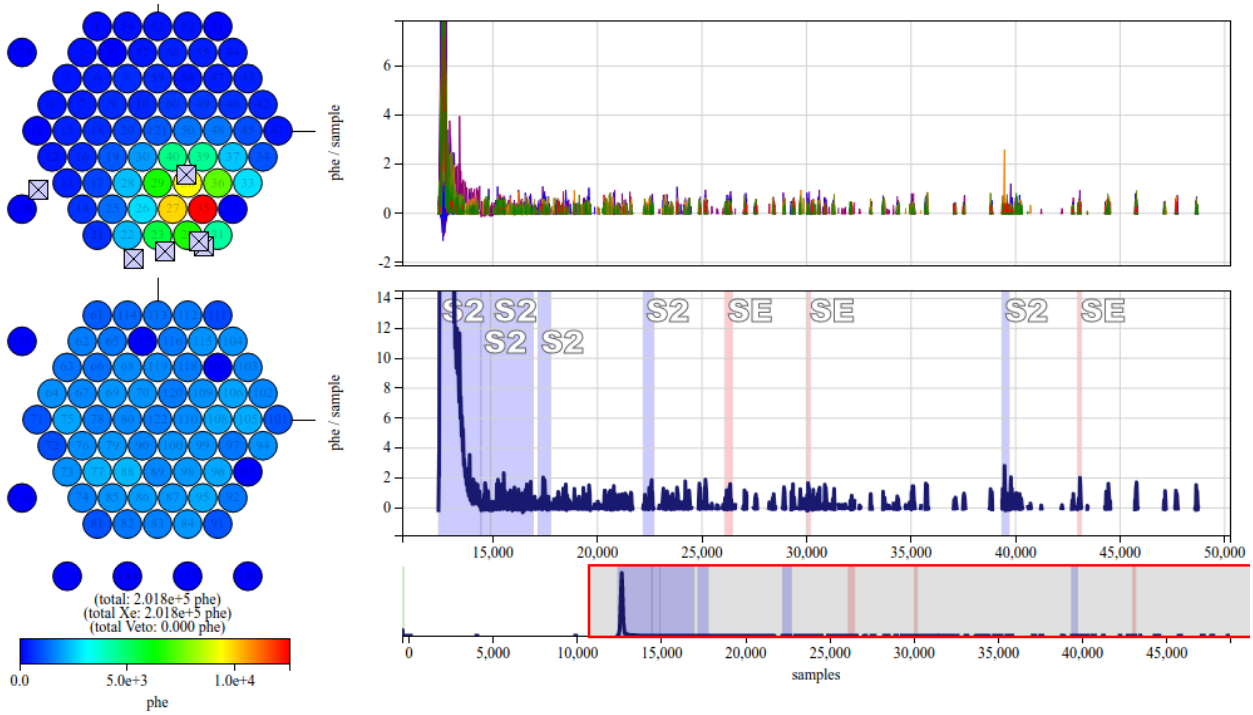


Figure 3.12: E-train

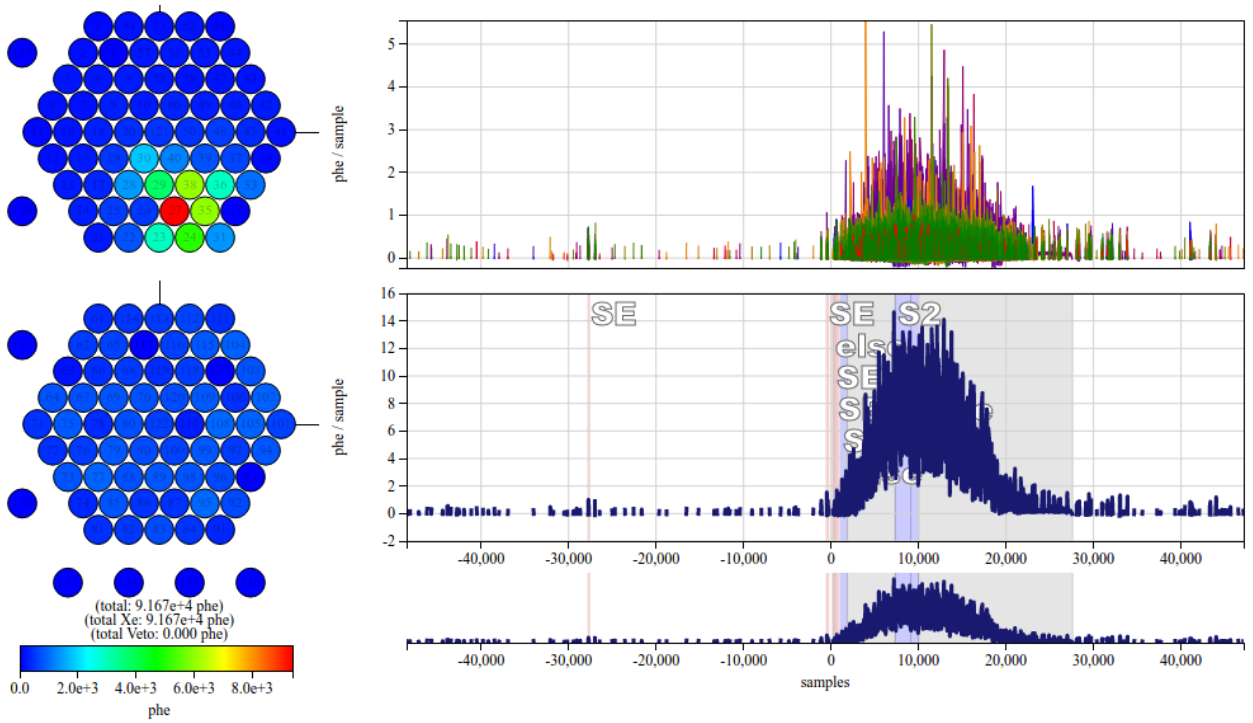


Figure 3.13: Electron burst

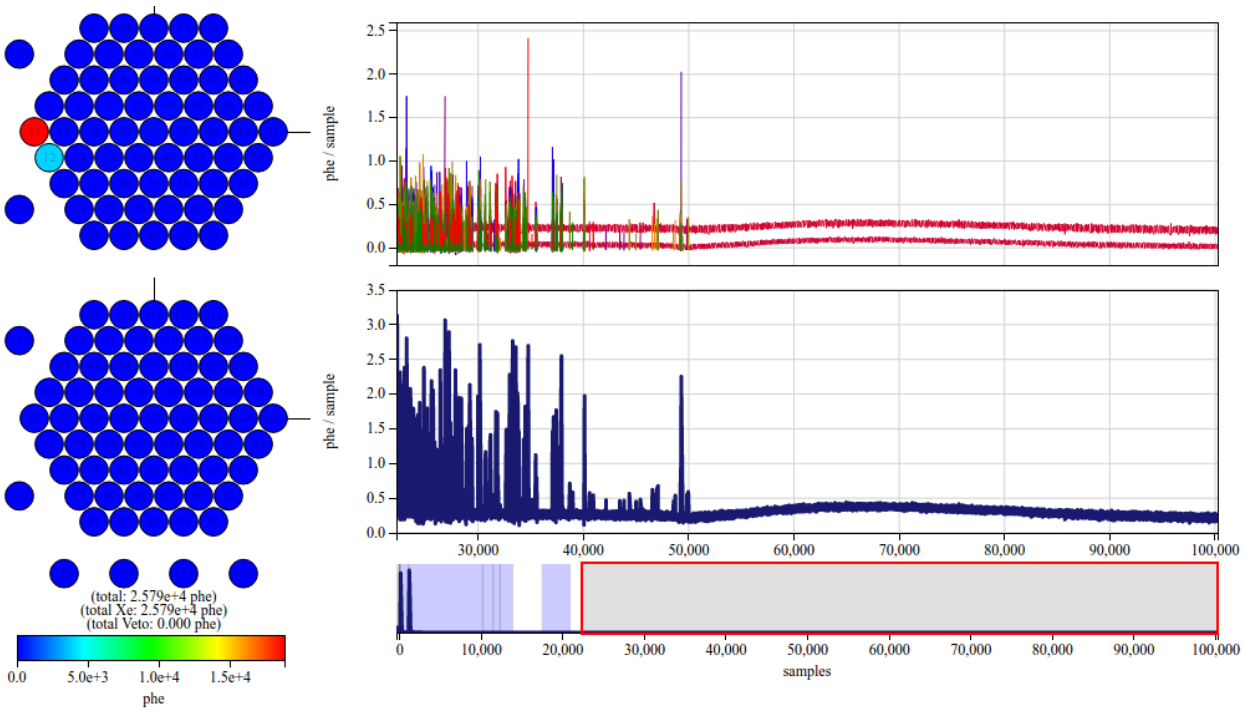


Figure 3.14: Baseline shift

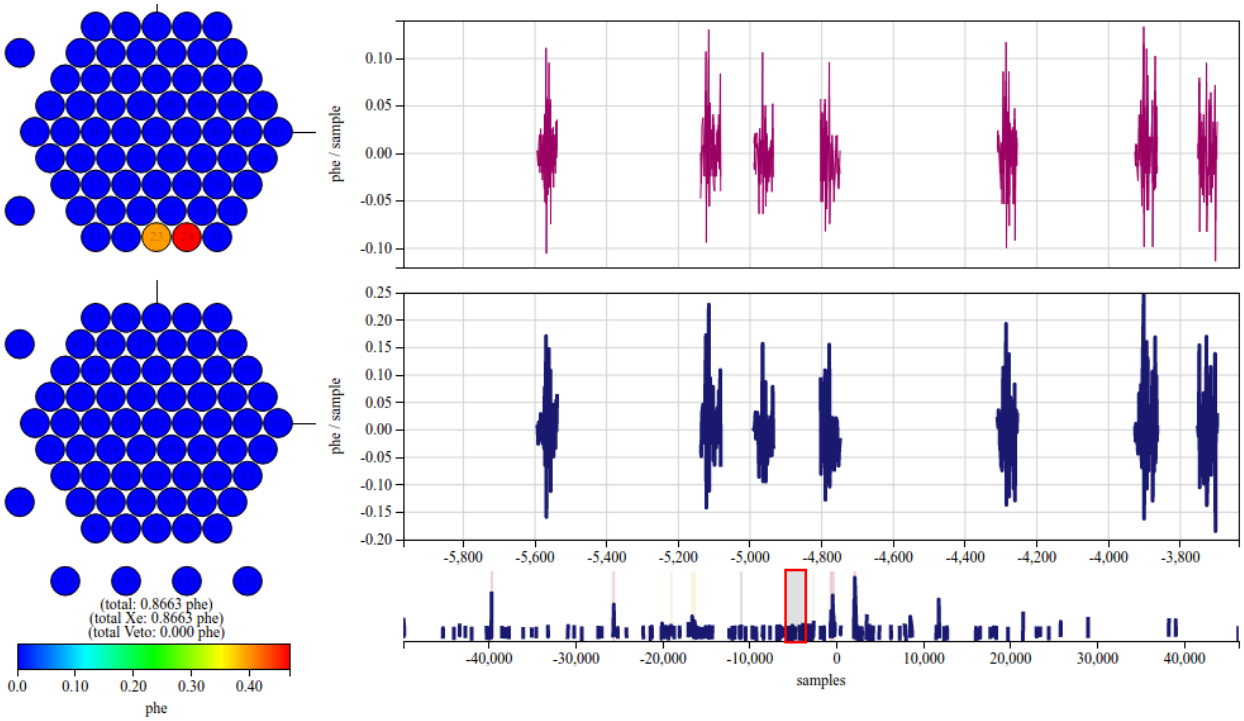


Figure 3.15: Electronic noise

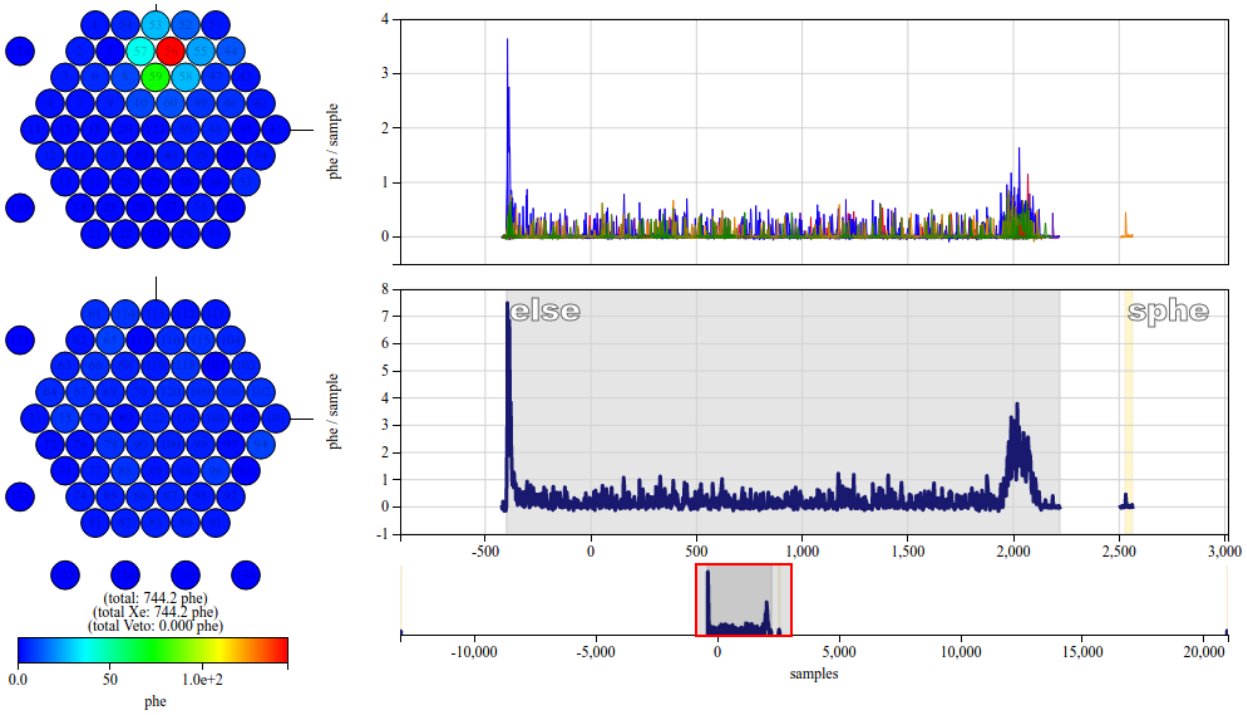


Figure 3.16: Gas event

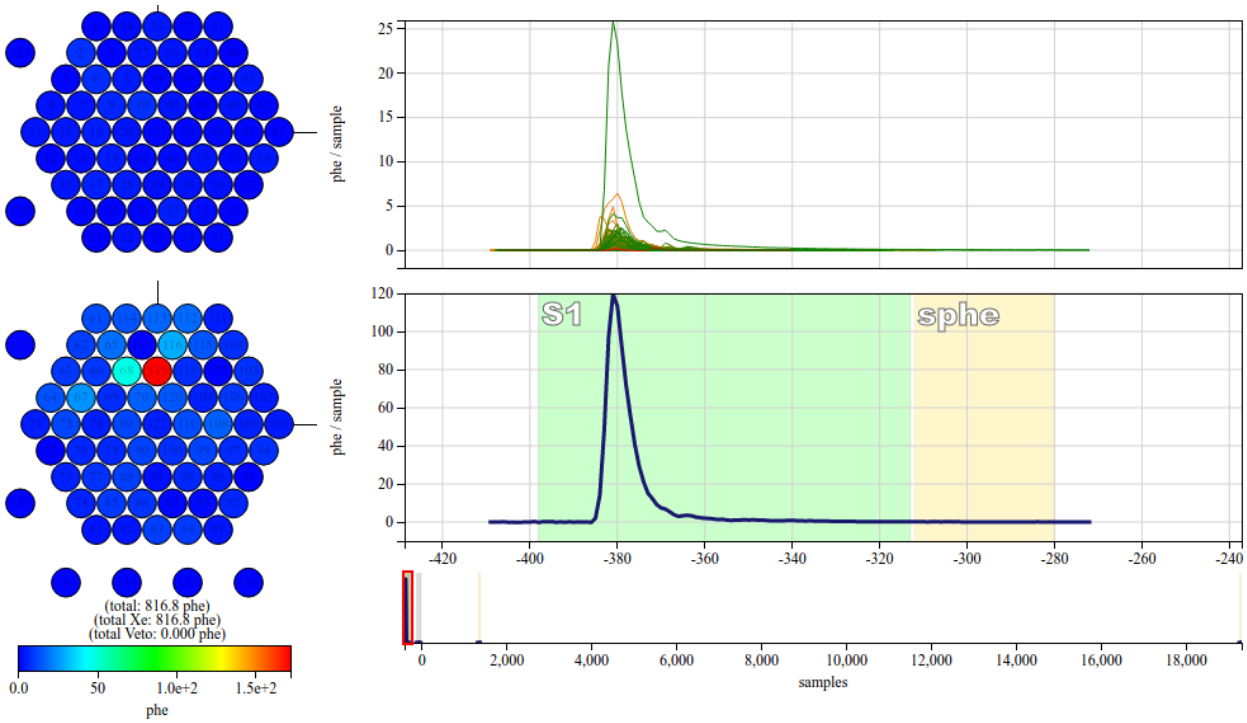


Figure 3.17: Inverse field region event

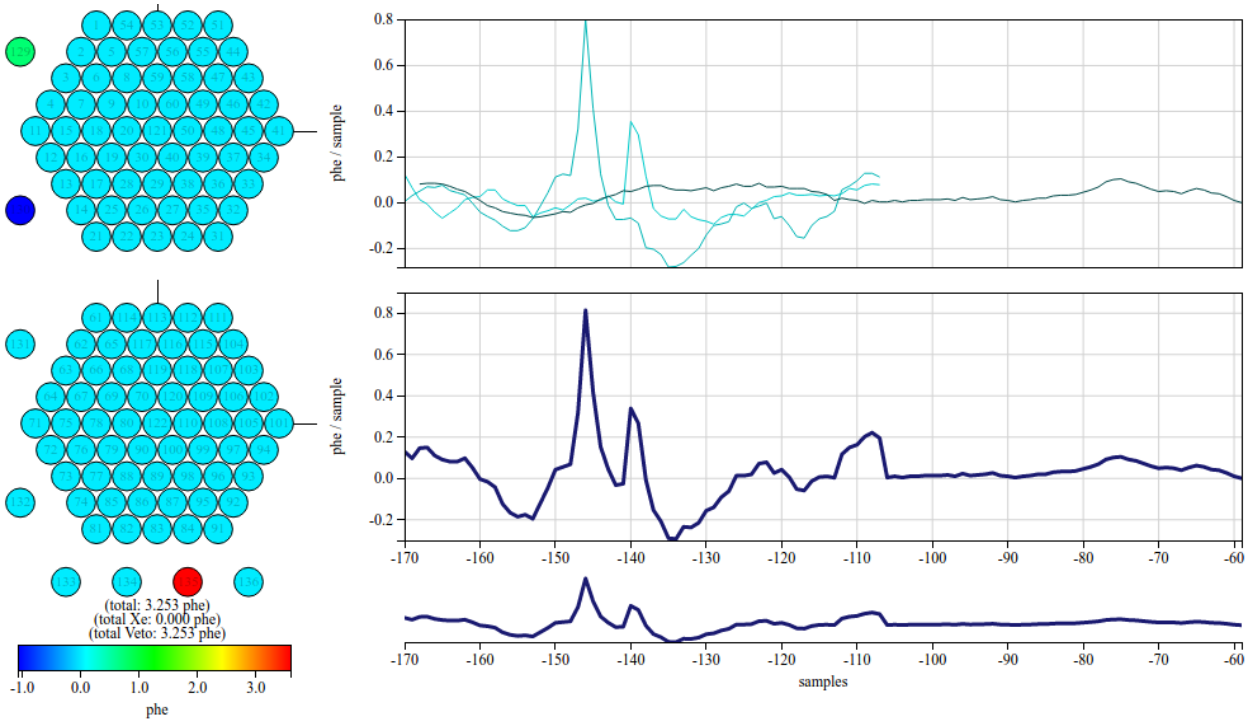


Figure 3.18: Water event

Chapter 4

Pulse Shape Discrimination

When a high-energy particle deposits energy in a xenon target, the resulting S1 scintillation pulse can be a powerful clue to classifying the interaction type and, thus, the source particle. Not only do the relative sizes of the S1 and S2 pulses provide classification evidence, but so does the time signature of the S1 pulse itself. The latter is the basis of the pulse shape discrimination (PSD) technique for helping to distinguish electron recoils (ER) and nuclear recoils (NR) in the xenon. After reviewing the underlying physics of S1 pulse shapes, we will develop techniques for distinguishing ER and NR S1s in LUX.

4.1 Physics of S1 Pulse Shapes

4.1.1 The Energy Deposition Process

Energy deposition in a xenon target begins with a recoiling electron (in the ER case) or nucleus (in the NR case). This particle may be the ionizing radiation itself, as in the case of a beta-minus particle causing an ER or an alpha particle causing an NR. Alternately, the recoiling particle may be a part of the xenon target to which energy is transferred by some other particle, as in the case of an ER induced by an incoming energetic photon via the photoelectric effect or an NR induced by a WIMP via a weak yet-to-be-discovered WIMP-nucleon coupling. Recoils are on the scale of nanometers for NRs and on the scale of many micrometers for ERs[68], but in either case the essential result is the same: the recoil leaves a trail of excited xenon atoms, ionized xenon atoms, and phonons.

What happens next is a series of energy transfers as atoms and electrons interact in the liquid xenon medium, as diagrammed in Fig. 4.1. Each excited xenon atom (or more accurately, each exciton) quickly binds with an adjacent un-excited xenon atom to form an excited dimer, or "excimer," on timescales of picoseconds[81].[1] These excimers decay after a few nanoseconds, resulting in the release of a photon and leaving two neutral xenon atoms behind.[95] The emitted photons comprise the S1 scintillation signal measured in LUX.

[1] These excimer states are called "self-trapped excitons" in some sources[81, 93].

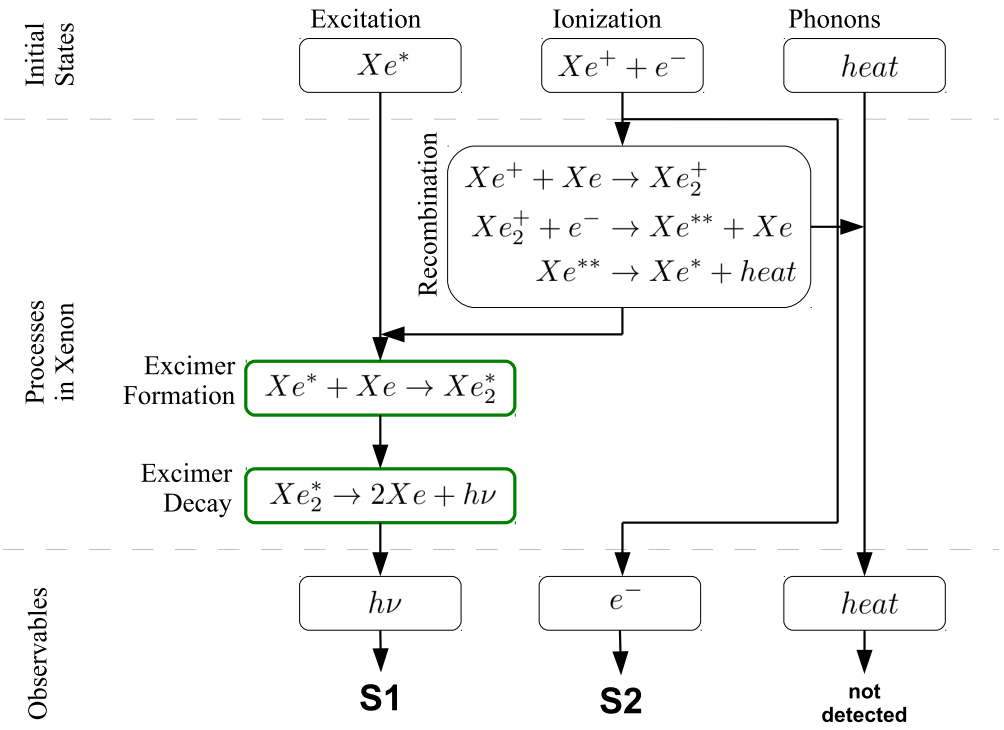


Figure 4.1: Energy deposition processes in xenon and resulting observable signals.

