SuperCDMS SNOLAB, HVeV Run 3, and Development of KIPM Detectors

Thesis by
Taylor Benjamin Aralis

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2024
Defended September 8, 2023

Taylor Benjamin Aralis
ORCID: 0000-0002-3501-6948

ACKNOWLEDGEMENTS

The research described in this thesis is a great deal more than I could ever accomplish alone. Instead, it is the product of many great scientists' ideas and efforts. Even my own small contributions would have been impossible were it not for many colleagues, mentors, friends, and family members (with significant overlap among the categories). To all who have supported this work and myself, thank you.

Professor Sunil Golwala is the constant force that drives the SuperCDMS collaboration's work at Caltech. I am extremely grateful to have received that first call from him back in 2016 inviting me to work in his group. Since that time, he has thoughtfully guided both my research and education. He has helped me grow both my in-lab and analytic capabilities. He has also greatly improved my understanding of experimental design and motivations. The Friday meetings he organizes have been a great place to share progress, ask questions, and receive feedback without undue pressure. Contrary to popular belief, these meetings do have a constant time of occurrence. We simply lack sufficient statistics to know that time within ~2.5 hours.

The SuperCDMS and dark-matter KIPM group at Caltech is filled with dynamic individuals I am honored to have worked with. Brett Cornell, the senior grad student upon my arrival, was always a patient mentor willing to share his valuable time and substantial knowledge with a profoundly unknowledgeable G1. Thank you for helping me understand SuperCDMS, MKIDs, and a variety of unrelated topics you somehow know everything about. It has been a joy to work with Yen-Yung Chang for the majority of my time at Caltech. Yen-Yung has always provided an easy optimism that has helped me through my own bouts of pessimism. Thank you for being open-minded and patient even when responding to my occasionally half-baked questions and ideas. I will always be proud of our work repairing the B101 dilution refrigerator (even if it is far from a groundbreaking scientific accomplishment). I would also thank Osmond Wen for the incredible work he has done continuing and improving the many group projects that myself and previous researchers started. Thanks to his work, a world-leading dark-matter search using KIPM detectors feels more achievable and imminent than ever. Thank you Ritoban Basu Thakur for frequently lending a helping hand with the work in B101 and for always being a reliable source of news on the newest publications. Thank you Karthik Ramanathan for both your help progressing the KIPM work and for helping me understand the

ionization model for HVeV Run 3. Thank you Brandon Sandoval for bringing fresh energy to and carrying on the KIPM work at Caltech. I am excited to see how the work progresses during your own tenure. I would also like to thank the Caltech ObsCos administrators who have assisted our group, including Kathy Deniston, Sheri Stoll, and Nancy Roth-Rappard. Without them, the group's research output would immediately cease. Thank you for organizing our purchases, conference travel, proposal submissions, and everything else.

The SuperCDMS collaboration has helped me through each step of my PhD process. In addition to benefiting from the collaboration's experimental apparatus and results, I have personally benefited from my interactions with many of the collaboration members. I thank Matt Hollister for his work prepping (and teaching me to use) the tabletop cryogenics at Caltech, without which the HEMT characterization would have been substantially more difficult. I would also like to thank my mentors at SLAC, who helped me get the most out of the time I spent working there. Specifically, thank you to Richard Partridge, Tsuguo Aramaki, and Paul Brink. Thank you to the entire HVeV Run 3 working group. I know it was a challenging analysis for all of us. It could not have been done without the incredible work of Alexander Zaytsev, Corey Bathurst, and Derek Sincavage. Thank you Valentina Novati for being our most dedicated analysis supervisor. Additionally, thank you all for dealing with my half-asleep self during the earliest of morning meetings. I would also like to thank Matthew Wilson for his generous contributions to the Run 3 analysis and specifically for helping me understand the Run 2 limit-setting procedure and software.

An enormous thank you to all my family. Thank you for being my best friends and my support system throughout this extended process. Our Friday calls helped me get through both my PhD and the year we all stayed inside. Thank you Mom and Dad for supporting my education at every step along the way. Thank you for being there whenever I reached out for help and for always being kind. Thank you for facilitating family gatherings and for sharing your home. Thank you Matt and Vivian (and Luca) for allowing me to stay with you during my month working at SLAC. You made that indispensable experience an incredibly enjoyable one as well. Thank you Hilary and Elan for hosting board-game nights, birthday parties, and other hangouts in LA. Thank you Nick for all the memes and jokes we have shared. Thank you Zach for being a constant online companion helping to keep me (relatively) sane.

Finally, thank you Andrea for sticking with me through these (often difficult) times. Thank you for coming to Pasadena with me and for being patient even when I couldn't always tell you what was going to happen next. You inspire me with the determination and adaptability you have demonstrated in your own career. Thank you for being a partner in so many things and for making our apartment in Pasadena a home (with a cute dog and everything).

ABSTRACT

Dark matter is the theorized source of many observed large-scale gravitational effects. It is dark in the sense that it lacks any heretofore measurable direct interaction with the electromagnetic spectrum. Being unable to rely on absorption, reflection, or emission of photons makes studying dark matter particularly challenging. Excluding neutrinos, which fail to explain the observed large-scale effects, dark matter has never been conclusively identified in a local laboratory experiment. There are many proposed models that could explain both our large-scale observations and our lack of local observations while still allowing for the possibility of local observation. Ultra-sensitive direct-detection experiments attempt to make precisely such observations. Confirmed detection of a new stable particle would provide important information for improving our understanding of both dark matter and cosmological models.

The SuperCDMS SNOLAB experiment is a direct-detection experiment designed with an initial focus on particle masses ≤ 10 GeV. The experiment will measure both phonon and ionization signals in kg-scale semiconductor crystals held at cryogenic temperatures. In this thesis, I describe the experiment with emphasis on the ionization readout. I also detail the characterization process I performed on the ionization amplifier's low-power high-electron-mobility transistors (HEMTs).

SuperCDMS high-voltage eV-resolution (HVeV) detectors are gram-scale detectors designed to achieve single electron-hole-pair sensitivity. The first HVeV direct-detection search produced world-leading exclusion limits for dark-matter masses down to $\sim$1 MeV. Here, I present my work analyzing the third search using such detectors. Run 3 was the first to include multiple detectors operated simultaneously and achieved an order of magnitude greater exposure than previous runs. I report the resulting exclusion limits for electron-coupled, dark-photon, and axion-like-particle dark matter.

Lastly, I discuss work performed at Caltech towards the development of kinetic-inductance phonon-mediated (KIPM) dark-matter detectors. KIPM detectors use frequency-multiplexed kinetic inductance detectors (KIDs) and have the potential for excellent event-position reconstruction and background rejection. KIPMs also present a clear path towards sub-eV resolution on event recoil energy. Such detectors could be used as part of a payload upgrade in SuperCDMS SNOLAB. KIPMs could also be used in smaller-scale experiments similar to SuperCDMS HVeV.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] SuperCDMS Collaboration et al. *A Strategy for Low-Mass Dark Matter Searches with Cryogenic Detectors in the SuperCDMS SNOLAB Facility.* 2023. arXiv: 2203.08463 [physics.ins-det].
T.A. was a SuperCDMS collaboration and author-list member.

[2] M. F. Albakry et al. "Investigating the Sources of Low-Energy Events in a SuperCDMS-HVeV Detector." In: *Physical Review D* 105.11 (June 2022). DOI: 10.1103/physrevd.105.112006. URL: https://doi.org/10.1103/physrevd.105.112006.
T.A. was a SuperCDMS collaboration and author-list member.

[3] Karthik Ramanathan et al. "Identifying Drivers of Energy Resolution Variation in a Multi-KID Phonon-Mediated Detector." In: *Journal of Low Temperature Physics* 209.3-4 (June 2022), pp. 457–463. DOI: 10.1007/s10909-022-02753-5. URL: https://doi.org/10.1007/s10909-022-02753-5.
T.A. took data and performed initial analysis that was later expanded upon to produce this LTD proceeding.

[4] O. Wen et al. "Performance of a Phonon-Mediated Detector Using KIDs Optimized for Sub-GeV Dark Matter." In: *Journal of Low Temperature Physics* 209.3-4 (July 2022), pp. 510–517. DOI: 10.1007/s10909-022-02764-2. URL: https://doi.org/10.1007/s10909-022-02764-2.
T.A. wrote much of the software for data taking, resonance fitting, and temperature-dependence fitting, and helped advise the data analysis.

[5] I. Alkhatib et al. "Constraints on Lightly Ionizing Particles from CDMSlite." In: *Physical Review Letters* 127.8 (Aug. 2021). ISSN: 1079-7114. DOI: 10.1103/physrevlett.127.081802. URL: http://dx.doi.org/10.1103/PhysRevLett.127.081802.
T.A. was a paper advocate, helped with some of the initial writing, and helped with editing during the review process.

[6] I. Alkhatib et al. "Light Dark Matter Search with a High-Resolution Athermal Phonon Detector Operated Above Ground." In: *Physical Review Letters* 127.6 (Aug. 2021). ISSN: 1079-7114. DOI: 10.1103/physrevlett.127.061801. URL: http://dx.doi.org/10.1103/PhysRevLett.127.061801.
T.A. was a SuperCDMS collaboration and author-list member.

[7] D. W. Amaral et al. "Constraints on Low-Mass, Relic Dark Matter Candidates from a Surface-Operated SuperCDMS Single-Charge Sensitive Detector." In: *Physical Review D* 102.9 (Nov. 2020). ISSN: 2470-0029. DOI: 10.1103/physrevd.102.091101. URL: http://dx.doi.org/10.1103/PhysRevD.102.091101.
T.A. was a SuperCDMS collaboration and author-list member.

[8] T. Aralis et al. "Constraints on Dark Photons and Axionlike Particles from the SuperCDMS Soudan Experiment." In: *Physical Review D* 101.5 (Mar. 2020). ISSN: 2470-0029. DOI: 10.1103/physrevd.101.052008. URL: http://dx.doi.org/10.1103/PhysRevD.101.052008.
T.A. was a SuperCDMS collaboration and author-list member.

[9] R. Agnese et al. "Production Rate Measurement of Tritium and Other Cosmogenic Isotopes in Germanium with CDMSlite." In: *Astroparticle Physics* 104 (Jan. 2019), pp. 1–12. ISSN: 0927-6505. DOI: 10.1016/j.astropartphys.2018.08.006. URL: http://dx.doi.org/10.1016/j.astropartphys.2018.08.006.
T.A. was a SuperCDMS collaboration and author-list member.

[10] R. Agnese et al. "Search for Low-Mass Dark Matter with CDMSlite Using a Profile Likelihood Fit." In: *Physical Review D* 99.6 (Mar. 2019). ISSN: 2470-0029. DOI: 10.1103/physrevd.99.062001. URL: http://dx.doi.org/10.1103/PhysRevD.99.062001.
T.A. was a SuperCDMS collaboration and author-list member.

[11] R. Agnese et al. "Energy Loss due to Defect Formation from 206Pb Recoils in SuperCDMS Germanium Detectors." In: *Applied Physics Letters* 113.9 (Aug. 2018), p. 092101. ISSN: 1077-3118. DOI: 10.1063/1.5041457. URL: http://dx.doi.org/10.1063/1.5041457.
T.A. was a SuperCDMS collaboration and author-list member.

[12] R. Agnese et al. "First Dark Matter Constraints from a SuperCDMS Single-Charge Sensitive Detector." In: *Physical Review Letters* 121.5 (Aug. 2018). ISSN: 1079-7114. DOI: 10.1103/physrevlett.121.051301. URL: http://dx.doi.org/10.1103/PhysRevLett.121.051301.
T.A. was a SuperCDMS collaboration and author-list member.

[13] R. Agnese et al. "Low-mass Dark Matter Search with CDMSlite." In: *Physical Review D* 97.2 (Jan. 2018). ISSN: 2470-0029. DOI: 10.1103/physrevd.97.022002. URL: http://dx.doi.org/10.1103/PhysRevD.97.022002.
T.A. performed the pre-release verification of the paper's public data release by reproducing the official spin-independent and spin-dependent WIMP cross-section limits using event energy and detector efficiency data.

[14] Y. -Y. Chang et al. "Development of a Massive, Highly Multiplexible, Phonon-Mediated Particle Detector Using Kinetic Inductance Detectors." In: *Journal of Low Temperature Physics* 193.5-6 (Dec. 2018), pp. 1199–1205. DOI: 10.1007/s10909-018-1900-9.
T.A. wrote much of the software for testing the CPW feedline and characterizing the KIDs on the 80-KID device.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

DARK MATTER

Dark matter is the label given to the theorized source of many observed gravitational effects. It is dark in the sense that it lacks any heretofore measurable direct interaction with the electromagnetic spectrum. Being unable to rely on absorption, reflection, or emission of photons makes studying dark matter particularly challenging. Instead, dark matter's existence and distribution are inferred by its large-scale gravitational effects on various astronomical and cosmological measurements. Since experimental work towards understanding the nature of dark matter is the primary focus of this thesis, I will discuss a few of such measurements in this section.

1.1 Astronomical Evidence

The astronomical evidence for dark matter includes observations of galactic rotation curves, galaxy clusters, and gravitational lensing.

1.1.1 Galactic Rotation Curves

The orbital velocities of visible objects (stars or gas) within a galaxy can be used in combination with the radial position of such objects to infer the total mass contained within that galaxy. When performed for many galaxies, this calculation has consistently implied greater mass than that explained by luminous matter alone. In particular, the observed velocities in spiral galaxies remain constant (or increase) at high radius, although they would be expected to decrease outside the core of luminous matter [1]. This particular effect can be explained by modifying the laws of gravity or by assuming the presence of non-luminous matter driving the velocity at high radius (dark matter). While there have been galaxy observations with velocity curves consistent with luminous matter only [2], these observations are rarities and may actually be the result of inaccurate distance measurements [3].

1.1.2 Galaxy Clusters

Similar to galaxies themselves, galaxy clusters show evidence of having more mass than that which luminous matter can account for. This result can be seen through the velocity dispersion of a galaxy cluster's component galaxies. The unknown,

non-luminous, matter driving velocity dispersion at high radius is what was first labeled dark matter [4]. The same result can be seen by observing the intracluster gas, which makes up the majority of a cluster's baryonic matter. The intracluster gas emits thermal x-rays, which can be measured to determine the gas pressure in the cluster. Assuming the cluster is stable, the gravitational force pulling a unit cell of gas towards the cluster center must be counterbalanced by the force of gas pressure pushing that cell radially outward. One can therefore use the gas pressure to calculate the gravitational force as a function of radius. Using the gravitational force, one can then calculate the cluster's mass [5]. Both measurements imply greater mass than explained by luminous matter alone.

1.1.3 Gravitational Lensing

Gravitational lensing is the deflection of light from a distant source caused by the bending of spacetime due to mass between the source and observer. Observing how an object (such as a galaxy cluster) lenses the light from various sources (such as background galaxies) is another way to measure an object's mass. When performed, these measurements also indicate the presence of non-luminous dark matter in objects such as galaxy clusters [5].

In the interesting case of the "Bullet Cluster" (1E 0657-56), gravitational lensing has been used to observe the movement of mass during a collision between two galaxy clusters. Lensing confirms that a majority of the mass from each cluster passes unobstructed through the other. Meanwhile, observations of x-ray emissions by the intracluster gas confirms that the majority of baryonic matter is slowed by interaction during the collision. This discrepancy reveals once again that the majority of mass must actually be non-baryonic [6]. Additionally, the separation of the lensing-derived mass and the luminous matter cannot be resolved with a simple modification to the theory of gravity. The Bullet Cluster has also been used to set limits on the self-interaction of dark matter since the two clusters' dark-matter contents must have passed through each other during the collision. With this method, the self-interaction cross section (per unit mass) is proposed to be $\sigma/m < 0.7$ cm^2/g [7]. Recent studies have shown that more careful treatment of the intracluster gas may weaken this limit [8].

1.2 Cosmological Evidence

Cosmology is the sub-field of astronomy that specifically deals with the origin and evolution of the large-scale properties of the universe. The existence of dark matter

(or a modified gravitational theory) would significantly impact large-scale evolution, so cosmological observations have important implications for dark-matter research. Cosmological evidence for dark matter includes the cosmic microwave background (CMB) and the formation of large-scale structure.

1.2.1 CMB Power Spectrum

The cosmic microwave background (CMB) is the highly uniform light remaining from the moment of recombination. Recombination occurred when electrons cooled enough to bind with protons and form neutral hydrogen, decoupling baryonic matter from photons and making the universe transparent to light for the first time. While uniform to one-hundredth of a percent, the CMB has small anisotropies, which carry significant information about the early universe.

When plotted as a function of angular separation, the CMB's temperature power spectrum has several peaks corresponding to wavelengths of acoustic waves with maximum density or maximum rarefaction at the time of recombination. Before recombination, acoustic waves in the baryon density occurred due to the competing forces of gravity and radiation pressure. Modes of different wavelengths reached different stages of oscillation before recombination, modulated by the speed of the baryon-photon plasma. Modes that reached maximum compression (density) at the time of recombination contribute to the first peak in the power spectrum. Modes that compressed once, then expanded to reach maximum rarefaction contribute to the second (and so on). Since dark matter is unaffected by radiation pressure, it did not oscillate like the baryons. Instead, it began a continuous accumulation that enhanced (via gravitational force) the compression-phase of baryons afterward. For this reason, the dark matter density in the early universe affects the height of all peaks beyond the second. By modeling dark matter's effect on the baryon oscillation, astronomers have confirmed that dark matter must make up about 85% of the mass in the universe [9].

1.2.2 Structure Formation

The stars, galaxies, and galaxy clusters observed today are the result of matter accumulating via the gravitational force. However, the force from luminous matter alone should not be sufficient to have formed the observed structure in the time since recombination (when the CMB confirms that baryonic matter had a highly uniform distribution). Before recombination, baryonic matter was ionized and oscillated due to radiation pressure (as described in Section 1.2.1). It therefore

could not begin accumulating before recombination. Dark matter provides an elegant solution. Dark matter began accumulating after radiation-matter equality (before which expansion prevented structure growth) and continued through recombination and beyond. After recombination, baryonic matter was gravitationally drawn to the dark-matter structure, making the structure visible to modern astronomers. This model essentially allows structure formation to begin earlier (via dark matter) than would be possible with baryonic matter alone. Simulation has shown this to be a viable explanation [10].

A requirement of this model is that most structure-forming dark matter was non-relativistic at the time of matter-radiation equality. Such dark matter is known as either cold or warm dark matter (CDM or WDM, respectively) depending on its mass. CDM includes particles with mass $\geq$ 1 GeV. WDM includes models with mass between 1 keV and 1 GeV. Both CDM and WDM would have sufficiently low velocities to recreate the gradual clumping of mass over time that eventually leads to large-scale structure [11]. This is the bottom-up formation model that is currently favored by cosmologists. The alternative model, known as top-down formation, describes massive objects breaking down to form the less-massive objects which constitute the universe today. Top-down models fit better with hot dark matter (HDM), which is any dark matter that had relativistic velocity at the time of matter-radiation equality. In order to attain such velocity, HDM particles must have $m \leq$ $\sim$10s eV. Neutrinos are HDM that has already been observed by direct detection. HDM alone cannot recreate bottom-up structure formation because its mean free path was too large to form structure as small as galaxies. It has been proposed that WDM may resolve some issues observed with CDM-formation models (such as the missing satellites problem) [11].

1.3 Candidates

While the effects of dark matter have been observed in these large-scale examples, it has never been observed directly in a laboratory or local experiment (excluding the observed neutrinos, which fail to explain the large-scale effects). Direct-detection experiments like SuperCDMS attempt to make precisely such observations. Assuming there is nothing unusual about our own galaxy, dark matter should be present with approximately the same abundance observed in others. If dark matter interacts via methods other than gravity, its interaction cross section must be small enough to have gone unnoticed throughout human history. There are many proposed models that could explain both our large-scale observations and our lack of local observa-

tions while still allowing for the possibility of local observation via ultra-sensitive direct-detection experiments.

1.3.1 WIMPs

Historically, the most popular model for searches has been the Weakly Interacting Massive Particle (WIMP). Rather than a single model with one definition, WIMPs are actually a class of particles proposed to have mass and interaction strength comparable in magnitude to the weak (W and Z) bosons. Since their inception, a number of specific WIMP models have been proposed. WIMP dark-matter candidates include the lightest supersymmetric particle (LSP) predicted by supersymmetry (SUSY), the lightest Kaluza-Klein particle (LKP) predicted by compactified universal extra dimensions (UED), and the lightest particle with odd T-parity (as defined in little Higgs theories) [12]. Additionally, SUSY, UED, and little Higgs theories each provide potential solutions to the hierarchy problem.

WIMP models can explain the relic dark-matter density observed today using the freeze-out production mechanism. With this mechanism, WIMPs are proposed to have been in thermal equilibrium with the high-temperature early universe. During this time, WIMPs were constantly produced via high-energy interactions between standard-model particles and annihilated via self-interaction. As temperature dropped, the relatively light standard-model particles no longer had sufficient energy to produce WIMPs, while the annihilation process continued. As the number of WIMPs decreased and the universe expanded, the probability of WIMP-WIMP interactions (and the corresponding annihilation) also decreased. The eventual ceasing of both production and annihilation can successfully reproduce the relic dark-matter density we observe today [13]. Specifically, this model reproduces the relic density when dark matter is assumed to have weak-scale interactions. This result only faintly depends on particle mass and is true for masses from ~1 GeV to ~100s of TeV. The fact that the interaction and mass scales both agree with a new weak-scale particle is what has historically been known as the "WIMP miracle."

Unitarity arguments provide an additional $\mathcal{O}(100\ \text{TeV})$ upper bound on individual particle mass for freeze-out CDM [14]. Such arguments show that for higher masses, dark matter would overclose the universe. To "overclose" the universe means to define a model that describes a greater density than cosmological measurements allow. Being weak-scale, WIMPs are expected to have mass within 1-2 orders of magnitude of the W and Z bosons (~100 GeV). This means that WIMPs satisfy the

freeze-out requirement, the unitarity bound, and are classified as CDM (see Section 1.2.2).

It is worth noting that the WIMP paradigm does not require that all of dark matter be WIMPs. There may be a large variety of dark-matter particles (WIMP and non-WIMP) with their own complex interactions. Varying the fraction of the relic density comprised of WIMPs allows one to justify a wider range of WIMP masses and cross sections [15].

Nuclear-recoil direct-detection searches have excluded significant mass and cross-section parameter space in the traditional mass range of WIMPs (see Figures 1.1 and 1.2). Collider experiments are also capable of excluding WIMP parameter space, particularly for the higher end of this mass region [16].

1.3.2 QCD Axion

Another historically popular dark-matter candidate is the axion. The axion was proposed as part of the Peccei-Quinn solution to the strong-CP problem from quantum chromodynamics (QCD). The QCD Lagrangian contains terms that violate CP, while no strong-interaction CP violation has been observed experimentally [11]. The usual method of probing strong-CP violation is to measure the neutron electric dipole moment (EDM). Failure to observe such an EDM has constrained the scale of strong-CP violation to a very small value. The Peccei-Quinn solution attempts to resolve this issue by proposing an additional symmetry that could absorb the CP-violating term via a field redefinition. The QCD axion is the resultant particle associated with that field. Axions are not expected to have undergone freeze-out, since their standard-model coupling would be too small to reach thermal equilibrium with the early universe. Instead, axion production is generally attributed to other mechanisms such as vacuum misalignment or emission from cosmic strings (not to be confused with superstrings of string theory) [17]. The misalignment mechanism describes the scenario where an initial fluctuation in the value of the axion field is frozen by expansion before the field can relax. With either production mechanism, axions would still be classified as CDM.

The historically accepted range for axion dark-matter mass is between 10^{-5} and 10^{-2} eV. This range is bound on the lower end by the requirement to not overclose the universe and on the upper end by previous laboratory experiments and astronomical observations [17]. It has been recognized more recently that the assumptions leading to the over-closure bound were unreasonably strong. Therefore, the lower bound

can be weakened and the mass window expanded using various other assumptions [18].

1.3.3 Hidden-Sector Dark Matter

More recently, the dark-matter community has begun considering a broader range of models with less emphasis on simultaneously resolving other outstanding questions in physics. This has lead to new models (such as dark photons) as well as generalizations of historically popular models (including the QCD axion and WIMP). Many of the resulting models are classified as "hidden-sector" dark matter. Hidden-sector dark matter is a general term for candidates that rely on a "dark portal" for interaction with the standard model. The "dark portal" would be a new force with a new mediator which interacts in some way with both the hidden and normal sectors of matter. Dark-portal interactions that couple the hidden and normal sectors must be sufficiently weak to have remained previously undetected.

1.3.4 Dark Photons

The typical example of a proposed "dark portal" mediator is the dark photon. Dark-photon models propose there are two kinetically mixed vector (spin 1) bosons. One of these bosons would directly interact with the standard model. The other boson would directly interact with the hidden sector. Mixing between the two bosons is defined by a kinetic mixing factor (ε). When the resultant Lagrangian is diagonalized, it produces two stable bosons: the standard massless photon and a dark photon which may be either massive or massless.

In the massive case, the dark photon may interact with both hidden-sector and standard-model particles, while the standard photon interacts only with standard-model particles. The lack of observed interactions with a massive dark photon can be attributed to the modulation of the interaction by ε, which may be very small.

In the massless case, the regular photon may interact with both sectors, while the dark photon only interacts with the hidden sector [19]. The possible coupling between regular photons and the hidden-sector may seem contradictory to the sector's "hidden" nature, but the small coupling would make dark matter only appear as very lightly or *milli-charged* particles. Although the massless dark photon would not directly couple to ordinary matter, it could still interact through its coupling to the regular photon. These interactions would be suppressed by the requirement that they be of dimension 5 or higher. A full description of the massive and massless cases is out of the scope of this thesis, but can be found in [19].

In some models, the massive dark photon itself is the missing dark matter. Similar to axion models, dark-photon dark matter could be generated with the right abundance via vacuum misalignment and would be considered CDM. Such dark matter would be detectable through photoelectric absorption of a photon by an atom in a detector. The only difference from regular photon absorption would be the presence of the mixing parameter [19]. The modified absorption cross section can be seen in Eq. 1.1.

$$\sigma_{A'} = \varepsilon^2 \sigma_{p.e.} \tag{1.1}$$

1.3.5 Axion-Like Particles

The QCD axion described in Section 1.3.2 is another hidden-portal candidate. It can also be generalized to the Axion-like particle (ALP), which uses the same coupling and production mechanisms, but is not bound by the constraints of solving the strong-CP problem [19]. With the constraints loosened, ALP models have a greater range of possible masses and interaction strengths for experiments to probe. There are also theoretical motivations for some non-QCD-axion ALP candidates, including the majoron and familon [20].

1.3.6 Hidden-Sector WIMPs

Proposed hidden-sector particles include hidden-sector equivalents to WIMPs. Instead of interacting via the weak or other visible-sector forces, these WIMP-like particles would interact via the dark-portal force or a gauge force exclusive to the hidden sector. Hidden-sector WIMPs could have undergone their own version of freeze-out (originally described in Section 1.3.1) to generate the observed relic abundance. Hidden-sector freeze-out differs from regular freeze-out in that the hidden-sector thermal history does not need to match that of the visible sector. Instead, constraints on the ratio of the sectors' temperatures can be set based on the hidden-sector's impact on cosmology. For example, hidden-sector particles would contribute to the number of relativistic degrees of freedom in the early universe. This is true even without a dark portal between the two sectors. Relativistic degrees of freedom affect the expansion rate of the universe. Big Bang nucleosynthesis (BBN) is highly sensitive to the early expansion rate. Therefore, the requirements of BBN constrain the hidden-sector thermal history [15]. We discussed above how freeze-out produces the correct relic density (Ω_χ) for weak-scale coupling and mass. Other combinations of coupling (g_χ) and mass (m_χ) can produce the same

relic density when scaled according to Eq. 1.2. By adjusting the strength of the proposed hidden-sector force, one can justify a large range of hidden-sector WIMP masses and still produce the correct relic abundance. This is sometimes referred to as the "WIMPless miracle" [15]. While this miracle does not propose to resolve any outstanding questions about the standard model, it does justify a new potential search region for WIMP-like interactions.

$$\Omega_\chi \sim \frac{m_\chi^2}{g_\chi^4} \tag{1.2}$$

Theoretical models which could explain hidden-sector WIMPs include gauge-mediated supersymmetry breaking (GMSB). GMSB naturally predicts multiple sectors which interact via messenger particles (our "dark portal"). GMSB also predicts superpartners with mass proportional to the proposed coupling squared. This is the same relationship required by Eq. 1.2. If one assumes the right scale for GMSB, they can then justify all the same combinations of hidden-sector WIMP properties allowed by the freeze-out model [15].

1.4 Direct-Detection Methodology and World-Leading Limits

Direct-detection experiments seek to observe local (in-lab or perhaps on-Earth) interactions with one or more of the proposed dark-matter particles. Confirmed detection of a new stable particle would provide important information that can be used to improve our understanding of dark matter. Measurements of the mass and interaction strength could be used to calculate how much of all dark matter might consist of such a particle. If the particle makes up a large amount of dark matter, it may be suitable for explaining the large-scale observations discussed in Sections 1.1 and 1.2. Explaining the aforementioned gravitational observations could remove any need for a modified theory of gravity on large (non-quantum) scales. Measurements could also be used to improve cosmological models and constrain various particle-physics theories. It is worth noting that there are theories with overlapping regions of proposed mass and coupling, and therefore even confirmed direct detection will not necessarily elevate one of the proposed models above all others.

1.4.1 Exclusion Plot Basics

Direct-detection experiments use target masses where energy depositions can be studied in order to identify interactions with dark matter. Often the target material

is chosen for its compatibility with a detection mechanism and its availability in a radiopure form. Typical materials include semiconductor crystals (as in the cases of SuperCDMS, SENSEI, or EDELWEISS) and liquid noble gases (as in the cases of XENON or LUX). The typical experimental goal is to record a statistically significant number of dark-matter interactions (with the target mass) which cannot be explained by backgrounds alone.

In the case that a discovery is not made, most experiments use their data to set an exclusion limit, or a value of interaction strength above which said experiment would have made a discovery (with some statistical confidence). Exclusion limits are generally set on whatever interaction parameter is relevant to the model being tested. In traditional WIMP searches looking for nuclear recoils, this might be the cross section for interaction between a WIMP and a nucleon. Another example would be the kinetic mixing parameter (ε) used in dark-photon models. Exclusion limits are usually set for a range of potential dark-matter masses which are then used as the x-axis in typical exclusion-limit plots. This section will display and discuss a number of example exclusion plots (Figures 1.1 – 1.5).

1.4.2 Neutrino Fog

A common feature on many dark-matter exclusion plots is a lower region labeled as the "neutrino fog." The neutrino fog represents the region of parameter space where one would observe a statistically-significant number of solar, atmospheric, and/or supernova background neutrinos [21]. Energy depositions from such neutrinos (through coherent elastic neutrino-nucleus scattering, CEνNS) will appear dark-matter-like in conventional direct-detection experiments. The high systematic uncertainty on the expected rate of neutrinos increases the difficulty in distinguishing them from dark matter and making a discovery in this region. This uncertainty is currently dominated by the uncertainty on atmospheric neutrinos. Having an improved model of such would lower the neutrino fog. The neutrino fog can also be navigated by considering characteristics other than the energy-deposition spectra. Dark matter is expected to undergo annual modulation, meaning that the precise rate of events should change based on the position (and therefore motion) of Earth relative to the sun. This modulation would not appear identically in neutrinos. The velocity of halo dark matter is also expected to predominately lie in the galactic plane, in the direction of the galaxy's rotation. This behavior can be exploited by directional detectors to differentiate dark matter from neutrinos.

1.4.3 Spin-Independent and Spin-Dependent Nuclear Recoils

Many experiments search for target-mass nuclei recoiling from interactions with dark matter. Resulting limits are generally set for two types of interactions: spin-independent (SI) interactions, where dark-matter couples to the nuclear mass (A), and spin-dependent (SD) interactions, where dark-matter couples to the nuclear spin (J). Historically, SI limits have been much stronger (and closer to the neutrino fog) than SD limits. The difference can partially be attributed to SI coupling being proportional to A^2. Experimentalists are therefore motivated to select the largest A material that is compatible with their experiment's other requirements. Example SI nuclear-recoil exclusion plots can be seen in Figure 1.1. Cross section is converted to cross-section per nucleon, allowing results from experiments with different target nuclei to be compared. The comparison implicitly assumes that protons and neutrons have identical dark-matter interactions.

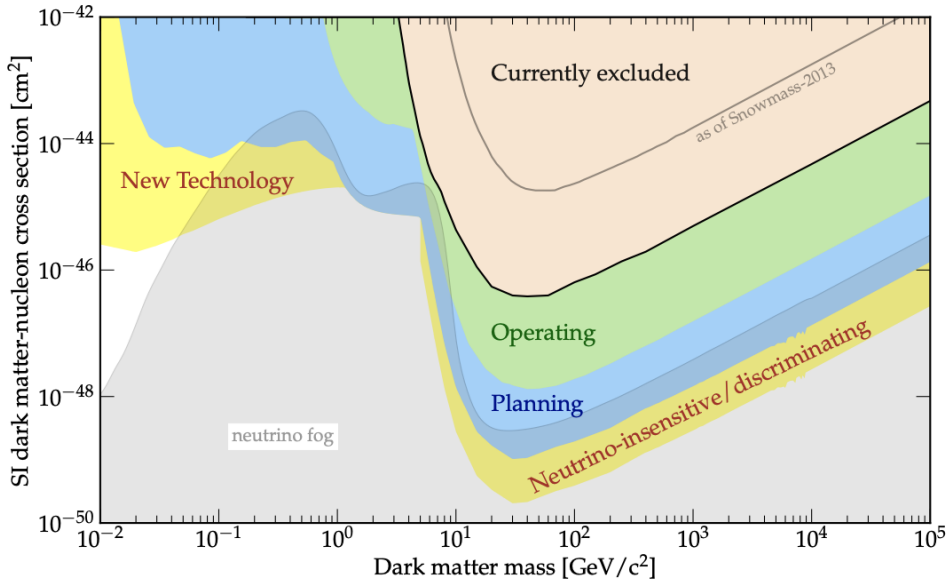


Figure 1.1: 90%-confidence exclusion limits for spin-independent nucleon cross section. The neutrino fog, where solar neutrino backgrounds begin to confound traditional dark-matter searches, is shown for a xenon target in grey. Taken from the Snowmass 2021 Particle Dark Matter Report [21].

For SD nuclear recoils, target nuclei with an odd number of neutrons, an odd number of protons, or an odd number of each are required. Such nuclei have $J \neq 0$ and are therefore sensitive to spin interactions. The coupling is proportional to $J(J+1)$ and can be optimized, but high-J materials are relatively rare (and the highest possible $J = 8$). Additionally, many typical direct-detection targets (such as silicon

or germanium) have an even number of protons and naturally low concentrations of odd-neutron isotopes. Experiments using liquid xenon targets are an exception, since natural xenon has large fractions of ^{129}Xe ($J = 1/2$) and ^{131}Xe ($J = 3/2$) [22]. SD limits are often calculated under one of two assumptions: dark matter only interacting with neutrons or only interacting with protons. The difference is significant because of the difference in spin contribution from the component quarks of each nucleon [23]. Example SD nuclear-recoil exclusion plots can be seen in Figure 1.2. In this example, the difference between protons and neutrons is ignored. Cross section is converted to cross-section for a $J = 1/2$ nucleon, allowing results from experiments with different J to be compared.

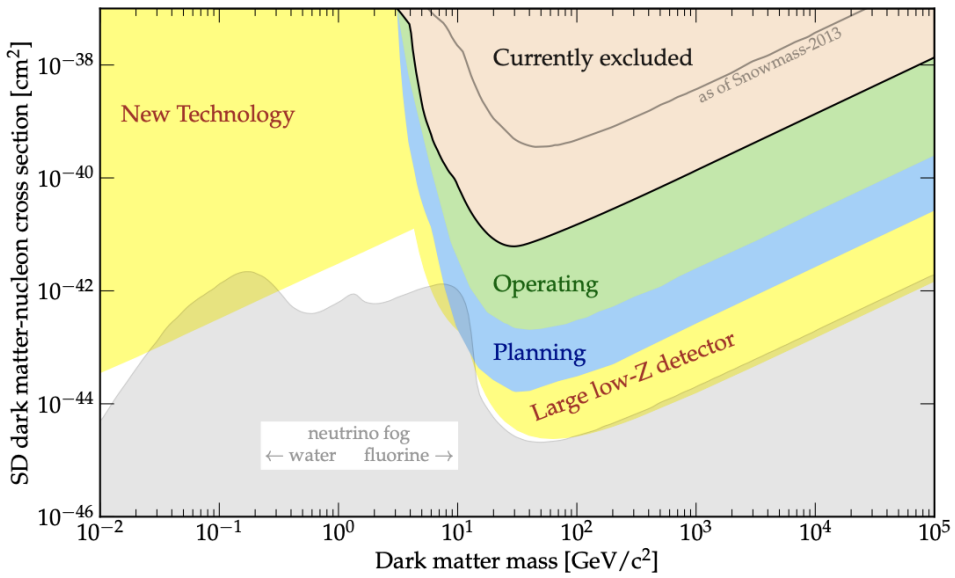


Figure 1.2: 90%-confidence exclusion limits for spin-dependent nucleon cross section. In this case, scattering with neutrons or protons. The neutrino fog, where solar neutrino backgrounds begin to confound traditional dark-matter searches, is shown for water and fluorine targets in grey. Taken from the Snowmass 2021 Particle Dark Matter Report [21].

1.4.4 Electron-Scattering Experiments

Dark-matter nucleon scattering is an effective way to search for higher mass dark matter (greater than ~10 times the nucleon mass), but becomes significantly less useful for lower masses. This can be understood using conservation of energy and momentum for elastic scattering in the rest frame of the target particle (see Eqs. 1.3 – 1.6).

$$m_\chi v_\chi = m_\chi v'_\chi + m_{target} v'_{target} \tag{1.3}$$

$$\frac{1}{2} m_\chi v_\chi^2 = \frac{1}{2} m_\chi v_\chi'^2 + \frac{1}{2} m_{target} v_{target}'^2 \tag{1.4}$$

$$v'_{target} = \frac{2 m_\chi}{m_\chi + m_{target}} v_\chi \tag{1.5}$$

$$\frac{1}{2} m_{target} v_{target}'^2 = \frac{4 m_\chi m_{target}}{\left(m_\chi + m_{target}\right)^2} \left(\frac{1}{2} m_\chi v_\chi^2\right) \tag{1.6}$$

Eq. 1.6 confirms that the target-particle recoil energy is maximized by selecting a target with mass equal to the anticipated dark matter. Target masses much larger than the dark matter will only receive a small fraction of the possible energy (e.g., a 10× larger or 10× smaller target receives only 1/3 the possible energy). This is one reason that direct-detection experiments have begun searching for electron recoils in addition to nuclear recoils. The lower mass of the electron allows for sensitivity to lower mass dark-matter scattering. It is worth noting that actual bound-electron scattering would not be purely elastic, but this calculation provides a reasonable explanation for why higher-mass targets become less useful. In semiconductor targets like those used by SuperCDMS, some deposited energy must go into excitation of the electron above the band gap.

Dark-matter-electron scattering limits are generally calculated for different choices of momentum-transfer form factor ($F_{DM}(q)$), which determine the dependence of the interaction on exchanged momentum. The most common choices of form factor are $F_{DM} = 1$ and $F_{DM} = (\alpha m_e / q)^2$. The former is appropriate for point-like interactions exchanging a heavy mediator. The latter is appropriate for interactions that exchange a light (or massless) mediator [24]. Example electron-scattering exclusion plots (for both form factors) can be seen in Figure 1.3.

1.4.5 Electron-Absorption Experiments

It is also common to set limits on various processes describing absorption by target electrons. Such processes include vector-Boson absorption for dark photons and pseudoscalar absorption for ALPs. For absorption processes, the entire rest-energy of the incident particle is absorbed during the interaction. The resulting distribution of energy depositions for a given mass of particle will be a delta function at that

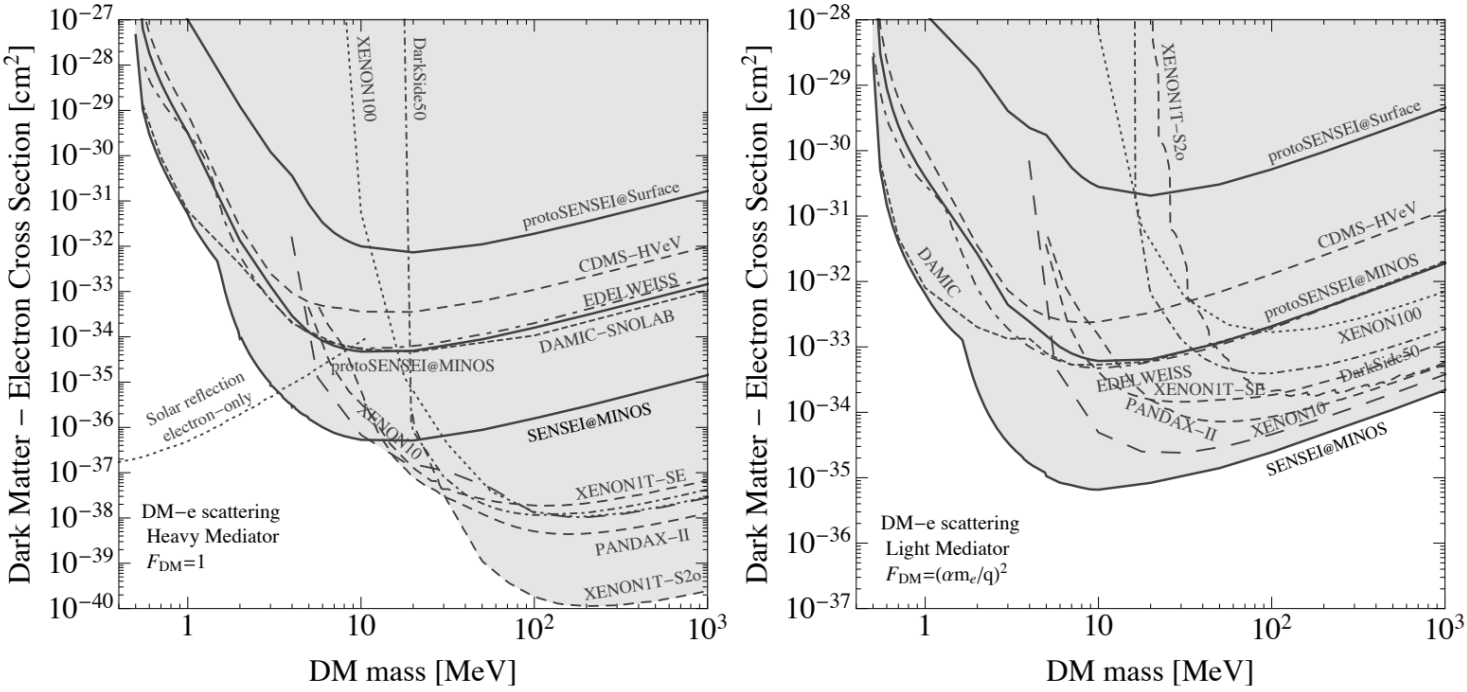


Figure 1.3: Dark-matter-electron scattering exclusion-limit plots taken from the Snowmass 2021 Low-Threshold Direct Detection Report [25]. The left and right plots assume a heavy and light mediator particle, respectively.

mass. In practice, a real detector will smear that delta function according to its energy resolution, but the final signal shape will still be significantly different from those of nuclear or electron scattering. Example dark-photon and ALP exclusion plots can be seen in Figures 1.4 and 1.5, respectively.

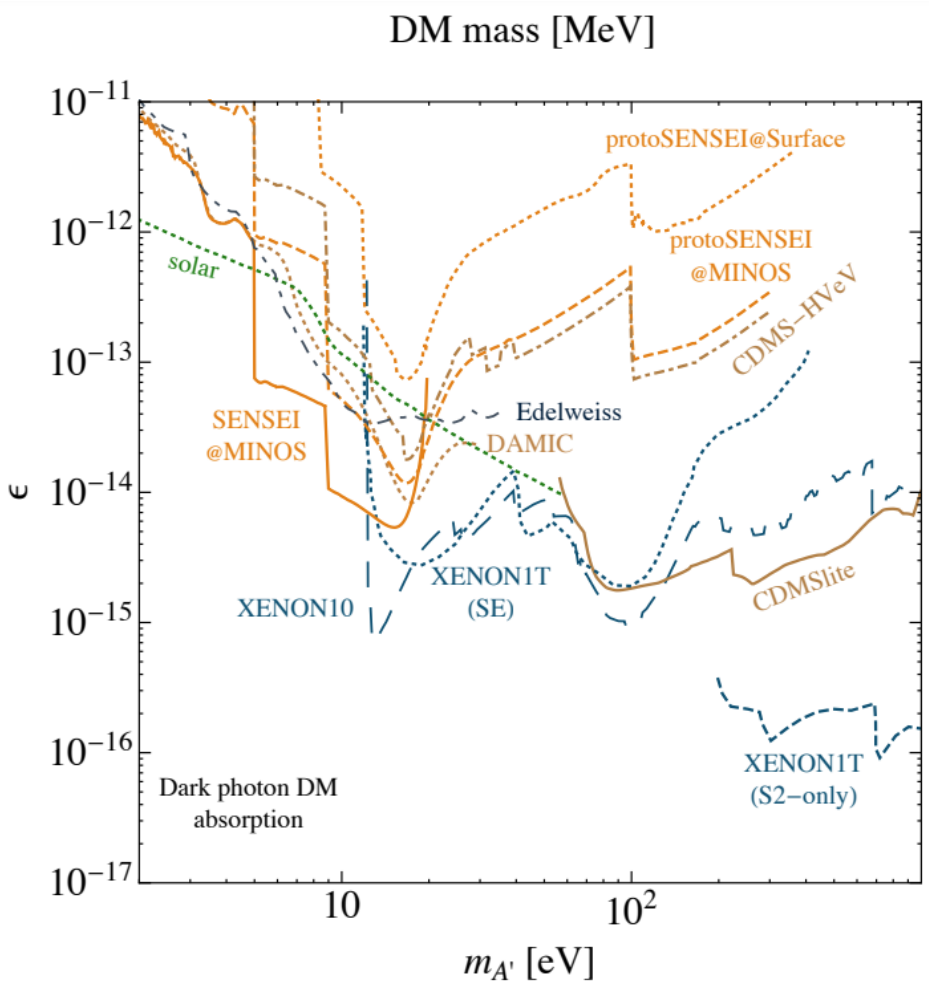


Figure 1.4: A dark-photon absorption exclusion-limit plot taken from the Snowmass 2021 Low-Threshold Direct Detection Report [25]. Limits are set on the kinetic mixing parameter between dark and standard-model photons.

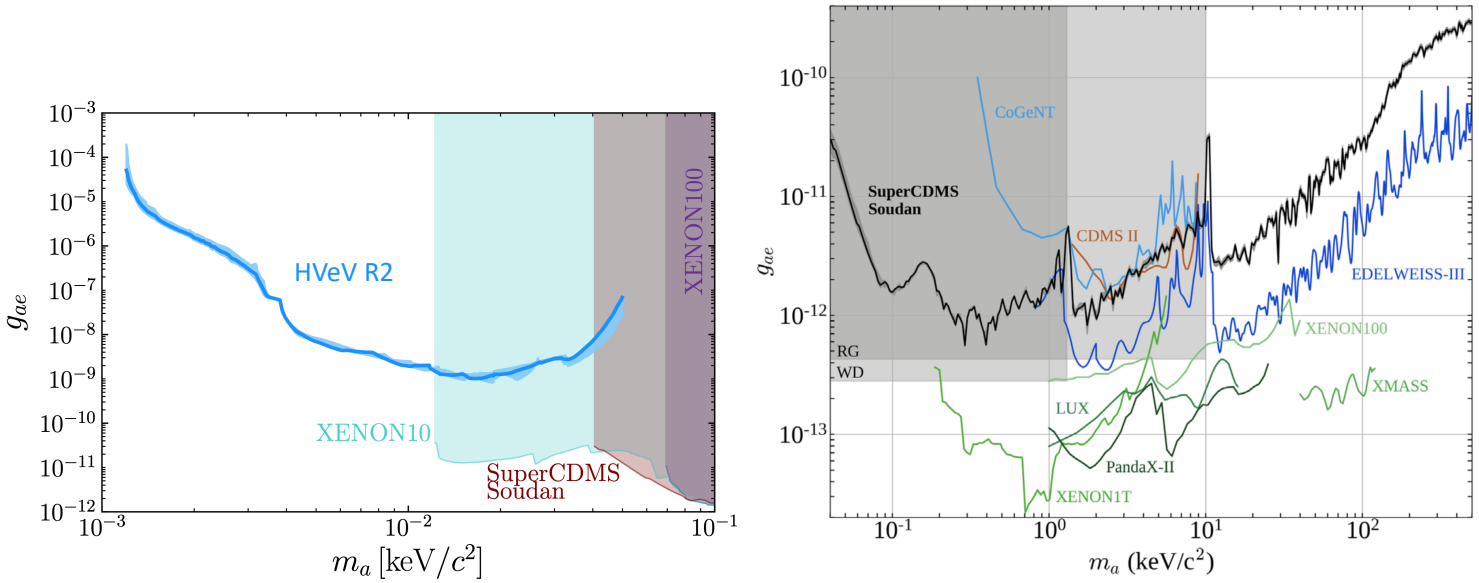


Figure 1.5: ALP exclusion limits set on the axioelectric coupling factor g_{ae}. (Left) Limits on the lower-mass region, taken from SuperCDMS HVeV Run 2 [26]. SuperCDMS HVeV detectors are sensitive to single ionizations in Si (~1.2 eV) and are therefore sensitive to absorption of particles of equivalently low mass. (Right) Limits on the higher-mass region, taken from the SuperCDMS Soudan dark photon and ALP paper [27]. The CDMS II, SuperCDMS Soudan, CoGeNT, and EDELWEISS-III experiments used germanium detectors. The discontinuities at 1 and 10 keV correspond to germanium atomic binding energies, which affect limits through the photoelectric cross section. Particles more massive than each binding energy have an additional means of absorption (exciting the associated bound electron), so the experiment's sensitivity increases above each. Each experiment looses sensitivity below its threshold. The SuperCDMS Soudan threshold was set by cuts removing regions of the detector with high rates of low-quality events. The plot's shaded regions are excluded by observations of stellar cooling, but require different assumptions than the direct detection limits and can be difficult to compare.

Chapter 2

SUPERCDMS SNOLAB

The SuperCDMS SNOLAB experiment is a direct-detection experiment designed with an initial focus on particle masses $\leq$ 10 GeV. The experiment will measure phonon and ionization signals in kg-scale semiconductor crystals held at cryogenic temperatures.

Phonon signals provide a precise way to measure nuclear-recoil energy with a very low intrinsic energy threshold ($O(1\,\text{meV})$ per acoustic phonon [28]). For example, Eq. 1.6 shows that an elastic recoil between a 5 GeV dark matter particle traveling 300 km/s (1e-3 c) and a ~30 GeV nucleus will deposit ~1 keV of energy in the phonon system of the crystal. This corresponds to ~1e6 phonons. In semiconductors, sufficiently energetic nuclear recoils will also generate ionization in the form of electron-hole pairs. This is due to the positive charge of the recoiling nucleus which can pull electrons from their sites.

The largest backgrounds in such a search will produce electron recoils rather than nuclear recoils. An electron recoil produces more ionization than a nuclear recoil that deposits equivalent energy. Therefore, electron recoils can be identified and removed by comparing the phonon and ionization signals. The SNOLAB iZIP detectors (which we will discuss in Section 2.1.2) are designed to measure both ionization and phonon energy in order to remove electron-recoil backgrounds.

As discussed in Sections 1.4.4 and 1.4.5, electron recoils may also result from dark-matter scattering or absorption. By considering such recoils, the SNOLAB experiment can also search for dark matter < 1 GeV without being inhibited by the kinematic-matching effect. The larger ionization signal from electron recoils increases the signal in such searches. The initial SNOLAB target crystals will be silicon and germanium. On average, only ~3 eV of electron-recoil energy is required to create an $e^- h^+$ pair in such crystals. The resulting ionization yield is $\gtrsim 5\times$ the yield in liquid noble experiments (xenon and argon have ionization energies of ~16 and ~24 eV, respectively) [29]. Using electron recoils, the initial SNOLAB payload will be sensitive to scattering with dark matter down to a few MeV and absorption of dark matter as light as a few eV.

The experiment is currently under construction and will be operated in the SNOLAB underground facility in Sudbury, Ontario. Operating in an underground lab provides a massive overburden to help minimize non-dark-matter signals from cosmic-ray secondaries and is common for direct-detection experiments. The SNOLAB experiment is a follow-up to the previous large-scale SuperCDMS experiment, which operated at the Soudan Underground Laboratory [30].

In the following sections, we will discuss the detectors, setup, backgrounds, and projected sensitivity of the SNOLAB experiment. We will then discuss the ionization-signal readout in detail. In order to ensure the readout functions as intended, I characterized each of the high-electron-mobility transistors (HEMTs) that will be used as the input stage of the ionization amplifier. We will discuss the HEMT characterization procedure and results.

2.1 Detector Design

SuperCDMS SNOLAB will use both silicon and germanium crystals as its target materials. The two materials have different nuclear masses and different intrinsic backgrounds. As discussed in Section 1.4.4, nuclear mass affects the energy transferred during dark-matter-nucleus scattering. With a nuclear mass of ~30 (~70) GeV for Si (Ge), we expect ~75% (~44%) of a 10 GeV particle's kinetic energy to be transferred to the recoiling nucleus. The transfer efficiency decreases for lighter dark-matter particles.

The SNOLAB experiment will also use two different designs of SuperCDMS detector (HV and iZIP). We will discuss the details and benefits of both designs in the following subsections. Each detector design has been produced using both Si and Ge, leading to four total device types (Si HV, Si iZIP, Ge HV, and Ge iZIP).

2.1.1 HV Detectors

The SuperCDMS high-voltage (HV) detectors consist of cylindrical crystals with arrays of phonon sensors lithographically patterned on the top and bottom flat surfaces. The phonon sensors are Quasiparticle-trapping-assisted Electrothermal-feedback Transition-Edge Sensors (QETs). During a phonon-generating event in the crystal, the QETs' superconducting aluminum fins collect phonon energy, generating quasiparticle excitations in the Al. The quasiparticle energy is then transferred (via quasiparticle diffusion and trapping) to tungsten Transition-Edge Sensors (TESs) in contact with each fin. Each TES is maintained (via electrothermal feedback) precisely in its transition between normal and superconducting. In this state, the

addition of quasiparticle energy will cause a sharp increase in TES resistance. In response, electrothermal feedback will provide less power to the TESs to return them to their original (mid-transition) state. The decrease in Joule dissipation in the TES is integrated to calculate the energy absorbed by the TES.

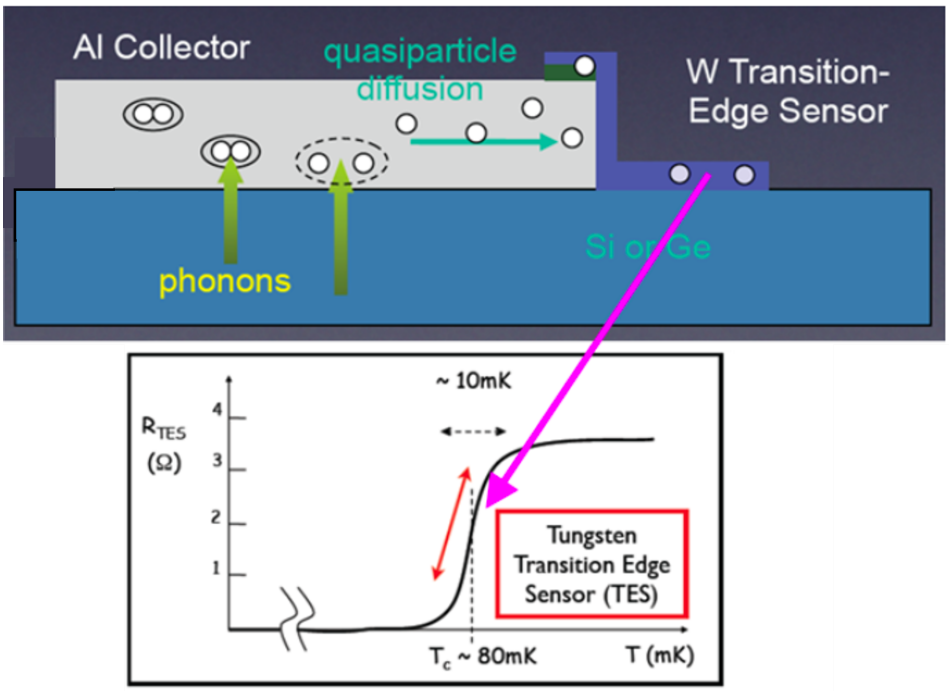


Figure 2.1: A diagram illustrating the operation of the SuperCDMS QET phonon sensors. Taken from the SuperCDMS SNOLAB Technical Design Report.

Phonon sensors are used to detect both recoil phonons and Neganov-Trofimov-Luke (NTL) phonons. NTL phonons are produced by accelerating charge across a semiconductor [31, 32]. The phonon energy is equivalent to the total work done on the charge. If charge is accelerated across a known voltage, the NTL-phonon energy can be used to extract the charge value. SuperCDMS detectors use NTL phonons to indirectly measure ionization signal using QETs. This is achieved by applying a large voltage difference (ΔV) between the top and bottom flat surfaces. Specifically, the phonon readout systems (QETs and readout lines) on each surface are biased with opposing DC voltages. The potential difference creates an electric field through the bulk of the detector pointing from one flat surface to the other. When electron-hole pairs are created by an ionizing event, the bulk electric field accelerates the opposing charges towards opposite detector surfaces and produces NTL phonons. For each electron-hole pair, the NTL-phonon energy (E_{NTL}) will be

equivalent to $e\Delta V$ (the work done accelerating a charge e across the entire crystal). For a uniform electric field, this result is independent of the event's depth within the detector. If n_{eh} electron-hole pairs are created, the total phonon energy (E_{total}) will follow Eq. 2.1.

$$\begin{aligned} E_{total} &= E_{recoil} + E_{NTL} \\ &= E_{recoil} + n_{eh} e\Delta V \end{aligned} \tag{2.1}$$

E_{recoil} is the phonon energy produced in the initial interaction including the energy that ionizes the electrons and holes. During an event, some fraction of E_{recoil} is stored in the excited $e^- h^+$ pairs. Upon reaching surface metal, e^-s and h^+s recombine with free charges and release the stored energy as phonons. Therefore, all of E_{recoil} is eventually observable as phonons.

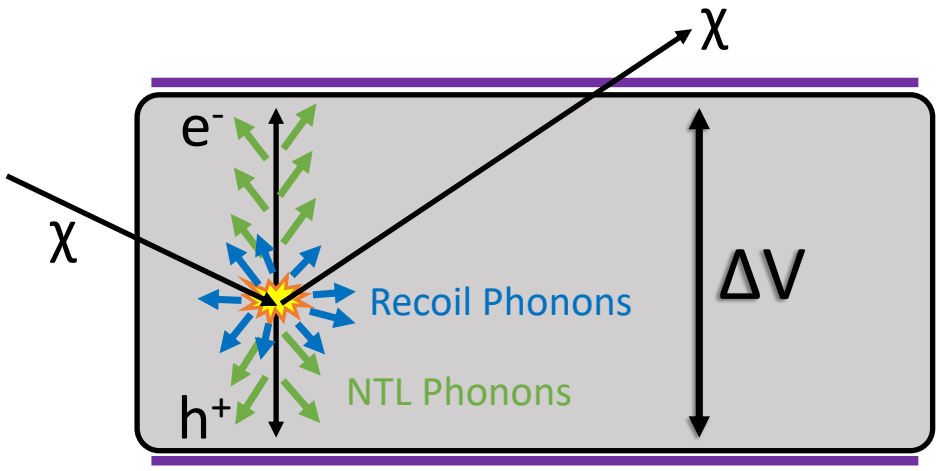


Figure 2.2: A visualization of the initial recoil phonons and the NTL phonons generated by the acceleration of one electron-hole pair. Reproduced from the SuperCDMS SNOLAB Technical Design Report.

We can see from Eq. 2.1 that increasing ΔV will increase the total energy of an ionizing event. Since the QET energy resolution is independent of ΔV, we can dramatically improve a detector's signal resolution by choosing a very large ΔV. In practice, this means selecting the largest ΔV that does not cause the semiconductor to break down and transmit significant leakage current. The NTL effect would convert such leakage current into a large phonon background. The SuperCDMS HV detectors are designed to achieve low energy threshold by using the maximum viable ΔV across a detector with QETs optimized for energy resolution. With anticipated

ΔV of 100 V and phonon-energy resolutions of a ~10s eV, the HV detectors are sensitive to dark-matter masses significantly below 1 GeV [33].

The QETs on each detector surface are divided into separate channels covering different regions of the surface. Each channel contains 100s of QETs, which would require extensive cryogenic and warm readout electronics to measure individually. Instead, all QETs in a channel are read out in parallel (combining their signals) and cannot be distinguished. For the HV detectors, each flat surface has six channels: An inner channel at the center of the surface, three wedge-shaped channels covering different azimuthal regions around the center, and two concentric ring-shaped channels covering the high-radius edge of the surface. An image of an HV detector and a diagram of its phonon channels are shown in Figure 2.3. The separate channels allow for basic position reconstruction of individual events. The reconstruction will be used to remove events that are close to the radial edge of the detectors. Near-surface events are more likely to originate from weakly-penetrating backgrounds. The bias electric field will also be less uniform near the radial edge and may produce events with incomplete charge drifting. For both reasons, it is beneficial to identify and remove events at high radius. The wedge-shaped channels also provide information on the azimuthal position of each event. This information can be used to remove events from a localized background in the testing apparatus (e.g., a radio-impure screw on one side of the detector housing).

A measurement of E_{total} with imprecise position reconstruction provides no means of differentiating the recoil and ionization signals. Without this information, the HV detectors are incapable of identifying and removing electron-recoil backgrounds. These backgrounds therefore limit the HV-detector sensitivity.

2.1.2 iZIP Detectors

SuperCDMS interleaved Z-sensitive Ionization Phonon (iZIP) detectors have the same form factor and phonon-sensing method as HV detectors. In addition, the iZIP detectors have lithographically patterned electrodes interleaved with the phonon channels. These electrodes collect charge for the ionization measurement and apply the iZIP's NTL voltage bias (while the iZIP QETs are maintained at DC ground). Interleaving biased electrodes with grounded QETs generates an electric field parallel to the surface near the top and bottom of each detector (see Figure 2.4). This is a purposeful feature that forces ionization generated near either flat surface to be entirely absorbed by metal on that surface. By removing events with asymmetric

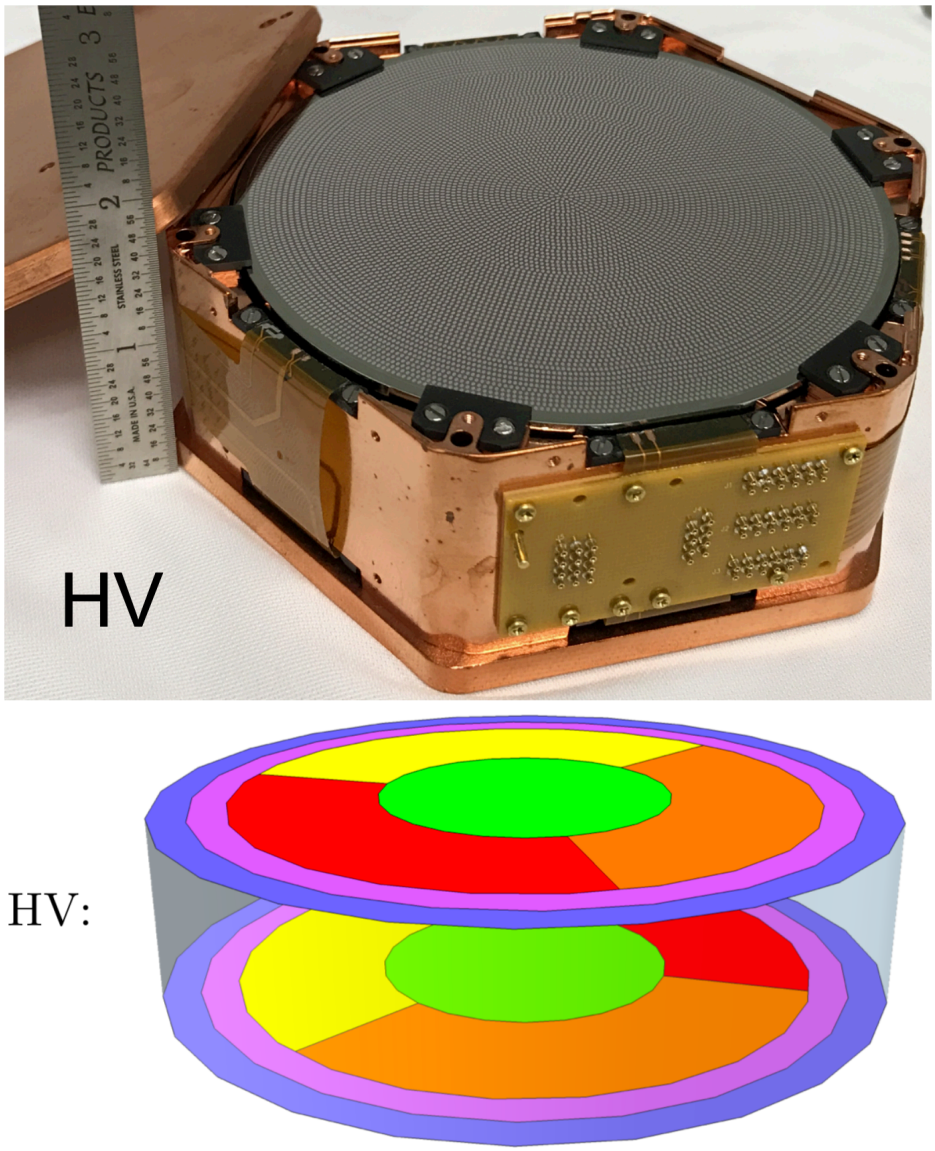


Figure 2.3: The SuperCDMS SNOLAB HV detector. (Top) A prototype SuperCDMS HV detector in its Cu housing. (Bottom) The locations of separate phonon-sensor channels on an HV detector. Channels of matching color on the top and bottom surfaces are not connected. The three azimuthally separated wedge channels are offset by 60° between top and bottom to allow for additional position information.

charge collection, iZIP detectors can therefore remove surface background events [34]. In the crystal bulk (> few mm deep), the electric field will be perpendicular to the flat surfaces and will accelerate electrons and holes to opposing detector surfaces (as in the HV detectors). iZIP detectors are "Z-sensitive" because of this different response to bulk and surface events.

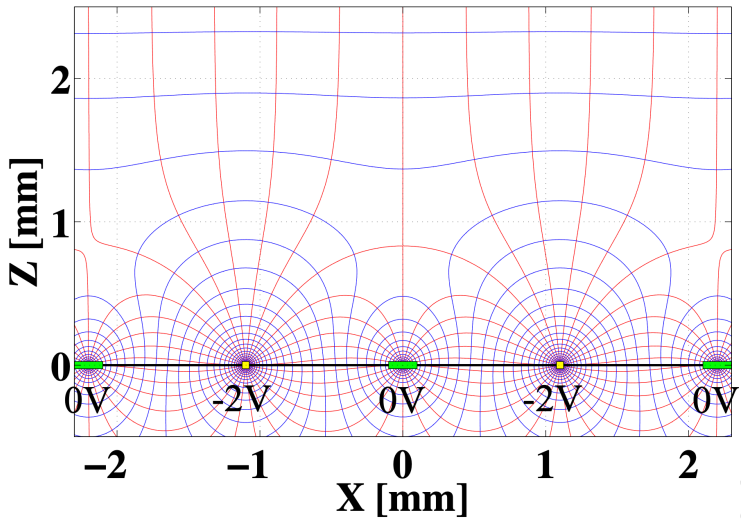


Figure 2.4: Example electric field near one surface of an iZIP detector with electrodes (yellow) biased to -2 V. The QETs (green) are grounded and the opposing-surface electrodes (not shown) are biased to +2 V. Electric field lines are red. Equipotential lines are shown in blue. Z is the depth into the detector with Z=0 being the detector's bottom surface. In this scenario, electrons produced within ~1 mm of the surface will be collected by the QETs rather than being accelerated to the opposing surface. Holes will be collected by the electrodes and measured as ionization signal. Therefore, ionization occurring within that region will only produce signal on the bottom surface and can be easily identified.

By measuring both ionization (with the electrodes) and phonon energy (with the QETs), iZIP detectors provide a way to discriminate between nuclear-recoil and electron-recoil events. Electron recoils produce greater ionization (per deposited energy) than nuclear recoils and can thus be identified by their higher value of n_{eh}/E_{recoil}. n_{eh} is provided by the ionization measurement and can be used in combination with the phonon measurement of E_{total} to calculate E_{recoil}. Typically, the resultant ratio is normalized to the average electron-recoil energy per electron-hole-pair produced (ε_{eh}) and is referred to as the ionization yield ($y\,(E_{recoil})$).

$$y\,(E_{recoil}) = \frac{n_{eh}}{E_{recoil}}\varepsilon_{eh} \tag{2.2}$$

Performing an accurate calculation of ionization yield requires that E_{NTL} not be so large that its uncertainties prevent accurate measurement of E_{recoil}. For this reason, iZIP detectors must be biased with a lower voltage than HV detectors ($\Delta V \approx 4$ V rather than 100 V). A smaller ΔV gives the iZIP detectors a higher threshold than the HV detectors. This higher threshold limits iZIP-detector sensitivity to dark-matter

masses $\gtrsim$3 GeV, but the iZIPs' ability to discriminate electron recoils gives them greater sensitivity than the HV detectors within this region.

Like HV detectors, the iZIP detectors have their phonon channels divided by area on each detector surface. Instead of having three wedges and two outer rings, each iZIP-detector surface has four wedges and one outer ring. An image of an iZIP detector and a diagram of its phonon channels are shown in Figure 2.5. The iZIP-detector electrodes are also divided into channels on each surface. The electrode channels consist of an outer ring (covering the same area as the outer-ring phonon channel) and a center (covering everything else). Having radial information from both charge and phonon channels provides an improved ability to identify and remove edge events. The increased number of wedge-shaped regions improves the ability to reconstruct azimuthal position (compared to the HV detectors).

With the iZIP detectors' ability to identify near-surface (flat or edge) and electron-recoil events, they are capable of removing the majority of backgrounds expected in the SNOLAB experiment [33]. Information on iZIP backgrounds can also be used to perform statistical background subtraction on the HV detectors, which are fabricated using the same methods and materials. In this way, the background-characterizing iZIP and low-threshold HV detectors make for a complementary initial payload for the SuperCDMS SNOLAB facility.

2.2 SNOLAB Experimental Setup

The infrastructure of a direct-detection experiment is arguably as important as the experiment's detection mechanisms. The experimental site, for example, will provide both backgrounds (from the building materials surrounding the experiment) and shielding from backgrounds (such as cosmic-ray secondaries). For some models, the site may shield from dark-matter signal as well as backgrounds. Additional shielding installed on-site will have a similar effect but can be more precisely controlled. The materials used for operating and housing detectors may introduce new backgrounds. The design of readout electronics may introduce or mitigate noise impacting the detector threshold and resolution.

In addition to the normal considerations, SuperCDMS SNOLAB also requires infrastructure to maintain the detectors at cryogenic temperatures ($\sim$15 mK) for at least a year of uninterrupted data taking.

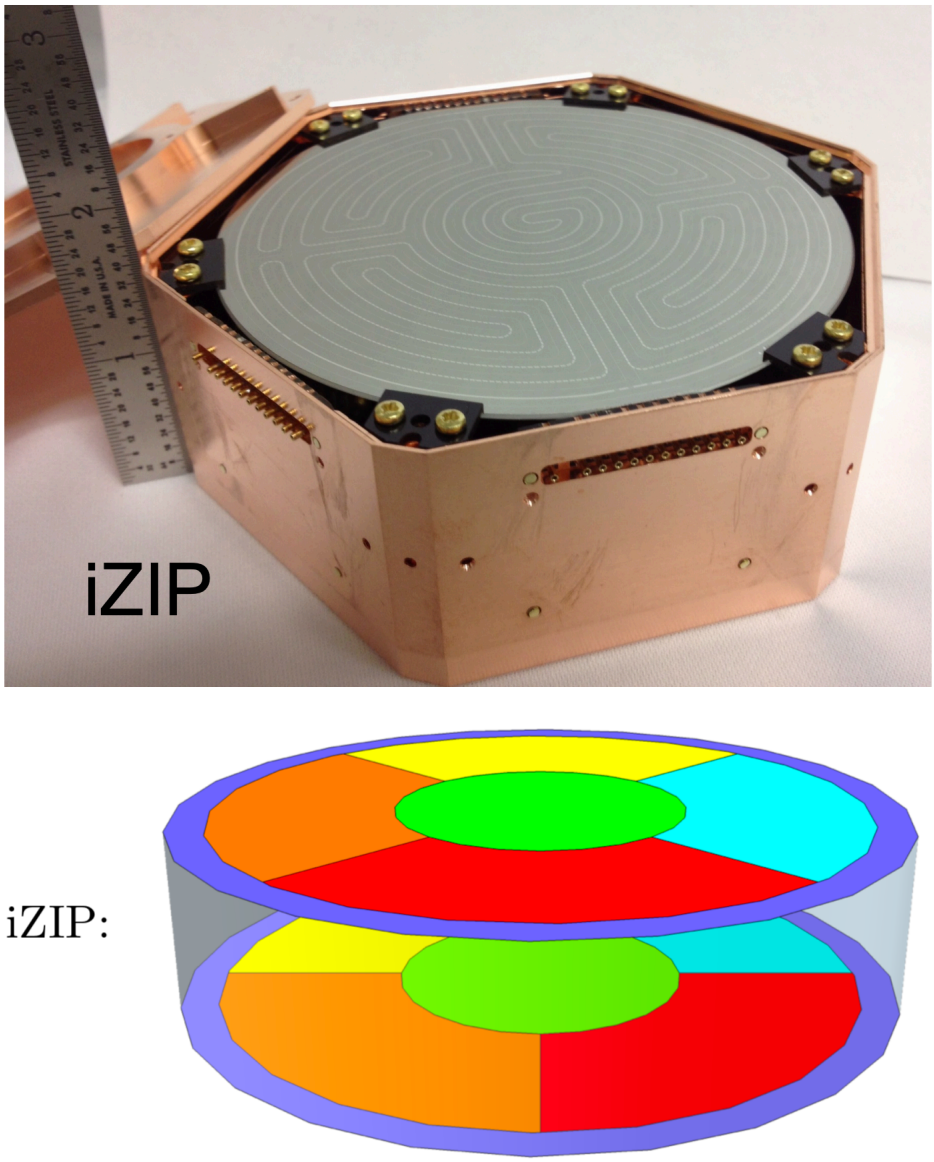


Figure 2.5: The SuperCDMS SNOLAB iZIP detector. (Top) A prototype SuperCDMS iZIP detector in its Cu housing. (Bottom) The locations of separate phonon-sensor channels on an iZIP detector. Channels of matching color on the top and bottom surfaces are not connected. The four azimuthally separated wedge channels are offset by 45° between top and bottom. iZIP detectors also have separate electrode channels. On each surface, the outer electrode channel shares its area with the outer phonon-sensor channel (purple). The inner electrode channel covers the rest of each surface.

2.2.1 SNOLAB Site

The SNOLAB site is an underground laboratory in Sudbury, Ontario, Canada. It is an expansion of the Sudbury Neutrino Observatory (SNO) experimental facility

and is located in a currently active nickel mine. The facility is located ~2 km underground. The ~2 km of rock overburden provides shielding from cosmic-ray secondaries equivalent to 6010 meters of water [33] (meter water equivalent or mwe is actually a standard unit for cosmic ray attenuation).

2.2.2 Cryogenic and Shielding Assembly

The SuperCDMS detectors will be stacked vertically in groups of six. All detector stacks will be cooled and maintained at 15 mK using a dilution refrigerator (DR). Additional thermal stages will be mounted above each detector stack. These stages are thermally isolated and will be maintained at 250 mK, 1 K, and 1 K using the same DR. The latter 1-K stage was originally intended to be held at 4 K but was changed in later designs. The stages support and thermally sink the cryogenic readout electronics including cables going between the detectors and significantly warmer ($\geq$50 K) stages. The combination of detector stack, low-temperature thermal stages, and mounted electronics is referred to as a detector tower.

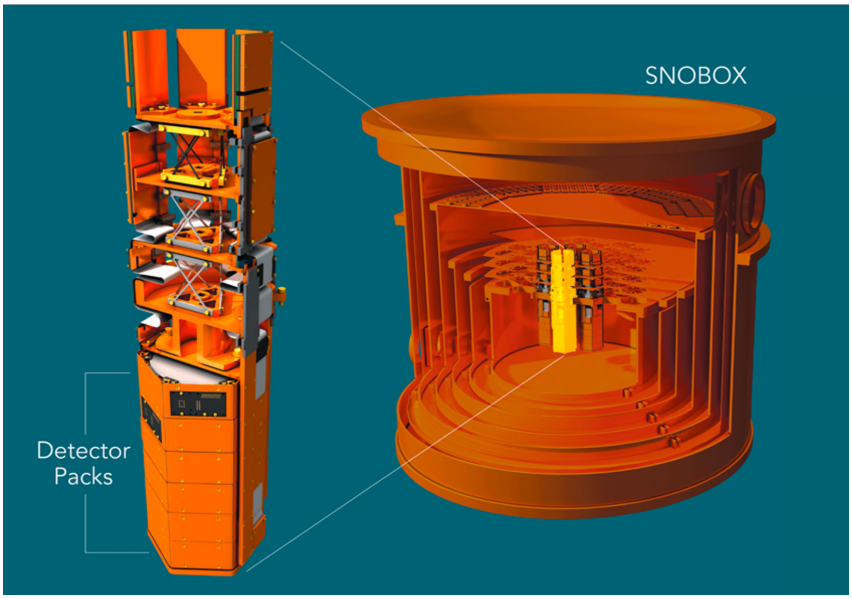


Figure 2.6: A rendering of the SuperCDMS SNOLAB experimental payload. (Left) A rendering of one detector tower with 6 detectors in its detector stack/pack. The x-shaped braces separate different thermal stages. (Right) A rendering of the SNOBOX with four detector towers installed. Renderings by Greg Stewart/SLAC National Accelerator Laboratory.

The initial SuperCDMS payload will include four detector towers (two iZIP and two HV). The towers will be mounted in seven nested Cu cans of various temperatures. The innermost can is connected to the detector stage and maintained at the same

temperature (15 mK). The next three cans connect to the higher detector-tower stages and are maintained at the same temperatures (250 mK, 1K, and 1K again). The next two cans do not directly connect to the towers and are maintained at 50 K and ~250 K. The outermost can maintains vacuum within the entire assembly and will sit at room temperature (~300 K). The entire can assembly is referred to as the SNOBOX. Apart from the outer vacuum can (OVC), each can is suspended by Kelvar rope with low thermal conductivity. The SNOBOX is designed to minimize vibrational noise at the detectors as well as to keep its contents safe during a seismic event. Each (non-OVC) can also shields its content from the higher-temperature black-body radiation of the cans outside it.

Surrounding the SNOBOX are several layers of shielding. The outermost shield is made of layers of polyethylene and polyethylene tanks filled with water. This shield is designed to protect from neutron backgrounds originating from radioactivity in the walls of the SNOLAB cavern. Inside the outer neutron shield will be an aluminum structure filled with boil-off nitrogen. Pure boil-off nitrogen is free of radon and will protect the assembly from collecting atmospheric radon. Otherwise, the decay of prompt radon daughters may present a considerable background. Inside this radon-purge layer will be a thick (23 cm) layer of low-activity lead to shield from external gamma radiation. Inside the lead shield will be another 40 cm polyethylene shield to provide further protection from neutrons including radiogenic neutrons from the lead shield. Inside the inner neutron shield will be a mu-metal magnetic shield used to limit the effect of Earth's magnetic field on the phonon sensors and readout electronics. The SNOBOX sits within this magnetic shield.

The dilution refrigerator sits outside the shielding apparatus with its thermal stages connected to those of the SNOBOX through a metal tube known as the C-stem (cryogenics stem). Another tube (the E-stem or electronics stem) connects the SNOBOX electronics to a separate E-tank (electronics tank) outside the shielding. The E-tank uses vacuum feedthroughs to connect the SNOBOX electronics to the rest of the lab. Both stems create holes in the shielding. The stems, the DR, and the E-tank share the same vacuum as the OVC.

