THE UNIVERSITY OF CHICAGO

CAMERAS IN THE DARK: DIRECT DETECTION OF DARK MATTER WITH CHARGE-COUPLED DEVICES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
RYAN THOMAS

CHICAGO, ILLINOIS
AUGUST 2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost I want to thank my advisor Paolo Privitera. He has been incredibly helpful and supportive throughout the past six years. He has made a challenging experience vastly easier through his advice and council, and I am forever grateful for the opportunity to work for and with him. I would also like to thank my committee members, Dan Hooper, Juan Collar, and Philippe Guyot-Sionnest.

I could not have done this work without the help and support of others work on the DAMIC project at UChicago: Karthik, Radomir, Dan, Ariel, Mircea, and all the others who have helped and worked with me throughout the years. Likewise I would like to thank those in DAMIC outside Chicago: Alvaro, Joao, and Romain especially. I owe a great debt to Richard Saldanha, for his work on the tritium activation measurement.

To all my fellow graduate students in my year, you have made a difficult journey easier. I would most especially thank Lipi, who has shown me more love and support than I sometimes feel I deserve, and who brought great joy and happiness to my life. I would also like to thank my office mates and fellow scientists in KICP, especially Evan, Aaron, John, and Joel, who have been good friends throughout this journey. My gratitude extends to all the KICP and physics support staff.

Finally, I would like to thank my family, who have always shown me love and care. My parents Ron and June, without whom I would not be here today, and my sister Megan who knows well the stresses higher education, and especially my aunt Diane, who has shown enormous support and interest in my work.

ABSTRACT

There are compelling motivations for the existence of dark matter particles which would explain the invisible gravitational mass present throughout the universe. Many direct-detection experiments have been built to look for these particles. All such searches so far have produced a null result, finding no conclusive sign of the elusive dark matter (although several have found hints of potential signals). Interest in the field has increasingly turned to searches for lighter dark matter candidates with mass $<10\text{GeV/c}^2$, which deposit less energy in the detector and are therefore difficult to see with most detectors.

The Dark Matter in CCDs (DAMIC) experiment has advanced the field of such light dark matter searches through the use of silicon charge-coupled devices (CCDs). CCDs combine high spatial resolution (with 15μm pixels) and low noise to provide a unique detector for multiple candidates for dark matter. The current generation of the experiment located at SNOLAB has already been used to look for several interesting varieties of dark matter. The next generation, DAMIC-M at the Modane Underground Laboratory in Modane, France will utilize a large array of new skipper CCDs, capable of achieving sub-electron noise resolution by using multiple non-destructive charge measurements, to search for multiple possible dark matter candidates.

Building such a large scale CCD detector comes with a number of significant challenges. One major challenge is simply operating many CCDs: existing CCD readout systems are not optimized for either skipper CCDs or their use as dark matter detectors, and scale poorly to large arrays in any case. A second challenge is understanding the backgrounds in silicon-based detectors. Such detectors have not only external backgrounds from radioisotope contamination in the detector assembly components, but also can contain radioactive tritium from cosmic ray neutron spallation of the silicon in the final fabricated sensor, a background peculiar to all silicon-based dark matter detectors.

The first part of this work describes the development of new high performance low noise electronics for CCD readout and control, the Online Digital Interface for Low Noise Electronics or "ODILE" system. We first describe the firmware used for these ODILE controllers, and follow this with a series of tests integrating the ODILE development

with skipper CCDs to achieve noise of $\ll 1\ e^-$ RMS/pixel. Use of these custom ODILE controllers will allow flexible and reliable operation of the $\mathcal{O}(100)$ CCDs necessary for DAMIC-M, running these CCDs in a steady state condition for months at a time to conduct extensive dark matter searches. The noise of the current prototype system has improved over the noise of commercially-available scientific CCD readout systems, with the readout electronics providing negligible noise compared to the intrinsic noise of the CCD itself.

In the second part of this work, we describe the development and validation of Monte Carlo simulations of the DAMIC at SNOLAB detector, which have been used to build an extensive model of its dark matter search backgrounds. This model was used to set limits on weakly interacting massive particle ("WIMP") dark matter cross sections in the 1-10 GeV c^2 mass region; an excess over the known sources of background was found, well-parameterized by an exponential with characteristic energy $\epsilon = 67 \pm 37$ $\mathrm{eV_{ee}}$. These simulations were also essential to measuring the bulk radioactive contamination of the silicon used in DAMIC CCDs, setting limits on the inherent contamination of silicon. Both of these results will help guide the development of the DAMIC-M detector, through a better understanding of our existing backgrounds and providing an interesting unknown excess to explore further.

Finally, we describe a novel measurement of the rate of activation of ^{3}H in silicon by cosmic ray neutron spallation. This measurement used a neutron beam at the LANSCE ICEHOUSE facility to irradiate silicon CCDs. The resulting tritium activation in these CCDs was converted into a measurement of the sea-level cosmic-ray neutron activation of silicon: 112 ± 24 atoms/kg/day. This is the first time such activation has been measured in silicon, and sets important constraints on the exposure to surface cosmic rays that any silicon-based dark matter detector can have. With this result, we can be assured that DAMIC-M will meet its ^{3}H background budget of 0.1 event/kg/keV/day in our region of interest (so long as we maintain less than 2 months of effective surface time during the production of the CCDs).

CHAPTER 1
OVERVIEW OF DARK MATTER AND DIRECT-DETECTION EXPERIMENTS

1.1 Introduction

In the past several decades, improved astrophysical and cosmological measurements of galactic rotations and the cosmic microwave background have shown that a significant fraction of the mass-energy in our universe cannot be explained by the baryonic matter contained in the Standard Model of particle physics. Most of this extra energy (~68% [1]) is in the form of so-called "dark energy", which seems to be a uniform constant throughout the universe and causes a repulsion that drives an accelerating expansion of the universe. Of the ~32% of the mass-energy that remains, only around 5% [2] seems to be the ordinary visible (or *baryonic*) matter we see every day. The remaining 27% is of an unknown nature, which doesn't seem to interact with light, and has therefore been named *dark matter*.

1.2 Evidence for Dark Matter

1.2.1 Galactic Rotation Curves

Beginning as far back as the 1940s [3], observations of astrophysicists like Jan Oort indicated that the mass distribution arising from gravitation kinematics (i.e. the orbital velocity of stars for a given galactic radius) did not follow the observed distribution of mass seen in visible light from star: the orbital velocity of stars should fall off farther from the center of the galaxy, following a so-called "Keplerian" profile (which follows a $1/\sqrt{r}$ dependency). It did not, and instead became nearly constant outside a certain radius (as seen in the curve for galaxy NGC 3198, shown in fig. 1.1 (left)). Questions about experimental uncertainties became settled when much more sensitive studies from Vera Rubin and Kent Ford [4] in the 1970s showed this to be a common phenomenon of many different galaxies (as seen in fig. 1.1 (right)).

Two explanations immediately spring to mind: either our theory of gravity is wrong (exhibiting some deviation from Newtonian models at galactic scales, the so-called "MOND" theory [5]), or there is a component of matter in these galaxies that does not emit or interact with visible light at all. This is termed *dark matter* (or "DM", for short).

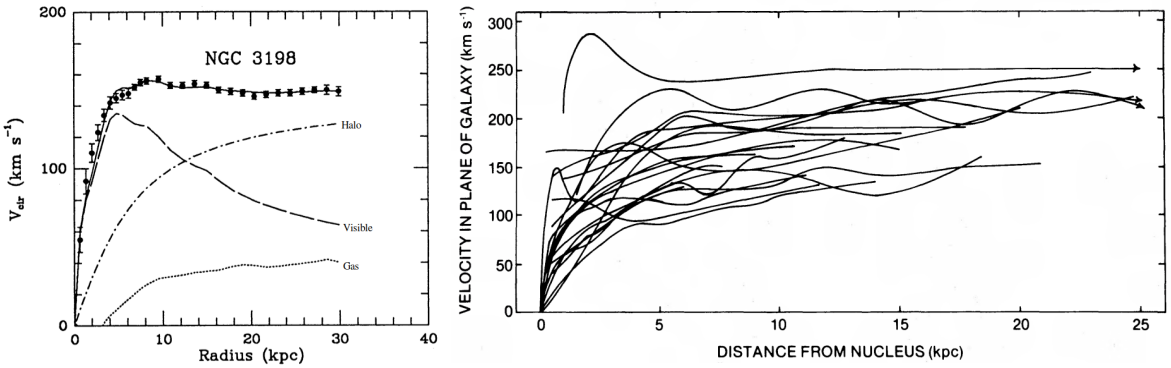


Figure 1.1: Left: An example of the galactic rotation curve for a spiral galaxy (data points), showing best-fit results for visible component (dashed), gas (dotted), and dark halo (dash-dot), and sum of all three (solid). Taken from Begeman et al. [6]. The visible component shows the expected Keplerian mass density profile from stars. **Right:** combined overlay of all 21 Sc galactic rotation curves measured in Rubin et al. [4].

1.2.2 CMB

The best measurements of the overall cosmological density of dark matter come from observations of the cosmic microwave background: leftover "relic" radiation from when the temperature of the universe dropped below that required to ionize hydrogen, and photons began to stream freely. This leaves a nearly-uniform, nearly perfect black body spectrum that has since redshifted into the microwave spectrum. Examinations of tiny ($\sim 10^{-6}$) [7] fluctuations in this spectrum (shown in fig. 1.2) allows a reconstruction of the large-scale structure of the universe. When combined with a model, such as the standard ΛCDM (which assumes the universe is dominated by dark energy + cold-dark matter), it provides a measurement of the absolute density of dark matter and baryonic matter in our universe. These observations give a ratio of dark to baryonic matter of 5.35.

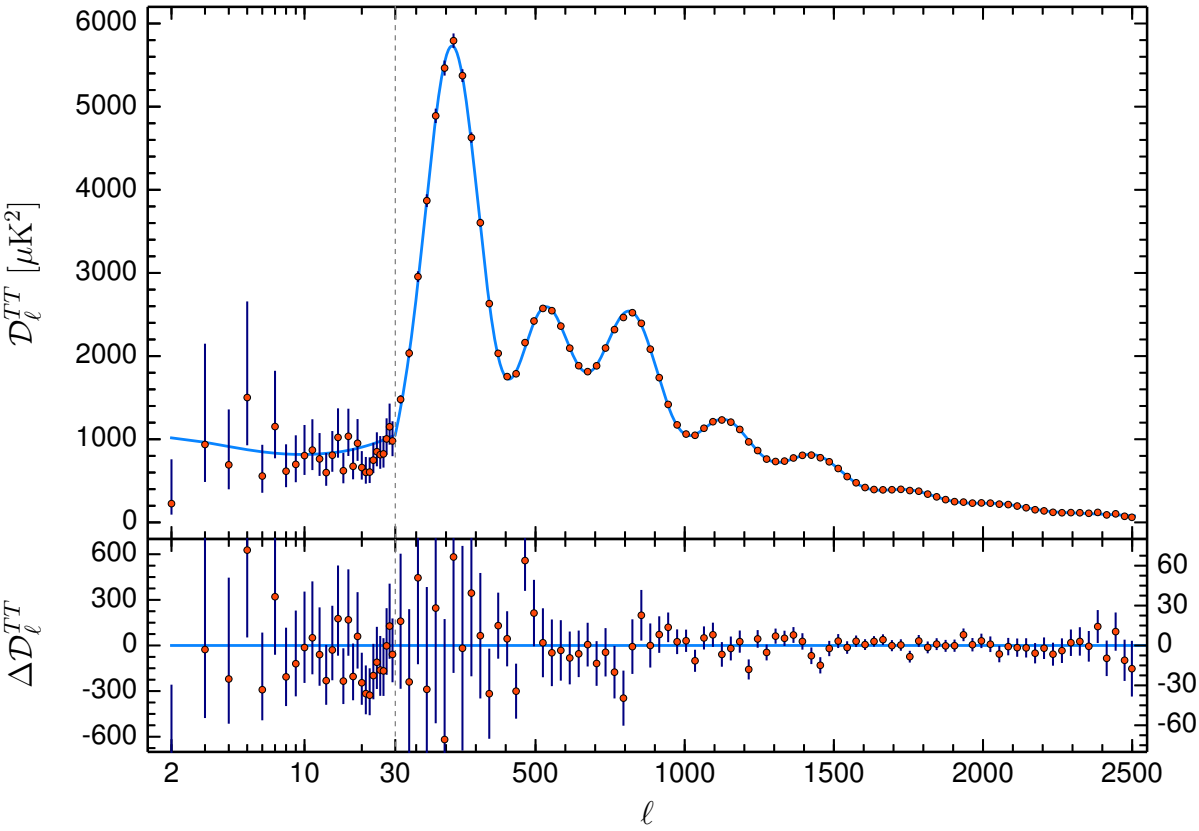


Figure 1.2: Results of the 2018 Planck satellite observation of the cosmic microwave background, with best fit results. The best fit parameters give a dark matter density $\Omega_c h^2 = 0.120 \pm 0.001$, baryon density of $\Omega_b h^2 = 0.0224 \pm 0.0001$. Figure and results from Aghanim et al. [2].

1.2.3 Lensing

One of the strongest evidence for DM as a form of matter (rather than a product of incomplete understanding of gravity such as MOND) has come from the observations of lensing produced by large galaxy clusters. These large galaxy clusters have sufficient mass to bend the light from sources behind them. The degree of this lensing is proportional to the gravitational mass of the cluster, allowing a direct measurement of the mass and mass distribution in these clusters. One of the most famous galaxy clusters is 1E 0657-56, aka the "Bullet Cluster". Observations of this cluster by Clowe et al. [8] and Markevitch et al. [9] showed that the gravitational mass of the cluster was displaced from the visible mass of the cluster, picture in fig. 1.3. This result is easily allowed by a dark matter model, but strongly contradicts most MOND models (which still require gravitational and visible mass to coincide).

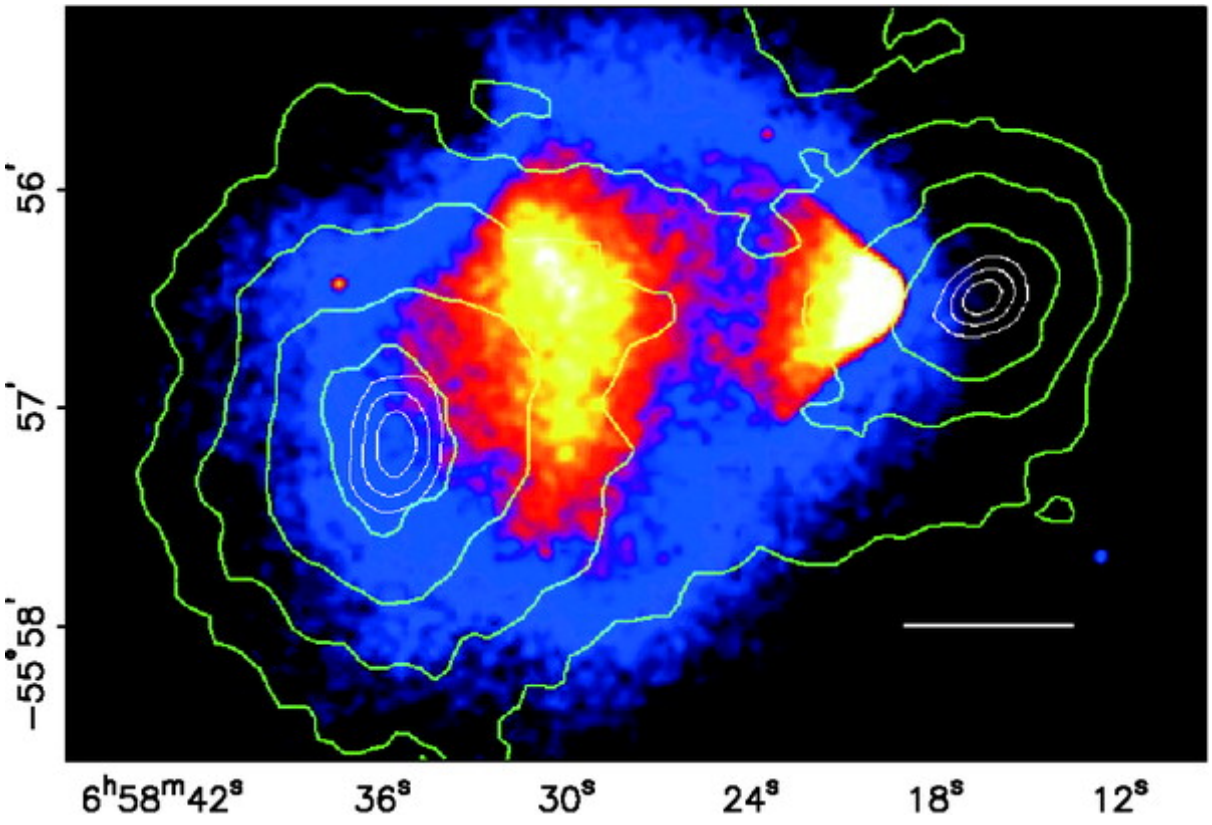


Figure 1.3: Weak lensing reconstruction of bullet cluster mass density (green contours) overlayed on the Chandra observatory image of the same cluster. Figure taken from [8].

1.3 What dark matter could be

A nearly universal property of all dark matter theories is that there exists, at some level, some non-gravitational coupling between the dark matter particles and standard baryonic matter. There is, as of yet, no direct evidence such a coupling exists. Any evidence for such coupling is usually model dependent. One general strong reason to believe such a coupling exists is that the dark matter mass density is within an order of magnitude of baryonic matter, and seems to have a "temperature" (i.e. mean velocity) reasonably close to that of visible matter. This is commonly referred to as *cold* (non-relativistic) dark matter, and is typically required for dark matter to form the halos seen from galactic rotation curve observations (if the dark matter had relativistic velocities, it would escape from the halo).

This suggests that baryonic and dark matter were at some point in the history of the universe coupled together, causing the density and thermal histories to also become coupled. Unfortunately, the existence of this coupling tells us very little about it's nature

or strength, and fails to constrain the DM particle mass (or even the number of possible species of DM[1]). The remainder of this discussion will follow the simplest assumption, that the universe is dominated by a single species of DM particle.

There exist a multitude of different theoretical dark matter models, ranging in mass from the so-called "ultralight" dark matter (mass $\ll eV/c^2$), including various QCD axion or axion-like particles, all the way to solar-mass objects (typically called massive compact halo objects or "MACHOs"). A review of many popular models can be found in [11].

Reasonably strict model-independent bounds exist for the mass of these particles, on both the lowest and highest possible. For fermionic DM particles, the highest possible number density is bounded by the available states in phase-space [12], $f < gh^{-3}$ (where g is the degrees of freedom). For the observed mass densities, this gives a lower bound on the mass of the particle of $m_f > 70$eV [13]. For bosonic DM, this limit does not exist, but we can instead note that the de Broglie wavelength for typical velocities is $\lambda \sim (eV/m)$ [14], and require that this be less than the size of dwarf galaxies (roughly a kiloparsec) to get a lower limit $m_B \gtrsim 10^{-22}$ eV.

Upper bounds on the mass are similarly weak: the primary limit is the requirement that the DM be sufficiently "fluid": we would certainly have noticed gravitational anomalies if 4/5th of the mass of the galaxy was contained only in a few thousand objects. Observations of wide halo binaries gives a limit of $M_{DM} < 5$ stellar masses [15] for MACHOs.

A few other observations can constrain the properties of dark matter. First it must be sufficiently "dark" to have escaped notice from astrophysical observations. Observations of the cosmological recombination era gives a charge for $m_\chi = 10$TeV of $< 10^{-4}$ of the electron charge [16]. Secondly, the particle must have a half-life at least comparable to the age of the universe[17]. Finally, the DM-DM self-interactions are limited by astrophysical observations [18].

A few popular dark matter models are described briefly below.

1. There has been much interest in recent years in the possibility of entire "hidden" dark matter sectors, possibly with as much or even more particle types and couplings as our own visible sector[10].

1.3.1 WIMPS

Probably the strongest dark matter candidate (at least in terms of popularity) has been the weakly-interacting massive particle (or "WIMP" for short). Aside from the catchy name, this particle was originally incredibly exciting due to the the so-called "WIMP miracle": a minimal supersymmetric extension to the Standard Model with a weak-scale annihilation cross section and a mass in the 10-1000 GeV/c^2 range creates a particle with a relic abundance that matches nearly exactly that expected from dark matter! [19] Unfortunately a true weak-scale supersymmetric WIMP has been ruled out by direct detection experiments decades ago, but more complex (non)-supersymmetric WIMP models (such as a particle that couples through a Higgs or Z portal) are still possible [20], and have decent theoretical motivation.

1.3.2 Axion(-like) Particles

Another popular extension to the Standard Model is the axion (or axion-*like*) particle (ALP for short). The QCD axion was originally proposed to solve the strong CP problem [21], and also doubles as a possible dark matter candidate [22]. In addition to providing an easy extension to the SM, axion-like models based on a very light (pseudo)scalar field potentially solve some of the observed small-scale issues with dark matter. It is also popular because the QCD axion models give a relation between their mass and coupling constants, allowing for definite (but model-dependent) targets for experimental searches [23, 24]. Axions have their own dedicated class of direct detection experiments such as ADMX [25], due to their unique detection methodology. However generic MeV scale ALPs can deposit energy through absorption by electrons, allowing limits to be set by more standard searches.

1.3.3 Dark Photons

The dark (or hidden) photon is a proposed massive (but stable) vector boson that has some kinetic mixing angle ϵ with the visible photon. A dark photon is present in many

extensions to the Standard Model (offering some theoretical basis for its existence), and has been proposed as possible dark matter candidate [26, 27]. The kinematic mixing of hidden photons with regular photons would allow for their production in the sun, which allows limits to be placed based on the (non)observation of solar events or loss of solar power into this dark photon.

1.4 What We Look for

There are usually two primary categories of dark matter detectors: indirect detectors, that look for products of the DM interaction with either themselves or other particles elsewhere in the universe, and direct detectors, which look for direct interaction of the dark matter with the detector itself. The former include both astrophysical searches (such as the searches for positron excess from the galactic center [28]) and collider searches that look for energy loss from DM production in the detector [29]. These searches are interesting in their own right and have produce highly useful limits to many models of DM, but in this work, we will focus exclusively on the direct detection mechanisms.

The concept behind direct detection is fairly straightforward. The density of dark matter in the local galactic neighborhood can be inferred from examinations of the halo profile of our galaxy. Data from the *Gaia* satellite place this at between 0.4 and 1.5 GeV/cm^3 [30]. These DM particles traveling with a velocity v can scatter off the detector and impart some recoil energy $E_R \propto \Delta v^2$. If this energy is greater than the threshold of the detector, this can produce an observable signal.

The dark matter is typically assumed to follow a Maxwell-Boltzmann velocity distribution

$$f(v, v_E) = e^{-(v+v_E)^2/v_0^2} \tag{1.1}$$

where v is the velocity of the dark matter in the local galactic halo, v_E is the velocity of the Earth with respect to the DM rest frame , and v_0 is the most probable velocity. This distribution is typically truncated, so that $v + v_E < v_{esc}$, where v_{esc} is the galactic escape velocity in Earth's vicinity (this is around 533 km/s [31]), as any particles which

exceed this velocity would escape the galaxy.

The differential particle density per unit velocity near Earth is then given by

$$dn = \frac{n_0}{k} f(v, v_E) d^3v \tag{1.2}$$

where n_0 is the mean dark matter particle density number (ρ_χ / m_χ), and k is the normalization constant such that $\int_0^{v_{esc}} dn = n_0$. We will assume $k = (\pi v_0^2)^{3/2}$ (this is the normalization for $v_{esc} = \infty$, but the correction for finite v_{esc} is $< 1\%$ [32]).

1.4.1 DM-Nucleus Scattering

For a full review of the mathematics of dark matter scattering off nuclei, see Lewin and Smith [32] and Jungman et al. [19]. The most relevant results are summarized below.

We can give the differential rate for a target nucleus of mass m_A as

$$dR = \frac{1}{m_A} \sigma_A v dn \tag{1.3}$$

where σ_A is the cross-section per nucleon. The recoil energy E_R of the nucleus can be written

$$E_R = \frac{v^2 m_r^2 (1 - cos\theta)}{m_A} \tag{1.4}$$

where $m_r = m_A m_X / (m_A + m_\chi)$ is the reduced mass of the nucleus-DM system. This scattering is typically assumed to be isotropic (uniform in θ). Note that using this equation we can immediately write the maximum nuclear recoil energy E_{max} created by incident DM particles,

$$E_{max} = \begin{cases} 400 \text{ keV} \left(\frac{m_A}{28 \text{ GeV}}\right) & m_\chi \gg m_A \\ 50 \text{ eV} \left(\frac{m_\chi}{1 \text{ GeV}}\right)^2 \left(\frac{28 \text{ GeV}}{m_A}\right) & m_\chi \ll m_A \end{cases} \tag{1.5}$$

where we have scaled the detector mass to the mass of silicon (28 GeV/c^2). These equations illustrate the importance of low-thresholds on dark matter searches: a 1 GeV dark matter particle can only impart a max energy of 50 eV in silicon, which is below

many practical direct detection thresholds.

We can then write the differential scattering rate in a detector with N_T target nuclei, as a function of the nuclear recoil energy E_R

$$\frac{dR}{dE_R} = N_T \frac{\rho_\chi}{k m_\chi} \int_{v_{min}}^{v_{esc}} d^3v f(v, v_E) v \frac{d\sigma(E_R, v)}{dE_R} \tag{1.6}$$

where v_{min} is the smallest velocity that can impart a recoil energy E_R ($v_{min} = (m_A E_R / 2m_r^2)^{\frac{1}{2}}$), and $\frac{d\sigma(E_R,v)}{dE_R}$ is the differential cross section.

This differential cross-section is model-dependent. If we assume a standard leading order spin-independent coupling proportional to the inverse square of the DM velocity, we can write

$$\frac{d\sigma(E_R, v)}{dE_R} = \frac{m_A}{2m_r^2 v^2} \sigma_0 F^2(q) \tag{1.7}$$

where σ_0 is the scattering cross section at zero momentum, and $F^2(q)$ is the (momentum dependent) nuclear form factor. For an isotope with atomic number Z and mass number A, the zero-momentum cross section is

$$\sigma_0 = \frac{4m_r^2}{\pi} [Z f^p + (A - Z) f^n]^2 \tag{1.8}$$

where f^p and f^n are the couplings to proton and neutron. Most experiments assume those couplings are the same, in which case the zero-momentum cross section simplifies to $\frac{4}{\pi} m_r^2 A^2 \sigma_{wn}^2$. If we combine this with eq. (1.6), we can note that our cross sectional rate becomes

$$\frac{dR}{dE_R} = N_T \frac{2 m_A \rho_\chi}{k \pi m_\chi} A^2 \sigma_{wn}^2 F^2(q) \int_{v_{min}}^{v_{esc}} d^3v \frac{f(v, v_E)}{v} \tag{1.9}$$

A plot of this rate for several typical target nuclei can be found in fig. 1.4. This equation contains a few notable features. First, $N_T m_A$ is the mass of our entire detector: as expected, our rate scales linearly with total detector mass. Second, the rate is generally inversely proportional to the DM mass, m_χ, due to the fixed mass density resulting in a decreased number density at higher masses, and vice versa. In practice, however, the integral over the kinematic term results in a significant decrease in rate for low mass

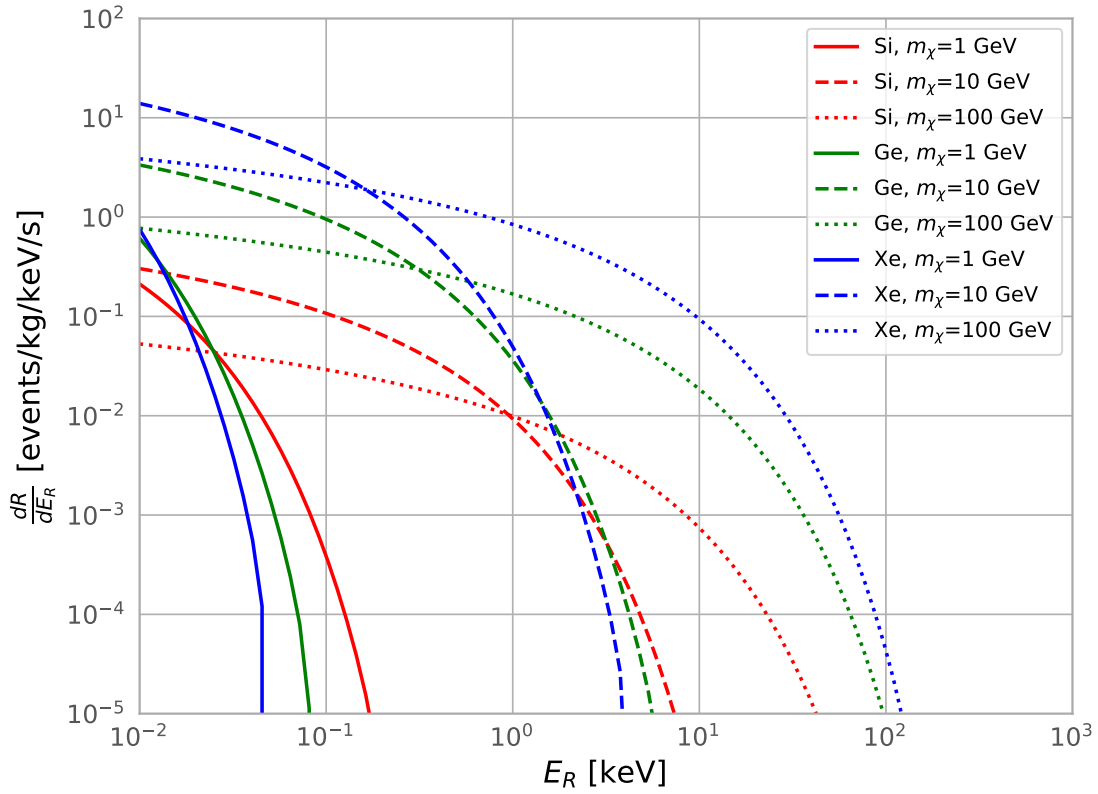


Figure 1.4: Differential scattering rate for silicon (red), germanium (green), and xenon (blue) targets, for WIMP masses of 1 (solid), 10 (dashed), and 100 (dotted) GeV/c^2. Plot is made with the assumption $F^2(q) = 1$, $\sigma_{wn} = 10^{-36}$ and shows only the kinematic, mass, and A^2 dependency on scattering rates.

WIMPs, since v_{min} is higher for lower mass WIMPS. Finally, for low mass WIMPs, lighter nuclei will have on average higher recoil energies, which makes detection easier.

1.4.2 DM-Electron scattering

For most practical detector, the energy threshold makes DM-nucleus scattering difficult to see for DM particles $m_\chi \lesssim 1$ GeV. Searches for sub-GeV dark matter therefore typically look for DM-electron scattering. DM-electron scattering is of particular interest for searches for dark photons and axion(-like) particles, which contain kinetic mixing and axioelectric couplings (respectively) with the electron. Computation of the differential rate is complicated, as it involves interactions with bound-state electrons. We will therefore

simply quote the result for ionization from Essig et al. [33]

$$\frac{dR_{ion}}{d\ln E_R} = \frac{6.2}{A}\frac{\rho_\chi}{0.4\text{GeV cm}^{-3}}\frac{\sigma_e}{10^{-40}\text{cm}^2}\frac{10\text{MeV}}{m_\chi}\frac{d\langle\sigma_{ion}v\rangle/d\ln E_R}{10^{-3}\bar{\sigma}_e} \tag{1.10}$$

where $d\langle\sigma_{ion}v\rangle/d\ln E_R$ is the thermally averaged ionization cross section, and σ_e is the DM-electron elastic scattering cross section. The constants are given so that each term is $\mathcal{O}(1)$ for typical values of each parameter.

1.5 How We Look For Dark Matter

Nearly all dark matter direct detection experiments are constructed using the same basic guidelines. First, the detector must have sufficiently low background in the relevant energy region to distinguish any potential signal DM from the background. The exact backgrounds varies by search and detector, but a handful are quite common. Cosmic ray muons are a nearly universal problem, and require placing most experiments deep underground to shield against. Activation of detector material from cosmic ray neutrons, radon in the air producing ^{210}Pb contamination of detector surfaces, and neutrons from spontaneous fission or (α,n) reactions are all also fairly common[2]. Generally, detectors use extensive shielding and materials carefully chosen and assayed for radioactive backgrounds. Polyethylene shielding against neutrons, ancient lead[3] recovered from shipwrecks for gamma rays, and even electroformed copper for inner parts are all common choices. Background rates are frequently given in the so-called "differential rate unit" (d.r.u.), defined as "events/keV/kg/day" (typical background rates for DM experiments are 0.01-10 d.r.u.).

Second, the detector is usually as large as possible, given the constraints of practicality, to maximize the total number of potential DM interactions. The practical size of a

2. Ironically placing detectors underground tend to make the latter two problems much worse, since underground labs tend to contain large amounts of uranium in the walls.

3. Naturally mined lead contains significant amounts of uranium, which decays into ^{210}Pb. The uranium is removed in the refining process, but the ^{210}Pb is not. Lead mined recently therefore contains very high amounts of ^{210}Pb, resulting in a significant background. Old lead has had time for the residual ^{210}Pb to decay away, and is therefore much more radiopure than modern lead.

detector is constrained by the background rate: backgrounds typically scale linearly with the detector size,[4] resulting in no signal/background gains from increased mass.

Third, the detector is usually operated for long periods of time, since expected DM rates are only tens of events per year. One year of operation is fairly standard for most searches. The operational time is tightly coupled to the mass of the detector, to the point where the exposure of most DM searches is typically given in mass·time, since this is the parameter of interest for the total number of expected interactions.

1.5.1 Types of Direct Detection Experiments

There exist a large range of different detectors looking for dark matter. Most of them focus on WIMP-nucleon scattering, but there has been a push in recent years to develop new technologies to look for other varieties of dark matter. A brief overview of some of the most common detectors is given below.

CCD Detectors. The focus of this work, CCD detectors use solid-state silicon charge-coupled devices (CCD) to look for ionization from dark matter interactions. DAMIC [34] and SENSEI [35] both use CCDs for dark matter searches. Silicon CCDs offer extremely low (∼10eV) energy thresholds and excellent (∼15μm spatial resolution), but have relatively low detector and atomic mass, making them primarily useful for low-mass WIMP.

Cryogenic Solid State Detectors. These experiments use either semiconductors [36] or scintillating crystals [37] to search for a combination of photons, phonons, and ionization from dark matter scatters. By using multiple detection channels, these experiments can often perform electron recoil/nuclear recoil discrimination, which makes them excellent at rejecting electromagnetic backgrounds (such as Compton scatters or beta decays). Unfortunately the efficiency for this discrimination tends to fall rapidly at low energies, and the difficulty of operating large detectors at cryogenic temperatures limits the mass that can be used in such detectors.

4. This is not always true: for example, noble liquid experiments benefit from a self-shielding effect. This means external backgrounds tend to scale as the surface area of the detector, while the effective mass scales as the volume of the detector, producing a signal to background gain from increasing the mass.

Noble Liquids. Offering the best overall limits on dark matter, noble liquid detectors such as XENON1T [38], LUX [39], and DARKSIDE [40] use a noble liquid (currently either xenon or argon, although helium and neon detectors have been proposed) in a time-projection chamber (TPC) to search for dark matter. Being a noble liquid makes the material much easier to purify and scale, and the nature of xenon and argon allows for electronic/nuclear recoil discrimination using either multiple collection channels (using a two-phase TPC) for xenon, or pulse-shape discrimination for argon. The easy scalability, simple purification, and relative maturity of the noble liquid detector has resulted in them offering some of the best overall limits on DM across a large range of masses.

Superheated detectors. Superheated detectors, such as PICO-60 [41], use a superheated liquid inside a bubble chamber to search for dark matter. These detectors are excellent at searches for spin-dependent interactions, as they can use targets (typically fluorine) with an odd number of nucleons to maximize spin-dependent couplings. They also have insensitivity to a number of standard radioactive backgrounds (such as muons). However, they have poor energy resolution, as bubble formation occurs for any events above threshold, limiting their spectral sensitivity.

1.5.2 Limit Setting

In practice, no direct detection experiment has yet seen a significant (uncontroversial) dark matter signal. Experiments instead set limits on the possible mass/cross section of any dark matter particles, by comparing the rate observed to the predicted rate for a given mass/cross section. These limits are highly specific to the specific detector in question, as it depends on the detector mass, material, energy threshold, background, and exposure times. Mass, for example, can vary between gram-scale for low-mass, eV-scale threshold detectors such as CRESST [43] to tonne size/keV threshold Xenon detectors [38]. To compare the WIMP limits from various detectors, the results are converted to a scattering cross section per nucleon (effectively the cross section for a hydrogen atom), using pre-agreed DM velocity and density parameters. The results from various experiments can be seen in fig. 1.5 for WIMP-nucleon scattering. Figure 1.6 shows limits for hidden photons.

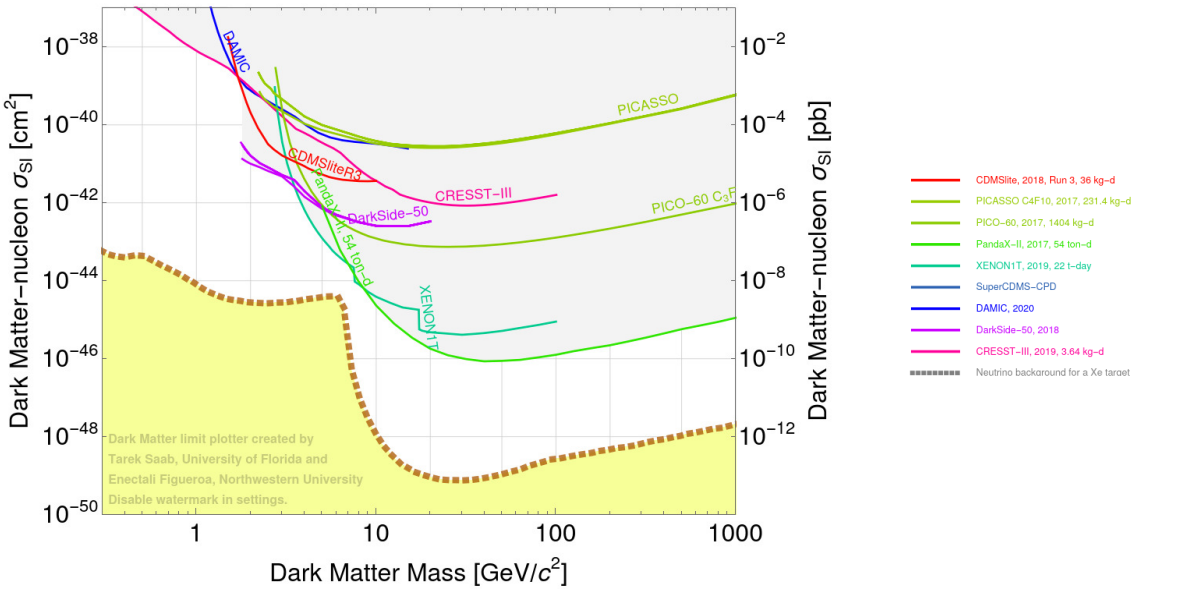


Figure 1.5: Selected spin-independent WIMP limits, from [42]. The limits shown focus on the low-mass region of the dark matter parameter space, as that is the region most relevant to this work.

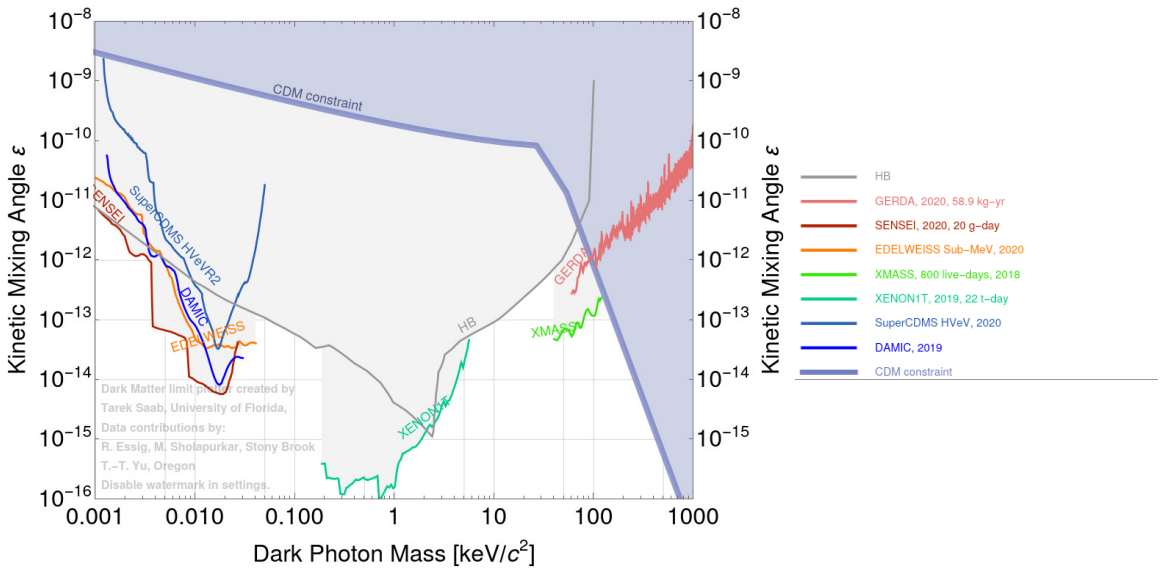


Figure 1.6: Hidden photon limits, from [42].

References

[1] T. M. C. Abbott, S. Allam, P. Andersen, C. Angus, J. Asorey, A. Avelino, S. Avila, B. A. Bassett, K. Bechtol, G. M. Bernstein, et al. First cosmology results using type ia supernovae from the dark energy survey: Constraints on cosmological parameters. *The Astrophysical Journal*, 872(2):L30, feb 2019. doi: 10.3847/2041-8213/ab04fa. URL `https://doi.org/10.3847/2041-8213/ab04fa`.

[2] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, et al. Planck 2018 results. *Astronomy & Astrophysics*, 641:A6, Sep 2020. ISSN 1432-0746. doi: 10.1051/0004-6361/201833910. URL `http://dx.doi.org/10.1051/0004-6361/201833910`.

[3] J. H. Oort. Some Problems Concerning the Structure and Dynamics of the Galactic System and the Elliptical Nebulae NGC 3115 and 4494. ApJ, 91:273, April 1940. doi: 10.1086/144167.

[4] V. C. Rubin, Jr. Ford, W. K., and N. Thonnard. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc). ApJ, 238:471–487, June 1980. doi: 10.1086/158003.

[5] M. Milgrom. A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis. ApJ, 270:365–370, July 1983. doi: 10.1086/161130.

[6] K. G. Begeman, A. H. Broeils, and R. H. Sanders. Extended rotation curves of spiral galaxies: dark haloes and modified dynamics. *Monthly Notices of the Royal Astronomical Society*, 249(3):523–537, 04 1991. ISSN 0035-8711. doi: 10.1093/mnras/249.3.523. URL `https://doi.org/10.1093/mnras/249.3.523`.

[7] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, et al. Structure in the COBE Differential Microwave Radiometer First-Year Maps. ApJ, 396:L1, September 1992. doi: 10.1086/186504.

[8] Douglas Clowe, Maruša Bradač, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *The Astrophysical Journal*, 648(2):L109–L113, aug 2006. doi: 10.1086/508162. URL `https://doi.org/10.1086/508162`.

[9] M. Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, W. Forman, C. Jones, S. Murray, and W. Tucker. Direct Constraints on the Dark Matter Self-Interaction Cross Section from the Merging Galaxy Cluster 1E 0657-56. ApJ, 606(2):819–824, May 2004. doi: 10.1086/383178.

[10] Dan Hooper, Rebecca K. Leane, Yu-Dai Tsai, Shalma Wegsman, and Samuel J. Witte. A systematic study of hidden sector dark matter: application to the gamma-ray and antiproton excesses. *Journal of High Energy Physics*, 2020(7), Jul 2020. ISSN 1029-8479. doi: 10.1007/jhep07(2020)163. URL `http://dx.doi.org/10.1007/JHEP07(2020)163`.

[11] Tongyan Lin. Tasi lectures on dark matter models and direct detection, 2019.

[12] Scott Tremaine and James E. Gunn. Dynamical role of light neutral leptons in cosmology. *Phys. Rev. Lett.*, 42:407–410, Feb 1979. doi: 10.1103/PhysRevLett.42.407. URL https://link.aps.org/doi/10.1103/PhysRevLett.42.407.

[13] Lisa Randall, Jakub Scholtz, and James Unwin. Cores in dwarf galaxies from fermi repulsion. *Monthly Notices of the Royal Astronomical Society*, page stx161, Jan 2017. ISSN 1365-2966. doi: 10.1093/mnras/stx161. URL http://dx.doi.org/10.1093/mnras/stx161.

[14] Particle Data Group, P A Zyla, R M Barnett, J Beringer, O Dahl, D A Dwyer, D E Groom, C J Lin, K S Lugovsky, E Pianori, et al. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), 08 2020. ISSN 2050-3911. doi: 10.1093/ptep/ptaa104. URL https://doi.org/10.1093/ptep/ptaa104. 083C01.

[15] Miguel A. Monroy-Rodríguez and Christine Allen. THE END OF THE MACHO ERA, REVISITED: NEW LIMITS ON MACHO MASSES FROM HALO WIDE BINARIES. *The Astrophysical Journal*, 790(2):159, jul 2014. doi: 10.1088/0004-637x/790/2/159. URL https://doi.org/10.1088/0004-637x/790/2/159.

[16] Samuel D. McDermott, Hai-Bo Yu, and Kathryn M. Zurek. Turning off the lights: How dark is dark matter? *Physical Review D*, 83(6), Mar 2011. ISSN 1550-2368. doi: 10.1103/physrevd.83.063509. URL http://dx.doi.org/10.1103/PhysRevD.83.063509.

[17] Benjamin Audren, Julien Lesgourgues, Gianpiero Mangano, Pasquale Dario Serpico, and Thomas Tram. Strongest model-independent bound on the lifetime of dark matter. *Journal of Cosmology and Astroparticle Physics*, 2014(12):028–028, Dec 2014. ISSN 1475-7516. doi: 10.1088/1475-7516/2014/12/028. URL http://dx.doi.org/10.1088/1475-7516/2014/12/028.

[18] D. Harvey, R. Massey, T. Kitching, A. Taylor, and E. Tittley. The nongravitational interactions of dark matter in colliding galaxy clusters. *Science*, 347 (6229):1462–1465, Mar 2015. ISSN 1095-9203. doi: 10.1126/science.1261381. URL http://dx.doi.org/10.1126/science.1261381.

[19] Gerard Jungman, Marc Kamionkowski, and Kim Griest. Supersymmetric dark matter. *Physics Reports*, 267(5-6):195–373, Mar 1996. ISSN 0370-1573. doi: 10.1016/0370-1573(95)00058-5. URL http://dx.doi.org/10.1016/0370-1573(95)00058-5.

[20] Leszek Roszkowski, Enrico Maria Sessolo, and Sebastian Trojanowski. Wimp dark matter candidates and searches—current status and future prospects. *Reports on Progress in Physics*, 81(6):066201, May 2018. ISSN 1361-6633. doi: 10.1088/1361-6633/aab913. URL http://dx.doi.org/10.1088/1361-6633/aab913.

[21] Steven Weinberg. A new light boson? *Phys. Rev. Lett.*, 40:223–226, Jan 1978. doi: 10.1103/PhysRevLett.40.223. URL https://link.aps.org/doi/10.1103/PhysRevLett.40.223.

[22] David J.E. Marsh. Axion cosmology. *Physics Reports*, 643:1–79, Jul 2016. ISSN 0370-1573. doi: 10.1016/j.physrep.2016.06.005. URL http://dx.doi.org/10.1016/j.physrep.2016.06.005.

[23] Jihn E. Kim. Weak-interaction singlet and strong CP invariance. *Phys. Rev. Lett.*, 43:103–107, Jul 1979. doi: 10.1103/PhysRevLett.43.103. URL https://link.aps.org/doi/10.1103/PhysRevLett.43.103.

[24] Michael Dine, Willy Fischler, and Mark Srednicki. A simple solution to the strong cp problem with a harmless axion. *Physics Letters B*, 104(3):199–202, 1981. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(81)90590-6. URL https://www.sciencedirect.com/science/article/pii/0370269381905906.

[25] T. Braine, R. Cervantes, N. Crisosto, N. Du, S. Kimes, L. J. Rosenberg, G. Rybka, J. Yang, D. Bowring, A. S. Chou, et al. Extended search for the invisible axion with the axion dark matter experiment. *Phys. Rev. Lett.*, 124:101303, Mar 2020. doi: 10.1103/PhysRevLett.124.101303. URL https://link.aps.org/doi/10.1103/PhysRevLett.124.101303.

[26] Haipeng An, Maxim Pospelov, Josef Pradler, and Adam Ritz. Direct detection constraints on dark photon dark matter. *Physics Letters B*, 747:331–338, Jul 2015. ISSN 0370-2693. doi: 10.1016/j.physletb.2015.06.018. URL http://dx.doi.org/10.1016/j.physletb.2015.06.018.

[27] Itay M. Bloch, Rouven Essig, Kohsaku Tobioka, Tomer Volansky, and Tien-Tien Yu. Searching for dark absorption with direct detection experiments. *Journal of High Energy Physics*, 2017(6), Jun 2017. ISSN 1029-8479. doi: 10.1007/jhep06(2017)087. URL http://dx.doi.org/10.1007/JHEP06(2017)087.

[28] M. Aguilar, L. Ali Cavasonza, G. Ambrosi, L. Arruda, N. Attig, P. Azzarello, A. Bachlechner, F. Barao, A. Barrau, L. Barrin, et al. Towards understanding the origin of cosmic-ray positrons. *Phys. Rev. Lett.*, 122:041102, Jan 2019. doi: 10.1103/PhysRevLett.122.041102. URL https://link.aps.org/doi/10.1103/PhysRevLett.122.041102.

[29] Stefano Giagu. Wimp dark matter searches with the atlas detector at the lhc. *Frontiers in Physics*, 7:75, 2019. ISSN 2296-424X. doi: 10.3389/fphy.2019.00075. URL https://www.frontiersin.org/article/10.3389/fphy.2019.00075.

[30] Jatan Buch, Shing Chau (John) Leung, and JiJi Fan. Using gaia dr2 to constrain local dark matter density and thin dark disk. *Journal of Cosmology and Astroparticle Physics*, 2019(04):026–026, Apr 2019. ISSN 1475-7516. doi: 10.1088/1475-7516/2019/04/026. URL http://dx.doi.org/10.1088/1475-7516/2019/04/026.

[31] T. Piffl, C. Scannapieco, J. Binney, M. Steinmetz, R.-D. Scholz, M. E. K. Williams, R. S. de Jong, G. Kordopatis, G. Matijevič, O. Bienaymé, et al. The rave survey: the galactic escape speed and the mass of the milky way. *Astronomy & Astrophysics*, 562:A91, Feb 2014. ISSN 1432-0746. doi: 10.1051/0004-6361/201322531. URL http://dx.doi.org/10.1051/0004-6361/201322531.

[32] J.D. Lewin and P.F. Smith. Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil. *Astroparticle Physics*, 6(1):87–112, 1996. ISSN 0927-6505. doi: https://doi.org/10.1016/S0927-6505(96)00047-3. URL https://www.sciencedirect.com/science/article/pii/S0927650596000473.

[33] Rouven Essig, Jeremy Mardon, and Tomer Volansky. Direct detection of sub-gev dark matter. *Physical Review D*, 85(7), Apr 2012. ISSN 1550-2368. doi: 10.1103/physrevd.85.076007. URL http://dx.doi.org/10.1103/PhysRevD.85.076007.

[34] A. Aguilar-Arevalo, D. Amidei, D. Baxter, G. Cancelo, B. A. Cervantes Vergara, A. E. Chavarria, J. C. D'Olivo, J. Estrada, F. Favela-Perez, R. Gaïor, et al. Results on low-mass weakly interacting massive particles from an 11 kg d target exposure of damic at snolab. *Phys. Rev. Lett.*, 125:241803, Dec 2020. doi: 10.1103/PhysRevLett.125.241803. URL https://link.aps.org/doi/10.1103/PhysRevLett.125.241803.

[35] Orr Abramoff, Liron Barak, Itay M. Bloch, Luke Chaplinsky, Michael Crisler, Dawa, Alex Drlica-Wagner, Rouven Essig, Juan Estrada, Erez Etzion, et al. Sensei: Direct-detection constraints on sub-gev dark matter from a shallow underground run using a prototype skipper ccd. *Physical Review Letters*, 122(16), Apr 2019. ISSN 1079-7114. doi: 10.1103/physrevlett.122.161801. URL http://dx.doi.org/10.1103/PhysRevLett.122.161801.

[36] R. Agnese, T. Aralis, T. Aramaki, I. J. Arnquist, E. Azadbakht, W. Baker, S. Banik, D. Barker, D. A. Bauer, T. Binder, et al. Search for low-mass dark matter with cdmslite using a profile likelihood fit. *Phys. Rev. D*, 99:062001, Mar 2019. doi: 10.1103/PhysRevD.99.062001. URL https://link.aps.org/doi/10.1103/PhysRevD.99.062001.

[37] CRESST collaboration, F. Petricca, G. Angloher, P. Bauer, A. Bento, C. Bucci, L. Canonica, X. Defay, A. Erb, F. v. Feilitzsch, et al. First results on low-mass dark matter from the cresst-iii experiment, 2017.

[38] E. Aprile, J. Aalbers, F. Agostini, M. Alfonsi, L. Althueser, F. D. Amaro, M. Anthony, F. Arneodo, L. Baudis, B. Bauermeister, et al. Dark matter search results from a one ton-year exposure of xenon1t. *Phys. Rev. Lett.*, 121:111302, Sep 2018. doi: 10.1103/PhysRevLett.121.111302. URL https://link.aps.org/doi/10.1103/PhysRevLett.121.111302.

[39] D. S. Akerib, S. Alsum, H. M. Araújo, X. Bai, A. J. Bailey, J. Balajthy, P. Beltrame, E. P. Bernard, A. Bernstein, T. P. Biesiadzinski, et al. Results from a search for dark matter in the complete lux exposure. *Phys. Rev. Lett.*, 118:021303, Jan 2017. doi: 10.1103/PhysRevLett.118.021303. URL https://link.aps.org/doi/10.1103/PhysRevLett.118.021303.

[40] P. Agnes, I. F. M. Albuquerque, T. Alexander, A. K. Alton, G. R. Araujo, D. M. Asner, M. Ave, H. O. Back, B. Baldin, G. Batignani, et al. Low-mass dark matter search with the darkside-50 experiment. *Phys. Rev. Lett.*, 121:081307, Aug 2018. doi: 10.1103/PhysRevLett.121.081307. URL https://link.aps.org/doi/10.1103/PhysRevLett.121.081307.

[41] C. Amole, M. Ardid, I.J. Arnquist, D.M. Asner, D. Baxter, E. Behnke, P. Bhattacharjee, H. Borsodi, M. Bou-Cabo, P. Campion, et al. Dark matter search results from the pico-60 c3f8 bubble chamber. *Physical Review Letters*, 118(25), Jun 2017. ISSN 1079-7114. doi: 10.1103/physrevlett.118.251301. URL http://dx.doi.org/10.1103/PhysRevLett.118.251301.

[42] Tarek Saab. Dark matter limit plotter v5.16. URL https://supercdms.slac.stanford.edu/dark-matter-limit-plotter.

[43] A.H. Abdelhameed, G. Angloher, P. Bauer, A. Bento, E. Bertoldo, C. Bucci, L. Canonica, A. D'Addabbo, X. Defay, S. Di Lorenzo, et al. First results from the cresst-iii low-mass dark matter program. *Physical Review D*, 100(10), Nov 2019. ISSN 2470-0029. doi: 10.1103/physrevd.100.102002. URL http://dx.doi.org/10.1103/PhysRevD.100.102002.

CHAPTER 2

CCDS AS DARK MATTER DETECTORS

2.1 Charge-Coupled Device Overview

Charge-coupled devices (CCDs) are relatively simple devices composed of metal-oxide-semiconductors (MOS) gates forming a grid of capacitors for charge storage. Initially conceived as a memory storage device for early computers, it was quickly realized that silicon CCDs could act as highly effective sensors for cameras [1]. As cameras, silicon CCDs have seen widespread use in the astronomical community. CCDs have been so important to astronomy that Willard Boyle and George Smith shared in the 2009 Nobel prize in physics for their invention. CCDs are widely used in visible and near-infrared astronomical telescopes, from ground based designs such as the Vera C. Rubin Observatory [2] to the space-based Hubble telescope [3]. For a complete and extremely thorough explanation of everything CCD related, see Janesick [4], which covers the subject of scientific CCDs exhaustively.

As cameras, CCDs have extremely high sensitivity (with a quantum efficiency in the visible and near infrared of close to 1) , a few e^- of readout noise, and charge transfer efficiencies of $\sim$1. While most of their development and use over the past 50 years has been in the field of photosensors for cameras, CCDs can also function as highly sensitive general purpose particle detectors. Their low noise and pixelized nature allow for high energy and spatial resolution, while the inherent design requirement for nearly pure silicon results in relatively low intrinsic detector backgrounds. The primary disadvantage to this use is CCDs relatively slow readout speeds, with readout times in the seconds to minutes for typical low-noise applications, and relatively low mass (as CCDs are typically only a few hundred microns thick). Nevertheless, experiments such as DAMIC [5] and CONNIE [6] have demonstrated that CCDs are powerful and versatile detectors for a range of different experimental searches requiring low-threshold detectors.