Ionized xenon atoms also bind with a neighboring atom to form a dimer – in this case, a charged unexcited dimer. For electrons released by the ionization of xenon atoms, there are two possible outcomes. One possibility is recombination: the electron binds with a nearby charged dimer, causing the two xenon atoms to separate and leaving one highly excited.[2] The highly excited xenon atom loses some energy to phonons and eventually forms an excimer, from which point the process proceeds just like the case of atoms excited by the initial recoil.[95]

For ionized electrons that do not recombine, the applied electric field draws them away from the vicinity of the recoil. Those that are not trapped in xenon impurities in the liquid volume or at the surface go on to contribute to the event's S2 electroluminescence signal. Recombination, in short, diverts one quantum of S2 signal into the S1 channel instead.

Although not shown in Fig. 4.1, Penning quenching is a mechanism with the opposite effect. Interaction between two xenon excimers (or one excimer and an impurity), can result in the excimer energy going into ionization and phonons instead of the usual scintillation[100]. Scintillation is also reduced by biexcitonic collisions, or collisions between two excitons prior to excimer formation[106].

[2]Prior to recombination, a sufficiently energetic electron will first shed energy by producing more ions and excitons and thermalizing[81].

4.1.2 ERs and NR Timescales

Both ERs and NRs induce all of the processes described above.[3] The difference between ER and NR S1 pulse shapes stems from the fact that two types of excimer can form[95]. In a singlet excimer, the excited electron is in a spin singlet state with its unexcited unpaired former partner electron. In a triplet excimer, the excited electron is in a spin triplet state with the unexcited unpaired electron. These two types of excimer have different decay times: 3.1 ± 0.7ns for the singlet and 24 ± 1ns for the triplet[108].

The key detail is this: Nuclear recoils preferentially populate the singlet state, and electron recoils preferentially populate the triplet state. For electron recoils induced by gammas, the singlet-to-triplet ratio is 0.17 ± 0.05 in the case of excitons from direct excitation and 0.8 ± 0.2 for excitons formed through recombination. On the other hand, the singlet-to-triplet ratio for neutron-induced nuclear recoils is 7.8 ± 1.5.[108]

In general, the singlet-to-triplet ratio increases with increasing linear energy transfer (LET, the energy loss per distance). The reason for this relationship is not well understood. Hitachi et al.[81] raise no fewer than six different hypotheses before settling on a mechanism related to recombination, which which a more recent author concurs[95]: The triplet state excimer is of slightly lower energy than the singlet state, so interactions with electrons produced via ionization can convert singlets to triplets. Because recombination happens more slowly in the low-LET case, thermal electrons have more time to cause such conversions, decreasing the singlet-to-triplet ratio.[81, 95].

The basic shape of the S1 pulse is roughly exponential. On the scale of nanoseconds, the contribution from direct excitation has essentially zero rise time, while the contribution from recombination has a short rise time related to the recombination timescale, which in turn is related to the LET and external E-field. These two contributions are both mixtures of singlet and triplet states, with each state having its characteristic falloff time.

This original signal is then temporally smeared out during the process of measuring the scintillation light. The major cause is that different photons travel very different distances before being detected by a photomultiplier. An individual photon may bounce many times off of the detector's UV-reflective PTFE walls before entering a PMT, spreading the arrival times of contemporaneously-emitted photons by as much as tens of nanoseconds. Temporal variations in electron-cascade times in the PMTs, as well as the process of pulse shaping and discrete 100 MHz digitization, further blur the original pulse shape. These blurring processes obscure the original difference between the ER and NR pulse shape. But as we will see, on average a subtle difference does remain.

4.1.3 Previous Pulse Shape Studies

The subject of pulse shape discrimination in a dual-phase xenon detector has been investigated previously. Kwong et al.[95] developed Xed1b, a dual-phase xenon detector with a 10g

[3]However, the extent to which each process occurs does differ. For example, the fraction of energy that winds up as phonons is much smaller for ERs than for NRs.

active volume. Although small, this detector matched the design of LUX in a number of key respects, including the same arrangement of five grids, and PMTs both above and below the active volume, with two PMTs in total to span the small detector. It was shown that using S2-vs-S1-size in concert with PSD resulted in superior ER/NR discrimination compared to S2-vs-S1-size alone. It was also shown that the effectiveness of PSD in liquid xenon decreases with increasing applied electric field up to 1 kV/cm and is roughly field-independent beyond that.[95, 94]

Around the same time, an analysis of PSD was undertaken using XENON10, a dual-phase xenon detector for dark matter searches with a 13.7kg active volume. As with Xed1b, the difference between ER and NR pulse shapes was found to be greater for higher recoil energies.[34]

In addition to dual-phase xenon detectors, related dark matter detector technologies have also benefited from pulse shape discrimination techniques. Because all-liquid xenon detectors do not need to extract electrons with an applied electric field, they show a larger PSD effect. Such was the case for the XMASS collaboration, which investigated PSD in a test setup with an active volume of 0.58 kg, in preparation for a similar analysis of their full 800 kg experiment[141]. Likewise, the ZEPLIN I liquid xenon dark matter detector with an active volume of 5kg (of which 3.2 kg was in the fiducial volume) also used PSD as the primary method to distinguish ER and NR events[29]. This is in contrast with ZEPLIN II, which was a two-phase detector[30].

Finally, pulse shape discrimination is routinely employed in argon detectors. That's because the argon singlet excimer has a decay time of 7.0 ± 1.0 ns while the triplet has a much larger decay time of 1.6 ± 0.1 μs[81]. This difference, exceeding two orders of magnitude, makes ER and NR pulse shapes very distinguishable in argon. This is used in experiments such as DarkSide[7].

Compared to the previous experiments in dual-phase xenon[95, 34], the PSD effect is smaller in LUX because the LUX detector's large size exacerbates the temporal spreading of the signal caused by different photons traveling paths of different lengths as they repeatedly bounce off the PTFE walls surrounding the active volume. This makes LUX an especially challenging system for extracting PSD information.

4.2 Pulse Shape Discrimination in LUX

In this section, we develop a method for determining whether a given event is more likely to be an electron recoil or a nuclear recoil, given the shape of its S1 pulse.

4.2.1 Calibration Data

In order to measure pulse shape discrimination in LUX, we require a collection of known ER events in the active volume, and a similar collection of known NR events. Fortunately, such events are available in the data captured as part of LUX's extensive calibration regime. As a

source of ERs, we will use the tritium calibration data discussed in Section 2.5.2. For NRs, we draw on the DD neutron calibration data discussed in Section 2.5.3.

Except where noted otherwise, this analysis will use data from Run 4, discussed in Section 2.6. Compared to the previous science run (Run 3), Run 4 offers a longer WIMP search livetime and a larger variety of calibration data. For nuclear recoils from DD, the specific calibration data used here comes from LUX's "Multi-Z" DD calibration campaign from 16 Sept. 2014 through 25 Sept. 2014. During this campaign, the height of the DD gun was periodically changed, to record data from the collimated neutron beam entering the detector at five different depths. The five depths were 7.5, 15, 22.5, 30, and 37.5 cm below the liquid surface, corresponding to drift times of about 56, 101, 151, 203, and 260 μs, respectively. For electron recoils from tritium, the specific calibration data used here was recorded from 25 Feb. 2015 through 26 Feb. 2015 following a tritiated methane injection. The five-month separation between these two calibration efforts is reasonably small compared to the 20-month duration of Run 4.

4.2.2 Calibration Cuts

Having selected the ER and NR calibration acquisitions to use for comparing S1 pulse shapes, the next step is to apply a series of cuts to the data. When analyzing WIMP search data, an event is cut (that is, excluded from the analysis) if it fails any of various tests designed to weed out events that are less likely to be from clean single scatters in the detector. Applying similar cuts to the calibration data ensures that the model of S1 pulse shape developed here will be applicable to the actual WIMP search candidate events that we ultimately want to study with the model. Therefore, the calibration cuts used here are very similar to the cuts used for Run 4 limit setting in Ref. [20], with the exception of the cut on event position in the detector, a cut which requires some extra consideration unique to the calibration data. We will save the discussion of the cut on event position (the "geometric" cut) for last, and discuss the other ("nongeometric") cuts first.

Nongeometric Cuts

To develop a model of pulse shape discrimination that is applicable to events passing Run 4 analysis cuts, the calibration data is subjected to essentially the same cuts, with only minor differences as discussed below. These cuts are:

- *Golden event cut*: Only golden events (discussed in Sec. 2.4.3) are considered. For the calibration data, the Run 3 reanalysis (Ref. [16]) definition of golden events is used instead of the Run 4 (Ref. [20]) definition, but the difference is negligible in practice.
- *Bad area cut*: For this cut, the "good area" of an event is taken to be the sum of the integrated signal collected during the S1 and S2 pulses, while the "bad area" is the integrated signal collected at all other times during the event. Events with too much bad area for a given good area are likely to be electron trains or other anomalies, so

they are excluded. This cut is implemented for calibration data in exactly the same way as for the Run 4 WIMP search data.

- *S1 requirements*: These checks on the S1 pulse are implemented identically for calibration data, with the following exception. Although the Run 4 analysis cuts restrict S1 size to be below 50phd, this requirement is waived for the calibration data in order to extend the analysis up to 100phd, near the calibration data's maximum available S1 size.

- *S2 requirements*: Although a minimum S2 size of 200phd is imposed, other Run 4 S2 quality cuts designed to flag difficulties with fitting the S2 shape were not used for the calibration data. These cuts only affect a small percentage of otherwise acceptable events.

- *Band cut*: Events more than three sigma above the ER band mean or more than seven sigma below the NR band mean in $\log_{10}$S2 vs. S1 space are assumed not to be correctly-reconstructed events of interest.

- *Post-unblinding cuts*: Two cuts implemented after Run 4 unblinding were not used for calibration data. Like many of the S2 requirements cuts, these were useful for removing small populations that could feign a dark matter signal, but they were not necessary for the large numbers of calibration events being analyzed in aggregate.

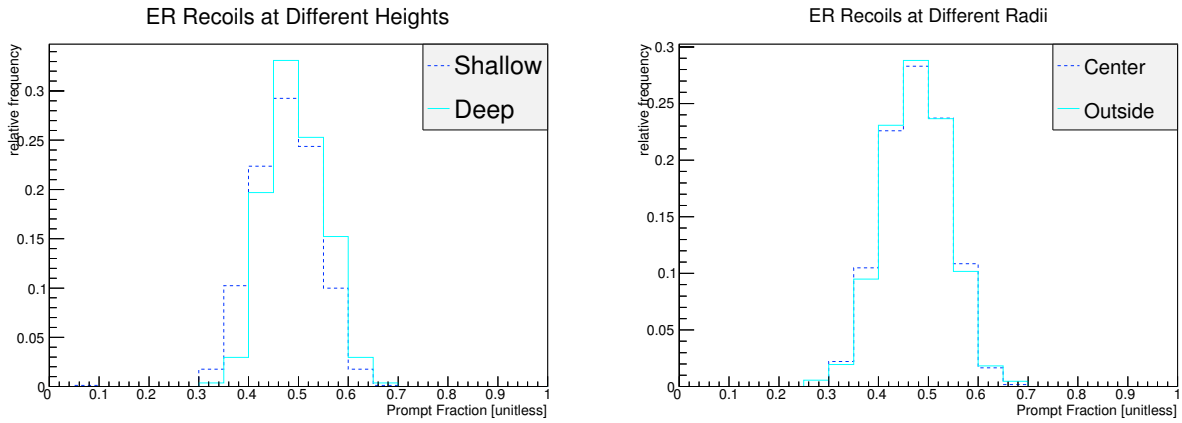
  Approximately 68% of golden ER calibration events and 43% of golden NR calibration events passed all of the implemented nongeometric cuts.

The Geometric Cut

There is one significant way in which nuclear recoils in the DD calibration data differ from nuclear recoils due to WIMPs. The WIMP signal would be uniform in the detector's active volume, whereas the DD neutrons are delivered in a collimated beam. (The tritium calibration presents no such difficulty because the tritiated methane source is also uniformly spread throughout the active volume following a tritium injection.) As will be shown next, this difference matters because pulse shape depends on event depth in the active volume, in addition to event type (ER or NR) and energy.

A number of variables affect the average shape of a golden event's S1 pulse. As discussed in Sec. 4.1.2, the shape is affected by whether the event is an electron recoil or nuclear recoil. The S1 pulse shape is also affected by the event's energy, for which S1 size can serve as an approximate proxy,[4] because initial event energy affects linear energy transfer. A related question is whether or not pulse shape is also dependent on the position of an interaction in the detector. A theoretical argument would suggest that shape should be affected by depth in the detector, but not by radial displacement from the detector's central

[4]The exact relationship between S1 size and energy is Eqn. 2.1.



(a) (b)

Figure 4.2: Electron recoil prompt fraction histograms for various regions of the active volume. (a) Shallow: 40μs < drift time < 105μs, Deep: 235μs < drift time < 300μs. (b) Center: delensed radial distance < 10cm, Outside: delensed radial distance > 20cm. All events used have S1 sizes in the 40 - 50phd range.

axis. The argument goes like this: If more photons travel directly to PMTs where they are detected, then more of the S1 signal should arrive right at the beginning of the pulse. On the other hand, if more photons bounce around the PTFE walls repeatedly, then the signal is more spread out in time. Most of the S1 signal is measured at the lower PMTs, so events happening low in the detector will see more of their photons (up to a maximum solid angle of almost 2π) impinge directly on a PMT, while events higher in the detector will see more of their photons bounce off the PTFE one or more times, temporally spreading the signal. Meanwhile, a radial variation in position shows no similar effect. Events nearer to a wall will have more photons undergo a single reflection than events near the center, but that does not change the photon arrival time by as much as repeated reflections.

This geometric argument is supported by calibration data. "Prompt fraction," defined precisely in Sec. 4.2.3 below, is a statistic that characterizes the pulse shape. Fig. 4.2 shows a histogram of prompt fraction for various subsets of the tritium calibration data passing the nongeometric cuts. A small systematic shift to higher prompt fractions can be seen for events lower in the detector as compared to those at shallower depth. However, there is no similar offset between small and large radial distances.[5] In light of this depth-dependence, we regard event type (ER or NR), S1 size, and drift time as the independent variables affecting the distribution of pulse shapes.

A natural way to apply a cut on event position is suggested by the spatial distribution

[5]In Fig. 4.2, the shallow, deep, center, and outside data sets have 800, 269, 544, and 875 events, respectively. In Fig. 4.2b a depth-dependent radial cut is imposed, and in Fig. 4.2a only events with drift times between 40 and 105 μs are used. Plots for NR are not shown because the lower statistics make the trend less evident.

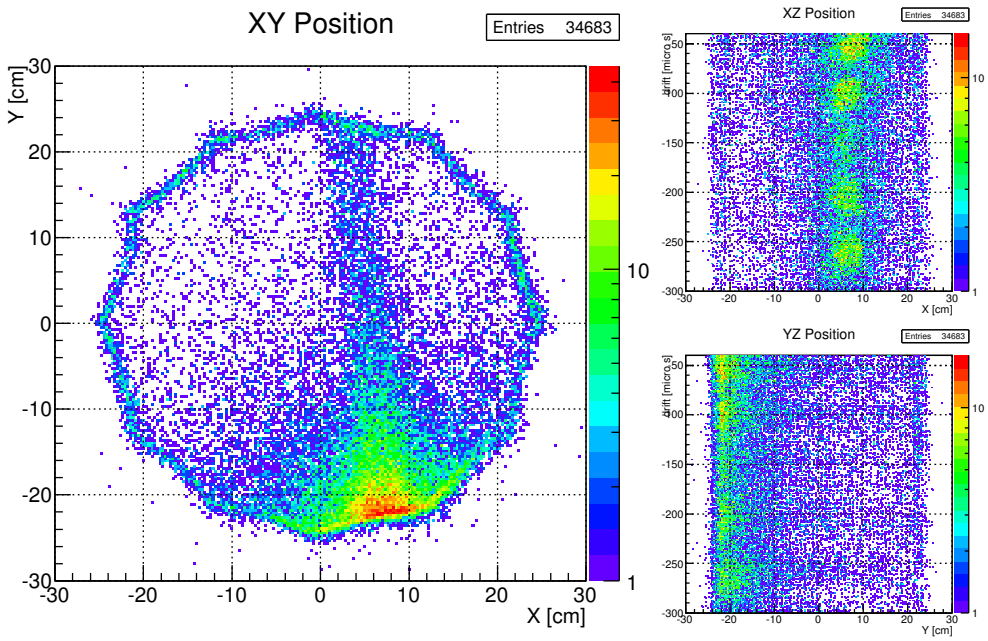


Figure 4.3: Delensed positions of all multi-Z DD calibration events passing nongeometric cuts.

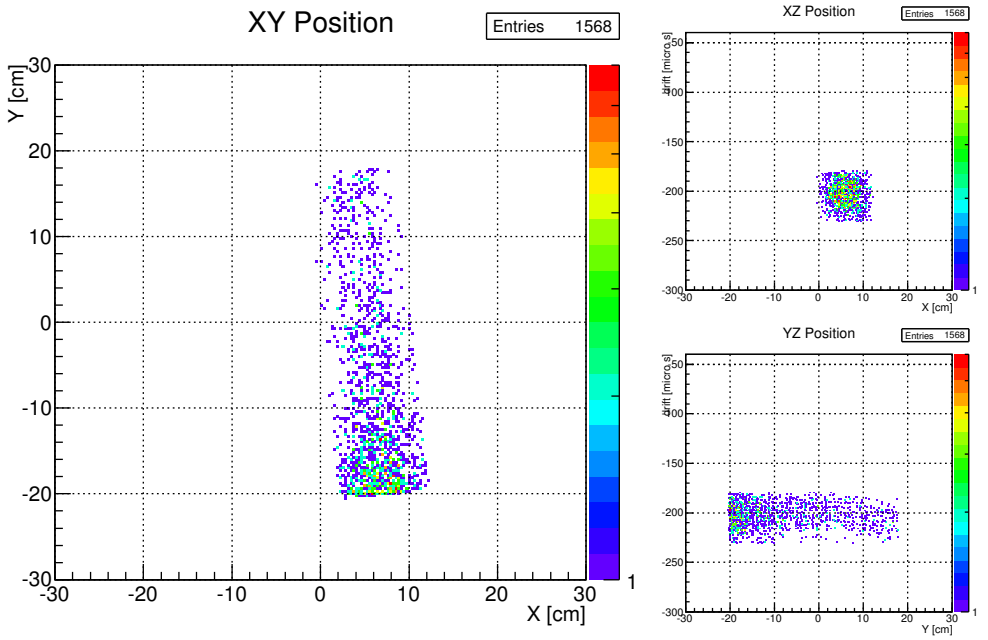


Figure 4.4: Delensed positions of DD calibration events from the second-lowest beam depth which pass both the nongeometric cuts and the geometric cut.

of the NR calibration data. Fig. 4.3 shows the positions of all NR calibration events passing the nongeometric cuts. Data was collected with the DD gun at five different depths (leading to the five bright spots in the data's projection onto the XZ plane), but in each case the horizontal orientation of the source was unchanged (as seen by the alignment of the beams in the XY plane). For each of the five beam positions, a geometric cut was designed to get as clean of an NR sample as possible. This was done by selecting events in the collimated neutron beam, while cutting away events near the wall and elsewhere in the xenon bulk. A cut on delensed X and Y coordinates[6] was used to rule out XY positions outside of the beam, while a drift time cut (specific to each DD gun position) similarly addressed the Z position. Fig. 4.4 shows just the NR data from the beam's second-lowest position, with the corresponding geometric cut having been applied.

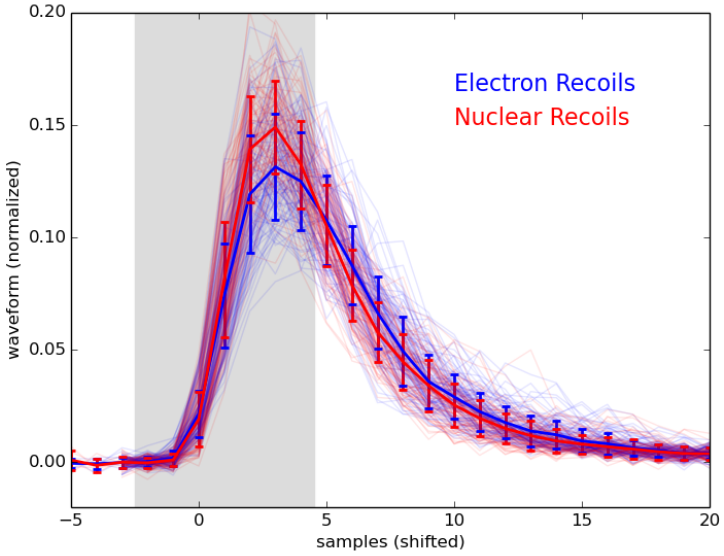
This geometric cut eliminates events in the NR calibration that are less likely to be initial single scatters from DD neutrons. For absolute consistency, the same geometric cut was also applied to the tritium data, even though the evidence presented above suggests that the XY cut on tritium was not really necessary (because of the radial independence of the pulse shape). Since the tritium ER calibration data has higher statistics than the NR data from the DD gun, the choice to use the same XY cut for ER does not substantially affect the analysis.

4.2.3 Prompt Fraction

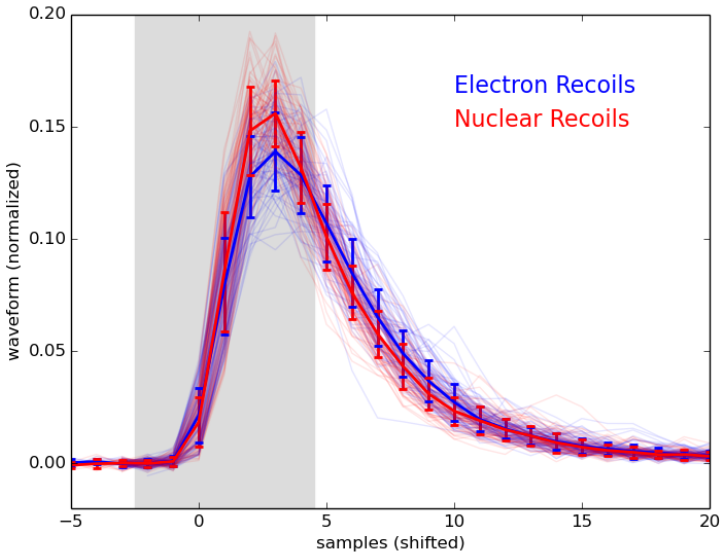
With the calibration datasets identified and appropriate cuts imposed, the events to be used for studying pulse shape have been fully specified. The next step is to identify a figure of merit for characterizing the the shape of an S1 scintillation pulse. This should be a single number for each event, with a statistical distribution for ER events that is as separated as possible from its distribution for NR events of the same depth and S1 size. To motivate the figure of merit used here, we start by inspecting the S1 pulses themselves.

Fig. 4.5 shows overlaid S1 pulses from the ER and NR calibration data. These are "sumpod" waveforms, meaning that they were generated by summing the signals from all of the detector's PMTs at each 10ns time step. In each plot, both the ER and NR pulses come from the same depth window (centered on the second-lowest depth used in the multi-Z DD calibration data, as shown in Fig. 4.4) and the same energy window (40-50phe for Fig. 4.5a and 70-80phe for Fig. 4.5b). To more easily compare the waveforms from different events, two transformations have been applied to the waveforms. Firstly, to line up the pulses temporally, each pulse has been shifted to give a value of zero to the RQ "aft_t05_samples," which is the sample number at which only the first 5% of the S1 pulse's integrated area has been measured. Secondly, each waveform has been normalized, so that differences in pulse shape are not obscured by the 10phd variation in overall pulse size among the pulses in each plot.

[6]The "delensed" X and Y coordinates are the true coordinates of the event, after subtracting out the effects of horizontal electron drift due to the detector's nonuniform electric field in Run 4, discussed in Sec. 2.6.