2.3 SNOLAB Backgrounds and Projected Sensitivity

The projected sensitivity of the SuperCDMS SNOLAB experiment was calculated in [33]. I will go over some of those results here for reference. Sensitivity is presented in the form of 90%-confidence exclusion limits calculated using the expected detector

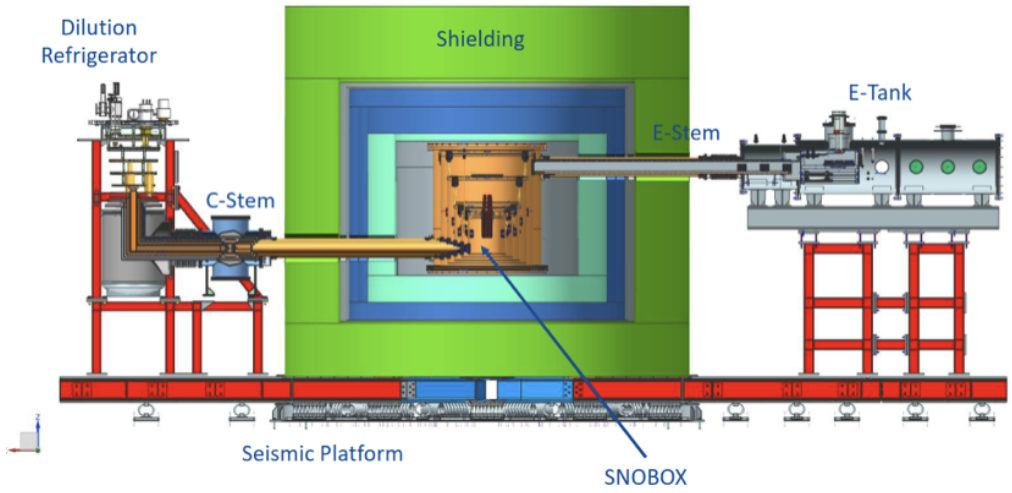


Figure 2.7: A diagram of the SuperCDMS SNOLAB experimental setup. The dilution refrigerator cools the SNOBOX stages through the C-stem. The SNOBOX holds the detector towers and is surrounded by various layers of shielding. The shielding consists of an inner neutron shield, a lead gamma shield, a radon purge barrier, and an outer neutron shield. The SNOBOX electronics are connected through the E-stem to the E-tank. The E-tank uses vacuum feedthroughs to connect the SNOBOX electronics to the rest of the lab. The complete assembly sits on a seismic platform to protect from seismic events. Diagram from [33].

responsivity, backgrounds, and analysis cuts. I will begin by discussing the relevant backgrounds in the SNOLAB experiment (Section 2.3.1). The analysis cuts applied to those backgrounds include removing surface events and electron recoils using position information and ionization yield, respectively. I will then present the sensitivity projections for multiple dark-matter models (Section 2.3.2).

2.3.1 Background Sources

Detector backgrounds can be subdivided by a number of factors. Some backgrounds produce electron-recoil (ER) events while others produce nuclear-recoil (NR) events. Some backgrounds produce events primarily near a detector's surfaces while others produce events throughout the detector (the latter are referred to as "bulk" backgrounds). These distinctions determine the SuperCDMS detectors' ability to reject each background (as discussed above). There are also backgrounds specific to germanium or silicon detector crystals.

Radioactive impurities within the detector crystals are one source of bulk backgrounds. Bulk backgrounds tend to be dominant for HV detectors, which are

capable of radial-surface discrimination but not of bulk electron-recoil removal. Impurities include cosmogenically produced tritium (^{3}H). Cosmic-ray secondaries interact with crystal nuclei (Si or Ge) to produce spallation products including tritium. Tritium β-decay leads to an 18.6 keV endpoint ER background. Another β-decay producing impurity is ^{32}Si, which naturally appears in Si crystals. The concentration is highly dependent on the environment in which the Si originated. Cosmic-ray spallation in Ge produces isotopes that decay via electron capture. The combination of different isotopes and different capture shells creates a total of 24 background spectral peaks in Ge.

Cosmic-ray interactions also produce trace radioactive isotopes in the SNOLAB housing materials. Specifically, the copper that makes up the detector housings and the SNOBOX can be cosmogenically activated. Housing materials can also accumulate radioactive isotopes naturally or during manufacturing. Relevant isotopes include ^{238}U, ^{232}Th, ^{40}K, and ^{60}Co. Of these, ^{238}U and ^{232}Th have radioactive daughters that contribute additional backgrounds. ^{238}U, ^{232}Th, and ^{40}K are also present in atmospheric dust and may accumulate on housing material surfaces. Radon is also present in the atmosphere. Specifically, ^{222}Rn can decay in air and deposit its daughter particles in the surface of materials. The ^{222}Rn daughters quickly decay to ^{210}Pb, which decays (over a long time) in the material to produce αs, βs, and X-rays. Housing backgrounds are generally subdivided into line-of-sight and non-line-of-sight backgrounds. For non-line-of-sight backgrounds, only penetrating radiation will reach the detector and will appear as bulk NR or ER backgrounds. Line-of-sight backgrounds may produce bulk or surface backgrounds depending on the radiation type and energy.

^{222}Rn daughters can also be implanted in detector surfaces. ^{210}Po (a daughter of implanted ^{210}Pb) decays to ^{206}Pb with enough energy to produce a nuclear recoil with comparable energy to dark-matter interactions.

The SNOLAB facilities are surrounded by norite rock covered in a layer of shotcrete. This cavern environment is expected to produce both gamma and neutron radiation. After shielding, the cavern gammas are expected to be subdominant and can be ignored. Cavern neutrons will produce additional bulk NRs.

The SNOLAB overburden and shielding will remove the majority of cosmic-ray events, but some cosmic rays and spallation products will still reach the detectors. Spallation neutrons producing bulk NRs are expected to be the most significant of these.

Scattering between solar neutrinos and nuclei (CEνNS) will not be a dominant background for the initial SNOLAB experiment. Such events will produce bulk NRs and are still considered in [33].

The expected background rates without analysis cuts applied are shown in the tops of Figures 2.9, 2.10, 2.12, 2.14, and 2.17. iZIP and HV detectors of the same semiconductor material share the same pre-cut backgrounds, but the backgrounds appear with different phonon energies due to the difference in ΔV. The different detector types also have different thresholds and background rejection ability, which can be seen in the same Figures.

2.3.2 Projected Sensitivity

Sensitivity limits were calculated separately for several dark-matter models. The models include nuclear-coupled, electron-coupled (through a light mediator), electron-coupled (through a heavy mediator), dark-photon, and axion-like-particle dark matter. The various sensitivity limits were calculated assuming an exposure equivalent to 4 years of data taking with 80% live time (the fraction of data-taking time actually recorded to data). The resultant sensitivity plots are shown in Figures 2.8, 2.11, 2.13, 2.15, and 2.16. The pre- and post-cut background spectra used to calculate these sensitivities are shown in Figures 2.9, 2.10, 2.12, 2.14, and 2.17.

Sensitivity to spin-independent nucleon-coupled dark matter is shown in Figure 2.8. Si detectors are sensitive to lower masses than their Ge counterparts because the Si nucleus has a lower mass. HV detectors are sensitive to lower masses, because their higher ΔV creates a lower recoil-energy threshold. For higher masses, the iZIPs' ability to remove backgrounds gives them better sensitivity. Although the Si iZIPs do not have the best sensitivity for any masses, they will provide useful information on the Si-HV backgrounds.

For nuclear-recoil dark matter, sensitivity was calculated using both the optimum interval method and using a profile-likelihood ratio. The latter utilizes information about known background spectra and is therefore capable of improving sensitivity. For the subsequent sensitivity calculations, only the profile-likelihood-ratio result is presented.

The post-cut backgrounds used to produce Figure 2.8 are shown in Figures 2.9 and 2.10. We can see the effect of the iZIP detectors' ER rejection. Both surface- and bulk-ER rates drop at high phonon energy for iZIP detectors. For lower energies, noise in the ionization measurement prevents accurate calculation of ionization

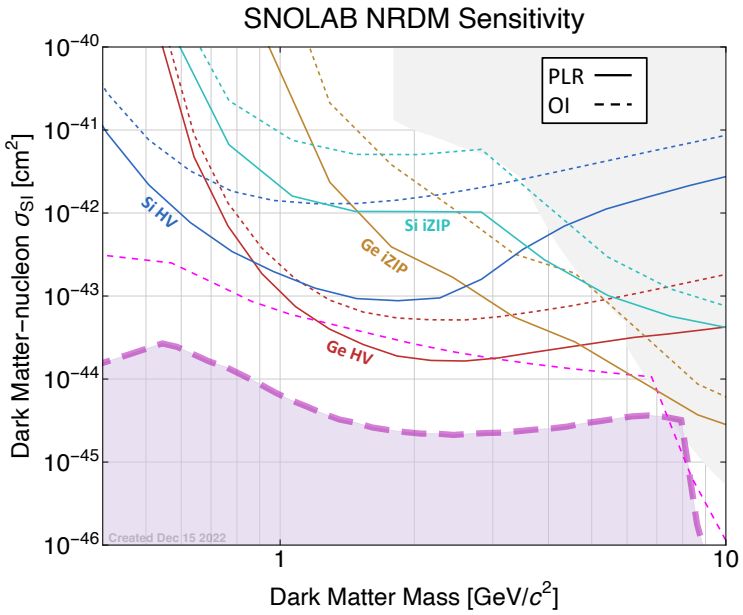


Figure 2.8: Projected SNOLAB sensitivity to spin-independent nucleon-coupled dark matter in the form of 90%-confidence exclusion limits. Produced by [33]. The pre- and post-cut backgrounds used to calculate the iZIP (HV) limits are shown in Figure 2.9 (2.10). Solid-line limits are calculated using a profile-likelihood ratio that utilizes information about background spectra. Dashed-line limits are calculated using the optimum interval method. The short-dashed pink line represents the single-neutrino sensitivity. An experiment with that sensitivity would see an average of one neutrino event over the course of taking data [35]. The long-dashed purple line and purple shaded region indicate the neutrino fog. In this mass range, the neutrino fog is dominated by solar fusion of ^{7}Be and β-decay of ^{8}B. The currently excluded parameter space is shaded in light grey.

yield. Surface rejection lowers the rate of surface NRs in both iZIP and HV detectors. Bulk NRs from neutron and neutrino interactions become significant for very low energies, but such energies are below the detector thresholds and do not impact the sensitivity.

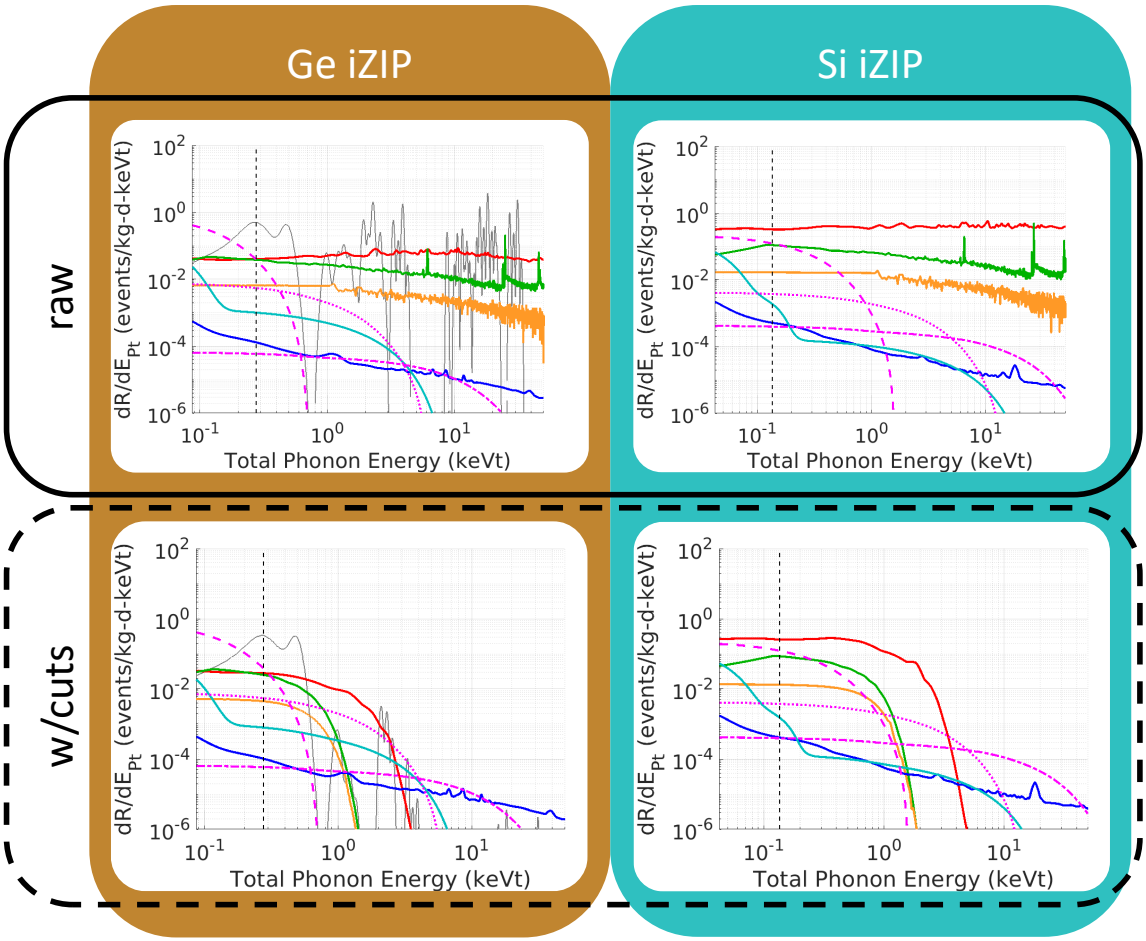


Figure 2.9: The expected backgrounds (in total phonon energy) used to calculate the iZIP nuclear-recoil sensitivities in Figure 2.8. (Top) The pre-cut spectra. (Bottom) The post-cut spectra. Produced by [33]. The vertical dashed lines indicate the estimated analysis thresholds (estimated to be seven times the phonon-energy resolution). Red backgrounds come from bulk ER events including Compton scattering and ^{3}H/^{32}Si β-decay. Green backgrounds come from surface ERs. Mustard backgrounds come from surface NR events including ^{222}Rn-descendent ^{206}Pb recoils. Blue backgrounds originate from neutron-induced bulk NRs. Cyan backgrounds originate from CEνNS. Grey backgrounds in Ge are spectral peaks due to cosmogenic activation. Example candidate signal models are shown in pink. Each example model is set to have equivalent cross section to the projected sensitivity. iZIP example candidates have masses of 1.6, 5, and 16 GeV. Higher-mass candidates produce higher-maximum energy.

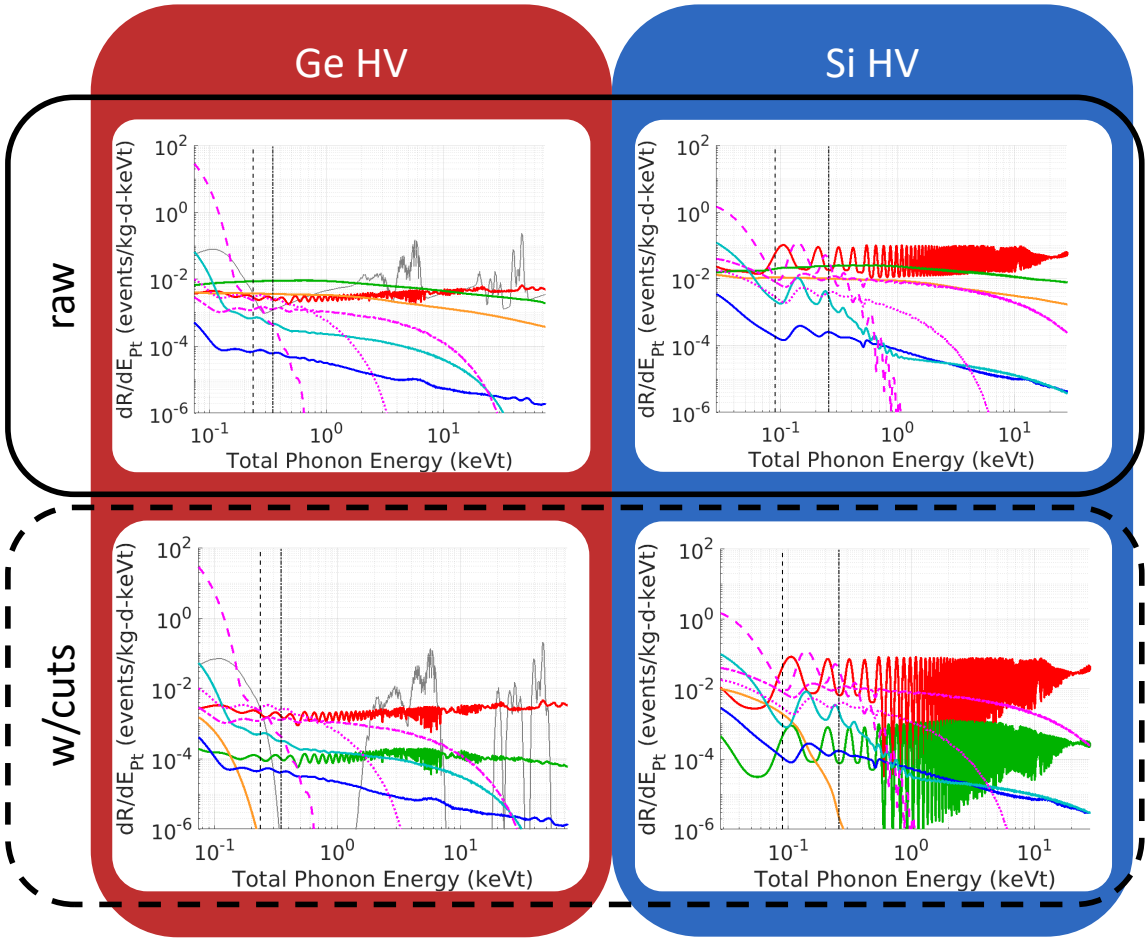


Figure 2.10: The expected backgrounds (in total phonon energy) used to calculate the HV nuclear-recoil sensitivities in Figure 2.8. The HV-style detectors use ΔV large enough to observe the quantization of the second term in Eq. 2.1. This quantization is the origin of the spectral peaks particularly noticeable in the ER-recoil spectra. (Top) The pre-cut spectra. (Bottom) The post-cut spectra. Produced by [33]. The vertical dot-dashed lines indicate the estimated degradation of HV-detector thresholds due to voltage-induced leakage events. Otherwise, uses the same legend as Figure 2.9. Example candidates have masses of 0.5, 1.6, and 5 GeV with higher-mass candidates producing higher-maximum energy.

The projected sensitivities to electron-coupled dark matter are shown in Figures 2.11 and 2.13 for heavy and light mediators, respectively. Projected limits are only shown using HV detectors because their lower threshold makes them superior in this mass region. The iZIP detectors' ability to reject ERs also serves no purpose for ER-search events.

The pre- and post-cut backgrounds used to produce Figures 2.11 and 2.13 are shown in Figures 2.12 and 2.14, respectively.

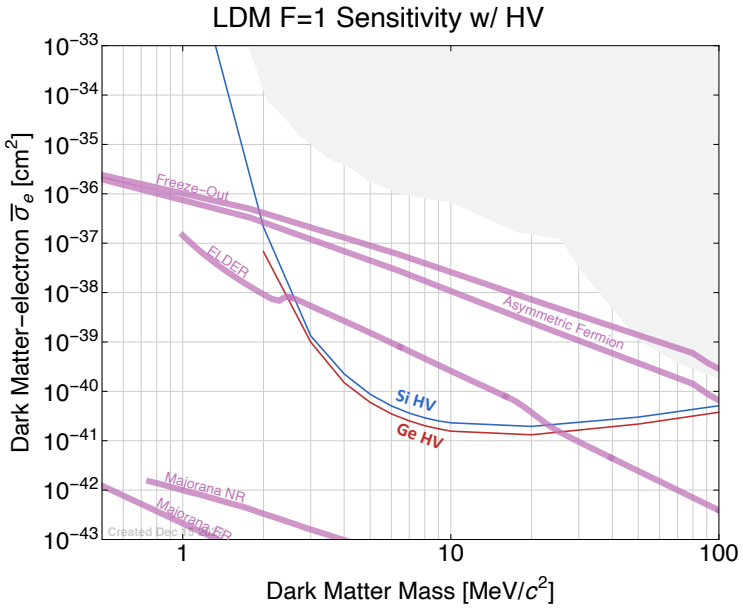


Figure 2.11: Projected SNOLAB sensitivity to electron-scattering dark matter with a heavy mediator (momentum-transfer form factor = 1). Produced by [33]. The pre- and post-cut backgrounds used to calculate the limits are shown in Figure 2.12. The currently excluded parameter space is shaded in light grey. Parameter space for some candidate models are shown in magenta.

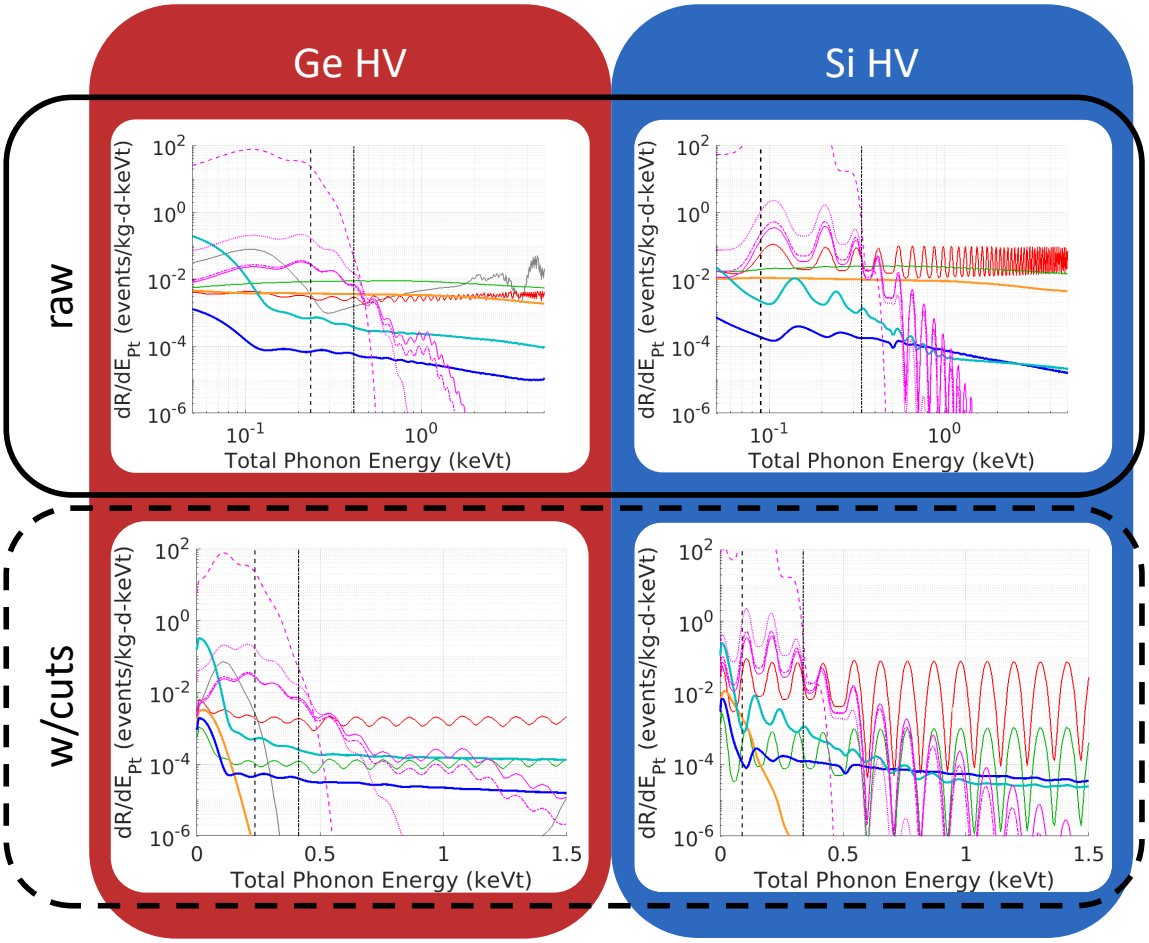


Figure 2.12: The expected backgrounds (in total phonon energy) used to calculate the electron-recoil sensitivity in Figure 2.11. (Top) The pre-cut spectra. (Bottom) The post-cut spectra. Produced by [33]. Uses the same legend as Figure 2.10. Example candidates have masses of 1, 3, 10, and 30 MeV with higher-mass candidates producing higher-maximum energy.

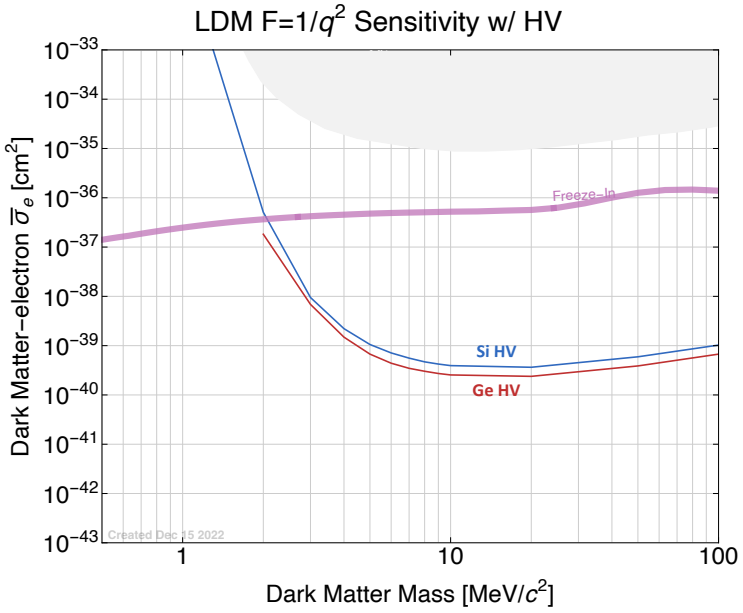


Figure 2.13: Projected SNOLAB sensitivity to electron-scattering dark matter with a light mediator (momentum-transfer form factor = $(\alpha m_e/q)^2$). Produced by [33]. The pre- and post-cut backgrounds used to calculate the limits are shown in Figure 2.14. The currently excluded parameter space is shaded in light grey. Parameter space for freeze-in dark matter is shown in magenta.

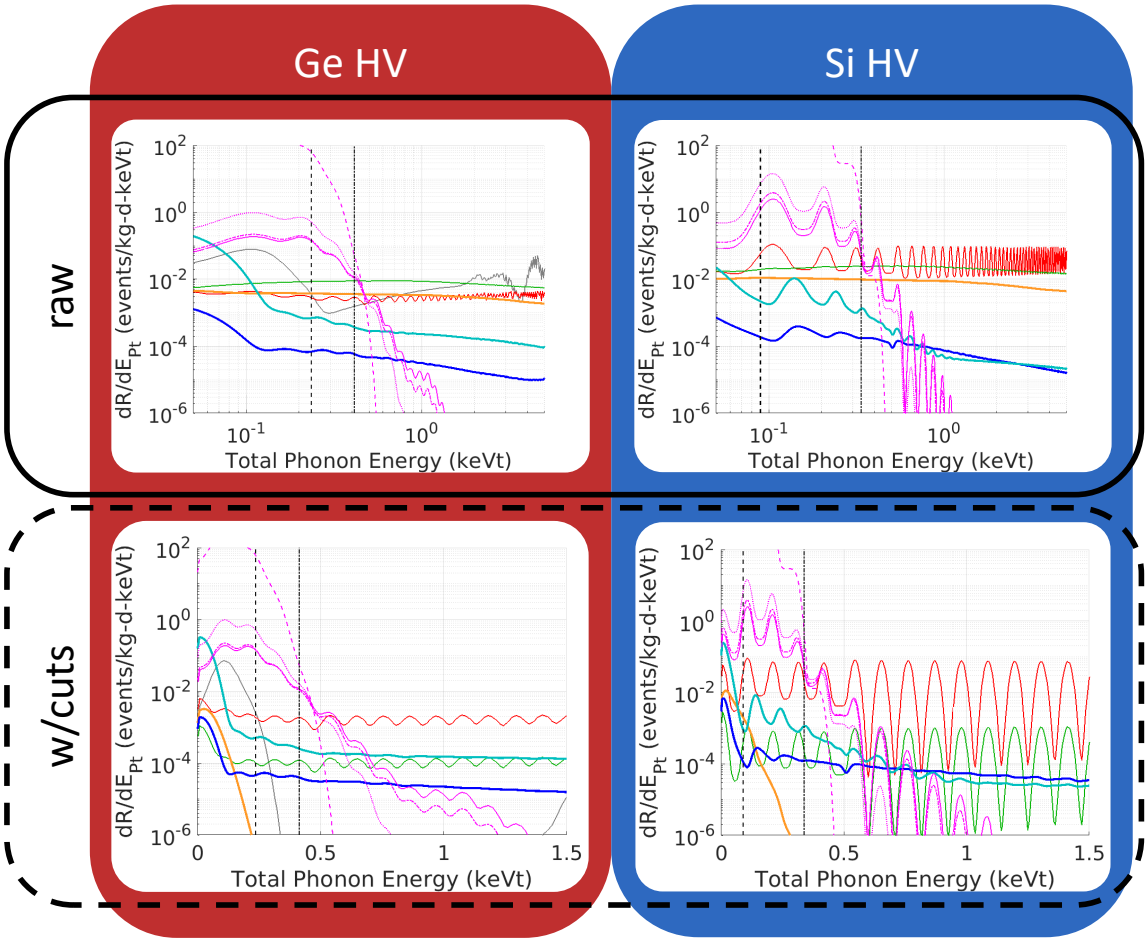


Figure 2.14: The expected backgrounds (in total phonon energy) used to calculate the electron-recoil sensitivity in Figure 2.13. (Top) The pre-cut spectra. (Bottom) The post-cut spectra. Produced by [33]. Uses the same legend as Figure 2.10. Example candidates have masses of 1, 3, 10, and 30 MeV with higher-mass candidates producing higher-maximum energy.

The projected sensitivities to dark-photon and axion-like-particle absorption are shown in Figures 2.15 and 2.16, respectively. The entire rest energy of either particle would be deposited as electron-recoil energy. Therefore, the projected limits extend down to the energy threshold of each detector. The HV detectors are superior in this range. Both absorption processes share the same backgrounds (shown in Figure 2.17).

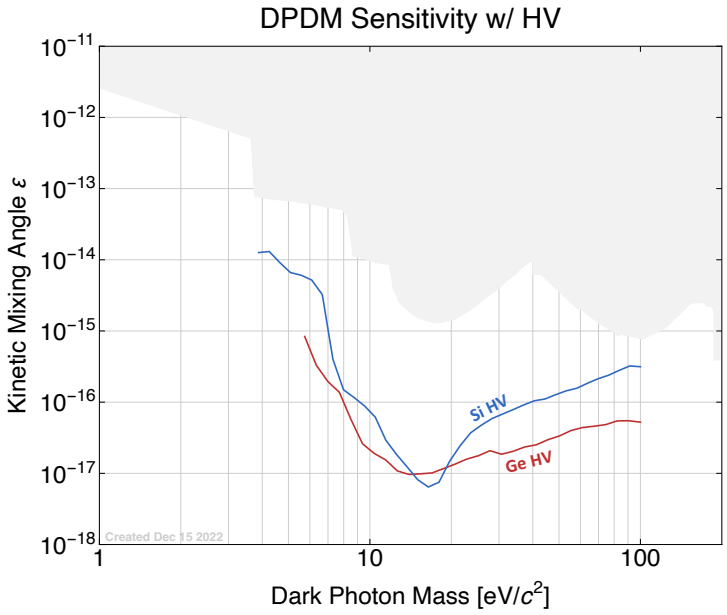


Figure 2.15: Projected SNOLAB sensitivity to dark-photon absorption. Produced by [33]. The pre- and post-cut backgrounds used to calculate the limits are shown in Figure 2.17. The currently excluded parameter space is shaded in light grey.

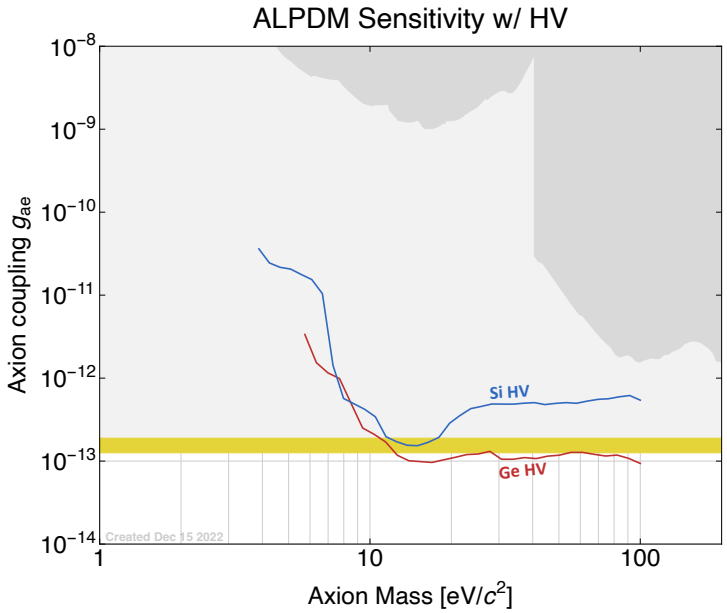


Figure 2.16: Projected SNOLAB sensitivity to axion-like-particle (ALP) absorption. Produced by [33]. The pre- and post-cut backgrounds used to calculate the limits are shown in Figure 2.17. The parameter space excluded by direct-detection experiments is shaded in dark grey. The parameter space excluded by stellar-cooling observations is shaded in light grey. The bottom of the latter region is marked in yellow and represents the best estimate based on stellar-cooling observations [36]. If ALPs do exist at the best-estimate level, the SNOLAB HV detectors should be able to detect them.

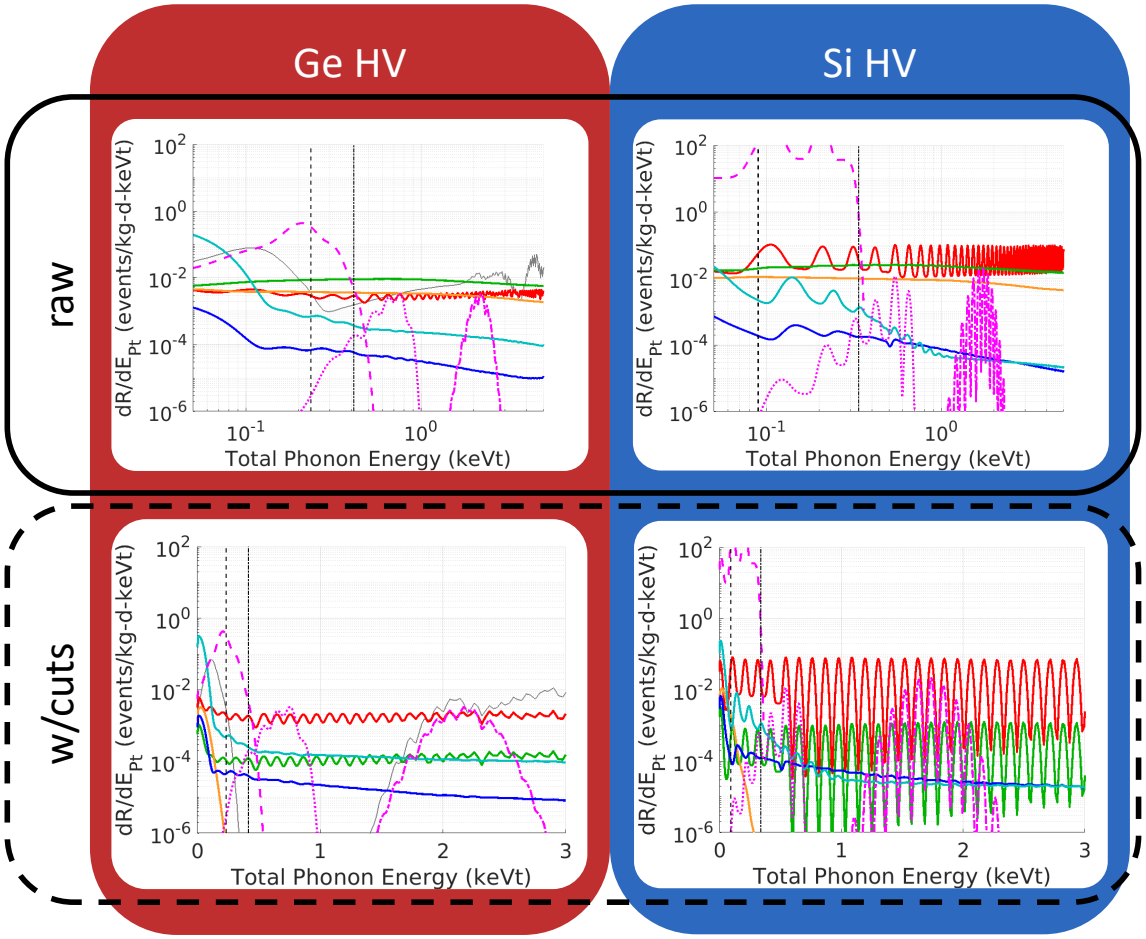


Figure 2.17: The expected backgrounds (in total phonon energy) used to calculate the electron-recoil (via particle absorption) sensitivities in Figures 2.15 and 2.16. (Top) The pre-cut spectra. (Bottom) The post-cut spectra. Produced by [33]. Uses the same legend as Figure 2.10. Example candidates have masses of 6, 20, and 60 eV with higher-mass candidates producing higher-maximum energy.

2.4 iZIP Ionization Readout

The iZIP detectors measure ionization by collecting electrons and holes on surface electrodes. Collection is driven by the bias applied to the electrodes and the resulting electric field (see Section 2.1.2). Each electrode channel can be modeled as a capacitance to ground. When charge is collected by a channel's electrodes, it appears on the equivalent capacitor and is measured and amplified using the iZIP ionization readout. Each iZIP requires four independent ionization readouts (one for each of the four ionization channels).

The SNOLAB experiment uses charge (current-integrating) amplifiers as the first-stage amplifier of each ionization readout. Each charge amplifier is split into a cryogenic input stage (mounted on the detector tower) and a room-temperature stage with the remaining components. We will discuss the purpose for the division in the following section. Before doing so, it will be useful to discuss the operation of a basic charge amplifier.

2.4.1 Basic Charge Amplifier Design

The diagram for a basic charge-amplifier is shown in Figure 2.18. In the case of sufficiently large open-loop gain (A_{ol}), these amplifiers convert charge to voltage with a conversion ratio defined by the inverse of the feedback capacitance. This can be proven with simple circuit analysis (Eqs. 2.3 – 2.6).

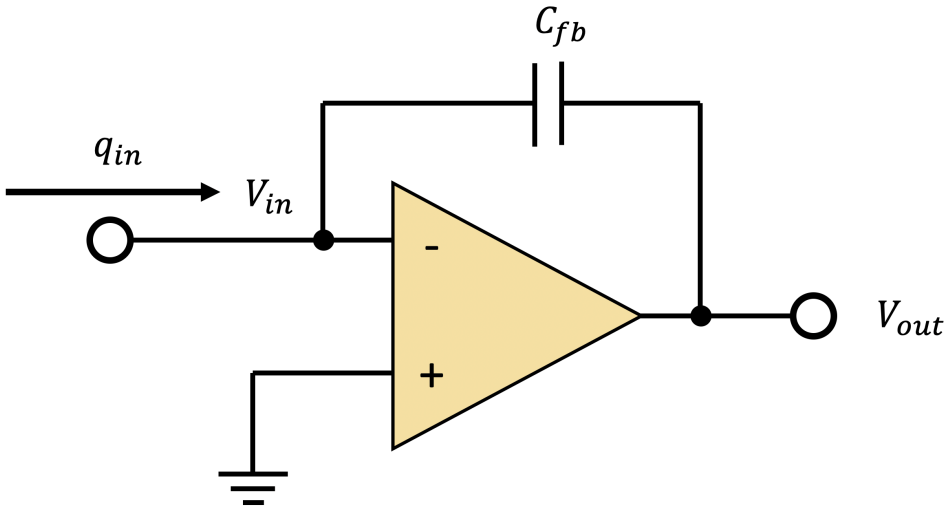


Figure 2.18: Block diagram for a basic charge amplifier.

First, we can relate the op-amp input and output voltages using the charge accumulated on the feedback capacitor. In this case, we will assume that the op-amp

input impedance is sufficiently large that all the input charge (q_{in}) accumulates on the feedback capacitor, and none is lost to the op-amp input or parasitic capacitance to ground.

$$\begin{aligned} V_{in} - V_{out} &= \frac{q_{fb}}{C_{fb}} \\ &= \frac{q_{in}}{C_{fb}} \end{aligned} \tag{2.3}$$

We can also relate the input and output voltages via the open-loop gain of the op amp.

$$\begin{aligned} V_{out} &= A_{ol}\,(V_+ - V_-) \\ V_{out} &= -A_{ol} V_{in} \end{aligned} \tag{2.4}$$

We can then combine Eqs. 2.3 and 2.4 to see that the output voltage is proportional to the input charge.

$$V_{out} = \frac{-1}{1 + \frac{1}{A_{ol}}} \frac{q_{in}}{C_{fb}} \tag{2.5}$$

In the scenario that $A_{ol} \gg 1$, the result simplifies and becomes constant with respect to small changes in A_{ol}.

$$\lim_{A_{ol} \to \infty} V_{out} = -\frac{q_{in}}{C_{fb}} \tag{2.6}$$

We now see that the conversion factor is equivalent to the negative inverse of the feedback capacitance. A charge amplifier's conversion factor is sometimes referred to as the circuit's "gain," although it is not a unitless ratio.

2.4.2 Charge Dissipation via Feedback Resistor

In practice, a high-value resistor is usually placed in parallel with the feedback capacitor. The feedback resistor provides a path for accumulated charge to dissipate. Without means of dissipation, static charge would accumulate until V_{out} reached the maximum op-amp output. In that state, the circuit would not be sensitive to additional charge. We can use the example of a δ-function deposition of charge on

the feedback capacitor (q_0) to calculate the dissipation rate. We start by equating the voltage across the capacitor and resistor.

$$\frac{q_{fb}}{C_{fb}} = -\frac{dq_{fb}}{dt}R_{fb} \tag{2.7}$$

We then solve the resulting ordinary differential equation.

$$\begin{aligned} \int_{q_0}^{q_{fb}(t)} \frac{dq_{fb}}{q_{fb}} &= \int_0^t -\frac{dt}{R_{fb}C_{fb}} \\ ln\left(\frac{q_{fb}(t)}{q_0}\right) &= -\frac{t}{R_{fb}C_{fb}} \\ q_{fb}(t \geq 0) &= q_0 e^{-t/R_{fb}C_{fb}} \end{aligned} \tag{2.8}$$

With the op-amp input held at virtual ground, we can use Eqs. 2.3 and 2.8 to calculate the output voltage.

$$V_{out}(t \geq 0) = -\frac{q_{fb}}{C_{fb}} = -\frac{q_0}{C_{fb}} e^{-t/R_{fb}C_{fb}} \tag{2.9}$$

The output response to an arbitrary input current can be calculated by analyzing the circuit in frequency space. In this case, we use the Fourier transforms of output voltage and signal current ($\tilde{V}_{out}$ and $\tilde{I}_{signal}$, respectively).

$$\begin{aligned} \tilde{V}_{out}(\omega) &= -\tilde{I}_{signal}(\omega) Z_{fb}(\omega) \\ &= -\tilde{I}_{signal} \frac{R_{fb}}{1 + j\omega R_{fb} C_{fb}} \end{aligned} \tag{2.10}$$

We can again use the example of a δ-function charge deposition.

$$I_{signal}(t) = q_0 \delta(t) \rightarrow \tilde{I}_{signal}(\omega) = q_0 \tag{2.11}$$

We can plug Eq. 2.11 into Eq. 2.10 and Fourier transform back[1] to find the same result as 2.9.

[1] In this case, the reverse Fourier transform requires a basic contour integral easily solved using the residue theorem.

$$\tilde{V}_{out} = -\frac{q_0 R}{1 + j\omega R_{fb} C_{fb}} \tag{2.12}$$

$$V_{out} = -\frac{q_0}{C_{fb}} e^{-t/R_{fb}C_{fb}} \tag{2.13}$$

From either example, we see that the charge dissipates with time constant $\tau_{fb} = R_{fb}C_{fb}$. Charge amplifiers are also known as current integrators, because the output voltage is proportional to the charge accumulated from an input current integrated over a time-scale on the order of τ_{fb}. The SNOLAB ionization readout will use R_{fb} = 100 MΩ and C_{fb} = 1 pF resulting in τ_{fb} = 100 µs.

For ionization measurements, the SNOLAB experiment will use a longer integration time than SuperCDMS Soudan. The measurement integration time is extended by constructing a larger τ_{fb} and by sampling the signal for a longer period of time. Longer integration improves resolution in charge amplifiers if the low-frequency noise introduced via the longer sampling period is sufficiently low. Therefore, minimizing the low-frequency (including $1/f$) noise is a priority in the ionization-readout design.

2.4.3 Amplifier Noise and Mitigation

Op amps like the one used in Figure 2.18 have a variety of internal noise sources producing voltage noise (including Johnson-Nyquist noise) and current noise (including shot noise). The combined noises can be represented using one voltage source and one current source at the op-amp input. Current noise at the op-amp input will be integrated and amplified with the same conversion factor as the charge signal. The resulting output voltage noise can only be lowered by lowering the conversion factor and therefore lowering the output signal by the same fraction. Since current noise cannot be mitigated with circuit design, it is best to pick an op amp with low current noise.

Conversely, voltage noise at the op-amp input will be converted to output voltage noise via a separate conversion factor. This factor is dependent on the circuit's input capacitance to ground. We can consider the scenario shown in Figure 2.19 with all input capacitance to ground described by one equivalent capacitor (C_{in}).

The noise voltage (V_N) will draw charge onto the input capacitor according to Eq. 2.14.

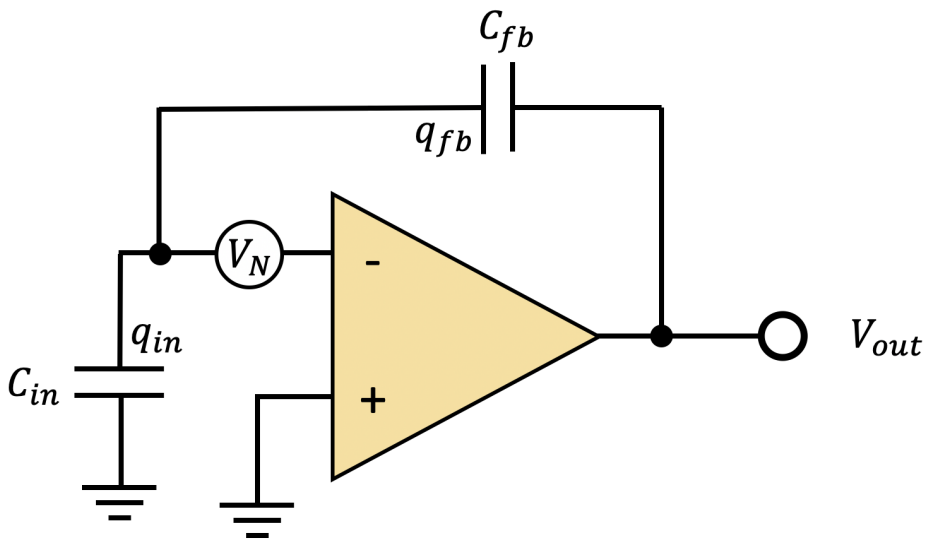


Figure 2.19: Block diagram for a basic charge amplifier with all input capacitance to ground described by one equivalent capacitor.

$$V_N = \frac{q_{in}}{C_{in}} \tag{2.14}$$

With no other source of charge present, an equal and opposite charge must accumulate on the feedback capacitor.

$$q_{fb} = -q_{in} \tag{2.15}$$

We can then use Eq. 2.3 to calculate the resulting output voltage.

$$\begin{aligned} V_N - V_{out} &= \frac{q_{fb}}{C_{fb}} \\ &= -\frac{q_{in}}{C_{fb}} \\ &= -V_N \frac{C_{in}}{C_{fb}} \end{aligned} \tag{2.16}$$

$$V_{out} = V_N \left(1 + \frac{C_{in}}{C_{fb}}\right) \tag{2.17}$$

Essentially, the input and feedback capacitors create a voltage divider with V_N as the central voltage. Eq. 2.17 can even be generalized by replacing C_{in} and C_{fb}

with Z_{in}^{-1} and Z_{fb}^{-1}, respectively (Eq. 2.18). This formulation makes it clear that Eq. 2.17 is only true for higher frequency noise (where capacitance dominates the input and feedback impedances). For lower frequency noise, $Z_{fb}/Z_{in} = R_{fb}/R_{in}$. With a high input impedance ($R_{in} \gg R_{fb}$), the voltage noise will not be amplified at all ($V_{out} \rightarrow V_N$). In the higher frequency region, we can minimize the voltage-noise amplification by minimizing the C_{in}. In practice, C_{in} will consist of the detector capacitance added to the signal-carrying wire capacitance, which increases with wire length.

$$V_{out} = V_N \left(1 + \frac{Z_{fb}}{Z_{in}}\right) \tag{2.18}$$

To minimize wire length and capacitance, the SNOLAB circuit was designed with its first stage as close as practically possible to the detector output. This required replacing the typical op-amp input with a transistor-based first stage capable of operating at cryogenic temperatures. High electron-mobility transistors (HEMTs) were chosen for the transistor input, due to their functionality at low temperature and good noise behavior relative to the SuperCDMS Soudan JFETs [37]. The HEMTs utilized also have very low power dissipation ($\sim 100\,\mu W$), which helps keep them sufficiently cold. The feedback capacitor and resistor must also sit at cryogenic temperatures. Having the feedback resistor at cryogenic temperatures has the additional benefit of lowering the resistor's Johnson-Nyquist noise. The circuit block-diagram for the SNOLAB design can be seen in Figure 2.20. The first stage components are visible in the dashed box labeled "4 K Stage," although this stage will actually be held at 1 Kelvin.

2.4.4 Current Mirror and HEMT Biasing

In order to achieve the high open-loop gain required to satisfy Eq. 2.6, the first-stage HEMT must have a sufficiently large transconductance (see Section 2.4.5). Transconductance is determined by the specific HEMT and its bias conditions (drain current, drain voltage, and gate voltage). For a given HEMT, setting any two conditions will require a specific value for the third. For the SNOLAB HEMTs and circuit design, we expect to achieve sufficient transconductance with 1 mA drain current and 100 mV drain voltage. These values will be the same for every ionization circuit, but slight differences between each circuit's HEMT will require a different gate voltage for each (see Figure 2.31). The gate voltages could be set by directly applying a DC voltage (via power supply or voltage divider), but such a method

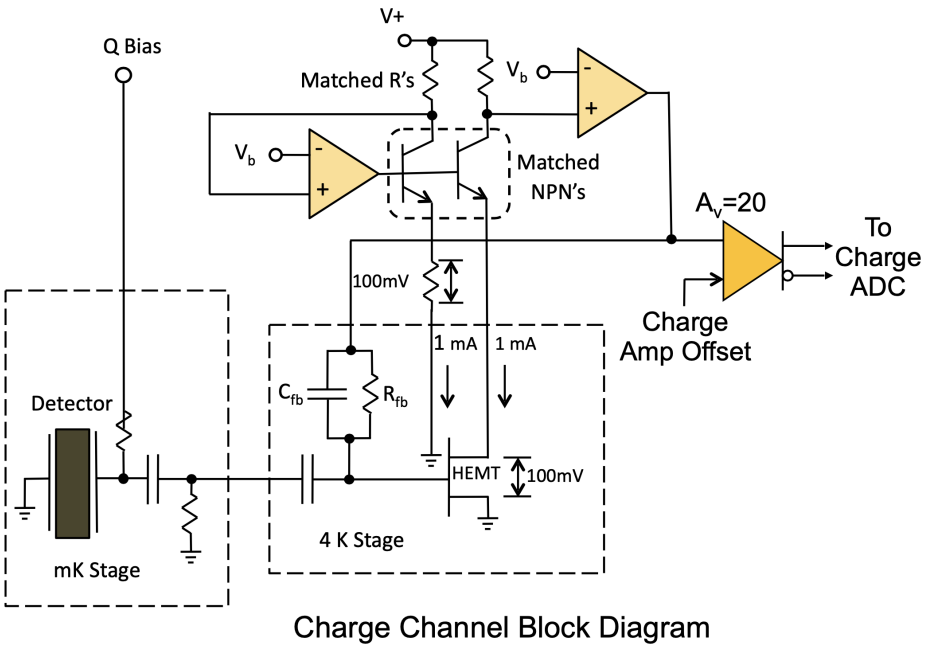


Figure 2.20: Block diagram for SuperCDMS SNOLAB's iZIP readout amplifier. A charge amplifier with a HEMT-based first stage kept at cryogenic temperatures ("4 K Stage" in the diagram, will actually be held at 1 Kelvin). A current mirror is used to bias the HEMT.

would require tuning each circuit according to its HEMT. Instead, the SNOLAB design uses an actively biased current mirror, which automatically adjusts the gate to reach the desired drain current and voltage.

In Figure 2.20), the pairs of matched NPN transistors and matched resistors, along with the left op amp, create the current mirror. By design, the left op amp uses negative feedback to keep its input-terminal voltages approximately equivalent. This results in a current of ~(V+ - V_b)/R flowing through the left matched resistor and into the left NPN's collector. The negative feedback can be understood using the following examples:

- If the op amp's positive-terminal voltage were $<V_b$, the op-amp output (and NPN base voltage) would be negative and the NPN would be put in the cut-off state. In cut-off, current would cease to flow into the collector (through the matched resistor), causing the positive-terminal voltage to increase until it was $>V_b$ (turning the NPN back on).

- If the op amp's positive-terminal voltage were $\gg V_b$, the op amp would output a large positive voltage and the NPN would move towards saturation. As the NPN approached saturation, greater current would flow into the collector (through the matched resistor), causing the positive-terminal voltage to drop. This would move the NPN away from saturation.

The only stable configuration is with the positive-terminal voltage slightly greater than V_b, leaving the NPN in its active state. When active, the emitter current is approximately equivalent to the collector current, depending on the precise transistor parameters (normally within 2%). Note that this design allows us to set the emitter current independent of the emitter resistance to ground. Choosing the emitter resistance then sets the emitter voltage.

The left emitter current and voltage are then mirrored on the right side of the diagram. This is guaranteed by the sameness of the matched NPN transistors (same design fabricated on the same substrate), which also receive equivalent base voltage, collector voltage, and collector current. Matching collector voltage and current are ensured by the matched resistors and the negative feedback of the right op amp. This feedback can be understood using the following examples:

- If the op amp's positive-terminal voltage were $>V_b$, the op-amp output (and HEMT gate voltage) would be positive. For the SNOLAB HEMTs, any positive gate voltage puts the HEMT into saturation. In saturation, a high current would flow through the HEMT drain, NPN emitter, and NPN collector, causing the positive-terminal voltage to decrease until it was $<V_b$. This would move the HEMT away from saturation.
- If the op amp's positive-terminal voltage were $\ll V_b$, the op amp would output a large negative voltage and the HEMT would be put in the cut-off state. In cut-off, current would cease to flow through the HEMT drain, NPN emitter, and NPN collector, causing the positive-terminal voltage to increase until the HEMT turned back on.

The only stable configuration is with the positive-terminal voltage slightly less than V_b, putting the HEMT into its linear regime. Note the difference from the left op amp's positive terminal, which will be held slightly above V_b. With this feedback, we can define the HEMT drain current and voltage using the mirror circuit, and the right op amp will set the appropriate gate voltage automatically.

This setup allows us to fully bias the HEMTs with minimum additional noise. Correlated noise between V+ and V_b will be partially removed via the op-amp subtraction. Such correlated noise is expected to be present since V_b is produced using V+ as the input to a voltage divider. The circuit is also robust to variation in ground voltage since the HEMT drain current and voltage are set using the current mirror rather than a direct reference to ground. There are actually two types of ground inside the cryostat. "Chassis ground" is directly connected to the metal bulk of the cryostat itself. "Analog ground" is connected (though conducting traces and wires) to ground on the warm electronics outside the cryostat. Both ground types are susceptible to noise from ground loops and other couplings. The SNOLAB ionization circuit is designed with the option to use either ground (depending on the placement of a jumper). The complete experiment will use whichever ground results in the lowest measured noise.

2.4.5 Amplification and HEMT Parameter Definitions

Let us briefly consider the SNOLAB amplification circuit if its negative feedback were disconnected. When charge signal arrived from the detector, it would generate a voltage change at the HEMT gate. This voltage would be converted to a change in current at the HEMT drain, with the conversion factor defined by the HEMT transconductance (g_m, the change in drain current due to a small change in gate voltage). The change in drain current would then change the right op amp's positive-input voltage, with the conversion factor defined by the matched resistor value ($R_{matched}$). With the op amp's feedback disconnected, its differential input would then be amplified by the op amp's open-loop gain (A_{opamp}) to produce a large output voltage. With the feedback connected, the circuit's total conversion factor is instead defined by the inverse of the feedback capacitance (akin to the simplified circuit in Figure 2.18). In fact, the closed-loop conversion from input charge to output voltage will be the same as Eq. 2.5 with the provision that A_{ol} is replaced with $g_m R_{matched} A_{opamp}$.

$$V_{out} = \frac{-1}{1 + \frac{1}{g_m R_{matched} A_{opamp}}} \frac{q_{in}}{C_{fb}} \tag{2.19}$$

Similarly to Eq. 2.6, we want $g_m R_{matched} A_{opamp}$ to be large in order to maximize the amplifier conversion factor and minimize the impact of $g_m R_{matched} A_{opamp}$ fluctuations. For this reason, we needed to ensure that all SNOLAB HEMTs have sufficient transconductance to satisfy our circuit design. The SNOLAB HEMTs are the same

design as the 100 pF HEMTs described in [37]. These HEMTs were measured (at 4.2 K) to have g_m = 35 millisiemens (mS) at 1 mA drain current and 100 mV drain voltage. The most recent SNOLAB circuit design (DCRC Rev-E) uses $R_{matched}$ = 1.6 kΩ. The op amps (part-number: LT1678) have minimum voltage gain of 0.3 V/μV. Combined with the HEMT g_m, these give a minimum open-loop voltage gain of 16.8 V/μV. When applied to equation 2.19, this implies the conversion factor will be within 1e-7 of the nominal value.

Similar to transconductance, drain conductance (g_d) is defined as the change in drain current due to a small change in drain voltage. Low g_d indicates that the HEMT is in or near its saturation region. This is where the SNOLAB HEMTs will operate, although there is not a specific constraint on g_d. A potential benefit to operating in saturation is that the drain current (which affects the output voltage) will be nearly insensitive to fluctuations in drain voltage. In [37], the 100 pF HEMTs were measured (at 4.2 K) to have g_d = 0.75 mS for the 100 mV and 1 mA bias point. We would like to know g_d for each HEMT in the SNOLAB experiment to ensure that none are behaving unexpectedly.

HEMTs also have a cutoff voltage (or threshold voltage, V_t). This is the minimum gate voltage that allows any drain current to flow through the HEMT. An ideal cutoff would be completely independent of drain voltage. In practice, there will always be a breakdown voltage of the transistor material, above which current will flow even in "cutoff." The cutoff voltage is not expected to have a significant effect on our amplification circuit, since we will bias our HEMTs well away from that region. However, it is useful to document the cutoff for future tests, as it provides a value below which you can reasonably expect no current through the HEMT, putting the HEMT in a safe state.

2.4.6 Thermal Considerations

As mentioned in Section 2.4.3, in order to minimize parasitic capacitance, the amplification circuit was designed with the first-stage components as near as practically possible to the detector. This means that the HEMTs, feedback resistors, and feedback capacitors for each detector will be placed on the 1-Kelvin stage of that detector's tower. This is the warmest thermal stage of each tower, where the components can be physically near to the detectors without applying a detrimental heat load. The tower model can be seen in Figure 2.21.

Notably, placing the feedback resistor at a lower temperature also lowers its Johnson-Nyquist noise. If not for the requirements outlined in Section 2.4.7, the resistor might have been placed on the detector packaging to lower its temperature further. The SuperCDMS Soudan experiment placed its feedback resistors at the detector stage (< 50 mK) for this purpose [38].

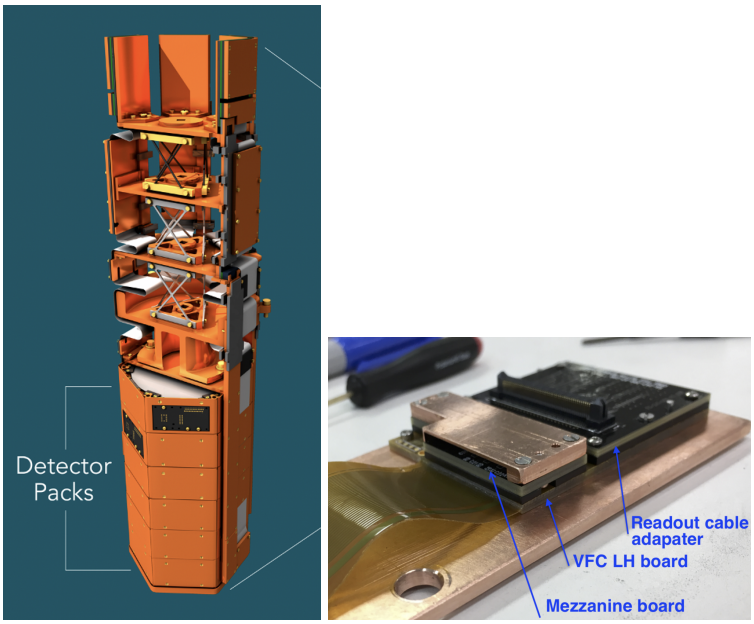


Figure 2.21: The HEMT mezzanine board and location on the detector tower. (Left) CAD model of a SNOLAB detector tower. HEMT mezzanine boards can be seen with copper covers near the top, which is the warmest thermal stage (1 Kelvin). (Right) Photo of one HEMT mezzanine board with its copper cover, attached to a detector tower's vertical flex cable on a copper storage plate.

What we now call the 1-Kelvin stage was initially designed to be 4 K. We redesigned the tower to thermally sink that stage to the cryostat still (at 1 K) in order to ensure the titanium (Ti 15-3-3-3 alloy[2]) wires that span the thermal stages of the tower are fully superconducting. It was found during early tests that the titanium had a transition temperature between 1 and 4 K, meaning the wires would have uncertain conductivity if mounted to a 4 K thermal stage. Simulations were used to confirm that the still cooling power will be sufficient to maintain the still and 1-Kelvin stage at 1 K. If expansions to the SNOLAB experiment require further heat load on the 1-Kelvin stage, NbTi cables with transition temperatures above 4 K could be used[3].

[2]Titanium alloy with 15% vanadium, 3% chromium, 3% tin, and 3% aluminum.

[3]NbTi wiring of sufficient robustness was not available in time to be used here. It can likely be developed if required.

This would allow us to revert the 1-Kelvin stage to 4 K without compromising conductivity.

The amplifier components on the 1-Kelvin stage are mounted on PCBs referred to as mezzanine boards. Each of these boards has 4 sets of components, to support the 4 ionization channels of a full SNOLAB iZIP detector (one inner and one outer channel per flat surface). Up to six mezzanine boards will be mounted on the 1-Kelvin stage of each tower, equal to the number of iZIP detectors installed. The mezzanine boards and their location can also be seen in Figure 2.21. The mezzanine PCBs are made of copper Cirlex, because it is less radioactive than FR4 PCB.

2.4.7 Vibration Considerations

We also considered microphonics while designing the circuit. Vibrations on the line between the detector and the HEMT create a varying line capacitance, which generates charge noise if the line carries non-zero voltage. This charge noise is equivalent to the capacitive noise multiplied by the DC voltage [39]. In order to minimize the effect, the majority of the line is grounded through a high-value resistor. Using a high value ensures that only static charge is removed while transient signals are still transmitted. Capacitors are placed (in series) on both ends of the grounded section to prevent the detector and HEMT voltage biases from being discharged. The remaining ungrounded portions are limited to the traces within the detector packages and those on the small mezzanine boards.

The presence of a capacitor separating the detector output and amplifier input does create the opportunity to lose some charge signal. Parasitic capacitance to ground on the grounded line will collect a fraction of charge, removing that fraction from the observable signal to be amplified. This further motivates keeping the first amplifier stage physically close to the detector, which minimizes the length of grounded line and resultant parasitic capacitance.

Wire motion could also cause noise via the triboelectric effect, which produces charge when different materials move while in physical contact or close proximity [39]. In this case, the concerning materials would be the line conductor and dielectric. In order to minimize this effect, the grounded line between detector and HEMT mostly consists of "vacuum coax," which has no direct contact with dielectric. Instead, the signal-carrying NbTi wires are surrounded by the vacuum maintained by the cryostat. The small sections of signal line which remain on circuit boards may still be susceptible to triboelectric noise. Vacuum coax also has much lower

capacitance to ground than a traditional coax (lowering the parasitic capacitance mentioned above).

2.5 HEMT Characterization Setup and Procedure

Using [37], we knew what to expect as the typical HEMT parameters. However, it was still important to characterize each HEMT going into the SNOLAB ionization readout individually. Characterization allows us to check for differences between the HEMTs tested in [37] and those produced for SNOLAB. If the SNOLAB HEMTs require different drain voltage or current to achieve sufficient transconductance, the biasing can be changed by altering the readout-circuit resistor values (seen in Figure 2.20). We can also identify and remove any HEMTs that do not work due to damage or fabrication issues.

2.5.1 HEMT Characterization Setup

For characterization, a Gifford-McMahon cryostat capable of reaching 4-Kelvin was commissioned at Caltech. The cryostat was designed and installed by Matt Hollister. Matt also designed and began installation of the readout wiring. I finished installing the readout wiring according to Matt's design. The HEMT parameter measurements in the following sections were all performed with the cryostat at its base temperature with no additional heating applied. In this state, the cryostat would typically stay between 4 and 6 K. We do not expect the HEMT characteristics to degrade significantly between this temperature and 1 K (where they will operate in SNOLAB). This expectation comes from [37], which shows the same HEMT design operating at both 4.2 and 77 K. In comparison, a change from 6 to 1 K seems minor.

The testing setup includes a pair of cables for connecting to a HEMT-testing motherboard designed at SLAC. The motherboard holds up to six HEMT mezzanine boards (two with shared readout traces, and therefore only five independently testable). During testing, the mezzanine boards were mounted with copper covers in-place for physical protection of the HEMTs. A testing motherboard with five mezzanine boards and their copper covers can be seen in Figure 2.22.

In order to protect the HEMT mezzanine boards and their copper covers (which will carry over to the SNOLAB experiment) from physical damage, and to minimize contamination by atmospheric radon and dust, I designed an aluminum motherboard case that encloses the HEMT-bearing portion of the motherboard while leaving the cable connections exposed. The case also acted as the thermal mount to securely sink the motherboard to our cryostat. The enclosed portion of the motherboard case

Figure 2.22: A photo of the HEMT-testing motherboard mounted in the bottom of its aluminum case. Five copper covers can be seen with HEMT mezzanine boards underneath. The two cable connectors can be seen on the top and bottom left.

included a meander cutout to allow gas to pump out before cooldowns while limiting radon ingress. The meander outlet was sealed with tape while outside the fridge. The case design can be seen in Figure 2.23.

Other protections included the use of gas from a high-pressure liquid nitrogen dewar to back-fill the cryostat after testing. This boil-off gas is very pure and effectively free of radon contamination. We also packaged each motherboard-assembly in two layers of heat-sealed nylon bag (filled with dewar nitrogen) for transport to and from Caltech. In total, this limits each motherboard's exposure to non-clean-room air to two specific time windows: the time between opening the bag and pumping out the cryostat, and the time between opening the cryostat and re-sealing with new nylon bags. In practice, I kept both time windows as brief as safely possible.

The basic readout circuit for testing the HEMTs can be seen in Figure 2.24. A Keithley 2220 power supply sets the gate-source voltage (V_{gs}), which is measured

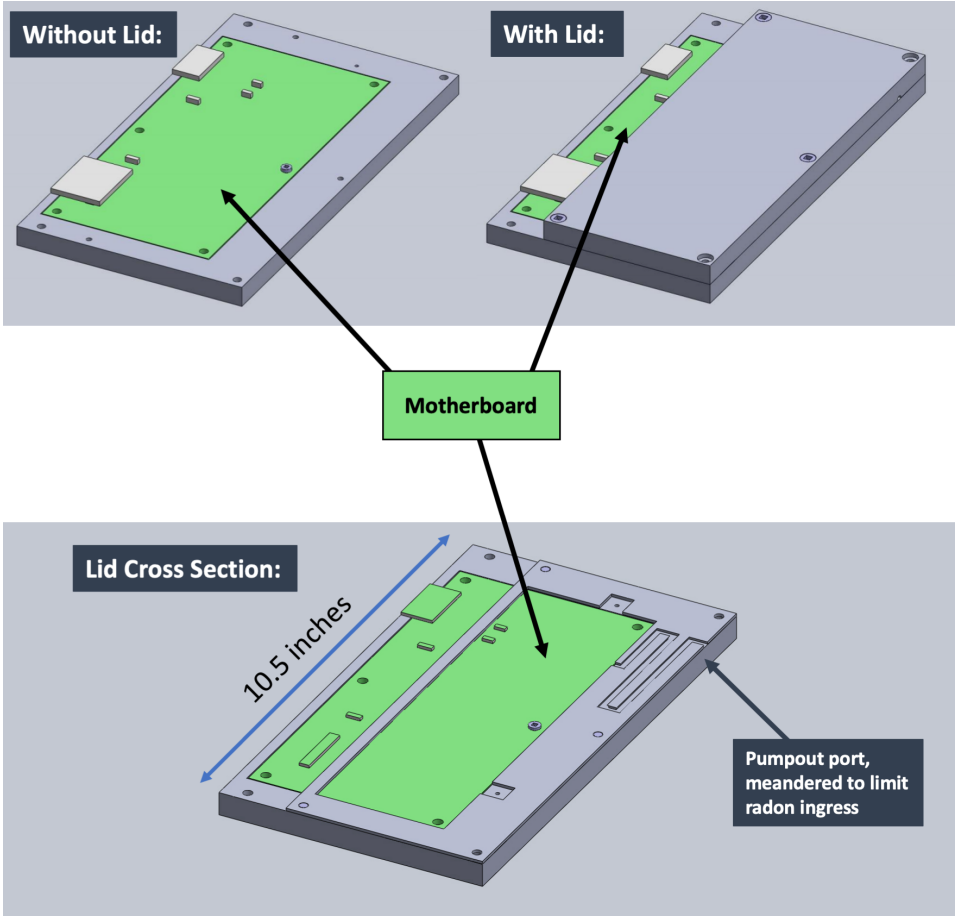


Figure 2.23: A CAD model of the HEMT-testing motherboard cover with a simple motherboard model. The lid cross section displays the meander used to release gas during cryostat pump-outs.

in parallel (outside the cryostat) by an Agilent 34401 multimeter. Setting and measuring V_{gs} outside the cryostat is accurate, because no current flows through the HEMT gate (as long as the gate-source voltage is below the breakdown voltage). The drain-source voltage (V_{ds}) is supplied and current (I_{ds}) is measured via a Keithley 2401 sourcemeter in a 4-wire measurement arrangement. The separate wire pairs connect on the motherboard. The return and negative-sense lines connect through the motherboard's ground plane, while the the supply and positive-sense lines connect through vias below the mezzanine-board fuzz-button pads. This wiring ensures that the only parasitic resistance in the measurement could come from traces on the small mezzanine boards themselves.

In order to test a fully loaded motherboard (five mezzanine boards of four HEMTs each) during a single cryostat cooldown, I connected the readout equipment to the

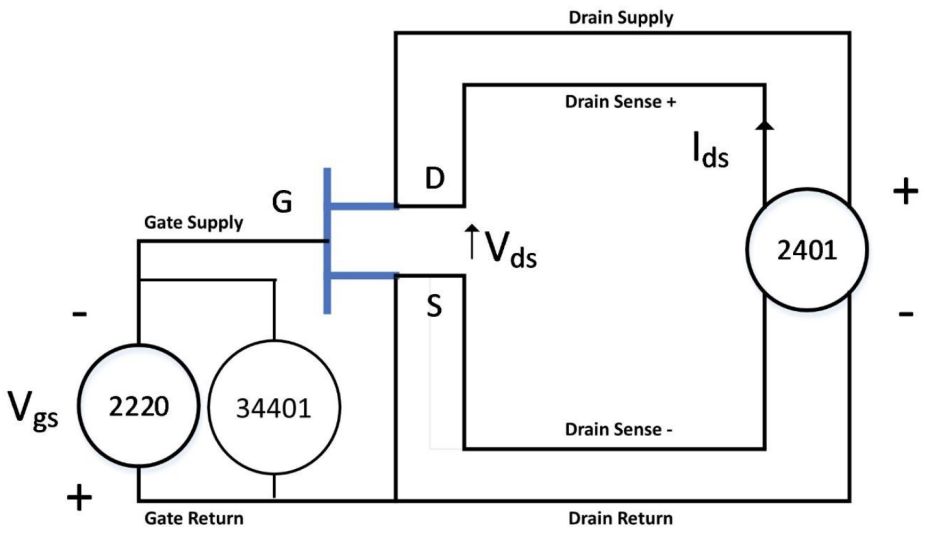


Figure 2.24: The basic HEMT testing circuit. The 2220 is a Keithley power supply which provides the gate-source voltage. The 34401 is an Agilent multimeter used to record the gate voltage. The 2401 is a Keithley sourcemeter which provides the drain voltage and measures the drain current via 4-wire measurement. This basic circuit is connected through a Keithley 7001 switch system, allowing one set of instruments to measure up to 20 HEMTs during one cooldown.

readout wiring (from the cryostat) via a Keithley 7001 switch system, which allows the user to connect the readout equipment to one HEMT at a time. During the cooldown process, all HEMT channels were connected simultaneously and shorted to the fridge chassis. This prevented the HEMTs from accumulating static charge due to the changing temperature. I tested one set of HEMTs without grounding during cooldown and found that the bias parameters changed noticeably. For this reason, I recommend that the SNOLAB HEMTs always be safely grounded during cooldowns to accurately reproduce their documented behavior. Details of this test will be discussed in Section 2.5.4.

2.5.2 Transconductance and IV Curves

I used this setup to measure g_m for all HEMTs allocated to the initial SNOLAB payload. Since the amplification circuit's bias points were not yet finalized, I performed the measurement for a matrix of candidate drain-current and drain-voltage values. Doing so required that I first establish what value of V_{gs} gives each desired (V_{ds}, I_{ds}) combination for a given HEMT. Specifically, I set V_{ds} to each of its candidate values and measured I_{ds} while increasing V_{gs} (from an initial large negative value). I stopped each measurement when I_{ds} exceeded the maximum

candidate I_{ds}[4]. I used the resulting data to identify the approximate V_{gs} required for each (V_{ds}, I_{ds}) combination. I then repeated the process for a finer array of V_{gs} data points around each combination's approximate V_{gs}. I calculated g_m for each bias point using the fine V_{gs} search data. To do so, I simply calculated the slope between the two data points closest to the desired I_{ds} value. Example data for one HEMT can be seen in Figure 2.25. I repeated the entire process for each HEMT separately, because different HEMTs require different V_{gs} for each (V_{ds}, I_{ds}) combination.

I_{ds} [mA]	g_m [mS] @ V_{ds} = 75 mV	g_m [mS] @ V_{ds} = 100 mV	g_m [mS] @ V_{ds} = 175 mV	g_m [mS] @ V_{ds} = 250 mV
0.5	32.315	35.605	32.073	35.809
1.0	50.705	50.949	52.753	45.85
2.0	72.971	82.479	79.677	81.05
3.0	91.163	108.248	99.144	97.054

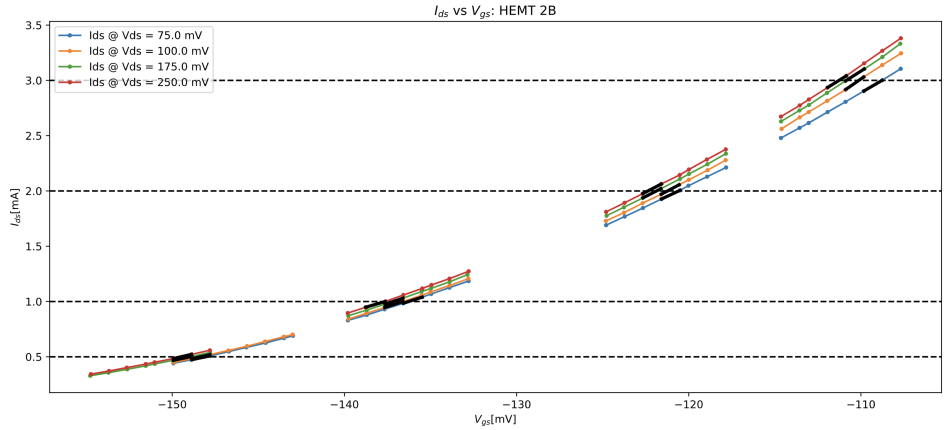
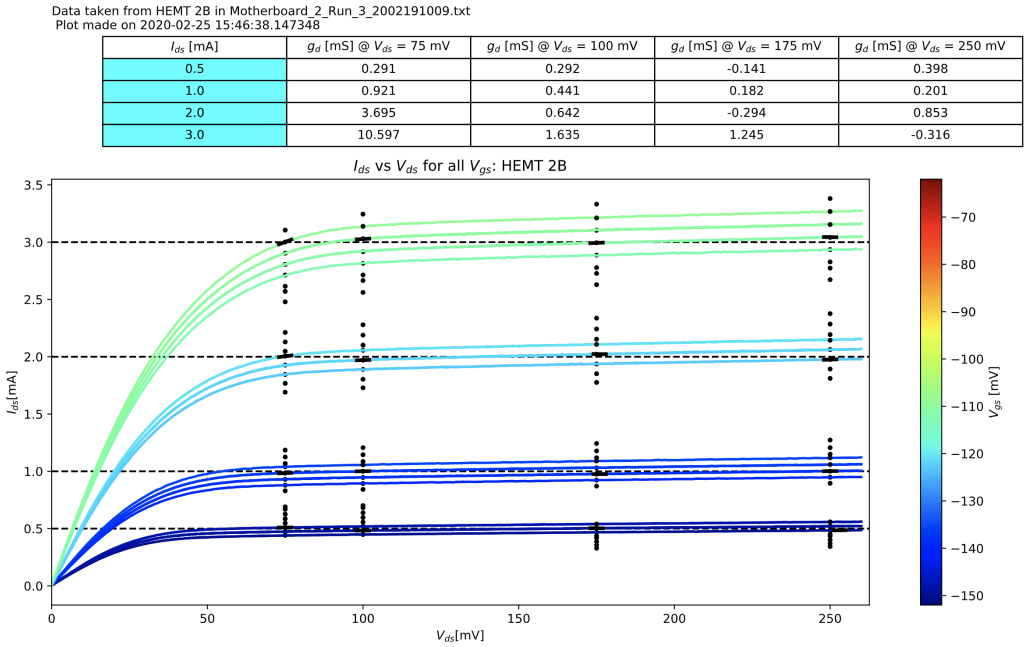


Figure 2.25: A typical HEMT transconductance measurement with results. (Top) Table of measured transconductance values for each candidate bias point. (Bottom) Plot of drain-current measurements for different drain and gate voltage combinations. Different colors denote different candidate drain voltages as described in the legend. Drain currents corresponding to potential bias points are noted with horizontal dashed lines. Black lines denote the data points used to calculate the transconductances for each bias point.

I also measured g_d for each HEMT and candidate bias point. Drain conductance measurements require measuring I_{ds} as a function of V_{ds}. Such measurements are known as IV curves (I and V standing for current and voltage, respectively). With the appropriate V_{gs} known (from the transconductance measurement) for each HEMT and bias point, I measured the IV curves by setting V_{gs}, sweeping over V_{ds}, and recording I_{ds}. The resulting plot for one HEMT can be seen in Figure 2.26. Note

[4]It is important to set a safe current limit during such a search since an inappropriate gate voltage could otherwise allow a large current through the HEMT drain and cause permanent damage. I damaged one HEMT this way, but the foresight to purchase spares prevented this from impacting the final SNOLAB experiment.

that two curves were actually measured for each bias setting, one with slightly-higher and one with slightly-lower I_{ds} than the desired value. This was necessary because of the inexact V_{gs}-finding method described above. I calculated g_d for each bias setting by performing a linear fit to the IV curve around the target V_{ds} and reporting the resulting slope.



I_{ds} [mA]	g_d [mS] @ V_{ds} = 75 mV	g_d [mS] @ V_{ds} = 100 mV	g_d [mS] @ V_{ds} = 175 mV	g_d [mS] @ V_{ds} = 250 mV
0.5	0.291	0.292	-0.141	0.398
1.0	0.921	0.441	0.182	0.201
2.0	3.695	0.642	-0.294	0.853
3.0	10.597	1.635	1.245	-0.316

Figure 2.26: A typical HEMT drain-conductance measurement with results. (Top) Table of fitted drain-conductance values for each candidate bias point. (Bottom) Plot of drain-current measurements for various gate voltages and a high density of drain voltages. Different colors denote different gate voltages as described by the color-bar. Drain currents corresponding to potential bias points are noted with horizontal dashed lines. Black points are earlier measurements used to find the required gate voltage. Black lines denote the fits used to calculate the drain-conductances for each bias combination.

2.5.3 Cutoff Voltage

I measured the HEMT cutoff voltages using the same setup. In order to do so, I set each desired drain voltage and raised the gate voltage (from an initial large negative value). This is similar to the initial gate-bias search, but with emphasis on fine binning near the gate voltage where current begins to flow. An example of the resulting data can be seen in 2.27. To estimate the cutoff voltage from each dataset, I iteratively removed all points which were statistically inconsistent with zero current and treated the highest remaining gate voltage as the cutoff.