2.2 CCD Design

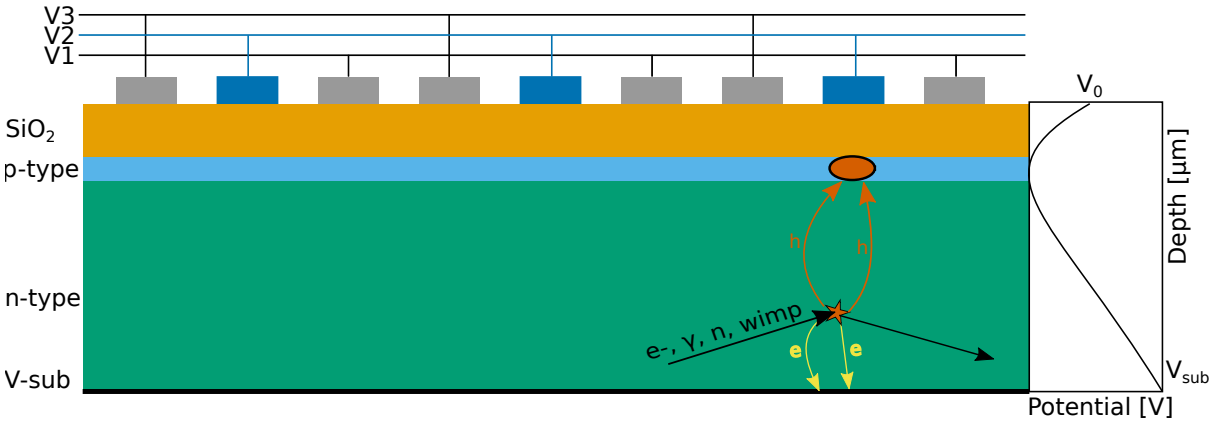


Figure 2.1: Schematic cross sectional view of a CCD, showing along a CCD column. Charge is produced when some particle strikes the bulk of the silicon substrate. This substrate is fully depleted by the application of the voltage V_{sub} across it, which causes the produced electron-hole pairs to drift apart. The electrons are collected on the backside contact (labeled "V-sub" in the diagram) and discarded, while the holes are collected under the V2 gate for each pixel. The voltage potential across the CCD is shown on the right hand side: the charge is collected in the voltage minimum inside the p-type silicon located at the top of the silicon substrate. The SiO_2 layer insulates this layer from the metal gates on the top, preventing charge loss.

DAMIC CCDs are developed by the Lawrence Berkeley National Laboratory (LBNL) [7]. The detection media of a CCD is a bulk layer of n-type silicon (the "substrate") sandwiched between a backside metal contact and the metal-oxide-silicon (MOS) gates on the front of the CCD, as shown in fig. 2.1. During image exposure[1], a voltage (the "substrate voltage" or V_{sub}) is applied across the backside contact and the front electrodes, creating an electric field inside the bulk of the detector. This electric fields depletes the silicon of free charge carriers, and prevents charge recombination when new carriers are produced. CCDs are fabricated using high-resistivity (> 10 kΩ cm) silicon, to allow for full depletion using low (~40V) bias voltages.

Particles that strike the silicon bulk cause ionization that releases electron-hole pairs from below the conduction band[2]. These pairs are split and drifted by V_{sub} to the surface (holes) and backside (electrons) of the CCD. In DAMIC CCDs, the charge is collected in a region of p-type silicon created above the bulk n-type silicon called the "buried channel".

1. Even when used as particle detectors, the language of "exposure" and "image" are commonly used.

2. In a CCD, either the electrons or the holes can act as charge carriers, depending on the design. In DAMIC CCDs, the positive charge holes are collected. For brevity, the term "charge" or sometimes "electrons" is used to refer to whatever carriers are collected, even if the carriers are holes.

This buried channel design keeps the charge packets away from the surface of the CCD, which increases the charge transfer efficiency and reduces the dark current produced in the CCD. While drifting to the surface of the CCD, the charge carriers are free to diffuse. For a typical low-energy deposit, this diffusion is driven by thermal motion, which causes the distribution of charge carriers on the surface to assume a Gaussian distribution centered at the location of the deposit, and with a width dependent on depth of the interaction (for more details see section 2.4.2). Charge-charge repulsion causes additional diffusion for higher energy clusters, and are especially dominant for charge produced by alpha particles inside the bulk or near the back of the CCD (this has a special term, the "plasma" effect [8]).

The surface structure of DAMIC CCDs is composed of sets of 3 MOS gates arranged vertically. Charge is prevented from transferring horizontally in the pixel array through channel stops (regions of high boron dopant). During an exposure, the central gate (nominally called "V2") is held at a higher voltage than the other two gates, causing charge in the pixel to accumulate under that gate. To readout the CCD, the gates are clocked in order to shift charge vertically, in a procedure called "vertical clocking", shown in fig. 2.2. Charge is shifted vertically along a column until it reaches the end of the CCD, where a special row of pixels called the "serial register" is located. Unlike the main frame of the CCD, pixels in the serial register have gates arranged horizontally (these gates are called "H1", "H2", and "H3"). These horizontal gates are clocked in order to shift the charge either left or right ("horizontal clocking", as shown in fig. 2.3. At either end of the serial register is the readout node. A micrograph picture of the CCD readout node, and schematic of vertical clocking process, can be seen in fig. 2.4

DAMIC CCDs contain 4 readout nodes, located at each of the four corners of the CCD, at the left and right ends of each serial register (these are labeled as L1, L2, U1, and U2 in fig. 2.5). The readout node is composed of several parts, shown schematically in fig. 2.6. First is the summing well, an adjustable gate which stores the charge just before readout. Next is an output gate that prevents charge from leaking from the summing well into the readout node itself until the sense node is ready for readout. After the

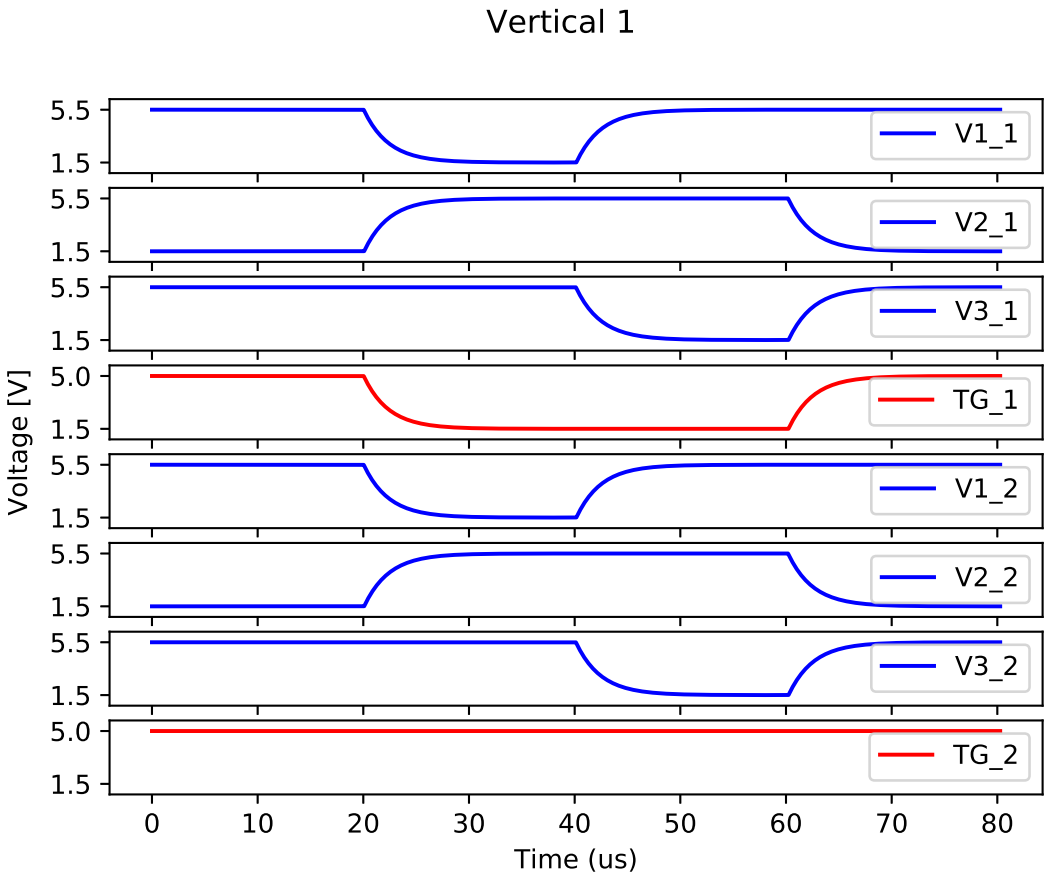


Figure 2.2: Plot of the vertical clocks during a readout process shifting charge towards serial register 1. In this diagram, charge is contained in the lowest well, starting under clock V2. Note that both sets of vertical clocks operate in the same sequence (charge is transferred V2→ V1→ V3→ V2). Only the transfer gate TG_1 is held open, as charge is not allowed to transfer into through TG_2 into serial register 2. Voltages and timing are typical. The clocks are shown after being passed through $\tau = 2.35\mu s$ RC filter (typical for DAMIC CCD clocks).

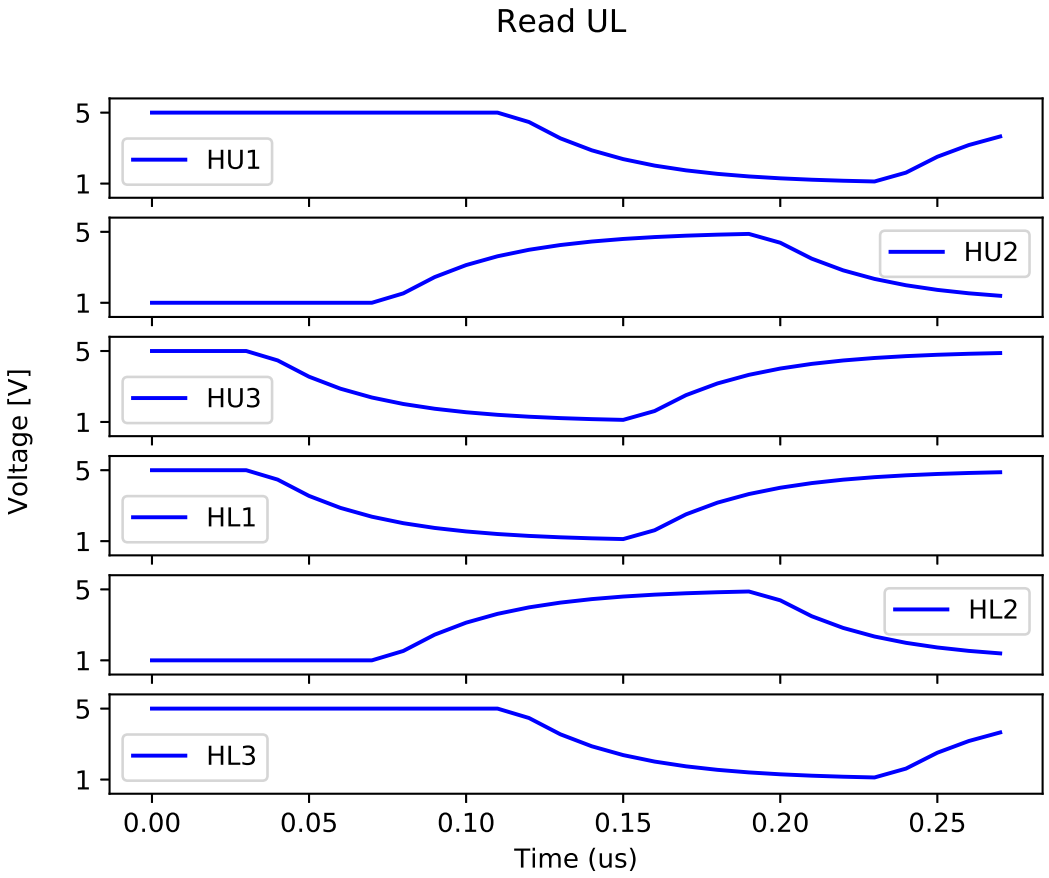


Figure 2.3: Plot of the horizontal clocking process. As in the vertical clocks, charge is held in the lowest clock gate. The shown clocking shifts the charge towards one specific side of the serial register: the left ("L") clocks and right (or "U") transfer charge in opposite order, both starting under H2, but shifting charge H2→ H3→ H1→ H2 for the U side (which shifts charge towards the "L" readout node), and H2→ H1→ H3→ H2 for the L side to shift charge towards that node. Clocks are shown passed through a $\tau = 33NP$ RC filter.

output gate is the sense node itself, a floating diffusion amplifier that holds the charge during readout. The sense node is a transistor that acts as a charge amplifier, producing a voltage proportional to the charge inside the sense node. Last is the drain, which is an electrode held at a low voltage. Once the charge has been read, a reset gate opens the connection between the sense node and the drain, allowing charge to clear from the node so that another pixel can be read.

2.3 CCD Readout

The readout process is as follows: once charge arrives in the summing well and the sense node has been reset from the previous read, a reference voltage is read from the readout

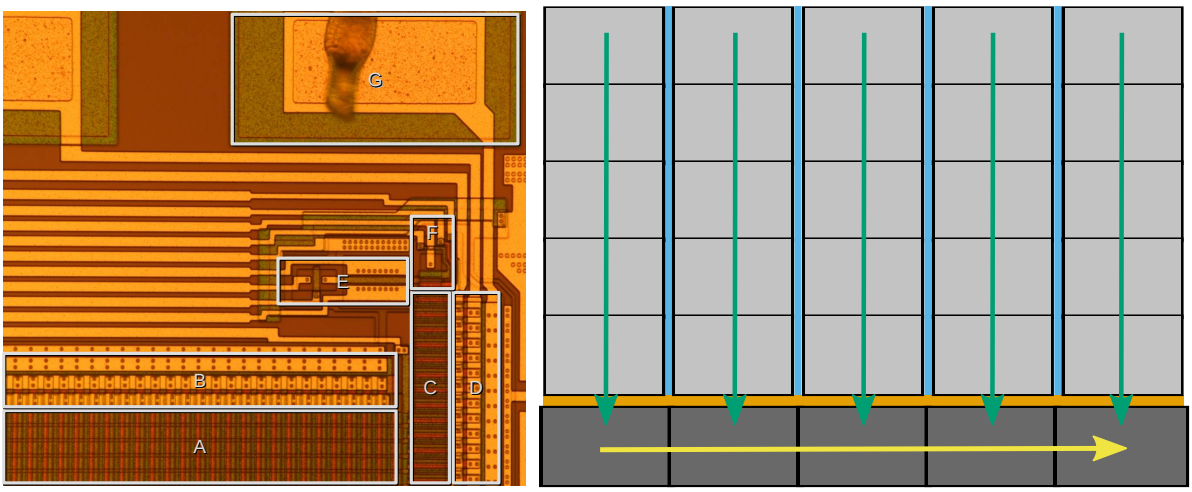


Figure 2.4: Left: Labeled micrograph of the readout region of a skipper CCD. Parts are: (A) the pixel array, with the 3 different colored bands showing the 3 vertical clocking regions. (B) the metal contacts for the three vertical clocks (V1,V2, and V3). (C) The serial register. (D) Metal contacts for the horizontal clocks. (E) The CCD amplifier and reset gate. (F) The summing well, output gate, sense node, and drain. (G) A wirebonded contact pad (the pad for clock "H3" is shown). In general, the copper regions show metal contacts, the dark reddish-brown is the bulk of the silicon, and differently colored stripes in the pixel array/serial register show different gates. **Right**: Top view of CCD. Light gray columns show the vertical columns of the CCD. Blue lines between columns represent channel stops that prevent charge from bleeding between pixels horizontally. The dark gray blocks at the bottom represent the serial register pixels, that hold charge to be read out by the sense node (not shown). The orange line between them shows the transfer gate that blocks charge from entering or leaving the serial register except during horizontal clocking procedure. Charge is shifted vertically down along the light green arrows, and horizontally along the yellow arrow. CCDs typically contain serial registers at both the top and bottom, and sense nodes at both left and right hand sides, often allowing for charge to be read from all 4 sense nodes simultaneously to speed up readout.

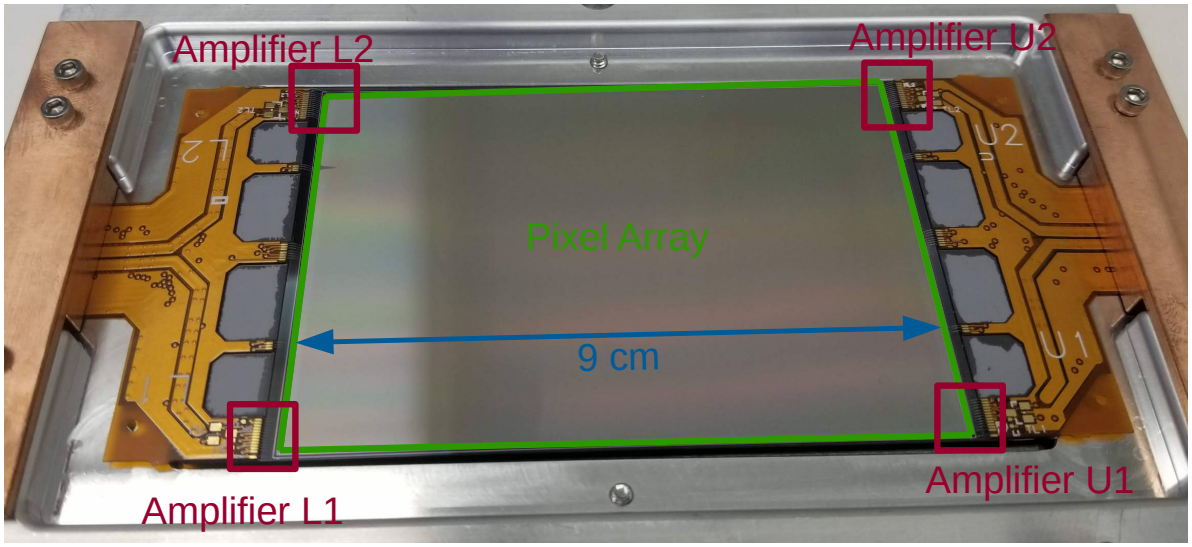


Figure 2.5: Picture of a 6k by 4k DAMIC-M development CCD, showing the pixel region and 4 amplifiers on the corners of the CCD. The CCD is shown in an aluminum frame, with copper bars holding the orange Kapton cables fingers in place.

amplifier before charge is transferred into the sense node. This reference voltage allows the elimination of "reset" or drift noise that occurs due to changes in the baseline output of the CCD amplifier. Once the reference voltage (or pedestal, as it is commonly called) has been measured, the summing well voltage is raised, allowing the charge inside the summing well to fall over the output gate and into the sense node. Now the voltage output of the amplifier is proportional to the charge inside the sense node, and a measurement of the pixel signal can be made. Subtracting the reference voltage from this measurement yields the absolute value of the charge in the pixel. This procedure is known as "correlated double sampling" (CDS), as two measurements are made of the amplifier output, one before and one after charge is dumped into the sense node.

Conventionally, this CDS is performed using an analog integrating circuit, integrating over the pedestal before flipping the polarity of the input amplifier and integrating over the signal. This technique allows for the removal of high frequency white noise, which

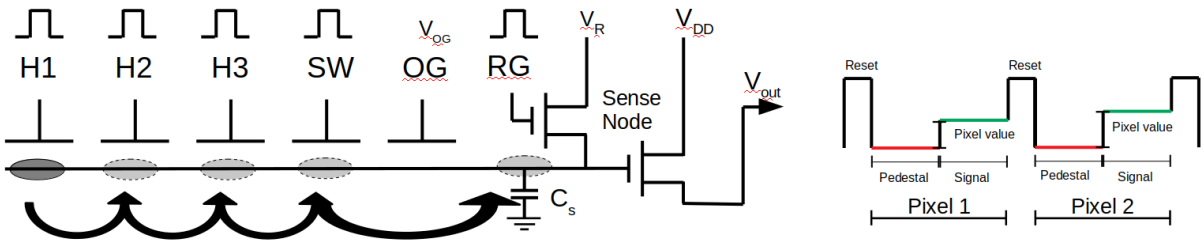


Figure 2.6: Schematic of the readout procedure for a single pixel, showing the process of charge being shifted through the horizontal serial register (the "H" clocks), past the output gate and into the summing well for readout. The left hand side shows the clocking procedure and simplified schematic of the sense node circuit, while the right hand shows a diagrammatic example of the output video signal for two pixels with some charge in them.

averages to zero over the integration time. Mathematically, the CDS transfer function is

$$|H_{CDS}(f)|^2 = 2 - 2\cos(2\pi f t_s) \tag{2.1}$$

where f is the input noise frequency, and t_s is the CDS integration window. For a real CCD video signal, the output is filtered by the amplifier circuit. If we assume a typical simple RC filter, we can write the filter transfer function as

$$|H_{RC}(f)|^2 = \frac{1}{1 + (2\pi f \tau)^2} \tag{2.2}$$

where τ is the RC time constant for the filter. The full gain of the system as a function of frequency is then the product of these two functions. A plot of the transfer function vs frequency for both the raw CDS transfer function, and for the CDS transfer function applied to an RC filtered signal can be found in fig. 2.7. If the power spectral distribution of the CCD amplifiers $|N_{CCD}(f)^2|$ is also know, the final noise for a given integration window can be computed from integrating the product of all three, i.e.

$$N_{CCD}(V) = \left(\int_0^\infty |N_{CCD}(f)|^2 |H_{RC}(f)|^2 ||H_{CDS}(f)^2|^2 \right)^{1/2} \tag{2.3}$$

The noise of the CCD is typically assumed to contain two components, a white noise and

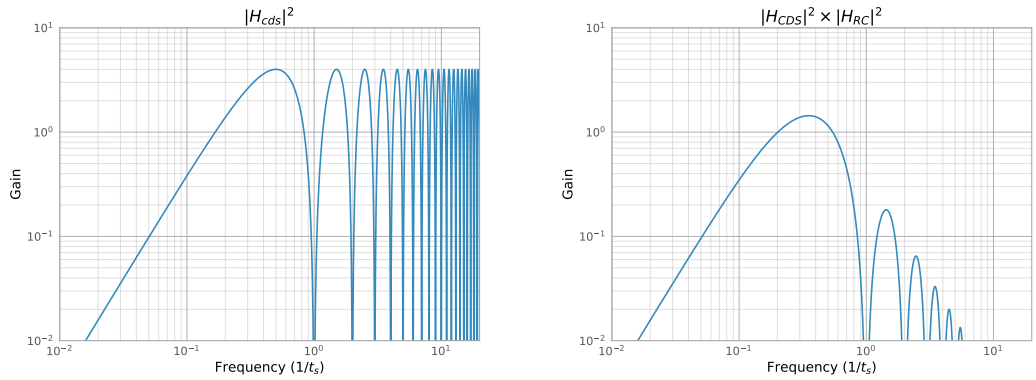


Figure 2.7: CDS transfer functions, for CDS alone (left) and CDS times RC filtering (right), in frequencies of units $1/t_s$. In both cases, the gain is highest for noise with frequencies at odd multiples of $1/2t_s$, and 0 at $1/t_s$. The CDS transfer function increases the gain of such frequencies by a factor of 4. When combined with the RC filter, it can be seen that higher frequency noise becomes negligible, but any noise with a frequency $f \approx 0.5/t_s$ has a gain >1, and is therefore amplified by the CDS function..

a 1/f noise. We can write this as

$$|N_{CCD}(f)|^2 = |W_{CCD}(f)|^2(1 + (f_c/f)) \tag{2.4}$$

where f_c is the corner frequency where the 1/f noise is equal to the white noise.

Extending the integration time allows the near complete elimination of white noise, and for analog integrators also increases the gain of the system (since the charge on the integrating capacitor and therefore resulting voltage is proportional to the integration time). Both effects are desirable for optimal performance in low-energy searches, and standard off the shelf CCD readout systems can often reach readout noise of around 1 electron RMS. However, the noise-reduction effect is limited by correlated noise effects, primarily 1/f noise from the CCD amplifiers. This noise actually increases in magnitude as you increase integration time, an effect that can be seen graphically in fig. 2.8. Following equation eq. (2.1), we can see that the contribution of this noise is maximal at $f_s = 0.5/t_s$. Removing this noise and achieving sub-electron resolution is the domain of so-called "skipper" CCDs (which are discussed in section 2.3.1. There has also been some research in the past few years into the possibility of using a digital CDS method to measure the 1/f noise and employ digital filtering algorithms that can remove this 1/f noise [9]. However, these techniques (while possible in theory) have proven difficult to implement in practice,

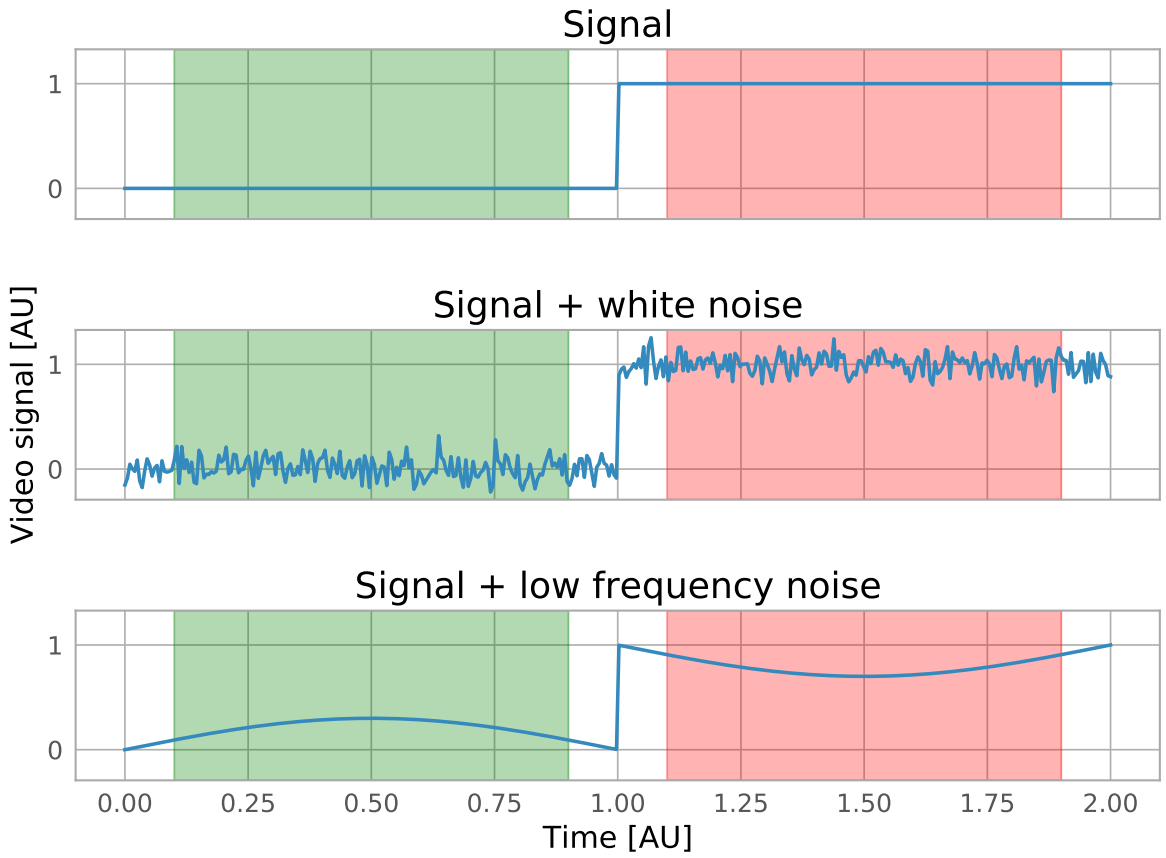


Figure 2.8: Plot showing an example of a noise-free signal (top), signal plus high frequency white noise (middle), and signal plus low frequency noise (bottom). The shaded regions show the CDS pedestal (green) and signal (red) integration windows. The value of the correlated double sampling is 1 for the top, 0.9885 for the middle, and 0.5465 for the bottom, illustrating how disruptive the low frequency noise is to the standard CDS algorithm.

and skipper CCD systems have proven more reliable.

An additional technique that can be used to lower the effective noise per pixel is *binning*. When reading out the CCD, the horizontal (vertical) clocking process can be repeated N (M) times before proceeding to the next readout step. This shifts charge from N (M) pixels into a single pixel, allowing charge from all of these pixels to be readout in a single operation. A readout of N horizontal and M vertical binning is termed *NxM* binning. The advantage of this is the aforementioned noise reduction, and an increase in readout speed $\propto$ the total binning. The disadvantage is a loss of positional information, since the pixels are effectively increased in size by a factor of N horizontally and M vertically. Typical binning for readout of DAMIC science data is 1x100 (i.e. binning 100 pixels together vertically, but performing no binning horizontally). This allows us to

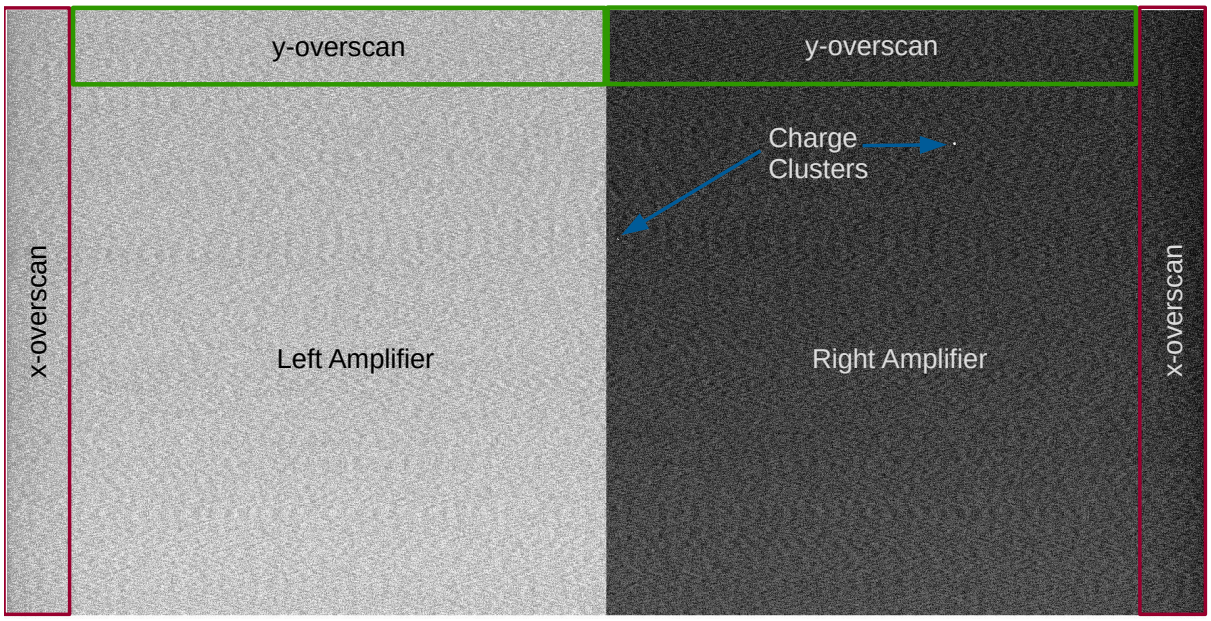


Figure 2.9: An example image from the DAMIC experiment showing the simultaneous readout of multiple channels, with the left amplifier performing a noise-only measurement, and the right amplifier reading charge. Additional x- and y- overscan is read using both amplifiers to further quantify the noise. Wavy patterns mirrored on the right and left amplifier shows the correlated noise.

retain all spatial information in the horizontal direction, while allowing a factor of 100 readout speed increase and noise reduction.

Another useful property of CCD readout is *overscan*. When reading out charge, the readout process can continue past the physical end of the CCD, in either the horizontal or vertical direction (producing *x-overscan* and *y-overscan*, respectively). These reads are of virtual pixels that don't exist and will therefore contain no charge, which allows a direct measurement of the readout noise alone. Similarly, multiple amplifiers on a CCD can be read out simultaneously. Typically this is used to speed up the readout process, but if the charge is shifted only to one amplifier, reading out another can produce a time-synchronous measurement of the electronic noise in the readout chain, allowing for reduction in correlated noise. For details of this subtraction process, see Aguilar-Arevalo et al. [10]. An example image showing this readout process can be seen in fig. 2.9. Only the right amplifier contains charge: the left serves only as a measurement of the noise.

2.3.1 Skipper CCDs

In the case of standard CCDs, after the pixel charge has been measured, the reset gate is opened, allowing the charge into the drain where it is discarded, and the sense node is cleared to be ready for another measurement. In the early 1990s, however, Janesick et al. [11] and Chandler et al. [12] realized that with some small modifications to the CCD readout node, this procedure could be changed to allow for ultra-low noise CCDs. Instead of immediately throwing away the pixel charge after a readout, the output gate voltage (and summing well itself) can be lowered to allow the charge to flow back from the sense node into the summing well again. Once there, the sense node can be reset without destroying the charge, which can be transferred back and forth into the sense node multiple times. This requires modifying the sense node itself to a floating gate amplifier (as a standard floating diffusion amplifier would inject charge into the signal).

Given proper circuit design and operational parameters, this measurement can be made efficiently without loss of charge, allowing multiple independent non-destructive charge measurements (*NDCMs* or "skips"). A schematic of this modified readout node and the "skipping" process is given in fig. 2.10, and an example of the clocking procedure can be seen in fig. 2.11. Since each measurement is independent, the noise of each measurement is uncorrelated, and taking the average of the measurements allows for a more precise measurement of the charge (following a $\frac{1}{\sqrt{N}}$ noise reduction law). In practice, actual skipper CCDs follow this expected noise reduction quite well, with little or no loss of charge during the measurement process. See for example fig. 4.14 in chapter 4. Skipper CCD readout is necessary to remove 1/f noise (given in eq. (2.3)) and low frequency noise in general (such as that shown in fig. 2.8), which creates a noise floor for standard CCD operation.

This skipper readout process allows for unprecedented single-electron charge resolution in CCD readout. This lowers the effective detection threshold for CCD detectors from the current noise threshold of $\sim 10e^-$ (50eV)[13], to a $2e^-$ (7eV) threshold (set by the presence of dark current which creates an unavoidable single-electron background).

Large-scale readout of skipper CCDs requires the development of new readout elec-

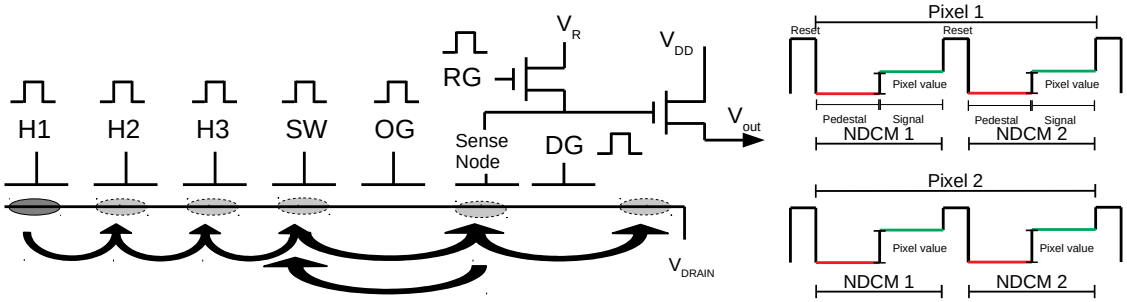


Figure 2.10: Schematic of skipper CCD readout process.

tronics, which is the focus of chapters 3 and 4.

2.4 CCD Operational Characteristics

2.4.1 Charge Transfer Efficiency

Modern CCDs are capable of highly efficient charge transfer between pixels, often exceeding 99.9999% charge transfer efficiency (CTE), which typically renders loss of charge in charge packets negligible. However, in some cases, such as when operated with non-ideal gate voltages or clock timings, or if the CCD is damaged extensively, charge loss (typically quantified as charge transfer *in*efficiency, defined as CTI $= 1 -$ CTE) can have a noticeable effect in the CCD. CTI can be caused by failure to properly transfer charge during clocking (causing some probability of leaving charge behind during each transfer). This is rarely a problem if the CCD is operated properly, and is usually ignored in DAMIC CCDs as a negligible process[3]. On the other hand, charge loss can also occur due to damage in the CCD. Such damage (for e.g. from neutrons striking silicon nuclei) can cause charge traps in some pixels of the CCD. These traps can absorb charge when charge packets are transferred through the pixel containing them, causing the charge to be left behind during the clocking process. This charge is then subsequently released, typically over timescales of a few hundred microseconds to milliseconds (the exact timescale depends on the nature of the dislocation in the CCD and the operating temperature of the

3. For a typical CTI value of 10^{-6}, for example, the worst-case scenario is around .8% of the charge will be shifted back one pixel and around .03% will be shifted back 2 pixels or more.

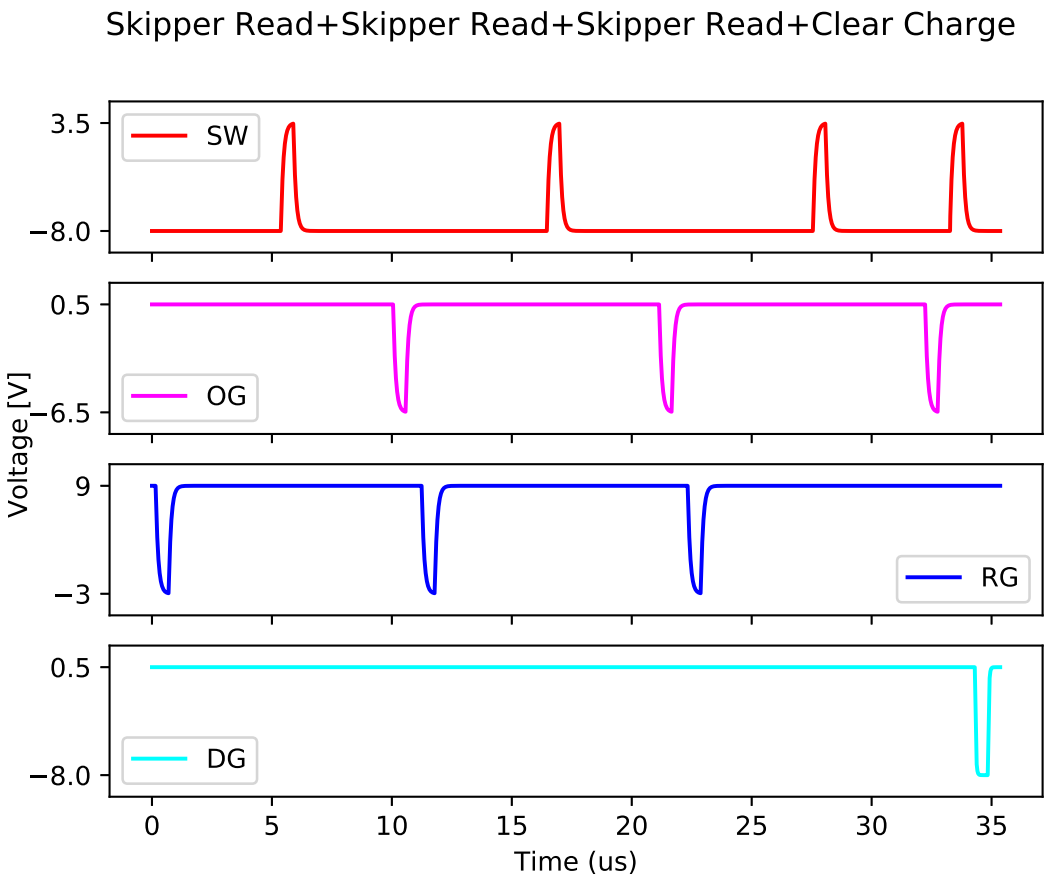


Figure 2.11: Typical clocks for a "skipper" NDCM readout, showing 3 skips followed by the charge dump process at the end.

CCD). This effect causes a non-binomial CTI effect (known as *deferred charge* [14]) that cannot be quantified using ordinary CTI models. More discussion of CTI can be found in section 6.5.

2.4.2 Diffusion

One of the interesting properties of CCDs is charge diffusion. When charge is produced in the detector, it is pulled to the buried channel by the applied electric field in the z-direction, but is left mostly free to diffuse through thermal motion in the x-y plane. The result is that the charge may end up in an adjacent pixel, rather than under the original location of the charge deposition. The distance traveled by a hole-pair in the diffusion process is calculated in [7] and found to be of the form

$$\sigma^2 = -Aln(1 - bz) \tag{2.5}$$

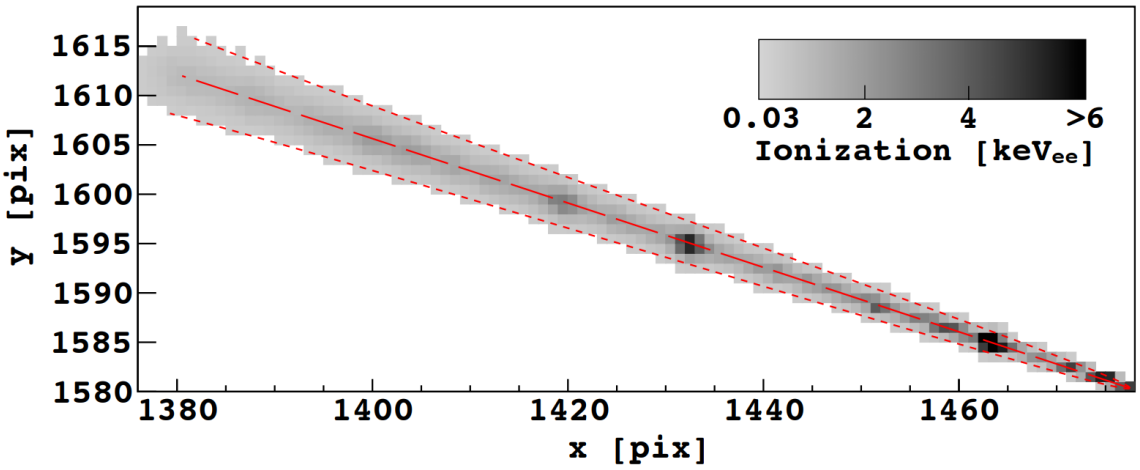


Figure 2.12: Example of a muon track used to extract diffusion parameters. The dashed (dotted) red lines show the center (edge) of the track, with the depth of the interaction proportional to the line-to-line distance. The track on the bottom right shows where the muon passes through the front of the CCD, and the top left is where the muon passes through the back.

where

$$A = \frac{2\epsilon_{Si} k_B T}{\rho_n e} \approx 150 \mu m^2 \tag{2.6}$$

$$b = \left(\frac{\epsilon_{Si} V_b}{\rho_n z_D} + \frac{z_D}{2} \right)^{-1} \approx 10^{-3} \mu m^{-}1 \tag{2.7}$$

Here, ϵ_{Si} is the permittivity of the silicon, k_B is the Boltzmann constant, T is the temperature, ρ_n is the volume charge density, e is the charge of an electron, V_b is the potential difference across the CCD, and z_D is the CCD thickness.

In practice, calculating A and b precisely from first principles is difficult, since it requires precise knowledge of the CCDs physical parameters. Instead, these parameters are typically extracted from data, such as fitting to the width of a muon track through the CCD (for e.g. fig. 2.12).

2.4.3 Dark Current

One of the major limitations of CCD operation is *dark current*. Dark current refers to any charge produced by internal effects of the detector itself (rather than by ionization from external sources). All forms of dark current increase the effective noise of the CCD (due to shot noise), and introduce a potential low-energy signal in the detector. Dark

current may derive from multiple sources. The primary sources relevant for DAMIC are from the thermal motion of the charge carriers, defects in the CCD, and glowing of the CCD readout amplifier.

Thermal dark current is dominant at higher temperatures, and derives purely from thermodynamic fluctuations in the energy states, and depends exponentially on the temperature of the CCD. The usual form of the dark current equation is

$$D_R(e^-) = CT^{1.5}e^{-E_g/(2k_BT)} \tag{2.8}$$

where $D_R(e^-)$ is the dark current rate in electrons/pixel/second, C is a CCD-dependent constant, T is the temperature, E_g is the silicon band-gap (~1.14eV at 150K), and k_B is Boltzmann's constant. Thanks to the exponential dependence on temperature, dark current can be suppressed by lowering the CCD operating temperature, but it cannot be entirely eliminated, resulting in an effective two electron (~7 eV) minimum threshold for CCD dark matter experiments. In DAMIC CCDs, dark current at typical operational temperatures of ~150K is $\mathcal{O}(10^{-1}e^-/\text{pixel/day})$ [15].

A second source of dark events are defects in the CCD. These defects are lattice impurities or dislocations that result in stress in the atomic lattice, lowering the band gap and allowing massive amounts of thermal dark current. Such defects can be produced either during the CCD fabrication process, or later through proton or neutron radiation interacting with silicon nuclei. Fortunately, these defects are localized to a specific site in the CCD, and can be removed by *masking* the defect sites (termed "hot pixels") in the images.

Another major source of dark current is amplifier glowing. The high electric fields involved in the readout process can result in the amplifier producing light. This glow causes noticeable dark current near the amplifier itself, and can impact any nearby CCDs. Glowing was seen in DAMIC CCDs (see fig. 2.13) during the commissioning process. It was eliminated by disabling the offending CCDs, and minimizing stress on the CCD during the cooldown process (which seemed to be a principle cause).

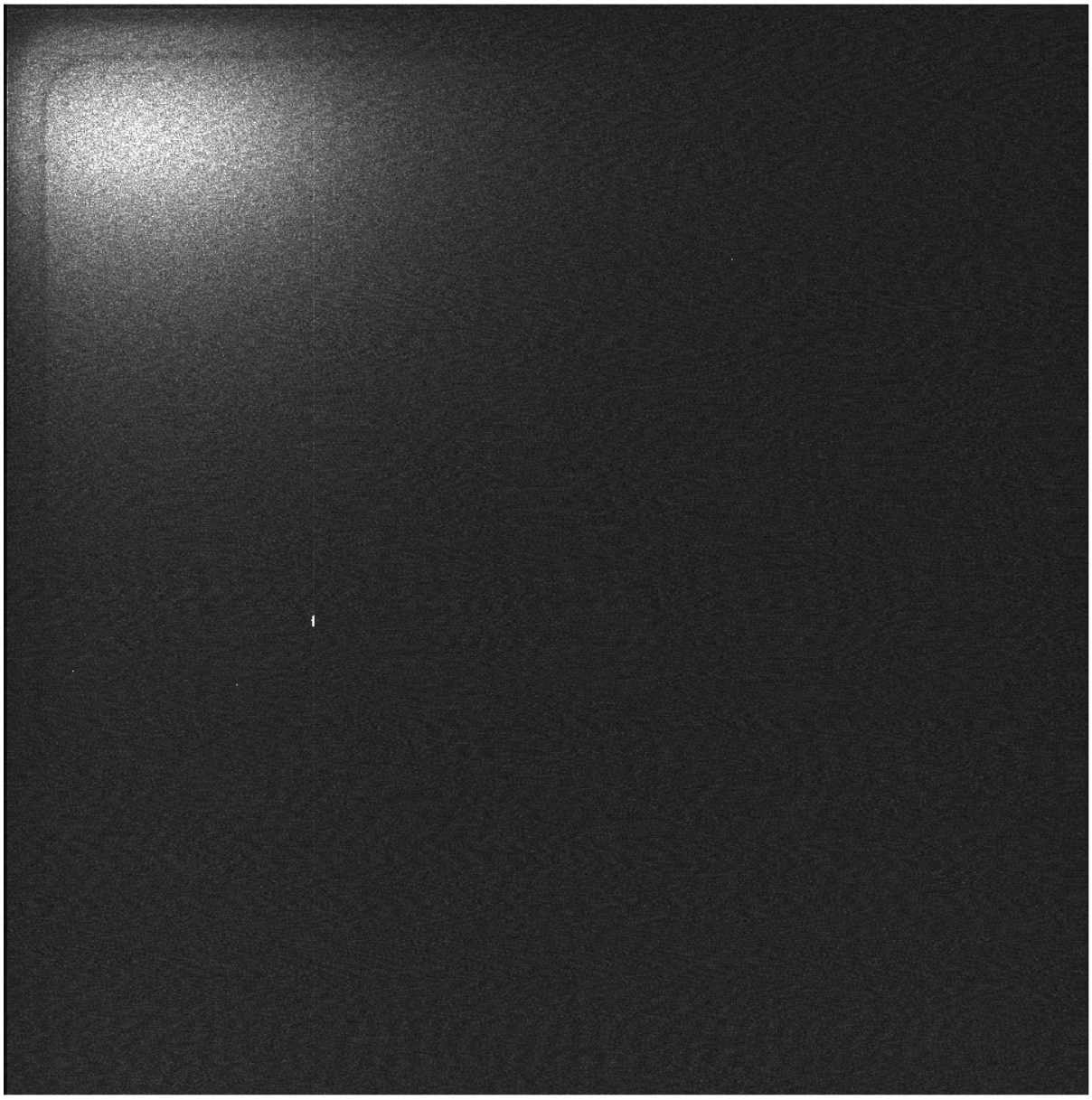


Figure 2.13: Example image clearly showing amplifier glow in the top right. It is unknown if the glowing was in the CCD itself. In the shown image, the glowing was most likely from a facing CCD, since the imprint of the curved frame can be seen in the image.

2.5 CCDs as Particle Detectors for DM Searches

The number of electron-hole pairs created in silicon per unit energy deposit by a high energy incident particles can be approximated as 3.77 $\frac{eV}{e^-}$. At low energies, the mean number of produced pairs follows a more complicated model [16].

In addition, the number of electrons produced has some fluctuation based on complicated statistical interactions in the medium. This fluctuation is usually assumed to follow a Poisson distribution with a mean value of the original energy. In practical detectors, however, the energy resolution is better than the expected Poisson variance, an improvement known as the " Fano Factor" F [17] (named after Ugo Fano, who discovered it). This factor F is always between 0 and 1, and is a temperature-dependent property of the material. For DAMIC CCDs operating at DAMIC at SNOLAB temperatures, F has been measured [18] to be 0.129, which is the value used in the remainder of this work.

As mentioned above, CCDs are primarily used in astronomy as sensors for visible and near infrared astronomy. However, any particle that deposits sufficient energy in the sensitive region of the CCD can create ionization, allowing CCDs to be used as general purpose particle detectors. The extraordinarily low noise and high spatial resolution allow for extremely precise measurements of deposits of energy all the way down to the eV scale. This allows the CCD to be sensitive to many different particles, from high energy alphas [19] down to low energy Compton [16] or neutron [5] scattering. CCDs do have several notable limitations. First, their thinness makes it likely that high energy gammas will scatter out or pass through the CCD entirely. Even high energy beta decays inside the CCD itself can escape, causing energy loss from the event. This makes the CCD inefficient as a beta or gamma detector above ~500 keV. This limitation is a major problem for many kinds of spectroscopic measurements, and renders them vulnerable to background contamination even in the low energy domain from Compton scattering of higher energy gammas. This thinness also limits the total mass of the CCD: building a detector with more active mass requires adding additional CCDs, which linearly increases the cost and complexity of the final design (compared to something like a Xenon experiment, which can be scaled by adding additional mass without requiring a

proportional scaling of readout and supporting electronics).

Second, the nature of CCD technology makes them unable to perform electron/nuclear recoil discrimination. Many dark matter searches use such discrimination to reduce electron recoil backgrounds [20] (such as the aforementioned Compton scattering), which would otherwise typically be a dominant background. However, at the low energy recoils that CCDs excel at, no developed detection technology is able to perform such electron/nuclear discrimination, making CCDs competitive as DM detectors.

Third, the exposure-readout scheme used in CCDs gives them extremely poor timing resolution compared to many other detection schemes. Many particle detectors have timing resolutions on the order of milliseconds all the way down to nanoseconds or faster. CCDs, however, must be read out relatively slowly in order to achieve their excellent noise performance, requiring on the order of seconds to hours to read a complete CCD (depending on readout configuration and parameters).

Finally, while the silicon used to make CCDs is incredibly pure (a result of the manufacturing process of the silicon wafer), ^{32}Si produced from spallation by cosmic ray neutrons of argon in the atmosphere can end up inside the CCD, producing intrinsic and unremovable contamination of the CCD. Also, ^{3}H can be produced *in situ* in the CCD from cosmic ray neutrons when the CCD is exposed on the surface. ^{32}Si can theoretically be mitigated by careful choice of silicon ingots when producing the CCD (although this is difficult in practice), but cannot be removed once the CCD is produced (and the relatively long half-life of ^{32}Si makes waiting for it to decay away impossible). ^{3}H production can be mitigated by keeping the CCDs underground as much as possible to reduce their exposure to cosmic ray neutrons, but some surface exposure is necessary when producing and transporting the CCDs during fabrication (currently, CCDs are too complex to produce on site in underground laboratories). The measurement of cosmogenic ^{3}H activation in the CCD is the subject of chapter 6.

2.5.1 Event Reconstruction

A key requirement of CCD usage as particle detectors is event and track reconstruction. Unlike when used in astronomy, the images taken are not used directly as images. Rather, we are concerned with identifying energy deposition from particle interactions in the detector. This necessitates some custom charge reconstruction code. Two methods are used for event reconstruction: the first is a simple threshold-based *seed* clustering, and the second uses a likelihood fitting. The threshold clustering methodology is relatively simple. In broad strokes, the steps are:

1. The median pixel value along each row and column is found by performing a Gaussian fit to the pixel values separately along each row and column. This value (the baseline or pedestal) is subtracted off from the rows and columns. This fit can be performed in either row or column first ordering.
2. The noise of the image is determined through a fit to the RMS pixel values.
3. "Seed" pixels that fall above the noise threshold (typically 4 times the RMS noise) are identified to begin the clustering process.
4. Pixels above a threshold (typically the same as the seed value, but this threshold may be different) and adjacent to a seed pixel are added together to form a cluster.

This seed clustering works well at identifying events at higher energies ($\gtrsim$ 500eV or so), but the efficiency drops at lower energies, since significant amounts of the charge from the event may have diffused into adjacent pixels, but with a charge value below the threshold. To identify these low energy events, we rely on the fact that such events will be nearly point-like, and result in a Gaussian spread of charge after diffusion. We can scan over the images using a 2 dimensional window (1 dimensional, in the case of binned data) and build two likelihoods. One is the likelihood $\mathcal{L}_N$ that the pixel values are noise-only, and the other is the likelihood $\mathcal{L}_G$ that the pixel values are described by a Gaussian of the form

$$f_G = I \text{ Gaus}(x, y, \mu_x, \mu_y, \sigma_x, \sigma_y) \tag{2.9}$$

where I is the total pixel value in the window, and x,y,μ_x, μ_y, σ_x, σ_y are the coordinates,

mean, and sigma of the Gaussian in x and y (for 1x100 binned data, only the x values are used). The ratio of these likelihoods $\Delta LL = -\ln \frac{\mathcal{L}_\mathcal{G}}{\mathcal{L}_N}$ then gives us a signal-to-noise likelihood ratio that can be used to identify clusters below the noise threshold of the seed clustering method.

Note that this likelihood fitting method can be applied to clusters identified using the seed clustering above. This yields better and more reliable estimates of the diffusion and total energy of the cluster.

2.6 The DAMIC Experiment

The Dark Matter in CCDs (DAMIC) experiment utilizes CCDs as particle detectors to search for dark matter. The experiment itself is located at the SNOLAB underground laboratory in Sudbury, Ontario, Canada. Located 2km underground in the Vale Creighton Mine, this laboratory is one of the deepest labs in the world, offering some of the best shielding from cosmic ray background achievable by any lab [21].

The DAMIC experiment used 675 micron thick 4k by 4k CCDs, fabricated by Lawrence Berkeley National Labs. Pictures of the detector setup can be found in fig. 2.14. The CCDs were epoxied to a silicon frame, wire bonded to a Kapton cable, and held in oxygen-free high conductivity (OFHC) copper trays inside a copper box. The box was held at cryogenic temperatures (in the range 125-150K) using a cryocooler (PCC Compressor with PT-30 Cooler) and inside a circular copper vacuum chamber with a pressure of $\sim 10^{-6}$mbar. Lead was used to provide additional shielding (with the inner ~1 inch lead nearest the CCDs being ancient lead recovered from sunken ships). A polyethylene shield around the entire detector provided shielding against neutron backgrounds from the walls of the laboratory. A N_2 gas purge from liquid nitrogen boil-off was used to reduce the radon contamination near the detector to <1 Bq m^{-3}, compared to the ~100 Bq m^{-3} rate in the SNOLAB facility.