(a)



(b)

Figure 4.5: Aligned and normalized S1 pulses for calibration events passing all cuts. Only events from the second-lowest DD gun depth, centered 30.0cm below the liquid xenon surface, are shown. Heavy lines are the averages with error bars indicating 1σ spread in the underlying populations. ER and NR are denoted by blue and red respectively, as labeled. (a) 75 ER waveforms and an equal number of NR waveforms in the energy range 40-50phd (b) 75 ER waveforms and an equal number of NR waveforms in the energy range 70-80phd

In addition to the raw samples, shown as thin, light lines, the average ER and NR pulse are shown as heavy lines. The associated error bars give the 1σ width of the underlying pulse populations. This should be distinguished from the actual uncertainly in the averages, which is smaller by a factor $\frac{1}{\sqrt{n}}$, where n is the number of waveforms used.

A careful inspection of the sample waveforms reveals what the averages make more explicit: For most of the rise time of the pulse, the normalized waveforms with the highest values are more likely to be NRs, and the normalized waveforms with the lowest values are more likely to be ERs. Over most of the pulse's falloff, the opposite is true.

The simplest way to capture this difference in shape is to use as the figure of merit a quantity called the prompt fraction. The prompt fraction is defined as the integrated area under the first part of the pulse, divided by the integrated area of the entire pulse. It is literally the fraction of light that is detected promptly. For the prompt fraction's numerator, the datapoints from 2 samples before aft_t05_samples to 4 samples after aft_t05_samples, inclusive, are summed together. For the denominator, all the samples making up the S1 waveform are used.[7] The region integrated to get the prompt fraction's numerator is shaded in light gray in Fig. 4.5. The region is chosen to extend from just before the pulse begins rising to the approximate time when the average normalized NR curve falls below the average normalized ER curve. As a result, NR events would be expected to have a larger mean prompt fraction than would ER events at the same depth with the same S1 size.

4.2.4 Model Building

The next step is developing an empirical model of the distribution of prompt fraction values as a function of event type, depth, and S1 size. This process starts by segmenting the calibration data into a variety of depth and S1 size bins, to study the distribution of ER prompt fractions and the distribution of NR prompt fractions in each bin. For depth, the five bins are bounded by drift times of 40, 80, 130, 180, 230, & 290 μs, chosen so that each interval brackets one of the multi-Z DD gun depths. For S1 size, the twelve bins are bounded by S1 sizes of 0, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, & 100 phd. This covers the full S1 range of the calibration data, with especially finely-spaced bins below 20phd where prompt fraction shows the strongest variation with respect to S1 size. The task of determining the prompt fraction statistical distribution must therefore be completed 120 times: for each of the $5 \times 12 = 60$ combinations of depth and S1 size, two distributions (one for ER and one for NR) must be characterized.

Fig. 4.6 shows a histogram of the calibration data used for each of these cases. Each column represents an S1 size interval, so the first column is data from 0-5phd, the second column is 5-10phd, and so on to the last column with data from 90-100phd. Each row represents a different population and depth interval. The first row shows all ER calibration data from all depths lumped together, and the subsequent five rows show ER data for

[7] In the terminology of the LUX data processing chain, samples ranging from RQ pulse_start_samples to RQ pulse_end_samples for the S1 pulse are used for the denominator.

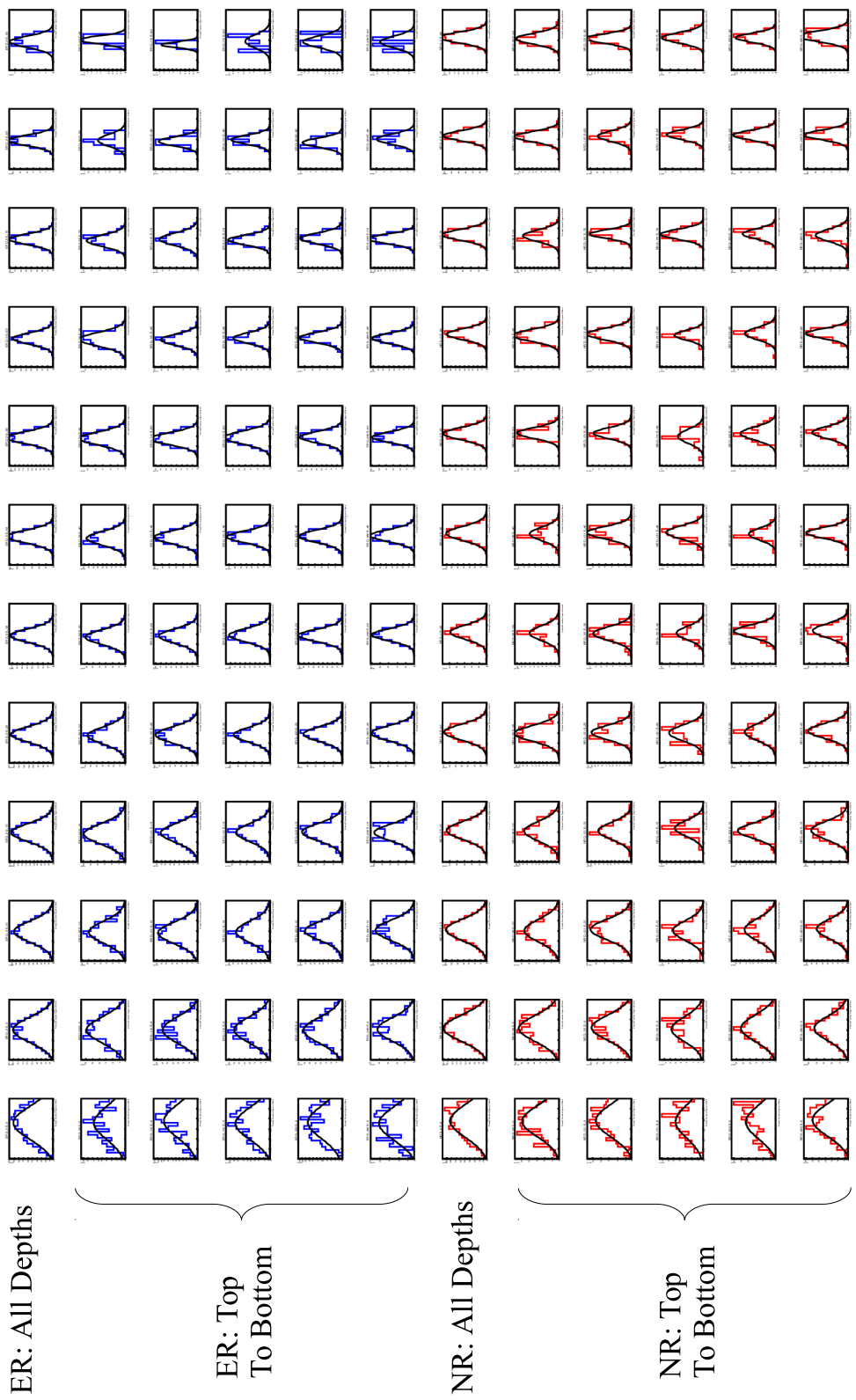


Figure 4.6: Gaussian fit to prompt fraction distribution for each event type, depth interval, and S1 size interval. See description in text.

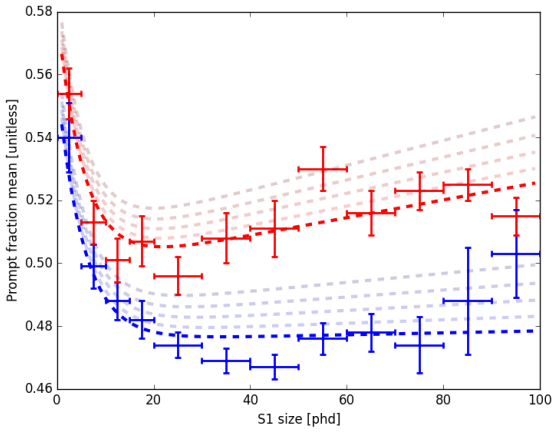
each depth bin, starting with the shallowest depth interval (40-80μs) and working down to the deepest one (230-290μs). The subsequent six rows give the same information but for nuclear recoils. The histograms of all depths combined (in the 1st and 7th rows) are included for comparison but are not used further in the analysis. The plots in the remaining rows correspond to the 120 cases tabulated in the previous paragraph. Although the small size of each plot precludes legibly marking their axes, the plots all have matching axes: The X-axis, prompt fraction, runs from 0.2 to 1 for each plot. The Y-axis is number of events, scaled so that the height of each plot's most populous histogram bin equals the height of the plot.

For each case, the histogram's shape is roughly gaussian. That is expected, because the prompt fraction numerator is the sum of the signal from a number of stochastic interactions. Therefore, a gaussian distribution is fit to the data in each case by calculating the prompt fraction mean and standard deviation. The resulting fitted gaussian distributions are overlaid on the raw data in Fig. 4.6.

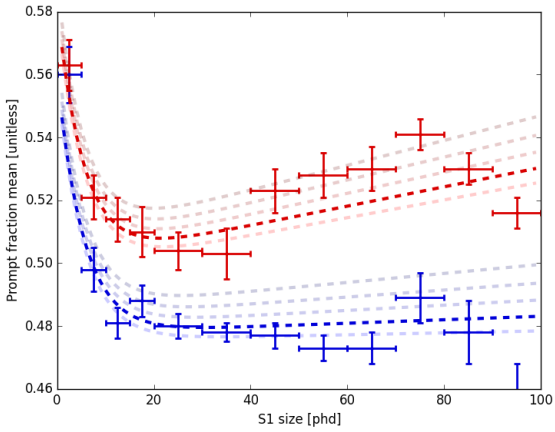
Figs. 4.7, 4.8, and 4.9 show the results of these computations in a different way. Figs. 4.7 and 4.8 show the calculated mean in each of the 120 cases analyzed. Each figure corresponds to a depth bin and the horizontal axis in each figure is S1 size. The 1σ statistical uncertainty of each mean, shown in the figures as vertical error bars, is found by dividing the calculated standard deviation by the square root of the number of samples. Fig. 4.9 shows the calculated standard deviation of each of the 120 cases analyzed, with S1 size again on the horizontal axis but the points from all depths overlaid on a single plot. The error bar for each point on this plot shows the approximate 1σ statistical uncertainty of that standard deviation. For gaussian distributions, this is approximated in the large n limit[11] as $\frac{\sigma}{\sqrt{2n}}$.

Looking at the the means and standard deviations of the distributions of prompt fraction for the various depths, S1 sizes, and calibration data populations, a number of overall trends are evident despite the noisiness of the data. To start, the mean prompt fraction does not vary too much from case to case – in only a handful of instances does the mean vary from 0.5 by more than 10% in either direction. However, the small variations in mean prompt fraction are not haphazard. For every depth/S1-size combination, the mean ER prompt fraction is lower than the mean NR prompt fraction. Mean prompt fraction varies with S1 size, rapidly falling for S1 sizes less than $\sim$20-30phd and gradually rising thereafter. Mean prompt fraction also varies with depth, with greater drift times corresponding to higher prompt fractions. As for the standard deviation of the prompt fraction distribution, this falls off with increasing S1 size but shows no evident dependence on depth or on the difference between ERs and NRs.

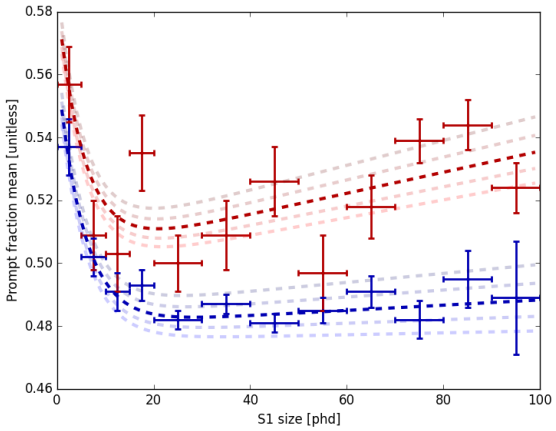
Rather than calculating prompt fraction distributions from first principles, we seek an empirical model that incorporates the properties described in the previous paragraph. The empirical model should take the form of formulas for prompt fraction mean and standard deviation as a function of event type, depth, and S1 size. In the discussion below, let T be a variable that is equal to 0 for electron recoils and equal to 1 for nuclear recoils. Let D be an event's drift time expressed in μs, and let x be its S1 size expressed in phd (detected photoelectrons). Guided by the observed shape of the data, suitable functional forms for



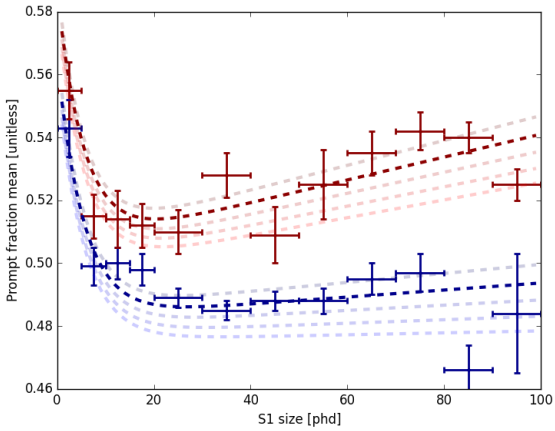
(a) Depth bin 1: 40-80μs drift time



(b) Depth bin 2: 80-130μs drift time

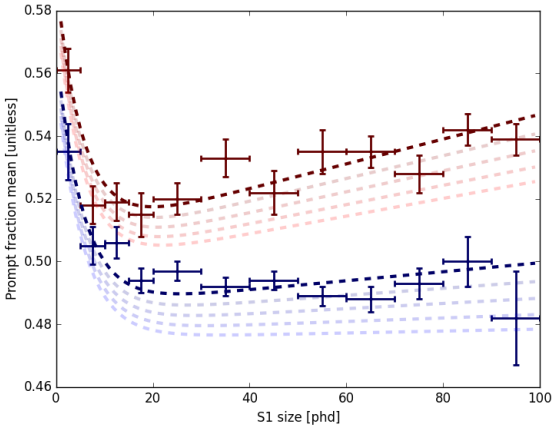


(c) Depth bin 3: 130-180μs drift time



(d) Depth bin 4: 180-230μs drift time

Figure 4.7: Prompt fraction for each combination of depth bin, S1 size bin, and event type (ER or NR). Each figure corresponds to a given depth bin as labeled. In each figure, the lower (blue) datapoints indicate ERs, and the higher (red) datapoints indicate NRs. Each horizontal error bar is half the energy bin width, and each vertical error bar is the 1σ statistical uncertainty in the vertical location of the data point. The dotted lines are the empirical model, shown in dark shades for the depth bin in question and light shades for all other depth bins.



(a) Depth bin 5: 230-290μs drift time

Figure 4.8: Fig. 4.7 continued.

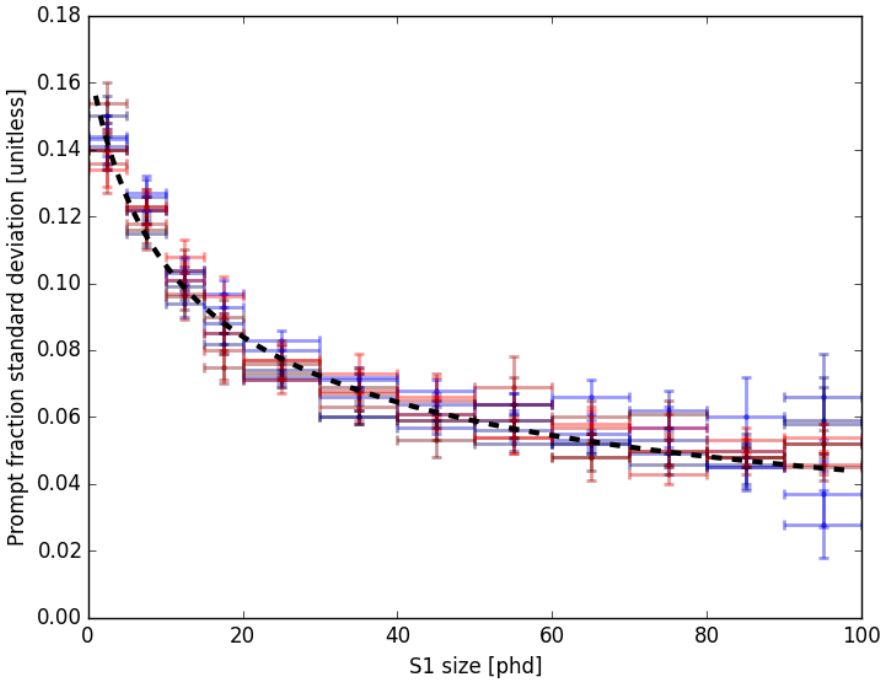


Figure 4.9: Prompt fraction standard deviation versus S1 size. Datapoints for all depth bins and populations (ER or NR) are shown, and the vertical error bars are the 1σ statistical uncertainty in the standard deviation. The dotted line is the empirical model.

Parameter	Value
a_1	$(4.73 \pm .02) \times 10^{-1}$
a_2	$(8.8 \pm 1.9) \times 10^{1}$
a_3	$(2.5 \pm .3) \times 10^{-4}$
a_4	$(5.5 \pm .9) \times 10^{-7}$
a_5	$(8.3 \pm .7) \times 10^{-2}$
a_6	$(1.8 \pm .2) \times 10^{-1}$
a_7	$(1.68 \pm .05) \times 10^{-1}$
a_8	$(5.8 \pm 1.2) \times 10^{0}$
a_9	$(4.6 \pm .3) \times 10^{-1}$

Table 4.1: Free parameters in the empirical model of S1 prompt fraction in LUX.

modeling the mean and standard deviation have been found. They are, respectively,

$$\mu_P = a_1 + (a_3 T + a_4 D)(x + a_2) + a_5 e^{-a_6 x}\,, \tag{4.1}$$

$$\sigma_P = a_7 \left(\frac{x}{a_8} + 1\right)^{-a_9}\,. \tag{4.2}$$

The expression for the mean is linear in the T and D variables, with an added nonlinear term to model the sharp change in prompt fraction at the lowest S1 sizes. The expression for the standard deviation is almost a power law decay, with a small added offset. The nine unitless free parameters are fit using a nonlinear least-squares optimization, with the results shown in Table 4.1. The resulting model is shown as dotted lines in Figs. 4.7, 4.8, and 4.9.

Finally, we consider whether this model is an acceptable fit to the data. Both the empirical model of mean prompt fraction, μ_P, and the empirical model of prompt fraction standard deviation, σ_P, are fit using 120 datapoints. Since μ_P has six free parameters, that leaves 114 degrees of freedom (DOF). With only three free parameters, σ_P has 117 degrees of freedom. Using the statistical uncertainties for the prompt fraction mean and standard deviation in each case as described above, it is found that the χ^2/DOF for the means is 1.62 and the χ^2/DOF for the standard deviations is 2.17. The high values of these statistics indicate that the empirical model does not fully capture the underlying variation in these values.[8] However, this is not surprising, because the model was developed ad hoc as an approximation, and there is no physical motivation for why it should fully capture the subtleties of the problem. The more relevant question is whether any systematic difference between the model and the data is large compared to the separation between the ER and NR means, and an examination of Figs. 4.7, 4.8, and 4.9 shows this not to be the case.

[8] In the large n limit, the reduced chi-squared distribution becomes concentrated around unity.

4.2.5 Model Implications

By using prompt fraction to characterize pulse shape and bringing to bear the model of prompt fraction distribution developed in the previous section, answers to a number of questions can be found: How effective is pulse shape discrimination at distinguishing electron recoils from nuclear recoils? Are the LUX WIMP-search data's pulse shapes consistent with a largely-ER background? Is ER/NR discrimination based on pulse shapes independent of discrimination based on $\log_{10}$(S2) vs. S1? These questions are addressed in this section.

Leakage Fraction

A simple metric for how well a figure of merit can distinguish two populations is the leakage fraction, the percentage of events in the first population with figure of merit values more extreme than the median of the second population. In the specific case under consideration, the leakage fraction is defined as the percentage of ER events with prompt fraction values smaller than the median NR prompt fraction. A leakage fraction of 0.5 (that is, 50%) is not inconsistent with having no discriminating power at all. After all, if the ER and NR prompt fractions had exactly the same distribution as each other, then 50% of the ER events would necessarily fall below their common median. On the other hand, a leakage fraction of 0 implies highly effective separation, with at least half the NR events characterized by prompt fractions smaller than even the smallest ER prompt fraction. The expected leakage can be calculated from the empirical model.[9] The result depends on S1 size, but it turns out to be depth-independent for reasons that will be further discussed below. This result is shown in Fig. 4.10.

This calculation shows that the efficacy of pulse shape discrimination is strongly dependent on S1 size. For low-energy recoils (which produce small S1 pulses), the leakage fraction gets close to 50%, meaning that there is almost no discrimination power. On the other hand, the leakage fraction drops with increasing S1 size, reaching a minimum value in the vicinity of 15% for the largest S1 pulses that could be meaningfully analyzed with LUX's tritium and DD calibration data.

Reduced Prompt Fraction

To see how actual Run 4 WIMP search data compares with this study of calibration data, we can take advantage of a convenient mathematical property of the model function of mean prompt fraction, μ_P. This function can be expressed as a sum of two terms, of which one depends on depth and S1 size but not event type, and the other depends on event type and S1 size but not depth. By subtracting away just the former, we arrive at what could be called a "reduced prompt fraction" – a depth-independent measure of pulse shape in LUX.

[9] In the calculation, the NR mean is used in place of the median, which is valid because the distribution is assumed to be gaussian.

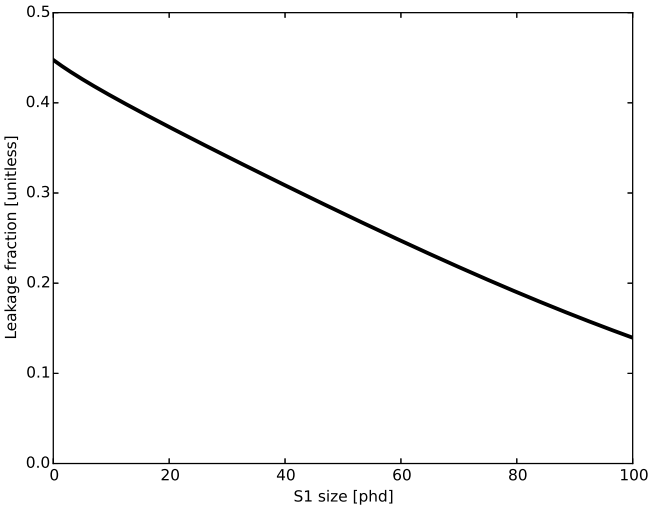


Figure 4.10: Leakage fraction versus S1 size.

Specifically, we define reduced prompt fraction f_r to be

$$f_r = f_p - \left(a_1 + (a_4 D)(x + a_2) + a_5 e^{-a_6 x}\right) \tag{4.3}$$

where f_p is the prompt fraction and the other variables are as defined in Sec. 4.2.4. The empirically modeled mean reduced prompt fraction is then:

$$\mu_R = (a_3 T)(x + a_2) \tag{4.4}$$

and the modeled standard deviation of reduced prompt fraction is the same as for prompt fraction itself, because just adding an offset does not change the distribution's spread. Conveniently, the standard deviation model is also depth-independent. Because T is zero for electron recoils and 1 for nuclear recoils, the mean reduced prompt fraction for ERs is simply zero, while the mean for NRs is a sloped line.

By using Eqn. 4.3, one can compute the reduced prompt fraction for any event, without knowing whether it's an ER or an NR. Using this equation amounts to subtracting away the model's prediction for the prompt fraction of an electron recoil at a given event's S1 size & depth in the detector, and then seeing what's left. The probability distributions for reduced prompt fraction still depend on S1 and whether the event is an ER or an NR, but the dependence on drift time has been entirely removed. As an additional advantage, Eqn. 4.4 is easier to work with computationally than Eqn. 4.1.

Because of this depth-independence, a reduced prompt fraction plot including events from different depths is still a fair "apples-to-apples" comparison. Such a plot is shown in Fig. 4.11. The two broad shaded bands, which have significant overlap, are the $\pm 2\sigma$ range for ERs and NRs. The points are Run 4 WIMP search events passing all analysis cuts, including a 50phd cut on S1 size.