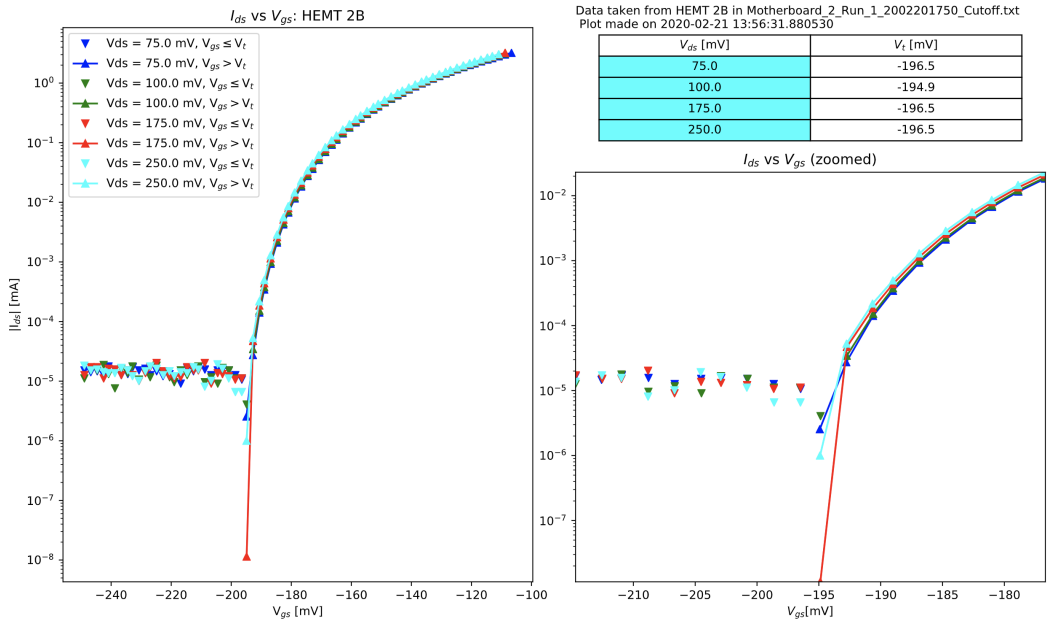


Figure 2.27: A typical HEMT voltage-cutoff measurement with results. (Left) Current measured as a function of drain and gate voltage. Different colors denote different drain voltages as described in the legend. Down-facing triangles are data points consistent with zero current. The highest gate voltage of the zero-current points is used to estimate the cutoff voltage for each drain voltage. (Top-right) Table of voltage cutoff estimates. (Bottom-right) A zoom-in of the left plot.

2.5.4 Ungrounded HEMT Characterization

In order to test the necessity of grounding the HEMTs during cooldown, I re-characterized three mezzanine cards of HEMTs without cooldown grounding. Each of these mezzanine cards had previously been characterized using the standard grounding method so a direct comparison could be made. Two of the mezzanine cards had been previously characterized multiple times (during different cooldowns) and were confirmed to have reproducible characteristics.

For the ungrounded test, I left all HEMT connections floating (connected to the Keithley 7001 switch system but not to ground) for the cooldown from ~300 to 4 K. I then performed the usual transconductance and IV-curve measurements. The results varied by mezzanine card.

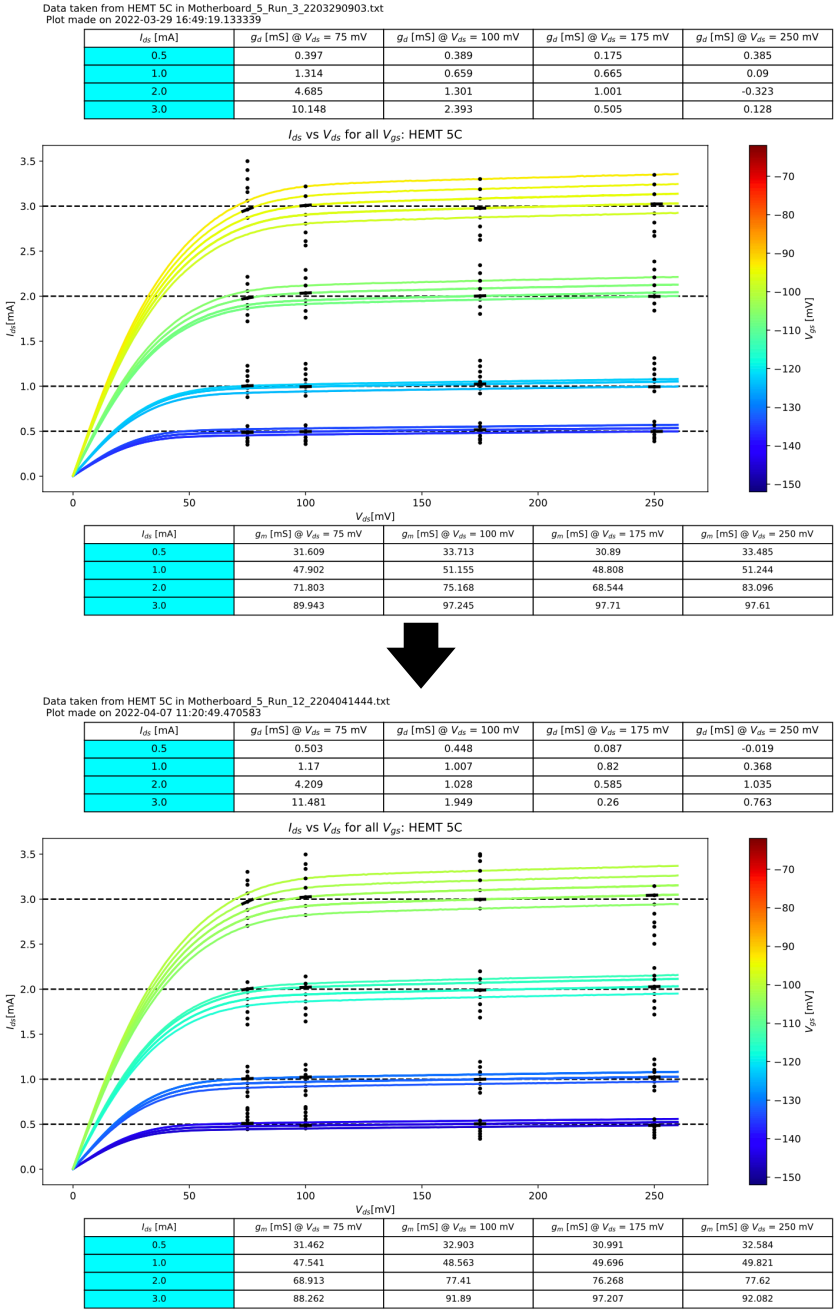
HEMTs on two of the three mezzanine cards saw little to no difference between the grounded and ungrounded results. The largest change among these HEMTs was one requiring ~5-10 mV lower V_{gs} to achieve each of the candidate (V_{ds}, I_{ds}) combinations. The g_m and g_d values at each bias point were not significantly

changed. The g_m tables, g_d tables, and IV curves for this HEMT (grounded and ungrounded) can be seen in Figure 2.28.

HEMTs on the third mezzanine card saw greater differences between the grounded and ungrounded results. For two HEMTs on the card, the V_{gs} required to achieve each (V_{ds}, I_{ds}) combination increased by ~10-20 mV. For the other two HEMTs, the required V_{gs} decreased. For one of these two, the required V_{gs} decreased so significantly that the measurement software failed to find V_{gs} values low enough to achieve the lower I_{ds} bias combinations. This result also produced noticeably higher g_d and lower g_m values for some bias points (see Figure 2.29).

While still at 4 K, I then explicitly grounded all HEMTs to chassis ground through a current-limiting 1.3 kΩ resistor. The resistor lowers the risk of a high current damaging HEMTs while discharging static charge. I then repeated the measurements and found that grounding at 4 K had no effect.

For the last test, I allowed the cryostat to warm up to 150 K. I then grounded all HEMTs through the 1.3 kΩ resistor and left the grounding in place while cooling back down to 4 K. 150 K was chosen because previous tests had confirmed that the SNOLAB HEMTs stop exhibiting transistor-like behaviour somewhere between 130-150 K and instead gain a high conductance even for low gate voltages. We theorized that the high conductance at 150 K may allow charge trapped in the conducting HEMT channel (where the 2-dimensional electron gas forms) to discharge. After returning to 4 K, I repeated the measurements and found that most of the HEMTs that were initially affected by the ungrounded cooldown had returned to their grounded-cooldown behavior (see Figure 2.30).



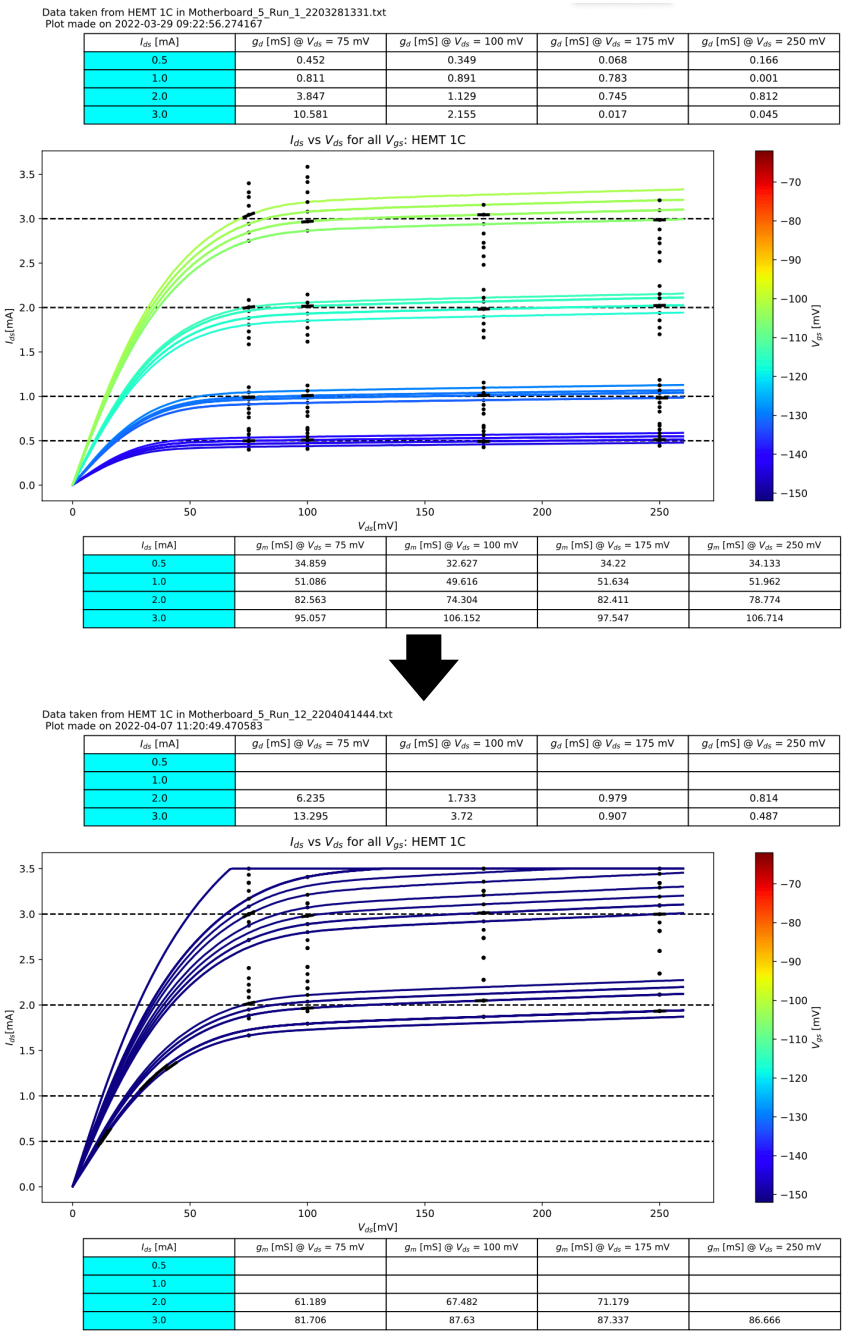
I_{ds} [mA]	g_d [mS] @ V_{ds} = 75 mV	g_d [mS] @ V_{ds} = 100 mV	g_d [mS] @ V_{ds} = 175 mV	g_d [mS] @ V_{ds} = 250 mV
0.5	0.397	0.389	0.175	0.385
1.0	1.314	0.659	0.665	0.09
2.0	4.685	1.301	1.001	-0.323
3.0	10.148	2.393	0.505	0.128

I_{ds} [mA]	g_m [mS] @ V_{ds} = 75 mV	g_m [mS] @ V_{ds} = 100 mV	g_m [mS] @ V_{ds} = 175 mV	g_m [mS] @ V_{ds} = 250 mV
0.5	31.609	33.713	30.89	33.485
1.0	47.902	51.155	48.808	51.244
2.0	71.803	75.168	68.544	83.096
3.0	89.943	97.245	97.71	97.61

I_{ds} [mA]	g_d [mS] @ V_{ds} = 75 mV	g_d [mS] @ V_{ds} = 100 mV	g_d [mS] @ V_{ds} = 175 mV	g_d [mS] @ V_{ds} = 250 mV
0.5	0.503	0.448	0.087	-0.019
1.0	1.17	1.007	0.82	0.368
2.0	4.209	1.028	0.585	1.035
3.0	11.481	1.949	0.26	0.763

I_{ds} [mA]	g_m [mS] @ V_{ds} = 75 mV	g_m [mS] @ V_{ds} = 100 mV	g_m [mS] @ V_{ds} = 175 mV	g_m [mS] @ V_{ds} = 250 mV
0.5	31.462	32.903	30.991	32.584
1.0	47.541	48.563	49.696	49.821
2.0	68.913	77.41	76.268	77.62
3.0	88.262	91.89	97.207	92.082

Figure 2.28: An example comparison between one HEMT's characterization after a regular grounded cooldown (Top) and an ungrounded cooldown (Bottom). The table above each plot shows the fitted drain-conductance values for each candidate bias point. The table below each plot shows the measured transconductance values for each candidate bias point. This example HEMT saw the greatest difference between grounded and ungrounded results among the two mezzanine cards which saw little to no change.



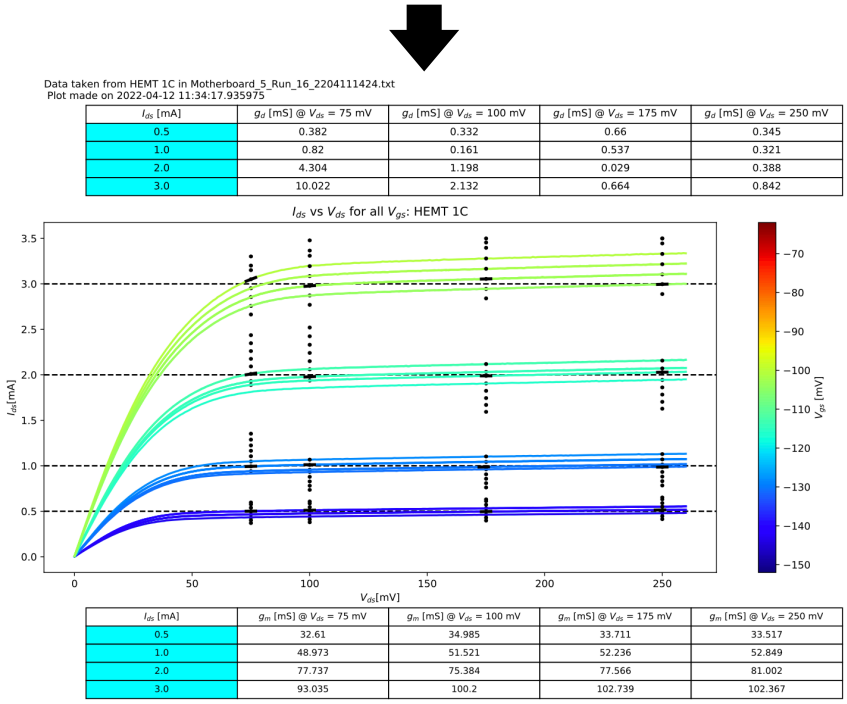
I_{ds} [mA]	g_d [mS] @ V_{ds} = 75 mV	g_d [mS] @ V_{ds} = 100 mV	g_d [mS] @ V_{ds} = 175 mV	g_d [mS] @ V_{ds} = 250 mV
0.5	0.452	0.349	0.068	0.166
1.0	0.811	0.891	0.783	0.001
2.0	3.847	1.129	0.745	0.812
3.0	10.581	2.155	0.017	0.045

I_{ds} [mA]	g_m [mS] @ V_{ds} = 75 mV	g_m [mS] @ V_{ds} = 100 mV	g_m [mS] @ V_{ds} = 175 mV	g_m [mS] @ V_{ds} = 250 mV
0.5	34.859	32.627	34.22	34.133
1.0	51.086	49.616	51.634	51.962
2.0	82.563	74.304	82.411	78.774
3.0	95.057	106.152	97.547	106.714

I_{ds} [mA]	g_d [mS] @ V_{ds} = 75 mV	g_d [mS] @ V_{ds} = 100 mV	g_d [mS] @ V_{ds} = 175 mV	g_d [mS] @ V_{ds} = 250 mV
0.5				
1.0				
2.0	6.235	1.733	0.979	0.814
3.0	13.295	3.72	0.907	0.487

I_{ds} [mA]	g_m [mS] @ V_{ds} = 75 mV	g_m [mS] @ V_{ds} = 100 mV	g_m [mS] @ V_{ds} = 175 mV	g_m [mS] @ V_{ds} = 250 mV
0.5				
1.0				
2.0	61.189	67.482	71.179	
3.0	81.706	87.63	87.337	86.666

Figure 2.29: Another example comparison between one HEMT's characterization after a regular grounded cooldown (Top) and an ungrounded cooldown (Bottom). The table above each plot shows the fitted drain-conductance values for each candidate bias point. The table below each plot shows the measured transconductance values for each candidate bias point. This example HEMT saw the greatest difference between grounded and ungrounded results among all HEMTs tested both ways. The measurement software failed to find V_{gs} values low enough to achieve I_{ds} of 0.5 and 1.0 mA. The blank table entries are where calculations failed due to data quality.



I_{ds} [mA]	g_d [mS] @ V_{ds} = 75 mV	g_d [mS] @ V_{ds} = 100 mV	g_d [mS] @ V_{ds} = 175 mV	g_d [mS] @ V_{ds} = 250 mV
0.5	0.382	0.332	0.66	0.345
1.0	0.82	0.161	0.537	0.321
2.0	4.304	1.198	0.029	0.388
3.0	10.022	2.132	0.664	0.842

I_{ds} [mA]	g_m [mS] @ V_{ds} = 75 mV	g_m [mS] @ V_{ds} = 100 mV	g_m [mS] @ V_{ds} = 175 mV	g_m [mS] @ V_{ds} = 250 mV
0.5	32.61	34.985	33.711	33.517
1.0	48.973	51.521	52.236	52.849
2.0	77.737	75.384	77.566	81.002
3.0	93.035	100.2	102.739	102.367

Figure 2.30: The result of warming up to 150 K, grounding, and cooling back to 4 K (while grounded) for the HEMT from Figure 2.29. Notice that the g_d, g_m, and V_{gs} for each (V_{ds}, I_{ds}) combination are very similar to those of the grounded cooldown.

Two HEMTs did not return to their grounded-cooldown behavior. One of these had already been observed to have anomalously high g_d (during grounded and ungrounded cooldowns). The other HEMT displayed an odd feature that I believe to be a readout effect and not a reflection of the HEMT. I discount both of these as being anomalous.

It is clear from these tests that the best choice for SNOLAB is to keep all HEMTs grounded during the entire cooldown to prevent buildup of static charge. If grounding is not done and unusual behavior is exhibited by the HEMTs, it may help to warm the HEMTs to 150 K and re-cool with grounding.

2.6 Properties of SuperCDMS SNOLAB HEMTs

The individual results for each of the tested SNOLAB HEMTs are available online to SuperCDMS collaborators[5]. The online documentation includes data and plots (equivalent to Figures 2.25, 2.26, and 2.27) for each HEMT. In this section, we will instead discuss the HEMTs' statistical properties and how they compare to those from [37]. Table 2.1 displays the means and standard deviations of the measured parameters for 60 of the SNOLAB HEMTs at one bias point. The 60 HEMTs included are those attached to the mezzanine cards of batches 1, 3, and 4. These are the batches allocated to the SNOLAB iZIP detectors. Twelve mezzanine cards (holding 48 HEMTs) from these batches will be used for the iZIP detectors of the initial SNOLAB payload.

parameter	mean	std dev	units
Gate Voltage	-124.36	14.57	mV
Transconductance	50.69	3.54	mS
Drain Conductance	0.57	0.38	mS
Cutoff Voltage	-185.86	11.74	mV

Table 2.1: The means and standard deviations of the measured parameters for 60 of the SNOLAB HEMTs. All parameters are shown for drain voltage of 100 mV. The gate voltage, transconductance, and drain conductance are shown for drain current of 1 mA. The 60 HEMTs included are those attached to mezzanine-card batches 1, 3, and 4.

2.6.1 Gate-Voltage Biasing

Histograms of the gate voltages required to achieve each bias setting with the SNOLAB HEMTs are shown in Figure 2.31. We can see that the required voltages

[5] `https://confluence.slac.stanford.edu/display/CDMS/HEMT+IV+Screening+Results`.

are on the order of -100 mV, while the total variation between HEMTs is about 60 mV for each setting. This means that the variation is a significant fraction of the total voltage and confirms that using feedback to bias the HEMTs (rather than manually setting gate voltages) is the more practical design. The required values are well within the range of what can be output by the circuit's biasing op amp.

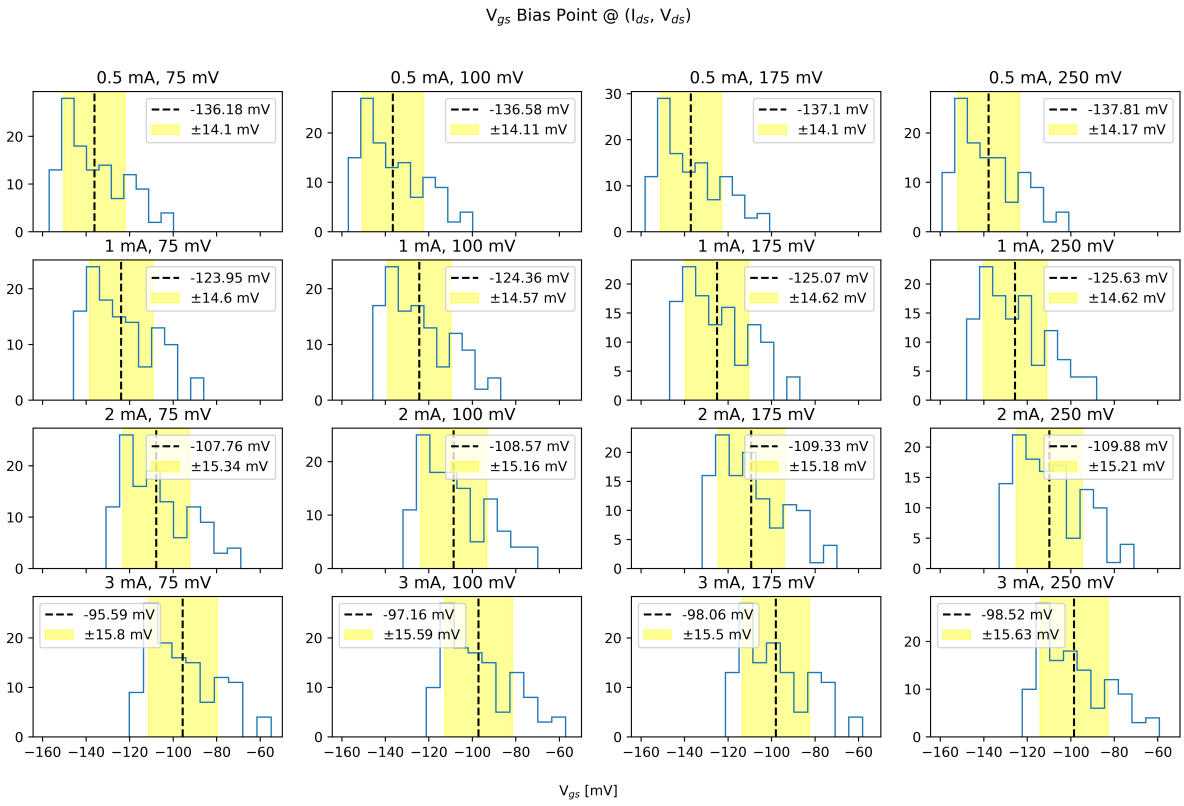


Figure 2.31: Histograms of the gate voltages required to produce each bias point in the SNOLAB HEMTs. Mean and standard-deviation of the results are also displayed.

2.6.2 Transconductance and Amplifier Gain

Histograms of the HEMT transconductances for each bias setting can be seen in Figure 2.32. For drain voltage of 100 mV and drain current of 1 mA, the mean transconductance was near 50 mS. As mentioned in Section 2.4.5, the 100 pF HEMTs from [37] were measured to have transconductance of 35 mS for the same bias point. The [37] transconductance is over four standard deviations lower than my own measurement. For the SNOLAB ionization circuit, either value of transconductance is sufficient to stabilize the ionization amplifier's gain (see Eq. 2.19 and the

subsequent paragraph). The difference is likely due to improvements in the HEMT design between [37] and production of the SNOLAB HEMTs in 2015.

I have also included a scatter plot of the HEMT transconductances, plotted in the order they were tested (Figure 2.33). The testing took place from February 2020 to June 2022, but no significant trend can be seen in the scatter plots. The plots support the conclusion that the testing setup and HEMTs did not suffer any unnoticed changes during the test period. Similar scatter plots displaying gate voltages, drain conductances, and cutoff voltages are included in Appendix A. None of the scatter plots suggest any concerning features or trends.

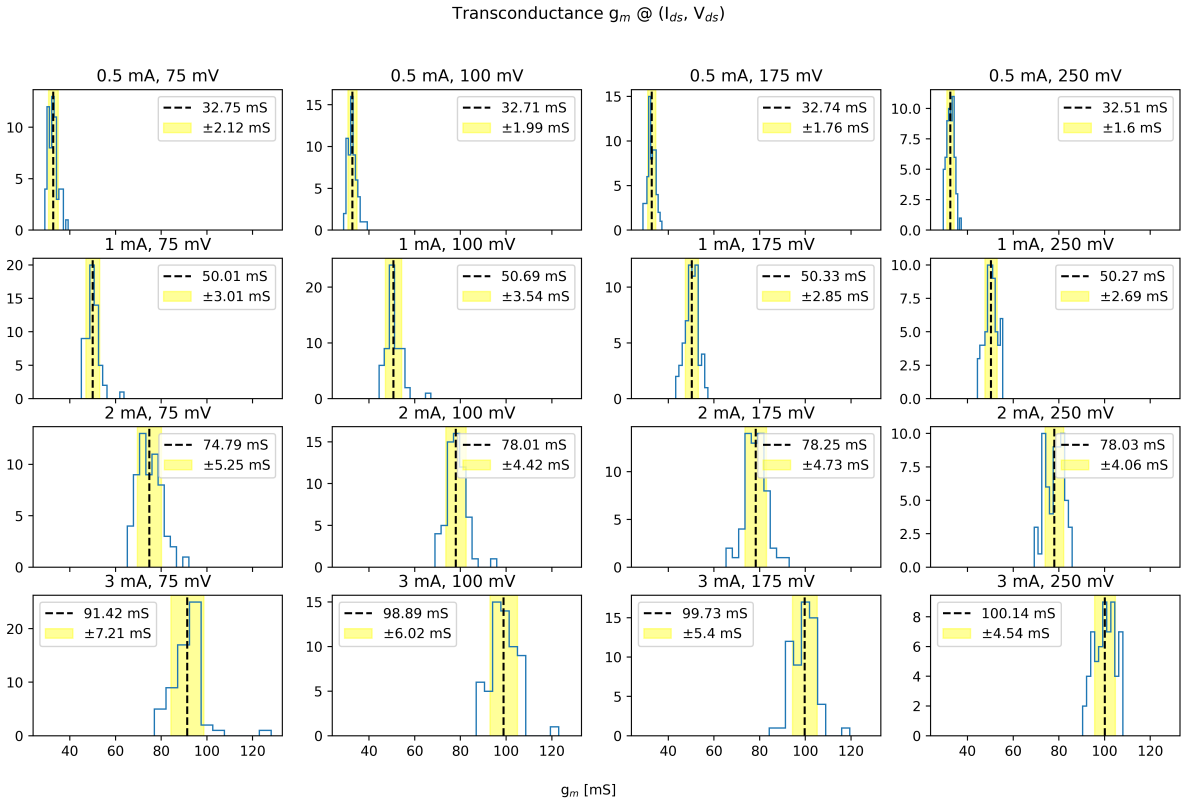


Figure 2.32: Histograms of the transconductances measured in the SNOLAB HEMTs for each potential bias point. Mean and standard-deviation of the results are also displayed.

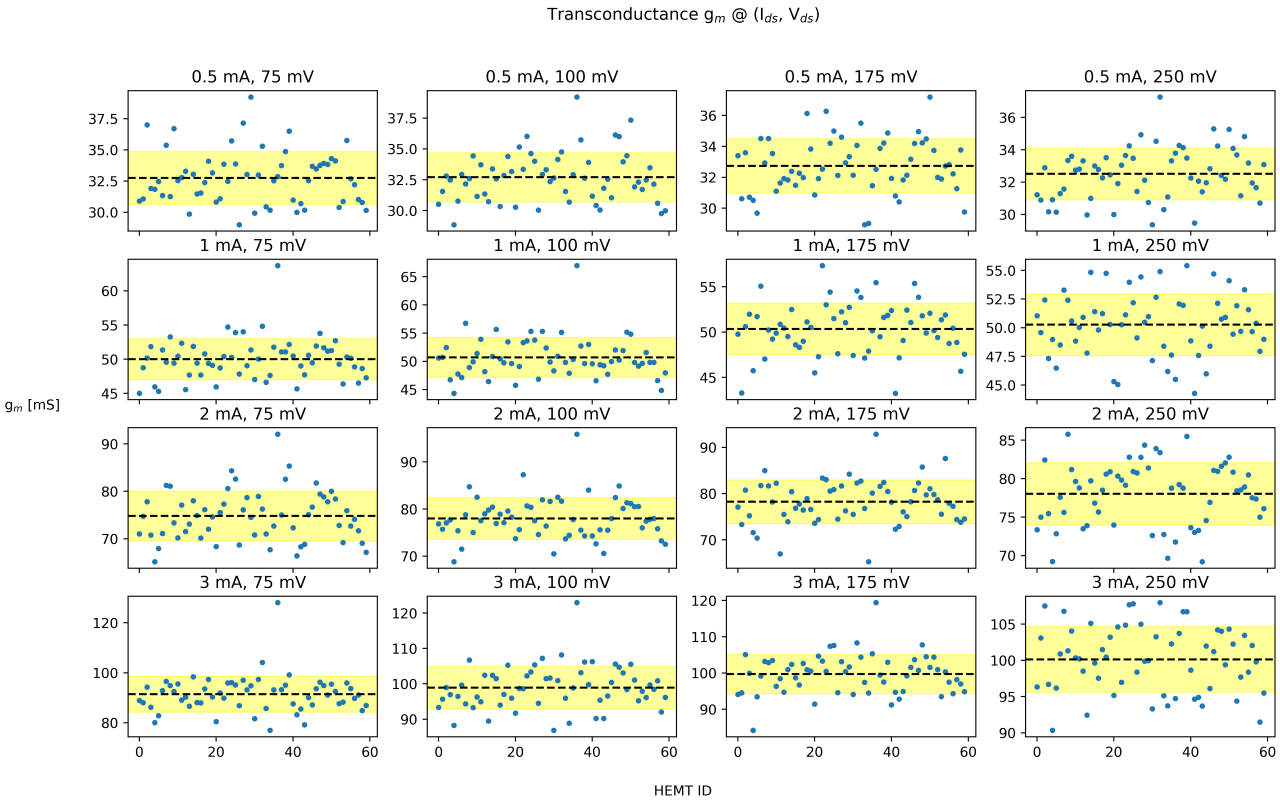


Figure 2.33: Scatter plots displaying the SNOLAB HEMT transconductances, in the order the HEMTs were tested. No significant trend can be seen over the 2 year testing period.

2.6.3 Drain Conductance

Histograms of the SNOLAB HEMT drain conductances for each bias setting can be seen in Figure 2.34. The data confirms that the HEMTs are near saturation (drain conductance near zero) for all but the low-voltage, high-current combinations. Those results are expected since a high drain conductance is required to achieve high current with only a low voltage. We will operate the SNOLAB HEMTs away from this region ensuring they are in or near saturation.

For drain voltage of 100 mV and drain current of 1 mA, the mean drain conductance was 0.57 ±0.38 mS. As mentioned in Section 2.4.5, the 100 pF HEMTs from [37] were measured to have drain conductance of 0.75 mS for the same bias point. The [37] drain conductance is within one standard deviation of my measurements at Caltech.

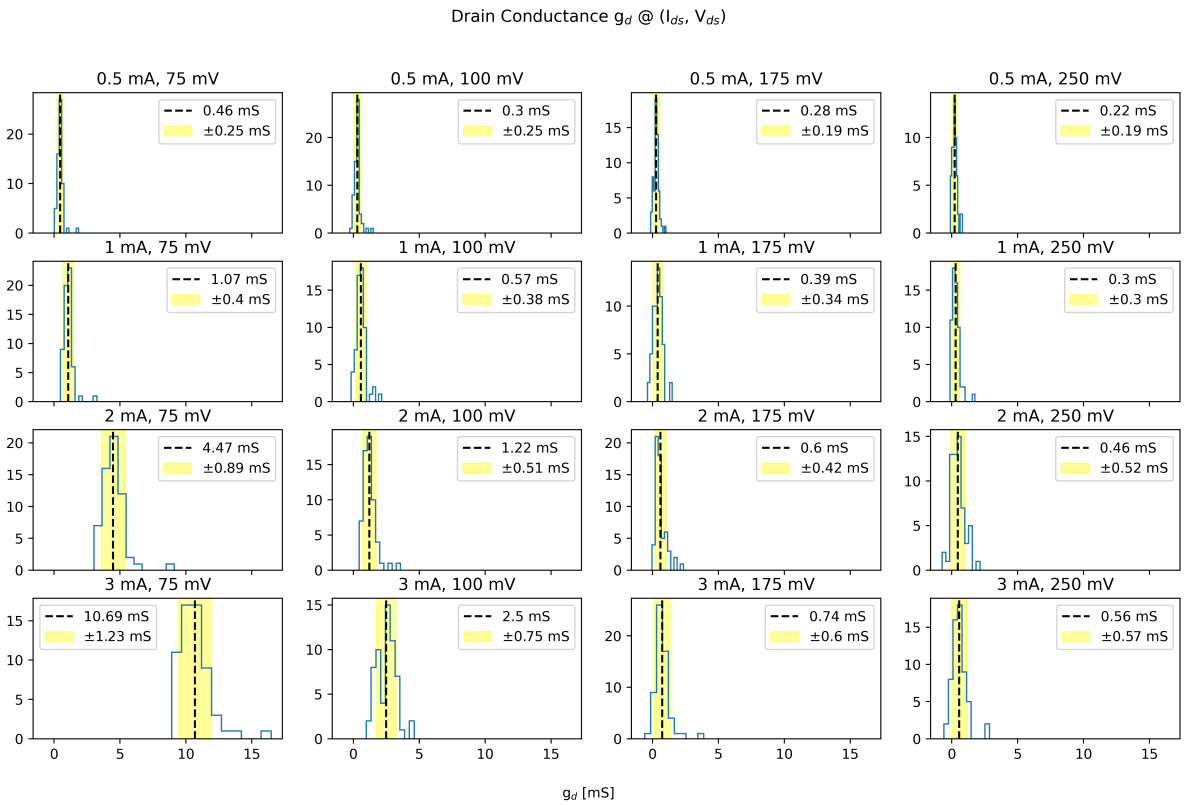


Figure 2.34: Histograms of the drain conductances measured in the SNOLAB HEMTs for each potential bias point. Mean and standard-deviation of the results are also displayed.

2.6.4 Cutoff Voltage

Histograms of the SNOLAB HEMT cutoff voltages for each drain voltage can be seen in Figure 2.35. As expected, there is no substantial change within our range of drain voltages. Simply setting any HEMT gate voltage to -250 mV should guarantee it remains safely in cutoff.

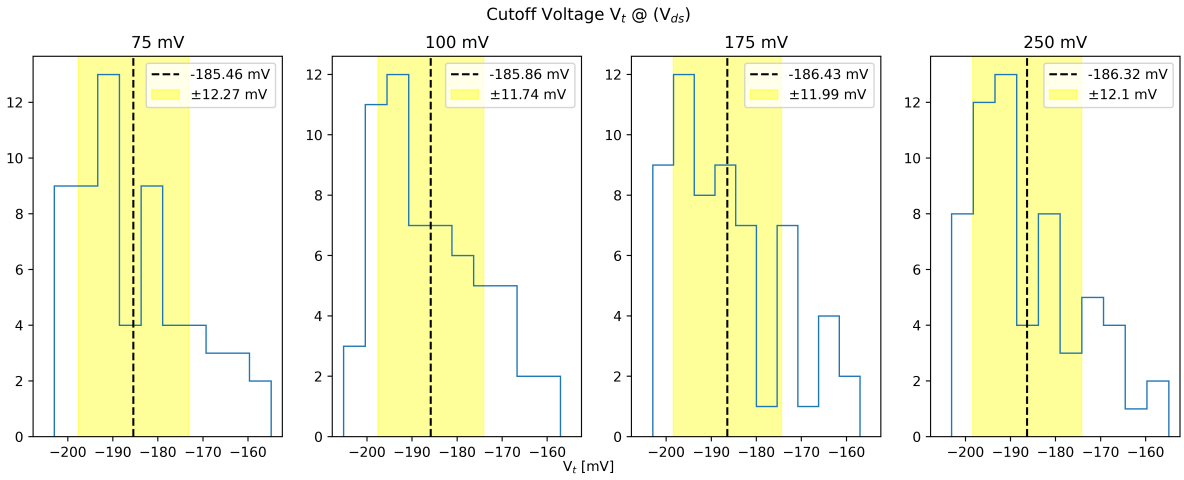


Figure 2.35: Histograms of the cutoff voltages measured in the SNOLAB HEMTs for each potential bias drain voltage. Mean and standard-deviation of the results are also displayed.

2.7 Ongoing Testing of the Ionization Readout

With each of the SNOLAB HEMTs characterized, construction of the iZIP detector towers progressed and was completed in 2022. Testing of the ionization readout continues as of Fall 2023. Ionization signals have been observed with the expected dissipation time (~100 µs). An example background signal, which was recorded above ground at the SLAC National Accelerator Laboratory, can be seen in Figure 2.36. For this event, both sides of the detector see signals simultaneously. As discussed in Section 2.1.2, this indicates that the event was not close to either of the detector's flat surfaces. The two sides of the detector see signals of opposite charge, because the electric field moves positively charged holes to one surface and negatively charged electrons to the other. The largest magnitude signal is seen in the outer channel of side 2 ("QOS2) indicating this may be an example of a high-radius background event.

The ionization readout's noise was also measured at SLAC. Some example measurements can be seen in Figure 2.37. In these examples, several complicated features

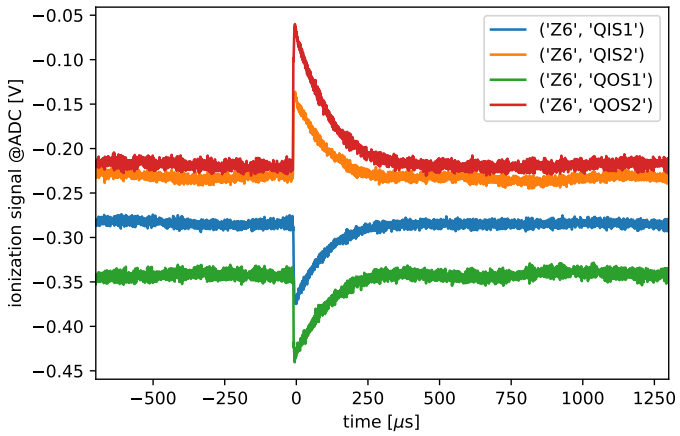


Figure 2.36: A background event with signal observed simultaneously in all four ionization channels of the sixth detector ("Z6") in SuperCDMS-SNOLAB tower 4. This data was taken above ground at SLAC National Accelerator Laboratory. In the legend, "QI" ("QO") indicates an inner (outer) charge channel. "S1" ("S2") indicates side 1 (side 2) of the detector. The signals are measured in units of voltage at the ADC. There is another amplifier between the output of the charge amplifier and the ADC input, so this is not precisely equivalent to the charge-amplifier output.

appear. Spectral lines appear around 10 kHz and above 200 kHz. Some of these are likely the result of the noisy lab environment and will be reduced in the carefully shielded SNOBOX. Other features around 10 kHz and at low frequency are suspected to be the result of filters installed on the warm electronics. Upcoming tests at SLAC will involve removing or altering these filters. Debugging will continue while the rest of the SNOLAB experiment is assembled.

After debugging, it is expected that the ionization noise will look more similar to that of Figure 2.38. This figure shows the charge-amplifier noise that was measured during the design stage of the SNOLAB experiment using a similar HEMT and a functionally similar (but not identical) amplifier design. The noise shows similar spectral peaks from the same noisy lab environment but otherwise fits well to the noise model. The noise model includes the HEMT voltage noise and Nyquist-Johnson noise from the feedback resistor.

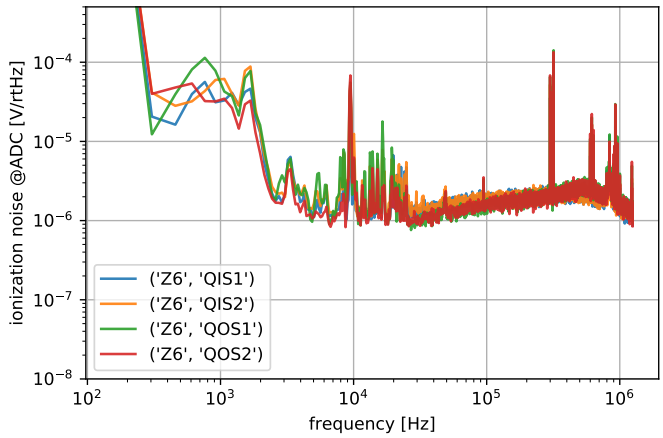


Figure 2.37: Charge-amplifier output noise (measured at the ADC) for the same setup and iZIPs as in Figure 2.36.

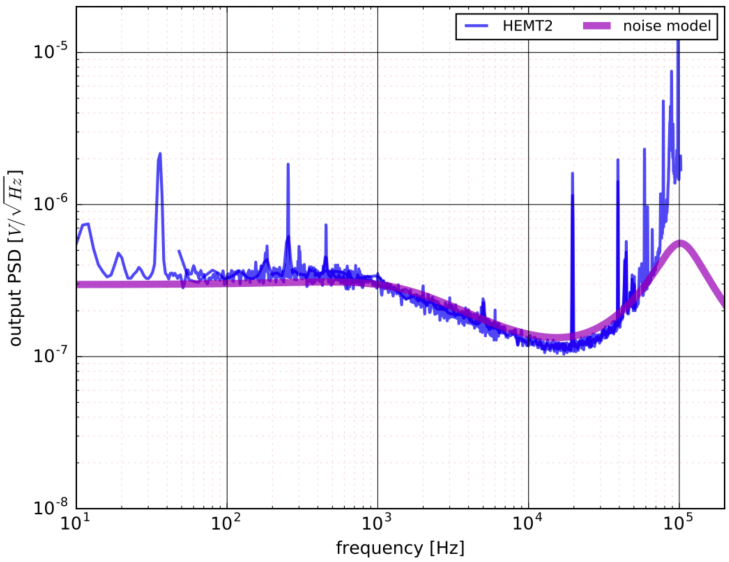


Figure 2.38: Charge-amplifier noise measured during the design stage of SuperCDMS SNOLAB (and featured in the SNOLAB Technical Design Report.) This measurement was taken using a different HEMT and circuit design than the final SNOLAB experiment. The HEMT utilized was also from [37] and the circuit should be comparable. The noise model (purple), which includes the HEMT voltage noise and Nyquist-Johnson noise, was confirmed to be a good match to the measurement (blue).

Chapter 3

HVEV RUN 3 AT NEXUS

During the design and construction stages of large-scale experiments such as SuperCDMS SNOLAB, it is beneficial to test newer detector designs in smaller scale experiments requiring less exposure and smaller data-taking and analysis teams. Such small-scale experiments can probe parameter space that is not targeted by longer-term experiments while also testing design concepts that may be used in future large-scale experiments. The HVeV Run 1 experiment was one such experiment (performed within the SuperCDMS collaboration) that successfully excluded new parameter space with just six days of data taking [40]. In this chapter, I will discuss the HVeV detector concept, the results of previous runs using such detectors, and the entirety of the HVeV Run 3 experiment.

3.1 HVeV Detectors

SuperCDMS high-voltage eV-resolution (HVeV) detectors operate similarly to SNOLAB HV detectors. Their target crystals are smaller than the SNOLAB design, typically $1 \times 1 \times 0.4$ cm rectangular prisms, equivalent to 0.93 grams of silicon per detector. Like HV detectors, they use quasiparticle-trap-assisted electrothermal-feedback Transition-edge-sensors (QETs) for phonon detection. HVeV detectors have two QET channels (one inner and one outer) on a single detector surface. The opposing detector surface has an aluminum grid that can be biased (relative to the grounded QETs) to induce NTL-phonon production. First-generation detectors, including the one used for HVeV Run 1 ([40]), have bias-grid electrodes with 20% coverage. Second-generation detectors, including the one used for HVeV Run 2 ([26]), have 5% coverage.

HVeV detectors are designed to achieve a combination of excellent energy resolution and high voltage bias sufficient to measure the NTL-phonon energy of individual electron-hole (e^-h^+) pairs. As seen in Eq. 2.1, each e^-h^+ pair will produce $e\Delta V$ of NTL-phonon energy. The HVeV Run 1 detector achieved consistent 14 eV resolution on energy deposited in the substrate. With ΔV = 140 V, the detector observed 1-e^-h^+ events at $10\sigma_E$. Events that ionized n e^-h^+ pairs were observed at $n10\sigma_E$. The resulting spectrum of observed event energies becomes quantized around these "e^-h^+ peaks." Example HVeV e^-h^+ peaks (produced using laser photons) can be

seen in Figure 3.1.

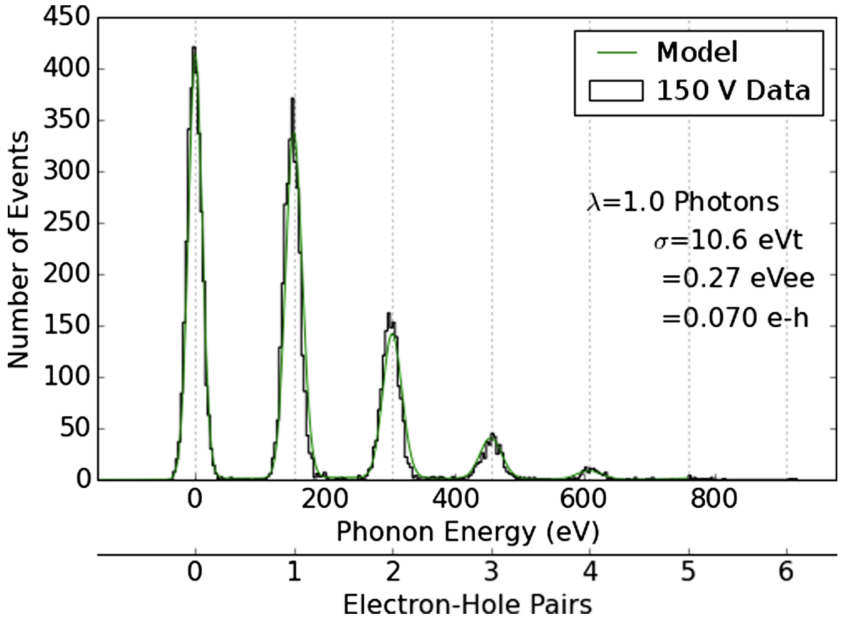


Figure 3.1: The energy spectrum observed by the HVeV Run 1 detector with ~10 eV energy resolution and 150V of applied bias. Each e^-h^+ pair produced is accelerated by the voltage bias to generate 150 eV of phonon energy observed in the detector. This example was made by the Run 1 analyzers using laser calibration data [40]. The fit to data is shown in green.

The HVeV detectors' sensitivity to NTL-phonon energy is a powerful tool for measuring event ionization. For the HVeV Run 1 detector, the ionization resolution was equivalent to $\sigma_{eh} = \sigma_E/e\Delta V = 0.1\ e^-h^+$ pairs. In silicon, an electron recoil of ~1.2 eV will produce a single e^-h^+ pair. Therefore, the HVeV detector's sensitivity to single e^-h^+ pairs corresponds to sensitivity to electron recoils as small as ~1.2 eV.

There have been a number of variations on the HVeV mask design. Most often, the adjusted parameters include the transition-edge-sensor (TES) length, the quasiparticle-trapping-fin length, and the TES number per channel (N_{TES}). The TES number is generally represented using the channel's total normal resistance, which is equivalent to the individual TES resistance divided by N_{TES}. Figure 3.2 displays a typical QET design with TES and fin lengths labeled. The various parameters affect important detector characteristics including energy resolution, dynamic range, and signal fall time. A large dynamic range is important for detector calibration. Signal fall time is important for distinguishing events from RF backgrounds. An example HVeV design can be seen in Figure 3.3. This design is the NF-C mask,

which maximizes dynamic range while maintaining good energy resolution. In theory, this choice makes NF-C detectors ideal for calibration.

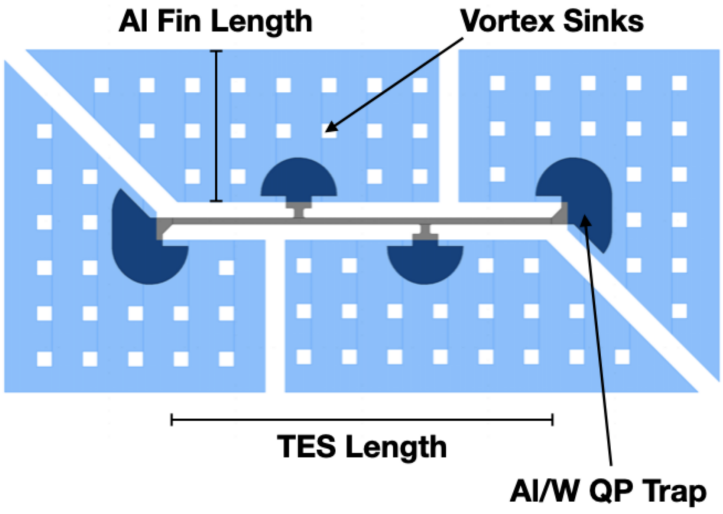


Figure 3.2: The QET design used in the NF-C detector design (from [41]). QET stands for Quasiparticle-trapping-assisted Electrothermal-feedback Transition-Edge Sensor (TES). Event phonons are absorbed and break Cooper pairs in the low-gap aluminum fins (light blue). Some fraction of the resulting quasiparticle electrons are trapped upon encountering the lower-gap Al/W traps (dark blue). Quasiparticles transitioning to the lower-gap material release excess energy as phonons that head the TES (grey).

A fabricated NF-C HVeV detector can be seen in Figure 3.4. The detector is mounted between two copper-plated FR-4 printed circuit boards (PCBs).

3.2 Previous HVeV Science

HVeV detectors have been used in two previous dark-matter searches. These are labeled HVeV Runs 1 and 2. In this section, we will discuss the experimental designs and results of both runs with a focus on how they impacted the design of HVeV Run 3.

3.2.1 Run 1 at Stanford

The first science data from an HVeV detector was taken over 6 days in an above-ground dilution refrigerator at Stanford University. The analysis considered 27 hours of data taken with the detector bias set to -140 V (using 20%-coverage bias electrodes). The refrigerator temperature was unstable, and periods with temperatures over 36 mK were removed during the analysis. Gain was corrected to account for the remaining temperature variations using the recorded fridge temperature. The

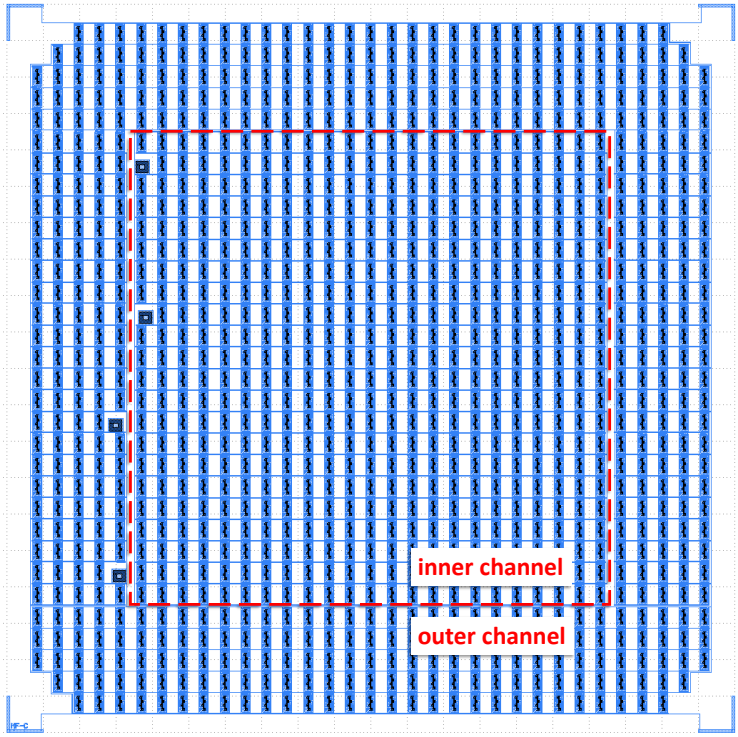


Figure 3.3: The detector-mask design used to produce the NF-C detector. At this scale, individual QETs appear as single blue and black pixels. Black is actually the AL/W traps and TESs. Blue is the AL fins and signal lines. The inner and outer channels are distinguished by the red dashed box (not part of the actual mask). Each channel is made up of ~500 QETs in parallel. The channel readout pads are the larger squares visible on the left side of the mask.

detector was calibrated using laser events, which showed the energy response to be nonlinear. A quadratic fit was used to calibrate the pulse-height to event energy. The total exposure after cuts was 0.49 gram-days. As mentioned above, the detector resolution was equivalent to 14 eV or 0.1 $e^{-}h^{+}$ pairs.

The first run set exclusion limits on dark-matter-electron scattering down to a dark-matter mass of 0.5 MeV (Figure 3.5). At the time, these limits were world-leading below 4 MeV [40]. Limits were also set on dark-photon kinetic mixing down to a dark-photon mass of 1.2 eV (Figure 3.6).

3.2.2 Run 2 at Northwestern

The second run using an HVeV detector used an updated design focused on improving the phonon energy resolution. The detector utilized had the same mass and volume as Run 1, but achieved energy resolution equivalent to 3 eV with a voltage

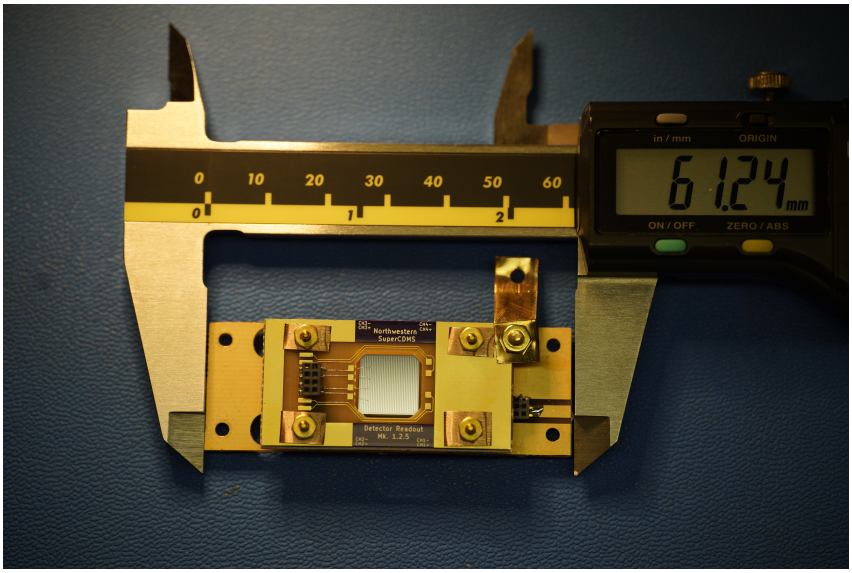


Figure 3.4: A photo of the NF-C detector in the copper and PCB (FR-4) housing used for Run 3. The detector face is visible through a window in the center of the PCB. The calipers show the housing as being ~6 cm wide. Wire-bonds connecting to the readout pads of the detector can be seen to the left of the detector.

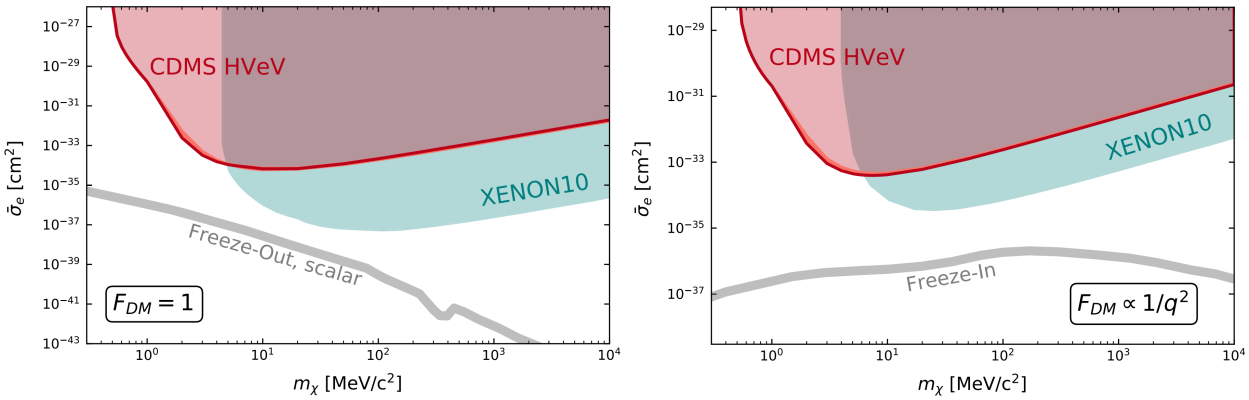


Figure 3.5: The dark-matter-electron scattering limits produced by HVeV Run 1 (from [40]). Red lines are the limits. The narrow pink region around each limit is the limit's uncertainty from varying the ionization model. The excluded regions are shaded in a transparent red. The parameter space excluded below 4 MeV was world-leading when Run 1 was published. (Left) The limit for scattering via a heavy mediator. (Right) The limit for scattering via a light mediator.

bias of 100 V (set using 5%-coverage bias electrodes). The resultant ionization resolution was equivalent to 0.03 $e^- h^+$ pairs. The detector was run in an above-ground adiabatic demagnetization refrigerator (at Northwestern University), which required daily cycling in order to maintain a temperature between 50 and 52 mK for 10-12

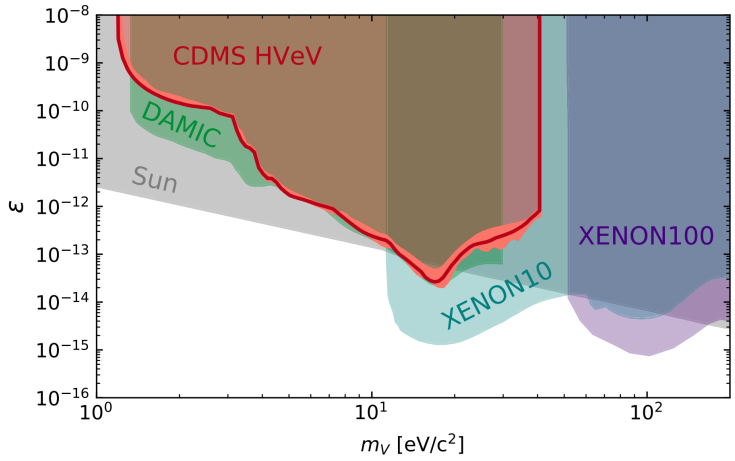


Figure 3.6: The dark-photon absorption limit produced by HVeV Run 1 (from [40]). The red line is the limit. The narrow pink region around the limit is the uncertainty from varying the ionization model and the photoelectric cross section. The excluded region is shaded in a transparent red.

hours each day. Outside these times, the fridge remained at 4 K. The run achieved a greater exposure of 1.2 gram-days after cuts [26].

Laser calibration data was taken each day after cooldown to account for the effects of fridge cycling on the calibration. Laser data was also taken for a range of fridge temperatures, in order to correct for temperature variation within the measured range. Time periods when the fridge temperature drifted outside the calibration range were removed.

Adjacent to its HVeV detector, Run 2 operated a veto detector designed to pickup RF background signals from the lab environment. The analysis removed all HVeV-detector events that were close in time to significant signals in the veto detector. It is statistically unlikely for weakly-interacting dark-matter to produce recoils in multiple detectors simultaneously (compared to radiogenic, cosmogenic, or RF backgrounds), so removing coincident events preferentially lowers the background rate. This type of cut is known as a live-time cut since the data and associated live-time were removed together. Alternatively, data-quality cuts remove events on the basis that the event was poorly characterized. Data-quality cuts lower the efficiency of the search since they quantify the experiment's inability to characterize all events precisely. The Run 2 veto detector was a second silicon wafer instrumented with a single TES. The transition temperature of the veto TES was near the fridge operating temperature, causing the veto detector to perform poorly and have only a small effect on the analysis.

Unlike Run 1, the Run 2 signal model considered effects we refer to as charge-trapping (CT) and impact ionization (II). Charge-trapping occurs when an electron or hole traveling through the crystal is captured by an oppositely ionized donor or acceptor impurity in the crystal. Donor (acceptor) impurities have donated (accepted) an electron to the conduction band, leaving the impurity positively (negatively) charged and capable of capturing an electron (hole). Conversely, impact ionization occurs when a traveling electron or hole frees a charge carrier that was previously captured by a donor or acceptor impurity. Since these effects remove or add charge carriers partway through the NTL-production process, they create events with non-quantized NTL-phonon energy that appear outside the typical e^-h^+ peaks. CT and II are visualized in Figure 3.7.

For Run 2, charge-trapping and impact-ionization were assumed to occur with fixed probabilities. The CT probability was defined as the probability that an e^-h^+ pair loses one of its charge carriers. The II probability was defined as the probability that an e^-h^+ pair produces one additional charge carrier. The model assumed that only one of the two processes could occur for a given e^-h^+ pair, and otherwise the pair would produce the full NTL-phonon energy [43]. The Run 2 probabilities were measured (using laser data) to be $11 \pm 3\%$ and $2^{+3}_{-2}\%$ for CT and II, respectively [26]. When the Run 1 detector was characterized using the same model, its CT was observed to be significantly lower at $0.713 \pm 0.093\%$ [44]. The Run 1 detector's II was comparable to Run 2 at $1.576 \pm 0.110\%$. The Run 1 probabilities were only

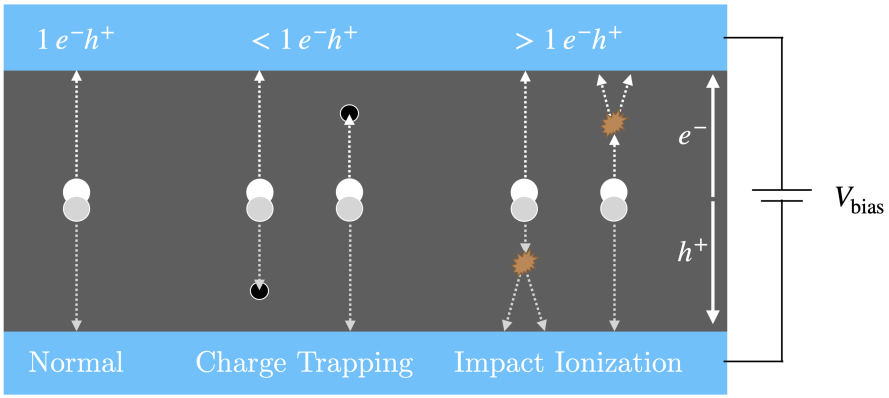


Figure 3.7: A diagram of charge transport in a voltage-biased detector. (from [42]). In each case, we see a bulk event that produces 1 e^-h^+ pair. On the left, the pair travels (in opposite directions) all the way across the detector. This will produce phonon energy equivalent to eV_{bias}. Many of such events will form the 1-e^-h^+ peak in the energy spectrum. In the middle, we see examples of one charge carrier being trapped by an oppositely ionized donor or acceptor impurity (charge trapping). Such cases will produce $< eV_{\text{bias}}$ of energy. On the right, we see examples of one charge carrier freeing a similar charge carrier that was previously captured by a donor or acceptor impurity (impact ionization). Such cases will produce $> eV_{\text{bias}}$ of energy.

measured after the Run 1 analysis and were therefore not used. Both probabilities were sufficiently low that we would not expect including them to significantly change the Run 1 limits.

The electron-scattering and dark-photon-absorption exclusion limits produced by the Run 2 analysis can be seen in Figures 3.8 and 3.9, respectively. For most of the mass range, the Run 2 limits were slightly worse than those of Run 1 (despite Run 2 having improved energy resolution and $\geq 2\times$ exposure). The weaker limits were actually the result of the Run 2 detector's higher CT rate and using a more conservative limit-setting method. Specifically, Run 2 used a Poisson limit-setting method rather than the Optimum-Interval (OI) method used by Run 1. The OI method was found to be highly dependent on the expected signal spectral (including CT and II effects) and therefore produced problematic systematic uncertainties. The Poisson method is more robust to uncertainty in CT and II because it only considers events within the quantized peaks.

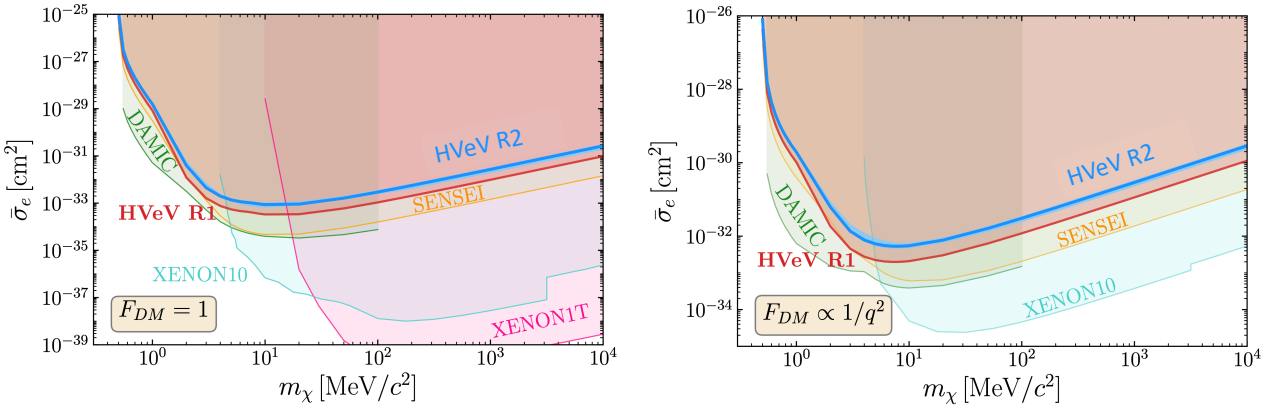


Figure 3.8: The dark-matter-electron scattering limits produced by HVeV Run 2 (from [26]). Blue lines are the Run 2 limits. The narrow light-blue region around each limit is the limit's uncertainty from varying the ionization model. (Left) The limit for scattering via a heavy mediator. (Right) The limit for scattering via a light mediator.

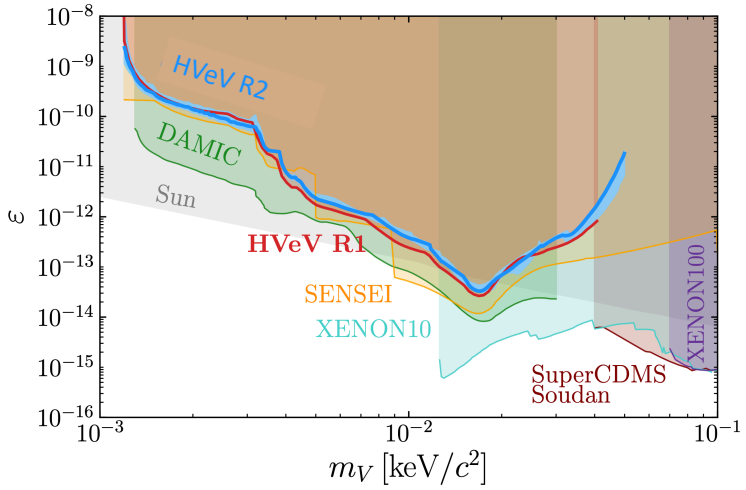


Figure 3.9: The dark-photon absorption limit produced by HVeV Run 2 (from [26]). The blue line is the limit. The narrow light-blue region around the limit is the uncertainty from varying the ionization model and the photoelectric cross section.

The Run 2 analysis also set an exclusion limit on the axioelectric coupling for axion-like particles (ALPs) down to an ALP mass of 1.2 eV. This limit was world-leading among direct detection experiments, but remains significantly less sensitive than the limit imposed by stellar-cooling observations ($g_{ae} \lesssim 10^{-13}$, see Figure 2.16).

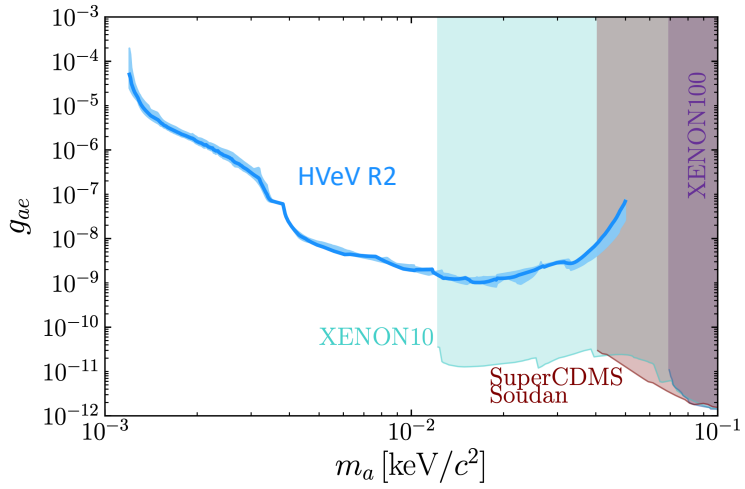


Figure 3.10: The ALP-absorption limit produced by HVeV Run 2 (from [26]). The blue line is the limit. The narrow light-blue region around the limit is the uncertainty from varying the ionization model and the photoelectric cross section. The entire region remains significantly less sensitive than the limit imposed by stellar-cooling observations ($g_{ae} \lesssim 10^{-13}$).

3.2.3 Low-Energy Excess

In HVeV Runs 1 and 2, the low-mass sensitivity was limited by an unexpected excess of events observed at low energy ($\sim$1 e^-h^+). The observed excesses were consistent with similar excesses in other sub-GeV dark-matter searches [45]. The Run 2 excess (compared to a dark-matter model that fits the $\geq$1 e^-h^+ region well) can be seen in Figure 3.11.

Additional data taken with the Run 2 detector at multiple bias voltages (0, 60, and 100 V) was used to study possible sources of this low-energy excess [45]. Both Run 2 and [45] recognized that a significant number of low-energy events occurred in groups immediately following larger events. These groups (including the initial larger event) were labeled "burst events." An example burst event can be seen in Figure 3.12. The [45] analysis proposed that such events could result from luminescence of the FR-4 PCBs used to clamp the detectors in Runs 1 and 2. Evidence included the low ionization yield observed in the larger events at the beginning of each burst, which is better explained by multiple simultaneous energy

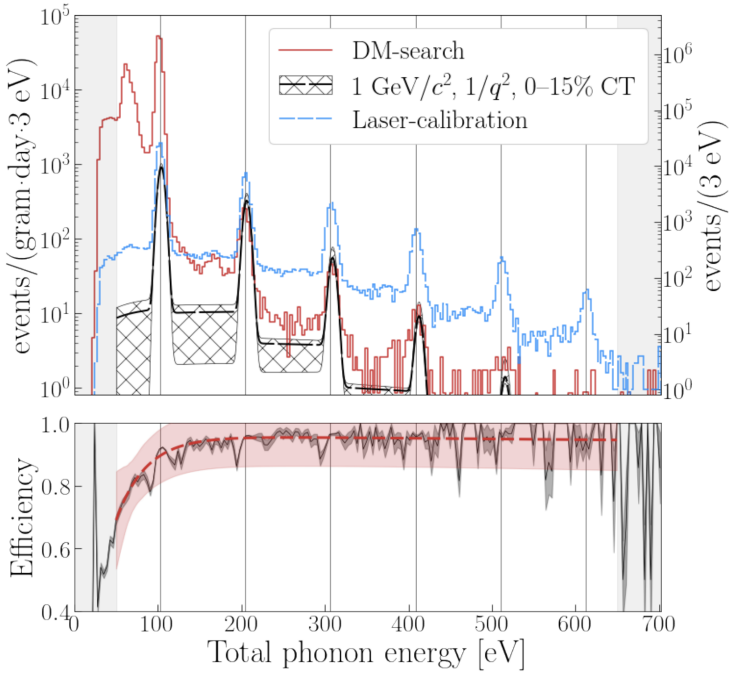


Figure 3.11: The HVeV Run 2 low-energy excess. (Top) The laser and dark-matter-search data from HVeV Run 2 with all live-time and quality cuts applied. Both spectra were taken with detector bias of 100 V. An attempt was made to fit the search data to a light-mediator electron-scattering signal model (black). Although the model is good for e^-h^+ peaks 2-5, it severely underestimates the rate of 1-e^-h^+ events. This led the analyzers to conclude that there is a significant excess of low-energy events. (Bottom) The cut efficiency of the Run 2 experiment. The black line is the measured efficiency by bin (with uncertainties in grey). The red line is the efficiency used. A large uncertainty (the pink shaded region) was used to account for the dips in the measured efficiency. Both plots and data from [26].

depositions than by a single larger energy deposition. FR-4 luminescence would also produce photons with decay time consistent with the rate observed in burst events.

Run 2 removed burst events by cutting periods of live time with heightened event rate. If the burst events were caused by an external background (like FR-4 luminescence), this cut would fail to remove any bursts that did not produce multiple recoils with distinct arrival times in the single detector. This would explain why the low-energy excess was still present in Run 2. In order to remove more of such events, HVeV Run 3 was designed to operate with four detectors mounted in adjacent pairs. Bursts that do not produce distinct events in a single detector may produce coincident events in nearby detectors. Such events can be removed with a coincidence cut (like Run 2's

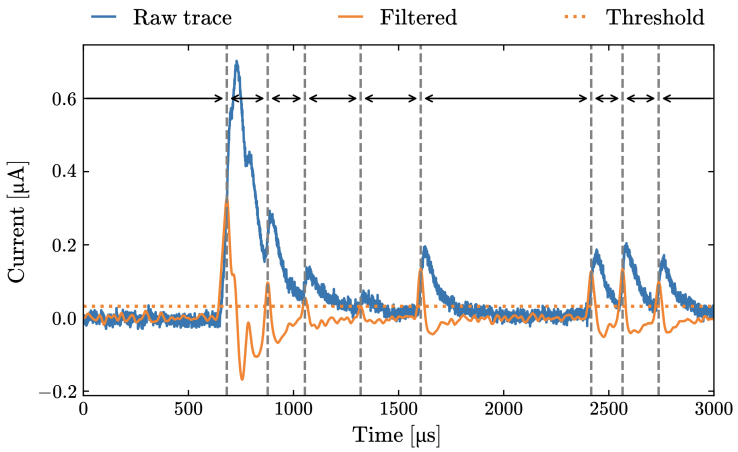


Figure 3.12: An example burst event before and after filtering (from [45]). We see the initial event is several times larger than the subsequent events. A Gaussian-derivative-filter was applied to produce the filtered signal. We see that the filter works well for identifying individual events in the burst via a post-filter threshold trigger. Identified events are marked with vertical dashed lines.

veto-detector cut but more functional). Run 3 was planned and executed before it was recognized that the FR-4 was both relatively radioactive (and thus a source of particle backgrounds) as well as the source of luminescence. Run 3 therefore used the same mounting method as Runs 1 and 2 (including FR-4 PCBs).

3.3 Setup at NEXUS

In addition to being the first HVeV run with multiple detectors, HVeV Run 3 included several improvements to the experimental setup (compared to previous runs). These improvements included being performed underground and utilizing a dilution refrigerator capable of maintaining stable temperature for multiple months. In this section, I will present an overview of the data-taking setup including these improvements.

3.3.1 NEXUS facility

Run 3 was taken at the Northwestern Experimental Underground Site (NEXUS), which is 107 m underground and has 300 mwe of rock overburden. Previous runs were exclusively taken above ground and were therefore exposed to greater cosmogenic backgrounds. During Run 3, cosmogenics were also considered a potential origin for burst events, so the underground site was part of the broader plan to reduce the low-energy excess. The facility utilized was also a class 10,000 clean-room, which lowered some backgrounds but did not affect those originating from FR-4 inside the detector chamber.

3.3.2 Cryogenics and Shielding

Data was taken inside a CryoConcepts dilution refrigerator capable of steady-state low-temperature operation. For Run 3, the fridge was maintained at 10.5 mK from December 2020 to February 2021. Temperature only changed during a software reset once every 2-3 weeks. The temperature promptly returned to 10.5 mK. Data taken during off-temperature periods was removed using a temperature-based live-time cut. Otherwise, the temperature was stable within 0.1 mK. The system stability allowed data to be calibrated once without need for a temperature correction.

The setup also included lead shielding inside and outside the cryostat. The total lead shielding blocked nearly all line-of-sight between the detectors and the surrounding environment.

3.3.3 Payload Arrangement

HVeV Run 3 was taken using a payload of four 0.93-gram Si HVeV detectors. The detectors were mounted in pairs, with each pair sharing one mounting bracket. The arrangement can be seen in Figure 3.13. Using multiple detectors raises the total target mass and therefore the exposure that can be acquired each day. We can also use the multiple detectors to implement the coincidence cut described in Section 3.2.3. With this arrangement, each detector acts as a veto detector for each of the others (particularly effective for those that share a bracket). All four detectors were mounted on the same horizontal axis. Unlike the veto detector for Run 2, the Run 3 detectors were operated well below their TES transition temperatures.

The detector payload was mounted in a copper enclosure that was made light-tight using epoxied feedthroughs and copper tape over mating surfaces (see Figure 3.14).

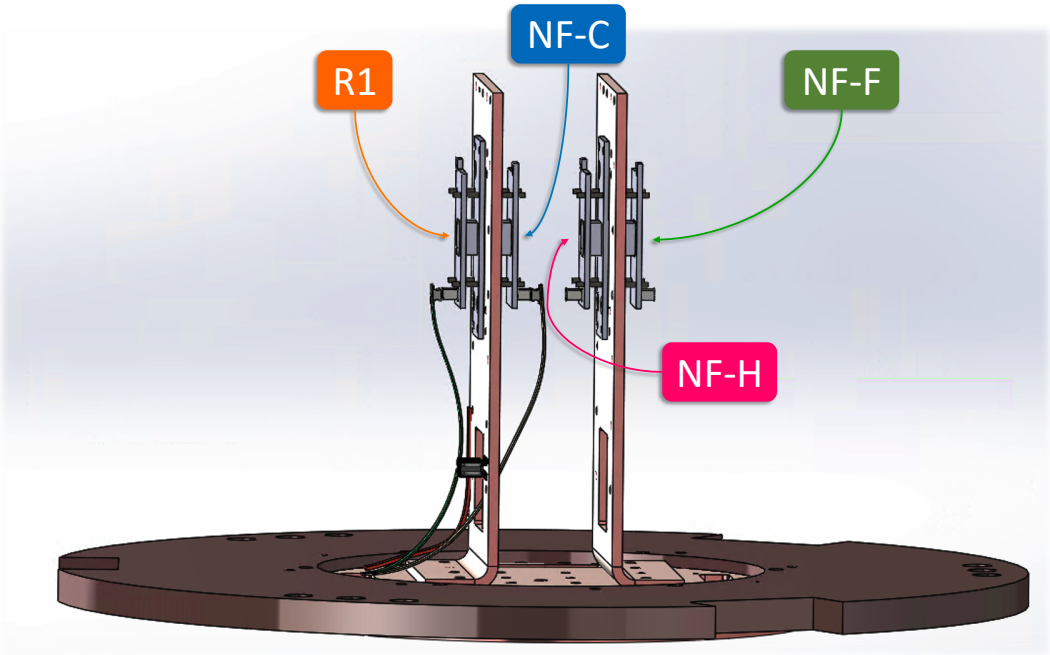


Figure 3.13: The arrangement and mounting of detectors in HVeV Run 3. Each detector is labeled with the mask design used for its production. The fiber optic used to input calibration photons came up through the center plate, between the two "L" brackets, and faced directly upwards. During the run, the tops of the "L" brackets were connected for better thermal coupling.

Figure 3.14: A photo of the copper enclosure holding the Run 3 detectors. The bottom copper plate is the same bottom plate from Figure 3.13. The Inner lead shield can be seen above the cylindrical copper housing. Three vertical bars thermally connect the bottom plate the the mixing-chamber stage of the dilution refrigerator (above the inner lead shield).