Readout electronics close to the CCD (the video interface board or *VIB*) are placed above the CCD, with lead shielding to block the radioactive background from these electronic components. Readout of the system was performed using a MONSOON [22]

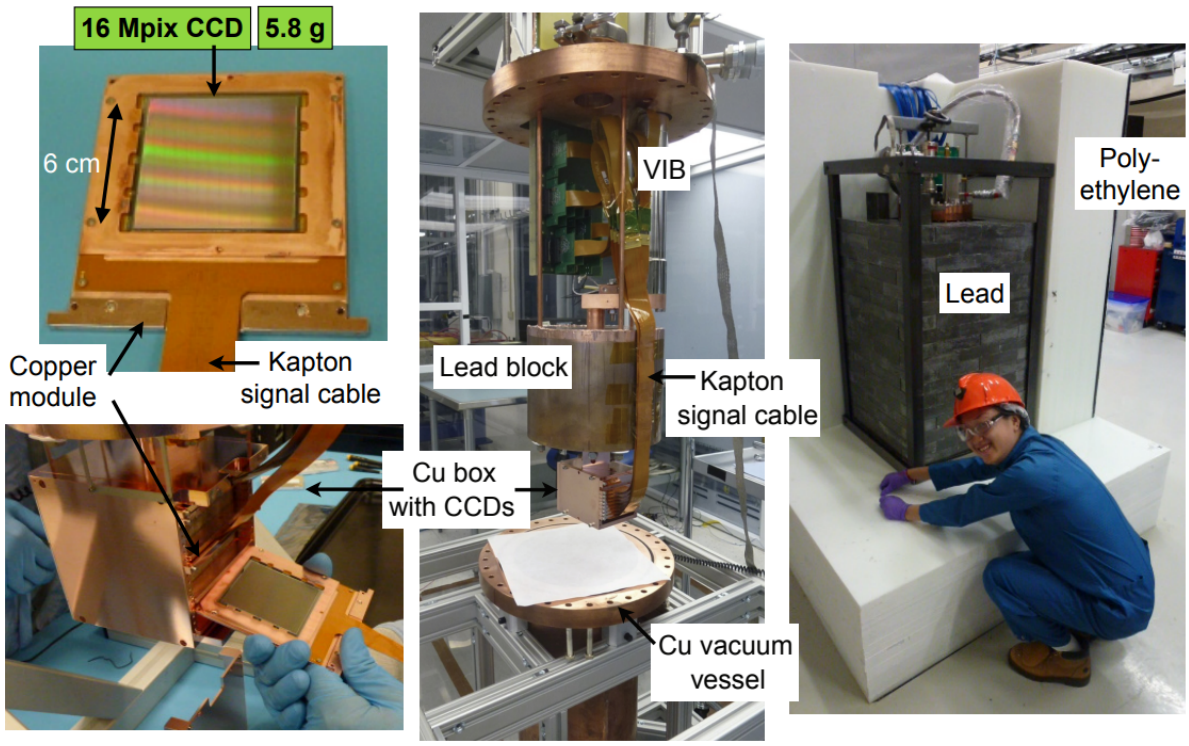


Figure 2.14: Pictures showing the DAMIC at SNOLAB setup. Clockwise from bottom left: closeup of CCD being inserted into copper box, 4k by 4k CCD, detector assembly above vacuum chamber, outside of detector showing lead and polyethylene shielding.

that provided all clocks, bias voltages, and digitization of the output of the CCDs.

An initial science run with 4k by 2k CCDs (each 2.9g) was performed in 2015. Several runs of data were taken in 2 and 3 CCD configurations over a period of 1 year. This science run was used to develop the tools for CCD analysis necessary for a DM search, with the results being published in Aguilar-Arevalo et al. [23], and also provided a proof-of-concept for the DAMIC project.

An expansion of the DAMIC at SNOLAB setup to 8 4k by 4k CCDs was performed in 2016. However, due to issues with amplifier glowing, CCD 5 was disabled, leaving only 7 functional CCDs. CCD 2 was initially non-functional, but began working during commissioning of the detector in 2017, and continued performing as expected for the remainder of the science run. During this upgrade, the detector parts (copper and lead) were extensively cleaned in an acid solution path intended to remove ^{210}Pb contamination from the surface of the components, resulting in a factor of reduction in background. The background in one CCD (CCD 1) was lowered even further by placing it in an

electroformed copper tray produced by Pacific Northwest National Lab (PNNL) [24], and placing ancient lead blocks inside the copper box above and below it. This reduced the background in that CCD to 5 d.r.u., compared to the 11 d.r.u. for the other 6 CCDs.

For the second stage experiment, 181 days of background and commissioning data was taken using 1x1 binned data to enhance the spatial discrimination capabilities of the setup. This data was used to help establish the background level of the experiment and perform coincidence studies [25]. After commissioning was completed, science data was taken in 1x100 binned mode, totaling 308 days of exposure. The WIMP search results from this science run are published in Aguilar-Arevalo et al. [13]. The development of simulations for the background model used in these results is the primary subject of chapter 5.

2.6.1 DAMIC-M

The next generation of the DAMIC experiment will be located at the Modane Underground Laboratory (LSM) in Modane, France. Called "DAMIC-M" (for Modane), this experiment will scale to deploying $\mathcal{O}(100)$ CCDs, with a target mass and operating time of 1 kg and 1 year (respectively). The background goal for DAMIC-M is ∼0.1 d.r.u., which will be achieved through careful use of ancient lead, electroformed copper, radon free environments, and extensive background assays and simulations. A test-bed called the Low Background Chamber (LBC) using a is scheduled for commissioning in fall of 2021, using 2 skipper CCDs and target background of 1-10 d.r.u., which will be used for background measurements, electronics component testing, and early DAMIC-M science results. Early schematics of the DAMIC-M LBC and main detector can be seen in fig. 2.15.

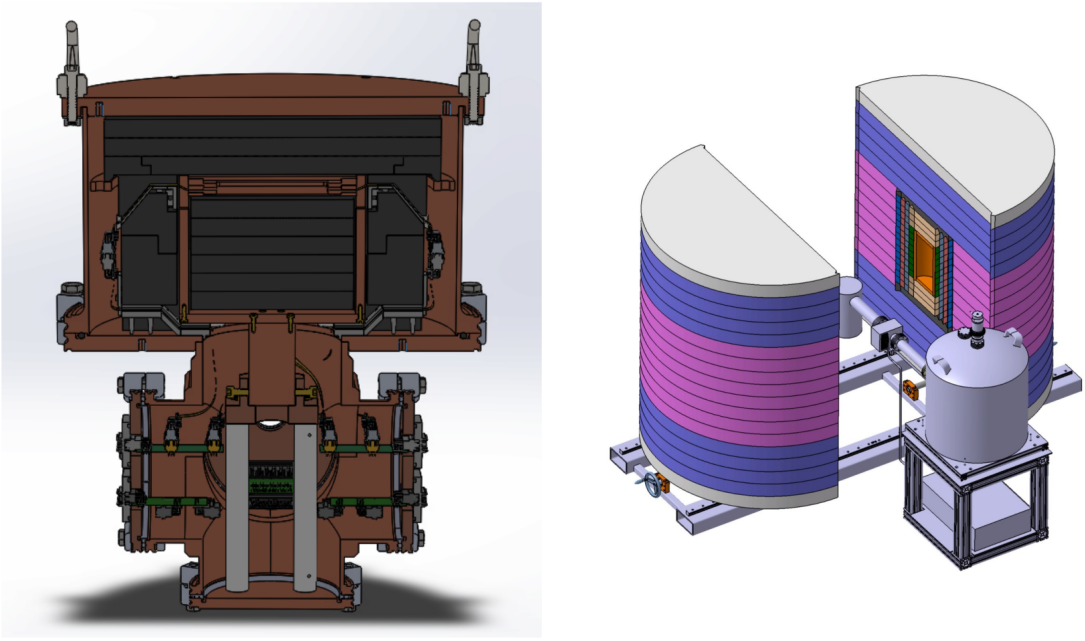


Figure 2.15: Left: Cross-sectional view of the low-background chamber (LBC) prototype DAMIC-M detector. The CCDs will be located in copper trays in the middle of the lead (dark gray). Only ancient lead will be used near the CCDs. The copper components will be OFHC initially, with an EFC (electroformed copper) upgrade planned later. The cold finger and electronics will come through the 5 flanges on the bottom of the copper cryostat. **Right:** Diagram of the full DAMIC-M detector design. Shown is an initial design, subject to further revisions from measurements made with the LBC.

References

[1] W. S. Boyle and G. E. Smith. Charge coupled semiconductor devices. *The Bell System Technical Journal*, 49(4):587–593, 1970. doi: 10.1002/j.1538-7305.1970.tb01790.x.

[2] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, D. Alonso, Y. AlSayyad, S. F. Anderson, J. Andrew, et al. LSST: From Science Drivers to Reference Design and Anticipated Data Products. ApJ, 873:111, March 2019. doi: 10.3847/1538-4357/ab042c.

[3] E.J. Chaisson and R. Villard. The science mission of the hubble space telescope. *Vistas in Astronomy*, 33:105–141, 1990. ISSN 0083-6656. doi: https://doi.org/10.1016/0083-6656(90)90018-4. URL `https://www.sciencedirect.com/science/article/pii/0083665690900184`.

[4] James R. Janesick. *Scientific charge-coupled devices*. SPIE Press, 2001. ISBN 0-8194-3698-4.

[5] A. E. Chavarria, J. I. Collar, J. R. Peña, P. Privitera, A. E. Robinson, B. Scholz, C. Sengul, J. Zhou, J. Estrada, F. Izraelevitch, et al. Measurement of the ionization produced by sub-kev silicon nuclear recoils in a ccd dark matter detector. *Phys. Rev. D*, 94:082007, Oct 2016. doi: 10.1103/PhysRevD.94.082007. URL `https://link.aps.org/doi/10.1103/PhysRevD.94.082007`.

[6] CONNIE Collaboration, A. Aguilar-Arevalo, X. Bertou, C. Bonifazi, M. Butner, G. Cancelo, A. Castaneda Vazquez, B. Cervantes Vergara, C. R. Chavez, H. Da Motta, et al. The connie experiment, 2016.

[7] S. E. Holland, D. E. Groom, N. P. Palaio, R. J. Stover, and Mingzhi Wei. Fully depleted, back-illuminated charge-coupled devices fabricated on high-resistivity silicon. *IEEE Transactions on Electron Devices*, 50(1):225–238, 2003. doi: 10.1109/TED.2002.806476.

[8] J. Estrada, J. Molina, J.J. Blostein, and G. Fernández. Plasma effect in silicon charge coupled devices (ccds). *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 665: 90–93, Feb 2011. ISSN 0168-9002. doi: 10.1016/j.nima.2011.10.060. URL `http://dx.doi.org/10.1016/j.nima.2011.10.060`.

[9] Konstantin Stefanov. Digital cds for image sensors with dominant white and 1/f noise. *Journal of Instrumentation*, 10:P04003–P04003, 04 2015. doi: 10.1088/1748-0221/10/04/P04003.

[10] A. Aguilar-Arevalo, D. Amidei, D. Baxter, G. Cancelo, B. A. Cervantes Vergara, A. E. Chavarria, E. Darragh-Ford, J. R. T. de Mello Neto, J. C. D'Olivo, J. Estrada, et al. Constraints on light dark matter particles interacting with electrons from damic at snolab. *Physical Review Letters*, 123(18), Oct 2019. ISSN 1079-7114. doi: 10.1103/physrevlett.123.181802. URL `http://dx.doi.org/10.1103/PhysRevLett.123.181802`.

[11] J. Janesick, T. Elliot, A. Dingizian, R. Bredthauer, C. Chandler, J. Westphal, and J. Gunn. New Advancements in Charge-Coupled Device Technology - Sub-Electron Noise and 4096 X 4096 Pixel CCDs. In George H. Jacoby, editor, *CCDs in astronomy*, volume 8, page 18, January 1990.

[12] Charles E. Chandler, Richard A. Bredthauer, James R. Janesick, and James A. Westphal. Sub-electron noise charge-coupled devices. In Morley M. Blouke, editor, *Charge-Coupled Devices and Solid State Optical Sensors*, volume 1242, pages 238 – 251. International Society for Optics and Photonics, SPIE, 1990. URL https://doi.org/10.1117/12.19457.

[13] A. Aguilar-Arevalo, D. Amidei, D. Baxter, G. Cancelo, B. A. Cervantes Vergara, A. E. Chavarria, J. C. D'Olivo, J. Estrada, F. Favela-Perez, R. Gaïor, et al. Results on low-mass weakly interacting massive particles from an 11 kg d target exposure of damic at snolab. *Phys. Rev. Lett.*, 125:241803, Dec 2020. doi: 10.1103/PhysRevLett.125.241803. URL https://link.aps.org/doi/10.1103/PhysRevLett.125.241803.

[14] Morley M. Blouke, Fanling H. Yang, Denis L. Heidtmann, and James R. Janesick. Traps and deferred charge in ccds. In Lloyd B. Robinson, editor, *Instrumentation for Ground-Based Optical Astronomy*, pages 462–485, New York, NY, 1988. Springer New York. ISBN 978-1-4612-3880-5.

[15] Karthik Ramanathan. *Particles Taking Selfies: Investigations into Light Dark Matter Using Silicon Charge Coupled Devices*. PhD thesis, Ph.D., 2020-12, 2020. URL http://knowledge.uchicago.edu/record/2738.

[16] K. Ramanathan and N. Kurinsky. Ionization yield in silicon for ev-scale electron-recoil processes. *Phys. Rev. D*, 102:063026, Sep 2020. doi: 10.1103/PhysRevD.102.063026. URL https://link.aps.org/doi/10.1103/PhysRevD.102.063026.

[17] U. Fano. Ionization yield of radiations. ii. the fluctuations of the number of ions. *Phys. Rev.*, 72:26–29, Jul 1947. doi: 10.1103/PhysRev.72.26. URL https://link.aps.org/doi/10.1103/PhysRev.72.26.

[18] K. Ramanathan, A. Kavner, A. E. Chavarria, P. Privitera, D. Amidei, T.-L. Chou, A. Matalon, R. Thomas, J. Estrada, J. Tiffenberg, and J. Molina. Measurement of low energy ionization signals from compton scattering in a charge-coupled device dark matter detector. *Phys. Rev. D*, 96:042002, Aug 2017. doi: 10.1103/PhysRevD.96.042002. URL https://link.aps.org/doi/10.1103/PhysRevD.96.042002.

[19] A. Aguilar-Arevalo, D. Amidei, X. Bertou, D. Bole, M. Butner, G. Cancelo, A. Castañeda Vázquez, A.E. Chavarria, J.R.T. de Mello Neto, S. Dixon, et al. Measurement of radioactive contamination in the high-resistivity silicon CCDs of the DAMIC experiment. *Journal of Instrumentation*, 10(08):P08014–P08014, aug 2015. doi: 10.1088/1748-0221/10/08/p08014. URL https://doi.org/10.1088%2F1748-0221%2F10%2F08%2Fp08014.

[20] E. Aprile, J. Aalbers, F. Agostini, M. Alfonsi, F. D. Amaro, M. Anthony, F. Arneodo, P. Barrow, L. Baudis, B. Bauermeister, et al. Signal yields of kev electronic recoils and their discrimination from nuclear recoils in liquid xenon. *Phys. Rev. D*, 97:

092007, May 2018. doi: 10.1103/PhysRevD.97.092007. URL https://link.aps.org/doi/10.1103/PhysRevD.97.092007.

[21] Alessandro Bettini. The world deep underground laboratories. *The European Physical Journal Plus*, 127:114, 2012. ISSN 2190-5444. doi: https://doi.org/10.1140/epjp/i2012-12114-y.

[22] Nick C. Buchholz and Philip N. Daly. The MONSOON Generic Pixel Server software design. In Hilton Lewis and Gianni Raffi, editors, *Advanced Software, Control, and Communication Systems for Astronomy*, volume 5496, pages 364 – 372. International Society for Optics and Photonics, SPIE, 2004. doi: 10.1117/12.550605. URL https://doi.org/10.1117/12.550605.

[23] A. Aguilar-Arevalo, D. Amidei, X. Bertou, M. Butner, G. Cancelo, A. Castaneda Vazquez, B. A. Cervantes Vergara, A.E. Chavarria, C.R. Chavez, J.R.T. de Mello Neto, et al. Search for low-mass wimps in a 0.6 kg day exposure of the damic experiment at snolab. *Physical Review D*, 94(8), Oct 2016. ISSN 2470-0029. doi: 10.1103/physrevd.94.082006. URL http://dx.doi.org/10.1103/PhysRevD.94.082006.

[24] E.W. Hoppe, C.E. Aalseth, O.T. Farmer, T.W. Hossbach, M. Liezers, H.S. Miley, N.R. Overman, and J.H. Reeves. Reduction of radioactive backgrounds in electroformed copper for ultra-sensitive radiation detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 764:116–121, 2014. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2014.06.082. URL https://www.sciencedirect.com/science/article/pii/S0168900214008365.

[25] A. Aguilar-Arevalo, D. Amidei, D. Baxter, G. Cancelo, B.A. Cervantes Vergara, A.E. Chavarria, E. Darragh-Ford, J.C. D'Olivo, J. Estrada, F. Favela-Perez, et al. Measurement of the bulk radioactive contamination of detector-grade silicon with DAMIC at SNOLAB. *Journal of Instrumentation*, 16(06):P06019, jun 2021. doi: 10.1088/1748-0221/16/06/p06019. URL https://doi.org/10.1088/1748-0221/16/06/p06019.

CHAPTER 3

DAMIC-M ELECTRONICS

3.1 Introduction

The DAMIC-M experiment has a targeted design goal of around 50 CCDs. Each CCD requires its own dedicated[1] bias and clock lines for image acquisition and charge readout, and up to 4 analog-to-digital converters (ADCs) to digitize the video output of the amplifiers on each CCD. These electronics must also be low noise, and capable of operating a CCD in skipper configuration stably for weeks (or months) at a time.

Existing commercially available systems cannot meet these requirements. Therefore, new CCD readout electronics for DAMIC-M have been in development for the past few years. This system, the Online Digital Interface for Low Noise Electronics ("ODILE") is designed with a primary motherboard card which contains a field programmable gate array ("FPGA") and digital communications hardware (to talk with the server and other ODILE boards). A high-speed mezzanine card (" HSMC") connector on the front of this motherboard allows for an ADC mezzanine card to be attached, and a pair of connectors on the back allow for connecting the CCD Clock And Bias Controller ("CABAC"). This modular design allows for development of the different components to occur in parallel, and for testing different ADC and CABAC electronics designs and revisions. The ODILE board is eventually intended to pair with a CCD Read Out Chip (the "CROC"), to form a CROC-ODILE pair. The CROC chip will be located near the CCD, and will provide enhanced video signal amplification and analog correlated double sampling (CDS) circuitry, with an interface powered through the CABAC card.

A diagram of the final CROC-ODILE design can be seen in fig. 3.1. The ADC digitizes the video signal from the CCD. The CROC provides either an amplified CCD video signal, or performs analog CDS (see section 2.3) and outputs the integrated result, depending on

1. Multiple CCDs can be run off a single set of bias and clock lines, but a malfunctioning CCD that produces glowing on the amplifiers could cause spurious charge in the entire experiment, and would need to be turned off, resulting in the loss of several CCDs (a problem that was encountered in DAMIC at SNOLAB). In addition, CCDs often have slightly different requirements for clock voltages for optimal performance. Having dedicated lines for each CCD is therefore vastly preferable.

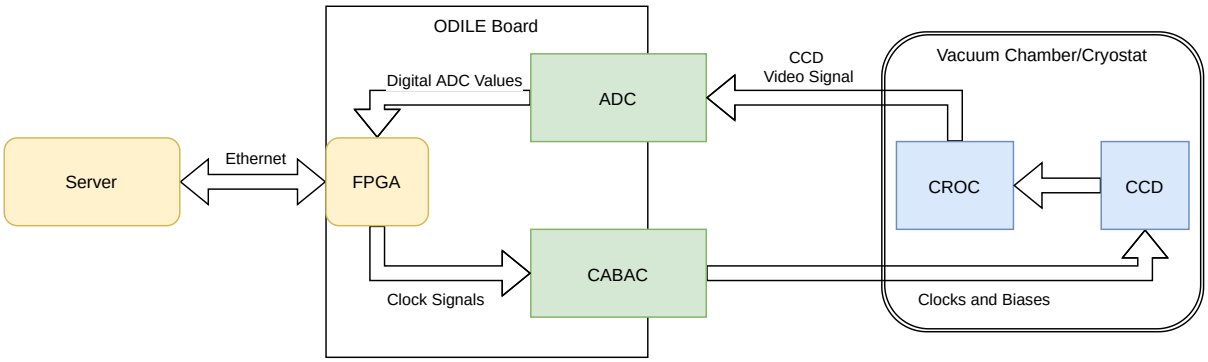


Figure 3.1: Simplified block diagram schematic of the final CROC-ODILE system design. Cold components are labeled in blue, mezzanine boards are in green, and programmable logic is in yellow.

operational mode. The CABAC provides all clock and DC voltages necessary to operate the CCD. The ODILE mainboard contains an FPGA with a programmable sequencer to operate the CCD, and communicates over Ethernet with the server.

The ODILE main board (pictured in fig. 3.2) contains an Arria V GX FPGA, which acts as a driver for all of the digital logic on the board (including the CABAC and ADC modules, and the serial I/O with the server). Multiple front panel connectors allow for digital I/O with the server or other ODILE boards, and the Versa Module Europa ("VME") connector on the back can be used for power (or optionally as another I/O channel). The board is designed with multiple power options: the main board can be powered with either a 4-pin connector or through the VME connector, while the CABAC and ADC boards can draw from either board power, or use their own front-panel connectors. The design also allows for split grounds, especially for the ADCs, which typically include optical couplers to allow complete electrical separation between digital and analog components, since digital clocks can be a significant source of noise in the ADC electronics.

3.2 Firmware

The FPGA that controls the ODILE is programmed through the use of a firmware file, designed and written in VHSIC Hardware Description Language ("VHDL"), and compiled with Altera's proprietary Quartus software. This firmware contains all of the logic for

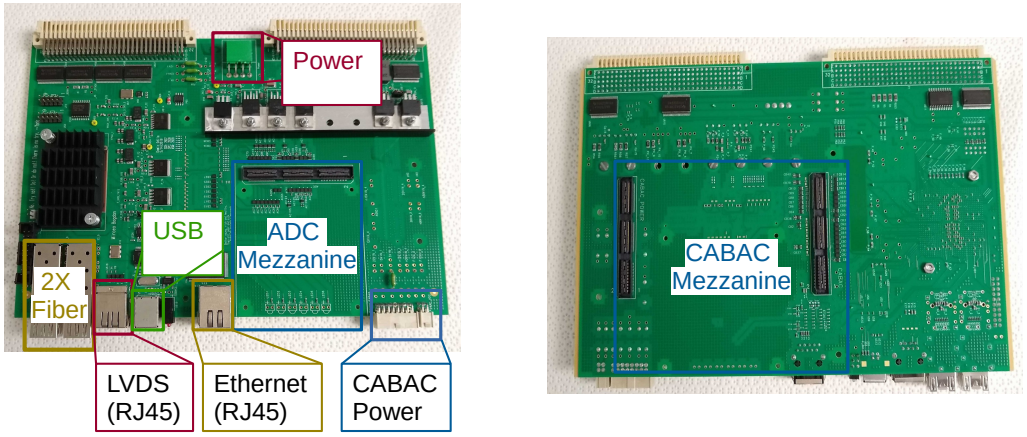


Figure 3.2: View of ODILE board from the front (left) and rear (right), with the different IO, power connectors, and mezzanine card slots highlighted.

server-side communication through the serial links, as well as drivers for ADC sampling and readout, digital CDS to integrate CCD video signal, and the CCD sequencer itself. It also contains Serial Peripheral Interface ("SPI") drivers to program the CABAC board voltages, and the CROC registers. Additional logic to monitor the CCD clock and bias values is currently in development.

3.3 Serial Interface

The ODILE contains several different physical interfaces for communication with the data acquisition ("DAQ") server. These are: 2 Small Form-factor Pluggable ("SFP") modules, an RJ-4 Ethernet port, and a Universal Serial Bus ("USB") 2.0 port. The USB interface is currently unused, but may be supported in the future.

The RJ-45 port is controlled by a Marvell 88E1111 transceiver chip that drives the magnetic interface, and therefore requires an Ethernet interface to use. The SFP modules, on the other hand, are driven directly by the FPGA, and can in principle support any arbitrary fiber-compatible protocol, up to a speed of 6 Gbps. These fiber modules are the preferred means of communication with the ODILE, as the optical nature of the communication allows for complete and total electrical isolation of the ODILE board. Given the widespread support of Ethernet, and to simplify supporting both the copper and fiber links, it was decided to use Gigabit Ethernet for both sets of links. This allows

for a flexible network topology (since ODILE boards can be connected to the DAQ server through a switch) and simplifies development of the DAQ-side system. Some further details of the Ethernet protocol can be found in appendix A.

The top level Ethernet block contains not only the firmware directly responsible for controlling the Ethernet interfaces themselves, but a large number of associated blocks, such as the remote update system and the main ODILE controller. It also contains a number of blocks that translate the standard 32-bit serial data from the Ethernet link into the data format expected by other top level blocks on the ODILE, such as the CCD sequencer or the ADC monitoring firmware.

The top level Ethernet block contains

1. Two copies of the fiber gigabit Ethernet block
2. The copper Ethernet block
3. The ODILE controller, which converts incoming commands into the appropriate signals
4. An Ethernet-CCDcontrol block, responsible for sending incoming data to the CCD-control block
5. A data arbiter, that handles sending data from the ODILE to the correct UDP port
6. A configuration register scanning block, which allows dynamically reading all configuration registers
7. The flash memory interface and remote update logic

The logical layout of these components can be seen in schematic form in fig. 3.3.

3.4 Ethernet Block

The Ethernet block itself comes in two varieties: one that contains an Altera TSE configured to control the fiber SFP modules for fiber Gigabit Ethernet, and one that is configured to talk with the Marvel 88E1111 chip for copper Ethernet. The remainder of the logic for the two block types is identical, with only some configuration settings being different. The components are shown in fig. 3.4.

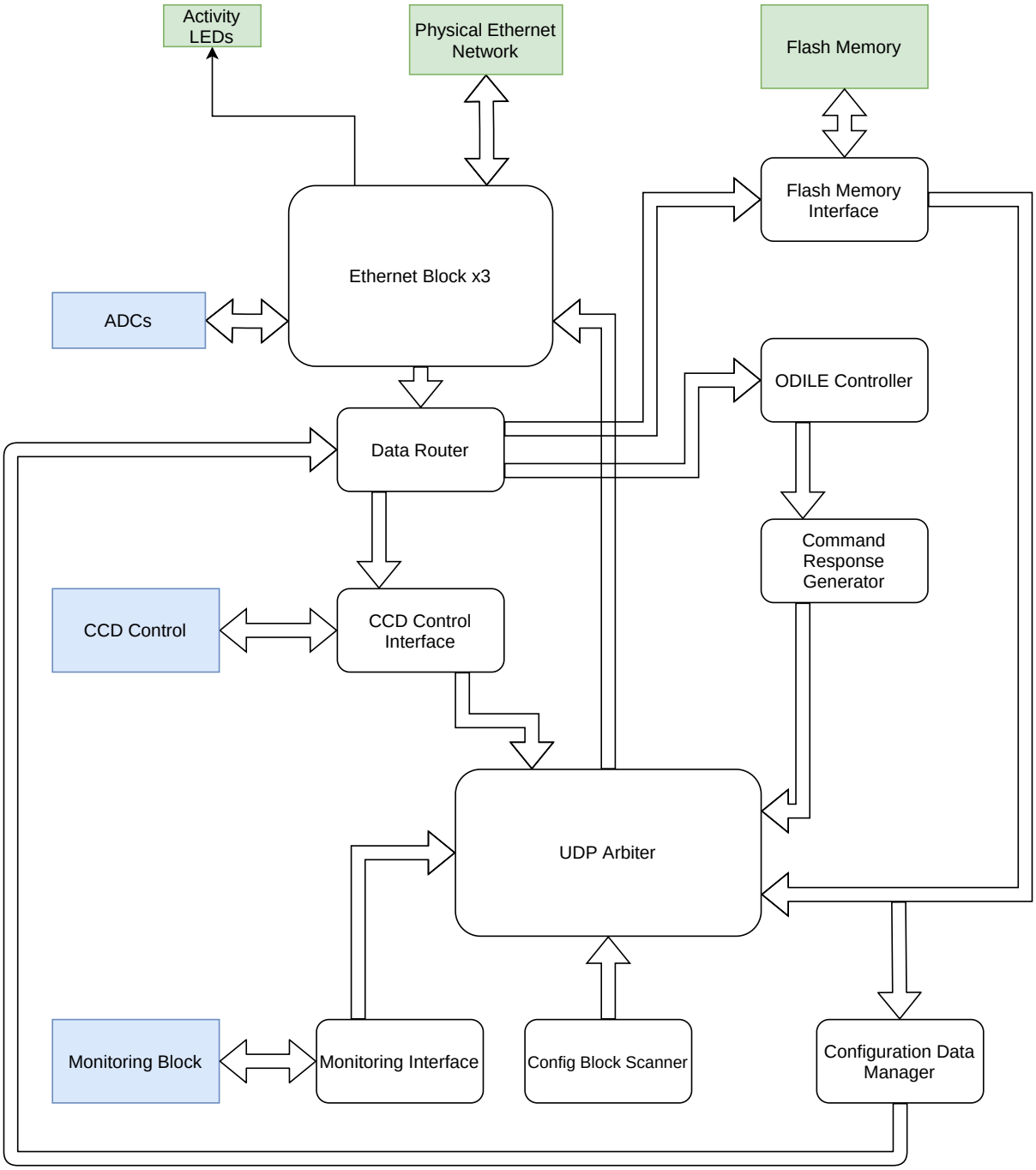


Figure 3.3: Schematic diagram and data connections between most components included in the top level Ethernet block. Arrows show logical directions of primary data flow. Other logical signals (such as the many signals from the controller, or the configuration data lines to various blocks) are not shown, so as not to crown the diagram. Blocks in blue indicate external logic, blocks in green indicate physical hardware directly connected to the Ethernet logic. The "Ethernet Block" contains all of the logic that directly handles the Ethernet interface, such as the frame generator or the Altera TSE MAC.

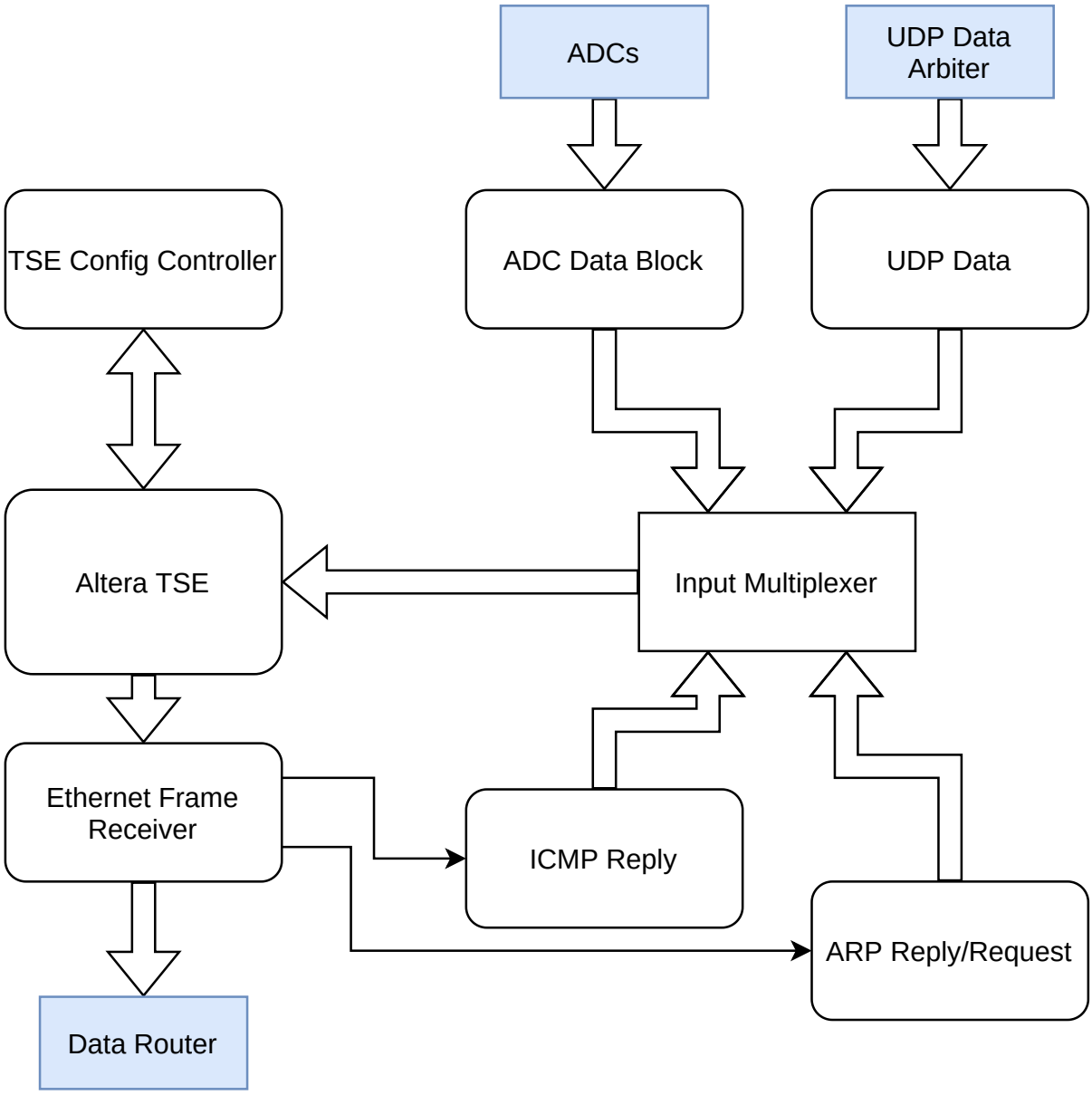


Figure 3.4: Schematic of data connections within the Ethernet block. Blue rectangles indicate connections to external logic. Thin black lines indicate trigger signals (these are driven by receiving specific types of Ethernet frames, specifically ICMP pings or ARP requests).

3.4.1 Altera TSE

The Altera Triple-Speed Ethernet (TSE) megafunction contains the logic for the Ethernet Media Access Control ("MAC") functionality, which is responsible for controlling the Ethernet hardware. It is capable of handling a huge variety of Ethernet features, only a handful of which are necessary for our application. Importantly, however, it handles nearly all of the hardware level interactions, from driving the physical hardware directly (in the case of the optical transceivers), to converting our 8-bit data words into the 10-bit words that are actually sent over the wire (8B10B encoding). A complete description of the TSE's functionality and parameters can be found in the user's guide [1].

For use in the ODILE, there are three groups of controls necessary. Those are the configuration lines, the transmit (or "tx") lines, and the receive ("rx") lines. Transmit and receive lines use an identical format: a 32-bit data bus, and 1 bit for each of the start of packet (SoP), end of packet (EoP), and data valid flags. Start and end of packet signals are used to delineate between different Ethernet frames, and data valid is used to signal that the other lines are valid. The configuration lines (used exclusively by the configuration controller described below) contain read and write request signals, an input address bus, and data in/out buses. The other important lines are the serial transmit/receive (txp/rxp) lines, which are used to send 1.25 Gbps serial data to the fiber transceivers or the copper physical layer ("PHY") driver.

The MAC also contains numerous other information lines, most of which are unused, with the exception of the LED status lines, used in generating an LED activity signal.

3.4.2 TSE Config Controller

The TSE config controller is responsible for setting the Altera TSE megafunction's various configuration registers to the correct values. The TSE MAC has a fairly large register space, since it can be configured for multiple Ethernet speeds and connection types (such as the Serial Gigabit Media Independent Interface (SGMII) we use for the copper connection, or 10/100 Mbps Ethernet speeds, which we do not use). It also supports multiple behavioral changes, such as automatically inserting the correct Ethernet frame check

sequence (FCS) at the end of each Ethernet frame, or magic packet detection. Most of these features are unused, but some configuration changes are still necessary for the MAC to function correctly. For a full description of the MAC register space, see the users guide [1].

The copper version of the TSE also contains a Management Data Input/Output (MDIO) bus capable of programming the physical layer (PHY) chip used for the RJ45 interface. This is especially important, because there are two physical interfaces with the copper Ethernet: the first, reduced Gigabit Media-Independent Interface (RGMII) has never worked successfully with the ODILE, so we use the second, the SGMII link mentioned above. The copper PHY must be programmed over MDIO to accept this signal before it will work, since the default setting is for RGMII. The MDIO registers on the PHY can be access through a dedicated sub-region of the MAC's configuration space. This same configuration space also serves to configure the optical fiber, since in that case the TSE MAC itself contains the PHY logic to drive the transceivers.

The actual configuration process is relatively straightforwards: the configuration controller sets the address and new configuration value on the data lines to the MAC and raises the "write" signal. It then waits until the "waitrequest" line goes low, indicating the process is finished, before proceeding to the next configuration register. In the case of MDIO programming, the controller first issues a read request for each register, before modifying the values using a series of "AND" and "OR" bitmasks (to leave bits not intended to be changed at their default values), following the recommended procedure described in the Marvell 88E1111 manual.

The configuration values set on both the TSE MAC, and the PHY over MDIO, have been determined through a mixture of reading the appropriate manuals, examination of the settings used in the Altera supplied testbench, and ultimately trial and error.

3.4.3 Ethernet Frame Receiver

The Altera TSE outputs Ethernet frames using a 32-bit wide serial data interface. In addition to the data itself, there are also start and end of packet flags (SoP and EoP,

respectively indicating when a frame is starting or ending), and data valid flags (used to indicate the other lines are valid). Since Ethernet packets multiple levels of header and payloads, some interpretation and unpacking of this data is required before any data can be forwarded to the appropriate block to handle. The Ethernet frame receiver handles this unpacking process.

The frame receiver will always read the incoming MAC and IP addresses and output those values, which can be used by other blocks in crafting any responses. It will also detect the presence of a IPv4/UDP packet (although it requires the IP header to be the standard length, with no additional fields), and read the UDP source and destination ports. The destination port is used by the rest of the FPGA logic to route incoming UDP data to the correct destination (the source port is unused). It can also detect both ARP and ICMP echo request ("ping") frames, and will output trigger signals to indicate the presence of either one. In the case of ICMP echo requests, the payload data will be forwarded to the responding block (as required by ICMP echo reply standard) [2]. The ARP block only requires the MAC and IP address information [3], which is read and output for all frames.

3.5 Ethernet Data Generators

The four remaining major sub-blocks within the Ethernet block all output complete Ethernet frames for transmission, using the standard 32+3 bit format. These outputs are multiplexed and passed through to the TSE MAC for transmission. In order to format complete Ethernet frames, each block includes 3 components: the Ethernet frame generator, a header generator, and some data source. While each block contains it's own frame and header generator, the code for them is identical, and only the data source (specific as it is to each block) is unique.

3.5.1 Frame Generator

Key to all Ethernet data generation is the frame generator. This block is responsible for generating the complete Ethernet frame (except the CRC32 at the end of the frame, which

is inserted automatically by the MAC). It does this by interfacing between the header generator, which generates not only the Ethernet frame header, but also the headers for IP/UDP/ARP blocks, and the data source, which is typically a First-In First-Out ("FIFO") buffer, but can also be any source with the correct 32-bit plus read request interface. The length of the Ethernet frame is determined at the start of the frame, using the "usedw_in" bus (with a maximum of 1500 bytes per frame, to obey the Ethernet standards). The frame generator does not technically need to know the length of the frame prior to starting frame generation, but the IPv4 header contained in the payload of most frames does, so it is considered easiest to wait in all cases. The input buffers for all data sources are deep enough that this does not result in a performance penalty or any potential for data loss.

The output of the frame generator is a 32-bit plus start/end of packet and data valid flags. The only directly configurable settings are the maximum payload length, and the delays on the input/output interfaces (to allow for variable delays on input FIFOs or other data source).

The other main component is the header generator. This mainly interfaces directly with the frame generator, but is configured by the external source with the correct MAC/IP addresses for both destination (i.e. the server) and source (i.e. the ODILE itself). It supports generating four header components: the Ethernet frame header, the IP header, the UDP header, and an arbitrary 32-bit application header (which is inserted as the first word of the payload). This application header is currently not used by any of the ODILE logic.

Note that when generating UDP headers, the header generator sets the checksum to "0x00_00", to simplify the header logic (this is allowed by the UDP specifications). It will however compute the IPv4 header checksum, as this is required by the IPv4 specification [4].

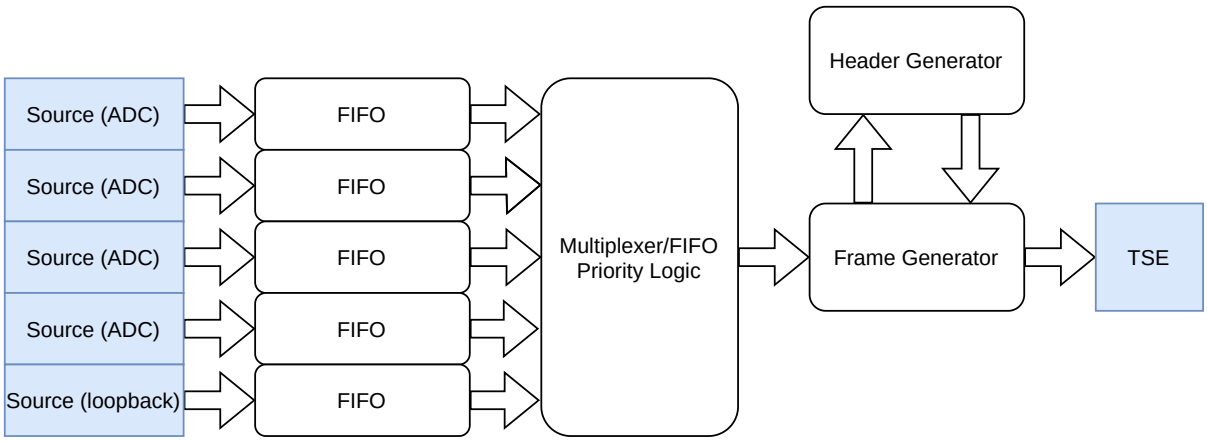


Figure 3.5: Schematic of ADC data block. Blue indicates external sources/destinations.

3.5.2 ADC Data Block

The primary source of data on the ODILE has always been the ADCs, whether the values are being read directly as an ADC data stream (allowing the ODILE to act as an oscilloscope), or after some processing, such as using digital CDS to convert the CCD video stream into a pixel stream. In either case, the data is expected to come as a continuous stream of independent values, each representing some value of the CCD signal. Since there are four such ADCs in all current ODILE mezzanine designs, the logic to handle these data streams must be able to accommodate at least 4 streams, all of which may be enabled simultaneously, and to transfer them from four parallel streams into a single 32-bit wide serial interface expected by the Ethernet transmission block. By default, this block includes inputs for 5 streams as shown in fig. 3.5. 4 of them are attached to the ADCs, and one is currently designated as a loop-back interface that sends any data received back to the sender, which allows testing the Ethernet functionality and throughput.

For testing purposes, all of these streams can be switched instead to a simple integer counter, to allow testing the serial interface. The speed of these counters can also be varied, which can be used with the loop-back interface to test the maximum throughput of Ethernet link (this has been tested to be 114 MBps).

The ADC data block (called the "ethernet_data_block" in the code, though it only handles part of the Ethernet data input) handles this by means of multiple independent

FIFO buffers. These FIFOs are dual-clock, allowing them to be driven by an external clock synchronous to the ADC data. They can then be read by the clock synchronous to the Ethernet logic, allowing the incoming data to operate at any speed (this is important since the ODILE design allows for different ADC mezzanine cards, which can operate at different sampling rates). The buffers also include write request lines, allowing data to only be written intermittently (such as after applying the CDS filtering or skipper averaging).

To read data from these buffers, some additional logic (called the "input_fifo_manager" or "IFM") compares the number of words contained in each buffer, and if any exceeds the programmable threshold (default 300 words), it will raise a signal to the frame generator to begin reading from that buffer.

The IFM includes two control flags for each buffer. The first enables or disables a specific buffer. Note that this will disable both reading and writing from the buffer: any data that is written to this buffer will simply be ignored. The second is a priority flag: any buffer with this flag enabled will be prioritized, even if it not the fullest. The minimum packet size is still respected, even for this priority flag. The use of the priority flag is discouraged, as it may result in unintended behavior or data loss, but is included for future use in case it may be useful.

The IFM will also automatically read ("flushed") from any "stale" FIFO, which is determined to be any FIFO which contains data, but has not been written to for some time (this is compile time parameter, set to approximately 1.25 ms[2] by default). This stale flush also takes priority over other data, similar to the priority flag above. This flushing is important in case of any changes to the incoming data stream: for example, switching from a constant ADC data value read to a triggered CDS value would cause the buffer to contain some raw ADC values, and some CDS values, which would be unexpected to the server. This can still happen if the new data stream is continuous with the old, but the server is expected to be able to handle (or ignore) such discontinuities.

2. Technically, this value is 125000 clock cycles, with a 100 Mhz clock

3.5.3 ARP/ICMP Blocks

The next two sub-blocks provide some additional support to enable standard Ethernet tools. The first is a block to handle the Address Resolution Protocol or ARP. This protocol is explained in more detail in the section A.2, but it is required for normal Ethernet networks, and so support for it is built in to the ODILE firmware.

Specifically, this block integrates with signals from the frame receiver to detect incoming ARP requests, and generate a reply to the correct MAC and IP address containing the current ODILE MAC address. The only required configuration inputs are the source and destination IP and MAC addresses, and signals that an incoming request requires a reply. Newer version of the firmware are also able to handle ARP *requests*: rather than using a MAC address loaded into the configuration registers, the ARP block can send out an ARP request, and the corresponding reply is stored in a special block that caches the reply. This cache is also intelligent enough to detect when an Ethernet frame is incoming from the server IP address, and will store the source MAC address from that frame.

To help eliminate any issues, this ARP cache is only considered valid for around 60 seconds. This timeout is refreshed by any packets incoming from the server. If the timeout is exceeded, an ARP request will be issued from the ODILE to update the cache. This relatively short timeout increases the reliability of the cache, with only a very minor performance overhead. The ARP request logic may be disabled entirely with a configuration setting, in which case the ODILE will fallback to the MAC address contained within its configuration registers.

The second sub-block provides replies to Internet Control Message Protocol ("ICMP") echo requests. The "ping" utility is a standard network tool that uses these echo requests to quickly diagnose network connectivity issues, and while technically optional, is often expected as a *de facto* requirement (failure to respond to pings is considered a sign the device is not working). The ICMP block uses the same input data as the ARP block: the MAC and IP address of the message source, and a signal to indicate that a reply is expected. It also requires the payload data from the ICMP packet, as the protocol requires the response to contain an identical payload. Note that the ICMP protocol

contains many other diagnostic and control messages which are not implemented.

3.5.4 UDP Data Block

The final source of Ethernet data is the UDP data block. This block is general purpose, and can generate frames addressed to any UDP, based on the inputs to the block. This block requires as inputs both the destination MAC and IP addresses and the data to send, but also the UDP port to send the data to. This port addressing is used to route the destination data on the server, similar to how data sent to the ODILE is routed based on the destination UDP port (and in fact as a general rule data sent to and from a specific port corresponds to a specific block within the ODILE).

This block is the second most complex data block, as it must contain a buffer to store not only the data itself, but also information about the packet[3]. Since incoming data may switch from one source to another (by changing the input UDP port), to properly create Ethernet frames, it stores the UDP port until either a) the port changes, indicating a new source is attempting to transmit, b) the data exceeds the maximum packet length, or c) the end-of-packet flag is raised. In either case, generation of an Ethernet frame is triggered (any new incoming data will continue to be written to the buffer). Like the ADC data block, this logic also includes a timeout to prevent data from going *stale* inside the data buffer.

3.6 Associated Logic

In addition to the firmware associated directly with sending and receiving data over the Ethernet links, some associated firmware logic is contained in the top Ethernet block. These blocks are responsible for translating data from the Ethernet into formats or signals used in the rest of the ODILE.

3. Unlike the ADC data block, we only have one data buffer, rather than one per potential data stream, since there may be many different source of UDP data on the ODILE

3.6.1 ODILE Controller

The ODILE controller interprets commands received by the Ethernet block into signals that trigger various functionality on the FPGA. This controller is only used for non-realtime signaling that does not require precise timing, as the Ethernet interface has a relatively high and non-deterministic latency. This block has no inherent link to Ethernet and could be easily adapted to USB or any serial interface with a 32-bit data width.

The controller operates using relatively simple logic: whenever valid data is received on the correct UDP port (port 17000, aka 0x4268), it checks a lookup table for that command, then performs the relevant function. Usually, this is to raise a trigger signal for some other block to perform a function, such as reading from a memory, or performing a firmware reconfiguration operation.

The actual commands used for ODILE instructions are 3 bytes long, typically formatted using ASCII abbreviations of the command. For example, the sequence "GCT" (in hex, 0x474354) tells the ODILE to "Get Compile Time", which will return the time (in 32-bit UNIX epoch 0format) when the firmware was compiled. The command structure uses the bottom 3 bytes of the 4 byte words from the Ethernet interface. The upper byte can be used as an optional extension to some commands to specify additional information used in the command (for instance to set the values of 8 hardware switches in the CCD control block). Commands that require more than 8 bits can use a full second 32-bit word. For example, this is used to specify the address of the start location of the firmware on the flash memory for update commands.

3.6.2 Command Response Generator

A closely connected block to the controller is the command response generator block. This is responsible for generating responses to commands to provide feedback to the server. Incoming commands are forwarded to the response generator, and responses are generated through a simple state machine. Most commonly, this is a generic "DON" reply to acknowledge the command was completed successfully, or an "INV" to report the ODILE received an invalid command. This block also contains logic to generate

responses to simple commands, such as responding with the compile time or the ODILE uptime. All data generated by this block is sent back to the control UDP port.

3.6.3 UDP Data Arbiter

While each Ethernet block contains dedicated lines for inputs from the ADC modules (since these are expected to send data nearly continuously), they only contain a single general purpose UDP data line that can transmit Ethernet data to any arbitrary UDP port. As a result, only one source can transmit data to this line at a time. There are multiple different modules on the ODILE that can generate such data (these modules are termed "clients" in the rest of this section). Therefore, some logic is necessary to control access to this line to prevent race conditions or data loss. This is the responsibility of the "udp_data_arbiter" block.

As the name implies, this block is responsible for arbitrating access to the UDP transmission lines. It uses a request-based access system, so that clients that require transmission access can request control of the UDP data lines. Clients may request access to send UDP data, and must wait for the arbiter to signal it is ready before sending data.

Each client interfaces to the data arbiter using a 52-bit input bus, and a 1-bit output signal. The lower 32 bits of the input bus contain the data to be transmitted. The next 16 bits contain the UDP port to send the data to (this is constant for most clients, but left as an input for any clients that may wish to change it).[4] The upper 5 bits (from bit 48 to 52) are data flags. Bit 48 acts as a data valid flag, indicating the data on bits 0 to 47 are valid. Bit 49 indicates an end-of-packet request by the client, to force the Ethernet block to send any new data in a new packet (this is not currently used by any client). Bit 50 is the transmit request line, and bit 51 is the transmit busy line, indicating the client wishes access to the UDP interface, and is currently or about to transmit data, respectively. Bit 52 is unused and left free for future use.

The data arbiter uses a simple state machine to control access to the UDP transmission

4. The firmware synthesis tool should optimize away any constant valued signals, so there will be no waste of resources if the port from a specific client is always constant.

lines. This state machine waits until any client request access to the UDP lines. If no other client is using the data lines, it sets the requesting client as the current client, and sets the "udp_ready" signal for that client to high, indicating the interface is ready to be used. Data sent to the arbiter then will be passed through to the Ethernet blocks. Note that the client has 1000 clock cycles to acknowledge the request by raising the respective "tx_busy" line, or the arbiter will reset itself to an idle state and the client will need to request access again. The arbiter will pass through data as long as the respective "tx_busy" line is high with no timeout.

UDP data can be sent to individual interfaces using the "dest_iface" signal. This acts as a simple bitmask: if the bit values for any of the 3 interfaces is low, UDP data is not written to that interface's data buffers. The IP/MAC address of the destination server can also be set using the "dest_addr" bus, which will control where the UDP data is sent. Note that these signals are not arbitrated, but instead take the values from the ODILE controller, based on which interface received the data and what server sent the data. Clients currently have no knowledge or control over the destination of their data: the assumption is that the destination for this data will be whatever server on the network last issued a command to the ODILE (this is unlike the ADC data, which is sent to the configured server address).

3.6.4 Configuration Register Block

Many of the blocks on the ODILE contain configurable parameters, such as the CCD sequencer, or the address of the Ethernet interfaces. Some of these blocks use their own dedicated systems, such as the CCD sequencer, which uses a number of memories, or the physical hardware like the CROC, which is configured over an SPI interface. However, many blocks require only a handful of configurable settings. There are many such blocks, and while they could be configured using a centralized block of registers, this would require many signals to be passed through multiple levels of logical blocks.

Instead, we developed a flexible and general-purpose configuration register block that can be used internally in any component, programmable through a standard 32-bit plus

Bit	31	30:34	24:16	15:0
Meaning	Write/read	Block Address	Register Address	Register Data

Figure 3.6: Format of configuration words.

data valid interface. Each of these 32-bit configuration words includes both the block and configuration register address that it is intended for, allowing this data to be sent to all the configuration blocks, without requiring any additional routing logic outside the blocks themselves. Each block can contain up to 254 16-bit registers, and the block addressing system allows up to 128 unique blocks on the ODILE.

The format of these 32-bit configuration words is shown in fig. 3.6. They are arranged as follows: the first bit is a read/write flag (write=0, read=1). The next 7 bits specifies the block address. Each block is given a block address when it is instantiated, and will ignore any data addressed to a different block. In principle, it is possible for multiple blocks to be instantiated with the same address, to allow for multiple blocks to be configured through a single configuration instruction. The next 8 bits specifies the configuration register that is being set, and the final 16 bits contains the new values for the configuration register.

The current values set in configuration blocks can be read through the same 32-bit plus data valid interface. Configuration registers can be read individually, or an entire block can be instructed to read all of its configuration registers by setting the register address to all 1s (i.e. "0xff"). Data is read from the registers with the read/write flag set to 0, to allow the registers to be programmed using this serial data with no alteration.

The configuration block contains logical signals to indicate if there is a read/write error (which can occur of the specified register is outside the valid register range for that block), as well as a "config_changed" signal, which will go high for several clock cycles if any configuration register in the block has been changed.

Both the initial values and number of configuration registers in a block are set at

compile time by setting the "DEFAULT_SETTINGS" array generic to the default values[5]. Due to a legacy design issue, these default settings are passed in as 32-bit registers. Likewise, the output register values are output as 32-bit wide registers, composed of 2 16-bit registers.

The configuration blocks contain both 32-bit data + 1 bit data valid input, and the same in output. The output is used in case of read commands, and is sent in the same identical format.

3.6.5 Configuration Register Scanner

It is possible to execute a read command to configuration registers simply by sending them configuration words with the 1st bit set to a "1", and the correct block and register address (or the register address set to all 1s to read the entire block). However, reading this data back to the server requires additional logic on the FPGA. Moreover, we prefer (where able) to condense read command logic into the standard command and control scheme described in section 3.6.1. We therefore developed a block whose sole purpose was to conduct reads of the configuration registers described in section 3.6.4, called the "Configuration Register Scanner". The logical design of this block is shown in fig. 3.7.

This scanner is capable of operating in two modes, either reading all values from a single configuration register block, or scanning over all the block address space to read all configuration blocks on the ODILE. Both modes operate quite similarly: first, the block issues a request to the UDP transmission pipeline for access to send data. Then, a read instruction is issued along the configuration pipeline (by issuing a configuration word with bits 31 and 23 through 16 set to 1s, and the block address in bits 30 to 24: this will be either the specified address in a single block scan, or 0, in a full scan). The block then waits for a response from the configuration block (we wait for 15 clock cycles, as the latency to and from the configuration blocks may be quite high, but should not exceed this). If it receives none, it assumes the block does not exist, and continues on. If

5. This allows changes to the number of registers in the block by simply extending the length of this array. The block will automatically recompute it's size. Any unused register bits should be automatically trimmed by the compiler, as long as they are not used by any logical blocks.

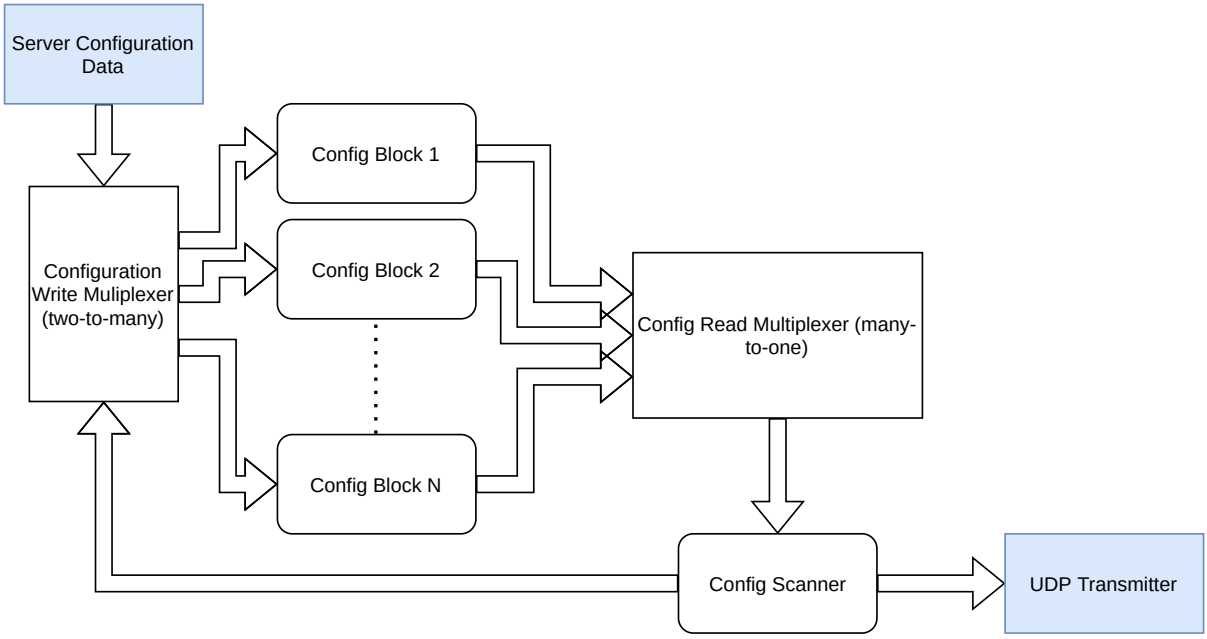


Figure 3.7: Schematic of configuration data pathways. The configuration blocks are only programmed by data incoming from the server over Ethernet: the lines from the scanner block exist only to issue read commands. This data is then branched out to all of the configuration blocks on the design (of which there may be up to 128). Read commands result in data being sent from these blocks to the scanner, which then forwards that data to the server.

it receives valid data (with the correct block address), it forwards that data to the UDP transmission pipeline to send to the server.