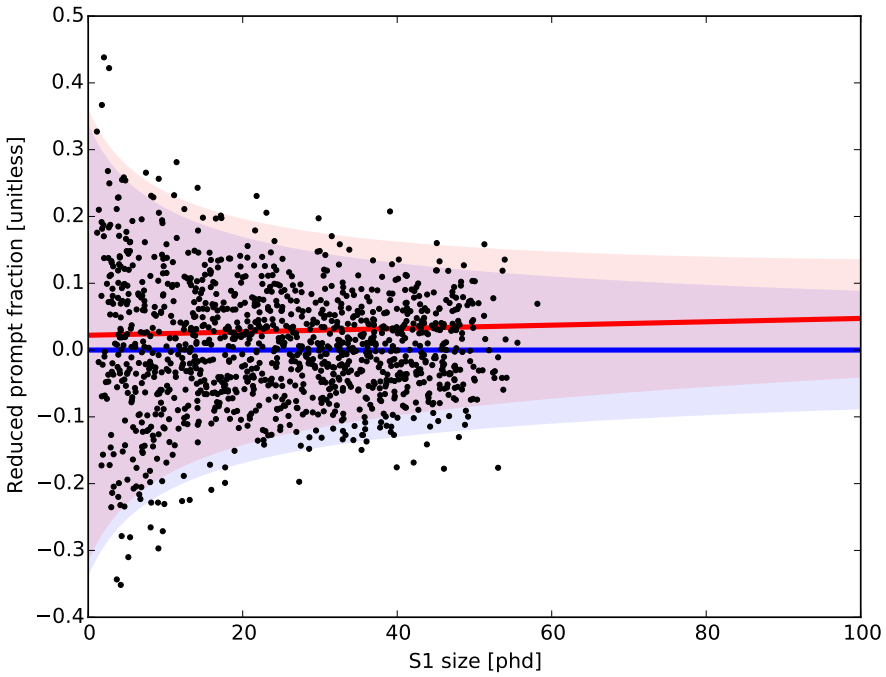


Figure 4.11: Reduced prompt fraction (prompt fraction with constant and depth-dependent terms subtracted away) vs S1 size. Of the two partially-overlapping shaded bands, the upper one (in red) is the modeled 2-sigma band for nuclear recoils and the lower one (in blue) is the modeled 2-sigma band for electron recoils. The upper (red) and lower (blue) heavy lines are the modeled means, and the points are Run 4 WIMP search events passing all analysis cuts.

Up till now, the analysis in this chapter has worked exclusively with calibration data. Fig. 4.11 represents the first comparison to WIMP search data.[10] It is expected, from both background studies and S2/S1 measurements, that virtually all these events should be ERs. This result is consistent with that expectation. Take the ribbon of parameter space falling within 2σ of the ER mean but not within 2σ of the NR mean, and compare it with the corresponding region within the 2σ NR band but not the ER band. The former region has more events than the latter, as would be expected for an ER population. The average reduced prompt fraction in each S1 size bin is also closer to the ER mean than to the NR mean.

[10] Although this pulse shape discrimination study is not, formally speaking, a blind analysis, this juncture represents the first application to WIMP search data of a procedure developed without using that same data.

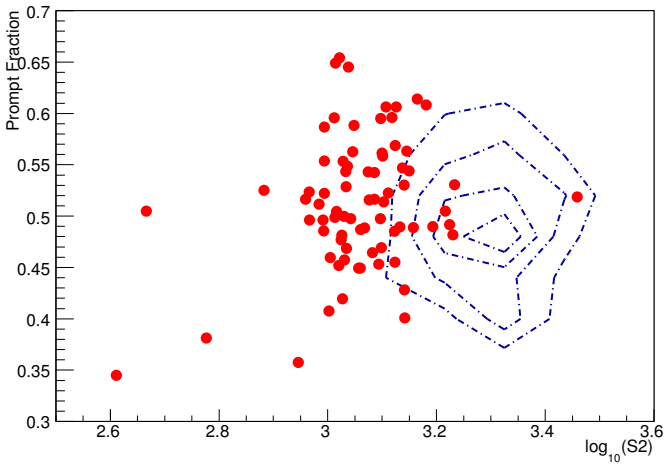


Figure 4.12: Prompt fraction vs. $\log_{10}$(S2) for a representative sample of calibration events, specifically those with 40phd < S1 size < 50phd and 180μs < drift time < 230μs. Contours indicate electron recoils and points are nuclear recoils.

Independence of Discrimination Methods

We now have two methods for distinguishing ERs and NRs. The first method is looking at $\log_{10}$(S2) vs. S1, as has been extensively used in LUX.[11] The second method is using pulse shape discrimination, as has been developed in this chapter. The former method provides a strong discrimination capability, while the latter furnishes a weaker but still meaningful capability. If the methods are to be used in concert, an important question is whether they generate results that are statistically independent of each other. If, for example, prompt fraction were perfectly correlated with $\log_{10}$(S2) for a given S1 size and depth, then using the two methods together would provide no more information than just using $\log_{10}$(S2) alone.

Fig. 4.12 compares the two methods in the case of a specific depth bin and S1 size bin. Compared to the nuclear recoils, the electron recoil population is well to the left, because of the difference in $\log_{10}$(S2), and it is slightly higher, because of the difference in prompt fraction. Most importantly, the events in each population form a structureless blob, with no apparent correlation or relationship between the x-axis value and y-axis value for individual points in each population. To quantify this observation, the coefficient of determination for the nuclear recoils shown in Fig. 4.12 is only 0.07 (and even that is unduly increased by outliers), while the coefficient for electron recoils is a negligibly small 7.3×10^{-5}. Although an arbitrary depth/S1 size combination was chosen for the figure and discussion, other depth/S1 size cases show similar behavior. From this it is concluded that pulse shape discrimination provides a new tool for distinguishing ERs and NRs in LUX that is statistically independent

[11] LUX Run 3 used $\log_{10}$(S2/S1) vs. S1 while LUX Run 4 used $\log_{10}$(S2) vs. S1, but this change of variables does not substantially alter the analysis.

of the previously-used method.

4.3 Other PSD Figures of Merit

Throughout the above analysis, prompt fraction has been used as the figure of merit for characterizing the shape of an S1 pulse. However, there are other methods for reducing an S1 signal to a single number describing its shape. In this section, two of these will be considered. The first involves looking at each PMT's waveform separately, to fit the arrival time of each individual photon. The second involves using machine learning techniques to classify pulse shapes. Although it will be seen that neither method substantially outperforms the simple prompt fraction, both involve the exploration of interesting data analysis techniques.

4.3.1 Individual Photon Fitting

A study was undertaken to characterize S1 pulse shapes by determining individual photon arrival times[18].[12] The first step to this technique is determining the typical response waveform in a PMT due to a single photon. Such a typical waveform, or "template," can be found by averaging many waveforms with integrated areas in the range 0.5-1.5phd, a range dominated by single-photon waveforms. Then, for each event to be studied, the waveform in each PMT is considered separately (instead of these waveforms being summed, as in Sec. 4.2). The given waveform is fit using five different models, each corresponding to the sum of a different number of template functions ranging from one to five. The best fit is then determined using Bayes' theorem, with priors generated from the waveform's integrated area. Fig. 4.13 shows an example of applying this technique to a simulated waveform. The method is notably precise, reconstructing photon arrival times with a 1σ uncertainty of 1.6ns, even though the LUX electronics only digitize the output signal every 10ns.

At that level of timing resolution, a number of small systematic timing offsets need to be corrected. For each fitted photon, the arrival time is shifted by the transit time for light to travel from the reconstructed photon position to the PMT assuming a direct path without reflection. Because the PMTs are set at different voltages, the average transit time for the signal to propagate through the PMT is different for each PMT, with differences of up to 9ns between channels. Even the variable lengths of the cables connecting the PMTs to the DAQ electronics can contribute an offset. To address these latter two effects, measurements made with UV calibration LEDs in the LUX detector are used to determine the channel-by-channel timing offset that remains after the xenon propagation time is subtracted away.

To turn the timing information into a figure of merit for pulse shape discrimination, a prompt fraction similar to that described in Sec. 4.2.3 is used. Time ranges for the numerator and denominator are found by an optimization procedure, with the goal of finding values that minimize the average leakage fraction over a distribution of events that is uniform in S1

[12]The work described in Section 4.3.1 was carried out by others in the LUX collaboration.

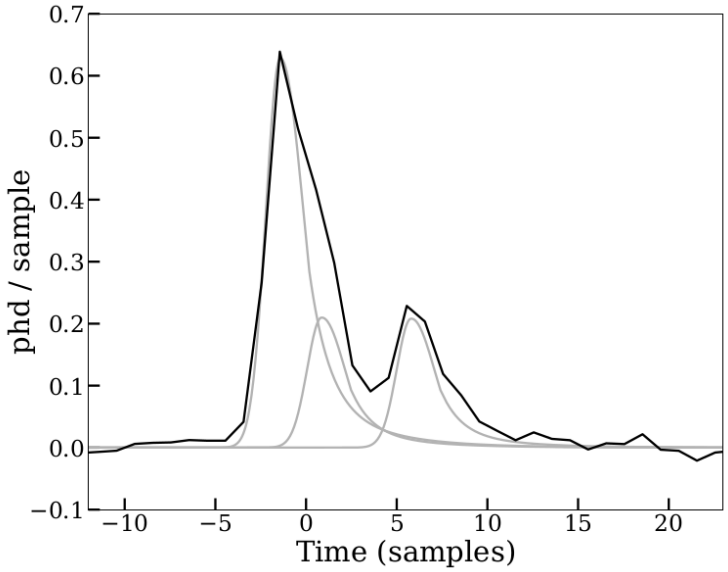


Figure 4.13: Example of fitting individual photon arrival times. The simulated signal in one PMT (black line) is due to three individual photons, two of which partially overlap in time, as determined by the fitting routine and shown with gray lines. Reprinted from [18].

size.[13] The values of the numerator and denominator are then essentially the total number of photons with arrival times falling into their respective windows. The sum of photons in each interval is weighted by the size of each photon peak as found by the template function fit. The weighting naturally takes into account complications such as the possibility of two photons arriving simultaneously. Fig. 4.14 shows the leakage fraction as a function of S1 size. It is comparable to the leakage fraction found by the empirical method, being only slightly stronger over most of the 0-100phd range.

The individual photon fitting method makes possible a novel measurement. By combining the photon times for many pulses of similar depth and S1 size, a continuous average pulse shape for ER, and one for NR, can be developed (contrast this with the average pulse shapes in Fig. 4.5, which are discretized to 10ns steps). This can then be fit using a theoretical model of pulse shape to determine various free parameters in the model. Such a theoretical model is developed by convolving several terms: First is a sum of two exponential decays due to the different decay times of singlet and triplet states. Second is a temporal spreading from geometric considerations, as some photons travel directly to PMTs while others bounce off the PTFE walls one or more times before being detected. Third is variation and uncertainty in PMT transit time and related instrumental effects. By fitting the combined model, for both ERs and NRs and across a range of S1 sizes, a number of parameters can be measured at once. These include a first measurement of the singlet-to-triplet ratio for low-energy nuclear

[13]The optimum is found to be a numerator from 8ns before the RQ aft_t05_samples to 32ns after it, and a denominator from 14ns before the same RQ to 134ns after it.

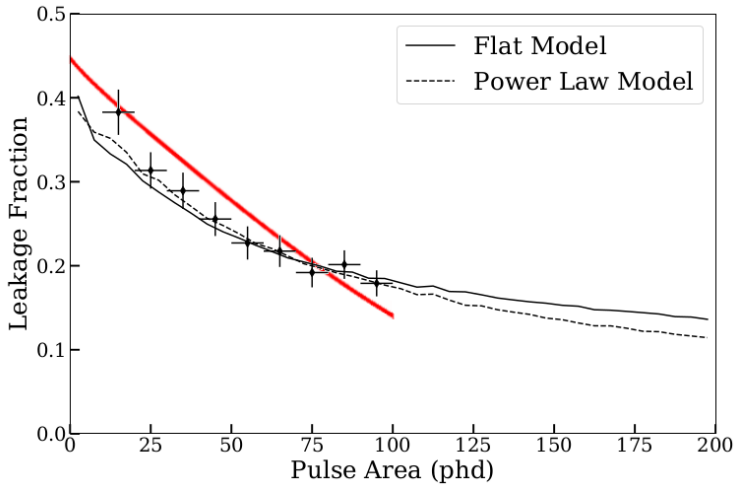


Figure 4.14: Leakage fraction versus S1 size for individual photon fitting method. Points with error bars are from data; thin black lines are from simulations. For comparison, the thick red line is the leakage fraction found by the empirical method as shown in Fig. 4.10. Reprinted from [18], with one added curve.

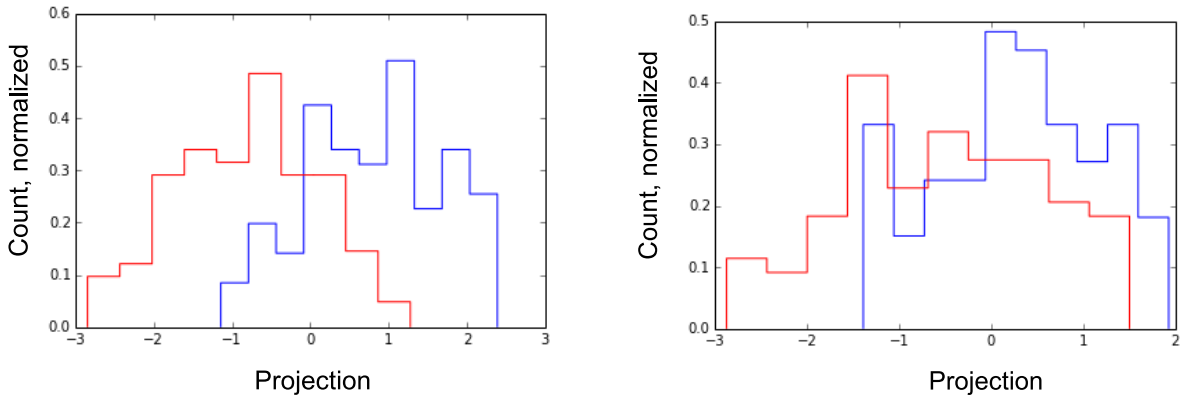
recoils. This measurement, as well as the individual photon fitting technique more broadly, is described in Ref. [18].

4.3.2 Machine Learning

A number of machine learning and statistical techniques warrant investigation as possible ways to improve pulse shape discrimination. One of these methods is linear discriminant analysis (LDA). In LDA, points in an n-dimensional space are classified into one of two classes.[14] This is done by using training data, points for which the class is known in advance, to compute a vector in the n-dimensional space such that projecting points onto the vector will produce a figure of merit that maximally separates the two classes. Once this optimal vector is "learned" from the training data, it can be put to use classifying events for which the class isn't necessarily known.

An LDA approach to classifying events as either ER or NR based on S1 pulse shape was investigated using Run 3 data. To do so, pulses are represented as 18-dimensional vectors, where each vector component is an integer equal to the number of photons detected in a corresponding 10ns time interval. This number is taken to be equal to the number of PMT channels crossing from below to above a 1.5mV threshold during the interval. Fig. 4.15 shows the result of this process, for the case of events in Run 3 calibration data with similar S1 size (70-80phe) and similar depth (90-125μs). Fig. 4.15a shows the result of LDA

[14]The technique can be generalized to a higher number of classes in so-called "multiclass LDA," but that is not needed here.



(a) (b)

Figure 4.15: Classifying pulse shapes with LDA. (a) Figure of merit histograms for the training data, comprised of 100 events of each class. The histogram more to the left (in red) is NRs, and the one more to the right (in blue) is ERs. (b) Analogous histograms from applying the learned vector to validation data, consisting of 100 different events from each class.

reduction of the training data itself, and Fig. 4.15b shows the result of applying the learned method to a different subset of the calibration data. While the former figure is expected to show good separation, because it uses the exact data the machine was trained with, the fact that the latter figure also shows acceptable separation indicates that the classification is working. However, this approach, and a similar analysis with a support vector machine (SVM) technique, does not show noticeable improvement in classification as compared to prompt fraction.

A subsequent analysis with Run 4 data looked at ways to strengthen the discriminating power of the individual photon fitting method. One possible area for improvement relates to the use of the aft_t05_samples RQ to temporally align various events' S1 pulses. In contrast with the high precision of fitted photon arrival times, the RQ itself only has a 10ns resolution. To more accurately estimate the time at which 5% of the integrated area of the pulse has passed by for a given event, a high-resolution waveform is constructed from the fitted times and their corresponding fitted areas using a weighted kernel density estimation (weighted KDE) method with an Epanechnikov kernel. Waveforms are then temporally aligned using the point at which the kernel density estimator function's cumulative distribution function (CDF) equals 0.05. Fig. 4.16 shows an example of this process, which results in an alignment time that is not constrained to 10ns steps.

Another question is whether a figure of merit other than prompt fraction could more fully take advantage of high-resolution timing information. Prompt fraction, after all, uses timing only to decide whether or not to include a pulse in the numerator (and, less importantly, the denominator), gaining nothing from improved timing information for pulses that are not near the cutoff points for these decisions. A possible alternative, which uses all timing information

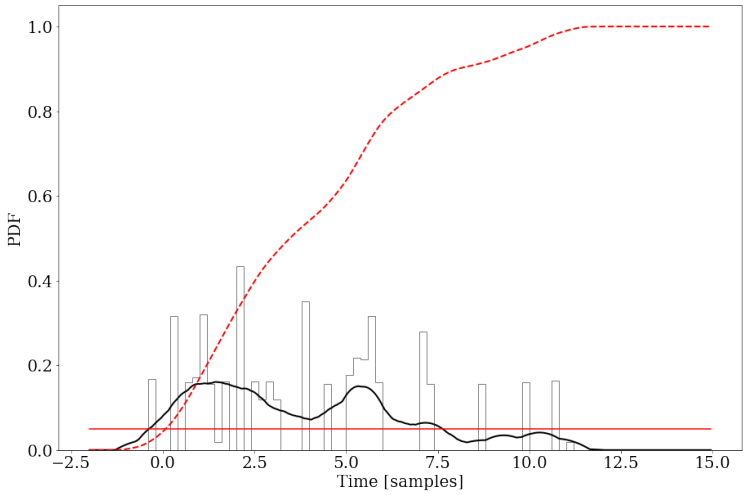


Figure 4.16: Example of fitting the waveform 5% percentile to $<$ 10ns precision. The histogram shows the photon arrival times, weighted by pulse area. The heavy solid black line is the KDE PDF and the heavy dashed red line is the KDE CDF. The time where the latter intersects the horizontal line of height 0.05 is used for alignment.

in a nontrivial way, is calculating the Kolmogorov-Smirnov (KS) statistic. In this technique, many waveforms are combined to produce an average CDF for electron recoils and an average CDF for nuclear recoils. For each event, an empirical distribution function is defined using the fitted photon arrival times.[15] The event's KS statistic for each hypothesis (ER and NR) is then taken to be the maximum vertical distance between the empirical distribution function and the corresponding average CDF. The event is more likely to belong to the class with the smaller KS statistic, so the difference of the statistic in the two cases can serve as a figure of merit. Fig. 4.17 shows an example of comparing the relevant CDFs.

Surprisingly, the pulse shape discrimination achievable with these techniques, either separately or together, is not materially different from that of the simple prompt fraction approach described in Sec. 4.2. The competitiveness of the standard technique compared to the novel methods of Sec. 4.3 suggests that the limit of pulse shape information that can be extracted in LUX may have been reached.

In the next chapter, we will momentarily put aside pulse shape discrimination and turn our attention to an analysis of limits that can be set on a new class of theoretical dark matter models. PSD will make only a brief appearance, in a discussion of generalizing LUX's limit-setting techniques. However, we will return to the topic of pulse shape discrimination in the concluding chapter, to consider the role it can play in an extension of that same analysis to higher energies.

[15] An empirical distribution function is a CDF built from a set of datapoints. It has a step at the location of each such point.

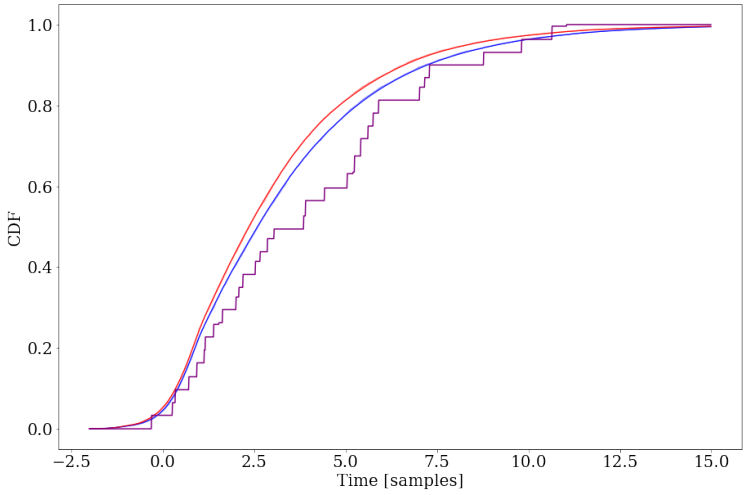


Figure 4.17: Calculating a Kolmogorov-Smirnov statistic. The upper (red) smooth line is the CDF for NRs, the lower (blue) smooth line is the CDF for ERs, and the step function is the empirical distribution function of the same example event shown in Fig. 4.16. The fluctuation of individual events is substantial compared to the small difference between the ER and NR averages; the event shown is actually an NR even though its time signature more closely resembles a typical ER.

Chapter 5

Inelastic Effective Field Theory

In the last five years, the field of dark matter direct detection has seen the widespread adoption of dark matter effective field theory (EFT). In this approach, limits are set on the coupling constants of a generalized theory of WIMP/nucleon interactions, superseding the previous approach of setting limits on just one or two specific interaction models. The new approach can be pushed even further with the development of inelastic effective field theory (IEFT), a theory which relaxes one of the assumptions of the previous version of EFT to create an even more expansive range of possible dark matter interactions with nuclei. This chapter will discuss IEFT and the analysis steps needed to set IEFT limits with LUX. It will conclude with limits on representative models in the IEFT parameter space, including some models that have been studied before and others that are given experimental limits for the first time here.

5.1 EFT Physics

This discussion of effective field theory builds on the discussion in Sec. 1.3.5. For clarity, we will refer to the original EFT theory as "elastic EFT" (eEFT), in contrast with the more generalized inelastic EFT, or IEFT.[1] While the former theory describes only elastic scattering, the latter allows for inelastic scattering where the outgoing WIMP's rest mass differs from the incoming WIMP's rest mass[2] by some mass difference δ. IEFT is introduced in Ref. [43], where it is called "model independent" inelastic dark matter. eEFT has been previously studied with LUX[97].

The physics of inelastic EFT parallels the physics of elastic EFT. In both cases, the non-relativistic interaction Lagrangian is assembled from various combinations of four Galilean-invariant terms: $\mathbf{S}_N$, $\mathbf{S}_\chi$, $i\mathbf{q}$, and/or $\mathbf{v}^\perp$. The first two of these are the respective spins of

[1] "EFT," without a modifying adjective, refers collectively to both eEFT and IEFT.

[2] We will not consider the complementary possibility of the nucleus being excited, an idea discussed in Ref. [44].

the nucleon[3] and the WIMP, while the third is proportional to the momentum transfer $\mathbf{q}$. The fourth term is equal to the component of the particles' relative velocity perpendicular to the direction of momentum transfer. In the inelastic case, that is

$$\mathbf{v}^{\perp} = \mathbf{v} + \frac{\mathbf{q}}{2\mu_N} + \frac{\delta}{|\mathbf{q}|^2}\mathbf{q} \tag{5.1}$$

where μ_N is the reduced mass of the nucleon and WIMP, and v is their relative velocity. For the elastic case, the expression is the same but with δ set to zero.

The full interactive Lagrangian is

$$\mathcal{L} = \sum_{N=p,n} \sum_{i} c_i^{(N)} \mathcal{O}_i \chi^+ \chi^- N^+ N^- \tag{5.2}$$

where the χ terms refer to the outgoing and incoming WIMPs and N terms refer to the outgoing and incoming nucleon. The list of operators $\mathcal{O}_i$ is shown in Table 5.1. Since each interaction can couple with different strengths to protons and neutrons, the expression includes a sum over these two nucleon types in addition to the sum over the operators. As a technical point, operator $\mathcal{O}_2$ is usually disregarded because it doesn't occur as a leading-order term in realistic theories,[4][32] so there are really only fourteen operators to consider. A specific dark matter model within the IEFT framework is fully specified by its 28 coupling constants $c_i^{(N)}$ along with the dark matter initial mass m_χ and mass splitting δ.