In order to perform the laser-photon calibration, a fiber-optic cable was routed through the fridge and into the detector chamber. The fiber came up through the bottom cold plate in Figure 3.13 and was left facing upward between the detector brackets. This was a change from previous runs, which used fiber optics pointed directly at the back of each run's detector. The assumption for Run 3 was that the laser photons would illuminate the entire detector chamber, producing calibration events in all detectors simultaneously. While laser data confirmed that events were generated in the detectors, it also showed a larger fraction of non-quantized events than seen in previous runs. We suspect that this excess was caused by the increased rate of laser photons striking the detectors' sidewalls. The HVeV electric field is less uniform near the sidewalls than in the detector bulk. The sidewalls are also expected to have a greater concentration of crystal impurities. Both factors would lead to sidewall events having a higher probability to produce non-quantized signals.

3.3.4 Detector Designs

Each of the four detectors used in Run 3 had different designs. The detector labeled "R1" was produced using the same design as the HVeV Run 1 detector (including 20% electrode coverage on the non-sensor surface). The R1 detector was also fabricated on the same substrate as the Run 1 detector and should have similar rates of charge trapping and impact ionization. The other three detectors were second-generation HVeV designs with improved energy resolution and 5% electrode coverage. These detectors were fabricated using the same silicon as the Run 2 detector. The second-generation "NF-C" detector was designed to have a large dynamic range for improved calibration. The second-generation "NF-F" and "NF-H" detectors were designed to have improved position resolution. R1 and NF-C were paired together on one mounting bracket. Likewise, NF-H and NF-F were paired together on the other mounting bracket.

After cooling down, we found that the NF-F detector was not functional. Upon inspection of the NF-F design, it became apparent that the inner channel of QETs had an unintentional short between input and output (see Figure 3.15). This design error rendered the NF-F detector unusable. We decided to continue the run with the remaining three detectors.

3.4 Data Taking

Science data was taken every day between December 27, 2020 and February 9, 2021 using a SuperCDMS Detector Control and Readout Card (DCRC). In order

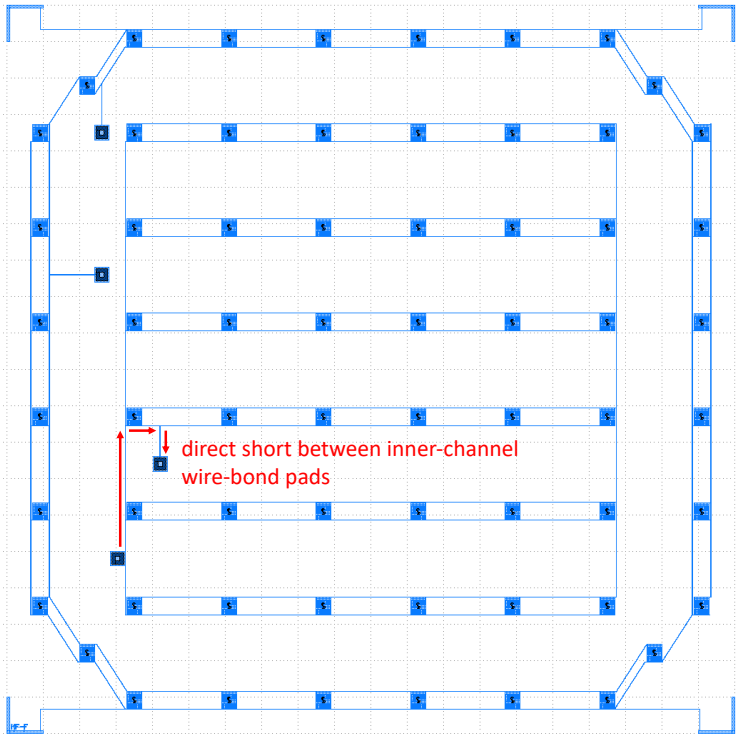


Figure 3.15: The detector-mask design used to produce the NF-F detector. An unintentional short between the input and output wire-bond pads of the inner channel is indicated with red arrows.

to study burst events, it was desired to continuously take data rather than to trigger on events above a set threshold. A set threshold could miss the low-energy events that are typically seen after the leading event in a burst. The data-throughput limit of the DCRC made this impossible for three detectors operated simultaneously. Instead, data was taken with 50% live time. Specifically, the system would record 0.5 seconds of data and then stop recording for a random length of time between 0 and 1 seconds (uniformly distributed). With 45 days of data taking, this resulted in approximately 22.5 days of live time. This corresponds to ~21 gram-days of exposure per detector. In reality, there were brief interruptions each day when data taking had to be manually stopped and restarted. Science-data taking was also stopped occasionally to record laser data.

On January 11th, the data-taking team attempted to take laser data but could not find signals coincident with the TTL trigger (the transistor logic that indicates when the laser diode is pulsed). The issue was rectified by reconnecting the serial port connected to the laser-diode driver. Laser data was then taken. It was later observed that reconnecting the driver altered the noise performance of the detectors. Low-

frequency noise, which had fluctuated significantly before January 11th, decreased and remained stable in all three operating detectors. A spectral line at 10 kHz also appeared in the R1 and NF-H detectors. The spectral line was not a signal-like background, so the post-January 11th noise performance was an overall improvement. During the analysis, we decided to only use data taken after the diode reset on the 11th. This lowered the data-taking days from 45 to 30. The approximate exposure was lowered from ∼21 to ∼14 gram-days per detector.

Laser data was also successfully taken on February 8th and 9th. The 9th was the last day of recording science data, so the analysis had access to laser data from the beginning and end of science-data taking.

The Run 3 analysis began as soon as data collection was completed (February 9, 2021). Run 3 was not expected to be sensitive to any previously unexplored parameter space. In part, this expectation was set by the increasing competition in the sub-GeV mass region (see the SENSEI result from 2020 [46]). Instead, the analysis was performed with the intention of qualitatively demonstrating the improvements from Runs 1 and 2 via the post-cut data spectrum and resultant exclusion limits. Accordingly, the data was not blinded or salted before the analysis began.

All science and laser data for Run 3 was taken with a voltage bias of 100 V (ΔV = 100 V).

3.5 Data Processing

Using the known DCRC design, the recorded ADC data was converted to units of current through each TES channel. Further use of the data required processing it to identify the timing and amplitude of individual events as well as converting the current signal to event energies. In the following section, we will discuss how the Run 3 data was processed. This includes event identification (triggering) and filtering to reduce non-signal-like backgrounds.

3.5.1 Triggering

Within each half second of data, events were identified using Gaussian-derivative-filter (GF) triggering. First, data was filtered using a kernel with the shape of a differentiated Gaussian. Then, events were identified as times when the GF-filtered data went above a set trigger threshold. The differentiated-Gaussian shape was chosen for its ability to distinguish events close in time. Figure 3.12 from the low-

energy event investigation shows GF triggering effectively identifying individual events within a burst [45]. For Run 3, a Gaussian with 25.6 µs standard deviation was used. With this filtering, events as close as ~120 µs could be distinguished. Figure 3.16 compares Run 3's GF to a matched filter (the type used in Run 2) using Run 3 data.

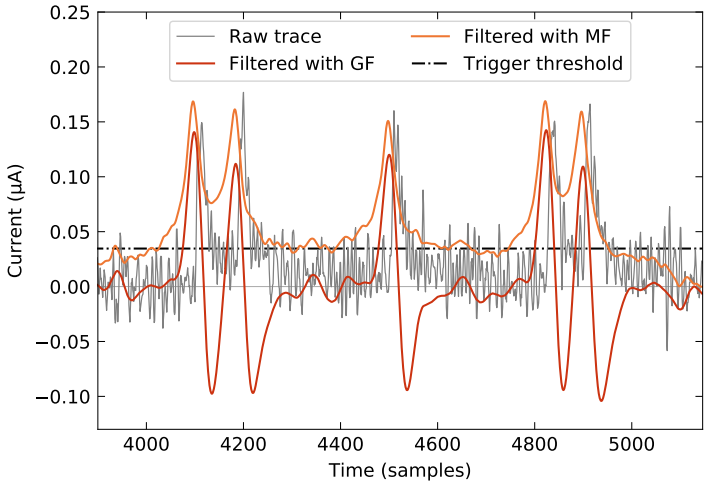


Figure 3.16: An example burst event in the Run 3 data. We see that the Gaussian-derivative filter (GF) works well for identifying individual events in the burst via a post-filter threshold trigger. The matched filter (MF) cleans the trace but is ineffective at identifying events that are near in time.

When tested using laser data, GF-trigger efficiency was observed to go to unity for event energies >35 eV. This indicated that only an insignificant number of events were missed in the energy range used for the limit setting (see Section 3.12.1).

Events were identified separately in each detector using the combined signal from inner and outer channels (relatively weighting each channel as described in Section 3.6). Each event's associated trace (the subset of time used for further analysis) was defined to include 6.55 ms to either side of the event's filtered maximum.

3.5.2 Optimal Filtering

The Gaussian-derivative filter was only applied when identifying event triggers. For the rest of the analysis, event traces were filtered using the Optimal Filtering (OF) method (see [47] and Section 4.2.1). This method creates a frequency-space filter proportional to the expected signal weighted by the measured noise. The relative weighting preferentially preserves frequencies consistent with signal and devalues

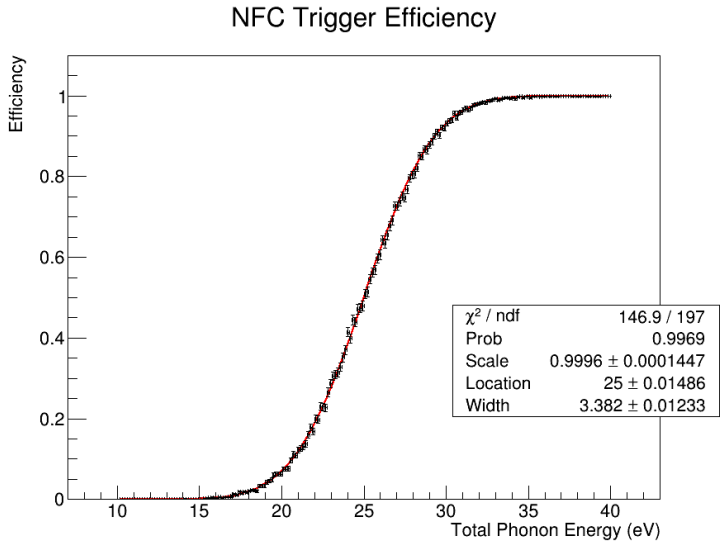


Figure 3.17: The Run 3 trigger efficiency fit to an error function. Trigger efficiency was calculated by applying the GF-trigger to the laser data and calculating the fraction of events identified. Efficiency goes to zero for low-energy events since they cannot be identified above noise. The important result is that efficiency goes to unity (with little uncertainty) above ~30 eV (well below the first quantized peak at 100 eV).

frequencies consistent with noise. The OF used for Run 3 was normalized such that signal-like events would have the same amplitude before and after filtering.

For Run 3, the expected-signal template for each detector was calculated using laser-generated events from January 11th. The events were aligned in time using the laser TTL signal and averaged to lower random noise. Events from traces with an overall slope were removed to avoid polluting the template with longer-timescale changes in signal. In total, just under 20k events were averaged for each detector. A low-pass Butterworth filter was used to remove remaining high-frequency noise. Specifically, a 10th-order forward-backward filter with 60 kHz cutoff frequency was used for NF-C and NF-H. The same type of filter was used for R1 but with a 20 kHz cutoff due to its higher noise. The NF-C template can be seen in Figure 3.18. The frequency-domain reduced χ^2, which quantifies discrepancies between data and the fitted OF template (in frequency space), was observed to be distributed about 1 for the Run 3 science data. This indicated that the laser-derived template was a good fit to the non-laser science data. The reduced χ^2 was also stable throughout the run, confirming that the signal shape did not change over time. The signal stability was attributed to the stable temperature and channel biasing maintained throughout the

run.

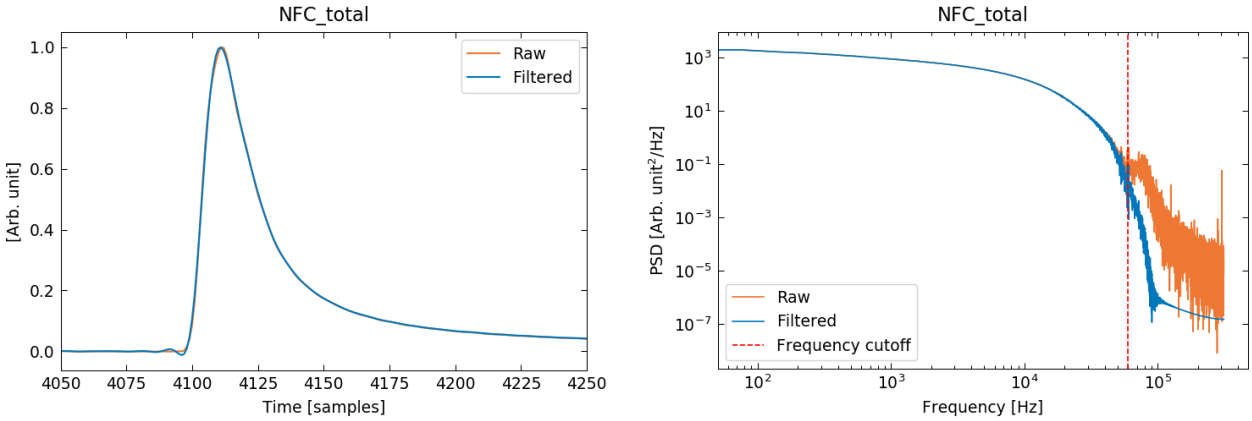


Figure 3.18: NF-C signal template

The raw and Butterworth-filtered signal template for the NF-C detector. (Left) The template in time space. Each time sample is equivalent to 1.6 µs. (Right) The template in frequency space. We see that the Butterworth filter effectively removes high-frequency noise.

While each detector's expected-signal template was constant throughout the run, the noise was re-estimated for every minute of data taking. For each minute, noise was calculated in each detector using traces triggered at random times. Randomly triggered traces with unusual mean, standard deviation, slope, or skewness were removed from the calculations. Typical noise power spectral densities (PSDs) for NF-C can be seen in Figure 3.19.

By combining templates and noise measurements, OFs were calculated for each detector and every minute of data taking. Each trace was then filtered using the appropriate OF. For each event, the amplitude of the filtered trace at the time of triggering was recorded as a reduced quantity (RQ) labeled the "OF0 amplitude". The reduced χ^2 of the fit between data and template at that time was recorded as the "OF0 χ^2". Amplitude and reduced χ^2 RQs were also recorded with the event timing allowed to shift to achieve the best fit (lowest χ^2) within ±24 µs of the trigger. These RQs were given the prefix "OFL" (OF with a Limited time shift). It is a feature of the OFL RQs that the maximum filtered amplitude in a trace will occur at the time that produces the lowest χ^2 fit [47]. Therefore, the OFL values also correspond to the highest filtered amplitude within the allowed time region. For events with little to no signal, this caused the OFL RQs to instead find the largest noise fluctuation resulting in biases in pulse height near zero amplitude. Accordingly, the analysis used OF0 values for sub-1 e^-h^+ events.

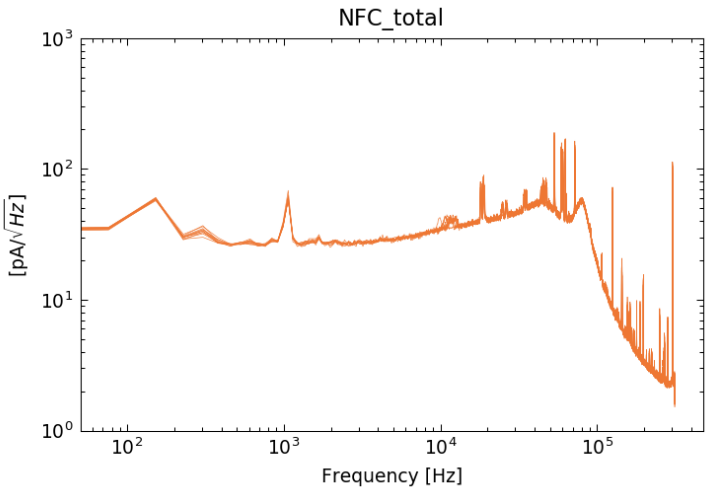


Figure 3.19: Typical noise power spectral density (PSD) in the NF-C detector. Many partially transparent PSDs are plotted on top of each other. Their similarity makes individual PSDs difficult to distinguish. Each PSD is calculated using random triggers from one minute of data taking. The noise was very stable, so the plot looks much like a single PSD.

3.6 Relative channel weighting

In HVeV detectors, calculating an event's total phonon energy requires combining the signals from all phonon-sensing channels. The Run 3 detectors had two channels each (one inner and one outer on a single surface). Inner and outer channels were not expected to have equivalent signal responses to absorbed energy[1]. With different responsivities, a simple sum of signals would depend on the partition of energy between the two channels. Energy partition depends on the initial event position within a detector, so partition dependence causes position dependence.

In order to produce a position-independent calibration of phonon energy, Run 3 calculated total signal using a weighted sum that minimized the dependence on energy partition. In this section, we will discuss the two methods used to calculate the weighting factors for each detector. The first method was flawed but produced the weighting values that were used in the final analysis. The second method was superior and confirmed that the first method's results were actually correct (despite the flawed method).

[1]The NF-C and NF-H detectors were actually designed to have equivalent responsivity (to absorbed energy) in the inner and outer channels. In practice, there will always be differences due to imperfect fabrication or biasing. The R1 detector was not designed to have equivalent responsivity.

3.6.1 Resolution-Optimizing Method

The first (flawed) method minimized the energy resolution in the 1 e^-h^+ quantized peak instead of directly removing partition dependence. The total signal (A) and relative weighting factor (α) were defined as in Eq. 3.1.

$$A \equiv \alpha A_{in} + A_{out} \tag{3.1}$$

where A_{in} and A_{out} are the inner and outer channel signals, respectively. For this method, we explicitly calculated the resolution on A for a range of choices of α. Resolution was calculated by fitting a Gaussian to the 1 e^-h^+ peak in A. The resolution (with an approximate energy calibration) would then be defined by Eq. 3.2.

$$\sigma_{\text{eV}} \sim \frac{\sigma_{Gaussian fit}}{\mu_{Gaussian fit}} 100 \text{ eV} \tag{3.2}$$

This calculation was performed using OF0 science data from each day of data taking. Laser data could not be used, since the large fraction of sidewall events altered the shape of the 1 e^-h^+ laser peaks and made them a poor estimator of detector resolution. Daily data was useful for checking that the resolution-optimizing α did not vary significantly over time.

The calculation was performed before the analysis χ^2 cut was finalized, so a basic energy-independent χ^2 cut was applied to the data instead. Another cut was applied to carefully select the 1 e^-h^+ peak events. The obvious choice would be to select these events using the quantized distribution of A, but such a selection would rely on prior knowledge of α. Instead, 1 e^-h^+ events were identified using a 2-dimensional cut dependent on both A_{in} and A_{out}. The cut retained events within a 3σ ellipse iteratively fit to data using principal-component analysis. The cut was 3σ under the assumption that the underlying distribution was bivariate normal. In that case, a 3σ cut would retain 98.9% of data from the underlying distribution. The 1 e^-h^+ selection was performed separately for each dataset. An example of the iterative 1 e^-h^+ cut applied to NF-C data can be seen in Figure 3.20.

After applying the cuts, σ_{eV} was calculated as a function of α for each dataset. An example result for NF-C can be seen in Figure 3.21. Each dataset's result was then fit to an approximate model for the variance of the weighted and normalized sum of

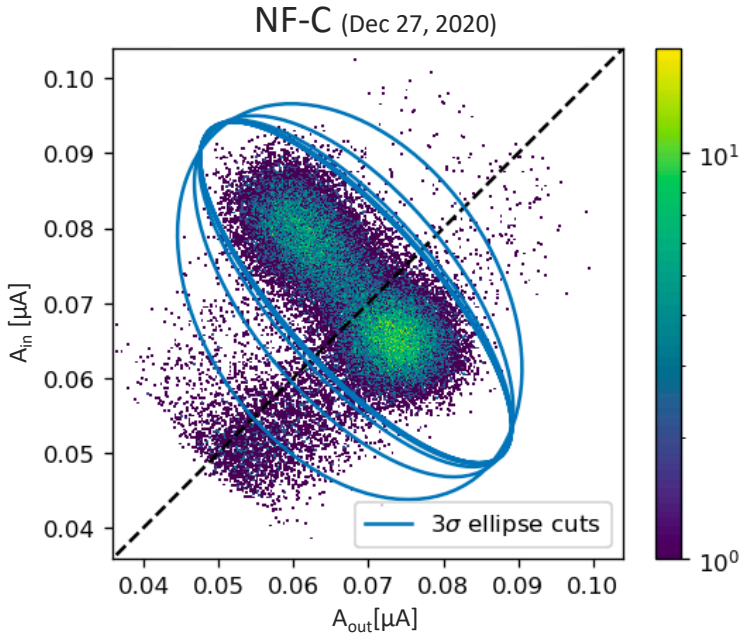


Figure 3.20: The result of using primary component analysis to iteratively fit an ellipse to the 1-e^-h^+ peak in one NF-C dataset. The heat map displays the number of events in a given bin. Each blue ellipse is one iteration of the cut (starting from the largest ellipse). We see that the cut converges after about four iterations.

two random (and potentially dependent) variables. The model utilized is shown in Eq. 3.3.

$$\sigma^2_{\text{eV}} = \left(\frac{100\text{ eV}}{\alpha m + 1}\right)^2 \left(\alpha^2 a + b + 2\alpha c\right) \tag{3.3}$$

The squared prefactor accounts for the normalization of the weighted sum to 100 eV. Specifically, m accounts for the difference in average signal observed in the two channels. a, b, and c are proportional to the inner-channel variance, the outer-channel variance, and the inner-outer-channel covariance, respectively. All four values were extracted via the fit and used to calculate $\alpha_{\text{optimized }\sigma}$ (the result of minimizing Eq. 3.3 with respect to α).

$$\alpha_{\text{optimized }\sigma} = \frac{mb - c}{a - mc} \tag{3.4}$$

Figure 3.21 also displays the fit to this model (with residuals) with the resulting $\alpha_{\text{optimized }\sigma}$ in the legend.

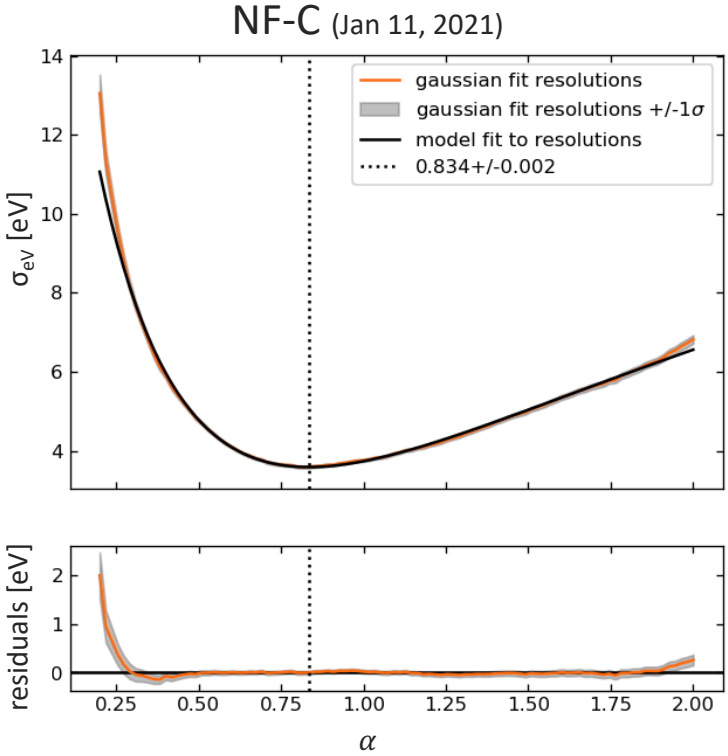


Figure 3.21: A typical result of applying the resolution-optimizing method to a dataset. (Top) The calculated resolution as a function of channel weighting (α) is shown in orange (with grey uncertainties). The fit to the resolution as a function of α (using Eq. 3.3) is shown in black. The α that gives the minimum resolution (using Eq. 3.4) is marked with a vertical dotted line and labeled in the legend.

The $\alpha_{\text{optimized }\sigma}$ for NF-C and NF-H were seen to be stable over the entirety of data taking (even going back to December 2020). For R1, $\alpha_{\text{optimized }\sigma}$ changed sharply on January 11th. This was likely caused by the change in noise observed on that day. Otherwise, the R1 $\alpha_{\text{optimized }\sigma}$ was stable but with relatively more variance than in the other two detectors. $\alpha_{\text{optimized }\sigma}$ for each detector and dataset can be seen in Figure 3.22.

Weighted means of $\alpha_{\text{optimized }\sigma}$ were calculated for NF-C, NF-H, R1 (before January 11th), and R1 (after January 11th). These can also be seen in Figure 3.22. The NF-C and NF-H values were applied to calculate the total signal and associated RQs in the final analysis. The R1 $\alpha_{\text{optimized }\sigma}$ was considered too uncertain and was not used. Instead, an unweighted sum was used for R1. At the time this decision was made, it was already decided that the R1 detector would only be used as a veto detector (due to its inferior resolution and noise). Partition independence and energy calibration were not necessary in a veto detector.

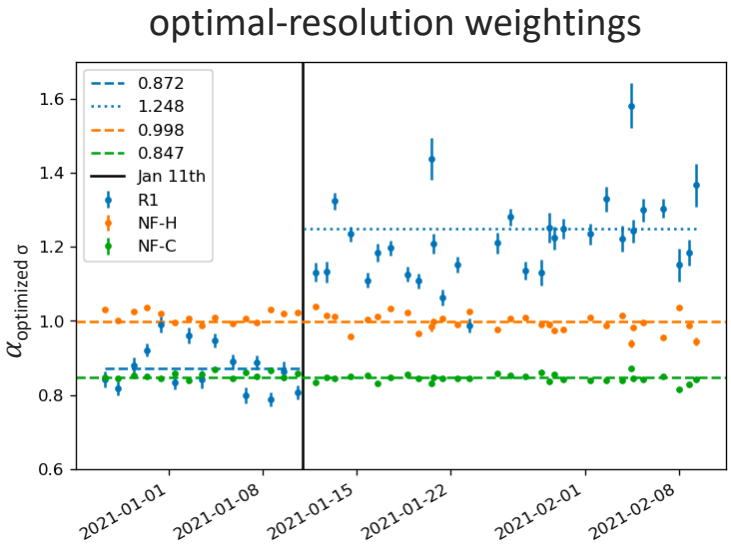


Figure 3.22: The result of calculating the resolution-minimizing channel weighting ($\alpha_{\text{optimized }\sigma}$) for each science dataset and each detector. Variance-weighted means were calculated for the NF-C and NF-H detectors (horizontal dashed lines). For the R1 detector, separate variance-weighted means were calculated for pre- and post-January 11th data (horizontal dashed and dotted lines). Variance weighting was used by mistake. Inverse variance weighting should have been used instead. The NF-C and NF-H results were shown to be valid anyways (see Figure 3.24).

The misguided emphasis of this method on resolution optimization was caught during the early stages of the analysis's inner-collaboration review. Even so, there was reason to believe that optimizing resolution may have also minimized the partition dependence. If the noise in both detector channels was subdominant to the variation in signal caused by varying energy partition, then partition variation would be the primary contributor to the signal variance of 1 $e^- h^+$ events. Figure 3.20 suggests this may be true for the NF-C detector. There are separate event clusters at either end of the distribution, each smaller than the total distribution. Presumably, the cluster size corresponds to the channel noise while the overall distribution (with a clear negative slope) is caused by the partition variation. In the case of partition-variation dominance, minimizing the 1 $e^- h^+$ resolution would also minimize the partition dependence.

3.6.2 Partition-Dependence-Minimizing Method

In order to test if the previously calculated weightings also minimized partition dependence, I proposed and applied a second method explicitly designed to minimize said dependence. We will now discuss the theory behind that method.

In an HVeV detector, each event's energy is divided between the inner and outer channels. Let us start by assuming the fraction of total energy (E) absorbed by the inner channel is equivalent to β. We will also assume that all the remaining energy is absorbed in the outer channel.

$$\begin{aligned} E_{in} &= \beta E \\ E_{out} &= (1-\beta)\, E \end{aligned} \tag{3.5}$$

The energy in each channel will be measured by QET's. Each channel will have different responsivity, dependent on channel biasing, number of TES's, etc. Here, we will represent responsivity with the conversion factor from absorbed energy to signal (C). Each channel will also have noise that appears in the observed signal.

$$\begin{aligned} A_{in} &= C_{in}E_{in} + Noise_{in} \\ A_{out} &= C_{out}E_{out} + Noise_{out} \end{aligned} \tag{3.6}$$

We can now apply the same definition of relative weighting factor (α) from Eq. 3.1.

$$\begin{aligned} A &\equiv \alpha A_{in} + A_{out} \\ &= \alpha\left(C_{in}\beta E + Noise_{in}\right) + C_{out}\left(1-\beta\right)E + Noise_{out} \end{aligned} \tag{3.7}$$

We want to choose α such that the observed total signal is independent of β.

$$\begin{aligned} \frac{dA}{d\beta} &= \alpha_{\beta\text{ indep}} C_{in}E - C_{out}E = 0 \\ \alpha_{\beta\text{ indep}} &= \frac{C_{out}}{C_{in}} \end{aligned} \tag{3.8}$$

Let us now consider what variation we would expect to see in each channel's signal among a group of mono-energetic events ($Var(E) = 0$). Ideally, each quantized $e^{-}h^{+}$ peak would count as one such group.

$$\begin{aligned} Var(A_{in}) &= Var(C_{in}\beta E + Noise_{in}) \\ &= C_{in}^2E^2Var(\beta) + 2C_{in}E\cancel{Cov(\beta, Noise_{in})} + \sigma^2_{noise,in} \\ &= C_{in}^2E^2Var(\beta) + \sigma^2_{noise,in} \end{aligned} \tag{3.9}$$

$$\begin{aligned} Var(A_{out}) &= Var(C_{out}(1-\beta)E + Noise_{out}) \\ &= C_{out}^2 E^2 Var(1-\beta) + 2C_{out}E\cancel{Cov(1-\beta, Noise_{out})} + \sigma_{noise,out}^2 \\ &= C_{out}^2 E^2 Var(\beta) + \sigma_{noise,out}^2 \end{aligned} \tag{3.10}$$

In Eqs. 3.9 and 3.10, we assume the channel noise is independent of event β. We can use these results to calculate $\alpha_{\beta\,\text{indep}}$.

$$\begin{aligned} \frac{Var(A_{out}) - \sigma_{noise,out}^2}{Var(A_{in}) - \sigma_{noise,in}^2} &= \frac{C_{out}^2}{C_{in}^2} \\ &= \alpha_{\beta\,\text{indep}}^2 \end{aligned} \tag{3.11}$$

While it might seem like one could instead extract C_{in}/C_{out} by fitting a line to the distribution in Figure 3.20, the process of doing so would really be the same as the first method. The resulting value would minimize the width perpendicular to the line which is equivalent to the resolution of the weighted sum of both channels. If the detector had dramatically different noise in its two channels, the line fit would be effected by the difference as well as the partition variation. The result would be equivalent to Eq. 3.11 with $\sigma_{noise,out}$ and $\sigma_{noise,in}$ assumed to be zero. The new method is independent of differences in channel noise.

For Run 3, I calculated $\sigma_{noise,in}^2$ and $\sigma_{noise,out}^2$ using random-triggered OF0 data. I again applied the iterative 3σ-ellipse cut to remove outlier events. I did not apply a χ^2 cut, because the intention was to measure noise and χ^2 cuts remove noisier events. An example dataset's OF0 random-trigger events in the NF-C detector can be seen in Figure 3.23. I followed the same procedure to select the 1 e^-h^+ events from triggered science data (also in Figure 3.23). I assumed the remaining 1 e^-h^+ events were mono-energetic and used them to calculate the inner and outer signal variances.

As with the previous method, I calculated the weighting factor ($\alpha_{\beta\,\text{indep}}$) for each detector and dataset (starting from January 11). The results can be seen in Figure 3.24. I performed the calculation using both OF0 and OFL amplitudes to calculate the signal variances. The results were found to be consistent with each other. Additionally, the results for NF-H and NF-C were observed to be stable and within

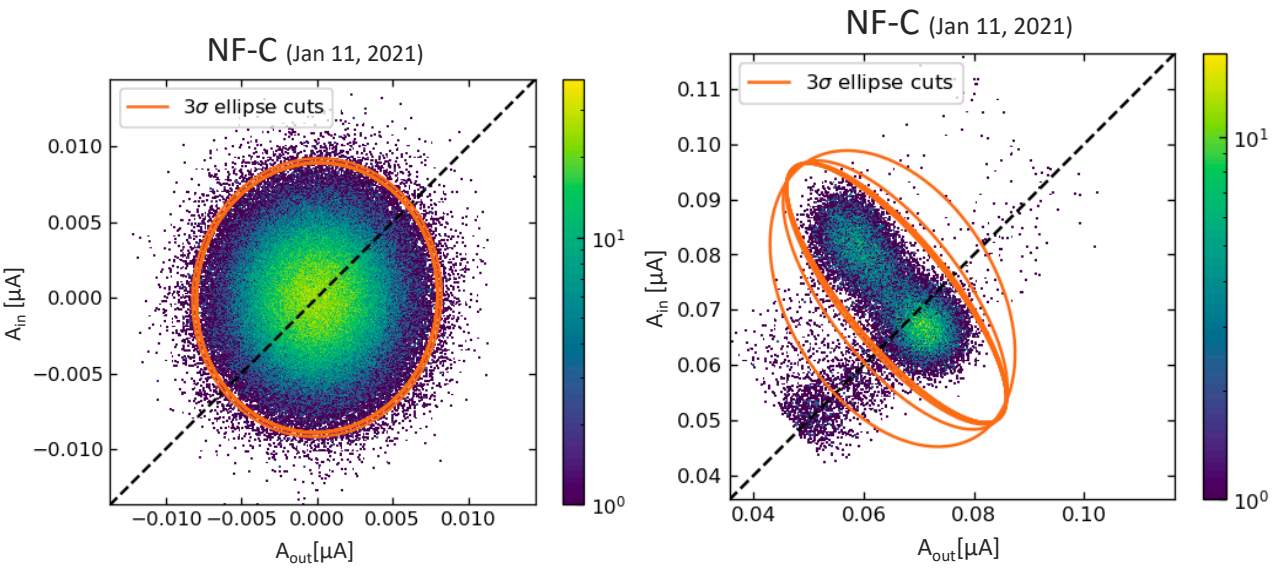


Figure 3.23: The result of using primary component analysis to iteratively fit ellipses to the 0-e^-h^+ (Left) and 1-e^-h^+ (Right) peaks in one NF-C dataset. For the 0-e^-h^+ peak, I used randomly triggered OF0 data. For the 0-e^-h^+ peak, I performed the calculation separately with OF0 and OFL GF-triggered data (OF0 shown here). We again see that the cut converges quickly (immediately for the 0-e^-h^+ peak).

~2% of the previous method's weightings (see Figure 3.22). Rather than being a coincidence, this is likely due to the 1 e^-h^+ variance being dominated by varying energy partition in those detectors. We determined the difference to be negligible and continued using the previous method's weightings (rather than reprocessing late in the analysis). The new method's R1 weighting was stable but significantly different from the previous method's. The difference was likely due to the R1 detector's higher noise. R1 was still only used as a veto detector and its channels were left unweighted.

3.7 Energy Calibration Using Laser Data

With the detector channels properly weighted, we proceeded to calibrate the detectors using laser data. The goal was to convert the calculated event OFL amplitudes (in units of current) to units of energy deposited in the detector substrate. We performed this calibration procedure for both NF-C and NF-H. R1 was only used as a veto detector and therefore did not require a full calibration. Later, we also decided to use NF-H as a veto only (due to its worse coincidence cutting ability). Hence, the plots in this section focus on NF-C.

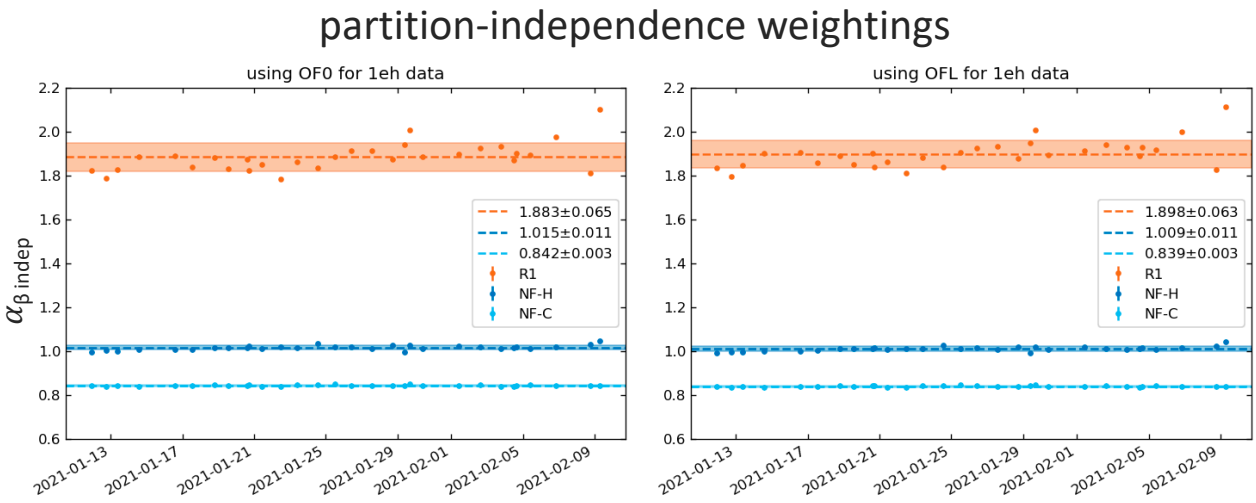


Figure 3.24: The result of calculating the partition-dependence-minimizing channel weighting ($\alpha_{\beta\ \mathrm{indep}}$) for each science dataset (January 11th onward) and each detector. Inverse-variance-weighted means were calculated for each detector (horizontal dashed lines). The NF-C and NF-H values were considered close enough to the results in Figure 3.22 that the old values were kept in use. The R1 value changed significantly from the previous result, but no relative weighting (α=1) was used for R1 (which was only a veto detector).

3.7.1 Low-λ Laser Data

Laser sidewall events presented a major impediment to using the laser data for calibration. Such events have amplitudes slightly greater than the amplitudes in each quantized peak. This led to the appearance of "tails" of higher-amplitude events attached to each peak which were not observed in previous runs' laser data nor the Run 3 science data. The sidewall events that comprised these tails were of uncertain energy and could therefore not be used in the calibration process. Even neglecting the tails proved difficult given the influence they had on any fit of the peaks.

It was observed during the run that laser data taken with higher laser λ produced peaks with relatively larger tails. In this discussion, λ is the Poisson parameter that best describes the mean number of e^-h^+ pairs produced by laser photons during each laser activation. Laser λ could be adjusted by varying the power supplied to the laser diode driver. On February 8th and 9th, we took laser data with particularly low λ in order to produce datasets with smaller tails that could be fit for calibration. The improvement in shape can be seen in Figure 3.25.

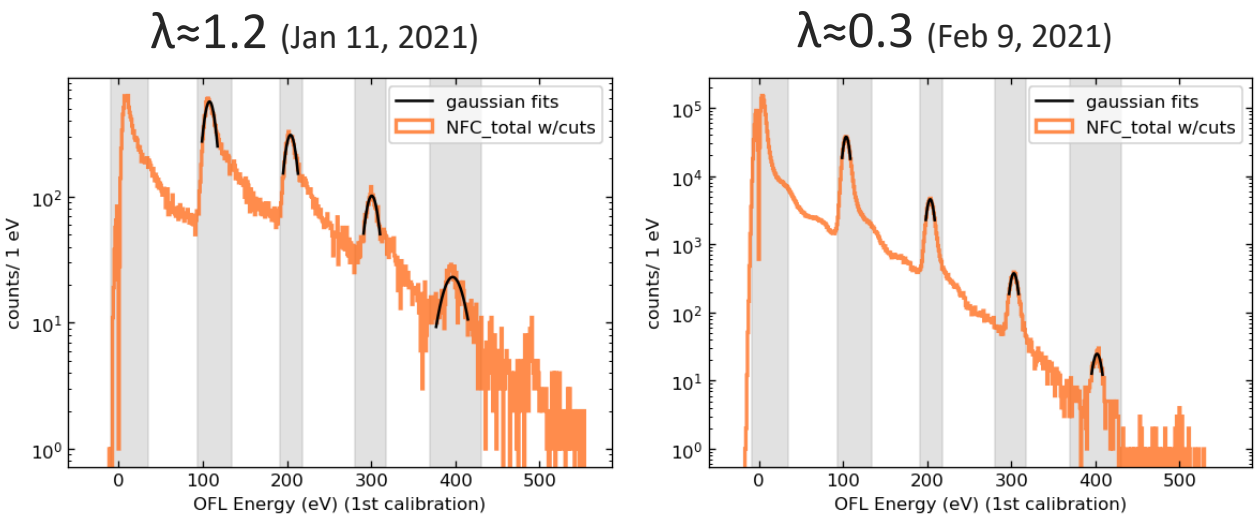


Figure 3.25: Two example laser spectra. In both plots, orange is the data in 1 eV bins with all live-time and calibration cuts applied. Energy on the x-axis is from a preliminary calibration but is very similar to the final energies. Grey regions are the initial search regions for peak fitting. Black curves are Gaussian distributions fit to each peak. (Left) Laser data with Poisson $\lambda > 1$. We see that each peak has a significant high-energy tail that could affect the Gaussian fit. (Right) Laser data with $\lambda < 0.5$. The between-peak regions are still sloped, but the underlying Gaussian peaks are clear for fitting.

By definition, the low-λ laser setting produced relatively few events in the higher e^-h^+ peaks. Having low statistics in the higher peaks makes them difficult to fit for calibration. Peaks with too few events could not be fit at all and would limit the upper energy range of the calibration. To improve the high-e^-h^+ statistics, one low-λ dataset was taken overnight with a higher frequency of laser activations (200 Hz rather than the previously used 20 Hz). This dataset achieved its goal of having sufficient statistics up to the 4 e^-h^+ peak. Unfortunately, the 200 Hz rate produced a laser trigger every 5 ms. This meant there were 3 triggers in every 13.1 ms trace. To account for the pileup, the overnight dataset received a special set of cuts designed to remove events with neighboring events of non-zero amplitude. Since the dataset had low λ, a large fraction of laser-triggered events had zero amplitude and the cut left sufficient data. The overnight dataset and its specialized cuts are shown in Figure 3.26.

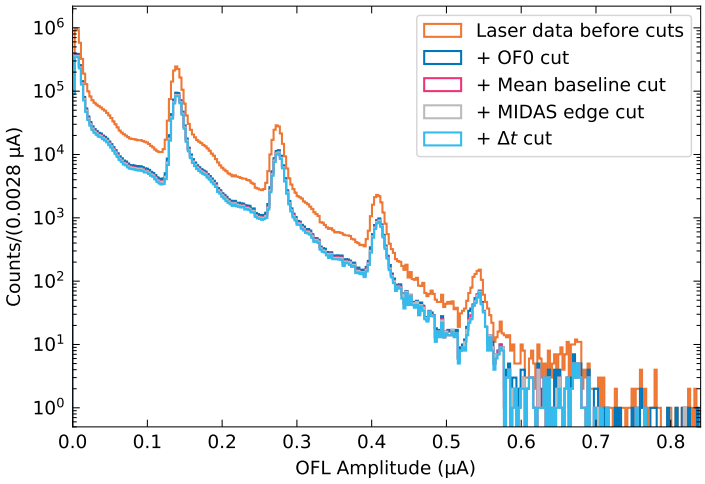


Figure 3.26: The low-λ laser dataset taken overnight with various cuts applied in order. The OF0 cut removed events with an adjacent laser-triggered event that had significant OF0 amplitude ($\gtrsim$12 eV). The mean-baseline cut removed events with changing baseline. The edge cut removed laser triggers that were too close to the edge of their time stream to be fully recorded. The Δt cut removed events with any GF-triggered (laser or non-laser) event within 7 ms. In this plot, the OF0 cut had the most significant effect. The Δt cut should remove many of the same events. The cuts did not significantly change the shape of the spectra. This is expected since the probability of a laser event having no neighbors should be amplitude independent.

Overnight-dataset cuts included an OF0 cut, which removed events with an adjacent laser-triggered event that had significant OF0 amplitude ($\gtrsim$12 eV). We also applied a Δt cut to remove events with any GF-triggered (laser or non-laser) event within 7

ms. A mean baseline cut was used to remove events with changing baseline (which may indicate an event occurring on the tail end of a previous event). An edge cut was used to remove laser triggers that were too close to the edge of their time stream to be fully recorded.

3.7.2 Daily Responsivity Variation

Three total laser datasets were used for the calibration. In addition to the overnight dataset from February 9th, we used two shorter low-λ datasets taken on February 8th. All three datasets were taken at the tail end of the run. We needed to test whether the detector responsivity was stable enough for such a calibration to be accurate for the whole run. As an estimator for detector responsivity, we calculated the 1-e^-h^+ event amplitude in each science dataset and compared it to the amplitude in all science datasets combined. The 1-e^-h^+ peak was chosen because it always had high statistics. The result can be seen in Figure 3.27.

The 1-e^-h^+ amplitude was stable, only varying by about 0.5% throughout the run. We modeled the distribution as a sum of asymmetric Gaussian distributions[2] with locations and widths defined by the individual dataset amplitudes and uncertainties (see Figure 3.27). We used the model to estimate the 68% confidence interval on daily responsivity. The confidence interval was calculated using the following conditions:

1. Include at least 68% of data generated by the model.
2. Include all bins with occupancy $\geq$ N.
3. The interval must be simply connected (practically defined as having no gaps between included bins greater than 0.005% of the combined 1-e^-h^+ amplitude).
4. Initially assume N = max(bin heights) and reduce N until all other conditions are met.

In the NF-C detector, the 68% bounds on variation in responsivity were +0.11% and -0.12% from the combined value. This information was later used when calculating the uncertainty on calibrated energy (see Section 3.7.5).

[2] Asymmetric distributions were used to represent the asymmetric uncertainties on each dataset's amplitude.

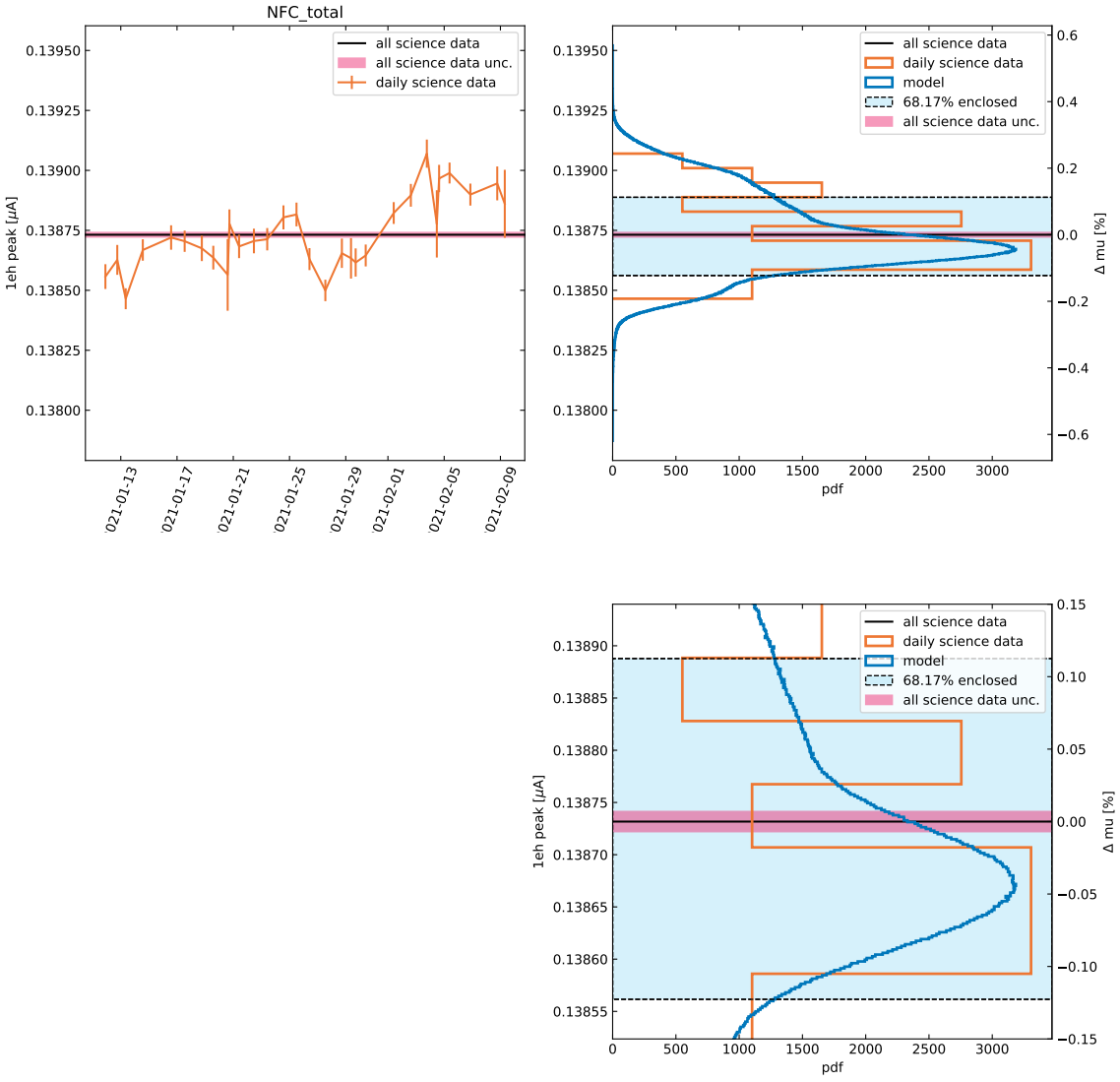


Figure 3.27: The measured variation in 1-$e^{-}h^{+}$ peak amplitude. (Top-left) The fitted amplitude of the 1-$e^{-}h^{+}$ peak in each science dataset. The vertical orange lines are individual amplitudes with asymmetric uncertainties. The horizontal black line is the result of fitting the 1-$e^{-}h^{+}$ peak in all science datasets summed together (with pink uncertainty). (Top-right) The individual amplitudes in histogram form. The blue model was produced by summing data from asymmetric Gaussian distributions representing each individual dataset with uncertainty. The blue region (enclosed in dashes) shows the 68% confidence region from the model. (Bottom-right) The same plot above zoomed-in to see the uncertainty and confidence region more closely.

To account for daily variation during the calibration calculation, the OFL and OF0 amplitudes from each of the three laser datasets were scaled by the combined 1-$e^{-}h^{+}$

amplitude divided by the 1-e^-h^+ amplitude from science data taken on the day that laser dataset was recorded (see Eq. 3.12). This scaling was only applied during the calibration calculation and was done to ensure that the resulting calibration parameters would be representative of the typical detector responsivity.

$$A'_{OF} = \frac{A_{1e^-h^+,\text{ combined}}}{A_{1e^-h^+,\text{ Feb 8 or 9}}} A_{OF} \tag{3.12}$$

3.7.3 Peak Fitting and Background Estimation

The Run 3 detectors were calibrated by fitting the quantized peaks (in current amplitude) to the expected phonon energy generated during laser-photon events. To obtain the current-amplitude peak locations, we fit each laser dataset's peak distributions (after scaling as in Eq. 3.12) to Gaussian distributions. Peak fitting was performed for each laser dataset separately. We used an un-binned log-likelihood fitting method to ensure the fits were not dependent on the choice of binning. The fitting was performed iteratively with the initial fit assuming that all events within a manually estimated range of amplitudes belonged to the peak of interest. Subsequent fits would only include events within an updated range derived from the Full Width at Half Maximum (FWHM) of the previous fit. The range selection was found to converge after only a few iterations. The careful event selection was done to minimize the influence of between-peak events on the fit. In a given dataset, if a peak was found to have fewer than 20 events in its FWHM, that combination of peak and dataset was not used in the calibration.

We fit peaks 1-4 using OFL amplitudes. As mentioned in Section 3.5.2, OFL could not be used to analyze the 0-e^-h^+ events (laser-triggered events with signals consistent with noise). Instead, fits were performed for the 0- and 1-e^-h^+ peaks in OF0 data, and their ratio was used to estimate the amplitude of the 0-e^-h^+ peak in OFL data. One might assume that the 0-e^-h^+ peak should have zero amplitude. In practice, we found that each of the three laser datasets had 0-e^-h^+ amplitude equivalent to between 1 and 3 eV. This offset was attributed to the presence of non-quantized events (such as sidewall events or direct absorption of photons by a TES). Example fits using OFL and OF0 data can be seen in Figure 3.28. The mean value (μ) of each Gaussian fit was used as the associated peak's current amplitude.

It was observed that the between-peak background had a significant dependence on energy (see the negative slopes visible in Figure 3.28). To estimate the effect of this background on the Gaussian fits, we fit decaying exponentials to the average value

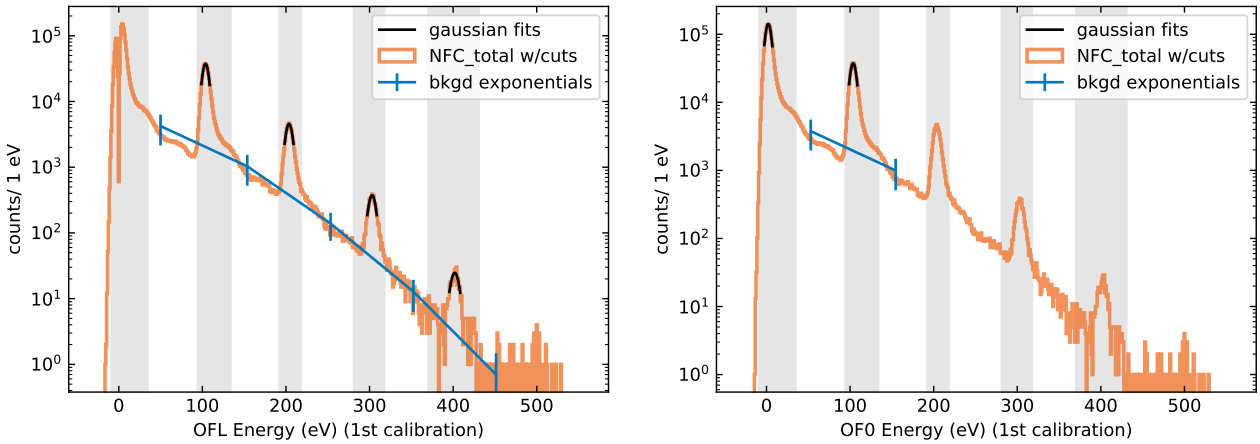


Figure 3.28: The overnight laser data with peaks fit to Gaussian distributions (see Figure 3.25). Shown in blue are the background estimates made using the between-peak data. (Left) The result using OFL data to fit peaks 1-4. (Right) The result using OF0 data to fit peaks 0-1.

of the between-peak regions around each peak. We then subtracted this estimated background from each Gaussian fit and used the change in maximum location as an additional systematic uncertainty on μ. The background estimations can also be seen in Figure 3.28. The added uncertainty due to this background was less then 0.05% for each peak.

3.7.4 Calibration Fitting

The final calibration required fitting the extracted peak locations to the phonon energy expected from laser-photon events. The laser diode used for Run 3 produced photons with wavelength of 635±5 nm (1.95±0.015 eV). The E_{gap} of silicon at low temperature is known to be ~1.1 eV. Since the photon energy was between 1 and 2 E_{gap}, we knew that each photon depositing energy in the substrate produced exactly 1 e^-h^+ pair. Therefore, the first e^-h^+ peak was the result of events with total phonon energy equivalent to 1.95 eV plus the NTL-phonon energy produced by 1 e^-h^+ pair. Similarly, the nth e^-h^+ peak was caused by simultaneous absorption of n photons and represented events with $n\times$ the first peak's phonon energy (see Eq. 3.13).

$$E_{\text{peak } n} = n(e\Delta V + 1.95eV) \tag{3.13}$$

where ΔV is the detector bias voltage. For Run 3, all science and laser data utilized was taken with ΔV = 100 V.

Additionally, we knew there was an offset in energy visible in the non-zero location of the 0-e^-h^+ peak. Since this offset was likely caused by non-quantized photon events, it was reasonable to expect that it would affect all peaks in some way. Specifically, we expected the offset to be constant and shift all peaks by the same amount. This was due to our belief that the number of photons in the detector chamber was significantly greater than 4 (the highest quantized peak we were interested in). In that case, the number of photons that produced fully quantized events did not reduce significantly the number of photons available to add non-quantized energy. The tails observed in laser data (discussed at the beginning of Section 3.7) along with the payload geometry (discussed in Section 3.3.3) convinced us that there were a large number of photons produced during each laser activation. To account for this offset, the peak locations were fit to the energies predicted by Eq. 3.14. HVeV Run 2 assumed a similarly constant offset.

$$E_{\text{peak } n} = n(e\Delta V + 1.95eV) + \text{offset} \tag{3.14}$$

The constant offset should therefore be equivalent to the 0-e^-h^+ peak energy. We estimated this value by converting the 0-e^-h^+ peak location in current amplitude to energy using an approximate conversion factor. The factor utilized was equivalent to the difference between the 1st and 0th peaks in energy divided by the same value in current amplitude (Eq. 3.15). The offset was calculated separately for each dataset since it should be laser-λ dependent.

$$\text{offset} \approx A'_{OFL,\text{ peak 0}} \left(\frac{e\Delta V + 1.95eV}{A'_{OFL,\text{ peak 1}} - A'_{OFL,\text{ peak 0}}} \right) \tag{3.15}$$

With the peak values known in both energy and OFL current amplitude, we could now fit to a calibration equation. Run 2 fit to a second order polynomial with the constant term removed. The constant term could be removed since the zero amplitude events must calibrate to zero energy. For Run 3, we found that a third order polynomial (with constant term removed) produced a better result without over-fitting (defined as producing a χ^2/ndf significantly below unity). The fit equation is shown in Eq. 3.16.

$$E = aA'_{OFL} + bA'^2_{OFL} + cA'^3_{OFL} \tag{3.16}$$

For the fit, the input uncertainties on current amplitude were those output by the peak Gaussian fits combined with the shift from the estimated background subtraction (both described in Section 3.7.3). For each dataset, the uncertainty on energy was equivalent to the uncertainty on offset. Offset uncertainties were propagated from uncertainties output by the peak Gaussian fits to $A'_{OF0,\text{ peak }0}$, $A'_{OF0,\text{ peak }1}$, and $A'_{OFL,\text{ peak }1}$. Additional uncertainties that could have been considered include the uncertainty on laser-photon energy (±0.015 eV/photon) and the uncertainty on the daily variation factor (in Eq. 3.12).

The results from all three of the laser datasets utilized were fit to the equation simultaneously. For NF-C, the shorter February 8th datasets only had sufficient statistics to fit peaks 1-3. The overnight February 9th dataset had sufficient statistics to fit peaks 1-4. In total, this meant the fit was performed with 10 data points. We felt comfortable fitting results from separate days simultaneously due to the stability of data taking and the daily variation being taken into account (see Section 3.7.2). The result of the fit can be seen in Figure 3.29. Error bars representing uncertainties on the input data points are plotted but too small to see (except in the one obvious case).

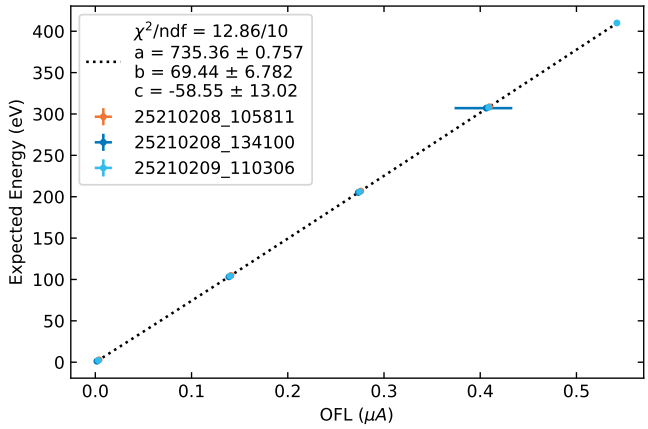


Figure 3.29: The result of fitting Eq. 3.16 to the fitted peak locations in three laser datasets. "25210208_105811," "25210208_134100," and "25210209_110306" are each low-λ laser datasets. "25210209_110306" is the overnight dataset and the only one with sufficient statistics to include the 4-e^-h^+ peak.

The resulting fit parameters were then used with Eq. 3.16 to calculate the energy of each NF-C event using the event's OFL amplitude in current units.

3.7.5 Residuals and Uncertainties

On the scale seen in Figure 3.29, the calibration fit looks ideal. The data points visually overlap the fit line and the χ^2/ndf is a little over unity. Figure 3.30 shows the fit in better detail by zooming in around each peak's data points. The green uncertainty band shown in Figure 3.30 displays one standard deviation of uncertainty from the fit result. The The orange uncertainty band displays the systematic uncertainty from the 68% daily responsivity variation measured in Section 3.7.2. Recall that 68% of the daily variation was within +0.11% and -0.12% of the combined value. Therefore, the orange band covers those variations from the fit value. The pink uncertainty band shows the previous two uncertainty bands combined in quadrature.

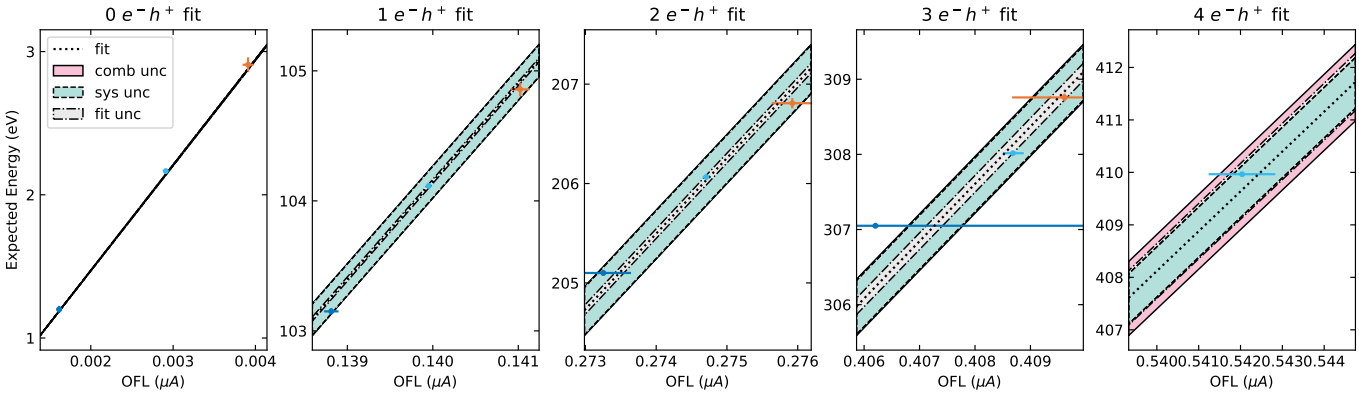


Figure 3.30: The results in Figure 3.29, zoomed in around each e^-h^+ peak being fit. Various uncertainty bands are also shown. The fit-uncertainty band ("fit unc") shows one standard deviation from the best fit. The systematic-uncertainty band ("sys unc") shows the 68% variation from Figure 3.27. The combined-uncertainty band ("comb unc") is calculated by combining the previous two in quadrature. Below 4 e^-h^+, the combined uncertainty is dominated by the systematic uncertainty. For 4 e^-h^+, the fit and systematic uncertainties contribute approximately equally.

Figure 3.31 shows the residuals of the fit as a function of calibrated energy. The same uncertainty bands as in Figure 3.30 are included. We see that all data points are inside or reasonably close to the combined uncertainty band. The 3-e^-h^+ data point for the second February 8th dataset appears to be an outlier but is acceptable when one considers its large uncertainty on current amplitude (easier to see in Figure 3.30).

The combined uncertainty band was used as the energy-dependent uncertainty on calibrated energy during dark-matter constraint setting. This is a change from Run 2, which assumed the uncertainty was ±0.5 eV for events of all energies.

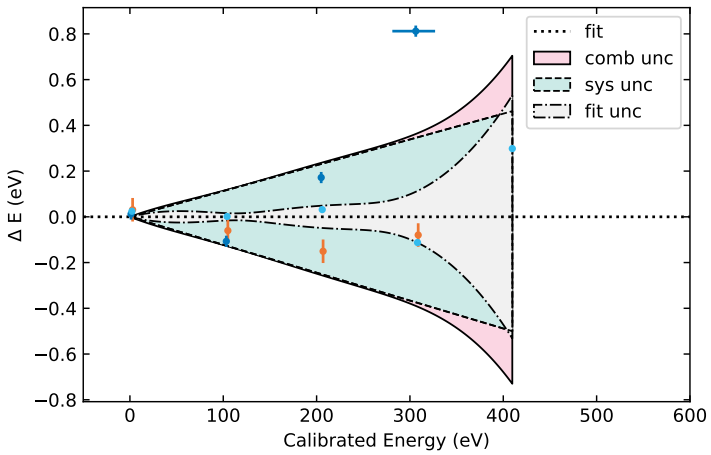


Figure 3.31: The same result and uncertainty bands described in Figure 3.30 plotted as residuals. The horizontal uncertainties (representing the peak-position uncertainties) were converted from units of current to energy using their end-point values in current. Since the conversion from current to energy is nonlinear (see Eq. 3.16), the converted uncertainties depend on the method of conversion. Converting the uncertainties by magnitude would yield different values. It is therefore preferable to use Figure 3.30 to visualize the horizontal uncertainties.

3.7.6 Calibration Extrapolation

During limit setting, we wanted to use events in peaks 1-4. This meant that we actually needed the energy calibration to be valid for energies up to 440 eV. This bound was calculated by adding the 400 eV from the NTL-phonon energy of a 4-e^-h^+ event to the maximum energy our ionization model predicted would generate a 4-e^-h^+ event (24 eV). We also added 3× our upper bound on energy resolution (3×5.49 eV) to account for the use of $3\sigma_E$ windows when counting events in each peak. Overall, extrapolating from 410 eV (the previous maximum energy as seen in Figure 3.30) to 440 eV constituted a 30 eV, or 7.3%, extrapolation.

To test if the previously described calibration was valid to 440 eV, we loosened the requirement during Gaussian fitting that each peak have 20 events. This allowed us to fit the 5-e^-h^+ peak in the overnight laser dataset reasonably well. Figure 3.32 shows the results of extrapolating the calibration fit and combined uncertainty to the fitted 5-e^-h^+ peak.

We see that the 5-e^-h^+ peak (at ~512 eV) is about 1.5 standard deviations away from the extrapolated prediction. The difference is equivalent to about 3 eV. If we naively assume that the misalignment grew linearly from 410 eV, this would indicate an uncertainty of 30 meV per eV of extrapolation. We would then expect the 440

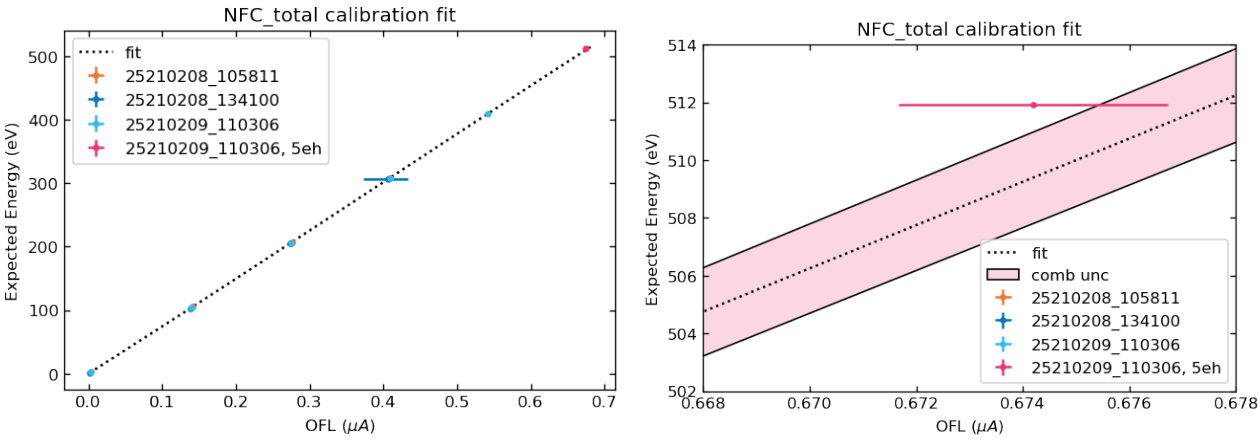


Figure 3.32: Extrapolation of the calibration fit to higher energies. (Left) The same data and fit in Figure 3.29 but extrapolated to the 5-e^-h^+ peak. The result of applying the Gaussian fit to the 5-e^-h^+ peak in the overnight dataset is also shown ("25210209_110306, 5eh"). (Right) A zoom-in around the 5-e^-h^+ peak showing that the extrapolated valued is only ~3 eV away from the peak fit. The extrapolation of the combined uncertainty is also shown.

eV events to be wrong by less than 0.9 eV (or 0.2%). We determined this error to be acceptable and used the calibration up to 440 eV.

3.8 Live-Time Selection

Live-time selections were used to select which time periods of data to consider in the final analysis.

3.8.1 Time-Stream Overlap Cut

Due to a data-taking bug, a small fraction of the 0.5 second time streams described in Section 3.4 overlapped in time. Overlapping time streams were removed from consideration if the extent of the overlap was large enough that the same event could be observed in both. On its own, this cut only removed 0.006% of data.

3.8.2 Mean-Baseline cut

Another cut was used to remove events with abnormal mean baseline (the unfiltered channel current when no events were occurring). Changes in baseline can indicate unwanted changes in detector performance and are therefore undesirable. Total mean baseline in each detector was calculated by summing the individual channel currents with the relative channel weightings calculated in Section 3.6. The baseline cut was calculated separately for each science dataset by fitting its mean-baseline

distribution to a Gaussian function. The cut then removed data that was more than 3σ away from the central value. An example mean baseline cut can be seen in Figure 3.33.

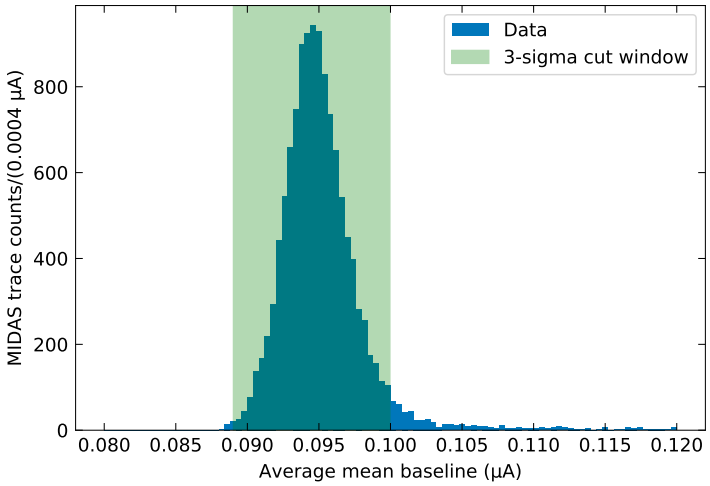


Figure 3.33: Histogram of mean-baseline values observed in one science dataset. The green region shows which time-stream baselines survived the cut. The region is determined by fitting a Gaussian to the distribution and cutting at $\pm 3\sigma$.

Applied independently, the mean-baseline cut would remove 6.0% of the total live time.

3.8.3 Temperature Cut

As mention in Section 3.3.2, the fridge temperature was extremely stable throughout the run. The temperature only changed during the software reset every 2-3 weeks. Otherwise, the temperature remained within 0.1 mK of the designated 10.5 mK. In the considered time (between January 11th and February 9th), there were two software resets (on January 15th and February 7th). These resets produced two datasets with temporary temperature differences that were removed with the temperature live time cut. These two series accounted for 4.4% of the total live time.

3.8.4 Coincidence and Δt Cut

The most substantial live-time cut was the aforementioned coincidence cut enabled by running multiple detectors simultaneously. The plots in Figure 3.34 were produced to visualize the coincidence rates between different detectors. For these plots and all matters concerning the coincidence cut, only R1 events with signal greater than 0.01 μA (equivalent to ~40 eV) were considered triggers. This was to prevent

the large number of noise events triggered in R1 from affecting the study. For NF-C and NF-H, all GF-triggered events were considered. The plots indicated a positive correlation in event rate (within half-second time streams) between the R1 and NF-C detectors. No correlation was visible between the NF-H detector and either of the others (in these plots).

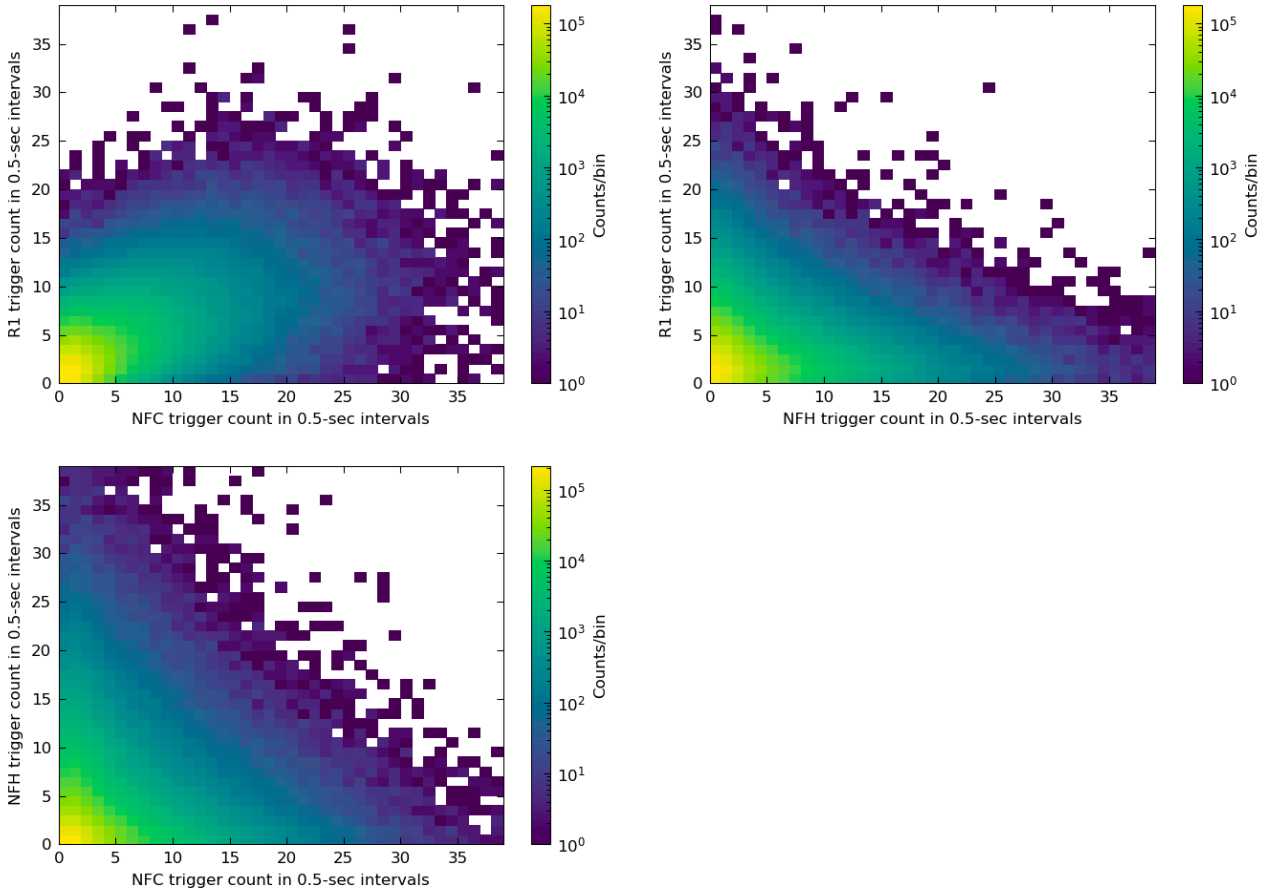


Figure 3.34: 2D histograms of trigger count within 0.5-second time streams. We see that time streams with high rate in the NF-C detector were likely to also have high rate in the R1 detector (and vice versa).

We analyzed coincidence more closely by plotting the distribution of time between adjacent triggered events. Figure 3.35 shows the Δt distribution for events triggered in the NF-C detector. Δt was defined as the distance in time between an event and the nearest triggered event in a specified detector. The Δt distribution was calculated between NF-C and each of the three detectors (including NF-C itself). Δt between NF-C and R1 showed a high rate of nearly simultaneous events. There was also a high rate of NF-C events that appeared near in time to other NF-C events. Such events included the traditional burst events (originally observed as multiple events in

a single detector). We also expected some pile-up events within each detector due to randomly occurring backgrounds. Removing such pile-up events was another goal, because their overlapping shapes make them difficult to analyze. Again, significantly less coincidence was observed with the NF-H detector.

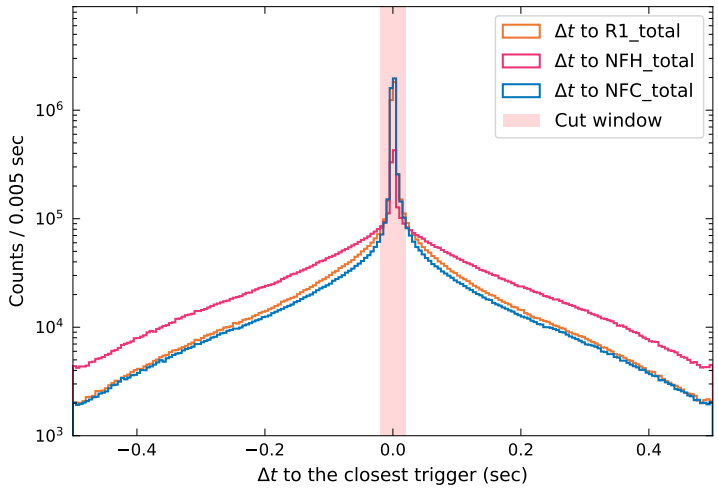


Figure 3.35: The distribution of times between each triggered event in the NF-C detector and the nearest triggered event in each other detector (Δt). We see a high rate of low Δt events between the NF-C detector and both itself and the R1 detector. The coincidence cut window (±20 ms) is indicated by the red shaded region. Events within the window will be removed with the coincidence live-time cut. Data was taken from all science datasets (January 11th onward.)

We attributed the enhanced coincidence between NF-C and R1 to the fact that they were spatially closer to each other than to NF-H (see Figure 3.13). In order to maximize the effect of the coincidence cut, we decided not to use data from the NF-H detector in the final analysis (except as a veto for the NF-C detector). As mentioned previously, we decided to do the same with the R1 detector due to its worse energy resolution and noise behavior (compared to NF-C and NF-H). This left us with one science detector (NF-C) and two veto detectors (one to either side of NF-C).

The actual coincidence cut was applied by removing 20 ms of live time (from the science detector) before and after each event in any detector. This meant that only events greater than 20 ms away from all others would be kept. Two events that were within 20 ms of each other would both be removed. This cut window is also shown in Figure 3.35. The 20 ms value was chosen to balance the reduction of coincidence

rate with the loss of live time. If applied on its own, this coincidence cut would remove 22.9% of the live time.

3.8.5 Live-Time Cut Summary

The total effect of live-time cuts on the calibrated science data can be seen in Figure 3.36. The combined cuts removed 30.6% of the live time. This does not include the effect of removing the R1 and NF-H detectors nor of removing the data before January 11th. Those were considered analysis decisions and not cuts of any type. The cuts lowered the January-11th-onward exposure time (in NF-C) from 13.14 to 9.12 days.

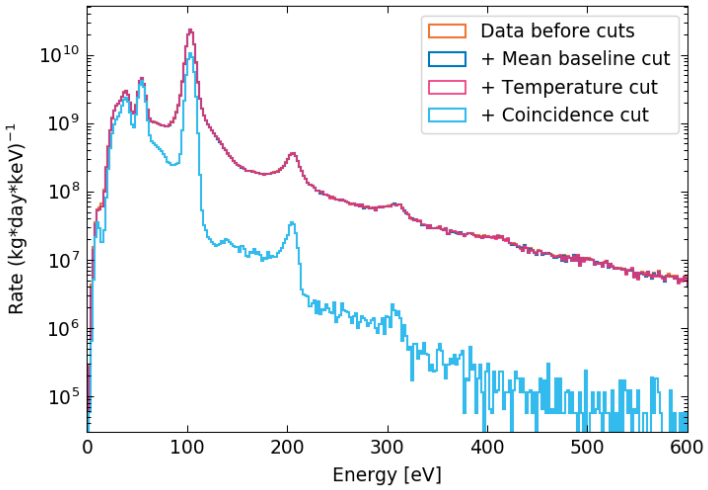


Figure 3.36: Each live-time cut applied in succession to the full Run 3 science data. The overlapping time-stream cut had a very small effect and was therefore included in all the spectra (but is unnoticeable). We see that the coincidence cut had the most significant effect on the post-cut spectrum and was the only one to alter the shape.