In the case of a single block read, once done, the block returns to idle mode. If the scan is over the entire block address, it will release the UDP transmission lines for 8 clock cycles, then request access again, before proceeding to the next block address, and issuing a request to read that block's entire configuration space. It repeats this process until it reaches the full range of the block address space (all 128 blocks). Reading a single block may take up to 2.7 microseconds (when operating on a 100 MHz clock). Reading the entire block address space could take up to 346 microseconds. This scanning schema allows the scanner to read every block in the ODILE automatically, without requiring any knowledge about how many blocks exist, or how large each block is. It will be slower than reading only existing blocks, but only slightly, and configuration reads/writes are not considered time-sensitive operations, so this should never be an issue.

Note that this scanner could theoretically overwhelm the Ethernet transmission pipeline, by generating data at nearly 3 Gigabits/second. So far, no logic exists in the

transmission pipeline to prevent this. In practice, it typically only generates data in very small bursts (which can be easily handled by the Ethernet FIFOs, which are currently 32 KB in size), but if the ODILE design ever utilizes most of the available address space, this could become an issue (this is easily solvable by increasing the wait time between block reads).

3.6.6 Flash Memory Interface

The ODILE uses a 32 megabyte "EPCQ-256" flash memory device [5] for all persistent data storage. This device can be used for data storage, but is primarily intended to store the firmware images for the ODILE FPGA. It can be accessed over dedicated pins using an Altera-provided megafunction (the "EPCQIO" block), which provides low-level access to the flash.

The nature of the EPCQ flash allows for read and write operations to both occur at the level of individual bytes. However, a byte can only be written to properly if it is first erased. The erase process sets the flash bits to high value ("1"s in binary), and byte-level write operations can only set bits to low ("0"s), not back to high. Erase operations, unlike reads and writes, can only be performed at the level of flash *sectors*, which are each 65,536 bytes in size (there are 512 of these sectors on the EPCQ256).

To simplify access to the flash, we use a wrapper function, called "epcqio_control". This controller interfaces with the odile_controller block to simplify the read/write process. It contains 2 512-word FIFO buffers for reading/writing data. Data can then be read or written to the flash by specifying the start address and the number of 32-bit words to read or write, then raising the read_data or write_data signals to initiate the read/write procedure. Data read from the flash will automatically be sent to the Ethernet block's UDP interface for transmission to the server.

A flash sector can be erased by specifying any address within the sector on the address bus, and then raising the "erase_sector" signal.

All operations cause the respective busy signal to go high for the duration of the operation. New instructions should not be issued to the block during this time (they

will be ignored). New data may be written to the write buffer irrespective of the current status, however care should be taken not to overfill the buffer (the "write_buffer_full" signal may be used for feedback to indicate the write buffer is full). This signaling is especially useful for writing data from the Ethernet interface, as the interface will send back a "DON" message when the flash finishes the current write or erase command, meaning it is safe to continue with further write operations.

Since the EPCQ flash operates at a maximum 20 MHz clock, the EPCQIO block handles sending data and read/write commands across clock domains automatically, using a 4-stage synchronization process for commands, and dual-clock FIFO buffers for data. There are two state machines, one that runs on the Ethernet logic clock, and one that runs on the flash interface clock. The former handles transmitted read/write/erase clocks across clock domains with the latter, which handles talking with the Altera megafunction to program the flash. The latter also converts the native 8-bit words that the flash uses to the 32-bit words used by the rest of the FPGA logic.

3.6.7 Remote Update Logic

Closely related to the flash memory interface is the remote update logic. The Arria V GX contains dedicated circuitry to load new firmware images from the flash memory, allowing the firmware to be updated remotely without requiring the use of the dedicated JTAG or serial programming interfaces.

The design of the remote update logic allows for two (or more) separate firmware images. These are typically separated into the "factory" firmware, which always begins at the start address of the flash (address 0x00000000), and the "application" firmware, which on the Quartus version we used may start at any address with the last 8 bits are zero.

In practice, the Arria V GX images are 22 MB in size, which would only allow a single image to be written to our 32 MB flash at a time. Fortunately the FPGA includes dedicated decompression logic, which allows us to compress these images. The compression yields variable size images, but the worst-case compression is given as 35% [6],

reducing the image size to < 15 MB (typically, the compressed image is around 11 MB in size). This allows us to program two firmware images to the flash memory, with a factor firmware at address 0x00000000, intended as a safe fallback that is not changed or updated after initial programming, and an "application" firmware starting at address 0x1000000 (i.e. halfway through the flash) that can be safely updated using the flash memory interface specified in section 3.6.6.

This memory arrangement leaves 1.5 MB at the end of the flash unused by firmware files and available to be used for arbitrary purposes (in principle, the 1.5 MB between application and firmware images is also available, but currently unused). These region is used for storing persistent configuration data for the ODILE, as described in section 3.6.8.

To perform a remote update, the binary firmware image must first be converted to a "raw programming data" file (.rpd) format using the Quartus software. Along with this should be a .map file that contains the end address of the firmware image (effectively, this is the size of the compressed firmware image in bytes). Then, this firmware image should be written to the flash, starting at either the application image start address (preferred), or the very start of the flash (to update the factory firmware). This can be done with the flash memory interface described in section 3.6.6.

In order to perform the firmware update, the update logic only needs to know the start address of the new firmware image, which can be loaded with the appropriate 2-word command. The firmware can then be loaded by issuing a "RUR" command, which will start the update process. The remote update controller will then load the new firmware address into the Altera megafunction, along with setting (and enabling) the watchdog timeout, before instructing the megafunction to load the new firmware. This process takes a few seconds, after which the new firmware should be ready for use.

The update controller also includes some logic for determining if the firmware is in "factory" or "application" state. However, this logic does not seem to function reliably, and should not be used for any significant purposes without further testing.

Byte	0	1	2	3
0	0xCD	Data Size	UDP Destination Port	
4-255	Data			

Figure 3.8: Format of generic flash configuration data. The UDP routing included in the header allows the data to be sent to any other block on the FPGA. The number of words is required to ensure only valid data is loaded, and the first byte header is used to indicate the data block contains some valid data.

3.6.8 Configuration Data Manager

In order to store persistent settings on the ODILE flash memory, we use the last 10 sectors of the flash memory. Each sector is designated a *page*, numbered from 0 to 9. Page number zero is automatically loaded on firmware power on, and can therefore be used for persistent power on configuration of the ODILE board (such as for e.g. setting a unique Ethernet MAC address).

Each of the configuration pages is further subdivided into blocks of 256 bytes (or 64 32-bit words), with a format shown in fig. 3.8. The first word of each block is used as a header for the block, for 3 purposes: The first byte should contain the hex value "0xCD" to indicate the block is valid. If it does not, the manager assumes it has reached the end of the configuration page, and will stop reading. The next byte indicates the number of words in the block that are used (including the header). The final 2 bytes indicate the UDP routing port for the rest of the block. This allows us to store settings or configuration data for any other block on the ODILE board, so long as that block is reachable through UDP port addressing.

The configuration data manager interfaces directly with the EPCQIO controller to handle reading the correct byte addresses. It does not perform write operations: it is the responsibility of the server side code to handle writing configuration pages to the flash correctly. Loading configuration pages can be done by specifying the 4-bit configuration page (valid values are from 0 to 9), then raising the "load_config" signal.

To minimize any additional logic or routing, the data manager hooks into the standard

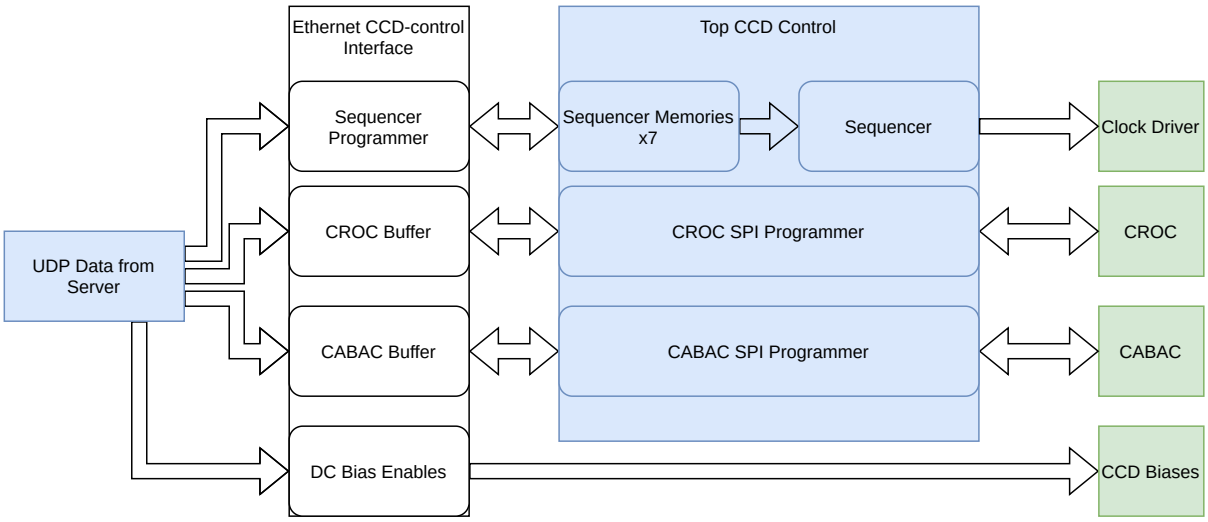


Figure 3.9: Schematic of interface with top level CCD control block.

flash-to-server communication to read data from the flash, and (after parsing the header) sends data to the standard UDP data router to forward it to other blocks.

To avoid the firmware from locking up in case there is an error in the flash reading process, the manager state machine will timeout if it does not receive data after a hard-coded time period (10 ms, by default).

3.6.9 Ethernet CCD-Control Interface

Some additional logic is required to interface between the 32-bit serial format that the Ethernet blocks produce, and the various data formats required by the CCD control logic. The CCD control has three main components: the sequencer, the CROC, and the CABAC. The logical layout of these blocks is shown in fig. 3.9.

The first is the CCD sequencer, which is drives the timing of the clock signals for the CCD. The sequencer itself is contained entirely within FPGA logic, and uses a 32-bit wide output to drive the various clocks for the CCD. The sequencer is a state machine driven by a custom instruction set designed for CCD operations, with basic CPU-like functionality, such as function calls and address jumps. Programming the sequencer uses seven memories, containing the sequencer instructions, with each memory serving a different purpose. Only three of these memories (the program, timing, and sequencer output) are essential to the basic sequencer functionality, but programming all of the

Bit	31:16	15:10	9:0
Word 1	Destination Memory	Unused	Memory Address
Word 2	Data		

Figure 3.10: Format of sequencer programming data words. The destination memory is optional (and ignored) if data is routed directly to the UDP port corresponding to that memory.

memories is supported, in case more advanced functionality is required. A full description of the sequencer is beyond the scope of this work: see [7] for details of the sequencer design and operation.

The sequencer memories are programmed through an interface that specifies the memory address (10 bits wide), the data to write to the memory (32 bits wide), and a write enable flag for each of the seven memories. The CCD control interface therefore needs to translate a 32-bit serial data signal into a 32 + 10 + 7 bit data format. This is done by using sets of two 32-bit words, the first containing the memory address in the bottom 10 bits, and the second containing the data. The memory to be written to can be specified in one of two ways: either by sending the data to a specific UDP address corresponding to that memory, or by adding the same UDP address to the upper 16 bits of the first word, and sending the programming data to the "sequencer serial program" UDP address (this latter method is preferred, as it allows the entire sequencer to be stored in one single stream of data, and programmed with a single write operation). Programming operations are conducted automatically when data is sent to the appropriate UDP destination port, without requiring any additional commands from the server. See fig. 3.10 for a graphical representation of this format.

The CCD control interface is also capable of both erasing these memories (which will set the value of all the memories to 0s), and reading the current values of a specific memory. Both erase and read operations are performed through standard ODILE commands. Data read from the sequencer will use the latter method specified above (and will send the read data to the same sequencer serial UDP port on the requesting server). It is recommended when performing a sequencer programming operation to first conduct an erase, as memories may contain values from previous sequencers that could interfere with

the sequencer, and to read the sequencer memories afterwards, to validate that they were programmed correctly.

The second component ran by the CCD control logic is the CROC chip. The CROC itself is programmed through an SPI link controlled by the CCD control logic. The control block receives data to program to the CROC through a 96-bit wide register (one input to write to the CROC, and one output that contains the last-read value). In addition, it contains a trigger signal to initiate a write to the CROC over the SPI link, which will take the current value of the 96-bit input register and write it to the CROC.

The final component is the CABAC chip, which is is responsible for applying clocks and bias voltages to the CCD. The external interface to the CABAC is similar to the CROC, with 32-bit input/output registers for data, and a write trigger signal. An additional signal to perform a reset of the CABAC is also available.

The Ethernet-CCD control interface talks to both CROC and CABAC with the same method: a buffer holds data values to be written to the component, and when that buffer contains enough data (either 32-bits, for the CABAC, or 96-bits for the CROC), the input register on the CCD control block is filled with that value, and the write signal is triggered. This process is triggered only by writing data to the corresponding buffer (addressed by UDP port): no additional command is required. The CABAC reset, on the other hand, is triggered using a standard ODILE command word. Likewise, reading the current value on the output registers is done with a command word. This assumes the current value is valid: at present, that is the responsibility of the server.

3.6.10 CCD Monitoring Interface

The final block contained in the top level Ethernet is the monitoring interface. The CABAC board is designed with several ADCs to sample the current voltage on the clock and bias lines (to provide monitoring and alarm generation in case of a hardware failure). This block requires a signal to begin the ADC readout process, and outputs a 18x32bit array containing the ADC values. The monitoring interface is responsible for converting this array into a serial stream of data for the server, and outputs data to the UDP arbiter

using the same transmit request standard defined above. The start of monitoring and readout process are both independently triggered by the controller block (the monitoring block will not read out register values on it's own after issuing a monitoring instruction).

3.7 VHDL Simulations

All of the above firmware was written in industry-standard VHDL. Creating a firmware image that can be programmed onto the FPGA requires a lengthy compilation process (since the VHDL code must be converted into a hardware equivalent, then mapped onto the FPGA hardware with signal routing to achieve the timing requirements of the design). Compilation requires around 1 hour even on a fairly powerful computer. Beyond that, testing the design on the FPGA itself can be tricky, since very little internal diagnostic information is available. The primary debugging tool on the FPGA itself is SignalTap (an example of which can be seen in fig. 3.11), which allows the acquisition of signals internal to the FPGA. This is useful to probe the internal state of the FPGA, but requires forethought, since the signals must be added prior to compilation, and still only offers limited information about the FPGA processes.

It is therefore standard practice to create a simulation of designs prior to attempting to compile them on the FPGA. This allows the running of the design in software, using simulated signals and clocks. In our case, this simulation only occurs at the "RTL" level, which tests the logical functionality of the design (and does not simulate the physical hardware itself: this is possible, but more difficult, and much slower). This RTL level simulation allows for fully probing the internal state of every component in the design, and rapid iteration or modification, since the simulation process takes seconds, compared to the hours required to compile and upload the design to hardware.

We therefore wrote simulations in the Modelsim toolkit for nearly all the modules listed above. An example of a simulation of the Ethernet-CCD control interface can be seen in fig. 3.12. Only very simple blocks or very minor changes are compiled and tested on the FPGA without first verifying the design in a Modelsim simulation.

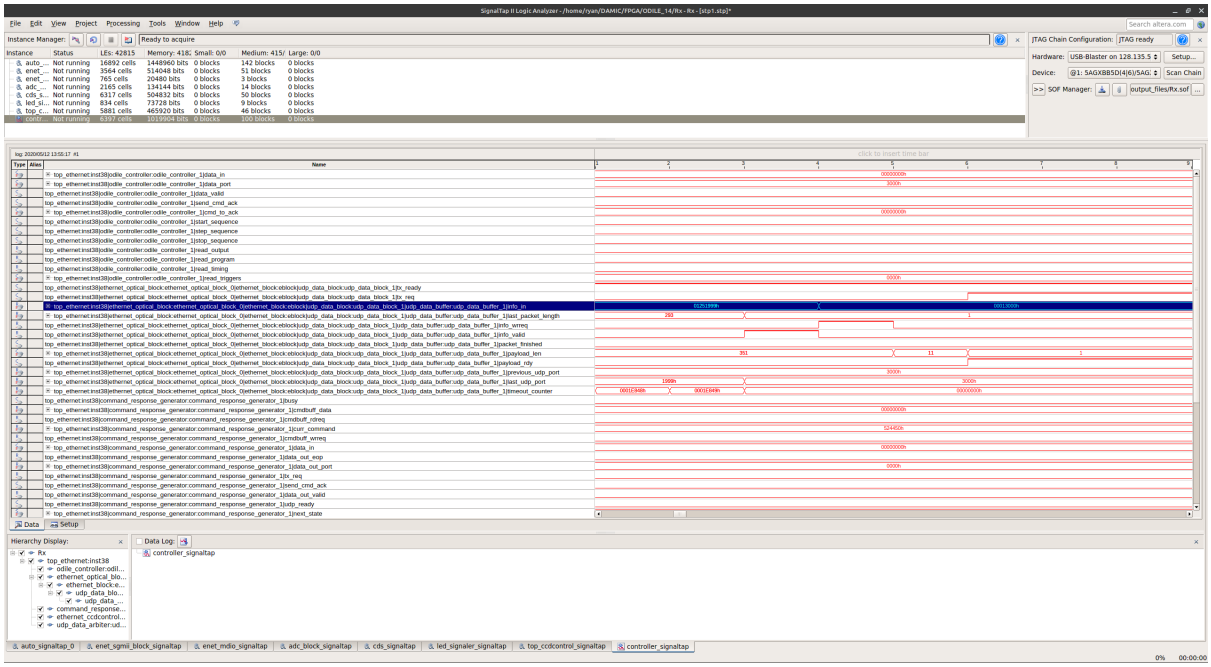


Figure 3.11: Example of a set of data acquired with SignalTap, showing the state of the UDP data arbiter near the end of a sequencer memory read operation. This information is highly useful in debugging issues on the FPGA that arise from hardware issues, or signals that cannot be easily simulated (such as most of the hardware components).

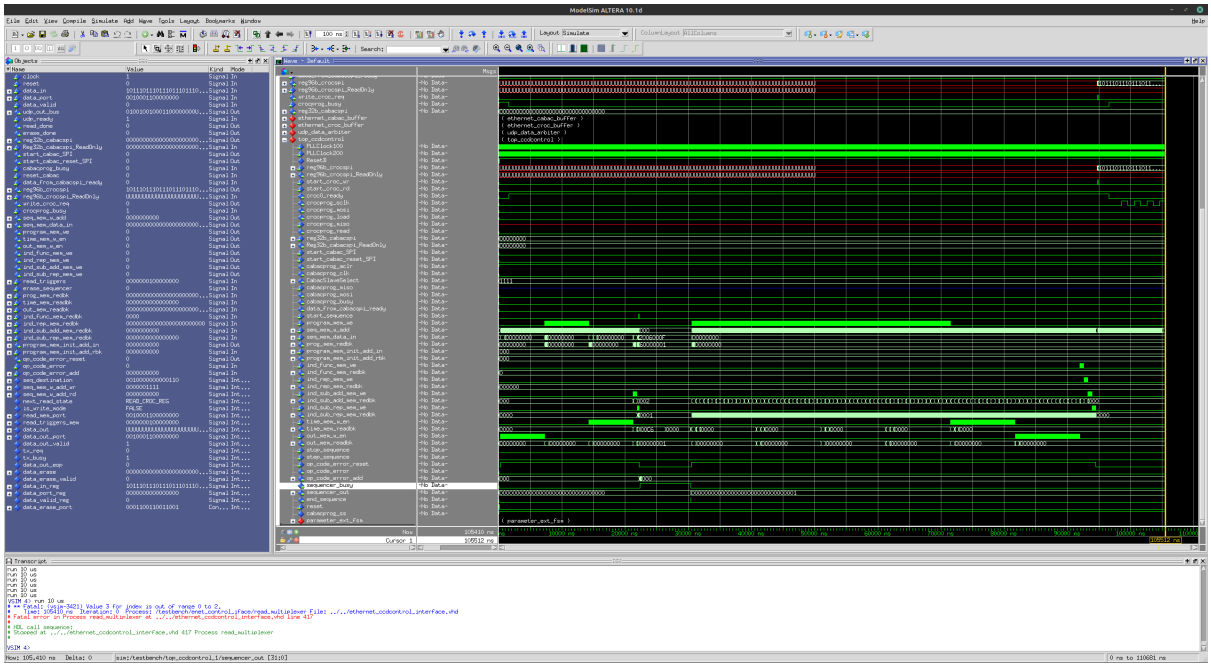


Figure 3.12: An example Modelsim simulation of the Ethernet-CCD control interface block, showing many of the CCD control internal signals in response to various operations. This type of simulation allows complete control of all external signals, and probes of all internal states. Some external hardware (such as the CROC lines) are not simulated, as this simulation is designed only to ensure the Ethernet-CCD control interface functionality.

3.8 ADC Driver

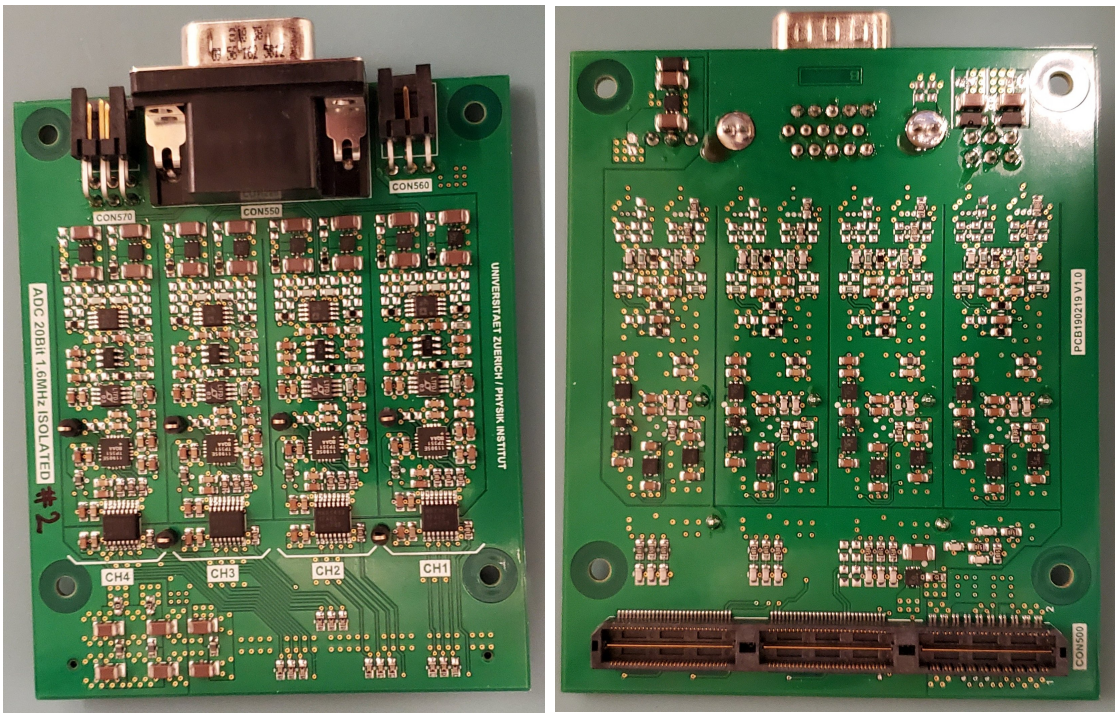


Figure 3.13: Picture of the 20-bit 1.6 MHz ADC mezzanine card from front (left) and back(right). The DB-15 connector used for the differential video signals is visible at the top of the image. Alongside this are two sets of pins to supply the ADC with external power, allowing the ADCs to operate on an analog power supply completely separate from the potentially noisy ODILE main board power (currently the system is ran entirely off the ODILE power). The mezzanine card contains 4 identical copies of the ADC chip, to allow readout of all 4 amplifiers on a CCD simultaneously.

The second major block on the ODILE firmware is the ADC driver logic. This block is responsible for driving the ADCs to perform sampling, and in the current iteration also handles the digital correlated double sampling (CDS) and skipper averaging for CCD readout. The layout of these components can be seen in fig. 3.14. The ADC block also includes some deglitching logic for the LVDS inputs, as these are used in the CCD image acquisition. These functions may move into a separate block in the future, to ease inter-compatibility with different ADC modules. Currently, the only available driver is for the 20-bit 1.6 Mhz MAX11905 SAR ADC (pictured in fig. 3.13), as that ADC was used in the predecessor to the ODILE board, and has legacy code available to drive it. This code implements a standard SPI interface to read the digitized values from the ADC, along with some additional logic to detect a sample is ready, and enforce the quiet time between

SPI and CNVST signals required by the ADC specification.

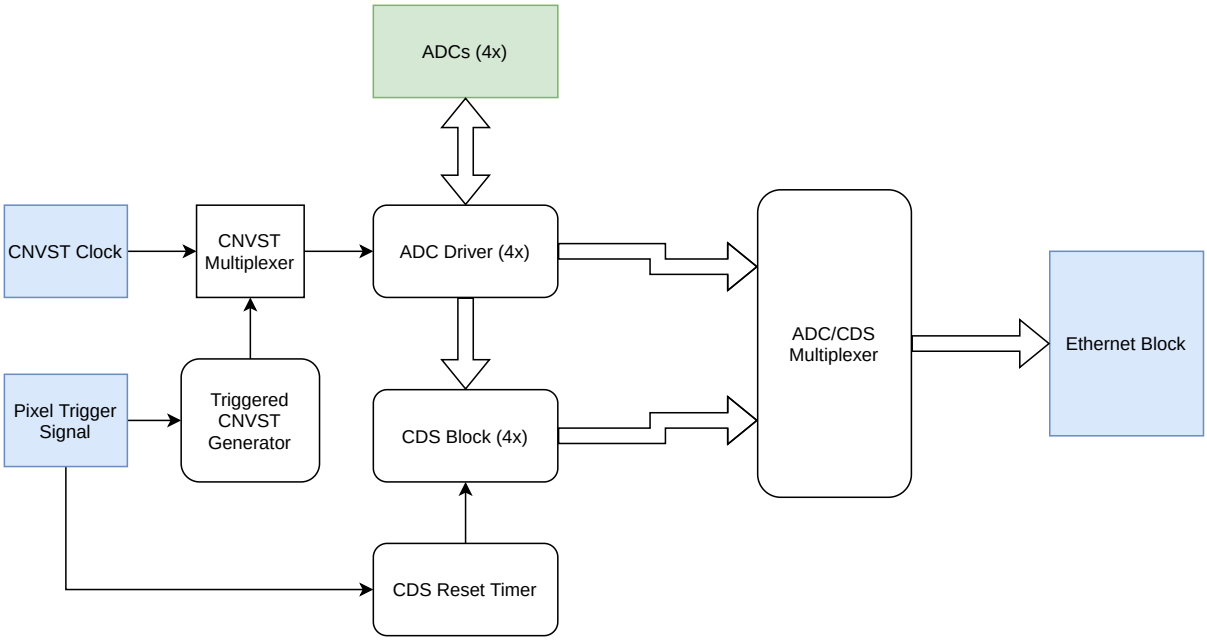


Figure 3.14: Schematic of ADC driver and associated logic. "CNVST" labels the conversion start signal, which can be driven by either a 1.6 Mhz clock, or a triggered signal driven by an external pixel integration signal. This signal is sourced from one of the three LVDS inputs to the ODILE, filtered to remove glitches. These LVDS signals can also be added to the top 3 bits of the ADC samples, allowing sampling of the LVDS signals at a particular sample. Data from the ADC driver is driven at the sampling rate. Data from the CDS driver is driven on the same clock, but using a data valid flag, since the averaged pixel value is only valid on some fraction of clock cycles. Data from either can be routed to the Ethernet block, based on the current configuration (only one can be chosen at a time). The reset timer activates after not receiving any pixel trigger signals for some time. Blue blocks indicate external logic, green are connected to physical hardware.

The CDS module handles averaging the incoming signals over both the individual pixels, and over multiple skipper samples. It is driven by two signals, a pixel integration signal, and a reset signal. When the pixel integration signal is driven high (to indicate the pixel pedestal), the block sums over the samples incoming on the ADC data lines. When the integration signal goes low, the logic averages over the number of samples it acquired. It then waits for the next integration window (the pixel signal), and repeats this sum/averaging process. It then makes the difference of these two values to make the pixel value. At this point, it can either output the pixel value (along with a pixel valid flag, to indicate the data is valid), or add this value to a running sum, to make the average over multiple NDCMs. In the latter case, the block will sum over the configured number of samples, then perform the average over all the samples to compute the average

of the skipper measurements.

Since only one signal is used for both pedestal and signal integration, the CDS logic assumes that the integration signal alternates between the two. Likewise, it assumes the skipper averaging occurs over the set number of pixels, starting with the first. The system includes a reset to override this, which will end the skipper averaging and reset the system to wait for the pedestal next. To increase the stability of the system, this reset is automatically triggered after a certain quiet period (this is around 6.2 milliseconds, in the current setup).

Due to the nature of how FPGAs handle arithmetic, floating point logic is difficult to handle, and would require a significant amount of additional work. Instead, all arithmetic is integer based, and any floating point remainder will be lost. This is not an issue when the range of integer is large, but can result in significant loss of information when the results occupy a small range of values. This seems to create issues with achieving single electron resolution with the skipper CCDs. Two methods of overcoming this problem were developed.

The first method is to configure the CDS block to output the sum of each pixel, rather than the average. However, running in this mode requires both pedestal and signal regions to contain the exact same number of samples, and so the block can be configured to sum only over a pre-defined number of samples. Due to the logical design, this value must be guaranteed to be equal to or less than the width of the pixel integration signal. This method eliminates the use of integer arithmetic entirely, and mimics the effect of analog CDS integration. However, it also results in the system gain being dependent on the integration window, which prevents direct comparison between images taken in different configuration.

The second method is to multiply the ADC samples by a constant before performing the average. This reduces the loss of information in integer division, since there are extra bits to store the remainder values in. This has the advantage of keeping the same calibration constant for all integration times (which greatly simplifies the process of comparing data sets). Both methods require increasing the amount of data sent to the server, since

we are using additional bits for the sum or multiplication, but the current scheme sends 32 bits for every sample (with the top 12 bits being zero), so this is unlikely to be an issue.

In an attempt to increase the flexibility of the CDS logic, it was split into two blocks: a driver block, that converts the integration signal into alternating pedestal and signal integration signals, and the actual CDS and skipper averaging logic. In practice these act as a single block.

References

[1] *Triple-Speed Ethernet Intel® FPGA IP User Guide*. Intel, 2021. URL https://datasheets.maximintegrated.com/en/ds/MAX11905.pdf.

[2] Internet Control Message Protocol. RFC 792, September 1981. URL https://rfc-editor.org/rfc/rfc792.txt.

[3] An Ethernet Address Resolution Protocol: Or Converting Network Protocol Addresses to 48.bit Ethernet Address for Transmission on Ethernet Hardware. RFC 826, November 1982. URL https://rfc-editor.org/rfc/rfc826.txt.

[4] Internet Protocol. RFC 791, September 1981. URL https://rfc-editor.org/rfc/rfc791.txt.

[5] Intel Corporation. *Quad-Serial Configuration (EPCQ) Devices Datasheet*, 2018. URL https://www.altera.com/en_US/pdfs/literature/hb/cfg/cfg_cf52012.pdf.

[6] Altera Corporation. *Arria V Device Handbook Volume 1: Device Interfaces and Integration*, 2018. URL https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/hb/arria-v/av_5v2.pdf.

[7] Laurent Le Guillou, Claire Juramy, Eric Aubourg, Stefano Russo, Eduardo Sepulveda, and Pierre Antilogus. *LSST REB Sequencer Language – User Manual*, 2017. URL http://supernovae.in2p3.fr/~llg/LSST/REB/sequencer/sequencer-language.pdf.

CHAPTER 4

CCD-ODILE INTEGRATION TESTS

4.1 Introduction

The performance of the ODILE system can only be properly evaluated when used for CCD readout. Tests of the system in isolation can be used for debugging or characterizing the performance of the ADCs alone, but actual performance requires attaching the ODILE to a CCD. Currently the CCD control electronics (the "CABAC" board) are under development at LPNHE in Paris. Until they are ready for testing, the ODILE cannot be used as a complete CCD controller. However, the ODILE can still be used to readout the CCD video signal, using one of the three prototype ADC mezzanine cards. Of these three cards, only the module using the 20-bit 1.6 Mhz MAX11905 SAR ADC has complete driver firmware available, as that ADC was used in the 20-bit ADC development board the ODILE is based upon. The setup and results of this integration test are detailed in this chapter.

4.2 Test Setup

The full test setup includes the ODILE board with an ADC mezzanine card, a controller for running CCD clocks and biases, the CCD itself inside a vacuum chamber, and several adapter and interface boards to convert signals between cables. The full setup is picture in fig. 4.1, and shown schematically in fig. 4.2. Details of the various components are given below. The inputs to the ODILE itself are shown in fig. 4.3.

4.2.1 CCD Setup

The CCD used for these integration tests was a 1k by 6k DAMIC-M development CCD, designed by Steve Holland and fabricated by Teledyne DALSA. The CCD was mounted inside a copper block (shown in fig. 4.4) designed to provide maximum cooling and shielding from external infrared radiation. This block was mounted on a Teflon table, inside a

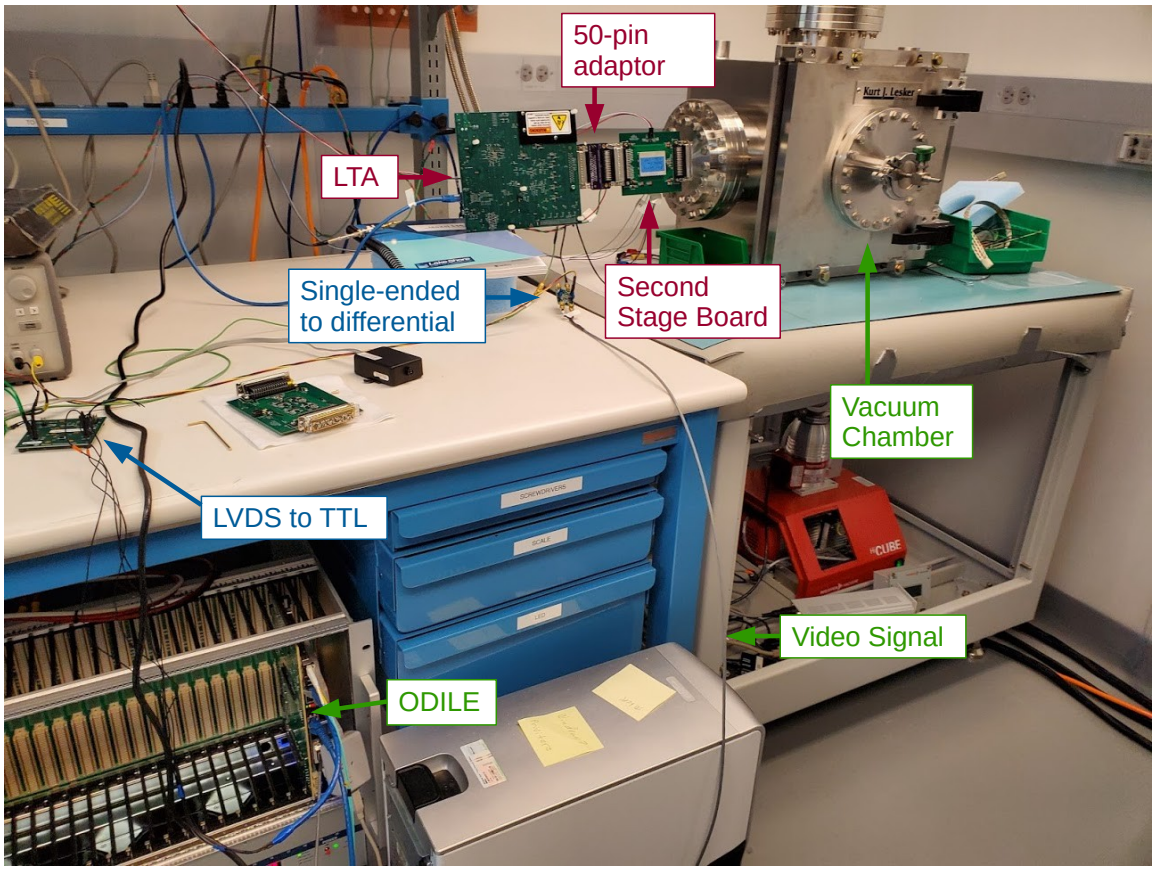


Figure 4.1: Picture of ODILE-CCD integration setup, using the LTA as the source for CCD clocks and biases. The second stage and adapter board are attached to the vacuum chamber, with the video signal sourced from the 50-pin second-stage-to-LTA adapter board.

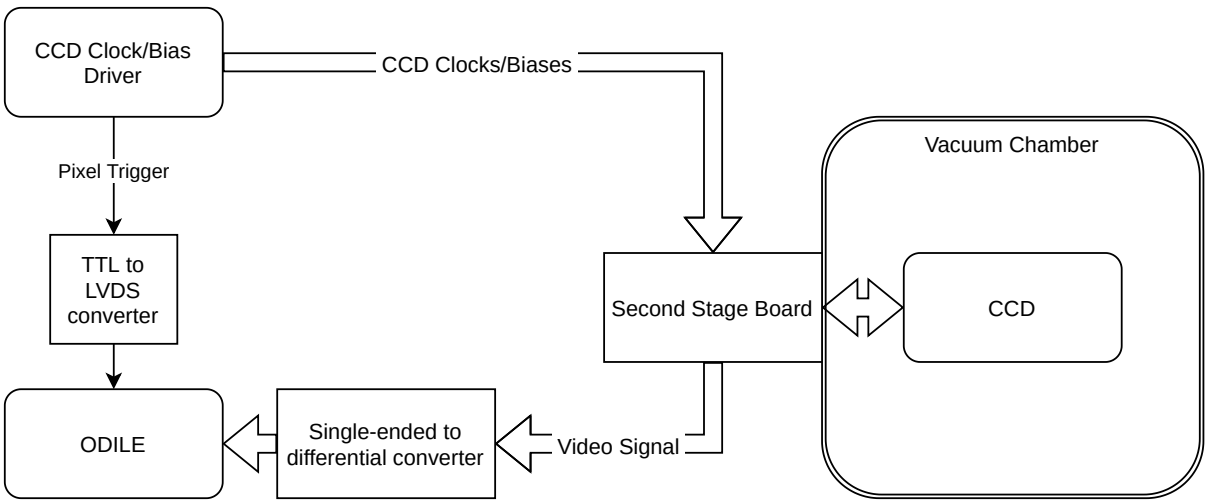


Figure 4.2: Schematic of integration test setup. The CCD clock and bias driver is either the ARC or LTA systems. The second stage board provides clock shaping and amplification. The single-ended to differential converter is an evaluation board used to convert to the differential signal expected by the ADC. The TTL to LVDS converter likewise converts to the LVDS signal usable as a pixel trigger signal.

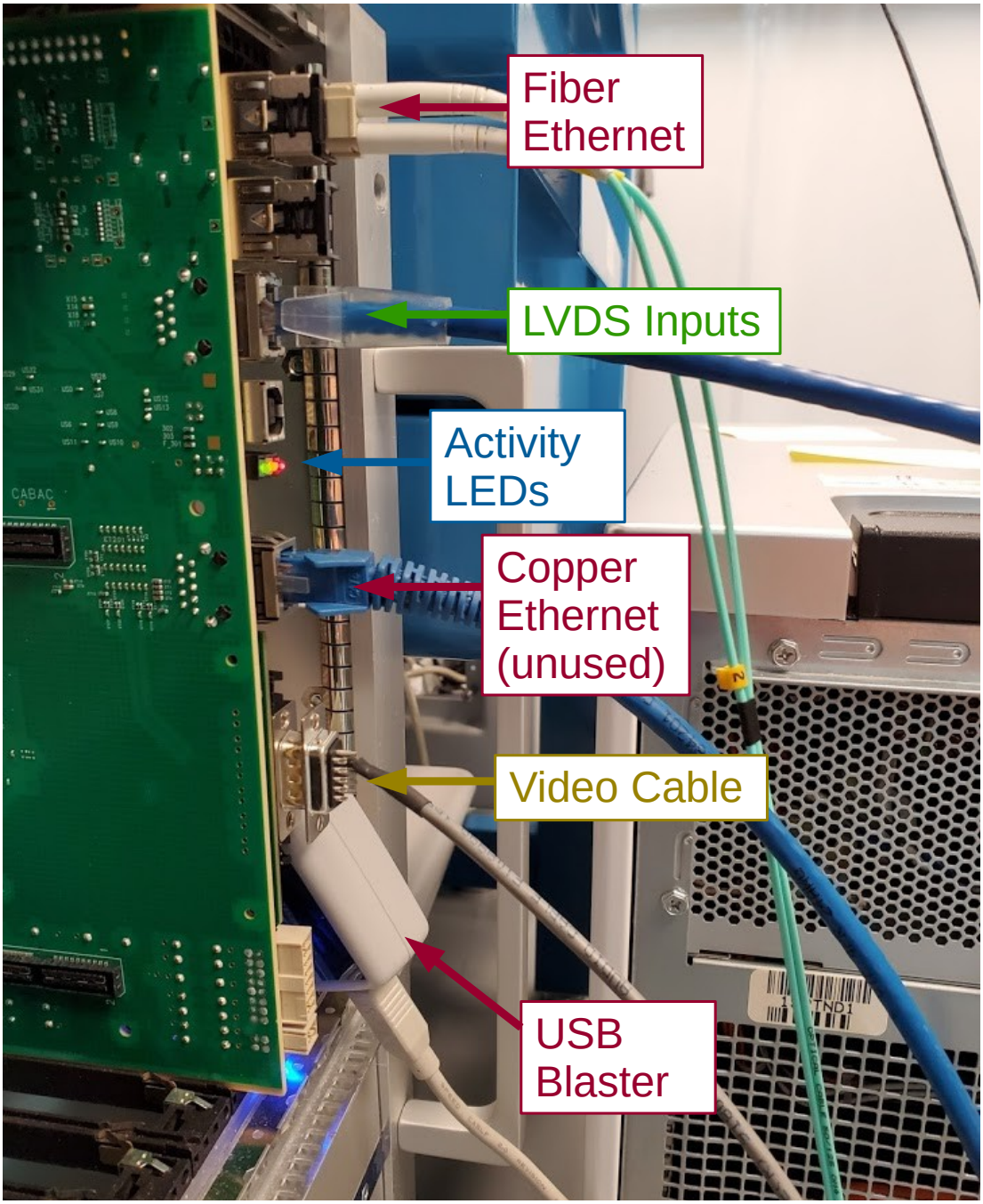


Figure 4.3: Picture of the ODILE in integration setup. The optical fiber Ethernet connection can be seen at the top. Below that is the LVDS signals used for the pixel integration trigger, and the activity LEDs. The copper RJ45 connector is installed (but not used for data acquisition), allowing us to test for any noise that may cause. The differential video signal is attached to the ADC mezzanine card using a DB-9 connector below that. The USB-blaster that is used for quick programming of the ODILE, as well as the SignalTap debugging, is visible at the bottom.

vacuum chamber build by the Kurt J. Lekser Company, as shown in fig. 4.5. The CCD was cooled to ~160 K by a Brooks Polycold PCC compressor (using PT30 blend coolant).

4.2.2 Electronics Setup

To power and clock the CCD while using the ODILE for video readout, there are two systems available. The first is the Astronomical Research Cameras (ARC) system, which has been used by the DAMIC experiment quite extensively [1, 2]. This system was not developed for skipper CCDs, but is flexible enough that the sequencer logic can be modified to run a skipper clocking sequence. Unfortunately, tests with the ODILE have demonstrated that the ARC system introduces some significant noise, which limits the performance of the ARC-ODILE integration.

The second system available for clocking is the Low Threshold Acquisition (LTA) system [3], developed at Fermi National Lab, with the explicit goal of low-noise readout using skipper CCDs. This system has shown better noise performance when used as the clock system, and was therefore used for the majority of the CCD-ODILE integration tests.

For either system, using the ODILE as a standalone video readout system requires some kind of synchronization signal between the clocking system and the ODILE, so that the ODILE can start/stop integration of the video signal.[1]

The ARC includes several unused clock lines and a pair of SMB outputs capable of multiplexing any of the clock lines, easily allowing the addition of a clock signal for the start/end of pixel integration window. The LTA does not technically include any extra clocks, but it does contain separate clocks for the two transfer gate clocks, of which only one is required during standard operation of the CCD. This extra TG clock was therefore re-purposed using a modified sequencer as a video integration signal. The ODILE has several LVDS inputs specifically designed for use as trigger or synchronization

1. Strictly speaking, the ODILE can be operated without synchronization, by simple digitization the entire video stream, and using offline analysis to perform the digital CDS procedure. However, this requires inferring pixel start times from the video signal itself. This procedure was done with the 20-bit ADC board that preceded the ODILE development, but the process was cumbersome and unreliable, and has not been repeated using the ODILE.

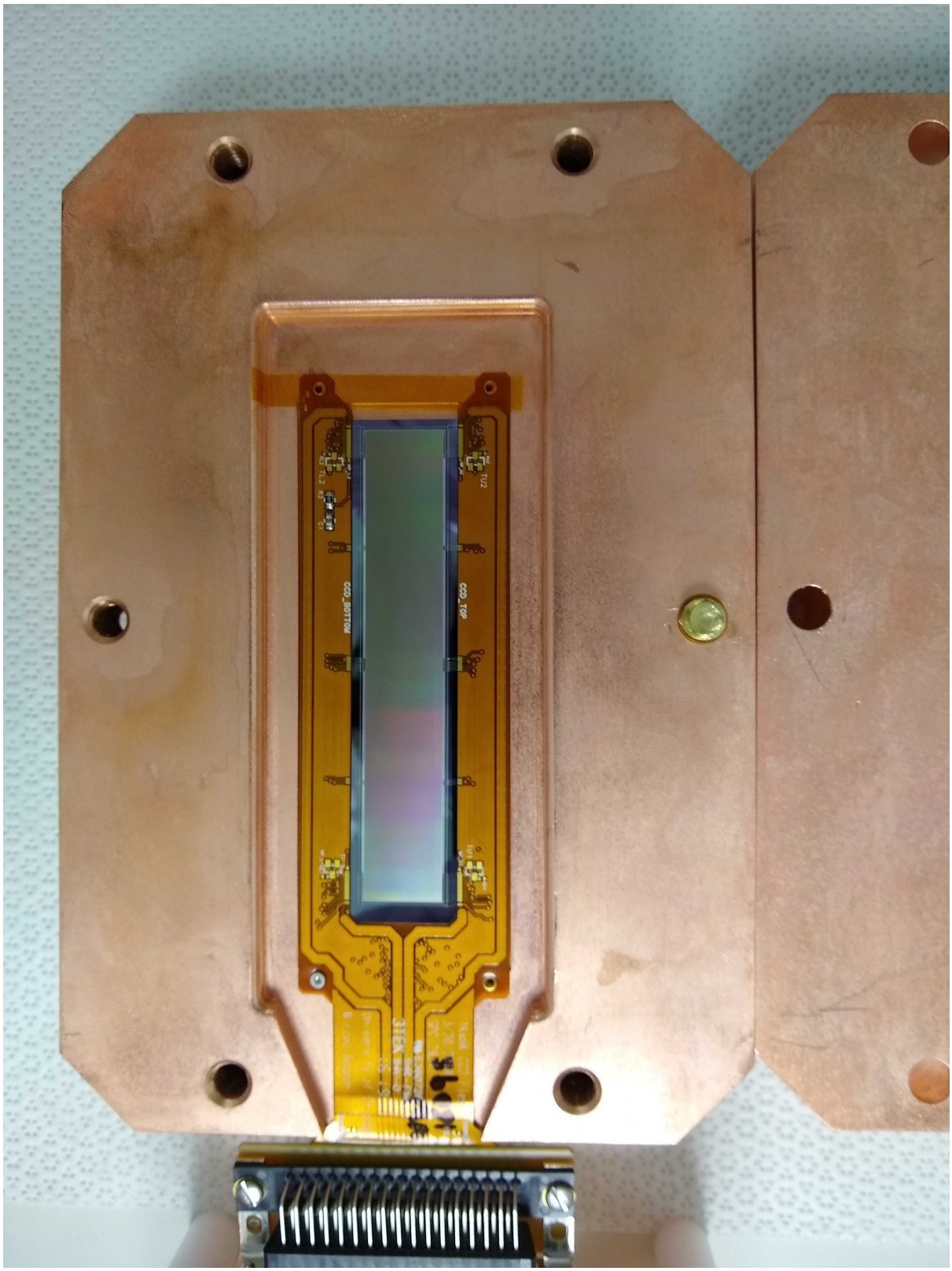


Figure 4.4: Picture of a 1k by 6k skipper CCD mounted inside the copper block used in test chamber 1. The CCD pixel array is the matte gray rectangle in the middle.

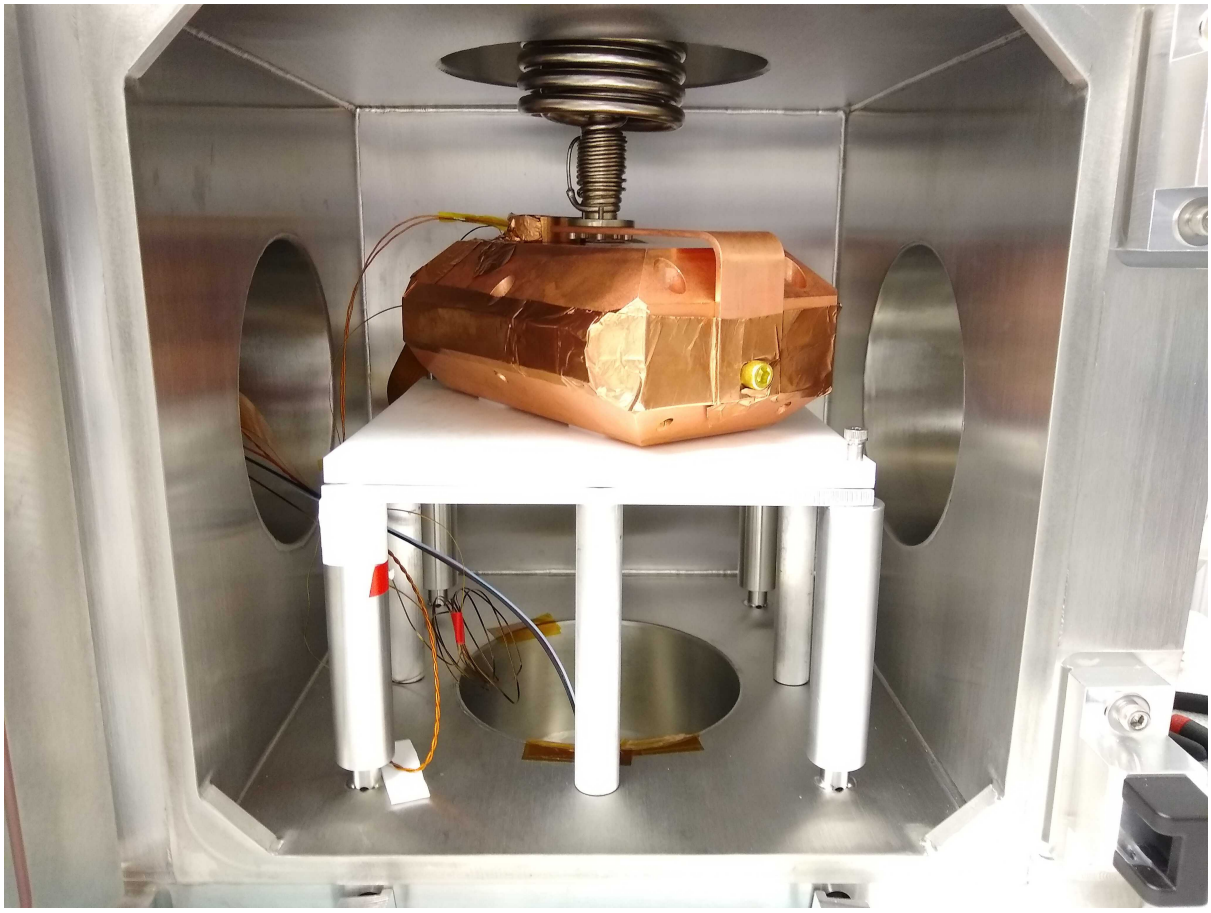


Figure 4.5: Picture of 1k by 6k DAMIC-M development skipper CCD, mounted in vacuum chamber. The large copper block provides additional shielding against external backgrounds, as well as any thermal radiation that could induced dark current in the CCD. The block is mounted on a Teflon table to provide thermal and electrical insulation from the chamber. The cold finger at the top is attached to the CCD through a copper bar.

Figure 4.6: Picture of the second stage board, designed at the University of Washington, along with the LTA adapter board.

signals. Converting from the single ended 3.3V TTL-like signal from the clock system to the differential signal expected by the ODILE requires the use of a TTL-to-LVDS converter. To perform this conversion, we used a dedicated LVDS line driver evaluation board (model "DS90LV027A-28AEVM") from Texas Instruments. This board includes 2 CMOS-to-LVDS lines, simplifying the process of integrating and testing the system.

Likewise, the current CCD readout electronics output a single ended video signal. The current ADC mezzanine cards expect a differential signal. Since driving a differential system with a single-ended input is not recommended, we used another differential amplifier evaluation board (model "ADA4945-1CP-EBZ") from Analog Devices, Inc. This board is capable of performing single-ended to differential conversion without modifications. Use of this board potentially increases the noise of the system (even if the amplifier is marketed as "low noise"), but the CCD signal is amplified by a factor of 31 in the second stage board, before the conversion process, so the impact of this converter is expected to be subdominant.

4.2.3 Second Stage Boards

All operation of the CCD uses a second-stage board, picture in fig. 4.6. This board attaches directly to the 50-pin connector on the vacuum chamber. It performs several functions:

Clock	Time constant
Output Gate	150 ns
Reset Gate	150 ns
Summing Well	150 ns
Vertical	10 μs
Horizontal	150 ns
Video	46 ns

Table 4.1: Typical clock shaping time constants for the second stage board used with DAMIC test CCDs.

1. Providing a common ground
2. Clock shaping
3. Amplification of CCD video signal

The clock shaping is highly important to reduce clock-induced noise. Constants for each clock are listed in table 4.1. Amplification is also important for skipper CCD operation: since the goal is sub-electron noise, the system gain must be $\gg 1$ ADU/e^-, or the signal will be buried under ADU noise. A common ground is also important to minimize ground loops, and has been determined experimentally to provide the optimal noise performance with our CCDs.

Integrating the ODILE, LTA, and CCD together to act as a single CCD readout system required an additional modification to the hardware. Connecting the LTA to the breakout board requires a small adapter board (since while both use 50-pin connectors, the expected ordering of the pins is different). This board contains no separate output for the video signal from the CCD: sending this signal to the ODILE instead of the LTA required removing a pair of pins and soldering in a new cable for the video signal.

4.3 20-bit 1.6 MHz ADC

Three ADC mezzanine cards have been developed for testing during ODILE development: a 20-bit 1.6 MHz board using the Maxim Integrated MAX11905 chip, a 20-bit 1.8 MHz board using the Analog Devices AD4020, and an 18-bit 15 MHz board using the Analog Devices LTC2387-18. All three chips are differential, successive-approximation ("SAR") ADCs. Only the first (the 20-bit 1.6 MHz board) has a fully developed firmware, and

has been used as the ADC in all CCD integration tests to date. Details of this chip can be found in [4].

The ADC conversion process is driven by a dedicated conversion-start signal ("CNVST"). This signal is required to operate at or below 1.6 MHz. Digital readout of the ADC code is performed through an SPI link (with the additional requirement that the SPI link stay quiet during the conversion process). The ADC is rated for 4.3 ADU of noise. The input range is -3 to 3 volts, for a conversion of 5.7μV/ADU, giving an input noise of 24μV RMS. The design used in testing includes no amplification at the input, an input bandwidth of 45 MHz (all filtering and amplification occurs before the ADC input), and uses fully differential inputs. Power spectral distributions for the input noise of the ADC alone can be seen in fig. 4.7. The noise with the CCD attached and powered (but without operating CCD clocks) can be seen in fig. 4.8. As can be seen comparing the two, the noise of the CCD and second-stage amplifiers is dominant across all frequency ranges, and the ADC adds no significant noise.

Operating the ODILE as a CCD video readout system requires some modifications to the firmware. The ADC system was originally designed for continuous readout: data is sent to the server at the ADC sampling rate (1.6 Mhz, with the current mezzanine card). The trigger signal from the clock system can be used by the ODILE CDS logic as an integration window timer, but this introduces significant opportunity for sampling jitter, since the sampling timescale of 640 nanoseconds is a fairly significant fraction of typically integration windows (which is on the scale of microseconds). This requires the use of fairly long wait times before pixel integration, and even so may introduce noise, as the CDS logic may sample over variable numbers of ADC samples.

We therefore modified the ADC driver to operate in a triggered mode: on the incoming trigger signal, the FPGA waits for a configurable amount of time, then triggers the ADC for a set number of samples. See fig. 4.9 for a timing diagram of these operating modes. This mode lowers the jitter between trigger and sampling window down to the timing of the primary FPGA clock (which is only a few tens of nanoseconds).