Of these operators, $\mathcal{O}_1$ and $\mathcal{O}_4$ have nonzero recoil spectra in the $v \to 0$ limit if $\delta = 0$, while the other operators have recoil spectra that always go to zero in the zero velocity limit.[5] In fact, $\mathcal{O}_1$ by itself is just spin-independent (SI) dark matter, while $\mathcal{O}_4$ by itself is the usual spin-dependent (SD) dark matter. Typical SD limit papers (such as Ref. [17]) place separate limits on WIMP-proton scattering and WIMP-neutron scattering, which in the EFT framework just means setting separate limits on c_4^p and c_4^n. Typical SI limit papers (such as Ref. [20]) assume equal coupling to protons and neutrons, which in the language of EFT means assuming that $c_1^p = c_1^n$. Such SI and SD limit papers ignore all operators aside from these two, as well as the possibility of nonzero δ.

One consideration with EFTs is the overall normalization, or scaling, of the operators. In the development of an effective field theory, scaling an operator by an overall multiplicative constant, even one that is not unitless, does not have much physical significance. Since each term includes a multiplicative coupling constant, any change in the magnitude of the operator can be simply absorbed by a corresponding change in the coupling constant.[6] An

[3] In this chapter, we use subscript N to refer to a single nucleon, following a convention of much of the EFT literature. This is a slight change from the EFT discussion in Sec. 1.3.5, where we instead used subscript n for consistency with other equations in the surrounding chapter.

[4] Anand et al.[32] observe that operator $\mathcal{O}_2$ "cannot be obtained from the leading-order nonrelativistic reduction of a manifestly relativistic operator."

[5] Note that $v \to 0$ implies $q \to 0$. $\mathcal{O}_1$ shows scalar coupling, while $\mathcal{O}_4$ is spin-coupled.

[6] Of course, these normalizations must be explicitly specified in order for coupling constant theories and limits to be well-defined.

$$\mathcal{O}_1 = \mathbf{1}$$
$$\mathcal{O}_2 = \left(\mathbf{v}^\perp\right)^2$$
$$\mathcal{O}_3 = i\mathbf{S}_N \cdot \left(\frac{\mathbf{q}}{m_N} \times \mathbf{v}^\perp\right)$$
$$\mathcal{O}_4 = \mathbf{S}_\chi \cdot \mathbf{S}_N$$
$$\mathcal{O}_5 = i\mathbf{S}_\chi \cdot \left(\frac{\mathbf{q}}{m_N} \times \mathbf{v}^\perp\right)$$
$$\mathcal{O}_6 = \left(\mathbf{S}_\chi \cdot \frac{\mathbf{q}}{m_N}\right)\left(\mathbf{S}_N \cdot \frac{\mathbf{q}}{m_N}\right)$$
$$\mathcal{O}_7 = \mathbf{S}_N \cdot \mathbf{v}^\perp$$
$$\mathcal{O}_8 = \mathbf{S}_\chi \cdot \mathbf{v}^\perp$$
$$\mathcal{O}_9 = i\mathbf{S}_\chi \cdot \left(\mathbf{S}_N \times \frac{\mathbf{q}}{m_N}\right)$$
$$\mathcal{O}_{10} = i\mathbf{S}_N \cdot \frac{\mathbf{q}}{m_N}$$
$$\mathcal{O}_{11} = i\mathbf{S}_\chi \cdot \frac{\mathbf{q}}{m_N}$$
$$\mathcal{O}_{12} = \mathbf{S}_\chi \cdot \left(\mathbf{S}_N \times \mathbf{v}^\perp\right)$$
$$\mathcal{O}_{13} = i\left(\mathbf{S}_\chi \cdot \mathbf{v}^\perp\right)\left(\mathbf{S}_N \cdot \frac{\mathbf{q}}{m_N}\right)$$
$$\mathcal{O}_{14} = i\left(\mathbf{S}_\chi \cdot \frac{\mathbf{q}}{m_N}\right)\left(\mathbf{S}_N \cdot \mathbf{v}^\perp\right)$$
$$\mathcal{O}_{15} = -\left(\mathbf{S}_\chi \cdot \frac{\mathbf{q}}{m_N}\right)\left(\left(\mathbf{S}_N \times \mathbf{v}^\perp\right) \cdot \frac{\mathbf{q}}{m_N}\right)$$

Table 5.1: Operators for effective field theory dark matter[32], applicable to either the elastic or the inelastic case. Operator $\mathcal{O}_2$ is not used further.

inconvenience when comparing different EFT papers is that there are two different normalization conventions in use. In one convention, used in Refs. [74, 97], the coupling constants for different operators have different units. In the other convention, used in Refs. [32, 43, 36], all coupling constants are unitless. This work will use the latter convention.[7]

For SI dark matter, limits are traditionally reported for the zero-velocity cross-section normalized to a nucleon, customarily denoted as σ_0.[8] It is useful to be able to convert back and forth between such limits and the equivalent EFT limits on the square of the magnitude of c_1. The conversion is $c_1^2 = f \cdot \sigma_0$, where

$$f = \frac{\pi \left(m_\chi + m_N\right)^2}{\left(m_\chi m_N\right)^2 H^4} \approx 3.69 \times 10^{37} \mathrm{cm}^{-2} \,. \tag{5.3}$$

Here, m_χ is the dark matter mass, m_N is the nucleon mass, and H is the Higgs field vacuum expectation value[9] of 246.2GeV. When $m_\chi \gg m_N$, which is typical of WIMPs, the conversion factor simplifies to the constant given in the equation.

5.2 IEFT Parameter Space

5.2.1 Regions of Parameter Space

Given that the IEFT framework defines a 30-dimensional parameter space, it is not feasible to compute an exhaustive list of limits for all possible models. Instead, the approach used here is to set limits for a collection of models that are in some way interesting, or representative, or tractable. Following the practice of some treatments of eEFT[97, 36], this analysis will be restricted to cases where only one operator at a time has nonzero coupling constants. That immediately reduces the 30-dimensional space to a 5-dimensional space. The five dimensions are: which operator mediates the interaction, its coupling constant, the dark matter mass, the mass splitting, and the extent to which isospin symmetry violation occurs. The choice of operator is a discrete finite variable with 14 possible choices, and the other four dimensions are continuous variables.

To further focus the analysis, and again following an eEFT precedent[36], we disregard isospin-violating dark matter models and assume equal coupling to protons and neutrons. This then is the basic plan for the analysis: Given a model specified by its IEFT operator, dark matter mass, and mass splitting, a limit will be set on the corresponding coupling constant.

[7] To be precise, Refs. [32] and [43] document computer codes that accept unitless coupling constants from users, before internally multiplying them by $(246.2\mathrm{GeV})^{-2}$ (the inverse square of the Higgs field vacuum expectation value) to produce coupling constants with units of GeV^{-2}. The limits we report here are limits on the unitless quantities as would be supplied by the users. Ref. [36] similarly divides out this same constant to present unitless limit plots.

[8] This is the quantity we previously denoted as σ_{norm} in Eqn. 1.65.

[9] The Higgs field vacuum expectation value does not otherwise make an explicit appearance in EFT; it is just used as a conveniently-sized normalization factor for making the coupling constants unitless.

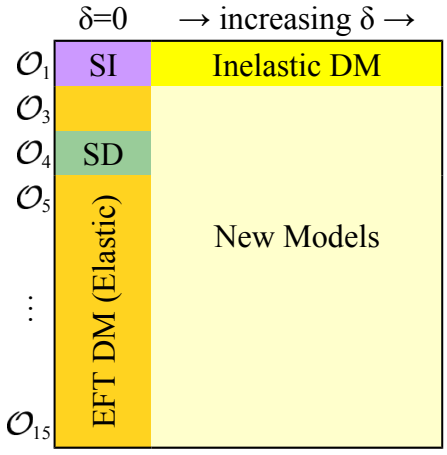


Figure 5.1: Two dimensions of IEFT parameter space, with the names of various families of models. Rows correspond to different operators. Leftmost column is models without mass splitting, while space further right represents models with nonzero mass splitting δ.

Even with these simplifying assumptions, the three-dimensional space of input variables (operator, mass, mass splitting) still covers an expansive range of dark matter interactions. As mentioned previously, the usual spin-independent and spin-dependent dark matter are there, with the former being equivalent to operator 1 with zero mass splitting, and the latter being equivalent to operator 4 with zero mass splitting. In the SD case, the assumption of isospin symmetry means that only one limit will be placed on a common coupling constant for protons and neutrons, instead of placing separate limits as is the case in most SD analyses.

Other areas within the three-dimensional parameter space also deserve mention. Dark matter that interacts by operator 1 and has positive mass-splitting is called "inelastic dark matter." (This is not to be confused with "inelastic EFT dark matter," of which inelastic dark matter is only one specific type.) Inelastic dark matter has gained attention as a possible mechanism for explaining how weak-scale dark matter can have lower-than-expected cross-sections[49].

On the other hand, if the mass splitting is set to zero, IEFT reduces exactly to eEFT. Elastic EFT and inelastic dark matter represent two limiting cases in the IEFT parameter space. For models with operators other than $\mathcal{O}_1$ and mass splittings greater than zero, neither limiting case is applicable, and we enter into the range of dark matter models that are unique to IEFT. These different possibilities are represented pictorially in Fig. 5.1.

Having laid out the range of models under consideration, we next consider how to determine the experimental effects of each.

5.2.2 Calculating IEFT Recoil Spectra

The experimental consequences of any particular WIMP-nucleus scattering model result from the model's nuclear recoil energy spectrum. This recoil spectrum is the rate of nuclear

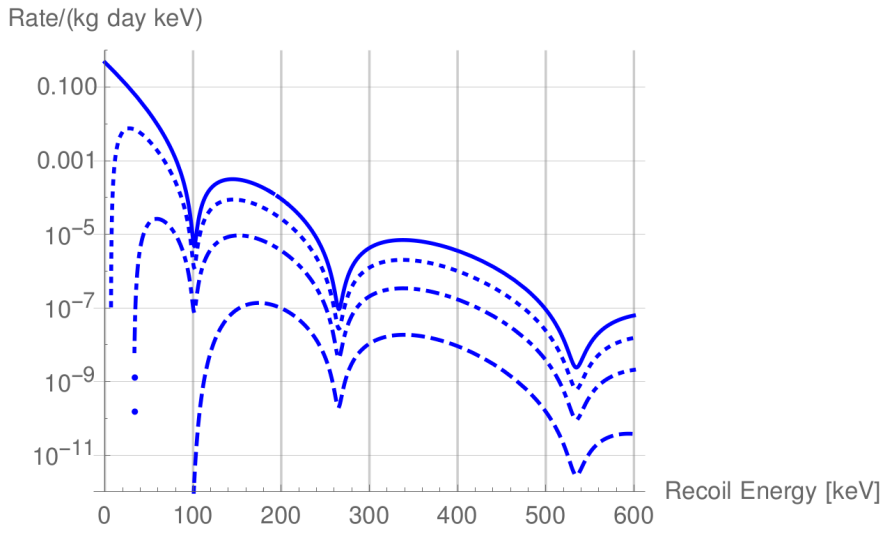


Figure 5.2: Recoil spectrum for inelastic dark matter with $m_\chi = 1\text{TeV}$ and $\sigma_0 = 10^{-40}\text{cm}^2$, in xenon. The four plotted curves, from top to bottom, correspond to mass splittings of 0keV, 100keV, 200keV, and 300keV.

recoils, as a function of recoil energy, caused by interactions with naturally-occurring halo WIMPs. It is often specified in units such as interactions per kilogram of target material per day of livetime per keV of recoil energy range.

Barello et al.[43] have developed a software library in the Mathematica language that can calculate the nuclear recoil energy spectrum for an arbitrary IEFT dark matter model.[10] A purpose-built driver program, in the form of a Mathematica notebook, uses that library to generate recoil spectra for the case of isospin-symmetric WIMP-nucleus scattering mediated by an arbitrary single operator, for any given combinations of WIMP mass and mass splitting. The calculations assume a xenon target with a naturally-occurring mix of isotopes. The calculations further assume the usual local WIMP density (0.3 GeV/cm^3) and the velocity distribution given in Eqn. 1.75.

As a quality-check of the Mathematica code, various published recoil spectra were reproduced using the software. For example, Fig. 5.2 shows the recoil spectrum for inelastic dark matter in xenon given various mass splitting values. This figure is a nearly identical recreation of the lefthand part of Figure 3 in Ref. [49]. Selected spectra from Refs. [76] and [97] have also been duplicated.

5.2.3 Selection of IEFT Models

The selection of specific models to set limits on is motivated by both theoretical and experimental considerations.

[10] This software library is in turn closely based on an eEFT software library by Anand et al.[32].

In this analysis, a subset of three operators has been selected for limit-setting: $\mathcal{O}_1$, $\mathcal{O}_3$, and $\mathcal{O}_4$. (Recall that $\mathcal{O}_2$ is not theoretically plausible.) This selection of three of the fourteen possible operators includes some of the most interesting models. $\mathcal{O}_1$ covers the important limiting cases of inelastic dark matter and traditional elastic spin-independent dark matter. Meanwhile, $\mathcal{O}_4$ reduces to the typical (elastic) spin-dependent dark matter in the $\delta = 0$ case. The operator $\mathcal{O}_3$, on the other hand, is an example of novel behavior beyond the usual SI/SD duality. $\mathcal{O}_3$ exhibits spin-orbit coupling between the nucleon spin and orbital angular momentum of the WIMP-nucleon system[96].

For each operator, limits will be set on the $\delta = 0$ (eEFT) case, as well as cases with positive mass splittings to explore behavior unique to IEFT. The range of δ's to use is driven by experimental modeling considerations. In order to reuse the cuts and background models developed for the LUX Run 4 SI analysis[20], this analysis will work with the same range of S1 sizes: 0–50phd. As a general property of IEFT dark matter, increasing the mass splitting δ suppress low-energy recoils and thereby shifts the nuclear recoil spectrum to higher energies. Therefore, sufficiently high δ values will shift the signal's S1 distribution above the usable range. To see when that happens, the fraction of nuclear recoils in the range 0–45keV was calculated for various combinations of WIMP mass and δ for each operator.[11] The results are shown in Fig. 5.3. For δ values above around 200keV, the fraction of signal in the low-energy window becomes negligible, so that value is used as a δ cutoff for the analysis.

Finally, a wide range of masses from 5GeV to 100TeV will be considered for each combination of operator and δ. Note that for a given WIMP mass, there is an upper bound on the mass splitting for which scattering is kinematically allowed.[12] This constraint is shown in Fig. 5.3 as the slightly-curved line running from lower-left to upper-right in each plot. Because scattering with mass & δ combinations to the upper-left of this curve cannot occur in the detector, limits can only be set for mass & δ combinations to the lower-right of the curve.

5.3 Profile Likelihood Ratio

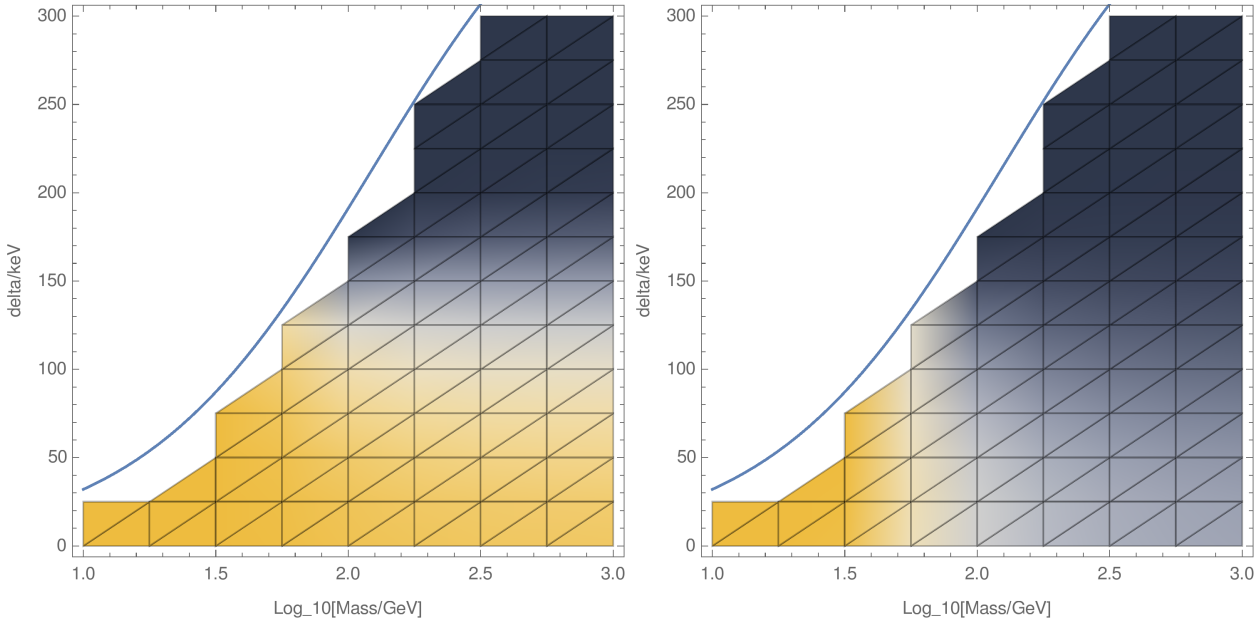
With the IEFT models selected and their nuclear recoil spectra calculated, the next step is to calculate the experimental limit on coupling constant for each model. In LUX, this is done with a profile likelihood ratio.

5.3.1 PLR Statistic

The profile likelihood ratio (PLR) statistical method allows one to determine whether an arbitrary physical model can be ruled out at an arbitrary confidence level, in light of a statistical sample that might or might not conform to the model's predictions. One of the

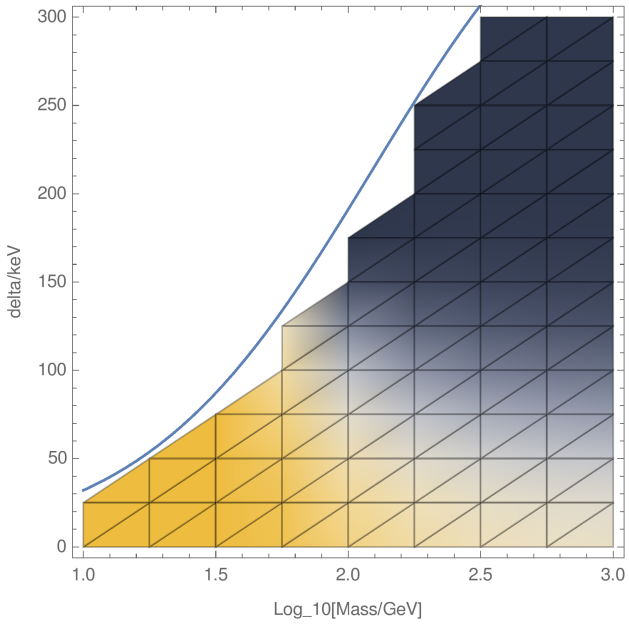
[11] 45keV nuclear recoils produce S1 sizes in the vicinity of 50phd.

[12] This happens because the escape velocity cutoff in the WIMP velocity distribution imposes an upper bound on the kinetic energy of an individual WIMP.



(a) Operator 1

(b) Operator 3



(c) Operator 4

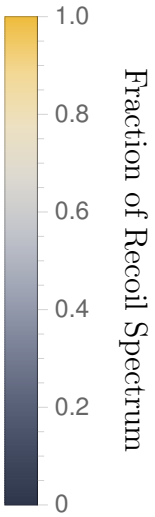


Figure 5.3: Fraction of recoil spectrum in the range 0–45keVnr as a function of WIMP mass (x-axis) and mass splitting (y-axis), for each of three operators. The smooth line in each figure shows the exact boundary of the kinematically-forbidden region.

strengths of the PLR method is that it can still give useful results even if the physical model depends on parameters that are not precisely known.[110]

In the case of LUX Run 4, the data used for the analysis is a list of five properties for each event that passes all analysis cuts. Those five properties are radial position r_{S2}, angle ϕ_{S2}, drift time z_{S2}, S1 size $S1$, and the logarithm of the S2 size $\log_{10}(S2)$. In other words, the output of the experiment is just a collection of points in a five-dimensional space.[13]

We define "physical model" to mean a full theory of what's happening in the detector. That includes which IEFT model is being analyzed, a putative value for the model's coupling constant c_i, and a complete description of the detector's backgrounds and signal response. To evaluate a specific physical model, one must generate a probability density function (PDF) of where the datapoints are expected to be if that physical model is in fact correct. This PDF has two parts: a contribution from the signal (i.e., WIMPs) and a contribution from known backgrounds. Prior to normalization, the contribution from the signal scales with the square of the coupling constant, c_i^2. The signal contribution's distribution in $(S1, \log_{10}(S2))$ space is calculated from the nuclear recoil spectrum using a simulation[136], and the signal distribution in $(r_{\mathrm{S2}}^2, \phi_{\mathrm{S2}}, z_{\mathrm{S2}})$ space is assumed to be uniform. Both the signal and background contributions depend on some values that are not precisely known. These are called "nuisance parameters." For example, the overall magnitude of each contribution to the background is treated as a nuisance parameter.

Once we have a set of datapoints and a PDF of where the data should be, the question is, "Do they match?" To answer that, we conduct a hypothesis test where the null hypothesis $\mathcal{H}_0$ is that our physical model is exactly correct, and the alternate hypothesis $\mathcal{H}_1$ is that the coupling constant value should be different.[14] For each hypothesis, a likelihood value $\mathcal{L}$ is calculated using the actual data.[15] For the null hypothesis, the likelihood value is calculated using the c_i^2 value dictated by the physical model being investigated. However, the values of all nuisance variables are allowed to vary to maximize the likelihood. Put mathematically,

$$\mathcal{L}\left(\mathcal{H}_0\right) = \mathcal{L}\left(c_i^2, \hat{\boldsymbol{\theta}}_{|c_i^2}\right) \tag{5.4}$$

where $\hat{\boldsymbol{\theta}}_{|c_i^2}$ is defined to be the set of nuisance parameter values that maximize the likelihood for the given c_i^2. For the alternate hypothesis, the calculation is done the same way except that the squared coupling constant c_i^2 is *also* allowed to vary, along with the nuisance parameters, to whatever set of values produces the greatest likelihood possible. Therefore,

$$\mathcal{L}\left(\mathcal{H}_1\right) = \mathcal{L}\left(\hat{c}_i^2, \hat{\boldsymbol{\theta}}\right) \tag{5.5}$$

[13]Technically, the event date functions as an additional dimension, although this is implemented as a "binned" (discrete) dimension with just four bins, unlike the other dimensions which are continuous. This information is used for modeling the time-dependent non-uniform electric field in Run 4.

[14]Counterintuitively, this means that the null hypothesis is *not* the absence of a signal. The null hypothesis is what one tries to reject, and we're trying to set a limit, not confirm a detection.

[15]Given a dataset $\{\mathbf{x}_i\}$ and PDF $f(\mathbf{x})$, the likelihood is defined to be $\prod_i f(\mathbf{x}_i)$.

where $\hat{c}_i^2$ and $\hat{\boldsymbol{\theta}}$ represent the values for c_i^2 and the nuisance parameters, respectively, that together maximize the likelihood function.

The hypothesis test's test statistic, denoted by λ, is the ratio of these two likelihoods:

$$\lambda\left(c_i^2\right) = \frac{\mathcal{L}\left(\mathcal{H}_0\right)}{\mathcal{L}\left(\mathcal{H}_1\right)}\,. \tag{5.6}$$

If our physical model's value for c_i^2 fits the data well, then λ will be close to unity, because then there's little to be gained by trying to further optimize c_i^2. On the other hand, if the physical model's value for c_i^2 is unrealistic, then the numerator will be much smaller than the denominator. For convenience, we transform the ratio using the monotonically decreasing function q, defined as

$$q = -2\ln\lambda\,. \tag{5.7}$$

A realistic value for c_i^2 gives a small q, while an unrealistic value gives a large q.

So how big does q have to be to reject the physical model? As the number of datapoints approaches infinity, the distribution of q approaches a χ^2 distribution. But LUX has a small number of events, so we must instead calculate the distribution of q for each physical model numerically. That can be done by using the physical model's PDF to generate a simulated set of experimental results and calculating the simulation's q-value. By repeating this step many times (5000, in this analysis), a distribution of q values can be constructed. If the actual data's q-value exceeds the 90^{th} percentile of the distribution, we reject the physical model with 90% confidence. And by testing a series of physical models that differ only in c_i^2 while leaving everything else (including the IEFT parameters of operator, mass, and mass splitting) the same, the lowest-possible rejected c_i^2 value can be found. This is the limit on c_i^2 for the given IEFT model.