We see that the baseline and temperature cuts had no effect on the spectrum shape. On the other hand, the coincidence cut had a large effect on both the shape and total rate. The 1-e^-h^+ event rate decreased by a factor of 2. This was equivalent to removing $\sim 1 \times 10^{10}$ DRU of 1-e^-h^+ events[3]. While the higher-peaks decreased by substantially less in absolute rate, their relative decreases in rate were greater. For example, the 2-e^-h^+ peak decreased by a factor of $\sim$10 (but only $\sim 4 \times 10^8$ DRU). Therefore, while the coincidence cut mostly removed low-energy events (as expected), it did not resolve the issue of a low-energy excess relative to the rates

[3]DRU stands for differential-rate unit and is equivalent to 1/(kg $\times$ day $\times$ keV).

in the higher e^-h^+-pair peaks (the relative peak heights remain similar to those in Figure 3.11).

3.9 Data-Quality Cuts

In addition to live-time cuts, two data-quality cuts were applied to the Run 3 data. In this section we will discuss each cut and its effect on the data.

3.9.1 χ^2 Cut

One cut was implemented on events' frequency-domain reduced χ^2 in order to remove events that were poorly fit by the signal template. Such events could consist of signal-like glitches or pile-up events. Specifically, the cut used a modified version of the reduced χ^2 that only considered frequencies up to 30 kHz. Higher frequencies contained significant noise and were less useful for characterizing the signals of interest (visible from Figure 3.18).

The χ^2 cut was produced using OFL data from the first half of the overnight laser dataset. Events within 3σ of each quantized peak were selected. For each peak's selected data, a cumulative data distribution (CDF) of the event's reduced χ^2 was calculated. Each CDF was then used to identify the reduced χ^2 for which 95% of events in that peak had superior fits. The 95% values as a function of peak energy were then fit to Eq. 3.17.

$$\chi^2_{\text{cut}}(E) = AE^n + B \tag{3.17}$$

The resulting function can be seen overlaying the overnight laser data in Figure 3.37.

The function was then used as Run 3's energy-dependent χ^2 cut. By design, it should have a 95% passage fraction. This was confirmed by applying the cut to the second half of the overnight laser dataset. The resulting cut efficiency (defined as the ratio of events surviving the cut to events before the cut) was calculated to be $94.71\% \pm 0.04\%$.

3.9.2 $\Delta\chi^2$ Cut

The other quality cut applied was the $\Delta\chi^2$ cut. This was a new type of cut that had not been used in any previous HVeV analysis. $\Delta\chi^2$ was defined as the difference between the usual reduced χ^2 and the reduced χ^2 of a two-signal fit. This definition of $\Delta\chi^2$ produces a value that is higher for events that are more likely to be pile up. The two-signal fit assumed two summed signals of the same signal shape discussed

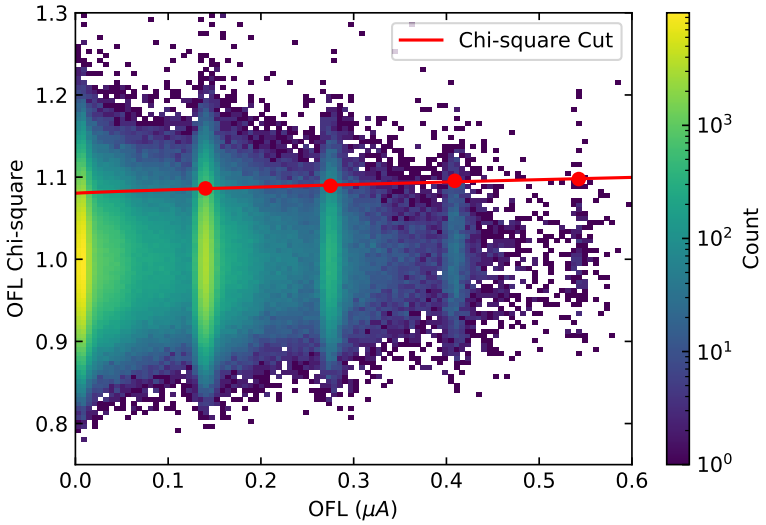


Figure 3.37: The 2D histogram of OFL χ^2 vs. amplitude for the first half of the overnight laser dataset. The energy-dependent 90% χ^2 cut is shown in red with individual peak values marked with dots.

in Section 3.5.2. The two signals' amplitudes were allowed to vary independently. One signal was forced to be within 0.4 ms of the trigger. The other signal was forced to be within 0.4 ms of the first. The two-signal fit was meant to identify pile-up events that were too close in time to be identified with separate triggers or removed with the χ^2 cut. The χ^2 cut was found to be effective in removing events further than 0.4 ms apart, so a broader time allowance was not necessary.

The $\Delta\chi^2$ cut was calculated using the exact same method and data as the χ^2 cut. The only difference was the RQ of interest ($\Delta\chi^2$ instead of χ^2). The $\Delta\chi^2$ cut overlaying laser data can be seen in Figure 3.38.

The $\Delta\chi^2$ cut was also designed to have 95% passage fraction. The actual efficiency was calculated to be 94.84 ± 0.03%.

3.9.3 Combined Quality-Cut Effect

The overnight laser dataset was also used to calculate the efficiency of the combined χ^2 and $\Delta\chi^2$ cuts. The resulting efficiency can be seen in Figure 3.39. The combined efficiency was calculated to be 90.05% ± 0.054%. This is almost exactly the individual efficiencies multiplied together, indicating that the cuts removed different sets of events. This was confirmed by testing that the $\Delta\chi^2$ cut did not reduce the

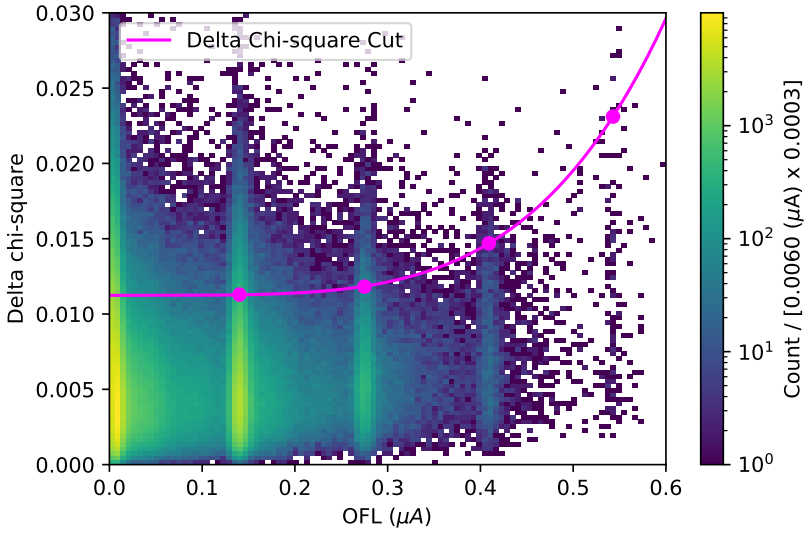


Figure 3.38: The 2D histogram of OFL $\Delta\chi^2$ vs. amplitude for the first half of the overnight laser dataset. The energy-dependent 90% χ^2 cut is shown in purple with individual peak values marked with dots.

higher end of the χ^2 distribution (and vice versa). $\Delta\chi^2$ and χ^2 were also shown to be uncorrelated.

The effect of the quality cuts on the NF-C science data can be seen in Figure 3.40. We see that the quality cuts remove a significant fraction of between-peak events while leaving the peaks intact. They specifically make the 3-e^-h^+ peak easier to identify.

3.10 Detector Energy Resolution

The energy resolution of the NF-C detector was calculated in a variety of ways. Baseline resolution was measured by fitting Gaussian functions to OF0 data from randomly triggered events (with all cuts applied). This was done separately for the overnight laser data and the combined science data. Higher e^-h^+-peak resolutions were calculated using the same method on triggered OFL data. The various resolution values and uncertainties can be seen in Figure 3.41.

Laser data was observed to have worse (larger) resolution for the baseline and peaks. This was likely a result of fitting to the non-Gaussian distributions produced by the laser sidewall events. The resolution was also observed to degrade for higher peak numbers. The same effect was observed in HVeV Run 2 but was never well

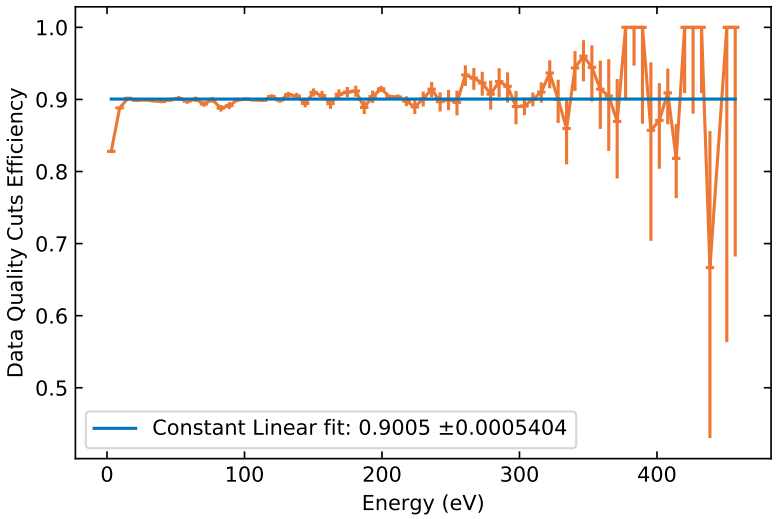


Figure 3.39: The Run 3 cut efficiency calculated using the second half of the overnight laser dataset. Orange data is the calculated efficiency with uncertainties. Each point was calculated by dividing the number of events surviving both the χ^2 and $\Delta\chi^2$ cuts by the number of pre-cut events (in each bin). The constant fit used for the limit-setting efficiency is shown in blue (with uncertainty quoted in the legend).

modeled. Rather than attempt to precisely model the effects of sidewall events and the energy dependence on resolution, we decided to use the central value of calculated resolutions with a high uncertainty. This decision was motivated by the belief that the observed variation in energy resolution would not have a large effect on the final limits. This was later confirmed with the limit results (see Section 3.12.7).

The largest calculated energy resolution (from the laser 4-e^-h^+ peak) was used as the upper bound. The smallest calculated energy resolution (from the science-data baseline) was used as the lower bound. The nominal value was taken to be the average of the upper and lower bounds. This led to an NF-C resolution of 4.26 ± 1.23 eV being used during limit setting. The resolution and uncertainty are also shown in Figure 3.41.

To ensure that energy resolution was stable during the run, the 1-e^-h^+-peak resolution was calculated for each science dataset individually. The results (compared to the combined-science-data resolution) can be seen in Figure 3.42. No trend in time was observed. The total variation was only about ±5%. This variation was deemed sufficiently small (much smaller than the total proposed uncertainty).

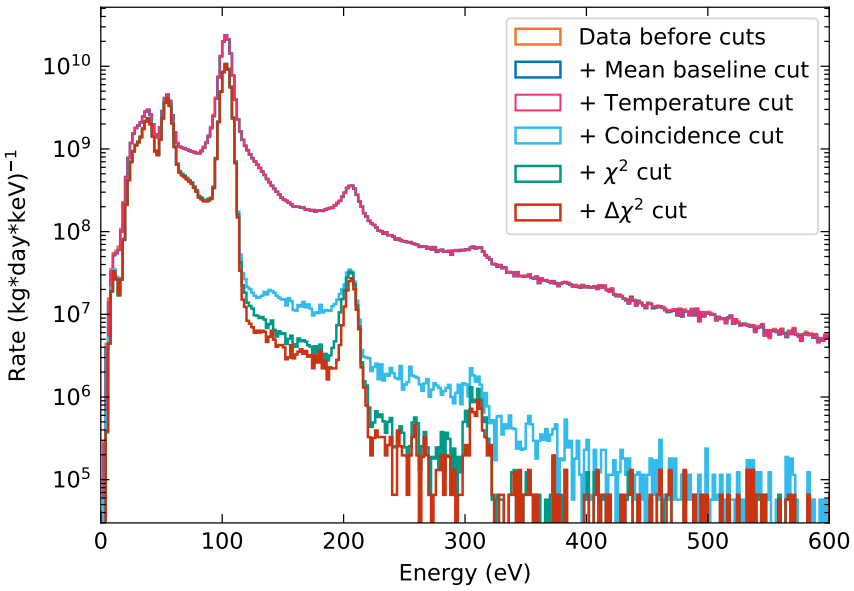


Figure 3.40: The Run 3 live-time and quality cuts applied sequentially to the full science data. We see that the coincidence cut still has the greatest effect on total event rates. The χ^2 and $\Delta\chi^2$ cuts are effective in lowering the rate of between-peak events.

3.11 Charge Trapping and Impact Ionization Model

For HVeV Run 2, laser data was fit to a model of charge trapping and impact ionization (CTII) in order to extract the associated probabilities and include both effects in the detector-response model. Since CT and II generate between-peak events, fitting such a model required data with uncontaminated between-peak distributions. In Run 3, the high rate of laser sidewall events made extracting these parameters from laser data infeasible. Instead, we noted that the NF-C detector was fabricated on the exact same substrate as the Run 2 detector and should have equivalent CTII behavior under the same biasing conditions. Therefore, we fit the CTII model for Run 3 to laser data from Run 2 and used the resultant probabilities when calculating the Run 3 limits.

The Run 2 CTII model assumed flat distributions of CT and II events between each peak. The Run 3 model used more accurate distributions produced by analytically modeling charge-carrier trajectories with CT and II taken into account. This model was referred to as the "exponential" CTII model since it assumed the probabilities exponentially decreased with distance from the original ionization depth (to account for charges being trapped before reaching greater distances). Although Runs 2 and

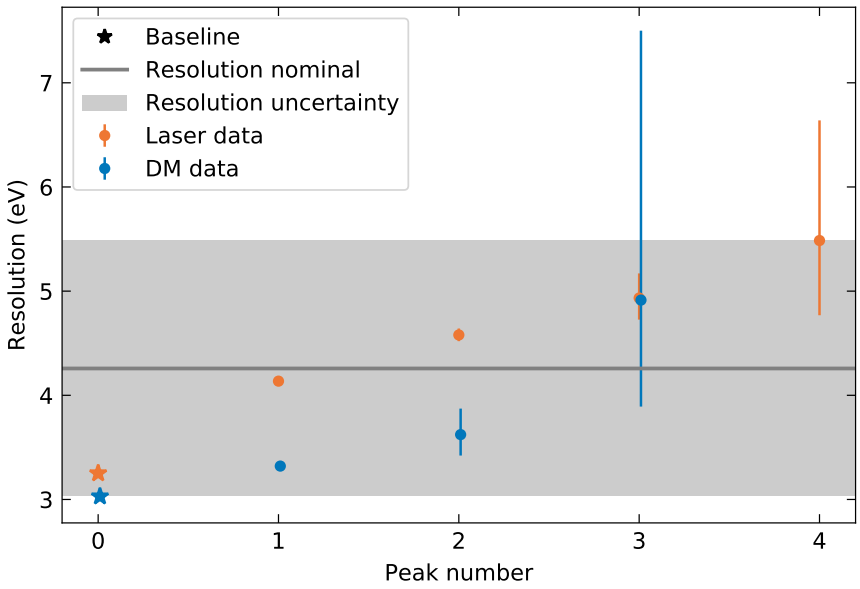


Figure 3.41: The energy resolution calculated by fitting Gaussian distributions to different peaks in various datasets. Baseline resolution was calculated using randomly triggered OF0 data. Higher peaks were calculated using OFL data. The laser data used was from the overnight laser dataset. The horizontal grey band encloses all calculated resolutions and represents the uncertainty on resolution used for limit setting. The horizontal grey line is the center of the band and was used for the official limit calculation. The resolution as a function of energy could be described using a linear fit, but doing so would not improve the Run 3 analysis (which only weakly depends on energy resolution below ~10 eV).

3 used the same laser data to estimate their CTII probabilities, the use of slightly different models lead to different results. The Run 3 (Run 2) CT probability was 12.8±1.5% (13±2%) for both electrons and holes. The Run 3 (Run 2) II probability was $1.6^{+1.8}_{-1.6}$% (1±1%) for production of an electron or a hole by an electron or a hole. For II, all four possibilities (e^- produces h^+, e^- produces e^-, h^+ produces e^-, and h^+ produces h^+) were assumed to be equal.

3.12 Poisson Limit Setting

To quantify the sensitivity of HVeV Run 3 (specifically the improvement from Run 2), we produced 90%-confidence-level limits using the Poisson limit-setting method. During Run 2, this method was found to handle uncertainties better than the Optimum Interval method used for Run 1 [26]. The Poisson method required assuming the number of observed events followed the Poisson probability mass function (PMF) shown in Eq. 3.18. The PMF returns the fractional probability

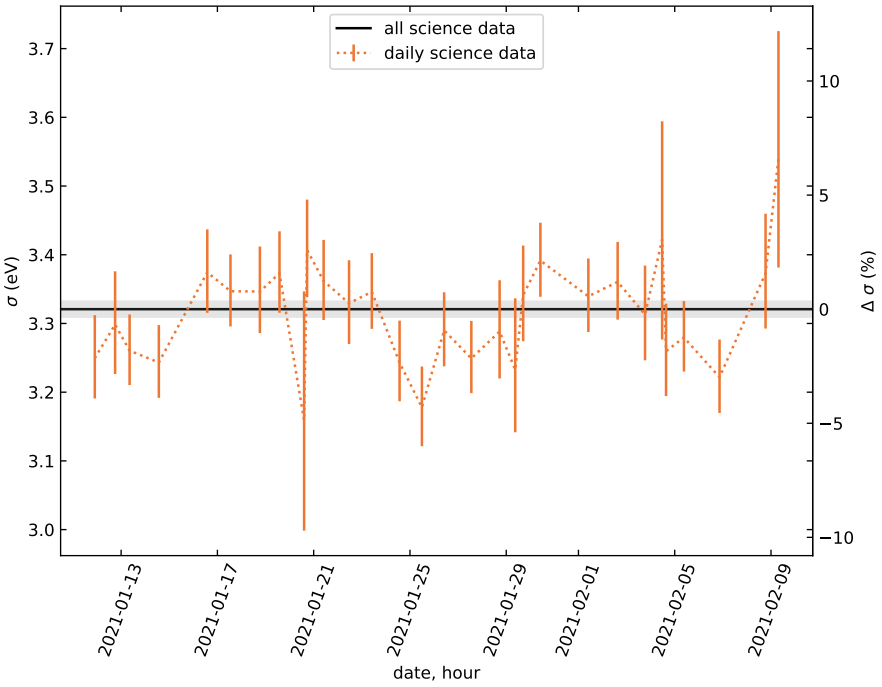


Figure 3.42: The fitted width of the 1-$e^{-}h^{+}$ peak in each science dataset. The results with uncertainties are shown as vertical orange bars. The fitted width in all science datasets summed together is displayed as a horizontal black line (with uncertainty in grey). We see that resolutions only vary (from the combined value) by about ±5%.

of observing k events in a counting experiment that measures λ events on average. The Poisson distribution is valid for events that occur with constant mean rate and independent arrival times. Both assumptions should be true for dark matter.

$$\mathrm{PMF}_{Pois}(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{3.18}$$

The Poisson cumulative distribution function (CDF) can then be used to place an upper limit on the underlying mean number of events with some confidence. The CDF returns the fractional probability of observing $\leq k$ events in a counting experiment that follows Poisson statistics with mean λ. The relationship between Poisson PMF and CDF is shown in Eq. 3.19.

$$\mathrm{CDF}_{Pois}(k;\lambda) = \sum_{n=0}^{k} \mathrm{PMF}_{Pois}(n;\lambda) \tag{3.19}$$

An upper limit is calculated by counting the number of observed events (N_{obs}) and solving Eq. 3.20 for λ_{lim} with CL equivalent to the desired confidence level (0.9 for 90%).

$$\mathrm{CDF}_{Pois}(N_{obs}; \lambda_{lim}) = 1 - \mathrm{CL} = 0.1 \tag{3.20}$$

We can then claim that the underlying mean value λ must be less than λ_{lim} with 90% confidence. Using the experiment's known efficiency and exposure, we can then convert λ_{lim} to a limit on the rate of dark-matter interactions. The interaction rate can then be converted to limits on various dark-matter interaction parameters using the dark-matter models combined with the model for detector performance.

3.12.1 Energy Regions of Interest

In practice, the described procedure should only be performed using events in the energy range where detector performance is well understood. Otherwise, uncertainties on cut efficiency, detector resolution, and calibrated energy will generate large uncertainties on the final limits. For Run 3, our ability to characterize the NF-C detector was limited by the laser-data energy range. Figures 3.28, 3.37, 3.38, and 3.41 show how the characterization was limited above the 4-e^-h^+ peak (due to low statistics). Accordingly, we only considered events between 50 and 440 eV. The range was set to include all events that produced between 1 and 4 e^-h^+ pairs. This range was smaller than Run 2, which used its superior laser data to consider data up to the 6-e^-h^+ peak.

Another source of uncertainty was the rate of events in between-peak regions. These regions consist of non-quantized events and are believed to be dominated by CT and II contributions. The uncertainties on CT and II probabilities create fractionally large uncertainties on the expected rates between peaks. Additionally, while the Run 3 CTII model was improved from Run 2, it was still not expected to be perfect. We therefore follow the procedure of Run 2 and count only events within windows around each quantized peak. Specifically, we count events within $\pm 3\sigma_E$ of each peak's expected location in the relevant dark-matter signal model.

The quantized nature of the energy spectrum also motivated setting limits using each peak individually. Summing all peaks would sacrifice information about event energies and degrade sensitivity. Specifically, sensitivity to models that are most likely to produce events in only one or a few peaks could be weakened by backgrounds in other peaks. Individual-peak limit setting takes advantage of the

HVeV detector's excellent ability to differentiate ionization peaks without having to include all the systematics of a full background model. For Run 3, this meant producing four distinct limits for each dark-matter model. As in Run 2, the separate limits were combined afterwards to give the final results.

3.12.2 Signal Models: Dark-Matter Rates

As in HVeV Run 2, the dark-matter models of interest were those of dark-matter-electron scattering (DMe), dark-photon absorption (DPA), and axion-like particle absorption (ALP). In this section, we will discuss the equations used to calculate the expected event rates from each model as a function of electron recoil energy (E_{er}).

The differential scattering rate used for DMe can be seen in Eq. 3.21. This rate was derived in [24] by considering the excitation of bound electrons in a static potential (the crystal structure) via scattering with dark matter from the local dark-matter halo.

$$\frac{dR_{DMe}}{d\ln E_{er}} = \frac{\rho_{DM}}{m_{DM}} \frac{1}{2m_{Si}} c\sigma_e \alpha \frac{m_e^2}{\mu_{DM}^2} I_{\text{crystal}}(E_{er}; F_{DM}) \tag{3.21}$$

Each parameter utilized is described below for reference.

- R_{DMe} is the rate of dark-matter-electron scattering in units of events/exposure.
- E_{er} is the recoil energy of the electron.
- ρ_{DM} is the local mass density of dark matter.
- m_{DM} is the mass of a single dark-matter particle.
- m_{Si} is the mass of a silicon nucleus.
- c is the speed of light in a vacuum.
- σ_e is the interaction cross section (the model parameter of interest for DMe).
- α is the fine structure constant.
- m_e is the mass of an electron.
- μ_{DM} is the reduced mass of the m_{DM}-m_e system.
- I_{crystal} is the (unitless) scattering integral for DM-electron scattering.

- F_{DM} is the momentum-transfer form factor, which determines the dependence of scattering on exchanged momentum (q). For Run 3, we consider $F_{DM} = 1$ and $F_{DM} = (\alpha m_e/q)^2$ for exchange of a heavy and light mediator, respectively.

As in Run 2, I_{crystal} was calculated using the QEdark code presented in [24] (made publicly available by the authors). For Run 3, the values were calculated using the dark-matter halo parameters recommended in [48]. This change (away from the Run 2 halo parameters) was made as part of a broader attempt to make SuperCDMS results directly comparable to those of other collaborations. The SuperCDMS, SENSEI, and DAMIC collaborations have all agreed to make use of these parameters. The Run 2 and Run 3 halo parameters are compared in Table 3.1.

Halo parameter	Run 2 value	Run 3 value	units
Local DM density	0.3	0.3	GeV/cm^3
Average Earth velocity w.r.t. DM halo	240	253.7	km/s
Average DM velocity w.r.t. galactic frame	230	238	km/s
Galactic Escape Velocity	600	544	km/s

Table 3.1: Table of halo parameters used for HVeV Runs 2 and 3. Except for density, the Run 2 values were taken from Section 6 of [24]. The Run 2 density was taken from [49]. The Run 3 values were taken from [48]. The Run 3 average Earth velocity (w.r.t. DM halo) was calculated using the average DM velocity (w.r.t. galactic frame), the solar peculiar velocity, and the Earth velocity (w.r.t the Sun) on March 9th.

The expected DPA rate for Run 3 was calculated using Eqs. 3.22 and 3.23. These equations were derived in [50] by modifying the equation for absorption of photons by electrons in a semiconductor. The equation was modified to account for the dark-photon number density, kinetic mixing (ε), and non-relativistic velocity.

$$R_{DPA} = \frac{1}{\rho_{Si}} \frac{\rho_{DM}}{m_\nu} \varepsilon^2_{\text{eff}} \frac{\sigma_1(\omega = m_\nu \frac{c^2}{\hbar})}{\hbar} \tag{3.22}$$

$$\varepsilon^2_{\text{eff}} = \frac{\varepsilon^2 m_\nu^2 c^4}{m_\nu^2 c^4 - 2 m_\nu c^2 \sigma_2 + \sigma_2^2 + \sigma_1^2} \tag{3.23}$$

Parameters that were not also used in Eq. 3.21 are described below for reference.

- R_{DPA} is the rate of dark-photon absorption in units of events/exposure.
- ρ_{Si} is the mass density of silicon.
- m_ν is the dark-photon mass. The dark photon is wholly absorbed, so $E_{er} = m_\nu c^2$.
- ε_{eff} is the (unitless) effective kinetic-mixing parameter in the silicon medium.
- $\sigma_1(\omega)$ and $\sigma_2(\omega)$ are the real and imaginary parts of the complex conductivity of silicon (in units of energy) as a function of frequency ω.
- $\hbar$ is the reduced Planck constant.
- ε is the (unitless) kinetic-mixing parameter in vacuum.

The expected ALP-absorption rate for Run 3 was calculated using Eq. 3.24. This equation was also derived in [50] by modifying the photon-absorption equation. The equation was similarly modified to account for the ALP number density, axioelectric coupling (g_{ae}), and non-relativistic velocity.

$$R_{ALP} = \frac{\rho_{DM}}{\rho_{Si}} \frac{3 g_{ae}^2 m_{ALP}}{16\pi\alpha m_e^2} \frac{\sigma_1(\omega = m_{ALP}\frac{c^2}{\hbar})}{\hbar} \tag{3.24}$$

Parameters that were not also used in previous equations are described below for reference:

- R_{ALP} is the rate of ALP absorption in units of events/exposure.
- g_{ae} is the axioelectric coupling factor.
- m_{ALP} is the ALP mass. The ALP is wholly absorbed, so $E_{er} = m_{ALP}c^2$.

Eq. 3.24 is slightly different from what was used in HVeV Run 2. The Run 2 equation can be found by setting $\sigma_1 \rightarrow \hbar c \sigma_{p.e.} \rho_{Si}$, where $\sigma_{p.e.}$ is the silicon photoelectric-absorption cross section in units of area/mass. The relationship between complex conductivity and photoelectric-absorption cross section (Eq. 3.25) can be used to

show that both equations are equivalent when the silicon index of refraction ($n(\omega)$) goes to unity.

$$\sigma_1(\omega) = n(\omega)\sigma_{p.e.}(\omega)\rho_{Si}\hbar c \tag{3.25}$$

The silicon $n(E = \hbar\omega)$ can be seen in Figure 3.43. We see that n goes to unity for deposited energies above 30 eV.

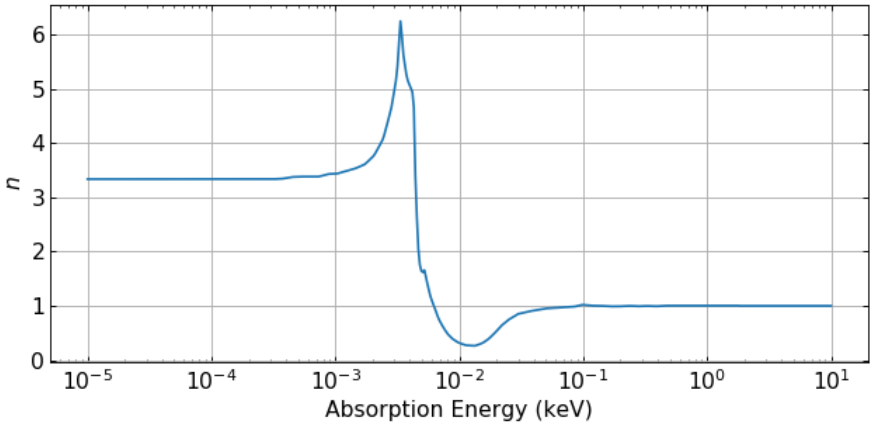


Figure 3.43: The silicon index of refraction as a function of photon energy ($E = \hbar\omega$). Figure is courtesy of Matt Wilson using data from [51].

Run 2 used the $\sigma_{p.e.}$ version of the Eq. 3.24 as a carry-over from the SuperCDMS Soudan dark-photon and ALP analysis [27]. The Soudan limit was set for ALP masses ≥40 eV and was therefore unaffected by the difference. We believe (following [50]) that Eq. 3.24 is the more accurate choice and ideally would have been used for Run 2. When comparing the Run 2 and 3 results, we scaled the Run 2 ALP limit to account for this difference.

For Run 3, we used silicon $\sigma_1(E = \hbar\omega)$ values curated by Matt Wilson (see Figure 3.44). Above 3.2 eV, these values were taken from literature ([51], [52], and [53]). Below 3.2 eV, the values were from measurements and a fitted model of indirect absorption [54]. The fitted model produced nominal, upper and lower results. We used the nominal result to calculate the Run 3 limits. We used the upper and lower results when estimating uncertainty on the limits.

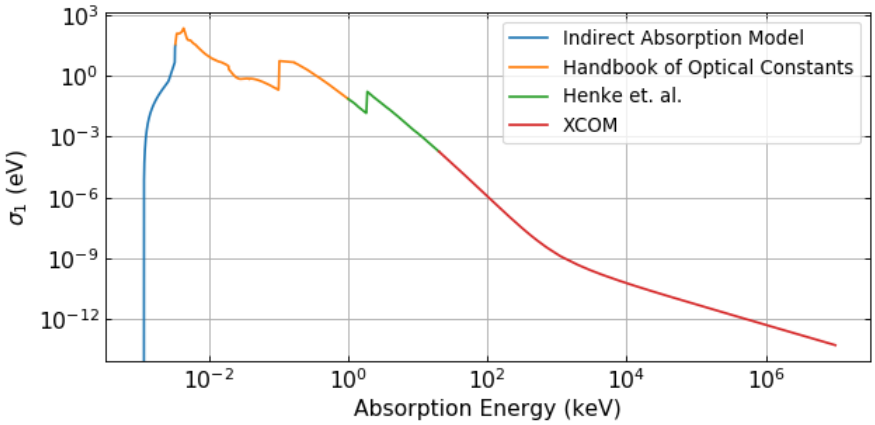


Figure 3.44: The real part of the complex conductivity of silicon used for the Run 3 absorption limits. Figure is courtesy of Matt Wilson. The colors and legend indicate where each region's data originated. The indirect absorption model is from [54]. The handbook of optical constants is [51]. The Henke et. al. paper is [52]. The XCOM data is from [53].

The silicon n and σ_1 values used for Run 3 (along with $\sigma_{p.e.}$ and all analogous values for germanium) can be easily loaded using a now publicly available software package[4].

3.12.3 Signal Models: Signal Ionization

Dark-matter models return the rate of events as a function of electron-recoil energy (E_{er}), but the total phonon energy observed by the detector (E_{det}) will be dominated by the energy from ionization-dependent NTL phonons. We therefore needed to understand how ionization, quantified by number of $e^- h^+$ pairs produced (n_{eh}), depended on E_{er}.

$$E_{det} = E_{er} + n_{eh} e \Delta V \tag{3.26}$$

The most basic ionization model consists of dividing E_{er} by the average energy per $e^- h^+$ pair produced for large energy depositions (ϵ_{eh}) and rounding down the resulting value to get n_{eh}. As our description implies, this method is only accurate for large E_{er} ($\gg E_{gap}$).

$$n_{eh} = \frac{E_{er}}{\epsilon_{eh}} \quad \text{for} \quad E_{er} \gg E_{gap} \tag{3.27}$$

Ionization models may also use a Fano factor (F) to define the variation in n_{eh} as a function of n_{eh}.

$$\sigma^2_{neh} = F n_{eh} \tag{3.28}$$

Since HVeV experiments are concerned with energy depositions near the silicon band gap, they require models that are more accurate at low energies. For Run 3, we used the ionization model described in [55]. This was another change from the signal model used in Run 2 that was adopted, in part, to make SuperCDMS results directly comparable to those of other collaborations. The new model was also expected to be more accurate for 12-50 eV recoils, which have historically been difficult to characterize due to a lack of sources capable of depositing such energies within detectors. The [55] analyzers navigated this issue by producing a phenomenological model and fitting its parameters to the available measurements

[4] `https://gitlab.com/supercdms/public/dmmodeling_opticalconstants`.

of ionization yield. The resulting model's dependence on E_{gap} (between 1.12 and 1.170 eV) was also considered.

For E_{er} below 50 eV, the [55] authors provided ionization probabilities produced via simulation for E_{gap} of 1.1692, 1.1627, and 1.1230 eV (see Figure 3.45). The parameter of interest is the probability to produce n-e^-h^+ pairs for a given recoil energy ($p_n(E_{er})$).

For Run 3, we used E_{gap} = 1.131 eV (the result of an updated fitting by Matt Wilson to the data from [54]). We interpolated the ionization results between 1.1627 and 1.1230 eV to obtain p_n for that E_{gap}. Karthik Ramanathan (of [55]) also produced a simulated result for E_{gap} = 1.134 eV that we used to verify the effectiveness of interpolation. The interpolated and simulated results for E_{gap} = 1.134 eV can be seen in Figure 3.46.

For E_{er} above 50 eV, the [55] authors provided an analytical method to calculate the ionization probabilities. The method uses functional forms for ϵ_{eh} and F (at high energy) that were fit to simulated results from the phenomenological model.

$$\epsilon_{eh} = 1.6989E_{gap} + 0.0843A + 1.2972, \quad E_{er} > 50eV \tag{3.29}$$

$$F = -0.0281E_{gap} + 0.0015A + 0.1383, \quad E_{er} > 50eV \tag{3.30}$$

where A is the ratio of phonon-carrier to carrier-carrier scattering. For Run 3, we used A = 5.2 from [55]. The coefficients shown in Eqs. 3.29 and 3.30 are of higher precision than those in [55]. The higher precision values were shared with us by Dr. Ramanathan and were required to recreate the results in [55]. We used the resultant ϵ_{eh} and F to calculate the ionization probabilities for E_{er} > 50 eV.

$$p_n(E_{er}) = \frac{1}{\sqrt{2\pi nF}} \exp\left(-\frac{1}{2}\left(\frac{n\epsilon_{eh} - E_{er}}{\epsilon_{eh}\sqrt{nF}}\right)^2\right), \quad E_{er} > 50eV \tag{3.31}$$

The ionization software used for Run 3 is now publicly available as a software package[5]. The package includes code I developed to use the [55] results as well as simpler ionization models.

[5] `https://gitlab.com/supercdms/public/dmmodeling_ehpairquantization`.

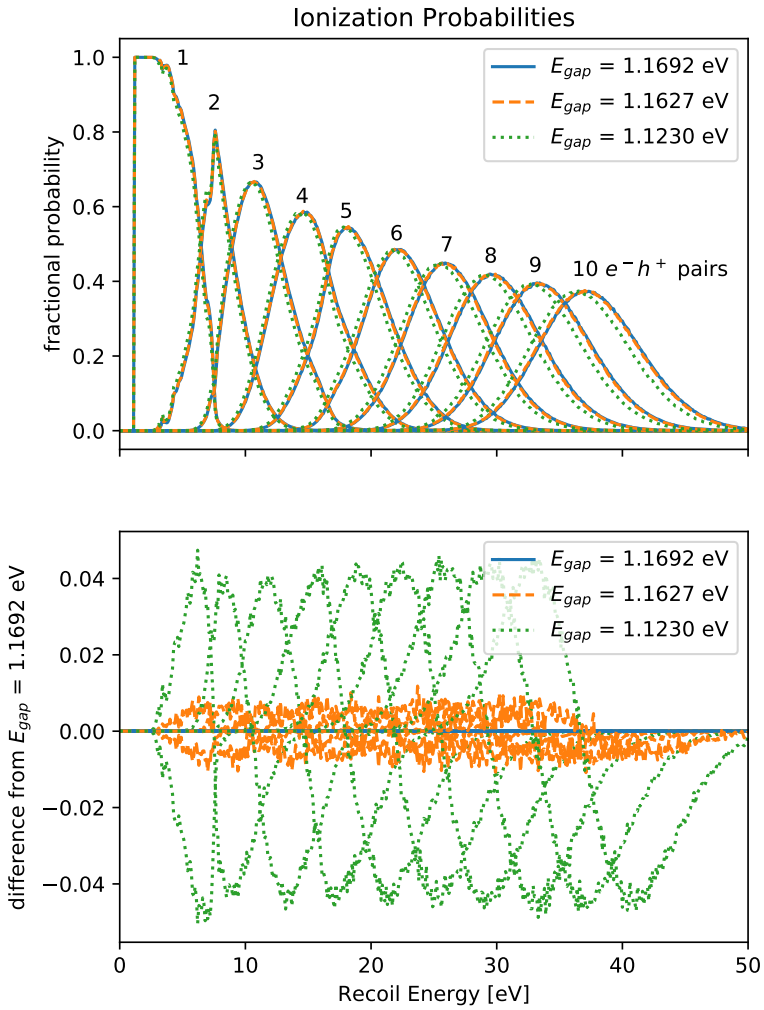


Figure 3.45: The ionization probability data used for Run 3. (Top) The pair-creation probabilities as a function of recoil energy for 1-10 e^-h^+ pairs. The results are from [55] and vary depending on the assumed value of E_{gap}. The probability also drops to zero below E_{gap}. (Bottom) The difference between probabilities for different E_{gap} values (compared to E_{gap} = 1.1692 eV).

3.12.4 Signal Models: Combined

Dark-matter and ionization models are combined to calculate the expected rate of events as a function of E_{det}. If CTII and detector resolution are ignored, the combined model is calculated using the following equation.

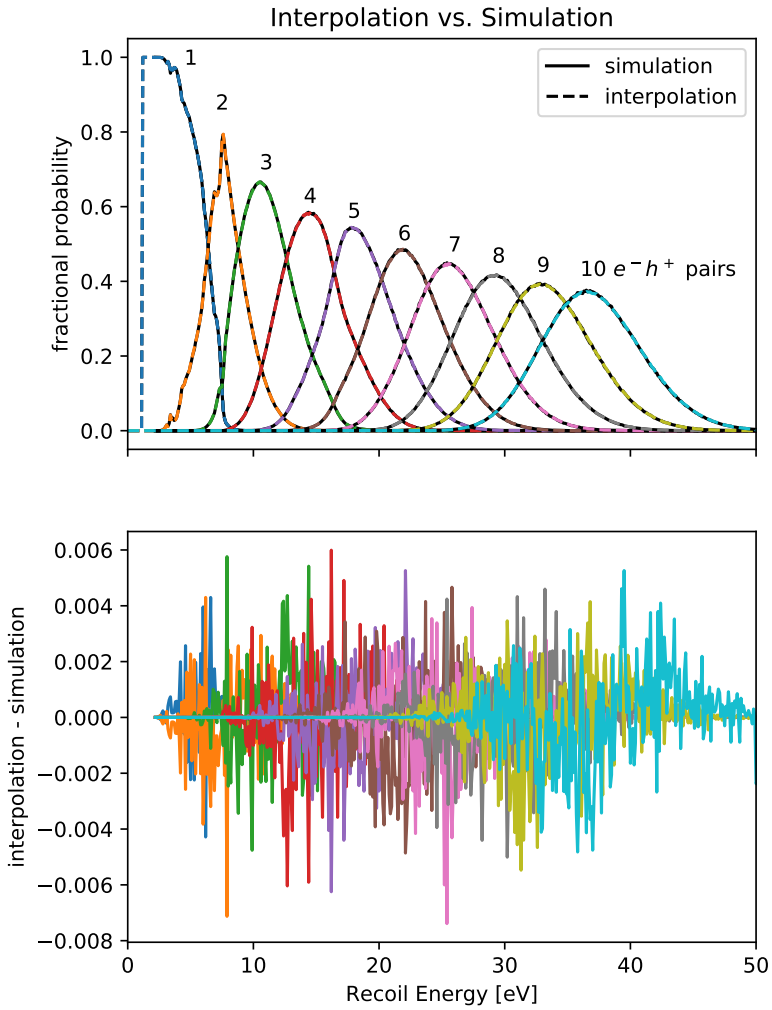


Figure 3.46: Results of testing the gap-based interpolation of ionization probabilities. (Top) The pair-creation probabilities for E_{gap}=1.134 eV calculated using interpolation and via simulation (courtesy of Karthik Ramanathan). (Bottom) The residuals between interpolated and simulated probabilities. We see that the difference is extremely small.

$$R_{det}(E_{det}) = \sum_{n=1} p_n \left(E_{det} - ne\Delta V\right) R_{DM} \left(E_{det} - ne\Delta V\right) \tag{3.32}$$

Theoretically, the sum over ionization (n) should be taken to infinity. For Run 3, we only summed up to n = 10. Without charge trapping, our energy range (up to 440

eV) would only allow us to observe events with initial ionization of ≤ 4 $e^- h^+$ pairs. We included up to 10 $e^- h^+$ pairs to account for significantly more charge trapping than was probable (CT probability of 12.8±1.5%).

In practice, the p_n values must be modified by the exponential CTII model before being convolved with the dark-matter model. The final models were also convolved with a Gaussian of width equivalent to the energy resolution of the NF-C detector. An example combined signal model for DMe with $F_{DM} = 1$ can be seen in Figure 3.47. This model accounts for ionization, CTII, and energy resolution.

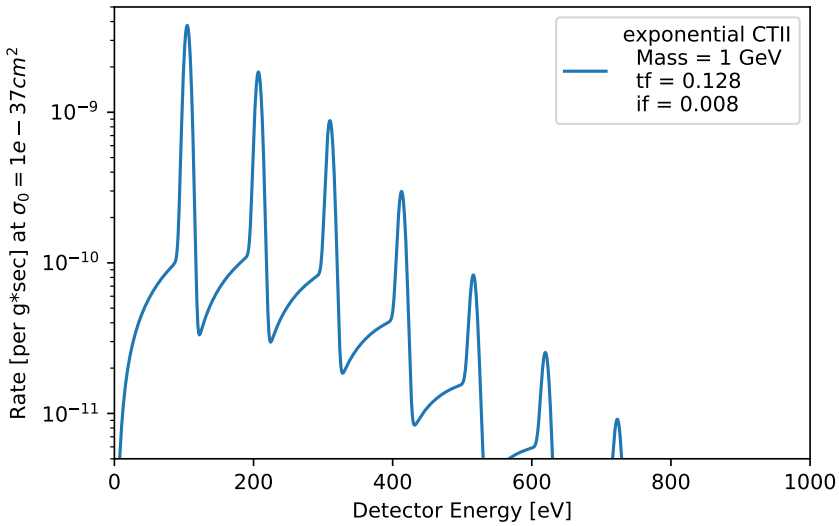


Figure 3.47: An example signal model for DMe ($F_{DM} = 1$) with m_{DM} of 1 GeV and energy resolution of 4.26 eV. Detector voltage bias was set to 100 V. Charge trapping and impact ionization fill in the between-peak regions. "tf" refers to the charge-trapping fraction. "if" refers to the impact-ionization fraction for both electron and hole production (0.8% probability of producing an electron, and a separate 0.8% probability of producing a hole). The model was produced assuming an interaction cross section (σ_0) of 1×10^{-37} cm^2.

3.12.5 Limit Setting

When setting the actual limits, each combined model was calculated using a reference interaction parameter. For DMe, the models were calculated with $\sigma_{e,0} = 10^{-37}$ cm^2. For DPA, the model was calculated with $\varepsilon_{\text{eff},0} = 5 \times 10^{-13}$. For ALP absorption, the model was calculated with $g_{ae,0} = 5 \times 10^{-11}$. The models were then multiplied by the detector efficiency (a function of E_{det}) and the post-live-time-cut exposure (a constant) to produce model spectra directly comparable to the final post-cut spec-

trum. The following procedure was then performed for each of the four quantized peaks in each model.

The model is used to calculate the average E_{det} of the quantized peak (E_{peak}). The model is then integrated from $E_{peak} - 3\sigma_E$ to $E_{peak} + 3\sigma_E$ to calculate the number of dark-matter events expected in that peak (N_0). Next, events within the same energy window are counted in the science data (to get N_{obs}). Then Eq. 3.20 is used to set an upper limit on this value (λ_{lim}). One of the following relationships (derived from the models in Section 3.12.2) is then used to extract the limit on the interaction parameter.

$$\begin{aligned} \sigma_e \propto R_{DME} &\rightarrow \frac{\sigma_{e,lim}}{\sigma_{e,0}} = \frac{R_{DME,lim}}{R_{DME,0}} = \frac{\lambda_{lim}}{N_0} \\ &\rightarrow \sigma_{e,lim} = \sigma_{e,0}\frac{\lambda_{lim}}{N_0} \end{aligned} \tag{3.33}$$

$$\begin{aligned} \varepsilon^2_{eff} \propto R_{DPA} &\rightarrow \frac{\varepsilon^2_{eff,lim}}{\varepsilon^2_{eff,0}} = \frac{R_{DPA,lim}}{R_{DPA,0}} = \frac{\lambda_{lim}}{N_0} \\ &\rightarrow \varepsilon_{eff,lim} = \varepsilon_{eff,0}\sqrt{\frac{\lambda_{lim}}{N_0}} \end{aligned} \tag{3.34}$$

$$\begin{aligned} g^2_{ae} \propto R_{ALP} &\rightarrow \frac{g^2_{ae,lim}}{g^2_{ae,0}} = \frac{R_{ALP,lim}}{R_{ALP,0}} = \frac{\lambda_{lim}}{N_0} \\ &\rightarrow g_{ae,lim} = g_{ae,0}\sqrt{\frac{\lambda_{lim}}{N_0}} \end{aligned} \tag{3.35}$$

The process was repeated for every combination of peak, mass, and model. This resulted in four limits (each a function of mass) for each dark-matter model.

3.12.6 Combining Limits

The Run 3 individual-peak limits were combined using the same method as Run 2. For each mass and model, the lowest value among all limits was used for the combined limit. This is equivalent to taking the lower envelope of all peak limits plotted simultaneously. Combining multiple limits in this manner produces a combined limit with lower confidence level than its component limits. When combining n limits (each with confidence level CL) the resulting limit will instead

have confidence level CL^n. This is similar to the "Look-elsewhere effect" and can be understood by considering the bias towards choosing the result with the most beneficial statistical fluctuation. In Runs 2 and 3, the effect was accounted for by setting individual limits with higher confidence levels. The confidence levels were chosen to keep the combined confidence level 90%. This technique is known as a Bonferroni correction [56].

$$\mathrm{CL}_{\mathrm{individual}} = \mathrm{CL}_{\mathrm{combined}}^{1/n} \tag{3.36}$$

For most model and mass combinations, only one or two peaks are likely to contribute the best limit. For such combinations, it is unnecessary to consider any other peaks. Considering additional peaks only worsens the individual limits by requiring harsher confidence levels. Therefore, Runs 2 and 3 used subsets of data to set preliminary limits. The preliminary limits were then used to select the peaks most likely to produce strong limits. For Run 3, the peak selection was done using 10% of the post-cut science data (data from every 10th 0.5-second time stream). The other 90% of science data was reserved for the final limit setting. The following procedure was performed for each mass and model combination:

1. For each peak, calculate the Poisson limit with CL=0.9 under each of the following assumptions:
   - N_{obs} was a Poisson under-fluctuation ($N_{obs} = N'_{obs} - \sqrt{N'_{obs}}$). Use N'_{obs} instead.
   - N_{obs} was a Poisson over-fluctuation ($N_{obs} = N'_{obs} + \sqrt{N'_{obs}}$). Use N'_{obs} instead.
2. For each peak, determine if its best limit (from the over-fluctuation assumption) is lower than each other peak's worst limit (from the under-fluctuation assumption). If so, select this peak for use during limit setting.
3. Repeat steps 1 and 2 with both N'_{obs} and the exposure scaled by 9.

In the procedure above, the over and under fluctuations were assumed to be one standard deviation away from the underlying mean (N'_{obs}) with standard deviation equivalent to that of a Poisson distribution ($\sigma_\lambda^2 = \sqrt{\lambda}$). Using two or three standard deviations was also proposed. Doing so would decrease the likelihood of excluding

a potentially good peak while raising the likelihood of including unnecessary peaks. It was eventually agreed to use one standard deviation (the same as Run 2), but 2 or 3 would have been equally valid choices.

The 9× scaling was used to account for the 9× greater exposure of the remaining limit-setting data. The scaling results produce lower limits for each peak, but the level of improvement depends on the original numbers of observed and expected events. Generally, limits set using peaks with many observed events are minimally improved while limits set using peaks with few observed events are greatly improved. The selected peaks can therefore change with exposure. For Run 3, peaks were used if they were selected under either scaling condition (original or 9× scaled). The original selection was included to account for the possibility that event counts did not scale linearly with exposure. Event counts could scale non-linearly if the detector efficiency or underlying event rate changed during data taking (neither of which we expect to be true). Therefore, including the original selection was likely unnecessary.

Figure 3.48 shows the Run 3 peak selection for DPA with the original and 9×-scaled results shown separately. The region between each over- and under-fluctuation limit is filled to visualize the range of limits produced by such fluctuations. We see how the ideal peak choice transitions smoothly from the first to fourth e^-h^+ peaks with small mass regions of overlap. By comparing both limit plots, one can confirm that the 1 and 2 e^-h^+ limits changed minimally from the scaling. This was due to those peaks having many events in the 10% data. The 3 and 4 e^-h^+ peaks had fewer events and their limits improved more significantly.

The total peak selection includes any peak selected with either scaling. The result for DPA can be seen in Figure 3.49. We see that only 1 or 2 peaks were chosen for each mass point. Therefore, the individual limits were produced with CL=0.9 or CL=$0.9^{1/2} \approx 0.95$ depending on the mass considered. This is a clear improvement from considering all 4 limits set with CL = $0.9^{1/4} \approx 0.975$.

3.12.7 Limit Sensitivity to Systematic Uncertainties

The final consideration for limit setting was the effect of various systematic uncertainties. Specifically, we wanted to understand how uncertainties on energy resolution, energy calibration, CTII probabilities, and cut efficiency could change the final limits. For absorption processes, the uncertainty on σ_1 below 3.2 eV (see Section 3.12.2) was also considered.

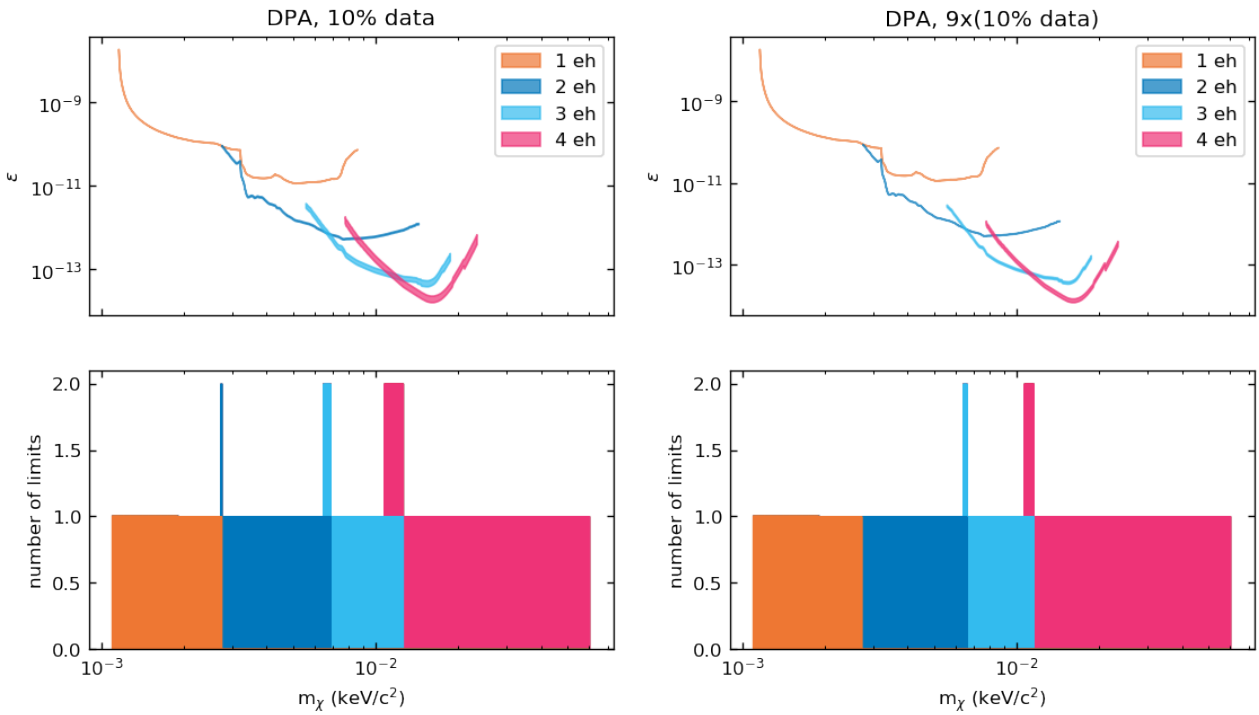


Figure 3.48: Peak-selection results for dark-photon absorption calculated using 10% of the science data. (Top) Limits calculated for each $e^- h^+$ peak with 90% confidence level. Each limit's shaded region is bounded by the limits produced under the over- and under-fluctuation assumptions. If a peak's lower limit is below all other peaks' upper limits, that peak will be included in the selection. (Bottom) The peaks selected using this method. Colors correspond to the labels in the top plot's legend. (Left) The result using 10% data as is. (Right) The result using 10% data with exposure and number of events scaled by 9.

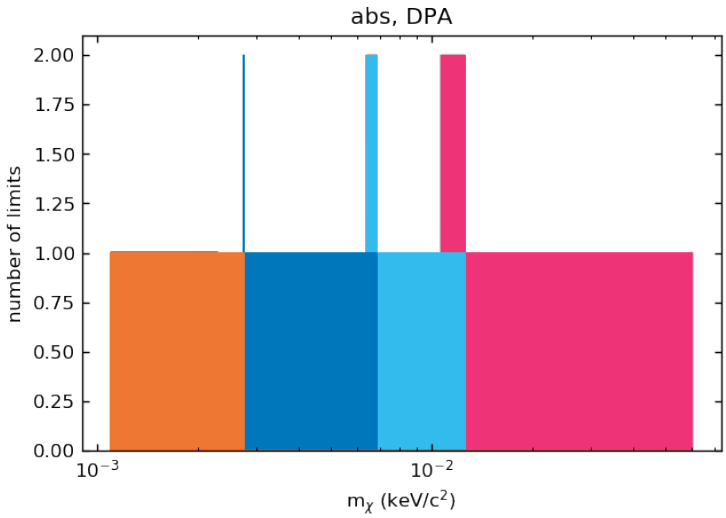


Figure 3.49: The peaks selected (as a function of mass) for dark-photon absorption. The result of combining both selections from Figure 3.48 by including any peaks selected in either. The number of limits defines the confidence level of individual limits used during the final limit setting (using Eq. 3.36).

As was done in Run 2, we quantified the effect by producing a systematic-uncertainty sensitivity "band" around each model's official limit result. The upper and lower bounds of the band were set to enclose the central 68.27% of 5000 limits produced with inputs randomly varied based on the input's uncertainties. Specifically, the following procedure was performed.

1. A probability density function (PDF) was defined for each input parameter with uncertainty. The mean value of each PDF corresponded to the value used for the official limit setting.

2. 5000 sets of input parameters were generated by randomly drawing each parameter from its associated PDF.

3. Individual peak limits were set (for each model) using each set of parameters and the same individual CL values as the official limits.

4. Each parameter set's limits were combined using the same peak for each mass as the official limit.

5. The distribution of all combined limits was used to calculate the upper and lower bounds on the central 68.27% region (as a function of mass).

The 68.27% value was chosen to emulate one standard deviation in a Gaussian distribution. An example distribution of limits can be seen in Figure 3.50. Such plots were produced for every model and mass considered (too many to include here).

The systematic uncertainties and their assigned PDFs are summarized in Table 3.2. For detector energy resolution, the PDF was a flat distribution bound by the minimum and maximum measured resolutions (see Section 3.10). This distribution was designed to conservatively consider all measured resolutions. For charge trapping (CT) and impact ionization (II), truncated Gaussian distributions were used to prevent probabilities from going below 0% or above 100%. The II value in Table 3.2 is the probability to produce a specific quasiparticle (0.8% probability to produce an electron, and a separate 0.8% probability to produce a hole). The PDF for cut efficiency was also a Gaussian and theoretically should have been truncated at 0 and 100%. However, the uncertainty on cut efficiency was so small that the probability of drawing a non-physical efficiency (<0% or >100%) was utterly negligible.

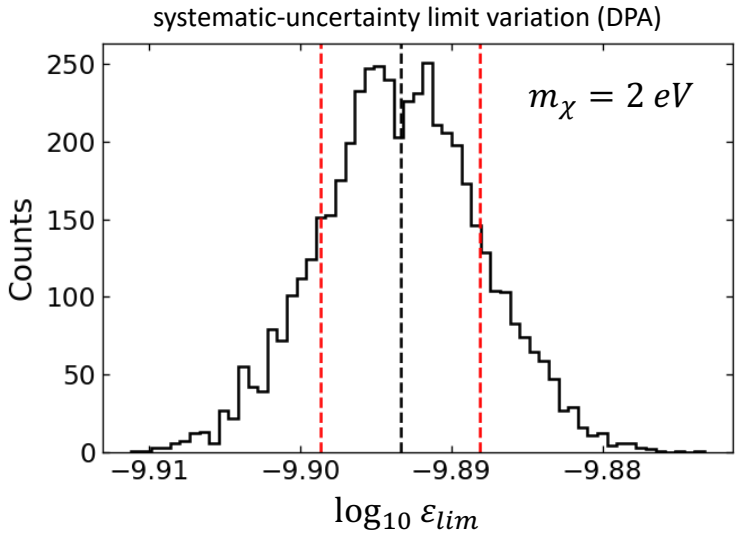


Figure 3.50: The distribution of limits produced by varying the inputs according to their uncertainties for absorption of a 2 eV dark photon. The median limit value is marked with a dashed black line. The upper and lower bounds of the 68.27% region are marked with dashed red lines.

As seen in Section 3.7.5, the uncertainty on energy calibration was a function of energy. We used a unit Gaussian distribution as its PDF. If the drawn value was positive (negative), we multiplied it by the upper (lower) combined calibration uncertainty and added the result to the calibrated energies. Therefore, a draw of σ_{PDF} = 0 would return the nominally calibrated energies. A draw of σ_{PDF} = +(-)1 would return calibrated energies equivalent to using the upper (lower) edge of the combined calibration uncertainty band.

For absorption processes, uncertainty on the real part of the complex conductivity of silicon (σ_1) was also energy dependent and generated using a unit Gaussian PDF. This time, positive (negative) drawn values were multiplied by the difference between the upper (lower) and nominal σ_1 values from [54]. The results were then added to the nominal value. Similarly to the energy calibration, a draw of σ_{PDF} = 0 would return the nominal σ_1. A draw of σ_{PDF} = +(-)1 would return the upper (lower) σ_1. Note that the nominal, lower, and upper σ_1 values are only different for absorption energies <3.2 eV.

Diagnostic plots were also produced to visualize how varying the parameters affected the resulting limits. An example for DPA of mass 2 eV is shown in Figure 3.51. The example results are typical among all models and masses with the exception of the σ_1 dependence (which is non-existent for DMe).

parameter	PDF shape	mean value	PDF parameters
energy resolution (σ_E)	flat	4.26 eV	bounds = [3.03,5.49] eV
CT probability	truncated Gaussian	12.8%	σ_{PDF} = 1.5%
			bounds = [0,100]%
II probability	truncated Gaussian	0.8%	σ_{PDF} = 0.9%
			bounds = [0,100]%
cut efficiency	Gaussian	90.05%	σ_{PDF} = 0.054%
energy calibration	unit Gaussian	0	σ_{PDF} = 1
conductivity of silicon (σ_1)	unit Gaussian	0	σ_{PDF} = 1

Table 3.2: The various systematic uncertainties considered for the Run 3 experiment. The PDFs used to calculate the sensitivity to systematic uncertainties are also described for each parameter.

The diagnostic plots agree with our assumption that energy resolution (within the considered range) would not significantly affect limit results. The plots also suggest that calibration and efficiency uncertainties had insignificant effects. Cut efficiency scales the expected number of events and could heavily affect limits, but the Run 3 efficiency uncertainty was very small (0.054%). CT and II had clear effects. Limits assuming higher CT or II rates produced worse (larger) limits. This was due to both effects moving events to between-peak regions that the limit-setting method did not consider. DPA and ALP limits for masses <3.2 eV also showed a strong dependence on complex conductivity. Greater σ_1 produced better (smaller) limits. This was expected since increasing σ_1 increases the expected rate of events (see Eqs. 3.22 and 3.24).

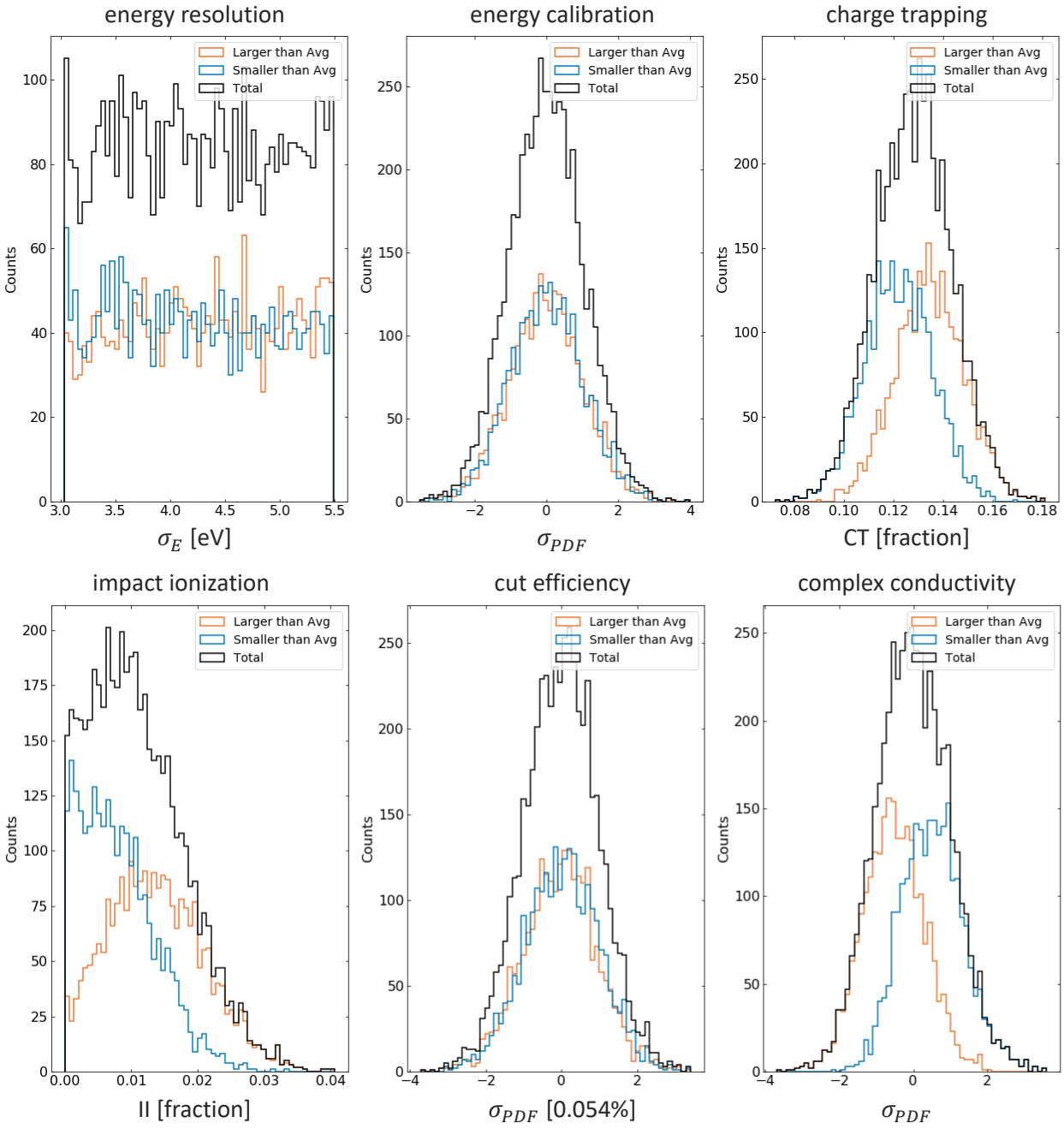


Figure 3.51: Diagnostic plots used to test the effect of various systematic uncertainties on the calculated limits. This example is for absorption of a 2 eV dark photon. Orange (blue) distributions show the parameters that contributed to larger (smaller) than average limits. These results are typical among all models and masses.

3.13 HVeV Run 3 Results

As mentioned above, the post-cut science data was split into 10 and 90% partitions. The two spectra are compared to each other (and the Run 2 data) in Figure 3.52. We see that the 10 and 90% spectra agree well in the first two quantized peaks. The third peaks are difficult to compare due to the low statistics of the 10% spectra. The fourth peaks have low statistics in both spectra, but the 10% data does suggest a higher rate. The rate differences between Runs 2 and 3 seem to be driven by the coincidence cut (see Figure 3.40). Rate in the first peak decreased by a factor of ~2. The higher-peak rates decreased by greater and greater factors.

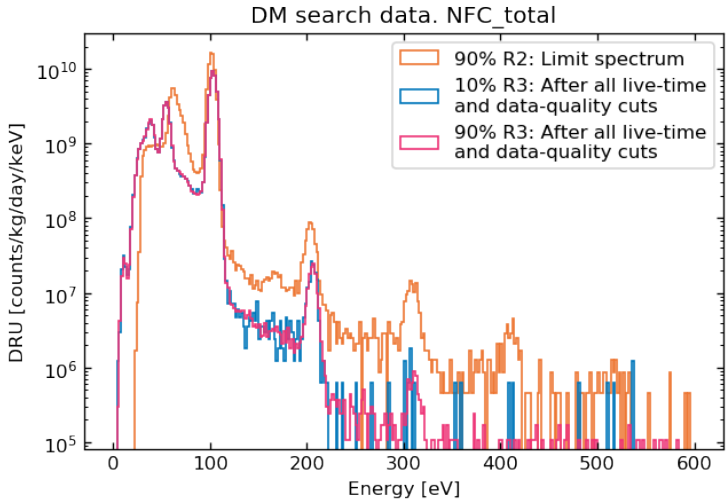


Figure 3.52: The 10 and 90% science-data spectra compared to the Run 2 limit-setting data (with all cuts applied). The spectra were not corrected for cut efficiency. Run 3 efficiency was ~90% for all peaks. Run 2 efficiency was ~90% for the first peak and ~95% for the others. The differences are small enough that the Run 2 and 3 spectra can still be easily compared.

3.13.1 Electron-Scattering Limits

After performing peak selection using the 10% data, the final limits were set using the 90% data. The resulting limits for dark-matter-electron scattering via a heavy and light mediator can be seen in Figures 3.53 and 3.54, respectively. The systematic-uncertainty-sensitivity bands are plotted with the limits and separately as fractional residuals.

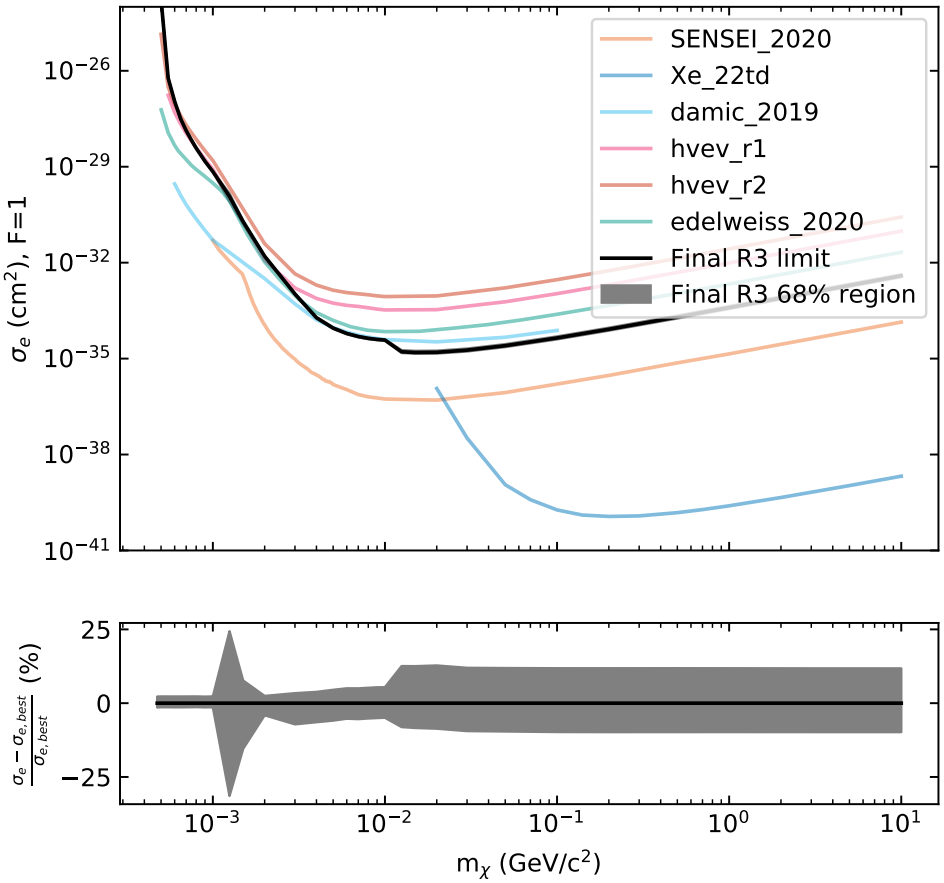


Figure 3.53: The official Run 3 dark-matter-electron-scattering limit assuming a heavy mediator (momentum-transfer form factor = 1). (Top) The limit with other experimental limits plotted for comparison. The region covering 68% of the sensitivity to systematic uncertainties is plotted but difficult to see on this scale. "SENSEI_2020" is the SENSEI Collaboration's 2020 result using a silicon Skipper CCD [46]. "Xe_22td" is the XENON Collaboration's 2019 result using 22 tonne-days of exposure in a liquid-xenon time projection chamber [57]. "damic_2019" is the DAMIC Collaboration's result 2019 using an array of silicon CCDs [58]. "hvev_r1" and "hvev_r2" correspond to HVeV Runs 1 and 2, respectively [40, 26]. "edelweiss_2020" is the EDELWEISS Collaboration's 2020 result using germanium detectors instrumented for phonon and ionization readout [59]. (Bottom) The region covering 68% of the sensitivity to systematic uncertainties plotted as a fractional residual of the official limit.

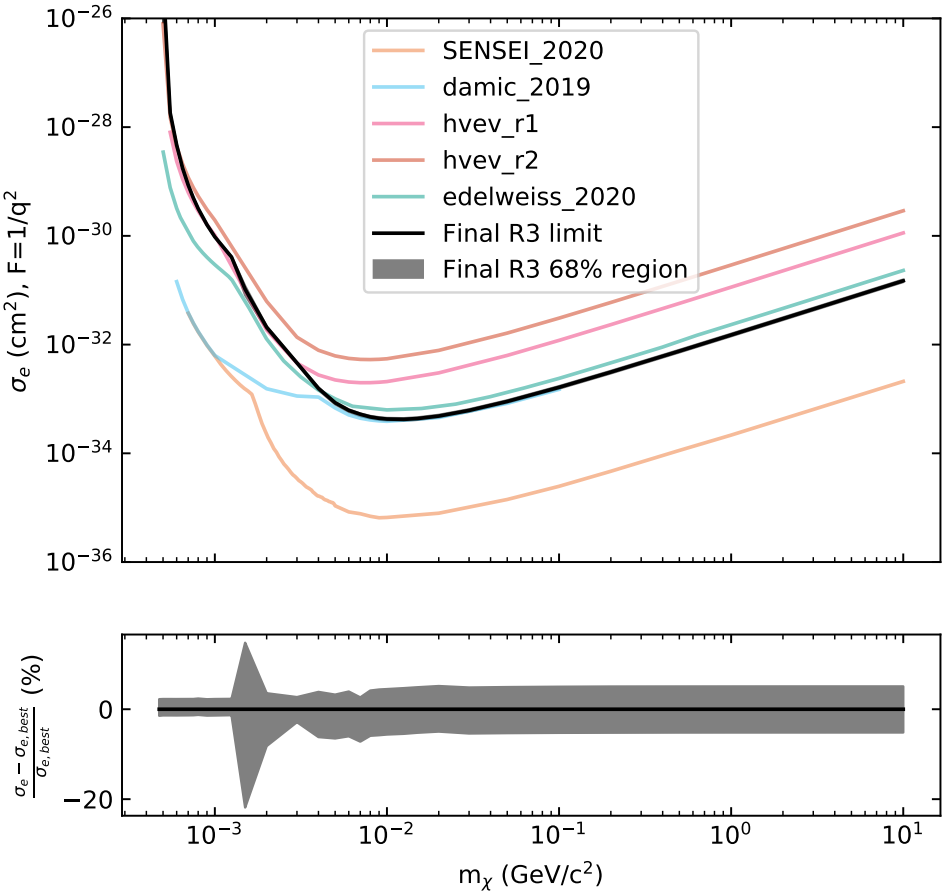


Figure 3.54: The official Run 3 dark-matter-electron-scattering limit assuming a light mediator (momentum-transfer form factor = $(\alpha m_e/q)^2$). (Top) The limit with other experimental limits plotted for comparison (see Figure 3.53). The region covering 68% of the sensitivity to systematic uncertainties is plotted but difficult to see on this scale. (Bottom) The region covering 68% of the sensitivity to systematic uncertainties plotted as a fractional residual of the official limit.

From the limits, we see that sensitivity was improved (compared to Runs 1 and 2) by an order of magnitude for dark-matter masses $\gtrsim$10 MeV. However, due to the remaining low-energy excess, the low-mass limits were not significantly improved. This is notable when comparing to similar experiments. The DAMIC experiment, which also used silicon detectors and operated underground (although much deeper at SNOLAB), produced limits extending to lower masses that were otherwise similar to the Run 3 limits [58]. FR-4 (the leading suspect for low-energy backgrounds in HVeV Runs 1-3) was not used near the DAMIC detectors. The EDELWEISS experiment used a germanium detector, but otherwise presents a similar comparison [59]. For all masses, the Run 3 limits were outdone by DAMIC, SENSEI, or XENON1T, so no new parameter space was excluded [58, 46, 57].

One notable feature in the final heavy-mediator limit was the sharp change just above 10 MeV. When investigated, this was found to be the transition from using the third to fourth quantized peak. Normally, such a transition would be smooth. In this case, the fourth quantized peak actually produced the best limit for all masses above ~6 MeV but was excluded by the peak selection below 10 MeV. The effect can be seen by plotting the final limit against the individual peak limits (Figure 3.55). Such plots also reveal that the effect was worse for the light-mediator limit. In this case, the fourth-peak limit was superior for all masses above 10 MeV but was entirely disallowed by the peak selection.

We discovered the fourth-peak exclusion was a result of deciding to only consider one standard deviation of fluctuation during the peak selection (see Section 3.12.6). Specifically, the rate in the fourth quantized peak was between one and two standard deviations lower in the 90% data than in the 10% data. The difference is visualized in Figure 3.56. The 90% data having a lower rate produced better fourth-peak limits than was predicted by the peak selection. Naturally, the difference also affected the absorption limits, but only the light-mediator electron-scattering limit excluded the fourth-peak limit entirely.

The electron-scattering limits showed very little sensitivity to systematic uncertainties. The sensitivity bands were too small to be seen on the log-scaled limit plots (Figures 3.53 and 3.54). The largest deviations were only ~25% away from the official limits. For both mediators, the maxima appeared as spikes between 1 and 2 MeV. They were confirmed to be the result of limits transitioning from using the first to second quantized peaks. Near the transition, the limit has a high dependence on impact ionization due to the potential for 1-e^-h^+ events to leak into the second

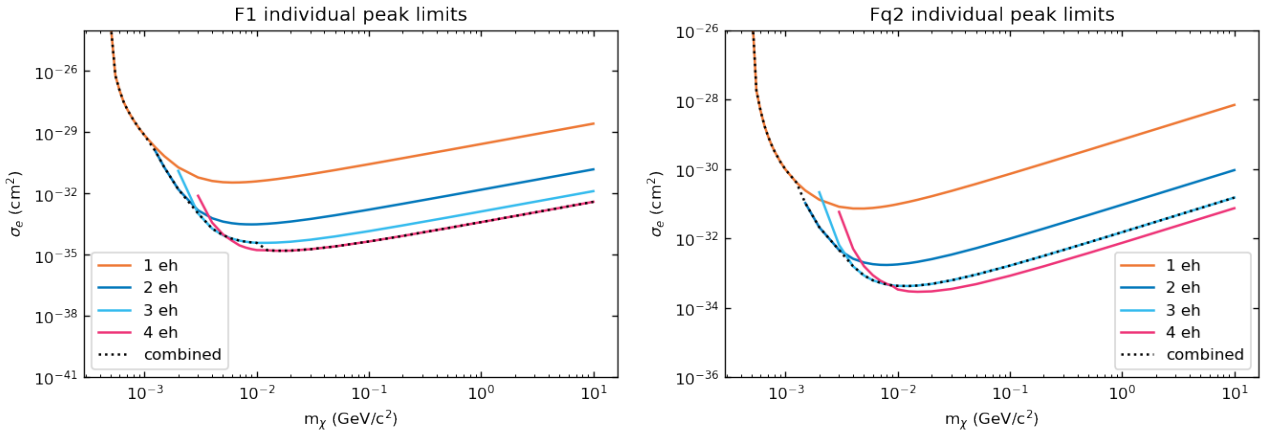


Figure 3.55: The Run 3 dark-matter-electron-scattering limits compared to the individual peak limits. (Left) The limits assuming a heavy mediator (momentum-transfer form factor = 1). (Right) The limits assuming a light mediator (momentum-transfer form factor = $(\alpha m_e/q)^2$). We see how the combined limits transition smoothly from using the first to third quantized peaks. For heavy mediators, the limit transitions to using the fourth peak as soon as the peak is allowed by the limit selection. For light mediators, the fourth peak is never allowed and thus never transitioned to.

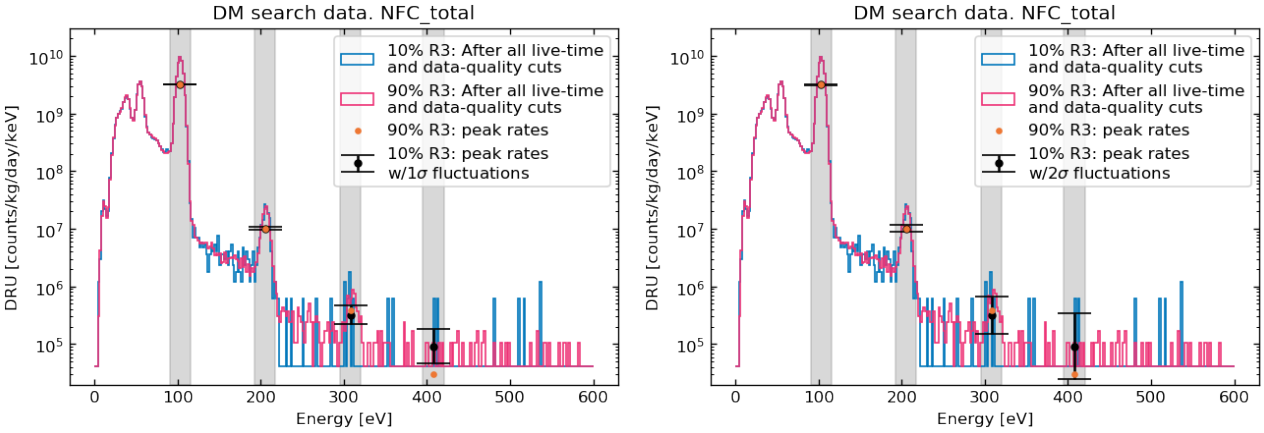


Figure 3.56: The Run 3 10 and 90% science-data spectra. Events were counted in each of the shaded regions and converted to rates in DRU. Error bars were placed on the 10% rates using the over- and under-fluctuation assumptions. (Left) Error bars assuming 1σ fluctuations. (Right) Error bars assuming 2σ fluctuations. We see that a 1σ fluctuation does not account for the difference in rate for the 4-e^-h^+ peak, while a 2σ fluctuation would account for such.

quantized peak. For higher masses, the number of 1-e^-h^+ events decreases and the dependence becomes weaker.

3.13.2 Electron-Absorption Limits

The Run 3 limit on dark-photon absorption can be seen in Figure 3.57. As with electron scattering, the limit improved most (compared to previous runs) for higher masses. However, the improvement stopped above ~20 eV due to the limited energy range of Run 3 (see Section 3.12.1). For all masses, the limits were outdone by the SENSEI experiment, so no new parameter space was excluded [46]. The third-to-fourth quantized-peak transition and associated drop in limit value can be seen just above 10 eV. The sensitivity to systematic uncertainties was again small ($\lesssim$25%). The sensitivity band also spiked near the first-to-second peak transition (around 3 eV) for the same reason discussed for electron-scattering.