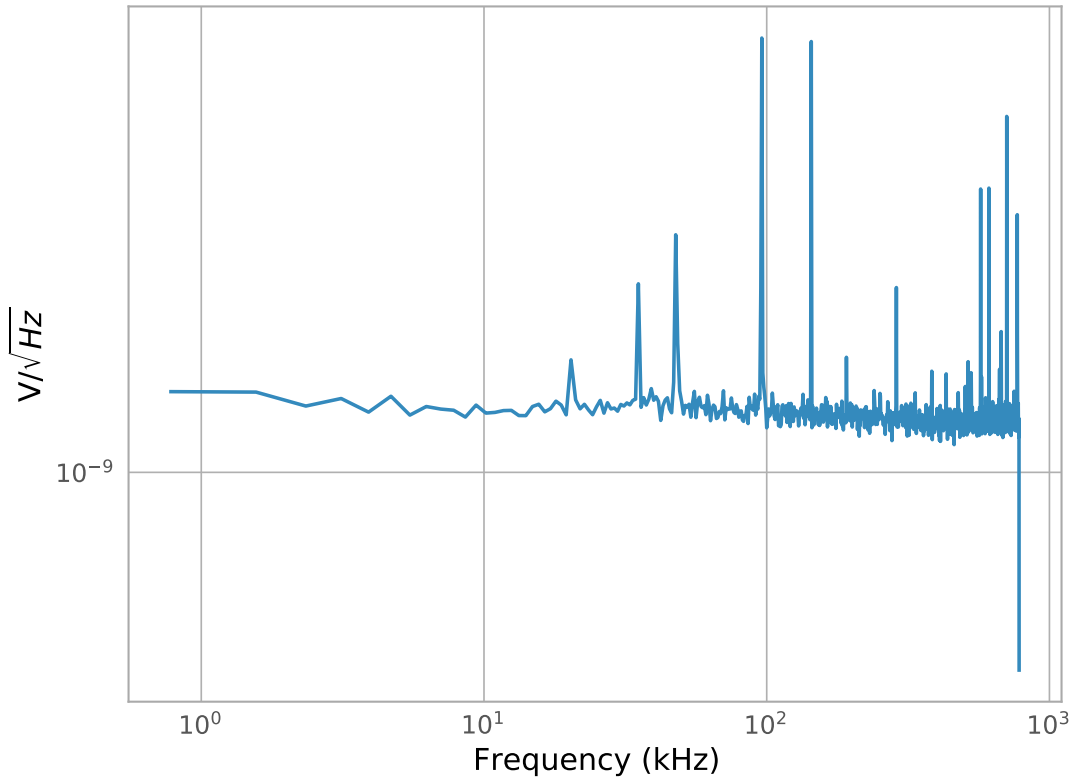


Figure 4.7: Power spectrum of the raw ADC noise. The relatively flat nature indicates the spectrum contains no significant non-white noise from outside sources, and the magnitude of 4.7 ADU (equivalent to 27 μV), which compares quite well with the expected inherent noise of the ADC, given as 4.3 ADU in the data sheet [4].

4.4 Integration Results

When operating in triggered mode, the ODILE is capable of performing the digital CDS, outputting the pixel values of the image. Custom written server code receives this data, and writes it to a .fits file for analysis or display. Images taken using a single measurement (NDCM) per pixel, and with 100 NDCMs/pixel can be seen in fig. 4.10.

A major advantage of the ODILE system over typical CCD data acquisition is access to the raw digital video signal. Typical CCD video readout systems only output the final averaged (signal-pedestal) value. In the case of the ARC system, for example, only the final value is available, since the analog integrator only makes one sample per pixel. Access to the underlying video stream allows more advanced noise analysis (such as examining the power spectrum for any bumps as in fig. 4.8, which was used in searching for and

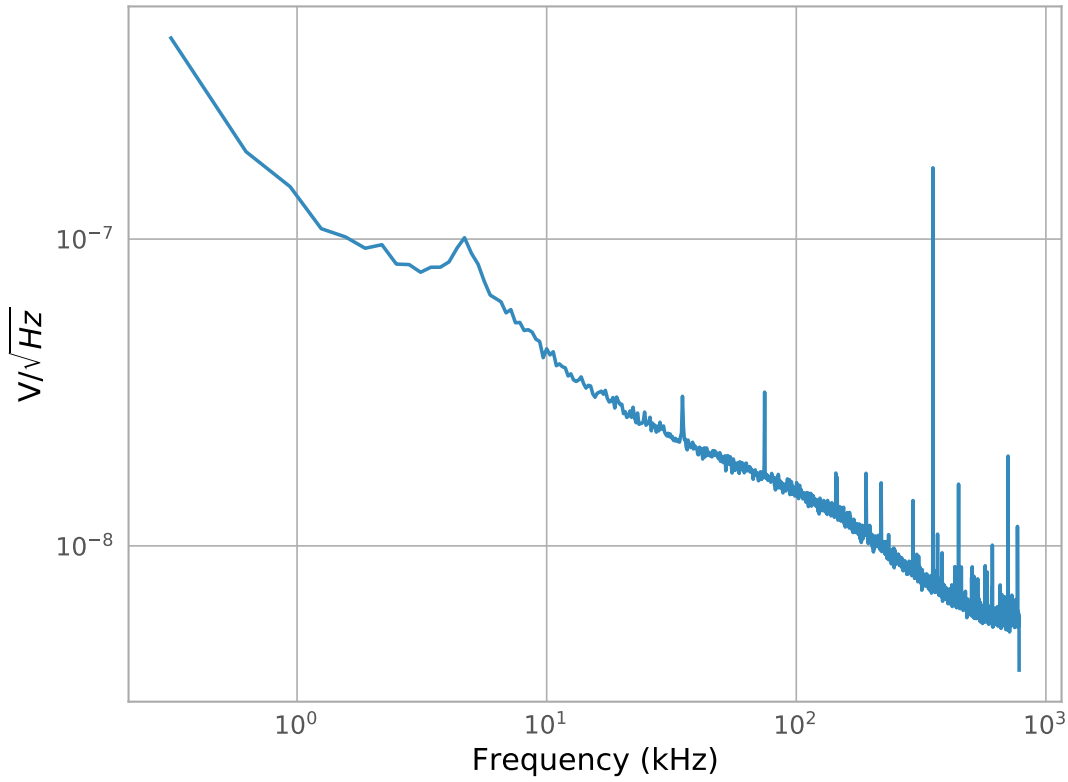


Figure 4.8: Power spectrum of the noise with CCD amplifier attached. The standard AC current noise bump at 60 Hz is clearly visible, but the noise otherwise follows a 1/f spectrum expected by the CCD amplifier. A few small spikes from environmental noise sources is visible, but there are no large sources of noise over most of the spectrum, indicating our setup has relatively low amounts of external electronic noise.

eliminating noise sources). We can also access the raw digitized video stream, which may prove useful in the future for developing more advanced digital filtering algorithms, and allows us to validate the results of the FPGA averaging by comparing it to the average performed offline, as in fig. 4.11.

The most relevant performance characteristic of a conventional CCD video readout system is the end noise of the pixel values, typically measured in units of electrons (e^-, though technically our CCDs measure positive holes), with a good system achieving around 2 e^- of RMS noise. Our system achieves 2.3 e^- with around 25 microseconds of integration time (as shown in fig. 4.12) with 1 NDCM/pixel, which is fairly reasonable for CCD systems of this type. After a minimum at around 30 microseconds, the noise actually starts to increase slightly, due to the 1/f noise from the CCD amplifier (this is

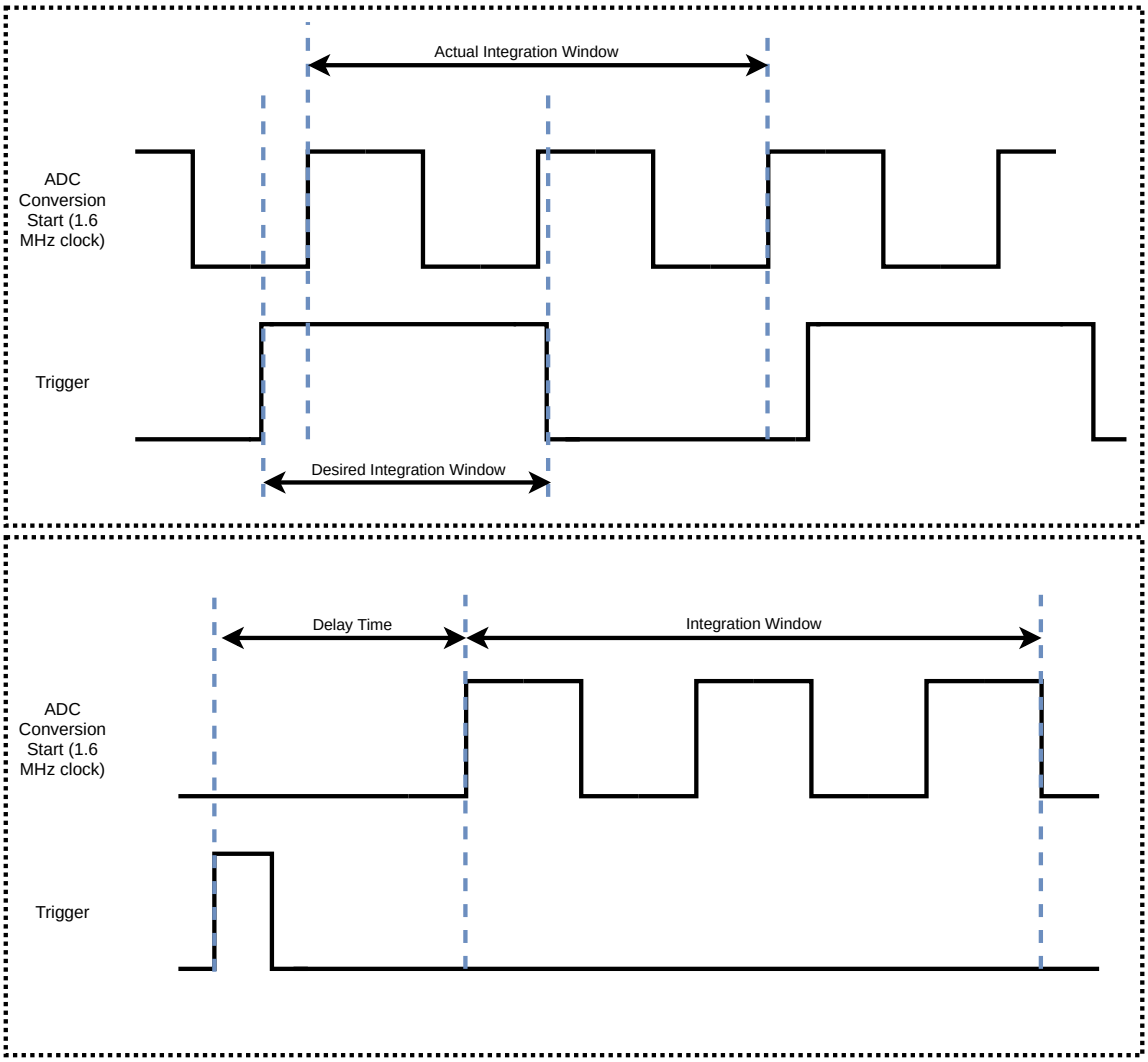


Figure 4.9: Timing diagrams showing the ADC running in free sampling (top) and triggered (bottom) modes, relative to the triggered signal. In the top, two triggers are shown, to illustrate the issue with jitter: the trigger signal has no relation to the conversion timing, and may occur anywhere relative to the 640 ns clock, resulting in an unknown (and inconsistent) timing between trigger signal and start of integration samples. This jitter may also result in +/- 1 sample being acquired for any particular pixel. In the triggered mode (bottom), the trigger signal has a fixed relationship to the conversion start, and the conversion start logic guarantees a specific number of samples per trigger. Both delay time and the number of samples per integration window are configurable.

also typical for CCD systems). For 4000 NDCMs, this gain with increased integration time is less dramatic, and reaches $0.05e^-$ at 10 microseconds.

For skipper CCDs, an even more important metric is the noise vs number of NDCMs. For an ideal system, the noise ought to decrease as $1/\sqrt{\mathrm{NCDM}}$. This will not be true if the skipper process is performed incorrectly (such as loss of charge in the skipper process), or if there is time-correlated noise that makes subsequent measurements no

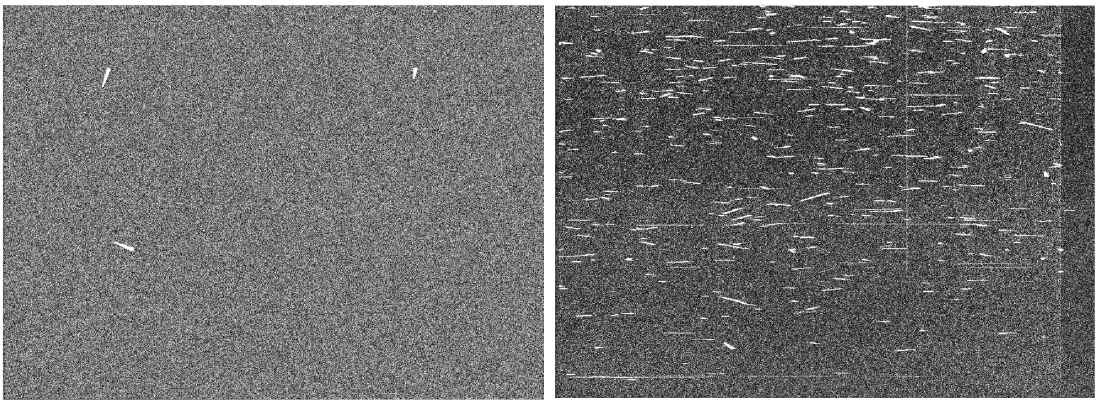


Figure 4.10: Left: Example image taken using the 1.6 MHz ADC, in a single-NDCM per pixel configuration. Several tracks of cosmic ray muons are clearly visible. **Right:** Example 100 NDCM image. The lower noise from skipper reduction can be seen even by eye in the overscan region on the right hand side of the skipper image, compared to the non-skipper image in fig. 4.10. The horizontal tracks are charge from particles striking the serial register during the readout process, which is quite long (around 1 hour minutes, just for this 800x1100 pixel image). The large number of events is due to the long readout time required by the skipping process.

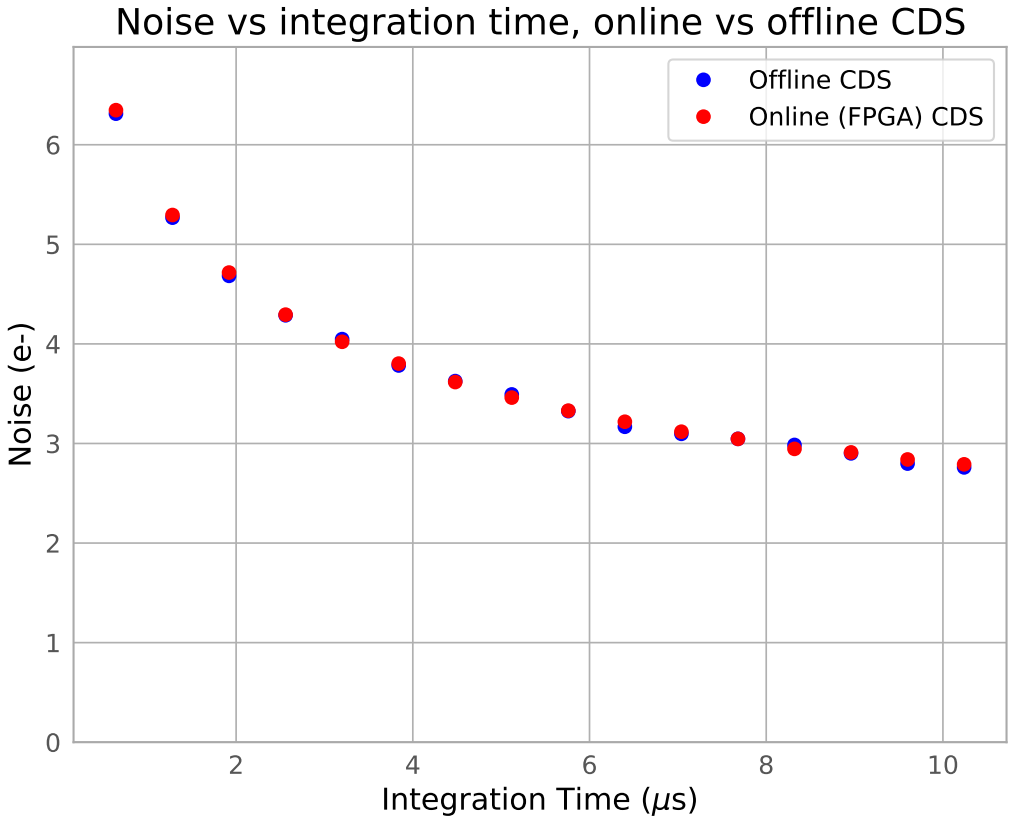


Figure 4.11: Comparison of noise when the CDS is calculated offline from the raw ADC values (blue) to the noise when the CDS is computed on the FPGA (red).

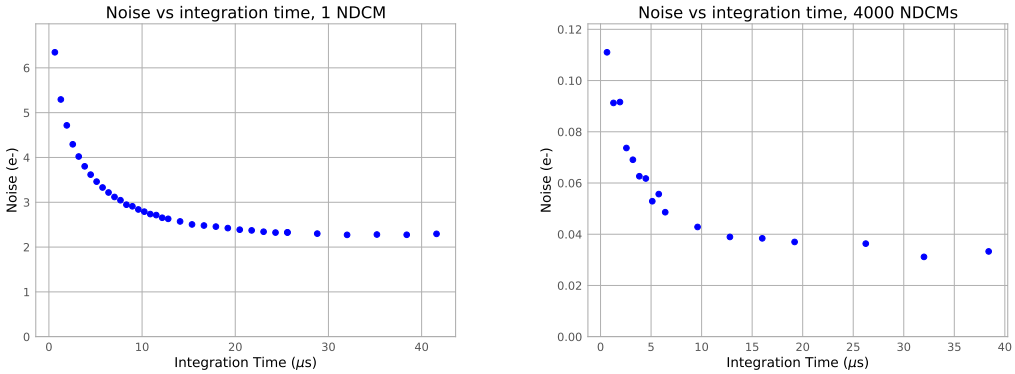


Figure 4.12: Plots of noise vs integration window for one (left) and 4000 (right) skip images.

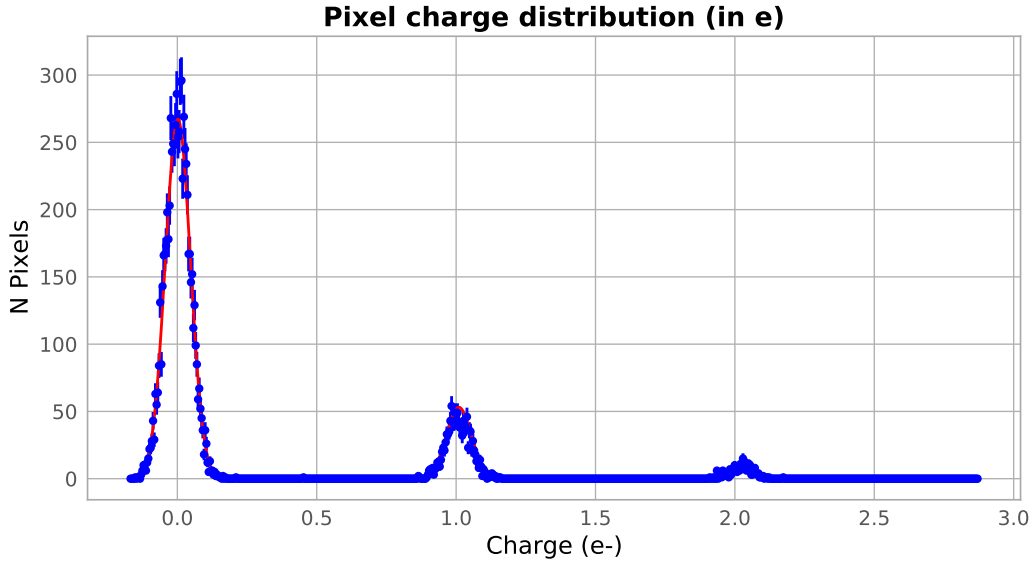


Figure 4.13: Example of pixel distribution for a 4000 skip image. The red line shows the fits to the 0, 1, and 2 electron peaks, with an RMS width for each of $0.04e^{-}$.

longer independent. As can be see in fig. 4.14, the noise of our system follows exactly the curve expected by theory, indicating our skipper system is performing exactly as expected.

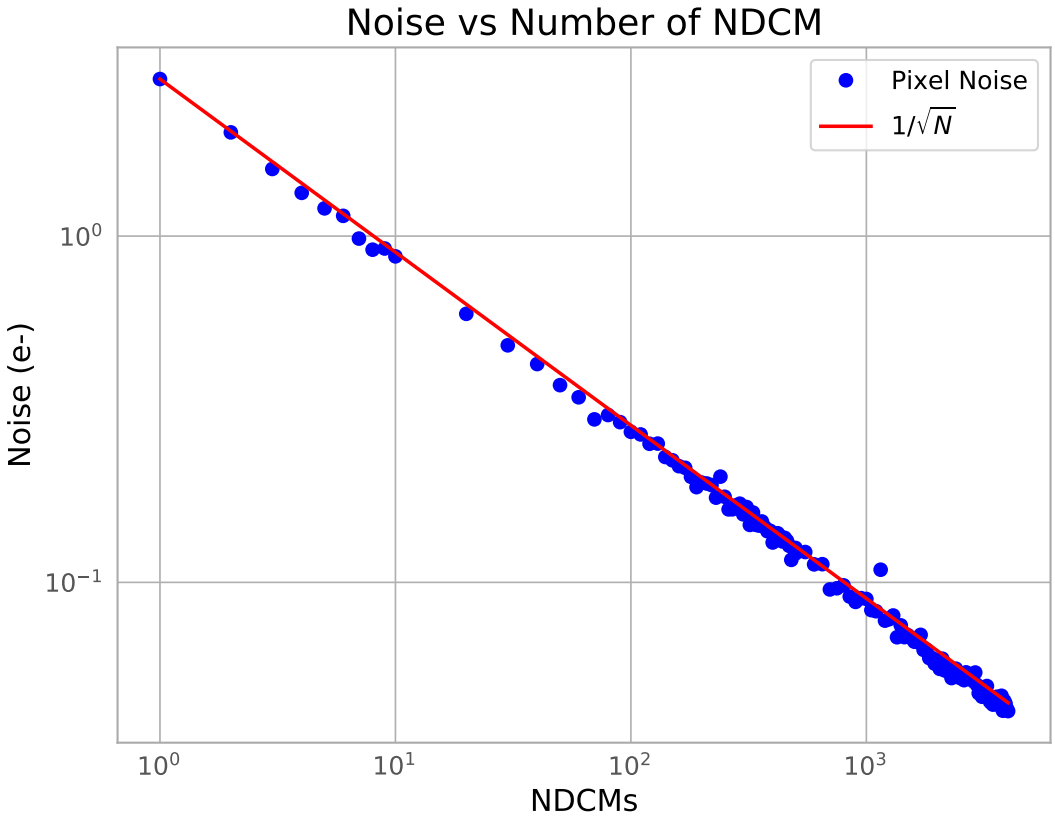


Figure 4.14: Noise versus number of NDCM, from 1 to 4000 NDCMs. The line in red shows the expected $\frac{1}{\sqrt{N_{\text{skips}}}}$ gain expected by theory, which is followed quite well by the actual data set.

References

[1] K. Ramanathan, A. Kavner, A. E. Chavarria, P. Privitera, D. Amidei, T.-L. Chou, A. Matalon, R. Thomas, J. Estrada, J. Tiffenberg, and J. Molina. Measurement of low energy ionization signals from compton scattering in a charge-coupled device dark matter detector. *Phys. Rev. D*, 96:042002, Aug 2017. doi: 10.1103/PhysRevD.96.042002. URL `https://link.aps.org/doi/10.1103/PhysRevD.96.042002`.

[2] A. E. Chavarria, J. I. Collar, J. R. Peña, P. Privitera, A. E. Robinson, B. Scholz, C. Sengul, J. Zhou, J. Estrada, F. Izraelevitch, et al. Measurement of the ionization produced by sub-kev silicon nuclear recoils in a ccd dark matter detector. *Phys. Rev. D*, 94:082007, Oct 2016. doi: 10.1103/PhysRevD.94.082007. URL `https://link.aps.org/doi/10.1103/PhysRevD.94.082007`.

[3] Guillermo Fernandez Moroni, Fernando Chierchie, Miguel Haro, L. Stefanazzi, A. Soto, Eduardo Paolini, G. Cancelo, K. Treptow, N. Wilcer, T. Zmuda, et al. Low threshold acquisition controller for skipper charge coupled devices. pages 86–91, 03 2019. doi: 10.1109/CAE.2019.8709274.

[4] *MAX11905 20-Bit, 1.6Msps, Low-Power, Fully Differential SAR ADC*. Maxim Integrated, 2018. URL `https://datasheets.maximintegrated.com/en/ds/MAX11905.pdf`. Rev. 4.

CHAPTER 5
DAMIC AT SNOLAB SIMULATIONS AND BACKGROUND MODEL

5.1 Introduction

Dark matter direct detection experiments are almost universally limited by the radioactive backgrounds present in the detector and surrounding environment. This background means that even with an unlimited exposure time and detector mass, the detector would be unable to distinguish between the background and any potential signal. The ideal solution is to lower the detector background as close to zero as possible. A certain level of background, however, is unavoidable. At that point, the background must be incorporated into your detection model. The greater your understanding of the background, the better the model you can create, and the more power your results will have to detect or exclude potential dark matter signals.

The standard method for background modeling is through Monte Carlo simulations. Previous generations of the DAMIC experiment have used a simplified simulation with three components: surface backgrounds (for e.g. from ^{210}Pb deposited on the surface of the CCD), bulk backgrounds (for e.g. Compton-scattered gamma rays), and potential signal (for e.g. the WIMP signal discussed in section 1.4.1). This technique is highly limited, because it does not utilize our full knowledge of the radioactive contamination of components of the detector, which have been measured through gamma ray counting and mass spectroscopy. We can improve on this simple model through a complete simulation of radioactive decays, particle tracking, energy deposition, and detector response. The DAMIC experiment uses the `GEANT4` [1] simulation toolkit to simulate the first 3 steps. `GEANT4` is a Monte Carlo library written in C++, capable of simulating the entire chain from radioactive decay through to the energy deposition in the detector. The final step (detector response) was performed using custom-built code.

A major advantage of the `GEANT4` toolkit is its modular and open sourced nature, which allows the user to chose the required physics models for their purposes, or to

modify the internal generator classes to create the correct simulations. For example, a key simulation of the DAMIC CCDs requires simulating the decay of ^{210}Pb embedded on the surface components of the CCD, following a complementary error function distribution (as expected based on literature studies of diffusion depths [2]). The ability to modify the internal source code of `GEANT4`'s source generator made this easy.

While `GEANT4` can handle the complete simulation from decay to ionization in the detector, it does not inherently include ability to model the detector response to the ionization, such as charge diffusion, pixelization, or readout noise. The tools to model these effects were created largely from scratch to model the full DAMIC at SNOLAB backgrounds.

Details of the `GEANT4` simulation used for the DAMIC at SNOLAB background model can be found in [3]. Here we will focus on the description of the processing of this data into a usable spectra for the DAMIC at SNOLAB background, as well as steps to verify the simulations model, and to create the final background model from simulations.

5.2 Detector Response

The output of `GEANT4` simulations gives a list of energy depositions, with the energy deposited (in keV) and the coordinates of the deposition. Converting this into a spectrum usable by the standard DAMIC analysis tools requires several additional steps. These are (in order):

1. Fano noise
2. Partial Charge Collection
3. Diffusion
4. Pixelization
5. Cluster identification
6. Pixel Binning
7. Readout noise
8. Pixel saturation

Typically, steps 1 through 5 are performed separately from steps 6 through 8. This

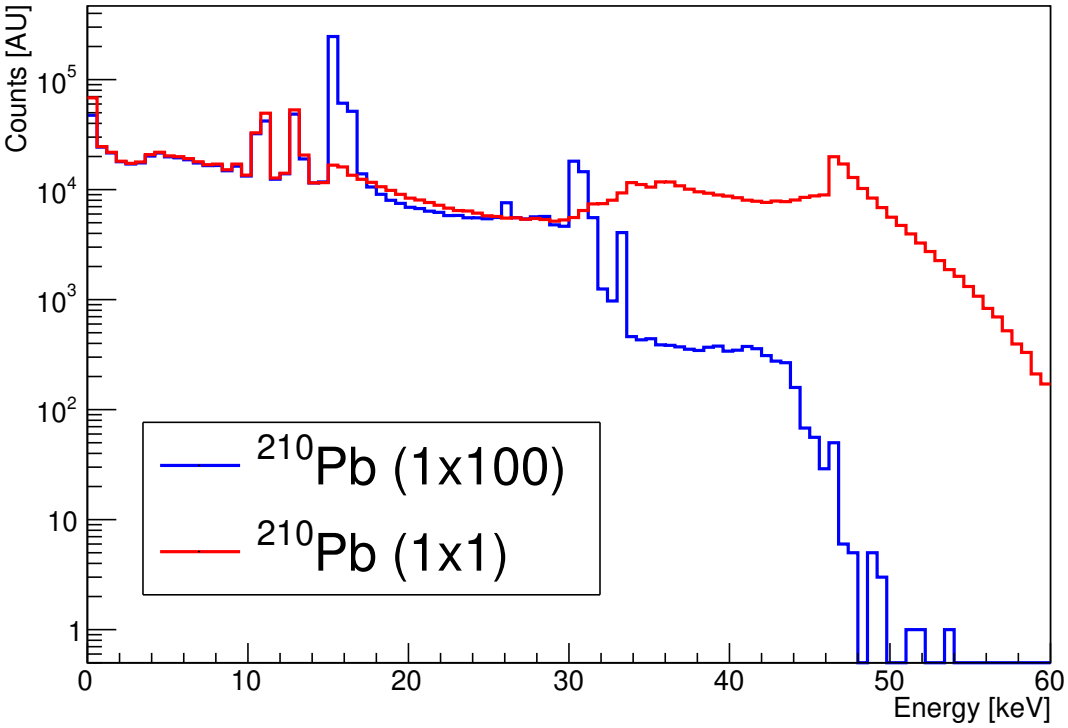


Figure 5.1: Comparison of front surface ^{210}Pb spectra with (blue) and without (red) noise, 1x100 binning, and saturation effects. The large peaks at 15 and 30 keV in the 1x100 spectrum is the one and two pixel saturation peaks, which shifts nearly all of the spectrum >30 keV into lower energy bins. These additional effects have very little impact in the spectrum below the saturation peak. The saturation effect is extremely pronounced, as this simulation is of decays on the front side of the CCD, resulting in nearly all of the charge being contained in 1 or 2 pixels.

allows us to separate the latter effects, which are caused by the readout process, from the first 4 effects, which are inherent to CCD technology (clustering is technically a processing effect, but is done in order to compress the data to a manageable level). A comparison of simulations with and without these readout effects can be seen in fig. 5.1.

The full Monte Carlo simulation of the SNOLAB detector requires simulating 23 different isotopes, in 64 different volumes, plus surface contaminants for all volumes. A wireframe and cross section of this geometry can be seen in fig. 5.2 and fig. 5.3, respectively. The number of simulations required for each isotope to generate a statistically useful result was tuned to optimize the number of events simulated, and not all isotopes are simulated for all volumes, but the result is still 1869 different simulations, taking up almost 3 TB of raw disk space Processing these simulations in a reasonable time-frame

required extensive optimization of the processing code, such as minimizing array allocation/deletion. The final code (aptly named "ReadSimFast") is capable of processing all 3TB of simulations in around one day, using a 32-core server. The processed simulations are only a few hundred GB in size, thanks to the reduction in data size during the pixelization process.

Fano Factor. The first step in processing the simulations is to convert the energy produced into a discrete number of electron-hole pairs (since the detector actually reads out the number of electron holes). The number of electrons produced is nominally a simple function of the deposited energy, so that $N_{e^-} = E \cdot 3.77\frac{eV}{e^-}$. As described in section 2.5, the actual number of electrons produced in each energy deposit has some fluctuation from the Fano noise of the process. In our simulation, we model this as a Gaussian distribution, with the mean N_{e^-} given above and a variance $\sigma^2 = FN_{e^-}$ (where F=0.129). The number of electrons produced is then rounded to the nearest whole integer, and then we apply diffusion.

Partial Charge Collection. Partial Charge Collection (PCC) is a phenomenon that is not typically considered in CCD data analysis. The backside of a CCD contains a region where the phosphorous content decreases to the bulk level. In this region, charge carriers have a significant probability of recombination, resulting in some charge loss. This results in significant changes to the spectrum and rate of events that occur in the back few microns of the CCD (see fig. 5.4), and must be included in our simulations in order to reproduce the correct spectra.

PCC in our CCD is modeled immediately after the Fano factor, and before propagating the electrons through diffusion. The partial charge collection efficiency at any distance z from the rear of the the CCD is parameterized through a polynomial

$$P(z) = \sum_{n=i}^{N} n_i z \tag{5.1}$$

where N is the degree of the polynomial (seven, in the current version), and n_i are the coefficients of our polynomial. P is capped in the range [0,1], since it represents a prob-

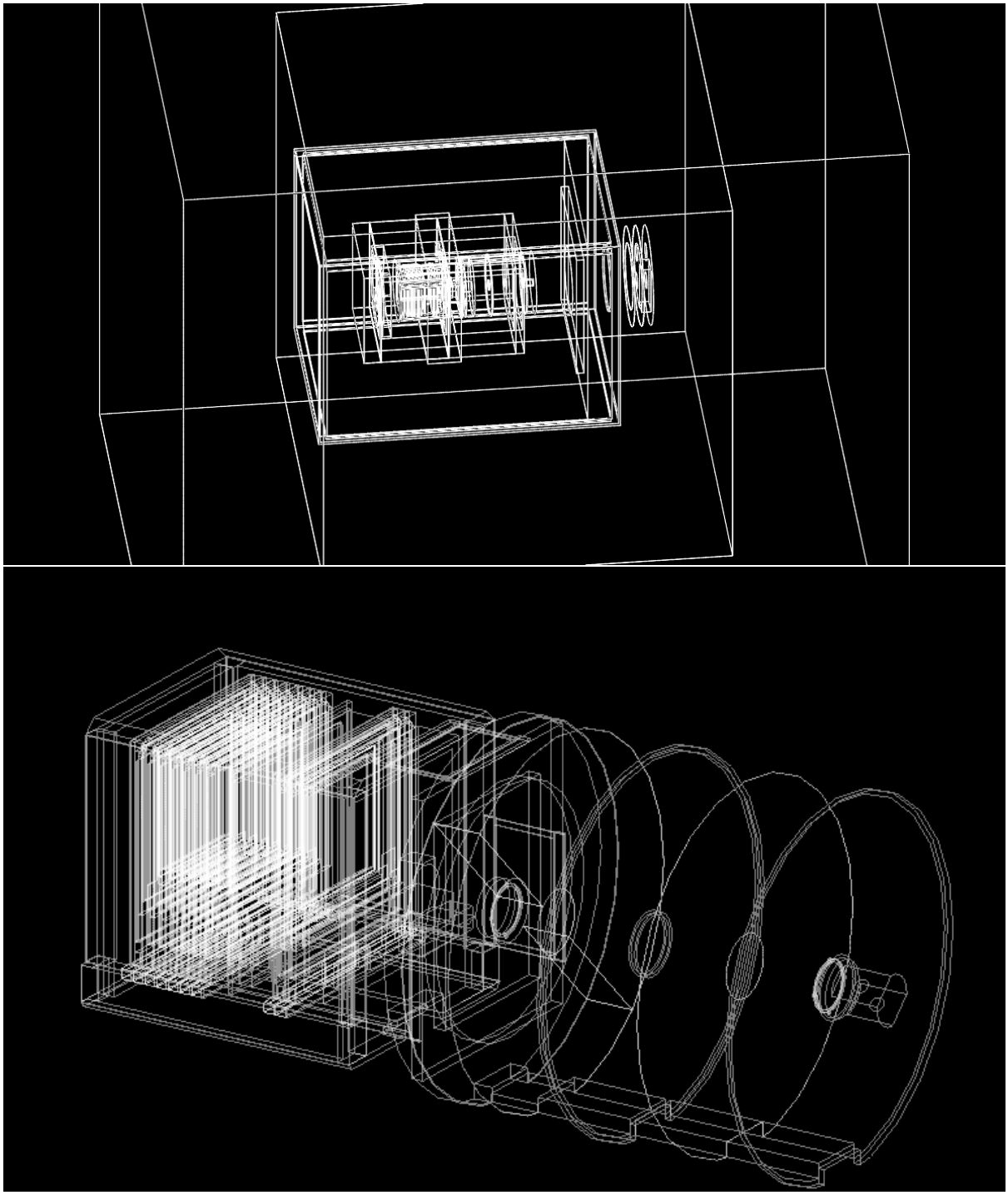


Figure 5.2: Top: Simulated DAMIC at SNOLAB geometry, showing the entire detector in wire-frame from the outer lead down to the inner CCDs. The larger outer box is regular lead, followed by ancient lead. Inside this is the copper vacuum vessel (which is circular, and can't be seen in this wire-frame). Inside that the box holding the detectors. **Bottom:** a zoomed wireframe of the DAMIC at SNOLAB inner box, which shows the CCDs inside their copper trays, held inside a copper box. The discs to the right (above, in the real detector) are lead (ancient for those near the detectors), to shield the CCDs from the vacuum interface board (VIB) electronics (not simulated). The lead discs contain a hole for the cold finger to cool the CCDs.

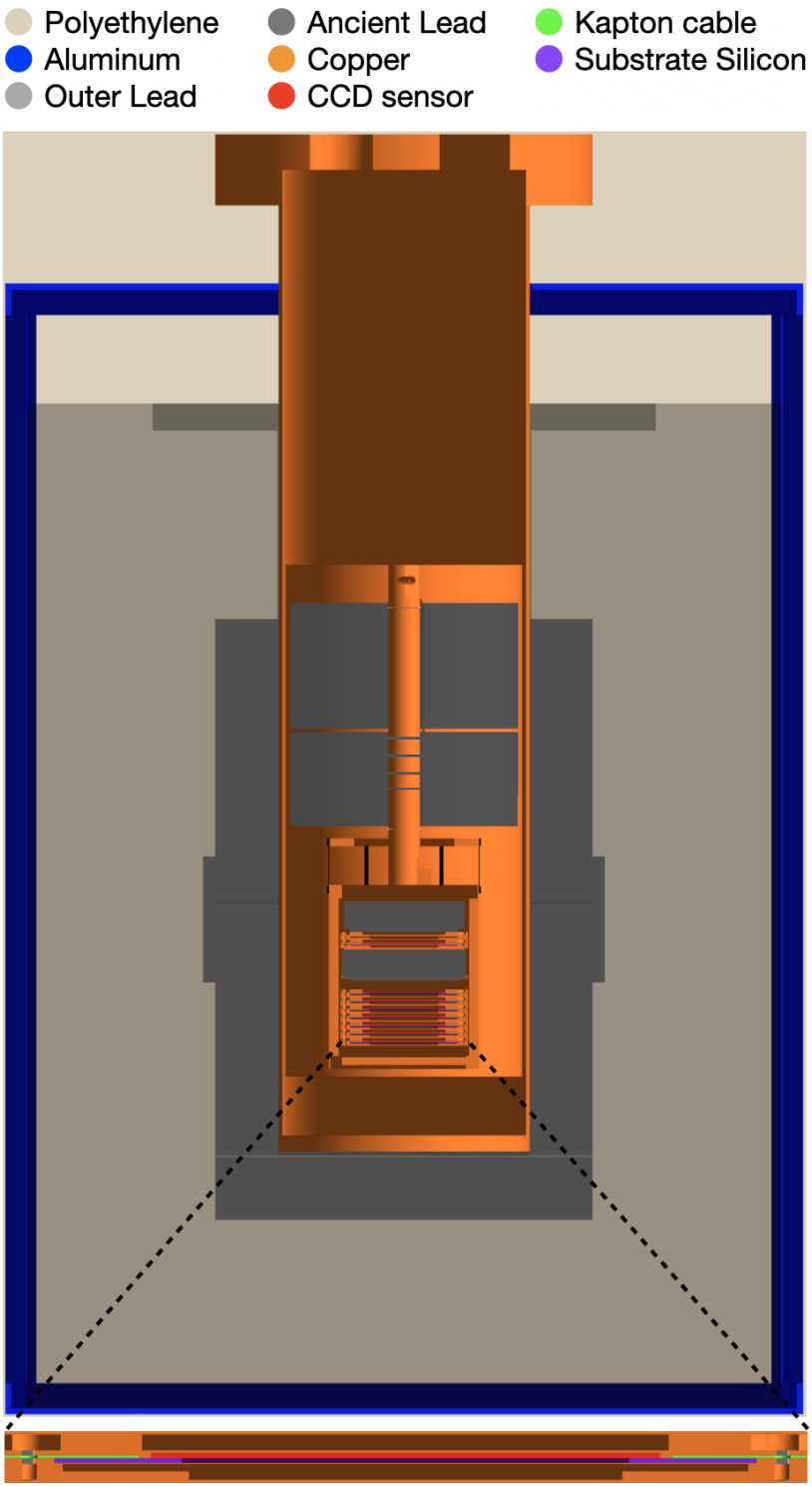


Figure 5.3: DAMIC at SNOLAB detector geometry cross section. The labeled colors show various detector components.

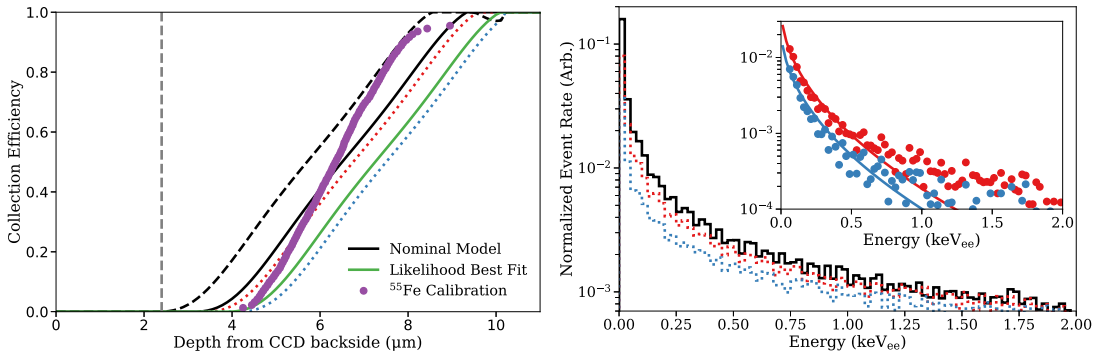


Figure 5.4: Left: A sample of various PCC models simulated for the backgrounds analysis, showing the range of possible efficiencies given the uncertainties in the analysis. **Right:** the effects of the different PCC models on ^{210}Pb event rate, for events simulated on the backside of the CCD. The nominal model is shown in dark black. The most significant effect on the spectral shape was observed when events start further into the CCD, suggesting that the collection "turnon" location was the most significant cause of uncertainty in the spectra arising from the PCC model.

ability. Likewise, the range of PCC is capped, and only calculated if z falls in a region where $0 < P < 1$. A large order polynomial was used in the simulation processing to allow the maximum flexibility in changing PCC model parameters in the simulation: the efficiencies were determined by simulating the propagation of charge using a numerical calculation, and a parameterization of the results used for the `GEANT4` simulations. Determining the correct coefficients was a significant undertaking, which is detailed in the appendix to [4].

Once this collection efficiency is determined, a random binomial is used to compute how many of the electrons initially created will survive, and the remaining electrons are passed to the diffusion stage.

Diffusion. The second step in simulation processing is to diffuse the individual electron-hole pairs. As mentioned in section section 2.2, thermal motion of the holes allows charge to drift in the x-y axis during the collection process. This process is simulated by allowing each hole to drift a Gaussian distance in both x and y. The variance of this distance is a function of the depth of the interaction, following the form

$$\sigma^2(z) = -A \cdot ln(1 - bz) \tag{5.2}$$

where A and b are constants whose parameters are derived from fits to the data,and z

is the depth of the energy deposit. An energy-dependent deviation from this diffusion model was noticed in our data-set. Corrections of the form

$$\sigma = \sqrt{-A \cdot ln(1 - bz)}\,(1 + \alpha E)^2 \tag{5.3}$$

with E being the total energy of the cluster, and α giving the energy dependence of the diffusion. However, applying this correction at such an early stage proved difficult: the energy dependence relies on the energy of the total cluster, which is not known until after the clustering step below. In addition, applying this correction early on makes tweaking the energy dependence parameter α more difficult. For this reason, it was decided to use the non-energy-dependent form given in eq. (5.2).

In our processing code, each electron is diffused independently, with a random Gaussian using the variance given above. Each electron is assigned a new position

$$x_{\text{new}} = x_{\text{old}} + \text{Gaus}(0, \sigma^2) \tag{5.4}$$

The shift occurs in both x and y directions. Once this new position is determined, the electron position is then converted into pixel coordinates using $n_x = \text{floor}(\frac{x_{\text{new}}}{15\mu m})$ in both x and y.

Charge Clustering. After all of the energy from an event has been converted into electrons and diffused, the event can be clustered. The clustering process is relatively simple, since the processing occurs on an event-by-event basis. Each pixel in the event that contains charge is scanned through, and if any adjacent pixels in the same CCD also contain charge, they are added together into a cluster. In this way all adjacent pixels containing charge form a single cluster. This step of the process is typically performed free of noise, which is instead added later on in the processing.[1] In this way, we do not eliminate any information or events from the simulation until it becomes absolutely

1. The algorithm does *allow* for the addition of noise at this stage. In this case, the value of each charge-containing pixel is modified by a random Gaussian value with a variance σ^2 equal to the noise, and then is only added to a cluster if the value after the noise is $>= 4\sigma$, to simulate the real noise cut on data.

necessary.

When we perform this processing we also compute the energy-averaged z-depth of the interaction,

$$z_{\text{avg}} = \frac{1}{E} \sum_{i=0}^{N_{\text{elec}}} z_i \tag{5.5}$$

where z_i is the depth of the ith electron in the pixel. This information is carried through for every pixel in a cluster, and allows us to reconstruct the original depth of events, especially for point-like events. This is used in for e.g. the partial charge collection model, to simulate the loss of energy for events near the rear dead region of the CCD, which requires knowing the true original location of the energy deposition.

We also associate every cluster with a original source event location that describes the position of the decay that produced the energy deposit. This can be used to create estimated heat maps and identify potentially "hot" components in the detector design. This was not used in our WIMP analysis but is used to inform the design of future generations of CCD experiments.

At this point we also add a unique event ID tag to all of the clusters produced by an event, so we can identify multi-scatter interactions or perform additional event-by-event processing later.

Pixel Binning. The above process (referred to as "1x1" simulation, as it reflects the readout from the CCD without performing any pixel binning) produces a complete simulation output which can be used in our analysis. However, the dark matter search uses 1x100 pixel binning, that collapses 100 pixels in the vertical direction into a single pixel read. This improves our signal to noise ratio for analysis (as the readout noise is only added once for every 100 pixels we read out).

To simulate this effect, we can use either raw simulations with 100-times larger vertical pixels, or we can perform secondary processing of the processed clusters. The latter is much easier and faster to perform, as processing raw `GEANT4` simulations takes large amounts of disk space and processing time, and is the preferred method for DAMIC at SNOLAB simulations.

To produce these 1x100 binned simulations from the processed clusters, the clusters

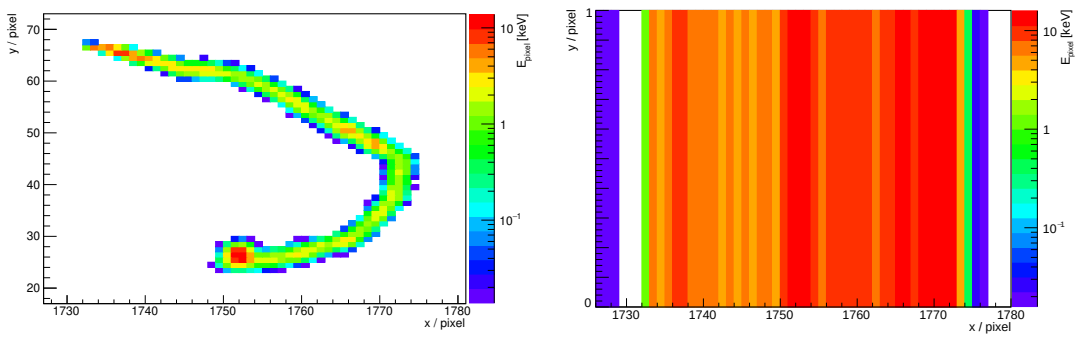


Figure 5.5: Example of the same simulated ^{32}P cluster, before (left) and after (right) applying 1x100 binning. The binned cluster also has noise and saturation effects, although both are negligible for this example.

are loaded into a dedicated C++ code which loads all pixel data on an event-by-event basis (using the unique event ID), since events with multi-site energy deposits may end up being binned into a single 1x100 pixel, if the event generates multiple separate energy deposits the in the same CCD. The effects of this binning process on a simulated ^{32}P decay can be seen in fig. 5.5.

Readout Noise. Readout noise may be applied in the original simulation processing, or applied at a later stage to noise-free simulations. This effect is per-pixel, and therefore must be done after the binning process to be accurately modeled. Applying noise is a relatively simple process: each pixel in a cluster is modified such that

$$P_{\text{new},i} = P_{\text{old},i} + Gaus(0, \sigma) \tag{5.6}$$

where P is the value of the ith pixel, and σ is the noise of the CCD (which is typically set at 2 e^- or 7.54 eV). This noise is applied to both pixels in the simulation that are known to have charge, and to a 5-pixel area around, to simulate a more realistic readout scenario (where the pixels with real charge may be difficult to identify among the noise).

After noise is applied, the clustering process is always re-ran, and only pixels with charge $>4\sigma$ are added to the cluster. This results in an effective cut of 30 eV on simulations.

Pixel Saturation. Saturation values for pixels are simulated by drawing random pixel baselines for the given CCD from SNOLAB data and computing the saturation

value for a 16-bit maximum pixel value (e.g. saturation = 65536 − baseline). This value (which is in ADC counts) is then converted to keV using the per-CCD calibration constants. Any pixel in a simulated cluster greater than this saturation value is set to the saturated value. This method allows us to account for the variation of saturation values between different CCDs, and for fluctuations in the saturation value over time due to changes in the baseline (as shown in fig. 5.6). Note that saturation is only applied to 1x100 binned data. The 1x1 background data was taken with different configuration parameters that result in a much higher saturation value which is not relevant for our analysis.

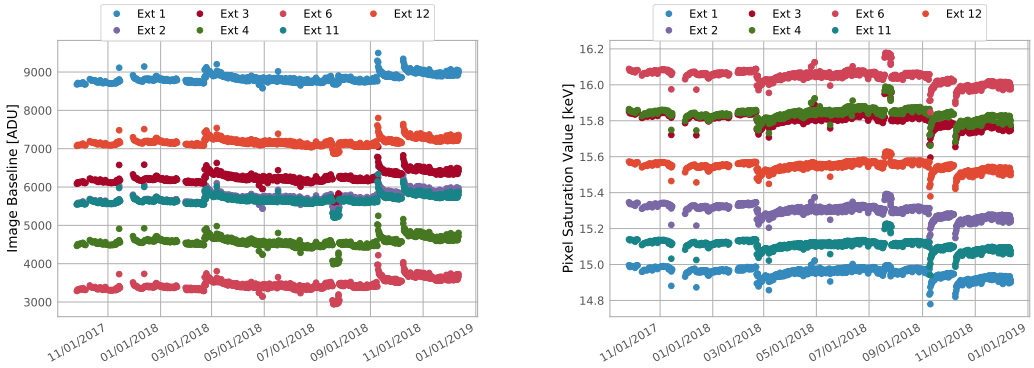


Figure 5.6: Baseline in ADU (left) and per-pixel saturation value in keV (right) of DAMIC at SNOLAB data over time, for 1x100 science data, from the beginning of the data run in September of 2017 until the end in December 2018. Gaps and large shifts in the baseline occur due to interruptions for mechanical issues (such as a failure of the vacuum system in November 2017) or calibration runs.

5.3 Development Simulations

Extensive simulations were done at University of Chicago, using local server resources (as well as some University of Chicago Research Computing Center resources, specifically the Midway cluster). These early simulations acted as both proof-of-concept, to demonstrate that the `GEANT4` simulation of the SNOLAB detector could produce accurate results (as shown in fig. 5.7 and fig. 5.8), as a development testbench for identifying errors in the simulations codebase, and for the various specific background studies (such as Aguilar-Arevalo et al. [5]). The modifications and scripts to parallelize and automate simulations

were also developed at this time, allowing large sets of simulations to be created from simple input files, and split to allow running across many threads and many processors simultaneously.

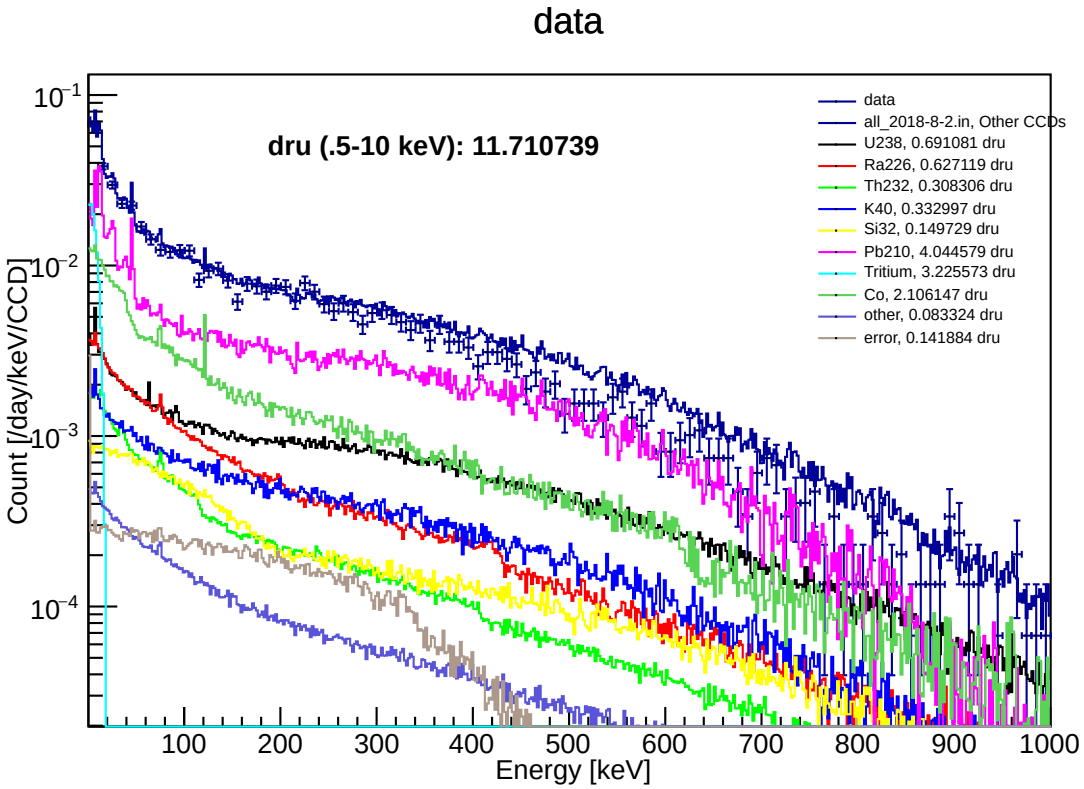


Figure 5.7: Example of an early development simulation, comparing the simulation results (various colored lines) to background 1x1 data from DAMIC at SNOLAB (the blue crosses). Data is from CCDs 2-7 (CCD 1 was excluded from the early analysis, due to the relatively low statistics). No fitting is employed in this example: the activity rates for various simulated components were the central value or upper limits taken from table 5.1. The rate in d.r.u. is that predicted from the simulation.

5.3.1 Simulation Model Verification

Three data comparisons were available to validate the `GEANT4` simulation results for CCD data sets. The first is a set of data taken with a ^{3}H source, an Eckart & Zeigler model AF-4070 source with a nominal 10 kBq activity. The ^{3}H is anodized to a 5 micron thick aluminum planchet, which is glued to a larger aluminum backing disc. A picture of the source (in the holder and frame used for taking data with the CCD) can be seen in fig. 5.9 (left). The source was placed directly over the CCD, and 2.1 days of data were taken.