Additional insight can be gained from a closely-related calculation. By calculating the q value for a large number of simulated background-only datasets, one can determine what the limit "should be" in the absence of a signal. Comparing the actual limit to this background-only distribution of expected limits (often shown on plots as shaded "Brazil bands"[16]) provides a quick check that the data is behaving as expected for a non-detection in a well-characterized detector. This also allows for seeing how much of a limit's strength is due to the experimental design, as opposed to a fortuitous statistical fluctuation during the experimental run.[110]

5.3.2 Modifications to PLR Code

For the previous LUX Run 4 analysis of spin-independent dark matter[103, 20], a profile likelihood ratio limit-setting code was implemented in C++ using the ROOT and RooFit libraries. In order to apply the PLR method to analyzing IEFT, a number of modifications to this previous code are made. Specifically, the code is generalized to support IEFT nuclear recoil spectra, and it is also generalized to take pulse shape discrimination into account.

[16]The bands are so-named because of the convention of coloring them green and yellow, which are the colors of the Brazilian flag. They are also called "sensitivity bands."

Support for IEFT Recoil Spectra

The most important PLR code modification is switching from using the recoil spectrum of elastic spin-independent dark matter (Eqn. 1.72) to accommodating an arbitrary IEFT recoil spectrum. To do that, recoil spectra are generated in advance for models of interest using the Mathematica code of Sec. 5.2.2 and stored in files.[17] Those files are then read in by the C++ code that generates signal model PDF files for the main limits code. Since the IEFT analysis uses the same range of S1 and S2 values as the SI analysis, the background PDFs do not need to be modified.

Switching from setting limits on σ_0 in the SI case, to setting limits on c_i^2 in the IEFT case, does not require any major changes to the architecture of the code, since both variables are simple multiplicative factors in their respective recoil spectra. However, in either case the limit-setting code must be given the maximum and minimum values to consider when searching for the limit. For the spin-independent analysis, these bounding values for various masses were all set by hand. But the large number of models we wish to study for the IEFT analysis makes that method impractical here. Instead, the c_i^2 value that gives an expected signal rate of 1 event in Run 4 is computed for each IEFT model. The lower and upper bounds are then taken to be 0.1 and 20 times this value, respectively.

Finally, as a practical matter, it is inconvenient to have IEFT model parameters hard-wired into the code, as this would require re-compiling and rerunning the code sequentially for each model. Instead, the code is modified so that the parameters are fed in as arguments then passed as needed in various function calls. That way, the signal model PDFs for all the desired models are easily generated in one nested for-loop. Afterwards, the actual limit-setting is carried out for many IEFT models simultaneously on a supercomputer cluster.

Support for Pulse Shape Discrimination

To build pulse shape discrimination (PSD) into the PLR code, a sixth continuous variable for each data point is added to the five that are already used in the analysis (the aforementioned r_{S2}, ϕ_{S2}, z_{S2}, $S1$, and $\log_{10}(S2)$). The sixth continuous variable is a PSD figure of merit. To be exact, the reduced prompt fraction described in Sec. 4.2.5 is used. Although the same limits would be found if the original prompt fraction were used instead, part of the code runs orders of magnitude faster with the reduced prompt fraction because this variable's empirical PDFs are mathematically simpler than those of the original prompt fraction.

Accommodating this extra variable entails adding it to both the data and to each physical model's PDF. To add the PSD dimension to the data, the reduced prompt fraction is calculated for each data point and these values are stored in a file along with the data points' other variables. In the few instances where an evt file is no longer available for a given data point (16 points out of 1221), the average reduced prompt fraction of the re-

[17] Because Mathematica and C++ use different conventions for formatting numbers, there is also a file-editing step.

maining points is used instead. To add the PSD dimension to the PDF, the background was assumed to consist of electron recoils (ERs) and the signal was assumed to consist of nuclear recoils (NRs). Therefore, the background PDF was modified by multiplying it by the reduced prompt fraction PDF for ERs, while the signal PDF was multiplied by the reduced prompt fraction PDF for NRs. Although the limit-setting code runs more slowly with the addition of another variable, no new nuisance parameters are added in the process, keeping the reduction in speed to less than a factor of two.

The modified PLR code takes as an argument a boolean value specifying whether or not to incorporate PSD. The limits in the remainder of this chapter are generated without PSD, with the next chapter addressing the effects of adding PSD.

5.4 Results

This section presents cross-section limits for various IEFT dark matter models described in Sec. 5.2, using the modified Run 4 PLR limit-setting code described in Sec. 5.3.

5.4.1 Spin-Independent Dark Matter

To begin, an 90% confidence level upper limit is calculated for elastic spin-independent dark matter as a function of WIMP mass. Elastic SI dark matter, the most conventional WIMP model, is equivalent to IEFT dark matter with operator $\mathcal{O}_1$ and mass splitting $\delta = 0$keV. A range of WIMP masses are analyzed: 5, 7, 10, 12, 14, 17, 21, 33, 50, 100, 200, 400, 1×10^3, 4×10^3, 20×10^3, and 100×10^3 GeV. For each mass, a limit on σ_0, the cross-section normalized to a nucleon, is calculated. The result is shown as the thick black line in Fig. 5.4.

In addition to this limit itself, five points in the distribution of expected limits in the absence of a signal are also calculated and overlaid as "Brazil bands" on Fig. 5.4. The thin gray line indicates the median expected limit in the absence of a signal. The green band indicates the range within one standard deviation of the median, and the wider yellow band indicates the range within two standard deviations.

As a consistency check, this limit can be compared to a limit previously set on the same dark matter model with the same data and similar methods, shown in Fig. 5.5. The two sets of results are broadly consistent, with only two significant differences. The first difference is an anomalously low result for the -2σ band for WIMP masses of 21 and 33 GeV in Fig. 5.4, and the second is difficulty setting limits at the lowest masses, at and below 12 GeV. Both errors arise from the same source: In the original Run 4 PLR code, the range of cross-sections to consider was hand-tuned for every data point to minimize convergence issues in the PLR calculation. However, for scalability the IEFT code uses automatically-generated cross-section bounds, as described in Sec. 5.3.2, which do not always perform as well. To prevent the occurrence of even more of these point-by-point problems, the IEFT PLR code is run five times for each mass and the median value used for the plot. This same method,

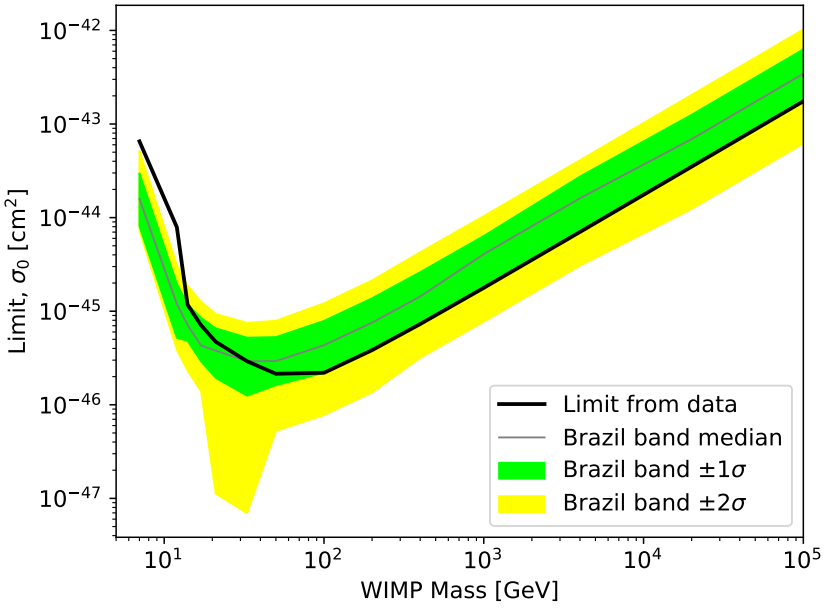


Figure 5.4: Limit on the normalized cross-section σ_0 for elastic spin-independent dark matter ($\mathcal{O}_1$ with $\delta = 0$keV), as a function of WIMP mass, computed with the generalized PLR code.

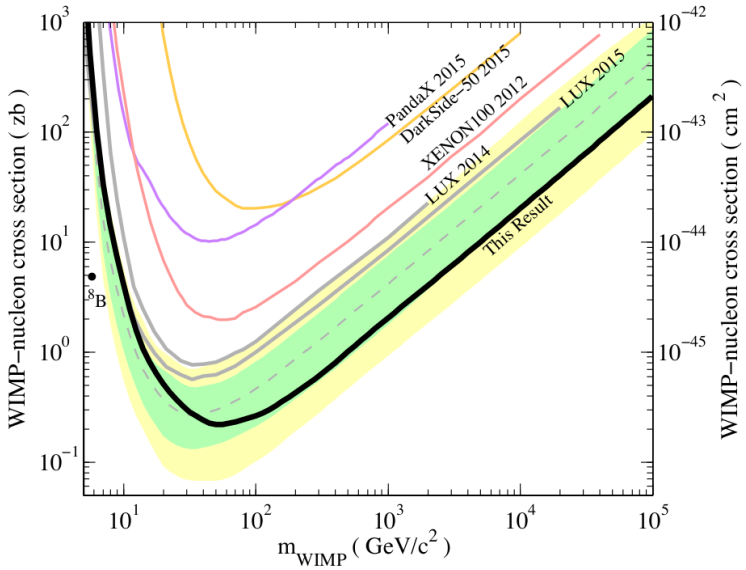


Figure 5.5: LUX Run 4 limit on σ_0 for elastic spin-independent dark matter, as a function of WIMP mass. This limit plot, computed with the original LUX Run 4 PLR code, aligns reasonably well with Fig. 5.4 computed with the generalized IEFT PLR code. Reprinted from [103].

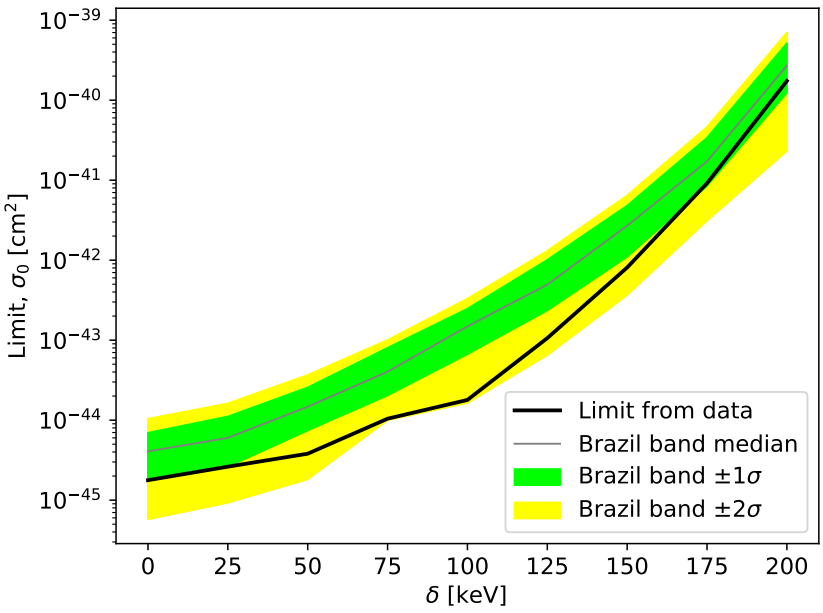


Figure 5.6: Limit on the normalized cross-section σ_0 for 1TeV inelastic dark matter ($\mathcal{O}_1$ with mass splitting $\delta > 0$keV), as a function of δ.

which mostly (but not entirely) resolves the issue, is also used for the plots in Sec. 5.4.4 below.

5.4.2 Inelastic Dark Matter

Increasing the mass splitting δ while sticking with operator $\mathcal{O}_1$ brings us into the regime of inelastic dark matter. Fig. 5.6 shows a limit on nucleon-normalized cross-section σ_0 as a function of δ for inelastic dark matter with a mass of 1TeV. Increasing δ suppresses scattering of all recoil energies, and even makes it kinematically impossible at an increasing range of low energies (see Fig. 5.2), which is why the limit becomes weaker as δ is increased.

This LUX limit can be compared with the limit on 1TeV inelastic dark matter from the PandaX collaboration, shown in Fig. 5.7. The LUX limit is stronger at low and intermediate δ values, such as 0keV and 100keV. But by $\delta = 200$keV, the PandaX limit becomes the stronger one. That turnover reflects a limitation of how the LUX data is being analyzed, not of the LUX experiment itself. Because this analysis only considers S1 sizes up to 50phd, it excludes most of the recoil spectrum of $\delta = 200$keV, artificially weakening the limit. An extension of the analysis to higher energies would especially strengthen the limit at high δ and increase the range over which LUX provides the world-leading result.

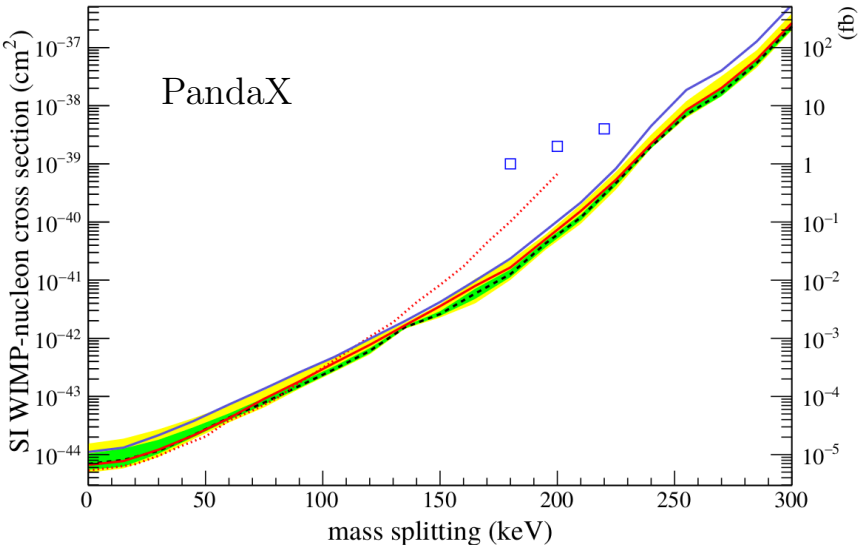


Figure 5.7: Limit on the normalized cross-section σ_0 for 1TeV inelastic dark matter, as a function of δ, as reported by the PandaX collaboration. The black dashed line is the limit, and the shaded bands are the 1- and 2-sigma Brazil bands. Compare with analogous LUX result for the $\delta = 0 - 200$keV range in Fig. 5.6. Reprinted from [55].

5.4.3 A Novel IEFT Example

Next we consider a novel model in IEFT parameter space – that is, a model that doesn't fall into either the eEFT limiting case or the inelastic dark matter limiting case. Specifically, we'll look at dark matter model with scattering mediated by operator $\mathcal{O}_3$ and a mass splitting of $\delta = 75$keV. Fig. 5.8 shows the limit as a function of WIMP mass. Here, the limit is set on the square of the operator's EFT coupling constant, c_3^2.

The most immediate observation about this limit plot is that it has the same basic shape as the limit on elastic SI dark matter (Fig. 5.4). In both cases, the limit becomes linearly proportional to the mass at high masses because, given the known local dark matter density, the number density of WIMPs is inversely proportional to WIMP mass. And in both cases, the detector sensitivity sharply falls off at low masses.

Another shared attribute of the two plots, as well as many of the other limit plots presented here, is that the actual limit falls below the median limit expected in the absence of a detection. In general, the number of background events observed in Run 4 was somewhat below the expected value, a chance statistical fluctuation which accounts for this common qualitative feature.

No previous direct detection experimental limits on c_i^2 values have been published for these novel IEFT models.[18] So unlike with the previous two result plots, there is no previously-published limit to which this result can be directly compared.

[18]However, some related quantities are presented in Ref. [43].

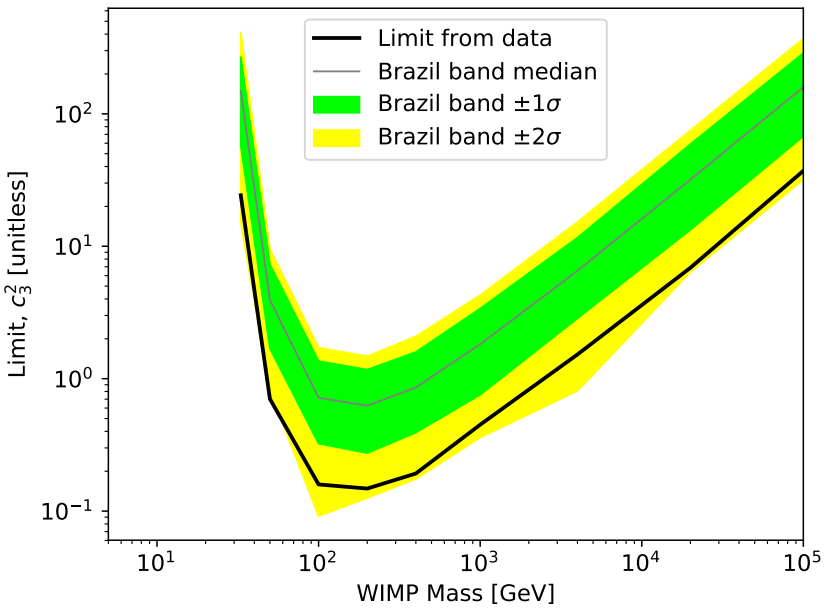


Figure 5.8: Limit on the squared coupling constant c_3^2 for IEFT dark matter with operator $\mathcal{O}_3$ and mass splitting $\delta = 75$keV, as a function of WIMP mass.

5.4.4 IEFT Dark Matter Limits

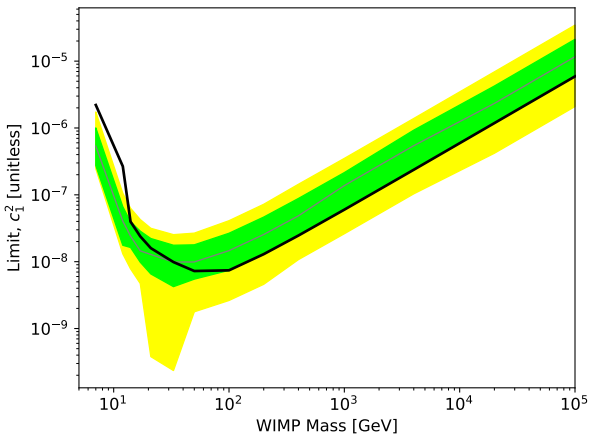
Finally, we present a systematic sweep across the subset of the IEFT parameter space selected in Sec. 5.2.3. Fig. 5.9 gives the limit on squared coupling constant versus WIMP mass for operators $\mathcal{O}_1$, $\mathcal{O}_3$, and $\mathcal{O}_4$, for the case of $\delta = 0$. Fig. 5.10 gives the corresponding limits for the case of $\delta = 100$keV, and Fig. 5.11 gives the corresponding limits when $\delta = 200$keV.

A few overall trends are evident. As δ increases, the limit on squared coupling constant c_i^2 becomes weaker for a given operator $\mathcal{O}_i$. This is because higher mass splittings suppress the recoil spectrum at all recoil energies for the same c_i value. Additionally, an increase in δ leads to an increase in the mass where the strongest limit can be set. Lighter WIMPs are less effective at causing high-energy recoils, and the kinematically forbidden range of low-energy recoils grows with δ.

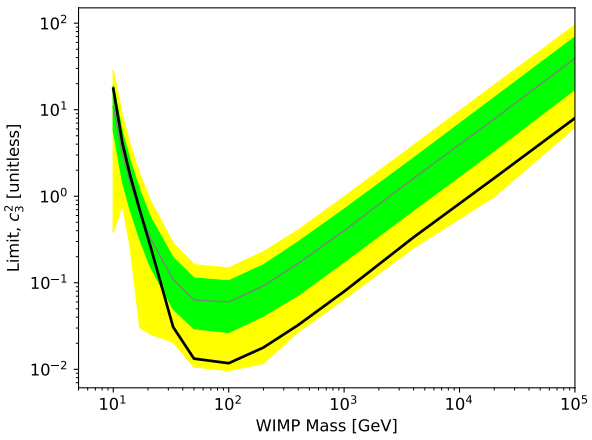
The $\delta = 0$ cases shown in Fig. 5.9 are elastic EFT dark matter. In fact, Fig. 5.9a is identical to the elastic spin-independent result previously shown in Fig. 5.4, with the only difference being that the limit is here set on c_1^2 instead of σ_0. For all three of these eEFT cases, the LUX Run 4 results shown here can be compared to the corresponding published eEFT results from the Xenon100 experiment[36], shown in Fig. 5.12. Note that the Xenon100 limits are expressed in terms of the same unitless squared coupling constants as are the LUX results, allowing for a direct comparison. The Xenon100 results show the same qualitative shape as the LUX results, but the LUX limits are stronger by roughly an order of magnitude.

The $\delta = 200$ limits shown in Fig. 5.10 represents an edge case of what can be analyzed at

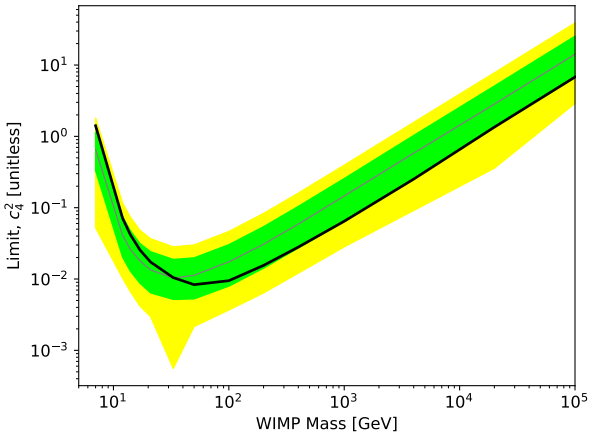
all by working with S1 sizes constrained to the range 0–50phd. For such high mass splittings, almost the entire recoil spectrum maps to typical S1 sizes above the 0–50phd range, with only a small tail of the distribution falling below 50phd. As a result, these results are especially sensitive to backgrounds in a small range of S1 sizes, and they are weaker than they could be. In general, extending the analysis to higher energies, and therefore to higher S1 and S2 sizes, increases the range of IEFT models for which limits can be set and also increases the strength of limits (like these) for models with recoil spectra that fall partially outside the original range of study. Steps that would be required to extend the analysis to higher energy are considered in the next chapter.



(a) Operator 1, $\delta = 0$keV



(b) Operator 3, $\delta = 0$keV



(c) Operator 4, $\delta = 0$keV

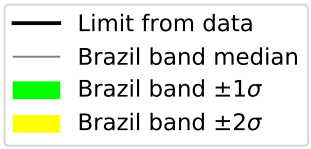
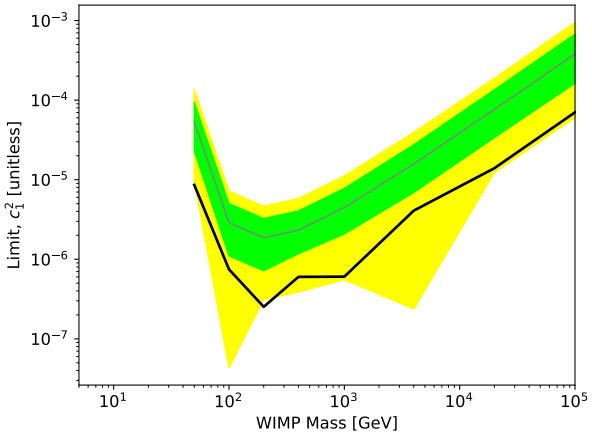
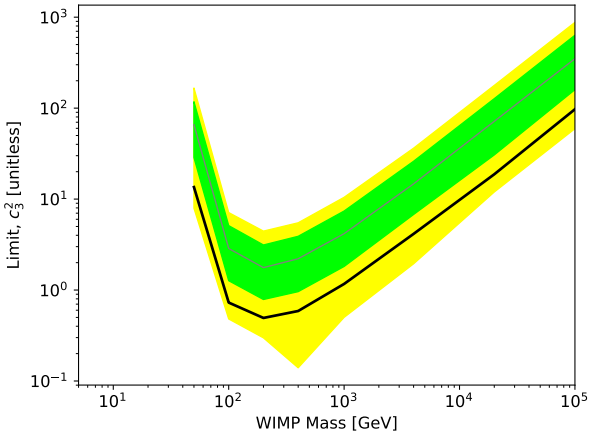


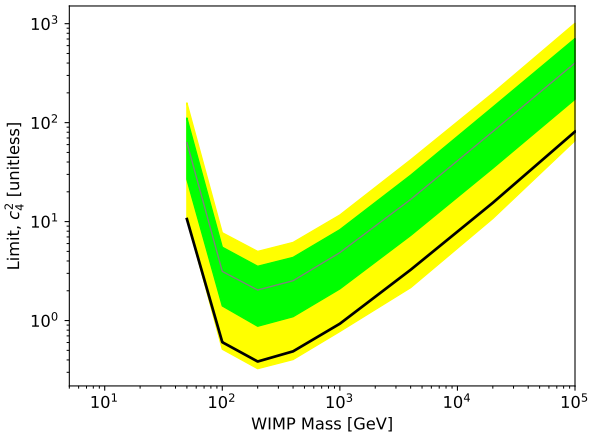
Figure 5.9: Limit on the squared coupling constant c_i^2 versus WIMP mass for elastic EFT dark matter (zero mass splitting) with interactions mediated by each of three different operators.