Additionally, the DPA sensitivity band shows notch-like features appearing above 10 eV. These were caused by single events moving in and out of the fourth quantized peak for different combinations of energy resolution and energy calibration. Energy resolution determined the span of events counted in the peak. Energy calibration determined which events were within that span. The low statistics of the fourth peak made single-event changes significant. The notches did not appear for electron scattering due to the more complicated relationship between particle mass and event energy.

The Run 3 limit on ALP absorption can be seen in Figure 3.58. The limit and sensitivity-band shapes share all the features described above for DPA. While the ALP limit was a substantial improvement over that of Run 2, it was still significantly less sensitive than the stellar-cooling limit ($g_{ae} \lesssim 10^{-13}$).

3.14 HVeV Run 4

A follow-up to HVeV Run 3 was planned and data-taking executed during the Run 3 analysis. The Run 4 experiment also used four HVeV detectors in the same dilution refrigerator underground at NEXUS. The high number of coincidence events observed in Run 3 confirmed that the low-energy excess was likely caused by an external background. This result, in conjunction with the results of [45] and the superior low-mass sensitivity of the DAMIC experiment, motivated redesigning the HVeV mounts to remove FR-4 completely. Instead, the detectors were mounted in an entirely copper housing. During science data taking, the copper housing was made light tight to minimize any remaining photon backgrounds. A separate cooldown was performed with an LED replacing the light-tight lid to produce data for the Run 4 calibration. The lead shielding was also improved since Run 3 to further

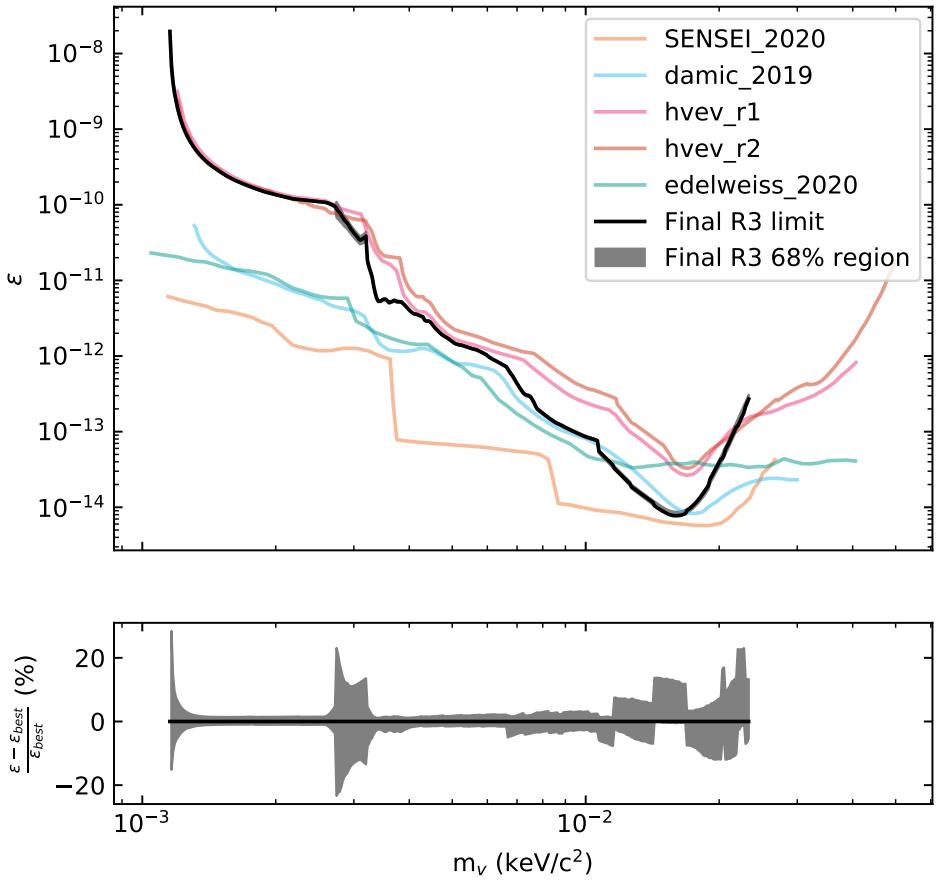


Figure 3.57: The official Run 3 dark-photon-absorption limit. (Top) The limit with other experimental limits plotted for comparison (see Figure 3.53). The region covering 68% of the sensitivity to systematic uncertainties is plotted but only visible around 3 eV. (Bottom) The region covering 68% of the sensitivity to systematic uncertainties plotted as a fractional residual of the official limit.

reduce environmental backgrounds. With each of these improvements, the Run 4 experiment is expected to observed significantly lower backgrounds than Run 3, particularly in the first quantized peak.

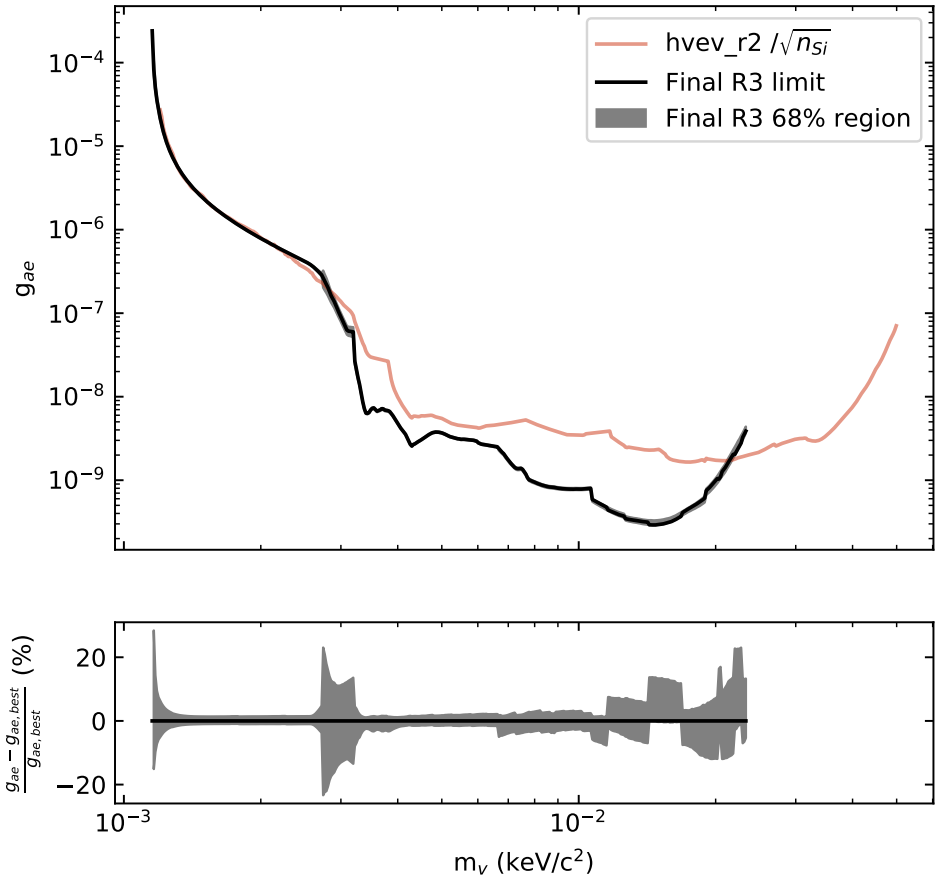


Figure 3.58: The official Run 3 ALP-absorption limit. (Top) The Run 3 limit with the Run 2 limit plotted for comparison. We scaled the Run 2 limit by the square root of the silicon index of refraction to account for the difference in model used for the limit setting (see Section 3.12.2). The stellar-cooling limit remains more sensitive than both limits ($g_{ae} \lesssim 10^{-13}$). The region covering 68% of the sensitivity to systematic uncertainties is plotted but only visible around 3 eV. (Bottom) The region covering 68% of the sensitivity to systematic uncertainties plotted as a fractional residual of the official limit.

Chapter 4

KIPM-DETECTOR DEVELOPMENT

In this chapter, we will discuss work I performed towards the development of kinetic-inductance phonon-mediated (KIPM) detectors for dark-matter detection. Similarly to SuperCDMS iZIP and HV detectors, the initial dark-matter interaction with a KIPM detector would occur within a target mass and produce athermal phonons. The athermal phonons could then be detected by one or many microwave kinetic inductance detectors (MKIDs) on the target surface. Compared to QETs, MKIDs provide several potential benefits to detector operation and readout that we will discuss in the following sections. MKIDs also have potential to achieve extremely low phonon-energy thresholds (using sub-eV resolution). KIPM detectors could be used as part of a payload upgrade in the SuperCDMS SNOLAB experiment. They could also be used in smaller-scale experiments similar to SuperCDMS HVeV Runs 1-4.

4.1 Overview of MKIDs

MKIDs are a well-established technology already used in many astronomical instruments. Specifically, they have been used to study star formation [60, 61, 62] and in exoplanet searches [63]. They are valued for their frequency multiplexability enabling increased sensor number in mm/sub-mm instruments. For near-infrared instruments, they provide timing information that can be used to remove speckle-like backgrounds [64]. When used to observe athermal phonons in a crystal, the increased sensor number has the potential to improve position reconstruction. In a SuperCDMS-style dark-matter detector, position reconstruction is beneficial in removing near-surface backgrounds (see Section 2.3.1). For MKIDs, a highly multiplexed readout does not require the addition of complicated electronics at any cryogenic stages. This simplifies experimental design by preserving volume and thermal-load capacity in the cryostat[1]. MKIDs are an excellent choice for future experiments pursuing a large exposure since their multiplexability (with minimal cryostat loading) is ideal for increasing scale. The simplicity of their cold elec-

[1]Typically, the most substantial thermal load for an MKID readout is the first-stage amplifier. From Section 4.4 onward, we use HEMT amplifiers that dissipate $O(50\,\mathrm{mW})$ at the 4 Kelvin stage. The HEMT bandwidth is sufficient to use a single one for ~100s of MKIDs (assuming our frequency separation of ~5 MHz).

tronics is also attractive when considering operation using existing experimental infrastructure (such as SuperCDMS SNOLAB)

4.1.1 Kinetic Inductance Basics

From an electronics perspective, MKIDs are inductor-capacitor (LC) resonators with a non-negligible portion of their characteristic inductance originating from the kinetic-inductance effect. Kinetic inductance is inductance originating from charge carriers resisting the change in motion caused by a time-varying electric field. The effect is analogous to traditional inductance caused by the Faraday's law tendency of a changing magnetic flux to create an electromotive force to drive a current to oppose the changing flux (Lenz's law). While traditional inductors store energy in their generated magnetic field, kinetic inductors store energy in the kinetic energy of the charge carriers.

In the example of a simple conductor (with no magnetic effects), the dependence of kinetic inductance on charge-carrier parameters can be approximately derived analytically. We can start by comparing the average kinetic energy of N charge carriers with mass m and root-mean-square (RMS) velocity v_{RMS} to the average energy stored in a traditional magnetic inductor (Eq. 4.1).

$$\begin{aligned} K = N\frac{1}{2}mv_{RMS}^2 = E_L = \frac{1}{2}L_K I_{RMS}^2 \\ \rightarrow L_K = \frac{Nmv_{RMS}^2}{I_{RMS}^2} \end{aligned} \tag{4.1}$$

I_{RMS} in Eq. 4.1 is the RMS current in the inductor. This variable can be removed under the assumption that I_{RMS} and v_{RMS} are directly proportional (Eq. 4.2). In reality, this assumption is only valid if the charge carriers travel coherently. If charge carriers travel in random directions (due to frequent collisions for example), the current across the conductor will remain zero independent of v_{RMS}. We will discuss in Section 4.1.2 how superconductors can effectively satisfy the coherence assumption.

$$I_{RMS} = \frac{Nqv_{RMS}}{\ell} = \frac{Nqv_{RMS}}{V/A} = nAqv_{RMS} \tag{4.2}$$

ℓ, V, and A are the length, volume, and cross-sectional area of the simple conductor, respectively. q is the electronic charge per carrier. It is generally more convenient

to describe the inductance using the density of charge-carriers (n) rather than their total number. We can then use equations 4.1 and 4.2 to calculate the inductance per unit length for this simple example (Eq. 4.3).

$$\rightarrow \frac{L_K}{\ell} = \frac{m}{nAq^2} \tag{4.3}$$

For thin-film conductors, it is common to consider kinetic inductance in units of inductance per geometric square. In this formalism, an inductor of length ℓ and width w would consist of ℓ/w squares. For a film with thickness d, the resulting inductance per square will be Eq. 4.4

$$\rightarrow \frac{L_K}{\ell/w} = \frac{m}{ndq^2} \tag{4.4}$$

If we plug in a film thickness of 100 nm, the mass and charge for electrons, and the free-electron density in Al ($6 \times 10^{10}\,\mu m^{-3}$), we find an inductance per square of ~6 fH/sq. The full calculation (Eq. 2.37 from [65]) for a 100 nm penetration depth in superconducting Al gives ~130 fH/sq. The naive calculation gives a reasonable estimation of the scale of L_K.

We also see from Eq. 4.4 that decreasing the carrier density increases the kinetic inductance. While true, decreasing the carrier density too much will limit the current that can be carried (by those carriers). We can use Eq. 4.2 to consider an extreme case. If there were only a single electron as charge carrier in an $\ell = 1$ mm conductor, the maximum I_{RMS} would be ~50 nA even if the electron moved with $v_{RMS} = c$. In a superconductor, applying current greater than what can be carried by the collision-less Cooper pairs will result in current being carried by broken Cooper pairs (quasiparticle electrons). Only current carried by the collision-less Cooper pairs will follow Eq. 4.2 and be effected by kinetic inductance.

By design, MKIDs usually have a combination of both kinetic and magnetic inductance, with the kinetic portion defined by the kinetic-inductance fraction (α).

$$L_{\text{total}} = L_K + L_M, \quad \alpha = \frac{L_K}{L_{\text{total}}} \tag{4.5}$$

An MKID's resonant frequency (f_r) will follow the usual definition for an LC resonator and will depend on the combined inductance (Eq. 4.6). The dependence

of f_r on L_K is one of two ways that MKIDs are used to detect changes in kinetic inductance.

$$f_r = \frac{\omega_r}{2\pi} = \frac{1}{2\pi\sqrt{L_{\text{total}}C}} \propto \frac{1}{\sqrt{L_K + L_M}} \tag{4.6}$$

4.1.2 Kinetic Inductance in Superconductors

A high kinetic-inductance fraction can be achieved in superconductors because bosonic Cooper pairs do not scatter as frequently as normal-state electrons. Cooper pairs condense into the ground state, forming a non-trivial energy gap between each pair and the lowest available excited state. As temperature decreases, more Cooper pairs condense and the energy gap grows. The gap reaches its maximum value at zero temperature. The maximum gap can be predicted according to the superconductor's critical temperature (T_c) [65].

$$\begin{aligned} \Delta_0 &\equiv \Delta\,(T = 0) \\ 2\Delta_0 &\approx 3.52 k_B T_c \end{aligned} \tag{4.7}$$

$\Delta\,(T)$ in Eq. 4.7 is the energy required to free one electron from its Cooper pair for a given material at temperature T. Electrons must be freed in pairs, so 2Δ is the more physically interesting value. k_B is the Boltzmann constant.

Scattering between Cooper pairs and the lattice only occurs where lattice vibrational energy overcomes 2Δ. Below T_c, the average vibrational energy is below 2Δ and scattering becomes infrequent. This allows Cooper pairs to travel without dissipation (the effect we call superconductivity). It also allows Cooper pairs to travel coherently when an alternating current is applied, justifying Eq. 4.2 for superconductors. Electrons in a normal-state conductor have no energy gap and scatter frequently, dissipating rather than storing energy.

In practice, the kinetic inductance of a material also depends on the input signal's driving frequency. The higher the frequency, the less time a carrier must travel (without scattering) before being decelerated and contributing to the inertial effect. Even normal-state electrons will contribute kinetic inductance when driven at frequencies greater than their scattering frequency. Kinetic inductance in real superconductors depends on a number of additional factors including material, geometry, and the associated magnetic fields. Eq. 4.3 therefore makes a useful reference but will not be relied upon in the following sections.

Incident particles (such as a photons or athermal phonons) can also deposit sufficient energy to break Cooper pairs in superconductors. This causes the Cooper-pair density to decrease, increasing the kinetic inductance (as seen in Eq. 4.3). The increased inductance lowers the resonant frequency of the MKID. Changing resonant frequency can be observed using a radio-frequency (RF) readout. Additionally, the electron quasiparticles released by broken Cooper pairs scatter and increase the MKID's total dissipation. The changes in resonant frequency and dissipation can be read out concurrently. In aluminum, breaking a Cooper pair requires less than 0.5 meV of energy, meaning the basic quanta of detection has a very low threshold.

4.1.3 MKID Readout Basics

The resonant frequency and dissipation of an MKID are typically measured by coupling the MKID to a feedline. The MKID then creates an effective notch-filter in the transmission of the feedline, centered on its resonant frequency. The quality factor (Q_r) of the resonance and its coupling to the feedline are reflected in the full-width at half maximum (FWHM) of the notch filter. MKIDs are natural candidates for frequency-domain multiplexing since multiple notch filters can be measured on the same feedline without interference. The only requirement is that each MKID has a distinct resonant frequency (separation by a few FWHMs is typical). For this reason, it is beneficial for multiplexability to have MKIDs with high Q_r.

$$Q_r \equiv \frac{f_r}{FWHM} \tag{4.8}$$

Typical MKID readouts use IQ up-mixing (see Appendix B) to produce a probing signal at the MKID's resonant frequency. This signal is sent through the feedline, where its amplitude and phase will be changed by the coupled MKID. The signal is then demodulated using another IQ mixer for recording and analysis. One reads out many MKIDs in parallel by inputting a "comb" of multiple resonant frequencies simultaneously. Typically, the comb is generated near zero-frequency using a DAC and then up-mixed to the frequencies of interest using an IQ mixer. The mixer multiplies all frequencies of the comb with the LO, creating an up-mixed version of each tone in the comb. After being sent through the feedline, the comb is down-mixed using the same LO and another IQ mixer. The resultant output will be more complicated than the DC-offset seen with a single-frequency readout (discussed in Appendix B) and must be demodulated again for each readout frequency. The second stage of demodulation is typically done using software or firmware rather

than additional IQ mixers. When designing such a system, it is important to have a readout bandwidth large enough to record up to twice the width (in frequency) of the desired comb. The factor of two is required to satisfy the Nyquist condition and avoid aliasing.

With this readout style, the ratio of output to input signal amplitude is equivalent to the magnitude of the transmission through the readout chain (including the MKID and coupled feedline). The difference between the output and input signal phase is defined as the phase of the transmission. In scattering-parameter terminology, the complex forward transmission is typically labeled as the scattering parameter S_{21}.

The transmission for an ideal LC-coupled feedline is shown in Eq. 4.9. A step-by-step derivation of this equation using a network-transmission model can be found in Chapter 4 of Jiansong Gao's thesis [65].

$$S_{21}(\omega) = 1 - \frac{\frac{Q_r}{Q_c}}{1 + 2jQ_r\left(\frac{\omega-\omega_r}{\omega_r}\right)} \tag{4.9}$$

For our purposes, the relevant LC-circuit is an MKID. ω is the angular frequency at which the transmission is being measured ($2\pi\times$ the readout frequency). Q_c is the coupling quality factor, which characterizes the energy lost from the MKID to the feedline. Q_r includes both Q_c and the internal quality factor (Q_i), which characterizes the internal loss of the MKID (Eq. 4.10).

$$\frac{1}{Q_r} = \frac{1}{Q_c} + \frac{1}{Q_i} \tag{4.10}$$

Internal loss is caused by the collision of quasiparticles in the superconductor, which creates a parasitic resistance in the inductor. This is equivalent to having a resistor in series with the inductor and capacitor. It is therefore appropriate to describe Q_i using the definition for a series RLC circuit (Eq. 4.11). With R constituting the parasitic resistance of the inductor.

$$\frac{1}{Q_i} = R\sqrt{\frac{C}{L_{\text{total}}}} = \frac{R}{\omega_r L_{\text{total}}} \tag{4.11}$$

For readout-frequencies far from resonance, the denominator of the second term in Eq. 4.9 becomes large and the transmission goes to unity (typical of a notch filter).

This feature is how many MKIDs can couple to the same feedline without interference. At precisely the resonant frequency ($\omega = \omega_r$), the transmission experiences its greatest loss (losing Q_r/Q_c of the total signal in voltage amplitude). When reading out real MKIDs, we will see that loss is not always greatest at the resonant frequency (see Figure 4.11).

4.1.4 MKID Responsivity

In this section, we will consider the response of an MKID to energy depositions. This is required to predict the energy resolution of the MKID (Section 4.2.1) and eventually the energy resolution of a full KIPM dark-matter detector (Section 4.2.2).

As discussed in Section 4.1.2, Cooper-pair breaking in an MKID will lower the resonant frequency and increase dissipation. An increase in dissipation will decrease the internal quality factor of the MKID. Specifically, the quality factor is inversely proportional to the dissipation (Eq. 4.11). Thus, to understand an MKID-based detector's sensitivity to energy depositions, it is necessary to understand how transmission changes with respect to changes in ω_r and $1/Q_i$ (Eqs. 4.12 and 4.14). These results can be derived starting with Eq. 4.9.

$$\frac{\partial}{\partial \omega_r} S_{21} = \frac{-2j \frac{Q_r^2}{Q_c} \frac{\omega}{\omega_r^2}}{\left(1 + 2jQ_r \left(\frac{\omega - \omega_r}{\omega_r}\right)\right)^2} \tag{4.12}$$

$$\rightarrow \left| \frac{\partial}{\partial \omega_r} S_{21} \right| = \frac{2 \frac{Q_r^2}{Q_c} \frac{\omega}{\omega_r^2}}{1 + 4Q_r^2 \left(\frac{\omega - \omega_r}{\omega_r}\right)^2} \tag{4.13}$$

$$\frac{\partial}{\partial 1/Q_i} S_{21} = \frac{\frac{Q_r^2}{Q_c}}{\left(1 + 2jQ_r \left(\frac{\omega - \omega_r}{\omega_r}\right)\right)^2} \tag{4.14}$$

$$\rightarrow \left| \frac{\partial}{\partial 1/Q_i} S_{21} \right| = \frac{\frac{Q_r^2}{Q_c}}{1 + 4Q_r^2 \left(\frac{\omega - \omega_r}{\omega_r}\right)^2} \tag{4.15}$$

With these results, we see that the absolute dissipation and frequency sensitivities are maximized on resonance ($\omega = \omega_r$)[2]. This is why MKIDs are generally read out

[2]The ω dependence of the numerator in Eq. 4.13 is insignificant to the sensitivity optimization for $Q_r \gg 1$.

as close to their resonant frequencies as possible. It is also worth noting that the ratio of sensitivities is a simple function of the readout and resonant frequencies (Eq. 4.16). The ratio is also an easy way to show that the frequency and dissipation responses are precisely perpendicular.

$$\frac{\frac{\partial}{\partial\omega_r}S_{21}}{\frac{\partial}{\partial 1/Q_i}S_{21}} = -2j\frac{\omega}{\omega_r^2} \tag{4.16}$$

The next step towards understanding the transmission response to an energy deposition is to understand how ω_r and $1/Q_i$ change with respect to changes in Cooper-pair density. By convention, this is usually done using the density of electron quasiparticles (n_{qp}) as the variable of interest. Each conduction electron in the metal exists as either a quasiparticle or half of a Cooper pair, so the two densities are entirely dependent (Eq. 4.17).

$$\begin{aligned} n_{\mathrm{e}} &= n_{qp} + 2n_{\text{Cooper-pairs}} \\ \partial n_{qp} &= -2\partial n_{\text{Cooper-pairs}} \end{aligned} \tag{4.17}$$

We can differentiate our definitions in equations 4.6 and 4.11 to begin the derivation. It is important to remember that the kinetic inductance fraction (α) is not constant but rather a function of L_K. If α is treated as a constant during differentiation, the resulting dL_K/dn_{qp} terms in Eqs. 4.18 and 4.19 will be wrong by a factor of α.

$$\frac{d\omega_r}{dn_{qp}} = -\alpha\frac{\omega_r}{2L_K}\frac{dL_K}{dn_{qp}} \tag{4.18}$$

$$\frac{d1/Q_i}{dn_{qp}} = \alpha\frac{1}{\omega_r L_K}\frac{dR}{dn_{qp}} - \alpha\frac{1/Q_i}{2L_K}\frac{dL_K}{dn_{qp}} \tag{4.19}$$

We now need to understand how changing quasiparticle density affects kinetic inductance (L_K) and parasitic resistance (R). The complexity of real MKIDs prevents us from using Eq. 4.3 to derive either. Instead, the result is found using BCS and Mattis-Bardeen theory. In Chapter 2 of his thesis, Jiansong Gao uses both to derive the relationship between quasiparticle density and a superconductor's complex impedance for various limits on film characteristics [65]. The film-characteristic

dependence is encoded in the factor γ. In Chapter 2 of Seth Siegel's thesis, he rewrote Gao's result in terms of surface resistance R_S and surface inductance L_S (Eqs. 4.20 and 4.21).

$$\frac{R_S - R_{S,0}}{\omega L_{S,0}} = |\gamma| \kappa_1 \left(T, \omega, \Delta_0\right) n_{qp} \tag{4.20}$$

$$\frac{L_S - L_{S,0}}{L_{S,0}} = |\gamma| \kappa_2 \left(T, \omega, \Delta_0\right) n_{qp} \tag{4.21}$$

One implication of the Mattis-Bardeen result is the exponential suppression of magnetic fields inside the superconductor, leading to a concentration of current density near the conductor surface. This effect is why the resistance and inductance above are labeled as "surface" values. The label is not meant to indicate that the resistance, inductance, or current are constrained precisely to the (infinitely thin) conductor surface. The resultant surface inductance originates from the kinetic-inductance effect and does not include magnetic inductance (originating from the device geometry). Therefore, $L_S = L_K$. The resultant surface resistance originates from quasiparticle scattering. Since quasiparticle scattering is the only source of dissipation, $R_S = R$. Siegel's results are described with respect to resistance and inductance at zero temperature with no external excitations (including photons or phonons). I have labeled such values with a "0" subscript. In Siegel's thesis, these values are labeled "dark," because photons are the typical excitation of concern for telescopes. In the 0 (or dark) state, no quasiparticles will exist to dissipate energy, so $R_{S,0} = 0$.

$$\frac{R}{\omega L_{K,0}} = |\gamma| \kappa_1 \left(T, \omega, \Delta_0\right) n_{qp} \tag{4.22}$$

$$\frac{L_K - L_{K,0}}{L_{K,0}} = |\gamma| \kappa_2 \left(T, \omega, \Delta_0\right) n_{qp} \tag{4.23}$$

The kappa functions (κ_1 and κ_2) encode the relationship between complex conductivity (σ) and quasiparticle density in a superconductor. Both are defined in Gao's thesis [65]. The functions are dependent on the temperature of the quasiparticle system. We typically assume quasiparticle temperature is equivalent to the fridge temperature (T), but exposure to persistent backgrounds such as RF photons can cause the effective quasiparticle temperature to be elevated.

$$\kappa = \frac{1}{|\sigma|}\frac{d\sigma}{dn_{qp}} \tag{4.24}$$

$$\kappa = \kappa_1 + j\kappa_2 \tag{4.25}$$

The complex conductivity is fully defined in integral form (Eqs. 4.26 – 4.28) in Gao's thesis [65]. It is a function of the normal-state conductivity (σ_n), superconducting energy gap (Δ), temperature (T), readout frequency (ω), and effective chemical potential for quasiparticle (μ^*).

$$\sigma = \sigma_1 + j\sigma_2 \tag{4.26}$$

$$\sigma_1 = \frac{2\sigma_n}{\hbar\omega}\int_{\Delta}^{\infty} \frac{\left[f\left(E;\mu^*,T\right) - f\left(E+\hbar\omega;\mu^*,T\right)\right]\left(E^2+\Delta^2+\hbar\omega E\right)}{\sqrt{E^2-\Delta^2}\sqrt{(E+\hbar\omega)^2-\Delta^2}}dE \tag{4.27}$$

$$\sigma_2 = \frac{\sigma_n}{\hbar\omega}\int_{\Delta-\hbar\omega}^{\Delta} \frac{\left[1-2f\left(E+\hbar\omega;\mu^*,T\right)\right]\left(E^2+\Delta^2+\hbar\omega E\right)}{\sqrt{\Delta^2-E^2}\sqrt{(E+\hbar\omega)^2-\Delta^2}}dE \tag{4.28}$$

The effective chemical potential is defined by requiring the integral of the Fermi-Dirac distribution with the superconducting density of states to give the correct quasiparticle density. The superconducting density of states is normalized to N_0 (the density of states for electrons of a single spin at the Fermi energy level).

$$n_{qp} = 4N_0\int_{\Delta}^{\infty} f\left(E;\mu^*,T\right)\frac{E\,dE}{\sqrt{E^2-\Delta^2}} \tag{4.29}$$

Eqs. 4.27 – 4.29 all use the well-known Fermi-Dirac distribution (Eq. 4.30).

$$f\left(E;\mu^*,T\right) = \frac{1}{1+e^{\frac{E-\mu^*}{k_BT}}} \tag{4.30}$$

Approximate forms for both kappa functions are also calculated in Gao's and Siegel's theses [65, 66]. These forms are functions of readout frequency through the factor ξ (Eq. 4.33). They use the zeroth-order modified Bessel functions of first and second kind (I_0 and K_0, respectively). The approximations below are specifically

for changes in conductivity due to athermal Cooper-pair-breaking events. Thermal changes will also affect quasiparticle number and conductivity but with a different approximate κ. Gao showed that the athermal and thermal κ are very similar for sub-Kelvin temperatures in Al [65].

$$\kappa_1\left(T,\omega,\Delta_0\right)=\frac{1}{\pi N_0\Delta_0}\sqrt{\frac{2\Delta_0}{\pi k_B T}}\sinh\left(\xi\right)K_0\left(\xi\right) \tag{4.31}$$

$$\kappa_2\left(T,\omega,\Delta_0\right)=\frac{1}{2N_0\Delta_0}\left[1+\sqrt{\frac{2\Delta_0}{\pi k_B T}}e^{-\xi}I_0\left(\xi\right)\right] \tag{4.32}$$

$$\xi=\frac{\hbar\omega}{2k_B T} \tag{4.33}$$

The approximate forms are appropriate when $\hbar\omega \ll \Delta$, $k_B T \ll \Delta$, and $e^{-\frac{\Delta-\mu^*}{k_B T}} \ll 1$.

Understanding the different film-characteristic assumptions encoded in γ requires a few additional definitions:

- The Mattis-Bardeen kernel is the kernel giving the current density (normal and supercurrent) from the vector potential. It has a characteristic decay length equivalent to the minimum of the coherence length (ξ_0) and the electron mean free path (ℓ_e).
- The coherence length can be thought of as the average distance between electrons in a Cooper pair or the average size of a Cooper pair. In either case, it is the characteristic length scale of the Cooper pairs.
- The effective penetration depth (λ_{eff}) is the characteristic decay length of the penetrating magnetic field.

The commonly used combinations of film-characteristic assumptions are listed below:

- **Thick-film, extreme anomalous limit:** The film is thicker than ℓ_e. This condition is referred to as the **thick-film limit**. Additionally, the decay length of the Mattis-Bardeen kernel is much greater than λ_{eff} because both ξ_0 and ℓ_e are much greater than λ_{eff}. This decay-length condition is referred to as the "**extreme anomalous limit**." γ is equivalent to $-1/3$.

- **Thick-film, local limit:** The **thick-film** limit is satisfied. Additionally, the decay length of the Mattis-Bardeen kernel is much less than λ_{eff} because either ξ_0 or ℓ_e is much less than λ_{eff}. This decay-length condition is referred to as the "**local limit**." γ is equivalent to $-1/2$.

- **Thin-film, local limit:** The film thickness is less than the unobstructed ℓ_e, limiting the actual mean free path to be equivalent to the film thickness. This condition is referred to as the **thin-film limit**. Additionally, the decay length is in agreement with the **local limit**. With this combination of assumptions, the field never decays to zero and the conductor has uniform current density. γ is equivalent to -1.

The summary of film-characteristic assumptions and associated γ values are shown in Eq. 4.34.

$$\gamma = \begin{cases} -1/3 & \text{Thick-film, extreme anomalous limit} \\ -1/2 & \text{Thick-film, local limit} \\ -1 & \text{Thin-film, local limit} \end{cases} \tag{4.34}$$

We can now differentiate Eqs. 4.22 and 4.23 to learn how the inductance and resistance change with quasiparticle density. The integral form of κ only weakly depends on n_{qp} through the factor μ^*. The approximate forms (Eqs. 4.31 and 4.32) do not depend on n_{qp} at all, so we do not consider that dependence in the differentiation.

$$\frac{dR}{dn_{qp}} = \omega L_{K,0} |\gamma| \kappa_1 \left(T, \omega, \Delta_0\right) \tag{4.35}$$

$$\frac{dL_K}{dn_{qp}} = L_{K,0} |\gamma| \kappa_2 \left(T, \omega, \Delta_0\right) \tag{4.36}$$

Now we can combine Eqs. 4.12, 4.14, 4.18, 4.19, 4.36, and 4.35 to write down the total response to changes in quasiparticle density.

$$\begin{aligned}\frac{dS_{21}}{dn_{qp}} &= \frac{\alpha|\gamma|\frac{Q_r^2}{Q_c}\frac{\omega}{\omega_r}\frac{L_{K,0}}{L_K}}{\left(1+2jQ_r\left(\frac{\omega-\omega_r}{\omega_r}\right)\right)^2}\left(\kappa_1 - \frac{\omega_r}{2Q_i\omega}\kappa_2 + j\kappa_2\right) \\ &= \frac{\alpha|\gamma|\frac{Q_r^2}{Q_c}\frac{\omega}{\omega_r}\frac{L_{K,0}}{L_K}}{\left(1+2jQ_r\left(\frac{\omega-\omega_r}{\omega_r}\right)\right)^2}\left(\kappa - \frac{\omega_r}{2Q_i\omega}\kappa_2\right)\end{aligned} \tag{4.37}$$

Those familiar with MKIDs will recognise Eq. 4.37 once a number of common assumptions are applied. The first assumption is that the MKID is always being read out precisely on resonance ($\omega = \omega_r$). This is a logical choice since (as discussed at the beginning of this section) the maximum responsivity is seen on-resonance. However, the assumption is not entirely realistic, since the exact resonant frequency will change with inductance (Eq. 4.18) during a Cooper-pair breaking event. Some MKID readouts use tone-tracking to read out on resonance during events of any size (improving dynamic range). For small signals, this is unnecessary and the assumption remains reasonable. Assuming $\omega = \omega_r$ gives Eq. 4.38.

$$\frac{dS_{21}}{dn_{qp}} = \alpha|\gamma|\frac{Q_r^2}{Q_c}\frac{L_{K,0}}{L_K}\left(\kappa - \frac{1}{2Q_i}\kappa_2\right) \tag{4.38}$$

Under the assumption (generally confirmed by measurement) of a high internal quality factor, the latter κ_2 term becomes trivial and can be ignored. This gives Eq. 4.39.

$$\frac{dS_{21}}{dn_{qp}} = \alpha|\gamma|\frac{Q_r^2}{Q_c}\frac{L_{K,0}}{L_K}\kappa \tag{4.39}$$

Lastly, L_K is assumed to be approximately equal to $L_{K,0}$. This can be achieved by operating at temperature significantly below T_c and by minimizing external excitations. The latter condition is desirable for low-background searches anyway. One can still make this assumption at higher temperatures if the resonant frequency is observed to make only a small fractional change from its lowest-temperature value. The resulting Eq. 4.40 is the typically used responsivity for MKIDs.

$$\frac{dS_{21}}{dn_{qp}} = \alpha|\gamma|\frac{Q_r^2}{Q_c}\kappa \tag{4.40}$$

Now, we see how MKID design can optimize sensitivity. Unsurprisingly, responsivity increases with α (kinetic-inductance fraction). A high Q_r (total quality factor) is also beneficial (in addition to being beneficial for multiplexing). When considering the dependence of Q_r on Q_c (coupling quality factor) and Q_i (internal quality factor), the optimal coupling is calculated to be $Q_c = Q_i = 2Q_r$. The dependence on γ indicates that a thicker film does not directly improve responsivity.

MKID sensitivity is also very stable. α, γ, and the quality factors are the result of MKID design and should be stable if the MKID is operated at relatively stable temperature. κ values are also relatively stable with temperature. Calculated values can be seen in Figure 4.1, taken from [66]. Response stability is what makes MKID calibration possible.

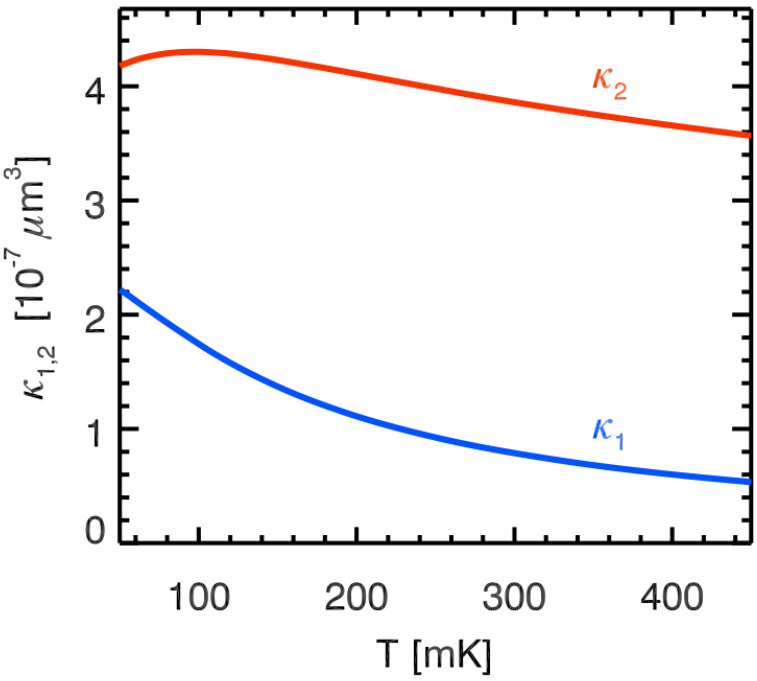


Figure 4.1: κ calculated using Mattis-Bardeen theory for MKIDs with frequency 3.2 GHz, Δ_0 = 0.21 meV (typical for aluminum), and various temperatures. Figure is taken from [66]

.

Eq. 4.40 can also be used to approximately convert small event signals from units of transmission (δS_{21}) to energy in the MKID (δE). The conversion requires multiplying by the superconducting energy gap (Δ).

$$\delta E = \Delta \delta n_{qp} \approx \Delta \frac{dn_{qp}}{dS_{21}} \delta S_{21} = \frac{Q_c \Delta}{\alpha |\gamma| Q_r^2 \kappa} \delta S_{21} \tag{4.41}$$

4.2 MKIDs for Dark Matter

In a dark-matter search, we could use MKIDs in a very similar way to the QET sensors used by SuperCDMS. Dark-matter interaction with a target substrate will produce phonons (either directly or through ionization and NTL-phonon generation), which can then be detected by MKIDs on the substrate surface. First, we must use our understanding of MKID responsivity to predict the resolution on energy deposited in the MKID. Then, we can predict the resolution of the total detector to energy deposited in the substrate.

4.2.1 MKID Energy Resolution

We will use the Optimal-Filtering method (as outlined in [47]) to estimate the best resolution we can achieve using our MKIDs. With this method, one applies an acausal filter equivalent to the complex conjugate of the Fourier-transformed signal template ($\tilde{s}^*(f)$) weighted by the noise spectral density ($J(f)$). The filter ($\phi(f)$) is defined in Eq. 4.42

$$\phi(f) = \frac{\tilde{s}^*(f))}{J(f)} \tag{4.42}$$

This filter is often re-normalized by the filtered template integrated over all frequencies. This choice ensures that events matching the template shape have equivalent amplitude before and after filtering.

$$\phi(f) = \frac{\frac{\tilde{s}^*(f))}{J(f)}}{\int_{-\infty}^{\infty} \frac{|\tilde{s}(f)|^2}{J(f)} df} \tag{4.43}$$

With such a filter, the optimal resolution can be calculated using only the template and noise [67].

$$\sigma^2 = \left[\int_{-\infty}^{\infty} \frac{|\tilde{s}(f)|^2}{J(f)} df \right]^{-1} \tag{4.44}$$

It is important to note that this equation requires the double-sided noise spectral density. If one prefers using the single-sided spectral density, they must change the bounds of integration to $[0, \infty]$ and multiply the integral by 2 twice. The first factor

of 2 comes from halving the region of integration. The second factor of 2 comes from the single-sided $J(f)$ being twice its double-sided equivalent for $f > 0$.

We will now consider the resolution in a simple case. Specifically, we will treat noise spectral density as constant ($J(f) = J_0$) and the signal template as an exponential decay which appears promptly at an initial time $t = 0$. This signal template is normalized to have a peak amplitude of 1 at $t = 0$.

$$s(t) = \begin{cases} 0 & t < 0 \\ e^{-t/\tau} & t \geq 0 \end{cases} \tag{4.45}$$

We will use Golwala's choice of definition for the Fourier transforms [67].

$$\begin{aligned} \tilde{g}(f) &= \int_{-\infty}^{\infty} dt g(t) e^{-j\omega t} \\ g(t) &= \int_{-\infty}^{\infty} df \tilde{g}(f) e^{j\omega t} \end{aligned} \tag{4.46}$$

We can easily calculate the Fourier transform of $s(t)$.

$$\tilde{s}(f) = \int_0^{\infty} dt e^{-t\left(\frac{1}{\tau}+j\omega\right)} = \frac{1}{\left(\frac{1}{\tau}+j\omega\right)} \tag{4.47}$$

We can then use Eq. 4.44 to calculate the resolution. The integration is simplified by observing that the integrand is proportional to the derivative of $\arctan(\omega\tau)$.

$$\begin{aligned} \sigma^2 &= \left[\int_{-\infty}^{\infty} \frac{df}{J_0\left(\frac{1}{\tau^2}+\omega^2\right)}\right]^{-1} \\ &= J_0\left[2\int_0^{\infty} \frac{d\omega/2\pi}{\left(\frac{1}{\tau^2}+\omega^2\right)}\right]^{-1} \\ &= J_0\left[\frac{1}{\pi}\tau\frac{\pi}{2}\right]^{-1} \\ &= J_0\frac{2}{\tau} \end{aligned} \tag{4.48}$$

The units of σ in Eq. 4.48 are defined by the units chosen for J_0. It is practical for us to first consider the noise in units of transmission (S_{21}). In this case, we need J_0 to have units of $[S_{21}]^2 / Hz = 1/Hz$.

Let us assume there is a first-stage amplifier coming directly after the MKID-coupled device. If the amplifier noise is an order of magnitude larger than all other noise sources we refer to the total noise as "amplifier-dominated." In this case, we can treat the amplifier as the only noise source.

$$J_0 = J_{amp} \tag{4.49}$$

A brief aside: In this thesis, I will use noise temperature as my primary measure of noise power. It is understood that a noise-temperature T is equivalent to the Johnson noise observed from an impedance-matched terminator held at temperature T. Noise temperature is proportional to noise power divided by the bandwidth over which the power is measured (Eq. 4.50). The proportionality constant is the Boltzmann constant.

$$k_B T_N = \frac{P_N}{BW} \tag{4.50}$$

If our first-stage amplifier has a constant input-referred noise temperature (T_N), it will produce a constant (with respect to frequency) noise power of $k_B T_N BW$ at the amplifier input (the device output). If the system has characteristic impedance Z_0, that noise power will correspond to a mean-square voltage of $k_B T_N BW Z_0$. Similarly, if our readout system generates a power of P_g at the device input, the input mean-square voltage will be $P_g Z_0$. Since the amplifier noise is random with respect to the input signal, we can write the real and imaginary parts of the transmission noise as follows.

$$\langle Re\,[S_{21}]^2 \rangle = \langle Im\,[S_{21}]^2 \rangle = \frac{1}{2}\frac{\langle V_N^2 \rangle}{\langle V_{in}^2 \rangle} = \frac{k_B T_N BW}{2P_g} \tag{4.51}$$

The factor of 2 in Eq. 4.51 arises from the division of noise power between the in-phase and quadrature-phase of S_{21}. A discussion of how this occurs for random noise is provided in Appendix C. We now divide out the measurement bandwidth to get J_{amp} in the correct units.

$$J_{amp} = \frac{\langle Re\,[S_{21}]^2 \rangle}{2BW} = \frac{k_B T_N}{4P_g} \tag{4.52}$$

The additional factor of 2 in the denominator arises from our use of a double-sided J_{amp}. In that case, the power within a bandwidth BW is evenly split between the positive and negative frequencies of that bandwidth.

Combining Eqs. 4.48 and 4.52 gives our resolution in S_{21} units.

$$\sigma_{S21} = \sqrt{\frac{k_B T_N}{2P_g \tau}} \tag{4.53}$$

We can now use Eq. 4.40 to convert this resolution to units of quasiparticle density.

$$\sigma_{n_{qp}} = \frac{Q_c}{\alpha |\gamma| Q_r^2 \kappa_{1,2}} \sqrt{\frac{k_B T_N}{2P_g \tau}} \tag{4.54}$$

$\kappa_{1,2}$ in Eq. 4.54 is either the real or imaginary part of κ, depending on whether we are measuring signal in the real or imaginary direction of S_{21}, respectively.

To convert to resolution on energy deposited in an MKID, we only need to multiply in the energy gap.

$$\sigma_{E,\,\mathrm{MKID}} = \frac{Q_c \Delta}{\alpha |\gamma| Q_r^2 \kappa_{1,2}} \sqrt{\frac{k_B T_N}{2P_g \tau}} \tag{4.55}$$

4.2.2 KIPM-Detector Energy Resolution

We would next like to estimate the resolution on measurements of phonon energy in the detector substrate. To do so, we must consider a number of additional factors. If our detector has N_r resonators on its surfaces, the phonon energy may be split among some or all of them. If the energy is split evenly, the total energy will be $N_r\times$ the energy observed in each MKID. For this reason, the full detector resolution will be increased by a factor of N_r. Conversely, if we record data from every MKID, we will have N_r independent measurements of the energy. By combining these measurements, we can remove random fluctuations and lower our resolution by a factor of $\sqrt{N_r}$. Both effects combine to raise the resolution by a factor of $\sqrt{N_r}$. This is a simplified handling of the energy partitioning, but gives a reasonable idea of the effect.

We must also consider the fact that not all phonon energy in the substrate will be absorbed by the active volume of an MKID. Some energy may be lost through the device mounting hardware or absorbed by insensitive metal on the substrate surface. Insensitive metal includes the feedline and MKID capacitors, neither of which will be sensitive to the breaking of Cooper pairs. While kinetic inductance will exist inside the MKID capacitors, their low current density (relative to the inductors) make changes in inductance insignificant to S_{21}[3]. Energy is also lost to sub-gap phonons (E<2Δ) that cannot break Cooper pairs and thus will not contribute observable energy. Phonon interactions within the substrate or surface metals can also cause phonons to down-convert to sub-gap energies. To account for the incomplete conversion of phonon to quasiparticle energy, we must define a collection efficiency factor (Eq. 4.56).

$$\eta_{ph} \equiv \frac{\text{energy absorbed by active MKID elements}}{\text{energy deposited in substrate}} \tag{4.56}$$

Total phonon energy will then be $1/\eta_{ph}$ times more than the combined energy observed in all MKIDs. Our resolution is therefore modified by the same factor.

When we combine all of these factors, we get the following resolution on total phonon energy in the MKID-based KIPM detector.

$$\sigma_{E,\,\text{KIPM}} = \frac{Q_c \Delta \sqrt{N_r}}{\alpha |\gamma| Q_r^2 \kappa_{1,2} \eta_{ph}} \sqrt{\frac{k_B T_N}{2 P_g \tau}} \tag{4.57}$$

We can make a number of observations using Eq. 4.57. Initially, it would appear that increasing readout power (P_g) will always improve resolution. In practice, the improvement is limited by other mechanisms. Even with readout frequency $f \ll 2\Delta/h$ ($\gtrsim$ 85 GHz in Al), Cooper-pair breaking by readout power is possible via multi-photon absorption, increasing quasiparticle density and lowering the quasiparticle lifetime. The signal decay time (τ from Eq. 4.45) depends on various time constants (including quasiparticle lifetime and phonon absorption time). If all other lifetimes are relatively short (imagine all phonon energy being absorbed instantaneously), τ is equivalent to just the quasiparticle lifetime. Therefore, a large P_g may lower τ depending on the other relevant lifetimes. Another consideration is our assumption of amplifier-noise domination. Increasing P_g will increase certain noise

[3]Section 4.3 of [65] shows that the effect on f_r and $1/Q_i$ due to changes in L_K and R is proportional to the current distribution squared.

sources intrinsic to the MKID (including quasiparticle generation-recombination noise[4]). Practically speaking, increasing P_g requires additional input amplification (or removal of input attenuation). Therefore, increasing P_g can add new and amplify existing input noise (including noise from the signal generator or amplifiers). If our device-intrinsic or input-readout noise grew to dominate the amplifier noise, Eq. 4.53 and its descendants would no longer be valid.

We also see that resolution is optimized when a device possesses only a single MKID ($N_r = 1$). This is true under the assumption that the number of MKIDs does not impact η_{ph}. The two values may be linked since additional MKIDs provide additional means for phonon energy to be absorbed. At Caltech, we have explored single and multi-MKID devices. We expect the multi-MKID devices to have poorer energy resolution and instead explore them for their position-reconstruction ability.

Another noteworthy feature is the importance of η_{ph}. If only 10% of phonon energy is absorbed into the MKID's quasiparticle system, the resolution will be 10× worse than its theoretical best. In reality, η_{ph} is limited to a maximum of ∼50% due to production of sub-gap phonons during quasiparticle generation [68]. Minimizing phonon-energy loss is a key consideration in our device and device-mount designs.

4.2.3 MKID and KIPM Designs

The MKIDs studied in this thesis were designed by Yen-Yung Chang and Osmond Wen. They use interdigitated capacitors and meandered inductors. In designs with multiple MKIDs, the inductor length is varied to differentiate each MKID in resonant frequency. The capacitor is kept the same in an attempt to keep the two-level system (TLS) noise consistent between MKIDs. It has been observed that TLS noise is capacitor-design dependent [66]. The inductor is designed to be symmetric about a central axis and is made of a pair of wound traces, rather than one long wound trace. These features were simulated to have good current uniformity and to minimize coupling between different MKIDs.

The MKID-coupled feedline is specifically a coplanar waveguide (CPW) with a finite ground plane. CPW feedlines have a central conductor with ground lines a distance away on both sides. All three sections lie on the same plane (the substrate surface). CPW was chosen to minimize the interaction distance of the dipole moments which appear between signal and ground lines. Previously tested coplanar striplines (CPS)

[4]Quasiparticle density in an MKID will fluctuate due to random breaking and thermal recombination of Cooper pairs. The fluctuation in number is assumed to follow Poisson statistics and leads to an intrinsic MKID noise we refer to as generation-recombination noise [65].

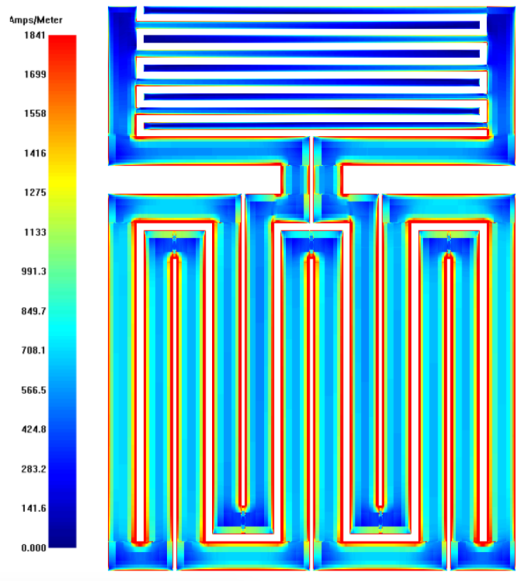


Figure 4.2: The general MKID design utilized throughout this thesis with simulated current density. The upper portion with relatively low current density (dark blue) is the interdigitated capacitor. The lower portion with higher current density (light blue/teal) is the symmetric coplanar-stripline inductor. Simulation and image produced by Yen-Yung Chang [69].

.

were observed to have transmission which varied significantly with frequency. The variation was partially attributed to the extended dipole field interacting with the device enclosure [70]. The CPS feedlines were also made with a very narrow gap between the signal and ground lines. The narrow gap introduced fabrication challenges and made it common for debris (such as small metal flakes) to cause shorts. The CPW ground line closest to each MKID is extended to surround the MKID entirely. This "ground shield" lowers crosstalk between MKIDs, which can cause resonance frequencies to shift from intended values [69]. The CPW ground plane is made finite (rather than covering the remaining surface of the substrate) in order to minimize the phonon energy absorbed and lost to this insensitive metal (preserving η_{ph}).

We primarily fabricated devices using niobium and aluminum. Nb has a T_c around 9 K and Δ_0 around 2 meV. Al has a T_c around 1 K and Δ_0 around 0.2 meV. Nb's higher T_c makes it useful for testing feedlines and MKIDs in refrigerators which are only capable of reaching liquid-helium temperature (4 K). In lower temperature refrigerators, Nb is useful for testing MKIDs significantly below their T_c. Al's lower Δ makes it a better choice for achieving lower energy resolutions and thresholds

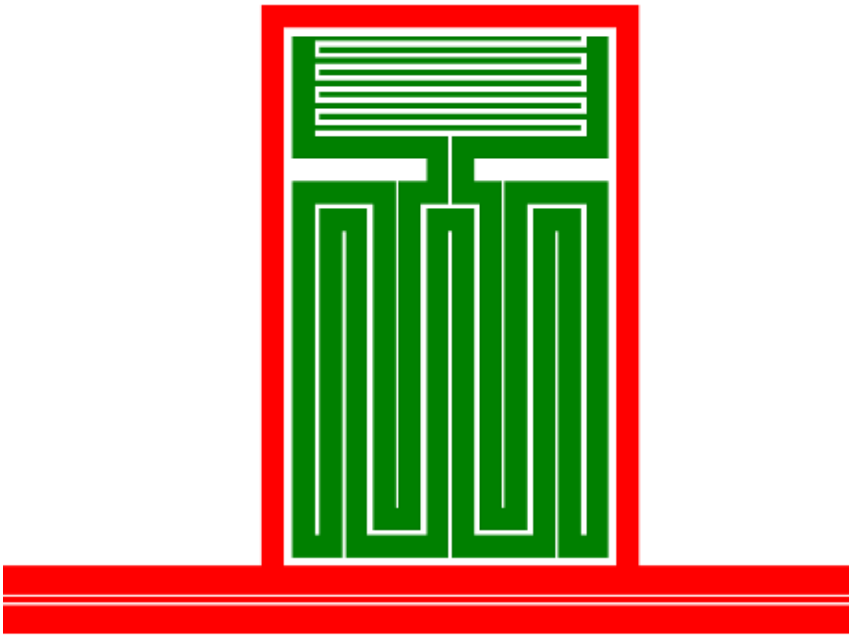


Figure 4.3: The general MKID design in green with the coplanar waveguide (CPW) feedline and ground shield shown in red.

as seen in Eq. 4.57. We fabricated some devices using both Nb and Al, typically making the feedline and MKID capacitors out of Nb while making the MKID inductors Al. The difference in Δ makes it more likely for phonon energy to be absorbed in the Al MKID inductors rather than the Nb feedline, improving η_{ph}.

4.2.4 Substrate Choice

A KIPM dark-matter detector would be fabricated on substrates similar to the SuperCDMS SNOLAB or HVeV detectors. As with those detectors, the material choice would be based on the expected dark-matter interaction rate for masses and models of interest. Silicon and germanium are likely materials, although others such as sapphire are becoming common as well [71]. The volume choice would be based on the required exposure and capability of the cryogenic setup. The first detector used for science data would likely be closer in size to an HVeV detector since it should probe new model parameters and is unlikely to be exposure limited. At Caltech, we have fabricated devices on silicon wafers. This allows us to benefit from the large amount of institutional knowledge about the material. Our test wafers are 1 mm thick. In this thesis, I will consider devices made on 3" round wafers and 2 cm square wafers.

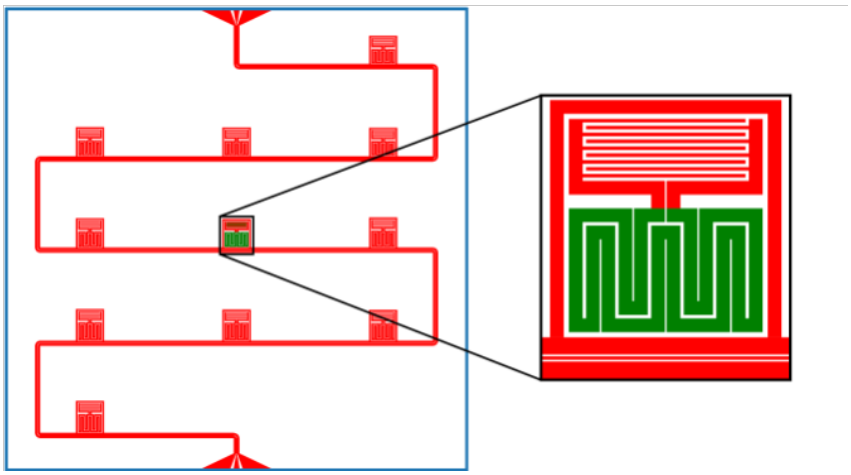


Figure 4.4: An example KIPM-detector design with CPW feedline and MKIDs. The feedline widens near either end to make wire-bonding possible. A zoom-in on the central MKID shows (in red) the parts that would be fabricated out of Nb (feedline, ground shield, and capacitor). Only the inductor (green) is sensitive to energy dispositions and would therefore be fabricated out of Al. In this design, only one MKID was intended to be sensitive. The others would be used for calibration purposes.

4.3 Early Device Tests

In the following section, we will discuss the KIPM testing performed at Caltech to improve our device design and characterization abilities. The feedlines and devices described below were all fabricated at the NASA Jet Propulsion Lab (JPL).

4.3.1 Feedline-only Tests

As outlined in Brett Cornell's thesis [70], early tests of complete detector designs showed a number of undesirable features. These included high transmission loss and variation with frequency. In order to isolate the issue, we fabricated and tested devices with only the CPW feedline (no MKIDs or ground shields). The design for one such device is shown in Figure 4.5. This feedline could couple to many MKIDs at many points across the detector. Other designs featuring shorter and simpler feedlines were tested for comparison. We fabricated these devices out of Nb so we could test them at half the Nb transition temperature in the 4-Kelvin cryostat (see Section 2.5).

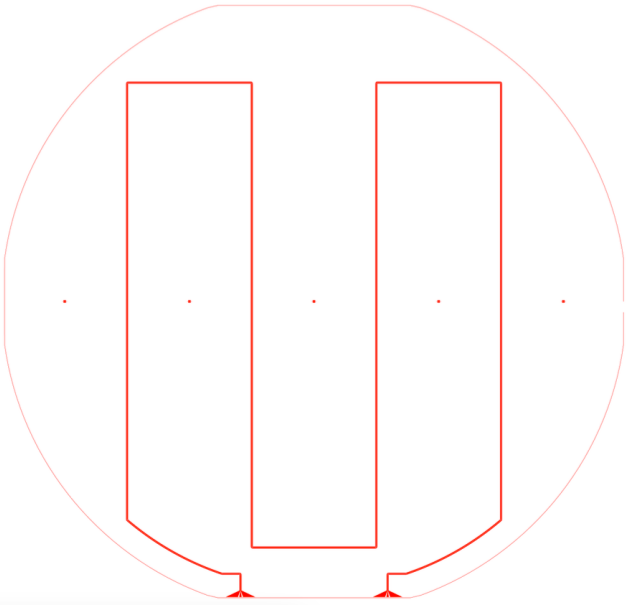


Figure 4.5: The mask design for producing a feedline-only device. This device was used to produce the results in Figure 4.7. The solid red trace going up and down the wafer is the CPW feedline (the separation between signal and ground is impossible to see at this scale). The wider pads at the bottom were used for wire-bonding.

After fabrication, the device was mounted in our 3" wafer testing box. The box is shown in Figure 4.6. The primary purposes of the box are thermal sinking, connecting coaxial cables to the feedline, and minimizing optical loading. The 3" box is gold-plated copper.

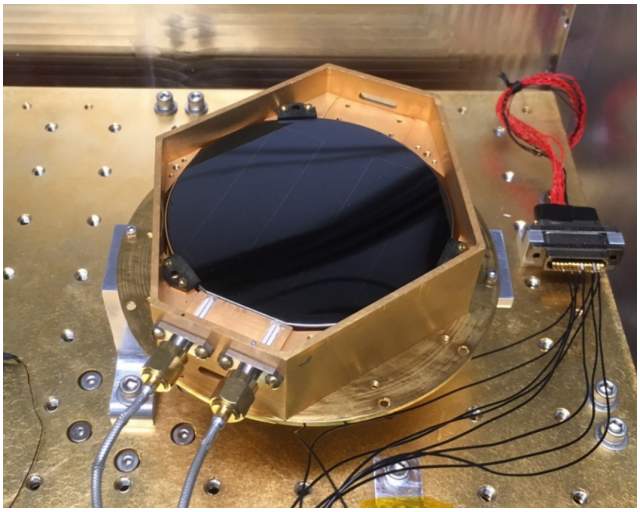


Figure 4.6: A feedline-only device (using the mask from Figure 4.5) mounted in the 3" device box. The device box is mounted in our 4-Kelvin cryostat. In this photo, the box lid was removed. The box lid is a flat hexagon which would otherwise close the box, blocking our view of the device and providing some protection from light.

After cooling the device in our cryostat, the test consisted of measuring RF transmission and reflection as functions of frequency using a vector-network analyzer (VNA). The setup only required connecting the two VNA ports to the two device ports using coaxial cables and cryostat feed-throughs.

Shorter feedlines with fewer turns were observed to have more uniform transmission [69]. Unfortunately, short feedlines are not sufficient to construct a multi-MKID KIPM capable of accurate position reconstruction. We suspect the uneven transmission to be (at least partially) due to feedline coupling to the resonant modes of the cavity formed by the device box. The copper device box creates E=0 boundary conditions and therefore defines the length-scale of the coupled modes. We experimented with minimizing the coupling by comparing the transmission of the device using various configurations of the box: box with both lids attached, box with both lids removed and raised off the cold-plate surface, and box with both lids attached and microwave-absorbing Eccosorb foam placed between the device and the feedline-facing lid. Changing the box configurations altered the boundary

conditions that form the resonant modes. The resulting transmission measurements are shown in Figure 4.7.

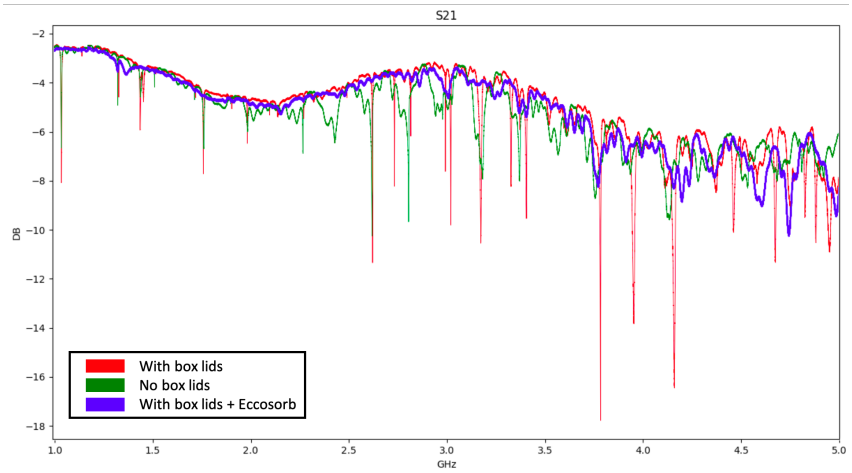


Figure 4.7: The feedline-only device transmission in various states of enclosure. The red transmission is mounted in the box with both lids attached. The green transmission is mounted in the box with both lids removed and raised off the cold-plate surface. The blue transmission is mounted in the box with both lids attached and microwave-absorbing Eccosorb foam placed between the device and the feedline-facing lid. Due to uncertain attenuation before and after the device, the transmissions should only be considered relative to each other.

The results confirmed that coupling to the box modes was one major contributor to feedline-transmission structure. Removing the lids removed structure at some frequencies and added structure at others. The structure that disappeared with the lids was likely caused by box-mode coupling. Structure that appeared when the lids were removed was likely caused by coupling to the resonant modes of the cavity formed by the cryostat itself (no longer shielded by the lids). Structure that appeared with and without the box lids may have been the result of coupling to modes defined by the non-lid parts of the box. While it is difficult to assign each transmission feature to a specific coupling, the results confirm that the box does have a large effect.

The Eccosorb configuration was included to approximate replacing the lid with infinite free space (another boundary condition of interest). Since Eccosorb is a microwave absorber, we would expect the box-mode coupling to disappear without exposing the device to additional cryostat-mode coupling. This seems to agree with the last result in Figure 4.7. The sharpest structures in the transmission disappeared without significant new structure appearing.

An outstanding concern is the temperature of the Eccosorb foam during the test. We did not have an adequate method to thermally sink the foam to the device box. The foam was therefore at a completely unknown temperature during the test. It may have remained significantly above 4 Kelvin. This would have caused the foam to radiate infrared (IR) photons directly at the substrate and feedline. IR photons should not affect the transmission except to lower it overall due to the feedline possibly having an effective temperature above 4 Kelvin. That said, the overall transmission being similar in magnitude to both other tests confirms the Nb remained superconducting. Radiating Eccosorb would not be acceptable for use near an actual device. Raising the effective temperature of the MKIDs would degrade their internal quality factors and quasiparticle lifetimes[5]. Eccosorb is also too radioactive for placement adjacent to a detector searching for dark matter.

Satisfied that the worst transmission features were not inherent to the CPW feedline design or fabrication, we proceeded to make (and test) full devices with MKIDs and ground shields included. It is possible that the transmission features may be removed in the future using a non-radioactive heat-sinkable absorber instead of Eccosorb. Another option may be to update the feedline with special corner designs intended to keep the signal and ground lines in phase thus minimizing the distance over which the feedline can couple [72].

4.3.2 80 MKID, Nb-feedline KIPM

One of our most exhaustively studied devices was fabricated by Yen-Yung Chang beginning on July 26th, 2018. As the second of two devices started by Yen-Yung on that date, it received the identifier YY180726.2. This device used 80 Al MKIDs with designed frequencies between 3.05 and 3.45 GHz. The feedline was fabricated out of Nb to minimize wasted phonon energy (as described in Section 4.2.3). The device substrate is a 1-mm thick, 3"-diameter Si wafer. The Al MKIDs are 30 nm thick. The Nb feedline is 300 nm thick to ensure good continuity across the wafer. An image of the mounted device YY180726.1 can be seen in 4.8. After fabrication, this device was used to for many of the initial MKID-characterization tests described below.

[5]We did see degraded internal quality factors when the Eccosorb test was repeated using a KIPM with previously characterized MKIDs.

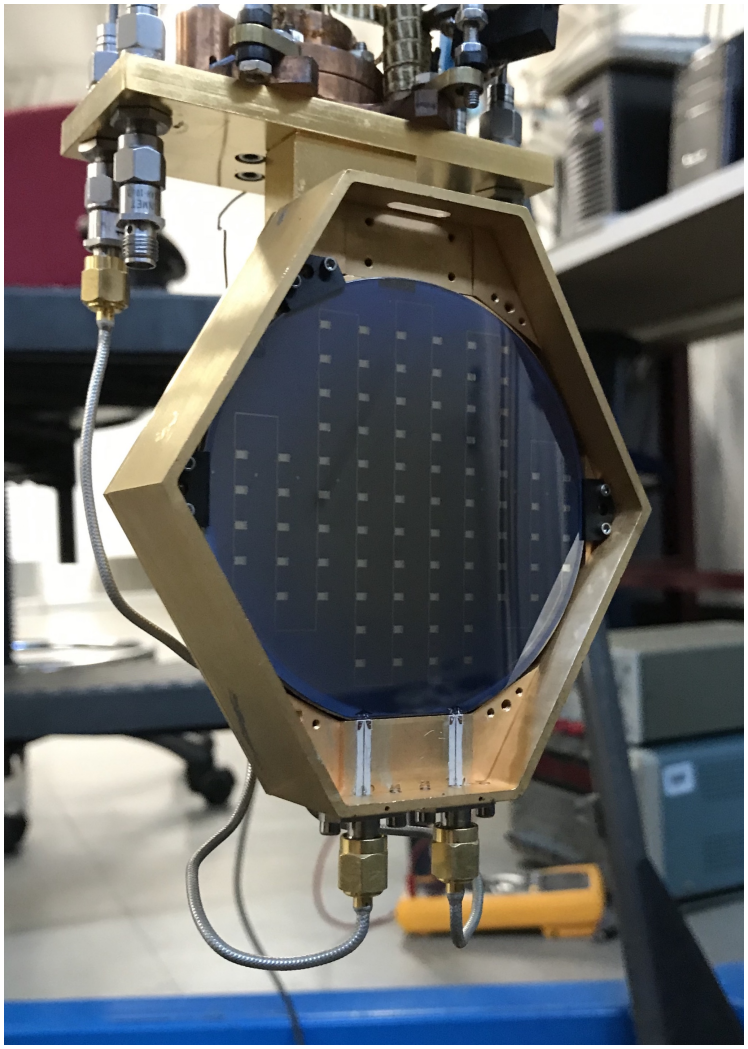


Figure 4.8: The device referred to with the identifier YY180726.2 mounted in a device box (with lid removed) in our Oxford Kelvinox 25. This device has 80 Al MKIDs and a Nb CPW feedline.

4.3.3 Resonance Identification and Characterization

The first step in characterizing a device is to cool it down and study its transmission. For Al devices, we primarily used our Oxford Kelvinox 25 dilution refrigerator, capable of achieving temperatures as low as 40 mK (with minimal load) and maintaining low temperature indefinitely (with regular liquid helium transfers). Operating at base temperature during testing is impractical, because the additional heat load from the cryogenic amplifier can cause the temperature to change during testing. For this reason, we normally used a PID-controlled heater to maintain a specific temperature above base (say 60 or 100 mK). The heater could then compensate for changes in amplifier power to maintain a constant temperature.

As with the feedline, the MKID transmission measurement is done using a VNA. The first transmission measurement of this device in our dilution refrigerator is shown in Figure 4.9. The temperature was not controlled during this run, but was observed to be relatively stable between 60 and 70 mK. The VNA was set to output -15 dBm of power with 40 dB of attenuators between it and the device. The cryostat coaxes and feed-throughs provide additional attenuation, so the power delivered to the device was <-55 dBm.

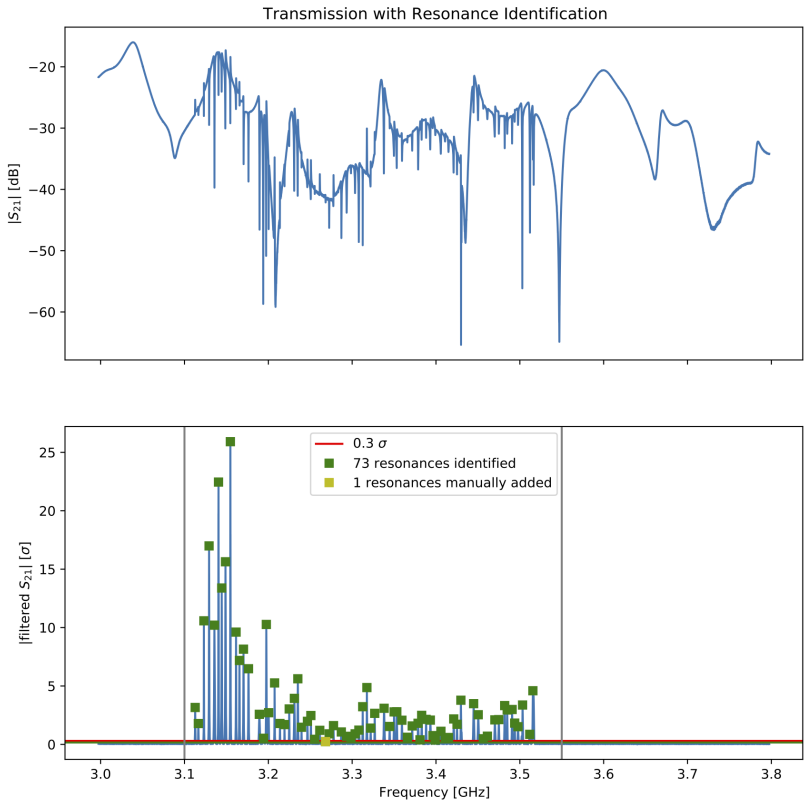


Figure 4.9: The 80-MKID KIPM characterization data. (Top) Transmission of YY180726.2 recorded on October 19th, 2018. VNA readout power was -15 dBm. 40 dB of attenuation was present between the VNA and feedline input. (Bottom) Magnitude of the transmission after being high-pass filtered. Resonances identified with the resonance-search algorithm are marked in green. A manually located resonance is marked in yellow. The threshold used to identify resonances is shown in red.

The next step in studying the device was locating and fitting the MKID resonances to extract various parameters. To expedite the locating process, I wrote and applied

a resonance-search algorithm. First, the algorithm high-pass filters the real and imaginary transmissions to remove the broad-frequency structure while leaving the sharp resonances. The real and imaginary filtered transmissions are then added in quadrature. The resulting value should be sharply peaked at any frequencies with sharp structure (including resonances). The algorithm then locates regions in frequency-space where the magnitude of the filtered transmission passes above a set threshold (before dropping below a different threshold). The first threshold is defined by the user in number of standard deviations of the total filtered transmission. The second threshold is set to 3× the mean filtered transmission of all data below the first threshold (and is therefore always lower). For each above-threshold region, the algorithm calculates the frequency with maximum filtered transmission and assigns that as the estimated resonant frequency. Resonance frequencies can also be added manually. The results of running the resonance-search algorithm on YY180726.2 can be seen in Figure 4.9. Figure 4.10 zooms in on the same result to display the thresholds.

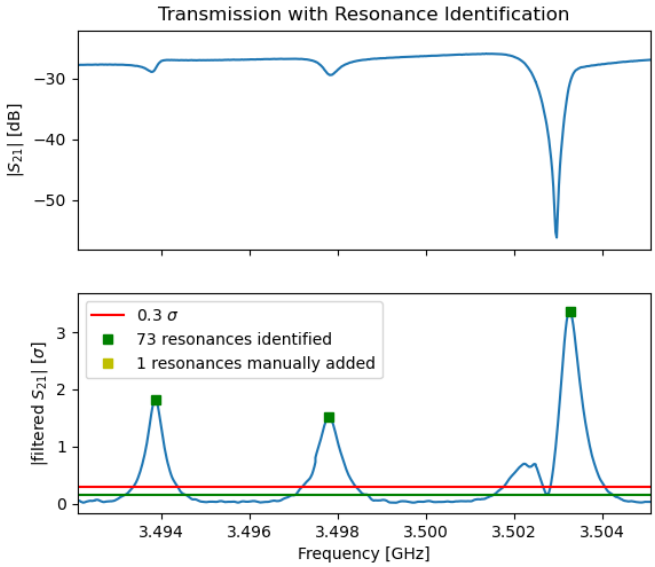


Figure 4.10: The same result shown in Figure 4.9 zoomed in to a frequency range covering three identified resonances.

The next step in device characterization is using the estimated resonance frequencies and measured transmission to fit each individual MKID resonance. We fit to Eq. 4.58 which is equivalent to Eq. 4.9 with various additional effects taken into consideration.

$$S_{21}(f) = ae^{-2\pi j(f-f_r)\tau}\left[1 - \frac{\frac{Q_r}{|\widehat{Q}_c|}e^{j\phi}}{1 + 2jQ_r\left(\frac{f-f_r}{f_r}\right)}\right] \tag{4.58}$$

First, the complex factor a is included to account for the overall transmission of the readout at frequencies near resonance. The factor τ is added to account for the frequency-dependent phase shift due to cable delay. $1/Q_c$ in Eq. 4.9 is replaced with $e^{j\phi}/|\widehat{Q}_c|$ to account for mismatched input or output impedance [73]. It is important to note that $Q_c \neq |\widehat{Q}_c|$ and is instead extracted from the fit using Eq. 4.59.

$$\frac{1}{Q_c} = Re\left\{\frac{1}{\widehat{Q}_c} = \frac{e^{j\phi}}{|\widehat{Q}_c|}\right\} = \frac{\cos(\phi)}{|\widehat{Q}_c|} \tag{4.59}$$

Eq. 4.58 is the same equation used by Jiansong Gao [65]. The only alteration for my fitting was changing τ from a real to a complex number. For this formulation, the cable delay is equivalent to the real part of τ. The imaginary part of τ accounts for the overall transmission's frequency dependence near each resonance (to first order). We see the result of fitting this equation to one resonance in Figure 4.11. The example resonance shows a downward slope (in transmission magnitude) around the resonance that could not be accounted for without an exponential dependence on frequency (provided by the imaginary component of τ).

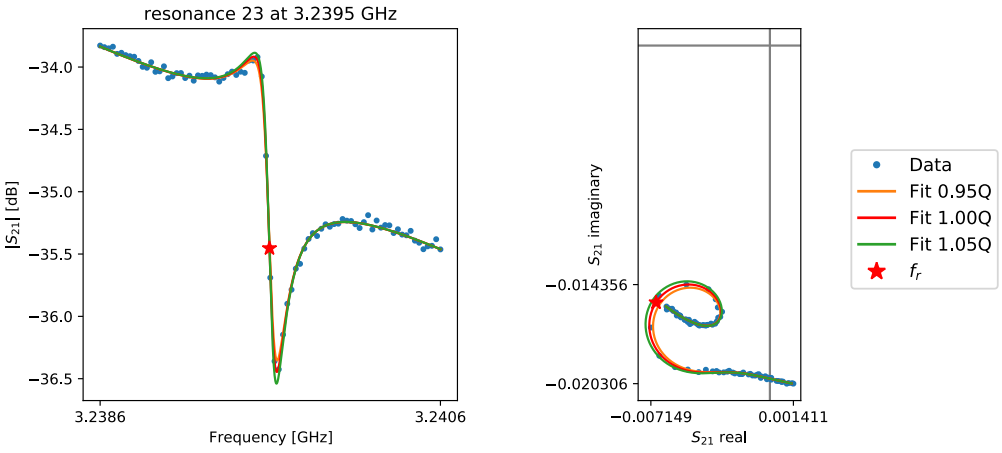


Figure 4.11: An example fit to one resonance from the data in Figure 4.9. The fit is plotted in red. The fit with Q_r varied by ±5% is shown in orange and green for comparison.

Eq. 4.58 includes a large number of fit parameters. Accordingly, the fit results can vary based on the initial parameter estimates. For all parameters except τ, I

calculated initial estimates using the methods described in Appendix E.2 of Gao's thesis [65]. For τ, I estimated the real and imaginary components by fitting first-order polynomials to the in-phase and quadrature-phase transmission for frequencies slightly above and below each resonance.

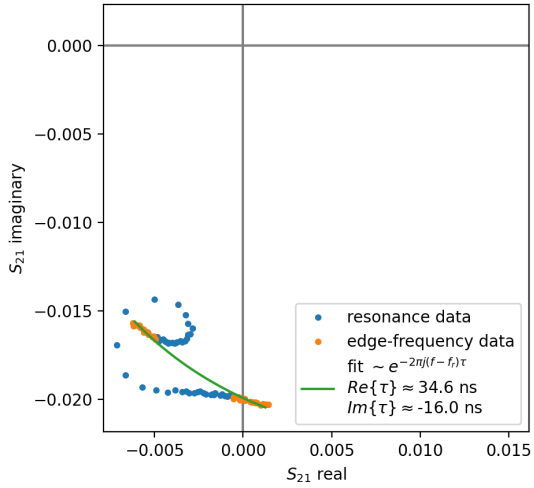


Figure 4.12: An example fit to frequencies slightly above and below a resonance to extract the real and imaginary components of τ. The same example resonance shown in Figure 4.11.

The resonance-fitting code was effective for nearly all resonances studied. Failures were almost exclusively due to large transmission features (or additional resonances) appearing too close in frequency to the resonance of interest. The distribution of quality factors for the resonances in Figure 4.9 can be seen in Figure 4.13. The median values are in agreement with Eq. 4.10.

We see that a significant number of resonances returned negative coupling quality factors. This is the result of evaluating Eq. 4.59 with ϕ between $\pi/2$ and $3\pi/2$. Such values were not considered in Khalil's analysis of resonator asymmetry caused by impedance mismatches [73]. One resonance was observed with negative internal quality factor. This was the result of Eq. 4.59 returning a coupling quality factor lower than the fitted total quality factor. Any negative quality factor is an nonphysical result and Eq. 4.59 therefore requires further investigation.