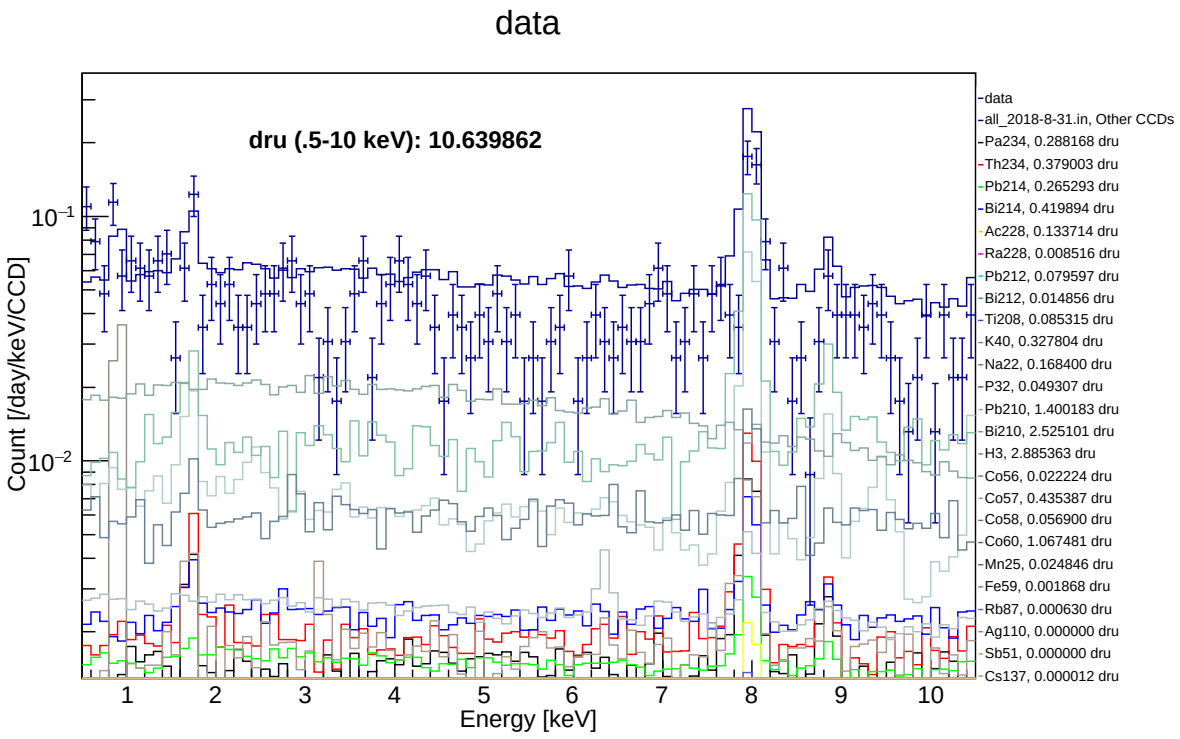


Figure 5.8: Development simulation comparing spectra for 1x100 science data to simulations. As above, only CCDs 2-7 were used, and the rates were the central value or upper limit taken from table 5.1.

To compare this data to our GEANT4 simulation, an *ad hoc* simulation was written using a simplified geometry that included only the CCD, source, and aluminum holder plate. A wire-frame picture of this geometry can be seen in fig. 5.9 (right). We then ran simulations using several different GEANT4 physics lists (Livermore[6], Penelope[7], and Microelectronics[8]) and compared them to the data, as shown in fig. 5.10. This provided three useful validations: first that the ^{3}H generator followed the correct energy spectrum (in GEANT4, the ^{3}H atom is stable, and so creating a ^{3}H beta decay spectrum requires manually generating electrons with the appropriate energy-momentum distribution). Second, it validates the GEANT4 physics list itself: we were able to identify a significant issue in the Microelectronics physics list[2] in this comparison, thanks to the massive disagreement of the simulation with the data. Finally, it validated our simulations processing code, since it compared the spectrum from processed CCD images to processed simulation images.

The second and third physics validations were comparisons to the Compton scattering

2. This physics list was of interest since it was created with the goal of simulating silicon-based semiconductor electronics similar to our CCDs.

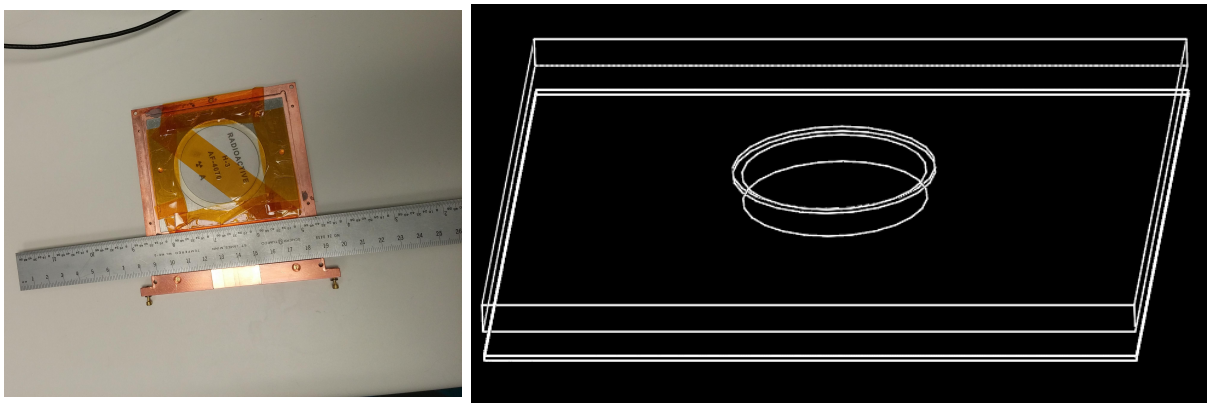


Figure 5.9: Left: Picture of tritium source installed in copper frame. This frame was positioned directly over the CCD, allowing a direct line of site between the CCD and the source. **Right:** Wire-frame schematic of ^{3}H simulation geometry

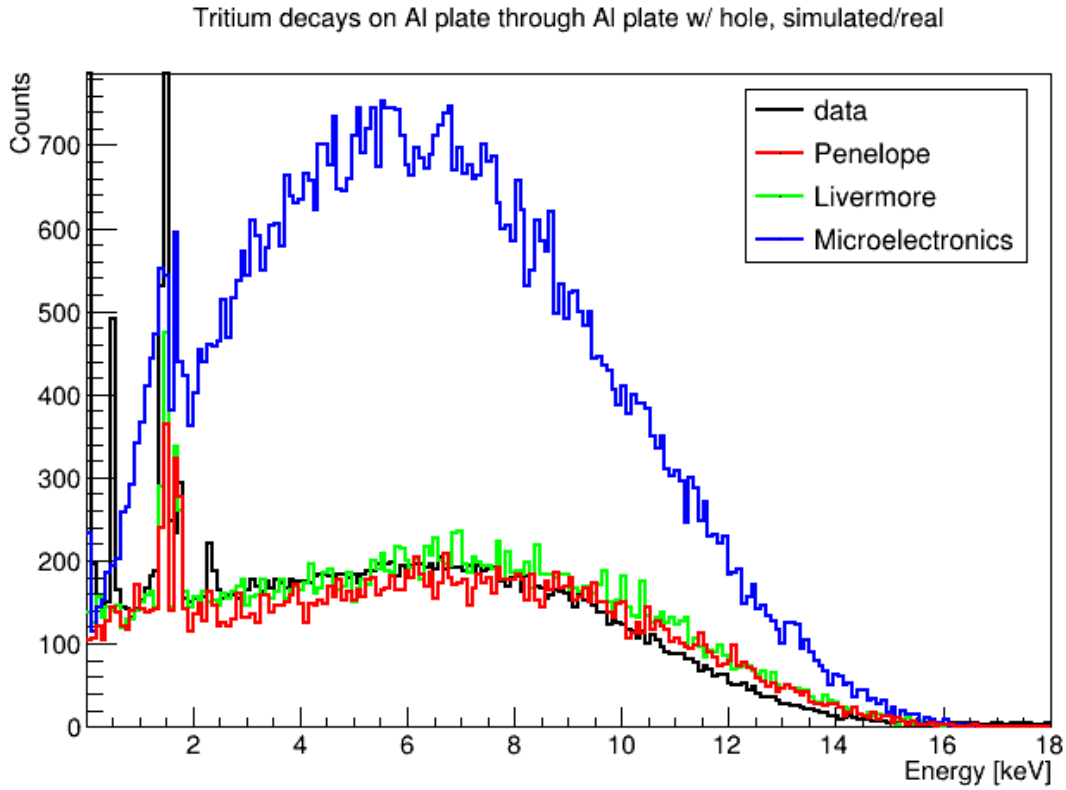


Figure 5.10: Comparison of `GEANT4` simulation to tritium data for several different `GEANT4` physics lists. Livermore (green) and Penelope (red) agree well with each other and with the data: the slight difference at higher energies is from a mismatch of the dead layer depth on the front of the CCD in this simulation. The Microelectronics (blue) results disagree with both data and the other lists due to a bug in this physics list.

models created from ^{57}Co and ^{241}Am data sets. Rather than comparing the simulations to the raw dataset, we compared the simulations to the model created from these datasets, as described in [9]. The results can be seen in fig. 5.11. This model for Compton scattering

	^{238}U	^{226}Ra	^{210}Pb	^{232}Th	^{40}K	^{32}Si
CCD/ Si frame	<11	<5.3	<160	<7.3	<0.5	140 ±30
Kapton Cable	58000 ± 5000	4900 ± 5700	...	3200 ± 500 $^{[M]}$	29000 ± 2000 $^{[M]}$	...
OFHC Copper	<120 $^{[M]}$	<130 $^{[G]}$	27000 ± 8000	<41 $^{[M]}$	<31 $^{[M]}$	...
Module Screws	16000 ± 44000 $^{[G]}$	<138 $^{[G]}$	27000 ± 8000	2300 ± 1600 $^{[G]}$	28000 ± 15000 $^{[G]}$	...
Ancient Lead	<23	<260 $^{[G]}$	33000	2.3	<5.8 $^{[M]}$	...
Outer Lead	<13	<200 $^{[G]}$	$(19 \pm 5)\times10^{6}$	<4.6	<220	...

Table 5.1: Measured radioactive activity of components in the DAMIC at SNOLAB detector, by parent isotope, in units of μBq/kg. A superscript M (G) is used to indicate a value measured with mass spectrometry (gamma counting).

was created directly from CCD data using external radioactive gamma ray sources at the University of Chicago. This model was built with the specific goal of providing a model of gamma ray backgrounds for silicon dark matter experiments at the low energies relevant to DAMIC at SNOLAB. Since the model was created using CCDs very similar to DAMIC CCDs, the model is expected to match real Compton scattering backgrounds very well, and the agreement between our `GEANT4` simulation and the model indicates the simulation is accurately reproducing this spectrum.

A final simulation used to help verify the `GEANT4` simulation chain is the ^{210}Pb decay. This decay has many interesting but known features, due to the complexity of the decay process (see fig. 5.12). No CCD dataset exists for quantitative comparison, but the simulation was useful to qualitatively verify that the simulation was producing the expected features. For example, some sets of simulations produced unexpected features at low energies that allowed us to identify an issue with the simulation geometry.

5.4 Simulation Activity Rates

A major source of backgrounds in the DAMIC experiment is from activation of materials due to exposure to cosmic ray neutrons. These cosmic ray neutrons can cause nuclear spallation of copper and silicon in the detector, resulting in backgrounds proportional to the time the components were on the surface (^{60}Co and ^{3}H are the most relevant backgrounds for Cu and Si, respectively). The average activity A for an isotope can be

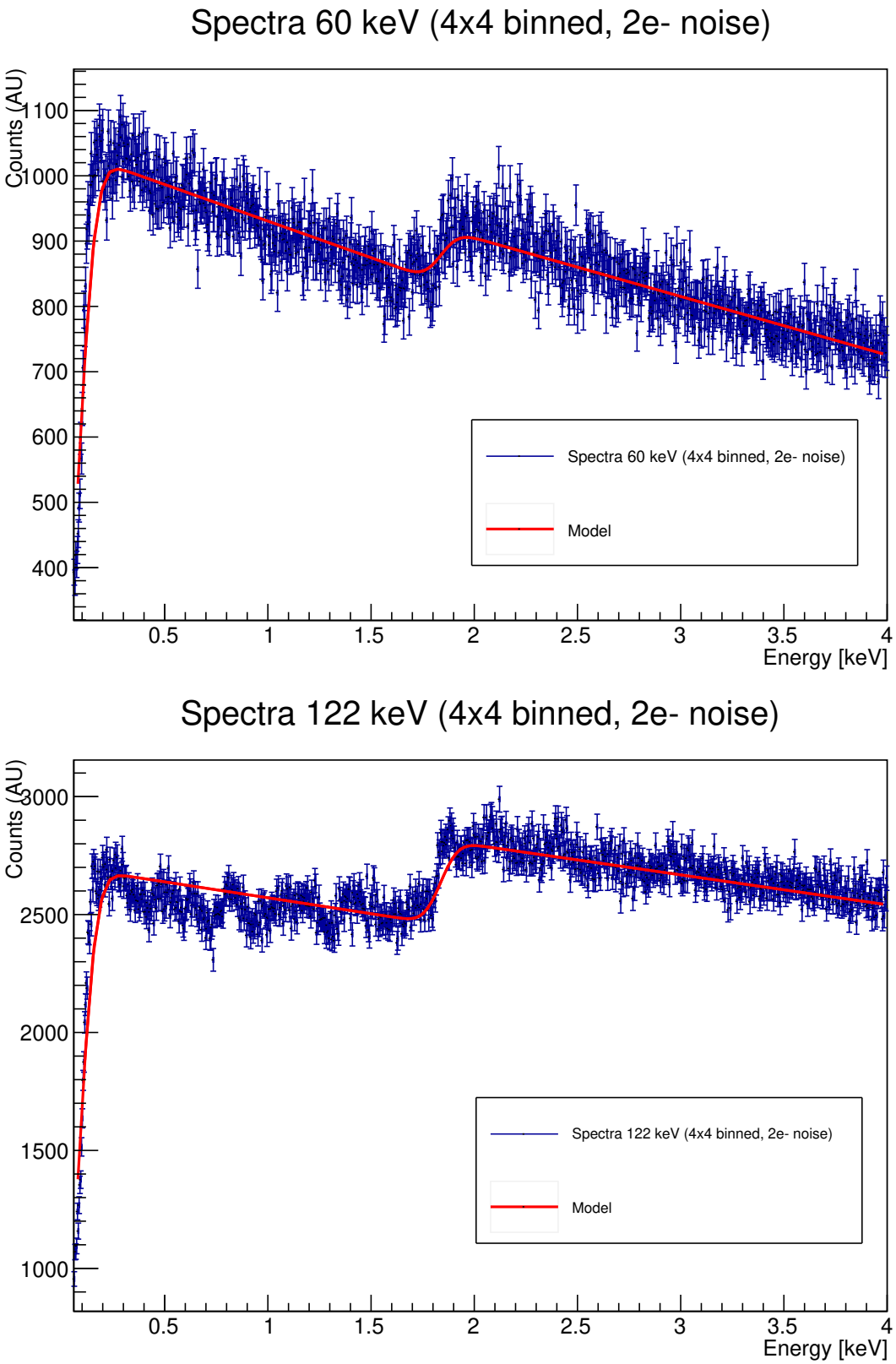


Figure 5.11: Comparison of `GEANT4` simulations to Compton scattering model developed in [9], using the ^{241}Am 60 keV γ rays spectrum (top) and the ^{57}Co 122 and 136 keV spectrum (bottom). The simulations were passed through the standard processing chains, and used 4x4 binning and $2e^-$ of noise to match the data acquisition parameters used in creating the model.

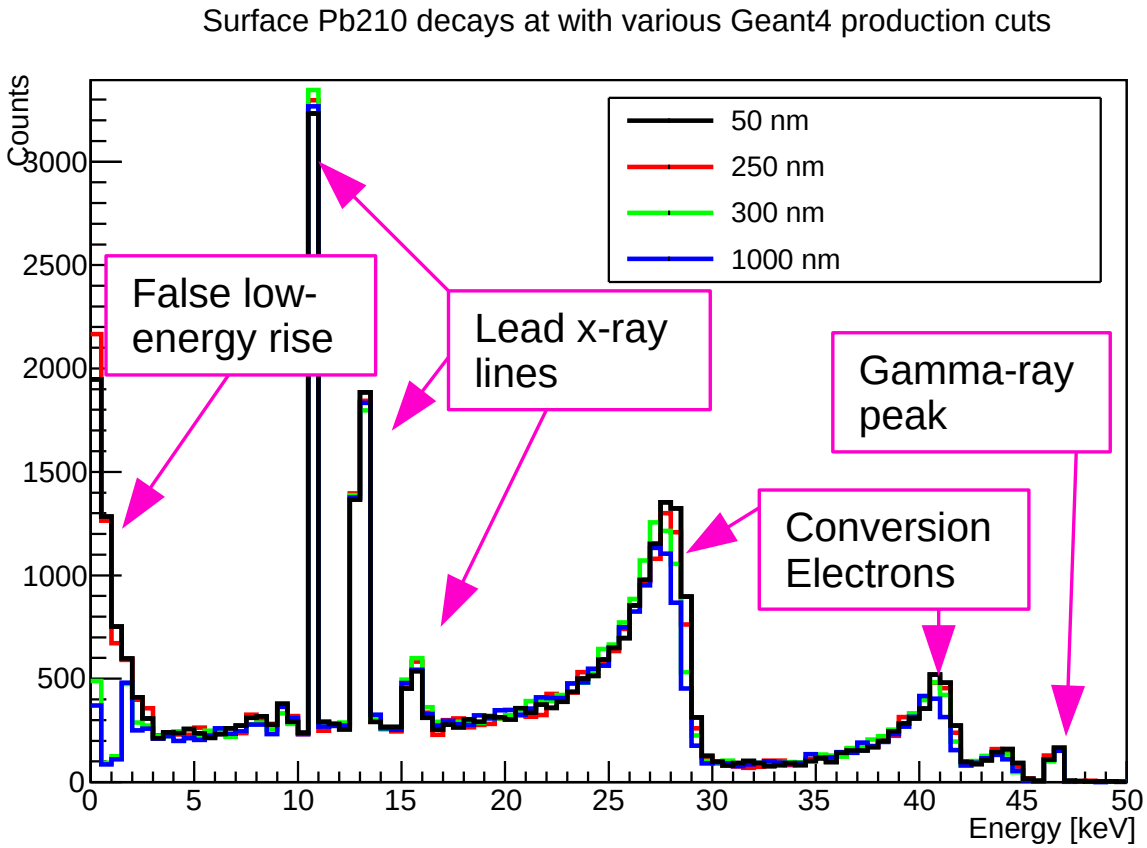


Figure 5.12: GEANT4 simulation of ^{210}Pb spectra for various GEANT4 range cuts. The false rise at low energies was due to a detector geometry error: the simulation contained multiple overlapping components which were not properly detected by GEANT4, resulting in multiple deposits per particle (this was fixed in later simulations). This was most noticeable for the low range cuts, but likely affected the entire simulation for all energies and ranges.

computed using the equation

$$A = \frac{S}{\lambda T_{exp}} \left(1 - e^{-\lambda T_{act}}\right) \left(e^{-\lambda T_{cool}}\right) \left(1 - e^{-\lambda T_{exp}}\right) \tag{5.7}$$

where $\lambda = \log(2)/\tau_{1/2}$ is the decay constant of the isotope (listed in table 5.2), S is the saturation activity of the isotope at sea level (given in table 5.2), and T_{act}, T_{cool}, and T_{exp} are the activation, cooldown, and exposure times respectively (which depend on the dataset used in the analysis). For the science dataset, the activation (cooldown) times of the copper modules, box, and vessel were 8 months, 16 months, and 1000 years[3] (540 days, 300 days, and 6.6 years) respectively, and the exposure time was 441 days.

3. The exact history of the copper in the vessel is unknown, so a 1000 year time period was used to force the rate to saturation values.

Parent Chain	Isotope	Q value	Half-Life	Saturation Rate
^{238}U	^{234}Th	274 keV	24 day	
	^{234m}Pa	2.27 MeV	1.2 min	
^{226}Ra	^{214}Pb	1.02 MeV	27 min	
	^{214}Bi	3.27 MeV	20 min	
^{210}Pb	^{210}Pb	63.5 keV	22 year	
	^{210}Bi	1.16 MeV	5 day	
^{232}Th	^{228}Ra	45.5 keV	5.8 year	
	^{228}Ac	2.12 MeV	6.1 hour	
	^{212}Pb	569 keV	10.6 hour	
	^{212}Bi	2.25 MeV	61 min	
	^{208}Tl	5.00 MeV	3.1 min	
^{40}K	^{40}K	1.31 MeV	$1.251{\cdot}10^{9}$ year	
Copper	^{60}Co	2.82 MeV	5.2 year	2100
Activation	^{59}Fe	1.56 MeV	44 day	455
	^{58}Co	2.31 MeV	71 day	1650
	^{57}Co	836 keV	271 day	1800
	^{56}Co	4.57 MeV	77 day	230
	^{54}Mn	1.38 MeV	312 day	215
	^{46}Sc	2.37 MeV	84 day	53
^{32}Si	^{32}Si	227 keV	172 year	
	^{32}P	1.71 MeV	14 day	
Silicon	^{22}Na	2.84 MeV	2.6 year	4.3
Activation	^{3}H	18.6 keV	12.32 year	10.7

Table 5.2: Isotopes simulated for the DAMIC at SNOLAB background analysis, grouped by parent decay chain with Q values and saturation rates in atoms/kg (for cosmogenic isotopes). Activation rate from [10] for copper, and [11] for ^{3}H and ^{22}Na.

5.4.1 ^{22}Na

The activation rate of ^{22}Na is well measured in silicon (by [11] and others). However, it's decay scheme allows a unique ability to measure the decay rate using our detector itself. 90.3% of ^{22}Na decays result in a β^+ emission, which escapes or produce high-energy annihilation γs that are unlikely to be seen by the CCD. However, 9.7% of decays occur by electron capture to the $\epsilon_{0,1}$ excited state of neon. The relaxing neon will emit a 1.274 MeV γ, which escapes the CCD, as well as Auger electrons or K-alpha x-rays, which form an observed line around 850 $\mathrm{eV_{ee}}$. This line can be seen in our data (see fig. 5.13), and counting the number of events divided by the exposure and correcting for the decay intensity yields an activity of 0.32±0.06 mBq/kg.

Interestingly, this also highlighted an error in the initial `GEANT4` simulation. This decay peak was seen in the development simulations at 960eV, over 100 eV higher than expected. This was due to a mistake in the `GEANT4` physics assumptions: the escaping 1.568 MeV neutrino and 1.274 MeV γ imparts an average kinetic energy to the nucleus of 100 eV (from straightforward kinematics). In a real silicon detector, this energy is almost entirely quenched to produce negligible ionization, since the energy is deposited by the recoiling nucleus through multiple different mechanisms (see [12]). The `GEANT4` simulation by default kills this recoiling nucleus, and adds the energy to the decay. This effect was corrected in the final simulation by eliminating the energy of the recoiling nucleus, to simulate the effect of quenching.

5.5 Surface Backgrounds

One of the major uncertainties and dominant background source in DAMIC at SNOLAB is the presence of ^{210}Pb contamination on various surfaces. ^{210}Pb originates when ^{222}Rn is produced in the ^{238}U chain. This radon diffuses into the air near detector components, where it decays ($\tau_{1/2} = 3.8$ days) through a series of short-lived ($\tau_{1/2} \approx$ minutes) daughter ions, which are attracted by ionic forces to nearby surfaces, before decaying into ^{210}Pb, with a $\tau_{1/2} = 22$ years. This decay process can drive the ^{210}Pb into the surface up-to a

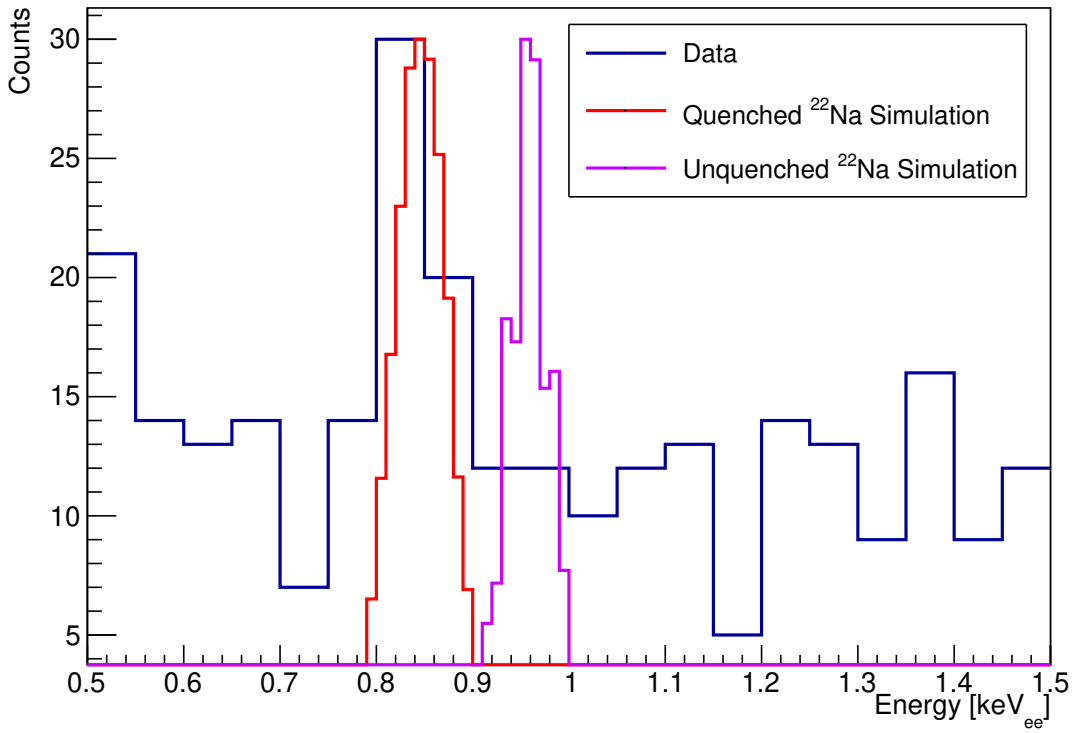


Figure 5.13: The ^{22}Na peak in our dataset, compared to the peaks in a quenched (red) and unquenched (violet) simulations. Note that the "quenching" employed here kills all energy from the recoil, rather than following the expected distribution (but at these energies, nuclear recoil is expected to produce negligible ionization, if any). The simulations are scaled to match the height of the ^{22}Na peak in data for reference. Note that the actual number of ^{22}Na events in the data is the peak minus the flat baseline of events.

depth of 100nm[2]. We assume that the depth distribution of such surface contamination follows the complementary error function:

$$\text{erfc}(\frac{z}{M_d}) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2} dt \tag{5.8}$$

with a characteristic depth $M_d = 50$ nm.

The contamination from ^{210}Pb is difficult to measure precisely, since the decay energy is too low for most gamma counters. The surfaces of all components in the detector near the CCD are cleaned to remove the top few microns of material before the detector is assembled. However the CCD itself cannot be cleaned in this manner (without significant damage), and this may have significant deposition of events. The spatial coincidence search detailed below offers some measurement for the activity, but the distribution of

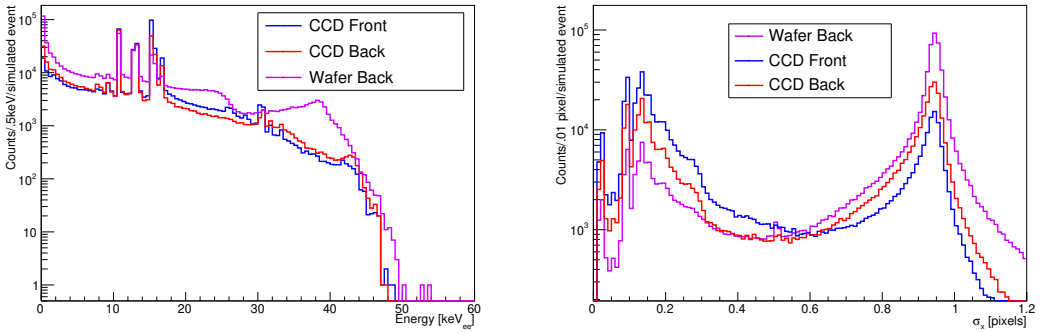


Figure 5.14: Comparison of the ^{210}Pb spectra in energy (left) and σ_x (right), for different locations. The different in energy between front and back of the CCD is very small, but there is a clear difference in the magnitude of the front and back events in the σ_x. Nevertheless, the two event profiles are fairly degenerate. The wafer back spectrum, however, is clearly distinct in both energy and σ_x space, which allows the fit to clearly differentiate between it and the CCD surfaces.

the contamination is still unknown. Before the CCD is fabricated, the silicon wafers that form the basis of the CCD were left exposed in a vertical position for several years. The front side of the wafer is polished to remove a few microns of material (effectively removing all contamination on the front wafer surface). The back side of the wafer, however, is unpolished, and may retain significant contamination from radon exposure. This original wafer corresponds to a depth ~3 μm into the backside of the CCD (the additional material is deposited during CCD fabrication). This allows for three locations for the ^{210}Pb on the CCD: the front and back of the CCD, and the back of the original wafer. The difference in spectral shapes between these three locations can be seen in fig. 5.14.

5.6 Coincidence Searches

One major advantage of pixelized solid state detectors, such as the CCDs used by DAMIC, is the ability to perform spatially correlated coincidence searches. Unlike liquid or gaseous experiments, decay daughter isotopes will remain in exactly (or nearly exactly) the same location as the parent decay. The micrometer-scale resolution of CCDs allows for precise spatial resolution to detect these coincidences. The only requirement is that both decays must be likely to be detected by the CCD (neutrinos or high energy gamma rays are more

likely to escape), and that the half-life of the daughter is sufficiently short to be likely to decay in the same data run as the parent.

For most parent-daughter isotopes chains, this search is a fairly simple analysis that can be performed without a Monte Carlo simulation. However there are two chains that are particularly relevant to CCDs that require simulations to properly quantify the spatial coincidence detection efficiency: these are the ^{22}Si-^{22}P decay, and the ^{210}Pb-^{210}Bi decays.

The first is solely a bulk contaminant: ^{32}Si is produced in the atmosphere by cosmogenic neutron spallation of Argon in the atmosphere , which can then end up inside the silicon used to produce the CCD wafer. This ^{32}Si is only a trace amount, but it is sufficient to produce a noticeable background in dark matter experiments, and cannot be removed except by isotopic enrichment (which is vastly too expensive, especially for such a trace isotope). ^{32}Si decays to ^{32}P through a beta decay with a Q-value of 227keV and a half-life of 172 years. ^{32}P then decays through a second beta decay, with a half-life of 14 days and a Q-value of 1.71 MeV. The extended tracks created by both parent and daughter result in a limited selection efficiency, which is quantified by the `GEANT4` simulation. Note that this search is split into two: the first requires the first event to have energy between $70 < E_\beta < 230$ keV, and time between events $\Delta t < 70$ days. The second requires $0.5 < E_\beta < 70$ keV, and $25 < \Delta t < 70$ days, to minimize overlap with the ^{210}Pb search. An example of a coincidence event can be seen in fig. 5.15.

The ^{210}Pb decays, on the other hand, can be either a bulk contaminant from uranium-chain isotopes remaining in the wafer after purification, or more dominantly as a surface contaminant, produced after radon in the air decays to lead and settles on any exposed surface of the detector. The latter is especially relevant, as the detectors in DAMIC at SNOLAB had no particular precaution against radon contamination of their environment, and could have years of contamination from radon dust during storage. ^{210}Pb has two possible beta decays (listed in table 5.4), both of which produce a ^{210}Bi daughter, which itself beta decays with a 5 day half-life and a Q-value of 1.16 MeV.

Both decays have relatively complicated decay signatures due to the extended nature of the β decay tracks, and a robust detection efficiency calculation requires a Monte Carlo

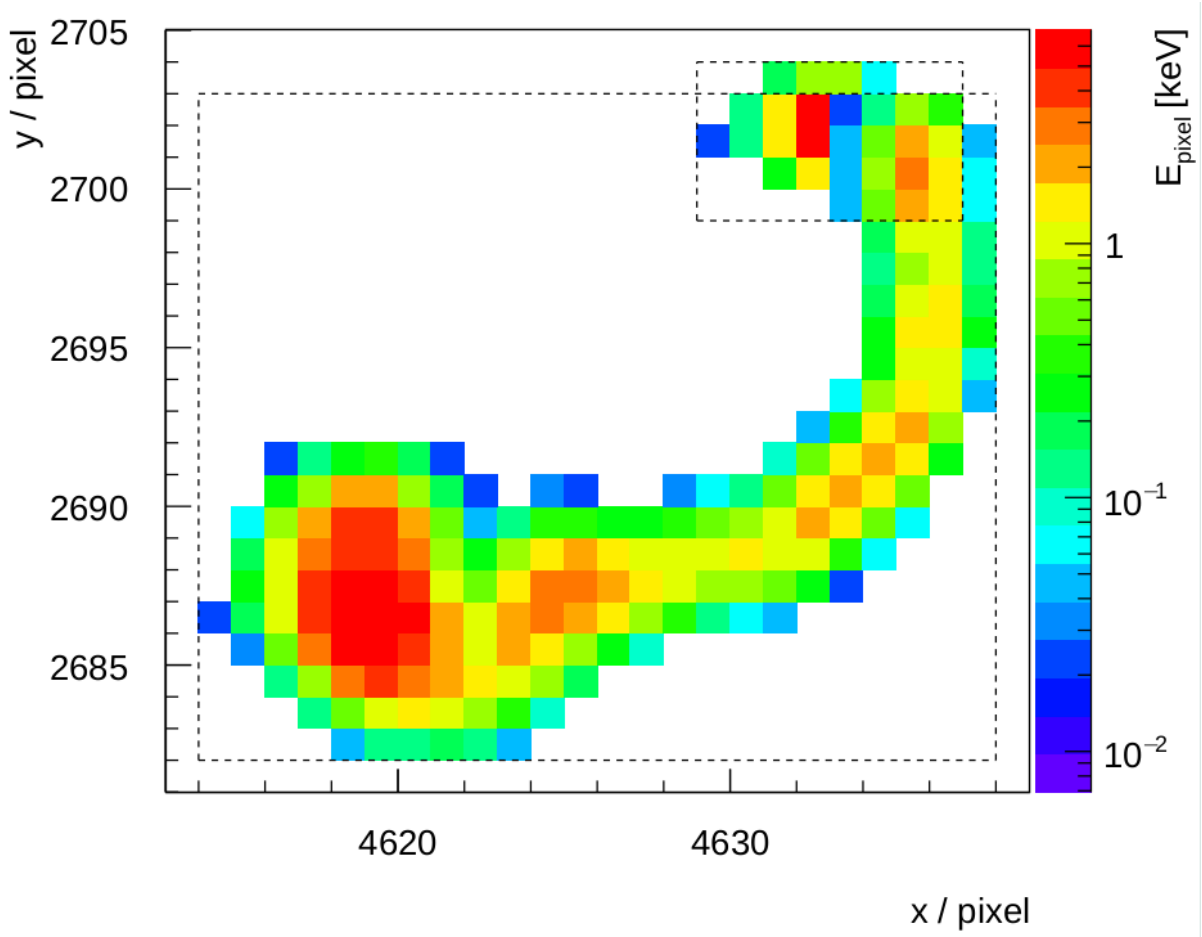


Figure 5.15: Example of a spatially-correlated ^{32}Si-^{32}P decay. The smaller box shows the ^{32}Si decay, which occurred first with an energy of 77 keV, followed by the 399 keV decay, 13.7 days later. The identification is made by the time separation, spatial coincidence, and energy of the first and second decay.

simulation. To produce this simulation, the `GEANT4` particle source was modified to allow specifying decay events from a plain-text file. We then ran ^{32}Si and ^{32}P simulations using the same set of origin coordinates (to simulate the decay occurring from the same physical location). These sets of simulations were then used to compute the detection efficiency for the coincidence search. The results for bulk decays have been published in [5], and are summarized in table 5.3.

This same procedure was employed for ^{210}Pb and ^{210}Bi. For bulk decays, this was less necessary (as the detection efficiency is near 1 anyways), but ^{210}Pb in particular has a complex decay scheme (see table 5.4) that usually produces multiple decay products. In addition, decays near the surface can produce events in multiple CCDs or fairly far from the originating decay (as seen in fig. 5.16).

Isotope		ϵ_{sel}	ϵ_t	N_{pair}	N_{acc}	N_{overlap}	Rate
^{32}Si	(1a)	0.398	0.701	26	6.5 ± 0.1	0	120 ± 30 μBq/kg
	(1b)	0.521	0.169	17	3.0 ± 0.1	1.3 ± 0.3	260 ± 80 μBq/kg
^{210}Pb	(2a)	0.981	0.748	69	2.5 ± 0.1	19.4 ± 5.1	< 160 μBq/kg
	(2b)	0.138	"	"	"	"	255 ± 28 nBq/cm^2
	(2c)	0.315	"	"	"	"	120 ± 12 nBq/cm^2

Table 5.3: Details for the $\beta \to \beta$ coincidence searches in DAMIC at SNOLAB background data, including pair selection and time efficiencies (ϵ_{sel} and ϵ_t, respectively), number of coincident pairs found (N_{pair}), number of accidental pairs expected (N_{acc}), and number of pairs expected from overlap with other decay searches (N_{overlap}). The (1a) and (1b) searches for ^{32}Si decays are split to reduce overlap with the ^{210}Pb. The ^{210}Pb search rates assume *all* of the 69 events (minus overlap and accidentals) occur in the bulk (2a), front (2b) or back (2c). The rates from the background model fit are 69 ± 12 (56 ± 8) nBq/cm^2 for front (back).

Emission Type	Energy (keV)	Frequency
γ	46.5	4.25
$\beta^-_{0,1}$	4.3 (17.0)	80.2
$\beta^-_{0,0}$	16.3 (63.5)	19.8
X-ray	9.4 - 15.7	22.0
e_{AL}	5.3-10.7	36.0
$ec_{1,0L}$	30.1-33	58
$ec_{1,0M}$	42.5-44.0	13.7
$ec_{1,0N}$	45.6-46.4	3.5

Table 5.4: Possible decay emissions from ^{210}Pb decays, showing the particle emitted and relative decay probability. A rare alpha decay, as well as the neutrino, are omitted. Energy is shown as average (max) for beta decays, and min-max range for x-rays or conversion electrons. Frequency is per 100 decays. "ec" indicates conversion electrons emitted by the excited Bismuth atom from the labeled shell (L,M, or N), following the beta decay to the excited (0,1) state (decays to this state result in either emission of a conversion electron, or of the 46.5 keV gamma ray). All numbers taken from [13].

5.7 Background Model

The final WIMP analysis simulations were run at the CC-IN2P3 computing center in France, before being transferred to Chicago for processing. After processing, the simulations were converted into a useful background model for the WIMP search by fitting them to our subset data. Since other experiments have well constrained the WIMP search space above 10 GeV, we chose to limit our dark matter search range to events below 6 keV$_{\text{ee}}$, leaving the higher energy region free to constrain our background model, while blinding the fit to our search region. Saturation effects limit the spectral sensitivity

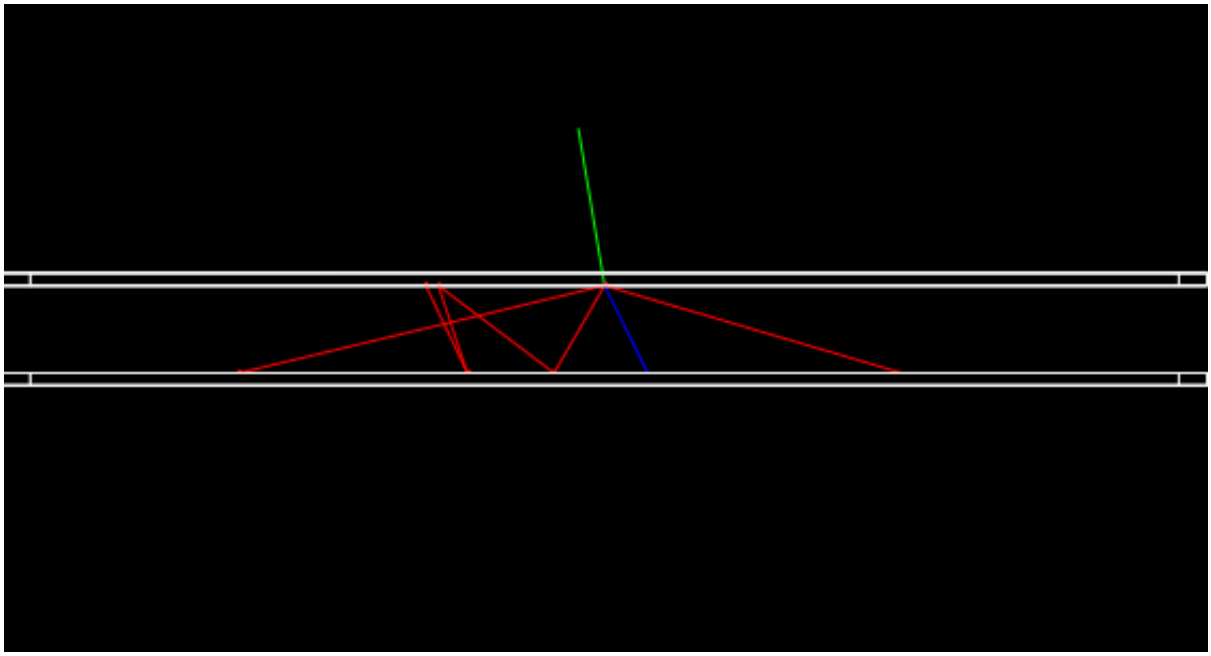


Figure 5.16: Example of GEANT4 simulation ^{210}Pb surface decay, using a simplified geometry with two CCDs. The white wire-frame planes show the CCDs themselves. The lines are different products of the decay: the red lines show electrons (beta particles and conversion electrons) from the decay, the green line shows the emitted neutrino (which is never seen in our detector, but simulated by GEANT4), and the blue line shows the ^{210}Bi ion after the decay.

above 14 keV$_{ee}$, but in order to cover the entire ^{3}H decay energies, we chose to include the spectrum up to 20 keV$_{ee}$. We also excluded the region [7.5, 8.5] keV$_{ee}$ from the fit, since this corresponds to the copper florescence peak not well simulated by GEANT4. The data from CCD 1 was not used in the fit, and instead kept as a cross-check to the fit results. The fitting process was developed on 64% of the total science data, with the remaining being held blinded until the analysis methodology was fixed, and added only afterwards to increase the statistics of the final result.

To construct our model, the simulations were converted into binned templates. Simulations were grouped into common normalizations by detector part and decay chain as specified in table 5.1 (with the further subdivision of copper into the copper box, vessel, and modules[4], since the copper for those components has different sources and histories). Daughter isotopes were assumed to be in secular equilibrium with parents, except were otherwise specified. To prevent simulations with low statistics from artificially driving the fit, any simulations resulting in fewer than 0.1 events/kg-day and less than 1000 events

4. The electroformed copper module used for CCD 1 was assumed to be perfectly radiopure

reaching the CCD were ignored. After these cuts, 289 simulations were left, grouped into 49 templates.

The background model fit is performed simultaneously in both energy and σ_x space to maximize the information available from the data. Each bin i in energy is 0.25keV$_{\rm ee}$ wide, and each bin j in σ_x is 0.025 pixels. The number of events for each template l, in each ij-bin is

$$\nu_{ijl} = \sum_m n_{ijm} \times A_l \; M_m \; \frac{\epsilon_{data} \; t_{exp}}{\epsilon_{sim} \; N_m}. \tag{5.9}$$

where n_{ijm} is the number of events in the mth simulation in that template, A_l is the initial guess of the activity from table 5.1, M_m is the mass of the part in the detector, $\epsilon_{data} \; t_{exp}$ is the efficiency-corrected exposure of the data, and $\epsilon_{sim} \; N_m$ is the efficiency-corrected number of decays simulated.

The templates are then combined into an expected number of events

$$\nu_{ij} = \sum_l C_l \nu_{ijl} \tag{5.10}$$

where C_l is the scaling factor for the lth template (initially set to 1). This scaling factor provides the relative activity of the template in the fit compared to the expected values given in table 5.1.

The fit itself is performed using a 2D log-likelihood

$$LL_{2D} = \sum_i \sum_j \left(k_{ij} \log(\nu_{ij}) - \nu_{ij} - \log(k_{ij}!) \right) \tag{5.11}$$

where k_{ij} is the number of events in the data in th ijth bin. For cases with k>100, Stirling's approximation is used, so that $\log(k!) \approx k \log(k) - k$. For bins with no simulated events ($\mu_{ij} = 0$), the likelihood is set to zero and the bin discarded.

The minimization of the likelihood is subject to a number of nuisance parameters, such that

$$LL_{total} = LL_{2D} - \sum_n \frac{(N_n^0 - N_n)^2}{2\sigma_n^2}, \tag{5.12}$$

where N_n^0 is the nominal value of the n^{th} nuisance parameter with uncertainty σ_n and

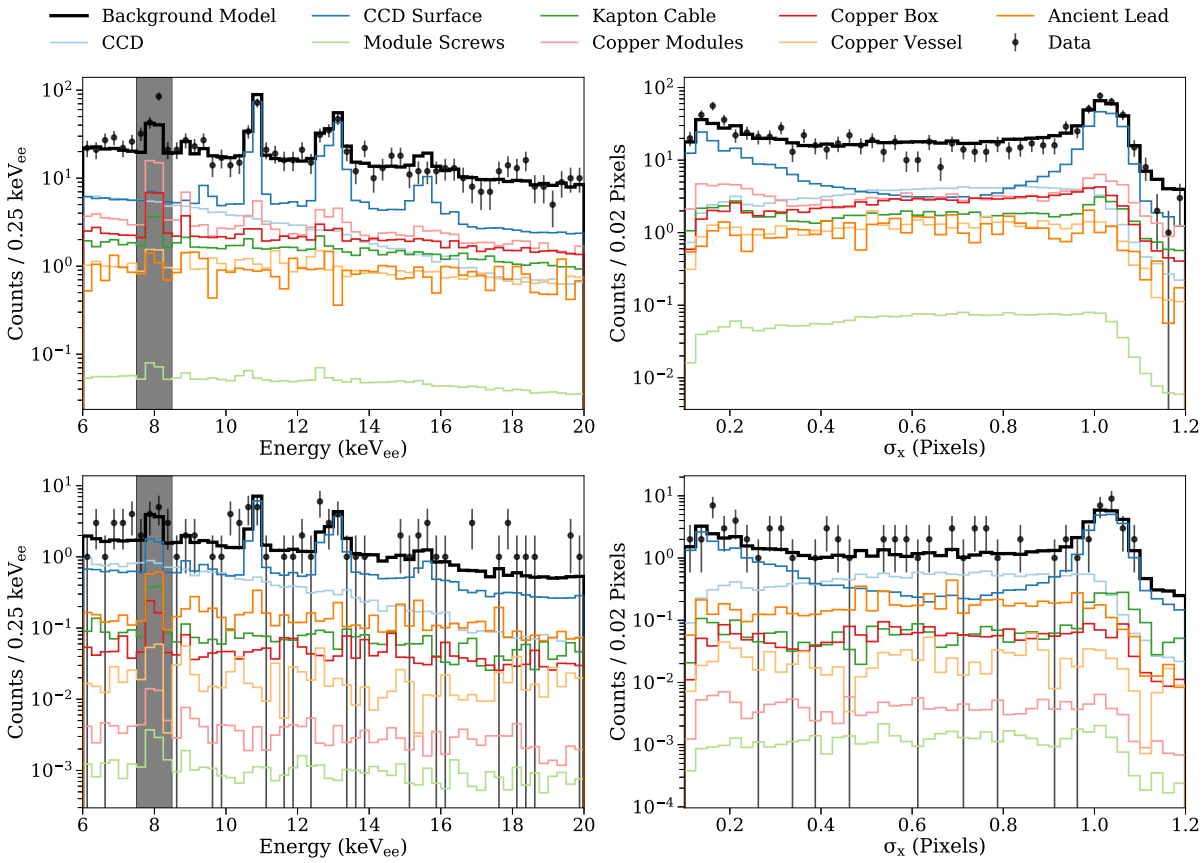


Figure 5.17: Fit results showing various components of DAMIC at SNOLAB background model, grouped by the detector part, for CCDs 2-7 (top), and 1 (bottom), broken up into energy (left) and σ_x (right). The largest contribution is from the CCD surface and CCD itself (from ^{210}Pb and ^{32}Si/^{3}H contamination of the surface and bulk, respectively), followed generally in order by the distance from the CCD (with the ancient lead and copper vessel contributing the least, and the copper modules holding the CCDs the most). The exception is the module screws, which have high activity but relatively small mass, resulting in them contributing very little total background.

N_n is the best fit value. These nuisance parameters use the uncertainty given in table 5.1 when available. When only an upper limit exists, the scaling parameter is allowed to vary freely from 0 to 1, with a 10% uncertainty applied if the fit value exceeds 1 (to prevent a hard rail in the fit). Likewise, for activation values, a 10% uncertainty is assumed. The ^{3}H and surface ^{210}Pb values are left completely unconstrained, since this analysis is blind to the cosmogenic activation history of the CCD, and the exact distribution of the surface ^{210}Pb is not well understood.

The results of this fit are shown broken into the individual components in table 5.5, and a plot of the model (grouped by detector part) compared to data is shown in fig. 5.17.

	Detector Part	Chain	C_l	Best Fit Activity	Rate (dru): CCDs 2–7 1–6 keV	CCDs 2–7 6–20 keV	Rate (dru): CCD 1 1–6 keV	CCD 1 6–20 keV
1	CCD	^{238}U	0.897	$\lesssim 9.86$ μBq/kg	0.01	0.01	< 0.01	0.01
2	CCD	^{226}Ra	0.900	$\lesssim 4.79$ μBq/kg	0.01	0.01	< 0.01	< 0.01
3	CCD	^{232}Th	0.900	$\lesssim 6.56$ μBq/kg	0.01	0.03	0.01	0.02
4	CCD	^{40}K	0.910	$\lesssim 0.42$ μBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
5	CCD	^{22}Na	1.066	340 ± 60 μBq/kg	0.17	0.16	0.10	0.09
6	CCD	^{32}Si	1.042	150 ± 30 μBq/kg	0.20	0.17	0.15	0.13
7	CCD	^{3}H	1.131	330 ± 90 μBq/kg	2.87	0.87	2.39	0.74
8	CCD (front surf.)	^{210}Pb	1.658	69 ± 12 nBq/cm^2	1.51	1.63	0.53	0.91
9	CCD (back surf.)	^{210}Pb	$< 10^{-4}$	< 0.1 nBq/cm^2	< 0.01	< 0.01	< 0.01	< 0.01
10	CCD (wafer surf.)	^{210}Pb	1.343	56 ± 8 nBq/cm^2	2.52	1.83	2.03	1.23
11	Copper Box	^{238}U	0.900	$\lesssim 110$ μBq/kg	0.01	0.01	< 0.01	< 0.01
12	Copper Box	^{226}Ra	0.900	$\lesssim 120$ μBq/kg	0.19	0.16	0.03	0.03
13	Copper Box	^{210}Pb	0.380	10 ± 6 mBq/kg	0.33	0.25	0.01	0.01
14	Copper Box	^{232}Th	0.900	$\lesssim 36$ μBq/kg	0.08	0.07	0.01	0.01
15	Copper Box	^{40}K	0.900	$\lesssim 28$ μBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
16	Copper Box	Act.	1.015	various	0.63	0.51	0.10	0.09
17	Copper Modules	^{238}U	0.900	$\lesssim 110$ μBq/kg	0.05	0.04	< 0.01	< 0.01
18	Copper Modules	^{226}Ra	0.900	$\lesssim 120$ μBq/kg	0.21	0.17	< 0.01	< 0.01
19	Copper Modules	^{210}Pb	0.557	15 ± 4 mBq/kg	1.22	0.89	< 0.01	< 0.01
20	Copper Modules	^{232}Th	0.900	$\lesssim 36$ μBq/kg	0.10	0.08	< 0.01	< 0.01
21	Copper Modules	^{40}K	0.900	$\lesssim 28$ μBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
22	Copper Modules	Act.	1.006	various	0.30	0.24	0.01	0.01
23	Kapton Cable	^{238}U	1.016	59 ± 5 mBq/kg	0.52	0.32	0.24	0.13
24	Kapton Cable	^{226}Ra	1.362	7 ± 5 mBq/kg	0.24	0.19	0.05	0.04
25	Kapton Cable	^{232}Th	1.010	32 ± 0.5 mBq/kg	0.17	0.13	0.04	0.03
26	Kapton Cable	^{40}K	1.003	29 ± 2 mBq/kg	0.09	0.06	0.04	0.02
27	Kapton Cable	Act.	1.000	various	0.01	0.01	< 0.01	< 0.01
28	Ancient Lead	^{238}U	0.911	$\lesssim 21$ μBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
29	Ancient Lead	^{226}Ra	0.900	$\lesssim 230$ μBq/kg	0.45	0.36	0.21	0.21
30	Ancient Lead	^{210}Pb	1.000	~ 33 mBq/kg	0.04	0.03	0.24	0.19
31	Ancient Lead	^{232}Th	1.000	~ 2.3 μBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
32	Ancient Lead	^{40}K	0.916	$\lesssim 5.3$ μBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
33	Outer Lead	^{238}U	0.916	$\lesssim 12$ μBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
34	Outer Lead	^{226}Ra	0.909	$\lesssim 190$ μBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
35	Outer Lead	^{210}Pb	1.000	18 ± 5 Bq/kg	< 0.01	< 0.01	< 0.01	< 0.01
36	Outer Lead	^{232}Th	0.907	$\lesssim 4.2$ μBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
37	Outer Lead	^{40}K	0.906	$\lesssim 200$ μBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
38	Module Screws	^{238}U	1.000	20 ± 40 mBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
39	Module Screws	^{226}Ra	0.900	$\lesssim 1.4$ mBq/kg	0.01	0.01	< 0.01	< 0.01
40	Module Screws	^{210}Pb	1.000	27 ± 8 mBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
41	Module Screws	^{232}Th	1.024	2.4 ± 1.6 mBq/kg	0.02	0.01	< 0.01	< 0.01
42	Module Screws	^{40}K	1.000	28 ± 15 mBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
43	Module Screws	Act.	1.000	various	< 0.01	< 0.01	< 0.01	< 0.01
44	Copper Vessel	^{238}U	0.903	$\lesssim 110$ μBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
45	Copper Vessel	^{226}Ra	0.900	$\lesssim 120$ μBq/kg	0.10	0.09	0.01	0.01
46	Copper Vessel	^{210}Pb	0.731	20 ± 8 mBq/kg	0.06	0.03	< 0.01	< 0.01
47	Copper Vessel	^{232}Th	0.900	$\lesssim 36$ μBq/kg	0.04	0.03	< 0.01	< 0.01
48	Copper Vessel	^{40}K	0.901	$\lesssim 28$ μBq/kg	< 0.01	< 0.01	< 0.01	< 0.01
49	Copper Vessel	Act.	0.486	various	0.33	0.27	0.05	0.04
	Total				12.51	8.67	6.27	3.96

Table 5.5: The fit results for the 49 templates used in the background model. Each template corresponds to a unique chain/detector part combination. The fit parameter C_l is the ratio of the best-fit result to the initial guess of each template activity. The rate for each CCD is given in d.r.u. (counts kg^{-1} day^{-1} keV^{-1}) separately for CCDs 2-7 and CCD 1, in both the 1-6 and 6-20 keV region. The upper limit activities indicate a fit component that is highly degenerate with another component: the upper bound given is the value chosen by the fit.

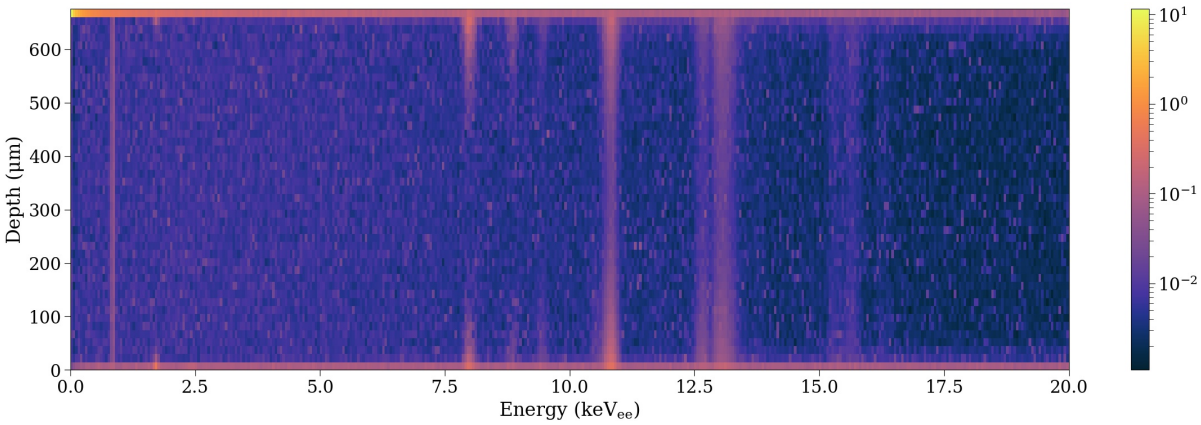


Figure 5.18: E-z plot of background model, over entire energy range used in this analysis. This is shown in simulated coordinates, using the averaged z value for events defined in eq. (5.4)

5.8 WIMP Search

After fitting to the 6–20 keV$_{ee}$ range, we can then use the resultant fit activities to build a background model for our WIMP search by combining the templates together using the fit parameters C_l. Summing the component templates together yields a single template in E-z space that contains our entire background model, shown in fig. 5.18. To convert this into a form directly comparable to our dataset, events from our background model are sampled at random, and pasted into *blank* images taken using the CCDs (these are images with the same readout parameters as science images, but zero exposure, creating images with the exact same noise and baseline behaviors as science images, but with no events). To avoid double-counting noise (which is contained in the blanks), a special set of "raw" 1x100 simulations with no noise was generated and used for this process. 250 events are pasted to each of the 83 blanks for each CCD, and repeated 10 times. These pasted events are then reconstructed using the exact same processing code as the real science dataset, to replicate all the effects included in this processing code. This includes the log-likelihood cluster detection, mask application, and any other cuts used in the real dataset. This provides a background model explicitly directly comparable with our real data. We then used this background model to build a profile likelihood ratio comparing background+WIMP to background alone. The results can be seen in energy-σ_x space in fig. 5.19, and in energy space alone in fig. 5.20. The result shows a clear agreement between background and data between 0.2–6 keV$_{ee}$, as well as some unexplained excess

of bulk events $< 200\text{eV}_{\text{ee}}$. In total, 17 ± 7.6 excess events are seen. These events are well-fit with an exponential of the form

$$f_s(E|\epsilon) = \frac{1}{\epsilon}\exp(-E/\epsilon). \tag{5.13}$$

with a characteristic energy $\epsilon = 67 \pm 37$ eV_{ee}.