(a) Operator 1, $\delta = 100$keV



(b) Operator 3, $\delta = 100$keV



(c) Operator 4, $\delta = 100$keV

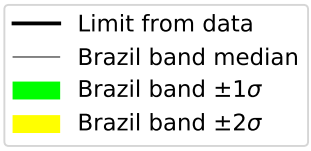
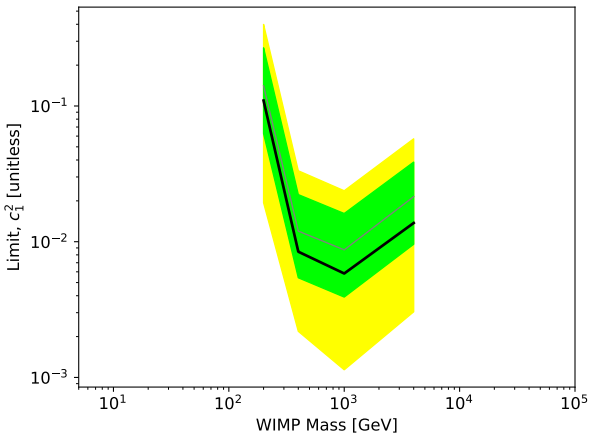
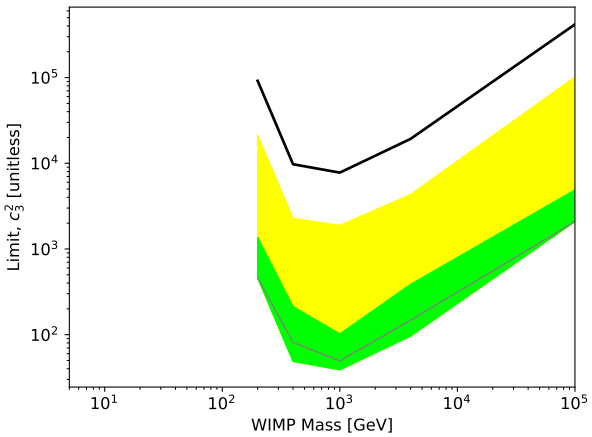


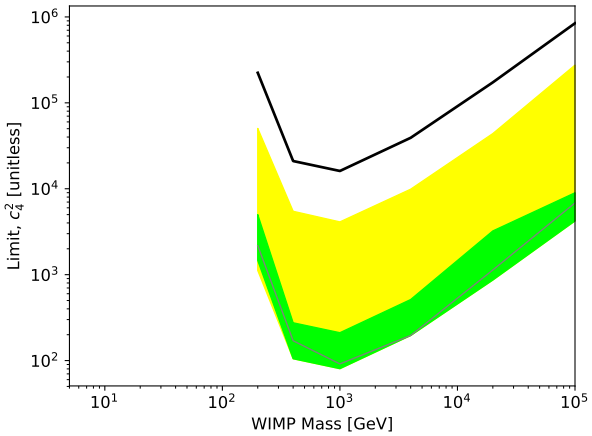
Figure 5.10: Limit on the squared coupling constant c_i^2 versus WIMP mass for IEFT dark matter with mass splitting $\delta = 100$keV and each of three different operators.



(a) Operator 1, $\delta = 200$keV



(b) Operator 3, $\delta = 200$keV



(c) Operator 4, $\delta = 200$keV

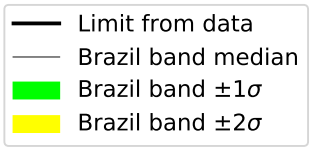
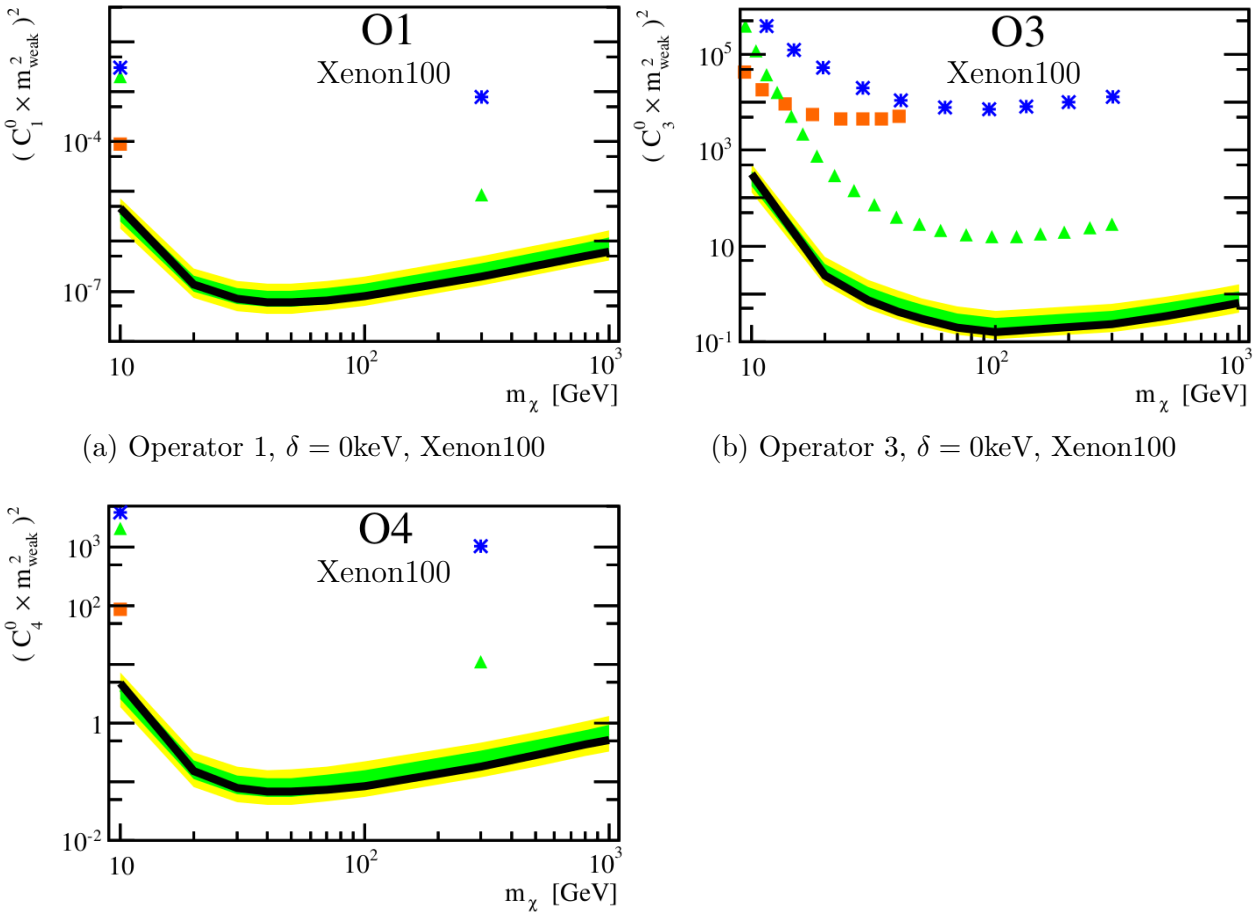


Figure 5.11: Limit on the squared coupling constant c_i^2 versus WIMP mass for IEFT dark matter with mass splitting $\delta = 200$keV and each of three different operators.



(a) Operator 1, $\delta = 0$keV, Xenon100

(b) Operator 3, $\delta = 0$keV, Xenon100

(c) Operator 4, $\delta = 0$keV, Xenon100

Figure 5.12: Xenon100 experimental limit on elastic EFT dark matter for operators 1, 3, and 4. Compare to Fig. 5.9, noting that the y-axes of both sets of plots are in terms of the same units. Reprinted from [36].

Chapter 6

Next Steps

In the previous chapter, a profile likelihood ratio (PLR) technique was used to set limits on a variety of inelastic effective field theory (IEFT) models of dark matter. Next, we consider the effects of pulse shape discrimination (PSD) on such limits, showing the effect to be minimal at low energies and motivating an ongoing effort to expand to higher energies.

6.1 PSD for Limit Setting

To evaluate the effect of incorporating PSD into the limit-setting code, a limit is set on the same IEFT model first without PSD sensitivity then with it. The model used for the limit is IEFT dark matter with interactions mediated by operator $\mathcal{O}_3$, initial WIMP mass of 50GeV, and a mass splitting of $\delta = 75$keV between the incoming and outgoing WIMPs. This is one of the models discussed in Sec. 5.4.3, for which limits are shown in Fig. 5.8. This specific model is chosen because it represents a best-case scenario for showing improvements from PSD. The PSD separation between electron recoils (ERs) and nuclear recoils (NRs) increases with S1 size, as shown in Fig. 4.11. Therefore, a greater PSD effect should be seen with recoil spectra having more events near the top of the S1 size range under study (0–50phd). This model's median S1 size is near 50phd (with about half the recoils actually getting missed because of their high energy), so it meets this condition.

For this IEFT model, the PLR limit-setting code is run 10 times without using PSD, and 10 times with PSD as described in Sec. 5.3.2. In the former case, the Brazil band median limit on squared coupling constant c_3^2 is 3.93 ± 0.02; in the latter case it is 3.92 ± 0.04. This does not indicate a meaningful improvement.[1]

To get a sense of just how much stronger pulse shape discrimination would need to be in order to get a noticeable improvement in the limits, we consider two hypothetical scenarios. In the first scenario, the PSD separation between the ER and NR bands is

[1]Nor is it statistically significant, with a p-value of 0.42 in a single-tailed two-sample T-test, although that could be resolved with a large number of additional runs to reduce statistical uncertainty. In a separate test, the limit on 50 TeV elastic SI dark matter also was not meaningfully strengthened by PSD.

artificially increased to half a standard deviation at all S1 sizes. Under this assumption, using PSD causes a 6% improvement in the expected limit. This separation is close to what actually occurs in the higher-energy range in the vicinity of 90–100phd. As a second hypothetical scenario, the PSD separation is artificially increased to one full standard deviation, in which case PSD would provide a 22% improvement in the expected limit. This scenario, while not very realistic, suggests by comparison with the previous scenario that the expected limit has a nonlinear dependence on PSD separation.

These observations suggest that while PSD may not be helpful in the low-energy range (with S1 sizes of 0–50phd), there remains a possibility for improving limits with the improved PSD separation found at higher energies. As noted in the previous chapter, extending LUX's analysis to higher energies would also strengthen current limits on many IEFT models[2] and allow for setting limits on a wider range of IEFT models.

6.2 Preliminary Work Towards Higher Energies

Within the broader LUX collaboration, an extension of the full analysis to higher energies is now under way. This has many elements, from extending the range of background PDFs to re-evaluating the analysis cuts to be used.[3] This section describes one small aspect of that effort: a consistency check between data and simulation.

To generate signal and background PDFs at high energies, it is necessary to know the distribution of S2 size versus S1 size out to high energies for both electron recoils and nuclear recoils. For the case of electron recoils, one way to study the distribution is to use the isotope carbon-14.

Shortly before the LUX detector's decommissioning, calibration data was recorded with a carbon-14 source. The C-14 was incorporated into methane, $^{14}CH_4$, which was injected into the detector. Because C-14 decays with the emission of β^- particles of up to 156keV, it produces electron recoils in the detector with S1 sizes of up to around 800phd.[42] Compared to tritiated methane ERs, with a maximum S1 size of just over 100phd, C-14 permits the study of a much greater range of electron recoil S1 sizes.

To calculate the ER band from the C-14 data, a subset of the C-14 events is selected so as to have a uniform energy distribution. The S2-vs-S1 data is then fit with an empirical model.[4] The result is shown as the curves in Fig. 6.1. In each plot, the solid line is the median, with the dot-dashed lines representing the 10% and 90% confidence levels.[42] Also shown in the figure are the Run 4 WIMP search datapoints above 50phd.[5] The background is expected to be almost entirely electron recoils, so the close alignment between the actual

[2]In fact, most of the LUX limits shown in the chapter would be strengthened, with the exception of those in Figs. 5.4 and 5.9a, and the leftmost parts of Figs. 5.8, 5.9b, and 5.9c.

[3]This is being done as a true blind analysis, which is possible because the Run 4 data was never unsalted in the 50–100phd range during the whole course of the previous 0–50phd analysis.

[4]The model has two power law terms and a linear term, all added together[42].

[5]Fig. 6.1 include salt events, but the data was only salted below 100phd.

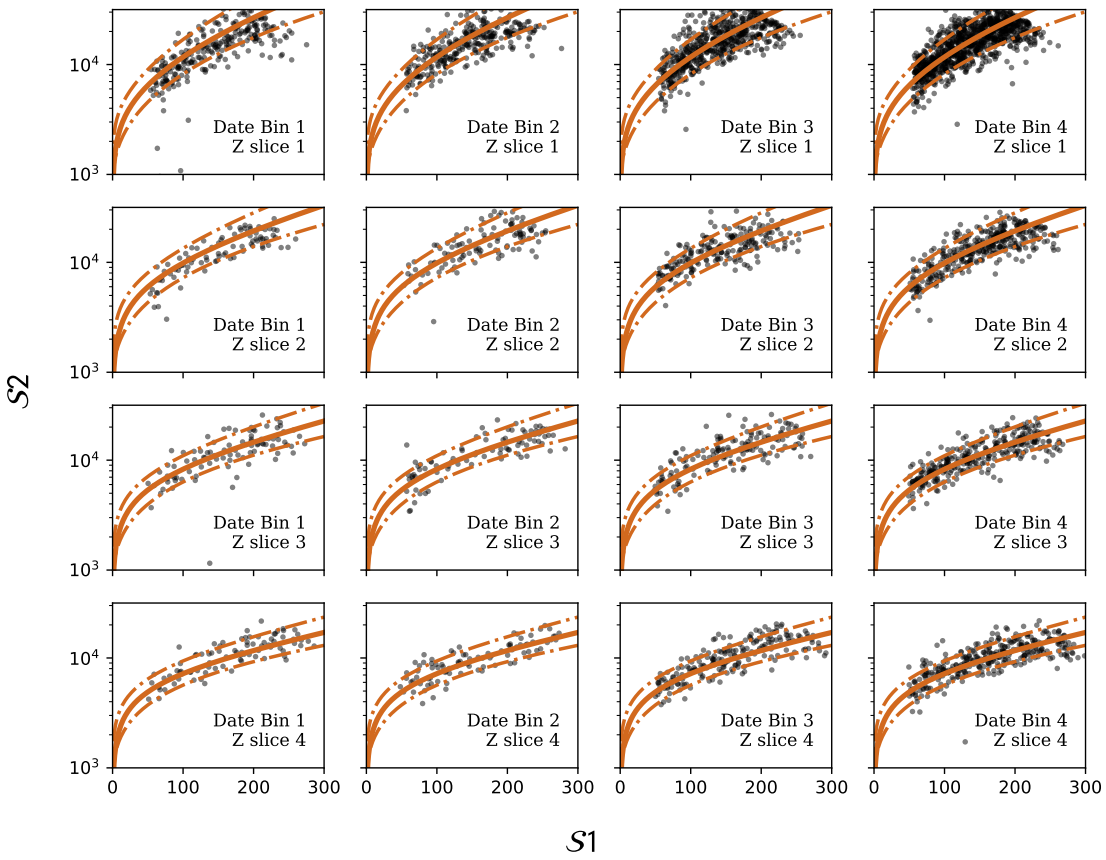


Figure 6.1: Plot of S2 versus S1 size for Run 4 WIMP search data above 50phd (dots), overlaid with distribution from C-14 calibration data (lines). To account for changing detector conditions, the data is divided into four time intervals and four depth bins. C-14 fit from [42].

background events and the ER distribution measured with calibration data is consistent with that expectation. To address the detector's varying electric field, the data was divided into four depth bins and four time intervals for some analyses, and it is similarly divided in Fig. 6.1 as well. However, the C-14 band varies only with depth, not time, because the C-14 data was collected only once.

As a second consistency check on the high-energy ER band, the experimentally measured band from the C-14 and WIMP search data was compared to a simulation. Fig. 6.2 shows the high-energy ER and NR bands as simulated with NEST[136]. As with the previous figure, the three lines for each band are the estimates of the median and 80% confidence interval. The simulated ER band is broadly consistent with the experimentally-measured band.[6] This agreement strengthens the evidence that high-energy electron recoils in LUX

[6]A slight downward offset in the simulation, as compared to the data, is found in the low-energy (S1<

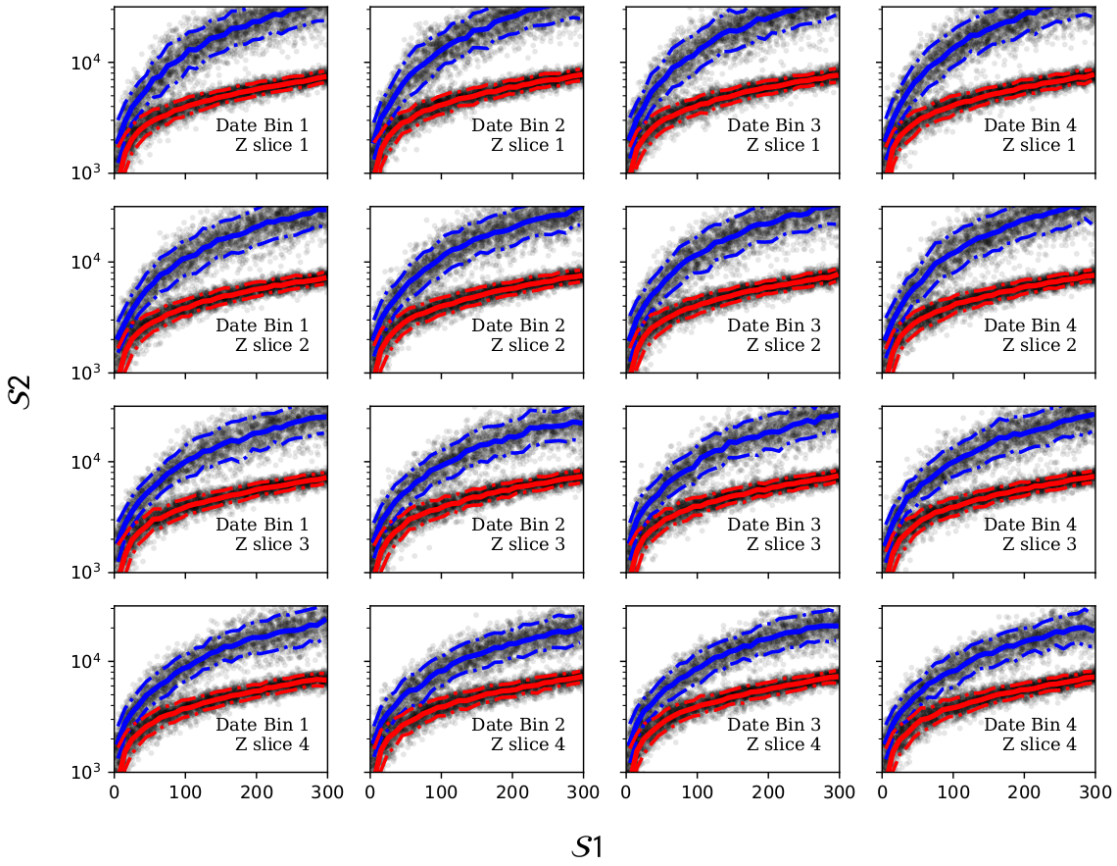


Figure 6.2: Plot of S2 versus S1 size from libNEST simulations of the ER band (higher, with blue lines) and NR band (lower, with red lines). The gray dots are individual simulation results. To account for changing detector conditions, the data is divided into four time intervals and four depth bins.

are behaving as expected.

Fig. 6.2 also has something to say about the usefulness of pulse shape discrimination at high energies. As energy increases, the S2 size gap between the ER and NR bands widens substantially. In fact, as S1 increases from 75phd to 150phd, the leakage fraction drops by two full orders of magnitude. This would hinder the usefulness of PSD for the highest S1 sizes. Put simply, there's little benefit to be gained from adding pulse shape information if the convention method for sorting ERs and NRs (looking at $\log_{10}$(S2) vs. S1 size) already provides almost perfect separation.

However, there could be a favorable intermediate range, where the increase in S1 size has strengthened the discriminating power of pulse shapes but not rendered it unnecessary through greatly improved $\log_{10}$(S2) vs. S1 separation. The utility of pulse shape discrimina-

50phd) regime as well as at higher energies, suggesting that some simulation setting is not optimal.

tion at S1 sizes above 50phd remains an open question, and a definitive answer must await the full extension of the LUX analysis to high energies.

6.3 Conclusion

As LUX data analysis expands to higher energies, a wider range of data recorded by this extraordinary instrument will be used to its fullest. Such efforts stand to benefit from the tools developed here to tackle problems ranging from working with inelastic effective field theory models to visually investigating events in the detector.

Working within a low-energy range, a variety of new dark matter limits have now been set. These include a world-leading limit on low-mass-splitting inelastic dark matter, as well as the first c_i^2 limits for novel IEFT models to be reported by a WIMP direct detection experiment. The eEFT limits shown here are stronger than those recently reported by another dual-phase xenon collaboration.

The scope of this project has involved much of the design and nearly all of the implementation of the computational tools discussed in Chapter 3, as well as the PSD and IEFT analysis work discussed in the subsequent three chapters. As noted throughout the latter chapters, existing tools and techniques (including the Run 4 PLR code, the Run 4 cut definitions, and the IEFT Mathematica code) have been employed in support of new analysis work. This work dovetails with related efforts by others in the collaboration, particularly as described in Secs. 4.3.1 and 6.2. Certain pedagogical approaches in Chapter 1 are also specific to this project.

Broadly, this discussion has followed a path from the theories of dark matter to the hardware and data management concerns of a real-world dark matter experiment. The visualization of signal pulses, among those data concerns, is a springboard to a deeper exploration of scintillation light pulse shapes. These in turn, in combination with powerful statistical techniques, bring us back to looking at the dark matter theories where the story began.

Bibliography

[1] J. Aalbers et al. (DARWIN Collaboration). "DARWIN: Towards the ultimate dark matter detector". In: *Journal of Cosmology and Astroparticle Physics* 11 (2016), p. 017.

[2] C.E. Aalseth et al. (CoGeNT Collaboration). "CoGeNT: A search for low-mass dark matter using p-type point contact germanium detectors". In: *Physical Review D* 88 (2013), p. 012002.

[3] M.G. Aartsen et al. (IceCube Collaboration). "Search for dark matter annihilations in the sun with the 79-string IceCube detector". In: *Physical Review Letters* 110 (Mar. 29, 2013), p. 131302.

[4] Lotty Ackerman et al. "Dark matter and dark radiation". In: *Physical Review D* 79 (Jan. 23, 2009), p. 023519.

[5] P.A.R. Ade et al. "Planck 2015 results. XIII. Cosmological parameters". In: *Astronomy and Astrophysics* 594 (2016), A13.

[6] N. Aghanim et al. "Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters". In: *Astronomy and Astrophysics* 594 (2016), A11.

[7] P. Agnes et al. (DarkSide Collaboration). "Results from the first use of low radioactivity argon in a dark matter search". In: *Physical Review D* 93 (2016), p. 081101.