Figure 4.14 shows that the typical frequency separation in YY180726.2 was ~5 MHz. This is in agreement with the ~4.9 MHz separation designed by Yen-Yung Chang (see Eq. 3.18 in [69]). The six largest separations were closer to twice the

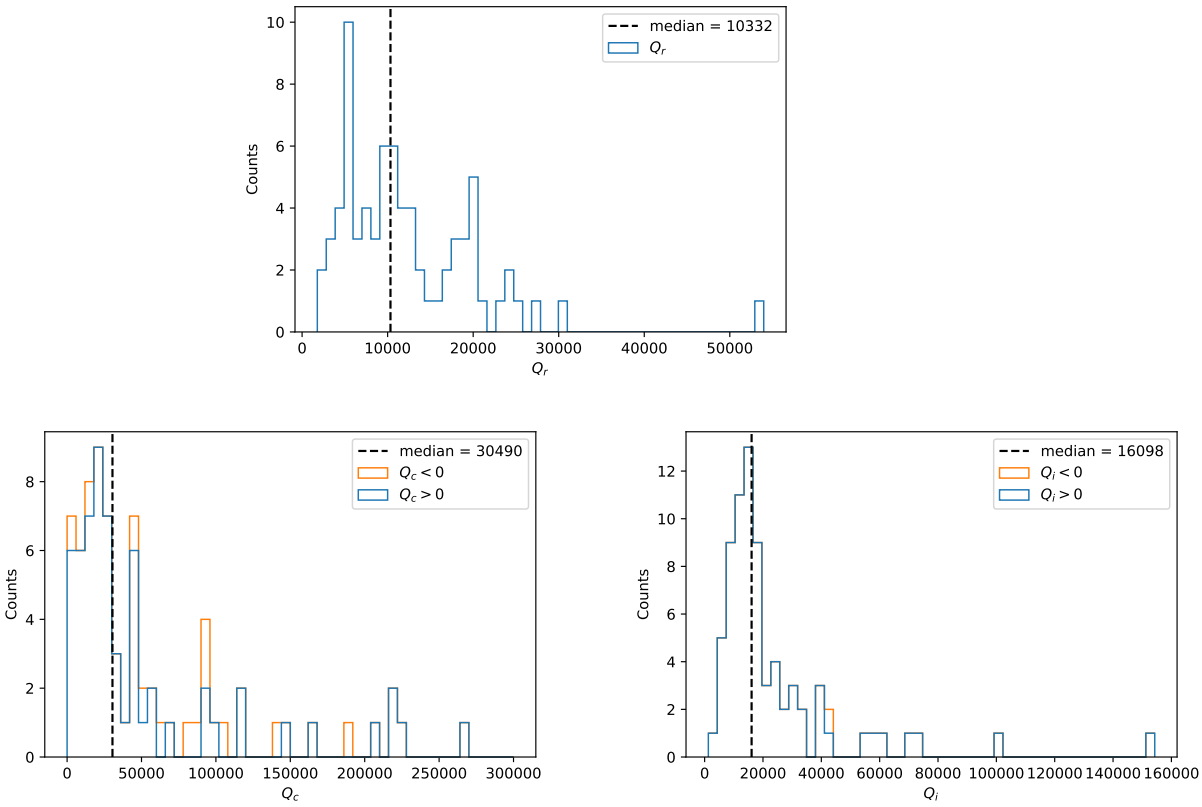


Figure 4.13: Distributions of quality factors for the resonances in Figure 4.9. (Top) The total resonance quality factors and median. (Bottom-left) The coupling quality factors and median. An additional value at $\sim 1.9 \times 10^6$ was excluded for plot clarity. Negative quality factors are plotted in absolute magnitude. The median was calculated using the absolute value of all factors. (Bottom-right) The internal quality factors and median. One negative quality factor was observed.

intended separation and may be gaps caused by the missing six resonances (the device has 80 MKIDs but only 74 identified resonances).

4.3.4 Readout-Power Sweeps

As seen in Eqs. 4.55 and 4.57, resolution will generally improve with readout power. The improvement will inevitably fail if power is raised too high. As discussed in Section 4.2.2, increased readout power will generate additional quasiparticles that degrade quasiparticle lifetime (eventually degrading resolution). The increased quasiparticle density will also increase dissipation and decrease the resonance quality factors (further degrading resolution). Sometimes, MKIDs have a narrow range where increasing readout power will increase internal quality factor. The additional

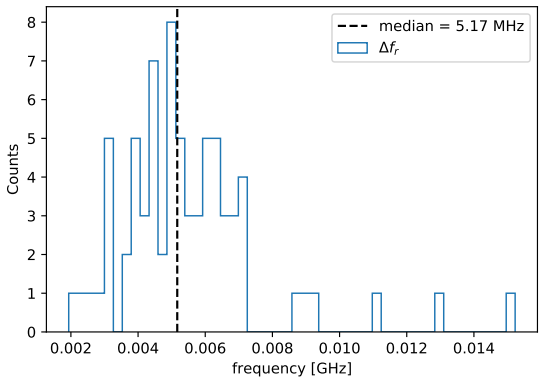


Figure 4.14: The spacing between adjacent resonant frequencies identified in Figure 4.9. The median separation is ~5 MHz.

power causes quasiparticles to block some final states involved in quasiparticle-photon scattering. The reduction in quasiparticle scattering reduces dissipation and increases internal quality factor [74]. Changes in quality factor can be studied by characterizing MKID resonances as a function of readout power. The result of performing such a study on the resonance from Figure 4.11 can be seen in Figure 4.15.

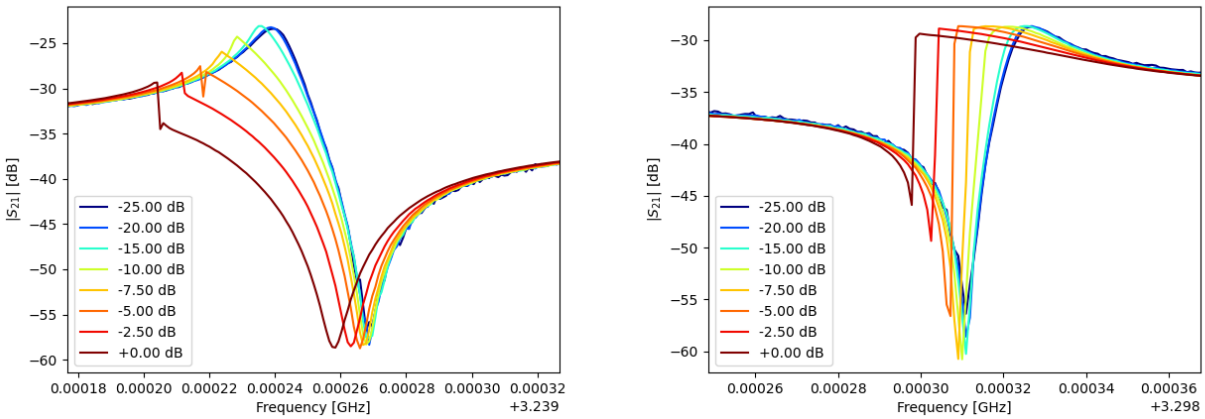


Figure 4.15: Transmission as a function of readout power for two MKIDs. The readout powers were in units of dBm from the VNA. There was at least 40 dB of cold attenuation (and unknown warm attenuation) between the VNA and MKIDs. Due to the uncertain attenuation, the powers should only be compared to each other (and are therefore labelled in dB). The y-axis is in dB referred to the VNA output power. For both MKIDs, we see clear signs of bifurcation for powers >-10 dB. (Left) The result for the MKID characterized in Figure 4.11. (Right) The result for another MKID from Figure 4.9.

Sufficiently high readout powers will cause resonances to undergo bifurcation. Specifically, this will occur when the readout-generated current becomes appreciable compared to the superconductor's critical current. For such powers, a resonance will appear differently depending on whether data is taken with increasing or decreasing frequency. In either case, the readout will generate current in the MKID before reaching the resonant frequency. The current will modify the MKID inductance due to nonlinear kinetic inductance [66]. When taking data with increasing frequency, the resonant frequency will move towards the readout frequency. As the two frequencies approach, coupled current will increase and the effect will intensify. This can lead to abrupt changes in recorded transmission. Such an abrupt change can be seen clearly in the right plot of Figure 4.15. The left plot is subject to the same effect but is more difficult to understand due to the asymmetric resonance shape. When taking data with decreasing frequency, the resonant frequency will move away from the measurement frequency. This can cause the recorded resonance to appear broader in frequency. Eventually the coupled current will decrease sufficiently that the resonant frequency shifts back to its initial state (possibly producing an abrupt change in recorded transmission). A more in-depth description and visualization of bifurcation can be found in Section 2.2.6 of [66].

4.3.5 Temperature Sweeps

Thermally generated quasiparticles will also alter the dissipation and kinetic inductance of MKIDs. This effect can be used to calculate the parameters required to calibrate absorbed phonon energy. In practice, this means recording transmission as a function of temperature and fitting the changing resonance parameters to a model based on Mattis-Bardeen theory. Such a fit will produce the MKID's kinetic inductance fraction (α) and the superconducting energy gap at zero temperature (Δ_0). These values can be used with Eq. 4.41 to convert small signals to units of energy.

I wrote the code to record transmission for each resonance of interest at each desired dilution-refrigerator temperature. Each temperature was set using a PID-controlled heater mounted on the mixing-chamber stage of the refrigerator. The resonance-fitting code was then used to extract the resonant frequencies and quality factors for each temperature. Example data and fits for one MKID can be seen in Figure 4.16.

The resultant $f_r(T)$ and $Q_i(T)$ could then be fit using Mattis-Bardeen theory. The fit equations can be derived using results from Section 4.1.4. First, plug Eq. 4.36

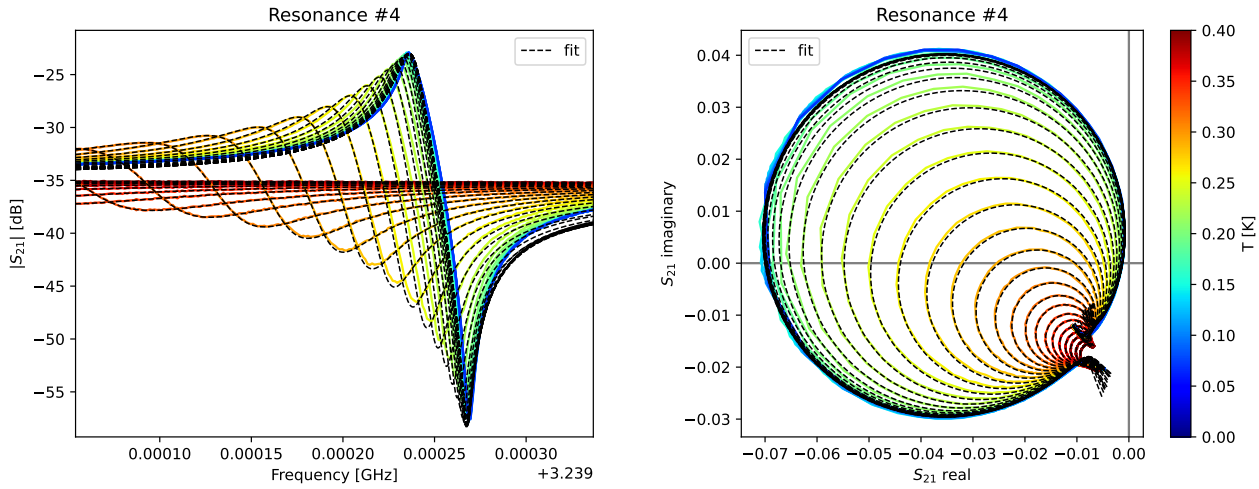


Figure 4.16: Transmission as a function of temperature for the MKID characterized in Figure 4.11. (Left) The transmission magnitude. (Right) The transmission in complex space. In this example, the fit is imperfect for the lowest temperatures. Data was taken with -15 dB readout power. From Figure 4.15, we see that -15 dB slightly distorted the resonance shape (compared to -20 dB). This likely caused the imperfect fit. The fitting still does a good job characterizing the change in shape as temperature increases.

into Eq. 4.18 and Eq. 4.35 into Eq. 4.19. As was done at the end of Section 4.1.4, we will assume $\omega = \omega_r$, $L_K \approx L_{K,0}$, and that Q_i is sufficiently large to ignore the second term in Eq. 4.19.

$$\begin{aligned}\frac{d\omega_r}{dn_{qp}} &= -\alpha\frac{\omega_r}{2L_K}\left(L_{K,0}|\gamma|\kappa_2\left(T,\omega,\Delta_0\right)\right)\\ &= -\alpha\frac{\omega_r}{2}|\gamma|\kappa_2\left(T,\omega_r,\Delta_0\right)\end{aligned} \tag{4.60}$$

$$\begin{aligned}\frac{d1/Q_i}{dn_{qp}} &= \alpha\frac{1}{\omega_r L_K}\left(\omega L_{K,0}|\gamma|\kappa_1\left(T,\omega,\Delta_0\right)\right) - \alpha\frac{1/Q_i\,\cancel{dL_K}}{\cancel{2L_K}\,dn_{qp}}\\ &= \alpha|\gamma|\kappa_1\left(T,\omega_r,\Delta_0\right)\end{aligned} \tag{4.61}$$

Eqs. 4.60 and 4.61 can then be integrated from $n_{qp} = 0$. We label the $n_{qp} = 0$ values of ω_r and Q_i as ω_0 and $Q_{i,0}$, respectively. For this integration and the subsequent fitting, we evaluate κ at $\omega = \omega_0$ and T at the the measured refrigerator temperature (as opposed to quasiparticle temperature). Figure 4.1 implies κ is a weak function of

T. κ depends on ω only through $\xi = \hbar\omega/2k_BT$, so the dependence on ω is similarly weak. Therefore, these choices are acceptable.

$$\begin{aligned} ln\left(\frac{\omega_r}{\omega_0}\right) &= -\alpha\frac{1}{2}|\gamma|\kappa_2\left(T,\omega_0,\Delta_0\right)n_{qp}(T) \\ &\approx \frac{\omega_r}{\omega_0} - 1 \end{aligned} \tag{4.62}$$

$$\rightarrow \omega_r(T) = \alpha\frac{\omega_0}{2}|\gamma|\kappa_2\left(T,\omega_0,\Delta_0\right)n_{qp}(T) + \omega_0 \tag{4.63}$$

The approximation of the logarithm in Eq. 4.62 is valid when $\omega_r/\omega_0 \approx 1$. This can be experimentally confirmed. In Figure 4.16, we see that f_r changes by ~0.5 MHz for f_0 ~3.2 GHz ($\omega_r/\omega_0 - 1 \lesssim 2\times10^{-4}$).

$$\frac{1}{Q_i} - \frac{1}{Q_{i,0}} = \alpha|\gamma|\kappa_1\left(T,\omega_0,\Delta_0\right)n_{qp}(T) \tag{4.64}$$

$$\rightarrow Q_i(T) = \left[\alpha|\gamma|\kappa_1\left(T,\omega_0,\Delta_0\right)n_{qp}(T) + Q_{i,0}^{-1}\right]^{-1} \tag{4.65}$$

Fitting to Eqs. 4.63 and 4.65 requires understanding how the density of thermal quasiparticles (n_{qp}) depends on temperature. This can be approximately derived using Eq. 4.29 with the assumptions previously described for κ (below Eq. 4.33). We will use the result from [75].

$$n_{qp}(T,\Delta_0) = 2N_0e^{\Delta_0/k_BT}\sqrt{2\pi k_BT\Delta_0} \tag{4.66}$$

During fitting, we assumed N_0 (the density of states for electrons of a single spin at the Fermi energy level) of $1.72\times10^{10}\,\mu m^{-3}eV^{-1}$ (from [65]). We also used the approximate forms for κ (Eqs. 4.31 and 4.32). It is worth remembering that these approximate forms are for athermally generated quasiparticles but were shown to be nearly equivalent to the thermal quasiparticle result by Gao [65]. A more accurate calculation could be done using the thermal κ approximation when fitting the temperature-sweep results then using the athermal κ approximation when calibrating events. The most accurate result would be obtained by using the integral form of κ during both steps. The MKIDs were 30 nm thick, so we used the thin-film and local limit assumptions for the Meissner effect ($\gamma = -1$).

The result of fitting the model to our example MKID is shown in Figure 4.17. The fit extracted the expected energy gap for aluminum (~0.2 meV) as well as the kinetic inductance fraction ($\alpha \approx 6\%$).

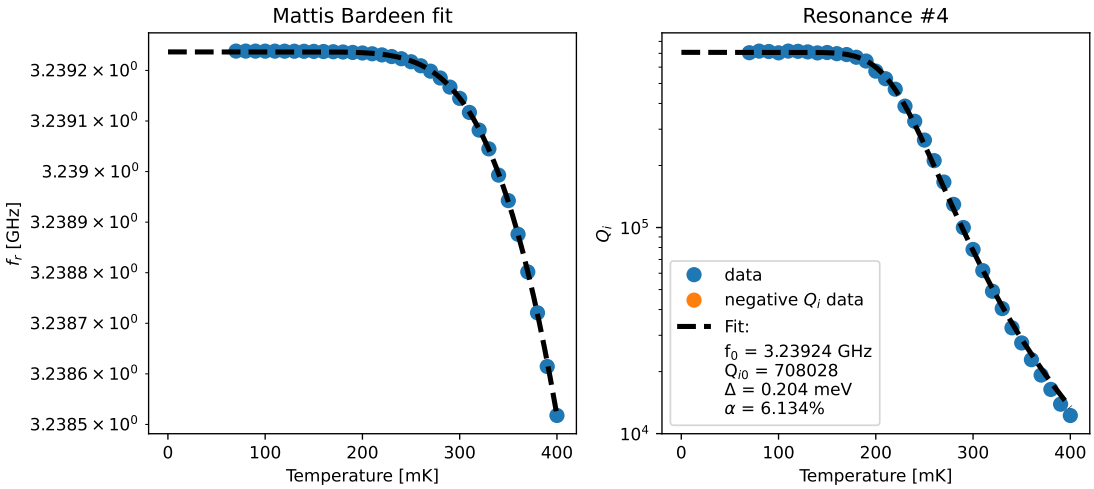


Figure 4.17: The result of using Mattis-Bardeen theory to fit the data from Figure 4.16. (Left) $f_r(T)$ with the fit result. (Right) $Q_i(T)$ with the fit result and legend. $f_r(T)$ and $Q_i(T)$ were fit simultaneously to produce a shared fit result.

All the software used for MKID characterization (data taking, identification, resonance fitting, Mattis-Bardeen fitting) is available online[6].

4.4 Readout Design

As described above, MKID characterization can be performed using only a VNA readout. During an event search, a more complicated readout is required to record signals in multiple MKIDs simultaneously with sufficient timing and energy resolution. Additionally, MKID characterization did not require particularly low noise, since time-averaging could be performed for as long as necessary. Averaging is not feasible during event-oriented data taking. It is therefore necessary to use a multi-tone readout system designed so that noise is dominated by the device and/or the low-noise first-stage amplifier. In this section, we will go over the basics of noise management in RF readout systems before discussing the specific system we used to measure our MKIDs' amplifier-dominated resolution.

[6] `https://github.com/mkid-caltech/PyMKID`.

4.4.1 MKID Readout Noise and Isolation

When reading out a low-noise cryogenic device, input-side noise coming from room-temperature components (including Johnson noise from room-temperature attenuators) can contribute significantly to the total noise. To mitigate the effect, it is common to include attenuators at each thermal stage in the readout chain. These attenuators reduce the noise coming from higher thermal stages while simultaneously contributing their own Johnson noise. The net effect is to decrease the total noise as Johnson noise decreases with temperature. Johnson noise is sometimes referred to as blackbody noise (or power) since it is equivalent to one-dimensional blackbody radiation produced by attenuators and propagated along the transmission line.

The noise temperature observed at the output of an attenuator is a function of attenuator temperature (T_A), attenuator loss (L_A), and noise temperature incident on the attenuator input (T_{in}).

$$T_{out} = \frac{T_{in}}{L_A} + T_A \left(1 - \frac{1}{L_A}\right) \tag{4.67}$$

Loss is defined as the ratio of power-in to power-out. The inverse of loss is gain. Gain is generally used for amplifiers while loss is used for attenuators. Both values are often described in decibels (dB).

$$L_A = \frac{power_{in}}{power_{out}} = \frac{1}{G_A} \tag{4.68}$$

$$G_A = \frac{power_{out}}{power_{in}} = \frac{1}{L_A} \tag{4.69}$$

$$L_A\,(dB) = 10\log_{10}(L_A) = 10\log_{10}\left(\frac{1}{G_A}\right) = -G_A\,(dB) \tag{4.70}$$

With Eq. 4.67, one can see that for a lossless attenuator ($L_A = 1$), the output noise is equivalent to the input noise. In this case, the attenuator acts as an ideal transmission line. In the case of complete loss ($L_A \to \infty$), the output noise is equivalent to the attenuator's physical temperature. This case is equivalent to the attenuator acting as a terminator. A more typical example would be using a 10 dB attenuator ($L_A = 10$) held at liquid-helium temperature (4 K) to reduce the noise of a room temperature device. If the room temperature device had an output noise-temperature of 300

K, the attenuator would output 33.6 K (using Eq. 4.67). In MKID readouts, a typical goal for attenuator placement is to keep the input noise at the MKID-device dominated by the device itself, while also achieving the required readout power. If the liquid-helium attenuator was followed by additional 10 dB attenuators at 1 K and 100 mK, the noise into the MKID device would be down to 516 mK (with 30 dB of total signal attenuation).

RF amplifiers are used to apply positive gain to signals after passing through an MKID-coupled device. These amplifiers apply identical gain to both signal and noise. They also contribute noise of their own. Therefore, amplifiers do not directly improve the signal-to-noise ratio (SNR) but instead are used to raise both signal and noise above the effects of subsequent noise sources. MKID readouts generally use multiple stages of amplification. The first-stage amplifier (or preamp) should have the lowest possible input noise. The preamp is used to raise signal and noise above the following amplifier's input noise. Cryogenic HEMTs are often used as first-stage amplifiers, although there is great potential for parametric amplifiers (which we will discuss later). Subsequent amplifiers are used to achieve total gain sufficient to make the readout system's digitizer noise subdominant.

It is common to include some form of isolation between an MKID device and its first-stage amplifier. MKID-coupled feedlines often have imperfect impedance matching (as do amplifiers), so the interface can have non-negligible reflection scattering coefficients (S_{11} or S_{22}). Reflections can create standing waves in the readout, as signal and noise are bounced between the device output and amplifier input. A simple solution is to include a small attenuation between the two components to damp reflections. However, an attenuator would also add noise and lower signal before the preamp. Better isolation options include circulators and the appropriately-named isolators.

Circulators are three-port devices which transmit signal from ports 1 to 2, 2 to 3, and 3 to 1 with minimal loss. Crucially, they strongly attenuate signals going in the opposite direction (1 to 3, 3 to 2, and 2 to 1). Therefore, a device routed to a circulator's port 1 will transmit its signal out port 2, but any reflected signal returned to port 2 would not reach the MKIDs on port 1. Port 3 should be terminated to dissipate the reflected signal. Isolators are circulators with their third port terminated. As an example, a Low Noise Factory isolator (LNF-ISC4_8A) can produce 22dB of isolation with only 0.17dB of forward attenuation over a large range of frequencies (4-8GHz). One benefit of using a circulator rather than a packaged isolator is the

ability to route port 3 to a terminator on a warmer thermal stage. Reflected power can then be dissipated without the additional thermal load affecting the payload temperature.

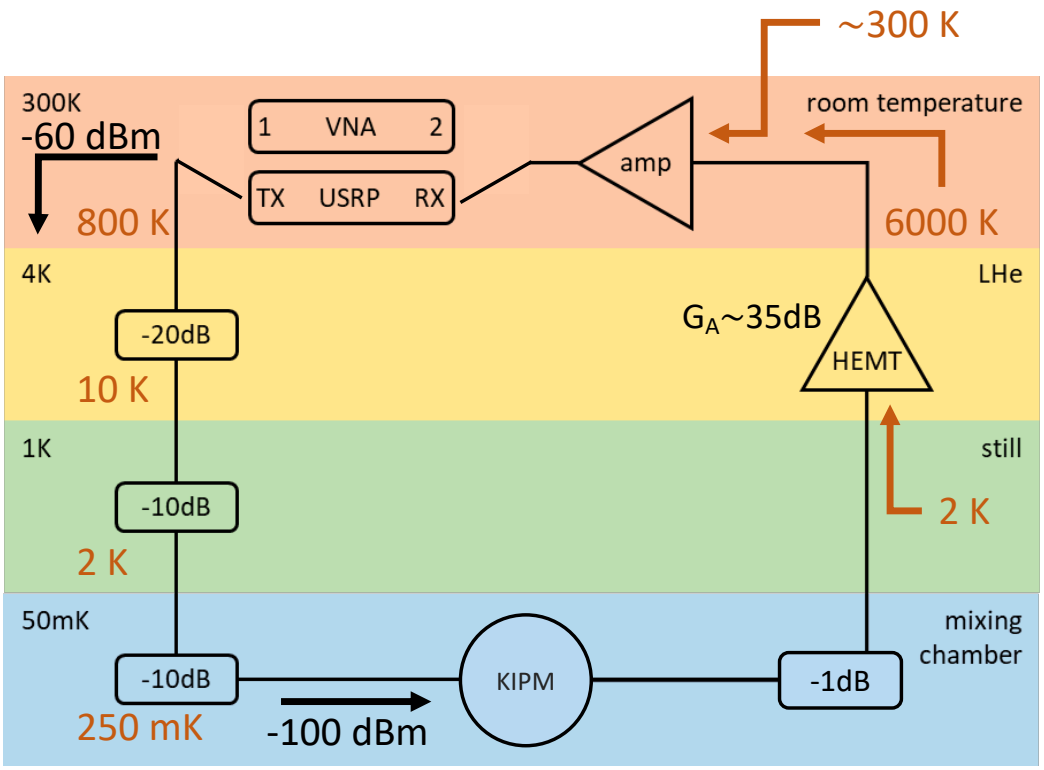


Figure 4.18: The readout chain used for most of the KIPM data-taking described herein. Adapted from a diagram by Osmond Wen. An example readout power (before and after attenuation) is shown in black. 800 K is used as an example readout noise going into the cryostat. The cold attenuators reduce this noise to 250 mK at the device. A 1 dB attenuator was used for isolation between the KIPM and preamp. As preamp, we used a few different cryogenic HEMTs (see Table 2.2 of [66]). The HEMTs had input noise between 2 and 5 K . In this example, the readout noise into the cryostat is dominated by even the lowest noise HEMT (2 K $\gg$ 250 mK). The HEMT gain is also sufficient to make the warm-amp noise insignificant (6000 K $\gg$ 300 K). The example combination of readout power and noise was calculated using a spreadsheet by Alvaro Loya Villalpando and is reasonable for the USRP readout (described in Section 4.4.2). In practice, the fridge coax cables also have some attenuation (particularly the ~1 m coaxes going from room temperature to 4 K). We did not account for coax attenuation here.

Another consideration is the magnetic field produced by circulators (and isolators) which often contain magnetized ferrite. It has been confirmed that magnetic flux present when an MKID transitions to its superconducting state can leave vortices

within the metal that degrade quality factor [76]. Some circulators, including the LNF-ISC4_8A, can be purchased off-the-shelf with magnetic shielded housings. Alternatively, one could magnetically shield the MKID-device itself to protect it from the circulator and other magnetic-field sources.

A non-magnetic option for isolation is a directional coupler. These three-port devices split incoming signal across two output ports (a "coupled" port that siphons off a fraction of the signal and a "transmitted" port that receives the rest). In the case of a power divider, this is done to split a signal into two equal (but 3 dB weaker) signals. Other directional couplers only siphon off a small fraction of the total signal for use elsewhere. For isolation, the important trait of a directional coupler is that most signal entering the transmitted port is redirected and dissipated in an internally terminated port 4.

The readout design used to take most of the KIPM data in this thesis is shown in Figure 4.18. Ideally, a readout should be tailored to specific power and noise requirements. We needed to vary readout power in order to test its effect on detector performance and to operate different MKIDs. Specifically, due to variation in quality factors and feedline transmission, one value of readout power is not optimal for all MKIDs. Changing readout power also changes the transmission-side readout noise (equivalent to adding or removing amplifiers or attenuators). Therefore, we designed our readout for flexibility. The cryogenic attenuation was chosen so noise would be dominated by the HEMT amplifier under typical operating conditions. The cryogenic attenuation was kept small enough that we could still deliver high powers to the KIPM (without excessive input amplification). We used a 1 dB attenuator for preamp isolation since it was the simplest option at the time.

4.4.2 USRP-GPU Readout System

Previous KIPM-detector work at Caltech was performed using an FPGA-based system known as the CASPER (Collaboration for Astronomy Signal Processing and Electronics Research) ROACH (Reconfigurable Open Architecture Computing Hardware) [68]. Since that time, there has been a loss of institutional knowledge on operating the ROACH within the KIPM group at Caltech. Specifically, the considerable expertise required to make changes in the ROACH firmware is no longer present. Instead of relearning the ROACH system, we decided to use a newer system that would be more flexible to different experimental designs and approachable to non-experts. Specifically, we made use of the multi-tone readout system described in [77].

The new system uses the Ettus Research X300 USRP (an off-the-shelf software-defined radio) to transmit and receive high-frequency analog signals used for RF readouts. The USRP itself uses digital-to-analog converters (DACs) to produce the in-phase (I) and quadrature-phase (Q) components of the transmitted signals. The I and Q values are up-mixed, filtered, and amplified through "daughterboards" installed in the USRP. The daughterboards are also produced and sold off-the-shelf by Ettus. The choice of daughterboard determines the readout frequencies and bandwidths available. For our readout, we predominately used a UBX daughterboard capable of 10-6000 MHz readout with 160 MHz bandwidth. When receiving signals, the daughterboards amplify, filter, and down-mix to produce the I and Q values recorded by the USRP analog-to-digital converters (ADCs). It is worth noting that the UBX daughterboard generates separate LOs (local oscillators) for up and down mixing. There is evidence to suggest the difference can cause additional phase noise in UBX-based measurements.

The USRP is connected to a desktop computer using a high-speed (10 Gigabit/s) Ethernet cable. The fast connection allows for all data from the ADCs to be transferred to desktop RAM. The data is then moved again to RAM on the desktop's GPU. The GPU is used to perform the most intensive processing tasks. Specifically, the GPU is used for the final demodulation that takes the I and Q values (already down mixed with the LO) and calculates individual values for each signal frequency of interest. This can be done using direct demodulation (multiplicatively down mixing each frequency individually) or using a polyphase filter bank (see [77]). We primarily use the direct-demodulation method to preserve greater signal bandwidth. The polyphase-filter-bank method is useful for demodulating larger numbers of tones

simultaneously. The ability to customize the demodulation method is another benefit of the USRP over the ROACH (which demodulated using the FPGA firmware rather than a GPU). The GPU also decimates the data for storage on desktop memory. For signal generation, the GPU calculates the transmission buffers sent to the USRP DACs to produce the desired analog signals.

In practice, the USRP system is operated using one desktop (with a powerful GPU) as the server. The server directly communicates with both the USRP and the GPU. A client can then be operated on another computer (local or remote) to control the system and receive the GPU-processed data using transmission control protocol (TCP). The client and server can be operated on the same computer.

Like the ROACH, the USRP is FPGA-based. However, for our purposes, the FPGA portion of the USRP never requires reprogramming. Instead, the readout is customized by changing daughterboards or altering code. The GPU code is written with CUDA and C++, both of which have extensive online documentation. The majority of the server and client code is written with C++. Lorenzo Minutolo has also written comprehensive Python packages for controlling the system through the client (documented online[7]). Python is a widely known and well-documented programming language. These packages make the system very approachable to beginners.

We primarily used the USRP system to to take continuous time-stream measurements of multiple on- and off-resonance frequencies simultaneously. We also used it to take VNA measurements without needing to switch between the USRP and a traditional VNA. Using the Python library, we wrote code to use the USRP as a spectrum analyzer to measure readout noise and test for HEMT-amplifier domination (see Section 4.5.2).

4.5 Resolution Measurements

After characterizing the RF behavior of MKIDs on the 80-MKID KIPM, we proceeded to measure the amplifier-dominated resolution on event-energy deposited in each MKID. As seen in Chapters 2 and 3, energy resolution is a key figure of merit in dark-matter search experiments. It was important to explicitly measure the MKID resolution in order to validate and compare with Eq. 4.55 which relies on various assumptions and does not account for the observed resonator asymmetry (discussed in Section 4.3.3).

[7] `https://www.its.caltech.edu/~minutolo/gpu_sdr_doc.html`.

To achieve the best possible resolution in our MKIDs, we use the Optimal-Filtering method previously discussed in Sections 3.5.2 and 4.2.1. Producing the filter requires both a signal template (describing the anticipated shape of a noiseless single-event signal) and a noise measurement. The filter can then be calculated using Eq. 4.43. To calculate the baseline resolution, one can apply the filter to noise data and measure the Gaussian width of the resultant distribution of zero-signal events. Alternatively, the resolution can be directly calculated using the template and noise with Eq. 4.44. We performed both methods.

4.5.1 Template Generation

Calculating a signal template required producing event-like signals and averaging to remove noise. Averaging can be improved if events are generated in large numbers. If events are produced with predictable timing, they can be accurately aligned during averaging to reduce timing jitter. It is also beneficial to produce events with approximately constant signal size comparable to the signal sizes expected during a search. We satisfied most of these conditions using phonons produced by recombination of readout-generated quasiparticles. Specifically, we used an arbitrary waveform generator (AWG) to input high power at the resonant frequency of one MKID in order to break a large number of Cooper pairs in that MKID's inductor. The readout-generated quasiparticles would then recombine and emit some fraction of their energy as above-gap phonons. Some phonons could then transfer from the pulsed MKID to the substrate. Above-gap substrate phonons could then be detected by MKIDs via the usual mechanism. Since we only needed to measure the signal shape, it was not necessary to know the efficiency of energy transfer between the various quasiparticle and phonon systems. We only needed to raise the AWG power until response could be seen in the non-pulsed MKIDs. The high-power AWG signal was only input briefly (in 10-20 μs pulses) to emulate signal from a delta-function energy deposition. The USRP was used to record transmission as a function of time at MKID resonant frequencies before and after each pulse. We compared the transmission response on- and off-resonance to confirm that the observed signals were not electrical artifacts caused by the sharp change in readout power. This method allowed us to produce many events with extremely predictable timing and approximately constant amplitude. The amplitude could also be adjusted by changing the AWG readout power and pulse duration.

Figure 4.20 shows the response of two MKIDs to one MKID being pulsed at its resonant frequency. Each response is shown via the change in complex transmission

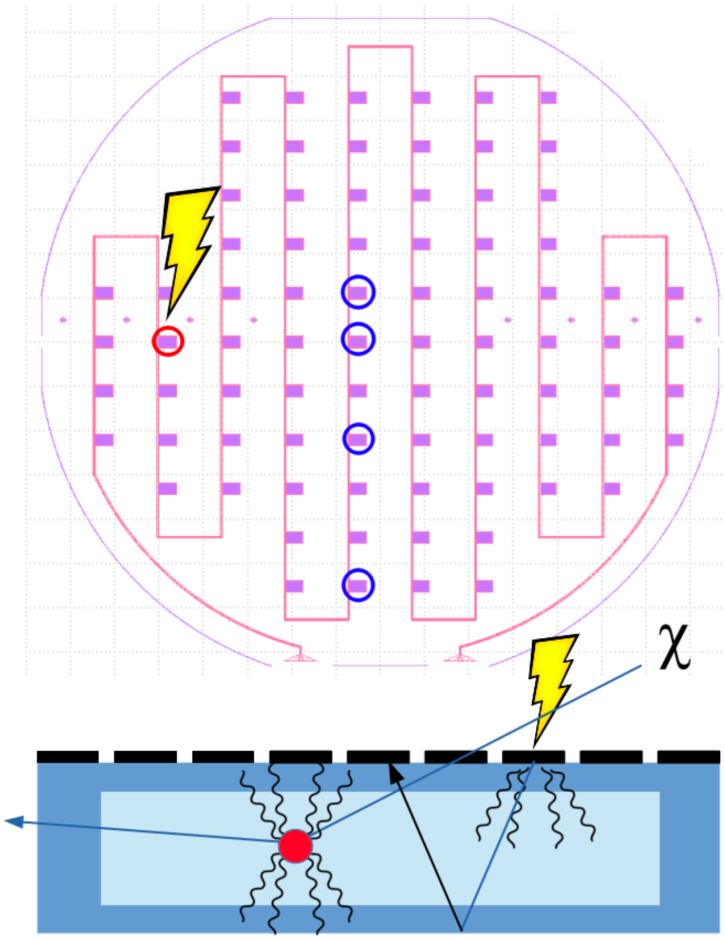


Figure 4.19: A diagram of the RF-pulsing technique borrowed from [69]. High-readout power in the "lightning"-targeted MKID breaks quasiparticles. Quasiparticle recombination produces phonons which can enter the substrate and be observed in other MKIDs. A dark-matter event that can similarly produce phonons is displayed for comparison. The MKIDs circled in this diagram do not correspond to the MKIDs used in Figure 4.20 and below.

measured at the initial resonant frequency. Responses were averaged in 30 ms intervals aligned to the beginning of the high-power pulse at 2 ms.

In order to calculate energy resolution, we used Eq. 4.41 to convert pulse responses to units of δn_{qp} (proportional to energy). First, it was necessary to convert the observed change in transmission to the "ideal resonator" basis. This is equivalent to translating from Eq. 4.58 (which describes the observed transmission) to Eq. 4.9 (which describes an ideal resonance for which Eq. 4.41 applies). Eq. 4.71 can be used to translate between the two bases.

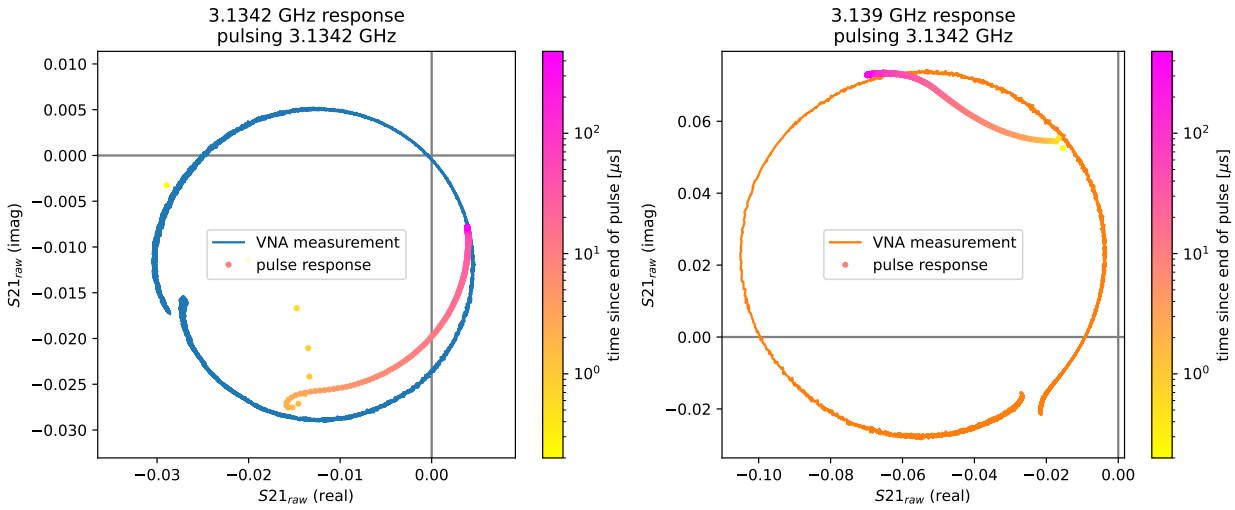


Figure 4.20: The complex-space response of two MKIDs to one MKID being pulsed at its resonant frequency. The 3.1342 GHz MKID was pulsed with high power for 20 µs. Response was calculated by averaging ~700 individual responses to the pulsing. The plotted response begins as soon as the pulsing ends. The complex transmission around each resonance (the result of a VNA-style measurement) is shown for reference. (Left) The pulsed MKID and its complex response. (Right) A neighboring MKID and its complex response.

$$S_{21,\,\text{ideal}} = 1 - \left[1 - \frac{S_{21,\,\text{observed}}}{ae^{-2\pi j(f-f_r)\tau}}\right]\frac{\cos(\phi)}{e^{j\phi}} \tag{4.71}$$

Pulse response was measured using the USRP's continuous data-taking mode. In this mode, the USRP records transmission as a function of time for one or more constant frequencies. During the measurement, it was observed that transmission differed noticeably between continuous and VNA-style USRP measurements. The source of the discrepancy remains unknown. It presented an issue when converting to the ideal-resonator basis, because Eq. 4.71 uses parameters extracted from the VNA-style measurements (including the overall gain and phase). To account for the difference, we applied a translation to each MKID's continuous measurements to make the average resonant-frequency transmission equal that of the VNA-style measurement. We also took continuous data slightly above and below each resonant frequency concurrent with the resonant-frequency data. We used this data to calculate the complex-space rotation (about the resonant-frequency transmission) that best aligns the continuous dS_{21}/df to the VNA-style dS_{21}/df. For each MKID, we applied the translation and rotation to all continuous measurements before con-

verting to the ideal-resonance basis. An example correction can be seen in Figure 4.21.

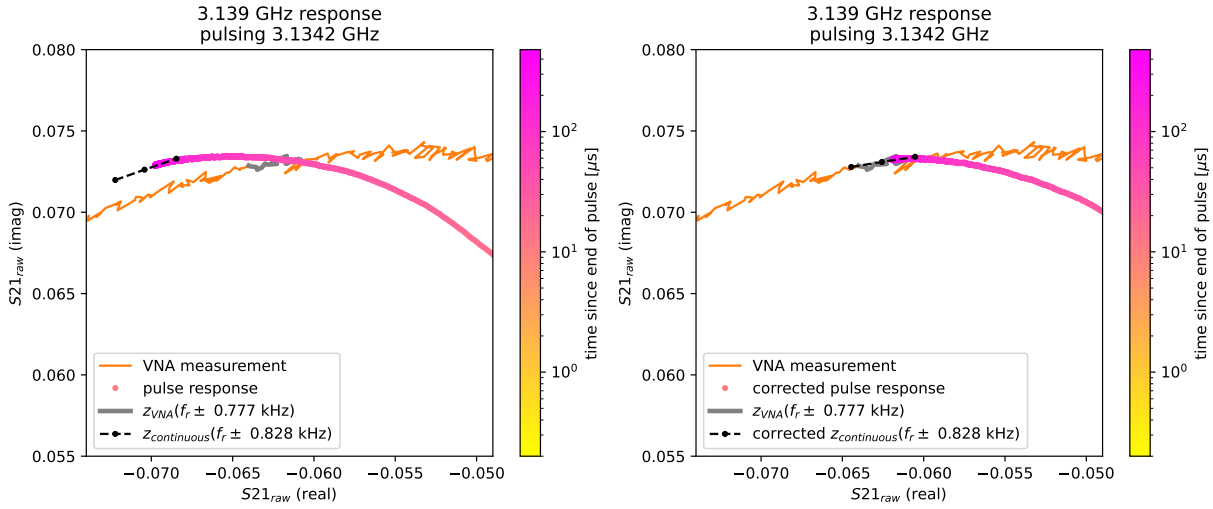


Figure 4.21: An example application of the continuous to VNA-style data correction. (Left) The complex-space response of an MKID to its neighbor being pulsed at its resonant frequency (a zoom in of the right side of Figure 4.20). Continuous and VNA-style measurements of transmission near resonance are shown in black and grey, respectively. We see that the continuous measurements disagree with the VNA-style measurements of the same transmissions. (Right) The same result with the continuous measurements translated and rotated to match the VNA-style measurements.

The corrected data was then converted to the ideal-resonator basis. Results for the data in Figure 4.20 are shown in Figure 4.22. We see that the conversion does not perfectly place the resonant-frequency transmission on the real axis (as implied by evaluating Eq. 4.9 for $f = f_r$). This reflects the imperfect fit of Eq. 4.58 to data. The difference appears large on the right side of Figure 4.22 but is actually equivalent to only a very small difference in frequency ($\sim$10 kHz). We accounted for this type of misalignment by applying an additional rotation to the pulse data when converting to units of δn_{qp}. The rotation was equivalent to changing the tangent of the resonance loop near f_r to be vertical.

We could then use Eq. 4.41 to convert the observed signals in ideal-resonator space to units of n_{qp}. Specifically, we treated the real and imaginary components of signal separately as shown in Eq. 4.72. This produces two separate measurements of δn_{qp} that should in theory be identical. The real signal component reflects the change in resonance-circle radius which is determined by quality factor. The real component

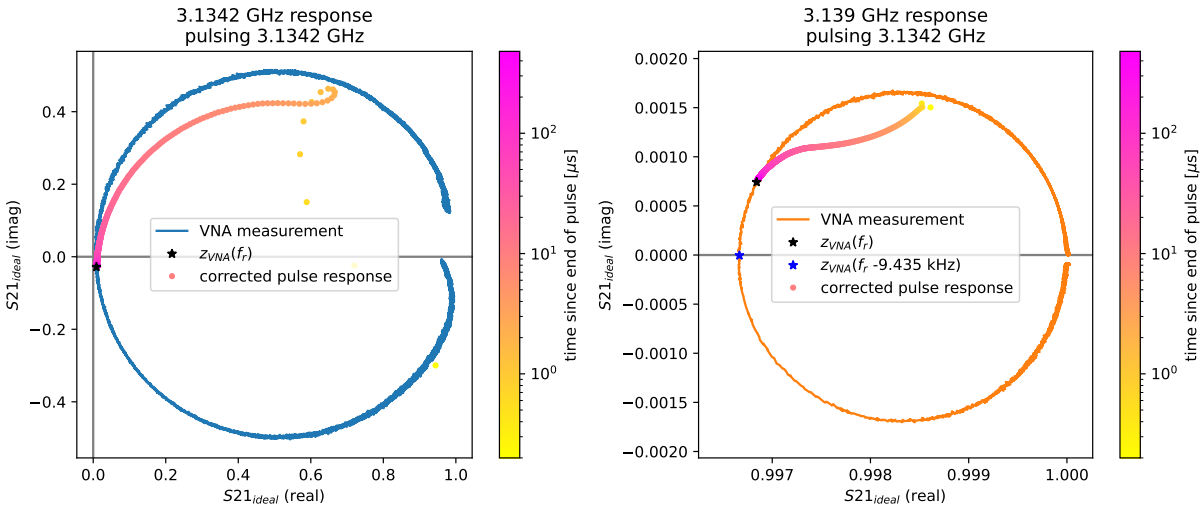


Figure 4.22: The results from Figure 4.20 after correcting for the difference between continuous and VNA-style measurements and converting to ideal-resonator space (using Eq. 4.71). Transmission at the resonant frequency is marked with a black star. (Left) The pulsed MKID and its response. (Right) A neighboring MKID and its response. The VNA-style frequency giving real ideal-space transmission is marked with a blue star. This is where the resonant-frequency transmission would be if the fit was perfect.

is therefore referred to as the dissipation signal. The complex component results from the changing resonant frequency and is therefore referred to as the frequency or phase signal.

$$\begin{aligned} \kappa\delta n_{qp} &= \frac{Q_c}{\alpha|\gamma|Q_r^2}\delta S_{21} \\ (\kappa_1 + j\kappa_2)\delta n_{qp} &= \frac{Q_c}{\alpha|\gamma|Q_r^2}\left(Re\{\delta S_{21}\} + jIm\{\delta S_{21}\}\right) \\ \rightarrow \delta n_{qp} &= \frac{Q_c}{\alpha|\gamma|Q_r^2\kappa_1}Re(\delta S_{21}) \equiv \delta n_{qp,\,\mathrm{diss}} \\ \rightarrow \delta n_{qp} &= \frac{Q_c}{\alpha|\gamma|Q_r^2\kappa_2}Im(\delta S_{21}) \equiv \delta n_{qp,\,\mathrm{freq}} \end{aligned} \tag{4.72}$$

In practice, κ is calculated in its approximate form and used as a constant with respect to n_{qp}. This produces a κ that should be in phase with small changes in n_{qp} but not larger signals. The changing phase for large signals can be seen in the curving δS_{21}s of Figure 4.22. Therefore, when calculating signal templates, we only considered signals after δS_{21} dropped below a small-signal threshold. The threshold

was defined to include δS_{21} for which the angle from the center of the resonance circle changed by less than 20% of θ_κ (defined in Eq. 4.73).

$$\theta_\kappa = \frac{\pi}{2} - \arctan\left(\frac{\kappa_2}{\kappa_1}\right) \tag{4.73}$$

A fraction of θ_κ was chosen to relate the allowed signal size to a physical parameter (in this case, the ideal signal deviation from a tangent). We see the effect of this selection in Figure 4.23. The signal-size limit appears reasonably small, but further work is required to truly understand when the conversion to δn_{qp} is and is not accurate.

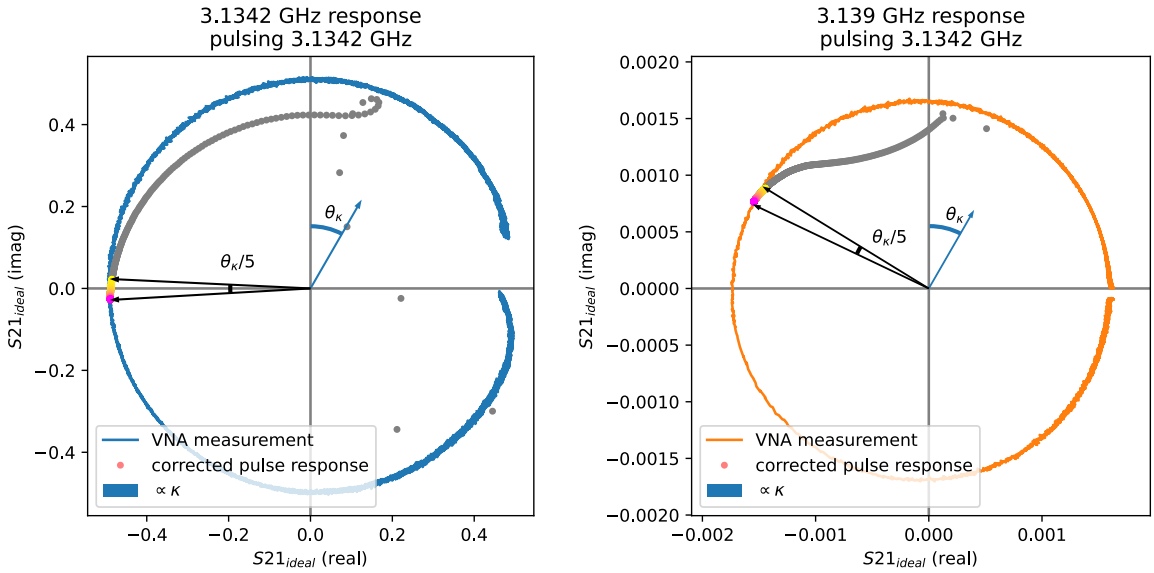


Figure 4.23: The small-signal selection applied to the post-correction data from Figure 4.22. A vector proportional to κ ($= \kappa_1 + j\kappa_2$) is included to visualize the calculation of θ_κ. Data failing the selection is left in grey. (Left) The pulsed MKID and its response. (Right) A neighboring MKID and its response.

The example pulse responses converted to δn_{qp} are shown in Figure 4.24. We see that the frequency and dissipation results do not match perfectly even for the small-signal data.

The pulsing method's ability to produce very large signals (relative to the small-signal threshold) could eventually be used to extract physical parameters describing the δn_{qp} response shape. The expected shapes for small and large signals are derived in Section 4.4.1 of [69]. Further comment on this result and the extraction of parameters are included in Appendix D.

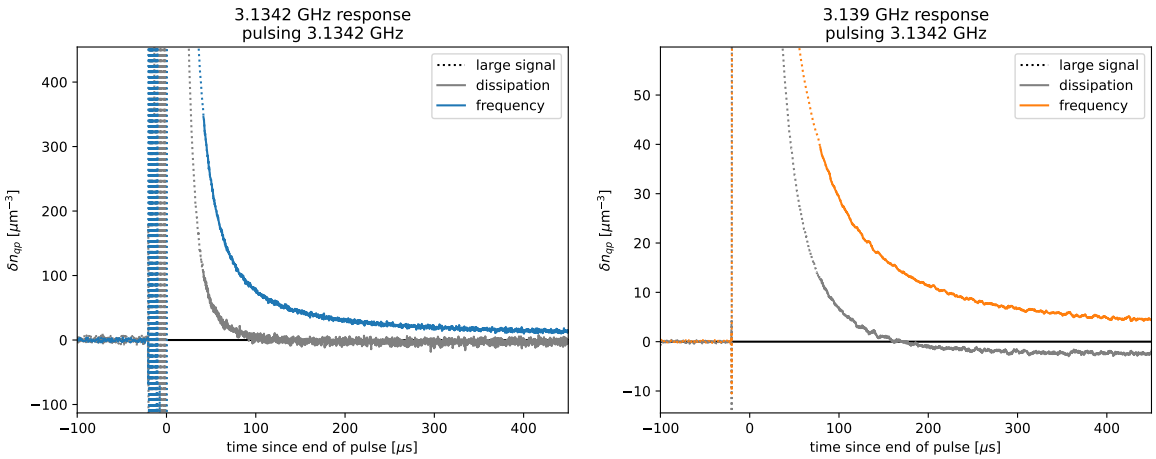


Figure 4.24: The quasiparticle-density (n_{qp}) response of two MKIDs to one MKID being pulsed at its resonant frequency. δn_{qp} is calculated separately using the frequency and dissipation response of the resonator (see Eq. 4.72). The two calculations do not agree precisely. Times with sufficiently small signal are shown with solid lines. Times with larger signal are shown with dotted lines. (Left) The pulsed MKID and its response. (Right) A neighboring MKID and its response.

When calculating resolution, the templates are peak-normalized to unity. To remove high-frequency noise remaining after averaging, we often use fits to the templates instead of the templates themselves. Using a fit which differs systematically from the template might affect energy calibration and resolution at high energy, but will have negligible effect near threshold (where noise is large compared to the difference between template and fit). The energy where the effect becomes significant (to 1σ) is the energy at which the mean of the reduced χ^2 distribution deviates by one from its zero-energy value. While it is beneficial for the fit to accurately describe the template, it is not necessary to use parameters with obvious physical meaning. For example, we sometimes use high-order polynomial fits. In the following sections, I use fits to Eq. D.3 (discussed in Appendix D) as the optimal-filter templates. Example normalized templates and fits can be seen in Figure 4.25.

4.5.2 Noise Measurements

When using the pulsing method to measure templates, we increased the pre-device warm amplification to achieve the high-power required for the pulse. We also decreased the post-device warm amplification to prevent the high-power pulse from damaging the USRP readout. The resulting readout chain was not first-stage amplifier dominated. Instead, the noise was likely dominated by the USRP receive-side

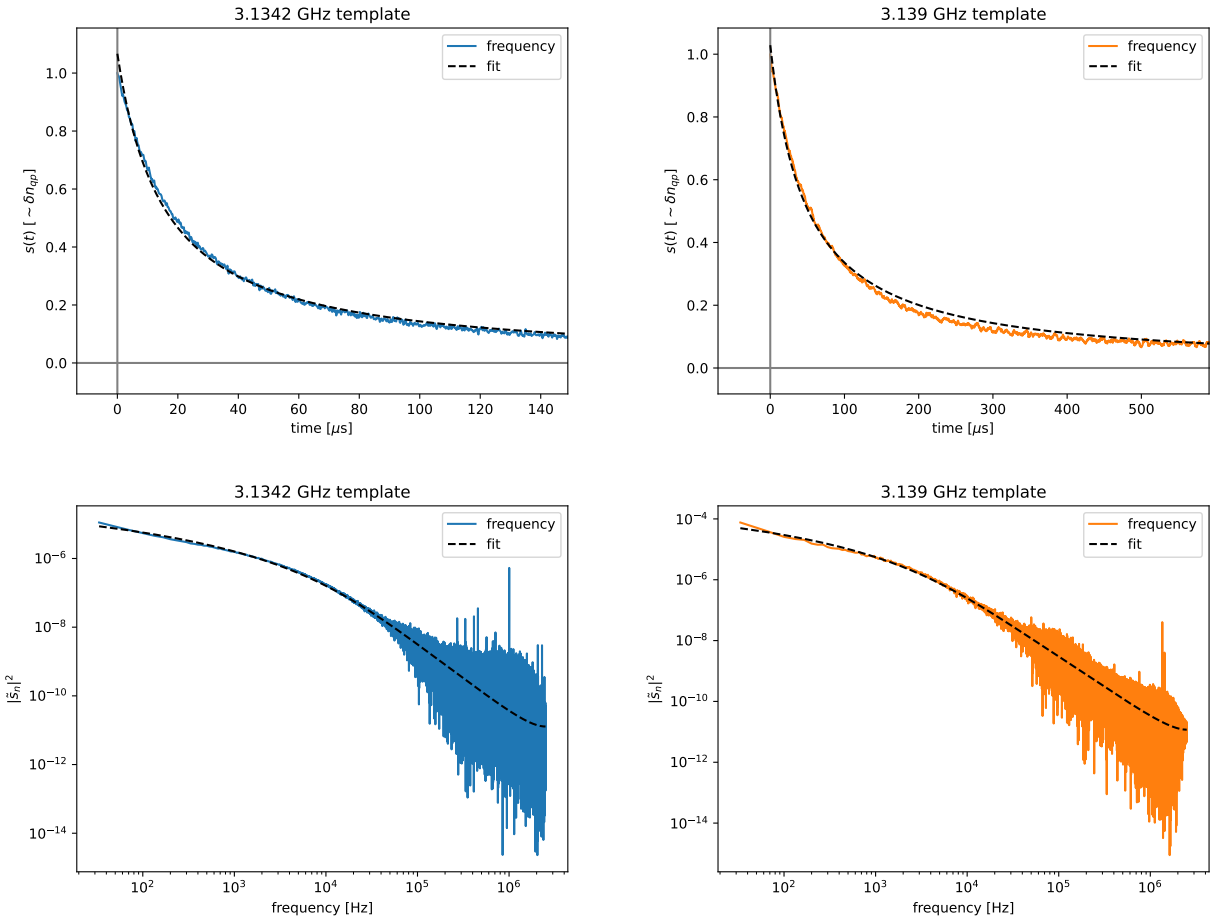


Figure 4.25: Normalized templates with fits resulting from the signals in Figure 4.24. Fits were made to Eq. D.3. (Left) The pulsed MKID and its template. (Right) A neighboring MKID and its template. (Top) The templates in the time domain. In addition to being normalized, templates are aligned to begin at $t = 0$. (Bottom) The unitless squared discrete-Fourier transform of the normalized templates. This value is used in calculating the resolution (see Eq. 4.75).

noise or the AWG noise into the cryostat. This was not an issue for template measurements, because we could measure and average as many events as necessary to recover the underlying signal shape. For noise measurements, we want to measure the best noise achievable during science data taking. A first-stage amplifier is required to elevate signal above the USRP receive-side noise, so first-stage-amplifier noise is unavoidable. Minimizing the total noise is therefore equivalent to choosing the lowest noise first-stage amplifier available and designing the readout chain to ensure that all other noise sources are subdominant to that amplifier's noise (using techniques discussed in Section 4.4.1). It is worth noting that device noise is also

unavoidable and can be difficult to distinguish from first-stage amplifier noise. We will treat the two noises as a combined value when testing for amplifier domination. To date, we expect to be amplifier-noise dominated at most frequencies. There is evidence that TLS noise may be relevant at low frequency for some device and power combinations [78], but not in the 80-MKID KIPM discussed here. The quasiparticle generation-recombination noise (estimated to be 3×10^{-6} $\mu m^{-6}/Hz$ in [69]) is expected to be just below amplifier noise (compare with Figure 4.27).

When testing noise contributions, we used a spectrum-analyzer mode I designed for the USRP. In this mode, the USRP performed a VNA-style measurement of transmission with the USRP-output amplitude set to zero. The result is an uncalibrated measurement of noise as a function of frequency that could be used to measure relative changes in transmission (in units of dB). The USRP output was set to zero by sending zero-amplitude transmission buffers to the USRP DACs. While there was zero signal, the USRP DACs, amplifiers, and variable attenuator were operated with the same settings used during normal data taking. This meant the USRP transmission-side noise was accurate to its value during data taking and could be included in noise-contribution measurements. The spectrum-analyzer code also plotted the absolute value of measured noise (in dB) in real time so the effects of changes to the readout design could be observed immediately.

During the noise-readout design, I defined noise domination as occurring when the dominant source is ≥6 dB larger than all other sources combined. This meant the dominant source had 4× greater noise power than all other sources. If two independent noise sources (with noise powers P_A and P_B in dB) are added together, the resulting noise will be equivalent to P_{A+B} in Eq. 4.74.

$$\begin{aligned} P_{A+B} &= 10\log_{10}\left(10^{\frac{P_A}{10}} + 10^{\frac{P_B}{10}}\right) \\ &= 10\log_{10}\left(10^{\frac{P_A}{10}}\left(1 + 10^{\frac{P_B-P_A}{10}}\right)\right) \\ &= 10\log_{10}\left(10^{\frac{P_A}{10}}\right) + 10\log_{10}\left(1 + 10^{\frac{P_B-P_A}{10}}\right) \\ &= P_A + 10\log_{10}\left(1 + 10^{\frac{P_B-P_A}{10}}\right) \end{aligned} \tag{4.74}$$

We can use this equation to confirm that adding a dominant noise source ($P_B - P_A \geq 6$ dB) produces an increase in combined noise of ≥7 dB. Adding a subdominant noise source ($P_B - P_A \leq -6$ dB) produces an increase in combined noise of <1 dB. This

methodology was used in combination with the spectrum-analyzer USRP mode to confirm amplifier-noise domination.

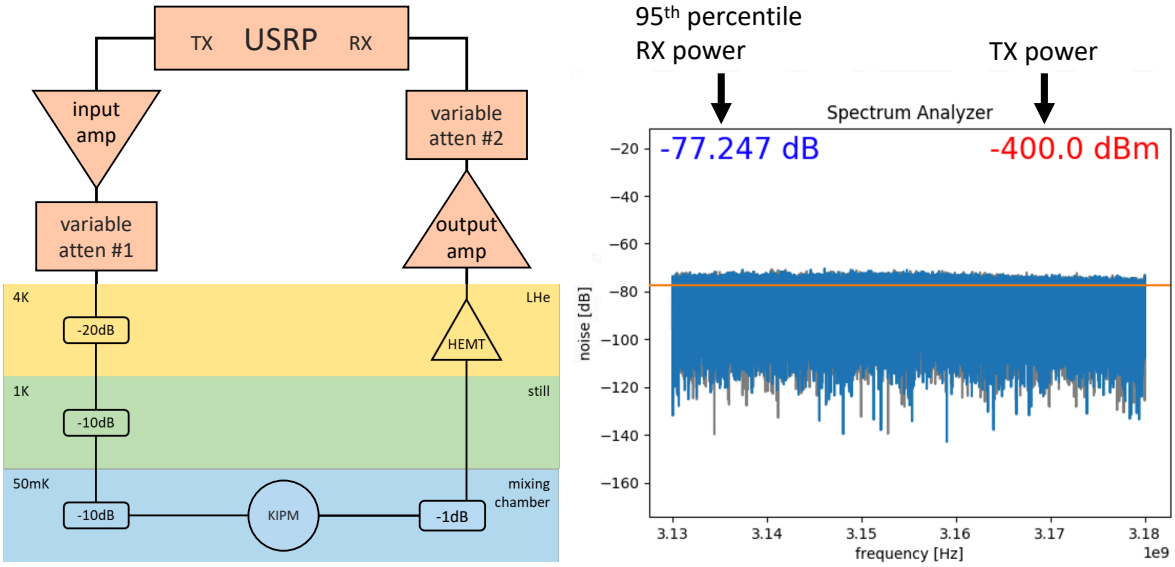


Figure 4.26: Our method for designing a readout with amplifier-dominated noise. (Left) The diagram from Figure 4.18 with further details on the room-temperature readout components (orange). TX (RX) is used to label the transmission (receive) side of the USRP. An input amplifier was used to achieve the desired readout power. Variable attenuators were used to adjust the input and output gains when designing for HEMT amplifier domination. (Right) An example measurement using the USRP to measure uncalibrated noise in spectrum-analyzer mode. The measured noise was summed of over the full measurement bandwidth and therefore did not differentiate between pink and white noise. The upper 95th percentile of measured power was used as the noise-power estimator.

The warm readout was set up as shown in the left side of Figure 4.26. USRP receive-side noise was measured by turning both warm amplifiers and the HEMT amplifier off and setting the variable attenuators to their maximum value (62 dB attenuation). The HEMT and output amplifiers could then be turned on. If the noise increases by 6 dB or more, then the noise is now dominated by the HEMT or output amplifier. This is unlikely to be true with variable attenuator 2 set to 62 dB. In that case, the attenuation should be lowered until the $\geq$6 dB increase is obtained. With amplifier noise dominating, we can toggle power to the HEMT to see which of the two amplifiers is dominant. This is best done with amplifier noise well above USRP noise, so it is a good idea to temporarily set variable attenuator 2 to 0 dB. Then, turning the HEMT off should decrease noise by $\geq$6 dB. If not, the warm output amplifier is dominating. This must be resolved by finding a warm amplifier with

lower input noise or a HEMT amplifier with greater gain. Assuming the HEMT amplifier dominates, the next test is to measure the effect of the USRP transmission-side noise. Variable attenuator 2 should be returned to the previously optimal value. All amplifiers should be powered and the change in noise from the output-and-HEMT-amplifier-only result should be considered. If the noise increased by <1 dB, then the transmission-side noise is subdominant. Variable attenuator 1 should be lowered until the increase is only just <1 dB. This ensures HEMT domination while allowing maximum readout power to the KIPM. If noise increases by ≥1 dB even with large attenuation, then attenuator Johnson noise is preventing amplifier domination. In this case, it may be necessary to redesign the cryogenic readout to include more cold attenuation before the device.

For the YY180726.1 KIPM, we successfully found HEMT-amplifier-dominated noise readout settings using a HEMT amplifier from Table 2.2 of [66] and the cold attenuation shown in Figure 4.26. Specifically, this was done using Mini-Circuits ZVA-183-S+ for both warm amplifiers. Variable attenuator 1 was set to 24 dB. Variable attenuator 2 was set to 20 dB. This design may not work for other KIPMs or at different frequencies (due to changing transmission) or for different settings on the USRP (including readout power). It is therefore preferential to repeat the above procedure and adjust the design whenever performing a new test.

Noise measurements were taken in the form of 2-second time streams recorded using the same USRP settings and frequencies used for the template measurements. The USRP output powers were adjusted to account for the change in transmission-side gain and match the power at the KIPM used during each template measurement. The noise measurements were converted to units of δn_{qp} using the same methods described for template measurements. This included measuring continuous-style data around each resonance and correcting for the difference from VNA-style measurements (like in Figure 4.21). Fits to VNA-style data were then used to convert to the ideal-resonator basis (see Eq. 4.71). Eq. 4.72 was used to convert to δn_{qp}. As with the templates, the converted noise has separate frequency- and dissipation-direction results in δn_{qp}.

4.5.3 Optimal Filter Resolution Calculation

With templates and noise measured for different MKIDs and readout powers, we were able to combine the results to calculate energy resolution. To do so, we used the discrete equivalent to Eq. 4.44. The integral form could not be used without analytic

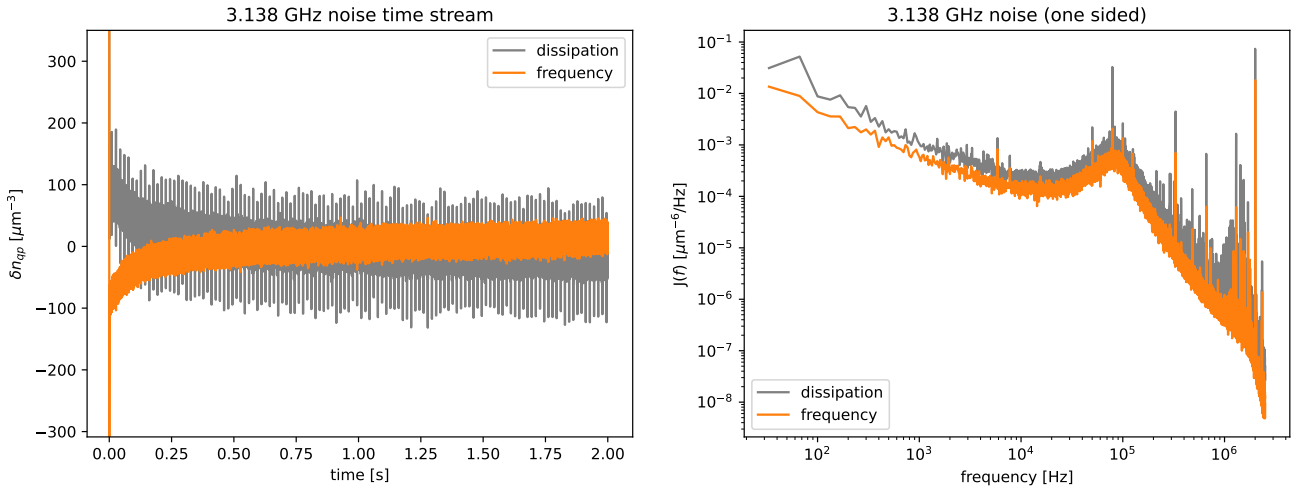


Figure 4.27: Example noise for the MKID on the right of Figure 4.20. (Left) Example time streams in δn_{qp}. The resonant frequency is 1 MHz lower than in previous plots, because the noise data was taken during a different cooldown of the fridge. The data drifts significantly early in the time stream. This is likely a USRP data-taking effect. We ignore the first second of data to account for the effect. We also see a noise feature every ~17 ms (60 Hz). This was likely the 60 Hz noise from outlet AC power exacerbated by a grounding issue. The feature disappeared later, possibly due to a electrical connection being secured. (Right) The average one-sided noise spectral density calculated using the latter half of the time streams from left.

functions for both the template ($s(t)$) and noise ($J(f)$). The discrete resolution is derived in Appendix B of [67].

$$\sigma^2 = \left[T \sum_{n=-N/2}^{N/2-1} \frac{|\tilde{s}_n|^2}{J_n} \right]^{-1} \tag{4.75}$$

In Eq. 4.75, T is the length of the template in time (the sampling time). $\tilde{s}_n$ is the discrete Fourier transform of the normalized signal template. We used Golwala's definition of the discrete Fourier transform (and inverse transform, see Eq. 4.76). In Eq. 4.75 and the Fourier transforms, N is the length of the template (in sample points).

$$\tilde{g}_n = \frac{1}{N}\sum_{k=-N/2}^{N/2-1} g_k e^{-j\omega_n t_k}$$
$$g_k = \sum_{n=-N/2}^{N/2-1} \tilde{g}_n e^{j\omega_n t_k} \tag{4.76}$$

J_n is the average discrete double-sided noise spectral density. J_n was calculated by dividing the noise data (after the initial 1 second of drift) into non-overlapping subsets each of length T. Each subset's mean was subtracted to remove any remaining drift. The subset noise time streams ($v(t)$) were then Fourier transformed, squared, and multiplied by T to calculate individual noise spectral densities (see Eq. 4.77). The subset results were averaged to produce a cleaner J_n for use in the resolution calculation.

$$J_n = T|\tilde{v}_n|^2 \tag{4.77}$$

We chose to limit the sum in Eq. 4.75 to exclude n corresponding to frequencies $\geq$100 kHz. This was done to prevent high-frequency noise remaining in the averaged template from artificially improving the predicted resolution. For the calculations in this section, I am already using a fit to the templates to remove high-frequency noise, so the limit should not have a significant effect. We also excluded $n = 0$, which corresponds to the integrated signal, since we expect the integral to be dominated by unremoved drift rather than individual events.

We used frequency and dissipation templates and noise measurements to calculate separate resolutions for frequency and dissipation signals. The resulting resolutions calculated for 8 MKIDs at various powers can be seen in Figure 4.28. Each of the MKID templates used for this result were measured by pulsing the MKID on the left of Figure 4.20.

The powers in Figure 4.28 were estimated using the recorded readout powers and records of the readout chain utilized. They are labelled as "upper estimates" since there was probably additional attenuation in the refrigerator coaxial cables that we did not measure or account for. The difference in powers should still be precise.

As a diagnostic, we often plot the sum from Eq. 4.75 as a cumulative function of frequency. Such plots visualize which frequencies contribute most to improving the

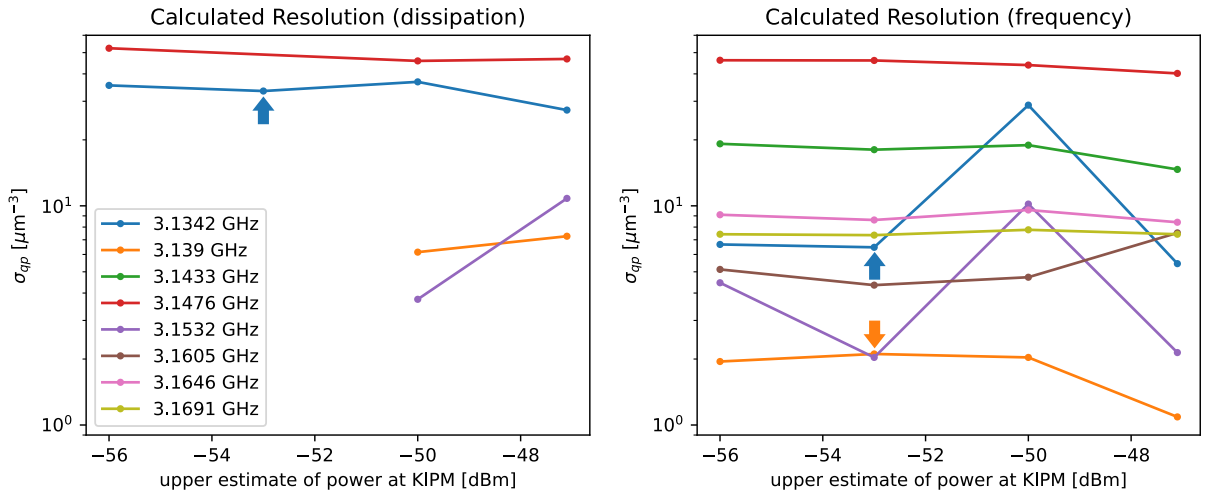


Figure 4.28: Resolutions calculated using Eq. 4.75 in the dissipation (Left) and frequency (Right) directions. Combinations of MKID and readout power that produced poor templates were excluded. The combinations corresponding to the left and right sides of Figure 4.20 are designated with blue and orange arrows, respectively.

resolution. Example normalized cumulative plots can be seen in Figure 4.29. The plots show that our useful signal is mostly between 1 and 10 kHz. This information is useful when deciding readout sample rate, decimation, and filtering.

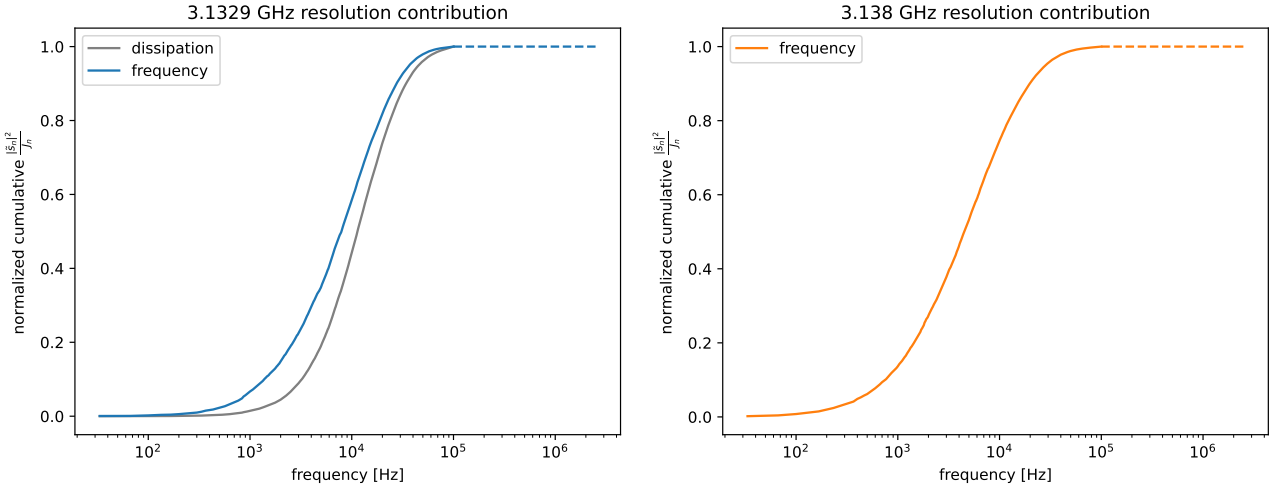


Figure 4.29: The normalized cumulative sums from Eq. 4.75 for calculating the resolutions designated in Figure 4.28. (Left) The MKID and power combination designated in blue. (Right) The MKID and power combination designated in orange.

We also calculated resolution by explicitly producing the optimal filters and applying them to noise time streams. The filter is calculated using the discrete version of Eq. 4.43 (also from [67]).

$$\phi_n = \frac{\frac{\tilde{s}_n^*}{J_n}}{\sum_{m=-N/2}^{N/2-1} \frac{|\tilde{s}_m|^2}{J_m}} \tag{4.78}$$

The filter is applied by convolution with data time streams. Specifically, a time stream ($v(t)$) is Fourier transformed, multiplied by the filter, and inverse-Fourier transformed to produce the filtered result ($v'(t)$). To simplify the final inverse transform, the template and time stream should be equivalent length in time.

$$v'(t) = \frac{\sum_{n=-N/2}^{N/2-1} e^{j\omega_n t} \frac{\tilde{s}_n^* \tilde{v}_n}{J_n}}{\sum_{m=-N/2}^{N/2-1} \frac{|\tilde{s}_m|^2}{J_m}} \tag{4.79}$$

As with the J_n calculation, we divided the noise data into non-overlapping subsets of length T (excluding the initial 1 second of drift). To calculate baseline resolution, we filtered each subset and sampled one data point from each 1 ms of filtered data. As with the Eq. 4.75 calculation, we only considered non-zero frequencies less than 100 kHz. We considered the sample distribution of filtered data from all subsets and fit a Gaussian to extract the resolution. Example results can be seen in Figure 4.30.

Resolutions calculated using the filtering and Gaussian-fit method can be seen in Figure 4.31 (for the same data as in Figure 4.28). As expected, the results are nearly identical to those of Figure 4.28. This serves as some confirmation that both methods are working as expected.

The filtering method can also be used to estimate the resolution on non-zero signals. This was done by adding an artificial pulse to each subset of noise data. The artificial pulses were made by scaling the signal template fit by the desired signal size and adding directly to the subset time streams. Each signal can then be identified as the maximum of the filtered subset data. An example result using artificial signals of $\delta n_{qp} = 500\,\mu\text{m}^{-3}$ can be seen in Figure 4.32.

In the Figure 4.32 example, the resolution slightly degraded from from the baseline resolution ($\sim 2.1\,\mu\text{m}^{-3}$ seen in Figure 4.30). If we fully trusted the parameter extraction described in Appendix D, we could use the parameters to calculate the

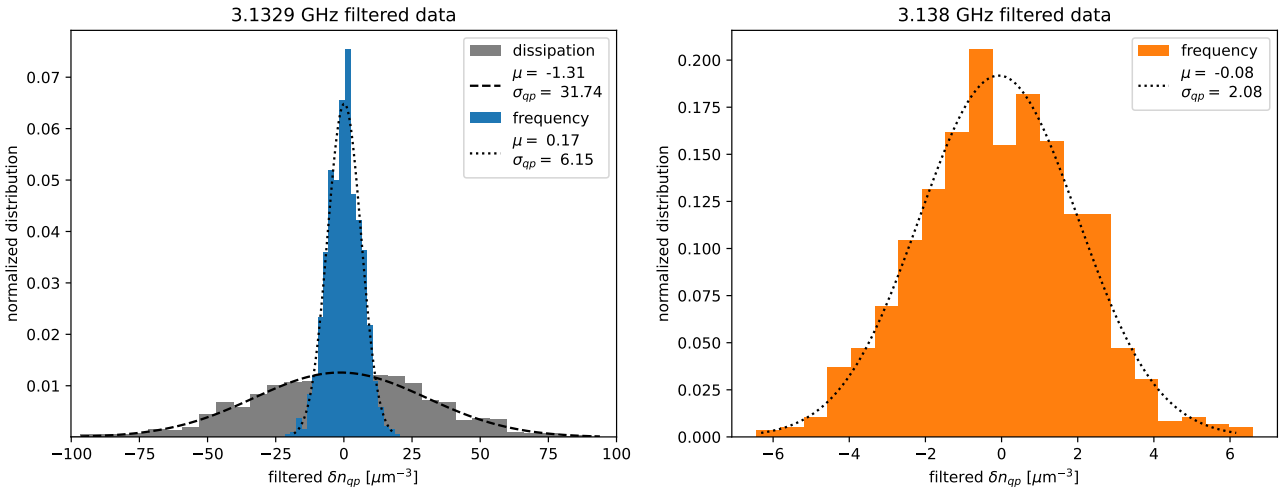


Figure 4.30: The distributions of filtered noise data for the MKID and power combinations designated in Figure 4.28. We see that the standard deviation of each distribution matches the resolution calculated using Eq. 4.75. (Left) The MKID and power combination designated in blue. (Right) The MKID and power combination designated in orange.