This unknown signal, while of limited statistics, motivates further work to investigate its origin, either as an unaccounted-for background or as a potential dark matter signal. DAMIC-M will use new low-threshold skipper CCDs, combined with advancements to the detector design and fabrication to lower known backgrounds, to further investigate this previously unobserved phenomenon.

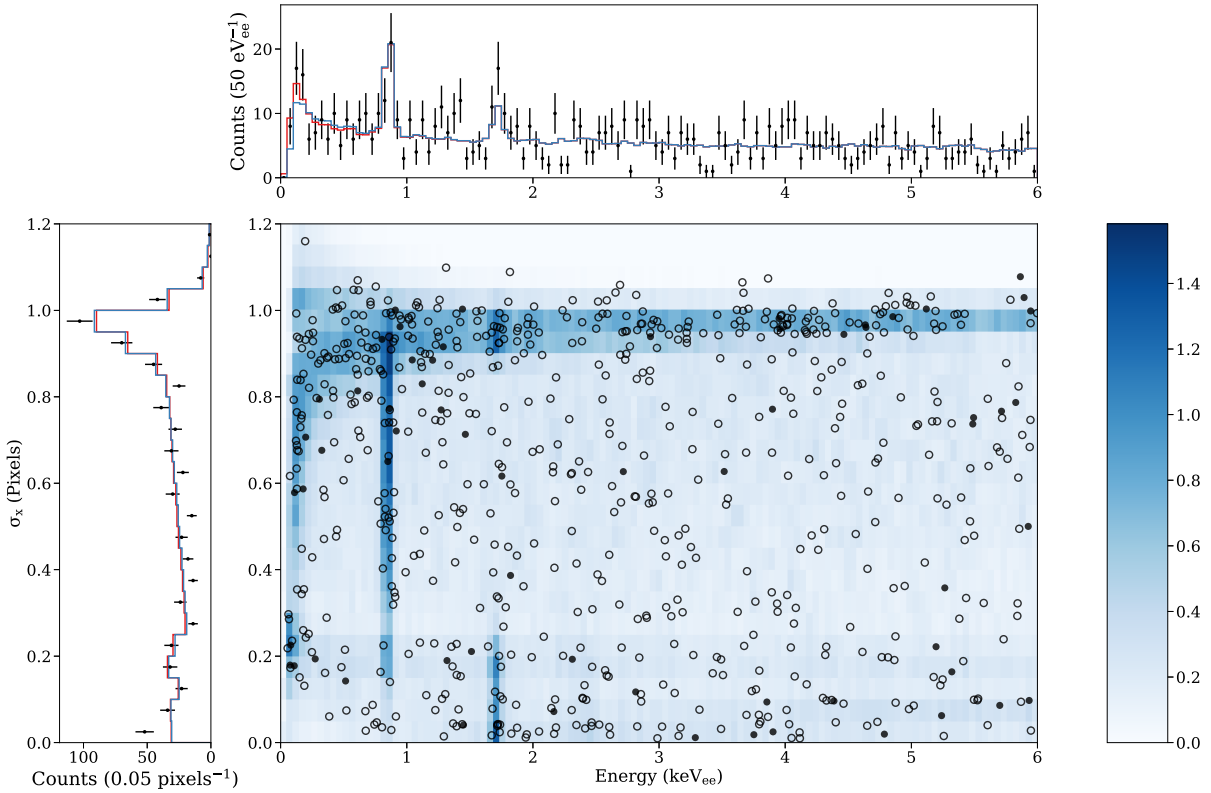


Figure 5.19: Plot of the likelihood fit results for our background vs signal models[14]. Open (filled) circles show events from CCDs 2-7 (1), overlayed on our background likelihood model. The colorbar is in units of expected counts per bin for the entire DAMIC at SNOLAB science dataset. Top and right side plots show the projection on the energy and σ_x axes (respectively), with the red line showing the best-fit signal+background model, and the blue the background only. The excess of events below 200 eV_{ee} is clear in the top projection, and contains no obvious σ_x dependency, indicating the excess is from bulk events. The likelihood fit masks the region between 1.6 and 1.8 keV_{ee} to avoid the silicon florescence line not well modeled by `GEANT4` (this mask is not shown to avoid cluttering the plot).

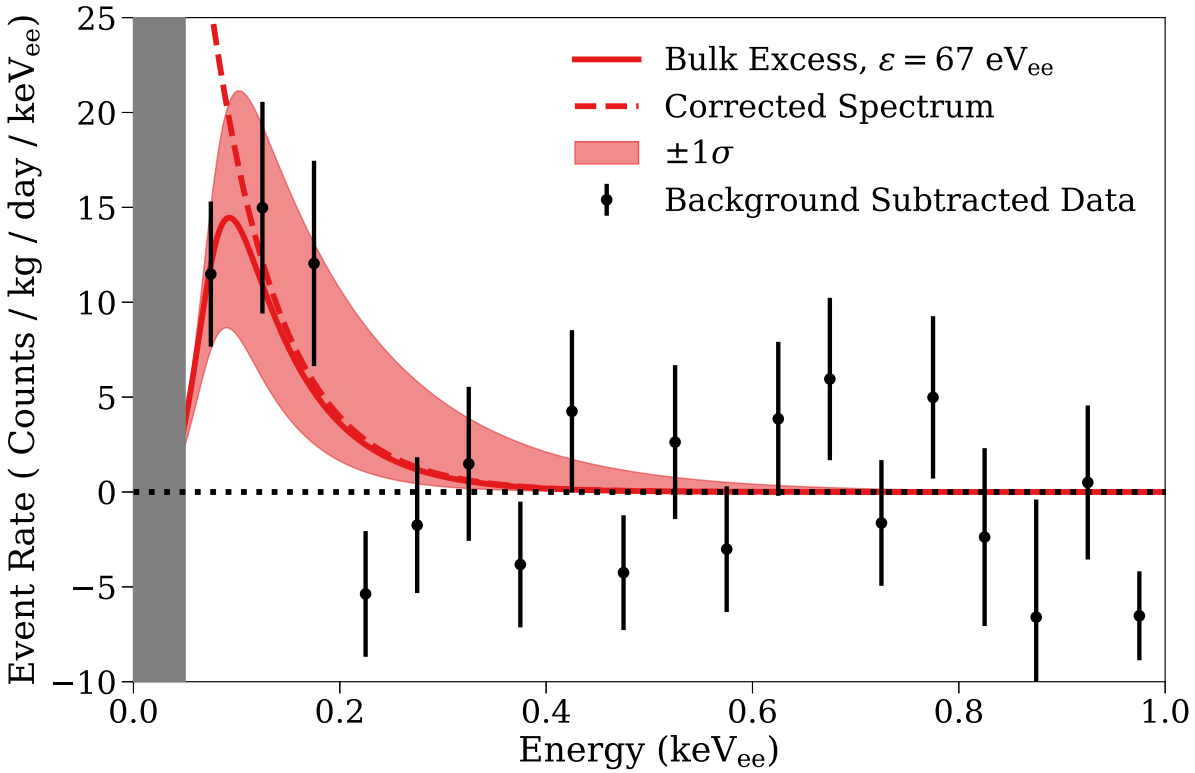


Figure 5.20: Plot of the best-fit bulk excess signal and $\pm 1\sigma$ region, along with the background-subtracted points themselves. The corrected spectrum (dashed-red line) shows the best-fit result after correcting for the energy-dependent efficiency of event reconstruction (which drops to zero below ~50 eV$_{ee}$).

References

[1] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003. ISSN 0168-9002. doi: https://doi.org/10.1016/S0168-9002(03)01368-8. URL https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[2] Andrea Pocar. *Low Background Techniques and Experimental Challenges for Borexino and its Nylon Vessels*. PhD thesis, Princeton University, November 2003.

[3] Joao Miguel Da Rocha. *Recherche de matière sombre légère avec des CCDs de l'expérience DAMIC à Snolab*. PhD thesis, LPNHE, 2019. URL http://www.theses.fr/2019SORUS517. Thèse de doctorat dirigée par Letessier-Selvon, Antoine Physique de l'Univers Sorbonne université 2019.

[4] A. Aguilar-Arevalo, D. Amidei, D. Baxter, G. Cancelo, B. A. Cervantes Vergara, A. E. Chavarria, J. C. D'Olivo, J. Estrada, F. Favela-Perez, R. Gaïor, et al. The radioactive background model for damic at snolab. Manuscript in preparation.

[5] A. Aguilar-Arevalo, D. Amidei, D. Baxter, G. Cancelo, B.A. Cervantes Vergara, A.E. Chavarria, E. Darragh-Ford, J.C. D'Olivo, J. Estrada, F. Favela-Perez, et al. Measurement of the bulk radioactive contamination of detector-grade silicon with DAMIC at SNOLAB. *Journal of Instrumentation*, 16(06):P06019, jun 2021. doi: 10.1088/1748-0221/16/06/p06019. URL https://doi.org/10.1088/1748-0221/16/06/p06019.

[6] D E Cullen, J H Hubbell, and L Kissel. Epdl97: the evaluated photo data library '97 version. 9 1997. doi: 10.2172/295438. URL https://www.osti.gov/biblio/295438.

[7] J. Baró, J. Sempau, J.M. Fernández-Varea, and F. Salvat. Penelope: An algorithm for monte carlo simulation of the penetration and energy loss of electrons and positrons in matter. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 100(1):31–46, 1995. ISSN 0168-583X. doi: https://doi.org/10.1016/0168-583X(95)00349-5. URL https://www.sciencedirect.com/science/article/pii/0168583X95003495.

[8] A. Valentin, M. Raine, J.-E. Sauvestre, M. Gaillardin, and P. Paillet. Geant4 physics processes for microdosimetry simulation: Very low energy electromagnetic models for electrons in silicon. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 288:66–73, 2012. ISSN 0168-583X. doi: https://doi.org/10.1016/j.nimb.2012.07.028. URL https://www.sciencedirect.com/science/article/pii/S0168583X12004314.

[9] K. Ramanathan, A. Kavner, A. E. Chavarria, P. Privitera, D. Amidei, T.-L. Chou, A. Matalon, R. Thomas, J. Estrada, J. Tiffenberg, and J. Molina. Measurement of low energy ionization signals from compton scattering in a charge-coupled device dark matter detector. *Phys. Rev. D*, 96:042002, Aug 2017. doi: 10.1103/PhysRevD.96.042002. URL https://link.aps.org/doi/10.1103/PhysRevD.96.042002.

[10] M. Laubenstein and G. Heusser. Cosmogenic radionuclides in metals as indicator for sea level exposure history. *Applied Radiation and Isotopes*, 67(5):750–754, 2009. ISSN 0969-8043. doi: https://doi.org/10.1016/j.apradiso.2009.01.029. URL https://www.sciencedirect.com/science/article/pii/S096980430900030X. 5th International Conference on Radionuclide Metrology - Low-Level Radioactivity Measurement Techniques ICRM-LLRMT'08.

[11] R. Saldanha, R. Thomas, R. H. M. Tsang, A. E. Chavarria, R. Bunker, J. L. Burnett, S. R. Elliott, A. Matalon, P. Mitra, A. Piers, et al. Cosmogenic activation of silicon. *Phys. Rev. D*, 102:102006, Nov 2020. doi: 10.1103/PhysRevD.102.102006. URL https://link.aps.org/doi/10.1103/PhysRevD.102.102006.

[12] A. E. Chavarria, J. I. Collar, J. R. Peña, P. Privitera, A. E. Robinson, B. Scholz, C. Sengul, J. Zhou, J. Estrada, F. Izraelevitch, et al. Measurement of the ionization produced by sub-kev silicon nuclear recoils in a ccd dark matter detector. *Phys. Rev. D*, 94:082007, Oct 2016. doi: 10.1103/PhysRevD.94.082007. URL https://link.aps.org/doi/10.1103/PhysRevD.94.082007.

[13] M.-M. Bé, V. Chisté, C. Dulieu, M.A. Kellett, X. Mougeot, A. Arinc, V.P. Chechev, N.K. Kuzmenko, T. Kibédi, A. Luca, and A.L. Nichols. *Table of Radionuclides*, volume 8 of *Monographie BIPM-5*. Bureau International des Poids et Mesures, Pavillon de Breteuil, F-92310 Sèvres, France, 2016. ISBN 978-92-822-2264-5. URL http://www.bipm.org/utils/common/pdf/monographieRI/Monographie_BIPM-5_Tables_Vol8.pdf.

[14] A. Aguilar-Arevalo, D. Amidei, D. Baxter, G. Cancelo, B. A. Cervantes Vergara, A. E. Chavarria, J. C. D'Olivo, J. Estrada, F. Favela-Perez, R. Gaïor, et al. Results on low-mass weakly interacting massive particles from an 11 kg d target exposure of damic at snolab. *Phys. Rev. Lett.*, 125:241803, Dec 2020. doi: 10.1103/PhysRevLett.125.241803. URL https://link.aps.org/doi/10.1103/PhysRevLett.125.241803.

CHAPTER 6

TRITIUM ACTIVATION RATE MEASUREMENT

6.1 Introduction

One of the major backgrounds in all silicon-based dark matter experiments is tritium. When the silicon ingots are initially produced, the silicon's temperature is raised until it reaches a molten state (> 1400 ℃). This process is necessary to produce the extremely pure silicon required for semiconductor fabrication, and has the added benefit of diffusing out any mobile contaminants in the silicon (such as tritium). The result is that, at formation, silicon ingots intrinsically contain negligible tritium contamination . However, once produced, the silicon ingot never again reaches such high temperatures (although temperatures up to ~1000 ℃ may be reached by the silicon wafers during the CCD fabrication process for brief periods of time). This means that any radioactive contaminants produced or added to the silicon will stay there until they decay.

While the fabrication processes do not themselves typically add any such contaminants, cosmic rays striking the upper atmosphere can produce neutrons that can penetrate through the atmosphere to reach sea level. These neutrons can impact the silicon and induce neutron spallation, breaking the silicon nucleus into components [1, 2]. Most of these spallation-induced components have a relatively short half life for most dark matter experiments (such as ^{7}Be with its 53 day half life), or produce high energy decay products that can be ignored (or are invisible) to dark matter experiments (such as ^{22}Na with β and γ decays $\gg$ 100keV). However, they can also produce ^{3}H, which has not only a relatively long half-life of 12.3 years, but also produces a single low-energy beta decay with a Q value of 18.3 keV, which lies exactly in the region of interest for silicon DM experiments.

There is some interest in the possibility of raising the temperature of the final detector to allow the tritium to diffuse out (since ^{3}H is expected to be relatively mobile in silicon at temperatures in the 100s of ℃), but such procedures are currently only theoretical. For the time being, tritium mitigation requires minimizing the exposure of the detector

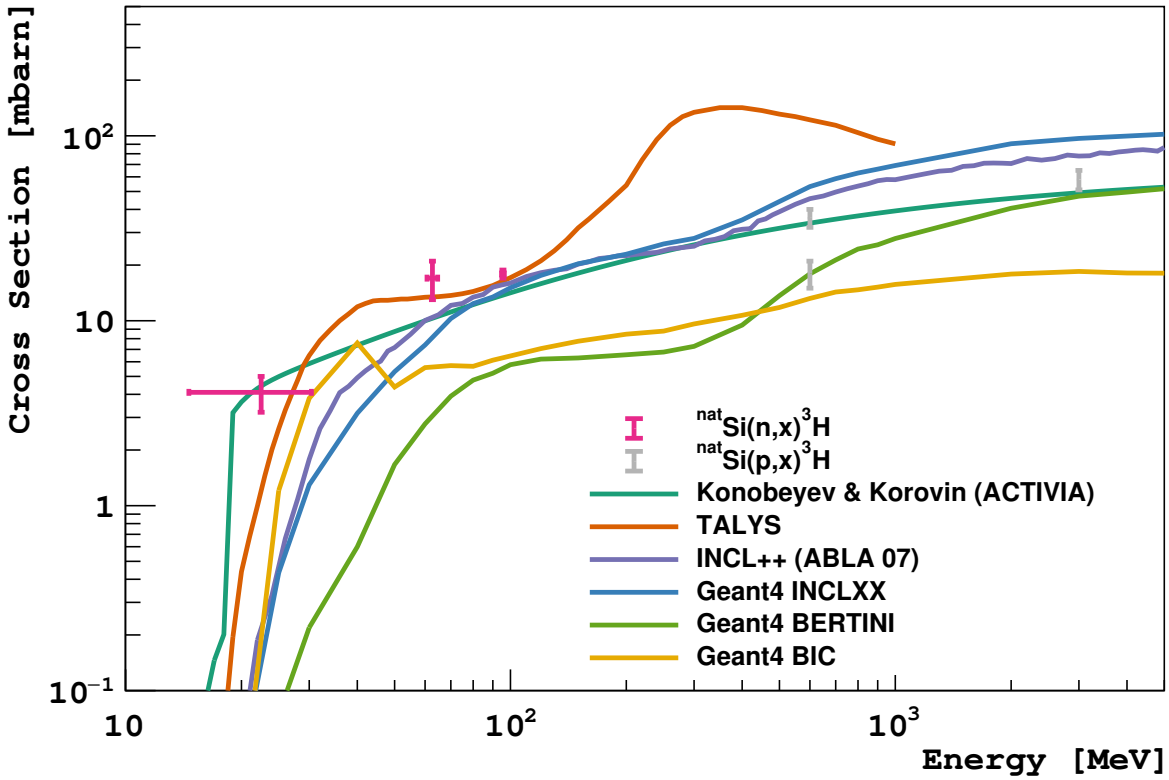


Figure 6.1: Various simulation and measurements of ^{3}H production in silicon. The lines correspond to various simulations using different physics models, while the error bars show measurements taken at specific neutron or proton energies. Figure from [3]. Shown are the K&K model [4, 5], TALYS[6], INCL[7], and ABLA[8] models, as well as the `GEANT4` simulations using INCLXX [7, 9], BERTINI [10, 11, 12, 13], and Binary Cascades (BIC) [14]

to cosmic ray neutrons. Practically, this means keeping the detector underground when possible, and minimizing the time on the surface. When surface travel is necessary (such as during transport and fabrication), devices are typically stored in heavy shielding to shield them from neutrons (such as the container constructed for DAMIC-M shown in fig. 6.2), and never transported by air, where neutron rates are several hundred times higher than on the surface.

The strictness of these handling procedures depends on the background requirement for the experiment, and (crucially) on the rate of tritium production in silicon by cosmic ray neutrons. Unfortunately for this analysis, this production rate is far too low to measure directly by any current experiment. It is possible future DM experiments may be able to measure the ^{3}H contamination with enough accuracy to produce a good direct measurement, but until then the only ways to estimate this rate are either by simulations

(which produce wildly differing results, as shown in fig. 6.1 or by raising the production rate, such as by exposure to a high-fluence neutron beam.

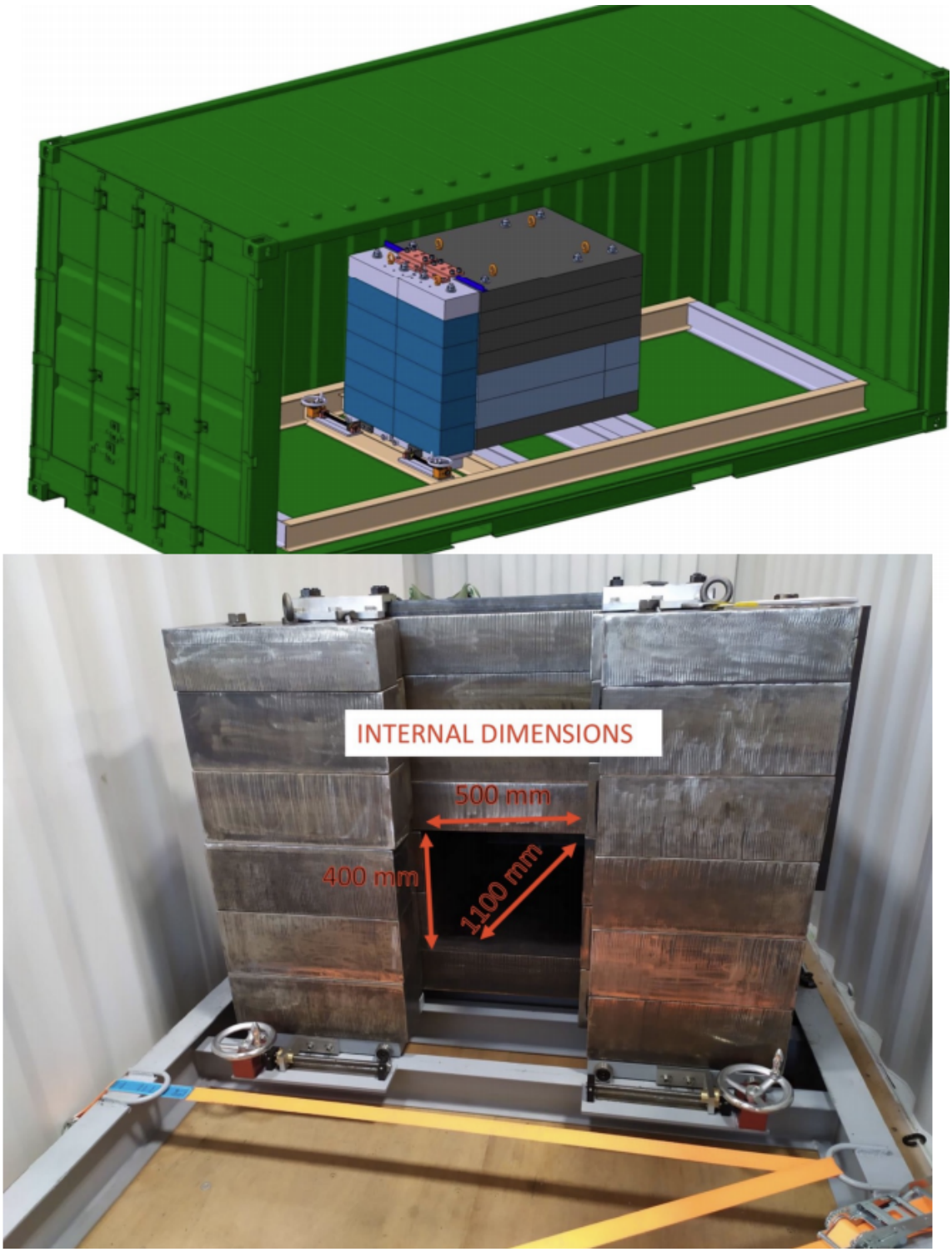


Figure 6.2: Schematic and picture of shipping container used in the production of DAMIC-M CCDs, with 15 tonnes of iron shielding. These containers are required during shipment and handling of CCDs or other sensitive detector components to reduce the exposure to cosmic ray neutrons. The requirements and design of these containers is highly dependent on the cosmogenic activation rate of the detector materials.

CCD	Beamline Exposure Time [hours]	Neutron Flux [10^{12} neutrons]
1	109.4	2.39 ± 0.18
2	62.7	1.42 ± 0.11
3	22.8	0.52 ± 0.04

Table 6.1: Neutron beam exposure time and total neutrons through target for each CCD.

6.2 Procedure

In September of 2018, we sent 3 CCDs (labeled 1,2, and 3) to the Los Alamos Neutron Science Center ("LANSCE") facility [1, 2] for exposure to their WNR ICE- HOUSE II high intensity neutron beam. This beam was explicitly developed to mimic the cosmic ray neutron spectrum, but with $\sim 10^6$ higher flux, allowing us to produce a quantity of tritium that can be measured even in a standard surface lab.

At the LANSCE facility, the three CCDs were exposed to the neutron beam for staggered time periods (see table 6.1). The CCDs were stored in specially designed aluminum boxes with thin front and back lids in order to minimize the shielding and scattering of the beamline. Several silicon, germanium, and copper components were also included at various stages for activation tests using a germanium counter (these measurements are not included in this analysis). A schematic of this setup in the three stages of adding CCDs can be found in fig. 6.3, and a picture of the final setup can found in fig. 6.4.

6.3 Measurement

6.3.1 Setup

All CCD data was taken at the University of Chicago. A picture of the setup can be seen in fig. 6.5. The CCD was placed inside a copper box, which acts both as mechanical support and as a thermal shield against infrared radiation. The copper box was cooled to a temperature of ~150 K through the cold finger attached at the top. This was driven by a Brooks Polycold compressor, connected to the cold finger through electrically isolating gas flanges that ensure the CCD is electrically isolated from the compressor. Vacuum

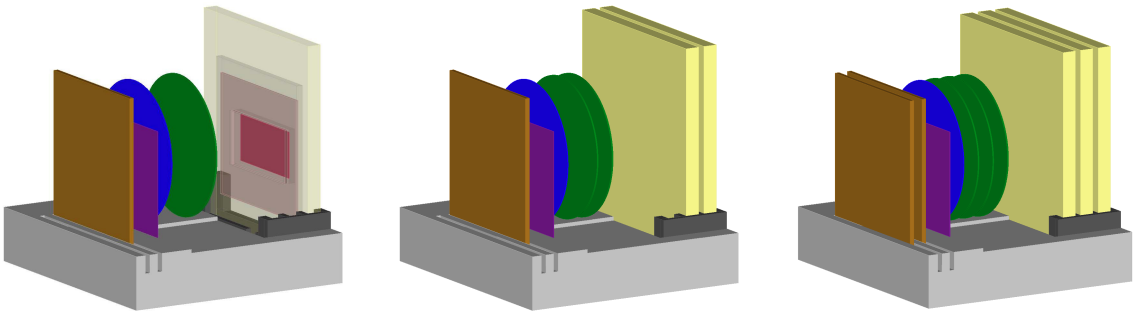


Figure 6.3: Schematic of the beamline setup as the three CCDs are added. The CCDs themselves are inside the yellow boxes. On the left is the initial setup with CCD #1, with a cross sectional view of the CCD itself (in red) inside the aluminum box. The middle shows the second setup including CCD #2, and the right shows the final setup with all 3 CCDs. The green discs are silicon wafers, the blue disc is a germanium wafer, the purple square is the wafer holder, and the brown copper shows the copper plates. The dark gray tray holding the CCD boxes is aluminum, while the light gray block holding the setup is acetal. The beam comes from the right to the left, so that it hits the CCDs first.

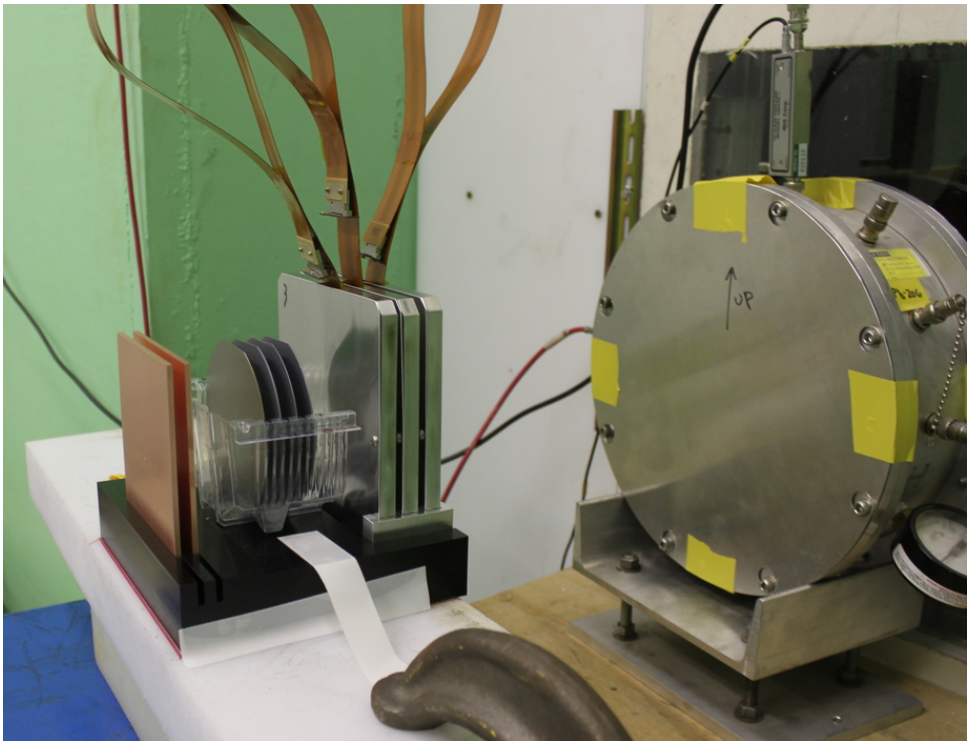


Figure 6.4: Picture of the final setup with all components in the beamline. The fission chamber that measures the beam intensity can be seen on the right hand side. The beam passes through this chamber before passing through the CCDs, which can be seen inside their aluminum boxes. The flex cables used to readout the CCDs can be seen coming from the top of the boxes.

inside the chamber is maintained using a HiCube 80 Eco turbopump, attached through the bottom flange. Data was taken using an Astronomic Research Camera's ("ARC") system composed of an ARC-32 clock driver to provide AC clocks, an LBNL designed

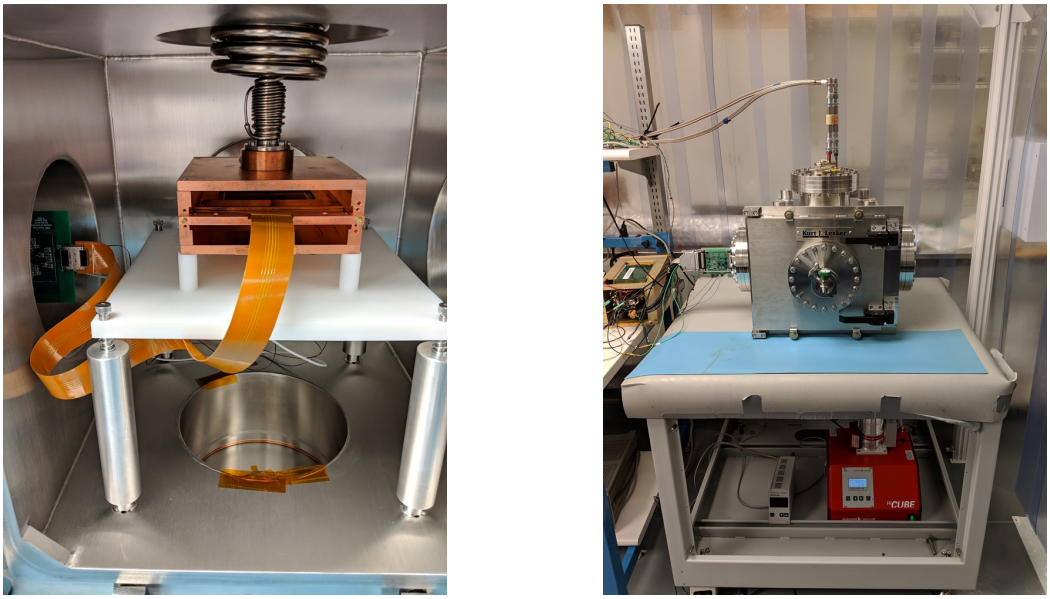


Figure 6.5: Left: picture of CCD inside copper box, inside vacuum chamber. The flex cable that is wire bonded to the CCD goes out through an interface board to a 50-pin D-sub feed-through connector on the side of the vacuum chamber. **Right:** external view of the same vacuum chamber. A custom designed external interface board with op-amp buffers and clock-filtering interfaces between the CCD and the ARC readout system ("Leach") that drives the CCD is visible attached to the side of the vacuum chamber itself.

bias board to provide DC voltages, and an ARC-45 video readout board to convert the analog CCD video signal into digital 16-bit values.

6.3.2 Background Data

Prior to shipment to the LANSCE beamline, data from all 3 CCDs was taken at the University of Chicago, to produce a background dataset. No statistically significant difference between the background in the CCDs was observed, as can be seen in table 6.2 and fig. 6.6. This is expected, as the dominant source of backgrounds for CCDs is external to the CCD. Images were taken using an exposure time of 500 seconds, and a readout time of 412 seconds, for a total per image exposure of 912 seconds.

6.3.3 Activated Data

After the exposure at LANSCE, the CCDs were left to let the short lived activated isotopes decay away for several months before being shipped back to the University of Chicago. There they were removed from the aluminum shipping boxes, and images were

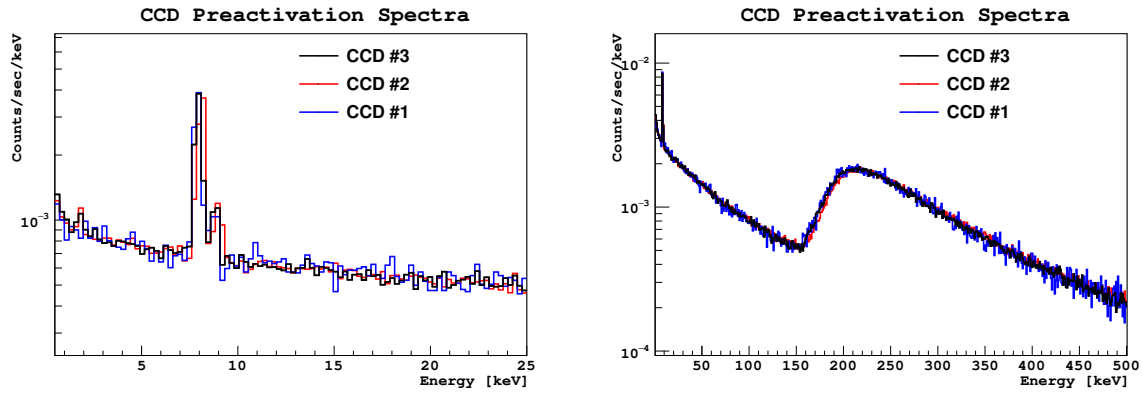


Figure 6.6: Plot of spectra for all 3 CCDs prior to activation in the neutron beam. Left shows the region from 0.5-25 keV, right shows from 0.5-500 keV. The peaks visible around 8 keV are from copper.

CCD	Background Data [days]	Background Rate (1-18 keV) [10^{-3} events/sec/kev]
#1	2.13	3.185 ± 0.032
#2	8.95	3.160 ± 0.016
#3	13.06	3.174 ± 0.013

Table 6.2: Measurement of backgrounds in the 3 CCDs before they were sent to LANSCE for activation. The most background data was taken for CCD #3 as that was intended to have the lowest activation rate, and would require the best background measurement to resolve the ^{3}H signal.

taken using a setup identical to the background data set. The exposure time for these images was lowered to 5 seconds to reduce charge pileup and dark current accumulation, but the readout parameters were otherwise identical. The results for the three CCDs can be seen in fig. 6.7. The damage to the CCDs from the neutron beam (which can be seen in the bright band in fig. 6.8) required some modifications to the processing of the data, which are detail in the next section.

Some additional data was taken later using CCD #3, after some significant modifications to the CCD chamber. These modifications caused a significant lowering of the CCD temperature, resulting in much lower dark current. This data set is shown in appendix B, since it is not directly comparable to the background data.

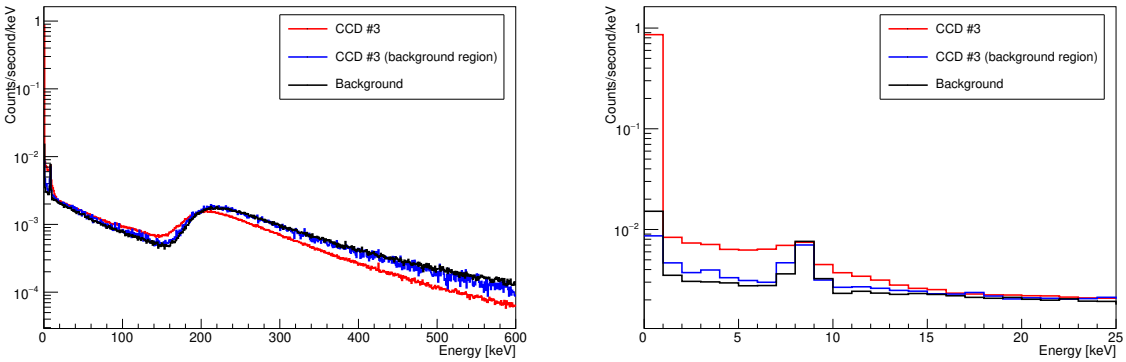
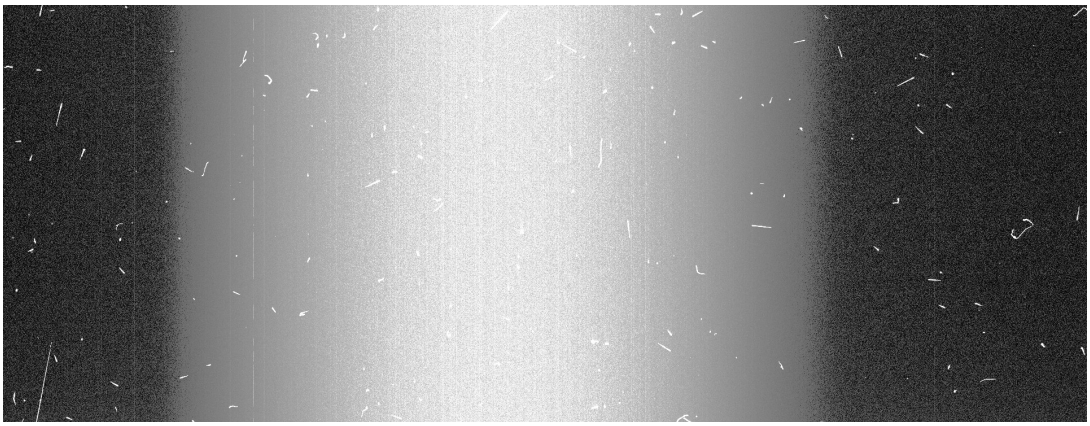


Figure 6.7: The spectrum of CCD #3 measured after activation (red) and before (black). The blue line shows the spectrum outside the activated region (in the first 400 columns). The left image shows the spectrum from 0-600 keV, while the right shows a zoom of the region from 0-25 keV. The rise in event rate below 18 keV is a clearly visible sign of ^{3}H activation in the CCD.

(a) Image taken before activation



(b) Image taken after activation

Figure 6.8: Example images taken using CCDs at University of Chicago. Top shows a background image taken before sending the CCDs to LANSCE for activation, bottom shows an image taken after activation. The large vertical column in the center happens due to the near-continuous readout process.

6.4 Data Processing

After the images are taken, they are stored as Flexible Image Transport System (".fits") files (a standard image format among the astronomical community). These images must be processed in order to perform event identification and reconstruction. In our case, we are using the CCDs as particle detectors rather than imaging devices, so these images themselves are not directly usable. Rather, they must be processed in order to perform event identification and reconstruction (these events are conventionally called "clusters"). There exists no standard framework for such event reconstruction in CCD images, but fortunately such software has been developed for the DAMIC at SNOLAB project , and can be used for our purposes with only slight modifications.

Processing CCD images requires several stages. These stages are image baseline fitting and subtraction, mask generation, and cluster identification.

6.4.1 Baseline fitting

In our CCD acquisition system, pixel values are acquired through the use of analog integration circuits. These circuits do not output a zero value for pixels that contain no charge, since the input voltage from the CCD is not typically centered around the zero reference for the ADC. In fact, such centering is not desirable, as it would result in saturating the lower end of the ADC for many pixels, resulting in a loss of information, and a DC offset voltage may be deliberately introduced during the integration to prevent this condition from occurring. Instead, zero pixels values are centered around a baseline value, and fluctuate with noise around this value. Processing the image to find real events requires identifying this value (and typically subtracting it off to provide a zero-baseline image for ease of charge identification later on).

In principle, the baseline should be constant across the entire image, in which case the baseline can be identified simply as the median pixel value in the image. If an image contains a large amount of charge, the x-overscan may be used instead for the same purposes. However, in practice our readout electronics do not always produce a

consistent baseline across the entire image. Large RC constants or simple voltage drift can result in changes to the baseline over time. The ARC readout system in particular often has a noticeably higher baseline near the bottom of the image (see fig. 6.8a). The exact cause of this has never been determined, but is driven by some large DC offset near the start of the image which decays over a period of a few seconds.

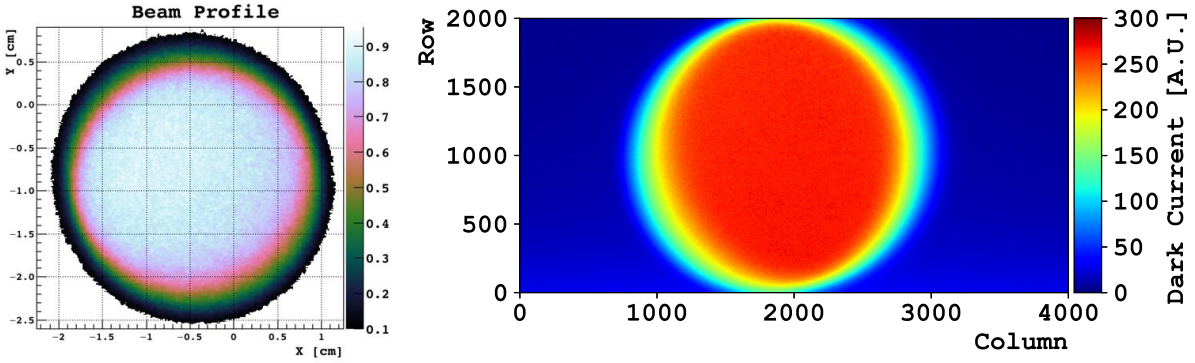


Figure 6.9: Left: LANSCE neutron beam profile. **Right:** Dark current profile for CCD #3. This profile is generated by taking two sets of images, one set with nearly zero exposure time (so that the only signal is generated during readout), and the second with a very long exposure time and the same readout time as the first. Both sets of images are then averaged (to remove non-dark-current charge), and then the first set of images is subtracted from the second to remove the dark current generated during readout, which would otherwise cause a smearing of charge that makes the dark current profile less visible. Finally a Gaussian smoothing filter is applied to the image to clean it.

Additionally, the CCDs that have been exposed to the neutron beam show extensive damage in the activated region. This damage has several effects , one of which is a significant amount of dark current ($\gg 1$ elec/pixel/image). This high level of dark current can be best be seen in fig. 6.9. When taking data, this dark current manifests as a large band of higher pixel values in the irradiated region (as seen in fig. 6.8b). This effectively creates a region of increased baseline in the CCD, since the pixel values in this region equal the standard baseline plus the dark current per pixel. The dark current varies slightly over time (especially at the start of a data run), but was surprisingly similar between all 3 CCDs, as can be seen in fig. 6.10. While these dark current events represent real charge, they are not the ionization events we are looking for, and can be safely subtracted.

To handle both these effects, the standard DAMIC baseline subtraction algorithm performs baseline subtraction on both a row-by-row and a column-by-column basis. Moreover, the fitting can split a column or row into multiple segments, in case the baseline is

not stable over an entire row.

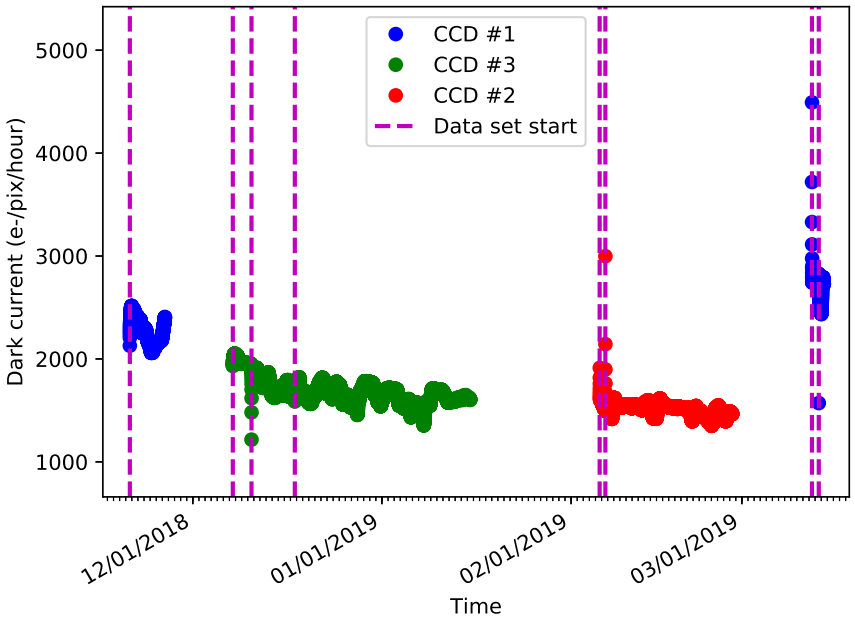


Figure 6.10: Dark current over time for all 3 CCDs. Dark current is estimated from the raw images by subtracting the median value of pixels with x>3500 (an estimate of the baseline well outside the activated region) from the median value from pixels in the range 1668<x<2120, the 452 columns near the beam center. Purple dashed lines indicate the start of a data acquisition run, which always caused the the dark current to vary considerably for several images. A second small set of data was taken for CCD #1 to check if the apparent decrease in dark current over time visible in the other CCDs also held for this CCD, which it did not (the dark current for this data set was slightly higher than for the earlier data set). Median dark current values for the 3 CCDs are: 2344, 1526, and 1662 e^-/pixel/hour for CCDs 1,2, and 3 (respectively).

6.4.2 Mask Generation

The next step in image reconstruction is mask generation. All CCDs inherently contain a certain number of defects in their silicon lattice. These defects can be caused by malformations of the silicon during ingot formation, damage during the fabrication process, or damage from radiation before or after production. In practice, the effect of these defects is to cause certain pixels to produce significant amounts of charge, or to trap charge for release later (as discussed in the next section). Pixels that produce charge are usually termed "hot", and can produce anywhere from several electrons per hour (if the pixel is merely "warm") to many thousands of electrons per second, depending on the severity

of the defect. The latter defect shows as not merely a single pixel but an entire defective column, as it produces enough charge to fill subsequent pixels after charge has been shifted away during during the readout process. Such a hot column can be seen in as a bright column on the left hand side of fig. 6.8a

While hot columns are easy to identify (at least visually) even from a single image, a typical hot or warm pixel can only be identified by looking at a set of images, as in a single image it will be indistinguishable from real ionization events. There are typically two metrics used to identify such pixels. The first is to find pixels that typically have values higher than normal. Since events are typically expected to be spread nearly uniformly through the CCD, pixels that consistently produce charge across an entire set of images are likely doing so because they contain a defect. The second metric is to look for pixels that are noisier than normal. This metric is less reliable, but can also identify pixels that (for whatever reason) may be consistently misbehaving.

The first metric is constructed by finding the median value of each pixel across your masking set. These median values are then histogramed, and a Gaussian fit is performed to find the typical range of median pixel values across the entire CCD. Pixels that fall significantly (> 5 sigma) far from the central value are assumed to be defective, and are added to the mask.

The second metric is similar, but looks at the RMS value of each pixel across the masking set. These values are then histogramed and fit, and again any that fall significantly far from the central value are added to the mask. Plots showing both of these distributions for the activated dataset can be seen in fig. 6.11.

In the case of our activated images, median and RMS values of pixels are not consistent across the image. Pixels inside the activated region have consistently higher values (due to higher dark current generation), and consistently higher noise (since Poissonian fluctuations in the dark current generation increases the RMS values of these pixels). We therefore split the image into 20-column segments, and treat each segment independently when performing the mask generation. This allows for a variation in median values and noise across the image, while still keeping a large (4000 pixel) sample to perform a fit.

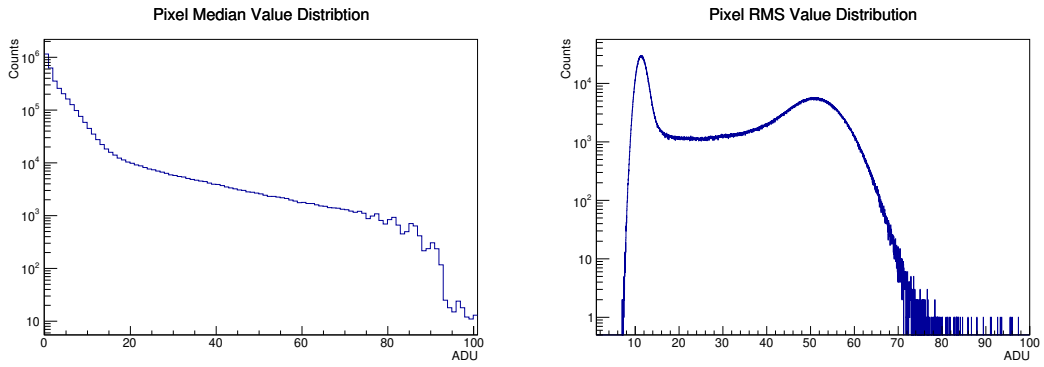


Figure 6.11: Plots of typical pixel median (left) and RMS (right) values for CCD #3. The large tail of median pixel values arise from hot pixels that consistently generate large amounts of charge across many images. Baseline subtraction eliminates dark current from contributing to the median value plot. The two-peaked plot of RMS values arises from dark current generation: since dark current follows a Poissonian process, the effective noise in pixels becomes the readout noise plus the fluctuation in dark current generation. The former is dominant in regions of the CCD outside the neutron beam (causing the peak at around 12 ADU, corresponding to our ~2 e- readout noise), while the latter is dominant in the activate region (causing a peak at round 55 ADU).

6.4.3 Clustering

Since charge produced inside the bulk of the CCD can diffuse into nearby pixels, reconstructing the energy of the original event requires clustering together all of the pixels containing charge from the event. Clustering low energy events is non-trivial since some pixels in the event may contain only a few electrons (a significant fraction of the total energy), and detecting such pixels above the noise is very difficult, requiring carefully tuned and calibrated log-likelihood searching algorithms. Fortunately, higher energy events are much easier to cluster, since almost all the charge is contained in pixels with dozens or even hundreds of electrons. This means that clustering can be performed with a simple threshold algorithm. This algorithm simply identifies pixels with charge $> 4\ \sigma$ above the noise, and then adds together all such pixels that touch into a single cluster. The total energy of the cluster is then simply the sum of all the pixel values in the cluster. To avoid the possibility of hot or warm pixels bleeding charge into real events, any event that contains or touches a masked pixel is eliminated. Since higher energy events typically span more pixels than lower energy events (as can be seen in fig. 6.12), this masking cut has an energy-dependent effect. This effect is corrected for through the Monte Carlo simulation.

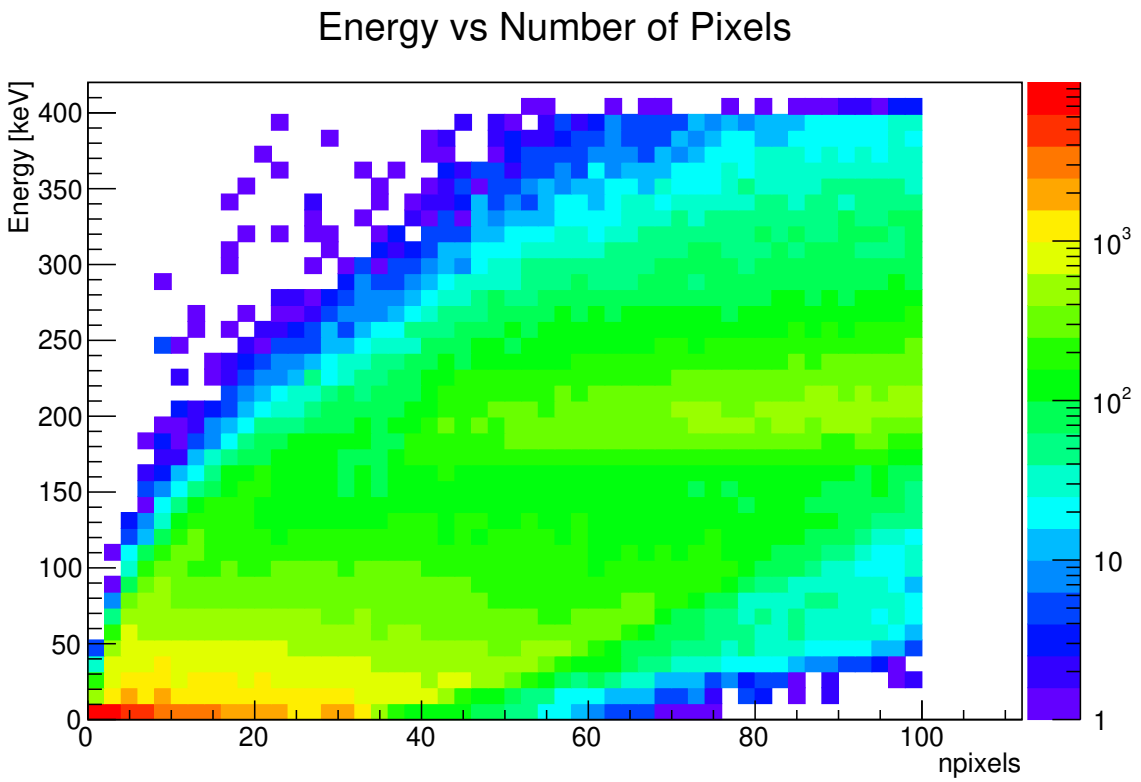


Figure 6.12: Energy of a cluster vs the number of pixels contained in that cluster. The increased number of pixels in higher energy events renders them more susceptible to effects of damage in the activated CCDs.

6.5 Charge Transfer Inefficiency

As mentioned in section 2.4, charge transfer inefficiency (CTI) is nominally caused in CCDs by a failure of the clocks to transfer all of the electrons in a charge packet. In the case of this clock-induced CTI, each transfer of charge has the same probability of causing some electron to be left behind. It is also usually assumed that the CTI probability is independent of the total charge in the packet (since in normal operating conditions, the voltage from the charge packet is $\ll$ the voltage from the clocks). In this case, CTI distribution occurring follows a binomial distribution, with the number of trials n equal to the number of charge transfers, the probability p equal to the CTI probability (which must be determined empirically from images), and the outcome k gives the number of pixels after the real event where the electron is finally read out (i.e. electrons that fall into the $k = 2$ bin end up being left behind twice during readout, and therefore appear 2 pixels behind where they would without CTI). To compute the CTI distribution for a

charge packet with N electrons in it, we can then write that

$$S(k; N, n, p) = N\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \tag{6.1}$$

where S(k;N,n,p) is the number of electrons in the kth trailing pixel behind the original pixel. For typical CCDs, p is extremely small ($\mathcal{O}(0.0001)$), and for 2k by 4k DAMIC CCDs, n ranges from 0 to 2000. It is therefore common to approximate this binomial distribution with the easier to use Poissonian distribution,

$$S(k; N, \lambda) = N\frac{\lambda^k e^{-\lambda}}{k!} \tag{6.2}$$

where our new variable $\lambda = np$.

CTI can be subdivided into two types: horizontal CTI, caused by charge being left behind in the previous column during clocking in the serial register, and vertical CTI, caused by charge being left behind in previous rows during vertical clocking. In our case, vertical CTI is predominant, while horizontal CTI is negligible.

While large amounts of CTI can be identified by eye, CTI is usually quantified by examining the spread of clusters, σ_x for horizontal and σ_y for vertical CTI. An example of both visible CTI, and the σ_y distribution for our data set (compared to several simulated CTI parameters), is shown in fig. 6.13. This spread may also be caused by diffusion (see section 2.2), but for clusters with energy > 1 keV the spread from diffusion is highly symmetric so that $\sigma_x \approx \sigma_y$. Since horizontal CTI is negligible in our CCDs even after the activation damage, the value σ_x can be still be used for depth reconstruction, while σ_y can be used as a metric of the CTI of events.

6.6 Simulation

To create our templates for fitting to find our final activation, multiple simulation steps are used. The intrinsic laboratory background is not directly simulated, and is instead derived from the pre-activation measurement (as it contains many different components, whose exact composition is not known). However, this spectrum must be modified to account

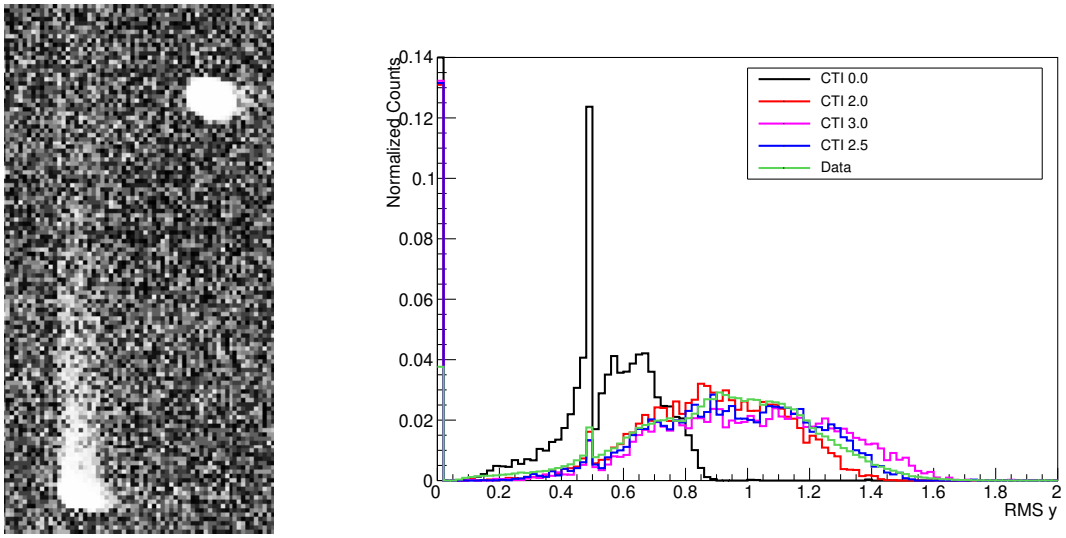


Figure 6.13: Left: Example of two events, one with large amounts of visible CTI (bottom left) and one with little or no visible CTI (top right). **Right:** Comparison of distribution of σ_y between data and simulation for various maximum CTI parameters. The agreement is not perfect for any parameter but the endpoint agrees well for $\text{CTI}_{max} = 2.5$. Events with $\sigma_y = 0$ are excluded from the analysis with the surface cuts, so the large difference in that bin is not important.

for the damage to the CCD cause by the activation process. The energy depositions from the activated components (^{22}Na and ^{3}H) can be simulated by the `GEANT4` library, but must also be passed through several processing stages to replicate the effects of charge reconstruction and CCD damage on the final spectra.