[8] R. Agnese et al. (SuperCDMS Collaboration). "Projected sensitivity of the SuperCDMS SNOLAB experiment". In: *Physical Review D* 95 (2017), p. 082002.

[9] S. Ahlen et al. (DMTPC Collaboration). "First dark matter search results from a surface run of the 10-L DMTPC directional dark matter detector". In: *Physics Letters B* 695 (2011), pp. 124–129.

[10] S.P. Ahlen et al. "Limits on cold dark matter candidates from an ultralow background germanium spectrometer". In: *Physics Letters B* 195.4 (Sept. 17, 1987), pp. 603–608.

[11] Sangtae Ahn and Jeffrey A. Fowler. *Standard Errors of Mean, Variance, and Standard Deviation Estimators*. `https://web.eecs.umich.edu/~fessler/papers/files/tr/stderr.pdf`. 2003.

[12] D.S. Akerib et al. (LUX Collaboration). *Chromatographic separation of radioactive noble gases from xenon*. arXiv:1605.03844. 2016.

[13] D.S. Akerib et al. (LUX Collaboration). "Data acquisition and readout system for the LUX dark matter experiment". In: *Nuclear Instruments and Methods in Physics Research A* 668 (2012), pp. 1–8.

[14] D.S. Akerib et al. (LUX Collaboration). "First results from the LUX dark matter experiment at the Sanford Underground Research Facility". In: *Physical Review Letters* 112 (Mar. 4, 2014), p. 091303.

[15] D.S. Akerib et al. (LUX Collaboration). *First searches for axions and axion-like particles with the LUX experiment*. arXiv:1704.02297. 2017.

[16] D.S. Akerib et al. (LUX Collaboration). "Improved limits on scattering of weakly interacting massive particles from reanalysis of 2013 LUX data". In: *Physical Review Letters* 116 (Apr. 22, 2016), p. 161301.

[17] D.S. Akerib et al. (LUX Collaboration). *Limits on spin-dependent WIMP-nucleon cross section obtained from the complete LUX exposure*. arXiv:1705.03380. 2017.

[18] D.S. Akerib et al. (LUX Collaboration). *Liquid xenon scintillation measurements and pulse shape discrimination in the LUX dark matter detector*. arXiv:1802.06162. 2018.

[19] D.S. Akerib et al. (LUX Collaboration). *Low-energy (0.7–74 keV) nuclear recoil calibration of the LUX dark matter experiment using D-D neutron scattering kinematics*. arXiv:1608.05381. 2016.

[20] D.S. Akerib et al. (LUX Collaboration). "Results from a search for dark matter in the complete LUX exposure". In: *Physical Review Letters* 118 (Jan. 13, 2017), p. 021303.

[21] D.S. Akerib et al. (LUX Collaboration). "Results on the spin-dependent scattering of weakly interacting massive particles on nucleons from the Run 3 data of the LUX experiment". In: *Physical Review Letters* 116 (Apr. 22, 2016), p. 161302.

[22] D.S. Akerib et al. (LUX Collaboration). "Signal yields, energy resolution, and recombination fluctuations in liquid xenon". In: *Physical Review D* 95 (2017), p. 012008.

[23] D.S. Akerib et al. (LUX Collaboration). "Technical results from the surface run of the LUX dark matter experiment". In: *Astroparticle Physics* 45 (2013), pp. 34–43.

[24] D.S. Akerib et al. (LUX Collaboration). "The Large Underground Xenon (LUX) experiment". In: *Nuclear Instruments and Methods in Physics Research A* 704 (2013), pp. 111–126.

[25] D.S. Akerib et al. (LUX Collaboration). "Tritium calibration of the LUX dark matter experiment". In: *Physical Review D* 93 (2016), p. 072009.

[26] D. Akimov et al. *Coherent scattering investigations at the Spallation Neutron Source: A Snowmass white paper*. arXiv:1310.0125. Oct. 1, 2013.

[27] D. Akimov et al. "Observation of coherent elastic neutrino-nucleus scattering". In: *Science* (Aug. 3, 2017), aao0990.

[28] Globus Alliance. *Globus: How it works*. `https://www.globus.org/how-it-works`. 2017.

[29] G.J. Alner et al. (UK Dark Matter Collaboration). "First limits on nuclear recoil events from the ZEPLIN I galactic dark matter detector". In: *Astroparticle Physics* 23 (2005), pp. 444–462.

[30] G.J. Alner et al. "First limits on WIMP nuclear recoil signals in ZEPLIN-II: A two-phase xenon detector for dark matter detection". In: *Astroparticle Physics* 28 (2007), pp. 287–302.

[31] C. Amole et al. (PICO Collaboration). "Dark matter search results from the Pico-2L C_3F_8 bubble chamber". In: *Physical Review Letters* 114 (June 12, 2015), p. 231302.

[32] Nikhil Anand, Liam Fitzpatrick, and W.C. Haxton. "Weakly interacting massive particle-nucleus elastic scattering response". In: *Physical Review C* 89 (2014), p. 065501.

[33] Adam Anderson. "Phonon-based position determination in SuperCDMS iZIP detectors". In: *Journal of Low Temperature Physics* 176 (2014), pp. 959–965.

[34] J. Angle et al. (XENON10 Collaboration). "Constraints on inelastic dark matter from XENON10". In: *Physical Review D* 80 (2009), p. 115005.

[35] G. Angloher et al. (CRESST Collaboration). "Results on light dark matter particles with a low-threshold CRESST-II detector". In: *European Physics Journal C* 76 (2016), p. 25.

[36] E. Aprile et al. (XENON Collaboration). "Effective field theory search for high-energy nuclear recoils using the XENON100 dark matter detector". In: *Physical Review D* 96 (2017), p. 042004.

[37] E. Aprile et al. (XENON Collaboration). "First dark matter search results from the XENON1T experiment". In: *Physical Review Letters* 119 (2017), p. 181301.

[38] E. Aprile et al. (XENON Collaboration). "Physics reach of the XENON1T dark matter experiment". In: *Journal of Cosmology and Astroparticle Physics* 4 (2016), p. 027.

[39] E. Armengaud et al. (EDELWEISS Collaboration). "Constraints on low-mass WIMPs from the EDELWEISS-III dark matter search". In: *Journal of Cosmology and Astroparticle Physics* 05 (2016), p. 019.

[40] The ATLAS Collaboration. "Search for dark matter at $\sqrt{s} = 13$ TeV in final states containing an energetic photon and large missing transverse momentum with the ATLAS detector". In: *European Physics Journal C [submitted]* (2017).

[41] E.A. Bagnaschi et al. *Supersymmetric dark matter after LHC Run 1*. Tech. rep. KCL-PH-TH/2015-33. arXiv:1508.01173. King's College, London, Aug. 5, 2015.

[42] Jon Balajthy. *Personal communication*. 2017.

[43] G. Barello, Spencer Chang, and Christopher A. Newby. "A model independent approach to inelastic dark matter scattering". In: *Physical Review D* 90 (2014), p. 094027.

[44] L. Baudis et al. "Signatures of dark matter scattering inelastically off nuclei". In: *Physical Review D* 88 (2013), p. 115014.

[45] R. Bernabei et al. (DAMA/LIBRA Collaboration). "Final model-independent results of DAMA/LIBRA-phase1 and perspectives for phase2". In: *Physics of Particles and Nuclei* 46.2 (Mar. 2015), pp. 138–146.

[46] Gianfranco Bertone and Dan Hooper. *A history of dark matter*. Tech. rep. FERMILAB-PUB-16-157-A. FermiLab, May 26, 2016.

[47] J. Billard, E. Figueroa-Feliciano, and L. Strigari. "Implication on neutrino backgrounds on the reach of next generation dark matter direct detection experiments". In: *Physical Review D* 89 (Jan. 27, 2014), p. 023524.

[48] Nassim Bozorgnia. "Dark matter direct detection - II". In: *School on Dark Matter*. São Paulo, Brazil, July 7, 2016.

[49] Joseph Bramante et al. "Inelastic frontier: Discovering dark matter at high recoil energy". In: *Physical Review D* 94 (2016), p. 115026.

[50] D.O. Caldwell et al. "Half-life limits on the zero-neutrino and two-neutrino double-β decay of ^{76}Ge". In: *Physical Review D* 33.9 (May 1, 1986), pp. 2737–2739.

[51] D.O. Caldwell et al. "Laboratory limits on galactic cold dark matter". In: *Physical Review Letters* 61.5 (Aug. 1, 1988), pp. 510–513.

[52] J. Calvo et al. (ArDM Collaboration). "Commissioning of the ArDM experiment at the CanFranc underground laboratory: First steps towards a tonne-scale liquid argon time projection chamber for dark matter searches". In: *Journal of Cosmology and Astroparticle Physics* 03 (2017), p. 003.

[53] Sean M. Carroll. "Cosmology for Particle Physicists". In: *CERN Academic Training Lectures*. May 10-13, 2005.

[54] Riccardo Catena and Piero Ullio. "A novel determination of the local dark matter density". In: *Journal of Cosmology and Astroparticle Physics* 8 (Aug. 3, 2010), p. 004.

[55] Xun Chen et al. (PandaX-II Collaboration). "Exploring the dark matter inelastic frontier with 79.6 days of PandaX-II data". In: *Physical Review D* 96 (2017), p. 102007.

[56] K. Choi et al. (Super-Kamiokande Collaboration). "Search for neutrinos from annihilation of captured low-mass dark matter particles in the sun by Super-Kamiokande". In: *Physical Review Letters* 114 (Apr. 10, 2015), p. 141301.

[57] Soo-Min Choi, Yoo-Jin Kang, and Hyun Min Lee. "On thermal production of self-interacting dark matter". In: *Journal of High Energy Physics* (Dec. 19, 2016), p. 99.

[58] Douglas Clowe, Anthony Gonzalez, and Maxim Markevitch. "Weak-lensing mass reconstruction of the interacting cluster 1E 0657-558: Direct evidence for the existence of dark matter". In: *Astrophysical Journal* 604.2 (Apr. 1, 2004), pp. 596–603.

[59] The CMS Collaboration. "Search for dark matter and unparticles in events with a Z boson and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* (Mar. 10, 2017), p. 061.

[60] P. Cushman et al. *Snowmass CF1 summary: WIMP dark matter direct detection*. arXiv:1310.8327v2. Nov. 3, 2013.

[61] Raymond Davis Jr., Don S. Harmer, and Kenneth C. Hoffman. "Search for neutrinos from the sun". In: *Physical Review Letters* 20.21 (May 20, 1968), pp. 1205–1209.

[62] E. Daw et al. (DRIFT Collaboration). *The DRIFT dark matter experiments*. arXiv: 1110.0222. Oct. 2, 2011.

[63] Anand Desai and Alex Moskowitz. *DMTOOLS dark matter plot limit generator*. `dmtools.brown.edu/plots`. 2013.

[64] A. Drukier and L. Stodolsky. "Principles and applications for neutrino physics and astronomy". In: *Physical Review D* 30.11 (Dec. 1, 1984), pp. 2295–2309.

[65] Daniel J. Eisenstein. "Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies". In: *Astrophysical Journal* 633 (Nov. 10, 2005), pp. 560–574.

[66] John Ellis, Andrew Ferstl, and Keith A. Olive. "Re-evaluation of the elastic scattering of supersymmetric dark matter". In: *Physics Letters B* 481 (May 25, 2000), pp. 304–314.

[67] Christoph Englert et al. "Exploring the Higgs portal". In: *Physics Letters B* 703.3 (Sept. 14, 2011), pp. 298–305.

[68] Carlos Faham. *Personal communication*. 2014.

[69] Carlos H. Faham. "Prototype, surface commissioning and photomultiplier tube characterization for the Large Underground Xenon (LUX) direct dark matter search experiment". PhD thesis. Brown University, 2014.

[70] C.H. Faham et al. "Measurements of wavelength-dependent double photoelectron emission from single photons in VUV-sensitive photomultiplier tubes". In: *Journal of Instrumentation* 10 (Sept. 9, 2015), p. 09010.

[71] Malcolm Fairbairn. *Velocity distribution of dark matter*. `http://www.hep.ph.ic.ac.uk/susytalks/fairbairnimperialoneday.pdf`.

[72] Nasim Fatemighomi (for the DEAP-3600 Collaboration). "DEAP-3600 dark matter experiment". In: *PIC 2015: Physics in Collision*. arXiv:1609.07990. University of Warwick, Sept. 15-19, 2015.

[73] Jonathan L. Feng et al. "Isospin-violating dark matter". In: *Physics Letters B* 703 (2011), pp. 124–127.

[74] A. Liam Fitzpatrick et al. "The effective field theory of dark matter direct detection". In: *Journal of Cosmology and Astroparticle Physics* 2 (2013), p. 004.

[75] D.J. Fixsen. "The temperature of the cosmic microwave background". In: *Astrophysical Journal* 707 (Dec. 20, 2009), pp. 916–920.

[76] Rick Gaitskell. "Noble travails: Noble liquid dark matter detectors". In: *American Physical Society April Meeting*. 2007.

[77] G.K. Giovanetti et al. (MAJORANA Collaboration). "A dark matter search with MALBEK". In: *Physics Procedia* 61 (2015), pp. 77–84.

[78] Mark W. Goodman and Edward Witten. "Detectability of certain dark-matter candidates". In: *Physical Review D* 31.12 (June 15, 1985), pp. 3059–3063.

[79] David Griffiths. *Introduction to Elementary Particles*. John Wiley & Sons, 1987.

[80] Francis Halzen and Alan D. Martin. *Quarks and Leptopns: An Introductory Course in Modern Particle Physics*. John Wiley & Sons, 1984.

[81] Akira Hitachi et al. "Effect of ionization density on the time dependence of luminescence from liquid argon and xenon". In: *Physical Review B* 27.9 (May 1, 1983), pp. 5279–5285.

[82] Yonit Hochberg et al. "Mechanism for thermal relic dark matter of strongly interacting massive particles". In: *Physical Review Letters* 113 (Oct. 22, 2014), p. 171301.

[83] D.P. Hogan. *Admin's guide to Visualux and the LBL NERSC data mirror*. Tech. rep. LuxDB00000478. LUX Collaboration, Apr. 25, 2016.

[84] J. Hoskins et al. (ADMX Collaboration). "Search for nonvirialized axionic dark matter". In: *Physical Review D* 84 (2011), 121302(R).

[85] Wayne Hu and Scott Dodelson. "Cosmic microwave background anisotropies". In: *Annual Reviews of Astronomy and Astrophysics* 40 (Sept. 2002), pp. 171–216.

[86] Edwin Hubble. "A relation between distance and radial velocity among extra-galactic nebulae". In: *Proceedings of the National Academy of Sciences* 15.3 (1929), pp. 168–173.

[87] John Huchra. *The Hubble Constant*. `https://web-beta.archive.org/web/20170216000944/https://www.cfa.harvard.edu/~dfabricant/huchra/hubble/`. 2008.

[88] P. Hut. "Limits on masses and number of neutral weakly interacting particles". In: *Physics Letters* 69B.1 (July 18, 1977), pp. 85–88.

[89] J.H. Jeans. "The motions of stars in a Kapteyn-universe". In: *Monthly Notices of the Royal Astronomical Society* 82.3 (Jan. 13, 1922), pp. 122–132.

[90] Gerard Jungman, Marc Kamionkowski, and Kim Griest. "Supersymmetric dark matter". In: *Physics Reports* 267 (1996), pp. 195–373.

[91] L.W. Kastens et al. "Calibration of a liquid xenon detector with $^{83}Kr^{m}$". In: *Physical Review C* 80 (2009), p. 045809.

[92] Edward W. Kolb and Michael S. Turner. *The Early Universe*. Addison-Wesley, 1990.

[93] Shinzou Kubota et al. "Dynamical behavior of free electrons in the recombination process in liquid argon, krypton, and xenon". In: *Physical Review B* 20.8 (Oct. 15, 1979), pp. 3486–3496.

[94] John Kwong. "The development of two-phase xenon dark matter detectors". PhD thesis. Princeton University, 2009.

[95] J. Kwong et al. "Scintillation pulse shape discrimination in a two-phase xenon time projection chamber". In: *Nuclear Instruments and Methods in Physics Research A* 612 (2010). arXiv:0908.0790, pp. 328–333.

[96] Nicole Larsen. *Personal communication*. 2016.

[97] Nicole A. Larsen. "An effective field theory analysis of the first LUX dark matter search". PhD thesis. Yale University, 2016.

[98] H.S. Lee et al. (KIMS Collaboration). "Search for low-mass dark matter with CsI(Tl) crystal detectors". In: *Physical Review D* 90 (2014).

[99] Benjamin W. Lee and Steven Weinberg. "Cosmological lower bound on heavy neutrino masses". In: *Physical Review Letters* 39.4 (July 25, 1977), pp. 165–168.

[100] Brian Lenardo. *Personal communication*. 2017.

[101] J.D. Lewin and P.F. Smith. "Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil". In: *Astroparticle Physics* 6 (1996), pp. 87–112.

[102] Juan Magaña and Tonatiuh Matos. "A brief review of the scalar field dark matter model". In: *Journal of Physics: Conference Series* 378 (2012), p. 012012.

[103] Aaron Manalaysay (for the LUX Collaboration). "Dark-matter results from 332 new live days of LUX data". In: *Identification of Dark Matter, Sheffield, UK*. July 21, 2016.

[104] Maxim Markevitch. "Constraints on phsyics of gas and dark matter from cluster mergers". In: *6 Years of Science with Chandra*. Nov. 4, 2005.

[105] Stephen P. Martin. *A supersymmetry primer*. arXiv:hep-ph/9709356v7. Jan. 27, 2016.

[106] D.-M. Mei et al. In: *Astroparticle Physics* 30 (2008), pp. 12–17.

[107] David Merritt et al. "Empirical models for dark matter halos. I. Nonparametric construction of density profiles and comparison with parametric models". In: *Astronomical Journal* 132 (Dec. 2006), pp. 2685–2700.

[108] J. Mock et al. "Modeling pulse characteristics in xenon with NEST". In: *Journal of Instrumentation* 9 (Apr. 3, 2014). arXiv:1310.1117, T04002.

[109] B.J. Mount et al. (LZ Collaboration). *LUX-ZEPLIN (LZ) technical design report*. Tech. rep. LBNL-1007256. arXiv:1703.09144. Lawrence Berkeley National Laboratory, Mar. 27, 2017.

[110] Curt Nehrkorn. *Personal communication*. 2017.

[111] NERSC. *About NERSC*. `http://www.nersc.gov/about/`. 2017.

[112] NERSC. *Storage by scientific discipline*. `http://www.nersc.gov/users/storage-and-file-systems/hpss/storage-statistics/scientific-discipline/`. 2017.

[113] NERSC. *What is NEWT?* `https://newt.nersc.gov/`.

[114] Jan H. Oort. "The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems". In: *Bulletin of the Astronomical Institutes of the Netherlands* 6.238 (Aug. 17, 1932), pp. 249–287.

[115] Kevin O'Sullivan. *Personal communication*. 2014.

[116] Antonio Palazzo. "Phenomenology of light sterile neutrinos: A brief review". In: *Modern Physics Letters A* 28 (Feb. 26, 2013), p. 1330004.

[117] David Parker. *The redshift of extragalactic nuclei*. Translation of Zwicky's 1933 paper. 2014.

[118] C. Patrignani et al. "Review of Particle Physics". In: *Chinese Physics C* 40 (2016), p. 100001.

[119] P.J.E. Peebles. "Growth of the nonbaryonic dark matter theory". In: *Nature Astronomy* 1.0057 (Mar. 2, 2017).

[120] Troy A. Porter and Simona Murgia (for the Fermi-LAT Collaboration). *Observations of high-energy gamma-ray emission toward the galactic centre with the Fermi Large Area Telescope*. arXiv:1507.04688. July 16, 2015.

[121] Joel R. Primack, David Seckel, and Bernard Sadoulet. "Detection of cosmic dark matter". In: *Annual Review of Nuclear and Particle Science* 38 (1988), pp. 751–807.

[122] Stefano Profumo. "TASI 2012 Lectures on Astrophysical Probes of Dark Matter". In: *Theoretical Advanced Study Institute in Elementary Particle Physics*. University of Colorado at Boulder, June 2012.

[123] Sobrah Rahvar. "Cooling in the universe". In: *Iranian Journal of Physics [submitted]* (2006).

[124] Surjeet Rajendran. "Detecting ultra-light dark matter with precision metrology". In: *UC Berkeley Physics Colloquium*. Berkeley, CA, 2014.

[125] John P. Ralston. *One model explains DAMA/LIBRA, CoGENT, CDMS, and XENON*. arXiv:1006.5255. June 28, 2010.

[126] J. Rich. "Direct detection of particle dark matter". In: *Dark Matter: Proceedings of the XXIIIrd Recontre de Moriond; Moriond Astrophysics Meetings*. Les Arcs, Savoie, France, Mar. 8-15, 1988.

[127] Leszek Roszkowski. "Particle dark matter – A theorist's perspective". In: *Pramana* 62.2 (Feb. 2004), pp. 389–401.

[128] Vera C. Rubin and Jr. W. Kent Ford. "Rotation of the Andromeda Nebula from a spectroscopic survey of emission regions". In: *The Astrophysical Journal* 159 (Feb. 1970), pp. 379–403.

[129] Vera C. Rubin, Jr. W. Kent Ford, and Norbert Thonnard. "Rotational properties of 21 Sc galaxies with a large range of luminosities and radii, from NGC 4605 (R = 4 kpc) to UGC 2885 (R = 122 kpc)". In: *Astrophysical Journal* 238 (June 1, 1980), pp. 471–487.

[130] Barbara Ryden. *Introduction to Cosmology*. Addison Wesley, 2003.

[131] Bernard Sadoulet. "Deciphering the nature of dark matter". In: *Review of Modern Physics* 71.2 (1999), S197–S204.

[132] Robert Schabinger. "Minimal spontaneously broken hidden sector and its impact on Higgs boson physics at the CERN Large Hadron Collider". In: *Physical Review D* 72 (2005).

[133] David Smith and Neal Weiner. "Inelastic dark matter". In: *Physical Review D* 64 (2001), p. 043502.

[134] Peter Sorensen and Carl Eric Dahl. "Nuclear recoil energy scale in liquid xenon with application to the direct detection of dark matter". In: *Physical Review D* 83 (Mar. 1, 2011), p. 063501.

[135] N. Suzuki et al. "The Hubble Space Telescope cluster supernova survey. V. Improving the dark-energy constraints above z>1 and building an early-type-hosted supernova sample". In: *Astrophysical Journal* 746.1 (Feb. 10, 2012), p. 85.

[136] M. Szydagis et al. "Enhancement of NEST capabilities for simulating low-energy recoils in liquid xenon". In: *Journal of Instrumentation* 8 (Oct. 8, 2013), p. C10003.

[137] Andi Tan et al. (PandaX-II Collaboration). "Dark matter results from first 98.7 days of data from the PandaX-II experiment". In: *Physical Review Letters* 117 (Sept. 16, 2016), p. 121303.

[138] Will Taylor. *Personal communication*. 2016.

[139] Mark Trodden and Sean Carroll. "TASI Lectures: Introduction to Cosmology". In: *Theoretical Advanced Study Institute in Elementary Particle Physics*. University of Colorado at Boulder, 2002 and 2003.

[140] Roberto Trotta et al. "The impact of priors and observables on parameter interfaces in the constrained MSSM". In: *Journal of High Energy Physics* 12 (Dec. 5, 2008), p. 024.

[141] K. Ueshima et al. (XMASS Collaboration). "Scintillation-only based pulse shape discrimination for nuclear and electron recoils in liquid xenon". In: *Nuclear Instruments and Methods in Physics Research A* 659 (2011), pp. 161–168.

[142] J.D. Vergados et al. *Light WIMP searches involving electron scattering*. arXiv:1605.05413. May 18, 2016.

[143] M.I. Vysotskiĭ, A.D. Dolgov, and Ya.B. Zel'dovich. "Cosmological limits on the masses of neutral leptons". In: *JETP Letters* 26.3 (Aug. 5, 1977), pp. 188–190.

[144] Andrea Zani. "The WArP experiment: A double-phase argon detector for dark matter searches". In: *Advances in High Energy Physics* (2014), p. 205107.

[145] Fritz Zwicky. "Die Rotverschiebung von extragalaktischen Nebeln". In: *Helvetica Physica Acta* 6 (1933), pp. 110–127.