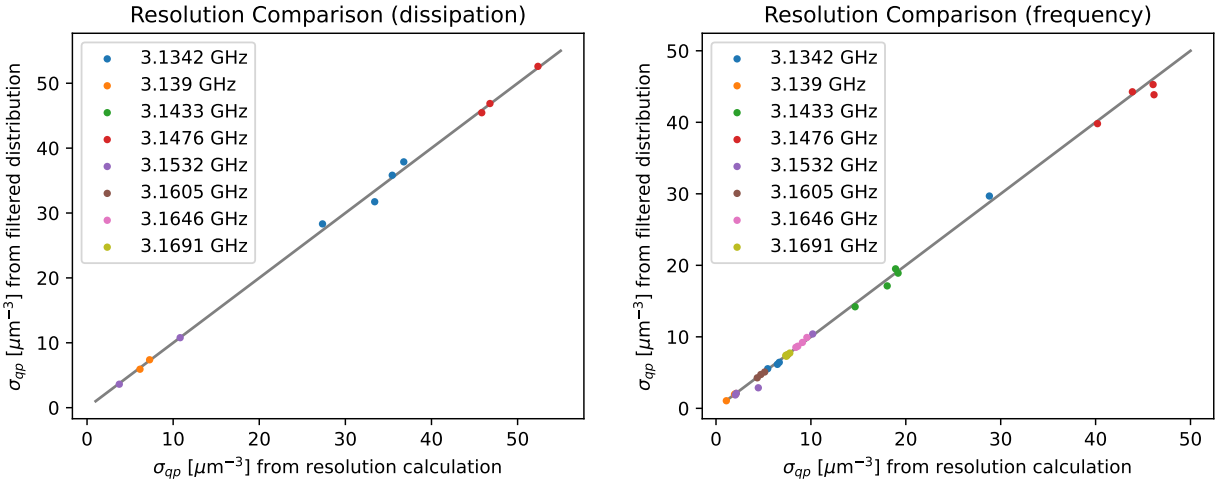


Figure 4.31: Resolutions calculated by filtering the noise time streams and measuring the resulting standard deviation compared to the resolutions in Figure 4.28. (Left) The dissipation-direction resolutions. (Right) The frequency-direction resolutions. Combinations of MKID and readout power that produced poor templates were excluded. A line with unity slope is included for comparison.

signal shape based on the desired signal size. In that case, the changing template shape might make the resolution's dependence on signal size more substantial.

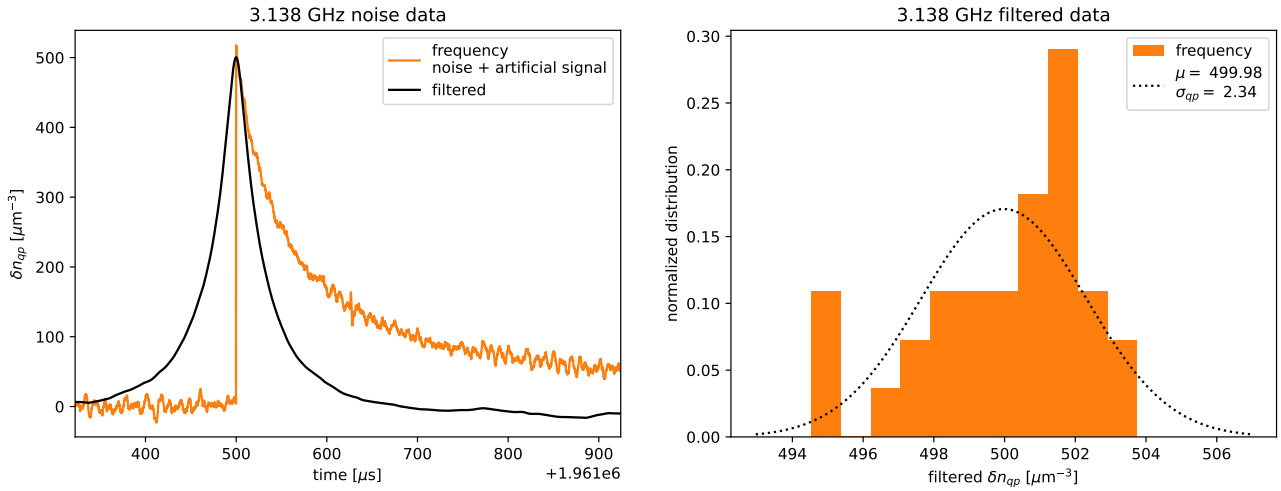


Figure 4.32: Resolution calculation using inserted artificial signals. (Left) An example noise time stream with an artificial $\delta n_{qp} = 500\,\mu\text{m}^{-3}$ signal inserted. The optimally filtered time stream has a clear peak at the location of the signal. The peak amplitude matches the input signal amplitude (an intentional feature of the filter normalization). (Right) The distribution of filtered peak amplitudes for ~30 artificial events like the one shown to the left. The MKID and power combination is the same one designated in orange in Figure 4.28. The predicted $500\,\mu\text{m}^{-3}$ resolution is slightly larger than the baseline resolution shown in that figure.

4.5.4 Correlated-Noise Removal

When calculating MKID quasiparticle resolutions, we observed significant pink noise (noise with power spectral density proportional to $1/f$) in the amplifier-dominated noise data (see the right side of Figure 4.27). Pink noise is typically caused by slow fluctuations in environmental or material conditions reflected in parameters such as amplifier gain or local-oscillator phase. Fluctuations in the readout will affect both the resonant-frequency MKID transmission and the off-resonance feedline transmission. If the on- and off-resonance fluctuations are correlated, simultaneous measurements can be used to reduce the correlated noise in both. The noise removal must be done using off-resonance transmission to ensure that actual signal (only seen on resonance) is not also removed.

Noise is more likely to be correlated between closer readout frequencies. High-Q MKIDs are therefore well suited to removing correlated noise since they require only a small shift in frequency to go from on to off resonance. We confirmed the presence of correlated noise in our readout by measuring coherence between data taken simultaneously on and off resonance. Coherence is defined as the squared

absolute value of cross spectral density between two signals normalized by the power spectral densities of both. A coherence of unity indicates complete correlation and is ideal for noise removal. The coherence measurement for one of our example MKIDs can be seen in Figure 4.33.

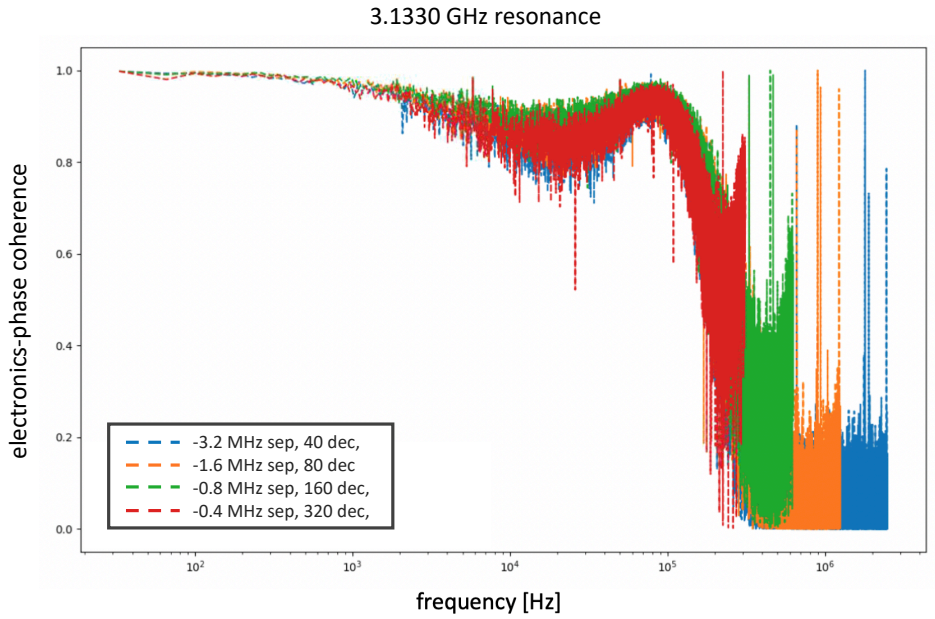


Figure 4.33: The transmission-phase-direction coherence as a function of frequency between one MKID and tracking-tones with various frequency separations. Each separation's coherence appears identical below 100 kHz. The difference in upper limit is caused by a difference in decimation. The amount of decimation was adjusted to prevent beating (between the resonance and tracking tones) from polluting the measurements. This is the same MKID from the left of Figure 4.20.

We measured coherence with off-resonance frequencies (referred to as "tracking tones") 0.4, 0.8, 1.6, and 3.2 MHz away from the resonant frequency. Below 100 kHz, all separations were observed to have significant coherence in the transmission phase and amplitude directions. It is desirable for a tracking-tone to be close enough to share all electronics noise sources with the resonant frequency but distant enough to avoid any quasiparticle signal from the associated MKID. A tracking-tone should also be distant enough that the beat frequency between it and the resonance is above the readout bandwidth (which can be reduced via decimation). Otherwise, the beat will appear like additional noise in the data. For our 80-MKID KIPM (with resonances separated by ~5 MHz), we decided to use tracking tones 2-3 MHz away from each resonance.

Since our correlated noise originates in the readout electronics (not the MKIDs), we performed the removal directly in the measured-transmission basis (rather than the ideal-resonator basis). We took concurrent on-resonance and tracking-tone datasets and separated them into transmission amplitude and phase components. We then decimated all data from a sampling rate of 5 MHz to 10 kHz. The decimation was done to ensure the removal was optimized for our signal region of interest. The removal was done separately for each component following Eq. 4.80.

$$\begin{aligned} v_{\text{clean}}(t) &= v_{\text{on-res}}(t) - A\left(v_{\text{tracking}}(t) - \overline{v_{\text{tracking}}}\right) \\ &\equiv v_{\text{on-res}}(t) - A\Delta v_{\text{tracking}}(t) \end{aligned} \tag{4.80}$$

where A is a constant used to scale the tracking-tone correlated noise to the noise on resonance. We chose our A value to minimize the variance of v_{clean} (Eq. σ^2_{clean}).

$$\begin{aligned} \sigma^2_{\text{clean}} &= \frac{1}{N}\sum_{t_n}\left[v_{\text{clean}}(t_n) - \overline{v_{\text{clean}}(t_n)}\right]^2 \\ &= \frac{1}{N}\sum_{t_n}\left[v_{\text{on-res}}(t_n) - A\Delta v_{\text{tracking}}(t_n) - \overline{v_{\text{on-res}}} - A\cancel{\overline{\Delta v_{\text{tracking}}}}\right]^2 \\ &\equiv \frac{1}{N}\sum_{t_n}\left[\Delta v_{\text{on-res}}(t_n) - A\Delta v_{\text{tracking}}(t_n)\right]^2 \end{aligned} \tag{4.81}$$

The optimal A to minimize Eq. 4.81 can be calculated explicitly (by differentiating with respect to A). The result is Eq. 4.82. When using Eq. 4.82, I chose to only sum times after the first second of data. This was done to avoid optimizing removal of the initial drift. Data after the first second is used in all other aspects of the analysis as well.

$$A = \frac{\sum_{t_n}\Delta v_{\text{on-res}}(t_n)\Delta v_{\text{tracking}}(t_n)}{\sum_{t_n}\left[\Delta v_{\text{tracking}}(t_n)\right]^2} \tag{4.82}$$

After noise removal, the cleaned amplitude and phase were recombined to produce the cleaned transmission. The result of applying the cleaning procedure to the noise in Figure 4.27 can be seen in Figure 4.34. In that example, pink noise was significantly improved but still present below 10 kHz. For other MKIDs, the cleaned noise appeared white (constant with respect to frequency) below 10 kHz. Above ~300 kHz in Figure 4.34, we see that cleaning actually increased the noise slightly.

The noise is uncorrelated at such frequencies (see Figure 4.33). Therefore, the noise subtraction method is equivalent to combining two independent noise sources (increasing total noise). The same procedure described above could be applied to clean science data during an experimental search.

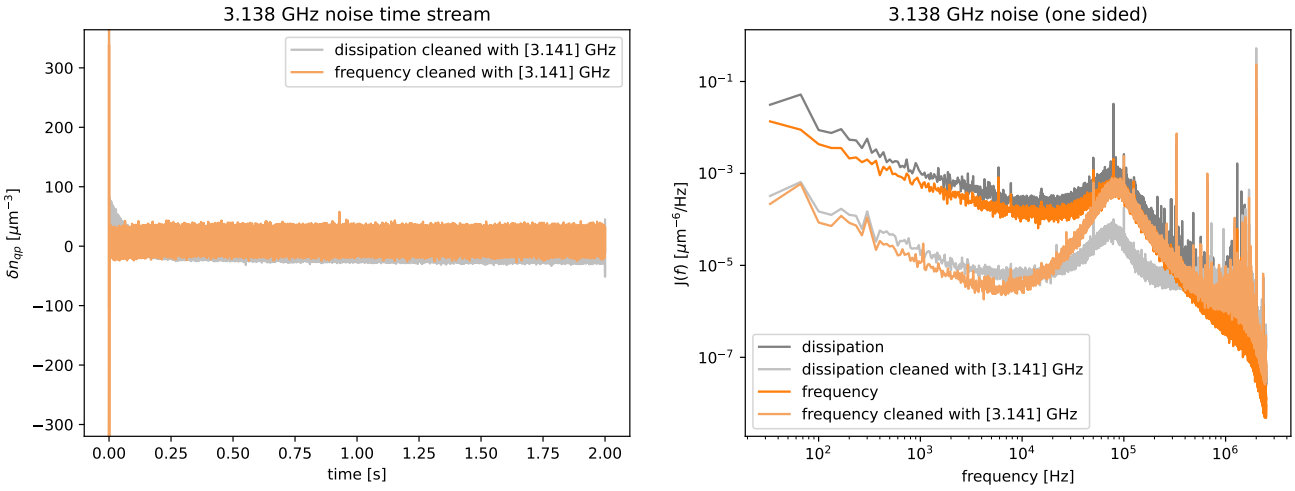


Figure 4.34: The same data and plots from Figure 4.27 but with correlated noise removal applied (referred to as cleaning). (Left) Cleaned time streams (in δn_{qp}). The 60 Hz feature has been reduced to the point of being insignificant in time streams. The small peak around 0.9 seconds is from a background event (possibly a muon). Its presence confirms our noise subtraction is successfully preserving real events. (Right) The average one-sided noise spectral density (before and after cleaning) calculated using the latter half of the time streams from left. We see the $1/f$ noise is significantly reduced but still present. The significant noise bump around 100 kHz also remains.

Cleaned noise can be used to calculate resolution using the same methods described above. The result of calculating resolution using Eq. 4.75 after noise cleaning can be seen in Figure 4.35 (for the same data used in Figure 4.28). Resolutions calculated by explicitly applying the filter to the cleaned time streams produced nearly identical improvements.

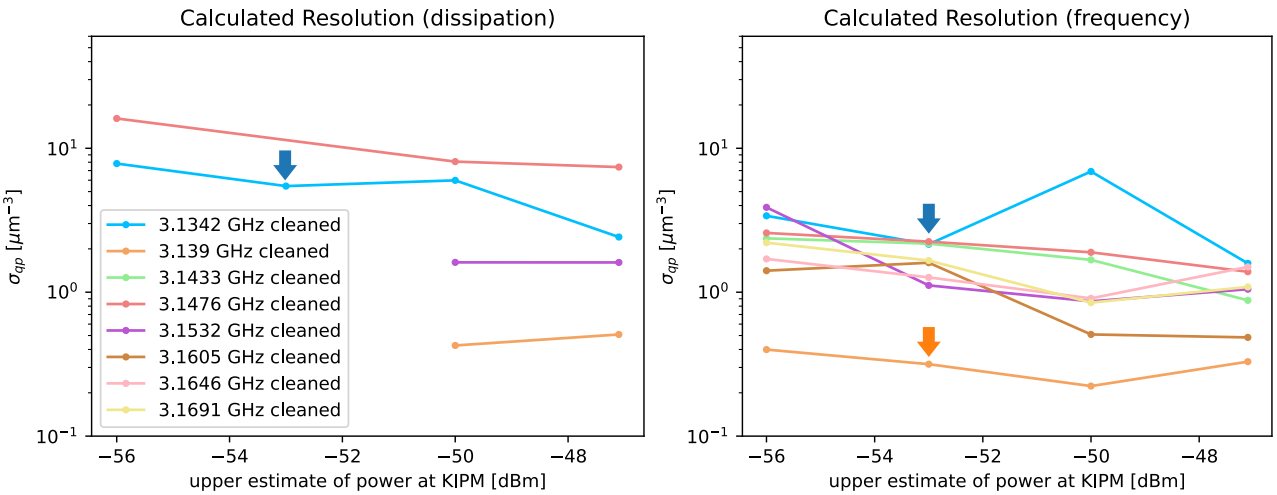


Figure 4.35: Resolutions calculated using Eq. 4.75 in the dissipation (Left) and frequency (Right) directions. Noise for each MKID was cleaned using tracking tones 2-3 MHz off resonance. Combinations of MKID and readout power that produced poor templates were excluded. The combinations corresponding to the left and right sides of Figure 4.20 are designated with blue and orange arrows, respectively.

We see significant improvement in many of the resolutions. Resolutions that were previously the largest seem to receive the greatest improvement. Consider the 3.1476 GHz MKID which improves from ~40 to ~2 μm^{-3} resolution in the frequency direction. Alternatively, the 3.1342 GHz resonance only improves from ~6 to ~3 μm^{-3}. The difference in improvement is likely dependent on MKID parameters including ϕ from Eq. 4.58. ϕ determines the rotation angle between the amplitude and phase noise components and the MKID frequency and dissipation directions.

4.5.5 Conversion to Energy

Up until now, we have reported resolutions in units of quasiparticle density. The conversion to energy within an MKID is simple. We only need to multiply by the sensitive volume of the MKIDs (to convert to quasiparticle number) and the superconducting energy gap (Δ from Eq. 4.41). For sensitive volume, we will use the area of the inductors on the 80-KIPM device. The area is a little different for each MKID, but they are all approximately 1.2 mm^2. With an aluminum thickness of 30 nm, we have an inductor volume of ~ 3.5×10^4 μm^3. For Δ, we use the value extracted via Mattis-Bardeen fitting for each MKID. The resulting Δs range from ~0.185 to ~0.202 meV. The resolutions from Figure 4.35 after conversion can be seen in Figure 4.36.

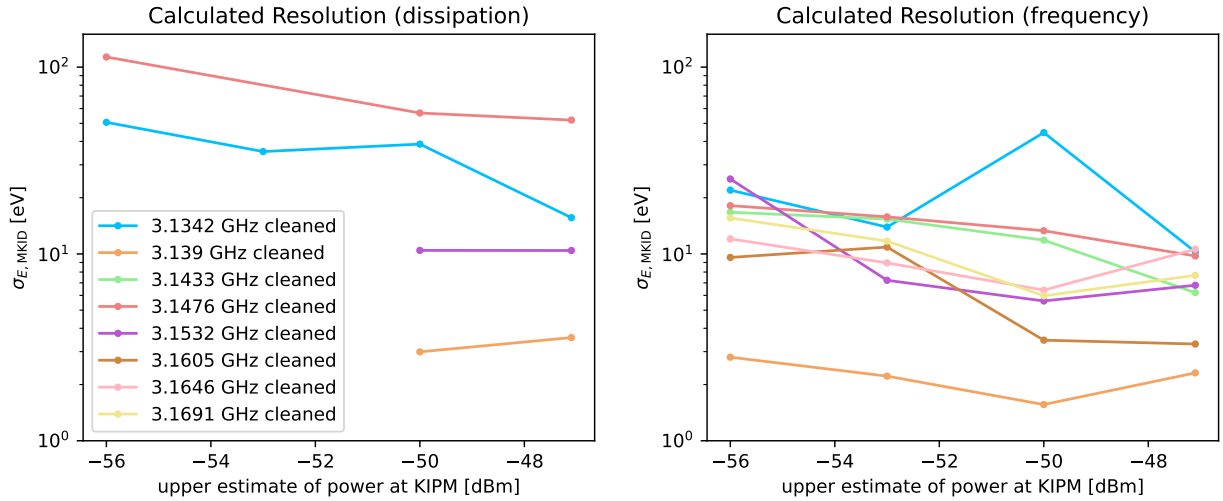


Figure 4.36: Resolutions calculated using Eq. 4.75 in the dissipation (Left) and frequency (Right) directions. The same result from Figure 4.35 converted to units of energy absorbed within the MKID. The best observed resolution was ~1.5 eV in the frequency direction using the 3.1390 GHz MKID.

4.5.6 Resonance-Parameter Dependence

Before and after noise cleaning, we observe significant variation in resolution between different MKIDs. Naturally, we would like to understand what parameters give certain MKIDs better resolution than others. We can use such information to design future MKIDs to achieve superior resolution. The majority of work towards understanding the variation was outlined by Karthik Ramanathan in [79]. A key result of that analysis and subsequent tests was that resolution among MKIDs operated at the same readout power is inversely proportional to the factor R_Q.

$$\sigma_{E,\,\mathrm{MKID}} \propto \frac{1}{R_Q} \tag{4.83}$$

$$R_Q \equiv |a| \frac{Q_r^2}{|\widehat{Q}_c|} \tag{4.84}$$

The parameters in Eq. 4.84 are the same extracted using fits to Eq. 4.58. We can use these parameters to compare our resolution results to the prediction.

In the case of Figure 4.37, the results seem reasonably consistent with the prediction. Other results (using a greater number of MKIDs) have matched the prediction to an even greater degree. Those results will be published in a separate paper soon.

From Eq. 4.83, we see that improving the overall transmission magnitude (encoded in $|a|$) would benefit our resolutions. We expect to do so in the future through better feedline impedance matching. Specifically, we would like to remove the large fluctuations visible in our current device transmissions (see Figure 4.9). With improved transmission and reasonably large Q, we would expect to see ~2 and ~1 eV MKID resolutions in the dissipation and frequency directions, respectively.

4.6 Ongoing and Future Work

Going forward, there are a number of ways in which KIPM detectors are being developed at Caltech and elsewhere. The typical goals are to improve individual MKID energy resolution, measure KIPM energy resolution, and improve KIPM energy resolution. With a number of realistic improvements, we think producing a single-MKID KIPM with sub-eV resolution should be possible in the near future. For multi-MKID KIPMs, the primary goal is to improve the uniformity of transmission so all MKIDs can be used simultaneously with good position resolution (which has already been shown in [68] for smaller substrates in which the feedline issues were less challenging).

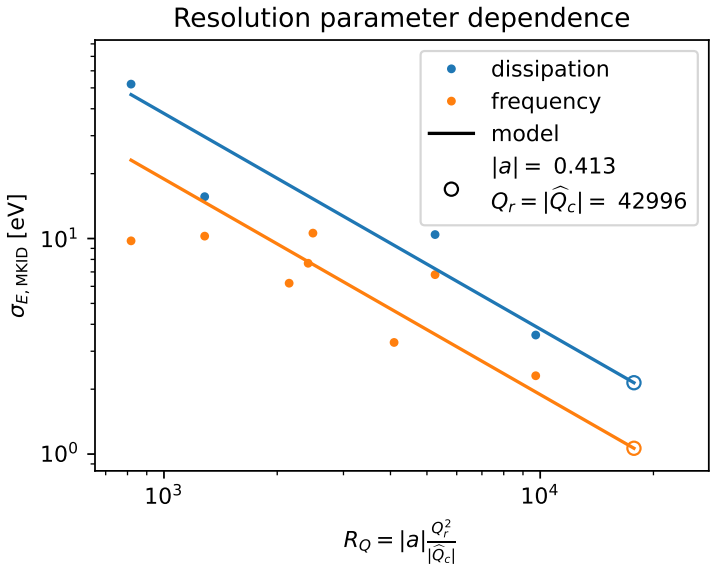


Figure 4.37: Resolutions from the highest-power data in Figure 4.36 compared to the predicted scaling factor (R_Q). As with previous plots, combinations of MKID and readout power that produced poor templates were excluded. The solid lines follow Eq. 4.83 multiplied by a constant to best match all data points. Unfilled circles indicate the model resolutions for optimal R_Q. Optimal R_Q uses the largest $|a|$ from all 8 MKIDs (at this power). In practice, this could be achieved by improving the feedline transmission uniformity. Optimal R_Q also assumes the largest observed Q_c, $Q_i \gg Q_c$, and $\phi = 0$.

At Caltech, we have installed an LED connected to a fiber-optic feed-through leading to the device housing in our fridge. With this arrangement, we can send LED photons with constant individual energy (but varying quantity) to trigger events within our devices. Photon events can be used to produce templates similar to those made with the RF-pulsing method. Additionally, laser photons can be used to perform an absolute energy calibration of our detectors. The calibration is done by producing many photon events using set LED-driver powers. Variation in the observed signals will arise from the system noise (readout and MKID noise) as well as photon shot noise. The photon shot noise comes from the Poisson variation in the quantity of photons per event. When shot noise is raised above system noise, its magnitude can be used to calculate the average number of photons arriving per event. The most accurate result can be found by fitting the combination of shot and system noise to the resolution as a function of event amplitude. The main challenge with this method is lowering the system noise such that the changing shot noise is clearly

distinguishable. Once the photon number is known for a given amplitude event, we can multiply by the individual photon energy to recover the total energy of such events.

Understanding the total event energy also allows us to calculate the phonon-collection efficiency (η_{ph} from Eq. 4.56). We simply calculate the event energy within each MKID (the change in quasiparticle number multiplied by the energy gap), sum over all sensitive MKIDs, and compare with the total event energy. This measurement has already been successfully performed by Dylan Temples at the NEXUS facility using one of our single-MKID devices [80]. The MKID was found to have frequency-direction $\sigma_{E,\,\mathrm{MKID}}$=2.6 eV and $\eta_{ph} \approx 0.8\%$. For a single-MKID device, this corresponds to a KIPM energy resolution of $\sigma_{E,\,\mathrm{KIPM}} \approx 320$ eV. If the MKID resolution was improved to that predicted in Figure 4.37, we would expect $\sigma_{E,\,\mathrm{KIPM}} \approx$ 120 eV

The collection efficiency observed at NEXUS is surprisingly low. Previous measurements using similar detectors have observed values of 7.0% [81]. Values as high as ~30% are expected to be achievable [47]. At Caltech, we have performed simulations that indicate phonon down conversion within niobium surface metal may be producing a large number of sub-gap phonons that our MKIDs are insensitive to. We plan to produce devices with less total niobium (thinner feedlines and no niobium MKIDs) to improve the collection efficiency. If $\eta_{ph} = 30\%$ could be achieved, it would improve our proposed resolution from ~120 to ~3 eV.

Another method to improve resolution is by lowering the first-stage-amplifier noise. As mentioned above, the HEMTs we use now have input noise temperature around 2-5 K. Novel kinetic-inductance-based traveling wave parametric amplifiers (TWPAs) can achieve input noise down to the standard quantum limit in our frequency range (200 mK) [82]. We can use Eq. 4.55 to confirm that a change in T_N from 2 K to 200 mK is expected to improve resolution by a factor of $\sqrt{10}$ (assuming amplifier noise domination can be maintained). Such an improvement would change the theoretical single-MKID KIPM resolution from ~3 to ~1 eV. We have previously operated one of our KIPM detectors with a kinetic-inductance TWPA at JPL and seen significant improvements. The results of that test will be published soon. Meanwhile, we are instrumenting our own dilution refrigerator to meet the operational requirements of the kinetic-inductance TWPA.

Aluminum MKIDs are relatively easy to fabricate at JPL and to test in our dilution refrigerator. The aluminum transition temperature ($T_c \approx 1.2$ K) is comfortably

below our refrigerator's base temperature (~50 mK with a detector payload). In a superior refrigerator (like what would be used in a full dark-matter search), we could improve resolution by using MKIDs made with lower T_c material. Hafnium is one of a few potentially suitable materials. The optical-MKID group at Santa Barbara has already fabricated and tested hafnium MKIDs with T_c of about 400 mK [83]. The lowering of T_c by a factor of 3 would lower the superconducting energy gap by the same factor. Using Eq. 4.55, we see that this should improve our resolution by a factor of 3 as well. Combined with the previously described improvements, this would result in a KIPM detector with ~0.3 eV resolution.

One potential issue going forward could be two-level system (TLS) noise. In some KIPM devices, we have begun observing TLS noise in the frequency direction of some MKIDs. This noise could limit the potential improvements we anticipate from increasing the R_Q parameter and lowering amplifier noise. We are therefore interested in studying device TLS noise to understand if anything can be done to minimize it. If TLS noise cannot be mitigated (as we suspect), we can instead focus on dissipation-direction measurements which remain unaffected. The dissipation resolution is expected to be a factor of ~2 worse (see Figure 4.37) and therefore could still achieve sub-eV resolution.

BIBLIOGRAPHY

[1] Yoshiaki Sofue and Vera Rubin. "Rotation Curves of Spiral Galaxies." In: *Annual Review of Astronomy and Astrophysics* 39.1 (Sept. 2001), pp. 137–174. DOI: 10.1146/annurev.astro.39.1.137. URL: https://doi.org/10.1146/annurev.astro.39.1.137.

[2] Pieter van Dokkum et al. "A Galaxy Lacking Dark Matter." In: *Nature* 555.7698 (Mar. 2018), pp. 629–632. DOI: 10.1038/nature25767. URL: https://doi.org/10.1038/nature25767.

[3] Ignacio Trujillo et al. "A Distance of 13 Mpc Resolves the Claimed Anomalies of the Galaxy Lacking Dark Matter." In: *Monthly Notices of the Royal Astronomical Society* 486.1 (Mar. 2019), pp. 1192–1219. DOI: 10.1093/mnras/stz771. URL: https://doi.org/10.1093/mnras/stz771.

[4] Heinz Andernach and Fritz Zwicky. *English and Spanish Translation of Zwicky's (1933) The Redshift of Extragalactic Nebulae.* 2017. DOI: 10.48550/ARXIV.1711.01693. URL: https://arxiv.org/abs/1711.01693.

[5] Steven W. Allen, August E. Evrard, and Adam B. Mantz. "Cosmological Parameters from Observations of Galaxy Clusters." In: *Annual Review of Astronomy and Astrophysics* 49.1 (Sept. 2011), pp. 409–470. DOI: 10.1146/annurev-astro-081710-102514. URL: https://doi.org/10.1146/annurev-astro-081710-102514.

[6] Douglas Clowe et al. "A Direct Empirical Proof of the Existence of Dark Matter." In: *The Astrophysical Journal* 648.2 (Aug. 2006), pp. L109–L113. DOI: 10.1086/508162. URL: https://doi.org/10.1086/508162.

[7] Scott W. Randall et al. "Constraints on the Self-Interaction Cross Section of Dark Matter from Numerical Simulations of the Merging Galaxy Cluster 1E 0657-56." In: *The Astrophysical Journal* 679.2 (June 2008), pp. 1173–1180. DOI: 10.1086/587859. URL: https://doi.org/10.1086/587859.

[8] Andrew Robertson, Richard Massey, and Vincent Eke. "What Does the Bullet Cluster Tell Us About Self-Interacting Dark Matter?" In: *Monthly Notices of the Royal Astronomical Society* 465.1 (Oct. 2016), pp. 569–587. DOI: 10.1093/mnras/stw2670. URL: https://doi.org/10.1093/mnras/stw2670.

[9] Planck Collaboration et al. "Planck 2015 Results - XIII. Cosmological parameters." In: *A&A* 594 (2016), A13. DOI: 10.1051/0004-6361/201525830. URL: https://doi.org/10.1051/0004-6361/201525830.

[10] Volker Springel et al. "Simulations of the Formation, Evolution and Clustering of Galaxies and Quasars." In: *Nature* 435.7042 (June 2005), pp. 629–636. DOI: 10.1038/nature03597. URL: https://doi.org/10.1038/nature03597.

[11] Leszek Roszkowski, Enrico Maria Sessolo, and Sebastian Trojanowski. "WIMP Dark Matter Candidates and Searches—Current Status and Future Prospects." In: *Reports on Progress in Physics* 81.6 (May 2018), p. 066201. DOI: 10.1088/1361-6633/aab913. URL: https://doi.org/10.1088/1361-6633/aab913.

[12] Gianfranco Bertone, Dan Hooper, and Joseph Silk. "Particle Dark Mmatter: Evidence, Candidates and Constraints." In: *Physics Reports* 405.5-6 (Jan. 2005), pp. 279–390. DOI: 10.1016/j.physrep.2004.08.031. URL: https://doi.org/10.1016/j.physrep.2004.08.031.

[13] Marc Kamionkowski. *WIMP and Axion Dark Matter.* 1997. DOI: 10.48550/ARXIV.HEP-PH/9710467. URL: https://arxiv.org/abs/hep-ph/9710467.

[14] Kim Griest and Marc Kamionkowski. "Unitarity Limits on the Mass and Radius of Dark-Matter Particles." In: *Phys. Rev. Lett.* 64 (6 Feb. 1990), pp. 615–618. DOI: 10.1103/PhysRevLett.64.615. URL: https://link.aps.org/doi/10.1103/PhysRevLett.64.615.

[15] Jonathan L. Feng. "Dark Matter Candidates from Particle Physics and Methods of Detection." In: *Annual Review of Astronomy and Astrophysics* 48.1 (Aug. 2010), pp. 495–545. DOI: 10.1146/annurev-astro-082708-101659. URL: https://doi.org/10.1146/annurev-astro-082708-101659.

[16] Tao Han et al. *WIMP Dark Matter at High Energy Muon Colliders –A White Paper for Snowmass 2021.* 2022. arXiv: 2203.07351 [hep-ph].

[17] Lars Bergström. "Dark Matter Candidates." In: *New Journal of Physics* 11.10 (Oct. 2009), p. 105006. DOI: 10.1088/1367-2630/11/10/105006. URL: https://doi.org/10.1088/1367-2630/11/10/105006.

[18] Luca Di Luzio et al. "The Landscape of QCD Axion Models." In: *Physics Reports* 870 (July 2020), pp. 1–117. DOI: 10.1016/j.physrep.2020.06.002. URL: https://doi.org/10.1016/j.physrep.2020.06.002.

[19] Marco Fabbrichesi, Emidio Gabrielli, and Gaia Lanfranchi. *The Physics of the Dark Photon.* Springer International Publishing, 2021. DOI: 10.1007/978-3-030-62519-1. URL: https://doi.org/10.1007/978-3-030-62519-1.

[20] Andreas Ringwald. *Axions and Axion-Like Particles.* 2014. DOI: 10.48550/ARXIV.1407.0546. URL: https://arxiv.org/abs/1407.0546.

[21] Jodi Cooley et al. *Report of the Topical Group on Particle Dark Matter for Snowmass 2021.* 2022. DOI: 10.48550/ARXIV.2209.07426. URL: https://arxiv.org/abs/2209.07426.

[22] Jeffrey Peter Filippini. "A Search for WIMP Dark Matter Using the First Five-Tower Run of the Cryogenic Dark Matter Search." PhD thesis. University of California, Berkeley, Jan. 2008. DOI: 10.2172/1415812. URL: https://www.osti.gov/biblio/1415812.

[23] Gerard Jungman, Marc Kamionkowski, and Kim Griest. "Supersymmetric Dark Matter." In: *Physics Reports* 267.5-6 (Mar. 1996), pp. 195–373. DOI: 10.1016/0370-1573(95)00058-5. URL: https://doi.org/10.1016/0370-1573(95)00058-5.

[24] Rouven Essig et al. *Direct Detection of Sub-GeV Dark Matter with Semiconductor Targets.* 2016. arXiv: 1509.01598 [hep-ph].

[25] Rouven Essig et al. *Snowmass2021 Cosmic Frontier: The Landscape of Low-Threshold Dark Matter Direct Detection in the Next Decade.* 2022. DOI: 10.48550/ARXIV.2203.08297. URL: https://arxiv.org/abs/2203.08297.

[26] D. W. Amaral et al. "Constraints on Low-Mass, Relic Dark Matter Candidates from a Surface-Operated SuperCDMS Single-Charge Sensitive Detector." In: *Physical Review D* 102.9 (Nov. 2020). ISSN: 2470-0029. DOI: 10.1103/physrevd.102.091101. URL: http://dx.doi.org/10.1103/PhysRevD.102.091101.

[27] T. Aralis et al. "Constraints on Dark Photons and Axionlike Particles from the SuperCDMS Soudan Experiment." In: *Physical Review D* 101.5 (Mar. 2020). ISSN: 2470-0029. DOI: 10.1103/physrevd.101.052008. URL: http://dx.doi.org/10.1103/PhysRevD.101.052008.

[28] Siné ad M. Griffin et al. "Multichannel Direct Detection of Light Dark Matter: Target Comparison." In: *Physical Review D* 101.5 (Mar. 2020). DOI: 10.1103/physrevd.101.055004. URL: https://doi.org/10.1103/physrevd.101.055004.

[29] Marc Schumann. *Dark Matter Search with Liquid Noble Gases.* 2012. arXiv: 1206.2169 [astro-ph.IM].

[30] R. Agnese et al. "Search for Low-Mass Weakly Interacting Massive Particles with SuperCDMS." In: *Physical Review Letters* 112.24 (June 2014). DOI: 10.1103/physrevlett.112.241302. URL: https://doi.org/10.1103/physrevlett.112.241302.

[31] B. S. Neganov and V. N. Trofimov. "Possibility of Producing a Bulky Supersensitive Thermal Detector at a Temperature Close to Absolute Zero." In: *Journal of Experimental and Theoretical Physics Letter (USSR) (Engl. Transl.); (United States)* 28:6 (Sept. 1978). URL: https://www.osti.gov/biblio/6230258.

[32] P. N. Luke. "Voltage-Assisted Calorimetric Ionization Detector." In: *Journal of Applied Physics* 64.12 (Dec. 1988), pp. 6858–6860. DOI: 10.1063/1.341976.

[33] SuperCDMS Collaboration et al. *A Strategy for Low-Mass Dark Matter Searches with Cryogenic Detectors in the SuperCDMS SNOLAB Facility.* 2023. arXiv: 2203.08463 [physics.ins-det].

[34] Rob Agnese et al. "Demonstration of Surface Electron Rejection with Interleaved Germanium Detectors for Dark Matter Searches." In: *Applied Physics Letters* 103 (May 2013). DOI: 10.1063/1.4826093.

[35] J. Billard, E. Figueroa-Feliciano, and L. Strigari. "Implication of Neutrino Backgrounds on the Reach of Next Generation Dark Matter Direct Detection Experiments." In: *Physical Review D* 89.2 (Jan. 2014). DOI: 10.1103/physrevd.89.023524. URL: https://doi.org/10.1103%2Fphysrevd.89.023524.

[36] Maurizio Giannotti et al. "Cool WISPs for Stellar Cooling Excesses." In: *Journal of Cosmology and Astroparticle Physics* 2016.05 (May 2016), pp. 057–057. DOI: 10.1088/1475-7516/2016/05/057. URL: https://doi.org/10.1088/1475-7516/2016/05/057.

[37] Q. Dong et al. "Ultra-Low Noise HEMTs for High-Impedance and Low-Frequency Preamplifiers: Realization and Characterization from 4.2 K to 77 K." In: *2014 11th International Workshop on Low Temperature Electronics (WOLTE).* 2014, pp. 21–24. DOI: 10.1109/WOLTE.2014.6881016.

[38] D. N. Seitz et al. "SuperCDMS Detector Readout Cryogenic Hardware." In: *The Thirteenth International Workshop on Low Temperature Detectors - LTD13*. Ed. by Betty Young, Blas Cabrera, and Aaron Miller. Vol. 1185. American Institute of Physics Conference Series. Dec. 2009, pp. 282–285. DOI: 10.1063/1.3292334.

[39] Thomas Alan Shutt. "A Dark Matter Detector Based on the Simultaneous Measurement of Phonons and Ionization at 20 mK." PhD thesis. University of California, Berkeley, Jan. 1993.

[40] R. Agnese et al. "First Dark Matter Constraints from a SuperCDMS Single-Charge Sensitive Detector." In: *Physical Review Letters* 121.5 (Aug. 2018). ISSN: 1079-7114. DOI: 10.1103/physrevlett.121.051301. URL: http://dx.doi.org/10.1103/PhysRevLett.121.051301.

[41] R. Ren et al. "Design and Characterization of a Phonon-Mediated Cryogenic Particle Detector with an eV-Scale Threshold and 100 keV-Scale Dynamic Range." In: *Physical Review D* 104.3 (Aug. 2021). DOI: 10.1103/physrevd.104.032010. URL: https://doi.org/10.1103/physrevd.104.032010.

[42] Matthew James Wilson. "A New Search for Low-mass Dark Matter and an Examination and Reduction of the Uncertainty due to the Photoelectric Absorption Cross Section using a Cryogenic Silicon Detector with Single-charge Sensitivity." PhD thesis. University of Toronto, June 2022. URL: http://hdl.handle.net/1807/123283.

[43] F. Ponce et al. "Modeling of Impact Ionization and Charge Trapping in SuperCDMS HVeV Detectors." In: *Journal of Low Temperature Physics* 199.3-4 (Feb. 2020), pp. 598–605. DOI: 10.1007/s10909-020-02349-x. URL: https://doi.org/10.1007/s10909-020-02349-x.

[44] F. Ponce et al. "Measuring the Impact Ionization and Charge Trapping Probabilities in SuperCDMS HVeV Phonon Sensing Detectors." In: *Phys. Rev. D* 101 (3 Feb. 2020), p. 031101. DOI: 10.1103/PhysRevD.101.031101. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.031101.

[45] M. F. Albakry et al. "Investigating the Sources of Low-Energy Events in a SuperCDMS-HVeV Detector." In: *Physical Review D* 105.11 (June 2022). DOI: 10.1103/physrevd.105.112006. URL: https://doi.org/10.1103/physrevd.105.112006.

[46] Liron Barak et al. "SENSEI: Direct-Detection Results on sub-GeV Dark Matter from a New Skipper CCD." In: *Physical Review Letters* 125.17 (Oct. 2020). DOI: 10.1103/physrevlett.125.171802. URL: https://doi.org/10.1103%2Fphysrevlett.125.171802.

[47] Noah Alexander Kurinsky. "The Low-Mass Limit: Dark Matter Detectors with eV-Scale Energy Resolution." PhD thesis. Jan. 2018. DOI: 10.2172/1472104. URL: https://www.osti.gov/biblio/1472104.

[48] D. Baxter et al. "Recommended Conventions for Reporting Results from Direct Dark Matter Searches." In: *The European Physical Journal C* 81.10 (Oct. 2021). DOI: 10.1140/epjc/s10052-021-09655-y. URL: https://doi.org/10.1140/epjc/s10052-021-09655-y.

[49] J.D. Lewin and P.F. Smith. "Review of Mathematics, Numerical Factors, and Corrections for Dark Matter Experiments Based on Elastic Nuclear Recoil." In: *Astroparticle Physics* 6.1 (1996), pp. 87–112. ISSN: 0927-6505. DOI: https://doi.org/10.1016/S0927-6505(96)00047-3. URL: https://www.sciencedirect.com/science/article/pii/S0927650596000473.

[50] Yonit Hochberg, Tongyan Lin, and Kathryn M. Zurek. "Absorption of Light Dark Matter in Semiconductors." In: *Physical Review D* 95.2 (Jan. 2017). DOI: 10.1103/physrevd.95.023013. URL: https://doi.org/10.1103/physrevd.95.023013.

[51] David F. Edwards. "- Silicon (Si)*". In: *Handbook of Optical Constants of Solids*. Ed. by Edward D. Palik. Burlington: Academic Press, 1997, pp. 547–569. ISBN: 978-0-12-544415-6. DOI: https://doi.org/10.1016/B978-

012544415-6.50027-3. URL: https://www.sciencedirect.com/science/article/pii/B9780125444156500273.

[52] B.L. Henke, E.M. Gullikson, and J.C. Davis. "X-Ray Interactions: Photoabsorption, Scattering, Transmission, and Reflection at E = 50-30,000 eV, Z = 1-92." In: *Atomic Data and Nuclear Data Tables.* 54.2 (1993), pp. 181–342. ISSN: 0092-640X. DOI: https://doi.org/10.1006/adnd.1993.1013. URL: https://www.sciencedirect.com/science/article/pii/S0092640X83710132.

[53] M. J. Berger et al. *XCOM: Photon Cross Section Database (version 1.5).* 2010. DOI: 10.18434/T48G6X. URL: http://physics.nist.gov/xcom (visited on 03/2019).

[54] C. Stanford et al. "Photoelectric Absorption Cross Section of Silicon Near the Bandgap from Room Temperature to Sub-Kelvin Temperature." In: *AIP Advances* 11.2 (Feb. 2021), p. 025120. ISSN: 2158-3226. DOI: 10.1063/5.0038392. eprint: https://pubs.aip.org/aip/adv/article-pdf/doi/10.1063/5.0038392/12991701/025120_1_online.pdf. URL: https://doi.org/10.1063/5.0038392.

[55] K. Ramanathan and N. Kurinsky. "Ionization Yield in Silicon for eV-Scale Electron-Recoil Processes." In: *Phys. Rev. D* 102 (6 Sept. 2020), p. 063026. DOI: 10.1103/PhysRevD.102.063026. URL: https://link.aps.org/doi/10.1103/PhysRevD.102.063026.

[56] Rupert G. Miller. *Simultaneous Statistical Inference.* Springer New York, NY, 1981. DOI: 10.1007/978-1-4613-8122-8. URL: https://doi.org/10.1007/978-1-4613-8122-8.

[57] E. Aprile et al. "Light Dark Matter Search with Ionization Signals in XENON1T." In: *Physical Review Letters* 123.25 (Dec. 2019). DOI: 10.1103/physrevlett.123.251801. URL: https://doi.org/10.1103/physrevlett.123.251801.

[58] A. Aguilar-Arevalo et al. "Constraints on Light Dark Matter Particles Interacting with Electrons from DAMIC at SNOLAB." In: *Physical Review Letters* 123.18 (Oct. 2019). DOI: 10.1103/physrevlett.123.181802. URL: https://doi.org/10.1103/physrevlett.123.181802.

[59] Q. Arnaud et al. "First Germanium-Based Constraints on Sub-MeV Dark Matter with the EDELWEISS Experiment." In: *Physical Review Letters* 125.14 (Oct. 2020). DOI: 10.1103/physrevlett.125.141301. URL: https://doi.org/10.1103/physrevlett.125.141301.

[60] Sunil R. Golwala et al. "Status of MUSIC, the MUltiwavelength Sub/millimeter Inductance Camera." In: *SPIE Proceedings*. Ed. by Wayne S. Holland. SPIE, Sept. 2012. DOI: 10.1117/12.926055. URL: https://doi.org/10.1117/12.926055.

[61] V. Asboth et al. "HerMES: a Search for High-Redshift Dusty Galaxies in the HerMES Large Mode Survey – Catalogue, Number Counts and Early Results." In: *Monthly Notices of the Royal Astronomical Society* 462.2 (July 2016), pp. 1989–2000. ISSN: 1365-2966. DOI: 10.1093/mnras/stw1769. URL: http://dx.doi.org/10.1093/mnras/stw1769.

[62] Jack Sayers et al. "Peculiar Velocity Constraints from Five-Band SZ Effect Measurements Towards RX J1347.5-1145 with MUSIC and Bolocam from the CSO." In: *The Astrophysical Journal* 820.2 (Mar. 2016), p. 101. ISSN: 1538-4357. DOI: 10.3847/0004-637x/820/2/101. URL: http://dx.doi.org/10.3847/0004-637X/820/2/101.

[63] Noah Swimmer et al. "SCExAO and Keck Direct Imaging Discovery of a Low-mass Companion Around the Accelerating F5 Star HIP 5319." In: *The Astronomical Journal* 164.4, 152 (Oct. 2022), p. 152. DOI: 10.3847/1538-3881/ac85a8. arXiv: 2208.00334 [astro-ph.SR].

[64] Alexander B. Walter et al. "The MKID Exoplanet Camera for Subaru SCExAO." In: *Publications of the Astronomical Society of the Pacific* 132.1018 (Nov. 2020), p. 125005. DOI: 10.1088/1538-3873/abc60f. URL: https://doi.org/10.1088/1538-3873/abc60f.

[65] Jiansong Gao. "The Physics of Superconducting Microwave Resonators." PhD thesis. California Institute of Technology, June 2008. DOI: 10.7907/RAT0-VM75. URL: https://resolver.caltech.edu/CaltechETD:etd-06092008-235549.

[66] Seth Siegel. "A Multiwavelength Study of the Intracluster Medium and the Characterization of the Multiwavelength Sub/millimeter Inductance Camera." PhD thesis. California Institute of Technology, July 2015. DOI: 10.7907/Z9Z31WJ7. URL: https://resolver.caltech.edu/CaltechTHESIS:10212015-211417853.

[67] Sunil Ramanlal Golwala. "Exclusion Limits on the WIMP Nucleon Elastic Scattering Cross-Section from the Cryogenic Dark Matter Search." PhD thesis. Jan. 2000. DOI: 10.2172/1421437. URL: https://www.osti.gov/biblio/1421437.

[68] David Craig Moore. "Dark Matter Direct Detection, Low-Mass WIMPs, Superconducting Particle Detectors, Microwave Kinetic Inductance Detectors." PhD thesis. California Institute of Technology, May 2012. DOI: 10.7907/X8JD-4R90. URL: https://resolver.caltech.edu/CaltechTHESIS:05172012-142723949.

[69] Yen-Yung Chang. "Dark Matter; SuperCDMS HVeV; Phonon-mediated Particle Detector With Kinetic Inductance Detector; Cryogenic Blackbody Radiation." PhD thesis. California Institute of Technology, May 2023. DOI: 10.7907/jswx-ad33. URL: https://resolver.caltech.edu/CaltechTHESIS:05212023-152144355.

[70] Brett Durcan Cornell. "A Dark Matter Search Using the Final SuperCDMS Soudan Dataset and the Development of a Large-Format, Highly-Multiplexed, Athermal-Phonon-Mediated Particle Detector." PhD thesis. California Institute of Technology, May 2018. DOI: 10.7907/M297-J716. URL: https://resolver.caltech.edu/CaltechTHESIS:06082018-171453167.

[71] S. Verma et al. "Low-Threshold Sapphire Detector for Rare Event Searches." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1046 (Jan. 2023), p. 167634. DOI: 10.1016/j.nima.2022.167634. URL: https://doi.org/10.1016/j.nima.2022.167634.

[72] Hosaeng Kim and Rhonda Franklin-Drayton. "Wire-Bond Free Technique for Right-Angle Coplanar Waveguide Bend Structures." In: *IEEE Transactions on Microwave Theory and Techniques* 57.2 (2009), pp. 442–448. DOI: 10.1109/TMTT.2008.2011207.

[73] M. S. Khalil et al. "An Analysis Method for Asymmetric Resonator Transmission Applied to Superconducting Devices." In: *Journal of Applied Physics* 111.5 (Mar. 2012). DOI: 10.1063/1.3692073. URL: https://doi.org/10.1063/1.3692073.

[74] P. J. de Visser et al. "Evidence of a Nonequilibrium Distribution of Quasiparticles in the Microwave Response of a Superconducting Aluminum Resonator." In: *Physical Review Letters* 112.4 (Jan. 2014). ISSN: 1079-7114. DOI: 10.1103/physrevlett.112.047004. URL: http://dx.doi.org/10.1103/PhysRevLett.112.047004.

[75] Jiachen Gao et al. "Equivalence of the Effects on the Complex Conductivity of Superconductor due to Temperature Change and External Pair Breaking." In: *Journal of Low Temperature Physics* 151 (Apr. 2008). DOI: 10.1007/s10909-007-9688-z.

[76] D. Flanigan et al. "Magnetic Field Dependence of the Internal Quality Factor and Noise Performance of Lumped-Element Kinetic Inductance Detectors." In: *Applied Physics Letters* 109.14 (Oct. 2016), p. 143503. DOI: 10.1063/1.4964119. URL: https://doi.org/10.1063/1.4964119.

[77] Lorenzo Minutolo et al. "A Flexible GPU-Accelerated Radio-frequency Readout for Superconducting Detectors." In: *IEEE Transactions on Applied Superconductivity* 29.5 (Aug. 2019), pp. 1–5. DOI: 10.1109/tasc.2019.2912027. URL: https://doi.org/10.1109/tasc.2019.2912027.

[78] O. Wen et al. "Performance of a Phonon-Mediated Detector Using KIDs Optimized for Sub-GeV Dark Matter." In: *Journal of Low Temperature Physics* 209.3-4 (July 2022), pp. 510–517. DOI: 10.1007/s10909-022-02764-2. URL: https://doi.org/10.1007/s10909-022-02764-2.

[79] Karthik Ramanathan et al. “Identifying Drivers of Energy Resolution Variation in a Multi-KID Phonon-Mediated Detector.” In: *Journal of Low Temperature Physics* 209.3-4 (June 2022), pp. 457–463. DOI: 10.1007/s10909-022-02753-5. URL: https://doi.org/10.1007/s10909-022-02753-5.

[80] Dylan J. Temples et al. “Performance of a Phonon-Mediated MKID Device at the NEXUS Low-Background Facility.” In: (Sept. 2023). DOI: 10.2172/2005196. URL: https://www.osti.gov/biblio/2005196.

[81] D. C. Moore et al. “Position and Energy-Resolved Particle Detection using Phonon-Mediated Microwave Kinetic Inductance Detectors.” In: *Applied Physics Letters* 100.23 (June 2012). DOI: 10.1063/1.4726279. URL: https://doi.org/10.1063%2F1.4726279.

[82] Nikita Klimovich et al. *Demonstration of a Quantum Noise Limited Traveling-Wave Parametric Amplifier.* 2023. arXiv: 2306.11028 [quant-ph].

[83] Nicholas Zobrist et al. “Design and Performance of Hafnium Optical and Near-IR Kinetic Inductance Detectors.” In: *Applied Physics Letters* 115.21 (Nov. 2019). DOI: 10.1063/1.5127768. URL: https://doi.org/10.1063%2F1.5127768.

[84] P. J. de Visser et al. “Fluctuations in the Electron System of a Superconductor Exposed to a Photon Flux.” In: *Nature Communications* 5.1 (Feb. 2014). DOI: 10.1038/ncomms4130. URL: https://doi.org/10.1038%2Fncomms4130.

Appendix A

FURTHER SNOLAB HEMT RESULTS

In this appendix, I have collected scatter plots of various measured SNOLAB-HEMT parameters. Measurement of these parameters took place from February 2020 to June 2022, but no significant trend can be seen in the scatter plots. In this way, the plots support the conclusion that the testing setup and HEMTs did not suffer any unnoticed changes during the test period. Additionally, there do not appear to be distinct distributions that could indicate differences between different HEMT batches.

A description of the measured parameters can be found in Section 2.4.5. A description of the measurement methods can be found in Section 2.5.

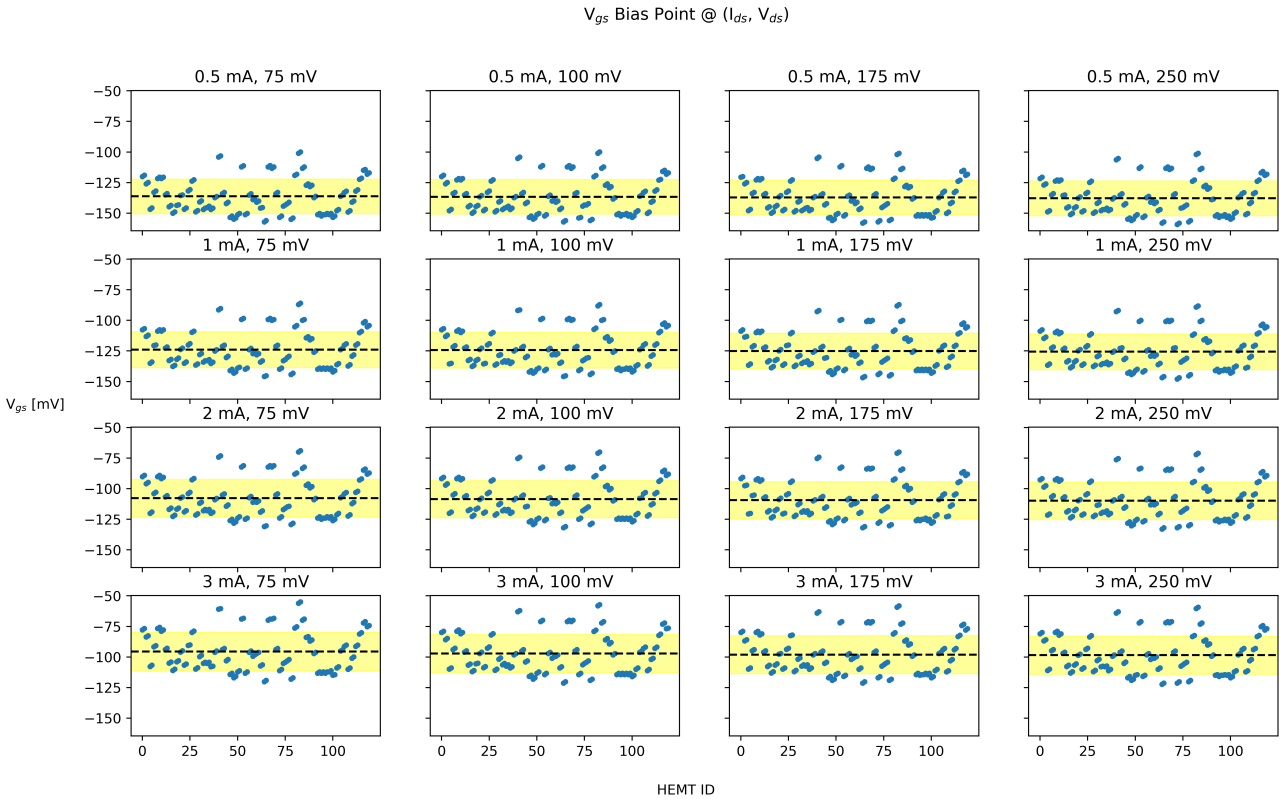


Figure A.1: Scatter plots displaying the gate voltages required to produce each bias point in the SNOLAB HEMTs, in the order the HEMTs were tested. No significant trend can be seen over the 2 year testing period.

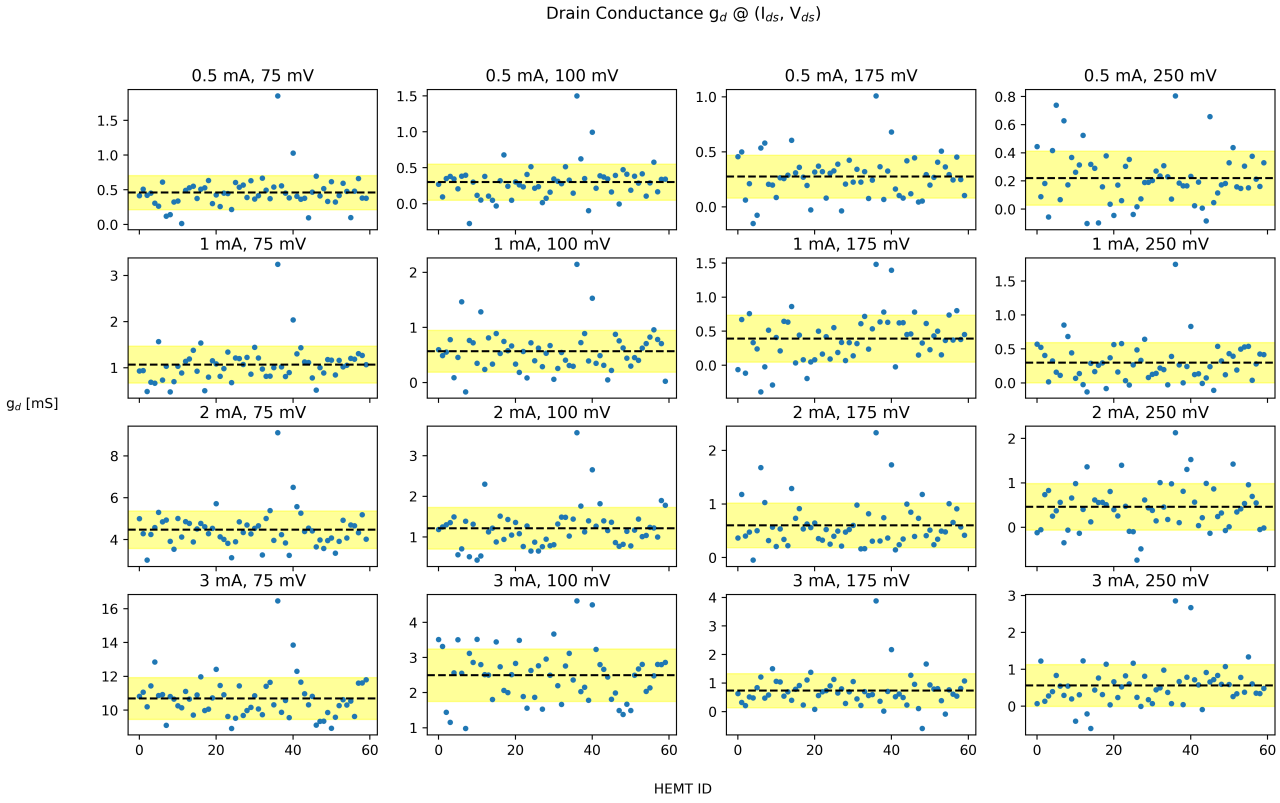


Figure A.2: Scatter plots displaying the SNOLAB HEMT drain conductances, in the order the HEMTs were tested. No significant trend can be seen over the 2 year testing period.

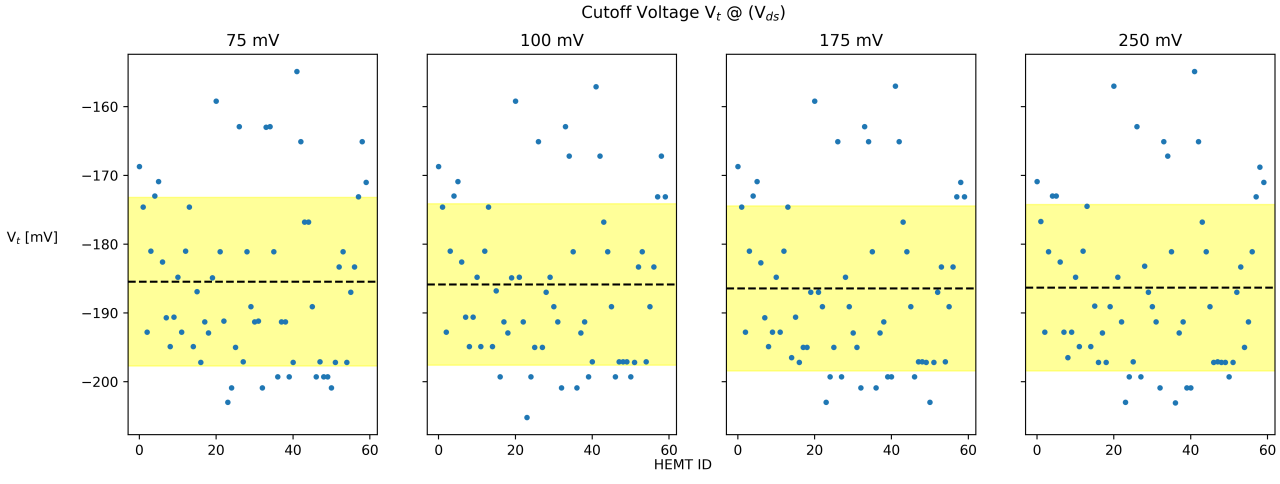


Figure A.3: Scatter plots displaying the SNOLAB HEMT cutoff voltages for each potential bias drain voltage, in the order the HEMTs were tested. No significant trend can be seen over the 2 year testing period.

Appendix B

RADIO-FREQUENCY READOUT BASICS

The basic unit of a traditional high-frequency readout is the frequency mixer. These devices take two analog input signals and multiply them together to form an output signal. If the inputs have different carrier frequencies, the output signal will be carried by two new frequencies: the sum and difference of those of the inputs (Eq. B.1)

$$A \sin(\omega_1 t) \cdot B \sin(\omega_2 t) = \frac{AB}{2} \cos((\omega_1 - \omega_2) t) - \frac{AB}{2} \cos((\omega_1 + \omega_2) t) \quad \text{(B.1)}$$

Often, one input frequency is significantly higher than the other, and is used to shift the lower-frequency input into a higher-frequency band. This is referred to as up-mixing. If two high-frequency inputs are used and the difference in their frequencies is desired, this is referred to as down-mixing. Unwanted signal frequencies can be attenuated using high- or low-pass filters. Many readouts use up- and down-mixing in equal parts. A generated signal is up-mixed to the high-frequency band of interest. The signal is used to probe a high-frequency device such as an MKID or antenna. Down-mixing can then be used to return the signal to a frequency-range that is easily recorded via ADC. To minimize noise in such readouts, one should use the same high-frequency source for both mixers. This source is normally referred to as the local oscillator (LO). By using a single LO, one avoids multiplying-in two independent noise sources during up- and down-mixing. Mixers are defined as unbalanced, single-balanced, or double-balanced. Unbalanced mixers allow both input signals to enter the mixer output (in addition to the multiplied signal). Single-balanced mixers remove one of the inputs from the output and double-balanced mixers remove both.

Frequency mixers are key components of IQ mixers (In-phase Quadrature mixers). For a down-mixing IQ mixer, two double-balanced frequency mixers receive the same radio-frequency (RF) signal as one input. One mixer's second input is the LO, while the other's is the LO shifted by 90° ($\pi/2$). The resultant mixer outputs are labeled as in-phase (V_i) and quadrature-phase (V_q), and are 90° out of phase with each other.

$$\begin{aligned} V_{RF} &\equiv A\,(t)\sin\left(\omega_{RF}t+\phi\,(t)\right) \\ V_{LO} &\equiv \sin\left(\omega_{LO}t\right) \\ V_{LO+90^\circ} &= \cos\left(\omega_{LO}t\right) \end{aligned} \tag{B.2}$$

$$\begin{aligned} V_i &= V_{RF}\cdot V_{LO} \\ &= A\sin\left(\omega_{LO}t\right)\sin\left(\omega_{RF}t+\phi\right) \\ &= \frac{A}{2}\left[\cos\left(\left(\omega_{LO}+\omega_{RF}\right)t+\phi\right)-\cos\left(\left(\omega_{LO}-\omega_{RF}\right)t-\phi\right)\right] \end{aligned} \tag{B.3}$$

$$\begin{aligned} V_q &= V_{RF}\cdot V_{LO+90^\circ} \\ &= A\cos\left(\omega_{LO}t\right)\sin\left(\omega_{RF}t+\phi\right) \\ &= \frac{A}{2}\left[\sin\left(\left(\omega_{LO}+\omega_{RF}\right)t+\phi\right)-\sin\left(\left(\omega_{LO}-\omega_{RF}\right)t-\phi\right)\right] \end{aligned} \tag{B.4}$$

If the LO frequency is chosen to match the RF-carrier frequency, the IQ mixer outputs will include DC-offsets proportional to the in-phase and quadrature components of the initial RF signal (Eq. B.5). The non-DC components will have twice the LO frequency and can be removed using low-pass filtering (Eq. B.6). In this way, an IQ mixer can demodulate an RF signal directly into an audio-band result carrying the time-dependence of the signal modulation. The resulting values are used as-is or to calculate the amplitude and phase of the RF signal as a function of time (Eq. B.7). By varying the LO frequency, one can use a single IQ-mixer setup to measure the amplitude and phase of many frequency components of the initial signal.

$$\omega_{LO}=\omega_{RF}\rightarrow\begin{cases} V_i=\frac{A}{2}\left[\cos\left(2\omega_{LO}t+\phi\right)-\cos\left(-\phi\right)\right] \\ V_q=\frac{A}{2}\left[\sin\left(2\omega_{LO}t+\phi\right)-\sin\left(-\phi\right)\right] \end{cases} \tag{B.5}$$

$$\text{low-pass filter}\rightarrow\begin{cases} LPF\left(V_i\right)=-\frac{A}{2}\cos\left(\phi\right) \\ LPF\left(V_q\right)=\frac{A}{2}\sin\left(\phi\right) \end{cases} \tag{B.6}$$

$$\begin{aligned} A\,(t)^2 &= 4\left[LPF\left(V_i\right)^2+LPF\left(V_q\right)^2\right] \\ \phi\,(t) &= -\arctan\left(\frac{LPF\left(V_q\right)}{LPF\left(V_i\right)}\right) \end{aligned} \tag{B.7}$$

For simplicity, I assumed lossless signal-splitting, phase-shifting, mixing, and low-pass filtering in the equations above.

IQ mixers can also be used in the reverse of the described procedure to instead produce a modulated RF signal. In this case, the user inputs a pair of DC values which are up-mixed with the LO and combined (90° phase shifted) to produce the desired signal.

Appendix C

TRANSMISSION NOISE FROM A RANDOM SOURCE

To understand how noise in S_{21} is defined when the dominant noise source has completely random phase, we must define S_{21} in terms of complex voltages.

$$S_{21} = \frac{v_{out}}{v_{in}} \tag{C.1}$$

The complex input voltage comes from our readout system. Let us consider it to be at a single frequency (ω).

$$v_{in} = V_{in,0} e^{j\omega t} \tag{C.2}$$

The complex output voltage is dominated by random noise (such as that from the input to an amplifier). We will assume its complex value is constant in amplitude with a random phase ($\phi(t)$).

$$v_{out} = V_{out,0} e^{j\phi(t)} \tag{C.3}$$

The actual measurable voltage is equivalent to the real part of each complex value. We can then relate the measurable RMS voltage to the amplitude of the complex voltage. This is trivial for the input voltage since the RMS of $\cos(\omega t)$ is just $1/2$.

$$\begin{aligned} \langle V_{in}^2 \rangle &= \langle Re\,\{v_{in}\}^2 \rangle \\ &= V_{in,0}^2 \langle \cos^2(\omega t) \rangle \\ &= \frac{1}{2} V_{in,0}^2 \end{aligned} \tag{C.4}$$

The result is actually identical for the output voltage, in the case that its phase is random.

$$\begin{aligned}\langle V_{out}^2 \rangle &= \langle Re\,\{v_{out}\}^2 \rangle \\ &= V_{out,0}^2 \langle \cos^2(\phi(t)) \rangle \\ &= \frac{1}{2} V_{out,0}^2 \langle 1 + \cos(2\phi(t)) \rangle \\ &= \frac{1}{2} V_{out,0}^2\end{aligned} \tag{C.5}$$

The average over $\cos(\phi(t))$ goes to zero because $\cos(x)$ is odd about $x = \pi/2$ and the random phase $\phi(t)$ will be evenly distributed about $\pi/2$ (or any other value).

We can now look for the RMS of the real part of S_{21}.

$$Re\,\{S_{21}\} = \frac{V_{in,0}}{V_{out,0}} \cos(\phi(t) - \omega t) \tag{C.6}$$

$$\begin{aligned}\langle Re\,\{S_{21}\}^2 \rangle &= \frac{V_{in,0}^2}{V_{out,0}^2} \langle \cos^2(\phi(t) - \omega t) \rangle \\ &= \frac{1}{2} \frac{V_{in,0}^2}{V_{out,0}^2} \langle 1 + \cos(2\phi(t) - 2\omega t) \rangle \\ &= \frac{1}{2} \frac{V_{in,0}^2}{V_{out,0}^2} \\ &= \frac{1}{2} \frac{\langle V_{out}^2 \rangle}{\langle V_{in}^2 \rangle}\end{aligned} \tag{C.7}$$

Here the average over cosine has gone to zero for the same reasons as above. We only have to acknowledge that multiplying by 2 and shifting the randomly distributed $\phi(t)$ by ωt has no effect. In the last equivalency of Eq. C.7, I used Eqs. C.4 and C.5.

The result is the same if we had considered the imaginary part of S_{21}.

$$\langle Im\,\{S_{21}\}^2 \rangle = \frac{1}{2} \frac{\langle V_{out}^2 \rangle}{\langle V_{in}^2 \rangle} \tag{C.8}$$

Both results show that the noise power (proportional to V^2) is evenly distributed between the real and imaginary parts of S_{21}. This was the only reasonable conclusion for a random noise source since it will have no preference for being in- or out-of phase with the (completely independent) input signal.

Appendix D

MKID SIGNAL SHAPES AND PARAMETER EXTRACTION

The response in quasiparticle density to a non-equilibrium initial value is derived in Section 4.4.1 of [69]. The response depends on the initial density ($n_{qp,0}$), the equilibrium density ($n_{qp,\infty}$), and the quasiparticle recombination constant (R).

$$n_{qp}(t) = \left(\frac{1+\chi}{1-\chi}\right) n_{qp,\infty} \tag{D.1}$$

$$\chi(t) = \left(\frac{n_{qp,0} - n_{qp,\infty}}{n_{qp,0} + n_{qp,\infty}}\right) e^{-2Rn_{qp,\infty}t} \tag{D.2}$$

We can rewrite Eq. D.1 to provide δn_{qp} which is what we actually measure in our MKIDs.

$$\begin{aligned} \delta n_{qp}(t) &= n_{qp} - n_{qp,\infty} \\ &= \left(\frac{2\chi}{1-\chi}\right) n_{qp,\infty} \end{aligned} \tag{D.3}$$

Eq. D.3 can then be directly fit to the signals found in Section 4.5.1. Extracting $n_{qp,\infty}$ allows us to convert our quasiparticle signals to absolute units (n_{qp} rather than δn_{qp}). Extracting R allows us to compare with typical values in aluminum and among different MKIDs. Example fits using Eq. D.3 can be seen in Figure D.1. The results describe the data well, but give suspiciously large R. From [84], one can extract an expected R value of about $10\,\mu m^3/s$. It is possible that our large R values are the result of inaccuracy in the conversion from δS_{21} to δn_{qp}. It is also possible that the full fit is failing to disentangle the R and $n_{qp,\infty}$ parameters. We will discuss another method to attempt to extract R below.

Also discussed in [69] are the limits of Eq. D.1 for very large and very small signals. In this context, small signals are defined as having $n_{qp,0} - n_{qp,\infty} \ll n_{qp,\infty}$. For small signals, the quasiparticle response approaches a basic exponential as shown in Eq. D.4.

$$\delta n_{qp}(t) \approx (n_{qp,0} - n_{qp,\infty}) e^{-2Rn_{qp,\infty}t}, \quad n_{qp,0} - n_{qp,\infty} \ll n_{qp,\infty} \tag{D.4}$$

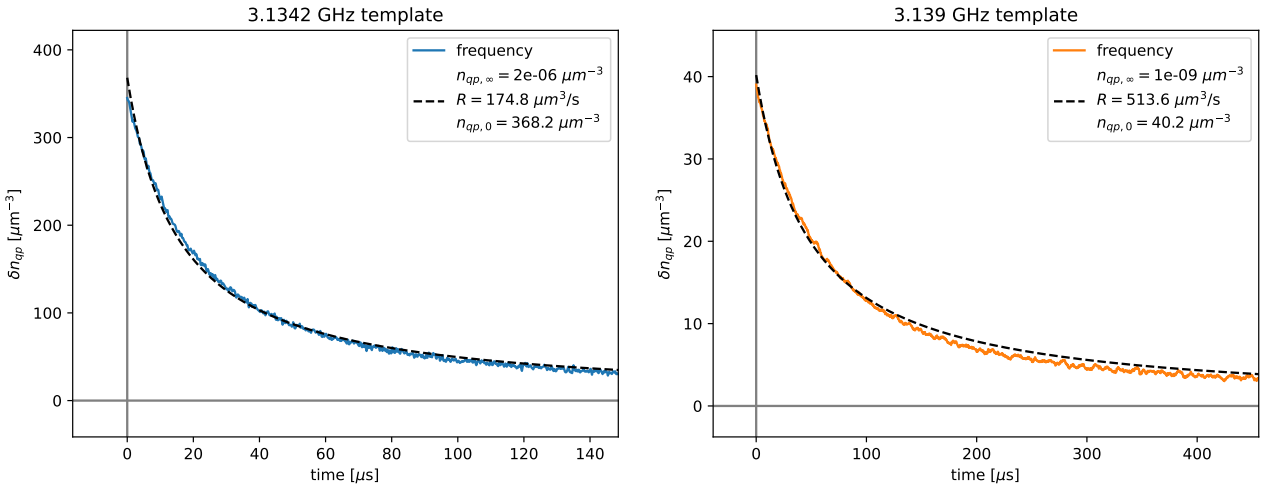


Figure D.1: Example fitting of Eq. D.3 to the frequency-direction templates calculated and displayed in Figure 4.24. Note that the "large signal" regions from Figure 4.24 are excluded from the templates (and plots and fits). (Left) Fit to the pulsed MKID's calculated template. (Right) Fit to a responding MKID's calculated template. The extracted R values are appear significantly different.

We see that the small-signal time constant is $\tau_{qp} = 1/2Rn_{qp,\infty}$. It is beneficial to lower the quiescent quasiparticle density if a longer quasiparticle lifetime is desired in small-signal events. It is also notable that using Eq. D.4 to fit a small signal will yield constraints on τ_{qp} but will not necessarily allow one to separate R from $n_{qp,\infty}$.

The results in Figure D.1 suggest that $n_{qp,0} \gg n_{qp,\infty}$. In this case, the large signal approximation from [69] is more appropriate. With $n_{qp,0} \gg n_{qp,\infty}$, the prefactor in Eq. D.2 goes to unity and $\chi(t) \rightarrow e^{-2Rn_{qp,\infty}t}$. This assumption alone does not hugely simplify Eq. D.3. We can do better by rewriting the equation as shown below.

$$\delta n_{qp}(t) = \left(\frac{2}{\frac{1}{\chi} - 1}\right) n_{qp,\infty} \tag{D.5}$$

We can then expand $1/\chi$ under the dual assumptions of $n_{qp,0} \gg n_{qp,\infty}$ and $2Rn_{qp,\infty}t \ll 1$. The latter condition can be attained in data by considering only the initial response in time ($t \ll 1/2Rn_{qp,\infty}$)

$$\frac{1}{\chi}(t) = \left(\frac{1+\frac{n_{qp,\infty}}{n_{qp,0}}}{1-\frac{n_{qp,\infty}}{n_{qp,0}}}\right) e^{2Rn_{qp,\infty}t} \tag{D.6}$$

$$\approx \left(1+2\frac{n_{qp,\infty}}{n_{qp,0}}\right)\left(1+2Rn_{qp,\infty}t\right) \tag{D.7}$$

$$\approx 1+2\frac{n_{qp,\infty}}{n_{qp,0}}+2Rn_{qp,\infty}t \tag{D.8}$$

When plugged back into Eq. 4.2.3, this gives the approximate form.

$$\delta n_{qp}(t) \approx \frac{1}{\frac{1}{n_{qp,0}}+Rt} \tag{D.9}$$

This result is analogous to equation A.28 in [69]. The addition of the $1/n_{qp,0}$ term gives the correct behaviour as $t \to 0$. With this formulation, it becomes possible to extract R by simply fitting a line to the inverse of δn_{qp}. The results of doing so for the same data in Figure D.1 can be seen in Figure D.2. To ensure the time assumption was satisfied, I only fit to the first 25 µs of data.

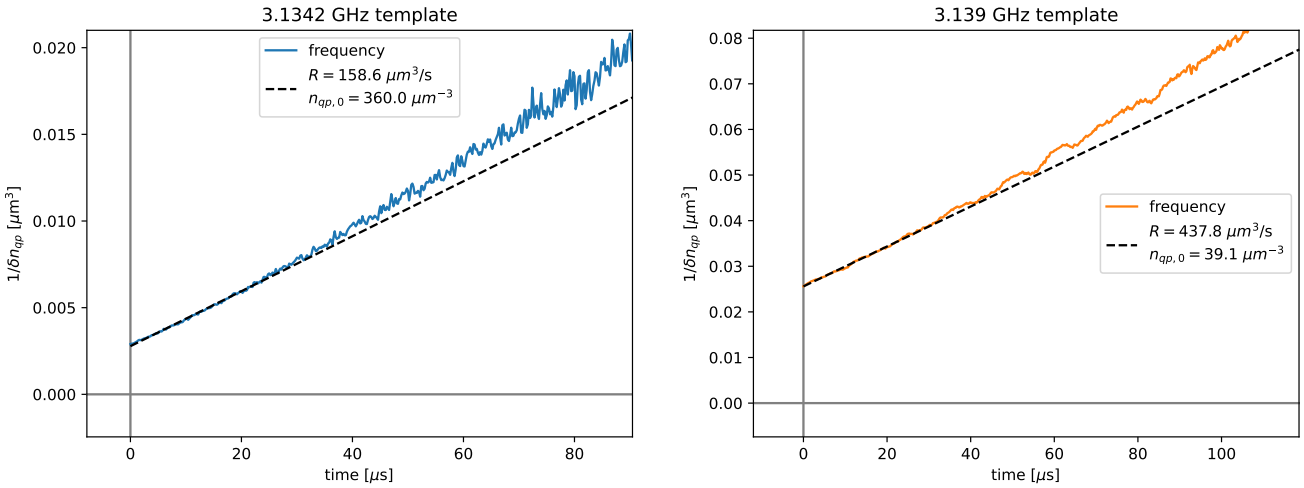


Figure D.2: Example linear fits to the inverted frequency-direction templates calculated and displayed in Figure 4.24. The fits only considered the first 25 µs of data. The extracted R values are consistent with those of Figure D.1. (Left) Fit to the pulsed MKID's inverted template. (Right) Fit to a responding MKID's inverted template.

We see the results are similar to those of the full fit. The inverse fitting method may be more robust in extracting R since it does not depend on simultaneously fitting $n_{qp,\infty}$.

Instead, the user must ensure the data satisfies the above discussed assumptions (large signal in a short time period). Since our R values remain suspiciously high, perhaps we are not truly satisfying the $n_{qp,0} \gg n_{qp,\infty}$ assumption. With this method, an insufficiently large signal would lead to a larger R than realistic. As an example, imagine fitting a line to the data in Figure D.2 for times between 60 and 85 µs. For those times, the slope (and R) would appear even larger. If signal size is the issue, we would need a signal larger than the small-signal condition permits to effectively extract R.

INDEX