6.6.1 GEANT4

The raw Monte Carlo simulations used in the analysis are produced by the same `GEANT4` code as is used in the DAMIC at SNOLAB experiment. The simulations are produced using a simplified geometry that includes only the CCD and copper box that holds it. This simulation is processed using the same tool-set as for the DAMIC at SNOLAB experiment, to simulate the effects of diffusion and Fano factor. The dark current and CTI of our images create additional effects which must be simulated separately.

6.6.2 Blank Generation

The standard CCD simulation pipeline converts `GEANT4` energy deposition into processed pixelized clusters directly, without creating full simulated images. Cluster reconstruction

of adjacent pixels is performed, but only to simulate the effect of events that deposit energy into multiple sites (splitting events across two or more distinct clusters). This process assumes that the images and reconstruction processes are well-behaved. In the case of our activated CCDs, this assumption is no longer safe.

To model any effects caused by image reconstruction in our damaged CCDs, we create a full image simulation chain. Simulated images (called "blanks") are produced from real images by using the baseline and per-pixel noise estimators given by the reconstruction chain. A python script creates a series of blank 2100x4200 images with these same properties (using the baseline as the median pixel values, and allowing individual pixels to fluctuate up or down by the per-pixel noise estimate). This creates simulated images with the same noise, dark current, and hot pixel effects as our real data set.

Once these blanks are created, events from either `GEANT4` or our background data set are "pasted" to the blank, by adding the charge of the pixels in the simulated cluster to the corresponding pixels in our simulated blank images. The effect of CTI is modeled by applying a Poissonian distribution to the events prior to pasting to the blank.

Mathematically, this distribution is applied by convolving the cluster with a discrete Poissonian kernel defined as

$$f_{i,\lambda} = \begin{cases} \frac{\lambda^i e^{-\lambda}}{i!} & i \geq 0 \\ 0 & i < 0 \end{cases} \tag{6.3}$$

where i runs from +/- 1/2 the total size of the kernel. The negative elements are zero because CTI can only shift charge in the positive y direction. The kernel is normalized to 1 before the convolution to conserve total charge. This kernel is 1 dimensional because we only model vertical CTI.

The magnitude of this kernel λ is allowed to freely vary between zero (to simulate events that occur very near the serial register, and therefore would show no CTI) and the maximum value (to simulate events that occurred near the top of the image and have the maximum possible CTI). Nominally the CTI is proportional to the vertical position of the original cluster, but due to the near-continuous readout process, events with either

low or high CTI can be reconstructed as having occurred anywhere in the CCD.

The position of simulated activation isotope events is distributed randomly in the y-axis (due to the aforementioned readout process). In the x-axis, events are distributed according to the dark current profile seen in fig. 6.9. Our pre-activation background distribution is assumed to be unchanged by the activation procedure, and is simulated uniformly across the CCD in both x and y.

These full simulated images are then processed using exactly the same data chain and masks as our real data set. This process incorporates the full effects of mask application and dark current on the spectra generated for our fitting template. The effect of this pasting process on the ^{3}H spectrum can be seen in fig. 6.14, comparing the simulated spectrum before and after this process.

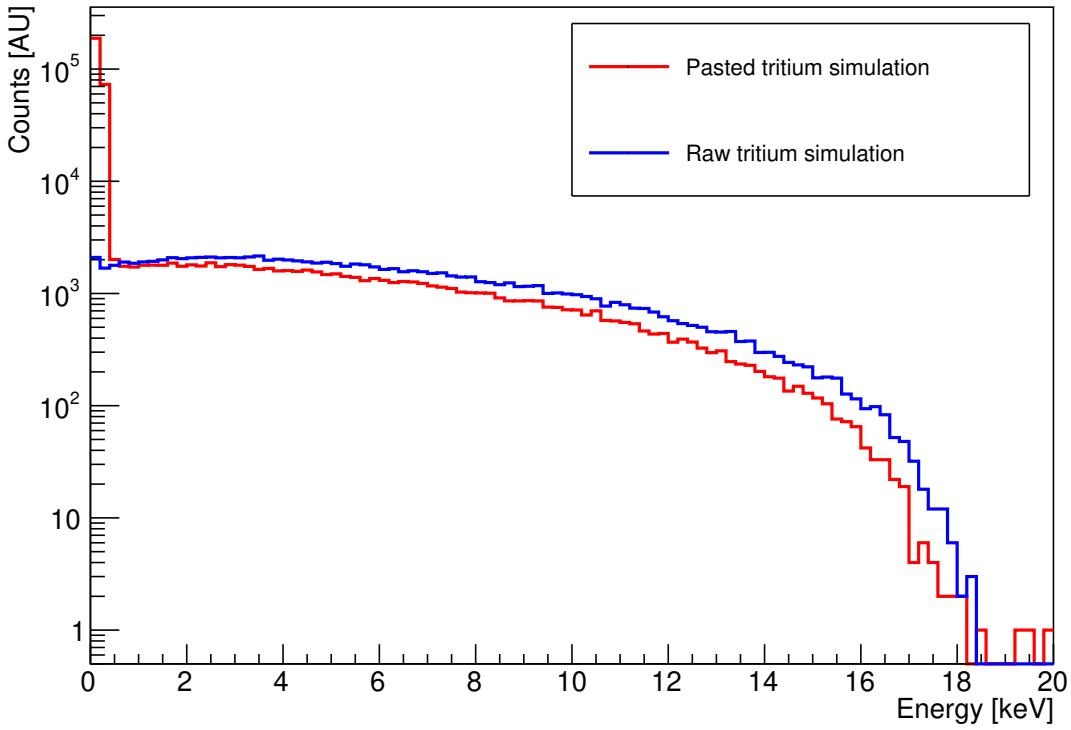


Figure 6.14: Comparison of ^{3}H simulation spectrum before and after effect of pasting to blanks, showing the significant effect on the spectrum from the modeled effects. This effect is highly prominent at low energies (<500 eV), which shows a rise similar to (but much more dramatic) than the real data. This indicates that we are modeling the nature of the effects well, but are not exactly able to reproduce their magnitude properly, which is why we do not use data at such low energies.

6.6.3 Data Selection

Several selection cuts were used in the final analysis. First, there was a substantial rise in events at low energy, below $\sim$ 500 eV , presumably driven by the high dark current and noise in the activated region (as this rise is not seen outside the activated region). We therefore limited the data analysis to only events $>$ 2 keV.

The spectrum at higher energies (>100 keV or so) also deviated from the expected spectrum. Events of these energies involve larger extended clusters that occupy many pixels (see fig. 6.12). Fortunately we are primarily interested in the ^{3}H spectrum, which peaks at ~4keV and ends at 18.6 keV, so limiting the range to 2-25 keV allows us to include all the relevant physics as well as a significant amount of non-^{3}H spectrum for background fitting, while excluding any spectrum with suspect features.

A second cut was made to exclude surface events. As mentioned in section 2.1, the RMS width of the clusters (in x, y, and x-y combined) is proportional to the depth of the interaction. In well-behaved CCDs, this allows for the identification of both front and back surface events. With the significant noise and CTI in our CCDs, the ability to identify rear surface events was considered suspect, but we can still identify and exclude front surface events as any events with RMS spread $\sigma_x = 0$ or $\sigma_y = 0$ (there was a large increase in such events compared to the background data, as shown in fig. 6.15).

A final cut was made around the region of copper fluorescence. The CCD is surrounded by copper that emits several x-ray florescence peaks between 7.8 and 9 keV. These peaks are not well simulated by the GEANT4 simulation, and would be highly shifted if there are any energy miscalibration effects, disproportionately affecting the likelihood or χ^2 value. We therefore excluded the entire energy region between 7 and 9.1 keV to eliminate any issues these lines could cause.

6.7 Fitting Procedure

To find the final ^{3}H activity in the data set, a fit of the model to the data is performed. Since our backgrounds and signal do not follow an analytical distribution, we use a binned

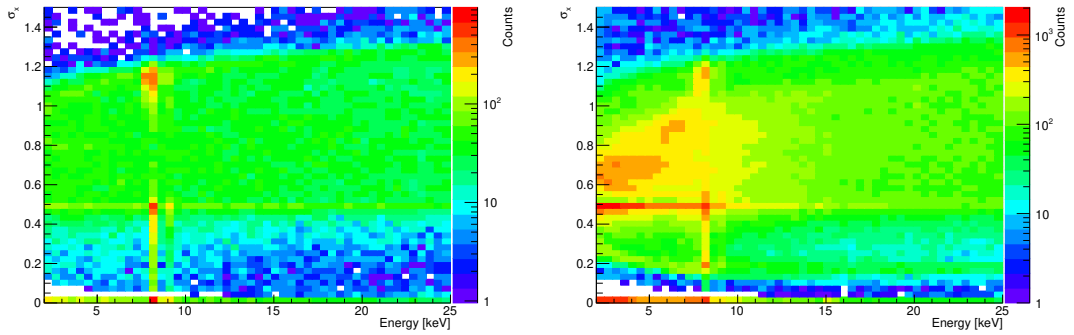


Figure 6.15: Left: σ_x vs energy for background data. **Right:** σ_x for data from CCD #3. The large excess of events near $\sigma_x = .7$ from bulk ^{3}H events is clearly visible and peaks around 2-3 keV as expected from ^{3}H, while the events with $\sigma_x = 0$ continue to rise towards lower energy, indicating they are not likely from ^{3}H decays.

likelihood fit, following the form

$$u_i = S_{trit}u_{i,trit} + S_{na}u_{i,trit} + S_{bkg}u_{i,bkg} \tag{6.4}$$

where $u_{i,n}$ is the number of events in the ith bin of the histogram for the nth component, and S_n is the fractional quantity of that component in the complete spectrum. In theory, this model allows for all components to be left free in the fit. In practice, it was found that the ^{22}Na and background components were highly degenerate (as both spectra are essentially flat in the fitting region), and so the background rate was fixed at the same rate as was measured before activation.

The actual fit itself was performed using the TMinuit [15] wrapper for the MINUIT [16] fitting library, with the results shown by component in fig. 6.16. Since our model was created from a measured background data set (whose per-bin uncertainty was comparable to the uncertainty in the measured data), a simple pure log-likelihood model would produce incorrect error estimations, since the standard binned likelihood assumes Poissonian uncertainties only on the data. Instead, 3 approaches were used to perform the fit and estimate the statistical uncertainties in the fit result.

The first method is to use a standard log likelihood fit, with our likelihood given by

$$-lnL = \sum_i u_i - k_i ln\left(\frac{u_i}{k_i}\right) - k_i \tag{6.5}$$

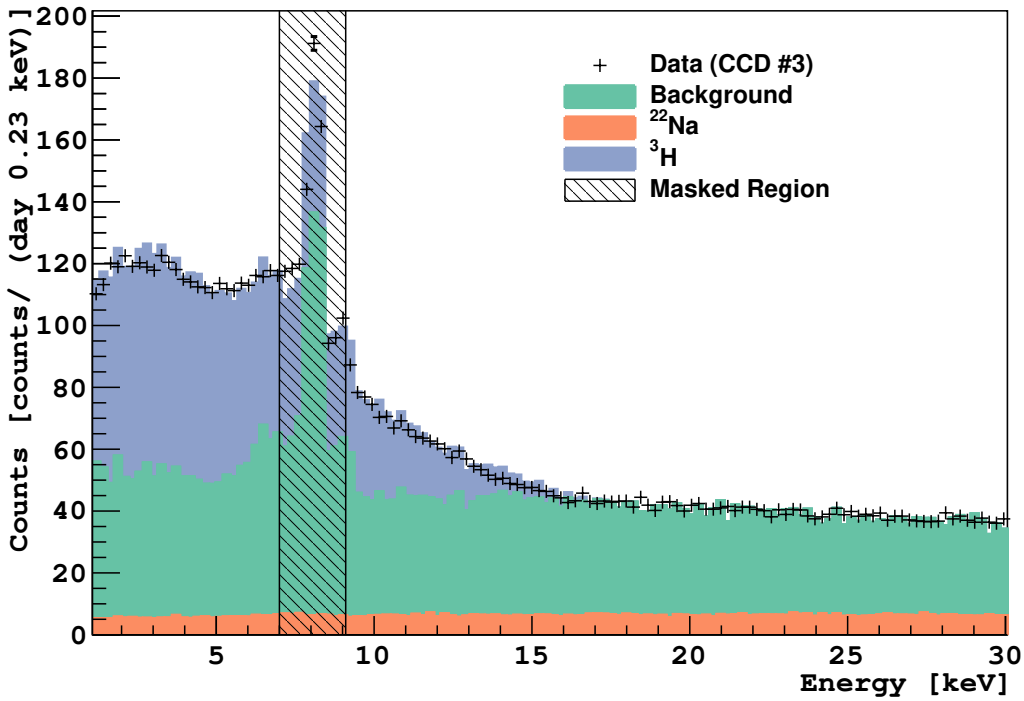


Figure 6.16: Plot of final fit results for CCD #3, broken into individual components. The shaded region shows the excluded copper peak region. The fit results are shown from 1 to 30 keV, while the fit is done from 2 to 25 keV. The ^{3}H background is clearly visible above the relatively flat spectra of the backgrounds and ^{22}Na.

where u_i is as defined above in eq. (6.4) and k_i is the number of data events in each bin. Here, Stirling's approximation has been used to eliminate the factorial in our likelihood (since all bins have >1000 events). This negative likelihood is then passed to the MINUIT minimizer to produce the best fit for our two free parameters.

Since the log likelihood fit has no free parameter for the per-bin uncertainty, this must be corrected for after the fit. To perform this correction, we then assume the error on the fit should be proportional to the increase in error from the uncertainty of the model, so that $\sigma_{true} = \sigma_{fit}\frac{\sqrt{\sigma_{data}^2+\sigma_{model}^2}}{\sigma_{data}}$. This is expected to produce a reasonable estimation of the error, but does not allow the fit to vary to account for the proper error on each bin.

The second technique is to use a χ^2 minimization to perform the fit. This allows for a variation of the fit due to the uncertainty in model bins. In this case our minimization function becomes

$$\chi^2 = \sum_i \frac{(u_i - k_i)^2}{\sigma_i^2} \tag{6.6}$$

where the uncertainty in each bin σ_i adds in quadrature, so that $\sigma_i^2 = \sigma_{data}^2 + \sigma_{model}^2$. In this case, the statistical uncertainty in the fit is expected to be produced correctly by the minimizer.

The third and final technique was to use the so-called "Barlow-Beeston" [17] method, which modifies the log-likelihood to allow each bin to fluctuate slightly, with a constraint, to account for the increased uncertainty. The full method requires adding an additional term for each of the three components in the fit, but to we can greatly simplify the fit by only introducing a single term per bin, with the uncertainty in that bin being the uncertainty in the model. Mathematically, this changes our log-likelihood to

$$-lnL = \sum_i \beta_i u_i - k_i ln\left(\frac{\beta_i u_i}{k_i}\right) - k_i + \left(\frac{\beta_i - 1}{\sigma_{model}}\right)^2 \tag{6.7}$$

where our new term β_i is a Gaussian-constrained parameter that is allowed to vary due to the uncertainty in our model σ_{model}. All three fitting techniques produce end results (both central values and uncertainties) that are consistent with each other (see table 6.3).

Uncertainty Computation Method	Fit result (No error adjustment) [mBq]	Uncertainty [mBq]	Fit Result (Error adjusted) [mBq]	Uncertainty [mBq]
After Fitting	45.2	0.3	45.2	0.6
χ^2	45.4	0.3	45.4	0.5
Barlow-Beeston	45.4	0.3	45.4	0.5

Table 6.3: Fit results and associated uncertainties for different fitting methods. Results are only shown for CCD #3, as that was the only CCD used in the final result.

6.8 Systematic Uncertanties

In our analysis we made several choices of parameters and configurations, which could reasonably have been made differently. To estimate our systematic uncertainty, we went through and varied these choices one at a time (while keeping the rest constant). Many of these parameters are distinct binary choices (such as to mask the copper fluorescence peak or not). When the parameter can take different values, values are varied within a reasonable range. A list of the parameters varied and the resulting fit is given in table 6.4,

and a scatterplot of the results is shown in fig. 6.17. The largest possible parameter change gives only a ~6% shift in the fit result, indicating the final result is robust against any changes to these fitting parameters. The final systematic uncertainty used was 1.5 mBq, or 3%. The uncertainty introduced by these systematics is small compared to the uncertainty introduced in the conversion to a cosmogenic rate, explained below.

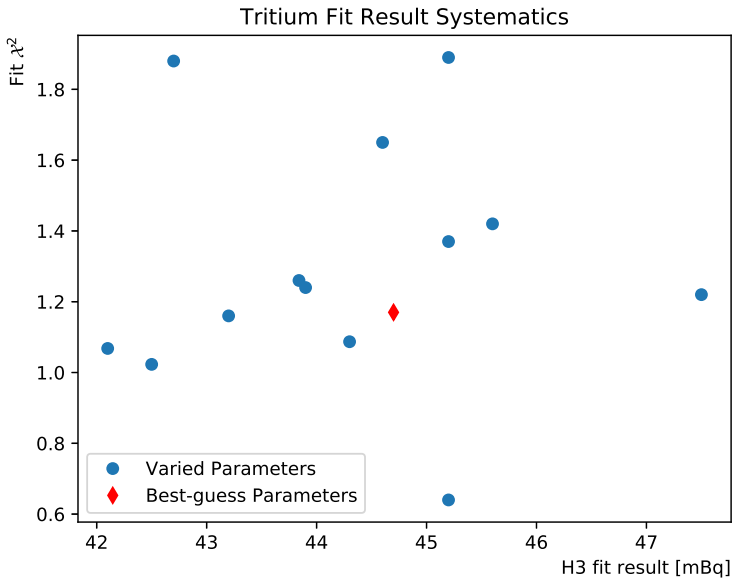


Figure 6.17: Scatter plot of tritium fit results for various different parameter choices.

We also verified that the result was reasonably reproduced with the data sets taken with the other activated CCDs. For our final measurement, we use CCD #3, which had the least neutron exposure (table 6.1). As a result, while it contained the least ^{3}H activation and therefore the lowest signal-to-noise ratio, it also had the least damage to the CCD. We applied the same methodology described in the previous sections to CCDs #1 and #2, with the only variation being the CTI parameter used in the blank pasting simulation (the blanks are also different but they are automatically generated from real data, not fixed parameters).

In both of these CCDs, presence of ^{3}H in the spectra is obvious, and its magnitude (given in table 6.5) is roughly proportional to the exposure for CCD #2. However, there are differences in the spectral shape of these two CCDs from that expected by the ^{3}H model, most notably for CCD #1 (show in fig. 6.18). We attributed these deviations to

Parameters	^{3}H Fit Result [mBq]	χ^2
Best-guess	44.7	1.17
Minimum energy 1 keV	43.9	1.24
Maximum energy 1 keV	43.2	1.16
Maximum energy 35 keV	42.5	1.02
Leave background floating	47.5	1.22
Include copper line	44.6	1.65
Smooth data spectrum	45.2	0.64
Include surface	45.6	1.42
Fix ^{22}Na rate	46.2	1.51
Decrease calibration constant 6%	42.7	1.88
Increase calibration constant 2%	45.2	1.89
RS=-0.01	45.2	1.37
RS=0.01	44.3	1.09
Max CTI 1.0	42.1	1.07
Max CTI 2.5	43.8	1.26

Table 6.4: Table of parameters varied for systematic uncertainty and resulting χ^2 value. Statistical uncertainties for all results was $\pm$0.5 mBq after the correction process described above. RS is an experimental parameter to account for any depth-dependent effects on charge reconstruction that shifts the energy of clusters events by $E_{\text{shift}} = E * (1 - \sigma_x * RS)$. Data smoothing uses ROOT's builtin "TH1::Smooth" function, and is used to check that the fit is not being driven by a single unexpectedly large bin.

the difficulty of properly parameterizing the effects of neutron induced damage in our simulations for such high neutron flux intensities. The possibility that damage to the CCDs would render the results unreliable had been anticipated prior to the measurement (and is one reason the staggered exposure with three CCDs was used in the first place). Nevertheless, the proportionality of the activation rate with the neutron flux (verified at the 5% level by CCD #2) provides confidence in the robustness of the analysis methodology.

6.9 Cosmogenic Activation Rate

The above analysis gives the activity of the CCD after the neutron beam exposure. Converting this into a cosmogenic activation rate requires two further steps. First, it was realized that the triton ions are produced in the silicon at high ($\mathcal{O}$(MeV)) energies. As a result, ions produced near the boundary of a volume can be ejected from the material it

CCD	Neutron Flux [10^{12} neutrons]	^{3}H Rate (mBq)
1	2.39 ± 0.18	138.6± 1.0
2	1.42 ± 0.11	129.3± 0.8
3	0.52 ± 0.04	45.0 ± 0.5

Table 6.5: Best-fit ^{3}H results for all 3 CCDs and total neutron flux (for comparison). All errors are statistical only, as systematic errors were not estimated for CCDs 1 and 2.

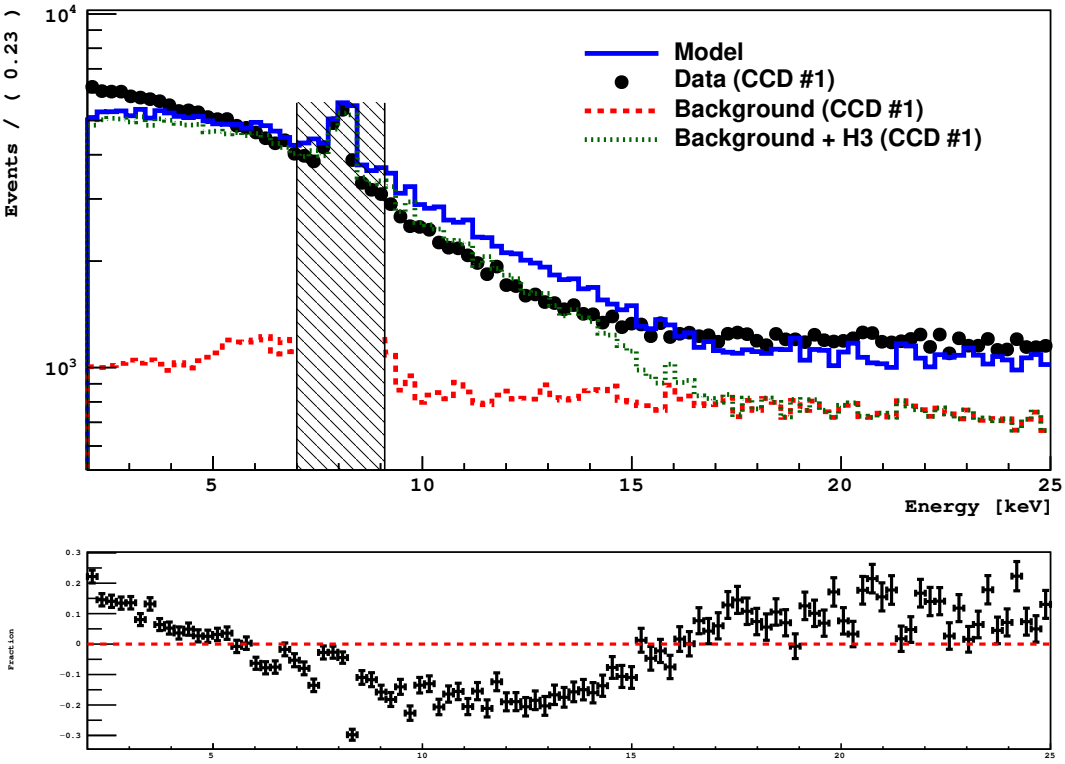


Figure 6.18: Data and best-fit model for CCD #1. The residuals can be seen at the bottom, and the various components in the model are shown with the labeled lines. The underestimation of the spectrum at below the copper peak, and overestimation beyond it until the end of the ^{3}H region can be explained if our model of the charge loss due to CTI and dark current does not fully capture charge loss from the increased damage to these CCDs.

was created in, and become implanted into another volume (generally one further along the beam, since the momentum of the tritons will be highly focused in that direction). After this, the rate measured in the material will be the creation rate plus the implantation rate minus the ejection rate. Secondly, the measured activity (after correcting for ejection/implantation) must be converted into an activation rate based on the cosmic ray neutron spectrum. Both these calculations were performed by colleagues at the Pacific Northwest National Lab (PNNL), and will only be summarized here in this work.

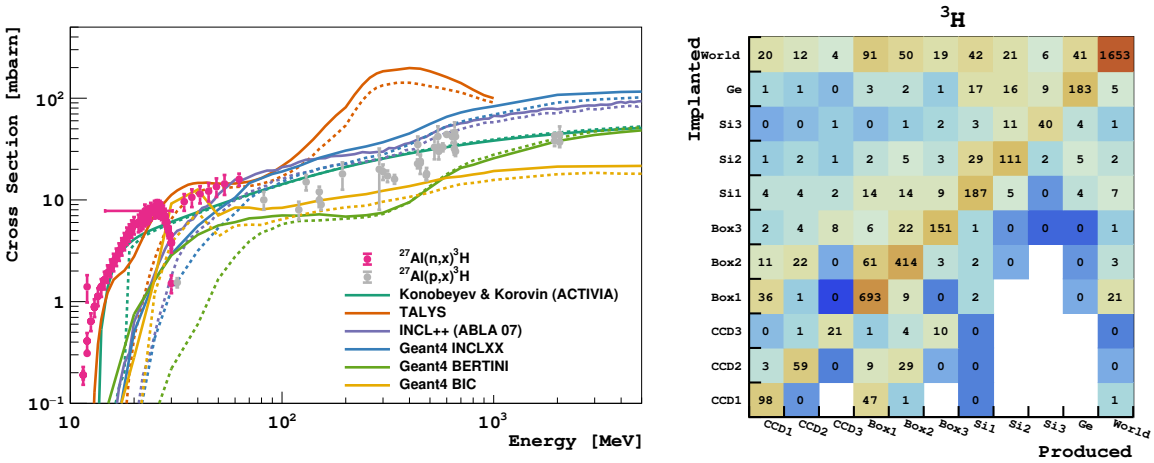


Figure 6.19: Left: Measured (data points) and simulated (solid lines) cross sections for neutron and proton-induced tritium production in aluminum. Dashed lines show the simulation results for silicon for comparison. **Right:** Computed ^{3}H transfer matrix for all volumes in the beam line.

6.9.1 Ejection and Implantation

The residual tritium rate R_i in a volume i after beam exposure can be written

$$R_i = \sum_j T_{ij} \cdot P_j \tag{6.8}$$

where P_j is the activity produced in volume j and T_{ij} is the transfer probability matrix, i.e. the probability that a triton produced in volume j will end up in i. This transfer matrix depends both on the kinematics of the produced tritons, and on the propagation of this triton through matter, both of which are computed numerically through Monte Carlo simulations. Our final goal is to compute P_j for the CCD: unfortunately, the measured rate R_i depends on the production P_j for other volumes, primarily the aluminum box. Fortunately, aluminum is fairly similar to silicon (being only one atomic number away), and has far more extensive measurements of neutron-induced production of tritium (as shown in fig. 6.19 (left)). As a result, the ejection probability of tritium produced in the silicon CCD is nearly counterbalanced by the implantation of tritium from the aluminum box. The simulated transfer matrices for all volumes are shown in fig. 6.19 (right). The ratio of tritons produced to tritons remaining is predicted to be 0.985 ± 0.078.

For nuclear models included in GEANT4 (INCLXX, BERTINI, and BIC), the transfer probability is calculated implicitly through the triton particle propagation in GEANT4. The TALYS model is not included in GEANT4 and does not compute the propagation of the particles by itself, but gives the kinematics of the produced triton, which are then given as distributions to GEANT4 to propagate the ion. The other two models (K&K (ACTIVIA) and INCL++ (ABLA07)) do not predict kinematic distributions. For those models, the average and standard deviation of the other models is used to estimate transfer probabilities. The estimate for production, ejection and implantation from the LANSCE beam in CCD #3 are shown by model in columns one, two, and three (respectively) of table 6.6.

6.9.2 Cosmogenic Neutron Activation

If the neutron beam had an energy spectrum identical to that of cosmic-ray neutrons, we could simply estimate the cosmogenic production rate by scaling the measured activity by the ratio of the cosmic-ray neutrons to that of the neutron beam. However the beam spectrum falls off faster at higher energies than that of cosmic rays (see fig. 6.20). Thus we must rely on a model for the production cross sections to extrapolate from the beam measurement to the cosmogenic production rate.

We can evaluate the accuracy of the different cross-section models by comparing the predicted ^{3}H activity produced by the LANSCE neutron beam irradiation to the decay-corrected measured residual activities. The ratio of the experimentally measured activities to the predictions of the residual activity from different models are shown in the final column of table 6.6. It can be seen that the predictions of the K&K and INCL models are in fairly good agreement with the measurement, while the TALYS model overpredicts and the GEANT4 BERTINI and BIC models underpredict the activity by more than a factor of two.

We can write the production rate for the cosmic ray spectrum P' as

$$P' = n \int \Phi(E) \cdot \sigma(E) dE \tag{6.9}$$

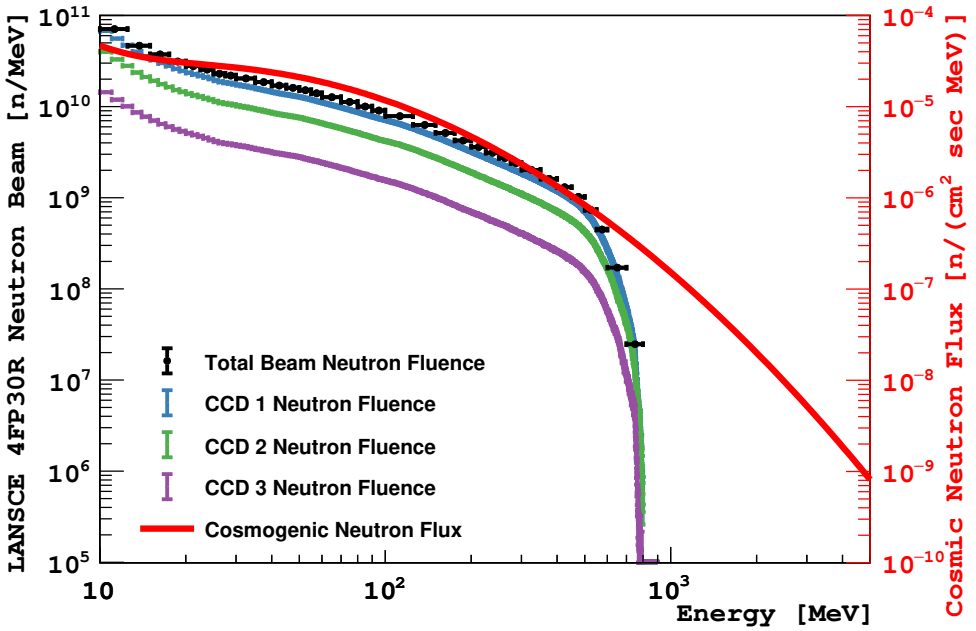


Figure 6.20: Comparison of the neutron fluence through the three CCDs as a function of the neutron energy. The scaled cosmic ray neutron spectrum from Gordon et al. [18] is shown in red for comparison.

Model	Pred. LANSCE ^{3}H prod. act. P_{CCD3} [mBq]	Ejected Activity E_{CCD3} [mBq]	Implanted Activity I_{CCD3} [mBq]	Pred. LANSCE ^{3}H res. act. R_{CCD3} [mBq]	Meas./Pred. ^{3}H res. act.
K&K (ACTIVIA)	40.8 ± 4.5			41.5 ± 5.6	1.10 ± 0.15
TALYS	116 ± 16	46.70 ± 0.12	53.8 ± 2.1	123 ± 17	0.370 ± 0.053
INCL++(ABLA07)	41.8 ± 4.8			42.5 ± 5.9	1.07 ± 0.15
GEANT4 BERTINI	13.0 ± 1.5	3.354 ± 0.072	3.699 ± 0.045	13.3 ± 1.6	3.43 ± 0.42
GEANT4 BIC	17.8 ± 1.8	4.995 ± 0.084	6.421 ± 0.059	19.2 ± 2.0	2.38 ± 0.26
GEANT4 INCLXX	42.3 ± 5.1	20.65 ± 0.11	16.94 ± 0.10	38.5 ± 4.6	1.19 ± 0.15

Table 6.6: The predicted ^{3}H rate in CCD 3 from the LANSCE beam, based on the different cross-section models used in this analysis. Shown is the predicted rate, the ejected and implanted rates, the final residual rate, and the ratio of the measured to predicted rates. Models without ejection/implantation use the average values from the other models.

where n is the number of atoms per unit mass in silicon, $\sigma(E)$ is the tritium production cross section, and $\Phi(E)$ is the energy-dependent flux of the cosmic ray neutrons. The integral is evaluated from 1 MeV to 10 GeV. While the cross section $\sigma(E)$ is not known across the entire energy range and each of the models predicts a different energy dependence, the overall normalization of each model is determined by the comparison to the measurements on the LANSCE neutron beam. The similar shapes of the LANSCE beam and the cosmic-ray neutron spectrum allow us to greatly reduce the systematic uncertainty arising from the unknown cross section.

The predicted production rates per unit target mass for the cross-section models considered are shown in the second column of table 6.7. Scaling these values by the ratio of the measured to predicted activities for the LANSCE neutron beam, we obtain our best estimates for the neutron-induced cosmogenic production rates per unit target mass, shown in the corresponding final column. The spread in the values for the different cross-section models is an indication of the systematic uncertainty in the extrapolation from the LANSCE beam measurement to the cosmic-ray neutron spectrum. If the LANSCE neutron-beam spectral shape was the same as that of the cosmic-ray neutrons, or if the cross-section models all agreed in shape, the central values in the final column of each table would be identical.

Our best estimate of the activation rate of tritium in silicon from cosmic-ray neutrons is $112 \pm 15_{\mathrm{exp}} \pm 12_{\mathrm{cs}} \pm 14_{\mathrm{nf}}$ atoms (^{3}H)/kg/day, where the first uncertainty listed is due to experimental measurement uncertainties (represented by the average uncertainty on the ratio of the measured to predicted activities from the LANSCE beam irradiation for a specific cross-section model), the second is due to the uncertainty in the energy dependence of the cross section (calculated as the standard deviation of the scaled cosmogenic production rates of the different models), and the third is due to the uncertainty in the sea-level cosmic-ray neutron flux (which we have estimated as 12.5% based on the uncertainties given in [18]).

In addition to activity induced by fast neutrons, interactions of protons, gamma-rays, and muons also contribute to the total production rate of ^{3}H, which is discussed further in [3]. These alternative production mechanisms are subdominant, and contribute an ~10% increase to the total production rate of ^{3}H, mostly due to protons. The estimated final production rate including all cosmogenic sources is 124 ± 25 atoms/kg/day. For silicon-based dark matter searches interested in the energy region 0–5 keV, this production rate corresponds to an activity of roughly 0.002 decays/keV/kg/day per day of sea-level exposure.

Model	Pred. Cosm. ^{3}H prod. rate [atoms/kg/day]	Scaled Cosm. ^{3}H prod. rate [atoms/kg/day]
K&K (ACTIVIA)	98 ± 12	108 ± 20
TALYS	259 ± 33	96 ± 18
INCL++(ABLA07)	106 ± 13	114 ± 22
G4 BERTINI	36.1 ± 4.5	124 ± 22
G4 BIC	42.8 ± 5.4	102 ± 17
G4 INCLXX	110 ± 14	130 ± 23

Table 6.7: The predicted sea-level cosmic ray neutron production of ^{3}H in silicon, for each model used in our analysis. Results are shown before (middle column) and after (last column) scaling by the measured/predicted ratio of the CCD counting measurement.

References

[1] S. Cebrian. Cosmogenic activation of materials. *International Journal of Modern Physics A*, 32:1743006, 2017. doi: 10.1142/S0217751X17430060.

[2] R. Agnese et al. Production Rate Measurement of Tritium and Other Cosmogenic Isotopes in Germanium with CDMSlite. *Astroparticle Physics*, 104:1–12, 2019. doi: 10.1016/j.astropartphys.2018.08.006.

[3] R. Saldanha, R. Thomas, R. H. M. Tsang, A. E. Chavarria, R. Bunker, J. L. Burnett, S. R. Elliott, A. Matalon, P. Mitra, A. Piers, et al. Cosmogenic activation of silicon. *Phys. Rev. D*, 102:102006, Nov 2020. doi: 10.1103/PhysRevD.102.102006. URL `https://link.aps.org/doi/10.1103/PhysRevD.102.102006`.

[4] A Yu Konobeyev and Yu A Korovin. Tritium production in materials from C to Bi irradiated with nucleons of intermediate and high energies. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 82(1):103–115, 1993.

[5] J.J. Back and Y.A. Ramachers. ACTIVIA: Calculation of isotope production cross-sections and yields. *Nuclear Instruments and Methods A*, 586:286–294, 2008.

[6] A.J. Koning, S. Hilaire, and M.C. Duijvestijn. TALYS-1.0. In O. Bersillon, F. Gunsing, E. Bauge, R. Jacqmin, and S. Leray, editors, *Proceedings of the International Conference on Nuclear Data for Science and Technology, April 22-27, 2007, Nice, France*, pages 211–214, 2008.

[7] A. Boudard, J. Cugnon, J.-C. David, S. Leray, and D. Mancusi. New potentialities of the Liege intranuclear cascade model for reactions induced by nucleons and light charged particles. *Physical Review C*, 87:014606, 2013.

[8] A. Kelic and K.-H. Schmidt. The de-excitation code ABLA07. In *Proceedings of Joint ICTP-IAEA Advanced Workshop on Model Codes for Spallation Reactions, ICTP Trieste*, pages 181–221, 2008.

[9] D. Mancusi, A. Boudard, J. Cugnon, J.-C. David, P. Kaitaniemi, and S. Leray. Extension of the Liège intranuclear-cascade model to reactions induced by light nuclei. *Physical Review C*, 90:054602, 2014.

[10] Hugo W Bertini. Low-energy intranuclear cascade calculation. *Physical Review*, 131 (4):1801, 1963.

[11] MP Guthrie, RG Alsmiller Jr, and HW Bertini. Calculation of the capture of negative pions in light elements and comparison with experiments pertaining to cancer radiotherapy. *Nuclear Instruments and Methods*, 66(1):29–36, 1968.

[12] Hugo W Bertini. Intranuclear-cascade calculation of the secondary nucleon spectra from nucleon-nucleus interactions in the energy range 340 to 2900 MeV and comparisons with experiment. *Physical Review*, 188(4):1711, 1969.

[13] Hugo W Bertini and Miriam P Guthrie. News item results from medium-energy intranuclear-cascade calculation. *Nuclear Physics A*, 169(3):670–672, 1971.

[14] G Folger, VN Ivanchenko, and JP Wellisch. The binary cascade. *The European Physical Journal A-Hadrons and Nuclei*, 21(3):407–417, 2004.

[15] Rene Brun, Fons Rademakers, Philippe Canal, Axel Naumann, Olivier Couet, Lorenzo Moneta, Vassil Vassilev, Sergey Linev, Danilo Piparo, Gerardo GANIS, et al. root-project/root: v6.18/02, August 2019. URL `https://doi.org/10.5281/zenodo.3895860`.

[16] F. James. MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1. 1994.

[17] Roger Barlow and Christine Beeston. Fitting using finite monte carlo samples. *Computer Physics Communications*, 77(2):219–228, 1993. ISSN 0010-4655. doi: https://doi.org/10.1016/0010-4655(93)90005-W. URL `https://www.sciencedirect.com/science/article/pii/001046559390005W`.

[18] M.S. Gordon, P. Goldhagen, K.P. Rodbell, T.H. Zabel, H.H.K. Tang, J.M. Clem, and P. Bailey. Measurement of the flux and energy spectrum of cosmic-ray induced neutrons on the ground. *IEEE Transactions on Nuclear Science*, 51:3427–3434, 2004.

CONCLUSION

We have developed the firmware responsible for operating and controlling the ODILE system, a custom controller designed for large arrays of CCDs in the DAMIC-M experiment. We also performed a number of tests integrating this system with skipper CCDs, achieving $\ll 1\ e^-$ of noise and demonstrating that the ODILE readout system meets its specifications. This system will allow long-term operation of possibly hundreds of CCDs in different operational configurations to search for dark matter with an $\mathcal{O}$(eV) threshold.

Beyond simply operating the CCD detectors, exploiting their full capability requires not only reducing any external or intrinsic radioactive backgrounds that may obscure potential signal, but understanding the backgrounds that remain to extract any signal above this background. We developed and validated a set of `GEANT4`-based simulations for DAMIC at SNOLAB that allowed us to build a detailed background model and set strong WIMP limits in spite of the presence of relatively high backgrounds. An excess of events at low energy of unknown origin, well-parameterized by an exponential spectrum with characteristic energy $\epsilon = 67 \pm 37$ $\mathrm{eV_{ee}}$ has been found. The source of this excess, either an unaccounted background or, more interestingly, a potential dark matter signal, will be explored further in the next phase of the DAMIC experiment (DAMIC-M) at the Modane Underground Laboratory. The design of this next-generation detector has been guided by the results from this background model, which has allowed the identification of individual isotopes and materials that contribute most highly to our background.

Simulations can only go so far in understanding backgrounds. One area simulations fall short in is physics involving nuclear reactions, such as spallation of silicon atoms by cosmic ray neutrons. This process produces multiple isotopes, with the most problematic for low-threshold DM searches being ^{3}H, due to its low energy beta decay. ^{3}H is produced in silicon detectors by cosmic ray neutrons for as long as they remain on the surface. The rate of this production has until now only been evaluated through nuclear physics simulations with almost an order of magnitude range in their estimates.

By using a high intensity neutron beam with an energy spectrum similar to that of cosmic ray neutrons to accelerate the ^{3}H activation process in silicon CCDs, we were able

to measure the neutron activation rate of ^{3}H. This measurement involved extensive work in correcting for the damage caused to the CCD by the neutron beam itself. From this measurement we have determined that the natural cosmogenic activation rate of ^{3}H in silicon at sea level is 112 ± 24 atoms/kg/day. This rate is sufficiently low that DAMIC-M will be able to meet its budgeted goal of 0.1 d.r.u. background from ^{3}H in the region $<$ 1keV (corresponding to approximately 2 months of activation) by following the existing development timeline. This measurement is also of interest for other silicon-based dark matter searches, which usually face similarly stringent requirements on ^{3}H background rates in the detector, and have so far been reliant on simulations of.

The work presented in this thesis has been of use not only for the current generation of the DAMIC experiment, but also lays much groundwork for further large-scale low-threshold silicon-based dark matter experiments.

APPENDIX A

ETHERNET

Ethernet is a standard protocol widely used in computer networking. This allows the use of standard commonly available components at every stage of the ODILE-server interface, from FPGA firmware to computer hardware and software. Altera even provides a "Megafunction", their "Triple-Speed Ethernet" (or "TSE") block which is can handle the physical interface with both the SFP fiber transceivers and the Marvell PHY chip, which greatly sped up and simplified development of the FPGA firmware. In addition, Ethernet is highly flexible, and easily allows many-to-one and one-to-many communication, allowing a single server to control many ODILE boards. It does not even limit the server to a maximum speed of 1 Gbps, as the server can attach to a switch that handles data at faster speeds (such as 10 Gig Ethernet), even if the ODILE itself is only capable of GigE links.

The physical layer of Ethernet (8B10B encoding, transmission hold-off, etc.) is handled by the PHY device (in our case, either the Marvell controller or the TSE megafunction), and not by our electronics firmware, and will therefore not be discussed here.

Our firmware is responsible for handling almost all the data transmitted over the Ethernet link, from the Ethernet frame header to the actual data itself (in OSI Model [1] terms, this is layer 2 and above). Ethernet is a packet-based protocol: data is transmitted not in a continuous stream, but in packets, where each packet can be anywhere from 64 to 1564 bytes long. Specifications for the Gigabit Ethernet protocol can be found in [2].

Each packet begins with a 14-byte header, composed of 3 parts: the destination MAC address, the source MAC address, and the EtherType designator (which specifies

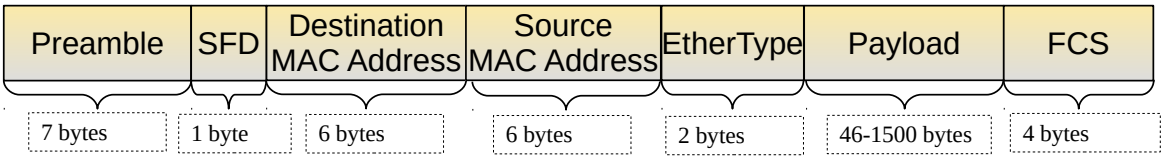


Figure A.1: Schematic of Ethernet header.

the nature of the payload). The MAC address is a 6-byte designator (usually given in the form "00:d8:61:a8:6e:dc") that specifies the unique physical address of the Ethernet device. This address is typically a globally unique, but Ethernet only requires that no two devices on a network share the same MAC address. The EtherType designates the nature of the payload data. We only use two types of payloads: IPv4, specified with the EtherType 0x0800, and ARP, designated by 0x0806. These are described in section A.3 and section A.2, respectively.

The end of each packet contains a 32 bit "Frame Check Sequence". This is a cyclic-redundancy check ("CRC") derived from the bits of the packet, which can detect most errors in data transmission. This CRC is generated and appended by the TSE megafunction, and is not described further here. After the header is the payload, which may be between 46 and 1500 bytes long. This payload may (in the case of IPv4 packets) contain it's own header-payload structure.

A.1 IPv4

The primary data protocol used in Ethernet communication is the so-called "Internet Protocol" (or IP). We specifically implement version 4[3], typically shortened to "IPv4", which uses 4 byte addresses, usually specified in the form "xxx.xxx.xxx.xxx", where each "xxx" is a decimal number between 0 and 255, representing 1 byte of the address.

The IPv4 header is by far the most complicated data structure used in our Ethernet packets, being 20 bytes long with 12 parts (although only a handful are relevant for our purposes, with the rest remaining set to reasonable hardcoded values). The structure of the header can be found in fig. A.2.

A.2 ARP

The Address Resolution Protocol ("ARP")[4] is used to map the specific MAC addresses (which are tied to hardware and the physical Ethernet link layer) to the higher level IPv4 addresses used by software to address data to and from the ODILE board. ARP

Byte	0	1	2	3
0	Version/IHL	DSCP/ECN	Total Length	
4	Identification		Flags/Offset	
8	TTL	Protocol	Header Checksum	
12	Source IP Address			
16	Destination IP Address			

Figure A.2: Schematic of IPv4 header. Except for the length, protocol, checksum, and addresses, all fields are set to reasonable hardcoded values, as they are not used or relevant to the ODILE communications. The only protocol values used are 0x01 (for ICMP echo responses), or 0x11, for UDP data.

is a request and reply protocol, where the requester sends out a packet asking for the MAC address of a specific IP address. This is sent using a special *broadcast* address, with the destination MAC set to all 1s (i.e. "FF:FF:FF:FF:FF"). All devices on the network see these packets, and if they have the IP address requested, respond an ARP reply, that contains their MAC address. This allows the Ethernet/IP link to be dynamically configured, so that devices can be added or change addresses in real time. This protocol is required for IPv4 communication, and so is implemented in the ODILE firmware.

A.3 UDP

The primary data type used in ODILE is the User Datagram Protocol ("UDP")[5]. UDP is one of the core data transmission protocols underlying the Internet (along with it's slightly more common cousin, Transmission Control Protocol or "TCP"). UDP is an extremely simple protocol that uses only 4 pieces of information: a source and destination IP address (used to route packets to the correct destination hardware), and a source and destination port (used to route the packet to the correct decoding software at the destination).

On the ODILE, the IP address is used along with the MAC address to specify which ODILE is data is being sent to and from, as well as which server receives the data (in cases which require multiple servers on the same network). The UDP port is used to designate the type of data being sent or received, so that the ODILE or server knows

Byte	0	1	2	3
0	Source port		Destination Port	
4	Length		Checksum (optional)	

Figure A.3: Schematic of UDP header.

Byte	0	1	2	3
0	Type	Code	Checksum	
4	Identifier		Sequencer Number	

Figure A.4: Schematic of ICMP header.

what to do with the data contained in the payload.

A.4 ICMP Ping

The Internet Control Message Protocol (ICMP) is a large and flexible protocol meant to support basic IPv4 communications with diagnostic and operational messages[6]. Several dozen message types are supported, but the most common and widely used is the echo reply/request message, commonly used with the "ping" command line tool. This is almost always the first and most reliable method of verifying basic communications functionality. Support is not technically required by IPv4 standard, but failure to respond to ICMP echo requests would be a *de facto* indicator that the device is not communicating.

ICMP packets are relatively simple. They are contained as payloads inside an IPv4 message. They are composed of an eight byte header (shown in fig. A.4), followed by a type-dependent payload. The only type supported by the ODILE has a type code of 8 (echo request), which will generate a reply with a type of 0 (echo reply). Both have a *code* field of 0. The *checksum* field is computed using the RFC 1071 [7] standard: the 1s complement sum of all 16-bit fragments of the ICMP packet (with the checksum field set to 0) is computed, and the 1s complement of this result is used in the checksum field. The ODILE does no validation of checksums for incoming echo requests. The remaining 4 bytes of the header data are composed of identifier and sequence fields: the identifier is typically unique to the "ping" process, while the sequence is incremented for subsequent

echo request/replies. The ODILE copies all identifier and sequence data directly from the echo request into the reply.

The ICMP packets also contain a data payload. For echo request and replies, this data is arbitrary: the reply is required to contain an identical copy of the request data payload. The payload is commonly filled with timestamp information, to easily compute round-trip packet times (by comparing the current time to the timestamp in the reply).

References

[1] Information technology — open systems interconnection — basic reference model: The basic model. *ISO/IEC 7498-1:1994*, 1994.

[2] H. Frazier. The 802.3z gigabit ethernet standard. *IEEE Network*, 12(3):6–7, 1998. doi: 10.1109/65.690946.

[3] Internet Protocol. RFC 791, September 1981. URL `https://rfc-editor.org/rfc/rfc791.txt`.

[4] An Ethernet Address Resolution Protocol: Or Converting Network Protocol Addresses to 48.bit Ethernet Address for Transmission on Ethernet Hardware. RFC 826, November 1982. URL `https://rfc-editor.org/rfc/rfc826.txt`.

[5] User Datagram Protocol. RFC 768, August 1980. URL `https://rfc-editor.org/rfc/rfc768.txt`.

[6] Internet Control Message Protocol. RFC 792, September 1981. URL `https://rfc-editor.org/rfc/rfc792.txt`.

[7] Computing the Internet checksum. RFC 1071, September 1988. URL `https://rfc-editor.org/rfc/rfc1071.txt`.

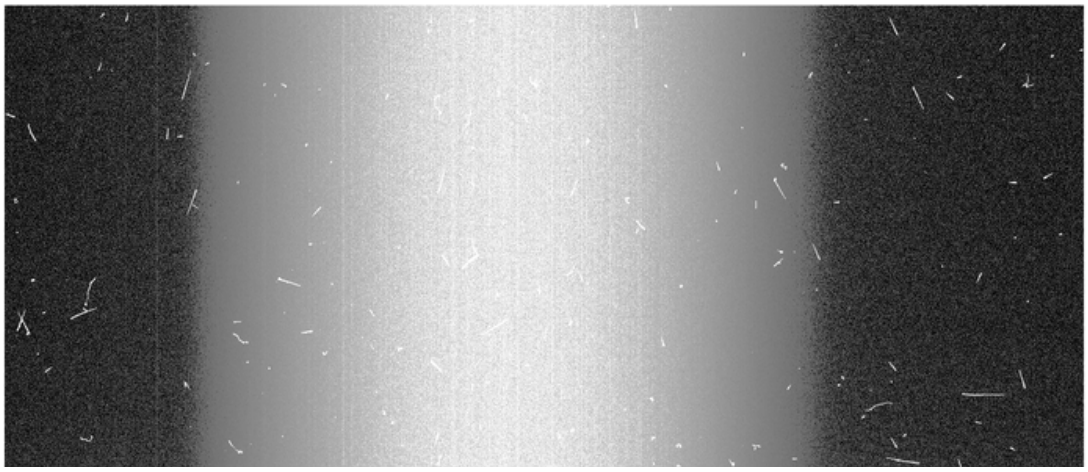
APPENDIX B

ADDITIONAL ACTIVATED CCD DATA

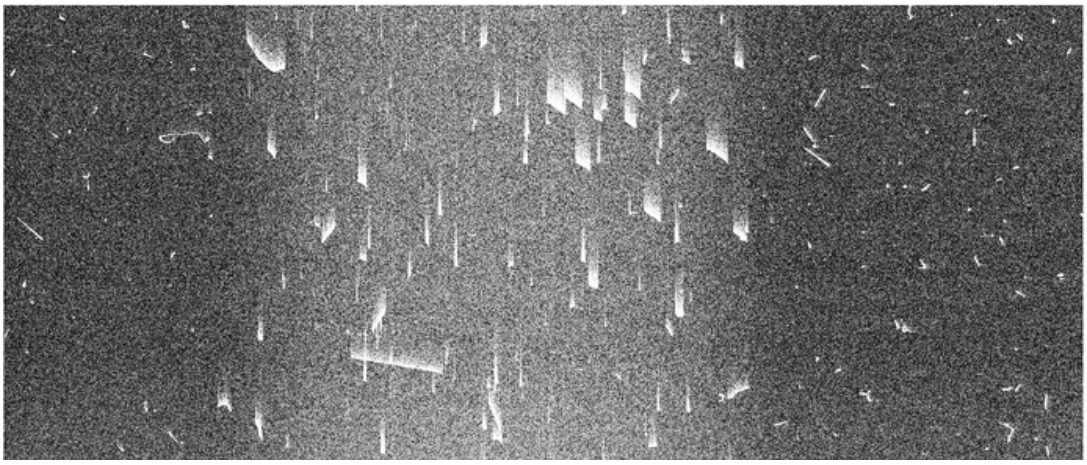
An additional set of data using the activated CCDs was taken using CCD #3 in November of 2019. This data set was taken with a ^{57}Co source to provide a high energy line that could be used for calibration at higher energies, and to explore charge reconstruction effects at this energy (with the goal of explaining the mismatch in the muon spectrum at energies >100keV). In the time between the science run and this second set of data, several changes had been made to the vacuum chamber which altered the thermal connectivity of the CCD to the cold finger. After this rework, the CCD temperature was ~ 20K cooler than the initial science data. The result was a roughly 200 fold decrease in the dark current (from ~250 e^-/pixel/image to ~1.2 e^-/pixel/image).

This reduction in dark current was expected to considerably simplify the analysis, but introduced an additional complication, which was the increase of vertical CTI in the images. This extremely prominent CTI is visible even in the raw images (see fig. B.1b), but can be seen most clearly in a comparison of the σ_y of the clusters vs row (see fig. B.2).

This CTI is particularly notable for being highly non-Poissonian: no Poissonian or binomial process can generate such long extended tails. This suggests that the process of CTI in damaged CCDs is much more complicated than the naive assumptions of CTI given above. An attempt was made to reproduce these long tails using exponentially distributed CTI, but this effort was ultimately abandoned due to lack of time. A full understanding of this data would require extensive work which was beyond the scope of this measurement.



(a) High dark current image. Note that even tracks within the activated region show little visible smearing from CTI.



(b) Low dark current image. The smearing from vertical CTI is clearly visible for elongated tracks inside the activated region of the CCD.

Figure B.1: Comparison of images taken with high (top) and low (bottom) dark currents.

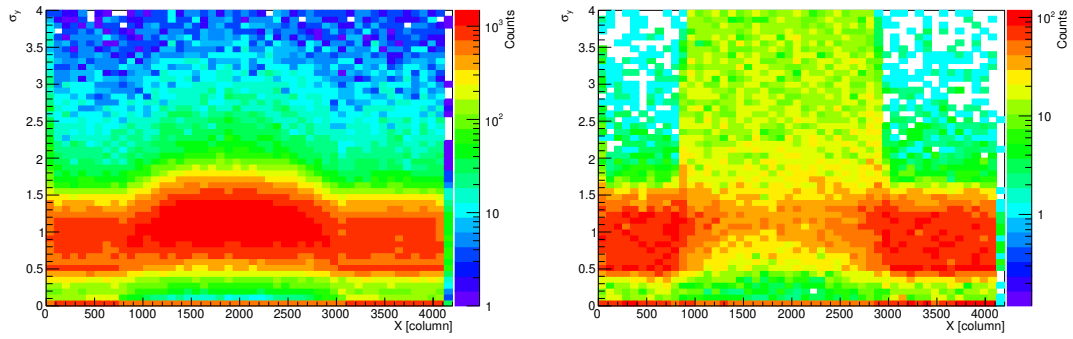


Figure B.2: Left: σ_y vs column for the science data set. Right: σ_y vs column for the low dark current data set. The extremely large σ_y is a characteristic of very high vertical CTI. Both plots only include events between 2 and 100 keV, to avoid the large rise in events at low energy and the effect of extended tracks at higher energies.